

Pregunta 1: La tecnología blockchain sólo sirve, esencialmente, para implementar criptomonedas.

- a) Verdadero
- b) **Falso**

La respuesta correcta es 'Falso'

Pregunta 2: ¿Qué podemos aseverar sobre Hadoop?

- a) Se desarrolló a partir de código liberado por Google
- b) El nombre viene de una marca de juguetes
- c) Su velocidad es inferior a la de Spark en la mayoría de las aplicaciones
- d) Sólo es aplicable para procesos en Batch
- e) Es una alternativa a MapReduce
- f) Está implementado usando varios lenguajes de programación

Las respuestas correctas son: Sólo es aplicable para procesos en Batch, Su velocidad es inferior a la de Spark en la mayoría de las aplicaciones

Pregunta 3: Si el formato en que están representados los datos de mi problema (almacenados en HDFS) no es en pares (clave, valor) el problema no puede ser resuelto con MapReduce

- a) Verdadero
- b) **Falso**

La respuesta correcta es 'Falso'

Pregunta 4: El principal concepto innovador de Spark es el de Resilient Distributed Dataset (RDD) que no es más que una implementación distribuida del concepto de dataframe de R, pero que soporta exactamente las mismas operaciones en formato distribuido.

- a) Verdadero
- b) **Falso**

La respuesta correcta es 'Falso'

Pregunta 5: Escribe el nombre de la unidad de información que es mil millones de veces mayor que 1 TB (una única palabra). Respuesta: Exabyte La respuesta correcta es: Zettabyte

Pregunta 6: El programador de aplicaciones MapReduce

- a) Tiene que encargarse de definir explícitamente el flujo de datos

- b) La principal herramienta para programar MapReduce son los PaaS
- c) **Debe concentrar la creatividad es en el diseño de las funciones Map y Reduce**
- d) Recibe el nombre de dummy

La respuesta correcta es: Debe concentrar la creatividad es en el diseño de las funciones Map y Reduce

Pregunta 7: Selecciona los componentes principales de Spark

- a) **Spark SQL**
- b) **Spark Streaming**
- c) Mahout
- d) **MLLib**
- e) HDFS
- f) Messos
- g) **GraphX**
- h) Hive
- i) **Spark Core**

Las respuestas correctas son: Spark SQL, Spark Streaming, MLLib, GraphX, Spark Core

Pregunta 8: MapReduce es un paradigma de programación caracterizado por:

- a) **Ser fácil de usar**
- b) **Robusto frente a fallos de los nodos del cluster**
- c) Sólo se puede usar con Java y Scala

Las respuestas correctas son: Ser fácil de usar, Robusto frente a fallos de los nodos del cluster

Pregunta 9: El enfoque tradicional para resolver problemas con cantidades ingentes de datos se basaba en el uso de recursos "High Performance Computing". ¿Cuál es la relación entre HPC y MapReduce?

- a) HPC siempre obtiene las soluciones más eficientes computacionalmente, pero el coste de desarrollo suele ser más elevado.
- b) MapReduce es una forma particular de HPC
- c) Por su facilidad, MapReduce es accesible para un número mayor de programadores
- d) Son esencialmente lo mismo

Las respuestas correctas son: HPC siempre obtiene las soluciones más eficientes computacionalmente, pero el coste de desarrollo suele ser más elevado., Por su facilidad, MapReduce es accesible para un número mayor de programadores

Pregunta 10: En su primera versión, Big Data se definía por tres Vs, correspondientes. a ... (escribe las tres palabras a que se refieren, separadas por espacios). Respuesta: volumen velocidad variedad
La respuesta correcta es: volumen velocidad variedad

Pregunta 11: Big Data se refiere al estudio y diseño de sistemas y métodos para computación de conjuntos de datos cuyas propiedades suponen un reto las capacidades de los ordenadores convencionales para analizar dichos datos y extraer valor no trivial de los mismos.

- a) Verdadero
- b) **Falso**

La respuesta correcta es 'Falso'

Pregunta 12: ¿Se puede implementar una aplicación blockchain sobre un servicio DaaS con un SGBD relacional?

- a) Verdadero
- b) **Falso**

La respuesta correcta es 'Falso'

Pregunta 13: Supongamos que disponemos de tantos recursos computacionales como se desee, entonces ¿se debe acelerar el cálculo de un proceso MapReduce usando el mayor número posible de procesos Map?

