

DEFINICIONES:

BI:

conjunto de modelos matemáticos y metodologías de análisis que explotan los datos para generar información y conocimiento útiles para procesos complejos de toma de decisiones.

El propósito principal de los sistemas BI es proporcionar a los trabajadores del conocimiento, las herramientas y metodologías que les permitan tomar decisiones efectivas y oportunas.

Si los responsables de la toma de decisiones pueden confiar en un sistema de inteligencia empresarial que facilite su actividad, podemos esperar que la calidad global del proceso de toma de decisiones mejore considerablemente.

DSS:

Es una aplicación informática interactiva que combina datos y modelos matemáticos para ayudar a los responsables de la toma de decisiones a resolver problemas complejos que se enfrentan en la gestión de las empresas y organizaciones públicas y privadas. Las herramientas de análisis proporcionadas por una arquitectura de inteligencia empresarial pueden considerarse como DSS capaces de transformar los datos en información y conocimiento útiles para los responsables de la toma de decisiones.

DW:

El término almacén de datos (DW) indica todo el conjunto de actividades interrelacionadas con el diseño, la implementación y el uso de un almacén de datos. En este sentido, el principal objetivo del DW es facilitar la consulta de los datos por parte de los usuarios. Con ello, es posible generar reportes históricos, análisis estadísticos, predicciones, comparaciones, proyecciones y demás. Esto con la finalidad de poder tomar decisiones estratégicas cada vez más acertadas.

Un Data Warehouse se crea al extraer datos desde una o más bases de datos de aplicaciones operacionales. Los datos extraídos son transformados para eliminar inconsistencias, si es necesario y luego ser cargados en el Data Warehouse. El proceso de transformar, crear el detalle de tiempo variante, resumir y combinar las extracciones de datos, ayudan a crear el ambiente para el acceso a la información Institucional.

DM:

Son sistemas que recopilan todos los datos requeridos por un departamento específico de la empresa, como marketing o logística, con el fin de realizar análisis de BI y ejecutar aplicaciones de apoyo a la toma de decisiones específicas para la propia función.

Por lo tanto, un DM puede considerarse como un almacén de datos funcional o departamental de un tamaño menor y un tipo más específico que el DW general de la empresa, con el fin de implementar aplicaciones de BI, algunas empresas se someten a diseñar y desarrollar de forma incremental una serie de DM integrados en lugar de un DW central, con el fin de reducir el tiempo de implementación y las incertidumbres relacionadas con el proyecto.

DL:

Es un repositorio de almacenamiento centralizado que contiene Big Data de varias fuentes en un formato granular y sin procesar. Puede guardar datos estructurados, semiestructurados o no estructurados, lo que significa que los datos pueden conservarse en un formato más flexible para usarlos en un futuro. Al guardar datos, un DL los asocia con identificadores y etiquetas de metadatos para poder extraerlos más rápidamente. Los DL suelen estar configurados sobre un clúster de hardware de consumo económico y escalable. Esto permite volcar los datos al DL por si se necesitan más adelante sin tener que preocuparse por la capacidad de almacenamiento. Los clústeres pueden existir localmente o en la cloud.

ETL:

Las herramientas de software que se dedican a realizar de forma automática tres funciones principales: extracción, transformación y carga de datos en el DW

• Extracción. Los datos se extraen de las fuentes internas y externas disponibles. La selección de los datos que se van a importar se basa en el diseño del almacén de datos, que a su vez depende de la información necesaria para los análisis de inteligencia empresarial y los DSS que operan en un dominio de aplicación específico.

• Transformación. El objetivo de la fase de limpieza y transformación es mejorar la calidad de los datos extraídos de las diferentes fuentes, a través de la corrección de incoherencias, inexactitudes y valores faltantes. Algunas de las principales deficiencias que se eliminan durante la fase de limpieza de datos son:

• incoherencias entre los valores registrados en diferentes atributos que tienen el mismo significado;

• duplicación de datos;

• datos faltantes;

• existencia de valores inadmisibles

Durante la fase de limpieza, se aplican reglas automáticas preestablecidas para corregir los errores actuales. En muchos casos, los diccionarios con términos válidos se utilizan para sustituir los términos supuestamente incorrectos, en función del nivel de similitud.

