Introduction To Big Data

Thaveewat Khanan

https://github.com/bobbylovemovie/trainbigdata

bobbylovemovie@gmail.com


โพสต์ทูเดย์
การเมือง
ปมคำถามพ่วง
ตั้งนายกฯคนนอก
"ต่อให้พรรคการเมืองเห็นก่อน"
A4
เศรษฐกิจดิจิทัล
'ข้อมูลคืออำนาจ'
big data
SOCIAL MEDIA
ผลตอบแทนลงทุนคอนโดวูบ


สกู๊ป
โพสต์ทูเดย์
สร้างธุรกิจจาก Big Data
เร่งประสิทธิภาพภาษี
Big Data


10 อาชีพที่ถูกคุกคาม
จาก Big Data และ Machine Learning
โดย...พ.อ.ดร.เศรษฐพงค์ มะลิสุวรรณ
www.nationtv.tv
NationTV22

What is Big Data ?

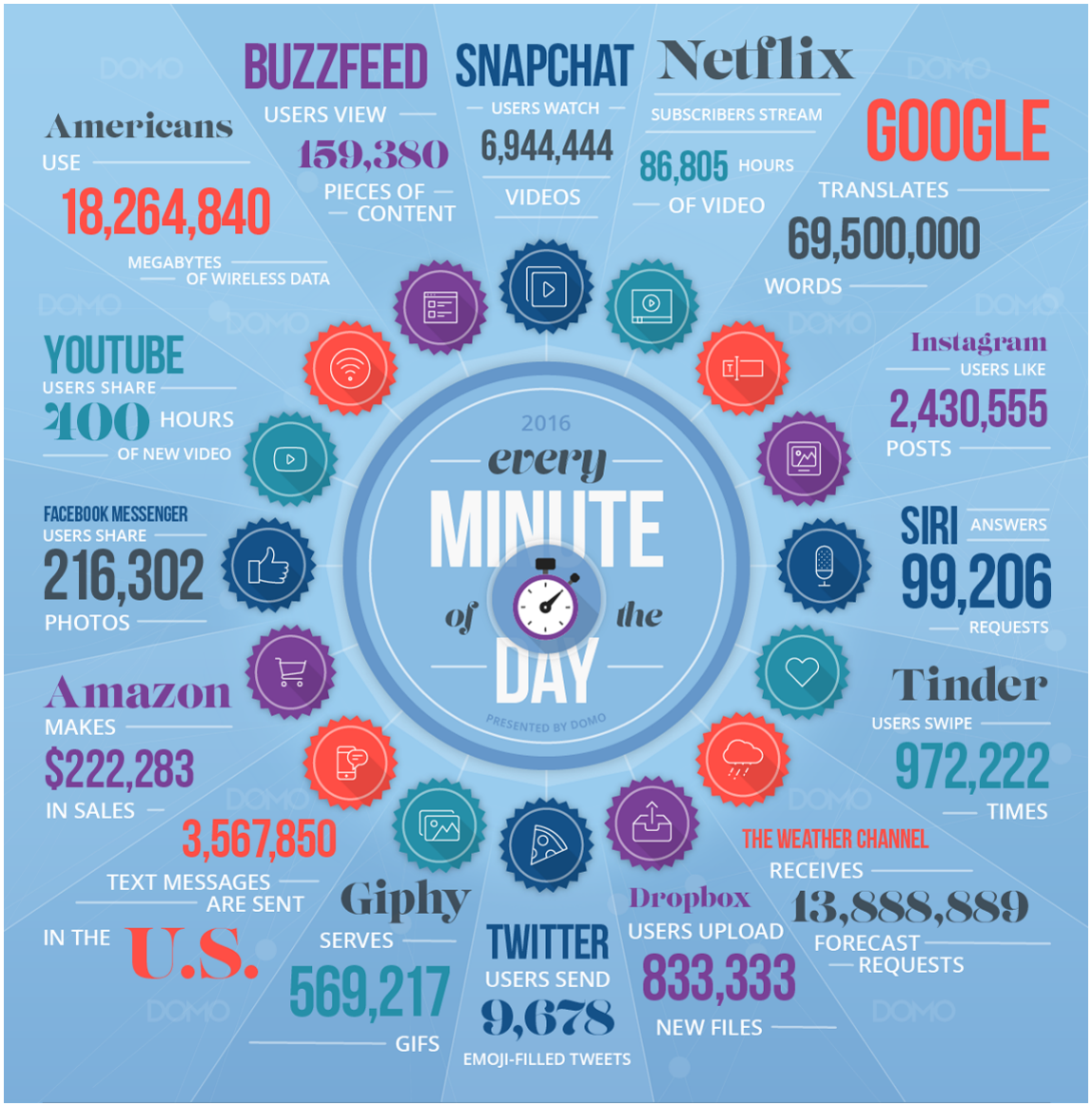
CHULALONGKORN BUSINESS SCHOOL
FLAGSHIP FOR LIFE

“ Big data is data that exceeds the processing capacity of conventional database systems.
The data is too big, moves too fast,
or doesn’t fit the structures of your database architectures.
To gain value from this data,
you must choose an alternative way to process it. ”

