Data explosion

SOURCE: BIG DATA
DAWN E. HOLMES

What is data?

- ‘data’ - a plural word of Latin origin, ‘datum’ being the singular
- ‘Data’, owes its origins to the scientific revolution in the 18th century led by intellectual giants such as Priestley, Newton, and Lavoisier; and, by 1809, following the work of earlier mathematicians, Gauss and Laplace were laying the highly mathematical foundations for modern statistical methodology

- 1854: cholera outbreak in Broad Street, London
- Allowing physician John Snow to chart the outbreak
- data on the birth and death rate, the frequency of various diseases, and statistics on income and crime is all now collected, which was not the case prior to the 19th century
- 1870: Census data became too large to handle
- 1890: Herman Hollerith's punched cards tabulator
- Reduced processing time from 8 years to 1 year

Data in the digital age

- On paper, before 1980
- 1989-WWW
- The explosion happened !

- Structured data, of the kind written by hand and kept in notebooks or in filing cabinets, is now stored electronically on spreadsheets or databases, and consists of spreadsheet-style tables with rows and columns
- Unstructured data is not so easily categorized and includes photos, videos, tweets, and word-processing documents
- Approximately 80 per cent of the world's data is unstructured in the form of text, photos, and images
- Semi-structured data

- The term ‘data explosion’, refers to the increasingly vast amounts of structured, unstructured, and semi-structured data being generated minute by minute
- According to a recent worldwide study conducted by IBM, about 2.5 *exabytes (Eb) of data are generated every day*.
- *One* Eb is 1018 (1 followed by eighteen 0s) bytes (or a million *terabytes (Tb))*
- Source?

Search engine data

- 2012: Over 3.5 billion searches made per day on Google
  - Logs
  - Cookies
- 2016: Facebook -1.71 billion active users/month
  - 1.5 *petabytes (Pb; es (Pb; or 1,000 Tb) of Web log data every day*
- Youtube: Over a billion users

Healthcare

- Increasingly computerized
- Through wearable or implantable sensors for health monitoring
- Personal fitness trackers of varying complexity which output ever more categories of data
- Measurements to include sleep tracking, arterial oxygen saturation rate and much more

- Researchers study the genes and sequencing genomes of a variety of species
- Structure of the deoxyribonucleic acid molecule (DNA), was first described as a double-helix by James Watson and Francis Crick in 1953
- One of the most highly publicized research projects in recent years is international human genome project, which determines the sequence, or exact order, of the three billion base-pairs that comprise human DNA

Real-time data

- Systems where response time is crucial
  - Data must be processed in a timely manner
- Autonomous car will generate on average 30 Tb of data daily much of which will have to be processed almost instantly
- Airplanes
- Global positioning systems
- Satellites

Astronomical data

- 2014: International Data Corporation report estimated that, by 2020, the digital universe will be 44 trillion *gigabytes* (10x size in 2013)
- Data produced by telescopes
  - Very Large Telescope in Chile : optical telescope, which actually consists of four telescopes, each producing huge amounts of data—15 Tb per night, every night in total.
  - Large Synoptic Survey, a ten-year project repeatedly producing maps of the night sky, creating an estimated grand total of 60 Pb

What use is all this data?

- Almost impossible to take part in everyday activities and avoid having some personal data collected electronically
  - Supermarket check-outs collect data on what we buy
  - Airlines collect information about our travel arrangements when we purchase a ticket
  - Banks collect our financial data
- Big data is used extensively in commerce and medicine and has applications in law, sociology, marketing, public health, and all areas of natural science
- Data in all its forms has the potential to provide a wealth of useful information if we can develop ways to extract it

- Determining which patterns are important is key to the success of big data analytics
- The big data revolution has given us smart cars and home monitoring
- Ultimate aim of working with big data is to extract useful information
- Ability to gather data electronically resulted in the emergence of data science, bringing together the disciplines of statistics and computer science
  - in order to analyse these large quantities of data to discover new knowledge in interdisciplinary areas of application.

- Satellite and ground sensors are used to monitor seismic activity.
  - Aim is to determine approximately where big earthquakes are *likely* to occur in the long-term.
- 2016: US Geological Survey (USGS) estimated that 'there is a 76% probability that a magnitude 7 earthquake will occur within the next 30 years in northern California'
  - Probabilities such as these help focus resources on measures
  - Ensuring that buildings are better able to withstand earthquakes
  - Having disaster management programmes in place

Term	Meaning
Bit	1 binary digit: 0 or 1
Byte	8 bits
Kilobyte (Kb)	1,000 bytes
Megabyte (Mb)	1,000 kilobytes
Gigabyte (Gb)	1,000 megabytes
Terabyte (Tb)	1,000 gigabytes
Petabyte (Pb)	1,000 terabytes
Exabyte (Eb)	1,000 petabytes
Zettabyte (Zb)	1,000 exabytes
Yottabyte (Yb)	1,000 zettabytes

Overview of Big Data

- Marriage of computers and telecommunications led to a boom of data.

- Social media
- Mass migration to digital technologies

- Data is the new oil
- Information is the new asset

- *“Hiding within those mounds of data is knowledge that could change the life of a patient, or change the world.”*

-Atul Butte, Stanford

Why detailed analysis was not possible before BigData era?

- Volume of data
- Granularity
- Computational resources
- Algorithms

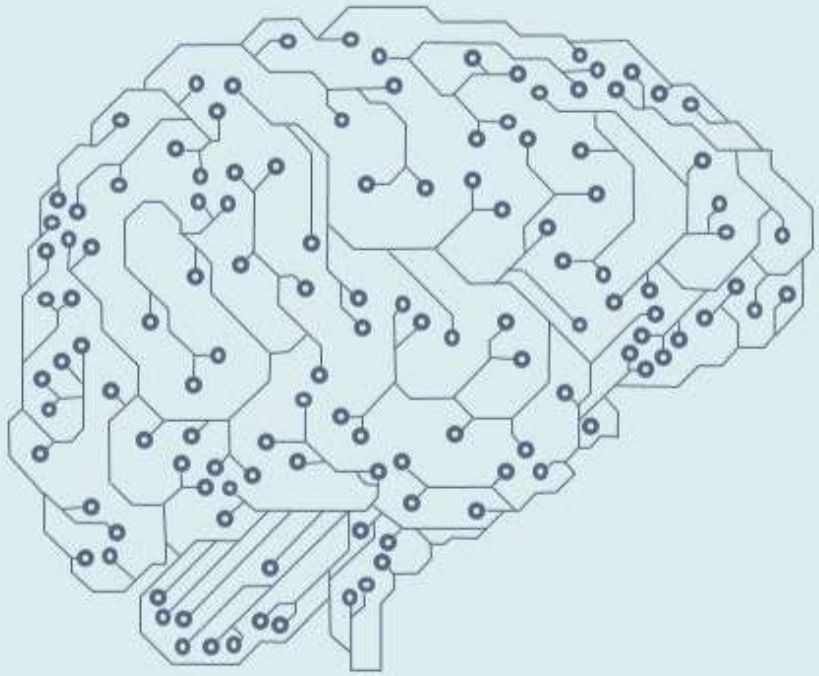
- BigData term evolved around 1990s and gained momentum post 2000s

- “ERA”

- Big Data era is described by rapidly expanding volumes of data, far beyond what people imagined would ever occur.

- 1 ZB = 1 billion TB
- Shift in behaviour of companies, businesses
- Simultaneous development, creation and maturity of other technologies to store, manipulate and analyze the data

- Data is not magic – its a valuable raw material that can be refined and distilled into valuable specific insights. These insights are then converted to information that eventually creates knowledge.

