

# Data Science and Applications

# Data All Around

- Lots of data is being collected and warehoused
  - Web data, e-commerce
  - Financial transactions, bank/credit transactions
  - Online trading and purchasing
  - Social Network



# Digital Advertisement

NETFLIX



One Month Free



www.netflix.com

SIGN UP

Get started



stackoverflow  
JOBS



SHEIN



Find your dream job  
on a career site built just for developers



Certificate  
Programme in  
**BUSINESS  
ANALYTICS**

Class of 2020, Summer  
Hyderabad | Mohali

Last Date to Apply: Dec 13, 2018

Apply Now



prime  
video

#KaleenBhaiya King of Mirzapur



I earn \$1,000 a month  
without any knowledge  
or experience.

Ankur I

# Internet Search



cats|

cats musical  
cats and dogs  
cats for sale  
cats musical songs  
cats breeds  
cats funny  
cats game  
cats protection  
cats meowing  
cats for adoption



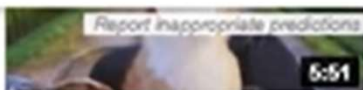
10:06

Cats are so funny you  
will die laughing - Funny  
cat compilation



7:00

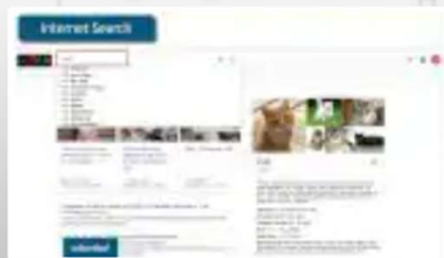
Why Humans Are  
Obsessed with Cats |  
Annals of Obsession |  
The ...



5:51

How I Trained My Cats

⬆ Pull up for precise seeking



The New Yorker  
YouTube - Mar 27, 2018

JunsKitchen  
YouTube - Nov 14, 2017

for Cats | Cat Breed Information, Cat ...

Your cat's online owners manual featuring articles about breed information, cat selection, training



More Images

## Cat

Animal

The cat, often referred to as the domestic cat to distinguish from other felids and felines, is a small, typically furry, carnivorous mammal. It is often called house cat when kept as indoor pet or feral/feral domestic cat when wild. It is often valued by humans for companionship and for its ability to hunt vermin. [Wikipedia](#)

Lifespan: 2 – 16 years (in the wild)

Scientific name: Felis catus

## Website Recommendation

### Compare with similar items



**This item** Bose SoundLink Wireless Around-Ear Headphones with Mic (Black)

Add to Cart



Sennheiser HD 4.40-BT Bluetooth Headphones (Black)

Add to Cart



Bose 741158-0020 SoundLink Wireless Around-Ear Headphones with Mic (White)

Add to Cart



Bose 789564-0030 Quiet Comfort 35 Wireless Headphone (Blue)-Special Edition

Add to Cart

Customer Rating

★★★★☆ (68)

★★★★☆ (349)

★★★★☆ (22)

★★★★☆ (200)

