

Universidad Central de Venezuela

Facultad de Ciencias

Escuela de Computación



**DESARROLLO DE UNA SOLUCIÓN DE
INTELIGENCIA DE NEGOCIOS PARA UN
DEPARTAMENTO DE FINANZAS DENTRO DEL
SECTOR PETROQUÍMICO**

Trabajo Especial de Grado
Presentado ante la Ilustre
Universidad Central de Venezuela
Por los Bachilleres
Contreras Guevara Lorena Marian
Vera Lozada Sergio

Para optar al título de
Licenciado en Computación

Tutor:
Prof. Rangel Wuilfredo

Caracas, octubre de 2009

ACTA

Quienes suscriben, miembros del Jurado designado por el Consejo de Escuela de Computación, para examinar el Trabajo Especial de Grado presentado por los bachilleres Lorena Marian Contreras Guevara CI. 17.144.108 y Sergio Vera Lozada CI. 16.507.402, con el título **Desarrollo de una solución de Inteligencia de Negocios para un departamento de finanzas dentro del sector petroquímico** a los fines de optar al título de Licenciado en Computación, dejan constancia de lo siguiente:

Leído como fue, dicho trabajo por cada uno de los miembros del jurado, se fijó el día 30 de octubre del 2009 a las 9:00 am, para que sus autores lo defendieran en forma pública, lo que se hizo en el salón de post grado de la Escuela de Computación, mediante una presentación oral de su contenido, luego de lo cual respondió las preguntas formuladas. Finalizada la defensa pública del Trabajo Especial de Grado, el jurado decidió aprobarlo.

En fe de lo cual se levanta la presente Acta, en Caracas a los 30 días del mes de octubre del año 2009.

Prof. Wuilfredo Rangel

Tutor

Prof. Robinson Rivas

Jurado

Prof. Antonio Silva

Jurado

DEDICATORIA

Lorena Contreras

Quiero dedicarle este T.E.G en un principio a mis padres Alberto Contreras y Lucia Guevara por darme la vida y ser parte de mi formación en especial a mi madre por el apoyo incondicional que siempre ha tenido conmigo para ayudarme en cada paso que doy, por sus consejos y dedicación, le doy gracias a Dios por darme una madre tan única y especial como lo eres tú. A cada uno de mis hermanos Richard y Patricia por creer en mis cualidades y apoyarme para lograr cada meta que me propongo mil gracias por su apoyo y dedicación que han tenido conmigo. Junto con ellos quiero dedicársela a esas personitas que alegran mi vida y que bajo su inocencia se preocupan por la realización de mis trabajos ellas son mi sobrina Drahcir y Carlota Contreras. A mi abuelita que está en el cielo, porque mientras te estuve a mi lado fuiste mi apoyo, mi consentidora se que estarías muy contenta en este momento por mi logro, a toda mi familia este logro también es de ustedes, los adoro y amo.

También quiero dedicarle este trabajo a mi novio Carlos Cabrera que me ha apoyado desde el primer momento que llego a mi vida. Novio gracias porque alegras cada uno de mis días y me ayudas con tu amor, dedicación y apoyo te amo.

A cada una de las personas que están en el Laboratorio de SI donde compartimos por muchas horas y formamos un buen equipo de trabajo, en especial a Carlos, José y mi tutor Wuilfredo por su ayuda, explicaciones y su dedicación incondicional para poder sacar adelante este T.E.G. Gracias muchachos, les deseo el mayor de los éxitos. A mi compañero Sergio, por el apoyo durante la carrera, donde siempre conté con su ayuda incondicional. Por último a mis amigos que siempre estuvieron apoyándome, motivándome para sacar adelante todo lo que me propongo. Amigos son un gran apoyo, gracias por eso y por más.

Sergio Vera

Comienzo dedicando este trabajo a mis padres Roberto Vera y Landys del Valle y a mi hermano Roberto Vera con quienes me he comido las verdes y las maduras, ellos siempre han sido una guía para seguir luchando a pesar de tantas cosas. Con ellos siempre he contado para hablar, reír y muchas veces para llorar, los quiero mucho. A mi tía Carmen Elena Vera, gracias a su ayuda pude comenzar en este mundo de la universidad y aparte de eso ella me abrió las puertas de su casa para que mi viaje a la universidad no fuera tan pesado, muchas gracias tía te quiero. Tengo mucho que agradecer a mi compañera de tesis Lorena Contreras, que desde que nos conocimos a sido una muy buena amiga y que siempre me estuvo apoyando, ayudando y hasta halándome para seguir adelante, muchas gracias Lorena. Al igual que Mayerling Mendoza, con quien siempre pude contar y quien siempre me escuchó y apoyo en todo, te quiero mucho amiga. A alguien especial que ya no está conmigo, que siempre me aconsejó, me acompañó, me dijo siempre esa palabra de apoyo que yo requería, es mi abuela Sila Antonia Roja, con quien compartí mucho, quien me enseñó tanto y no hablo de los estudios sino de la vida que son conocimientos mucho más importantes, espero que aunque tarde ella pueda verme desde donde este, ya que siento que este título es de los dos. Te quiero y te extraño demasiado abuela. También quiero agradecerle a una persona especial, que en estos últimos años ha creído más en mí que yo mismo, mi novia Corina Goncalves quien llegó a mi vida en momentos de mucha turbulencia, me estabilizo con amor y paciencia; ha sido desde ese momento una luz que me muestra el camino a seguir, Te adoro. A toda mi familia que de una u otra manera me han apoyado. Y por último quiero agradecer a la gente que no creía en mí, que decían que yo no podría, que no me graduaría, ellos me dieron mucha fuerza para seguir y llegar hasta aquí y decirles que sus palabras aunque las dijeron por mal a mi me hicieron un bien, gracias.

RESUMEN

El presente Trabajo Especial de Grado (TEG) consiste en el desarrollo de una solución de inteligencia de negocios (BI, Business Intelligence por sus siglas en Inglés) para un departamento de finanzas del sector petroquímico, el cual maneja una gran cantidad de información y esta es utilizada para realizar los análisis, reportes e informes necesarios. En la actualidad estas tareas son realizadas de manera manual apoyándose en hojas de cálculos, lo que puede traer como consecuencia errores de manipulación en los datos y una gran pérdida de tiempo en todo este proceso. Nuestra solución de inteligencia de negocio disminuye los tiempos empleados para la extracción y carga de los datos así como para la generación de reportes, informes, gráfica, entre otros y aseguran la entrega de información exacta, útil y necesaria para apoyar la toma de decisiones.

El objetivo de este TEG es aprovechar las ventajas que nos ofrece una plataforma BI la cual es una arquitectura y colección de operaciones integradas y bases de datos que proporcionan a la comunidad empresarial el acceso fácil a los datos; es decir se aprovecha los proceso analítico (OLAP) el cual nos permite un mejor rendimiento para acceder a los datos multidimensionales proporcionando una base técnica para los análisis y cálculos requeridos, ayudando a que se tomen decisiones de manera rápida y precisa.

Para alcanzar dicho objetivo se utilizó una metodología de desarrollo de soluciones de inteligencia de negocio llamada "*The Business Dimensional Lifecycle*" propuesta por Ralph Kimball. Y como plataforma de desarrollo se utilizó una herramienta de código abierto (Open Source) llamada Pentaho BI la cual es una plataforma que ofrece funcionalidades como los son la integración de datos (ETL), reportes, análisis, herramientas OLAP, minería de datos, entre otras.

ÍNDICE

DEDICATORIA	3
RESUMEN	5
INTRODUCCIÓN	11
CAPÍTULO 1. PROBLEMA DE INVESTIGACIÓN	13
1.1 TÍTULO	13
1.2 SITUACIÓN ACTUAL	13
1.3 PLANTEAMIENTO DEL PROBLEMA	14
1.4 SOLUCIÓN PROPUESTA	15
1.5 OBJETIVOS	17
1.5.1 OBJETIVO GENERAL	17
1.5.2 OBJETIVOS ESPECÍFICOS	17
1.6 ALCANCE	17
1.7 IMPORTANCIA Y JUSTIFICACIÓN	18
CAPÍTULO 2. MARCO CONCEPTUAL	19
2.1 SISTEMAS DE INFORMACIÓN	19
2.1.1 TIPOS DE SISTEMAS EN CUANTO A LA ORGANIZACIÓN	20
2.2 ALMACÉN DE DATOS (DATAWAREHOUSE)	22
2.2.1 DEFINICIÓN	22
2.2.2 ARQUITECTURA DE UN ALMACÉN DE DATOS	22
2.2.3 BASES DE DATOS MULTIDIMENSIONALES- BDM.....	23
2.2.4 MODELO DE DATOS DIMENSIONAL	23
2.3 INTELIGENCIA DE NEGOCIO (BI, BUSINESS INTELLIGENCE POR SUS SIGLAS EN INGLÉS)	27
2.3.1 DEFINICIÓN	27
2.3.2 CARACTERÍSTICAS.....	28
2.3.3 COMPONENTES O ARQUITECTURA	28
2.3.4 VENTAJAS	30
2.4 PLATAFORMA DE INTELIGENCIA DE NEGOCIO PENTAHO	30
2.4.1 DEFINICIÓN	30
2.4.2 CARACTERÍSTICAS DE LA PLATAFORMA PENTAHO BI	31
2.4.3 ARQUITECTURA.....	34
2.4.4 COMPONENTES	37
2.5 METODOLOGÍA DE RALPH KIMBALL (THE BUSINESS DIMENSIONAL LIFECYCLE)	41
2.5.1 PLANIFICACIÓN DEL PROYECTO	41
2.5.2 DEFINICIÓN DE LOS REQUERIMIENTOS DEL NEGOCIO	42
2.5.3 DISEÑO TÉCNICO DE LA ARQUITECTURA	43
2.5.4 SELECCIÓN DE LAS HERRAMIENTAS	43
2.5.5 MODELO DIMENSIONAL	43

2.5.6 DISEÑO FÍSICO.....	44
2.5.7 DISEÑO DE PROCESOS ETL.....	44
2.5.8 ESPECIFICACIÓN DE LA APLICACIÓN DE USUARIO FINAL.....	44
2.5.9 DESARROLLO DEL LA APLICACIÓN DE USUARIO FINAL.....	45
2.5.10 DESPLIEGUE.....	45
CAPÍTULO 3. MARCO APLICATIVO.....	46
3.1 METODOLOGÍA DE DESARROLLO.....	46
3.1.1 PLANIFICACIÓN DEL PROYECTO.....	47
3.1.2 DEFINICIÓN DE LOS REQUERIMIENTOS DEL NEGOCIO.....	48
3.1.2.1 FICHA TÉCNICA DEL REQUERIMIENTO A.....	49
3.1.2.2 FICHA TÉCNICA DEL REQUERIMIENTO B.....	49
3.1.2.3 FICHA TÉCNICA DEL REQUERIMIENTO C.....	50
3.1.2.4 FICHA TÉCNICA DEL REQUERIMIENTO D.....	51
3.1.2.5 FICHA TÉCNICA DEL REQUERIMIENTO E.....	51
3.1.2.6 FICHA TÉCNICA DEL REQUERIMIENTO F.....	52
3.1.2.7 FICHA TÉCNICA DEL REQUERIMIENTO G.....	53
3.1.2.8 FICHA TÉCNICA DEL REQUERIMIENTO H.....	54
3.1.2.9 FICHA TÉCNICA DEL REQUERIMIENTO I.....	55
3.1.2.10 FICHA TÉCNICA DEL REQUERIMIENTO J.....	56
3.1.2.11 FICHA TÉCNICA DEL REQUERIMIENTO K.....	57
3.1.3 SELECCIÓN DE LAS HERRAMIENTAS.....	57
3.1.4 DISEÑO TÉCNICO DE LA ARQUITECTURA.....	58
3.1.4.1 COMPONENTES DE LA ARQUITECTURA DEL SISTEMA.....	58
3.1.4.2 INTEGRACIÓN DE LOS COMPONENTES DEL SISTEMA.....	68
3.1.5 MODELO DIMENSIONAL.....	69
3.1.6 DISEÑO FÍSICO.....	71
3.1.6.1 DIMENSIÓN: DIM_PRODUCTO.....	71
3.1.6.2 DIMENSIÓN: REGION.....	72
3.1.6.3 DIMENSIÓN: CORPORACION.....	72
3.1.6.4 DIMENSIÓN: DIM_BALANCESHEET.....	73
3.1.6.5 DIMENSIÓN: DIM_PRODUCION.....	74
3.1.6.6 DIMENSIÓN: DIM_UNIT.....	75
3.1.6.7 DIMENSIÓN: DIM_INSUMO.....	75
3.1.6.8 DIMENSIÓN: DIM_CLOSING.....	76
3.1.6.9 DIMENSIÓN: DIM_FECHA.....	77
3.1.6.10 DIMENSIÓN: DIM_OPERACION.....	78
3.1.6.11 TABLA DE HECHOS: FACT_TABLE_NETBACK.....	78
3.1.6.12 TABLA DE HECHOS: FACT_TABLE_MENTOR.....	79
3.1.7 DISEÑO DE PROCESOS ETL.....	79
3.1.8 ESPECIFICACIÓN DE LA APLICACIÓN DE USUARIO FINAL.....	85
3.1.9 DESARROLLO DE LA APLICACIÓN DE USUARIO FINAL.....	86
3.1.10 DESPLIEGUE.....	112
CAPÍTULO 4. CONCLUSIONES, RECOMENDACIONES Y LIMITACIONES.....	113
REFERENCIAS.....	116

ÍNDICE DE FIGURAS

Figura 1. Sistemas de Información.....	19
Figura 2. Tipos de Sistemas de Información.....	21
Figura 3. Cubo representativo de dimensiones.	24
Figura 4. Plataforma de Business Intelligence Pentaho usando Dashboards	33
Figura 5. Arquitectura de Pentaho Data Integration.....	34
Figura 6. Plataforma BI Pentaho	35
Figura 7. Pentaho Open BI Suite (Componentes)	37
Figura 8. The Data Warehouse Lifecycle Toolkit. (Ciclo de vida multidimensional del negocio)	41
Figura 9. Ejemplo de una consulta sencilla del lenguaje MDX.	59
Figura 10. JPivot Navegador.....	60
Figura 11. Pentaho Reporting Designer	62
Figura 12. Shema Workbeach.....	63
Figura 13. Estructura del archivo .modrian.xml.....	64
Figura 14. Estructura del archivo .xaction	65
Figura 15. Interfaz gráfica de Kettle	66
Figura 16. Modelo del diseño multidimensional con sus dimensiones, tabla de hechos y relaciones.	70
Figura 17. Primer subproceso de ETL para la fact_table.....	82
Figura 18. Segundo subproceso de ETL para la fact_table.....	83
Figura 19. Tercer subproceso de ETL para la fact_table.....	84
Figura 20. Proceso general de ETL para la fact_table.....	84
Figura 21. Template de los Reportes.....	85
Figura 22. Visualización de un cubo en JPivot.	86
Figura 23. Query MDX generado por JPivot	87
Figura 24. Analítico de Sales.	87
Figura 25. Analítico de Operation.....	88
Figura 26. Analítico de Balance Sheet	88
Figura 27. Analítico de FSA.	89
Figura 28. Analítico de Inventory.	90
Figura 29. Analítico de Profit and Loss (Total).	90
Figura 30. Analítico de OPEX Production.....	91
Figura 31. Analítico de Analítico de OPEX Administrative.....	92
Figura 32. Sales Region Production MT	93
Figura 33. Sales by Off taker Production MT	93
Figura 34. Analítico de Unit Cost	94

Figura 35. Analítico de OPEX Plant	95
Figura 36. Diseño de un reporte en Report Designer	96
Figura 37. Interfaz del servidor Web de Pentaho.....	97
Figura 38. Interfaz de la ubicación de la solución en Pentaho.	98
Figura 39. Parámetros de los reportes.	99
Figura 40. Despliegue de los parámetros de los reportes.	99
Figura 42. Reporte de Operations.	100
Figura 44. Reporte de Sales by Region Production Volumen MT (*000).....	101
Figura 45. Reporte de Sales by Region Net Back US\$/MT.	102
Figura 46. Reporte de Sales by off taker Production Volumen MT (*000).....	103
Figura 47. Reporte de Sales by off taker Net Back US\$/MT.	104
Figura 48. Reporte de Profit and Loss (OPEX Production)	105
Figura 49. Reporte de Profit and Loss (OPEX PLANT).....	106
Figura 50. Reporte de Profit and Loss (OPEX Administrative).....	107
Figura 51. Reporte de Unit Cost.....	108
Figura 52. Reporte de Balance Sheet	109

ÍNDICE DE TABLAS

Tabla 1. Plan de trabajo.....	47
Tabla 2. Ficha técnica del requerimiento A.....	49
Tabla 3. Ficha técnica del requerimiento B.....	49
Tabla 4. Ficha técnica del requerimiento C.....	50
Tabla 5. Ficha técnica del requerimiento D.....	51
Tabla 6. Ficha técnica del requerimiento E.....	51
Tabla 7. Ficha técnica del requerimiento F.....	52
Tabla 8. Ficha técnica del requerimiento G.....	54
Tabla 9. Ficha técnica del requerimiento H.....	54
Tabla 10. Ficha técnica del requerimiento I.....	55
Tabla 11. Ficha técnica del requerimiento J.....	56
Tabla 12. Ficha técnica del requerimiento K.....	57
Tabla 13. Integración de los componentes del sistema.....	68
Tabla 14. Dimensión dim_producto.....	71
Tabla 15. Dimensión region.....	72
Tabla 16. Dimensión corporacion.....	72
Tabla 17. Dimensión dim_balancesheet.....	73

Tabla 18. Dimension dim_produccion	74
Tabla 19. Dimensión dim_unit.....	75
Tabla 20. Dimensión dim_insumo	76
Tabla 21. Dimensión dim_closing.....	76
Tabla 22. Dimensión dim_fecha	77
Tabla 23. Dimensión dim_operacion	78
Tabla 24. Tabla de hechos fact_table_netback.....	78
Tabla 25. Tabla de hechos fact_table_metor	79

INTRODUCCIÓN

En la actualidad los sistemas de información son muy utilizados para obtener una mayor cantidad de datos, mejorar el flujo de información y poder tomar decisiones que produzcan una mejora en la rentabilidad y también una ventaja competitiva.

También es cierto que cada vez se genera una mayor cantidad de información, la cual luego debe ser tratada, organizada y llevada a tablas, gráficos o informes para poder ser analizados, pero todo este trabajo es realizado en su mayoría de forma manual apoyándose en hojas de cálculo lo que implica que el proceso es lento y puede tener muchos errores debido a la manipulación de los datos.

Por estas razones es importante contar con sistemas que permitan analizar todos esos datos de forma rápida y también visualizarlas de manera que se puedan tomar decisiones más rápidas, en el momento exacto y con la mayor cantidad de data posible. Una de las herramientas que permite analizar una gran cantidad de información y ofrece la generación los reportes, tablas, graficas, entre otros, son las de inteligencia de negocio (BI, Business Intelligence por sus siglas en inglés).

En este momento existen muchas herramientas de inteligencia de negocio que te ofrecen diversas funciones y características que se deben estudiar para elegir la que ofrezca mayores beneficios. Las herramientas basadas en la filosofía de software libre ofrecen las funciones generales de un BI privado y a su vez tienen las siguientes ventajas particulares como los son, todos los usuarios de software libre pueden ver los códigos fuentes para ser estudiados y adaptados a sus necesidades, al hacer mejoras y hacerlas públicas toda la comunidad se ve beneficiada, el software puede ser utilizada por cualquier organización que lo requieran, dado que la licencia del mismo es gratuita.

La plataforma BI escogida ofrece herramientas necesarias que satisfacen la inteligencia de negocios en las organizaciones, aumentando así su ventaja competitiva, numerosas organizaciones que han adoptado la aplicación BI Open Source Pentaho, se han abierto paso en el mercado con relativa facilidad, dando respuestas concretas a las carencias con respecto a la presentación de la información.

Con el propósito de alcanzar los objetivos planteados, este documento se estructura de la siguiente manera:

En el Capítulo I: Se inicia describiendo el problema de investigación y la situación actual especificando así las necesidades y luego se presentan los objetivos tanto específico como general, el alcance y la justificación.

El Capítulo II se refiere al marco conceptual aquí se le muestra al lector los conocimientos fundamentales para la elaboración del trabajo. Además se especifican las definiciones, características, procesos y relaciones que sirven de apoyo a las bases teóricas del desarrollo de una solución de Inteligencia de Negocios para un departamento de finanzas dentro del sector petroquímico.

El Capítulo III comprende el marco aplicativo, se presenta la metodología de desarrollo propuesta por Ralph Kimball en su libro Data Warehouse Lifecycle Toolkit adaptado a nuestro trabajo especial de grado.

El Capítulo IV contiene las conclusiones, donde se menciona que se logró luego del desarrollo de la solución de inteligencia de negocio.

CAPÍTULO 1. Problema de investigación

En este capítulo se detalla el contexto del problema, describiendo la situación actual de cómo son manejados sus procesos, especificado así el problema a resolver y la solución adecuada que se adapta a las necesidades de la organización. También se detalla el objetivo general y los objetivos específicos que determinaran así su alcance, importancia y justificación del problema.

1.1 Título

Desarrollo de una solución de inteligencia de negocios para un departamento de finanzas dentro del sector petroquímico.

1.2 Situación Actual

En los departamentos de finanzas se encargan de controlar e informar la ejecución de los recursos económicos financieros de la empresa lo cual implica el manejo de un gran volumen de información la cual es manipulada y procesada por una variedad de hojas de cálculo y también son procesadas manualmente usando fórmulas, gráficos para poder analizar sus métricas.

Luego esta información es transcrita o llevada a herramientas que permitan su publicación o presentación, todo esto trae como consecuencia una serie de limitaciones asociadas como lo es:

- Una fuerte inversión en tiempo de trabajo para la realización de un informe completo de las distintas métricas que se desea ver en un departamento de finanzas.
- Posibilidad de ocurrencia de errores.

- Mucho retardo de las operaciones realizadas manualmente, lo que conlleva que cuando se tenga disponible la información para tomar decisiones importantes sea ya muy tarde y se tomen decisiones no adecuadas para el momento.

Los procesos de extracción, transformación y carga de los datos se realizan todos los meses y los procesos de realización y publicación de reportes son realizados tres veces al año, lo cual hace que se inviertan gran cantidad de recursos en la generación de cada uno de los informes con las métricas que evalúa una empresa de finanzas.

1.3 Planteamiento del Problema

En la actualidad los departamentos de finanzas poseen sistemas que le permiten obtener datos de todo los movimientos financieros de la empresa pero a la hora de visualizarlos para tomar decisiones con ellos se pierde mucho tiempo ya que tienen que ser extraídos de diferentes archivos Excel y seleccionar ciertas variables para realizar así los documentos, las tablas y los informes de manera manual para ser presentados y discutidos.

Para la toma de decisiones este departamento requiere de una serie de tablas y gráficos que son creados actualmente de forma manual copiando datos de hojas de cálculo para ser colocadas en presentaciones de power point, lo cual generaba errores de transcripción a la hora de seleccionar las variables de los archivos Excel para realizar la representación en power point. Las tablas y gráficas que se requirieron son:

- Summary
- Operations
- Sales
- Sales by Region Production MT y Sales by Region Net Back (US\$)

- Sales by Off taker MT y Sales by Off Taker Net Back (US\$)
- Methanol Inventory
- Profit and Loss
- Profit and Loss (Total)
- Unit Cost
- Balance Sheet
- FSA

Para la realización de todas estas tablas y gráficos se invierten más de una semana de trabajo para que los mismos estén terminados y listos para ser mostrados en las juntas con los directivos.

Es por esto que en este trabajo se realizará una solución de inteligencia de negocio para el departamento de finanzas que le permita mejorar los tiempos de respuesta a la hora de generar los informes que requieran para visualizar y analizar la información del mismo para que de esta manera puedan tomar sus decisiones.

1.4 Solución Propuesta

Sabiendo la problemática y las necesidades que presenta el departamento de finanzas en cuanto a la generación de reportes en un menor tiempo y con datos precisos, se necesita realizar los diferentes analíticos y reportes que ellos necesitan para evaluar sus finanzas, partimos de que ya el departamento posee una estructura para estos reportes y la forma en que muestran sus gráficas a la hora de estudiar los datos.

Tomando en cuenta lo expuesto anteriormente se propondrá realizar un conjunto de estrategias y herramientas que satisfagan sus necesidades a la hora de trabajar la data para poder generar sus reportes y gráficas.

Para satisfacer los requerimientos del departamento se propone realizar una solución de inteligencia de negocio, donde se realizará el modelo multidimensional orientado a responder todos sus requerimientos, al igual que se desarrollarán los distintos ETL para poblar el almacén de datos de donde se obtendrá la información que se necesita para generar los distintos reportes que serán utilizados por el departamento apoyándolo así a la toma de decisiones.

Utilizaremos una serie de herramientas que son necesarios para implementar las soluciones propuestas, entre ellas están:

- MySQL 5.1
- Kettle
- Mondrian
- Esquema Workbench
- JPivot
- Reporting Designer (Diseñador de reportes)

Partiendo de estas herramientas obtendremos como resultado la generación de un conjunto de reportes siguiendo la estructura que ellos actualmente utilizan cuyas medidas serán Budget, Real y Var, estos reportes son:

- Summary
- Operations
- Sales
- Sales by Region Production MT y Sales by Region Net Back (US\$)
- Sales by Off taker MT y Sales by Off Taker Net Back (US\$)
- Methanol Inventory
- Profit and Loss
- Profit and Loss (Total)
- Unit Cost
- Balance Sheet

- FSA

1.5 Objetivos

Como objetivo general y específicos nos planteamos:

1.5.1 Objetivo General

Desarrollar una solución de Inteligencia de Negocio con la plataforma Business Intelligence Pentaho para el apoyo de la toma de decisiones desarrollando el caso de estudio de un departamento de finanzas dentro del sector petroquímico.

1.5.2 Objetivos Específicos

- Estudiar el modelo de negocio de un departamento de finanzas.
- Definición de los requerimientos para la presentación de reportes.
- Análisis y diseño del repositorio multidimensional.
- Diseño y construcción del proceso de extracción, transformación y carga (ETC) para la unificación de los datos de un departamento de finanzas.
- Desarrollo de la solución de inteligencia de negocio en la plataforma de inteligencia de negocio Pentaho.
- Despliegue de la solución de inteligencia de negocio.
- Realizar pruebas que permitan validar el correcto funcionamiento de la solución.