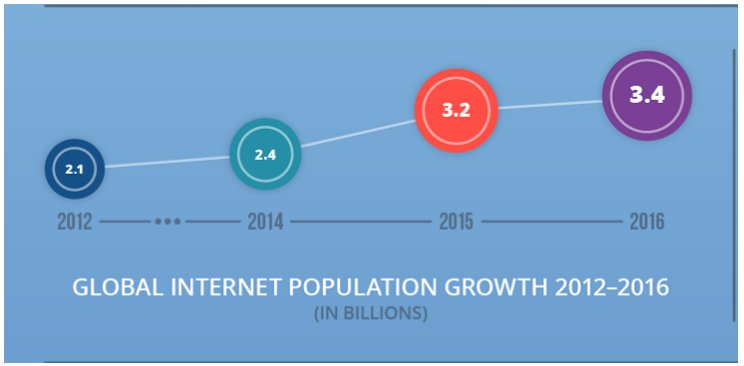
Big Data Now: O'Reilly Media

Why Big Data ?

CHULALONGKORN BUSINESS SCHOOL
FLAGSHIP FOR LIFE



DATA NEVER SLEEP 4.0



https://www.domo.com/blog/2016/06/data-never-sleeps-4-0/

Facebook Usage Statistics

CHULALONGKORN BUSINESS SCHOOL
FLAGSHIP FOR LIFE

June 2014

- 829 million daily active users on average
- 654 million mobile daily active users on average
- 1.32 billion monthly active users
- 1.07 billion mobile monthly active users
- Approximately 81.7% of our daily active users are outside the US and Canada

June 2016

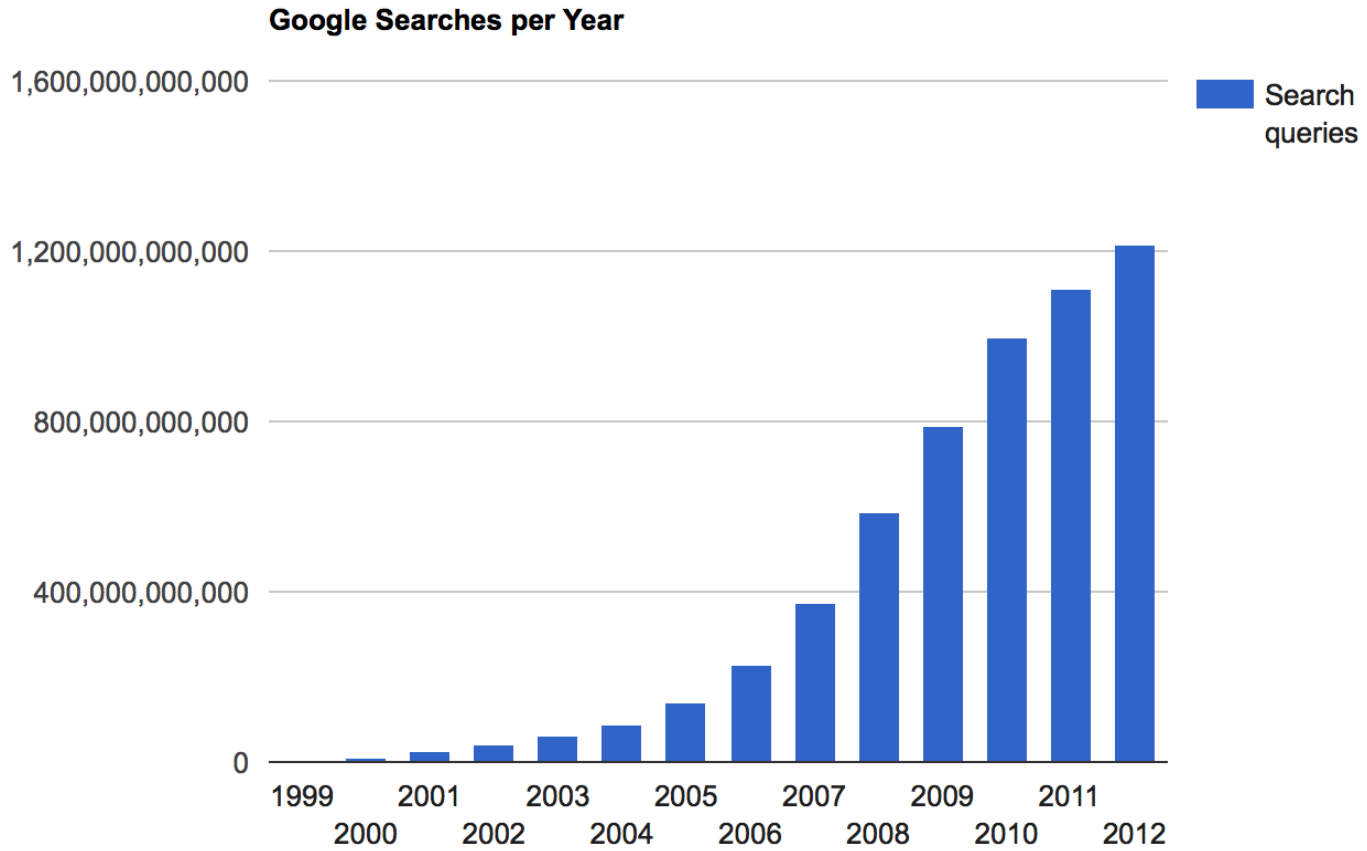
- 1.13 billion daily active users on average for June 2016
- 1.03 billion mobile daily active users on average for June 2016
- 1.71 billion monthly active users as of June 30, 2016
- 1.57 billion mobile monthly active users as of June 30, 2016
- Approximately 84.5% of our daily active users are outside the US and Canada