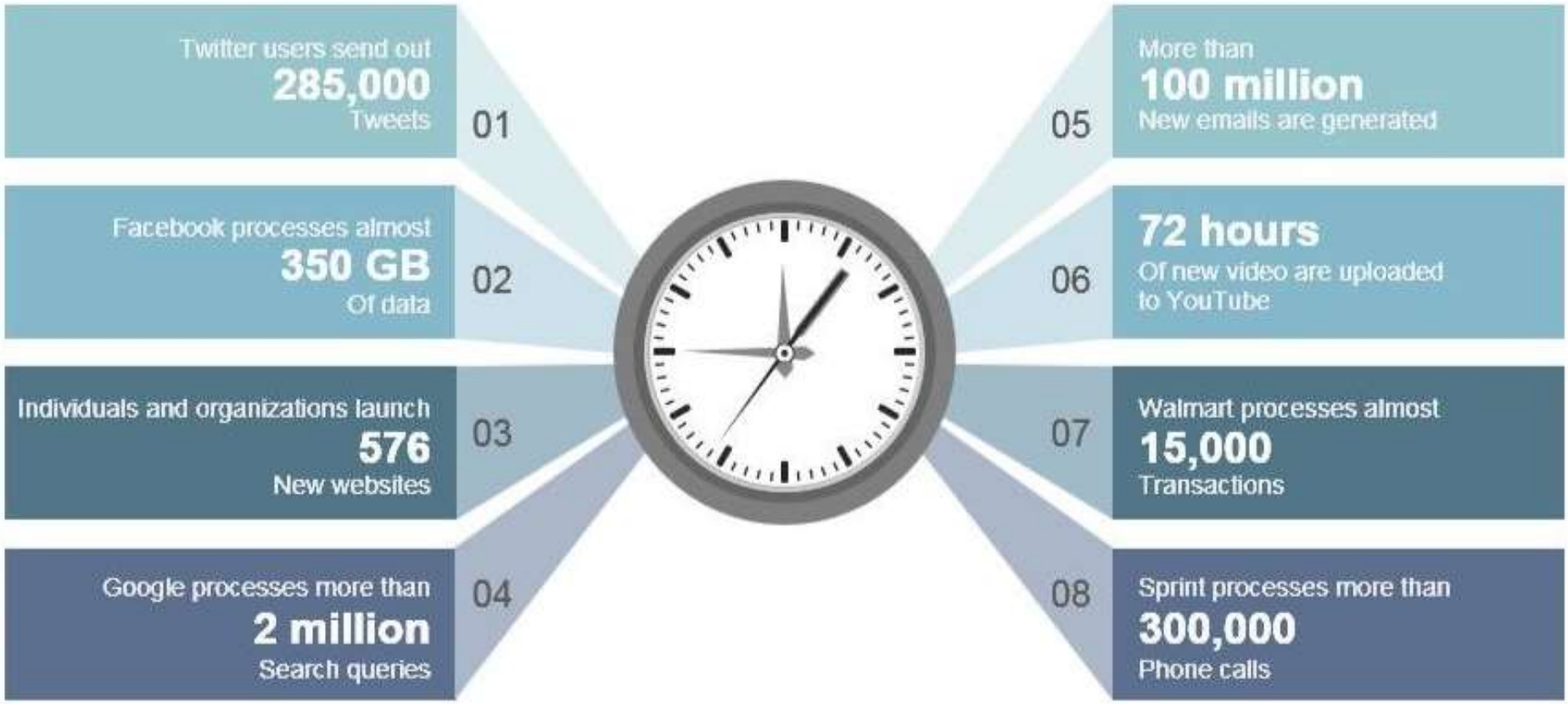


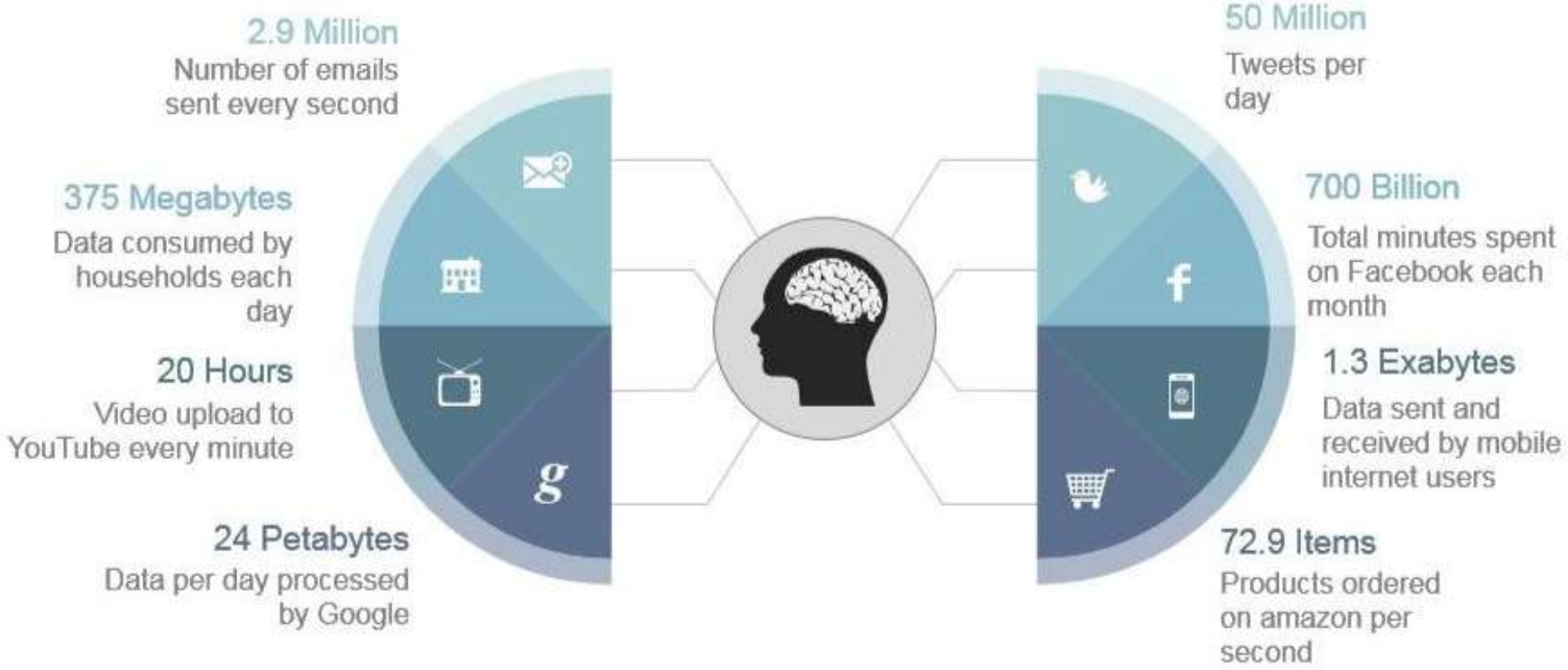
Big Data

Big data is characterized by large volumes of different types of data (e.g. Social, web, transaction, etc.) That builds very quickly.

It exceeds the reach of commonly used hardware environments and software tools to capture, manage and process in a timely manner for its users.

How big is Big Data?




2.9 Million
Number of emails sent every second
375 Megabytes
Data consumed by households each day
20 Hours
Video upload to YouTube every minute
24 Petabytes
Data per day processed by Google
50 Million
Tweets per day
700 Billion
Total minutes spent on Facebook each month
1.3 Exabytes
Data sent and received by mobile internet users
72.9 Items
Products ordered on amazon per second

Sources of Big Data

- Images
- Media
- Database
- Locations
- Email
- Click stream
- Social network
- HTML
- Sensors

Sources of Big Data

Media

Media and communication outlets (articles, podcasts, audio, video, email, blogs)

Social

Digital material created by social media (text, photos, videos, tweets)

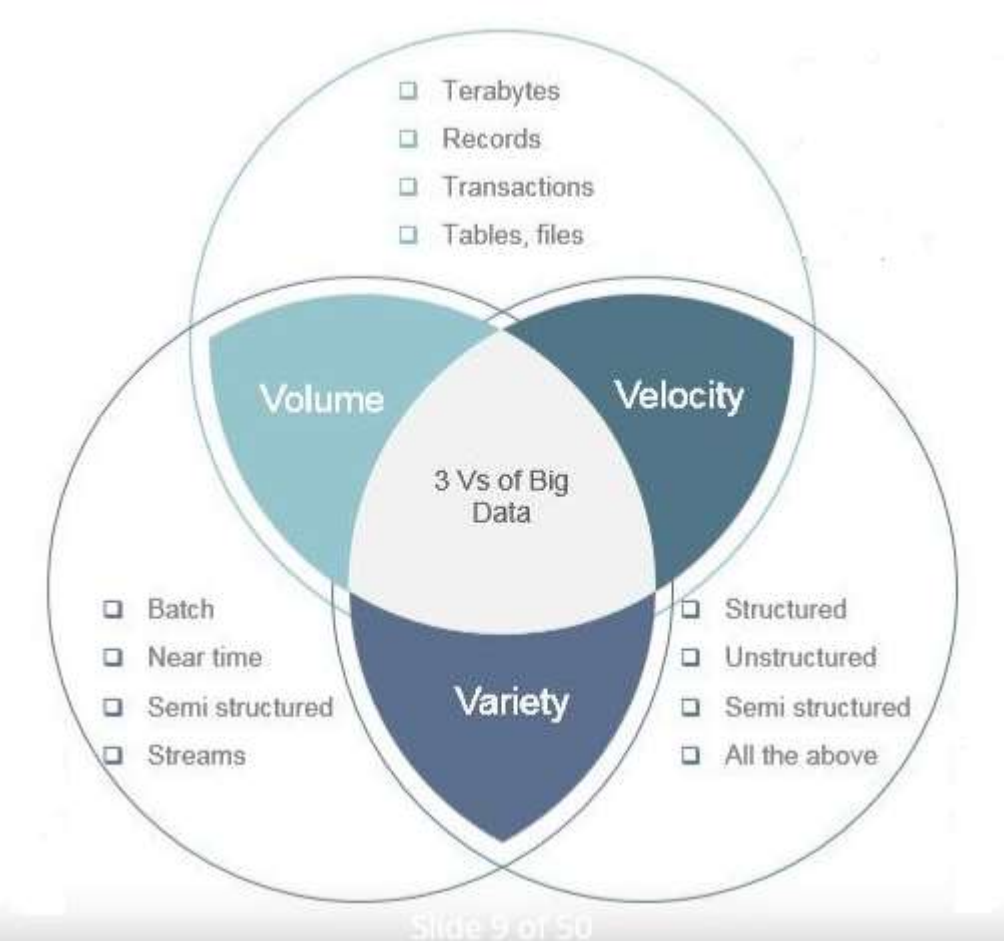
Machine

Data generated by computers and machines generally without human intervention (business process logs, sensors, phone calls)

Historical

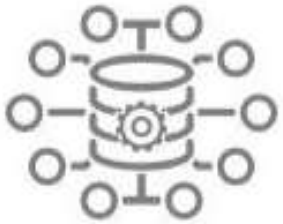
Data about our environment (weather, traffic, census) and archived documents, forms or records

V's of Big Data



Small Data and Big Data

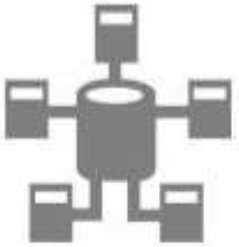
Small Data



- Low Volumes
- Batch Velocities
- Structured Varieties

Vs

Big Data



- Into Petabyte Volumes
- Real-time Velocities
- Multistructured Varieties

Objectives of Big Data

- Be more responsive to the market
- Analyzing consumer behaviour
- Combining multiple data sources
- Improving customer service
- Generate additional revenue

Big Data Technologies

- Machine learning
- Genetic algorithm
- Simulation
- Signal processing
- Natural Language Processing
- Crowd sourcing
- Data fusion
- Data integration

Impact of Big Data

Sports predictions

Big data was used in 2012 to famously predict the US would win 108 medals in the summer Olympics. They won 104.

Advanced healthcare

Big data is used to predict where outbreaks of potentially epidemic viruses will occur.

Easier commutes

Big data streamlines traffic, using data from GPS systems. Traffic lights now react to traffic conditions, weather & accidents.

Presidential campaigns

Statistician Nate silver used big data to predict the outcome of the 2012 presidential election.

Smartphones

Smartphone users utilize big data when asking for directions & verbally asking their phone questions.

Personalized advertising

Big data is used to recommend purchases & personalized ads more relevant.

Healthcare

It allow us to find new cures and better understand and predict disease patterns. This leads to saving more lives.

Science

It creates new possibilities and ways to conduct research which would otherwise be impossible, helping us to make new discoveries.

Security

Police forces use big data tools to predict criminal activities, conduct investigations and ultimately to catch criminals faster.

Business

It helps us to improve and optimize the ways we do business by making data-driven decisions.

Benefits

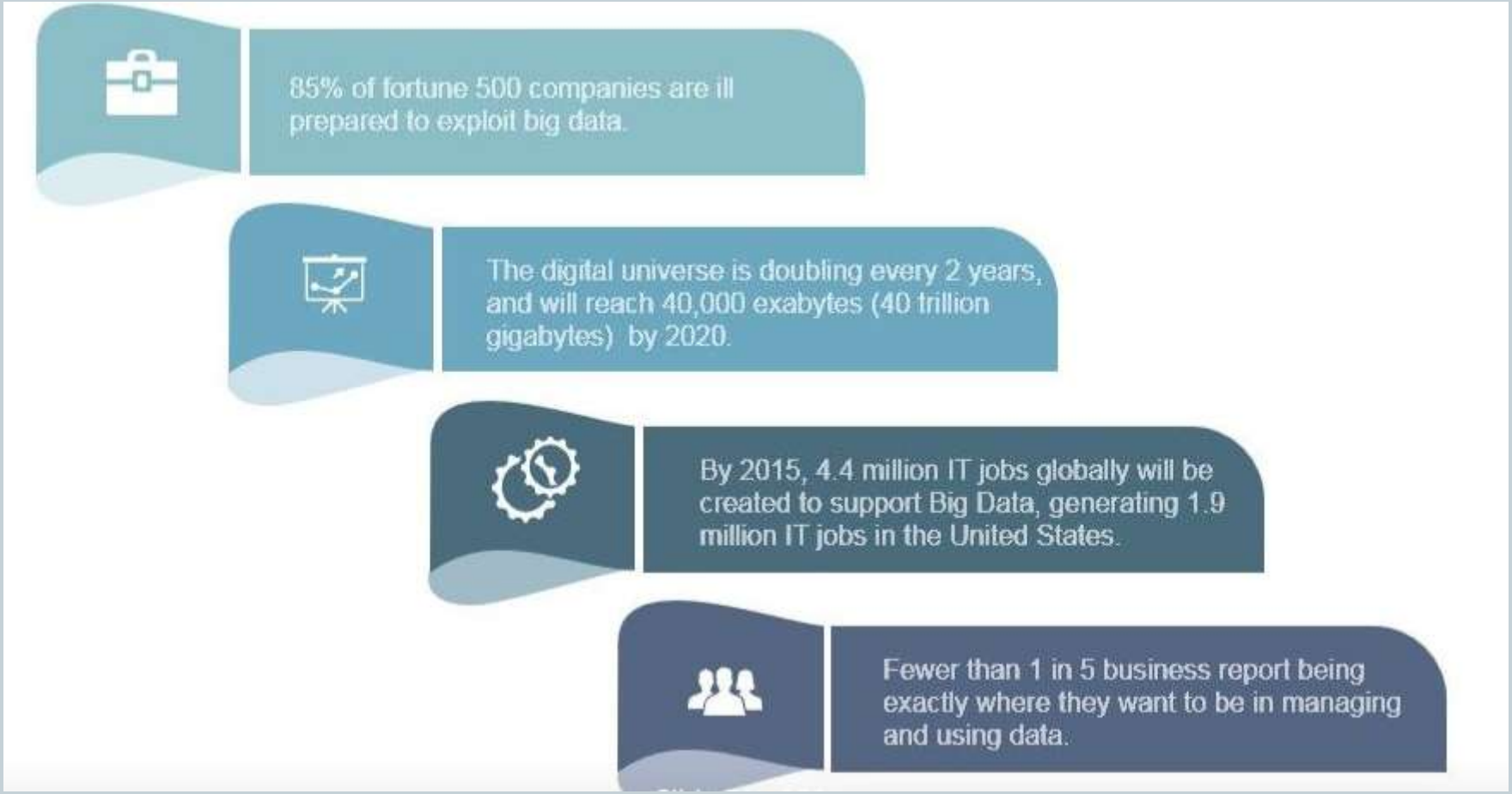
Increased efficiency

Better business decision making

Improved customer experience and engagement

Achieved financial savings

Future of Big Data



30% of companies looking for skills in advanced analytics/ predictive analytics in the next 12 months.

