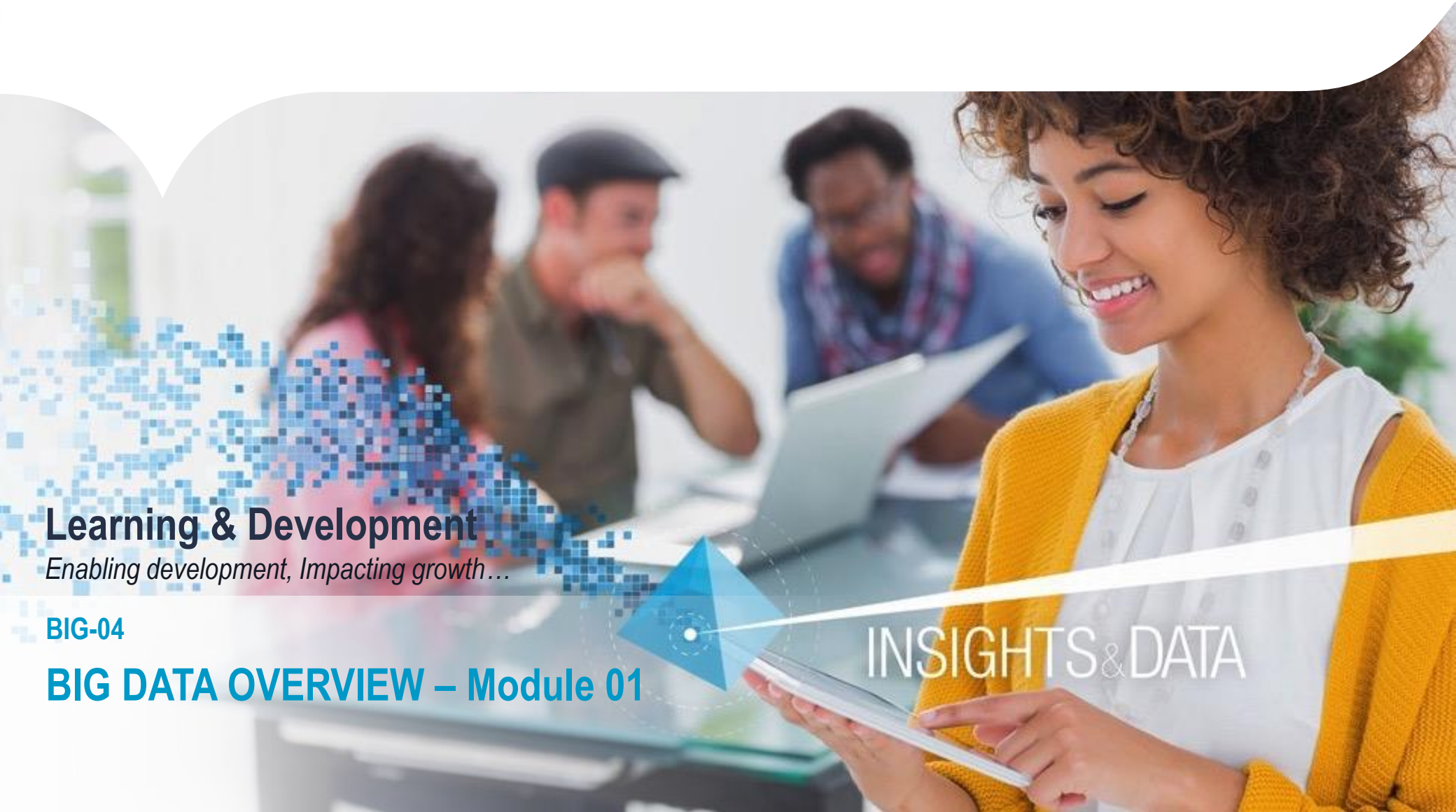
Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

People matter, results count.



Learning & Development

Enabling development, Impacting growth…

BIG-04

BIG DATA OVERVIEW – Module 01

Module Outline

- What's Big Data?
- Big Data: 3V's
- Explosion of Data
- What's driving Big Data
- Applications for Big Data Analytics
- Big Data Use Cases
- Benefits of Big Data
- Q & A

What's Big Data?

