

Introducción al Almacenamiento de Datos

Análisis de datos

Objetivos de la Presentación

- 1. Introducir los conceptos clave del almacenamiento de datos.
- 2. Explicar los diferentes tipos de sistemas de almacenamiento.
- 3. Resaltar la importancia del almacenamiento de datos en el análisis.
- 4. Analizar las tecnologías actuales de almacenamiento.

¿Qué es el Almacenamiento de Datos?

- - Proceso de guardar y gestionar grandes cantidades de información.
- - Permite su posterior recuperación y análisis eficiente.
- - Soporte para el análisis de datos y la toma de decisiones.

Tipos de Almacenamiento de Datos

- 1. Bases de Datos Relacionales
- 2. Bases de Datos NoSQL
- 3. Data Warehousing
- 4. Data Lakes
- 5. Almacenamiento en la Nube

Bases de Datos Relacionales

- - Modelo estructurado basado en tablas.
- - Ejemplos: MySQL, PostgreSQL, SQL Server.
- - Características: ACID (Atomicidad, Consistencia, Aislamiento, Durabilidad).
- - Ideal para datos transaccionales.

Bases de Datos NoSQL

- - Datos no estructurados o semi-estructurados.
- - Ejemplos: MongoDB, Cassandra, Redis.
- - Características: CAP Theorem (Consistencia, Disponibilidad, Tolerancia a Particiones).
- - Útil para alta escalabilidad y flexibilidad.

Data Warehousing (Almacenes de Datos)

- - Almacena grandes volúmenes de datos históricos.
- - Optimizado para análisis de datos a gran escala.
- - Ejemplos: Amazon Redshift, Google BigQuery, Snowflake.
- - Esquemas estructurados: Esquema estrella, copo de nieve.

Data Lakes

- - Repositorios centralizados de datos sin transformar.
- - Almacenan datos estructurados y no estructurados.
- - Ejemplos: Hadoop, Azure Data Lake, Amazon S3.
- - Flexibilidad en el manejo de datos en su forma nativa.

Almacenamiento en la Nube

- - Uso de plataformas en la nube: AWS, Google Cloud, Azure.
- - Ventajas: Elasticidad, escalabilidad, y facilidad de acceso.
- - Seguridad y replicación de datos.

Proceso ETL (Extract, Transform, Load)

- - Extracción: Obtención de datos de diversas fuentes.
- - Transformación: Limpieza y ajuste de los datos.
- - Carga: Almacenamiento en un sistema centralizado (data warehouse o data lake).
- - Herramientas: Apache Nifi, Talend, Informatica.

Importancia del Almacenamiento en el Análisis de Datos

- - Acceso rápido a datos correctos en el momento adecuado.
- - Impacto en rendimiento, escalabilidad y costos.
- - Base fundamental para la toma de decisiones basada en datos.

Conclusiones

- 1. El almacenamiento de datos es crucial en el análisis de datos.
- 2. La selección del tipo adecuado de almacenamiento depende de las necesidades de la organización.
- 3. La evolución hacia soluciones en la nube y sistemas distribuidos ofrece mayor flexibilidad y escalabilidad.