55% of companies looking for skills in advanced analytics/ predictive analytics in the next 12 months.

75% of companies looking for skills in advanced analytics/ predictive analytics in the next 12 months.

Opportunities and Challenges

- Lack of sufficiently skilled IT staff
- Cost of technology
- Managing data quality
- Data integration

Types of Digital Data

BOOK: FUNDAMENTALS OF BUSINESS ANALYTICS (WILEY PUB)

BY R N PRASAD AND SEEMA ACHARYA

Types of data

- Structured
- Unstructured
- Semi-structured
  - Meta Data is also a type of data

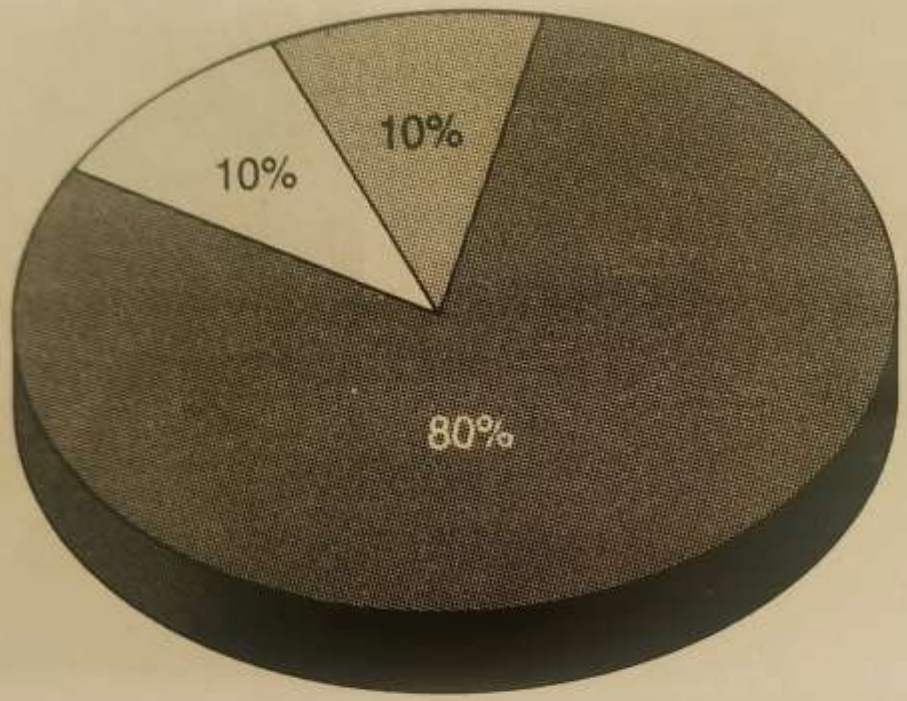


Figure 2.1 Distribution of digital data in three forms.

Structured data

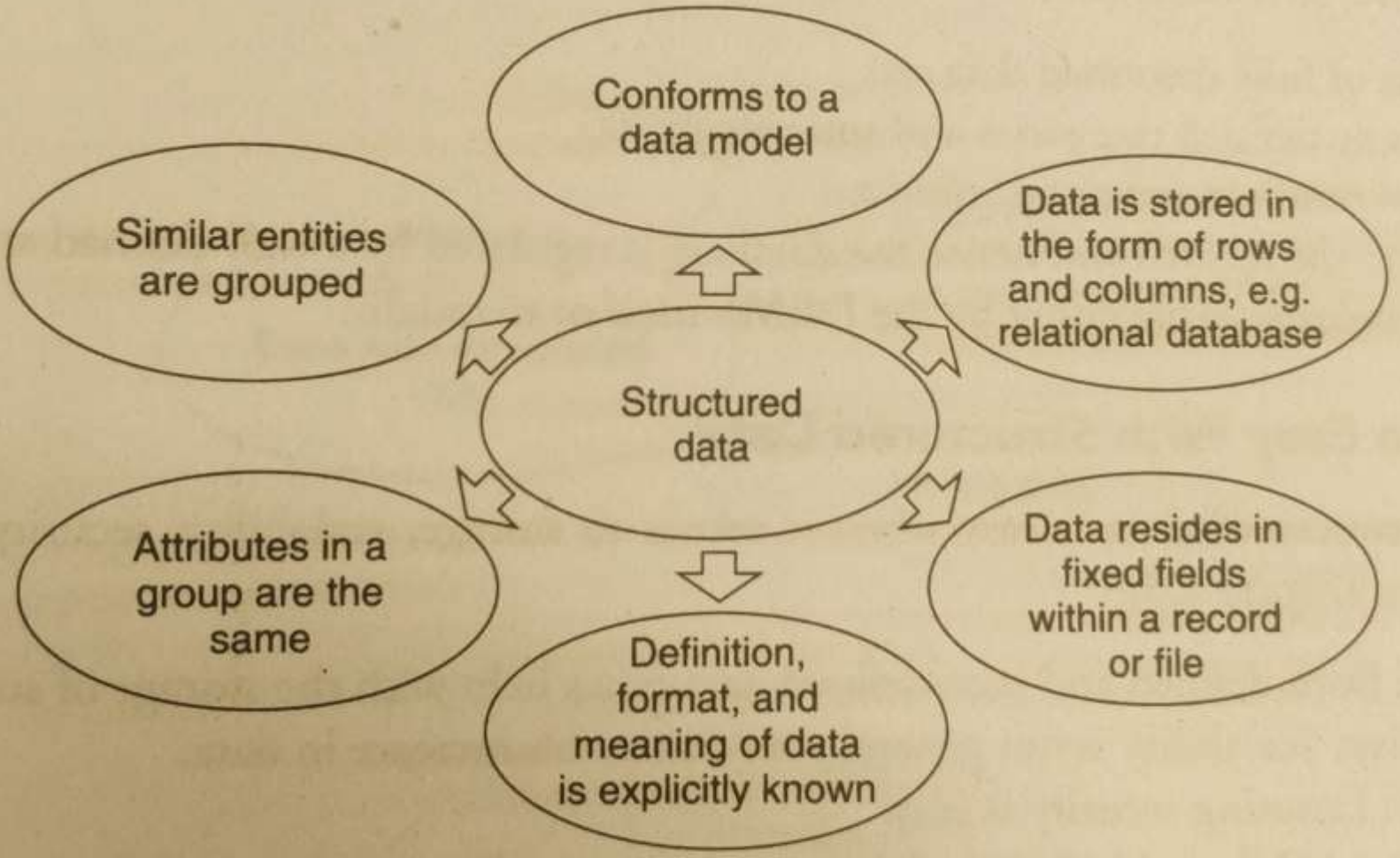


Figure 2.3 Characteristics of structured data.

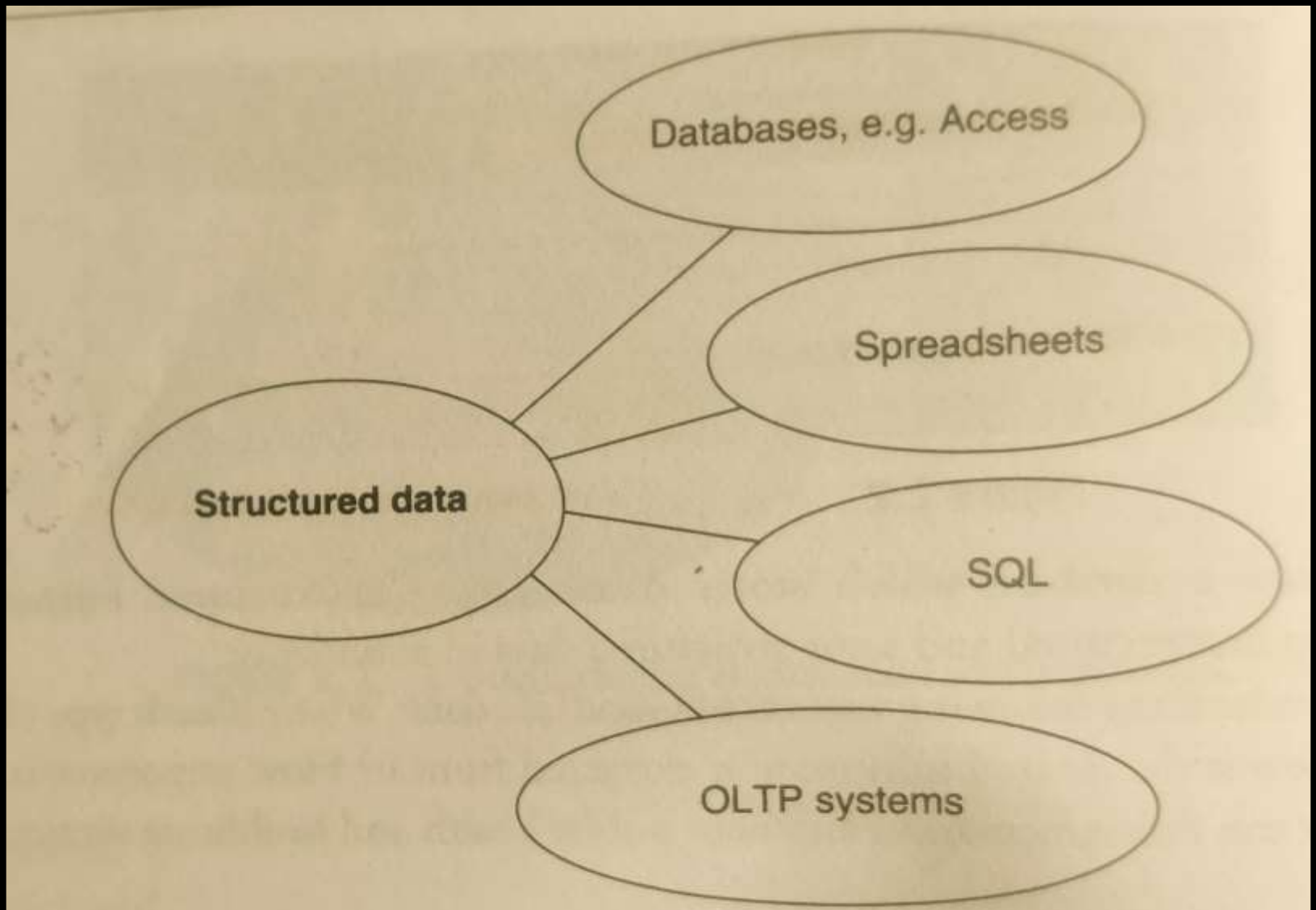


Figure 2.4 Sources of structured data.