1.6 Alcance

Para el alcance de la solución de inteligencia negocio se tiene:

- Elaboración del modelo multidimensional, sus métricas y las dimensiones asociadas a un departamento de finanzas dentro del sector petroquímico.

- Se realizarán los procesos de extracción, transformación y cargar de los distintos archivos Excel para el almacén de datos.
- Se realizarán un conjunto de reportes parametrizados por fecha con salidas al servidor OLAP, Web, PDF o Excel, los cuales cumplen con las métricas requeridas de un departamento de finanzas dentro del sector petroquímico.

1.7 Importancia y Justificación

En la actualidad en general los departamentos de finanzas se enfrentan a la difícil tarea de realizar sus reportes de forma manual, lo que trae como consecuencia una demora considerable en este proceso y también se pueden producir errores en la manipulación de los datos, lo que podría reflejarse en tomar decisiones con información que no es real.

Por lo antes mencionado toma una gran importancia realizar una plataforma de inteligencia de negocio que agilice la creación de reportes, gráficas e informes en el área del departamento de finanza para ahorrar tiempo es estos procesos, así como evitar errores en la manipulación de los datos y con esto poder apoyar de manera adecuada a la toma de decisiones con datos reales y en poco tiempo.

CAPÍTULO 2. Marco Conceptual

2.1 Sistemas de Información

Saroka (2002), define sistemas de información “*es un conjunto de recursos humanos, materiales, financieros, tecnológicos, normativos y metodológicos, organizados para brindar, a quienes operan y a quienes adoptan decisiones en una organización, la información que requieren para desarrollar sus respectivas funciones*”.

Por otro lado Laudon K. y Laudon J. (2004), definen sistemas de información como un “*conjunto de componentes interrelacionados que recolectan (o recuperan), procesan, almacenan y distribuyen información para apoyar la toma de decisiones y el control de una organización*”.

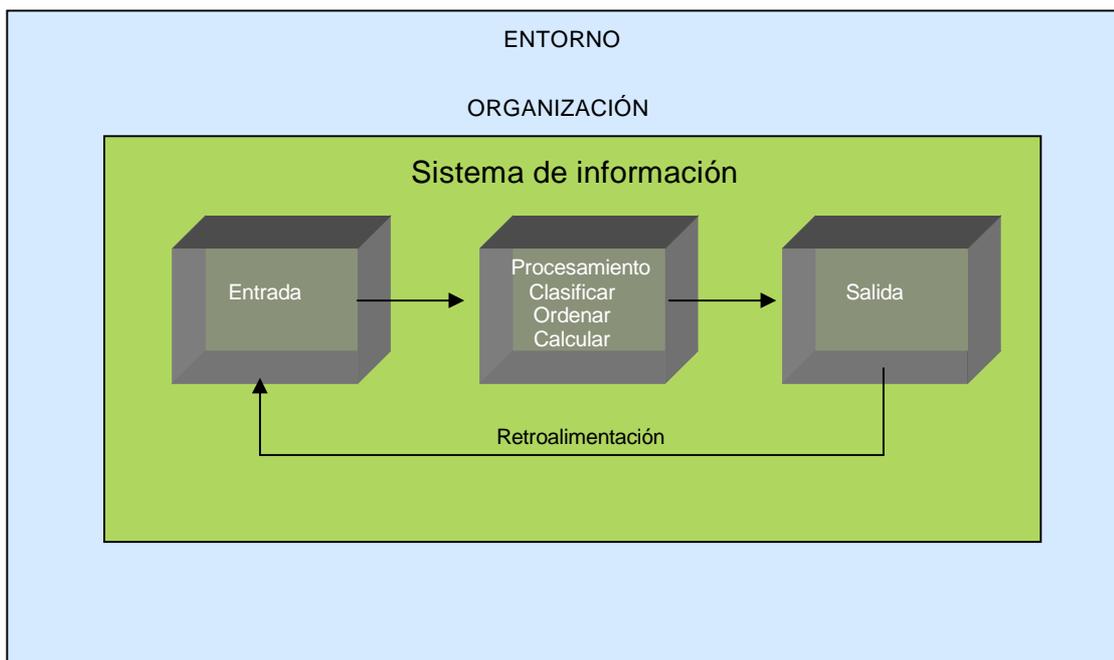


Figura 1. Sistemas de Información
Fuente. Laudon K. y Laudon J. (2004)

2.1.1 Tipos de sistemas en cuanto a la organización

Laudon K. y Laudon J. (2004), consideran que los sistemas de información se pueden dividir en cuanto a los niveles de la organización. (En la figura 2 se muestran gráficamente los tipos de sistemas de información que se encuentran en una organización). A continuación se describen los niveles.

- *Sistemas a nivel operativo.* Este sistema apoya a los gerentes operativos en el seguimiento de las actividades y transacciones elementales de la organización como ventas, ingresos, depósito en efectivo, nómina, decisiones de crédito y flujos de materiales. En este nivel se cuenta con los sistemas de procesamiento de transacciones (TPS, por sus siglas en inglés) los cuales son sistemas computarizados que efectúan y registran transacciones diarias de los negocios, tales como nómina e inventario.
- *Sistemas a nivel del conocimiento.* Este tipo de sistemas apoyan la integración de nuevos conocimientos del negocio, ayudando a controlar el flujo del trabajo dentro del negocio. En este nivel se cuenta con los sistemas de trabajo del conocimiento (KWS, por sus siglas en inglés) los cuales auxilian a los trabajadores profesionales tales como científicos, ingenieros y doctores, ayudando a crear nuevos conocimientos que contribuyan a la organización o a toda la sociedad y los sistemas de oficina auxilian principalmente a los trabajadores de datos quienes no crean un nuevo conocimiento si no que usan la información para analizarla y transformar datos.
- *Sistemas a nivel administrativo.* Estos sistemas de información apoyan las actividades de supervisión, control y toma de decisiones. Este nivel de sistema es apoyado por los sistemas de información gerencial (MIS, por sus siglas en inglés) los cuales son sistemas que dan soporte a las tareas organizacionales proporcionando informes semanales, mensuales y anuales, que ayuden a los gerentes a la toma de decisiones, el control y planeación a nivel administrativo. Otro de los sistemas que dan servicio a nivel administrativo

son los sistemas de apoyo a la toma de decisiones (DSS, por sus siglas en inglés), estos sistemas ayudan a los gerentes a tomar decisiones que son exclusivas, cambiantes y no fáciles de anticipar.

- *Sistema a nivel estratégico.* Estos sistemas apoyan a las actividades de planificación a largo plazo de las organizaciones ayudan a los directores de las organizaciones enfrentar y resolver aspectos estratégicos; dando así los cambios necesarios del entorno externo de la organización. Los directores para la toma de decisiones se ayudan con los sistemas de apoyo a ejecutivos (ESS, por sus siglas en inglés), donde estos sistemas dan soporte a las decisiones no rutinarias que requieren juicio, evaluación y comprensión, es decir, dan soporte a la toma de decisiones con gráficos y comunicaciones avanzadas.

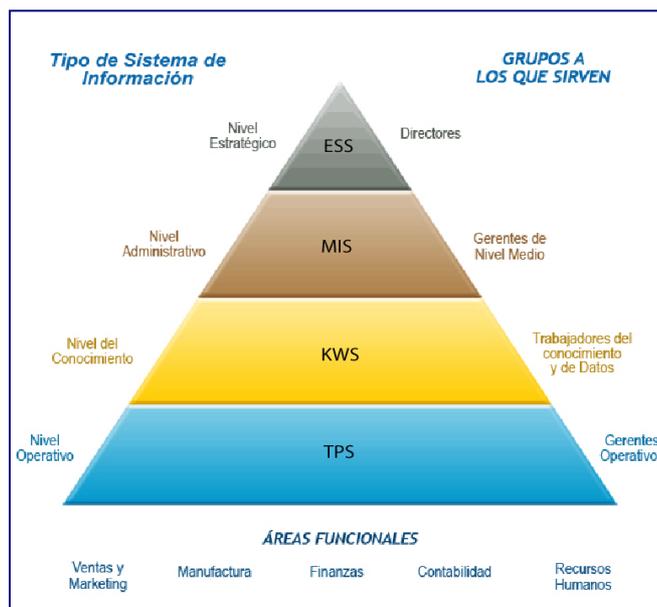


Figura 2. Tipos de Sistemas de Información
Fuente. Laudon K. y Laudon J. (2004)

2.2 Almacén de Datos (Datawarehouse)

2.2.1 Definición

Almacén de Datos, es definido como “*una colección de datos integrados, no volátil y variable en el tiempo que ayuda a la toma de decisiones de la empresa u organización*”. (Cipher, 2008)

Se encarga de guardar datos históricos, es decir datos a través del tiempo por lo que obviamente necesita grandes recursos de almacenamiento para su funcionamiento. Principalmente tiene solo consultas debido a que por lo general están conectadas con sistemas que extraen la información para después mostrarla a un gerente o un analista para tomar una decisión o acción.

Es decir es una base de datos centralizada que almacena una gran cantidad de información, que se encarga de guardar los datos históricos y que principalmente se usan para realizar consultas, las cuales permiten tomar decisiones de importancia para la organización.

2.2.2 Arquitectura de un Almacén de Datos

La arquitectura de un almacén de datos se encuentra dividida en: según Hernández J (s.f)

- Herramientas de Transformación/Propagación para cargar el Almacén de Datos.
- Sistema ETL (Extracción, Transformación y Carga): Realiza las funciones de extracción de las fuentes de datos (transaccionales o externas), transformación (limpieza, consolidación, etc.) y la carga del Almacén de Datos, realizando:
 - *Extracción de los datos.*
 - *Filtrado de los datos: limpieza, consolidación, etc.*

- *Carga inicial del almacén: ordenación, agregación, etc.*
- *Actualización del almacén: operación periódica que propaga los cambios de las fuentes de datos al Almacén de datos.*
- El servidor del almacén de datos: Repositorio propio de datos que proporcionan características de Almacenamiento, Gestión y Recuperación.
- Interfaces y Gestores de Consulta que permiten acceder a los datos y sobre ellos se conectan herramientas más sofisticadas (OLAP, minería de datos) y proporcionan Consultas, Informes, Análisis y Extracción de conocimiento (Data Mining)
- Herramientas para manejar la información acerca del almacén de datos (metadata)
- Sistemas de Integridad y Seguridad: se encargan de un mantenimiento global, copias de seguridad, etc.

2.2.3 Bases de datos Multidimensionales- BDM

Son diseñadas para los sistemas de soporte de decisiones en la cual los datos tienen una estructura matricial (multidimensional) para su almacenamiento. Este tipo de organización admite consultas más complejas. (Cipher, 2008).

Estas BD permiten ver la información de manera precisa ya que puede utilizar varias tablas para ir filtrando la data o proporcionar una salida más precisa y más exacta de lo que se quiere ver en los movimientos de una organización.

2.2.4 Modelo de Datos Dimensional

En el modelo de datos dimensional se tiene:

2.2.4.1 Dimensiones

Son cada uno de los ejes en un espacio multidimensional. (Kimball M. y Ross R, 2002). Esta información es relativamente estática en el tiempo estas representan formas de ver la información, en el modelado dimensional se conoce como tablas dimensiones.

Como se puede observar en la Figura 3, se tiene un cubo que representa un espacio multidimensional y en cada uno de los ejes o aristas de este cubo se encuentran las dimensiones donde cada una de estas se pueden interceptar.

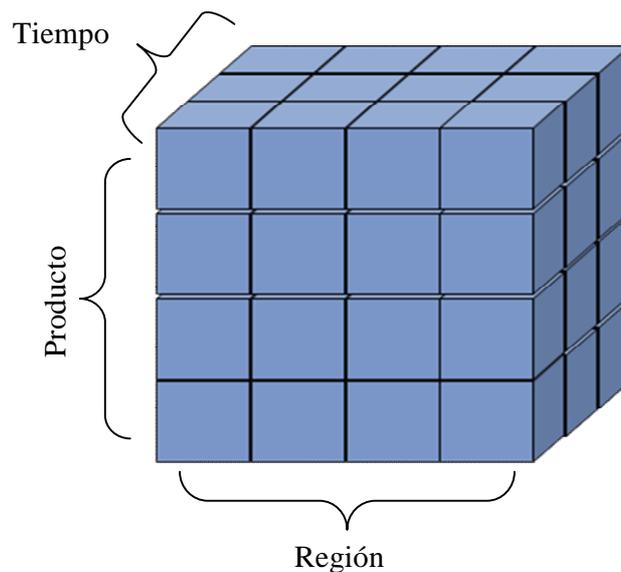


Figura 3. Cubo representativo de dimensiones.

Fuente. (Wolff C., 2002). Modificado

En este caso tenemos las dimensiones: Producto, Región y Tiempo. Dependiendo desde el punto de vista que se quiera el dato, podemos obtenerlo cruzando estas tres dimensiones y tener los datos que nos permitan responder algunas de las preguntas.

Las dimensiones permiten ver y analizar la información desde diferentes formas y al cruzar varias dimensiones funcionan como filtro para obtener la información precisa que se requiere para ser utilizada.

2.2.4.2 Atributos de Dimensión

Son los atributos de tipo cualitativo que suministra el contexto en el que se obtienen las medidas. Ejemplos: ciudad, estado, país, continente, etc. (Kimball M. y Ross R, 2002)

Estos nos permiten comprender con claridad la data que se está mostrando o consultando.

2.2.4.3 Jerarquías

Una jerarquía es la unión de varios atributos de dimensión mediante una relación de tipo Jerárquico. Ejemplos: ciudad - estado - país - continente. (Kimball M. y Ross R, 2002)

Las jerarquía nos permite dar una relación de contención a los atributos (un continente está compuesto por países, un país tiene estados, etc.) y un orden haciendo más legible los datos para ser analizados.

2.2.4.4 Celda

Una celda es una posición formada por la intersección de cada uno de los elementos de las dimensiones que forman el cubo. La celda puede contener cero, uno o varios datos. (Cipher, 2008)

De la figura 3 se puede obtener una celda que contenga información de un determinada región y de un producto específico y en un instante de tiempo dado.

La celda es la una posición donde se encuentra una data que se obtiene del cruce de varias dimensiones que sirven de filtro y nos permiten llegar a la data exacta que se requiere analizar.

2.2.4.5 Tabla de Hechos (Fact Table)

Son las tablas centrales dentro del modelo dimensional y contienen los valores de las medidas del negocio que se desean analizar. Estas comúnmente poseen una clave primaria compuesta formada por claves foráneas provenientes de las dimensiones, por lo que se estaría expresando una relación o cardinalidad de 'n' a 'n' entre las tablas de hechos y las dimensiones respectivas. (Wolff C., 2002)

Las tablas de hechos además de estar conformadas por una clave primaria compuesta, poseen uno o más campos numéricos llamados medidas o hechos, que se hacen sobre la granularidad de dicha tabla. La granularidad se entiende como el nivel de detalle con el que se almacenarán los datos en la tabla de hechos.

En general la tabla de hechos es una tabla central donde se tiene una clave primaria compuesta por claves foráneas y por medidas que son campos numéricos que describen la granularidad de la tabla.

2.2.4.6 Tabla Look-up o Tabla Dimensión

Estas se conectan a la tabla de Hechos y permiten ver la información de distintas formas. Almacena un conjunto de valores o atributos que están relacionados con una dimensión particular. Esta relación es expresada en forma de

jerarquía, ya que los miembros o valores de estas tablas están agrupados por niveles de detalle de la información. (Wolff C., 2002)

Estas tablas contienen información textual descriptiva, los atributos de estas son utilizados como filtros de las consultas o análisis realizados sobre el almacén de datos para obtener una información más precisa.

Una tabla dimensión permite representar una perspectiva de los datos y almacenar atributos que tienen alguna relación en particular y esta puede representarse como una jerarquía, por ejemplo: la dimensión tiempo que presenta la siguiente jerarquía: día, semana, mes y año.

2.3 Inteligencia de Negocio (BI, Business Intelligence por sus siglas en inglés)

2.3.1 Definición

Most y Atre (2003). Definen Inteligencia de Negocio como “una arquitectura y colección de operaciones integradas, así como la aplicación al apoyo de la toma de decisión y bases de datos que proporcionan a la comunidad empresarial el acceso fácil a los datos”, es decir, mediante las bases de datos de la organización la Inteligencia de Negocio le proporcionan aplicaciones y herramientas que le permiten a la organización tomar decisiones apoyándose en las mismas.

Por otro lado Larson (2006), define inteligencia de negocios como “la entrega de información exacta, información útil y necesaria para apoyar la toma de decisiones”.

En general se puede decir que los sistemas de BI son el conjunto de metodologías, aplicaciones y tecnologías que permiten a la organización reunir, depurar y transformar datos en información organizada y estructurada que sirve como apoyo a la toma de decisiones de la organización.

2.3.2 Características

Este conjunto de herramientas y metodologías tienen en común las siguientes características:

- **Accesibilidad a la información.** Los datos son la fuente principal de este concepto. Lo primero que deben garantizar este tipo de herramientas y técnicas será el acceso de los usuarios a los datos con independencia de la procedencia de estos.
- **Apoyo en la toma de decisiones.** Se busca ir más allá en la presentación de la información, de manera que los usuarios tengan acceso a herramientas de análisis que les permitan seleccionar y manipular sólo aquellos datos que les interesen.
- **Orientación al usuario final.** Se busca independencia entre los conocimientos técnicos de los usuarios y su capacidad para utilizar estas herramientas.

2.3.3 Componentes o arquitectura

A continuación se describirán los elementos que conforman una herramienta de inteligencia de negocios.

- **Consultas y Reportes,** son herramientas para la elaboración de informes y listados, tanto en detalle como sobre información agregada, a partir de la información de los datawarehouses y datamarts. (Cortes E, 2007)

- **Cuadro de mando analítico (EIS tradicionales)**, es elaborado a partir de datamarts, informes, resúmenes e indicadores clave para la gestión (KPI, Key Performance Indicators), que permitan a los gestores de la empresa analizar los resultados de la misma de forma rápida y eficaz. En la práctica es una herramienta de consulta orientada a la obtención y presentación de indicadores para la dirección (frente a la obtención de informes y listados). (Cortes E, 2007)
- **Cuadro de mando integral o estratégico (Balanced Scorecard)**, este modelo parte de que la estrategia de la empresa es el punto de referencia para todo proceso de gestión interno. Con él los diferentes niveles de dirección y gestión de la organización disponen de una visión de la estrategia de la empresa traducida en un conjunto de objetivos, iniciativas de actuación e indicadores de evolución. (Cortes E, 2007). Los objetivos estratégicos se asocian mediante relaciones causa-efecto y se organizan en cuatro áreas o perspectivas: financiera, cliente, procesos y formación o desarrollo. El cuadro de mando integral es una herramienta que permite alinear los objetivos de las diferentes áreas o unidades con la estrategia de la empresa y seguir su evolución.
- **OLAP (on-line analytical processing)**, son herramientas que manejan interrogaciones complejas de bases de datos relacionales, proporcionando un acceso multidimensional a los datos, capacidades intensivas de cálculo y técnicas de indexación especializadas. Permiten a los usuarios analizar sus datos planteando consultas sobre diferentes atributos o ejes. Utilizan un servidor intermedio para almacenar los datos multidimensionales pre calculado de forma que la explotación sea rápida. (Cortes E, 2007)
- **Datamining (Minería de datos)**, son auténticas herramientas de extracción de conocimiento útil, a partir de la información contenida en las bases de datos de cualquier empresa. El objetivo que se persigue es descubrir patrones ocultos, tendencias y correlaciones, y presentar esta información de forma sencilla y accesible a los usuarios finales, para solucionar, prever y

simular problemas del negocio. Este incorpora la utilización de tecnologías basadas en redes neuronales, árboles de decisión, reglas de inducción, análisis de series temporales y visualización de datos. (Cortes E, 2007)

2.3.4 Ventajas

- Disponer de la capacidad de extraer, depurar, consolidar, sintetizar y presentar datos, información y conocimiento de forma automatizada.
- Poner a disposición de los usuarios la información necesaria para el análisis y la toma de decisiones: ágil, flexible y fiable en el formato adecuado a cada uno. La información se obtiene sin dependencias de otros departamentos, con posibilidad de navegación OLAP por los propios usuarios que permite profundizar en el análisis de forma interactiva en base a cualquiera de las dimensiones disponibles.
- Homogeneidad en la utilización de la información (interna y externa): utilización de la misma información al medir las mismas cosas.
- Sistema soportado sobre plataformas tecnológicas sólidas y escalables.

2.4 Plataforma de Inteligencia de Negocio Pentaho

2.4.1 Definición

Se define como una plataforma de BI “*orientada a la solución*” y “*centrada en procesos*” que incluye los principales componentes requeridos para implementar soluciones basados en procesos y ha sido concebido desde el principio para estar basada en procesos.

Esta plataforma BI ofrece una infraestructura de herramientas de análisis e informes integrado con un motor de workflow de procesos de negocio. La plataforma será capaz de ejecutar las reglas de negocio necesarias, expresadas en forma de procesos y actividades así como también de presentar y entregar la

información adecuada en el momento adecuado. Su modelo de ingresos parece estar orientado a los servicios de soporte, formación y consultoría.

En la Web se presenta una organización por productos: Reporting, Análisis, Dashedboards, Data Mining y se hace mención específica al workflow como una de las capacidades BI claves de la plataforma.

2.4.2 Características de la plataforma Pentaho BI

- Multiplataforma tanto a nivel de cliente como a nivel del servidor no hay dependencia, utiliza lenguaje interpretado.
- Está construido en torno al servidor de aplicaciones J2EE JBoss y Jboss Portal, Tomcat y es portable en cualquier servidor de aplicaciones como glassfish, habilitando que toda la información sea accesible mediante un browser en la intranet de la empresa.
- Pentaho presenta informes en los formatos habituales (Html, Excel, PDF, entre otros) mediante *JfreeReport* u otras plataformas como *BIRT* o *JasperReports*. Para la generación de PDFs utilizan el conocido Apache FOP. Asimismo incorpora la librería *JPivot*, la cual permite visualizar tablas OLAP a través de un browser y realizar las operaciones típicas de análisis OLAP (drill down, slice y dice).
- Se cuenta con áreas de reporte llamado *Pentaho Reporting* en el cual existe tres productos con diferentes enfoques y dirigidos a diferentes tipos de usuarios:
 - *Pentaho Report Designer*: Editor basado en eclipse con capacidad de personalización de informes a las necesidades de negocio destinado a desarrolladores, está incluye asistente para facilitar las configuración de propiedades.
 - *Pentaho Report Design Wizard*: Herramienta de diseño de informe que facilita el trabajo y permite a los usuarios obtener resultados de forma inmediata. Está destinada a usuarios con pocos conocimientos técnicos.

- *Web ad-hoc reporting*: Es similar a la herramienta anterior pero vía Web. Extiende la capacidad de los usuarios finales para la creación de informes a partir de plantillas pre configuradas y siguiendo un asistente de creación.
- Para generar gráficos se apoyan en JFreeChart, una librería para generar los gráficos más comunes (2D, 3D, barras, líneas series temporales, entre otros), interfaces para acceder a diferentes fuentes de datos, exportación a PNG, JPEG y PDF, soporte para servlets, JSPs, applets y aplicaciones clientes.
- Los dashboards son un desarrollo propio de Pentaho, recogen información de todos los componentes de la plataforma incluyendo aplicaciones externas, feeds RSS y páginas Web. Incluyen gestión y filtrado del contenido, seguridad basada en roles y drill down. Pueden ser integrados en terceras aplicaciones, en portales o dentro de la plataforma Pentaho. En la figura 4 se puede observar el portal de la plataforma BI Pentaho usando dashboards.
- Los WebServices constituye una característica esencial de Pentaho. Las acciones que son las tareas más sencillas que constituyen una solución de Pentaho pueden publicarse como WebServices. Se utiliza como motor de WebServices Apache Axis, quedando los servicios descritos en el lenguaje de definición de servicios Web WSDL.
- Proporciona amplia capacidad de preparación de informes de clases y muy flexibles, desde simples informes en un sitio Web de gran formato a los informes para aplicaciones, como la presentación de informes financieros.
- Capacidad de conectarse directamente a las fuentes de datos o para conectar a través de una capa de metadatos centralizados. Se cuenta con una amplia gama de formatos incluyendo Adobe PDF, HTML, Microsoft Excel, RTF o texto plano, informes con gran contenido que incluyen tablas, gráficos, plantillas, galerías, formato para cabeceras, pies de página, entre otros.

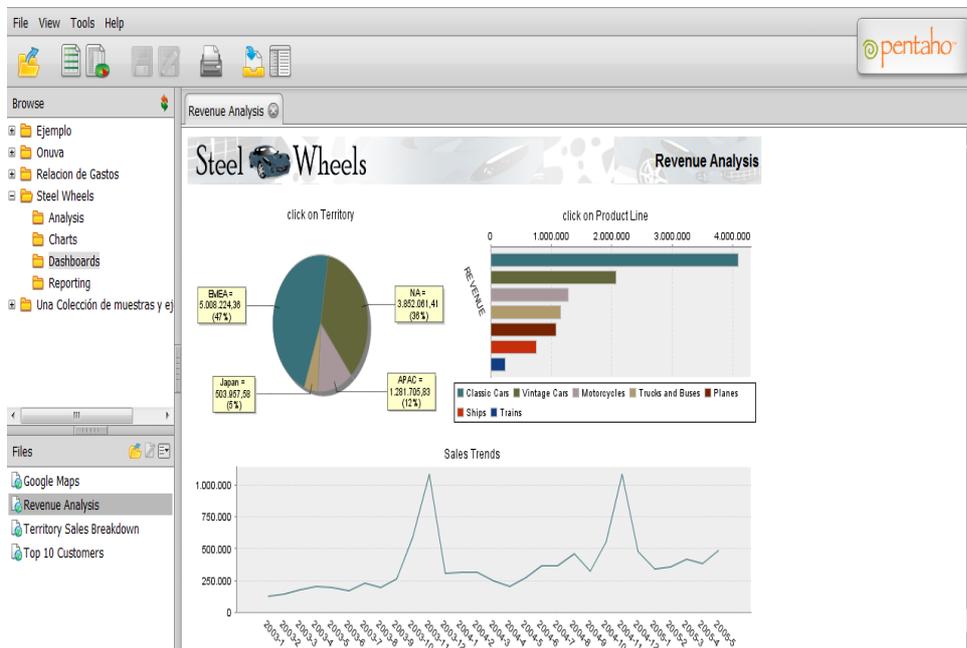


Figura 4. Plataforma de Business Intelligence Pentaho usando Dashboards.
Fuente. www.pentaho.com

- Es compatible con amplia fuente de datos, compatible con conectividad de base de datos de Java (JDBC) 2.0 o posterior, soporta bases de datos propietarias incluyendo Oracle, IBM, DB2, Microsoft SQL Server, NCR entre otras. Soporta populares bases de datos de código abierto incluyendo MySQL, PostgreSQL, EnterpriseDB, entre otras.
- Se cuenta con el Pentaho Data Integration que permite una poderosa extracción, transformación y carga (ETL), ya que muchas organizaciones tienen información disponible en aplicaciones y base de datos separados. Pentaho Data Integration abre, limpia e integra esta información y la pone en manos del usuario, provee una sola versión de todos los recursos de información. La arquitectura de Pentaho Data Integration se puede observar en la figura 5.