Price

₹ 19,000.00

₹ 7,490.00

₹ 19,000.00

₹ 29,363.00

Shipping

FREE Shipping

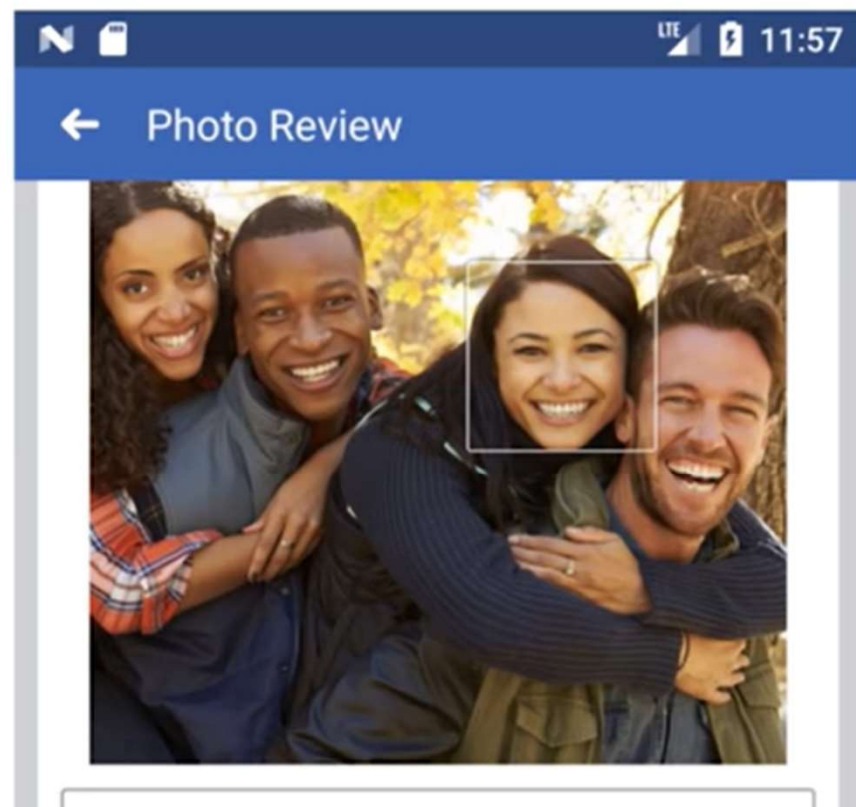
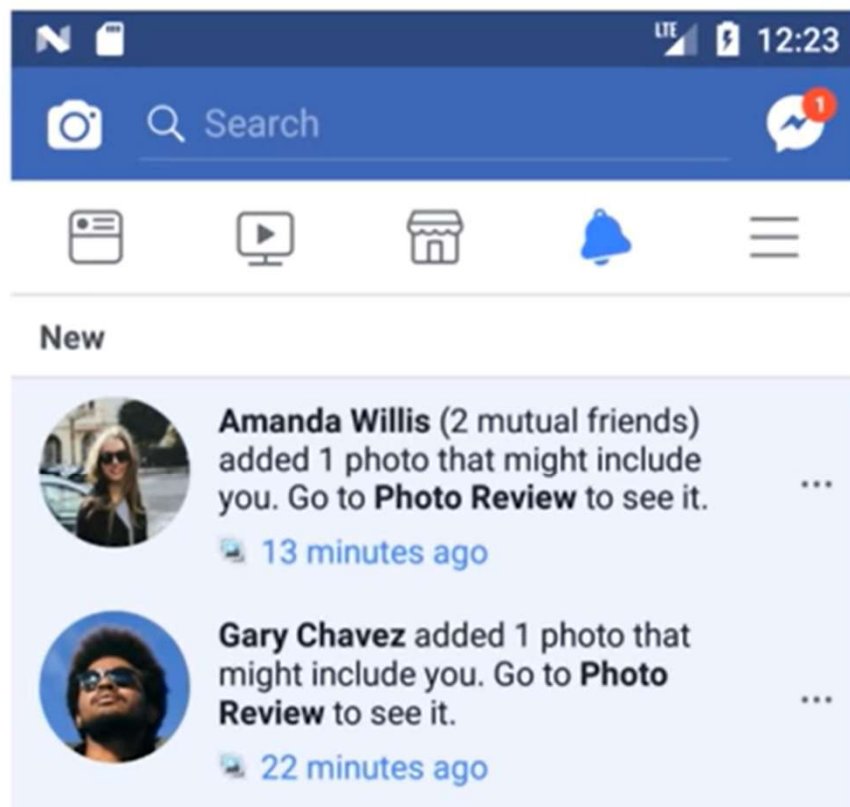
FREE Shipping

