



INSTITUTO POLITÉCNICO NACIONAL

**UNIDAD PROFESIONAL INTERDISCIPLINARIA DE INGENIERÍA
Y CIENCIAS SOCIALES Y ADMINISTRATIVAS**

SECCIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

***METODOLOGÍA DE USO DE HERRAMIENTAS DE
INTELIGENCIA DE NEGOCIOS COMO ESTRATEGIA PARA
AUMENTAR LA PRODUCTIVIDAD Y COMPETITIVIDAD DE
UNA PyME***

T E S I S

**QUE PARA OBTENER EL GRADO DE
MAESTRO EN CIENCIAS
EN INFORMÁTICA**

P R E S E N T A :

PAMELA GUTIÉRREZ MELÉNDEZ

DIRECTORES DE TESIS:

DR. FERNANDO VÁZQUEZ TORRES

DR. JAVIER GARCÍA GARCÍA



MÉXICO, D.F.

2012



SIP-14-BIS

INSTITUTO POLITÉCNICO NACIONAL SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REVISIÓN DE TESIS

En la Ciudad de México, D.F. siendo las 18:00 horas del día 23 del mes de octubre del 2012 se reunieron los miembros de la Comisión Revisora de Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación de UPIICSA para examinar la tesis titulada:
"METODOLOGÍA DE USO DE HERRAMIENTAS DE INTELIGENCIA DE NEGOCIOS COMO ESTRATEGIA PARA AUMENTAR LA PRODUCTIVIDAD Y COMPETITIVIDAD DE UNA PYME"

Presentada por el alumno:

GUTIÉRREZ
Apellido paterno

MELÉNDEZ
Apellido materno

PAMELA
Nombre(s)

Con registro:

B	0	9	1	9	6	6
---	---	---	---	---	---	---

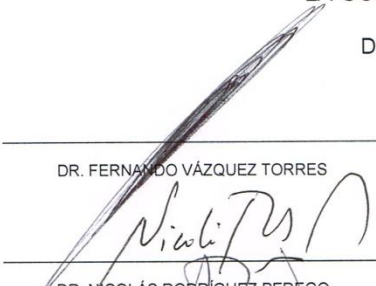
aspirante de:

MAESTRO EN CIENCIAS EN INFORMÁTICA

Después de intercambiar opiniones los miembros de la Comisión manifestaron **APROBAR LA TESIS**, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

LA COMISIÓN REVISORA

Directores de tesis


DR. FERNANDO VÁZQUEZ TORRES



DR. JAVIER GARCÍA GARCÍA


DR. NICOLÁS RODRÍGUEZ PEREGO


M. EN C. ABRAHAM GORDILLO MEJÍA.


M. EN C. GUILLERMO PÉREZ VÁZQUEZ

EL PRESIDENTE DEL COLEGIO DE PROFESORES


M. EN C. GUSTAVO MAZCORRO TÉLLEZ

S. E. P.
I. P. N.



C. P. T. I. C. S. A
SECCIÓN DE ESTUDIOS
DE POSGRADO E
INVESTIGACIÓN



INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

CARTA CESION DE DERECHOS

En la Ciudad de México D.F. el día 23 del mes de Octubre del año 2012, la que suscribe Pamela Gutiérrez Meléndez con el número de registro B091966, adscrita a la Maestría en Ciencias en Informática, manifiesta que es autora intelectual del presente trabajo de Tesis bajo la dirección del Dr. Fernando Vázquez Torres y el Dr. Javier García García cede los derechos del trabajo titulado — Metodología De Uso De Herramientas De Inteligencia De Negocios Como Estrategia Para Aumentar La Productividad Y Competitividad De Una PyME, al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o director del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección madefer@gmail.com. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar fuente del mismo.

ATENTAMENTE:

Pamela Gutiérrez Meléndez

Agradecimientos

A mis padres por impulsarme a llegar hasta este punto, por la paciencia, por la comprensión, por el apoyo que siempre me han dado. GRACIAS, este logro principalmente es suyo.

A mi sunshine, que ha estado a mi lado alentándome todos estos años y que continua haciéndolo. GRACIAS por esas noches que pasaste en vela junto a mí.

A mis amigos y compañeros que compartimos esta experiencia enriquecedora y que nos quedan muchas experiencias más por vivir.

A mis maestros que me apoyaron para que lograra cerrar este ciclo tan importante en mi vida. GRACIAS a todos por sus recomendaciones.

GRACIAS a todos los que aportaron comentarios, consejos, tiempo para el logro de este trabajo.

INDICE GENERAL

GLOSARIO	i
-----------------------	----------

RESUMEN	ii
----------------------	-----------

ABSTRACT.....	iii
----------------------	------------

INTRODUCCIÓN.....	iv
--------------------------	-----------

Capítulo I Marco Teórico	1
1.1 Historia.....	1
1.2 Estado del Arte	2
1.3 Inteligencia de Negocios	7
1.3.1. Componentes	8
1.3.1.1. Fuentes de Datos.....	8
1.3.1.2. ETL	10
1.3.1.3. Bodegas de Datos.....	12
1.3.1.4. KPI.....	14
1.3.1.5. Cubo	15
1.3.1.6. Minería de Datos	16
1.3.1.7. Presentación.....	20
1.4 PYME.....	22
1.4.1. Competitividad	22
1.4.2. Productividad	23
Capítulo II Marco de Referencia	25
2.1 Entorno PYME	25
2.1.1. Competitividad, Productividad y BI.....	25
2.1.2. BI para Pymes.....	28
2.2 Metodologías existentes.....	31
2.2.1. CRISP-DM.....	31
2.2.2. HEFESTO 2.0	35

2.2.3.	METODOLOGIAS AGILES.....	38
2.2.4.	METODOLOGIA KIMBALL	42
2.2.5.	METODOLOGIA INMON.....	46
2.2.6.	X-Warehousing.....	49
2.2.7.	EBAF Conversion Model	50
2.2.8.	Comparativa	50
2.3	Herramientas BI OS.....	52
2.3.1.	Jaspersoft (01)	52
2.3.2.	Pentaho (02).....	53
2.3.3.	OpenI (03).....	54
2.3.4.	SpagoBI (04)	54
2.3.5.	CloverETL (05).....	55
2.3.6.	Palo (06)	55
2.3.7.	PocOLAP (07).....	56
2.3.8.	KNIME (08)	56
2.3.10.	Weka (10)	57
2.3.11.	Rapid Miner (YALE) (11)	57
2.4	Herramientas BI Propietarias.....	58
2.4.1.	Oracle BI (12).....	58
2.4.2.	MicroStrategy (13)	58
2.4.3.	Cognos (14).....	59
2.4.4.	Microsoft (15).....	60
2.4.5.	QlikView (16)	61
2.5	Mejores prácticas	61
2.6	FCE	62
Capítulo III	Propuesta de Solución “Metodología Propuesta Pyme-PyC”	64
3.1	Negocio	66
3.1.1.	Planeación estratégica	67
3.1.1.1.	Misión.....	67
3.1.1.2.	Visión	67
3.1.1.3.	Objetivos	67
3.1.1.4.	Metas.....	67

3.1.1.5.	Factores Críticos de éxito	68
3.1.2.	Procesos de negocio.....	68
3.1.3.	Necesidades de información	68
3.1.4.	KPI.....	68
3.1.5.	Lista de prioridades	68
3.2	Planeación.....	69
3.2.1.	Alcance	69
3.2.2.	Actividades	69
3.2.3.	Recursos	69
3.2.4.	Riesgos.....	70
3.2.5.	Plan.....	70
3.3	Desarrollo.....	70
3.3.1.	Preparación Técnica	70
3.3.1.1.	Selección Herramientas.....	71
3.3.1.2.	Instalación	71
3.3.2.	Modelado DWH.....	71
3.3.2.1.	Análisis de fuente de datos	71
3.3.2.2.	HLDM	72
3.3.2.3.	MLDM	72
3.3.2.4.	LLDM.....	72
3.3.2.5.	Construcción.....	72
3.3.3.	Datos	73
3.3.3.1.	Análisis de datos.....	73
3.3.3.2.	Desarrollo proceso ETL.....	73
3.3.3.3.	Pruebas.....	74
3.3.3.4.	Población histórica DWH y Calendarización ETL	74
3.3.4.	Minería	74
3.3.4.1.	Selección de datos / Conformar Indicadores	74
3.3.4.2.	Modelado de datos / Establecer correspondencias.....	74
3.3.4.3.	Evaluar resultados	74
3.3.5.	Cubo MOLAP	75
3.3.5.1.	Dimensiones y Jerarquías	75

3.3.5.2.	Indicadores	75
3.3.5.3.	Cubos	75
3.3.6.	Presentación	76
3.3.6.1.	Reportes	76
3.3.6.2.	Dashboards	76
3.3.6.3.	ScoreCards	76
3.4	Validación	76
3.5	Implementación	77
Capítulo IV	Caso de Estudio: Pizzería	78
4.1	Negocio	78
4.1.1.	Planeación estratégica	78
4.1.1.1.	Misión	78
4.1.1.2.	Visión	79
4.1.1.3.	Objetivos	79
4.1.1.4.	Metas	79
4.1.1.5.	Factores Críticos de éxito	79
4.1.2.	Procesos de negocio	80
4.1.3.	Necesidades de información	81
4.1.4.	KPI	81
4.1.5.	Lista de prioridades	81
4.2	Planeación	82
4.2.1.	Alcance	82
4.2.2.	Actividades	82
4.2.3.	Recursos	82
4.2.4.	Restricciones y Supuestos	83
4.2.5.	Riesgos	84
4.2.6.	Plan	84
4.3	Desarrollo	84
4.3.1.	Preparación Técnica	85
4.3.1.1.	Selección Herramientas	85
4.3.1.1.1.	Jaspersoft	85
4.3.1.1.2.	SpagoBI	85

4.3.1.1.3.	Openl	86
4.3.1.1.4.	Microstrategy	86
4.3.1.2.	Instalación	86
4.3.1.2.1.	Jaspersoft	86
4.3.1.2.2.	SpagoBI	95
4.3.1.2.3.	Microstrategy	95
4.3.2.	Modelado DWH.....	102
4.3.2.1.	Análisis de fuente de datos	102
4.3.2.2.	HLDM	103
4.3.2.3.	MLDM	103
4.3.2.4.	LLDM.....	104
4.3.2.5.	Construcción.....	106
4.3.3.	Datos	107
4.3.3.1.	Análisis de datos.....	107
4.3.3.2.	Desarrollo proceso ETL.....	109
4.3.4.	Modelo multidimensional	110
4.3.5.	Presentación.....	114

CONCLUSIONES..... 117

ANEXO I. RapidMiner 119

BIBLIOGRAFÍA 124

GLOSARIO

PyME.	Son empresas con un número reducido de ingresos y empleados.
Productividad.	Es la relación que existe entre lo que genera una empresa y los recursos humanos con los que cuenta.
Competitividad.	Es la capacidad de las empresas de hacer frente a la competencia que existe en el mercado.
Dato.	Características aisladas de entidades.
Información.	Son un conjunto de datos que al relacionarse tienen un significado.
Conocimiento.	Es información que es almacenada y puesta a disposición de los interesados para que pueda realizar y/o mejorar sus actividades, permitiéndoles tener un aprendizaje.
OLAP.	Procesamiento analítico en línea, permite utilizar estructuras multidimensionales que permiten agilizar las consultas.
OLTP.	Procesamiento transaccional en línea, permite la administración de la información transaccional generada por aplicaciones operativas.
ROLAP.	Tipo de procesamiento analítico que está basado en un esquema relacional.
MOLAP.	Tipo de procesamiento analítico basado en un esquema multidimensional.
HOLAP.	Tipo de procesamiento analítico que mezcla las características de un esquema relacional con un esquema multidimensional.
KDD.	Acrónimo para Knowledge Discovery in Databases, que consiste en el proceso que se sigue para poder descubrir conocimiento de las bases de datos.
XML.	Lenguaje de marcas extensible, es un lenguaje de etiquetas que permite almacenar información.

RESUMEN

En este trabajo de tesis podremos encontrar la propuesta de una metodología para la implementación de soluciones de inteligencia de negocios enfocada en PyMEs. Esta metodología se realizó en base a la investigación realizada en la que se documentaron diversas metodologías como CRISP-DM ó X-Warehousing.

A lo largo de este trabajo encontramos primeramente contamos con un repaso de los antecedentes de la inteligencia de negocios permitiéndonos conocer cómo ha evolucionado este tipo de soluciones. Seguido de esto podemos conocer el estado del arte de la inteligencia de negocios por la documentación de los trabajos que se han realizado en este campo del manejo de la información. También se hace mención de los principales conceptos relacionados a la inteligencia de negocios con la finalidad que se comprenda el resto de este trabajo. Así mismo como parte del contexto de este trabajo se mencionan la PyME y los principales conceptos que son de relevancia para este trabajo de tesis como son la productividad y la competitividad.

Con la finalidad de tener una mejor referencia y justificación de este trabajo, se hace un repaso de cómo afecta la inteligencia de negocios la productividad y competitividad en las PyMEs. Así mismo se hace un recuento de los beneficios que puede traer para las PyMEs la inteligencia de negocios. Se documentan diversas metodologías de inteligencia de negocios que se utilizan de referencia para la metodología propuesta en este trabajo de tesis. Recopilamos información básica de diversas herramientas de inteligencia de negocios que se encuentran en el mercado siendo propietarias y de tipo Open Source.

Se hace la propuesta de una metodología de inteligencia de negocios enfocada en las PyMEs. Esta metodología pretende que de acuerdo al nivel de madurez de la empresa, se introduzcan ciertas prácticas de inteligencia de negocios, de tal manera que cuando la empresa llegue a un alto nivel de madurez, esta pueda contar con una solución de inteligencia de negocios lo suficientemente robusta.

Finalmente en este trabajo encontraremos la implementación de esta metodología en una empresa dedicada a la comida rápida. Vemos el uso de diversas herramientas como SpagoBI, Jaspersoft y Microstrategy.

Así mismo se hace una pequeña muestra del uso de otro tipo de herramientas open source para la inteligencia de negocios enfocadas en la parte de minería de datos como lo es rapidminer.

ABSTRACT

In this thesis work we find the proposal of a methodology for the implementation of business intelligence solutions focused on SMEs. This methodology was performed based on research conducted in the various methodologies that were documented as CRISP-DM or X-Warehousing.

Throughout this work, we first find an overview of the history of business intelligence has evolved allowing us to know how this type of solution. Following this we can know the state of the art business intelligence for the documentation of the work that has been done in the field of information management. Mention is also made of the main concepts related to business intelligence in order to understand the rest of this paper. Also as part of the context of this work are listed SMEs and the main concepts that are relevant to this thesis, such as productivity and competitiveness.

In order to have a better reference and justification of this work, there is a review of how intelligence affects business productivity and competitiveness of SMEs. It also takes account of the benefits it can bring to SMEs business intelligence. Documented various business intelligence methodologies that are used as a reference for the methodology proposed in this thesis. We collect basic information on various business intelligence tools that are in the market to be proprietary and Open Source.

The objective of this work is to propose a methodology of Business Intelligence focused on SMEs. This methodology pretends to introduce practices of Business Intelligence in accordance to a maturity level in the SME, in order to evolve the solution matching to its maturity level such that it can have a robust enough solution.

Finally in this work we can find the implementation of this methodology on a SME dedicated to sell fast food. We can observe the use of different tools as SpagoBI, Jaspersoft and Microstrategy. Also we can read about the usage of another tool of data mining called RapidMiner.

INTRODUCCIÓN

La inteligencia de negocios es una solución que permite a las empresas acceder a información que les permite tomar decisiones, esta información generalmente proviene de diversas fuentes por lo que la solución de inteligencia de negocios se encarga de integrarla y transformarla para almacenarla en una estructura especial conocida como Data Warehouse. El Data Warehouse cumple con ciertas características para su construcción contando principalmente con 2 tipos de elementos, las tablas de hechos y las dimensiones. Este tipo de modelado nos facilita el usar herramientas multidimensionales que nos permiten hacer el cruce de los hechos con diversas dimensiones que les den un sentido real para el negocio. El resultado de una solución de inteligencia de negocios generalmente lo vemos de manera visible por medio de herramientas web que nos permiten visualizar los reportes que se generen con la información obtenida de la solución.

Este tipo de soluciones es muy popular para empresas de gran tamaño, sin embargo y debido a que gran parte de las empresas que existen en México son PyMEs. Este trabajo de tesis busca ofrecer a este amplio mercado una metodología que les permita integrar prácticas de inteligencia de negocios.

Se busca cumplir principalmente con dos objetivos:

- Proporcionar a las PyMEs una metodología que les permita la integración de una solución de inteligencia de negocios que sea adecuada a su nivel de madurez, con lo cual buscamos que la implementación sea amigable para que sean aceptados los cambios por los empleados de la empresa de una manera más transparente.
- Permitir que estas empresas al tener a su alcance este tipo de metodologías y por lo tanto este tipo de soluciones, sean capaces de incrementar su productividad lo que las llevara a ser más competitivas.

Otro de los puntos que hace que este tipo de soluciones se ven inalcanzables para las PyMEs, son los costos que implica el construir una solución de este tipo. Es por esto que este trabajo de tesis incorpora la mención de diversas herramientas de tipo Open Source como una opción para tener accesible herramientas que permiten el desarrollo y explotación de una solución de inteligencia de negocios.

Capítulo I Marco Teórico

En este capítulo encontraremos los elementos básicos que se necesitan para comprender el tema objeto de este trabajo de tesis. Tenemos una breve reseña histórica de la inteligencia de negocios que nos permite posicionarnos en una situación actual. Así mismo se puede consultar el estado del arte lo que nos permite conocer en qué dirección se están dando los trabajos más recientes relacionados. Están disponibles los elementos que componen un esquema de una solución de inteligencia de negocios y por estar este trabajo enfocado en las PyMEs se hace mención del concepto de PyME, productividad y competitividad.

Historia

Encontramos el primer antecedente de la inteligencia de negocios en siglo V antes de Cristo un escrito llamado “El arte de la guerra” escrito por Sun Tzu puesto que este documento muestra la importancia de la inteligencia y la información para la solución de problemas.

Desde este punto y hasta la década de los 60's la Inteligencia de Negocios comienza a tomar un rumbo dentro de las Tecnologías de Información es con el surgimiento del concepto de los Sistemas de Soporte a Decisiones, que provee a los usuarios con un acceso sencillo a modelos de decisión con la finalidad de apoyar las tareas de toma de decisiones (Benki & Papastathopoulos, 2009). Así mismo a finales de esta década surge un concepto principal para la evolución de la Inteligencia de negocios, surge el concepto de base de datos.

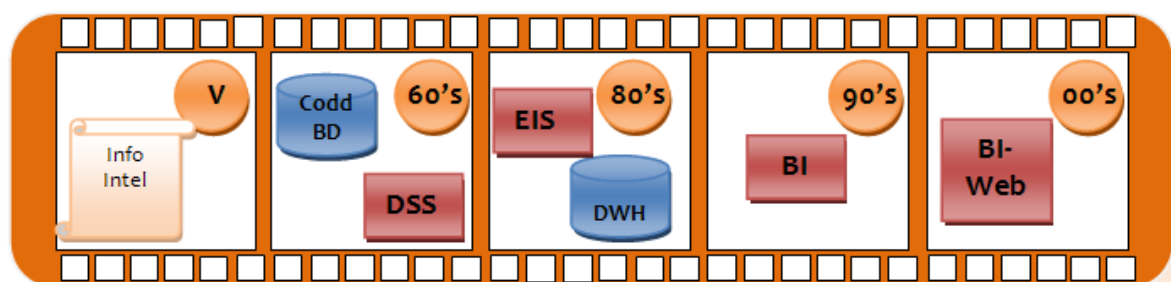
En la década de los 80's Ralph Kimball y Bill Inmon crean el concepto de DataWarehouse, proporcionando una visión del concepto muy completa y la cual se sigue considerando como base en los desarrollos de Inteligencia de Negocios. En esta misma década surgieron las primeras aplicaciones de reporte, sin embargo tanto las aplicaciones de base de datos como de reporte no permitían la independencia del usuario final por la complejidad en su manejo. A finales de esta década se formaliza el concepto de Inteligencia de Negocios por Howard Dresner quien considera el BI como un conjunto de metodologías cuyo principal objetivo es aumentar la eficiencia de las empresas. De esta misma manera durante esta década existe una evolución de los sistemas surgiendo los sistemas de Información Ejecutivos.

Para la década de los 90's contando con un poco más de claridad sobre lo que consistía la Inteligencia de Negocios, lo que provoca un surgimiento desmedido de herramientas de BI que hacían que el acceso a la información fuera más sencillo. Para entonces ya se identifican Sistemas de BI como parte de la evolución.

Durante el año de 1998 en México se empiezan a utilizar herramientas de BI con finalidades de reporte, sin embargo la funcionalidad de dichas herramientas era un poco compleja lo que complicaba su uso y administración, así mismo era complicado el desarrollo de cubos. Además las herramientas de BI contaban con ciertas limitantes lo que forzaba a las empresas a realizar desarrollos extraordinarios para poder cubrir completamente la necesidad del negocio.

Con la llegada del año 2000 se consolidan algunas herramientas de BI y se amplía el espectro de fuentes de información involucradas en el proceso de BI, tomando no solo fuentes de información estructuradas. Con el auge de las redes surge la necesidad de los sistemas de BI de evolucionar para poder presentar los resultados de BI vía Web y los avances en tecnología permiten que los usuarios puedan acceder a su información a través de diversos dispositivos, los usuarios pueden definir el formato en el que quieren ver su información. En conclusión en esta última evolución se ha buscado independizar al usuario del departamento de Sistemas.

En México ha existido una tendencia por generar una gran cantidad de consultores de calidad especializados en Inteligencia de Negocios, principalmente expertos en ciertas herramientas de BI, aunque también hay expertos en ciertas fases de la Inteligencia de Negocios como puede ser el front-end, ETL, definición y administración.



Fuente: Diseño propio

I-1. Historia BI

Estado del Arte

En el año de 2007 (Watson & Wixom, 2007) consideraron que el marco del BI se conformaba principalmente por 2 actividades:

1. Datawarehousing
2. BI

En la parte de Datawarehousing consideraron todo lo que involucra el poner la información dentro del DataWarehouse, considerando que esta actividad absorbía 80% del tiempo de cada proyecto. La parte de BI involucraba la explotación de la información por parte de los usuarios.

Entre los beneficios que resaltaban la reducción de inversión en infraestructura debido a que se eliminaban los datos redundantes y los datos duplicados, haciendo la explotación de la información más eficiente, permitiendo la estrategia de negocios.

Desde ese entonces se ha trabajado en el desarrollo de BI en tiempo real y en soportar la Administración de Procesos de Negocio (BPM).

John Maning nos presenta las siguientes tendencias de BI:

1. Se pretende hacer más amigable el front-end para los usuarios, y que este llegue a más usuarios cada vez, sin que estos requieran de alguna capacitación.
2. También se busca no solo apoyar la toma de decisiones en niveles estratégicos, sino que se incluyan los procesos del BI en procesos operativos lo que también impactaría en toda la empresa, esto bajo un esquema de tiempo real, entregando información oportuna y relevante a cada usuario dependiendo de sus necesidades reales.
3. Aprovechando los beneficios del BI ciertas empresas lo pueden utilizar como una oportunidad para obtener ingresos ofreciendo mantener y administrar toda la infraestructura de BI para ofrecer como servicio la entrega de la información necesaria por los clientes.
4. Aprovechar la integración de los datos para la implementación de motores de búsqueda potentes que permita a los usuarios realizar búsquedas sobre temas de la empresa y obtener resultados tan acertados como si realizaran la búsqueda en su motor de búsqueda predilecto.
5. Siguiendo el avance de la tecnología hay que explotar los ítems móviles como son los Smartphone y los ipads, para que los usuarios tengan en cualquier momento un resumen de la información que realmente les sea útil.

Por otro lado, Diego Arenas en su artículo “Predicciones BI 2010-2012” indica las siguientes tendencias:

- Existirá una tendencia por realizar proyectos de BI pequeños que tengan un gran impacto en poco tiempo,
- Identifica la necesidad de trabajar proyectos de integración de datos para ser trabajados como soluciones por si solas.
- Una apertura a las Pymes ofreciendo costos de adquisición de las herramientas menores.
- Se explotará el uso de los webinar por parte de herramientas propietarias.
- Cada vez será más sencilla la administración de las herramientas de BI.
- Crecimiento y tal vez estandarización de la jerga de Business Intelligence

- Uso de herramientas propietarias en Pymes
- Uso de herramientas tipo OS.
- Crecimiento del mercado
- Proyectos con mayor calidad debido a las experiencias acumuladas.

Para el mercado de las PyMEs se sabe que estas requieren estructuras sencillas de Inteligencia de Negocios a bajo costo y que no les requieran hardware y software adicional. Es por esto que se tiene la opción de ofrecer una mezcla conceptos con la finalidad de generar una solución integral que cubra sus necesidades, apoyándose en el uso de herramientas Open Source. Basándose en arquitecturas ROLAP, herramientas ETL y el modelado y análisis de procesos (Grabova, Darmont, Chauchat, & Zolotaryova, 2010).

Existe un trabajo de investigación en el que se habla sobre el uso de agentes en sistemas organizacionales de soporte a la toma de decisiones (Fu, Xu, Song, & Xin, 2008). En este trabajo se muestra la siguiente arquitectura:

1. Capa de interfaz. Es la presentación de la información.
2. Capa de decisión. Donde se encuentra el Data Warehouse, las bases de datos, los modelos y una base de problemas.
3. Capa núcleo. Elige el modelo de decisión y realiza todo el proceso de decisión generando el resultado para el usuario.

Basando todo su funcionamiento en 2 agencias:

- Agencia de aplicación: Recibe las peticiones y se encarga de la comunicación entre agentes. Está formada por:
 - Agente de coordinación.
 - Agente de información.
 - Agente de análisis.
 - Agente de búsqueda.
 - Agente de distribución.
- Agente de decisión: Responsable del proceso de las actividades de decisión proporcionando resultados al agente de distribución. Está formado por:
 - Agente de coordinación.
 - Agente de procesamiento.
 - Agente de razonamiento.
 - Agente de aprendizaje.

Existe otro estudio, con una arquitectura similar al anterior, enfocado en aplicar Inteligencia de Negocios con agentes con el objetivo de reducir los costos de una solución de Inteligencia de Negocios. En uno de estos estudios (Feng, Li, Gao, Liu, & Xu, 2010) la arquitectura del sistema propuesto se divide en tres niveles:

1. Capa de interfaz de usuario. Está enfocada en la estandarización de peticiones y la visualización de resultados.
2. Capa de función núcleo. Está enfocada en realizar tareas de sentencias de consulta, análisis y minado.
3. Capa de recursos de datos. Está enfocada en proporcionar servicios básicos de datos a la capa de función núcleo, estableciendo y manteniendo metadata.

Apoyándose estas capas en el uso de 7 tipos de agentes: User agent, User management agent, Task assignment agent, OLAP agent, DM agent, Data source management agent, Data source agent. El funcionamiento del sistema consiste en que el User management agent hace la validación de logeo del usuario creándole un User agent. Entre estos agentes se comunican las peticiones y el User management agent lo transfiere al Task assignment agent el cual envía el mensaje al agente correspondiente de acuerdo a la petición finalmente la respuesta a la petición regresa al user agent para mostrar el resultado al usuario.

Así mismo existen otros trabajos (Li, Li, & Zhu, 2010) enfocados en el minado del conocimiento con la finalidad de dar soporte a la toma de decisiones. De acuerdo a este trabajo, proponen un marco basado en Extenics (una nueva disciplina enfocada en el trato de problemas contradictorios formulando un modelo, analizándolos en objetos, las características y las métricas relacionadas) y la administración de conocimiento que consiste en 6 pasos:

1. Colección de información
2. Expresar todo el conocimiento e información en elementos de materia y son almacenados en la base de elementos básicos
3. Se elige un elemento materia como punto de partida
4. Se calcula el siguiente elemento de materia
5. Se hace un ciclo hasta considerar todos los elementos materia
6. Se valida las relaciones realizadas

La toma de decisiones requiere de información comprensible y conocimiento. De acuerdo a este artículo se puede construir un cubo de información que contenga integrada la información de la fuente de datos al cual se le pueda aplicar el método de extensión que nos permita generar estrategias que resolverán los problemas contradictorios.

Existe otro estudio (Li, Zhu, & Pan, 2010) que se basa en la combinación de Extenics y la minería de datos y cuyo objetivo es cultivar conocimiento en las PyMEs. El método de siembra del conocimiento consiste en identificar el problema, para posteriormente identificar el conocimiento relativo a este problema.

El método para cultivar el conocimiento parte de la colección de información detallada. Para esto es útil un cubo de información que permite integrar la información de diversas fuentes de datos de manera sistemática. Después se puede aplicar el método de extensión de tal manera

que el conocimiento sembrado se extienda generando un conocimiento de más nivel, mejorando la toma de decisiones en las PyMEs.

El proceso consiste en lo siguiente:

1. Distribuir la información sobre el problema de negocio y las condiciones bajo las cuales se establece el objetivo.
2. Agregar información o conocimiento relacionado a problemas en la plataforma.
3. Conectar información o conocimiento por elemento, características y métricas.
4. Evaluar el conocimiento clasificado para seleccionar el conocimiento factible para transformaciones.
5. Recibe entregables
6. Validar que se puedan tomar decisiones y resolver problemas

Proponen un árbol de conocimiento para almacenar el conocimiento generado por el cultivo de conocimiento por medio de un algoritmo propuesto. Este algoritmo localiza el problema con sus elementos base para después recolectar información y conocimiento resultado del proceso de minería de datos. Se selecciona el conocimiento como semilla y se cultiva como elementos base. Entonces se evalúa si el árbol de conocimiento crecerá o si es necesario más conocimiento mediante interacción humana. Es un proceso iterativo hasta cubrir las necesidades de los tomadores de decisiones.

En otro estudio (Xie, 2009) se propone un modelo que este enfocado en la mejora del conocimiento. Este modelo de administración de conocimiento está basado en minería de datos. El modelo inicia con la preparación de los datos, para esto primero se debe definir el tema que se trabajará, esto para poder establecer los requerimientos y propósitos que cubrirá la minería de datos y finalmente se define el problema que se quiere resolver. Se seleccionan las fuentes de datos de acuerdo a los requerimientos, se procesan los datos integrándolos y transformándolos. Hace uso de diversas técnicas de minado de datos como son redes neuronales, árboles de decisión, modelos lógicos, clasificación y análisis cluster entre otros modelos estadísticos.

Hace mención de un portal de conocimiento empresarial en el que los empleados podrán contactarse entre ellos y puedan ayudarse. Este portal provee 6 funciones:

1. Adquisición de conocimiento
2. Almacenamiento del conocimiento
3. Concentración y compartimento del conocimiento
4. Aplicación e innovación del conocimiento
5. Recuperación del conocimiento
6. Evaluación del desempeño del conocimiento

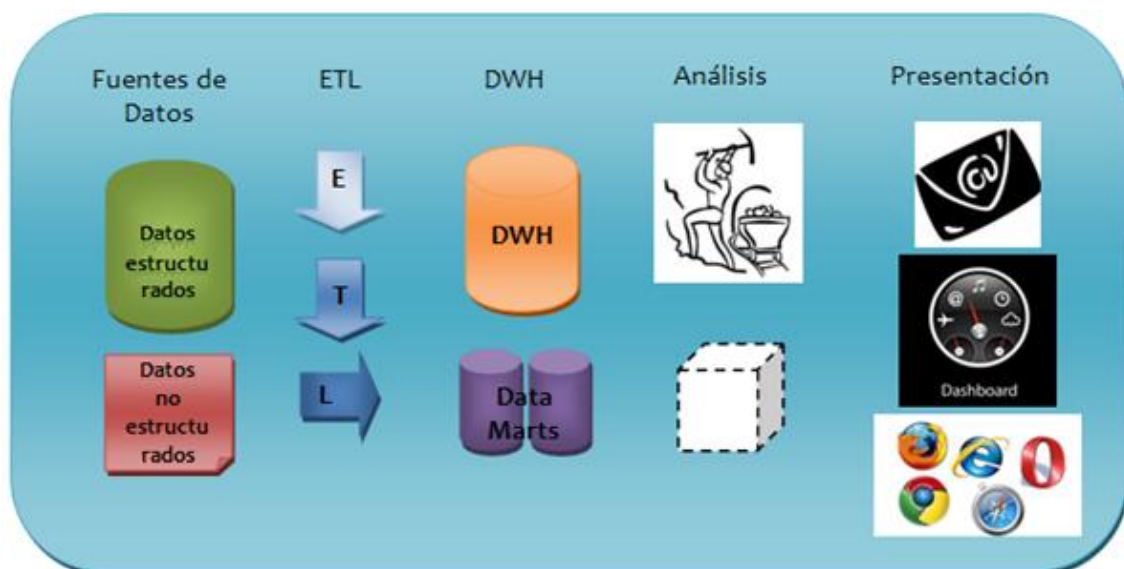
Para que esta propuesta funcione también se deben implementar ciertas estrategias que permitan un buen funcionamiento de la administración del conocimiento. Por ejemplo menciona la importancia de tener una buena administración de personal puesto que el personal es quien mantendrá el conocimiento en constante actualización y estos deben estar acostumbrados a cooperar con sus compañeros. La plataforma debe ser de fácil uso proporcionando información en tiempo real. Es necesario mantener motivados a los empleados por lo que recomienda la implementación de un sistema de reconocimiento y castigo para motivar a los empleados en la protección intelectual así como reducir la pérdida de talentos. Se debe regular el proceso de aprendizaje de los empleados.

Inteligencia de Negocios

La Inteligencia de Negocios es una solución tecnológica cuyo objetivo primordial es el apoyar la toma de decisiones, esta solución tecnológica está conformada por metodologías y herramientas que permiten cumplir dicho objetivo. Sin embargo, la Inteligencia de negocios puede adaptarse a diversos alcances dependiendo de las necesidades del negocio.

La inteligencia de Negocios busca transformar los datos en información para finalmente transformar la información en conocimientos.

La Inteligencia de Negocios en términos generales debiera contener los elementos que se muestran en la imagen.



Fuente: Diseño propio

I-2. Contexto BI

Las herramientas de Inteligencia de Negocios están clasificadas principalmente en dos grupos:

- Soluciones enfocadas en Web

Este tipo de herramientas permiten la implementación de una solución de BI a partir de herramientas Web. Esto permite el uso de un nuevo concepto llamado Web warehousing, en el cual se conservan las principales características del Warehousing organizando y administrando los datos con la diferencia de que estos no se almacenan. Esto se logra con el uso de navegadores, servicios web y XML. Este tipo de soluciones son una buena opción para los datos no estructurados. Un ejemplo de esto es Cloud Computing.

- Soluciones enfocadas en Memoria

Este tipo de soluciones están enfocadas en el almacenamiento de los datos, este almacenamiento puede ser semántico, sintáctico y aproximado. En cuestión de cubos el principal es el semántico para el cual existe una técnica llamada "Quotient Cube". El problema de este tipo de soluciones es que se requiere de grandes espacios para el almacenamiento de la información, aunque en últimas fechas los costos de memoria han reducido lo que permite que estas soluciones sigan siendo una opción. Este tipo de soluciones se apoyan de bases de datos MMDB las cuales permiten un tiempo de respuesta corto pero sufre de vulnerabilidad con los datos. Dentro de esta solución encontramos las bases de datos de vectores que permiten que cualquier dato juegue el rol de dimensión y métrica al mismo tiempo, el problema es que se dificulta el contar con visibilidad sobre la metadata (Grabova, Darmont, Chauchat, & Zolotaryova, 2010).

1.1.1. Componentes

En esta sección se hace mención de los componentes básicos que conforman una solución de inteligencia de negocios con la finalidad de tener una mejor comprensión de este trabajo de tesis.

1.1.1.1. Fuentes de Datos

Las fuentes de datos son aquellas que contienen los datos en su mayor nivel de detalle, generalmente vienen directamente de la operación diaria del negocio pero puede tener diferentes presentaciones. Estas conservan la información que el cliente considera relevante según lo que ha definido pudiendo apoyarse o no de Tecnologías de Información. Por lo tanto las fuentes de datos se pueden clasificar en 2 grupos:

- Fuentes de Datos Estructuradas. Son aquellas fuentes de información que tienen cierto orden y son fácilmente manipuladas para todo el proceso de BI.

- Fuentes de Datos Estructuradas. Son principalmente textos.

Así podemos decir que las fuentes de datos que pueden existir son:







Fuente	Descripción	Ilustración
Archivos de Texto Planos	Son archivos de tipo texto los cuales pueden ser estructurados o no estructurados.	
Hojas de cálculo	Son hojas que pueden simular la estructura de una tabla por medio de columnas y por manejar filas.	
XML	Es un lenguaje que maneja etiquetas que permite estructurar los datos.	
Bases de Datos Transaccionales OLTP	Son almacenes de datos que están diseñados para trabajar de manera eficiente en inserciones actualizaciones y borrados de registros. Generalmente son alimentados por aplicaciones que funcionan a nivel de operación del negocio.	
Bases de Datos de Aplicaciones Especializadas	Las bases de datos alimentadas por aplicaciones especializadas como pueden ser CRM, ERP, SCM, BPM, etc. Pueden servir como fuente de información.	
MDX	La información que pueda traer un cubo de información pudiera servir también de fuente de información	

Tabla 1. Fuentes de Datos

Las fuentes de datos listadas anteriormente solo son algunas de las existentes, siendo estas las más comunes. Como podemos ver dependiendo de la complejidad que conlleve los procesos de negocio puede que este requiera de diversas fuentes de datos, además de que lo ideal para cualquier desarrollo sería que estas fuentes de datos se encontraran bien documentadas para poder trabajar sobre ellas, aunque en la vida real, es poco común encontrar fuentes de datos documentadas.

Así mismo existe la posibilidad de que no se utilicen todas las fuentes de datos existentes, tal vez exista redundancia de datos y solo se deba usar alguna fuente de datos o tal vez el alcance del desarrollo no haga necesario que se utilicen todas las fuentes. Sin embargo es necesario realizar un buen análisis de las necesidades de información del negocio para hacer la mejor selección de las fuentes de información y poder cubrir las necesidades de los usuarios finales.

Estas fuentes de información son de suma importancia porque no solo nos proporcionaran los datos necesarios para cubrir las necesidades de información de los usuarios finales, sino que también nos ayudara a comenzar a modelar nuestro destino de dicha información que principalmente seria una bodega de datos.

