

GUÍA DE LA SESIÓN DE ANALÍTICA DE DATOS

LÍNEA ANALÍTICA DE DATOS		
II. NÚCLEO ESPECÍFICO		Fundamentos y potencialidades del Machine Learning (ML)
Sesión #: 07	Conceptos de base de Big Data: Ejemplos de aplicación de Big Data a casos a casos prácticos de la vida real	Tiempo: 3 horas
Observaciones especiales: N/A		Fecha: 21/09/2021

Con antelación verificar si se cuenta con:

Aula virtual	X	Accesos	X	Presentación	X	Recursos		Cámara y micrófono	X
--------------	---	---------	---	--------------	---	----------	--	--------------------	---

Con antelación verificar si se cuenta con:

Aula virtual	X	Accesos	X	Presentación	X	Recursos		Cámara y micrófono	X
--------------	---	---------	---	--------------	---	----------	--	--------------------	---

- *Navegador web - Preferiblemente Google Chrome.*

Objetivos de la sesión

- Conocer el origen del big data
- Conocer las 3 v's del big data (volumen, variedad, velocidad)
- Comprender el teorema CAP (consistencia, disponibilidad, tolerante a la partición)
- Identificar las distintas bases de datos (Relacionales, MongoDB, Cassandra)
- Relacionar las distintas bases de datos con los posibles usos dentro de la industria

Introducción al tema

En esta sesión se presentará un contexto general del origen del concepto de Big Data, de su fundamentación desde el software y desde los técnicas más comunes para el procesamiento de información de grandes volúmenes de datos. Se explorarán las bases de datos más relevantes desde el punto de vista práctico y se explicará el fundamento teórico del teorema CAP (consistencia, disponibilidad, tolerante a la partición) para dar contexto y claridad de la naturaleza de las bases de datos para grandes volúmenes de datos. También se dará un contexto general del big data así como sus usos en empresas de la industria. En la fundamentación conceptual es importante entender las distintas bases de datos y sus posibles usos. Las posibles bases de datos a explicar son:

- Base de datos relacionales
- MongoDB (Bases de datos documentales)
- Cassandra (Base de datos tipo Columna)
- HBase (Base de datos tipo Columna)
- Neo4J (Base de datos de tipo grafo)

Método: Propuesta pedagógica

La sesión se realizará de forma sincrónica a distancia a través de la plataforma TEAMS o Google Meet.

- **Primera parte:** Presentación a los participantes de los Conceptos base de grandes volúmenes de datos, se espera que en el transcurso de la presentación se cuente con interacción que permita enfocar las técnicas en problemas particulares de las empresas.

Duración: 1 horas.

- **Segunda parte:** Presentación del teorema CAP (consistencia, disponibilidad, tolerante a la partición) y su utilización en la selección de las distintas bases de datos. Explicación de algunas de las bases de datos más importantes del mercado.

Duración: 0.5 horas.

- **Tercera parte:** Los participantes se dirigen a sus respectivos grupos con su asesor asignado. El asesor contextualizará y profundizará el tema de la sesión, y responderá a las preguntas de los participantes que tengan relación con la conferencia. Se realizarán las actividades previstas con acompañamiento del asesor a través de la práctica-búsqueda.

Duración: 1.5 horas.

Habilidades y competencias desarrolladas

- Conocer y comprender los fundamentos de grandes volúmenes de datos
- Diferenciar las características principales de las bases de datos sql y no-sql
- Identificar el campo de aplicación de cada base de datos sql y no-sql

SECCIÓN PRÁCTICA

Recursos del taller

- *Navegador web - Preferiblemente Google Chrome.*

Instrucciones

Actividad 1

Utilizar la herramienta gráfica de kahoot con un cuestionario de 10 preguntas acerca del Big Data