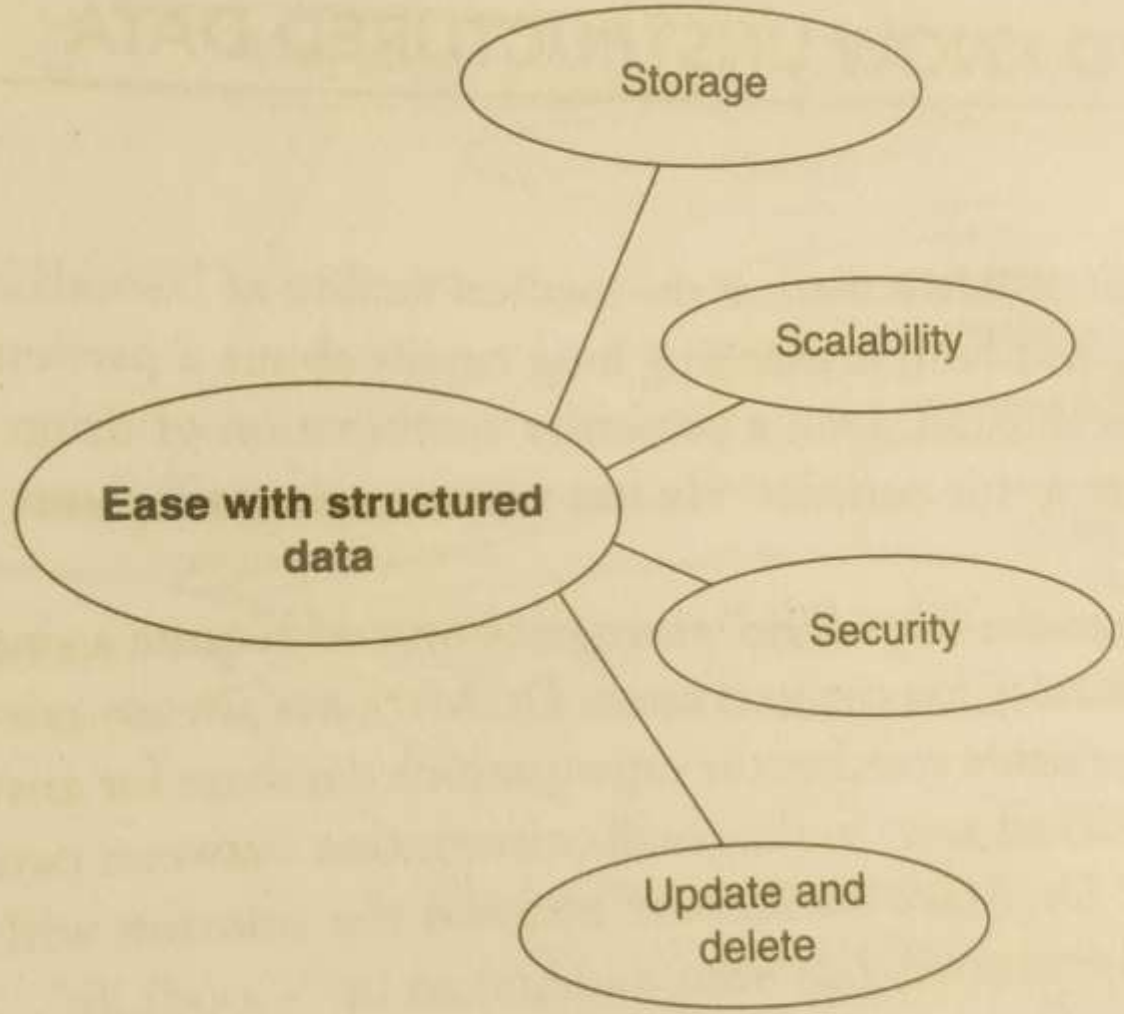


Figure 2.5 Ease with structured data.

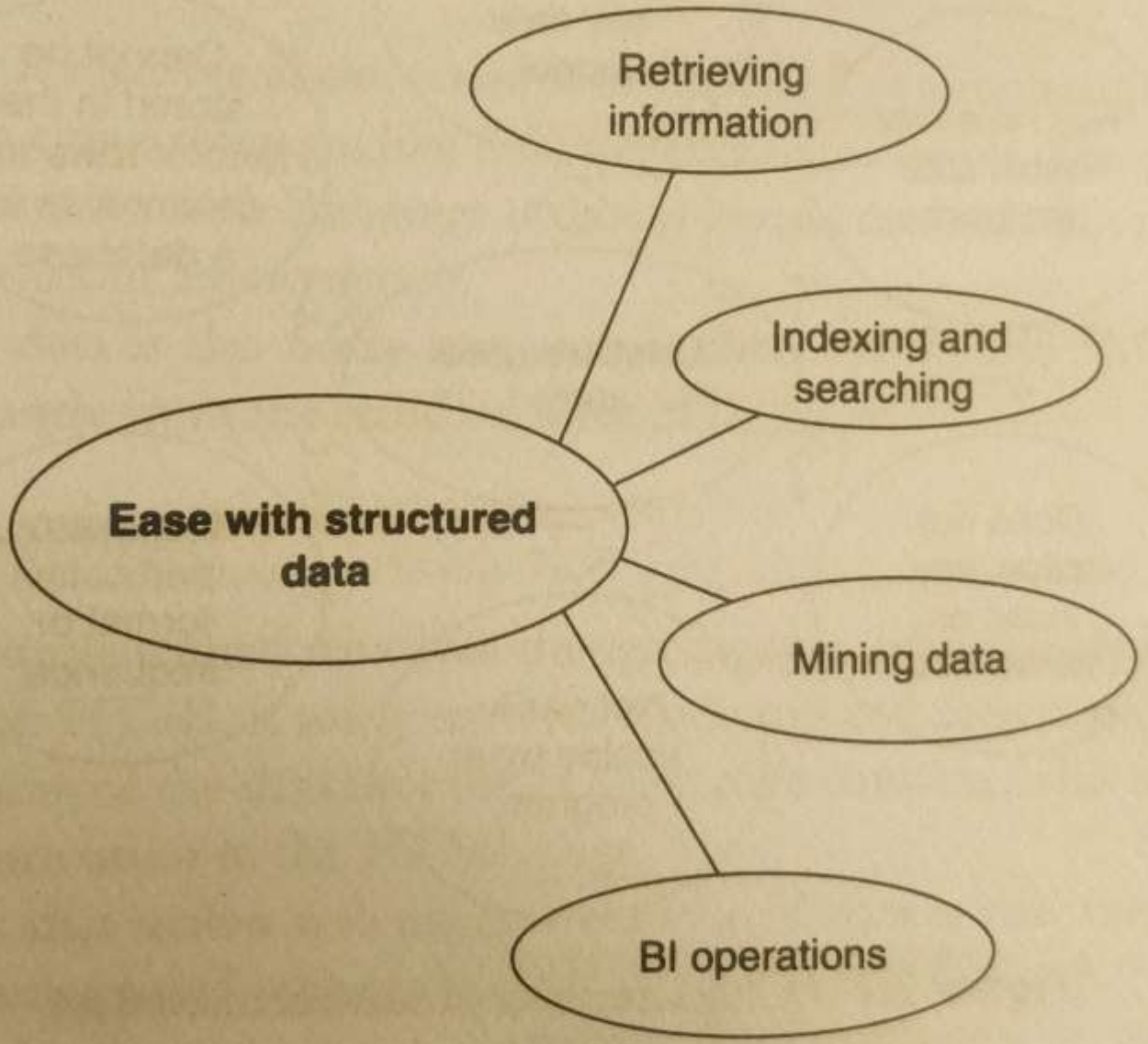


Figure 2.6 Ease of retrieval of structured data.

Unstructured data

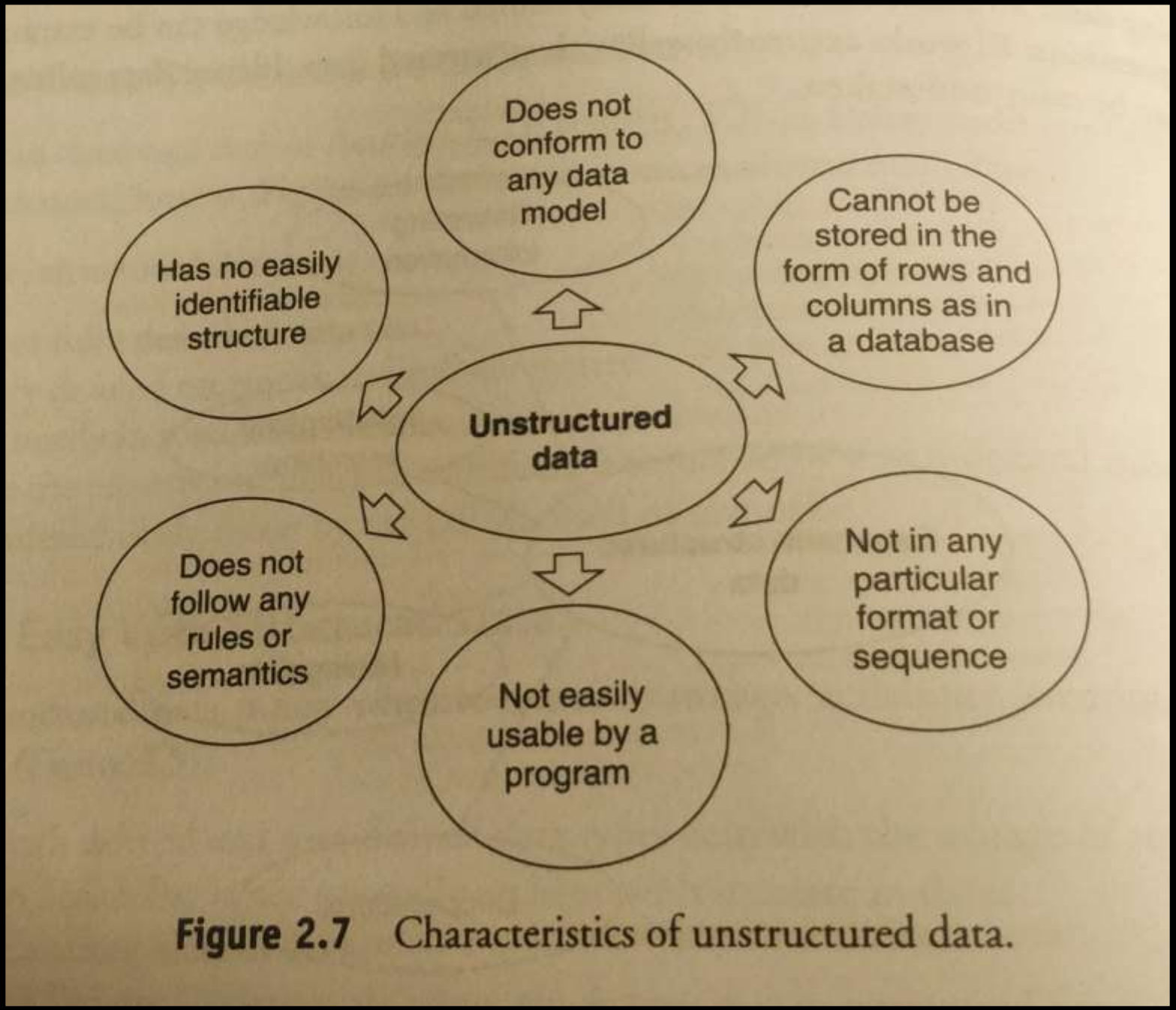


Figure 2.7 Characteristics of unstructured data.

