

MIDTERM CODING EXAMINATION

Course: Introduction to Big Data

Class: 22CQ_21

1. Nội dung thực hiện

Trong bài kiểm tra giữa kỳ này, sinh viên sẽ thực hiện bài làm theo cả hai hình thức: cá nhân và làm việc nhóm. Nội dung chi tiết của các câu hỏi được trình bày trong tập tin **Questions.pdf**.

Yêu cầu cụ thể:

- Mỗi sinh viên cần hoàn thành:
 - **02 câu hỏi bất kỳ ở Mức độ 1 (Level 1)**
 - **01 câu hỏi bất kỳ ở Mức độ 2 (Level 2)**
 - **Câu hỏi số 13**, thực hiện theo nhóm cùng với các thành viên khác.

Phân công trong nhóm:

- Các thành viên trong nhóm cần tự phân chia câu hỏi Mức độ 1 và Mức độ 2 sao cho **không có câu hỏi nào được làm trùng bởi hai người**.
- Ví dụ: nếu nhóm có 4 thành viên (A, B, C, D) và có 8 câu hỏi Mức độ 1, có thể phân chia như sau:
 - A: câu hỏi 1 và 2
 - B: câu hỏi 3 và 4
 - C: câu hỏi 5 và 6
 - D: câu hỏi 7 và 8

Ngôn ngữ lập trình:

- Sinh viên được lựa chọn giữa **Python hoặc Java** để viết chương trình Hadoop MapReduce.
- Tuy nhiên, yêu cầu sử dụng **phương pháp cài đặt cơ bản, không được dùng thư viện hoặc mã nguồn nâng cao**.

Về test case:

- Mỗi sinh viên phải tạo ít nhất **3 test case cho mỗi câu hỏi** đã thực hiện.
- Các test case này cần có **quy mô và độ phức tạp cao hơn** so với dữ liệu mẫu (sample input) được cung cấp trong đề bài.

Qui định về đạo văn:

- Bài làm sẽ bị xem là **đạo văn** nếu **mã nguồn của bất kỳ câu hỏi nào có mức độ giống nhau trên 50%** so với:
 - Bài làm của nhóm khác, hoặc
 - Tài liệu có sẵn trên Internet

- **Hình thức xử lý:**

- Nếu phần đạo văn thuộc về **bài làm cá nhân** ⇒ sinh viên nhận **0 điểm cho toàn môn học**
- Nếu phần đạo văn thuộc về **bài làm nhóm** ⇒ **tất cả thành viên trong nhóm nhận 0 điểm cho toàn môn học**

2. Định danh cá nhân

Sinh viên cần đưa file **hadoop-env.jar** vào môi trường đã cài đặt Hadoop bằng cách tải trực tiếp file này trên trình duyệt của môi trường hoặc sử dụng các thao tác **upload/mount** nếu là môi trường ảo để đưa file này từ hệ thống chủ vào.

Tiếp theo, sinh viên cần xác định vị trí của file hadoop-env.jar trong filesystem của môi trường thông qua đường dẫn **"/path/to/jar/file.jar"** cùng với vị trí mà sinh viên muốn chạy file, ví dụ **"/current/path"**. Lưu ý đảm bảo rằng tài khoản mà mình đang dùng có quyền ghi vào thư mục ở vị trí hiện tại **/current/path** đồng thời cũng có quyền chạy file jar ở đường dẫn **/path/to/jar/file.jar** để quá trình chạy thành công.

Sau đó, sinh viên chạy lệnh sau trên command line: **java -jar /path/to/jar/file.jar <YOUR_HDFS_PORT> <StudentID>** Sau khi lệnh kết thúc, sinh viên có thể thấy 2 dòng xác nhận MSSV được in ra màn hình.

Cuối cùng, sinh viên tìm file **<MSSV>_verification.txt** trong thư mục của vị trí chạy **/current/path** và nộp file này.

3. Tổ chức bài nộp

Mỗi nhóm chỉ nộp một tập tin nén (.zip) duy nhất vào đường dẫn nộp bài được chỉ định trên Moodle. Tên tập tin zip phải đặt theo tên nhóm, ví dụ: **CQ00.zip**, được tạo từ **thư mục có cùng tên (CQ00)**.

Cấu trúc thư mục CQ00 như sau:

1. **Tập tin readme.txt**

Ghi rõ:

- Họ tên và MSSV của từng thành viên
- Danh sách các câu hỏi ở Mức độ 1 và Mức độ 2 mà mỗi thành viên đã thực hiện

2. **Thư mục con cho từng thành viên, đặt tên theo MSSV**

- Ví dụ, nhóm có 4 thành viên thì thư mục CQ00 sẽ có 4 thư mục con tương ứng với 4 MSSV.

3. **Thư mục Level3 (dành cho câu hỏi nhóm – Câu hỏi số 13)**

Bao gồm:

- Mã nguồn chương trình
- 03 test case
- Video trình bày kết quả
- Tập tin định danh của **thành viên thực hiện chạy chương trình**

Mỗi thư mục con của sinh viên bao gồm:

a. Thư mục cho mỗi câu hỏi ở Mức độ 1

Tên thư mục: Level1 – Question X, trong đó X là mã số câu hỏi tương ứng.

Mỗi sinh viên sẽ có **2 thư mục như vậy**, tương ứng với 2 câu hỏi mức 1.

Nội dung trong mỗi thư mục:

- Mã nguồn chương trình
- **03 test case** (lớn hơn sample input)
- Video trình bày kết quả: bao gồm dữ liệu sử dụng, quá trình chạy, và kết quả đầu ra
- Tập tin định danh **<MSSV>_verification.txt**

b. Thư mục cho câu hỏi ở Mức độ 2

Tên thư mục: Level2 – Question Y, với Y là mã số câu hỏi tương ứng.

Nội dung tương tự như thư mục cho mỗi câu hỏi ở Mức độ 1:

- Mã nguồn chương trình
- 03 test case
- Video trình bày kết quả
- Tập tin định danh **<MSSV>_verification.txt**

Yêu cầu định dạng video:

- Video phải được lưu dưới định dạng phổ biến (như .mp4, .avi, v.v.)
- Đảm bảo độ phân giải đủ cao để dễ dàng quan sát nội dung hiển thị
- Nếu tổng dung lượng của các video vượt quá 18MB, sinh viên có thể tải video lên Google Drive hoặc YouTube. Đường dẫn video phải thể hiện rõ thời gian cập nhật không vượt quá hạn nộp bài (thông qua dấu thời gian của nền tảng lưu trữ).

4. Thang điểm

STT	Nội dung	Điểm	STT	Nội dung	Điểm
1	Câu hỏi thứ nhất của Mức độ 1		3	Câu hỏi Mức độ 2	
1.1	Mã nguồn	0.5	3.1	Mã nguồn	1
1.2	03 test cases	0.5	3.2	03 test cases	1
1.3	Video	0.5	3.3	Video	1
2	Câu hỏi thứ hai của Mức độ 1		3	Câu hỏi Mức độ 3	
2.1	Mã nguồn	0.5	3.1	Mã nguồn	1
2.2	03 test cases	0.5	3.2	03 test cases	1
2.3	Video	0.5	3.3	Video	2

* Nội dung tương ứng sẽ không được chấm nếu thiếu mã nguồn, video, hoặc tập tin định danh.

* Câu hỏi Mức độ 3 được đánh giá chung cho mọi thành viên.

CHÚC CÁC BẠN THỰC HIỆN BÀI KIỂM TRA THÀNH CÔNG.