

Clase 12 de Febrero

viernes, 12 de febrero de 2021

12:11 p. m.

Profesor Christian Socha

Link de meet: <https://meet.google.com/hnt-zeaq-pix>

Q:

Temas revisados:

- ¿Qué es una base datos?
- ¿Qué ventajas tiene una base datos?
- ¿Qué tipos de base de datos hay?
- ¿Qué papel cumplen las BD en los sistemas de información (SI)?
- ¿Qué es un diccionario de datos? ¿Cuáles son los roles en las BD?

A:

Se realizaron las preguntas en grupos o individualmente.

https://docs.google.com/document/d/1R-OVW108_xajpolw1JZmKhv1Wj0cSIafee2w7Z2TAn8/edit?usp=sharing

Actividad para la próxima clase:

- ¿Qué es Data Cloud?
- ¿Qué es una bodega de datos?
- ¿Qué es un Datamart?
- ¿Qué es un proceso ETL?
- ¿Qué es Big Data?
- ¿Funcionamiento del Big Data?
- Historia de Big Data

¿Qué son las Bases de Datos?

En informática las bases de datos son una biblioteca digital de información organizada y clasificada para su uso según sea requerido. Ya sea para únicamente ser visualizada de manera más fácil o manejada y modificada de ser necesario.

¿Qué ventajas tienen las Bases de Datos?

La ventaja principal que tiene el la forma dinámica de visualizar y modificar la información en comparación de los libros en físico que requieren más tiempo para ser revisados y hacer modificaciones suele ser tedioso. También se descarta información redundante que únicamente afecta la eficiencia del manejo de datos.

¿Qué tipos de DBs existen?

Hay bases de datos estáticas y dinámicas.

La estática tiene la información fija para ser visualizada junto con sus cambios a lo largo del tiempo, con el tiempo se alimenta de más datos y estos se pueden usar para crear un modelo de negocios o una estrategia de mercado de una empresa. La dinámica permite que su información vaya modificándose con el tiempo. También las hay por su modo de administración: **DBs Jerárquicas**, de **red**, o **relacionales**.

Base de datos relacional

La base de datos relacional es una recopilación de la información empresarial organizada de tal forma que se puede consultar, actualizar, analizar y sacar los datos fácilmente. La información se encuentra en tablas y campos relacionados entre sí.

Base de datos distribuida

La base de datos distribuida consiste en el almacenamiento de porciones de la base de datos en diferentes ubicaciones físicas y, por tanto, el procesamiento está distribuido o replicado entre los distintos puntos de una red de trabajo.

Base de datos NoSQL

El nombre de la base de datos **NoSQL** proviene de **Not only SQL** o, en español, no solo SQL. Esto se debe a que este tipo de base de datos suele evitar el uso del SQL o lo usa de apoyo, pero no como consulta. El hecho de evitar el SQL es porque se usa para proyectos en los que se necesita trabajar en la base de datos con un gran volumen. En las bases de datos con lenguaje SQL, los distintos atributos de un elemento, están en diferentes columnas, mientras que en una **NoSQL** todos los atributos se encuentran en una misma columna, ahorrando espacio.

- <https://www.ticportal.es/glosario-tic/base-datos-database>

¿Qué papel cumplen las DBs en los Sistemas de Información?

El papel principal es facilitar a las personas el manejo de la información con un almacenamiento organizado y relacionable.

¿Qué es un diccionario de datos?

Es un repositorio único para todo el proyecto de desarrollo en un nuevo sistema, este ayuda a conceptualizar el sistema y la forma en que necesitara trabajar. Permite tener una notación formal del contenido de los flujos y estructuras de datos. Básicamente el DD recopila definiciones de todos los datos que se manejan en el sistema

¿Cuáles son los roles en las DBs?

Un rol es el nivel de privilegio que tiene un usuario en la DB para visualizar o modificar información del sistema. Los principales son el administrador, el diseñador y el usuario final.

¿Qué es Data Cloud?

Cloud Computing y Big Data:

La **Big Data** se refiere a un conjunto de información cuyo **volumen**, **complejidad**, **variabilidad** y **velocidad de crecimiento** dificultan su **captura**, **gestión**, **procesamiento** y **análisis** mediante las herramientas convencionales como bases de datos relacionales y estadísticas convencionales.

La importancia de la Big Data yace en su análisis ya que ayuda a las organizaciones a aprovechar sus datos e identificar nuevas oportunidades en las tendencias y comportamiento de los usuarios y consumidores de determinada organización.

El **Cloud computing** es el entorno bajo el cual almacenamos las base de datos o Big Data para procesarla según se requiera. Este es el método mas común de almacenarla porque abarata los costos en comparación de tener servidores propios en las oficinas, que ocupan espacio físico, con el tiempo deben ser reemplazados, y requieren también de mantenimiento, para estos se utiliza el termino **On premise** (on prem).

¿Qué es una bodega de datos?

Una bodega de datos o Data **Warehouse** es un repositorio físico o lógico(Cloud) donde se almacenan todos los datos del punto anterior para su mantenimiento y revisión, tiene tres estructuras:

Básica: Sistemas operativos y archivos planos que proporcionan datos y metadatos en bruto. El **usuario final** puede acceder para su análisis, generación de informes y minería.

Básica con un área de ensayo: Proporciona un lugar generalmente en el almacén donde los datos se limpian antes de ingresar, es posible personalizar la arquitectura para diferentes grupos dentro de la organización.

Agregando Data **marts**: Son sistemas diseñados para líneas de negocios particulares. Los puede haber para ventas, inventario y compras y así el usuario final puede acceder desde cada departamento.

¿Qué es un Datamart?

Es un almacén de datos orientado a un área en específico como puede ser **distribucion**, inventarios, recursos humanos etc. Estos datos están almacenados en las data **warehouses**.

¿Qué es un proceso ETL?

Es el proceso en el cual los datos se analizan en una sola fuente, el termino describe las tres fases:

E: **Extraction**; Aquí se capturan todos los datos de las diversas fuentes que se dispongan y se centralizan.

T: **Transformation**; En este punto se estandarizan los datos, algo denominado Data **Cleaning** así parece que los datos fueran de una sola fuente para su fácil manejo. Para esto se escribe un algoritmo que nos ayude a hacerlo o también nos podemos apoyar en el machine **learning**.

L: Load; O carga es donde ya han sido extraídos esos datos y se almacenan en un Data **warehouse** para su posterior análisis.

Historia del Big Data:

Se inicia la escritura en Mesopotamia y Egipto en el 3100 AC. Desde ese momento las personas empiezan a darle gran importancia a la información porque es una manera de llevar registro de las cosas que hacían.

➤ La información se registraba en tablillas, papiros y jeroglíficos. Ahora los almacenamos en la nube o discos duros.

Freemont Ride bibliotecario de la Wesleyan University, en 1944 publico un libro donde estimaba que las bibliotecas universitarias estadounidenses duplicarían en tamaño cada dieciséis años.

Luego de Ride hubo muchas otras personas que Utilizaron los datos que se habían almacenado para sacar conclusiones o dar explicaciones de lo que se podía conseguir gracias a ellos.

Estas fechas facilitaron la aparición del Big Data:

- 1950: Aparece la memoria de tambor magnético.
- 1956: Memoria Virtual
- 1974: Aparición de SQL.
- 1985: Bases de datos relacionales
- 1989: Business Intelligence
- 1995: Expansión del World Wide Web (WWW)
- 1996: Business Intelligence 2.0
- 1999: IoT (Internet of Things)
- En 2013 se hizo el lanzamiento de IPv6 por la demanda de Ips
- 2014 Se lanza Apache Spark
-