

1. Cuál es la definición de ..RDDS...

- Resilient Distributed Datasets (RDDs)

- Resilient Distributed Databases (RDDs)
- Resilient Distributed Dataworks (RDDs)

2. Las funciones lambda de python...

- Son anónimas

- Son públicas.
- Ambas

3. Podemos crear un RDD a partir del sistema local de archivos, HDFS, Cassandra, HBase, Amazon S3 ?

- Verdadero

- Falso

4. La instrucción de Transformación map(func) ... realiza :

- Un nuevo RDD a partir de otro aplicando una transformación a cada elemento original

- Un nuevo RDD aplicando una transformación y eliminando el elemento original
- Un nuevo RDD aplicando una agrupación dinámica del elemento original

5. ¿Podemos cambiar el tipo de los elementos de los RDDs con un map?

- NO, son inmutables

- Si con map (flat)

6. ¿Cuál es el tamaño del rdd de salida en log.filter?

- Menor o igual que el original
- Mayor o igual que el original
- Siempre mayor

7. ¿Qué acción realiza count() en una instrucción de Spark ?

- Devuelve el número de elementos del RDD

- Devuelve el número de esquemas del RDD

8. ¿Qué acción realiza collect en una instrucción de Spark?

- Devuelve una lista con todos los elementos del RDD

- Devuelve una instancia de instalación con los elementos del RDD
- Devuelve los logs de instalación con los elementos del RDD

9. ¿Qué acción realiza joint en una instrucción de Spark?

- Realiza una operación unión de dos RDD

- Realiza una operación listada de dos RDD

10. cache() mantiene en memoria el RDD y usa MEMORY_ONLY

- FALSO

- VERDADERO

11. ¿Que son los Closures?

- Son las funciones que se ejecutan en las transformaciones y se pasan a cada nodo junto con las variables necesarias
- Son las transformaciones que se ejecutan en las funciones y se pasan a cada script junto con las variables
- Son instalaciones que se ejecutan en las funciones y se pasan a cada script junto con las variables

12. ¿Cómo hacemos si queremos contar el número de líneas corruptas de un fichero?

- Variables compartidas de tipo accumulator
- Variables compartidas de tipo terminator
- Variables tipo Select Count :

13. ¿Cómo hacemos si queremos compartir cierta información con todos los workers?

- No se puede. Los Workers son Inmutables

- Variables compartidas de tipo broadcast

14. Las Variables accumulators son de sólo Lectura...

- FALSO
- VERDADERO

15. La Consola de Spark está dividida en :

- jobs, stages, environment y executors

- jobs, tasks, environment y connectors

16. ¿Qué sucede cuando se hace un reduceByKey?

- Reducimos la clave.

- Agrupamos en un nodo los elementos con una misma clave

17. Qué es SPARK

- Apache Spark es un motor de código abierto desarrollado específicamente para el procesamiento y análisis de datos a gran escala

- Apache Spark es un motor licenciado por Amazon desarrollado específicamente para el procesamiento y análisis de datos a gran escala

18. Qué entendemos por un proceso Batch

- Es un proceso de Python y SQL

- Es un proceso independiente de código

- Es Procesado por lotes

19. Qué entendemos por un Driver en SPARK

- El "driver" es el proceso donde se ejecuta el método main del programa que queremos ejecutar
- El "driver" es el proceso donde se ejecuta el método start del programa que queremos ejecutar

20. Que es la Evaluación Perezosa

- Eso no existe en Spark
- Es un proceso de Spark no lo ejecutará hasta que no sea necesario en una acción.

21- ¿Qué es YARN?

Un lenguaje de programación

Un tipo de gestor de recursos

22- ¿Qué es un cluster?

Es una Agrupación de Ordenadores o bases de datos

Es una Agrupación de servicios

23- ¿Qué es la inmutabilidad?

Es una característica de Spark

Es una capacidad de Python

24- ¿Qué son las vistas en SQL?

Es una tabla virtual que se crea con un resultado específico

Es una tabla que se crea automáticamente en SQL

25- ¿Qué operaciones soportan los RDD?

Transformaciones y Acciones

Selección y Microservicios.