Durante la fase de transformación, se realizan conversiones de datos para garantizar la homogeneidad y la integración con respecto a las diferentes fuentes de datos.

Se realizan la agregación y consolidación de datos con el fin de obtener los resúmenes que reducirán el tiempo de respuesta necesario para las consultas y análisis posteriores.

Carga. Por último, después de ser extraídos y transformados, los datos se cargan en las tablas del DW para ponerlos a disposición de los analistas y las aplicaciones de apoyo a la toma de decisiones. Se debe considerar la carga de datos de la siguiente manera: • Replicar las tablas de dimensiones • Replicar la tabla de hechos • Realizar una carga automática • Indexar los datos para un mayor rendimiento

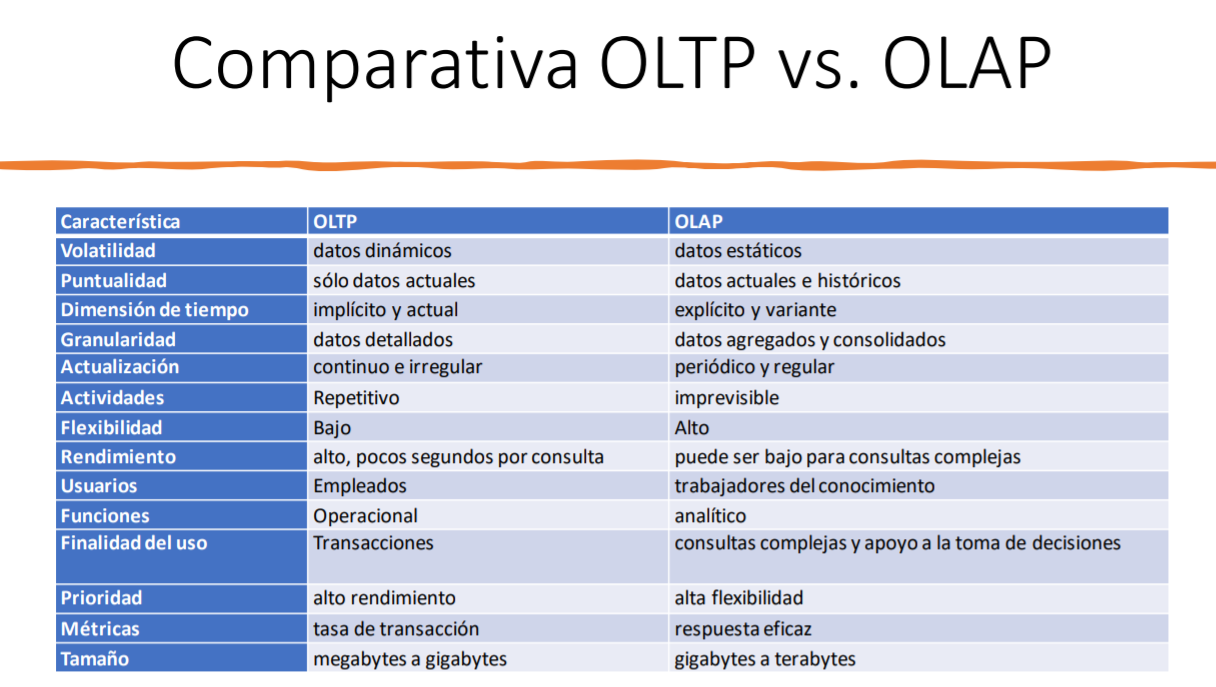
OLAP:

Es una solución utilizada en la Inteligencia de Negocios cuyo objetivo es agilizar la consulta de grandes cantidades de datos. Para ello utiliza estructuras de datos diversas, normalmente multidimensionales (o Cubos OLAP), que contienen datos resumidos de Sistemas Transaccionales (OLTP). La principal característica que potencia OLAP, es que es lo más rápido al ejecutar sentencias SQL de tipo SELECT, en contraposición con OLTP que es la mejor opción para operaciones de tipo INSERT, UPDATE Y DELETE.

