Tema 1. Analítica de datos

Introducción

López y Zarza (2017) describieron que el requerimiento más claro para llevar a cabo una mejor toma de decisiones es que existan los datos, es decir, que cada transacción, operación, decisión o consecuencia genere algún tipo de efecto identificable en el tiempo, de modo que pueda archivarse para futuras referencias.

Por tanto, el dato vivo es el dato que se consume de manera recurrente. En este aspecto, la infraestructura que soporte la generación, ingesta, almacenamiento y consumo de los datos será crítica para la organización.

El dato debe buscarse antes de tomar una decisión, y solo tienen sentido aquellos datos que la posibilitan. El término relevante, entonces, es **la decisión en relación con los datos**.

El siguiente elemento en la toma de decisiones es la evaluación de las consecuencias o resultados de las decisiones tomadas. El impacto de las decisiones se da en tres variables:

- La **operación** o el valor de negocio en la adaptación, creación o extinción de procesos y maneras de llevar adelante el día a día.
- La **cultura** en el tipo y combinación de habilidades analíticas y sociales requeridas para aportar valor de manera consistente y la mecánica de cooperación o trabajo en equipo.
- La **infraestructura** en las aplicaciones, tecnologías, así como micro y macroarquitecturas que soporten efectivamente las dos anteriores.

En este tema comprenderás el concepto de *big data*, cómo se conforma y la importancia de su uso en la toma de decisiones.

Explicación

De inicio, sabes que los datos generan información, conocimiento y más datos. Un entendimiento tradicional de la relación entre estas tres entidades se puede presentar de la siguiente manera:

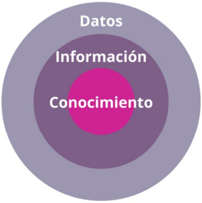


Figura 1. Núcleos tradicionales, desde los datos al conocimiento.

Sin embargo, en el big data, la generación del conocimiento no es tan lineal, sino que el análisis de datos permite tener una imagen clara del pasado y del presente para reconocer tendencias y cambios verdaderos, significa evaluar y explorar los propios criterios para exponer los sesgos (López y Zarza, 2017). Gráficamente se vería como en la figura 1.

El conocimiento en un sistema, por lo tanto, es la información en otro y tal vez el dato en un tercer sistema como muestra la figura 2.

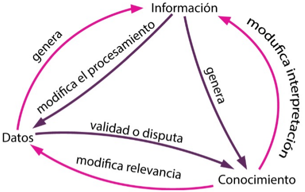


Figura 2. Sistema retroalimentado.
Fuente: López, J., y Zarza, G. (2017). LA INGENIERÍA DEL BIG DATA: CÓMO TRABAJAR CON DATOS. España: Editorial UOC.

Teniendo como base lo anterior, se puede comenzar a distinguir qué significa en la práctica ser big data o *data driven*.

Cabría entonces preguntarse: ¿Entiendes los reportes que utilizas? ¿Comprendes la variabilidad de los datos, los rangos de incertidumbre o lo confiable que es la fuente? ¿Puedes imaginar más de una posible explicación a los datos que lees? ¿Tomas decisiones y asumes riesgos controlados? ¿O tan solo buscas tener información para cubrir las decisiones que ya habías considerado?

La creación del conocimiento mediante el big data no tiene ninguna utilidad cuando este no se comparte con los involucrados en la toma de decisiones. Por lo tanto, debe verse como un proceso que amplifica lo generado por las personas en lo individual y se cristaliza como parte del conocimiento de la empresa.

El manejo apropiado de los datos deriva en una ventaja competitiva para la organización por la cantidad de interacciones con los clientes, las transacciones por medio de los dispositivos y los distintos canales de distribución.

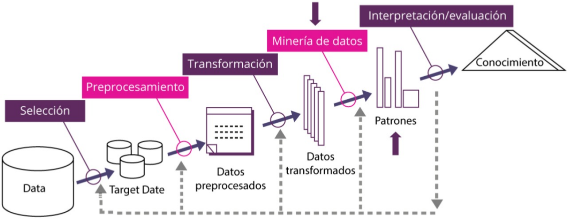


Figura 3. Creación del conocimiento a través del big data.
Fuente: López, J., y Zarza, G. (2017). LA INGENIERÍA DEL BIG DATA: CÓMO TRABAJAR CON DATOS. España: Editorial UOC.

1.1 ¿Qué es el big data?

Para explicar lo que significa big data, primero hay que señalar lo que no es, esto para que puedas dimensionar el concepto en tu situación personal.

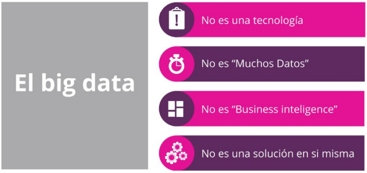


Figura 4. Qué no es el big data.
Fuente: Facultad de Administración y Ciencias Sociales - Universidad ORT Uruguay. (2018, 17 de julio). *Analítica de datos aplicada a negocios: ¿Es para mí? ¿Es ahora?* [Archivo de video]. Recuperado de https://www.youtube.com/watch?v=n7bwlqyhPa8

El big data se considera como tal a partir del cumplimiento de las siguientes características:

Big Data
Velocity
Variety

Figura 5. Condicionantes del big data.

