

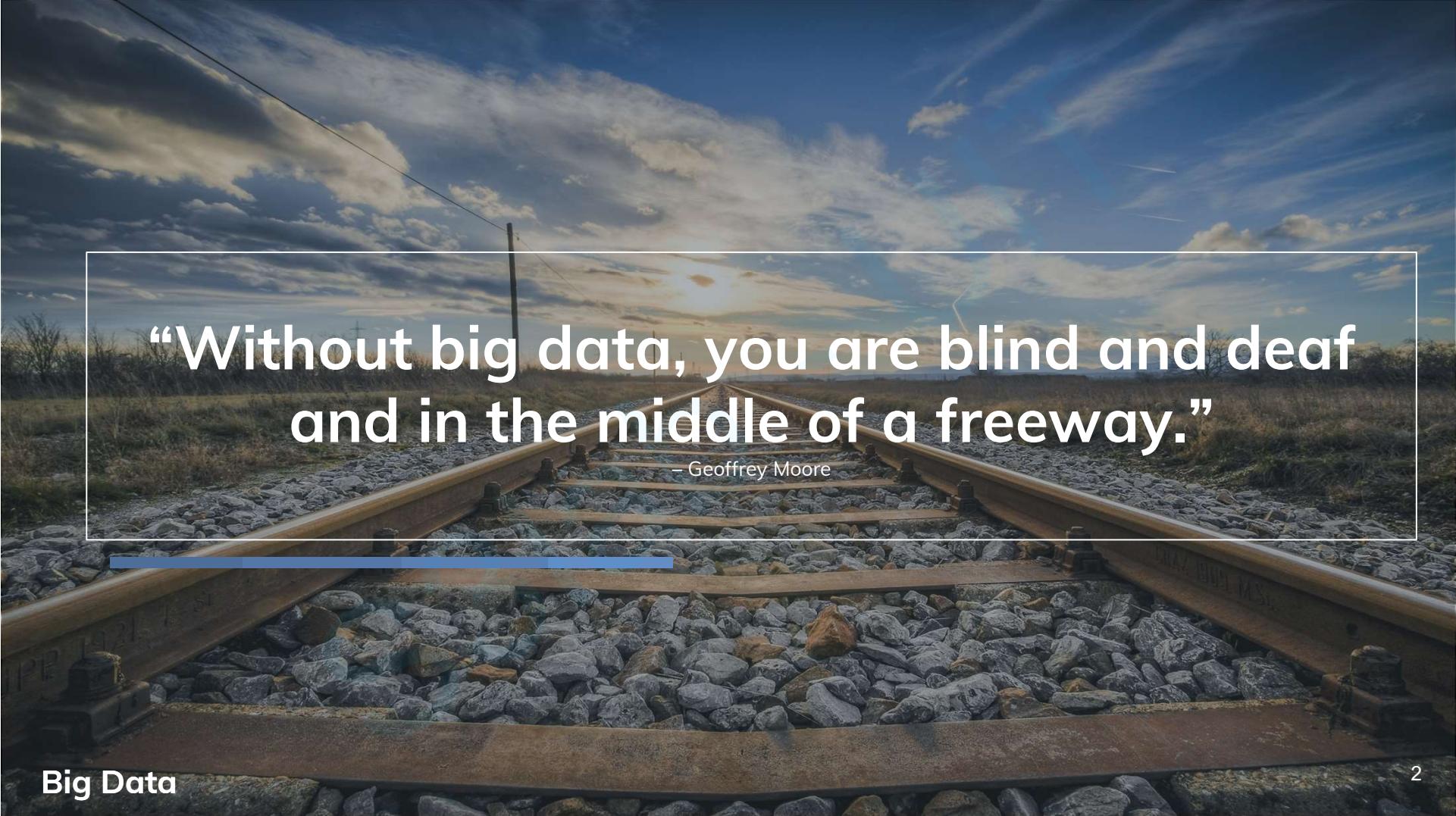


Big Data

(Understanding about Big data)

Trong-Hop Do
September 8th, 2020

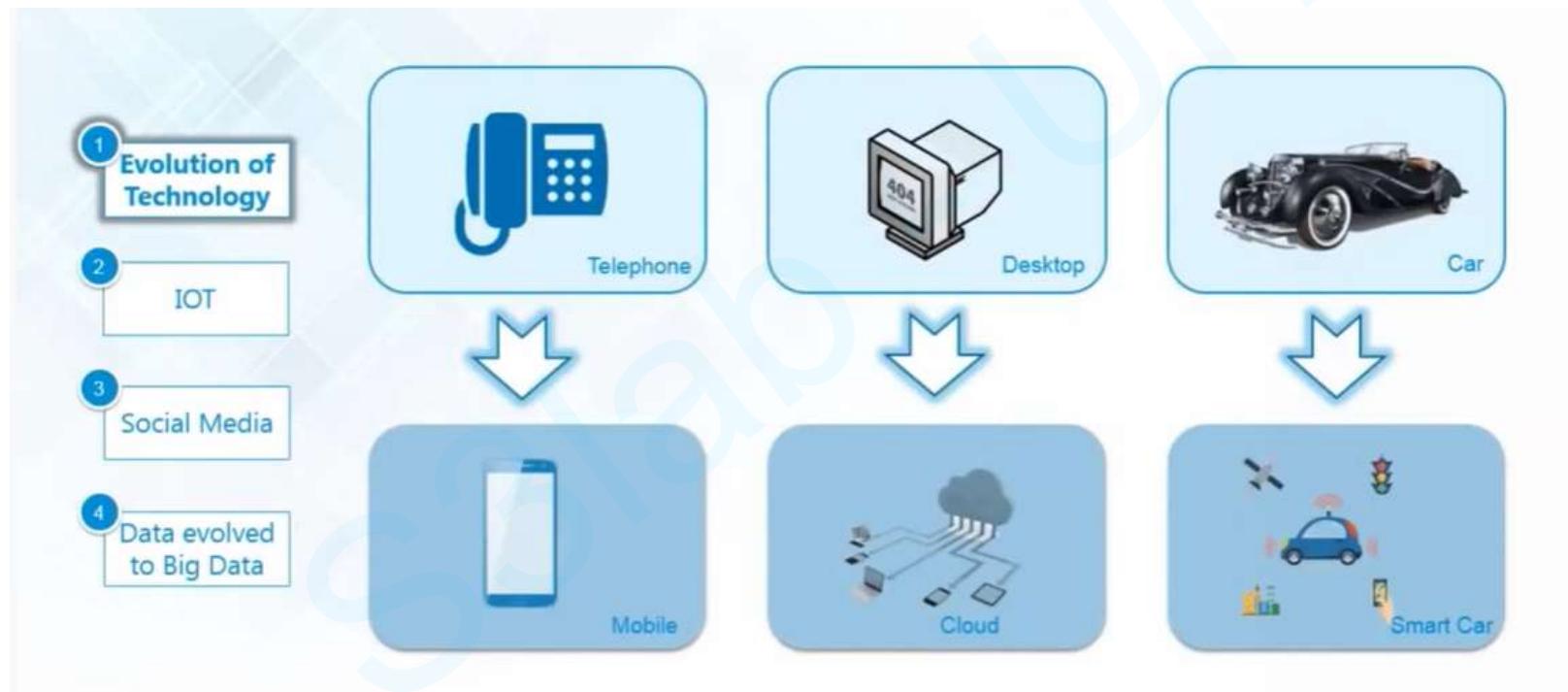
S³Lab
Smart Software System Laboratory



**“Without big data, you are blind and deaf
and in the middle of a freeway.”**

– Geoffrey Moore

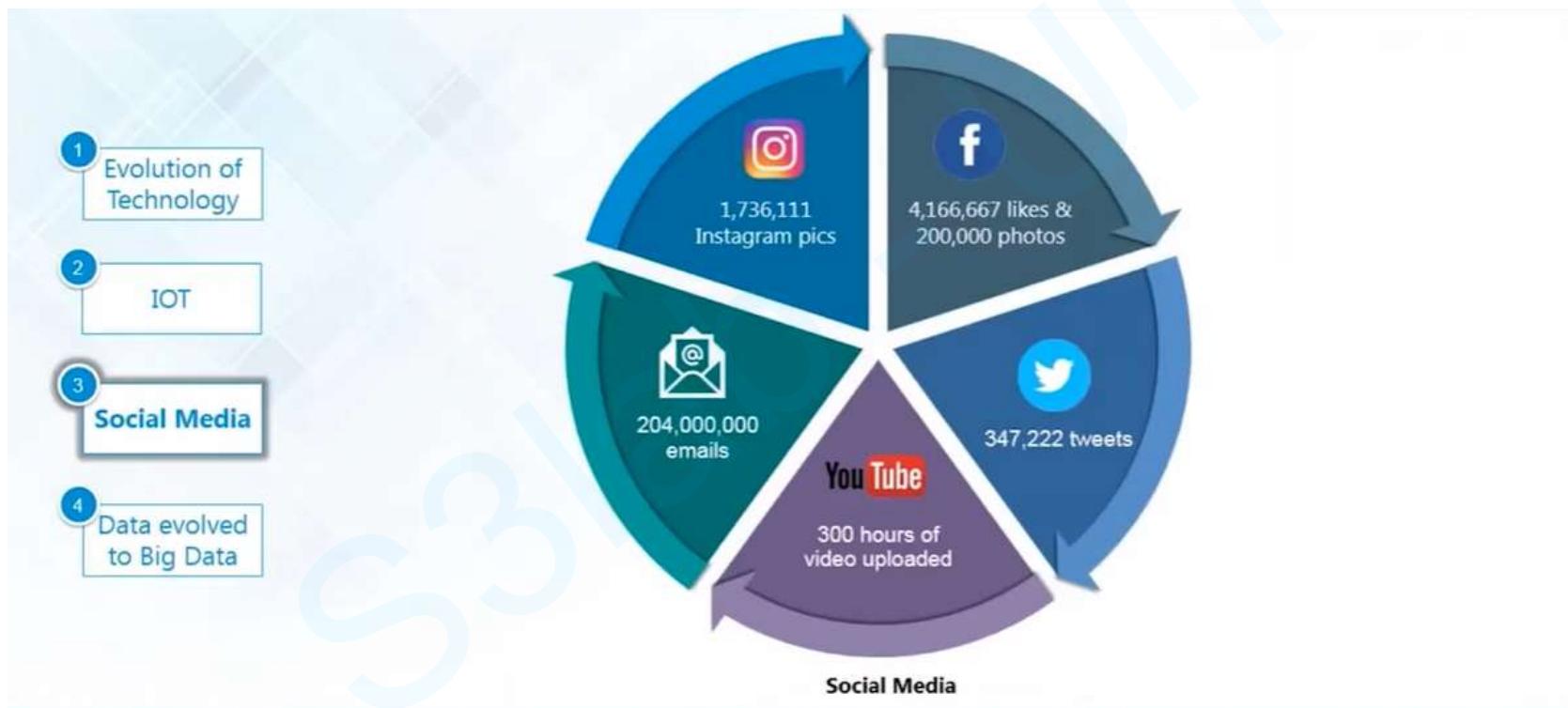
Evolution of Technology



IOT



Social media



Other factors



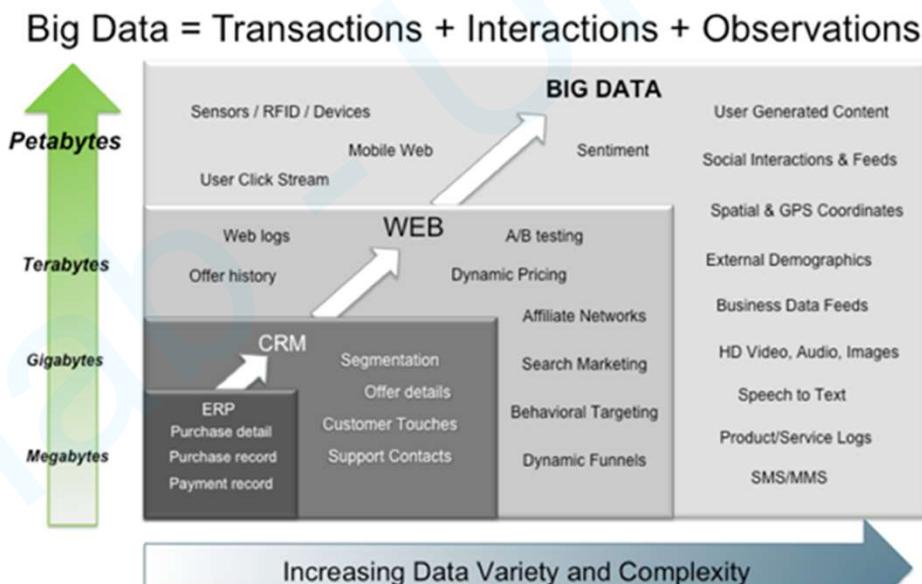
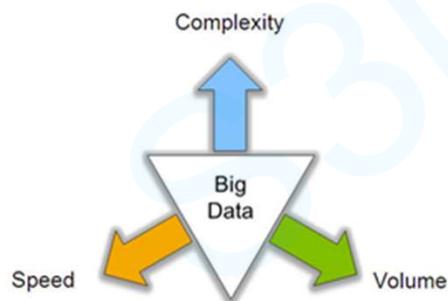
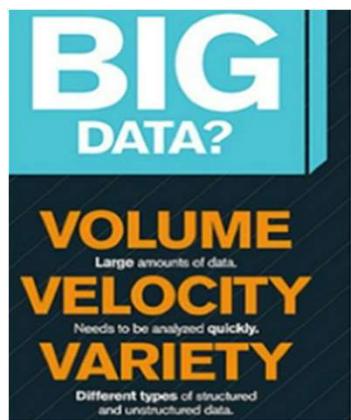


What is Big Data

- 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.
- Challenges: Capture, Curation, Storage, Search, Sharing, Transfer, Analysis, and Visualization.



