

Question 1

Complete

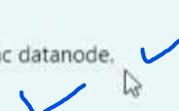
Marked out of
1.00

Flag question

Cơ chế chịu lỗi của datanode trong HDFS?

Select one or more:

- a. Sử dụng cơ chế heartbeat, Namenode định kỳ hỏi các datanode về trạng thái tồn tại của các datanode.
- b. Sử dụng cơ chế heartbeat, định kỳ các datanode thông báo về trạng thái cho Namenode.
- c. Sử dụng Zookeeper để quản lý các thành viên datanode trong cụm.

**Question 2**

Complete

Marked out of
1.00

Flag question

Các mục tiêu chính của Apache Hadoop?

Select one or more:

- a. Lưu trữ dữ liệu khả mở, xử lý dữ liệu lớn mạnh mẽ và trực quan hóa dữ liệu hiệu quả
- b. Lưu trữ dữ liệu khả mở và Xử lý dữ liệu lớn mạnh mẽ
- c. Trực quan hóa dữ liệu hiệu quả
- d. Lưu trữ dữ liệu khả mở
- e. Xử lý dữ liệu lớn mạnh mẽ

**Question 3**

Complete

Chọn phát biểu sai

Question 3

Complete

Marked out of

1.00

Flag question

Chọn phát biểu sai

Select one or more:

- a. NoSQL yêu cầu lược đồ CSDL phải được định nghĩa trước khi thêm dữ liệu
- b. NoSQL cho phép thêm vào dữ liệu mà không cần định nghĩa trước lược đồ dữ liệu
- c. NoSQL được đưa ra nhằm bổ sung các giải pháp mà CSDL truyền thống không đáp ứng tốt

Question 4

Complete

Marked out of

1.00

Flag question

Phát biểu nào sau sai về Kafka?

Select one or more:

- a. Tiến trình quảng bá message lên cụm Kafka gọi là publishers
- b. Kafka quản lý các luồng thông điệp (messages) thành các nhóm gọi là các Topics.
- c. Tiến trình đăng ký theo dõi các topics gọi là consumers.
- d. Các máy chủ chạy Kafka gọi là các brokers.

Question 5

Complete

Marked out of

1.00

Flag question

Đâu là đặc điểm của RDD (Resilient distributed dataset) của Spark?

Select one or more:

- a. Người sử dụng không thể quyết định số các phân mảnh của mỗi RDD
- b. Có khả năng chịu lỗi
- c. Được chia thành các phân mảnh (partition)
- d. Người lập trình có thể quyết định số các phân mảnh của mỗi RDD

Question 6

Complete

Marked out of
1.00

Flag question

Đâu là đặc điểm của Spark streaming?

Select one or more:

- a. Có thể nhận đầu vào là các tệp tin trên HDFS
- b. Không thể thực hiện các truy vấn SQL
- c. Có thể nhận đầu vào là các luồng dữ liệu từ Kafka

**Question 7**

Complete

Marked out of
1.00

Flag question

Kiến trúc xử lý dữ liệu Lambda có đặc điểm gì?

Select one or more:

- a. Giúp giải quyết vấn đề độ trễ từ khi dữ liệu được thập tới kết quả phân tích của mô hình xử lý theo lô
- b. Kết hợp xử lý dữ liệu theo lô và theo luồng
- c. Giúp giải quyết vấn đề nhược điểm của xử lý theo luồng là kết quả phân tích không khai thác được toàn bộ dữ liệu trong lịch sử.
- d. Có kiến trúc gồm 2 tầng: tầng xử lý theo lô và tầng xử lý theo luồng
- e. Bao gồm các tiến trình ETL (extract, transform, load) đưa dữ liệu vào hồ dữ liệu (data lake)

**Question 8**

Complete

Marked out of
1.00

Flag question

Cơ chế tổ chức dữ liệu của Datanode trong HDFS?

Select one or more:

- a. Các chunk là các tệp tin trong hệ thống tệp tin cục bộ của máy chủ datanode.
- b. Các chunk được lưu trữ tin cậy trên datanode theo cơ chế RAID.
- c. Các chunk là các vùng dữ liệu liên tục trên ổ cứng của máy chủ datanode.



Question 9

Complete

Marked out of
1.00

Flag question

Phát biểu nào sai về Presto?

Select one or more:

- a. Presto là một engine truy vấn SQL hiệu năng cao, phân tán cho dữ liệu lớn
- b. Presto được quản lý bởi Apache Software foundation
- c. Presto cho phép tích hợp với các công cụ Business Intelligence
- d. Presto được quản lý bởi Presto Software foundation

Question 10

Complete

Marked out of
1.00

Flag question

Phát biểu nào sai về Hfile trong Hbase?