Fuente: Facultad de Administración y Ciencias Sociales - Universidad ORT Uruguay. (2018, 17 de julio). *Analítica de datos aplicada a negocios: ¿Es para mí? ¿Es ahora?* [Archivo de video]. Recuperado de https://www.youtube.com/watch?v=n7bwlqyhPa8

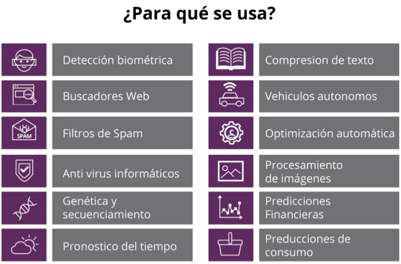
De acuerdo con Ladredo (2020), las tres V del big data significan lo siguiente:

- **Volumen:** es la cantidad masiva de datos a analizar.
- **Velocidad:** se refiere al ritmo en que los datos de entrada fluyen desde las diversas fuentes, tales como los procesos de negocio, las máquinas y sensores, las redes sociales, los dispositivos móviles, etcétera.
- **Variedad:** se refiere a las diferentes fuentes y tipos de datos, tanto estructurados como no estructurados, provenientes de emails, fotos, videos, sistemas de monitorización, PDF, ficheros de sonido, etcétera.

Esto significa que las organizaciones llegan a un punto en el que las hojas de cálculo y las formas tradicionales de gestionar les es insuficiente para la toma de decisiones, por lo que el siguiente paso es transitar hacia nuevas formas de administrar los recursos: el big data.

La literatura especializada describe que un exabyte alcanza para registrar todas las palabras pronunciadas por todos los seres humanos que hayan existido. La mayor parte de la catarata de datos se crea porque sí de forma espontánea.
Cada vez que haces algo que involucra una transacción, registro o un dispositivo electrónico, se puede saber el tipo de personalidad, mapas detallados de cerebros de acuerdo con las neuronas conectadas, planos de comportamiento criminal, manejo de crisis y catástrofes naturales, entre otras cosas, y esto solamente cuando las personas comparten opiniones en las redes sociales (Sosa, 2019).

La figura 6 es clara para describir algunos otros usos que se le da al big data.



¿Para qué se usa?

Detección biométrica	Compresion de texto
Buscadores Web	Vehiculos autonomos
Filtros de Spam	Optimización automática
Anti virus informáticos	Procesamiento de imágenes
Genética y secuenciamiento	Predicciones Financieras
Pronostico del tiempo	Preducciones de consumo

Figura 6. Usos del big data.

Fuente: Facultad de Administración y Ciencias Sociales - Universidad ORT Uruguay. (2018, 17 de julio). *Analítica de datos aplicada a negocios: ¿Es para mí? ¿Es ahora?* [Archivo de video]. Recuperado de https://www.youtube.com/watch?v=n7bwlqyhPa8

¿Qué hacer entonces?

Comprender el proceso.

Hay que recordar que el big data es una nueva disciplina que aún está en vías de desarrollo y que en las organizaciones falta personal especializado para su manejo, ya que no se ha dado el salto del *Small Data* a nuevos modelos de análisis de datos.

La siguiente figura muestra la secuencia que te permitirá visualizar la totalidad del proceso para que puedas decidir más acertadamente mediante el análisis.

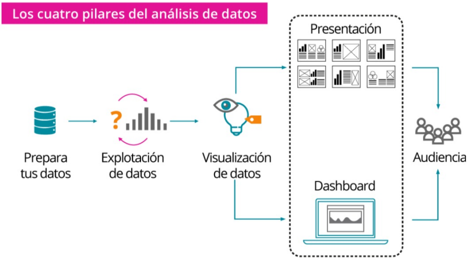


Figura 7. Los cuatro pilares del análisis de datos.

Fuente: Comunicación Numérica. (2020, 7 de septiembre). *Fundamentos del Análisis de Datos para Toma de Decisiones* [Archivo de video]. Recuperado de https://www.youtube.com/watch?v=qvZxvMWMvDo&t=1s

Hay que ir por partes.

El inicio, como puedes observar en la figura 8, es hacerte de tres elementos importantes para analizar tus insumos informativos, los cuales son los siguientes:

La **estrategia** se centra en definir las preguntas que el análisis deberá responder.

Por ejemplo: ¿Cómo han evolucionado los ingresos en el último año?, ¿qué relación existe entre la rentabilidad y la cantidad de ingresos por categoría? Una vez definidas las preguntas, el siguiente paso es **construir** los conjuntos de **datos** que te permitirán contestarlas.

