

ESCUELA POLITÉCNICA DEL EJÉRCITO

DPTO. DE CIENCIAS DE LA COMPUTACIÓN

CARRERA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

**“ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UNA
SOLUCIÓN BUSINESS INTELLIGENCE PARA LA
GENERACIÓN DE INDICADORES Y CONTROL DE
DESEMPEÑO, EN LA EMPRESA OTECEL S.A, UTILIZANDO
LA METODOLOGÍA HEFESTO V2.0”**

**Previa a la obtención del Título de:
INGENIERO EN SISTEMAS E INFORMÁTICA**

POR:

**SOFÍA ANABEL BUSTOS BARRERA
VERÓNICA NATHALY MOSQUERA ARTIEDA**

SANGOLQUÍ, Marzo del 2013

CERTIFICACIÓN

Certifico que el presente trabajo fue realizado en su totalidad por las Srta(s). BUSTOS BARRERA SOFÍA ANABEL y MOSQUERA ARTIEDA VERÓNICA NATHALY como requerimiento parcial a la obtención del título de INGENIEROS EN SISTEMAS E INFORMÁTICA.

Marzo, 20 del 2013

ING. MAURICIO CAMPAÑA

AUTORIZACIÓN

Autorizamos a la ESCUELA POLITÉCNICA DEL EJERCITO la publicación en la Biblioteca Virtual de la Institución, del trabajo titulado **“ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UNA SOLUCIÓN BUSINESS INTELLIGENCE PARA LA GENERACIÓN DE INDICADORES Y CONTROL DE DESEMPEÑO, EN LA EMPRESA OTECEL S.A, UTILIZANDO LA METODOLOGÍA HEFESTO V2.0”**, que es de nuestra propia autoría y responsabilidad.

Marzo del 2013

Sofía Anabel Bistos B.

Verónica Nathaly Mosquera A.

DEDICATORIA

Dedicamos esta Tesis a nuestros padres quienes fueron nuestro apoyo incondicional a lo largo de nuestras vidas con sus consejos, sus valores, por la motivación constante que nos ha permitido ser personas de bien, pero más que nada, por su amor, a nuestros ingenieros que nos brindaron las bases de conocimiento necesarias para enfrentarnos a la vida profesional. A Dios que nos ha permitido llegar hasta este punto y habernos dado sabiduría para lograr nuestros objetivos, además de su infinita bondad y amor.

BUSTOS BARRERA SOFÍA ANABEL

MOSQUERA ARTIEDA VERÓNICA NATHALY

AGRADECIMIENTOS

Agradecemos a la Escuela Politécnica del Ejército, puesto que ha acogido a los mejores docentes, mismos que han compartido su conocimiento y han sembrado la semilla para la investigación. El agradecimiento especial al Ing. Mauricio Campaña, Director de Tesis, Ing. Lorena Duque Codirectora de Tesis, quienes han sido guía a lo largo de este proyecto desinteresadamente, hasta culminarlo. A nuestros padres por su apoyo, comprensión, quienes complementaron nuestra educación con valores, para ser profesionales con ética. A los integrantes del área de Pruebas de la Gerencia de Construcción de Otecel S.A., quienes colaboraron en todo el proceso para la culminación del proyecto y a todas aquellas personas que estuvieron pendientes de nuestra preparación a lo largo de la carrera.

BUSTOS BARRERA SOFÍA ANABEL

MOSQUERA ARTIEDA VERÓNICA NATHALY

ÍNDICE DE CONTENIDOS

CERTIFICACIÓN	ii
DEDICATORIA.....	iv
AGRADECIMIENTOS.....	v
GLOSARIO	xv
ACRÓNIMOS	xvii
RESUMEN.....	xviii
CAPÍTULO 1.....	1
1. TEMA.....	1
1.1. INTRODUCCIÓN	1
1.2. FORMULACIÓN DEL PROBLEMA.....	2
1.3. JUSTIFICACIÓN	3
1.4. OBJETIVOS	4
1.4.1. Objetivo General	4
1.4.2. Objetivos Específicos.....	5
1.5. ALCANCE.....	5
1.6. METODOLOGÍA.....	6
1.6.1. Características	7
1.7. FACTIBILIDAD.....	7
1.7.1. Factibilidad Técnica	8
1.7.1.1. Hardware.....	8
1.7.1.2. Software	9
1.7.2. Factibilidad Operativa	9
1.7.3. Factibilidad Económica.....	10
CAPÍTULO 2.....	12
MARCO TEÓRICO.....	12
2.1. INTRODUCCIÓN	12
2.2. SITUACIÓN ACTUAL.....	13
2.3. ORGANIGRAMA ESTRUCTURAL TELEFÓNICA	15
2.4. MAPA DE PROCESOS DE GESTIÓN DE TECNOLOGÍA DE INFORMACIÓN	17
2.5. GERENCIA DE TECNOLOGÍA.....	18

2.6.1. ORGANIGRAMA ESTRUCTURAL TECNOLOGÍA	18
2.6.2. MODELO DE PROCESOS GESTIÓN DE TECNOLOGÍA	19
2.7.1. Objetivo.....	19
2.7.2. Ámbito de Aplicación	20
2.7.3. Estructura del Área de pruebas.....	21
2.7.4. Descripción del Proceso de Área de Pruebas.....	22
2.8. BUSINESS INTELLIGENCE	26
2.8.1. DEFINICIONES	26
2.8.2. Características BI	28
2.8.3. Componentes de Business Intelligence.....	29
2.8.3.1. Datawarehouse	30
2.8.3.1.1. Características DataWarehouse	31
2.8.3.1.2. Estructura de DatawareHouse	32
2.8.3.2. Query & Reporting	33
2.8.3.2.1. Características de las Herramientas Reporting	34
2.8.3.2.2. Entornos de gestión de informes.....	34
2.8.3.3. PROCESO ANALÍTICO EN LÍNEA (OLAP).....	34
2.8.3.3.1. DEFINICIÓN.....	34
2.8.3.3.2. CARACTERÍSTICAS.....	35
2.8.3.3.3. ESQUEMAS	36
2.8.3.4. CUADROS DE MANDO	37
CAPÍTULO 3.....	38
METODOLOGÍA PARA EL DESARROLLO DE BUSINESS INTELLIGENCE ...	38
3.1. METODOLOGÍAS PARA EL DESARROLLO DE BI	38
3.1.1. Metodología Inmon	38
3.1.2. Metodología Kimball.....	39
3.1.3. Tipos de Arquitectura.....	41
3.1.4. Selección de Metodología.....	44
3.2. METODOLOGÍA HEFESTO	44
3.2.1. FASES DE LA METODOLOGÍA HEFESTO V2.0	46
3.2.1.1. ANÁLISIS DE REQUERIMIENTOS	46
3.2.1.3. MODELO LÓGICO DEL DW.....	51
3.2.1.4. INTEGRACIÓN DE DATOS	55
CAPÍTULO 4.....	57

HERRAMIENTAS	57
4.1. Motor de Base de Datos Oracle	57
4.1.1. Introducción.....	57
4.1.2. Características	58
4.2. Solución BI	61
4.2.1. Pentaho	62
4.2.1.1. Características de Pentaho.....	63
4.2.1.2. Arquitectura Pentaho	65
4.2.1.3. Pentaho Reporting	66
4.2.1.6. Pentaho Data Mining	71
4.2.1.7. KETTLE - Pentaho Data Integration.....	72
4.2.1.8. Pentaho BI Platform	75
4.2.1.9. Comparación entre Pentaho Community Edition y Enterprise Edition	76
4.2.2. Data Warehouse y Data Mart.....	80
4.2.2.1. Introducción.....	80
4.2.2.2. Arquitectura Data Warehouse.....	82
4.2.2.3. Componentes Básicos.....	83
4.2.3. Cubos OLAP	89
4.2.4. Estructura Cubo OLAP	89
4.3. Cuadros de Mando	90
4.3.1. Arquitectura de CDF	91
4.3.2. Online Analytical Processing - OLAP	92
4.3.2.1. Clasificaciones entre las implementaciones OLAP	93
4.3.3. Servidor OLAP Mondrian	94
4.4. Cubos de Información	95
4.5. Esquema Workbench de Mondrian.....	96
CAPÍTULO 5.....	97
DESARROLLO DE LA SOLUCIÓN	97
5.1. Definición del Proyecto.....	97
5.2. Justificación, Objetivos y Alcance	97
5.3. Identificación de Roles.....	98
5.4. Estrategia y método de Implementación.....	99

5.5. Situación Actual	100
5.6. Origen de la fuente de Datos	100
5.7. Análisis de Requerimientos	100
5.7.1. Entrevistas	101
5.7.2. Requerimientos	102
5.7.3. Identificar Indicadores Y Perspectivas De Análisis	105
5.7.3.1. Indicadores	105
5.7.3.2. Perspectivas	105
5.7.4. Diseño de la Arquitectura	107
5.7.5. Ambiente Estación de Trabajo	108
5.8. Análisis de los OLTP	109
5.8.1. Conformación de Indicadores	109
5.8.2. Establecer Correspondencias con los requerimientos.....	111
5.8.3. Modelo Dimensional	111
Estándar Modelo Dimensional	111
Lista de Dimensiones.....	112
5.8.4. Nivel de granularidad	120
5.8.5. Modelo Conceptual Ampliado	122
5.9. Selección e Instalación de Herramientas	124
5.9.1. Selección de Herramientas.....	124
5.9.2. Instalación de las Herramientas.....	125
5.10. INTEGRACIÓN DE DATOS	129
Estándar de ETL	129
5.10.1. Carga Inicial	131
5.10.2. Data Integration	131
5.10.3. Creación del Cubo	141
5.10.4. Creación de Reportes	145
CAPÍTULO 6.....	153
CONCLUSIONES Y RECOMENDACIONES	153
6.1. Conclusiones	153
6.2. Recomendaciones	155
BIBLIOGRAFÍA	157

ÍNDICE DE TABLAS

Tabla 1.1. Costos de Material.....	10
Tabla 1.2. Costos de Operación.....	11
Tabla 1.3. Costo Total del Proyecto	11
Tabla 3.1 Comparación de Metodologías – Tesistas	43
Tabla 4.1. Productos Pentaho. Tesistas.....	64
Tabla 4.2: Feature Comparison: Community & Enterprise Editions	76
Tabla 4.3: Pentaho Community & Commercial Product Comparison	77
Tabla 5.1. Ingenieros - Roles	98
Tabla 5.2. Análisis de Requerimientos	102
Tabla 5.2. Descripción de Dimensión Tiempo ejecuta	113
Tabla 5.3. Descripción de Dimensión Requerimientos.....	114
Tabla 5.4. Descripción de Dimensión Directivos	115
Tabla 5.5. Descripción de Dimensión Tiempo Transacción	116
Tabla 5.6. Descripción de Dimensión Ambiente.....	117
Tabla 5.7. Descripción Métrica	118
Tabla 5.8. Descripción de Tabla Requerimiento - BD Origen.....	118
Tabla 5.9. Descripción de Tabla Ambiente - BD Origen.....	119
Tabla 5.10. Descripción de Tabla Directivos - BD Origen	119
Tabla 5.11. Descripción de Tabla Tareas - BD Origen.....	120

ÍNDICE DE FIGURAS

Figura 2.1: Organigrama Estructural Telefónica – OTECEL S.A.....	15
Figura 2.2: Mapa de Procesos Telefónica Ecuador – Otecel S.A.....	16
Figura 2.3: Gestión de Tecnología de Información.....	17
Figura 2.4: Organigrama Estructural Tecnología – OTECEL S.A.....	18
Figura 2.5: Gestión de Tecnología – OTECEL S.A.	19
Figura 2.6: Estructura Área de Pruebas – OTECEL S.A.	21
Figura 2.7: Proceso BI.....	28
Figura 2.8: Datawarehouse	30
Figura 2.9: Formación de Cubos OLAP: Tesistas	35
Figura 3.1: Arquitectura Inmon	39
Figura 3.2: Kimball	41
Figura 3.3: Top Down.....	41
Figura 3.4: Bottom Up	42
Figura 3.5: Híbrida.....	42
Figura 3.6: Fases de la metodología Hefesto v2.0: Tesistas.....	45
Figura 3.7: Modelo Conceptual: Tesistas	48
Figura 3.8: Modelo Conceptual con atributos: Tesistas.....	50
Figura 3.9: Modelo Conceptual ampliado: Tesistas.....	51
Figura 3.10: Diseño de tablas de dimensiones: Tesistas	52
Figura 3.11: Tabla de hechos (Esquema estrella y copo de nieve): Tesistas	53
Figura 3.12: Caso 1, preguntas.	54
Figura 3.13: Caso 1, preguntas.	54
Figura 3.14: Caso 1, preguntas.	54
Figura 4.1. Arquitectura Oracle	58
Figura 2.2: Estructuras de memoria y procesos de Oracle	61
Figura 4.3: Componentes Pentaho.....	64
Figura 4.4: Arquitectura Pentaho.....	66
Figura 4.5: Arquitectura de Pentaho Data Integration	73
Figura 4.7: Esquema en Estrella	85
Figura 4.8: Esquema copo de Nieve	87
Figura 4.9: Esquema Constelación	88

Figura 4.10: Arquitectura CDF.....	92
Figura 4.11: Funcionamiento de Mondrian.....	94
Figura 5.1: Estructura Actual de Proceso de Información en el Área de Pruebas 100	
Figura 5.2: Diagrama de Perspectivas - Tesistas.....	107
Figura 5.3: Proceso Solución: Tesistas	107
Figura 5.4: Ambiente Estación de Trabajo - Tesistas.....	108
Figura 5.5: Indicadores Generales: Tesistas	109
Figura 5.6: Indicador Ambiente: Tesistas	109
Figura 5.7: Indicador Tareas: Tesistas	110
Figura 5.8: Indicador Responsables: Tesistas.....	110
Figura 5.9: Proceso Análisis Solución: Tesistas.....	111
Figura 5.10: Modelo Dimensional: Tesistas	112
Figura 5.11: Análisis Indicador Número de Horas Trabajadas: Tesistas.....	121
Figura 5.12: Análisis Número de Requerimientos: Tesistas	121
Figura 5.13: Análisis Porcentaje de Requerimientos: Tesistas	122
Figura 5.14: Análisis Número Horas Trabajadas: Tesistas	122
Figura 5.15: Análisis Número Requerimientos: Tesistas.....	123
Figura 5.16: Análisis Tareas Asignadas: Tesistas.....	123
Figura 5.17: Análisis Porcentaje de Requerimientos: Tesistas	124
Figura 5.18: Carpeta Utilitarios Pentaho	126
Figura 5.19: Pentaho Data Integration	127
Figura 5.20: Configuración conexión Base de Datos	127
Figura 5.21: Configuración Base de Datos.....	128
Figura 5.22: Interfaz de Usuario Pentaho.....	128
Figura 5.23: Archivo CSV	131
Figura 5.24: Selección Archivo.....	132
Figura 5.25: Carpeta Utilitarios Pentaho	132
Figura 5.26: Cabeceras Archivo	133
Figura 5.27: Java Script	133
Figura 5.28: Proceso de Transformación	134
Figura 5.29: Código JavaScript	134
Figura 5.30: Configuración Dimensión	135
Figura 5.31: Configuración Dimensión	135

Figura 5.32: Configuración Dimensión	136
Figura 5.33: Configuración Dimensión	136
Figura 5.34: Configuración Dimensión	137
Figura 5.35: Configuración Dimensión	137
Figura 5.36: Configuración Dimensión	138
Figura 5.37: Configuración Dimensión	138
Figura 5.38: Configuración Dimensión	139
Figura 5.39: Configuración Dimensión	139
Figura 5.40: Configuración Dimensión	140
Figura 5.41: Configuración Dimensión	140
Figura 5.42: Creación DataSource	141
Figura 5.43: Creación DataSource	141
Figura 5.43: Configuración DataSource	142
Figura 5.44: Configuración Vista de Análisis	142
Figura 5.45: Menú Vista de Análisis	143
Figura 5.46: Selección Esquema.....	143
Figura 5.47: Número de Requerimientos por Ambiente y Tipo de Horas	144
Figura 5.48: Vista de Análisis	144
Figura 5.49: Requerimientos por Ambiente	145
Figura 5.50: Requerimientos por Ambiente	146
Figura 5.51: Total Horas Trabajo – Ambiente.....	147
Figura 5.52: Número Requerimiento – Ambiente	147
Figura 5.53: Número Requerimiento – Ambiente	148
Figura 5.54: Tipo de Pruebas por Requerimiento.....	149
Figura 5.55: Número Requerimiento – Ambiente	149
Figura 5.56: Tiempo Total Horas Ambiente.....	150
Figura 5.57: Número de Requerimientos	151
Figura 5.58: Control Tipo Prueba – Ambiente	151
Figura 5.59: Control Tipo Prueba – Ambiente	152

ÍNDICE DE ANEXOS

Anexo A – Script de Base de Datos de Dimensiones..... **Error! Bookmark not defined.**

Anexo B - Script Base de Datos Temporal..... **Error! Bookmark not defined.**

Anexo C - Manual de Instalación de Herramientas que contiene Instalación de Producto Pentaho, Configuración y Ejecución **Error! Bookmark not defined.**

Anexo D - Manual Creación de Dimensiones..... **Error! Bookmark not defined.**

Anexo E – Creación de Dashboard **Error! Bookmark not defined.**

Anexo F – Creación de Vista de Análisis **Error! Bookmark not defined.**

GLOSARIO

Dato	Representación simbólica de un atributo o característica de una entidad.
Información	Conjunto de datos procesados que transmite conocimiento.
Procesamiento	Técnicas usadas para manipular datos, para lograr en ellos una transformación.
Transformación	Acción o procedimiento mediante el cual algo se modifica, altera o cambia de forma manteniendo su identidad.
Conocimiento	Conjunto de información almacenada mediante la experiencia o el aprendizaje.
Gestión de	Herramienta que se enmarca dentro de los
Tecnología	procesos generales de innovación en las empresas.
Query	cadena de consulta, que se ejecuta en una sentencia SQL con la interacción de una base de datos.
Arquitectura	Diseño de una solución de Bussines Intelligence
Esquema	Modelo Lógico que se utiliza para el diseño de dimensiones.

Fichero Plano	Un archivo de texto llano, texto simple, texto plano, texto sencillo o texto pelado.
Usuario	Perona(s) que interactúan con la implementación.
Carga	Proceso por el cual se realiza el Almacenamiento de Datos.

ACRÓNIMOS

DM:	Data Mart
DW:	Data Warehouse
OLTP:	On Line Transaction Processing
PRD:	Pentaho Report Designer
RDW:	Pentaho Report Design Wizard
MDX:	Expresiones multidimensionales
RDBMS:	Base De Datos Relacional
CDF:	Community Dashboard Framework
BDD:	Base de Datos
CE:	Community Edition
SVC:	Servicio al Cliente
ETL:	Extraer, transformar, cargar
OLAP:	Procesamiento analítico en línea
KPI:	Indicadores Clave de Desempeño
SGA:	Área Global del Sistema
PGA:	Área Global del Programa
SOA:	Arquitecturas Orientadas a Servicios
PDI:	Pentaho Data Integration
CDF:	Community Dashboard Framework
CDE :	Community Dashboard Editor

RESUMEN

El uso de una herramienta de Business Intelligence (BI) permite organizar y analizar los datos alojados en bases de datos de distintas fuentes, para obtener así el conocimiento, que facilite la interpretación y correcta comprensión que ayude a la toma de decisiones para el negocio, dando una ventaja competitiva. La presente tesis de grado propone la implementación de una solución BI para el manejo de Datos de la empresa de telefonía celular OTECEL. S.A., basado en la metodología Hefesto que es flexible y permite que la solución sea escalable de acuerdo a los nuevos cambios requeridos, haciendo uso de la herramienta Open Source denominada Pentaho, que permite realizar el tratamiento de los datos para el análisis. Los resultados de la implementación de la solución BI, permitieron obtener en menor tiempo información que podrá ser analizada por Jefes e Ingenieros del Área de Pruebas, que mediante reportes dinámicos prediseñados y la vista de análisis obtienen acceso a los indicadores y control de las tareas.

Business Intelligence permite reunir, transformar y depurar los datos que se encuentren en forma desestructurada, proveniente de bases de datos operacionales, evitando la existencia de datos irrelevantes. Los datos constituyen un elemento primordial para la empresa y sobre todo para el desarrollo de una solución Business Intelligence (BI). Estas necesidades hacen que se requiera implementar una tecnología de software, basada en las nuevas tendencias y herramientas de análisis de información. Vivimos en una época en la que la información es la clave para mantenerse competitivo, los gerentes y analistas requieren tener un acceso rápido y fácil a información útil y valiosa, una forma de solucionar este problema es por medio del uso de Business Intelligence (BI).

CAPÍTULO 1

1. TEMA

“ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UNA SOLUCIÓN BUSINESS INTELLIGENCE PARA LA GENERACIÓN DE INDICADORES Y CONTROL DE DESEMPEÑO, EN LA EMPRESA OTECEL S.A, UTILIZANDO LA METODOLOGÍA HEFESTO V2.0”

1.1. INTRODUCCIÓN

Business Intelligence hace referencia a la capacidad de organizar y explotar la información que se encuentra alojada en bases de datos de distintas fuentes, las cuales mediante procesos iterativos, permiten el análisis de información para obtener así el conocimiento, que facilite la interpretación y correcta comprensión por parte de los usuarios que se encargan de la toma de decisiones de negocio, así como la transformación de los datos de la compañía en conocimiento para obtener una ventaja competitiva.

Business Intelligence permite reunir, transformar y depurar los datos que se encuentren en forma desestructurada, proveniente de bases de datos operacionales, evitando la existencia de datos irrelevantes.

Los datos constituyen un elemento primordial para la empresa y sobre todo para el desarrollo de una solución Business Intelligence (BI) y cuyo objetivo central es la información relevante sobre la situación de la empresa, la misma que será de gran apoyo a las personas encargadas de la toma de decisiones.

OTECEL S.A es una empresa privada de telefonía celular, actualmente el enfoque de esta empresa es prestar servicios en el ámbito de telecomunicaciones. Dentro de las actividades realizadas por el personal técnico y administrativo que trabaja en la empresa existe una serie de necesidades, sin embargo nos enfocaremos en el área de pruebas de la Gerencia de Construcción.

Estas necesidades hacen que se requiera implementar una tecnología de software, basada en las nuevas tendencias y herramientas de análisis de información. Vivimos en una época en la que la información es la clave para mantenerse competitivo, los gerentes y analistas requieren tener un acceso rápido y fácil a información útil y valiosa, una forma de solucionar este problema es por medio del uso de Business Intelligence (BI).

1.2. FORMULACIÓN DEL PROBLEMA

En la actualidad, el área de pruebas de la Gerencia de Construcción no cuenta con una herramienta que permita generar ágilmente reportes a los jefes y directivos encargados de la toma de decisiones, este proceso se lo maneja de manera manual, lo cual genera retrasos en tiempos de respuesta y cargas operacionales innecesarias.

La toma de decisiones dentro del área de pruebas de la Gerencia de Construcción debe manejarse de forma estratégica e inmediata, pero al realizarse estos procesos de manera manual, existe un retardo en tiempos de entrega de resultados, lo que repercute en demoras en tomas de decisiones que ocasionan pérdida competitiva.

Mediante los indicadores y resultados obtenidos se puede determinar medidas de cambio y mejora basados en el análisis de información y datos procesados, que generan una visión más clara del estado actual del negocio.

La implantación de una solución BI representa un apoyo de gran utilidad al momento de realizar informes, puesto que las personas encargadas de la obtención de reportes, dejan de lado sus tareas asignadas para entregar los resultados a tiempo, al personal encargado de tomar decisiones las decisiones, quienes requieren contar con información relevante para definir estrategias claves.

En reuniones gerenciales o interdepartamentales imprevistas que se llevan a cabo, existe información vital, que se necesita conocer, la cual sin una herramienta de apoyo apropiada podría llegar a demorarse incluso días debido al gran volumen de datos que se manejen. Gerentes, directivos y jefes departamentales deben tener acceso rápido a la información para realizar un análisis personalizado de los diferentes escenarios existentes que se manejan en el área de pruebas de la Gerencia de Construcción.

La información obtenida de la herramienta debe ser oportuna, veraz, explícita y confiable, para que de esta manera las resoluciones tomadas, sean las más acertadas en beneficio del crecimiento de la empresa.

1.3. JUSTIFICACIÓN

En la actualidad el manejo de la información es vital para obtener una ventaja competitiva, y para ello, los gerentes y analistas, encargados de la toma de decisiones, requieren un acceso fácil y rápido a información útil y valiosa de la empresa para dar soluciones operativas, tácticas y estratégicas de mejora, basados en el análisis de esta información.

La implementación de una solución de Business Intelligence (BI) dentro de la empresa OTECEL S.A., para el área de pruebas de la Gerencia de Construcción, representa una estrategia corporativa de crecimiento, optimización, productividad y competitividad, que no sólo les permitirá tener tiempos de respuesta menores en la obtención de datos procesados, sino el mejor manejo de los mismos, la toma de decisiones basados en datos precisos y de calidad para su análisis.

Uno de las principales problemáticas se halla en la generación de reportes por parte del área de pruebas, razón por la cual, los gerentes y analistas requieren de herramientas que les permitan de manera rápida realizar un análisis de información y generación de reportes, ya que los mismos en la actualidad tardan alrededor de cuatro días en ser elaborados.

El área de pruebas actualmente, no cuenta con una solución BI, para manejar el procesamiento de datos, este proceso se lo hace con la ayuda del recurso humano, que se encarga de realizar de manera manual, el manejo de los datos, utilizando tablas dinámicas en Excel para la generación de los reportes, haciendo que este proceso no sea óptimo, y que los tiempos de obtención de resultados finales, se demoren más, en un aproximado de cuatro días, mientras que con la utilización de una solución BI, se puede obtener estos resultados en tiempo de horas o minutos, este tiempo sin embargo variará de acuerdo a la cantidad de datos que se vayan a procesar, pero serán mucho menores a los que actualmente se posee, motivo por el cual la implementación de una solución BI representa una necesidad de gran relevancia dentro del negocio.

El uso de la Herramienta de Pentaho Community (CE), al ser open source permitirá que la implementación de la solución BI, sea de baja costo y además cuenta con otras herramientas que a su vez, ayudarán en la generación de reportes, cubos. Las herramientas Pentaho ofrecen muchas características de gran alcance, utilizadas en BI como son las vistas de análisis, presentación de informes que tienen la capacidad de personalización de los mismos, acorde a las necesidades de negocio.

1.4. OBJETIVOS

1.4.1. Objetivo General

Implementar una solución BI para el área de pruebas de la Gerencia de Construcción de la empresa Otecel S.A, mediante la investigación, aplicación de tecnologías actuales, utilización de herramientas de software libre, empleando la metodología Hefesto.

1.4.2. Objetivos Específicos

- Analizar la información de activaciones de planes comerciales y registro de tareas por parte de proveedores , entregada por área de pruebas de la Gerencia de Construcción
- Diseñar la estructura de que tendrán las dimensiones y cubos pertinentes para crear el diseño de vista de análisis y la navegabilidad que tendrá la información pertinente a cada Datamart.
- Implementación de una solución BI, utilizando las herramientas Pentaho Community Edition (CE).
- Utilizar la herramienta Report Designer para crear y diseñar los reportes correspondientes a cada Datamart.
- Determinar las fuentes de datos ,con las que se trabajará en la solución BI

1.5. ALCANCE

Se realizará la implementación de la Solución de Business Intelligence, para la generación de indicadores y el control de desempeño, en el área de pruebas de la Gerencia de Construcción de la empresa OTECEL S.A.

La implementación de una Solución BI con la herramienta Pentaho Community permitirá el desarrollo de dos reportes prediseñados en el Pentaho Report Designer, para el área de pruebas de OTECEL S.A., que tiene la capacidad de personalización de informes de acuerdo a las necesidades de negocio utilizando Oracle, como motor de base de datos, del departamento de Gerencia de Construcción del área de pruebas.

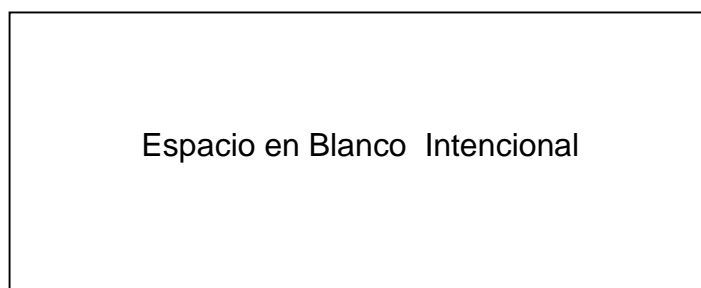
Se trabajará con el Datamart del área de pruebas del departamento de Gerencia de Construcción, el cual será alimentado por las bases de datos de la empresa.

Se implementará un cubo, para poder realizar las distintas combinaciones, y otorgar la información requerida por el usuario. El proyecto no incluye capacitaciones al usuario final.

El proyecto concluirá con la elaboración de reportes, generación de los indicadores y control de desempeño para el área de pruebas de OTECEL S.A. Las vistas de análisis diseñadas y los reportes prediseñados, elaborados desde la consola de Pentaho, serán utilizados por los ingenieros de pruebas, para facilitar la toma de decisiones.

1.6. METODOLOGÍA

Se utilizará la metodología Hefesto V2.0 , durante la implementación de la solución BI, que es una metodología, propia para el trabajo con soluciones bajo la plataforma de Pentaho, la idea de utilizar Hefesto es que no se requieran fases extensas de reunión de requerimientos y análisis, fases de desarrollo monolítico que conlleve demasiado tiempo y fases de despliegue muy largas. Lo que se busca, es entregar una primera implementación que satisfaga una parte de las necesidades, para demostrar las ventajas y motivar a los usuarios.



1.6.1. Características

Esta metodología cuenta con las siguientes características:

- Los objetivos y resultados esperados en cada fase se distinguen fácilmente y son sencillos de comprender.
- Se basa en los requerimientos de los usuarios, por lo cual su estructura es capaz de adaptarse con facilidad y rapidez ante los cambios del negocio.
- Reduce la resistencia al cambio, ya que involucra a los usuarios finales en cada etapa para que tomen decisiones respecto al comportamiento y funciones.
- Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.
- Es independiente del tipo de ciclo de vida que se emplee para contener la metodología.
- Es independiente de las herramientas que se utilicen para su implementación.
- Cuando se culmina con una fase, los resultados obtenidos se convierten en el punto de partida para llevar a cabo el paso siguiente.

1.7. FACTIBILIDAD

Después de definir la problemática presente y establecer las causas que ameritan el implementar un nuevo sistema, es pertinente realizar un estudio de factibilidad para determinar la infraestructura tecnológica y la capacidad técnica que implica la implantación de una solución BI, así como los costos, beneficios y el grado de aceptación que la propuesta genera en la Empresa. Este análisis permitirá determinar las posibilidades de implantar la solución y su puesta en marcha, los aspectos tomados en cuenta para este estudio fueron clasificados en tres áreas, las cuales se describen a continuación:

1.7.1. Factibilidad Técnica

La Factibilidad Técnica consistió en realizar una evaluación de la tecnología existente en la organización, este estudio estuvo destinado a recolectar información sobre los componentes técnicos que posee la organización y la posibilidad de hacer uso de los mismos en el desarrollo e implementación del sistema propuesto y de ser necesario, los requerimientos tecnológicos que deben ser adquiridos para el desarrollo y puesta en marcha del sistema en cuestión.

De acuerdo a la tecnología necesaria para la implantación de la Solución BI para el departamento de pruebas de OTECEL S.A, se evaluó bajo dos enfoques: Hardware y Software.

1.7.1.1. Hardware

En cuanto a Hardware, específicamente el servidor donde debe estar instalado el sistema propuesto, este debe cubrir con los siguientes requerimientos mínimos:

- Dos computadores de escritorio:
- Intel Core 2 Duo
- 3 Gb memoria RAM
- 320 Gb de Disco Duro
- Conexión a Internet
- Impresora, Lexmark 1100
- Unidad de Protección UPS.

Por características físicas de la Red de OTECEL S.A, la plataforma a implementar está disponible solamente en la intranet, la misma que permita la conexión de todos los usuarios que se encuentren en la misma red.

1.7.1.2. Software

En cuanto al software, la empresa cuenta con todas las aplicaciones y licencias que se emplearan para el equipo donde se implementara la solución BI.

- Sistema Operativo: Windows XP
- Internet Explorer (Navegador Web) para consultas y documentación.
- Pentaho BI suite.
- Oracle Developer.
- SybasePower Designer 12.
- Ofimática: Microsoft Office 2007.

1.7.2. Factibilidad Operativa

La Factibilidad Operativa permite predecir, si se pondrá en marcha la Solución propuesta, aprovechando los beneficios que ofrece, a todos los usuarios involucrados con el mismo, ya sean los que interactúan en forma directa con este, así también aquellos que reciben información producida por la solución BI. Por otra parte, el correcto funcionamiento del sistema en cuestión, siempre estará supeditado a la capacidad de los empleados encargados de dicha tarea.

La necesidad y deseo de un cambio en la forma actual de elaborar los reportes para el posterior análisis, expresada por los usuarios y el personal involucrado con el mismo, llevó a la aceptación de una nueva solución, que de una manera más sencilla y amigable, que proporcione la información en forma oportuna y confiable. Basándose en las entrevistas y conversaciones sostenidas con el personal involucrado se demostró que estos no representan ninguna oposición al cambio, por lo que el sistema es factible operacionalmente.

En el proceso de adiestramiento se detallan los aspectos de actualización de conocimientos y nuevas formas en el manejo del nuevo sistema.

