Những thứ chuẩn bị cho Powerpoint thuyết trình

# Về Big data (trong khoảng 1 – 3 slide, có thể cắt giảm cũng được)

* Giới thiệu về Big data, về tính chất, đặc điểm
* ứng dụng của Bigdata trong đời sống (không cần chi tiết quá, để dành chỗ cho các thứ khác)
* giới thiệu về các công cụ để sử lý big data (làm tối giản nhất có thể)

# Về Spark

## Giới thiệu (trong khoảng 2 – 4 slide)

* giới thiệu về spark,
* có thể giới thiệu qua về kiến trúc của spark, hoặc các thành phần của nó
* đặc điểm nổi bật của spark : tốc độ, dễ sử dụng, dễ mở rộng, có tính module

## Đi sâu (còn lại)

Giới thiệu qua (1 slide)

### Kiến trúc của spark (4 - 5 silde)

* giới thiệu qua về kiến trúc (ở đây làm chi tiết về RDD khoảng 2 slide dữ liệu trong file doc có)
* về thành phần và thư viện

### Chi tiết về thành phần mỗi cái một slide

* spark core
* spark sql (chi tiết)
* spark mlib
* spark streamming
* ..

### Cách thức hoạt động của spark (khoảng 2 slide)

### Các tính năng chính của spark

* về sử lý dữ liệu phân tán
* về sử lý thời gian thực
* ….

### Ứng dụng của spark (1 slide)

### Về ưu nhược điểm

### Ví dụ đơn giản (rồi tôi gửi demo cái kết quả thi thpt)

Mọi thứ chỉ mang tính tham khảo, con số chỉ là tương đối, tùy thuộc vào nội dung, có thể thêm bớt, chỉnh sửa

Nếu trong doc có những khái niệm dài dòng, thì có thể tóm gọn lại hoặc có thể tham khảo trên các kênh Internet

### Cách thức hoạt động, định nghĩa, và một số khái niệm có thể tham khảo trên internet của spark (có thể tham khảo trên intenet)

<https://vinahost.vn/apache-spark-la-gi/#ftoc-heading-8>