Select one or more:

- a. Nhiều Hfile có thể được gộp lại thành 1 Hfile lớn theo những khoảng thời gian nhất định
- b. Hfile chứa một tập hợp các dòng bản ghi trong Hbase table
- c. Một version của 1 dòng hay 1 bản ghi trong Hbase table có thể được phân rã trên nhiều Hfile khác nhau
- d. Nhiều Hfile có thể được gộp lại thành 1 Hfile lớn khi cần thiết

Question 11

Complete

Marked out of
1.00

Flag question

Công cụ nào có thể sử dụng để hỗ trợ import, export dữ liệu vào ra hệ sinh thái Hadoop?

Select one or more:

- a. Flume
- b. Sqoop
- c. Hive
- d. Oozie

Question 12

Complete

Marked out of
1.00

Flag question

Thế nào là UNIX semantic?

Select one or more:

- a. Tệp tin là chỉ đọc, không cho phép cập nhật và ghi đè. Mọi tiến trình đều có thể đọc tệp tin đồng thời.
- b. Cập nhật tới tệp tin có thể được nhìn thấy ngay lập tức bởi các tiến trình khác mà mở tệp tin đó cùng thời điểm với tiến trình ghi.
- c. Cập nhật tới tệp tin chỉ có thể thấy được bởi các tiến trình khác sau khi tiến trình ghi thực hiện thao tác đóng tệp.

Question 13

Complete

Marked out of
1.00

Flag question

Cơ chế nhận bản dữ liệu trong HDFS?



Select one or more:

- a. Datanode là primary quyết định vị trí các nhân bản của các chunk tại các secondary datanode.
- b. Namenode quyết định vị trí các nhân bản của các chunk trên datanode.
- c. Client quyết định vị trí lưu trữ các nhân bản với từng chunk.

Question 14

Complete

Marked out of
1.00

Flag question

Đâu là vấn đề khi xử lý dữ liệu lớn với MapReduce?

Select one or more:

- a. Xử lý dữ liệu lớn theo lô (Bulk processing)
- b. Xử lý dữ liệu lớn trong thời gian tương tác
- c. Xử lý chuỗi các công việc
- d. Xử lý luồng dữ liệu lớn

Question 15

Complete

Marked out of
1.00

Flag question

Đâu là một dạng của NoSQL

Select one or more:

- a. OLAP
- b. JSON
- c. Key-value store
- d. MySQL

Question 16

Complete

Marked out of
1.00

Flag question

Phát biểu nào sai về cơ chế scheduling của Presto?



Select one or more:

- a. Stage có thể được lập lịch theo giai đoạn
- b. Một task có thể được lập lịch chạy trên bất kỳ worker nào
- c. Split được gán cho task theo cơ chế lazy
- d. Stage có thể được lập lịch all-at-once

Question 17

Complete

Marked out of
1.00

Flag question

Thao tác nào không được hỗ trợ bởi Hbase

Select one or more:

- a. Put
- b. Multiput
- c. Scan
- d. Join

e. Get

Question 18

Complete

Marked out of
1.00

Flag question

Hbase có thể được sử dụng cho kiểu dữ liệu nào

Select one or more:

- a. Dữ liệu có cấu trúc
- b. Dữ liệu phi cấu trúc
- c. Dữ liệu bán cấu trúc
- d. Tất cả các phương án được đưa ra.

Question 19

Complete

Marked out of
1.00

Flag question

HDFS giải quyết bài toán một điểm hỏng hóc duy nhất (single-point-of-failure) cho Namenode bằng cách nào?

Select one or more:

- a. Sử dụng Secondary namenode theo cơ chế active-passive. Secondary namenode chỉ hoạt động khi có vấn đề với Namenode.
- b. Sử dụng thêm secondary namenode theo cơ chế active-active. Cả Namenode và Secondary namenode cùng online trong hệ thống

Question 20

Complete

Marked out of
1.00

Flag question

Chọn phát biểu đúng về NoSQL

Select one or more:

- a. Không hỗ trợ các truy vấn SQL
- b. Không thể được sử dụng kết hợp với các CSDL quan hệ
- c. Đáp ứng khả năng xử lý giao dịch với tính nhất quán chặt
- d. Rất phù hợp cho các tập dữ liệu phân tán quy mô lớn

Question 21

Complete

Marked out of
1.00

Flag question

Phát biểu nào sau đây không đúng về Apache Hadoop?

