

# Cloud and Big Data

## Big Data Overview

Gaurav Parashar  
gauravparashar24@gmail.com  
Lecture 1-3

December 9, 2019

# Table of Contents

- 1 Objectives
- 2 Introduction
  - What does "Big Data" Mean?
  - Applications of Big Data
  - Generation of Big Data
- 3 5 Vs of Big Data
  - Limitations of Existing Systems
- 4 Application of Big Data in Industry Verticals
  - Telecom
    - Call Detail Record
- 5 References

# Objectives

- Understand What is Big Data?

# Objectives

- Understand What is Big Data?
- Analyse limitations of existing systems

# Objectives

- Understand What is Big Data?
- Analyse limitations of existing systems
- Understand Hadoop and its features

# Objectives

- Understand What is Big Data?
- Analyse limitations of existing systems
- Understand Hadoop and its features
- Hadoop Ecosystem

# Objectives

- Understand What is Big Data?
- Analyse limitations of existing systems
- Understand Hadoop and its features
- Hadoop Ecosystem
- Understand Hadoop 2.x core components

# Objectives

- Understand What is Big Data?
- Analyse limitations of existing systems
- Understand Hadoop and its features
- Hadoop Ecosystem
- Understand Hadoop 2.x core components
- Perform read and write in Hadoop

# Objectives

- Understand What is Big Data?
- Analyse limitations of existing systems
- Understand Hadoop and its features
- Hadoop Ecosystem
- Understand Hadoop 2.x core components
- Perform read and write in Hadoop
- Big Data and Cloud Lab Guidelines

# Table of Contents

## 1 Objectives

## 2 Introduction

- What does "Big Data" Mean?
- Applications of Big Data
- Generation of Big Data

## 3 5 Vs of Big Data

- Limitations of Existing Systems

## 4 Application of Big Data in Industry Verticals

- Telecom
  - Call Detail Record

## 5 References

Objectives

**Introduction**

5 Vs of Big Data

Application of Big Data in Industry Verticals

References

What does "Big Data" Mean?

Applications of Big Data

Generation of Big Data

# What does "Big Data" Mean?

# What does "Big Data" Mean?

- Collecting large amounts of data

# What does "Big Data" Mean?

- Collecting large amounts of data
  - Computers, Databases, Video, Voice , tweets, comments, blogs, web pages, logs, calls, messages (Text + Whatsapp + ...) ,...

# What does "Big Data" Mean?

- Collecting large amounts of data
  - Computers, Databases, Video, Voice , tweets, comments, blogs, web pages, logs, calls, messages (Text + Whatsapp + ...) ,...

What do we do with this data?

# What does "Big Data" Mean?

- Collecting large amounts of data
  - Computers, Databases, Video, Voice , tweets, comments, blogs, web pages, logs, calls, messages (Text + Whatsapp + ...) ,...

What do we do with this data?

- Early Fraud Detection, Credit Card Frauds, Tick [1] Analytics

# What does "Big Data" Mean?

- Collecting large amounts of data
  - Computers, Databases, Video, Voice , tweets, comments, blogs, web pages, logs, calls, messages (Text + Whatsapp + ...) ,...

What do we do with this data?

- Early Fraud Detection, Credit Card Frauds, Tick [1] Analytics
- Content personalisation, Recommendation System

# What does "Big Data" Mean?

- Collecting large amounts of data
  - Computers, Databases, Video, Voice , tweets, comments, blogs, web pages, logs, calls, messages (Text + Whatsapp + ...) ,...

What do we do with this data?

- Early Fraud Detection, Credit Card Frauds, Tick [1] Analytics
- Content personalisation, Recommendation System
- Insurance: Personalised Pricing

# What does "Big Data" Mean?

- Collecting large amounts of data
  - Computers, Databases, Video, Voice , tweets, comments, blogs, web pages, logs, calls, messages (Text + Whatsapp + ...) ,...

What do we do with this data?

- Early Fraud Detection, Credit Card Frauds, Tick [1] Analytics
- Content personalisation, Recommendation System
- Insurance: Personalised Pricing
- Customer Loyalty Data