FREE Shipping

FREE Shipping



# Image Recognition



## Speech Recognition



# Gaming





## Medical Sciences



Medical  
Image  
Analysis



Genetics  
and  
Genomics



Creation  
Of  
Drugs



Virtual  
Assistance for  
Patients

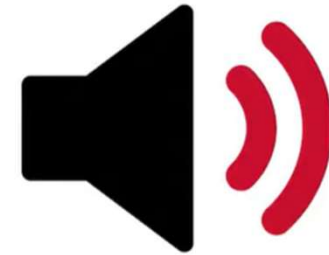
# **Data** can be



field notes



videos



audio recordings



photographs



documents

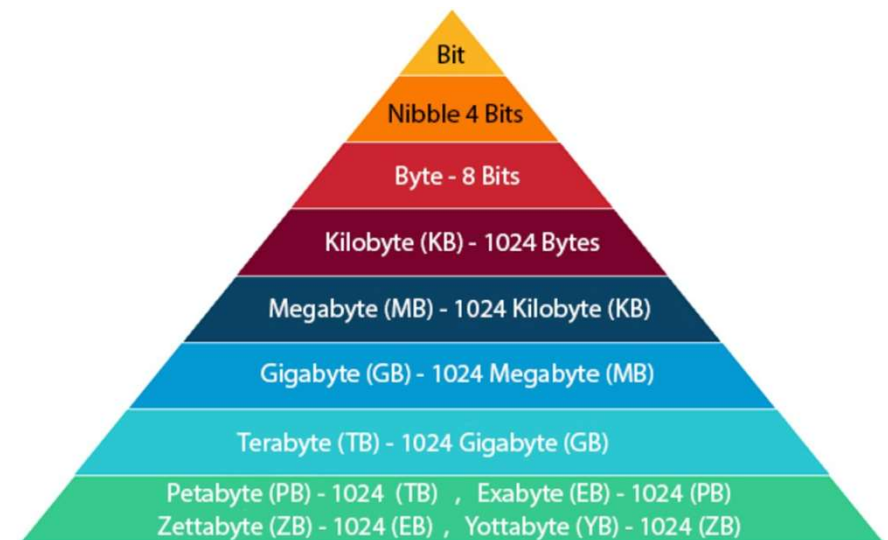


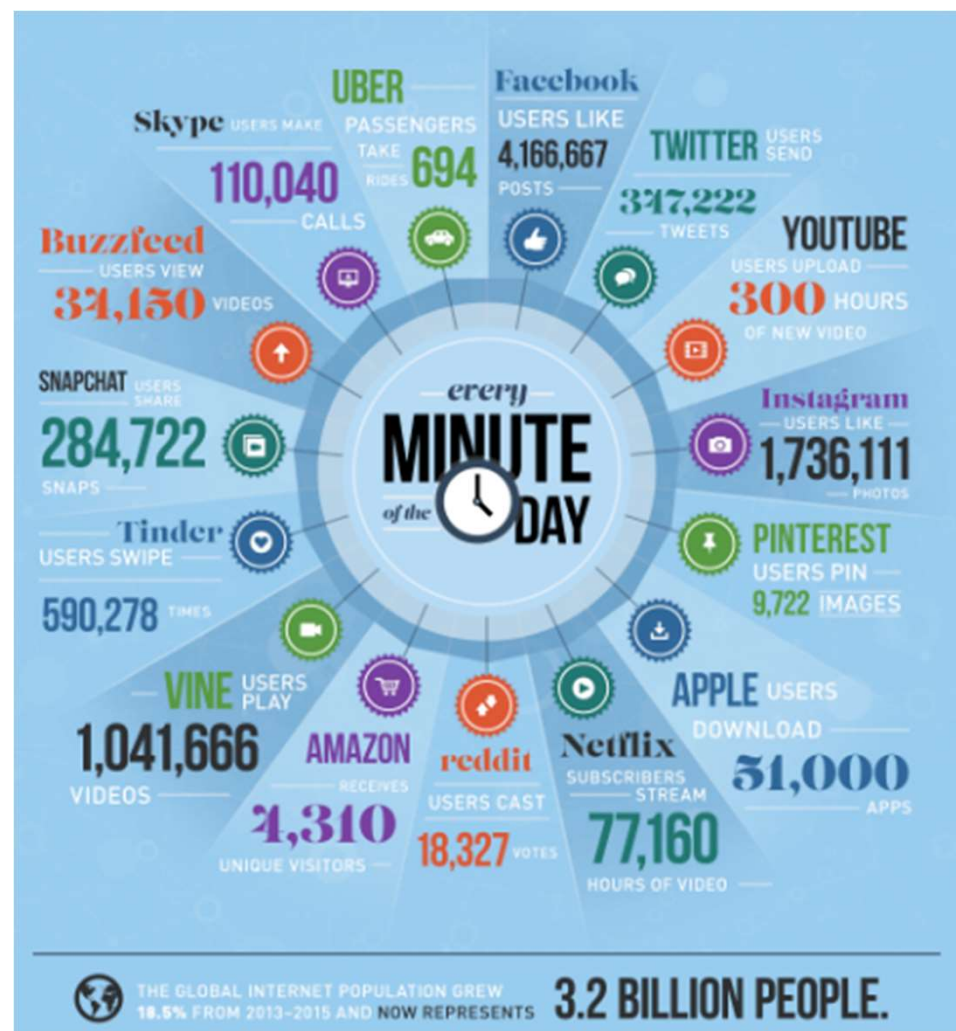
transcripts

# How Much Data Do We have?

- Google processes 20 PB a day (2008)
- Facebook has 60 TB of daily logs
- eBay has 6.5 PB of user data + 50 TB/day (5/2009)
- 1000 genomes project: 200 TB

- Cost of 1 TB of disk: \$35
- Time to read 1 TB disk: 3 hrs  
(100 MB/s)