Source :http://newsroom.fb.com/company-info/

Google Usage Statistics

CHULALONGKORN BUSINESS SCHOOL
FLAGSHIP FOR LIFE

Google now processes over **40,000 search** queries every second which translates to over **3.5 billion searches per day** and **1.2 trillion searches per year** worldwide



Source: http://www.internetlivestats.com/google-search-statistics/

in 1 second, each and every second, there are...

CHULALONGKORN BUSINESS SCHOOL
FLAGSHIP FOR LIFE

- **7,370 Tweets sent in 1 second**
- **745 Instagram photos uploaded in 1 second**
- **56,645 Google searches in 1 second**
- **132,881YouTube videos viewed in 1 second**
- **2,529,971Emails sent in 1 second**
- **38,625 GB of Internet traffic in 1 second**

Source : http://www.internetlivestats.com/one-second/#google-band

Three Characteristics of Big Data

CHULALONGKORN BUSINESS SCHOOL
FLAGSHIP FOR LIFE

Volume

- Volumes of data are larger than those conventional relational database infrastructures can cope with

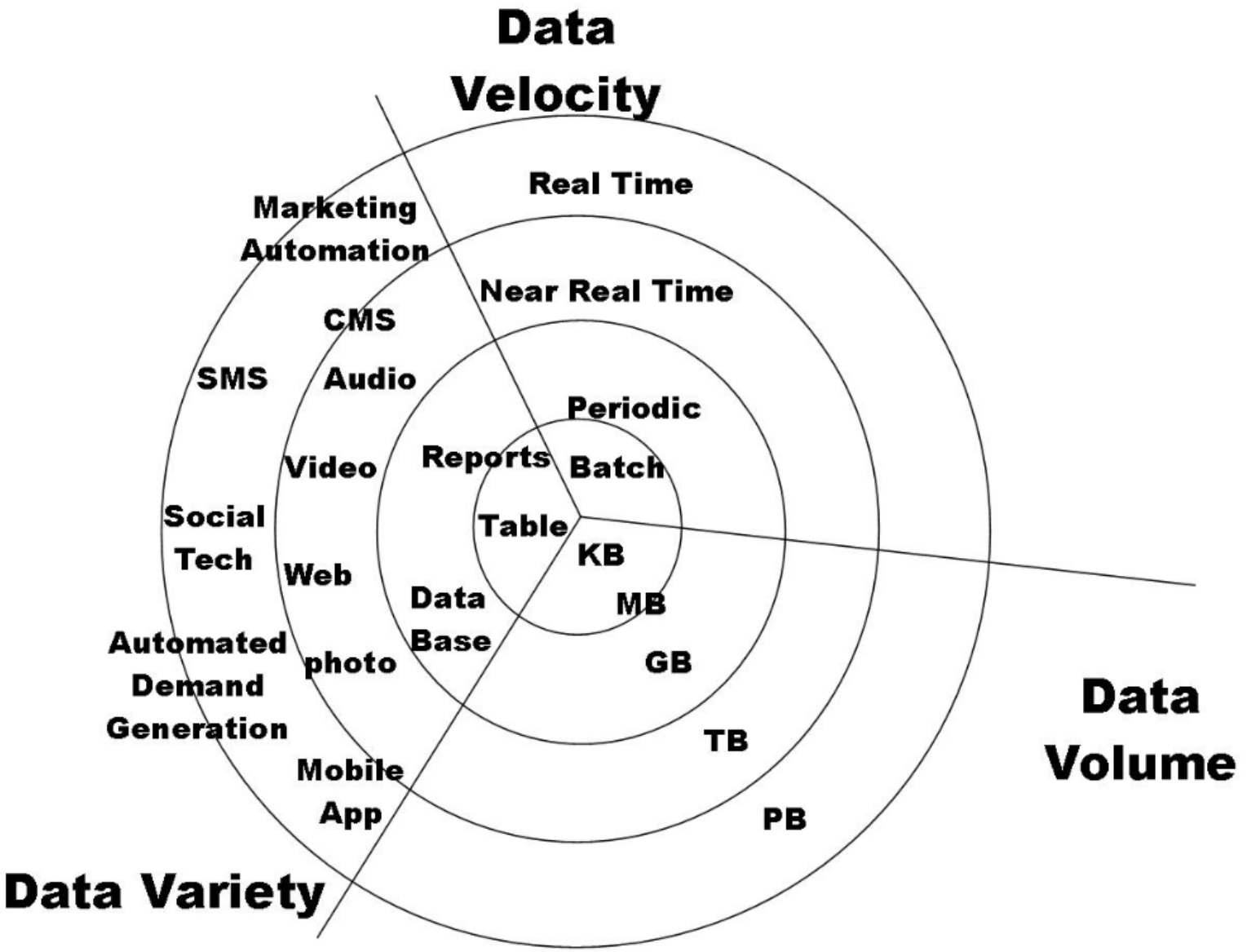
Velocity

- Rate at which data flows in is much faster.
  - Mobile event and interaction by users.
  - Video, image , audio from users