Con la finalidad de garantizar el buen funcionamiento del sistema y que este impactará en forma positiva a los usuarios, la plataforma presenta interfaz amigable al usuario, lo que se traduce en una herramienta de fácil manejo y comprensión.

El seguimiento de la realización del proyecto estará a cargo del jefe de la Gerencia de Construcción del área de pruebas de OTECEL S.A., que será quien garantice la entrega de los equipos en los cuales se realizará las instalaciones.

1.7.3. Factibilidad Económica

COSTOS DE MATERIAL

Tabla 1.1. Costos de Material

Descripción	Cantidad	Costo Unitario	Costo Total
Computador	2	700	1400
Utilitarios de Microsoft Office 2007	1		300
Pentaho Suite Open Source	-	-	0
Suministros de oficina		360	360
Servicios básicos		200	200
Egresados en Sistemas	2	800	1600
		Total	\$ 3860

COSTOS DE OPERACIÓN

Tabla 1.2. Costos de Operación

Descripción	Tiempo/Meses	Costo Mensual	Costo Total
Alimentación	6	50	300
Movilización	6	105	630
		Total	\$ 930

COSTO TOTAL DEL PROYECTO

Tabla 1.3. Costo Total del Proyecto

TIPO DE COSTO	VALOR
Materiales	\$3860
Operación	\$930
COSTO TOTAL	\$4790

Los valores mencionados serán cubiertos por la empresa OTECEL S.A., quien auspicia el proyecto hasta su respectiva implementación

CAPÍTULO 2

MARCO TEÓRICO

2.1. INTRODUCCIÓN

Telefónica es uno de los operadores integrados de telecomunicaciones líder a nivel mundial en la provisión de soluciones de comunicación, información y entretenimiento, con presencia en Europa y Latinoamérica. Está presente en 25 países y cuenta con una base de clientes que supera los 311,8 millones a junio de 2012.

La compañía dispone de uno de los perfiles más internacionales del sector al generar más de un 75% de su negocio fuera de su mercado doméstico, y se constituye como el operador de referencia en el mercado de habla hispano-portuguesa.

El Grupo ocupa la octava posición en el sector de telecomunicaciones a nivel mundial por capitalización bursátil, la primera como operador europeo integrado, y la décimo cuarta en el ranking Eurostoxx 50, que agrupa las mayores compañías de la zona Euro (30 de junio de 2012).

Telefónica es una empresa totalmente privada, con más de 1,5 millones de accionistas directos y cotiza en el mercado continuo en las bolsas españolas (Madrid, Barcelona, Bilbao y Valencia) y en las de Londres, Nueva York, Lima y Buenos Aires.

En Latinoamérica, la compañía presta servicios a más de 207,9 millones de clientes a 30 de junio de 2012, posicionándose como operador líder en Brasil, Argentina, Chile y Perú y contando con operaciones relevantes en Colombia,

Costa Rica, Ecuador, El Salvador, Guatemala, México, Nicaragua, Panamá, Puerto Rico, Uruguay y Venezuela. En Europa, la compañía tiene presencia, además de en España, en el Reino Unido, Irlanda, Alemania, República Checa y Eslovaquia, dando servicio a más de 103,1 millones de clientes al cierre de junio de 2012.¹

La visión de Telefónica es transformar en realidad las posibilidades que ofrece el nuevo mundo digital y ser uno de sus principales protagonistas. Uno de los pasos que ha dado la compañía ha sido la creación de diferentes unidades de desarrollo de productos y servicios globales, que abarcan, entre otros, Cloud Computing (servicios en la nube), eHealth, Servicios Financieros, M2M (conexiones máquina a máquina), Vídeo y Hogar Digital, Aplicaciones y Seguridad. Su objetivo es aprovechar economías de escala y adaptar desarrollos globales a necesidades regionales.

Estas unidades son los ejes de un modelo de innovación globalizado ideados para capturar el máximo de este crecimiento del sector de las telecomunicaciones.

2.2. SITUACIÓN ACTUAL

Está en las 24 provincias del país y tiene 4.6 millones de usuarios hasta febrero de 2012. Su plantilla cuenta con más de 1.160 colaboradores, de los cuales un 4% son personas con discapacidad. Sus ingresos representan casi el 1% del PIB nacional. Mantiene un Índice de Satisfacción del Cliente situado en 9

¹Telefónica, Telefónica, [En línea] 2012
http://www.telefonica.com/es/about_telefonica/html/quienessomos/quienessomos.shtml

puntos sobre 10. El 98.8% de los requerimientos de usuarios se pueden realizar vía telefónica o web y el 60.5% de sus proveedores son ecuatorianos.²



Espacio en Blanco Intencional

²Telefónica, Telefónica, [En línea] 2012 http://www.telefonica.com.ec/at_2.php

2.3. ORGANIGRAMA ESTRUCTURAL TELEFÓNICA

Figura 2.1: Organigrama Estructural Telefónica – OTECEL S.A.

2.4. MAPA DE PROCESOS DE GESTIÓN DE TECNOLOGÍA DE INFORMACIÓN

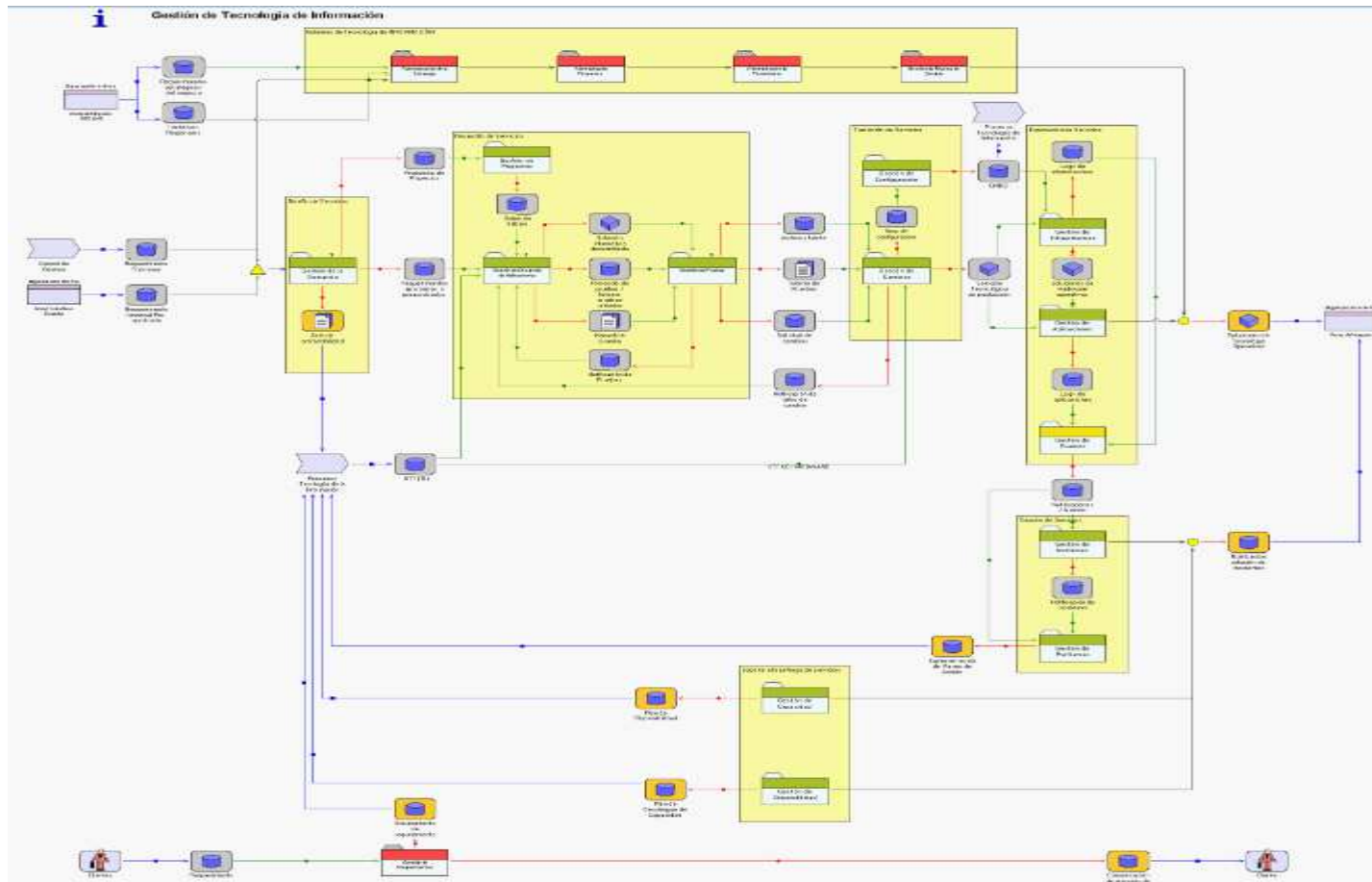


Figura 2.3: Gestión de Tecnología de Información

2.5. GERENCIA DE TECNOLOGÍA

2.6.1. ORGANIGRAMA ESTRUCTURAL TECNOLOGÍA

Figura 2.4: Organigrama Estructural Tecnología – OTECEL S.A.

2.6.2. MODELO DE PROCESOS GESTIÓN DE TECNOLOGÍA

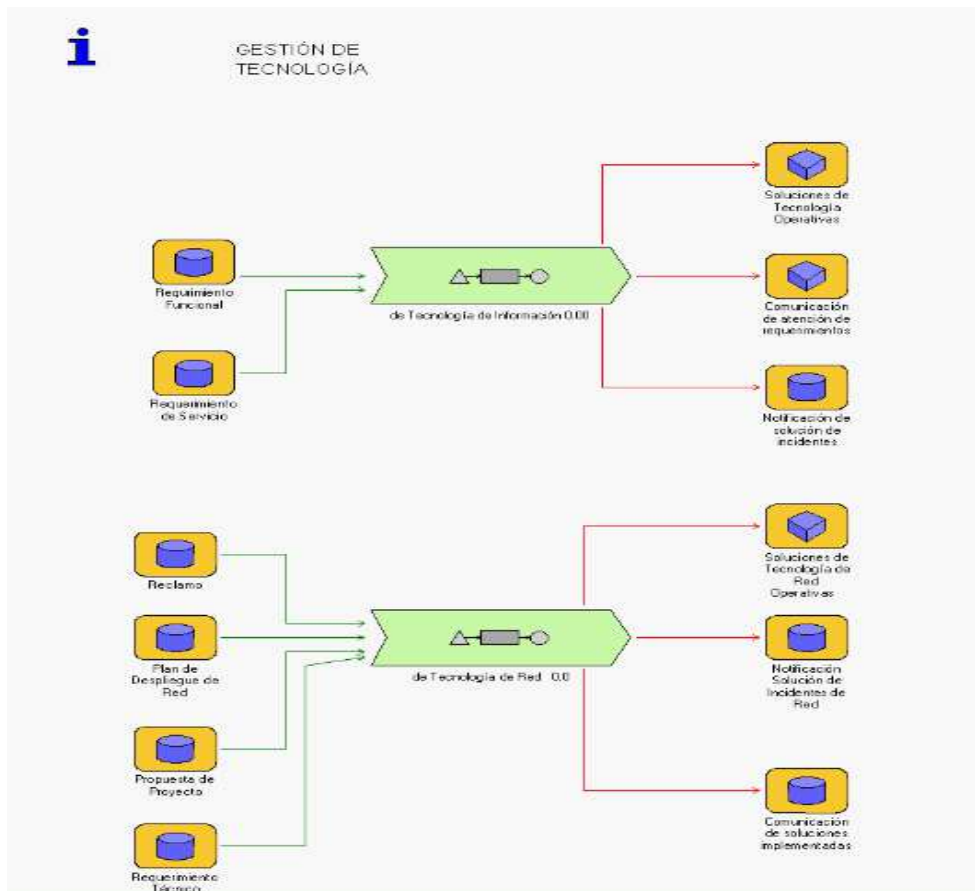


Figura 2.5: Gestión de Tecnología – OTECEL S.A.

2.7. ÁREA DE PRUEBAS

2.7.1. Objetivo

Asegurar la eficacia y eficiencia de los requerimientos de OTECEL S.A., verificando la calidad del funcionamiento al momento de la entrega de las soluciones informáticas.

2.7.2. Ámbito de Aplicación

Aplica a los requerimientos de desarrollo y configuración de aplicaciones emitidos por las diferentes dependencias de Telefónica Ecuador que necesitan ser liberados para su puesta en producción.

Espacio en Blanco Intencional

2.7.3. Estructura del Área de pruebas

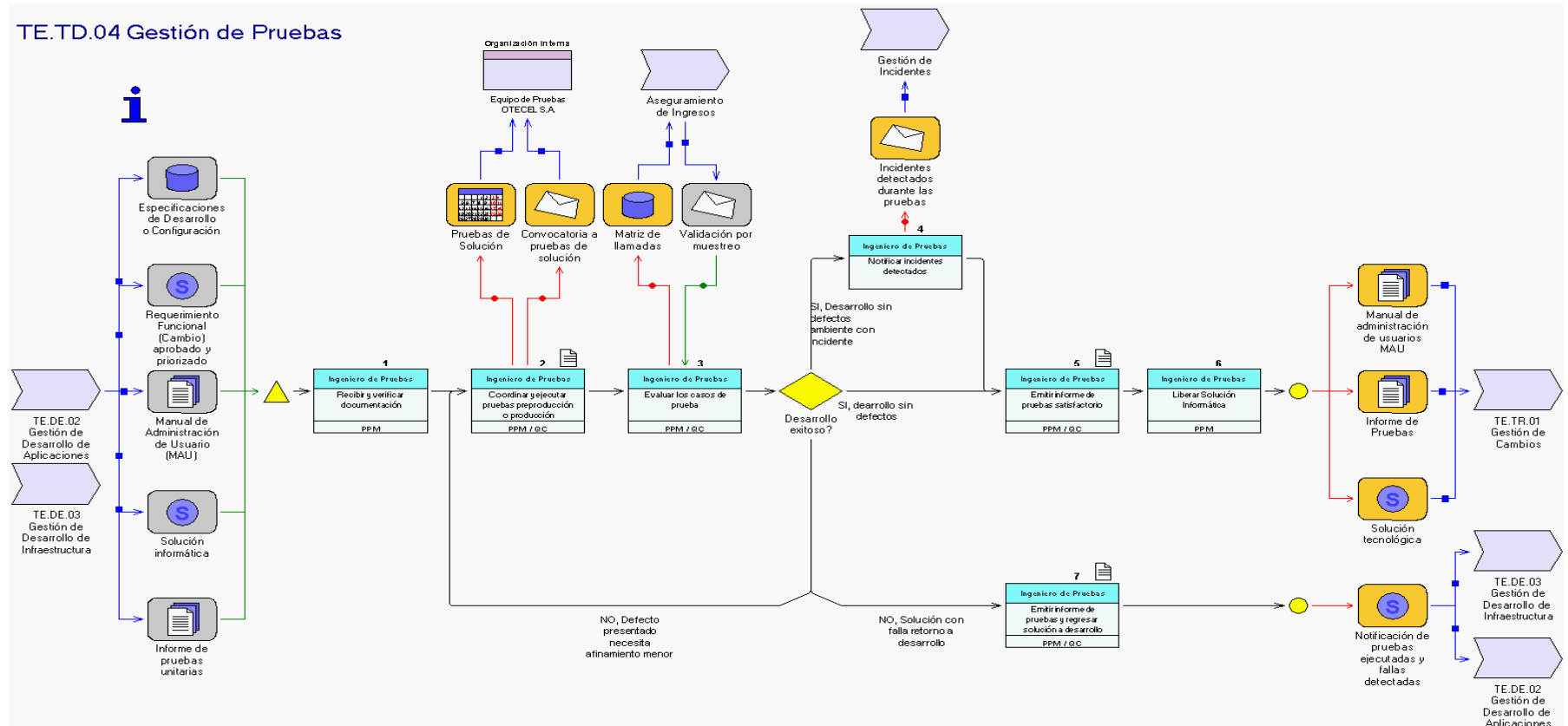


Figura 2.6: Estructura Área de Pruebas – OTECEL S.A.

2.7.4. Descripción del Proceso de Área de Pruebas

Recibir y Verificar documentación

- Recibir los documentos del requerimiento y sus especificaciones técnicas.
- Verificar que los documentos se encuentren debidamente completos y aprobados.
- Verificar que la solución tecnológica se encuentre depositada y disponible en el ambiente de pruebas correspondiente. En el caso de no disponer del ambiente confirmar la ubicación en donde se desarrollarán las pruebas.
- Registrar en la herramienta de gestión la aceptación de los documentos y el inicio de la etapa pruebas.

Coordinar y Ejecutar Pruebas Preproducción o Producción

- Convocar a las sesiones para la ejecución de las pruebas a todos los colaboradores definidos como equipo de pruebas en el protocolo a aplicar.
- Gestionar la disponibilidad de inventario de pruebas asegurando su existencia y funcionalidad conforme con lo descrito con el procedimiento "Gestión y Control del Inventario de Pruebas".
- Ejecutar las pruebas, conjuntamente con el equipo de pruebas del Negocio, conforme con lo planificado y definido en el protocolo.
- Documentar las pruebas realizadas identificando si el resultado es exitoso o presenta defectos en la herramienta de control de calidad.

- En el caso de que no se disponga de ambiente de preproducción para una o varias plataformas, las pruebas correspondientes a estas plataformas se deberán ejecutar en el ambiente de producción. Las pruebas en producción básicas de servicio, se realizarán siempre que pueda existir afectación al servicio al cliente final, y se deben ejecutar en la ventana de mantenimiento abierta para este fin, las pruebas producción con el equipo de pruebas y usuario del requerimiento se realizan una vez concluidas el paso a producción y las pruebas internas producción.

Evaluar los casos de prueba

- Evaluar cada caso de prueba identificando cuales han resultado casos exitosos o casos con defectos. Durante la evaluación verificar el registro existente en la herramienta de control de calidad y la descripción de los impactos acorde con el método de evaluación establecido para el protocolo de pruebas. Determinar la calidad de la solución tecnológica identificando la existencia de defectos y estableciendo su liberación conforme los siguientes criterios de aceptación:
- Funcionalidad, cuando cumple con los objetivos y requisitos del requerimiento de negocio.
- Cumplimiento con el índice de aceptación definido para los Casos de Prueba.
- Siempre que se han detectado defectos en los casos de prueba ejecutados, comunicar al constructor el defecto detectado y tipo de error que arroja la solución tecnológica. Inmediatamente, coordinar con el constructor la corrección del mismo confirmando su complejidad, decidiendo en conjunto continuar con la evaluación o el retorno de fase.

- Cuando la decisión es continuar con la evaluación, el Ingeniero de pruebas prepara el listado con los casos de prueba que se repetirán. El Constructor afina el error y notifica la solución tecnológica lista para ejecutar casos de prueba seleccionados.
- Cuando el afinamiento no surte efecto o cuando el defecto detectado requiere de una corrección de mayor complejidad, la actividad continúa en "Regresar solución a Construcción".
- En el caso de considerar inconforme el desarrollo, elaborar el informe de pruebas notificando la devolución del desarrollo a la etapa de construcción. En este caso la gestión continúa en la actividad "Regresar Solución tecnológica a Construcción".
- Cuando se considera que el desarrollo está conforme con las especificaciones o requisitos del requerimiento, aun cuando se presentan fallos de operación en la aplicación o sistema en ambiente de pruebas, notificar los fallos operacionales en pruebas funcionales a el área de infraestructura con la finalidad de asegurar la correcta configuración del ambiente preproducción de pruebas y acordar la liberación con el usuario solicitante y el constructor.
- Siempre que el desarrollo se califique satisfactorio, acordar la liberación de la solución tecnológica con el usuario solicitante y el constructor.

Notificar Incidentes Detectados

- En el caso de que el desarrollo probado no presente defectos y existan incidentes del ambiente de pruebas, sea este en pre-producción o en producción, notificar el incidente identificado y registrarlo en la herramienta de gestión vigente.

Emitir Informe de pruebas satisfactorio

- Siempre que la solución tecnológica supera las pruebas, detallar las condiciones en que se desarrollaron los casos de pruebas y los resultados obtenidos en el formato "Informe de Pruebas" de la herramienta de calidad vigente. El detalle incluye los casos ejecutados, el resultado de la ejecución de los mismos.

Liberar Solución Informática

- Siempre que una solución tecnológica se califique como satisfactorio en pruebas, confirmar que se ha anexado toda la documentación de esta etapa en la herramientas vigentes y registrar en la herramienta de gestión la aceptación y liberación de la solución tecnológica conjuntamente con el usuario solicitante, el constructor y miembro del equipo de SVC (cuando se lo considere en la herramienta de gestión vigente).

Emitir informe de pruebas y regresar solución a desarrollo

- Siempre que la solución tecnológica presente defectos y su liberación no sea aceptada, detallar las condiciones en que se desarrollaron los casos de pruebas y los resultados obtenidos en el formato "Informe de Pruebas" de la herramienta de calidad vigente describiendo el avance de las pruebas de la solución tecnológica y el o los defectos detectados que no permiten liberar la misma.
- Coordinar con el desarrollador las fechas para la ejecución de las pruebas nuevamente.

Gestión de Proveedores

- Los ingenieros del Área de Pruebas, realizan el control de tareas realizadas por los proveedores, dependiendo del cronograma entregado para el cumplimiento de los requerimientos, los cuales deben seguir y estar alineados a los procesos establecidos por la compañía, de esta manera se asegura la obtención del producto final.

2.8. BUSINESS INTELLIGENCE

2.8.1. DEFINICIONES

“Se denomina inteligencia de negocios o BI (del inglés *Business Intelligence*) al conjunto de estrategias y herramientas enfocadas a la administración y creación de conocimiento mediante el análisis de datos existentes en una organización o empresa.”³

“Las metodologías Business Intelligence utilizan la información para mejorar la gestión de las empresas. Gracias al software de BI, los usuarios pueden acceder y analizar los datos con facilidad, y tomar mejores decisiones.”⁴

³Wikipedia. Wikipedia, http://es.wikipedia.org/wiki/Inteligencia_empresarial.

⁴Gartner. Gartner, [En línea] 2012 <http://www.businessintelligence.info/mercado/gartner-business-intelligence-2012>.

“BI es una alternativa tecnológica y de administración de negocios, que cubre los aspectos del manejo de información para la toma de decisiones, desde su extracción en los sistemas, depuración, transformación, diseño de estructuras de datos o modelos especiales para el almacenamiento de datos, hasta la explotación de la información mediante herramientas comerciales de fácil uso para los usuarios. Éste concepto es llamado también DataWarehouse (DWH).”⁵

“BI es un proceso interactivo para explorar y analizar información estructurada sobre un área (normalmente almacenada en un datawarehouse), para descubrir tendencias o patrones, a partir de los cuales derivar ideas y extraer conclusiones. El proceso de Business Intelligence incluye la comunicación de los descubrimientos y efectuar los cambios. Las áreas incluyen clientes, proveedores, productos, servicios y competidores.”⁶

“Business Intelligence (BI) es una categoría amplia de soluciones de software de ordenador que permite a una empresa u organización para profundizar en sus operaciones a través de la presentación de informes críticos de las aplicaciones y herramientas de análisis.”⁷

Con las definiciones propuestas se concluye que Business Intelligence brinda herramientas que ayudan a la toma de decisiones, puesto que presenta la información de manera ordenada, rápida y oportuna, permite a las empresas profundizar en su análisis, ya que se tiene indicadores claves que ayudan a la gestión y manejo de la misma.

⁵**Gopac.** Gopac [En línea] 2012 <http://www.gopac.com.mx/v3/gopacbi/quees.asp>

⁶**Business Intelligence; Competir con Información** [En línea] 2012 http://www.iwith.org/pdf/Libro_BI_Competer_con_Informacion.pdf

⁷**Information Builders.** Ingormation Builders [En línea] 2012 <http://www.informationbuilders.com/business-intelligence>

BI permite la recolección de información efectiva, de manera que combinando datos y análisis, se pueda obtener el conocimiento necesario para tomar estrategias de mejora continua, y esto permita apalancar el buen funcionamiento de la planificación para cumplir los objetivos propuestos.

2.8.2. Características BI

Las principales características son:

- Ayuda en la toma de decisiones: posee herramientas de visualización avanzadas como: gráficas, tablas, velocímetros, que ayudan a obtener rápidos tiempos de respuesta, permite una gran navegabilidad, seleccionar y manipular información que le interese al usuario.
- Acceso a la información: brinda datos de calidad, completos, correctos y coherentes. Permite el ingreso a los datos de manera independiente.
- Orientación al usuario final: se busca el manejo de interfaces amigables, que permitan al usuario cierta intuición, sin necesidad de conocimiento técnico para su uso.

Proceso de BI

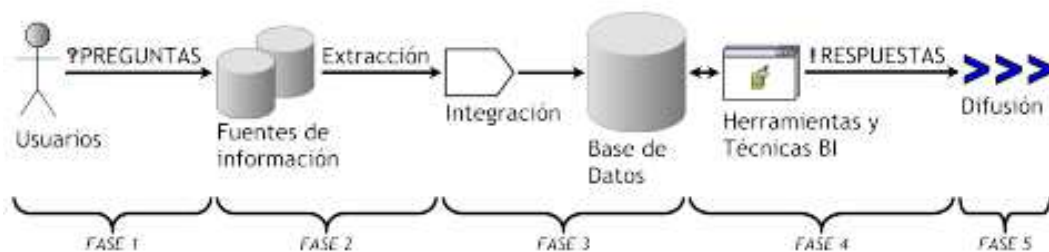


Figura 2.7: Proceso BI⁸

⁸Data Prix; [En línea] 2012 <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/1-business-intelligence/13-proceso-bi>

- 1. Dirigir y Planear:** Fase inicial en la cual se recolectan los requerimientos de la información específicos de los distintos usuarios, de esta manera se comprenderán sus distintas necesidades, que ayuden a generar las distintas preguntas que ayudarán a alcanzar los objetivos.
- 2. Recolección de Información:** Se realiza el proceso de extracción de distintas fuentes de información de la empresa, puede ser de manera interna o externa, esto ayudará a encontrar las respuestas a las preguntas planteadas en el paso anterior.
- 3. Procesamiento de Datos:** Se carga e integra los datos en su forma más rústica en un formato que se utiliza para el análisis, se crea una base de datos completamente nueva o se consolida la información.
- 4. Análisis y Producción:** Se procede a trabajar sobre datos extraídos e integrados, se utiliza las herramientas y técnicas que brinda BI, para crear la inteligencia, el resultado a las preguntas planteadas en un inicio se generan mediante creación de reportes, indicadores, entre otros.
- 5. Difusión:** En la fase final se entrega a los usuarios las herramientas adecuadas que les permitirá interactuar con los datos de manera sencilla y rápida.

2.8.3. Componentes de Business Intelligence

Es importante definir en un proyecto de Business Intelligence la información que se requiere para la gestión y toma de decisiones, además de donde obtendremos los datos y cuál es la disponibilidad y periodicidad que se requiere, establecer el formato y la navegabilidad que el usuario necesita.

2.8.3.1. Datawarehouse

“Un DataWarehouse es un conjunto de datos orientado a temas, integrado, no volátil, estable y que se usa para el proceso de toma de decisiones”. Un DataMart es un subconjunto sectorial del DW a menudo perteneciente a un departamento concreto.⁹

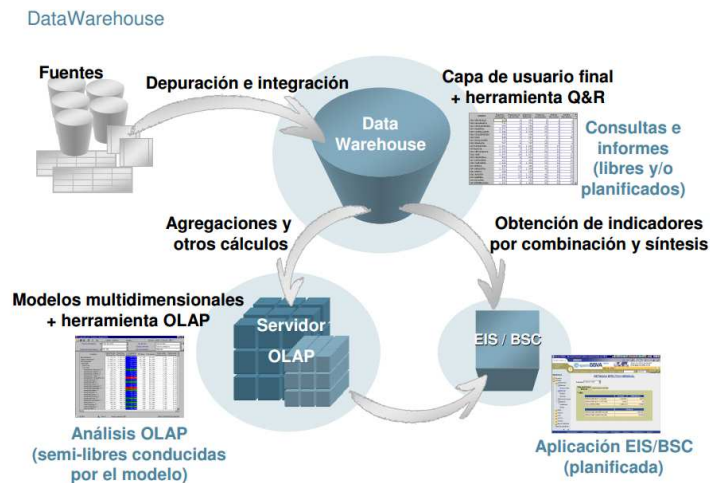


Figura 2.8: Datawarehouse¹⁰

Datawarehouse almacena una gran cantidad de información histórica del negocio, aísla los sistemas operacionales de las necesidades de información para la gestión. Un cambio en los sistemas operacionales no debe afectar al DW/DM. El DataWarehouse se alimenta a partir de los datos operacionales mediante las herramientas ETL (extract, transform, load).

^{10,9} **Gestión de Proyectos**, Gestión de Proyectos [en línea]
http://www.larioja.org/upload/documents/458269_gestion_proyectos.pdf

Extraer: hace referencia a la captura de información desde los sistemas operacionales, necesidad de integración:

- Bases de datos
- Ficheros planos
- Documentos impresos

Transformar: hace referencia a la adaptación de los datos fuente al formato destino definido en el DataWarehouse:

- Agregar datos numéricos
- Transponer información
- Componer unos textos a partir de otros

Cargar: hace referencia al proceso en que los nuevos datos son finalmente almacenados en el DataWarehouse en su formato definitivo.

2.8.3.1.1. Características DataWarehouse

Orientada al Negocio: la información está clasificada de acuerdo a los aspectos que son de interés para la empresa, la clasificación que se realiza afecta al diseño y la implementación de los datos encontrados, puesto que la estructura difiere de los procesos operacionales orientados a las aplicaciones.

Integrada: Implica que los datos que vienen de distintas fuentes, los cuales son producidos por departamentos, aplicaciones, entre otros, los mismos que pueden ser internos o externos, deben ser consolidados en una instancia antes que se agreguen a un DW.

Variante en el tiempo: existe la posibilidad que en el DW, se maneje grandes volúmenes de datos por esta razón el acceso puede demorar al realizar la consulta, pero esta característica es muy distinta a la información que se encuentra en el ambiente operacional, la cual se obtiene en el momento mismo del acceso.

No volátil: en el ambiente operacional del DW existe la carga y el acceso a los datos, por esta razón no cambian, no se requiere mecanismos de control de recuperación o concurrencia.

2.8.3.1.2. Estructura de Datawarehouse

DW está compuesto por diversos tipos de datos, los mismos que se dividen según el nivel de detalle:

1. Detalle de datos actuales: son datos de gran tamaño, se almacenan a nivel de detalle, la administración es compleja, puesto que se requiere el acceso sencillo y veloz a la información reflejan los eventos más recientes.
2. Detalle de datos históricos: este tipo de datos se almacenan a nivel de detalle y contiene información antigua para consulta no muy frecuente.
3. Datos ligeramente resumidos: son datos que tienen un bajo nivel de detalle y agrupan datos bajo una condición establecida de análisis.
4. Datos altamente resumidos: tienen fácil acceso, reducen aún más a los datos que ya han sido resumidos.
5. Metadatos: son una representación de la información acerca de los datos, el contenido no es tomado desde el ambiente operacional.

2.8.3.2. Query & Reporting

Son las herramientas que ayudan en la construcción de informes, los mismos que pueden ser en detalle o información agregada, estos datos se obtienen a partir de la información de DataWarehouse y DataMarts.

Este tipo de reportes se caracterizan por lo complejo de las consultas y altos tiempos de respuesta.

La explotación del Data Warehouse mediante "Query& Reporting" debe permitir una gradación de la flexibilidad de acceso, proporcional a la experiencia y formación del usuario. A este respecto, se recomienda el mantenimiento de al menos tres niveles de dificultad:

1. Los usuarios poco expertos podrán solicitar la ejecución de informes o consultas predefinidas según unos parámetros predeterminados.
2. Los usuarios con cierta experiencia podrán generar consultas flexibles mediante una aplicación que proporcione una interfaz gráfica de ayuda.
3. Los usuarios altamente experimentados podrán escribir, total o parcialmente, la consulta en un lenguaje de interrogación de datos.

Hay una extensa gama de herramientas en el mercado para cumplir esta funcionalidad sobre entornos de tipo Data Warehouse, por lo que se puede elegir el software más adecuado para cada problemática empresarial concreta.

2.8.3.2.1. Características de las Herramientas Reporting

- Informes dinámicos
- Información según niveles de usuario.
- Disponibilidad de formatos de salida de informes
- Distintos canales de salida

2.8.3.2.2. Entornos de gestión de informes

- Reglas de negocio
- Consultas asistidas
- Clusters de informes
- Perfiles de usuario

2.8.3.3. PROCESO ANALÍTICO EN LÍNEA (OLAP)

2.8.3.3.1. DEFINICIÓN

Herramientas OLAP se define como una tecnología de software que permite realizar el análisis de información en línea, ayuda a realizar la administración y ejecución de consultas, es así como infiere en la capa de negocio, para la toma de decisiones, brinda respuestas rápidas a preguntas complejas, se tiene la opción del análisis histórico de la información visualizando de esta manera el comportamiento y evolución de la misma.

2.8.3.3.2. CARACTERÍSTICAS

- Permite disponer la información en diferentes formatos como: tablas, gráficos, reportes lo que ayuda a mantener organizada la información necesaria para los usuarios.
- Ayuda en el análisis complejo de un gran volumen de datos, los mismos que pueden ser pesados y no predecibles
- Permite definir dimensiones (sin límites) para analizarlas mediante jerarquías y combinaciones.
- Los cubos OLAP se generan mediante esquemas sobre el DW, las dimensiones definen dominios, los miembros de una dimensión se agrupan de forma jerárquica, cada celda del cubo contienen datos que relacionan elementos de las dimensiones.

