

# Introducción a la *business intelligence*

Josep Curto Díaz

PID\_00238512



# Índice

<b>Introducción</b> .....	5
<b>1. ¿Qué es la inteligencia de negocio?</b> .....	7
1.1. Diferencias entre <i>business intelligence</i> , <i>business analytics</i> y <i>big data</i> .....	9
1.2. Beneficios de un sistema de <i>business intelligence</i> .....	11
1.3. ¿Cuándo es necesaria la inteligencia de negocio? .....	12
<b>2. Estrategia de <i>business intelligence</i></b> .....	14
2.1. ¿Cómo detectar que no existe una estrategia? .....	14
<b>3. Soluciones <i>open source business intelligence</i></b> .....	18
3.1. Pentaho .....	19
<b>Abreviaturas</b> .....	21
<b>Bibliografía</b> .....	22



## Introducción

La gestión de una organización se fundamenta en tomar decisiones adecuadas respecto a clientes, productos, empleados, proveedores y procesos de negocio. Por lo tanto, es necesario tener mecanismos que den soporte a una toma de decisiones eficiente.

En los últimos años, ha emergido una nueva forma de competir que se fundamenta en tomar decisiones basadas en datos y evidencias dejando atrás la intuición. Esta forma de competir combina diferentes estrategias para generar valor de negocio: *business intelligence* (BI), *business analytics* y *big data*. No extraño que los CIO de las principales empresas del mundo destaquen que su principal prioridad tecnológica son este tipo de iniciativas.

Así, la explotación de la información en el contexto de las organizaciones ha pasado de ser una necesidad a una prioridad de máxima relevancia. El objetivo es poder tomar mejores y más rápidas decisiones informadas de negocio.

Aunque vivimos en la era de los datos complejos, conocida como *big data*, y la progresiva automatización de nuestras capacidades predictivas, la inteligencia de negocio sigue siendo relevante. Muchas organizaciones aún no han desplegado este tipo de iniciativas y entre aquellas que lo han hecho no todas han logrado alcanzar el éxito esperado. No en vano, este tipo de iniciativas son complejas puesto que suponen una transformación de gran calado en la organización. Por ello, en esta asignatura nos centraremos tan solo en la inteligencia de negocio.

El mercado de *business intelligence* existe desde hace bastantes años; ha evolucionado hacia soluciones con mayores prestaciones y podemos considerar que ha alcanzado una significativa madurez. Destacamos, por ejemplo, que:

- Se ha producido una consolidación mediante la compra de empresas pequeñas por parte de los principales agentes del mercado (entre ellas SAP, IBM, Microsoft u Oracle).
- Se ha enriquecido con soluciones *open source* que cubren el espectro de necesidades de una organización para la explotación de la información.
- Han aparecido nuevas empresas con foco en la innovación cubriendo nuevos nichos en el mercado de la inteligencia de negocio, como la visualización, el análisis predictivo, las *virtual appliances* y/o el *real-time business intelligence* (entre ellas Tableau, Qlikview o Yellowfin).

### CIO

Son las siglas de *chief information officer*, que equivalen a director de sistemas de una organización.

- Las principales soluciones *BI open source*, a saber Pentaho, JasperSoft y Actuate, han sido adquiridas por Hitachi Data Systems, Tibco y OpenText, respectivamente.
- Hemos asistido a la aparición de una nueva generación de soluciones enfocadas a la generación de valor de conjuntos de datos complejos, lo que frecuentemente se refiere como *big data*, que expande el valor de la inteligencia de negocio.

Esta asignatura se centrará en introducir los diferentes conceptos que engloba la inteligencia de negocio. Se discutirán las diferencias entre *business intelligence*, *business analytics* y *big data*. Otro de los objetivos es ejemplificar el desarrollo de un proyecto de *business intelligence* mediante herramientas *open source* para facilitar la comprensión de los conceptos presentados en la asignatura.

Algunas de las herramientas en el contexto de la inteligencia de negocio acumulan varios años de desarrollo y evolución, y están respaldadas por organizaciones que tienen un claro modelo de negocio y que generan sinergias entre ellas en forma de ecosistemas. Podemos encontrar tanto herramientas de bases de datos como de minería de datos. Tal es la madurez de dichas soluciones, que es posible desarrollar e implementar proyectos de inteligencia de negocio para todo tipo de organizaciones, tanto pymes como grandes organizaciones.

