



**UNIVERSIDAD AUTÓNOMA
DE AGUASCALIENTES**

**“Business Intelligence aplicada al Sistema
Integral de Administración en el INEGI”**

**Centro de Ciencias Básicas
Departamento de Sistemas de Información**

**Trabajo Práctico que para obtener el grado de Maestro en
Informática y Tecnologías Computacionales**

PRESENTA:

L.I. Juana Patricia Hernández Acosta

DIRECTOR DE TESIS:

M.C. Carlos Argelio Arévalo Mercado

SINODALES:

M.C. Fernando Flores Mejía

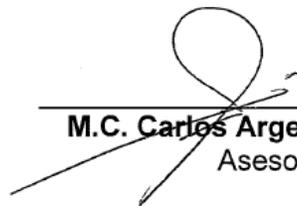
M.C. Francisco Javier Pinales Delgado

Ciudad universitaria 2008

Por este conducto autorizamos a la tesista:

L.I. Juana Patricia Hernández Acosta

La impresión de su documento de tesis, ya que cumple con los requisitos de contenido y forma exigidos por la Universidad Autónoma de Aguascalientes.



M.C. Carlos Argelio Arévalo Mercado
Asesor Principal



M.C. Fernando Flores Mejía
Sinodal



M.C. Francisco Javier Pinales Delgado
Sinodal

AGRADECIMIENTOS.

A DIOS por haberme brindado la maravillosa experiencia de la vida.

A mis padres, pues no hay manera de agradecer el esfuerzo y dedicación con que me han guiado, lo que hoy me permite el logro de una meta más.

A Yakamoko, por su apoyo incondicional a lo largo de mi vida.

A mis amigas Laura y Edna, por sus palabras de aliento y todos los momentos compartidos.

RESUMEN.

El Instituto Nacional de Estadística Geografía e Informática (INEGI), es el encargado de proveer al país de información confiable y oportuna en aspectos geográficos y estadísticos, para el logro de los procesos involucrados en la generación y distribución de esa información se toman decisiones basadas en información de múltiples fuentes, con los riesgos de confiabilidad que esto conlleva e incurriendo en prolongados periodos de tiempo.

Dado ese escenario es necesario implementar un sistema integral que permita contar con una base de datos única, y a partir del cual se disponga de una herramienta que proporcione información ejecutiva para una efectiva y rápida toma de decisiones.

Business Intelligence (BI) cuenta con la habilidad de consolidar información y analizarla con la suficiente velocidad y precisión para descubrir ventajas y tomar mejores decisiones de negocios, INEGI implementa con esta tecnología un sistema basado en la plataforma Pentaho, que logra proporcionar a sus directivos información de relevancia en el momento preciso en que es requerida como base para una toma de decisiones.

Contar con un sistema que facilitara la toma de decisiones fue de enorme utilidad para los directivos del instituto, ya que tienen la certeza de que la información tomada como fundamento en su toma de decisiones es confiable, a la vez que se disminuyeron tiempos, costos del personal que participaba en el tratamiento de la información y sobre todo riesgos; INEGI busca mantenerse dentro de las Instituciones a la vanguardia, y el análisis correcto y oportuno de los datos juega un rol primordial para esto.

Palabras clave: Business intelligence, toma de decisiones, dashboards, datawarehouse, tecnologías de información, INEGI.

INDICE

1. DESCRIPCIÓN DEL PROBLEMA.	1
1.1. Problemática general.	1
1.2. Problemática particular.	8
1.2.1. Contexto de la organización.	8
1.2.2. Sistemas integrales para la toma de decisiones.	13
1.2.3. Sistema Integral de Administración (SIA).	16
2. OBJETIVOS.	20
2.1. Objetivos generales.	20
2.2. Objetivos específicos.	20
3. ESTUDIOS DE CASOS SIMILARES.	21
3.1. Casos de BI en México.	21
3.2. Casos de BI internacionales.	29
4. TEORÍAS BASE Y TECNOLOGÍAS UTILIZADAS.	30
4.1. Teorías base.	30
4.1.1. Inteligencia de Negocios	30
4.1.1.1. La cadena de suministros.	32
4.1.1.2. CRM.	32
4.1.2. Datawarehouse (DW).	32
4.1.3. Tableros de control o Dashboards.	33
4.1.4. ERP.	34
4.2. Tecnologías utilizadas.	34
4.2.1. Pentaho.	34
4.2.2. Java.	37
4.2.3. Mondrian.	39
4.2.4. JPivot.	39
4.2.5. Jfreechart.	40
4.2.6. XML.	41
4.2.7. AJAX.	41

INDICE

5. DESCRIPCIÓN DE LA SOLUCIÓN.	43
5.1. Transformación de la plataforma.	45
5.2. Cambios a la configuración del servidor OC4J.	45
5.3. Deploy del archivo Pentaho .ear.	45
5.4. Ejecución general de una tarea en la plataforma Pentaho.	46
5.5. Desarrollo de tableros de control utilizando la información del SIA.	47
5.5.1. Diagrama de estrella.	47
5.5.2. Muestra de plazas ocupadas por Regional a través de un tablero de control integrando Google Maps.	48
5.5.3. Funcionamiento de un dashboard.	49
5.5.4. Funcionamiento de JPivot.	53
6. CONCLUSIONES.	55
7. GLOSARIO.	59
8. BIBLIOGRAFÍA.	67
9. ÍNDICE DE TABLAS.	73
10. ÍNDICE DE FIGURAS O DIAGRAMAS.	74
A. Apéndice A	76
A.1. Descarga de archivos.	76
A.2. Construcción del archivo .ear para Oracle OC4J.	76
B. Apéndice B	80
B.1. Conflictos con TopLink	80
B.2. Out of Memory Errors.	80
B.3. Agregar el archivo hsqldb.jar en el directorio [OC4J_Home/lib].	80
B.4. Modificar el archivo data-sources.xml.	80
B.5. Problema en la solución XSLTs.	82
B.6. Editar el archivo: [pentaho-solutions]/system/custom/xsl/files-list.xml.	83
B.7. El archivo antlr-2.7.6.	83
B.8. Colocar el hsqldb.jar, log4j.jar, log4j-boot.jar en la carpeta lib de oracle.	84

INDICE

B.9. Definición de DataSources en JDeveloper.	84
B.9.1. DS para Oracle.	85
B.9.2. DS para Hibérnate.	86
B.9.3. DS Sampledata.	87
B.9.4. DS para Quartz.	88
B.9.5. DS para Shark.	89
B.10. Configuración para mostrar las imágenes generadas automáticamente por la plataforma.	90
B.11. Pentaho se conforma de varios proyectos externos a su suite.	90
B.12. Archivos Orion-web.xml y Orion-application.xml.	91
B.13. Ejecución de Pentaho.	92
C. Apéndice C	93
C.1. Abrir el IDE.	93
C.2. Seleccionar el origen de la aplicación.	93
C.3. Carga del archivo.	94
C.4. Desglose de módulos.	94



DESCRIPCION DEL PROBLEMA

1. DESCRIPCIÓN DEL PROBLEMA.

1.1. Problemática general.

Sin duda, hoy en día las bases de datos son el punto neural para las grandes empresas e instituciones puesto que constituyen uno de los soportes fundamentales para el proceso de toma de decisiones; de ahí la importancia de que la información almacenada en ellas sea confiable y proporcione información de calidad. Estableciendo que la tecnología de la información (TI) se entiende como aquellas herramientas y métodos empleados para recabar, retener, manipular o distribuir información. La TI se encuentra generalmente asociada con las computadoras y las tecnologías afines aplicadas a la toma de decisiones (Bologna y Walsh, 1997). Las bases de datos (BD) en la mayoría de las organizaciones se depositan en una infraestructura básica TI, y se caracteriza por procesos manuales, con un mínimo control central, estándares de TI y políticas, contando además con débiles esquemas de seguridad, respaldo y administración como prácticas comunes.

Por otro lado los sistemas de información (SI) que operan en la infraestructura básica de las TI presentan múltiples problemas debido a la falta de integración de los datos. Los usuarios de estos SI se enfrentan a problemas relacionados con el elevado tiempo que tienen que dedicar a la obtención de información, en detrimento del que le deberían dedicar al análisis de la misma. En este contexto surge la necesidad de modificar los sistemas operacionales a nivel transaccional de la información. Se trata en definitiva de proporcionar a los directivos de las empresas las herramientas adecuadas para la explotación y análisis de los datos que les permitan obtener el conocimiento necesario en el proceso de toma de decisiones estratégicas.

Aunado a las practicas deficientes en el tratamiento de la información, cabe destacar que el volumen de información que se produce diariamente dentro de las organizaciones ha crecido tanto, que la eficiencia y calidad de la información depende cada vez más de la manera en que se administra y planifica el ciclo de creación y utilización de la misma. Contrariamente al comportamiento de otras ventajas, no necesariamente ganamos el mayor valor por tener más información, la cantidad de información producida todos los años es casi increíble; y tratar de integrarla que con la existente parece un esfuerzo colosal (David Loshin, 2003). La información es un activo y constituye un valor, igual que el dinero, los bienes de equipo, el trabajo o las materias primas. El valor de la información estriba, por paradójico que parezca en una sociedad definida como <sociedad de la información> y de la <opulencia comunicacional>, y en su recurso estratégico, ligado a los procesos de la toma de decisiones (Jesús García, 1998), no está ligada a los procesos que la generan: se produce en un tiempo y lugar distinto al de su utilización y aplicación, por lo que su valor está determinado por quien la usa, y no por quien la produce, “ El valor de la información es casi siempre proporcional al número (y el número proporcional al estatus) de las personas que forman el circuito por el que circula” (Jesús García, 1998). Los gobiernos no están exentos de esta situación.

La Organización de las Naciones Unidas (ONU) indica que uno de los factores clave para el éxito de los e-gobiernos es que estos aprendan a administrar información para crear conocimiento. Dicho de otro modo: para que los gobiernos generen valor público deben utilizar eficientemente la información, esto es, información que genere conocimiento. También señala que los gobiernos son el mayor, retenedor y proveedor de información, y agrega que en un mundo globalizado donde los mercados enfrentan una creciente competencia, la innovación es más importante que la productividad o la eficiencia en costos, y que tanto los individuos como los gobiernos no solo tienen que resolver nuevas ecuaciones para alcanzar sus objetivos, sino que

estas ecuaciones contienen un porcentaje cada vez mayor de variables desconocidas. Este entorno exige que los gobiernos eleven su capacidad para administrar información y crear conocimiento; de lo contrario enfrentarán una situación de abundancia información y déficit de conocimiento [1].



Figura 1. Origen del conocimiento.

Datos, información, conocimiento. Quizás la forma más sencilla de diferenciar los términos sea pensar que los datos están localizados en el mundo y el conocimiento está localizado en agentes de cualquier tipo (personas, empresas, máquinas, etc.), mientras que la información adopta un papel mediador entre ambos.

Los datos son la mínima unidad semántica, y se corresponden con elementos primarios de información que por sí solos son irrelevantes como apoyo a la toma de decisiones. También se pueden ver como un conjunto discreto de valores, que no dicen nada sobre el por qué de las cosas y no son orientativos para la acción.

Por ejemplo, un número telefónico o un nombre de una persona, son datos que, sin un propósito, una utilidad o un contexto no sirven como base para apoyar la toma de una decisión. Los datos pueden ser una colección de hechos almacenados en algún lugar físico como un papel, un dispositivo electrónico (CD, DVD, disco duro...), o la mente de una persona. En este sentido las tecnologías de la información han aportado mucho a recopilación de datos.

Como cabe suponer, los datos pueden provenir de fuentes externas o internas a la organización, pudiendo ser de carácter objetivo o subjetivo, o de tipo cualitativo o cuantitativo, etc.

La información se puede definir como un conjunto de datos procesados y que tienen un significado (relevancia, propósito y contexto), y que por lo tanto son de utilidad para quién debe tomar decisiones, al disminuir su incertidumbre.

Los datos se pueden transformar en información añadiéndoles valor:

- Contextualizando: se sabe en qué contexto y para qué propósito se generaron.
- Categorizando: se conocen las unidades de medida que ayudan a interpretarlos.
- Calculando: los datos pueden haber sido procesados matemática o estadísticamente.
- Corrigiendo: se han eliminado errores e inconsistencias de los datos.
- Condensando: los datos se han podido resumir de forma más concisa (agregación).

Por tanto, la información es la comunicación de conocimientos o inteligencia, y es capaz de cambiar la forma en que el receptor percibe algo, impactando sobre sus juicios de valor y sus comportamientos.

Información = Datos + Contexto (añadir valor) + Utilidad (disminuir la incertidumbre)

El conocimiento es una mezcla de experiencia, valores, información y know-how que sirve como marco para la incorporación de nuevas experiencias e información, y es útil para la acción. Se origina y aplica en la mente de los conocedores. En las organizaciones con frecuencia no sólo se encuentra

dentro de documentos o almacenes de datos, sino que también esta en rutinas organizativas, procesos, prácticas, y normas.

El conocimiento se deriva de la información, así como la información se deriva de los datos. Para que la información se convierta en conocimiento es necesario realizar acciones como:

- Comparación con otros elementos.
- Predicción de consecuencias.
- Búsqueda de conexiones.
- Conversación con otros portadores de conocimiento.

(Davenport y Prusak, 1998).

Cuando la supervivencia de una organización depende de cada vez más de su habilidad de extraer información correcta de datos, la calidad de los datos en si misma llega a ser muy importante, la calidad no está restringida a la exactitud y consistencia de los datos, abarca desde el punto en que la organización integra los datos hasta que hace un buen uso de los mismos a través de toda la estructura organizacional e implica también el escrutinio en el nivel inicial de la entrada de datos a la organización y en los valides de datos geográficamente dispersos, todo esto con referencia a otras bases de datos a través de la organización.

Estudios como los mencionados a continuación, han determinado que los problemas de la calidad de datos cuestan el 10 por ciento de sus costos totales. Si la calidad de los datos es pobre, el personal de la organización puede gastar el 25 por ciento de su tiempo manejando las quejas causadas por datos irregulares, inclusive corregir y encontrar faltantes en los datos que deben ser correctos al llegar a un departamento en particular, no tiene sentido.

En años anteriores, varios desastres han ocurrido por la calidad pobre de los datos, destacándose en los titulares de las noticias:

- "Correo del INS deja aturdida a escuela del vuelo". A consecuencia de documentos mal administrados, una escuela del vuelo de Florida recibió noticias del Servicio de Inmigración y Naturalización sobre la aprobación de visas de estudiante para dos de los terroristas del 11 de septiembre - seis meses después al día del ataque -[2].
- "Error en la relación Habitación-Precio del Hotel W Beneficia a Algunos Huéspedes de Nueva York". La calidad pobre de los datos impacto las ganancias de una cadena de hoteles cuando el costo de la habitación fue citada incorrectamente en su sitio web [3].
- "Respuesta correcta, mal Resultado: fallas en prueba cobran cuentas". En los últimos años, las compañías de pruebas educativas han experimentado averías graves en el control de calidad. Los errores en la industria de pruebas han afectado en los últimos tres años a millones de estudiantes que tomó pruebas de aptitud estandarizadas en por lo menos 20 estados. La compañía que dio califco las pruebas en Minnesota dio a 47.000 estudiantes resultados más bajos que los que merecían. Casi 9.000 estudiantes en la Ciudad de Nueva York fueron asignados erróneamente a la escuela de verano en 1999 a causa de un error por otra gran compañía [4].

Algunos ejemplos señalando específicamente la importancia de nombres de clientes y direcciones (todo en los Estados Unidos) [5]:

1. En una estimación, más de 175.000 IRS y cheques de reembolso de impuestos estatal fueron marcados como "imposibles de entregar" por el departamento postal en un año.
 2. En una auditoría compararon las listas de registro de votantes a los datos recogidos de la oficina de correos de reubicación de datos, se
-

estima que 15 a 20 por ciento de los votantes en las listas de registro de votantes se han trasladado o han fallecido.

3. Una sociedad absorbente cerró un acuerdo para un nuevo negocio con sólo la mitad de los clientes previstos. ¿El problema? La duplicación de datos.
4. Una empresa de fibra óptica EE.UU. perdió \$ 500,000 después de un envío en donde las etiquetas incorrectas causaron que el cable inapropiado se instalara en el fondo de un lago.
5. Gobierno de los EE.UU. estima que miles de millones de dólares se pierden cada año debido a la mala calidad de los datos.

Estos ejemplos ilustran los severos problemas causados por la calidad de datos en general, los problemas corporativos que se relacionan con la toma de decisiones a nivel directivo basadas en la información con la que cuentan a través de su estructura de TI, se observa que la mayoría de directores no disponen a tiempo de la información que necesitan para evitar costosas caídas. ¿Cómo puede un director justificar sus decisiones sin tener hechos tangibles? Si el acceso a la información no es claro ni rápido, la mayoría de decisiones pueden tomarse demasiado tarde.

En México los gobiernos hacen esfuerzos para que cualquier persona y en cualquier momento disponga de información pública, pero se hace cada vez más evidente que nuestra sociedad requiere, mas bien, no solo el acceso a la abundante información pública, sino que sea confiable y precisa, lo que significa confiar en la fuente que la provee.

Existen importantes logros y avances en materia de Gobierno Digital para México, lo que le ha permitido mejorar su posicionamiento en el contexto internacional. Por ejemplo, el Reporte Mundial 2003 de la ONU, lo ubica en el sitio 30 de entre cerca de 200 países evaluados. Lo coloca, además, en el cuarto lugar en materia de sofisticación y madurez de servicios

gubernamentales en línea, en noveno lugar en e-Participación y reconoce que México es una de las diez naciones que ha progresado de manera más efectiva y más rápida en sus estrategias de e-Gobierno.

Así mismo, en el estudio e-Government Leadership: Engaging the Customer, 2002-2003, publicado por la firma consultora ACCENTURE, México se posiciona como el país con mayor grado de avance en materia de Gobierno Electrónico entre las 24 naciones estudiadas [6].

1.2. PROBLEMÁTICA PARTICULAR.

1.2.1. Contexto de la organización.



Figura 2. Presencia de INEGI en el territorio nacional.

El Instituto Nacional de Estadística, Geografía e Informática (INEGI) es un órgano desconcentrado de la Secretaría de Hacienda y Crédito Público, con autonomía técnica y administrativa [INEGI].

Los objetivos del Instituto son:

- Generar e integrar información estadística y geográfica sobre el territorio, la población y la economía de México;
- Proporcionar a la sociedad el servicio público de información estadística y geográfica; así como
- Normar, coordinar y promover el desarrollo de los Sistemas Nacionales Estadístico y de Información Geográfica, con el objeto de
- Satisfacer las necesidades de información de los diversos sectores de la sociedad.

Para prestar el servicio público de información estadística y geográfica a una población superior a los 100 millones de habitantes, el INEGI cuenta con una organización conformada por:

La junta de gobierno. Conformada por el titular de la Secretaría de Hacienda y Crédito Público, quien la presidirá, tres funcionarios públicos de la SHCP, nombrados por el titular de la misma, un funcionario público de la Secretaría de Economía, nombrado por el titular de la dependencia anteriormente citada, el Presidente del INEGI y tres vocales; y que tiene como algunas de sus funciones: Aprobar anualmente la propuesta del Presidente del Instituto, el plan de trabajo del mismo y, el presupuesto anual; así como conocer el informe anual de actividades del Instituto y emitir sugerencias de mejora y dar seguimiento a los puntos anteriores y aprobar las políticas para la administración de los recursos humanos, financieros y materiales del Instituto.

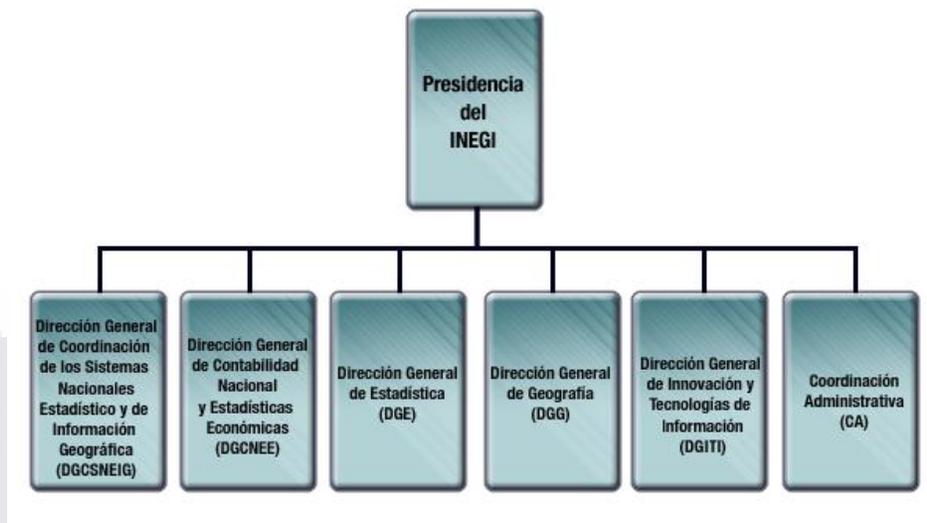


Figura 3. Estructura organizacional del INEGI.

Presidencia del INEGI. La sede de la Presidencia del Instituto está establecida en la ciudad de Aguascalientes. Entre las atribuciones que tiene el titular, están: representar al Instituto ante la Junta de Gobierno y establecer con esta última las políticas para la administración de los recursos humanos, financieros y materiales del INEGI, de acuerdo con sus programas y objetivos; así como la planeación estratégica y los lineamientos generales para la consecución de las funciones del Instituto, y los demás acuerdos emanados de la Junta de Gobierno.

Unidades administrativas. Para el ejercicio de sus funciones a nivel nacional y de acuerdo al artículo 102 del Reglamento Interior de la SHCP, el INEGI cuenta con las siguientes unidades administrativas.

Ámbito central

1. Dirección General de Coordinación de los Sistemas Nacionales Estadístico y de Información Geográfica
2. Dirección General de Contabilidad Nacional y Estadísticas Económicas
3. Dirección General de Estadística
4. Dirección General de Geografía
5. Dirección General de Innovación y Tecnologías de Información
6. Coordinación administrativa

Ámbito regional y estatal

Dirección Regional. Tiene entre sus facultades y atribuciones las siguientes:

1. Ejecutar, coordinar y supervisar los programas del Instituto, ajustándose a las disposiciones aplicables y a las normas, sistemas y procedimientos que al efecto establezcan las unidades administrativas centrales, y de acuerdo con los lineamientos que al efecto dicte el Presidente del propio Instituto.
2. Representar al Instituto ante las autoridades estatales y municipales que correspondan a su jurisdicción, para la ejecución de los programas y concertaciones en que aquéllas participen conjuntamente con el Instituto.
3. Supervisar y controlar el correcto ejercicio de las partidas que integran su presupuesto autorizado.

