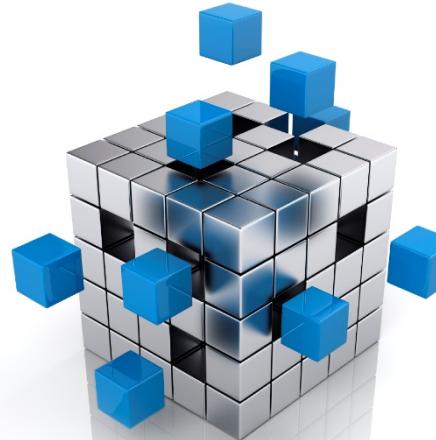

Lecture 12

Big Data

Summary – last week

Summary

- Real-time Data Warehousing
 - Real-time Data Processing Challenges
 - Real-time Join Algorithms



This week

- Big Data
 - Big Data Introduction
 - Characteristics of Big Data
 - Challenges



What is Big Data?

- Definitions
 - Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, manage, and process data within a tolerable elapsed time¹.
 - Big data is a set of techniques and technologies that require new forms of integration to uncover large hidden values from large datasets that are diverse, complex, and of a massive scale².

¹Snijders, C.; Matzat, U.; Reips, U.-D. (2012). "Big Data": Big gaps of knowledge in the field of Internet". *International Journal of Internet Science* 7: 1–5

²Ibrahim; Targio Hashem, Abaker; Yaqoob, Ibrar; Badrul Anuar, Nor; Mokhtar, Salimah; Gani, Abdullah; Ullah Khan, Samee (2015). "big data" on cloud computing: Review and open research issues". *Information Systems* 47: 98–115

Who's Generating Big Data?

Social Media



Social media and networks
(all of us are generating data)



Scientific instruments
(collecting all sorts of data)



Mobile devices
(tracking all objects all the time)



Sensor technology and networks
(measuring all kinds of data)

- The progress and innovation is no longer hindered by the ability to collect data
- But, by the ability to manage, analyze, summarize, visualize, and discover knowledge from the collected data in a timely manner and in a scalable fashion

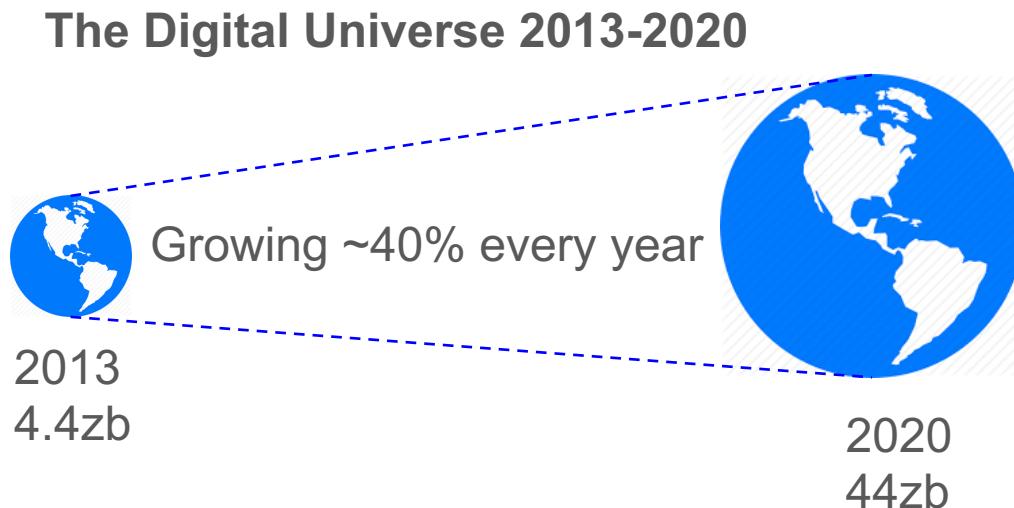
How much data?

2018 *This Is What Happens In An Internet Minute*



How much data?

- We are living in a digital universe.
- Data volume is increasing exponentially.
 - Growing 40% a year into the next decade¹
 - 4.4 zettabytes to 44zb from 2013 to 2020



¹<http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm>

How much Data?

- Google processes 100 PB/day; 3 million servers
- Facebook has 300 PB + 500 TB/day; 35% of world's photos
- YouTube 1000 PB video storage; 4 billion views/day
- Twitter processes 124 billion tweets/years
- SMS message – 6.1 trillion/year
- US credit cards – 1.4 billion Cards; 20 billion transactions/year



640K ought to be
enough for
anybody.