Variety

- the source data is diverse, and doesn't fall into neat relational structures eg. text from social networks, image data, a raw feed directly from a sensor source.

Source Introduction to Big Data: Dr. Putchong Uthayopas

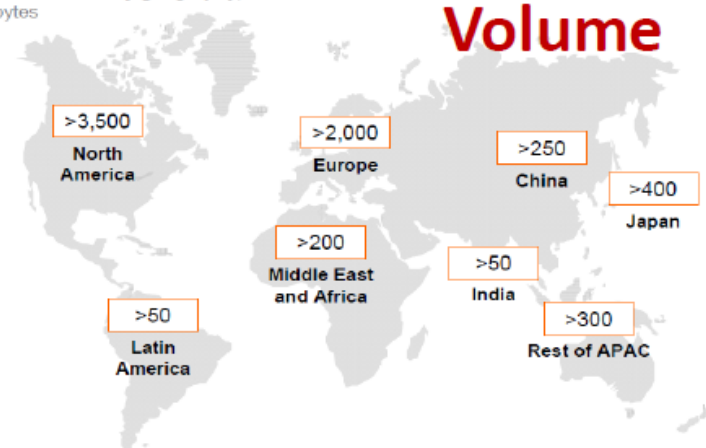

Data
Velocity
Real Time
Near Real Time
Periodic
Batch
Marketing
Automation
CMS
SMS
Audio
Video
Reports
Table
Social
Tech
Web
Data
Base
KB
MB
GB
TB
PB
Automated
Demand
Generation
photo
Mobile
App
Data Variety
Data
Volume

Big Data = Volume, Variety and Velocity (3Vs)

Volume

Amount of new data stored varies across geography

New data stored[1] by geography, 2010
Petabytes

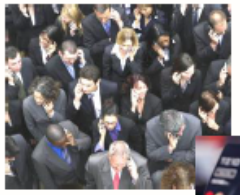


1 New data stored defined as the amount of available storage used in a given year; see appendix for more on the definition and assumptions.
SOURCE: IDC storage reports; McKinsey Global Institute analysis

Variety

People to Machine

People to People



Machine to Machine

facebook, StumbleUpon It!, LinkedIn, plaxo, YAHOO! 360° BETA, twitter, Windows Live Spaces, utterli, remember the milk, YouTube, reddit

Velocity

30 billion pieces of content are shared on Facebook every month.