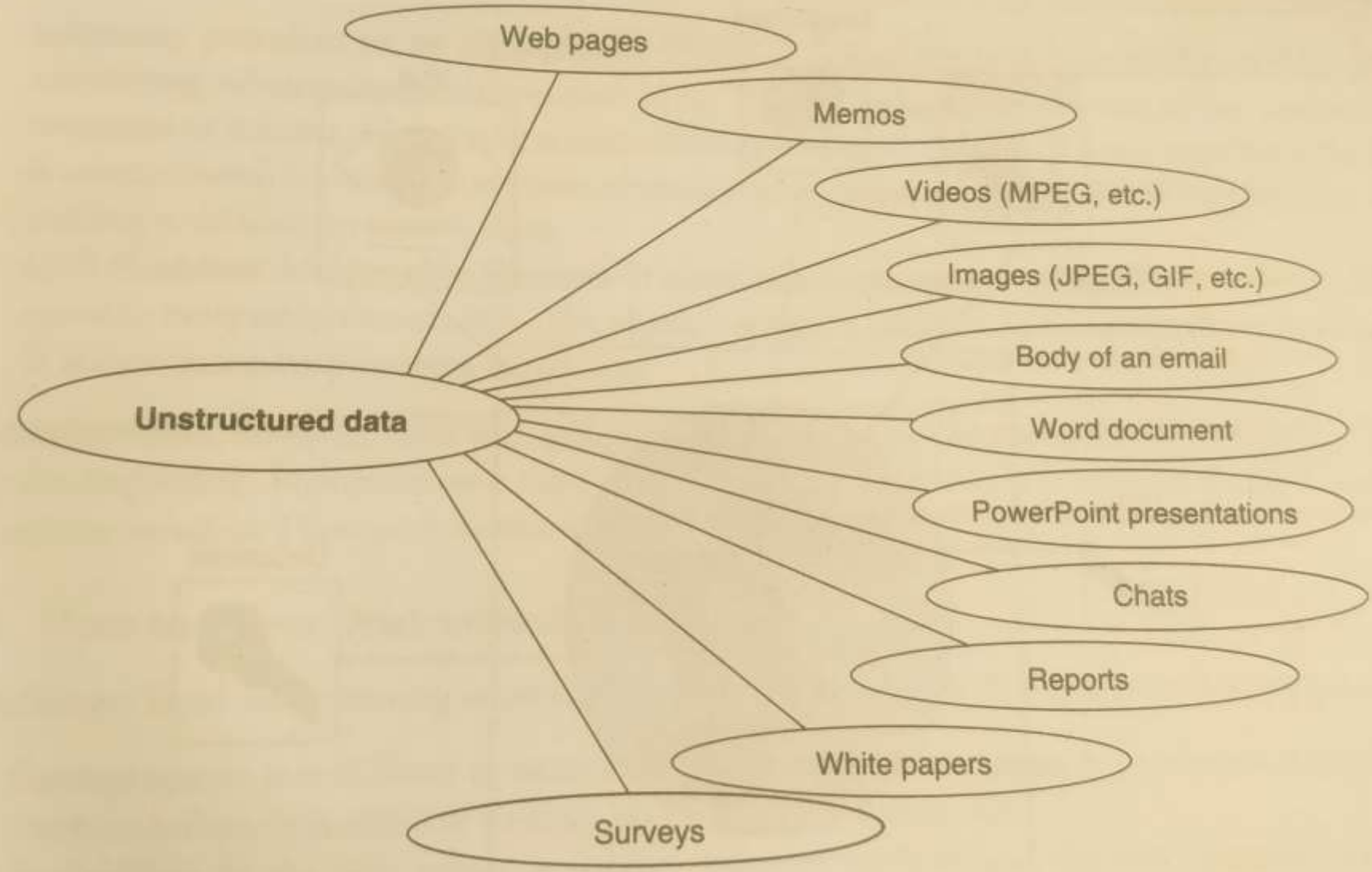


Figure 2.8 Sources of unstructured data.

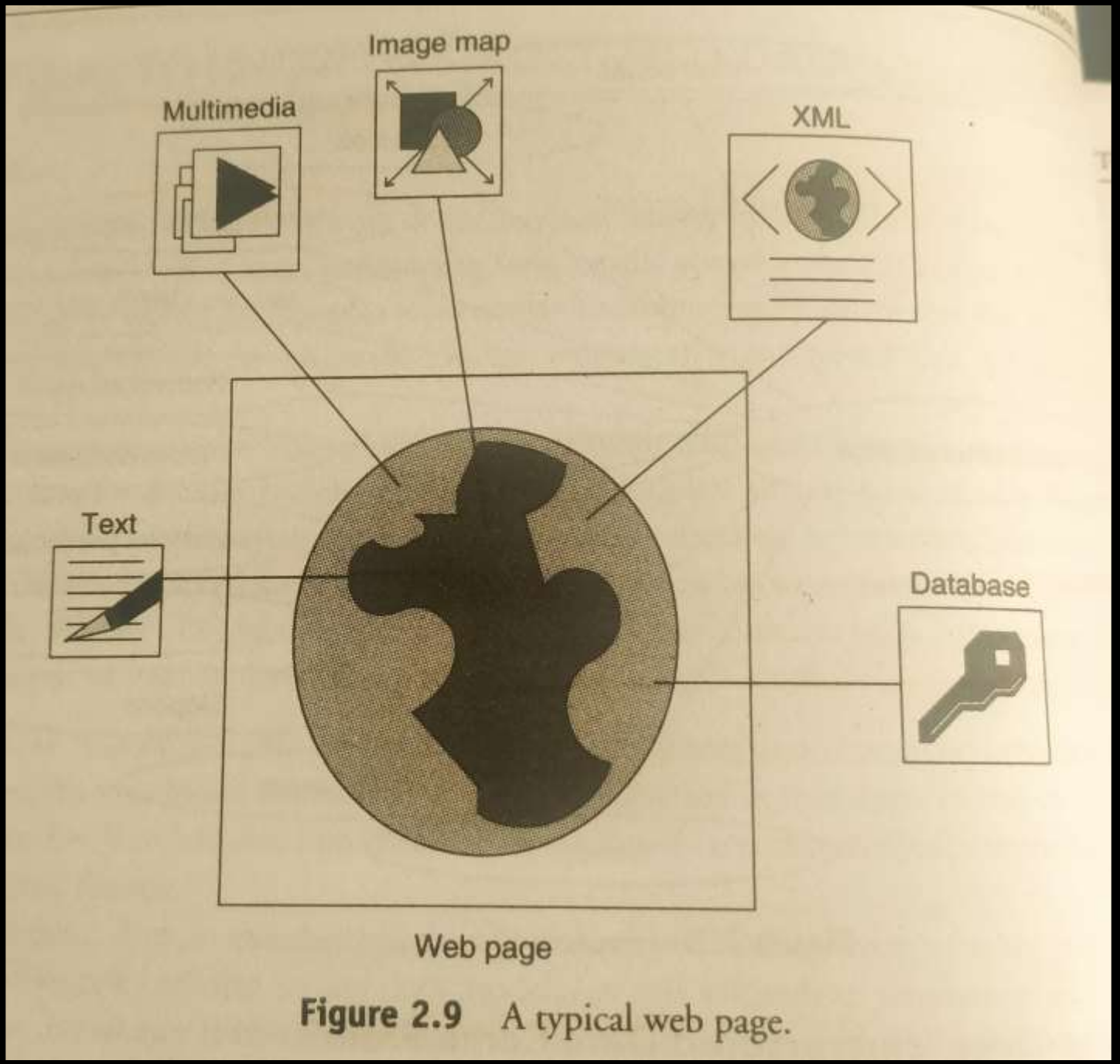


Figure 2.9 A typical web page.

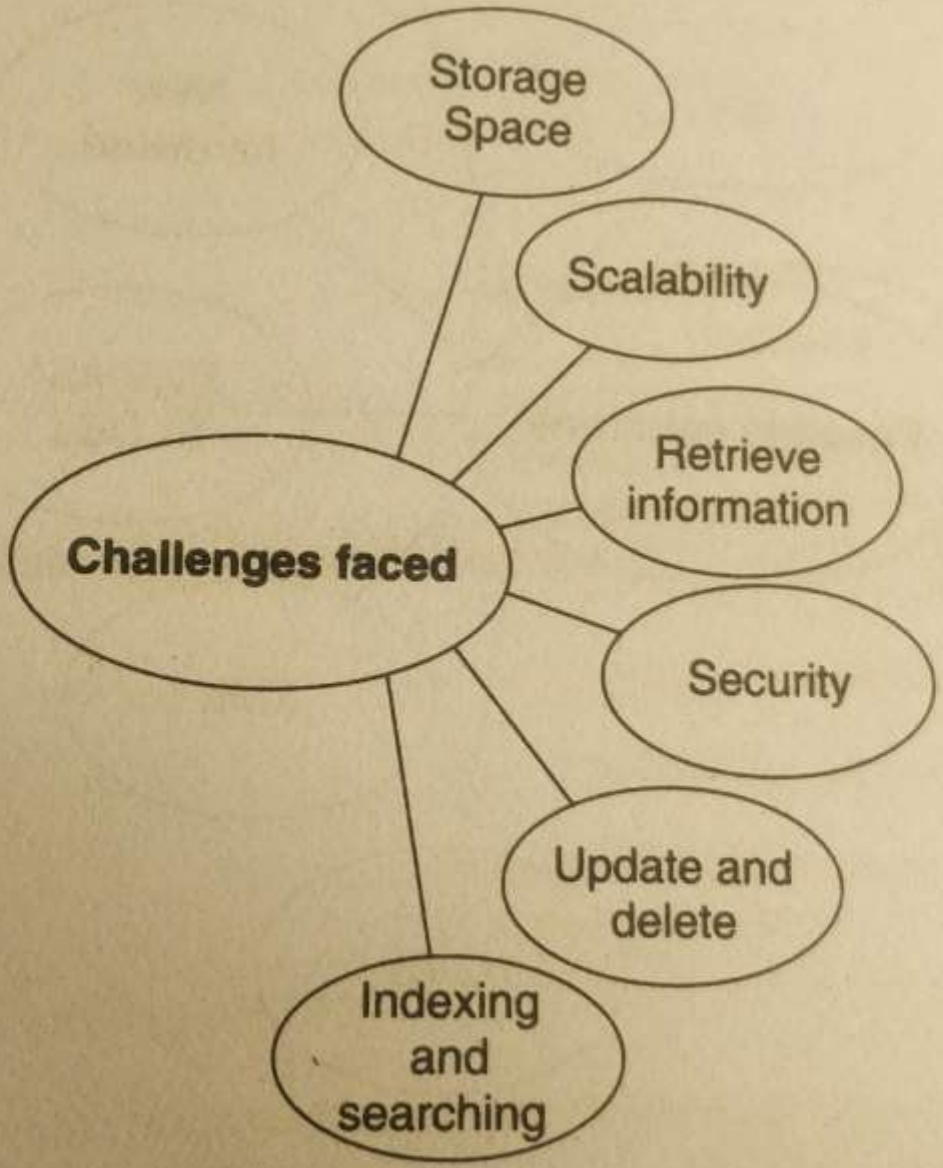


Figure 2.10 Challenges faced while storing unstructured data.