# 1. ¿Qué es la inteligencia de negocio?

El objetivo de este primer apartado es introducir la inteligencia de negocio y enumerar las tecnologías que engloba.

El contexto de la sociedad de la información ha propiciado la necesidad de tener mejores, más rápidos y más eficientes métodos para extraer y transformar los datos de una organización en información y distribuirla a lo largo de la cadena de valor.

## La cadena de valor

Descrita y popularizada por Michael E. Porter en su obra *Competitive Advantage: Creating and sustaining superior performance*, la cadena de valor empresarial es un modelo teórico que permite describir las actividades que generan valor en una organización.

La inteligencia de negocio (*business intelligence*; BI) responde a dicha necesidad y podemos entender, en una primera aproximación, que es una evolución de los sistemas de soporte a las decisiones (*decision support systems*; DSS). Sin embargo, este concepto, que actualmente se considera crítico en la gran mayoría de las empresas, no es nuevo. En octubre de 1958, Hans Peter Luhn, investigador de IBM, acuñó el término en el artículo *A Business Intelligence System*:

«La habilidad de aprehender las relaciones de hechos presentados de forma que guían las acciones hacia una meta deseada».

No es hasta 1989 cuando Howard Dresden, en dicho momento analista de Gartner, propone una definición formal del concepto:

«Conceptos y métodos para mejorar las decisiones de negocio mediante el uso de sistemas de soporte basados en hechos».

Desde entonces, el concepto del que estamos hablando ha evolucionado aunando diferentes tecnologías, metodologías y términos bajo su paraguas. Es, por tanto, necesario establecer una definición formal de uso en el presente material.

Se entiende por ***business intelligence*** el conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización.

En esencia, mediante la inteligencia de negocio, podemos romper con la siguiente máxima:

### Lectura recomendada

M. Castells (1997-1998). *La era de la información: economía, sociedad y cultura* (3 vols.). Madrid: Alianza Editorial.

### Gartner

Es una de las principales empresas de prospección de mercado.

«Lo que no se define no se puede medir. Lo que no se mide, no se puede mejorar. Lo que no se mejora, se degrada siempre».

William Thomson

Algunas de las tecnologías que forman parte de la *business intelligence* son estas:

- *Data warehouse*.
- *Reporting*.
- Análisis OLAP (*online analytical processing*).
- Análisis visual.
- Análisis predictivo.
- Cuadro de mando.
- Cuadro de mando integral.
- Minería de datos.
- Gestión del rendimiento.
- Previsiones.
- Reglas de negocio.
- *Dashboards*.
- Integración de datos (que incluye *extract, transform and load*; ETL).

A lo largo de la asignatura entraremos en detalle en algunas de ellas para tener claro los componentes mínimos que deben tener este tipo de sistemas.

El siguiente diagrama, creado por Forrester Research, ilustra las diferentes tecnologías y aspectos que forman parte de la inteligencia de negocio:

**Forrester Research**

Se trata de una de las principales empresas de prospección de mercado.



Figura 1. Conceptos y tecnologías en *business intelligence*

	Desktop gadgets	Office suites	Mobile	Disconnected	Apps Form factors PSO	Industry vertical applications	Appliance	Strategy	Methodology	
Delivery	Portals		Interactive voice response, ATM, point-of-sale							
Reporting	Dashboards		Alerts	Advanced data visualization						
	Search	Geospatial	Reporting – ad hoc, analytical, production							
Performance management	Metrics/KPIs		Planning	Scorecards						
	Strategy/objectives management									
Supporting applications	Collaboration	Life-cycle mgt.	Localization	QA						Version control
	Metadata – integration, repositories			ECM						eLearning
Analytics	Data/text mining		Guided decisions	NLP						Guided search
	Time series	OLAP	Operational DSS	Predictive analytics						
	Usage analytics	Statistical analysis		Web analytics						
Discovery and integration	Accelerators/query optimization		Adapters/tool kits							
	BAM/CEP	BPM/BRE integration		Discovery accelerators						
	DQ – cleansing, profiling		EAI/SOA	EII	ETL/CDC					
	Integration – third party applications									
	Operational data scores (ODS), data warehouse (DW), data marts (DM)									
Data	Report mining		Services registry and repository							
	Columnar DBMS		Hierarchical/XML	In-memory DBMS						
	Multidimensional OLAP		Multivalued DBMS	RDBMS						
	Streaming DBMS		Search DBMS							
	Network		Servers	Storage						
Infrastructure					Enterprise applications: ERP, CRM, SCM, ERM	BI SaaS	MSP/application outsourcing	Governance	BPO	