4 billion hours of video are watched on YouTube each month

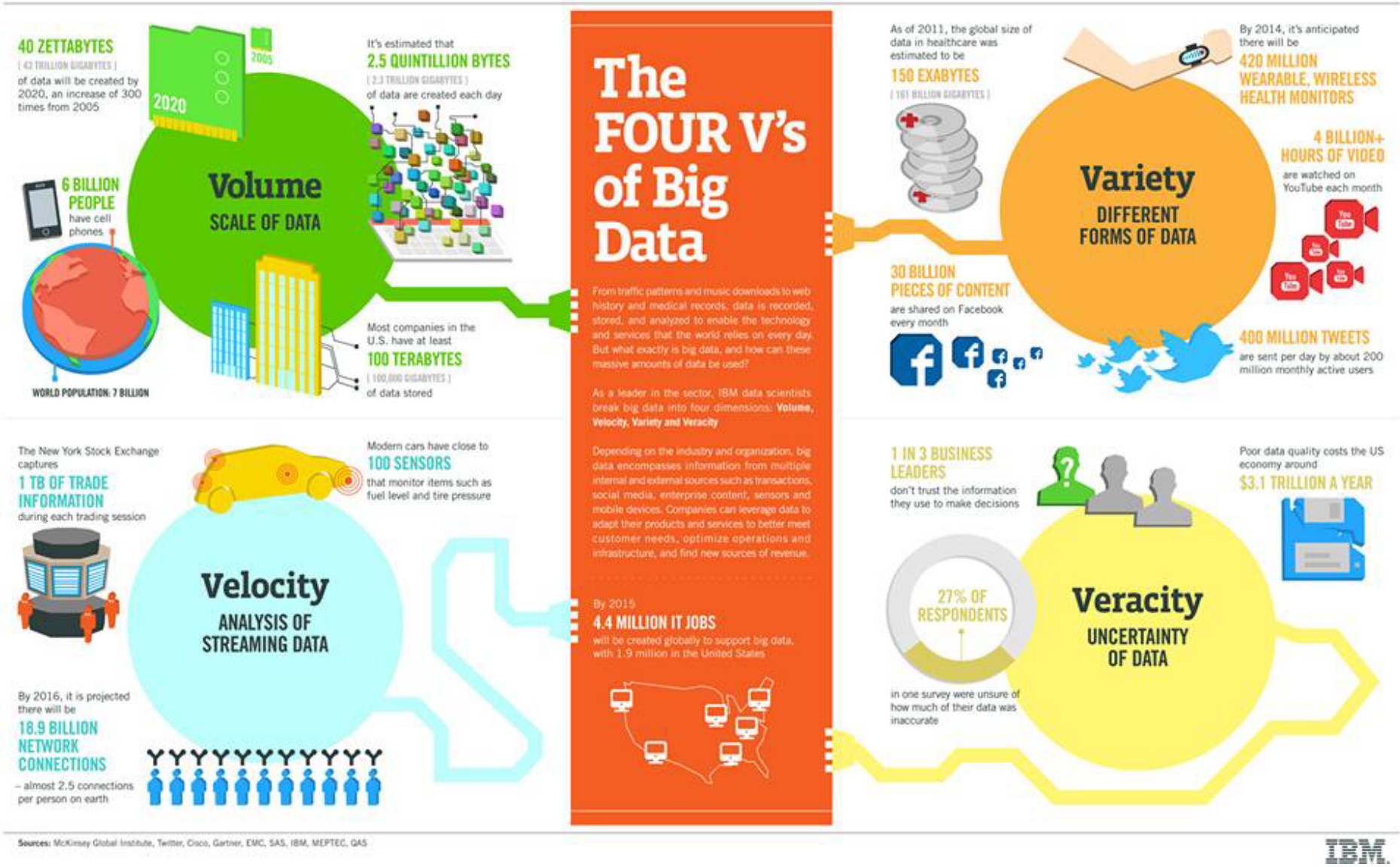
400 Million Tweets are sent per day 200M monthly active users

Source: IBM

Scale

1000 **kilobytes** = 1 Megabyte
1000 **Megabytes** = 1 Gigabyte
1000 **Gigabytes** = 1 Terabyte
1000 **Terabytes** = 1 Petabyte
1000 **Petabytes** = 1 Exabyte
1000 **Exabytes** = 1 Zettabyte
1000 **Zettabytes** = 1 Yottabyte
1000 **Yottabytes** = 1 Bronobyte
1000 **Bronobytes** = 1 Geopbyte

4Vs of Big Data



Volume
SCALE OF DATA

40 ZETTABYTES
[43 TRILLION GIGABYTES]
of data will be created by 2020, an increase of 300 times from 2005

It's estimated that
2.5 QUINTILLION BYTES
[2.3 TRILLION GIGABYTES]
of data are created each day

6 BILLION PEOPLE
have cell phones

WORLD POPULATION: 7 BILLION

Most companies in the U.S. have at least
100 TERABYTES
[100,000 GIGABYTES]
of data stored

The New York Stock Exchange captures

1 TB OF TRADE INFORMATION

during each trading session

Modern cars have close to

100 SENSORS

that monitor items such as fuel level and tire pressure

Velocity

ANALYSIS OF STREAMING DATA

By 2016, it is projected there will be

18.9 BILLION NETWORK CONNECTIONS

– almost 2.5 connections per person on earth

As of 2011, the global size of data in healthcare was estimated to be

150 EXABYTES

[161 BILLION GIGABYTES]