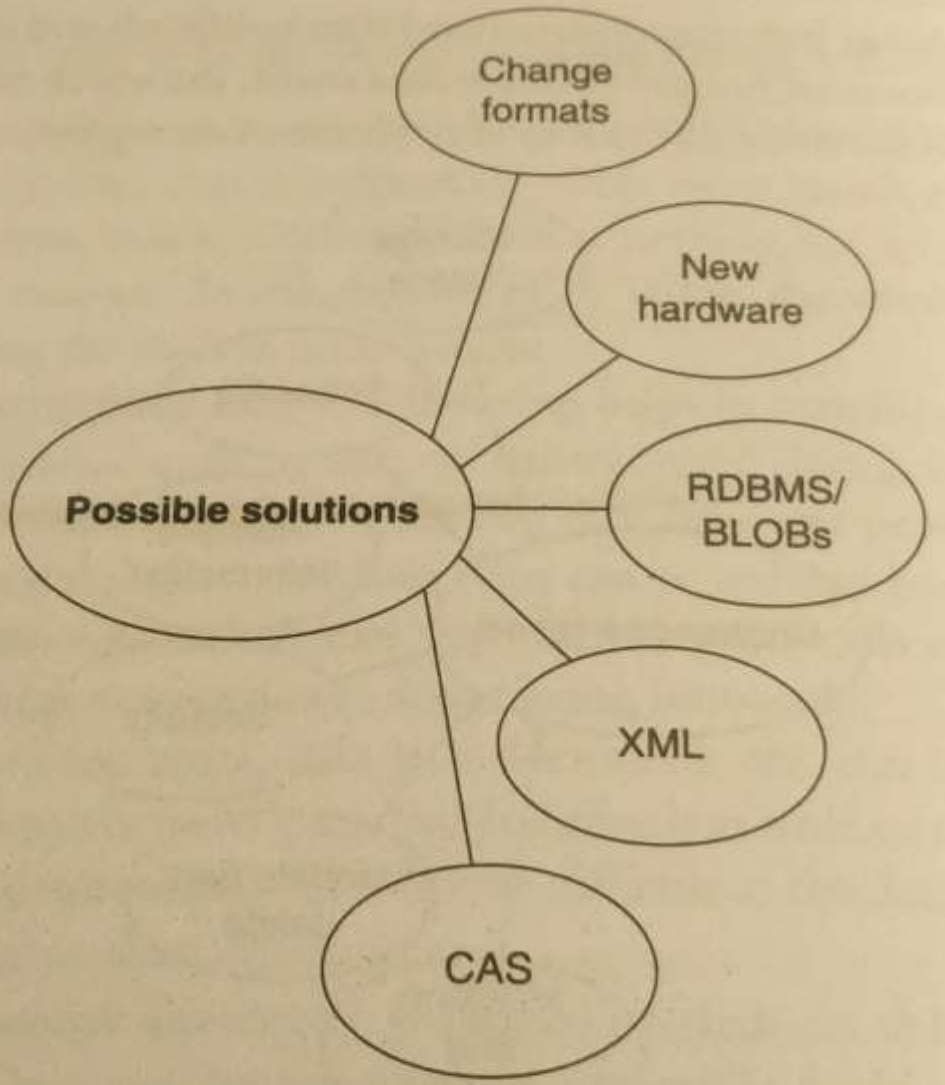


Figure 2.11 Possible solutions for storing unstructured data.

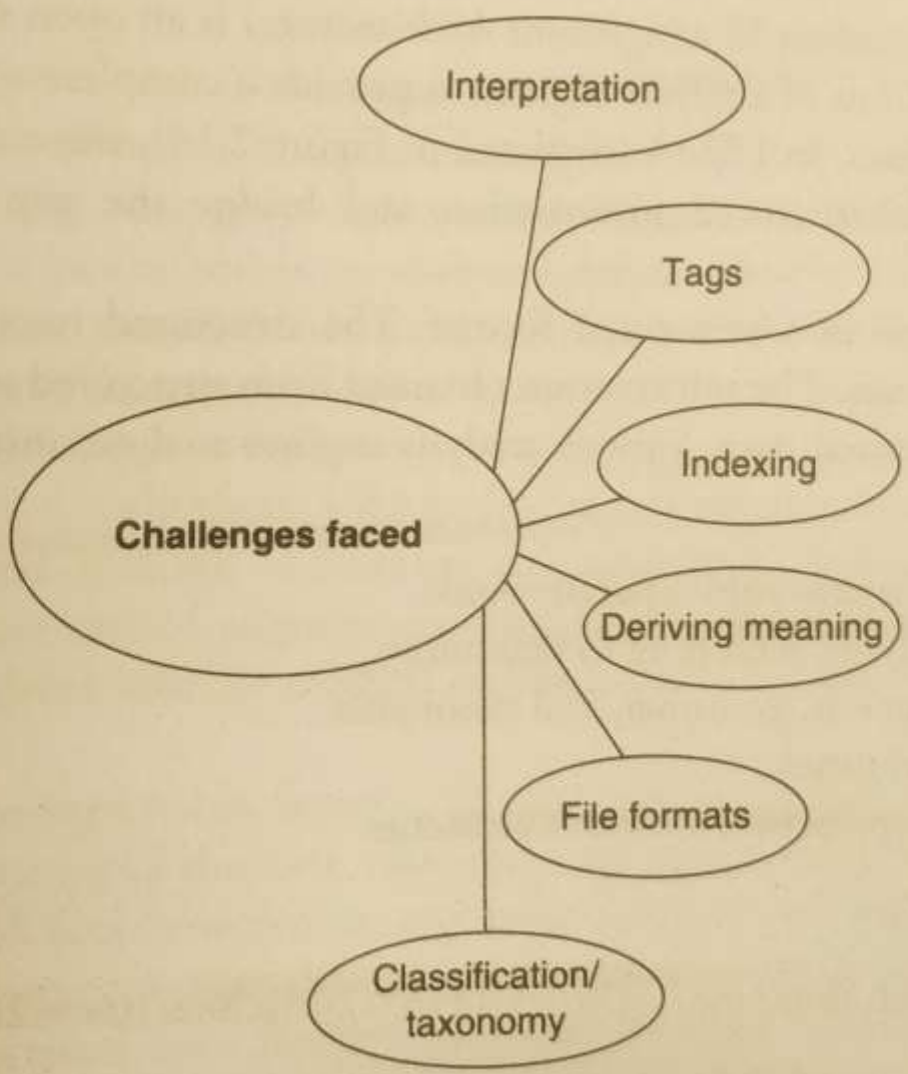


Figure 2.12 Challenges faced while extracting information from stored unstructured data.

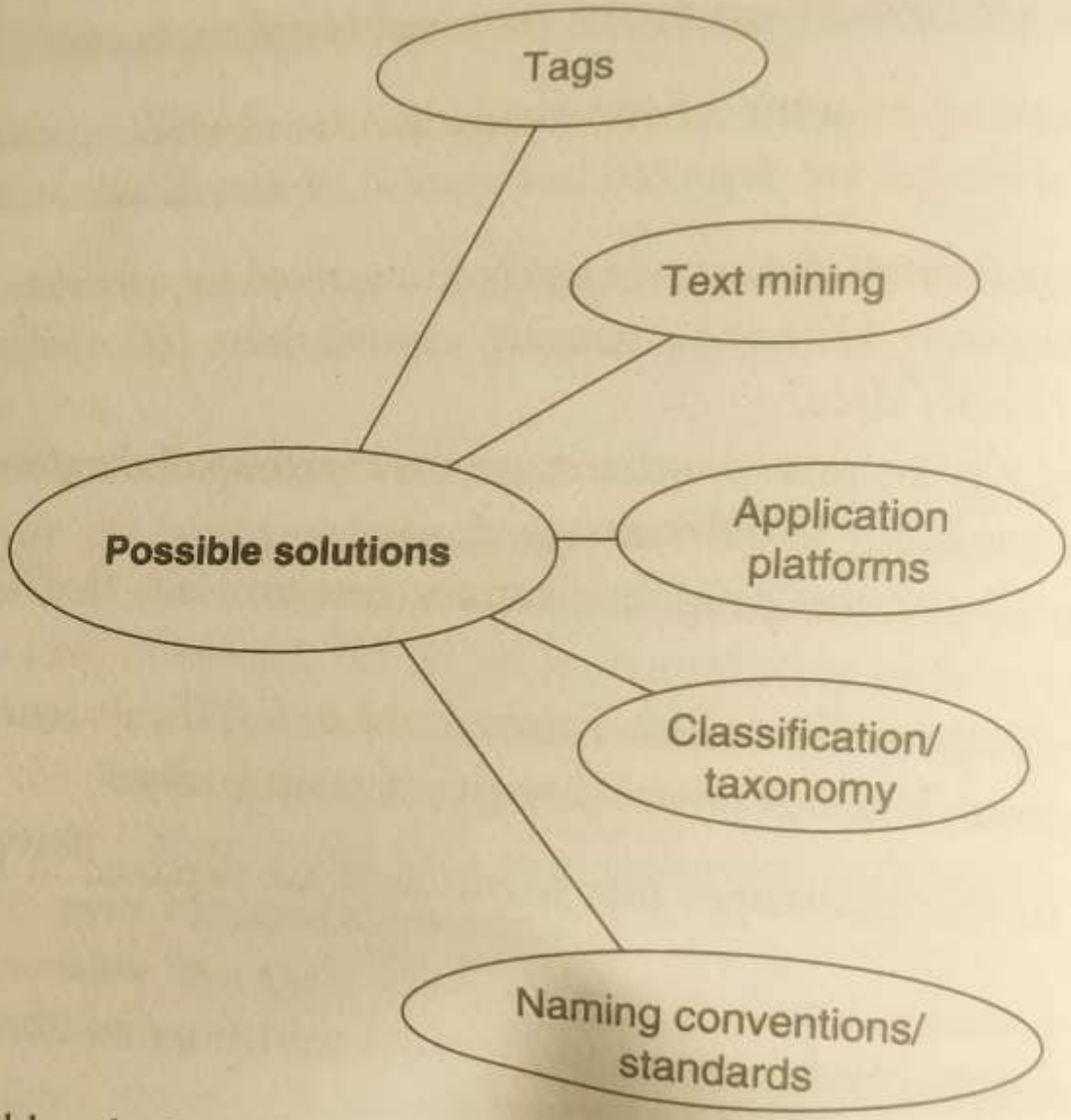


Figure 2.13 Possible solutions for extracting information from stored unstructured data.