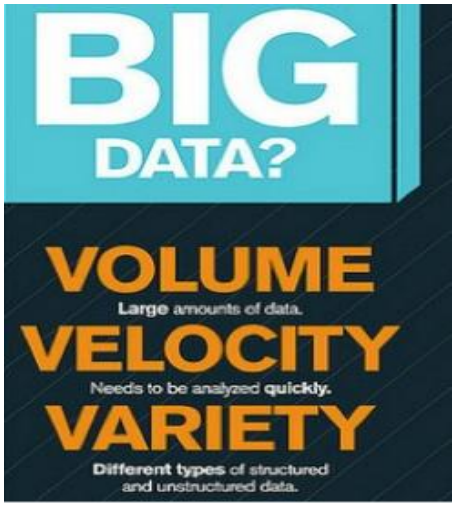
No single definition; here is from Wikipedia:

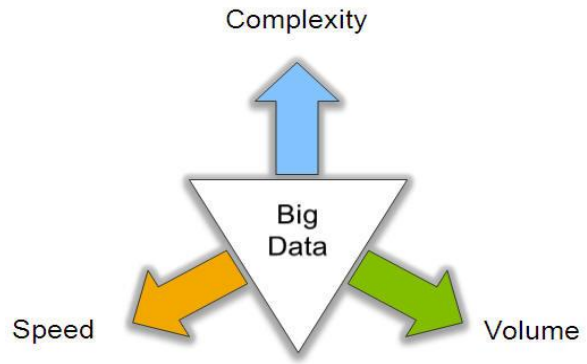
- **Big data** is the term for a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications.

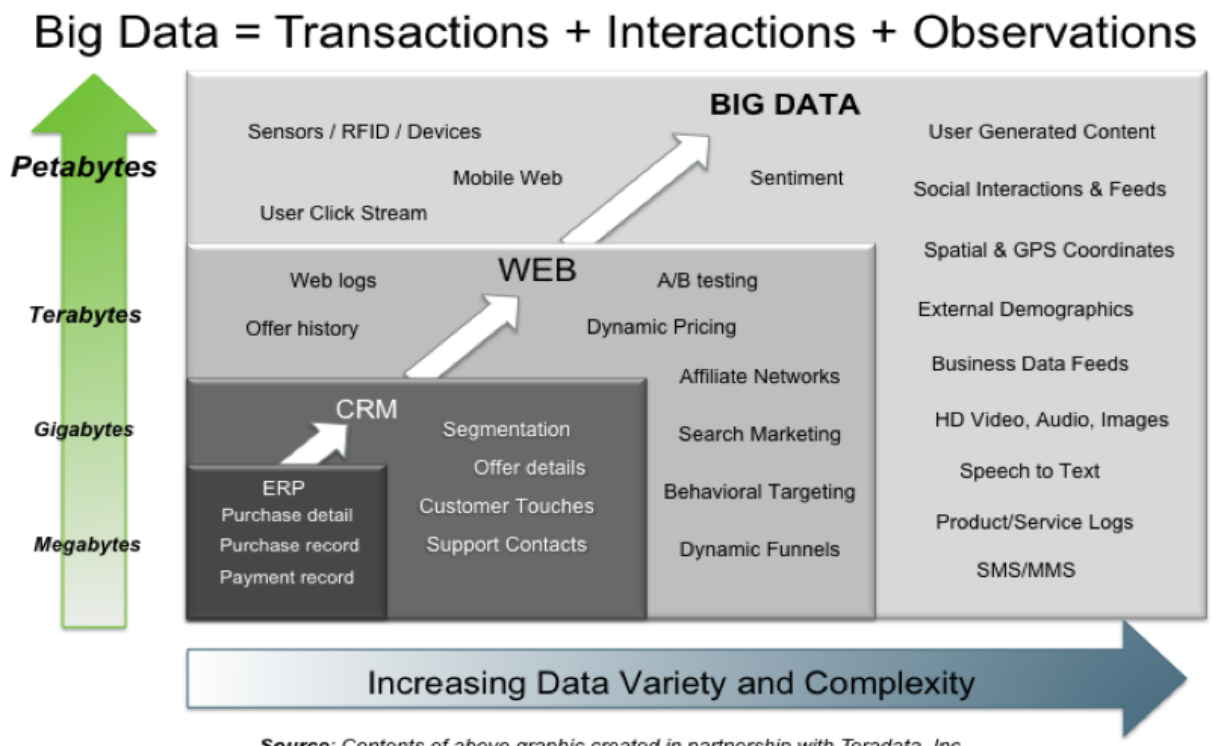
- The challenges include capture, curation, storage, search, sharing, transfer, analysis, and visualization.

- The trend to larger data sets is due to the additional information derivable from analysis of a single large set of related data, as compared to separate smaller sets with the same total amount of data, allowing correlations to be found to "spot business trends, determine quality of research, prevent diseases, link legal citations, combat crime, and determine real-time roadway traffic conditions."

