



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO

FACULTAD DE INFORMÁTICA Y ELECTRÓNICA

ESCUELA DE INGENIERÍA EN SISTEMAS

**“ESTUDIO COMPARATIVO DE HERRAMIENTAS OPEN SOURCE PARA ANÁLISIS
MULTIDIMENSIONAL CASO PRÁCTICO: PROASETEL S.A., ANÁLISIS MULTIDIMENSIONAL
DEL RUB-ECUADOR”**

TESIS DE GRADO

Previa la obtención del título de:

INGENIERO EN SISTEMAS INFORMÁTICOS

Presentado por:

JUAN CARLOS ANASICHA TADAY

JORGE RUBÉN MOYÓN MOYÓN

RIOBAMBA – ECUADOR

2009

Nuestro agradecimiento a nuestra directora, Ingeniera Ivonne Rodríguez por su tiempo, paciencia, consejos y apoyo durante la realización de este proyecto de tesis. Muchas gracias.

Al presidente del tribunal de nuestra tesis, Ingeniero Iván Ménes Camejo por su tiempo, atención y valiosas sugerencias durante la revisión de esta tesis.

A todos los que contribuyeron de una u otra forma apoyándonos desinteresadamente durante el desarrollo de esta tesis de grado, especialmente al Ingeniero Daniel Murray un contacto de la web quién nos asesoró en algunos parámetros utilizados en la realización de esta tesis.

Dedico el presente trabajo de tesis a mi familia, en especial a mis padres María y Manuel, mis hermanos y hermanas quienes siempre me manifestaron su apoyo incondicional y con su ejemplo supieron inspirar en mí un espíritu de superación constante.

Juan Carlos Anasicha Taday.

Esta tesis va dedicada de manera especial a mi madre Maria por su apoyo, consejos y comprension que me inspiraron a superarme constantemente a mi padre Jacinto y hermanos por brindarme su apoyo absoluto e incondicional formando siempre en mí la clase de persona que aspiro ser.

Jorge Rubén Moyón Moyón.

FIRMAS DE RESPONSABILIDADES

NOMBRE	FIRMA	FECHA
Dr. Romeo Rodríguez DECANO FACULTAD DE INFORMATICA Y ELECTRONICA
Ing. Iván Menes DIRECTOR DE ESCUELA DE INGENIERIA EN SISTEMAS
Ing. Ivonne Rodríguez DIRECTORA DE TESIS
Ing. Iván Menes MIEMBRO DEL TRIBUNAL
Lcdo. Carlos Rodríguez DIR. CENTRO DE DOCUMENTACIÓN
NOTA DE LA TESIS	

“Nosotros, JUAN CARLOS ANASICHA TADAY Y JORGE RUBÉN MOYÓN MOYÓN somos responsables de las ideas, doctrinas y resultados expuestos en esta tesis; y, el patrimonio intelectual de la Tesis de Grado pertenece a la ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO”.

Juan Carlos Anasicha Taday

Jorge Rubén Moyón Moyón

INDICE GENERAL

PORTADA

AGRADECIMIENTO

DEDICATORIA

ÍNDICE GENERAL

ÍNDICE DE ABREVIATURAS

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

INTRODUCCIÓN

CAPÍTULO I

1. MARCO REFERENCIAL

1.1	PROBLEMATIZACIÓN	- 18 -
1.2	JUSTIFICACIÓN.....	- 19 -
1.3	OBJETIVOS	- 20 -
1.3.1	<i>Objetivo General</i>	- 20 -
1.3.2	<i>Objetivos Específicos</i>	- 20 -
1.4	HIPÓTESIS	- 21 -
1.5	MÉTODOS Y TÉCNICAS	- 21 -
1.5.1	<i>Métodos</i>	- 21 -
1.5.1.1	Inductivo – Deductivo.....	- 21 -
1.5.1.2	Método Comparativo	- 21 -
1.5.1.3	Científico.....	- 21 -
1.5.1.4	Experimental	- 21 -
1.5.2	<i>Técnicas</i>	- 22 -
1.5.2.1	Fuentes	- 22 -

CAPÍTULO II

2. MARCO TEÓRICO

2.1	BUSINESS INTELLIGENCE (BI)	- 23 -
2.2	ARQUITECTURA DE BUSINESS INTELLIGENCE.....	- 24 -
2.2.1	<i>Sistemas Fuentes</i>	- 25 -
2.2.2	<i>Capa DATA WAREHOUSE</i>	- 26 -
2.2.3	<i>Capa BI</i>	- 26 -
2.3	ANÁLISIS OLAP.....	- 26 -
2.3.1	<i>Características Principales de las Tecnologías OLAP</i>	- 27 -
2.3.2	<i>Tecnologías OLTP vs. OLAP</i>	- 28 -
2.3.3	<i>Conceptos y Componentes</i>	- 30 -
2.3.3.1	Cubo	- 30 -
2.3.3.2	Medida	- 30 -
2.3.3.3	Dimensión.....	- 31 -
2.3.3.4	Nivel.....	- 31 -
2.3.4	<i>Operaciones con OLAP</i>	- 31 -

2.3.5	Arquitectura OLAP	- 33 -
	Figura II.5. Arquitectura básica OLAP	- 33 -
2.3.6	El Modelo de Datos OLAP	- 34 -
2.3.6.1	Modelo o esquema de Estrella (“star schema”) tradicional.....	- 35 -
2.3.6.2	Modelo Copo de Nieve (“snow flake schema”).....	- 36 -
2.3.6.3	Modelo Constelación.....	- 36 -
2.3.7	Implementación del OLAP	- 37 -
2.3.7.1	Sistemas MOLAP.....	- 38 -
2.3.7.2	Sistemas ROLAP	- 39 -
2.3.7.3	ROLAP vs. MOLAP	- 40 -
2.3.7.4	Sistemas HOLAP (Sistema Híbrido).....	- 41 -
2.3.8	El Entorno OLAP en la Actualidad	- 41 -
2.3.8.1	OLAP y el código propietario	- 41 -
2.3.8.2	OLAP y el Open Source	- 43 -
2.4	OPERADORES OLAP	- 43 -
2.4.1	Características de las Herramientas OLAP	- 45 -
2.4.2	Componentes de las Herramientas OLAP	- 45 -

CAPÍTULO III

3. HERRAMIENTAS OLAP OPEN SOURCE

3.1	JUSTIFICACIÓN	- 48 -
3.2	PALO 2.5	- 49 -
3.2.1	<i>Antecedentes y Visión General</i>	- 49 -
3.2.2	<i>Arquitectura de Palo 2.5</i>	- 49 -
3.2.3	<i>Instalación de Palo.</i>	- 50 -
3.2.4	<i>Primeros pasos con Palo</i>	- 51 -
3.2.5	<i>Sobre las Reglas de Palo 2.5</i>	- 54 -
3.2.6	<i>Usuarios y Permisos</i>	- 54 -
3.2.7	<i>Soporte y Comunidad</i>	- 55 -
3.3	MONDRIAN 3.0.4	- 56 -
3.3.1	<i>Antecedentes y Visión General</i>	- 56 -
3.3.2	<i>Arquitectura de Mondrian 3.0.4</i>	- 57 -
3.3.3	<i>Instalación de Mondrian 3.0.4</i>	- 60 -
3.3.4	<i>Funcionalidades de Mondrian 3.0.4</i>	- 61 -
3.3.5	<i>Flujo de Ejecución de Mondrian 3.0.4</i>	- 62 -
3.3.6	<i>Definición del Esquema Multidimensional XML</i>	- 63 -
3.3.6.1	Definición del driver de conexión.....	- 63 -
3.3.7	<i>Expresiones Multidimensionales</i>	- 65 -
3.3.7.1	Children	- 65 -
3.3.7.2	Count.	- 67 -
3.3.7.3	Avg.....	- 67 -
3.3.7.4	Max.....	- 67 -
3.3.7.5	Min	- 68 -
3.3.7.6	Crossjoin	- 68 -

CAPÍTULO IV

4. ESTUDIO COMPARATIVO

4.1 INTRODUCCIÓN	- 69 -
4.2 DETERMINACIÓN DE LAS HERRAMIENTAS OPEN SOURCE PARA EL ANÁLISIS MULTIDIMENSIONAL.	- 70 -
4.2.1 Palo	- 70 -
4.2.2 Mondrian.....	- 70 -
4.3 DETERMINACIÓN DE LOS PARÁMETROS DE COMPARACIÓN.	- 70 -
4.3.1 Datos Cuantitativos	- 71 -
4.3.2 Datos Cualitativos	- 72 -
4.3.3 Escalas.....	- 73 -
4.4 DESCRIPCIÓN DE LAS FUENTES DE DATOS.	- 75 -
4.4.1 Palo	- 75 -
4.4.1.1 Modelamiento dimensional	- 75 -
4.4.2 Mondrian.....	- 77 -
4.4.2.1 Modelamiento dimensional	- 77 -
4.5 DESARROLLO DE PRUEBAS	80
4.5.1 Palo	80
4.5.1.1 Crear el esquema para el modelo OLAP o cubo	80
4.5.2 Mondrian.....	82
4.5.2.1 Modificar Mondrian.jsp.....	82
4.5.2.2 Crear el esquema para el modelo OLAP o cubo	83
4.5.2.3 Crear una consulta mdx.....	84
4.5.2.4 Resultados presentados en pantalla	85
4.6 ANÁLISIS CUALITATIVO Y CUANTITATIVO.....	88
4.6.1 Análisis Cuantitativo.....	88
4.6.1.1 Verificación de existencia de elementos	88
4.6.1.2 Interpretación de los resultados	89
4.6.2 Análisis Cualitativo	92
4.6.2.1 Calificaciones alcanzadas.....	92
4.6.2.2 Interpretación de los resultados	99
4.7 RESULTADO TOTAL DEL ANÁLISIS COMPARATIVO	103
4.7.1 Promedios parciales alcanzados	103
4.7.2 Comprobación de la Hipótesis.....	104
4.7.2.1 Hipótesis	104
4.7.2.2 Comprobación	104
4.7.3 Conclusión del Análisis Comparativo	105

CAPÍTULO V

5. DISEÑO E IMPLEMENTACIÓN

5.1 INTRODUCCIÓN	106
5.2 METODOLOGÍA PARA DESARROLLO DEL PROYECTO.....	107
5.2.1 Planeación y Administración del Proyecto	107
5.2.1.1 Objetivo del Proyecto	107
5.2.1.2 Definición del Proyecto	107
5.2.1.3 Alcance del Proyecto	108
5.2.1.4 Justificación del Proyecto en el Negocio	109

5.2.2	Análisis de Requerimientos.....	109
5.2.2.1	Levantamiento de Requerimientos.....	109
5.2.2.2	Documentación de Requerimientos.....	111
5.2.3	Modelamiento Dimensional	114
5.2.3.1	El Datamart.....	114
5.2.3.2	Definición de la granularidad	115
5.2.3.3	Dimensiones	116
5.2.3.4	Tablas Virtuales de Hechos.....	124
5.2.3.5	Diseño del modelo dimensional	129
5.2.4	Diseño Técnico de la Arquitectura.....	134
5.2.4.1	Datos	134
5.2.4.2	Back Room.....	141
5.2.4.3	Front Room.....	142
5.2.4.4	Infraestructura.....	142
5.2.5	Construcción de los Cubos.....	147
5.2.5.1	Archivos JSP	148
5.2.5.2	Estructuras XML.....	158
5.2.6	Reportes Implementados.....	162

CONCLUSIONES

RECOMENDACIONES

RESUMEN

SUMMARY

GLOSARIO

BIBLIOGRAFIA

BIBLIOGRAFÍA GENERAL

BIBLIOGRAFÍA RELACIONADA AL TEMA

ANEXOS

ANEXO 1

NAVEGANDO EN EL DATAMART

ANEXO 2

MANUAL TÉCNICO

INDICE DE ABREVIATURAS

BI	Business Intelligence. (Inteligencia de Negocios).
DBMS	Data Base Management System. (Sistema Manejador de Base de Datos).
DM	Datamart
DWH	Data Warehouse
E – R	Entidad – Relación
ETL	Extracción Transformación y Carga
HOLAP	Hybrid On Line Analytical Processing. (Proceso Analítico en Línea Híbrido).
I/O	Operaciones de Entrada Salida
JDBC	Origen de Datos (Conector de Base de Datos en Java)
JOLAP	Java OLAP.
JSP	Java Service Protocol (Protocolo de Servicios Java)
MCDS	Ministerio de Coordinación de Desarrollo Social
MDDB	Base de Datos multidimensional
MDX	Multidimensional Expressions. (Expresiones Multidimensionales).
MOLAP	Multidimensional On Line Analytical Processing. (Proceso Analítico en Línea Multidimensional).
OLTP	On Line Transaction Processing. (Proceso Transaccional en Línea).
OLAP	On Line Analytical Processing. (Proceso Analítico en Línea).
POO	Programación Orientada a Objetos
PSNM	Programa Sistema Nacional de Microfinanzas

PSND	Programa Social Nutriendo el Desarrollo
PSSA	Programa Social Socio Ahorro
PSSS	Programa Social Socio Solidario
RDBMS	Sistema Manejador de Base de Datos Relacional.
RIPS	Registro Interconectado de Programas Sociales
ROLAP	Relational On Line Analytical Processing. (Proceso Analítico en Línea Relacional)
RUB	Registro Único de Beneficiarios
RRHH	Recursos Humanos.
SQL	Structured Query Language
URL	Uniform Resource Locator. (Localizador Uniforme de Recursos).
XML	Extended Markup Lenguaje. (Lenguaje de Marcado Extendido).

INDICE DE TABLAS

Tabla II.1. Tecnologías OLTP vs. OLAP.	- 28 -
Tabla II.2. Ejemplos de requerimientos de reportes en OLTP vs. OLAP	- 29 -
Tabla II.3. Sistemas ROLAP vs. MOLAP	- 40 -
Tabla IV.4. Determinación de los parámetros de comparación	- 71 -
Tabla IV.5. Variables cuantitativas.....	- 72 -
Tabla IV.6. Variables cualitativas	- 73 -
Tabla IV.7. Escalas con sus equivalencias	- 74 -
Tabla IV.8. Verificación de la existencia de elementos cuantitativos.....	88
Tabla IV.9. Evaluación acerca del servidor OLAP	93
Tabla IV.10. Evaluación acerca de la Interfaz web que muestra los reportes	94
Tabla IV.11. Evaluación acerca de la visión multidimensional de los datos	95
Tabla IV.12. Evaluación acerca de la administración de dimensiones.....	96
Tabla IV.13. Evaluación acerca de los operadores de manipulación	97
Tabla IV.14. Evaluación acerca de la transparencia en el almacén de datos.....	98
Tabla IV.15. Promedios parciales alcanzados	103
Tabla IV.16. Promedio total.....	104
Tabla V.17. Campos de la vista Dimensión Centros de Acopio.....	116
Tabla V.18. Campos de la vista Dimensión Periodos del Programa	117
Tabla V.19. Campos de la vista Dimensión Precio Venta por Litro Leche.....	117
Tabla V.20: Campos de la vista Dimensión Beneficiarios por Ubicación y Datos	118
Tabla V.21: Campos de la vista Dimensión Producción Diaria Leche	119
Tabla V.22: Campos de la vista Dimensión Beneficiarios por Capacitación	120
Tabla V.23: Campos de la vista Dimensión Descuentos Compras	121
Tabla V.24: Campos de la vista Dimensión Ingresos por Ventas	122
Tabla V.25: Campos de la vista Dimensión Créditos en General	123
Tabla V.26: Campos de la vista Fact Centros de Acopio	124
Tabla V.27: Campos de la vista Fact Periodos del Programa	124
Tabla V.28: Campos de la vista Fact Precio Venta por Litro Leche	125
Tabla V.29: Campos de la vista Fact Beneficiarios por Ubicación y Datos.....	125
Tabla V.30: Campos de la vista Fact Producción Diaria Leche.....	126
Tabla V.31: Campos de la vista Fact Beneficiarios por Capacitación.....	126
Tabla V.32: Campos de la vista Fact Descuentos Compras.....	127
Tabla V.33: Campos de la vista Fact Ingresos por Ventas.....	128
Tabla V.34: Campos de la vista Fact Créditos en General	128
Tabla V.35. Características de los servidores para BD y Análisis.	147
Tabla V.36. Herramientas software de los servidores para BD y Análisis.....	147

INDICE DE FIGURAS

Figura II.1. Entorno Business Intelligence.....	- 24 -
Figura II.2. Arquitectura típica de soluciones BI	- 25 -
Figura II.3. Cubo tridimensional: Geografía, Producto y Tiempo.	- 30 -
Figura II.4. Operaciones sobre cubos: especializar y generalizar.	- 32 -
Figura II.5. Arquitectura básica OLAP	- 33 -
Figura II.6. Modelos de datos OLAP	- 34 -
Figura II.7. Modelo de estrella con 3 dimensiones y una tabla de hechos.	- 35 -
Figura II.8. Modelo de Copo de Nieve	- 36 -
Figura II.9. Modelo o esquema de constelación	- 37 -
Figura II.10. Ejemplo de un Informe OLAP.....	- 44 -
Figura II.11. Interfaz de una herramienta OLAP	- 46 -
Figura III.12. Arquitectura de PALO 2.5	- 50 -
Figura III.13. PALO 2.5: Paste View	- 51 -
Figura III.14. PALO 2.5: Paste Elements	- 52 -
Figura III.15. PALO 2.5: Paste Data Funtion	- 52 -
Figura III.16. PALO 2.5: Modeller.....	- 53 -
Figura III.17. Reglas de PALO 2.5	- 54 -
Figura III.18. PALO 2.5: Usuarios y Permisos	- 55 -
Figura III.19. Arquitectura del servidor de procesamiento analítico en línea Mondrian.....	- 59 -
Figura III.20. Funcionamiento de Mondrian	- 63 -
Figura III.21. Definición de la conexión a la Base de Datos en Workbench.....	- 64 -
Figura III.22. Desarrollo de un cubo OLAP en Workbench	- 64 -
Figura III.23. Diagrama de la estructura básica MDX.....	- 65 -
Figura III.24. Resultado de la consulta MDX Ventas de todos los productos	- 66 -
Figura III.25. Resultado de la consulta MDX Ventas de todos los productos utilizando “Children” ..	- 66 -
Figura IV.26. Diseño dimensional del data warehouse Ventas para el análisis de PALO	- 76 -
Figura IV.27. Diseño dimensional del data warehouse FoodMart para el análisis de Mondrian-	- 79 -
Figura IV.28. Herramienta de administración de bases de datos y cubos en Palo.....	81
Figura IV.29: Creación del cubo Sales en Palo	81
Figura IV.30. Líneas de conexión al warehouse FoodMart.....	83
Figura IV.31. Creación de un cubo OLAP “Sales”	84
Figura IV.32. Creación de varias medidas para el cubo “Sales”	84
Figura IV.33. Creación de dimensiones para el cubo “Sales”	84
Figura IV.34. Consulta mdx por defecto para el cubo “Sales”	85
Figura IV.35. Reporte dinámico del cubo “Sales” utilizando Palo	86
Figura IV.36. Reporte dinámico del cubo “Sales” utilizando Mondrian	87
Figura IV.37. Promedios Parciales alcanzados.....	103
Figura IV.38. Promedio total.....	104
Figura V.39. Datamart Centros de Acopio vigentes por ubicación.	129
Figura V.40. Datamart períodos del programa.	129
Figura V.41: Datamart precios vigentes del litro de leche.....	130
Figura V.42. Datamart beneficiarios datos personales.....	130
Figura V.43. Datamart producción diaria de leche	131

Figura V.44. Datamart beneficiario por capacitación	131
Figura V.45. Datamart descuentos en sus compras	132
Figura V.46. Datamart ingresos por venta de productos socio solidario.....	132
Figura V.47. Datamart créditos concedidos en general.....	133
Figura V.48. V.Arquitectura de la Solución	134
Figura V.49. Tablas para desarrollo de datamart Centros de Acopio por ubicación	135
Figura V.50. Tabla para desarrollo del datamart periodos del programa	136
Figura V.51. Tablas para desarrollo del datamart precios vigentes de litro de leche.....	136
Figura V.52. Tablas para desarrollo del datamart beneficiarios datos personales.....	137
Figura V.53. Tablas para desarrollo del datamart producción diaria de leche	137
Figura V.54. Tablas para desarrollo del datamart beneficiarios por capacitación recibida.....	138
Figura V.55. Tablas para desarrollo del datamart descuentos en sus compras	139
Figura V.56. Tablas para desarrollo del datamart ingresos por venta de productos	140
Figura V.57. Tablas para desarrollo del datamart Créditos concedidos en general	141
Figura V.58. Infraestructura de la Solución	143
Figura V.59. Líneas de código del archivo mqCantidadCentrosdeAcopio.jsp.....	148
Figura V.60. Driver de conexión a la Base de Datos y su Url	148
Figura V.61. Líneas de código del archivo RUB_PSND_1.xml	159
Figura V.62. Ejemplo de desarrollo de un esquema de cubo OLAP con una dimensión: Provincia y dos medidas: Numero Tanques y Capacidad Litros Leche.....	161
Figura V.63. Archivo de consulta MDX: mqCantidadCentrosdeAcopio.jsp	163
Figura V.64. Resultados gráficos del reporte: Centros de Acopio por ubicación	163
Figura V.65. Archivo de consulta MDX: mqCantidadPeriodos.jsp	164
Figura V.66. Resultados gráficos del reporte: Periodos del Programa	164
Figura V.67. Archivo de consulta MDX: mqPreciosVentaLeche.jsp	165
Figura V.68. Resultados gráficos del reporte: Precios vigentes de litro de leche.....	165
Figura V.69. Archivo de consulta MDX: mqCantidadBenCentrosProvinciasDatos.jsp.....	166
Figura V.70. Resultados gráficos del reporte: Beneficiarios Datos Personales	166
Figura V.71. Archivo de consulta MDX: mqCantidadBenCentrosProvinciasLitrosLecheProduc.jsp	167
Figura V.72. Resultados gráficos del reporte: Producción diaria de leche	167
Figura V.73. Archivo de consulta MDX: mqCantidadBenCentrosProvinciasCapacitacion.jsp	168
Figura V.74. Resultados gráficos del reporte: Beneficiarios por la capacitación recibida	168
Figura V.75. Archivo de consulta MDX: mqDescuentosCompras.jsp	169
Figura V.76. Resultados gráficos del reporte: Descuentos a los usuarios en sus Compras	169
Figura V.77. Archivo de consulta MDX: mqIngresosVentasProductos.jsp.....	170
Figura V.78. Resultados gráficos del reporte: Ingresos por la venta de producto sociocolidarios	170
Figura V.79. Archivo de consulta MDX: mqCreditosNivelPobreza.jsp	171
Figura V.80. Resultados gráficos del reporte: Beneficiarios por nivel de pobreza	171
Figura V.81. Archivo de consulta MDX: mqCreditosUbicacion.jsp	172
Figura V.82. Resultados gráficos del reporte: Beneficiarios por género y ubicación	172
Figura V.83. Archivo de consulta MDX: mqCreditosEntidadesMicrofinancieras.jsp	173
Figura V.84. Resultados gráficos del reporte: Beneficiarios por Entidad Microfinanciera	173
Figura V.85. Archivo de consulta MDX: mqCreditosActividad.jsp	174
Figura V.86. Resultados gráficos del reporte: Beneficiarios por Actividad Microempresarial	174
Figura V.87. Archivo de consulta MDX: mqCreditosEntidadActividad.jsp.....	175

Figura V.88. Resultados gráficos del reporte: Beneficiarios por Entidad Microfinanciera y Actividad Microempresarial	175
Figura V.89. Archivo de consulta MDX: mqCreditosForma.jsp.....	176
Figura V.90. Resultados gráficos del reporte: Beneficiarios por la forma de crédito	176
Figura V.91. Archivo de consulta MDX: mqCreditosFormaEntidad.jsp	177
Figura V.92. Resultados gráficos del reporte: Beneficiarios por la forma de crédito y Entidad Microfinanciera	177
Figura V.93. Archivo de consulta MDX: mqCreditosGeneral.jsp	178
Figura V.94. Resultados gráficos del reporte: Beneficiarios del programa	178

INTRODUCCIÓN

En la actualidad el ente empresarial de los diferentes sectores buscan extender su productividad y ventajas competitivas proporcionándole a la gerencia información analítica y estratégica para el negocio. Esto se logra al aprovechar la información (datos procesados) que a diario es almacenada en sus bases de datos operativas.

Al intentar utilizar esta información de las bases de datos operativas para tomar decisiones, se presentan varios problemas:

- Ausencia de información personalizada y/o relevante para los distintos perfiles que existen en un negocio.
- Existe demasiada información, muy genérica de la cual no se pueden sacar conclusiones.
- Falta de retroalimentación oportuna para la mejora de los negocios.
- La información muchas veces es irrelevante para el área interesada en mejorar sus decisiones.
- La organización termina por desaprovechar todos estos datos, perdiendo un proceso de aprendizaje de sus propios logros e información.

Por lo tanto, se plantea realizar una unión entre el mundo de los datos y el de los negocios, por medio de la inteligencia de negocios BI (Business Intelligence) con una solución basada en data warehousing (bodega de datos). Esta solución permite utilizar los datos operativos de una empresa para producir información relevante y que soporte la toma de decisiones empresariales.

Para el proyecto de data warehousing se toman como fuente los sistemas de información que tenga la empresa, pueden ser varios y en diferentes formatos, como bases de datos o archivos de texto. Luego de extraer los datos relevantes, son transformados de ser necesario y son cargados a una nueva base de datos, diseñada para soportar la inteligencia de negocios, que luego será analizada de forma multidimensional con análisis OLAP que es el motivo fundamental de este proyecto de tesis.

Mediante este proceso se producirá información relevante para los ejecutivos de una empresa. Preguntas como: ¿Se va a lograr una cuota de ventas en un trimestre determinado?, ¿En cuál ciudad tiene mayor potencial determinado producto?, ¿Qué tal se está vendiendo un producto con respecto a períodos de tiempo anteriores? , ¿Cual es el producto más rentable en determinada ciudad?, ¿Cuáles son mis mejores clientes? y por lo tanto, sus

decisiones correspondientes se toman a diario en una empresa, basándose en muchas ocasiones en intuiciones o conjeturas. Mediante el tipo de análisis proporcionado en este proyecto, estas preguntas serán resueltas con base en hechos y cifras rescatadas de las fuentes de datos operativas de la organización.

Grandes empresas como EPM, Telmex e IBM han utilizado la inteligencia de negocios para estos propósitos, permitiéndoles conocer mejor a sus clientes, sus productos, ventas, costos y otros factores determinantes en sus negocios. Las pymes han estado ajenas a estas tecnologías por el alto costo que una solución de inteligencia de negocios como lo es el Data warehouse implica. Es por esto que el proyecto plantea un estudio comparativo de herramientas Open Source para el análisis multidimensional del data warehouse RUB-ECUADOR para luego desarrollar una aplicación Web para la explotación de cubos y/o datamarts creados sobre ese data warehouse.

Al abordar este documento, se expone un conocimiento teórico del proceso de Inteligencia de Negocio de manera que se pueda dar a conocer claramente las fases que comprenden el análisis multidimensional OLAP y el proceso de explotación de la información plasmada en los cubos y/o datamarts en reportes para la toma de decisiones. Luego se da a conocer la metodología utilizada para el desarrollo del proyecto.

Más adelante se exponen los elementos más importantes de las herramientas libres utilizadas para la solución, para después entrar a detallar los procesos realizados en el proyecto, desde la planeación del proyecto hasta su implantación. Finalmente se exponen las conclusiones y recomendaciones que ha dejado el proyecto para dar claridad a los conceptos desarrollados y servir de base a futuros desarrollos de la misma área.

CAPÍTULO I

1. MARCO REFERENCIAL

1.1 PROBLEMATIZACIÓN

En la actualidad el tratamiento de la información transaccional, generada en el día a día de las empresas, para evaluar la situación de negocio presente y extraer tendencias de futuro necesita de herramientas informáticas que faciliten dicho análisis y la posterior toma de decisiones. Es así que hoy en día se cuenta con las herramientas de análisis multidimensional, las mismas que fueron desarrolladas para minimizar el tiempo que conlleva el análisis y la toma de decisiones en una aplicación tanto de escritorio como Web.

Así, nace la necesidad de un sistema que permita disponer de esta información rápida y fácilmente, y que la mantenga actualizada y centralizada. Las herramientas OLTP y otras más tradicionales se han mostrado de difícil uso, con inconsistencias a la hora de la obtención de datos y en ocasiones excesivamente lentas para la extracción de estos datos, por lo que en la actualidad se maneja el análisis de datos OLAP (On-Line Analytical Processing).

Este Procesamiento Analítico en Línea es la técnica que permite ver y manipular los datos por dimensiones, proveyendo a los gerentes y analistas fácil acceso a la información con el fin de soportar el proceso de toma de decisión.

Las herramientas OLAP pueden soportar requerimientos complejos de análisis, analizar datos desde diferentes perspectivas y soportar análisis complejos contra un volumen

determinado de datos. Su objetivo fundamental es proveer al usuario final el fácil análisis de los datos, con la potencia y confiabilidad de una base de datos corporativa, y con la posibilidad de ver los datos desde diversos puntos de vista o dimensiones.

El MCDS, es una institución pública que requiere analizar su información para la toma de decisiones, para tal requerimiento es necesario realizar un estudio comparativo de Herramientas Open Source que permitan Análisis Multidimensional (OLAP).

Para luego desarrollar una aplicación web que permita la explotación de los cubos y/o datamarts realizados en el análisis multidimensional mencionado de forma tal que la solución sea adecuada para la toma de decisiones del ministerio.

1.2 JUSTIFICACIÓN

La razón principal para el desarrollo del proyecto se sustenta en el decreto No. 1014 emitido por el presidente de la república del Ecuador Rafael Correa en donde se establece como política pública para las entidades de la administración pública central la utilización de Software Libre en sus sistemas y equipamientos informáticos.

Se conoce como Software Libre a los programas de computación que se pueden utilizar y distribuir sin restricción alguna y que permiten su acceso a los códigos y fuentes y que sus aplicaciones pueden ser mejoradas.

Los programas de computación incluyen las siguientes libertades; utilización del programa con cualquier propósito de uso común; distribución de copias sin restricción alguna; estudio y modificación del programa (requisito: código fuente disponible); y publicación del programa mejorado (requisito; código fuente disponible).

Asimismo, las entidades de la administración pública central, previa la instalación del Software Libre en sus equipos, deberán verificar la existencia de la capacidad técnica que brinde el soporte necesario para el uso de este tipo de software. Además, el decreto faculta la utilización de software propietario (no Libre), únicamente cuando no exista una solución de Software Libre que supla las necesidades requeridas, o cuando esté en riesgo la seguridad nacional, o cuando el proyecto informático se encuentre en un punto de no retorno.

El artículo 5 del Decreto, precisa que tanto para Software Libre como para software propietario, siempre y cuando satisfagan los requerimientos, se debe preferir 6 soluciones: nacionales que permitan autonomía y soberanía tecnológica; regionales con componente nacional; regionales con proveedores nacionales; internacionales con componente nacional; internacionales con proveedores nacionales; e internacionales.

Según el Decreto 1014, la Subsecretaría de Informática será el órgano regulador y ejecutor de las políticas y proyectos informáticos y realizará el control y seguimiento del decreto en mención. El uso de software libre contempla ventajas como:

Contar con una licencia. Siempre será mejor usar un producto Open Source a usar uno propietario pirateado.

Tener la fuente. Siempre podemos modificarlo y adaptarlo a nuestras necesidades. Incluso podemos reparar errores que detectemos o incluir parches realizados por otros usuarios. O modificarlo para que se ejecute en otro sistema operativo, contra otra base de datos, etc.

Como resultado del Estudio se seleccionará la herramienta más adecuada para efectuar el respectivo análisis multidimensional en el data warehouse RUB del MCDS, este trabajo se realizará a través de la empresa Proasetel S.A. Siendo necesario mencionar que se realizarán las debidas pruebas sobre datos reales tanto en las fuentes como en el data warehouse.

De la misma manera es importante indicar que este análisis multidimensional consiste en el desarrollo de cubos y/o datamarts provenientes del citado data warehouse, para su explotación tanto en la Intranet como en la extranet a través de una aplicación web.

1.3 OBJETIVOS

1.3.1 Objetivo General

Realizar un estudio comparativo de Herramientas Open Source para Análisis Multidimensional caso práctico: Proasetel S.A., Análisis Multidimensional del RUB-ECUADOR

1.3.2 Objetivos Específicos

- Estudiar las herramientas para el análisis multidimensional Open Source Mondrian y Palo.
- Definir las métricas de comparación que intervendrán en el estudio.
- Realizar un análisis comparativo para verificar los resultados obtenidos sobre estas herramientas, para seleccionar la más óptima para su utilización en el Análisis Multidimensional del DWH RUB-ECUADOR del MCDS realizado por Proasetel S.A.

- Implementar cubos y/o datamarts basados en el Datawarehouse RUB-ECUADOR para desarrollar una aplicación Web que permita su explotación.

1.4 HIPÓTESIS

“El Estudio comparativo para el análisis multidimensional con herramientas Open Source permitirá determinar la herramienta que brinda las mejores prestaciones para el desarrollo del análisis multidimensional del data warehouse RUB de tal forma que facilite el proceso de toma de decisiones”.

1.5 MÉTODOS Y TÉCNICAS

1.5.1 Métodos

1.5.1.1 Inductivo – Deductivo

Sigue un proceso analítico sintético, en el cual la inducción y la deducción se complementan.

El método inductivo es aquel que parte de lo particular a lo general, de las partes al todo.

El método deductivo va de lo general a lo particular. Parte del concepto a los principios, definiciones o afirmaciones de las cuales se extrae conclusiones y consecuencias.

1.5.1.2 Método Comparativo

La presente investigación no se limita al estudio de una sola herramienta sino al estudio comparativo de varias para el análisis multidimensional de datos.

1.5.1.3 Científico

Permite descubrir variedades científicas. Se utilizará este método para la recolección de información y desarrollo de la investigación.

1.5.1.4 Experimental

Se fundamenta en el método científico, comprueba en forma objetiva una ley o una verdad científica, enriquece la calidad de información, datos y vivencias que contribuyen a interpretar la realidad y a actuar sobre ella conscientemente.

1.5.2 Técnicas

1.5.2.1 Fuentes

Son hechos o documentos a los que accede el investigador y que le permiten obtener información que puede ser primaria o secundaria.

Primaria: Se utilizarán las siguientes técnicas:

- Observación
- Entrevista
- Encuestas
- Lluvia de ideas

Secundarias: Se tomará información de medios como:

- Internet
- Revistas
- Prensa

CAPÍTULO II

2. MARCO TEÓRICO

2.1 BUSINESS INTELLIGENCE (BI)

La Inteligencia del Negocio BI o Business Intelligence por sus siglas en inglés, suele definirse como la transformación de los datos de la compañía en conocimiento para obtener una ventaja competitiva.

Otra definición puede ser que esta se refiere al proceso de recolectar, consolidar y almacenar múltiples fuentes de datos para la toma de decisiones estratégicas. En una sola palabra "ANALISIS"

Desde un punto de vista más pragmático, y asociándolo directamente a las tecnologías de la información, podemos definir a la Inteligencia del Negocio como el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada (interna y externa a la compañía) en información estructurada, para su explotación directa (reporting, análisis OLAP...) o para su análisis y conversión en conocimiento soporte a la toma de decisiones estratégicas sobre el negocio.

Esta definición pretende abarcar y describir el ámbito integral del entorno BI ya que en la presente se describe en detalle solo una parte de Business Intelligence, es decir lo que se refiere al análisis multidimensional y a su explotación en reportes para la toma de decisiones, reflejado resumidamente en el gráfico que aquí se muestra.

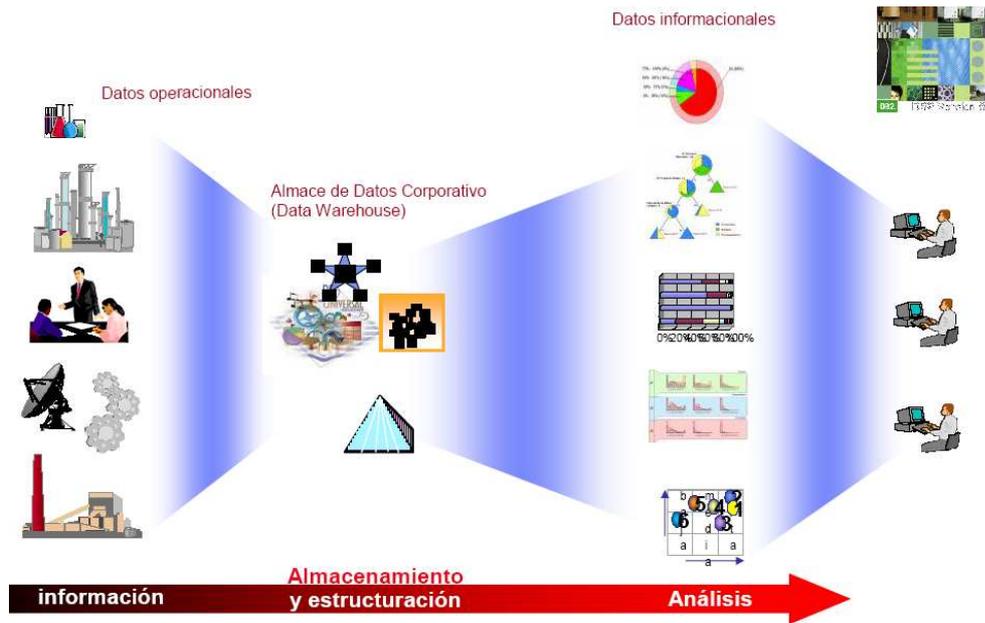


Figura II.1. Entorno Business Intelligence

2.2 ARQUITECTURA DE BUSINESS INTELLIGENCE

La arquitectura de las soluciones BI han ido evolucionando ya que antes no bastaba con tener el manejador de base de datos más robusto, porque la problemática era la extracción de datos almacenados en otros sistemas.

Es aquí donde surge el concepto del Enterprise Service Bus (ESB), infraestructura de middleware que habilita y gestiona la comunicación para la integración entre componentes aplicativos heterogéneos. ESB se basa en estándares como Web services, XML y otros.

Un Business Intelligence se compone de diferentes tecnologías que se integran para formar una solución empresarial.

Los componentes se orientan a transformar los datos en información oportuna y confiable para el usuario final.

Por otro lado una arquitectura Business Intelligence BI se divide en las siguientes capas:

- Sistemas Fuentes.
- Capa Data Warehouse y
- La Capa BI

Siendo en una parte de la capa data warehouse referente al análisis multidimensional, y la capa BI que centraremos el estudio.

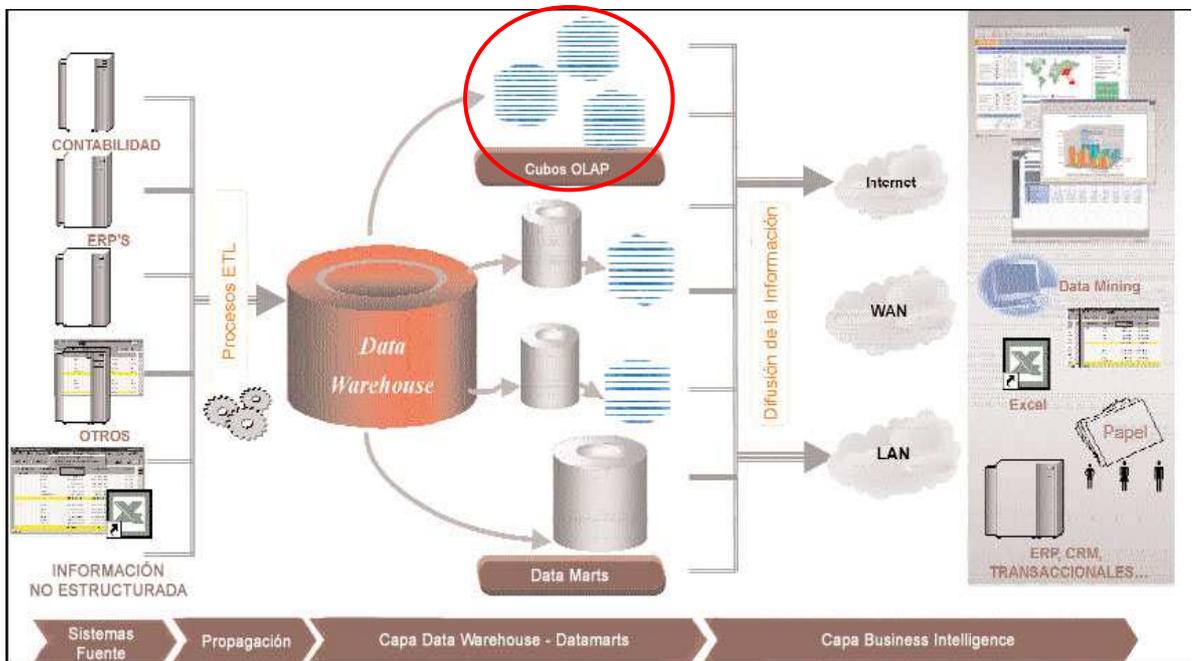


Figura II.2. Arquitectura típica de soluciones BI

La figura anterior además de mostrar la arquitectura típica de soluciones BI muestra el objeto básico de nuestro estudio en este capítulo, el cual está señalado con un círculo, y se refiere al análisis multidimensional mediante la solución en *cubos*.

2.2.1 Sistemas Fuentes

Los datos administrados por los sistemas de aplicación operacionales son la fuente principal de datos para el data warehouse. Estos sistemas son los encargados de recolectar información diaria de las tareas operativas de la organización

Estos datos operacionales constituyen la base de todo sistema Business Intelligence ya que de estos depende la calidad de información que se entregue al usuario final.

En muchas ocasiones también se requiere de datos externos para alimentar al sistema como hojas electrónicas, archivos de texto, debido a que algunas áreas de la institución no se encuentran automatizadas por lo que no cuentan con un sistema transaccional donde almacenar los datos.

2.2.2 Capa DATA WAREHOUSE

La capa data warehouse es el centro de la arquitectura en un sistema Business Intelligence, que se encarga de organizar y almacenar los datos para el análisis de los mismos.

En el presente documento de tesis de grado no se tomará muy en cuenta la creación de un data warehouse, sino más bien se tratará acerca del análisis multidimensional que se realiza sobre el mismo.

Es necesario mencionar que este análisis se realiza en la misma capa Data Warehouse, y se trata de implementación de cubos para representar los datos transformados en información lista para su explotación.

2.2.3 Capa BI

Existen herramientas necesarias para comprender los datos y utilizarlos para tomar decisiones inteligentes. Proporcionan toda la funcionalidad de análisis y creación de informes empresariales dirigidos a una amplia gama de posibles usuarios. Estas herramientas son las siguientes:

- Consultas e Informes.
- Cuadros de mando.
- Sistemas de Información Ejecutiva.
- OLAP.
- Minería de datos.
- Visualización Geográfica.

2.3 ANÁLISIS OLAP

OLAP (On-Line Analytical Processing). El Procesamiento Analítico en Línea es la técnica que permite ver y manipular los datos por dimensiones, proveyendo a los gerentes y analistas fácil acceso a la información con el fin de soportar el proceso de toma de decisión.

En esta técnica de análisis, en lugar de ejecutar múltiples consultas, los datos son estructurados para permitir un acceso rápido y fácil a las respuestas de las preguntas que son típicamente formuladas. De esta manera, OLAP, brinda flexibilidad en la visualización de la información.

Las herramientas OLAP pueden soportar requerimientos complejos de análisis, analizar datos desde diferentes perspectivas y soportar análisis complejos contra un volumen determinado de datos.

Su objetivo fundamental es proveer al usuario final el fácil análisis de los datos, con la potencia y confiabilidad de una base de datos corporativa, y con la posibilidad de ver los datos desde diversos puntos de vista o dimensiones.

Permite vistas reformateadas y calculadas sin el riesgo de perder o corromper los datos originales y hace que la información pueda ser compartida por varios usuarios sin tener que duplicar archivos. En muchos casos los usuarios pueden adicionar o cambiar datos sin el riesgo de sobrescribir la información original.

El uso más común de estas herramientas en una empresa se da en el análisis de ventas y compras de materia prima. Gracias a este análisis se evalúa la rentabilidad de productos, capacidad de producción o la demanda. Estos aspectos dependen directamente de los requerimientos del negocio específicos para cada empresa.

Las herramientas OLAP están dirigidas principalmente a los usuarios finales por lo que requieren de una interfaz simple y deben tener una buena integración con los sistemas que las alimentan.

2.3.1 Características Principales de las Tecnologías OLAP

Rápido: proporciona la información al usuario a una velocidad constante. La mayoría de las peticiones se deben de responder al usuario en cinco segundos o menos.

Análisis: realiza análisis estadísticos y numéricos básicos de los datos, predefinidos por el desarrollador de la aplicación o definido "ad hoc" por el usuario.

Compartida: implementa los requerimientos de seguridad necesarios para compartir datos potencialmente confidenciales a través de una gran población de usuarios.

Multidimensional: llena la característica esencial del OLAP, que es ver la información en determinadas vistas o dimensiones.

Información: acceden a todos los datos y a la información necesaria y relevante para la aplicación, donde sea que ésta resida y no esté limitada por el volumen. El OLAP es un componente clave en el proceso de almacenamiento de datos (data warehousing) y los servicios OLAP proporcionan la funcionalidad esencial para una gran variedad de aplicaciones que van desde reportes corporativos hasta soporte avanzado de decisiones.

2.3.2 Tecnologías OLTP vs. OLAP

Estas tecnologías tienen que ver con el almacenamiento de datos transaccional u operacional como OLTP y analítico como OLAP.

La tecnología OLTP es la implementada en las bases de datos operacionales que mayormente se encuentran en el ente empresarial del Ecuador, las cuales no pueden ser utilizadas para realizar un análisis adecuado sobre sus datos residentes, para solucionar este problema se tiene el procesamiento analítico en línea OLAP que es la solución que permite realizar análisis inteligente de los datos de un data warehouse para convertir esto en información pertinente de manera que se pueda explotar dicha información de cubos multidimensionales en reportes adecuados para la toma de decisiones estratégicas.

A continuación se ilustra en la siguiente tabla las características más sobresalientes de estas dos tecnologías:

	OLTP (On Line Transaccional Processing)	OLAP (On Line Analytical Processing)
Usuario típico	Empleados	Gerentes o analistas
Uso del sistema	Operación día a día	Análisis del negocio
Interacción con el usuario	Predeterminada	Ad-hoc
Datos	Datos actuales	Datos actuales + históricos
Características de los datos	Atómicos	Resumidos
Características del trabajo	Lectura/escritura	Lectura (excepto actualizaciones off-line)
Unidad de trabajo	Transacción	Query
Procesamiento	Orientado a procesos	Orientado a requerimientos
Actualizaciones	Un registro a la vez	Varios registros a la vez

Tabla II.1. Tecnologías OLTP vs. OLAP.

Como se puede observar claramente en la tabla anterior la tecnología OLAP es muy necesaria a nivel de las altas gerencias en una empresa, realizando análisis del negocio con un alto grado de interacción con los usuarios Ad-hoc que se refiere a la habilidad para los usuarios de

cambiar los parámetros sobre un reporte para crear sus propias versiones personalizadas de ese reporte. De esta forma se dará respuesta a necesidades como:

- Quiero ver este reporte, pero por mes en lugar de trimestre
- Puedo ver este reporte a nivel provincia en lugar de región
- Puedo ver esta evolución de ventas mensuales pero sólo de mi equipo de ventas

Más importante aún es que las respuestas a estos interrogantes lo resuelven los propios usuarios sin la necesidad de la intervención del departamento sistemas y maximizando el tiempo de análisis por sobre el tiempo de construcción e integración de la información.

Otro parámetro importante en las tecnologías de análisis multidimensional OLAP es el hecho de que se puede realizar actualizaciones de varios registros a la vez y no se limita como en el caso de las tecnologías OLTP a actualizaciones de un registro a la vez.

Por otro lado las tecnologías OLTP son únicamente necesarias en las bases de datos operacionales o transaccionales mientras que OLAP es análisis de los datos de un data warehouse para convertir en información estratégica. En la siguiente tabla, algunos ejemplos de preguntas para reportes que se hace en cada una de las mencionadas tecnologías:

OLTP	OLAP
• ¿Cuándo se envió la orden?	• ¿Qué factores afectaron el tiempo de entrega de las órdenes?
• ¿Cuántos fueron los ingresos el último año?	• ¿Cómo contribuyó cada línea (o producto) a las ganancias por distrito (o tienda) durante el último trimestre (o mes o año)?
• ¿Cuánto fue el inventario total el último año?	• ¿Cuales productos tienen el más bajo Margen Bruto de retorno en inventario?

Tabla II.2. Ejemplos de requerimientos de reportes en OLTP vs. OLAP

2.3.3 Conceptos y Componentes

2.3.3.1 Cubo

OLAP efectúa el almacenamiento lógico de los datos en arreglos ó matrices multidimensionales denominadas cubos. El cubo contiene los datos que son de interés para los usuarios; organiza la información dentro de dimensiones y medidas en una estructura multidimensional para soportar las preguntas que tienen los usuarios acerca de los datos de su compañía. Además proporcionan un mecanismo para la consulta de datos con rapidez y con una respuesta uniforme ante diferentes cantidades de datos que contenga el cubo o por la complejidad de una consulta.

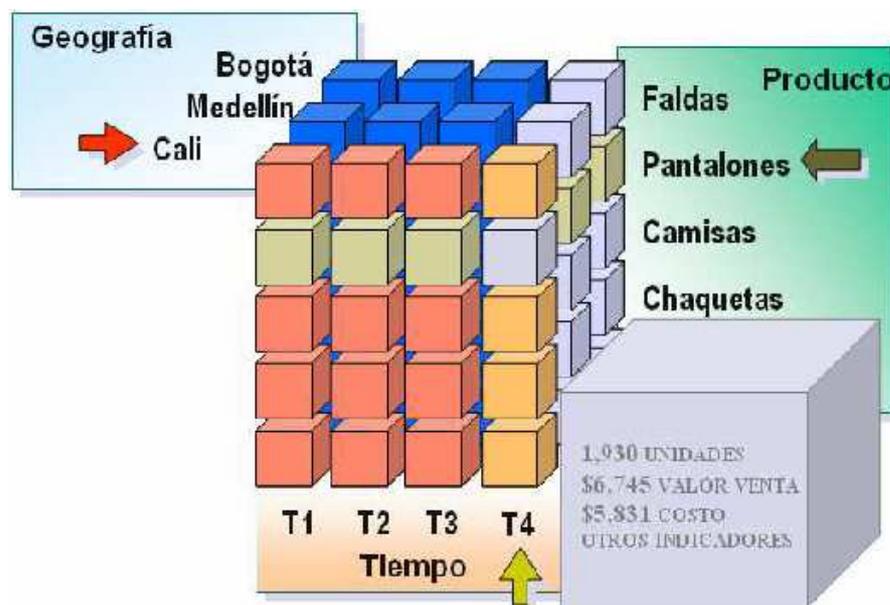


Figura II.3. Cubo tridimensional: Geografía, Producto y Tiempo.

Un cubo se compone de dimensiones, jerarquías (niveles) y medidas. En el ejemplo de la figura 3 se tiene un cubo con tres dimensiones: Geografía, Producto y Ciudad.

Además se tienen tres medidas: Unidades, Valor venta y Costo. En la celda de la parte inferior derecha de la imagen se muestran los datos para una posible pregunta gerencial:

- ¿Cuántas unidades, a qué valor y con qué costo se vendieron pantalones en la ciudad de Cali en el tiempo T4? Con su respectiva información.

2.3.3.2 Medida

La medida es el valor que toma determinada variable que se está analizando. Las medidas son resultados cuantificables, o indicadores clave de desempeño usados para

determinar el éxito de una operación de negocios. Orientan las respuestas a preguntas relacionadas con cuestiones numéricas como la cantidad, valor o costo.

En el caso de la figura 3 se tienen tres medidas, indicando que se vendieron 1930 unidades, a un valor de venta de 6745 y con costo de 5831. Un cubo puede contener una o varias medidas, dependiendo del diseño y los requerimientos. Existen dos tipos de medidas:

Medida regular: toma su dato directamente de una fuente disponible. Es un compendio de información que ya se tiene, tal como el número de unidades vendidas, ingresos, gastos, niveles de inventario.

Medida calculada: obtiene como resultado un nuevo dato numérico para medidas que no están en una fuente directa disponible. Es derivada de otras medidas. Ejemplos de este tipo de medidas son: ganancia (ingresos – costos), margen de ganancia (ingreso – costo /ingreso), tiempo promedio de espera (fecha de entrega – fecha de la orden), etc.