Select one or more:

- a. Hadoop thiết kế để mở rộng thông qua kỹ thuật scale-out, tăng số lượng máy chủ
- b. Xử lý dữ liệu phân tán với mô hình lập trình đơn giản, thân thiện hơn như MapReduce.
- c. Thiết kế để vận hành trên siêu máy tính, cấu hình mạnh, độ tin cậy cao
- d. Thiết kế để vận hành trên phần cứng phổ thông, có khả năng chống chịu lỗi phần cứng

**Question 22**

Complete

Marked out of
1.00

Flag question

Ưu điểm của hệ thống tệp tin phân tán là gì?



Select one or more:

- a. Tập trung hóa việc quản trị.
- b. Đơn giản hóa việc chia sẻ dữ liệu.
- c. Cho phép người dùng có cái nhìn hợp nhất (như nhau) về toàn bộ dữ liệu trong hệ thống.

**Question 23**

Complete

Marked out of
1.00

Flag question

Phát biểu nào đúng về Amazon DynamoDB

Select one or more:

- a. DynamoDB là multi-hop DHT
- b. DynamoDB là one-hop DHT
- c. DynamoDB là zero-hop DHT



Question 24

Complete

Marked out of
1.00[Flag question](#)

Đâu là các thao tác có thể thực hiện trên RDD (Resilient distributed dataset) của Spark?

Select one or more:

- a. Yêu cầu Spark lưu RDD ở bộ nhớ đệm ✓
- b. Thực hiện các hành động (action)
- c. Thực hiện các biến đổi mà cập nhật các bản ghi trong RDD
- d. Thực hiện các biến đổi (transformation) ✓
- e. Thực hiện các biến đổi mà xoá các bản ghi trong RDD

Question 25

Complete

Marked out of
1.00[Flag question](#)

Phát biểu nào sai về Kafka?



Select one or more:

- a. Thứ tự của message trong mỗi partition do key của message quyết định.
- b. Kafka producer quyết định message sẽ được gửi đến partition nào trong topic. ✓
- c. Kafka producer có thể gửi message đến nhiều broker khác nhau.

Question 26

Complete

Marked out of
1.00[Flag question](#)

CSDL nào dưới đây không phải là NoSQL

Select one or more:

- a. MongoDB
- b. Không phải các đáp án đã đưa ra
- c. Cassandra
- d. Microsoft SQL server ✓

- d. Microsoft SQL server

Question 27

Complete

Marked out of
1.00

Flag question

Cơ chế mà NoSQL sử dụng để tăng khả năng chịu lỗi

Select one or more:

- a. Phân mảnh và phân tán dữ liệu ra nhiều máy chủ
- b. Nhân bản (Replication) 
- c. Giao diện truy vấn đơn giản hơn so với CSQL quan hệ truyền thống

Question 28

Complete

Marked out of
1.00

Flag question

Phát biểu nào sai về Kafka?



Select one or more:

- a. Các message trên Kafka được lưu lại theo thời gian (time-based)
- b. Các message trên Kafka được lưu lại theo kích thước partition (size-based)
- c. Các message trên Kafka được lưu lại trước khi thực hiện compaction
- d. Message sau khi được tiêu thụ bởi tất cả các consumer thì bị xoá. 

Question 29

Complete

Marked out of
1.00

Flag question

Các biến đổi (transformation) trên Spark có đặc điểm gì?

Select one or more:

- a. Mỗi phép biến đổi trên RDD được thực thi bởi một hay nhiều Spark worker 
- b. Thực hiện theo cơ chế lười biếng, khi nào một hành động (action) cần tới phép biến đổi trước đó phải thực hiện thì mới phải thực hiện 
- c. Các biến đổi (transformation) luôn tạo ra RDD mới có cùng số partition với RDD đầu vào

Question 30

Complete

Marked out of
1.00

Flag question

Các đặc điểm của virtual node trên AmazonDB. Chọn phương án sai

Select one or more:

- a. Số lượng các node ảo đối với mỗi node vật lý là khác nhau tùy vào từng node vật lý.
- b. Số lượng các node ảo bắt buộc cần phải căn cứ vào khả năng lưu trữ của node vật lý.
- c. Node ảo đóng vai trò quan trọng trong bài toán cân bằng tải và hiệu năng khi một node vật lý ra hoặc kết nối vào cụm.
- d. Mỗi node vật lý có thể được ánh xạ thành nhiều node ảo, nằm liên tiếp nhau trong vòng tròn không gian khoá.

Question 31

Complete

Marked out of
1.00

Flag question

Thể nào là Session semantic?



Select one or more:

- a. Tệp tin là chỉ đọc, không cho phép cập nhật và ghi đè. Mọi tiến trình đều có thể đọc tệp tin đồng thời.
- b. Cập nhật tới tệp tin có thể được nhìn thấy ngay lập tức bởi các tiến trình khác mà mở tệp tin đó cùng thời điểm với tiến trình ghi.
- c. Cập nhật tới tệp tin chỉ có thể thấy được bởi các tiến trình khác sau khi tiến trình ghi thực hiện thao tác đồng tệp.

