Business School
UNIVERSITY OF COLORADO DENVER

Information Systems Program

Module 5
Physical Design and Governance of Data Warehouses

Lesson 3: Big data issues

Lesson Objectives

- Provide definitions of big data
- Discuss dimensions of big data
- Define units of big data with examples
- Reflect on value areas for big data

Big Data Definitions

- Doug Laney (2001)

  "E-commerce, in particular, has exploded data management challenges along three dimensions: volume, velocity and variety."

- McKinsey Global Institute (2011)

  "datasets whose size is beyond the typical ability of database software to capture, store, and analyze"

- John Akred, Founder, Silicon Valley Data Science

  " "Big Data" refers to a combination of an approach to informing decision making with analytical insight derived from data, and a set of enabling technologies that enable that insight to be economically derived from at times very large, diverse sources of data."

Big Data Timeline

NASA paper
•1997

How much information?
•2000

Laney report
•2001

How much information?
•2003

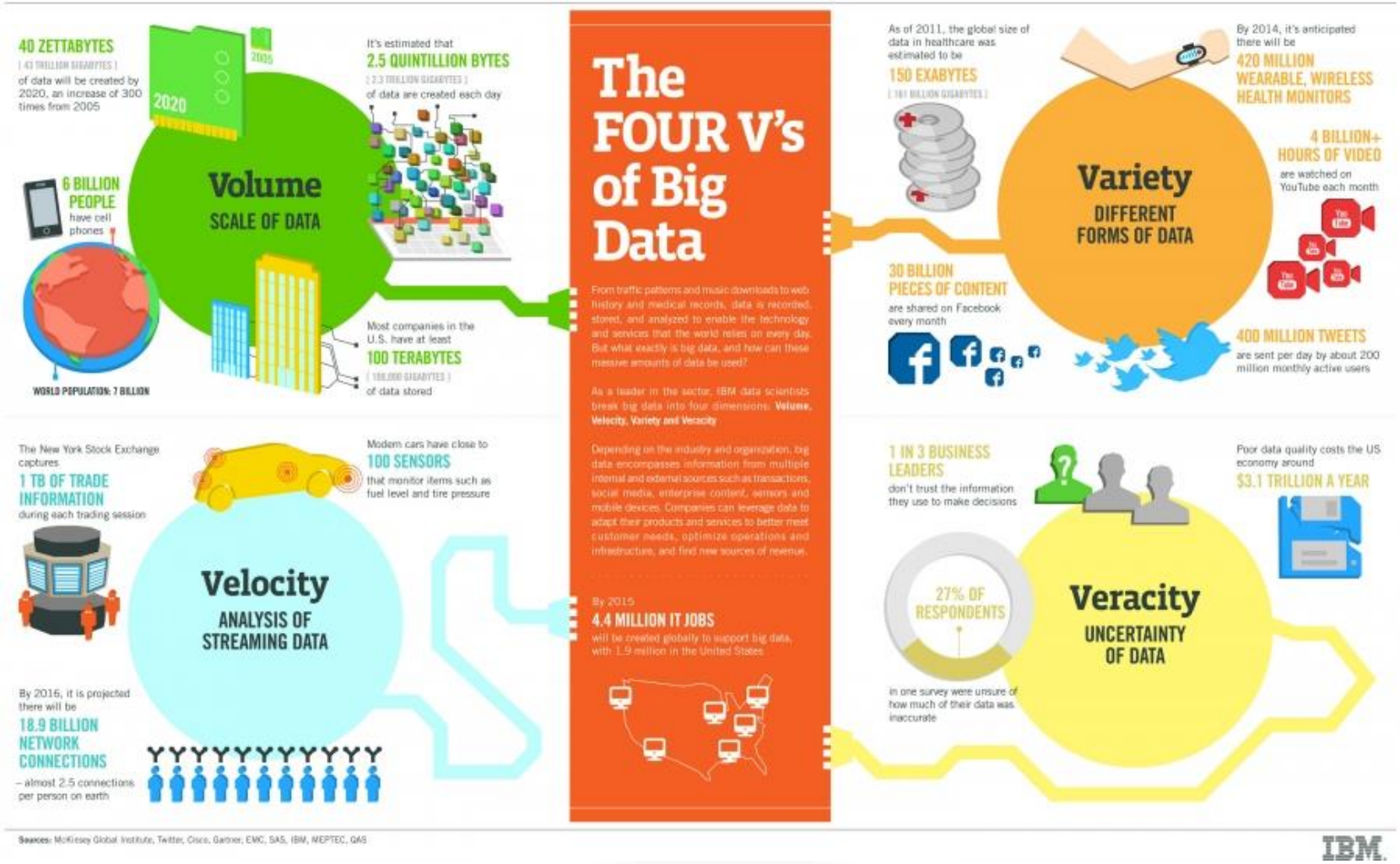
Google MapReduce
•2004

Hadoop
•2005

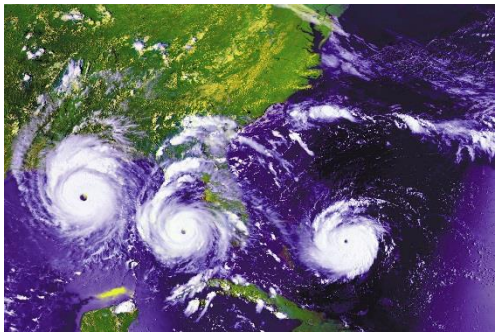
Government funding essay
•2008

Hadoop 2
•2013

IBM Big Data Dimensions



Sources of Big Data



Data Unit Sizes

- Units
  - Kilobyte: 1,024 bytes or 1,000 bytes
  - Megabyte: 1,024 (1,000) KB
  - Gigabyte: 1,024 (1,000) MB
  - Terabyte: 1,024 (1,000) GB
  - Petabyte: 1,024 (1,000) TB
  - Exabyte: 1,024 (1,000) PB
  - Zettabyte: 1,024 (1,000) EB
  - Yottabyte: 1,024 (1,000) ZB
- Confusion between base 2 and base 10 units

Examples of Big Data Units

Data Unit	Big Data Example
Terabyte (TB) 1,024 (1,000) GB	Typical hard drive capacity on a personal computers
Petabyte (PB) 1,024 (1,000) TB	Teradata Database 14 capacity of 50 PB
Exabyte (EB) 1,024 (1,000) PB	Estimate of global IP traffic in 2021 by Cisco: 278 EB/month
Zettabyte (ZB) 1,024 (1,000) EB	Cisco estimate of total volume of IP traffic in 2021: 3.3 ZB. IDC estimate of digital universe in 2020: 40 ZB.
Yottabyte (YB) 1,024 (1,000) ZB	Estimate storage capacity of U.S. National Security data center capacity is 1.0 YB; High definition video of all human activity: 100 YB

Big Data Trends and Examples

- 40 percent data growth projected by IDC in 2014
  - 1.7 megabytes per second per individual by 2020
  - 4.4 zettabytes to 44 zettabytes by 2020
- 3.5 billion queries per day using Google
- 60 billion messages per day and 64 billion video views per day on Facebook

Big Data Value Areas

Promotions

Risk management

Inventory management

Surveillance

Military

Entertainment

Summary

- Much hype but also substantial importance about big data
- Know data units
- Relative concept partially dependent on organization
- Understand drivers of big data and opportunities