2.3.3.3 Dimensión

Los atributos de tipo texto que describen cosas son organizados en dimensiones. Es necesario establecer un criterio puramente de diseño y basado en los requerimientos del negocio para establecer los atributos que se incluyen como dimensiones y los que se pueden descartar al realizar la bodega de datos.

2.3.3.4 Nivel

Las dimensiones están construidas por niveles. Estos niveles representan la jerarquía establecida por las estructuras organizacionales y modelos de datos que la organización usa.

Cada nivel inferior provee cada vez datos más detallados que relaciona a la dimensión.

Las herramientas especializadas para análisis OLAP permiten fijar este nivel de granularidad en forma dinámica mientras el usuario final navega por su reporte. La dimensión tiempo provee un claro ejemplo del uso de niveles. Se tiene el año en un nivel superior, luego le siguen el semestre, trimestre, mes y por último en el nivel más inferior se encuentra el día.

2.3.4 Operaciones con OLAP

La información que se analiza con OLAP debe estar estructurada de tal forma que se puedan realizar las siguientes operaciones:

- **Drill Down y Roll Up (profundizar y escalar):** Estas dos operaciones permiten visualizar la información a un nivel de detalle distinto del actual. Drill Down

permite ver un nivel mayor de detalle, es decir de lo general se va a lo particular. Roll Up permite al usuario desplazarse entre los niveles superiores para obtener información agregada, ver acumulados y sumalizaciones.

- Alterar las filas por columnas (permutar dos dimensiones de análisis). Rotar (Swap).
- Obtener interactivamente respuestas desde diferentes perspectivas.
- Realizar consultas que requieren combinación de diferentes fuentes contenidas en el data warehouse.
- Efectuar cálculos relativamente complejos (ranking, porcentajes, sumas, etc.)
- Slice and Dice (Cortar y Rotar): Estas dos operaciones permiten navegar a través de un cubo visualizado. La operación Slice corta el cubo para que el usuario pueda enfocarse solamente en algunas perspectivas. La operación Dice hace que el cubo rote para poder apreciar la información desde otra perspectiva. Por ejemplo si se tiene un reporte que muestra el número de productos vendidos por cada sucursal al final del último trimestre, se puede cortar y rotar la información para mostrar los ingresos sobre los últimos dos meses por cada línea de producto
- Generalizar y Especializar, o como en la biografía se conoce: abstracción y concentración (generalizar y especializar).

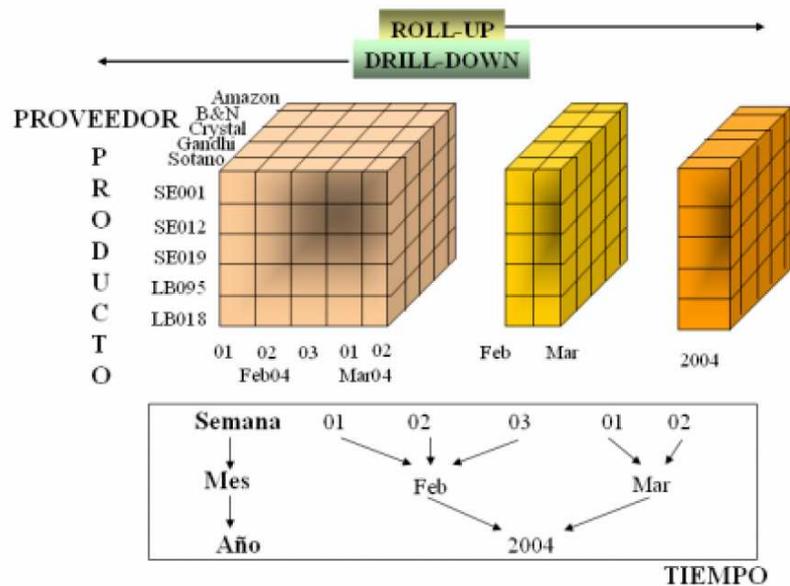


Figura II.4. Operaciones sobre cubos: especializar y generalizar.

2.3.5 Arquitectura OLAP

Aunque en ocasiones se utilizan indistintamente, los términos data warehouse y proceso analítico en línea (OLAP) se aplican a diferentes componentes de sistemas de ayuda a la toma de decisiones o sistemas de inteligencia empresarial. Los datos contenidos en un data warehouse se encuentran organizados para permitir el análisis por medio de herramientas OLAP.

La tecnología OLAP permite un uso más eficaz de los almacenes de datos para el análisis en línea, lo que proporciona respuestas rápidas a consultas analíticas complejas e iterativas. Los modelos de datos multidimensionales de OLAP y las técnicas de agregados de datos organizan y resumen grandes cantidades de datos para que puedan ser evaluados con rapidez mediante el análisis en línea y las herramientas gráficas.

La respuesta a una consulta realizada sobre datos históricos a menudo suele conducir a consultas posteriores en las que el analista busca respuestas más concretas o explora posibilidades. Los sistemas OLAP proporcionan la velocidad y la flexibilidad necesarias para dar apoyo al analista en tiempo real.

La figura 5 muestra la integración del data warehouse y los procesos OLAP, que generalmente se implementan por medio de una aplicación servidora que accede al data warehouse y realiza los procesos de análisis. A través de este servicio OLAP, los usuarios acceden a la información residente en las bases de datos, de la siguiente manera:

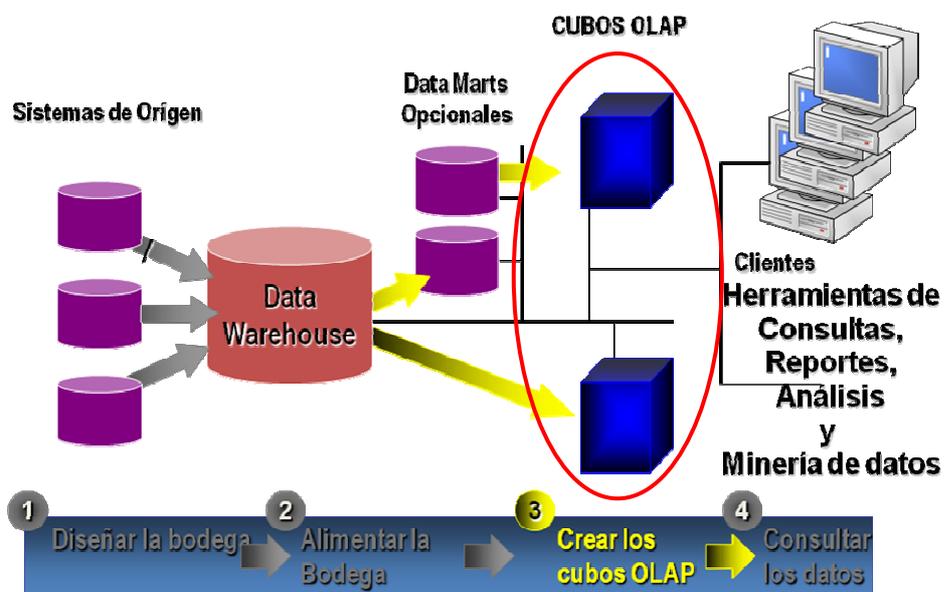


Figura II.5. Arquitectura básica OLAP

2.3.6 El Modelo de Datos OLAP

Un reto fundamental en la implementación del OLAP es mapear el esquema inicial de la base de datos a un modelo multidimensional. Esto requiere de un significativo esfuerzo de programación con muchos de los productos en el mercado hoy en día.

En la evolución de los productos OLAP, el diseño de la base de datos OLAP se ha vuelto un proceso especializado y arcano, intrincadamente enlazado a la tecnología específica del OLAP que se esté implementando.

Consecuentemente, los desarrolladores de bases de datos OLAP son muy especializados, lo cual ha llevado a altos costos de desarrollo de aplicaciones concentrados en la etapa de diseño de datos.

En la mayoría de las implementaciones de OLAP, se asume que los datos han sido preparados para el análisis a través del almacenamiento de datos (data warehousing) y que la información se ha extraído de sistemas operacionales, limpiado, validado y resumido antes de incorporarse en una aplicación OLAP.

Este es un paso vital en el proceso, que asegura que los datos que son vistos por el usuario OLAP son correctos, consistentes y que llenan las definiciones organizacionales para los datos.

Los datos se organizan en los siguientes esquemas:

- Modelo Estrella
- Modelo Copo de nieve
- Constelación

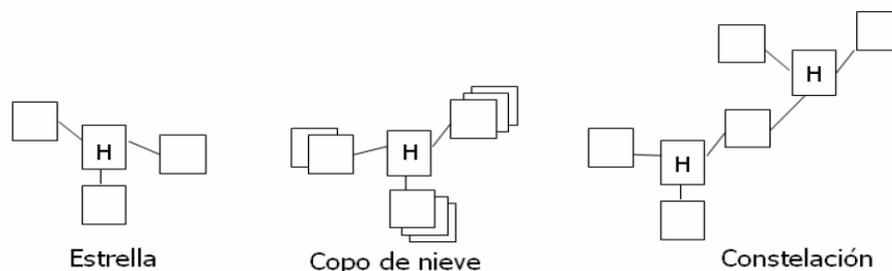


Figura II.6. Modelos de datos OLAP

2.3.6.1 Modelo o esquema de Estrella (“star schema”) tradicional

El esquema de estrella es una técnica de modelado usada para hacer corresponder un modelo multidimensional a una base de datos relacional, su nombre es porque gráficamente se parece a una estrella.

El esquema de estrella tiene cuatro componentes: hechos, dimensiones, atributos y jerarquías de atributos. Los hechos y dimensiones son representados por tablas físicas en el almacén de datos, la tabla de hechos está relacionada a cada dimensión en una relación uno a muchos.

Las tablas y dimensiones están relacionadas por llaves foráneas y están sujetas a las restricciones de las llaves foráneas y primarias.

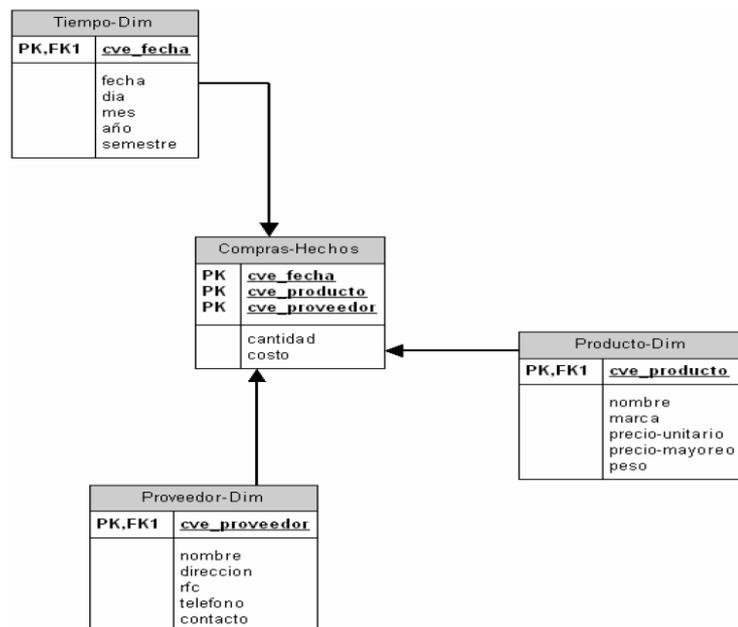


Figura II.7. Modelo de estrella con 3 dimensiones y una tabla de hechos.

Para calcular el total de compras realizadas a proveedores de Internet el esquema estrella tendría que realizar los siguientes pasos:

1. De la dimensión Proveedor, seleccionar todos los proveedores donde el canal es Internet.
2. De la tabla de hechos, seleccionar y calcular la suma de todas las cantidades y costos de las transacciones de compra a los proveedores del paso 1.

2.3.6.2 Modelo Copo de Nieve (“snow flake schema”)

El esquema de copo de nieve es una variación de la estrella tradicional, lo que se hace es que en cada dimensión se almacenan jerarquías de atributos o bien simplemente se separan atributos en otra entidad por razones de desempeño y mejor utilización del espacio. En la siguiente figura la dimensión producto se ha modificado separando sus datos generales de sus otras características.

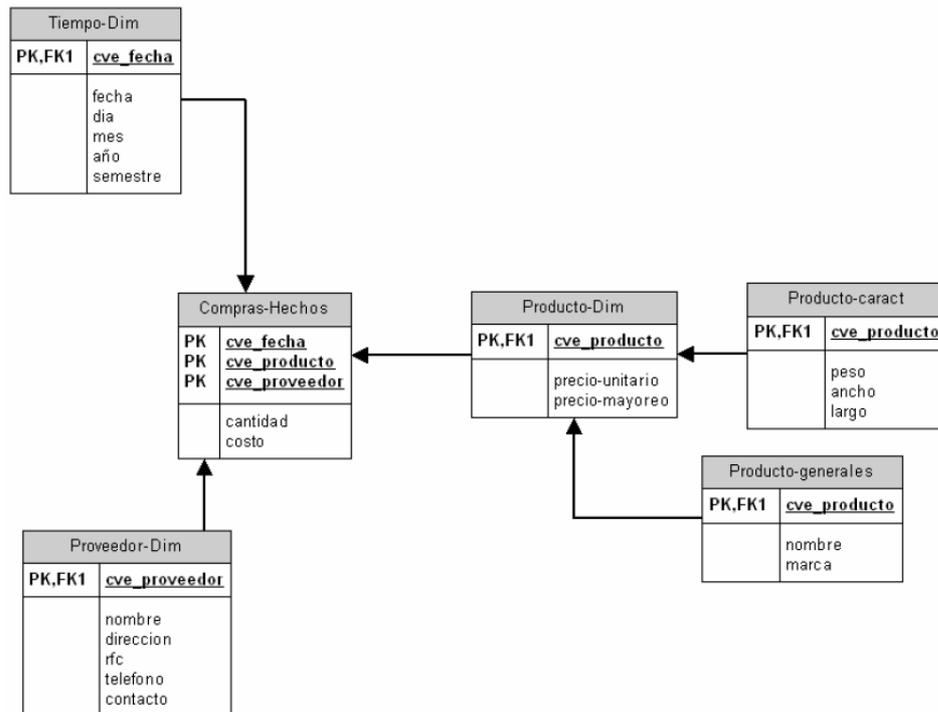


Figura II.8. Modelo de Copo de Nieve

2.3.6.3 Modelo Constelación

El modelo constelación nuevamente es una variación del esquema estrella tradicional, en este modelo algunos atributos de las dimensiones se separan formando una nueva entidad que puede ser compartida con otros cubos.

La utilidad principal de este modelo es que al tener dimensiones que pueden ser compartidas por diferentes cubos se tendrá un mejor uso del espacio de almacenamiento evitando la redundancia.

En la Figura 9 se puede observar como está implementado un ejemplo de un modelo o esquema de constelaciones:

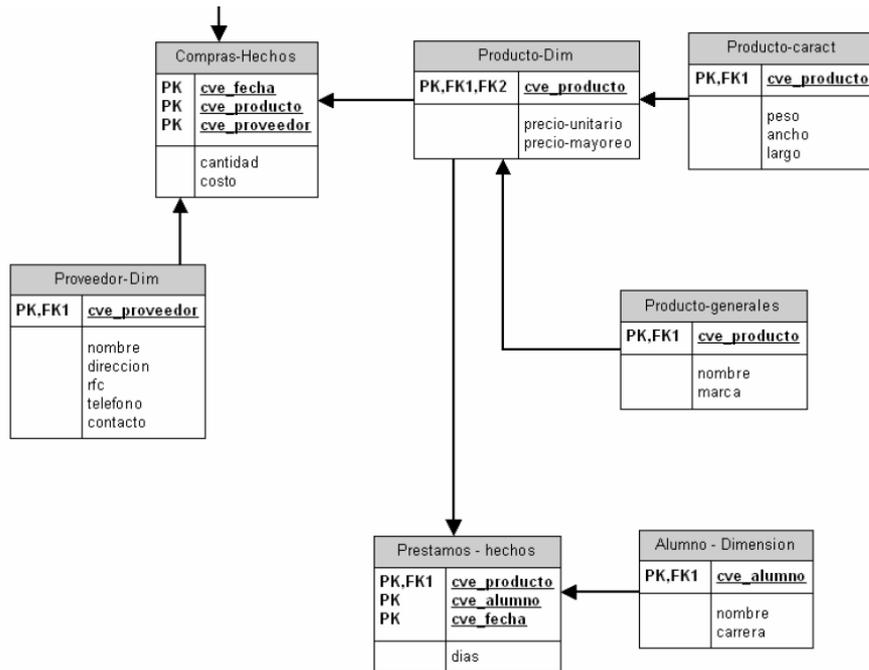


Figura II.9. Modelo o esquema de constelación

2.3.7 Implementación del OLAP

Los cubos, las dimensiones y las jerarquías son la esencia de la navegación multidimensional del OLAP. Al describir y representar la información en esta forma, los usuarios pueden navegar intuitivamente en un conjunto complejo de datos. Sin embargo, el solo describir el modelo de datos en una forma más intuitiva, hace muy poco para ayudar a entregar la información al usuario más rápidamente.

Un principio clave del OLAP es que los usuarios deberían de ver tiempos de respuesta consistentes para cada vista de datos que requieran. Dado que la información se colecta en el nivel de detalle solamente, el resumen de la información es usualmente calculado por adelantado. Estos valores pre-calculados, son la base de las ganancias de desempeño del OLAP.

En los primeros días de la tecnología OLAP, la mayoría de las compañías asumía que la única solución para una aplicación OLAP era un modelo de almacenamiento no relacional. Después, otras compañías descubrieron que a través del uso de estructuras de base de datos (esquemas de estrella y de copo de nieve), índices y el almacenamiento de agregados, se podrían utilizar sistemas de administración de bases de datos relacionales (RDBMS) para el OLAP.

Estos vendedores llamaron a esta tecnología OLAP relacional (ROLAP). Las primeras compañías adoptaron entonces el término OLAP multidimensional (MOLAP), estos conceptos, MOLAP y ROLAP, se explican con más detalle en los siguientes párrafos. Las implementaciones MOLAP normalmente se desempeñan mejor que la tecnología ROLAP, pero tienen problemas de escalabilidad. Por otro lado, las implementaciones ROLAP son más escalables y son frecuentemente atractivas a los clientes debido a que aprovechan las inversiones en tecnologías de bases de datos relacionales preexistentes.

Un desarrollo reciente ha sido la solución OLAP híbrida (HOLAP), la cual combina las arquitecturas ROLAP y MOLAP para brindar una solución con las mejores características de ambas: desempeño superior y gran escalabilidad. Un tipo de HOLAP mantiene los registros de detalle (los volúmenes más grandes) en la base de datos relacional, mientras que mantiene las agregaciones en un almacén MOLAP separado.

2.3.7.1 Sistemas MOLAP

Los datos que subyacen en los hipercubos son almacenados junto con las agregaciones en una estructura multidimensional es decir en los cubos fuera del data warehouse o Datamart (duplicación).

Ofrecen mejor respuesta a las consultas pues:

- Tienen los datos calculados y los join realizados.
- No hay bloqueos, son sólo lectura.
- Las celdas vacías no se almacenan.
- Se utiliza procesador OLAP que trabaja con índices bitmap y caché de datos.

Los sistemas MOLAP son portables. Además tienen buen comportamiento con no más de 10 dimensiones y volumen inferior a 5 GB.

Un sistema MOLAP usa una base de datos propietaria multidimensional, en la que la información se almacena multidimensionalmente, para ser visualizada en varias dimensiones de análisis.

El sistema MOLAP utiliza una arquitectura de dos niveles: la bases de datos multidimensionales y el motor analítico.

La base de datos multidimensional es la encargada del manejo, acceso y obtención del dato. El nivel de aplicación es el responsable de la ejecución de los requerimientos OLAP. El nivel de presentación se integra con el de aplicación y proporciona una interfaz a través de la cual los

usuarios finales visualizan los análisis OLAP. Una arquitectura cliente/servidor permite a varios usuarios acceder a la misma base de datos multidimensional.

La información procedente de los sistemas operacionales, se carga en el sistema MOLAP, mediante una serie de rutinas por lotes. Una vez cargado el dato elemental en la Base de Datos multidimensional (MDDDB), se realizan una serie de cálculos por lotes, para calcular los datos agregados, a través de las dimensiones de negocio, rellenando la estructura MDDDB.

Tras rellenar esta estructura, se generan unos índices y algoritmos de tablas hash para mejorar los tiempos de accesos a las consultas. Una vez que el proceso de compilación se ha acabado, la MDDDB está lista para su uso. Los usuarios solicitan informes a través de la interface, y la lógica de aplicación de la MDDDB obtiene el dato.

La arquitectura MOLAP requiere unos cálculos intensivos de compilación. Lee de datos pre-compilados, y tiene capacidades limitadas de crear agregaciones dinámicamente o de hallar ratios que no se hayan pre-calculados y almacenados previamente.

2.3.7.2 Sistemas ROLAP

Los datos se mantienen en la base de datos relacional y las agregaciones se mantienen en tablas dentro de la base de datos donde se encuentra el Datamart.

Ventajas de su uso:

- No se duplican los datos.
- Se puede usar el lenguaje SQL.
- No hay limitación en uso de dimensiones.

No son portables. Además tiene sentido utilizarse para datos poco consultados.

El sistema ROLAP utiliza una arquitectura de tres niveles. La base de datos relacional maneja los requerimientos de almacenamiento de datos, y el motor ROLAP proporciona la funcionalidad analítica. El nivel de base de datos usa bases de datos relacionales para el manejo, acceso y obtención del dato. El nivel de aplicación es el motor que ejecuta las consultas multidimensionales de los usuarios.

El motor ROLAP se integra con niveles de presentación, a través de los cuáles los usuarios realizan los análisis OLAP.

Después de que el modelo de datos para el Data Warehouse se ha definido, los datos se cargan desde el sistema operacional. Se ejecutan rutinas de bases de datos para agregar el

dato, si así es requerido por los modelos de datos. Se crean entonces los índices para optimizar los tiempos de acceso a las consultas.

Los usuarios finales ejecutan sus análisis multidimensionales, a través del motor ROLAP, que transforma dinámicamente sus consultas a consultas SQL. Se ejecutan estas consultas SQL en las bases de datos relacionales, y sus resultados se relacionan mediante tablas cruzadas y conjuntos multidimensionales para devolver los resultados a los usuarios.

La arquitectura ROLAP es capaz de usar datos pre-calculados si estos están disponibles, o de generar dinámicamente los resultados desde los datos elementales si es preciso. Esta arquitectura accede directamente a los datos del Data Warehouse, y soporta técnicas de optimización de accesos para acelerar las consultas. Estas optimizaciones son, entre otras, particionado de los datos a nivel de aplicación, soporte a la desnormalización y joins múltiples.

2.3.7.3 ROLAP vs. MOLAP

ROLAP	MOLAP
<ul style="list-style-type: none">• Delega la negociación entre tiempo de respuesta y el proceso por lotes al diseño del sistema.	<ul style="list-style-type: none">• Requiere que sus bases de datos se pre-compilen para conseguir un rendimiento aceptable en las consultas, incrementando, por tanto los requerimientos por lotes.
<ul style="list-style-type: none">• Soportan con facilidad la alta volatilidad de los datos, es decir pueden realizar consolidación ad-hoc	<ul style="list-style-type: none">• Están más orientados hacia consolidaciones por lotes
<ul style="list-style-type: none">• Pueden crecer hasta un gran número de dimensiones	<ul style="list-style-type: none">• Generalmente son adecuados para diez o menos dimensiones.
<ul style="list-style-type: none">• Soportan análisis OLAP contra grandes volúmenes de datos elementales	<ul style="list-style-type: none">• Se comportan razonablemente en volúmenes más reducidos (menos de 5Gb)

Tabla II.3. Sistemas ROLAP vs. MOLAP

Por esto, ROLAP es una arquitectura flexible y general, que crece para dar soporte a amplios requerimientos OLAP. MOLAP es una solución particular, adecuada para soluciones

departamentales con unos volúmenes de información y número de dimensiones más modestos.

2.3.7.4 Sistemas HOLAP (Sistema Híbrido)

Los datos que subyacen en los hipercubos son almacenados en una estructura o base de datos relacional y las agregaciones en una estructura multidimensional o lo que se denomina cubos.

Ventajas de su uso:

- Consumen menos espacio de disco
- Consultas más ágiles al utilizar motor OLAP

Los sistemas HOLAP no son portables. Ofrecen solución intermedia (coste almacenamiento – rendimiento consultas).

Pos cierto la portabilidad se refiere a la habilidad que tienen los diferentes software de segmentar la información en diferentes plataformas hardware.

2.3.8 El Entorno OLAP en la Actualidad

2.3.8.1 OLAP y el código propietario

Hoy en día no es una sorpresa que Microsoft OLAP está transformando la industria OLAP. Está siendo distribuido en una escala gigantesca, a precios que son una fracción de los demás servidores OLAP que además, no son mejor en calidad. Presenta Wizards como todos los productos modernos de Microsoft (aunque el desarrollo de una aplicación es más trabajoso de lo esperado).

Presenta una irrelevante herramienta para el cliente (muchas de las funcionalidades no están bien optimizadas), está siendo adoptado por muchos proveedores de soluciones (de Microsoft) y especialistas, en la creación de aplicaciones. Aunque el producto no sea el mejor, está siendo instalado en gran escala, muchas de las personas que lo compran meramente por el precio han sido sorprendidos por su calidad, rendimiento y funcionalidad.

Ahora Microsoft posee un servidor OLAP que es tan funcional como otros competidores establecidos, en algunos aspectos mejor que cualquiera de ellos y débil en algunas áreas. Esto a un precio muy bajo comparado con los demás productos comerciales que existen. Particularmente para la Edición Empresarial (Enterprise Edition) es aún mucho más barata que

cualquier otro servidor o funcionalidad OLAP. Para muchos clientes, Microsoft OLAP es una posible alternativa para la mayoría de las aplicaciones.

Otros propietarios de servidores OLAP descartan la amenaza de Microsoft, señalando que no están compitiendo directamente con las ventas de Microsoft, además que Microsoft tiene poca experiencia en el tema OLAP. Los dos planteamientos son validos pero estos no han impedido que muchos sitios experimenten con el producto ya que es muy barato. El efecto de esto es visible en los resultados públicos de los demás vendedores después de la llegada de Microsoft OLAP. Por ejemplo, los servidores Prodea y STG ROLAP están totalmente descontinuados, MicroStrategy está luchando por mantenerse, Hyperion y Applix tienen ventas menores que las de Microsoft, Express y Holos han disminuido sus ventas, Pilot y Gentia desaparecen de la escena. Solo Bussiness Objects y Congnos han tenido significativas mejoras.

Los competidores están encontrando que Microsoft OLAP es una solución para más aplicaciones de las que ellos mismos esperaban, lo cual está tentando más desarrolladores de aplicaciones analíticas a usar SQL Server y Análisis Services como plataforma.

Los compradores tienen una gama de herramientas clientes para escoger, aunque están algo decepcionados por la poca cantidad de productos que explotan el rico API que ha surgido; la mayoría de los clientes OLAP en el mercado se conectan con Analysis Services, pero muy pocos explotan su poder. Vergonzosamente, el propio cliente de Microsoft, Excel 2002, es capaz a medias de tomar los avances de Microsoft OLAP Server. Microsoft adquirió un producto más sofisticado Data Analyzer, pero está más dirigido a las necesidades de los especialistas en análisis, no siendo una herramienta de propósitos generales.

Microsoft no ha capturado todo el mercado OLAP, y se espera que este monopolio nunca se imponga. Pero sus competidores deben hacer mejor un trabajo presentado opciones viables, que son en algunos aspectos hoy en día más funcionales que las de Microsoft. Sin ir más lejos, ni IBM ni Oracle han respondido bien. IBM DB2 OLAP Server, una versión de Essbase, fue un fracaso técnico con ventas modestas que fueron decayendo con el paso del tiempo, mientras que Oracle se ha conducido en el área de forma incorrecta. Un grupo de productos basados en el viejo Express salieron para el nuevo 9i, esta generación está más integrada con RDBMS.

Esto ocurrió en octubre del 2000, lo que produjo una disminución de las ventas de sus viejos productos ya probados. Pero esta primera versión en agosto del 2001 era incompleta e inútil, y los viejos productos no fueron reemplazados hasta el 2006. Más aún, los nuevos se han convertido en una desilusión. Siendo muy complejos y con pocas ventajas sobre los viejos que Oracle había comprado diez años antes.

Esto no solo ha perjudicado a Oracle en las ventas sino que también ha llevado a las manos de Microsoft muchos sitios que estaban en las de Oracle y Express. Finalmente, Oracle ha

terminado todos los componentes necesarios para hacer su nuevo servidor OLAP completamente viable.

2.3.8.2 OLAP y el Open Source

La utilización de soluciones Open-Source han aumentado inmensamente en años recientes, pero en el auge del mercado de análisis extrañamente no ha mantenido el paso. Existe alguna variedad de productos los cuales pueden ser encontrados buscando en SourceForge.net.

Desafortunadamente, muchos de estos intentos carecen de la calidad requerida para su aplicación en el mundo de la Inteligencia de Negocios (BI) y el análisis. Sin embargo, si prometen un buen futuro para el Open-Source en esta área.

Primero que todos en esta lista de productos Open Source para la inteligencia de negocios se encuentra Mondrian, un servidor OLAP. Surgió en el 2002 y entre sus características se pueden mencionar que soporta agregaciones, cálculos y categorización de datos provenientes de bases de datos relacionales, incluyendo Oracle, MySQL, Microsoft SQL Server, IBM DB2, entre otras.

Usa como editor gráfico de modelos OLAP el popular IDE Elipse. También implementa MDX el lenguaje estándar de consultas OLAP, y trabaja con XML for Analysis al igual que con la especificación JOLAP. Presenta como interfaz a los usuarios finales otro proyecto llamado JPivot, el cual es una librería JSP que renderiza una tabla OLAP y permite a los usuarios las navegaciones típicas OLAP (*slice, dice, drill down y roll up*).

En noviembre del 2005 Pentaho, un proveedor de soluciones Open Source, agregó a Mondrian a su plataforma BI, convirtiéndose así Mondrian en una ventaja más de la plataforma, la cual provee reportes, análisis, dashboards, workflow, y data mining. Pentaho está obligado a brindar soporte para el usuario final de Mondrian y las comunidades de desarrollo, y garantizar que permanezca como un producto libre. Usuarios que anteriormente fueron forzados a pagar por un producto tradicional de BI ahora tiene la oportunidad poseer una plataforma de análisis robusta, a un precio muy bajo, utilizando bases de datos relacionales como MySQL junto con Pentaho OLAP Server.

2.4 Operadores OLAP

Lo interesante no es poder realizar consultas que, en cierto modo, se pueden hacer con selecciones, proyecciones, concatenaciones y agrupamientos tradicionales.

Lo realmente interesante de las herramientas OLAP son sus operadores de refinamiento y manipulación de consultas como los que vimos en la sección 2.3.4 y a continuación se enumeran algunas ellas:

- DRILL , ROLL , SLICE & DICE , PIVOT

Aunque en la sección 2.3.4 ya se detallaron los conceptos de estos operadores, en la presente se describe un resumen de estos, pues es fundamental conocerlos para tener claro en nuestra mente lo que una herramienta OLAP nos debe proporcionar.

- **Roll up (dril-up):** resumir los datos. Subir en la jerarquía o reducir las dimensiones.
- **Drill Down (roll-down):** el contrario del anterior. Bajar en la jerarquía o introducir nuevas dimensiones.
- **Slice and Dice:** Selección y proyección.
- **Pivot (rotar):** Reorientar el cubo.
- **Drill:** Se utilizan las coordenadas dimensionales especificadas por un usuario para una celda en un cubo para moverse a otro cubo a ver información relacionada.
- **Drill across:** implica utilizar más de una tabla de hechos.
- **Drill through:** Ir desde el nivel de máximo detalle del cubo a sus tablas relacionales (Utilizando SQL).

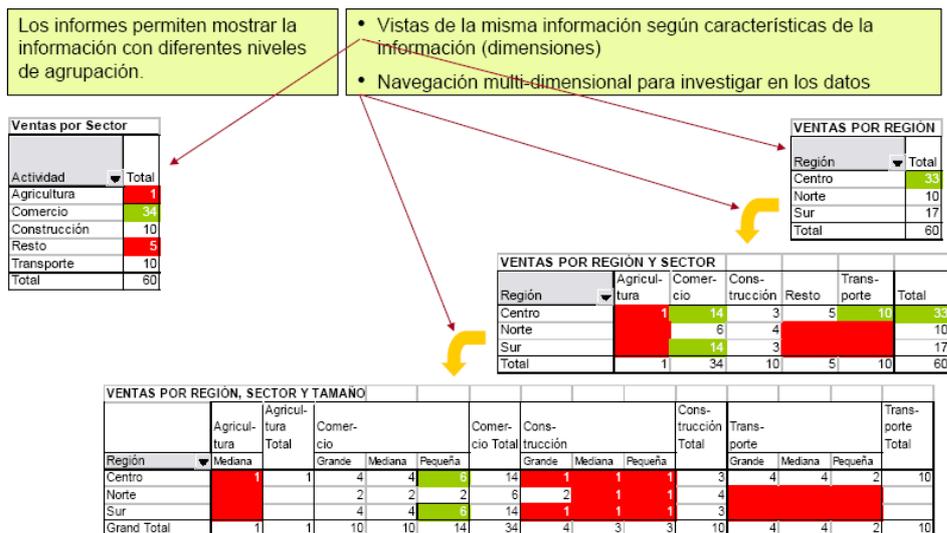


Figura II.10. Ejemplo de un Informe OLAP.

2.4.1 Características de las Herramientas OLAP

- Ofrecer una visión multidimensional de los datos (matricial).
- No imponer restricciones sobre el número de dimensiones.
- Ofrecer simetría para las dimensiones.
- Ofrecer operadores intuitivos de manipulación: drill-down, roll-up, slice-and-dice, pivot.
- Permitir definir de forma flexible (sin limitaciones) sobre las dimensiones: restricciones, agregaciones y jerarquías entre ellas.
- Ser transparentes al tipo de tecnología que soporta el almacén de datos (ROLAP o MOLAP).

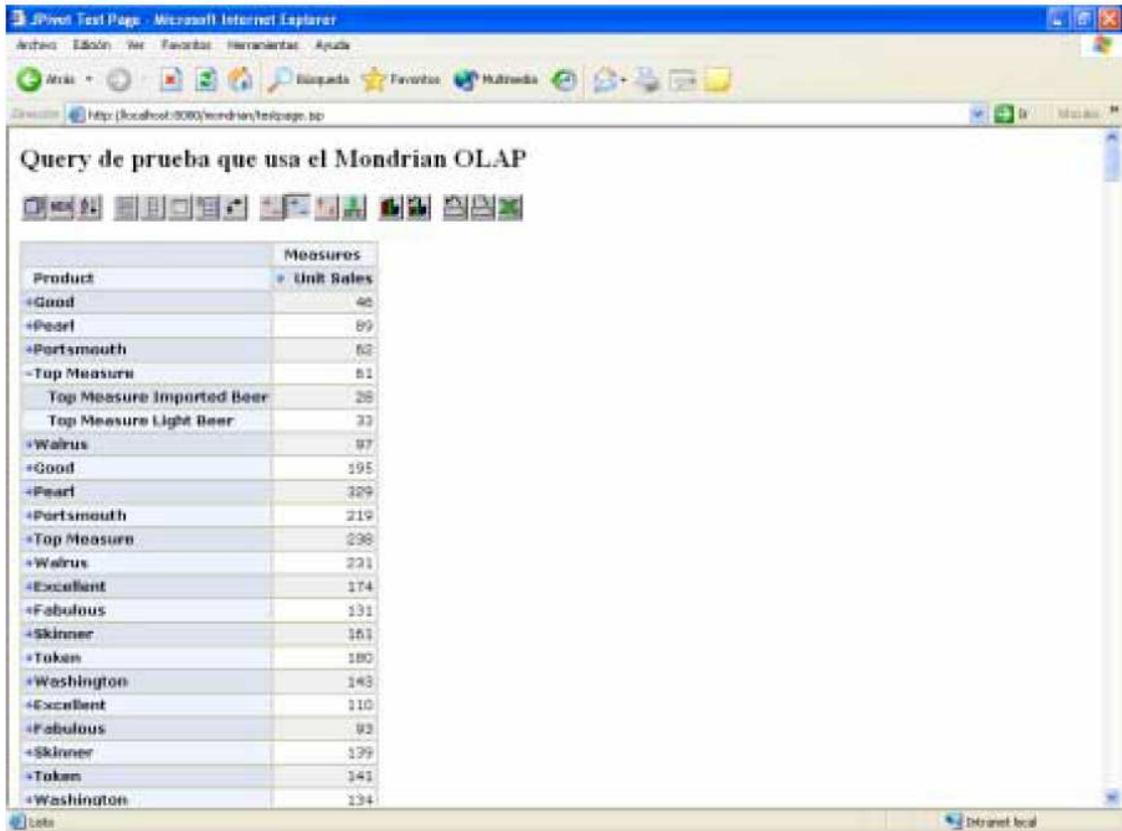
2.4.2 Componentes de las Herramientas OLAP

Las herramientas de procesamiento analítico en línea están integradas por dos componentes principales:

- El servidor OLAP
El cual permite la definición, creación y acceso a modelos de datos multidimensionales.
- La interfaz
La cual permite visualizar grandes volúmenes de datos y navegar sobre ellos interactivamente.

Para esto se utilizan tablas dinámicas como la que se muestra en la figura 11.

En este ejemplo podemos ver cuántas unidades se han vendido de diferentes marcas de productos, el reporte únicamente muestra las ventas de los primeros 3 meses del año. En el caso de la marca "*Top measure*" vemos que se han vendido 61 unidades, interactivamente se solicito ver las ventas de cada uno de los productos de esa marca, en este caso se muestra "*Top measure imported beer*" con 28 productos vendidos y "*Top measure light bee*" con 33 productos vendidos.



The screenshot shows a web browser window titled "JPIVOT Test Page - Microsoft Internet Explorer". The address bar shows "http://localhost:8080/mondrian/testpage.jsp". The main content area displays the title "Query de prueba que usa el Mondrian OLAP" and a table of data. The table has two columns: "Product" and "Measures". The "Measures" column is expanded to show "Unit Sales". The data is organized into several groups, each with a plus sign in the "Product" column. The groups and their values are: "Good" (46), "Pearl" (89), "Portsmouth" (62), "Top Measure" (81), "Top Measure Imported Beer" (28), "Top Measure Light Beer" (33), "Watus" (97), "Good" (195), "Pearl" (329), "Portsmouth" (219), "Top Measure" (238), "Watus" (221), "Excellent" (174), "Fabulous" (131), "Skinner" (161), "Token" (180), "Washington" (143), "Excellent" (110), "Fabulous" (93), "Skinner" (139), "Token" (141), and "Washington" (134).

Product	Measures
+Good	46
+Pearl	89
+Portsmouth	62
+Top Measure	81
Top Measure Imported Beer	28
Top Measure Light Beer	33
+Watus	97
+Good	195
+Pearl	329
+Portsmouth	219
+Top Measure	238
+Watus	221
+Excellent	174
+Fabulous	131
+Skinner	161
+Token	180
+Washington	143
+Excellent	110
+Fabulous	93
+Skinner	139
+Token	141
+Washington	134

Figura II.11. Interfaz de una herramienta OLAP

El procesamiento analítico en línea utiliza el concepto de cubos como ya lo habíamos explicado en secciones anteriores, para manipular y visualizar la información. La tecnología de procesamiento en línea en general se refiere a las herramientas de reporte y graficado que permiten desplegar información agregada, permiten detectar alguna o algunas variables de interés y descender hacia los datos de detalle. Estos reportes permiten identificar semáforos o indicadores que posteriormente pueden analizarse con cuidado para identificar causas de los mismos o tendencias que puedan estar implicando alguna situación favorable o desfavorable para la organización.

El proceso consiste en (1) crear un subconjunto de datos estructurado en un modelo de datos con características especiales para alto desempeño, normalmente se utilizan modelos dimensionales [Chaudhuri 1997] también conocidos como estrellas o cubos [Kimball 1996, 1997, 2003] y un conjunto de operaciones sobre ellos y (2) crear la interfaz de usuario, es decir, el reporte o gráfico que permiten visualizar el resultado de las operaciones sobre este modelo dimensional indicando las jerarquías y dimensiones por las cuales se podrá llevar a cabo la navegación sobre las dimensiones deseadas.

Un buen ejemplo de este tipo de software es Mondrian, un servidor de procesamiento en línea que incluye todas las funcionalidades previamente mencionadas y es el único prototipo cuyo código fuente está disponible para experimentación.

En el presente documento de tesis de grado presentaremos un estudio comparativo de dos herramientas Open Source de procesamiento analítico en línea, los cuales son denominados PALO y Mondrian.

Luego de haber dado una introducción y algunos detalles acerca del análisis multidimensional OLAP a continuación se muestra la información teórica necesaria de las dos herramientas anteriormente mencionadas para en el siguiente capítulo enfatizar la comparación de estas.

CAPÍTULO III

3. HERRAMIENTAS OLAP OPEN SOURCE

3.1 JUSTIFICACIÓN

Las herramientas OLAP de estudio son: Mondrian y Palo mencionadas anteriormente, en este capítulo analizaremos los puntos teóricos más importantes acerca de cada una de estas herramientas.

Por otro lado es necesario indicar que ambas herramientas son de análisis multidimensional Open Source que como su naturaleza misma lo indica, se las puede conseguir de manera totalmente gratuita, cumpliendo de esta forma con el justificativo general de este proyecto de tesis que es estudiar herramientas OLAP de código abierto.

Además son las más utilizadas mundialmente en varias plataformas, refiriéndonos al área de Business Intelligence (**BI**) de código abierto para la toma de decisiones empresariales. En el caso de Palo se utiliza en el proyecto de JEDOX la empresa que promueve su utilización. Por otro lado el servidor OLAP Mondrian es más utilizado en varias plataformas de renombre como Pentaho en donde se adoptó un nuevo nombre para Mondrian el cual es Pentaho Analysis (Mondrian - Jpivot), de la misma manera sucede en la suite BI JasperSoft.

A continuación se describe las herramientas empezando por Palo 2.5 y luego de ello se finaliza el presente capítulo con Mondrian 3.0.4, las cuales son las versiones actuales de cada uno de los mencionados proyectos de análisis multidimensional, de la siguiente manera:

3.2 PALO 2.5

El objetivo de esta sección es describir de forma exhaustiva las características y funcionalidades de PALO en su versión 2.0. Dicha sección intentará analizar el motor MOLAP desde el punto de vista técnico y sus posibles vinculaciones o integraciones con Pentaho.

3.2.1 Antecedentes y Visión General

La discusión entre las tecnologías ROLAP y MOLAP en el data warehousing ha generado desde siempre mucha polémica y mucha literatura. En esta sección no se pretende explicar de forma las diferencias entre ROLAP y MOLAP puesto que ya lo hicimos anteriormente. Sin embargo conviene tener claros los siguientes conceptos MOLAP:

Una arquitectura MOLAP almacena los datos en formato multidimensional, en ficheros binarios, generalmente no accesibles desde otras aplicaciones. El sistema MOLAP se basa en los cálculos en memoria y en la posibilidad de escritura en el mismo cubo. Generalmente presenta un mejor rendimiento en tiempo de ejecución, pero necesita más tiempo de procesado y carga de datos.

Otra de las ventajas de PALO 2.5 es que incorpora reglas que permiten cálculos complejos al estilo MDX con la ventaja que se pueden escribir en el mismo cubo permitiendo así hacer proyecciones de futuro sobre datos del pasado.

3.2.2 Arquitectura de Palo 2.5

La arquitectura de PALO es cliente servidor. En el siguiente diagrama se puede observar los diferentes componentes disponibles para PALO 2.5 así como las diferentes capas que componen su arquitectura. Palo es una herramienta escrita en C.

No todos los componentes disponibles son nativos de los fabricantes. Ha salido una implementación de la API de PALO en java (JPalo) que ha sentado las bases para futuros desarrollos sobre PALO como por ejemplo el cliente web de PALO para explotación de los cubos.

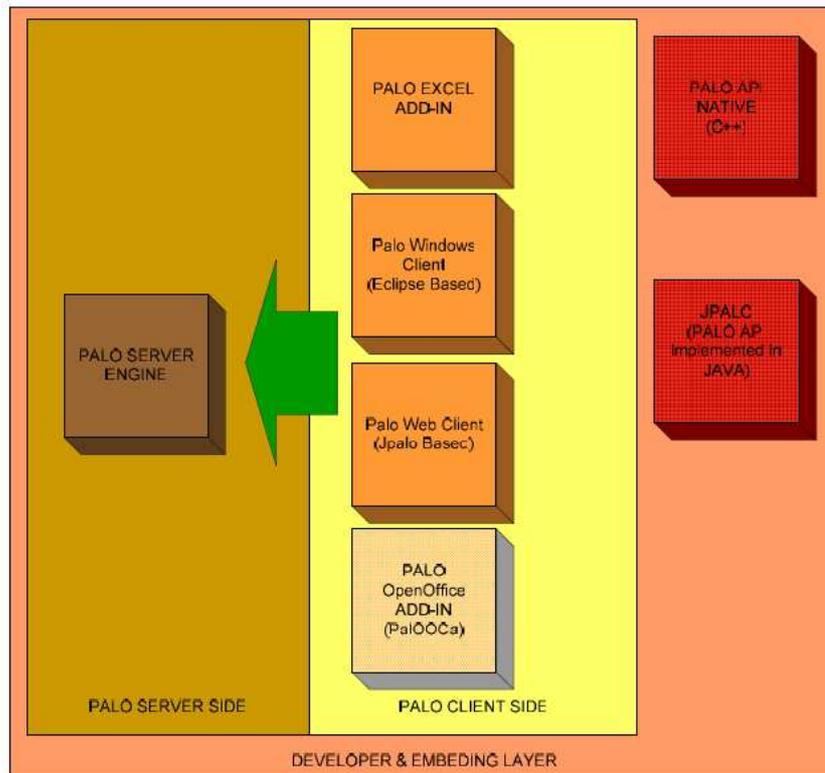


Figura III.12. Arquitectura de PALO 2.5

3.2.3 Instalación de Palo.

Para disponer de un entorno MOLAP operativo con PALO hay que realizar los siguientes pasos:

- Instalar o tener disponible en nuestra red un servidor PALO MOLAP. Podemos obtener el instalador de la web de PALO <http://www.jedox.com/>
- Instalar cualquiera de las herramientas cliente para PALO, estas herramientas permiten la creación de cubos para análisis multidimensional:

PALO Excel Add-in. Incluido con el servidor de PALO, descargable de la web: <http://www.jedox.com/>

PALO Client. Se puede descargar de la web de JPALO: <http://www.jpalo.net/download.htm>

- Instalar el PALO web Client que de igual forma se puede descargar de <http://www.jpalo.net/download.htm>, esta herramienta permitirá publicar en el servidor web de palo los reportes de análisis, pero se debe tomar en cuenta que será necesario instalar primero el servidor web, que será el mismo que utilizará Mondrian para publicar

sus reportes, este es el Apache Tomcat que para Palo es necesario desde la versión 5.5 en adelante, además es necesario conocer que este servidor web se puede descargar de internet ya que es totalmente gratuito.

Crear un entorno de datos MOLAP o cargar una base de datos MOLAP existente. Se puede descargar la base de datos demo desde: <http://worksheetserver.com/download/zip/Demo2.zip> para realizar las pruebas necesarias.

3.2.4 Primeros pasos con Palo

Para la creación de cubos se puede utilizar tanto el Excel addin como el Palo Client. El Excel addin incluye algunas funcionalidades que pueden resultar cómodas para trabajar con los datos.

Las diferentes opciones del plugin son:

Paste View: Permite Configurar hojas del Excel para rellenar datos en el cubo de forma fácil y rápida.

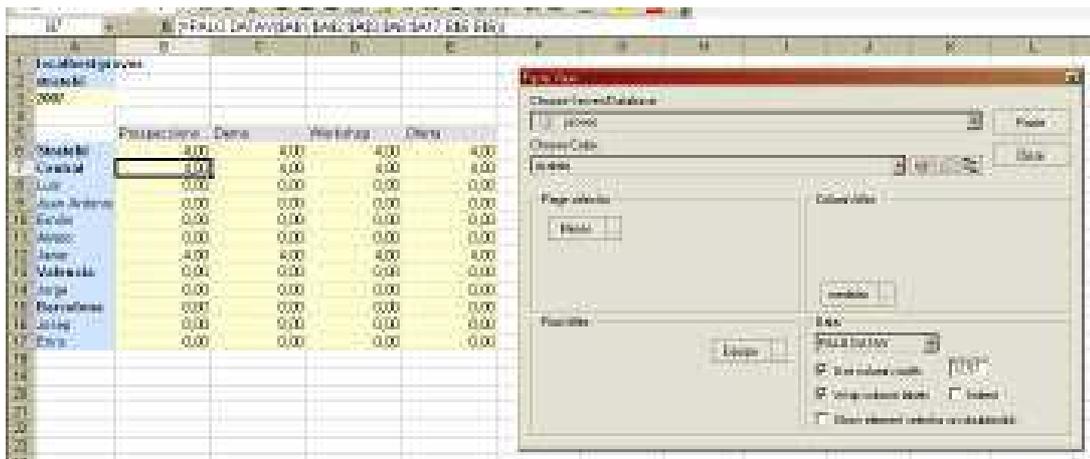


Figura III.13. PALO 2.5: Paste View

Paste Elements: Permite insertar los valores de los elementos de las dimensiones en las hojas Excel para trabajar con ellos.

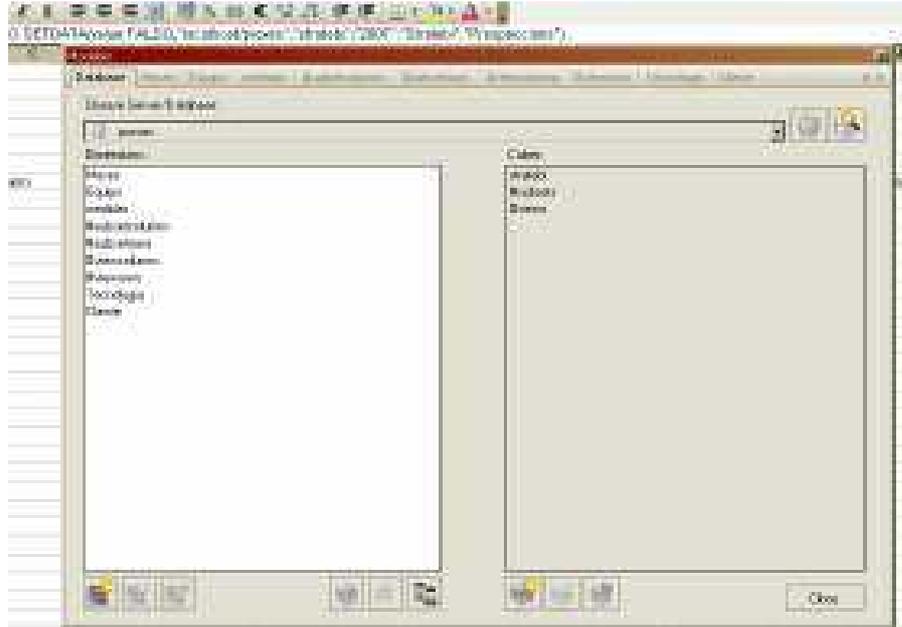


Figura III.16: PALO 2.5: Modeller

Data Import: Permite Importar datos desde orígenes de datos distintos (CSV,ODBC,MOLAP,TXT)

Save as Snapshot: Guarda una instantánea de la base de datos en su estado actual para recuperarla en cualquier momento.

Palo Wizard: Asistente para crear, borrar o modificar bases de datos. Permite también gestionar las instancias de Palo Server.

Los pasos necesarios para construir un cubo en PALO son:

- 1.- Crear o conectarse a una base de datos.
- 2.- Crear dimensiones, consolidaciones y jerarquías
- 3.- Crear hechos
- 4.- Definir cubos
- 5.- Rellenar Datos

Todos ellos muy sencillos e intuitivos con las herramientas del Excel Addin.

3.2.5 Sobre las Reglas de Palo 2.5

Palo 2.5 incorpora un sistema de reglas o formulas avanzadas para trabajar con los elementos de los cubos. Para editar las reglas utilice Rule Editor del Excel Addin.

