



Trịnh Thanh Hải TH24,11

Thiết kế hệ thống (Trường Đại học Kinh doanh và Công nghệ Hà Nội)

# BÀI TẬP TRẮC NGHIỆM BIG DATA

Họ tên:.....Trình Thanh Hải .....

Lớp:.....TH24.11.....

**Dữ liệu của Big Data là loại nào ?**

- A. Structured Data
- B. Semi-Structured Data
- C. Unstructured Data
- D. Tất cả**

**Đặc trưng nào không phải của Big Data**

- A. Volume
- B. Variety
- C. Vision**
- D. Velocity

**NoSQL là ?**

- A. Database
- B. Field
- C. Document**
- D. Collection

**Mục tiêu của NoSQL là gì?**

- A. NoSQL cung cấp một giải pháp thay thế cho cơ sở dữ liệu SQL để lưu trữ dữ liệu dạng văn bản.
- B. Cơ sở dữ liệu NoSQL cho phép lưu trữ dữ liệu không có cấu trúc.
- C. NoSQL không thích hợp để lưu trữ dữ liệu có cấu trúc.
- D. NoSQL là một định dạng dữ liệu mới để lưu trữ các tập dữ liệu lớn.**

**Cloudera phát triển công cụ nào ?**

- A. HCatalog**
- B. Hbase
- C. Impala
- D. Oozie

**Loại nào không phải là CSDL NoSQL ?**

**A. SQL Server**

B. MongoDB

C. Cassandra

D. Không có

**Đâu là một kiểu của CSDL NoSQL**

A. SQL

B. Document databases

C. JSON

**D. Tất cả**

**Chọn đúng 5 đặc trưng cho Big Data ?**

**A. Volume, Velocity, Variety, Veracity, Value**

B. Volume, Videos, Velocity, Variability, Value

C. Volume, Variability, Veracity, Visualization, Value

D. Volume, Velocity, Veracity, Visualization, Value

**Velocity (Tốc độ) là đặc trưng nói về?**

A. Tốc độ gia tăng khối lượng dữ liệu lớn

B. Tốc độ cập nhật dữ liệu lớn

C. Tốc độ xử lý dữ liệu lớn

**D. Tốc độ lưu trữ dữ liệu lớn**

**Variety (Tính đa dạng) là đặc trưng về ?**

A. Kiểu dữ liệu thu thập

B. Kiểu nội dung dữ liệu

**C. Nguồn thu thập dữ liệu**

D. Phương thức xử lý dữ liệu

**Đặc trưng quan trọng nhất của Big Data?**

A. Khối lượng

B. Tính đa dạng

C. Tốc độ

**D. Tất cả**

**Đâu là đặc trưng quyết việc triển khai Dữ liệu lớn?**

A. Khối lượng

B. Giá trị

**C. Tốc Độ**

D. Tính đa dạng

**Tích hợp dữ liệu là quá trình ?**

**A. Kết hợp các dữ liệu không đồng nhất từ nhiều nguồn khác nhau**

B. Sao chép dữ liệu vào CSDL hệ thống để tiến hành phân tích

C. Làm sạch các dữ liệu thu thập được từ các hệ thống thành phần

D. Tăng giá trị từ các tài nguyên dữ liệu đang lưu trữ phân tán

**Thuật ngữ Thuật ngữ Dữ liệu lớn ra đời năm nào?**

A. 1997

B. 2000

C. 1998

**D. 1941**

**Các dạng thức khoa học dữ liệu của Jim Gray?**

A. Thực nghiệm

B. Lý thuyết

C. Tính toán

**D. Tất cả các ý**

**Phát biểu nào không phải dạng thức nghiên cứu Khoa học dữ liệu của Jim Gray?**

A. Khai thác dữ liệu

**B. Mô phỏng**

C. Tính toán

D. Thực nghiệm

**Đâu không phải là phương thức xử lý dữ liệu lớn?**

A. Thu thập (acquire)

**B. Đánh giá (reviews)**

C. Tổ chức (organize)

D. Phân tích (analyze)

**RDBMS là gì ?**

**A. Relational Database Management System**

B. Relat Data Management System

C. Relational Database Microsoft System

D. Tất cả

**Thị trường Big Data bao gồm:**

A. Phần cứng

B. Phần Mềm

**C. Tất cả**

D. Dịch vụ

**Nhược điểm của tích hợp dữ liệu theo phương pháp Tight Coupling**

**A. Độ trễ**

B. Phản hồi truy vấn

C. Phụ thuộc vào nguồn dữ liệu

D. Tất cả

**Nhược điểm của tích hợp dữ liệu theo phương pháp Loose Coupling**

A. Tất cả

B. Chi phí cao

C. Độ trễ