Semi structured data

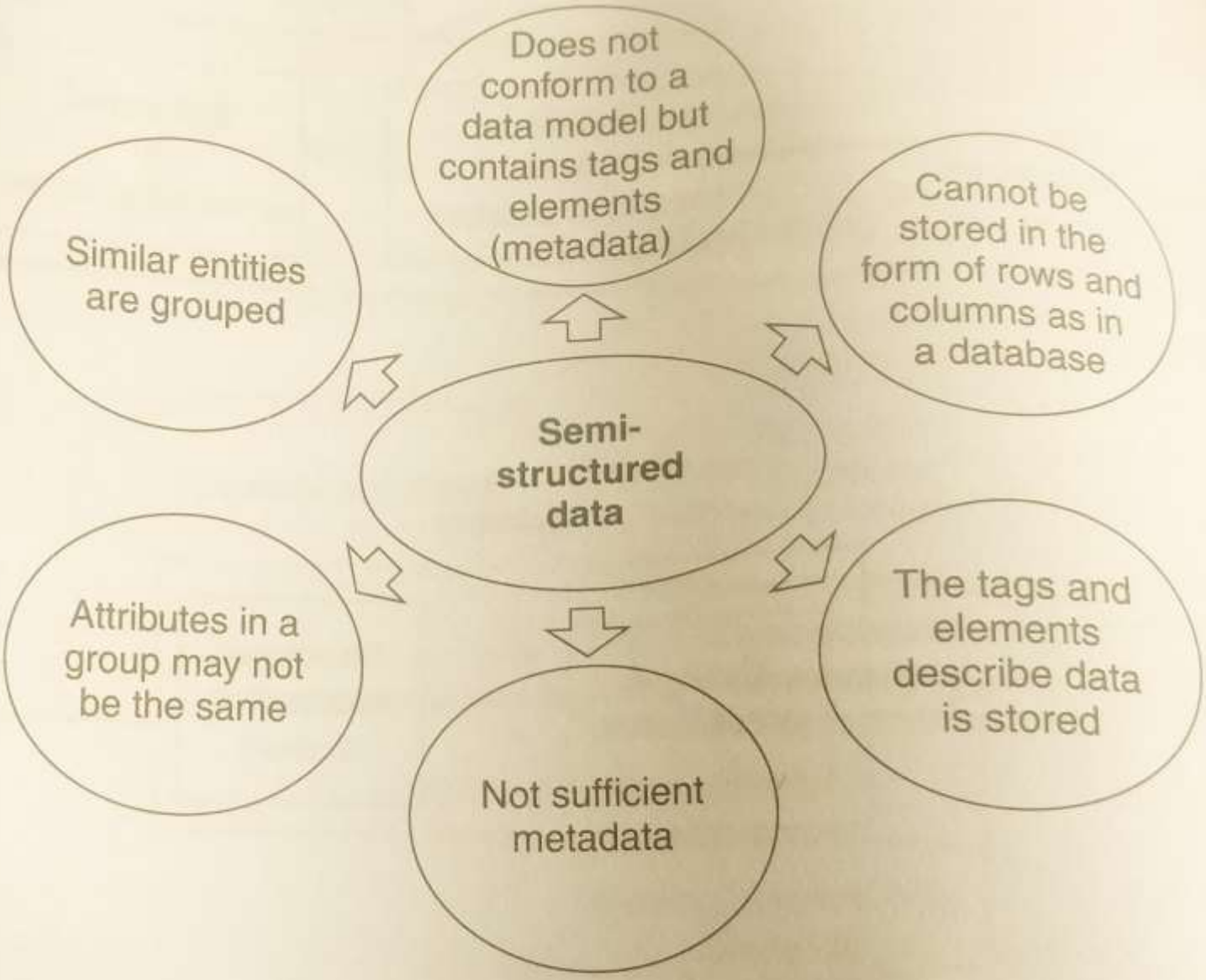


Figure 2.15 Characteristics of semi-structured data.

attributes.

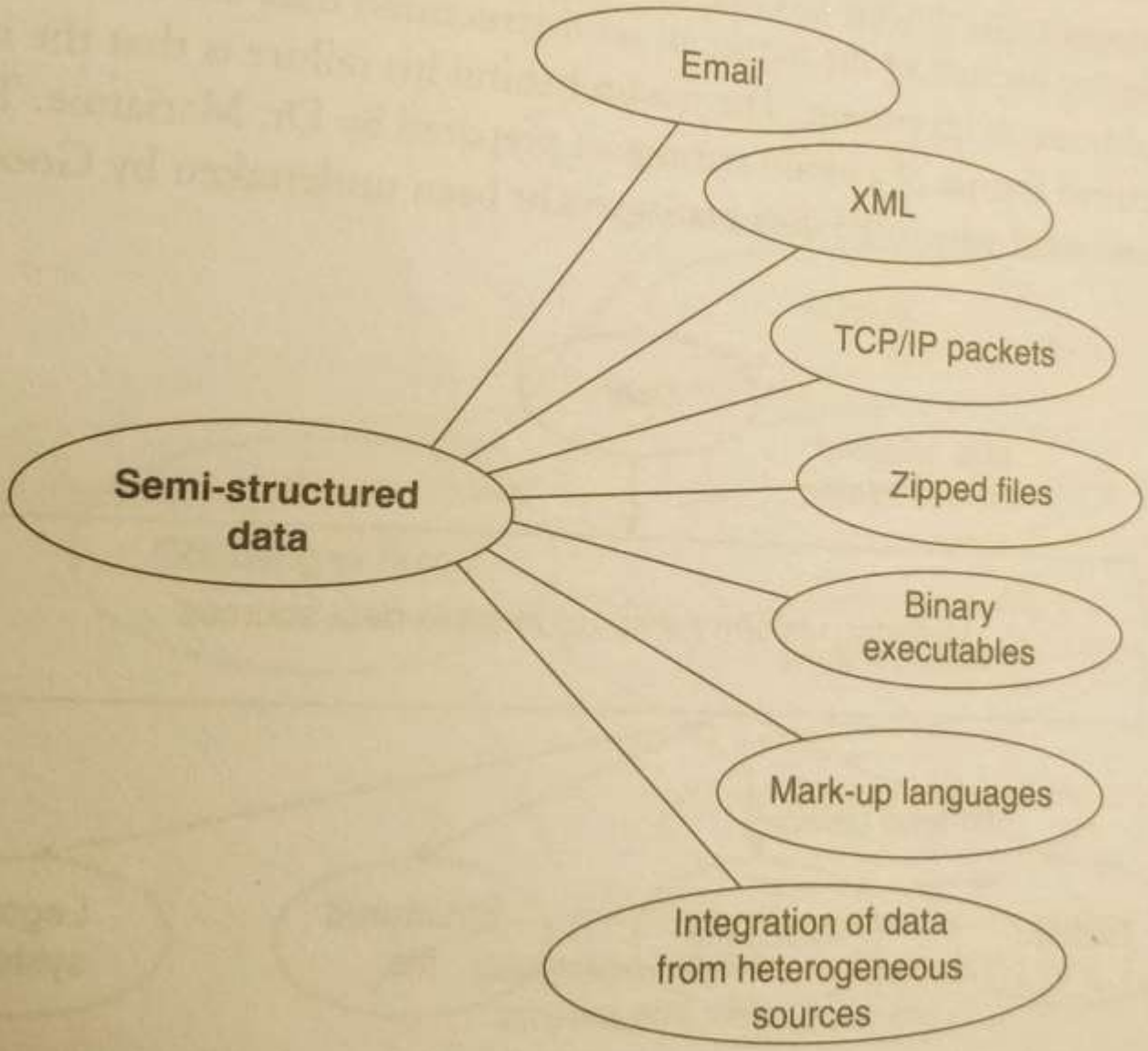


Figure 2.17 Sources of semi-structured data.

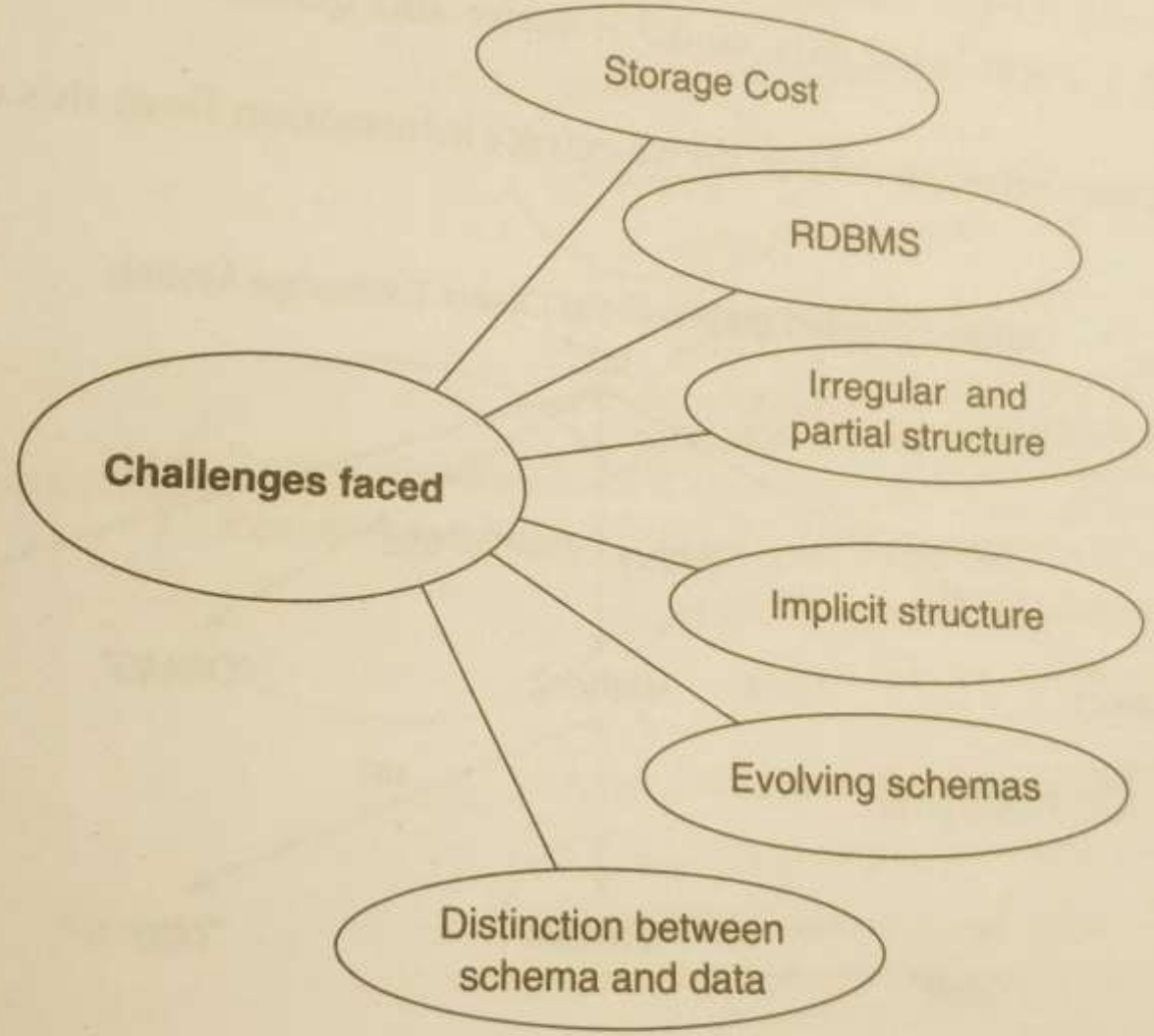


Figure 2.19 Challenges faced in storing semi-structured data.