Para las fuentes de información debiera existir un experto sobre la temática que trate cierto grupo de entidades presentes en la fuente de información. Esto nos permitirá encontrar un sentido a la arquitectura de las fuentes de datos en dimensiones y hechos, para proceder a modelar el destino y poder trabajar sobre las transformaciones necesarias y sus mapeos correspondientes con el destino diseñado.

1.1.1.2. ETL

El proceso ETL en pocas palabras consiste en la Extracción, Transformación y Carga de la información desde nuestras fuentes de datos hasta nuestra bodega de datos. Esto con la finalidad de que garantizar que los datos que sean almacenados en nuestro destino cumplan con ciertas validaciones y formatos que nos permitan asegurar la integridad, consistencia y no redundancia de la información.

Todo el proceso ETL se ajusta a los requerimientos definidos por el usuario final quien es el que expresa que preguntas quiere responder con la información y quien nos puede indicar de donde viene esta información y probablemente en qué condiciones debiera cumplir estos datos para ser información para ellos.

La extracción de datos consiste en realizar una copia de los datos mediante una selección de lo requerido. Se pueden extraer tablas completas, ciertos campos, archivos completos, ciertos campos, ciertos registros, etcétera dependiendo de los requerimientos establecidos. Estas extracciones se hacen sobre las fuentes de información con las que se cuente y que hayan sido seleccionadas para alimentar la bodega de datos.



Fuente: Diseño propio

I-3. Proceso ETL - Extracción

La transformación es la fase del proceso que implica un mayor esfuerzo pues es en esta fase en la que se hace la limpieza de los datos para cumplan con los requisitos especificados antes de iniciar el proceso.

Entre las principales transformaciones que se pueden hacer son cambios en los tipos de datos, obtener nuevas columnas de las existentes, manejo de los nulos, entre muchas otras. Esta fase del proceso es la que nos permitirá llevar los datos de las fuentes de tal manera que estos cumplan con los requisitos necesarios para poder llegar a la siguiente fase del proceso y no se generen errores que lleven a la pérdida de datos permitiéndonos entregar datos de calidad para los usuarios finales.

Esta fase de proceso debe poder garantizar que la información sea correcta, consistente, completa y no ambigua.



Fuente: Diseño propio

I-4. Proceso ETL - Transformación

La carga de los datos consiste en insertar los datos transformados y que son considerados datos de calidad en nuestra bodega de datos. Esta carga generalmente se debe realizar en

cierto orden, puesto que la bodega de datos pudiera contener ciertas restricciones que si no se cumplen durante la carga el proceso fallaría y no se podría hacer la carga.



Fuente: Diseño propio

I-5. ETL - Carga

1.1.1.3. Bodegas de Datos

Las bodegas de datos tienen son acepciones principales:

1. Colección de datos que están orientados a temas, integrados, no volátiles y que varían en el tiempo y cuya finalidad es servir de soporte en la toma de decisiones. Estos datos contienen granularidad de los datos corporativos. (Inmon, 2002)
2. Es el conglomerado de datos organizacionales en áreas de desarrollo y presentación, donde dichos datos provienen de la operación y son manipulados para el análisis que el usuario final requiera. (Kimball & Ross, 2002)

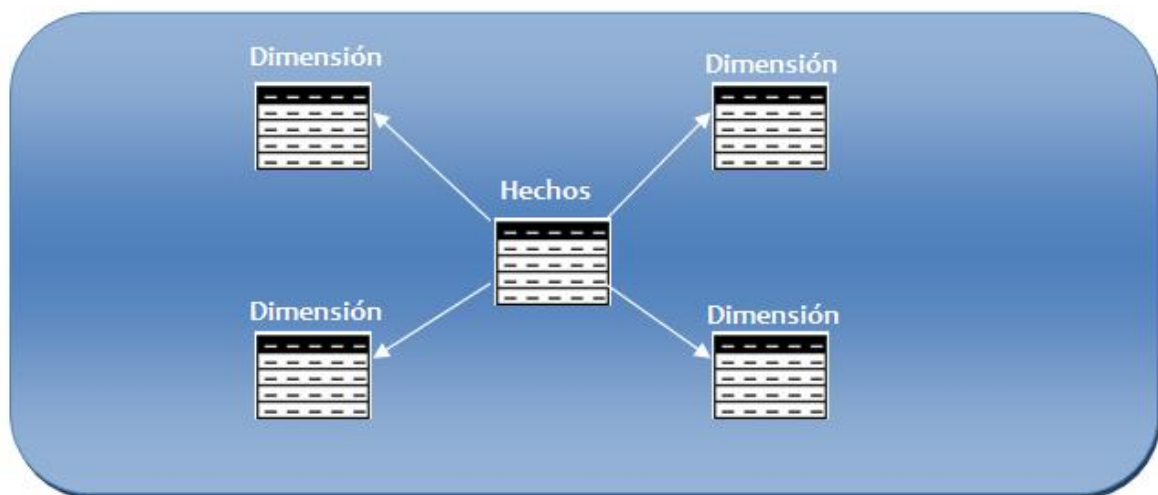
De acuerdo con estas definiciones sabemos que el Data Warehouse debiera contener las principales áreas del negocio las cuales de preferencia deberían estar identificadas en el modelo de negocio. Generalmente estas áreas son comunes y pueden ser:

- Clientes
- Productos
- Transacción o actividad
- Políticas
- Quejas
- Cuenta

Para cada una de dichas áreas se debe identificar las tablas de dimensión y las tablas de hechos. Las tablas de dimensión son aquellas tablas que contienen atributos de los datos, que nos permiten darle sentido a los datos numéricos, por ejemplo Clientes, Productos. Las tablas de hechos contienen datos con los que se pueden hacer operaciones matemáticas por ejemplo Ventas.

Estas tablas se pueden organizar en dos tipos de modelado:

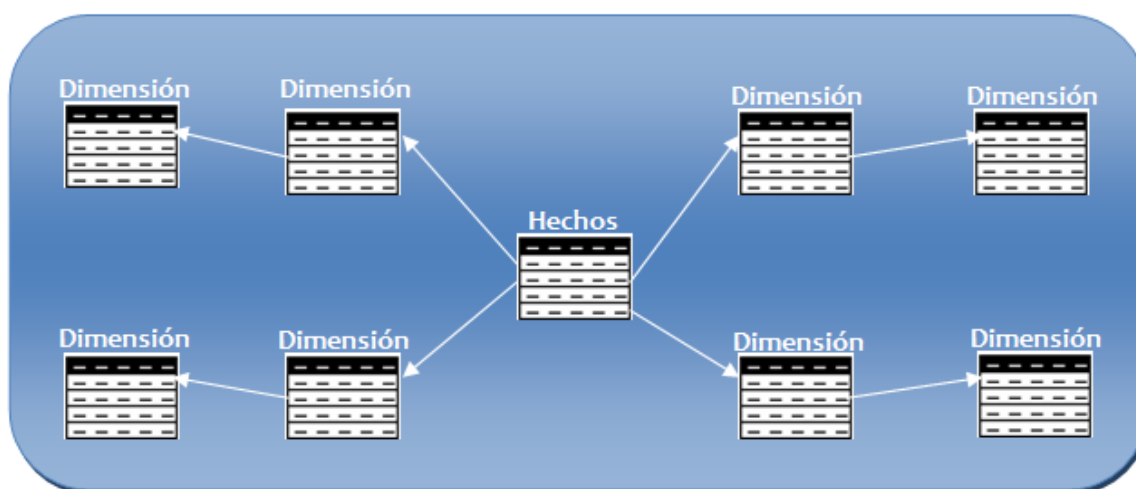
- Estrella. Las tablas de dimensiones no se encuentran normalizadas por lo que una tabla de dimensión contiene diversos campos identificadores con sus respectivos campos de descripción. Cada llave de una tabla de hechos se encuentra ligada a una tabla de dimensión. Este tipo de modelado permite la creación de jerarquías lo que permite navegar por la información (Boussaid, Ben Messaoud, Choquet, & Anthoard, 2006).



Fuente: Diseño propio

I-6. Esquema estrella

- Snowflake Las tablas de dimensiones se encuentran normalizadas por lo que cada tabla contiene un campo identificador, un campo de descripción y un campo que funcione como llave foránea que la une a otra tabla de dimensión. Esta estructura permite que las tablas se unan a la tabla de hechos mediante este campo identificador. Las llaves de la tabla de hechos se encuentran ligadas a varias dimensiones de manera directa e indirecta (Boussaid, Ben Messaoud, Choquet, & Anthoard, 2006).



Fuente: Diseño propio

I-7. Esquema Copo de Nieve

1.1.1.4. KPI

Para poder entender lo que son los KPI (Key Performance Indicators) primeramente debemos comprender lo que es una métrica. Una métrica es una medida numérica que representa una parte de los datos del negocio en relación a una o varias dimensiones a través de sus jerarquías.

Entonces un KPI es una métrica que está ligada a los objetivos de la empresa y normalmente se presentan en forma de porcentajes y tienen la finalidad de facilitar al usuario de negocio el identificar si están funcionando los planes (Gonzalez, 2006).

Los KPI al ser los determinantes del éxito o fracaso del negocio requieren una atención especial y constante para monitorear los avances de la empresa y tomar decisiones inmediatas que corrijan el rumbo del negocio hacia lo establecido en los objetivos.

Para facilitar su lectura, los KPI's son presentados mediante ScoreCards, y están dirigidos a los altos mandos de las organizaciones.

Según (Parmenter, 2010) existen principalmente 4 tipos de métricas con las cuales se debiera seguir la regla 10/80/10 que establece que el negocio debe contar con:

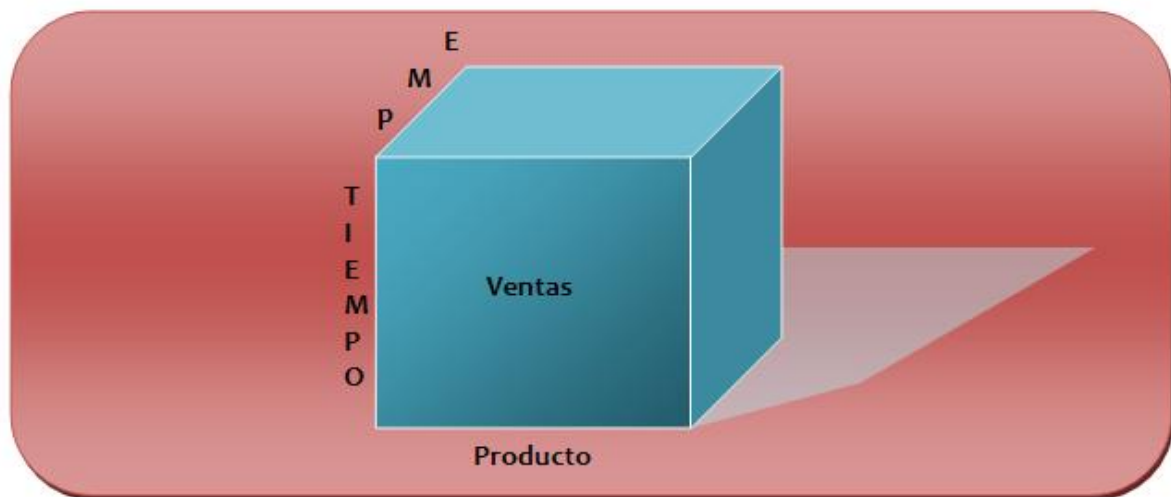
1. 10 IRC. Indicadores Clave de Resultado que Reflejan los resultados históricos y sirven para mostrar el crecimiento de la empresa hasta el día de hoy. Ejemplo: Satisfacción del cliente, Tasa de retorno de capital.
2. 80
 - a. IP. Son los que indican al personal que es lo que debe hacer y miden el performance del negocio, permitiendo la alineación con la estrategia de la empresa. Ejemplo: Quejas recibidas de clientes selectos, Entregas demoradas a clientes selectos
 - b. IR. Muestran que es lo que se ha hecho, muestran resultados concretos. Ventas diarias, cobranzas realizadas en la semana.
3. 10 KPI. Reflejan el performance inmediato del negocio y las acciones que hay que llevar a cabo para poder cumplir con los objetivos establecidos. Ejemplo: Renuncias de Empleados.

Para la definición de KPI's se debe buscar que estos cumplan con 7 características:

1. Métricas no financieras
2. Requieran ser medidas constantemente
3. Son dirigidas para los altos mandos
4. Claramente indica que acción debe ser tomada
5. Se identifica claramente a los responsables de cada KPI
6. Tienen un impacto significativo
7. Impulsa a tomar acciones apropiadas

1.1.1.5. Cubo

Un cubo de datos permite a los datos ser modelados y vistos en diferentes dimensiones para lo cual es necesario que trabaje bajo un modelo multidimensional, basado en dimensiones y hechos. En otras palabras un cubo procesa la información de acuerdo a un diseño específico que empata con los requisitos de información determinado de tal manera que nos permite ver de acuerdo a las necesidades establecidas, por ejemplo podemos ver las ventas por empleado por producto durante el mes de Septiembre, de esta manera mezclamos 3 dimensiones y un hecho.



Fuente: Diseño propio

I-8. Cubo

Esta organización de los datos nos permitirá conocer la productividad de los empleados según las ventas que realice de cada producto para identificar si los empleados no están siendo productivos y tomar medidas, identificar que empleado es más productivo y proporcionarle un bono o un reconocimiento a su esfuerzo.

Los cubos se manejan bajo un esquema OLAP (Online Analytical Processing), que a diferencia del esquema OLTP (Online Transaction Processing), su principal objetivo es optimizar el tiempo de las consultas que se realizan sobre una base de datos. De este existen dos versiones:

- A. MOLAP. Trabaja con un motor especial basado en almacenamiento multidimensional que permite visualizar la información en diversas dimensiones.
- B. ROLAP. Trabaja con un manejador de base de datos, por lo que funciona como intermediario entre el sistema OLTP y el front-end del usuario final.

1.1.1.6. Minería de Datos

La minería de datos es parte de un proceso cuyo objetivo es la extracción de conocimiento a partir de un conjunto muy grande de datos. El proceso general es conocido como KDD (Knowledge Discovery from data) y está conformado de diversas fases entre las cuales encontramos a la minería de datos.

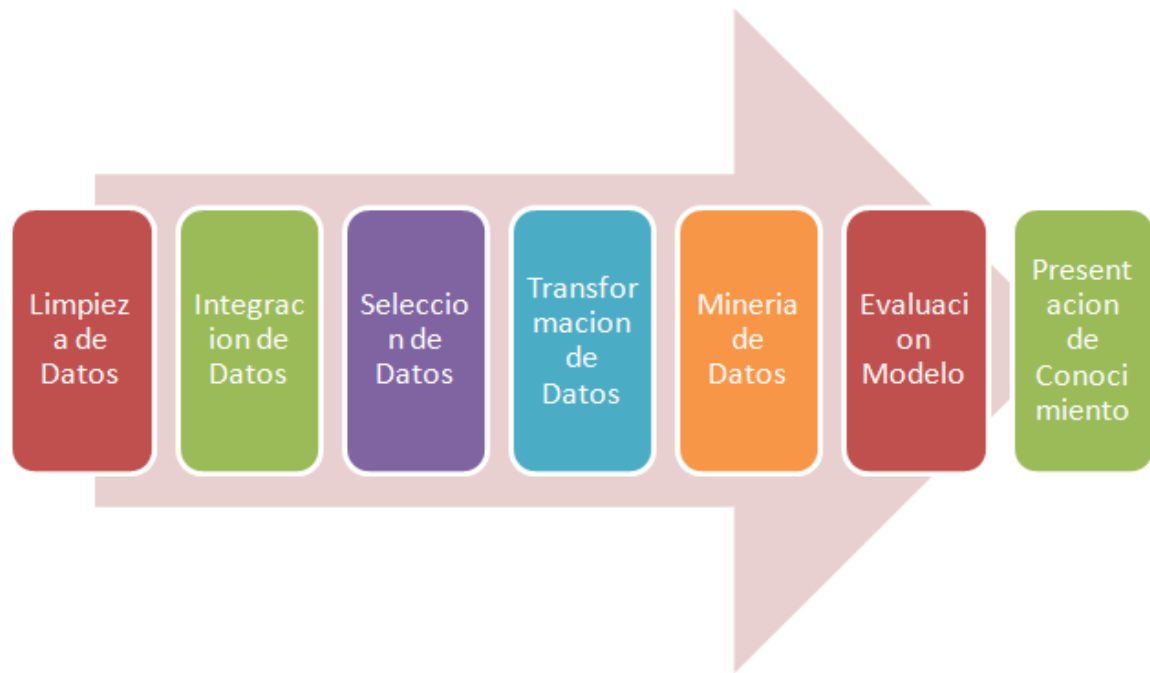


Fuente: Diseño propio

I-9. Minería de Datos

El proceso KDD consiste en lo siguiente:

1. Limpieza de Datos
2. Integración de Datos
3. Selección de Datos
4. Transformación de Datos
5. Minería de Datos
 - a. Definición de objetivos
 - b. Recolección de datos e integración
 - c. Análisis de exploración
 - d. Selección de atributos
 - e. Desarrollo del modelo y validación
 - f. Predicción e interpretación
6. Evaluación de Patrones
7. Presentación del conocimiento



Fuente: Diseño propio

I-10. Proceso KDD

Se dice que la minería de datos debe cumplir con 6 tareas principales:

1. Clasificar. Asignar a cada entidad un grupo.
2. Estimación. Asignar valores numéricos a variables.
3. Predicción. Clasificación de entidades de acuerdo a su comportamiento esperado en el futuro.
4. Agrupado de afinidad. Evaluación de relaciones entre elementos de datos.
5. Clustering. Dividir un universo de datos en conjuntos pequeños que tengan similitudes.
6. Descripción. Caracterizar lo que se haya descubierto a lo largo del proceso de minería de datos.

Para que este proceso se lleve a cabo con éxito se requiere de un gran compromiso por parte de los expertos en el dominio de los datos y los mineros de datos. Mientras que el experto en el dominio de los datos responde todas las posibles dudas que le puedan surgir al minero de datos, el minero de datos hace uso de métodos matemáticos que nos lleven a un aprendizaje inductivo a partir de los datos (Luis Paulo Vieira Braga, 2009).

Generalmente la minería de datos puede tener dos enfoques:

- Interpretación. Es la identificación de patrones para mostrarlos en términos de reglas que sean fácilmente entendidas por los expertos en el dominio de los datos y que de esta manera representen conocimiento para ellos.

- Predicción. Es el asumir que una variable tendrá cierto valor en el futuro

Para lograr cumplir con estos enfoques existen diversos algoritmos que permiten llegar a los modelos deseados, por mencionar algunos tenemos C4.5, K-Means, Support Vector Machines, A priori, EM, PageRank, AdaBoost, KNN, Naives Bayes.

El proceso puede ser mediante un aprendizaje guiado en el cual existe una previa clasificación de los datos o aprendizaje no supervisado en el cual no existe ninguna clase de clasificación de los datos (Vercellis, 2009).

Una aportación importante en este campo se dio a partir de un estudio enfocado en las tareas de marketing (Kumar Kar, Kumar, & Kumar De, 2010), en el cual se hace un resumen breve de como algunas tareas de Minería de datos ayudan a mejorar el resultado de los esfuerzos de mercadotecnia. El estudio menciona lo siguiente:

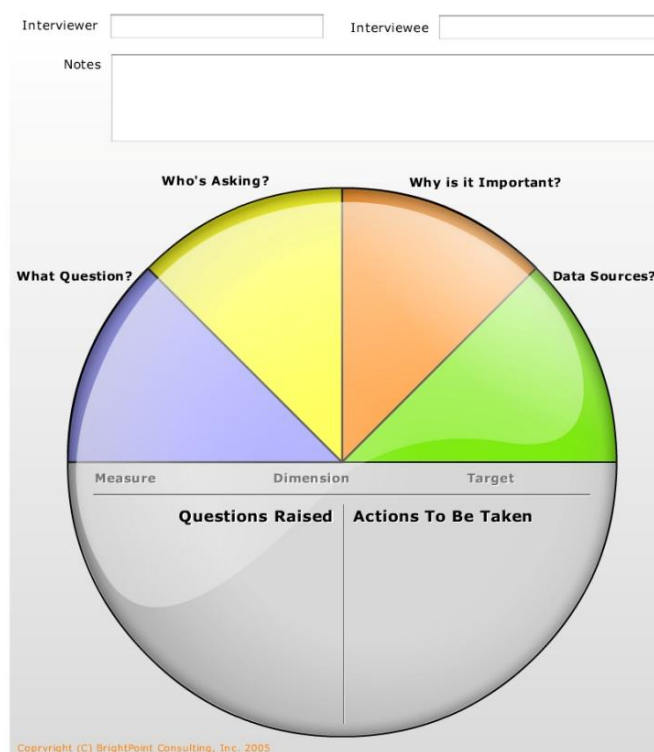
- Clustering. Se agrupan los datos en clases de acuerdo a sus similitudes, esta técnica puede ser utilizada para segmentar un mercado.
- Clasificación. Se realiza a partir de la teoría de decisión Bayesiana, redes neuronales, etcétera. Esto puede permitir desarrollar publicidad selectiva acorde a una segmentación previa.
- Asociación de patrones. Sirve para predecir patrones basados en secuencias de datos en base a los cuales la herramienta seleccionada ha sido entrenada. En marketing permite la predicción de preferencias de los clientes, que productos o publicidad pueden ser interesantes.
- Sumarización. Son métodos que permiten agrupar la información de acuerdo a ciertas variables que le dan sentido a métricas establecidas con las que se relaciona. Permite la identificación de la utilidad por segmentos lo que permite que distribuya los recursos acorde a este análisis.
- Modelo Predictivo. Es un proceso mediante el cual un modelo es creado y elegido para hacer mejores predicciones de una probabilidad. Este tipo de procesos permiten que se identifiquen las probabilidades de ocurrencia de respuesta de los clientes de tal manera que se les pueda proporcionar promociones especiales.
- Análisis de liga. Es una metodología para mapear y medir el flujo de la información mediante la interacción de sus nodos. Permite tener visibilidad que persona es un líder y puede influenciar a los seres a su alrededor.

1.1.1.7. Presentación

La finalidad de tener un Data Warehouse es el que llegue información al usuario final para que pueda tomar las decisiones pertinentes para mejorar el negocio, por lo tanto existen varias maneras de presentar esta información de manera que sea fácil su lectura.

Entra las maneras en las que se puede presentar tenemos:

- Scorecards. Son usados para alinear las operaciones con la estrategia del negocio. Este tipo de presentación de la información contiene principalmente los KPI y está dirigido a los altos mandos. Este tipo de herramienta como lo especifica un artículo sobre PyMEs japonesas (Aoki & Hasebe, 2012), es importante para monitorear el funcionamiento de una estrategia y dichas estrategias se deben evaluar constantemente para que se adapten al ambiente. En este mismo estudio se indica un proceso para la implementación de un BSC:
 - Generar Caso de la compañía
 - Identificar problema
 - Generar y planear la solución
 - Revisión y aprobación del cliente para su implementación
 - Validación (este paso puede hacer que se repita nuevamente todo el proceso)



Fuente: Brighpoint Consulting

I-11. Scorecards

- Dashboards. Está destinado a medir el desempeño de los procesos de negocio para asegurar que cumplan los objetivos establecidos, el dashboard contiene métricas y KPI's que permiten medir el cumplimiento de los objetivos proporcionando información para la toma de decisiones sobre acciones que encaminen el negocio al cumplimiento de sus objetivos.





Fuente: <http://adpu.wordpress.com>

I-12. Dashboard

- Reportes. Se muestra la información en forma de tablas que contienen información para ser analizada por los especialistas del negocio, este tipo de presentación de la información permite presentar la información a diferentes niveles de detalle.

Así mismo esta información dependiendo de la herramienta que se utilice, puede llegar al usuario de diferentes maneras, las principales son:

Presentación	Descripción	Ilustración
Archivos	Los reportes se pueden generar en archivos de algún tipo, generalmente y por comodidad se utiliza Excel.	 
Correo	Los reportes pueden ser enviados por correo electrónico.	

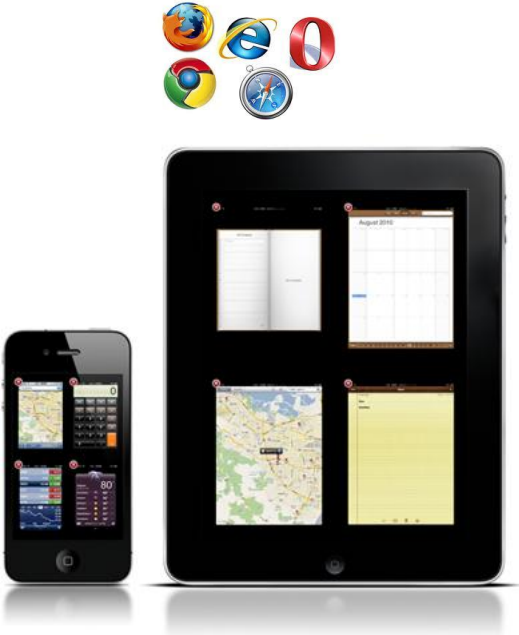
Presentación	Descripción	Ilustración
Web	Pueden conectarse a un portal empresarial o al front end web de la herramienta implementada.	
Móvil	Existen herramientas que permiten que los reportes sean consultados desde los móviles con conexión a internet y que contengan las aplicaciones móviles necesarias.	

Tabla 2. Presentación de Datos

PYME

Las PyMEs son empresas de pequeño tamaño en cuanto a los ingresos que generan y los empleados con los que cuentan. Este es un concepto que se tiene considerado en el mundo aunque se difiere entre países entre los rangos que permiten clasificar a una empresa como PyME.

1.1.2. Competitividad

La competitividad es la capacidad para posicionarse de manera ventajosa, permanecer y desarrollarse en el mercado. Para lograr esto se debe tomar en cuenta diversos conceptos como:

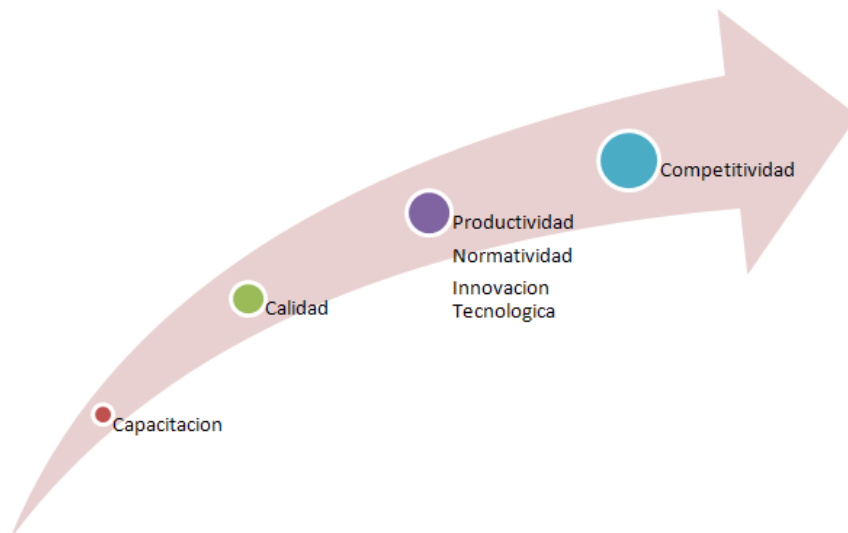
- Precio
- Diferenciación
- Innovación
- Desarrollo Tecnológico

Algunos de las metodologías que se usan para medir la competitividad son:

- Índice de Tipo de Cambio Real Efectivo
- Productividad

- Índice WEF del Global Competitiveness Report
- Índice IMD
- Metodología CAN

Uno de los elementos importantes en la competitividad de cualquier empresa es la capacitación y desarrollo de competencias para el trabajo.



Fuente: Diseño propio

I-13. Competitividad

1.1.3. Productividad

La productividad es el aprovechamiento del tiempo de trabajo. Una persona es productiva cuando se mantiene trabajando, desarrollando correctamente su labor durante la jornada.

La baja productividad puede causar una alteración en los costos de operación y propiciar merma en las utilidades.

Es necesario identificar métricas que nos permitan tener un control sobre la productividad de la empresa, para tomar las decisiones adecuadas que nos permitan incrementar la productividad. Un ejemplo de estos indicadores es:

Desempeño de la productividad = recursos/resultados

Los recursos pueden ser:

- Primarios: Recursos humanos, materiales, técnicos, etc.

- Secundarios: tiempo

Los resultados pueden ser:

- Primarios: producción
- Secundarios: dinero

La productividad permite conocer la eficiencia de una empresa y es el principal requisito para competir en el mercado.

Existen ciertos factores que afectan la productividad:

1. Externos. Disponibilidad materias primas, mano de obra calificada, políticas estatales relativas a tributación y aranceles, infraestructura, capital, control.
2. Internos: bienes inmuebles, materiales, energía, maquinas, recurso humano.

Existe un bajo aprovechamiento de la productividad del capital humano, por lo que se sabe que el nivel de productividad de empleados y obreros en México es de 50 a 60%. Incluso se han encontrado grupos con una productividad que no rebasa el 50%.

Existen algunas acciones que pueden ayudar a fortalecer la productividad:

- Actualizaciones
- Apoyos gubernamentales
- Capacitación

Capítulo II Marco de Referencia

En este capítulo se documenta la información referente al estado de las PyMEs en México y como es su interacción con la inteligencia de negocios, mostrando los beneficios que puede traer a las PyMEs su implementación. Esto nos otorga una visión de la importancia de la inteligencia de negocios como una oferta posible para incrementar la competitividad de las PyMEs. Así mismo se documentan diversas metodologías de inteligencia de negocios, esto nos permite tener un punto de referencia para generar la propuesta de este trabajo de tesis. De la misma manera se hace un recuento de las herramientas que existen en el mercado tanto Open Source como propietarias.

Entorno PYME

Esta sección tiene por objetivo mostrar hechos relacionados a las PyMEs como lo son datos que demuestran cómo se afecta de manera positiva la competitividad y productividad de una PyME al emplear soluciones de BI. Se realizó la documentación de diversos estudios que guardan una relación con este tema.

2.1.1. Competitividad, Productividad y BI

En México el 99% de las empresas están identificadas como MiPymes, estas representan una enorme fuente de empleo, sin embargo el esquema nacional no está diseñado para el desarrollo de este tipo de empresas.

De acuerdo al Diario Oficial de la Federación las empresas PyME se clasifican como se muestra en la ilustración II-1 - Clasificación de PyME.

Estratificación				
Tamaño	Sector	Rango de número de trabajadores	Rango de monto de ventas anuales (mdp)	Tope máximo combinado*
Micro	Todas	Hasta 10	Hasta \$4	4.6
Pequeña	Comercio	Desde 11 hasta 30	Desde \$4.01 hasta \$100	93
	Industria y Servicios	Desde 11 hasta 50	Desde \$4.01 hasta \$100	95
Mediana	Comercio	Desde 31 hasta 100	Desde \$100.01 hasta \$250	235
	Servicios	Desde 51 hasta 100	Desde \$100.01 hasta \$250	235
	Industria	Desde 51 hasta 250	Desde \$100.01 hasta \$250	250

Fuente: Inegi

II-1 - Clasificación de PyME

Se sabe que las Pymes en México no emplean la tecnología por lo que comenzar a hacerlo podría representar una enorme ventaja competitiva. Actualmente solo 25% de las empresas PYME utilizan internet y lo utilizan con fines de búsqueda de información, solo el 10% cuenta con una página web.

Además de esto según algunos estudios (Li, Li, & Zhu, 2010) de los principales problemas de las PyMEs es que no cuentan con empleados con los suficientes conocimientos y habilidades para tomar decisiones de manera correcta cuando los problemas ocurren. Generalmente las PyMEs dedicadas a la consultoría, son empresas que cuentan con personal especializado en alguna materia, por las PyMEs en este giro cuentan con personal calificado, aunque no siempre calificado para la toma de decisiones estratégicas.

Algunos de los puntos que pueden ayudar a mejorar la competitividad de las empresas es que estás identifiquen bien los sectores de oportunidad y que se establezcan en una ubicación física que les permita sobresalir teniendo perfectamente identificado el nicho de mercado al cual se van a dirigir y de ser posible formar una cadena de valor.

Están establecidas las siguientes características que tienen las PyME en China sobre la toma de decisiones (Zhao & Yao, 2012):

1. “La toma de decisiones no es de manera científica, se basa en experiencias personales e información personal”. Este punto aplica de igual manera en México debido a que un gran porcentaje de los tomadores de decisiones no cuentan con una formación formal al respecto de la toma de decisiones.
2. “Debido a la gran capacidad para generar empleos se debe considerar la toma de decisiones de manera macro económica, considerando el entorno nacional”. Se ha notado que las PyMEs carecen muchas veces de conocimiento de su entorno externo,

debido a que centran su atención en sus operaciones diarias restando importancia al contexto en el que se desenvuelven. Al igual que en China representan una gran fuente de trabajo por lo que este punto también debe considerarse.

3. “La locación de las PyMEs también representa un factor importante en su desarrollo pues existe un desbalance entre el desarrollo de las PyME y su ubicación”. Es parte de la formalización de la toma de decisiones el elegir una ubicación adecuada para una PyME.

Según un estudio sobre el impacto de la implementación de herramientas de inteligencia demostró que los trabajadores con conocimiento pueden generar una ventaja competitiva cuando son asistidos por herramientas eficientes y competitivas (Sadok & Lesca, 2009).

“Si se desea mantener la competitividad es necesario incrementar la productividad y para ello el uso de las TIC es uno de los instrumentos más eficaces que se puede utilizar. “ (Castillo, 2007) . En esta época de crisis se ha mostrado una tendencia por que las empresas que se apoyan de las TIC para apoyar su proceso de toma de decisiones sobrevivan. El que una empresa se apoye de las TIC y que sumado a esto conozca su negocio les permite generar la inteligencia necesaria para contar con una ventaja competitiva en su industria.

Otro de los puntos importantes para mantener la competitividad de una empresa está en la capacitación de sus miembros. Esto se debe no solo a la especialización de los empleados sino a que estos se ven motivados a partir de estas capacitaciones lo que los incita a ser más creativos y disfrutan más sus labores. El apoyarse en las tecnologías para aumentar la productividad de sus miembros también es una buena opción,

Al aumentar la productividad, se eleva la competitividad puesto que la empresa con empleados más productivos permite el uso de tecnologías, realizar nuevos desarrollos y un crecimiento constante.

La Inteligencia de negocios tiene como principal materia prima los datos de una empresa por lo cual podemos asegurar que cualquier empresa que cuente con datos sin importar en qué forma estos estén organizados, pueden implementar una solución de inteligencia de negocios que les permita visualizar escenarios, realizar pronósticos y por lo tanto tomar decisiones.

De acuerdo a un artículo de Andrés de Alcazar el considera que una empresa PyME puede beneficiarse inicialmente de cuatro soluciones propias de la inteligencia de negocios (Del Alcazar, 2010):

1. Reporteo Avanzado
2. Análisis de datos

3. Cuadros de Mando
4. Minería de datos

Ariadna Cruz menciona que el invertir en soluciones BI representa para una PyME un aumento en su competitividad puesto que este tipo de soluciones le permite tener la información de manera accesible y completa de todo el negocio, esto enriquecido con las posibilidades que ofrece la inteligencia de negocios como es el análisis de los datos, la minería de datos, Reporting avanzado o cuadros de mando, representa un gran aporte para la toma de decisiones dentro del negocio (Cruz, 2010).

Como conclusión de estudio realizado por (Benki & Papastathopoulos, 2009), se confirma que la implementación de herramientas como los DSS ha tenido un impacto positivo en las utilidades de las PyMEs permitiéndoles crecer mientras más exploten este tipo de herramientas. Para que las PyMEs sobrevivan al ambiente cambiante, es necesario que sean capaces de utilizar sistemas que les faciliten la combinación de datos, utilizar modelos sofisticados de análisis y herramientas.

La información con la que se pueda contar nos podría proporcionar respuestas a ciertas interrogantes relacionadas con la productividad, lo que nos llevaría a la toma de decisiones dirigidas a aumentar las utilidades de la compañía mediante la reducción de la improductividad de los recursos humanos de la empresa.

Según una entrevista realizada en algunas PyMEs, se encontró que las empresas utilizan la inteligencia de negocios de manera fragmentada, espontánea y de manera aislada por algunos individuos. No existe una formalización del almacenamiento de los datos, las empresas se basan principalmente en el conocimiento tácito y no cuentan con criterios de selección definidos. Es por esto que los proyectos de implementación de inteligencia de negocios en empresas PyME es necesario que busquen la optimización de los procesos que utilicen de inteligencia de negocios y que los tiempos y costos no sean elevados (Sadok & Lesca, 2009).

Con la tecnología idónea se lleva a la PyME por el camino del desarrollo, evolucionando esta de tal manera que pueda incrementar su competitividad. Además se ha demostrado que es necesario que las PyME cuenten además del capital necesario, con el conocimiento necesario para explotar las herramientas que se decida adquirir.

2.1.2. BI para Pymes

Mediante la implementación de soluciones de BI se busca contar con un conjunto de estrategias que nos permitan crear conocimiento a partir del análisis de los datos existentes en la empresa. Los objetivos de implementar BI en una PYME son:

1. Accesibilidad de la información
2. Apoyo en la toma de decisiones
3. Orientación al usuario final

De acuerdo a un estudio realizado en empresas PyMEs alemanas, se identificaron tres principales beneficios al implementar una solución de inteligencia de negocios (Scholz, Schieder, Kurze, Gkuchowski, & Boehringer, 2010):

1. Existe una mejora en el soporte de los datos, debido a que se reducen los esfuerzos por analizar y reportear los datos.
2. Se mejora la toma de decisiones, esto porque el análisis de los datos se puede hacer de manera más precisa al contar con más datos disponibles.
3. Se logra generar ahorros de personal y costos, por medio de la disminución de los esfuerzos.

Considero que este tipo de beneficios también se pueden ver con la implementación de soluciones de inteligencia de negocios en las PyMEs Mexicanas, debido a que se estarían atacando los mismos problemas de falta de integración de la información, lo que facilitará el reporte de los datos y su correspondiente análisis.

En las Pymes podemos encontrar ciertas particularidades que tal vez en proyectos de implementación sobre empresas grandes no encontraríamos como lo es que la probabilidad de que no se cuente con un sistema de información lo que represente un reto para la integración de los datos, el que no tengan totalmente definido su proceso de toma de decisiones de manera estructurada.