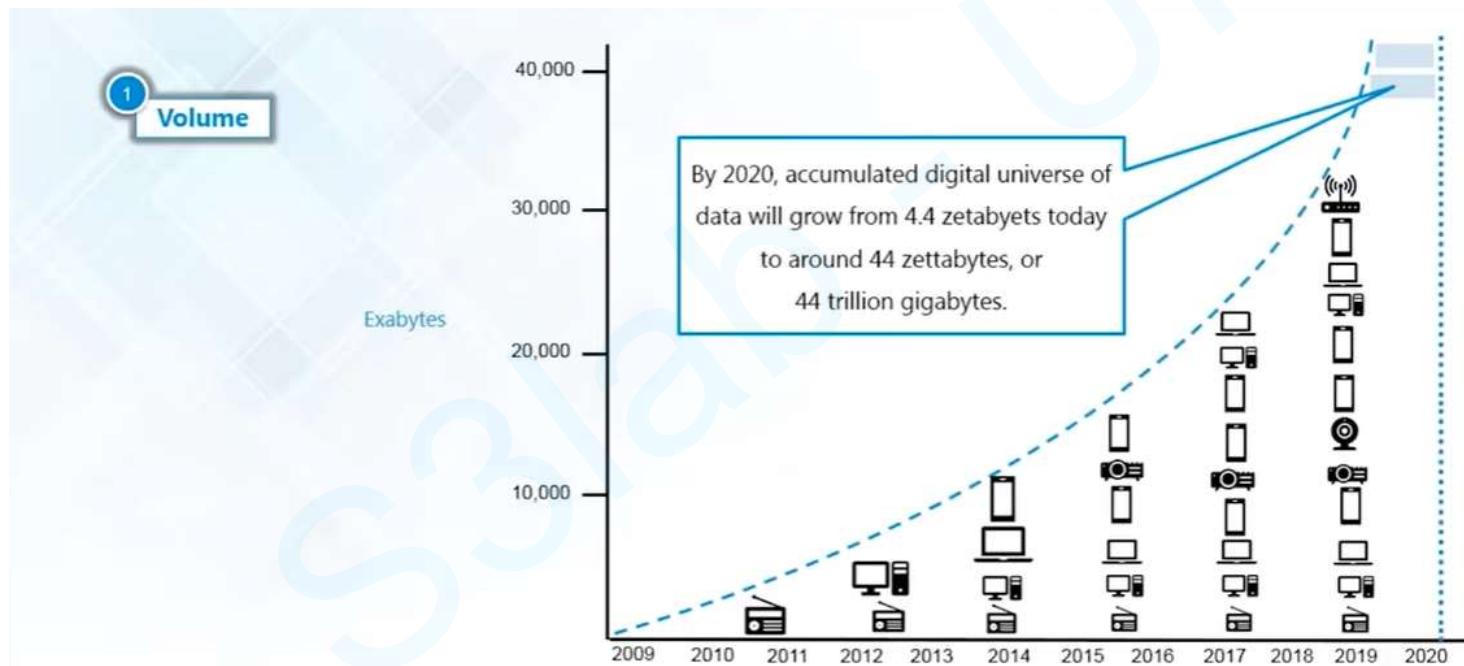
Big Data: 3V's





Big Data: 3V's

Volume (scale)

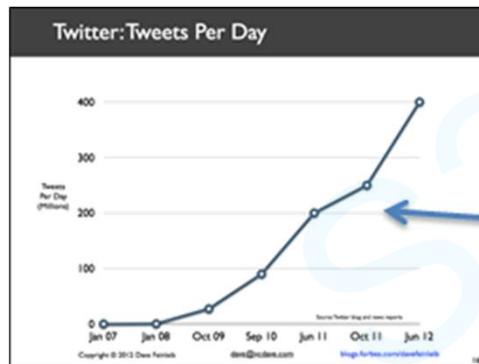
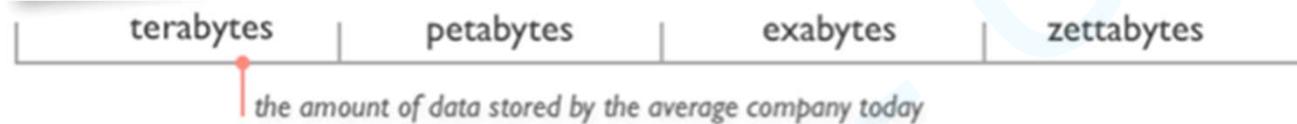


Big Data

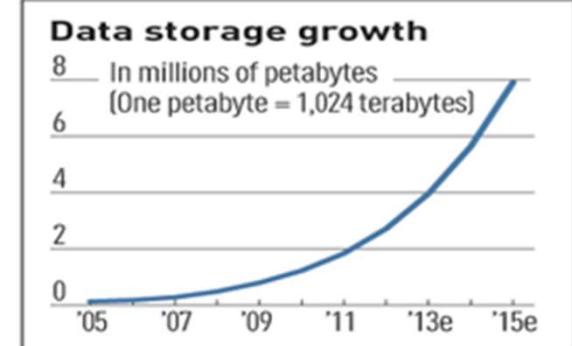


Big Data: 3V's

Volume (scale)



*Exponential increase in
collected/generated data*





Big Data: 3V's

Volume (scale)

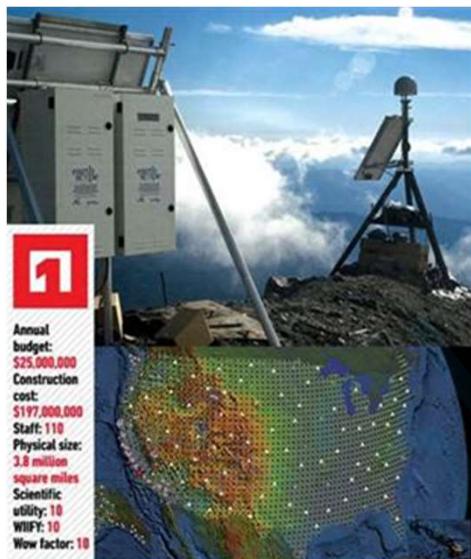


Big Data



Big Data: 3V's

Volume (scale)



Earthscope - 67 terabytes of data

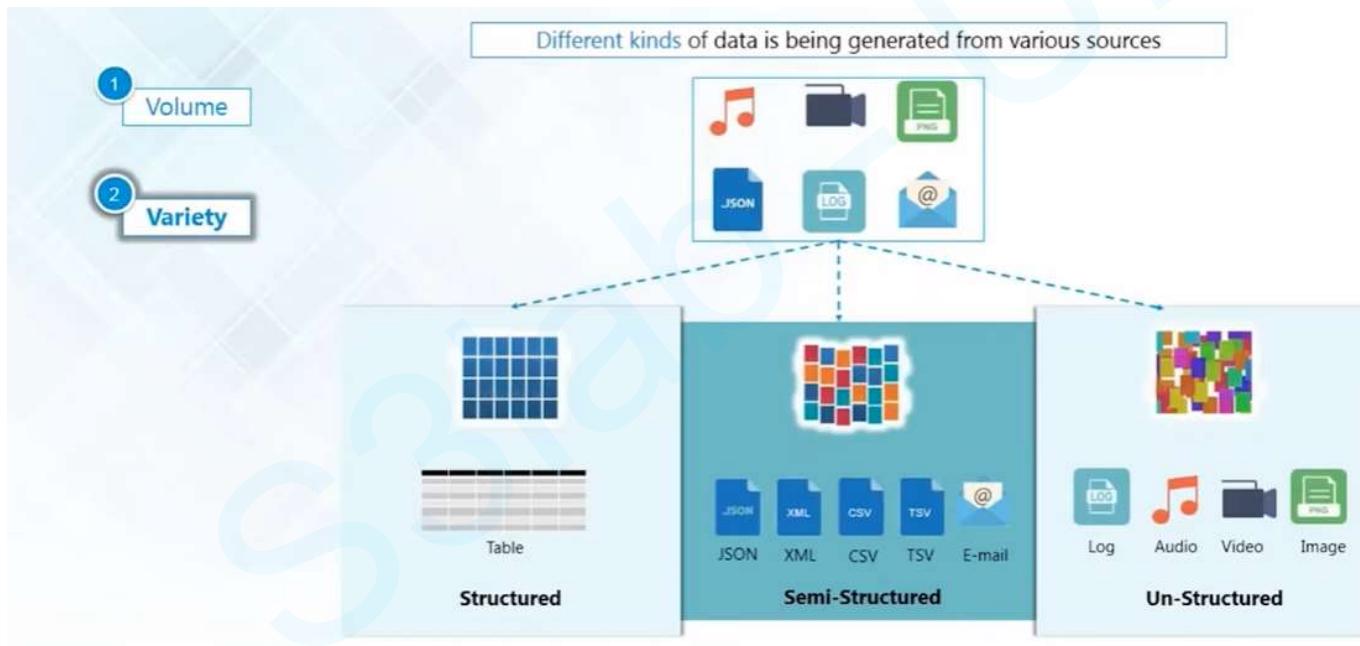


CERN's Large Hadron Collider (LHC) generates 15 PB a year



Big Data: 3V's

Variety (Complexity)





Big Data: 3V's

Variety (Complexity)

- Big data could be of three types
 - **Structured:** The data that can be stored and processed in a fixed format (fixed schema) is called as Structured Data. Ex. RDBMS
 - **Semi-Structured:** not have a formal structure of a data model, but nevertheless it has some organizational properties like tags and other markers to separate semantic elements that makes it easier to analyze. Ex. XML files or JSON documents.
 - **Unstructured:** Text Files and multimedia contents like images, audios, videos are example of unstructured data. The unstructured data is growing quicker than others, experts say that 80 percent of the data in an organization are unstructured.