The Model has Changed

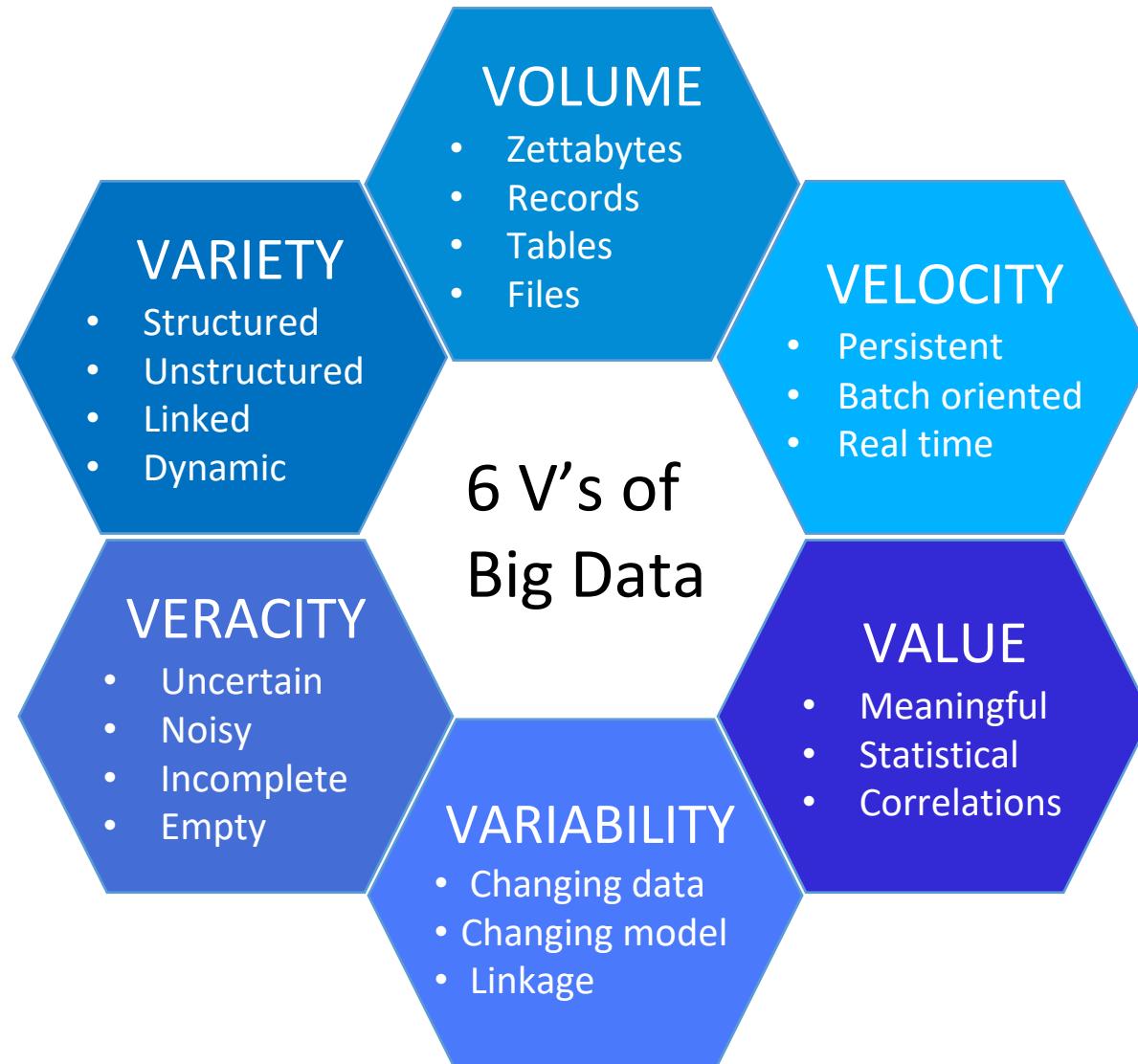
- The Model of Generating/Consuming Data has Changed
- Old Model: Few companies are generating data, all others are consuming data



