



Big Data

Question 1

CompleteMarked out of 1.00Flag question

Question text

NoSQL có đặc điểm nào dưới đây?Select one or more:

- ☐ a.Mở rộng theo chiều dọc, thiết kế phức tạp, tinh chỉnh được tính sẵn sàng của hệ thống
- ☐ b.Không thể sử dụng SQL để truy vấn dữ liệu NoSQL
- ☒ c.Mở rộng theo chiều ngang, tinh chỉnh được tính sẵn sàng của hệ thống
- ☐ d.Mở rộng theo chiều dọc, thiết kế đơn giản, khó tinh chỉnh tính sẵn sàng của hệ thống

Question 2

CompleteMarked out of 1.00Flag question

Question text

Cơ chế nhân bản dữ liệu trong HDFS?Select one or more:

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

Question 3CompleteMarked out of 1.00Flag question

Question text

Các mục tiêu chính của Apache Hadoop?Select one or more:

- ☐ a.Xử lý dữ liệu lớn mạnh mẽ
- ☐ b.Trực quan hoá dữ liệu hiệu quả
- ☐ c.Lưu trữ dữ liệu khả mở

☒ d.Lưu trữ dữ liệu khả mở và Xử lý dữ liệu lớn mạnh mẽ

☐ e.Lưu trữ dữ liệu khả mở, xử lý dữ liệu lớn mạnh mẽ và trực quan hoá dữ liệu hiệu quả

Question 4 CompleteMarked out of 1.00Flag question

Question text

Thành phần nào không thuộc thành phần lõi của Hadoop?Select one or more:

☐ a.Mapreduce framework

☒ b.Apache Zookeeper

☐ c.Hệ thống tệp tin phân tán HDFS

☐ d.YARN: yet another resource negotiator

☒ e.Apache Hbase

Question 5 CompleteMarked out of 1.00Flag question

Question text

Các đặc điểm của virtual node trên AmazonDB. Chọn phương án saiSelect one or more:

☐ a.Mỗi node vật lý có thể được ánh xạ thành nhiều node ảo, nằm liên tiếp nhau trong vòng tròn không gian khoá.

☐ b.Số lượng các node ảo đối với mỗi node vật lý là khác nhau tùy vào từng node vật lý.

☒ c.Số lượng các node ảo bắt buộc cần phải căn cứ vào khả năng lưu trữ của node vật lý.

☐ d.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.

Question 6 CompleteMarked out of 1.00Flag question

Question text

Phát biểu nào đúng về Amazon DynamoDBSelect one or more:

☒ a.DynamoDB là zero-hop DHT

☐ b.DynamoDB là one-hop DHT

☐ c.DynamoDB là multiple-hop DHT

Question 7 CompleteMarked out of 1.00Flag question

Question text

Phát biểu nào sai về PrestoSelect one or more:

☐ a.Presto có thể truy vấn nhiều data storages khác nhau như HDFS, Cassandra

☐ b.Presto thường nhanh hơn Hive hay Pig

[x] c.Presto không truy vấn được dữ liệu trong MySQL, MS SQL và các CSDL quan hệ truyền thống

Question 8CompleteMarked out of 1.00Flag question

Question text

Thao tác nào không được hỗ trợ bởi HbaseSelect one or more:

- ☐ a.Scan
- ☐ b.Multiput
- ☐ c.Put
- ☒ d.Join
- ☐ e.Get

Question 9CompleteMarked out of 1.00Flag question

Question text

Ưu điểm của kiến trúc NAS (Network attached Storage)?Select one or more:

- ☐ a.Máy khách có thể kết nối tới NAS bằng đường truyền Ethernet thông thường (Chuẩn kết nối TCP/IP).
- ☒ b.Đơn giản hoá việc chia sẻ dữ liệu.
- ☒ c.Tính khả mở cao.