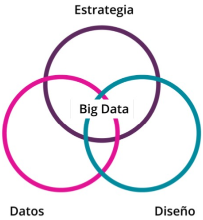


Figura 8. Elementos para el análisis de datos.
Fuente: López, J., y Zarza, G. (2017). *LA INGENIERÍA DEL BIG DATA: CÓMO TRABAJAR CON DATOS*. España: Editorial UOC

Este proceso puede ser muy sencillo si los datos se encuentran ya disponibles de manera estructurada. Pero si no es el caso, la preparación de los datos puede suponer hasta el 80% del tiempo de un proyecto de analítica. Es importante, por lo tanto, que los datos cuenten con:

- Disponibilidad: significa que los datos lleguen a tiempo dentro de un periodo determinado, de tal forma que puedan actualizarse de la manera correcta y continua.
- Formato: según el IEP (s.f.), existen diferentes formas de considerar los datos para su análisis, estas son las siguientes:
  a. **Datos estructurados**: datos ordenados y bien definidos en cuanto a su formato, tamaño y longitud. Por ejemplo, base de datos, hojas de cálculo, etcétera.
  b. **Datos no estructurados**: los datos están desorganizados y no tienen valor hasta que se ordenan, se identifican y se almacenan. Algunos ejemplos son los videos, audios, imágenes, PDF, etcétera.
  c. **Semiestructurados**: no están perfectamente estructurados, pero sí tienen una organización definida. En este caso, serían datos con los formatos HTML, XML o JSON.
- Limpieza: se refiere a la depuración de los datos erróneos de la base de datos.
- Procesamiento: de acuerdo con Tokio School (2021), el procesamiento en big data se lleva a cabo mediante:
  a. **Disputa de datos**: se limpia el conjunto de datos y se transforman en datos más accesibles y utilizables.
  b. **Compresión de datos**: los datos cambian a un formato que se pueda almacenar de manera más eficiente.
  c. **Cifrado de datos**: los datos se traducen a otro código para poder protegerlos de problemas de seguridad.

El proceso de **diseño** consta de tres fases:

- Esbozar: definir el contorno o estructura del big data.
- Prototipar: hacer la versión inicial del big data.
- Finalizar: la última fase del diseño consiste en preparar tu análisis visual para comunicarlo.

1.2 Definición de bases de datos para analítica

HEAVY.AI (s.f.) describe que una base de datos analítica almacena y administra big data, incluidos datos comerciales, de mercado y de clientes para el análisis de inteligencia comercial.

Las bases de datos analíticas están optimizadas para proporcionar tiempos de respuesta de consulta rápidos y análisis avanzados. Sus características incluyen almacenamiento basado en columnas, carga en memoria de datos comprimidos y la capacidad de buscar datos a través de múltiples atributos.

Los modelos de big data se consideran buenos cuando te hacen ganar tiempo en la obtención de información.

De acuerdo con López (2021), la programación tradicional (modelos relacionales) hace que se consideren soluciones tecnológicas muy sobredimensionadas. Algunas organizaciones de tecnologías ofrecen soluciones excelentes para proyectos de big data con una memoria de dos terabytes de almacenamiento.

Estos tipos de soluciones pueden procesar hasta 500 millones de registros en un lapso de dos a tres minutos, que es un muy buen tiempo de procesamiento, además de que la herramienta es muy fácil de manejar.

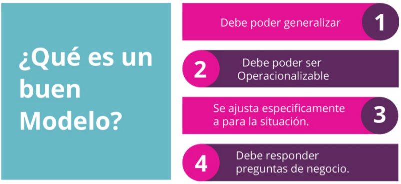

¿Qué es un buen Modelo?
Debe poder generalizar 1
2 Debe poder ser Operacionalizable
Se ajusta especificamente a para la situación. 3
4 Debe responder preguntas de negocio.

Figura 9. Características de un buen modelo de big data.

Fuente: Facultad de Administración y Ciencias Sociales - Universidad ORT Uruguay. (2018, 17 de julio). *Analítica de datos aplicada a negocios: ¿Es para mí? ¿Es ahora?* [Archivo de video]. Recuperado de https://www.youtube.com/watch?v=n7bwIqyhPa8

En el siguiente esquema puedes observar que la información se alimenta de distintas fuentes (Holistics, 2020).

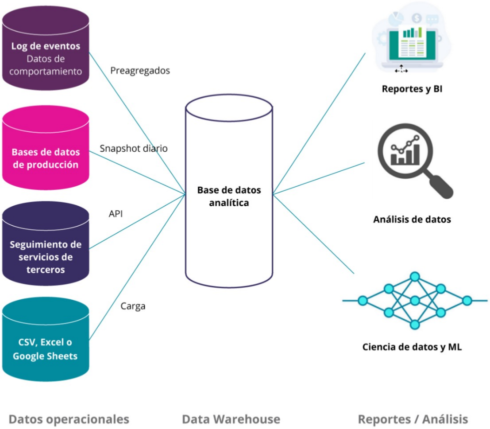


Figura 10. Configuración común para una base de datos analítica.

Fuente: Holistics. (2020). *The Analytics Setup Guidebook*. Recuperado de https://www.holistics.io/books/setup-analytics/a-modern-analytics-stack/

Como puedes observar, las **fuentes** de las que se alimenta la base de datos pueden ser de distinta naturaleza, tales como:

- **Log de eventos**: estos datos son los que pueden reflejar comportamientos de sistemas, usuarios o componentes.
- **Bases de datos de producción**: son los datos centrales del negocio, por ejemplo, las ventas, la información de clientes, la de inventario, facturas y/o cotizaciones, esa información con la que la empresa trabaja en el día a día y que refleja su operación central.
- **Seguimiento de servicios de terceros**: como anuncios en Facebook, seguimiento geográfico de barcos o camiones, seguimiento de sensores de algún tipo, incluso dispositivos IoT.
- **CSV** (valores separados por comas), **Excel** o **Google Sheets** (hojas de cálculo): formato de archivos que son necesarios incluir en la base de datos cada determinado tiempo.