# What does "Big Data" Mean?

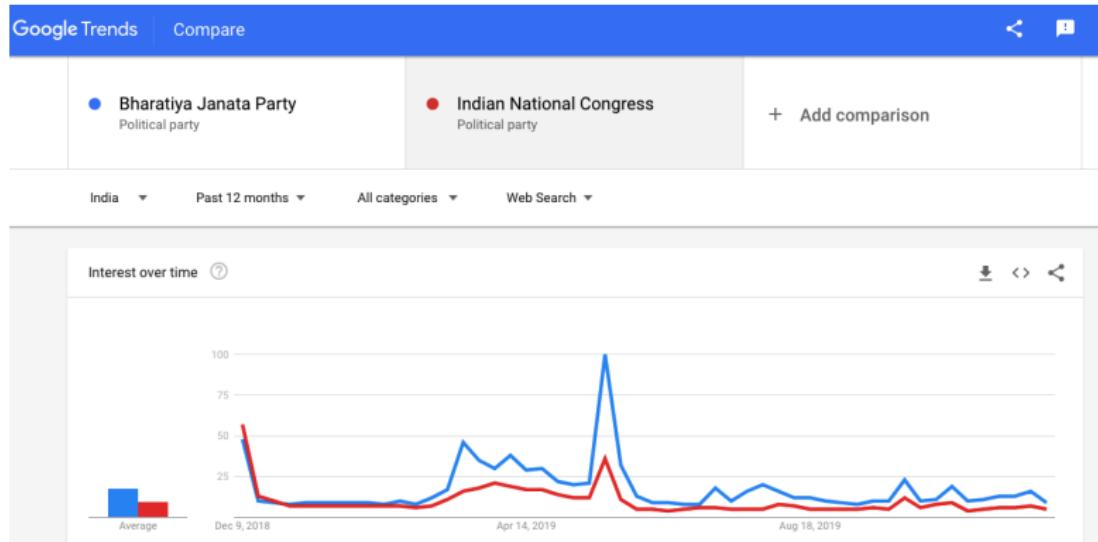
- Collecting large amounts of data
  - Computers, Databases, Video, Voice , tweets, comments, blogs, web pages, logs, calls, messages (Text + Whatsapp + ...) ,...

What do we do with this data?

- Early Fraud Detection, Credit Card Frauds, Tick [1] Analytics
- Content personalisation, Recommendation System
- Insurance: Personalised Pricing
- Customer Loyalty Data
- Predicting Future

# Applications of Big Data

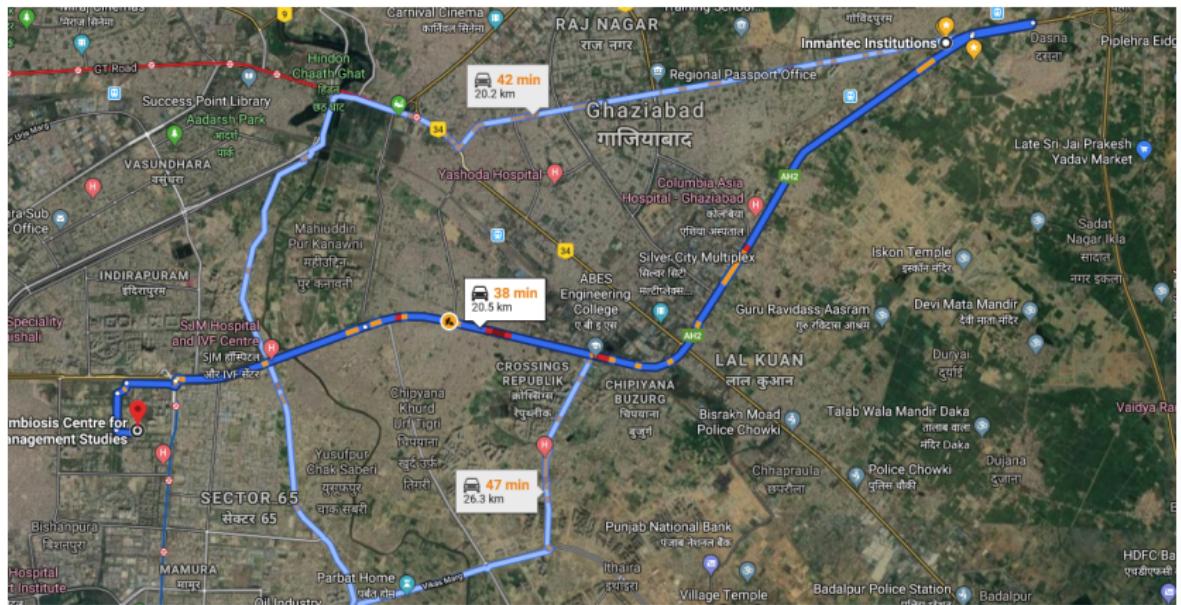
Examples:  
Google Trends



<sup>1</sup>Google Trends for BJP and INC

# Applications of Big Data

## Maps



# Applications of Big Data

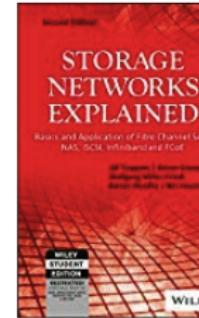
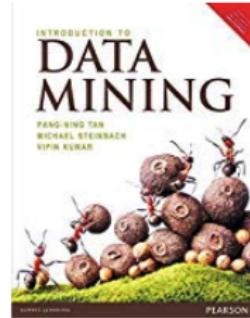
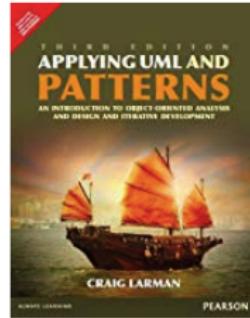
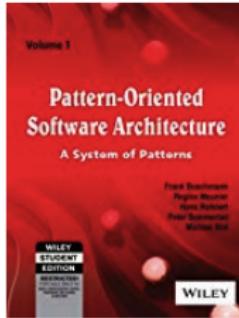
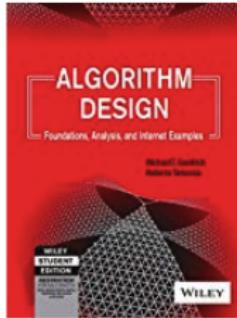
## Recommendation Systems