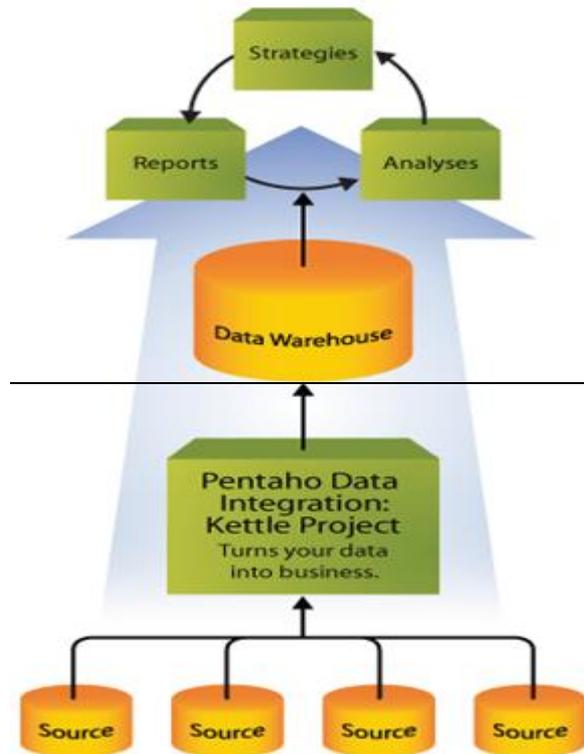


Figura 5. Arquitectura de Pentaho Data Integration
Fuente. www.pentaho.com

2.4.3 Arquitectura

La plataforma de Business Intelligence Pentaho cuenta con la siguiente arquitectura:

- Plataforma cien por ciento J2EE asegurando la escalabilidad, integración y portabilidad.
- Puede correr en servidores compatibles con J2EE como JBoss AS, WebSphere, Tomcat, WebLogic y Oracle AS.

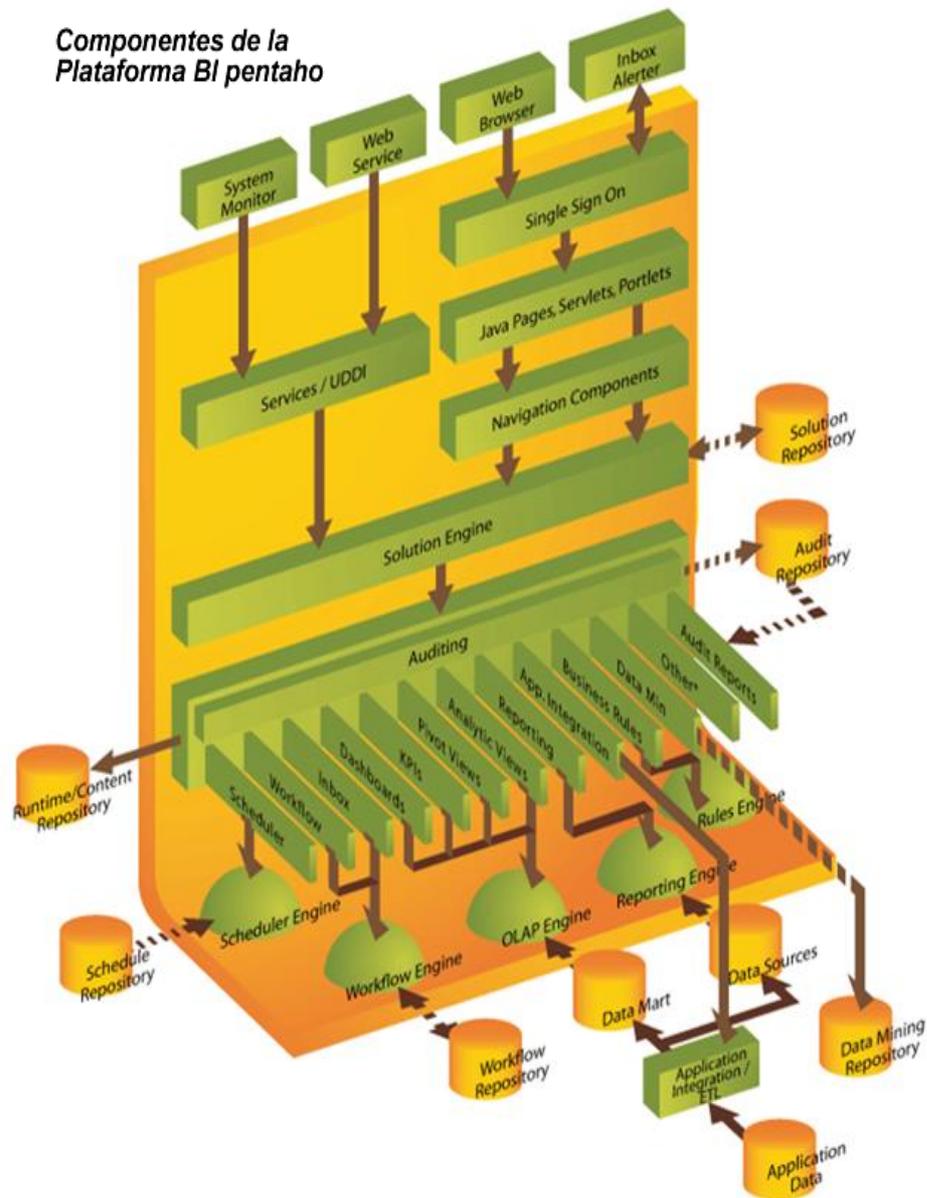


Figura 6. Plataforma BI Pentaho

Fuente. www.pentaho.com

- Base de datos, vía JDBC, IBM DB2, Microsoft SQL Server, MySQL, Oracle, PostgreSQL, NCR Teradata, Firebird.
- No hay dependencia del sistema operativo, utiliza lenguaje interpretado.

- Lenguaje de programación Java, Javascript, JSP, XSL (XSLT/XPath/XSL-FO).
- Interfaz de desarrollo basado en Java SWT, Eclipse, Web-based.
- Repositorio de datos basado en XML.
- Todos los componentes están expuestos vía Web Services para facilitar la integración con arquitecturas orientadas a servicios (SOA).

La figura 6 muestra cómo interactúan los componentes y la arquitectura de Pentaho. El servidor de Pentaho contiene los motores y componentes para la presentación de informes, análisis, reglas de negocio, correo electrónico, notificaciones de escritorio y flujo de trabajo. Estos componentes se integran en conjunto a fin de que puedan ser utilizados para resolver un problema de Inteligencia de Negocio. Dentro de una solución de la conducta, la interoperabilidad, y la interacción del usuario de cada sub-sistema se define por una colección de documentos de definición de soluciones y estos son documentos XML que contienen:

- Definiciones de las actividades que se ejecutan como parte de los procesos, en la demanda o llamado por los servicios web. Estas actividades incluyen las definiciones de: fuentes de datos, consultas, plantillas de informes, las normas de notificación, las normas de negocios, tableros de control, análisis y puntos de vista.

El servidor de Pentaho incluye la infraestructura que proporciona la administración avanzada del sistema. Esto incluye la supervisión del sistema (SMNP, Simple Network Management Protocol) los servicios, informes de uso, soporte de servicios Web, herramientas de validación de configuración y herramientas de diagnóstico, también incluye una herramienta para la Integración de Aplicaciones Empresariales (EAI) / extracción, transformación y carga (ETL).

2.4.4 Componentes

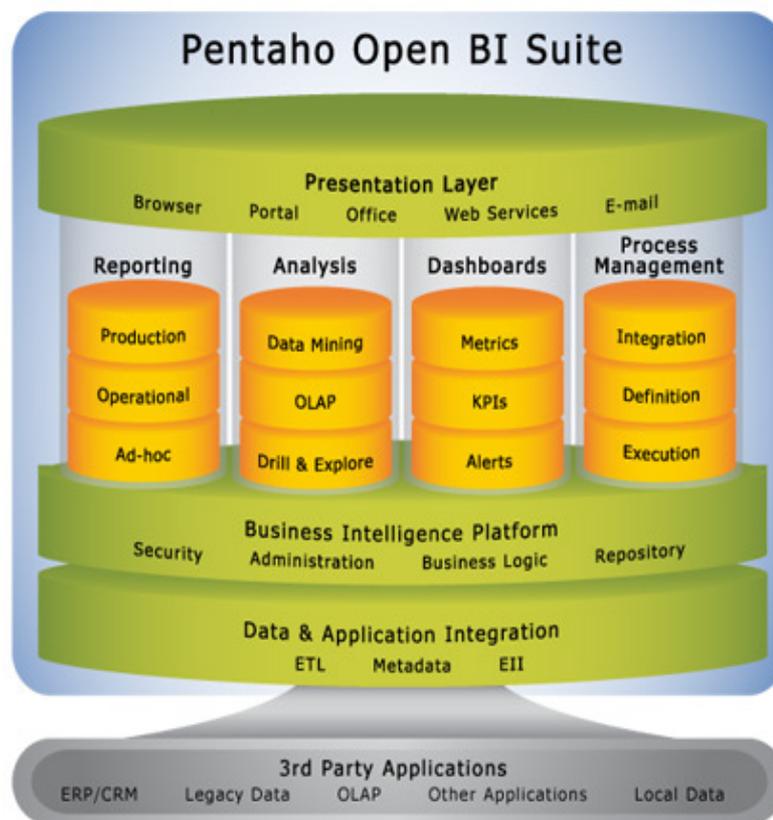


Figura 7. Pentaho Open BI Suite (Componentes)
Fuente. www.pentaho.com

La solución Business Intelligence Open Source Pentaho incluye todos aquellos componentes que nos podemos encontrar en las soluciones BI propietarias más avanzadas como lo son: reporting, análisis, dashboards, workflow, data mining, integración de datos, single sign-on. (Procedimiento de autenticación que habilita al usuario para acceder a varios sistemas con una sola instancia de identificación), auditoría de uso y rendimiento, planificador, notificador, seguridad y perfiles.

La figura 7 se muestra la estructura de los diferentes módulos que forman parte de Pentaho.

Reporte (reporting)

Pentaho Reporting es una solución basada en el proyecto JFreeReport que permite generar informes ágiles y de gran capacidad, este permite la distribución de los resultados del análisis en múltiples formatos todos los informes incluyen la opción de imprimir o exportar a formato PDF, XLS, HTML y texto. Pentaho Reporting permite también la programación de tareas y ejecución automática de informes con una determinada periodicidad. Soporta amplia fuente de datos relacionales, OLAP y XML basados en fuentes de datos.

Análisis (Analysis)

Pentaho Analysis suministra a los usuarios un sistema avanzado de análisis de información. Con uso de las tablas dinámicas (pivot tables, crosstabs) generadas por Mondrian y JPivot, el usuario puede navegar por los datos ajustando la visión de los datos y los filtros de visualización añadiendo o quitando los campos de agregación. Los datos pueden ser representados en una forma de SVG o Flash, dashboards widgets o también integrados con los sistemas de minería de datos y los portales Web (portlets). Además, con el Microsoft Excel Analysis Services se puede analizar los datos dinámicos en Microsoft Excel (usando la conexión a OLAP server Mondrian).

Dashboards

Todos los componentes del módulo Pentaho Reporting y Pentaho Analysis pueden formar parte de un Dashboard. En Pentaho Dashboards es muy fácil incorporar una gran variedad en tipos de gráficos, tablas y velocímetros (dashboard widgets) e integrarlos con los Portlets JSP, en donde podrá visualizar informes, gráficos y análisis OLAP.

Data Mining

Pentaho incorpora la tecnología Weka, es un conjunto integral de herramientas para aprendizaje automático y minería de datos. Su amplio conjunto de clasificación, regresión, reglas de asociación y algoritmos de agrupación se puede utilizar para ayudarle a comprender mejor la empresa y también ser aprovechado para mejorar el rendimiento futuro a través de análisis predictivo. Como principal características tenemos:

- Motor de minería de datos provee de un set de algoritmos de aprendizaje del proyecto Weka incluyendo clustering, segmentación, árboles de decisión, random forests, redes neuronales y análisis de componentes principales, entre otros.
- Integración de Pentaho Data Integration, automatización de procesos de transformación de datos al formato que requiere el motor de data mining.
- Los algoritmos pueden ser aplicados directamente a un set de datos o invocados desde código Java
- Provee filtros para normalización, re-muestreo, selección de atributos, transformación y combinación de atributos
- Modelos para predicción de variables nominales o numéricas
- Herramientas gráficas de diseño para pre procesamiento de datos, regresión, clustering, reglas de asociación

Integración de Datos

Pentaho Data Integration (PDI, también llamada Kettle) es el componente de Pentaho responsable de la extracción, transformación y carga (ETL) de procesos. Aunque las herramientas ETL son los más usados en entornos de almacenes de datos, Kettle también se puede utilizar para otros fines:

- Migración de datos entre las aplicaciones o bases de datos.

- Exportación de datos desde bases de datos para archivos planos.
- Carga de datos de forma masiva en bases de datos.
- Limpieza de datos.

Como principal características se tiene:

- Cada proceso es creado con una herramienta gráfica donde se especifica qué se va hacer sin necesidad de escribir código que indique cómo hacerlo.
- Admite una amplia gama de formatos de entrada y salida, incluyendo archivos de texto, hojas de datos, archivos XML, propiedades de Java y los motores de base de datos open source y propietarios.
- Arquitectura extensible permite desarrollar conectores y plugins fácilmente
- Basado en repositorio facilita la reutilización de componentes de transformación, colaboración y administración de modelos, conexiones, logs, etc.
- Debugger integrado
- Todos los pasos de base de datos existe un botón SQL que generará el código requerido para las tablas y/o índices. También está disponible a nivel de transformación y trabajo.

El workflow de procesos de negocio

Dos son los fundamentos del workflow de procesos de negocio: el motor de workflow Enhydra Shark y el estándar WPDL, ayudado por la Workflow Management Coalition (WFMC), organismo que declara tener más de 300 empresas asociadas incluyendo a IBM, Oracle, BEA, Adobe, SAP, TIBCO o SUN, entre otros. Dentro del proyecto Enhydra se encuentra también Enhydra JaWE un editor de workflow XPDL según las especificaciones de WFMC.

2.5 Metodología de Ralph Kimball (The Business Dimensional Lifecycle)

Tomaremos como base la metodología de desarrollo de sistemas de inteligencia de negocio (Business Dimensional Lifecycle) definido por Kimball, R. (2002) en la figura 8 se ilustra las diferentes etapas de la metodología. Este diagrama ilustra las distintas tareas de alto nivel requeridas para el efectivo diseño, desarrollo e implementación del sistema.

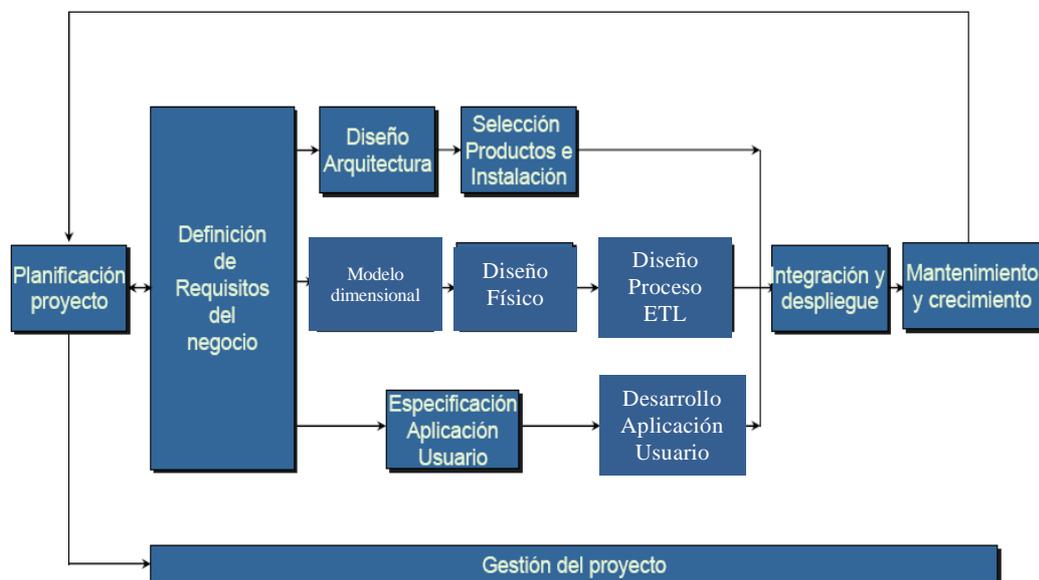


Figura 8. The Business Dimensional Lifecycle (Ciclo de vida multidimensional del negocio)
Fuente. Ralph Kimball.

A continuación se describe cada una de las etapas de la metodología:

2.5.1 Planificación del Proyecto

El ciclo de vida comienza con la planificación del proyecto y está dirigido a la definición y alcance del proyecto, incluyendo la evaluación y la justificación del negocio. Estas tareas deben ser tomadas en cuenta de forma temprana debido a los altos costos asociados a la mayoría de los proyectos de los almacenes de datos. Por

esto, el proyecto de planificación está enfocado en los recursos y habilidades a nivel de las necesidades de personal, unido con la asignación de tareas, duración y secuencia. El resultado integrado identifica todas las tareas asociadas con el Ciclo de Vida Dimensional del negocio. La planificación del proyecto depende de los requerimientos del negocio, como indican con flechas en ambos sentidos entre estas actividades en la figura 8.

Al hacer la planificación del proyecto se debe tener en cuenta:

- Hacer partícipes a los usuarios operacionales, técnicos y analistas, ya que los usuarios aseguran la calidad de los datos; y los técnicos y analistas aseguran la calidad del producto final.
- Seleccionar una aplicación piloto con una alta probabilidad de éxito. Se debe estudiar la viabilidad, el origen de los datos, el hardware y software disponible y planificar un entrenamiento para los usuarios.
- Construir prototipos rápida y frecuentemente.
- Reportar activamente y publicar casos exitosos.
- Herramientas de visualización fáciles de usar.

2.5.2 Definición de los requerimientos del Negocio

El incremento del éxito de un proyecto viene dado por la buena comprensión del negocio y de los requerimientos de los usuarios. Si no existe tal comprensión, el proyecto puede convertirse en un ejercicio inútil para el equipo. Para ello se deben contemplar visitas particularizadas a los usuarios.

Se deben determinar indicadores, los cuales deben reflejar lo que se quiere medir, para ello hay que tener una definición clara y concisa de lo que se quiere para poderlo medir. Además de lo nombrado anteriormente, se deben fijar políticas de seguridad de datos, roles, usuarios, datos y procesos.

2.5.3 Diseño técnico de la arquitectura

Para la solución se requiere de la integración de numerosas tecnologías. Se debe tener en cuenta tres factores: los requerimientos del negocio, los actuales ambientes técnicos y las directrices técnicas estratégicas futuras planificadas para de esta forma poder establecer el diseño de la arquitectura técnica del ambiente de la solución.

2.5.4 Selección de las herramientas

Usando el diseño de la arquitectura, se especifican tanto los componentes de la arquitectura como la plataforma del hardware, el sistema manejador de base de datos, las herramientas de montaje o las herramientas de accesos a datos que se necesitaran evaluar y seleccionar. Una vez que las herramientas han sido evaluadas y seleccionadas, son instaladas y probadas para asegurar la integración apropiada dentro del ambiente de trabajo.

2.5.5 Modelo dimensional

La definición de los requerimientos del negocio determina los datos necesarios para hacer frente a los Requisitos analíticos del negocio del usuario. Diseñando el modelo de datos que soportan los diferentes análisis, requiere un enfoque diferente que se utiliza para el diseño de sistemas operativos. Con todos estos datos se va construyendo una matriz asegurando que el proyecto es extensible en toda la organización a través del tiempo.

Luego de conducirnos a un análisis más detallado de los datos del sistema operativo correspondiente, se logra el acoplamiento de este análisis de datos con el análisis de los requisitos del negocio, desarrollando así un modelo dimensional.

2.5.6 Diseño Físico

El diseño físico de las base de datos se focaliza sobre la selección de las estructuras necesarias para soportar el diseño lógico. Algunos de los elementos principales de este proceso son la definición de convenciones estándares de nombres y del ambiente de la base de datos. La indexación y las estrategias de particionamiento son también determinadas en esta etapa.

2.5.7 Diseño de procesos ETL

En este paso de la metodología se realiza los procesos de extracción a aquellos datos que son requeridos y que permitirán efectuar la carga del modelo físico acordado. Así mismo, el procesos de transformación para convertir o recodificar los datos fuente a fin de poder efectuar la carga efectiva del modelo físico. Por otra parte, los procesos de carga de datos las cuales son los procesos requeridos para poblar el Data Warehouse.

Todas estas tareas son altamente críticas ya que tienen que ver con la materia prima del data warehouse: los datos. La desconfianza y pérdida de credibilidad del data warehouse serán resultados inmediatos e inevitables si el usuario choca con información inconsistente. Es por ello que la calidad de los datos es un factor determinante en el éxito de un proyecto de data warehousing. Es en esta etapa donde deben sanearse todos los inconvenientes relacionados con la calidad de los datos fuente.

2.5.8 Especificación de la aplicación de usuario final

Se recomienda definir un conjunto estándar de aplicaciones para el usuario final. La especificación de la aplicación del usuario final describe las plantillas de reportes, los parámetros que maneja el usuario y los calculo que requieran. Estas

especificaciones aseguran que el equipo de desarrollo y el negocio del usuario entienden lo mismo de la aplicación que se va a entregar.

2.5.9 Desarrollo del la aplicación de usuario final

Siguiendo las especificaciones de la aplicación, el desarrollo de la aplicación del usuario final las herramientas de configuración y construcción de reportes. Esta aplicación se utiliza para construir usando una herramienta avanzada que provee de una significativa productividad para el grupo de desarrollo de la aplicación. Adicionalmente, ofrece un poderoso mecanismo para el usuario de negocios que facilita la modificación de las plantillas de reportes ya existentes.

2.5.10 Despliegue

Representa la convergencia de las tecnologías, datos, y la aplicación de usuario final que es accesible desde el escritorio del usuario de negocios.

CAPÍTULO 3. Marco Aplicativo

3.1 Metodología de desarrollo

En la actualidad las empresas generan una gran cantidad de datos e información para mantener el control de las mismas, pero es muy común que el tratamiento de dicha data se hace muy pesado ya que se realiza manualmente lo que trae como consecuencia un gran empleo de tiempo para dicha labor además de que se pueden generar errores humanos por y esto se hace muy complicado generar reportes que les permitan ver de manera rápida como se está comportando la organización. En nuestro caso se hace imprescindible que la información obtenida esté disponible y colocada en un almacén de datos para luego realizar los diferentes analíticos y reportes y así los gerentes poder tomar decisiones correctas en el momento exacto.

Los usuarios que pertenecen a un departamento de finanza requieren de esta herramienta que les facilite los procesos de extracción y carga de datos y a su vez la generación de reportes para ver sus estadísticas de comportamiento financiero y así poder tomar una mejor decisión frente a cualquier escenario que se pueda presentar en este departamento.

Como se explicó en el apartado 2.5 la metodología de desarrollo es la de Ralph Kimball "*The Business Dimensional Lifecycle*", la cual nos permite abarcar todos los elementos de desarrollo de la aplicación de inteligencia de negocio. A continuación se muestra paso a paso su desarrollo implementado en la solución de inteligencia de negocio para nuestro caso de uso.

3.1.1 Planificación del Proyecto

Esta organización posee sistemas que le permiten obtener gran cantidad de información en el área financiera, pero no poseen una herramienta que le permita extraer, transformar y cargar los datos de los diferentes archivos de Excel para luego realizar los diferentes informes y analíticos que ellos requieren en su organización para tomar decisiones oportunamente, es en este punto donde se desarrollara nuestro trabajo, el cual será desarrollar una solución de inteligencia de negocios.

OBJETIVO	ACTIVIDADES	RESPONSABLE	CRONOGRAMA (semanas)
Análisis Requerimientos	Pedir al cliente los archivos de Excel. Entender el modelo del negocio, para cualquier duda preguntar al cliente.	El quipo de trabajo de la solución y el cliente.	Primera semana de inicio del proyecto
Diseño de la solución	Diseño del modelo multidimensional. Diseño de los procesos ETL. Diseño del patrón de los reportes	El quipo de trabajo de la solución	2 semanas
Desarrollo del Sistema	Implementar el modelo multidimensional Implementar el proceso ETL para la población del datawarehouse. Desarrollo de los reportes.	El quipo de trabajo de la solución	De 3 a 4 semanas

Tabla 1. Plan de trabajo.

Se pautaron 8 reuniones con el cliente para realizar el levantamiento de requerimientos y entrega de datos necesarios para el desarrollo de la herramienta BI. También en estas reuniones se fueron entregando adelantos de la herramienta para ir teniendo feedback con el cliente en cada evolución del sistema.

En la tabla 1 se muestra el plan de trabajo que se realizó para el cumplimiento de dichas tareas, donde se describe el objetivo junto con sus actividades el responsable y el cronograma.

