

La Inteligencia de Negocios: Etapas del proceso

Vásquez Castrillón John Bayron, Sucerquia Osorio Andrés

Maestría en ingeniería de sistemas, Universidad Tecnológica de Pereira

Pereira, Colombia 2011

bayron733@utp.edu.co

sucerquia@utp.edu.co

Resumen

La explotación y el aprovechamiento del conocimiento generado en las organizaciones se convierten en la ventaja competitiva, factor diferenciador y clave del éxito.

Este artículo presenta las tecnologías y herramientas que conforman el proceso de inteligencia de negocios y sus etapas de extracción, consolidación, explotación y visualización. También se analiza la situación actual del mercado de proveedores de soluciones de inteligencia de negocios y las tendencias futuras.

I. INTRODUCCIÓN

Las organizaciones producen grandes volúmenes de información que por sí sola no agregan valor al negocio, se encuentran en diferentes fuentes de datos, no hay consistencia entre ellos, no se puede acceder a información en línea, la toma de decisiones se basa en información desactualizada. Como respuesta a estas necesidades de información surge la inteligencia de negocios como un proceso que permite aprovechar al máximo la información, transformándola en conocimiento que ayuda a la toma de mejores decisiones gerenciales.

II. DESARROLLO DE CONTENIDOS

¿Qué es Business Intelligence?

Business Intelligence (BI) o inteligencia de negocios se define como la habilidad corporativa para tomar decisiones. Esto se logra mediante el uso de metodologías, aplicaciones y tecnologías que permiten reunir, depurar, transformar datos, y aplicar en ellos técnicas analíticas de extracción de conocimiento (Parr, 2000).

BI se puede definir como el uso de los datos recopilados con el fin de generar mejores decisiones de negocio, esto implica accesibilidad, análisis y revelar nuevas oportunidades (Almeida, Ishikawa, Reinschmidt, & Roeber, 1999).

BI es el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada en información estructurada, para su explotación directa o para su análisis y conversión en conocimiento, dando así soporte a la toma de decisiones sobre el negocio (Sinnexus, 2007).

BI se puede ver como el proceso en el que intervienen personas y sistemas con la meta de obtener, recopilar, analizar y presentar la información que soporte de mejor forma la toma de decisiones de negocio. El proceso se puede dividir en 4 etapas: *extracción, consolidación, explotación y visualización* (Dávila, 2006).

¿Qué datawarehouse?

Muchos expertos en el tema de la inteligencia de negocios han realizado diversas definiciones acerca de lo que es una bodega de datos o datawarehouse, quizá las más famosas sean la de Bill Inmon:

“Datawarehouse es una colección de datos orientados al tema, integrados no volátiles e históricos cuyo objetivo es servir de apoyo en el proceso de toma de decisiones gerenciales” (Inmon,1996).

y la de Ralph Kimbal:

“La Bodega de Datos es un colección de datos en forma de una base de datos que guarda y ordena información que se extrae directamente de los sistemas operacionales (ventas, producción, finanzas, marketing, etc.) y de datos externos”.(Kimball,2002).

Para nuestro enfoque definimos data warehouse como un gran repositorio que integra diferentes fuentes de datos (Bases de datos de producción OLTP, archivos de Excel, archivos planos etc.), con el fin de que con su posterior análisis y procesamiento, esta pueda ser usada para la toma de decisiones a nivel gerencial.

Así como lo define Bill Inmon, el data warehouse debe caracterizarse por ser:

- **Integrado:** En el data warehouse la información debe estar estructurada consistentemente.
- **Temático:** Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales.
- **Histórico:** El tiempo es parte importante de la información contenida en los almacenes de datos, con esto se pueden realizar tendencias.
- **No volátil:** El data warehouse existe para ser leído, pero no modificado.

para el género irá en la bodega de datos si F o M o si 1 o 2.

I. Etapa de extracción

El proceso de implementación de un sistema de inteligencia de negocios en una organización debe iniciar por seleccionar la información relevante para la toma de decisiones, esto requiere contar con la participación de personal en los niveles operativo, táctico y gerencial.

Una vez identificada la información relevante se pasa a la etapa de consolidación donde se realiza el proceso de Extracción, Transformación y Carga ETL que según Kimball es un conjunto de procesos por medio de los cuales los datos de la fuente operacional son preparados para colocarse en el data warehouse (Kimball, 2002).