Estas unidades administrativas contarán con coordinaciones estatales que ejecutarán los programas regionales, dentro de la entidad federativa correspondiente.

Para la consecución de sus objetivos en el país, el INEGI cuenta con diez direcciones regionales y 34 coordinaciones estatales, que a continuación se señalan:

Regional	Sede	Coordinaciones estatales
Noreste	Monterrey, Nuevo León	<ul style="list-style-type: none"> • Nuevo León • Coahuila de Zaragoza • Tamaulipas
Norte	Durango, Durango	<ul style="list-style-type: none"> • Durango • Chihuahua • Zacatecas
Noroeste	Hermosillo, Sonora	<ul style="list-style-type: none"> • Sonora • Sinaloa • Baja California • Baja California Sur
Occidente	Guadalajara, Jalisco	<ul style="list-style-type: none"> • Jalisco • Colima • Michoacán de Ocampo • Nayarit
Centro Norte	San Luis Potosí, S.L.P.	<ul style="list-style-type: none"> • San Luis Potosí • Guanajuato • Querétaro de Arteaga • Aguascalientes
Centro Sur	Toluca, estado de México	<ul style="list-style-type: none"> • Estado de México • Guerrero • Morelos
Oriente	Puebla, Puebla	<ul style="list-style-type: none"> • Puebla • Hidalgo • Tlaxcala • Veracruz de Ignacio de la Llave
Sur	Oaxaca, Oaxaca	<ul style="list-style-type: none"> • Oaxaca • Chiapas • Tabasco
Sureste	Mérida, Yucatán	Yucatán Campeche Quintana Roo
Centro	México, Distrito Federal	Distrito Federal

Tabla 1. Relación de estados por ámbito Regional-Estatal.

1.2.2. Sistemas integrales para la toma de decisiones.

Dada la creciente demanda de información estadística del país, INEGI como un Órgano Descentralizado que apoya la generación de esta información a través de la realización de diferentes proyectos censales como encuestas, registros administrativos o censos, los directivos del Instituto requieren tomar decisiones sobre la estructura organizacional que deberá soportar a estos eventos en tiempo y en forma, por lo que tienen necesidad de información del personal que se encuentra en las diferentes Unidades Administrativas; para satisfacer dichas necesidades se apoyan en diferentes productos como:

1. Sistema Integral de Administración (SIA) mismo que por ser la base transaccional del presente estudio se describirá a detalle en el punto 1.2.3.
2. Sistema de Control de Producción y Ventas (ConProVe). El cual incluye los siguientes subprocesos:
 - Integración del Programa Anual Editorial, en el cual se recopilan los requerimientos editoriales.
 - Planeación de la producción, donde se establecen las prioridades en la generación de productos de acuerdo a los requerimientos editoriales, se prevén los insumos necesarios para la producción y se determinan los procesos productivos para la generación de cada producto.
 - Cotización de productos donde se puede estimar el costo de generación de los mismos.
 - Generación y control de órdenes de producción, con el que se da inicio a la producción editorial y procesos de impresión indicando la línea a seguir.
 - Seguimiento a producción, donde se registran los avances en cada uno de los procesos y actividades realizadas.

- Control de la producción, donde se registran tanto los tiempos (mano de obra) e insumos utilizados para la generación de cada producto.
 - Trámite de Derechos de Autor de cada producto.
 - Estimación de tiraje de productos, donde se establece el total de productos a generar de acuerdo a la demanda de las Áreas Generadoras y la Red Nacional de Consulta.
 - Almacenaje y distribución de productos, que considera el control de inventarios y seguimiento a la distribución a nivel nacional.
 - Comercialización, en el cual se llevan las actividades de venta de productos y servicios.
 - Registros contables derivados de los costos de producción, movimientos de almacén y ventas de productos y servicios.
3. Control de acceso. Proceso para llevar a cabo la administración automatizada de los accesos y por consecuencia de la asistencia del personal del Instituto y de las personas que los visitan.
 4. Diversos registros administrativos tratados comúnmente con herramientas de office tradicional.

Para controlar la información concerniente a los procesos administrativos involucrados en el funcionamiento del instituto, INEGI cuenta ya con el SIA que lleva el papel del Sistema de Planificación de Recursos Empresariales (ERP).

La característica principal de esta aplicación, es que fue desarrollado de manera interna por trabajadores del Instituto. Para llegar a tal decisión se evaluaron alternativas de solución, dentro de las que se encontraban los siguientes GRP's (Government Resource Planning): PeopleSoft, SAP, Oracle, entre otros, de igual manera se visitaron oficinas de gobierno que se encontraban en el proceso de implantación de alguno de estos sistemas, con

el fin de contar con métricas como: costo, tiempo de implementación, soporte técnico ofrecido por el fabricante, que permitieran sondear y evaluar más acertadamente sus alcances y limitaciones.

Las métricas de interés en las alternativas analizadas contaban con desventajas substanciales, mismas que las descartaron al momento de la selección para el desarrollo del ERP para el instituto, a continuación se mencionan las más relevantes:

1. Alto costo de las soluciones comerciales.
2. Tiempo de implementación excesivo.
3. Ajuste de los procesos de la empresa, a la aplicación.
4. Costo extra por desarrollo o ajustes de nuevos módulos.
5. Costo por capacitación o asesorías.

1.2.3. Sistema Integral de Administración (SIA).

El SIA integra la información de diversos servicios que facilitan la labor cotidiana del instituto, entre los cuales se destacan los que se identifican en el siguiente diagrama.

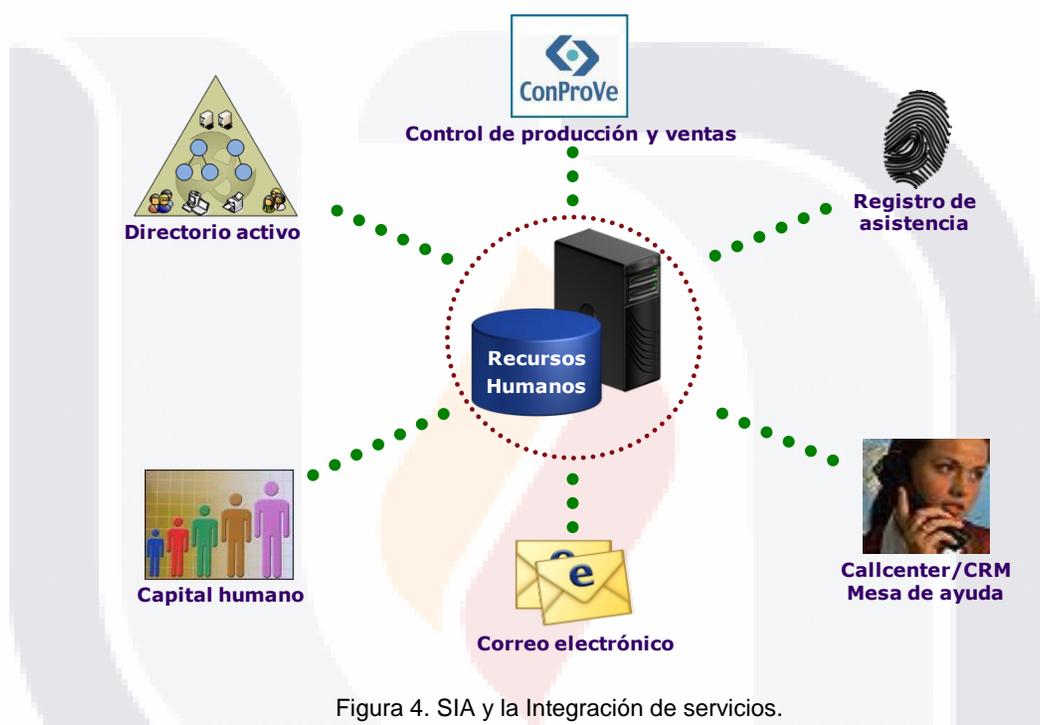


Figura 4. SIA y la Integración de servicios.

Al realizar un análisis de los procesos existentes en el SIA y la operación de las áreas administrativas a efecto de cubrir con los requerimientos de información de los directivos, se encontraron las siguientes áreas de oportunidad:



Figura 5. Esquema para la toma de decisiones anterior a la implantación de BI.

1. El SIA, no cuenta con un módulo de extracción de indicadores para la toma de decisiones.
2. Cruzar o relacionar información de la organización es muy difícil, por ejemplo, costos de la plantilla poblada, comparativos del mes con registros históricos, etc.
3. Recaptura y duplicidad de información-procedimiento en diversas áreas a efecto de proveer de información valiosa al ejecutivo.
4. Desgaste de las áreas de enlace administrativas en realizar un reporte ad-hoc de cada ejecutivo.
5. Diversidad de reportes locales para dar respuesta a peticiones del área administrativa central.
6. No hay homologación entre los productos generados en concepto y contenido.

Ante la importancia de las áreas de oportunidad identificadas, era necesario establecer un esquema en el que los usuarios pudieran autoabastecerse de la información que requerían, sin depender de un área experta.

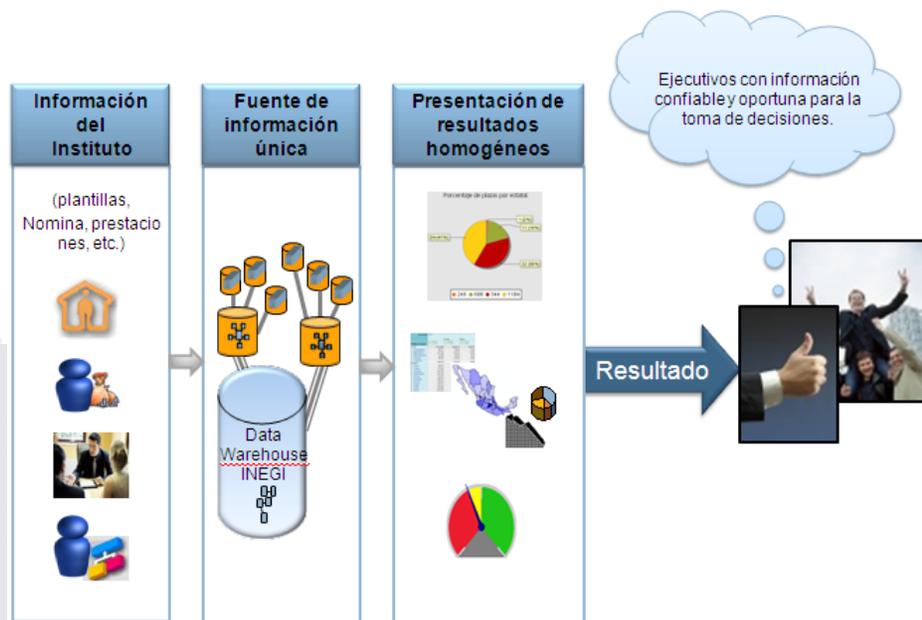


Figura 6. Esquema esperado para la toma de decisiones utilizando BI.

Al aplicar una mejora al proceso de extracción de información para proporcionar información precisa que respalde y apoye la labor de los directivos, impacta en los aspectos que se detallan en el siguiente caso muestra de la toma de decisiones con el proceso actual.

CASO MUESTRA. Consulta sobre el total de la plantilla poblada y su costo en nomina a las direcciones generales a nivel central del instituto.

1. Número de direcciones regionales: 5
2. Número mínimo de personas atendiendo la solicitud: 1
3. Puesto promedio del personal que realiza las funciones sobre el control y manejo de plantillas: Profesional ejecutivo de servicios especializados
4. Sueldo mensual:
5. Procedimiento. La Coordinación Administrativa solicita a las áreas de recursos humanos de cada dirección general, el total de su plantilla

poblada, así como el total de los pagos realizados por nomina a efecto de realizar un concentrado para el presidente del instituto.

ACCION	RESULTADO
• Tiempos de respuesta a transacciones	3 a 5 días
• Costos	\$4,300 Por Técnico operativo
• Número de procesos	5 Uno por cada dirección general
• Número de posibles errores	6 Uno por cada dirección general El de la CA
• Cantidad de personas involucradas en el proceso	7 El solicitante por parte de presidencia El solicitante de CA El personal de cada dirección gral.
• Número de pasos requeridos	6 Solicitud de presidencia Solicitud de CA a las direcciones gcales. Proceso por dirección Envío de resultados a CA Concentrado por parte de CA Envío del concentrado a presidencia
• Fuentes de información	6 Posibilidad de que cada dirección general tenga su propio repositorio de información sobre concentrados

Tabla 2. Acciones necesarias actualmente para la toma de decisiones.

Estos resultados muestran un panorama general de lo que puede ser la toma de decisiones en el instituto. Debe considerarse que dicho panorama puede agravarse al involucrar la totalidad de alcance a nivel nacional con que cuenta INEGI; por ejemplo, realizar la misma consulta ahora involucrando las diez direcciones regionales y las 34 coordinaciones estatales correspondientes.

TESIS TESIS TESIS TESIS TESIS



OBJETIVOS

TESIS TESIS TESIS TESIS TESIS

2. OBJETIVOS.

2.1. Objetivos generales.

1. Contar con un sistema informático integral que conforme una base de datos única como resultado de las actividades relacionadas con el manejo de personal, recursos materiales y financieros, a partir del cual se disponga de una herramienta para la toma de decisiones a fin de lograr una mayor eficiencia en el manejo de los recursos del Instituto.
2. Proporcionar información ejecutiva para una efectiva y rápida toma de decisiones.

2.2. Objetivos específicos.

1. Contar con un sistema de toma de decisiones que explote los datos transaccionales del SIA en las actividades relacionadas con el manejo de los diferentes módulos que lo componen.
2. Generar el modelo de datawarehouse (DW) con sus respectivas dimensiones y datamarts (DM).
3. Simplificar procesos.

TESIS TESIS TESIS TESIS TESIS



ESTUDIOS DE CASOS SIMILARES

TESIS TESIS TESIS TESIS TESIS

3. ESTUDIOS DE CASOS SIMILARES.

3.1. Casos de BI en México.

El mercado de BI en México.

El mercado de BI en México alcanzará un valor de \$45 millones de dólares en 2008 [InformationWeek]. Con herramientas de Business Intelligence (BI), México ajusta el Presupuesto de Egresos de la Federación que envía la Secretaría de Economía a la Cámara de Diputados. Los diferentes Centros de Estudios (CE) que apoyan a los diputados (como el CE de las finanzas públicas, el de derecho e investigaciones parlamentarias y el CE para el adelanto de las mujeres y la equidad de género, entre otros) manejan importantes cúmulos de información. Estos datos son gestionados para después analizarlos de manera cualitativa y cuantitativa en las diversas materias de interés para el trabajo legislativo. Cada año, a través de los CE, que se enfocan en el análisis de información, se apoya a la Cámara de Diputados en la revisión de la propuesta del Presupuesto de Egresos de la Federación que hace la Secretaría de Hacienda.

Por otro lado Publio Rábago Riquer, director de sistemas del CE para el desarrollo rural sustentable y soberanía alimentaria, explica cómo ocurre este proceso: “El presupuesto llega en formato de papel, con muchos PDF dispersos, que fragmentan la información. Lo primero que se hace es entender la composición del presupuesto; es decir, cómo se propone el Ejecutivo ejercer los recursos de la Federación, en qué proporción se ejercerá el gasto en empleo, egresos, medio ambiente y el aspecto social. Así, en cuestión de horas, los diputados pueden empezar a aportar opiniones y a tomar decisiones de cómo ajustar el presupuesto original”. Recientemente esta herramienta logró acoplar el presupuesto para un programa al que originalmente se le habían asignado \$120,000 millones de pesos, quedando finalmente en

\$180,000 millones una vez que se analizó la información, se previeron los cambios y modificaciones, y se añadieron a dicho programa.

El impacto económico. Según una nueva investigación llevada a cabo por Information Builders en mayo de 2007, los empleados pierden aproximadamente 67 minutos al día buscando información. El estudio revela que la falta de sistemas de toma de decisiones en las organizaciones cuesta millones de dólares al año en pérdida de productividad, dificultando así la competitividad.

¿Qué representa que cada empleado destine más de una hora para hallar datos? Es el equivalente a 15.9% de su tiempo de trabajo diario perdido. Para una organización de 1,000 trabajadores que eroga \$70,000 dólares diarios en salarios, equivaldría a casi \$11 millones de dólares anuales de salario perdido por los empleados buscando información para tomar una decisión.

Según la investigación, los directivos también están frustrados y desmotivados por el mal acceso a datos precisos, y 54% de los encuestados manifiestan que la falta de información indicada, consistente y completa es la mayor barrera para tomar buenas decisiones. Como resultado de esta situación tienen que confiar mucho más en su experiencia que en buena información para tomar decisiones, lo que 63% de los participantes en la encuesta calificó como factor clave para tomar una buena decisión.

Asimismo, los encuestados consideran que el fácil acceso a la información (55%), el fácil uso de herramientas de búsqueda (40%) y los informes oportunos (25%) les ayudarían a ser más productivos en su trabajo.

Gerry Cohen, CEO de Information Builders, indica que el hecho de que decisiones importantes sean tomadas con base en información inconsistente, imprecisa y prematura está alarmando a las organizaciones. “Es también

preocupante que los empleados estén perdiendo un tiempo excesivo de su trabajo en buscar datos, en lugar de hacer su trabajo”, añade.

Es por esto que las organizaciones del mundo deben analizar los sistemas y tecnologías que pueden suponer barreras en el modo en el que se accede, distribuye y actualiza la información, a fin de asegurarse de que los usuarios de la empresa y terceros pueden tomar decisiones óptimas y rápidas que tengan un impacto positivo en el rendimiento del negocio.

Al usar BI de forma correcta, las organizaciones podrán maximizar los beneficios de su inversión en TI existente y de los datos que tienen disponibles para convertirlos en información significativa, y para prueba está la Cámara de Diputados. Ahora, sólo falta que se dejen de conflictos partidistas.

CIFRAS DE INTERÉS.

- Sólo 44% de los negocios tiene una solución de información empresarial global.
- El 68% de los encuestados considera que recibir la información de toda la empresa de forma conjunta y fácil le ayudaría a tomar mejores decisiones.
- Sólo 22% emplea soluciones de BI para proveer información a sus usuarios
- Para 35% de los encuestados, la información proporcionada por los sistemas existentes falla en su objetivo de apoyar la toma de decisiones, y otro 32% considera que dichos sistemas no apoyan específicamente las decisiones relacionadas con retención de clientes.
- El 63% de los participantes en la investigación ha facilitado información a clientes, proveedores y socios, así como a 54% de los trabajadores móviles, pero en la mayoría de los casos lo que distribuyen son datos obsoletos e incompletos.

BI en GM México.

Con ventas anuales por 10,000 millones de dólares, la visión de General Motors (GM) México es ofrecer a las personas un portafolio de vehículos acorde con sus necesidades y estilos. Su objetivo es poner énfasis en el servicio post venta, ya que de eso dependerá la satisfacción del cliente.

Para cumplir con este objetivo, el área de Sistemas jugara un papel muy importante en la compañía, pues debe mantener una operación estable de todas las herramientas que auxilian el negocio, así como proporcionar la TI necesaria para que la subsidiaria se mantenga competitiva.

“Con ayuda de la tecnología analizamos la viabilidad de los negocios, atendemos las necesidades del mercado mexicano de manera oportuna y conocemos el estado de los almacenes para aumentar o disminuir el inventario con base en la demanda de los vehículos”, explica Fernando Durán López, director de Sistemas de General Motors México.

GM México opera bajo el esquema de outsourcing en toda la ejecución de su operación; el desarrollo e implantación de sistemas se lleva a cabo por un tercero, EDS, con la intención de que el personal del área de Sistemas se concentre exclusivamente en crear proyectos que soporten los objetivos del negocio: satisfacción de clientes, costo competitivo, por mencionar solo ejemplos.

A decir del directivo de GM, la labor de las 40 personas que laboran en esa área consiste en controlar proyectos y en desarrollar estrategias que aporten valor a la compañía; por ello, la ejecución técnica esta bajo el modelo de outsourcing, pues no ofrece un valor competitivo:

"Como nuestros recursos económicos son limitados, tenemos que priorizar las inversiones tecnológicas de acuerdo con los requerimientos del negocio; es decir, en aquellas áreas donde tendrán el mayor impacto".

Comenta que un factor clave para el éxito de la empresa es manejar dos tipos de presupuesto, el primero enfocado en reducir los costos operativos con ayuda de la tecnología, y el segundo en crear nuevos proyectos tecnológicos que aporten valor al negocio.

"Ambos presupuestos, recalca, se manejan en forma independiente para tener una eficiencia en la operación y para crear nuevos valores y herramientas en beneficio de los clientes".

En GM México la tecnología es muy importante para apoyar la operación diaria; por eso existen diferentes sistemas informáticos en cada área del corporativo y de las plantas automotrices; por ejemplo, en la parte de manufactura posee sistemas que controlan toda la producción en forma automatizada.

En el área de ingeniería se cuenta con sistemas tecnológicos para que los empleados operen en forma eficiente; es decir, existen robots que sueldan las unidades, un área automatizada de pintura, así como otra para instalar partes internas y mecánicas de las unidades.

En la parte administrativa, las herramientas informáticas controlan la operación del negocio en forma efectiva y segura; de hecho, cumple totalmente con los requerimientos de nuevas leyes de control financiero como la Sarbanes Oxley.

Inteligencia En GM. Debido a que el comportamiento del mercado mexicano es diferente de otros países en los que GM tiene presencia, la solución de inteligencia de negocios o BI es fundamental para la subsidiaria mexicana

pues, como lo asegura Fernando Durán, le permite tomar mejores decisiones en el negocio, así como atender los requerimientos de sus clientes:

"Con BI entendemos perfectamente al mercado y, de esta forma, sabemos que pasa con nuestros productos y donde hay que poner mayor énfasis".

Su estrategia de BI comenzó en 2002 a raíz de que la compañía creció en el uso de la información y, por tanto, fue necesario un software para apoyar las decisiones del negocio. El proveedor elegido para tal fin fue MicroStrategy, ya que se especializa en ofrecer soluciones de inteligencia de negocios integrales.

A decir del director de Sistemas de General Motors México, no solo las áreas administrativas utilizan BI, también las plantas paulatinamente comienzan a ocupar esa solución para cuestiones operativas y de inventarios. La idea es que en los próximos años sea utilizada por toda la organización.

Actualmente, 90% de la solución de BI está enfocada exclusivamente al área de Ventas. Su objetivo consiste en pasar de la etapa de la administración de datos a la de explotación, que es precisamente el sentido de BI.

"Mi objetivo, enfatiza Fernando Durán, es lograr una mayor participación de mercado con ayuda de BI; por ello, existe un grupo de inteligencia de negocios encargado de explotar esta herramienta al máximo".