3.1.2 Definición de los requerimientos del Negocio

Se requiere del desarrollo de una solución de inteligencia de negocio para llevar a un departamento de finanzas a repuestas más rápidas a la hora de presentar reportes y extraer los datos de diferente fuentes de datos. En reuniones con el cliente se llegó a la definición de los requerimientos del negocio determinando así el alcance del proyecto, para ello se realizaron fichas técnicas que describen cada uno de los analíticos que requieren especificando los datos que desean visualizar y las métricas que requieren por cada reporte y así como también su organización. A continuación se muestran las fichas técnicas:

En la tabla 2 se describe el requerimiento de Summary junto con la organización de las variables que se requieren y sus medidas, así como también la fecha de inicio y culminación del requerimiento.

3.1.2.1 Ficha técnica del requerimiento A.

FICHA TÉCNICA / MEMORIA DESCRIPTIVA DEL INSTRUMENTO ANALÍTICO				
Fecha Inicio:	07/09/2009	Fecha Finalización:	26/10/2009	
Nombre del Instrumento	Summary			
Gerencias que lo utilizan	DEPARTAMENTO DE FINANZAS DENTRO DEL SECTOR DE PETROQUÍMICA.			
Descripción del Producto Entregado:	El producto es un sistema que permite ver de manera gráfica y en tablas las variables de resumen. Se anexa los datos y su organización.			
	Summary	Budget	Real	Var
	Production MT (´000)			
	Total Sales US\$ Million			
	Weigth Avg. FOB Jose Price (US\$/MT)			
	Total Unit Cost			
	Net Profit			

Tabla 2. Ficha técnica del requerimiento A.

En la tabla 3 se describe el requerimiento de Operations junto con la organización de las variables que se requieren y sus medidas, así como también la fecha de inicio y culminación del requerimiento

3.1.2.2 Ficha técnica del requerimiento B.

FICHA TÉCNICA / MEMORIA DESCRIPTIVA DEL INSTRUMENTO ANALÍTICO				
Fecha Inicio:	07/09/2009	Fecha Finalización:	26/10/2009	
Nombre del Instrumento	Operations			
Gerencias que lo utilizan	DEPARTAMENTO DE FINANZAS DENTRO DEL SECTOR DE PETROQUÍMICA.			
Descripción del Producto Entregado:	El producto es un sistema que permite ver de manera gráfica y en tablas las variables de Operaciones. Se anexa los datos y su organización.			
	Operations	Budget	Real	Var
	Production MT (´000)			
	Scheduled Maintenance Days			
	Non Programmed Maintenance Days			
	Production Days			
	Service Factor			

Tabla 3. Ficha técnica del requerimiento B.

En la tabla 4 se describe el requerimiento de Sales junto con la organización de las variables que se requieren y sus medidas, así como también la fecha de inicio y culminación del requerimiento

3.1.2.3 Ficha técnica del requerimiento C.

FICHA TÉCNICA / MEMORIA DESCRIPTIVA DEL INSTRUMENTO ANALÍTICO				
Fecha Inicio:	07/09/2009	Fecha Finalización:	26/10/2009	
Nombre del Instrumento	Sales			
Gerencias que lo utilizan	DEPARTAMENTO DE FINANZAS DENTRO DEL SECTOR DE PETROQUÍMICA.			
Descripción del Producto Entregado:	El producto es un sistema que permite ver de manera gráfica y en tablas las variables de Ventas. Se anexa los datos y su organización.			
	Sales	Budget	Real	Var
	Total Sales Volumen TM('000)			
	Domestic			
	Export			
	Weigth Avg. FOB Jose Price (US\$/MT)			
	Total Sales US\$ Million			
	Hydrogen Rich Gas			
	Total Methanol Sales US\$ Million			
	Domestic			
	Export			

Tabla 4. Ficha técnica del requerimiento C.

En la tabla 5 se describe el requerimiento de Sales by Region Production MT y Sales by Region Net Back (US\$) junto con la organización de las variables que se requieren y sus medidas, así como también la fecha de inicio y culminación del requerimiento

3.1.2.4 Ficha técnica del requerimiento D.

FICHA TÉCNICA / MEMORIA DESCRIPTIVA DEL INSTRUMENTO ANALÍTICO				
Fecha Inicio:	07/09/2009	Fecha Finalización:	26/10/2009	
Nombre del Instrumento	Sales by Region Production MT y Sales by Region Net Back (US\$)			
Gerencias que lo utilizan	DEPARTAMENTO DE FINANZAS DENTRO DEL SECTOR DE PETROQUÍMICA.			
Descripción del Producto Entregado:	El producto es un sistema que permite ver de manera gráfica y en tablas las variables de Ventas por región. Se anexa los datos y su organización.			
	Sales by Region Production MT / Sales by Region Net Back	Budget	Real	Var
	Local			
	Europe			
	USA			
	Brazil			
	Total			

Tabla 5. Ficha técnica del requerimiento D.

En la tabla 6 se describe el requerimiento de Sales by off taker MT y Net Back (US\$) junto con la organización de las variables que se requieren y sus medidas, así como también la fecha de inicio y culminación del requerimiento

3.1.2.5 Ficha técnica del requerimiento E.

FICHA TÉCNICA / MEMORIA DESCRIPTIVA DEL INSTRUMENTO ANALÍTICO				
Fecha Inicio:	07/09/2009	Fecha Finalización:	26/10/2009	
Nombre del Instrumento	Sales by Off taker MT y Sales by Off Taker Net Back (US\$)			
Gerencias que lo utilizan	DEPARTAMENTO DE FINANZAS DENTRO DEL SECTOR DE PETROQUÍMICA.			
Descripción del Producto Entregado:	El producto es un sistema que permite ver de manera gráfica y en tablas las variables de Ventas por Corporación. Se anexa los datos y su organización.			
	Sales by off- taker Production MT / Sales by off- taker Net Back (US\$)	Budget	Real	Var
	PQV			
	MC			
	MQC			
	Total			

Tabla 6. Ficha técnica del requerimiento E.

En la tabla 7 se describe el requerimiento de Methanol Inventory junto con la organización de las variables que se requieren y sus medidas, así como también la fecha de inicio y culminación del requerimiento

3.1.2.6 Ficha técnica del requerimiento F.

FICHA TÉCNICA / MEMORIA DESCRIPTIVA DEL INSTRUMENTO ANALÍTICO			
Fecha Inicio:	07/09/2009	Fecha Finalización:	26/10/2009
Nombre del Instrumento	Methanol Inventory		
Gerencias que lo utilizan	DEPARTAMENTO DE FINANZAS DENTRO DEL SECTOR DE PETROQUÍMICA.		
Descripción del Producto Entregado:	El producto es un sistema que permite ver de manera gráfica y en tablas las variables de Inventario de Metanol. Se anexa los datos y su organización.		
	Methanol Inventory	Budget	Real
	Closing		
	 Opening		
	 Production		
	 Methanol Swap		
	 Sales		

Tabla 7. Ficha técnica del requerimiento F.

En la tabla 8 se describe el requerimiento de Profit and Loss junto con la organización de las variables que se requieren y sus medidas, así como también la fecha de inicio y culminación del requerimiento

3.1.2.7 Ficha técnica del requerimiento G.

FICHA TÉCNICA / MEMORIA DESCRIPTIVA DEL INSTRUMENTO ANALÍTICO				
Fecha Inicio:	07/09/2009	Fecha Finalización:	26/10/2009	
Nombre del Instrumento	Profit and Loss			
Gerencias que lo utilizan	DEPARTAMENTO DE FINANZAS DENTRO DEL SECTOR DE PETROQUÍMICA.			
Descripción del Producto Entregado:	El producto es un sistema que permite ver de manera gráfica y en tablas las variables de pérdida y ganancia. Se anexa los datos y su organización.			
	Profit and Loss – OPEX Production	Budget	Real	Var
	Total OPEX Production			
	Total Variable Cost			
	Natural Gas			
	Chemicals			
	Electricity			
	Water			
	FSA			
	Total Fixed Cost			
	Industrial Services			
	Condominium			
	Laboratory Supplies			
	Profit and Loss – OPEX Plant	Budget	Real	Var
	Total Cost of Sales			
	Total Production Cost			
	Total OPEX Plant			
	Total Personnel			
	Total Maintenance			
	Total General			
	Total Professional Services			
	Depr. Amort & Specials Prov.			
	Inventory Variation			
	Profit and Loss – OPEX Administrative	Budget	Real	Var
	Total Sales Administrative and Others			
	Total OPEX Sales and Administrative			
	Total Sales			
	Total Personnel			
	Total General			
	Total Professional Services			

	Social Fund Total Depreciation and Amortization Depreciation and Amortization Amortization - Expansion Total Financial (income) expenses Financial Income Financial expenses Financial expenses - expansion Total Other (income) expenses Other (income) expenses			
--	--	--	--	--

Tabla 8. Ficha técnica del requerimiento G.

En la tabla 9 se describe el requerimiento de Profit and Loss (Total) junto con la organización de las variables que se requieren y sus medidas, así como también la fecha de inicio y culminación del requerimiento

3.1.2.8 Ficha técnica del requerimiento H.

FICHA TÉCNICA / MEMORIA DESCRIPTIVA DEL INSTRUMENTO ANALÍTICO				
Fecha Inicio:	07/09/2009	Fecha Finalización:	26/10/2009	
Nombre del Instrumento	Profit and Loss (Total)			
Gerencias que lo utilizan	DEPARTAMENTO DE FINANZAS DENTRO DEL SECTOR DE PETROQUÍMICA.			
Descripción del Producto Entregado:	El producto es un sistema que permite ver de manera gráfica y en tablas las variables de pérdida y ganancia (total). Se anexa los datos y su organización.			
	Profit and Loss (Total) Sales Total Cost of Sales Gross Margin Total Sales, Admin and Other Profit Before Tax Total Tax Net Profit	Budget	Real	Var

Tabla 9. Ficha técnica del requerimiento H.

En la tabla 10 se describe el requerimiento de Unit Cost junto con la organización de las variables que se requieren y sus medidas, así como también la fecha de inicio y culminación del requerimiento

3.1.2.9 Ficha técnica del requerimiento I.

FICHA TÉCNICA / MEMORIA DESCRIPTIVA DEL INSTRUMENTO ANALÍTICO			
Fecha Inicio:	07/09/2009	Fecha Finalización:	26/10/2009
Nombre del Instrumento	Unit Cost		
Gerencias que lo utilizan	DEPARTAMENTO DE FINANZAS DENTRO DEL SECTOR DE PETROQUÍMICA.		
Descripción del Producto Entregado:	El producto es un sistema que permite ver de manera gráfica y en tablas las variables de Costo Unitario. Se anexa los datos y su organización.		
	Unit Cost	Budget	Real
	Total Unit Cost		
	Production Unit Cost		
	Total Variable Cost		
	Sub Total Gas		
	Gas		
	FSA		
	Chemicals and Others		
	Fixed cost and + plant expenses		
	Total Gas Cost (US\$/MMBTU)		
	Total Gas Cost Budget (US\$/MMBTU)		

Tabla 10. Ficha técnica del requerimiento I.

En la tabla 11 se describe el requerimiento de Balance Sheet junto con la organización de las variables que se requieren y sus medidas, así como también la fecha de inicio y culminación del requerimiento

3.1.2.10 Ficha técnica del requerimiento J.

FICHA TÉCNICA / MEMORIA DESCRIPTIVA DEL INSTRUMENTO ANALÍTICO				
Fecha Inicio:	07/09/2009	Fecha Finalización:	26/10/2009	
Nombre del Instrumento	Balance Sheet			
Gerencias que lo utilizan	DEPARTAMENTO DE FINANZAS DENTRO DEL SECTOR DE PETROQUÍMICA.			
Descripción del Producto Entregado:	El producto es un sistema que permite ver de manera gráfica y en tablas las variables de Balance Se anexa los datos y su organización.			
	Balance Sheet	Budget	Real	Var
	Total Current Assents			
	Cash and Securities			
	Account Receivables			
	Inventories			
	Tax Credits			
	Deferred income taxes			
	Prepaid expenses			
	Total Assets			
	Fixed assets			
	Projects in progress			
	Deferred charges			
	Others assets			
	Total Liabilities and Equity			
	Total Liabilities			
	Sub Total			
	Long term loans			
	Total Shareholders Equity			
	Capital stocks & reserves			
	Treasury stock			
	Retained earnings			
	Net profit of the year			

Tabla 11. Ficha técnica del requerimiento J.

En la tabla 12 se describe el requerimiento de FSA junto con la organización de las variables que se requieren y sus medidas, así como también la fecha de inicio y culminación del requerimiento

3.1.2.11 Ficha técnica del requerimiento K.

FICHA TÉCNICA / MEMORIA DESCRIPTIVA DEL INSTRUMENTO ANALÍTICO				
Fecha Inicio:	07/09/2009	Fecha Finalización:	26/10/2009	
Nombre del Instrumento	FSA			
Gerencias que lo utilizan	DEPARTAMENTO DE FINANZAS DENTRO DEL SECTOR DE PETROQUÍMICA.			
Descripción del Producto Entregado:	El producto es un sistema que permite ver de manera gráfica y en tablas las variables de FSA. Se anexa los datos y su organización.			
	FSA	Budget	Real	Var
	Total Cash Revenue			
	Gross Revenue			
	Variation in account receivable			
	Total Cash Cost			
	Sub Total			
	Production Expenses			
	Sales and Administrative Expenses			
	Financial Expenses and Other Inco.			
	Variation in account payables			
	Total net cash margin			
	% of suppl. amount according to level			
	Feedstock Supplemental Amount			

Tabla 12. Ficha técnica del requerimiento K.

3.1.3 Selección de las herramientas.

En el seminario “Estudio comparativo de herramientas software libre para el apoyo a la toma de decisiones en las pequeñas y medianas empresas”; se realizó estudios comparando las principales características, funcionalidades y ventajas de las principales plataformas de código abierto en el área de inteligencia de negocios como los son Pentaho, Jaspersoft y Birt. Luego de ese estudio se llegó a la conclusión que la plataforma de BI Pentaho, es una de las principales aplicaciones de código abierto que ofrece herramientas necesarias que satisfacen la inteligencia de negocios en las organizaciones aumentando así sus ventajas competitivas (en el apartado 2.4 se estudia a mayor profundidad las características de la herramienta Pentaho); numerosas organizaciones que han adoptado la aplicación Business

Intelligence Open Source Pentaho se han abierto paso en el mercado con relativa facilidad dando respuestas concretas a las carencias con respecto a la presentación de la información de las organizaciones.

3.1.4 Diseño técnico de la arquitectura

La arquitectura es la requerida por el tipo de implementación que se va a realizar y es la de un sistema de inteligencia de negocios y cuenta con los siguientes componentes:

3.1.4.1 Componentes de la arquitectura del sistema

Mondrian

El componente central de la solución que permite visualizar los cubos es Mondrian, conocido también como Pentaho Analysis Services. Es el motor OLAP integrado en la suite de Inteligencia de negocios de código abierto de Pentaho. La razón de usar OLAP para las consultas es la velocidad de respuesta. Una base de datos relacional almacena entidades en tablas discretas y normalizadas. Esta estructura es buena en un sistema transaccional pero para las complejas consultas que incluyen distintos tipos de datos (almacenados cada uno en una tabla) es relativamente lenta.

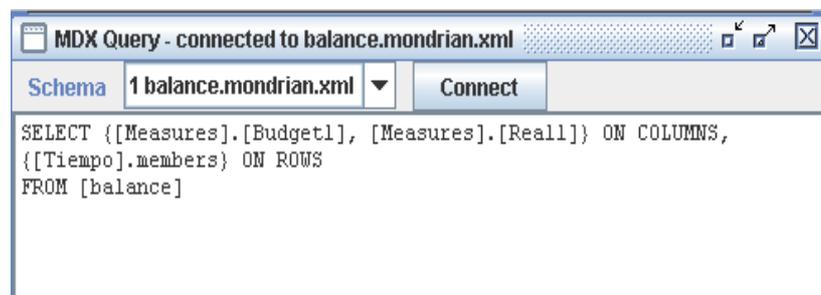
Mondrian es un motor ROLAP con caché, lo cual lo sitúa cerca del concepto de OLAP híbrido. ROLAP significa que en Mondrian no residen datos (salvo en la caché) sino que estos residen en una Sistema de Gestión de Bases de Datos externo. Es en esta base de datos en la que residen las tablas que conforman la información multidimensional con la que Mondrian trabaja.

Mondrian se encarga de recibir consultas multidimensionales (en lenguaje MDX, estándar de expresiones multidimensionales) y devolver los datos de un cubo, sólo que este cubo no es algo físico sino un conjunto de metadatos que definen como se van a “mapear” estas consultas que tratan conceptos multidimensionales a sentencias SQL de manera de satisfacer la consulta multidimensional.

Requerimientos, JRE (Java Runtime Environment, o Ambiente de Ejecución Java) 1.5 o superior. Bases de datos compatibles con SQL (MySQL, Postgres).

MDX

Es un lenguaje para consultar bases de datos multidimensionales, de la misma manera que SQL se utiliza para la consulta de bases de datos relacionales. En un principio se había definido como parte de la OLE DB para OLAP especificación y un lenguaje similar, mdXML, es parte de la especificación XML para Análisis. A pesar de que MDX se ve muy similar a SQL no es lo mismo, ya que MDX es un lenguaje ejecutado por Mondrian y de acuerdo con la base de datos que esté enlazado, realiza una traducción y genera la consulta que es capaz de interpretar el manejador de BD en uso. En la figura 9 podemos observar un ejemplo de la sintaxis del query MDX.



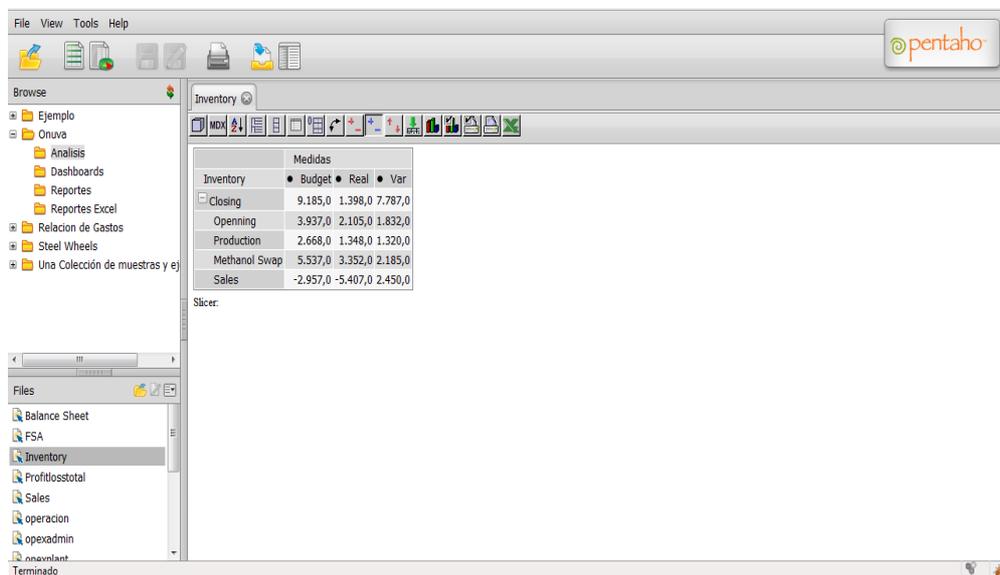
```
MDX Query - connected to balance.mondrian.xml
Schema 1 balance.mondrian.xml  Connect
SELECT {[Measures].[Budget1], [Measures].[Reall]} ON COLUMNS,
{[Tiempo].members} ON ROWS
FROM [balance]
```

Figura 9. Ejemplo de una consulta sencilla del lenguaje MDX

JPivot

Para la visualización de la información Multidimensional se tiene la herramienta llamada Jpivot el cual es un componente de la plataforma Pentaho (En la figura 10 podemos observar la interfaz de JPivot, junto con todos sus componentes). Esta herramienta permite seleccionar las dimensiones que se quieren analizar. Luego permite ejecutar las operaciones OLAP típicas como lo son:

- Dril-down: ir al detalle, por ejemplo, año -> trimestre -> mes.
- Roll – up: permite ir del detalle a un nivel más general.
- Slice: seleccionar un subconjunto con un valor único (todas las ventas correspondientes a un trimestre)
- Dice: es como un Slice pero con dos dimensiones.
- Píivot: cambiar la orientación del reporte, por ejemplo intercambiar el eje de las x por el eje de las y.



The screenshot shows the JPivot Navigator interface. The main window displays a pivot table for 'Inventory' with the following data:

	Budget	Real	Var
Closing	9.185,0	1.398,0	7.787,0
Openning	3.937,0	2.105,0	1.832,0
Production	2.668,0	1.348,0	1.320,0
Methanol Swap	5.537,0	3.352,0	2.185,0
Sales	-2.957,0	-5.407,0	2.450,0

The interface also includes a 'Browse' pane on the left with a tree view of folders like 'Ejemplo', 'Onuva', 'Análisis', 'Dashboards', 'Reportes', 'Reportes Excel', 'Relacion de Gastos', 'Steel Wheels', and 'Una Colección de muestras y ej...'. A 'Files' pane at the bottom left shows files such as 'Balance Sheet', 'FSA', 'Inventory', 'Profitos total', 'Sales', 'operacion', 'opevadmin', 'anaviant', and 'Terminado'. The top menu bar includes 'File', 'View', 'Tools', and 'Help'. The Pentaho logo is visible in the top right corner.

Figura 10. JPivot Navigator

Pentaho Reporting Designer (Diseñador de reportes de Pentaho)

Otro de los componentes a usar es el Reporting Designer de Pentaho, permite a las organizaciones un fácil acceso, formato y distribución de información. Este provee flexibilidad de despliegue incomparable.

- Amplia capacidad de reportes desde simples reportes web hasta reportes altamente personalizados con formatos complejos.
- Publicación de reportes tanto planificados como bajo demanda.
- Entrega en línea, vía correo electrónico, web portales, aplicaciones personalizadas, servicios web o notificación mediante la barra de tareas.
- Amplio rango de formatos, incluyendo Adobe PDF, HTML, Microsoft Excel o texto plano.
- Contenido interactivo en los reportes incluyendo tablas, gráficos, plantillas, galerías, vistas previas, formato de encabezados y pies de página, etc.
- Soporte de Sub Reportes
- Creación de Reportes Ad-Hoc mediante Wizards que ocultan la complejidad al usuario final mediante una interfaz web
- Integración con hojas de cálculo.
-

A continuación en la figura 11 se muestra la interfaz del ambiente de trabajo de Pentaho Reporting Designer.

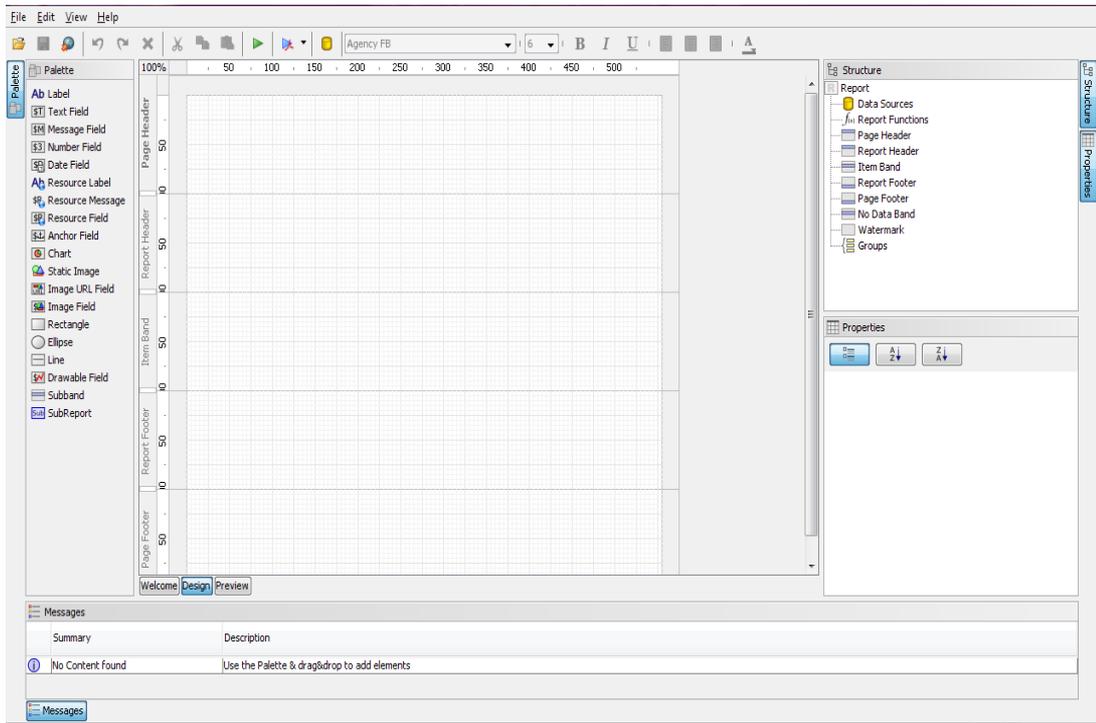


Figura 11. Pentaho Reporting Designer

Schema Workbench

El esquema de trabajo de Mondrian (workbench) le permite visualmente crear y probar esquemas cubo OLAP Mondrian (En la figura 12 se puede observar la interfaz del esquema Workbench). Como se menciono anteriormente Mondrian es un OLAP Server que permite hacer consultas multidimensionales mediante un lenguaje de consultas llamado MDX habilitando un análisis interactivo para el estudio de grandes cantidades de datos guardados en base de datos relacionales. Los esquemas de Mondrian son almacenados en archivos XML y pueden ser utilizados por los otros componentes de la plataforma.

