

## Examen para BDA03

### Intento 1

#### Pregunta 1

Hive es escalable porque pueden añadirse más nodos HiveServer2 para que trabajen en paralelo, ejecutando una query en los diferentes HiveServer2. ¿Verdadero o falso?

Seleccione una:

Verdadero

**Falso**

#### Pregunta 2

Apache Spark es una tecnología que sustituye a Hadoop. ¿Verdadero o falso?

Seleccione una:

Verdadero

**Falso**

#### Pregunta 3

¿Cuál de las siguientes afirmaciones sobre Impala es correcta?

- a. **Permite acceder a los datos de HDFS como si fuera una tabla.**
- b. Permite administrar un clúster Hadoop.
- c. Permite importar datos de otras bases de datos relacionales.
- d. Ninguna de las anteriores es correcta.

#### Pregunta 4

¿Cuál es la sentencia de Hive con la que se crea una nueva tabla?:

- a. NEW TABLE.
- b. MAKE TABLE.
- c. INSERT.
- d. **CREATE TABLE.**

#### Pregunta 5

¿Cómo se llama la principal abstracción de datos de Apache Spark:

- a. File.
- b. Dataset.
- c. **RDD.**
- d. Dataframe.

#### Pregunta 6

Apache Spark tiene una latencia mayor que Apache Flink como norma general para procesamiento en tiempo real. ¿Verdadero o falso?

Seleccione una:

**Verdadero**

Falso

### Pregunta 7

Una partición de una tabla es un subconjunto de los registros de la tabla que comparten el valor de alguna columna en común. ¿Verdadero o falso?

Seleccione una:

**Verdadero**

Falso

### Pregunta 8

Apache Spark puede leer y procesar datos que se encuentran en una base de datos relacional. ¿Verdadero o falso?

Seleccione una:

**Verdadero**

Falso

### Pregunta 9

Para automatizar la ejecución de trabajos que se debe realizar en Hadoop, por ejemplo, para validar los datos ingestados, ¿qué componente se debe utilizar?

- a. Pig.
- b. YARN.
- c. Zookeeper.
- d. **Oozie.**

### Pregunta 10

Hive permite conectar Excel con ficheros que están en HDFS. ¿Verdadero o falso?

Seleccione una:

**Verdadero**

Falso

### Intento 2

#### Pregunta 1

¿Cuál es la sentencia de Hive con la que añadimos un registro en una tabla dándole los valores de los campos en la misma sentencia?

- a. LOAD DATA.
- b. **INSERT INTO.**
- c. ADD DATA.
- d. ADD INTO.

#### Pregunta 2

HBase ofrece un interfaz de consultas parecido a SQL, llamado HQL. ¿Verdadero o falso?

Seleccione una:

Verdadero

**Falso**

### Pregunta 3

Si quiero parar un nodo de un clúster Hadoop porque está dando problemas, ¿qué herramienta podré utilizar?

- a. Ambari.
- b. Oozie.
- c. Hue.
- d. YARN.

### Pregunta 4

¿Cómo se llama el componente que ofrece, desde una web, acceso a los ficheros de HDFS y poder lanzar consultas Hive?

- a. Hue.
- b. Hive.
- c. Impala.
- d. Spark.

### Pregunta 5

Hive es una herramienta muy útil, pero tiene como aspecto negativo es que toda la seguridad la delega en los permisos de los ficheros de HDFS. ¿Verdadero o falso?

Seleccione una:

Verdadero

Falso

### Pregunta 6

¿Cuál de las siguientes funcionalidades no es ofrecida por Apache Spark?:

- a. Computación en paralelo de procesos de transformación de datos.
- b. Algoritmos y utilidades para manejar datos como grafos.
- c. **Una base de datos NoSQL.**
- d. Interfaz para interactuar con programas escritos en R.

### Pregunta 7

¿Desde qué componente de Hadoop se puede enviar un email al administrador cuando una ingesta automática ha fallado?

- a. Cloudera Manager.
- b. Spark.
- c. Ambari.
- d. **Oozie.**

### Pregunta 8

Apache Flink es una buena herramienta de procesamiento en tiempo real, pero tiene como principal inconveniente que no se puede ejecutar sobre YARN. ¿Verdadero o falso?

Seleccione una:

Verdadero

Falso

#### Pregunta 9

¿Qué componente del ecosistema Hadoop permite utilizar sintaxis SQL para manejar datos que están almacenado en HBase?

- a. Oozie.
- b. **Phoenix.**
- c. Storm.
- d. Pig.

#### Pregunta 10

Para ingestar datos que se generan en tiempo real, ¿qué componente de Hadoop se utiliza?

- a. Storm.
- b. Spark.
- c. Flink.
- d. **Flume.**