Fuente: Forrester Research (2008).

### 1.1. Diferencias entre *business intelligence*, *business analytics* y *big data*

Acabamos de definir lo que es la inteligencia de negocio. Para poder entender las diferencias entre las tres estrategias de explotación a las que hemos hecho referencia, necesitamos definir las también.

Primero definamos qué es *business analytics* (BA) o analítica de negocio.

Se entiende por *business analytics* el conjunto de estrategias, tecnologías y sistemas que permiten analizar el rendimiento pasado de una organización para poder predecir comportamientos futuros, así como para detectar patrones ocultos en la información.

Ahora definamos que es *big data*.

Se entiende por *big data* el conjunto de estrategias, tecnologías y sistemas para el almacenamiento, procesamiento, análisis y visualización de conjuntos de datos complejos, que frecuentemente, pero no siempre, viene definida por **volumen**, **velocidad** y **variedad**.

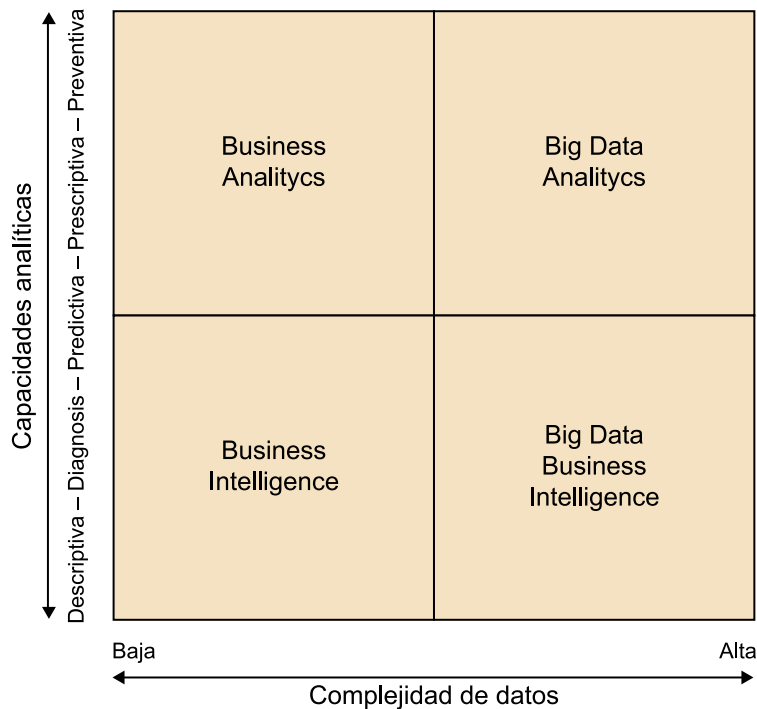
Vamos a comparar estas estrategias respecto a diferentes factores: **herramientas, foco, uso, tipo del dato, complejidad del dato y alcance**. Además, indicamos su nivel de madurez en el mercado. La siguiente tabla describe las diferencias entre estas estrategias:

Tabla 1. Diferencias entre *business intelligence*, *business analytics* y *big data*

	<b>BI (madura)</b>	<b>BA (madura)</b>	<b>Big data (emergente)</b>
<b>Herramientas</b>	Consultas, alertas, <i>reporting</i> , OLAP, etc.	Clasificación, <i>clustering</i> , regresión	<i>Machine learning</i> , visualización, etc.
<b>Foco</b>	Qué y cómo pasó, cuántos, con qué frecuencia, cuál es el problema, qué es necesario hacer	Por qué está pasando, qué pasaría si todo continúa igual, qué pasará a continuación, qué es lo mejor que puede pasar	Capturar, almacenar, procesar, analizar
<b>Uso</b>	Reactivo	Proactivo Predictivo Prescriptivo	Todos
<b>Tipo de datos</b>	Estructurados	Estructurados Semiestructurados	Todo tipo
<b>Complejidad del dato</b>	Baja	Baja/media	Alta
<b>Alcance</b>	Dirección	Procesos	Vertical/procesos