Big Data: 3V's



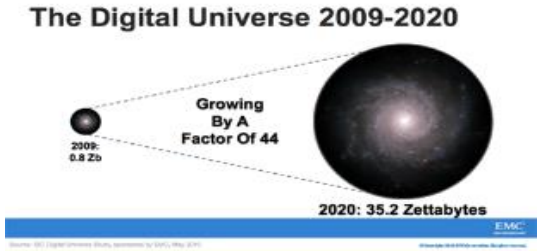


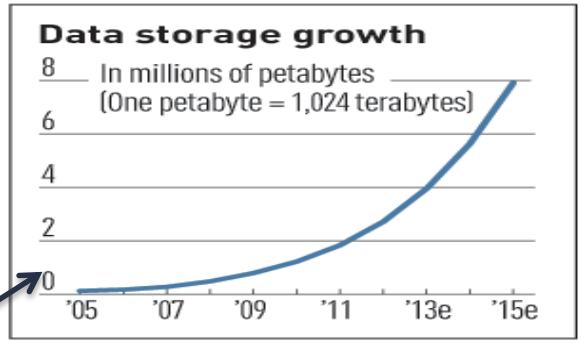


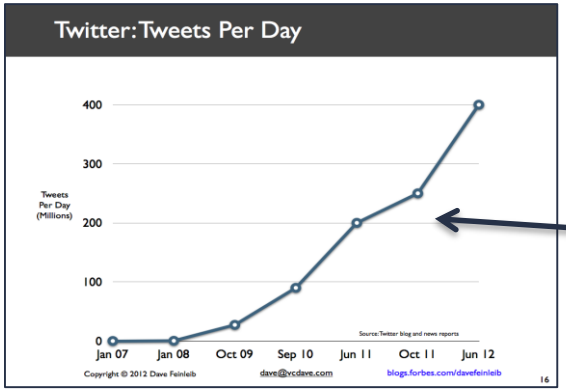
***Source**: Contents of above graphic created in partnership with Teradata, Inc.*

Volume (Scale)

- **Data Volume**
  - 44x increase from 2009 2020
  - From 0.8 zettabytes to 35zb
- Data volume is increasing exponentially

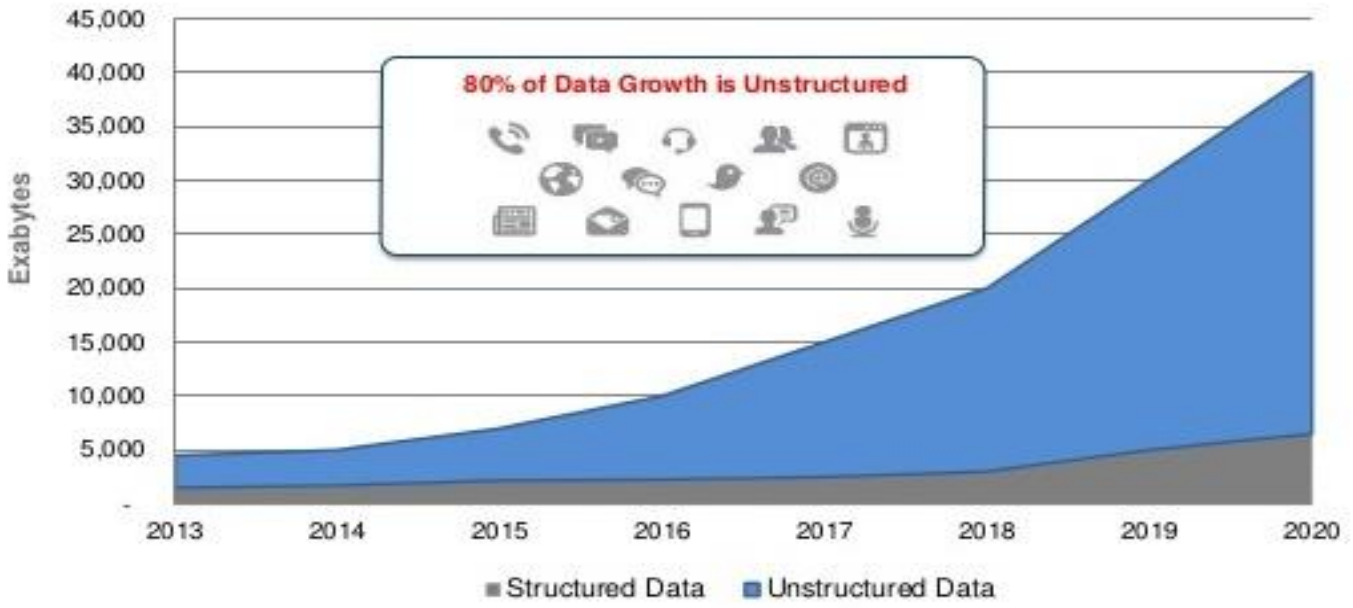






Exponential increase in collected/generated data

Volume (continued)



- By 2020, International Data Corporation predicts the number will reach 40,000 EB, or 40 Zettabytes (ZB) .
- The world's information is doubling every two years. By 2020, there will be 5,200 GB of data for every person on Earth.
- By 2020, the amount of high-value data worth analyzing will double and 60% of information delivered to decision makers will be actionable.