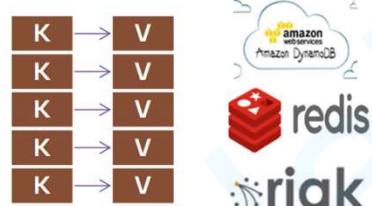


Big Data: 3V's

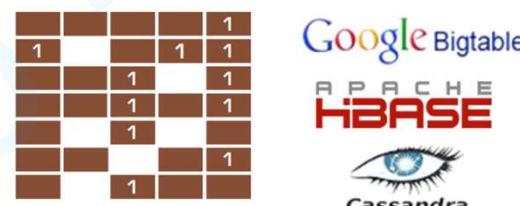
Variety (Complexity)

- Semi-Structured, NoSQL

Key-Value Stores



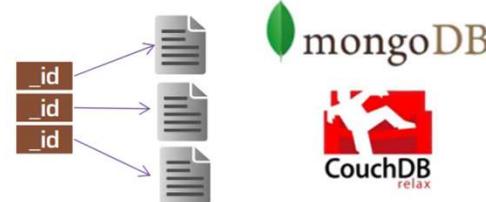
Column Stores



Graph Databases



Document Stores



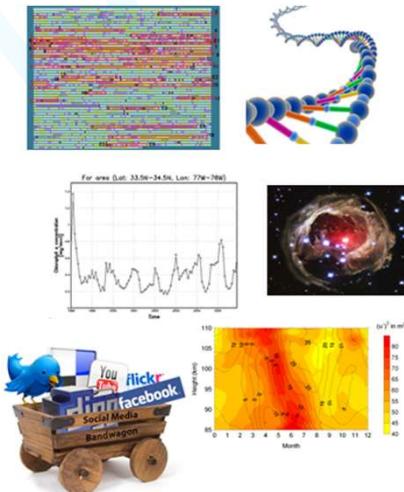


Big Data: 3V's

Variety (Complexity)

- Relational Data (Tables/Transaction/Legacy Data)
- Text Data (Web, log)
- Semi-structured Data (XML)
- Graph Data: Social network, Semantic web (RDF)...
- Streaming Data: You can only scan the data once
- A single application can be generating / collecting many types of data
- Big Public Data (online, weather, finance, etc.)

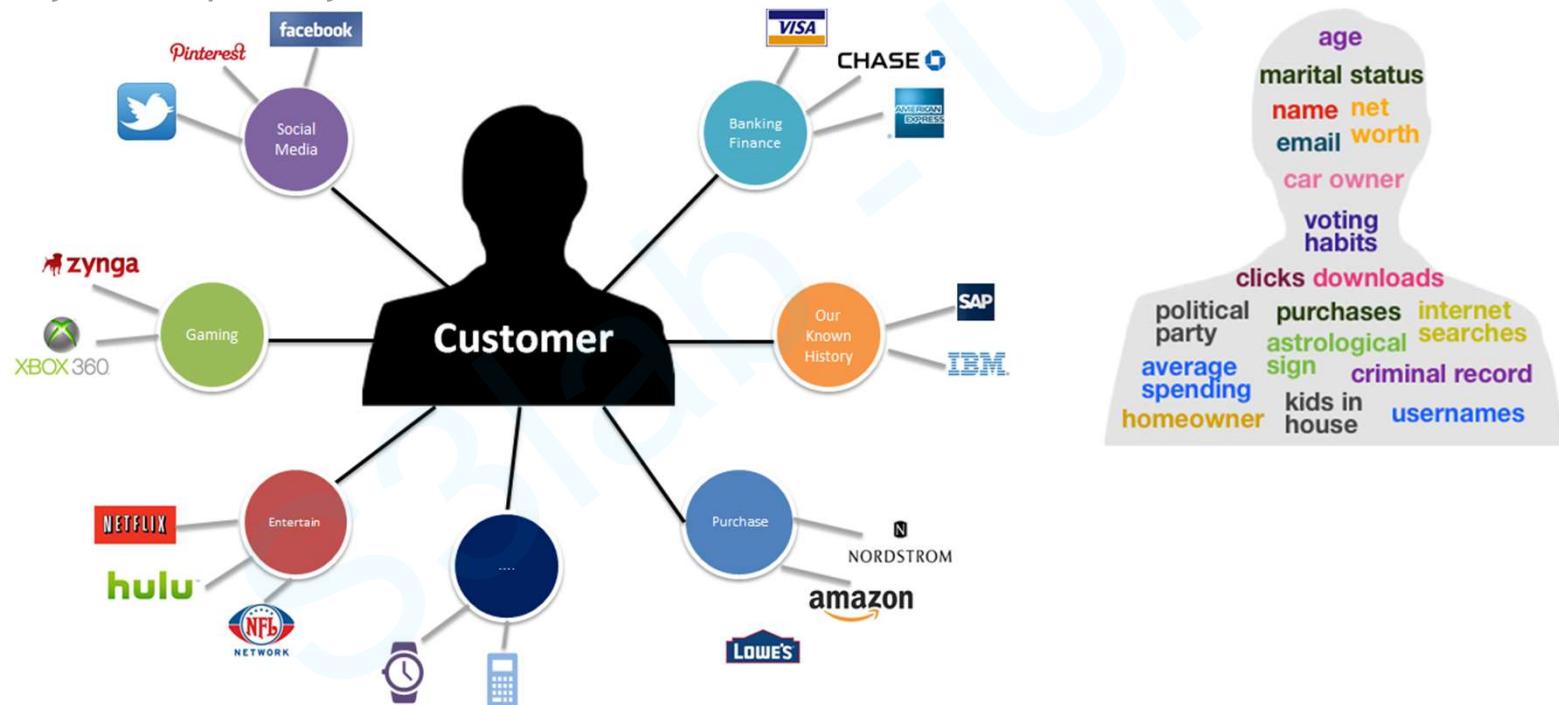
To extract knowledge ➡ all these types of data need to linked together





Big Data: 3V's

Variety (Complexity)



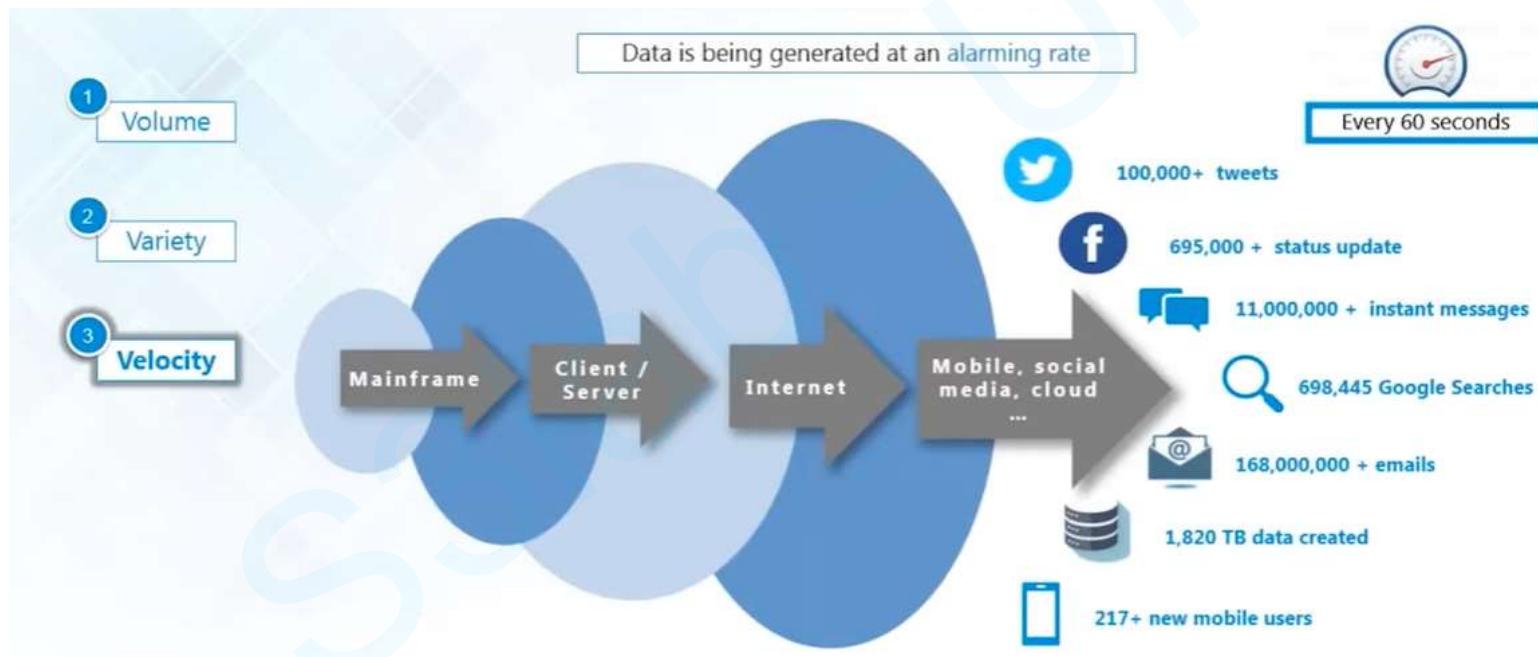
Big Data

17



Big Data: 3V's

Velocity (Speed)





Big Data: 3V's

Velocity (Speed)

- Data is begin generated fast & need to be processed fast
- Online Data Analytics
- Late decisions ➡ missing opportunities
- Examples
 - **E-Promotions:** Base 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: 3V's

Velocity (Speed)



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)



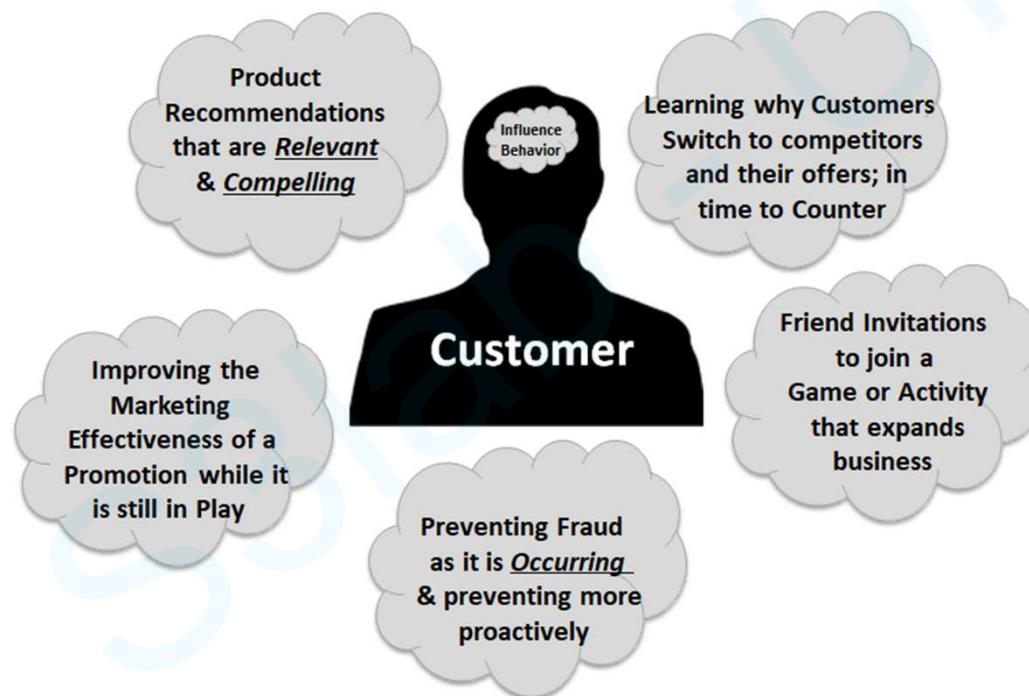
Sensor technology and networks
(measuring all kinds of data)