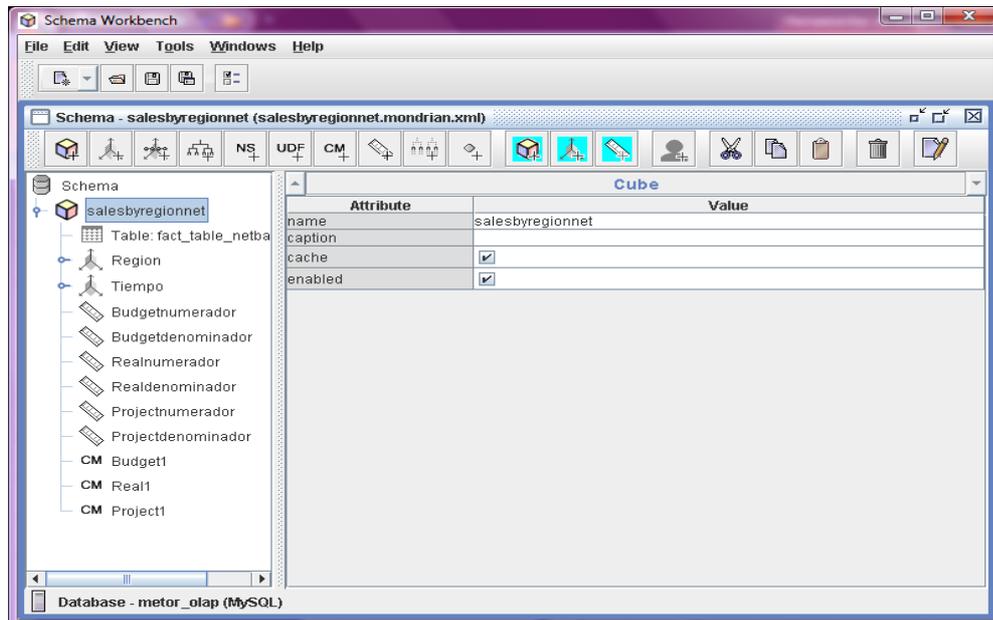


Figura 12. Shema Workbeach

El esquema Workbeach ofrece las siguientes funcionalidades:

- Editor Schema integrado con la fuente de datos subyacente para su validación.
- Prueba de consultas MDX y la base de datos
- Navegador de la estructura de la base de datos subyacentes.

Como elementos principales para la creación de los cubos tenemos:



Cube: Componente para crear un cubo.



Dimension: Componente para crear una dimensión dentro del cubo.



Hierarchy: Componente para agregar una jerarquía dentro de una dimensión.



Measure: Componente para crear una métrica.



Calculated Member: Componente para crear una medida calcula, esta usa las métricas y se le aplica operaciones.



Level: Componente para crear un nivel dentro de una jerarquía definida en la dimensión

Esta herramienta genera los archivos .modrian.xml que es el archivo que contiene la estructura de los cubos con sus dimensiones, atributos y métricas. Para poder visualizar los analíticos en el servidor de Pentaho también es necesario crear el archivo .xaction el cual contendrá los parámetros, query que se requieren para visualizar dichos analíticos en el servidor de Pentaho. En la figura 13 se muestra parte de la estructura del archivo .modrian.xml.

```
<Schema name="balance">
  <Cube name="balance" cache="true" enabled="true">
    <Table name="fact_table_mentor">
    </Table>
    <Dimension type="TimeDimension" foreignKey="id_fecha" name="Tiempo">
      <Hierarchy name="Tiempo" hasAll="true" allMemberName="All tiempo" primaryKey="id_fecha">
        <Table name="dim_fecha">
        </Table>
        <Level name="Anio" table="dim_fecha" column="year" type="String" uniqueMembers="false" levelType="TimeYears">
        </Level>
        <Level name="Mes" table="dim_fecha" column="anio_mes" type="String" uniqueMembers="true" levelType="TimeMont">
        </Level>
      </Hierarchy>
    </Dimension>
    <Dimension foreignKey="id_balancesheet" name="Balance">
      <Hierarchy name="Balance" hasAll="true" allMemberName="All balance" primaryKey="id_balancesheet">
        <Table name="dim_balancesheet">
        </Table>
        <Level name="Nivel 1" table="dim_balancesheet" column="tipo_balance" ordinalColumn="orden" type="String" uni>
        </Level>
        <Level name="Nivel 2" table="dim_balancesheet" column="caracteristica" ordinalColumn="orden2" type="String">
        </Level>
        <Level name="Nivel 3" table="dim_balancesheet" column="caracteristica1" ordinalColumn="orden3" type="String">
        </Level>
        <Level name="Nivel 4" table="dim_balancesheet" column="caracteristica2" ordinalColumn="orden4" type="String">
        </Level>
      </Hierarchy>
    </Dimension>
  </Cube>
</Schema>
```

Figura 13. Estructura del archivo .modrian.xml

En la figura 14 se muestra parte de la estructura del archivo xaction, lo más importante de este archivo son las conexiones a la base de datos y la referencia que se hace al archivo .modrion.xml y el query para poder visualizar el analítico en el servidor Web.

```

<action-outputs>
  <model type="string"/>
  <connection type="string"/>
  <mdx type="string"/>
  <options type="list"/>
  <title type="string"/>
  <url type="string"/>
</action-outputs>
<component-name>PivotViewComponent</component-name>
<component-definition>
  <title>balance</title>
  <viewer>Pivot</viewer>
  <model>Onuva/Analisis/balance.mondrian.xml</model>
  <jndi>onuva</jndi>
  <query>with member [Measures].[Budget] as 'IIf((NOT IsEmpty([Measures].[Budget1])), [Measures].[Budget
member [Measures].[Real] as 'IIf((NOT IsEmpty([Measures].[Real])), [Measures].[Real], 0.0)'
member [Measures].[Var] as 'IIf((NOT IsEmpty([Measures].[Var1])), [Measures].[Var1], 0.0)'
select {[Measures].[Budget], [Measures].[Real], [Measures].[Var]} ON COLUMNS,
  Hierarchize(Union(Union(Union(Union(Union(Union({[Balance.Balance].[All balance].[Total Current Assents], [Balance.F
from [balance]</query>
  <options>
    <personal/>
    <cube-nav/>
    <mdx-edit/>
    <sort-conf/>
    <level-style/>
    <hide-spans/>

```

Figura 14. Estructura del archivo .xaction

ETL Kettle

Herramienta para la transformación, extracción y carga de los datos que provienen de los archivos de Excel utilizados por el departamento, para alimentar el datawarehouse. A continuación se muestra su la interfaz gráfica de Kettle (figura 15)

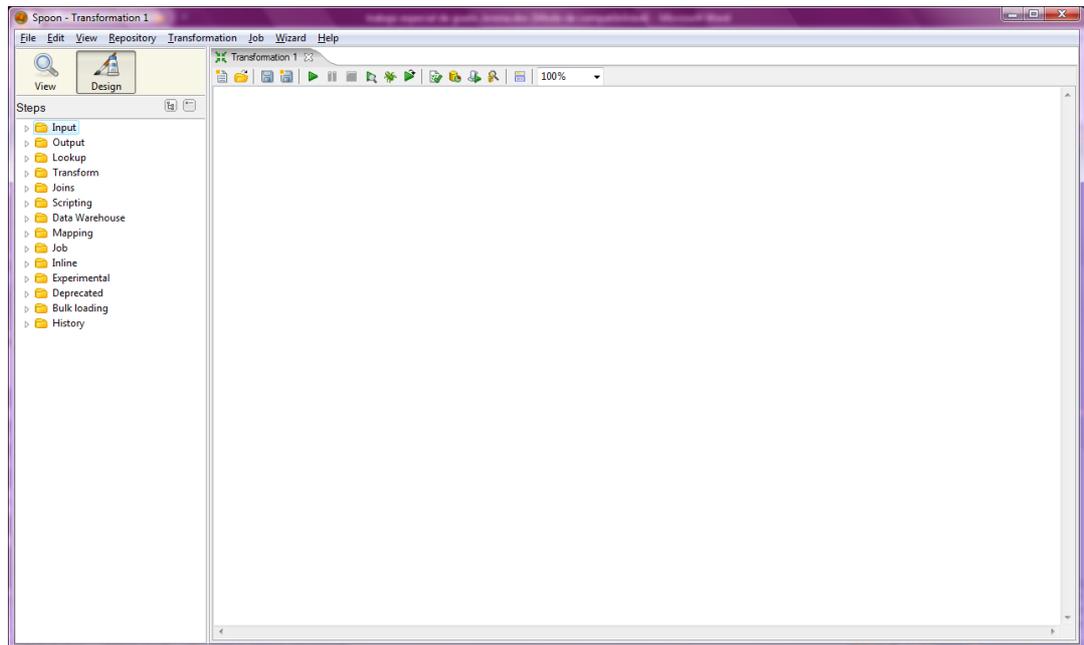


Figura 15. Interfaz gráfica de Kettle

ETL kettle se compone de 4 herramientas:

- SPOON: permite diseñar de forma gráfica la transformación ETL.
- PAN: ejecuta las transformaciones diseñadas con SPOON.
- CHEF: permite mediante una interfaz gráfica, diseñar la carga de datos incluyendo un control de estado de los trabajos.
- KITCHEN: permite ejecutar los trabajos batch diseñados con Chef.

MySQL

MySQL es el Sistema Manejador de Bases de Datos (SMBD), proporciona rapidez y facilidad de uso. Es un manejador de gran aceptación y es debido en parte, a que existen infinidad de librerías y otras herramientas que permiten su uso a través de una gran cantidad de lenguajes de programación, además de su fácil instalación y configuración.

Las características más importantes de MySQL son:

- Multihilo, núcleo de aplicación completamente multihilo. Puede usar múltiples CPUs, si están disponibles.
- Disponibilidad en diversas plataformas y sistemas operativos (Linux, Windows, Solaris, MacOS X, OS/2, entre otros).
- Provee motores de almacenamiento que tratan con distintos tipos de tablas, tanto transaccionales como no transaccionales. Entre ellos tenemos MyISAM, Memory, Merge, InnoDB, BDB.
- APIs disponibles para C, C++, Java, Perl, PHP, Ruby entre otros.
- El código de MySQL está comprobado por Purify (herramienta para verificación de problemas de memoria).
- Soporta un ambiente Cliente/Servidor.

Plataforma BI de Pentaho

La plataforma de inteligencia de negocios Pentaho provee un ambiente para la administración de usuarios, roles, autenticación y autorización unificada, reportes, análisis, dashboards e indicadores de desempeño (KPI), entre otras funcionalidades.

Requerimientos, JRE (Java Runtime Environment, o Ambiente de Ejecución Java) 1.5 o superior, 2GB de RAM, 1GB en disco.

JRE (Java Runtime Environment)

JRE (Java Runtime Environment) permite ejecutar aplicaciones desarrolladas en Java permitiendono nos así crear visualmente esquemas de cubos OLAP y validar el esquema contra la base de datos del cubo, ejecutar consultas MDX, usando el esquema y la base de datos.

3.1.4.2 Integración de los componentes del sistema

En la tabla 13 se describe los componentes de Pentaho, junto con su función y las interfaces que utiliza para poder funcionar dicho componente, de esta manera se muestra la integración que existen entre los componente de BI Pentaho.

Componente	Función	Interfaces utilizadas
Mondrian	Leer los cubos diseñados con el Workbench y grabados en un archivo XML. Permite ejecutar las operaciones OLAP típicas anteriormente descritas.	Utiliza el conector java SQL para leer los datos de los cubos grabados en la BD. Utilizando las librerías JRE y la plataforma Tomcat genera las instrucciones necesarias para que le Servidor Tomcat presente las páginas con las que se visualizarán las consultas de los cubos
Diseñador de Reportes (Reporting Desinger)	Leer el dominio creado con Pentaho Metadata Editor en formato XML para extraer los datos necesarios para la generación de los reportes.	Pentaho Report Designer Pentaho contenido XML. Archivos de estilo XSL son usados para generar los archivos HTML.
Esquema Workbeach	Schema Workbench permite crear visualmente esquemas de cubos OLAP y validar el esquema contra la base de datos del cubo, para permitir la visualización de la data posteriormente con jpivot	Utilizando el conector java mysql el Workbench accede al esquema de la base de datos, el usuario puede producir.
Plataforma de Inteligencia de Negocio	Administración de usuarios, roles, autenticación y autorización unificada, agendamiento de reportes, análisis, tableros de control e indicadores de desempeño.	Los archivos de configuración son almacenados en XML, utiliza XSL para guardar los estilos de las páginas (totalmente configurable), puede utilizar LDAP para administración de autenticación y manejo de perfiles. Utiliza el JRE como entorno de programación y finalmente en la implementación de de la solución BI se utilizará el conector java mysql.

Tabla 13. Integración de los componentes del sistema.

3.1.5 Modelo dimensional

Luego de definir todos los requerimientos del negocio determinamos los datos necesarios para cumplir esos requerimientos analíticos planteados el departamento de finanzas. Comenzamos a definir como las dimensiones de nuestra solución y luego especificamos los diferentes grados de detalle (atributos) dentro de cada dimensión, como así también la granularidad de cada indicador (variable o métrica) y las diferentes jerarquías que dan forma al modelo dimensional de un departamento de finanzas.

La solución dimensional a la que se llegó se muestra en la figura 16, aquí se observan las dimensiones, con su tabla de hechos con su granos y su relación entre ellas.

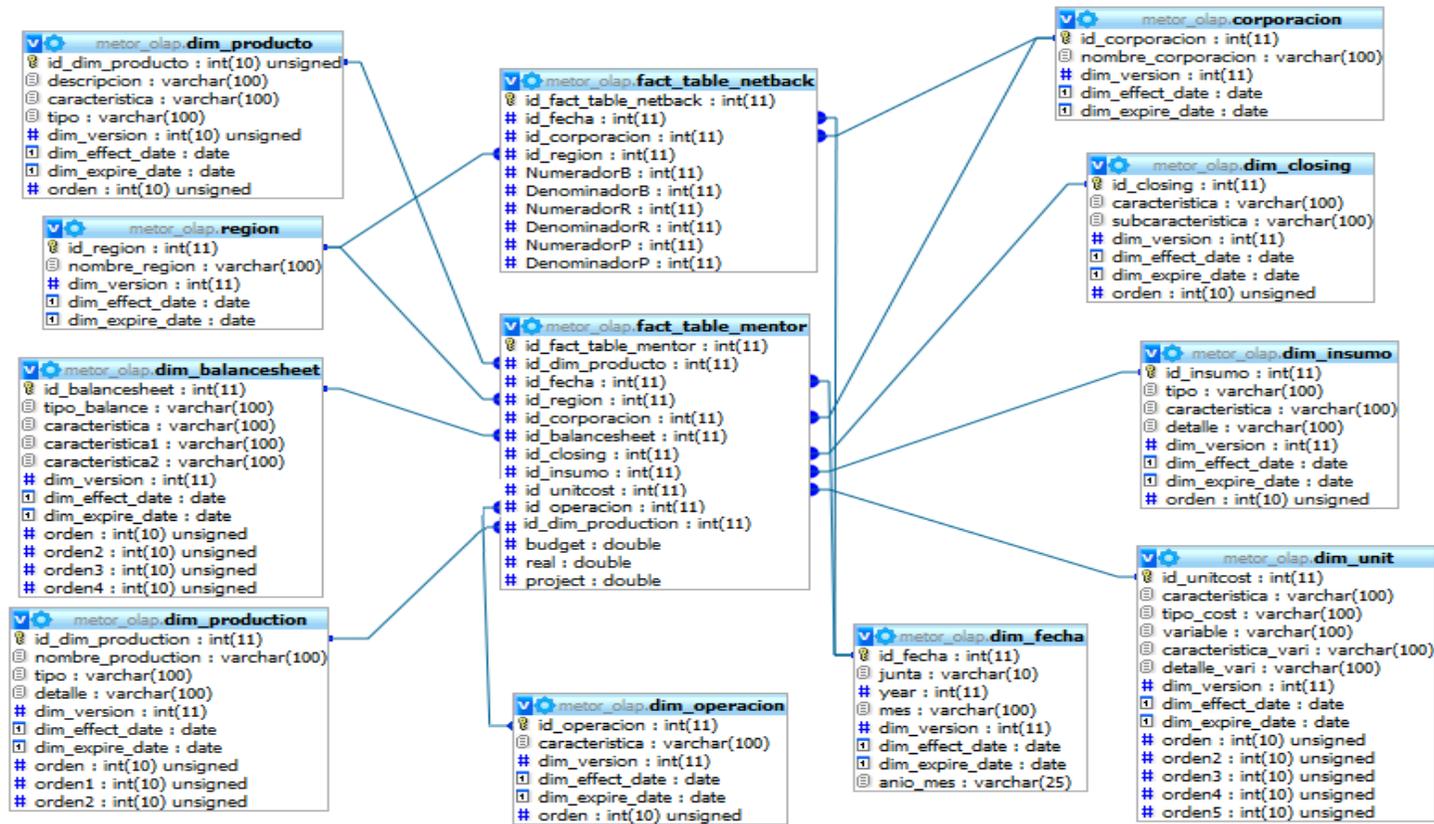


Figura 16: Modelo del diseño multidimensional con sus dimensiones, tabla de hechos y relaciones.

3.1.6 Diseño Físico

En este paso de la metodología se describir la estructura del modelo dimensional. A continuación se muestra en la tabla 14 la estructura de la dimensión de producto con su definición, atributos y tipos de datos.

3.1.6.1 Dimensión: dim_producto

dim_producto		
Descripción:	Tabla que se almacena la información relacionada con los productos.	
Tipo	Tabla de dimensión	
Columnas		
Id_dim_producto	INT (11) NOT NULL AUTO_INCREMENT	Identificador único de la tabla dimensión producto
Descripcion	VARCHAR (100) NOT NULL	Descripción
Caracteristica	VARCHAR (100) NOT NULL	Características
Tipo	VARCHAR (100) NOT NULL	Tipo
dim_version	INT (11) NOT NULL	Campo para la versión del datos
dim_effect_date	DATE NOT NULL	Fecha de creación
dim_expire_date	DATE NOT NULL	Fecha de expiración
Orden	INT (11) NOT NULL	Campo para preservar el orden en que se almacenan los registros en la dimensión

Tabla 14. Dimensión dim_producto.

En la tabla 15 se muestra la estructura de la dimensión región con su definición, atributos y tipos de datos.

3.1.6.2 Dimensión: region

Region		
Descripción	Tabla que se almacena la información de las regiones para mostrar las ventas por región en Production Volumen TM y Net Back.	
Tipo	Tabla de dimensión	
Columnas		
Id_region	INT (11) NOT NULL AUTO_INCREMENT	Identificador único de la tabla dimensión región
nombre_region	VARCHAR (100) NOT NULL	Nombre de las regiones, las cuales son Local (Venezuela), Europa, USA y Brasil.
dim_version	INT (11) NOT NULL	Campo para la versión del dato
dim_effect_date	DATE NOT NULL	Fecha de creación
dim_expire_date	DATE NOT NULL	Fecha de expiración

Tabla 15. Dimensión region.

En la tabla 16 se muestra la estructura de la dimensión corporación con su definición, atributos y tipos de datos.

3.1.6.3 Dimensión: corporacion

Corporación		
Descripción	Tabla que se almacena la información de las corporaciones para mostrar las ventas por corporación en Production Volumen TM y Net Back.	
Tipo	Tabla de dimensión	
Columnas		
Id_corporacion	INT (11) NOT NULL AUTO_INCREMENT	Identificador único de la tabla dimensión corporación
Nombre_corporacion	VARCHAR (100) NOT NULL	Nombre de las corporaciones, las cuales son Pequiven, MITSUBISHI CORPORATION y MITSUBISHI GAS CHEMICAL.
dim_version	INT (11) NOT NULL	Campo para la versión del dato
dim_effect_date	DATE NOT NULL	Fecha de creación
dim_expire_date	DATE NOT NULL	Fecha de expiración

Tabla 16. Dimensión corporacion.

En la tabla 17 se muestra la estructura de la dimensión balance sheet con su definición, atributos y tipos de datos.

3.1.6.4 Dimensión: dim_balancesheet

dim_balancesheet		
Descripción:	Tabla que se almacena la información de las distintas variables que se evalúan en el requerimiento Balance Sheet.	
Tipo	Tabla de dimensión	
Columnas		
Id_dim_balancesheet	INT (11) NOT NULL AUTO_INCREMENT	Identificador único de la tabla dimensión dim_balancesheet
tipo_balance	VARCHAR (100) NOT NULL	Tipo
Caracteristica	VARCHAR (100) NOT NULL	Características
characteristical	VARCHAR (100) NOT NULL	Características1 (Sub nivel 1)
caracteristica2	VARCHAR (100) NOT NULL	Características2 (Sub Nivel 2)
dim_version	INT (11) NOT NULL	Campo para la versión del dato
dim_effect_date	DATE NOT NULL	Fecha de creación
dim_expire_date	DATE NOT NULL	Fecha de expiración
Orden	INT (11) NOT NULL	Campo para preservar el orden por tipo_balance en que se almacenan los registros en la dimensión
orden2	INT (11) NOT NULL	Campo para preservar el orden por característica en que se almacenan los registros en la dimensión
orden3	INT (11) NOT NULL	Campo para preservar el orden por characteristical en que se almacenan los registros en la dimensión
orden4	INT (11) NOT NULL	Campo para preservar el orden por caracteristica2 en que se almacenan los registros en la dimensión

Tabla 17. Dimensión dim_balancesheet.

En la tabla 18 se muestra la estructura de la dimensión produccion con su definición, atributos y tipos de datos.

3.1.6.5 Dimensión: dim_produccion

dim_produccion		
Descripción	Tabla que se almacena la información de las distintas variables que se evalúan en el requerimiento Profit and Loss	
Tipo	Tabla de dimensión	
Columnas		
Id_plant	INT (11) NOT NULL AUTO_INCREMENT	Identificador único de la tabla dimensión dim_produccion
Nombre	VARCHAR (100) NOT NULL	Nombre
nombre_produccion	VARCHAR (100) NOT NULL	Nombre Producción
Tipo	VARCHAR (100) NOT NULL	Tipo
Detalle	VARCHAR (100) NOT NULL	Detalle
dim_effect_date	DATE NOT NULL	Fecha de creación
dim_expire_date	DATE NOT NULL	Fecha de expiración
Orden	INT (11) NOT NULL	Campo para preservar el orden por nombre_plant en que se almacenan los registros en la dimensión
orden2	INT (11) NOT NULL	Campo para preservar el orden por tipo en que se almacenan los registros en la dimensión
orden3	INT (11) NOT NULL	Campo para preservar el orden por detalle en que se almacenan los registros en la dimensión

Tabla 18. Dimensión dim_produccion.

En la tabla 19 se muestra la estructura de la dimensión unit cost con su definición, atributos y tipos de datos.

3.1.6.6 Dimensión: dim_unit

dim_unit		
Descripción	Tabla que se almacena la información de las distintas variables que se evalúan en el requerimiento Unit Cost	
Tipo	Tabla de dimensión	
Columnas		
Id_unitcost	INT (11) NOT NULL AUTO_INCREMENT	Identificador único de la tabla dimensión dim_unit
Característica	VARCHAR (100) NOT NULL	Característica
tipo_cost	VARCHAR (100) NOT NULL	Tipo Costo
Variable	VARCHAR (100) NOT NULL	Variable
caracteristica_vari	VARCHAR (100) NOT NULL	Caracateristica de la Variable
detalle_vari	VARCHAR (100) NOT NULL	Detalle de la Variable
dim_version	INT (11) NOT NULL	Campo para la versión del dato
dim_effect_date	DATE NOT NULL	Fecha de creación
dim_expire_date	DATE NOT NULL	Fecha de expiración

Tabla 19. Dimensión dim_unit.

En la tabla 20 se muestra la estructura de la dimensión insumo con su definición, atributos y tipos de datos.

3.1.6.7 Dimensión: dim_insumo

dim_insumo		
Descripción	Tabla que se almacena la información de las distintas variables que se evalúan en el requerimiento Profit and Loss Total	
Tipo	Tabla de dimensión	
Columnas		
Id_insumo	INT (11) NOT NULL AUTO_INCREMENT	Identificador único de la tabla dimensión dim_insumo

Característica	VARCHAR (100) NOT NULL	Característica
tipo_cost	VARCHAR (100) NOT NULL	Tipo de Costo
Detalle	VARCHAR (100) NOT NULL	Detalle
dim_version	INT (11) NOT NULL	Campo para la versión del dato
dim_effect_date	DATE NOT NULL	Fecha de creación
dim_expire_date	DATE NOT NULL	Fecha de expiración

Tabla 20. Dimensión dim_insumo.

En la tabla 21 se muestra la estructura de la dimensión closing con su definición, atributos y tipos de datos.

3.1.6.8 Dimensión: dim_closing

dim_closing		
Descripción	Tabla que se almacena la información de las distintas variables que se evalúan en el requerimiento de Methanol inventory	
Tipo	Tabla de dimensión	
Columnas		
Id_closing	INT (11) NOT NULL AUTO_INCREMENT	Identificador único de la tabla dimensión dim_insumo
Característica	VARCHAR (100) NOT NULL	Característica
Subcaracterística	VARCHAR (100) NOT NULL	Sub- Característica
dim_version	INT (11) NOT NULL	Campo para la versión del dato
dim_effect_date	DATE NOT NULL	Fecha de creación
dim_expire_date	DATE NOT NULL	Fecha de expiración
Orden	INT (11) NOT NULL	Campo para preservar el orden por Sub-Característica en que se almacenan los registros en la dimensión, ya que se quiere en un orden en específico.

Tabla 21. Dimensión dim_closing.

En la tabla 22 se muestra la estructura de la dimensión fecha con su definición, atributos y tipos de datos.

3.1.6.9 Dimensión: dim_fecha

dim_fecha		
Descripción	Tabla que se almacena los periodos de los presupuestos aprobados y financiados de las ventas, ingresos y egresos.	
Tipo	Tabla de dimensión	
Columnas		
Id_fecha	INT (11) NOT NULL AUTO_INCREMENT	Identificador único de la tabla dimensión fecha
Year	VARCHAR (100) NOT NULL	Año
Mes	VARCHAR (100) NOT NULL	Mes
dim_version	INT (11) NOT NULL	Campo para la versión del dato
dim_effect_date	DATE NOT NULL	Fecha de creación
dim_expire_date	DATE NOT NULL	Fecha de expiración

Tabla 22. Dimensión dim_fecha.

En la tabla 23 se muestra la estructura de la dimensión operación con su definición, atributos y tipos de datos.

3.1.6.10 Dimensión: dim_operacion

dim_operacion		
Descripción	Tabla que se almacena la información de las distintas variables que se evalúan en el requerimiento de Operation	
Tipo	Tabla de dimensión	
Columnas		
Id_operacion	INT (11) NOT NULL AUTO_INCREMENT	Identificador único de la tabla dimensión dim_operacion
Característica	VARCHAR (100) NOT NULL	Característica
dim_version	INT (11) NOT NULL	Campo para la versión del datos
dim_effect_date	DATE NOT NULL	Fecha de creación
dim_expire_date	DATE NOT NULL	Fecha de expiración

Tabla 23. Dimensión dim_operacion.