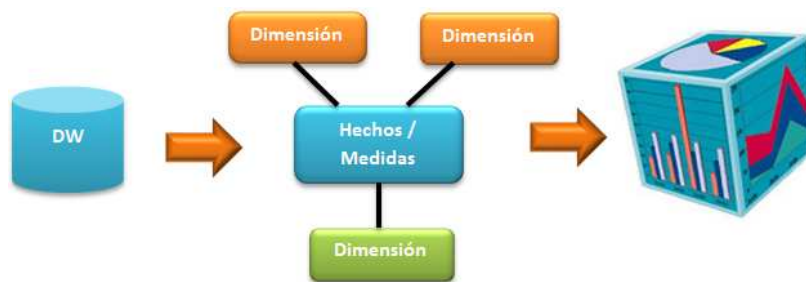


Figura 2.9: Formación de Cubos OLAP: Tesistas

2.8.3.3.3. ESQUEMAS

Esquema en Estrella

Se tiene una sola tabla de hechos (tabla fact), la cual tiene datos para realizar el análisis rodeada de tablas dimensiones, esta tabla se envuelve de tablas más pequeñas, se asemeja a una estrella, es importante destacar que la tabla de dimensiones tiene una clave primaria simple a diferencia de la tabla de hechos, en donde la clave principal está compuesta por claves primarias de las tablas dimensionales.

La principal ventaja de este esquema es su velocidad y simplicidad para realizar análisis multidimensionales, además para el usuario final las consultas no son complicadas. El esquema estrella por su rendimiento permite indexar dimensiones de manera individual sin que afecte al rendimiento de la base de datos.

Esquema Copo de Nieve

En este esquema las tablas de las dimensiones son normalizadas de tal manera que se simplifiquen las operaciones de selección de datos, se presenta la información sin que exista redundancia.

Usar este esquema permite segregar datos de las dimensiones y otorgar un esquema que sustente los requerimientos de diseño, a pesar que posee una mayor complejidad en su estructura, hace una mejor utilización en el espacio, se puede desarrollar clases de jerarquías fuera de las dimensiones, no se debe utilizar este esquema en donde el tiempo de respuesta sea un factor de criticidad para el usuario final.

2.8.3.4. CUADROS DE MANDO

Los cuadros de mando manejan gran cantidad de información de manera visual permitiendo que sean prácticos y entendibles, otorgando al usuario recursos visuales, el entorno es muy intuitivo.

Es muy factible medir el funcionamiento de una empresa con métricas clave (KPI's, Key Performance Indicators), de esta manera se tiene una clara visión si se está consiguiendo los objetivos propuestos, se realizan planes de acción o se toma medidas para que no afecte en el resultado global

En los cuadros de mando se manejan distintos indicadores que se los mencionarán a continuación:

Los indicadores deben cumplir con los siguientes requisitos:

- **Selectivos:** Se centra en lo que es importante, evitando analizar indicadores que no son relevantes.
- **Informativos:** Permiten concluir rápidamente la información asociada.
- **Oportunos:** Se refiere a la disponibilidad de la información en el momento preciso para la toma de decisiones.
- **Sencillos:** Realizarlos de tal manera que resulten explicativos e intuitivos, evitando sofisticación en el método de obtención y contenido
- **Aceptados:** Los indicadores deben ser útiles y permitidos por los miembros de la organización.

CAPÍTULO 3

METODOLOGÍA PARA EL DESARROLLO DE BUSINESS INTELLIGENCE

3.1. METODOLOGÍAS PARA EL DESARROLLO DE BI

Existen varias tecnologías, utilizadas para el desarrollo e implementación de Soluciones de BI, entre las que encontramos:

3.1.1. Metodología Inmon

Esta metodología fue desarrollada por Bill Inmon en el año 1992 en su libro “Building the Data Warehouse”. Para esta metodología un Data Warehouse es considerado como un conjunto de datos orientados por temas, integrados, variantes en el tiempo y no volátiles, cuyo objetivo es ser apoyo importante para la toma de decisiones estratégicas. El Data Warehouse es una parte de un sistema de BI y los Data Marts obtienen su información a partir de este Data Warehouse y para ser almacenada la información es necesario normalizarla antes de almacenarse.

Debido a las grandes cargas de datos que debe realizarse, Inmon considera que el ambiente de origen de los datos y el ambiente de acceso de datos deben estar físicamente separados en diferentes bases de datos y en equipos separados, y los datos deben estar cuidadosamente gestionados y condensados. En esta metodología los DM son considerados como capas del

DWH y los DM son dependientes del depósito central de datos o DWH Corporativo y por lo tanto se construyen después de él.

Arquitectura Inmon

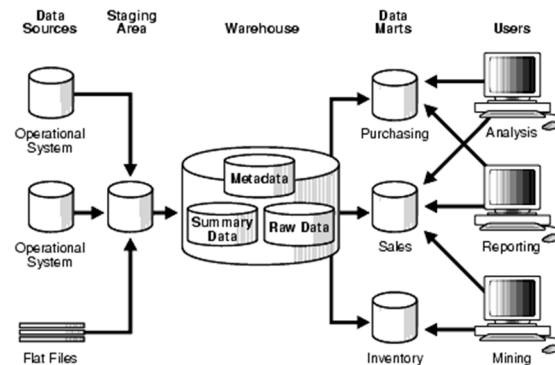


Figura 3.1: Arquitectura Inmon¹¹

3.1.2. Metodología Kimball

En esta metodología el Data Warehouse se forma del conjunto de todos los Data Marts generados existentes en una empresa. La información siempre se almacena en un modelo dimensional. Otra forma de denominar ésta aproximación es como "Bottom-up"

En la metodología Kimball un Data Warehouse se considera como una copia de los datos transaccionales para realizar la consulta y el análisis de los mismos. La metodología de Kimball, también conocida como Modelo Dimensional (Dimensional Modeling), está basado en el Ciclo de Vida Dimensional del Negocio (Business Dimensional Lifecycle).

¹¹ **Catarina. 2010.** [En línea] 2010. [Citado el: 05 de Septiembre de 2012.] http://catarina.udlap.mx/u_dl_a/tales/documentos/lis/rosette_u_jm/capitulo3.pdf.

Para almacenar los datos, las estructuras de estos requieren una presentación dimensional. En el Modelo Dimensional se constituyen modelos de tablas y relaciones con el propósito de optimizar la toma de decisiones, basándose en las sentencias Sql realizadas en una base de datos relacional que están ligadas con la medición o un conjunto de mediciones de los resultados de los procesos de negocio.

Este Modelo Dimensional, tiene como objetivo presentar los datos dentro de un marco de trabajo estándar e intuitivo, para permitir su acceso con un alto rendimiento. Cada Modelo Dimensional está compuesta por una tabla con una llave combinada, llamada tabla de hechos, y con un conjunto de tablas más pequeñas llamadas tablas de dimensiones. Los elementos de estas tablas se pueden definir de la siguiente manera:

- **Hechos:** es una colección de piezas de datos y datos de contexto. Cada hecho representa una parte del negocio, una transacción o un evento.
- **Dimensiones:** es una colección de miembros, unidades o individuos del mismo tipo.
- **Medidas:** son atributos numéricos de un hecho que representan el comportamiento del negocio relativo a una dimensión.

Cada punto de entrada a la tabla de hechos está conectado a una dimensión, lo que permite determinar el contexto de los hechos.

Una base de datos dimensional se puede concebir como un cubo de tres o cuatro dimensiones (OLAP), en el que los usuarios pueden acceder a una porción de la base de datos a lo largo de cualquiera de sus dimensiones.

Dado que es muy común representar a un modelo dimensional como una tabla de hechos rodeada por las tablas de dimensiones, frecuentemente se le denomina también modelo estrella o esquema de estrella-uni3n

Arquitectura Kimball

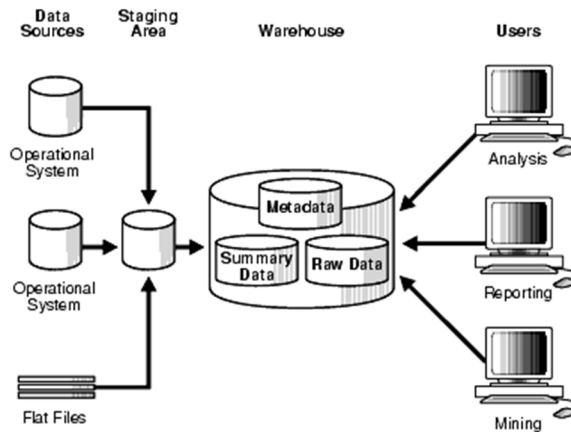


Figura 3.2: Kimball¹²

3.1.3. Tipos de Arquitectura

Top-Down: con esta arquitectura se define inicialmente DW y despu3s se realiza el desarrollo, y se construyen y cargan los DM

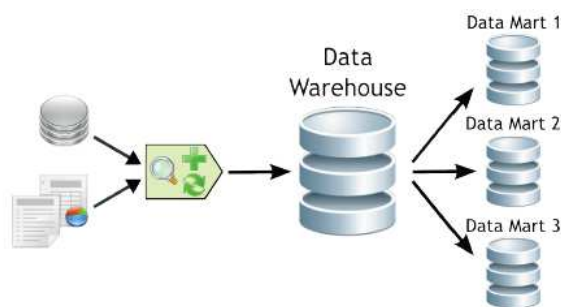


Figura 3.3: Top Down¹³

¹² Ing. Bernabeu, Ricardo Darío.(2010).HEFESTO: Metodología para la Construcción de un Data Warehouse. [En Línea][Citado el 26,08 de 2012]

¹³Ing. Bernabeu, Ricardo Darío.(2010).HEFESTO: Metodología para la Construcción de un Data Warehouse. [En Línea][Citado el 26,08 de 2012]

Bottom-Up: Con este tipo de arquitectura, se definen previamente los DM y luego se integran en un DW centralizado.

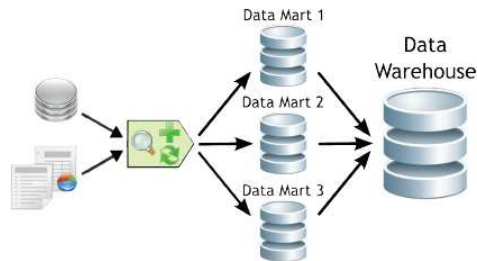


Figura 3.4: Bottom Up¹⁴

Híbrida: con este tipo de arquitectura se puede realizar el desarrollo de un DM o un DW, de acuerdo al alcance empresarial y la magnitud de datos y fuentes de análisis.

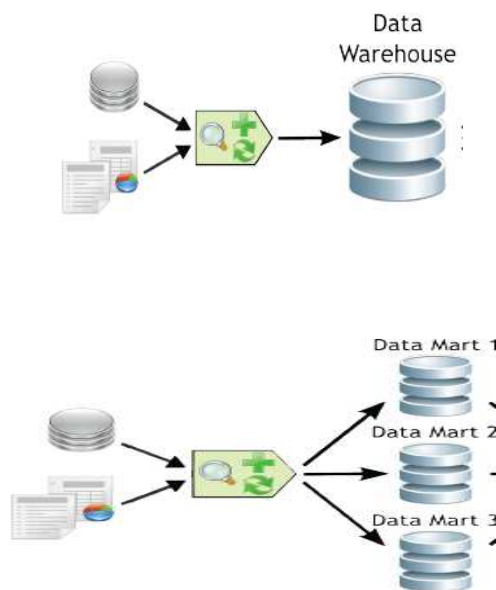


Figura 3.5: Híbrida¹⁵

¹⁴Ing. Bernabeu, Ricardo Darío.(2010).HEFESTO: Metodología para la Construcción de un Data Warehouse. [En Línea][Citado el 26,08 de 2012] <http://www.businessintelligence.info/docs/hefesto-v2.pdf>.

¹⁵Ing. Bernabeu, Ricardo Darío.(2010).HEFESTO: Metodología para la Construcción de un Data Warehouse. [En Línea][Citado el 26,08 de 2012] <http://www.businessintelligence.info/docs/hefesto-v2.pdf>.

Comparación de Metodologías

Tabla 3.1 Comparación de Metodologías – Tesistas

Nombre Metodología	INMON	KIMBALL	HEFESTO
Autor	Bill Inmon	Ralph Kimball	Ing. Bernabeu Ricardo Darío
Arquitectura	Top-Down	Bottom-Up	Híbrida
Énfasis	Data Warehouse	Data Mart	Data Mart y Data Warehouse
	Los problemas que se van a solventar ya son conocidos de antemano , tiene un mayor costo ya que cada fase alcanza un alto nivel de detalle	Está basada en experimentos y prototipos , es más flexible y menos costosa	Se enfoca en el análisis de los requerimientos para identificar indicadores , perspectivas y realizar el análisis de los datos

3.1.4. Selección de Metodología

Se ha seleccionado la Metodología Hefesto, debido a que las fases que esta comprende, se acopla al desarrollo y está enfocada al análisis de requerimientos, identificación de identificadores y perspectivas, que permitirán realizar el análisis de datos respectivos y obtener resultados que apoyen a la toma de definiciones de la organización.

3.2. METODOLOGÍA HEFESTO

HEFESTO es una metodología propia para la construcción e implementación de un DW y se puede adaptar a cualquier ciclo de vida de desarrollo de software. Esta metodología permite realizar una primera implementación que satisface una parte de las necesidades, para demostrar sus ventajas y motivar a los usuarios.

Características:

- Los objetivos y resultados finales de cada fase se diferencian fácilmente y son sencillos de comprender.
- El usuario se encuentra involucrado estrechamente al final en cada etapa, para que tome decisiones respecto al comportamiento y funciones del DW.
- Utiliza modelos conceptuales y lógicos, sencillos de interpretar y analizar.
- Es independiente de las herramientas que se utilicen para su implementación.
- Es independiente de las estructuras físicas que contengan el DW y de su respectiva distribución.

- Cuando se finaliza con una fase o paso, los resultados obtenidos al terminar, sirven como punto de partida para llevar a cabo el siguiente paso.
- Se puede aplicar tanto para la implementación de DM como para DW.
- Está basada en los requerimientos del usuario, lo que permite adaptarse con facilidad y rapidez ante los cambios en el negocio.
- La metodología HEFESTO permite la realización de Data Warehouse o un Data Mart al mismo tiempo

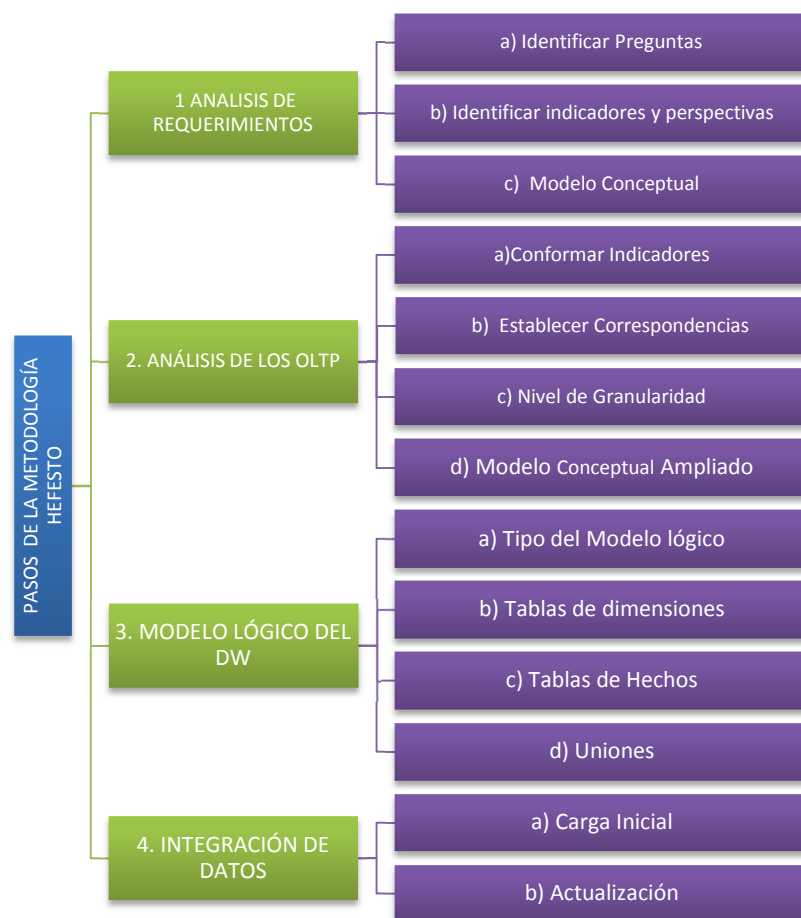


Figura 3.6: Fases de la metodología Hefesto v2.0: Tesistas

3.2.1. FASES DE LA METODOLOGÍA HEFESTO V2.0

3.2.1.1. ANÁLISIS DE REQUERIMIENTOS

Inicialmente se debe identificar los requerimientos de los usuarios, esta información se obtendrá mediante la elaboración de preguntas, que permitan conocer los objetivos de la organización. Para después de un análisis, poder determinar cuáles serán los indicadores y las perspectivas que se utilizarán que para la construcción del DW o DM.

Al final se realizará un modelo conceptual que consolide la información de los indicadores, perspectivas, información que permitirá tener una visualización global del análisis inicial.

3.2.1.1.1. Identificar Preguntas

Por la necesidad de adquirir información sobre los requerimientos de los usuarios, es posible aplicar diferentes técnicas, tales como entrevistas, cuestionarios, observaciones, entre otros.

Como punto de partida, se realiza el análisis de los requerimientos de los diferentes usuarios, ya que de ello depende, que el desarrollo muestre lo que se espera del depósito de datos, en relación a sus funciones y cualidades.

Principalmente se busca obtener e identificar las necesidades de información clave de alto nivel que proveerá el soporte para desarrollar las siguientes etapas, para cumplir con las metas y estrategias de la empresa lo que apoyará en los procesos de toma de decisiones.

Un análisis adecuado, permitirá conocer claramente los objetivos estratégicos planteados por la empresa, y sus principales procesos.

Para realizar el estudio y análisis de la información, se debe formular preguntas complejas sobre el negocio, que contemplen variables de análisis que se consideren relevantes, y permitan obtener fuentes de información notables e importantes.

3.2.1.1.2. Identificar indicadores y perspectivas de análisis

Con la información obtenida de las preguntas realizadas, se procede a determinar cuáles serán los indicadores y las perspectivas con los que se va a trabajar.

Al referirse a **Indicadores**, se engloba todos aquellos valores numéricos que representan lo que se desea analizar concretamente (saldos, promedios, cantidades, sumatorias, fórmulas) mientras que las **Perspectivas** son aquellos objetos de los que se analizará los indicadores (clientes, productos, proveedores).

3.2.1.1.3. Modelo Conceptual Ampliado

En esta etapa, se construirá un modelo conceptual a partir de los indicadores y perspectivas obtenidas del análisis anterior para poder determinar los alcances del proyecto sobre los cuales se va a trabajar y que puede ser presentado a los usuarios

En la parte izquierda del modelo se listan las perspectivas identificadas, que se unen a un óvalo central que representa la relación que existe entre ellas. La **relación**, es el proceso o área de estudio elegida. De dicha relación, se desprenden flechas que indican cuales son los indicadores o medidas, que se ubican a la derecha del esquema.

El modelo conceptual muestra una vista global sobre los resultados finales, las variables que se utilizarán para analizarlos y cuál es la relación que existe entre ellos.

Representación gráfica del modelo conceptual:

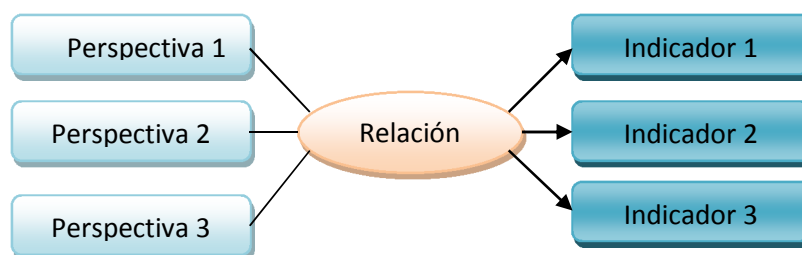


Figura 3.7: Modelo Conceptual: Tesistas

ANÁLISIS DE LOS OLTP

En esta fase se efectuara un análisis de las fuentes OLTP para determinar cómo serán calculados los indicadores y así establecer las respectivas correspondencias que existen entre el modelo conceptual creado y las fuentes de datos que se tiene.

Después se definirá los campos que deben estar incluidos en las perspectivas, esta información estará contenida en un modelo conceptual ampliado.

3.2.1.2.1. Conformar Indicadores

Se explicara cuál será la forma de cálculo para obtener los indicadores definiendo de cada uno:

- Hecho/s que lo componen, con su respectiva fórmula de cálculo.
- Función de sumarización que se utilizará para su agregación.

3.2.1.2.2. Establecer Correspondencias con los requerimientos

En esta etapa, se examinan los OLTP con los que se va a trabajar, que contienen características e información necesaria, para definir las correspondencias existentes entre el modelo conceptual y las fuentes de datos.

De los indicadores, se deberá definir cuál será la forma de cálculo, sea mediante fórmulas u operaciones complejas.

Nivel de Granularidad

Se seleccionan los campos que integrarán las perspectivas encontradas. Cuando se ha determinado las relaciones con los OLTP se procederá a definir los campos que conformarán cada perspectiva, por los que se manipularán y filtrarán a los indicadores.

Con las correspondencias establecidas, se mostrará al usuario los datos de análisis para las perspectivas. Se debe tener claro el significado de los campos o valores de los OLTP, para lo cual se debe programar reuniones con el personal encargado del sistema, para el análisis de los datos.

Se debe presentar al o los usuarios, los datos existentes y el significado de los mismos, con sus posibles valores y características para que los usuarios puedan determinar qué datos tienen mayor importancia para obtener indicadores.

En el modelo conceptual se añadirá bajo cada perspectiva los campos o atributos para poder mostrar gráficamente los resultados obtenidos.

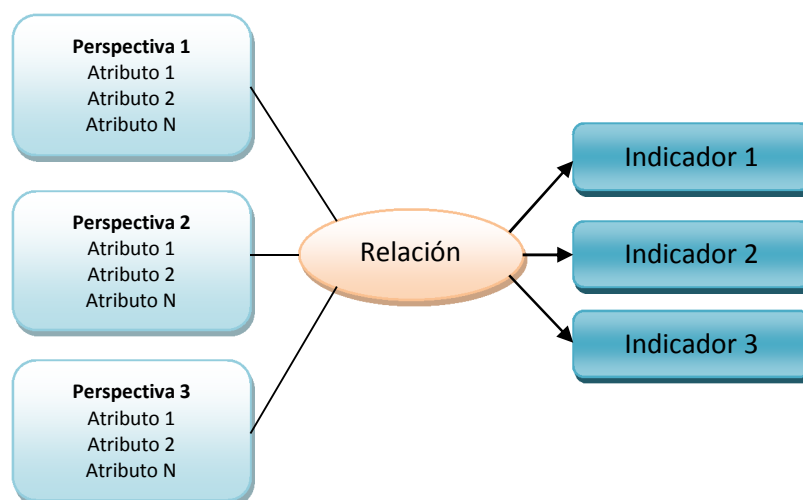


Figura 3.8: Modelo Conceptual con atributos: Tesistas

3.2.1.2.3. Modelo Conceptual Ampliado

Al modelo Conceptual definido en pasos anteriores, se lo ampliará añadiendo los campos y atributos en las perspectivas, y se colocará bajo cada indicador, su fórmula de cálculo respectiva

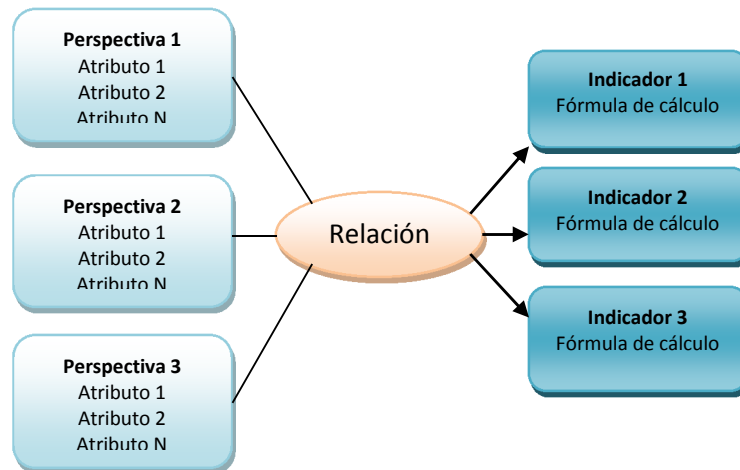


Figura 3.9: Modelo Conceptual ampliado: Tesistas

3.2.1.3. MODELO LÓGICO DEL DW

En esta etapa se creará un modelo lógico de la estructura de DW, el cual se basará en el modelo conceptual ampliado, para lo cual se define inicialmente el tipo de modelo lógico, se diseñan las tablas de dimensiones y de hechos, para realizar las uniones necesarias

3.2.1.3.1. Tipo del Modelo lógico

Se debe definir el esquema (estrella, constelación o copo de nieve) que se ajuste estrechamente a las necesidades y requerimientos de los usuarios.

Tablas de dimensiones

Se debe crear las dimensiones del esquema lógico, tomando cada perspectiva con sus atributos correspondientes para:

- Elegir un nombre que identifique la dimensión.
- Añadir un campo que represente su clave principal.
- Redefinir nombres de los atributos si no son explicativos.

Representación Gráfica:

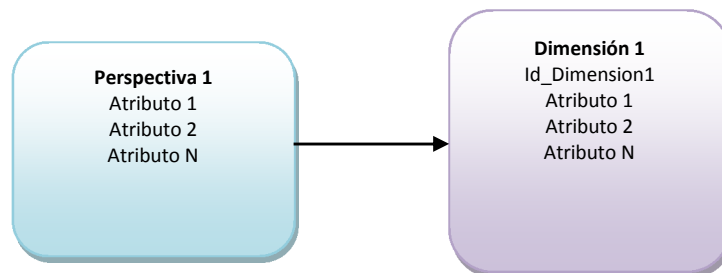


Figura 3.10: Diseño de tablas de dimensiones: Tesistas

3.2.1.3.2. Tablas de Hechos

Las tablas de hechos contienen, los hechos que serán utilizados por los analistas de negocio para apoyar el proceso de toma de decisiones. Incluyen datos cuantitativos de los cuales se construirán los indicadores de estudio.

Esquemas en Estrella y Copo de nieve:

- Asignar un nombre a la tabla de hechos que indique la información analizada, área de estudio, negocio enfocado.
- Definir una clave primaria, compuesta por la combinación de las claves primarias de cada tabla de dimensión relacionada.
- Crear campos de hechos e indicadores, de acuerdo a los definidos en el modelo conceptual.

Representación Gráfica:

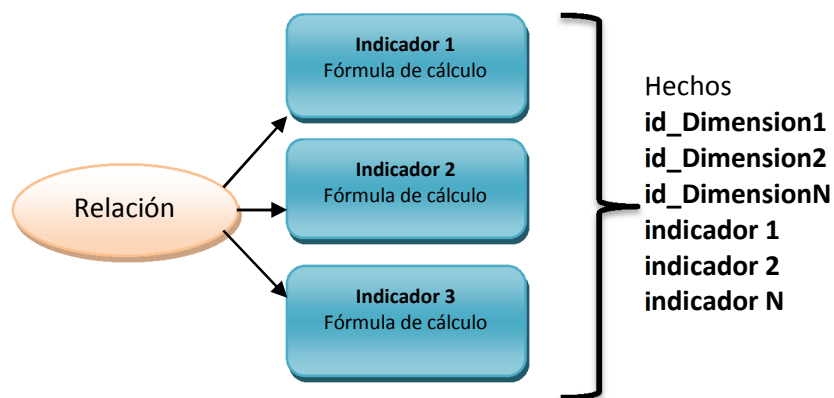


Figura 3.11: Tabla de hechos (Esquema estrella y copo de nieve): Tesistas

Esquemas Constelación:

- Las tablas de hechos se construyen tomando en cuenta el análisis de las preguntas realizadas por los usuarios
- Los nombres que lleven las tablas de hechos deben ser claros e identificativos, estas tablas contendrán los hechos correspondientes y su clave estará formada por la combinación de las claves de las tablas de dimensiones relacionadas.

Para el diseño de tablas de hechos, se debe tomar en cuenta los siguientes casos:

Caso 1: Si en dos o más preguntas del negocio se encuentran los mismos indicadores, con diferentes perspectivas de análisis, entonces el número de tablas de hecho existente será igual al número de preguntas que cumplan esta condición.

"Analizar el **Indicador1** por Perspectiva1 y por Perspectiva2"
"Analizar el **Indicador1** por Perspectiva2 y por Perspectiva3"

Figura 3.12: Caso 1, preguntas.¹⁶

Caso 2: Si existen distintos indicadores que se encuentran incluidos en dos o más preguntas del negocio, con diferentes perspectivas de análisis, existirán tantas tablas de hechos como preguntas cumplan esta condición.

"Analizar el **Indicador1** por Perspectiva1 y por Perspectiva2".
"Analizar el **Indicador2** por Perspectiva2 y por Perspectiva3".

Figura 3.13: Caso 1, preguntas.¹⁷

Caso 3: Si el conjunto de preguntas de negocio cumplen con los casos 1 y 2, se deberán unir los interrogantes que tengan diferentes indicadores pero iguales perspectivas de análisis, para luego reanudar el estudio de las preguntas

"Analizar el **Indicador1** por Perspectiva1 y por Perspectiva2".
"Analizar el **Indicador2** por Perspectiva1 y por Perspectiva2".

Figura 3.14: Caso 1, preguntas.¹⁸

^{13 17 15}Ing. Bernabeu, Ricardo Darío.(2010).HEFESTO: Metodología para la Construcción de un Data Warehouse. [En Línea][Citado el 26,08 de 2012] <http://www.businessintelligence.info/docs/hefesto-v2.pdf>.

3.2.1.3.3. Uniones

Las uniones correspondientes entre las tablas de dimensiones y tablas de hechos, se realizan para los tres tipos de esquemas.

3.2.1.4. INTEGRACIÓN DE DATOS

El modelo lógico que se ha construido debe ser puesto a prueba con datos, utilizando técnicas calidad de datos como ETL, para definir reglas y normas para actualización

3.2.1.4.1. Carga Inicial

Se realiza una carga Inicial del DW, alimentando el modelo creado. Se debe llevar a cabo tareas de limpieza de datos, calidad de datos, procesos ETL, que tienen una lógica compleja.

En el DW debe contener datos válidos, quitando valores faltantes o anómalos, para así utilizar únicamente los datos útiles

En un *Esquema Constelación*, varias tablas de dimensiones comparten diferentes tablas de hechos, donde se cargan los datos de las dimensiones y luego los de las tablas de hechos, manteniendo, la correspondencia entre cada elemento.

En un *Esquema Copo de Nieve*, cuando al haber jerarquías de dimensiones, se comienza haciendo la carga de las tablas de dimensiones desde el nivel más general al más detallado.

3.2.1.4.2. Actualización

La actualización se la realizada teniendo la carga de datos en el DW:

- Establecimiento de políticas y estrategias ejecutar actualizaciones o refresco de datos.
- Especificar tareas de limpieza de datos, calidad de datos, procesos ETL, para actualizar los datos del DW.
- Especificar de forma general y detallada las acciones que deberá realizar cada software.

CAPÍTULO 4

HERRAMIENTAS

4.1. Motor de Base de Datos Oracle

4.1.1. Introducción

Es un sistema de gestión de base de datos que brinda un modelo objeto-relacional. Es una herramienta cliente/servidor que permite almacenar y recuperar información relacionada.

Oracle maneja instancias las cuales consta de una estructura de memoria, llamada Área Global del Sistema (SGA) y de procesos background que son utilizados por el servidor para el manejo de la base de datos. Es decir que cada instancia puede utilizar una base de datos en cualquier punto o momento.

La Base de Datos Oracle se almacena físicamente en ficheros, la correspondencia entre ficheros y tablas se realiza gracias a la estructura interna de la base de datos, que permite que distintos tipos de datos se almacenen físicamente separados, esta división lógica se realiza por los espacios de las tablas (tablespaces).

4.1.2. Características

- Usuarios concurrentes
- Portabilidad
- Gestión de seguridad
- Gestión de grandes bases de datos
- Bajos tiempos de respuesta
- Compatibilidad

4.1.3. Arquitectura Oracle

Tenemos tres componentes básicos:

- Estructuras de memoria
- Procesos
- Archivos



Figura 4.1. Arquitectura Oracle¹⁹

¹⁹Arquitectura de Oracle; Arquitectura Oracle [En línea] 2012
<http://xue.unalmed.edu.co/~mfcabrera/db/arqoracle.pdf>

Estructuras de Memoria

Existen dos clases de memoria, una que comparten todos los usuarios conectados y otra que se dedica al trabajo de cada uno de ellos.

El área global del sistema (SGA), es el área compartida por todos los usuarios, también se define como una región de memoria compartida, la cual contiene datos e información de control para una instancia, es así como procesos en segundo plano más SGA constituyen una instancia en Oracle, cabe recalcar que Oracle asigna el SGA al momento que se inicia una instancia y desasigna la misma cuando la instancia se detiene.

Se recomienda para un rendimiento adecuado el SGA debe ser tan grande como sea posible.