- 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



Big Data: 3V's

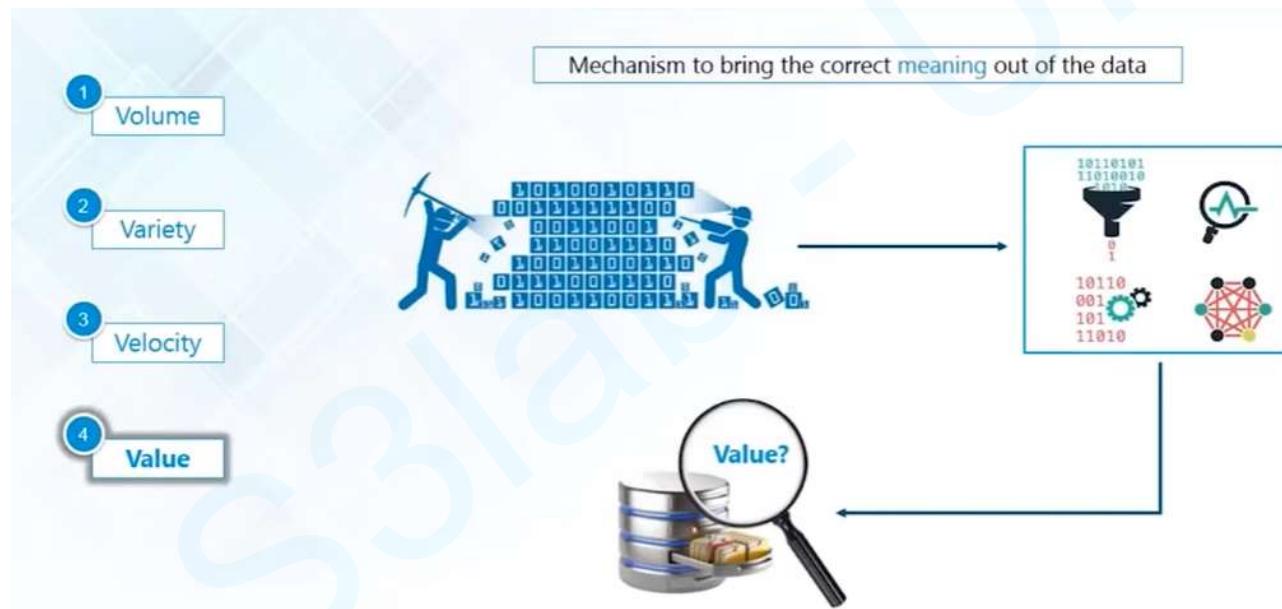
Velocity (Speed)





Big Data: 5V's

Value

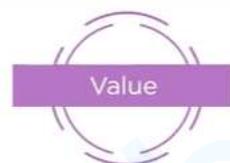




Big Data: 5V's

Value

How much data is useful and meaningful



Value refers to the ability to turn your data useful for business





Big Data: 5V's

Veracity





Big Data: 5V's

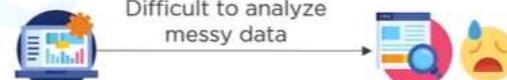
Veracity

Trustworthiness of data in terms of quality and accuracy

Veracity

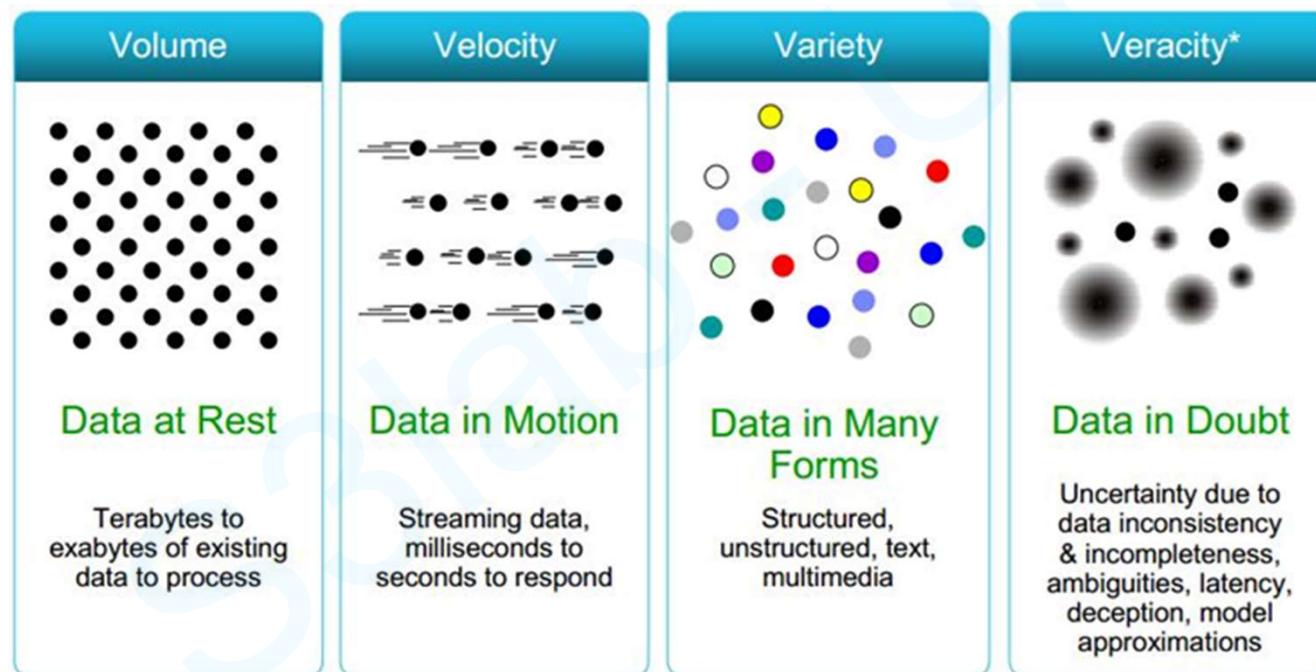
Extracting loads of data is not useful if the data is messy and poor in quality

Twitter posts with abbreviations, spelling mistakes, etc.



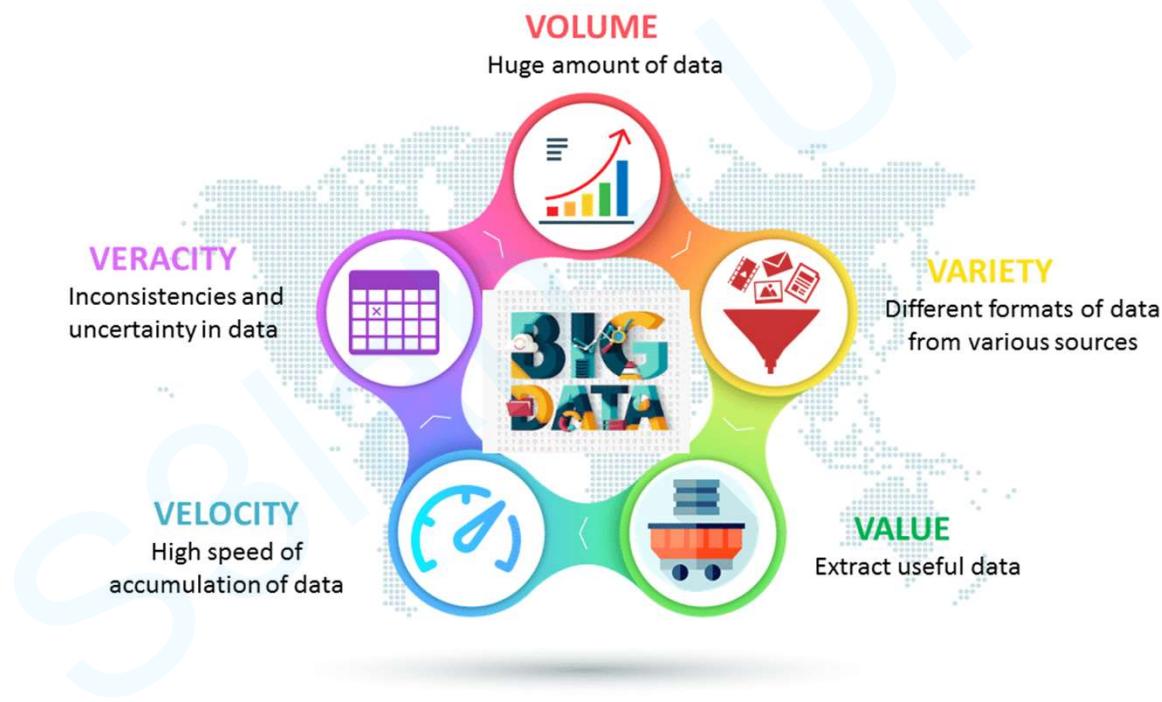


Big Data: 4V's





Big Data: 5V's



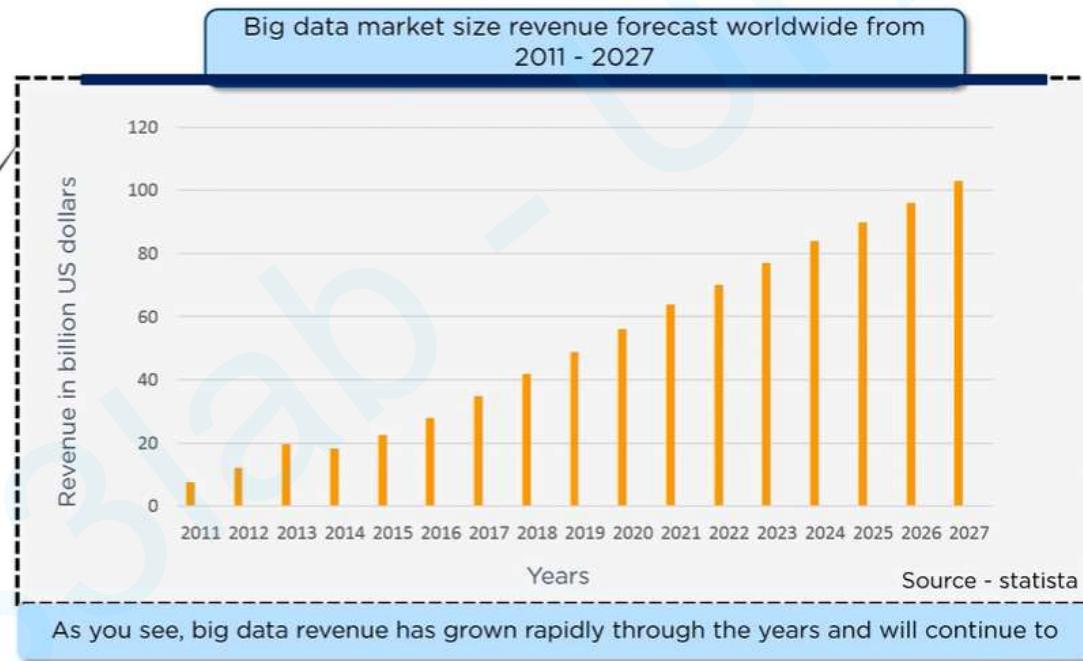


Big Data: NV's

- The above image depicts the five V's of Big Data but as and when the data keeps evolving so will the V's. I am listing five more V's which have developed gradually over time:
 - Validity: correctness of data
 - Variability: dynamic behaviour
 - Volatility: tendency to change in time
 - Vulnerability: vulnerable to breach or attacks
 - Visualization: visualizing meaningful usage of data



Big Data: Applications





Big Data: Applications





Big Data: Applications



BANKING AND SECURITIES

Challenges:

- Early warning for Securities fraud and Trade visibility.
- Card fraud detection and audit trails.
- Enterprise credit risk reporting.
- Customer data transformation and analytics.

The Securities Exchange Commission (SEC) is using big data to monitor financial market activity by using network analytics and natural language processors. This helps to catch illegal trading activity in the financial markets.

COMMUNICATIONS, MEDIA & ENTERTAINMENT

Challenges:

- Collecting, analyzing and utilizing consumer insights.
- Leveraging mobile and social media content.
- Understanding patterns of real-time, media content usage.

Wimbledon Championships leverages big data to deliver detailed sentiment analysis on the tennis matches to TV, mobile and web users in real-time.



Big Data: Applications



3

HEALTHCARE PROVIDERS

Challenges:

- Rising Medical costs.
- Unavailability/inadequate/unusable Data.

Free public health data and Google Maps have been used by the University of Florida to create visual data that allows for faster identification and efficient analysis of healthcare information, used in tracking the spread of chronic disease.



4

EDUCATION

Challenges:

• Incorporating data from varied sources.	• Untrained Staff and Institutions about Big Data	• Issues of privacy and data protection.
---	---	--

The University of Tasmania, Australia with over 26000 students has deployed a Learning and Management System that tracks, log time, time spent on different pages and the overall progress of a student over time.



Big Data: Applications

The infographic is divided into two main sections: Manufacturing & Natural Resources (Section 5) and Government (Section 6). Each section includes a challenge list and a callout bubble.

MANUFACTURING & NATURAL RESOURCES (Section 5)

Challenges:

- Increase in the volume, complexity and velocity of data due to rising demands of Natural resources.
- Large volumes of untapped data from the manufacturing industry.
- Underutilization of data prevents improved quality, energy efficiency, reliability and better profit margins.