Por otro lado, todo el conjunto de herramientas destinadas a realizar análisis de inteligencia de negocios y soporte los procesos de toma de decisiones se den por el nombre de procesamiento analítico en línea (on line analitycal processing, OLAP).

OLTP:

Se hace referencia a las aplicaciones de software que están en el corazón de los sistemas operativos como procesamiento de transacciones en línea (on line transactional processing, OLTP).



INTEGRAN:

Partes del BI:

Fuentes de datos

Los datos proceden en mayor parte de transacciones internas y fuentes externas. Sin embargo, aunque se hayan recopilado y almacenado de forma sistemática y estructurada, estos datos no pueden utilizarse directamente con fines de toma de decisiones, deben procesarse mediante herramientas de extracción adecuadas y métodos analíticos capaces de transformarlos en información y conocimientos que puedan ser posteriormente utilizados por los responsables de la toma de decisiones.

DW/DM

Permite, mediante el concepto de modelo multidimensional, estructurar los datos en un repositorio empresarial, para aplicar consultas OLAP. El DM mantiene los mismos conceptos, pero a una escala reducida de capacidad y funcionamiento.

Exploración de datos

Consta de las herramientas para realizar un análisis pasivo de BI, que consisten en sistemas de consulta e información, así como métodos estadísticos. Se pide a los responsables de la toma de decisiones que generen preguntas previas o definan criterios de extracción de datos, y luego utilicen las herramientas de análisis para encontrar respuestas y confirmar su conocimiento original.

Minería de Datos

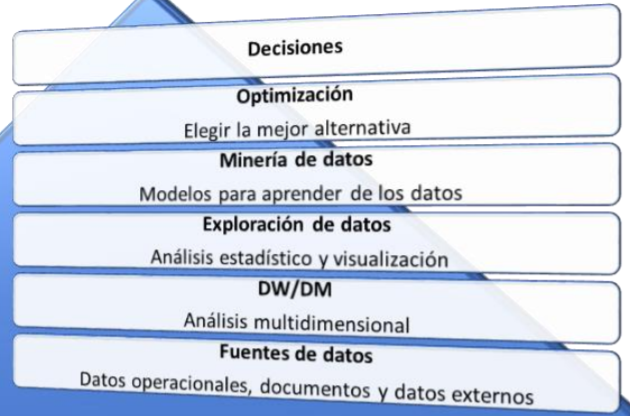
Incluye metodologías activas de BI, cuya finalidad es la extracción de información y conocimiento a partir de los datos. Se consideran modelos matemáticos para el reconocimiento de patrones, el aprendizaje automático y las técnicas de minería de datos

Optimización

Los modelos de optimización permiten determinar la mejor solución a partir de un conjunto de acciones alternativas.

Decisiones

Corresponde a la elección y la adopción real de una decisión específica, y de alguna manera representa la conclusión natural del proceso de toma de decisiones. La elección de una decisión es por parte de los responsables de la toma de decisiones, que también pueden aprovechar la información informal y no estructurada disponible para adaptar y modificar las recomendaciones y las conclusiones alcanzadas mediante el uso de modelos matemáticos.



CARACTERÍSTICAS:

Data Lake:

Un DL funciona a partir de un principio llamado schema-on-read o esquema contra escritura. Esto significa que no existe un esquema predefinido en el que deban encajarse los datos antes de almacenarlos.

Un único repositorio compartido de datos, normalmente almacenado en el Sistema de archivos distribuido (DFS).

Incluye funcionalidades de orquestación y programación de trabajos (por ejemplo, a través de YARN).

Contiene un conjunto de aplicaciones o flujos de trabajo para consumir, procesar o actuar sobre los datos.