# Data Sources

Evolution of  
Technology

IOT

Social Media

Other factors



Telephone



Desktop



Car



Mobile



Cloud



Smart Car





# Data Sources

---

Evolution of  
Technology

IOT

Social Media

Other factors



1,736,111 pictures



347,222 tweets



204,000,000  
emails



4,166,667 likes &  
200,000 photos



300 hours of video  
uploaded

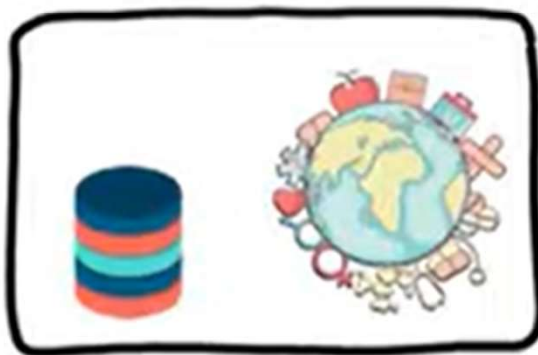
# Data Sources



# This is possible with the concept of

# 5 V's

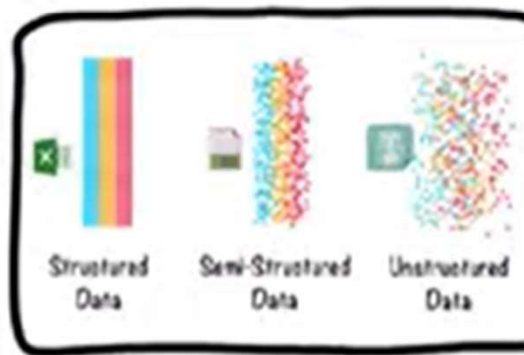
Volume



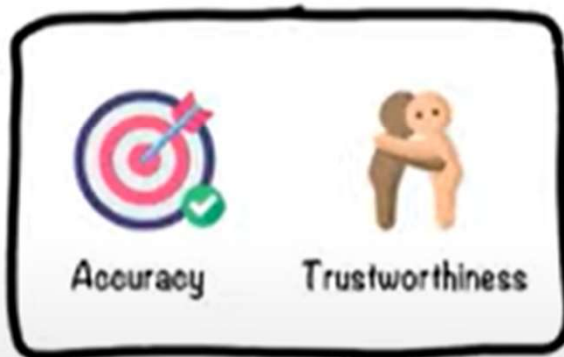
Velocity



Variety



Veracity



Value



# Data Analysis At Walmart

Halloween and cookie sales



*Data scientist at Walmart found a connection between Halloween and the sales of cookies.*