1.3 Distintos tipos de análisis

Dependiendo de tus propias necesidades sobre lo que deseas hacer, puedes obtener información de tipo:

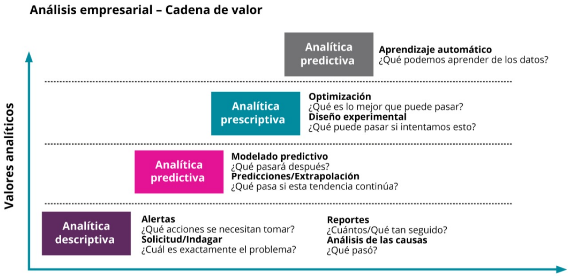

Análisis empresarial – Cadena de valor
Valores analíticos
Analítica predictiva
Aprendizaje automático
¿Qué podemos aprender de los datos?
Analítica prescriptiva
Optimización
¿Qué es lo mejor que puede pasar?
Diseño experimental
¿Qué puede pasar si intentamos esto?
Analítica predictiva
Modelado predictivo
¿Qué pasará después?
Predicciones/Extrapolación
¿Qué pasa si esta tendencia continúa?
Analítica descriptiva
Alertas
¿Qué acciones se necesitan tomar?
Solicitud/Indagar
¿Cuál es exactamente el problema?
Reportes
¿Cuántos/Qué tan seguido?
Análisis de las causas
¿Qué pasó?

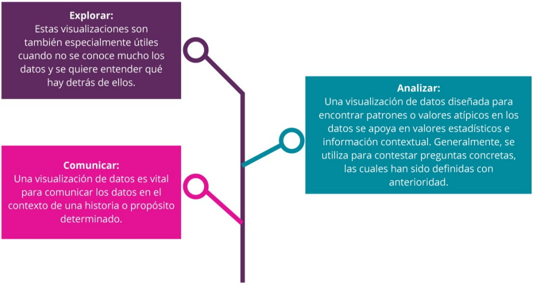
Figura 11. Análisis de datos- Valor agregado.

Fuente: Facultad de Administración y Ciencias Sociales - Universidad ORT Uruguay. (2018, 17 de julio). *Analítica de datos aplicada a negocios: ¿Es para mí? ¿Es ahora?* [Archivo de video]. Recuperado de https://www.youtube.com/watch?v=n7bwIqyhPa8

- **Descriptiva**: porque lo relevante es la capacidad de comprender el estado de las cosas por medio de un reporte o visualización, qué ocurrió en el pasado y qué factores intervinieron.
- **Predictiva**: porque se intentan explotar los sucesos pasados para poder decir qué es lo que ocurrirá.
- **Prescriptiva**: porque además de intentar adelantar la ocurrencia de determinados sucesos, se busca dilucidar cuál es el mejor curso de acción posible.
- **Asistida**: consiste en decisiones automatizadas para casos de bajo riesgo y bajo valor, apoyo para descubrir eventos o tendencias relevantes y otras funcionalidades, pero respondiendo básicamente a las preguntas metanalíticas: ¿qué información debería estar mirando?, ¿qué decisión debería

estar considerando?

Principalmente, las visualizaciones de datos permiten conseguir tres objetivos:



Explorar:
Estas visualizaciones son también especialmente útiles cuando no se conoce mucho los datos y se quiere entender qué hay detrás de ellos.

Analizar:
Una visualización de datos diseñada para encontrar patrones o valores atípicos en los datos se apoya en valores estadísticos e información contextual. Generalmente, se utiliza para contestar preguntas concretas, las cuales han sido definidas con anterioridad.

Comunicar:
Una visualización de datos es vital para comunicar los datos en el contexto de una historia o propósito determinado.

Figura 12. Objetivos del análisis visual.

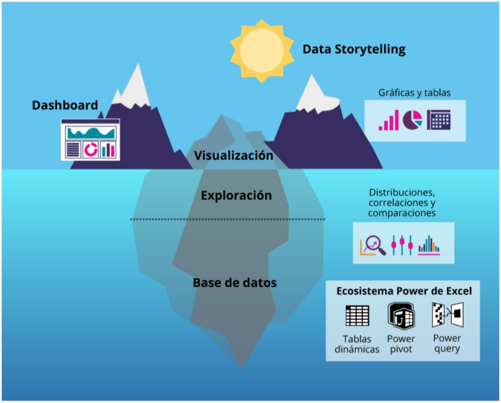
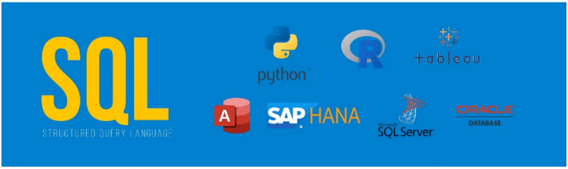


Figura 13. El iceberg de la analítica de datos.
Fuente: López, J., y Zarza, G. (2017). *LA INGENIERÍA DEL BIG DATA: CÓMO TRABAJAR CON DATOS*. España: Editorial UOC.