Aunque queda claro que estas estrategias son diferentes, lo normal es que se combinen en los proyectos de explotación del dato. La siguiente gráfica permite identificar casos de uso respecto a la complejidad del dato y las capacidades analíticas que desarrollar en la organización.

Figura 2. Combinación de estrategias



Fuente: Josep Curto.

La lectura de este gráfico se realiza a partir de sus ejes. Por ejemplo, cuando necesitemos desarrollar capacidades analíticas descriptivas y la complejidad de dato sea alta combinaremos *big data* con *business intelligence* formando un único sistema.

## 1.2. Beneficios de un sistema de *business intelligence*

La implantación de estos sistemas de información proporciona diversos beneficios, entre los que podemos destacar:

- Crear un círculo virtuoso de la información: los datos se transforman en información que genera un conocimiento que permite tomar mejores decisiones que se traducen en mejores resultados y que generan nuevos datos.
- Proporcionar una visión única, conformada, histórica, persistente y de calidad de toda la información relevante para la organización.
- Crear, manejar y mantener métricas, indicadores claves de rendimiento (*key performance indicator*; KPI) e indicadores claves de metas (*key goal indicator*; KGI) fundamentales para la empresa.
- Habilitar el acceso a información actualizada tanto a nivel agregado como en detalle.

- Reducir el diferencial de orientación de negocio entre el departamento TI y la organización.
- Mejor comprensión y documentación de los sistemas de información en el contexto de una organización.
- Mejor comprensión de las opiniones y necesidades de los potenciales clientes a partir del análisis de redes sociales.
- Mejora de la competitividad de la organización como resultado de ser capaces de:
  - Diferenciar lo relevante de lo superfluo.
  - Acceder más rápido a información.
  - Tener mayor agilidad en la toma de las decisiones.

### 1.3. ¿Cuándo es necesaria la inteligencia de negocio?

Existen situaciones en las que la implantación de un sistema de *business intelligence* resulta adecuada. Destacamos, entre todas las que existen:

- La toma de decisiones se realiza de manera intuitiva en la organización.
- Identificación de problemas de calidad de información.
- Uso de Excel como repositorios de información corporativos o de usuario. Lo que se conoce como Excel caos.
- Necesidad de cruzar información de manera ágil entre departamentos.
- Evitar silos de información.
- Las campañas de marketing no son efectivas por la información base usada.
- Existe demasiada información en la organización para ser analizada de la manera habitual. Se ha alcanzado la masa crítica de datos.
- Es necesario automatizar los procesos de extracción y distribución de información.

#### Excel caos

Se entiende por Excel caos el problema resultante del uso intensivo de Excel como herramienta de análisis. Cada usuario trabaja con un fichero personalizado. Como resultado, la información no cuadra entre departamentos y el coste de sincronización es sumamente elevado.

En definitiva, los sistemas de *business intelligence* buscan responder a las preguntas:

- ¿Qué pasó?
- ¿Qué pasa ahora?
- ¿Por qué pasó?

- ¿Qué pasará?

Tal y como explica Thomas Davenport en su libro *Competing on Analytics* (2007), una nueva forma de estrategia competitiva está emergiendo basada en el uso de la estadística descriptiva, modelos productivos y complejas técnicas de optimización, datos de alta calidad y una toma de decisiones basada en hechos. En dicho contexto, la inteligencia de negocio es el paso previo para dicha estrategia dado que ayuda a sentar las bases para su futuro despliegue.

## 2. Estrategia de *business intelligence*

Desplegar un proyecto de inteligencia de negocio en el seno de una organización no es un proceso sencillo. Las buenas prácticas indican que, para llegar a buen puerto, es necesario tener una estrategia de inteligencia de negocio que coordine las tecnologías, el uso, los procesos de madurez, la metodología que emplear, etc.