Callout Bubble: Enhancement in Supply chain capabilities from big data being used to increase productivity

GOVERNMENT (Section 6)

Challenges:

- Integration and
- Interoperability of big data.

The Food and Drug Administration (FDA) is using big data to detect and study patterns of food-related illnesses and diseases, allowing for faster response to treatments.



Big Data: Applications

INSURANCE

Challenges:

- Lack of personalized services, pricing, targeted services to new market segments.
- Underutilization of data gathered by loss adjusters.
- Hunger for better insight.

Opportunities:

- Customer insights for transparent and simpler products.
- Predicting customer behavior through data derived from social media, GPS-enabled devices and CCTV footage.
- Claims management, predictive analytics from big data has been used to offer faster service

RETAIL & WHOLESALE TRADE

Unutilized Data derived from customer loyalty cards, POS scanners, RFID etc.

Opportunities:

- Optimized staffing through data from shopping patterns, local events etc.
- Reduced fraud and
- Timely analysis of inventory



Big Data: Applications

9

TRANSPORTATION

Challenges:

- Data from location-based social networks and high speed data from telecoms have affected travel behavior.
- Transport demand models are still based on poorly understood new social media structures.

Some applications of big data by governments, private organizations and individuals include:

Governments use of big data:
traffic control, route planning,
intelligent transport systems,
congestion management (by
predicting traffic conditions)

Private sector use of big data
in transport:
revenue management,
technological enhancements,
logistics and for competitive
advantage (by consolidating
shipments and optimizing
freight movement)

Individual use of big data
includes:
route planning to save on fuel
and time, for travel arrange-
ments in tourism etc.

10

ENERGY & UTILITIES

Challenges:

- 60% of electric grid assets will need replacement in this decade.
- Global installed wind capacity increased by 12.4%.
- Smart meters become main-stream, while consumers want more control & insights into energy consumption.

Smart meter readers allow data to be collected almost every 15 minutes. This granular data is being used to analyze consumption of utilities better which allows for improved customer feedback and better control of utilities use.

Big Data

35



Big Data: Applications

Weather forecast



Challenge

It is very inconvenient when the weather changes suddenly. Imagine having storms, hurricanes, floods without a warning.

Solution

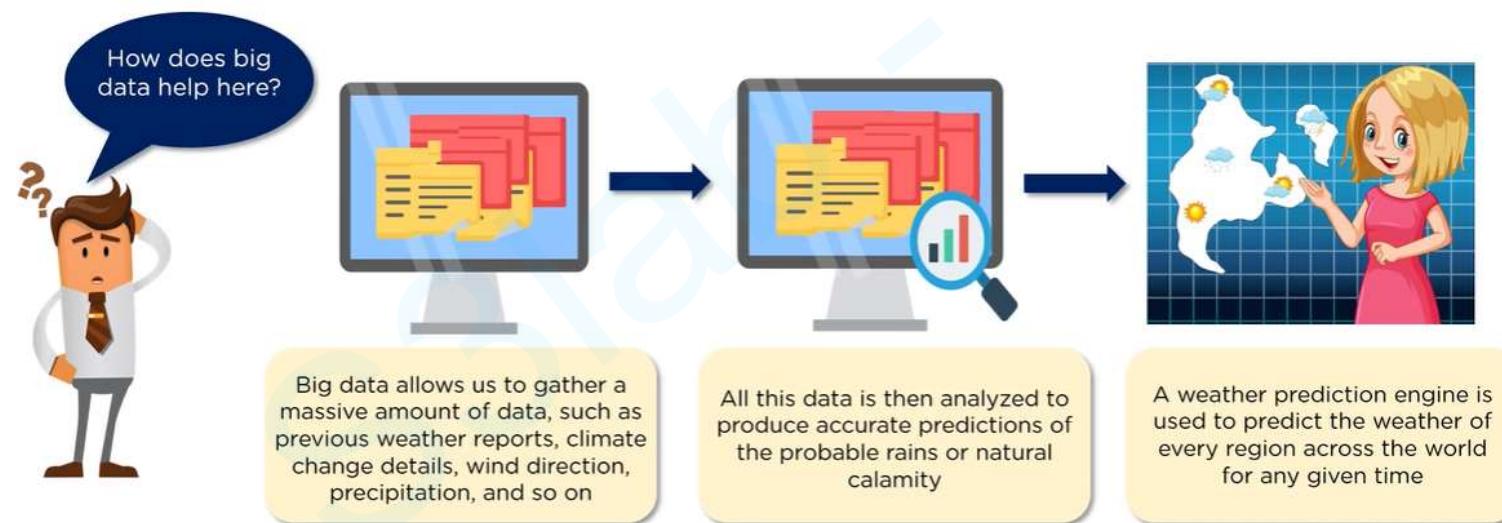


The solution is to design a system which predicts the weather conditions accurately. Big data is used in designing such a system



Big Data: Applications

Weather forecast





Big Data: Applications

Weather forecast



Predicting a landslide is very difficult with just the basic warning signs



The University of Melbourne developed an advanced tool which predicts the boundary where a landslide is likely to occur two weeks before



The predictions are made two weeks before the landslide occurs. The tool works on big data and applied mathematics



Big Data: Applications

Media and entertainment



Publishers are now able to gather more data about their visitors, which in turn enables them to serve relevant advertising

Used for targeted advertisement

Customer sentiment analysis

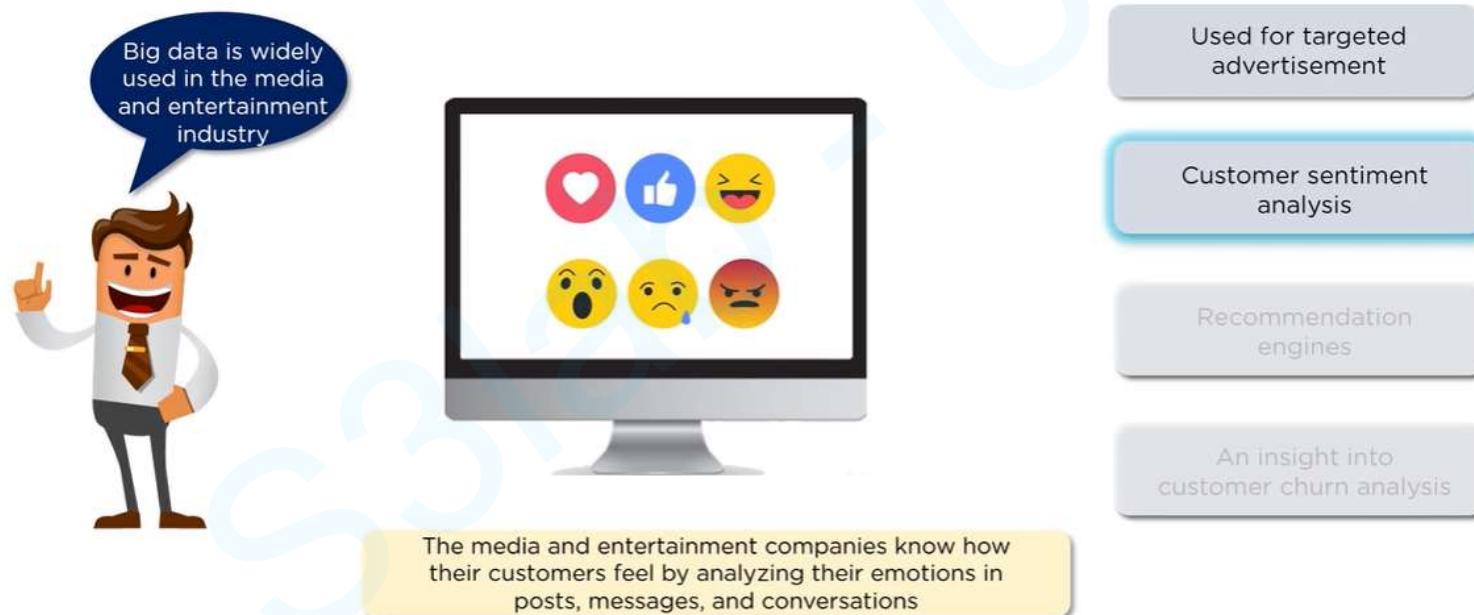
Recommendation engines

An insight into customer churn analysis



Big Data: Applications

Media and entertainment





Big Data: Applications

Media and entertainment



Recommendation engines can accurately predict what a customer would like to watch next depending on their past watch and the browse history

Used for targeted advertisement

Customer sentiment analysis

Recommendation engines

An insight into customer churn analysis



Big Data: Applications

Media and entertainment





Big Data: Applications

Media and entertainment



Did you know that Starbucks app uses big data?





Big Data: Applications

Media and entertainment



Did you know that Starbucks app uses big data?



Based on the customer's ordering history, the app suggests new products, sends personalized offers such as a birthday discount and so on



Big Data: Applications

Health care



With the available big data, medical researches are done very efficiently, and new treatments and medicines are discovered

Medical research

Personalized treatment

Cost reduction

Health of the population



Big Data: Applications

Health care





Big Data: Applications

Health care





Big Data: Applications

Health care





Big Data: Applications

Health care



United health care uses big data to detect medical fraud and identity threat



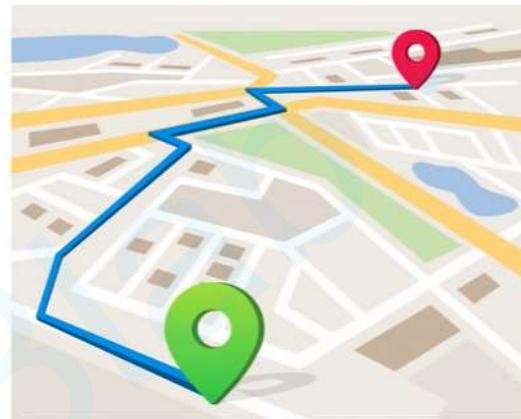
It predicts the likelihood of disease management programs to succeed, depending on how patients respond to it





Big Data: Applications

Logistic



The sensors within the vehicles analyze the fastest route to reach the destination. This dynamic routing plan will take into consideration the weather, traffic, and orders

Flexible routing

Capacity planning

Smart warehousing

Customer satisfaction



Big Data: Applications

Logistic





Big Data: Applications

Logistic



Using big data analytics and tracking sensors, warehouse robotics is improved resulting in efficient resource allocation and reduced cost

Flexible routing

Capacity planning

Smart warehousing

Customer satisfaction



Big Data: Applications

Logistic

Your package will now reach you faster!

Semantic analysis and text processing helps to analyze customer reactions which will eventually create an instant feedback loop

- Flexible routing
- Capacity planning
- Smart warehousing
- Customer satisfaction



Big Data: Applications

Logistic



UPS is one of the biggest package shipping company in the world. They use a wide variety of data every moment

They use big data to optimize routes dynamically. The system can automatically change routes in real time. By doing so, UPS can anticipate traffic jams, bad weather, and so on



Big Data: Applications

Travel and tourism

The illustration shows a cartoon character pointing upwards with a speech bubble saying "We have the best packages!". A hand holds a tablet displaying a bar chart with an upward arrow, representing revenue management. To the right is a globe with an airplane, symbolizing global travel. A callout box at the bottom states: "Looking into past occupancy rates, room tariffs, school holidays, peak seasons, the tourism industry can anticipate demand and maximize the revenue". To the right of the globe are four boxes: "Revenue management" (highlighted in blue), "Market research", "Personalized offers", and "Investment opportunities".

We have the best packages!

Revenue management

Market research

Personalized offers

Investment opportunities

Looking into past occupancy rates, room tariffs, school holidays, peak seasons, the tourism industry can anticipate demand and maximize the revenue

©Simplilearn. All rights reserved.



Big Data: Applications

Travel and tourism



Big data is used by companies to analyze information about their competitors which gives an understanding of what other hotels are offering their customers

Revenue management

Market research

Personalized offers

Investment opportunities



Big Data: Applications

Travel and tourism

We have the best packages!

