

OpenCourseWare

Sistemas Paralelos y Distribuidos

Félix García Carballeira

Alejandro Calderón Mateos

Ejercicios de autoevaluación

Tema 9



1. Según el material, ¿cuál era una de las ventajas principales de los sistemas centralizados iniciales (alrededor de 1960)?
 - a. Alta tolerancia a fallos de red.
 - b. Acceso ubicuo a los datos desde cualquier dispositivo.
 - c. Facilidad para el mantenimiento y la compartición de recursos.
 - d. Escalabilidad ilimitada para soportar millones de usuarios.
2. ¿Cuál de las siguientes afirmaciones captura la esencia de la famosa cita de Leslie Lamport sobre la realidad de los sistemas distribuidos?
 - a. Un sistema distribuido es una colección de ordenadores que aparece como un único sistema coherente.
 - b. El fallo de un ordenador del que no tenías conocimiento puede inutilizar tu propio ordenador.
 - c. En un sistema distribuido, la latencia es cero y el ancho de banda es infinito.
 - d. Los componentes de un sistema distribuido se coordinan únicamente mediante el paso de mensajes.
3. En el contexto de Big Data, ¿cuál de las siguientes NO es una de las "Cuatro V" mencionadas en la presentación?
 - a. Virtualización.
 - b. Veracidad.
 - c. Volumen.
 - d. Velocidad.
4. ¿Cuál fue el principio fundamental que inspiró el diseño de Hadoop, basado en las publicaciones de Google sobre GFS y MapReduce?
 - a. Procesar los datos directamente en el lugar donde se almacenan para minimizar el movimiento de datos por la red.
 - b. Utilizar hardware especializado y redes de muy alta velocidad para acelerar la transferencia de datos.
 - c. Almacenar todos los datos en bases de datos relacionales para facilitar las consultas.
 - d. Mover los datos a un clúster de cómputo centralizado para su procesamiento.
5. Dentro de la arquitectura básica de Hadoop, ¿cuál es la función del componente MapReduce?
 - a. Permitir la ejecución de consultas SQL sobre los datos almacenados.
 - b. Almacenar los datos en bloques replicados a través de múltiples máquinas.
 - c. Transferir datos entre Hadoop y bases de datos relacionales tradicionales.
 - d. Proporcionar un framework para escribir aplicaciones que procesan grandes conjuntos de datos en paralelo.
6. Si un analista de datos necesita ejecutar consultas con una sintaxis similar a SQL sobre terabytes de datos almacenados en HDFS, ¿qué proyecto del ecosistema Hadoop sería el más adecuado?
 - a. HBase.
 - b. Pig.
 - c. Hive.
 - d. Flume.

7. ¿Cuál es la diferencia fundamental en el manejo de los datos intermedios entre Apache Spark y el MapReduce de Hadoop?
 - a. Spark no genera datos intermedios, ya que procesa todo en un único paso.
 - b. Hadoop MapReduce escribe los resultados intermedios de cada etapa en el disco (HDFS), mientras que Spark intenta mantenerlos en memoria RAM.
 - c. Spark siempre escribe los resultados intermedios en una base de datos relacional, mientras que Hadoop los escribe en HDFS.
 - d. Hadoop comprime los datos intermedios, mientras que Spark los procesa sin comprimir para mayor velocidad.
8. ¿Qué abstracción de datos es fundamental en Apache Spark, representando una colección de elementos distribuida, inmutable y que puede recuperarse de fallos?
 - a. DAG (Directed Acyclic Graph).
 - b. DataFrame.
 - c. HDFS Block.
 - d. RDD (Resilient Distributed Dataset).
9. Al comparar los modelos de procesamiento, ¿qué ventaja ofrece el DAG (Grafo Acíclico Dirigido) de Spark sobre el modelo MapReduce de Hadoop?
 - a. Permite crear flujos de trabajo con múltiples etapas de forma flexible y optimizada, evitando pasos intermedios fijos.
 - b. El DAG garantiza que todos los datos se procesen en un único nodo para evitar la comunicación de red.
 - c. El modelo MapReduce es más adecuado para el procesamiento iterativo y en tiempo real.
 - d. El DAG es más simple, consistiendo siempre en solo dos etapas: Map y Reduce.
10. En el contexto de la evolución hacia sistemas de gran escala, ¿qué es el "Internet of Things" (IoT)?
 - a. Un sistema para que todas las personas estén conectadas a través de teléfonos móviles.
 - b. La interconexión de objetos y dispositivos cotidianos a Internet, permitiéndoles enviar y recibir datos.
 - c. Una red de alta velocidad que conecta exclusivamente los grandes centros de datos.
 - d. Un nuevo protocolo de red que reemplaza a TCP/IP para mejorar la velocidad.