**Data Lake**

Un data lake es un repositorio de almacenamiento que contienen una gran cantidad de datos en bruto y que se mantienen allí hasta que sea necesario. A diferencia de un data warehouse jerárquico que almacena datos en ficheros o carpetas, un data lake utiliza una arquitectura plana para almacenar los datos.

A cada elemento de un data lake se le asigna un identificador único y se etiqueta con un conjunto de etiquetas de metadatos extendidas. Cuando se presenta una cuestión de negocios que debe ser resuelta, podemos solicitarle al data lake los datos que estén relacionados con esa cuestión. Una vez obtenidos podemos analizar ese conjunto de datos más pequeño para ayudar a obtener una respuesta.

El data lake se asocia a menudo con el almacenamiento de objetos orientado a Hadoop. En este escenario, los datos de una organización se cargan primero en la plataforma Hadoop y, a continuación, se aplican las herramientas de análisis y de minería de datos a los datos que residen en los nodos clúster de Hadoop.

**Big data**

Big data es un término que describe el gran volumen de datos – estructurados y no estructurados – que inundan una empresa todos los días. Pero no es la cantidad de datos lo importante. Lo que importa es lo que las organizaciones hacen con los datos. El big data puede ser analizado para obtener insights que conlleven a mejores decisiones y acciones de negocios estratégicas.

**OLAP**

Es el acrónimo en inglés de procesamiento analítico en línea (On-Line Analytical Processing). Es una solución utilizada en el campo de la llamada Inteligencia de negocios (o Business Intelligence) cuyo objetivo es agilizar la consulta de grandes cantidades de datos.

**Datawarehouse**

Un data warehouse es un repositorio unificado para todos los datos que recogen los diversos sistemas de una empresa. El repositorio puede ser físico o lógico y hace hincapié en la captura de datos de diversas fuentes sobre todo para fines analíticos y de acceso.

Normalmente, un data warehouse se aloja en un servidor corporativo o cada vez más, en la nube. Los datos de diferentes aplicaciones de procesamiento de transacciones Online (OLTP) y otras fuentes se extraen selectivamente para su uso por aplicaciones analíticas y de consultas por usuarios.

Data Warehouse es una arquitectura de almacenamiento de datos que permite a los ejecutivos de negocios organizar, comprender y utilizar sus datos para tomar decisiones estratégicas. Un data warehouse es una arquitectura conocida ya en muchas empresas modernas.

**Data Mining**

Data Mining es un término que está de moda y que se utiliza, en muchas ocasiones, de manera incorrecta. Algunas de las referencias erróneas que se le atribuyen son que se trata de cualquier forma de datos a gran escala, o de procesamiento de la información.

La enorme cantidad de datos que se generan y procesan a diario en las organizaciones lleva de la mano la necesidad de analizarlos en tiempo real, el problema es que los métodos tradicionales de análisis de datos no son suficientes. De ahí surgen los conceptos de Data Mining y Big Data. Ambos se refieren al hecho de saber aprovechar convenientemente las gigantescas masas de información de que disponen las organizaciones en su día a día, para extraer información útil que les ayude en la toma de decisiones

**Bunisness Intelligence**

El término Business Intelligence (BI por sus siglas en inglés) hace referencia al uso de estrategias y herramientas que sirven para transformar información en conocimiento, con el objetivo de mejorar el proceso de toma de decisiones en una empresa.

En plena era digital, tomar decisiones bien informadas es uno de los principales factores de diferenciación de las empresas. En este post vamos a explicar con precisión qué es Business Intelligence, o inteligencia de negocio, y qué herramientas de BI existen.

**ETL**

Extract, Transform and Load («extraer, transformar y cargar», frecuentemente abreviado ETL) es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, data mart, o data warehouse para analizar, o en otro sistema operacional para apoyar un proceso de negocio.

Los procesos ETL también se pueden utilizar para la integración con sistemas heredados. Se convirtieron en un concepto popular en los años 1970.

**KDD**

Knowledge Discovery implica la evaluación e interpretación de patrones y modelos para tomar decisiones con respecto a lo que constituye conocimiento y lo que no lo es. Por lo tanto, el KDD requiere de un amplio y profundo conocimiento sobre tu área de estudio.