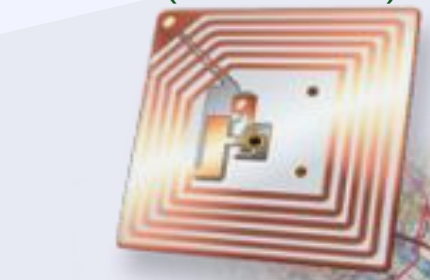
Data Sources for large volume

12+ TBs
of tweet data every day

? TBs of data every day

25+ TBs **of log data every day**

30 billion RFID tags today (1.3B in 2005)



4.6 billion camera phones world wide

100s of millions of GPS enabled devices sold annually

2+ billion people on the Web by end 2011

Volume(continued)…The Earthscope

- The Earthscope is the world's largest science project. Designed to track North America's geological evolution, this observatory records data over 3.8 million square miles, amassing 67 terabytes of data. It analyzes seismic slips in the San Andreas fault, sure, but also the plume of magma underneath Yellowstone and much, much more.

- *(http://www.msnbc.msn.com/id/44363598/ns/technology_and_science-future_of_technology/#.TmetOdQ--uI)*

Velocity (Speed)

- **Velocity**: Speed at which data is generating day by day
- 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

Velocity(continued)…Real-time/Fast Data

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)

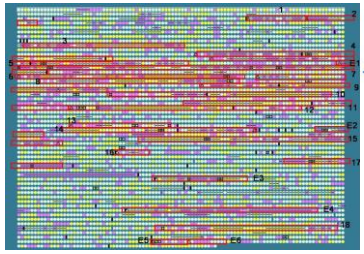
Sensor technology and networks
(measuring all kinds of data)

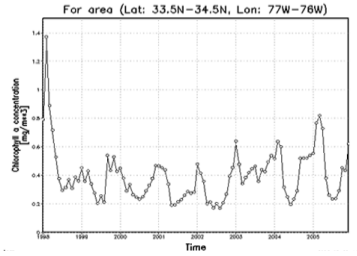
- The progress and innovation is no longer hindered by the ability to collect data
- But, by the ability to manage, analyze, summarize, visualize, and discover knowledge from the collected data in a timely manner and in a scalable fashion

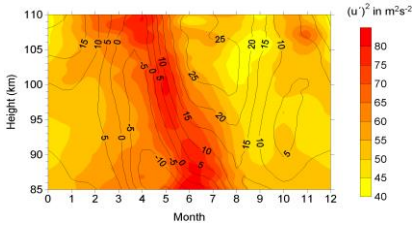
Variety (Complexity)

- Relational Data (Tables/Transaction/Legacy Data)
- Text Data
- XML Data
- Streaming Data
  - Data changing within fraction of seconds
- Audio Data
- Video Data
- Logs Data
- Graph Data
  - Social Network
- A single application may generate/collect different types of data

- Big Public Data (online, weather, finance, etc)