30 BILLION PIECES OF CONTENT

are shared on Facebook every month

Variety

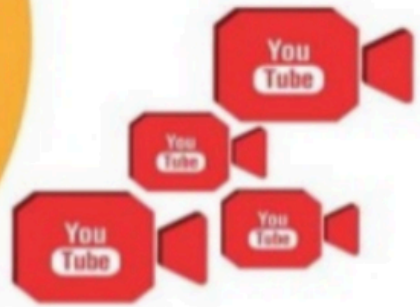
DIFFERENT FORMS OF DATA

By 2014, it's anticipated there will be

420 MILLION WEARABLE, WIRELESS HEALTH MONITORS

4 BILLION+ HOURS OF VIDEO

are watched on YouTube each month



400 MILLION TWEETS

are sent per day by about 200 million monthly active users

CHULALONGKORN BUSINESS SCHOOL
FLAGSHIP FOR LIFE

1 IN 3 BUSINESS LEADERS

don't trust the information they use to make decisions

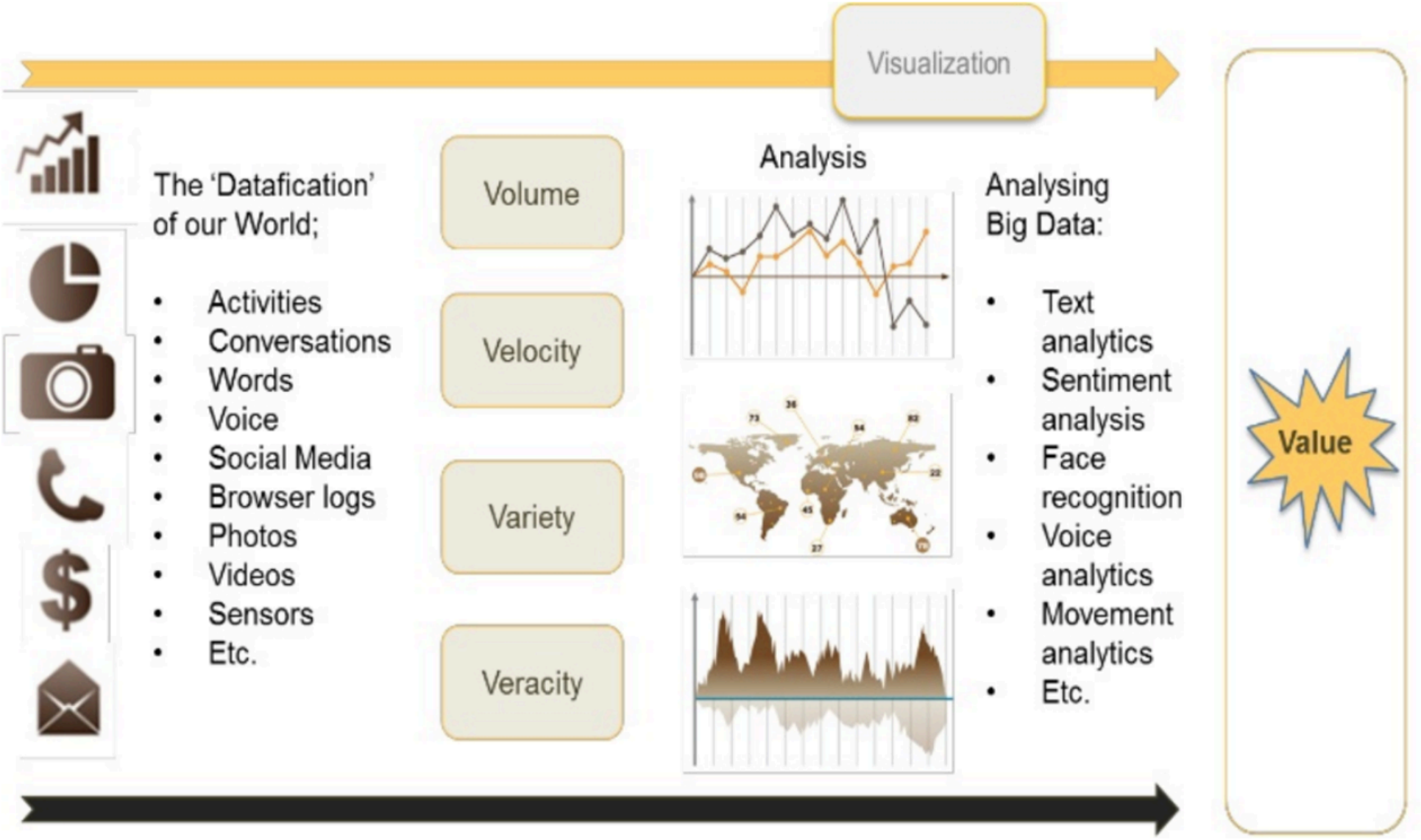
Poor data quality costs the US economy around

