Chapter 3 Part 2 (B) Organizational Information Systems and Their Impact

MIS 02500 Issues in MIS

Lecture Overview

- ☐ What is big data?
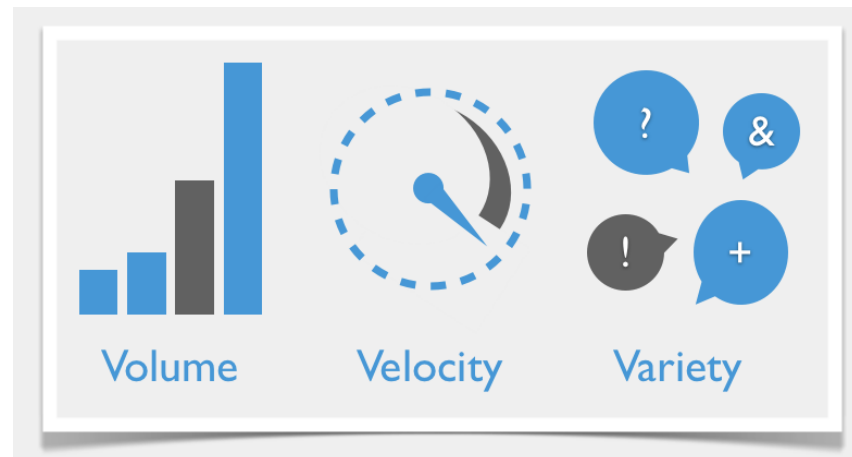
- ☐ The challenges of big data
  - ■ Technology challenges
  - ■ Management challenges

- ☐ Reading material
  - ■ Piccoli & Pigni text Chapter 3, *From Business Intelligence and Big Data* and *Big Data: Technology* sections
  - ■ Two articles (that you can find on weekly overview)
    - ☐ Big Data: The Management Revolution (*Harvard Business Review*)
    - ☐ Big data: are we making a big mistake? (*Financial Times*)

Quiz - Big Data Articles

Big Data

- The massive quantities of unstructured and semi-structured data from Internet and networked services and applications

  - **Volume** is the amount of digital data that organizations have to store and manage

  - **Velocity** is the speed of creation and use of new digital data. This dimension relates to the need to gather, process and communicate information in real-time

  - **Variety** is the kind of digital data that organizations have to store and manage

(For details on the three Vs of big data, see this week's article "Big Data: The Management Revolution" and Piccoli & Pigni text Chapter 3, *From Business Intelligence and Big Data* section)