Como puedes observar en la figura 13, el big data es un iceberg lleno de oportunidades que se encuentra en un mar de datos muy extenso y peligroso, pues muchas empresas solo se enfocan en la punta del iceberg.

Para muchas personas, el big data es solo una nueva tecnología que usa datos masivos, sobre todo de las redes sociales, para tomar decisiones sobre sus estrategias de ventas y publicidad. Sin embargo, es mucho más que eso, dado que engloba distintas ciencias, como la estadística y el *machine learning*, para diferentes propósitos en finanzas, industria, medicina, recursos humanos. Por ello, si no se tiene cuidado con el resguardo de los datos, su análisis puede generar grandes dolores de cabeza.

1.4 Herramientas para hacer análisis de datos



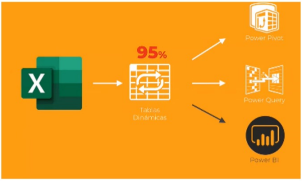


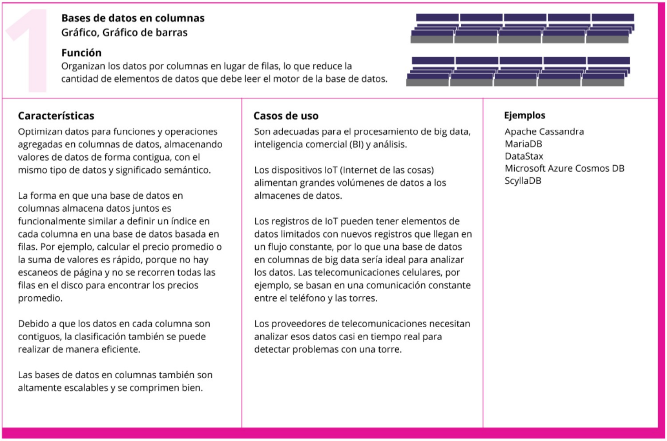
Figura 14. Herramientas para la analítica de datos.

Fuente: López, J., y Zarza, G. (2017). *LA INGENIERÍA DEL BIG DATA: CÓMO TRABAJAR CON DATOS*. España: Editorial UOC.

1.5 Tipos de bases de datos analíticas en el mercado

Según Indicative (2022), las características de las bases de datos analíticas incluyen almacenamiento basado en columnas, carga en memoria de datos comprimidos y la capacidad de buscar datos a través de múltiples atributos.

A continuación, se describen algunas de las principales bases de datos analíticas en el mercado:



1

Bases de datos en columnas
Gráfico, Gráfico de barras

Función

Organizan los datos por columnas en lugar de filas, lo que reduce la cantidad de elementos de datos que debe leer el motor de la base de datos.

Características

Optimizan datos para funciones y operaciones agregadas en columnas de datos, almacenando valores de datos de forma contigua, con el mismo tipo de datos y significado semántico.

La forma en que una base de datos en columnas almacena datos juntos es funcionalmente similar a definir un índice en cada columna en una base de datos basada en filas. Por ejemplo, calcular el precio promedio o la suma de valores es rápido, porque no hay escaneos de página y no se recorren todas las filas en el disco para encontrar los precios promedio.

Debido a que los datos en cada columna son contiguos, la clasificación también se puede realizar de manera eficiente.

Las bases de datos en columnas también son altamente escalables y se comprimen bien.

Casos de uso

Son adecuadas para el procesamiento de big data, inteligencia comercial (BI) y análisis.

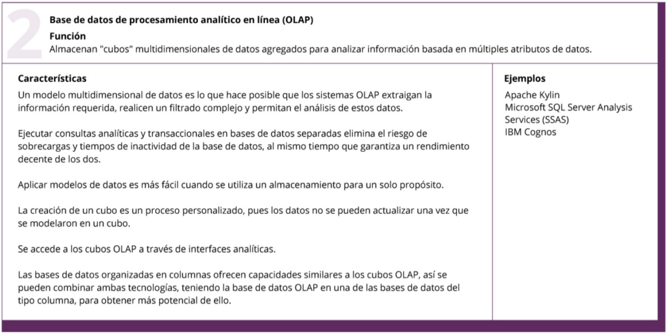
Los dispositivos IoT (Internet de las cosas) alimentan grandes volúmenes de datos a los almacenes de datos.

Los registros de IoT pueden tener elementos de datos limitados con nuevos registros que llegan en un flujo constante, por lo que una base de datos en columnas de big data sería ideal para analizar los datos. Las telecomunicaciones celulares, por ejemplo, se basan en una comunicación constante entre el teléfono y las torres.

Los proveedores de telecomunicaciones necesitan analizar esos datos casi en tiempo real para detectar problemas con una torre.