Existe un estudio (Aversano, Grasso, & Tortorella, 2011) que afirma que las necesidades de información vienen de diferentes niveles de toma de decisiones. Uno de estos niveles busca el soporte a la operación y el otro está enfocado a un nivel más alto considerando las tareas del manejo de la empresa.

Así mismo puede que cuenten con mucho conocimiento tácito construido a través de la experiencia, lo cual complica la transferencia del conocimiento y su explotación (Sadok & Lesca, 2009).

La cantidad de datos que pueda manejar una PYME puede permitir que se puedan analizar en más del 20% lo que permite contar con un mejor análisis, así mismo el volumen de información que pueden manejar les permite utilizar en conjunto con otro tipo de soluciones como son Cloud Computing o en su defecto existen herramientas que tratan de minimizar el impacto de la infraestructura necesaria para la implementación de soluciones de este tipo.

Debido a las condiciones del mercado en la actualidad en la que todo sucede en poco tiempo, es necesario que las Pymes cuenten con información relevante que les permita seguir el ritmo que marca el mercado actual, permitiéndoles competir frente a cualquier tipo de empresa mejorando su toma de decisiones apoyados por una solución de Inteligencia de negocios.

Según un estudio realizado en Grecia, el 13% de las Microempresas utilizan herramientas DSS, contra 45.5% de las pequeñas empresas y un 77.8% de las empresas medianas (Benki & Papastathopoulos, 2009). En México se puede suponer que las PyMEs siguen esta misma tendencia, pero con una menor representación.

Las herramientas que el mercado de BI ofrece en el mercado permiten hacer las mediciones necesarias para cada negocio, de manera rápida lo que permite tomar acciones inmediatas cuando así es requerido. Además que permiten tener un amplio conocimiento de todos los aspectos del negocio contando con visiones estratégicas y operativas, esto con la finalidad de tomar decisiones más efectivas y de calidad en cualquier nivel de la empresa.

Samuel Benarroch cita las siguientes ventajas del uso de inteligencia de negocios:

- Mejora la productividad de los empleados.
- Simplifica la gestión de los PKI
- Comunica los objetivos de manera consistente a todos los miembros de la empresa.
- Potencia las inversiones hechas en tecnología y facilitan el ahorro de recursos
- Permiten acceder a información que facilite la toma de decisiones
- Identificación y extracción de datos no estructurados
- Ofrece análisis de la información

El principal objetivo de la inteligencia de negocios es apoyar a las empresas en la toma de decisiones mediante la integración, análisis y presentación de sus datos. Lo que para una PYME representa en una mejor toma de decisiones basadas en hechos.

El estudio realizado por (Benki & Papastathopoulos, 2009) arroja que la implementación de herramientas como los DSS está asociada con las utilidades que puede generar una PyME.

El que una PyME pueda tomar mejores decisiones le permitirá le traerá entre otras ventajas las siguientes (Zhao & Yao, 2012):

1. Permitirá generar valor a sus clientes, empleados, dueños y sociedad
2. Mantener estructuras simples y flexibles
3. Tomar ventaja de la formalización del proceso de toma de decisiones.

Para poder ofrecer a una solución que se adapte a sus necesidades reales, por lo que Lesca propone en 2003 un modelo llamado VAS-IC, en el cual se identifican los siguientes pasos (Sadok & Lesca, 2009):

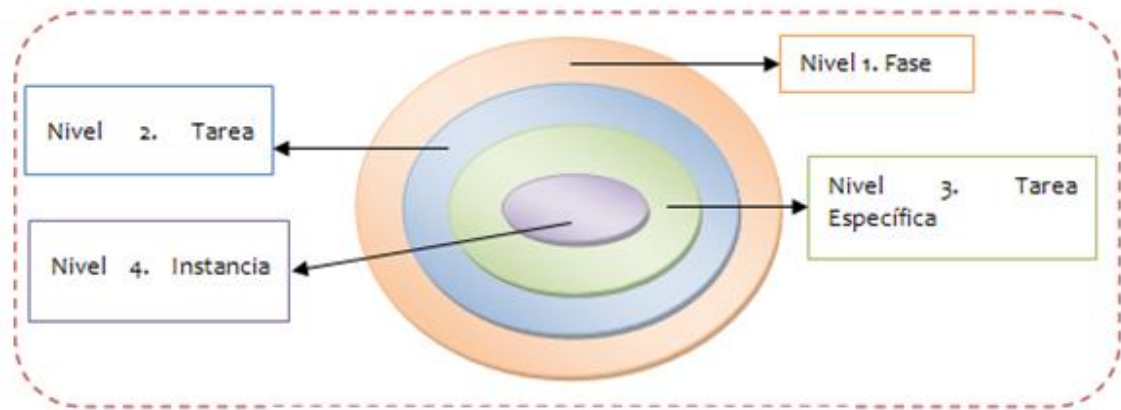
- Targeting. Identificación de actores y procesos a ser monitoreados para optimizar los costos y tiempos dedicados.
- Tracking. Es la recolección de la información que permita identificar a los dueños de los procesos.
- Knowledge memorization. En este paso se almacena el conocimiento generado como resultado del paso “Hacer Sentido Colectivo”.
- Diffusion. Es la distribución del conocimiento mediante el paso de “Hacer Sentido Colectivo” a los usuarios apropiados.
- Action. Si el conocimiento generado es significativo se toma en cuenta para la toma de decisiones.
- Collective Sense-Making. Es el proceso en el cual se crea conocimiento mediante la realización de ligas entre los datos recolectados.

Metodologías existentes

En esta sección podemos encontrar la documentación de diversas metodologías que están enfocadas en la Inteligencia de Negocios, así como otras que además están enfocadas en PyMEs. Esta documentación se hace con el fin de tener una referencia sobre las metodologías que existen actualmente.

2.1.3. CRISP-DM

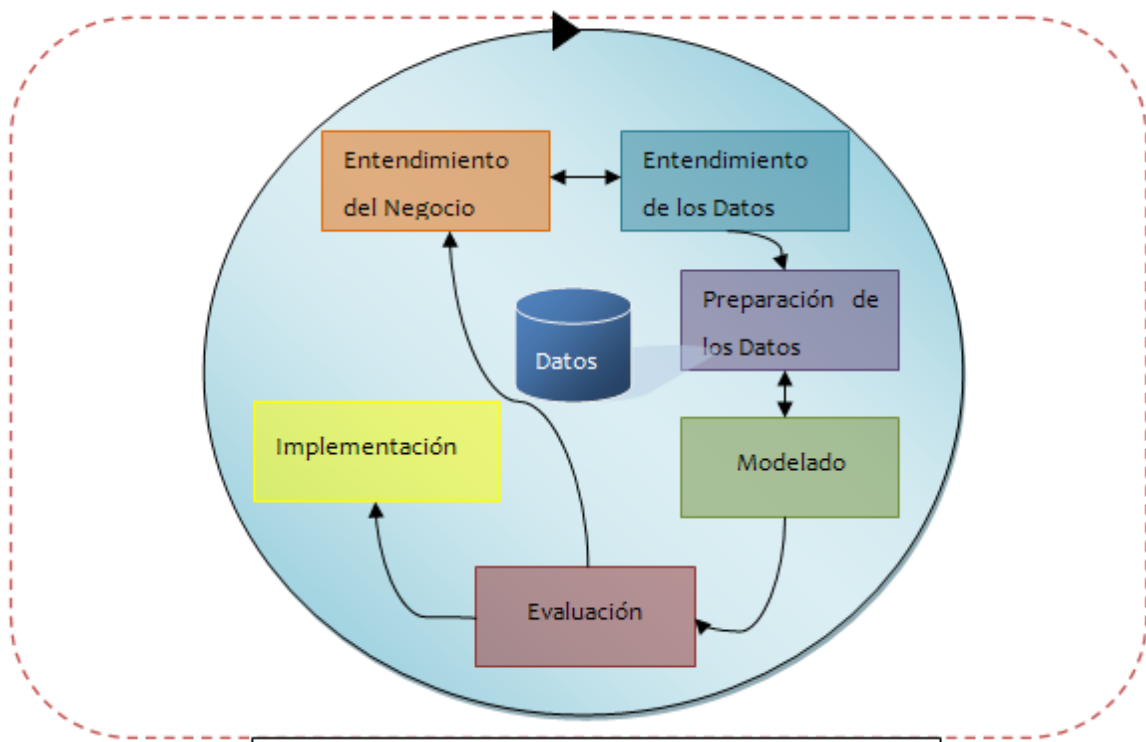
La metodología CRISP (Cross Industry Standard Process for Data Mining) fue desarrollada con la finalidad de asegurar el éxito de los proyecto de Minería de Datos. Esta metodología esta descrita en un modelo de un proceso jerárquico yendo de lo general a lo particular.



Fuente: Diseño propio

II-2. Proceso Jerárquico CRISP DM

Los proyectos de minería de datos siguen un ciclo de vida de acuerdo al cual existe la definición de un modelo de referencia, en este modelo se pueden identificar las fases y tareas requeridas, sin embargo las relaciones entre estos elementos puede variar de acuerdo a los objetivos y necesidades que pueda presentar cada proyecto.



Fuente: Diseño propio

II-3. Ciclo de Vida CRISP DM

Para un mayor entendimiento a continuación describiremos cada fase que comprende la metodología.

Fase	Tarea	Salida
Entendimiento del negocio	Determinar Objetivos del negocio. Entender que es lo que realmente quiere el cliente, descubrir factores importantes que puedan influenciar el proyecto.	<ul style="list-style-type: none"> • Situación actual • Objetivos del negocio • Criterios de éxito del negocio.
	Evaluar Situación. Recursos, restricciones, supuestos y otros factores importantes para cubrir el objetivo del análisis de datos y el plan del proyecto.	<ul style="list-style-type: none"> • Inventario de recursos. • Requerimientos, Supuestos y Restricciones. • Riesgos y Contingencias • Terminología • Costos y beneficios
	Determinar objetivos de Minería de Datos. Determinar los objetivos del proyecto de manera técnica.	<ul style="list-style-type: none"> • Objetivos de minería de datos • Criterios de éxito de minería de datos
	Plan del Proyecto. Desarrollar el plan de proyecto que nos lleve a lograr los objetivos de minería de datos y por lo tanto a cubrir los objetivos del negocio.	<ul style="list-style-type: none"> • Plan del proyecto • Evaluación inicial de herramientas y técnicas
Entendimiento de los Datos	Recolección inicial de los datos. De acuerdo a los recursos mencionados en la fase anterior se inicia la recolección de los datos.	<ul style="list-style-type: none"> • Reporte de recolección de datos
	Descripción de los datos. Se describen las propiedades de los datos recolectados	<ul style="list-style-type: none"> • Reporte de descripción de datos
	Explorar los datos. Aborda las preguntas de minería de datos por medio de sentencias, visualizaciones o	<ul style="list-style-type: none"> • Reporte de exploración de datos

Fase	Tarea	Salida
	reportes.	
	Verificar calidad de los datos. Se realiza la verificación y validación de la información de los datos obtenidos	<ul style="list-style-type: none"> • Reporte de calidad de datos
Preparación de los datos	Conjunto de datos. Serán necesarios para el modelado	<ul style="list-style-type: none"> • Conjunto de datos • Descripción del conjunto de datos
	Selección de los datos. Seleccionar los datos a ser usados en el análisis de acuerdo a su relevancia para cumplir los objetivos de minería de datos.	<ul style="list-style-type: none"> • Razón de ser de los datos
	Limpiar datos. Elevar la calidad de los datos.	<ul style="list-style-type: none"> • Reporte de limpieza de datos
	Construir datos. Producción de atributos derivados y registros.	<ul style="list-style-type: none"> • Atributos derivados • Generación de registros
	Integrar datos. Combinación de tablas o registros.	<ul style="list-style-type: none"> • Combinar datos
	Formatear datos. Modificaciones sintácticas que no modifican su significado.	<ul style="list-style-type: none"> • Datos reformateados
Modelado	Seleccionar técnica de modelado.	<ul style="list-style-type: none"> • Técnica de modelado
	Evaluar modelo. De acuerdo los criterios de éxito de minería de datos determinados en fases anteriores.	<ul style="list-style-type: none"> • Evaluación del modelo • Revisión de parámetros de configuración
Evaluación	Evaluar resultados. Permite evaluar el nivel en que el modelo cumple con los objetivos del negocio	<ul style="list-style-type: none"> • Evaluación de los resultados de minería de datos respecto a los criterios de éxito del negocio.

Fase	Tarea	Salida
		<ul style="list-style-type: none"> Modelos aprobados
	Revisar proceso. Revisión a profundidad en búsqueda de factores importantes	<ul style="list-style-type: none"> Revisión del proceso
	Determinar pasos siguientes. Se debe determinar si después de este paso se puede pasar a la implementación o realizar alguna otra iteración.	<ul style="list-style-type: none"> Listado de posibles acciones Decisiones
Implementación	Planeación de Implementación. Estrategia para llevar a cabo la implementación.	<ul style="list-style-type: none"> Plan de implementación
	Plan de monitoreo y mantenimiento. Estrategia para monitorear y mantener la solución de minería de datos.	<ul style="list-style-type: none"> Plan de monitoreo y mantenimiento
	Producción del reporte final. Un documento que resuma los resultados de minería de datos y la presentación de los mismos.	<ul style="list-style-type: none"> Reporte final Presentación
	Revisión del proyecto. Lecciones aprendidas	<ul style="list-style-type: none"> Documentación de experiencias

Tabla 3. Metodología CRISP DM

2.1.4. HEFESTO 2.0

Es una metodología que toma su nombre de la mitología griega y está enfocada en la construcción de un Data Warehouse.

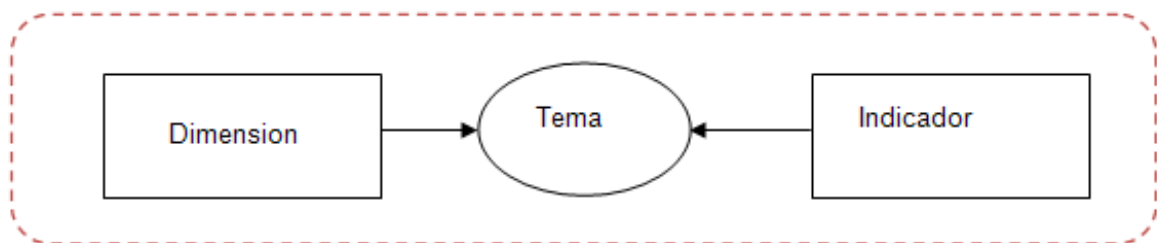
Esta metodología consiste de 4 fases:

La fase 1 ANÁLISIS DE REQUERIMIENTOS consiste principalmente en obtener de los usuarios los requerimientos que permitirán definir los indicadores que le permitan medir el rendimiento de la empresa para la toma de decisiones que permita encaminar a la empresa al cumplimiento de los objetivos.

Para cubrir la fase 1 se deben realizar ciertas tareas específicas. La primera tarea a realizar será conocer las necesidades de información del usuario para el logro de sus objetivos, para lo cual se pueden utilizar diversas técnicas. La siguiente tarea es desmenuzar lo obtenido en la tarea de conocimiento de las necesidades de información con la finalidad de identificar los indicadores y las dimensiones que le den sentido a dichos indicadores. En el manual de referencia se ejemplifica como sigue:

“Unidades vendidas de cada producto a cada cliente en un tiempo determinado”

La siguiente actividad es generar un modelo conceptual que represente de manera visual los hallazgos mencionados. Para esta metodología se propone utilizar rectángulos y elipses, los rectángulos a la izquierda representaran las dimensiones, la elipse deberá indicar el proceso en cuestión y los rectángulos a la derecha se especifican los indicadores.



Fuente: Diseño propio

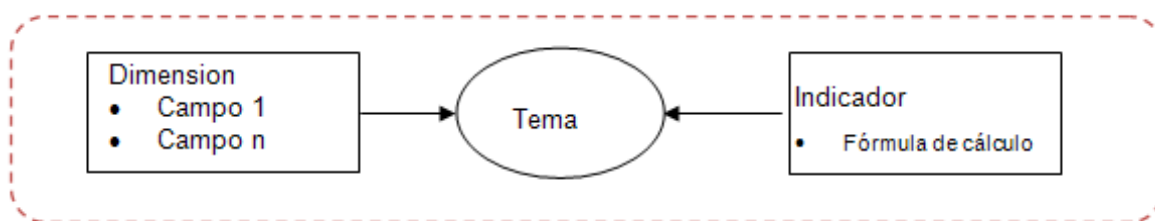
II-4. Análisis de Requerimientos

Una vez que se cuenta con el modelo conceptual se pasa a la fase 2 ANALISIS DE LOS OLTP, en dicha fase la primer tarea a realizar será conformar los indicadores, esto se hace mediante la definición de las fórmulas y conceptos que permitan obtener los indicadores. Por lo que para conformar los indicadores es necesario conocer que elementos o hechos nos permitirán hacer los cálculos y conocer qué tipo de operación es necesario realizar a dichos hechos para obtener el valor del indicador.

La segunda tarea consiste en Establecer correspondencias. Esto es en pocas palabras realizar un mapeo de las fuentes de datos proporcionadas por el usuario y el modelo conceptual realizado a partir de las necesidades de información del mismo.

Después de este mapeo existe otra tarea que es necesario realizar, esta tarea es el nivel de granularidad y consiste en definir los campos que formaran parte de las dimensiones, esto se realiza con el apoyo del usuario puesto que será este el que defina dichos campos así como será quien nos proporcione una explicación de los campos lo que nos permitirá formar un diccionario de datos o en su defecto nos podremos apoyar de uno ya existente para realizar esta tarea.

Ya que se cuenta con la información del nivel de granularidad, se procede a agregarlo al Modelo Conceptual por lo que se extiende este modelo como se muestra en la imagen.



Fuente: Diseño propio

II-5. Detalle de Análisis de Requerimientos

Cuando ya se cuenta con el Modelo Conceptual Extendido se puede avanzar a la fase 3 MODELO LÓGICO DEL DW para poder construirlo se deben realizar ciertas tareas, la primera de ellas es definir el tipo de Modelo Lógico, puede ser en estrella, copo de nieve y constelación. Una vez decidido esto procedemos a la siguiente tarea que es el diseño de las Tablas de Dimensiones, tomando en cuenta el nombre de la tabla, claves primarias, nombres de campos intuitivos.

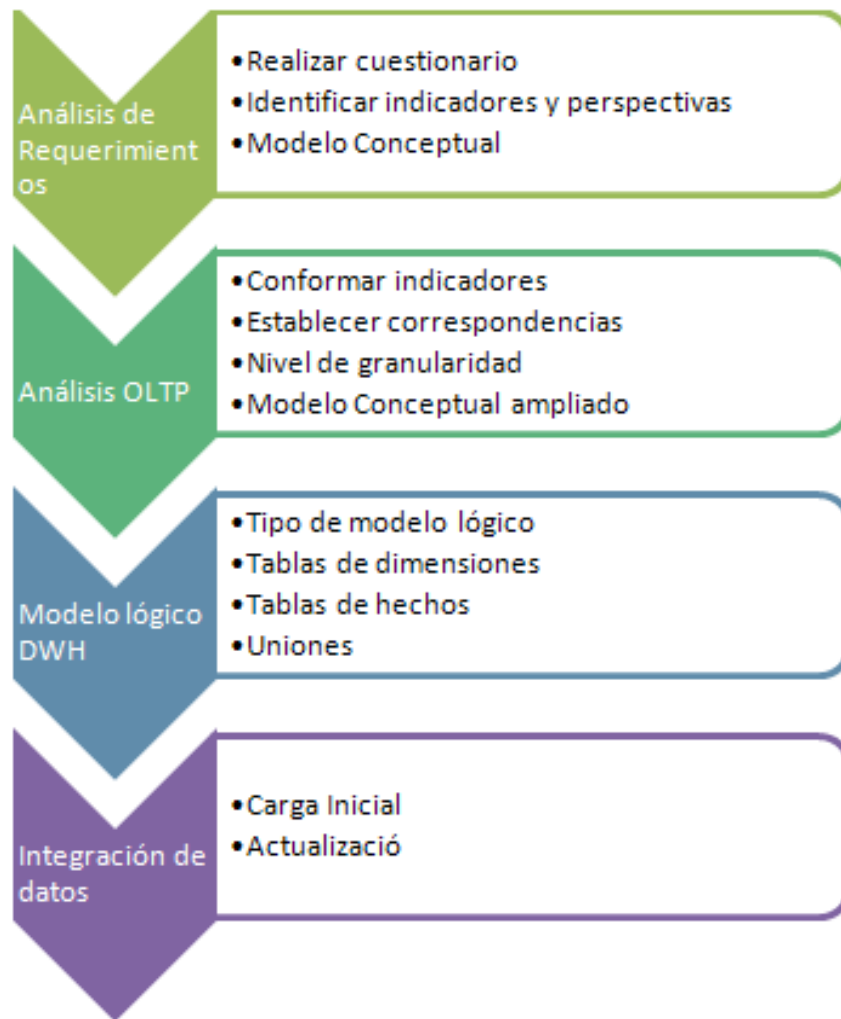
Seguidamente se definen las Tablas de Hechos que contendrán principalmente los valores que necesarios para determinar los indicadores, estas tablas deberán tener un nombre representativo, se deben verificar las llaves primarias, por cada campo que guarde valores de hechos tendrá su correspondiente indicador.

Finalmente la última tarea de esta fase consiste en hacer las relaciones indicadas en el modelo conceptual, es decir relacionar Dimensiones con hechos,

La última fase es la INTEGRACIÓN DE DATOS esta fase consiste en poblar el Data Warehouse, el cual se debe construir después de obtener el Modelo Lógico usándolo como guía para la construcción.

La primera tarea es la Carga Inicial, para lo cual nos debemos apoyar de los procesos ETL que nos permitirán garantizar la calidad de los datos. Para hacer una carga correcta, primero se deberán poblar las dimensiones siguiendo su orden en la jerarquía (de lo general a lo particular) para después cargar las tablas de hechos.

Después de haber realizado la primera carga se debe de proceder a hacer la siguiente tarea la cual consiste en establecer la arquitectura necesaria para poder actualizar el Data Warehouse, para lo cual se debe continuar con actividades de limpieza de datos.



Fuente: Diseño propio

II-6. HEFESTO

2.1.5. METODOLOGIAS AGILES

Las metodologías ágiles están diseñadas para el modelado y la documentación de proyectos y se basan en 4 valores y en 12 principios (Infante, 2009).

Valores:

- El primer principio está enfocado en enfatizar que el éxito del proyecto depende de las personas por lo que es importante mantenerlas motivadas.
- El segundo principio menciona que aunque la documentación es importante no es el objeto de un proyecto, es importante que los proyectos cumplan con su objetivo y que la documentación realmente sea solo para soporte.

- El tercer principio habla sobre la colaboración que debe de existir entre el cliente y TI sin ser tan rigurosos con lo establecido en el contrato buscando siempre la colaboración y no el atacarse.
- El cuarto principio menciona que los desarrollos se deben adaptar a los cambios, no tomar los planes como una imposición sino más bien como una guía.

Principios:

1. Lograr la satisfacción del cliente.
2. Adaptarse según los cambios que sean requeridos.
3. Entregar resultados funcionales.
4. Equipo de trabajo formado por representantes del negocio y desarrolladores en TI.
5. Trabajar con gente motivada, y mantener dicha motivación
6. Comunicación personales las más recomendable
7. Avance medido en términos de resultados.
8. Soporte continuo para asegurar la calidad
9. Desarrollo sustentable
10. Simplicidad
11. Organización de los equipos
12. Lecciones aprendidas



Fuente: Diseño propio

II-7. Principios Metodologías ágiles

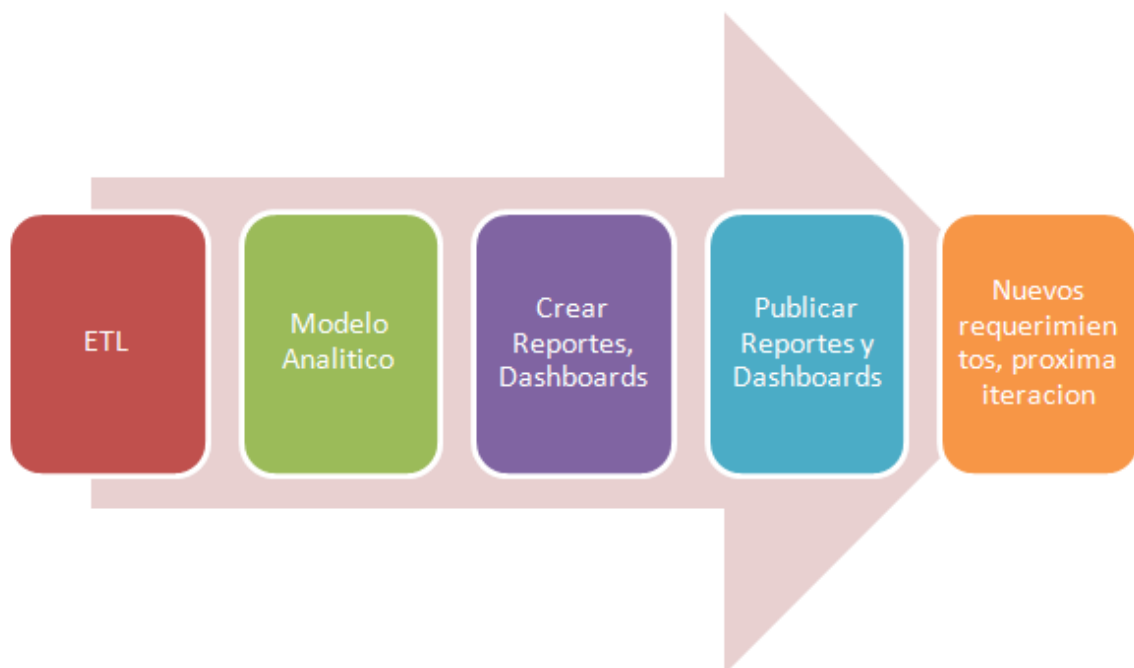
Estos valores y principios pueden ser usados para un proyecto de Business Intelligence. Por ejemplo la metodología Scrum que esta enfocada a la administracion de proyectos que busca elevar la productividad del equipo participante en dichos proyectos.

Esta metodologia consiste en lo siguiente:

1. Listado de funcionalidades. En este listado se incluyen los requerimientos ordenados de tal manera que podemos conocer cuales son las prioridades, este documento debe ser responsabilidad del cliente quien se debe comprometer a mantenerlo actualizado.
2. Listado de tareas. Son el conjunto de tareas a realizar a partir del listado de funcionalidades, de esta manera el equipo SCRUM traduce el QUE en el COMO, este equipo debe ser reducido a lo mas de 9 personas.
3. Ciclo de desarrollo. Dado que los desarrollos cumplen con un ciclo y se hacen de manera incremental se ocupa un día para la planeacion a lo mas 4 semanas de trabajo y 1 día para la entrega de funcionalidades al cliente.

A lo largo de todo el desarrollo debe estar presente el lider de proyecto por parte del equipo SCRUM quien se encargara de asegurarse de que se logren los objetivos, ayudando al equipo SCRUM a eliminar obstaculos, protege al equipo. A demas de que se hace un seguimiento diario en reuniones de 15 minutos, en las cuales se determina que se hizo, que se hara y que problemas existen.

Asi este tipo de metodologias pueden ser ampliamente adaptabas para proyectos de Business Intelligence. Es asi como encontramos otra metodologia agil desarrollada por GoodData un empresa que ofreece solucione BI en web, los cuales describen su metodologia agil como sigue:



Fuente: Diseño propio

II-8. Metodología Ágil BI

Ericka Chickowski nos muestra 6 pasos para convertir nuestros proyectos de Business Intelligence en ágiles (Chickowski, 2009):

1. Involucrar a los usuarios de manera fácil y regular
2. Priorizar de acuerdo al valor.
3. Automatizar las pruebas
4. Motivar la cultura de colaboración más allá de los desarrolladores
5. Empezar rápido fuera de la puerta

2.1.6. METODOLOGIA KIMBALL

La metodología de Ralph Kimball está enfocada principalmente en la construcción del Data Warehouse. La metodología conocida como Ciclo de vida del Road Map Dimensional de Negocio establece lo siguiente:

La razón de ser de los proyectos de Business Intelligence y de muchos otros, es el negocio, por lo tanto uno de los puntos importantes es tener claro que las necesidades del negocio son las que nos guiarán a lo largo de todo el proyecto. En general se contempla que el ciclo de vida dimensional del negocio se puede expresar en términos de lo que se muestra en la siguiente ilustración:



Fuente: Diseño propio

II-9. Metodología Ralph Kimball

1. PM

El primer paso es la planeación de proyectos, esto como una buena práctica usada en prácticamente la mayoría de los proyectos de TI. Esta planeación contempla los siguientes puntos:

- Evaluación de preparación.
 - Tener un patrocinador exigente que tenga visión y pasión por el negocio
 - Tener una fuerte motivación y compromiso con el negocio para la construcción de su Data Warehouse
 - Factibilidad técnica, de operación y de datos, la cual es la más crítica en términos de proyectos de este tipo. Esta consiste en identificar si los datos que están contenidos en las operaciones del negocio realmente cubren los requerimientos del negocio. Se debe trabajar con datos limpios y al nivel correcto de granularidad.

- Relación TI-Negocio
- Cultura analítica de la compañía
- Definición del alcance. Se establecen los límites que existirán alrededor del proyecto. Este alcance es definido en conjunto por TI y el negocio.
- Justificación. Principalmente es la estimación de los costos y los beneficios.

Recursos. Se refiere al equipo de trabajo que participara en el proyecto y que debe estar conformado tanto por personas de TI como por personas del negocio.

Recursos Humanos	
Negocio	TI
○ Patrocinador. Cliente final.	○ PM
○ Operador. Gerente de nivel medio al cual el patrocinador le delega responsabilidades.	○ Arquitecto técnico
○ Líder. Es aquel que está sumamente envuelto en lo que respecta al proyecto	○ Especialista técnico de soporte
○ Usuario	○ Modelador de datos
○ Analista de sistemas	○ Coordinador de meta data
○ Experto en el área	○ Administrador de datos
○ Desarrollador de la aplicación analítica	○ Diseñador de staging
○ Capacitador en Data Warehouse	○ Soporte de Data Warehouse

Tabla 4. Recursos - Metodología Ralph Kimball

Se deben identificar todas las actividades que se deben realizar y se debe dar un seguimiento al proyecto mínimo a la llegada de cada hito del proyecto, para obtener el visto bueno del cliente. Contemplados estos puntos se debe cuidar mucho el que se desarrolle de acuerdo a este plan y que este plan se mantenga. Esto se logra mediante los planes de comunicación y cuidando los posibles cambios que puedan ser requeridos a lo largo del proyecto, para que estos no se salgan de nuestro control.

2. Requerimientos del negocio

Para poder obtener los requerimientos del negocio debemos planear el cómo obtendremos dichos requerimientos. Existen 2 técnicas principales para la recolección de requerimientos:

- Las entrevistas
- Las sesiones facilitadoras

Estas se deben realizar principalmente con 3 roles del negocio:

1. Representantes del negocio
2. Expertos en los sistemas fuente
3. Los expertos en la materia

Con la finalidad de obtener información de que es lo que hacen, como lo hacen y porque lo hacen y poder relacionar estas respuestas con los datos.

Para lograr este punto con éxito se debe seleccionar a las personas correctas para fungir como entrevistadoras, las cuales deberán ir acompañadas de una tercera persona que les apoye en la toma de notas y que aporte otra visión de lo que se pueda generar en la entrevista.

También se debe realizar una selección de los representantes del negocio con los que se agendará una cita de preparación para obtener los requerimientos del negocio.

3. Ciclo de Vida del track tecnológico
 - Diseño de la arquitectura

Es la definición de los planos que nos permitirán contar con un diseño integral que tome en cuenta los aspectos técnicos y elementos del Data Warehouse. Estos elementos son representados por medio de modelos que van en diferentes niveles de detalle mostrando los requerimientos inmediatos. Este diseño sigue 8 pasos:

- a. Establecer la fuerza de Arquitectura. Es conveniente definir a 3 personas en el diseño de la Arquitectura, estas tres personas son: arquitecto técnico, diseñador del área de staging y el desarrollador de aplicaciones.
- b. Se colectan los requerimientos del negocio. Esto se hace de acuerdo a las necesidades críticas del negocio, como pueden ser tiempos, disponibilidad, performance, etc.
- c. Documentación de los requerimientos de arquitectura. Se deberán documentar los hallazgos obtenidos a partir de las entrevistas, enfocado en los aspectos que pudieran impactar en la arquitectura.
- d. Desarrollo del modelo de arquitectura de alto nivel. Se clasifican los requerimientos de acuerdo a Datos de staging, accesos de datos, metadata e infraestructura.
- e. Diseño y especificación de los subsistemas. Se lleva a mayor cada uno de los grupos incluidos en el modelo de alto nivel, mostrando las capacidades y requerimientos específicos de cada sección.

- f. Determinar las fases de implementación de la Arquitectura. Se deben establecer prioridades para la implementación de las definiciones hechas de acuerdo a los requerimientos del negocio.
- g. Documentación técnica de la Arquitectura. Este documento debe contener la información necesaria para que se lleve a cabo la implementación del Data Warehouse.
- h. Revisión y finalización de la Arquitectura Técnica. Debe ser distribuido por los miembros de TI y el negocio con la finalidad de que sea retroalimentado este plan y quede completo para su validación.

- Selección del producto e instalación

De acuerdo a lo establecido en la planeación de la arquitectura, se busca por un producto que encaje con lo mencionado en dicho plan. Para hacer una buena selección se pueden realizar las siguientes actividades:

- a. Realizar una matriz de evaluación.
- b. Hacer una búsqueda en el mercado.
- c. Reducir opciones al mínimo para realizar evaluaciones detalladas.
- d. Requerir prototipos.
- e. Seleccionar producto, instalar prueba y negociar.
4. Ciclo de vida del track de datos

- Modelado dimensional

En el modelado dimensional identificamos las dimensiones que darán información de carácter cualitativo y los hechos que ofrecen información cuantitativa sobre el negocio. Para llegar a este modelado se realiza lo siguiente:

- Se hace una lista de las posibles dimensiones con sus intersecciones,
- Se identifican los procesos de negocio
- Se evalúa la granularidad, la consistencia, valore validos y la disponibilidad de los atributos.
- Se crea el esquema dimensional
- Se valida el esquema dimensional
- Se documenta el modelo

- Diseño físico

El modelado dimensional es traducido en un modelo físico, es muy probable que el modelado dimensional no se respete del todo puesto que en el modelo físico se deben tomar en cuenta ciertas estrategias que pueden hacer que la implementación del modelado dimensional en el modelo físico no sea tan transparente, esto debido a temas de agregación, índices, etc.

- Diseño y desarrollo del área de staging

Esta parte implica el diseño y desarrollo del proceso de ETL. Para este proceso primero se deben trabajar las dimensiones. Este proceso generalmente se divide en 2 secciones: las dimensiones y los hechos.

5. Ciclo de vida del track de aplicaciones analíticas

- Especificaciones. De acuerdo a todo lo desarrollado hasta este punto, es necesario generar las vistas que los usuarios accederán, mediante herramientas de reporte. Se hace toda la definición del front end
- Desarrollo. El desarrollo de las actividades analíticas definidas se lleva a cabo bajo ciertos estándares.

6. Implementación. Es el llevar todo el desarrollo del Data Warehouse al día a día de las operaciones para lo cual se requiere educar a los usuarios y ofrecer cierto tiempo de soporte por cualquier contingencia que pudiera ocurrir.

Mantenimiento y crecimiento. Para el mantenimiento se realizan tareas de soporte, soporte técnico, educación, todo siguiendo un programa de soporte. Además de que se puede crecer el proyecto cuando exista la necesidad de hacerlo.

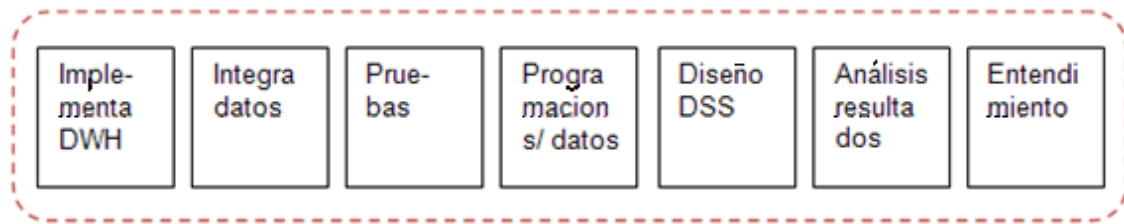
2.1.7. METODOLOGIA INMON

Bill Inmon es considerado el padre del concepto Data Warehouse, el menciona que un Data Warehouse debe cumplir con las siguientes características:

- Dirigido a un área. Datos sobre un área específica en lugar de operaciones de la compañía
- Integrado. Unión de diferentes fuentes de datos de manera coherente
- Variable en el tiempo. Todos los datos pertenecen a un periodo de tiempo determinado
- No volátil. Los datos no son eliminados.

La metodología que Bill Inmon propone es iterativa la cual sigue un esquema contrario al clásico de desarrollo de sistemas ya que lo primero con lo que se trabaja son datos, estos se integran para ser probados y programar de acuerdo a ellos para analizar los resultados y de

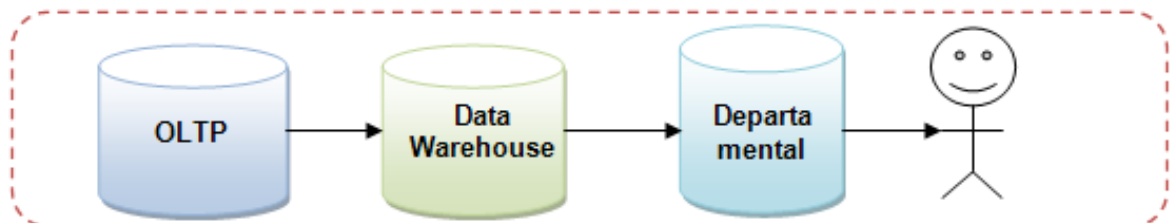
esta manera comprender los requerimientos. La metodología principalmente consiste en lo siguiente:



Fuente: Diseño propio

II-10. Metodología Bill Inmon

Dentro de esta metodología se menciona que la construcción de toda la arquitectura de un Data Warehouse toma bastante tiempo, puesto que su desarrollo inicial está relacionado con necesidades genéricas empresariales, a lo largo del tiempo este tipo de necesidades son cubiertas por el Data Warehouse para mas personas por lo que la demanda del uso del Data Warehouse aumenta y esto hace que el performance se vea afectado. Es por esto que al llegar a este punto se comienzan a construir segmentos del Data Warehouse que se alimentaran del Data Warehouse y que permitirán tener la información almacenada de manera que esta vaya dirigida a departamentos, con esto se logra disminuir la demanda sobre el Data Warehouse debido a que por ejemplo para estos momento en lugar de tener a 100 usuarios requiriendo de manera directa los servicios del Data Warehouse tendré 5 departamentos.