VENTAJAS COMPETITIVAS.

De acuerdo con Fernando Durán, las ventajas competitivas que le dio la herramienta de MicroStrategy fueron tener más informados a los ejecutivos y directivos, reaccionar en forma rápida a las exigencias del mercado, conocer cada día las ventas logradas por la empresa y compartir información con los distribuidores:

"El hecho de que los distribuidores estén alineados con la estrategia tecnológica de GM México, se traduce en beneficios para ellos, ya que saben la forma en la cual se comporta el mercado".

Con el análisis de la información, GM evalúa los productos que lanzará al mercado y sabe que áreas deben ser reforzadas; incluso, planea hacer marketing dirigido para utilizar sus recursos técnicos y humanos de manera más efectiva, y así lograr un mayor impacto en los consumidores.

Mientras en años anteriores la empresa tenía datos de tiempo atrás; hoy con ayuda de BI, cuenta con información diaria para una correcta toma de decisiones.

"Decidimos implantar BI para unificar la información de la compañía; el hecho de que exista sólo una fuente de información nos permite enfocarnos en los objetivos del negocio y responder de forma más fácil a las exigencias del mercado", reitera Durán [Monografías].

BI BBVA BANCOMER.

La iniciativa privada también ha tomado ventaja del análisis de la información. El negocio de banca mayorista y de inversiones de BBVA mejoró su conocimiento y servicio al cliente mediante un desarrollo de CRM que incluyó herramientas de BI. El objetivo: contar con un conocimiento íntegro y detallado de sus clientes, de las relaciones de estos con el banco, de sus necesidades financieras y transaccionales y otros hechos relevantes.

Hoy, la institución está inmersa en su actualización. Los desarrollos realizados soportan la tecnología de *thin client* de Information Builders. La tecnología WebFocus, de esta firma, le ha permitido a BBVA reducir los tiempos de desarrollo e implantación, con una navegabilidad sencilla y atractiva y unas necesidades de mantenimiento posterior reducidas.

Entre otros beneficios están los avances en la información de gestión, la optimización de procesos, la eficacia en las ventas y en el aumento de los contactos comerciales y el aumento en los ingresos. Además, se ha mejorado el conocimiento del cliente, se afianzaron las técnicas de retención de negocio y creció el porcentaje de efectividad.

El mayor reto para BBVA fue crear conciencia a la fuerza comercial de la necesidad de alimentar el sistema con información de visitas, eventos y hechos relevantes relacionados con el cliente. La solución a estos problemas se está consiguiendo con el desarrollo de herramientas que aporten un servicio a los vendedores, así como con sistemas de incentivos basados en el convencimiento y no en la imposición [InformationWeek].

3.2. Casos de BI internacionales.

BI BBVA Horizonte.

La compañía del grupo BBVA de pensiones y cesantías, ha implementado con Gattaca S.A. una solución de Business intelligence y Balance Scorecard para la gestión de la gerencia nacional de ventas. Logrando grandes eficiencias con los indicadores de gestión, alertas y seguimiento a metas de ventas. Esta herramienta fue implementada bajo un esquema de usuarios ilimitados que permite la visualización y seguimiento a indicadores de gestión en las diferentes áreas de la compañía.

Adicionalmente, a través de la transferencia de conocimiento y las herramientas de diseño, BBVA Horizonte está actualmente expandiendo el uso de esta herramienta a otras áreas que requieren de la gestión soportada en indicadores [Gattaca].

BI de Capel.

Aplicación de soluciones móviles como apoyo a la Inteligencia de Negocios. Capel, la mayor compañía productora de pisco en Chile, adoptó recientemente una solución inalámbrica para ventas, recaudación y entrega en terreno, que en conjunto con el proyecto Inteligencia de Negocios permite procesar grandes volúmenes de datos para analizarlos y transformarlos en valor. Gracias a estas dos modificaciones tecnológicas, hoy puede monitorear a diario todas las variables claves del negocio, destinando más tiempo a actividades que generen valor para sus clientes [Ingeniería Industrial Chile].

TESIS TESIS TESIS TESIS TESIS



TEORIAS BASE Y
TECNOLOGIAS UTILIZADAS

TESIS TESIS TESIS TESIS TESIS

4. TEORÍAS BASE Y TECNOLOGÍAS UTILIZADAS.

4.1. Teorías base.

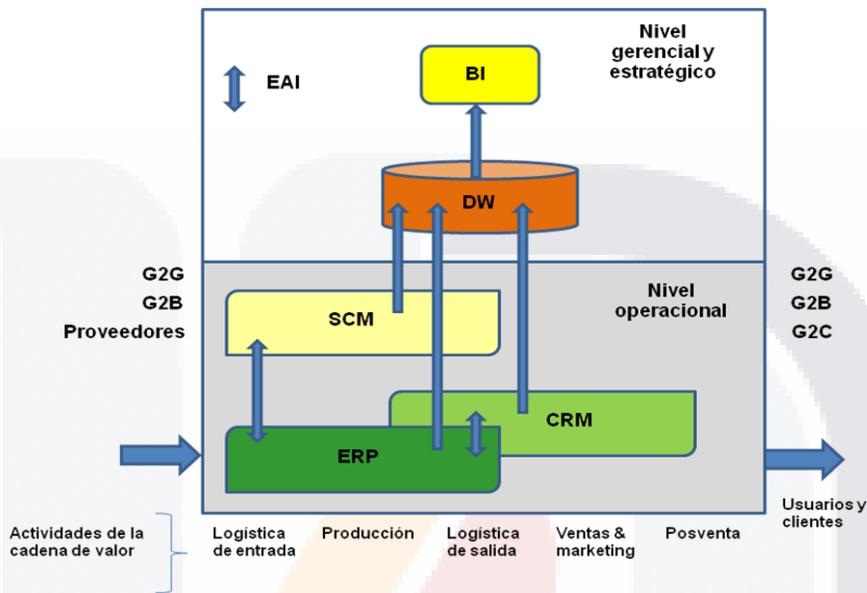


Figura 7. Tecnologías que apoyan el proceso de toma de decisiones a través del BI.

4.1.1. Inteligencia de Negocios. BI por sus siglas en inglés (Business Intelligence). El análisis de datos en las empresas, ha sido desde el inicio de éstas, una práctica cotidiana, la llegada de herramientas tecnológicas para el soporte a la toma de decisiones ha facilitado ésta tarea a los niveles intermedios y estratégicos de la organización, así es como nace el concepto de BI.

Así pues BI es la habilidad de consolidar información y analizarla con la suficiente velocidad y precisión para descubrir ventajas y tomar mejores decisiones de negocios. Definición compatible con la necesidad actual de los negocios que ante la presión de ser cada día más competitivos, para mantenerse tienen la doble tarea no sólo de permanecer sino de ser lucrativos (Cano C., 1999).

En BI, la "facilidad de uso" es un término relativo. Un informe mensual del usuario es otro proyecto de análisis sobre las tendencias de la persona. Más probablemente los usuarios tienen necesidades individuales de acceso a los datos y su forma de presentación. La comprensión y finalmente la clasificación de estas necesidades pueden disparar la redefinición de los requerimientos del negocio, el refinamiento y enriquecimiento de los datos, y la selección de una tecnología mucho más fácil de usar. En la siguiente figura se muestra una clasificación de usuarios según sus requerimientos de información.

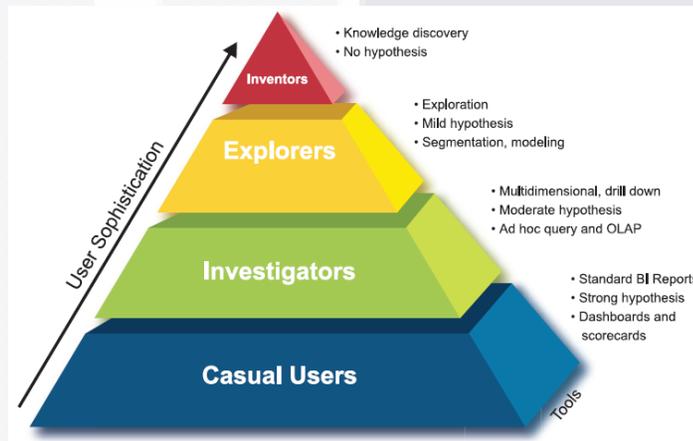


Figura 8. Clasificación de los usuarios de BI.

La siguiente pirámide representa la evolución del BI propuesta para compañías con una estructura vertical [7].

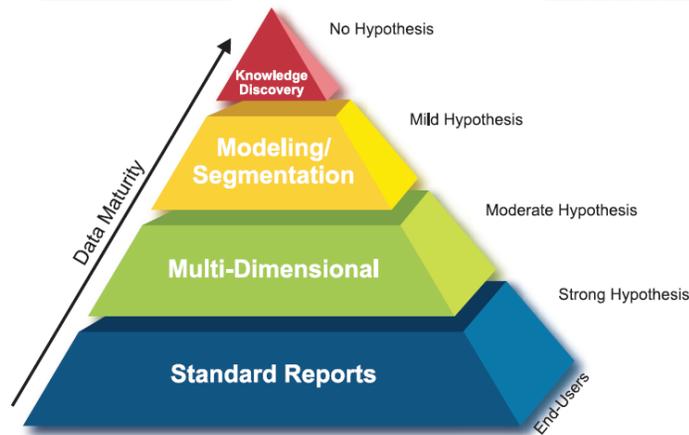


Figura 9. Evolución del BI.

BI se puede definir como el proceso de analizar los bienes o datos acumulados en la empresa y extraer una cierta inteligencia o conocimiento de ellos. Dentro de la categoría de bienes se incluyen las bases de datos de clientes, información de la cadena de suministro, ventas personales y cualquier actividad de marketing, estructura o fuente de información relevante para la empresa, por lo que resultan importantes los siguientes términos:

4.1.1.1. La cadena de suministros. SCM. Por sus siglas en inglés Supply Chain Management, se ha definido como la integración, desde el consumidor hasta los primeros proveedores, de los procesos de negocio clave que proporcionan los productos, servicios e información que añaden valor a los clientes y accionistas (Lambert, Cooper & Pagh, 1998).

4.1.1.2. CRM. Responde a las siglas “Customer Relationship Management” o “Gestión de las relaciones con los clientes”. CRM es toda estrategia empresarial que implica un cambio de modelo de negocio centrado en la gestión automatizada de todos los puntos de contacto con el cliente, cuyo objetivo es captar, fidelizar y rentabilizar al cliente ofreciéndole una misma cara mediante el análisis centralizado de sus datos. (Ignacio García Valcárcel, 2001).

4.1.2. Datawarehouse (DW). Las bases de datos de una organización almacenan información de todo tipo y en grandes cantidades. Sin embargo, esta información sin ninguna herramienta que facilite el análisis de los datos y la toma de decisiones ejecutivas tendría escaso valor. Por lo tanto son necesarios mecanismos que permitan extraer la información relevante para el proceso de la toma de decisiones. Estos mecanismos se conocen con el nombre genérico de DW.

Un DW es una copia de datos transaccionales específicamente estructurada para la consulta y el análisis (Ralph Kimball, 2002).

Con la implementación de un DW se obtienen las siguientes ventajas (José Manuel Huidobro, Rubén de la Peña, David Roldan & José Antonio de la Fuente, 2005):

- Representación visual de la estrategia de negocio y transformación de esta en acción.
- Comunicación de los objetivos de negocio a toda la organización
- Ajuste dinámico de la estrategia de negocio para dar respuesta a los cambios del mercado.
- Control del rendimiento de los principales factores de éxito utilizando benchmarks internos y externos.
- Análisis y optimización del rendimiento basándose en el modelo de negocio unificado.

4.1.3. Tableros de control o Dashboards. "No puedes dirigir lo que no puedes medir."



Figura 10. Muestras de dashboards.

Las personas dentro de una compañía necesitan acceso a una página que despliegue aspectos clave de operación, características financieras, etc., de las cuales son responsables. Estas páginas son conocidas como "Dashboards" porque muestran un vistazo de los datos que son más importantes (Michael Hugos, 2006).

Cuando los datos complejos son presentados en un formato visual sencillo a nivel sumarial, se obtiene en una sola vista las tendencias y relaciones más significativas en los datos del desempeño del negocio.

Gráficos de análisis altamente intuitivos permiten entender rápidamente las actividades del negocio para así tomar decisiones efectivamente [Cognos].

4.1.4. ERP. Enterprise Resource Planning o Planeación de Recursos de la Empresa. Es un sistema integral de gestión empresarial que está diseñado para modelar y automatizar la mayoría de procesos en la empresa (área de finanzas, comercial, logística, producción, etc.). Su misión es facilitar la planificación de todos los recursos de la empresa. El ERP es un conjunto de paquetes de sistemas configurables de información dentro de los cuales se integra la información a través de áreas funcionales de la organización. Los sistemas ERP son extremadamente costosos, y una vez que los sistemas ERP se implantan con éxito trae una serie de beneficios importantes para las empresas (Kumar & Hillengersberg, 2000).

4.2. Tecnologías utilizadas.

4.2.1. Pentaho. La plataforma Open Source Pentaho Business Intelligence cubre ampliamente las necesidades de análisis de los datos y de los informes empresariales.

Las soluciones de Pentaho están escritas en Java y tienen un ambiente de implementación también basado en Java. Eso hace que Pentaho es una solución muy flexible para cubrir una amplia gama de necesidades empresariales tanto las típicas como las sofisticadas y específicas al negocio.

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

Las soluciones que Pentaho pretende ofrecer se componen fundamentalmente de una infraestructura de herramientas de análisis e informes, integrado con un motor de workflow de procesos de negocio. La plataforma será capaz de ejecutar las reglas de negocio necesarias, expresadas en forma de procesos y actividades y de presentar y entregar la información adecuada en el momento adecuado [Todo BI].

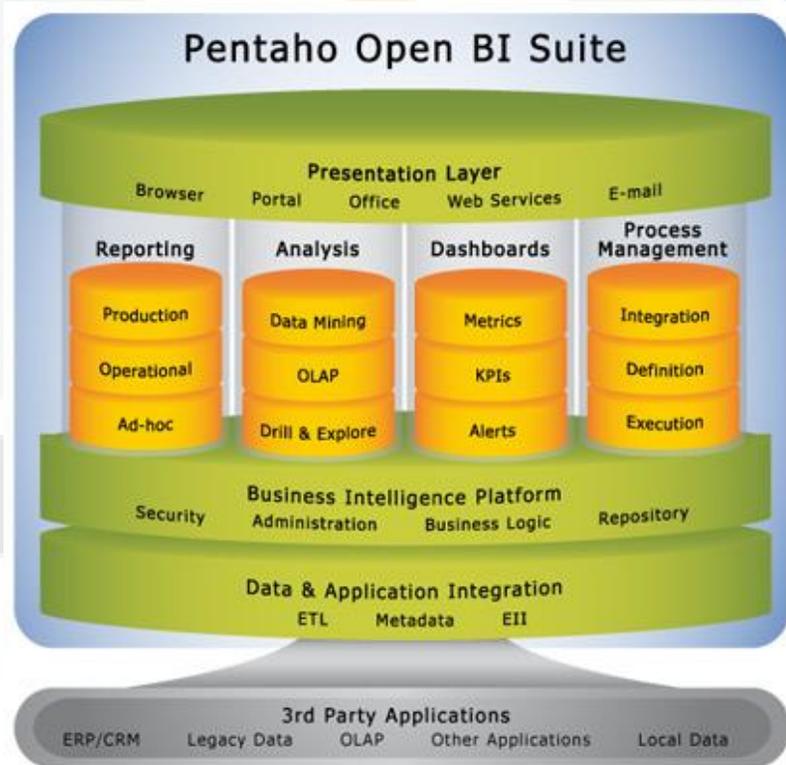


Figura 11. Pentaho Open BI suite.

Los módulos de la plataforma o suite Pentaho BI son:

1. Reporting - un modulo de los informes ofrece la solución adecuada a las necesidades de los usuarios. Pentaho Reporting es una solución basada en el proyecto JFreeReport y permite generar informes ágil y de gran capacidad. Pentaho Reporting permite la distribución de los resultados del análisis en múltiples formatos - todos los informes incluyen la opción de imprimir o exportar a formato PDF, XLS, HTML y texto. Los reportes Pentaho permiten también programación de tareas y ejecución automática de informes con una determinada periodicidad.
2. Análisis - Pentaho Análisis suministra a los usuarios un sistema avanzado de análisis de información. Con uso de las tablas dinámicas (pivot tables, crosstabs), generadas por Mondrian y JPivot, el usuario puede navegar por los datos, ajustando la visión de los datos, los filtros de visualización, añadiendo o quitando los campos de agregación. Los datos pueden ser representados en una forma de SVG o Flash, los dashboards widgets, o también integrados con los sistemas de minería de datos y los portales web (portlets).

Data Mining - análisis en Pentaho se realiza con la herramienta WeKa.

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

4. Administración del proceso. Que incluye la integración, definición y ejecución de la información y la forma en que habrá de tratarse. Este apartado suele hacer uso del proyecto Kettle (Pentaho Data Integration) que permite implementar los procesos ETL (Extract, Transform and Load) o Extracción, Transformación y Carga.

Kettle es un proyecto belga que incluye un conjunto de herramientas para realizar ETL. Uno de sus objetivos es que el proyecto ETL sea fácil de generar, mantener y desplegar.

Se compone de 4 herramientas:

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

4.2.2. Java. Es un lenguaje de programación orientado a objetos desarrollado por Sun Microsystems a principios de los años 90. El lenguaje en sí mismo toma mucha de su sintaxis de C y C++, pero tiene un modelo de objetos más simple y elimina herramientas de bajo nivel, que suelen inducir a muchos errores, como la manipulación directa de punteros o memoria.

Las aplicaciones Java están típicamente compiladas en un *bytecode*, aunque la compilación en código máquina nativo también es posible. En el tiempo de ejecución, el *bytecode* es normalmente interpretado o compilado a código nativo para la ejecución, aunque la ejecución directa por hardware del *bytecode* por un procesador Java también es posible [Wiki Java].

Java es (Douglas Bell, Doug Bell, Mike Parr, Alonso Vidal & Guillermo Levine, 2003):

1. Orientado a objetos. Java está completamente orientado a objetos. No es un lenguaje al que se le haya agregado esta orientación después de haber sido creado.
2. Compatible con internet. La principal motivación de la creación de java fue permitir que se desarrollen programas que utilicen Internet y Web. Los programas en java pueden invocarse fácilmente desde los diferentes exploradores; además, pueden transmitirse a través de internet y ejecutarse en cualquier computadora.
3. De propósito general. Aunque está diseñado para crear aplicaciones para la World Wide Web (www), cualquier cosa que otro lenguaje de programación puede hacer, también puede hacerla java.
4. Independiente de la plataforma. Los programas en Java pueden ejecutarse en casi todas las computadoras ¡Sin necesidad de modificarlos! A esto agregue el concepto “escríbalo solo una vez, ejecútelo en cualquier lugar”.
5. Robusto. Si un programa java falla (y realmente todos los programas tienen esa tendencia), no provocara destrozos, daños ni incertidumbre. Como los programas java se ejecutan dentro de una “jaula” de protección, los efectos de cualquier error están confinados y controlados, incluso están protegidos contra la filtración de virus.

4.2.3. Mondrian. Es un servidor OLAP escrito en java. Mondrian habilita interactuar con conjuntos de datos voluminosos almacenados en una BD sin escribir sentencias de SQL [Mondrian].

Mondrian es usado para:

1. Mejorar el rendimiento, análisis interactivo de pequeños o grandes volúmenes de información.
2. Exploración "Dimensional" de los datos, por ejemplo, el análisis de ventas por línea de producto, por región, por período de tiempo.
3. Análisis sintáctico de Multi-Dimensional eXpression (MDX). Estructurado en lenguaje de consultas (SQL) para obtener respuestas a las preguntas sobre las dimensiones.
4. Alta velocidad de las consultas mediante el uso de tablas agregadas en el RDBMS.
5. Cálculos avanzados utilizando las expresiones de cálculo del lenguaje MDX.

4.2.4. JPivot. Es una etiqueta JSP personalizada que permite a los usuarios realizar la típica navegación OLAP como slice and dice, dril down, dril up y roll up en una tabla OLAP.



Dimension Espacial	Imágenes				
	Todas las imágenes	Consultadas	Descargadas	Bin Publicar	Visualizadas
-Todas las Regiones	240,600,080	62,264,264	62,729,536	62,166,820	61,519,452
-Colombia	208,965,840	52,723,131	52,727,376	51,959,571	51,555,762
+AMAZONAS	1,736,147	461,522	390,696	492,590	391,339
+ANTIOQUIA	19,313,915	4,874,155	4,957,819	4,772,861	4,709,080
+ARAUCA	1,308,675	311,118	355,838	330,951	310,768
+ATLANTICO	3,306,338	825,401	836,741	864,344	779,852
+BOGOTA D.C.	157,037	48,603	39,503	28,733	40,198
+BOLIVAR	7,740,374	1,876,493	1,948,412	1,922,650	1,992,819
+BOYACA	20,457,802	4,827,376	5,350,075	5,208,642	5,071,709
+CALDAS	5,124,794	1,346,469	1,391,377	1,268,933	1,188,015
+CAQUETA	2,260,525	561,044	619,439	554,560	525,482

Figura 12. Muestra de un cubo visualizado a través de JPivot.

JPivot utiliza Mondrian como su servidor OLAP y soporta también XMLA como datasource [JPivot].

4.2.5. Jfreechart.



Figura 13. Muestra de tipos de gráficos generados por JFreeChart.

Es código 100% java, es una librería para gráficos que hace fácil a los desarrolladores el desplegar gráficos de calidad en sus aplicaciones.

El set de JFreeChart incluye [JFreeChart]:

1. Una API bien documentada, soportando una gran variedad de gráficos.
2. Diseño flexible y fácil de extender, ya sea en aplicaciones del lado del servidor o del lado del cliente.
3. Soporta muchos tipos de salidas, incluyendo componentes swing, archivos de imágenes (incluyendo PNG y JPEG), formatos de vectores gráficos (incluyendo PDF, EPS y SVG)
4. JFreeChart es “Open source”, más específicamente, software (SW) libre.

4.2.6. XML. Extensible Markup Language o Lenguaje de Marcado Extensible. Contiene formas, etiquetas, estructuras, y protege la información. Lo hace con símbolos incrustados en el texto, llamados marcas. Las marcas aumentan el sentido de la información en algunos aspectos, identificando sus partes y cómo se relacionan entre sí (Erik T. Ray, 2003).

XML es muy simple, pero estricto que juega un papel fundamental en el intercambio de una gran variedad de datos. Es un lenguaje muy similar a HTML pero su función principal es describir datos y no mostrarlos como es el caso de HTML. XML es un formato que permite la lectura de datos a través de diferentes aplicaciones [WC3].