Ejemplos

Apache Cassandra
MariaDB
DataStax
Microsoft Azure Cosmos DB
ScyllaDB



2

Base de datos de procesamiento analítico en línea (OLAP)

Función

Almacenan "cubos" multidimensionales de datos agregados para analizar información basada en múltiples atributos de datos.

Características

Un modelo multidimensional de datos es lo que hace posible que los sistemas OLAP extraigan la información requerida, realicen un filtrado complejo y permitan el análisis de estos datos.

Ejecutar consultas analíticas y transaccionales en bases de datos separadas elimina el riesgo de sobrecargas y tiempos de inactividad de la base de datos, al mismo tiempo que garantiza un rendimiento decente de los dos.

Aplicar modelos de datos es más fácil cuando se utiliza un almacenamiento para un solo propósito.

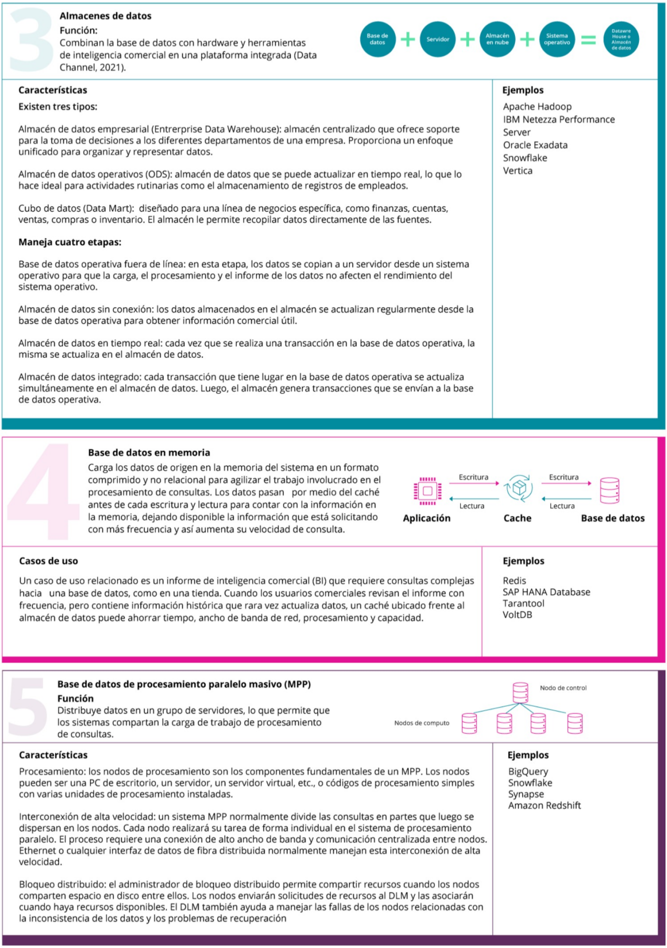
La creación de un cubo es un proceso personalizado, pues los datos no se pueden actualizar una vez que se modelaron en un cubo.

Se accede a los cubos OLAP a través de interfaces analíticas.

Las bases de datos organizadas en columnas ofrecen capacidades similares a los cubos OLAP, así se pueden combinar ambas tecnologías, teniendo la base de datos OLAP en una de las bases de datos del tipo columna, para obtener más potencial de ello.

Ejemplos

Apache Kylin
Microsoft SQL Server Analysis Services (SSAS)
IBM Cognos



3

Almacenes de datos

Función:

Combinan la base de datos con hardware y herramientas de inteligencia comercial en una plataforma integrada (Data Channel, 2021).

Base de datos
+
Servidor
+
Almacén en nube
+
Sistema operativo
=
Datawre House o Almacén de datos

Características

Existen tres tipos:

Almacén de datos empresarial (Entrerprise Data Warehouse): almacén centralizado que ofrece soporte para la toma de decisiones a los diferentes departamentos de una empresa. Proporciona un enfoque unificado para organizar y representar datos.

Almacén de datos operativos (ODS): almacén de datos que se puede actualizar en tiempo real, lo que lo hace ideal para actividades rutinarias como el almacenamiento de registros de empleados.

Cubo de datos (Data Mart): diseñado para una línea de negocios específica, como finanzas, cuentas, ventas, compras o inventario. El almacén le permite recopilar datos directamente de las fuentes.

Maneja cuatro etapas:

Base de datos operativa fuera de línea: en esta etapa, los datos se copian a un servidor desde un sistema operativo para que la carga, el procesamiento y el informe de los datos no afecten el rendimiento del sistema operativo.

Almacén de datos sin conexión: los datos almacenados en el almacén se actualizan regularmente desde la base de datos operativa para obtener información comercial útil.

Almacén de datos en tiempo real: cada vez que se realiza una transacción en la base de datos operativa, la misma se actualiza en el almacén de datos.

Almacén de datos integrado: cada transacción que tiene lugar en la base de datos operativa se actualiza simultáneamente en el almacén de datos. Luego, el almacén genera transacciones que se envían a la base de datos operativa.

Ejemplos

Apache Hadoop
IBM Netezza Performance Server
Oracle Exadata
Snowflake
Vertica

4

Base de datos en memoria

Carga los datos de origen en la memoria del sistema en un formato comprimido y no relacional para agilizar el trabajo involucrado en el procesamiento de consultas. Los datos pasan por medio del caché antes de cada escritura y lectura para contar con la información en la memoria, dejando disponible la información que está solicitando con más frecuencia y así aumenta su velocidad de consulta.