Fuente: Diseño propio

II-11. Implementaciones

1. Implementación del Data Warehouse.

- a. OLTP. El primer paso para la implementación de un Data Warehouse es el identificar las fuentes de datos, analizarlas y mapear sus elementos de acuerdo al estándar que hayamos definido. Esto en el orden de tratar de homologar los datos que sea posible para su entrada al Data Warehouse.
- b. Modelos de Procesos. Se debe tener conocimiento de los procesos que sigue la información y para eso nos sirve el modelo de procesos. Este modelo contiene información como:
 - Descomposición funcional
 - Diagrama de contexto

- Diagrama de flujo de datos
 - Diagrama de transición de estados
 - Pseudocódigo
- c. Modelo de datos. Se trabaja con 2 tipos de modelos:
- El Modelo de datos nos muestra los datos primitivos, tomando en cuenta el elemento tiempo, se plasman los cálculos que se realicen y finalmente se muestran sus relaciones.

El Modelo de Datos del Data Warehouse. Los modelos anteriores nos deberán entregar la definición de los sujetos a los que estará orientado el Data Warehouse. Debe venir en 3 perspectivas y son explicadas en la siguiente tabla:

Modelo	Descripción
Modelos de alto nivel - DER	Representa entidades y sus relaciones. Es el nivel más alto de abstracción. Y debe tener definido un alcance de integración.
Modelo de nivel medio - DIS	Para cada entidad definida en el DER se detalla más su composición de manera que para cada entidad tendremos un conjunto de elementos que se relacionan a esta entidad. Estos elementos tienen un dato que los identifica y un dato que los relaciona con otro elemento.
Modelo de nivel bajo - PM	Es una representación de cómo quedaran las tablas en el Data Warehouse con sus campos, tipos de datos, relaciones, etc.

Tabla 5. Modelado de datos

- d. Una vez que se tiene conocimiento de este modelo se deben tomar ciertas decisiones sobre el diseño del Data Warehouse. Entre estas decisiones tenemos las siguientes:
- Normalización, debemos decidir el grado al que nuestro Data Warehouse
 - Granularidad
 - Particiones
 - Minería de Datos
- e. Al haber tomado estas decisiones, se debe generar un documento que contenga estas decisiones que hemos tomado para la definición del Data Warehouse. Este documento debe contener un concepto de Data Warehouse, una descripción de los sistemas que lo alimentan, como se debe usar el Data Warehouse, como obtener ayuda, responsables, plan de migración, mapeo de datos entre los datos operacionales y el data Warehouse, etc.

- f. Metadata. Contiene información sobre nuestro Data Warehouse. En pocas palabras es un diccionario de datos. Es pieza clave para el mejor aprovechamiento del Data Warehouse. Facilita las tareas de análisis ya que funciona como un índice del contenido del Data Warehouse.
2. Integración de datos. Implica el implementar procesos ETL que nos permitan extraer la información de los ambientes transacciones para cargarlo dentro del Data Warehouse. Esto puede implicar un cambio en la tecnología, selección de los datos que residirán en el Data Warehouse, cambios de llaves en los objetos, formato de los datos, sumalizaciones, estandarización de nomenclaturas,
3. Pruebas. Se hacen pruebas al respecto de la implementación del Data Warehouse. Se realizan los ajustes necesarios para poder obtener los resultados esperados en nuestro Data Warehouse.
4. Programación. Se hacen las programaciones necesarias para que se ejecuten ciertos procesos, para que exista la posibilidad de paralelismo, se administra la Meta Data, índices, particiones, monitoreo, etc.
5. Diseño DSS. Se trabaja sobre un esquema multidimensional para poder generar la información que realmente soporte la toma de decisiones.
6. Análisis. El tomador de decisiones analiza la información obtenida a partir del DSS.
7. Requerimientos. A partir del análisis de los datos obtenidos el tomador de decisiones llegue al entendimiento de los requerimientos que tiene su negocio para mejorar.

A grandes rasgos esta es la metodología que Bill Inmon propone y que forma parte del marco de referencia de este trabajo de investigación.

2.1.8. X-Warehousing

El objetivo de esta metodología es el homogenizar documentos XML, está enfocada principalmente en los datos y no tanto en las fuentes de los mismos. Para lograr su cometido, se inicia con el análisis de objetivos definidos por los usuarios para poder generar un modelo conceptual multidimensional. En base a esto se puede modelar un Data Warehouse por medio de esquemas XML que hacen referencia a un modelo de cubos teniendo en un documento XML la información de los hechos correspondientes, asignando las dimensiones (en esquema estrella o copo de nieve) correspondientes. Una vez que se logra esto, es necesario generar los árboles de atributos mediante la aplicación de algoritmos. Después se genera el cubo XML, mediante la comparación del MCM y los documentos XML, esto se logra mediante la aplicación de las funciones de punning y grafting que permiten obtener los atributos en común. Para validar que el árbol de atributos generado es correcto, se evalúa contra un Documento XML de contenido mínimo el cual permite garantizar que el árbol generado cuenta con la información mínima necesaria para permitir el análisis. (Boussaid, Ben Messaoud, Choquet, & Anthoard, 2006).

Para esta metodología sus creadores generaron una herramienta en Java que está dividida en dos módulos:

- Modulo de carga. Carga los XML de entrada y el árbol de atributos.
- Modulo de combinación. Permite generar el documento XML homogéneo.

Por lo tanto en base a lo modelado que está relacionado con lo definido por el usuario, se procede a homogenizar los documentos XML mediante su comparación con un árbol de atributos.

2.1.9. EBAF Conversion Model

Esta metodología está enfocada en negocios basados en Web (Pesaran Behbahani, 2012). Se basa en el uso de estructuras y modelos de minería de datos para cumplir con el objetivo de medir la eficiencia de la mercadotecnia aplicada al negocio para poder generar clientes recurrentes.

Para lograr esto se basa en 5 fases:

- Conciencia. Es una métrica definida para medir la eficiencia de los esfuerzos de marketing para dar a conocer el negocio al público.
- Contacto. Métrica que permite medir la eficiencia de las actividades de marketing para influenciar lo suficiente al público en general para que se interese en entrar en contacto con la empresa.
- Compromiso. Métrica que permite medir la eficiencia de mantener la atención del público en general que ya entablo contacto con la empresa.
- Conversión. Métrica que permite medir la eficiencia de convertir a un sujeto en cliente.
- Retención. Métrica que permite medir la eficiencia de convertir un cliente en cliente recurrente.

2.1.10. Comparativa

Todas las metodologías presentadas están relacionadas con aspectos de Inteligencia de Negocios, a continuación mostraremos un breve resumen en forma de comparativa de dichas metodologías.

METODOLOGIA	ENFOQUE	CARACTERISTICA	RESUMEN
CRISP – DM	Minería de Datos	Visión de Negocio	1. Entendimiento del negocio 2. Entendimiento de los datos 3. Preparación de los datos 4. Modelado

METODOLOGIA	ENFOQUE	CARACTERISTICA	RESUMEN
			<ol style="list-style-type: none"> 5. Evaluación 6. Implementación
HEFESTO 2.0	DWH	Construcción de DWH	<ol style="list-style-type: none"> 1. Análisis de requerimientos 2. Análisis OLTP 3. Modelo lógico DWH 4. Integración de datos
Metodologías Agiles	Software – BI	Desarrollo de Soluciones de BI	<ol style="list-style-type: none"> 1. Listado de prioridades (cliente) 2. Listado de tareas (desarrolladores) 3. Ciclo de desarrollo <ol style="list-style-type: none"> a. ETL b. Modelo c. Presentación d. Publicación
KIMBALL	DWH	Diseño de DWH	<ol style="list-style-type: none"> 1. Planeación del proyecto 2. Requerimientos del negocio 3. Diseño Arquitectura Técnica 4. Selección e instalación de herramientas 5. Modelado dimensional 6. Diseño físico 7. Especificaciones herramienta analítica 8. Desarrollo herramienta analítica 9. Área de Staging 10. Desarrollo 11. Mantenimiento
INMON	DWH	Conceptos de DWH	<ol style="list-style-type: none"> 1. Implementación DWH 2. Integración de Datos 3. Pruebas 4. Programación sobre datos 5. Diseño DSS 6. Análisis resultados 7. Entendimiento
X-Warehousing	DWH y Cubos	Modelado Conceptual Multidimensional	<ol style="list-style-type: none"> 1. Modelado conceptual multidimensional 2. Árboles de atributos 3. Funciones de combinación 4. Homogenización de documentos

METODOLOGIA	ENFOQUE	CARACTERISTICA	RESUMEN
			XML
EBAF Conversion Model	Minería de Datos	Aplicación Minería de Datos para medir la eficiencia de mercadotecnia	<ol style="list-style-type: none"> 1. Estructuras de Minería de datos 2. Modelos de Minería de datos <ol style="list-style-type: none"> a. Conciencia b. Contacto c. Compromiso d. Conversión e. Recurrencia

Tabla 6. Comparativa de Metodologías

Herramientas BI OS

Esta sección tiene por objetivo recopilar algunas de las herramientas de Inteligencia de tipo open source que se encuentran en el mercado con la finalidad de mostrar que el mercado de herramientas open source es lo suficientemente amplio para competir con el mercado propietario.

2.1.11. Jaspersoft (01)

Es una plataforma de Business Intelligence que está destinada para generar soluciones en empresas pequeñas y medianas. Esta plataforma cuenta con herramientas que permiten hacer el desarrollo completo de una solución de BI, partiendo desde la extracción de la información de las fuentes de datos para ser almacenadas en un repositorio de datos, para que posteriormente estos sean explotados con herramientas de análisis para que finalmente sean visualizados por los usuarios de negocio en diferentes niveles. Para lograr esto la plataforma cuenta con las siguientes herramientas:

- Jaspersoft ETL.

Permite desarrollar, administrar y documentar los procesos de ETL en una organización. Dichos procesos servirán para poblar el repositorio de datos.

- Jaspersoft OLAP

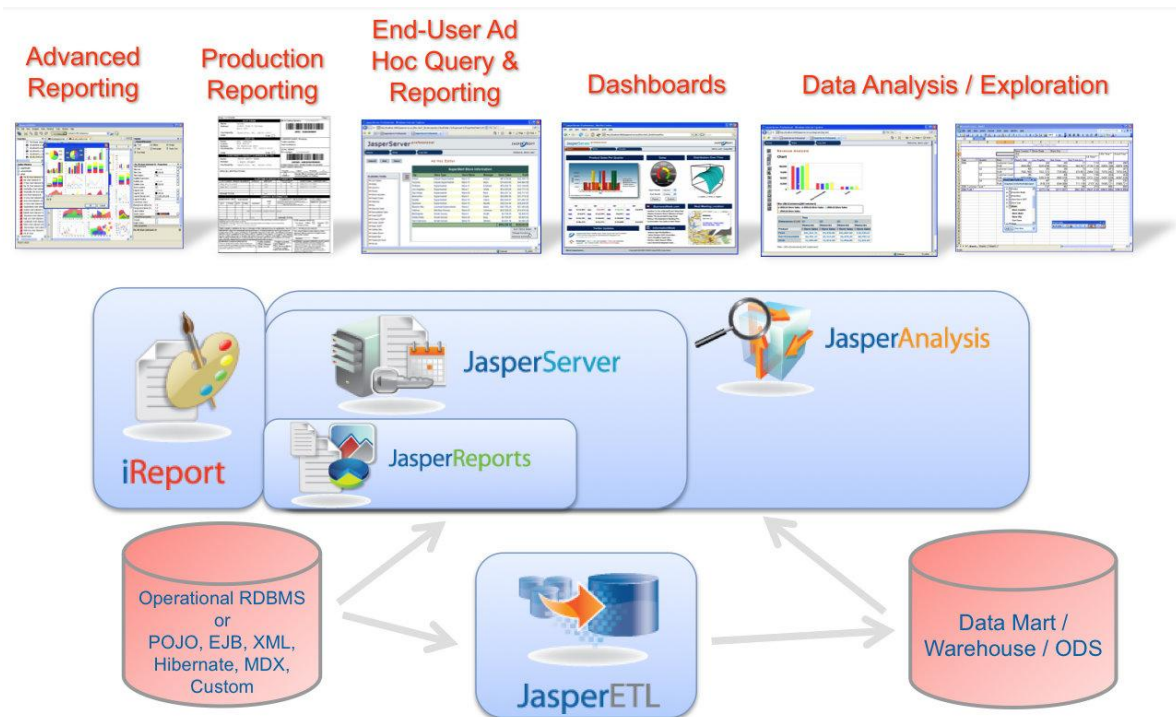
Esta herramienta permite hacer análisis multidimensional sobre los datos. Lo que permite obtener una mejora en la respuesta de las consultas de dicha información debido a las agregaciones que son almacenadas en la metadata del servidor OLAP.

- JaspersoftReport Server

Es el servidor que permite la administración de la publicación de los reportes generados para su visualización en Web.

- iReports Designer

Herramienta que permite diseñar reportes que pueden contener imágenes, sub-reportes, gráficas y matrices.



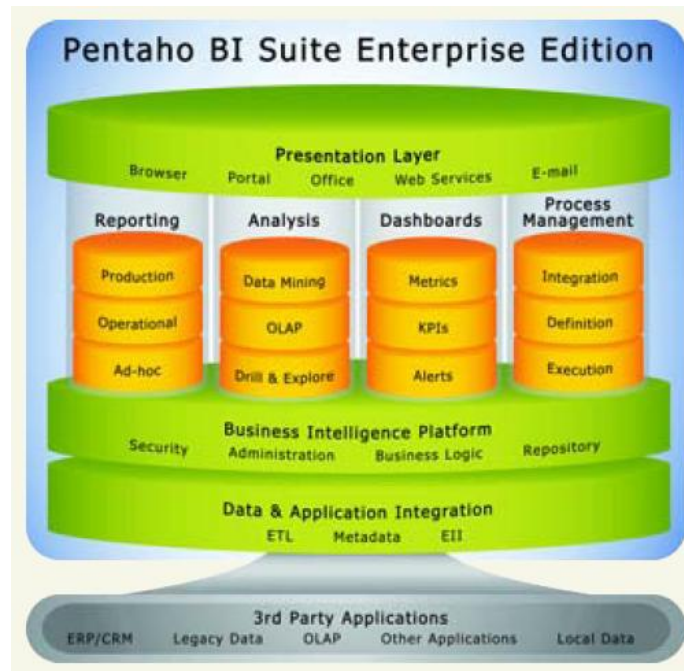
Fuente: Jaspersoft

II-12. Arquitectura Jaspersoft

2.1.12. Pentaho (02)

Es una suite de Business Intelligence pionera ofrece servicios críticos como calendarización seguridad, integración, navegación por contenido y provee las siguientes funcionalidades:

- Data Integration
- Reporting
- Analysis
- Dashboard
- Data Mining



Fuente: Pentaho

II-13. Pentaho

2.1.13. OpenI (03)

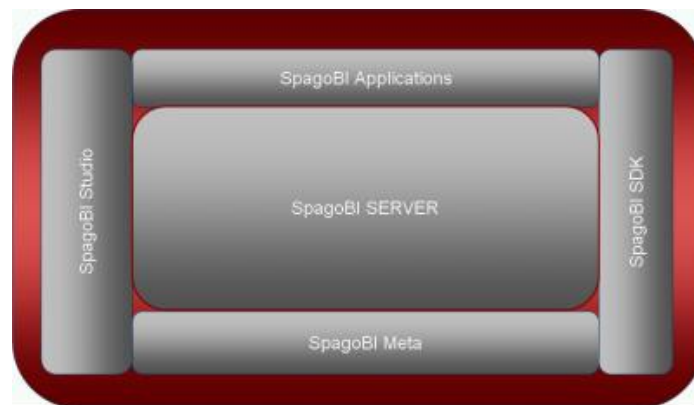
Es una aplicación de código abierto para soluciones Business Intelligence a la medida. Esta desarrollado en J2EE y puede trabajar con datos OLAP o bases de datos relacionales. Actualmente es un plugin que trabaja junto con Pentaho para visualizar datos multidimensionales, este plugin permite el desarrollo y publicación de reportes XMLA con fuentes OLAP.

2.1.14. SpagoBI (04)

Es una plataforma de código abierto que cubre todas las aéreas analíticas de Business Intelligence. SpagoBI ofrece lo siguiente:

- Reporting
- OLAP
- Chart
- Dashboard
- KPI
- Cockpits
- GEO/GIS
- Data Mining
- QuerybyExample

- Smart Filter
- Accesibility Reporting
- RT Console
- Dossier
- ETL
- Office



Fuente: SpagoBI

II-14. SpagoBI

2.1.15. CloverETL (05)

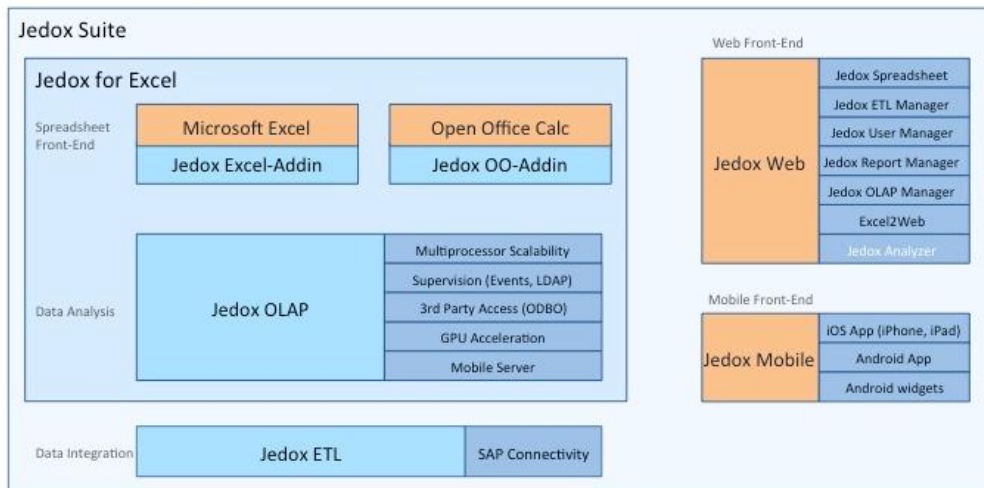
Es una plataforma de Integración de Datos que permite definir y crear transformaciones de datos en Java (Grabova, Darmont, Chauchat, & Zolotaryova, 2010). Es considerada por sus creadores como una plataforma de integración de datos de alto desempeño que permite mover datos de diferentes localidades.

Cuenta con una interfaz gráfica que facilita tanto el desarrollo como la administración de las soluciones. También ofrece diversas ediciones que buscan ajustarse a los presupuestos y necesidades existentes.

2.1.16. Palo (06)

Es un servidor de bases de datos multidimensionales que permite almacenar los datos en cubos, dimensiones elementos y elementos de atributos. Así mismo cuenta con herramienta para la integración de datos, servidor web, para presentación de datos en móvil, y con una suite completa para soluciones de BI.

Jedox Suite Overview



Fuente: Palo

Ilustración II-15 - Palo

2.1.17. PocOLAP (07)

Es una solución ligera para datamining. Provee una interfaz web, y visibilidad de hoja de cálculo que permite un uso más amigable de la herramienta. Su objetivo es ser una herramienta amigable, su desarrollo está realizado en Java.

2.1.18. KNIME (08)

De acuerdo a un estudio realizado por (Chen, Ye, Williams, & Xu, 2007), está fue una de las herramientas mejor evaluadas. Es una plataforma que permite a integración, procesamiento, análisis y exploración de datos por medio de una interfaz amigable. Cuenta con diversas versiones enfocadas a diferentes necesidades de análisis de los datos.



Fuente: Diseño propio

Ilustración II-16 - Versiones KNIME

2.1.19. AlphaMiner (09)

De acuerdo a un estudio realizado por (Chen, Ye, Williams, & Xu, 2007), está fue una de las herramientas mejor evaluadas. Es una plataforma de minería de datos considerada totalmente enfocada al mercado de las PyMEs. Cuenta con una herramienta para la construcción de casos, tiene la opción de extender las capacidades de la herramienta mediante plug-ins, y cuenta con funciones poderosas de minado de datos.



Fuente: AlphaMiner

Ilustración II-17 - Logo AlphaMiner

2.1.20. Weka (10)

De acuerdo a un estudio realizado por (Chen, Ye, Williams, & Xu, 2007), está fue una de las herramientas mejor evaluadas. Es una colección de algoritmos de minería de datos, permite el pre.procesamiento de los datos, clasificación, regresión, clustering, reglas de asociación y la visualización.



Fuente: Weka

Ilustración II-18 - Logo WEKA

2.1.21. Rapid Miner (YALE) (11)

De acuerdo a un estudio realizado por (Chen, Ye, Williams, & Xu, 2007), está fue una de las herramientas mejor evaluadas. En aquel entonces era conocida como YALE, es una plataforma de minería de datos considerada la herramienta líder. Permite la integración de datos, el

análisis de datos y el reporte de los resultados. Cuenta con una interfaz gráfica, tiene un repositorio para el manejo de procesos, datos y metadatos.



Fuente: RapidMiner

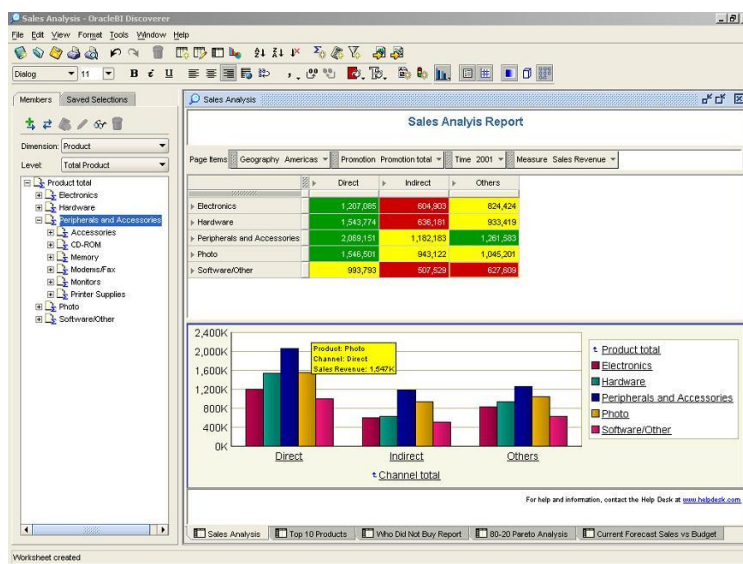
Ilustración II-19 - Logo RapidMiner

Herramientas BI Propietarias

Esta sección tiene por objetivo el mostrar algunas de las herramientas propietarias de inteligencia de negocios que se encuentran disponibles en el mercado. Las herramientas enlistadas tienen ya una larga trayectoria y son reconocidas mundialmente.

2.1.22. Oracle BI (12)

Es la plataforma más completa para la inteligencia de negocios (BI) disponible en la actualidad, cubriendo un amplio espectro de necesidades de inteligencia de negocios, incluidos los tableros interactivos, el análisis ad-hoc, alertas e inteligencia proactivas, publicación e informes avanzados, análisis predictivo en tiempo real, análisis de tecnología móvil, y mucho más.



Fuente: OracleBI

II-20. Oracle BI

2.1.23. MicroStrategy (13)

Es una plataforma que ofrece productos que ayudan a las empresas a soportar las necesidades cambiantes del negocio para grupos de trabajo y aplicaciones departamentales de BI, facilita la migración de aplicaciones Bi departamentales a una arquitectura empresarial. Su principal virtud es la capacidad que tiene para visualizar los datos de manera atractiva, es una herramienta que se ve limitada por no contar con elementos que permitan transformar datos. Esta totalmente enfocada en la presentación de los mismos.

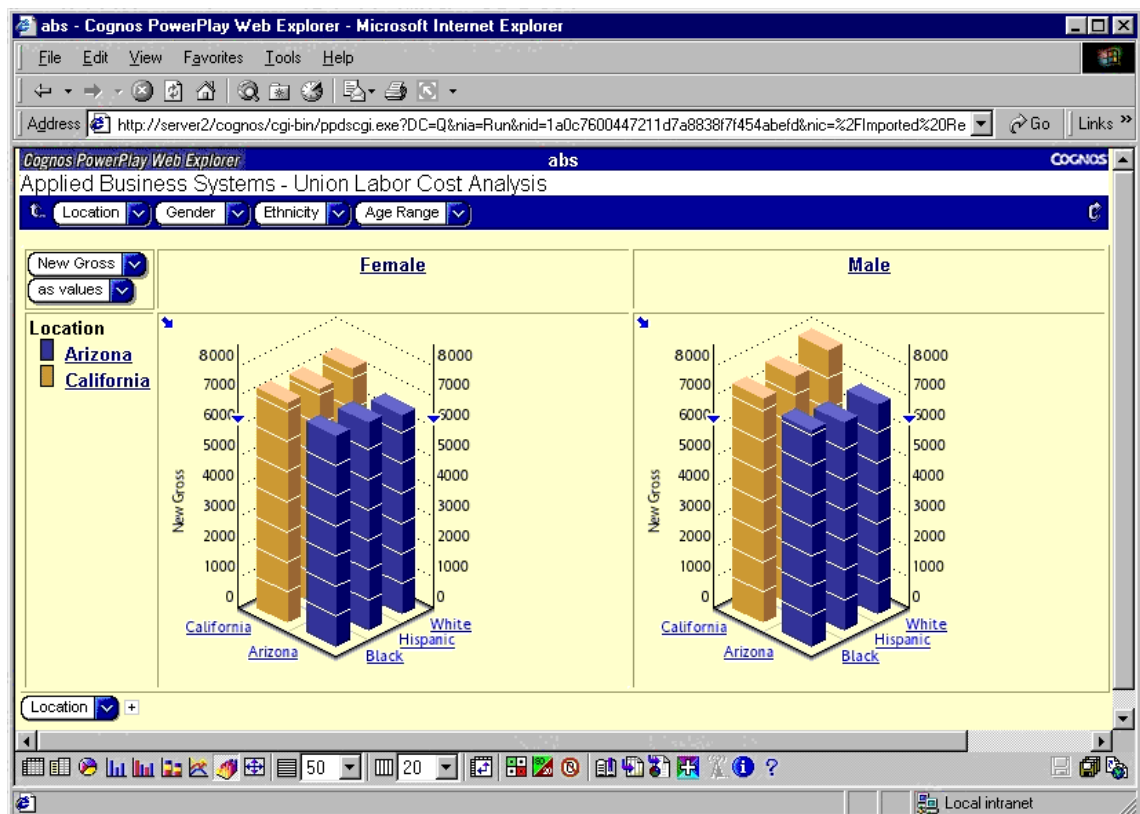


Fuente: Microstrategy

II-21. Microstrategy

2.1.24. Cognos (14)

Es una plataforma empresarial, permite la entrega completa consistente y en tiempo de la información para todos los usuarios dentro de una infraestructura altamente estable. Satisface las necesidades de inteligencia de negocios como reportes, dashboards, scorecards, análisis y planeación reduciendo la complejidad del ambiente de Business Intelligence.



Fuente: COGNOS

II-22. COGNOS

2.1.25. Microsoft (15)

Microsoft BI Platform, es una plataforma completa para el análisis de datos e inteligencia de negocios, está basada en Microsoft SQL Server y proporciona capacidades de reportes, análisis e integración de datos sin precedente.

Report Manager - Windows Internet Explorer

http://reah/Reports_KATMAI/Pages/Report.aspx?ItemPath=%2fAdventureWorks+2008+Sample+Reports%2fSales+Trend+2008

SQL Server Reporting Services
Home > AdventureWorks 2008 Sample Reports >
Sales Trend 2008

Home | My Subscriptions | Site Settings | Help

Search for:

View Properties History Subscriptions

1 of 1 100% Find | Next Select a format Export

Adventure Works Sales Analysis

Products by Subcategory, Quarter & Trend		2002			2003					2004				
		Q3	Q4	Total	Q1	Q2	Q3	Q4	Total	Y / Y %	Q1	Q2	Total	Y / Y %
Accessories	Bike Racks						\$75,920	\$60,883	\$136,804		\$37,150	\$63,084	\$100,234	-36 %
	Bike Stands						\$6,996	\$11,925	\$18,921		\$8,268	\$10,653	\$18,921	0 %
	Bottles and						\$11,854	\$15,968	\$27,822		\$16,105	\$18,659	\$34,764	-20 %
	Cleaners						\$5,137	\$4,724	\$9,862		\$3,358	\$4,949	\$8,307	-19 %
	Fenders						\$7,649	\$11,759	\$19,408		\$11,583	\$13,276	\$24,859	22 %
	Helmets	\$33,653	\$24,671	\$58,724	\$11,660	\$25,524	\$81,538	\$89,595	\$208,317	72 %	\$72,597	\$101,789	\$174,386	-19 %
	Hydration Packs						\$31,577	\$27,110	\$58,687		\$17,904	\$28,254	\$46,158	-27 %
	Locks	\$6,325	\$3,780	\$10,105	\$2,205	\$3,839		\$15	\$6,159	-64 %				
	Pumps	\$5,157	\$3,226	\$8,383	\$1,763	\$3,382			\$5,145	-63 %				
	Tires and Tubes						\$41,940	\$81,948	\$103,889		\$61,519	\$66,617	\$128,337	-19 %
Bikes	Mountain Bikes	\$3,141,467	\$2,837,647	\$5,979,114	\$2,517,500	\$2,908,659	\$3,617,012	\$3,808,656	\$12,851,826	53 %	\$3,473,750	\$4,268,134	\$7,741,884	-66 %
	Road Bikes	\$4,930,693	\$4,189,622	\$9,120,315	\$3,584,255	\$4,119,659	\$3,844,124	\$3,734,892	\$15,282,929	40 %	\$3,391,876	\$4,001,737	\$7,393,614	-107 %
	Touring Bikes						\$3,298,008	\$3,765,585	\$7,064,592		\$3,414,027	\$4,068,455	\$7,480,482	6 %
Clothing	Bib-Shorts	\$66,860	\$35,323	\$102,183	\$21,544	\$43,458	\$351		\$65,353	-56 %	\$234	\$234	\$468	-13,866 %
	Cans	\$3,001	\$3,076	\$7,026	\$1,793	\$3,647	\$8,678	\$8,618	\$71,617	40 %	\$7,316	\$5,965	\$13,581	-17 %

Local Intranet 100%

Fuente: Microsoft

II-23. SQL Server Reporting Services

2.1.26. QlikView (16)

Provee un motor de ETL y no requiere de datos pre-agregados. Permite el análisis a cualquier nivel de detalle. Permite una conexión automática de las tablas para lo cual es necesario que el modelo de datos contenga los mismos nombres en todas las tablas cuyos conceptos se deban ligar (Grabova, Darmont, Chauchat, & Zolotaryova, 2010).

Está enfocado en soluciones de inteligencia de negocios de autoservicio lo que quiere decir que buscan dar autonomía al usuario final para que pueda generar los reportes que requiera en el momento que así lo decida.

Mejores prácticas

Las mejores prácticas son aquellas que les han funcionado a otras empresas y que es casi un hecho que te funcionarían a ti. En cuanto a lo relacionado con soluciones de Business Intelligence tenemos las siguientes mejores prácticas:

1. Conocer y evaluar la estrategia de Business Intelligence de la empresa
2. Se deben definir los KPI necesarios para la empresa

3. La calidad de los datos es fundamental por lo que debe de existir toda una infraestructura que garantice la exactitud, calidad y velocidad de disponibilidad de los datos.
4. Se deben usar los KPI para alinear al negocio y TI
5. Seguir la regla de tener pocos KPI que definan el estado del negocio.
6. Asegurarse de que exista un diccionario de datos o en su caso generar uno.
7. Los procesos de Business Intelligence deben estar fuertemente relacionados con los KPI's que sean definidos.
8. Monitoreo constante de la solución de Business Intelligence para asegurar el cumplimiento de los objetivos.
9. Debe existir la flexibilidad y se debe de contar con un gobierno para asegurar la transformación de los datos en información.
10. Guardar planes de consulta en los cuales tengamos un registro de los tiempos de las ejecuciones realizadas de manera que podamos medir el rendimiento del Data Warehouse.
11. Se deberán guardar los errores ocurridos durante el proceso de ETL en tablas especiales para dicho fin.

Se debe de contar con modularidad en los procesos con la finalidad de que estos no tengan duraciones excesivas y que se acaben los recursos antes de que otros procesos puedan ser ejecutados.

Para la construcción de un Data Warehouse se deben hacer ciertos cuestionamientos que se deben considerar (Gameiro, 2011):

- Costos.
- Tiempo
- Usuarios
- Recursos Humanos
- Hardware
- Servicios

FCE

Hay algunos aspectos que se deben tomar muy en cuenta cuando se realiza la implementación de una solución de Business Intelligence. Entre estos tenemos (Yeoh & Koronios, 2009):

Punto de vista organizacional.

1. Un fuerte compromiso del equipo de trabajo (TI y negocio) y una buena administración de proyectos, así mismo se considera importante que el patrocinador del proyecto pertenezca al negocio propiamente y no a TI.
2. Visión clara y un caso de negocios bien establecido, debe existir una alineación con la estrategia del negocio.
3. Nunca se debe perder el foco sobre los requerimientos del negocio y sus procesos, recordemos que lo que se busca es soportar la toma de decisiones, pero no podremos soportar dicha toma de decisiones si no se conocen los requerimientos del negocio y los procesos que generan dichos requerimientos.
4. Se debe tener como un objetivo alcanzar niveles aceptables de ROI. Es uno de los puntos más delicados y generalmente es uno de los puntos que nos permiten vender la idea de implementar una solución de Business Intelligence por lo que se debe cuidar que se logre cumplir con dicha promesa.

Punto de vista de procesos

5. Un equipo de trabajo bien conformado. Contando con un experto en el negocio y un equipo de desarrollo balanceado en conocimientos es fundamental para el éxito de los proyectos de BI.
6. Una buena definición del proyecto de tal manera que permita realizar tareas de manera iterativa permitirán lograr el éxito en la implementación de los proyectos de Business Intelligence.
7. Se deben soportar los cambios que puedan surgir a partir del usuario y deben de ser administrados. Es esencial la participación de los usuarios a los que se dirigirá la solución puesto son estos a quienes se debe complacer y para lograrlo es de suma importancia contar con su participación.

Punto de vista tecnológico

8. Se debe contar con la posibilidad de escalar las soluciones por lo que nuestro esquema de trabajo debe ser flexible que nos permita dirigir al negocio a donde lo requiere.
9. Para que las soluciones de Business Intelligence tengan éxito, los datos deben contar con calidad para que estos puedan ser integrados sin problemas y se logren los resultados esperados.

Capítulo III Propuesta de Solución “Metodología Propuesta Pyme-PyC”

La metodología propuesta en este trabajo de tesis pretende entre otras cosas ajustarse a la realidad de las PyMEs. Es por esto que siguiendo lo indicado en un artículo (Gameiro, 2011), se establece que de acuerdo al nivel de madurez de la empresa es el requerimiento de decisiones, por lo que a mayor nivel de madurez se requiere una mejor toma de decisiones.

De acuerdo a la investigación realizada, diversas fuentes consultadas concluyen en que se debe mantener la sencillez en la solución que se proponga para una PyME, así mismo no debe de representar un incremento en los costos de la empresa. Debe de ser amigable permitiendo la optimización de los procesos actuales.

Según el estudio realizado por Sadok y Lesca, existen 7 condiciones necesarias de aceptación para una buena implementación de una solución de inteligencia de negocios (Sadok & Lesca, 2009):

1. Debe ser una solución simple.
2. Debe evitar el almacenamiento de los datos.
3. Debe estar basado en el uso de fuentes de datos relacionales.
4. Debe considerar la explotación e interpretación de datos informales
5. Debe tener una organización que permita la reducción de tiempos.
6. Debe evitar la formalización de la implementación del conocimiento para la interpretación de los datos.
7. Debe proporcionar resultados de manera inmediata.

Otros aspectos importantes a tomar en cuenta para la propuesta de metodología de este trabajo, son resultado del estudio realizado por (Scholz, Schieder, Kurze, Gkuchowski, & Boehringer, 2010), en el cual se identificaron los principales retos a los que se enfrenta una implementación de inteligencia de negocios en una empresa alemana:

- De acuerdo al uso de la solución de inteligencia de negocios, se debe considerar que no debe ser muy complicada, y se debe contemplar una capacitación que permita un mejor entendimiento de la solución a implementar.

- Se debe elegir el software correcto que no genere más problemas de los que solucione, así como se debe guardar especial cuidado con la calidad de los datos y debe dar tiempos de respuesta satisfactorios.
- Debe contener las funcionalidades que requiera el usuario final.

Para introducir a las empresas PyME en el uso de la inteligencia de negocios, de acuerdo a un estudio realizado en empresas Poblanas, utilizaron las siguientes métricas base para determinar medir el desempeño de una empresa (Bruhn, Karlan, & Schoar, 2012):

1. Beneficio
 - a. Utilidad = Ventas del mes – Costos del mes
 - b. Margen de Utilidad = (Ventas del mes / Costo del mes)*100
2. Productividad
 - a. ROA = beneficios / activos

De acuerdo a la experiencia de un consultor plasmada en un artículo (Williamw & Williams, 2003), para que la implementación de una solución de inteligencia de negocios realmente aporte valor al negocio es necesario que este impacte en los procesos operativos a partir de proporcionar información al área administrativa. De tal manera que al hacer esto se mejoren tanto los procesos administrativos como operativos.

Según el artículo recomienda que se realice mediante un esquema en el cual se contemple la ingeniería de procesos para entender a la perfección la manera en la que la Inteligencia de Negocios va a interactuar para generar el valor deseado.