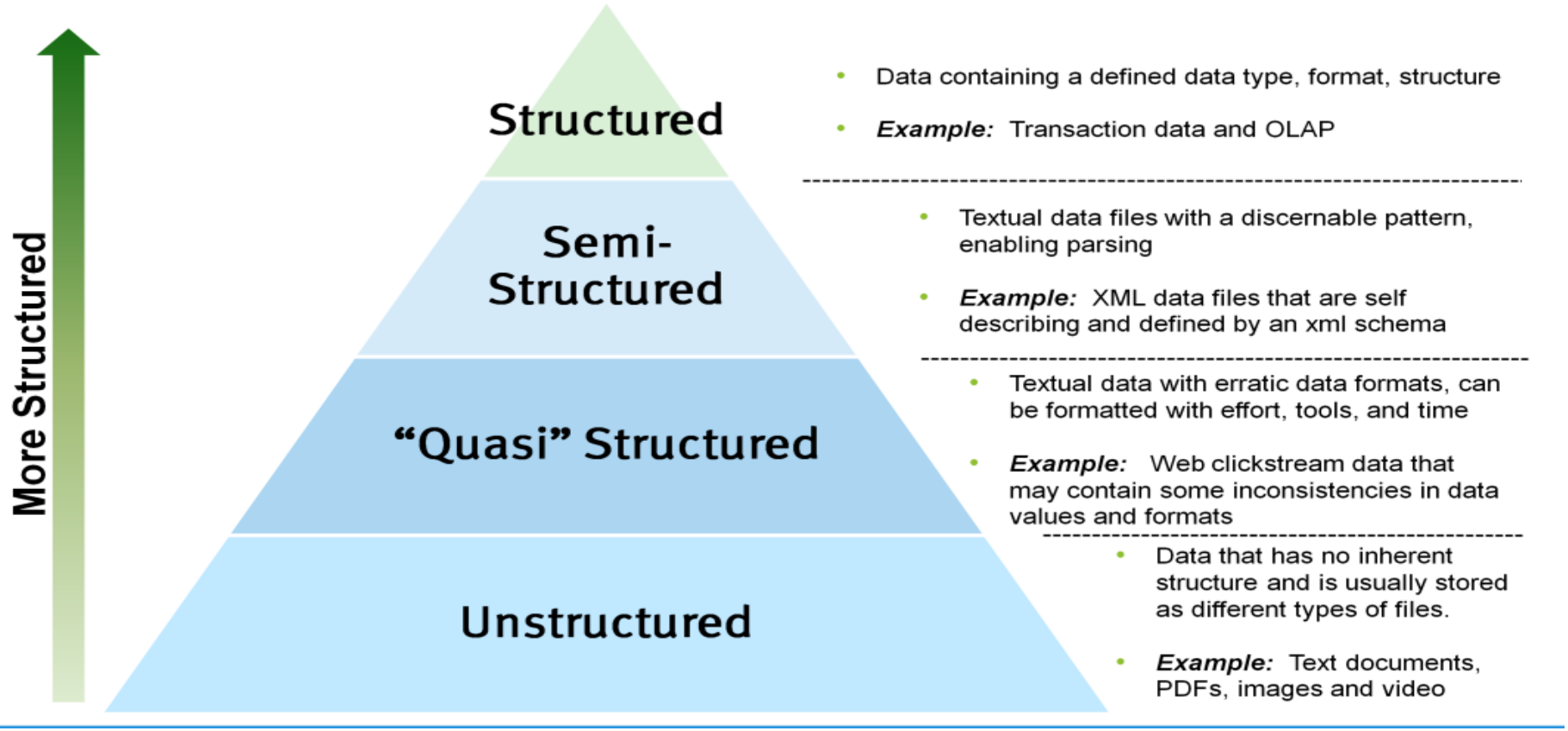


To extract knowledge➔ All these types of data need to linked together

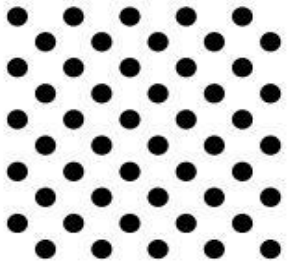
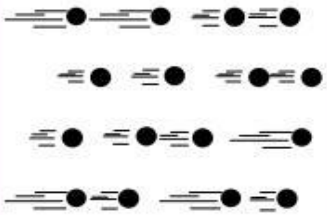
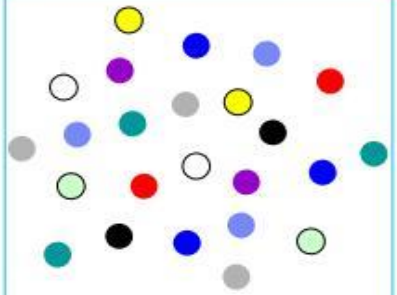
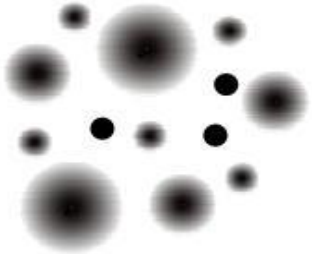
Variety(continued)…Types of Data in Big Data