**D. Phụ thuộc mạng / băng thông**

**Công cụ nào hỗ trợ tốt NoSQL?**

A. SAP Data Services

B. Oracle Data Integrator

C. SQL Server Integration Services

**D. Tất cả**

**Công cụ tích hợp dữ liệu Boomi là của?**

**A. Dell**

B. IBM

C. Microsoft

D. SAP

**Công cụ SQL Server Integrator do ai phát triển ?**

A. Oracle

B. IBM

**C. Microsoft**

D. SAP

**SAP Data Service là công cụ để làm gì ?**

A. Lưu trữ dữ liệu

**B. Tích hợp dữ liệu**

C. Phân tích dữ liệu

D. Tất cả

**Hệ quản trị CSLD DynamoDB là kiểu nào ?**

**A. Key value**

B. Wide Column based

C. Document based

D. Graph based

**Hệ quản trị CSLD MongoDB là kiểu nào ?**

**A. Document based**

B. Key value

C. Wide Column based

D. Graph based

**Hệ quản trị CSLD Neo4J là kiểu nào ?**

A. Key value

B. Wide Column based

C. Document based

**D. Graph based**

**Hệ quản trị CSLD IBM Graph là kiểu nào ?**

A. Key value

B. Document based

**C. Graph based**

D. Wide Column based

**Hệ quản trị CSLD Google Big Table là kiểu nào ?**

**A. Key value**

B. Wide Column based

C. Document based

D. Graph based

**Đâu không phải là RDBMS ?**

A. IBM DB2

B. MS SQL Server

C. MS Access

**D. Cassandra**

**Loại nào là dữ liệu Bán cấu trúc ?**

**A. Tất cả**

B. JSON

C. CSV

D. XML

**Loại nào là dữ liệu Unstructured ?**

A. Video

B. XML

**C. Table**

D. Tất cả

**Yếu tố nào quyết định để sử dụng NoSQL**

- A. Tốc độ gia tăng CSDL
- B. Tính đa dạng của dữ liệu
- C. Tốc độ truy cập dữ liệu

**D. Tất cả**

**CSDL nào không phải kiểu Key-Value:**

**A. MongoDB**

- B. DynamoDB
- C. Redis
- D. Riak

**CSDL NoSQL nào không phải kiểu Document:**

- A. MongoDB
- B. CouchDB
- C. Elasticsearch

**D. Riak**

**CSDL nào không phải kiểu Graph:**

**A. InfoGrid**

- B. Hbase
- C. InfiniteGraph
- D. IBM Graph

**CSDL nào không phải kiểu Wide-Column:**

- A. Hbase
- B. Cassandra
- C. BigTable

**D. Dex**

**Ưu điểm của hệ thống HDFS là gì ?**

- A. Lưu trữ phân tán, xử lý song song, khả năng chịu lỗi cao
- B. Lưu trữ song song, xử lý phân tán, tính sẵn sàng cao
- C. Xử lý phân tán song song, khả năng chịu lỗi chấp nhận sai sót

**D. Tất cả**

**Ưu điểm của DFS là gì ?**

- A. Hệ thống lưu trữ song song nên tránh được ảnh hưởng khi một máy chủ hoặc bộ nhớ bị lỗi

B. Hệ thống được sao lưu tại Server thứ hai nên đảm bảo việc cung cấp dữ liệu

**C. Khi một máy chủ hoặc bộ nhớ bị lỗi, hệ thống tệp phân tán vẫn đảm bảo có thể cung cấp dữ liệu ổn định**

- D. Tất cả

## **Ưu điểm của công nghệ Cluster**

### **A. Tất cả**

- B. Hiệu quả chi phí
- C. Tính sẵn sàng cao
- D. Khả năng mở rộng linh hoạt

## **Cluster node có mấy loại ?**

- A. 1
- B. 2
- C. 3**
- D. 4

## **Nguồn của kiến trúc HDFS trong Hadoop có nguồn gốc là**

### **A. Hệ thống tệp phân phối của Google**

- B. Hệ thống tệp phân tán của Yahoo
- C. Hệ thống tệp phân tán của Facebook
- D. Hệ thống tệp phân tán Azure

## **Loại dữ liệu mà Hadoop có thể xử lý là**

- A. Structred (Có cấu trúc)
- B. Semi-structured (Bán cấu trúc)
- C. Unstructured (Không có cấu trúc)
- D. All of the above (Tất cả những điều trên)**

## **YARN là viết tắt của**

- A. Yahoo's another resource name
- B. Yet another resource negotiator**
- C. Yahoo's archived Resource names
- D. Yet another resource need.

