

Lección 2.5 Herramientas de BI para Big Data.

En las lecciones anteriores hemos detallado las distintas partes de un sistema de BI:

- Fuentes de datos.
- Procesos de Integración.
- El almacén de datos y los servidores OLAP.
- Las aplicaciones de usuario final: Cubos OLAP, generación de informes (Reporting), cuadros de mando (Dashboards) y cuadros de mando integrales (Balanced Scorecards), procesos de Minería de Datos...

Las aplicaciones de usuario son el producto final de un desarrollo BI, pero para desarrollar una solución BI completa son necesarias herramientas y técnicas que permitan implementar las distintas partes del sistema de BI.

Además, es necesario el uso de herramientas que soporten las nuevas características introducidas por el Big Data (las 5'vs).

1. Soporte para Big Data

- Algunas de las características a las que han de dar soporte las herramientas de BI para Big Data son:
 - Carga y gestión de grandes volúmenes de datos de forma eficiente (Volumen)
 - Baja latencia requerida para Dashboards y análisis OLAP.
 - Facilitar la integración de un mayor número de fuentes y soportar el amplio abanico de formatos existentes (Variedad).
 - Nuevos formatos: XML, JSON, BD's NoSQL, API's de servicios web...
 - Facilitar el diseño de procesos de verificación de la calidad (Veracidad).
 - Integración de fuentes en tiempo real (Velocidad).
 - Tanto recolección como integración en tiempo real.
- Por otro lado, necesitamos de herramientas que simplifiquen la aplicación de las técnicas para el desarrollo de aplicaciones de Inteligencia de Negocio - Big Data de forma lo más eficiente y efectiva posible.

2. Clasificación de herramientas BI

Los distintos tipo de aplicaciones BI para Big Data se pueden clasificar de forma simplificada en:

- Herramientas de gestión de las fuentes y diseño de procesos ETL.

- Para soportar la carga de datos en los repositorios Big Data desde los nuevos tipos de fuentes y permitan el diseño y ejecución de procesos ETL (Extracción, Transformación y Carga).
- Herramientas de diseño de cubos o tablas OLAP, informes (Reporting) y cuadros de mando (Dashboard).
 - Herramientas que simplifiquen el desarrollo de visualizaciones y que permitan conectar con los almacenes o repositorios Big Data para la carga de los datos.
- Herramientas de minería de datos.
 - Procesos que intentan descubrir patrones en grandes volúmenes de datos aplicando métodos de la inteligencia artificial, aprendizaje automático, o la estadística.
- Herramientas de servidor.
 - Para permitir la ejecución y compartición de las visualizaciones e informes a través de una intranet o internet.

Además de las herramientas anteriores tenemos las herramientas propias de una arquitectura Big Data, como las que permiten el almacenamiento, gestión y consulta de la información (ej. Hadoop y las BD NoSQL).

3. Herramientas de BI que soportan Big Data

En el mercado existen distintos tipos de soluciones, que nos permiten afrontar el desarrollo de aplicaciones BI-Big Data tanto de forma completa como para alguna de las partes del proceso (ej. Procesos ETL, Visualización,...).

- En general podemos clasificar las soluciones existentes en 3 grandes grupos:
 - Open source: Pentaho Community Edition, Spago BI, Jaspersoft, Talend (solo ETL).
 - Resultados profesionales pero mayor complejidad de desarrollo.
 - Propietarias: Tableau, Excel 2013, Qlikview, SQL Server Reporting Services.
 - Resultados profesionales con menor complejidad a costa de pagar los precios de las licencias (en algunos casos solo asequibles para grandes empresas).
 - Soluciones en la nube: Microstrategy Cloud, Google Big Query, Google Fusion Tables, Microsoft Azure.

- Resultados profesionales pagando solo por aquello que se necesita en cada momento. A cambio, perdemos control sobre el sistema (dependemos del proveedor del servicio y aplicaciones en la nube).
- Además algunas de estas soluciones son de tipo “Self Service”, es decir, simplifican la realización del desarrollo de la solución BI.
 - Excel 2013 es un ejemplo de ello con sus plugins:
 - Power Query: Diseño y ejecución de ETL para la carga de datos en una hoja de excel o un modelo de Power Pivot.
 - Power Pivot: Soporta la creación de complejos modelos de datos en memoria, superando las limitación de filas de una hoja de excel.
 - Power View: Diseño y publicación de Dashboard a partir de los datos de una hoja de Excel o un modelo de Power Pivot.

4. Adaptación de la técnicas de BI a Big Data

- No solo hay que adaptar las herramientas si no también la forma de proceder.
 - Técnicas y metodologías específicas para el desarrollo de soluciones BI - Big Data.
- Aunque aún queda mucho por hacer, la industria proporciona conjuntos de mejores prácticas.
 - Para el diseño del modelo de datos
 - Podemos aplicar el modelado multidimensional (el cual vimos en la lección 2.2) para el diseño del repositorio o almacén Big Data, al igual que se ha hecho con éxito en el diseño de Almacenes de Datos (Data Warehouses), durante más de 30 años.
 - Para la gestión de los datos.
 - Para la gobernanza de datos.
 - Para el diseño de la arquitectura.