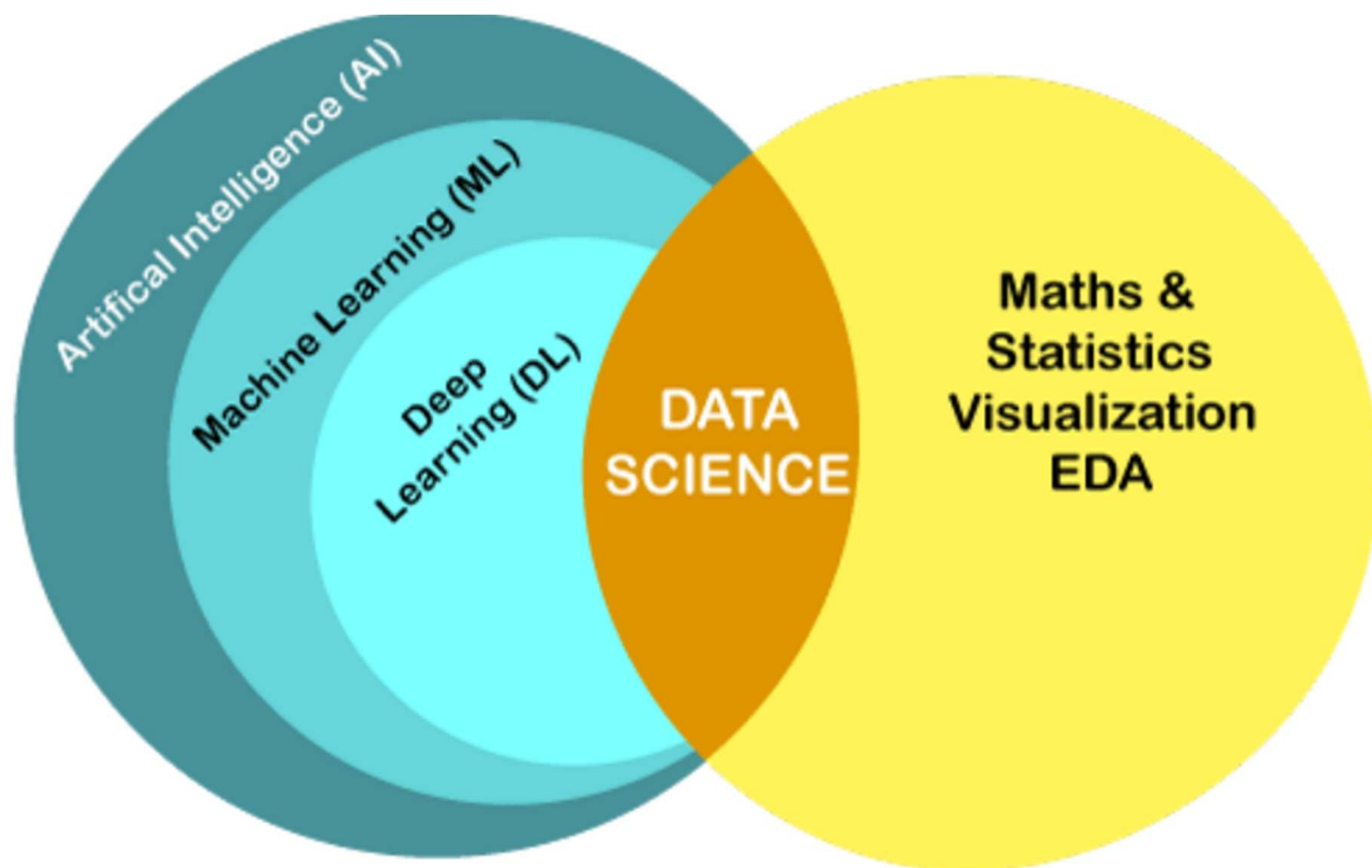
# Data Analysis At Walmart

---

Social media and cake pops



*Walmart is leveraging social media data to find about the trending products so that they can be introduced to the Walmart stores across the world*



# What is Data Science?



**"Torture the data, and it will confess to anything."**  
~ Ronald Coase, Economics, Nobel Prize

**Data Science** is the process of extracting knowledge and insights from data by using *scientific methods*.



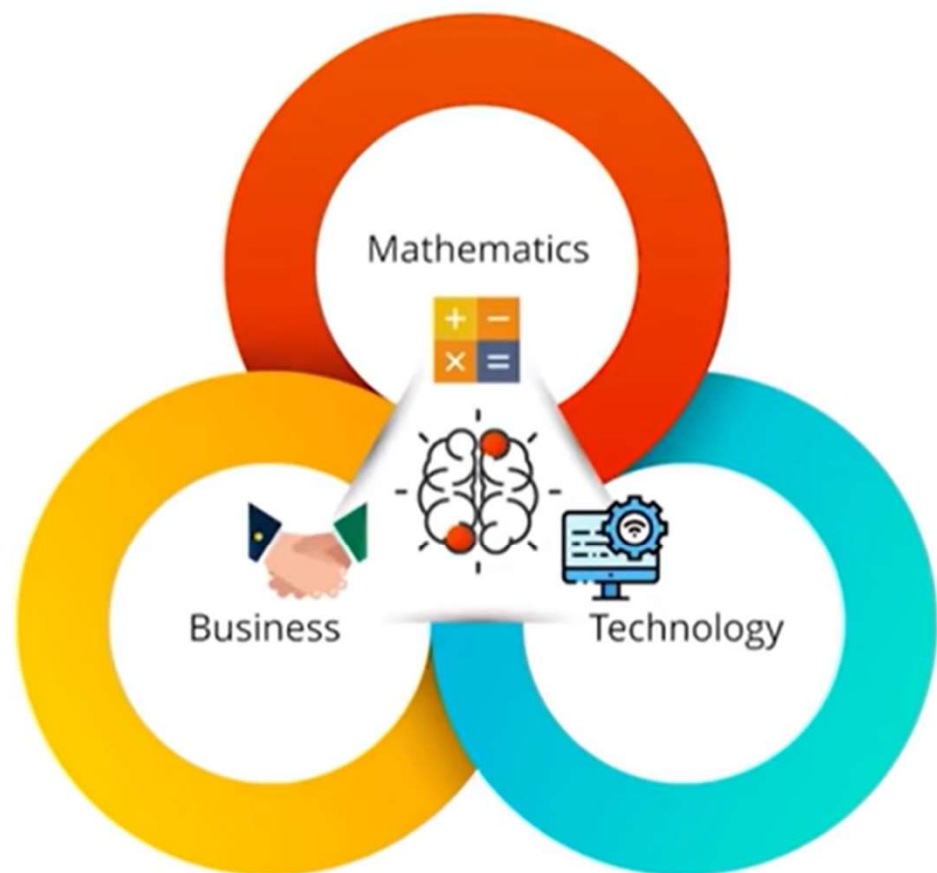
Scientific methods:

**Programming** + **Statistics** + **Business**



# Who Is A Data Scientist?

---



# Data Science – Skill Set

---



Statistics



Programming  
languages



Data extraction &  
processing



Data wrangling &  
exploration



Machine Learning



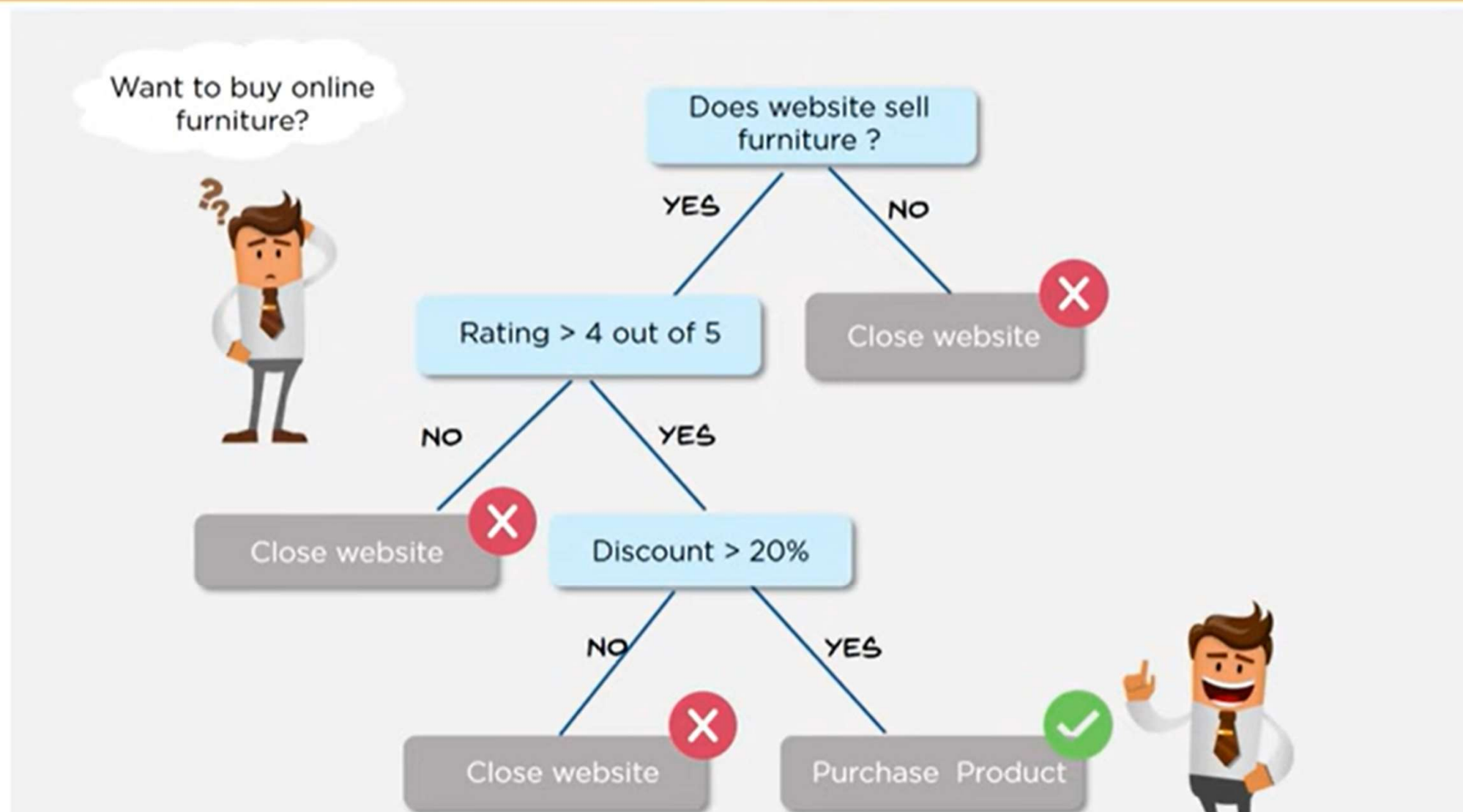
Big Data processing  
frameworks



Data visualisation



# What is Data Science?



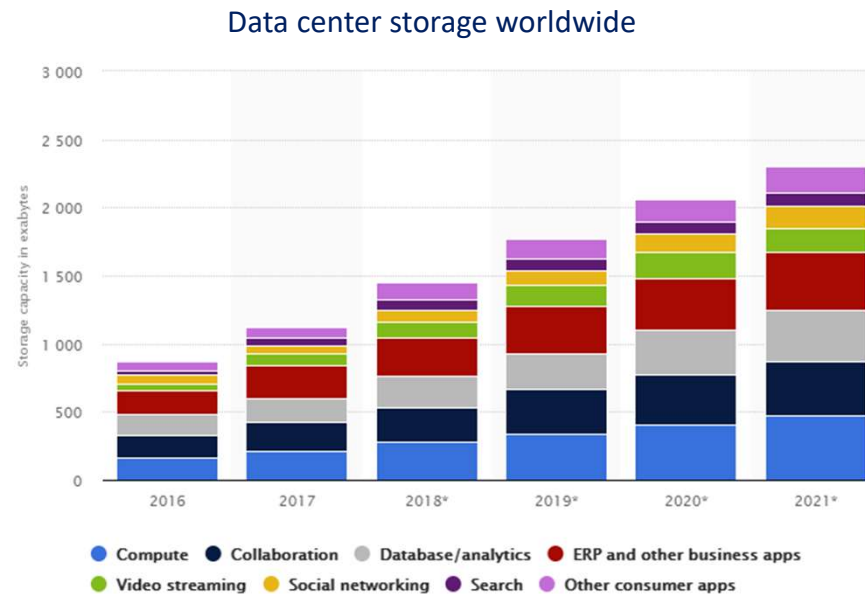
# Big Data

◆ Big Data is any data that is expensive to manage and hard to extract value from

- Volume
  - The size of the data
- Velocity
  - The latency of data processing relative to the growing demand for interactivity
- Variety and Complexity
  - the diversity of sources, formats, quality, structures.

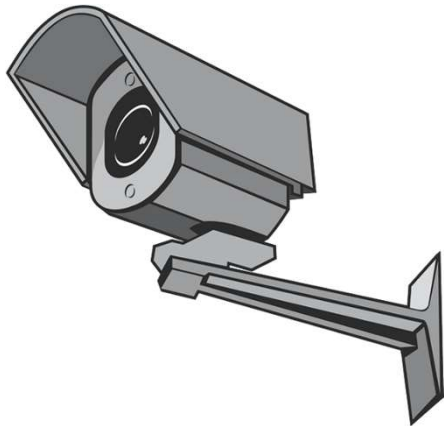
# The 3 Vs: Volume

- Scale of the data must be „big“
  - No clear definition
  - „that demand [...] innovative forms of information processing“ (Gartner)



