**TRƯỜNG ĐH CÔNG NGHỆ THÔNG TIN**

**KHOA CÔNG NGHỆ PHẦN MỀM**

-----🙞🙜🕮🙞🙜-----

****

**BÁO CÁO**

**Đề tài:**

**Xây dựng hệ thống recommendation**

**dựa trên Content base filtering**

**trên nền tảng Big Data với**

**Hadoop Mapreduce**

Giảng viên hướng dẫn:

Thạc sĩ: Phạm Thi Vương

Sinh viên thực hiện:

12520053 Nguyễn Viết Danh

# TÌM HIỂU VỀ BIG DATA

## Big data

### Giới thiệu

Big data là một thuật ngữ rộng cho bộ dữ liệu lớn hay phức tạp mà các ứng dụng xử lý dữ liệu truyền thống là không đủ để thực hiện. Những thách thức bao gồm phân tích, nắm bắt, quản lý dữ liệu, tìm kiếm, chia sẻ, lưu trữ, truyền, trực quan hóa, và bảo mật thông tin. Thuật ngữ này thường dùng chỉ đơn giản để sử dụng các phân tích hoặc các phương pháp tiên tiến khác để thu được chính xác giá trị từ dữ liệu và hiếm khi tập dữ liệu đó có kích thước cụ thể. Độ chính xác trong big data có thể hỗ trợ cho việc ra quyết định tự tin. Và quyết định tốt hơn có thể có nghĩa là hiệu quả hoạt động cao hơn, giảm chi phí và rủi ro.

Phân tích các tập dữ liệu có thể tìm thấy mối tương quan mới, để tìm ra xu hướng kinh doanh, ngăn ngừa bệnh tật, chống tội phạm,... Các nhà khoa học, doanh nhân, học viên truyền thông, quảng cáo cũng như chính phủ thường xuyên gặp khó khăn với các tập dữ liệu lớn trong các lĩnh vực bao gồm tìm kiếm trên Internet, tài chính và kinh doanh qua mạng. Các nhà khoa học gặp phải những hạn chế trong công tác e-Science, bao gồm khí tượng, gen, sự kết nối, mô phỏng vật lý phức tạp, nghiên cứu sinh học và môi trường.

Data sets grow in size

## Ứng dụng

## Các công cụ

# tài liệu tham khảo

Wikipedia

<https://en.wikipedia.org/wiki/Big_data>