Las tecnologías XML son un conjunto de módulos que ofrecen servicios útiles a las demandas más frecuentes por parte de los usuarios. XML sirve para estructurar, almacenar e intercambiar información.

4.2.7. AJAX JavaScript Asíncrono y XML. Ajax es una nueva etiqueta para describir la rica experiencia de una aplicación de escritorio corriendo sobre un explorador web estándar, sin requerir ningún plug-in en especial (Michael Mahemoff, 2006).

Es un término nuevo para describir dos capacidades de los navegadores que han estado presentes por años, pero que habían sido ignoradas por muchos desarrolladores Web.

Las dos capacidades en cuestión son:

1. La posibilidad de hacer peticiones al servidor sin tener que volver a cargar la página.
2. La posibilidad de analizar y trabajar con documentos XML.

AJAX no es una tecnología, sino la unión de varias tecnologías, incorpora (Michael Mahemoff, 2006):

- ❑ Presentación basada en estándares usando XHTML y CSS
- ❑ Exhibición e interacción dinámicas usando el Document Object Model (DOM), para modificar dinámicamente los documentos.
- ❑ Intercambio y manipulación de datos usando XML y XSLT
- ❑ Recuperación de datos asincrónica usando XMLHttpRequest
- ❑ JavaScript poniendo todo junto.

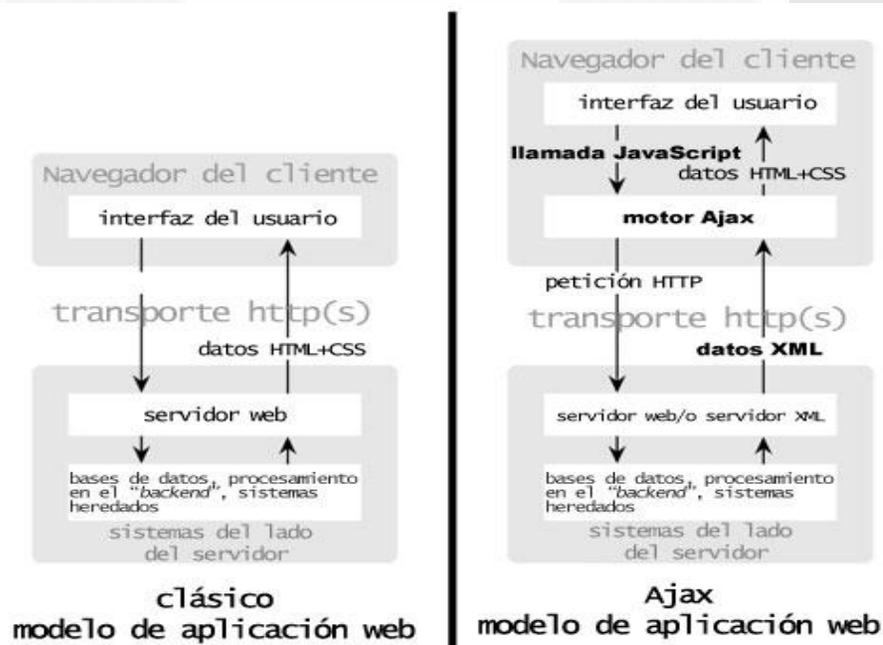
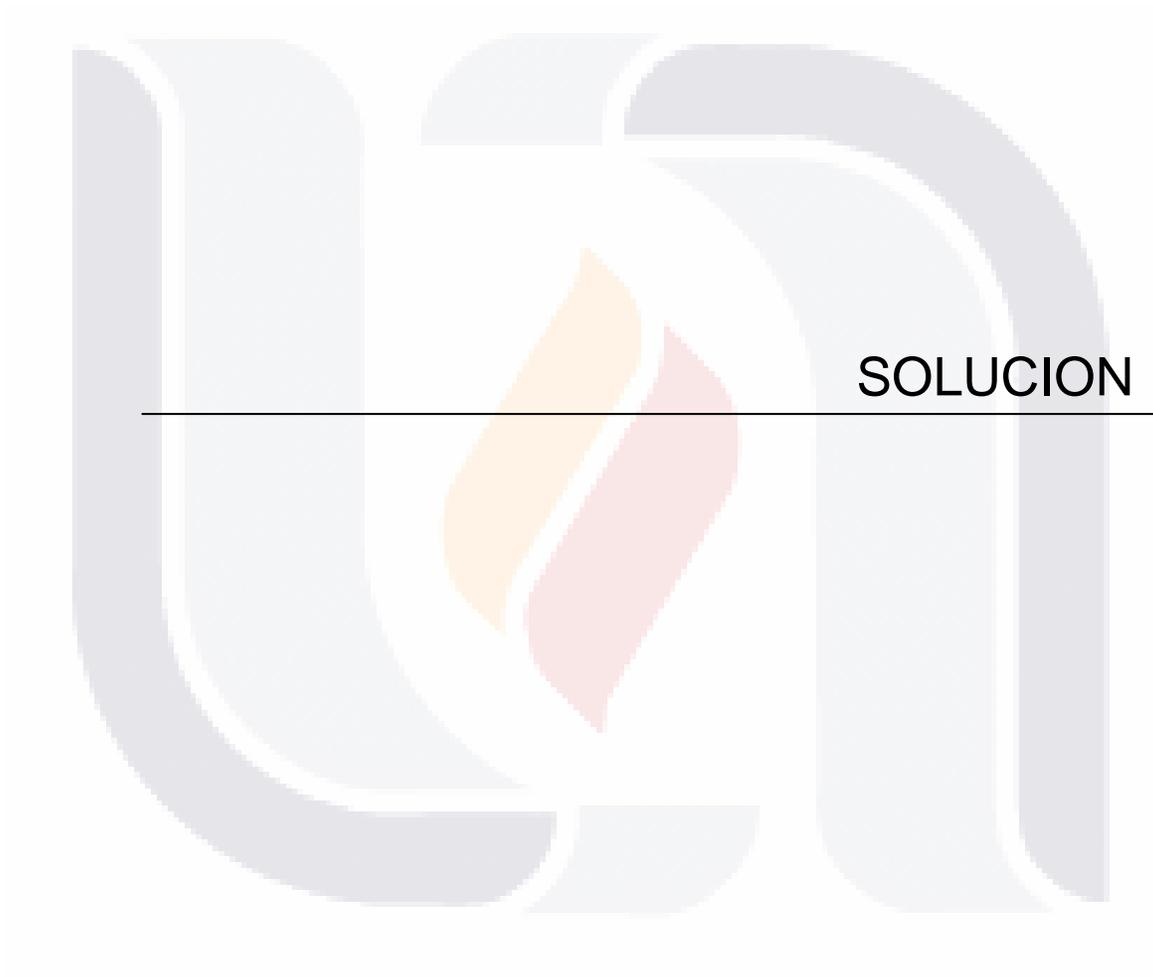


Figura 14. Esquema de aplicaciones web. Clásico vs. Ajax.



SOLUCION

5. DESCRIPCIÓN DE LA SOLUCIÓN.

La documentación de este caso práctico tiene como finalidad mostrar las ventajas del uso de la plataforma open source de Pentaho, por lo que en el siguiente diagrama de arquitectura, se muestra la relación entre los componentes del servidor de BI de pentaho y las interfaces con el mundo externo, en donde se puede identificar los servicios a los que el usuario podrá acceder a efecto de hacerse de información relevante para sus actividades cotidianas referente a la toma de decisiones.

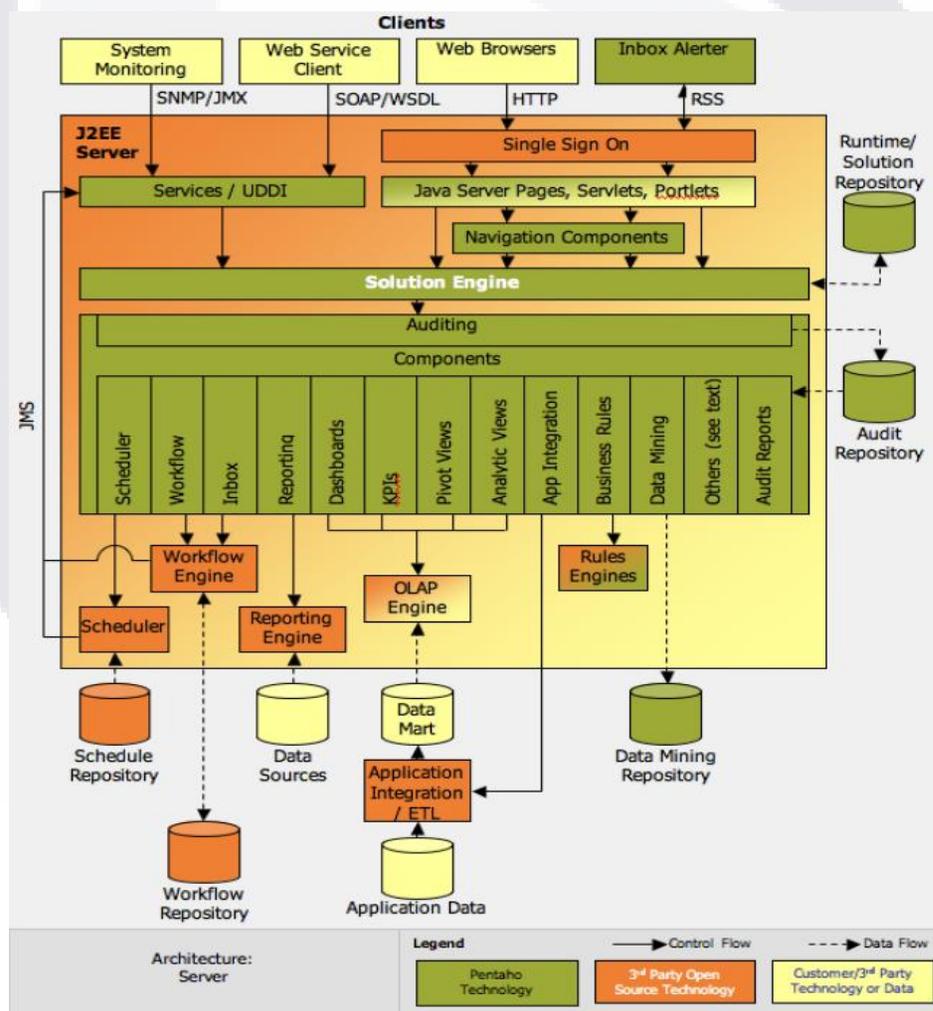


Figura 15. Arquitectura de la plataforma Pentaho.

El corazón del servidor es el Solution Engine, ya que es el punto principal de actividad en la plataforma, y se encuentra entre el mundo externo (clientes web, servicios, sistemas de monitoreo, etc.) y la capa de componentes. Solicitando tareas y redireccionandolas al o los componentes apropiados para su ejecución.

La investigación y transformación que se documentan en este caso práctico abarca solamente algunos de los aspectos de la plataforma, como son: Interacción con el cliente a través de JSP a través de exploradores web, dashboards, Pivot, accesos a la BD.

Ventajas para los usuarios finales:

1. Menor tiempo de respuesta en la extracción de información de la BD.
2. Diferentes vistas y versiones de la información.
3. Representaciones gráficas con mayor legibilidad a la que un reporte textual puede ofrecer.
4. Síntesis y agregados de la información mediante cubos.
5. Generación de reportes parametrizados.

Cabe mencionar que para el funcionamiento de la plataforma Pentaho debe contarse con el siguiente software (SW) instalado:

- Ant previamente instalado en el sistema y agregado a la variable de entorno PATH.
- JDK
- JRE.

El proceso de instalación de los mismos son pasos sencillos y se obviarán en esta documentación.

5.1. Transformación de la plataforma.

La plataforma Open Source Pentaho BI es la base para la implementación de BI explotando la información del SIA; sin embargo está diseñada para trabajar con el servidor de aplicaciones JBoss, a efecto de compatibilidad con las aplicaciones del Instituto se requirió configurar la plataforma para montarla en un servidor Oracle Application Server (OAS).

En el apéndice A se explica con mayor detalle el proceso de transformación que sufrió la suite.

5.2. Cambios a la configuración del servidor OC4J.

Hasta el momento, tenemos la versión que Pentaho proporciona como demo, montada sobre nuestro servidor OAS. Sin embargo, el trabajo de configuración no termina, para comenzar a explotar nuestro propio DW, deben realizarse conexiones a la BD Oracle en donde se encuentra la información del SIA.

En el apéndice B, se detalla puntualmente cada cambio necesario en la configuración del OC4J.

5.3. Deploy del archivo Pentaho .ear.

Finalmente, se puede hacer el deploy del archivo .ear obtenido en el apéndice A, para lo que utilizaremos el IDE JDeveloper previamente configurado siguiendo los pasos del apéndice B.

En el apéndice C, se muestra paso a paso como hacer el deploy del archivo.

5.4. Ejecución general de una tarea en la plataforma Pentaho.

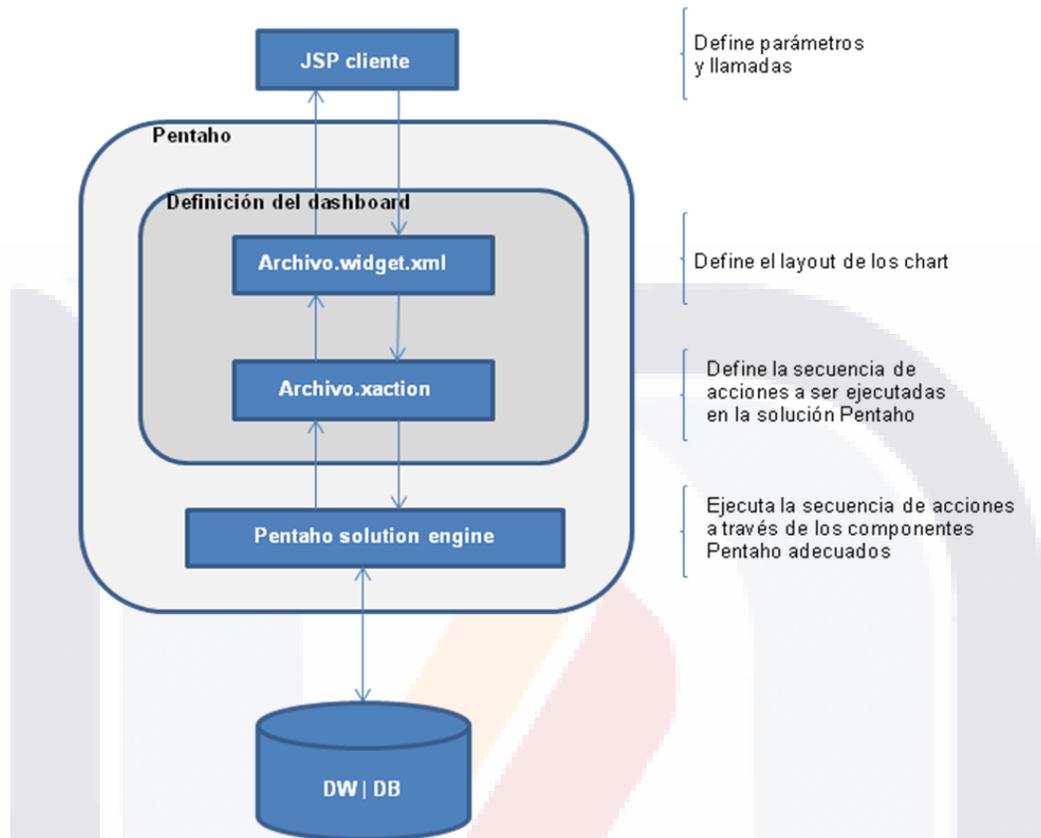


Figura 32. Ejecución general de una tarea en la plataforma Pentaho.

Donde:

- **JSP cliente.** Es la pagina en que el cliente parametriza sus búsquedas o elige el dashboard que desea visualizar.
- **Archivo widget.** Es un archive xml en el que se definen las características particulares de los charts que se embeberán en el dashboard.
- **Archivo xaction.** Es un document xml que define completamente las pequeñas tareas que el solution engine debe realizar. Este es un proceso ligero que define el orden de ejecución de uno o más componentes de la plataforma BI de Pentaho, y tiene la capacidad de

utilizar un result set o bien llamar otro archivo de secuencias y ejecutar condicionalmente otros componentes. Este documento debe tener la extensión “.xaction”.

- **Pentaho Solution Engine.** Es el motor de la plataforma que se encarga de realizar las tareas solicitadas.
- **DW| DB.** Repositorio de la información a explotarse a través de la aplicación de BI.

5.5. Desarrollo de tableros de control utilizando la información del SIA.

Para la explotación de la información del sistema transaccional es necesario definir la forma en que se mostrará la información. Para ello pueden diseñarse esquemas con agregados, que faciliten la extracción al llevarse a cabo sobre un grupo bien identificado de información relevante.

A efecto de mostrar las características de BI se documenta:

- A. Diseño de un diagrama de estrella.
- B. Vínculo de pentaho con la API de google maps.
- C. Funcionamiento de un dashboard.
- D. Funcionamiento de JPivot.

Puntos que tienen como finalidad cumplir con los objetivos planteados en el caso de estudio que se documenta.

5.5.1. Diagrama de estrella.

Este diagrama se diseñó a efecto de cumplir con requerimientos constantes de 2 principales operaciones del SIA, que son: consultas a plantilla, involucrando plazas, total de plantilla ocupada, plantilla vacante y el costo de las mismas, pudiendo identificarse por Unidad Ejecutora en diferentes intervalos de tiempo.

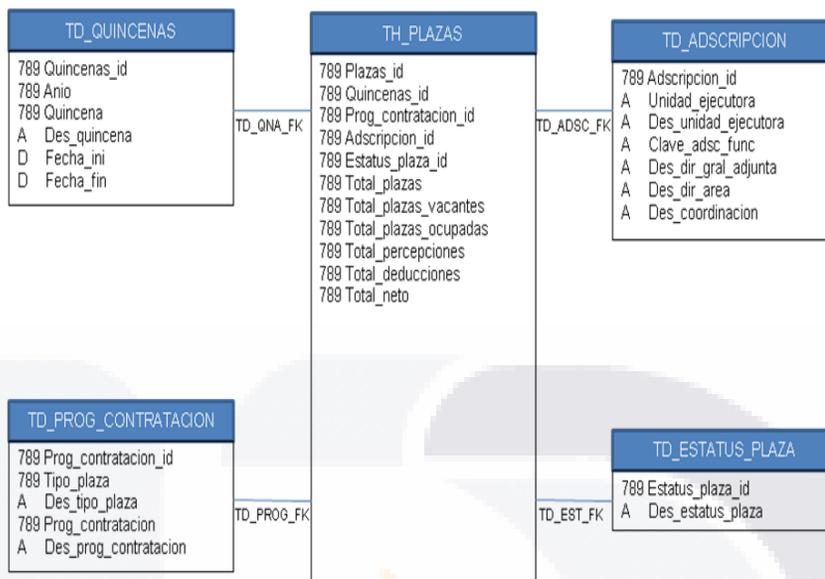


Figura 33. DW. Diagrama de estrella con información del SIA.

5.5.2. Muestra de plazas ocupadas por Regional a través de un tablero de control integrando google maps.

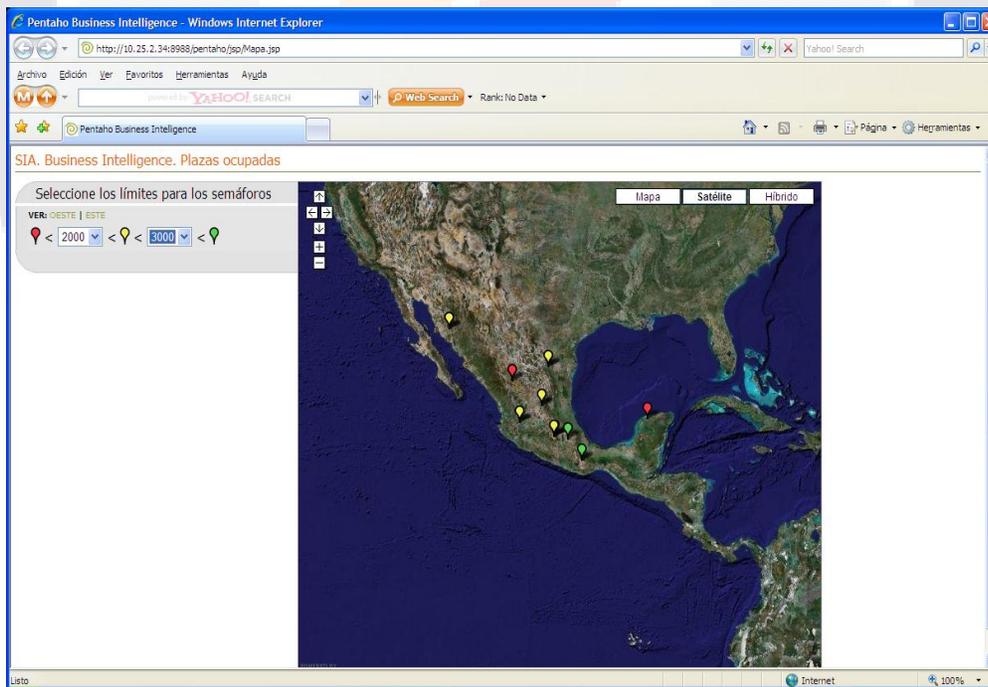


Figura 34. Regionales semaforizadas en base a parámetros.

La plataforma de Pentaho permite la inclusión de varios proyectos, uno de estos es el uso de la API de Google Maps. Debido a que INEGI cuenta con presencia a nivel nacional, es de gran utilidad identificar rápidamente el estado en el que se encuentran las diferentes variables de interés, por lo que se implemento el uso de google maps para realizar una semaforización a nivel regional para verificar el seguimiento del número de plazas ocupadas en las diferentes regionales con que cuenta el Instituto.

5.5.3. Funcionamiento de un dashboard.

SIA. Business Intelligence. Plazas ocupadas

Seleccione los límites para los semáforos

VER: OESTE | ESTE

 < <  < < 

Figura 35. Controles para actualizar los semáforos.

A través de los controles para actualizar los semaforos se pueden definir los rangos de valores con el que se actualizaran los globos que representan cada una de las regionales. Esta actualización se realiza por medio de llamadas asincronas al servidor utilizando Ajax. Cada vez que cambia un valor dentro de los controles se realiza un barrido al datasource que contiene el numero de plazas ocupadas por regional y se determina si debe actualizarse el color del globo que le corresponde, de forma que este muestre la selección del usuario.