- <https://create.kahoot.it/details/514a455b-be0f-40f3-a265-8770c54db06b>

Actividad 2

- Identificar con 4 casos de uso que tipo de base de datos utilizar

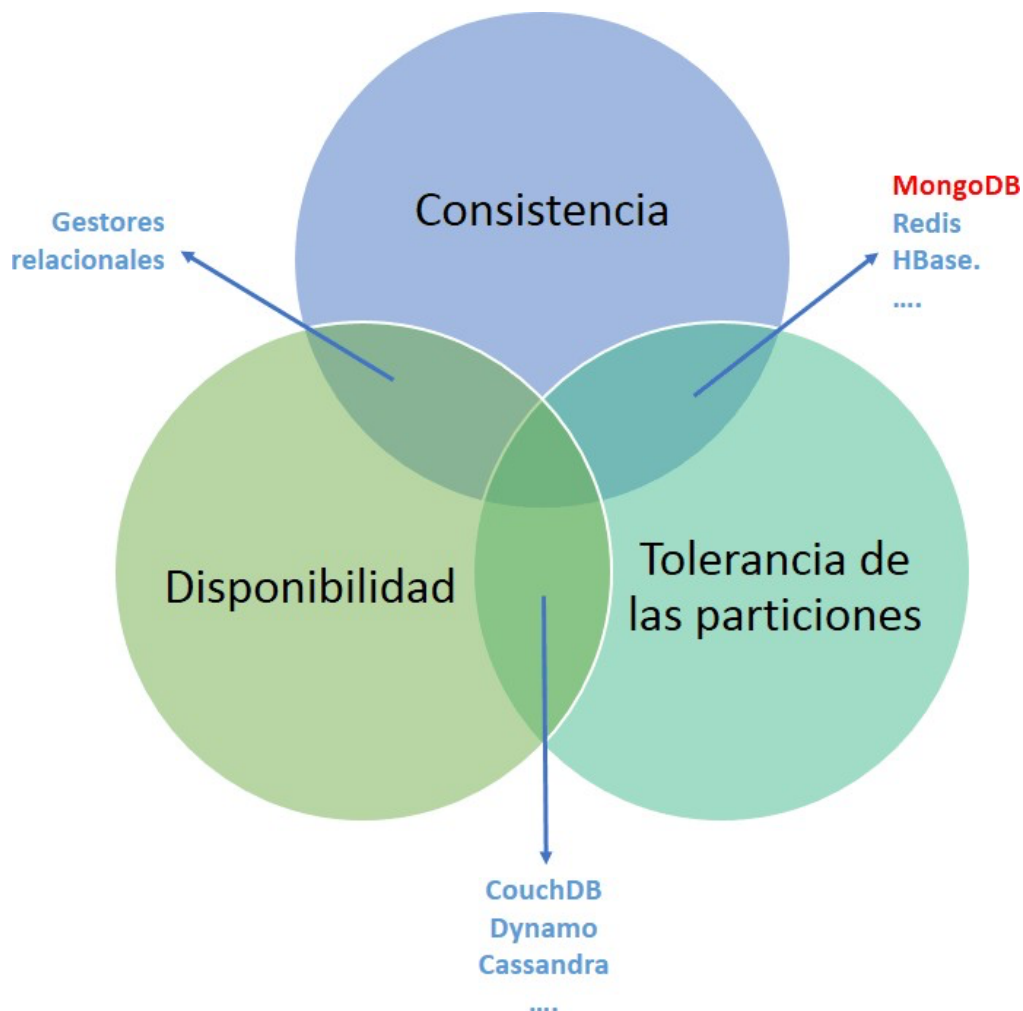
Actividad 3

- Realizarme preguntas acerca del estado mi empresa con respecto a grandes volúmenes de datos.

BALANCE DEL DÍA

- Los participantes identificarán y comprenderán el uso de grandes volúmenes de datos.
- Se conocerán las aplicaciones más comunes y los problemas que se resuelven con estas técnicas.
- Se identificará un problema de la empresa o los interesados que pueda ser resuelto o optimizado con una base de datos sql o no-sql

Teorema CAP



Preguntas para participantes:

- ¿Hoy en día cómo almaceno mis datos?
- ¿Necesito una base de datos relacional?
- ¿Mis datos se generan en un gran volumen de datos?
- ¿Mis datos tienen una gran variedad?
- ¿Mis datos se generan a gran velocidad?
- ¿Qué infraestructura tengo hoy para enfrentar el Big Data?
- ¿Qué desafíos tengo dentro de mi empresa para afrontar el Big Data?
- ¿Nos conviene utilizar una base de datos particionada?
- ¿Debe mantener mi empresa consistencia en los datos al particionarlos?
- ¿Qué nos haría falta para iniciarnos en el mundo del Big Data?
- ¿Necesitaría dinero externo a mi compañía para iniciar en Big Data?
- ¿Me conviene utilizar servicios en la nube para almacenar datos?

Base de datos relacional:

Mis datos son de tipo transaccional. Mi software se encarga de llevar el inventario de una compañía así como las ventas y compras. En cada momento se debe garantizar la integridad de los datos y no puedo permitir que un usuario reciba información desactualizada acerca del inventario.

¿Qué tipo de base de datos debo utilizar?

¿Qué partes del teorema CAP cumple esta base de datos?

¿Debo almacenar esta información en la nube o lo puedo hacer en mis servidores locales?

MongoDB:

Mis datos son transaccionales. Todos los usuarios deben recibir la misma información en cada momento. Se generan muchísimos datos por lo que un sólo servidor no sería suficiente para aceptar el gran volumen de datos.

¿Qué tipo de base de datos debo utilizar?

¿Qué partes del teorema CAP cumple esta base de datos?

¿Debo almacenar esta información en la nube o lo puedo hacer en mis servidores locales?



Cassandra:

Mis datos no necesitan transaccionalidad. Puedo permitir que un usuario reciba información desactualizada acerca del estado de la aplicación. Se generan muchísimos datos por lo que un sólo servidor no sería suficiente para aceptar el gran volumen de datos.

¿Qué tipo de base de datos debo utilizar?

¿Qué partes del teorema CAP cumple esta base de datos?

¿Debo almacenar esta información en la nube o lo puedo hacer en mis servidores locales?