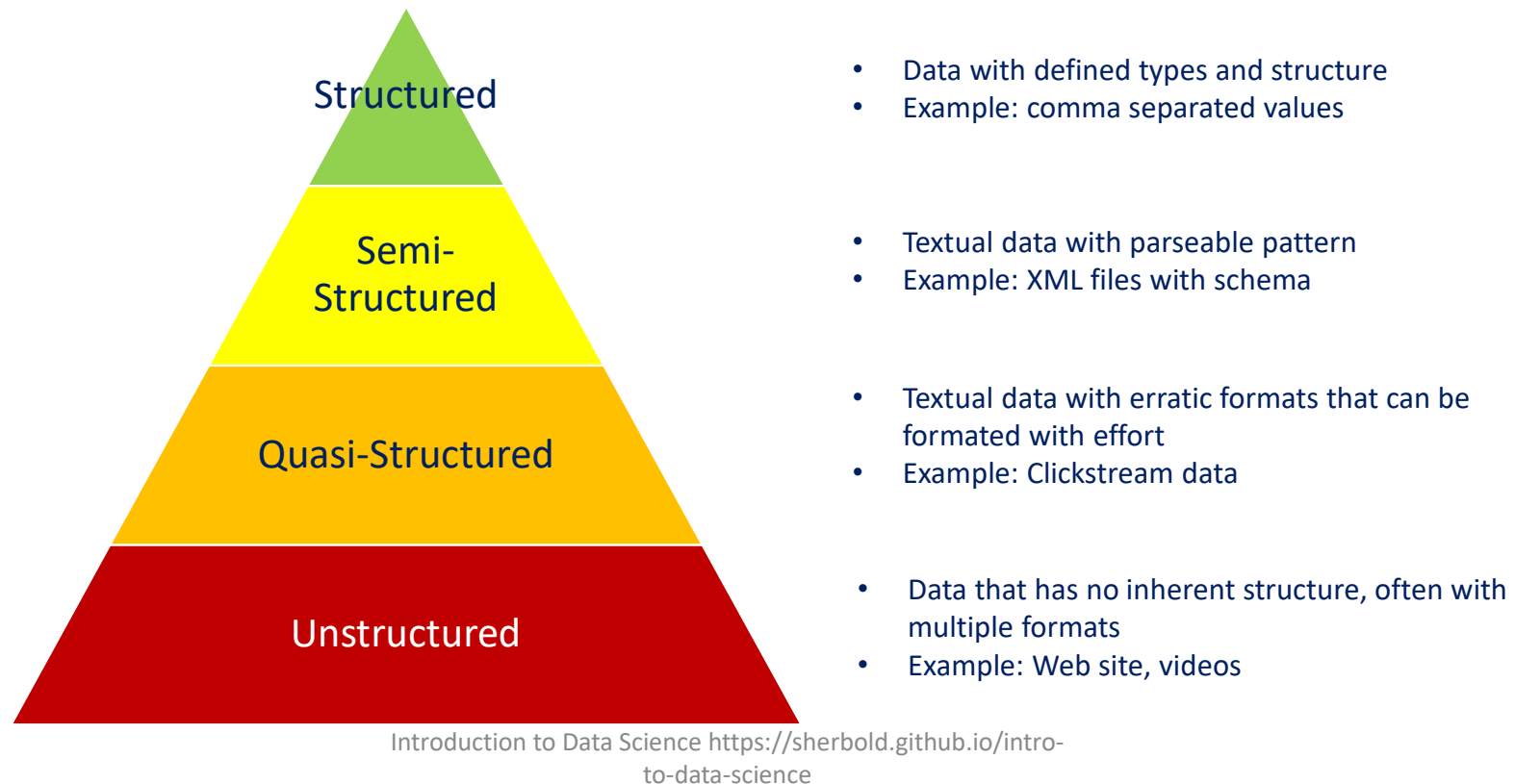
# The 3 Vs: Velocity

- Speed at which new data is created
- Speed at which data must be processed and analyzed
  - Often close to real-time



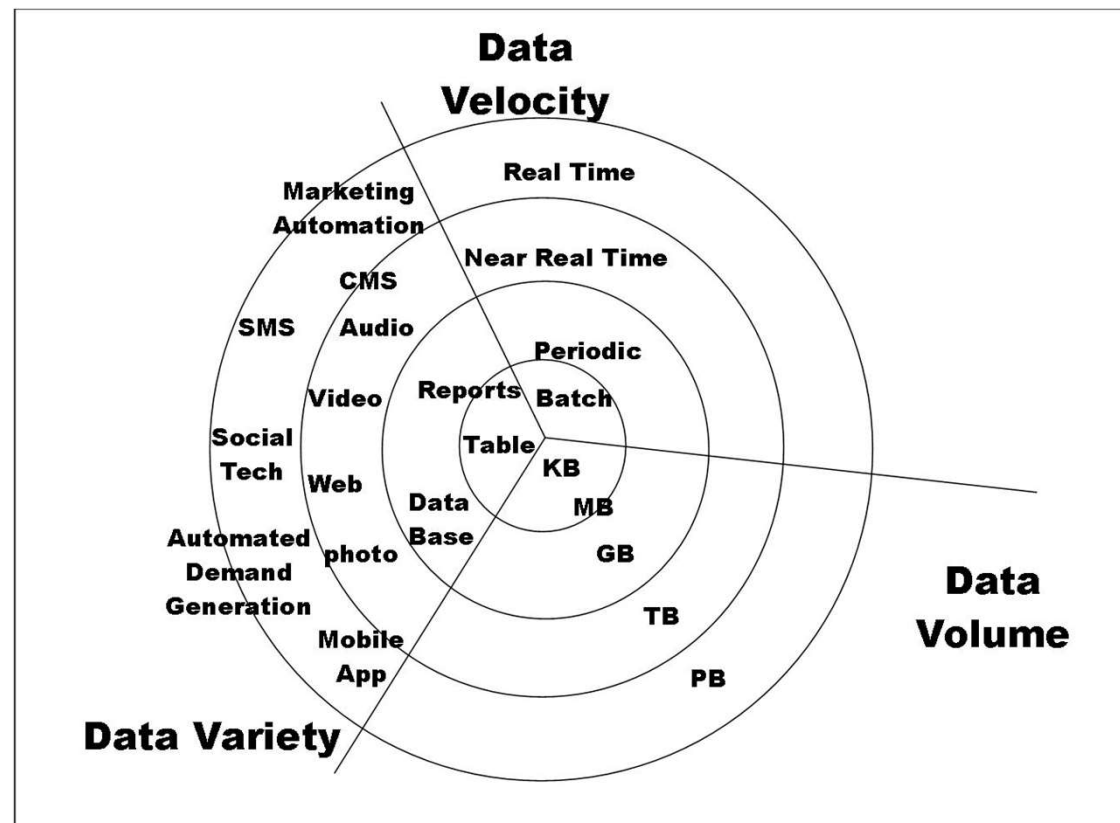
# The 3 Vs: Variety

- Diversity in data types and data sources





# Big Data



# Types of Data We Have

- Relational Data (Tables/Transaction/Legacy Data)
- Text Data (Web)
- Semi-structured Data (XML)
- Graph Data
- Social Network, Semantic Web (RDF), ...
- Streaming Data
- You can afford to scan the data once

# What To Do With These Data?

- Aggregation and Statistics
  - Data warehousing and OLAP
- Indexing, Searching, and Querying
  - Keyword based search
  - Pattern matching (XML/RDF)
- Knowledge discovery
  - Data Mining
  - Statistical Modeling