

Examen para BDA01.

Intento 1.

Pregunta 1.

¿Qué componente del ecosistema Hadoop permite acceder a los ficheros de HDFS con lenguaje SQL, como si fueran tablas de una base de datos?:

- a. **Hive.**
- b. MapReduce.
- c. Pig.
- d. Kafka.

Pregunta 2.

Storm permite procesar datos que llegan en streaming. ¿Verdadero o falso?

Seleccione una:

Verdadero

Falso

Pregunta 3.

¿Cuál es la herramienta o componente del ecosistema Hadoop con la que los Data Scientists realizan análisis interactivos?:

- a. Ambari.
- b. Hive.
- c. HDFS.
- d. **Zeppelin.**

Pregunta 4.

¿Qué tipo de nodo en Hadoop tiene discos en RAID?:

- a. Los nodos worker.
- b. **Los nodos master.**
- c. Ningún nodo.
- d. Todos los nodos.

Pregunta 5.

La arquitectura de Hadoop se basa en tener nodos worker, que hacen el trabajo, y nodos master que coordinan ese trabajo. ¿Verdadero o falso?

Pregunta 6.

Hadoop requiere hardware commodity, lo que significa que puedo montarme un clúster en casa con los PCs obsoletos que tengo. ¿Verdadero o falso?

Seleccione una:

Verdadero

Falso

Pregunta 7.

Hadoop no es un sistema eficiente para casos de uso de poca volumetría ¿Verdadero o falso?

Seleccione una:

Verdadero

Falso

Pregunta 8.

El coste de Hadoop es similar al de las bases de datos relacionales, la diferencia está en que Hadoop es escalable. ¿Verdadero o falso?

Seleccione una:

Verdadero

Falso

Pregunta 9.

Google está utilizando Hadoop para la generación de los índices de búsqueda ¿Verdadero o falso?

Seleccione una:

Verdadero

Falso

Pregunta 10.

Hadoop es un proyecto liderado por la empresa Cloudera. ¿Verdadero o falso?

Seleccione una:

Verdadero

Falso

Intento 2.

Pregunta 1.

Aproximadamente, cuánto costaría una infraestructura Hadoop de 10 nodos worker, 2 nodos master y 1 nodo frontera?:

- a. 1.000.000 euros.
- b. 10.000 euros.
- c. 300.000 euros.
- d. **100.000 euros.**

Pregunta 2.

Hadoop cubre bien los casos de uso donde se quiere analizar datos con algoritmos de Machine Learning. ¿Verdadero o falso?

Seleccione una:

Verdadero

Falso

Pregunta 3.

Los nodos frontera realizan parte del trabajo de procesamiento con los datos en una plataforma Hadoop. ¿Verdadero o falso?

Seleccione una:

Verdadero

Falso

Pregunta 4.

¿Cuál es el componente que se utiliza para planificar, orquestar o automatizar flujos de trabajo en Hadoop?:

- a. YARN.
- b. Hive.
- c. **Oozie.**
- d. Impala.

Pregunta 5.

¿Cuál de las siguientes distribuciones Hadoop se encuentra activa hoy en día:

- a. Sólo existen distribuciones en cloud.
- b. **Cloudera.**
- c. MAPR.
- d. Hortonworks.

Pregunta 6.

¿Cómo se consigue escalar o ganar mayor capacidad en una plataforma Hadoop?:

- a. **Añadiendo más nodos worker al clúster.**
- b. Cambiando los servidores por otros con CPUs más potentes.
- c. Añadiendo más nodos master al clúster.
- d. Realizando optimizaciones en los sistemas operativos.

Pregunta 7.

Sobre el tipo de red en el que conectar Hadoop, ¿qué afirmación es correcta?:

- a. Al necesitar hardware commodity, no importa el tipo de red.
- b. **La red debería tener un buen ancho de banda, por ejemplo 10-20 Gbps.**
- c. Hadoop no necesita red de comunicaciones.
- d. No importa el ancho de banda de la red, pero sí que tenga un ping muy bajo.

Pregunta 8.

¿Qué cantidad de memoria es recomendada para la instalación de Hadoop?:

- a. 64 Gb en los nodos worker y 256 Gb en los nodos master.
- b. Al menos 1 Tb en todos los nodos.
- c. **Se recomienda 256 Gb de memoria RAM en todos los nodos.**
- d. Al necesitar hardware commodity, no importa la cantidad de memoria.

Pregunta 9.

Hadoop es la única tecnología capaz de almacenar y procesar grandes volúmenes de datos hasta un volumen de petabytes. ¿Verdadero o falso?

Seleccione una:

Verdadero

Falso

Pregunta 10.

Se suele considerar a Doug Cutting como el "padre de Hadoop". ¿Verdadero o falso?

Seleccione una:

Verdadero

Falso