# Nội Dung tìm hiểu:

## Big Data:

* Khái niệm
* Vai trò
* Đặc điểm
* Ứng dụng

## Phương pháp, thuật toán sử lý BigData

* Mô tả
* Ưu điểm
* Các công nghệ phổ biến

Lấy khoảng 2 đến 3 ví dụ về các phương pháp và thuật toán sử lý Big Data

## Công cụ sử lý BigData

* Apache Hadoop
* Apache Spark

Tìm hiểu về: Khái niệm, thành phần, hiệu năng, chi phí triển khai, thêm nếu có…

## Apache Spark

### Định Nghĩa

* Định nghĩa
* Lịch sử phát triển
* Tính năng

Tìm hiểu chi tiết

### Kiến trúc

* Spark Core
* RDDs
* Driver program
* Spark Context
* Cluster Manager
* Executors

Tìm hiểu về

* Định nghĩa
* Vai trò
* Cách khởi tạo

### Thành phần và thư viện của Spark

* Spark SQL, các thành phần như Dataset, DataFrame, công cụ như Catalyst Optimizer và Tungsten Execution Engine để tối ưu hóa truy vấn
* Spark Streamming
* SparkMlib
* Spark GraphX
* Giới thiệu qua về SparkR, PySpark

Tìm hiểu về định nghĩa, vai trò cách thức triển khai đối với thư viện và thành phần của nó , còn các thư viện khác thì chỉ cần tìm hiểu định nghĩa và vai trò, đối với Spark SQL thì tìm hiểu thêm về cách thức triển khai các câu lệnh truy vấn

### Nguồn dữ liệu

Tìm hiểu về cách kết nối Spark kết nối đến nguồn dữ liệu như:

* HDFS
* JSON
* JDBC
* NoSQL
* … có thể thêm nếu thích

### Cài đặt Spark

Tìm hiểu về cách cài đặt Spark trên (có hình ảnh minh họa, yêu câu là cài được Spark trên máy):

* Local
* Cluster

### Ứng dụng của Spark

Tìm hiểu chi tiết về:

* Phân tích dữ liệu lớn
* Machine Learnning
* Sử lý thời gian thực
* Các lĩnh vực khác,…

## So sánh với Hadoop

* Hiệu năng
* Chi phí
* Cách sử dụng

## Tổng kết

* Ưu điểm
* Nhược điểm

Phân chia:

Thao tìm hiểu về phần Big data, thuật toán sử lý bigdata

Tâm tìm hiểu về phần cài đặt spark, so sánh hiệu năng với hadoop

Quang tìm hiểu về phần công cụ sử lý Bigdata, tổng kết

Sơn tìm hiểu về phần nguồn dữ liệu, ứng dụng của Spark,

Thành tìm hiểu về kiến trúc, thư viện của Spark