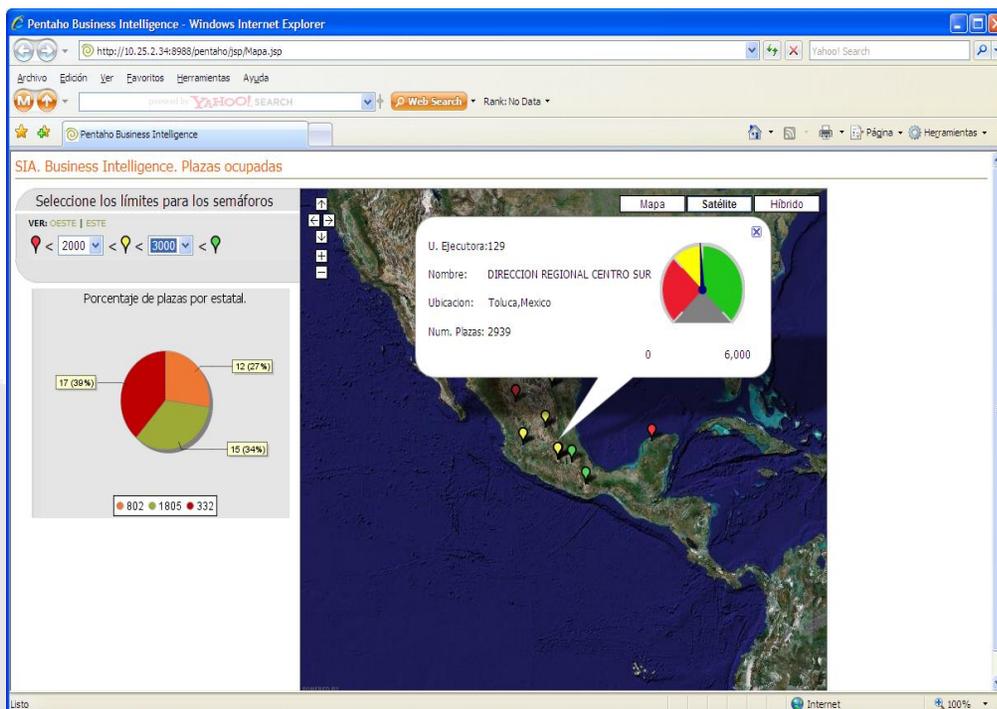


Figura 36. Desglose del semáforo seleccionado.

Una vez que el usuario da un clic sobre el globo de cualquier regional para ver su información, se dispara la ejecución de archivos .xaction, cada uno de ellos mostrando diferentes datos, tal como puede observarse en la imagen anterior.



Figura 37. Pop up. Identificación de la Regional seleccionada.

El pop up de la figura 37, proporciona datos generales de la regional, tales como: Nombre, ubicación física, total de número de plazas ocupadas y un dial que está directamente relacionado con los controles de los semáforos; este dial se considero de 0 a 6000 plazas puesto que es un rango razonable de acuerdo a las plazas que puede manejar el Instituto en cada una de sus regionales.

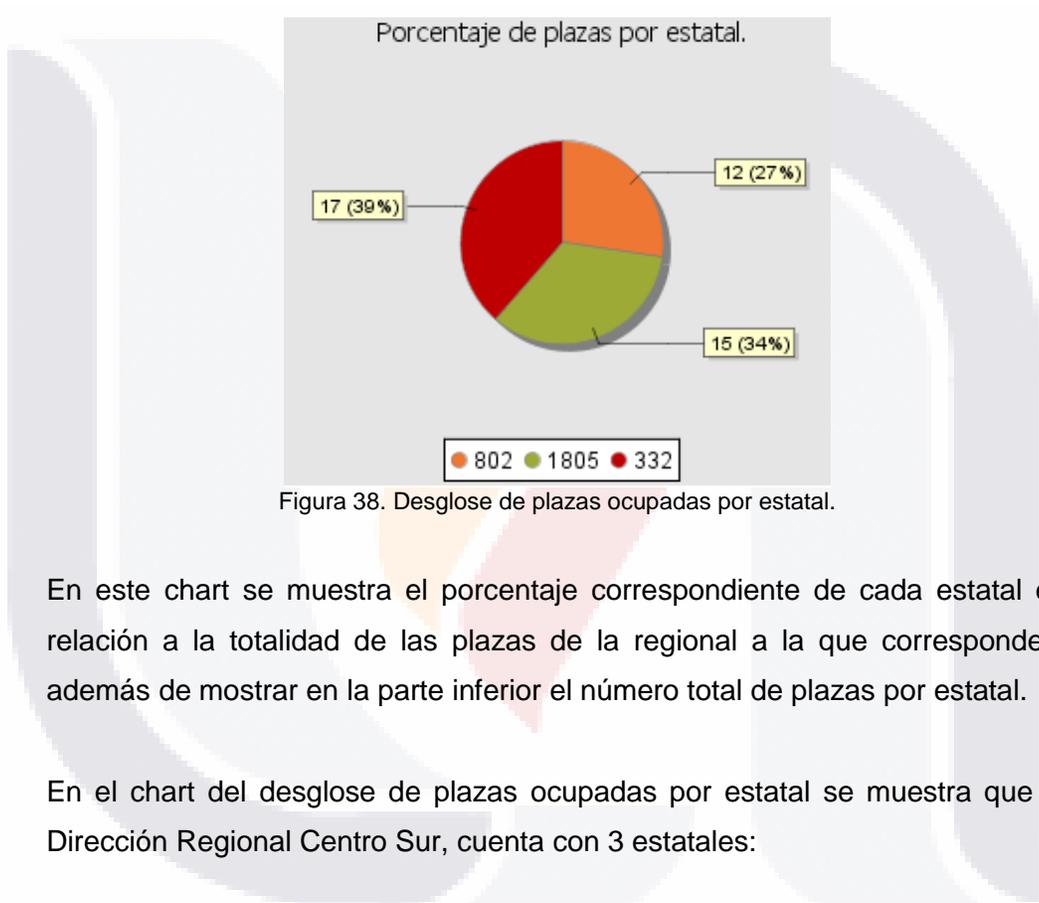


Figura 38. Desglose de plazas ocupadas por estatal.

En este chart se muestra el porcentaje correspondiente de cada estatal en relación a la totalidad de las plazas de la regional a la que corresponden, además de mostrar en la parte inferior el número total de plazas por estatal.

En el chart del desglose de plazas ocupadas por estatal se muestra que la Dirección Regional Centro Sur, cuenta con 3 estatales:

CLAVE	NOMBRE	%EN RELACION A LA REGIONAL	PLAZAS OCUPADAS
12	Estado de México	27	802
15	Guerrero	34	1805
17	Morelos	39	332
N° de plazas en la regional			2939

Tabla 3. Resultados obtenidos a través del chart de BI sobre plazas ocupadas.

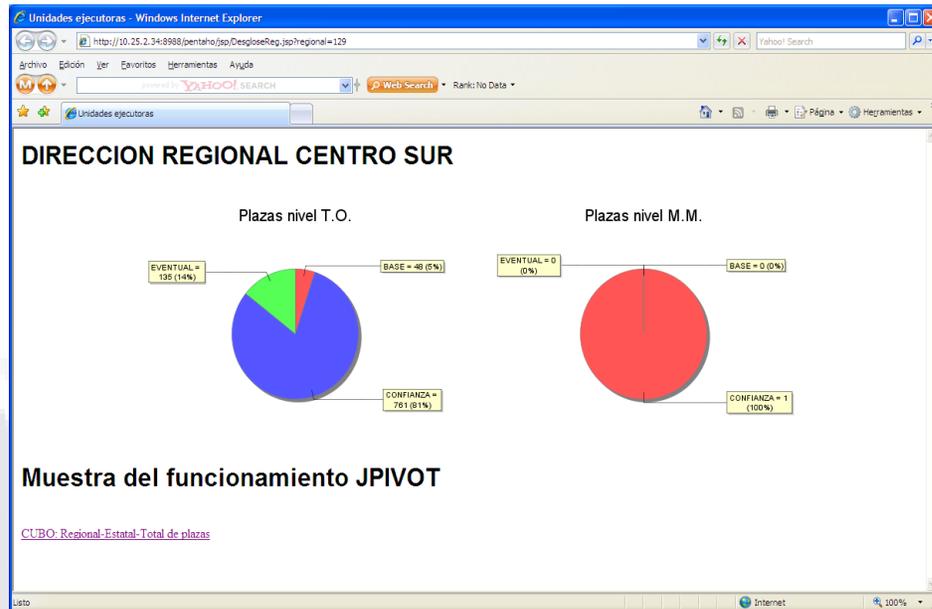


Figura 39. Pop up. Dashboard de la regional seleccionada.

En esta ventana que se lanza a la vez que se generan el pop up y el chart anterior, y en ella se puede observar a que niveles de mando pertenecen las plazas ocupadas para la regional previamente seleccionada.

Donde:

- T.O. Tecnico operativo.
- M.M Mando medio.

5.5.4. Funcionamiento de JPivot.

La información puede ser desplegada con la variedad de estilos como sea necesaria. A continuación veremos el concentrado de plazas por regional-estatal a través de un cubo.

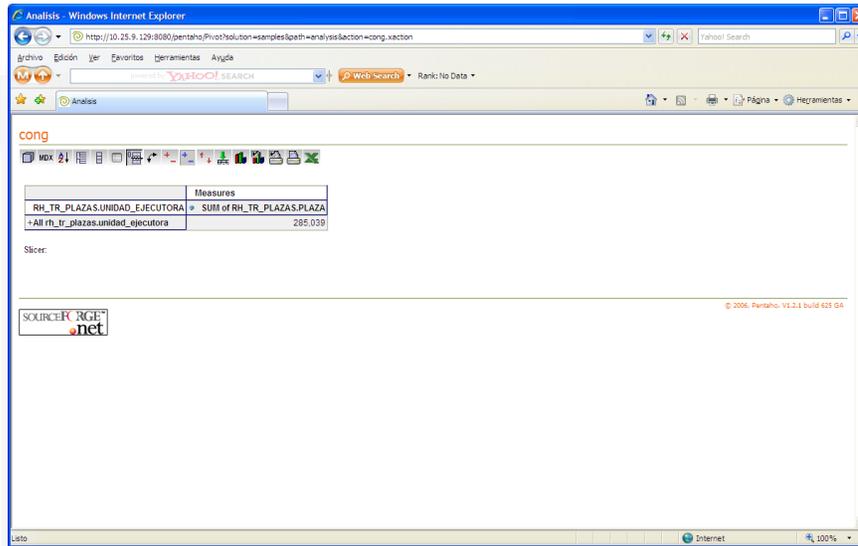


Figura 40. Cubo. Plazas por Regional-Estatal.

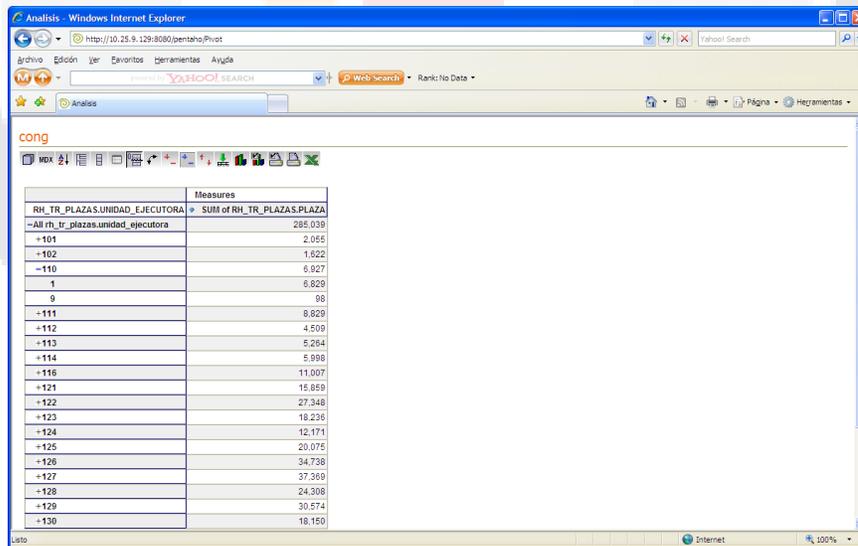


Figura 41. Desglose del cubo: Plazas por Regional-Estatal.

De esta forma el usuario al dar un solo clic sobre el gráfico puede obtener información de relevancia, y que puede ser útil para la nueva asignación de recursos financieros, materiales, incrementos y todas aquellas decisiones que contemplen como parámetro el número de plazas.





CONCLUSIONES

6. CONCLUSIONES.

Pese a que INEGI cuenta con el poder adquisitivo como para obtener un producto de marcas como COGNOS, MicroStrategy, etc., puede y debe aprovechar las características que ofrecen los productos Open Source a través de los grupos de desarrollo con que cuenta el Instituto y de esta manera fomentar la investigación en sus integrantes.

A lo largo de este documento se identifican claramente las ventajas del uso adecuado de la tecnología, misma que sin ser necesariamente costosa puede ser aprovechada para satisfacer las diversas necesidades de información en cualquier nivel de la estructura del Instituto.

La implementación de la solución BI documentada cumple con los objetivos planteados en el capítulo 3. Describiéndose a continuación de manera puntual cada uno de ellos.

Objetivos generales.

- Contar con un sistema informático integral que conforme una base de datos única como resultado de las actividades relacionadas con el manejo de personal, recursos materiales y financieros, a partir del cual se disponga de una herramienta para la toma de decisiones a fin de lograr una mayor eficiencia en el manejo de los recursos del Instituto. Los diferentes sistemas administrativos, particularmente el SIA el cual provee la información más relevante para la toma de decisiones en el Instituto, hacen uso de una BD Oracle, y esta controla la integridad y los diferentes perfiles de acceso a los datos almacenados. Dentro de esta BD se almacenan también las diferentes estructuras de datos utilizados para la implementación de BI, logrando así tener un solo repositorio que se toma como base al momento de tomar las

decisiones encaminadas a dirigir las diferentes actividades que coadyuvan al cumplimiento de los objetivos, metas y visión del Instituto.

- Proporcionar información ejecutiva para una efectiva y rápida toma de decisiones. Al contar con un solo punto de acceso y a través de la implementación de un modulo de BI dentro del SIA, el ejecutivo requiere hacer menos trámites administrativos para hacerse de la información que necesita, adicionalmente el modulo de BI esta desarrollado para ser accedido de manera remota por medio de cualquier explorador web, por lo que los tiempos de respuesta disminuyen considerablemente.

Objetivos específicos.

- Contar con un sistema de toma de decisiones que explote los datos transaccionales del SIA en las actividades relacionadas con el manejo de los diferentes módulos que lo componen. El presente caso práctico ejemplifica como se puede explotar concentrados de información que generen un conocimiento legible y fácilmente interpretable, basándose en datos capturados a través de los sistemas transaccionales.

Los casos muestra de una toma de decisiones apoyada en el desarrollo de una aplicación de BI se centraron en el número de plazas ocupadas por regional; tipos de plazas con que cuenta identificándolas por niveles de mando, así como el desglose de las plazas por estatales. Mostrando con ello las ventajas de una aplicación de BI sobre el modulo de Recursos Humanos, ventajas que pueden fácilmente ampliarse para abarcar aspectos financieros, materiales y cualquier otro modulo que requiera realizar una extracción rápida y concisa de información.

- Generar un modelo de datawarehouse (DW) con sus respectivas dimensiones y datamarts (DM). A través de diagramas de estrella, como el que se muestra en la figura 33, se muestra como una forma sencilla de esquematizar la información proporciona grandes beneficios al momento de hablar de BI, con dicha estrella se puede obtener datos relevantes sobre los costos de plantilla poblada por unidad ejecutora, costos netos por los diferentes ámbitos en que se divide el Instituto, costos de plantilla por diferentes periodos de tiempo (quincena, ejercicio), y las diversas combinaciones de los aspectos que se encuentran identificados en la misma.

Cabe mencionar, que mientras se cuente con sistemas que manejen la información necesaria para tomar decisiones, el uso del BI proporciona aspectos que no podemos obtener mediante un sistema transaccional, de allí la importancia de su implementación.

- Simplificar procesos. Se puede apreciar con claridad, que la implementación y uso de BI en el Instituto simplifica una gran variedad de procesos involucrados en la toma de decisiones, puesto que se disminuye el número de personas que intervienen para la explotación de información y generación de reportes, se eliminan trámites administrativos, se reducen tiempos de espera, y el ejecutivo que requiere la información puede extraerla personalmente en el momento adecuado.

Este caso práctico muestra como la configuración y concepción de una estructura adecuada de varias tecnologías puede ayudar a extraer información que nos permitan tomar mejores decisiones, basadas en información confiable y obtenida con un mínimo tiempo de respuesta. Se comprueba también que la implementación de soluciones gráficas proporciona una visión más representativa y fácil de interpretar,

factores que brindan un plus que facilita las labores de los directivos al momento de proveerse de información necesaria para sus actividades.

Las recomendaciones sobre investigaciones futuras respecto a la implementación de BI en INEGI documentado en este caso práctico, pueden manejar como siguiente etapa la integración de las herramientas de Kettle ETL con que cuenta la plataforma de Pentaho, a efecto de realizar la extracción, transformación y carga de la información a los esquemas de BD que se diseñen para cubrir las necesidades de los tableros de control; así como la implementación de los flujos de trabajo a través del Work Flow Engine, cubriendo de esta manera varios módulos de los procesos involucrados en la creación y administración de la información base para la toma de decisiones.



GLOSARIO

7. GLOSARIO.

Analysis Services Ofrece funciones de procesamiento analítico en línea (OLAP) y minería de datos para aplicaciones de Business Intelligence. Analysis Services admite OLAP y permite diseñar, crear y administrar estructuras multidimensionales que contienen datos agregados desde otros orígenes de datos, como bases de datos relacionales. En el caso de las aplicaciones de minería de datos, Analysis Services permite diseñar, crear y visualizar modelos de minería de datos que se construyen a partir de otros orígenes de datos mediante el uso de una gran variedad de algoritmos de minería de datos estándar del sector.

ANT Apache Ant es una herramienta usada en programación para la realización de tareas mecánicas y repetitivas, normalmente durante la fase de compilación y construcción (build).

Esta herramienta, hecha en Java, tiene la ventaja de no depender de los órdenes de shell de cada sistema operativo, sino que se basa en archivos de configuración XML y clases Java para la realización de las distintas tareas, siendo idónea como solución multi-plataforma. Ant es un proyecto de código abierto de la Apache Software Foundation.

BDR Base de Datos Relacional. Sistema de almacenamiento de datos basado en un conjunto de tablas unidas mediante relaciones.

Benchmark El benchmark es una técnica utilizada para medir el rendimiento de un sistema o componente de un sistema, frecuentemente en comparación con algún parámetro de referencia.

Business Intelligence Son las herramientas o estrategias que se implementan para hacer más eficiente el proceso de toma de decisiones.

Classpath Variable de entorno que permite a la máquina virtual java saber donde localizar sus clases.

CSS Hojas de Estilo en Cascada (Cascading Style Sheets), es un mecanismo simple que describe cómo se va a mostrar un documento en la pantalla, o cómo se va a imprimir, o incluso cómo va a ser pronunciada la información presente en ese documento a través de un dispositivo de lectura. Esta forma de descripción de estilos ofrece a los desarrolladores el control total sobre estilo y formato de sus documentos.

Contenedor de aplicaciones Dicho de forma genérica no es más que un entorno de ejecución estandarizado que ofrece unos servicios por medio de componentes. Los componentes externos al contenedor tienen una forma estándar de acceder a los servicios de dicho contenedor, con independencia del fabricante.

Crosstabs. Efecto de transformar filas en columnas.

Dashboard o tableros de control. Están diseñados para generar el máximo impacto visual posible en un formato optimizado para lograr mostrar la información rápidamente, por medio de una combinación de tablas, gráficos, reglas de medición, cuadrantes y otros indicadores gráficos

Datasource Los objetos DataSource permiten manejar un pool de conexiones a una base de datos.

Deploy Efecto de colocar un archivo .war o .ear para que el servidor coloque los archivos en las carpetas correspondientes y así poder publicar la aplicación.

Dimensión Criterio de clasificación de información. Eje de análisis. Lista de valores que proporciona un índice a los datos. Por ejemplo: <Tiempo>, <Geografía>, <Producto>. Elementos de una dimensión. Posibles valores de un eje de análisis. Por ejemplo, "Enero de 1998", "Trimestre 4 de 1998", o "1996" para la dimensión <Tiempo> y "Bilbao", "Andalucía" o "Zona Norte" para la dimensión <Geografía>

DM Datamart. Sistema que mantiene una copia de parte de un DataWarehouse para un uso departamental. Almacén de datos históricos relativos a un departamento de una organización, utilizado por una herramienta OLAP para procesar información, elaborar informes y vistas.

DOM Document Object Model (DOM). El Modelo de Objetos del Documento (DOM) es una interfaz de programación de aplicaciones (API) para documentos HTML y XML. Define la estructura lógica de los documentos y el modo en que se accede y manipula un documento. En la especificación del DOM, el término "documento" se utiliza en un sentido amplio. XML se utiliza cada vez más como un medio para representar muchas clases diferentes de información que puede ser almacenada en sistemas diversos, y mucha de esta información se vería, en términos tradicionales, más como datos que como documentos. Sin embargo, XML presenta estos datos como documentos, y se puede usar el DOM para manipular estos datos.

Con el Modelo de Objetos del Documento los programadores pueden construir documentos, navegar por su estructura, y añadir, modificar o eliminar elementos y contenido. Se puede acceder a cualquier cosa que se encuentre en un documento HTML o XML, y se puede modificar, eliminar o añadir usando el Modelo de Objetos del Documento, salvo algunas excepciones.

Drill down Descomponer (visualmente) en detalle un dato según una jerarquía de una dimensión.

Dril up Agregar (visualmente) un dato según una jerarquía de una dimensión.

DW DataWarehouse. Sistema almacén de datos que reúne la información generada por los distintos departamentos de una organización. Pretende conseguir que cualquier departamento pueda acceder a información de cualquiera de los otros mediante un único medio, así como obligar a que los mismos términos tengan el mismo significado para todos. Es un almacén de datos históricos, utilizado por una herramienta OLAP para procesar información, elaborar informes y vistas. También se define como un conjunto de datos orientados por tema, integrados, variables en el tiempo y no volátiles que se emplea como apoyo a la toma de decisiones.

ERP Los sistemas de planificación de recursos de la empresa (en inglés ERP, Enterprise Resource Planning) son sistemas de gestión de información que integran y automatizan muchas de las prácticas de negocio asociadas con los aspectos operativos o productivos de una empresa.

G2B Government To Business Relación entre la Administración y las empresas. Buscando la participación activa y comprometida del sector privado en la modernización gubernamental. Ser refiere a los tramites administrativos que las empresas pueden realizar con la Administración a través de medios electrónicos.