En la tabla 24 se muestra la estructura de la tabla de hechos net back con su definición, atributos y tipos de datos.

3.1.6.11 Tabla de Hechos: fact_table_netback

fact_table_netback		
Descripción	Tabla que almacena toda la información de los presupuestos, gastos reales y proyectados de las ventas por región y corporación en Net Back US\$	
Tipo	Tabla de Hechos	
Columnas		
Id_fact_table_netback	INT (11) NOT NULL AUTO_INCREMENT	Identificador único de la tabla de hechos fact_table_netback
Id_fecha	INT (11) NOT NULL	Apuntador (foreigner key) a la tabla de fecha.
Id_corporacion	INT (11) NOT NULL	Apuntador (foreigner key) a la tabla de corporación.
Id_region	INT (11) NOT NULL	Apuntador (foreigner key) a la tabla de región
NumeradorB	INT (11) NOT NULL	Medida del Budget para el numerador.
DenominadorB	INT (11) NOT NULL	Medida del Budget para el denominador.
NumeradorR	INT (11) NOT NULL	Medida del Real para el numerador.
DenominadorR	INT (11) NOT NULL	Medida del Real para el denominador.

Tabla 24. Tabla de hechos fact_table_netback.

En la tabla 25 se muestra la estructura de la tabla de hechos mentor con su definición, atributos y tipos de datos.

3.1.6.12 Tabla de Hechos: fact_table_mentor

fact_table_mentor		
Descripcion	Tabla que almacena toda la información de los presupuestos, gastos reales y proyectados de las ventas, costo unitario, egresos e ingresos, balance.	
Tipo	Tabla de Hechos	
Columnas		
Id_fact_table_mentor	INT (11) NOT NULL AUTO_INCREMENT	Identificador único de la tabla de hechos fact_table_mentor
Id_dim_producto	INT (11) NOT NULL	Apuntador (foreigner key) a la tabla de producto.
id_fecha	INT (11) NOT NULL	Apuntador (foreigner key) a la tabla de fecha.
Id_region	INT (11) NOT NULL	Apuntador (foreigner key) a la tabla de región.
Id_corporacion	INT (11) NOT NULL	Apuntador (foreigner key) a la tabla de corporación.
Id_balancesheet	INT (11) NOT NULL	Apuntador (foreigner key) a la tabla de balance sheet.
Id_closing	INT (11) NOT NULL	Apuntador (foreigner key) a la tabla de closing.
Id_insumo	INT (11) NOT NULL	Apuntador (foreigner key) a la tabla de insumo.
Id_unitcost	INT (11) NOT NULL	Apuntador (foreigner key) a la tabla de unit cost.
Id_dim_production	INT (11) NOT NULL	Apuntador (foreigner key) a la tabla de producción.
Butget	DOUBLE	Medida de presupuesto
Real	DOUBLE	Medida de gasto real

Tabla 25. Tabla de hechos fact_table_meteor.

3.1.7 Diseño de procesos ETL

Para la transformación e integración de los datos que serán almacenados en el datawarehouse se realizaron una serie de pasos de transformación y depuración de datos para poder cargarlos en dicho datawarehouse. Entre los pasos más resaltantes para realizar las diferentes transformaciones fueron:



Excel Input

Paso de Entrada

Excel Input: Lee los datos de un archivo de Excel.



Get System Info

Paso de Entrada

Get Ssystem Info: Obtiene información de la fecha del sistema.



Table output

Paso de Salida

Table output: Escribe información en una base de datos usando una conexión a la misma.



Calculator

Paso de Transformación

Calculator: Proporciona operaciones para crear nuevos campos mediante la realización de cálculos simples.



Filter rows

Paso de Transformación

Filter rows: Filtra filas mediante ecuaciones simples.



Append streams

Paso de Transformación

Append streams: Une dos bloques de datos con la misma estructura.



Row denormaliser

Paso de Transformación

Row denormaliser: Nos permite transformas filas en columnas dado un valor clave. Dado que queremos operar entre filas, para ello lo llevamos a columnas.



Row Normaliser

Paso de Transformación

Row Normaliser: Nos permite transformas columnas en filas.



Group by

Paso de Transformación

Group by: Agrupar valores por una característica determinada.



Select values

Paso de Transformación

Select values: Escoge los valores que interesan. Se utiliza para seleccionar y modificar el orden exacto y el nombre en que los campos tienen que ser colocados en las filas de salida, también se utiliza para eliminar los campos que se quiere retirar.



Sort rows

Paso de Transformación

Sort rows: Ordena ascendente y descendente una serie de campos que se especifican.



Dimension lookup/update

Paso de Datawarehouse

Dimension Lookup/update Actualización de una dimensión que cambian lentamente en el almacén de datos. Y también es utilizado para buscar información en una dimensión.



Value Mapper

Paso de Transformación

Value Mapper Busca valores en específico y lo transforma en uno solo.

Para la carga de los datos en el datawarehouse estos pasan por tres procesos de extracción, transformación y carga (ETL) para llegar a ser organizados dentro del mismo. A continuación se describen los procesos principales, ya que en su mayoría se repiten para cumplir con la población del datawarehouse.

En la figura 17 podemos observar uno de los procesos de ETL, primero se cargan los datos desde los archivos .txt que son nuestras fuentes de datos, se leen las variables manejadas y sus valores para budget y real, el año y mes. Luego se busca una serie de valores en específicos y se transforma en uno solo, ya que existen variables que se necesitan unificar porque en el datawarehouse se requiere esa variable unificada (sumadas), luego éstas son ordenadas, se eligen los valores que se requieren para ser colocadas en las filas de salida y se renombran los campos con el nombre que se necesitan para dichas filas de salida, para este caso podemos observar que se tienen dos entradas de archivos .txt es por ello que se realiza un paso de unión de esos bloques. Otro de los pasos fundamentales de dicho ETL es el proceso de des-normalización y normalización, este proceso se realiza ya que se necesita operar entre filas, es por ello que éstas filas son pasadas para columnas (procesos de des-normalización), se realizan las operaciones necesarias y luego son pasadas nuevamente a filas (proceso de normalización).

Como se requiere llenar la tabla de hechos, se realiza una búsqueda de claves sustitutas para llenar los valores Budget, Real en la tabla de hechos.

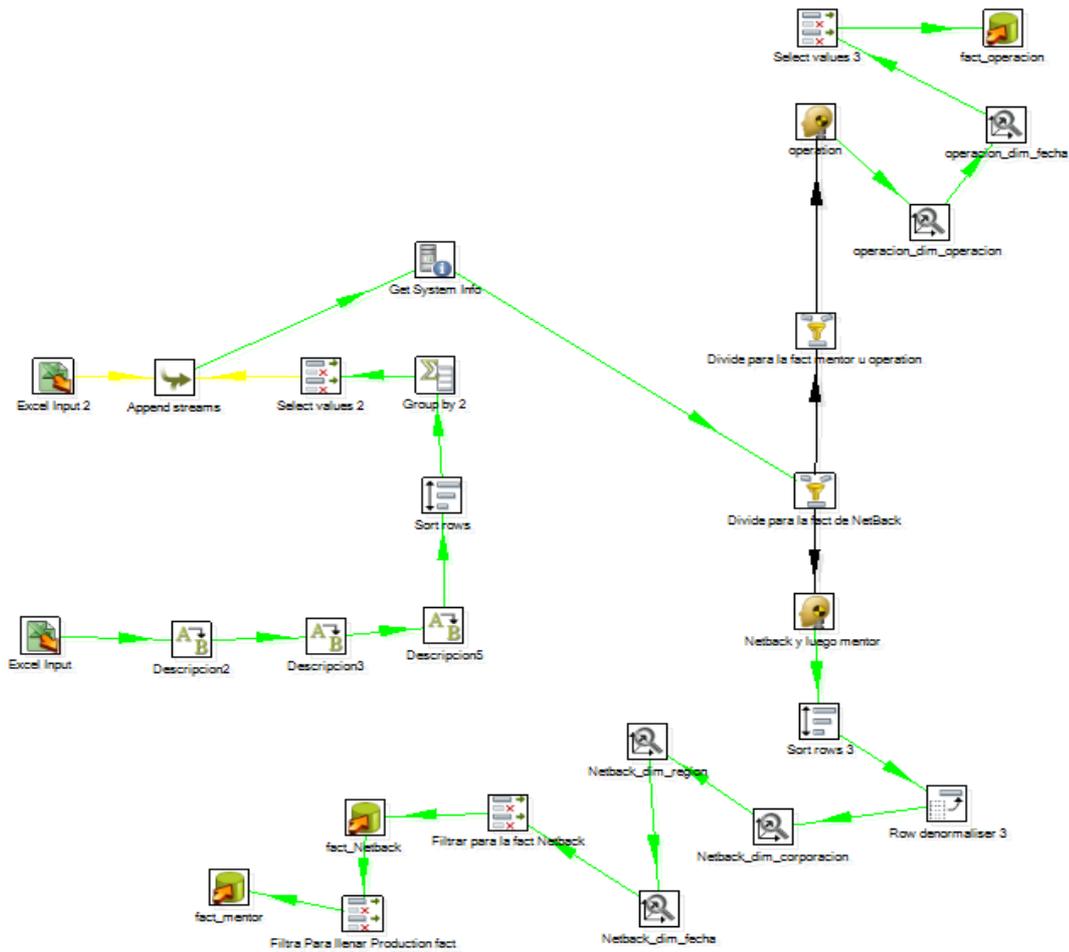


Figura 17: Primer subproceso de ETL para la fact_table

En el siguiente proceso se cargan los datos desde los archivos .txt, se leen las variables manejadas y sus valor para budget y real, el año y mes, luego se busca los valores en específicos y se transforma en uno solo, ya que existen variables que se necesitan unificar ya que en el datawarehouse se requiere de una sola, luego éstas son ordenadas y luego se eligen los valores que se requieren para ser colocadas en las filas de salida y se renombrar los campos con el nombre que se necesitan para dichas filas de salida. Como se requiere llenar la tabla de hechos, se

realiza una búsqueda de claves sustitutas para llenar los valores Budget, Real en la tabla de hechos. En la figura 18 podemos ver los pasos anteriormente descritos.

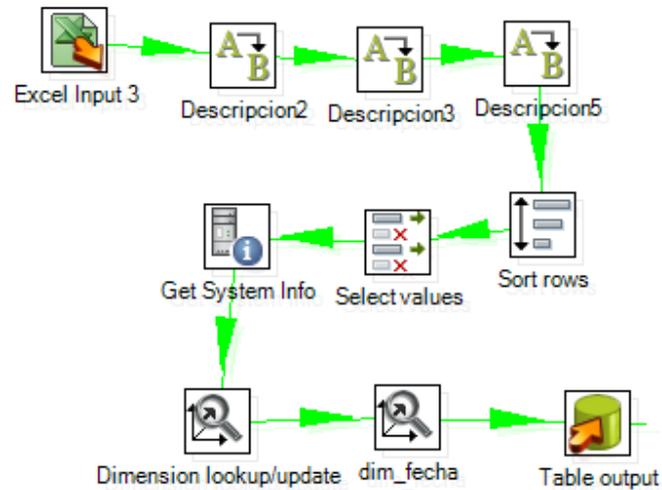


Figura 18. Segundo subprocesso de ETL para la fact_table.

En este proceso destaca el uso de los filtros para poder bifurcar filas en las cuales se necesitan realizar operaciones, una vez filtrado los datos en uno de los flujos se realiza operaciones de multiplicación entre los elementos de las filas, es por ello que anteriormente se aplica el proceso de des-normalización, para poder operar con esos elementos de fila. En la figura 19 podemos observar estos pasos.

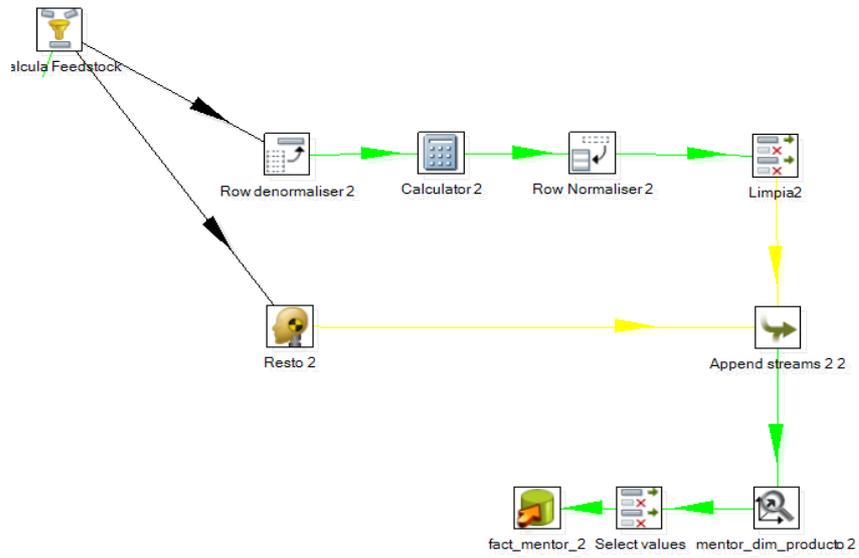


Figura 19. Tercer subproceso de ETL para la fact_table.

A continuación se muestra los procesos unificados para la población del datawarehouse.

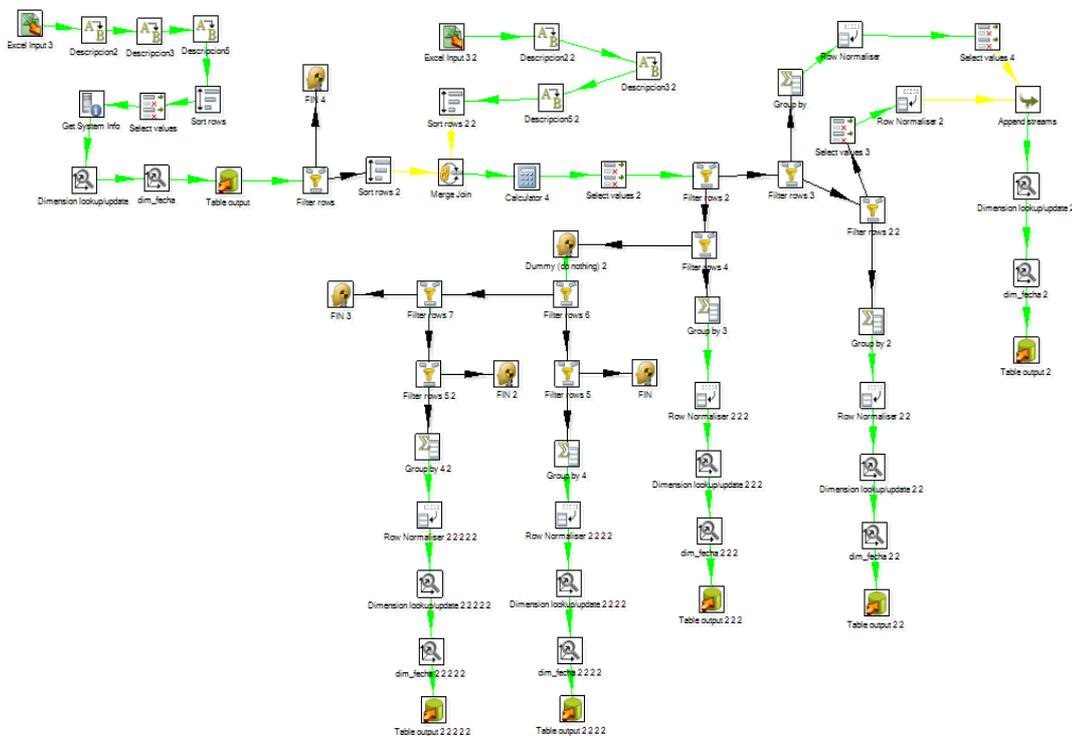


Figura 20. Proceso general de ETL para la fact_table.

3.1.8 Especificación de la aplicación de usuario final

Uno de los puntos importantes del desarrollo de la solución de inteligencia de negocio, es el desarrollo de los reportes, es por ello que se determina como serán estos reportes, como estarán parametrizados para que el usuario de respuestas a sus necesidades.

Para nuestro caso, los reportes tienen un formato estándar, el cual contiene su título, una gráfica y las diferentes variables que dan respuesta a una necesidad del departamento, éstas variables están asociadas a métricas como lo son Budget, Real y la Variación. El usuario debe de determinar que período de fecha quiere el reporte, este período de fecha va a tener el formato de año-mes (Ejemplo: 2009-01), además debe de determinar que formato requiere el reporte el cual puede ser como contenido Web, PDF, Excel, entre otros. A continuación en la figura 21 se muestra el *template* de cómo serían los reportes que manejara el usuario final.

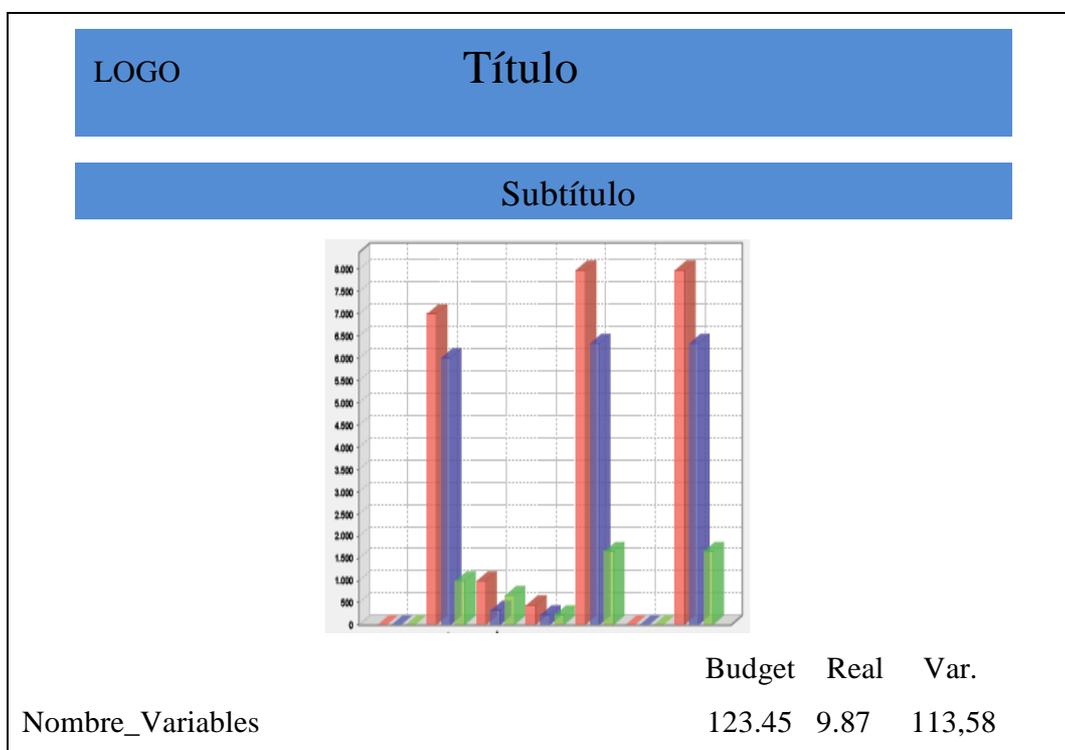


Figura 21. Template de los Reportes.

El usuario también cuenta con la opción de que reportes desea ver, esto los puede seleccionar mediante el servidor Web de Pentaho, donde se le mostrará una carpeta que contiene cada uno de los reportes que desea, solo el usuario escogerá que reporte desea ver.

3.1.9 Desarrollo de la aplicación de usuario final

Para realizar la aplicación de usuario se procede primero a la creación de los cubos, los cuales contienen las dimensiones, tablas de hechos y métricas que fueron descritas en el apartado 3.1.5 y 3.1.6, para la realización de dichos cubos se utilizó el esquema Workbeach. Estos cubos pueden ser visualizados al ingresar en el servidor Web de Pentaho a través de la herramienta JPivot, una vez visualizado el cubo se debe realizar los filtros, seleccionando las dimensiones con los atributos y métricas que desea visualizar en un analítico, es decir, nos permite modelar las representaciones en base a consultas para crear un cuadro de mando integral que sirve de base para la presentación de informes., como se muestra en la figura 22.

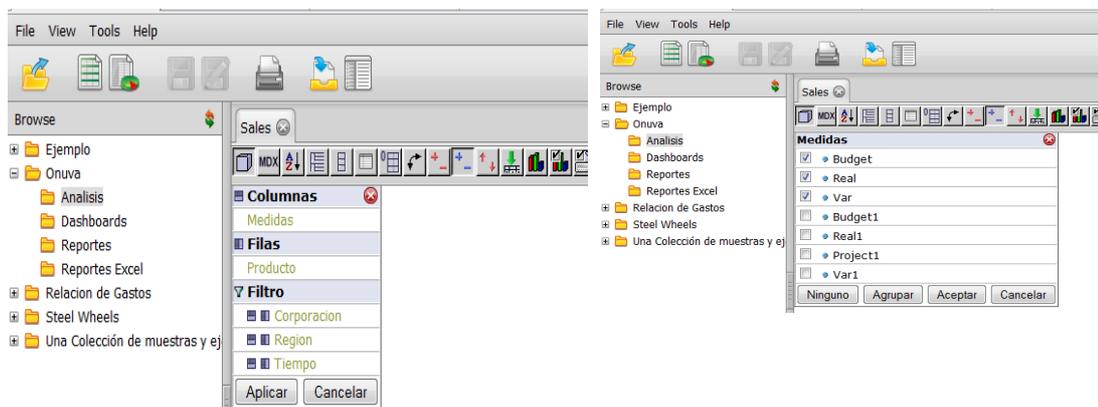


Figura 22. Visualización de un cubo en JPivot.

Una vez seleccionada las dimensiones con sus atributos y métricas, la herramienta nos generará un código MDX que es el que se copia en los archivos

que nos permite visualizar los analíticos en el servidor Web, para presentar informes. En la figura 23 se muestra cómo la herramienta nos genera el código MDX.

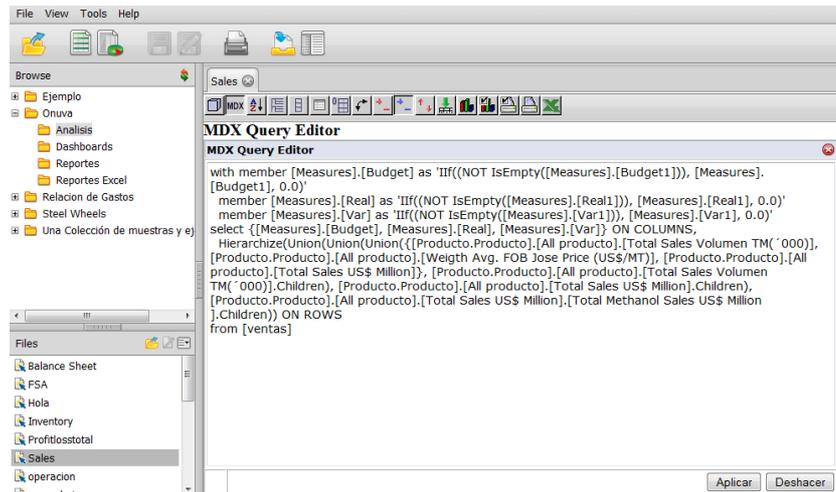


Figura 23. Query MDX generado por JPivot.

A continuación en la figura 24 se muestra en Jpivot el requerimiento de Sales, el cual nos muestra el cubo que contiene la consulta multidimensional.

Producto	Medidas		
	Budget	Real	Var
Total Sales Volumen TM('000)	3.670,0	4.446,0	-776,0
+ DOMESTIC	1.369,0	1.337,0	32,0
+ EXPORT	2.301,0	3.109,0	-808,0
+ Weigth Avg. FOB Jose Price (US\$/MT)	2.005,0	789,0	1.216,0
Total Sales US\$ Million	10.312,0	7.671,0	2.641,0
+ Hydrogen Rich Gas	1.332,0	672,0	660,0
+ Total Methanol Sales US\$ Million	8.980,0	6.999,0	1.981,0
DOMESTIC	6.995,0	6.007,0	988,0
EXPORT	1.985,0	992,0	993,0

Slicer:

Figura 24. Analítico de Sales.

En la figura 25 se muestra en Jpivot el requerimiento de Operación, el cual nos muestra el cubo que contiene la consulta multidimensional.

Operacion	Medidas		
	Budget	Real	Var
Production MT ('000)	5.666,0	7.427,0	-1.761,0
Scheduled Maintenance Days	3.202,0	1.441,0	1.761,0
Non Programmed Maintenance Days	4.535,0	2.682,0	1.853,0
Production Days	1.022,0	790,0	232,0
Service Factor %	90,0	80,0	10,0
Capacity Factor %	90,0	80,0	10,0

Slicer:

Figura 25. Analítico de Operation.

En la figura 26 se muestra en Jpivot el requerimiento de balance sheet, el cual nos muestra el cubo que contiene la consulta multidimensional.

Balance	Medidas		
	Budget	Real	Var
Total Current Assents	8.624,0	8.306,0	318,0
Cash and Securities	2.879,0	1.549,0	1.330,0
Account Receivables	986,0	325,0	661,0
Inventories	909,0	569,0	340,0
Tax Credits	0,0	0,0	0,0
Deferred income taxes	1.853,0	2.881,0	-1.028,0
Prepaid expenses	1.997,0	2.982,0	-985,0
Total Assets	889,0	449,0	440,0
Fixed assets	444,0	224,0	220,0
Projects in progress	445,0	225,0	220,0
Deferred charges	0,0	0,0	0,0
Others assets	0,0	0,0	0,0
Total Liabilities and Equity	6.474,0	4.151,0	2.323,0
Total Liabilities	2.676,0	3.105,0	-429,0
Sub Total	2.676,0	3.105,0	-429,0
Account payables & other current liabilit	1.116,0	340,0	776,0
Deferred taxes	1.113,0	2.538,0	-1.425,0

Figura 26. Analítico de Balance Sheet.

En la figura 27 se muestra en Jpivot el requerimiento de FSA, el cual nos muestra el cubo que contiene la consulta multidimensional.