### 2.1. ¿Cómo detectar que no existe una estrategia?

Es posible detectar que no existe una estrategia definida a través de los siguientes puntos y percepciones en el seno de una organización:

- Los usuarios identifican el departamento de informática (*information technology*; IT) como el origen de sus problemas de inteligencia de negocio.
- La dirección considera que la inteligencia de negocio es otro centro de coste.
- El departamento de IT continúa preguntando a los usuarios finales sobre las necesidades de los informes.
- El sistema de BI está soportado por *help desk*.
- No hay diferencia entre BI y gestión del rendimiento.
- No es posible medir el uso del sistema de inteligencia de negocio.
- No es posible medir el retorno de la inversión (*return on invest*; ROI) del proyecto de *business intelligence*.
- Se considera que la estrategia para el *data warehouse* es la misma que para que el sistema de inteligencia de negocio.
- No hay un plan para desarrollar, contratar, retener y crecer el equipo de BI.
- No se sabe si la empresa tiene una estrategia para el BI.
- No existe un responsable funcional (o bien el asignado no es el adecuado).
- No existe un centro de competencia.

- Existen múltiples soluciones en la organización distribuidas en diferentes departamentos que repiten funcionalidad.
- No hay un plan de formación real y consistente de uso de las herramientas.
- Alguien cree que es un éxito que la información consolidada esté a disposición de los usuarios finales al cabo de dos semanas.
- Los usuarios creen que la información del *data warehouse* no es correcta.
- No existe una cultura analítica en la que el dato y los hechos son relevantes para tomar decisiones sea cual sea el nivel de la organización.

El desarrollo de una estrategia de negocio es un proceso a largo plazo que incluye múltiples actividades, entre las que es conveniente destacar:

- Crear un centro de competencia (o de excelencia) de BI (BICC). Tiene el objetivo de aunar conocimiento en tecnologías, metodologías, estrategia, con la presencia de un patrocinador a nivel ejecutivo y con analistas de negocio implicados y que tenga responsabilidad compartida en éxitos y fracasos.
- Establecer los estándares de BI en la organización para racionalizar tanto las tecnologías existentes como las futuras adquisiciones.
- Identificar qué procesos de negocio necesitan diferentes aplicaciones analíticas que trabajen de manera continua para asegurar que no existen silos de funcionalidad.
- Desarrollar un *framework* de métricas a nivel empresarial como el pilar de una gestión del rendimiento a nivel corporativo.
- Incluir los resultados de aplicaciones analíticas (minería de datos u otras) en los procesos de negocio con el objetivo de añadir valor a todo tipo de decisiones.
- Revisar y evaluar el portafolio actual de soluciones en un contexto de riesgo / recompensas.
- Considerar inversiones tácticas cuyo retorno de inversión estén dentro de un periodo de tiempo de un año; además tener en cuenta los diferentes análisis de mercado, de soluciones e incluso el *hype cycle* de Gartner para conocer el estado del arte.

- Aprender de los éxitos y fracasos de otras empresas revisando casos de estudio y consultado a las empresas del sector para determinar qué ha funcionado y qué no.
- Evangelizar la organización.
- Alinear el departamento IT y el negocio en caso de no poder organizar un BICC, fundamental para trabajar como equipo integrado. El departamento de IT debe entender las necesidades y entregar la mejor solución ajustada a la necesidad particular y escalable a otras futuras.
- Poner atención a las necesidades que requieren BI en la organización porque se suele satisfacer a los usuarios o departamentos que gritan más fuerte, y esto no significa que den mayor valor a la compañía. Por ejemplo, los departamentos de finanzas son un caso típico de baja atención en soluciones BI.

### **El *hype cycle* de Gartner**

Es una representación gráfica de la madurez, adopción y aplicación de negocio de una o varias tecnologías específicas. Es decir, muestra el ciclo de vida de dichas tecnologías. Las etapas que componen el ciclo son las siguientes:

- Disparador tecnológico: cuando aparece el concepto en el mercado.
- Pico de expectativa inflada: cuando se habla mucho del concepto, pero está poco aplicado.
- Valle de la desilusión: cuando la herramienta está por debajo de lo que se esperaba de ella.
- La pendiente de tolerancia: el camino hacia la madurez.
- *Plateau* de productividad: cuando alcanza la madurez.