El ETL consiste en extraer los datos de la fuente de origen, transformarla, cargarla e indexarla, asegurando su integridad, coherencia y disponibilidad en el destino.



Fig. 1 Proceso de la información en la organización, (Dávila, 2006).

II. Etapa de consolidación

Esta etapa consiste de la recopilación de los datos de las diferentes fuentes, ya sean internas o externas de manera automática o semiautomática con el fin de normalizarlos, depurarlos y estructurarlos, almacenándolos en la bodega de datos, todo ello teniendo en cuenta que se debió haber hecho un análisis exhaustivo de las necesidades de información de la organización (etapa previa).

Uno de los objetivos de esta etapa es lograr un consenso entre las diferentes fuentes de datos, por ejemplo si tengo una base de datos de producción OLTP, allí se almacena la información de los empleados y el género está representado como M para Masculino y F para Femenino, además tengo información de estudiantes en un archivo de Excel en donde el género está representado 1 para Masculino y 2 para Femenino, la idea de normalizar es tomar la decisión de qué tipo de representación

En esta etapa se requiere una combinación de metodologías, técnicas, hardware y los componentes de software que proporcionan en conjunto la infraestructura para soportar el proceso de información.

Una metodología muy usada es la creación de Datamarts.

Kimball define un **Datamart** como el subconjunto lógico y físico del área de presentación de datos en un Data Warehouse. Originalmente, los datamart fueron definidos como un subconjunto altamente agregado de datos, normalmente usados para resolver preguntas específicas del negocio.

La principal diferencia entre las bodegas de datos y las bases de datos transaccionales OLTP radica en que las primeras están optimizadas para realizar consultas y las segundas para hacer inserciones y modificaciones de los datos, esto sugiere un cambio en el modelado lógico y físico de la información.

Después de finalizada la consolidación de la información pasamos a la etapa de explotación.

III. Etapa de explotación

En ésta etapa es donde se comienzan a aplicar las herramientas existentes para dejar listos los datos de la bodega en manos de los usuarios, quienes deben estar en capacidad de empezar a aprovechar y explotar la información ya depurada y filtrada que hay en la bodega de datos. En esta etapa tenemos dos tecnologías que nos permiten realizar un proceso de explotación de los datos.

La primera son los cubos OLAP, los cuales los podemos definir como una tecnología de bases de datos que maneja más de dos dimensiones y que permite ver desde diferentes vistas los datos almacenados en la bodega de datos.

El término OLAP fue presentado en 1993 (Codd & Salley, 1993). Según la definición que le dio Codd, OLAP es un tipo de procesamiento de datos que se caracteriza, entre otras cosas, por permitir el análisis multidimensional (Kimball, et al., 1998). Dicho análisis consiste en modelar la información en medidas, dimensiones y hechos.

Las medidas son los valores de un dato, en particular, las dimensiones son las descripciones de las características que definen dicho dato y los hechos corresponden a la existencia de valores específicos de una o más medidas para una combinación particular de dimensiones. Este modelo se representa vectorialmente: los hechos se ubican lógicamente en una celda que queda en la intersección de ciertas coordenadas según el modelo (x, y, z, ...), donde cada una de las coordenadas de la celda representa una dimensión (Tory &

Moller, 1998).

Para materializar el análisis multidimensional en una base de datos se usa la correspondencia entre los elementos del modelo (hechos y coordenadas) y los de la base (tabla de hechos y dimensiones). Como la tabla de hechos y las dimensiones se implementan en tablas en una base de datos, se puede usar el lenguaje SQL (ISO, 1992) para la definición de las tablas de un modelo multidimensional en una base de datos relacional, pero se hizo necesario extender el modelo relacional con el fin de dar soporte a las funcionalidades propias de análisis multidimensional (Agrawal, Gupta y Sarawagi, 1997).

Tales funcionalidades incluyen:

1. La declaración de dimensiones y jerarquías, ya que en el modelo relacional no se manejaban dichos conceptos;
2. Un acceso más rápido a los datos, a través de métodos de generación de índices para datos espaciales desde el punto de vista multidimensional (Papadias, et al., 2002).
3. El cálculo de valores preagregados que optimicen las consultas.
4. La definición de operaciones de navegación en las dimensiones y de agrupación de medidas (Bédard et al., 2000).