Big Data Characteristics: Data Structures

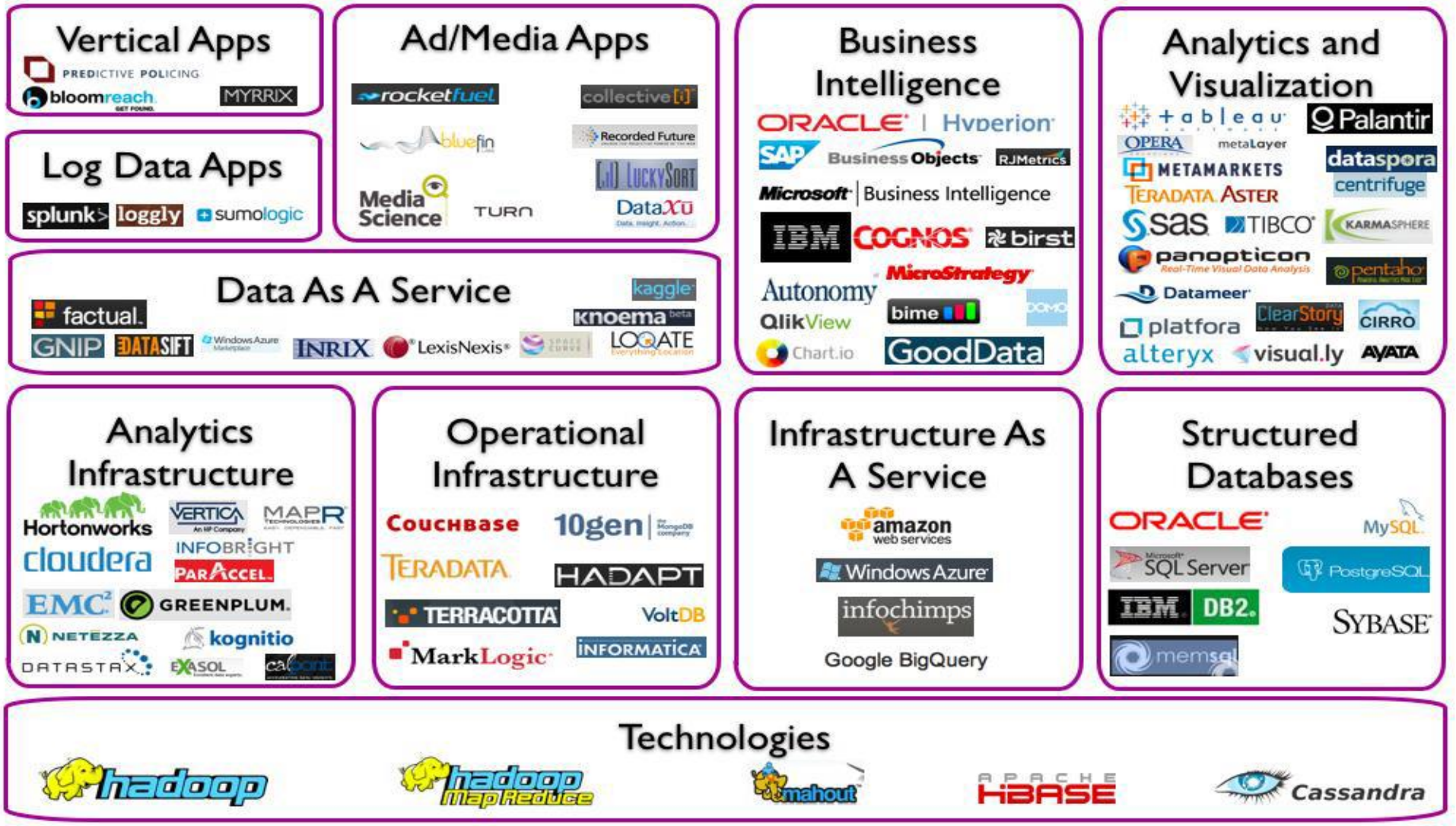
Data Growth is Increasingly Unstructured



Some Make it 4V's

Volume	Velocity	Variety	Veracity*
			
Data at Rest	**Data in Motion**	**Data in Many Forms**	**Data in Doubt**
Terabytes to exabytes of existing data to process	Streaming data, milliseconds to seconds to respond	Structured, unstructured, text, multimedia	Uncertainty due to data inconsistency & incompleteness, ambiguities, latency, deception, model approximations

The Big Data Landscape



Quiz

1. What type of data Big Data deals with?

A. Only Structured data
B. Only Unstructured data
C. All types of Data

2. Which of following characteristic of big data deals with Speed?

A. Volume
B. Variety
C. Velocity

Quiz-Answers

1. What type of data Big Data deals with?

A. Only Structure data
B. Only Unstructured data
C. All types of Data

C: All types of Data

2. Which of following charactistics of big data deals with Speed?

A. Volume
B. Variety
C. Velocity

C: Velocity

Quiz(continued)