Fuente: Diseño propio

Ilustración III-1 - Esquema para dar valor al negocio

También es importante involucrar al personal en los cambios que sean necesarios para implementar una solución de BI. De acuerdo a un estudio (Laumer, Maier, Weitzel, & Eckhardt, 2012) se deben de considerar como principales variables en el éxito de la implementación de un sistema: la salud, la satisfacción con el empleo, el compromiso con la empresa y las intenciones de cambiar de empleo. Todas estas variables son dependientes entre sí puesto que los empleados frente a un nuevo sistema de información pueden ver afectada su salud lo

que a su vez puede afectar la satisfacción del empleado ante su trabajo afectando el compromiso con la organización viéndose afectadas las intenciones de cambio de empleo.

De acuerdo a una evaluación de los servicios de información (Lee, Ko, & Jeong, 2009), menciona que se debe considerar la eficiencia y efectividad del sistema. Consideran conveniente la evaluación de la calidad del sistema y la información, el uso, la satisfacción del usuario, el impacto. Las conclusiones de esta evaluación dicen que la calidad en la información, el sistema y el servicio fomentaran el uso del sistema. También establecen que a mayor madurez en el uso de las TI mayor será el uso.

Tomando como referencia todos estos antecedentes, en este trabajo de tesis se propone que de acuerdo a un nivel de madurez en el manejo de la información es la recomendación que se hace sobre la metodología que se propone y detalla a continuación.

La metodología propuesta está dividida en 5 fases:

- Negocio
- Planeación
- Desarrollo
- Validación
- Implementación

De acuerdo al nivel de madurez en el manejo de la información en la empresa, se recomienda la implementación de cada fase.

Fases de la Metodología	Bajo No cuenta con sistemas de información	Medio Cuenta con sistemas de información	Alto Cuenta con sistemas de información y un Data Warehouse
Negocio	Ok	Ok	Ok
Planeación	-	Ok	Ok
Desarrollo			
Modelado	OK	OK	OK
Datos	OK	OK	OK
Cubos	-	OK	OK
Minería de Datos	-	-	OK
Presentación	Tablas pivote	Web	Web
Validación	OK	OK	OK
Implementación	OK	OK	OK

Tabla 7 - Solución de acuerdo al nivel de madurez

Negocio

El principal objetivo de esta sección es tener un conocimiento del negocio y en el caso de ser necesario trabajar a la par con los expertos del negocio para establecer los conceptos principales que ayudaran a la empresa a dar claridad sobre su rumbo para poder establecer las

estrategias que le permitan dirigirse a los objetivos que se planteen, así como definir las métricas necesarias que apoyen en la determinación del cumplimiento de dichos objetivos.

Esta fase es recomendada en cualquiera de los niveles de madurez en el manejo de la información. El ejecutar esta fase permite llegar a un entendimiento del negocio.

3.1.1. Planeación estratégica

La planeación estratégica permite que las empresas puedan definir claramente quienes son, a donde se quiere ir y como llegar.

3.1.1.1. Misión

La misión permite delimitar quiénes somos y que hacemos, describiendo en un pequeño párrafo los principales servicios y/o productos que se ofrecen, así como el mercado al que nos dirigimos. Es un reflejo de nuestro presente y es la carta de presentación con los clientes.

3.1.1.2. Visión

Ya que sabemos la misión de la empresa, se puede establecer su visión. La visión es una vista al futuro de lo que esperamos lograr con la empresa, el crecimiento que esperamos que esta tenga, todo descrito en un pequeño párrafo. Esta definición será crucial para la definición de los “cómo” que guiarán a la empresa.

3.1.1.3. Objetivos

A partir de la visión se pueden definir los objetivos que servirán de base para hacer que la visión de la empresa se vuelva una realidad, Los objetivos se definen en frases cortas que se escriben de manera infinitiva. Son descritos de manera genérica y su finalidad es meramente de guía y para tener plasmado el camino que se debe seguir para cumplir la visión, de esta manera se pueden consultar cada que sea necesario.

3.1.1.4. Metas

Las metas están relacionadas con los objetivos, son frases cortas que complementan los objetivos incluyendo valores numéricos que faciliten la medición del cumplimiento de los objetivos por medio del cumplimiento de las metas. Las metas son más específicas que los objetivos.

3.1.1.5. Factores Críticos de éxito

Los factores críticos de éxito son aquellos que forzosamente deben o no presentarse para el cumplimiento de las metas y por ende para el logro de los objetivos definidos.

3.1.2. Procesos de negocio

Los procesos de negocio son las actividades de la empresa que guardan cierta relación y que requieren de un insumo para poder generar algún resultado. La definición de los procesos de negocio es de gran utilidad para la definición de estrategias y para generar las mejoras necesarias para la operación de la empresa.

Este conjunto de actividades están íntimamente ligadas con la operación de la empresa tomando en cuenta las principales funciones que debe cumplir la empresa de acuerdo a la misión definida.

3.1.3. Necesidades de información

El conocer los procesos de negocio nos dará claridad en cuáles son las necesidades de información que se requieren en la empresa, lo cual también permitirá que se definan las fuentes de información necesarias para la construcción de una solución de inteligencia de negocios.

Las necesidades están relacionadas con los insumos que alimentan a los procesos de negocio para el logro de los objetivos planteados por la empresa.

3.1.4. KPI

Los indicadores de desempeño clave (KPI), permiten medir el estado en el que se encuentra la empresa, para determinar si se están logrando los objetivos definidos. Este indicador es esencial para la toma de decisiones y para poder hacer ajustes en las estrategias ejecutadas para estar en un proceso de mejora continua.

3.1.5. Lista de prioridades

Una vez que ya se cuenta con el conocimiento del negocio se procede a realizar una lista de prioridades de los indicadores de desempeño clave, para de esta manera tener claro el orden en el cual se trazara la solución de inteligencia de negocios, de tal manera que se pueda segmentar la solución con la finalidad de poder hacer un desarrollo ágil.

Planeación

En esta sección se definirán los aspectos relacionados con la planeación para el desarrollo de la solución de inteligencia de negocios. Esto tomando en cuenta algunos de los aspectos más relevantes que propone la metodología de administración de proyectos PMI.

Esta fase de la metodología es recomendada para aquellas empresas cuyo nivel de madurez es medio o alto, esto debido a que es a partir de estos niveles que es recomendable invertir tiempo en organizar los proyectos que se presenten. Cuando una empresa tiene un nivel de madurez bajo, tal vez no sea tan recomendable la inversión del tiempo en la planeación del desarrollo de la solución debido a su baja complejidad.

3.1.6. Alcance

Primeramente se definirá el alcance del proyecto, estableciendo que aspectos estarán incluidos en el desarrollo, tratando de delimitar de manera adecuada el proyecto para evitar problemáticas innecesarias. En este apartado se deberán incluir los entregables que se comprometerán.

3.1.7. Actividades

Una vez definidos los entregables se deben definir las actividades que se deben realizar para poder generar dichos entregables, estas actividades deberán incluir una cronología que permita conocer los tiempos que son requeridos para realizar cada actividad, tomando en cuenta que es recomendable establecer a lo más 5 días en las actividades de lo contrario una actividad con más de 5 días deberá ser segmentada. De la misma manera las actividades deberán tener el indicador de precedencia para conocer las dependencias que existan entre las actividades.

3.1.8. Recursos

Ya que están definidas las actividades se deben definir los recursos que necesitamos para la ejecución de dichas actividades. De esta manera debemos designar responsables para las actividades así como roles y responsables para gestión del proyecto. También se debe especificar que recursos materiales se requieren, así como los recursos técnicos necesarios.

Es importante definir el plan de comunicación en el cual se plasmen los roles y responsabilidades así como los medios de contacto. Así mismo establecer las fechas en las que se realizarán las juntas de avance y establecer el formato en el cual se hará dicha presentación.

3.1.9. Riesgos

Es necesario identificar los riesgos que puedan interferir con el cumplimiento de los tiempos establecidos para las actividades definidas, estos riesgos tienen que ser evaluados para determinar las probabilidades de que ocurran y el impacto que puede tener sobre el proyecto. Así mismo se debe asignar un responsable para cada riesgo así como establecer actividades que permitan mitigar los riesgos identificados.

3.1.10. Plan

Finalmente se plasma en un documento las actividades con sus tiempos, responsables, porcentajes de progreso, fechas compromiso, dependencias. Este documento deberá facilitar el dar seguimiento a los avances del proyecto, permitiendo actualizar de manera fácil el estado de las actividades.

Desarrollo

Ya que se cuenta con un plan para la ejecución de las actividades necesarias para desarrollar una solución de Inteligencia de negocios. Se precede con la ejecución de dichas actividades que en grandes rasgos se describirán en los siguientes apartados.

De acuerdo al nivel de madurez de la empresa, serán las tareas que se recomienda que se ejecuten. En el caso de un nivel de madurez bajo, solo se requiere del modelado de la información, siendo esta presentada en hojas de cálculo, con lo cual se puede jugar con la información a través de tablas dinámicas. En el caso de un nivel de madurez medio, es recomendable la dar un paso más después del modelado, es decir se recomienda la construcción del Data Warehouse, así mismo es recomendable implementar cubos que permitan un manejo de la información multidimensional, la presentación de estos resultados se recomienda que sean en web. En el caso de un nivel de madurez alto, se recomienda la implementación de minería de datos como un paso más delante de la creación de cubos, recomendándose la presentación en web.

3.1.11. Preparación Técnica

Se debe realizar una preparación de los ambientes para poder iniciar con la ejecución de las actividades designadas durante la planeación, pero para lograr esa preparación se debe realizar un análisis de las herramientas que sean más adecuadas para el desarrollo de la solución de Inteligencia de Negocios de acuerdo a los volúmenes de información y a la capacidad de compra que se tenga.

3.1.11.1. Selección Herramientas

Cuando se trata de desarrollos ambiciosos, con riesgos grandes y presupuestos holgados, se puede considerar el uso de herramientas propietarias por la solides que representan aunque representan altos costos, sin embargo para aquellos casos en los que se cuenta con un presupuesto muy reducido se puede considerar el utilizar herramientas open source. Cuando se recurre a herramientas open source es necesario hacer una buena selección pues no todas las herramientas disponibles son adecuadas (Gameiro, 2011).

Existen diversas herramientas en el mercado que permiten desarrollar soluciones de inteligencia de negocios. Para elegir las herramientas es necesario considerar los requerimientos para su instalación, las ventajas y desventajas que representa cada herramienta.

Se debe generar el documento que refleje las ventajas y desventajas de usar la herramienta seleccionada y debe contener las firmas aceptación que sirva de respaldo. Así mismo este documento deberá contener los requisitos de hardware y software para la instalación así como una guía para la instalación.

3.1.11.2. Instalación

Una vez seleccionada la herramienta se debe realizar la instalación de las herramientas tomando en consideración los requerimientos de software y hardware para que esta instalación sea exitosa.

En primera instancia se deben cubrir los requerimientos de hardware, por lo que se deben realizar las compras necesarias para cumplir con dichos requerimientos. Ya que se cuentan con los requerimientos de hardware se procede con la instalación de los pre-requisitos para que funcione adecuadamente la herramienta para finalmente realizar la instalación de la herramienta.

3.1.12. Modelado DWH

En este punto de la metodología busca que se logre generar un gráfico con 3 niveles diferentes de detalle, con la finalidad de generar las tablas con las restricciones y relaciones necesarias para construir el datawarehouse. Estos niveles nos permiten conocer los componentes de manera general y con un gran detalle los objetos que forman parte del datawarehouse.

3.1.12.1. Análisis de fuente de datos

Se realiza un análisis a profundidad de las fuentes de datos, tomando como referencia la información recolectada en la fase 0 Conocimiento del negocio. Esto es para determinar qué información de nuestras fuentes de datos es útil para la construcción de nuestro DataWarehouse.

3.1.12.2. HLDM

Una vez que se han revisado las fuentes de datos y se han documentado, se procede con el modelado del DataWarehouse. En primera instancia se realiza el modelado de datos de alto nivel, en el cual se vera de manera genérica las entidades que formarán parte del modelo y sus relaciones. Este modelado se basará en las conclusiones obtenidas durante el análisis de las fuentes de datos.

3.1.12.3. MLDM

Ya que se cuenta con el modelado de datos de alto nivel, se procede a detallarlo de tal manera que las entidades se amplíe su descripción a través de atributos, por lo que el modelo de datos de medio nivel, contendrá las entidades, sus atributos correspondientes y las relaciones entre dichas entidades.

3.1.12.4. LLDM

Como ultima parte del modelado de datos, se debe construir el modelo de datos de bajo nivel. En este modelo se representan los atributos de las entidades con todas sus características. Prácticamente este modelo de datos nos permitirá visualizar la estructura que tendrán las tablas requeridas para la construcción del DataWarehouse.

3.1.12.5. Construcción

Finalmente se generan los scripts con la estructura del DataWarehouse para construirlo con el manejador de base de datos seleccionado en la fase de preparación técnica, tomando en cuenta la sintaxis que utilice dicha herramienta.

Como documentación en este punto se deberá generar el diccionario de datos que permita tener el entendimiento del objetivo que se persigue con cada tabla y cuáles son sus características principales.

3.1.13. Datos

El objetivo de esta sección es trabajar con los datos analizándolos para diseñar y desarrollar el proceso ETL el cual terminara poblando el datawarehouse. Esta sección es una de las bases para los siguientes pasos de la metodología pues la salida de este paso será el datawarehouse poblado el cual es una entrada para el resto de los procesos.

3.1.13.1. Análisis de datos

En este punto se parte del análisis de la fuente de datos para conocer el origen de los datos y comenzar a analizar y planear las transformaciones que son necesarias para que se inserten en el DataWarehouse.

Se realizan los mapeos necesarios entre las fuentes de datos y las tablas finales contenidas en el DataWarehouse para tener una visión clara del destino que tendrá cada dato que será utilizado.

3.1.13.2. Desarrollo proceso ETL

Ya que se cuenta con el conocimiento del origen y destino de los datos se procede con el desarrollo del proceso ETL, el cual se recomienda tenga bien identificadas estas partes.

Se recomienda primeramente hacer una extracción de los datos sin mayores transformaciones e insertarla en una base de datos de paso (staging area). Ya que se tenga la copia fiel de los datos extraídos en la base de datos de paso, se procederá a realizar las transformaciones necesarias para lo cual nos podremos apoyar de tantas tablas temporales como consideremos necesario. Ya que se tienen los datos como los queremos, se introducen en el DWH.

Se debe tomar en cuenta que un ETL cuya finalidad sea la de poblar dimensiones deberá considerar la actualización de dichas dimensiones y la inserción de nuevos registros, esto es recomendable para tener siempre los datos disponibles aunque el proceso ETL falle. En el caso de poblar una tabla de hechos generalmente solo se realizan inserciones de los nuevos registros.

Se debe tomar en cuenta la creación de un proceso ETL para la población histórica de las tablas de hechos del DataWarehouse y otro proceso ETL para las cargas periódicas que se vayan a realizar.

Se deben documentar estos procesos para saber el camino que siguen los datos y las transformaciones que se le hacen a los datos.

3.1.13.3. Pruebas

Ya que se cuenta con los procesos ETL se realizan las pruebas necesarias para comprobar que funcionan adecuadamente. Una vez que se han probado y se valida que funcionan adecuadamente se considera que están listos para ser usados en un ambiente productivo.

3.1.13.4. Población histórica DWH y Calendarización ETL

Finalmente se realiza la carga de información en el DataWarehouse utilizando el ETL de carga histórica y se realiza un proceso que permita programar el proceso ETL para cargas periódicas para que sea ejecutado con la frecuencia que se considere necesario y que de esta manera este poblando el DataWarehouse con la información más reciente.

3.1.14. Minería

Cuando se tiene identificado algún problema que requiera la identificación de patrones de la información se puede hacer uso de la minería de datos para apoyar en la toma de decisiones de las empresas.

3.1.14.1. Selección de datos / Conformar Indicadores

Ya que se encuentra construido el DataWarehouse se selecciona el conjunto de datos que se utilizarán para ser analizados por los algoritmos de minería de datos.

Esta selección dependerá del problema que se busque atacar ya que la información que se utilice para estos fines deberá ser puntual para obtener los mejores resultados.

3.1.14.2. Modelado de datos / Establecer correspondencias

De acuerdo a los algoritmos de minería de datos seleccionados se generan los modelos de datos que serán el resultado de la aplicación de dichos algoritmos.

3.1.14.3. Evaluar resultados

Una vez obtenidos los modelos de datos es necesaria su revisión con el personal experto para determinar que modelos se acercan más a la realidad de la empresa, de manera tal que se puedan tomar la totalidad de los datos para generar los reportes necesarios para la toma de decisiones.

3.1.15. Cubo MOLAP

Para poder tener un análisis de los datos más potente y significativo es importante explotar el DataWarehouse con herramienta de análisis multidimensional, para lo cual se generan cubos de información que permiten tener los datos lineales del DataWarehouse en un esquema multidimensional lo que permite generar consultas más enriquecedoras para la toma de decisiones.

3.1.15.1. Dimensiones y Jerarquías

Dentro de un modelo multidimensional uno de los puntos importantes a desarrollar son las dimensiones y las jerarquías involucradas en el modelo.

Las dimensiones son los elementos cualitativos que permitirán evaluar la información. Son aquellos elementos que le dan sentido a los hechos. Las jerarquías es la manera en la que se puede navegar a través de la información.

Dentro del modelo multidimensional es importante definir de manera correcta las dimensiones y sus jerarquías para poder consultar los indicadores necesarios en distintos niveles de agregación.

3.1.15.2. Indicadores

Los Indicadores son las métricas utilizadas en el modelo multidimensional y son todos aquellos elementos que nos permiten hacer mediciones de manera cuantitativa.

Estos indicadores surgen a partir de las tablas de hechos del DataWarehouse y consiste en todos aquellos datos numéricos, estos son conocidos como indicadores base. Estos indicadores base permiten realizar los cálculos necesarios para generar los Indicadores principales de rendimiento, los cuales son claves para la toma de decisiones.

3.1.15.3. Cubos

Finalmente se genera la estructura que uniendo las dimensiones y los indicadores mediante un esquema de estrella, formará un cubo. Una vez formada esta estructura se podrán realizar las consultas multidimensionales que enriquecerán los reportes que se puedan generar para la toma de decisiones.

3.1.16. Presentación

El usuario final podrá explotar el DataWarehouse a partir de los reportes que se le presenten los cuales generalmente se presentan en formato web. Existen tres maneras de presentar el resultado del proceso de inteligencia de negocios al usuario final.

3.1.16.1. Reportes

En los reportes se incluye la información con cierto grado de detalle. Generalmente se presentan como tablas pivote en las cuales se pueden mezclar las dimensiones e indicadores que se lean de los cubos.

3.1.16.2. Dashboards

Es un resumen de los indicadores de rendimiento, mostrando sus resultados de manera visual mediante gráficas, tacómetros, entre otros. De manera que a primera vista se pueda observar el estado general de la empresa. Está más enfocado a medir el desempeño de los procesos. Generalmente un dashboard engloba la información por un tema específico.

3.1.16.3. ScoreCards

Contiene un resumen de la información de los indicadores de rendimiento clave, mostrando mediante semáforos su estado. Este tipo de reportes es generado con una temporalidad más amplia debido a que permite comparar períodos de tiempo, por ejemplo cuatrimestres. Está más enfocado en medir el estado del cumplimiento de la estrategia de la empresa. Debe incluir todos los aspectos que permitan conocer el estado global de la empresa.

Validación

Cuando el usuario final es capaz de ver la información mediante un reporte, dashboard o ScoreCard, entonces puede comenzar a validar si la información que visualiza es correcta. Este proceso es de suma importancia para saber si el proceso completo es correcto.

Una vez que el usuario indica que la información es correcta se procede con la implementación, en caso contrario se hacen los ajustes necesarios para hacer las correcciones necesarias.

Es recomendable hacer validaciones entre cada bloque del proceso de desarrollo para facilitar el proceso de validación y que este no sea tan pesado al final del desarrollo de la solución.

Implementación

Con el visto bueno del usuario, se puede continuar con la implementación de la solución en un ambiente productivo. Una vez implementado es recomendable monitorearlo y dar soporte a la solución hasta que esta funcione de manera estable.

Capítulo IV Caso de Estudio: Pizzería

Negocio

Domino's Pizza es una empresa internacional que surge en Michigan en 1960, cuenta con tiendas propias y franquicias. Este trabajo está enfocado en una franquicia.

La empresa "Domino's" es una PyME dedicada principalmente a la preparación y venta de Pizzas, actualmente utilizan un sistema POS para realizar el registro de su inventario así como de sus empleados y por supuesto las ventas realizadas.

Este sistema les genera reportes genéricos que no representan un gran apoyo para el gerente y su toma de decisiones, puesto que no están diseñados para cubrir las necesidades de información requeridas para el cumplimiento de los objetivos de la empresa.

Así mismo a pesar de que el sistema les permite tener información sobre el desempeño de sus empleados, no explota esta información debido al no tan conveniente diseño de los reportes actuales, los cuales además son en formato texto lo que dificulta más su utilización.

Debido a esto la empresa tampoco cuenta con un sistema que les proporcione real ventaja competitiva como lo podría ser la implementación de la Minería de Datos y en general a una solución integral de BI, al ser sus reportes tan planos y a la falta de procesos que le permitan explotar de una mejor manera su información, pierde cierta competitividad.

Como podemos observar actualmente la empresa "Domino's" funge como una empresa tradicional la cual no utiliza todas las tecnologías de información que se encuentran hoy disponibles en el mercado y que están al alcance de cualquier empresa.

El nivel de madurez con el que cuenta esta empresa es medio, por lo que las implementaciones que se presentan van de acuerdo a este nivel de madurez. Se hace mención de opciones de implementación para los otros niveles de madurez en el ANEXO I y ANEXO II.

4.1.1. Planeación estratégica

4.1.1.1. Misión

“Gente Excepcional sirviendo la mejor pizza del mundo”. Es la misión presentada en la tienda, sin embargo no expresa en su totalidad los productos y servicios que ofrece la empresa, por lo que se trabajo una nueva misión interna que permitiera conocer un poco mejor a la empresa.

“Somos una empresa dedicada a la preparación y venta de productos de comida rápida como son pizzas, botanas y postres así como bebidas, contando con servicio de restaurant o de entrega a domicilio”.

4.1.1.2. Visión

La visión con la que trabaja la empresa es “Ser el mejor operador del sistema Domino’s Pizza con el mejor talento humano”.

4.1.1.3. Objetivos

De acuerdo a la misión y visión de la empresa se definieron los siguientes objetivos:

- Incrementar las ventas.
- Incrementar la productividad de los empleados.
- Mejorar el servicio al cliente.

4.1.1.4. Metas

Siguiendo los objetivos planteados, se definieron las siguientes metas, que nos permitirán lograr los objetivos establecidos.

- Incrementar las ventas en un 20%
- Incrementar la productividad de los empleados en un 30%. Atendiendo mejor y a más comensales.
- Mejorar el servicio al cliente, aumentando en un 20% la recepción de comensales.

4.1.1.5. Factores Críticos de éxito

Para que la empresa considere que las acciones que planea ejecutar en el futuro próximo podemos decir que será de acuerdo a los resultados que tenga a partir de la toma de decisiones que puede realizar.

Por lo tanto uno de los principales Criterios será que el gerente de la tienda sea capaz de tomar decisiones a partir de los reportes que se generen con el desarrollo de este proyecto.

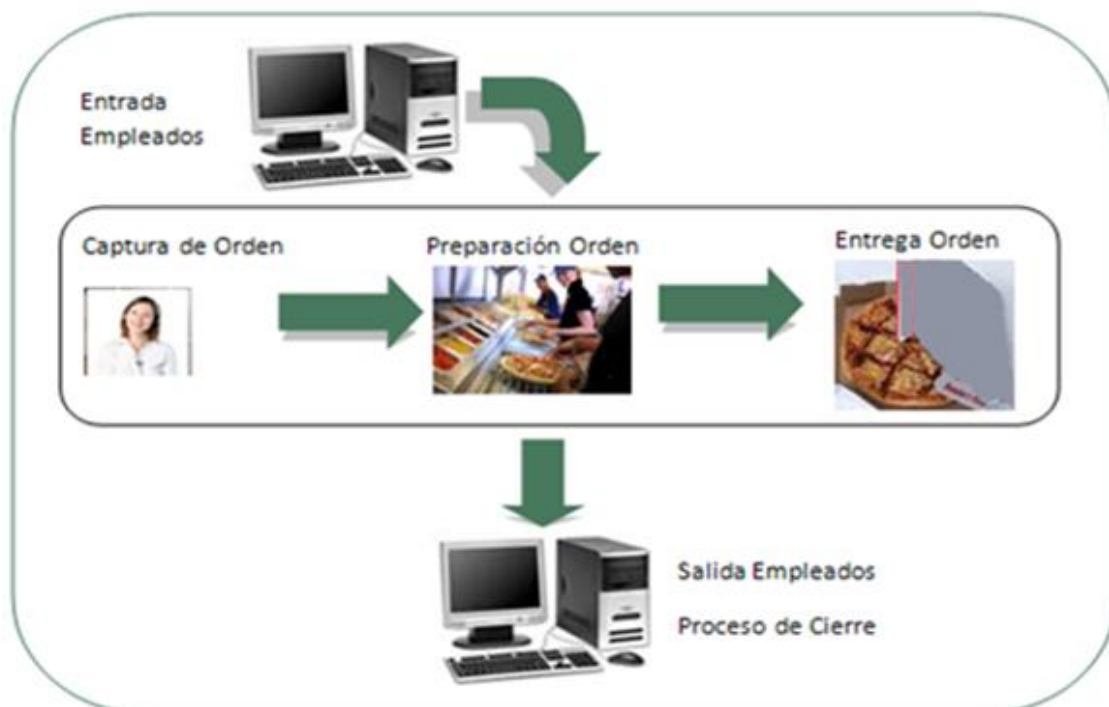
Otro criterio será el cumplimiento de los objetivos de acuerdo a la información generada por las necesidades de información identificadas a partir de los objetivos definidos por la empresa.

Los factores críticos de éxito son aquellos que forzosamente deben o no presentarse para el cumplimiento de las metas y por ende para el logro de los objetivos definidos.

4.1.2. Procesos de negocio

Al ser una franquicia, siguen procesos y procedimientos ya pre-establecidos. La empresa “X” sigue el siguiente proceso:

Se registran los empleados a su llegada, cuando llega un cliente registran la orden la cual es pasada a la línea de producción donde se prepara la pizza y finalmente esta es entregada. Cuando ha finalizado el turno del empleado este debe registrar su salida, así mismo al final del día se hace un proceso de cierre con el cual se actualizan los inventarios y se generan los reportes en texto plano.



Fuente: Diseño propio

IV-1. Proceso

Para el cumplimiento de los objetivos establecidos, es necesario fijarse principalmente en dos procesos de negocio: las ventas y la producción. Las ventas se realizan principalmente en dos modalidades “a domicilio” y “restaurante”, cuando las ventas son a domicilio se tiene la ventaja de contar con los datos del cliente lo que permite realizar un análisis sobre los hábitos de compra de los clientes para establecer una estrategia que permita que los clientes que hayan dejado de comprar vuelvan a comprar.

En el proceso de producción podremos analizar los tiempos de cada empleado para medir su productividad y establecer estrategias que permitan tomar las decisiones correspondientes que permita que los empleados sean más productivos.

4.1.3. Necesidades de información

Para poder tener las bases necesarias para la toma de decisiones será necesario contar con cierta información que nos sirva de guía para la adecuada toma de decisiones.

Para esto es necesario contar con la información proveniente principalmente del sistema de venta que se utiliza actualmente que nos proporcionara la información de las ventas y la información relacionada al desempeño de los empleados. También se podrá extraer la información de los clientes para poder dar un seguimiento a estos y poder mejorar el servicio al cliente.

4.1.4. KPI

Los principales indicadores de rendimientos que se definieron para evaluar el desempeño de la empresa son:

- Ventas
- Productividad
- Servicio al Cliente

4.1.5. Lista de prioridades

Se definieron tres objetivos, por lo que es necesario fijar prioridades para poder cubrirlos progresivamente de tal manera de que se logre se cumplan en tiempo y forma.

Se establecieron las prioridades de la siguiente manera:

- Se desarrollaran los informes necesarios para monitorear las ventas.

- Se deberá trabajar en el desarrollo de los informes necesarios para la toma de decisiones sobre el desempeño de los empleados.
- Se desarrollaran los informes necesarios para el seguimiento de los clientes.

Planeación

4.1.6. Alcance

El desarrollo de este proyecto contempla el análisis, diseño y generación de un prototipo de solución de BI, tomando en cuenta lo siguiente:

- Análisis, diseño y construcción de un DWH.
- Análisis, diseño y construcción del proceso ETL
- Análisis, diseño y construcción de un reporte estratégico.

4.1.7. Actividades

Actividades	Descripción
Análisis	Se realizará un análisis de las fuentes de datos y las necesidades del negocio.
Diseño	Se realiza el diseño del DWH y del proceso ETL, así como el reporte a generar.
Desarrollo	Se creara el DWH y su proceso de ETL y se creará el reporte.
Pruebas	Se harán las pruebas necesarias para verificar que el desarrollo funciona de manera adecuada.

Tabla 8. Actividades generales

4.1.8. Recursos

Actualmente para el desarrollo de este proyecto se cuenta con ciertos recursos, los cuales se detallan en la Tabla 9. Recursos.

Concepto	Detalle
Personal:	Pamela Gutiérrez Meléndez Consultor BI
Fuentes de Datos:	POS
Hardware:	Procesador: Intel Core Duo 2.4ghz Memoria RAM: 2.75 GB Disco Duro: 80 GB

Software:	Jaspersoft SpagoBI OpenI Microstrategy
-----------	---

Tabla 9. Recursos

4.1.9. Restricciones y Supuestos

Para poder llevar a cabo el proyecto que deberá estar alineado a los objetivos del negocio se han establecido los siguientes requerimientos:

- Software. Se debe contar con el software necesario para la implementación de la solución de BI.
- Hardware. Se debe contar con el hardware que cubra con los requerimientos mínimos de las herramientas seleccionadas.
- Accesos. Se deben proporcionar los accesos necesarios a las fuentes de información.
- Comunicación. Se debe mantener una estrecha comunicación entre los expertos del negocio y los implementadores de BI

Para cumplir en los objetivos establecidos por el negocio en tiempo y forma, se definieron los siguientes supuestos:

- Se cuenta con la infraestructura idónea para poner el proyecto en producción
 - Hardware
 - Software
 - Red
- Se cuenta con el apoyo del personal de la empresa para resolver cuestiones de definición de datos.
- Se cuenta con la información necesaria para la conexión a las fuentes de datos.

Así mismo se definieron las siguientes restricciones, que permiten acotar el alcance del proyecto, estas restricciones son:

- Se utilizaran las herramientas de desarrollo seleccionadas.
- Se desarrollara la solución de acuerdo a las prioridades definidas por el negocio.
- Se creara el repositorio de datos según las definiciones iniciales.
- Se hará la limpieza de datos según lo establecido por el repositorio de datos.
- Se seguirá la metodología propuesta.

4.1.10. Riesgos

Se han identificado ciertos riesgos para los cuales se han definido ciertas acciones que mitiguen o en su defecto minimicen las probabilidades de que sucedan. A continuación se muestra un listado de dichos riesgos:

Riesgo	Probabilidad	Acción	Responsable
Falta de tiempo de los colaboradores	Media	Concertar citas con colaboradores con anticipación	Negocio, Consultor
Fallas en conexiones de red con fuentes de datos	Media	Comunicar fallas al departamento de redes para dar pronta solución	Negocio (Proveedor Red)
Fallas Equipo Hardware de desarrollo	Baja	Se deberán hacer respaldos semanales de avances y se deberá contar	Negocio (Proveedor HW), Consultor
Fallas con instalaciones y configuraciones de SW	Media	Se deberá contactar con los proveedores correspondientes y realizar la investigación necesaria.	Consultor

4.1.11. Plan

De acuerdo a las actividades generales definidas se establece el siguiente calendario dividido en semanas para la ejecución de dichas actividades:

Actividades	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16
Análisis																
Diseño																
Desarrollo																
Pruebas																

Desarrollo

4.1.12. Preparación Técnica

4.1.12.1. Selección Herramientas

El desarrollo de esta solución se hará sobre el Sistema operativo Windows Server 2003, con el cual ya cuenta la franquicia debido a que al adquirir dichas franquicias les proporcionan equipos con este sistema operativo por lo que la franquicia no realizara una inversión adicional.

Como motor de base de datos se trabajara con SQL Server 2008 edición express que es gratuita por lo que la empresa no tendrá que realizar un desembolso económico con esta herramienta, además de que los volúmenes de información que maneja hacen que esta herramienta sea adecuada.

Debido al alcance de este trabajo se realizará la instalación de cuatro herramientas para el desarrollo de la solución con la finalidad de obtener una visión de que herramienta es la más funcional para una empresa PyME.

4.1.12.1.1. Jaspersoft

La suite de BI Jaspersoft debido a que es de licencia abierta y es lo suficientemente completa para el desarrollo de esta solución, Se debe instalar principalmente 3 herramientas de la suite:

- Jasper Server
- Jasper iReports
- Jasper ETL

4.1.12.1.2. SpagoBI

Herramienta Open Source que ofrece toda una suite de BI, entre los componentes que incluye son:

- SpagoBI Server, es el núcleo que ofrece funciones analíticas.
- SpagoBI Studio, ambiente de desarrollo.
- SpagoBI Meta, ambiente de desarrollo de metadata.
- SpagoBI SDK, librerías de desarrollo basadas en Java.
- SpagoBI Applications, modelos analíticos desarrollados para trabajarse en SpagoBI Studio.



Fuente: Diseño propio

IV-2. Componentes SpagoBI

Existen disponibles para su descarga las herramientas Server, Meta y Studio. Las cuales pueden trabajar para sistemas Linux y Windows. En el caso de Windows contiene archivos ejecutables que realizan el inicio del Servidor y abre la herramienta Studio donde se puede genera el modelo de metadata.

4.1.12.1.3. OpenI

Al inicio del desarrollo de este trabajo OpenI se conocía como una herramienta que ofrecía una suite completa para el desarrollo de soluciones de BI, sin embargo en la actualidad ya solo se conserva un plug-in que funciona junto con la herramienta Pentaho. Es por eso que no fue posible realizar las pruebas necesarias con dicha herramienta.

4.1.12.1.4. Microstrategy

Es una herramienta propietaria para el desarrollo de soluciones BI, según el cuadrante de Gartner se encuentra como una de las herramientas líderes de BI. Es una herramienta costosa, sin embargo no deja de ser una opción para una PyME esto debido a que ofrece una suite de reporte gratuita que permite que las Pymes tengan acceso a dicha herramienta. El adquirir esta herramienta da la opción de que en el momento de que la PyME evolucione al nivel de convertirse en una empresa grande ya no será necesario hacer un cambio drástico de herramientas de BI. Esta herramienta ofrece un servidor y un ambiente de desarrollo.

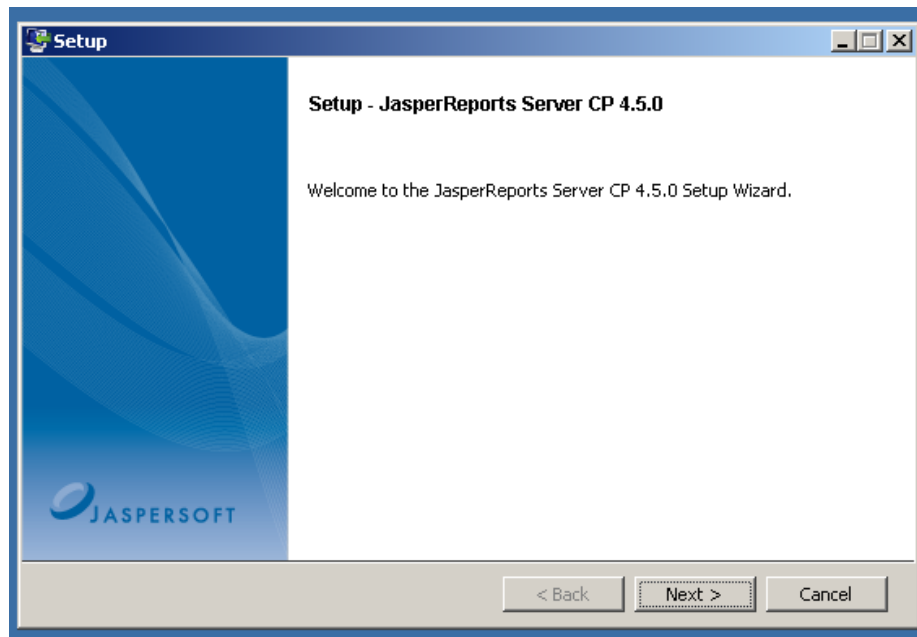
4.1.12.2. Instalación

4.1.12.2.1. Jaspersoft

Para la instalacion de jaspersoft Server es necesario la descarga del instalador de la siguiente pagina:

<http://jasperforge.org/projects/jasperserver>

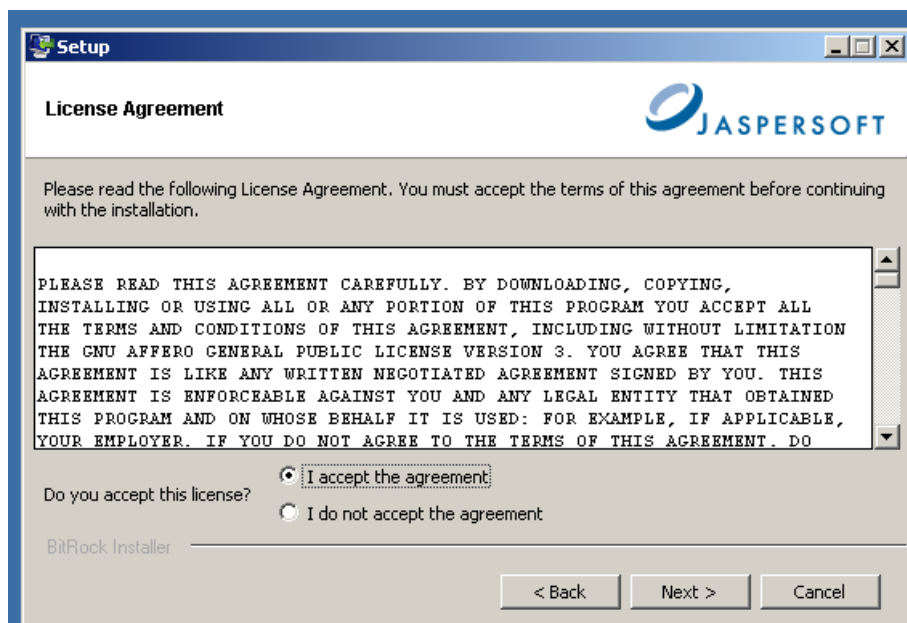
Se ejecuta el instalador y nos mostrara la pantalla de bienvenida, a la cual se le da siguiente para pasar a la siguiente pantalla.