- a) Sí, si varios procesos MAP se ejecutan en el mismo nodo
- b) Sí, porque a mayor número de MAP, menor tamaño de partición a procesar y mayor rapidez
- c) No, porque la ganancia por mayor número de procesos no se compensa con el mayor número de comunicaciones
- d) **No siempre dado que podemos encontrarnos con particiones con un número de datos demasiado bajo (poco representativos)**

La respuesta correcta es: No siempre dado que podemos encontrarnos con particiones con un número de datos demasiado bajo (poco representativos)

Pregunta 14: ¿Cuál es el factor de aceleración (speedup) que, habitualmente, se obtiene en un algoritmo al codificarlo en Spark frente al mismo algoritmo en Hadoop? (valor numérico) Respuesta: 100 La respuesta correcta es: 100

Pregunta 15: Selecciona las herramientas que forman parte del ecosistema Apache Hadoop

- a) **Pig**
- b) **Hive**
- c) **Hbase**
- d) Frink
- e) **Zookeeper**
- f) MLlib
- g) Pork
- h) Hachis
- i) Spark
- j) Malamente

Las respuestas correctas son: Hbase, Hive, Pig, ZooKeeper

Pregunta 16: A partir de la versión 2.0 Hadoop introdujo Yarn (Data Operating System) como una forma de mejorar el uso y rendimiento de HDFS.

- a) Verdadero
- b) **Falso**

La respuesta correcta es 'Falso'

Pregunta 17: Selecciona los tipos de problema que no sean (fácilmente) resolubles con Hadoop

- a) **Procesos iterativos**
- b) **Procesos interactivos**
- c) Procesos en batch
- d) Ajedrez
- e) Algoritmos de Programación Dinámica basados en memorización
- f) **Procesamiento de grados**

Ha seleccionado correctamente 2. Las respuestas correctas son: Procesos iterativos, Procesos interactivos, Procesamiento de grados, Ajedrez

Pregunta 18: De entre las técnicas clásicas de diseño de algoritmos la más parecida en filosofía a MapReduce es la descomposición propia de la Programación Dinámica.

- a) Verdadero

b) **Falso**

La respuesta correcta es 'Falso'

Pregunta 19: Sobre Big Data podemos indicar:

- a) **Todas las demás respuestas son falsas**
- b) Ha supuesto, fundamentalmente, una revolución en el contexto de los negocios, pero no de la tecnología
- c) Sólo contempla el procesamiento de datos con un volumen masivamente grande
- d) Sólo se aplica a métodos de analítica de datos y extracción de conocimiento

La respuesta correcta es: Todas las demás respuestas son falsas


Pregunta 20: Hadoop File System (HDFS):

- a) Sólo es aplicable para nodos que comparten sus discos mediante NFS
- b) Sólo se puede utilizar para aplicaciones hechas en Hadoop
- c) Es más eficaz si los datos sólo se almacenan en memoria principal
- d) **Es robusto y puede usarse con múltiples marcos de trabajo MapReduce**

Pregunta 21: ¿Se puede desplegar una infraestructura para big data usando contenedores?

- a) Sí, pero sólo si HDFS también se despliega con contenedores
- b) No porque los contenedores son demasiado ligeros para permitir cálculos complejos
- c) Sólo si los cálculos no requieren el uso de GPUs
- d) **Sí**

La respuesta correcta es: Sí

Pregunta 22: Indica el nombre del framework de Big Data del que es propio el esquema de operaciones que se refleja en la figura: 

La respuesta correcta es: Hadoop

Pregunta 23: En relación con MapReduce:

- a) La operación Map, viene de la similitud en la organización de datos que necesitan con el tipo de dato de C++ y Java para representar diccionarios: map.

- b) Es aplicable a todos los conjuntos de datos que puedan representarse como pares {clave, valor}.
- c) Sus etapas principales son Map, Reduce and Restoring
- d) **No es aplicable a todo tipo de problemas**

La respuesta correcta es: No es aplicable a todo tipo de problemas, Es aplicable a todos los conjuntos de datos que puedan representarse como pares {clave, valor}.

Pregunta 24: La principal ventaja de Spark sobre Hadoop fue el cambio del lenguaje de programación elegido para su implementación.

- a) Verdadero
- b) **Falso**

La respuesta correcta es 'Falso'