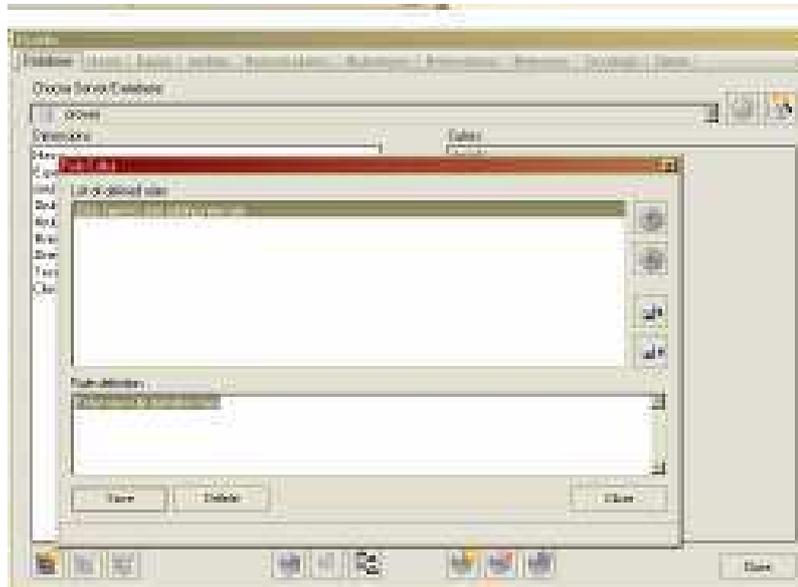


Figura III.17. Reglas de PALO 2.5

Las reglas de negocio permiten realizar cálculos avanzados y operaciones de lectura y escritura de celdas de los cubos. Se puede establecer un paralelismo con las fórmulas del lenguaje MDX que permiten acceder a otros cubos, miembros de otras jerarquías, etc. La gran diferencia es que el sistema MOLAP combinado con las reglas de PALO permite hacer proyecciones de futuro y escribir los resultados de dichas reglas en celdas del cubo.

3.2.6 Usuarios y Permisos

Palo incluye también la implementación de un sistema de seguridad para la explotación de los cubos. Permite crear:

- Usuarios
- Roles
- Permisos sobre objetos

De esta forma se pueden controlar los accesos a los objetos de la base de datos así como manejar diferentes roles según los requerimientos de cada usuario (lectura, escritura, control total) Se puede también filtrar la visibilidad a subconjuntos de datos, etc... Se puede gestionar también con el Modeller del Excel Addin.

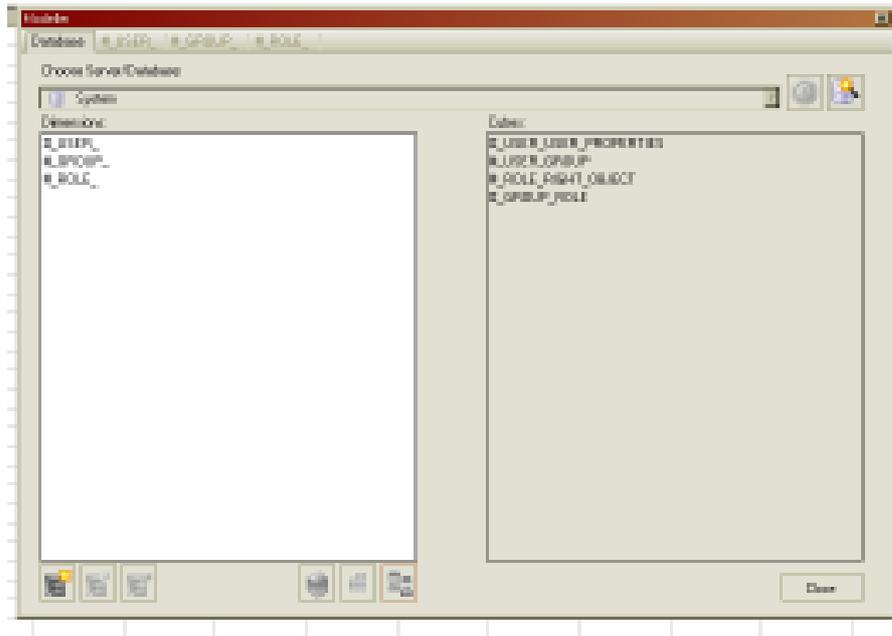


Figura III.18. PALO 2.5: Usuarios y Permisos

3.2.7 Soporte y Comunidad

Palo es un producto de la empresa jedox. Es un producto Open Source sin embargo la comunidad que le da soporte es muy reducida. La empresa cobra para obtener los manuales completos de la aplicación y tiene servicios como el Excel Enterprise Spreadsheets que permite hacer dashboards con el Excel para explotarlos vía web.

La empresa Tensegrity software ha implementado una API en java de PALO y también un cliente java de palo basado en eclipse. Todo Open Source.

3.3 MONDRIAN 3.0.4

3.3.1 Antecedentes y Visión General

La herramienta open source llamada Mondrian, es un servidor de procesamiento analítico en línea desarrollado para uso gratuito y experimental.

Este servidor ofrece los servicios de creación y administración de modelos dimensionales así como los de acceso e interacción con los mismos. Mondrian, ahora rebautizado como Pentaho Analysis Services, es el motor OLAP integrado en la suite de Business Intelligence Open Source Pentaho.

Además está licenciado bajo la Mozilla Public License (MPL). Esta licencia es una de las "Business Friendly" lo cual implica que es de las menos restrictivas para su uso desde la mayor parte de los puntos de vista (al igual que el resto de la suite de Pentaho), permitiendo Modificar, Embeber, Modularizar, el software sin restricciones; dejando al parecer de la organización el aporte o no de los cambios realizados al proyecto.

La página actual del proyecto (junto con la documentación) es:

<http://Mondrian.pentaho.org/>

Mondrian es un motor ROLAP con caché, lo cual lo sitúa cerca del concepto de HOLAP (HYBRID OLAP).

ROLAP significa que en Mondrian no residen datos (salvo en la caché) sino que estos residen en un Sistema de Gestión de Bases de Datos externo, que para este trabajo de tesis se ha utilizado SQL Server 2005.

Es en esta base de datos en la que residen las tablas que conforman la información multidimensional con la que Mondrian trabaja (los modelos en estrella de nuestros Datamarts por ejemplo). MOLAP en cambio es el nombre que reciben los motores OLAP en los que los datos residen en una estructura dimensional.

Este servidor OLAP se encarga de recibir consultas dimensionales (lenguaje MDX) y devolver los datos de un cubo, solo que este cubo no es algo físico sino un conjunto de metadatos que definen como se han de "mapear" estas consultas que tratan conceptos dimensionales a sentencias SQL ya tratando con conceptos relacionales que obtengan de la base de datos la información necesaria para satisfacer la consulta dimensional.

Alguna de las ventajas de este modelo son:

- El no tener que generar cubos estáticos ahorrando lo que cuesta generarlos y la memoria que ocupan.
- La posibilidad de utilizar siempre los datos residentes en la base de datos, de forma que se trabaja con datos actualizados. Muy útil en entorno de BI Operacional.
- Pese a que tradicionalmente los sistemas MOLAP tienen una cierta ventaja de rendimiento, la aproximación híbrida de Mondrian, el uso de caché y de tablas agregadas, hace que se puedan obtener muy buenos rendimientos, sin perder las ventajas del modelo ROLAP clásico. Es muy importante aprovechar bien las ventajas de las bases de datos donde residen las tablas.

3.3.2 Arquitectura de Mondrian 3.0.4

Un servidor como Mondrian consiste de cuatro capas: la capa de presentación, la capa de cálculo, la capa de agregación y la capa de almacenaje.

La capa de presentación, es el visualizador y determina que es lo que el usuario final verá en su monitor, y como puede interactuar para realizar operaciones sobre lo que ve. Existen muchas formas de presentar conjuntos de datos multidimensionales, incluyendo tablas pivote (como las que se pueden desarrollar con MS Excel), gráficas de pastel, diagramas de barras y líneas, y herramientas avanzadas de visualización como mapas "clickables" y gráficas dinámicas como el árbol hiperbólico [Shneiderman 1998].

Esas herramientas pueden escribirse en Swing o JSP, los gráficos se pueden poner en formato JPG o GIF, o ser transmitidos a una aplicación remota vía XML. Lo que tendrán en común todas las formas de presentación es la "gramática" multidimensional de las dimensiones, medidas y celdas en las cuales la capa de presentación hace la pregunta y el servidor de datos de procesamiento analítico en línea regresa la respuesta. Para Mondrian se ha desarrollado un prototipo de interfaz llamado Jpivot, un ejemplo de esta interfaz se muestra en la Figura III.11.

La segunda capa es **la capa de cálculos**, esta capa hace un análisis sintáctico, valida y ejecuta las consultas escritas en un lenguaje de consultas que el servidor de datos pueda entender, Mondrian utiliza un lenguaje llamado MDX ("Multi-Dimensional eXpressions") , cuya sintaxis es similar a la del lenguaje SQL.

Una consulta sobre un modelo dimensional se evalúa en varias fases. Los ejes se calculan primero y luego los valores de las celdas dentro de los ejes. Por eficiencia, la capa de cálculos envía las peticiones de celdas a la capa de agregación en lotes.

Los metadatos describen el modelo dimensional y como se hace el mapeo hacia el modelo de datos relacional, esto porque Mondrian es un servidor "ROLAP" que utiliza un administrador de base de datos relacional como MySQL para albergar los datos. La descripción de los metadatos se realiza en XML siguiendo unas reglas para la construcción de esquemas definidas por los autores de Mondrian.

En resumen esta capa recibe instrucciones en MDX y las traduce a SQL para que puedan ser ejecutadas en la base de datos relacional, hace uso de los metadatos para saber cuáles campos de las tablas relacionales corresponden a dimensiones o hechos.

La tercera capa es **la capa de agregación**, una agregación es un conjunto de valores de medidas (celdas) en memoria (cache), calificadas por un conjunto de valores de columnas (dimensiones). La capa de cálculo recibe solicitudes de conjuntos de celdas. Si las celdas solicitadas no están en cache, o no son derivables por un cálculo de una agregación en cache, el administrador de la capa de agregación envía una petición a la capa de almacenaje.

La capa de almacenaje, es un RDBMS en el caso de Mondrian, o puede ser un repositorio de datos desarrollado para funciones de almacenamiento y recuperación de información con características similares a un DBMS. Esta capa es responsable de proporcionar datos agregados en celdas y miembros de las tablas de dimensión. Es conveniente usar un RDBMS como lo hace Mondrian, sin embargo, es posible desarrollar un sistema de almacenamiento optimizado para datos multidimensionales.

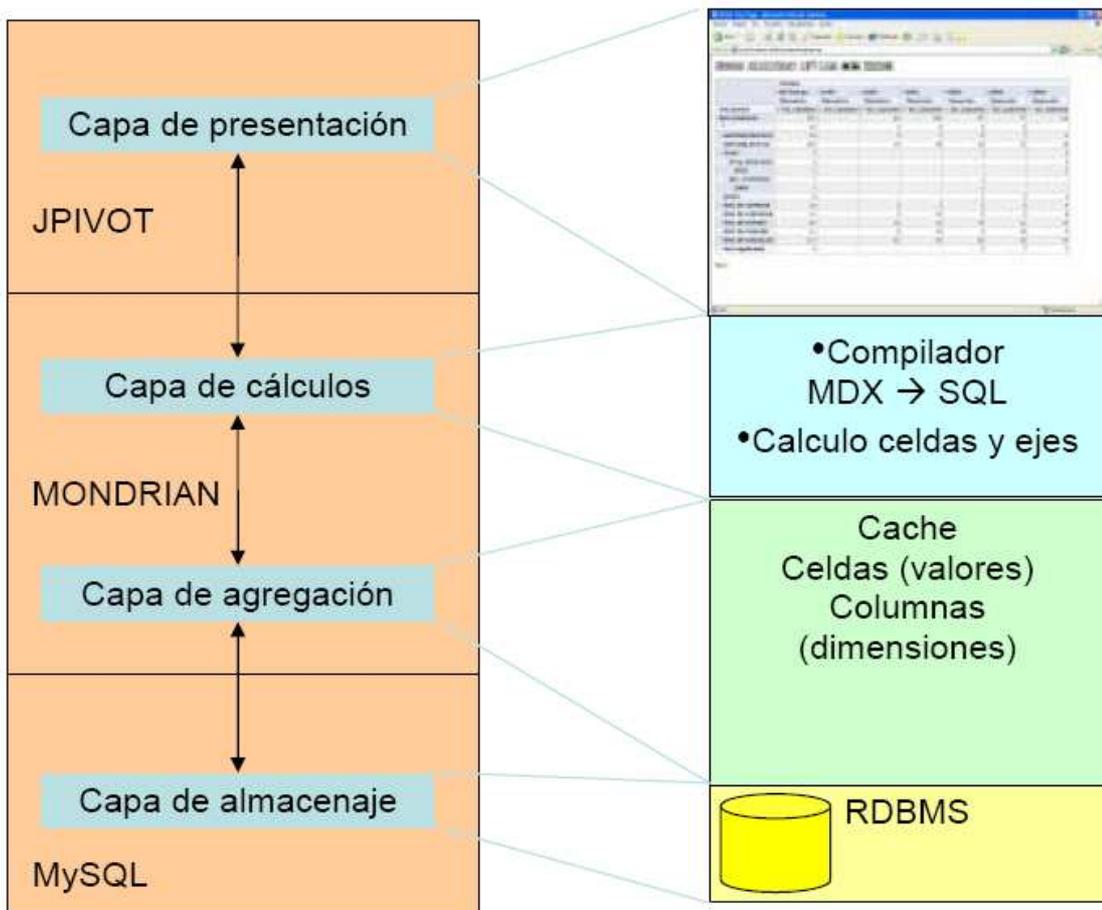


Figura III.19. Arquitectura del servidor de procesamiento analítico en línea Mondrian.

Los componentes antes mencionados pueden convivir en la misma máquina o pueden distribuirse entre máquinas. Las capas 2 y 3, las cuales forman a Mondrian, pueden estar en la misma máquina.

La capa de almacenaje puede estar en otra máquina y ser accesada usando una conexión remota con JDBC. En un sistema multiusuario, la capa de presentación puede existir en cada máquina de cada usuario final (excepto en el caso de páginas **JSP** o **ASP** generadas en el servidor como en nuestro caso)

Los agregados en Mondrian se administran con un cache de memoria. Ese cache guarda agregaciones pre-calculadas en memoria de tal forma que las consultas siguientes puedan acceder a los valores de celdas sin ir al disco. Si el cache almacena los datos requeridos en un nivel bajo de agregación entonces podrá calcular los datos requeridos haciendo operaciones de generalización. Es difícil elegir un conjunto de agregaciones a pre-calculas las cuales le den mayor velocidad al servidor sin usar grandes cantidades de espacio en disco, particularmente las de alta dimensionalidad o si los usuarios están lanzando consultas impredecibles.

En un sistema donde los datos están cambiando en tiempo real, es impráctico mantener los agregados pre-calculados.

Un cache de tamaño razonable puede permitir a un sistema desempeñarse adecuadamente al enfrentarse a consultas impredecibles, con pocos o sin agregados pre-calculados.

Mondrian realiza las agregaciones de datos de la manera siguiente:

- Los datos de hechos son almacenados en un RDBMS.
- Lee datos agregados en el cache lanzando consultas con "GROUP BY".
- Si el RDBMS soporta vistas materializadas y se decide crear vistas materializadas para algunas agregaciones, entonces el servidor de procesamiento analítico en línea las utilizara implícitamente. El manejador de agregaciones debe ser capaz de darse cuenta de que esas vistas materializadas existen y que esas agregaciones son más baratas de calcular.

La idea general de los desarrolladores de Mondrian fue delegar a la base de datos todas las operaciones que pueda realizar. El almacenamiento multidimensional puede reducir I/O, pero desde el punto de vista de los autores de Mondrian los beneficios no son muchos.

Mondrian proporciona una API que sirve para que las aplicaciones cliente que se desarrollen puedan ejecutar consultas sobre los modelos de datos dimensionales. En el caso de Mondrian nos ofrece la posibilidad de hacerlo vía JDBC y usando Java.

El lenguaje de consultas del prototipo Mondrian es llamado MDX, se requiere un JDBC que puede manejar SQL. Un programa que utiliza el API crea una conexión a la base de datos multidimensional de manera similar a una conexión con JDBC.

Una consulta en MDX se realiza de manera similar a una instrucción JDBC y se crea compilando una instrucción en MDX. El resultado es similar a un "ResultSet" de JDBC. Debido a que estamos tratando con datos multidimensionales, la consulta en MDX consiste de ejes y celdas, en lugar de columnas y filas de un SQL normal.

3.3.3 Instalación de Mondrian 3.0.4

Para trabajar con Mondrian 3.0.4 primero se lo debe bajar de la página web que anteriormente se mencionó, como la siguiente: <http://Mondrian.pentaho.org/>

Una vez tengamos la herramienta, se procede a instalar en el orden siguiente:

- 1) La plataforma java sobre la cual se trabaja, en el caso de esta tesis es la jdk-6u7-windows-i586-p que se lo puede descargar de internet.

- 2) Posteriormente el servidor Web sobre el cual se trabaja, en nuestro caso se utiliza el apache-tomcat-6.0.16, que igual se lo puede descargar de internet.
- 3) Si ya tenemos listo e instalado el DBMS sobre el cual se va a trabajar proceder con la instalación de Mondrian 3.0.4, en nuestro caso se está utilizando como DBMS el SQL Server 2005.
- 4) Procedemos a la instalación de Mondrian 3.0.4, donde una vez obtenido de internet el paquete del servidor OLAP, lo descomprimos y por medio de la consola de administración de Apache Tomcat ejecutamos el archivo Mondrian.war.
- 5) Y por último en la carpeta donde se instaló Apache Tomcat 6.0.16 ya sea en Windows o Linux procedemos a grabar debajo de la carpeta /lib, el sqjjdbc para trabajar con las bases de datos de SQL Server 2005 en este caso.

Luego de esto ya tenemos listo nuestro entorno de trabajo. Siendo importante mencionar que Mondrian trabaja desde jdk-1.4 en adelante, apache-tomcat-5.0.27 en adelante.

Para crear y modificar los esquemas de cubos de Mondrian se puede utilizar la herramienta "Schema Workbench", Cube Designer de Pentaho. O de otra manera algún IDE de desarrollo NetBeans 6.1, Eclipse u otros.

3.3.4 Funcionalidades de Mondrian 3.0.4

En general Mondrian ofrece las siguientes funcionalidades:

- 1) Cuenta con un compilador y ejecutor de instrucciones en lenguaje MDX. Lo que hace el compilador es traducir el MDX al SQL del manejador de base de datos relacional que se está usando.
- 2) Los datos del modelo dimensional deben ser documentados en XML para ser reconocidos por Mondrian, formando con esto una especie de metadatos.
- 3) Utiliza JDBC para la conexión a los RDBMS.
- 4) Los resultados de las consultas sobre modelos dimensionales son almacenados en una memoria caché administrada por Mondrian.
- 5) Mondrian trata de calcular los agregados con los datos que se encuentran en la cache como técnica para mejorar el desempeño de las consultas.

3.3.5 Flujo de Ejecución de Mondrian 3.0.4

Veamos para que quede más claro un ejemplo de un flujo de ejecución de una consulta MDX contra Mondrian:

1 – Un cliente (por ejemplo la interfaz web JPivot) lanza una consulta MDX contra Mondrian, solicitando una serie de datos y hablando de concepto dimensionales (ej. “Quiero el gasto del último año para todas las provincias”).

2 – Mondrian recibe la sentencia MDX (referida a un cubo en concreto) busca en sus metadatos (esquema de Mondrian, un fichero xml que define cubos) que conceptos relacionales (tablas, columnas) se asocian con estos conceptos dimensionales. Busca si ya tiene esos datos en la caché (los obtiene muy rápidamente), si los tiene los devuelve al interfaz, sino ejecuta el siguiente paso.

3 - Genera las sentencias SQL necesarias (mirando en su esquema con la definición del cubo) para obtener esos datos. (ej. Una consulta SQL que obtiene los nombres de todas las provincias, otra que obtiene los gastos asociados a cada provincia ya agregados, etc...)

4 – La base de datos ejecuta las sentencias SQL (paso que más tiempo consume del proceso) y devuelve los datos a Mondrian

5 – Mondrian almacena los datos recibidos en la cache (para agilizar posteriores consultas).

6 – Mondrian devuelve el resultado a la interfaz (al cliente que lo solicitó, por ejemplo JPivot).

Veamos en cualquier caso un diagrama explicativo:

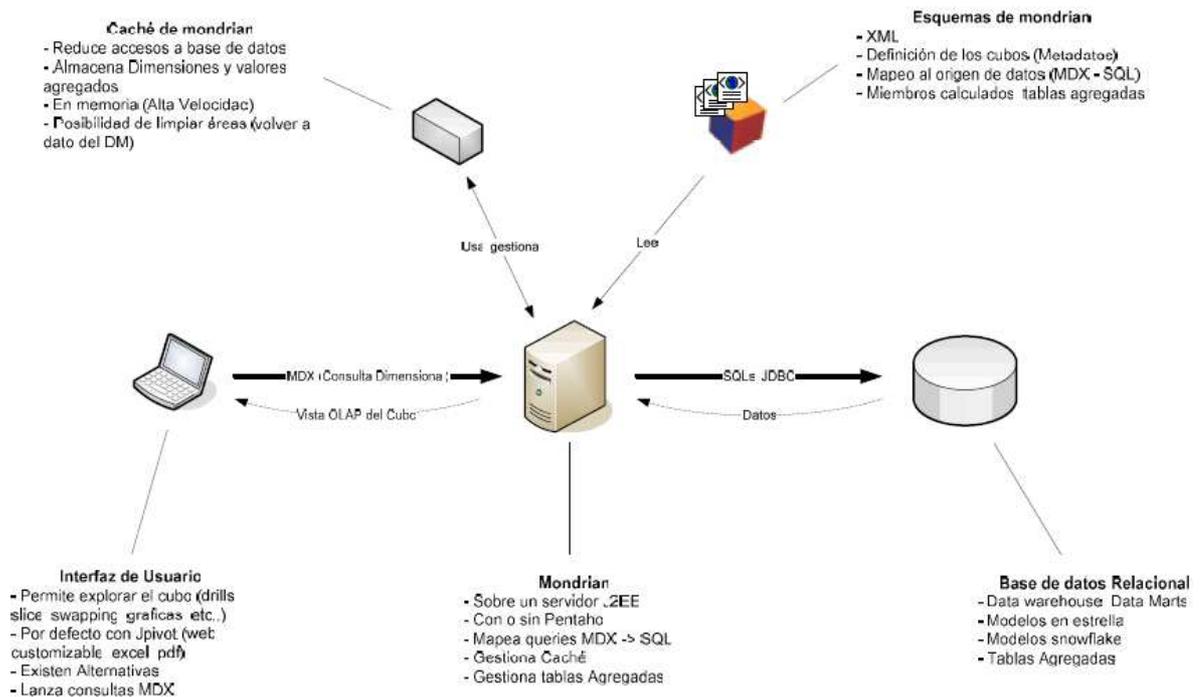


Figura III.20. Funcionamiento de Mondrian

3.3.6 Definición del Esquema Multidimensional XML

Como se ha explicado ya en el flujo de ejecución de Mondrian, es necesario definir el esquema multidimensional en un formato XML, teniendo el usuario desarrollador la posibilidad de definirlo en cualquier editor de documentos XML que como mínimo será en un bloc de notas.

Por otro lado se debe mencionar que la Plataforma Business Intelligence denominada Pentaho la misma que utiliza Mondrian como servidor OLAP, ha desarrollado un proyecto que facilite al usuario la creación de estos esquemas XML, y se denomina Schema Workbench en la cual se define de manera un poco mejor por decirlo así, los cubos OLAP con sus respectivas medidas, dimensiones, jerarquías y niveles.

A continuación se mostrará las interfaces de desarrollo de cubos OLAP que ofrece el software Schema Workbench

3.3.6.1 Definición del driver de conexión

En Schema Workbench hay que definir la conexión a la base de datos sobre la cual se realizará el análisis OLAP, la misma que es JDBC.

En el menú “Tools” se debe escoger la opción “Preferences” para definir la conexión a la base de datos. A continuación se muestra un ejemplo de ello:

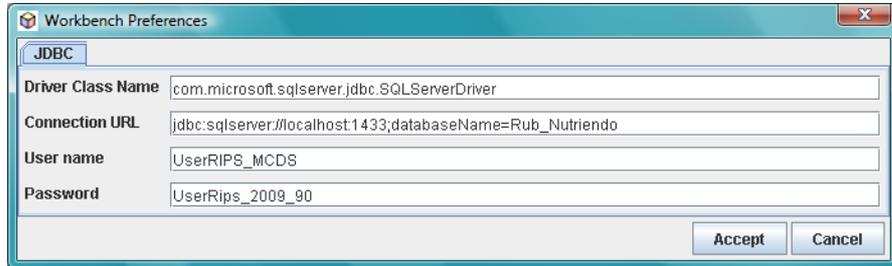


Figura III.21. Definición de la conexión a la Base de Datos en Workbench

Luego se define el cubo OLAP, escogiendo “Schema” del menú “File” y de la opción “New”, para obtener un cuadro de diálogo muy intuitivo por cierto para el desarrollo del cubo. A continuación se muestra un ejemplo de ello:

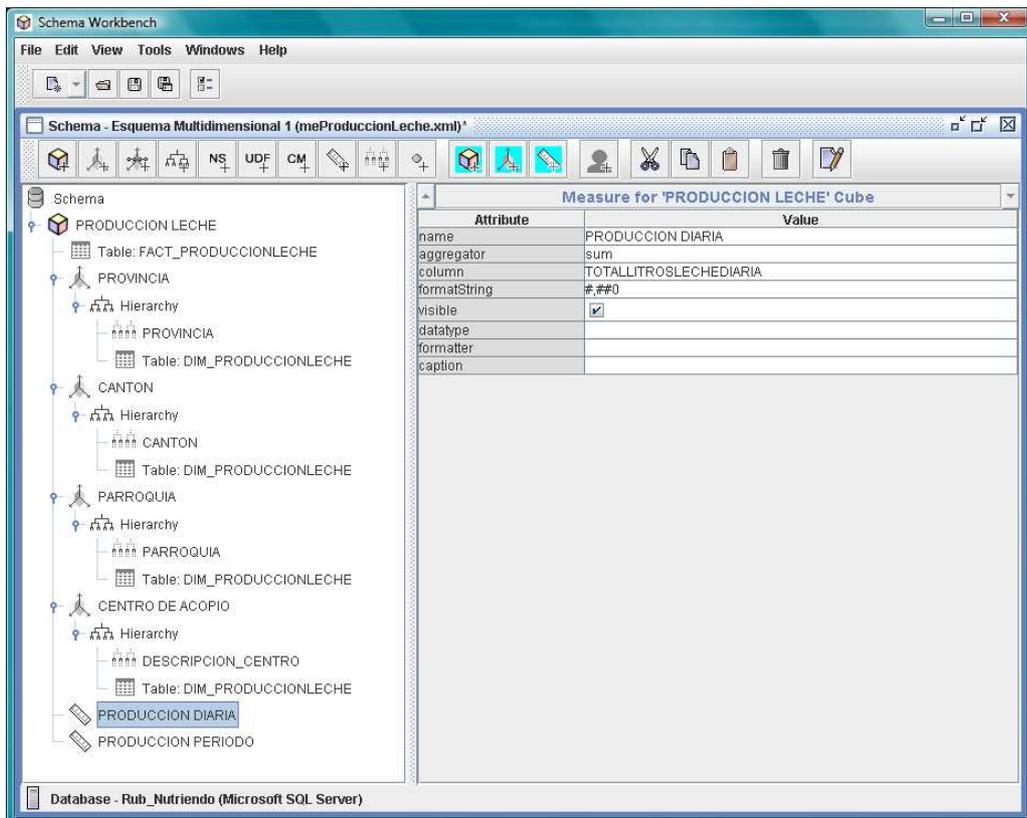


Figura III.22. Desarrollo de un cubo OLAP en Workbench

3.3.7 Expresiones Multidimensionales

Las expresiones multidimensionales MDX son las que se han mencionado en el flujo de ejecución de Mondrian, las mismas que son sentencias parecidas al lenguaje transact-sql de las bases de datos relacionales, sino que la diferencia radica en que este lenguaje MDX es para el análisis multidimensional.

El funcionamiento ya se explicó en la sección 3.3.5 del flujo de ejecución, ahora en este apartado se mostrará ciertos aspectos sobre este lenguaje.

La estructura detallada de una consulta MDX se expone en el siguiente diagrama de ejemplo:

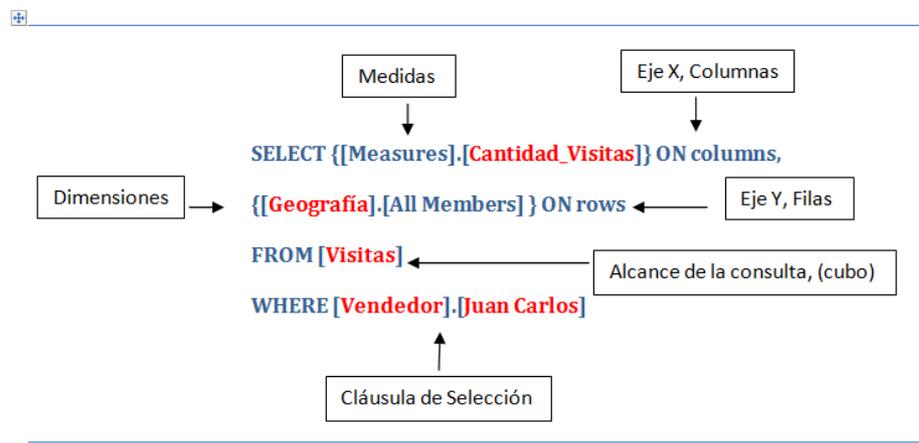


Figura III.23. Diagrama de la estructura básica MDX

Existe una similitud entre las consultas MDX y SQL, debido a que comparten sintaxis como lo son las cláusulas SELECT, FROM y WHERE. Sin embargo existen ciertos elementos que se distancian de SQL para dar soporte a las necesidades multidimensionales.

En MDX se identifican varios miembros, que cumplen un rol determinado por su posición en la jerarquía.

3.3.7.1 Children

La primera función es Children la cual, identifica el miembro o los miembros de un nivel por debajo del nivel inicial. Para usar la función Children se debe especificar el miembro, para que la función retorne todos los niveles por debajo de ella.

Ejemplo: Se tiene la siguiente consulta MDX:

```
Select {[Measures].[Valor]} ON columns,
```

Hierarchize ({{[Producto].[All Productos]}}) ON rows

from [cuboVentas]

De esta consulta se obtiene el siguiente resultado:

Measures	
Producto	Valor
All Productos	1,480,238,151

Figura III.24. Resultado de la consulta MDX Ventas de todos los productos

Que es el valor de las ventas de todos los productos por ejemplo. Ahora se utilizará la función children aplicada a la dimensión producto para comparar su resultado.

Ejemplo de consulta:

Select {[Measures].[Valor]} ON columns,

Hierarchize(Union({[Producto].[All Productos]}, [Producto].[All Productos].Children)) ON rows

from [cuboVentas]

Al ejecutar la anterior consulta se da el siguiente resultado:

Measures	
Producto	Valor
All Productos	1,480,238,151
Asesora	5,510,000
Capacitación	110,569,754
Desarrollo	1,068,367,852
Instalación	147,968,876
Soporte	88,168,533

Figura III.25. Resultado de la consulta MDX Ventas de todos los productos utilizando "Children"

Esta nueva consulta muestra el valor vendido de todos los productos como un total y también de cada uno, pues cada producto es hijo de la categoría Producto.

3.3.7.2 Count.

La función Count retorna el número de items de un cubo, por ejemplo las dimensiones, las tuplas, las celdas, niveles en una dimensión o una jerarquía.

Se tiene el siguiente ejemplo:

```
with member [Measures].[total] as 'count([Producto].[All Productos].Children)'

select {[Measures].[total]} ON columns

from [cuboVentas]
```

La función realiza un conteo de los componentes de Producto, que son cinco pues hay cinco categorías de productos en nuestro ejemplo anterior.

3.3.7.3 Avg.

La función Avg retorna el promedio de un número total de miembros. En la siguiente consulta de ejemplo se utiliza esta función.

```
with member [Measures].[total] as 'Avg([Producto].[All Productos].Children)'

select {[Measures].[total]} ON columns

from [cuboVentas]
```

3.3.7.4 Max.

La función Max retorna el máximo valor de un número total de miembros.

```
Max(«Set»[, «Numeric Expression»])
```

En el siguiente ejemplo se muestra esto:

```
with member [Measures].[total] as 'Max([Producto].[All Productos].Children)'

select {[Measures].[total]} ON columns

from [cuboVentas]
```

En el resultado se ve el valor de ventas de la categoría de producto con mayor valor en ventas.

3.3.7.5 *Min*

La función min retorna un mínimo valor de un número total de miembros.

```
Min({Set}[, {Numeric Expression}])
```

Para la cual se da el siguiente ejemplo:

```
with member [Measures].[total] as 'Min([Producto].[All Productos].Children)'  
  
select {[Measures].[total]} ON columns  
  
from [cuboVentas]
```

El resultado muestra el valor de la categoría de productos con menores ventas.

3.3.7.6 *Crossjoin*

Frecuentemente en MDX, los resultados de una consulta deberán necesitar más miembros de más de una dimensión.

```
Crossjoin({Set1}, {Set2})
```

Esta función retorna el producto cartesiano de distintos miembros de dos sets, que retorna todas las posibles combinaciones de esos miembros. El siguiente ejemplo muestra el uso de esta característica.

```
select      Crossjoin(Hierarchize(Union({[Ciudad].[All Ciudads]}, [Ciudad].[All  
Ciudads].Children)),  
  
{[Measures].[Valor]} ON columns,  
  
{[Producto].[All Productos]} ON rows  
  
from [cuboVentas]
```

Se hace un producto cartesiano entre el miembro Producto y Ciudad, para conocer el valor de las ventas de los productos en cada ciudad. Estas son las funciones más básicas del lenguaje de expresiones multidimensionales MDX.

CAPÍTULO IV

4. ESTUDIO COMPARATIVO

4.1 INTRODUCCIÓN

El presente capítulo tiene como objetivo fundamental, analizar y evaluar cada una de las herramientas de análisis multidimensional open source que se han seleccionado para el estudio comparativo las denominadas PALO y MONDRIAN, las mismas que se han explicado anteriormente.

Al citar la frase análisis en detalle nos referimos al hecho de estudiar las herramientas mencionadas en el párrafo anterior, en un sentido práctico de forma tal que se pueda apreciar su funcionalidad, rendimiento, habilidades y demás prestaciones que estos dos paquetes de software open source para análisis multidimensional ofrecen.

Para el desarrollo de este capítulo partiremos de dos data warehouse demos que ofrecen cada una de las herramientas OLAP a estudiar las mismas que serán nuestras fuentes de datos ya que el análisis multidimensional no comprende el desarrollo de un data warehouse sino mas bien el tratamiento multidimensional expresado en cubos OLAP de la información de forma que se obtenga como resultado los respectivos reportes de Business Intelligence para la toma de decisiones empresariales.

Por último se ha de mencionar que con las fuentes ya definidas se aplicará el proceso de análisis multidimensional sobre esos datos, para determinar la mejor herramienta en base a la evaluación de cada una de ellas bajo una serie de parámetros de evaluación que mostraremos en el mismo.

4.2 DETERMINACIÓN DE LAS HERRAMIENTAS OPEN SOURCE PARA EL ANÁLISIS MULTIDIMENSIONAL.

Las herramientas open para el análisis multidimensional son las denominadas Palo y Mondrian, a continuación ofrecemos una descripción rápida de las mismas:

4.2.1 Palo

PALO (leído al revés OLAP) se trata de un motor multidimensional OLAP, que trabaja fundamentalmente con los datos de grandes hojas de cálculo, tipo Excel. Mediante un add-in integrado se realizan cálculos complejos a gran velocidad, permitiendo además el write-back, esto es, no solo consulta, sino escritura contra la base de datos.

La herramienta es Open Source y el add-in es gratuito, por lo que se trata de una gran utilidad para todos aquellos que manejen gran cantidad de datos en Excel.

4.2.2 Mondrian

Mondrian es un motor ROLAP con caché, lo cual lo sitúa cerca del concepto de Hybrid OLAP, ROLAP significa que en Mondrian no residen datos (salvo en la caché) sino que estos residen en un Sistema de Gestión de Bases de Datos externo permitiendo analizar grandes cantidades de datos.

Es en esta base de datos en la que residen las tablas que conforman la información multidimensional con la que Mondrian trabaja. MOLAP es el nombre que reciben los motores OLAP en los que los datos residen en una estructura dimensional.

Mondrian se encarga de recibir consultas dimensionales (lenguaje MDX) y devolver los datos de un cubo, sólo que este cubo no es algo físico sino un conjunto de metadatos que definen como se han de "mapear" estas consultas que tratan conceptos dimensionales a sentencias SQL ya tratando con conceptos relacionales que obtengan de la base de datos la información necesaria para satisfacer la consulta dimensional.

4.3 DETERMINACIÓN DE LOS PARÁMETROS DE COMPARACIÓN.

Es importante definir los parámetros de comparación entre Palo y Mondrian que son las herramientas Open Source de análisis multidimensional a estudiar.

Con el resultado final se puede elegir la mejor herramienta que más se adapte y ofrezca mayores prestaciones en el desarrollo de las aplicaciones web de explotación de cubos OLAP y reporte de la información.

Con lo anterior en mente definiremos el estudio comparativo en base a los siguientes parámetros:

PARÁMETROS	JUSTIFICACIÓN
Datos cuantitativos	De manera de contabilizar y calificar cada uno de los elementos utilizados en el software de análisis multidimensional Palo y Mondrian, para llevar a cabo una evaluación correcta.
Datos cualitativos	De manera de calificar los datos cuantitativos bajo ciertas variables como: consistencia, funcionalidad, interface completa, interfaz amigable, facilidad de uso, que sea tangible, que contenga elementos multimediales.
Escalas	Se calificará en una escala de 0...20. Donde 20 es la máxima nota.

Tabla IV.4. Determinación de los parámetros de comparación

4.3.1 Datos Cuantitativos

En esta sección de datos cuantitativos, se contabiliza los elementos que deben existir necesariamente y son utilizados en cada una de las herramientas open source de análisis multidimensional a comparar explicando si la herramienta utiliza o no ese elemento, es decir son parámetros que podemos ver.

Al mencionar los elementos utilizados nos referimos a los componentes y características que debe tener una herramienta de análisis OLAP, como se explico en la sección 2.4. Estos se listan bajo el nombre de variables en la siguiente tabla:

VARIABLES	DESCRIPCIÓN
Servidor OLAP	El cual permite la definición, creación y acceso a

	modelos de datos multidimensionales
Interfaz	La cual permite visualizar grandes volúmenes de datos y navegar sobre ellos interactivamente
Visión multidimensional de los datos	Las herramienta OLAP debe ofrecer una vista matricial de los datos
Sin restricciones sobre el número de dimensiones	No debe imponer restricciones acerca del número de dimensiones que se esté utilizando
Simetría para las dimensiones	Correspondencia exacta en la disposición regular de las partes o puntos de un cuerpo o figura con relación a un centro, un eje o un plano
Operadores intuitivos de manipulación	Toda herramienta OLAP debe ofrecer operadores para: drill-down, roll-up, slice-and-dice, pivot
Flexibilidad en la definición de dimensiones	Permitir definir de forma flexible (sin limitaciones) sobre las dimensiones: restricciones, agregaciones y jerarquías entre ellas
Transparencia en el almacén de datos	Ser transparentes al tipo de tecnología que soporta el almacén de datos (ROLAP o MOLAP)

Tabla IV.5. Variables cuantitativas

Todos y cada uno de estos elementos debe tener una herramienta de análisis multidimensional para ser llamada de esa forma.

Con esto pretendemos verificar que existan primeramente estos ocho elementos en cada una de las herramientas OLAP, para luego calificarlos bajo una escala que definiremos en la sección 3.3.3

4.3.2 Datos Cualitativos

Los parámetros cualitativos se refieren a las cualidades de las herramientas, es decir que no se las ve pero se las puede percibir.

En la siguiente tabla se definen las variables cualitativas que se utiliza en este estudio comparativo:

VARIABLES	DESCRIPCIÓN
Consistencia	Se refiere a la estabilidad, solidez, coherencia, lugar, espacio y color de la iconografía y gráficos utilizados.
Funcionalidad	El usuario tiene que saber qué es lo que se espera hacer, es decir que por ejemplo los botones funcionen correctamente y no haya sorpresas
Interface completa	Que tenga un inicio claro y un fin esperado
Interfaz amigable	Ofrecer el poder y la capacidad para invitar al uso, es decir que uno pueda recomendar
Facilidad de uso	Que ofrezca optimizar los pasos, es decir que no hayan pasos redundantes
Tangibilidad	Que se pueda percibir de forma precisa y que el usuario sienta que el contenido se refiera al tema
Elementos multimediales	Si los hay deben estar acordes al tema, es decir al análisis multidimensional

Tabla IV.6. Variables cualitativas

Estas variables cualitativas también se verifican que existan en cada una de las herramientas de análisis multidimensional a estudiar para ponerlas una nota en base a la escala que a continuación la definiremos.

4.3.3 Escalas

Tanto los parámetros cuantitativos como los cualitativos se calificarán utilizando una escala que va desde 1 hasta 20, la misma que luego se ofrecerá su respectiva descripción correspondiente que obedezca a la siguiente tabla:

RANGOS	EQUIVALENCIAS
19 - 20	Excelente
17 -18	Muy Bueno
15 - 16	Bueno
13 - 14	Regular
0 - 12	Insuficiente

Tabla IV.7. Escalas con sus equivalencias

Vale la pena indicar que se asignará notas a cada una de las variables o parámetros ya sean cuantitativos y/o cualitativos para finalmente elegir la herramienta OLAP open source que alcance las mayores calificaciones de manera que podamos con ella aplicar el análisis multidimensional y/o la explotación de cubos sobre las bases de datos transaccionales de RUB-ECUADOR.

4.4 DESCRIPCIÓN DE LAS FUENTES DE DATOS.

Las fuentes de datos para el análisis multidimensional de cada una de estas herramientas OLAP son data warehouse demos puesto que este proyecto es solo una parte de lo que se realiza en toda una solución BI, de manera que lo que se trata de decir es que a partir de tener ya listas las fuentes de datos se realiza las pruebas necesarias de análisis multidimensional sobre cada una ellas. A continuación se dará una descripción de las fuentes de datos que se utilizarán para el estudio comparativo de las herramientas OLAP open source:

4.4.1 Palo

Los data warehouse demos incorporados en estos software de uso libre, los mismos que han sido creados especialmente para este tipo de pruebas.

Para este software, el data warehouse a utilizar se denomina "Ventas", pues es una empresa que se encarga de las ventas de productos computacionales, el mismo que tiene y utiliza un esquema estrella con su tabla de hechos central y compuesta además de varias dimensiones.

Es necesario aclarar que este software admite fuentes data warehouse solo implementados en Excel, además permite hacer conexiones ODBC con otros DBMS pero únicamente a nivel de consultas siendo de momento la primera limitante encontrada, puesto que las bases de datos transaccionales de los programas sociales que al final necesitarán una solución de análisis multidimensional, están implementadas en SQL Server.

4.4.1.1 Modelamiento dimensional

El Modelamiento dimensional para este warehouse, esta implementado como un esquema en estrella pura con una tabla central de hechos o fact table que contiene las medidas necesarias, y además varias tablas llamadas dimensiones con su respectiva información descriptiva, a continuación detallamos sobre esto:

Tabla de Hechos

La tabla de hechos es la denominada Ventas dicha tabla de hechos tiene información de las ventas registradas que son sus respectivas medidas, a continuación las listamos:

1. El número de unidades vendidas
2. El beneficio bruto

Debemos indicar que esta tabla es la que tiene los indicadores numéricos que son las medidas sobre las cuales se deberá realizar el análisis, las mismas que se van agrupando según la dimensión deseada, esto depende del diseño del data warehouse.

Dimensiones

Hay varias dimensiones en el demo del data warehouse, la información descriptiva de cada valor que tome la medida a analizar, por ejemplo se listan las dimensiones que actualmente existen en el mismo:

1. Regiones
2. Productos
3. Años
4. Meses
5. Tipos de Datos

Siendo en este punto importante mencionar que es bueno analizar con dimensiones Tiempo, pues estas ofrecen un mayor nivel de análisis, en este caso por años y por meses. Esto depende exclusivamente del diseño del data warehouse.

Diseño Dimensional

A continuación se muestra un gráfico del diseño dimensional del data warehouse que utilizaremos en el análisis de PALO:

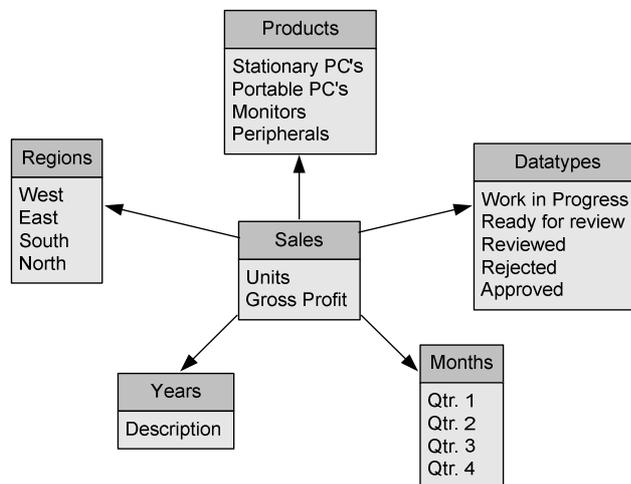


Figura IV.26. Diseño dimensional del data warehouse Ventas para el análisis de PALO

En la figura III.21, el warehouse que se ha utilizado en el análisis de Palo utiliza un esquema en estrella pura.

4.4.2 Mondrian

Para el análisis comparativo en la herramienta OLAP Mondrian, se utilizó el data warehouse demo que ofrece esta herramienta, con el fin de dedicarnos a lo que esta investigación se dirige esto es solo al análisis multidimensional sobre un data warehouse ya listo para utilizarse.

El warehouse que ofrece el servidor OLAP Mondrian para el desarrollo de pruebas es el denominado "foodmart" que además es muy parecido al que se utiliza en Palo puesto que se trata de una empresa encargada de ventas pero en este caso como su nombre lo indica de alimentos, además está implementado en Microsoft Access pero tiene un script para ejecutarlo sobre otro DBMS como SQL Server a diferencia del warehouse anterior que está implementado únicamente en Excel.

Por otro lado es importante aclarar que este servidor OLAP puede tener su warehouse en un DBMS cuales quiera ya sea SQL Server, PostgreSQL, MySQL, Access, entre otros puesto que utiliza los drivers de conexión como ODBC y JDBC al contrario de PALO que solo utiliza ODBC pero a nivel de solo consultas. Para el análisis en el warehouse implementado en SQL Server se utiliza un JDBC.

4.4.2.1 Modelamiento dimensional

Para este warehouse por el mismo hecho de utilizar un DBMS el esquema es un poco más complejo, no se utiliza un esquema en estrella sino más bien el esquema denominado copo de nieve como se vio en el capítulo del marco teórico.

El warehouse está diseñado con dos tablas de hechos con sus respectivas medidas y varias dimensiones, como son:

Tabla de Hechos

Hay dos tablas de hechos las cuales listamos a continuación:

1. Sales_fact_1997
2. Inventory_fact_1998

Estas tablas de hechos son las necesarias para almacenar las medidas del warehouse que son las unidades vendidas, las cantidades de las ventas, las cantidades de los costos

Dimensiones

Existen una variedad de dimensiones en el warehouse con esquema copo de nieve, las cuales se listan a continuación:

1. Region

2. Store
3. Employee
4. Customer
5. Salary
6. Promotion
7. Time_by_day

Siendo las principales dimensiones que utiliza el warehouse que ofrece el servidor OLAP Mondrian para realizar las pruebas necesarias.

Diseño Dimensional

A continuación se muestra el gráfico del diseño dimensional del data warehouse que se utilizará en el análisis de Mondrian:

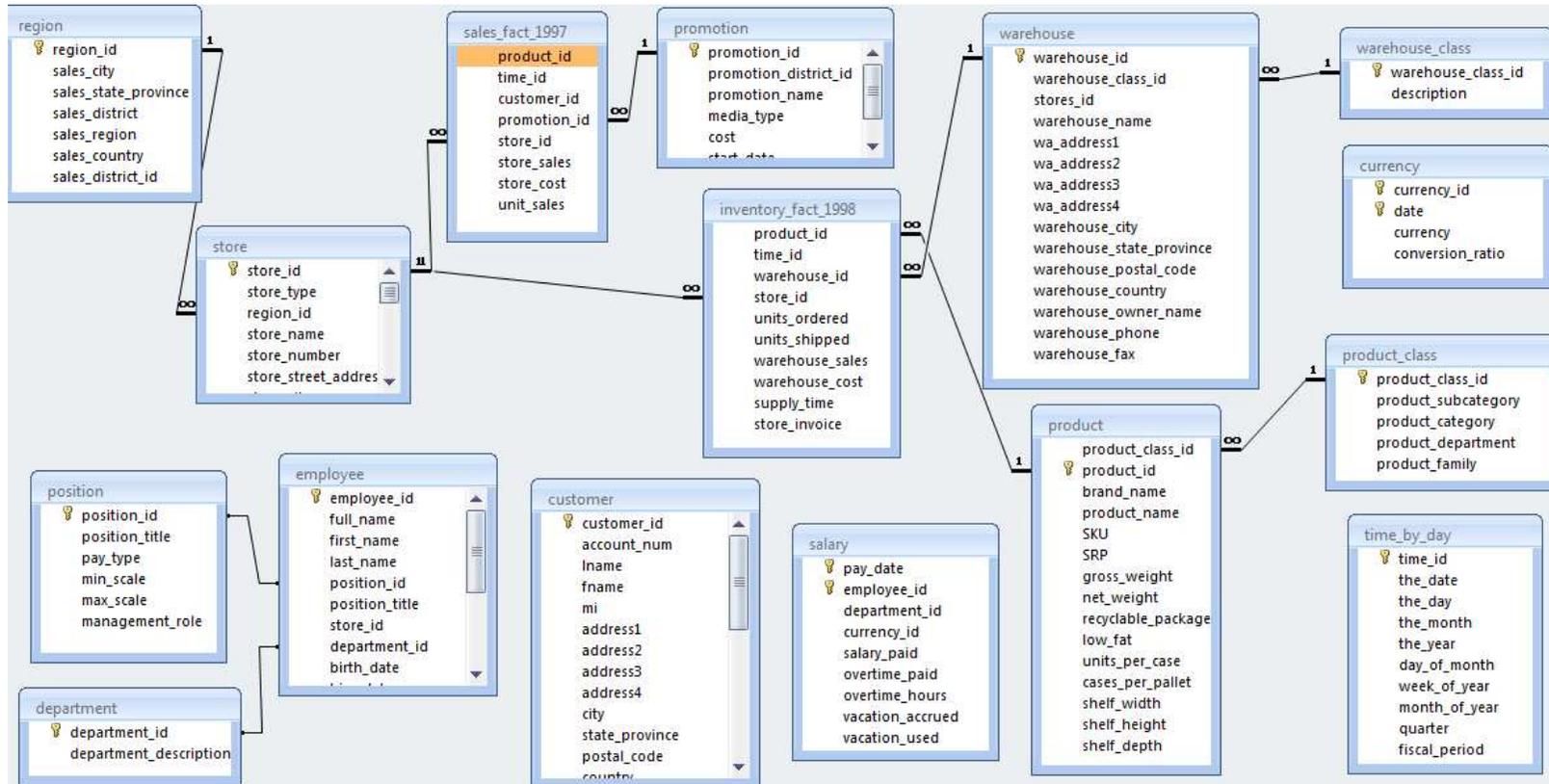


Figura IV.27. Diseño dimensional del data warehouse FoodMart para el análisis de Mondrian

4.5 DESARROLLO DE PRUEBAS

4.5.1 Palo

Desplegar web-palo.war en la interfaz de administración de Apache Tomcat, para poder crear un Cubo de ejemplo y este se publicará automáticamente para poder analizar el servidor OLAP Palo, para ello es necesario lo siguiente:

- Crear el cubo de análisis multidimensional en la herramienta de administración de PALO, ya sea el Palo Excel Add-in o a su vez el Palo Client que básicamente hacen lo mismo, con la diferencia que la primera es más intuitiva para los usuarios.
- Mostrar los resultados en el browser.