Question 10CompleteMarked out of 1.00Flag question

Question text

Đây là một dạng của NoSQLSelect one or more:

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

Question 11CompleteMarked out of 1.00Flag question

Question text

Đây là vấn đề khi xử lý dữ liệu lớn với MapReduce?Select one or more:

- ☒ a.Xử lý dữ liệu lớn trong thời gian tương tác
- ☐ b.Xử lý luồng dữ liệu lớn
- ☐ c.Xử lý chuỗi các công việc

☐ d.Xử lý dữ liệu lớn theo lô (Bulk processing)

Question 12 CompleteMarked out of 1.00Flag question

Question text

Phát biểu nào sai về Kafka?Select one or more:

☐ a.Kafka producer có thể gửi message đến nhiều broker khác nhau.

☐ b.Thứ tự của message trong mỗi partition do key của message quyết định.

☒ c.Kafka producer quyết định message sẽ được gửi đến partition nào trong topic.

Question 13 CompleteMarked out of 1.00Flag question

Question text

Kiến trúc xử lý dữ liệu Lambda có đặc điểm gì?Select one or more:

☐ a.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.

☐ b.Bao gồm các tiến trình ETL (extract, transform, load) đưa dữ liệu vào hồ dữ liệu (data lake)

☐ c.Có kiến trúc gồm 2 tầng: tầng xử lý theo lô và tầng xử lý theo luồng

☒ d.Kết hợp xử lý dữ liệu theo lô và theo luồng

☒ e.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

Question 14 CompleteMarked out of 1.00Flag question

Question text

Phát biểu nào sau sai về Kafka?Select one or more:

☒ a.Tiến trình quảng bá message lên cụm Kafka gọi là publishers.

☐ b.Các máy chủ chạy Kafka gọi là các brokers.

☐ c.Tiến trình đăng ký theo dõi các topics gọi là consumers.

☐ d.Kafka quản lý các luồng thông điệp (messages) thành các nhóm gọi là các Topics.

Question 15 CompleteMarked out of 1.00Flag question

Question text

Phát biểu nào đúng về Presto?Select one or more:

☐ a.Presto có cơ chế chịu lỗi khi thực thi truy vấn

☒ b.Các stage được thực thi theo cơ chế pipeline, không có thời gian chờ giữa các stage như Map Reduce

☐ c.Presto cho phép xử lý kết tập dữ liệu mà kích thước lớn hơn kích thước bộ nhớ trong

Question 16 CompleteMarked out of 1.00Flag question

Question text

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.Thiết kế để vận hành trên siêu máy tính, cấu hình mạnh, độ tin cậy cao

☐ c.Xử lý dữ liệu phân tán với mô hình lập trình đơn giản, thân thiện hơn như MapReduce.

☐ d.Thiết kế để vận hành trên phần cứng phổ thông, có khả năng chống chịu lỗi phần cứng

Question 17 CompleteMarked out of 1.00Flag question

Question text

Hadoop giải quyết bài toán chịu lỗi thông qua kỹ thuật gì. Chọn đáp án sai.Select one or more:

☐ a.Các công việc cần tính toán được phân mảnh thành các tác vụ độc lập.

☐ b.Hadoop chịu lỗi thông qua kỹ thuật dư thừa

☒ 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 18 CompleteMarked out of 1.00Flag question

Question text

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.Một version của 1 dòng hay 1 bản ghi trong Hbase table có thể được phân rã trên nhiều Hfile khác nhau

☐ c.Hfile chứa một tập hợp các dòng bản ghi trong Hbase table

☐ d.Nhiều Hfile có thể được gộp lại thành 1 Hfile lớn khi cần thiết

Question 19 CompleteMarked out of 1.00Flag question

Question text

Các biến đổi (transformation) trên Spark có đặc điểm gì?Select one or more:

- ☐ a.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
- ☐ b.Mỗi phép biến đổi trên RDD được thực thi bởi một hay nhiều Spark worker
- ☒ c.Các biến đổi (transformation) luôn tạo ra RDD mới có cùng số partition với RDD đầu vào

