

## Sección primera. HADOOP

1. ¿Cuáles de las siguientes opciones incluyen componentes de YARN con su cardinalidad correcta?

- **NodeManager, uno por cada nodo del cluster.**
- ResourceManager, uno por cada tarea que se ejecute. FALSO
- ClusterManager, uno por cluster. FALSO

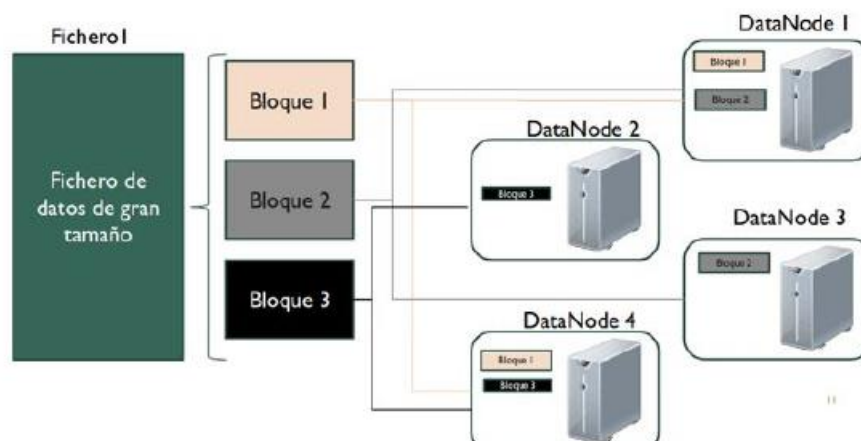
2. ¿Para qué sirve el “rack-awareness”?

- Para evitar que problemas en un “rack” afecten a todas las copias de un bloque.**
- Para minimizar las distancias entre nodos en la red. FALSO

3. Para averiguar el tamaño de una carpeta almacenada en el cluster HDFS, ¿qué comando debemos utilizar?

- Ninguna de las restantes respuestas.
- **hdfs dfs -du <carpeta>**
- hadoop -getsize <carpeta>
- hfs -size <carpeta>

4. En el sistema de la figura, ¿cuál será el contenido del NameNode de HDFS en referencia al bloque 1?



**Ninguna de las restantes respuestas.**

- Nodo 2, nodo 3.
- Nodo 2, nodo 4.
- Nodo 1, nodo 2.

2. ¿Cuál de los siguientes parámetros es obligatorio especificar al realizar una exportación de datos desde HDFS hacia una base de datos relacional utilizando Sqoop?

**--connect.**

--password. FALSO

--username. FALSO

#### Sección segunda. MR

1. ¿Cuál de las siguientes opciones es cierta sobre MapReduce + HDFS?

I. Ejecuta aplicaciones arbitrarias. VERDADERO

II. No tolera fallos, reinicia tareas por un fallo. FALSO

2. ¿Cuál es el protocolo de salida por defecto de mrjob?

- JSONProtocol

3. ¿Cuál de las siguientes afirmaciones es cierta sobre MapReduce?

I. Es deseable llevar la complejidad de los cálculos a la función reduce siempre que sea posible.

II. La salida del reduce es la entrada del map, donde cada par <clave ,valor> se asigna a un map aleatorio.

- I. Falsa. II. Falsa.

- I. Correcta. II. Falsa.

- I. Correcta. II. Correcta.

- I. Falsa. II. Correcta

#### Sección tercera. HIVE

1. ¿Dónde se almacenan los datos de las tablas internas de Hive?

- En cualquier directorio de HDFS especificado en la creación de la tabla.

- En el metastore.

- En un directorio de HDFS dedicado específicamente para este fin.

2. ¿Cuál es el componente de Hive que permite que los esquemas de bases de datos sean accesibles fuera de Hive?

- HCatalog

- Ninguna de las opciones

-HUE

-Thrift

2.- ¿Cuál de las siguientes afirmaciones es cierta sobre Hive?

I. Una consulta HiveQL genera procesos MapReduce.

II. Hive es extensible, por lo que permite utilizar plugins desarrollados por terceros.

I. Correcta. II. Falsa.

.

I. Falsa. II. Falsa.

.

I. Falsa. II. Correcta.

.

I. Correcta. II. Correcta.

Sección cuarta. PIG

1. Para dividir una relación existente en varias partes en función de algún criterio, ¿qué orden se utiliza?

- filter

- split.

- cut

2. ¿Cómo se llama el Shell de Pig que permite la ejecución de consultas interactivas?

- Bash. FALSO

- Grunt. VERDADERO

- Pig. FALSO

Sección quinta. SERDE, INYECCIÓN

1. ¿Cuál de las siguientes afirmaciones es cierta sobre Parquet?

- I. Es un formato binario de codificación de datos basado en columnas.
- II. Permite alcanzar tamaños de archivo más pequeños de Avro.

Para el fichero de datos que se muestra abajo, ¿qué devuelve el siguiente código Pig?

```
data = load 'empleados.txt' using
```

```
PigStorage(';') as
```

```
(nombre:chararray,  
titulos:bag{título:  
(nombretítulo:chararray, anyo:int)}},  
puestotrabajo:map[]);
```

```
nombreciudad = foreach data  
generate nombre,  
puestotrabajo#'oficina' as oficina;  
dump nombreciudad;
```

Fichero de datos:  
Miguel;{(licenciado,1999),  
(diplomado,1997)};  
[cargo#director,depto#compras,oficina#Madrid]

(Miguel,Madrid)

.

(Miguel,cargo#director,depto#compras,oficina#Madrid))

.

(Miguel,oficina#Madrid)

.

Ninguna de las restantes respuestas.

.

2.- ¿Qué componente de Pig chequea la sintaxis de los scripts?

Parser.

Ninguna de las restantes respuestas.

Compiler.

Optimizer.

¿Cuál de las siguientes afirmaciones es cierta?

- Un agente Flume solamente puede transportar información, pero no puede modificar la información que transporta.

Ninguna de las restantes respuestas.

Al realizar una exportación desde HDFS hasta una base de datos relacional con Sqoop, la tabla relacional destino debe existir antes de realizar la exportación.

Los canales en un agente Flume son componentes activos que sirven para decidir a qué sink envían la información.

.- ¿Cuál de las herramientas de inyección de datos vistas genera trabajos MapReduce?

I. Sqoop.

II. Flume.

I. Falsa. II. Correcta.

II. Correcta. II. Correcta.

I. Falsa. II. Falsa.

II. I. Correcta. II. Falsa.