Volume   Velocity   Variety

The Technology Challenges of Big Data

- ☐ Challenges
  - ■ How should the data be stored?
  - ■ How do data from various sources be integrated?
  - ■ How are data retrieved and disseminated?
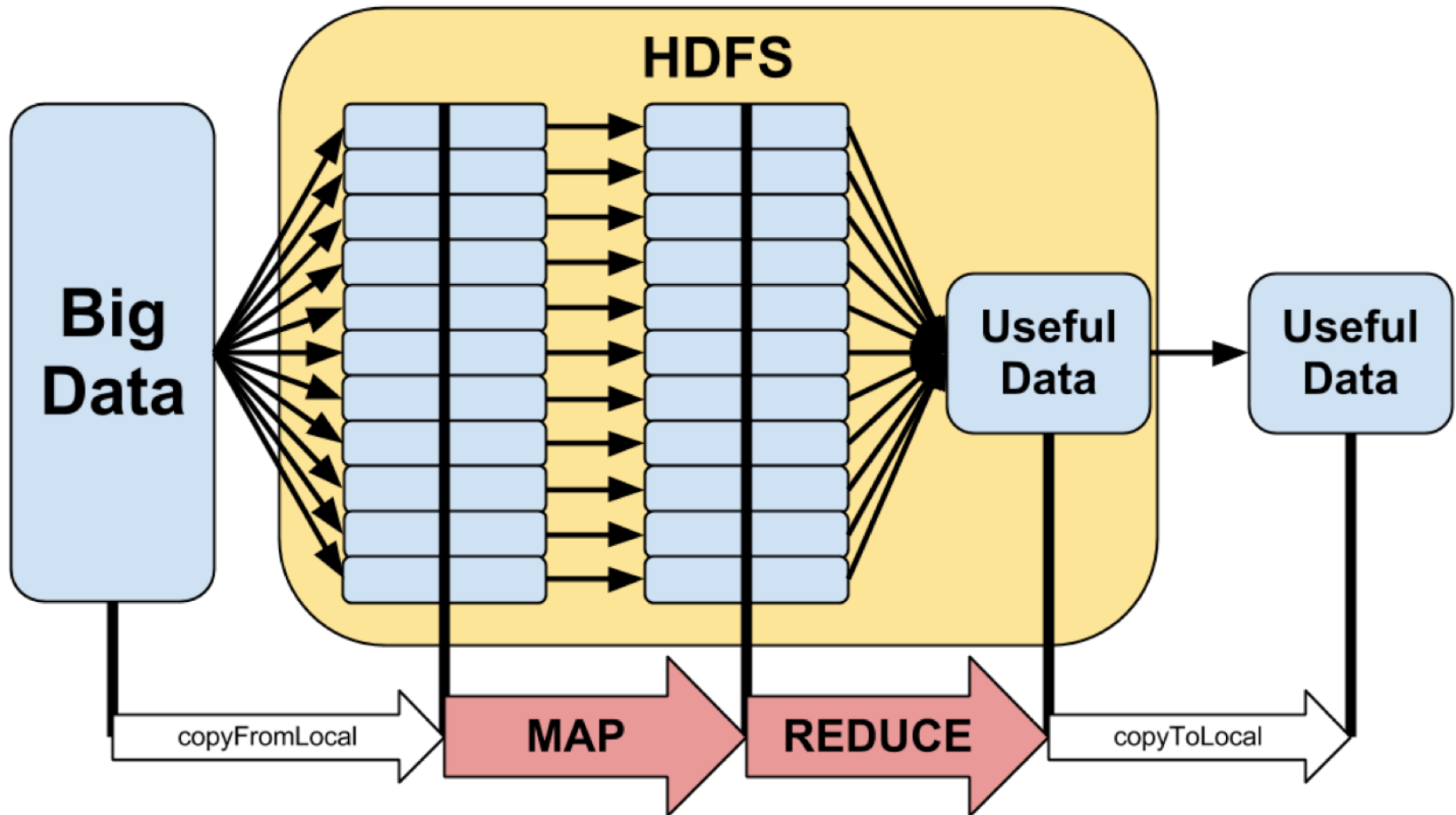- ☐ Requires new technologies and tools
  - ■ Hadoop

Hadoop

- Open-source software framework from Apache
  - http://hadoop.apache.org/
- Designed for big data
- Breaks data task into sub-problems and distributes the processing to many inexpensive computer processing nodes
- Combines result into smaller data set that is easier to analyze
- Key services
  - Hadoop Distributed File System (HDFS)
  - MapReduce
- Video: What is Hadoop?
  - Question: What are the two fundamental concepts that help understand Hadoop?

YouTube link for the video is provided **in this week's overview (Google Doc)**

Conceptual Overview of Map-Reduce

Source: http://www.glennklockwood.com/data-intensive/hadoop/overview.html

Big Data: Five Management Challenges

- Leadership

- Talent management

- Technology

- Decision making

- Company culture

Source: "Big Data: The Management Revolution"

- https://hbr.org/2012/10/big-data-the-management-revolution

Big Data: Four Pieces of Cautionary Advice

- Data analysis results can be wrong
  - Do not ignore false positive
- Not every single data point can be captured
  - Be cautious of sampling error and sampling bias
- Causation still matters
- Scientific and statistical models are still needed

Source: Big data: are we making a big mistake?
  - https://www.ft.com/content/21a6e7d8-b479-11e3-a09a-00144feabdc0