Based on a tourists' past travel history and likes, they can receive personalized experiences that are focused on their needs

- Revenue management
- Market research
- Personalized offers
- Investment opportunities



Big Data: Applications

Travel and tourism



Few countries use Big Data to examine tourism flows and discover investment opportunities in their country.
Investment in such areas is profitable

Revenue management

Market research

Personalized offers

Investment opportunities



Big Data: Applications

Travel and tourism



Airbnb is one of the best online homestay networks. To help travelers around the world find the best properties, Airbnb analyzes all the big data to produce this result

It lists properties through customers preferences, keywords, and pricing. By doing so, it delivers its customers with the best stay



Big Data: Applications

Government and law enforcement





Big Data: Applications

Government and law enforcement

Lesser crime rates and more development!

JOB

By analyzing the number of students graduating every year and the number of job openings, governments can have an idea of the unemployment in the country

- Predictive policing
- Tackling unemployment
- Poverty
- Improving social and economic policies



Big Data: Applications

Government and law enforcement



Governments use big data tools to discover areas which fall under the poverty line and does the needful

Predictive policing

Tackling unemployment

Poverty

Improving social and economic policies



Big Data: Applications

Government and law enforcement





Big Data: Applications

Government and law enforcement



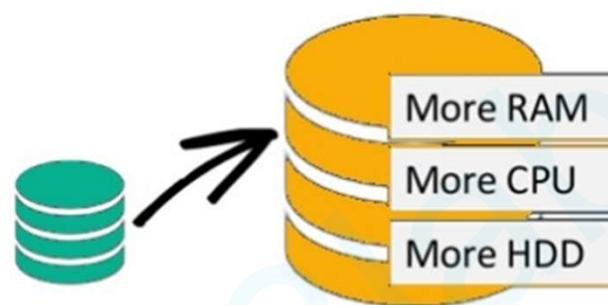
The New York Police Department uses big data analytics to protect its citizens

The department identifies crime trends, threats, and prevent crimes by analyzing data such as certain emails, fingerprints, and records from police investigations, and other public databases

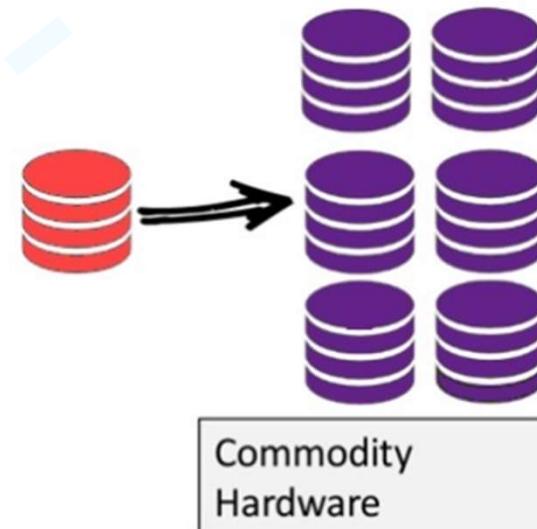


Big Data: Scale

Scale-Up (*vertical* scaling):



Scale-Out (*horizontal* scaling):





Big Data: Evolution

- The Model of Generating / Consuming Data has changed
 - Old Model: a few companies are generation data, all others are consuming data

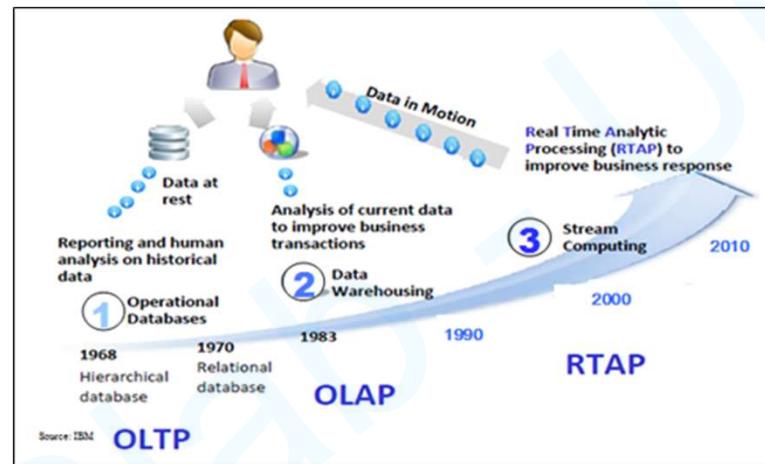


- New Model: All of us are generating data, and all of us are consuming data





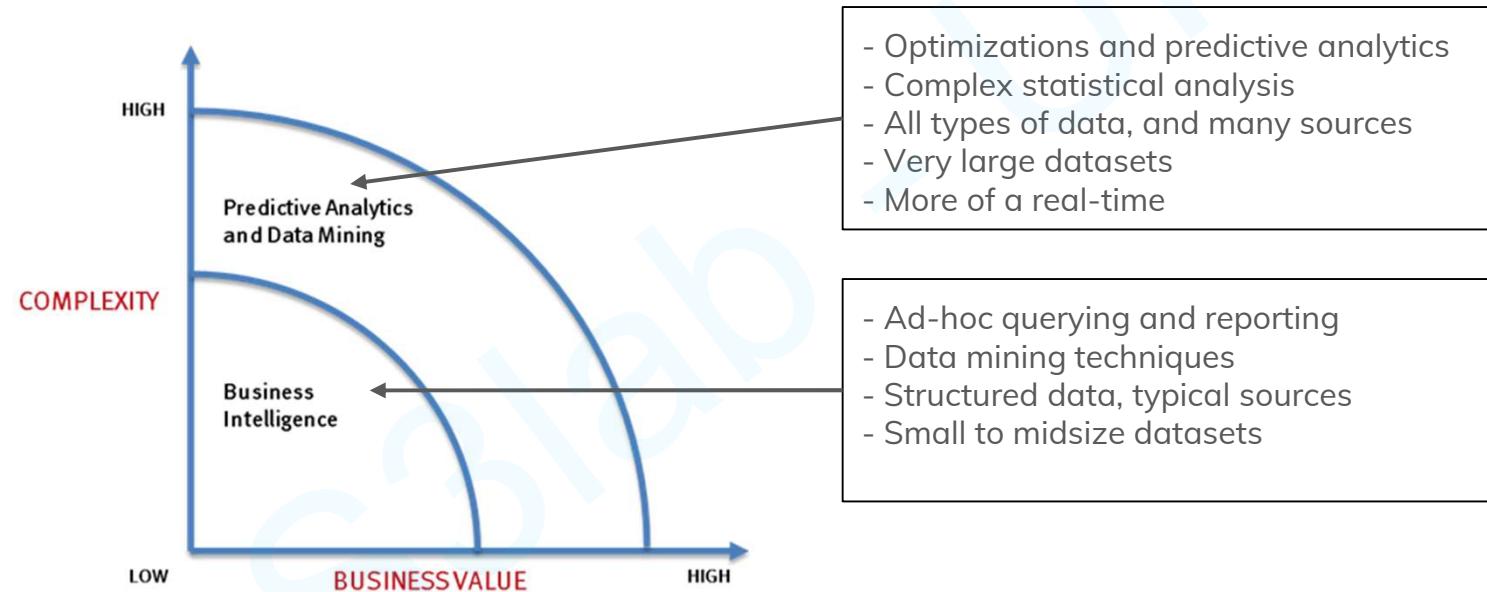
Big Data: Evolution



- **OLTP:** Online Transaction Processing (DBMSs)
- **OLAP:** Online Analytical Processing (Data Warehousing)
- **RTAP:** Real-time Analytics Processing (Big Data Architecture & Technology)