[Explore all](#)

[See more](#)

[See more](#)

### Recommendations for you in Books



# Applications of Big Data

## Twitter Trends

The image shows a Twitter feed with a tweet from Sumit Kumar (@sumitmeetg2007) dated Dec 5. The tweet content is: "16K Continuous Hill Run followed by drills and strength exercises ./ #MarathonTraining #running 💪💪". Below the tweet, there is a link to Strava: "Check out my activity on Strava: strava.app.link/EVnBA3x091". The Strava activity image shows five people standing together, with one person running. The activity details are: "Run 15.0 km Pace 6:33 /km". To the right of the main feed, there is a "Trends for you" sidebar. It features three trending topics: "#ChangeWhatYouCan" (MG ZS EV - India's First Pure Electric Internet SUV - Coming Soon!), "#hyderabadpolice" (130K Tweets), and "#HumanRights" (25K Tweets). There are also three "Trending in India" sections: "#NDTV" (Opinion divided on killing of accused in Telangana vet's ...), "#Encounter" (205K Tweets), and "#RIPDisha" (5,200 Tweets).

4

## <sup>4</sup>Trending in Twitter

# Applications of Big Data

## Big Data in Sports [2]

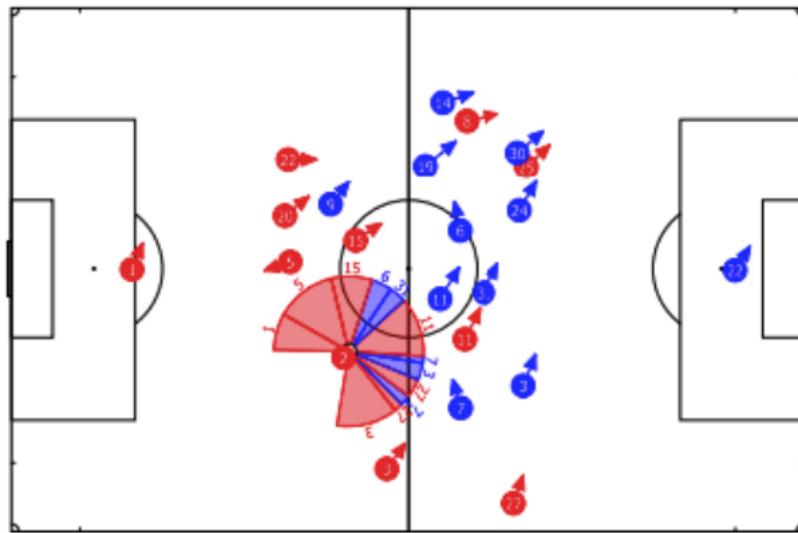


