

MASTERCLASS ORACLE

1. ¿Qué es Big Data? Big Data es en el sector de tecnologías de la información y la comunicación hace referencia a grandes conjuntos de datos que por la velocidad a la que se generan; las múltiples fuentes y formatos; y la capacidad de procesos; es realmente complicado trabajar con ellos. recogerlos · almacenarlos · buscar · compartir · analizar · visualizar · procesar · entender.

2. Historia de Big Data. Diversos puntos por los que ha ido pasando esta tecnología.

3. ¿Qué es Oracle Big Data? Solución integrada por Oracle para trabajar con Cloudera Hadoop y un ecosistema específico dentro del mundo Oracle.

4. ¿El Volumen importa? Qué cantidad de información puedo manejar mediante Big Data y qué tipo de información es propensa a ser procesada mediante este modelo.

5. ¿La velocidad importa? A qué velocidad puedo procesar la información, necesito ser rápido a la hora de procesar los datos?

6. ¿La variedad importa? Diferentes orígenes y destinos de los datos. Como parsear la información entre los diferentes orígenes de datos y su posterior composición.

7. ¿Por qué el Big Data es un término tan pegadizo desde 2012?

8. Qué elementos actuales se pueden beneficiar del Big Data. Lista de ejemplo genéricos, tipo redes sociales, procesamiento de televisión y vídeo en general, análisis de información (telefonía), Marketing, reconocimiento facial, ocr, etc. (Imágenes genéricas de casos más generales en el mundo Big Data)

9. ¿Cuánta información digital se genera en el mundo? Ciento diez mil millones de iPads de 16 GB completamente llenos. Ésa es la cantidad de datos que EMC e IDG calculan que se habrá producido en el mundo en 2011. O, en otros términos, 1,8 zettabytes (1,8 por 10 elevado a 21 bytes). Y se espera que la cantidad de datos almacenados se duplique cada dos años. En 2015 estaremos produciendo yottabytes (1 por 10 elevado a 24). Ya se están buscando nuevos nombres para los múltiplos que se espera ver llegar. 1 ZB = 10³ EB = 10⁶ PB = 10⁹ TB = 10¹² GB = 10¹⁵ MB = 10¹⁸ KB = 10²¹ bytes.

10. La información es poder. Un informe del Foro Económico Mundial declaró que los datos constituyen una nueva clase de activo económico, como la moneda o el oro.

11. ¿Cuál es la complejidad de recoger esos datos? La variedad de formatos implica que haya datos no estructurados y que no siguen las características ACID (Atomicity, Consisten-

cy, Isolation, Durability) como fotos, vídeos o información proveniente de redes sociales.

Esto genera la necesidad de nuevas tecnologías y mecanismos para poder alojarlos y poder ser procesados.

12. Principal estructura de Oracle Big Data basada en Cloudera Hadoop. MapReduce Framework de hardware distribuido (cluster o grids) que divide los problemas en subproblemas (Map) y luego recopilas las mini-respuestas (Reduce) para generar conclusiones. La solución más común es Hadoop. Este modelo fue creado y promovido por Google.

13. Sql y NoSQL Database. NoSQL (tipos de bases de datos). Amplia clase de sistemas de gestión de bases de datos que difieren del modelo clásico del sistema de gestión de bases de datos relacionales (RDBMS) en múltiples aspectos: no usan SQL como el principal lenguaje de consultas, los datos almacenados no requieren estructuras fijas como tablas, no garantizan ACID y escalan bien

horizontalmente (ej:MongoDB, Cassandra, BigTable)

14. Estructura de Cluster dentro de Big Data y clusters NoSQL. Cluster de aplicación, Big Data y NoSQL.

15. Oracle Big Data y Linux. Componentes. Elementos con los que se trabaja en Cloudera Hadoop de manera general.

Flume.

Sqoop.

Impala

Hive.

Spark

16. Qué aporta Oracle Big Data. Aportaciones propias de Oracle al mundo Big Data y Clou-dera.

Conectores.

Librerías específicas.

17. Casos Prácticos. Ejemplos prácticos con máquina virtual de lo que se puede llegar a hacer.

1. Twitter. Colectar tweets para recabar información.
2. Reconocimiento Facial, reconocer a un personaje en un video.
3. Mover datos entre diferentes plataformas de bases de datos.