G2C Government To Consumer Relación entre la Administración y los ciudadanos. Ser refiere a los trámites administrativos que los ciudadanos pueden realizar con la Administración a través de medios electrónicos.

G2G Relaciones Gobierno-Gobierno para que este actúe más eficientemente en la prestación de los servicios y en la administración de la información, reforzando su liderazgo frente a la sociedad.

J2EE Java Platform, Enterprise Edition o Java EE (anteriormente conocido como Java 2 Platform, Enterprise Edition o J2EE hasta la versión 1.4), es una plataforma de programación—parte de la Plataforma Java—para desarrollar y ejecutar software de aplicaciones en Lenguaje de programación Java con arquitectura de n niveles distribuida, basándose ampliamente en componentes de software modulares ejecutándose sobre un servidor de aplicaciones.

JavaScript Lenguaje de programación interpretado, es decir, que no requiere compilación, utilizado principalmente en páginas web, con una sintaxis semejante a la del lenguaje Java y el lenguaje C.

JBoss Es un servidor de aplicaciones J2EE de código abierto implementado en Java puro. Al estar basado en Java, JBoss puede ser utilizado en cualquier sistema operativo que lo soporte.

JDK Java Development Kit o (JDK), es un software que provee herramientas de desarrollo para la creación de programas en java. Puede instalarse en una computadora local o en una unidad de red.

En la unidad de red se puede tener la aplicación distribuida en varias computadoras y trabajar como una sola aplicación.

JRE Es el acrónimo de Java Runtime Environment (entorno en tiempo de ejecución Java) y se corresponde con un conjunto de utilidades que permite la ejecución de programas java sobre todas las plataformas soportadas.

El JRE está constituido por una JVM (máquina virtual Java) que es el programa que interpreta el código Java y además por las librerías de clases estándar que implementan el API de Java. Ambas JVM y API deben ser consistentes entre sí, de ahí que sean distribuidas de forma conjunta.

Un usuario sólo necesita el JRE para ejecutar las aplicaciones desarrolladas en lenguaje Java, mientras que para desarrollar nuevas aplicaciones en dicho lenguaje es necesario un entorno de desarrollo, denominado JDK, que además del JRE (mínimo imprescindible) incluye, entre otros, un compilador para Java.

OC4J es el acrónimo de **O**racle **A**pplication **S**erver **C**ontainers for **(4)** **J**2EE . Es decir, el servidor de aplicaciones J2EE de oracle.

OLAP On Line Analytical Processing. Procesamiento Analítico En Línea. Se trata de procesos de análisis de información. Estos sistemas están orientados al acceso en modo consulta.

Outsourcing Proceso en el cual una firma identifica una porción de su proceso de negocio que podría ser desempeñada más eficientemente y/o más efectivamente por otra corporación, la cual es contratada para desarrollar esa porción de negocio. Esto libera a la primera organización para enfocarse en la parte o función central de su negocio.

Roll up Proceso que calcula para un indicador, y para una o más de las dimensiones por las que ese indicador se mueve, los valores agregados o padres sucesivos a partir de la suma de sus hijos, según las jerarquías especificadas, pudiendo poseer cada dimensión más de una jerarquía. Por ejemplo, es el proceso que suma los ingresos por cada provincia acumulándolos en los ingresos de la comunidad autónoma correspondiente. Se trata de una función que relaciona los valores de dos niveles jerárquicos distintos y adyacentes en una dimensión, transformando un grupo de datos de un nivel en un único dato asignable a otro valor del nivel superior.

Root En sistemas de ficheros se refiere al directorio raíz. En Unix se refiere al usuario principal.

SERVIDOR DE APLICACIONES. Se denomina servidor de aplicaciones a un servidor en una red de computadores que ejecuta ciertas aplicaciones

Usualmente se trata de un dispositivo de software que proporciona servicios de aplicación a las computadoras cliente. Un servidor de aplicaciones generalmente gestiona la mayor parte (o la totalidad) de las funciones de lógica de negocio y de acceso a los datos de la aplicación. Los principales beneficios de la aplicación de la tecnología de servidores de aplicación son la centralización y la disminución de la complejidad en el desarrollo de aplicaciones. Si bien el término es aplicable a todas las plataformas de software, hoy en día el término *servidor de aplicaciones* se ha convertido en sinónimo de la plataforma J2EE de Sun Microsystems.

TI Tecnologías de Información. Término general que se refiere al conocimiento y uso de computadoras y sistemas de comunicación electrónicos en las organizaciones.

XHTML (Lenguaje de Marcado de Hipertexto Extensible) es una versión más estricta y limpia de HTML, que nace precisamente con el objetivo de remplazar a HTML ante su limitación de uso con las cada vez más abundantes herramientas basadas en XML. XHTML extiende HTML 4.0 combinando la sintaxis de HTML, diseñado para mostrar datos, con la de XML, diseñado para describir los datos.

XML *Extensible Markup Language* («lenguaje de marcas extensible»), es un metalenguaje extensible de etiquetas desarrollado por el World Wide Web Consortium (W3C).

XMLHttpRequest es una interfaz empleada para realizar peticiones HTTP y HTTPS a servidores WEB. Para los datos transferidos se usa cualquier codificación basada en texto, incluyendo: texto plano, XML, JSON, HTML y codificaciones particulares específicas. La interfaz se presenta como una clase

de la que una aplicación cliente puede generar tantas instancias como necesite para manejar el diálogo con el servidor.

El uso más popular, si bien no el único, de esta interfaz es proporcionar contenido dinámico y actualizaciones asíncronas en páginas WEB mediante tecnologías construidas sobre ella como por ejemplo AJAX.

XSLT Es un lenguaje que se usa para convertir documentos XML en otros documentos XML; puede convertir un documento XML que obedezca a un DTD a otro que obedezca otro diferente, un documento XML bien formado a otro que siga un DTD, o, lo más habitual, convertirlo a "formatos finales", tales como WML (usado en los móviles WAP) o XHTML.

WIDGETS En el contexto de la programación de aplicaciones visuales, los *widgets* tienen un significado más amplio como componente o control visual que el programador reutiliza y tienen un gran valor para el usuario, idea íntimamente ligada con el concepto de interfaz gráfica de usuario (GUI por sus siglas en inglés).

TESIS TESIS TESIS TESIS TESIS



BIBLIOGRAFIA

TESIS TESIS TESIS TESIS TESIS

8. BIBLIOGRAFÍA.

Bologna, J. y Walsh, A. M. (1997) *The Accountant's Handbook of Information Technology*. John Wiley and Sons.

Cano, C. (1999) *Business Intelligence, decisiones de negocio basadas en tecnología: ruta crítica del negocio moderno*. Retomado el 5 de Junio del 2001, de ProQuest Direct en el World Wide Web <http://www-cib.mty.itesm.mx>.

Davenport, T.; Prusak, L. (1998), *Working Knowledge: How Organizations Manage What They Know*. Harvard Business School Press.

David Loshin (2003) *Business Intelligence: The Savvy Manager's Guide*. Morgan Kaufmann Publishers.

Douglas Bell] Douglas Bell, Doug Bell, Mike Parr, Alfonso Vidal Romero Elizondo, Guillermo Levine Gutiérrez (2003) *Java Para Estudiantes*. Pearson Educación.

Erik T. Ray (2003) *Learning XML*. O'Really.

Ignacio García Valcárcel (2001) *CRM: Gestión de la relación con los clientes*. FC Editorial.

Jesús García (1998) *La comunicación interna*. pág.259. Ediciones Díaz de Santos S.A.

José Manuel Huidobro Moya, Rubén de la Peña Martínez. David Roldan Martínez, José Antonio de la Fuente (2005). *La tecnología e-business*. Cengage Learning Editores.

Kumar, K., y Hillegersberg, J. v. (2000). *Enterprise resource planning: Introduction. Communications of the ACM* pag. 22-26.

Lambert, D. M., M. C. Cooper, y J. D. Pagh (1998) *Supply Chain Management: Implementation Issues and Research Opportunities.* International Journal of Logistics Management.

Michael Hugos (2006) *Essentials of Supply Chain Management.* Wiley.

Michael Mahemoff (2006) *Ajax design patterns.* O'Really.

Ralph Kimball (2002) *The Data Warehouse Toolkit* pag. 310. Wiley.

REFERENCIAS A DOCUMENTOS.

[1] Reporte Mundial del Sector Publico, 2003.

[2] USA Today, 13 de marzo del 2002.

[3] The Wall Street Journal On Line, 3 de enero del 2002.

[4] The New York Times, 20 de mayo de 2001.

[5] White Paper Javed Beg and Shadab Hussain Data Warehouse Consultants, Wipro Technologies, "Data Quality: A Problem and an Approach",

[6] Tomás Orozco La Roche. Experiencias de buenas prácticas en el desarrollo del e-Gobierno en México. IX Congreso Internacional del CLAD sobre la Reforma del Estado y de la Administración Pública, Madrid, España, 2 – 5 Nov. 2004.

[7] White Paper Jill Dyché, baseline consulting, "BI Adoption Evolves".

DOCUMENTOS CON ACCESO EN LA WWW.

[Cognos] Dave Fiack, Jeppesen

“Análisis visual con inteligencia de negocios Cognos. Cognos visualizer: Tablero de control para una visión crítica y rápida.”

Recuperado de:

http://www.cognos.com/latin/sp/products/ebi/fs_visualizer_spanish.pdf

[Gattaca] Gattaca S.A. 2006

“Casos de éxito”

Recuperado de:

<http://www.e-gattaca.com/eContent/NewsDetail.asp?ID=68&IDCompany=4>

[INEGI] Instituto Nacional de Estadística Geografía e Informática, 2008

Recuperado de:

<http://www.inegi.gob.mx/inegi/contenidos/espanol/instituto/default.asp?c=1503>

[InformationWeek] Juan Antonio Osequeda, Jueves 06 de septiembre de 2007.

“En busca del dato perdido”

Recuperado de:

<http://www.informationweek.com.mx/articulo-59-6817-394.html>

[Ingeniería Industrial Chile] Ingeniería Industrial Universidad de Chile, 2006

“Cómo generar ventajas competitivas aplicando Business Intelligence, (BI)”.

Recuperado de:

http://www.tis.cl/2007/futurosTalleres/2006/Taller_1/Programa/?Taller_12006

[JFreeChart] David Gilbert's

“JFreeChart”

Recuperado de:

<http://www.jfree.org/jfreechart/>

[JPivot] SourceForge, 17 de marzo del 2008

“JPivot”

Recuperado de:

<http://pivot.sourceforge.net/>

[Mondrian] Pentaho

“Pentaho Analysis Services: Mondrian Project”

Recuperado de:

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

[Monografías] Gabriel Alberto Trejo Islas

“Business Intelligence. Decisiones inteligentes en General Motors”

Recuperado de:

<http://www.monografias.com/trabajos17/decisiones-inteligentes/decisiones-inteligentes.shtml>

[Pentaho] Pentaho Open Source Business Intelligence

Recuperado de:

<http://www.pentaho.com/>

[Todo BI] Todo BI, domingo, 21 de mayo del 2006

“Pentaho: la solución Open Source Business Intelligence”

Recuperado de:

<http://todobi.blogspot.com/2006/05/pentaho-la-solucion-open-source.html>

[WC3] World Wide Web Consortium, Oficina española. 9 de enero del 2008

“Guía Breve de Tecnologías XML”

Recuperado de:

<http://www.w3c.es/divulgacion/guiasbreves/tecnologiasXML>

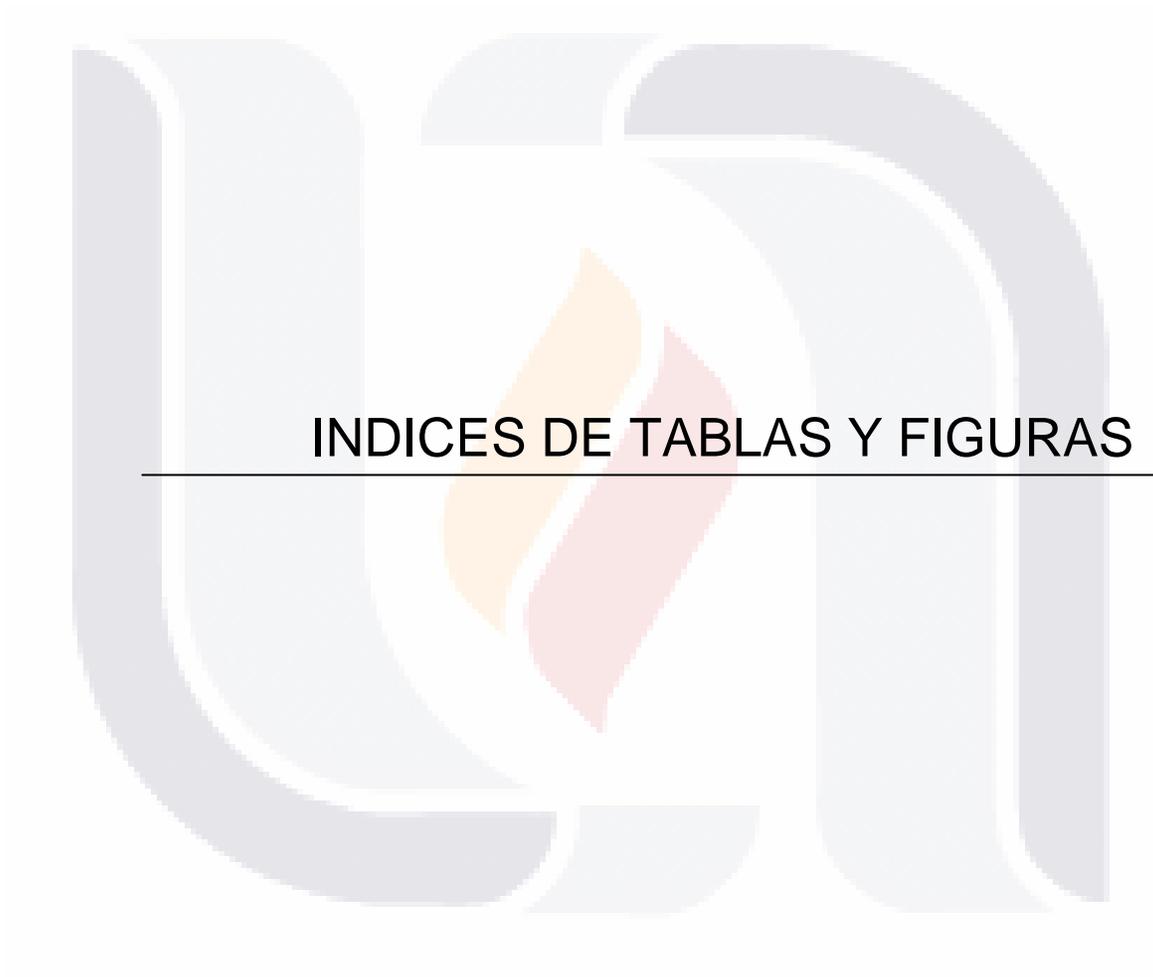
[Wiki Java] Wikipedia la enciclopedia libre, 23 abril del 2008

“Lenguaje de programación Java”

Recuperado de:

http://es.wikipedia.org/wiki/Lenguaje_de_programaci%C3%B3n_Java





INDICES DE TABLAS Y FIGURAS

9. ÍNDICE DE TABLAS.

N° de tabla	Descripción	Origen
1	Relación de estados por ámbito Regional-Estatal	http://intranet.inegi.gob.mx/default.aspx
2	Tabla de acciones necesarias actualmente para la toma de decisiones	
3	Resultados obtenidos a través del chart de BI sobre plazas ocupadas	

10. ÍNDICE DE FIGURAS O DIAGRAMAS.

N° de figura	Descripción	Origen
1	Origen del conocimiento	
2	Presencia de INEGI en el territorio nacional	
3	Estructura organizacional del INEGI	http://intranet.inegi.gob.mx/default.aspx
4	SIA y la Integración de servicios	
5	Esquema para la toma de decisiones anterior a la implantación de BI	
6	Esquema esperado para la toma de decisiones utilizando BI	
7	Tecnologías que apoyan el proceso de toma de decisiones a través del BI	
8	Clasificación de los usuarios de BI	[7]
9	Evolución del BI	[7]
10	Muestras de dashboards	
11	Pentaho Open BI suite	[Todo BI]
12	Muestra de un cubo visualizado a través de JPivot	
13	Muestra de tipos de gráficos generados por JFreeChart	
14	Esquema de aplicaciones web. Clásico vs. Ajax	(Michael Mahemoff, 2006).
15	Arquitectura de la plataforma Pentaho	Creating Pentaho Solutions.pdf obtenido de: http://www.pentaho.com
16	Deploy de un proyecto con un archive .ear	
17	Asignación de un nombre al proyecto	
18	Listado de módulos	
19	Resumen del Nuevo proyecto.	
20	Carga del proyecto al IDE.	
21	Conclusión del deploy	
22	Carga de nuevas librerías.	
23	Configuración de DS al IDE.	
24	Definición de DS para Oracle.	
25	Definición de DS para Hibernate.	
26	Definición de DS para SampleData	
27	Definición de DS para Quartz	
28	Definición de DS para Shark	
29	Configuración de la url base.	

ÍNDICE DE FIGURAS O DIAGRAMAS.

N° de figura	Descripción	Origen
30	Configuración del password para publicar.	
31	Ejecución de la plataforma Pentaho.	
32	Ejecución general de una tarea en la plataforma Pentaho	http://www.hisp.info/inf5750/wiki/DashboardManual#Step
33	DW. Diagrama de estrella con información del SIA.	
34	Regionales semaforizadas en base a parámetros.	
35	Controles para actualizar los semáforos.	
36	Desglose del semáforo seleccionado.	
37	Pop up. Identificación de la Regional seleccionada.	
38	Desglose de plazas ocupadas por estatal.	
39	Pop up. Dashboard de la regional seleccionada.	
40	Cubo. Plazas por Regional-Estatal.	
41	Desglose del cubo: Plazas por Regional-Estatal.	



APENDICE A

A.1. Descarga de archivos.

Descargar y descompactar en cualquier ruta.

1. pentaho_j2ee_deployments-1.6.0.GA.863.zip
2. pentaho_sample_data-1.6.0.GA.863.zip
3. pentaho_solutions-1.6.0.GA.863.zip

Pentaho_j2ee_deployments-1.6.0.GA.863.zip contiene un script Ant que permite construir varios archivos, como: .war, .ear, etc. configurados para diferentes servidores de aplicaciones.

A.2. Construcción del archivo .ear para Oracle OC4J.

1. Se considera root a la carpeta donde se descompacto [pentaho_j2ee_deployments]. En este directorio se encuentra el archivo build.xml.
2. En el archivo build.xml ubicar y eliminar la etiqueta "build-orion2.0.5-ear". Reemplazar por el siguiente XML:

```
<!-- =====  
target: build-orion2.0.5-ear  
===== -->  
<target name="build-orion2.0.5-ear" depends="zip-pentaho-style-war">  
  <antcall target="war-pentaho-tomcat">  
    <param name="rdbms" value="hsqldb" />  
  </antcall>  
  <mkdir dir="{build.ears.dir}/orion" />  
  <ear destfile="{build.ears.dir}/orion/pentaho.ear"  
    appxml="{ear.dir}/application.xml">  
    <manifest>  
      <attribute name="Implementation-Title"  
        value="{impl.title}" />  
      <attribute name="Implementation-Version"  
        value="{impl.version}" />  
      <attribute name="Implementation-Vendor"  
        value="{impl.vendor}" />  
    </manifest>  
    <fileset dir="{build.wars.dir}">  
      <include name="pentaho-style.war" />  
      <include name="sw-style.war" />  
    </fileset>  
    <fileset dir="{build.wars.dir}/tomcat/hsqldb"  
      includes="pentaho.war" />  
    <metainf dir="pentaho-res/orion" includes="*.xml" />  
  </ear>  
</target>
```

3. En root, se encuentra el archivo `override.properties`. Agregar la siguiente línea:

```
ear.dir=pentaho-res/ear
```

Con estos pasos basta para construir el archivo `pentaho.ear` para el OC4J.

4. Se necesita hacer unos cuantos cambios al archivo `web.xml`:

- a. El default del contenedor de aplicaciones OC4J está configurado sobre el puerto 8080. Así que debe cambiarse para que el default sea el Puerto 8988.

NOTA. Los cambios al `web.xml` deben replicarse en el `web.xml` que se encuentra en el subdirectorio para el deploy.

```
[pentaho_j2ee_deployments]/pentaho-precompiled-  
jsps/pentaho.war/WEB-INF/web.xml
```

```
Y [pentaho_j2ee_deployments]pentaho-webapp/WEB-  
INF/web.xml
```

Localizar el param-name `base-url`, con el valor <http://localhost:8080/pentaho/> y cambiarlo por `http://localhost:8888/pentaho/`.

- b. Debe asegurarse que el servidor encontrara el directorio de `Pentaho-solutions`. Localizar el param-name `pentaho-solutions`, y remplazarlo con el valor de la ruta absoluta del directorio en donde desempacamos `pentaho-solutions`. Ejemplo `d:\work\pentaho-solutions`.

5. El archivo `application.xml` se encuentra en el directorio `[pentaho_j2ee_deployments]/pentaho-res/ear`.

a. Eliminar todos los módulos que se encuentran en la etiqueta

`<!--additional web apps -->` , y el modulo `web-uri` para el `style.war` que se encuentran arriba de la etiqueta anterior. Los únicos módulos que debe contener el archive `application.xml` son `pentaho.war` y `pentaho-style.war`.

6. El archivo `orion-web.xml` no existe dentro de los archivos que se descargaron y descompactaron, por lo que debe crearse.

a. Crear el archivo `orion-web.xml` en el directorio `[pentaho_j2ee_deployments]/pentaho-webapp/WEB-INF`.

b. Agregar el siguiente xml al archivo:

```
<orion-web-app>
<resource-ref-mapping location="jdbc/HibernateDS" name="jdbc/Hibernate"/>
<resource-ref-mapping location="jdbc/SampleDataDS" name="jdbc/SampleData"/>
<resource-ref-mapping location="jdbc/QuartzDS" name="jdbc/Quartz"/>
<resource-ref-mapping location="jdbc/SharkDS" name="jdbc/Shark"/>
<resource-ref-mapping location="jdbc/SampleDataAdminDS"
name="jdbc/SampleDataAdmin"/>
<resource-ref-mapping location="jdbc/SampleDataDS" name="jdbc/datasource1"/>
<resource-ref-mapping location="jdbc/SampleDataDS" name="jdbc/datasource2"/>
<resource-ref-mapping location="jdbc/SampleDataDS" name="jdbc/datasource3"/>
<resource-ref-mapping location="jdbc/SampleDataDS" name="jdbc/datasource4"/>
<resource-ref-mapping location="jdbc/SampleDataDS" name="jdbc/solution1"/>
<resource-ref-mapping location="jdbc/SampleDataDS" name="jdbc/solution2"/>
<resource-ref-mapping location="jdbc/SampleDataDS" name="jdbc/solution3"/>
<resource-ref-mapping location="jdbc/SampleDataDS" name="jdbc/solution4"/>
</orion-web-app>
```

7. Finalmente construir el archivo pentaho.ear para realizar el deploy posterior a JDeveloper.

1. Debe abrirse una ventana de comando, y situarse en el directorio [pentaho_j2ee_deployments], ejecutar:

```
ant build-orion2.0.5-ear
```





APENDICE B

Antes de hacer el deploy del archivo .ear, debe darse de baja la instancia del OC4J, para realizar algunas adecuaciones a la configuración del servidor.