SGA se divide en tres partes:

1. **Fondo común compartido (Shared pool)**, la cual contiene el diccionario de datos, órdenes SQL que se solicitan para el procesamiento, además de buffers para mensajes de ejecución paralelas y estructuras de control.
2. **Área de Memoria Rápida (Database buffer cache)**, contiene copias de los bloques de datos de lectura de las páginas, mantiene datos de las órdenes SQL de los usuarios conectados concurrentemente a la base de datos, permitiendo así se reduzca la contención de sistemas multiprocesador.

3. **Área de Registros de rehacer (Redo log buffer)**, contiene información de los cambios realizados a la base de datos, posee operaciones INSERT, UPDATE, DELETE, CREATE, ALTER, DROP, permitiendo recuperar datos si así se lo requiere.

Por otro lado por cada sesión de usuario que se establece, se crea también en memoria, un área que es específica denominada área global de programa PGA, esta área no es compartida con sesiones de otros usuarios.

PGA es una región de memoria que contiene datos e información que permiten el control de procesos de los servidores.

La memoria PGA se clasifica de la siguiente manera:

1. **Memoria de Sesión:** se asigna para mantener variables de una sesión y otra información que sea relativa a la sesión.
2. **Área SQL privada:** contiene consultas de información de ejecuciones y consultas de ejecuciones, la sesión que establece una sentencia tiene un área privada de SQL, así mismo el usuario que realiza una sentencia también tiene un área SQL privada y usa un área SQL compartida.

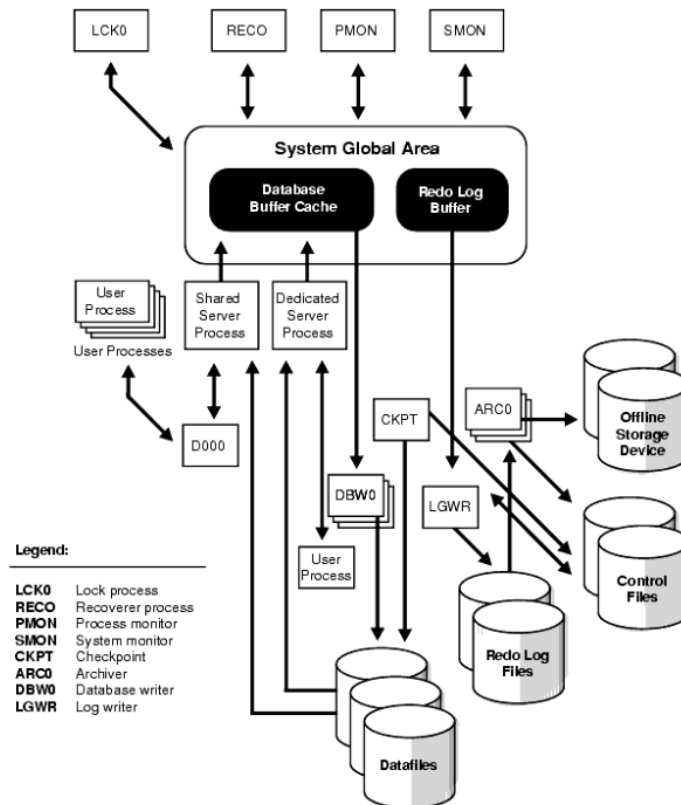


Figura 2.2: Estructuras de memoria y procesos de Oracle²⁰

4.2. Solución BI

En la actualidad la exigencia de mercado y la demanda de clientes sumado con la competencia global, se ha vuelto crítico la necesidad de obtener información de manera rápida y con la mayor veracidad, por esta razón cada vez más organizaciones acuden a soluciones de inteligencia de negocios y así mejorar la toma de decisiones y la agilidad comercial.

²⁰Estructura de Memoria; ; Arquitectura Oracle [En línea] 2012 <http://space.itpub.net/9537053/viewspace-735280>

Por esta exigencia la comunidad open source brinda programas libres para generar inteligencia artificial las cuales se orientan a la solución y se centran en los procesos, es así como una plataforma permite crear soluciones complejas a problemas de negocio, las soluciones que muestra este tipo de herramientas se componen de infraestructura de herramientas de análisis e informes integrados, lo cual asegura que se pueda crear las reglas de negocio necesarias. Una muestra clara de solución es la herramienta Pentaho.

4.2.1. Pentaho

Open Source Pentaho Business Intelligence, se define como una plataforma “orientada a la solución” y “centrada en procesos”, su objetivo principalmente se basa en cubrir amplias necesidades de análisis de datos y reportes empresariales, agilitando de esta manera la toma de decisiones. Las soluciones de Pentaho están escritas en Java al igual que su ambiente de implementación, lo que permite que sea flexible y cubra una amplia gama de necesidades ya sean específicas o sofisticadas. Brinda la posibilidad de que la información sea clocada en la intranet de la empresa.

Pentaho abarca las soluciones brindando diferentes opciones como son: reportes, análisis, dashboards, minería de datos, consultas, procesos ETL.

4.2.1.1. Características de Pentaho

- Pentaho es una plataforma orientada a la solución y centrada en los procesos
- Pentaho permite presentar información rápida y oportuna
- Permite realizar análisis de datos para generar informes empresariales
- Las soluciones Pentaho, está escrita en Java y su ambiente de implementación está basado también en Java
- Pentaho está construido bajo un servidor de aplicaciones compatibles con J2EE como JBOSS AS, WebSphere, Tomcat, WebLogic y Oracle AS, lo que asegura la escalabilidad, integración y portabilidad.
- Pentaho maneja servicios críticos tales como programación, seguridad, integración automatización y flujo de trabajo.
 - Maneja la Integración con los procesos del negocio
 - Permite administrar y programar reportes
 - Maneja administración para la seguridad de usuarios
- Se puede trabajar con base de datos: vía JDBC, IBM DB2, Microsoft SQL Server, MySQL, Oracle, PostgreSQL, NCR Teradata, Firebird.
- No tiene dependencia de sistema operativo
- Lenguaje de programación: Java, Javascript, JSP, XSL (XSLT/XPath/XSL-FO).
- Interfaz de desarrollo: Java SWT, Eclipse, Web-based.
- Repositorio de datos basado en XML.
- Sus componentes están expuestos vía Web Services lo que apoya en la integración con Arquitecturas Orientadas a Servicios (SOA).

Tabla 4.1. Productos Pentaho. Tesistas

Herramienta	Característica	
Pentaho Reporting	Desarrollo de informes	
Pentaho Analysis	Análisis de la información	
Pentaho Dashboard	Mejorar el rendimiento organizacional	
Pentaho Data Integration	Integrar información	
Pentaho Data Mining	Entender de mejor manera al negocio	
Pentaho BI Platform	Administra y programa reportes	

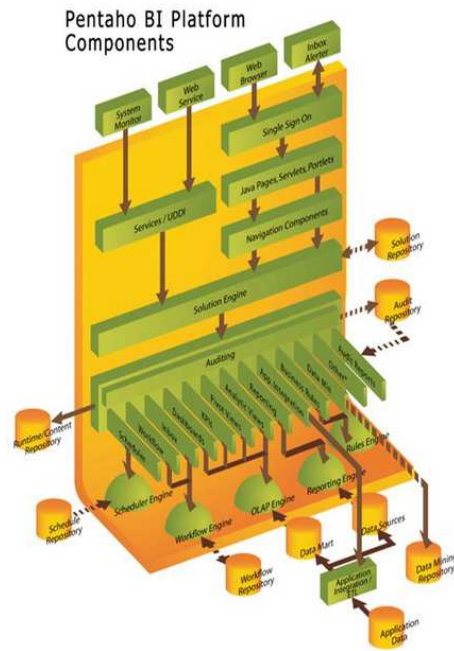


Figura 4.3: Componentes Pentaho²¹

²¹**Componentes Pentaho;** Componentes Pentaho [En línea] 2012
<http://www.gravitar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/>

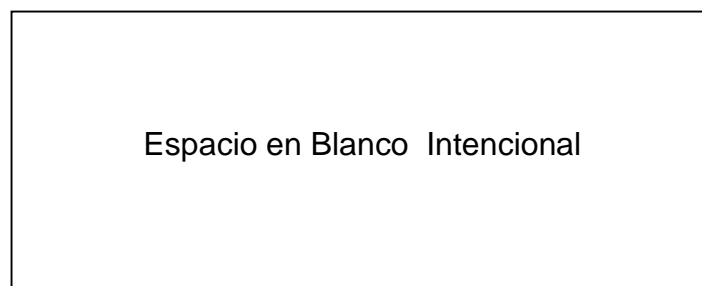
4.2.1.2. Arquitectura Pentaho

La solución Business Intelligence OpenSource Pentaho es una alternativa a las soluciones de negocio tradicionales más completas por lo que incluye varios componentes para agilizar la manera de obtener información y analizarla.

Los informes se generan mediante la herramienta JfreeReport en formatos (HTML, Excel, pdf). En cuanto a la generación de PDFs lo realiza mediante Apache FOP, gracias a la librería JPivot, con la cual se pueden ver tablas OLAP a través de un browser.

Integra Pentaho Report Design Wizard, es una herramienta que permite el diseño de informes, para la generación de gráficos es a través de JFreeChart, una librería para generar los gráficos, interfaces para acceder a diferentes fuentes de datos, exportación a PNG, JPEG y PDF y soporte para servlets, JSPs, applets y aplicaciones clientes.

Las soluciones que se elaboran se pueden publicar sin ningún inconveniente como servicio web, gracias al motor de servicios web Apache Axis, esta es una característica fundamental de Pentaho.



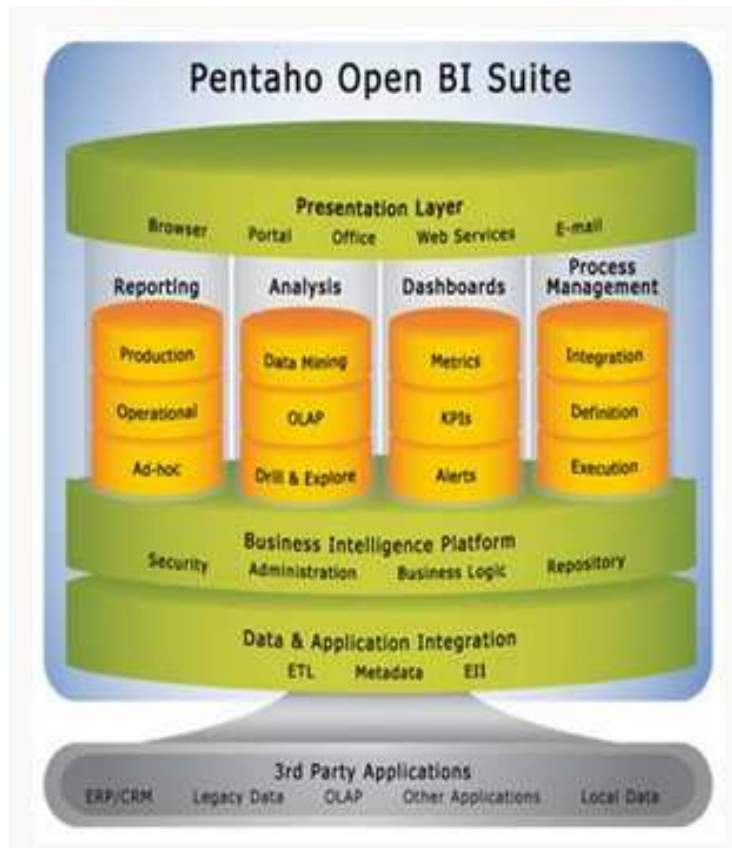


Figura 4.4: Arquitectura Pentaho²²

4.2.1.3. Pentaho Reporting

El desarrollo de Pentaho Reporting es una solución proporcionada por Pentaho que viene integrada en su suite y tiene como objetivo, tener un motor de informes flexible y fácil de usar. Pentaho Reporting permite transformar los datos en información significativa y adaptarla a los usuarios.

Pentaho Reporting incluye **Pentaho Report Designer (PRD)** y **Pentaho Report Design Wizard (RDW)** que son utilizados para crear reportes definidos en un editor gráfico. Pentaho Report Designer puede además ser utilizado como una potencial herramienta de reportes.

²²**Arquitectura Pentaho;** Arquitectura Pentaho [En línea] 2012
<http://www.gravitar.biz/index.php/herramientas-bi/pentaho/arquitectura-pentaho/>

Los reportes que se han definido son ejecutados por el **Pentaho Reporting Engine** (PRE), que es una librería de reportes en Java que puede ser embebida tanto en aplicaciones Java de escritorio o servidor. El Reporting Engine maneja tanto el Pentaho Report Designer como el Pentaho BI-Server, que es una aplicación web de Business Intelligence, usada para publicar reportes y otros contenidos vía web

Formatos de salida para exportar Reportes:

Los reportes de Pentaho Reporting pueden ser exportados en:

- PDF
- Excel (97)
- XHTML
- De texto enriquecido (RTF)
- CSV
- XML (vía API)

Características

- Permite la creación de reportes con la herramienta Report Designer.
- Permite agregar un dominio de metadatos fácil de usar con una fuente de datos a través del editor de metadatos.
- Pentaho Web Console, permite tener una fácil gestión de informes y vistas de análisis.
- Maneja conectividad entre las herramientas de cliente y el servidor de informes de BI, lo que permite que el contenido creado con el Editor de Metadatos, Report Designer y Design Studio, pueda ser publicado directamente en el servidor de BI.

1. Pentaho Report Designer:

Este editor se basa en eclipse, y tiene la capacidad de personalizar informes dependiendo de las necesidades del negocio. Incluye asistentes de ayuda para tener una fácil configuración de sus propiedades. El editor de consultas, ayuda de manera rápida, a seleccionar los datos que serán utilizados en los reportes.

Características

- Es un diseñador gráfico (drag & drop) arrastrar y soltar que permite tener un acceso a gráficas, cálculos, datos, agrupaciones para la creación y diseño de reportes de alta resolución.
- Incluye plantillas predeterminadas, para agilizar el proceso de diseño de reportes.
- Los reportes pueden ser ejecutados localmente con el PRD o pueden ser publicados directamente en el servidor Pentaho de tal manera que los usuarios puedan acceder a ellos.
- Cuenta con un editor de consultas que facilita la obtención de los datos que serán utilizados en el reporte.

2. Pentaho Report Design Wizard:

Es una herramienta para el diseño de informes, y permite obtener resultados de manera rápida para:

- Conexión a bases relacionales
- Integrar el resultado dentro del portal Pentaho
- Posibilidad de montar codificación semafórica

3. Web ad-hoc Reporting:

Esta una herramienta web con características similares a Report Design Wizard, y permite a los usuarios crear reportes, haciendo uso de plantillas predeterminadas.

Características Generales²³

- Funcionalidad crítica para usuarios finales:
 - Acceso vía web
 - Informes parametrizados
 - Scheduling
 - Suscripciones
 - Distribución (bursting)
- Ventajas a especialistas en informes:
 - Acceso a fuentes de datos heterogéneos: relacional (vía jdbc), OLAP, XML, transformaciones de Pentaho data Integration.
 - Capacidad de integración en aplicaciones o portales: jsp, portlet, web service.
 - Definición modular de informes (distinción entre presentación y consulta)

²³ **Gravitar. 2012.** Gravitar. [En línea] 2012. [Citado el: 19 de Septiembre de 2012.] <http://www.gravitar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/>.

- Diseño de informes flexible
 - Entorno de diseño gráfico
 - Capacidad de uso de templates
 - Acceso a datos relacionales, OLAP y XML

- Compatible para:
 - Ser embebible
 - Ser fácil de extender
 - No consumir muchos recursos
 - 100% Java: portabilidad, escalabilidad e integración

- Multiplataforma (tanto a nivel de cliente como servidor): mac, linux/unix y Windows

4.2.1.4. Pentaho Analysis

Procesa la información de manera ágil, y apoya la toma de decisiones óptimas.

Características Generales:

- Vista dimensional de datos (por ventas, por periodo, entre otros)
- Navegar y explorar
 - Análisis Ad Hoc
 - Drill-down (permite profundizar en el detalle de los datos)
 - Seleccionar un específico miembro para el análisis
- Alto rendimiento
 - Con tecnología optimizada se obtiene una rápida respuesta interactiva.

4.2.1.5. Pentaho Dashboard

Provee información crítica necesaria para entender y mejorar continuamente el rendimiento organizacional a través de métricas clave en una intuitiva interfaz visual, lo que permite un inmediato rendimiento individual, departamental o empresarial.

De esta manera permite:

- Identificación y monitoreo de métricas (KPI's)
- Drill a reportes (profundiza los detalles)
- Alertas que se basan en las reglas de negocio asignadas

4.2.1.6. Pentaho Data Mining

Incluye herramientas que permiten realizar transformaciones sobre los datos, o tareas de manera que se corre datos en algoritmos complejos, se intensifica patrones y correlaciones escondidas, permite entender de mejor manera el negocio

Se caracteriza por:

- Descubrir patrones ocultos y correlaciones en los datos
- Se previene futuros eventos basados en patrones históricos.
- Rendimiento y escalabilidad.

4.2.1.7. KETTLE - Pentaho Data Integration

Pentaho Data Integration (PDI) es un motor de transformación, diseñado para realizar procesos de integración de datos.

KETTLE - “K Extraction Transformation Transportation Load E”, es una potente herramienta capaz de manipular y transformar información con gran volumen de datos. Realiza procesos de extracción, manipulación, validación y carga de datos desde múltiples fuentes de origen y en diferentes entornos, lo que permite evitar grandes cargas de trabajo manual frecuentemente difícil de mantener y de desplegar.

Hace uso de una interfaz gráfica “SPOON” que permite diseñar “jobs” o tareas de transformación de datos para ser procesados bajo las herramientas de Kettle (PAN y KITCHEN).



Espacio en Blanco Intencional

Arquitectura de Pentaho Data Integration:

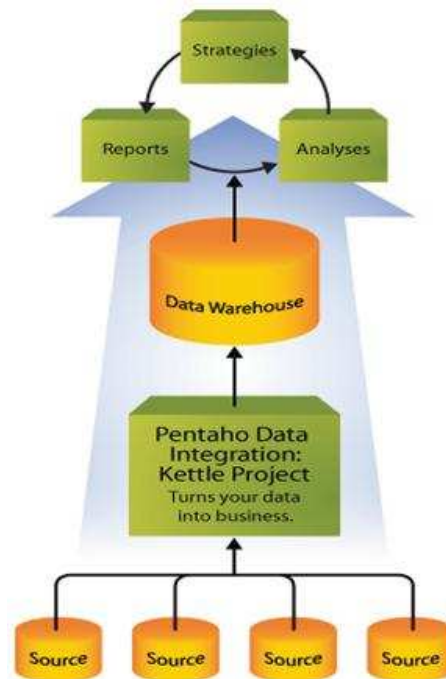


Figura 4.5: Arquitectura de Pentaho Data Integration ²⁴

Propiedades básicas

- Incluye un entorno gráfico de desarrollo
- Maneja tecnologías: Java, XML, JavaScript
- Es independiente del sistema operativo
- Está Basado en dos tipos de objetos: Transformaciones (colección de pasos en un proceso ETL) y trabajos (colección de transformaciones)
- Incluye cuatro herramientas:

²⁴Gravitar. 2012. Gravitar. [En línea] 2012. [Citado el: 19 de Septiembre de 2012.] <http://www.gravitar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/>.

1. **SPOON:** Diseñar transformaciones ETL usando el entorno gráfico
2. **PAN:** Ejecuta transformaciones diseñadas con spoon, es un motor de transformación de datos que permite realizar acciones de la lectura y escritura sobre diversas fuentes de datos
3. **CHEF:** Permite crear trabajos o “Jobs” programados
4. **KITCHEN:** Es una aplicación que permite ejecutar “Jobs” o tareas planificadas en modo “batch” (serie, lote) que corren automáticamente en intervalos de tiempo establecidos.

Muchas organizaciones manejan gran cantidad de datos alojada en distintas aplicaciones y base de datos diferentes, Pentaho Data Integration abre, limpia e integra esta información poniéndola disponible para el usuario, generando consistencia al poder manejar una sola versión de todos los recursos de información

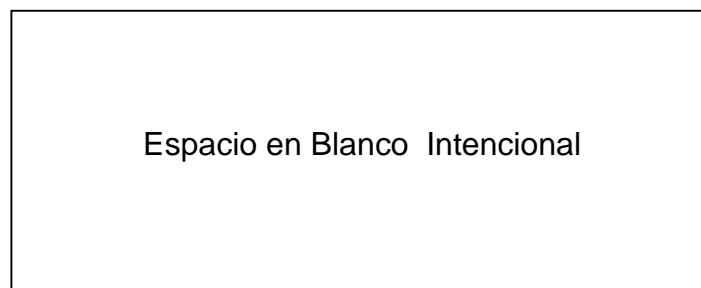
Características

- Proporciona algoritmos que automatizan los procesos de transformación de datos a la forma en que la minería de datos puede explotarlos. Los resultados se pueden visualizar en modo gráfico ya sea agrupado, segmentado, de árbol de decisión, bosque aleatorio, redes neurales y componentes de análisis.
- Utiliza filtros especiales para realizar procesos de discretización, normalización, re-muestreo “re-sampling”, selección y transformación de atributos.
- Kettle es una herramienta de PDI que permite implementar los procesos extracción, transformación y carga de datos (ETL).

- Los datos y sus transformaciones están separados. Las transformaciones y Jobs son almacenadas en archivos de formato XML, donde se detallan las acciones a realizar en los datos. Para construir las transformaciones, se utilizan los componentes, que se enlazan entre sí mediante saltos (steps), para determinar el flujo de datos entre los diferentes componentes.
- Incluye herramientas para realizar consultas, generación de reportes, análisis interactivo, tableros de mando, extracción, transformación y carga de datos (ETL) , integración de datos, minería de datos (data mining), y un servidor para la plataforma de BI.

4.2.1.8. Pentaho BI Platform

La plataforma Pentaho brinda automatización, programación, seguridad e integración, lo que permite a los usuarios finales tener un ambiente para administrar y mantener el despliegue de la empresa, integración procesos de negocio y administración reportes.



4.2.1.9. Comparación entre Pentaho Community Edition y Enterprise Edition

Tabla 4.2: Feature Comparison: Community & Enterprise Editions²⁵

Software and Services	Community Edition	Enterprise Edition
Reporting	Open Source	Certified
Analysis	Open Source	Certified
Dashboards	Open Source	Certified
Data Integration / ETL	Open Source	Certified
Business Intelligence Platform	Open Source	Certified
Data Mining	Open Source	Certified (Add-On)
Community Forums Interaction	✓	✓
Community Web Documentation (wiki)	✓	✓
Professional Support		
• Telephone support (toll-free)		✓
• E-mail support		✓
• Service Level Agreement		✓

• Unlimited support cases		✓
Software Maintenance		
• Software maintenance	By in-house staff	✓ By Pentaho Engineers
• Patch releases		✓
• Fixes included in future releases		✓
Enhanced Functionality		
• Pentaho Enterprise Console		✓
• Single Sign-On		✓
• Streamlined security configuration		✓
• Application diagnostics		✓
• Repository utilities		✓
• Lifecycle management		✓
• Audit reports		✓
• Automated content expiration		✓
• Clustering		✓
• Performance monitoring		✓
• ETL management and monitoring		✓
Certified Software		
• Stabilized software		✓
• Managed release cycle		✓
• Optimized builds		✓

²⁵Corporation, Pentaho. 2008. Pentaho. [En línea] 14 de 9 de 2008. [Citado el: 22 de 09 de 2012.] http://www.osbi.fr/wp-content/pentaho_editions_feature_comparison.pdf.

Product Expertise		
• Professional documentation		✓
• Knowledge base		✓
• Consultative support		✓
• Remote assistance packages		✓
• Installation/configuration packages		✓
• Design and integration packages		✓
• Troubleshooting and optimization packages		✓
• Enterprise Edition online forum		✓
• Web based training		✓
Software Assurance		
• Intellectual Property Indemnification		✓
• Warranty for services		✓

Tabla 4.3: Pentaho Community & Commercial Product Comparison²⁶

	Open Source	Basic	Professional	Enterprise
FUNCTIONALITY				
Enterprise Reporting › For Report Developers › Enterprise Pixel-Perfect Reports › Rich Graphical Designer › Wizard-based Creation	✓	✓	✓	✓
Interactive Reporting › For Business Users › Ad Hoc Operational Reports › Drag & Drop Design	⊘	⊘	✓	✓
Interactive Analysis & Visualization › For Business Analysts › Rich Visualizations › Powerful Analytics › Highly Interactive	⊘	⊘	✓	✓

²⁶ Corporation, Pentaho. 2008. Pentaho. [En línea] 14 de 9 de 2008. [Citado el: 22 de 09 de 2012.] http://www.osbi.fr/wp-content/pentaho_editions_feature_comparison.pdf.

FUNCTIONALITY				
Dashboards > For Executives > KPIs & Performance Metrics > Self-Service Designer	⊘	✓	✓	✓
Data Integration	Basic > Graphical Designer	Enterprise > Graphical Designer > Automated Scheduler > Shared Repository > Team Collaboration > Versioning and Locking	Enterprise > Graphical Designer > Automated Scheduler > Shared Repository > Team Collaboration > Versioning and Locking	Enterprise > Graphical Designer > Automated Scheduler > Shared Repository > Team Collaboration > Versioning and Locking
Secure Data Integration > Integration with LDAP/ Active Directory > User / Role Administration > Content Permissions	⊘	✓	✓	✓
Connectivity	Basic > Relational Databases > NoSQL Sources > Analytic Databases	Expanded > Relational Databases > Enterprise Applications such as Google Analytics, Google Docs, and Java Message Services	Expanded > Relational Databases > Enterprise Applications such as Google Analytics, Google Docs, and Java Message Services	Expanded > Relational Databases > NoSQL Sources > Analytic Databases > Enterprise Applications such as Google Analytics, Google Docs, and Java Message Services

FUNCTIONALITY				
Install, Deploy, Admin > Automated Installers > Centralized Administration > Performance Monitoring > Troubleshooting & Diagnostics > Audit Reporting > Automated Content Expiration > Backup and Recovery > Single Sign-on	⊘	✓	✓	✓
Predictive Analytics	✓	⊘	⊘	✓
Big Data	Big Data Components > Graphical Designer for Big Data	⊘	⊘	Enterprise Grade Big Data > Graphical Designer for Big Data > Automated Scheduler > Shared Repository > Team Collaboration > Versioning and Locking > Security
Mobile	⊘	⊘	✓	✓
SUPPORT & MAINTENANCE				
Certified and Quality Assured Software	⊘	✓	✓	✓
Maintenance Releases	⊘	✓	✓	✓

SUPPORT & MAINTENANCE				
Support	⊘	Standard Support > Business Hours Only > Support Portal Only > Severity 1 Response Time: 4 Business Hours > Severity 2 Response Time: 1 Business Day > Severity 3/4 Response Time: 2 Business Day > Named Support Contacts: 2	Premium Support > 24x7x365 > Telephone & Support Portal > Severity 1 Response Time: 1 Business Hour > Severity 2 Response Time: 2 Business Hours > Severity 3/4 Response Time: 4 Business Hours > Named Support Contacts: 3 > Standard Support Option Available	Premium Support > 24x7x365 > Telephone & Support Portal > Severity 1 Response Time: 1 Business Hour > Severity 2 Response Time: 2 Business Hours > Severity 3/4 Response Time: 4 Business Hours > Named Support Contacts: 3
Maintenance Releases	⊘	✓	✓	✓
Community Online Forums	✓	✓	✓	✓
Customer Online Forums	⊘	✓	✓	✓
PROFESSIONAL SERVICES				
Solutions and Services from Pentaho and Certified Partners	⊘	✓	✓	✓
Free Consulting Checkpoints	⊘	⊘	3 (with Premium Support)	3

4.2.1.9.1. Pentaho Community

Pentaho es una suite BI abierta, que cuenta con un entorno, accesible vía web. Cada usuario tiene la posibilidad de visualizar todos los elementos habilitados para su perfil. Permite generar nuevas vistas de análisis o nuevos reportes y guardarlos asociados al perfil, para consultarlos más tarde, así como exportarlos a Excel, PDF o realizar su impresión.

Es propicio para utilizarlo bajo ambientes de evaluación o pre-producción. Para utilizar esta opción se requieren de recursos internos para gestionar los propios ciclos de lanzamiento, pruebas, apoyo.

4.2.1.9.2. Pentaho Enterprise Edition

Se lo emplea en ambientes de producción ya que, no requiere de recursos internos para el desarrollo y pruebas y garantiza rápidos tiempos de resolución y los acuerdos de nivel de servicio

4.2.2. Data Warehouse y Data Mart

4.2.2.1. Introducción

En la actualidad las empresas han centralizado su información por lo que requieren procesos automatizados y rapidez en el resultado de la obtención de los mismos, por esta razón se requiere innovadoras estrategias de gestión que permitan día a día ser más competitivos, para la toma de decisiones es indispensable contar con mecanismos y herramientas que brinden información sólida y coherente, es ahí de donde se parte para la creación de un Data Warehouse, permitiendo así que todos los individuos de la empresa obtengan datos procesados de manera ágil y sencilla.

El objetivo principal de un Data Warehouse es satisfacer requerimientos de información interna de una empresa, facilitando a los usuarios analistas y encargados de la toma de decisiones, realizar sus tareas de manera más rápida, en el DW las personas acceden a sus datos.

Los datos en el DW se deben mantener actualizados y guardar bases con una perspectiva histórica, permitiendo así consultas operativas e información para llevar a cabo un análisis multidimensional

Un Data Mart es un almacén de datos limitado a un área específica de una organización, tiene la misión de cumplir con requerimientos determinados de un grupo concreto de usuarios, los mismos que se sentirán familiarizados con los términos que serán utilizados para el análisis, comprensión, contenido y facilidad de uso. Se puede generar un data mart obteniendo los datos de un data Warehouse corporativo o también pueden ser creados totalmente independientes, aunque cabe recalcar que no es lo más adecuado realizarlo de manera independiente, puesto que se originan círculos autónomos de información, perdiendo así el concepto de la creación de almacenes de datos, que pretenden precisamente agrupar toda la información.

Un data mart está orientado a realizar consultas específicas de un área determinada en la organización, para esto utiliza distintos procesos batch, mediante herramientas OLAP que ofrecen una visión amplia de la información, sobre esto se puede construir información para los directivos y de esta manera se tome mejores decisiones. Por estas razones se puede decir que un data mart son pequeños data warehouse centrados en un tema o área específica de la organización.

Los principales beneficios de utilizar Datamarts son:

- Acelerar las consultas reduciendo el volumen de datos a recorrer
- Estructurar los datos para su adecuado acceso por una herramienta
- Dividir los datos para imponer estrategias de control de acceso
- Segmentar los datos en diferentes plataformas hardware
- Permite el acceso a los datos por medio de un gran número de herramientas del mercado, logrando independencia de estas.

4.2.2.2. Arquitectura Data Warehouse

La estructura básica incluye:

- **Datos operacionales:** un origen o fuente de datos para poblar el componente de almacenamiento físico DW. El origen de los datos son los sistemas transaccionales internos de la organización como también datos externos a ésta.
- **Extracción de Datos:** selección sistemática de datos operacionales usados para poblar el componente de almacenamiento físico DW.
- **Transformación de datos:** procesos para agregar y realizar otros cambios en los datos operacionales para reunir los objetivos de orientación a temas e integración principalmente.
- **Carga de Datos:** inserción sistemática de datos en el componente de almacenamiento físico DW.
- **Data warehouse:** almacenamiento físico de datos de la arquitectura DW.
- **Herramientas de Acceso al componente de almacenamiento físico DW:** herramientas que proveen acceso a los datos. Estas herramientas pueden ser herramientas específicas de mercado para visualización de bases multidimensionales almacenadas en data warehouses como también aplicaciones desarrolladas dentro de la organización del tipo EIS/DSS.

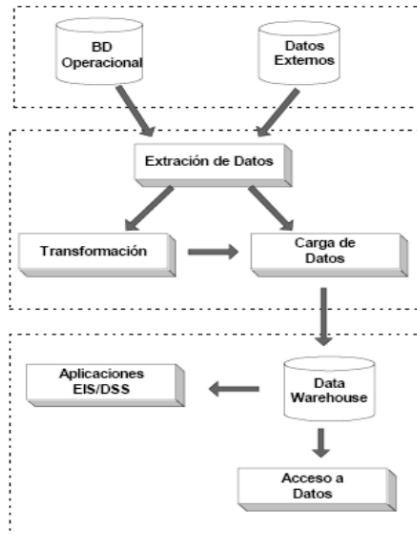


Figura 4.6: Estructura Básica de un Data Warehouse²⁷

4.2.2.3. Componentes Básicos

Cuando se realiza un modelo dimensional de negocio se presentan los siguientes componentes:

- Dimensiones
- Atributos
- Elementos
- Relaciones
- Jerarquías
- Tablas de Hechos o Indicadores

²⁷**Estructura Básica de un Data warehouse;** Estructura Básica de un Data warehouse[En línea] 2012 http://biblioteca.usac.edu.gt/tesis/08/08_0350_CS.pdf

Dimensiones: Son líneas del negocio, las cuales proveen un método general para organizar la información corporativa. Se define como un grupo de uno o más atributos, los cuales son distintos unos de otros, no se pueden encontrar según el esquema de implementación elegido sino que se presenten sólo como un recurso lógico que ayuda a mostrar múltiples perspectivas de los datos permitiendo realizar análisis por diferentes dimensiones o incluso cruzando información entre distintas dimensiones.

Atributos: Los atributos son una agrupación de elementos dentro de una dimensión. El objetivo de los atributos es ver la información de cada dimensión a diferentes niveles de detalle y agrupar los datos para ser analizados.