Figure: Available receivers of pass by Red2

# Applications of Big Data

- Weather Prediction
- Medical Diagnosis
- Smart Cities and Buildings

# Generation of Big Data

## What Happens in an Internet Minute?



# Table of Contents

## 1 Objectives

## 2 Introduction

- What does "Big Data" Mean?
- Applications of Big Data
- Generation of Big Data

## 3 5 Vs of Big Data

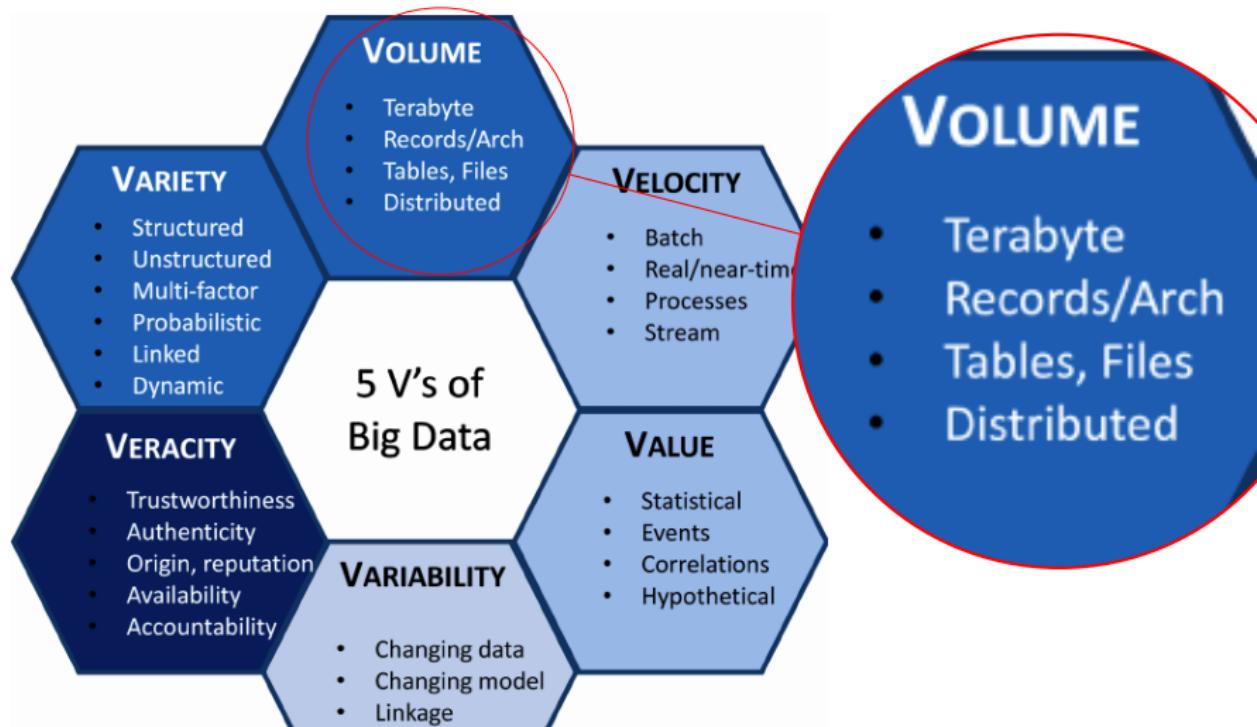
- Limitations of Existing Systems

## 4 Application of Big Data in Industry Verticals

- Telecom
- Call Detail Record

## 5 References

# 5 Vs of Big Data[3]: Volume 24



## Scenario: Volume

What is the starting limit of Big Data?

A > 1 GB - < 1TB

## Scenario: Volume

What is the starting limit of Big Data?

- A > 1 GB - < 1TB
- B > 1 TB - < 1PB

## Scenario: Volume

What is the starting limit of Big Data?

- A > 1 GB - < 1TB
- B > 1 TB - < 1PB
- C > 1 PB - < 1EB

## Scenario: Volume

What is the starting limit of Big Data?

- A > 1 GB - < 1TB
- B > 1 TB - < 1PB
- C > 1 PB - < 1EB
- D > 1 EB - < 1ZB

## Scenario: Volume

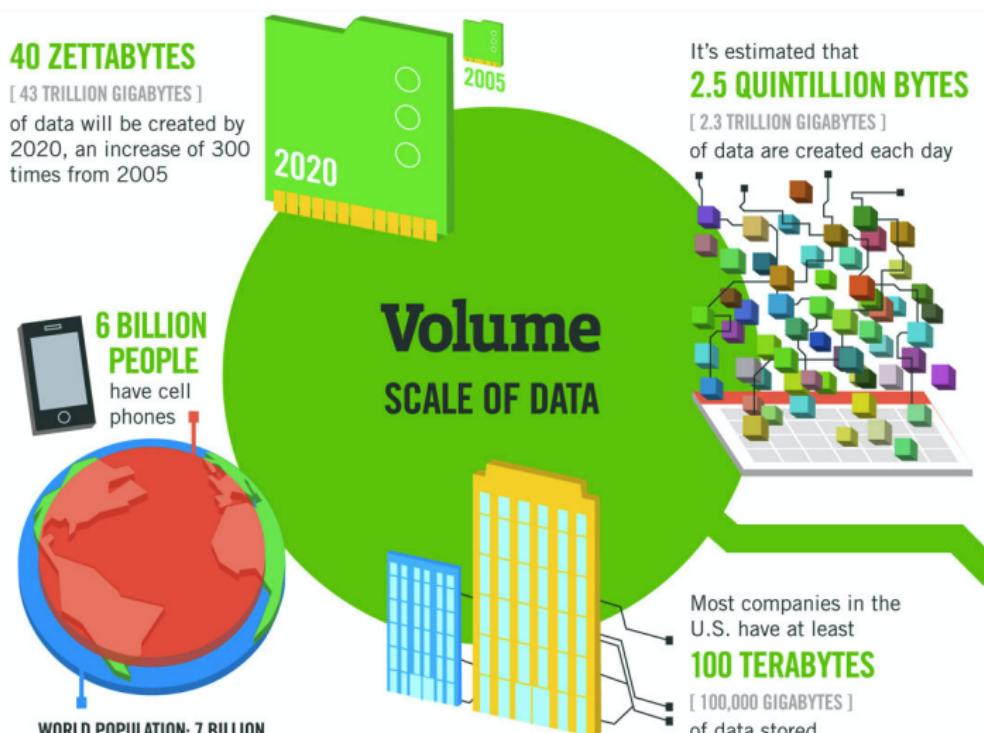
What is the starting limit of Big Data?