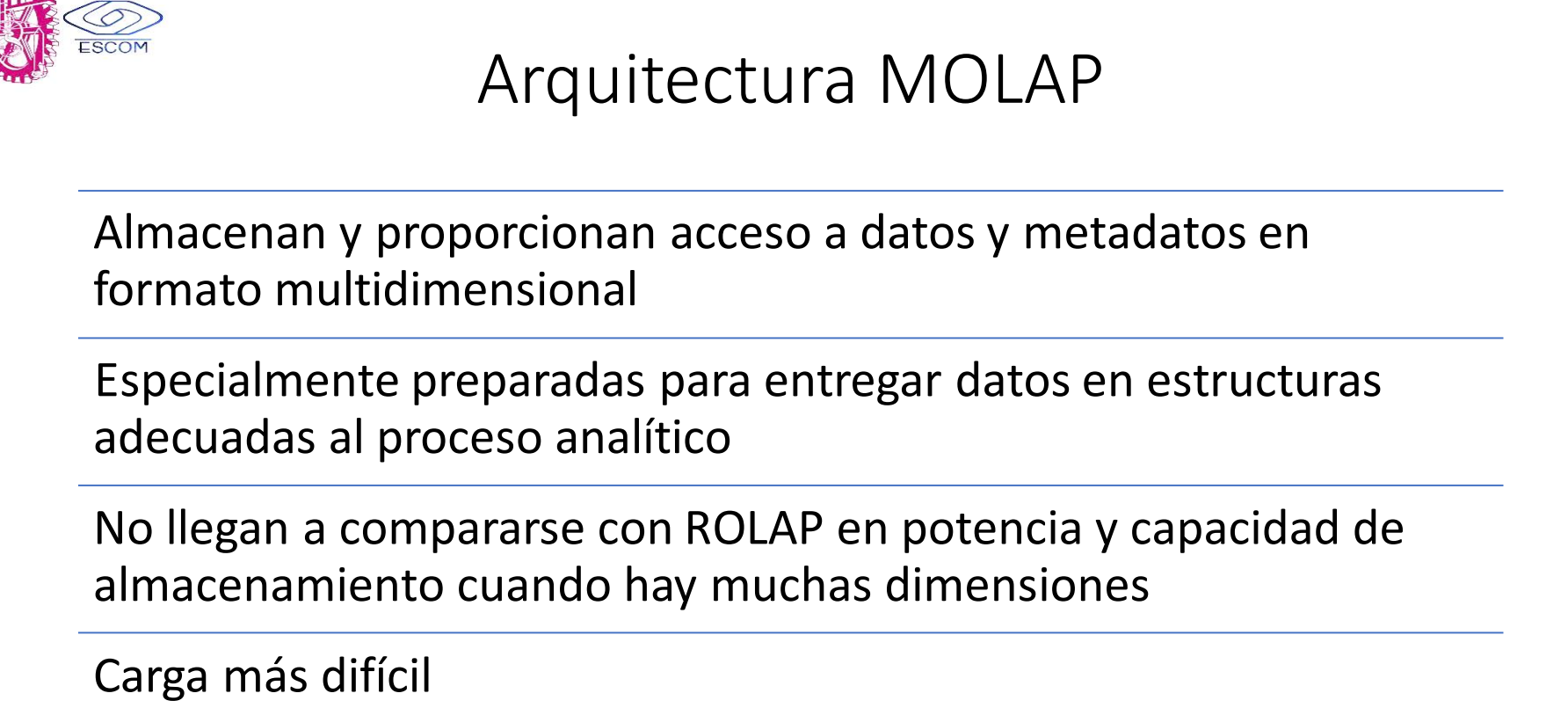
El fácil acceso de los usuarios, ya que los propietarios de datos pueden entonces consolidar datos de clientes, proveedores y operaciones, eliminando barreras técnicas e incluso políticas para compartir datos.

COMPLETAR:

Modelo multidimensional:

PREGUNTA:

MOLAP:



Referencia arquitectural de los DW

Características de un concepto, tenía “transporte y +”