- New Model: all of us are generating data, and all of us are consuming data



Big Data V's

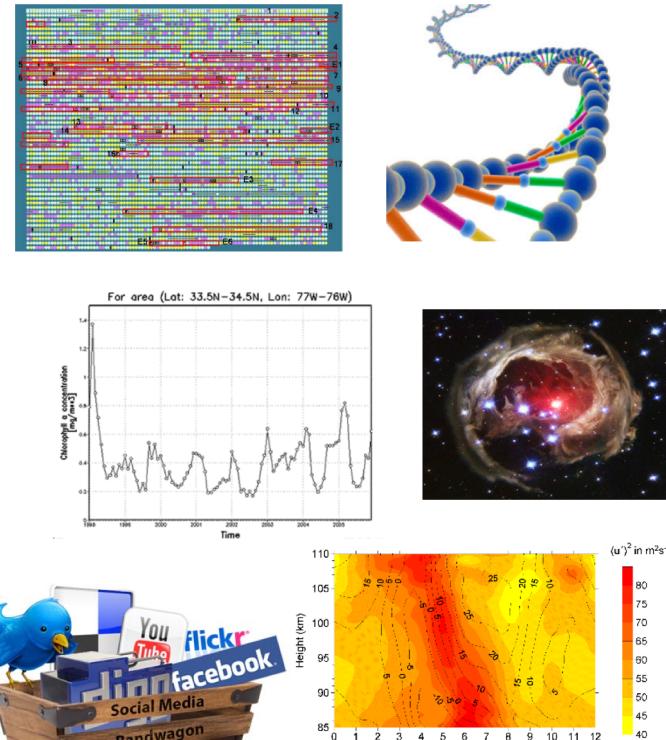


7th V is for Visualisation.

Big Data Challenges

- Various formats, types, and structures
- Text, numerical, images, audio, video, sequences, time series, social media data, multi-dim arrays, etc...
- Static data vs. streaming data
- A single application can be generating/collecting many types of data

To extract knowledge → all these types of data need to be linked together



Big Data Challenges (continued)

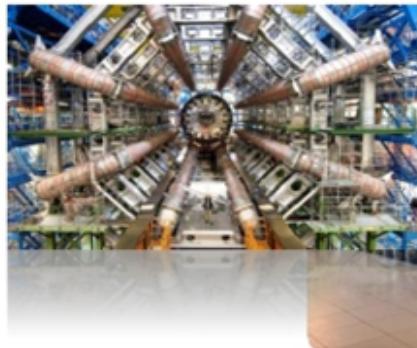
- Data is being generated fast and need to be processed fast
- Online Data Analytics
- Late decisions → missing opportunities
- Examples

E-Promotions: Based on your current location, your purchase history, what you like → send promotions right now for store next to you

Healthcare monitoring: sensors monitoring your activities and body → any abnormal measurements require immediate reaction



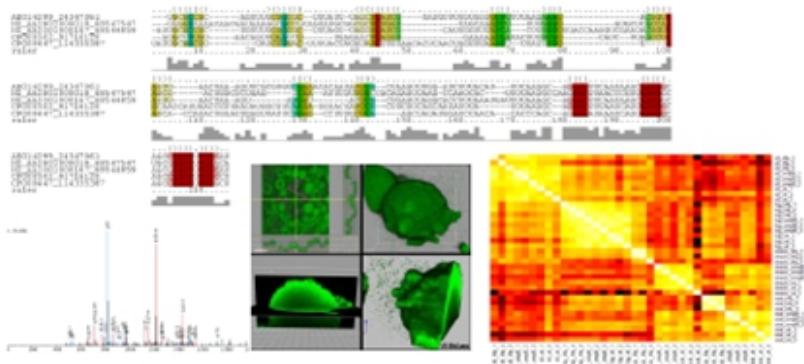
Big Data Challenges (continued)