Question 32

Complete

Marked out of
1.00

Flag question

Phát biểu nào sai về Hbase

Select one or more:

- a. Hbase có lề thuộc vào các dịch vụ cung cấp bởi Zookeeper
- b. Hbase hỗ trợ truy vấn dạng SQL
- c. Hbase không hỗ trợ versioning
- d. Hbase có lề thuộc vào các dịch vụ cung cấp bởi HDFS

Question 33

Complete

Marked out of
1.00

Flag question

Giữa Pig và Hive, công cụ nào có giao diện truy vấn gần với ANSI SQL hơn?

Select one or more:

- a. Pig
- b. Hive ✓
- c. Pig và Hive đều không có giao diện truy vấn gần với SQL.

Question 34

Complete

Marked out of
1.00

Flag question

Ưu điểm của kiến trúc SAN (Storage area network)?

Select one or more:

- a. Quản trị dễ dàng hơn so với NAS.
- b. Hiệu năng, băng thông tốt hơn với NAS. ✓
- c. Máy khách có thể kết nối tới SAN bằng đường truyền Ethernet thông thường (Chuẩn kết nối TCP/IP).

Question 35

Complete

Marked out of
1.00

Flag question

Hadoop giải quyết bài toán khai mỏ bằng cách nào? Chọn đáp án sai.

Select one or more:

- a. Các node thêm vào cụm cần có cấu hình, độ tin cậy cao
- b. Các node tham gia vào cụm đóng cả 2 vai trò tính toán và lưu trữ
- c. Thiết kế hướng phân tán ngay từ đầu, mặc định triển khai trên cụm máy chủ
- d. Các node tham gia vào cụm Hadoop được gán vai trò hoặc là node tính toán hoặc là node lưu trữ dữ liệu ✓

Question 36

Complete

Marked out of
1.00

Flag question

Đâu là ưu điểm của Spark so với MapReduce?

Select one or more:

- a. Hỗ trợ tốt cho xử lý chuỗi các biến đổi ✓
- b. Khai thác bộ nhớ trong thay vì sử dụng hệ thống lưu trữ ngoài như HDFS ✓
- c. Có khả năng chịu lỗi ✓
- d. Có thể khai phá dữ liệu trong thời gian tương tác ✓

Question 37

Complete

Marked out of
1.00

Flag question

Thành phần nào không thuộc thành phần lõi của Hadoop?

Select one or more:

- a. Apache Zookeeper ✓
- b. Hệ thống tệp tin phân tán HDFS
- c. Apache Hbase ✓
- d. Mapreduce framework
- e. YARN: yet another resource negotiator

Question 38

Complete

Marked out of
1.00

Flag question

Hadoop giải quyết bài toán chịu lỗi thông qua kỹ thuật gì. Chọn đáp án sai.

Select one or more:

- a. Hadoop chịu lỗi thông qua kỹ thuật dư thừa
- b. Các công việc cần tính toán được phân mảnh thành các tác vụ độc lập.
- c. Các tệp tin được phân mảnh, các mảnh được lưu trữ tin cây trên ổ cứng theo cơ chế RAID ✓
- d. Các tệp tin được phân mảnh, các mảnh được phân tán ra các node khác trên cụm

- e. YARN: yet another resource negotiator

Question 38

Complete

Marked out of
1.00

Flag question

Hadoop giải quyết bài toán chịu lỗi thông qua kỹ thuật gì. Chọn đáp án sai.

Select one or more:

- a. Hadoop chịu lỗi thông qua kỹ thuật dư thừa
- b. Các công việc cần tính toán được phân mảnh thành các tác vụ độc lập.
- c. Các tệp tin được phân mảnh, các mảnh được lưu trữ tin cậy trên ổ cứng theo cơ chế RAID
- d. Các tệp tin được phân mảnh, các mảnh được nhân bản ra các node khác trên cụm



Question 39

Complete

Marked out of
1.00

Flag question

Phát biểu sau đây đúng hay sai: Trong cụm Kafka, 1 server đóng vai trò leader, các server còn lại đóng vai trò follower.

Select one or more:

- a. Sai
- b. Đúng



Question 40

Complete

Marked out of
1.00

Flag question

Phát biểu nào sai về Presto

Select one or more:

- a. Presto thường nhanh hơn Hive hay Pig
- b. Presto không truy vấn được dữ liệu trong MySQL, MS SQL và các CSDL quan hệ truyền thống
- c. Presto có thể truy vấn nhiều data storages khác nhau như HDFS, Cassandra