$3.1 TRILLION A YEAR

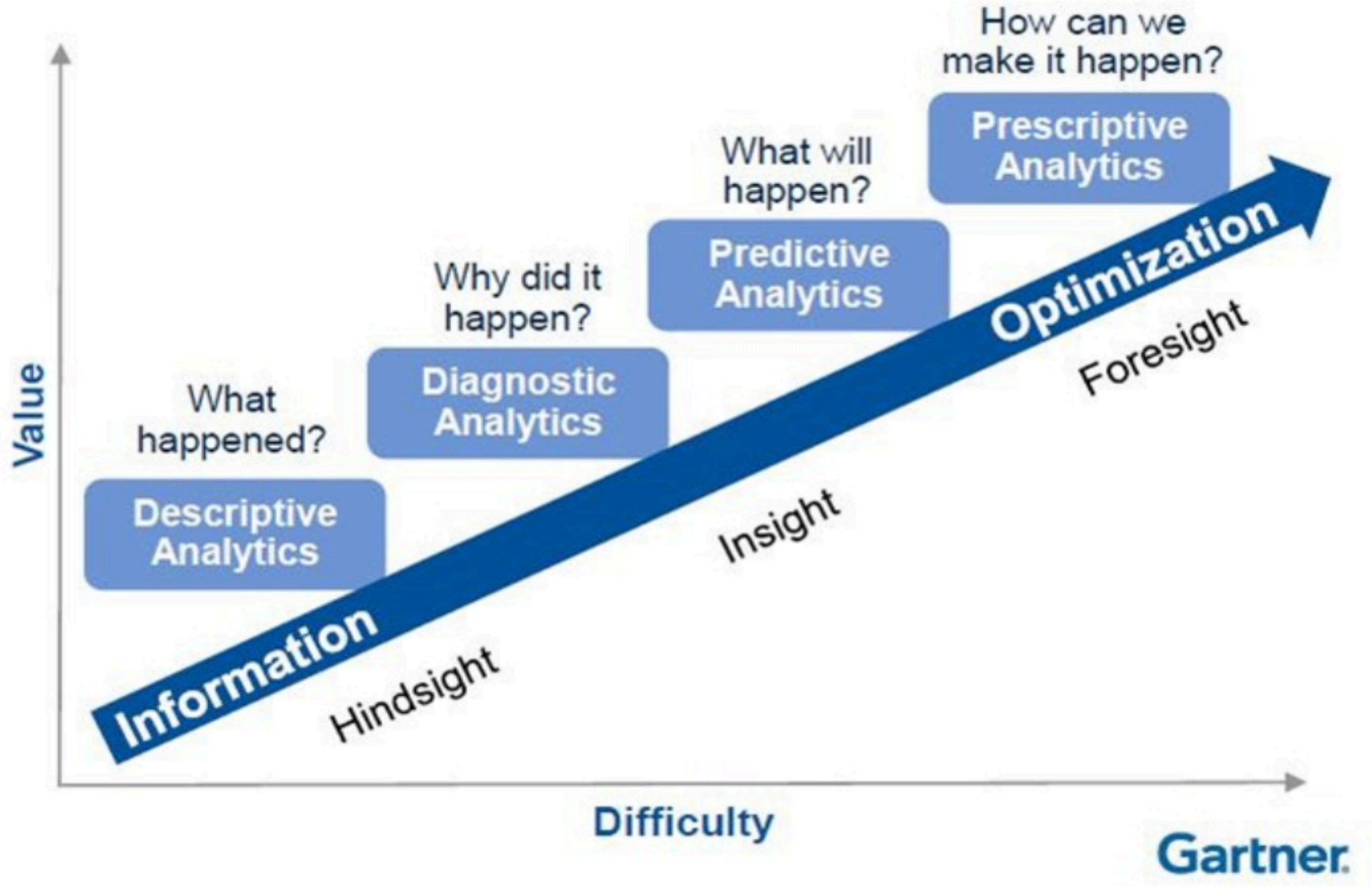
in one survey were unsure of how much of their data was inaccurate

Veracity

UNCERTAINTY OF DATA

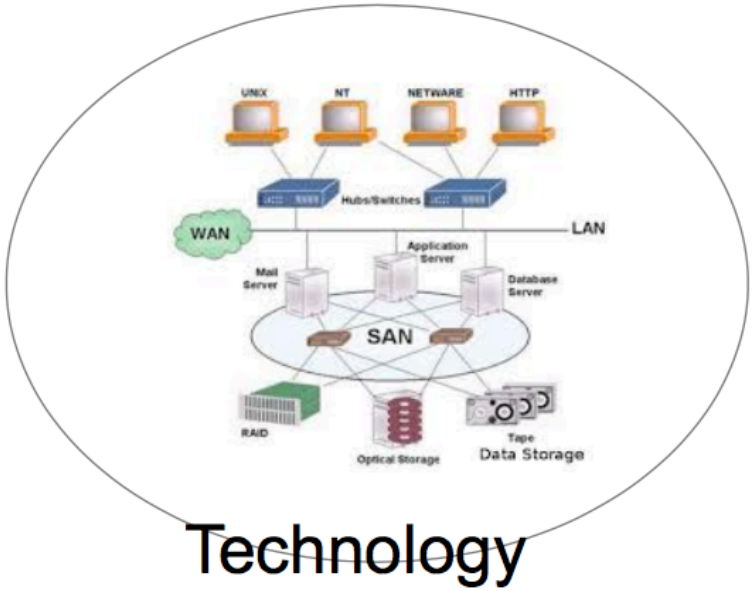


Source: Bernard Marr

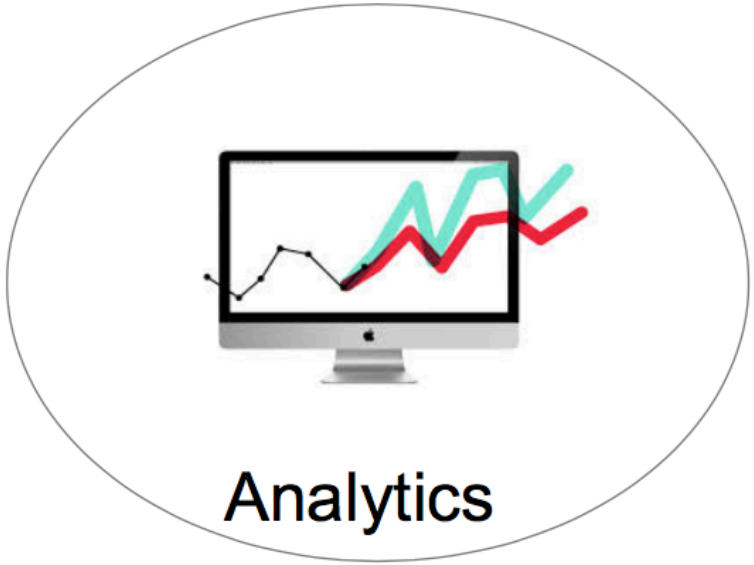

How can we make it happen?
Prescriptive Analytics
What will happen?
Predictive Analytics
Why did it happen?
Diagnostic Analytics
What happened?
Descriptive Analytics
Optimization
Foresight
Insight
Hindsight
Information
Value
Difficulty
Gartner

