

# **Môn học “Nhập môn Phân tích Dữ liệu Lớn – 71ITDS40303”**

## **Câu hỏi Thi Kết hợp (Tự luận + Vấn đáp) – HK242**

### **Phần I Tổng quan**

**Câu 1:** Hãy trình bày định nghĩa và các thuộc tính cơ bản liên quan tới khái niệm Dữ liệu Lớn, Tại sao cần áp dụng các công cụ lưu trữ và phân tích dữ liệu đặc thù (Hadoop HDFS, MapReduce) khi thực hiện phân tích Dữ liệu Lớn so với Dữ liệu truyền thống ?

**Câu 2:** Hãy trình bày định nghĩa và các thuộc tính cơ bản liên quan tới Dữ liệu Lớn. Liệt kê một số công nghệ và giải pháp phổ biến trong phân tích Dữ liệu Lớn. Cho một thí dụ thực tế áp dụng công nghệ và giải pháp phân tích Dữ liệu Lớn mà sinh viên đã tìm hiểu. Đây là khó khăn chính khi thực hiện phân tích Dữ liệu Lớn ?

**Câu 3:** Hãy trình bày định nghĩa và các thuộc tính cơ bản liên quan tới Dữ liệu Lớn. Mô tả hoạt động của công cụ phân tích Dữ liệu Lớn phổ biến trong môi trường điện toán đám mây Google Cloud Platform.

**Câu 4:** Hãy trình bày định nghĩa và các thuộc tính cơ bản liên quan tới Dữ liệu Lớn. Mô tả hoạt động của công cụ phân tích Dữ liệu Lớn phổ biến trong môi trường điện toán đám mây Microsoft Azure

**Câu 5:** Hãy trình bày định nghĩa và các thuộc tính cơ bản liên quan tới Dữ liệu Lớn. Mô tả hoạt động của công cụ phân tích Dữ liệu Lớn trong môi trường điện toán đám mây Amazon Web Servers.

**Câu 6:** Hãy trình bày định nghĩa và các thuộc tính cơ bản liên quan tới Dữ liệu Lớn. Mô tả hoạt động của công cụ phân tích Dữ liệu Lớn trong môi trường Hadoop DevOps trong môi trường điện toán đám mây riêng, kết nối MS Visual Studio Code bằng SSH với máy chủ của Khoa CNTT.

**Câu 7:** Hãy trình bày định nghĩa và các thuộc tính cơ bản liên quan tới Dữ liệu Lớn. Vai trò của ngôn ngữ lập trình Python, các thư viện Python hỗ trợ phân tích dữ liệu lớn là gì ? Cho thí dụ thực tế mô tả hoạt động phân tích dữ liệu lớn bằng lập trình Python.

**Câu 8 :** Hãy trình bày định nghĩa và các thuộc tính cơ bản liên quan tới Dữ liệu Lớn. Hãy trình bày mô hình và giải thích cơ chế hoạt động của Hadoop Map/ Reduce.

**Câu 9 :** Hãy trình bày định nghĩa và các thuộc tính cơ bản liên quan tới Dữ liệu Lớn. Mô tả kiến trúc của hệ thống tập tin phân tán HDFS và vai trò của HDFS trong phân tích Dữ

liệu Lớn. Để quản trị HDFS, chúng ta có thể sử dụng công cụ nào ? Hãy liệt kê một số lệnh cơ bản để quản trị HDFS.

**Câu 10:** Hãy trình bày định nghĩa và các thuộc tính cơ bản liên quan tới Dữ liệu Lớn. Các thành phần chính trong kiến trúc phân tích dữ liệu lớn Hadoop là gì ? Giải thích các khái niệm HDFS, Namenode, Datanode, Resource Manager

## **Phần 2 Thực hành Phân tích Dữ liệu Lớn (Hadoop)**

- + Trình bày quy trình đăng nhập vào máy chủ bằng tài khoản cá nhân (định danh mã sinh viên), **khởi tạo** môi trường DevOps Hadoop
- + Trình bày các lệnh **kiểm tra** cấu hình hoạt động của DevOps Hadoop
- + Trình bày các lệnh **khởi động** dịch vụ DevOps Hadoop
- + Trình bày các lệnh thực thi **nhập xuất dữ liệu** với HDFS
- + Trình bày nguyên tắc hoạt động của **mapper** và **reducer** trong bài toán phân loại và đếm từ trong tập tin văn bản.

## **Phần 3 Đồ án Môn học**

Cho biết tên Chủ đề, nội dung và kết quả chính đã thực hiện trong Đồ án Môn học

**Sinh viên trả lời 01 Câu hỏi trong Phần 1, trình bày các nội dung liên quan tới thực hành LAB trong Phần 2 và Đồ án Môn học trong Phần 3**