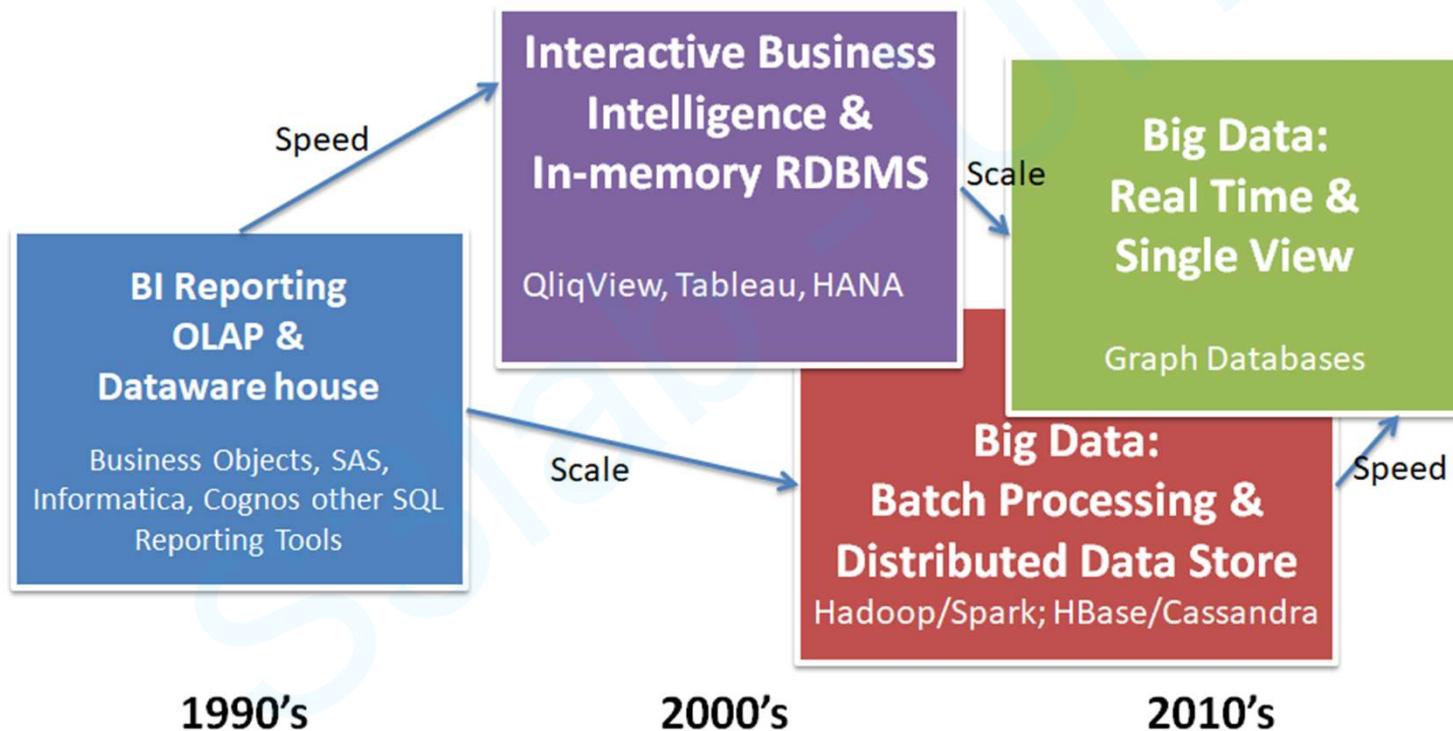


Big Data: Evolution





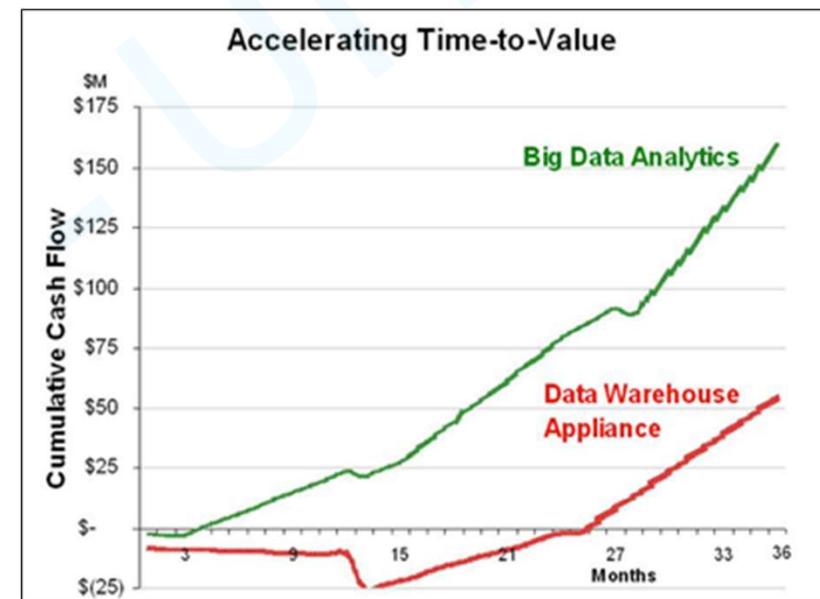
Big Data: Evolution





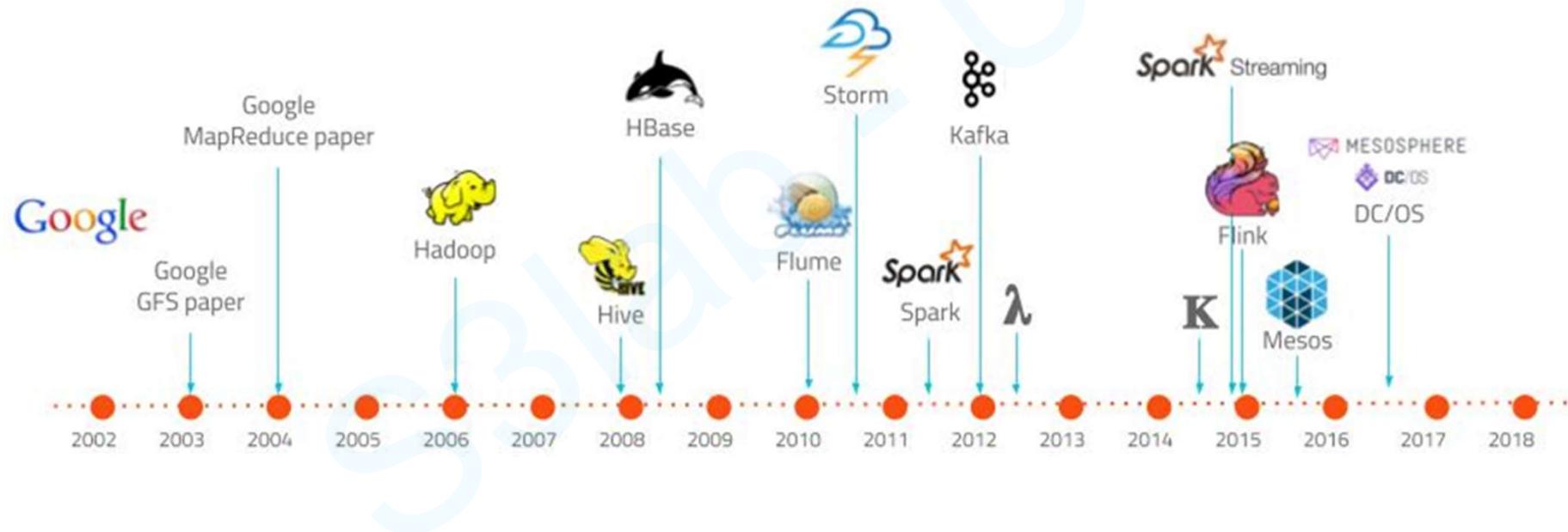
Big Data: Evolution

- Big data is more real-time in nature than traditional DW applications
- Traditional DW architectures (e.g. Exadata, Teradata) are not well-suited for big data apps
- Shared nothing, massively parallel processing, scale out architectures are well-suited for big data apps





Big Data: Evolution





Big Data: Evolution

FUN FACTS?

Initially, decoding human genome took **10 years**, now it can be achieved in **1 week**.

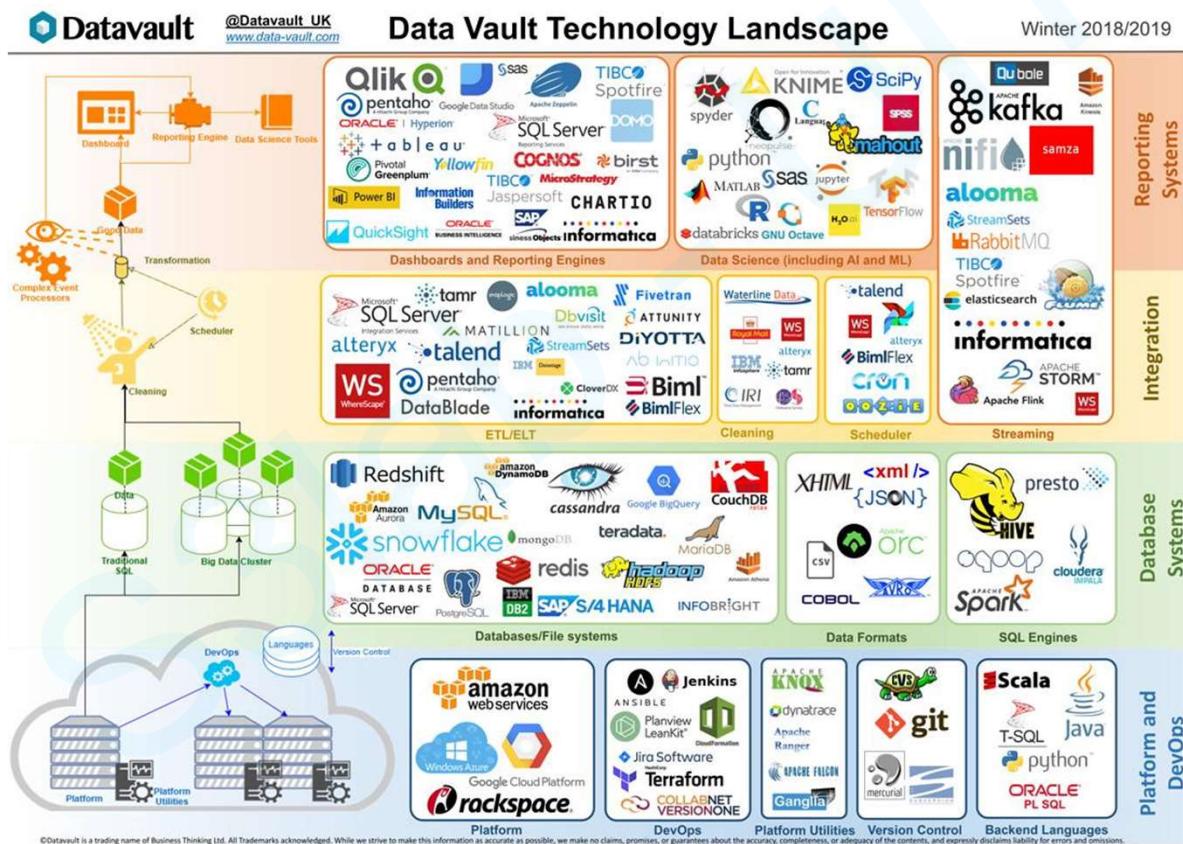
Big Data market is projected to grow from **\$42 B** in 2018 to **\$103B** in 2027 attaining a 10.48% CAGR.

★ IN 2020 ★

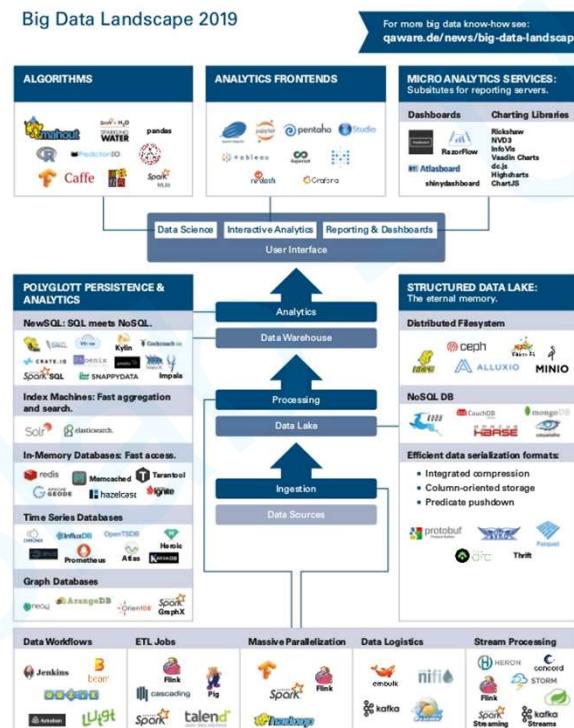
44 Zettabytes data	1.7 MB data created by each user/sec	6.1 Billion Smartphone users	21 Billion IOT Devices	1/3rd data passing through Cloud
---------------------------------	---	---	----------------------------------	--



Big Data: Landscape



Big Data: Landscape





Big Data: Landscape (Open sources)



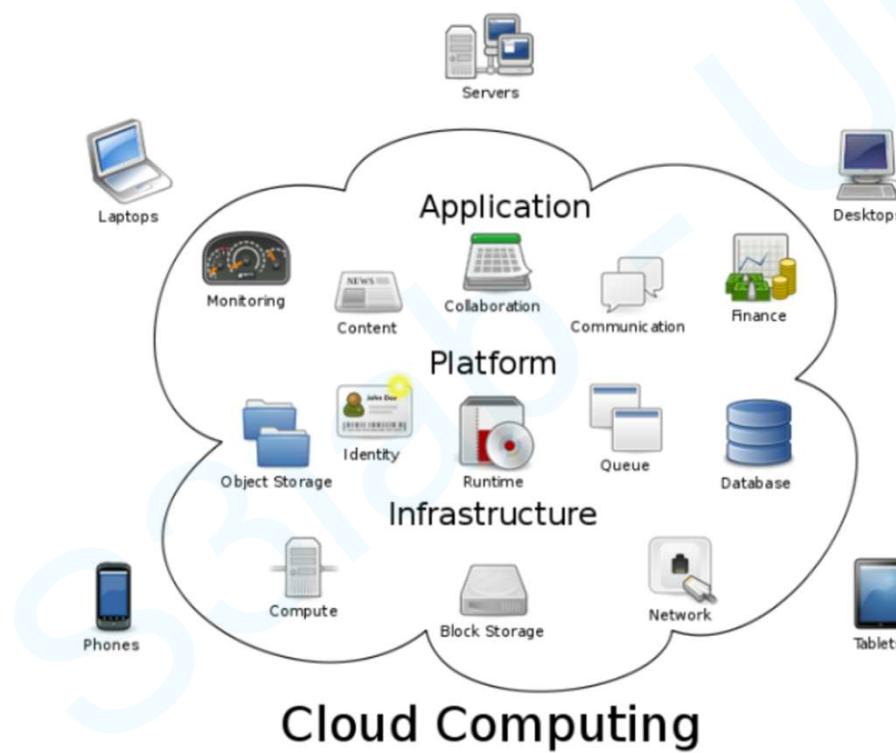


Big Data: Cloud Computing

- IT resources provided as a service
 - Compute, storage, databases, queues
- Clouds leverage economies of scale of commodity hardware
 - Cheap storage, high bandwidth network & multi-core processors
 - Geographically distributed data centers
- Offerings from Microsoft, Amazon, Google, ...



Big Data: Cloud Computing





Big Data: Cloud Computing

Benefits

- Cost & Management
 - Economies of scale, “outsourced” resource management
- Reduced time to deployment
 - Ease of assembly, works “out of the box”
- Scaling
 - On demand provisioning, co-locate data and compute
- Reliability
 - Massive, redundant, shared resources
- Sustainability: Hardware not owned



Big Data: Cloud Computing

Benefits





Big Data: Cloud Computing

Issues

- Data Security
 - Agree with the cloud service provider ensure data security.
- Performance
 - Service-Level Agreement (SLA) should be clear
- Compliance
 - Depend on the service provider
- Legal Issues
 - Data stored in multiple **locations**
- Cost: pay as per usage, use services in a controlled manner



Big Data: Cloud Computing

Deployment Models

- **Public:** computing infrastructure is hosted at the vendor's premises
- **Private:** Computing architecture is dedicated to customer and is not shared with other organizations.
- **Hybrid:** Host some critical, secure applications in private clouds. The not so critical applications are hosted in the public cloud
 - Cloud bursting: the organization uses its own infrastructure for normal usage, but cloud is used for peak loads.