La segunda tecnología que permite explotar y descubrir información oculta, no evidente, es la minería de datos o de manera más generaliza proceso de descubrimiento de conocimiento KDD.

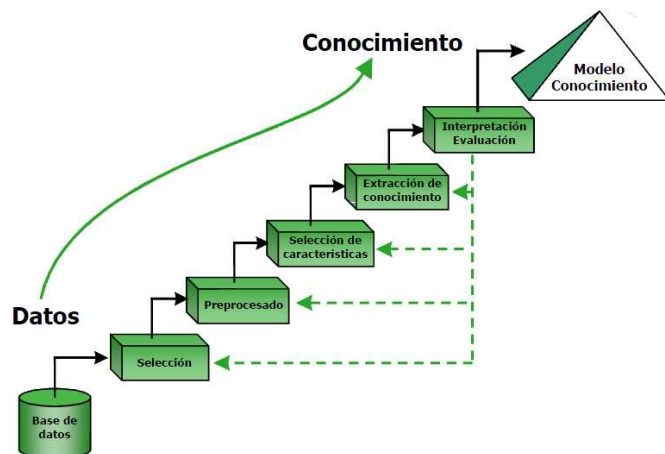


Fig. 1 Proceso de Minería de Datos. Tomada de Villena, J. 2009. Inteligencia en Redes de Comunicaciones: Minería de Datos

La minería de datos consiste en la aplicación de un conjunto de métodos, para el procesamiento y análisis de datos. Para esto se basa en dos conceptos: “escarbar y explotar”. Así, grandes volúmenes de datos son tratados mediante diversos procesos para permitir el descubrimiento de información no evidente, elementos de utilidad y comportamientos interesantes como: cambios, anomalías, estructuras

significativas y patrones de comportamiento para aplicarlos a nuevos conjuntos de datos. El objetivo primordial de la minería de datos es el aprovechamiento de las características hombre-máquina, es decir, la mezcla entre flexibilidad, creatividad y conocimiento general con la potencia de cálculo y almacenamiento. Esto para poder realizar una exploración de datos efectiva y veraz, con el fin de construir un sistema computacional que sea capaz de extraer y modelar el conocimiento no visible.

Las técnicas de minería de datos provienen de la inteligencia artificial y de la estadística. Dichas técnicas son algoritmos más o menos sofisticados que se aplican sobre un conjunto de datos para obtener un resultado, dichos algoritmos pueden clasificarse según su finalidad en:

- Regresión
- Clasificación
- Agrupamiento

Y se usan algoritmos de redes neuronales, algoritmos genéticos, arboles de decisión, maquinas de soporte vectorial, series de tiempo, cadenas de markov, simulación de montercarlo, redes bayesianas y análisis de varianza, análisis discriminante, etc.

Algunos de los algoritmos requieren supervisión humana para obtener resultados confiables.

Una vez explotados los datos esta listos para ser presentados.

IV. Etapa de visualización

Después de realizar la explotación, continua una etapa de visualización, donde los usuarios a través de ciertas herramientas gráficas pueden conocer de primera mano lo que está sucediendo en la organización, esta etapa involucra las siguientes metodologías y/o herramientas: Balance Score Card, Sistemas de Soporte a la Decisión (DSS), Sistemas de Información Ejecutiva (EIS).

Balance Score Card

El Cuadro de Mando Integral (CMI, o Balanced Scorecard), desarrollado por Kaplan y Norton(), es un instrumento útil para la planificación estratégica de las empresas. Pretende evitar que los objetivos de la dirección se centren sólo en el corto plazo y los rendimientos actuales para introducir una perspectiva más integral que incluye a los accionistas, los clientes, los procesos clave y el crecimiento y el desarrollo de la organización.

También se puede considerar como una aplicación que ayuda a una compañía a expresar los objetivos e iniciativas necesarias para cumplir con su estrategia, mostrando de forma continuada cuándo la empresa y los empleados alcanzan los resultados definidos en su plan estratégico. (Sinnexus, 2007).

Sistemas de Soporte a la Decisión (DSS)

Un Sistema de Soporte a la Decisión (DSS) es una herramienta de Business Intelligence enfocada al análisis de los datos de una organización.