B.1. Conflictos con TopLink. Pentaho no usa Toplink, y el archivo antlr.jar tiene conflictos con la librería de hibernate3, por lo que debe eliminarse el directorio Toplink en la ruta [OC4J_Home].

B.2. Out of Memory Errors. Pentaho es una gran ingeniería, por lo que consume mucha memoria, al estar corriendo puede generar un mensaje de "out of memory". Por lo que debe modificarse el parámetro JVM_ARGS en el archivo [OC4J_Home]/bin/oc4j.cmd para incluir:

```
set JVMARGS=%OC4J_JVM_ARGS% -Xms128m -Xmx512m -XX:MaxPermSize=128m
```

B.3. Agregar el archivo hsqldb.jar en el directorio [OC4J_Home/lib].

Esto es para probar la aplicación sobre la base de datos que maneja la suite pentaho en la solución de ejemplo. Por lo que se requiere agregar el driver del JDBC para Hypersonic al directorio [OC4J_Home/lib]. El archivo hsqldb.jar se encuentra en:

```
[pentaho_j2ee_deployments]/pentaho-third-party.
```

Al momento de hacer el deploy del archivo .ear debe agregarse una entrada al classpath para este .jar.

B.4. Modificar el archivo data-sources.xml. Debe agregarse los datasources al archivo:

```
[OC4J_Home]/j2ee/home/config/data-sources.xml.
```

Agregando el siguiente xml, antes de la etiqueta final<datasources/>:

```
<data-source
class="com.evermind.sql.DriverManagerDataSource"
name="Hibernate"
location="jdbc/HibernateDS"
xa-location="jdbc/xa/HibernateDS"
ejb-location="jdbc/Hibernate"
connection-driver="org.hsqldb.jdbcDriver"
username="hibuser"
password="password"
url="jdbc:hsqldb:hsq://localhost/hibernate"
inactivity-timeout="30"
/>
<data-source
class="com.evermind.sql.DriverManagerDataSource"
name="SampleData"
location="jdbc/SampleDataDS"
xa-location="jdbc/xa/SampleDataDS"
ejb-location="jdbc/SampleData"
connection-driver="org.hsqldb.jdbcDriver"
username="pentaho_user"
password="password"
url="jdbc:hsqldb:hsq://localhost/sampledata"
inactivity-timeout="30"
/>
<data-source
class="com.evermind.sql.DriverManagerDataSource"
name="SampleDataAdmin"
location="jdbc/SampleDataAdminDS"
xa-location="jdbc/xa/SampleDataAdminDS"
ejb-location="jdbc/SampleDataAdmin"
connection-driver="org.hsqldb.jdbcDriver"
username="pentaho_admin"
password="password"
url="jdbc:hsqldb:hsq://localhost/sampledata"
inactivity-timeout="30"
/>
```

```
<data-source
class="com.evermind.sql.DriverManagerDataSource"
name="Quartz"
location="jdbc/QuartzDS"
xa-location="jdbc/xa/QuartzDS"
ejb-location="jdbc/Quartz"
connection-driver="org.hsqldb.jdbcDriver"
username="pentaho_user"
password="password"
url="jdbc:hsqldb:hsq://localhost/quartz"
inactivity-timeout="30"
/>
<data-source
class="com.evermind.sql.DriverManagerDataSource"
name="Shark"
location="jdbc/SharkDS"
xa-location="jdbc/xa/SharkDS"
ejb-location="jdbc/Shark"
connection-driver="org.hsqldb.jdbcDriver"
username="sa"
password=""
url="jdbc:hsqldb:hsq://localhost/shark"
inactivity-timeout="30"
/>
```

B.5. Problema en la solución XSLTs. Oracle tiene algún orden en la construcción XSLT, al tratar de hacer una referencia a una clase de Java como el valor de un namespace. Debe verificarse en cada xml dentro del directorio [pentaho-solutions]/system/custom/xsl, por cada ocurrencia. Son al menos 20, y hacer los siguientes cambios.

1. Abrir el archivo xslt.
2. Encontrar la clase Java referenciada en la declaración del namespace.

Están siempre al inicio del archivo y lucen parecidas a la siguiente línea.

```
xmlns:msg="org.pentaho.messages.Messages"
```

Nota. xmlns:msg no es la única posibilidad es solo un ejemplo.

3. Anteponer al nombre de la clase Java con:

<http://www.oracle.com/XSL/Transform/java/>. El ejemplo corregido debe ser:

```
xmlns:msg="http://www.oracle.com/XSL/Transform/java/org.pentaho.messages.Messages".
```

4. Repetir los pasos anteriores para cada archivo xslt en el directorio que contenga una ocurrencia de namespace.

B.6. Editar el archivo: [pentaho-solutions]/system/custom/xsl/files-list.xml. Buscar y eliminarlas siguientes líneas del XML.

Nota. No están una cerca de la otra.

```
<xsl:param name="target" />
```

```
<xsl:with-param name="target" select="$target"/>
```

B.7. El archivo antlr-2.7.6. Archivo que se encuentra en pentaho-third-party se renombra como antlr y se coloca en el directorio jlib del JDeveloper ejemplo:

```
C:\software\jdeveloper\toplink\jlib
```

B.8. Colocar el hsqldb.jar, log4j.jar, log4j-boot.jar en la carpeta de lib de oracle.

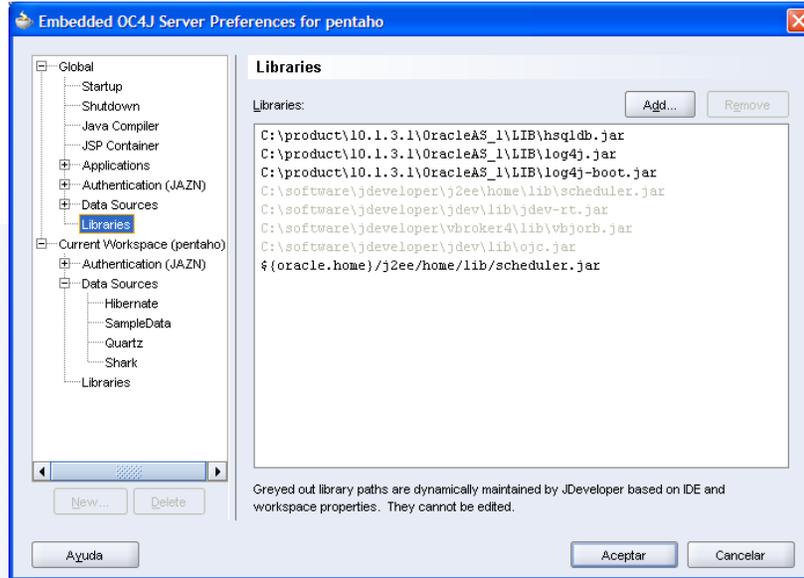


Figura 22. Carga de nuevas librerías.

B.9. Definición de DataSources en JDeveloper.

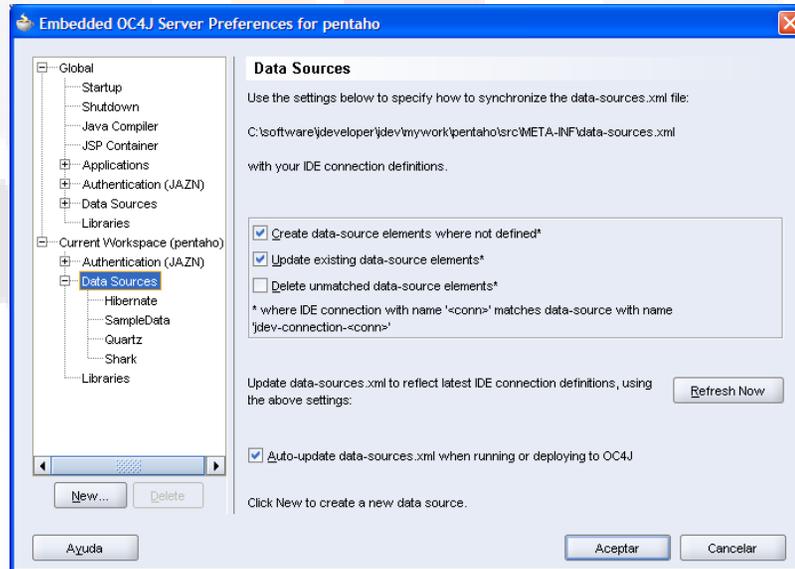


Figura 23. Configuración de DS al IDE.

B.9.1. DS para Oracle.

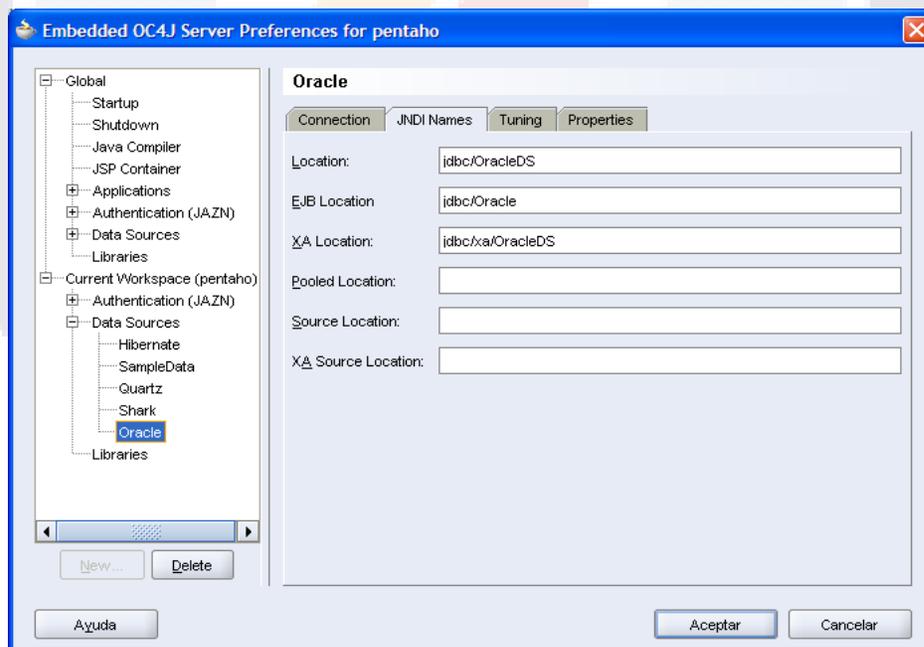
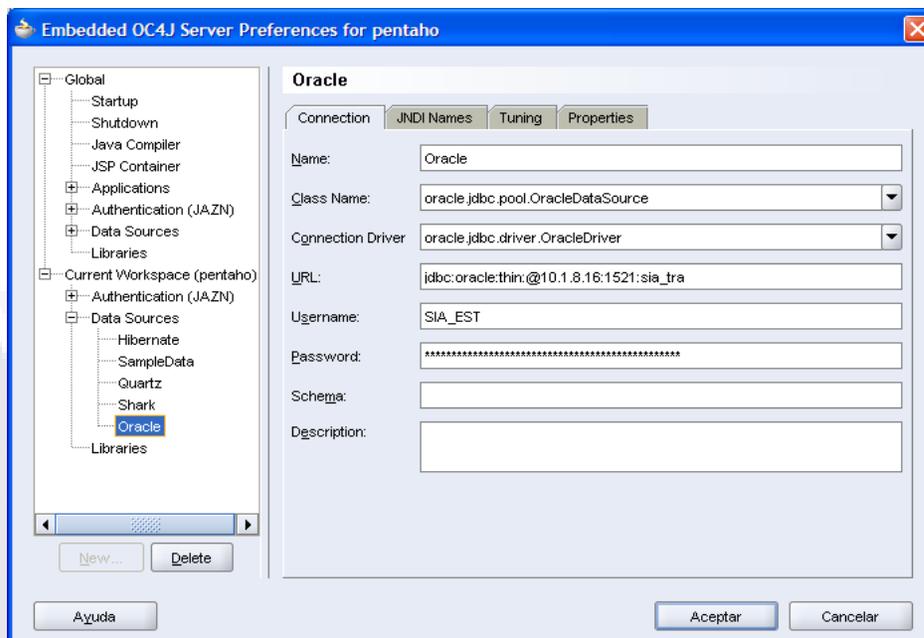


Figura 24. Definición de DS para Oracle.

B.9.2. DS para Hibérnate.

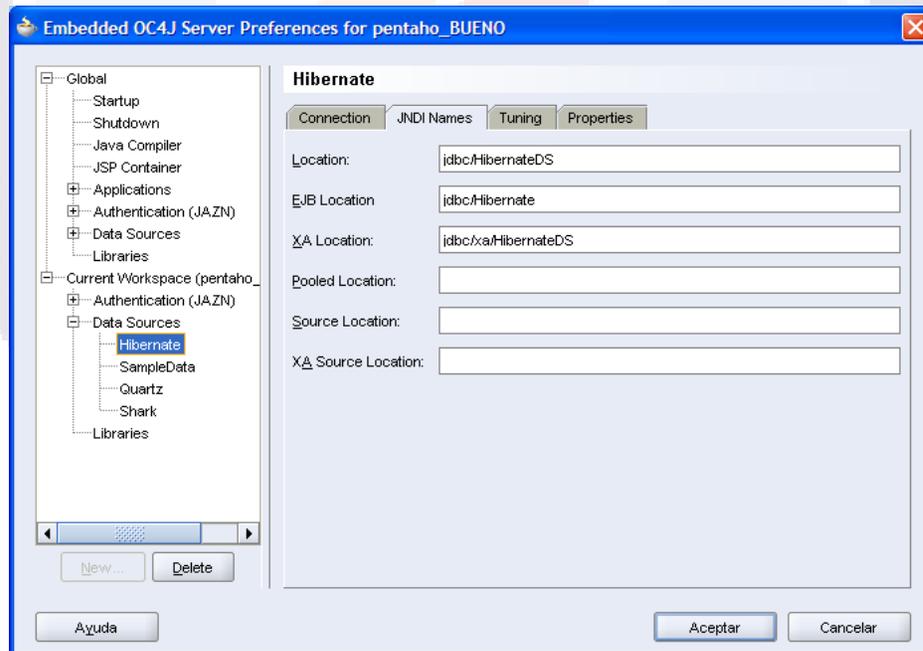
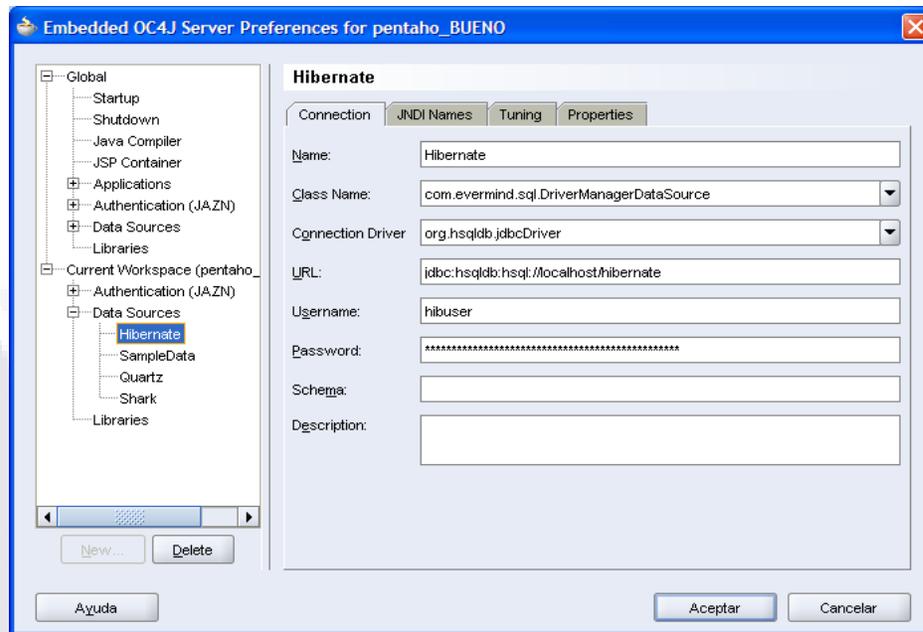


Figura 25. Definición de DS para Hibérnate.

B.9.3. DS Sampledata.

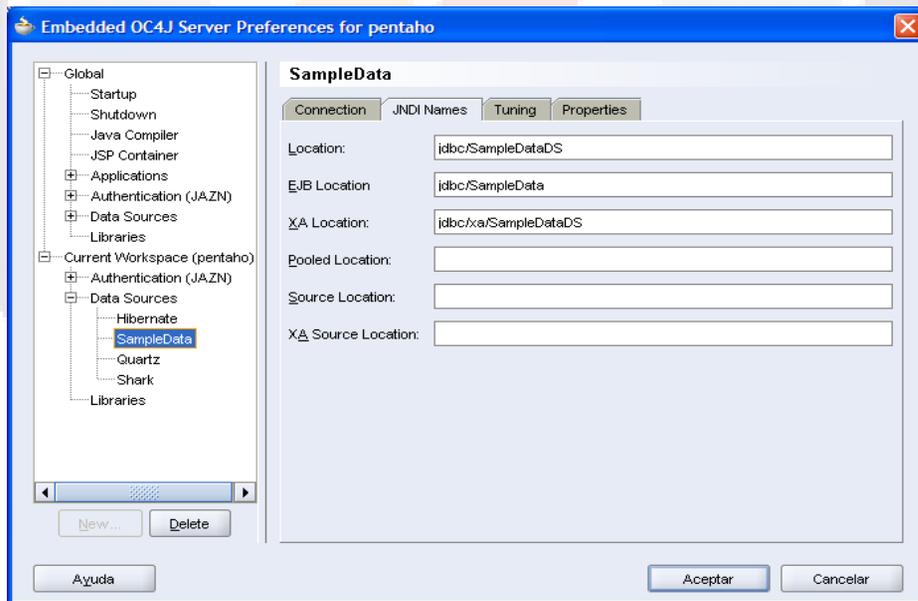
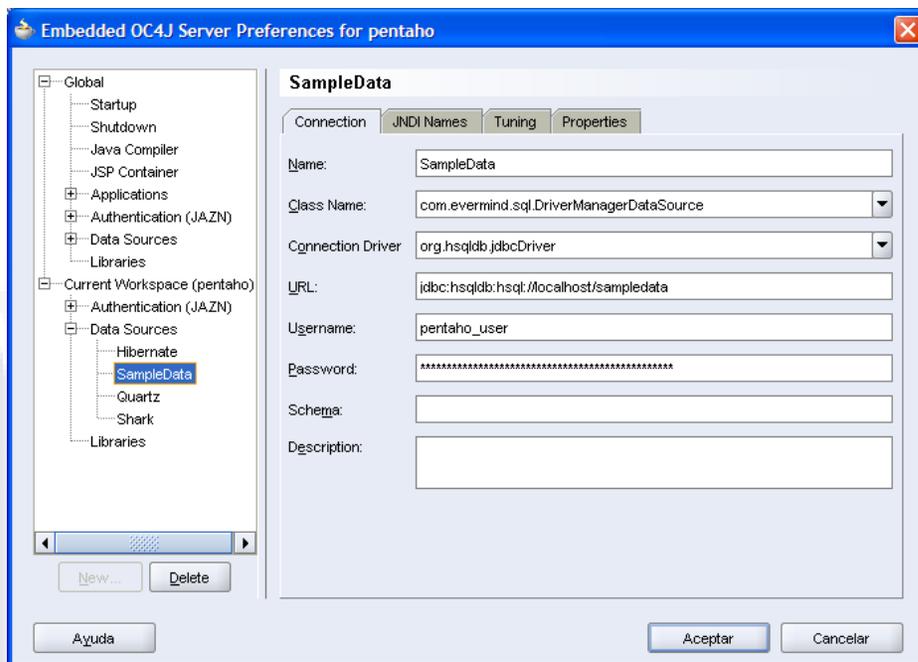


Figura 26. Definición de DS para SampleData.

B.9.4. DS para Quartz.

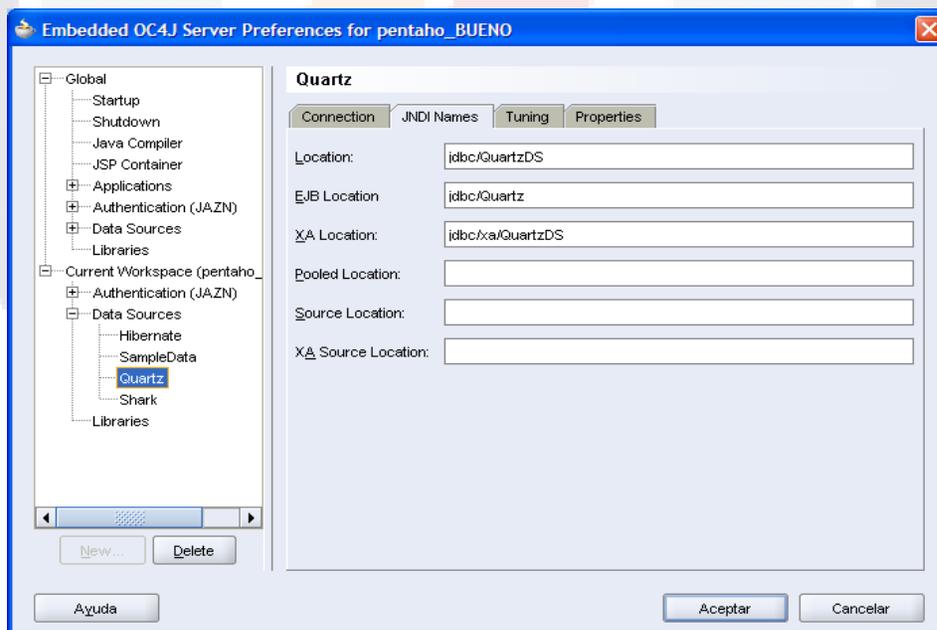
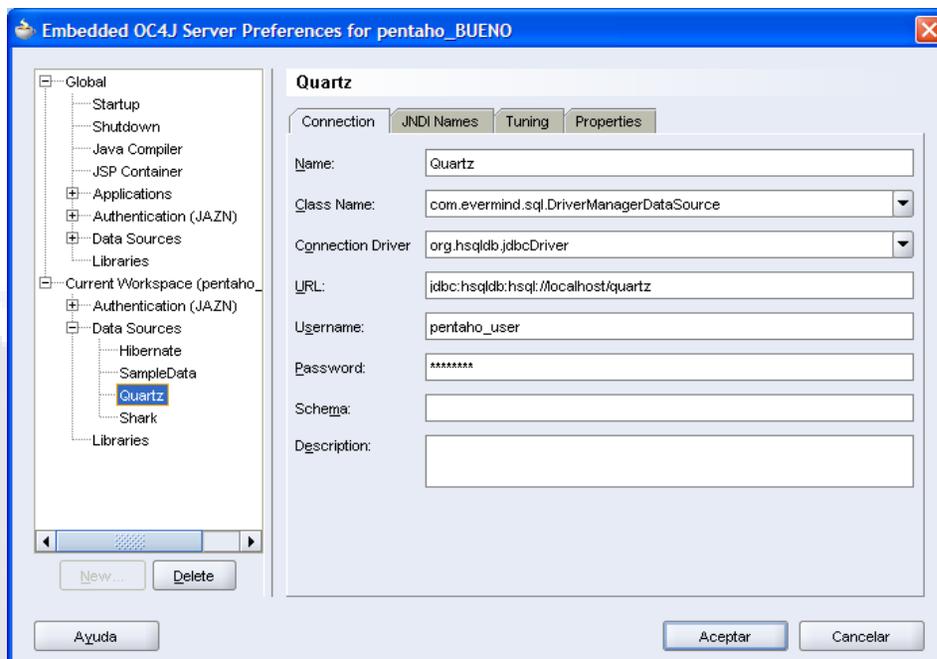


Figura 27. Definición de DS para Quartz.