Big Data: Cloud Computing

Type of Services

- Infrastructure as a service (**IaaS**):
 - Why buy machines when you can rent cycles?
 - Amazon's EC2, Rackspace
- Platform as a service (**PaaS**):
 - Give me nice API and take care of the maintenance, upgrades, ...
 - Google App Engine (GAE), Windows Azure
- Software as a service (**SaaS**):
 - Just run it for me
 - Gmail, Salesforce, dropbox



Big Data: Lambda Architecture

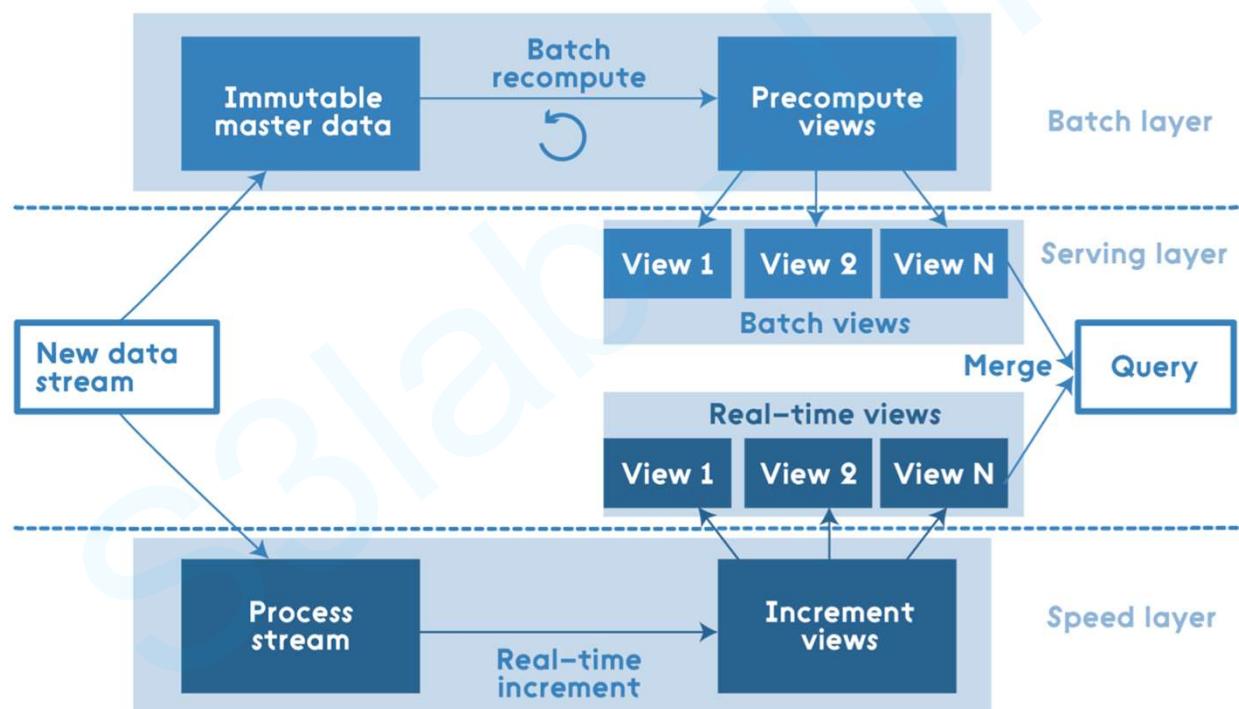
What is Lambda Architecture?

- This is the new big data architecture.
 - Designed to ingest and process
 - Query both fresh and historical (batch) data in a single architecture.
 - Solve the problem of computing arbitrary functions, contains 3 layers:
 - **Batch layer** (Data lake): historical archive, batch query, batch processes for general analytics or ad hoc.
 - **Serving layer**: handles serving up results. Also, combined with both the speed and batch layer.
 - **Speed layer**: queuing, stream, and do the same analytics as batch but in real-time on only the most recent data.



Big Data: Lambda Architecture

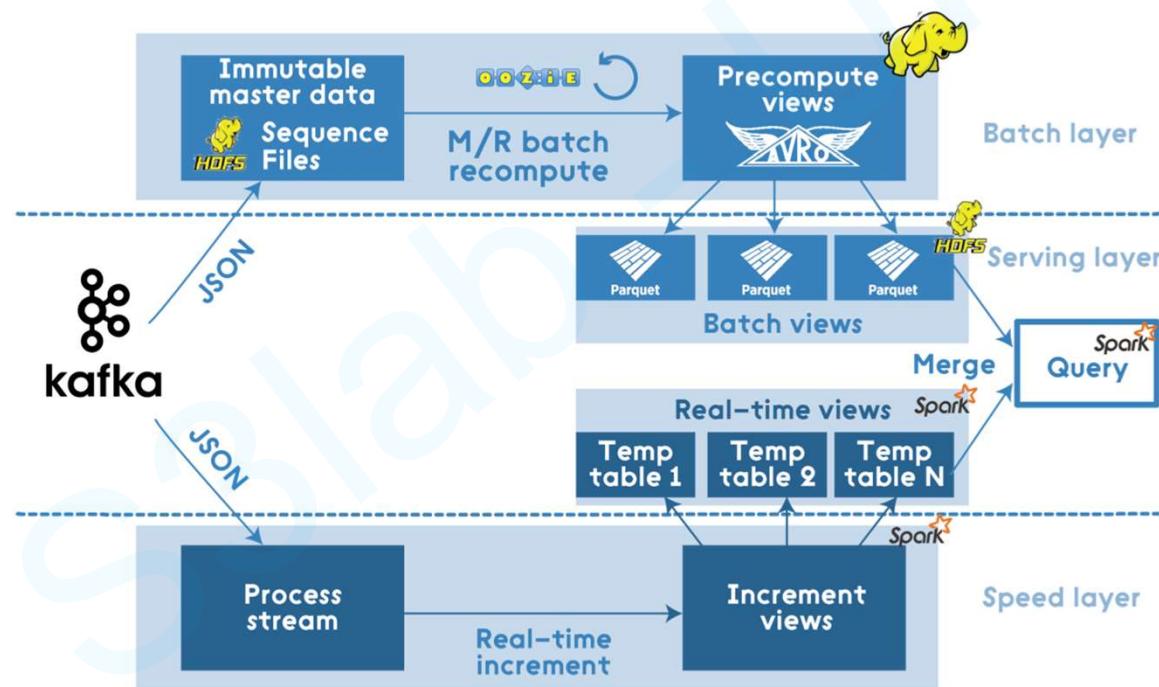
What is Lambda Architecture?





Big Data: Lambda Architecture

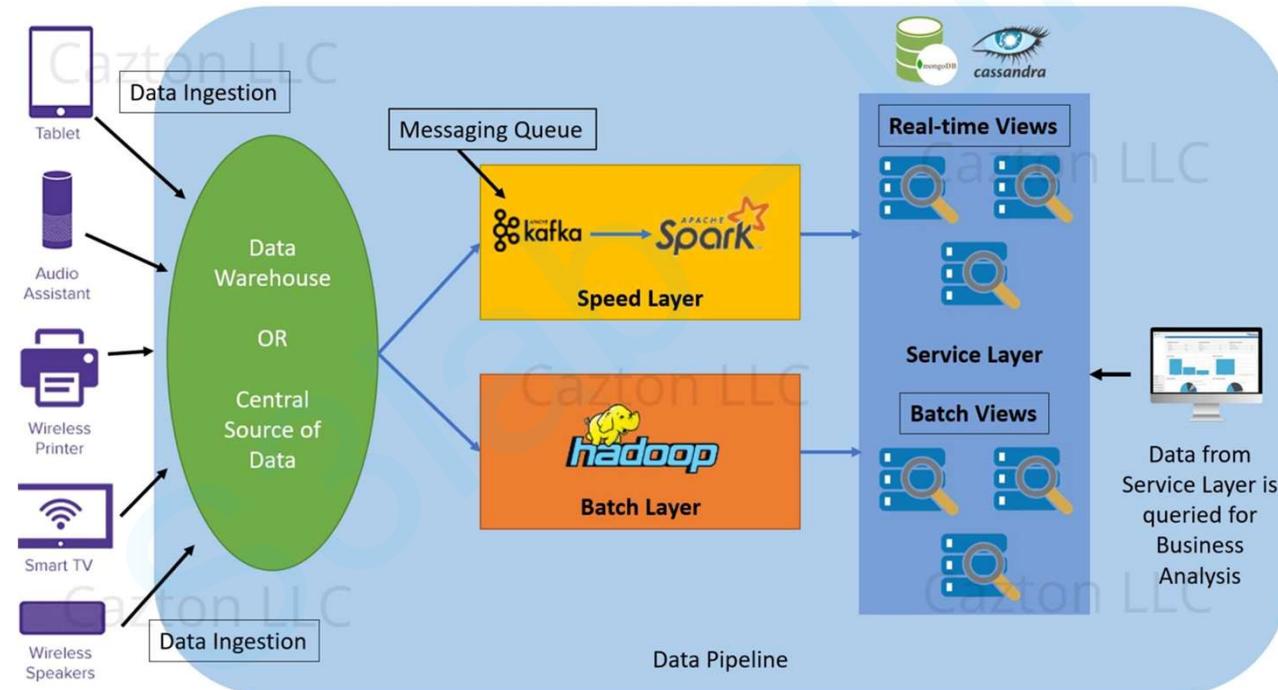
real-approach





Big Data: Lambda Architecture

real-approach



Q & A



Cảm ơn đã theo dõi

Chúng tôi hy vọng cùng nhau đi đến thành công.

S3 LAB

Nghiên cứu là để sẻ chia



INTRO
GIỚI THIỆU

PROJECTS
DỰ ÁN

OPENSOURCE
MÃ NGUỒN MỞ

SHOWCASE
TRÌNH DIỄN

INTER-JOB
TT VÀ CÔNG VIỆC

NEWS
THÔNG BÁO

CONTACT
LIÊN HỆ



Latest news



Mã Nguồn Mở

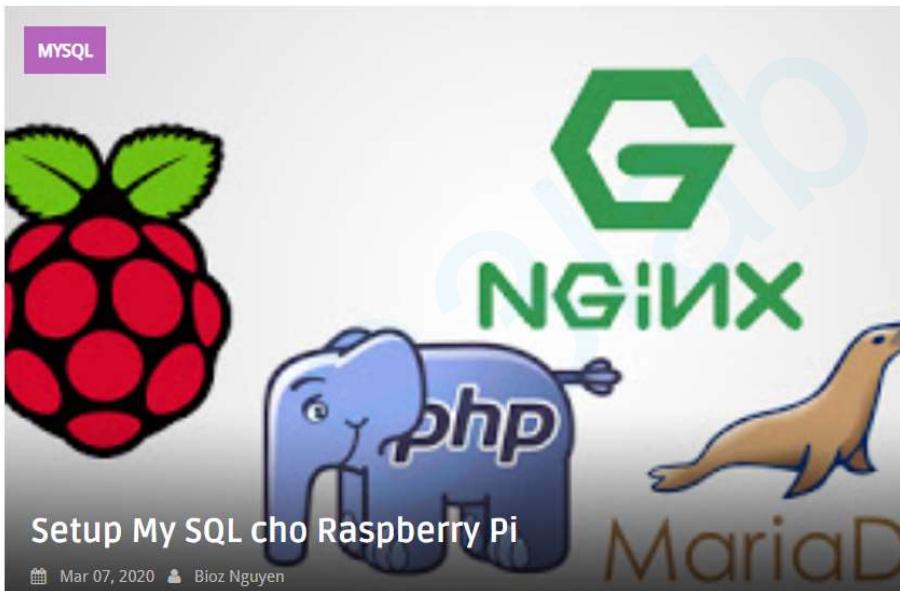
Video Wall Với Raspberry Pi

MySQL

Setup MySQL cho Raspberry Pi

Tài Liệu

Information Technology Project Management



Setup MySQL cho Raspberry Pi

Mar 07, 2020 Bioz Nguyen



Information Technology Project Management, Schwalbe

Aug 20, 2019 Bioz Nguyen



Quản lý dự án CNTT

Aug 20, 2019 Bioz Nguyen

8