En principio, puede parecer que el análisis de datos es un proceso sencillo, y fácil de conseguir mediante una aplicación hecha a medida o un ERP sofisticado. Sin embargo, no es así: estas aplicaciones suelen disponer de una serie de informes predefinidos en los que presentan la información de manera estática, pero no permiten profundizar en los datos, navegar entre ellos, manejarlos desde distintas perspectivas.. etc. (Sinnexus, 2007).

Los DSS permiten a los usuarios generar informes dinámicos, flexibles sin necesidad de recurrir al departamento de sistemas de la organización y sin ser un experto técnico, pudiendo tener a mano la información histórica necesaria para realizar su trabajo.

Sistemas de Información Ejecutiva (EIS)

EIS como una aplicación informática que muestra informes y listados (query & reporting) de las diferentes áreas de negocio, de forma consolidada, para facilitar la monitorización de la empresa o de una unidad de la misma.

El EIS se caracteriza por ofrecer al ejecutivo un acceso rápido y efectivo a la información compartida, utilizando interfaces gráficas visuales e intuitivas. Suele incluir alertas e informes basados en excepción, así como históricos y análisis de tendencias. También es frecuente que permita la domiciliación por correo de los informes más relevantes. (Sinnexus, 2007).

El EIS permiten al usuario final tener una herramienta para tener un panorama del negocio, con la cual pueda consultar indicadores y particularizar aquellos en los que se desee, se puede pasar de lo general a lo particular y viceversa.

Un objetivo muy importante en esta etapa es darle al usuario la posibilidad de personalizar la visualización de los informes y consultas de manera transparente y sobre todo fácil.

En este campo se requiere mayor investigación, ya que las herramientas actuales poseen una experiencia de usuario baja.

V. Análisis del mercado

Las principales empresas del sector Business Intelligence son:

- Microsoft
- Oracle
- Microstrategy
- IBM Cognos
- Information Builders
- SAS
- QlikTech
- SAP Business Objects
- Teradata
- Hyperion
- Tableau
- Netezza
- Pentaho



Fig. 2 El Cuadrante mágico de Gartner para las plataformas BI. As of January 2011

Según el informe de enero de 2011 de Gartner¹ sobre las plataformas Business Intelligence, las suite de Business Intelligence de los proveedores líderes son complejas, caras, y presentan importantes deficiencias en cuanto a rendimiento y usabilidad.

Gartner reconoce que los “líderes” no salen muy bien parados en ninguna de las encuestas. Ni SAP, ni IBM, ni Oracle, ni casi nadie puede estar contento con estos resultados.

Por ejemplo, según Gartner, los clientes de SAP Business

¹ www.gartner.com/

Objects afirman que las migraciones, implantaciones e integraciones son complejas, que el rendimiento y la dificultad de despliegue son problemas detectados por gran parte de sus clientes (el 40%), y que la experiencia de usuario y la calidad del soporte es de los peores de toda la comparativa.

De Oracle, por poner otro ejemplo, se destaca su falta de innovación, y sus larguísimos ciclos de desarrollo. En la versión 11g, por citar algún caso, las funcionalidades in-memory o las visualizaciones interactivas son escasas o inexistentes.

El caso de IBM Cognos es incluso más llamativo. El 33% de los usuarios afirman tener problemas de rendimiento, resultado que se viene repitiendo desde hace años. Sus propios clientes afirman que la implantación de Cognos es aún más compleja que la de sus competidores, y que tienen importantes deficiencias de usabilidad por parte del usuario final. El 20% de los clientes de IBM se plantea discontinuar su uso en menos de 5 años. Han pasado 3 años desde la adquisición de Cognos por parte de IBM, y el soporte ofrecido sigue siendo de los peores entre todos los fabricantes de la comparativa.