Elementos: Son las instancias de los atributos que permiten clasificar el rendimiento del negocio. Es importante aclarar que si bien no forman parte indispensable del modelado multidimensional es aconsejable su incorporación para un mayor entendimiento del modelo en etapas tempranas del ciclo de vida.

Relaciones: Son asociaciones lógicas de atributos dentro de una jerarquía, los atributos dentro de una dimensión están directamente relacionados unos con otros a través de los diferentes tipos de relaciones, por otro lado los atributos en diferentes dimensiones están relacionados uno con otros a través de los indicadores o variables del negocio definidas como intersección de las dimensiones.

Jerarquías: Representadas por un ordenamiento lógico dentro de la dimensión, se encuentran formadas por los diferentes tipos de relaciones entre los atributos de una misma dimensión. Pueden existir múltiples jerarquías dentro de una dimensión pero siempre es posible identificar una jerarquía principal o columna vertebral de la dimensión y jerarquías secundarias o descriptivas compuestas por atributos característicos definidos desde la jerarquía principal.

Indicadores o tabla de hechos: Son las variables o métricas que ayudarán a medir la representación del negocio. Existen dos tipos de indicadores: básicos y derivados.

4.2.2.4. Tipos de Modelamiento de un DW

Esquema en estrella

El esquema en estrella, consta de una tabla de hechos central y de varias tablas de dimensiones relacionadas a esta, a través de sus respectivas claves.

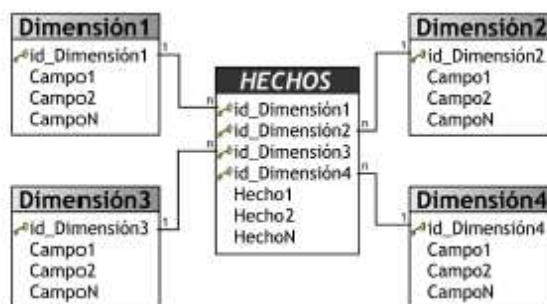


Figura 4.7: Esquema en Estrella²⁸

²⁸ **Esquema en Estrella;** Esquema en estrella[En línea] 2012
http://www.linuxbi.com/files/datawarehouse_hefesto.pdf

Este esquema pretende normalizar de esta manera se pretende eliminar la redundancia, la repetición de datos y que las claves sean independientes de las columnas, pero en este tipo de modelos se requiere no evitar precisamente esto.

Las ventajas de la des normalización, son las de obviar uniones entre las tablas cuando se realizan consultas, procurando así un mejor tiempo de respuesta y una mayor sencillez con respecto a su utilización. El punto en contra, es que se genera un cierto grado de redundancia, pero el ahorro de espacio no es significativo.

El esquema en estrella es el más simple de interpretar y optimiza los tiempos de respuesta ante las consultas de los usuarios.

Las características principales:

- Tiene los mejores tiempos de respuesta
- El diseño es modificable fácilmente
- Existe una relación entre su diseño y la forma en que los usuarios visualizan y manipulan los datos.
- Minimiza el análisis.
- Facilita la interacción con herramientas de consulta y análisis.

Esquema Copo de Nieve

Este esquema representa una extensión del modelo en estrella cuando las tablas de dimensiones se organizan en jerarquías de dimensiones.

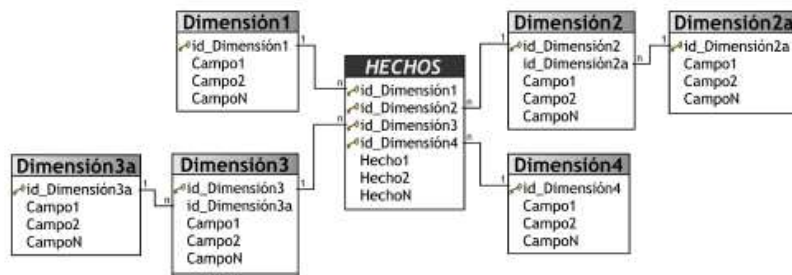


Figura 4.8: Esquema copo de Nieve²⁹

Este modelo es más cercano a un modelo de entidad relación, que al modelo en estrella, puesto que sus tablas de dimensiones están normalizadas.

Este modelo tiene la posibilidad de segregar los datos de las de las tablas de dimensiones y proveer un esquema que sustente los requerimientos de diseño, además es muy flexible y puede implementarse después de que se haya desarrollado un esquema en estrella.

Las características principales de este modelo son:

- Es más complejo en su estructura.
- Utiliza de manera más adecuada el espacio.
- Las tablas de dimensiones están normalizadas, por lo que requiere menos esfuerzo de diseño.
- Desarrolla clases de jerarquías fuera de las tablas de dimensiones, que permiten realizar análisis de lo general a lo detallado y viceversa.

²⁹Esquema Copo de Nieve; Esquema Copo de Nieve[En línea] 2012
http://www.linuxbi.com/files/datawarehouse_hefesto.pdf

Los principales inconvenientes de este esquema:

- Cuan se trabaje con múltiples tablas de dimensiones, cada una de ellas con varias jerarquías, se creará un número de tablas bastante considerable, que pueden llegar al punto de ser inmanejables.
- Al existir muchas uniones y relaciones entre tablas, el desempeño puede verse reducido.

Esquema Constelación

El esquema constelación posee una serie de esquemas en estrella, una tabla principal y por una o más tablas de hechos auxiliares, las cuales pueden ser sumalizaciones de la principal. Dichas tablas yacen en el centro del modelo y están relacionadas con sus respectivas tablas de dimensiones. No es necesario que las diferentes tablas de hechos compartan las mismas tablas de dimensiones, puesto que, las tablas de hechos auxiliares pueden vincularse con solo algunas de las tablas de dimensiones asignadas a la tabla de hechos principal, y también pueden hacerlo con nuevas tablas de dimensiones.

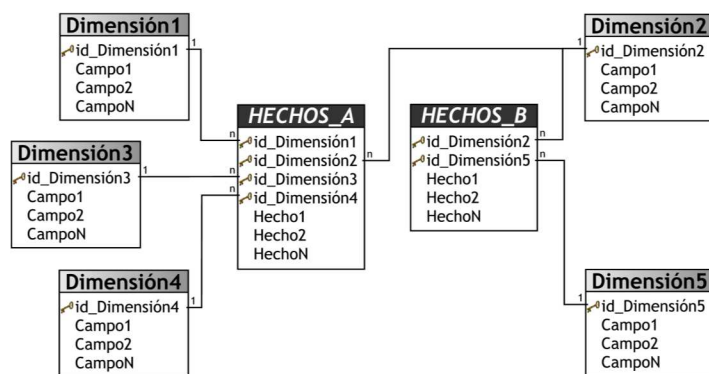


Figura 4.9: Esquema Constelación³⁰

³⁰Esquema Constelación; Esquema Constelación[En línea] 2012
http://www.linuxbi.com/files/datawarehouse_hefesto.pdf

4.2.3. Cubos OLAP

Un Cubo OLAP se denomina como una base de datos que tiene diversas dimensiones, el almacenamiento físico de los datos se lo realiza en un vector multidimensional. Este concepto rebasa a la posibilidad que los usuarios solo dispongan de una base de datos relacional, la cual limitaba el análisis instantáneo de grandes cantidades de datos.

En un sistema OLAP pueden existir más de tres dimensiones; los datos se sumarizan de acuerdo a factores de negocio seleccionados, provee un mecanismo para una respuesta rápida de consultas complejas.

Para la creación de un cubo primero de lo define, seleccionando una tabla objetivo y sus medidas, para luego especificar la estrategia de sumarización diseñando agregaciones.

4.2.4. Estructura Cubo OLAP

Tabla de Hechos:

Se encarga de la captura de los datos que miden las operaciones del equipo, normalmente se componen de un gran número de filas, poseen datos numéricos que sirven para obtener información de la organización de manera resumida, cabe destacar que incluye claves externas, claves primarias de las tablas dimensiones relacionadas, atributos de registros de hechos.

Las tablas de hechos no deberían contener información que sea descriptiva, tampoco datos que no provengan de los campos de medición numéricos y campos de índice que relacionan los hechos con las correspondientes entradas en las tablas de dimensiones.

Medida: Son valores de la tabla de hechos, los mismos que se analizan y agregan.

- Medida regular: es un resumen de la información ya existente, es decir, se toma los datos de una información disponible.
- Medida calculada: realiza un cálculo, obteniendo como resultado un nuevo dato numérico que no se encuentra dentro de la información disponible.

Dimensión: son datos descriptivos de tipo texto, los cuales tienen la funcionalidad de seleccionar datos a un nivel de detalle deseado.

Niveles: son una jerarquía dentro de la dimensión, es decir se establece por la estructura organizacional y modelos de datos que la empresa utiliza. Categoriza desde el resumen hasta el detalle

4.3. Cuadros de Mando

Los cuadros de Mando muestran vistas rápidas del estado de la organización, agrupando indicadores, gráficos, gráficos geoespaciales, tablas, listados de informes y cubos en un sólo documento. Maneja una navegación arrastrar y soltar (drill and down), para llegar a la fuente de los datos.

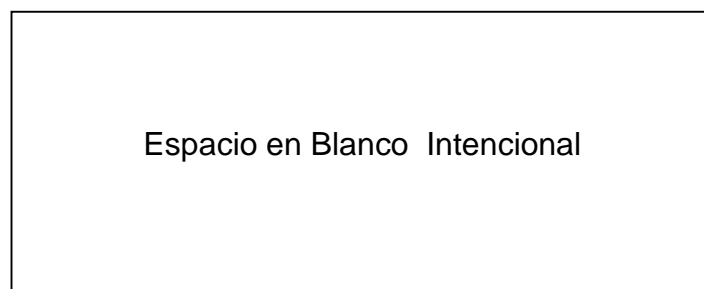
En la construcción de cuadros de mando dinámicos y tableros (Dashboards) para la plataforma BI de Pentaho se hace uso de CDF (Community Dashboard Framework) que es un conjunto de tecnologías Open Source diseñadas para su creación, al ser una aplicación web que utiliza tecnología Ajax permite combinar informes, gráficos, tablas OLAP y mapas de manera dinámica.

En Pentaho se incluye el plugin correspondiente para hacer uso de la herramienta (CDF) tanto en la versión Community como en la Enterprise.

4.3.1. Arquitectura de CDF

Los dashboards CDF son páginas web que contienen áreas llamadas componentes, donde se visualiza contenido BI (informes, gráficos, tablas Olap, etc). Cuando ejecutamos un dashboard en la plataforma BI, se produce las siguientes secuencias de acciones para ejecutarlo en el servidor:

1. El usuario utiliza el navegador en la plataforma BI para abrir un tablero. Esto genera una request HTTP que es enviada al servidor BI de Pentaho.
2. El servidor reconoce una petición de Dashboard e intenta localizar el fichero .xcdf asociado.
3. El fichero .xcdf determina el template o plantilla del dashboard. Es un fichero HTML parcial que contiene los huecos para los componentes y las instrucciones Javascript para llenar estos componentes. El dashboard template se combina con la plantilla del documento (outertemplate) para generar una página web (documento HTML).³¹



³¹Espinosa, Roberto. 2010. [En línea] 20 de Julio de 2010. <http://churriwifi.wordpress.com/category/pentaho>.

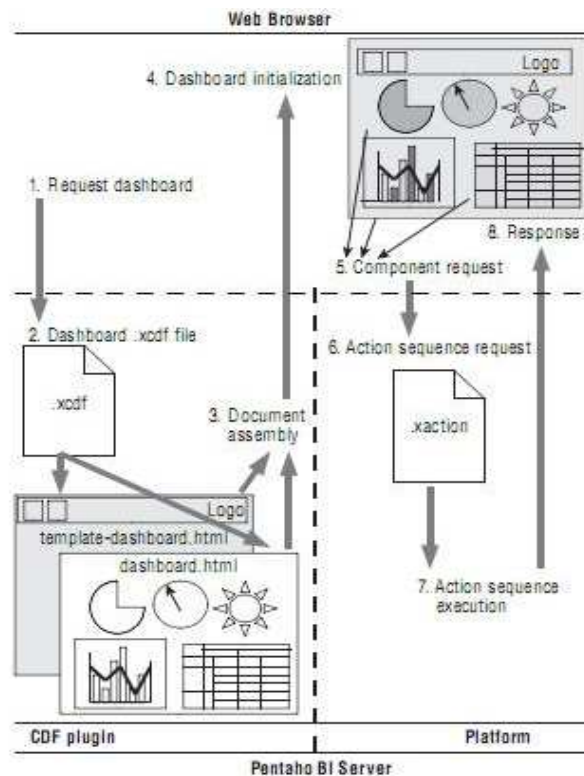


Figura 4.10: Arquitectura CDF³²

4.3.2. Online Analytical Processing - OLAP

Online Analytical Processing (OLAP) .El Procesamiento Analítico en línea permite analizar grandes cantidades de datos en tiempo real. A diferencia de procesamiento de transacciones en línea (OLTP), donde se ejecuta operaciones típicas de leer y modificar individuales y pequeñas cantidades de los registros, mientras que OLAP se ocupa de los datos de forma masiva, y las operaciones son generalmente de sólo lectura. El término 'Online' implica que a pesar de que existan enormes cantidades de datos involucrados y pese a los millones de registros que ocupan gran cantidad de almacenamiento, el sistema debe responder a las consultas realizadas lo suficientemente rápido, y permitir una exploración interactiva de los datos.

³²Espinosa, Roberto. 2010. [En línea] 20 de Julio de 2010. <http://churriwifi.wordpress.com/category/pentaho>.

4.3.2.1. Clasificaciones entre las implementaciones OLAP

- **ROLAP** Se encarga de realizar el almacenamiento de los datos en un motor relacional. Los mismos que serán detallados, para evitar las agregaciones y hace uso de tablas se encontraran normalizadas. Los esquemas de tipo estrella o copo de nieve son los más utilizados, aunque es posible trabajar sobre cualquier base de datos relacional. La arquitectura está compuesta por un servidor de banco de datos relacional y el motor OLAP se encuentra en un servidor dedicado. La principal ventaja de esa arquitectura es que permite el análisis de una enorme cantidad de datos.
- **MOLAP** Permite almacenar los datos en una base de datos multidimensional. Los resúmenes de la información son previamente calculados, lo que permite optimizar los tiempos de respuesta, estos valores pre calculados o agregaciones son la base de las ganancias de desempeño de este sistema. Algunos sistemas utilizan técnicas de compresión de datos para disminuir el espacio de almacenamiento en disco debido a los valores pre calculados.
- **HOLAP (Hybrid OLAP)** almacena algunos datos en un motor relacional y otros en una base de datos multidimensional
- **DOLAP (Desktop OLAP)**. Es un OLAP destinado para el uso en equipos de escritorio. Extrae toda la información que necesita analizar desde la base de datos relacional y la guarda en el escritorio. Desde ese momento, todas las consultas y análisis son hechas contra los datos guardados en el escritorio.

4.3.3. Servidor OLAP Mondrian

Mondrian es un motor OLAP basado en Java que ejecuta consultas escritas en lenguaje MDX. Lee los datos de una **base de datos relacional** (RDBMS), y presenta los resultados en un formato multidimensional en Java API

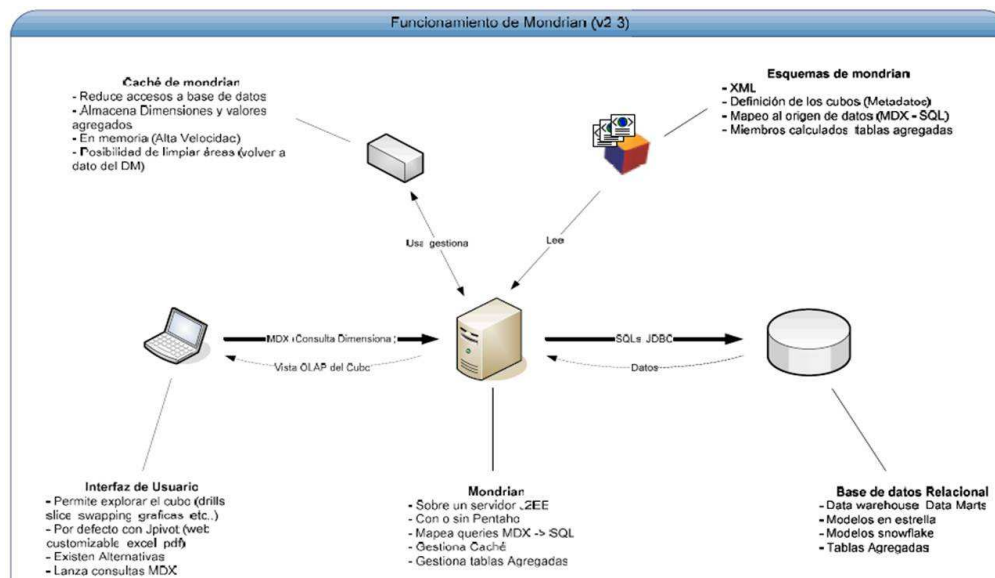


Figura 4.11: Funcionamiento de Mondrian³³

Mondrian es un servidor OLAP open source encargado de gestionar los proceso de comunicación entre una aplicación OLAP (escrita en Java) y la base de datos con los datos fuente, es decir, Mondrian actúa como "JDBC para OLAP".

En la arquitectura de Mondrian se ejecuta sobre un servidor web y permite tener una comunicación entre aplicaciones OLAP con las bases de datos. El núcleo del servidor Mondrian es similar a JDBC pero exclusivo para OLAP. Proporciona la conexión a la base de datos y ejecuta las sentencias SQL.

³³ Salinas, Alexandro. 2008. [En línea] 12 de Marzo de 2008. [Citado el: 22 de Septiembre de 2012.] <http://www.gravitar.biz/index.php/bi/introduccion-pentaho-parte-1/>.

Mondrian es un servidor OLAP escrito en Java. Permite analizar grandes cantidades de datos almacenados en bases de datos SQL de una forma interactiva sin necesidad de escribir las sentencias que serían necesarias para ello en SQL. Entre otras características destaca la compatibilidad con el MDX (expresiones multidimensionales) y el lenguaje de consulta XML(XMLA) para análisis y olap4j.

Kettle. Es una herramienta de Pentaho Data Integration que permite implementar los procesos de extracción, transformación y carga de datos. El uso de kettle permite evitar grandes cargas de trabajo manual frecuentemente difícil de mantener y de desplegar.

4.4. Cubos de Información

En la plataforma Open Source de PENTAHO se incluye la herramienta OLAP Mondrian, que permite crear **cubos de información** para análisis multidimensional. Los cubos de información están compuestos de archivos XML en los que se encuentran definidas las dimensiones y las conexiones de los datos. Los archivos XML por lo general son complejos para desarrollarlos manualmente, por lo que es conveniente utilizar herramientas gráficas para realizar la edición de estos, entre algunas herramientas Open Source Pentaho se tiene a Cube Designer que permite crear cubos y utiliza el Workbench para la edición de los mismos. Son subconjuntos de datos de un almacén de datos, organizado y agrupado dentro de una estructura multidimensional. Los datos se agregan de acuerdo a factores de negocio seleccionados, proveyendo el mecanismo para la rápida y uniforme tiempo de respuesta de las complejas consultas. Estos datos organizados en Cubos, son la materia prima que el DataWareHouse almacena.

4.5. Esquema Workbench de Mondrian

El esquema Workbench es utilizado para definir esquemas (con cubos OLAP) que pueden ser utilizados en las Vistas de Análisis

El Esquema Workbench de Mondrian es un diseñador de interfaz que permite crear y probar cubos OLAP Mondrian visualmente. El motor Mondrian procesa las solicitudes de MDX con los esquemas ROLAP (OLAP Relacional). Estos archivos de esquema XML de metadatos son modelos que se crean en una estructura específica utilizada por el motor de Mondrian. Estos modelos XML pueden ser considerados de forma de cubo estructuras que utilizan tablas de HECHO y tablas de dimensiones se encuentran en el RDBMS. No se requiere que de un cubo físico real; sólo que el modelo de metadatos se cree.

Funcionalidades:

- Editor de esquemas integrados con la fuente de datos subyacente para la validación.
- Prueba MDX consultas contra la base de datos y los esquemas

CAPÍTULO 5

DESARROLLO DE LA SOLUCIÓN

5.1. Definición del Proyecto

Para desarrollar el proyecto se determinó la necesidad de implementar una solución BI, por parte de los ingenieros de Pruebas del área de la Gerencia de Construcción, en la vicepresidencia de Tecnología de la Empresa Telefónica Ecuador, con el fin de facilitar el control y la toma de decisiones por parte de los coordinadores y jefes principalmente, que requieren mantener monitoreadas las tareas que realizan los proveedores para verificar el funcionamiento de los diferentes cambios o soluciones que requiere la empresa como tal y los clientes, de esta manera se genera independencia al momento de extraer la información y no se necesitará de procesos manuales u operacionales que sólo retrasan las tareas diarias de los ingenieros.

5.2. Justificación, Objetivos y Alcance

La definición de la justificación, objetivos y alcance se encuentran detallados en el capítulo I en este documento, mismo que enfatiza en la solución del problema que ya se definió.

La solución se focaliza en la elaboración de indicadores, que permitan bajo el esquema de reportes visualizar: tiempos, cantidades, responsables, lo cual facilitará tener tiempos de respuesta menores para la toma de decisiones y planes de acción en el caso que se requieran, de esta manera se reduce la carga operativa de los ingenieros de pruebas, quienes actualmente deben sacar estos datos manualmente solicitando reportes al administrador de base de datos.

5.3. Identificación de Roles

- Patrocinador de Negocio: Telefónica Ecuador
- Administrador de Bases de Datos: Ingeniero Senior Área de Pruebas Gerencia de Construcción.
- Analista del negocio en la Gerencia de Construcción: Jefe de sección de desarrollo tecnológico de sistemas administrativos.
- Personal involucrado en el negocio: ingenieros del área de pruebas.
- Los roles de: desarrollador, modelador de datos, diseñador, líder del proyecto estarán asignados a los egresados encargados de la solución.

Tabla 5.1. Ingenieros - Roles

Ingeniero	Nombre del Puesto	Puesto que depende Jerárquicamente	Función
Efrén Zambrano	Jefe Área de Pruebas	Gerencia de Construcción	Jefe y Coordinador Proyectos
Roberto Zumba	Ingeniero de Pruebas	Jefe de Pruebas	DBA
Eduardo Loaiza	Ingeniero de Pruebas	Jefe de Pruebas	Coordinador Proyectos
Lourdes Huilcatoma	Ingeniero de Pruebas	Jefe de Pruebas	Coordinador Proyectos Express

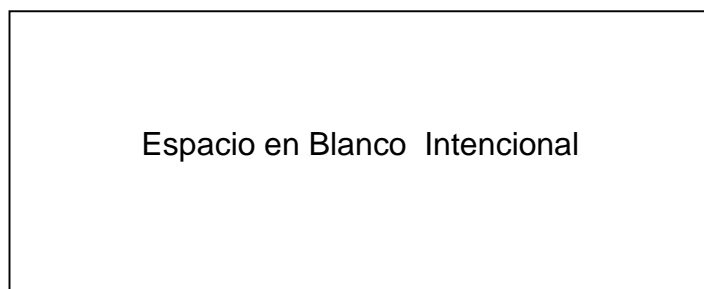
5.4. Estrategia y método de Implementación

Se realizará la implementación de la Solución de Business Intelligence, para la generación de indicadores y el control de desempeño, en el área de pruebas de la Gerencia de Construcción de la empresa OTECEL S.A.

La implementación de una Solución BI con la herramienta Pentaho Community permitirá el desarrollo de dos reportes prediseñados en el Pentaho Report Designer, para el área de pruebas de OTECEL S.A., que tiene la capacidad de personalización de informes de acuerdo a las necesidades de negocio utilizando Oracle, como motor de base de datos, del departamento de Gerencia de Construcción del área de pruebas.

Se implementará un cubo, para poder realizar las distintas combinaciones, y otorgar la información requerida por el usuario.

Se utilizará la metodología Hefesto V2.0 , durante la implementación de la solución BI, que es una metodología, propia para el trabajo con soluciones bajo la plataforma de Pentaho, la idea de utilizar Hefesto es que no se requieran fases extensas de reunión de requerimientos y análisis.



5.5. Situación Actual

Actualmente el proceso de control por parte de los Ingenieros de pruebas, se lo realiza manualmente, de tal manera que para obtener un reporte de los avances de pruebas en cualquier proyecto, lo realizan de manera manual, lo cual disminuye el tiempo que podrían utilizar en tareas de análisis que beneficien al área y a la empresa.



Figura 5.1: Estructura Actual de Proceso de Información en el Área de Pruebas

5.6. Origen de la fuente de Datos

Se trabajará con los datos obtenidos de la base de datos Tareas del área de pruebas del departamento de Gerencia de Construcción de la empresa Telefónica Movistar Ecuador, que se encuentra físicamente en Telefónica Colombia.

5.7. Análisis de Requerimientos

Se realizaron reuniones con el personal encargado de los procesos de pruebas, para determinar las necesidades y requerimientos para poder definir qué información necesitan obtener en tiempo real una vez realizado el desarrollo.

5.7.1. Entrevistas

Se realizaron entrevistas con los ingenieros involucrados directamente en el proceso de control de sus proveedores para determinar, que información requieren revisar constantemente, de tal manera que facilite sus actividades diarias y no lo realicen de manera manual u operativa.

Los requerimientos fueron analizados con:

- El jefe de Área de Pruebas de Tecnología.
- Los ingenieros de pruebas involucrados en el proceso.
- Los administradores de las bases de datos.

Espacio en Blanco Intencional

5.7.2. Requerimientos

Con las entrevistas realizadas se entendió el proceso de control, y se definió los requerimientos de acuerdo a lo realizado, de la siguiente manera:

Tabla 5.2. Análisis de Requerimientos

Nombre Reporte	Título Reporte	Descripción	Filtros			
Control Requerimientos	Requerimientos por Ambiente	Permite obtener la cantidad de requerimientos realizados de manera: - Mensual - Anual Agrupado por el Tipo de Ambiente en donde se realiza las prueba: - Producción - Preproducción - Proyectos	Año	Mes	Ambiente	
		Permite obtener la cantidad de requerimientos realizados de manera: - Mensual - Anual Agrupado por el Tipo de Ambiente en donde se realiza las prueba: - Producción - Preproducción - Proyectos	Año	Mes	Semana	Ambiente
Control Horario Tarea	Total Horas Trabajo – Ambiente	Permite obtener un resumen mensual de la cantidad de horas requeridas para realizar pruebas: - Normal - Extraordinaria Agrupado por el Tipo de Ambiente: - Producción - Preproducción - Proyectos	Año	Mes	Tipo Hora	

Nombre Reporte	Título Reporte	Descripción	Filtros		
Control Req Ambiente Reprueba	Número Requerimiento – Ambiente	1. Permite obtener un resumen mensual de los requerimientos que se han realizado pruebas por cada ambiente: - Producción - Preproducción - Proyectos	Año	Mes	
		2. Obtiene un resumen mensual del número de repruebas que se tuvo que realizar por ambiente.	Año	Mes	Ambiente
		3. Resumen mensual del número de horas requeridas para realizar pruebas por el tipo de ambiente.	Año	Mes	Ambiente
Control Tipo Pruebas por Requerimiento	Tipo de Pruebas por Requerimiento	1. Permite obtener el tipo de pruebas realizadas mensualmente por la clasificación de requerimientos	Año	Mes	Requerimiento
		2. Permite obtener el número de requerimientos realizados por el tipo de prueba: - PE -> Pruebas Especiales - PI -> Pruebas Internas - EP -> Estrategias Especiales - NA -> Pruebas SVC - PF -> Pruebas Formales	Año	Mes	Tipo Prueba
Responsables Tarea	Control Tareas Responsables	Control Tareas Responsables	Año	Mes	Requerimiento

Nombre Reporte	Título Reporte	Descripción	Filtros			
Control Tipo Prueba - Ambiente	Control Tipo Prueba - Ambiente	Permite obtener un resumen semanal del número de tipos de pruebas realizados en cada ambiente	Año	Mes	Semana	Ambiente
		Permite obtener un resumen mensual del número de tipos de pruebas realizados en cada ambiente	Año	Mes	Ambiente	
Tiempo Total Horas Ambiente	Total Horas de Ejecución - Requerimientos	Permite obtener un resumen mensual de horas requeridas para realizar pruebas por ambiente: - Producción - Preproducción - Proyectos	Año	Mes		
Tiempo Transacción	Número de Requerimientos	Permite obtener el número de requerimientos realizados mensualmente	Año	Mes		

5.7.3. Identificar Indicadores Y Perspectivas De Análisis

5.7.3.1. Indicadores

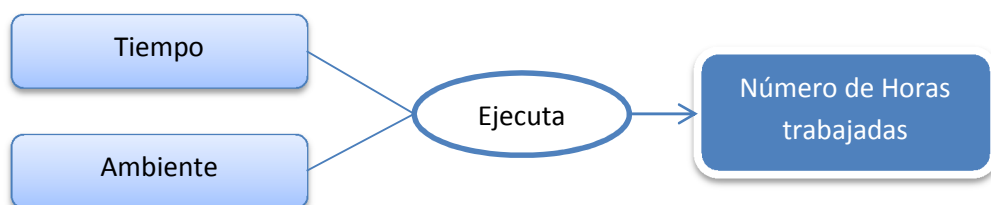
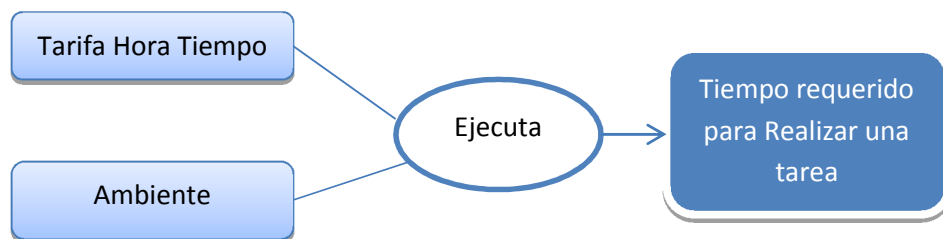
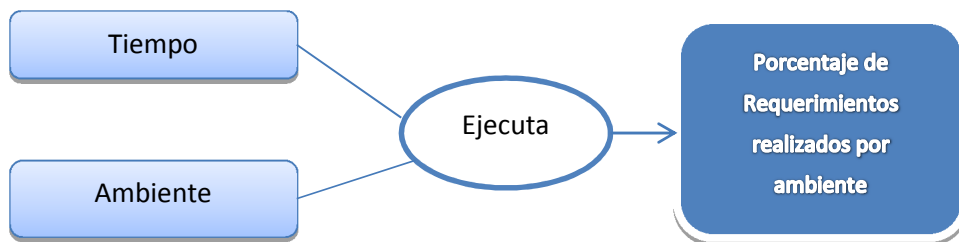
- Número de Requerimientos realizados.
- Cantidad de Tareas asignadas a cada Directivo.
- Número de horas trabajadas en horario normal y extraordinario.
- Porcentaje de requerimientos que existen por en Producción, Preproducción y Proyectos.
- Número de Pruebas realizadas en Producción, Preproducción y Proyectos.
- Numero de requerimientos realizados por ambiente.
- Número de Repuebas realizadas por ambiente.
- Sumatoria del tiempo requerido para realizar la prueba.

5.7.3.2. Perspectivas

- Año
- Mes
- Ambiente
 - Producción
 - Preproducción
 - Proyectos

- Requerimiento
- Responsable Tarea
- Hora
 - Ordinaria
 - Extraordinaria
- Tarifa Hora

MODELO CONCEPTUAL



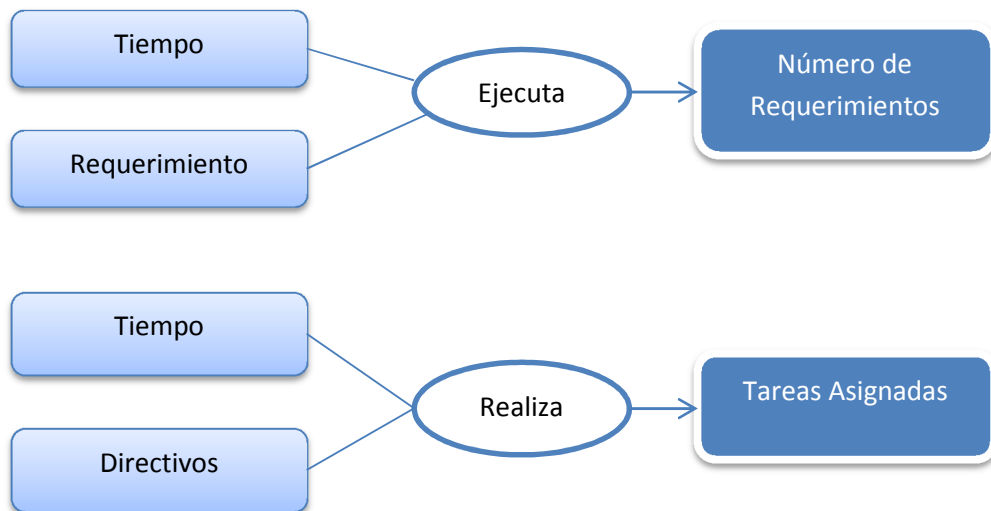


Figura 5.2: Diagrama de Perspectivas - Tesistas

5.7.4. Diseño de la Arquitectura

Con los OLTP, que contienen las fuentes de datos, se realizara los procesos ETL, para ingresar los datos del DataMart. De los cuales se obtendrán los datos necesarios para la generación de reportes, para generar los indicadores requeridos por el área de pruebas, para su respectivo análisis



Figura 5.3: Proceso Solución: Tesistas

La base de datos de origen será la de registro de tareas en Oracle, de la cual se exportarán los datos en un archivo Excel, en la misma actualmente se almacena todos los datos con los que operacionalmente los ingenieros de pruebas realizan los cálculos de control y verifican estadísticamente el avance de los proyectos que se encuentra en fase de prueba.