## **Điều nào sau đây không phải là mục tiêu của HDFS?**

- A. Phát hiện lỗi và khôi phục
- B. Xử lý tập dữ liệu khổng lồ
- C. Ngăn chặn việc xóa dữ liệu**
- D. Cung cấp băng thông mạng cao để di chuyển dữ liệu

## **Trong HDFS, các tệp không thể**

- A. Đọc
- B. Xóa
- C. Thực thi**
- D. Lưu trữ



**So với RDBMS, Hadoop**

- A. Có tính toàn vẹn dữ liệu cao hơn.
- B. Có giao dịch ACID không
- C. Thích hợp để đọc và truy vấn nhanh
- D. Hoạt động tốt hơn trên dữ liệu phi cấu trúc và bán cấu trúc.**

**Vấn đề chính gặp phải khi đọc và ghi dữ liệu song song từ nhiều đĩa là gì?**

- A. Xử lý khối lượng lớn dữ liệu nhanh hơn.
- B. Kết hợp dữ liệu từ nhiều đĩa.**
- C. Phần mềm cần thiết để thực hiện nhiệm vụ này là cực kỳ tốn kém.
- D. Phần cứng cần thiết để thực hiện tác vụ này là cực kỳ tốn kém.

**Tính năng định vị dữ liệu trong Hadoop có nghĩa là**

- A. lưu trữ cùng một dữ liệu trên nhiều nút.
- B. chuyển vị trí dữ liệu từ nút này sang nút khác.
- C. đồng định vị dữ liệu với các nút tính toán.**
- D. Phân phối dữ liệu trên nhiều nút.

**Các tệp HDFS được thiết kế cho**

- A. Nhiều người viết và sửa đổi ở các hiệu số tùy ý.
- B. Chỉ nối vào cuối tệp.**
- C. Chỉ ghi thành tệp một lần.
- D. Truy cập dữ liệu có độ trễ thấp.

**Công việc khung hadoop được viết bằng**

- A. C ++
- B. Python
- C. Java**
- D. Go

**Cái nào không phải là một trong đặc trưng 3Vs của dữ liệu lớn?**

- A. Vận tốc Velocity
- B. Tính xác thực Veracity**
- C. Khối lượng Volume
- D. Đa dạng Variety

**Điều nào sau đây đúng đối với Hadoop?**

- A. Đây là một khung phân tán.
- B. Thuật toán chính được sử dụng trong đó là Map Reduce
- C. Nó chạy có thể thực thi trên hạ tầng Cloud Computing
- D. Tất cả đều đúng**

**Loại nào sau đây lưu trữ dữ liệu?**

- A. Name node
- B. Data node**
- C. Master node
- D. None of these

**Node nào sau đây quản lý các nút khác?**

- A. Name node
- B. Data node
- C. slave node
- D. None of these**

**Hadoop xử lý khối lượng lớn dữ liệu như thế nào?**

- A. Hadoop sử dụng song song rất nhiều máy. Điều này tối ưu hóa việc xử lý dữ liệu.
- B. Hadoop được thiết kế đặc biệt để xử lý lượng lớn dữ liệu bằng cách tận dụng phần cứng MPP.**
- C. Hadoop gửi mã đến dữ liệu thay vì gửi dữ liệu đến mã.
- D. Hadoop sử dụng các kỹ thuật bộ nhớ đệm phức tạp trên NameNode để tăng tốc độ xử lý dữ liệu.

**MapReduce do công ty nào phát triển ?**

- A. Apache**
- B. Google
- C. IBM
- D. Amazon

**Dịch vụ đám mây nào hỗ trợ tốt cho Big Data ?**

- A. Amazon AWS
- B. Google Cloud
- C. Tất cả**
- D. Microsoft Azure

**Nhược điểm khi triển khai Big Data trên nền tảng Cloud Computing là?**

- A. Nhiều rủi ro cho hệ thống**
- B. Không có ý đúng
- C. Tốc độ xử lý không đảm bảo
- D. Phân tích theo thời gian thực kém

**Mô hình chính dành cho Điện toán đám mây ?**

- A. Cơ sở hạ tầng dưới dạng dịch vụ (IaaS)

**B. Tất cả**

C. Nền tảng dưới dạng dịch vụ (PaaS)

D. Phần mềm dưới dạng dịch vụ (SaaS)

Vũ Thái Hoàng