B.9.5. DS para Shark.

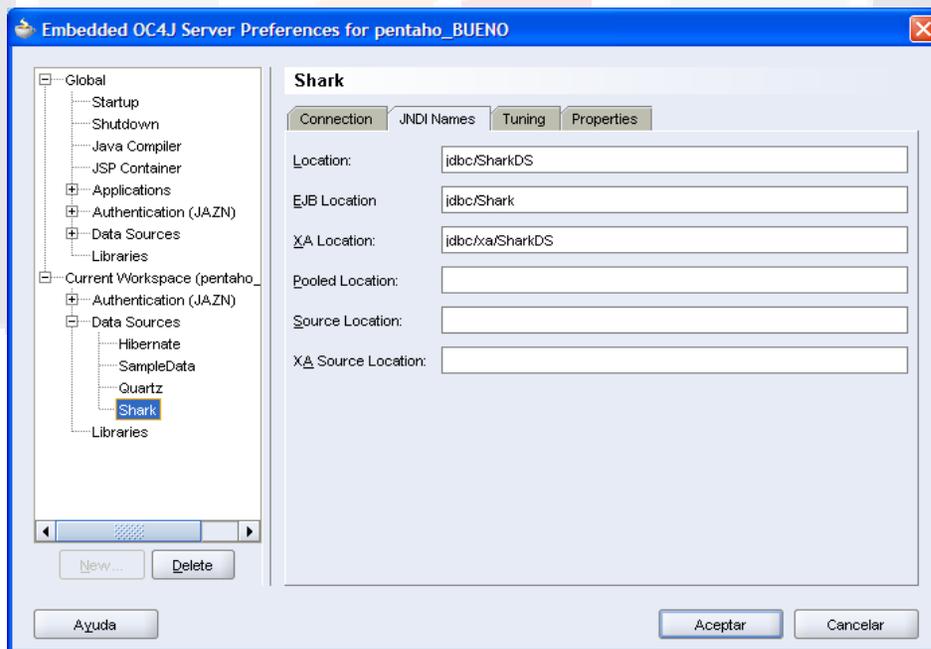
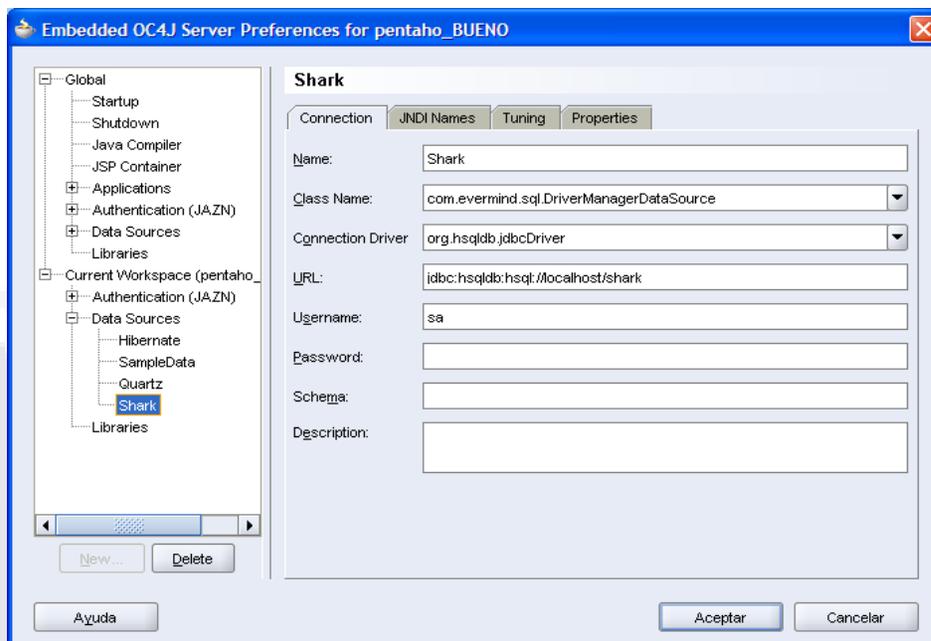


Figura 28. Definición de DS para Shark.

B.10. Configuración para mostrar las imágenes generadas automáticamente por la plataforma. Debe modificarse la url base en el archivo web.xml

```
<context-param>
  <param-name>base-url</param-name>
  <param-value>http://10.25.2.34:8988/pentaho/</param-value>
</context-param>
```

Figura 29. Configuración de la url base.

B.11. Pentaho se conforma de varios proyectos externos a su suite. Entre los cuales están ReportDesign, CubeDesigner, Etc, para poder realizar una publicación de los programas generados en estos proyectos es necesario proporcionar un password de publicación en el archivo Publisher_config.xml, ubicado en [pentaho-solutions\system].

```
1  <?xml version="1.0"?>
2  <!--
3      This file must be configured before client tools
4      will be able to publish onto the server. Publishing
5      involves accepting a file (action sequence, xml document,
6      anything) and placing it into the solution. This
7      action should only be allowed if the tool is configured
8      with a server/deployment specific password. The
9      RepositoryFilePublisher is now coded to expect an MD5
10     hash of the password entered here before it will allow
11     publish into the server.
12
13     This is a critical security measure designed to prevent
14     malicious actions from being published into your server.
15     Please take care to make the password sufficiently
16     difficult to guess. Also take care that the contents of
17     this file are not generally available
18 -->
19  <publisher-config>
20     <publisher-password>sia</publisher-password>
21  </publisher-config>
22
```

Figura 30. Configuración del password para publicar.

B.12. Archivos Orion-web.xml y Orion-application.xml. Una vez que realizamos las modificaciones anteriores a la configuración de nuestro servidor, los siguientes archivos deben lucir como a continuación se indica:

Orion-web.xml

```
<orion-web-app xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="http://xmlns.oracle.com/oracleas/schema/orion-web-
10_0.xsd" deployment-version="10.1.3.1.0"
deployment-time="1193079343906"
jsp-cache-directory="./persistence"
jsp-cache-tlds="standard"
temporary-directory="./temp"
context-root="/pentaho"
schema-major-version="10" schema-minor-version="0" >
<web-app-class-loader search-local-classesfirst="true"/>
<classpath path="C:\product\10.1.3.1\OracleAS_1\LIB"/>
<resource-ref-mapping location="jdbc/HibernateDS" name="jdbc/Hibernate"/>
<resource-ref-mapping location="jdbc/SampleDataDS" name="jdbc/SampleData"/>
<resource-ref-mapping location="jdbc/QuartzDS" name="jdbc/Quartz"/>
<resource-ref-mapping location="jdbc/SharkDS" name="jdbc/Shark"/>
<resource-ref-mapping location="jdbc/SampleDataAdminDS" name="jdbc/SampleDataAdmin"/>
<resource-ref-mapping location="jdbc/SampleDataDS" name="jdbc/datasource1"/>
<resource-ref-mapping location="jdbc/SampleDataDS" name="jdbc/datasource2"/>
<resource-ref-mapping location="jdbc/SampleDataDS" name="jdbc/datasource3"/>
<resource-ref-mapping location="jdbc/SampleDataDS" name="jdbc/datasource4"/>
<resource-ref-mapping location="jdbc/SampleDataDS" name="jdbc/solution1"/>
<resource-ref-mapping location="jdbc/SampleDataDS" name="jdbc/solution2"/>
<resource-ref-mapping location="jdbc/SampleDataDS" name="jdbc/solution3"/>
<resource-ref-mapping location="jdbc/SampleDataDS" name="jdbc/solution4"/>
</web-app>
</orion-web-app>
```

Orion-application.xml

```
<orion-application xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="http://xmlns.oracle.com/oracleas/schema/orion-application-
10_0.xsd" deployment-version="10.1.3.1.0" default-data-source="jdbc/OracleDS" component-
classification="external"
schema-major-version="10" schema-minor-version="0" >
<web-module id="pentaho" path="pentaho.war" />
<web-module id="pentaho-style" path="pentaho-style.war" />
<persistence path="persistence" />
<library path="C:\product\10.1.3.1\OracleAS_1\LIB" />
<imported-shared-libraries>
<remove-inherited name="oracle.toplink"/>
</imported-shared-libraries>
<principals path="principals.xml" />
<jazn provider="XML" />
<log>
<file path="application.log" />
</log>
<data-sources path="./data-sources.xml" />
</orion-application>
```

B.13. Ejecución de Pentaho. Si se han generado los archivos y configurado correctamente el servidor, solo basta ejecutar el proyecto desde el IDE, con lo que se debe observar una página como la siguiente.

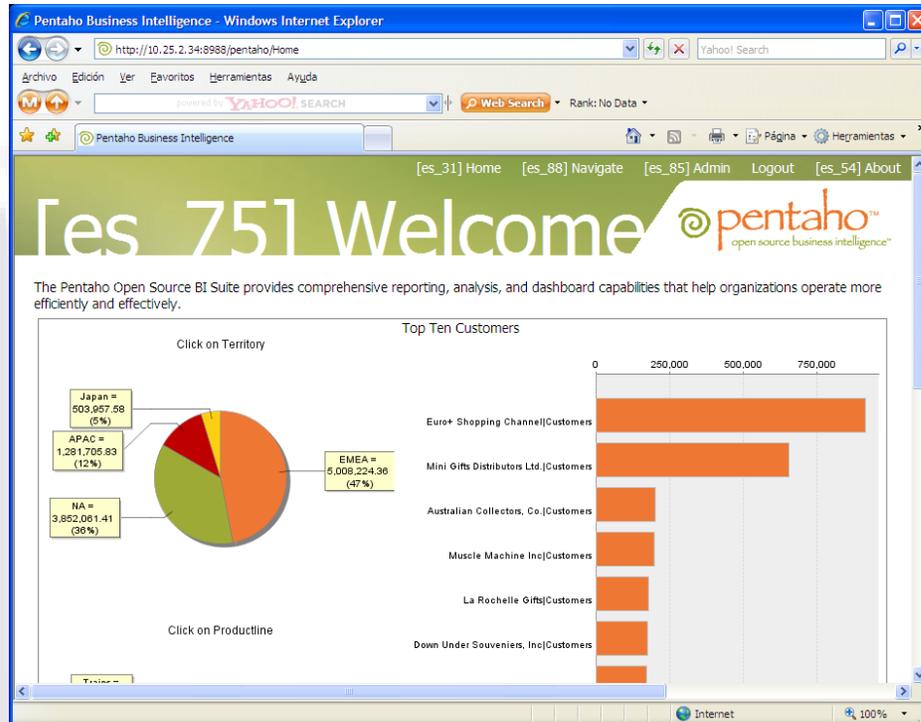


Figura 31. Ejecución de la plataforma Pentaho.

Misma que indica, que la suite de pentaho ha sido transformada correctamente para su uso sobre un servidor OC4J.



APENDICE C

C.1. Abrir el IDE. Para cargar el proyecto creado.

C.2. Seleccionar el origen de la aplicación. A través de la siguiente secuencia de opciones.

File-New-General-Applications\Application from EAR File

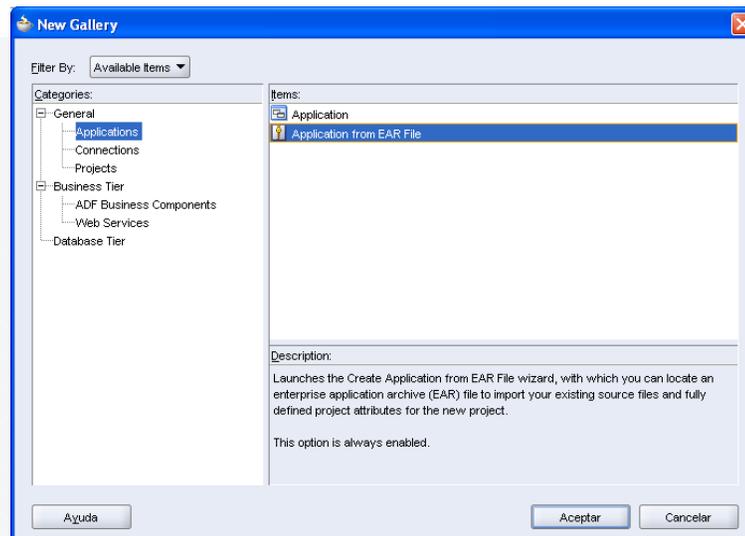


Figura 16. Deploy de un proyecto con un archive .ear

C.3. Carga del archivo. Indicar la ubicación física del archivo pentaho.ear y asignarle el nombre con el cual se identificara la suite de pentaho dentro de JDeveloper.

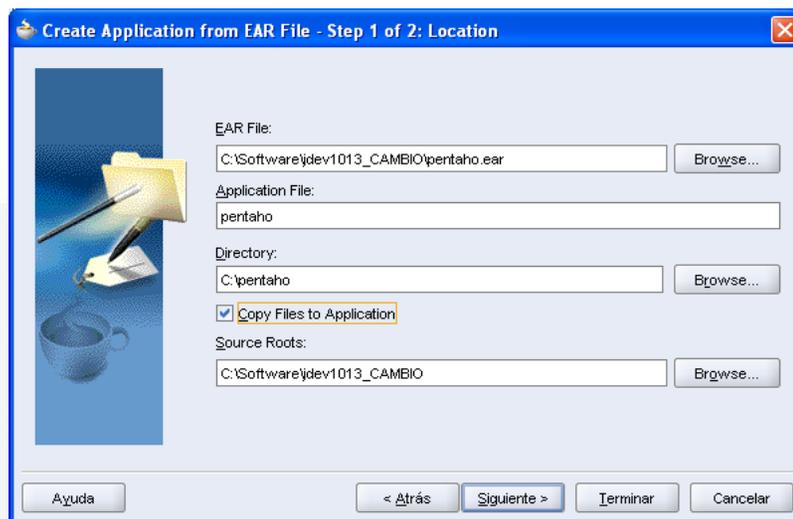


Figura 17. Asignación de un nombre al proyecto.

C.4. Desglose de módulos. Se desplegara el listado de los diferentes módulos de la suite que se instalaran dentro del proyecto, al dar clic en el botón Terminar se realiza la carga de los archivos al IDE.

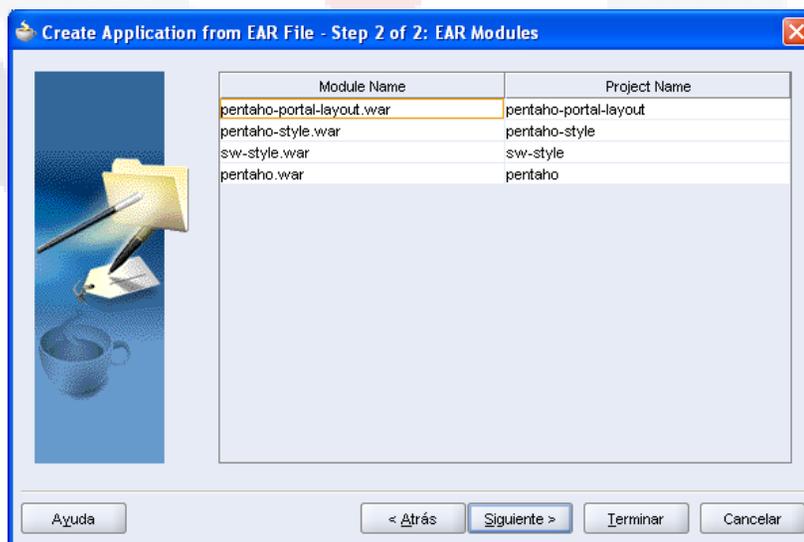


Figura 18. Listado de módulos

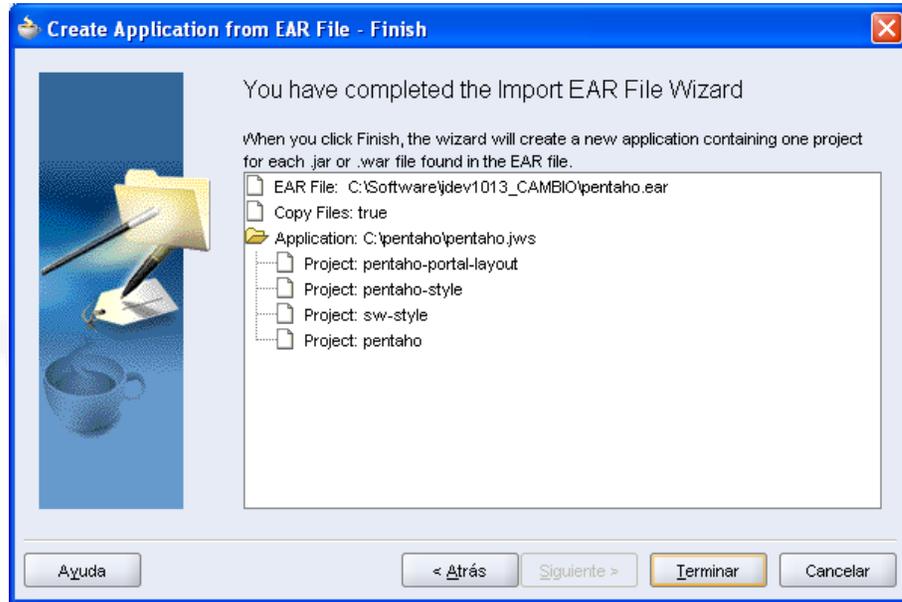


Figura 19. Resumen del Nuevo proyecto.

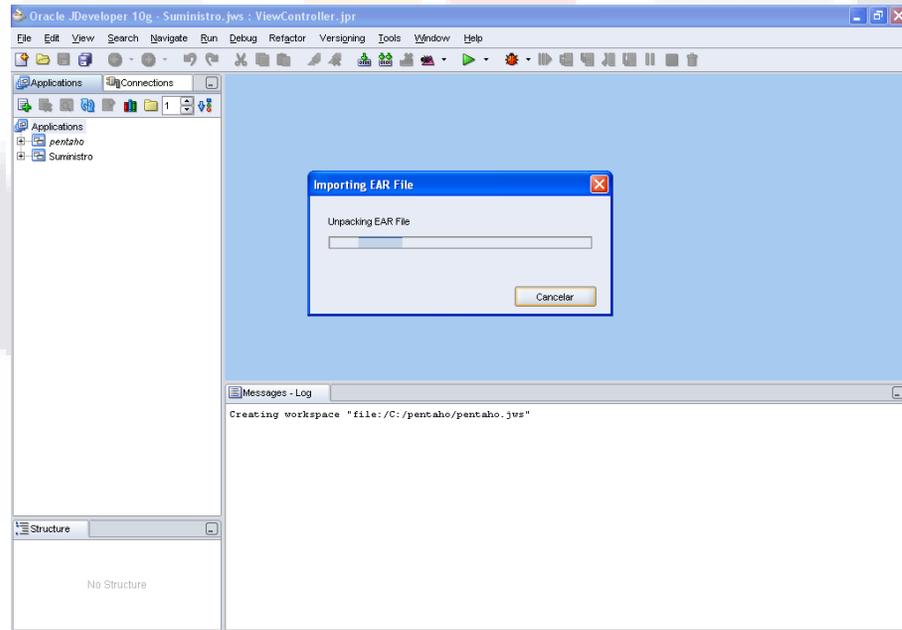


Figura 20. Carga del proyecto al IDE.

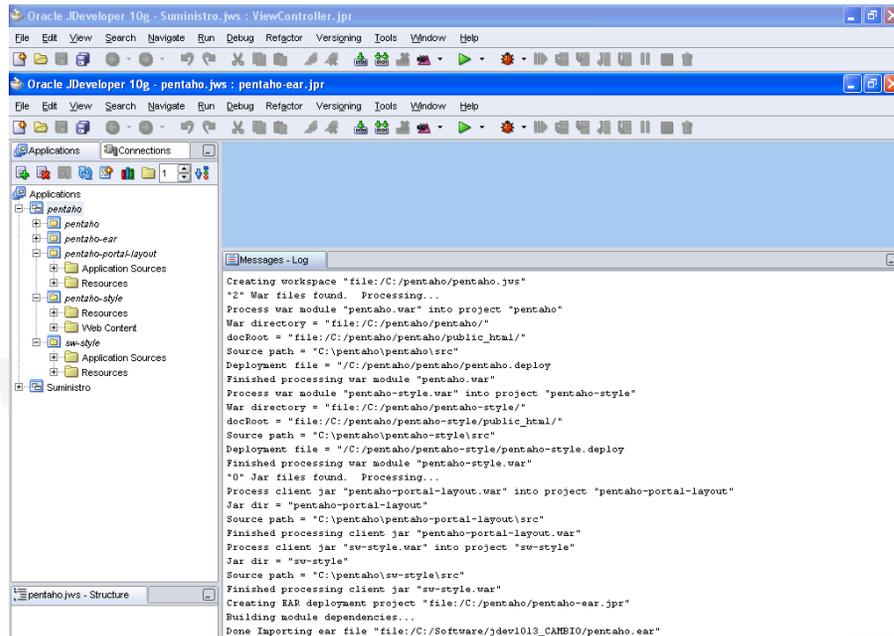


Figura 21. Conclusión del deploy.