Para realizar los procesos OLTP, se utilizará la herramienta Pentaho Data Integration, la cual generará los reportes y cálculo de los indicadores que se requieren.

5.7.5. Ambiente Estación de Trabajo

En el gráfico se mostrará el ambiente en la estación de trabajo donde se implementará el proyecto.

Los datos se obtendrán de la base de datos del registro de tareas, el cálculo de los indicadores y la generación de los reportes se realizarán a través de Pentaho User Console, la cual permitirá realizar los procesos OLTP.

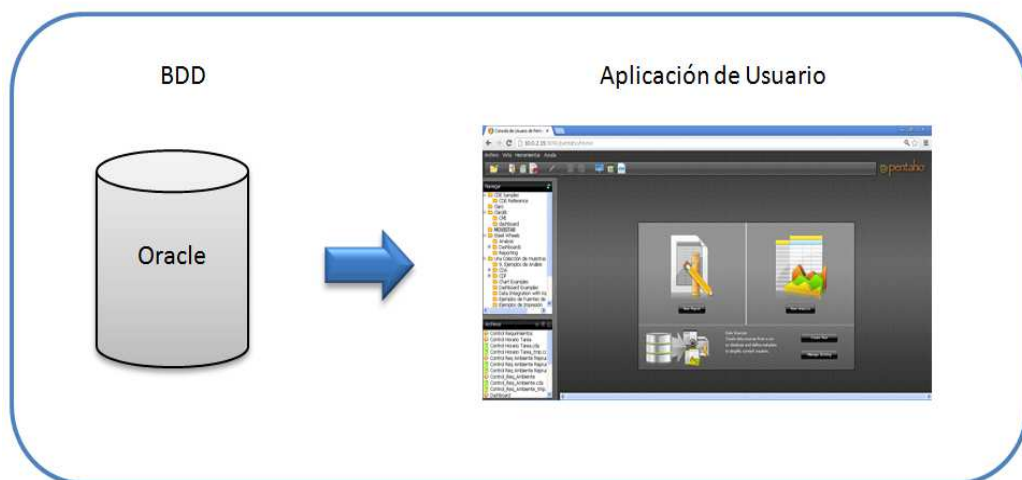


Figura 5.4: Ambiente Estación de Trabajo - Tesistas

5.8. Análisis de los OLTP

5.8.1. Conformación de Indicadores



Figura 5.5: Indicadores Generales: Tesistas

Ambiente

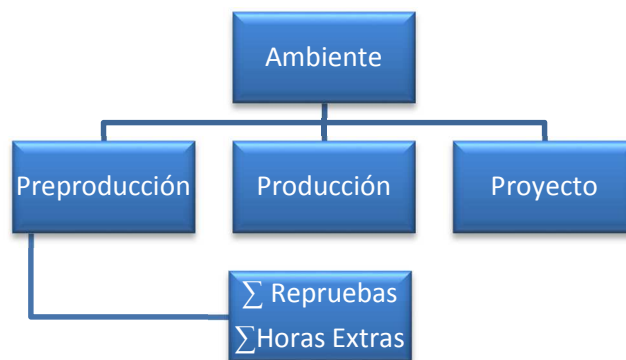


Figura 5.6: Indicador Ambiente: Tesistas

Tareas

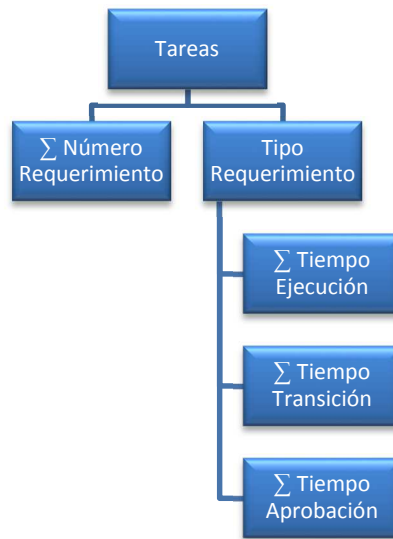


Figura 5.7: Indicador Tareas: Tesistas

Responsables



Figura 5.8: Indicador Responsables: Tesistas

5.8.2. Establecer Correspondencias con los requerimientos

Lo que se pretende específicamente para lograr una consolidación de los datos, es definir previamente lo que se necesita calcular, en este caso los indicadores son el foco en el negocio, para lo cual se requiere reportes continuos para el control y toma de decisiones.



Figura 5.9: Proceso Análisis Solución: Tesistas

5.8.3. Modelo Dimensional

Estándar Modelo Dimensional

El estándar que se utilizara para dar nombre a las dimensiones creadas y para el caso de creación de nuevas Dimensiones será:

- ✓ **Dim:** Para representar una Dimensión.
- ✓ **Nombre Tabla:** Representa el nombre de la tabla a la que hacer referencia la dimensión.

Formato: Dim Nombre de la tabla a la que hace referencia

Lista de Dimensiones

- ✓ DimAmbiente
- ✓ DimRequerimientos
- ✓ DimDirectivos
- ✓ DimTiempoTransaccion
- ✓ DimTiempoEjecuta

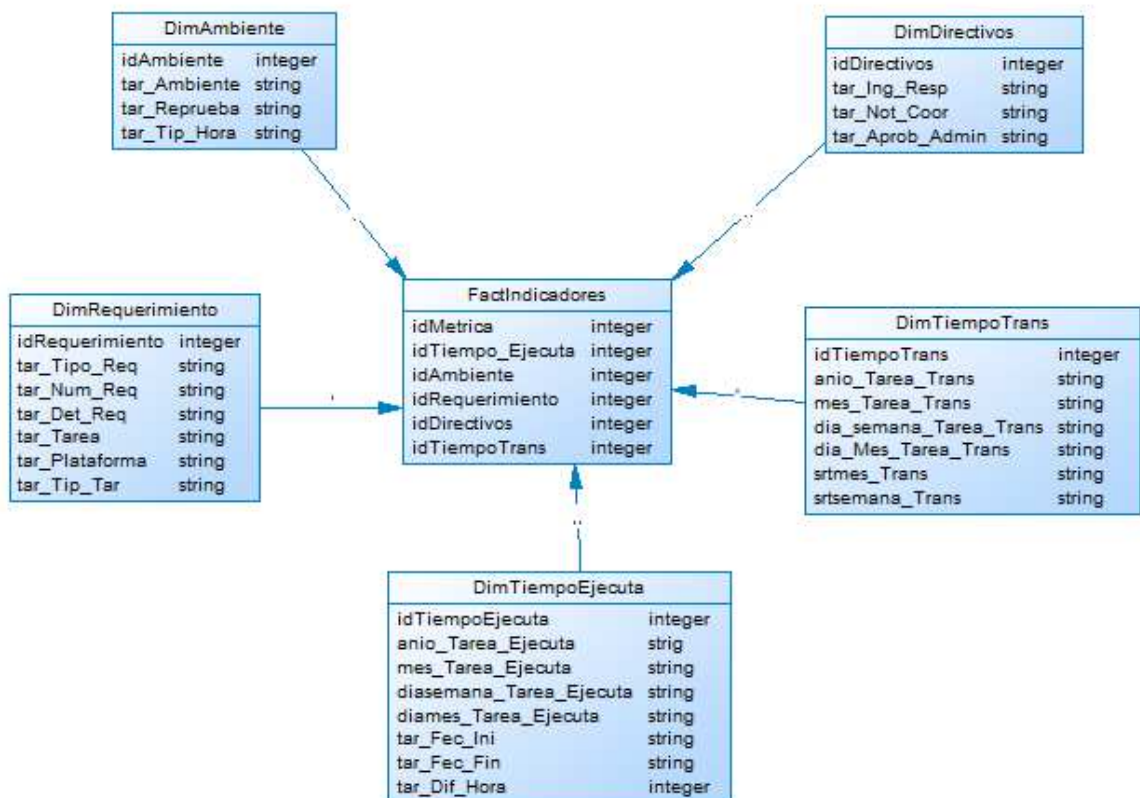


Figura 5.10: Modelo Dimensional: Tesis

Tabla 5.2. Descripción de Dimensión Tiempo ejecuta

Tabla	DimTiempoEjecuta								
Tipo de Tabla	Dimensional								
Descripción	Tabla de registros de la ejecución de pruebas								
Destino							Origen		
Columna	Descripción	Tipo de Dato	Tamaño	Clave	FK para	Null	Tabla	Campo	Tipo de Dato
idTiempoEjecuta	Código de la acción	integer	256	pk	Metrica	No	-	-	-
anio_tarea_eje	Año realización prueba	varchar	200	n			tareas	TAR_FEC_TAREA_EJEC	Date
mes_tarea_eje	Mes realización prueba (Valor numérico)	integer	256	n			tareas	TAR_FEC_TAREA_EJEC	Date
diaSemana_tarea_eje	Semana realización prueba (Valor numérico)	integer	256	n			tareas	TAR_FEC_TAREA_EJEC	Date
diaMes_tarea_eje	Día realización prueba	integer	256	n			tareas	TAR_FEC_TAREA_EJEC	Date
srtMesEje	Nombre de mes de realización de la prueba	varchar	200	n			-	Proceso transformación Java Script	-
srtSemanaEje	Nombre de semana de realización de la prueba	varchar	200	n			-	Proceso transformación Java Script	-
tar_dif_hora	Tiempo requerido para ejecutar una prueba	varchar	200	n				TAR_DIF_HOR	String
tar_fec_ini	Fecha de inicio de la ejecución de la prueba	varchar	200	n			tareas	TAR_FEC_INI	Date
tar_fec_fin	Fecha de fin de la ejecución de la prueba	varchar	200	n			tareas	TAR_FEC_FIN	date

Tabla 5.3. Descripción de Dimensión Requerimientos

Tabla	DimRequerimiento								
Tipo de Tabla	Dimensional								
Descripción	Tabla de registros del tipo de requerimientos realizados en pruebas								
Destino							Origen		
Columna	Descripción	Tipo de Dato	Tamaño	Clave	FK para	Null	Tabla	Campo	Tipo de Dato
idRequerimiento	Código de la acción	integer	256	pk	Metrica	No	-	-	-
tar_tipo_req	Clasificación del tipo de requerimiento para realizar pruebas	varchar	2000	n			requerimiento	TAR_TIPO_REQ	varchar
tar_num_req	Número de requerimiento asociado al proyecto	varchar	200	n			requerimiento	TAR_FEC_TAREA_EJEC	varchar
tar_det_req	Detalle del requerimiento para realizar pruebas	varchar	200	n			requerimiento	TAR_DET_REQ	varchar
tar_tarea	Descripción de la tarea ejecutada	varchar	200	n			requerimiento	TAR_TAREA	varchar
tar_plataforma	Nombre de la plataforma donde se realiza la prueba	varchar	200	n			requerimiento	TAR_PLATAFORMA	varchar
tar_tip_tar	Nombre del tipo de pruebas realizadas	varchar	200	n			requerimiento	TAR_TIP_TAR	varchar

Tabla 5.4. Descripción de Dimensión Directivos

Tabla	DimDirectivos								
Tipo de Tabla	Dimensional								
Descripción	Tabla de registro de coordinadores y responsables de pruebas								
Destino							Origen		
Columna	Descripción	Tipo de Dato	Tamaño	Clave	FK para	Null	Tabla	Campo	Tipo de Dato
idDirectivos	Código de la acción	integer	256	pk	Metrica	No	-	-	-
tar_ing_res	Ingeniero coordinador responsable de pruebas	varchar	2000	n			requerimiento	TAR_TIPO_REQ	varchar
tar_num_req	Número de requerimiento asociado al proyecto	varchar	200	n			requerimiento	TAR_NUM_REQ	varchar
tar_nota_coord	Detalle de las notas del coordinador responsable	varchar	200	n			requerimiento	TAR_NOTA_COORD	varchar
tar_aprob_admin	Descripción de la tarea ejecutada	varchar	200	n			requerimiento	TAR_APROB_ADMIN	varchar

Tabla 5.5. Descripción de Dimensión Tiempo Transacción

Tabla		DimTiempoTrans							
Tipo de Tabla		Dimensional							
Descripción		Tabla de registros de la ejecución de pruebas							
Destino							Origen		
Columna	Descripción	Tipo de Dato	Tamaño	Clave	FK para	Null	Tabla	Campo	Tipo de Dato
idTiempoTrans	Código de la acción	integer	256	pk	Metrica	No	-	-	-
anio_tarea_trans	Año realización prueba	varchar	200	n			tareas	TAR_FEC_TRAN	date
mes_tarea_trans	Mes transición prueba (Valor numérico)	integer	256	n			tareas	TAR_FEC_TRAN	date
diaSemana_tarea_trans	Semana transición prueba (Valor numérico)	integer	256	n			tareas	TAR_FEC_TRAN	date
día_Mes_tarea_trans	Día transición prueba	integer	256	n			tareas	TAR_FEC_TRAN	date
srtMes_trans	Nombre de mes de realización de la prueba	varchar	200	n			tareas	Proceso transformación Java Script	-
srtSemana_trans	Nombre de mes de realización de la prueba	varchar	200				tareas	Proceso transformación Java Script	

Tabla 5.6. Descripción de Dimensión Ambiente

Tabla	DimAmbiente								
Tipo de Tabla	Dimensional								
Descripción	Tabla de registros de los ambiente en donde se realizan las pruebas								
Destino							Origen		
Columna	Descripción	Tipo de Dato	Tamaño	Clave	FK para	Null	Tabla	Campo	Tipo de Dato
tar_ambiente	Ambiente en el que se realizó la prueba	varchar	200	n			ambiente	TAR_AMBIENTE	varchar
tar_reprueba	Comprobar si requirió de una nueva prueba	varchar	200	n			ambiente	TAR_REPRUEBA	varchar
tar_tip_hora	Horario en el que se realizó la prueba , en horario normal o en horas suplementarias	varchar	200	n			ambiente	TAR_TIP_HORA	varchar
idAmbiente	Código de la acción	integer	256	pk	Métrica	no			

Tabla 5.7. Descripción Métrica

Tabla	Metrica					
Tipo de Tabla	Dimensional					
Descripción	Tabla de hechos , conformada por las claves primarias de las tablas de dimensión					
Destino						
Columna	Descripción	Tipo de Dato	Tamaño	Clave	FK para	Null
idMetrica	Código de la acción	integer	256	pk		no
Metrica	Campo Contador para las Dimensiones	integer	256			no
idTiempoEjecuta	FK - tabla DimTiempoEjecuta	integer	256	fk		no
idTiempoTrans	FK - tabla DimTiempoTrans	integer	256	fk		no
idAmbiente	FK - tabla DimAmbiente	integer	256	fk		no
idRequerimiento	FK - tabla DimRequerimiento	integer		fk		no
idDirectivos	FK - tabla DimDirectivos	integer	256	fk		no

Base de Datos Origen

Tabla 5.8. Descripción de Tabla Requerimiento - BD Origen

Tabla	Requerimiento					
Tipo de Tabla	BD Origen					
Descripción	Tabla de registro de requerimiento para realizar pruebas					
Descripción de Campos						
Campo	Descripción	Tipo de Dato	Tamaño	Clave	Null	
idRequerimiento	Código Requerimiento	integer	256	pk	no	
tar_tipo_req	Clasificación del tipo de requerimiento para realizar pruebas	varchar	2000	n		
tar_num_req	Número de requerimiento asociado al proyecto	varchar	200	n		
tar_det_req	Detalle del requerimiento para realizar pruebas	varchar	200	n		
tar_tarea	Descripción de la tarea ejecutada	varchar	200	n		
tar_plataforma	Nombre de la plataforma donde se realiza la prueba	varchar	200	n		
tar_tip_tar	Nombre del tipo de pruebas realizadas	varchar	200	n		

Tabla 5.9. Descripción de Tabla Ambiente - BD Origen

Tabla	Ambiente				
Tipo de Tabla	BD Origen				
Descripción	Tabla de registros de los ambiente en donde se realizan las pruebas				
Descripción de Campos					
Campo	Descripción	Tipo de Dato	Tamaño	Clave	Null
idAmbiente	Código Ambiente	integer	256	pk	No
tar_ambiente	Ambiente en el que se realizó la prueba	varchar	2000	n	
tar_reprueba	Comprobar si requirió de una nueva prueba	varchar	200	n	
tar_tip_hora	Horario en el que se realizó la prueba , en horario normal o en horas suplementarias	varchar	200	n	
idAmbiente	Código de la acción	varchar	200	n	

Tabla 5.10. Descripción de Tabla Directivos - BD Origen

Tabla	Directivos				
Tipo de Tabla	BD Origen				
Descripción	Tabla de registro de coordinadores y responsables de pruebas				
Descripción de Campos					
Campo	Descripción	Tipo de Dato	Tamaño	Clave	Null
idDirectivos	Código de la acción	integer	256	pk	No
tar_ing_res	Ingeniero coordinador responsable de pruebas	varchar	2000	n	
tar_num_req	Número de requerimiento asociado al proyecto	varchar	200	n	
tar_nota_coord	Detalle de las notas del coordinador responsable	varchar	200	n	
tar_aprob_admin	Descripción de la tarea ejecutada	varchar	200	n	

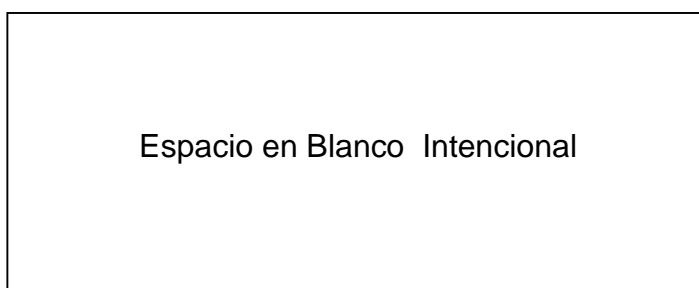
Tabla 5.11. Descripción de Tabla Tareas - BD Origen

Tabla	Tareas				
Tipo de Tabla	BD Origen				
Descripción	Tabla de registros de la ejecución y transición de pruebas				
Descripción de Campos					
Campo	Descripción	Tipo de Dato	Tamaño	Clave	Null
idTarea	Código de la acción	integer	256	pk	no
tar_fec_tarea_ejec	Fecha de la llegada a la cancha de pruebas	date		n	
tar_fec_ini	Fecha de inicio de ejecución de prueba	date		n	
tar_fec_fin	Fecha de finalización de prueba	date		n	
tar_fec_tran	Fecha de llegada de realización de prueba	date		n	
tar_dif_hora	Tiempo requerido para ejecutar una prueba	varchar	200	n	

5.8.4. Nivel de granularidad

De acuerdo a las correspondencias establecidas, se analizaron los campos que serán utilizados para el desarrollo de los reportes de los cuales se determinaron las perspectivas y los indicadores.

Primero se examinó las fuentes de datos entregadas para el análisis, para definir los significados de cada campo, y luego se consultó con el personal de sistemas que maneja los datos, para comprender y definir los campos que permitirán obtener los indicadores deseados.



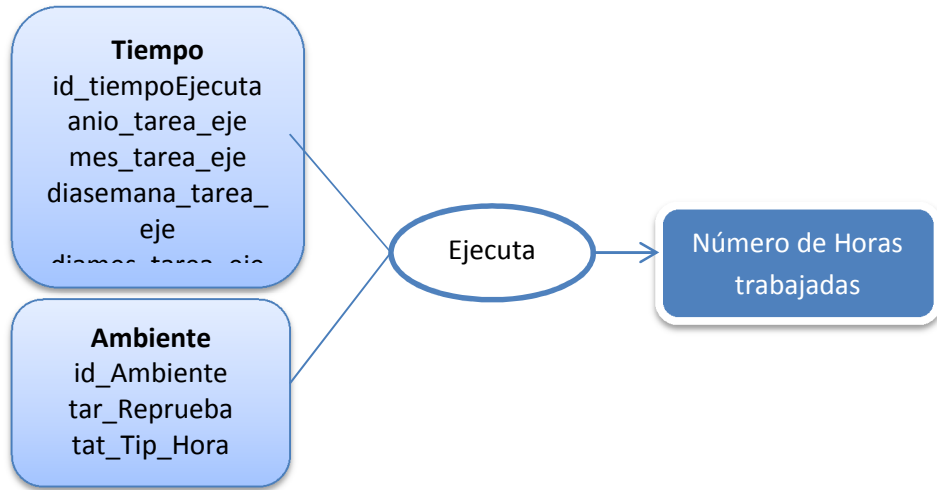


Figura 5.11: Análisis Indicador Número de Horas Trabajadas: Tesistas

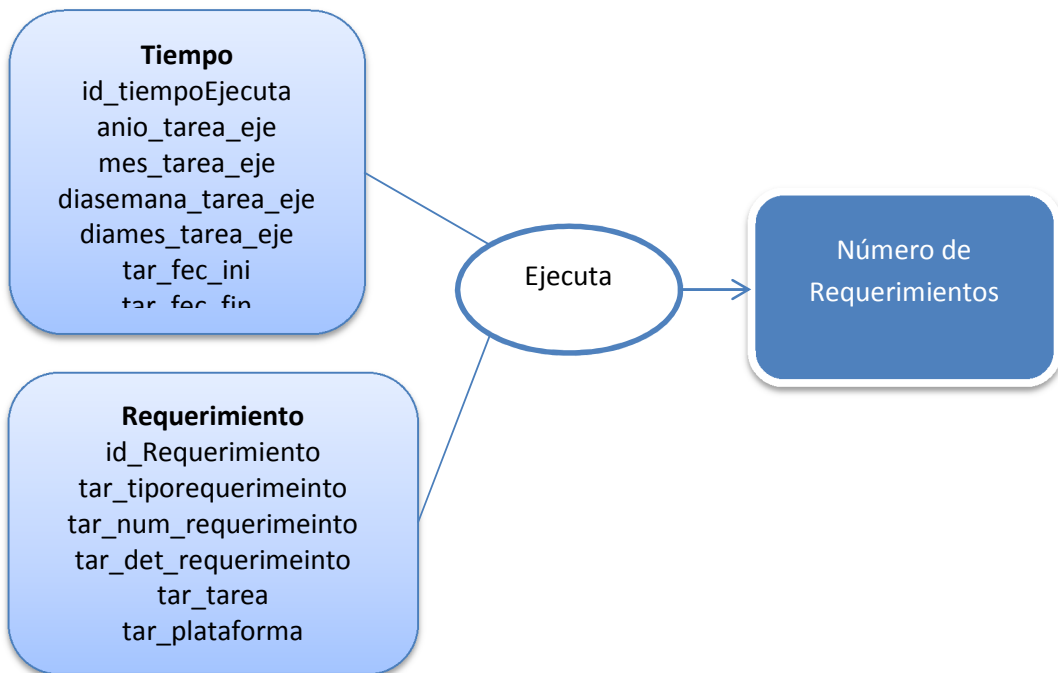


Figura 5.12: Análisis Número de Requerimientos: Tesistas

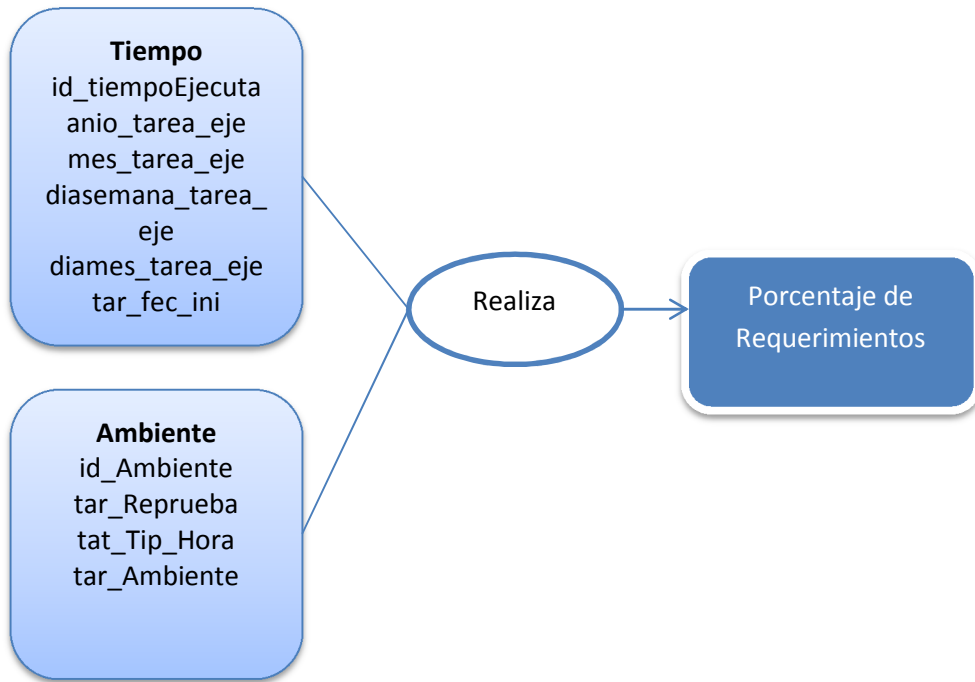


Figura 5.13: Análisis Porcentaje de Requerimientos: Tesistas

5.8.5. Modelo Conceptual Ampliado

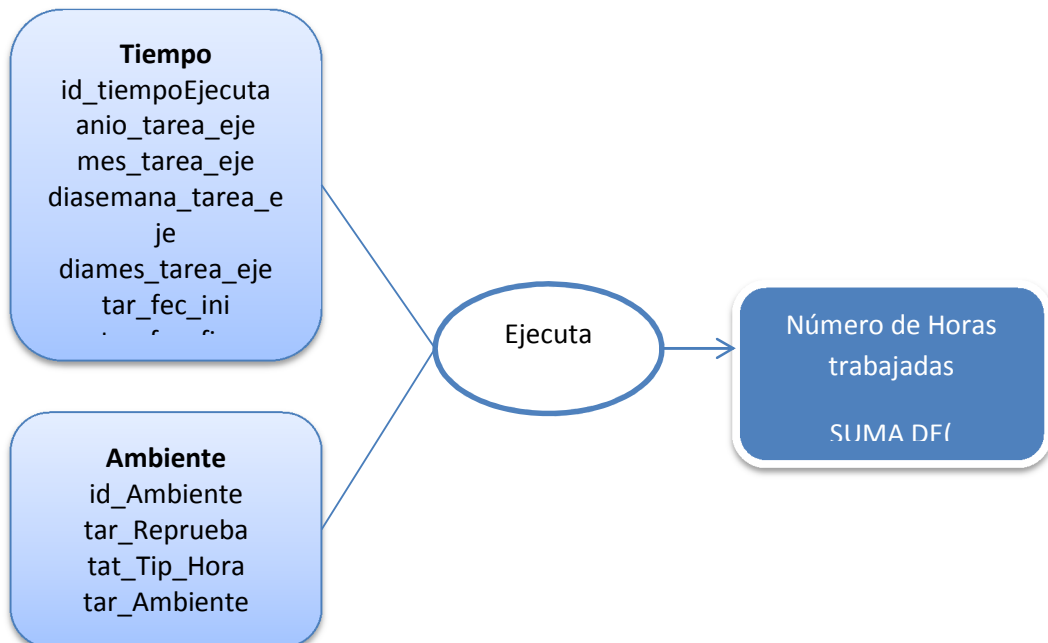


Figura 5.14: Análisis Número Horas Trabajadas: Tesistas

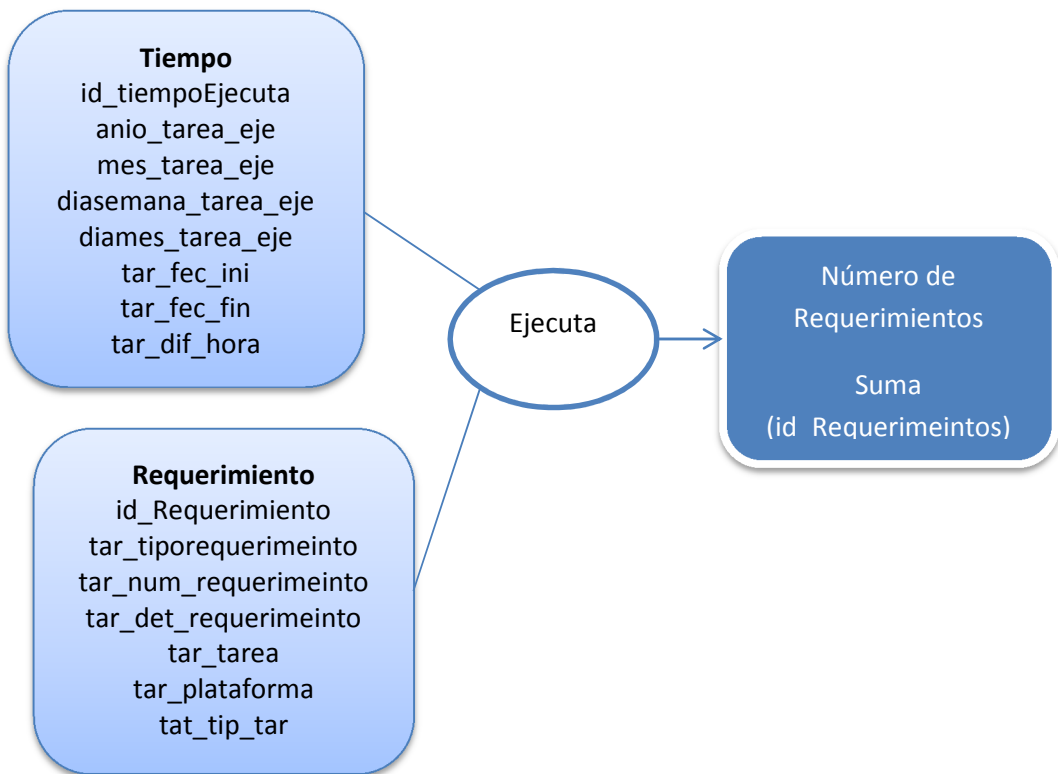


Figura 5.15: Análisis Número Requerimientos: Tesistas

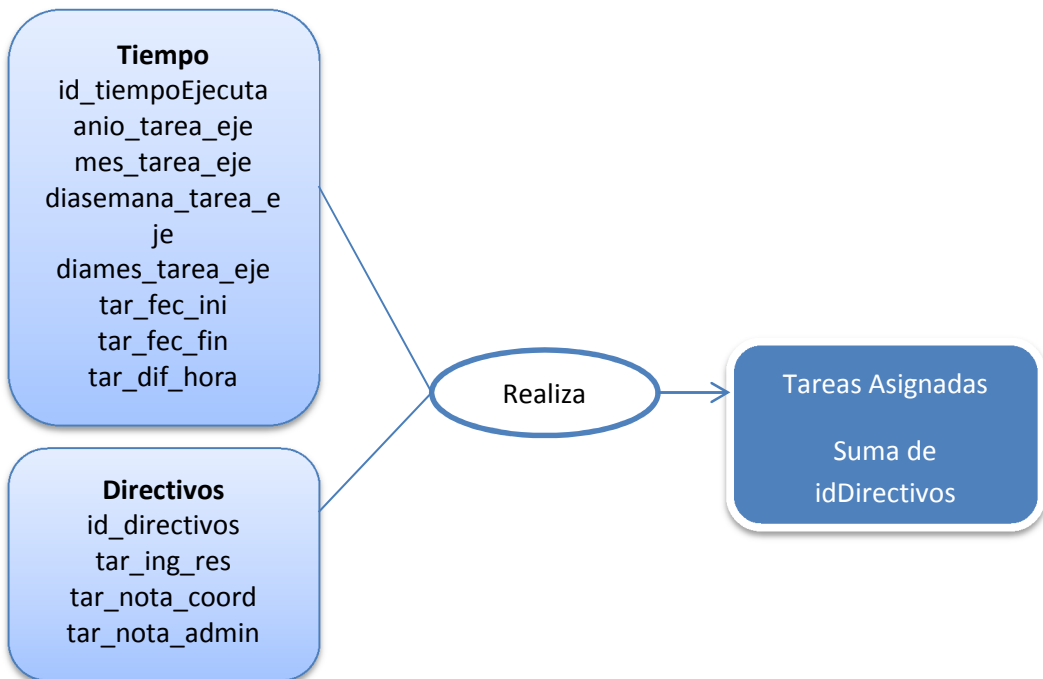


Figura 5.16: Análisis Tareas Asignadas: Tesistas

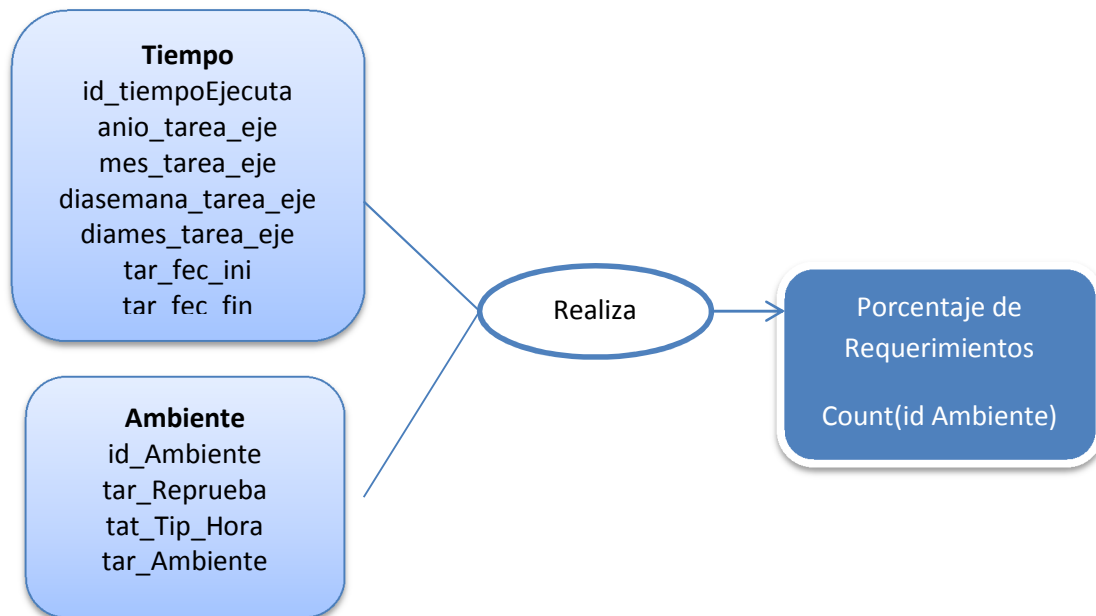


Figura 5.17: Análisis Porcentaje de Requerimientos: Tesistas

5.9. Selección e Instalación de Herramientas

5.9.1. Selección de Herramientas

Para mantener la integridad de la información se ha seleccionado Oracle 10i en donde estarán los datos que se requiere, correspondientes al registro de tareas.