Production	Medidas		
	● Budget	● Real	● Var
<input type="checkbox"/> Total Cash Revenue	0,0	0,0	0,0
<input type="checkbox"/> Gross Revenue	0,0	0,0	0,0
<input type="checkbox"/> Variation in account receivable	0,0	0,0	0,0
<input type="checkbox"/> Total Cash Cost	0,0	0,0	0,0
<input type="checkbox"/> Sub Total	0,0	0,0	0,0
Production Expenses	0,0	0,0	0,0
Sales and Administrative Expenses	0,0	0,0	0,0
Financial Expenses and Other Income	0,0	0,0	0,0
Other	0,0	0,0	0,0
<input type="checkbox"/> Variation in account payables	0,0	0,0	0,0
<input type="checkbox"/> Total net cash margin	0,0	0,0	0,0
<input type="checkbox"/> % of supplemental amount according to level	0,0	0,0	0,0
<input type="checkbox"/> Feedstock Supplemental Amount	0,0	0,0	0,0

Figura 27. Análisis de FSA.

En la figura 28 se muestra en Jpivot el requerimiento de Inventory, el cual nos muestra el cubo que contiene la consulta multidimensional.

Inventory	Medidas		
	● Budget	● Real	● Var
[-] Closing	9.185,0	1.398,0	7.787,0
Openning	3.937,0	2.105,0	1.832,0
Production	2.668,0	1.348,0	1.320,0
Methanol Swap	5.537,0	3.352,0	2.185,0
Sales	-2.957,0	-5.407,0	2.450,0

Figura 28. Análítico de Inventory.

En la figura 29 se muestra en Jpivot el requerimiento de Profit and Loss (Total), el cual nos muestra el cubo que contiene la consulta multidimensional.

Insumo	Medidas		
	● Budget	● Real	● Var
[+] Sales	0,0	0,0	0,0
[+] Total Cost of Sales	0,0	0,0	0,0
[+] Gross Margin	0,0	0,0	0,0
[+] Total Sales, Admin and Other	0,0	0,0	0,0
[+] Profit Before Tax	0,0	0,0	0,0
[+] Total Tax	0,0	0,0	0,0
[+] Net Profit	0,0	0,0	0,0

Slicer:

Figura 29. Análítico de Profit and Loss (Total).

En la figura 30 se muestra en Jpivot el requerimiento de OPEX Production, el cual nos muestra el cubo que contiene la consulta multidimensional.

Production	Medidas		
	● Budget	● Real	● Var
<input type="checkbox"/> Total OPEX Production	15.987,0	12.906,0	3.081,0
<input type="checkbox"/> Total Variable Cost	10.628,0	6.137,0	4.491,0
Natural Gas	4.102,0	2.104,0	1.998,0
Chemicals	1.430,0	549,0	881,0
Electricity	3.198,0	256,0	2.942,0
Water	344,0	469,0	-125,0
FSA	1.554,0	2.759,0	-1.205,0
<input type="checkbox"/> Total Fixed Cost	5.359,0	6.769,0	-1.410,0
Industrial Services	2.665,0	5.295,0	-2.630,0
Condominium	1.357,0	797,0	560,0
Laboratory Supplies	1.337,0	677,0	660,0

Slicer:

Figura 30. Análítico de OPEX Production.

En la figura 31 se muestra en Jpivot el requerimiento de OPEX Administrative, el cual nos muestra el cubo que contiene la consulta multidimensional.

Opexadmin	Medidas		
	● Budget	● Real	● Var
<input type="checkbox"/> All opexadmin	25.172,0	14.304,0	10.868,0
<input type="checkbox"/> Total Sales Administrative and Others	13.519,0	7.824,0	5.695,0
<input type="checkbox"/> Total OPEX Sales and Administrative	11.733,0	6.918,0	4.815,0
<input type="checkbox"/> Total Sales	1.429,0	548,0	881,0
Product Inspection	443,0	223,0	220,0
Custom Broker	986,0	325,0	661,0
Demurrages	0,0	0,0	0,0
<input type="checkbox"/> Total Personnel	4.932,0	1.587,0	3.345,0
Personnel & Benefits	443,0	223,0	220,0
Personnel & Benefits - Expansion	0,0	0,0	0,0
Training	855,0	566,0	289,0
Assigned personnel	0,0	0,0	0,0
<input type="checkbox"/> Total General	3.371,0	3.998,0	-627,0
Business Travel expenses	0,0	0,0	0,0
Expansion project	0,0	0,0	0,0

Figura 31. Analítico de OPEX Administrative.

En la figura 31 se muestra en Jpivot el requerimiento de Sales by región Production MT, el cual nos muestra el cubo que contiene la consulta multidimensional.

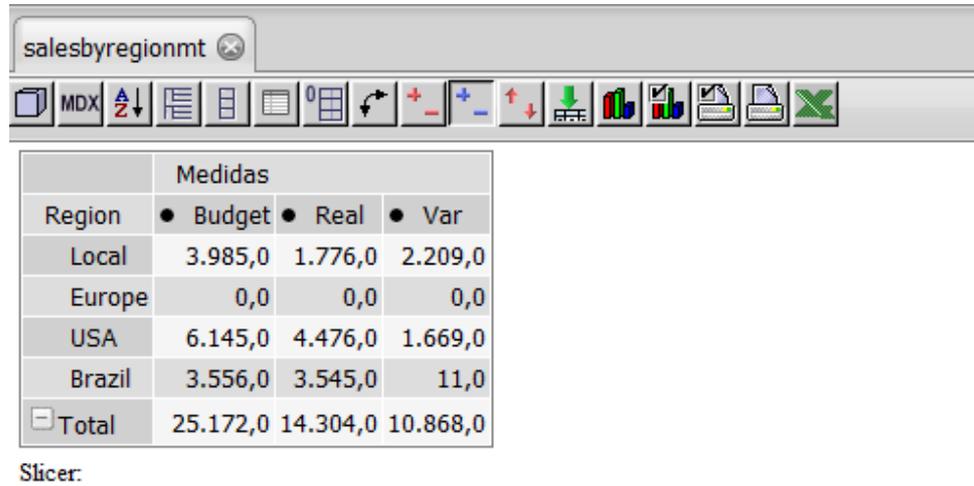


Figura 32. Analítico de OPEX Sales Region Production MT.

En la figura 33 se muestra en Jpivot el requerimiento de Sales by off Taker Production MT, el cual nos muestra el cubo que contiene la consulta multidimensional.

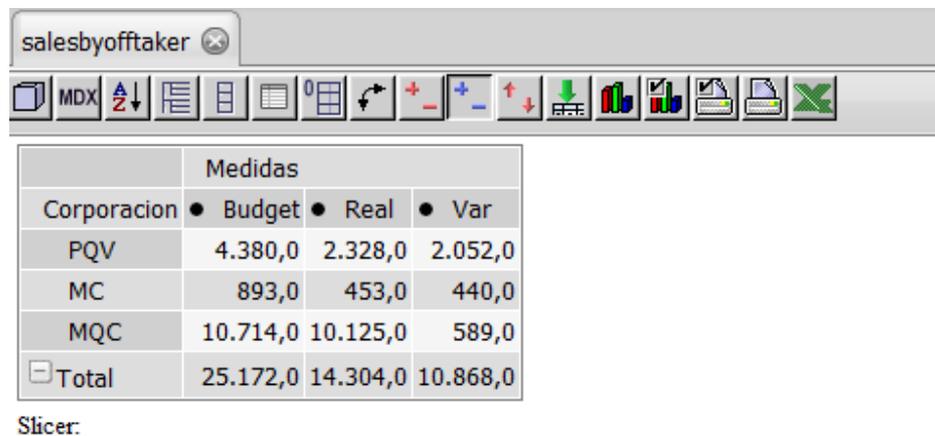


Figura 33. Sales by Off taker Production MT.

En la figura 34 se muestra en Jpivot el requerimiento de Unit Cost, el cual nos muestra el cubo que contiene la consulta multidimensional.

unitcost	
unit	Medidas
	● Budget ● Real ● Var
+ Production MT (000)	1.597,0 3.104,0 -1.507,0
- Total Unit Cost	14.390,0 9.802,0 4.588,0
- Production Unit Cost	12.366,0 6.693,0 5.673,0
- Total Variable Cost	10.361,0 5.904,0 4.457,0
- Sub Total Gas	6.719,0 5.424,0 1.295,0
Gas	2.720,0 1.656,0 1.064,0
FSA	3.999,0 3.768,0 231,0
+ Chemicals and Others	3.642,0 480,0 3.162,0
+ Fixed cost and + plant expenses	2.005,0 789,0 1.216,0
+ S&A expenses	466,0 346,0 120,0

Figura 34. Análisis de Unit Cost.

En la figura 35 se muestra en Jpivot el requerimiento de Opex Plant, el cual nos muestra el cubo que contiene la consulta multidimensional.

Plant	Medidas		
	● Budget	● Real	● Var
<input type="checkbox"/> Total Cost of Sales	15.987,0	12.906,0	3.081,0
<input type="checkbox"/> Total Production Cost	15.987,0	12.906,0	3.081,0
<input type="checkbox"/> Total OPEX Plant	15.987,0	12.906,0	3.081,0
<input type="checkbox"/> Total Personnel	15.987,0	12.906,0	3.081,0
Personnel & Benefits	15.987,0	12.906,0	3.081,0
Training	0,0	0,0	0,0
<input type="checkbox"/> Total Maintenance	0,0	0,0	0,0
Spare parts & Supplies	0,0	0,0	0,0
Contrated Services	0,0	0,0	0,0
No Capitalizable investments	0,0	0,0	0,0
<input type="checkbox"/> Total General	0,0	0,0	0,0
No recoverable VAT	0,0	0,0	0,0
Building and office maintenance	0,0	0,0	0,0
Cleaning Services	0,0	0,0	0,0
Safety supplies	0,0	0,0	0,0
Business travel expenses	0,0	0,0	0,0
Telephone, fax and courier	0,0	0,0	0,0
Vehicles maintenance	0,0	0,0	0,0

Figura 35. Análisis de OPEX Plant.

El usuario también se le proporcionan reportes con diferentes salidas de formatos los cuales pueden ser página Web, PDF, Excel, entre otros. Estos fueron realizados en Pentaho Reporting Designer el cual después de realizar el diseño del reporte como se puede ver en la figura 36 una interfaz de diseño de reportes, éste nos genera un archivo que es usado para configurar los parámetros de entradas, query, conexiones, entre otras cosas.

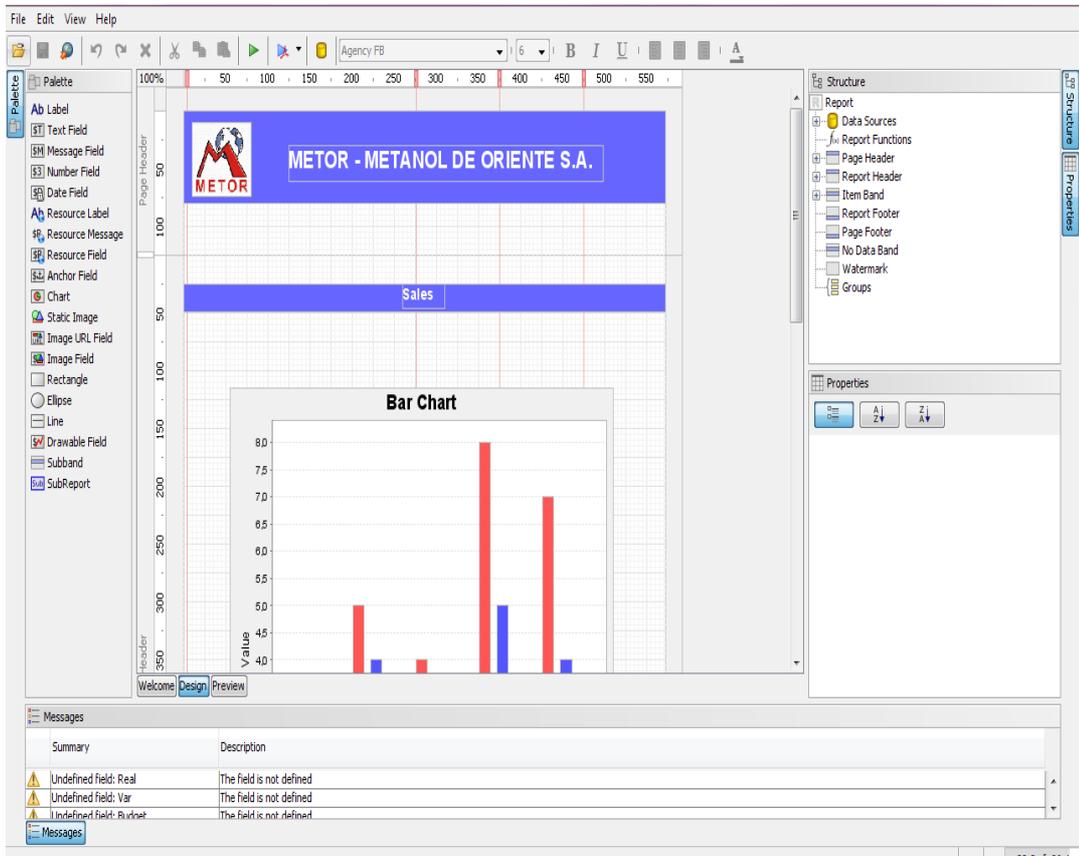


Figura 36. Diseño de un reporte en Report Designer.

Una vez realizados los reportes el usuario cuenta con una interfaz del servidor Web de Pentaho, el cual le permite en un principio escoger que reporte desea ver. En la figura 37 se muestra la interfaz del servidor Web de Pentaho, donde se puede visualizar cada uno de los reportes que se generan de la solución de inteligencia de negocio.

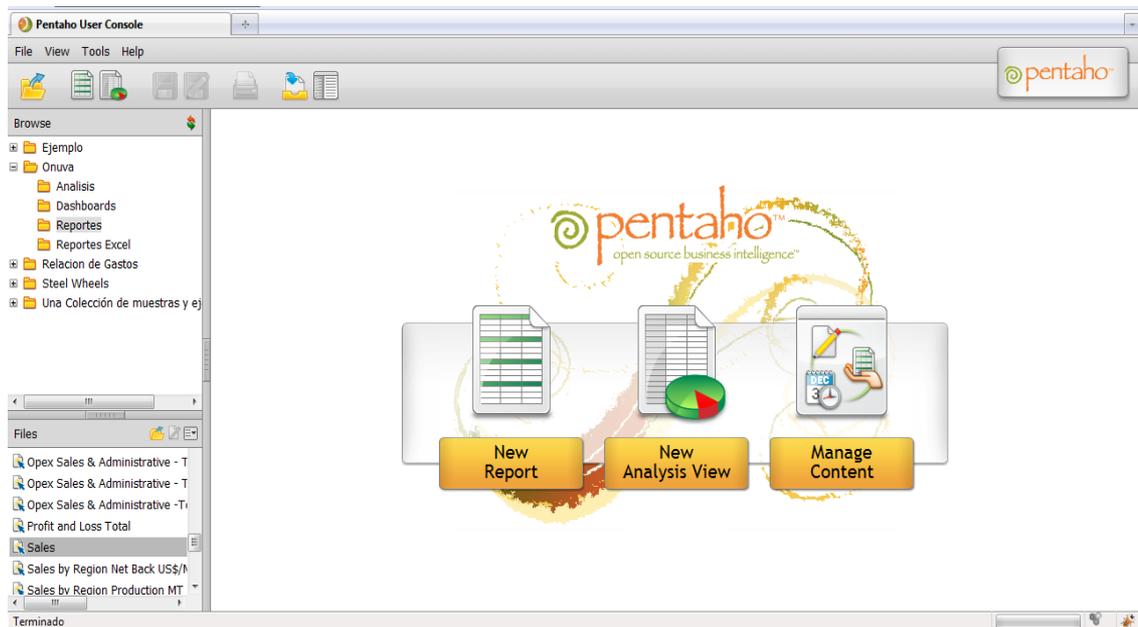


Figura 37. Interfaz del servidor We de Pentaho.

En la figura 38 se muestra la interfaz donde el usuario selecciona la carpeta que contendrá sus reportes, una vez seleccionada la carpeta de los reportes, en la parte inferior del lado izquierdo aparecerán cada uno de los reportes que cumple con los requerimientos ya planteados, el usuario solo tiene que escoger cuál de ellos quiere ver.

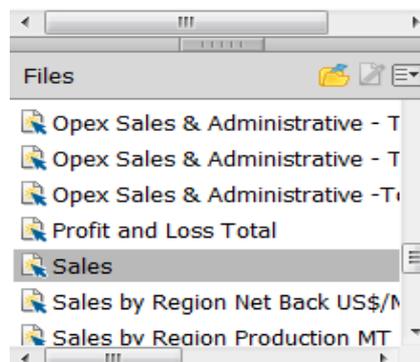


Figura 38. Interfaz de la ubicación de la solución en Pentaho.

Una vez seleccionado el reporte que desea ver, el usuario debe indicar la fecha en la que desea ver el reporte, esta fecha es presentada en una lista desplegable de manera que en esa lista solo aparecerá el año y mes que contiene data de dicho reporte y también el usuario debe indicar cuál es el tipo de salida que requiere ya que puede ver el reporte en formato PDF, Web, Excel, entre otros, en la figura 39 se muestra la interfaz de lo que se explico anteriormente.

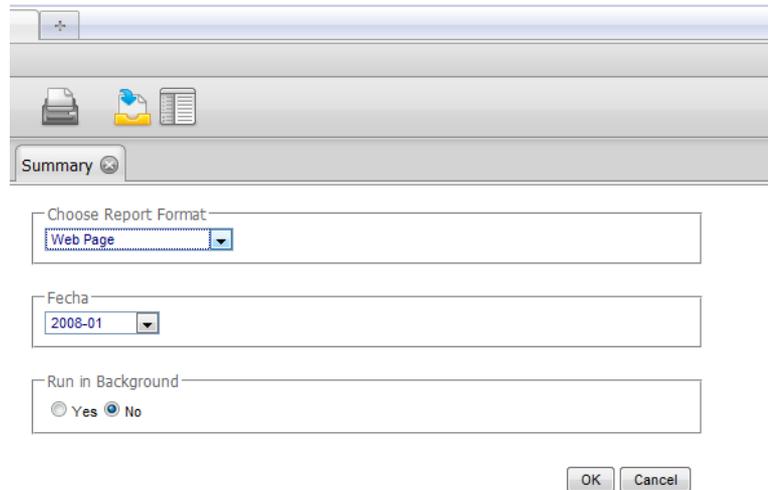


Figura 39. Parámetros de los reportes.

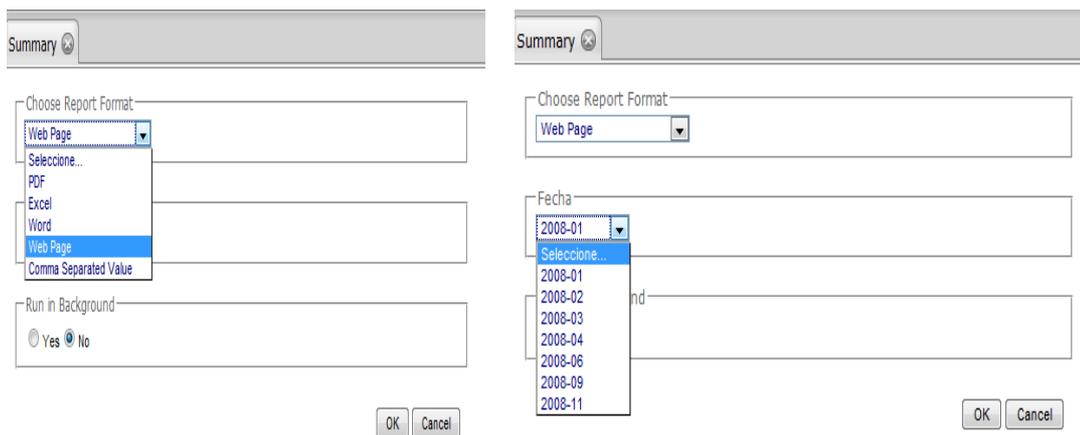


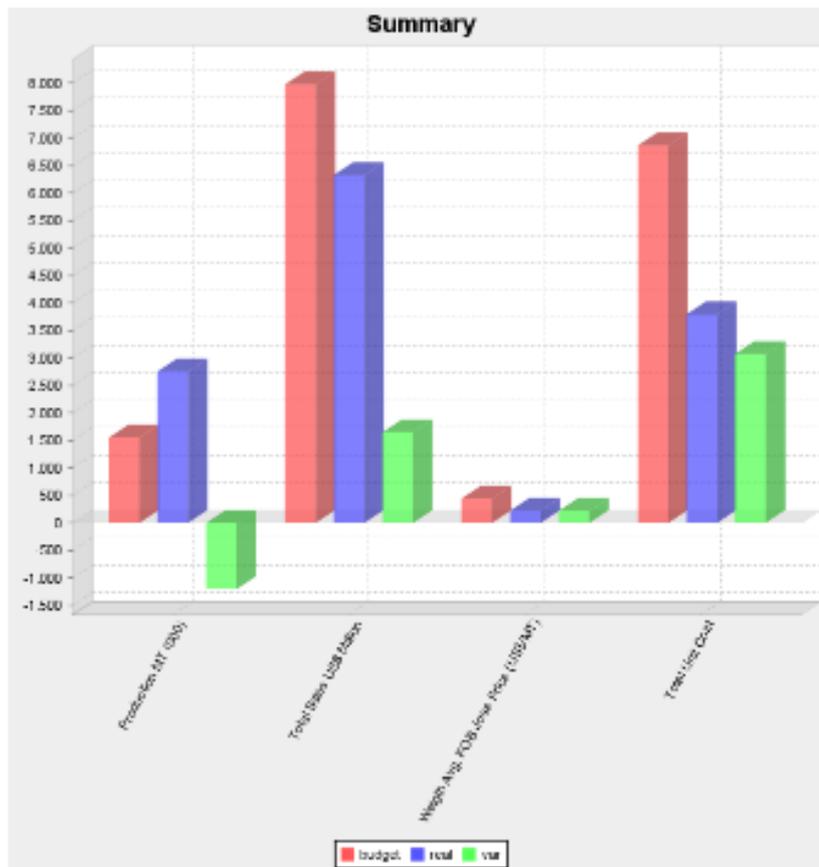
Figura 40. Despliegue de los parámetros de los reportes .

Una vez que el usuario seleccione los parámetros que se necesitan para mostrar los reportes, se muestra el reporte. A continuación se muestra en la figura 41 el reporte de summary, dado así repuestas a los requerimientos planteados.



METOR - METANOL DE ORIENTE S.A.

Summary



	Budget	Real	Var
Production MT (000)	1553.0	2758.0	(-1205.0)
Total Sales US\$ Million	7981.0	6332.0	1649.0
Weight Avg. FOB Jose Price (US\$/MT)	443.0	223.0	220.0
Total Unit Cost	6871.0	3797.0	3074.0

Figura 41. Reporte de Summary.

En la figura 42 se muestra el reporte de operations, dado así repuestas a los requerimientos planteados.

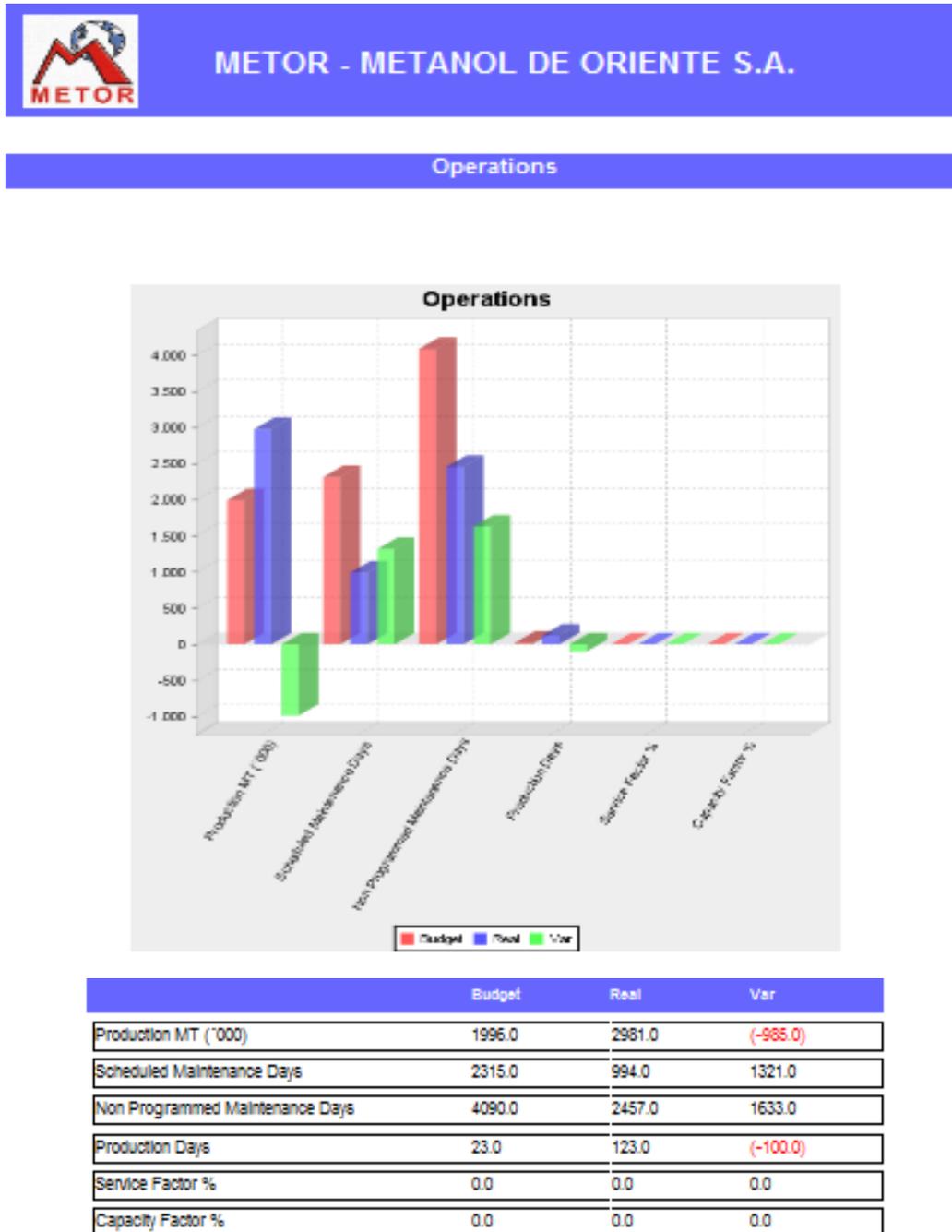


Figura 42. Reporte de Operations.