De Microstrategy se dice que tiene una curva de aprendizaje muy pronunciada, incluso para desarrolladores experimentados. Los usuarios carecen de capacidades de reporting ad-hoc (salvo los tristes informes parametrizables...). A pesar de disponer de versión "gratuita", los clientes de Microstrategy afirman que la mayoría de proyectos de Microstrategy son muy caros y complejos. (Business Intelligence fácil, 2011)

IV. CONCLUSIONES

- Las empresas de hoy necesitan herramientas tecnológicas que les permita tomar decisiones de negocio precisas y de forma rápida, ya que los sistemas transaccionales actuales tradicionales, suelen presentar una estructura muy inflexible para este fin.
- Las herramientas de BI no vienen a remplazar los sistemas transaccionales actuales, se trata de sistemas con objetivos distintos, eficientes en sus respectivas funciones, pero que deben complementarse para optimizar el valor de los sistemas de información.
- BI es una herramienta que se constituye en una ventaja tecnológica ya que con ella se logra centralizar, depurar y afianzar los datos, descubrir información no evidente para las aplicaciones actuales, optimizar el rendimiento de los sistemas.
- BI es una herramienta que se constituye en una ventaja competitiva ya que con ella se logra seguimiento real del plan estratégico, aprender de

errores pasados, mejorar la competitividad, obtener el verdadero valor de las aplicaciones de gestión.

- El mercado ofrece soluciones que son complejas, caras, y presentan importantes deficiencias en cuanto a rendimiento y usabilidad. Por esto se requiere que los proveedores sigan investigando y desarrollando tecnología que mejore la experiencia del usuario.

REFERENCIAS

- [1] Almeida, M. S., Ishikawa, M., Reinschmidt, J., & Roeber, T. (1999). *Getting Started with Data Warehouse and Business Intelligence*. Obtenido de <http://www.redbooks.ibm.com/redbooks/pdfs/sg245415.pdf>.
- [2] Davila, F. (2006). *LA INTELIGENCIA DEL NEGOCIO*. Obtenido de <http://sigma.poligran.edu.co/politecnico/apoyo/cuadernos/intelligence.pdf>.
- [3] Dayal, U. (2009). *Data Integration Flows for Business Intelligence*. Obtenido de <http://www.edbt.org/Proceedings/2009-StPetersburg/edbt/papers/p0001-Dayal.pdf>.
- [4] Parr, O. (2000). *Data mining cookbook*. Obtenido de <http://books.google.com.co/books?id=L3w0loZrcU0C&printsec=frontcover&dq=Data+Mining+Cookbook#v=onepage&q=&f=false>.
- [5] Sinnexus. (2007). *Business Intelligence*. Obtenido de http://www.sinnexus.com/business_intelligence/index.aspx.
- [6] Inmon, W. H (1996). *Building the data warehouse* Jhon Wiley & Sons
- [7] Kimball Ralph & Ross Margy (2002). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. Wiley,. Obtenido: <http://www.sigmod.org/publications/sigmod-record/0309/R17.AnisimovBookReview.pdf>
- [8] Codd, E. F., Codd, S. B. and Salley, C. T.(1993). Providing OLAP to user-analysts: An IT mandate., E. F. Codd and Associates
- [9] Kimball, R., Reeves, L., Ross, M. and Thornthwaite, W.(1998). *The Data Warehouse Lifecycle Toolkit: expert methods for designing, developing and deploying data warehouses.*, John Wiley & Sons, Inc.
- [10] Tory, M. and Moller, T.(1998). *A model-based visualization taxonomy.*, School of computing science, Simon Fraser University, Burnaby, Canada.
- [11] ISO (1992). International Standard Organization., *ISO/IEC9075/1992 Information technology ? database languages- Structured Query Language (SQL)* Geneva, Switzerland.
- [12] Agrawal, R., Gupta, A. and Sarawagi, S (1997). *Modeling Multidimensional Databases.*, In: ICDE 97: Proceedings of the 13th International Conference on Data Engineering, IEEE Computer Society.
- [13] Papadias, D, Tao, Y., Kalnis, P. and Zhang, J.(2002). *Indexing Spatio-Temporal Data Warehouses.*, In: Proceedings. 18th International Conference on Data Engineering. IEEE Computer Society.
- [14] Bédard, Y., Merrett, T. and Han, J.(2000). *Fundamentals of spatial data warehousing for geographic knowledge discovery.*, In: H. Miller and J. Han (Editors), *Geographic data mining and knowledge discovery*.
- [15] *Business Intelligence fácil.* (s.f.). La cruda realidad del Business Intelligence. Obtenido de <http://www.businessintelligence.info/mercado/gartner-business-intelligence-2011.html>