4.5.1.1 Crear el esquema para el modelo OLAP o cubo

Es importante aclarar que se utilizó el Palo Excel Add-in en la administración multidimensional puesto que ofrece la misma funcionalidad que el Palo Client pero es más intuitivo y fácil de utilizar y además prácticamente hace lo mismo que la otra herramienta.

Una vez instalado en el computador esta herramienta (Ver sección 2.5.1.3 de la instalación), se debe proceder a realizar las siguientes acciones:

1. Abrir Palo Excel Add-in desde el menú Inicio o a su vez desde el escritorio, el mismo que ofrece un menú denominado Palo en la misma interfaz de Excel.
2. Seleccione el menú Palo, que es la herramienta de análisis multidimensional con sus respectivas opciones como ya explicamos en la sección 2.5.1
3. Escoja la opción Modeller, en la cual podemos administrar la conexión al servidor, las bases de datos, además realizar el esquema mismo del análisis multidimensional.

Seleccione la base de datos Demo que Palo ofrece para realizar pruebas. A continuación defina y cree las tablas que servirán como dimensiones y hechos.

Palo ofrece un data warehouse con sus dimensiones y tabla de hechos, por lo que solo interesa ahora definir nuestro cubo para realizar el análisis multidimensional.

Modeller posee una interfaz para la creación de cubos OLAP de manera fácil e intuitiva la misma que se puede observar en la Figura 28.

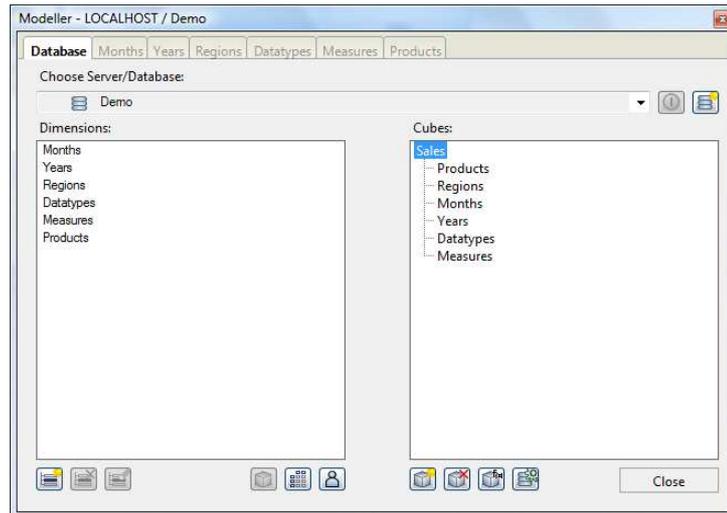


Figura IV.28: Herramienta de administración de bases de datos y cubos en Palo

Como se puede observar en la figura anterior de ha creado el cubo de nombre Sales, el mismo que se encuentra marcado, y de la misma forma podemos fijarnos que a la izquierda se encuentran todas las dimensiones del data warehouse, las mismas que únicamente es necesario escogerlas mientras se crea el cubo como se puede ver a continuación:

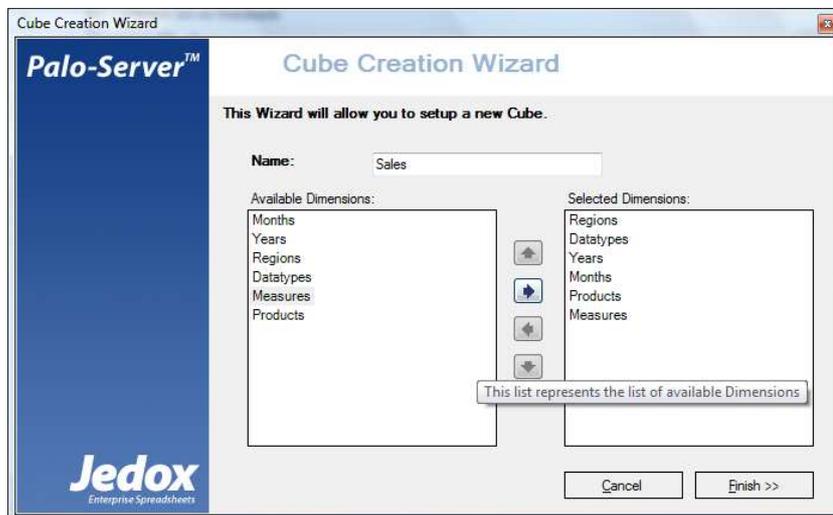


Figura IV.29: Creación del cubo Sales en Palo

Con esto se puede ver los resultados en el browser, y si es posible el usuario desarrollador puede utilizar y manipular la forma de presentación o vistas de sus datos o crear paso a paso sus cubos con las jerarquías y niveles que desee.

4.5.2 Mondrian

Despliegue el archivo Mondrian.war en la interfaz de administración de Apache Tomcat, para luego poder crear un Cubo de ejemplo y publicarlo para poder analizar el servidor OLAP Mondrian.

Para realizar el análisis multidimensional se requiere lo siguiente:

- Modificar el código fuente de Mondrian.jsp, de tal forma que se ingrese el código necesario para establecer la conexión a la base de datos en este caso al warehouse.
- Crear un esquema para el modelo OLAP, puede ser en eclipse o a su vez en la herramienta que ofrece Pentaho la denominada Schema Workbench que es la que se utiliza por su facilidad para el diseño del esquema del modelo OLAP o cubo o a su vez se puede simplemente realizar en cualquier editor de documentos XML,

Además se puede realizar el esquema XML del cubo OLAP en un bloc de notas para los usuarios más avanzados.

4.5.2.1 Modificar Mondrian.jsp

Copie el driver de conexión a la base de datos en la carpeta de las librerías del proyecto de Mondrian, y utilizar el sqljdbc1.2 que podemos descargar de la red ya que es totalmente gratuito.

Modificar el archivo Mondrian.jsp en las líneas referentes a la cadena de conexión a la base de datos, en este caso al warehouse FoodMart ubicado en SQL Server, como a continuación se muestra en el gráfico:

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<?xml-stylesheet type="text/xsl" href="http://www.tonbeller.com/jpivot" ?>
<?xml-stylesheet type="text/xsl" href="http://java.sun.com/jstl/core" ?>

<!-- uses a dataSource -->
<!-- jp:mondrianQuery id="query01" dataSource="jdbc/MondrianFoodmart"
catalogUri="/WEB-INF/demo/FoodMart.xml" -->

<!-- uses mysql -->
<!-- jp:mondrianQuery id="query01" jdbcDriver="com.mysql.jdbc.Driver"
jdbcUrl="jdbc:mysql://localhost/foodmart" catalogUri="/WEB-
INF/queries/FoodMart.xml" -->

<!-- uses a role defined in FoodMart.xml -->
<!-- jp:mondrianQuery role="California manager" id="query01"
jdbcDriver="sun.jdbc.odbc.JdbcOdbcDriver"
jdbcUrl="jdbc:odbc:MondrianFoodMart" catalogUri="/WEB-
INF/queries/FoodMart.xml" -->

<jp:mondrianQuery id="query01"
jdbcDriver="com.microsoft.sqlserver.jdbc.SQLServerDriver"
jdbcUrl="jdbc:sqlserver://localhost:1433;databaseName=FoodMart;user=User
RIPS_MCDS;password=UserRips_2009_90;" catalogUri="/WEB-
INF/queries/FoodMart.xml">

select
  {[Measures].[Unit Sales], [Measures].[Store Cost], [Measures].[Store
Sales]} on columns,
  {[([Promotion Media].[All Media], [Product].[All Products])} ON rows
from Sales
where ([Time].[1997])
</jp:mondrianQuery>

<c:set var="title01" scope="session">Test Query uses Mondrian OLAP
</c:set>
```

Figura IV.30. Líneas de conexión al warehouse FoodMart

4.5.2.2 Crear el esquema para el modelo OLAP o cubo

Crear el esquema del modelo OLAP es decir el cubo mismo, para ello seguimos los siguientes pasos:

1. Crear el nuevo esquema de cubo que lo llamaremos FoodMart, teniendo en cuenta que es un documento XML así que si deseamos lo podemos hacer en un bloc de notas, de la siguiente forma:

```
<? Xml version="1.0"?>
<Schema name="FoodMart">
</Schema>
/>
```

2. Crear el cubo con sus dimensiones y medidas, además es necesario indicar que no solo podemos crear un cubo sino varios cubos de acuerdo a nuestras necesidades.

Para este ejemplo crearemos el cubo Sales es decir de las ventas realizadas, para ello primero especificamos el nombre del cubo con su respectiva tabla origen de la siguiente manera:

```
<!-- Sales -->
- <Cube name="Sales" defaultMeasure="Unit Sales">
- <Table name="sales_fact_1997">
```

Figura IV.31. Creación de un cubo OLAP "Sales"

3. Crear medidas del cubo Sales, por ejemplo:

```
<Measure name="Unit Sales" column="unit_sales" aggregator="sum" formatString="Standard" />
<Measure name="Store Cost" column="store_cost" aggregator="sum" formatString="#,###.00" />
<Measure name="Store Sales" column="store_sales" aggregator="sum" formatString="#,###.00" />
<Measure name="Sales Count" column="product_id" aggregator="count" formatString="#,###" />
<Measure name="Customer Count" column="customer_id" aggregator="distinct-count" formatString="#,###" />
```

Figura IV.32. Creación de varias medidas para el cubo "Sales"

4. Posteriormente crear cada una de las dimensiones, en un formato como el que presentamos a continuación:

```
- <Dimension name="Promotion Media" foreignKey="promotion_id">
- <Hierarchy hasAll="true" allMemberName="All Media" primaryKey="promotion_id" defaultMember="All Media">
  <Table name="promotion" />
  <Level name="Media Type" column="media_type" uniqueMembers="true" />
</Hierarchy>
</Dimension>
- <Dimension name="Promotions" foreignKey="promotion_id">
- <Hierarchy hasAll="true" allMemberName="All Promotions" primaryKey="promotion_id" defaultMember="[All Promotions]">
  <Table name="promotion" />
  <Level name="Promotion Name" column="promotion_name" uniqueMembers="true" />
</Hierarchy>
</Dimension>
```

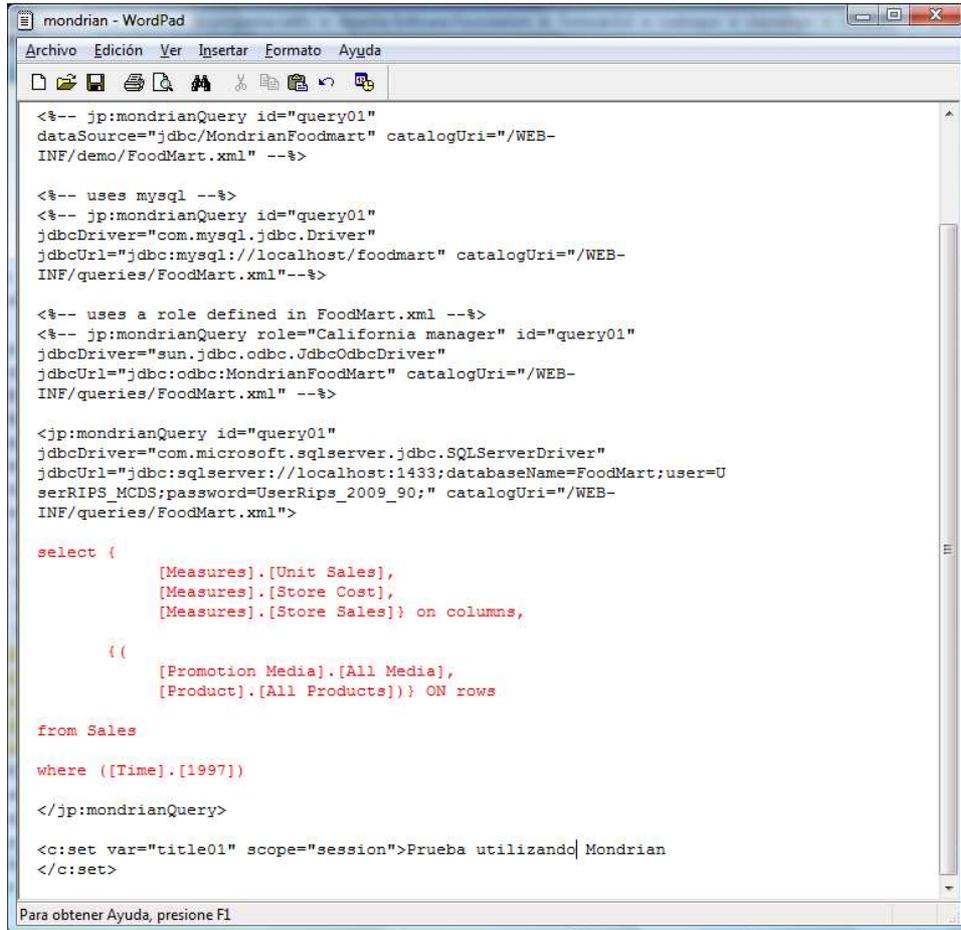
Figura IV.33. Creación de dimensiones para el cubo "Sales"

En la figura IV.28 mostramos las creaciones de dos de las dimensiones que estamos utilizando para este cubo Sales, son las dimensiones para las promociones.

El objetivo fundamental es mostrar el formato para crear las dimensiones.

4.5.2.3 Crear una consulta mdx

Para ello debemos modificar una vez más el archivo Mondrian.jsp y escribir en lenguaje mdx que es el lenguaje multidimensional de todo data warehouse, una consulta que será por defecto, como se muestra a continuación.



```
<!-- jp:mondrianQuery id="query01"
dataSource="jdbc/MondrianFoodmart" catalogUri="/WEB-
INF/demo/FoodMart.xml" -->

<!-- uses mysql -->
<!-- jp:mondrianQuery id="query01"
jdbcDriver="com.mysql.jdbc.Driver"
jdbcUrl="jdbc:mysql://localhost/foodmart" catalogUri="/WEB-
INF/queries/FoodMart.xml"-->

<!-- uses a role defined in FoodMart.xml -->
<!-- jp:mondrianQuery role="California manager" id="query01"
jdbcDriver="sun.jdbc.odbc.JdbcOdbcDriver"
jdbcUrl="jdbc:odbc:MondrianFoodMart" catalogUri="/WEB-
INF/queries/FoodMart.xml" -->

<jp:mondrianQuery id="query01"
jdbcDriver="com.microsoft.sqlserver.jdbc.SQLServerDriver"
jdbcUrl="jdbc:sqlserver://localhost:1433;databaseName=FoodMart;user=U
serRIPS_MCDS;password=UserRips_2009_90;" catalogUri="/WEB-
INF/queries/FoodMart.xml">

  select {
    [Measures].[Unit Sales],
    [Measures].[Store Cost],
    [Measures].[Store Sales]} on columns,
    {
      {
        [Promotion Media].[All Media],
        [Product].[All Products]} ON rows
    }
  from Sales
  where ([Time].[1997])