- A > 1 GB - < 1TB
- B > 1 TB - < 1PB
- C > 1 PB - < 1EB
- D > 1 EB - < 1ZB

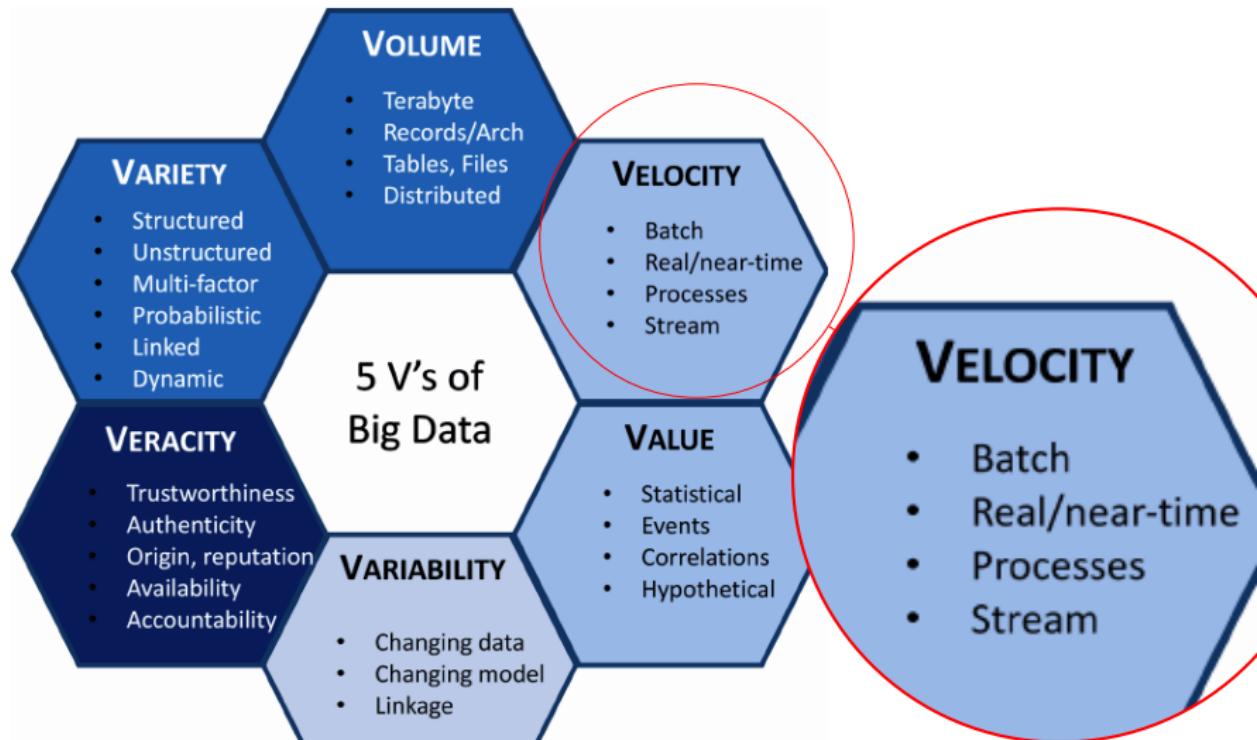
Answer:

Depends upon the organisation definition of Big Data. Relative term.

# Scenario: Volume 24



# 5 Vs of Big Data: Velocity 24



# 5 Vs of Big Data: Velocity

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

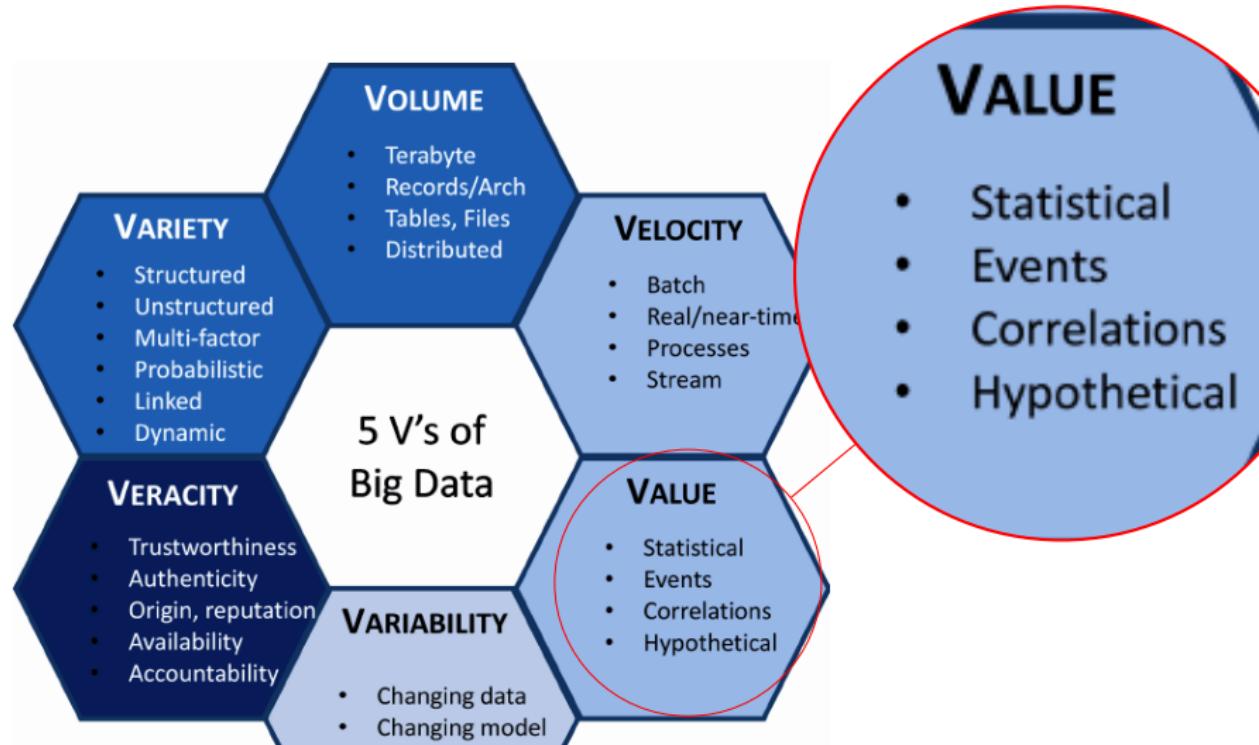
By 2016, it is projected  
there will be