Fuente: Instalación herramienta

IV-3. Jaspersoft - Pantalla bienvenida

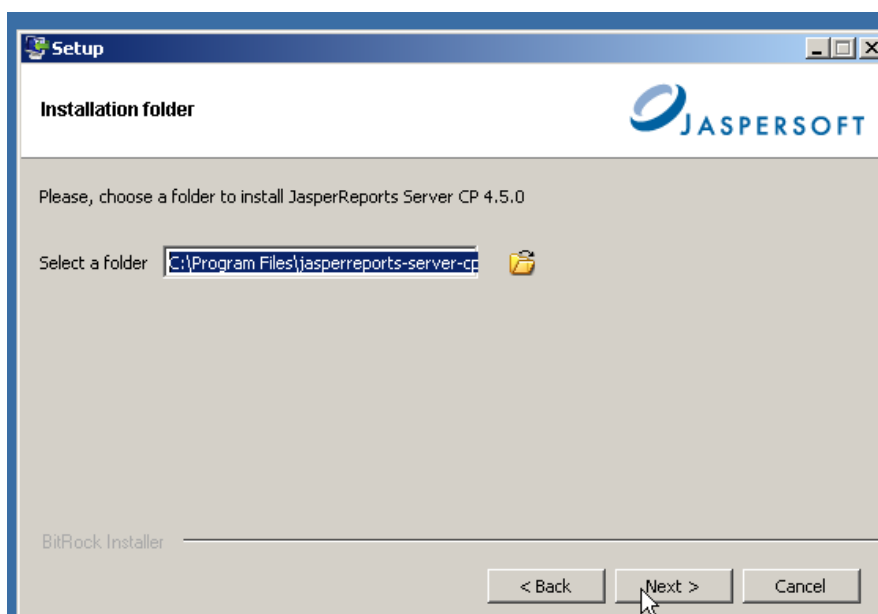
En la siguiente pantalla veremos el acuerdo de licencia que para continuar se debe de aceptar. Se presiona siguiente para pasar a la siguiente pantalla.



Fuente: Instalación herramienta

IV-4. Jaspersoft - Acuerdo de licencia

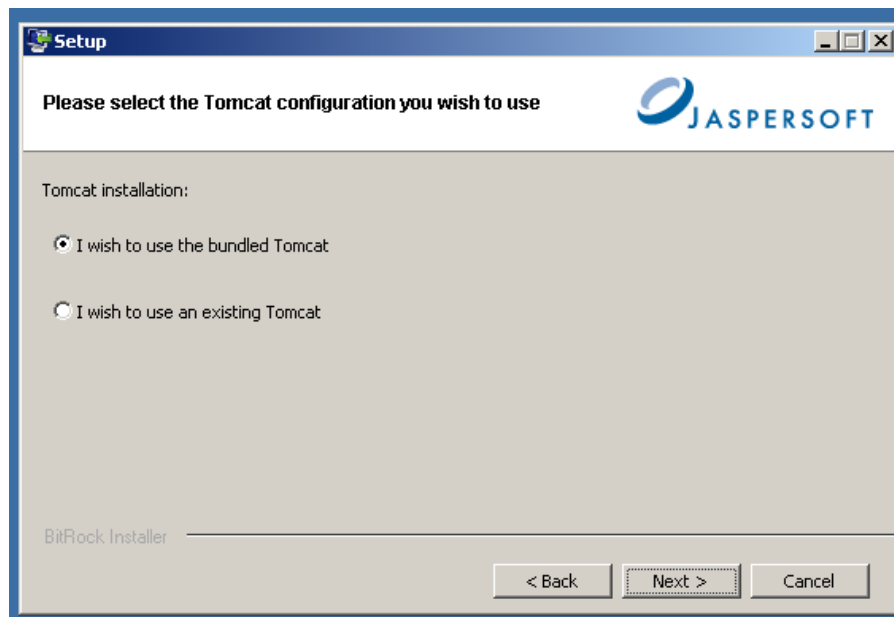
En la siguiente pantalla observaremos la ruta en la que se hará la instalación, una vez definida esta ruta daremos click en el botón de siguiente.



Fuente: Instalación herramienta

IV-5. Jaspersoft - Ruta de instalación

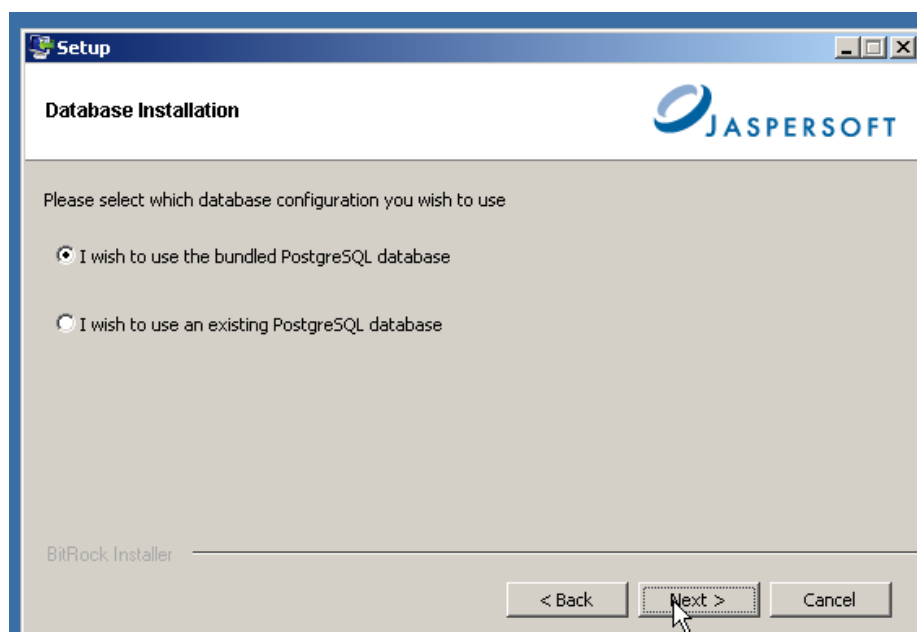
En la pantalla siguiente podremos elegir si queremos que se instale Tomcat o si ya existe una instalación de Tomcat. Seleccionamos la instalación incluida seleccionamos siguiente.



Fuente: Instalación herramienta

IV-6. Jaspersoft - Instalación TOMCAT

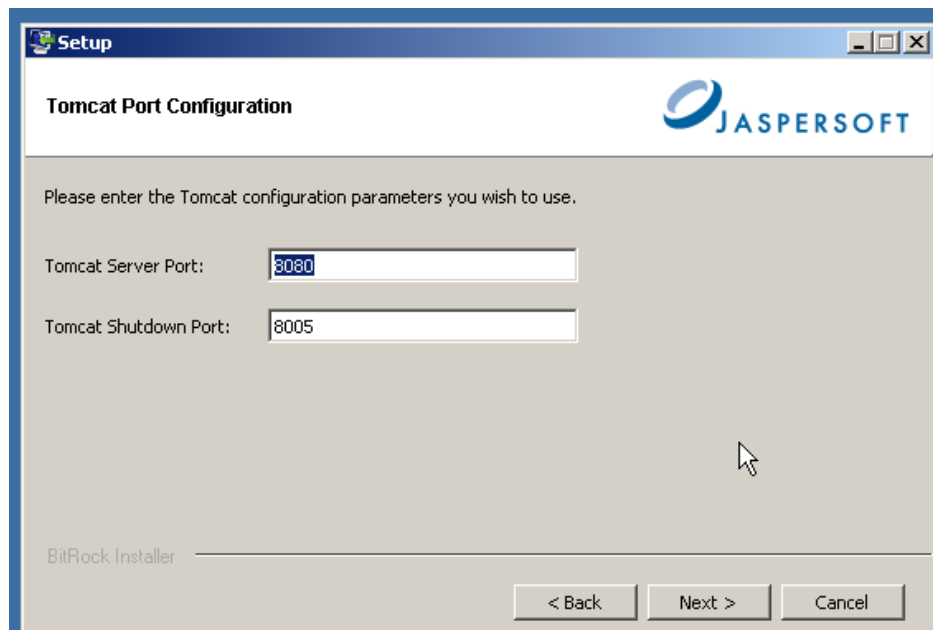
Seleccionamos si queremos que se instale el motor de base de datos PostgreSQL o trabajar con una instalación existente. Damos click en siguiente.



Fuente: Instalación herramienta

IV-7. Jaspersoft - Base de datos para metadatos

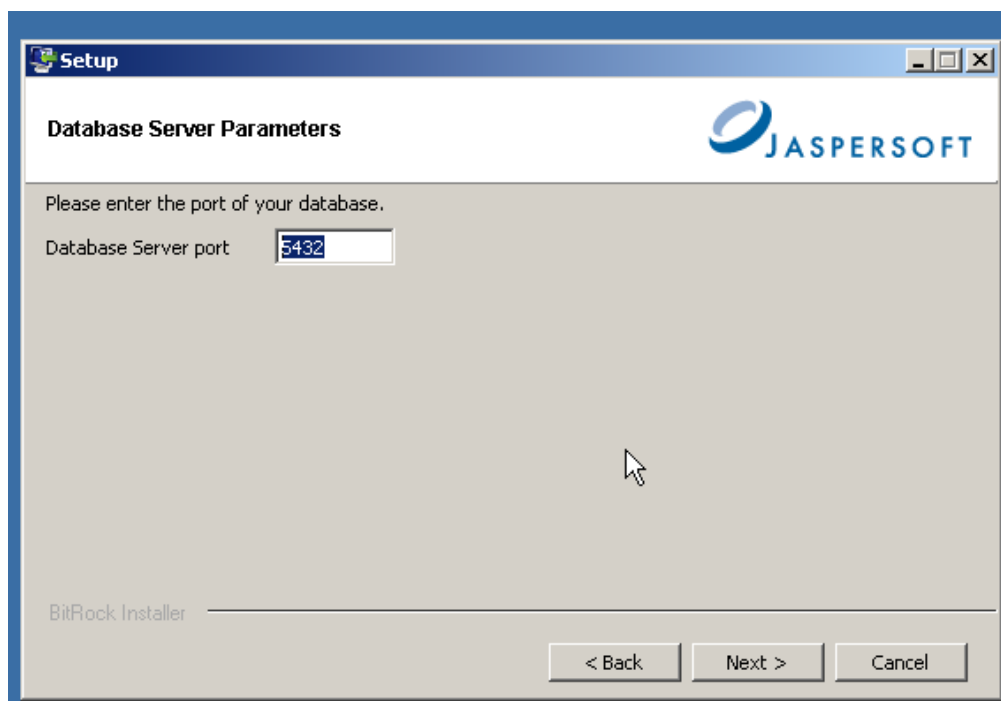
Se configura los puertos para el servidor Tomcat.



Fuente: Instalación herramienta

IV-8. Jaspersoft - Configuración TOMCAT

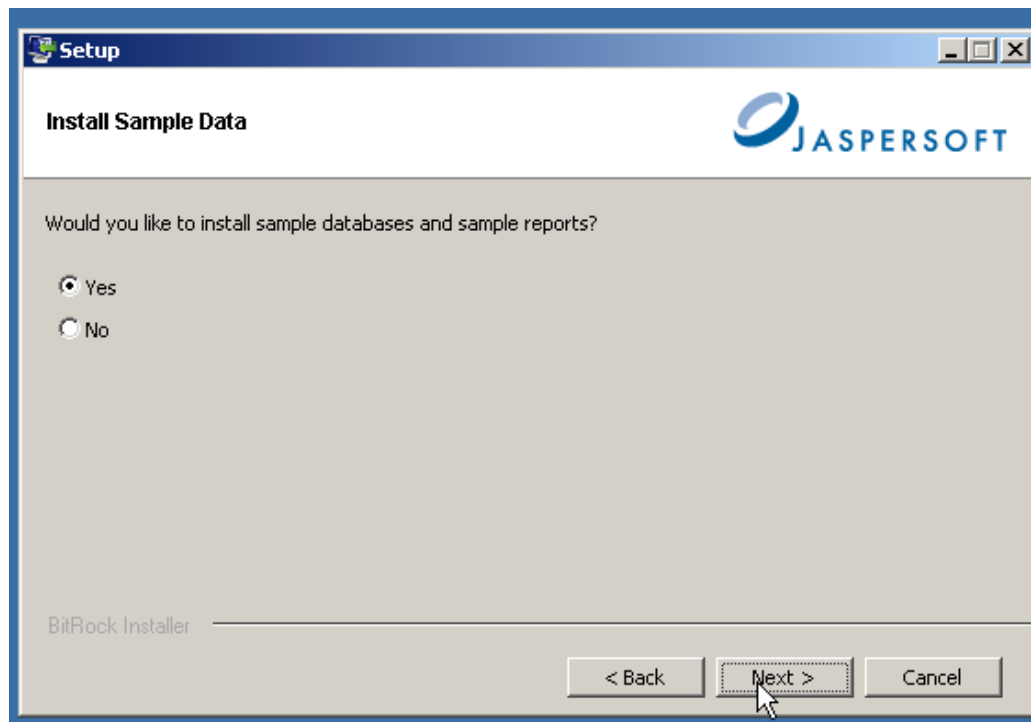
Se configura el puerto del servidor de base de datos.



Fuente: Instalación herramienta

IV-9. Jaspersoft - Configuración base de datos

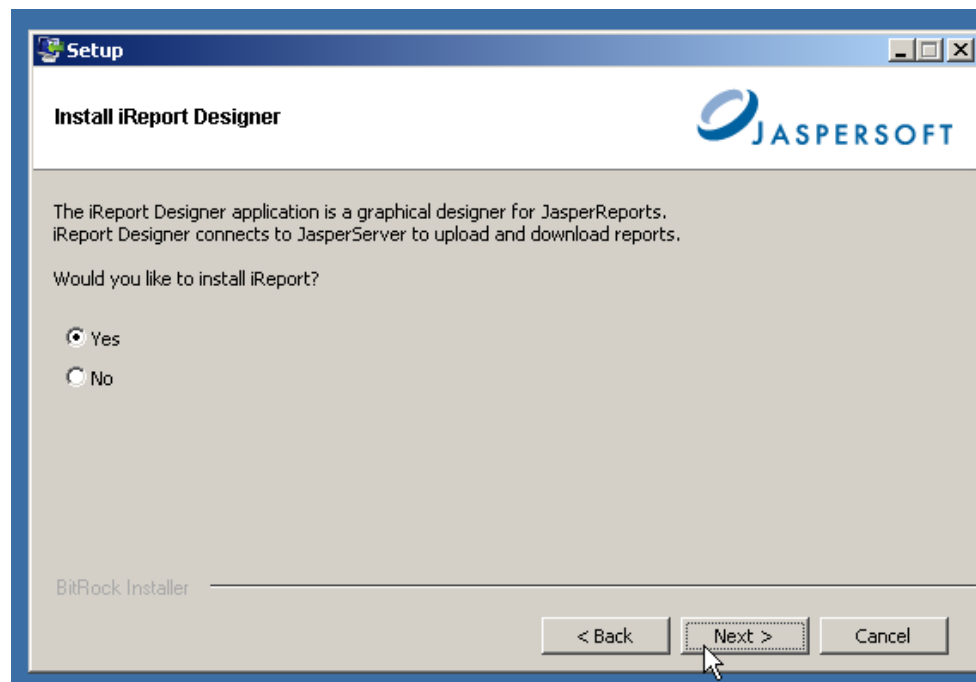
Seleccionamos si queremos instalar datos de ejemplo



Fuente: Instalación herramienta

IV-10. Datos ejemplo

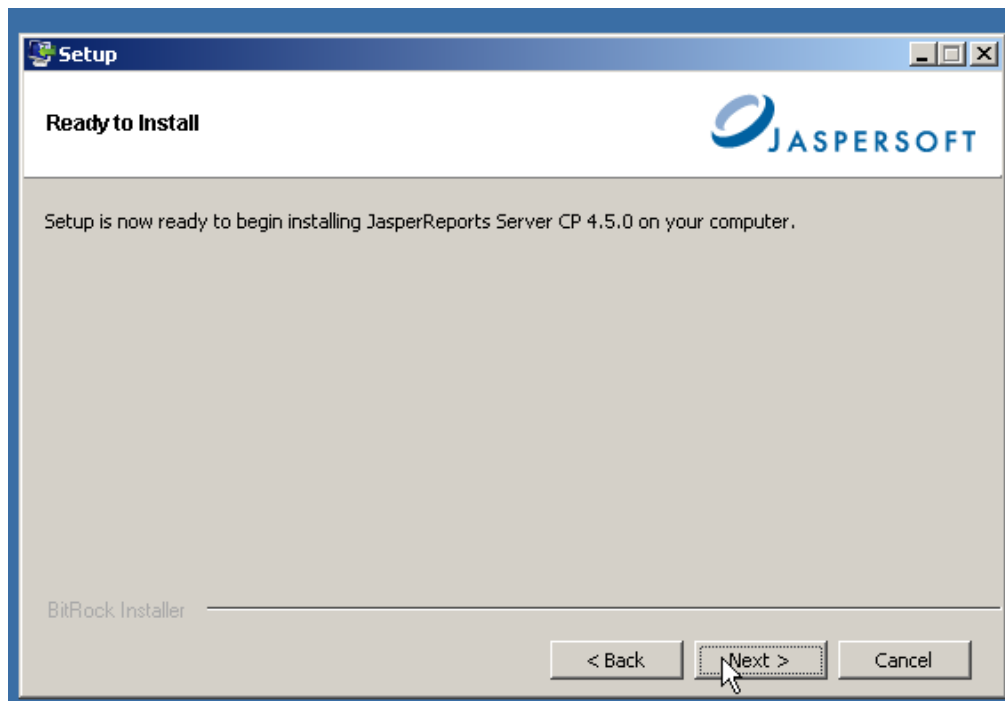
Instalaremos la herramienta iReport.



Fuente: Instalación herramienta

IV-11. iReport

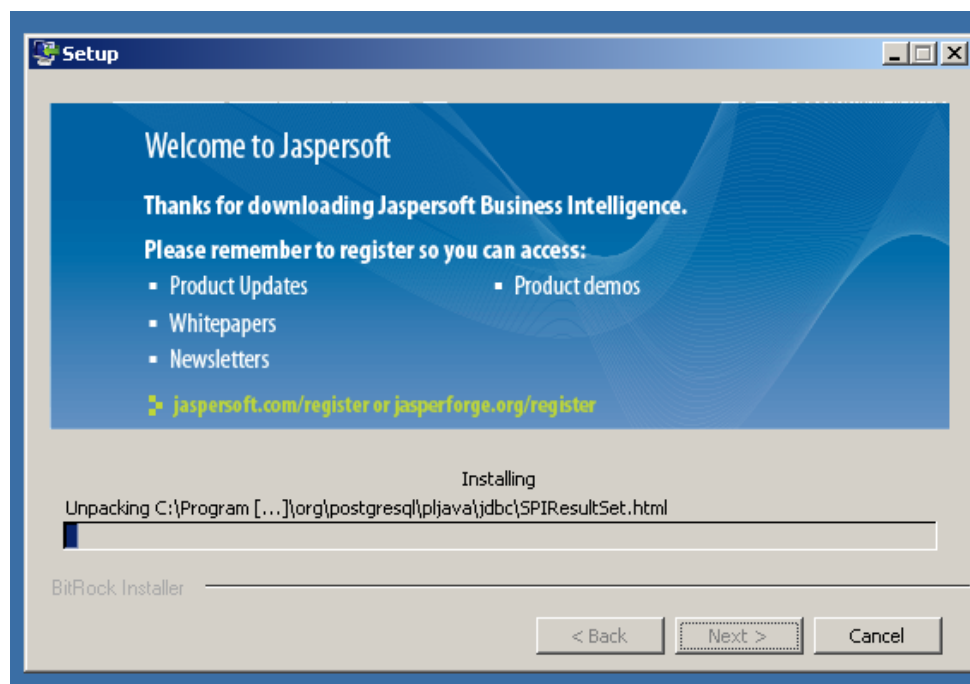
Se confirma que está todo listo para la instalación



Fuente: Instalación herramienta

IV-12. Confirmación de instalación

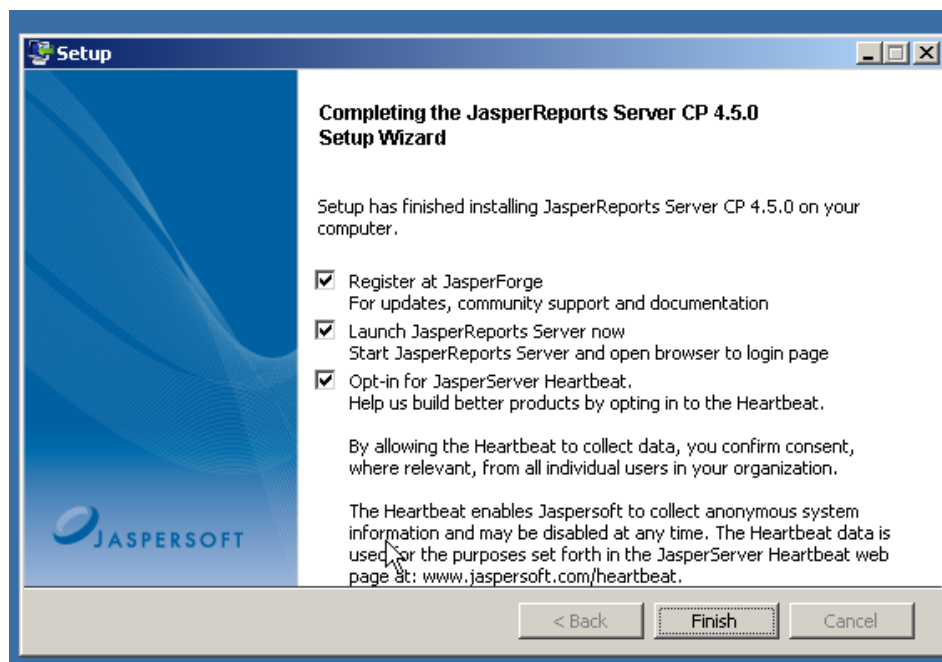
Comienza la instalación y la barra de progreso nos ira indicando el avance de dicha instalación.



Fuente: Instalación herramienta

IV-13. Progreso instalación

Una vez completada la instalación nos dará la opción para registrar el producto, iniciar el servidor y ayudar para mejorar la herramienta.



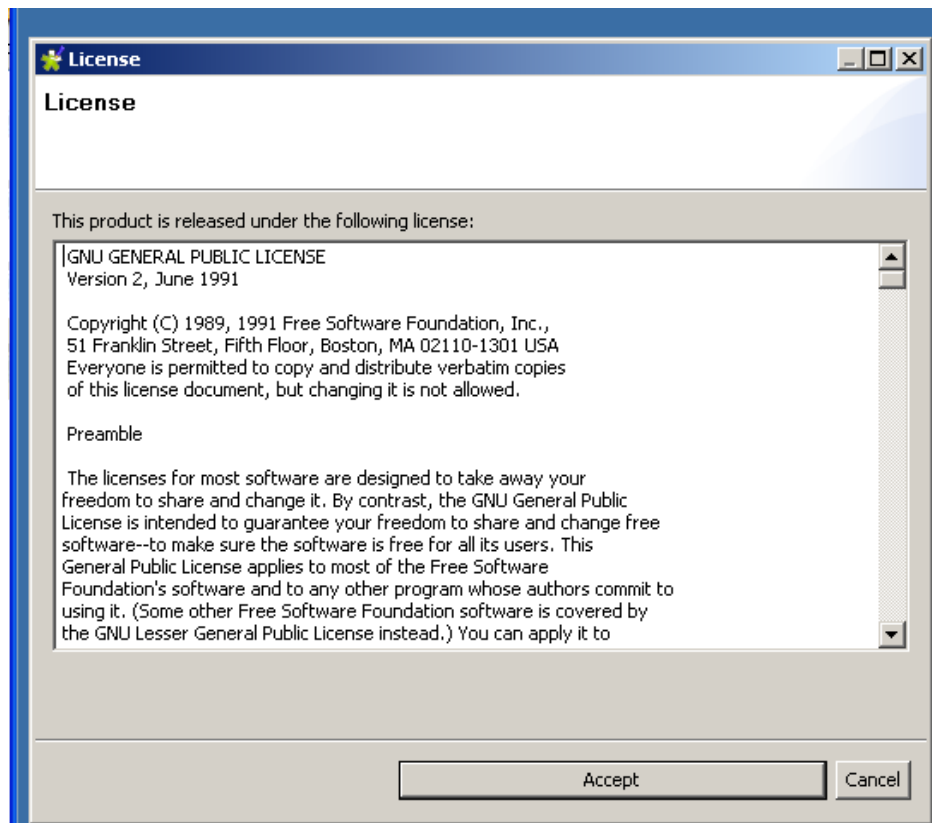
Fuente: Instalación herramienta

IV-14. Instalación completada

Con esto finaliza la instalación del servidor web de servicios de Jaspersoft. Para instalar JasperETL se debe descargar el instalador de la siguiente página:

<http://jasperforge.org/projects/jasperetl>

Esta herramienta no hay necesidad de instalarla, ya que solo basta con descomprimir el archivo comprimido que se descarga y utilizar el ejecutable que viene incluido en los archivos. Es necesario que aceptemos los términos de la licencia de la herramienta.



Fuente: Instalación herramienta

IV-15. JasperETL - Acuerdo de licencia

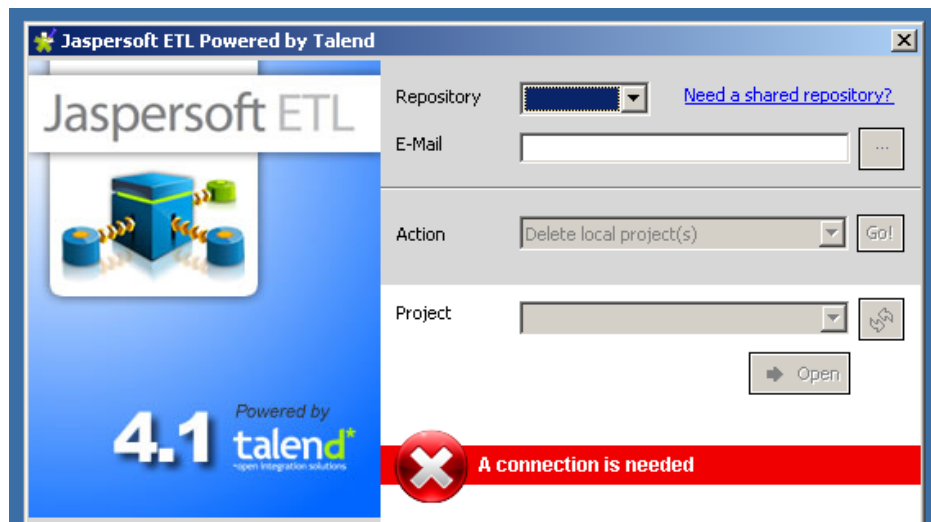
Nos dará la opción de unirnos a la comunidad Talend, que son los que desarrollan toda la documentación relacionada a las herramientas de Jaspersoft.



Fuente: Instalación herramienta

IV-16. JasperETL - Comunidad Talend

Finalmente accederemos a la pantalla de configuración de Jaspersoft ETL



Fuente: Instalación herramienta

IV-17. JasperETL - Configuración

4.1.12.2.2. SpagoBI

Para poder trabajar con SpagoBI es necesario descargar principalmente dos de sus herramientas principales:

1. SpagoBI Server
(http://forge.ow2.org/project/download.php?group_id=204&file_id=18088)
2. SpagoBI Studio
(http://forge.ow2.org/project/download.php?group_id=204&file_id=18084)
3. SpagoBI Meta
(http://forge.ow2.org/project/download.php?group_id=204&file_id=18082)

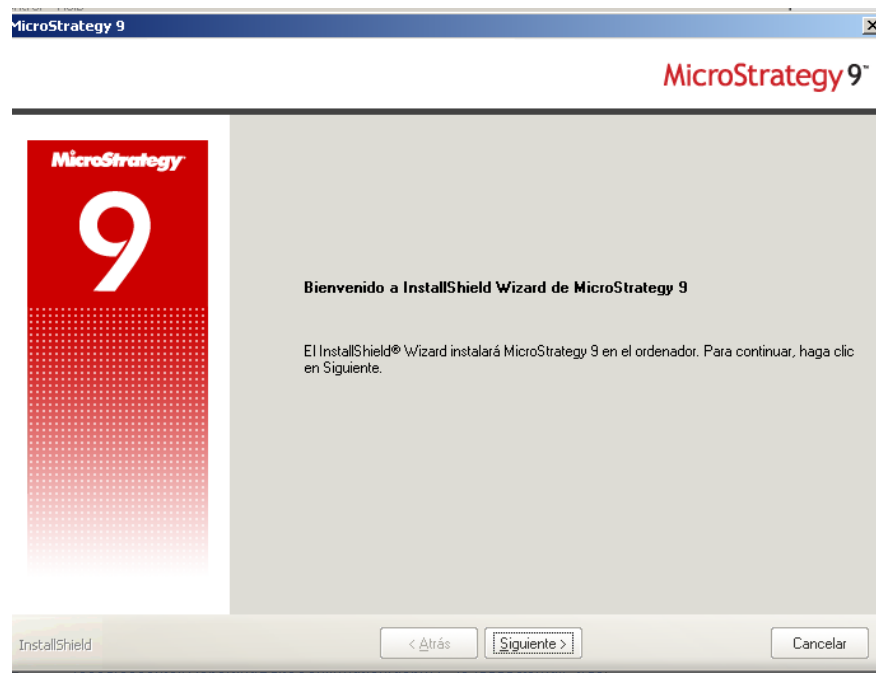
Al realizar la descarga de estas aplicaciones no habrá necesidad de hacer mayores instalaciones puesto lo que se descarga son ejecutables, que abrirán las herramientas en cuestión para su uso.

4.1.12.2.3. Microstrategy

Microstrategy ofrece de manera gratuita una suite de reporte (<http://www.microstrategy.com.mx/softwaredereportinggratuito/>), para tener acceso a ella es necesario acceder con una cuenta de correo corporativa por lo que la empresa PyME en cuestión es necesario que cuente con un dominio propio para poder registrar su correo de manera exitosa y poder contar con la suite de reporte.

Una vez descargada es necesario descomprimir el archivo en formato .zip y ejecutar el archivo .exe contenido en la carpeta donde se descomprimió el archivo descargado.

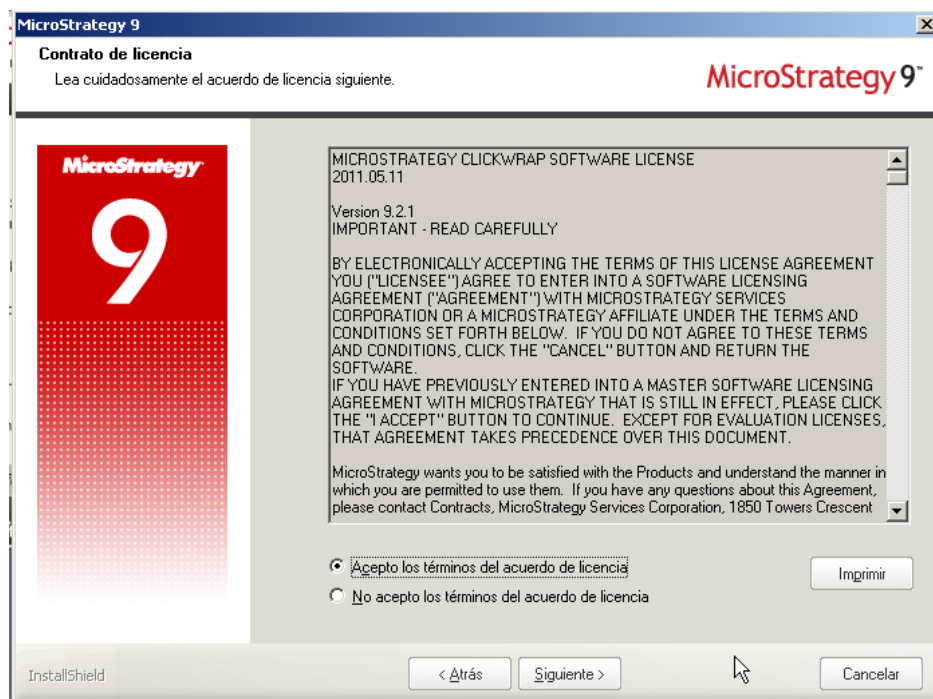
En primera instancia observaremos la pantalla de bienvenida. Se da selecciona el botón “siguiente”.



Fuente: Instalación herramienta

IV-18. Microstrategy - Pantalla bienvenida

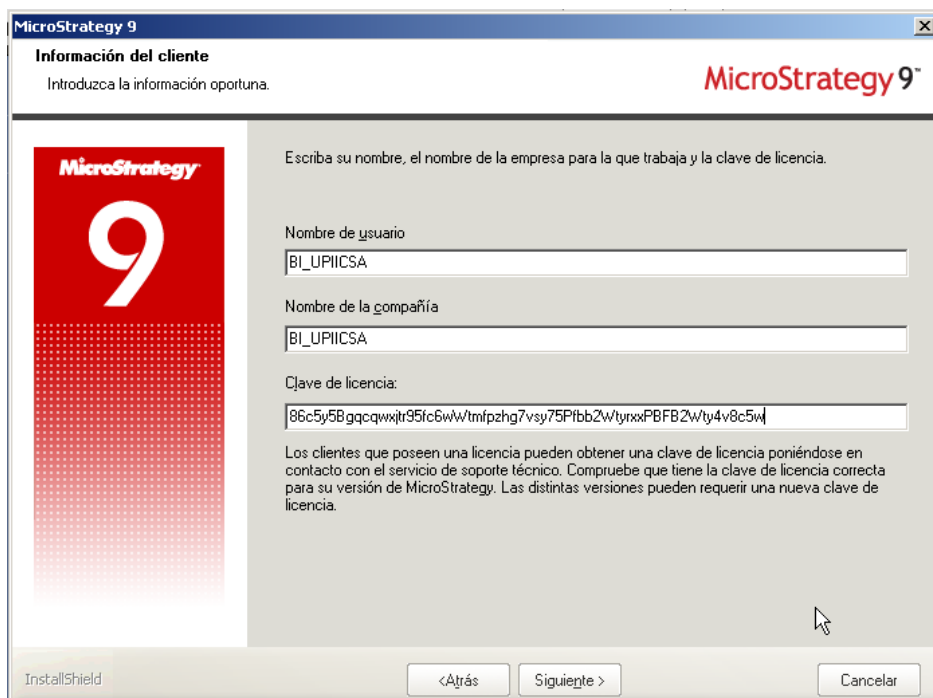
Nos mostrará la licencia del software, la cual tendremos que aceptar para poder continuar con la instalación.



Fuente: Instalación herramienta

IV-19. Microstrategy - Acuerdo de licencia

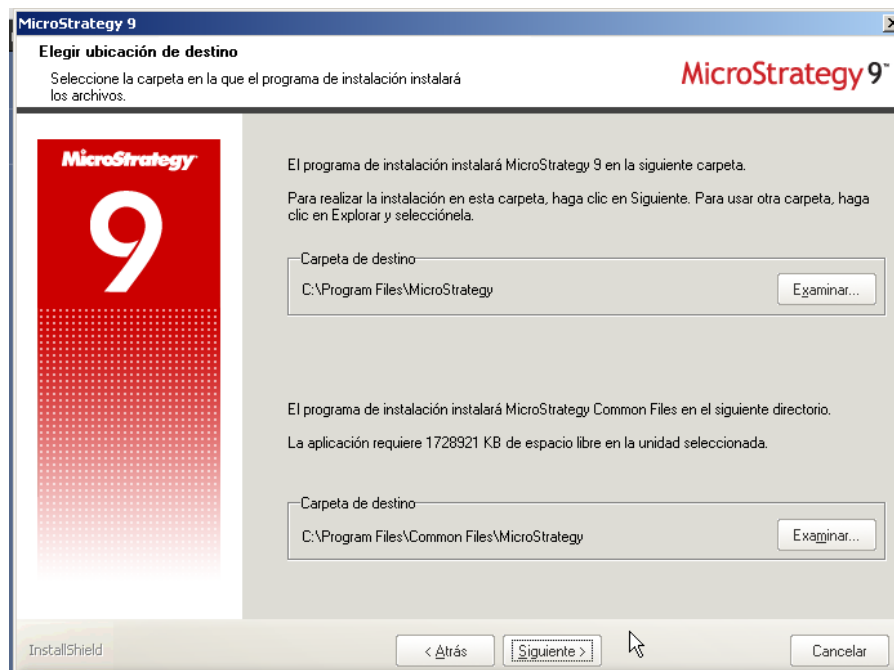
Se introduce la información del cliente que se usará para la configuración del servidor y en donde se debe introducir la clave de licencia proporcionada.



Fuente: Instalación herramienta

IV-20. Microstrategy - Clave de licencia

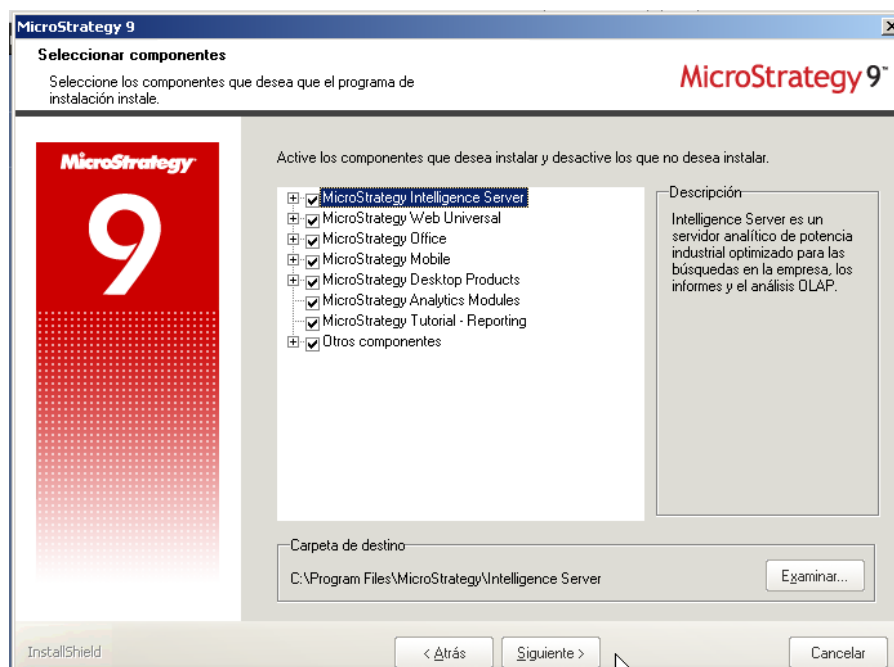
Se elige la ruta en la cual se almacenarán los archivos que se instalarán.