### ***Business intelligence maturity model***

Si bien el objetivo de esta asignatura no es dar pautas para definir una estrategia de *business intelligence*, sino una introducción de conceptos, un buen punto de partida es identificar cuál es el grado de madurez de la organización con respecto a la inteligencia de negocio.

El BIMM (*business intelligence maturity model*) es un modelo de madurez que permite clasificar nuestra organización desde el punto de vista del grado de madurez de implantación de sistemas *business intelligence* en ella (en relación directa con *frameworks* como COBIT).

1) **Fase 1: No existe BI.** Los datos se hallan en los sistemas de procesamiento de transacciones en línea (*online transaction processing*; OLTP), desperdigados en otros soportes o incluso solo contenidos en el *know how* de la organización.

#### **COBIT**

*Control objectives for information and related technology* (COBIT) es un conjunto de mejores prácticas para el manejo de información creado por la Information Systems Audit and Control Association (ISACA) y el IT Governance Institute (ITGI) en 1992.



Las decisiones se basan en la intuición, la experiencia, pero no en datos consistentes. El uso de datos corporativos en la toma de decisiones no ha sido detectado y tampoco el uso de una herramienta adecuada al respecto.

**2) Fase 2: No existe BI, pero los datos son accesibles.** No existe un procesado formal de los datos para la toma de decisiones, aunque algunos usuarios tienen acceso a información de calidad y son capaces de justificar decisiones con dicha información. Frecuentemente, este proceso se realiza mediante Excel o algún tipo de *reporting*. Se intuye que deben existir soluciones para mejorar este proceso pero se desconoce la existencia del *business intelligence*.

**3) Fase 3: Aparición de procesos formales de toma de decisiones basada en datos.** Se establece un equipo que controla los datos y que permite hacer informes contra estos, que permiten tomar decisiones fundamentadas. Los datos son extraídos directamente de los sistemas transaccionales sin *data cleansing*, ni modelización ni existe un *data warehouse*.

**4) Fase 4: *Data warehouse*.** El impacto negativo contra los sistemas OLTP lleva a la conclusión de que un repositorio de datos es necesario para la organización. Se percibe el *data warehouse* como una solución deseada. El *reporting* sigue siendo personal.

**5) Fase 5: El *data warehouse* crece y el *reporting* se formaliza.** El *data warehouse* funciona y se desea que todos se beneficien de él, de modo que el *reporting* corporativo se formaliza. Se habla de OLAP, pero solo algunos identifican realmente sus beneficios.

**6) Fase 6: Despliegue de OLAP.** Después de cierto tiempo, ni el *reporting* ni la forma de acceso al *data warehouse* es satisfactoria para responder a preguntas sofisticadas. OLAP se despliega para dichos perfiles. Las decisiones empiezan a impactar de manera significativa en los procesos de negocio a lo largo de la organización.

**7) Fase 7: El *business intelligence* se formaliza.** Aparece la necesidad de implantar otros procesos de inteligencia de negocio, como *data mining*, *balanced scorecard*, etc., y procesos de calidad de datos impactan en procesos como *customer relationship management* (CRM), *supply chain management* (SCM), etc. Se ha establecido una cultura corporativa que entiende claramente entre sistemas OLTP y DSS.

Existen otros modelos de madurez, como el modelo Delta analítico de Thomas Davenport o el de TDWI.

#### TDWI

Es el acrónimo de The Data Warehouse Institute, organización que se dedica a la investigación del *business intelligence*. En este enlace es posible autoevaluarse contra el modelo de madurez.

#### Data cleansing

Consiste en el proceso de detectar y mejorar (o borrar) registros incorrectos e incompletos con el objetivo de conseguir datos coherentes y consistentes.

#### Lectura complementaria

T. H. Davenport; J. G. Harris; R. Morison (2010). *Analytics at Work: Smarter Decisions, Better Results*. Nueva York: Harvard Business Press.