Pentaho Community Edition es la herramienta libre que con la cual se implementará la solución para el cálculo de indicadores y generación de reportes para el control de tareas.

Cada una de las características, funcionalidades y herramientas están descritas en el Capítulo IV.

5.9.2. Instalación de las Herramientas

La instalación de la base de datos Oracle ya se realizó con anterioridad en la empresa Telefónica Movistar Ecuador

De la página <http://community.pentaho.com/> se descargó:

- Pentaho BI server platform community edition 4.2.1 stable
- Pentaho Data Integration (Kettle)

Instalación de Pentaho BI Community Edition:

- Instalación jdk

Configurar las variables de entorno: classpath y path

La herramienta Pentaho Community Edition, no posee un asistente de instalación como tal, por el contrario contiene dentro de sus carpetas un ejecutable .bat que permite iniciar su servicio, la ubicación será en el directorio o carpeta donde se descomprima los componentes.

- Se creó una carpeta en el disco local C, en la cual se descomprimen todas las herramientas descargadas



Espacio en Blanco Intencional

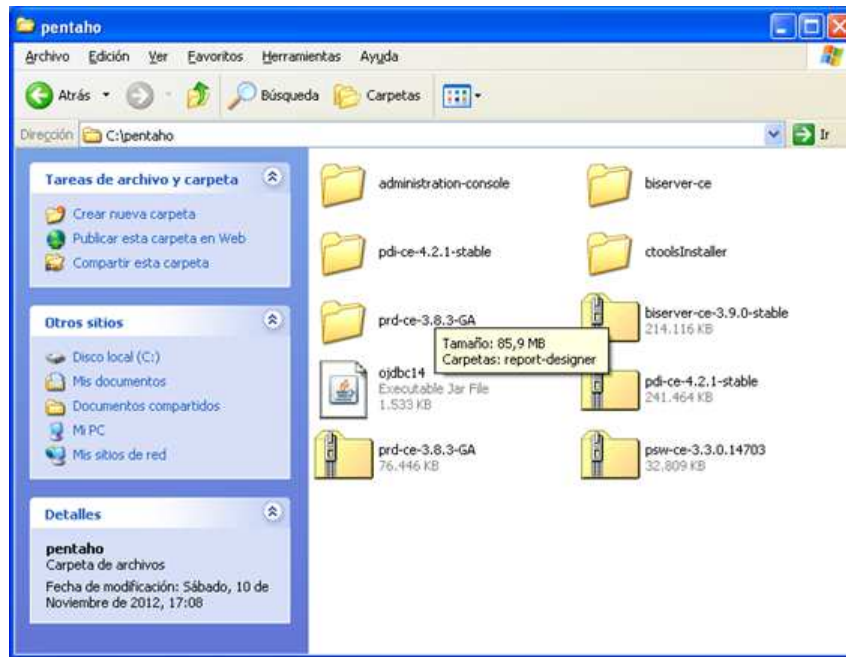


Figura 5.18: Carpeta Utilitarios Pentaho

- En la Carpeta **C:\pentaho\pdi-ce-4.2.1-stable\data-integration** se encuentra la herramienta Spoon con la cual se realizará los proceso de transformación ETL
- Se debe copiar el driver en la dirección **C:\pentaho\pdi-ce-4.2.1-stable\data-integration\lib** para la conexión entre Pentaho y la base de Datos Oracle, el driver se descargó de la página oficial de Oracle <http://www.oracle.com/us/products/index.html>
- **Pentaho BI server:** dentro de la carpeta biserver-ce ubicada en la dirección **C:\pentaho\biserver-ce** ejecutar el archivo start-pentaho.bat, este archivo levanta el servicio con el servidor Tomcat.
- **Pentaho Data Integration (kettle):** dentro de la carpeta descomprimida Pentaho Data Integration ejecutar el archivo **Spoon.bat** que se encuentra en la dirección **C:\pentaho\pdi-ce-4.2.1-stable\data-integration** que iniciará la aplicación para generar gráficamente los procesos ETL.



Figura 5.19: Pentaho Data Integration

- Se debe configurar la conexión para poder realizar la los procesos de transformación en la herramienta en donde se encontraran los datos procesados

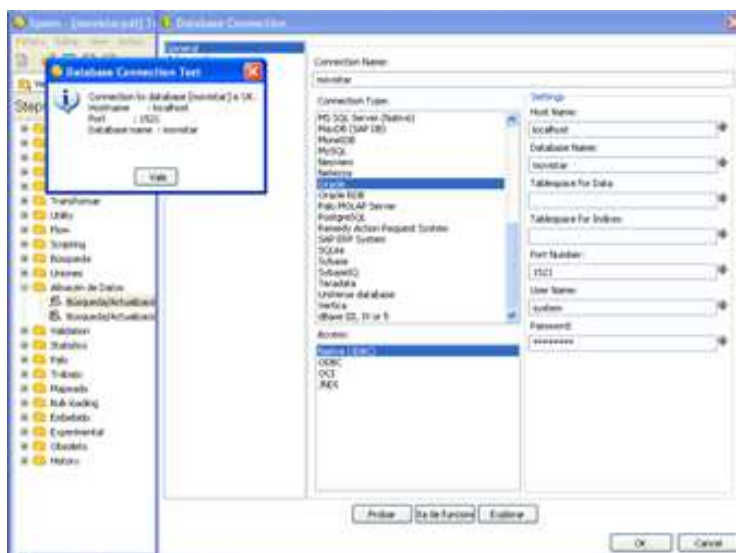


Figura 5.20: Configuración conexión Base de Datos

- Administration Console: dentro de la carpeta administration-console ubicada en **C:\pentaho\administration-console**; ejecutar el archivo **start-pac.bat** que permite ingresar directamente a la administración de Pentaho para realizar la conexión con la base de datos correspondiente.

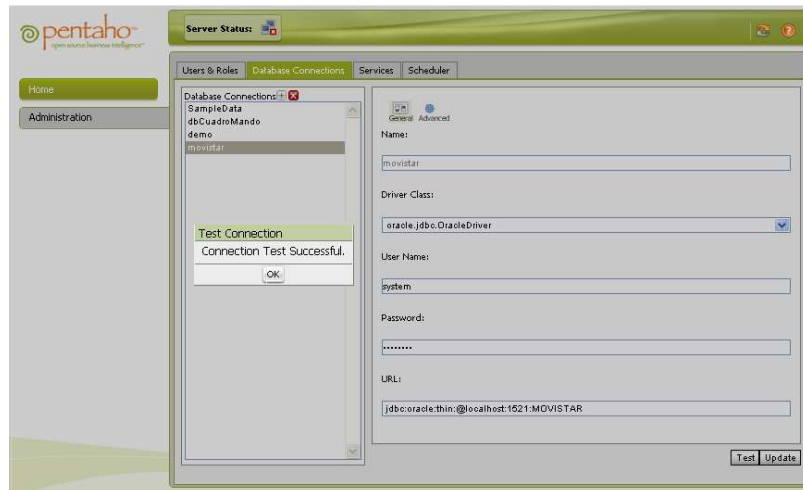


Figura 5.21: Configuración Base de Datos

- Para iniciar el desarrollo de los reporte y el diseño del cubo se debe ejecutar **start-pentaho** ubicado en la carpeta **C:\pentaho\biserver-ce** , una vez levantado , podemos iniciar con la herramienta , ingresando la dirección **http://10.0.2.15:9090/pentaho/Home**

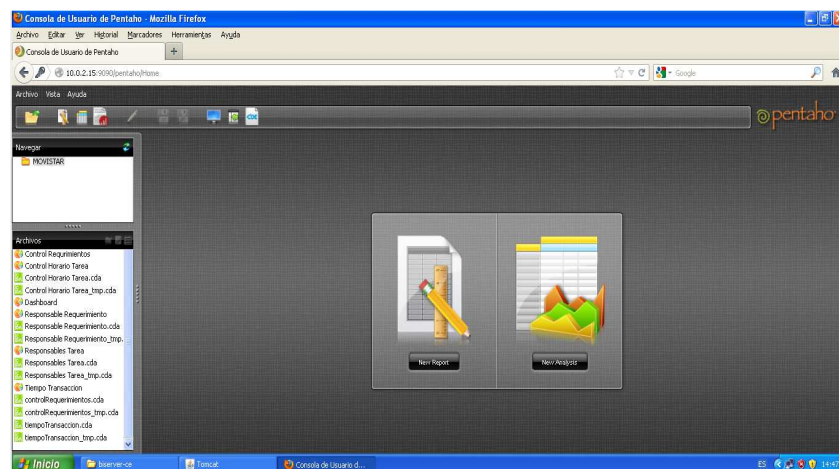


Figura 5.22: Interfaz de Usuario Pentaho

5.10. INTEGRACIÓN DE DATOS

Estándar de ETL

Para la creación de ETL el estándar que se llevara para asignar el nombre a los mismos es el siguiente:

- ✓ **Transformación:** Sera el nombre que se dará, para indicar que se trata de un proceso de ETL o transformación.

- ✓ **Nombre Proceso:** Ira el nombre del proceso al cual se aplicara los las distintas transformaciones.

ETL

- ✓ TransformacionTarea

Lista de Procesos

Los tipos de entrada que de los cuales se hará la carga de información serán archivos CSV, debido a la cantidad de data que se debe realizar la carga. El nombre que se dará a los archivos de entrada será:

- **Fuentes de Entrada:**



Las fuentes de entrada, son herramientas que permiten indicar que tipo de fuente contendrán los datos a ser transformados, en este caso se utilizara archivos CSV.

- ✓ **CSV:** Son archivos de texto plano que contiene los datos fuente extraídos de la Base de datos, en los cuales se utilizara como separador de columnas el carácter “,”.

- ✓ **Nombre Proceso:** El nombre del proceso al cual están asociados los datos contenidos en el archivo.

Formato: CSV Nombre_Proceso

- **Valor Java Script Modificado:**



Sirve para agregar validación o transformación a los campos,

el nombre que se asignara a los objetos de este tipo será:

- ✓ **JS_:** Representa un objeto Java Script.
- ✓ **Nombre Validador:** Representa el nombre del tipo de transformación que se está Realizando

Formato: JS_ Nombre Validador

- **Calculadora:**



Esta herramienta, se utiliza para realizar transformaciones

de campos, pudiendo hacer uso de las funciones propias

de esta herramienta que ayudan a realizar una

discriminación en los campos o llegar a un nivel de detalle más específico, en

el caso que se requiera.

- ✓ **C_:** Indica que se trata de la herramienta Calculadora.
- ✓ **Nombre Cálculos:** Indica que realizara se realizara con la herramienta Operaciones.

Formato: C_ Nombre Cálculos

- **Almacén de Datos:**



a

Esta herramienta, se utiliza para realizar la carga de los datos, de acuerdo a los campos a los que se les asociaran cada Dimensión, los cuales se insertaran en la Base de Datos que alojara, los resultados obtenidos del Proceso de transformación

- ✓ **Dim:** Es el indicador, para definir que se está realizando la carga de los datos a una Dimensión.
- ✓ **Nombre Dimensión:** Indica el Nombre de la dimensión a la cual están asociados los datos cargados.

Formato: DimNombre Dimensión

5.10.1. Carga Inicial

Se realiza la carga inicial del archivo OLTP, utilizando técnicas de limpieza y calidad de datos, procesos ETL, con la herramienta **Pentaho**

5.10.2. Data Integration

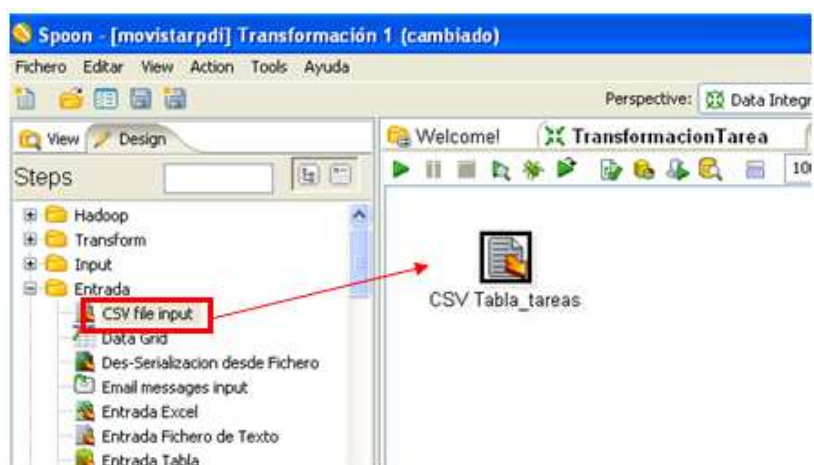


Figura 5.23: Archivo CSV

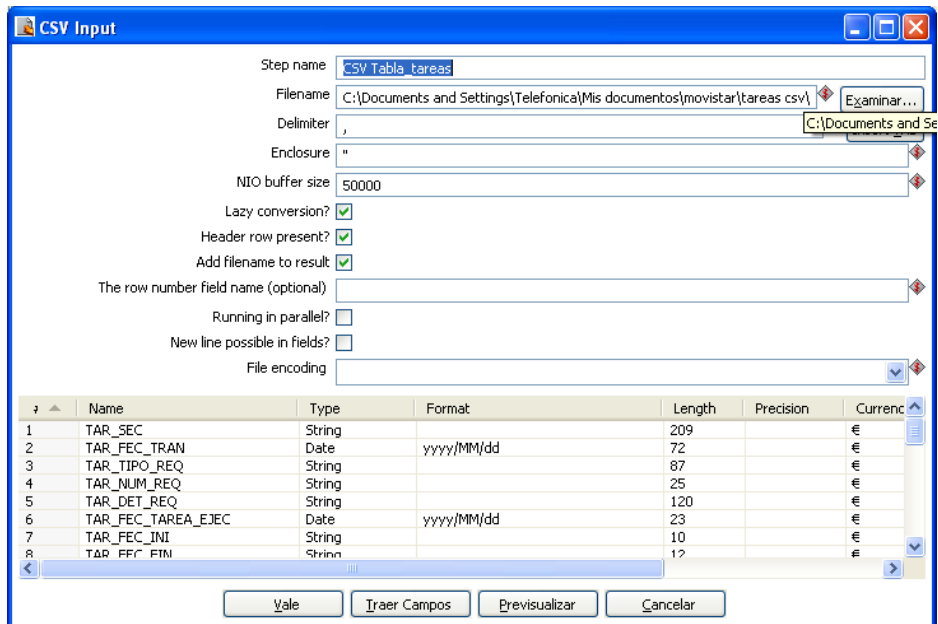


Figura 5.24: Selección Archivo

Con la fuente de datos cargada, se procede a realizar las operaciones necesarias para generar los datos calculados

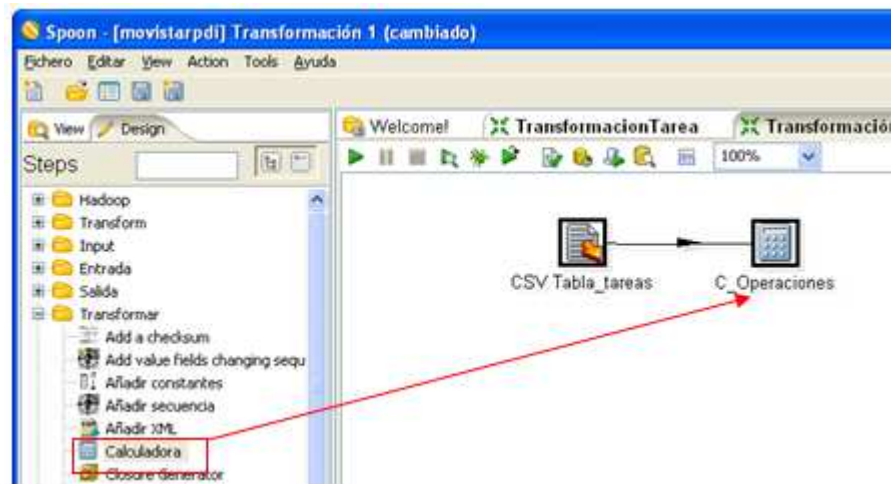


Figura 5.25: Carpeta Utilitarios Pentaho

Calculadora								
Nombre paso: C_Operaciones								
Campos:								
	Nuevo campo	Cálculo	Campo A	Campo B	Campo C	Tipo de valor	Longitud	Precisi
1	Métrica	Valor constante A	1			Number		
2	año_tarea_eje	Año de fecha A	TAR_FEC_TAREA_EJEC			String		
3	mes_tarea_eje	Mes de fecha A	TAR_FEC_TAREA_EJEC			Integer		
4	diaSemana_tarea_eje	Día de la semana de fecha A	TAR_FEC_TAREA_EJEC			Integer		
5	año_tarea_trans	Año de fecha A	TAR_FEC_TRAN			String		
6	mes_tarea_trans	Mes de fecha A	TAR_FEC_TRAN			Integer		
7	diaSemana_tarea_trans	Día de la semana de fecha A	TAR_FEC_TRAN			Integer		
8	diaMes_tarea_eje	Día del mes de fecha A	TAR_FEC_TAREA_EJEC			Integer		
9	diaMes_tarea_trans	Día del mes de fecha A	TAR_FEC_TRAN			Integer		

Figura 5.26: Cabeceras Archivo

Se realizó scripts con java en los cuales se parametrizó los datos de mes, año y semana, para que se muestren como texto de acuerdo al dato numérico que se encuentra en los OLTP

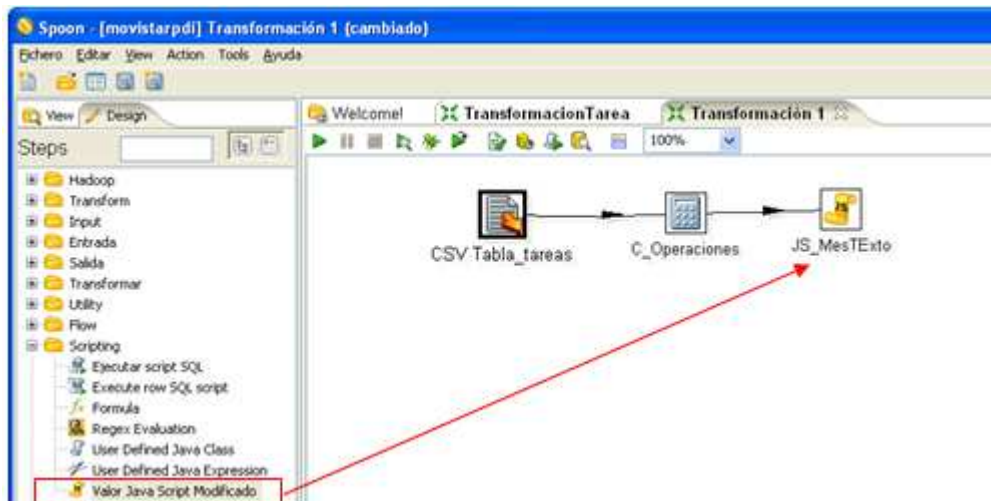


Figura 5.27: Java Script

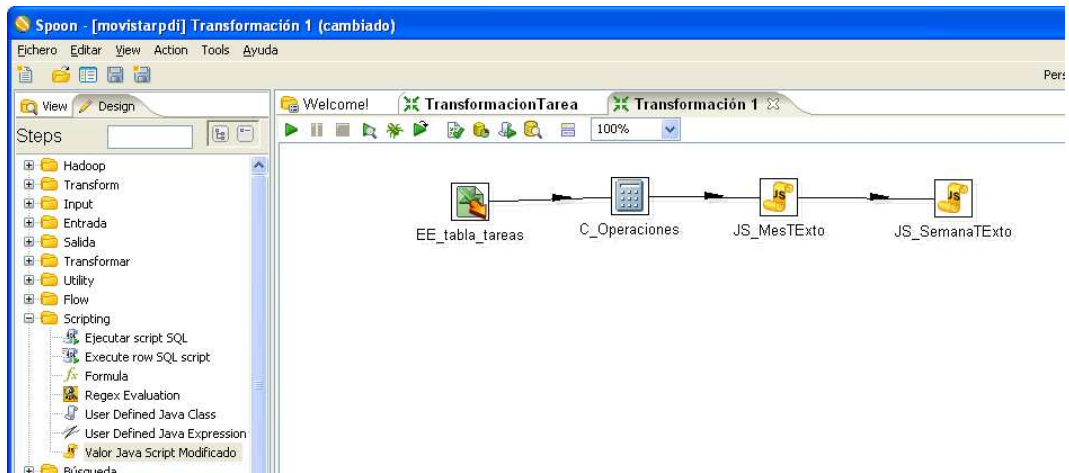


Figura 5.28: Proceso de Transformación

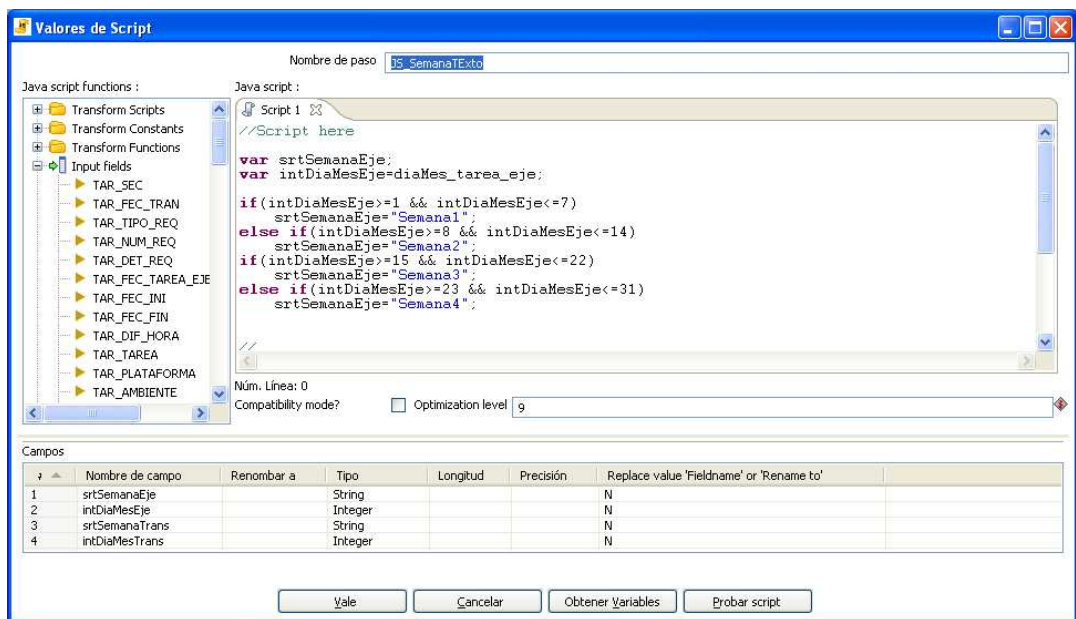


Figura 5.29: Código JavaScript

Se inicia con creación de las dimensiones

DimTiempoEjecuta

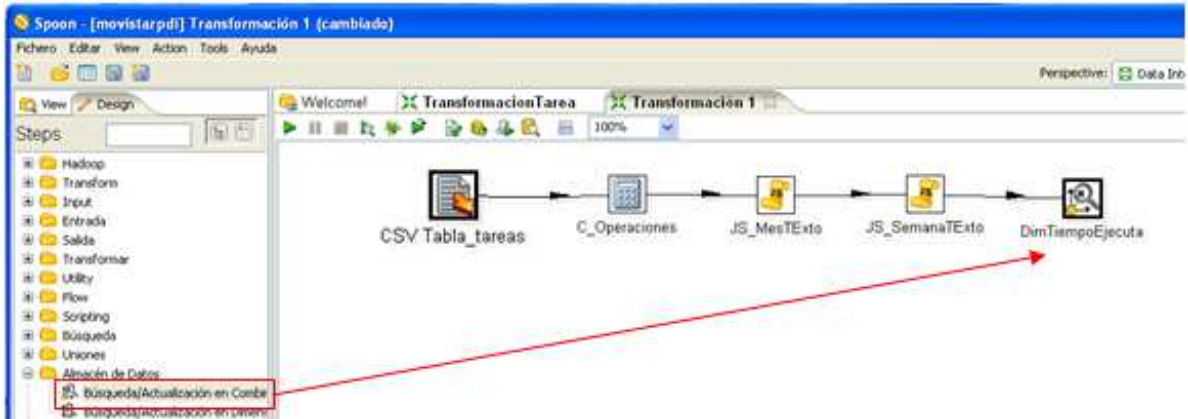


Figura 5.30: Configuración Dimensión

Se ingresa los campos de la dimensión y los campos del flujo

The dialog box 'Búsqueda/Actualización en Combinación' is shown with the following configuration:

- Nombre paso: DimTiempoEjecuta
- Conexión: movistar
- Esquema destino: [empty]
- Tabla destino: DimTiempoEjecuta
- Tamaño transacción: 100
- Tamaño cache: 9999

Campos clave (para buscar fila en tabla):

#	Campo de Dimensión	Campo en flujo
1	anio_tarea_eje	anio_tarea_eje
2	mes_tarea_eje	mes_tarea_eje
3	diaSemana_tarea_eje	diaSemana_tarea_eje
4	diaMes_tarea_eje	diaMes_tarea_eje
5	srtMesEje	srtMesEje
6	srtSemanaEje	srtSemanaEje
7	TAR_FEC_INI	TAR_FEC_INI
8	TAR_FEC_FIN	TAR_FEC_FIN
9	TAR_DIF_HORA	TAR_DIF_HORA

Campo de clave técnica: IdTiempoEjecuta

Creación de clave técnica:

- Utilizar máximo tabla + 1
- Utilizar secuencia
- Utilizar campo auto-incrementativo

¿Eliminar campos de búsqueda?

¿Utilizar código hash?

Campo con código Hash en tabla: [empty]

Date of last update field (optional): [empty]

Buttons: Vale, Cancelar, Obtener Campos, SQL

Figura 5.31: Configuración Dimensión

DimTiempoTrans



Figura 5.32: Configuración Dimensión

Ingreso de Campos de la dimensión DimTiempoTrans

Nombre paso: DimTiempoTrans

Conexión: movistar

Esquema destino: [Examinar...]

Tabla destino: DimTiempoTrans [Examinar...]

Tamaño transacción: 100

Tamaño cache: 9999

Campos clave (para buscar fila en tabla):

	Campo de Dimensión	Campo en flujo
1	anio_tarea_trans	anio_tarea_trans
2	mes_tarea_trans	mes_tarea_trans
3	diaSemana_tarea_trans	diaSemana_tarea_trans
4	diaMes_tarea_trans	diaMes_tarea_trans
5	srtMesTrans	srtMesTrans
6	srtSemanaTrans	srtSemanaTrans

Campo de clave técnica: idTiempoTrans

Creación de clave técnica:

- Utilizar máximo tabla + 1
- Utilizar secuencia
- Utilizar campo auto-incrementativo

¿Eliminar campos de búsqueda?

¿Utilizar código hash?

Campo con código Hash en tabla: []

Date of last update field (optional): []

[Vale] [Cancelar] [Obtener Campos] [SQL]

Figura 5.33: Configuración Dimensión

DimAmbiente

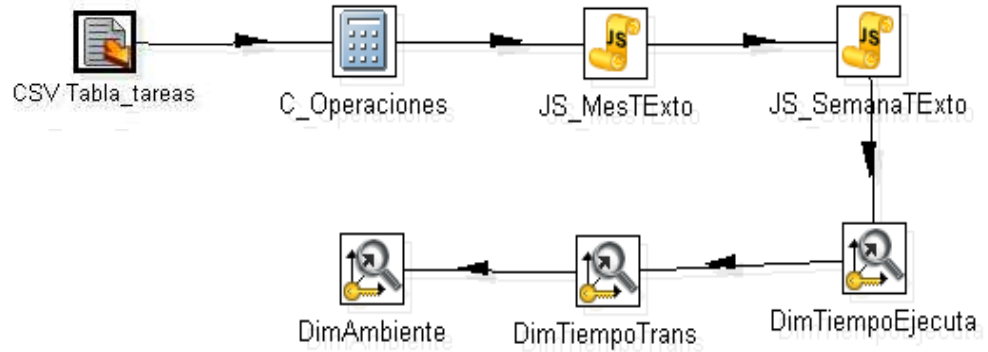


Figura 5.34: Configuración Dimensión

Ingreso de Campos en la dimensión DimAmbiente

Nombre paso: DimAmbiente

Conexión: movistar

Esquema destino: [Examinar...]

Tabla destino: DimAmbiente [Examinar...]

Tamaño transacción: 100

Tamaño cache: 9999

Campos clave (para buscar fila en tabla):

#	Campo de Dimensión	Campo en flujo
1	TAR_AMBIENTE	TAR_AMBIENTE
2	TAR_REPRUEBA	TAR_REPRUEBA
3	TAR_TIP_HORA	TAR_TIP_HORA

Campo de clave técnica: idAmbiente

Creación de clave técnica:

- Utilizar máximo tabla + 1
- Utilizar secuencia
- Utilizar campo auto-incrementativo

¿Eliminar campos de búsqueda?

¿Utilizar código hash?

Campo con código Hash en tabla: []

Date of last update field (optional): []

[Vale] [Cancelar] [Obtener Campos] [SQL]

Figura 5.35: Configuración Dimensión

DimRequerimiento

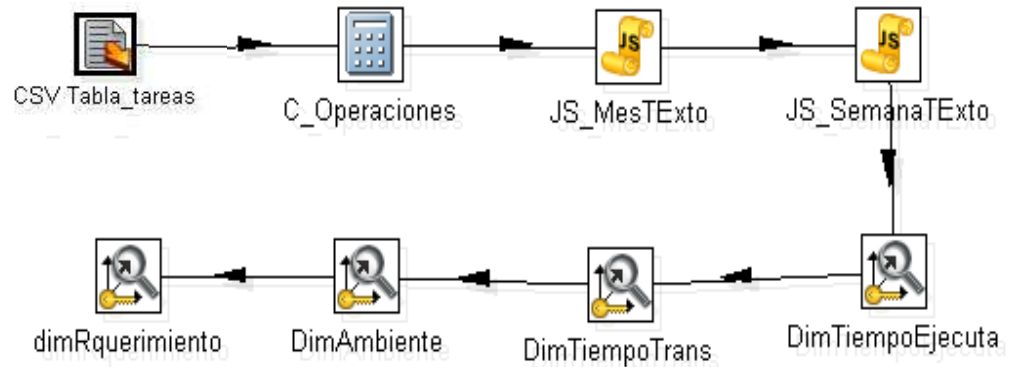


Figura 5.36: Configuración Dimensión

Ingreso de campos en la dimensión DimRequerimiento

Nombre paso: dimRequerimiento

Conexión: movistar

Esquema destino: []

Tabla destino: dimRequerimiento

Tamaño transacción: 100

Tamaño cache: 9999

Campos clave (para buscar fila en tabla):

	Campo de Dimensión	Campo en flujo
1	TAR_TIPO_REQ	TAR_TIPO_REQ
2	TAR_NUM_REQ	TAR_NUM_REQ
3	TAR_DET_REQ	TAR_DET_REQ
4	TAR_TAREA	TAR_TAREA
5	TAR_PLATAFORMA	TAR_PLATAFORMA
6	TAR_TIP_TAR	TAR_TIP_TAR

Campo de clave técnica: idRequerimiento

Creación de clave técnica:

- Utilizar máximo tabla + 1
- Utilizar secuencia
- Utilizar campo auto-incrementativo

¿Eliminar campos de búsqueda?

¿Utilizar código hash?