Casos de uso

Un caso de uso relacionado es un informe de inteligencia comercial (BI) que requiere consultas complejas hacia una base de datos, como en una tienda. Cuando los usuarios comerciales revisan el informe con frecuencia, pero contiene información histórica que rara vez actualiza datos, un caché ubicado frente al almacén de datos puede ahorrar tiempo, ancho de banda de red, procesamiento y capacidad.

Ejemplos

Redis
SAP HANA Database
Tarantool
VoltDB

5

Base de datos de procesamiento paralelo masivo (MPP)

Función

Distribuye datos en un grupo de servidores, lo que permite que los sistemas compartan la carga de trabajo de procesamiento de consultas.

Características

Procesamiento: los nodos de procesamiento son los componentes fundamentales de un MPP. Los nodos pueden ser una PC de escritorio, un servidor, un servidor virtual, etc., o códigos de procesamiento simples con varias unidades de procesamiento instaladas.

Interconexión de alta velocidad: un sistema MPP normalmente divide las consultas en partes que luego se dispersan en los nodos. Cada nodo realizará su tarea de forma individual en el sistema de procesamiento paralelo. El proceso requiere una conexión de alto ancho de banda y comunicación centralizada entre nodos. Ethernet o cualquier interfaz de datos de fibra distribuida normalmente manejan esta interconexión de alta velocidad.

Bloqueo distribuido: el administrador de bloqueo distribuido permite compartir recursos cuando los nodos comparten espacio en disco entre ellos. Los nodos enviarán solicitudes de recursos al DLM y las asociarán cuando haya recursos disponibles. El DLM también ayuda a manejar las fallas de los nodos relacionadas con la inconsistencia de los datos y los problemas de recuperación

Ejemplos

BigQuery
Snowflake
Synapse
Amazon Redshift

Revisando cada uno de los tipos de bases de datos, puedes elegir, de acuerdo con el proyecto, presupuesto, empresa y datos que se tengan, el que puede brindar la solución para el momento actual e ir viendo a cuál ir evolucionando con el paso del tiempo. Los ejemplos de cada una de estas tecnologías dan una idea de cómo el mercado se va abriendo a estas posibilidades y nuevas opciones.

Cierre

puedes ver, se tiene una variedad de opciones en cuanto a bases de datos de analítica se refiere, siendo esta elección uno de los puntos clave del Departamento de Analítica, en conjunto con la arquitectura y el gobierno de datos.

Sin lugar a duda, el aprovechamiento de los datos es la piedra angular que las diferentes industrias han buscado aprovechar durante décadas. Finalmente, gracias a la consolidación del big data y su movimiento hacia la inteligencia artificial, se está en condiciones de poder lograrlo y realmente aprovechar los datos.

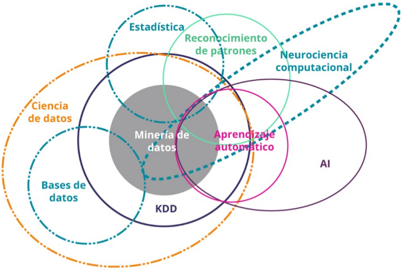


Figura 15. Big data, data mining y machine learning.
Fuente: López, J., y Zarza, G. (2017). *LA INGENIERÍA DEL BIG DATA: CÓMO TRABAJAR CON DATOS*. España: Editorial UOC.

El gran reto de las compañías radica en poder implementar soluciones que les permitan desarrollar y aprovechar esa piedra angular dentro del marco de procesos que pueden escalar, que sean eficientes y cuya utilización y mantenimiento no genere un nuevo problema en sí mismo.

Las dimensiones de la empresa, de los datos y de lo que se quiere lograr con ellos son los indicadores que deben guiar esta decisión, teniendo en cuenta toda la tecnología de la que ya se puede echar mano, así como de las experiencias previas.

Además, cada vez son más las empresas que, al ofrecer servicios en la nube, brindan esquemas de estos tipos con la flexibilidad que se requiere y con la capacidad indicada.

Ahora que conoces estos nuevos tipos de bases de datos, te puedes hacer las siguientes preguntas: ¿Cuál pudieras aplicar a tu proyecto? ¿Por qué? ¿Qué ventajas le daría?