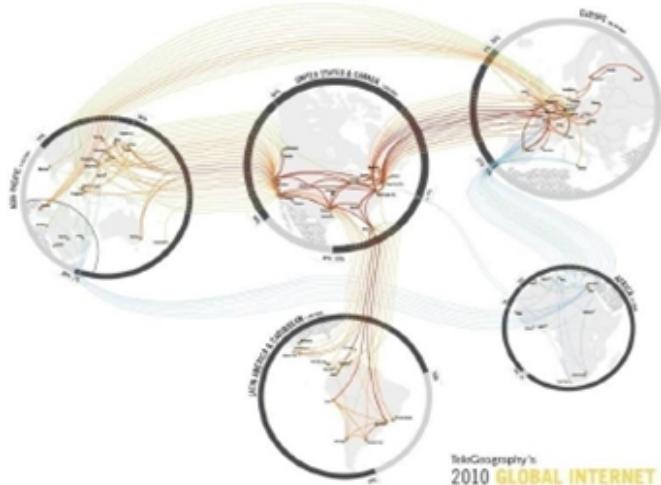
Large volumes



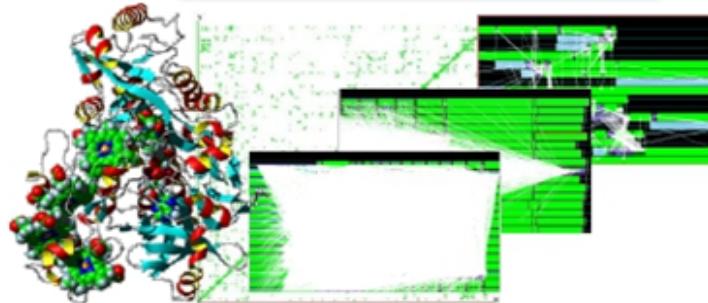
High-throughput



Heterogeneity



Relating and Linking



Complexity

Value of Big Data Analytics

Big Data in Changing Healthcare



Value of Big Data Analytics

Smart Cities – Big Data to Ease Transport Congestion



Value of Big Data Analytics

Big Data in Social Media



90% the available data in the world was collected over just the previous two years and 80% of that data comes from “unstructured” sources, like social media¹.