Campo con código Hash en tabla: []

Date of last update field (optional): []

[Vale] [Cancelar] [Obtener Campos] [SQL]

Figura 5.37: Configuración Dimensión

DimDirectivos

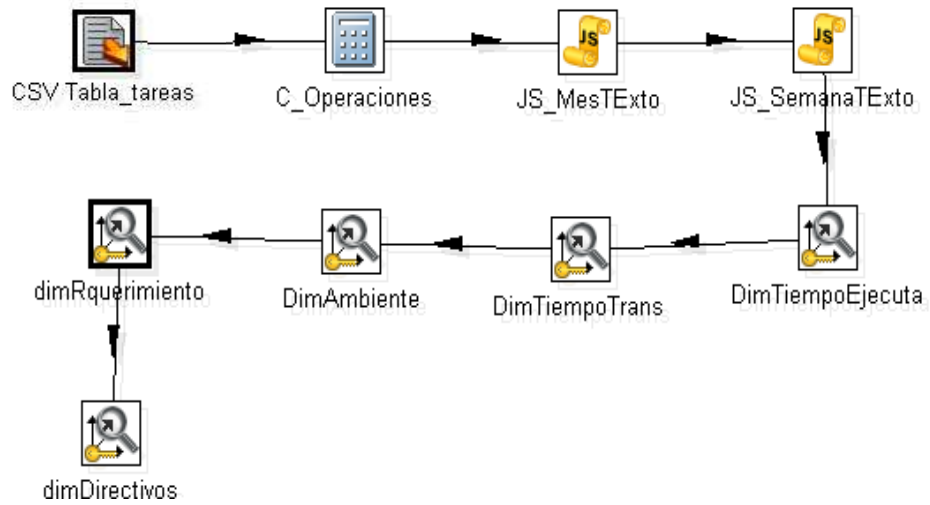


Figura 5.38: Configuración Dimensión

Ingreso de campos en la dimensión DimDirectivos

Nombre paso: DimDirectivos

Conexión: movistar

Esquema destino:

Tabla destino: dimDirectivos

Tamaño transacción: 100

Tamaño cache: 9999

Campos clave (para buscar fila en tabla):

#	Campo de Dimensión	Campo en flujo
1	TAR_ING_RES	TAR_ING_RES
2	TAR_NOTA_COORD	TAR_NOTA_COORD
3	TAR_APROB_ADMIN	TAR_APROB_ADMIN

Campo de clave técnica: idDirectivos

Creación de clave técnica:

- Utilizar máximo tabla + 1
- Utilizar secuencia
- Utilizar campo auto-incrementativo

¿Eliminar campos de búsqueda?

¿Utilizar código hash?

Campo con código Hash en tabla:

Date of last update field (optional):

Botones: Vale, Cancelar, Obtener Campos, SQL

Figura 5.39: Configuración Dimensión

DimMetrica

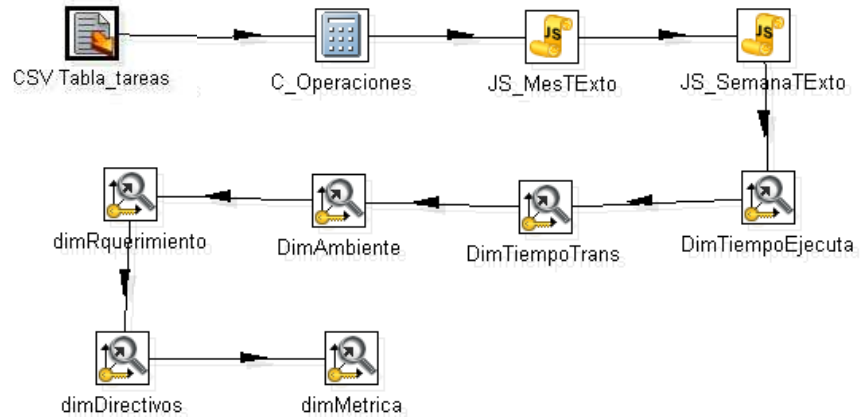


Figura 5.40: Configuración Dimensión

Ingreso de los campos de la dimensión DimMetrica

Búsqueda/Actualización en Combinación

Nombre paso:

Conexión:

Esquema destino:

Tabla destino:

Tamaño transacción: Tamaño cache:

Campos clave (para buscar fila en tabla):

#	Campo de Dimensión	Campo en flujo
1	Metrica	Metrica
2	idTiempoEjecuta	idTiempoEjecuta
3	idTiempoTrans	idTiempoTrans
4	idAmbiente	idAmbiente
5	idRequerimiento	idRequerimiento
6	idDirectivos	idDirectivos

Campo de clave técnica:

Creación de clave técnica:

Utilizar máximo tabla + 1

Utilizar secuencia

Utilizar campo auto-incrementativo

¿Eliminar campos de búsqueda?

¿Utilizar código hash?

Campo con código Hash en tabla:

Date of last update field (optional):

Figura 5.41: Configuración Dimensión

5.10.3. Creación del Cubo

Para la creación del Cubo se hace uso de la herramienta Pentaho, en el cual se asociara los campos de las dimensiones que se han creado, para generar las vistas de análisis

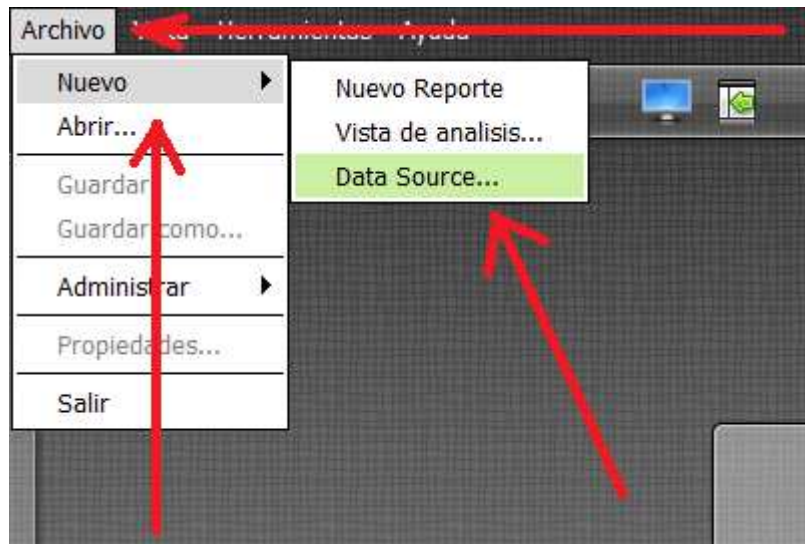


Figura 5.42: Creación DataSource

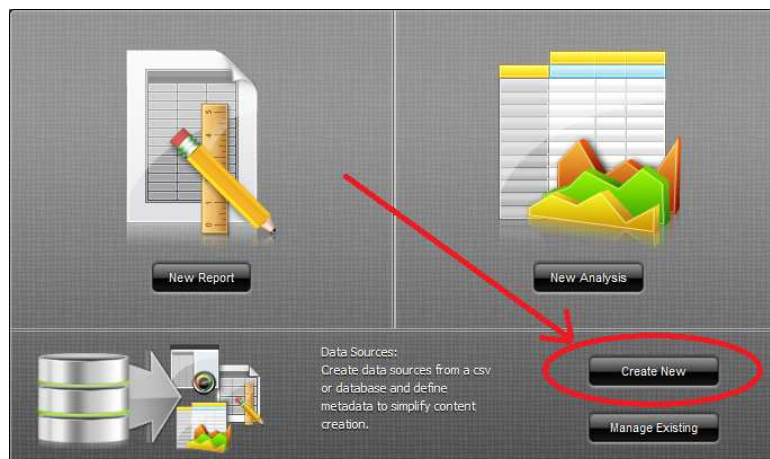


Figura 5.43: Creación DataSource

Creación del datasource, para generar la vista de análisis, se selecciona la conexión con la que se está trabajando 'movistar'.

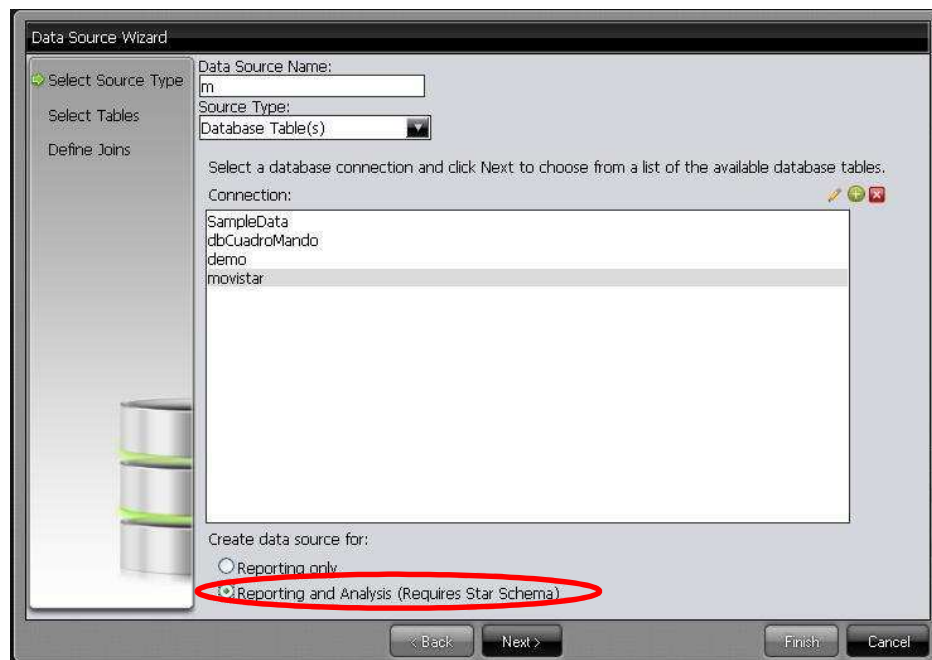


Figura 5.43: Configuración DataSource

Una vez seleccionado el datasource con el que se hará la vista de análisis, se deberá seleccionar las dimensiones que conformaran el cubo.

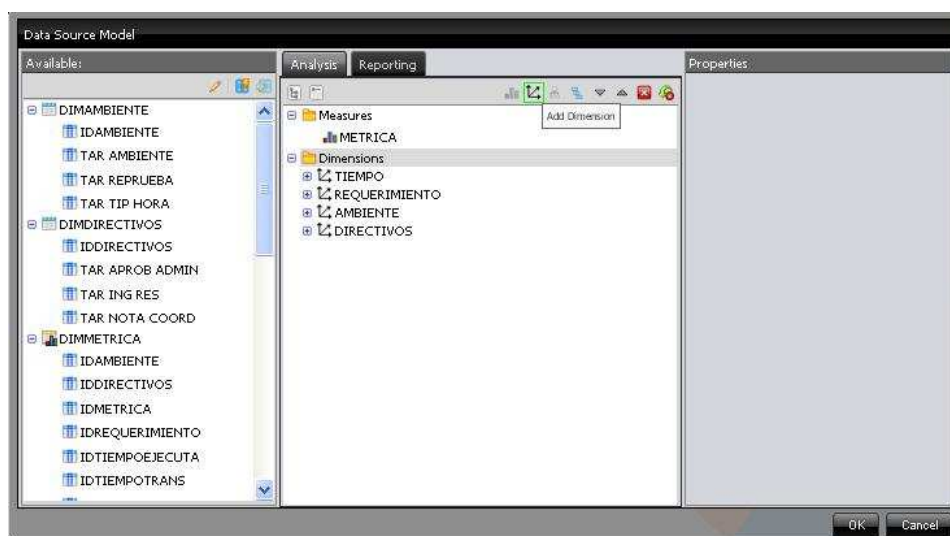


Figura 5.44: Configuración Vista de Análisis

Para poder observar la vista de análisis creada se debe seleccionar en la ventana de Pentaho Archivo/ Abrir/ Vista de Análisis



Figura 5.45: Menú Vista de Análisis

Se mostrará la vista un recuadro para seleccionar el esquema y cubo configurado anteriormente

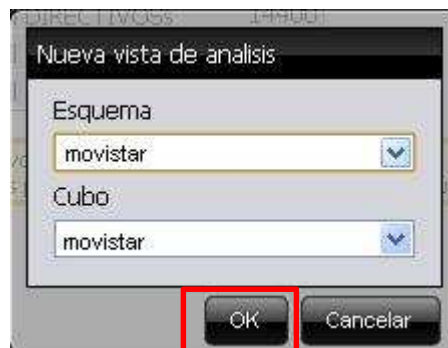


Figura 5.46: Selección Esquema

Se mostrará la vista de análisis se encontraran las dimensiones seleccionadas para la vista de análisis

Se pueden seleccionar o trabajar dinámicamente con las distintas dimensiones, de acuerdo a la información que se quiera obtener de la vista

5.10.4. Creación de Reportes

➤ El reporte **Requerimientos por Ambiente** permite obtener la cantidad de requerimientos realizados de manera:

- Mensual
- Semanal
- Anual

Agrupado por el **Tipo de Ambiente** en donde se realiza las prueba:

- Producción
- Preproducción
- Proyectos

Reporte por Mes y Año



Figura 5.49: Requerimientos por Ambiente

Reporte por Semana, Mes y Año

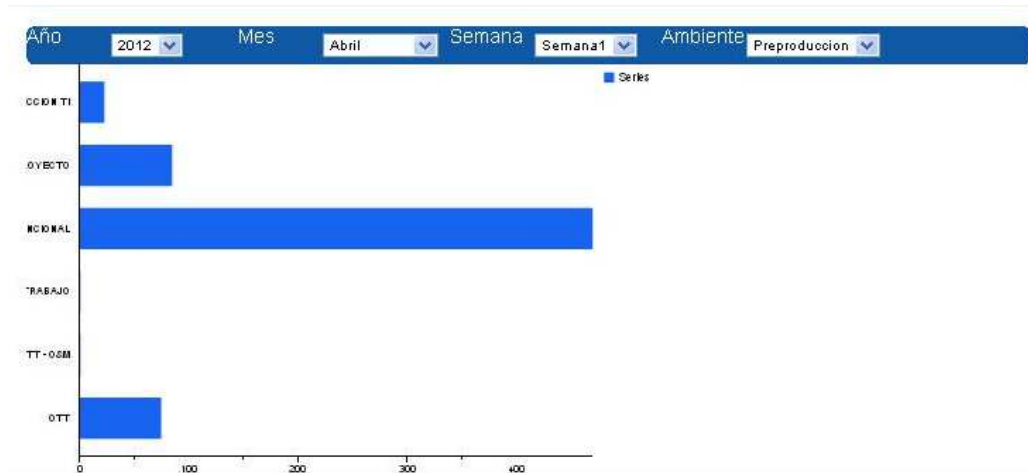


Figura 5.50: Requerimientos por Ambiente

- El reporte **Total Horas Trabajo – Ambiente** Permite obtener un resumen mensual de la cantidad de horas requeridas para realizar pruebas:
 - Normal
 - Extraordinaria

Agrupado por el Tipo de Ambiente:

- Producción
- Preproducción
- Proyectos

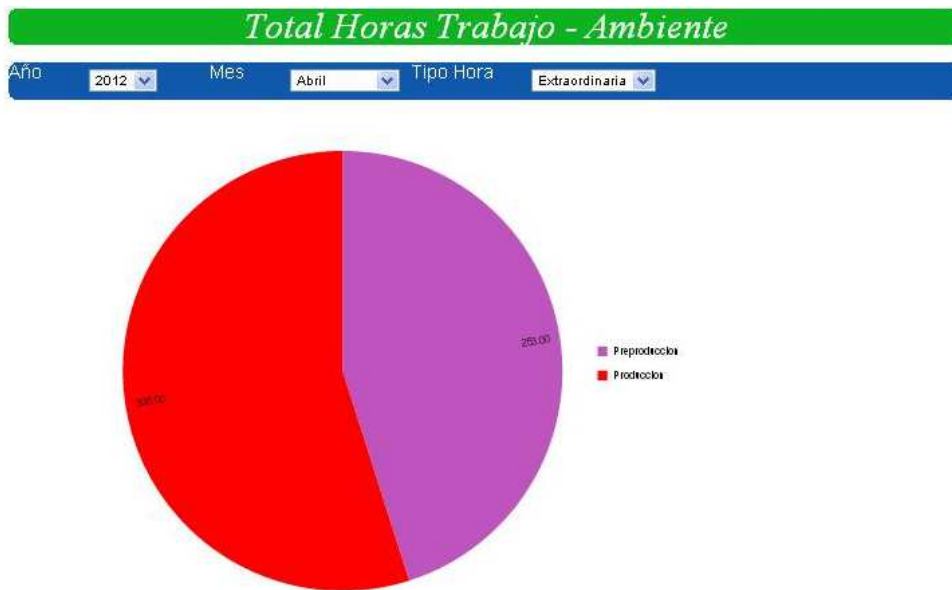


Figura 5.51: Total Horas Trabajo – Ambiente

➤ El reporte de **Número Requerimiento – Ambiente** permite obtener un resumen mensual de los requerimientos que se han realizado pruebas por cada ambiente:

- Producción
- Preproducción
- Proyectos



Figura 5.52: Número Requerimiento – Ambiente

En este reporte también se especifica, tal como se requiere la información, si existen repuebas (es decir que en el mismo requerimiento se realiza la misma prueba nuevamente por algún error detectado) y el conteo, si las horas en las que se realizaron las pruebas fueron normales o extraordinarias.

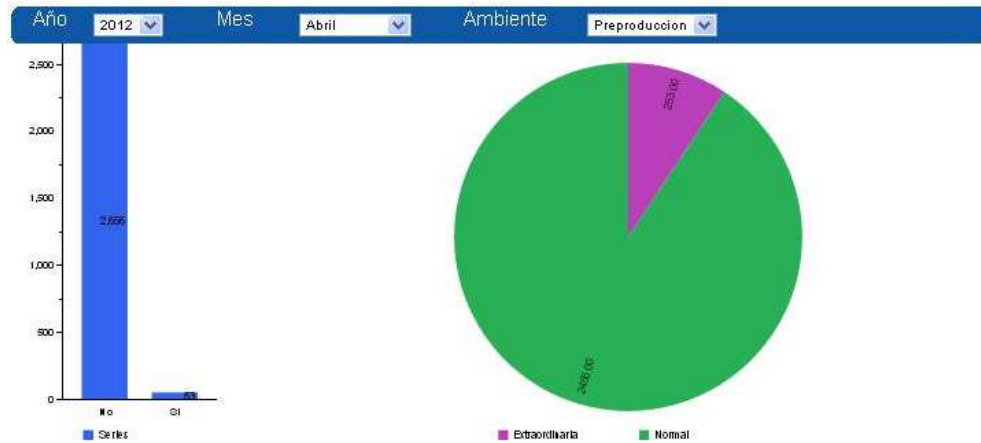


Figura 5.53: Número Requerimiento – Ambiente

- El reporte de **Tipo de Pruebas por Requerimiento** permite obtener el tipo de pruebas realizadas mensualmente por la clasificación de requerimientos ya sea:
 - Año
 - Mes
 - Requerimiento

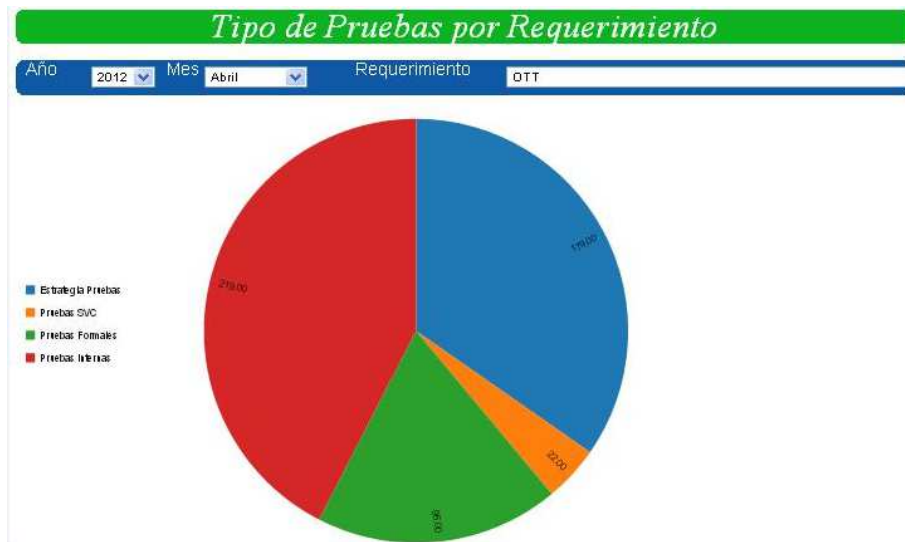


Figura 5.54: Tipo de Pruebas por Requerimiento

Obtiene el número de requerimientos realizados por el tipo de prueba:

- PE -> Pruebas Especiales
- PI -> Pruebas Internas
- EP -> Estrategias Especiales
- NA -> Pruebas SVC
- PF -> Pruebas Formales



Figura 5.55: Número Requerimiento – Ambiente

➤ El reporte **Tiempo Total Horas Ambiente** Permite obtener un resumen mensual de horas requeridas para realizar pruebas por ambiente:

- Preproducción
- Producción
- Proyectos



Figura 5.56: Tiempo Total Horas Ambiente

Espacio en Blanco Intencional

- El reporte **Número de Requerimientos** permite obtener el número de requerimientos realizados mensualmente.



Figura 5.57: Número de Requerimientos

- El reporte **Control Tipo Prueba – Ambiente** permite obtener un resumen mensual del número de tipos de pruebas realizadas en cada ambiente

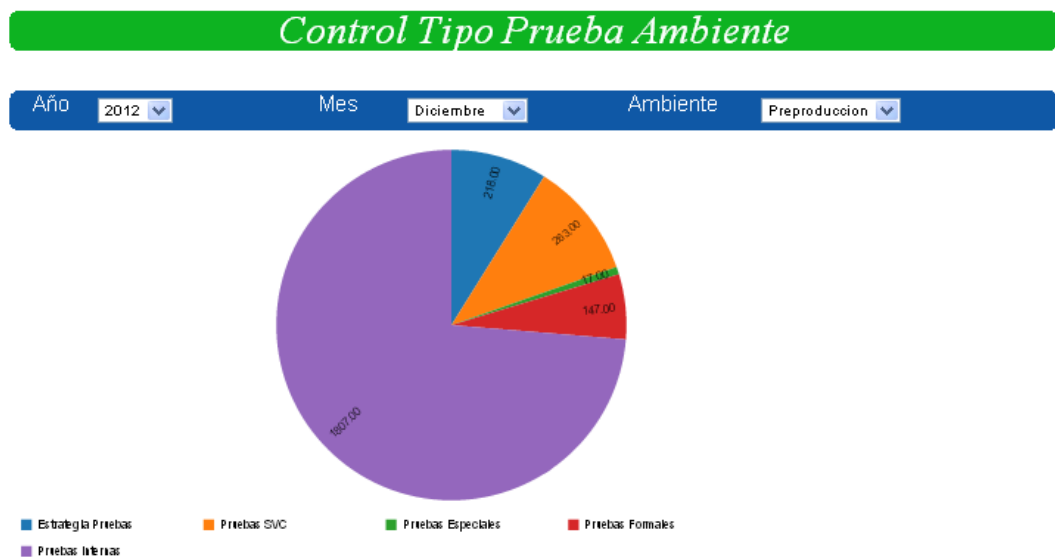


Figura 5.58: Control Tipo Prueba – Ambiente

Permite obtener un resumen semanal del número de tipos de pruebas realizados en cada ambiente

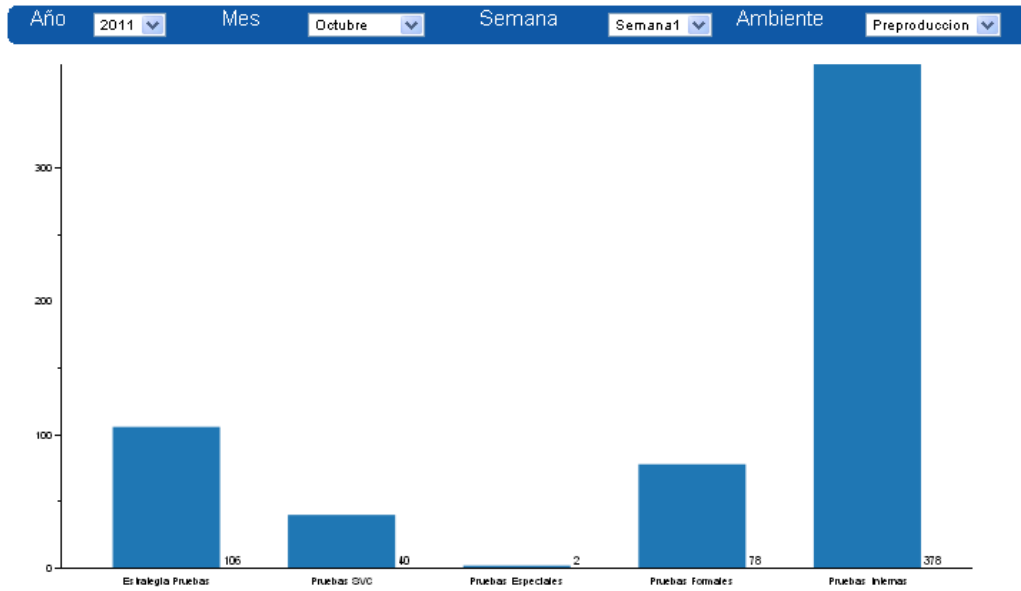


Figura 5.59: Control Tipo Prueba – Ambiente

CAPÍTULO 6

CONCLUSIONES Y RECOMENDACIONES

6.1. Conclusiones

- La implementación de la solución BI para el Área de Pruebas de la Gerencia de Construcción de Telefónica Movistar S.A., permite obtener indicadores en tiempo real, lo que ayuda a los ingenieros de pruebas a mantener un control continuo, sobre las tareas y los tiempos de respuesta en su realización, para de acuerdo a los resultados obtenidos se pueda determinar acciones de mejora o implementación de nuevas soluciones, en cuanto a los procesos que se manejan actualmente.
- Para la instalación de los utilitarios necesarios para implementar la solución BI, se investigó el funcionamiento de la herramienta Pentaho Community, basándose en manuales, búsquedas en línea y lectura de libros, para configurar adecuadamente el entorno en el que se requería trabajar con el desarrollo de la aplicación. Determinando que es una herramienta intuitiva, manejable, multiplataforma, sencilla al momento de integrar los datos, y lo más importante que al ser Open Source, los componentes requeridos para la generación de información adicional es accesible.

- La utilización de la Metodología Hefesto, permitió identificar fácilmente objetivos y resultados alcanzados, los cuales son sencillos de comprender, con lo que se determinó las necesidades del negocio, basándose en los requerimientos del usuario, involucrándolo así en cada etapa para la toma de decisiones oportunas ante los cambios del negocio, gracias a su estructura adaptable.
- Para definir la fuente de datos necesaria e identificar su funcionalidad, se mantuvo una estrecha relación con el usuario en la recopilación de los requerimientos, de esta manera se entendió la lógica del negocio y se definió cuáles serían los indicadores clave para el análisis y toma de decisiones.
- Se procedió a la implementación de la solución BI, siguiendo las fases comprendidas por la Metodología Hefesto, acoplando la solución a la infraestructura tecnológica manejada por Telefónica Movistar, utilizando la base de datos de producción, la red interna para que los usuarios puedan acceder a la aplicación alojada en uno de sus servidores.
- La herramienta de Pentaho Community (Data Integration), permite realizar la actualización de datos de manera sencilla, para que al momento de la obtención de los indicadores se obtenga resultados finales basándose en información al día, manteniendo las transformaciones y cálculos realizados en la carga inicial, los cuales se aplicarán automáticamente en las actualizaciones posteriores.

6.2. Recomendaciones

- Para la obtención y comprensión de la lógica del negocio se debe mantener una relación estrecha con el usuario en cada una de las fases que comprende la metodología aplicada, para obtener la información necesaria e identificar los puntos clave para el desarrollo de la solución.
- La iniciativa de uso de esta herramienta Pentaho Community, utilizada actualmente en el Área de Pruebas de la Gerencia de Construcción, serviría como un referente para que otras áreas puedan reducir tiempos en la obtención de indicadores, respecto a los procesos llevados en cada una de ellas y así contribuir a la toma de decisiones y planes de acción de manera oportuna.
- Se debe mantener reuniones constantes con los custodios de las bases de datos, para definir claramente los significados de los campos necesarios para obtener los indicadores y así precisar que tablas contienen los datos requeridos para el análisis, puesto que se detectaron que los nombres de algunos campos no eran tan específicos y en otros casos los datos eran inconsistentes o no constaban, por lo que es necesario verificar que los datos entregados sean los requeridos para realizar el análisis antes de ejecutar la carga inicial.

- Para la implementación de una solución BI, es importante definir la Metodología con la que se va a trabajar, puesto que se conoce cuáles serán los pasos a seguir para identificar gradualmente la información adecuada, obteniendo un resultado final efectivo sin cambios críticos en la fase final del proyecto.
- Al inicio del proyecto es necesario definir los recursos que se dispondrán para preparar el ambiente sobre el cual se construirá la solución, en la cual se instalarán todos los utilitarios necesarios y se hará uso de la base de datos existente.
- Se debe definir el formato de archivos para realizar las cargas iniciales en la herramienta, para que al momento de realizar las actualizaciones, únicamente se cambie el archivo existente por el actual, sin la necesidad de cambios bruscos en los demás procesos de transformación.

BIBLIOGRAFÍA

- **Arrayan. 2009.** Blogspot. [En línea] 14 de Abril de 2009. <http://abdarrayan.blogspot.com/>.
- **Bautista, César. 2011.** Blog BI. [En línea] 23 de Diciembre de 2011. [Citado el: Diciembre de 22 de 2012.] http://www.bi-ppr.com.mx/index.php?option=com_content&view=article&id=58&Itemid=115.
- **Bernabeu, Ricardo Dario. 2009.** [En línea] 04 de Mayo de 2009. <https://sites.google.com/site/magm33332/hefesto>.
- **Cano, Jose. 2007.** [En línea] Marzo de 2007. [Citado el: 1 de Octubre de 2012.] http://itemsweb.esade.edu/biblioteca/archivo/Business_Intelligence_competir_con_informacion.pdf.
- **Catarina. 2010.** [En línea] 2010. [Citado el: 05 de Septiembre de 2012.] http://catarina.udlap.mx/u_dl_a/tales/documentos/lis/rosette_u_jm/capitulo3.pdf.
- **Cavazos, Eliezer. 2008.** Gravatar. [En línea] 30 de Septiembre de 2008. <http://www.gravatar.biz/index.php/bi/metodologia-business-intelligence/>.
- **Corporation, Pentaho. 2005-2012.** *Pentaho*. [En línea] 2005-2012. [Citado el: 24 de 09 de 2012.] <http://www.pentaho.com/assets/pdf/3YOaBbQNjmeeyLc37nYG.pdf>.
- **—. 2008.** Pentaho. [En línea] 14 de 9 de 2008. [Citado el: 22 de 09 de 2012.] http://www.osbi.fr/wp-content/pentaho_editions_feature_comparison.pdf.
- **Dario, Ricardo. 2010.** Businessi.Intelligence. [En línea] 19 de Julio de 2010. [Citado el: 12 de 09 de 2012.] <http://www.businessintelligence.info/docs/hefesto-v2.pdf>.
- **Espinosa, Roberto. 2010.** [En línea] 20 de Julio de 2010. <http://churriwifi.wordpress.com/category/pentaho>.
- **—. 2010.** El Rincon del BI. [En línea] 19 de Abril de 2010. [Citado el: 22 de Agosto de 2012.] <http://churriwifi.wordpress.com/2010/04/19/15-2-ampliacion-conceptos-del-modelado-dimensional/>.
- **Gravatar. 2012.** Gravatar. [En línea] 2012. [Citado el: 19 de Septiembre de 2012.] <http://www.gravatar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/>.

- —. **2012**. Gravatar. [En línea] 2012. [Citado el: 25 de 09 de 2012.] <http://www.gravatar.biz/index.php/herramientas-bi/pentaho/caracteristicas-pentaho/>.
- **Ibermática. 2007**. [En línea] Octubre de 2007. [Citado el: 28 de Agosto de 2012.] http://www.larioja.org/upload/documents/458269_gestion_proyectos.pdf.
- **Ing. Bernabeu, Ricardo Dario. 2007**. linuxbi. [En línea] 07 de Noviembre de 2007. [Citado el: 08 de 25 de 2012.] http://www.linuxbi.com/files/datawarehouse_hefesto.pdf.
- **Jiménez, Claudia.** [En línea] <http://xue.unalmed.edu.co/~mfcabrera/db/oracle.pdf>.
- Pentaho Corporation. [En línea] [Citado el: 29 de 9 de 2012.] <http://www.pentaho.com/assets/pdf/yIDhC8GGezCNqR1VcAy4.pdf>.
- **Reyes Marroquín, Mario Roberto y Rosales Tejada, Pablo Augusto. 2007**. [En línea] Agosto de 2007. http://biblioteca.usac.edu.gt/tesis/08/08_0350_CS.pdf.
- **Rodríguez Sanz, Miguel. 2010**. [En línea] 22 de Julio de 2010. [Citado el: 2012 de Octubre de 27.] <http://es.scribd.com/doc/52203545/16/Metodologia-propuesta-por-Bill-Inmon>.
- **Salinas, Alexandro. 2008**. [En línea] 12 de Marzo de 2008. [Citado el: 22 de Septiembre de 2012.] <http://www.gravatar.biz/index.php/bi/introduccion-pentaho-parte-1/>.
- **Sansó, Bartomeu Vives. 2007**. [En línea] 2007. [Citado el: 25 de Septiembre de 2012.] http://dmi.uib.es/~labsoft/Labsg/4003_Labsg_Tema2-1.pdf.
- **TodoBI. 2011**. Todo BI. [En línea] 02 de Marzo de 2011. <http://todobi.blogspot.com/2011/03/metodologias-agiles-en-proyectos-bi.html>.
- **Tufiño, Jose. 2011**. bitstream. [En línea] Agosto de 2011. [Citado el: 27 de Agosto de 2012.] <http://bibdigital.epn.edu.ec/bitstream/15000/4101/1/CD-3861.pdf>.
- **Vicente, Instituto San.** [En línea] <http://www.iessanvicente.com/colaboraciones/oracle.pdf>.