Fuente: Instalación herramienta

IV-21. Microstrategy - Ruta de instalación

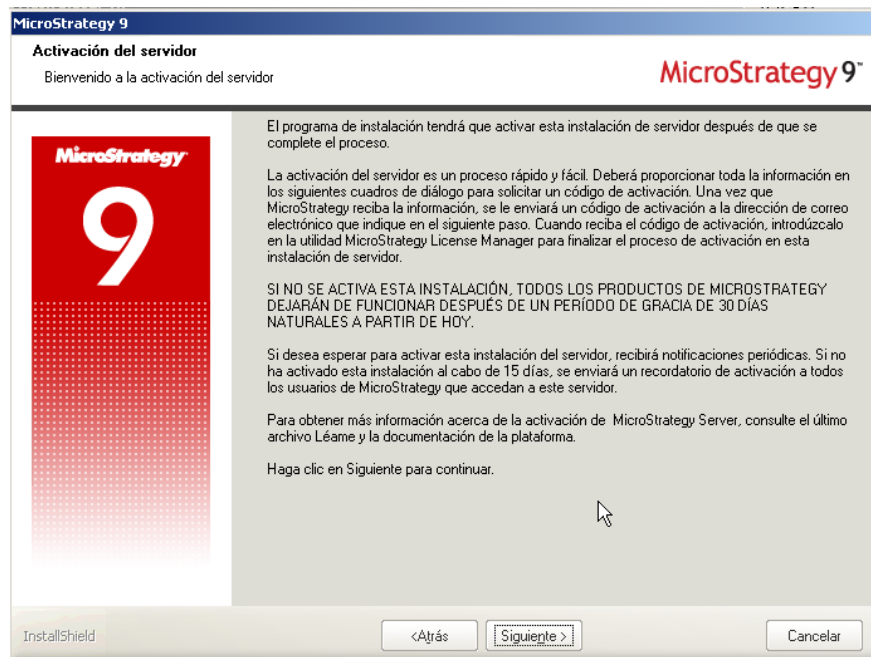
Se seleccionan los componentes a instalar.



Fuente: Instalación herramienta

IV-22. Microstrategy - Selección de componentes

Se nos indica que una vez finalizada la instalación es necesario que la herramienta sea activada de lo contrario una vez pasados 30 días naturales la herramienta se bloqueará.



Fuente: Instalación herramienta

IV-23. Microstrategy - Mensaje informativo de activación

Se selecciona el modo en el que será utilizada la herramienta, en este caso se utilizó para pruebas sin embargo existen otros escenarios bajo los cuales se puede utilizar.

Fuente: Instalación herramienta

IV-24. Microstrategy - Uso del software

Se introduce la información necesaria para la activación del producto.

MicroStrategy 9

Activación del servidor

Introduzca la información del contacto del cliente

MicroStrategy 9™

Ha indicado que está instalando el software en nombre de un cliente de MicroStrategy. Debería haber introducido la información del contacto en el paso anterior.

Además, también hay que introducir la información de contacto de un empleado de la empresa autorizada. Introduzca esta información a continuación.

Información del usuario Se requieren todos los campos

Nombre de la Compañía: Departamento:

Nombre: Apellido:

Título: Teléfono:

Correo electrónico: Confirme correo electrónico:

Dirección:

Ciudad: Estado (opcional):

Código Postal: País:

Para mayor información acerca de cómo MicroStrategy protege la privacidad y seguridad de sus datos, por favor pulse en el botón de Directiva de Privacidad.

Directiva de privacidad

InstallShield

<Atrás Siguiente > Cancelar

Fuente: Instalación herramienta

IV-25. Microstrategy - Datos para activación

Se autoriza el envío del código de activación.

MicroStrategy 9

Activación del servidor

Solicitar código de activación

MicroStrategy 9™

El programa de instalación ha reunido la información necesaria para solicitar un Código de activación. Una vez finalizada la instalación, el programa de instalación transferirá esta información a MicroStrategy en un formato codificado y enviará automáticamente una solicitud de código de activación. El archivo de activación (C:\Program Files\Common Files\MicroStrategy\Activate.xml) generado después de la instalación contendrá toda la información que se enviará a MicroStrategy.

Al enviar la información, indica que conoce y está de acuerdo con los términos contenidos en la directiva de privacidad de MicroStrategy, de la cual podrá encontrar una copia en <http://www.microstrategy.com/Privacy.asp>.

Usted, o el usuario individual, tienen derecho de acuerdo con las leyes pertinentes (es decir, las directivas europeas 1995/46 y 2002/58), a acceder, actualizar, rectificar/modificar, insertar, eliminar o quitar los datos suministrados en cualquier momento, poniéndose en contacto con MicroStrategy mediante correo electrónico enviado a la dirección privacy@microstrategy.com, por fax al número 703.848.8610 o por carta a MicroStrategy (1850 Towers Crescent Drive, Vienna, VA 22182, USA). Si se encuentra fuera de Estados Unidos, comprende que la información proporcionada a MicroStrategy se transferirá fuera del área económica europea o del país de origen (para propósitos relacionados con el uso del software de MicroStrategy). Da su consentimiento a esta transferencia internacional. Además certifica que el usuario final le ha autorizado a transmitir dicha información.

Nota: Se enviará el código de activación a PGUTIERREZM@INTELLEGO.COM.MX.

¿Desea solicitar un código de activación una vez que termine la instalación?

☒ Sí, deseo solicitar un Código de activación.

☐ No, solicitaré el código de activación más adelante.

InstallShield

<Atrás Siguiente > Cancelar

Fuente: Instalación herramienta

IV-26. Microstrategy - Solicitud clave de activación

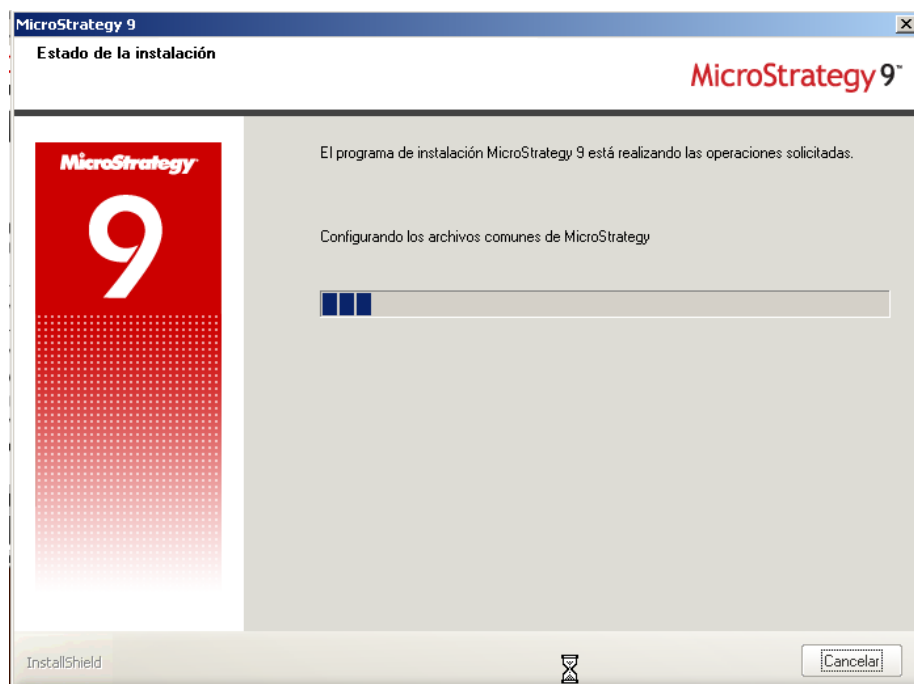
Se confirma los componentes a instalar.



Fuente: Instalación herramienta

IV-27. Microstrategy - Confirma componentes a instalar

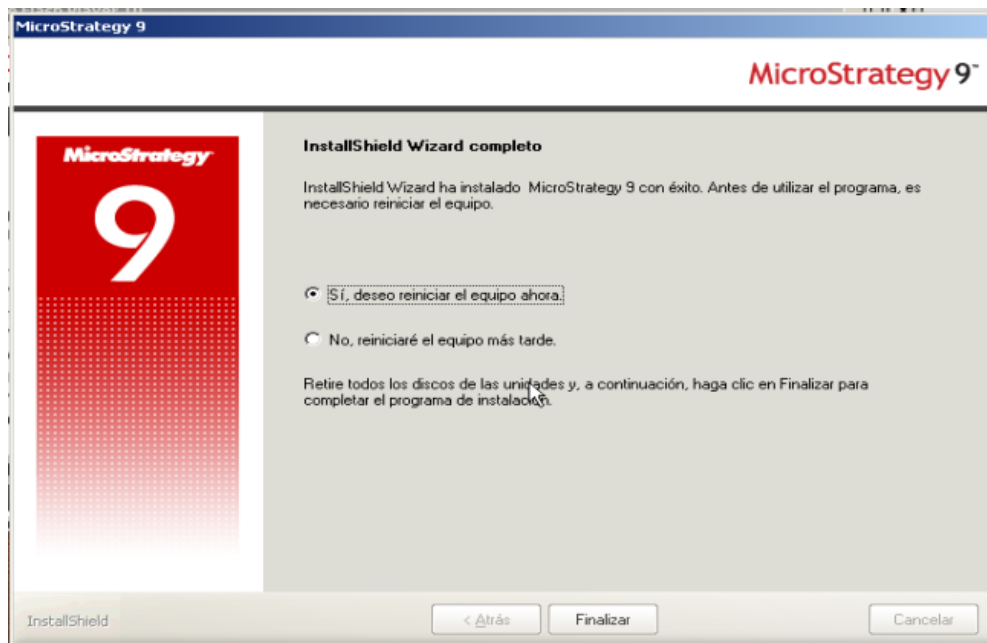
Comienza la instalación y se muestra su progreso.



Fuente: Instalación herramienta

IV-28. Microstrategy - Progreso de instalación

Una vez finalizada la instalación se solicitará el reinicio del equipo.



Fuente: Instalación herramienta

IV-29. Microstrategy - Fin instalación

4.1.13. Modelado DWH

4.1.13.1. Análisis de fuente de datos

La principal fuente de información con la que contamos es la base de datos del sistema POS que utiliza la pizzería en sus operaciones diarias. Esta base de datos se encuentra en SQL Server, y para su análisis ha sido cargada en SQL Server 2008 express edition.

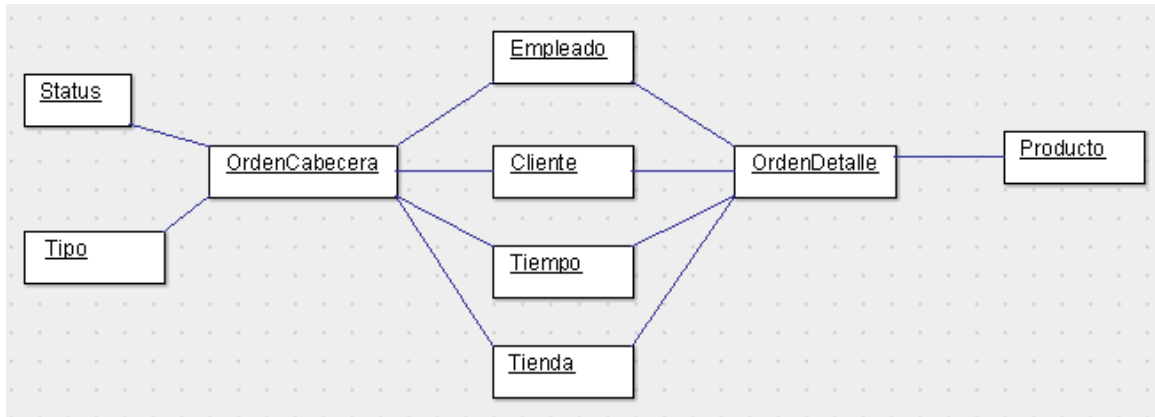
Para comenzar con el análisis de estos datos se nos proporciono un respaldo de la base de datos, la cual fue restaurada en el laboratorio. Visualizando el contenido de dicha base de datos pudimos observar que existen muchos objetos que no son necesarios para cumplir con los objetivos planteados.

Se realizó una revisión de cada una de las tablas en búsqueda de datos que nos pudieran ser útiles en el logro de los objetivos planteados. Estos objetos están relacionados con lo siguiente:

- Productos
- Clientes
- Empleados
- Ordenes

4.1.13.2. HLDM

Con conocimiento de esto se pudo generar lo que podemos observar en la ilustración IV-30. HLDM, el modelo de datos de alto nivel que muestra principalmente los elementos involucrados en la generación de información.



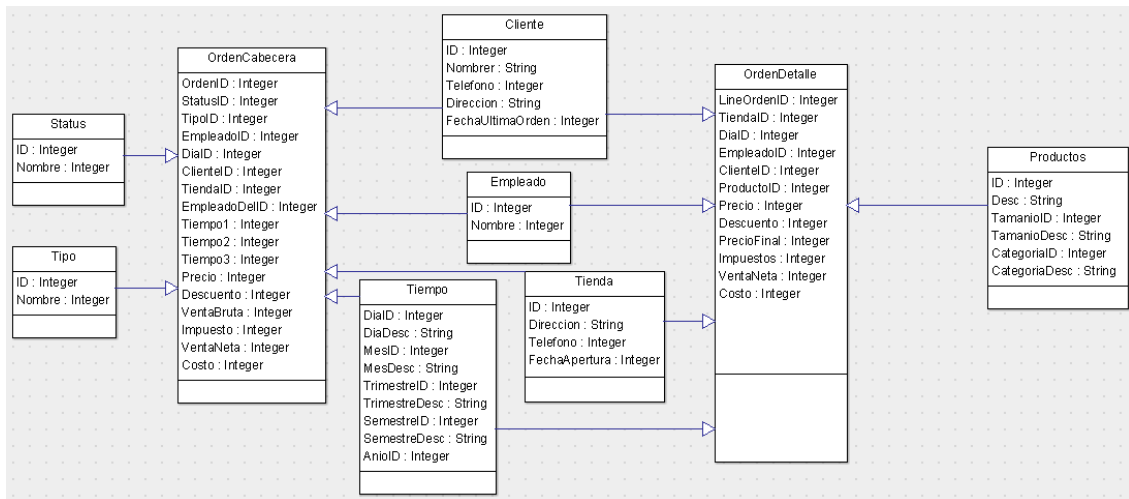
Fuente: Diseño propio

IV-30. HLDM

4.1.13.3. MLDM

Una vez que se tuvo identificados los principales elementos que se debieran involucrar, se procedió a determinar aquellos atributos correspondientes a las entidades identificadas que son útiles para responder a las interrogantes planteadas.

En la ilustración IV-31. MLDM, podemos observar el modelo de datos de medio nivel, el cual es una extensión del HLDM. En este diagrama podemos ver los atributos que fueron identificados en la base de datos del sistema POS y que se consideran de gran importancia.



Fuente: Diseño propio

IV-31. MLDM

4.1.13.4. LLDM

Extendiendo aún más el MLDM obtuvimos el modelo de datos de bajo nivel (modelo físico), en el cual podemos observar un mayor detalle en la descripción de los datos. Definiendo para cada atributo un alias (campo) así como un tipo y tamaño de datos. Este alias sigue una nomenclatura especial en la cual se define la ENTIDAD_ATRIBUTO. Cada una de las entidades ilustradas anteriormente es descrita a mayor detalle a continuación:

EMPLEADO			
CAMPO	DESCRIPCION	TIPO DE DATO	TAMAÑO
EMPLEADO_ID	Identificador del empleado	Numérico	5
EMPLEADO_NOM	Nombre del empleado	alfabético	20

Tabla 10. LLDM - Empleado

TIENDA			
CAMPO	DESCRIPCION	TIPO DE DATO	TAMAÑO
TIENDA_ID	Identificador de tienda	Numérico	5
TIENDA_NOM	Nombre de tienda	alfabético	15
TIENDA_DIR	Dirección de tienda	alfabético	50
TIENDA_TEL	Teléfono tienda	Numérico	10
TIENDA_FA	Fecha de apertura	fecha	

Tabla 11. LLDM - Tienda

ORDEN DETALLE			
CAMPO	DESCRIPCION	TIPO DE DATO	TAMAÑO
ORDEN_DTL_ID	Identificador de línea de orden	numérico	15

ORDEN DETALLE			
CAMPO	DESCRIPCION	TIPO DE DATO	TAMAÑO
ORDEN_ID	Identificador de orden	numérico	10
ORDEN_DTL_LINE	Línea de orden	numérico	5
PRODUCTO_ID	Identificador de producto	numérico	5
STATUS_ID	Status de línea	numérico	3
ORDEN_DTL_CAN	Cantidad	numérico	9,2
ORDEN_DTL_PRE	Precio	numérico	9,2
ORDEN_DTL_DES	Descuento	numérico	9,2
ORDEN_DTL_IMP	Impuestos	numérico	9,2
ORDEN_DTL_VTA	Venta	numérico	9,2

Tabla 12. LLDM - Orden detalle

ORDEN CABECERA			
CAMPO	DESCRIPCION	TIPO DE DATO	TAMAÑO
ORDEN_ID	Identificador de orden	numérico	10
CLIENTE_ID	Identificador de cliente	numérico	5
TIENDA_ID	Identificador de tienda	numérico	5
EMPLEADO_ID_M	Identificador empleado mostrador	numérico	5
EMPLEADO_ID_E	Identificador empleado entrega	numérico	5
TIPO_ID	Identificador tipo de orden	numérico	3
STATUS_ID	Identificador status de orden	numérico	3
ORDEN_PRE	Precio de orden	numérico	9,2
ORDEN_DES	Descuento de orden	numérico	9,2
ORDEN_IMP	Impuesto de orden	numérico	9,2
ORDEN_VTA	VENTA	numérico	9,2
ORDEN_FEC	Fecha y Hora	date time	
ORDEN_TTO	Tiempo de toma orden en seg	numérico	10
ORDEN_TPR	Tiempo de preparación en seg	numérico	10
ORDEN_TDE	Tiempo de despacho en seg	numérico	10
ORDEN_TEN	Tiempo de entrega en seg	numérico	10

Tabla 13. LLDM - Orden Cabecera

PRODUCTO			
CAMPO	DESCRIPCION	TIPO DE DATO	TAMAÑO
PRODUCTO_ID	Identificador del producto	numérico	5
PRODUCTO_DESC	Nombre del producto	alfanumérico	20
PRODUCTO_TAM	Tamaño del producto	alfanumérico	10
PRODUCTO_CAT	Categoría del producto	alfanumérico	10

Tabla 14. LLDM - Producto

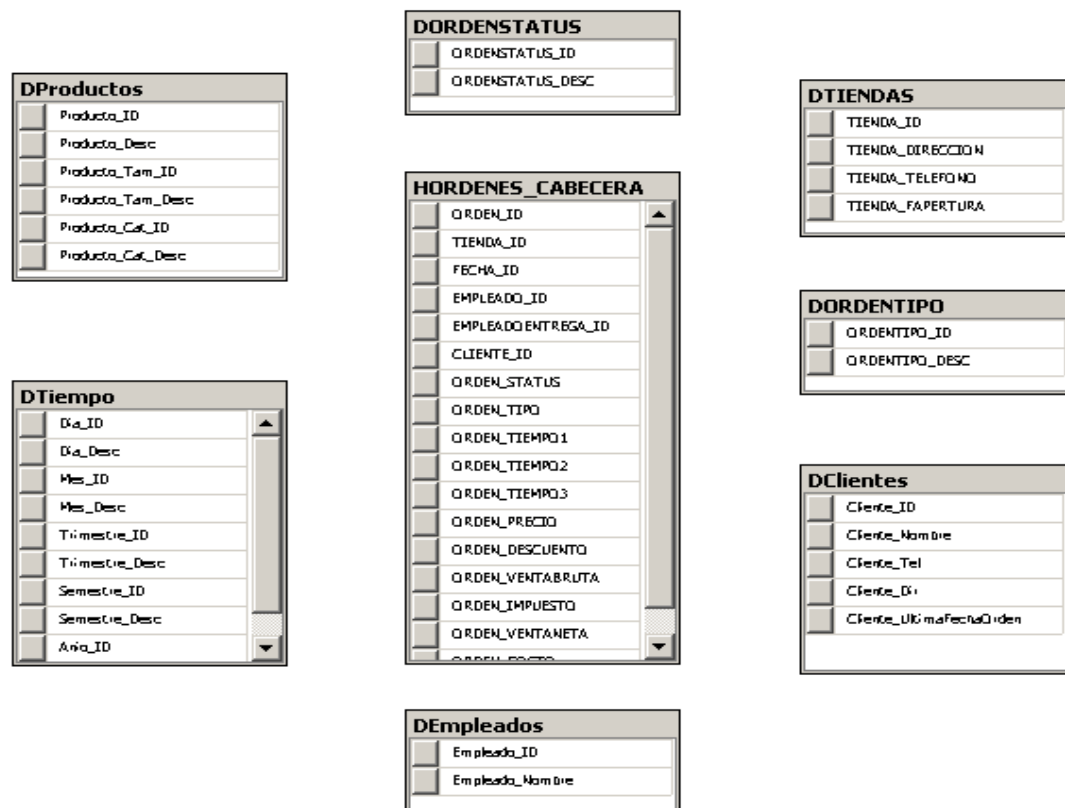
CLIENTE			
CAMPO	DESCRIPCION	TIPO DE DATO	TAMAÑO
CLIENTE_ID	Identificador del cliente	Numérico	5
CLIENTE_NOMBRE	Nombre del cliente	Alfabético	50
CLIENTE_DIRECCION	Dirección del cliente	Alfanumérico	50
CLIENTE_TELEFONO	Teléfono del cliente	Numérica	10

Tabla 15. LLDM - Cliente

4.1.13.5. Construcción

De acuerdo a las necesidades del negocio se realizaron dos modelaros principales siguiendo esquemas de estrella con la finalidad de que las consultas sean lo más optimizadas posibles. Estos modelados se hicieron tomando en cuenta las necesidades de información que se analizaron del negocio.

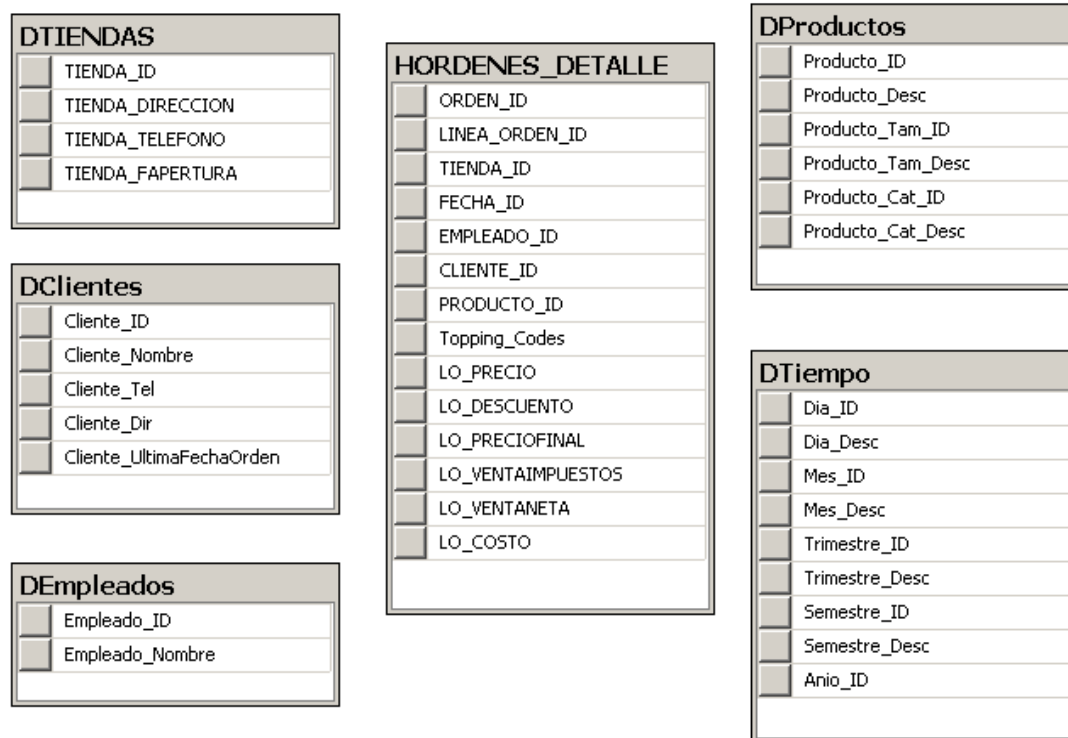
En primera instancia tenemos la cabecera de las órdenes como tabla de hechos en la cual tenemos las principales métricas de tiempos en los que se despacho una orden así como las ventas. Estas métricas se pueden analizar por una jerarquía de tiempo, por empleado, por cliente, tipo de orden, estatus de la orden,



Fuente: Diseño propio

IV-32. Modelo de datos - Cabecera

El otro modelado está enfocado en el detalle de las ordenes como tabla de hechos en la cual las principales métricas que se tienen son las ventas, la diferencia con la cabecera de ordenes es que el detalle nos permite hacer el cruce con la dimensión de productos por lo que podemos analizar las ventas por producto, además de las dimensiones de cliente, empleado y tiempo.



Fuente: Instalación herramienta

IV-33. Modelo de datos - Detalle

4.1.14. Datos

4.1.14.1. Análisis de datos

Dentro de la exploración de los datos encontramos que para poder medir la productividad de los empleados contamos con el registro de los tiempos que se manejan en la toma de órdenes y la entrega de dichas ordenes.

Para cada orden se registran la fecha y hora en la que se registra la orden, el tiempo que tomo registrar la orden, el tiempo de entrega, tiempo que pasa la orden en estantes, tiempos de preparación.

Con estos registros se puede medir los tiempos efectivos de los empleados en cuanto al servicio que ofrecen, así mismo estos datos se pueden relacionar con las ventas para saber de los empleados que menos venden que tiempos de servicio tienen registrados. En la ilustración

IV-34. Análisis de datos, podemos observar una ejemplificación de los datos con los que se cuenta. Con estos datos podemos definir estrategias que nos permitan motivar a nuestros empleados de tal manera que estos se vuelvan más productivos, así como evaluar otras posibles causas de la baja productividad e implementar acciones correctivas.

	Order_Number	Customer_Code	OrderFinalPrice	OrderRoyaltySales	Sales_Tax1
1	148048	0	3000.00	2608.70	391.30
2	160553	16051	2687.00	2316.43	-0.08
3	156745	6398	2383.00	2054.29	0.00
4	165354	3246	2250.00	1939.50	0.00
5	152239	0	2118.00	1841.75	276.25
6	158313	4404	1990.00	1715.52	0.00
7	165766	20485	1990.00	1715.51	0.00
8	158058	0	1794.00	1546.56	-0.01
9	151229	19485	1791.00	1557.42	233.58
10	148635	6908	1640.00	1426.08	213.92
11	161513	7152	1633.00	1407.74	0.00
12	165995	20508	1625.00	1411.19	0.00
13	163954	0	1608.00	1386.23	-0.01
14	159549	20096	1596.00	1375.88	0.00
15	165113	17501	1592.00	1372.48	-0.08
16	165083	6497	1559.00	1343.95	0.00
17	152304	19044	1550.00	1347.90	202.10
18	158199	9236	1542.00	1329.22	0.00
19	164327	0	1540.00	1327.59	-0.02
20	147710	0	1526.00	1326.96	199.04
21	161831	20000	1480.00	1281.37	0.00

Fuente: Diseño propio

IV-34. Análisis de datos – Importes de venta

Los datos relacionados con las ventas contamos con las órdenes, el cliente relacionado, la venta total, el impuesto asociado y la venta sin impuesto. Así mismo estos datos se pueden relacionar con los tiempos de sus ordenes correspondientes para determinar si el servicio fue proporcionado en los tiempos mínimos requeridos, así mismo podemos identificar los clientes que nos han generado un mayor número de ventas así como sus hábitos de consumo para poder determinar ofertas especiales para estos clientes. Los datos que podemos visualizar sobre esta temática los vemos en la ilustración IV-35. Análisis de datos - .

	Order_Number	customer_code	Driver_ID	Added_By	Order_type_code	Actual_order_date	Order_Time	OrderDeliveryTimeSecs	OrderLoadTimeSecs	OrderRackTimeSecs	OrderDispatchTimeSecs	OrderDeliveryTimeSecs
1	148048	0	1638	0867	C	2009-11-13 21:13:19.000	274	NULL	41	NULL	NULL	NULL
2	160553	16051	4963	4963	D	2010-01-28 13:10:00.000	501	1519	45	152	680	1519
3	156745	6398	4963	4963	D	2010-01-04 13:00:00.000	330	1444	506	30	1019	1444
4	165354	3246	1347	1767	D	2010-02-27 14:37:16.000	4146	1284	18	287	788	1284
5	152239	0	NULL	4118	C	NULL	110	NULL	0	0	NULL	NULL
6	158313	4404	9876	9876	D	2010-01-14 13:30:00.000	1204	1419	201	274	958	1419
7	165766	20485	4102	7624	D	2010-02-28 21:21:42.000	365	1530	205	106	794	1530
8	158058	0	NULL	4963	C	NULL	58	NULL	0	0	NULL	NULL
9	151229	19485	4963	1638	D	2009-12-04 13:45:00.000	521	1852	263	109	855	1852
10	148635	6908	9876	4225	D	2009-11-17 13:19:52.000	429	1385	163	190	836	1385
11	161513	7152	4152	0867	D	2010-02-02 13:46:31.000	140	1584	111	276	870	1584
12	165995	20508	0893	8659	D	2010-03-02 23:28:41.000	224	757	33	282	798	757
13	163954	0	3955	3955	C	2010-02-17 19:23:39.000	341	NULL	11	NULL	NULL	NULL
14	159549	20096	1972	1767	D	2010-01-21 19:54:16.000	4567	1261	12	398	893	1261
15	165113	17501	4118	3005	D	2010-02-26 12:52:01.000	854	1535	148	186	817	1535
16	165083	6497	4655	4102	D	2010-02-25 21:04:59.000	396	887	195	201	879	887
17	152304	19044	1189	9876	D	2009-12-10 13:20:00.000	396	1880	552	-115	920	1880
18	158199	9236	NULL	4959	D	NULL	805	NULL	0	0	NULL	NULL
19	164327	0	NULL	4963	D	NULL	576	NULL	0	0	NULL	NULL
20	147710	0	9876	9876	C	2009-11-11 13:30:00.000	328	NULL	74	NULL	NULL	NULL

Fuente: Diseño propio

IV-35. Análisis de datos - Tiempos

4.1.14.2. Desarrollo proceso ETL

Con la finalidad de poblar el DWH construido se diseño y construyo un proceso de ETL. Para lo cual se utilizo la herramienta Jaspersoft ETL. Para utilizar dicha herramienta se configuro la conexión a un repositorio y se creó el proyecto correspondiente.

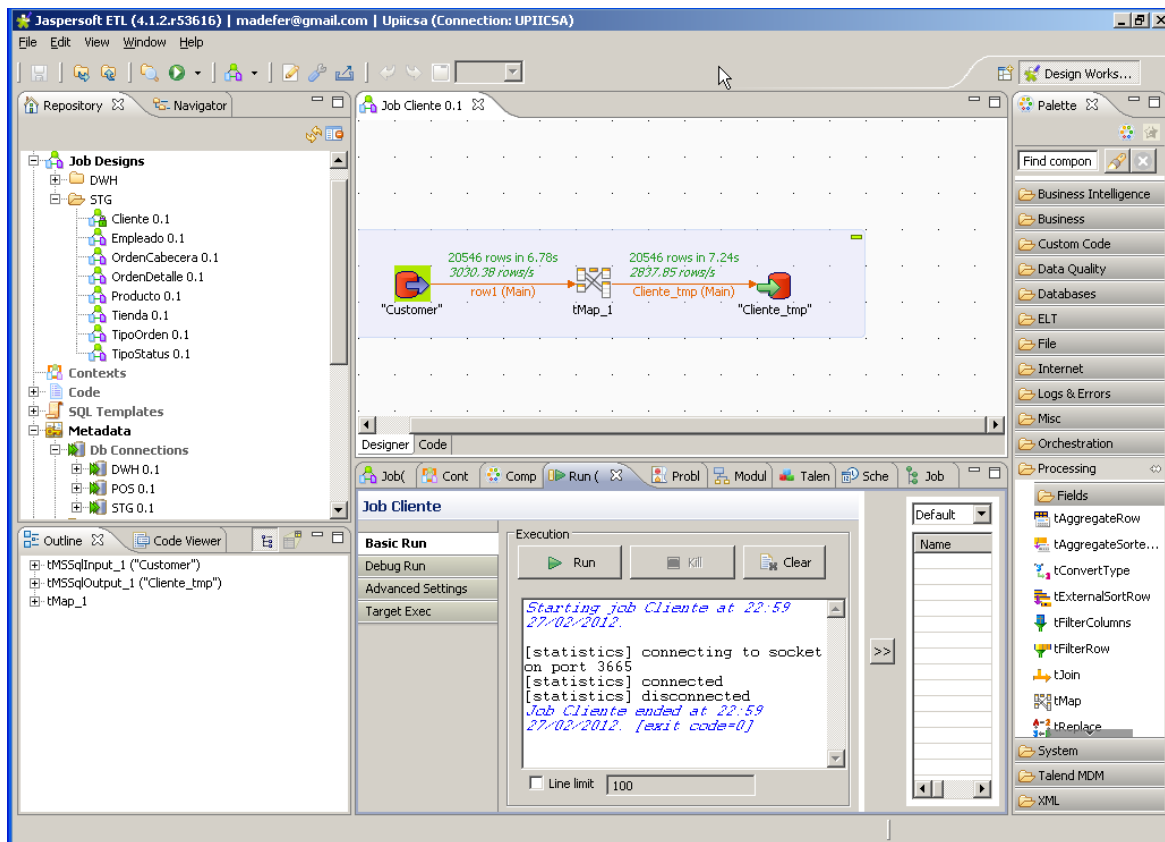
Una vez contando con el proyecto se crearon las conexiones a las bases de datos correspondientes. De acuerdo a la arquitectura definida para el proceso ETL, se crean las conexiones a la fuente de datos POS, a la base de datos de staging en la que se realizan las transformaciones necesarias y finalmente la base de datos DWH, la cual contiene los datos finales.

Ya que se tenían las conexiones se importa la metadata de las tablas correspondientes para poder crear los procesos ETL. Para iniciar con las transformaciones se crean los jobs necesarios. Siguiendo la arquitectura definida, en primera instancia se extrajeron los datos de la fuente de datos POS a la base de datos STG sin realizar mayores transformaciones. Con los datos en la base de datos de staging, se realizan las transformaciones necesarias para ajustarse al modelado de datos del DWH.

Ya que se tienen los datos con el formato adecuado para ser insertados en el DWH, son insertados de tal manera que el DWH es poblado con la información correspondiente.

SpagoBI no tiene una herramienta propia para ETL, se basa en el mismo motor que utiliza Jaspersoft por lo que el desarrollo del proceso ETL solo se realizo con Jaspersoft, Microstrategy es

una herramienta enfocada a trabajar con un DataWarehouse ya construido por lo que no cuenta con una aplicación para la construcción de procesos ETL.



Fuente: Diseño propio

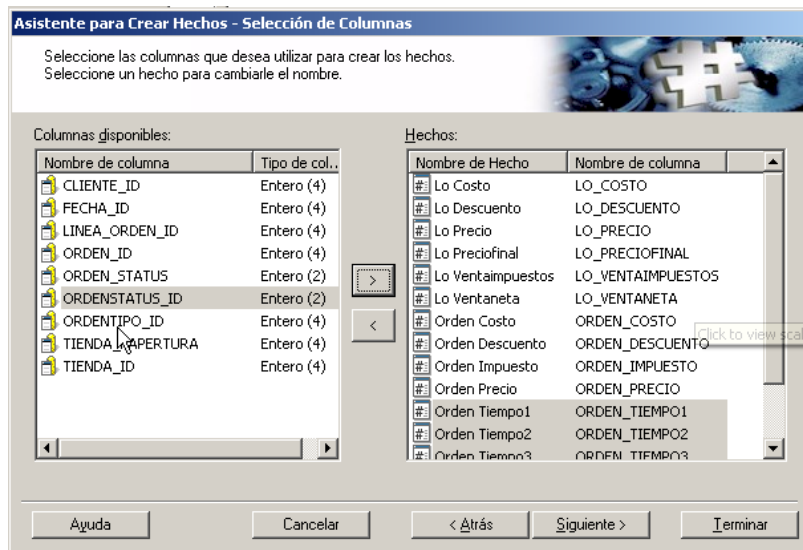
IV-36. Jaspersoft - ETL

4.1.15. Modelo multidimensional

El modelo multidimensional para la solución de BI se generó con la herramienta Microstrategy. Esta herramienta permite la creación de elementos base a partir de las tablas construidas en el proceso ETL.

Durante la configuración de la herramienta está establecido un paso para la definición de dos de los elementos base para el modelo multidimensional, estos elementos son:

- Hechos que consiste en todos aquellos campos del DataWarehouse que representen un valor cuantitativo de los datos, es decir aquellos datos numéricos referentes al negocio.



Fuente: Diseño propio

IV-37. Microstrategy - Hechos

- Atributos que consiste en todos aquellos campos del DataWarehouse que representan un valor cualitativo que permite dar mayor significado a los hechos, permitiendo hacer un análisis de los hechos por sus diversos atributos.