**18.9 BILLION  
NETWORK  
CONNECTIONS**

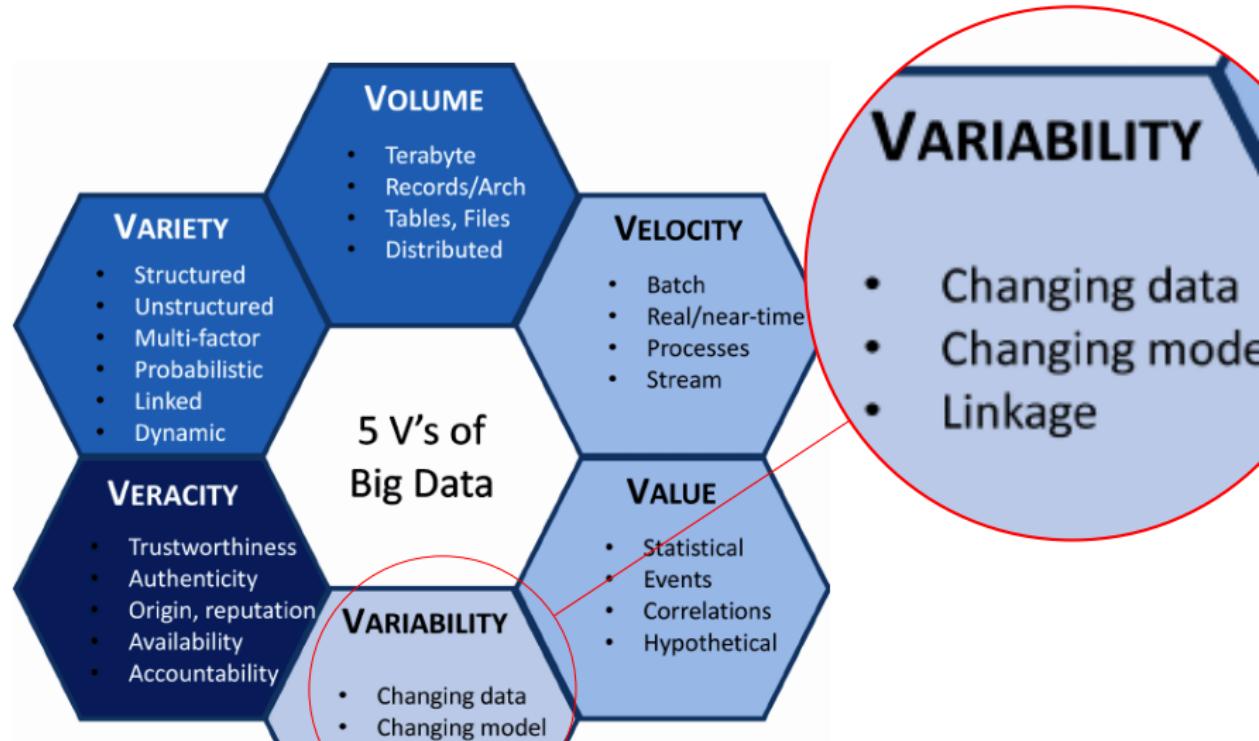
– almost 2.5 connections  
per person on earth