</jp:mondrianQuery>

<c:set var="title01" scope="session">Prueba utilizando Mondrian
</c:set>
```

Figura IV.34. Consulta mdx por defecto para el cubo "Sales"

4.5.2.4 Resultados presentados en pantalla

En esta sección mostraremos los resultados que presentan cada una de las interfaces de las herramientas OLAP estudiadas:

Palo

La siguiente es la interfaz web dinámica que ofrece Palo para el análisis multidimensional.

The screenshot shows the Palo Web Client interface in a Windows Internet Explorer browser. The address bar displays the URL: `http://localhost:8080/web-palo/com.tensegrity.paloweviewer.modules.application.Application/Application.html`. The page title is "Palo Web Client".

The main content area is titled "Cube 'Sales'" and shows a multidimensional report. The dimensions are configured as follows:

- Months: Year
- Years: 2002
- Datatypes: Variance
- Measures: Gross Profit

The report is organized by Regions. The visible regions are Europe, East, South, and Italy. The data is presented in a table with columns for each region and rows for various product categories. The product categories are:

- All Products
- Stationary PCs
- Desktop
- Server
- Portable PCs
- Notebook

The table shows numerical values for each product category across the different regions. For example, for "All Products", the values are: Europe (32,831.21), East (12,320.86), South (62,336.62), and Italy (39.9).

Figura IV.35. Reporte dinámico del cubo "Sales" utilizando Palo

Mondrian

La interfaz dinámica que ofrece Jpivot que es el navegador OLAP con el que trabaja el servidor de análisis multidimensional "Mondrian".

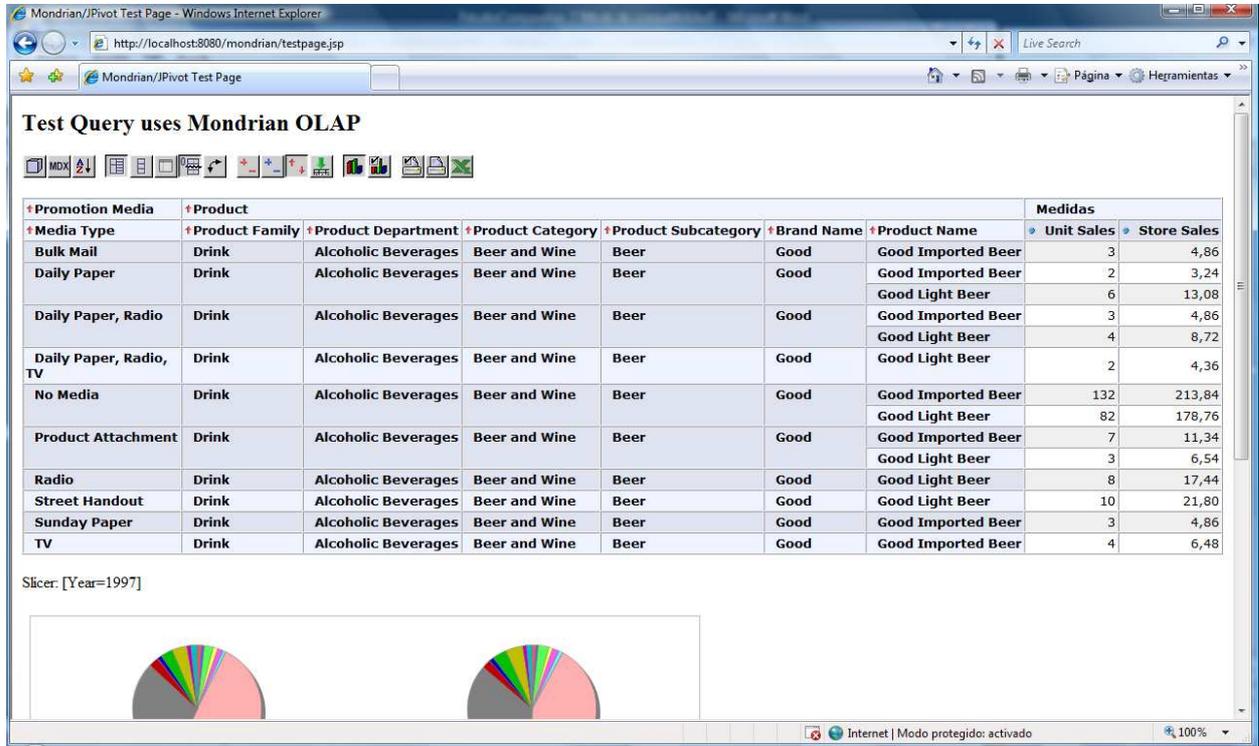


Figura IV.36. Reporte dinámico del cubo "Sales" utilizando Mondrian

4.6 ANÁLISIS CUALITATIVO Y CUANTITATIVO

4.6.1 Análisis Cuantitativo

4.6.1.1 Verificación de existencia de elementos

Es interesante que una herramienta OLAP posea elementos que garanticen el mejor rendimiento y prestaciones en su utilización.

En la tabla 8 se listan los elementos y servidores OLAP de evaluación en las cuales se cuantifican cada uno de ellos con los respectivos comodines como son:

- Un visto si el elemento existe en la herramienta y
- Una X si no existe.

Al final se contabiliza los elementos existentes por cada servidor OLAP.

VARIABLES	SOFTWARE OLAP A EVALUAR	
	PALO	MONDRIAN
Servidor OLAP	✓	✓
Interfaz	✓	✓
Visión multidimensional de los datos	✓	✓
Sin restricciones sobre el número de dimensiones	✗	✓
Simetría para las dimensiones	✓	✓
Operadores intuitivos de manipulación	✓	✓
Flexibilidad en la definición de dimensiones	✓	✓
Transparencia en el almacén de datos	✗	✓
TOTALES	6	8

Tabla IV.8. Verificación de la existencia de elementos cuantitativos

4.6.1.2 Interpretación de los resultados

En la presente sección se define, analiza e interpreta los resultados de cada uno de los elementos de evaluación de las herramientas Open Source de análisis multidimensional que se contemplaron en la tabla 8.

a. Servidor OLAP

Debido a que el servidor OLAP permite la definición, creación y acceso a modelos de datos multidimensionales, se convierte en el elemento fundamental que la herramienta de análisis multidimensional debe poseer.

Este vital e imprescindible elemento es el que permite la creación de los cubos OLAP con sus medidas, dimensiones, jerarquías y niveles.

Las herramientas de estudio poseen este elemento principal por lo que se convierten en claros paquetes software de análisis multidimensional en el sentido de que permiten crear y alojar los esquemas multidimensionales sobre los cuales se analizarán los datos que provienen de un warehouse y/o datamart.

b. Interfaz

Es el elemento que permite visualizar grandes volúmenes de datos y navegar sobre ellos interactivamente.

Para lo cual es necesario indicar según las pruebas realizadas anteriormente que las dos herramientas utilizadas en el estudio comparativo tienen una interfaz para visualizar volúmenes de datos grandes y además si permiten navegar sobre ellos de forma muy interactiva.

Para esta sección solo se enfatizará en la existencia de cada uno de estos elementos en las herramientas OLAP que estamos analizando, ya que en el análisis cualitativo se verá un estudio más detallado de cada uno de estos elementos.

c. Visión multidimensional de los datos

Esta característica se refiere a que las herramientas OLAP deben ofrecer una vista matricial de los datos de tal forma que se pueda apreciar tanto las descripciones de las dimensiones así como las medidas de manera combinada en columnas y filas en donde además el usuario puede interactuar con ellas.

d. Sin restricciones sobre el número de dimensiones

Con esto se puntualiza que las herramientas OLAP no deben imponer restricciones acerca del número de dimensiones que se esté utilizando en el análisis multidimensional por ello la palabra.

En la tabla 8, se puede observar que los dos software Palo y Mondrian tienen una X y un visto respectivamente en esta característica y esto se debe a que:

1. En el caso de Palo, tiene una X porque este software sí presenta restricciones en cuanto al número de dimensiones.

Esta herramienta OLAP permite realizar análisis multidimensional sobre los datos pero únicamente hasta 256 dimensiones.

2. En el caso de Mondrian, tiene un visto porque de verdad que es un software de análisis multidimensional puesto que se maneja sin restricciones en cuanto al número de dimensiones, es decir los desarrolladores business Intelligence tienen la oportunidad de configurar el modelo pertinente de análisis OLAP que permitirá navegar interactivamente sobre un esquema multidimensional.

e. Simetría para las dimensiones

Con ello se puntualiza que una herramienta OLAP debe ser capaz de permitir a los desarrolladores la creación de esquemas multidimensionales en donde sus dimensiones se correspondan exactamente con el punto central que es la tabla de hechos con sus medidas de forma que se ofrezca información exacta y verídica en los informes.

En la tabla 8, se puede observar que tanto Palo como Mondrian ofrecen esta característica puesto que en las pruebas realizadas se manejó estrellas puras como esquemas multidimensionales para cada herramienta OLAP con sus dimensiones y tabla de hechos central.

f. Operadores intuitivos de manipulación

Este elemento básicamente debe tener toda herramienta de análisis multidimensional, esto es que deben ofrecer operadores para: drill-down, roll-up, slice-and-dice, pivot que se explicó en la sección 2.4.

Estos operadores permiten el análisis en sí ya que estas son las que permiten navegar interactivamente en el cubo OLAP de tal forma que les ofrecerá a los usuarios finales una vista más clara y estadística de su información para la adecuada toma de decisiones empresariales.

Como lo hemos visto en la tabla 8, las dos herramientas OLAP Palo y Mondrian ofrecen estos operadores cada uno de una manera un poco diferente pero que en esencia hacen lo mismo y esto puede verse en la sección 3.5 del desarrollo de pruebas sobre cada uno de ellos.

g. Flexibilidad en la definición de dimensiones

Permitir definir de forma flexible (sin limitaciones) sobre las dimensiones: restricciones, agregaciones y jerarquías entre ellas lo que se explico igualmente en la sección 2 referente al marco teórico, y lo podemos ver claramente que tanto Palo como Mondrian tienen esta característica según las pruebas realizadas en la sección 3.5.

h. Transparencia en el almacén de datos

Ser transparentes al tipo de tecnología que soporta el almacén de datos (ROLAP o MOLAP).

En el capítulo II del marco teórico hemos dejado claro la parte teórica relacionada con el análisis multidimensional independientemente de la herramienta.

Luego en el mismo capítulo en la sección 2.5 se explico que Palo es una herramienta MOLAP y Mondrian es ROLAP tendiente a ser HOLAP.

Pero hay que notar algo sumamente importante y es que mientras PALO almacena los datos fuentes y multidimensionales en sus propias estructuras físicas de almacenamiento de datos tal como lo hace un DBMS y además permite conectarse con otros DBMS por medio de ODBC pero únicamente a nivel de llamadas a consultas (queries) pero de manera diferente a Mondrian.

Mondrian por su parte al ser ROLAP, permite el almacenamiento de los datos fuentes en RDBMS cualesquiera ya sea SQL Server, PostgreSQL, MySQL entre otros por medio de conexiones JDBC, y para realizar su análisis multidimensional filtra la información que necesita y lo almacena en memoria caché en esquemas multidimensionales con formato XML lo cual le acerca al concepto de HOLAP.

Con esto lo que pretendemos decir es que con Palo no es posible tomar los datos fuentes de otros DBMS de la misma forma que lo hace Mondrian. Puesto que con Palo si queremos hacer análisis multidimensional de una base de datos que se encuentre en un DBMS cualesquiera

que sea, se debe primero rellenar manualmente los registros de las fuentes en la base de datos multidimensional de Palo.

Mientras que con Mondrian lo único que se debe hacer es definir un esquema multidimensional XML en donde al invocar ya desde el reporte OLAP esto es en el resultado final al usuario, se rellena en forma automática el esquema multidimensional de datos en memoria caché y no necesitamos pasarlos manualmente como sucede en PALO. Razón por la cual concluimos para este apartado que Palo no es transparente en el almacén de datos mientras que Mondrian si lo es.

4.6.2 Análisis Cualitativo

En esta sección se dará una calificación a las variables cuantitativas en forma detallada que se ha podido verificar en Palo y Mondrian en base a ciertos parámetros definidos como datos o variables cualitativas en el apartado 3.3.2.

Este análisis cualitativo tiene como objetivo fundamental evaluar las herramientas OLAP Open Source que están siendo objeto del presente estudio, asignándoles notas en cada uno de los elementos y características que debe tener una herramienta de análisis multidimensional, de tal forma que se pueda obtener como resultado la herramienta que obtenga las mejores notas, situándole en la cúspide para de esta forma dedicarnos en el próximo capítulo al desarrollo de las aplicaciones web de explotación de los cubos a realizar con la herramienta resultante de este estudio comparativo.

4.6.2.1 Calificaciones alcanzadas

Las calificaciones que vemos a continuación están dadas en la escala definida con anterioridad en la sección 3.3.3 específicamente en la tabla 7 que describe los rangos de la escala a utilizar en el presente estudio comparativo de herramientas Open Source para análisis multidimensional.

Además hemos de mencionar que se ha puesto las notas según nuestra experiencia en la utilización de las herramientas descritas en la sección 3.5 que habla acerca del desarrollo de pruebas en cada software OLAP los denominados Palo y Mondrian en el mismo orden.

a. Servidor OLAP

En la siguiente tabla se muestra las calificaciones alcanzadas por cada una de las herramientas OLAP que estamos analizando en cuanto al servidor OLAP que se utiliza.

VARIABLES	SOFTWARE OLAP A EVALUAR	
	PALO	MONDRIAN
Consistencia	15	17
Funcionalidad	17	18
Interface completa	18	20
Interfaz amigable	18	16
Facilidad de uso	18	17
Tangibilidad	18	18
Elementos multimediales	18	17
PROMEDIO	17	18

Tabla IV.9. Evaluación acerca del servidor OLAP

b. Interfaz

En la siguiente tabla se muestra las calificaciones alcanzadas por cada una de las herramientas OLAP que estamos analizando en cuanto a la interfaz web que se utiliza en la presentación de los resultados.

VARIABLES	SOFTWARE OLAP A EVALUAR	
	PALO	MONDRIAN
Consistencia	18	19
Funcionalidad	19	19
Interface completa	18	20
Interfaz amigable	18	18
Facilidad de uso	18	19
Tangibilidad	19	19
Elementos multimediales	17	20
PROMEDIO	18	19

Tabla IV.10. Evaluación acerca de la Interfaz web que muestra los reportes

c. Visión multidimensional de los datos

En la siguiente tabla se muestra las calificaciones alcanzadas por cada una de las herramientas OLAP que estamos analizando en cuanto a la visión multidimensional de los datos.

VARIABLES	SOFTWARE OLAP A EVALUAR	
	PALO	MONDRIAN
Consistencia	17	19
Funcionalidad	18	18
Interface completa	17	19
Interfaz amigable	19	19
Facilidad de uso	18	18
Tangibilidad	19	19
Elementos multimediales	17	19
PROMEDIO	18	19

Tabla IV.11. Evaluación acerca de la visión multidimensional de los datos

d. Administración de las dimensiones

A continuación se muestra una tabla en donde podremos observar la calificación para las herramientas OLAP en cuanto al hecho de que una herramienta de análisis multidimensional no debe presentar restricciones sobre el número de dimensiones y además la forma general de administración de las mismas en la interfaz web.

Es necesario mencionar también como Palo por ser una herramienta MOLAP tiene obligatoriamente restricciones en cuanto al número de dimensiones que se puede utilizar, entonces se observa cómo nos afectó este hecho en el desarrollo de pruebas.

VARIABLES	SOFTWARE OLAP A EVALUAR	
	PALO	MONDRIAN
Consistencia	16	19
Funcionalidad	16	19
Interface completa	16	19
Interfaz amigable	18	19
Facilidad de uso	19	19
Tangibilidad	19	19
Elementos multimediales	16	19
PROMEDIO	17	19

Tabla IV.12. Evaluación acerca de la administración de dimensiones

e. Operadores intuitivos de manipulación

En la siguiente tabla se muestra las calificaciones alcanzadas por cada una de las herramientas OLAP que estamos analizando en cuanto a la utilización de los operadores intuitivos de manipulación, es decir los que permiten realizar el drill-down, roll-up, slice-and-dice, pivot. Esto es en la interfaz web que muestra los reportes Business Intelligence.

VARIABLES	SOFTWARE OLAP A EVALUAR	
	PALO	MONDRIAN
Consistencia	17	19
Funcionalidad	17	19
Interface completa	17	19
Interfaz amigable	17	18
Facilidad de uso	18	19
Tangibilidad	18	18
Elementos multimediales	17	19
PROMEDIO	17	19

Tabla IV.13. Evaluación acerca de los operadores de manipulación

f. Transparencia en el almacén de datos

En este inciso se describe una nota de acuerdo a la experiencia adquirida en la utilización de cada una de las herramientas: Palo y Mondrian en cuanto a la transparencia en el almacén de datos esto es como afecta en los resultados finales.

Hemos visto ya que Palo por su mismo concepto de ser una herramienta MOLAP, no utiliza un DBMS externo para la capa de almacenamiento de datos como ocurre con Mondrian, razón por la cual si afecta en los resultados de los reportes, y más aun cuando los usuarios deseen realizar análisis multidimensional de una gran cantidad de datos. Esto es parecido a cuando Palo tenía restricciones en cuanto al número de dimensiones.

VARIABLES	SOFTWARE OLAP A EVALUAR	
	PALO	MONDRIAN
Consistencia	16	18
Funcionalidad	17	18
Interface completa	17	18
Interfaz amigable	18	19
Facilidad de uso	18	18
Tangibilidad	18	18
Elementos multimediales	18	19
PROMEDIO	17	18

Tabla IV.14. Evaluación acerca de la transparencia en el almacén de datos

4.6.2.2 Interpretación de los resultados

a. Servidor OLAP

En promedio en cuanto al servidor OLAP como se observa en la tabla 9 se analiza los resultados a continuación:

Palo: obtuvo una nota de 17, lo que le ubica en un software muy bueno en cuanto a la utilización que hace del servidor OLAP. Este hecho se debe a su naturaleza misma, ya que este software de análisis multidimensional es MOLAP por lo que ofrece un buen rendimiento en cuanto a ciertos parámetros como la administración de los esquemas multidimensionales o cubos OLAP pero básicamente la fuente de datos no debe ser muy grande para su correcto rendimiento por lo que es una gran limitante.

Mondrian: obtuvo una nota de 18, lo que le ubica en un software muy bueno en cuanto a la utilización que hace del servidor OLAP. Esto se debe mayormente a la dificultad que el software requiere para administrar los esquemas OLAP puesto que se debe tener muy en claro los conceptos de business Intelligence ya que Mondrian tiene algunas herramientas para la administración del servidor OLAP, pero la mejor es la denominada Schema Workbench que la verdad no es muy intuitiva como la que ofrece Palo.

b. Interfaz

En promedio en cuanto a la Interfaz como se observa en la tabla 10 se analiza los resultados a continuación:

Palo: obtuvo una nota de 18, lo que le ubica en un software muy bueno en cuanto a la utilización que hace de la interfaz web que muestra los resultados. Este hecho se debe a que el software no ofrece en su interfaz de resultados una interfaz completa, además sus elementos multimediales son limitados por lo que no es muy intuitivo en cuanto a su manejo.

La interfaz web que muestra los resultados de los reportes business Intelligence que ofrece Palo si es interactiva con los usuarios pero no al nivel de Mondrian. Pero por otro lado por ser MOLAP, Palo permite ingresar en tiempo real valores para las medidas lo que lo hace superior a Mondrian en este punto, esto es que se puede trabajar directamente con la base de datos teniendo la capacidad de ingresar datos en los registros de las medidas.

Mondrian: obtuvo una nota de 19, lo que le ubica en un software sobresaliente en cuanto a la utilización que hace de la interfaz web que muestra los resultados. Este hecho se debe a que el software es muy completo ofreciéndole al usuario prácticamente todas las características que debe ofrecer una herramienta OLAP en cuanto a sus resultados, es decir presenta un alto nivel

de interacción con el usuario debido a que le da posibilidades para personalizar sus reportes, ofreciéndole una amplia gama de elementos multimediales con alto funcionalidad, además permite realizar gráficos estadísticos en tiempo real.

Además Mondrian en sus múltiples botones de control que ofrece en su interfaz web de los reportes permite exportar los resultados en otros formatos esto es a EXCEL y PDF lo que Palo no ofrece para la web.

c. Visión multidimensional de los datos

En promedio en cuanto a la visión multidimensional de los datos como se observa en la tabla 11 se analiza los resultados a continuación:

Palo: obtuvo una nota de 18, lo que le ubica en un software muy bueno en cuanto a la visión multidimensional de los datos. Esto se debe a que esta herramienta si muestra los datos en forma matricial pero con el inconveniente que de manera predeterminada muestra solo una medida a la vez.

Por otro lado es necesario mencionar que hay ocasiones en las que no va a existir una medida es decir va a ser vacía, para una dimensión o un conjunto de dimensiones dadas, y se debe mencionar que Palo no ofrece el poder ocultar o habilitar los registros de las dimensiones con medidas vacías.

Además es muy reducido en cuanto a los elementos multimediales de control para que el usuario personalice sus reportes dinámicos.

Mondrian: obtuvo una nota de 19, lo que le ubica en un software sobresaliente en cuanto a la visión multidimensional de los datos. Esto se debe a que como la herramienta anterior también Mondrian muestra los resultados de sus reportes dinámicos en forma matricial y aunque al igual que Palo primero se debe configurar la consulta mdx para ver varias medidas, pero este software es superior en cuanto a que tiene una variedad de elementos multimediales de control para ver por ejemplo los datos origen de una medida dada haciendo más grande el nivel de detalle.

Lo mejor de Mondrian en cuanto a la visión multidimensional de los datos es que si permite habilitar y ocultar los registros de las dimensiones cuyas medidas estén vacías. Esto es algo muy importante que no debe faltar en una herramienta de análisis OLAP.

d. Administración de las dimensiones

En promedio en cuanto a la administración de las dimensiones como se observa en la tabla 12 se analiza los resultados a continuación:

Palo: obtuvo una nota de 17, lo que le ubica en un software muy bueno en cuanto a la administración de las dimensiones. Esto se debe a que esta herramienta no ofrece una interfaz completa para manipular interactivamente las dimensiones, sin mostrar intuitivamente cuando existen varios niveles y jerarquías en una dimensión dada.

Además no muestra muchos elementos de control para manipular las dimensiones como lo hace Mondrian.

Mondrian: obtuvo una nota de 19, lo que le ubica en un software sobresaliente en cuanto a la administración de las dimensiones. Esto se debe a que esta herramienta es muy interactiva con el usuario, la misma que le presenta varias opciones que puede hacer con sus dimensiones y esto es desde ocultar y mostrarlas cuando lo desee como lo hace también Palo, pero además permite realizar filtros intuitivamente sobre las dimensiones y las medidas.

Por otro lado este software permite también ordenar alfabéticamente ya sea en forma ascendente como descendentemente y no solo por dimensiones sino también por medidas.

e. Operadores intuitivos de manipulación

En promedio en cuanto a los operadores intuitivos de manipulación como se observa en la tabla 13 se analiza los resultados a continuación:

Palo: obtuvo una nota de 17, lo que le ubica en un software muy bueno en cuanto a la utilización de los operadores intuitivos de manipulación. Esto se debe a que esta herramienta no ofrece muchos elementos multimediales de control que son los que necesariamente hacen estas operaciones aunque i las hace pero no de forma intuitiva.

Mondrian: obtuvo una nota de 19, lo que le ubica en un software sobresaliente en cuanto a la utilización de los operadores intuitivos de manipulación. Esto se debe a que esta herramienta aunque es superior a Palo en cuanto a que ofrece los botones de control necesarios para realizar las operaciones de business Intelligence, pero no son muy amigables y a veces el usuario tiende a perderse en este punto.

Además Mondrian permite realizar dos operaciones de manipulación en el análisis que son muy importantes en OLAP y que además Palo no lo hace, estas son: pivot (permite la rotación del cubo en ejes) y además el drill through (permite los datos origen)

f. Transparencia en el almacén de datos

En promedio en cuanto a la transparencia en el almacén de datos como se observa en la tabla 14 se analiza los resultados a continuación:

Palo: obtuvo una nota de 17, lo que le ubica en un software muy bueno en cuanto a la transparencia en el almacén de datos. Esto se debe a que esta herramienta no es transparente en la capa de almacenaje puesto que no permite traer datos de un DBMS diferente ya que Palo no es ROLAP como Mondrian sino MOLAP.

Mondrian: obtuvo una nota de 18, lo que le ubica en un software muy bueno en cuanto a la administración de las dimensiones. Esto se debe a que esta herramienta aunque es un ROLAP si trabaja ya con más de un millón de datos presenta problemas de memoria caché y hay que tener un servidor con buenas características de hardware para que Mondrian de su mejor rendimiento.

4.7 RESULTADO TOTAL DEL ANÁLISIS COMPARATIVO

4.7.1 Promedios parciales alcanzados

A continuación se muestra en la siguiente tabla el promedio parcial de todos y cada uno de los parámetros que son elementos primordiales y características que se deben necesariamente tomar en cuenta en el estudio exhaustivo de una herramienta de análisis multidimensional.

VARIABLES	PROMEDIOS SOFTWARE OLAP	
	PALO	MONDRIAN
Servidor OLAP	17	18
Interfaz	18	19
Visión multidimensional de los datos	18	19
Administración de las dimensiones	17	19
Operadores intuitivos de manipulación	17	19
Transparencia en el almacén de datos	17	18

Tabla IV.15. Promedios parciales alcanzados

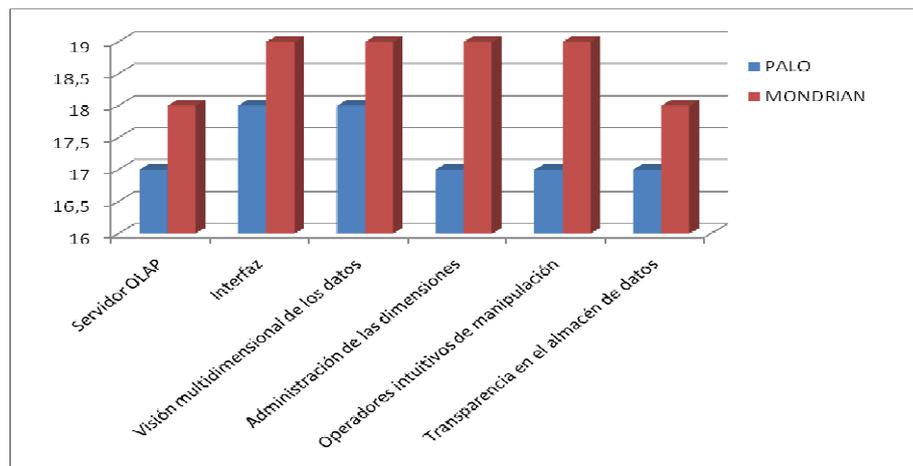


Figura V.37. Promedios Parciales alcanzados

4.7.2 Comprobación de la Hipótesis

4.7.2.1 Hipótesis

“El Estudio comparativo para el análisis multidimensional con herramientas Open Source permitirá determinar la herramienta que brinda las mejores prestaciones para el desarrollo del análisis multidimensional del data warehouse RUB de tal forma que facilite el proceso de toma de decisiones”.

4.7.2.2 Comprobación

Para la comprobación de la hipótesis anteriormente citada y basado en la evaluación comparativa de cada uno de estas herramientas software de análisis OLAP, se ofrece a continuación el promedio total el mismo que esta dado por las notas parciales anteriormente mencionadas, de tal forma que se obtenga como resultado la mejor herramienta de análisis multidimensional, de la siguiente manera:

VARIABLES	SOFTWARE OLAP	
	PALO	MONDRIAN
Promedio	17	19
Equivalencia	Muy Bueno	Sobresaliente

Tabla IV.16. Promedio total

La tabla 16 y la figura IV.33 muestran claramente la supremacía de Mondrian como herramienta de análisis multidimensional sobre Palo, en cuanto a la plataforma Open Source se refiere.

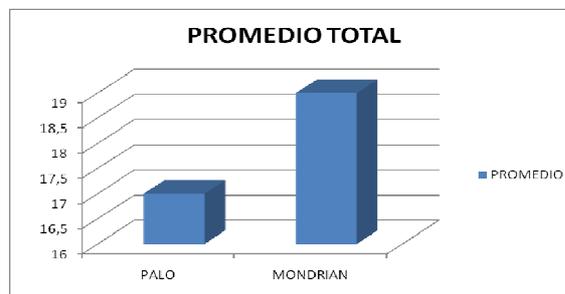


Figura IV.38. Promedio total

4.7.3 Conclusión del Análisis Comparativo

El resultado del estudio comparativo sobre las herramientas de análisis multidimensional Open Source: Palo y Mondrian, ha dado como resultado que Mondrian ofrece mejor rendimiento así como mayores prestaciones que Palo, para la plataforma de Business Intelligence de código abierto.

De manera que con este software se dará la solución BI para el proyecto RUB – ECUADOR, cumpliendo con la hipótesis planteada, lo cual está documentado en el capítulo 5.

Por otro lado se ofrece la siguiente lista de conclusiones a las cuales se ha podido llegar gracias al estudio comparativo, acerca de Mondrian que se estableció como la mejor herramienta de análisis multidimensional Open Source:

- 1) El modelo de datos multidimensional tiene que ser diseñado y construido por el desarrollador. Se tiene que definir el modelo o esquema en XML, realizar la carga de datos y se deben tener conocimientos de cómo construir un modelo de datos multidimensional.
- 2) Tiene problemas de desempeño.
- 3) La instalación de Mondrian es una tarea que no está automatizada y por lo tanto es compleja puesto que se tiene que adecuar y recompilar el código.
- 4) Si se requiere una interfaz de usuario, esta tendrá que ser construida desde cero. Un proyecto que se ha venido desarrollando junto con Mondrian para ofrecer una capa de visualización es el proyecto Jpivot el mismo que lo hemos utilizado.

Con esto lo que se trata de decir es que si bien es cierto Mondrian como herramienta de análisis multidimensional Open Source está muy por encima de la herramienta OLAP Palo como se ha demostrado en este capítulo, pero también tiene sus desventajas y esto se debe al mismo hecho de ser Open Source, es decir vemos que todavía falta de complementarla para competir con las herramientas de código propietario.

Por último es necesario indicar que con la herramienta resultante de este análisis comparativo en este caso Mondrian, se procede a desarrollar e implementar las aplicaciones web de análisis multidimensional para los programas sociales de Rub - Ecuador.

CAPÍTULO V

5. DISEÑO E IMPLEMENTACIÓN

5.1 INTRODUCCIÓN

En el presente capítulo se describe en detalle el diseño e implementación de cada uno de los *productos software* que darán como resultado las aplicaciones Web de explotación de los cubos OLAP implementados para el proyecto RUB-ECUADOR.

El citado proyecto está conformado por varios sistemas independientes, cada uno de estos con sus respectivas bases de datos operacionales implementadas en SQL Server 2005. Razón por la cual se necesita desarrollar reportes BI que como su nombre lo indica son reportes inteligentes con un alto nivel de interacción con los usuarios finales, para cada una de las bases de datos operacionales ya existentes en RUB-ECUADOR.

Además se debe aclarar que al citar la frase “productos software” anteriormente señalada, nos referimos al hecho del diseño e implementación de los Datamart, cubos OLAP, y los reportes finales a partir de cada una de las bases de datos operacionales existentes en RUB-ECUADOR, que son las fuentes para nuestra solución BI, la misma que consiste en reportear la información de los cubos OLAP.

Finalmente se enfatiza el hecho de que los reportes inteligentes que resultarán de este trabajo serán completamente independientes para cada uno de los sistemas de RUB-ECUADOR, es decir una solución BI para cada uno de ellos por decirlo así, con sus fuentes que son las bases de datos transaccionales u operacionales desarrolladas en SQL Server 2005, los Datamart desarrollados por medio de vistas en SQL Server 2005, los cubos OLAP desarrollados en Mondrian 3.0.4, y los reportes inteligentes utilizando el navegador diseñado para Mondrian denominado Jpivot.

5.2 METODOLOGÍA PARA DESARROLLO DEL PROYECTO

Para el desarrollo de la construcción de los Datamart, cubos OLAP y reportes inteligentes, se sigue la metodología de Ralph Kimball, dado que establece claros procesos para todo el ciclo del desarrollo del proyecto y garantiza la calidad y eficiencia de la solución de inteligencia de negocios.

Esta metodología fue desarrollada desde el inicio del proceso de construcción, hasta llegar a las etapas de interacción con el usuario y documentación del proyecto.

En las siguientes secciones se describen los procesos realizados para cada fase del proyecto que garantizan su calidad y cumplimiento.

5.2.1 Planeación y Administración del Proyecto

5.2.1.1 Objetivo del Proyecto

Construir los Datamart, cubos OLAP y reportes de Business Intelligence, necesarios para cada uno de los sistemas existentes en el proyecto RUB-ECUADOR, los mismos que serán aplicaciones Web independientes que estarán invocados por medio de enlaces (links) desde cada aplicación Web de los sistemas.

5.2.1.2 Definición del Proyecto

Para el proyecto desarrollado se ha identificado un alto interés por parte de los directivos del Ministerio de Coordinación de Desarrollo Social (MCDS), para el éxito de su implementación.

La demanda del proyecto se encuentra en la necesidad de obtener mejor información de cada uno de los sistemas que conforman RUB-ECUADOR, para tomar mejores decisiones a nivel del ministerio y así mejorar su competitividad y rendimiento. Esta demanda se satisface al implementar para cada sistema una solución de inteligencia de negocios.

Dicha Solución BI está compuesta por uno o varios Datamarts, desarrollados por medio de tablas temporales o también conocidos como vistas esto en SQL Server 2005, en donde se encuentran ya no los simples datos operacionales de la base de datos sino la información útil lista para ser reportada o como en nuestro caso información lista para el respectivo análisis multidimensional por medio de cubos OLAP.

Posteriormente el debido análisis multidimensional para por último definir el reporte de la información desarrollando con herramientas Open Source como Mondrian 3.0.4 su navegador OLAP Jpivot y Java 6.0 como plataforma, además del servidor Web también de código libre denominado Apache Tomcat 6.0.16.

5.2.1.3 Alcance del Proyecto

El presente proyecto de tesis tiene todos los componentes de una solución BI con análisis multidimensional en herramientas Open Source.

Como fuentes se tiene algunas de las bases de datos desarrolladas por PROASETTEL S.A. de cada sistema independiente perteneciente a RUB-ECUADOR como lo explicamos anteriormente, las mismas que a continuación se enumeran:

1. RUB_NUTRIENDO
2. RUB_SOCIOAHORRO
3. RUB_SOCIOSOLIDARIO
4. RUB_MICROFINANZAS

Dichas bases de datos almacenan una gran cantidad de datos operacionales que satisfacen los requerimientos funcionales y no funcionales de cada uno de los sistemas del mismo nombre.

A partir de esto se define mediante la implementación de vistas todos y cada uno de los Datamart que se utilizan para el análisis multidimensional en cubos OLAP y su posterior explotación. Estas vistas hemos desarrollado en SQL Server 2005 que es el DBMS en donde están implementadas las bases de datos.

Finalmente, el análisis OLAP es realizado utilizando el servidor OLAP Mondrian 3.0.4. Todas estas herramientas soportan varios sistemas operativos, incluyendo Microsoft Windows y Linux.

Esto permite una gran portabilidad y la posibilidad de realizar el análisis OLAP vía web, gracias a que el análisis OLAP se realiza utilizando un navegador de Internet.

Los Datamart implementados por medio de vistas, proporcionan información sobre cada uno de los sistemas mencionados en la lista anterior. Para luego presentar el desarrollo y la implementación de los requerimientos de los reportes del Ministerio de Coordinación de Desarrollo Social.

5.2.1.4 Justificación del Proyecto en el Negocio

Además de ser un proyecto académico, el proyecto busca mejorar la productividad y rendimiento en cuanto al reporte de información se refiere en cada uno de los sistemas citados anteriormente del Ministerio de Coordinación de Desarrollo Social MCDS. Como justificación para el negocio se prevén los siguientes beneficios:

1. Soluciones Business Intelligence económicas

Ya que con la realización del denominado análisis multidimensional con herramientas Open Source, estamos dando una solución BI libre en cuanto a la capa BI se refiere para cada uno de los sistemas antes mencionados.

2. Mejora de proceso de toma de decisiones.

Es fundamental considerar este beneficio, puesto que los ejecutivos del MCDS podrán obviamente obtener un mayor nivel en cuánto a tomar decisiones importantes, al tener reportes BI con información actualizada, clara y precisa cuyos reportes además ofrece un alto nivel estadístico y de interacción con el usuario final.

5.2.2 Análisis de Requerimientos

Para el levantamiento de requerimientos se realizaron entrevistas a personas del área técnica y de negocio del Ministerio. Los requerimientos identificados se implementan con la herramienta OLAP Mondrian.

5.2.2.1 Levantamiento de Requerimientos

Para realizar la recolección de requerimientos PROASETTEL S.A. realizó entrevistas con los futuros usuarios de la solución y además con personas del área de tecnología del Ministerio de Coordinación de Desarrollo Social MCDS. Los requerimientos del negocio se listan a continuación:

1. Programa Social Nutriendo el Desarrollo **PSND**

- a. Ver Centros de Acopio vigentes por ubicación.

- b. Ver los períodos del programa.
- c. Ver los precios de venta vigentes del litro de leche por ubicación y Centros de Acopio.
- d. Ver beneficiarios con sus datos personales por ubicación y Centros de Acopio en un período.
- e. Ver la producción diaria de leche por ubicación y Centros de Acopio en un período.
- f. Ver beneficiarios por la capacitación recibida.

2. Programa Social Socio Ahorro **PSSA**

- a. Ver descuentos otorgados a los usuarios en sus compras por ubicación en un rango de fechas.

3. Programa Social Socio Solidario **PSSS**

- a. Ver estadísticas de los ingresos por la venta de productos Socio Solidario por ubicación en un período.

4. Programa Sistema Nacional de Microfinanzas **PSNM**

- a. Ver beneficiarios por nivel de pobreza en un rango de fechas.
- b. Ver beneficiarios por género y ubicación en un rango de fechas.
- c. Ver beneficiarios por Entidad Microfinanciera en un rango de fechas.
- d. Ver beneficiarios por Actividad Microempresarial en un rango de fechas.
- e. Ver beneficiarios por Entidad Microfinanciera y Actividad Microempresarial en un rango de fechas.
- f. Ver beneficiarios por la forma de crédito en un rango de fechas.
- g. Ver beneficiarios por forma de crédito y Entidad Microfinanciera en un rango de fechas.
- h. Ver los beneficiarios del programa en un rango de fechas.

Estos requerimientos fueron los acordados para implementar en reportes estadísticos con la empresa, todos ellos soportados con datos encontrados en las bases de datos respectivamente que funcionan como fuentes.

5.2.2.2 Documentación de Requerimientos

La documentación de estos requerimientos pretende presentar una descripción de cada uno y las fuentes de donde se extraerán los datos para producir la información de la cual cada caso es responsable.

El detalle de cada fuente de datos se encuentra en el apartado 4.2.4. Para conocer la forma como se accede a cada reporte que implementa a cada requerimiento desde la herramienta, ver el anexo “Navegando en el Datamart”.

1. Programa Social Nutriendo el Desarrollo *PSND*

- a.** Ver Centros de Acopio vigentes por ubicación.

Descripción: Listado y cantidad de Centros de Acopio registrados y vigentes por el PSND en general/por provincia.

Fuentes de Datos: Base de datos “RUB_NUTRIENDO”.

- b.** Ver los períodos del programa.

Descripción: Listado y cantidad de periodos registrados en el sistema PSND.

Fuentes de Datos: Base de datos “RUB_NUTRIENDO”.

- c.** Ver los precios de venta vigentes del litro de leche por ubicación y Centros de Acopio.

Descripción: Listado de los precios de venta vigentes del litro de leche por ubicación y Centros de Acopio en general/por provincia.

Fuentes de Datos: Base de datos “RUB_NUTRIENDO”.

- d.** Ver beneficiarios con sus datos personales por ubicación y Centros de Acopio en un período.

Descripción: Listado y cantidad de beneficiarios con sus datos personales registrados y vigentes por el PSND en general/por provincia en un periodo dado.

Fuentes de Datos: Base de datos "RUB_NUTRIENDO".

e. Ver la producción diaria de leche por ubicación y Centros de Acopio en un período.

Descripción: Listado y cantidad de beneficiarios de los Centros de Acopio registrados y vigentes por el PSND en general/por provincia por la producción diaria de litros de leche y su ingreso mensual en un periodo dado.

Fuentes de Datos: Base de datos "RUB_NUTRIENDO".

f. Ver beneficiarios por la capacitación recibida.

Descripción: Listado y cantidad de beneficiarios de los Centros de Acopio registrados y vigentes por el PSND en general/por provincia.

Fuentes de Datos: Base de datos "RUB_NUTRIENDO".

2. Programa Social Socio Ahorro **PSSA**

a. Ver descuentos dados a los usuarios en sus compras por ubicación en un rango de fechas.

Descripción: Mostrar las estadísticas de los descuentos dados a los usuarios en sus compras por provincia, cantón, parroquia y sucursal en un rango de fechas dado.

Fuentes de Datos: Base de datos "RUB_SOCIOAHORRO".

3. Programa Social Socio Solidario **PSSS**

a. Ver estadísticas de los ingresos por la venta de productos Socio Solidario por ubicación en un período.

Descripción: Mostrar las estadísticas de los ingresos por la venta de productos Socio Solidario por provincia, cantón, parroquia y producto en un periodo dado.

Fuentes de Datos: Base de datos "RUB_SOCIOSOLIDARIO".

4. Programa Sistema Nacional de Microfinanzas **PSNM**

- a. Ver beneficiarios por nivel de pobreza en un rango de fechas.

Descripción: Listado y cantidad de beneficiarios por nivel de pobreza, región, provincia, cantón y parroquia en un rango de fechas dado.

Fuentes de Datos: Base de datos "RUB_MICROFINANZAS".

- b. Ver beneficiarios por género y ubicación en un rango de fechas.

Descripción: Listado y cantidad de beneficiarios por estado de crédito, género, región, provincia, cantón y parroquia en un rango de fechas dado.

Fuentes de Datos: Base de datos "RUB_MICROFINANZAS".

- c. Ver beneficiarios por Entidad Microfinanciera en un rango de fechas.

Descripción: Listado y cantidad de beneficiarios por Entidad Microfinanciera y estado de crédito en un rango de fechas dado.

Fuentes de Datos: Base de datos "RUB_MICROFINANZAS".

- d. Ver beneficiarios por Actividad Microempresarial en un rango de fechas.

Descripción: Listado y cantidad de beneficiarios por actividad Microempresarial, estado de crédito, región, provincia, cantón y parroquia en un rango de fechas dado.

Fuentes de Datos: Base de datos "RUB_MICROFINANZAS".

- e. Ver beneficiarios por Entidad Microfinanciera y Actividad Microempresarial en un rango de fechas.

Descripción: Listado y cantidad de beneficiarios por entidad Microfinanciera, actividad Microempresarial y estado de crédito en un rango de fechas dado.

Fuentes de Datos: Base de datos "RUB_MICROFINANZAS".

f. Ver beneficiarios por la forma de crédito en un rango de fechas.

Descripción: Listado y cantidad de beneficiarios por la forma de crédito, estado de crédito, región, provincia, cantón y parroquia en un rango de fechas dado.

Fuentes de Datos: Base de datos "RUB_MICROFINANZAS".

g. Ver beneficiarios por forma de crédito y Entidad Microfinanciera en un rango de fechas.

Descripción: Listado y cantidad de beneficiarios por la entidad Microfinanciera, forma de crédito y estado de crédito en un rango de fechas dado.

Fuentes de Datos: Base de datos "RUB_MICROFINANZAS".

h. Ver los beneficiarios del programa en un rango de fechas.

Descripción: Listado y cantidad de beneficiarios del programa por región, provincia, cantón y parroquia en un rango de fechas dado.

Fuentes de Datos: Base de datos "RUB_MICROFINANZAS".

5.2.3 Modelamiento Dimensional

Para iniciar el Modelamiento dimensional se debe tener en cuenta el principal objetivo de cualquier datamart: el análisis de la información. Este análisis es realizado por medio de reportes, por lo tanto al modelar el datamart se debe tener como objetivo la información deseada en los reportes.

Además se tomará en cuenta que los datamart realizados para cada uno de los sistemas de RUB-ECUADOR ha sido realizado por medio de la creación de vistas en SQL Server 2005 como ya lo hemos explicado anteriormente. En las siguientes secciones se detalla cada componente del modelo dimensional.

5.2.3.1 El Datamart

El modelo diseñado e implementado es de varios datamart. Este concepto se diferencia al de bodega de datos en que el datamart busca centrarse en un objetivo único, o en el análisis de un área específica de la empresa.

En el caso de este proyecto de tesis es analizar cada una de las bases de datos operacionales de los sistemas independiente cada uno de ellos entre sí de RUB-ECUADOR.

Sin embargo, no es posible clasificar a un datamart como una “pequeña bodega de datos”, pues el concepto de datamart o bodega es solo de diseño.

Un datamart de un área de una gran empresa puede ser más complejo y contener un mayor número de datos que toda una bodega de datos de una empresa de menor tamaño, o con diferentes bases de datos.

Los datamart implementados en este proyecto tienen cada uno de ellos como fuentes las bases de datos de cada sistema de RUB-ECUADOR en SQL Server 2005, por esta razón y por la facilidad en el diseño se implementarán vistas utilizando la esquematización que se maneja en los datamart.

Es importante mencionar que en cada una de las dimensiones y tablas de hechos de nuestros datamart se agregó información adicional la misma que sirve para permitir al usuario final el poder necesario para personalizar sus reportes, de manera que pueda habilitar y/o ocultar en tiempo de ejecución las dimensiones y medidas adicionales en cada uno de los reportes estadísticos Business Intelligence que se ha desarrollado.

5.2.3.2 Definición de la granularidad

Se definió la granularidad de las tablas de hechos como las más bajas o granulares posibles. Esta granularidad corresponde a por ejemplo:

1. Los ingresos, producción diaria de litros de leche, precios de litro de leche vigente, entre otros dado un periodo, en el caso del PSND.
2. Los descuentos dados a los usuarios en un rango de fechas determinado, es decir una transacción individual en el caso de PSSA.
3. Los ingresos por la venta de cada uno de los productos socio solidarios en el caso del sistema con el mismo nombre PSSS, todo ello en un periodo dado.
4. Los montos y las tasas de crédito totales y promedio de los beneficiarios para el Programa Sistema Nacional de Microfinanzas PSNM, en un rango de fechas dado por el usuario.

De esta forma será posible llegar al grado de detalle de consultar registros de manera específica, aunque este no sea el objetivo de un datamart. Las medidas, que son los campos de valor de las tablas de hechos son los valores con la granularidad establecida.

5.2.3.3 Dimensiones

Se definen las dimensiones que soportan los requerimientos definidos, cumpliendo con la granularidad de la tabla de hechos. Las siguientes secciones relacionan las tablas diseñadas para la base de datos con su dimensión correspondiente.

1. PROGRAMA SOCIAL NUTRIENDO EL DESARROLLO (PSND)

a. Dimensión Centros de Acopio

Vista: DIM_CENTROSACOPIO

Contiene la información necesaria acerca de los Centros de Acopio registrados y vigentes en el programa, tales como la provincia, cantón, parroquia, la descripción del Centro de Acopio seguido de su teléfono, y más información adicional como el período, las fechas de inicio y fin de ese periodo, el responsable de ese Centro de Acopio.

Campo Nombre
Fact_PK
PRONOMBRE
CANNOMBRE
PARNOMBRE
PERCODIGO
PERFECHAINICIO
PERFECHAFINAL
PERVIGENTE
PERDESCRIPCION
CENARESPONSABLE
CENATELEFONOS
CENADIRCALLE2
CENADIRNUMERO
CENADIRCALLE1
CENACODIGO
CENADESCRIPCION

Tabla V.17. Campos de la vista Dimensión Centros de Acopio

b. Dimensión Periodos del Programa

Vista: DIM_PERIODOSPROGRAMA

Contiene la información necesaria acerca de la totalidad de periodos registrados en el sistema PSND, además se adiciona una columna para que el usuario final observe si el periodo está vigente o no.

Campo Nombre
PERCODIGO
PERVIGENTE
PERDESCRIPCION
PERFECHAINICIO
PERFECHAFINAL

Tabla V.18. Campos de la vista Dimensión Periodos del Programa

c. Dimensión Precio Venta por Litro Leche

Vista: DIM_PRECIOVENTALECHE

Contiene la información necesaria acerca de los precios de venta del litro de leche vigente por los centros de acopio por provincia, cantón, parroquia y descripción de cada centro de acopio.

Campo Nombre
Fact_PK
PERVIGENTE
CENACODIGO
CENADESCRIPCION
PERCODIGO
PERDESCRIPCION
PRONOMBRE
CANNOMBRE
PARNOMBRE

Tabla V.19. Campos de la vista Dimensión Precio Venta por Litro Leche

d. Dimensión Beneficiarios por Ubicación y Datos

Vista: DIM_BENEFICIARIOSDATOS

Contiene la información necesaria acerca los beneficiarios del PSND con sus datos personales como teléfonos, género, nivel de instrucción, si recibe o no el bono de desarrollo humano, el número de su quintil si lo tiene, el estado civil y la descripción del nivel de pobreza.

Además se cita el período ya que al final en el reporte BI estadístico Web se necesitará esta información porque se da al usuario final la opción de escoger en que periodo quisiera ver los datos, teniendo por defecto el periodo vigente. A continuación en la siguiente página se muestra los campos de esta vista:

Campo Nombre
Fact_PK
CENACODIGO
CENADESCRIPCION
PRONOMBRE
PERVIGENTE
PERCODIGO
PERFECHAINICIO
PERFECHAFINAL
PROLCEDULA
NOMBRESCOMPLETOS
SEXDESCRIPCION
NIVDESCRIPCION
PROLFECHAINGPROGRAMA
PROLRECIBEBDH
PROLTELEFONOS
PERDESCRIPCION
ESTCDESCRIPCION
NIVPDESCRIPCION
CANNOMBRE
PARNOMBRE
QUIDDESCRIPCION

Tabla V.20: Campos de la vista Dimensión Beneficiarios por Ubicación y Datos

e. Dimensión Producción Diaria Leche

Vista: DIM_PRODUCIONLECHE

Contiene la información necesaria acerca los beneficiarios del PSND acompañada del Centro de Acopio, el periodo, y también la producción de litros de leche en el tanque, para su casa y otros, esto para que el usuario final en el reporte BI la posibilidad de verlos y mostrar u ocultar. Además se da al usuario final la opción de escoger en que periodo quisiera ver los datos, teniendo por defecto el periodo vigente.

Campo Nombre
Fact_PK
CENACODIGO
CENADESCRIPCION
PRONOMBRE
PERVIGENTE
PERCODIGO
PERFECHAINICIO
PERFECHAFINAL
PROLCEDULA
NOMBRESCOMPLETOS
SEXDESCRIPCION
NIVIDESCIPCION
PROLFECHAINGPROGRAMA
PROLTELEFONOS
PROLPRODCASA
PROLPRODTANQUE
PROLOTROS
PERDESCRIPCION
CANNOMBRE
PARNOMBRE

Tabla V.21: Campos de la vista Dimensión Producción Diaria Leche

f. Dimensión Beneficiarios por Capacitación

Vista: DIM_BENEFICIARIOS_CAPACITACION

Contiene la información necesaria acerca los beneficiarios del PSND por la capacitación recibida adjunto la información del curso, como su nombre, lugar donde se desarrolló el curso, fecha de inicio y fecha de finalización del curso.

Además se da al usuario final la opción de escoger en que periodo quisiera ver los datos, teniendo por defecto el periodo vigente.

Campo Nombre
NOMBRESCOMPLETOS
SEXDESCRIPCION
NIVIDESCIPCION
PROLFECHAINGPROGRAMA
PROLTELEFONOS
CURCFECHAINICIO
CURCFECHAFINALIZACION
PROLCEDULA
Fact_PK
PRONOMBRE
CANNOMBRE
CURCLUGAR
CURCNOMBRE

Tabla V.22: Campos de la vista Dimensión Beneficiarios por Capacitación

2. PROGRAMA SOCIAL SOCIO AHORRO (PSSA)

a. Dimensión Descuentos Compras

Vista: DIM_DESCUENTOS

Contiene la información necesaria acerca los beneficiarios del PSSA por los descuentos dados a los mismos en sus compras, añadiendo la provincia, cantón, parroquia, sucursal y la fecha de la transacción.

Esta fecha es necesaria puesto que se da al usuario final en el reporte BI estadístico web la posibilidad de elegir un rango de fechas.

Campo Nombre
FACT_PK
PRONOMBRE
CANNOMBRE
PARNOMBRE
SUCDESCRIPCION
BENNOMBRECOMPLETO
TRAFECHAHORA

Tabla V.23: Campos de la vista Dimensión Descuentos Compras

3. PROGRAMA SOCIAL SOCIO SOLIDARIO (PSSS)

a. Dimensión Ingresos por Ventas

Vista: Dimension_1

Contiene los ingresos por la venta de productos socio solidarios de los beneficiarios, adjunto la información del nombre de la empresa, la marca del producto, el producto con su código de barras.

Además se da la posibilidad a los usuarios finales de que en el reporte estadístico tenga la posibilidad de elegir el periodo.

Campo Nombre
Fact_PK
PERDESCRIPCION
PRONOMBRE
CANNOMBRE
PARNOMBRE
EMPNOMBRE
MARDESCRIPCION
PROCODIGOBARRA
PRODUCTO
PROVENTASUPERMERCADOS
PROVENTATIENDAS
PREDESCRIPCION
PERCODIGO
PERFECHAINICIO
PERFECHAFINAL

Tabla V.24: Campos de la vista Dimensión Ingresos por Ventas

4. PROGRAMA SISTEMA NACIONAL DE MICROFINANZAS (PSNM)

a. Dimensión Créditos en General

Vista: Dimension_1

Contiene básicamente toda la información de los beneficiarios de los microcréditos, junto con la forma de crédito, el estado de ese crédito, y la fecha de crédito, la misma que sirve para dar la posibilidad al usuario final de escoger un rango de fechas.

Campo Nombre
Fact_Pk
SOLNUMERO
POSBNUMSOCIO
ORDENREDITO
CREDITO
REGNOMBRE
PRONOMBRE
CANNOMBRE
PARNOMBRE
SEXDESCRIPCION
POSBPLAZOCREDITO
POSBFECHACREDITO
POSBFECHACREDITO1
ENTMCODIGO
ENTMNOMBRE
TIPIDESCIPCION
ENTMFECHAINGPROGRAMA
ACTMDESCRIPCION
NIVPDESCRIPCION
BENEFICIARIOS
POSBCEDULA
POSBTELEFONOS

Tabla V.25: Campos de la vista Dimensión Créditos en General

5.2.3.4 Tablas Virtuales de Hechos

A continuación se detallan las tablas de hechos para cada uno de los sistemas de igual forma como anteriormente se lo hizo con las dimensiones:

1. PROGRAMA SOCIAL NUTRIENDO EL DESARROLLO (PSND)

a. Fact Centros de Acopio

Vista: FACT_CENTROSACOPIO

Contiene la información necesaria acerca de los valores numéricos que son las medidas para los Centros de Acopio, como son: el número de tanques que existe en cada uno de los centros y la capacidad de litros de leche de cada Centro. Su dimensión es la denominada DIM_ CENTROSACOPIO de la base de datos RUB_NUTRIENDO.

Campo Nombre
Fact_PK
CENANUMTANQUES
CENACAPLITROSLECHE

Tabla V.26: Campos de la vista Fact Centros de Acopio

b. Fact Periodos del Programa

Vista: FACT_PERIODOSPROGRAMA

Contiene la medida periodo, pues de esta forma se cuenta estos campos de manera que el usuario final sepa el número de periodos existentes y los pueda desplegar. Su dimensión es la denominada DIM_ PERIODOSPROGRAMA de la base de datos RUB_NUTRIENDO.

Campo Nombre
PERCODIGO

Tabla V.27: Campos de la vista Fact Periodos del Programa

c. Fact Precio Venta por Litro Leche

Vista: FACT_PRECIOVENTALECHE

Contiene la información necesaria para mostrar al usuario final un promedio del precio de litro de leche, de manera que dicho usuario al final puede ver en su reporte en general y también en forma desplegada es decir cuánto es el precio del litro de leche en cada Centro de Acopio, provincia, cantón y parroquia. Su dimensión es la denominada DIM_PRECIOVENTALECHE de la base de datos RUB_NUTRIENDO.

Campo Nombre
Fact_PK
PRELPRECIOLITROLECHE

Tabla V.28: Campos de la vista Fact Precio Venta por Litro Leche

d. Fact Beneficiarios por Ubicación y Datos

Vista: FACT_BENEFICIARIOSDATOS

Contiene la medida cédula puesto que en este datamart se necesita saber solo el número de usuarios por provincia, cantón y parroquia, donde el mayor nivel de granularidad es la cédula. Su dimensión es la denominada DIM_BENEFICIARIOSDATOS de la base de datos RUB_NUTRIENDO.

Campo Nombre
Fact_PK
PROLCEDULA

Tabla V.29: Campos de la vista Fact Beneficiarios por Ubicación y Datos

e. Fact Producción Diaria Leche

Vista: FACT_PRODUCIONLECHE

Contiene las medidas del total de litros diarios de leche, la producción de litros de leche que hay en el tanque, el precio por litro, y además el ingreso promedio mensual, esto para dar luz al requerimiento del reporte estadístico BI acerca de la producción diaria de leche, y su dimensión es la denominada DIM_ PRODUCCIONLECHE de la base de datos RUB_NUTRIENDO.

Campo Nombre
Fact_PK
TOTALLITROSLECHE
PROLPRODTANQUE
PROLINGPROMMENSUAL
PRELPRECIOLITROLECHE

Tabla V.30: Campos de la vista Fact Producción Diaria Leche

f. Fact Beneficiarios por Capacitación

Vista: FACT_BENEFICIARIOS_CAPACITACION

Contiene las medidas de las cédulas del beneficiario en primer lugar para ver el número de ellos en un curso de capacitación, el número de cursos que tiene cada capacitación y además el costo total subsidiado para dar un total en cuanto a provincias, cantones y nombres de los cursos de capacitación cuya información se encuentra en la dimensión, la misma que es la denominada DIM_ BENEFICIARIOS_CAPACITACION de la base de datos RUB_NUTRIENDO.

Campo Nombre
Fact_PK
PROLCEDULA
CURCNUMERO
CURCCOSTOTOTALSUBSIDIADO

Tabla V.31: Campos de la vista Fact Beneficiarios por Capacitación

2. PROGRAMA SOCIAL SOCIO AHORRO (PSSA)

a. Fact Descuentos Compras

Vista: FACT_DESCUENTOS

Contiene las medidas relacionada a la fecha de la transacción realizada esto es para poder desplegar dado un rango de fechas, luego contiene otra medida que es llamada ValorDescuento y como su nombre lo indica es la que determina los descuentos dados a los usuarios en sus transacciones diarias, para este sistema. Su dimensión es la denominada DIM_ DESCUENTOS de la base de datos RUB_SOCIOAHORRO.

Campo Nombre
FACT_PK
TRAVALORDESCUENTO
TRAFECHAHORA

Tabla V.32: Campos de la vista Fact Descuentos Compras

3. PROGRAMA SOCIAL SOCIO SOLIDARIO (PSSS)

a. Fact Ingresos por Ventas

Vista: FACT_1

Contiene las medidas del precio de venta de cada producto, el porcentaje de descuento mínimo, la cantidad de unidades vendidas para las tiendas y supermercados, el volumen de las ventas para las tiendas y supermercados.

De manera que el resultado es un reporte con un alto nivel de interacción con los usuarios finales puesto que ellos pueden desde ir desplegando su información de acuerdo a las dimensiones del cubo OLAP, hasta ir graficando sus resultados mostrados al igual que todo reporte estadístico implementado en este proyecto.

Su dimensión es la denominada DIMENSION_1 de la base de datos RUB_SOCIOSOLIDARIO.

Campo Nombre
Fact_PK
PROPVP
PROPORCDESCUENTOMIN
REGEANTUNIDVENDIDASSUP
REGEVOLUMENVENTASSUP
REGEANTUNIDVENDIDASTIE
REGEVOLUMENVENTASTIE
PERCODIGO

Tabla V.33: Campos de la vista Fact Ingresos por Ventas

4. PROGRAMA SISTEMA NACIONAL DE MICROFINANZAS (PSNM)

a. Fact Créditos en General

Vista: FACT

Contiene las medidas relacionadas al monto y tasas de crédito que se da a los usuarios en el PSNM, además hay una medida para las fechas, la misma que sirve para dar la opción a los usuarios finales de filtrar sus resultados en un rango de fechas dado.

Estos montos y tasas sirven para determinar en el reporte final el total así como el promedio de cada uno de ellos ya sea montos como tasas de crédito. Su dimensión es la denominada DIMENSION_1 De la base de datos RUB_MICROFINANZAS.

Campo Nombre
Fact_Pk
POSBFECHACREDITO1
POSBMONTOCREDITO
POSBTASACREDITO

Tabla V.34: Campos de la vista Fact Créditos en General

5.2.3.5 Diseño del modelo dimensional

Luego de haber determinado las vistas que funcionarán como dimensiones y tablas de hechos para los diferentes programas es decir PSND, PSSA, PSSS y PSNM y de haber presentado una solución de varios datamarts por cada sistema en este caso siete para PSND, y un datamart para PSSA, PSSS y PSNM con relación de uno a uno por decirlo así.

Ahora se presenta el diseño multidimensional de cada datamart creado de acuerdo cada uno de los sistemas detallados a continuación:

1. PROGRAMA SOCIAL NUTRIENDO EL DESARROLLO (PSND)

a. Datamart: Centros de Acopio vigentes por ubicación.

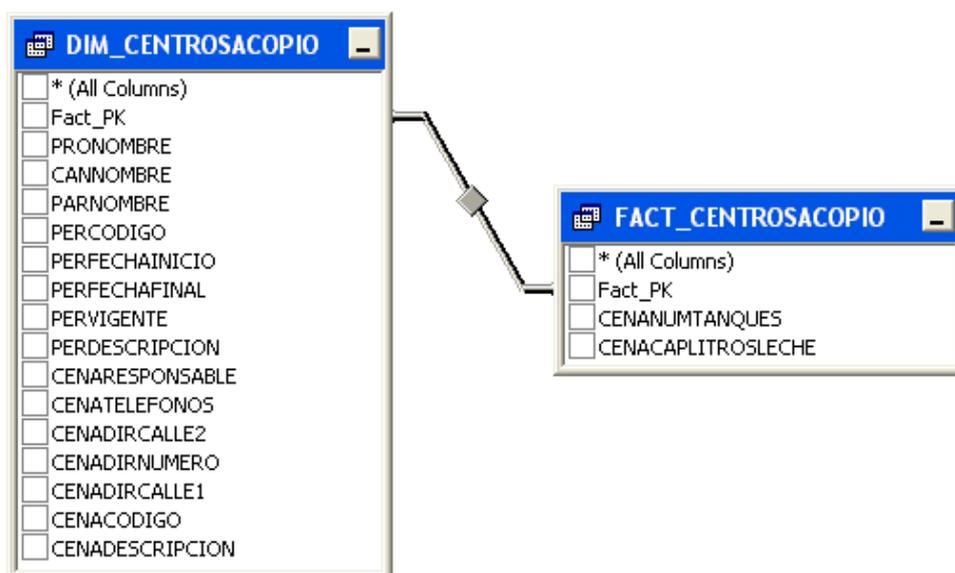


Figura V.39. Datamart Centros de Acopio vigentes por ubicación.

b. Datamart: períodos del programa.

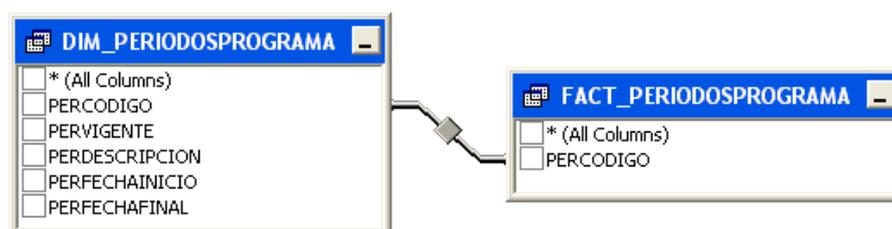


Figura V.40. Datamart períodos del programa.

- c. Datamart: precios de venta vigentes del litro de leche por ubicación y Centros de Acopio.

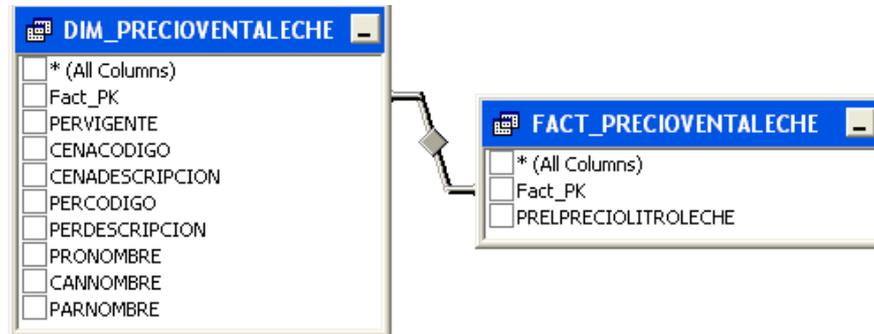


Figura V.41: Datamart precios vigentes del litro de leche.

- d. Datamart: beneficiarios con sus datos personales por ubicación y Centros de Acopio en un período.

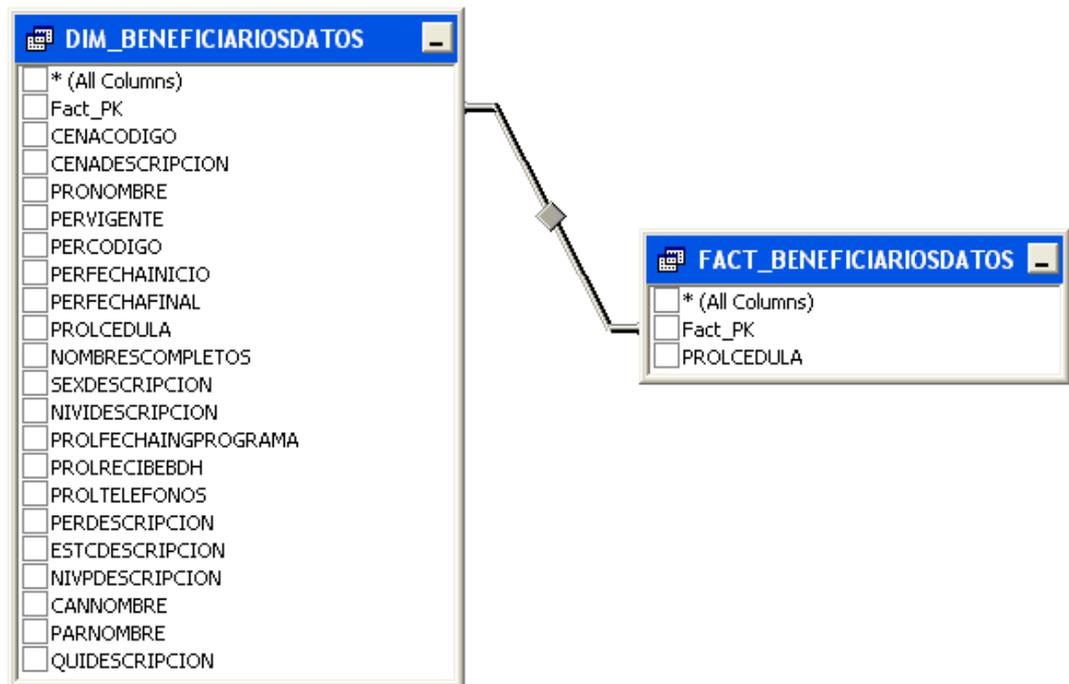


Figura V.42. Datamart beneficiarios datos personales

- e. Datamart: producción diaria de leche por ubicación y Centros de Acopio en un período.

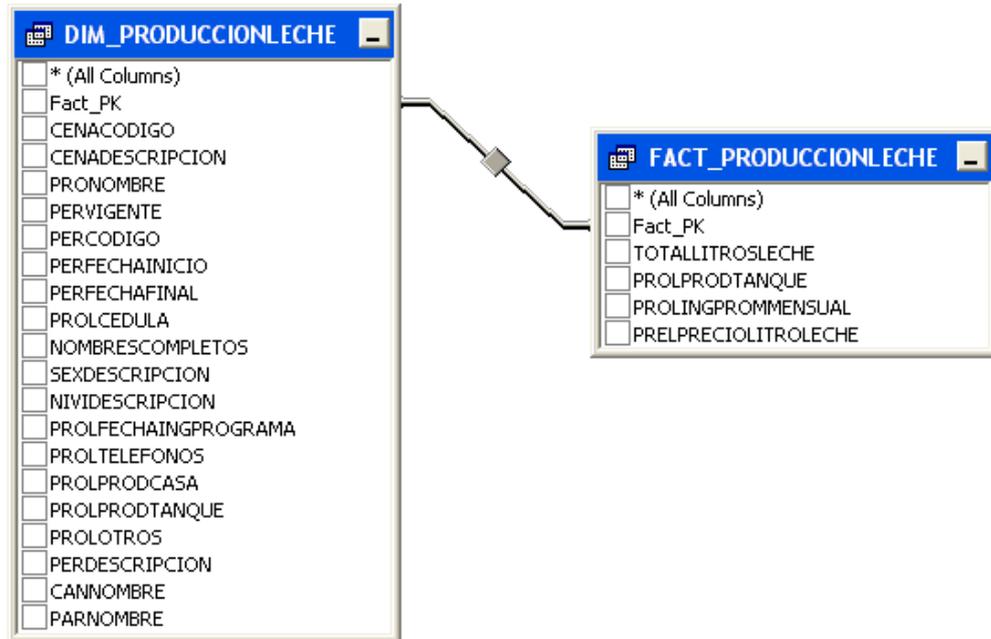


Figura V.43. Datamart producción diaria de leche

- f. Datamart: beneficiarios por la capacitación recibida.

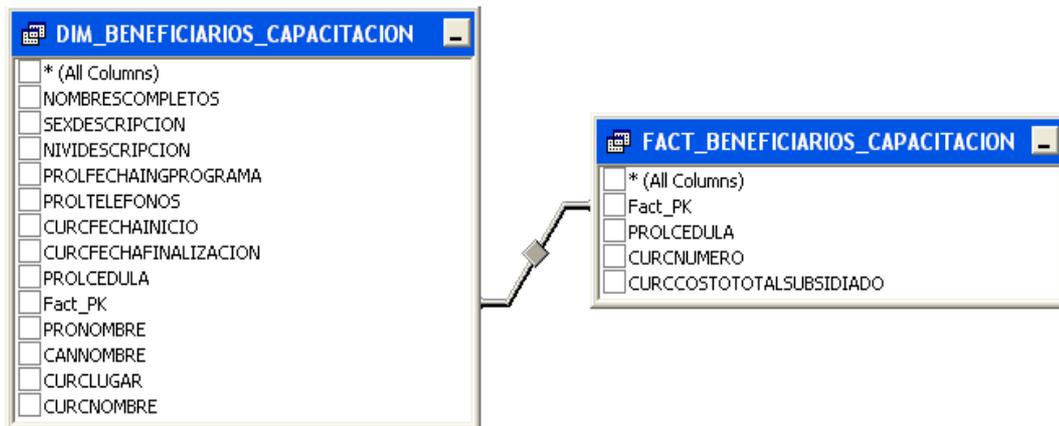


Figura V.44. Datamart beneficiario por capacitación

2. PROGRAMA SOCIAL SOCIO AHORRO (PSSA)

- a. Datamart: descuentos dados a los usuarios en sus compras por ubicación en un rango de fechas.

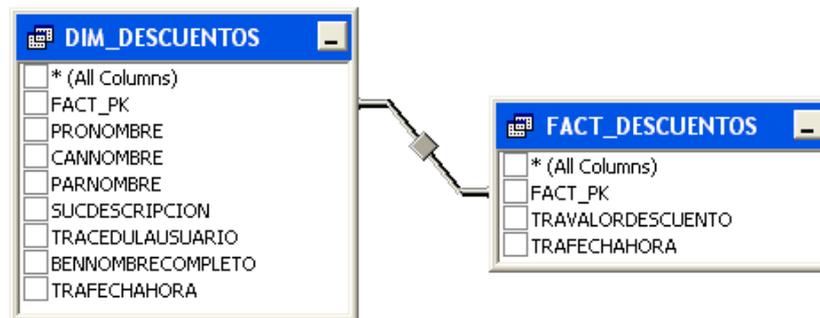


Figura V.45. Datamart descuentos en sus compras

3. PROGRAMA SOCIAL SOCIO SOLIDARIO (PSSS)

- a. Datamart: estadísticas de los ingresos por la venta de productos Socio Solidario por ubicación en un período.

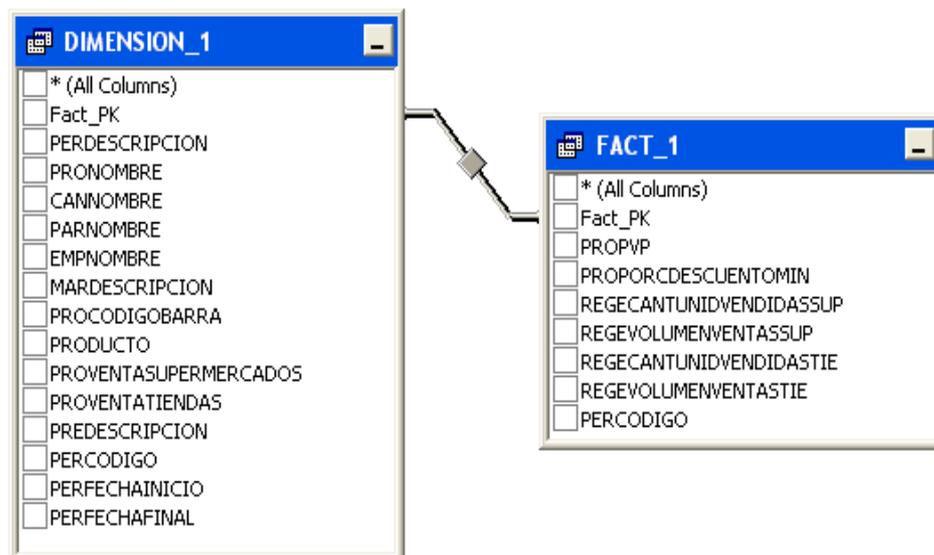


Figura V.46. Datamart ingresos por venta de productos socio solidario

4. PROGRAMA SISTEMA NACIONAL DE MICROFINANZAS (PSNM)

a. Datamart: créditos concedidos a los beneficiarios en general.

Aquí es importante mencionar que se desarrolló un solo datamart para los 8 reportes de requerimientos anotados anteriormente del Programa Sistema Nacional de Microfinanzas puesto que para todos ellos se necesita tener las mismas medidas es decir de créditos y tasas de créditos además de las fechas necesarias a la hora de filtrar la información por rangos de fechas.

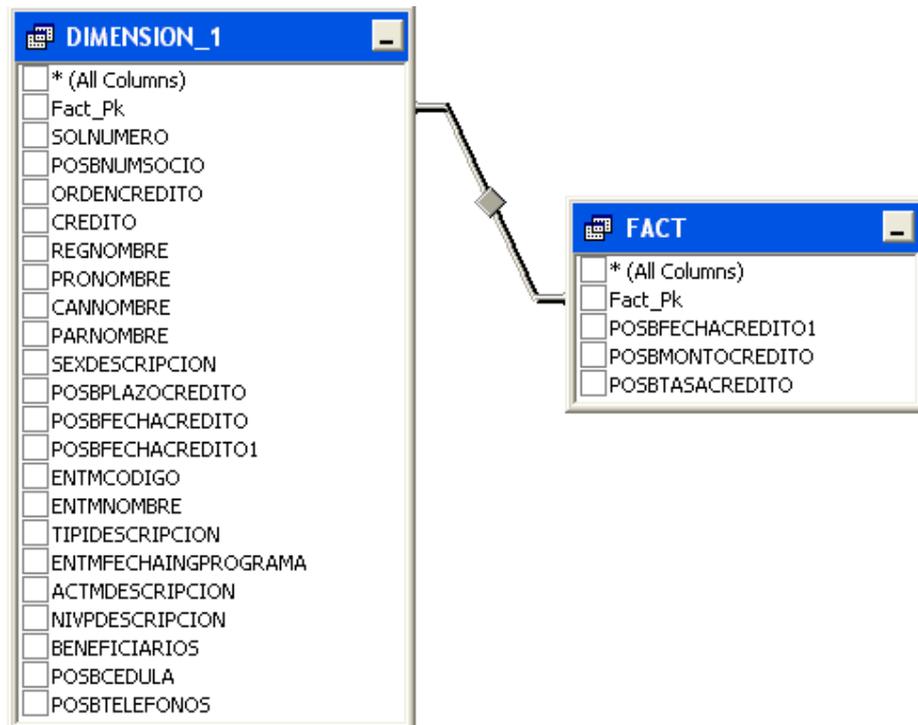


Figura V.47. Datamart créditos concedidos en general

5.2.4 Diseño Técnico de la Arquitectura

La solución es un sistema de información que se conforma de varias tecnologías utilizadas para implementar la solución orientada al usuario final, con la capacidad de integrar los datos que produce cada Programa Social y transformarlos en información activa y productiva para la toma de decisiones.

Por lo expuesto, el sistema de información se enmarca en la categoría de un sistema del tipo "Inteligencia de Negocio" (Business Intelligence) apoyada en las Bases de Datos de los Programas Sociales y dentro de ellas Vistas que definen las estructuras de los Datamarts por cada Programa para luego formar el cubo con Mondrian y poder observar y navegar en el Cubo mediante el Visor Web JPivot.

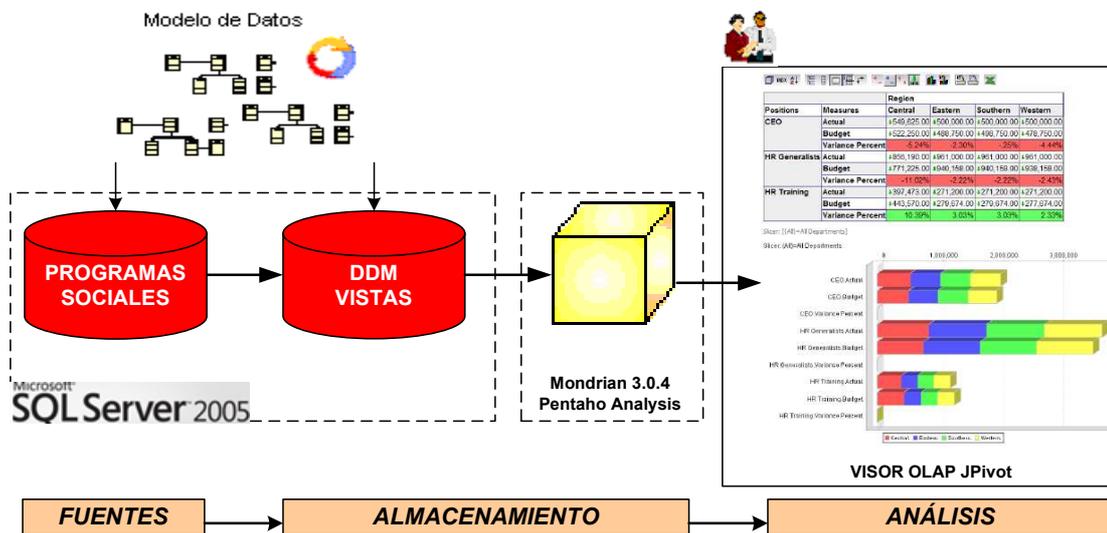


Figura V.48. V.Arquitectura de la Solución

5.2.4.1 Datos

Los datos constituyen la información de las bases de datos operacionales que son las fuentes, estos datos son los que alimentan las vistas creadas como datamarts las mismas que son siempre actualizadas puesto que cada vez que una vista es invocada, ella devuelve únicamente los datos actualizados de la base de datos.

Es decir para esta solución no es necesario realizar procesos ETL. Entonces para el análisis de datos se empieza por analizar los datos fuentes que están en las tablas de las bases de datos operacionales citados anteriormente.

Proasetel S.A. ha desarrollado modelos E – R para los sistemas OLTP también desarrollados por la misma entidad.

Estas bases de datos se encuentran implementadas en SQL Server 2005 y son llamadas RUB_NUTRIENDO, RUB_SOCIOAHORRO, RUB_SOCIOSOLIDARIO y RUB_MICROFINANZAS.

MAPEO DE LOS DATOS EN LOS MODELOS DIMENSIONALES

Para consolidar los datos que formaran parte de los datamart, se ha mapeado los datos en vistas denominadas "GENERALES" y "ESTADISTICAS" en otros casos, las mismas que tienen las columnas que se convertirán en dimensiones y medidas de los datamarts.

Con el objetivo de facilitar el desarrollo de las dimensiones y medidas de cada uno de los datamarts construidos por medio de vistas para los sistemas OLTP que se mencionaron anteriormente, y son las siguientes:

1. RUB_NUTRIENDO

- a. Tablas necesarias para el desarrollo del datamart: Centros de Acopio vigentes por ubicación

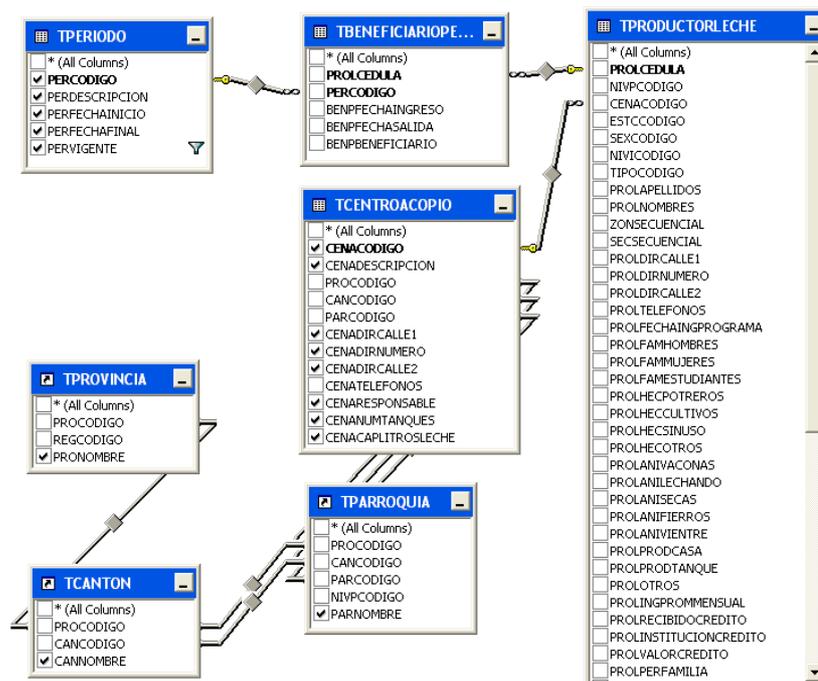


Figura V.49. Tablas para desarrollo de datamart Centros de Acopio por ubicación

- b. Tabla necesaria para el desarrollo del datamart: períodos del programa

TPERIDO	
<input type="checkbox"/>	* (All Columns)
<input checked="" type="checkbox"/>	PERCODIGO
<input checked="" type="checkbox"/>	PERDESCRIPCION
<input checked="" type="checkbox"/>	PERFECHAINICIO
<input checked="" type="checkbox"/>	PERFECHAFINAL
<input type="checkbox"/>	PERVIGENTE

Figura V.50. Tabla para desarrollo del datamart periodos del programa

- c. Tablas necesarias para el desarrollo del datamart: precios de venta vigentes del litro de leche por ubicación y Centros de Acopio

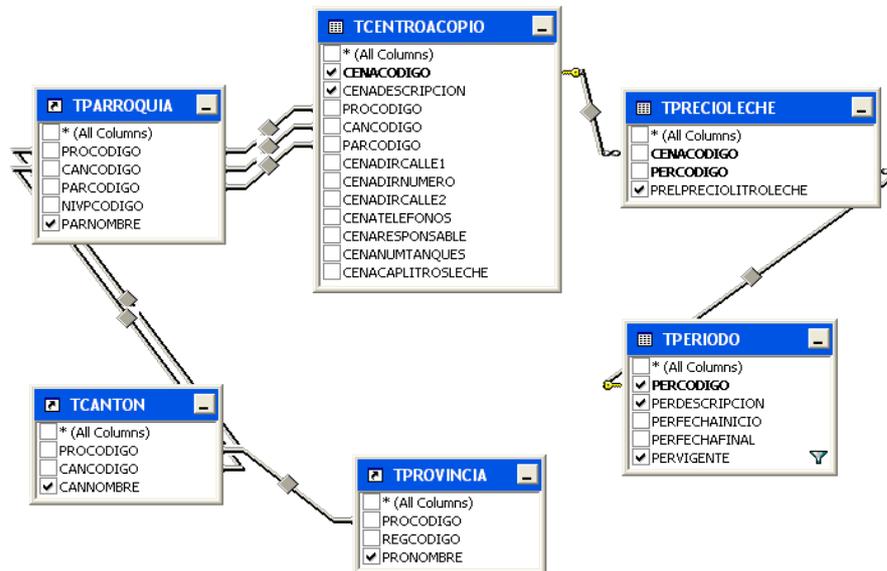


Figura V.51. Tablas para desarrollo del datamart precios vigentes de litro de leche

- d. Tablas necesarias para el desarrollo del datamart: beneficiarios con sus datos personales por ubicación y Centros de Acopio en un período

En la siguiente página se muestra el diagrama de las tablas necesarias para la construcción de las vistas que representan las dimensiones y hechos para el datamart de los beneficiarios con sus datos personales, de la base de datos RUB_NUTRIENDO.

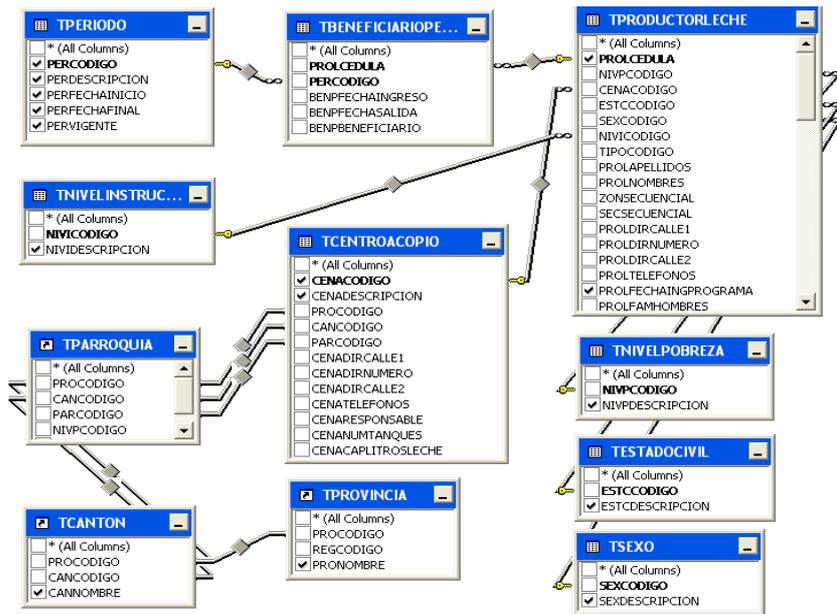


Figura V.52. Tablas para desarrollo del datamart beneficiarios datos personales

- e. Tablas necesarias para el desarrollo del datamart: producción diaria de leche por ubicación y Centros de Acopio en un período

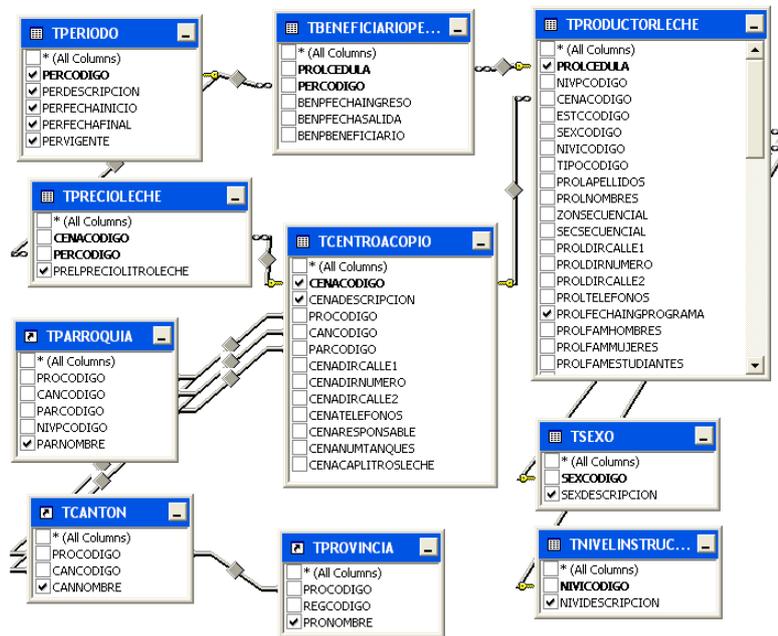


Figura V.53. Tablas para desarrollo del datamart producción diaria de leche

- f. Tablas necesarias para el desarrollo del datamart: beneficiarios por la capacitación recibida

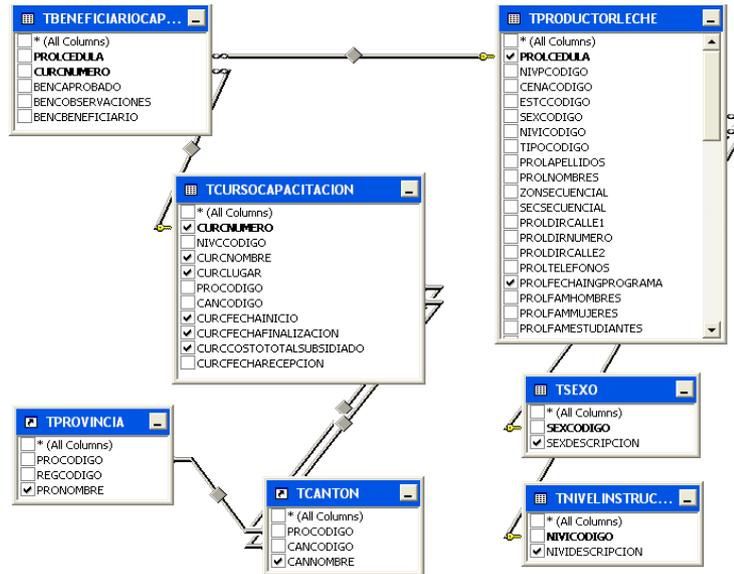


Figura V.54. Tablas para desarrollo del datamart beneficiarios por capacitación recibida

2. RUB_SOCIOAHORRO

A continuación mostraremos las tablas necesarias para el desarrollo del datamart para el PSSA, donde además se da a los usuarios finales la posibilidad de filtrar los datos en un rango de fechas.

Al igual que para el sistema anterior es decir el PSND, los reportes estadísticos business intelligence desarrollados para los sistemas PSSA, PSSS y PSNM están desarrollados con información geográfica como provincias, cantones y parroquia además regiones en el caso de PSNM, esto se logra con la ayuda de los SINÓNIMOS de SQL Server 2005, puesto que esto está en otra base de datos denominada RUB_GEOREFERENCIAL.

- a. Tablas necesarias para el desarrollo del datamart: descuentos dados a los usuarios en sus compras por ubicación en un rango de fechas

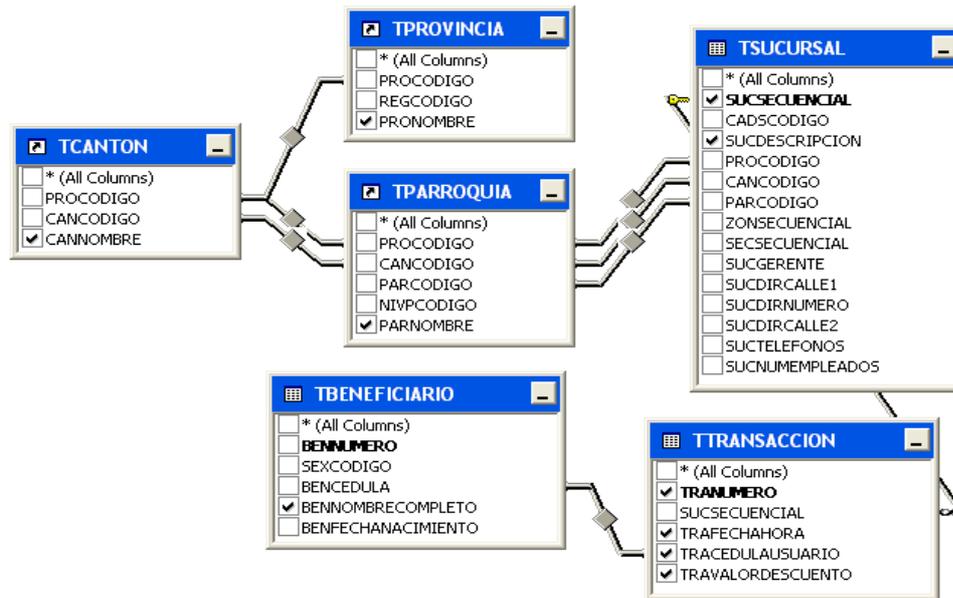


Figura V.55. Tablas para desarrollo del datamart descuentos en sus compras

3. RUB_SOCIOSOLIDARIO

- a. Tablas necesarias para el desarrollo del datamart: estadísticas de los ingresos por la venta de productos Socio Solidario por ubicación en un período

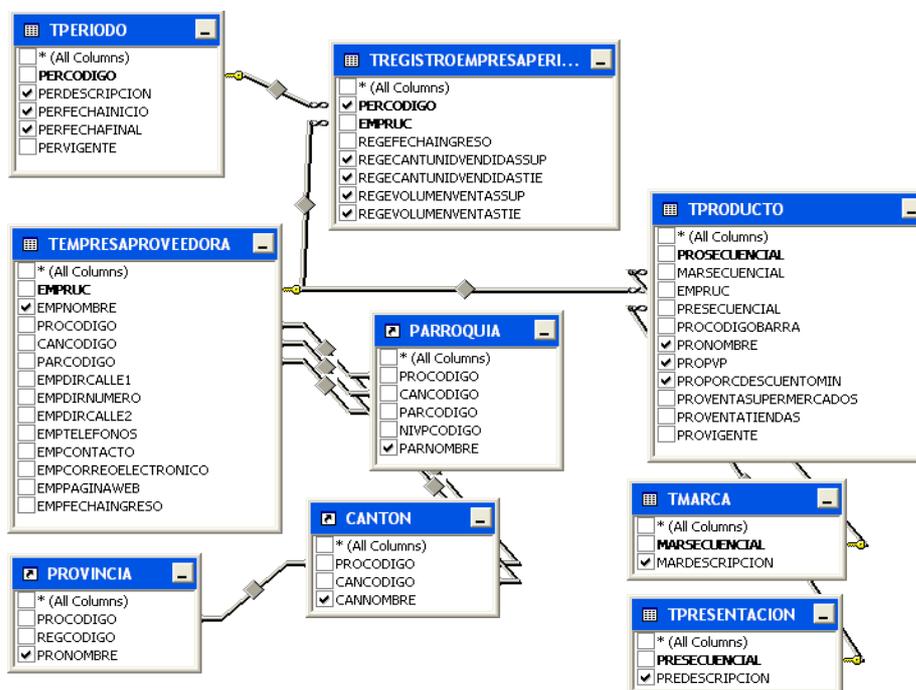


Figura V.56. Tablas para desarrollo del datamart ingresos por venta de productos

4. RUB_MICROFINANZAS

- a. Tablas necesarias para el desarrollo del datamart: créditos concedidos a los beneficiarios en general

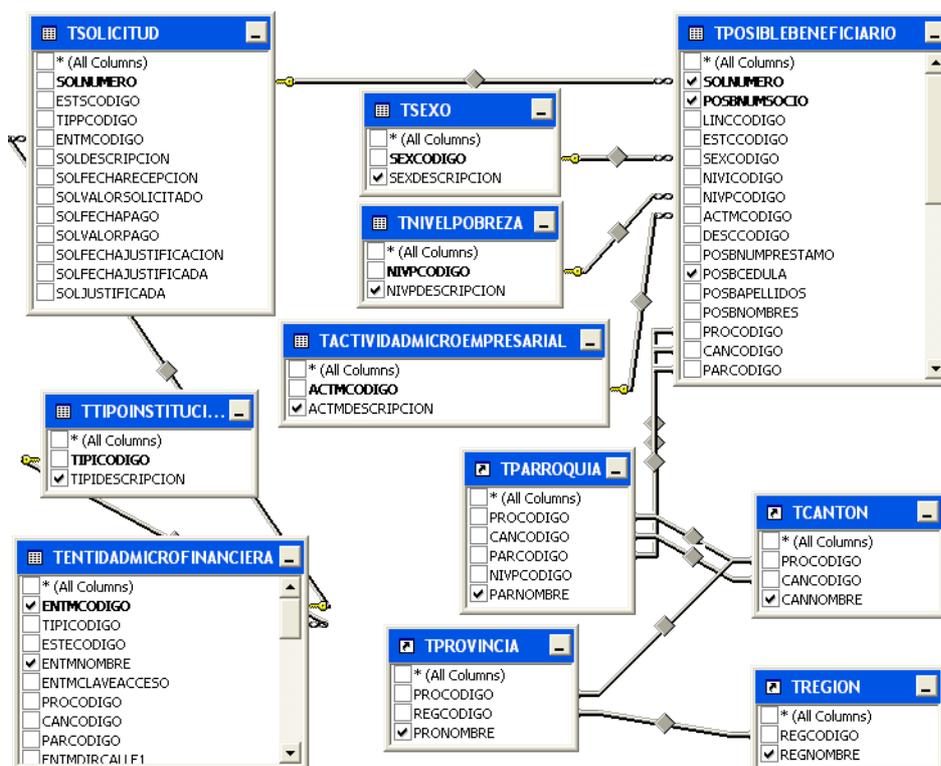


Figura V.57. Tablas para desarrollo del datamart Créditos concedidos en general

5.2.4.2 Back Room

En esta solución no es necesario realizar o desarrollar una herramienta de ETL (Extracción, Transformación y Carga) para los datamart. Como en la mayoría de veces para cargar los datos de un datamart primero se hace los procesos ETL para cargar la información en el data warehouse, posteriormente si la solución BI es de datamarts entonces hay que realizar otro proceso ETL de carga de datos para los respectivos datamarts.

Se ha desarrollado varios datamarts que soporten el análisis multidimensional a partir de vistas en SQL Server 2005 y como ya conocemos cada vez que una vista es invocada en este caso por los esquemas multidimensionales es decir los cubos OLAP que hemos desarrollado en el servidor Mondrian 3.0.4, al ejecutarse esas vista devuelven un resultado actualizado de los datos de las bases de datos que en este caso estamos utilizando esto es lo mismo que

persigue una herramienta ETL puesto que esta se encarga como su nombre lo dice de extraer los datos, transformar los datos y cargar los mismos en los esquemas que se cree necesario de una forma actualizada.

Con esto que hemos descrito estamos ya listos para utilizar una herramienta OLAP o de análisis multidimensional. La herramienta que se ha escogido para esto es la herramienta Open Source de análisis multidimensional Mondrian 3.0.4, teniendo como resultados esquemas multidimensionales en estrellas para su posterior reporte de la información business intelligence.

5.2.4.3 Front Room

Cada datamart realizado para el MCDS (Ministerio de Coordinación de Desarrollo Social) están estructurados de tal forma que se puede observar la información solicitada de cada requerimiento, en dimensiones como productos, beneficiarios, geografía (región, provincia, cantón y parroquia), centros de acopio, periodos, sucursales entre otras y medidas relacionadas a precios, descuentos, ingresos, montos de crédito, tasa de interés entre otras. Toda la información de las columnas descritas en el párrafo anterior provienen de los datamarts de las bases de datos:

1. RUB_NUTRIENDO,
2. RUB_SOCIOAHORRO,
3. RUB_SOCIOSOLIDARIO,
4. RUB_MICROFINAZAS

Cada uno de estos esquemas datamarts son invocados por los otros esquemas OLAP es decir los cubos desarrollados en el servidor OLAP Mondrian 3.0.4.

Para posteriormente obtener reportes estadísticos business intelligence utilizando el navegador propio de Mondrian denominado Jpivot cuyas herramientas son totalmente Open Source.

5.2.4.4 Infraestructura

Las bases de datos operacionales o también denominadas transaccionales o sistemas OLTP y el análisis multidimensional junto con los reportes web business intelligence se ejecutarán en un servidor de base de datos y un servidor de análisis respectivamente que actualmente posee el Ministerio de Coordinación de Desarrollo Social MCDS.

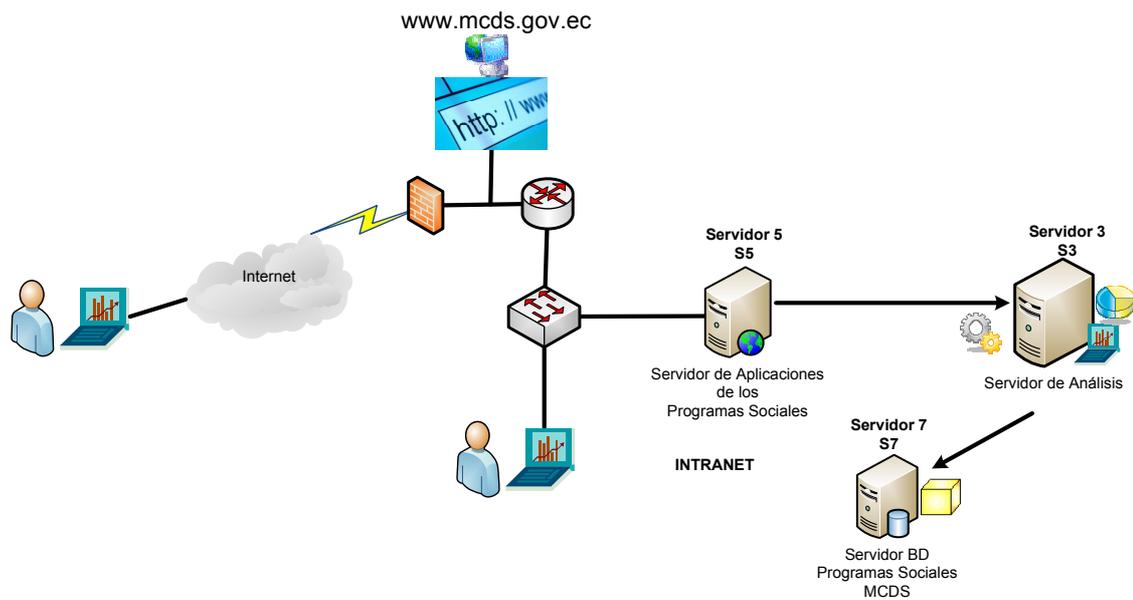


Figura V.58. Infraestructura de la Solución

Las características actuales de estos equipos se muestran a continuación en la siguiente tabla ilustrativa de los servidores:

Servidores Blade	
Chasis Blade H	Descripción
Chasis Blade Center	Un chasis IBM Blade Center Modelo H con 14 Slots cada uno para Servidores Blade.
Componentes de Chasis 	<ul style="list-style-type: none"> ✓ 2 Módulos Switch LAN Ethernet redundantes, cada uno con 6 puertos Gigabit Ethernet. ✓ Soporte switches Ethernet de capa 4 a capa 7. ✓ 2 Módulos Switch SAN Fibra Canal, cada uno con 3 puertos externos de 4Gbps. ✓ Midplane redundante. ✓ Ventiladores y fuentes de poder y con capacidad de cambio en caliente. ✓ Cumpla con al menos dos circuitos de energía ✓ Unidad óptica DVD, debe ser local

	<p>accesible por cada blade sin necesidad de intervención física para que cada blade lo acceda</p> <ul style="list-style-type: none"> ✓ Módulo de Administración y KVM incluido, debe cumplir la administración via web sin necesidad de instalar software adicional ✓ Capacidad de virtualizar MAC Address de las tarjetas de red y WWN de las tarjetas de fibra (HBAs) con cualquiera sea la marca de switch interno que tenga el chasis
Servidor Blade S3	Descripción
Servido Blade HS21XM	IBM Blade Server HS21, ocupa 1 slot en el Chasis
<p>Componentes del Servidor Blade</p> 	<ul style="list-style-type: none"> ✓ 2 Procesadores QuadCore Intel Xeón de 3.0 Ghz, 1333 MHz FSB, 12MB L2 Caché. ✓ Memoria RAM de 32GB. ✓ 1 Disco de 146GB y 10K rpm. ✓ 2 Puertos Gigabit Ethernet internos. ✓ 1 Tarjeta interna Dual Port Fibra Canal de 4 Gbps.
Servidor Blade S5	Descripción
Servido Blade HS21	IBM Blade Server HS21, ocupa 1 slot en el Chasis
<p>Componentes del Servidor Blade</p> 	<ul style="list-style-type: none"> ✓ 2 Procesadores DualCore Intel Xeón de 3.0 Ghz, 1333 MHz FSB, 6MB L2 Caché. ✓ Memoria RAM de 16GB. ✓ Dos Discos de 146GB y 10K rpm. ✓ 2 Puertos Gigabit Ethernet internos.

Servidor Blade S7	Descripción
Servido Blade HS21	IBM Blade Server HS21, ocupa 1 slot en el Chasis
Componentes del Servidor Blade 	<ul style="list-style-type: none"> ✓ 1 Procesadores DualCore Intel Xeón de 3.3 Ghz, 1333 MHz FSB, 6MB L2 Caché. ✓ Memoria RAM de 16GB. ✓ Dos Discos de 146GB y 10K rpm en Raid 1. ✓ 2 Puertos Gigabit Ethernet internos.
Almacenamiento FC	Descripción
Storage DS 4700	1 Storage DS 4700 Modelo 70, con 1 TB de disco con capacidad de crecimiento hasta 84 TB.
Componentes del Storage 	<ul style="list-style-type: none"> ✓ 2 Controladoras de Fibra de 4Gbps. 2 GB Memoria Cache, exclusivo para datos. ✓ Tecnología Fibra Canal de 4Gbps end to end. ✓ 4 puertos de 4Gbps hacia la SAN ✓ 4 puertos de 4Gbps hacia los discos. ✓ 8 Discos de 146GB de 15 K rpm de 4 Gbps. ✓ Soporte niveles de RAID: 0, 1, 5, 6 y 10 ✓ 2 Particiones Activadas, con crecimiento a 256 particiones. ✓ Storage Manager licenciado para la capacidad maxima a la que escala el sistema de almacenamiento (112 TB)
Servidor x3650 Backup	Descripción
Servidor x3650	Servidor de Rack x3650, de 2 Unidades de Rack,

<p>Componentes del servidor</p> 	<ul style="list-style-type: none"> ✓ 1 Procesadores Xeon Quad Core 5450 3.0GHz/1333MHz/12MB L2 Caché, crecimiento hasta 2 procesadores. ✓ 2 puertos Ethernet Gigabit. ✓ 8 GB de memoria, crecimiento hasta 48GB. ✓ 6 Discos de 146GB SAS de 15 K rpm. ✓ 1 Controladora RAID ✓ 1 tarjeta HBA SAS ✓ Ventiladores y fuentes de poder y con capacidad de cambio en caliente. ✓ Unidad óptica DVD/CD WR
<p>Unidad de Respaldo TS3200</p>	<p>Descripción</p>
<p>Librería TS3200</p>	<p>IBM System Storage TS3200 Tape Library</p>
<p>Componentes de la librería</p> 	<ul style="list-style-type: none"> ✓ 1 Drive LTO 4, con crecimiento hasta 2 Drives. Capacidad máxima de almacenamiento Nativo 38,4GB / comprimido 76,4GB. ✓ Soporte hasta 48 cartuchos de data. ✓ Velocidad de lectura y escritura de 120 MB/seg nativo y 240 MB/seg comprimido. ✓ 20 Cartuchos LTO 4 ✓ 1 Cartucho de limpieza
<p>RACK</p>	<p>Descripción</p>
<p>Rack Standar</p>	<p>NetBAY S2 42U Standard Rack Cabinet, Monitor Flat Panel, Teclado para Rack, Switch console KVM, Unidades de distribución de Poder, cables.</p>

Garantía	✓ 3 años con cobertura 7 x 24.
----------	--------------------------------

Tabla V.35. Características de los servidores para BD y Análisis.

Las herramientas software que se utilizan en estos servidores son las siguientes:

DESCRIPCIÓN	LICENCIA	
	TIPO	CANTIDAD
Microsoft Windows Server 2008 Enterprise x64	Propietario	2
Microsoft SQL Server 2005 Standard Edition	Propietario	1
Jdk-6u7-windows-i586-p	GNU (Libre)	2
Mondrian-Jpivot (PENTAHO Analysis)	GNU (Libre)	1
Apache-tomcat-5.5	GNU (Libre)	1

Tabla V.36. Herramientas software de los servidores para BD y Análisis

5.2.5 Construcción de los Cubos

El análisis OLAP de la solución es realizado con la herramienta Mondrian 3.0.4. Por lo tanto, los cubos que reflejan el diseño del modelo dimensional de los datamarts serán construidos de acuerdo a los requerimientos de Mondrian para tal fin.

Se deben crear dos archivos para utilizar un cubo en Mondrian 3.0.4, en las siguientes secciones se describen las estructuras de estos archivos. Para conocer más detalles de la creación de estos archivos, ver anexo “Navegando en el Datamart”.

Luego de haber establecido las estructuras de estos archivos, un usuario puede navegar por estas estructuras (dimensiones, atributos y medidas) de forma intuitiva e interactiva, utilizando únicamente el Mouse desde la interfaz gráfica ofrecida por Mondrian-Jpivot.

Mondrian maneja dimensiones, niveles, categorías y medidas. Estas determinan la ruta posible de drill down y drill up en la herramienta OLAP Mondrian. Los valores que toma cada nivel son conocidos como las categorías. Las Medidas representan los indicadores de gestión del negocio para analizar (datos numéricos).

5.2.5.1 Archivos JSP

El archivo JSP es utilizado para definir la ruta y una consulta inicial de un cubo Mondrian. Las siguientes líneas de código son las líneas esenciales para establecer esta ruta. Como ejemplo para el reporte que muestra los Centros de Acopio vigentes en el caso de PSND tenemos el siguiente archivo jsp:

```
<%@ page session="true" contentType="text/html; charset=ISO-8859-1" %>
<%@ taglib uri="http://www.tonbeller.com/jpivot" prefix="jpp" %>
<%@ taglib prefix="c" uri="http://java.sun.com/jstl/core" %>

<jpp:mondrianQuery id="query01"
  jdbcDriver='<%=getServletContext().getInitParameter("jdbcSQLServer2005Driver")%>'
  jdbcUrl='<%= getServletContext().getInitParameter("jdbcSQLServer2005Url")%>'
  catalogUri="/Mondrian/queries/RUB_PSND_1.xml">

  select NON EMPTY (
    [Measures].[NUMERO TANQUES],
    [Measures].[CAPACIDAD LITROS LECHE]) ON COLUMNS,
  NON EMPTY ((
    [PROVINCIA.PROVINCIAS].[PROVINCIAS],
    [CANTON.CANTONES].[CANTONES],
    [PARROQUIA.PARROQUIAS].[PARROQUIAS],
    [CENTRO DE ACOPIO.NOMBRES].[NOMBRES],
    [TELEFONO.TELEFONOS].[TELEFONOS]) ON ROWS
  from [REPORTE1]

</jpp:mondrianQuery>

<c:set var="title01" scope="session">
  CENTROS DE ACOPIO VIGENTES EN GENERAL Y POR PROVINCIA
</c:set>
```

Figura V.59. Líneas de código del archivo mqCantidadCentrosdeAcopio.jsp

Donde las líneas señaladas se definen en un archivo web.xml como:

```
<context-param>
  <param-name>jdbcSQLServer2005Driver</param-name>
  <param-value>com.microsoft.sqlserver.jdbc.SQLServerDriver</param-value>
  <description>Driver de conexion con la base de datos de SQL Server 2005</description>
</context-param>

<context-param>
  <param-name>jdbcSQLServer2005Url</param-name>
  <param-value>jdbc:sqlserver://localhost:1433;databaseName=RUB_NUTRIENDO;
  user=rubecuador;password=rubecuador;</param-value>
  <description>URL de conexion con la base de datos del SQL Server 2005</description>
</context-param>
```

Figura V.60. Driver de conexión a la Base de Datos y su Url

Por otro lado es necesario indicar que para resolver el tema de los parámetros es decir en algunos requerimientos de reportes que los usuarios han pedido para poder filtrar sus resultados, también se lo hace utilizando archivos jsp y cada vez que para un programa social de MCDS que requiere parámetros se implementa un archivo denominado Datos.jsp, el mismo que es el encargado de filtrar la información de acuerdo a lo que el requerimiento demande, en algunos casos un rango de fechas como es en los reportes de PSNM y PSSA o en otros por periodos como son los casos de PSND y PSSS.

A continuación se muestra el código necesario para enviar por parámetros periodos a la aplicación de Nutriendo el Desarrollo:

```
<%@page contentType="text/html" pageEncoding="ISO-8859-1"%>

<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"

"http://www.w3.org/TR/html4/loose.dtd">

<%@ page import="java.text.SimpleDateFormat" %>

<html>

<head>

<title>Programa Social Nutriendo el Desarrollo - Datos del reporte</title>

<link rel="stylesheet" type="text/css" href="css/Plantilla.css">

<link rel="stylesheet" type="text/css" href="css/ColorLetra.css">

<link rel="stylesheet" type="text/css" href="css/Menu.css">

<link rel="stylesheet" type="text/css" href="css/TablaDatos.css">

<script language="javascript" src="js/Strings.js"> </script>

</head>

<jsp:useBean id="ubDimension4" scope="request" class="Acceso_Nutriendo.Dimension4" />

<jsp:useBean id="ubDimension5" scope="request" class="Acceso_Nutriendo.Dimension5" />

<jsp:useBean id="ubDimension7" scope="request" class="Acceso_Nutriendo.Dimension7" />

<body topmargin="0" leftmargin="0" marginwidth="0" marginheight="0" bgcolor="#F2F8F9">
```

```

<%

try {

%>

<form action="Procesar_2.jsp" method="post" onsubmit="return
Valida_Combo('cbPeriodo',0,'idlblErrorPeriodo','* Seleccione período','true');">

<input type="hidden" id="PaginaFuente" name="PaginaFuente" value="Datos.jsp" />

<input type="hidden" id="query" name="query" value="<%= request.getParameter("query")%>"
/>

<input type="hidden" id="query1" name="query1" value="<%=
request.getParameter("query")%>" />

<input type="hidden" id="Opcion" name="Opcion" value="<%=
request.getParameter("Opcion")%>" />

<%

SimpleDateFormat jtsdfFecha = new SimpleDateFormat("dd/MM/yyyy");

%>

<table width="1003px" align="center" cellpadding="0" cellspacing="0">

<tr>

<td align="center"><div id="cabecera"></div></td>

</tr>

<tr><td>&nbsp;</td></tr>

<tr>

<td align="CENTER" valign="top">

<table width="50%">

<%

if (request.getParameter("ColorMensaje") != null) {

```

```
%>

<tr>

<td>

<fieldset class="ColorMenu">

<legend class="LetraFecha">Resultados</legend>

<table align="center">

<tr align="justify">

<%

switch (Integer.parseInt(request.getParameter("ColorMensaje"))) {

case 0:

%>

<td class="LetraCorrectoA"><%= request.getParameter("Mensaje")%> </td>

<%

break;

case 1:

%>

<td class="LetraIncorrectoA"><%= request.getParameter("Mensaje")%> </td>

<%

break;

default:

break;

}

%>
```

```
</tr>

</table>

</fieldset>

</td>

</tr>

<%

}

%>

<tr>

<td align="center">

<fieldset class="ColorMenu">

<legend align="center" class="LetraFecha">Descripción del Reporte</legend>

<%

if (request.getParameter("Opcion") != null && !request.getParameter("Opcion").equals("null")) {

switch (Integer.parseInt(request.getParameter("Opcion"))) {

case 4:

out.print("Beneficiarios por Ubicaci&oacute;n");

break;

case 5:

out.print("Beneficiarios - Datos Personales");

break;

case 7:

out.print("Producción Diaria Leche");
```

```

break;

default:

break;

}

} else {

out.println("<span class='LetraIncorrectoA'>No definida</span>");

}

%>

</fieldset>

<br>

</td>

</tr>

<tr>

<td>

<table align="CENTER" width="100%">

<tr>

<td width="100%" align="left">

<fieldset class="ColorMenu">

<legend align="center" class="LetraFecha">Ingreso de datos</legend>

<table width="100%" align="left">

<%

java.sql.ResultSet rssPeriodos = null;

if (request.getParameter("Opcion") != null && !request.getParameter("Opcion").equals("null")) {

```

```

%>

<tr class="LetraCampos">

<td align="right"><span class="LetraRequerido">* </span>Período: </td>

<td width="80%">

<div>

<select id="cbPeriodo" name="cbPeriodo">

<%

switch (Integer.parseInt(request.getParameter("Opcion"))) {

case 4:

rssPeriodos = ubDimension4.ListarPeriodosFechasDesc();

if (rssPeriodos != null) {

out.print("Descuentos dados a los usuarios en sus compras");

while (rssPeriodos.next()) {

%>

<option value="<%=rssPeriodos.getString("PERCODIGO")%>">

<%=rssPeriodos.getString("PERDESCRIPCION") + " (" +
jtsdfFecha.format(rssPeriodos.getDate("PERFECHAINICIO")) + " - " +
jtsdfFecha.format(rssPeriodos.getDate("PERFECHAFINAL")) + ")"%>

</option>

<%

}

rssPeriodos.close();

}

break;

```

case 5:

```
rssPeriodos = ubDimension5.ListarPeriodosFechasDesc();
```

```
if (rssPeriodos != null) {
```

```
    while (rssPeriodos.next()) {
```

```
%>
```

```
<option value="<%=rssPeriodos.getString("PERCODIGO")%>">
```

```
<%=rssPeriodos.getString("PERDESCRIPCION") + " (" +  
jtsdfFecha.format(rssPeriodos.getDate("PERFECHAINICIO")) + " - " +  
jtsdfFecha.format(rssPeriodos.getDate("PERFECHAFINAL")) + ")"%>
```

```
</option>
```

```
<%
```

```
    }
```

```
    rssPeriodos.close();
```

```
}
```

```
break;
```

case 7:

```
rssPeriodos = ubDimension7.ListarPeriodosFechasDesc();
```

```
if (rssPeriodos != null) {
```

```
    while (rssPeriodos.next()) {
```

```
%>
```

```
<option value="<%=rssPeriodos.getString("PERCODIGO")%>">
```

```
<%=rssPeriodos.getString("PERDESCRIPCION") + " (" +  
jtsdfFecha.format(rssPeriodos.getDate("PERFECHAINICIO")) + " - " +  
jtsdfFecha.format(rssPeriodos.getDate("PERFECHAFINAL")) + ")"%>
```

```
</option>
```

```
<%  
  
}  
  
rssPeriodos.close();  
  
}  
  
break;  
  
default:  
  
break;  
  
}  
  
%>  
  
</select>  
  
<label id="idlblErrorPeriodo" class="LetraIncorrectoA"></label>  
  
</div>  
  
</td>  
  
</tr>  
  
<tr>  
  
<td colspan="2" align="center">  
  
<input type="reset" value="Reestablecer"/>  
  
<input type="submit" value="Aceptar"/>  
  
</td>  
  
</tr>  
  
<%  
  
} else {  
  
%>
```

<tr>

<td class="LetraIncorrectoA" align="justify">

La página no ha recibido los datos necesarios para la petición de datos los mismos que servirán para la generación del reporte

Por favor seleccione nuevamente el enlace que le llevo e este sitio para que pueda observar nuevamente el reporte deseado.

Si el problema persiste por favor consulte con el administrador del sistema.

</td>

</tr>

<% }>

%>

</table>

</fieldset>

</td>

</tr>

</table>

</td>

</tr>

</table>

</td>

</tr>

<tr><td> </td></tr>

<tr>

<td colspan="2">

```

<%@ include file="Pie.jsp" %>

</td>

</tr>

</table>

</form>

<%

} catch (Exception exc) {

out.println("<span class='LetraIncorrecta'>Ha ocurrido un error inesperado en el ingreso de
datos para la generación del reporte <br/><br/><br/>Por favor consulte con el administrador del
sistema. PROASETEL.</span>");

}

%>

</body>

</html>

```

El código que se mostró arriba representa a la programación en un archivo Datos.jsp para poder hacer uso de los parámetros antes de invocar un reporte de Mondrian, es decir para filtrar los resultados en lenguaje MDX

Este archivo Datos.jsp es el que enviará el o los parámetros que en él se hayan programado a la consulta MDX sobre el cubo OLAP del respectivo archivo XML que contiene el esquema de cubos.

5.2.5.2 Estructuras XML

Mondrian, para entenderse con los datamarts desarrollados por medio de vistas en SQL Server 2005, diseñada con el modelo dimensional, usa un archivo XML.

Este archivo es en el cual se describen las dimensiones, medidas y cubos que se usan en el proceso de análisis multidimensional de los datamarts de cada uno de los programas del MCDS.

En estas descripciones se asocian los nombres de los campos que se utilizaron en los datamarts de las bases de datos operacionales.

El archivo XML que define las dimensiones y las medidas usadas en el cubo que soporta el datamart por ejemplo para el reporte que muestra los Centros de Acopio vigentes en el caso del Programa Social Nutriendo el Desarrollo PSND es como sigue en la figura 54 de la siguiente página:

```
<Schema name="New Schema1">
  <Cube name="REPORTE1" cache="false" enabled="true">
    <Table name="FACT_1" schema="dbo">
    </Table>
    <Dimension type="StandardDimension" foreignKey="Fact_PK" name="CODIGO PERIODO">
      <Hierarchy>
    </Dimension>
    <Dimension type="StandardDimension" foreignKey="Fact_PK" name="PERIODO">
      <Hierarchy>
    </Dimension>
    <Dimension type="StandardDimension" foreignKey="Fact_PK" name="FECHA INICIO">
      <Hierarchy>
    </Dimension>
    <Dimension type="StandardDimension" foreignKey="Fact_PK" name="FECHA FINAL">
      <Hierarchy>
    </Dimension>
    <Dimension type="StandardDimension" foreignKey="Fact_PK" name="PROVINCIA">
      <Hierarchy>
    </Dimension>
    <Dimension type="StandardDimension" foreignKey="Fact_PK" name="CANTON">
      <Hierarchy>
    </Dimension>
    <Dimension type="StandardDimension" foreignKey="Fact_PK" name="PARROQUIA">
      <Hierarchy>
    </Dimension>
    <Dimension type="StandardDimension" foreignKey="Fact_PK" name="CENTRO DE ACOPIO">
      <Hierarchy>
    </Dimension>
    <Dimension type="StandardDimension" foreignKey="Fact_PK" name="TELEFONO">
      <Hierarchy>
    </Dimension>
    <Measure name="NUMERO TANQUES" column="CENANUHTANQUES" formatString="##0.##" aggregator="sum">
    </Measure>
    <Measure name="CAPACIDAD LITROS LECHE" column="CENACAPLITROSLECHE" formatString="##0.##" aggregator="sum">
    </Measure>
  </Cube>
</Schema>
```

Figura V.61. Líneas de código del archivo RUB_PSND_1.xml

Este archivo XML debe ser direccionado por todos los archivos .jsp que contengan una consulta mdx pero que necesite la información de este CUBO OLAP, de manera que se realiza

un reporte business intelligence sobre un cubo OLAP desarrollado en Mondrian cuyo esquema multidimensional es un archivo XML.

Ahora bien para esta solución hemos implementado varios cubos OLAP para explotarlos por medio de la realización de archivos .jsp que contienen en sus líneas de código lenguaje MDX, que es el lenguaje que este servidor OLAP entiende como se explicó en el capítulo teórico.

Los esquemas XML que se han desarrollados uno por cada requerimiento de reportes citados al inicio de este capítulo es decir seis cubos OLAP para PSND, uno para PSSA, uno para PSSS y ocho cubos OLAP para el Programa Sistema Nacional de Microfinanzas PSNM.

1. Estructura de los cubos

Para aclarar la implementación de cada cubo OLAP en primer lugar luego de tener listos nuestros datamarts en SQL Server 2005, procedemos a desarrollar el esquema XML del cubo como vimos por ejemplo en la figura 54.

Posteriormente a esto para reportear ya en un entorno Web como la interfaz que ofrece Jpivot, debemos desarrollar un archivo .jsp que va a tener algunas líneas de código esenciales entre ellas el driver de conexión al DBMS que estemos utilizando en nuestro caso SQL Server 2005 en donde residen nuestros datamarts como ya lo vimos en las figuras 52 y 53, además el Url de conexión con el mismo DBMS en donde especificamos el nombre de la base de datos, el usuario y contraseña definidos, y también otra de las líneas fundamentales es el direccionamiento del archivo XML que contiene el esquema del cubo OLAP al que deseamos explotar.

Y por último para reportear la información en la interfaz web Jpivot de Mondrian, procedemos a escribir las líneas de código referentes al lenguaje MDX que sirve para navegar por el cubo ya invocado como vimos anteriormente en la figura 52, todo ello en el mismo archivo jsp que estuvimos hablando en el párrafo anterior.

Por último se detalla un ejemplo de un cubo OLAP desarrollado en Mondrian, con sus dimensiones y medidas:

```

<Schema name="New Schema1">
  <Cube name="REPORTE1" cache="false" enabled="true">

    <Table name="FACT_1" schema="dbo">
    </Table>
    <Dimension type="StandardDimension" foreignKey="Fact_PK" name="PROVINCIA">
      <Hierarchy name="PROVINCIAS" hasAll="true" allMemberName="PROVINCIAS"
        primaryKey="Fact_PK" primaryKeyTable="DIMENSION_1">
        <Table name="DIMENSION_1" schema="dbo">
        </Table>
        <Level name="PROVINCIA" table="DIMENSION_1" column="PRONOMBRE"
          type="String" uniqueMembers="true" levelType="Regular"
          hideMemberIf="Never">

        </Level>
      </Hierarchy>
    </Dimension>

    <Measure name="NUMERO TANQUES" column="CENANUMTANQUES" formatString="##0.###"
      aggregator="sum" visible="true">
    </Measure>

    <Measure name="CAPACIDAD LITROS LECHE" column="CENACAPLITROSLECHE"
      formatString="##0.###" aggregator="sum" visible="true">
    </Measure>

  </Cube>
</Schema>

```

Figura V.62. Ejemplo de desarrollo de un esquema de cubo OLAP con una dimensión: Provincia y dos medidas: Numero Tanques y Capacidad Litros Leche.

Donde en primer lugar se define el esquema del cubo, teniendo en cuenta que un esquema puede tener uno o varios cubos y de igual manera el cubo tiene una o varias dimensiones las cuales tienen una o varias jerarquías con sus niveles, y sin olvidarnos por cierto de las medidas y también las medidas calculadas del cubo.

En la figura V.55 se puede apreciar claramente la sintaxis para el desarrollo de esquemas de cubos multidimensionales en el servidor OLAP Mondrian.

5.2.6 Reportes Implementados

Los reportes estadísticos business intelligence implementados cumplen con los requerimientos planteados en la sección 4.2.2.1. Su implementación está hecha sobre Mondrian el servidor OLAP de Pentaho, y con el uso de su interfaz gráfica es posible ver estos reportes y analizar la información con tablas de datos o gráficas de diversos tipos (barras, pies, líneas, etc).

Para este análisis de información se dan diferentes posibilidades de navegación, como el drill up y drill down. Además es posible analizar solo algunas dimensiones, o incluir a varias para un reporte. También es posible restringir en un análisis a determinados miembros de una dimensión que sean del interés del análisis, en vez de incluir a todos. Para conocer más detalles de la forma de utilizar el análisis OLAP Mondrian ver anexo “Navegando en el Datamart”.

En las siguientes secciones se documentan estos reportes en forma gráfica, aunque en la herramienta también se muestra la tabla de resultados de los datos. Se incluye también el query MDX necesario para realizar los reportes. Es importante notar que no es necesario escribir un query MDX para navegar por el cubo OLAP.

Esta navegación puede ser hecha de forma sencilla e interactiva con el usuario usando solo el mouse, seleccionando las opciones que se quieren (ver anexo Navegando en el Datamart). Además de los reportes acordados con MCDS es posible realizar muchos más, haciendo uso de las características de Mondrian y de las dimensiones implementadas. Se mostrará un solo reporte por página para dar claridad al documento.

A continuación se muestra una lista que indica el orden de los reportes para los respectivos sistemas programas sociales del MCDS:

1. Programa Social Nutriendo el Desarrollo **PSND**
2. Programa Social Socio Ahorro **PSSA**
3. Programa Social Socio Solidario **PSSS**
4. Programa Sistema Nacional de Microfinanzas **PSNM**

1. Programa Social Nutriendo el Desarrollo PSND

a. **Reporte:** Ver Centros de Acopio vigentes por ubicación.

La consulta MDX que soporta este reporte es la siguiente:

```
select NON EMPTY {  
    [Measures].[NUMERO TANQUES],  
    [Measures].[CAPACIDAD LITROS LECHE]} ON COLUMNS,  
NON EMPTY { (  
    [PROVINCIA.PROVINCIAS].[PROVINCIAS],  
    [CANTON.CANTONES].[CANTONES],  
    [PARROQUIA.PARROQUIAS].[PARROQUIAS],  
    [CENTRO DE ACOPIO.NOMBRES].[NOMBRES],  
    [TELEFONO.TELEFONOS].[TELEFONOS] ) } ON ROWS  
from [REPORTE1]
```

Figura V.63. Archivo de consulta MDX: mqCantidadCentrosdeAcopio.jsp

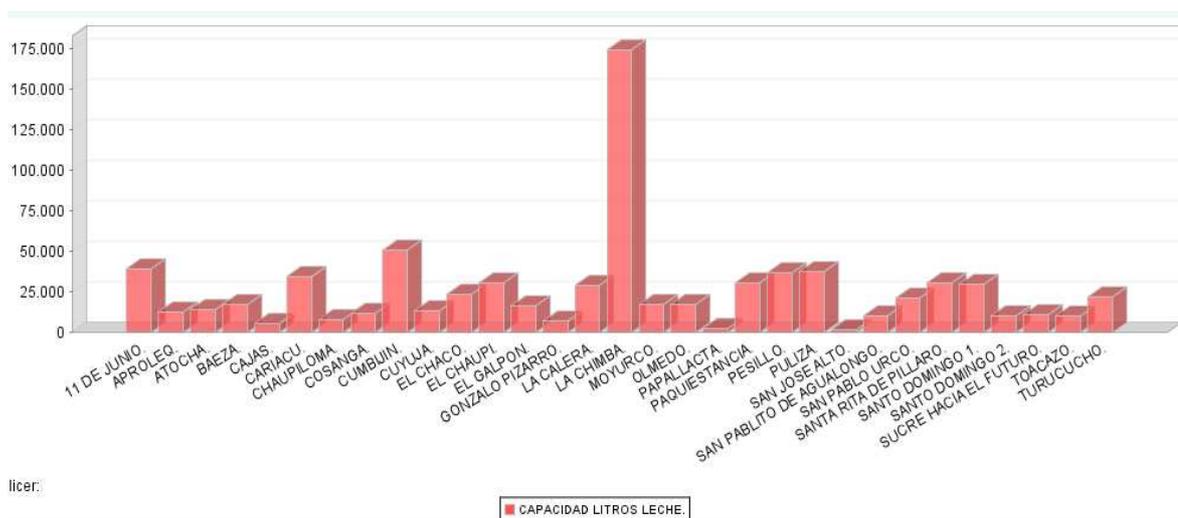


Figura V.64. Resultados gráficos del reporte: Centros de Acopio por ubicación

Esta consulta representa a los Centros de Acopio vigentes del PSND, con las medidas referentes al número de tanques que tiene cada uno de los centros registrados y además la capacidad de litros de leche que dicho centro tiene.

b. Reporte: Ver los periodos del programa.

La consulta MDX que soporta este reporte es la siguiente:

```
select NON EMPTY {[Measures].[PERIODOS]} ON COLUMNS,  
NON EMPTY {(  
[PERIODO.NOMBRES].[NOMBRES],  
[FECHA INICIO PERIODO.FECHAS].[FECHAS],  
[FECHA FINAL PERIODO.FECHAS].[FECHAS],  
[VIGENTE.VIGENTES].[VIGENTES])} ON ROWS  
from [REPORTE2]
```

Figura V.65. Archivo de consulta MDX: mqCantidadPeriodos.jsp

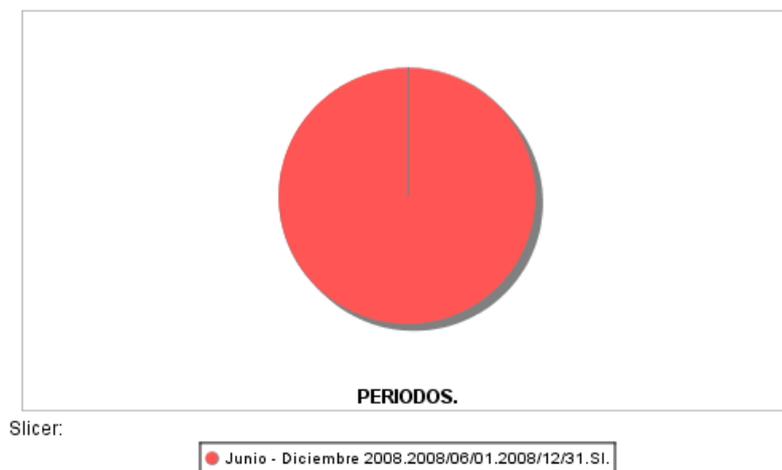


Figura V.66. Resultados gráficos del reporte: Periodos del Programa

Esta consulta como su nombre lo indica muestra los periodos del programa, actualmente en la base de datos RUB_NUTRIENDO se encuentra un solo periodo registrado en el programa, razón por la cual el gráfico cuya medida del cubo es el número de periodos y las dimensiones del cubo son el nombre del periodo, las fechas de inicio y fin del periodo, y además la vigencia del mismo, en este caso me dice el reporte que el periodo si está vigente.

Para realizar el gráfico hemos escogido en tiempo de ejecución un tipo de gráfico "Pastel por Columna".

c. **Reporte:** Ver los precios de venta vigentes del litro de leche por ubicación y Centros de Acopio.

La consulta MDX que soporta este reporte es la siguiente:

```
select NON EMPTY (
    [Measures].[PROM LITRO LECHE] ON COLUMNS,
NON EMPTY ({
    [PROVINCIA.PROVINCIAS].[PROVINCIAS],
    [CANTON.CANTONES].[CANTONES],
    [PARROQUIA.PARROQUIAS].[PARROQUIAS],
    [CENTRO DE ACOPIO.NOMBRES].[NOMBRES]}) ON ROWS
from [REPORTE3]
```

Figura V.67. Archivo de consulta MDX: mqPreciosVentaLeche.jsp

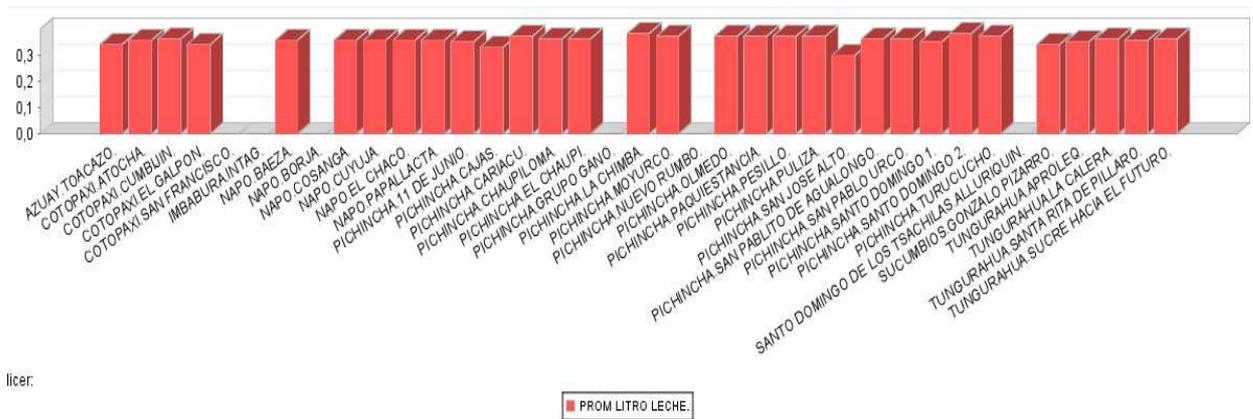


Figura V.68. Resultados gráficos del reporte: Precios vigentes de litro de leche

Esta consulta presenta en pantalla los precios vigentes de litro de leche por provincia, cantón, parroquia y centro de acopio, donde cuanto el reporte no está desplegado en su mayor nivel de detalle se va mostrando el promedio del precio del litro de leche hasta que cuando el usuario final despliega todas las medidas se abrirá también ya no un promedio sino el precio actual del litro de leche de ese centro de acopio.

Para realizar el gráfico hemos escogido en tiempo de ejecución un tipo de gráfico “Barra Vertical 3D Apilada”.

d. Reporte: Ver beneficiarios con sus datos personales por ubicación y Centros de Acopio en un período.

La consulta MDX que soporta este reporte es la siguiente:

```
select NON EMPTY (
    [Measures].[CANTIDAD] ON COLUMNS,
NON EMPTY ( (
    [PERIODO.NOMBRES].[NOMBRES],
    [PROVINCIA.PROVINCIAS].[PROVINCIAS],
    [CANTON.CANTONES].[CANTONES],
    [PARROQUIA.PARROQUIAS].[PARROQUIAS],
    [CENTRO DE ACOPIO.NOMBRES].[NOMBRES],
    [CEDULA.CEDULAS].[CEDULAS],
    [BENEFICIARIO.NOMBRES].[NOMBRES] ) ON ROWS
from [REPORTE5]
where [CODIGO PERIODO.CODIGOS].[CODIGOS].[<%=request.getParameter("cbPeriodo") %>]
```

Figura V.69. Archivo de consulta MDX: mqCantidadBenCentrosProvinciasDatos.jsp

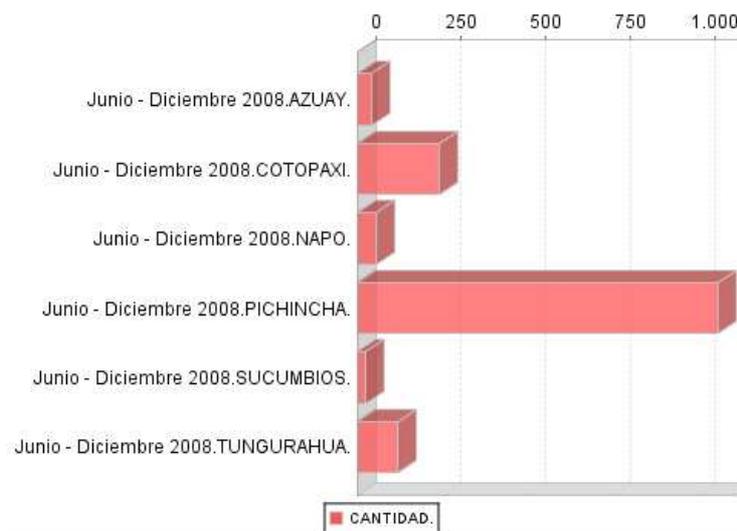


Figura V.70. Resultados gráficos del reporte: Beneficiarios Datos Personales

Este reporte estadístico muestra de acuerdo al periodo escogido por el usuario final, un listado con el número de beneficiarios que existe en RUB_NUTRIENDO cuya medida se llama "CANTIDAD", por provincia, cantón, parroquia, centro de acopio, y el nivel de detalle máximo es el nombre del beneficiario.

e. **Reporte:** Ver la producción diaria de leche por ubicación y Centros de Acopio en un período.

La consulta MDX que soporta este reporte es la siguiente:

```
select NON EMPTY {  
    [Measures].[PRODUCCION DIARIA],  
    [Measures].[PRODUCCION TANQUE],  
    [Measures].[PRECIO LITRO],  
    [Measures].[INGRESO PROM]} ON COLUMNS,  
NON EMPTY {(  
    [PERIODO.NOMBRES].[NOMBRES],  
    [PROVINCIA.PROVINCIAS].[PROVINCIAS],  
    [CANTON.CANTONES].[CANTONES],  
    [PARROQUIA.PARROQUIAS].[PARROQUIAS],  
    [CENTRO DE ACOPIO.NOMBRES].[NOMBRES],  
    [CEDULA.CEDULAS].[CEDULAS],  
    [BENEFICIARIO.NOMBRES].[NOMBRES])} ON ROWS  
from [REPORTE7]  
where [CODIGO PERIODO.CODIGOS].[CODIGOS].[<%=request.getParameter("cbPeriodo") %>]
```

Figura V.71. Archivo de consulta MDX:
mqCantidadBenCentrosProvinciasLitrosLecheProduc.jsp

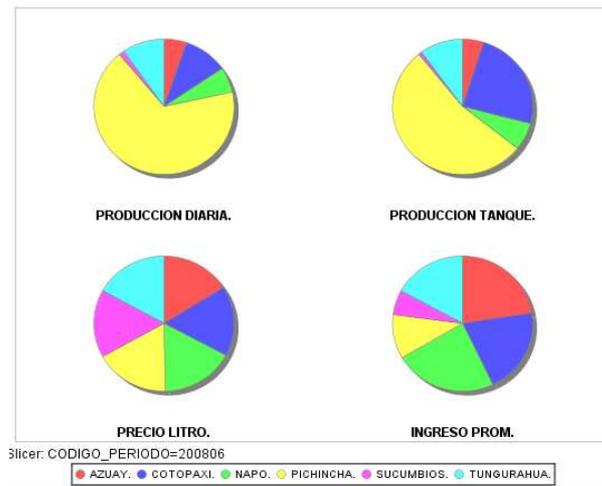


Figura V.72. Resultados gráficos del reporte: Producción diaria de leche

Esta consulta muestra la producción diaria de leche, de cada uno de los productores de leche que han sido beneficiarios del PSND. Para dar una mejor vista en el gráfico se ha escogido solo por provincias, con todas las medidas que el reporte contiene.

Para realizar la forma del gráfico hemos escogido en tiempo de ejecución un tipo de gráfico "Pastel por Columna".

f. **Reporte:** Ver beneficiarios por la capacitación recibida.

La consulta MDX que soporta este reporte es la siguiente:

```
select NON EMPTY {  
    [Measures].[NUM_CURSO],  
    [Measures].[TOTAL SUBSIDIADO]} ON COLUMNS,  
NON EMPTY { (  
    [PROVINCIA.PROVINCIAS].[PROVINCIAS],  
    [CANTON.CANTONES].[CANTONES],  
    [LUGAR.LUGARES].[LUGARES],  
    [FECHA_INICIO.FECHAS].[FECHAS],  
    [FECHA_FIN.FECHAS].[FECHAS],  
    [CEDULA.CEDULAS].[CEDULAS],  
    [BENEFICIARIO.NOMBRES].[NOMBRES] ) } ON ROWS  
from [REPORTE9]
```

Figura V.73. Archivo de consulta MDX: mqCantidadBenCentrosProvinciasCapacitacion.jsp

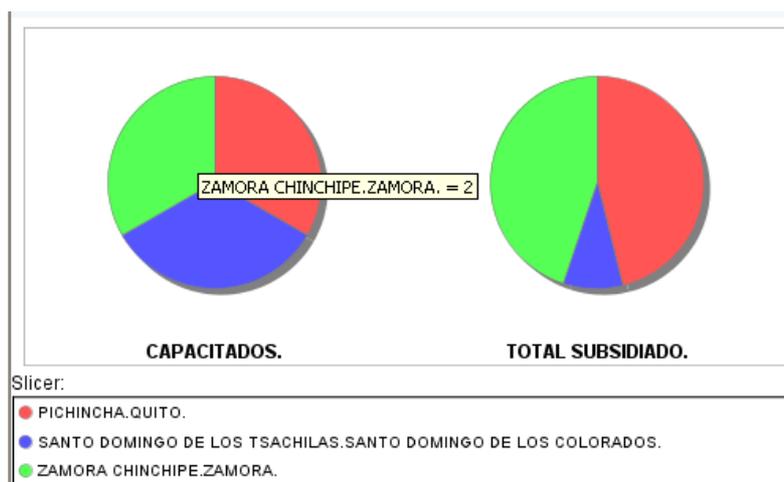


Figura V.74. Resultados gráficos del reporte: Beneficiarios por la capacitación recibida

Este reporte muestra los beneficiarios del PSND por la capacitación recibida, razón por la cual en el gráfico se escogerá por provincia y cantón donde vemos que en las tres provincias hay el mismo número de capacitados =2 por ejemplo.

Para la forma del gráfico escogimos en tiempo de ejecución un "Pastel por Columna".

2. Programa Social Socio Ahorro PSSA

- a. **Reporte:** Ver descuentos dados a los usuarios en sus compras por ubicación en un rango de fechas.

La consulta MDX que soporta este reporte es la siguiente:

```
select NON EMPTY {
    [Measures].[DESCUENTO COMPRAS]} ON COLUMNS,
NON EMPTY
    Filter(
        Crossjoin([PROVINCIA.PROVINCIAS].[PROVINCIAS].Children,
        Crossjoin([CANTON.CANTONES].[CANTONES].Children,
        Crossjoin([PARROQUIA.PARROQUIAS].[PARROQUIAS].Children,
        Crossjoin([SUCURSAL.SUCURSALES].[SUCURSALES].Children,
        Crossjoin({ [CEDULA.CEDULAS].[CEDULAS]},
        Crossjoin({ [BENEFICIARIO.BENEFICIARIOS].[BENEFICIARIOS]},
        [FECHA_TRANSACCION.FECHAS].[FECHAS].Children))))),
        ({<%= strFechaDesde %> <= [Measures].[FECHAINT]}
        AND
        ([Measures].[FECHAINT] <= <%= strFechaHasta %>})) ON ROWS
from [REPORTE1]
```

Figura V.75. Archivo de consulta MDX: mqDescuentosCompras.jsp

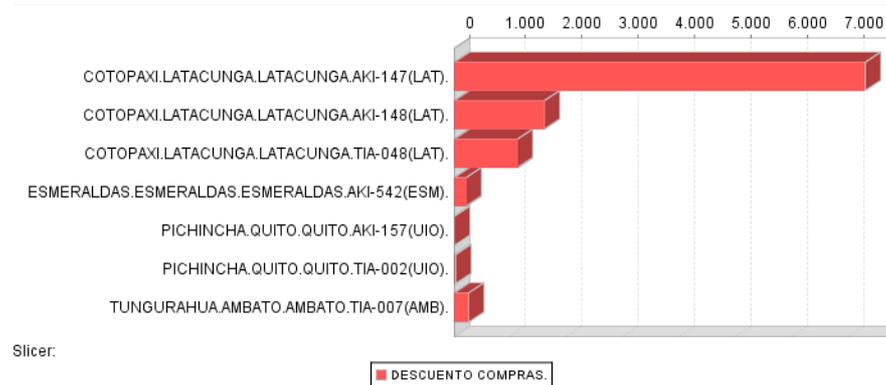


Figura V.76. Resultados gráficos del reporte: Descuentos a los usuarios en sus Compras

Este reporte muestra los descuentos que cada usuario ha recibido en una sucursal por concepto de una transacción realizada, para ello existen varias dimensiones como vemos en la consulta MDX de la figura 68, y una sola medida la cual se llama DESCUENTO COMPRAS.

Para realizar la forma del gráfico hemos escogido en tiempo de ejecución un tipo de gráfico "Barra Horizontal 3D Apilada".

3. Programa Social Socio Solidario PSSS

- a. **Reporte:** Ver estadísticas de los ingresos por la venta de productos Socio Solidario por ubicación en un período.

La consulta MDX que soporta este reporte es la siguiente:

```
select NON EMPTY (
  [Measures].[PRECIO VENTA PUBLICO],
  [Measures].[PORCENTAJE DESCUENTO MINIMO],
  [Measures].[UNIDADES VENDIDAS SUPER],
  [Measures].[VOLUMEN VENTAS SUPER],
  [Measures].[UNIDADES VENDIDAS TIENDAS],
  [Measures].[VOLUMEN VENTAS TIENDAS]) ON COLUMNS,
NON EMPTY
({ ([PERIODO.PERIODOS].[PERIODOS],
  [PROVINCIA.PROVINCIAS].[PROVINCIAS],
  [CANTON.CANTONES].[CANTONES],
  [PARROQUIA.PARROQUIAS].[PARROQUIAS],
  [EMPRESA.NOMBRES].[NOMBRES],
  [MARCA.MARCAS].[MARCAS],
  [CODIGO BARRA.CODIGOS].[CODIGOS],
  [PRODUCTO.PRODUCTOS].[PRODUCTOS],
  [PRESENTACION.PRESENTACIONES].[PRESENTACIONES],
  [VENTA SUPER.VENTAS].[VENTAS],
  [VENTA TIENDA.VENTAS].[VENTAS]}) ON ROWS
from [REPORTE1]
where [CODIGO PERIODO.CODIGOS].[CODIGOS].[<%=request.getParameter("cbPeriodo")%>]
```

Figura V.77. Archivo de consulta MDX: mqlIngresosVentasProductos.jsp

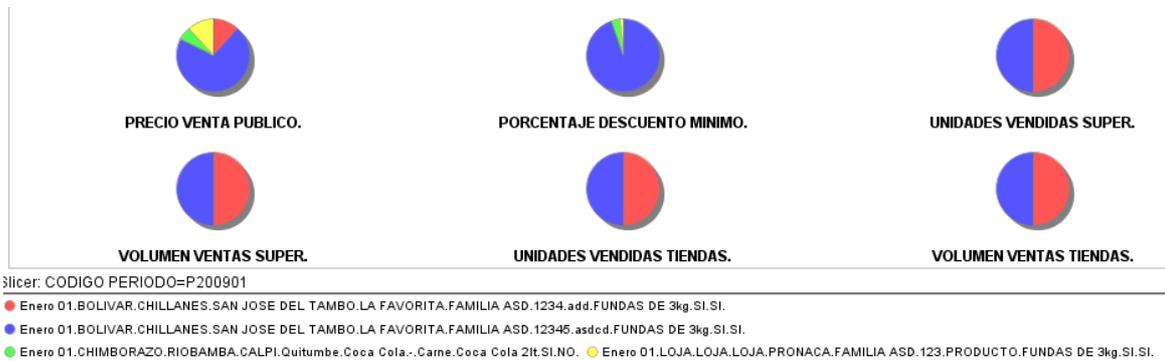


Figura V.78. Resultados gráficos del reporte: Ingresos por la venta de producto sociocolidarios

Este reporte muestra los ingresos de los beneficiarios del PSSS por la venta de productos sociocolidarios, donde como vemos en el gráfico están todas las dimensiones y medidas de la consulta MDX de la figura 70.

Para realizar la forma del gráfico hemos escogido en tiempo de ejecución un tipo de gráfico "Pastel por Columna".

4. Programa Sistema Nacional de Microfinanzas PSNM

a. **Reporte:** Ver beneficiarios por nivel de pobreza en un rango de fechas.

La consulta MDX que soporta este reporte es la siguiente:

```
select Non Empty{
  [Measures].[SOL-SOCIO],
  [Measures].[MN TOTAL],
  [Measures].[MN PROM],
  [Measures].[TASA MIN %],
  [Measures].[TASA MAX %],
  [Measures].[TASA PROM %]} on columns,
NON EMPTY
FILTER(
  Crossjoin([NIVEL DE POBREZA.NIVELES].[NIVELES].Children,
  Crossjoin([REGION.REGIONES].[REGIONES].Children,
  Crossjoin([PROVINCIA.PROVINCIAS].[PROVINCIAS].Children,
  Crossjoin([CANTON.CANTONES].[CANTONES].Children,
  Crossjoin([PARROQUIA.PARROQUIAS].[PARROQUIAS].Children,
  Crossjoin([CEDULA.CEDULAS].[CEDULAS].Children,
  [FECHA DE CREDITO.FECHAS].[FECHAS].Children))))),
  ({<strFechaDesde> <= [Measures].[FECHACREDITO1]
  AND
  ([Measures].[FECHACREDITO1] <= <strFechaHasta> })) ON rows
from REPORTE3
```

Figura V.79. Archivo de consulta MDX: mqCreditosNivelPobreza.jsp

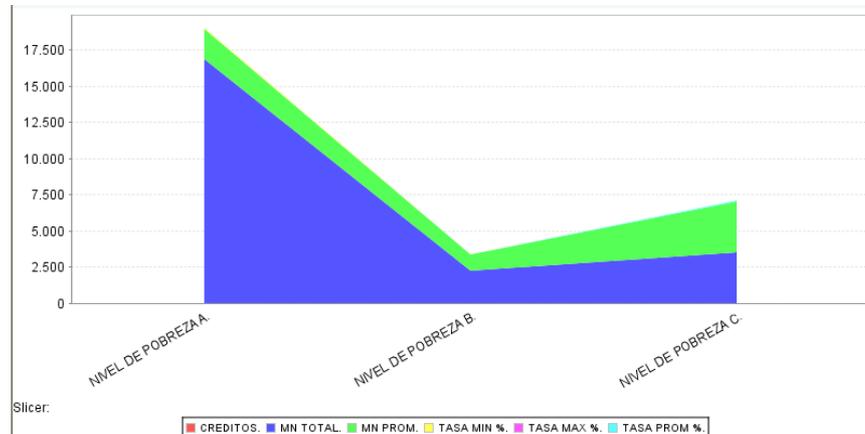


Figura V.80. Resultados gráficos del reporte: Beneficiarios por nivel de pobreza

Este reporte muestra los beneficiarios del PSNM por el nivel de pobreza, y otra información como la provincia, cantón y parroquia, tal como se muestra en la figura 72 de las dimensiones y medidas como los montos totales y promedio y las tasas de crédito que se les concedió. Para realizar la forma del gráfico hemos escogido en tiempo de ejecución un tipo de gráfico “Área Vertical Apilada”.

b. **Reporte:** Ver beneficiarios por género y ubicación en un rango de fechas.

La consulta MDX que soporta este reporte es la siguiente:

```

select NON EMPTY {
    [Measures].[SOL-SOCIO],
    [Measures].[MN TOTAL],
    [Measures].[MN PROM],
    [Measures].[TASA MIN %],
    [Measures].[TASA MAX %],
    [Measures].[TASA PROM %]} ON COLUMNS,
NON EMPTY
FILTER(
    Crossjoin([ESTADO CREDITO.CREDITOS].[CREDITOS].Children,
    Crossjoin([GENERO.GENEROS].[GENEROS].Children,
    Crossjoin([REGION.REGIONES].[REGIONES].Children,
    Crossjoin([PROVINCIA.PROVINCIAS].[PROVINCIAS].Children,
    Crossjoin([CANTON.CANTONES].[CANTONES].Children,
    Crossjoin([PARROQUIA.PARROQUIAS].[PARROQUIAS].Children,
    Crossjoin([CEDULA.CEDULAS].[CEDULAS].Children,
    [FECHA DE CREDITO.FECHAS].[FECHAS].Children)))))),
    ({<%= strFechaDesde %> <= [Measures].[FECHACREDITO1]
    AND
    ([Measures].[FECHACREDITO1] <= <%= strFechaHasta %>})) ON rows
from [REPORTE5A]

```

Figura V.81. Archivo de consulta MDX: mqCreditosUbicacion.jsp

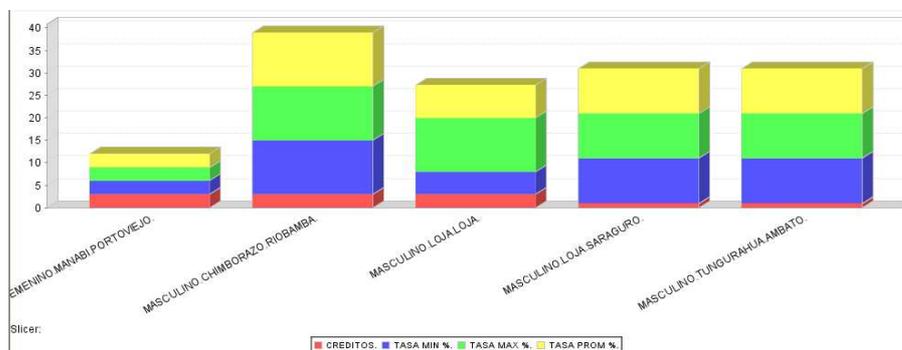


Figura V.82. Resultados gráficos del reporte: Beneficiarios por género y ubicación

Este reporte muestra los montos de crédito y tasas de interés de los beneficiarios de PSNM agrupados por ubicación y por género, para dar una mejor vista al gráfico del reporte se escogerá sólo las medidas relativas a tasas de créditos para que se vea la diferencia, ya que si el gráfico lo hacemos en un “Barra Vertical 3D Apilada” no se va a distinguir las tasas de interés puesto que los montos de crédito sobrepasan los miles de dólares mientras que las tasas de interés son solo porcentajes, ver la figura V.75.

c. **Reporte:** Ver beneficiarios por Entidad Microfinanciera en un rango de fechas.

La consulta MDX que soporta este reporte es la siguiente:

```

select Non Empty(
    [Measures].[SOL-SOCIO],
    [Measures].[MN TOTAL],
    [Measures].[MN PROM],
    [Measures].[TASA MIN %],
    [Measures].[TASA MAX %],
    [Measures].[TASA PROM %} on columns,
Non Empty
FILTER(
    Crossjoin([TIPO INSTITUCION.TIPOS].[TIPOS].Children,
    Crossjoin([NOMBRE ENTIDAD.NOMBRES].[NOMBRES].Children,
    Crossjoin([ESTADO CREDITO.CREDITOS].[CREDITOS].Children,
    Crossjoin([ESTADO CREDITO.CREDITOS].[CREDITOS].Children,
    Crossjoin([CEDULA.CEDULAS].[CEDULAS].Children,
    [FECHA DE CREDITO.FECHAS].[FECHAS].Children))))),
    (<%= strFechaDesde %> <= [Measures].[FECHACREDITO1]
    AND
    ([Measures].[FECHACREDITO1] <= <%= strFechaHasta %>)) ON rows
from REPORTE5B

```

Figura V.83. Archivo de consulta MDX: mqCreditosEntidadesMicrofinancieras.jsp

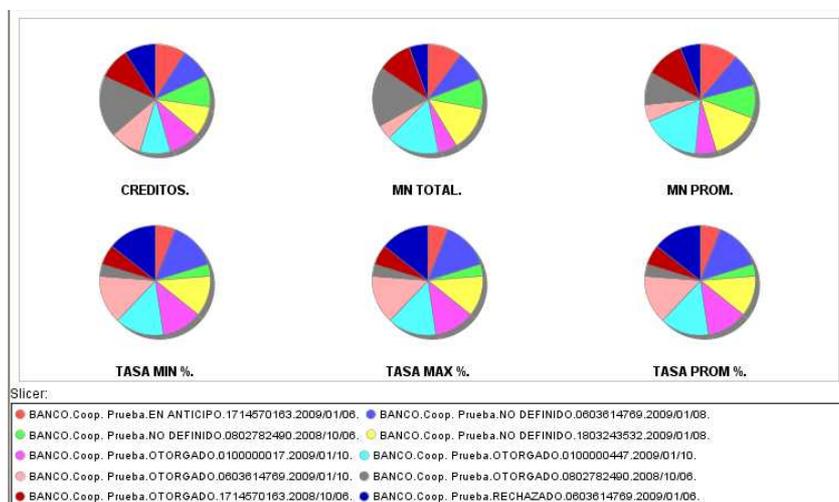


Figura V.84. Resultados gráficos del reporte: Beneficiarios por Entidad Microfinanciera

El reporte anterior las mismas medidas que todos en PSNM con la diferencia de que presenta sus resultados por Entidades Microfinancieras. La forma del gráfico es de tipo “Pastel por Columna” escogido mientras el reporte web está en ejecución.

d. **Reporte:** Ver beneficiarios por Actividad Microempresarial en un rango de fechas.

La consulta MDX que soporta este reporte es la siguiente:

```
select Non Empty{
  [Measures].[SOL-SOCIO],
  [Measures].[MN TOTAL],
  [Measures].[MN PROM],
  [Measures].[TASA MIN %],
  [Measures].[TASA MAX %],
  [Measures].[TASA PROM %]} on columns,
Non Empty
FILTER(
  Crossjoin([ACTIVIDAD.ACTIVIDADES].[ACTIVIDADES].Children,
  Crossjoin([ESTADO CREDITO.CREDITOS].[CREDITOS].Children,
  Crossjoin([REGION.REGIONES].[REGIONES].Children,
  Crossjoin([PROVINCIA.PROVINCIAS].[PROVINCIAS].Children,
  Crossjoin([CANTON.CANTONES].[CANTONES].Children,
  Crossjoin([PARROQUIA.PARROQUIAS].[PARROQUIAS].Children,
  Crossjoin([CEDULA.CEDULAS].[CEDULAS].Children,
  [FECHA DE CREDITO.FECHAS].[FECHAS].Children)))))),
  ((<%= strFechaDesde %> <= [Measures].[FECHACREDITO1])
  AND
  ([Measures].[FECHACREDITO1] <= <%= strFechaHasta %>))) ON rows
from REPORTE6A
```

Figura V.85. Archivo de consulta MDX: mqCreditosActividad.jsp

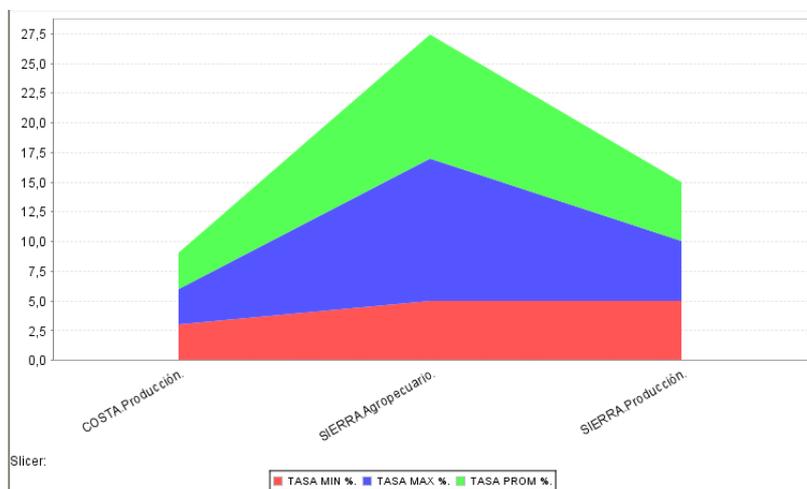


Figura V.86. Resultados gráficos del reporte: Beneficiarios por Actividad Microempresarial

Este reporte devuelve los resultados según una actividad Microempresarial. En el gráfico se escogió un tipo de “Área Vertical Apilada” y las dimensiones región, y actividad con las medidas tasas de crédito.

e. Reporte: Ver beneficiarios por Entidad Microfinanciera y Actividad Microempresarial en un rango de fechas.

La consulta MDX que soporta este reporte es la siguiente:

```

select Non Empty(
    [Measures].[SOL-SOCIO],
    [Measures].[MN TOTAL],
    [Measures].[MN PROM],
    [Measures].[TASA MIN %],
    [Measures].[TASA MAX %],
    [Measures].[TASA PROM %} on columns,
Non Empty
FILTER(
    Crossjoin([TIPO INSTITUCION.TIPOS].[TIPOS].Children,
    Crossjoin([NOMBRE ENTIDAD.NOMBRES].[NOMBRES].Children,
    Crossjoin([ACTIVIDAD.ACTIVIDADES].[ACTIVIDADES].Children,
    Crossjoin([ESTADO CREDITO.CREDITOS].[CREDITOS].Children,
    Crossjoin([CEDULA.CEDULAS].[CEDULAS].Children,
    [FECHA DE CREDITO.FECHAS].[FECHAS].Children))))),
    {(<%= strFechaDesde %> <= [Measures].[FECHACREDITO1]
    AND
    ([Measures].[FECHACREDITO1] <= <%= strFechaHasta %>)) ON rows
from REPORTE6B

```

Figura V.87. Archivo de consulta MDX: mqCreditosEntidadActividad.jsp

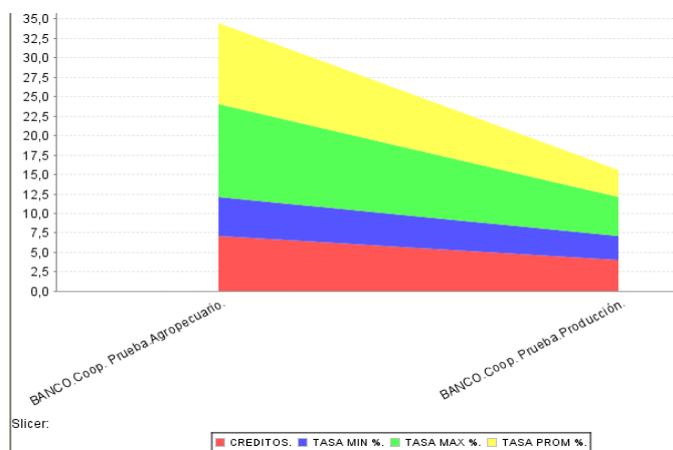


Figura V.88. Resultados gráficos del reporte: Beneficiarios por Entidad Microfinanciera y Actividad Microempresarial

Esta consulta muestra los montos de crédito y tasas de interés concedidos a los beneficiarios del PSNM por entidad Microfinanciera y actividad Microempresarial, aunque para graficar escogimos solo el número de créditos concedidos y las tasas de interés como medidas, y además la institución, la entidad Microfinanciera y la actividad Microempresarial como dimensiones.

f. **Reporte:** Ver beneficiarios por la forma de crédito en un rango de fechas.

La consulta MDX que soporta este reporte es la siguiente:

```

select Non Empty(
  [Measures].[SOL-SOCIO],
  [Measures].[MN TOTAL],
  [Measures].[MN PROM],
  [Measures].[TASA MIN %],
  [Measures].[TASA MAX %],
  [Measures].[TASA PROM %] } on columns,
Non Empty
  FILTER(
    Crossjoin([FORMA CREDITO.CREDITOS].[CREDITOS].Children,
      Crossjoin([ESTADO CREDITO.CREDITOS].[CREDITOS].Children,
        Crossjoin([REGION.REGIONES].[REGIONES].Children,
          Crossjoin([PROVINCIA.PROVINCIAS].[PROVINCIAS].Children,
            Crossjoin([CANTON.CANTONES].[CANTONES].Children,
              Crossjoin([PARROQUIA.PARROQUIAS].[PARROQUIAS].Children,
                Crossjoin([CEDULA.CEDULAS].[CEDULAS].Children,
                  [FECHA DE CREDITO.FECHAS].[FECHAS].Children)))))),
    {(<%= strFechaDesde %> <= [Measures].[FECHACREDITO1]
      AND
      ([Measures].[FECHACREDITO1] <= <%= strFechaHasta %>))} ON rows
from REPORTE7A

```

Figura V.89. Archivo de consulta MDX: mqCreditosForma.jsp

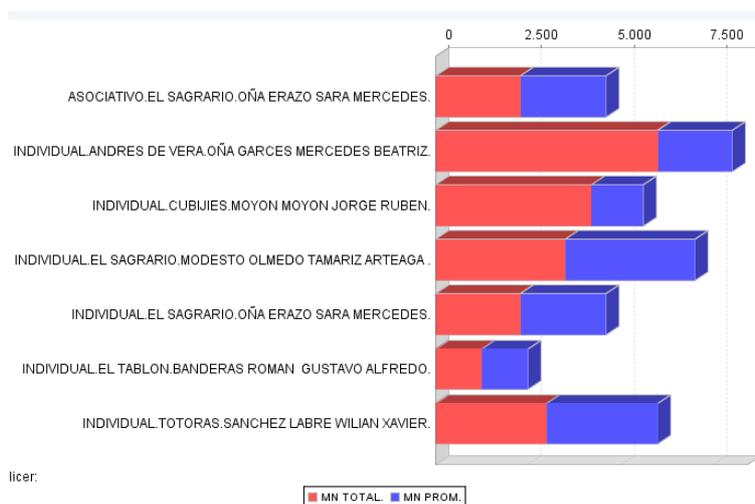


Figura V.90. Resultados gráficos del reporte: Beneficiarios por la forma de crédito

Este reporte muestra los resultados de acuerdo a la forma de crédito, donde escogimos el tipo “Barra Horizontal Apilada” para el gráfico, presentando los resultados estadísticos gráficos con las dimensiones de la forma de crédito, el nombre del beneficiario y el monto total y promedio concedido para el mismo.

- g. **Reporte:** Ver beneficiarios por forma de crédito y Entidad Microfinanciera en un rango de fechas.

La consulta MDX que soporta este reporte es la siguiente:

```

select Non Empty(
  [Measures].[SOL-SOCIO],
  [Measures].[MN TOTAL],
  [Measures].[MN PROM],
  [Measures].[TASA MIN %],
  [Measures].[TASA MAX %],
  [Measures].[TASA PROM %]) on columns,
Non Empty
FILTER(
  Crossjoin([TIPO INSTITUCION.TIPOS].[TIPOS].Children,
  Crossjoin([NOMBRE ENTIDAD.NOMBRES].[NOMBRES].Children,
  Crossjoin([FORMA CREDITO.CREDITOS].[CREDITOS].Children,
  Crossjoin([ESTADO CREDITO.CREDITOS].[CREDITOS].Children,
  Crossjoin([CEDULA.CEDULAS].[CEDULAS].Children,
  [FECHA DE CREDITO.FECHAS].[FECHAS].Children))))),
  ((<%= strFechaDesde %> <= [Measures].[FECHACREDITO1]
  AND
  ([Measures].[FECHACREDITO1] <= <%= strFechaHasta %>))) ON rows
from REPORTE7B

```

Figura V.91. Archivo de consulta MDX: mqCreditosFormaEntidad.jsp

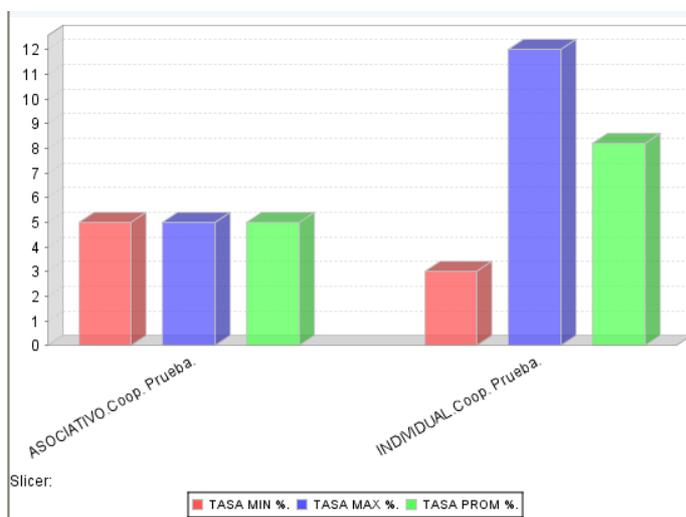


Figura V.92. Resultados gráficos del reporte: Beneficiarios por la forma de crédito y Entidad Microfinanciera

Este reporte muestra los créditos concedidos a los beneficiarios del PSNM por la forma de crédito y por entidad Microfinanciera, para el gráfico se ha escogido una “Barra Vertical 3D” donde se tomo en cuenta solo las tasas de interés.

h. Reporte: Ver los beneficiarios del programa en un rango de fechas.

La consulta MDX que soporta este reporte es la siguiente:

```

select Non Empty(
    [Measures].[SOL-SOCIO],
    [Measures].[MN TOTAL],
    [Measures].[MN PROM],
    [Measures].[TASA MIN %],
    [Measures].[TASA MAX %],
    [Measures].[TASA PROM %]) on columns,
Non Empty
FILTER(
    Crossjoin([REGION.REGIONES].[REGIONES].Children,
    Crossjoin([PROVINCIA.PROVINCIAS].[PROVINCIAS].Children,
    Crossjoin([CANTON.CANTONES].[CANTONES].Children,
    Crossjoin([PARROQUIA.PARROQUIAS].[PARROQUIAS].Children,
    Crossjoin([CEDULA.CEDULAS].[CEDULAS].Children,
    Crossjoin([TELEFONO.TELEFONOS].[TELEFONOS].Children,
    [FECHA DE CREDITO.FECHAS].[FECHAS].Children))))),
    ({<%= strFechaDesde %> <= [Measures].[FECHACREDITO1]
    AND
    ([Measures].[FECHACREDITO1] <= <%= strFechaHasta %>)})) ON rows
from REPORTE7A

```

Figura V.93. Archivo de consulta MDX: mqCreditosGeneral.jsp

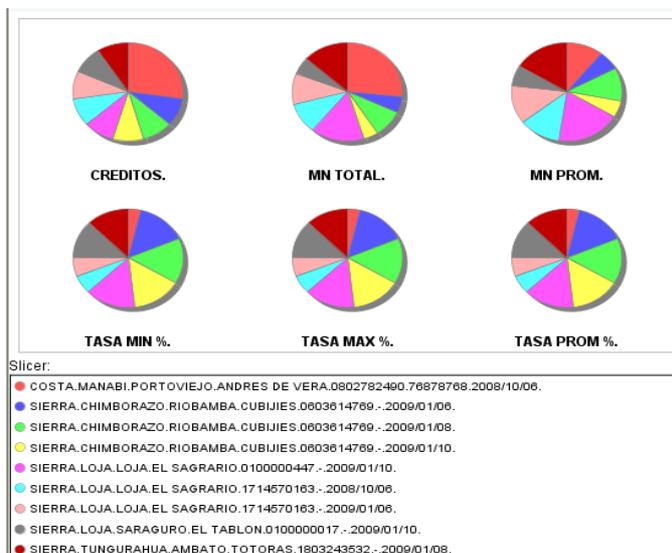


Figura V.94. Resultados gráficos del reporte: Beneficiarios del programa

Este reporte muestra todos los beneficiarios del Programa Sistema Nacional de Microfinanzas, en el gráfico dejamos activas todas las dimensiones y medidas que originalmente se declaro en la consulta MDX de la figura 73.

CONCLUSIONES

Como resultado del desarrollo del presente documento de tesis en donde se realizó como primer punto un estudio comparativo entre dos de las herramientas de análisis multidimensional Open Source denominadas Palo y Mondrian para luego implementar las aplicaciones web que muestren reportes dinámicos y estadísticos Business Intelligence, se muestra las siguientes conclusiones:

- La herramienta de análisis multidimensional Palo, por ser un software MOLAP no es transparente en cuanto al almacén de datos, razón por la cual no se puede conectar con una base de datos externa, si bien es cierto Palo si permite hacer conexiones ODBC pero solo a nivel de consultas pero no para rellenar los datos en el cubo de forma automática como lo hace Mondrian.
- Palo por estar desarrollado en conjunto con EXCEL no permite el análisis multidimensional en grandes volúmenes de datos, sino se restringe hasta 60000 registros que admite EXCEL 2003.
- Palo tiene limitantes muy grandes en el análisis multidimensional, y las más importantes son que tiene restricciones en cuanto al número de dimensiones y no permite realizar algunas operaciones de manipulación como son pivotar el cubo e intercambiar ejes. Hay que tomar en cuenta que Mondrian no tiene ninguna de estas limitantes.
- Es difícil para un usuario desarrollador manejar la herramienta de análisis multidimensional Mondrian si no tiene los conocimientos bien cimentados en el campo de business Intelligence y en cuanto al desarrollo web en jsp es decir en la plataforma de código abierto Java.
- Mondrian presenta problemas de rendimiento cuando se realiza análisis multidimensional sobre volúmenes de datos muy grandes, hablando de más de un millón de registros. Esto se debe a que debe ejecutarse en un servidor realmente potente en cuanto a memoria RAM puesto que su acercamiento a un software HOLAP hace que las estructuras dimensionales y la operación misma lo haga en caché.
- Si utilizamos un servidor con características normales es decir ni tan altas ni tan bajas se puede ver un buen rendimiento de Mondrian si la base de datos no sobrepasa los 100000 registros, teniendo en cuenta que esto es configurable para admitir volúmenes de datos más grandes, pero tiende a hacer lento sus procesos.

Por último hay que indicar que al iniciar esta tesis, el proyecto sobre el cual se implementó la solución BI Open Source utilizando Mondrian – Jpivot se denominaba RUB-ECUADOR (Registro Único de Beneficiarios del Ecuador), pero por disposiciones y políticas del MCDS a cambiado a un nuevo nombre **RIPS** (Registro Interconectado de Programas Sociales).

RECOMENDACIONES

A partir del desarrollo del presente trabajo se dará algunas conclusiones respecto del mismo, que los usuarios deberán tomar en cuenta para la correcta utilización de los software OLAP que han sido objeto de este estudio:

- Se recomienda a los usuarios que trabajan con hojas de cálculo Excel y que no necesitan realizar análisis OLAP sobre volúmenes grandes y complejos de datos, hacerlo con la herramienta de análisis multidimensional Palo, puesto que esta se adapta muy bien con Excel pero se debe tomar en cuenta que el análisis multidimensional que esta herramienta ofrece es muy limitado tanto en la administración que esta herramienta hace de las dimensiones como de los operadores de manipulación.
- Actualmente una sola combinación de usuario / contraseña está disponible para Palo. Aunque existe una ilimitada cantidad de usuarios que pueden usar las mismas credenciales. Si usted necesita más que un usuario nuestra sugerencia actualmente es que se utiliza múltiples carpetas de "Palo Web Client"
- Se recomienda también que si usted utiliza bases de datos en cualquier DBMS difundido en el mercado, y necesita realizar análisis multidimensional sobre esos datos que además tienen estructuras complejas, lo haga con Mondrian el mismo que permite realizar conexiones JDBC.
- Implementar Mondrian como herramienta OLAP en una empresa, requiere de un equipo que como mínimo tenga 16 Gb de memoria RAM y un buen procesador para que ofrezca un mejor rendimiento en su reporte dinámico, además de configurar la maquina virtual de java en el archivo de inicio del servidor sea este Tomcat el cual se utilizo para la implementación de esta solución o cualquiera para que el procesamiento y presentación de resultados sea eficiente.
- Cuando se trata de reportes que utilizan una gran cantidad de información el usuario debe tener claro las dimensiones y medidas que van a interactuar en el reporte BI, para mejorar el rendimiento.
- Al trabajar con Mondrian y si no se conoce tan a fondo el desarrollo de un análisis OLAP se recomienda utilizar la herramienta Schema Workbench de Pentaho que es totalmente Open Source y es una gran ayuda para el desarrollo de cubos OLAP con Mondrian.

RESUMEN

El objetivo de la tesis fue seleccionar, mediante un estudio comparativo entre Palo 2.5 y Mondrian 3.0.4, la herramienta más adecuada para el análisis multidimensional y la toma de decisiones, aplicado a los Programas Sociales del Ministerio de Coordinación y Desarrollo Social del Ecuador a través de la Consultora PROASETTEL S.A. ubicado en la ciudad de Quito.

La investigación se basó en el método científico general y se utilizó como herramienta de desarrollo e implantación Excel 2003, jdk 6, JPivot, JPalo, Schema Workbench 3.0, NetBeans IDE 6.5, Apache Tomcat 5.5, servidor de base de datos SQL Server 2005, sobre la plataforma Windows Server 2008.

Mediante parámetros de comparación de la Herramienta OLAP, se determinó un 95% para Mondrian y 85% para Palo, seleccionándose a Mondrian por ser la mejor opción para el desarrollo de reportes estadísticos de Business Intelligence que servirán a directivos y coordinadores de los programas sociales para la toma de decisiones.

Se desarrolló una solución realizando esquemas Datamart en los que residirán los datos para el análisis a nivel de Vistas en la Base de Datos; construyendo en Workbench los Cubos que son esquemas lógicos en la que residirán los datos en cache para su análisis y cálculos; diseñando páginas Web y reportes estadísticos de Business Intelligence para explotar los cubos desarrollados en NetBeans; publicando estas aplicaciones en el servidor Tomcat para la utilización de los usuarios y logrando obtener reportes estadísticos en tiempo real.

La herramienta implantada brinda mejores prestaciones para el desarrollo de reportes estadísticos, cabe mencionar que toda la solución se desarrolló con Software Libre a excepción de la Base de Datos.

SUMMARY

This thesis has had as its main objective to choose through a comparative study between Palo Mondrian 2.5 and 3.0.4, the most appropriate tool for multidimensional analysis and decision making, as applied to social programs of the Ministerio de Coordinacion de Desarrollo Social of Ecuador to through the consulting PROASETTEL S.A. located in the city of Quito.

The investigation was based on scientific method and was used as a tool for development and deployment of Excel 2003, jdk 6, JPivot, JPalo, Schema Workbench 3.0, NetBeans IDE 6.5, Apache Tomcat 5.5, database server SQL Server 2005 on Windows Server platform 2008.

By benchmarking the OLAP tool, was determined for 95% to Mondrian and 85% to Palo selected as the best option for the development of statistical reports that will provide business intelligence to managers and coordinators of programs for social decision making.

We developed a solution made in schemes Datamart staying data for the analysis of views in the database, built on the Cubes Workbench are schema staying in the cached data for analysis and calculations, design Web pages of statistical reports and business intelligence to exploit the cube developed in NetBeans; running these applications on the Tomcat server for the use of users and achieved statistical reports in real time.

The implemented tool provides better benefits for the development of statistical reports, which include the entire solution was developed with Open Source except for the Database.

GLOSARIO

DWH: Por el inglés Data warehouse. Representa a una bodega de datos.

Datamart: Es un subconjunto especializado de una bodega de datos. Los diferentes datamarts contienen diferentes combinaciones y selección de los mismos datos detallados identificados en la bodega de datos, por esto puede decirse que los datamarts vienen a ser como una extensión natural de la bodega de datos.

Granularidad: Se refiere al mayor nivel de detalle.

JDBC: Java Database Connectivity, API de Java que permite hacer conexiones a bases de datos para manipular la información que están contienen.

Metadata: Información sobre los datos almacenados en la bodega de datos, permite conocer las fuentes de donde es extraído y reglas de negocio.

OLAP: ON LINE ANALYTICAL PROCESSING. Es la tecnología que permite a los usuarios tener una visión de los datos de una forma rápida, interactiva y fácil de usar.

OLTP: ON LINE TRANSACTIONAL PROCESSING. Las aplicaciones de OLTP están organizadas para ejecutar las transacciones para los cuales fueron hechos, como por ejemplo: mover dinero entre cuentas, un cargo o abono, una devolución de inventario, etc. Por otro lado, una bodega de datos está organizada en base a conceptos, como por ejemplo: clientes, facturas, productos, etc.

Query: Consultas sobre una base de datos o sobre una bodega de datos. Permiten manipular la información de las bases de datos y las bodegas.

RDBMS: Manejador de bases de datos relacional. Constituyen la base para los sistemas OLTP.

Sistemas Legacy: Sistemas fuente de la empresa con los que manejan la operación de las transacciones diarias. Su característica principal es el enfoque en los procesos de operación diaria de la empresa.

BIBLIOGRAFIA

BIBLIOGRAFÍA GENERAL

- (1) KIMBALL, R. The Data Warehouse Toolkit. New York: Wiley, 1996
- (2) KIMBALL, R. The Data Warehouse Lifecycle Toolkit. New York: Wiley, 1998
- (3) CERVO, L. Metodología Científica. Bogotá: Mc Graw, 1998
- (4) METODOS DE INVESTIGACION

<http://www.monografias.com/trabajos/metoinves/metoinves.zip>

200809

- (5) METODO CIENTIFICO

<http://www.hondurassilvestre.com/Reportajes/MetodoCientifico.htm>

200809

- (6) METODO CIENTIFICO

<http://www.monografias.com/trabajos21/metodo-cientifico/metodo-cientifico.shtml>

200809

BIBLIOGRAFÍA RELACIONADA AL TEMA

- (7) KRISHNAMURTHY, S. An Analysis of Open Source Business Models.

Washington: MIT Press, 2005

- (8) SOBRE PALO 2.0

http://www.telefonica.net/web2/todobi/Nov07/Sobre_PALO20.pdf

200901

- (9) PALO FEACTURES

<http://www.jedox.com/en/enterprise-spreadsheet-server/excel-olap-server/palo-features.html>

200901

(10) ANALISIS DEL ESTADO DE MONDRIAN (OLAP OPEN SOURCE)

http://www.telefonica.net/web2/todobi/Mayo07/Analisis_de_Mondrian.pdf

200902

(11) DATAMART (DATA WAREHOUSE) TOOL: MONDRIAN + JRUBIK

<http://www.icis.cgiar.org/icis/images/e/ea/DataMart-Tool.pdf>

200903

(12) AN OLAP SOLUTION USING MONDRIAN AND JPIVOT

http://eric.univ-lyon2.fr/~sbimonte/doc/presentation_2007-02.pps

200903

(13) CREATING INTERACTIVE OLAP APPLICATIONS WITH MYSQL ENTERPRISE AND MONDRIAN

http://eric.univ-lyon2.fr/~sbimonte/doc/presentation_2007-02.pps

200904

(14) HERRAMIENTAS OPENSOURCE J2EE DE DW

https://www.u-cursos.cl/ingenieria/2007/1/IN830/1/material_docente/objeto/1298

200904

ANEXOS

Anexo 1:

NAVEGANDO EN EL DATAMART

Mondrian es una herramienta supremamente robusta en cuanto al manejo de hipercubos y al tratamiento de bases de datos se refiere. El usuario de Mondrian debe tener conocimientos básicos de OLAP, generación y uso de hipercubos e inteligencia de negocio y administración.

Por lo general, las herramientas OLAP son dirigidas a usuarios que toman decisiones, como ejecutivos, gerentes o buzos de información. Por lo tanto son herramientas que permiten una interacción relativamente sencilla con el usuario, permitiéndole generar un cubo gráficamente o hacer gráficas con los resultados sin mayores conocimientos técnicos.

Para este ejemplo se cita uno de los reportes realizados en la aplicación web Programa Social Socio Ahorro (PSSA) el mismo que ofrece el siguiente menú:



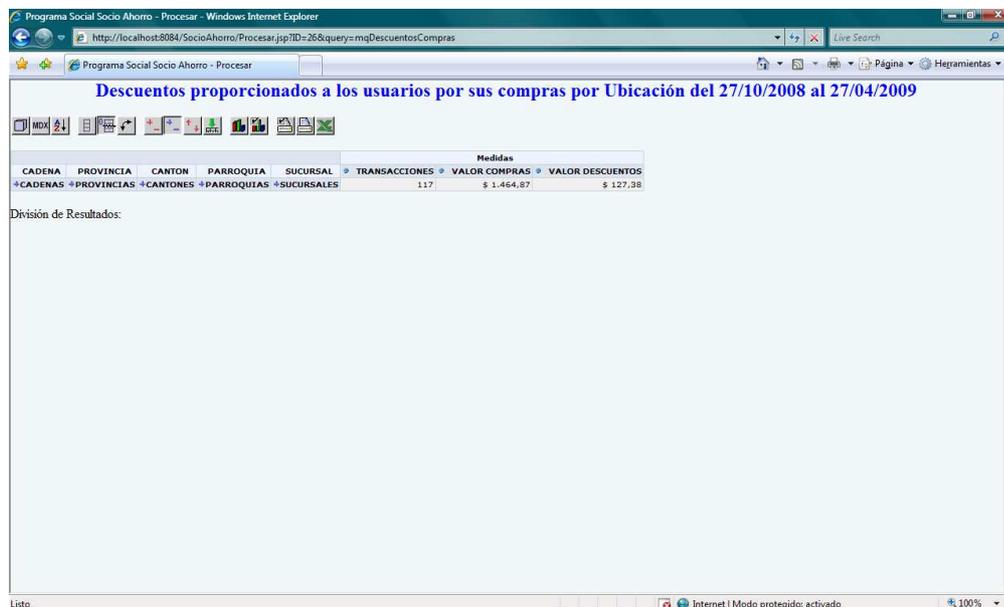
En la figura anterior se escogerá el botón “Reportes” para ir a la ventana que muestra los reportes implementados:



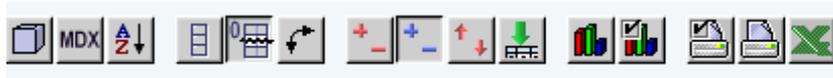
La figura anterior muestra los enlaces a los reportes implementados en el Programa Social Socio Ahorro, a continuación escogemos el reporte que muestra los “Beneficiarios por Provincia” y saldrá como resultado la siguiente ventana:



Esta muestra los parámetros que se envía a la base de datos en SQL Server 2005 para que se construya el datamart mediante esas fechas de inicio y de fin, teniendo la posibilidad de escoger la fecha interactivamente en un calendario desarrollado en jsp.



Para explicar los modos de opciones de navegación en la herramienta se seguirá el proceso de realizar un reporte típico, para ello empezaremos utilizando cada uno de los botones de la barra de navegación OLAP que ofrece Mondrian.



Los botones que se mostraron en la figura son los siguientes, en su respectivo orden:

- Navegador OLAP
- Mostrar editor MDX
- Configurar tabla OLAP
- Ocultar Repeticiones
- Suprimir filas/columnas vacías
- Detallar miembro
- Abrir detalle
- Entrar en detalle
- Mostrar datos origen
- Mostrar Gráfico
- Configurar Gráfico
- Configurar impresión
- Exportar a PDF
- Exportar a Excel

La tabla a continuación descrita, muestra una lista de la simbología empleada en la barra de navegación interactiva del reporte estadístico implementado en esta aplicación.

Esta simbología está acompañada de una descripción para cada uno de los botones y/o símbolos que componen la misma, de la siguiente manera:

Barra de Navegación de los Reportes Estadísticos	
SIMBOLOGÍA	DESCRIPCIÓN
	<p>Navegador OLAP, le da varias opciones al usuario para personalizar su reporte estadístico, habilitando y/o deshabilitando las dimensiones o filas informativas y las medidas o columnas de valores numéricos.</p> <p>Además ayuda al usuario para filtrar sus resultados estadísticos por dimensiones y/o medidas.</p>
	<p>Mostrar Editor MDX, muestra un editor avanzado para realizar una consulta mdx, para ello el usuario debe tener un buen conocimiento en construcción de cubos utilizando el lenguaje mdx.</p>
	<p>Configurar tabla OLAP, muestra un cuadro de dialogo que permite configurar la tabla de forma que se pueda ordenar los registros de las dimensiones y también las medidas, por grupos y por jerarquía.</p>
	<p>Ocultar Repeticiones, sirve para ocultar o habilitar los nombres repetidos en el reporte estadístico.</p>
	<p>Suprimir Filas/Columnas vacías, permite ocultar o mostrar los registros de las dimensiones con medidas vacías.</p>
	<p>Intercambiar ejes, permite intercambiar los ejes de las dimensiones mostradas en el reporte estadístico.</p>
	<p>Detallar miembro, muestra el detalle de todos los miembros del reporte estadístico</p>
	<p>Abrir detalle, muestra el detalle por cada miembro del reporte estadístico seleccionado</p>
	<p>Entrar en detalle, muestra en detalle el reporte estadístico, sin repeticiones</p>
	<p>Mostrar datos origen, permite mostrar los datos origen para una medida o columna de valor numérico dada</p>

	<p>Mostrar gráficos, muestra un gráfico estadístico de acuerdo al tipo de gráfico y especificaciones configuradas por el usuario</p>
	<p>Configurar gráficos, muestra un cuadro de diálogo de configuración de gráficos estadísticos, con varias opciones como:</p> <ul style="list-style-type: none"> • Tipo de gráfico, que son barras, barras apiladas, áreas y pasteles en dos y tres dimensiones. • Título del gráfico. • Dimensiones del gráfico • Etiquetas, para los ejes horizontal y vertical
	<p>Configurar impresión, permite configurar la impresión que se exportará a PDF, y se puede ingresar el título para el reporte, sus dimensiones y damos la facilidad al usuario de escoger si desea el reporte de forma horizontal o vertical.</p>
	<p>Exportar a PDF, exporta el resultado del reporte estadístico business Intelligence en formato PDF con las especificaciones de impresión configuradas por medio del botón anterior</p>
	<p>Exportar a Excel, exporta el resultado del reporte estadístico business Intelligence en formato EXCEL</p>

Anexo 2:

MANUAL TÉCNICO

Programa Sistema Nacional de Microfinanzas

INTRODUCCIÓN

Este manual tiene como objetivo fundamental explicar los diferentes mecanismos que permiten la conexión a los reportes estadísticos del Programa Sistema Nacional de Microfinanzas de tal forma que el usuario de la aplicación web, pueda tener un punto de vista técnico acerca de la misma.

BASES DE ESPECIFICACIÓN

OBJETIVOS DE DESARROLLO

- Implementar una aplicación web que permita brindar al Ministerio de Coordinación y Desarrollo Social información específica y detallada, con posibilidad de realizar análisis multidimensional sobre la misma correspondiente al Programa Sistema Nacional de Microfinanzas.
- Construir una interfaz eficiente y amigable que permita la visualización de la información de análisis.

ABREVIATURAS Y SIGLAS

- DBMS: Sistema de Administración de Bases de Datos.
- JDK: Java Development Kit. Entorno de Ejecución Java.
- MCDS: Ministerio de Coordinación de Desarrollo Social.
- PSNM: Programa Sistema Nacional de Microfinanzas.

VISIÓN GENERAL DEL PRODUCTO

- PLATAFORMA DE PROGRAMACION: jdk-6u7-windows-i586-p
- MOTOR DE BASE DE DATOS: SQL Server 2005
- MOTOR OLAP: Pentaho Analysis (Mondrian Jpivot)

- SISTEMA OPERATIVO: Microsoft Windows, Linux.

VENTAJAS DE LA CONEXIÓN POR INTERFAZ WEB

Actualización permanente: Mediante el uso de la interfaz WEB de administración y consulta usted siempre tendrá a disposición la última versión de la aplicación de consulta en línea sin necesidad de reinstalaciones de nuevos programas.

Seguridad: Al utilizar navegadores estándar se tiene disponible el protocolo de seguridad SSL (Capa Segura de Sockets) que proporciona un ambiente de navegación seguro.

Múltiples sistemas operativos: Con los programas tradicionales de administración y consulta usted sólo contaba con versiones para plataformas Windows. La interfaz web puede ser utilizada desde cualquier sistema operativo siempre y cuando cuente con un programa navegador como MS Internet Explorer o Netscape Navigator por ejemplo.

Ambiente de trabajo estándar: Los programas de navegación web son programas muy populares. Por lo tanto los analistas no necesitan enfrentarse a un ambiente de trabajo radicalmente diferente al que están acostumbrados.

La interfaz web de consulta y análisis se convierte en el medio versátil y eficaz para la toma de decisiones por parte de los directivos del Programa de acuerdo a la información disponible en el sistema fuente.

BENEFICIOS O CARACTERÍSTICAS DE IMPLEMENTACIÓN

La implantación de la aplicación web de análisis para el Programa Sistema Nacional de Microfinanzas permitirá obtener los siguientes beneficios:

- Conocer y administrar dinámicamente los reportes Business Intelligence que se han implementado en la aplicación Web para el Programa Sistema Nacional de Microfinanzas.
- Ofrecer un alto nivel de portabilidad, al permitir que la aplicación web implementada se ejecute en cualquier tipo de navegador web (browser).

DESCRIPCIÓN GENERAL

PERSPECTIVA DEL PRODUCTO

La aplicación web de análisis del Programa Sistema Nacional de Microfinanzas **PSNM**, cubre todos y cada uno de los parámetros de, reportes de información procesada, a los cuales se entregan una documentación clara.

Además presenta un buen nivel de seguridad, tanto a nivel de las bases de datos implementadas para esta aplicación así como también en la interfaz web misma de manera que se garantice los esquemas fundamentales de seguridad requeridos.

ATRIBUTOS DE LA APLICACIÓN

La aplicación web de análisis de información del Programa Sistema Nacional de Microfinanzas (**PSNM**), es un Software de propósito específico, adaptable y configurable que de manera general tendrá atributos específicos como:

Disponibilidad

El sistema deberá estar considerado para estar disponible para todos sus usuarios, por ser una aplicación Web.

Seguridad y Confiabilidad

La seguridad estará orientada a la creación de sesiones y al acceso de las aplicaciones mediante un login contenida en la aplicación web de administración o gestión de datos del Programa, que tendrá 2 factores a evaluar, la cuenta de usuario y la contraseña, para comprobar si se trata de un usuario autorizado y desde esta mediante enlaces podrá acceder a los reportes de análisis de información del Programa.

Si el identificador introducido no coincide con la almacenada, se dará una indicación de error y no podrá acceder al análisis de datos. Los tipos de usuario que se van a contemplar y las labores que corresponden a cada uno de ellos son:

- **Súper Usuario**, es aquel usuario que tiene acceso al sistema sin limitaciones y permisos de acceso al área de reportes estadísticos en este tipo de usuario podrán estar: Ministros, Asesores, etc.

- **Usuario coordinador del PSNM**, el mismo que tendrá acceso al sistema mediante permisos para la visualización y el análisis de la información requerida de acuerdo a las opciones solicitadas.

Integridad de Datos

La integridad de los datos se precautela a través de: el esquema de base de datos implementada, la aplicación web desarrollada.

Mantenibilidad

Se pretende que la documentación sea la necesaria y provea de claros parámetros para obtener facilidad en el mantenimiento en caso de ser requerido.

Portabilidad

El sistema al ser una aplicación Web correrá en cualquier navegador Web, logrando de esta manera un alto nivel de portabilidad para la aplicación.

INTERFACES

INTERFACES EXTERNAS

INTERFACES DE USUARIO	INTERFACES DE SOFTWARE
<p>El producto se desenvolverá en un ambiente WINDOWS, LINUX.</p> <p>El usuario podrá visualizar varias presentaciones y diseño de pantallas en un navegador web.</p> <p>Además contara con especificaciones básicas tales como cajas, botones.</p>	<p>PLATAFORMA DE PROGRAMACION:</p> <p style="padding-left: 40px;">jdk-6u7-windows-i586-p</p> <p>MOTOR DE BASE DE DATOS:</p> <p style="padding-left: 40px;">SQL Server 2005</p> <p>MOTOR OLAP:</p> <p style="padding-left: 40px;">Pentaho Analysis (Mondrian-Jpivot)</p> <p>SISTEMA OPERATIVO:</p> <p style="padding-left: 40px;">Microsoft Windows, Linux</p>

INTERFACES DE HARDWARE

<ul style="list-style-type: none">• La Visualización gráfica, mostrada en el monitor y generada por el GDI de Windows• Conexión directa a la impresora, teclado, Mouse haciendo uso de los diferentes puertos, dando la opción de crear un reporte, realizar consultas y la entrada de datos.	<ul style="list-style-type: none">• Distribución elevada de recursos a nivel de procesador (Servidor – Requerimiento de Instalación).• Protocolo de Conexión TCP – IP.• Tarjetas de Red de alta confiabilidad.
--	---

INTERFACES DE COMUNICACIÓN

- Protocolo TCP / IP.
- DNS (Domain Name System)
- DHCP(Dinamic Host Configuration Protocol)

LIMITACIONES GENERALES

LIMITACIONES DE SOFTWARE

El MCDS cuenta con el respectivo Software para la implementación e implantación, dentro de los requerimientos planteamos los siguientes:

1. SISTEMA OPERATIVO

servidor:	Windows, Linux
Cliente:	Windows, Linux

2. DESARROLLADOR

PROGRAMACIÓN:	jdk-6u7-windows-i586-p
---------------	------------------------

3. SISTEMA DE ADMINISTRACIÓN DE BASE DE DATOS

DBMS	SQL Server 2005
------	-----------------

4. ANÁLISIS MULTIDIMENSIONAL

PROGRAMACIÓN:	Pentaho Analysis (Mondrian - Jpivot)
---------------	--------------------------------------

LIMITACIONES HARDWARE

El entorno hardware del MCDS no posee limitantes debido a que en su infraestructura o backbone posee todos los equipos requeridos para la implementación e implantación del sistema a desarrollarse. Los clientes (usuarios externos al Programa Social) deben contar con conexión a Internet.

LIMITACIONES DE ADAPTACIÓN DEL LUGAR

El Lugar de Implementación cubre a cabalidad con las necesidades del nuevo sistema por lo que cumple con un lugar destinado solamente para el manejo de servidores con las respectivas seguridades.

DEFICIENCIAS DE PERSONAL INFORMÁTICO

El Ministerios de Coordinación y Desarrollo Social (**MCDS**) debe contar con personal que administre el sistema.

Dicho personal debe reunir una serie de características en su perfil como se lista en la tabla de la siguiente página:

PERFIL TECNICO I : Desarrollador de Aplicaciones Web con herramientas Open Source
Formación Académica
Título Universitario en el área Informática o Sistemas
Certificación en herramientas de desarrollo Open Source, o cursos sobre el tema

Conocimientos amplios acerca de Business Intelligence

Experiencia

Sólidos conocimientos en bases de datos relacionales, énfasis en Microsoft SQL Server 2005

Experiencia en Programación en Java

Experiencia en Instalación y configuración de servidores Web Apache Tomcat

Sólidos conocimientos en Linux CentOS

Sólidos conocimientos de Web Services: énfasis Java Web Services

Experiencia específica en proyectos para ambientes WEB

Tareas

Revisar y mantener actualizados los sistemas de análisis de los programas sociales del MCDS

Instalación y configuración de Linux en servidores de la Institución

Construcción de web Services requeridos.

Implantación, configuración y optimización de servidores Web.

Soporte a usuarios de los sistemas de análisis de los programas sociales del MCDS

Realizar otras tareas afines y complementarias, conforme a lo asignado por su superior inmediato.

Cumplir las metas individuales que le sean asignadas y los compromisos que ellas conlleven, conforme a la naturaleza del cargo.

FASES

Fase 1. PLANIFICACIÓN Y ESPECIFICACIÓN DE REQUISITOS

Introducción

Esta fase corresponde con la Especificación de Requisitos Tradicional ampliada con una sección de análisis / gestión de riesgos y con una definición de casos de uso de alto nivel.

Entonces es necesario mencionar que el objetivo fundamental de esta fase, es mostrar para el Programa Sistema Nacional de Microfinanzas la respectiva documentación acerca del análisis y gestión del riesgo y además un conjunto de requisitos en forma modelada utilizando los mencionados Casos de Uso de alto nivel.

Estos casos de uso han sido modelados a partir de los requerimientos de los usuarios que a su vez han sido tomados mediante la interacción que ha habido con los usuarios coordinadores de este programa.

Análisis y Gestión del Riesgo

Para el desarrollo correcto de la aplicación web del Programa Sistema Nacional de Microfinanzas que por cierto es la principal aplicación de administración para el mismo ya que además de ella hemos desarrollado una aplicación de escritorio la misma que ayuda a crear un archivo con los registros de posibles beneficiarios de los microcréditos, para que con este archivo el coordinador del programa pueda subir para que la aplicación web del PSND los ingrese o los actualice automáticamente en la base de datos que reside en SQL Server 2005.

Con lo anterior en mente es necesario identificar y analizar los riesgos que puedan suscitarse durante el desarrollo, implementación y puesta en producción del mismo, los cuales pueden afectar levemente y hasta convertirse en un verdadero problema, en cuanto a la planificación, calidad y factibilidad del proyecto, el cual deberá ser gestionado adecuadamente de acuerdo a las estrategias y supervisión que se plantee a continuación para cada riesgo.

Identificación del Riesgo

IDENTIFICACIÓN	DESCRIPCIÓN DEL RIESGO	CATEGORÍA	CONSECUENCIA
R1	No existe disponibilidad de tecnología	Riesgo técnico	- No se podría implementar el software
R2	No existe bases de datos	Riesgo técnico	- No se podría obtener la información del programa
R3	No existen herramienta de desarrollo de software	Riesgo técnico	- No se podría desarrollar e implementar el software rápidamente
R5	Los usuarios se resisten a utilizar el sistema	Riesgo del Negocio	- Proyecto desarrollado en vano pérdida de recursos
R6	Los usuarios no se adaptan al sistema	Riesgo del Negocio	- Los objetivos del proyecto no se cumplirían
R7	Dificultad en operar el software	Riesgo del Negocio	- Pérdida de beneficiarios por ejemplo
R8	Interfaces inadecuadas	Riesgo Técnico	- Demora en la comprensión para la operación del software
R9	Verificaciones incompletas	Riesgo Técnico	- Inestabilidad del software
R10	Validación incompleta	Riesgo del Negocio	- Errores en el desempeño del software
R12	Falta de conocimiento en el funcionamiento de las herramientas case	Riesgo Técnico	-Retrasos en el proyecto.
R13	La base de datos no se adapta al lenguaje de programación utilizado	Riesgo Técnico	-Retrasos en el proyecto.

R15	No existe compatibilidad del software con la plataforma a utilizar	Riesgo Técnico	-Buscar alternativas de implementación del software.
R16	No existe una identificación correcta del usuario que inicia la sesión	Riesgo Técnico	-No puede ingresar a su sesión porque no existe identificación correcta del usuario.
R17	Existe una conexión inestable con las bases de datos	Riesgo Técnico	-Información incompleta.
R21	Mala información recolectada de los usuarios	Riesgo del Proyecto	-Proyecto de mala calidad.

Análisis del Riesgo

a. Determinación de la Probabilidad

ID	PROBABILIDAD
R1	40%
R2	10%
R3	50%
R5	30%
R6	25%
R7	60%
R8	70%
R9	40%
R10	25%
R12	5%
R13	40%
R15	40%
R16	35%
R17	38%
R21	40%

RANGO PROBABILIDAD	DESCRIPCIÓN	VALOR
1...33	BAJA	1
34...67	MEDIA	2
68...99	ALTA	3

b. Determinación del Impacto

DESCRIPCIÓN	VALOR
Bajo	1
Media	2
Alto	3
Crítico	4

c. Determinación de la Exposición al riesgo

EXPOSICIÓN AL RIESGO	COLOR	VALOR
BAJA	AMARILLO	1 o 2

MEDIA	VERDE	3 o 4
ALTA	CELESTE	Mayor o igual que 6

d. Determinación de la Prioridad del riesgo

ID	Probabilidad			Impacto		Exposición al riesgo		Prioridad
	%	Valor	Posibilidad	Valor	Impacto	Valor	Exposición	
R17	38	Media	2	Critico	4	Alta	8	1
R8	70	Alta	3	Moderado	2	Alta	6	2
R13	40	Media	2	Alto	3	Alta	6	2
R21	40	Media	2	Alto	3	Alta	6	2
R1	40	Media	2	Moderado	2	Media	4	3
R2	10	Baja	1	Critico	4	Media	4	3
R9	40	Media	2	Moderado	2	Media	4	3
R15	40	Media	2	Moderado	2	Media	4	3
R16	35	Media	2	Moderado	2	Media	4	3
R3	50	Media	2	Bajo	1	Baja	2	4
R7	60	Media	2	Bajo	1	Baja	2	4
R5	30	Baja	1	Bajo	1	Baja	1	5
R6	25	Baja	1	Bajo	1	Baja	1	5
R10	26	Baja	1	Bajo	1	Baja	1	5
R12	5	Baja	1	Bajo	1	Baja	1	5

Plan de Reducción, Supervisión y Gestión del Riesgo

En esta sección vamos a mostrar un plan de reducción, supervisión y gestión de los riesgos que hemos considerado más probables, y son los que anteriormente se han señalado con color celeste.

HOJA DE GESTIÓN DEL RIESGO			
ID. DEL RIESGO: R17		FECHA: 17-11-2008	
Posibilidad: 2 Valor: MEDIA	Impacto: 4 Valor: CRITICO	Exposición: 8 Valor: ALTO	Prioridad: 1 Valor: ALTO

DESCRIPCIÓN: Existe una conexión inestable con las bases de datos
REFINAMIENTO: <p>Causa 1: No hay compatibilidad de la aplicación web con el DBMS utilizado</p> <p>Causa 2: No se estudio a fondo los requerimientos del sistema</p> <p>Consecuencia: Retraso en la planificación.</p>
REDUCCIÓN Y SUPERVISIÓN: <ul style="list-style-type: none"> - Realizar un análisis de la compatibilidad existente entre el DBMS a utilizar y la plataforma en la que se desarrolla la aplicación web. - Supervisar en lo posible si el sistema se <i>cae por causa</i> de la conexión utilizada entre la aplicación web implementada y el DBMS utilizado en este caso SQL Server 2005
GESTIÓN: <ul style="list-style-type: none"> - La aplicación web desarrollada en la plataforma java (JSP) permite una conexión estable con el DBMS utilizado (SQL Server 2005) mediante sqldrbc 2.0 de Microsoft que se puede descargar totalmente gratis desde el sitio oficial de Microsoft.
ESTADO ACTUAL: <p>Fase de reducción realizada</p> <p>Fase de Supervisión realizada</p> <p>Fase de Gestión realizada</p>

HOJA DE GESTIÓN DEL RIESGO			
ID. DEL RIESGO: R8		FECHA: 17-11-2008	
Posibilidad: 3 Valor: ALTO	Impacto: 2 Valor: MODERADO	Exposición: 6 Valor: ALTO	Prioridad: 2 Valor: ALTO
DESCRIPCIÓN: Interfaces inadecuadas			

REFINAMIENTO:

Causa 1: No se ha tomado el tiempo adecuado para el desarrollo de las interfaces.

Causa 2: No se ha definido correctamente los requerimientos.

Causa 3: Las encuestas no fueron desarrolladas a las personas correctas

Consecuencia: Demora en la comprensión para la operación del software

REDUCCIÓN Y SUPERVISIÓN:

- Realizar una interfaz amigable con el usuario.
- Que la interfaz sea de fácil comprensión
- Presentar a las personas que operarán el software para que opinen sobre la interfaz que se está desarrollando.

GESTIÓN:

- Contratar a un diseñador gráfico para que diseñe las interfaces.

ESTADO ACTUAL:

Fase de reducción realizada

Fase de Supervisión realizada

Fase de Gestión realizada

HOJA DE GESTIÓN DEL RIESGO

ID. DEL RIESGO: R13

FECHA: 17-11-2008

Posibilidad: 2

Impacto: 3

Exposición: 6

Prioridad: 2

Valor: MEDIA

Valor: ALTO

Valor: ALTO

Valor: ALTO

DESCRIPCIÓN: La base de datos no se adapta al lenguaje de programación utilizado

REFINAMIENTO:

Causa 1: No se realizó un estudio adecuado sobre el DBMS que se necesitaba en la implementación.

Causa 2: La base de datos existente no es compatible con la base de datos a implementar.

Consecuencia: Proyecto de mala calidad

REDUCCIÓN Y SUPERVISIÓN:

- Verificar la compatibilidad del DBMS a utilizar en este caso SQL Server 2005 con la plataforma de programación Web a utilizar (JSP)
- Realizar pruebas para ver la compatibilidad
- Comprobar que se estén cumpliendo con los requisitos

GESTIÓN:

- Buscar controladores para realizar la comunicación entre el lenguaje y el DBMS a utilizar u otras alternativas que nos presenta el mercado.

ESTADO ACTUAL:

Fase de reducción realizada

Fase de Supervisión realizada

Fase de Gestión realizada

HOJA DE GESTIÓN DEL RIESGO

ID. DEL RIESGO: R21		FECHA: 17-11-2008	
Posibilidad: 2 Valor: MEDIA	Impacto: 3 Valor: ALTO	Exposición: 6 Valor: ALTO	Prioridad: 2 Valor: ALTO
DESCRIPCIÓN: Mala información recolectada de los usuarios			
REFINAMIENTO: Causa 1: El entrevistador no tomó las muestras adecuadas Causa 2: El cuestionario de las encuestas no estaban bien desarrolladas Causa 3: Las encuestas no fueron desarrolladas a las personas <i>correctas</i> Consecuencia: Proyecto de mala calidad			
REDUCCIÓN Y SUPERVISIÓN: <ul style="list-style-type: none">- Realizar el cuestionario y formato de las encuestas correctamente de modo que estos sean claros y <i>precisos</i> para la rápida comprensión.- Comprobar que se estén cumpliendo con los formatos adecuados al realizar las encuestas.- Verificar que las encuestas para la recolección de requerimientos se hagan a los usuarios que operan el Programa Sistema Nacional de Microfinanzas.			
GESTIÓN: <ul style="list-style-type: none">- Recolectar los requerimientos con los coordinadores del Programa Sistema Nacional de Microfinanzas.			
ESTADO ACTUAL: Fase de reducción realizada Fase de Supervisión realizada Fase de Gestión realizada			

Requerimientos Funcionales

Supuestos y Dependencias del Sistema

La aplicación web de análisis de información del Programa Sistema Nacional de Microfinanzas no posee dependencia del sistema operativo Linux, solo se lo especifica como requisito para su implementación por la seguridad y facilidad de manejo que brindaría al sistema, dando fiabilidad en el manejo de la información.

Las ventajas principales del sistema operativo son la restricción de accesos y administración de usuarios, seguridades que el sistema requiere en forma opcional de acuerdo al grado de factibilidad.

Requisitos Funcionales Específicos

IDENTIFICADOR	RF05-PSNM	
FUNCIÓN:	Validar datos o parámetros de análisis	
CATEGORIA:	Oculto	Proceso
DESCRIPCION:	El sistema debe realizar controles lógicos y de existencia sobre los parámetros de análisis: <ul style="list-style-type: none">- Verificar si el o los parámetros solicitados y proporcionados se encuentran en la Base de Datos para construir el reporte de análisis de información.	
CONSIDERACION:	<ul style="list-style-type: none">- Si el o los parámetros no se encuentran en la Base de Datos este le indicara el mensaje de no existir información necesaria para construir el reporte de análisis.	

IDENTIFICADOR	RF08-PSNM	
FUNCIÓN:	Emitir reportes y exportación de información	
CATEGORIA:	Evidente	Salida
DESCRIPCION:	El sistema debe permitir al usuario consultar e imprimir reportes con la	

	<p>capacidad de exportar a otros formatos:</p> <ul style="list-style-type: none">- Lista de posibles beneficiarios no calificados (rechazados) en una solicitud enviada por la Entidad Microfinanciera.- Listado y cantidad de Entidades Microfinancieras calificadas y vigentes en general/por provincia.- Listado y cantidad de créditos colocados (beneficiarios) y monto total en general y por Entidades Microfinancieras, en un rango de fechas.- Listado y cantidad de beneficiarios según el nivel de pobreza (A, B, C) y monto total, en general y por Entidades Microfinancieras, en un rango de fechas.- Listado y cantidad de beneficiarios que reciben el Bono de Desarrollo Humano y monto total, en general y por Entidades Microfinancieras, en un rango de fechas.- Listado y cantidad de beneficiarios según la calificación de quintiles obtenidas de la Base de Datos del SELBEN y monto total, en general y por Entidades Microfinancieras, en un rango de fechas.- Listado y cantidad de personas no calificadas (rechazadas) por solicitud con la razón por la cual fue negada.- Cantidad de créditos colocados (beneficiarios) y monto total por género/provincia/cantón/parroquia, en general y por Entidades Microfinancieras, en un rango de fechas.- Cantidad de créditos colocados (beneficiarios) y monto total por actividad microempresarial/provincia/cantón/parroquia, en general y por Entidades Microfinancieras, en un rango de fechas.- Cantidad de créditos colocados (beneficiarios) y monto total a los beneficiarios del Bono de Desarrollo Humano por provincia/cantón/parroquia, en general y por Entidades Microfinancieras, en un rango de fechas.- Cantidad de créditos colocados (beneficiarios) y monto total asociativos e individuales por provincia/cantón/parroquia, en general y por Entidades Microfinancieras, en un rango de fechas.
--	---

	- Información estadística de montos, plazo y tasa promedio en general y por Entidad Microfinanciera.
CONSIDERACION:	

Fase 2. CONSTRUCCIÓN

Introducción

Esta fase tiene como objetivo fundamental presentar una documentación clara acerca del diseño de alto nivel así como del diseño de bajo nivel, los mismos que a continuación se explican.

En la fase de Diseño de Alto Nivel de un ciclo de desarrollo se investiga sobre el problema, intentando diseñar lo que va a ser la interacción del sistema con el exterior.

Se identifican los conceptos relacionados con el subconjunto de casos de uso que se esté tratando. Los detalles de implementación se dejan para la fase de Diseño de Bajo Nivel.

Cuando el ciclo de desarrollo no es el primero, antes de la fase de Diseño de Alto Nivel hay una serie de actividades de planificación.

Estas actividades consisten en actualizar los modelos que se tengan según lo que se haya implementado, pues siempre se producen desviaciones entre lo que se ha diseñado y lo que finalmente se construye.

Diseño de alto Nivel

Refinamiento de Casos de Uso

En el refinamiento de casos de uso, tratamos de dar una mayor explicación a los usuarios acerca de la visión general de cada proceso que se va a realizar en el Programa Sistema Nacional de Microfinanzas de tal forma que se pueda entender y conocer su funcionamiento básico, así como de los actores que intervienen en el mismo.

a. Casos de Uso esenciales en formato expandido

Aquí mostramos los casos de uso con una visión más allá de la que vimos en la fase uno, es decir le adjuntamos una tabla denominada "Curso Normal de Eventos", en donde los usuarios pueden ver como inicialmente van a interactuar con el sistema.

Caso de uso:	Validar datos
Requerimiento que implementa	RF05-PSNM RF01-G
Actores	Sistema PSNM
Descripción:	Cada vez que se ingresa o modifica datos es necesaria su validación.
Referencias cruzadas:	
Precondición :	
Postcondición:	Actualiza datos y asigna la respectiva observación.

Curso normal de eventos:

Acción de los actores	Respuesta del sistema
1. Este caso de uso comienza cuando el sistema del PSNM recepta datos	2. Valida datos según restricciones establecidas
	3. Actualiza la base de datos, si está o no Calificado y le asigna Observaciones

Caso de uso:	Emitir reportes y exportación de información
Requerimiento que implementa	RF08-PSNM
Actores	Usuario PSNM
Descripción:	El Usuario PSNM puede consultar e imprimir reportes a través de la web, con la capacidad de exportar a otros formatos.
Referencias cruzadas:	Casos de uso : Validar datos

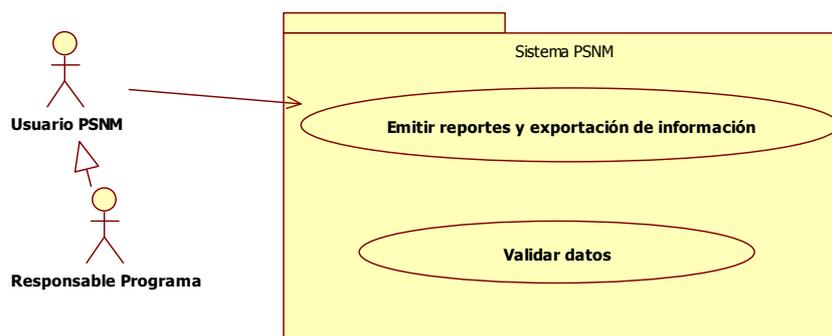
Precondición :	- Datos del Usuario PSNM correctos
Postcondición:	- Reporte listo para consultar, analizar, configurar variables disponibles, imprimir y/o exportar.

Curso normal de eventos:

Acción de los actores	Respuesta del sistema
1. Usuario PSNM ingresa al sistema	2. Muestra opciones: - Reportes por categoría
3. Usuario PSNM escoge la categoría del reporte que necesita	4. Solicita datos o parámetros para generar el reporte estadístico.
	5. Muestra el reporte de acuerdo a la categoría seleccionada, con opciones para realizar el análisis de datos disponible en el reporte.
6. Configura las variables del reporte estadístico de acuerdo a las variables disponibles en el mismo.	
7. Observa y selecciona opción de imprimir y/o exportar el reporte	8. Imprime y/o exporta el reporte si el usuario desea.

b. Refinamiento del diagrama de Casos de Uso

Diagrama de Casos de Uso para el Programa Sistema Nacional de Microfinanzas



El gráfico anterior muestra el diagrama de casos de uso del Programa Sistema Nacional de Microfinanzas el mismo que consta de los procesos fundamentales que se realizan para esta aplicación, dando un punto de vista más claro de cómo es que esta implementada tanto la

aplicación web como la de escritorio que como ya se ha mencionado anteriormente sirve para ingresar Posibles Beneficiarios.

Debemos mencionar que todos los diagramas se han realizado en una herramienta case Open Source denominada Star UML.

Diagrama de Actividades (calles)

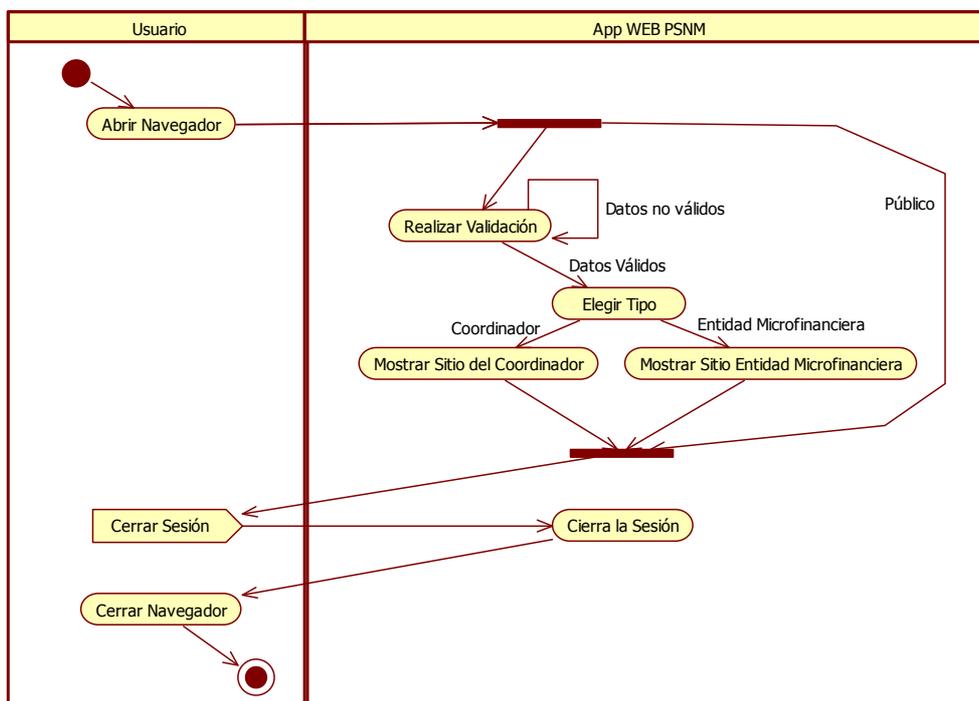
Como ya lo hemos dicho anteriormente, el sistema para el Programa Sistema Nacional de Microfinanzas se compone de dos subsistemas por decirlo así, y estos son los siguientes:

a. Aplicación WEB del Programa Sistema Nacional de Microfinanzas

Vamos a aclarar que en esta aplicación existen tres sitios:

1. Sitio del usuario Coordinador y Súper Usuario del PSNM

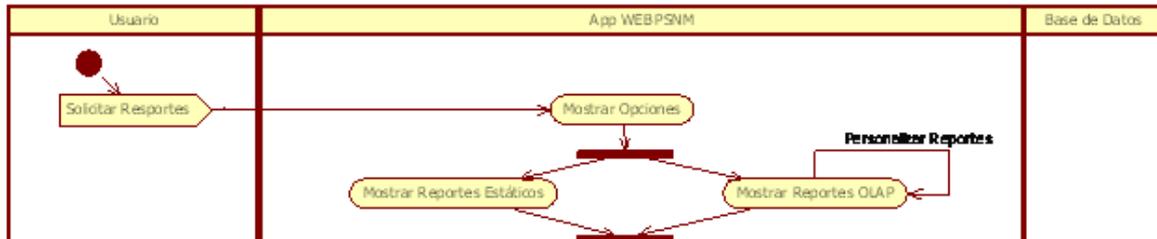
Para entrar a cada uno de estos sitios simplemente al abrir el navegador y por consiguiente este programa, la ventana de inicio nos llevará a cada uno de ellos de la siguiente forma:



A continuación vamos a mostrar un diagrama de actividades respecto de las acciones que se realizan en esta aplicación por cada uno de los sitios desarrollados ya que este programa es muy extenso:

1. Sitio del usuario Coordinador del PSNM

Para finalizar con la aplicación web, vamos a mostrar el diagrama de actividades con todas las acciones que se realizan en el sitio desarrollado para el usuario coordinador del PSNM.



Diseño de bajo Nivel

La fase de Diseño de Bajo Nivel tiene como objetivo fundamental crear una solución a nivel lógico para satisfacer los requisitos. Existen una gran cantidad de actividades que se realizan en esta sección, como son la definición de los diagramas de interacción, los de clases de diseño (en paralelo con los diagramas de interacción), el esquema de base de datos y por último se define la arquitectura del sistema.

Informes de Interfaz de Usuario

El objetivo de este apartado es mostrarle al usuario que la interfaz diseñada, cumple los siguientes objetivos:

- Efectividad lograda por medio de diseño de interfaces que permitan a los usuarios a acceder el sistema en una forma que sea congruente con sus necesidades individuales.
- Efectividad mostrada por medio de interfaces que aumenten la velocidad de la captura de datos y reduzca errores.
- Demostrar consideración al usuario diseñando interfaces adecuadas y que el sistema les proporcione la retroalimentación adecuada.
- Productividad mostrada por su adecuación a los principios ergonómicos establecidos en el diseño de interfaces y espacios de trabajo para los usuarios.

A continuación presentamos el diseño de la interfaz principal de la aplicación web:



Diagramas de Interacción

Los Diagramas de Interacción muestran el intercambio de mensajes entre instancias del modelo de clases para cumplir las post-condiciones establecidas en un contrato. Hay dos clases de Diagramas de Interacción: diagramas de Colaboración y diagramas de Secuencia.

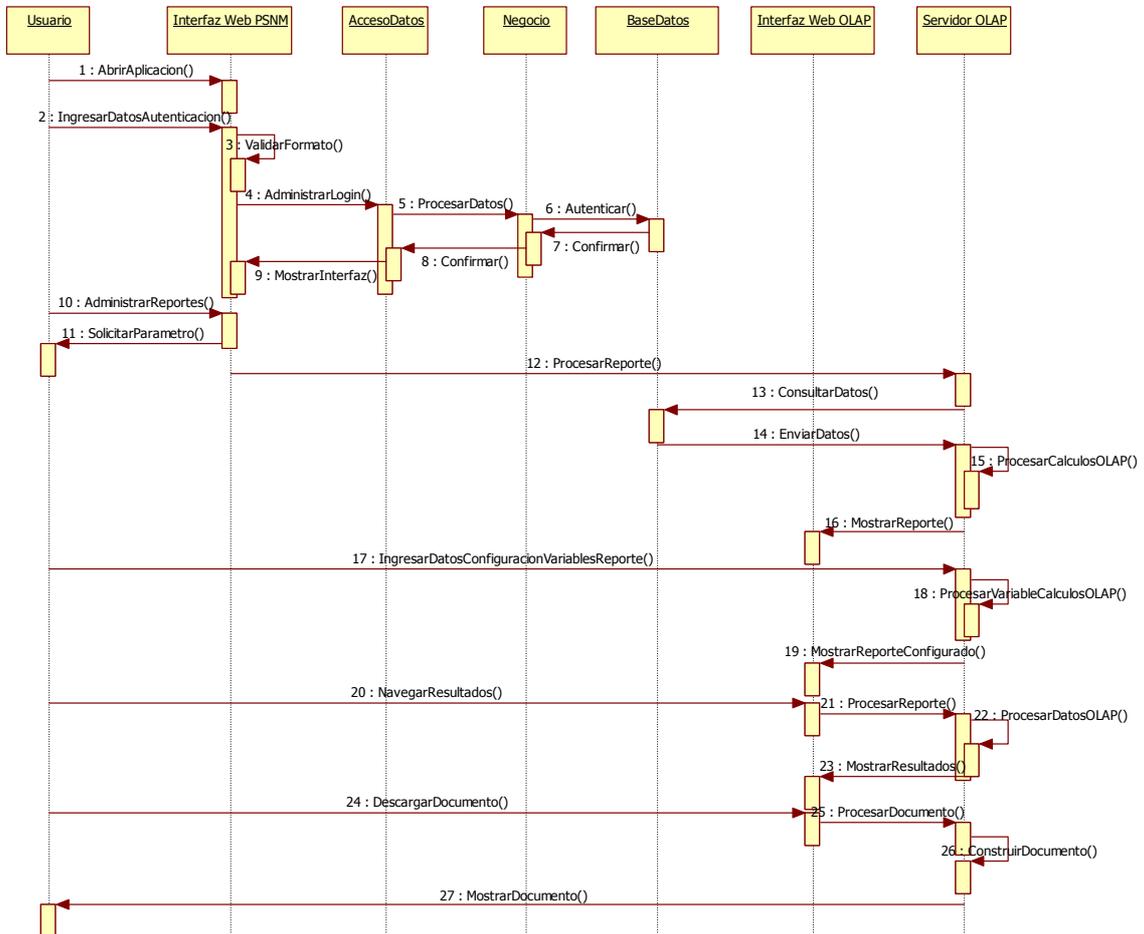
a. Diagrama de Secuencia

Destaca la ordenación temporal de los mensajes. El Diagrama de Secuencia es más adecuado para observar la perspectiva cronológica de las interacciones.

1. Aplicación WEB del Programa Sistema Nacional de Microfinanzas

A continuación mostramos un diagrama de secuencia por cada sitio desarrollado para el Programa Sistema Nacional de Microfinanzas:

i. Sitio del usuario Coordinador y Súper Usuario del PSNM



b. Diagrama de Colaboración

Destaca la organización estructural de los objetos que envían y reciben los mensajes. El Diagrama de Colaboración ofrece una mejor visión espacial mostrando los enlaces de comunicación entre objetos.

1. Aplicación WEB del Programa Sistema Nacional de Microfinanzas

A continuación vamos a mostrar un diagrama de colaboración respecto de las acciones que se realizan en esta aplicación:

i. Sitio del usuario Coordinador del PSNM

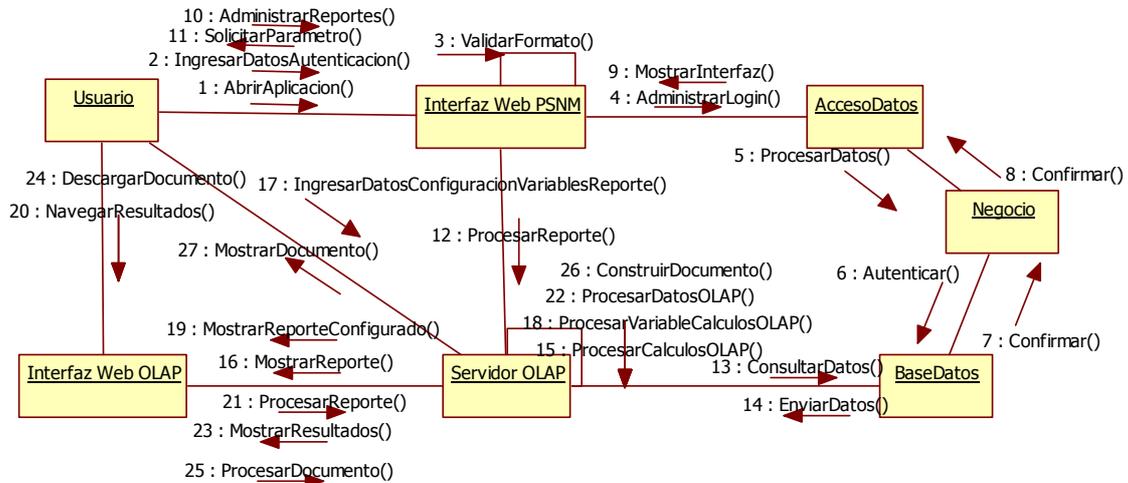


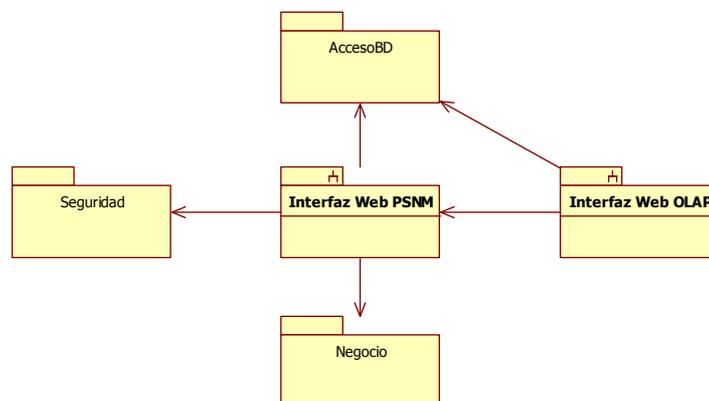
Diagrama de clases de diseño

El conjunto de todas las clases usadas, junto con sus relaciones, forma el Diagrama de Clases de Diseño. Un Diagrama de Clases de Diseño muestra la especificación para las clases software de una aplicación.

A diferencia del Modelo Conceptual, un Diagrama de Clases de Diseño muestra definiciones de entidades software más que conceptos del mundo real.

a. Diagrama de Clases para la aplicación web del PSNM

Este Diagrama está compuesto en su totalidad de paquetes, donde cada uno de los paquetes tiene sus respectivas clases, además debemos mencionar que el paquete “AccesoBD” y “Negocio” son paquetes que tienen subpaquetes y luego cada uno de estos subpaquetes tiene clases implementadas.



Las siguientes tablas muestran las clases de cada uno de los paquetes citados siendo en conjunto el diagrama general de clases para la aplicación web PSNM.

En las tablas que a continuación se muestran, vamos a presentar las clases de los paquetes que tienen sub-paquetes.

Paquete: ACCESOBD

PAQUETE	SUB-PAQUETE	DIAGRAMA DE CLASES
AccesoBD	ProgramasSociales	

Paquete: NEGOCIO

PAQUETE	SUB-PAQUETE	DIAGRAMA DE CLASES
Negocio	Administración	

b. Descripción de las Clases.

En este apartado se dará a conocer una descripción de las clases que se han utilizado, siendo las más importantes la aplicación web, puesto que para la de escritorio se ha manejado la misma lógica:

CAPA	PAQUETE	CLASES O ARCHIVO	DESCRIPCIÓN
Acceso a Base de Datos	AccesoBD.Master	Entidad.java	Clase hereda de AccesoBD.Entidad.java y además obtiene los datos de conexión con la base de datos de la Master del archivo de configuración que se encuentra dentro de Administracion.Configuracion. DataBase.properties
		UsuarioSistema (UsuarioPrograma).java	Logeo del usuario con el programa social
		Proceso.java	Obtención de los procesos del programa social [súper usuario]
		Permiso.java	Obtención de los permisos del usuario a los procesos del programa social
Administración Operacional	Administracion.Configuracion	DataBase.properties	Archivo de configuración o propiedad de las base de datos como son: usuario, contraseña y cadena de conexión de las bases de datos que intervendrán en el sistema.
		Logs.properties	Configuración de la ubicación donde se almacenaran las excepciones u errores generados por el sistema en cualquiera de sus capas
		TipoUsuario.properties	Configuraciones de los tipos de usuarios del sistema utilizada para la generación de la interfaz de los usuarios
		TipoMensaje.properties	Tipos de mensajes de operaciones
	Administracion.Metadata	TipoMensaje.java	Tipos de datos y mensajes que se utilizaran en el sistema

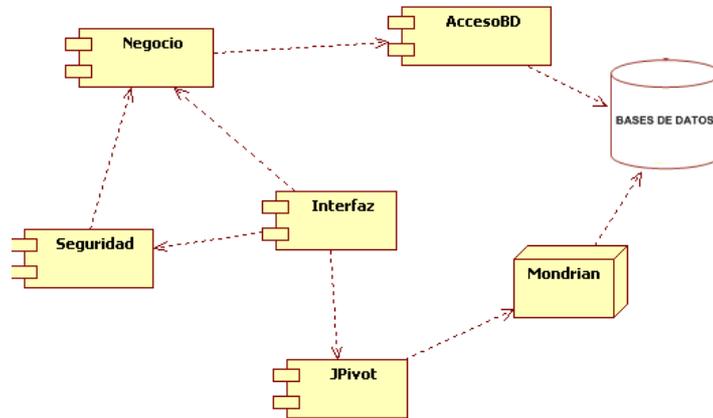
Diagramas de Implementación

a. Diagrama de Componentes

Estos diagramas describen los elementos físicos del sistema y sus relaciones. Los componentes representan todos los tipos de elementos software que entran en la fabricación de aplicaciones informáticas. A continuación mostramos los diferentes diagramas de componentes:

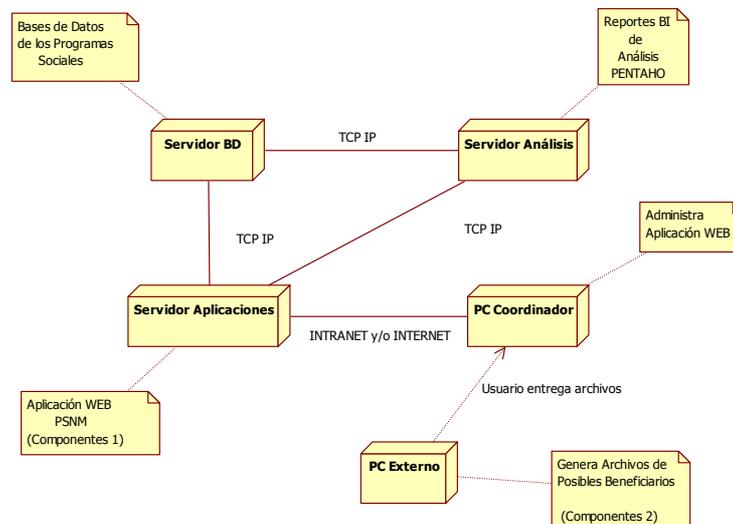
1. Componentes de la aplicación estadística del Programa Sistema Nacional de Microfinanzas

El gráfico los principales elementos y sus relaciones que intervienen en esta aplicación:



b. Diagrama de Despliegue

Los Diagramas de Despliegue muestran la disposición física de los distintos nodos que componen un sistema y el reparto de los componentes sobre dichos nodos.

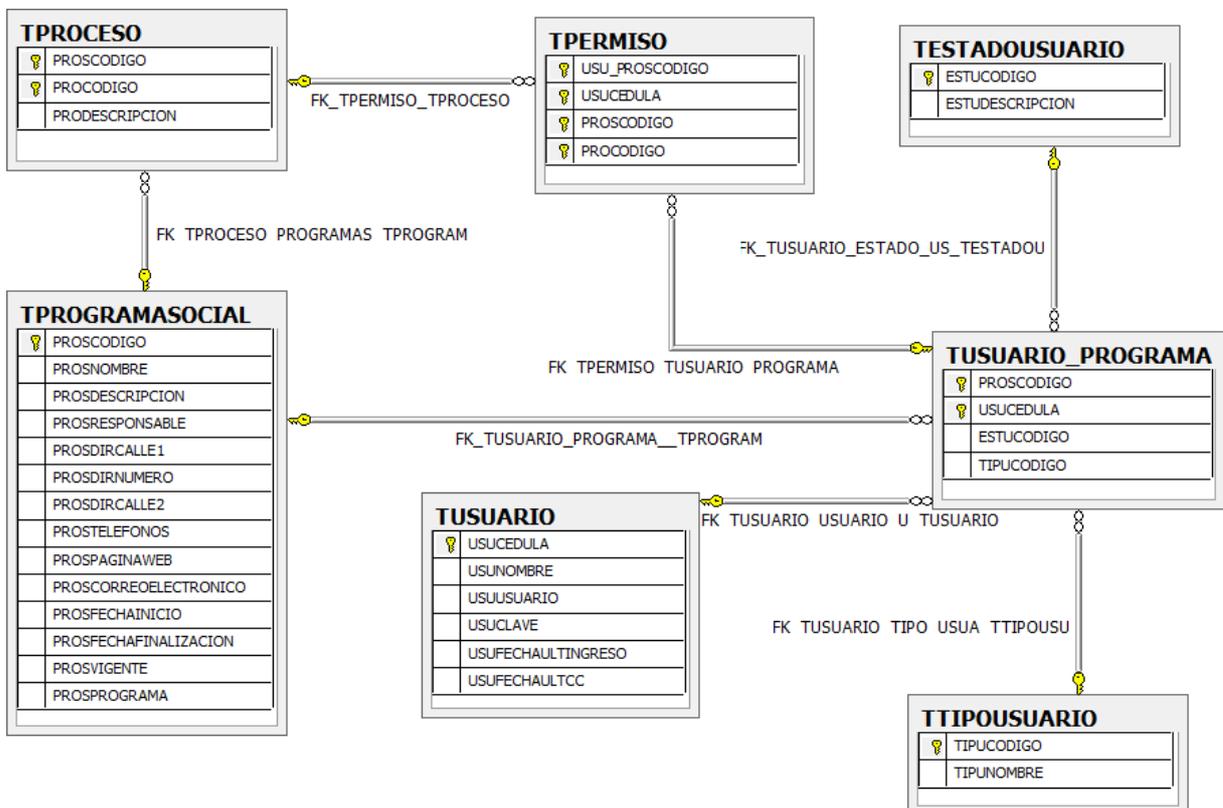


Diccionario de Datos

Rub_Master

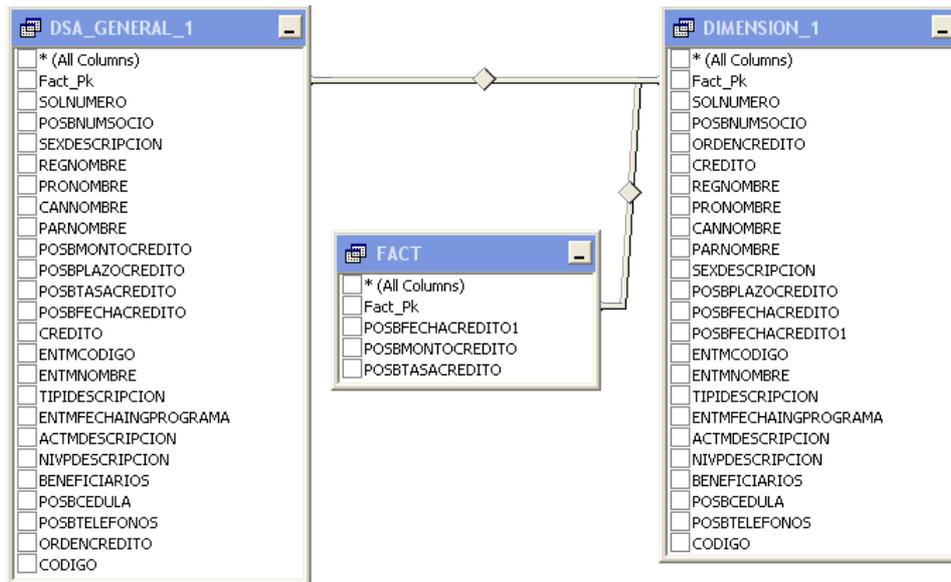
A continuación presentamos el modelo físico de la base de datos “Rub_Master”, la misma que está encargada de almacenar toda la información de los usuarios de los programas sociales, sus tipos, permisos, estados y procesos.

Es donde como objetivo fundamental hemos definido todos y cada uno de los esquemas de seguridad planteados para las aplicaciones. En esta sección se mostrará para Rub_Master solo el esquema del modelo físico de la base de datos, puesto que se dará mayor énfasis a la base de datos misma del Programa Sistema Nacional de Microfinanzas que se analizará luego de la presentación de este esquema:



Rub_Microfinanzas

A continuación en la siguiente página mostramos el esquema del modelo físico de la base de datos que hemos utilizado para implementar cada uno de los procesos que implican el Programa Sistema Nacional de Microfinanzas (**PSNM**), denominada “Rub_Microfinanzas”:



LISTADO DE VISTAS

NOMBRE	DESCRIPCIÓN
DSA_GENERAL_1	Información de las fuentes de datos de las tablas del Sistema.
DIMENSION_1	Información de dimensiones a analizar campos que no son cuantificables.
FACT	Vista de hechos en la se encuentran valores cuantificables además de crear nuevos campos calculados.

DESCRIPCIÓN DE COLUMNAS

VISTA 1: DSA_GENERAL_1

COLUMNA	TIPO DE DATOS	DESCRIPCIÓN
Fact_Pk	bigint	Código autogenerado
SOLNUMERO	int	Número de la Solicitud
POSBNUMSOCIO	int	Número de socio en la Entidad Microfinanciera
SEXDESCRIPCION	varchar(50)	Sexo
REGNOMBRE	varchar(50)	Nombre de la Región
PRONOMBRE	varchar(50)	Nombre de la Provincia

CANNOMBRE	varchar(50)	Nombre del Cantón
PARNOMBRE	varchar(50)	Nombre de la Parroquia
POSBMONTOCREDITO	money	Cantidad de crédito
POSBTASACREDITO	float	Tasa de Crédito
POSBPLAZOCREDITO	int	Plazo de Crédito
POSBFECHACREDITO	date	Fecha de Crédito
POSBCEDULA	varchar(10)	Cedula del beneficiario
ACTMDESCRIPCION	varchar(50)	Nombre de la Actividad Microempresarial
NIVPDESCRIPCION	varchar(50)	Nivel de Pobreza

VISTA 2: DIMENSION_1

COLUMNA	TIPO DE DATOS	DESCRIPCIÓN
Fact_Pk	bigint	Código autogenerated
SOLNUMERO	int	Número de la Solicitud
POSBNUMSOCIO	int	Número de socio en la Entidad Microfinanciera
SEXDESCRIPCION	varchar(50)	Sexo
REGNOMBRE	varchar(50)	Nombre de la Región
PRONOMBRE	varchar(50)	Nombre de la Provincia
CANNOMBRE	varchar(50)	Nombre del Cantón
PARNOMBRE	varchar(50)	Nombre de la Parroquia
POSBMONTOCREDITO	money	Cantidad de crédito
POSBTASACREDITO	float	Tasa de Crédito
POSBPLAZOCREDITO	int	Plazo de Crédito
POSBFECHACREDITO	date	Fecha de Crédito
POSBCEDULA	varchar(10)	Cedula del beneficiario
ACTMDESCRIPCION	varchar(50)	Nombre de la Actividad Microempresarial
NIVPDESCRIPCION	varchar(50)	Nivel de Pobreza

VISTA 2: FACT

COLUMNA	TIPO DE DATOS	DESCRIPCIÓN
Fact_Pk	bigint	Código autogenerated
POSBMONTOCREDITO	money	Cantidad de crédito

POSBTASACREDITO	float	Tasa de Crédito
POSBPLAZOCREDITO	int	Plazo de Crédito
POSBFECHACREDITO	date	Fecha de Crédito

TABLA 2: TBENEFICIARIO

COLUMNA	TIPO DE DATOS	DESCRIPCIÓN
BENNUMERO	bigint	Código de Beneficiario, autogenerado
SEXCODIGO	char(1)	Código del sexo del beneficiario, cuyos valores permitidos son: 1: Masculino 2: Femenino
BENCEDULA	char(10)	Número de cédula de un beneficiario
BENNOMBRECOMPLETO	varchar(100)	Nombres y apellidos completos de un beneficiario
BENFECHANACIMIENTO	datetime	Información de la fecha de nacimiento de un beneficiario