Question 20CompleteMarked out of 1.00Flag question

Question text

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.Có khả năng chịu lỗi
- ☒ c.Có thể khai phá dữ liệu trong thời gian tương tác
- ☒ d.Khai thác bộ nhớ trong thay vì sử dụng hệ thống lưu trữ ngoài như HDFS

Question 21CompleteMarked out of 1.00Flag question

Question text

Đâu không phải là tính năng mà NoSQL nào cũng đáp ứngSelect one or more:

- ☐ a.Khả năng mở rộng linh hoạt
- ☐ b.Phù hợp với dữ liệu lớn
- ☒ c.Tính sẵn sàng cao

Question 22CompleteMarked out of 1.00Flag question

Question text

Các đặc trưng của HDFS. Chọn đáp án sai.Select one or more:

- ☐ a.Hỗ trợ cơ chế phân quyền và kiểm soát người dùng của UNIX
- ☒ b.Hỗ trợ thao tác đọc ghi tương tranh tại chunk (phân mảnh) trên tệp tin
- ☐ c.Hỗ trợ nén dữ liệu để tiết kiệm chi phí
- ☐ d.Tối ưu cho các tệp tin có kích thước lớn

Question 23CompleteMarked out of 1.00Flag question

Question text

Giữa Pig và Hive, công cụ nào có giao diện truy vấn gần với ANSI SQL hơn?Select one or more:

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

☒ c.Hive

Question 24 CompleteMarked out of 1.00Flag question

Question text

Phát biểu nào sai về cơ chế scheduling của Presto?Select one or more:

☒ a.Một task có thể được lập lịch chạy trên bất kỳ worker nào

☐ b.Stage có thể được lập lịch all-at-once

☐ c.Split được gán cho task theo cơ chế lazy

☐ d.Stage có thể được lập lịch theo giai đoạn

Question 25 CompleteMarked out of 1.00Flag question

Question text

Ưu điểm của hệ thống tệp tin phân tán là gì?Select one or more:

☒ a.Cho phép người dùng có cái nhìn hợp nhất (như nhau) về toàn bộ dữ liệu trong hệ thống.

☐ b.Tập trung hoá việc quản trị.

☒ c.Đơn giản hoá việc chia sẻ dữ liệu.

Question 26 CompleteMarked out of 1.00Flag question

Question text

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 27 CompleteMarked out of 1.00Flag question

Question text

Chọn phát biểu đúng khi nói về MongoDBSelect one or more:

☐ a.Các văn bản có thể chứa nhiều cặp key-value hoặc key-array, hoặc các văn bản lồng (nested documents).

☐ b.MongoDB hay các NoSQL có khả năng mở tốt hơn các CSDL quan hệ truyền thống.

☐ c.MongoDB có các trình điều khiển driver cho nhiều ngôn ngữ lập trình khác nhau.

☒ d. Tất cả các phương án đã đưa ra.

Question 28 CompleteMarked out of 1.00Flag question

Question text

Cơ chế chịu lỗi của datanode trong HDFS?Select one or more:

☒ a. Sử dụng cơ chế heartbeat, định kỳ các datanode thông báo về trạng thái cho Namenode.

☐ b. Sử dụng Zookeeper để quản lý các thành viên datanode trong cụm.

☐ c. Sử dụng cơ chế heartbeat, Namenode định kỳ hỏi các datanode về trạng thái tồn tại của các datanode.