3. Map each of the below data as structured ,semi structured or unstructured.

A. PDF
B. E-mail
C. Database table
D. XML

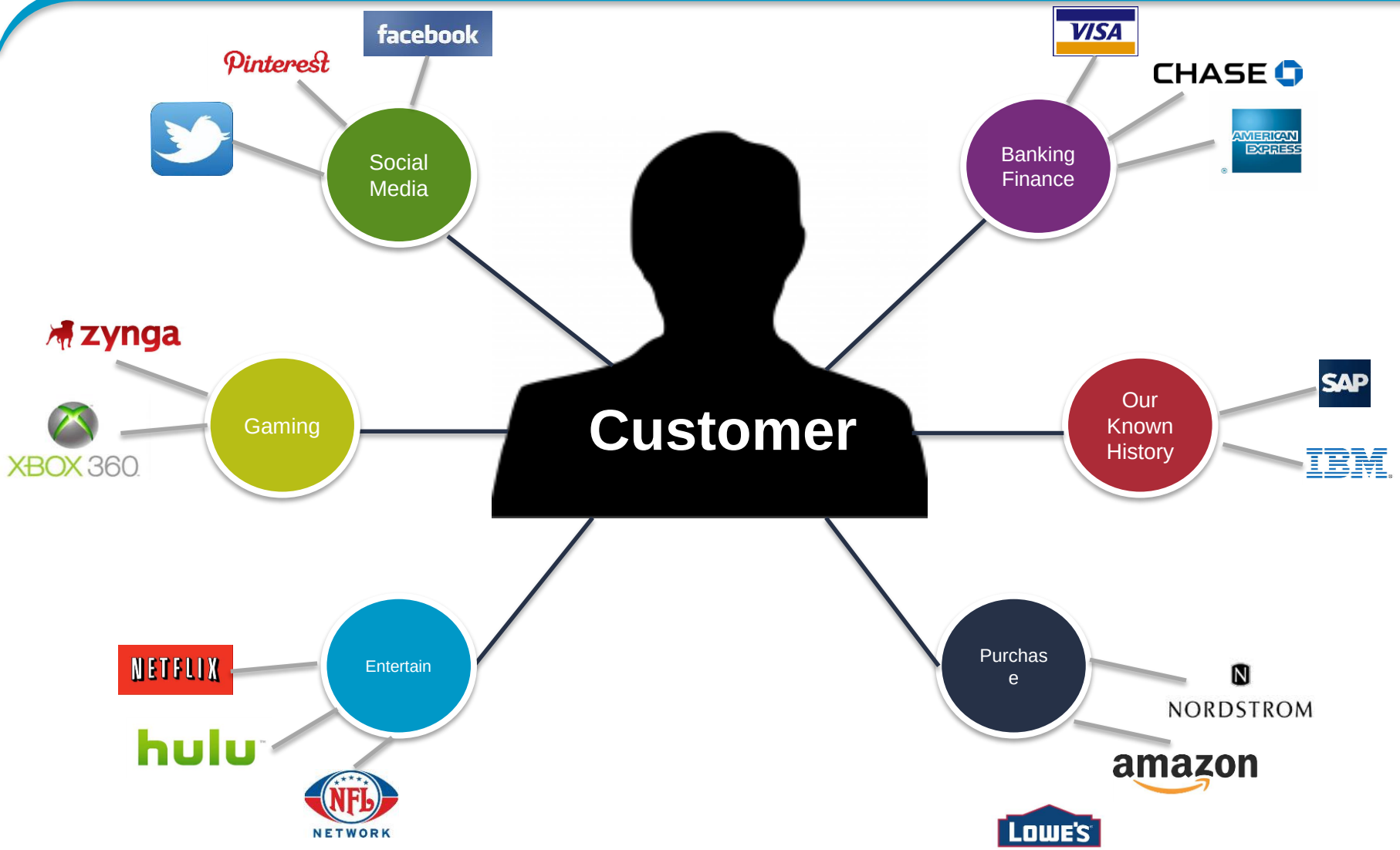
Quiz-Answer

3. Map each of the below data as structured ,semi structured or unstructured.

A. PDF
B. E-mail
C. Database tables
D. XML

PDF→ Unstructured
E-mail → Unstructured
Database tables → **Structured**
XML File → **Semi structured**