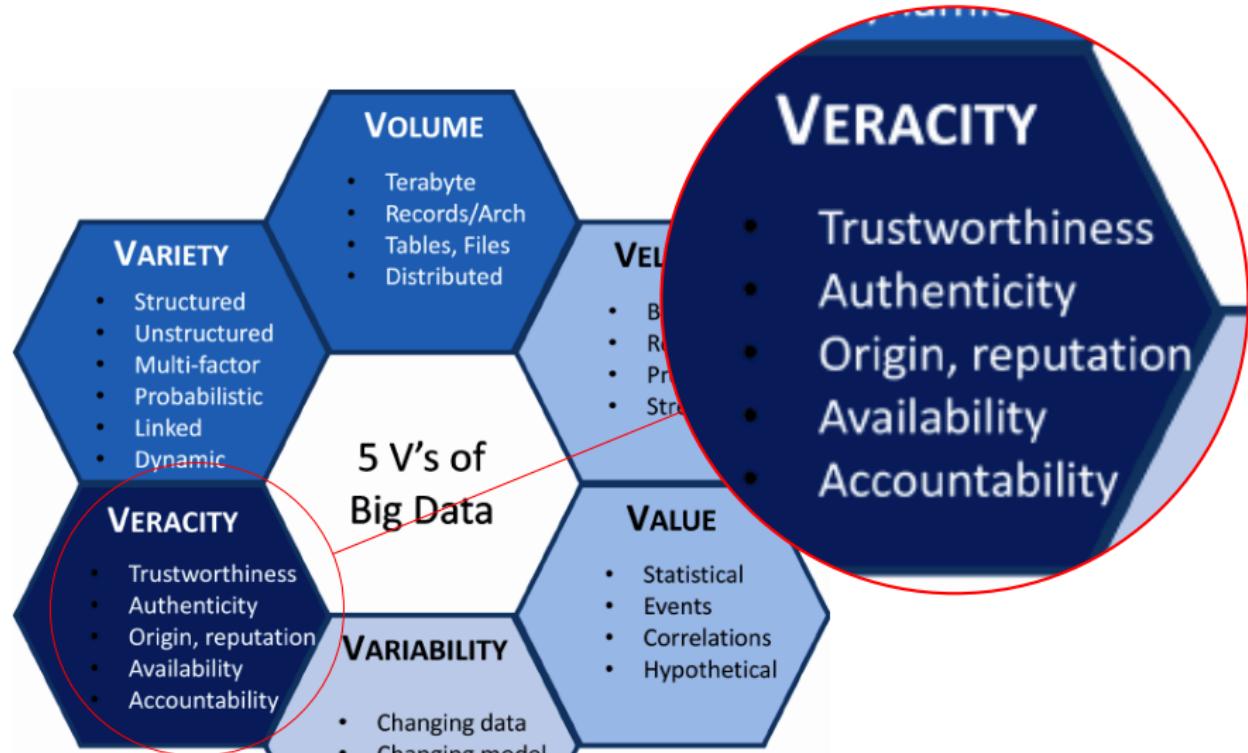
# 5 Vs of Big Data: Value 24



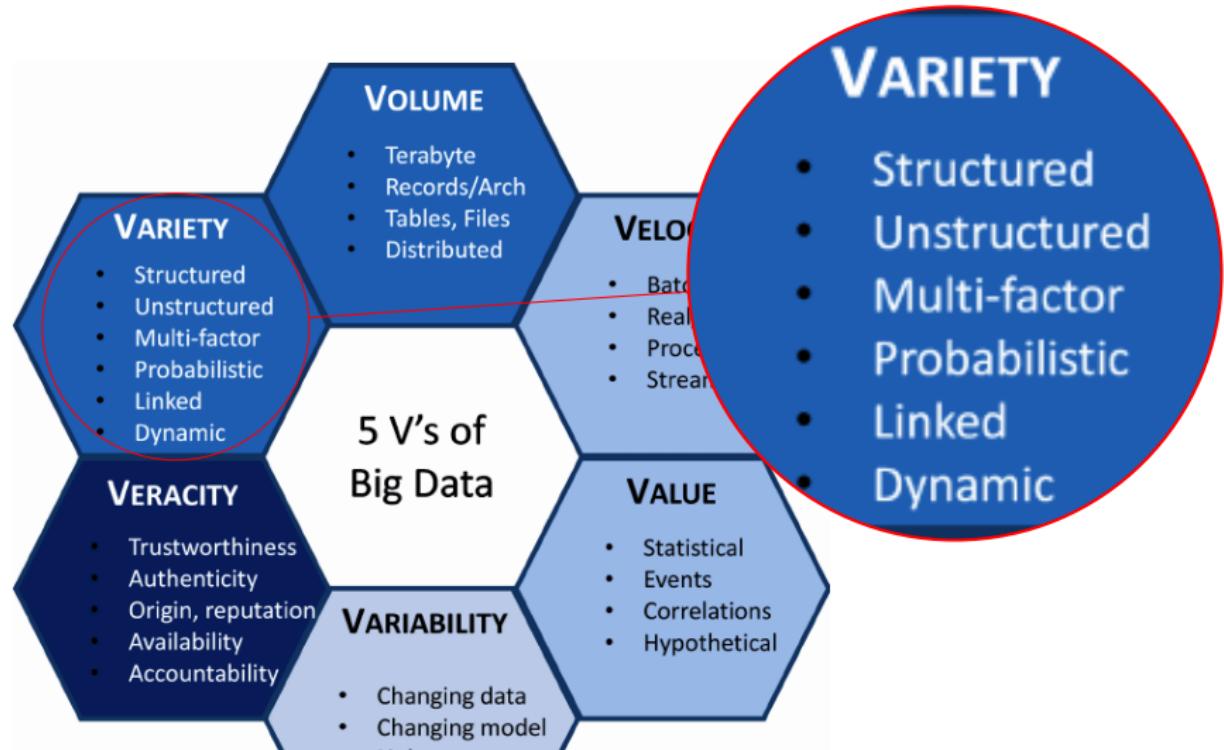
# 5 Vs of Big Data: Variability 24



# 5 Vs of Big Data: Veracity 24



# 5 Vs of Big Data: Variety 24



# Limitations of Existing Systems

- ① Cost of scaling

# Limitations of Existing Systems

- ① Cost of scaling
- ② Vertical Scaling

# Limitations of Existing Systems

- ① Cost of scaling
- ② Vertical Scaling
- ③ Integration with legacy systems

# Limitations of Existing Systems

- ① Cost of scaling
- ② Vertical Scaling
- ③ Integration with legacy systems
- ④ Rapid Change in type of data

# Limitations of Existing Systems

- ① Cost of scaling
- ② Vertical Scaling
- ③ Integration with legacy systems
- ④ Rapid Change in type of data
- ⑤ Lack of skills