Referencias bibliográficas

- Abdul, I. (2022). *What are MPP Systems? Benefits, Types and Examples*. Recuperado de https://www.royalcyber.com/blog/data-services/what-is-massively-parallel-processing-mpp/
- Altexsoft. (2021). *What is OLAP: A Complete Guide to Online Analytical Processing*. Recuperado de https://www.altexsoft.com/blog/olap-online-analytical-processing/
- Comunicación Numérica. (2020, 7 de septiembre). *Fundamentos del Análisis de Datos para Toma de Decisiones* [Archivo de video]. Recuperado de https://www.youtube.com/watch?v=qvZxvMWMvDo&t=1s
- Data Channel. (2021). *What is Data Warehousing? How it Works, Types, and General Stages*. Recuperado de https://datachannel.co/blogs/introduction-to-data-warehousing/
- Facultad de Administración y Ciencias Sociales - Universidad ORT Uruguay. (2018, 17 de julio). *Analítica de datos aplicada a negocios: ¿Es para mí? ¿Es ahora?* [Archivo de video]. Recuperado de https://www.youtube.com/watch?v=n7bwIqyhPa8
- HEAVY.AI. (s.f.). *Analytical Database Definition*. Recuperado de https://www.heavy.ai/technical-glossary/analytical-database
- Holistics. (2020). *The Analytics Setup Guidebook*. Recuperado de https://www.holistics.io/books/setup-analytics/a-modern-analytics-stack/
- IEP. (s.f.). *5 tipos de datos en el Big Data*. Recuperado de https://www.iep.edu.es/5-tipos-de-datos-en-el-big-data/#:~:text=El%20Big%20Data%20seg%C3%BAn%20su,manera%2C%20es%20m%C3%A1s%20f%C3%A1cil%20procesarlos.
- Ladredo, I. (2020). *Las 4 V´s del Big Data*. Recuperado de https://www.baoss.es/las-4-vs-del-big-data/#:~:text=La%20velocidad%20en%20big%20data,datos%20es%20masivo%20y%20continuo.
- López, J., y Zarza, G. (2017). *LA INGENIERÍA DEL BIG DATA: CÓMO TRABAJAR CON DATOS*. España: Editorial UOC.
- López, M. (2021). *¿Qué debemos tener en cuenta para implementar modelos de big data?* Recuperado de https://www.pragma.com.co/blog/que-debemos-tener-en-cuenta-para-implementar-modelos-de-big-data
- Rovira, P., y Pascual, V. (2021). *ANALÍTICA VISUAL. CÓMO EXPLORAR, ANALIZAR Y COMUNICAR DATOS*. España: ANAYA MULTIMEDIA.
- Sosa, W. (2019). *Big data: Breve manual para conocer la ciencia de datos que ya invadió nuestras vidas.* Argentina: Siglo XXI.
- Tokio School. (2021). *Fases Big Data: conoce todo el proceso*. Recuperado de https://www.tokioschool.com/noticias/fases-big-data/

Para saber más

Los siguientes enlaces son externos a la Universidad Tecmilenio, al acceder a ellos considera que debes apegarte a sus términos y condiciones.

Videos

Para conocer más acerca de **bases de datos analíticas**, te sugerimos ver lo siguiente:

- IGERENCIA. (2022, 10 de enero). *¿Qué viene para el sector de analítica de datos en el 2022?* [Archivo de video]. Recuperado de https://youtu.be/rC4egM3fAWc?t=59
- Salvador Ramos - Power BI. (2022, 14 de marzo). *SQL vs Power BI [Descubre las diferencias]* [Archivo de video]. Recuperado de https://www.youtube.com/watch?v=d5hlZWol-uk

Lecturas

Para conocer más acerca de **bases de datos analíticas**, te sugerimos leer lo siguiente:

- Holistics. (2020). *The Analytics Setup Guidebook*. Recuperado de https://www.holistics.io/books/setup-analytics/a-modern-analytics-stack/
- Kryptonsolid. (2021). *¿Qué es una base de datos analítica?* Recuperado de https://kryptonsolid.com/que-es-una-base-de-datos-analitica/

Checkpoint

Asegúrate de:

- Distinguir las características que debe tener una base de datos dedicada a la analítica.
- Identificar los tipos de bases de datos para analítica que hay en el mercado actualmente.
- Reconocer qué empresas y programas ya ofrecen este tipo de bases de datos para poder adquirirlas.

"Tecmilenio no guarda relación alguna con las marcas mencionadas como ejemplo. Las marcas son propiedad de sus titulares conforme a la legislación aplicable, estas se utilizan con fines académicos y didácticos, por lo que no exísten fines de lucro, relación publicitaria o de patrocinio".

La obra presentada es propiedad de ENSEÑANZA E INVESTIGACIÓN SUPERIOR A.C. (UNIVERSIDAD TECMILENIO), protegida por la Ley Federal de Derecho de Autor; la alteración o deformación de una obra, así como su reproducción, exhibición o ejecución pública sin el consentimiento de su autor y titular de los derechos correspondientes es constitutivo de un delito tipificado en la Ley Federal de Derechos de Autor, así como en las Leyes Internacionales de Derecho de Autor.

El uso de imágenes, fragmentos de videos, fragmentos de eventos culturales, programas y demás material que sea objeto de protección de los derechos de autor, es exclusivamente para fines educativos e informativos, y cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por UNIVERSIDAD TECMILENIO.

Queda prohibido copiar, reproducir, distribuir, publicar, transmitir, difundir, o en cualquier modo explotar cualquier parte de esta obra sin la autorización previa por escrito de UNIVERSIDAD TECMILENIO. Sin embargo, usted podrá bajar material a su computadora personal para uso exclusivamente personal o educacional y no comercial limitado a una copia por página. No se podrá remover o alterar de la copia ninguna leyenda de Derechos de Autor o la que manifieste la autoría del material.