Value of Big Data Analytics

Big Data in Retail Sector

amazon

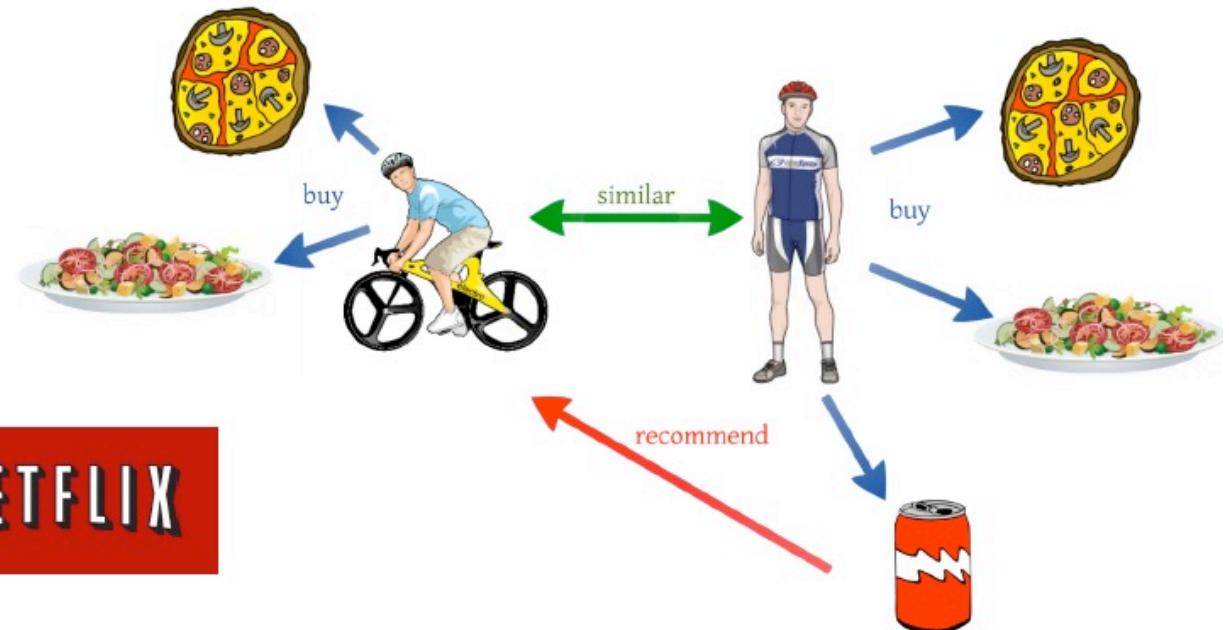
Spotify

P

LinkedIn

NETFLIX

ebay



Big Data Landscape

Big Data Landscape

Vertical Apps



Log Data Apps



Ad/Media Apps



Business Intelligence



Analytics and Visualization



Data As A Service



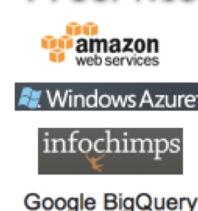
Analytics Infrastructure



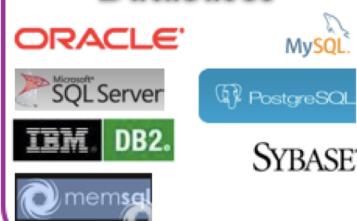
Operational Infrastructure



Infrastructure As A Service



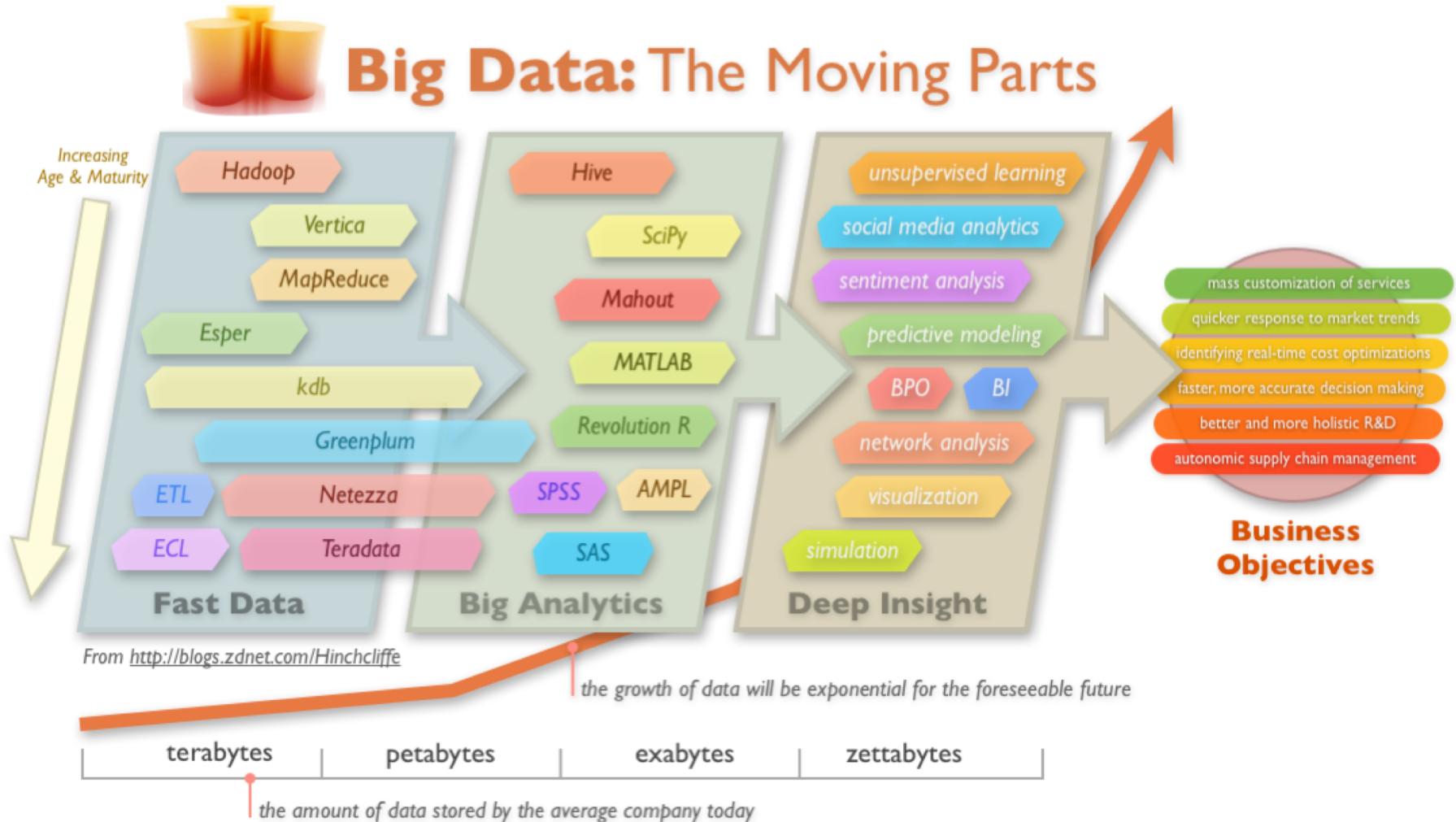
Structured Databases



Technologies



Big Data Technology



Summary

Summary

- Big Data
 - What is Big Data
 - Who Generate Big Data
 - Characteristics of Big Data - 3V's
 - Challenges
 - Big Data Technologies