### 3. Soluciones *open source business intelligence*

El *open source* es una filosofía de desarrollo de software que cumple los siguientes principios:

a) **Abierto:** la comunidad tiene libre acceso, uso y participación del código fuente, así como la posibilidad de uso de foros para proporcionar *feedback*.

b) **Transparencia:** la comunidad tiene acceso al *roadmap*, documentación, defectos y agenda de las *milestones*.

c) **Early & often:** la información se publica de manera frecuente y pronto a través de repositorios públicos (incluyendo el código fuente).

El *open source* ya no es una tendencia emergente, sino que es un enfoque que tiene un impacto profundo y una presencia importante en todos los sectores, tal y como comentaban los analistas de Gartner (Plummer y otros, 2008):

«2012, el 80 % del SW comercial incluirá algún componente *open source*. Incluir componentes *open source* en los productos para abaratar costes es considerado la mínima estrategia que las compañías pueden llevar a cabo para mantener su ventaja competitiva en 5 años».

En los últimos años, el mercado *business intelligence* se ha enriquecido con soluciones *open source* que cubren todo el espectro de necesidades de una organización para la explotación de la información. Algunas de estas herramientas tienen ya a sus espaldas varios años de recorrido y actualmente se encuentran respaldadas por organizaciones que tienen un claro modelo de negocio orientado a los servicios de valor añadido. Es posible encontrar herramientas solventes y maduras desde el nivel de base de datos hasta el de procesos de minería de datos que pueden, en algunos casos, adaptarse a las necesidades de una organización.

De hecho, el advenimiento de las tecnologías de *big data*, cuyo núcleo está fundamentado en tecnologías con licencia libre, ha reforzado la idea de que el movimiento *open source* es una fuerza de mercado a tener en cuenta para la generación de ventajas competitivas y, sobre todo, en la explotación de los datos.

En el mercado de soluciones *open source business intelligence* (OSBI) existen múltiples soluciones. Las principales son: Pentaho, Tibco Jaspersoft, Birt, que forma parte de OpenText Analytics, y SpagoBI. Entre ellas destaca la solución de

#### SpagoBI

Cabe comentar que SpagoBI es actualmente la única solución en el mercado BI puramente *open source*.

Pentaho, que es la que usaremos en esta asignatura porque contiene todas las componentes mínimas de un sistema de *business intelligence* y es posible explicarlas y usarlas de manera unitaria e independiente.

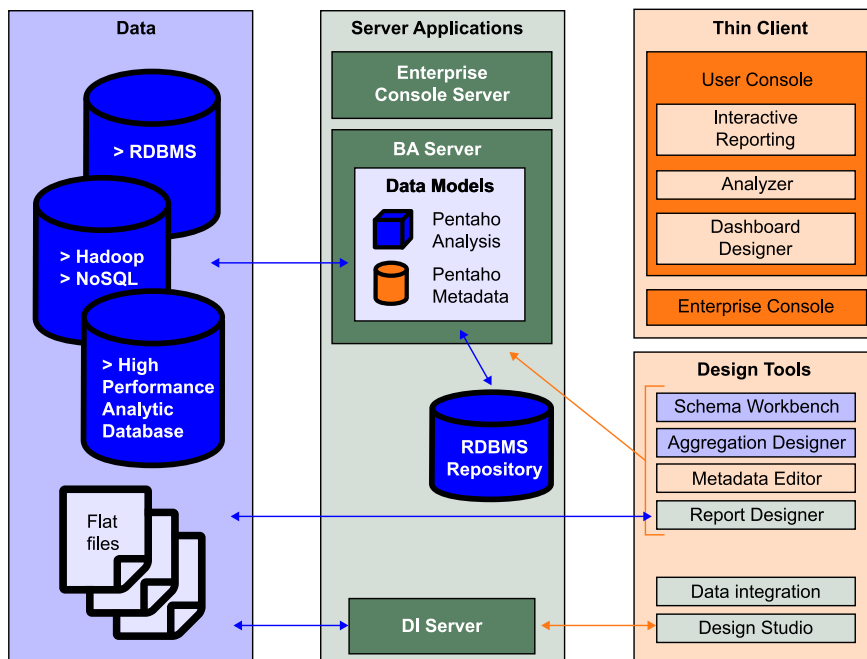
### 3.1. Pentaho

Nacida en el año 2006, Pentaho es una de las *suites* más conocidas del mercado OSBI. Existen dos versiones: *community* y *enterprise*. Serán comentadas a lo largo de los diferentes módulos. Está compuesta por diferentes motores incluidos en el servidor de Pentaho:

- **Reporting:** soporta informes estáticos, paramétricos y *ad hoc*.
- **Análisis:** soporta OLAP (mediante Mondrian) y minería de datos (mediante Weka).
- **Cuadros de mando:** mediante CDF (*community dashboard framework*) y *dashboard designer* (Enterprise).
- **ETL:** mediante la herramienta Pentaho Data Integration (anteriormente llamada Kettle). Incluye un servidor.
- **Metadata:** que proporciona una capa de acceso de información basada en lenguaje de negocio.

El siguiente esquema representa la arquitectura de Pentaho:

Figura 3. Arquitectura de Pentaho



Actualmente, Pentaho está siguiendo la estrategia *open core*, que consiste en, a partir de un núcleo *open source*, ofrecer servicios y módulos mejorados, lo que frecuentemente también se conoce como *commercial open source*. Esta es la razón de encontrarnos dos versiones. La principal diferencia entre ambas es que la versión *enterprise* está bajo una modalidad de suscripción (que incluye funcionalidad extendida) y la versión *community* es completamente gratuita. A lo largo de la asignatura la gran mayoría de las veces se hará referencia a la versión *community*.

Las principales páginas de referencia de Pentaho son estas:

- Página web de la compañía: <http://www.pentaho.com/>.
- Página de la comunidad: <http://community.pentaho.com/>.
- Documentación: <https://help.pentaho.com/Documentation>.
- Wiki comunidad: <http://wiki.pentaho.com/display/COM/Community+Wiki+Home>.
- Jira Pentaho: <http://jira.pentaho.com/>.
- Sourceforge Pentaho: <http://sourceforge.net/projects/pentaho>.
- Sourceforge Mondrian: <http://sourceforge.net/projects/mondrian>.
- Github Pentaho: <https://github.com/pentaho>.

## Abreviaturas

- BI** *Business intelligence.*
- BICC** *Business intelligence competency center.*
- BIMM** *Business intelligence maturity model.*
- COBIT** *Control objectives for information and related technology.*
- CRM** *Customer relationship management.*
- DSS** *Decision support systems.*
- EII** *Enterprise information integration.*
- ETL** *Extract, transform and load.*
- IBM** *International business machines.*
- KGI** *Key goal indicator.*
- KPI** *Key performance indicator.*
- ODS** *Operational data store.*
- OLAP** *Online analytical processing.*
- OLTP** *Online transaction processing.*
- OSBI** *Open source business intelligence.*
- SI** *Sistemas de información.*
- SQL** *Structured query language.*
- TI** *Tecnologías de la información.*

## Bibliografía

**Castells, M.** (1997-1998). *La era de la información: economía, sociedad y cultura* (3 vols.). Madrid: Alianza Editorial.

**Davenport, T. H.; Harris, J. G.** (2007). *Competing on Analytics: The New Science of Winning*. Nueva York: Harvard Business Press.

**Davenport, T. H.; Harris, J. G.; Morison, R.** (2010). *Analytics at Work: Smarter Decisions, Better Results*. Nueva York: Harvard Business Press.

**Miller, D.** (2007). *Measuring Business Intelligence Success: A Capability Maturity Model*. Nueva York: D. M. Morrissey.

**Miller, G. J.; Brautigam, D. V.; Gerlach, S.** (2006). *Business Intelligence Competency Centers: A Team Approach to Maximizing Competitive Advantage*. Hoboken: Wiley and SAS Business Series.

**Plummer, D. C.; Smulders, Ch.; Fiering, L.; Natis, Y. V.; Mingay, S.; Driver, M.; Fenn, J.; McLellan, L.; Wilson, D. R.** (2008, 8 de enero). «Gartner's Top Predictions for IT Organizations and Users, 2008 and Beyond: Going Green and Self-Healing». Gartner.

**Porter, M. E.** (1985). *The Competitive Advantage: Creating and Sustaining Superior Performance*. Nueva York: Free Press.