En la figura 43 se muestra el reporte de sales, dado así repuestas a los requerimientos planteados.



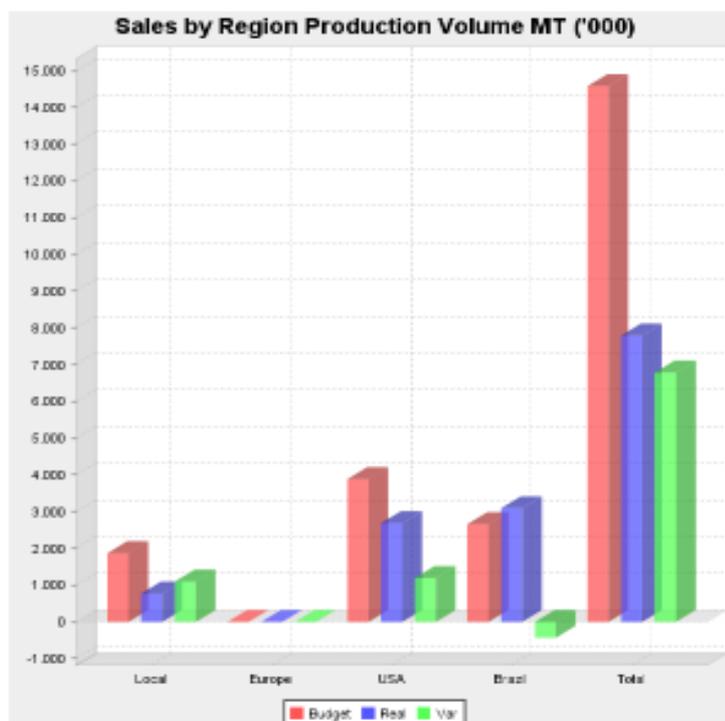
Figura 43. Reporte de Sales.

En la figura 44 se muestra el reporte de sales by región Productio Volumen MT, dado así repuestas a los requerimientos planteados.



METOR - METANOL DE ORIENTE S.A.

Sales by Region Production Volume MT ('000)



	Budget	Real	Var
Local	1872.0	771.0	1101.0
Europe	0.0	0.0	0.0
USA	3888.0	2691.0	1197.0
Brazil	2664.0	3093.0	(-429.0)
Total	14582.0	7794.0	6788.0

Figura 44. Reporte de Sales by Región Production Volumen MT ('000).

En la figura 45 se muestra el reporte de sales by región net back, dado así repuestas a los requerimientos planteados.

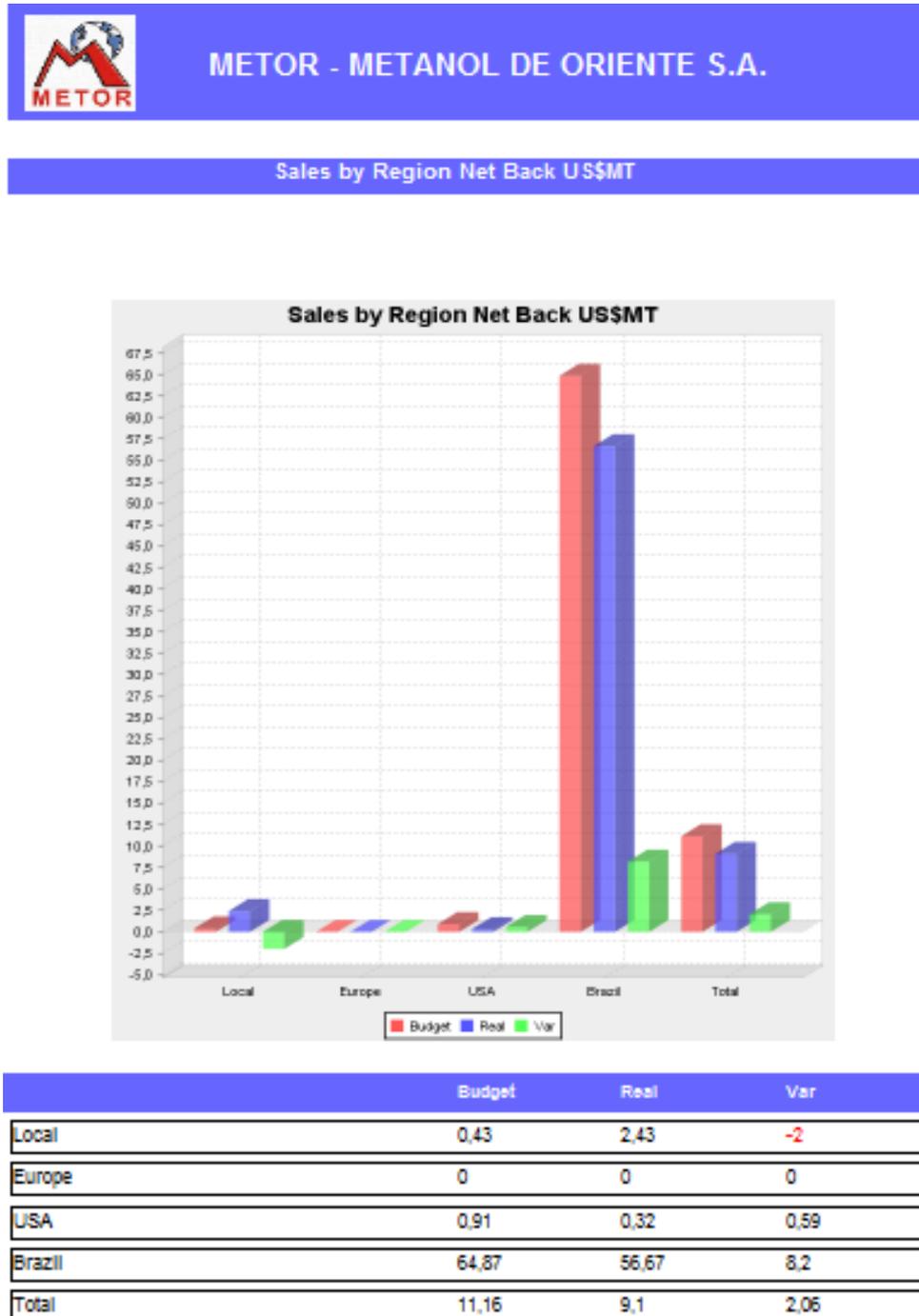


Figura 45. Reporte de Sales by Región Net Back US\$/MT.

En la figura 46 se muestra el reporte de sales by off taker production MT, dado así repuestas a los requerimientos planteados.



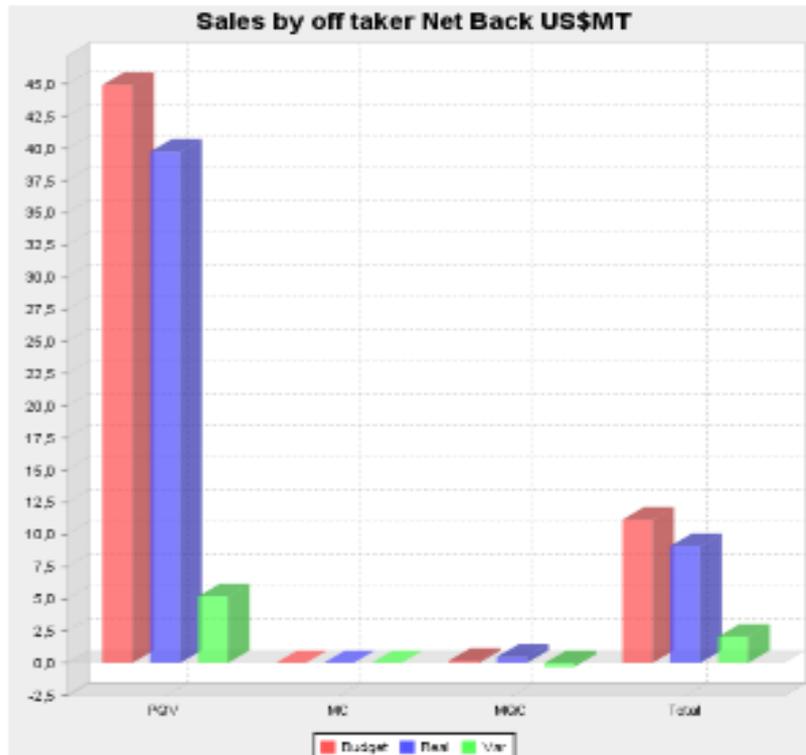
Figura 46. Reporte de Sales by off taker Production Volumen MT ('000).

En la figura 47 se muestra el reporte de sales by off taker net back, dado así repuestas a los requerimientos planteados.



METOR - METANOL DE ORIENTE S.A.

Sales by off taker Net Back US\$MT

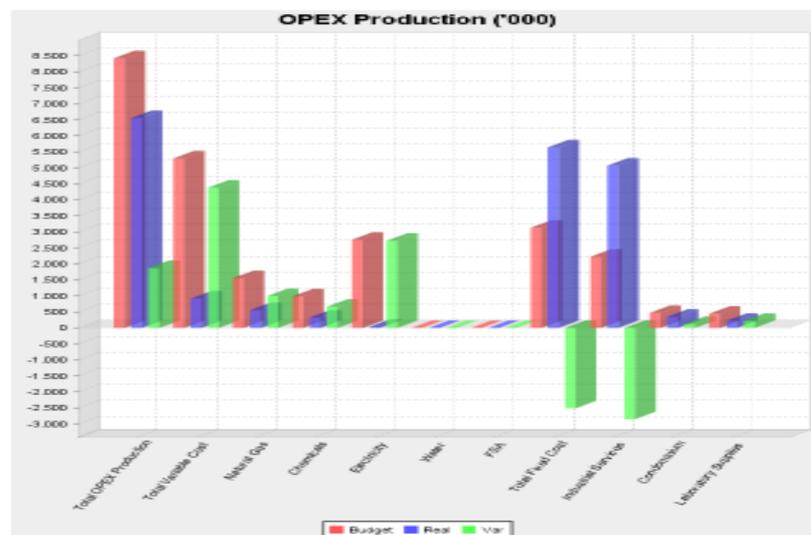


	Budget	Real	Var
PQV	44,95	39,76	5,19
MC	0	0	0
MQC	0,19	0,48	-0,28
Total	11,16	9,1	2,06

Figura 47. Reporte de Sales by off taker Net Back US\$/MT.

En la figura 48 se muestra el reporte de Opex production, dado así repuestas a los requerimientos planteados.

METOR - METANOL DE ORIENTE S.A.
OPEX Production ('000)



	Budget	Real	Var
Total OPEX Production	8424.0	6555.0	1869.0
Total Variable Cost	5295.0	916.0	4379.0
Natural Gas	1554.0	558.0	996.0
Chemicals	996.0	325.0	661.0
Electricity	2755.0	33.0	2722.0
Water	0.0	0.0	0.0
FSA	0.0	0.0	0.0
Total Fixed Cost	3129.0	5639.0	(-2510.0)
Industrial Services	2220.0	5070.0	(-2850.0)
Condominium	466.0	346.0	120.0
Laboratory Supplies	443.0	223.0	220.0

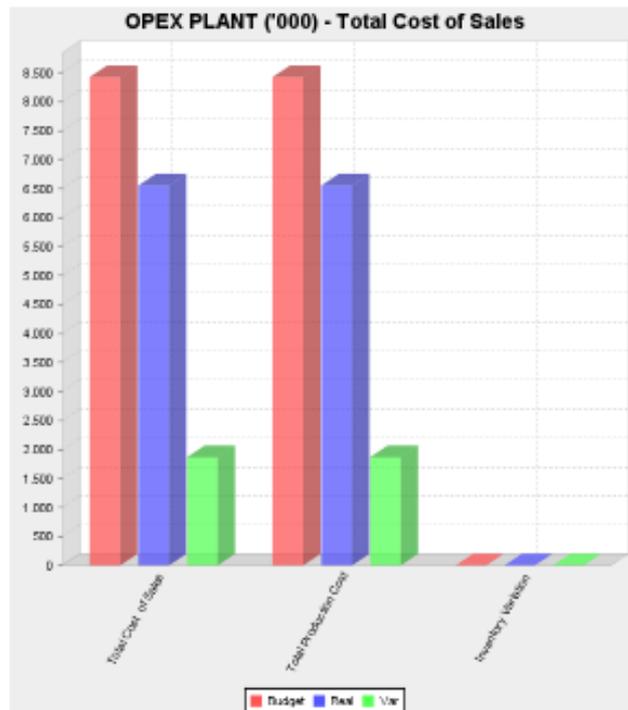
Figura 48. Reporte de Profit and Loss (OPEX Production).

En la figura 49 se muestra el reporte de Opez plant, dado así repuestas a los requerimientos planteados.

METOR - METANOL DE ORIENTE S.A.

Profit & Loss

OPEX PLANT ('000) - Total Cost of Sales



	Budget	Real	Var
Total Cost of Sales	8424.0	6555.0	1869.0
Total Production Cost	8424.0	6555.0	1869.0
Inventory Variation	0.0	0.0	0.0

Figura 49. Reporte de Profit and Loss (OPEX PLANT).

En la figura 50 se muestra el reporte de Opex Administrative, dado así repuestas a los requerimientos planteados.

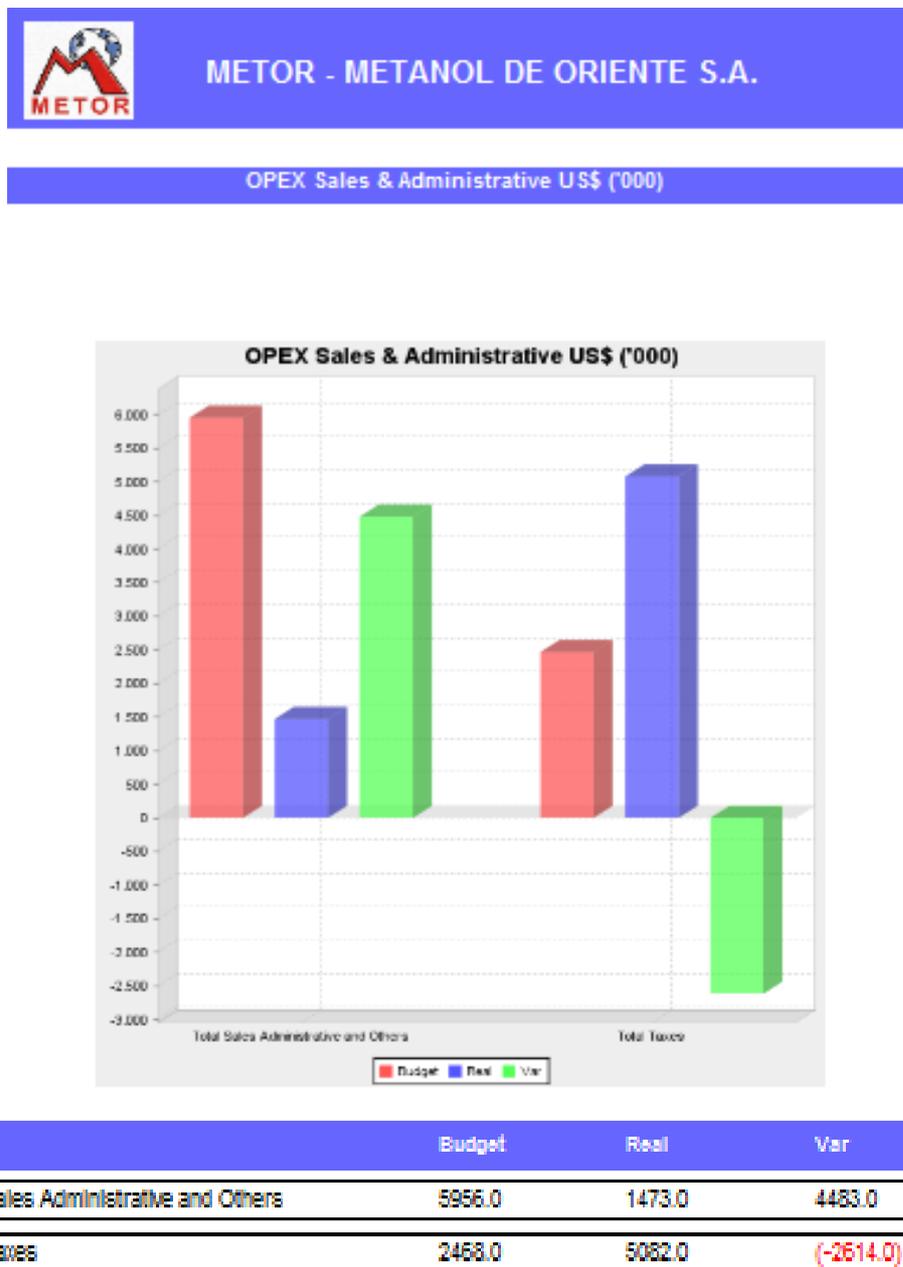
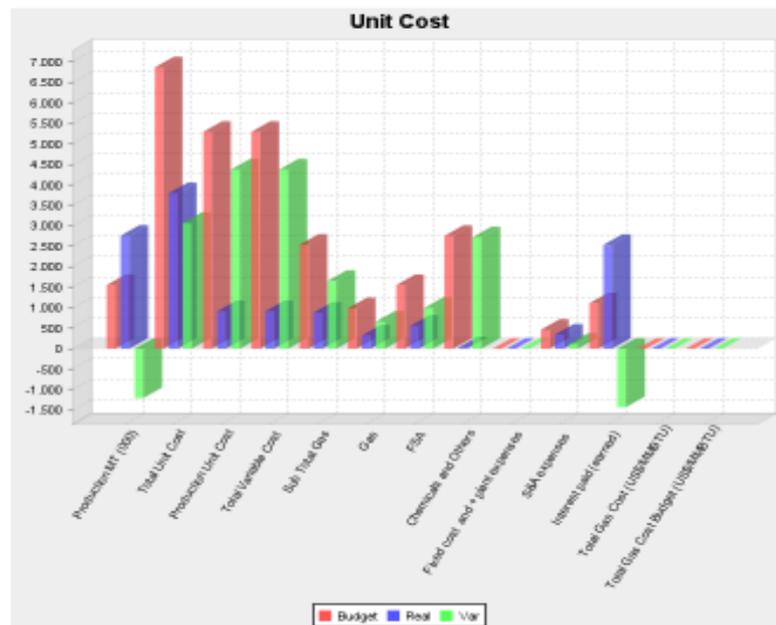


Figura 50. Reporte de Profit and Loss (OPEX Administrative).

En la figura 51 se muestra el reporte de unit cost, dado así repuestas a los requerimientos planteados.



Unit Cost



	Budget	Real	Var
Production MT (000)	1553.0	2758.0	(-1205.0)
Total Unit Cost	6871.0	3797.0	3074.0
Production Unit Cost	5295.0	916.0	4379.0
Total Variable Cost	5295.0	916.0	4379.0
Sub Total Gas	2540.0	883.0	1657.0
Gas	996.0	325.0	661.0
FSA	1554.0	558.0	996.0
Chemicals and Others	2755.0	33.0	2722.0
Fixed cost and + plant expenses	0.0	0.0	0.0
S&A expenses	466.0	346.0	120.0
Interest paid (earned)	1110.0	2535.0	(-1425.0)
Total Gas Cost (US\$/MMBTU)	0.0	0.0	0.0
Total Gas Cost Budget (US\$/MMBTU)	0.0	0.0	0.0

Figura 51. Reporte de Unit Cost.

En la figura 52 se muestra el reporte de balance sheet, dado así repuestas a los requerimientos planteados.



Figura 52. Reporte de Balance Sheet.

3.1.10 Despliegue

Una vez desarrollado cada uno de los reportes, se necesita verificar la infraestructura que tiene el cliente, en este paso se va a verificar la configuración de hardware, conexión a las bases, acceso a intranet o internet, direcciones LAN (si no son dinámicamente asignadas), auditorías de tecnología sobre las configuraciones en las que se encuentran las PCs, se prevé actualizaciones de hardware y software, se verifica la seguridad (login de red y base de datos), se realizan pruebas de procedimientos de instalación en una variedad de máquinas, y se planifica la instalación con la correspondiente educación a los usuarios, para que los usuarios tengan un dominio del uso de la herramienta.

Es importa este último punto, ya que es imprescindible que el usuario conozca la herramienta para realizar su trabajo con fluidez y los reportes salgan sin ningún inconveniente, y el desarrollo de la solución de inteligencia de negocio se obtenga el éxito esperado.

CAPÍTULO 4. Conclusiones, recomendaciones y limitaciones

Luego de terminar este TEG podemos decir que se lograron alcanzar los objetivos planteados, en los cuales llegamos a concluir que:

Se estudio el modelo de negocio del departamento de finanzas y se pudo comprender la forma en que trabajan y los datos que manejan, los cuales nos lleva a mejorar su proceso de extracción, transformación y carga, así como sus procesos de publicación de reportes para ser analizados en las distintas juntas que ellos tienen.

Se definieron correctamente las métricas que se son medidas que utiliza el departamento de finanzas, para estudiar el comportamiento de la empresa en cuanto a los gastos que estos manejan.

Se analizó la posibilidad de una solución de inteligencia de negocio y se pudo determinar que era una buena opción para este departamento y que si cumplía con sus expectativas, ya que disminuía el tiempo de la realización de la extracción de los datos en un 90% y la realización de los reportes.

Se diseño la Base de Datos Multidimensional la cual responde todas las preguntas que se generaron en el departamento de finanzas de manera satisfactoria.

Se diseño y se construyó los ETL que nos permitieron realizar operaciones requeridas en los datos y depurarla para alimentar al almacén de datos.

Se pudo desarrollar la solución de inteligencia de negocio para un departamento de finanzas en el sector petroquímico, utilizando la plataforma de inteligencia de

negocios Pentaho, permitiendo que se tomen decisiones de manera más rápida y oportunas.

Se realizó pruebas de la solución y las mismas fueron satisfactorias ya que cumplieron con los requerimientos del cliente y estas actualmente están en fase de prueba para su posterior puesta en producción.

La metodología que se uso para el desarrollo de la solución de inteligencia nos ayudó ir por tareas de alto nivel que se requiere para un efectivo diseño, desarrollo e implementación permitiendo cubrir los pasos necesarios para el desarrollo de la solución, por lo tanto esta metodología fue de gran ayuda ya que nos permitió recorrer cada uno de esos pasos necesarios para la elaboración de la solución de inteligencia de negocio de un departamento de finanzas.

La herramienta BI Pentaho, provee un completo espectro de funcionalidades de Inteligencia de Negocios, incluyendo reportes, análisis, dashboards, minería de datos, integración de datos. Al aprovechar la riqueza de las tecnologías de código abierto y las contribuciones de la comunidad de desarrollo de código abierto, Pentaho es capaz de innovar mucho más rápido que los proveedores comerciales. Como resultado, Pentaho ofrece una alternativa de código abierto que supera a las soluciones de inteligencia de negocio propietarias en muchas áreas como arquitectura, soporte de estándares, funcionalidad y simplicidad de implantación.

Recomendaciones

Como recomendación tenemos:

- Integrar la solución con el sistema Web de la organización
- Crear un sistema transaccional dado lo que se uso fueron archivos Excel que por lo general son elementos analíticos.

- Analizar la posibilidad de construir una mejor relación entre sus procesos para poder establecer una mejor relación entre ellos.

Limitaciones

Como principales limitaciones se tiene que para la realización de la solución de inteligencia de negocio es que no se cuenta con un sistema transaccional lo que se dificulta ver la relación entre las variables ya que desde los archivos fuentes está relación no se ve claramente.

Referencias

Ackoff, R. (1999). *Ackoff's Best His Classic Writings on Management*. John Wiley & Sons.

Arquitectura de una solución de Business Intelligence. (s.f.). Extraído el 30 de Octubre de 2008 desde

http://www.sinnexus.com/business_intelligence/arquitectura.aspx

Ballard, C., Abdel-Hamid, A., Frankus, R., Hasegaw, F., Larrechart, J., Leo, P. & Ramos, J. (2006). *Improving Business Performance Insight. With Business Intelligence and Business Process Management*. RedBooks.

Cano, J. L. (s.f.). *Business Intelligence: Competir con Información*. Extraído el 14 de Febrero de 2009 desde <http://www.dataprix.com/files/BI-Business-Intelligence-Competir-Con-Informacion.pdf>

Cortes E. (2007). Copyright SAP BI. Business Intelligence. Extraído el 14 de octubre de 2009 desde <http://www.sap.com/spain>

Hernández J. Almacenes de Datos (s.f). Extraído el 15 de septiembre desde <http://users.dsic.upv.es/~jorallo/cursoDWDM/dwdm-II.pdf>

Kimball M. y Ross R. (2002). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling-Second Edition*. United States of America: Wiley Computer Publishing.

Moss, L. T. & Atre, S. (2003). *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications*. Boston: Addison Wesley.

- Larson, B. (2006). *Delivering Business Intelligence with Microsoft SQL Server 2005*. California: McGraw-Hill/Osborne.
- Laudon, K. C. y Laudon J. P. (2004). *Sistemas de Información Gerencial*. Mexico: Pearson Prentice Hall.
- Saroka, R. (2002). *Sistemas de Información en la era digital*. Buenos Aires: Fundación OSDE.
- Langit, L. (2007). *Foundations of SQL Server 2005 Business Intelligence*. New York: Apress.
- López, J.C. (2005). *Open Source software e IBM*. Extraído el 6 de Junio de 2009 desde <http://www.mityc.es>
- Pentaho Open Source Business Intelligence. Extraído el 10 de Junio de 2009 desde <http://www.pentaho.com>
- Reinschmidt, J. & Françoise A. (2000). *Business Intelligence Certification Guide*. California: IBM International Technical Support Organization.
- Wolff, C. (2002). *Modelamiento Multidimensional*. Extraído el 29 de septiembre de 2009 desde <http://www.inf.udec.cl/revista/ediciones/edicion4/modmulti.PDF>.