Question 29 CompleteMarked out of 1.00Flag question

Question text

Phát biểu nào sai về HbaseSelect one or more:

☐ a. Hbase có hệ thuộc vào các dịch vụ cung cấp bởi HDFS

☐ b. Hbase có lệ thuộc vào các dịch vụ cung cấp bởi Zookeeper

☒ c. Hbase không hỗ trợ versioning

☐ d. Hbase hỗ trợ truy vấn dạng SQL

Question 30 CompleteMarked out of 1.00Flag question

Question text

Hệ thống nào cho phép đọc ghi dữ liệu tại vị trí ngẫu nhiên, thời gian thực tới hàng terabyte dữ liệuSelect one or more:

☐ a. Flume

☐ b. Pig

☐ c. HDFS

☒ d. Hbase

Question 31 CompleteMarked out of 1.00Flag question

Question text

Đâu là đặc điểm của RDD (Resilient distributed dataset) của Spark?Select one or more:

☒ a. Người lập trình có thể quyết định số các phân mảnh của mỗi RDD

☐ b. Người sử dụng không thể quyết định số các phân mảnh của mỗi RDD

☒ c. Được chia thành các phân mảnh (partition)

☒ d. Có khả năng chịu lỗi

Question 32 CompleteMarked out of 1.00Flag question

Question text

Đâu là cơ chế chịu lỗi của Apache Spark?Select one or more:

☐ a. Chịu lỗi qua cơ chế huyết thống

☐ b. Chịu lỗi qua cơ chế nhân bản

☒ c. Chịu lỗi qua cơ chế lưu lại lịch sử nhiều phiên bản

Question 33 CompleteMarked out of 1.00Flag question

Question text

Ưu điểm của kiến trúc SAN (Storage area network)?Select one or more:

☒ a. Hiệu năng, băng thông tốt hơn với NAS.

☒ b. Quản trị dễ dàng hơn so với NAS.

☐ c. Máy khách có thể kết nối tới SAN bằng đường truyền Ethernet thông thường (Chuẩn kết nối TCP/IP).

Question 34 CompleteMarked out of 1.00Flag question

Question text

Đâu là kỹ thuật có thể được dùng để thích nghi các giải thuật học máy cho dữ liệu lớn? Select one or more:

☐ a. (2) Song song hoá trên Mapreduce hay Spark

☐ b. (3) Các kiến trúc mới xử lý luồng liên tục như mini-batch, complex event processing

☒ c. Tất cả các ý (1), (2), (3)

☐ d. (1) Sub-sampling, principal component analysis, feature extraction và feature selection

☐ e. Các ý (2) và (3)

Question 35 CompleteMarked out of 1.00Flag question

Question text

Phát biểu nào sau đây sai về Kafka?Select one or more:

☒ a. Kafka bảo đảm thứ tự của các message với mỗi topics.

☐ b. Message sau khi được tiêu thụ (consume) thì không bị xoá.

☐ c. Partition được nhân bản ra nhiều brokers.

☐ d.Các topic gồm nhiều partition

Question 36 CompleteMarked out of 1.00Flag question

Question text

Phát biểu nào sau đây sai về Kafka?Select one or more:

☐ a.Tất cả các thao tác ghi, đọc được xử lý bởi leader, follower làm theo leader.

☒ b.Mỗi partition có 1 leader và nhiều followers.

☐ c.Nếu leader bị lỗi, 1 follower sẽ thay thế trở thành leader mới

Question 37 CompleteMarked out of 1.00Flag question

Question text

Thế nào là UNIX semantic?Select one or more:

☒ a.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.

☐ 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.Tệp tin là chỉ đọc, không cho phép cập nhật và ghi đè. Mọi tiến trình đều có thể đọc tệp tin đồng thời.

Question 38 CompleteMarked out of 1.00Flag question

Question text

Vai trò của YARN?Select one or more:

☐ a.Cung cấp các chức năng phối hợp phân tán độ tin cậy cao như quản lý thành viên, bầu cử, giám sát trạng thái hệ thống

☒ b.Quản lý và phân phối tài nguyên trong cụm Hadoop

☐ c.Cung cấp giao diện người dùng mức cao, biến đổi truy vấn thành các job Mapreduce

Question 39 CompleteMarked out of 1.00Flag question

Question text

Đâu là đặc điểm của Spark streaming?Select one or more:

☐ a.Không thể thực hiện các truy vấn SQL

☒ b.Có thể nhận đầu vào là các luồng dữ liệu từ Kafka