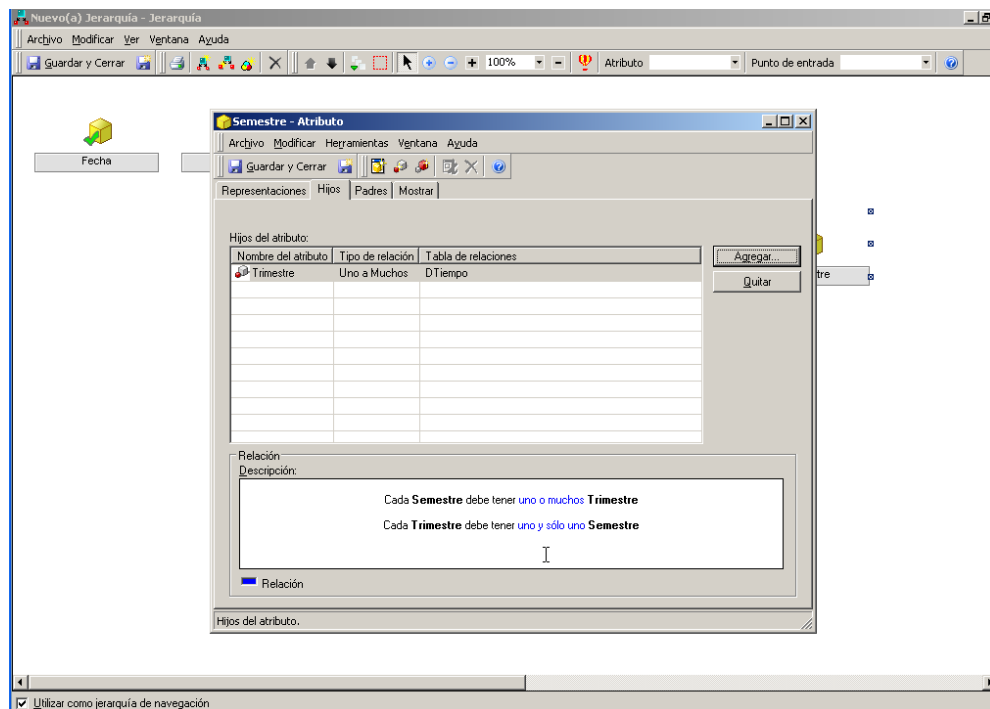


Fuente: Diseño propio

IV-38. Microstrategy - Atributos

Estos elementos son la base de un modelo multidimensional ya que permiten generar los elementos de siguiente nivel utilizados en los modelos multidimensionales utilizados por Microstrategy. A partir de los elementos base, podemos construir los elementos necesarios en la generación de reportes.

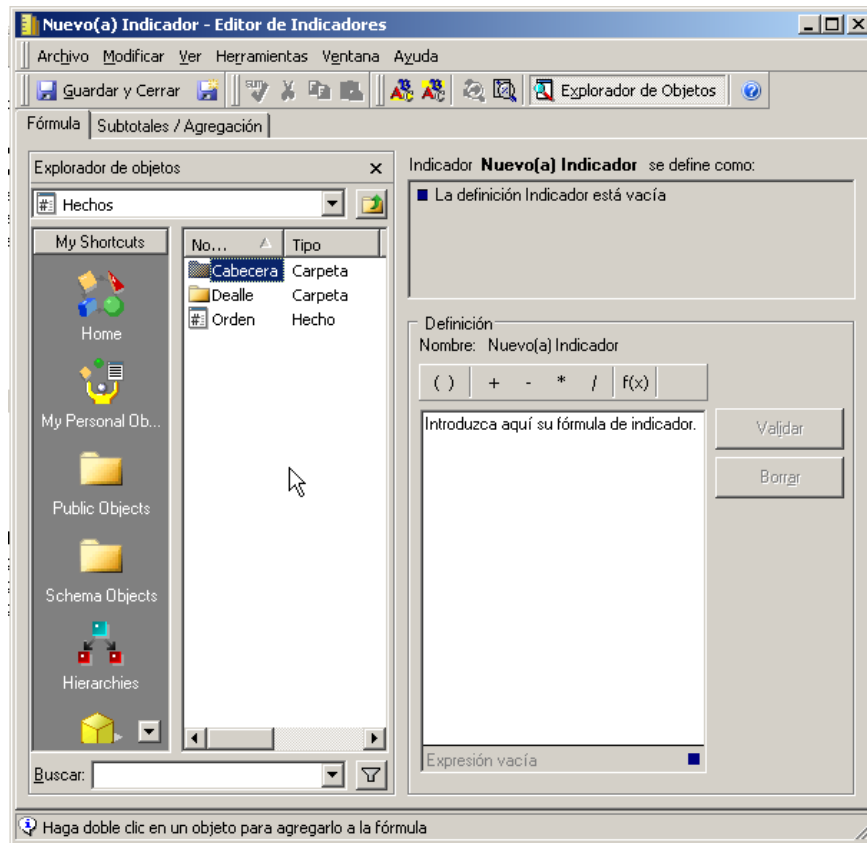
Los atributos permiten la generación de jerarquías de navegación, a través de las cuales se acceder a atributos de diferentes niveles de granularidad, lo que permite ver las métricas agrupadas a diferentes niveles.



Fuente: Diseño propio

IV-39. Microstrategy - Jerarquías

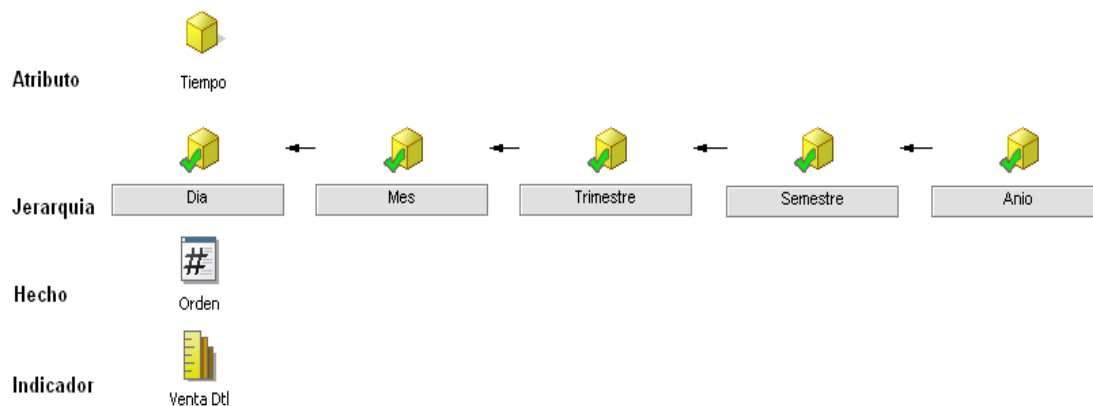
Los hechos permiten la creación de métricas, los cuales son valores calculados a partir de los hechos obtenidos directamente del DataWarehouse.



Fuente: Diseño propio

IV-40. Microstrategy - Indicadores

Con esto se construyen los elementos básicos del modelo multidimensional que permite la generación de los reportes.

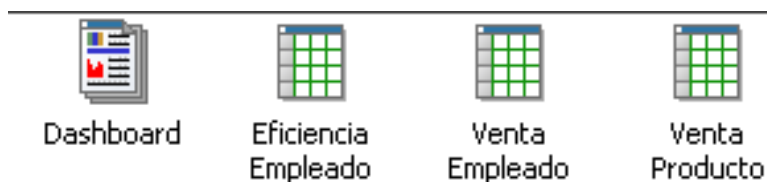


Fuente: Diseño propio

IV-41. Microstrategy - Componentes modelo multidimensional

4.1.16. Presentación

Una vez generado el modelo multidimensional es posible generar los reportes necesarios mediante la utilización de sus elementos. En este caso se generaron tres reportes.



Fuente: Diseño propio

IV-42. Reportes generados

Uno de ellos mostraba principalmente al empleado que genera el mayor número de ventas, esto sirve de indicador para poder conocer que es lo que motiva a este empleado para generar dichos resultados e incentivarlo a continuar con el mismo o mayor nivel de productividad.

The screenshot shows a software window titled 'Eficiencia Empleado'. It has a menu bar (Archivo, Modificar, Ver, Insertar, Formato, Datos, Cuadrícula, Mover, Ventana, Ayuda) and a toolbar. On the left, there's a pane for 'Objetos de informe' with a tree view showing 'Empleado' and 'Tiempo'. The main area displays a table with the following data:

Empleado	Indicadores	Tiempo
4655 CERVANTES ARRONA JUAN		1,713
4102 RAMIREZ MARTINEZ RUBEN		1,623
GOMEZ FRIAS MARIA DE JESUS		1,607
RODRIGUEZ VICTOR		1,584
MARTINEZ SEPULVEDA ABIGAIL		1,577
GOMEZ GUADALUPE		1,571
4118 ROJAS HERNANDEZ JOSE MARTIN		1,566
JIMENEZ HURTADO OSCAR		1,542
GONZALEZ MICHELLE		1,540
DIAZ HERNANDEZ RENE		1,536
4962 SANTIAGO MACIAS JULIO		1,535
1347 LOPEZ DIAZ JOSE RODRIGO		1,526
9876 MENDOZA DIAZ LUIS EDHER		1,521
MARTINEZ RODRIGUEZ 4152 OMAR		1,510
GONZALEZ BERNAL ALFREDO		1,506
ROSAS CABELLO TANIA		1,486
ARELLANO JULIO		1,479
0893 SANCHEZ PEREZ DANTE ALFREDO		1,448
9384 TAPIA PINEDA GUILLERMO		1,434
4963 ECHEVARRIA LOPEZ JONATHAN		1,408
GUERRA EDUARDO		1,390
ARREGUIN ESTRADA XOCHITL		1,386
1972 SORIANO MARTINEZ ESTEBAN		1,373
ALEMAN GRACIDA NIMCI HEDREY		1,369
HDEZ LIZBET		1,368
GARCIA MARQUEZ JUANA		1,355
LOPEZ LIZBETH		1,309
HERNANDEZ ESTUDILLO KARIM		1,304
ENGAVI MORENO ARELI BERENICE		1,230

At the bottom of the window, there's a status bar showing 'Ejecución completa', 'Tiempo de Ejecución: 00:00:03', 'Filas: 29', 'Columnas: 1', and 'Plantilla Local Estándar'.

Fuente: Diseño propio

IV-43. Tiempos por empleado

Así mismo se genero un reporte para conocer que empleado es el que le toma más tiempo completar una orden, esto nos da la visibilidad de conocer que empleados están siendo menos productivos para poder realizar una investigación sobre los porqué y atender las causas que generen que dichos empleados produzcan por encima del promedio establecido.

Empleado	Indicadores	Venta				
		Noviembre 2009	Diciembre 2009	Enero 2010	Febrero 2010	Total
ALEMAN GRACIDA NIMCI HEDREY		7,501,443	10,110,641	10,110,641	9,132,192	36,854,916
GARCIA MARQUEZ JUANA		8,996,044	12,125,103	12,125,103	10,951,706	44,197,954
4962 SANTIAGO MACIAS JULIO			5,319,562	5,319,562		10,639,123
HDEZ LIZBET		15,790,116	21,282,330	21,282,330	19,222,750	77,577,525
ENGAVI MORENO ARELI BERENICE			6,693,906	6,693,906	6,046,108	19,433,920
Total		32,287,603	55,531,540	55,531,540	45,352,755	188,703,438

Fuente: Diseño propio

IV-44. Top 5 Ventas por empleado

Finalmente se generó un reporte que nos de la visibilidad para conocer cuál de nuestros productos es el más exitoso, esto nos permitirá establecer estrategias que permitan que se le ofrezca a los clientes promociones adecuadas a este habito de consumo, lo que implica un aumento en las ventas.

Producto	Indicadores	Venta Dtl
GDEPIZ		968,144
MEDPIZ		446,845
C14PIZ		354,431
D4PIZ		289,785
PAPOT		166,450
CANELAZ		149,530

Fuente: Diseño propio

IV-45. Top 6 de ventas por producto

Contando con estos tres reportes se pudo generar un reporte tipo resumen que engloba estos tres resultados.



Fuente: Diseño propio

IV-46. Dashboard

CONCLUSIONES

Para la realización de este trabajo se realizó una investigación en la que se logró integrar la información de 34 fuentes, que se consideraron compatibles con este trabajo de tesis, permitiendo enriquecer este trabajo. La metodología presentada en este trabajo, toma en cuenta los contenidos de estas fuentes para tratar de cubrir los hallazgos actuales de las PyMEs.

La metodología propuesta es enfocada en principalmente conocer el negocio para el cual se pretende desarrollar cierto conocimiento, esto permite guiar al dueño del negocio para establecer prioridades y realizar un desarrollo gradual que sea congruente con su nivel de madurez. De tal manera que se genere una conciencia de los tiempos que implicará tener lista cierta información y no se realicen los trabajos en tiempos exageradamente cortos que impidan hacer un buen trabajo. Lo que se pretende con esta metodología es dar la pauta para que las empresas puedan introducirse al mundo de la inteligencia de negocios y que esto les permita ser más competitivas.

La metodología propuesta abarca todos los puntos conocidos de la inteligencia de negocios para ser aplicados, engloba aspectos de negocio, un poco de administración de proyectos y los tres puntos medulares de la inteligencia de negocios que son la construcción de un DataWarehouse, la construcción de los procesos ETL que permiten poblar el DataWarehouse, los procesos de análisis como son cubos de información y minería de datos y la explotación de dicha información por medio de reportes, los cuales permitirán a los usuarios finales tomar las decisiones necesarias para alcanzar los objetivos establecidos durante la familiarización con el negocio.

El desarrollo de soluciones de inteligencia de negocios con herramienta open source implica un proceso complicado con una curva de aprendizaje bastante amplia, además de no existir toda la documentación que uno quisiera y la que existe no siempre suele ser tan clara. Es por eso que a mi consideración es importante que se aproveche que las herramientas propietarias estén viendo a las PyME como un mercado potencial, abriéndoles las puertas al mundo de la toma de decisiones ofreciendo versiones de sus herramientas de manera gratuita. De esta manera se puede aprovechar la estabilidad que ofrece una herramienta propietaria versus una herramienta libre.

No es necesario estar casado con una sola herramienta en el desarrollo de este trabajo al final del día se trabajó en conjunto una herramienta libre y una herramienta propietaria demostrando que estos universos no están peleados.

De las herramientas opensource elegidas para el desarrollo de este trabajo, puedo concluir que la más avanzada es Jaspersoft. Ha alcanzado altos niveles de manejabilidad, le faltan mejoras en aspectos de análisis de los datos pero no dudo que en un futuro no tan lejano evolucione a tal grado que se facilite la explotación de dicho aspecto.

SpagoBI, va por buen camino, sin embargo aun le falta mucho que madurar y tiene varias áreas de oportunidad a ser explotadas, comparte varios desarrollos con Jaspersoft entre ellos la herramienta de ETL y el Studio para la generación de reportes. Además de eso hace falta documentación más clara para poder desarrollar con dicha herramienta. No se logra hacer una instalación muy adecuada del servidor y las herramientas no mandan mensajes de errores claros.

OpenI no fue posible su instalación y no se pudo probar realmente la herramienta debido a que en la página ya solo está disponible un plug-in que trabaja con Pentaho.

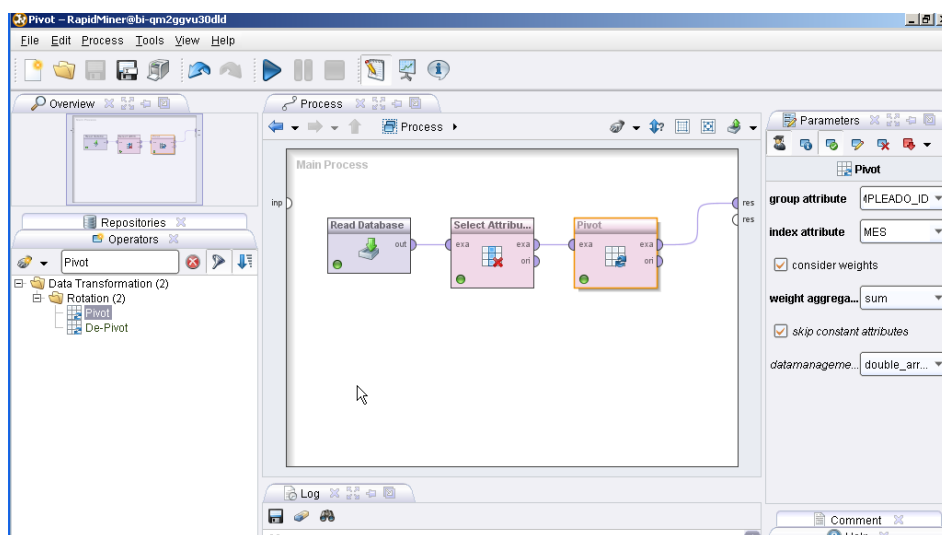
ANEXO I. RapidMiner

Es recomendable que una PyME se apoye de tecnología Open Source para construir soluciones lo suficientemente confiables para poder manipular y explotar la información. Por ejemplo la herramienta Rapidminer cuenta con una instalación bastante sencilla, tiene un amplio catalogo con diversos procesos que permiten el desarrollo e implementación de minería de datos.

Cuando el nivel de madurez de una empresa PyME es muy bajo, la recomendación que se hace para que pueda iniciar a introducir prácticas de inteligencia de negocios sería:

- Se debe comenzar a hacer un registro de las operaciones que se realicen en la empresa, generalmente es bueno comenzar por las ventas para después de manera paulatina continuar con los demás temas importantes a considerar como costos, recursos humanos, presupuestos, etc.
- Se debe guardar un registro de los productos y/o servicios que ofrezca la empresa. Conforme vaya transcurriendo el tiempo y sea necesario conservar una clasificación de productos, se podrán construir jerarquías.
- Si es posible, se debe de guardar un registro de los clientes con los que se cuenta, asignándoles un identificador único. Conforme vaya transcurriendo el tiempo y sea necesario conservar una clasificación de clientes, se podrán construir jerarquías.

RapidMiner cuenta con un operador que nos permite pivotear la información para su consulta.



Fuente: Diseño propio

Después de ejecutado el proceso de pivote se despliega la información como se puede observar en la siguientes información. En este ejemplo se está revisando la venta neta por empleado a través del tiempo a nivel mensual.

The screenshot shows the Pivot application window. The main area displays a table with the following data:

Row No.	EMPLEADO	ORDEN_VENTANETA_FEBRERO 2010	ORDEN_VENTANETA_DICIEMBRE 2009	ORDEN_VENTANETA_ENERO 2010	ORDEN_VENTANETA_MARZO
1	0059	142.250	?	?	?
2	0867	131.030	0	142.250	176.720
3	0893	107.760	117.390	176.720	131.030
4	1060	193.950	?	184.490	151.720
5	1106	343.120	310.430	142.250	343.120
6	1189	?	103.480	171.560	?
7	1347	205.170	173.050	131.030	158.620
8	1638	?	155.660	343.120	?
9	1767	131.030	293.050	176.720	131.030
10	1968	?	?	?	111.210
11	1972	176.720	?	142.250	381.030
12	3005	161.200	132.170	171.560	81.900
13	3150	?	?	189.650	?
14	3955	158.620	346.100	171.560	171.560
15	4016	?	?	?	?
16	4102	176.720	?	171.560	171.560
17	4118	176.720	206.090	284.490	81.900
18	4149	133.620	86.090	274.140	?
19	4152	342.240	132.170	262.930	?

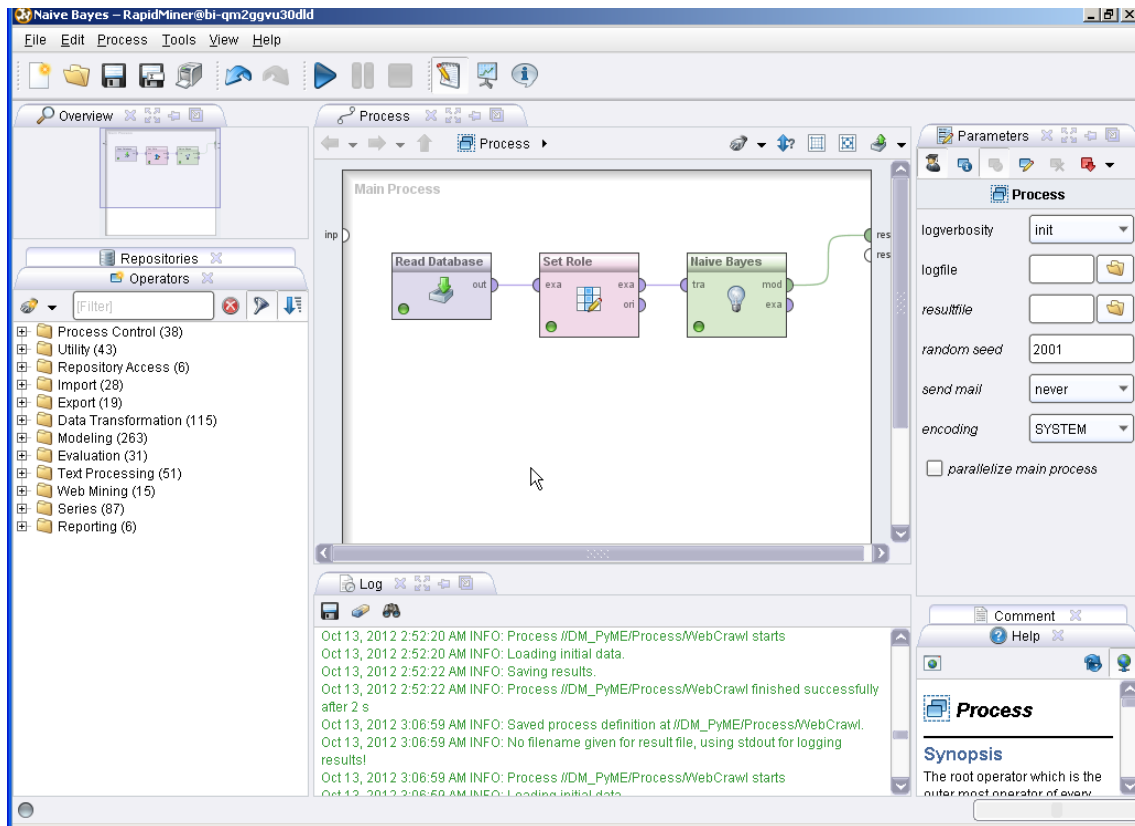
The interface also includes a menu bar (File, Edit, Process, Tools, View, Help), a toolbar, and a log window at the bottom showing system messages.

Fuente: Diseño propio

IV-48 - Resultado operador Pivot

Cuando el nivel de madurez con el que cuenta la empresa es alto, es mayor el beneficio de utilizar métodos de minería de datos que nos permitan hacer análisis más avanzado de la información.

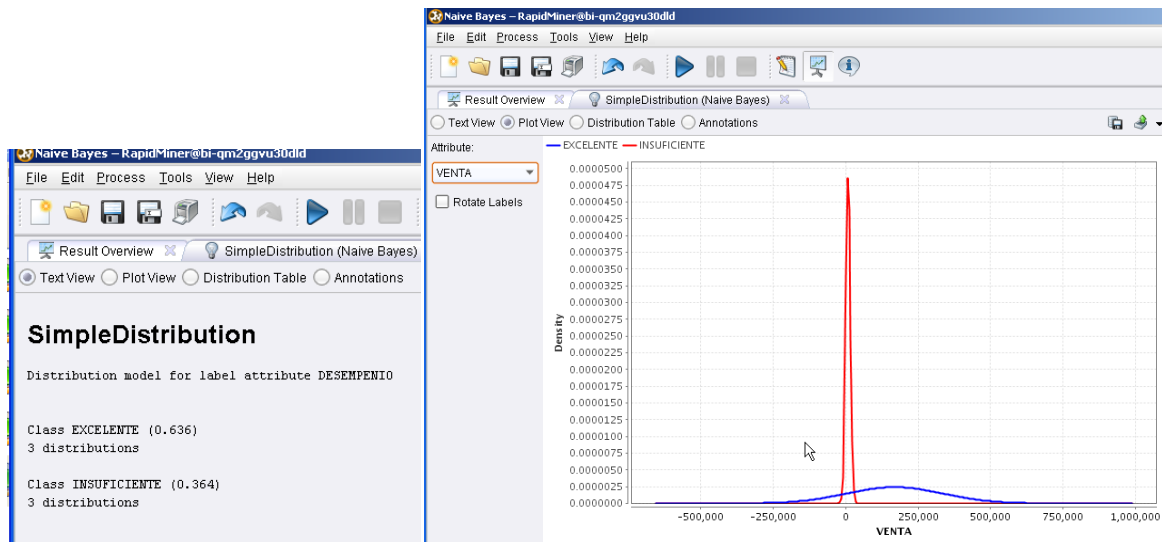
Esta aplicación cuenta con diversas herramientas para la aplicación de la minería de datos, en el caso que podemos ver en la imagen IV-49 - Naives Naves, se clasifico el desempeño de los empleados con el uso del algoritmo Naives Bayes. Se hace una lectura a una tabla de SQL Server y se clasifican los atributos como ID, label, regular y finalmente se ejecuta el algoritmo en cuestión.



Fuente: Diseño propio

IV-49 - Naives Naves

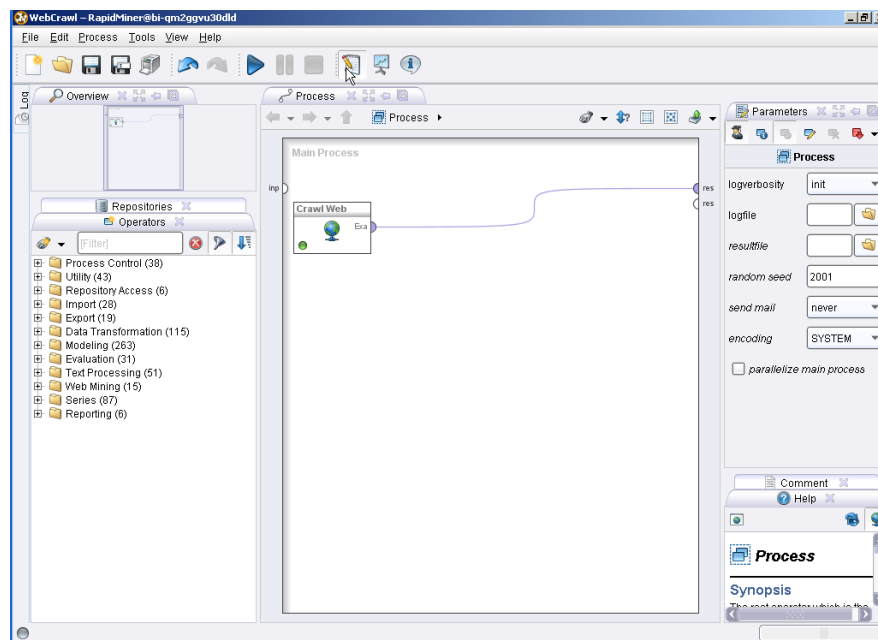
Una vez que termina el procesamiento de la información, nos muestra los resultados, en este caso tenemos que existe un 63.6% de que el desempeño de los empleados sea excelente contra un 36.4%. Y nos da la opción de ver la densidad del desempeño de acuerdo a los atributos regulares establecidos que fueron: Venta, Tiempo de ejecución, Número de órdenes levantadas.



Fuente: Diseño propio

IV-50 – Resultados

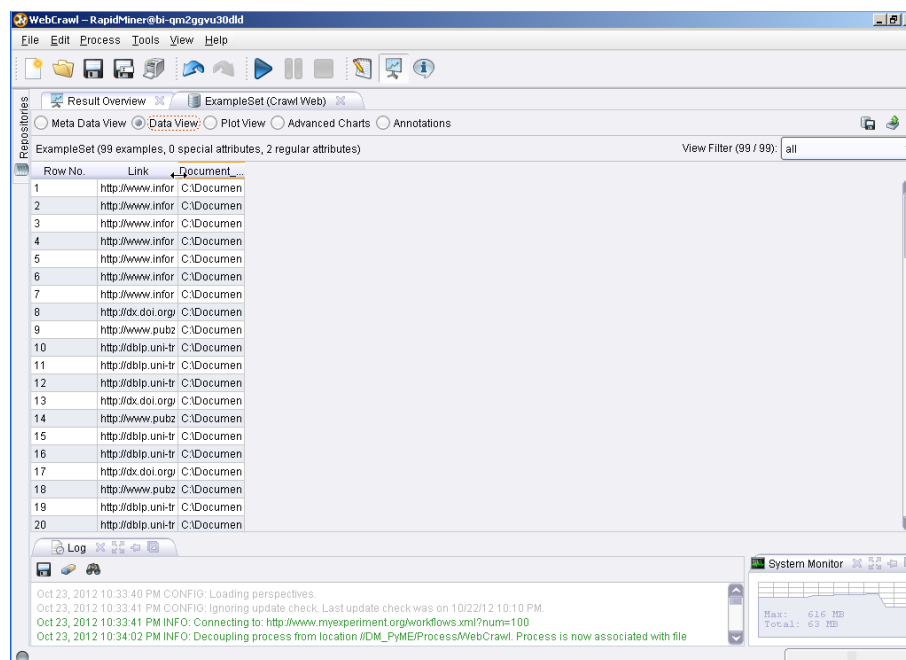
Existe una opción adicional para obtener mayor información para el minado de datos. Esta opción consiste en acudir a páginas web que nos permitan extraer información para el minado de datos. El objetivo del ejercicio realizado era obtener las páginas que cumplieran con cierto patrón. En este caso se hizo mediante la url, indicándole una página como base y de acuerdo a las ligas que tuviera referenciadas tomaría en cuenta aquellas que tuvieran la palabra “IJBIDM” y las almacenaría.



Fuente: Diseño propio

IV-51 - Minado Web

Cuando el proceso termina, regresa los resultados como se puede observar en la pantalla de la imagen IV-52 - Resultados Minado Web. Así mismo debería almacenar en una carpeta que se haya definido previamente los archivos en el formato que nosotros elegimos que puede ser html o xml.



The screenshot shows the 'WebCrawl - RapidMiner' application window. The main view is 'Data View' showing a table with 20 rows of crawled data. The table has columns 'Row No.', 'Link', and 'Document...'. The 'Link' column contains various URLs, and the 'Document...' column shows the file path 'C:\Documents'.

Row No.	Link	Document...
1	http://www.infor	C:\Documents
2	http://www.infor	C:\Documents
3	http://www.infor	C:\Documents
4	http://www.infor	C:\Documents
5	http://www.infor	C:\Documents
6	http://www.infor	C:\Documents
7	http://www.infor	C:\Documents
8	http://dx.doi.org	C:\Documents
9	http://www.pubz	C:\Documents
10	http://dblp.uni-tr	C:\Documents
11	http://dblp.uni-tr	C:\Documents
12	http://dblp.uni-tr	C:\Documents
13	http://dx.doi.org	C:\Documents
14	http://www.pubz	C:\Documents
15	http://dblp.uni-tr	C:\Documents
16	http://dblp.uni-tr	C:\Documents
17	http://dx.doi.org	C:\Documents
18	http://www.pubz	C:\Documents
19	http://dblp.uni-tr	C:\Documents
20	http://dblp.uni-tr	C:\Documents

At the bottom of the window, there is a 'Log' panel showing system messages and a 'System Monitor' panel showing memory usage (Max: 616 MB, Total: 63 MB).

Fuente: Diseño propio

IV-52 - Resultados Minado Web

BIBLIOGRAFÍA

(s.f.). Obtenido de <http://www.jaspersoft.com/>

(s.f.). Obtenido de <http://www.pentaho.com/>

(s.f.). Obtenido de <http://openi.org/>

(s.f.). Obtenido de <http://www.spagoworld.org/>

(s.f.). Obtenido de <http://www.cloveretl.com/>

(s.f.). Obtenido de <http://www.palo.net/>

(s.f.). Obtenido de <http://www.pocolap.org/>

(s.f.). Obtenido de <http://www.knime.org/>

(s.f.). Obtenido de <http://www.eti.hku.hk/alphaminer/>

(s.f.). Obtenido de <http://www.cs.waikato.ac.nz/ml/weka/>

(s.f.). Obtenido de <http://rapid-i.com/content/view/181/190/>

(s.f.). Obtenido de <http://www.oracle.com/us/solutions/business-analytics/business-intelligence/enterprise-edition/overview/index.html>

(s.f.). Obtenido de <http://www.microstrategy.com.mx/>

(s.f.). Obtenido de <http://www-01.ibm.com/software/analytics/cognos/>

(s.f.). Obtenido de <http://www.microsoft.com/en-us/bi/GetMicrosoftBI.aspx>

(s.f.). Obtenido de <http://www.qlikview.com/>

- Aoki, M., & Hasebe, M. (2012). Significance of Learning Process in BSC Introducing Process in Japanese Small and Medium Enterprises. *Tohoku Management - Accounting Research Group*.
- Aversano, L., Grasso, C., & Tortorella, M. (2011). Evolving Enterprise Information System in Small and Medium Enterprise: an opportunistic approach and a case study. *IEEE Conference on Commerce and Enterprise Computing*.
- Benki, C., & Papastathopoulos, A. (2009). The Impact of Structured, Unstructured and Integrated Decision Support Systems on SME Economic Performance. An Empirical Study. *3rd International Conference on Communications and information technology*.
- Boussaid, O., Ben Messaoud, R., Choquet, R., & Anthoard, S. (2006). X-warehousing: an XML-based approach for warehousing complex data. *10th East-European Conference on Advances in Databases and Information Systems*.
- Bruhn, M., Karlan, D., & Schoar, A. (2012). The Impact of Consulting Services on Small and Medium Enterprises: Evidence from Randomized Trial in México. *Department of Economics Yale University*.
- Castillo, A. (2007). Las TIC, la clave para el desarrollo de las PYME. *TECNO-ECONOMIA*.
- Chen, X., Ye, Y., Williams, G., & Xu, X. (2007). A survey of open source data mining systems. *International conference on Emerging technologies in knowledge discovery and data mining*.
- Chickowski, E. (25 de 08 de 2009). *smartertechnology*.
<http://www.smartertechnology.com/c/a/Technology-For-Change/Six-Steps-to-Agile-BI/1/>.
 Recuperado el 2010
- Cruz, A. (2010). PyMEs poco competitivas. *www.pymempresario.com*.
- Del Alcazar, A. (2010). Como hacer mas competitiva su empresa.

- Feng, Y., Li, X.-x., Gao, C., Liu, Y., & Xu, H.-y. (2010). Design of the Low-cost Business Intelligence System Based on Multi-agent. *International Conference of Information Science and Management Engineering*.
- Fu, R., Xu, C., Song, M., & Xin, Z. (2008). Research on the Organizational Decision Support System for Small & Medium-Size Enterprise Based on Agent. *International Symposiums on Information Processing*.
- Gameiro, C. (2011). Implementation of Business Intelligence tools using Open Source Approach. *Workshop on Open Source and Design of Communication*.
- Gonzalez, T. (2006). Dashboard Design: Key Performance Indicators and Metrics. Choosing the right data to display. <http://www.brightpointinc.com/Articles.asp?File=Dashboard%20Design%20Metrics%20and%20KPIs.htm>. Obtenido de <http://www.brightpointinc.com/Articles.asp?File=Dashboard%20Design%20Metrics%20and%20KPIs.htm>
- Grabova, O., Darmont, J., Chauchat, J.-H., & Zolotaryova, I. (2010). Business Intelligence for Small and Middle-Sized Entreprises. *ACM SIGMOD Record*, 39.
- Infante, L. (2009). Business Intelligence Latin America. <http://www.bi-la.com/profiles/blogs/metodologia-agil-introduccion>.
- Inmon, H. (2002). *Building the DataWarehouse*. Wiley.
- Kimball, R., & Ross, M. (2002). *The Data Warehouse Toolkit*. Wiley.
- Kumar Kar, A., Kumar, A., & Kumar De, S. (2010). A Study On Using Business Intelligence For Improving Marketing Efforts. *Business Intelligence Journal*.
- Laumer, S., Maier, C., Weitzel, T., & Eckhardt, A. (2012). The Implementation of Large-scale Information Systems in Small and Medium-Sized Enterprises – A Case Study of Work-and Health-related Consequences. *45th Hawaii International Conference on System Sciences*.

- Lee, C., Ko, I., & Jeong, C. (2009). Evaluating the Effectiveness of Information Service for SMEs on Information Orientation and Firm Performance. *42nd Hawaii International Conference on System Sciences*.
- Li, H., Li, X., & Zhu, Z. (2010). Knowledge Mining for Intelligent Decision Making in Small and Middle Business. *Third International Symposium on Intelligent Information Technology and Security Informatics*.
- Li, X., Zhu, Z., & Pan, X. (2010). Knowledge cultivating for intelligent decision making in small & middle businesses. *International Conference on Computational Science*.
- Luis Paulo Vieira Braga, L. I. (2009). *Introducción a la Minería de Datos*. Editora E-papers.
- Parmenter, D. (2010). *Key Performance Indicators Developing, Implementing, and Using Winning KPIs*. Wiley. Recuperado el 2010, de Gestipolis.
- Pesaran Behbahani, M. (2012). A Business Intelligence Framework to Provide Performance Management Through a Holistic Data Mining View. *School of Computing and Information Systems*.
- Sadok, M., & Lesca, H. (2009). A Business Intelligence Model for SMEs Based on Tacit Knowledge. 7.
- Scholz, P., Schieder, C., Kurze, C., Gkuchowski, P., & Boehringer, M. (2010). Benefits and Challenges of Business Intelligence Adoption in Small And Medium-Sized Enterprises. *18th European Conference on Information Systems*.
- Vercellis, C. (2009). *Business Intelligence Data Mining and Optimization for Decision Making*. Wiley.
- Watson, H. J., & Wixom, B. H. (2007). The Current State of Business Inetelligence. *IT SYSTEMS PERSPECTIVES*.
- Williamw, N., & Williams, S. (2003). The Business Value of Business Intelligence. *Business Intelligence Journal*.

- Xie, H. (2009). The Research on Knowledge Management of Small and Medium-Sized IT Enterprises Based on Data Mining. *International Conference on Information Management, Innovation Management and Industrial Engineering*.
- Yeoh, W., & Koronios, A. (2009). Critical Success Factors for Business Intelligence Systems. *Journal of Computer Information Systems*.
- Zhao, Y., & Yao, Y. (2012). Bridging Theories and Practices: An Evaluation Framework of Small and Medium Sized Enterprises' (SMEs) Decision-making in China. *Fifth International Joint Conference on Computational Sciences and Optimization*.