# Table of Contents

- 1 Objectives
- 2 Introduction
  - What does "Big Data" Mean?
  - Applications of Big Data
  - Generation of Big Data
- 3 5 Vs of Big Data
  - Limitations of Existing Systems
- 4 Application of Big Data in Industry Verticals
  - Telecom
    - Call Detail Record
- 5 References

# Telecom

A telco[4] serving 8 million prepaid mobile subscribers

- ① Volume 13: amounting to 11 billion records annually
- ② Velocity 16: generates approximate 30 million CDR[5]s daily
- ③ Value 18: \_\_\_\_\_
- ④ Variability 19: \_\_\_\_\_
- ⑤ Veracity 20: \_\_\_\_\_
- ⑥ Variety 21: \_\_\_\_\_

---

UID, DATE (yyyy-MM-dd), TRIP\_SEQUENCE\_ID, MOBILITY\_TYPE, TRANSPORT\_MODE,  
TOTAL\_DISTANCE, TOTAL\_TIME, START\_TIME, END\_TIME, TOTAL\_POINTS, POINT\_LIST  
422a837717,2015-06-06,1,STAY,STAY,0.000,57749.000,00:00:00,16:02:29,1,1|2015-06-01  
00:00:00|6.373743|-10.772951

422a837717,2015-06-06,2,MOVE,WALK,3153.708,2323.000,16:02:29,16:41:12,39,1|2015-  
06-01 16:02:29|6.373497|-10.773267;2|2015-06-01 16:03:30|6.374243|-10.773447;3|2015-06-  
01 16:04:31|6.374983|-10.773652;4|2015-06-01 16:05:32|6.375711|-10.773898;5|2015-06-01  
16:06:33|6.376103|-10.774265;6|2015-06-01 16:07:34|6.375691|-10.774913;7|2015-06-01  
16:08:35|6.375280|-10.775561;8|2015-06-01 16:09:36|6.374868|-10.776209;9|2015-06-01  
16:10:38|6.374457|-10.776858;10|2015-06-01 16:11:39|6.374046|-10.777506;11|2015-06-01  
16:12:40|6.373634|-10.778154;12|2015-06-01 16:13:41|6.373223|-10.778802;13|2015-06-01  
16:14:42|6.372842|-10.779469;14|2015-06-01 16:15:43|6.372488|-10.780150;15|2015-06-01  
16:16:44|6.372098|-10.780811;16|2015-06-01 16:17:45|6.371690|-10.781461;17|2015-06-01  
16:18:47|6.371286|-10.782115;18|2015-06-01 16:19:48|6.370883|-10.782768;19|2015-06-01  
16:20:49|6.370479|-10.783421;20|2015-06-01 16:21:50|6.370076|-10.784074;21|2015-06-01  
16:22:51|6.369672|-10.784727;22|2015-06-01 16:23:52|6.369238|-10.785360;23|2015-06-01  
16:24:53|6.368795|-10.785988;24|2015-06-01 16:25:55|6.368358|-10.786618;25|2015-06-01  
16:26:56|6.368053|-10.787320;26|2015-06-01 16:27:57|6.368022|-10.787936;27|2015-06-01  
16:28:58|6.368780|-10.788057;28|2015-06-01 16:29:59|6.369542|-10.788139;29|2015-06-01  
16:31:00|6.370310|-10.788158;30|2015-06-01 16:32:01|6.371075|-10.788096;31|2015-06-01  
16:33:02|6.371611|-10.788376;32|2015-06-01 16:34:04|6.371883|-10.789094;33|2015-06-01  
16:35:05|6.372137|-10.789817;34|2015-06-01 16:36:06|6.372195|-10.790581;35|2015-06-01  
16:37:07|6.372183|-10.791348;36|2015-06-01 16:38:08|6.372054|-10.791999;37|2015-06-01  
16:39:09|6.371293|-10.791979;38|2015-06-01 16:40:10|6.371313|-10.792746;39|2015-06-01  
16:41:12|6.371295|-10.793513

422a837717,2015-06-06,3,STAY,STAY,0.000,34295.000,16:41:12,23:59:59,1,1|2015-06-06  
16:41:12|6.371295|-10.793513

---

# Table of Contents

- 1 Objectives
- 2 Introduction
  - What does "Big Data" Mean?
  - Applications of Big Data
  - Generation of Big Data
- 3 5 Vs of Big Data
  - Limitations of Existing Systems
- 4 Application of Big Data in Industry Verticals
  - Telecom
  - Call Detail Record
- 5 References

-  [Investopedia, “Financial advisory website.”](#)  
Accessed on 2019-12-06.
-  [J. Gudmundsson and M. Horton, “Spatio-temporal analysis of team sports - A survey,” \*CoRR\*, vol. abs/1602.06994, 2016.](#)
-  [IBM, “ibm big data platform - bringing big data to the enterprise.”](#)  
Accessed on 2019-12-06.
-  [T. T. B. Services, “Big data and the telecom industry.”](#)  
Accessed on 2019-12-09.
-  [I. T. Union, “Call detail record \(cdr\) analysis: Republic of liberia.”](#)  
Accessed on 2019-12-09.