	Traditional Analytics (BI)	vs	Big Data Analytics
Focus on	• Descriptive analytics • Diagnosis analytics		• **Predictive analytics** • **Data Science**
Data Sets	• Limited data sets • Cleansed data • Simple models		• Large scale data sets • More types of data • Raw data • Complex data models
Supports	**Causation:** what happened, and why?		**Correlation**: new insight More accurate answers

Data Sources



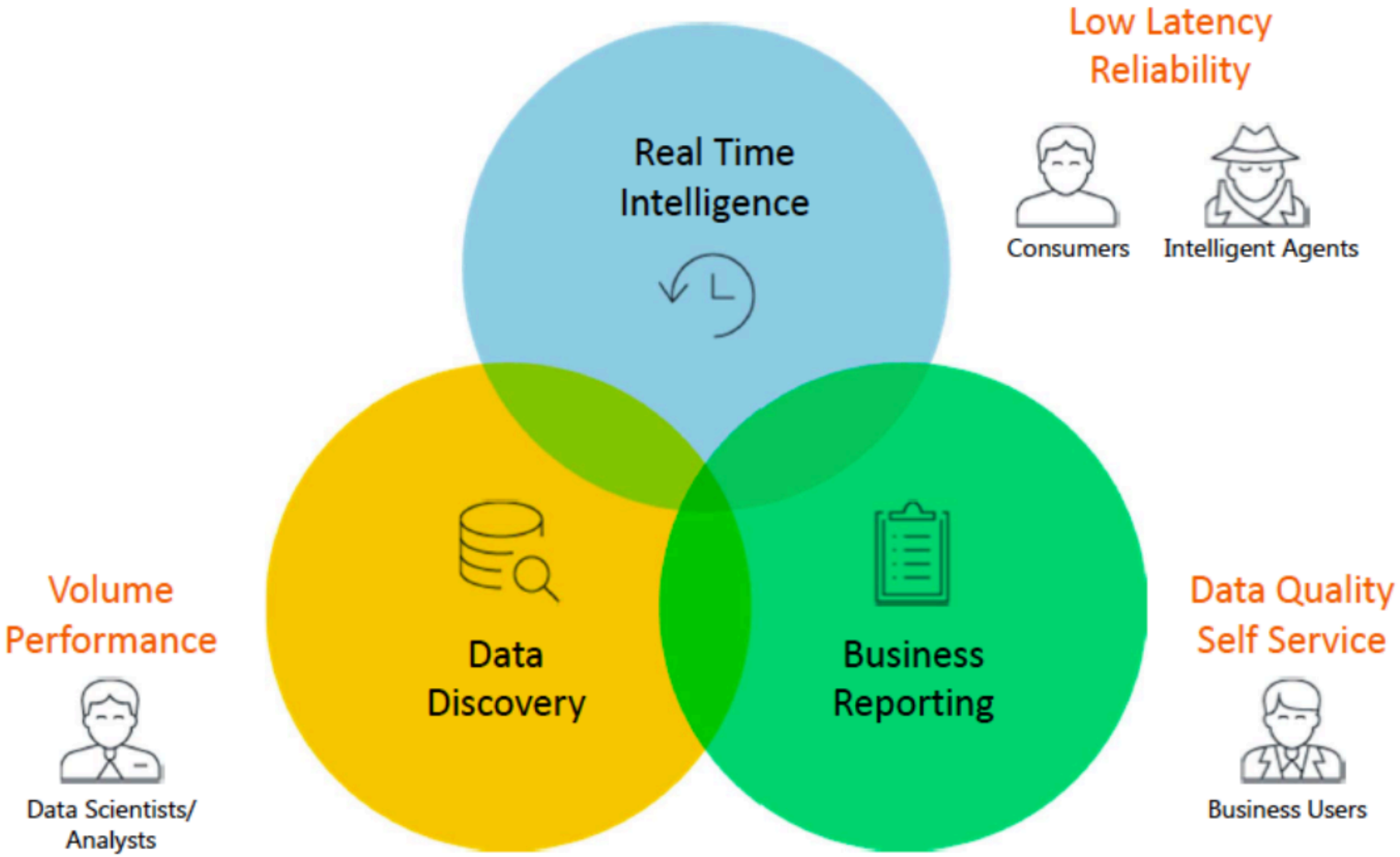
Technology



Analytics

Big Data Analytics

CHULALONGKORN BUSINESS SCHOOL
FLAGSHIP FOR LIFE



Source: SoftServe