Explosion of Data- A Single View to the Customer



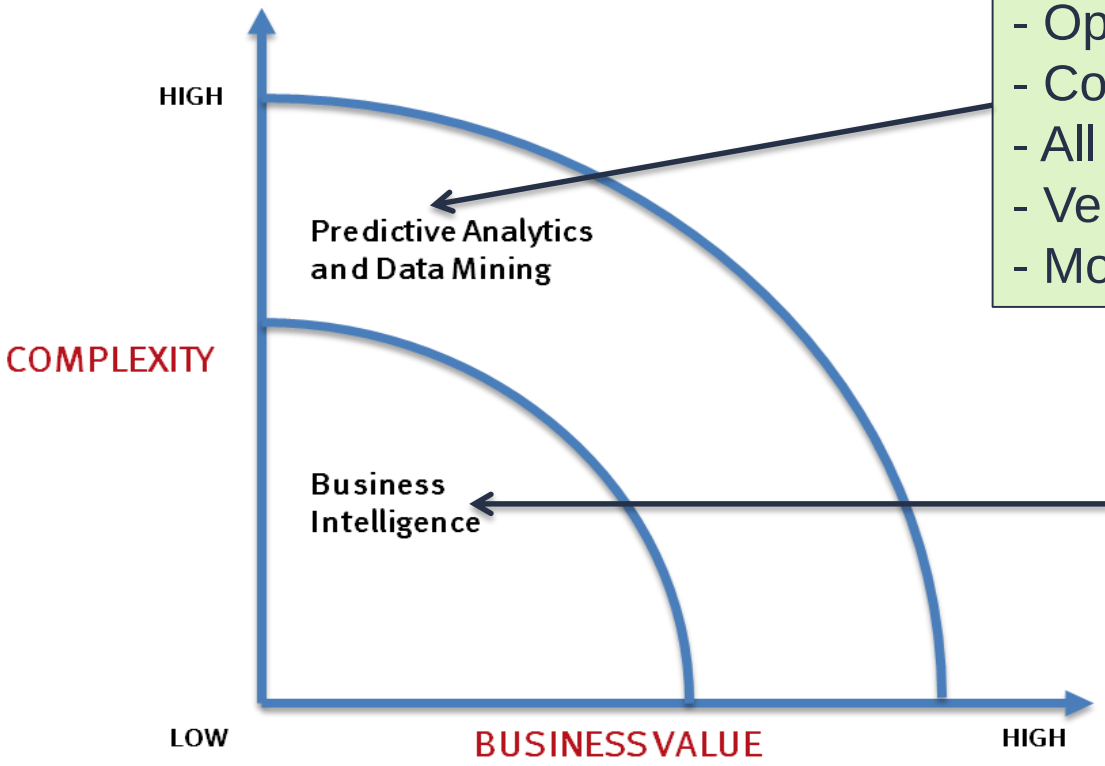
Explosion of Data(continued)- The Model Has Changed…

- **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

What's driving Big Data



- Optimizations and predictive analytics
- Complex statistical analysis
- All types of data, and many sources
- Very large datasets
- More of a real-time

- Ad-hoc querying and reporting
- Data mining techniques
- Structured data, typical sources
- Small to mid-size datasets

What is really innovative behind the Big Data buzz word ?

Hadoop & Cloud : Low cost ways to store, manage and analyze massive volumes of data

No SQL : New ways to organize and analyze non-structured data

Cost

Event Processing tools : Ability to analyze and detect trends in real time streaming events (monitoring, Next Best Action, Fraud...)

In-memory technologies : A new way to guarantee response time even for very complex calculations

Speed

Explosion of Analytics usage : R is open source, High performance statistics make them usable by every needed process

Multiple external data sources : Easier to tackle new data sources e.g. social media, traffic, GPS sensors, open data.

Insights

Applications for Big Data Analytics

Smarter Healthcare

Multi-channel

Finance

Log Analysis

Homeland Security

Traffic Control

Telecom

Search Quality

Manufacturing

Trading Analytics

Fraud and Risk

Retail: Churn

Big Data Use Cases

- Telecommunications
  - Network Performance Optimization
  - Customer churn prevention
  - Call records analysis

TELECOMMUNICATIONS

BT

- Healthcare
  - Gathering patient's complete information
  - Service quality improvements
  - Personalized treatment planning

- Banking and Financial Services
  - Fraud Detection
  - Customer segmentation analysis
  - Credit risk assessment

FINANCIAL SERVICES

JP Morgan Chase

Big Data Use Cases(continued)

- Retail
  - Customer churn prevention
  - Point of sales transaction analysis
  - 360 degree customer view

- E-commerce
  - Click stream analysis
  - Recommendation engine
  - Ad targeting

- Government Sector
  - UID enrollment
  - Social welfare schemes

Benefits of Big Data

- Increase in operational efficiency
- Improved strategic direction
- Better customer service
- Enhanced customer experience
- Early identification of business opportunities
- Reduced time to market
- Compliance with regulations

Quiz

1. Name different types of data in big-data.

A. Structured and Semi-structured

B. Quasi-Structured and Unstructured

C. None of the above

D. A&B

Quiz-Answer

1. Name different types of data in big-data.

A. Structured and Semi-structured

B. Quasi-Structured and Unstructured

C. None of the above

D. A&B

D: A & B

Q & A

Capgemini

CONSULTING.TECHNOLOGY.OUTSOURCING

People matter, results count.

About Capgemini

With almost 145,000 people in over 40 countries, Capgemini is one of the world's foremost providers of consulting, technology and outsourcing services. The Group reported 2014 global revenues of EUR 10.573 billion.

Together with its clients, Capgemini creates and delivers business and technology solutions that fit their needs and drive the results they want. A deeply multicultural organization, Capgemini has developed its own way of working, the Collaborative Business Experience™, and draws on Rightshore®, its worldwide delivery model.

Learn more about us at **www.capgemini.com**.

www.capgemini.com