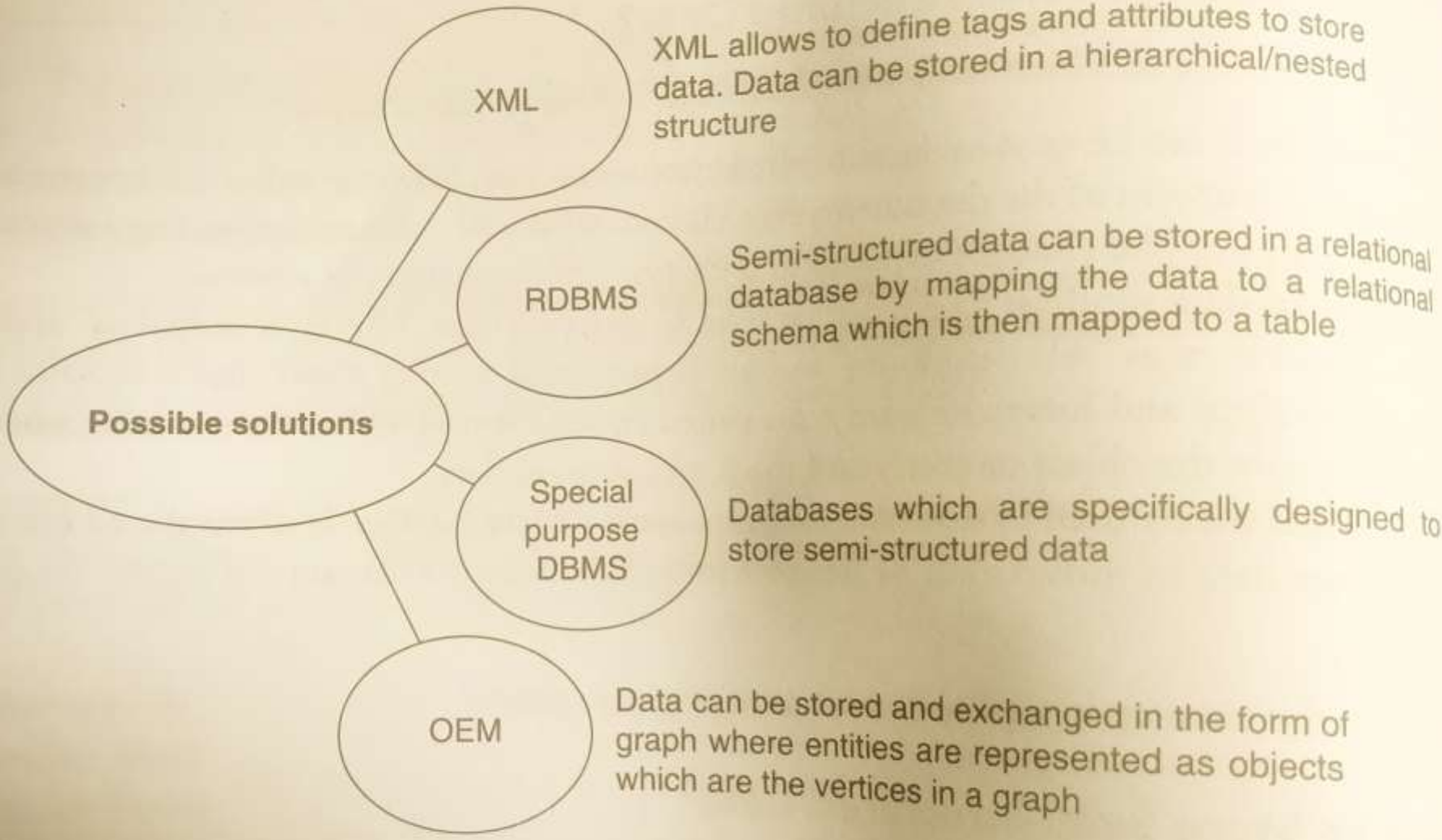


Figure 2.20 Possible solutions for storing semi-structured data.

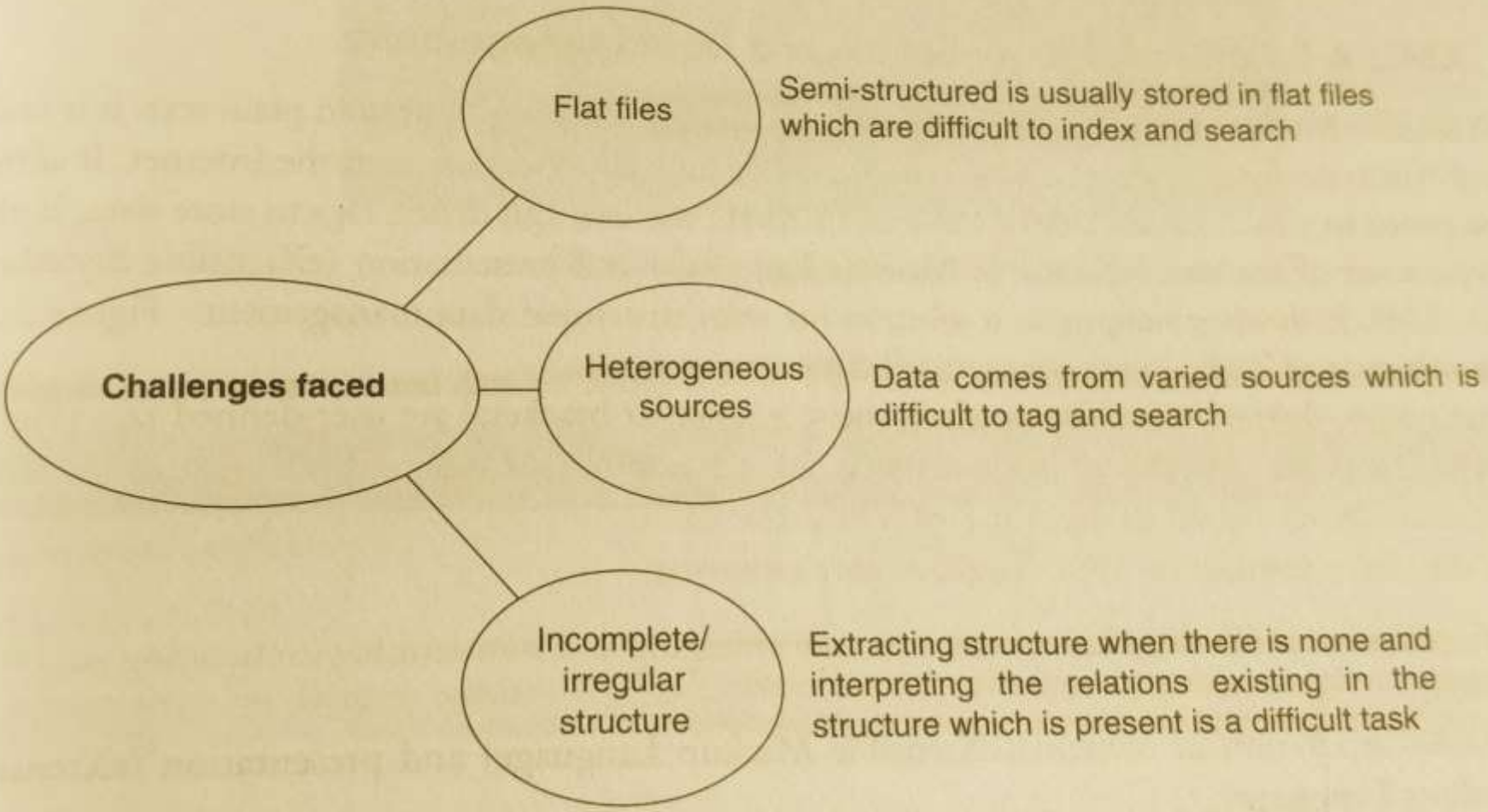


Figure 2.22 Challenges faced while extracting information from semi-structured data.

Figure 2.23 Possible solutions for extracting information from semi-structured data.

Big Data applications

Healthcare

- Big Data has already started to create a huge difference in the healthcare sector. With the help of predictive analytics, medical professionals are now able to provide personalized healthcare services to individual patients.
- Fitness wearables, telemedicine, remote monitoring – all powered by Big Data and AI – are helping change lives for the better.

Academia

- Big Data is also helping enhance education today.
- Education is no more limited to the physical bounds of the classroom – there are numerous online educational courses to learn from.
- Academic institutions are investing in digital courses powered by Big Data technologies to aid the all-round development of budding learners.

Banking

- The banking sector relies on Big Data for fraud detection.
- Big Data tools can efficiently detect fraudulent acts in real-time such as misuse of credit/debit cards, archival of inspection tracks, faulty alteration in customer stats, etc

Manufacturing

- **According to TCS Global Trend Study, the most significant benefit of Big Data in manufacturing is improving the supply strategies and product quality.**
- In the manufacturing sector, Big data helps create a transparent infrastructure, thereby, predicting uncertainties and incompetencies that can affect the business adversely.

IT

- One of the largest users of Big Data, IT companies around the world are using Big Data to optimize their functioning, enhance employee productivity, and minimize risks in business operations.
- By combining Big Data technologies with ML and AI, the IT sector is continually powering innovation to find solutions even for the most complex of problems.

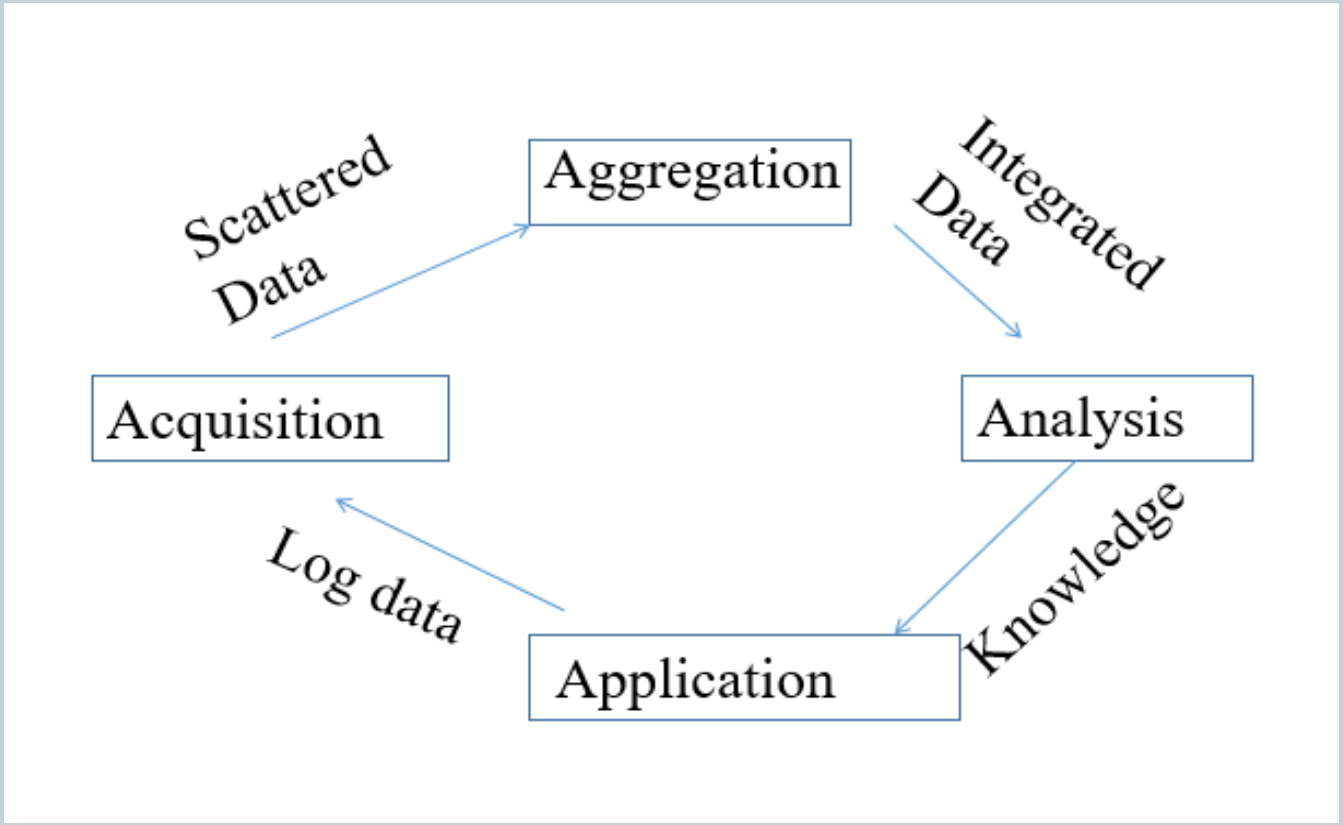
Retail

- Big Data has changed the way of working in traditional retail stores. Over the years, retailers have collected vast amounts of data from local demographic surveys, POS scanners, RFID, customer loyalty cards, store inventory, and so on. Now, they've started to leverage this data to create personalized customer experiences, boost sales, increase revenue, and deliver outstanding customer service.
- Using smart sensors and Wi-Fi to track the movement of customers, the most frequented aisles, for how long customers linger in the aisles, among other things.
- social media data to understand what customers are saying about their brand, their services, and tweak their product design and marketing strategies accordingly.

Transportation

- Big Data Analytics holds immense value for the transportation industry. In countries across the world, both private and government-run transportation companies use Big Data technologies to optimize route planning, control traffic, manage road congestion, and improve services.
- use Big Data to revenue management, drive technological innovation, enhance logistics to gain the upper hand in the market.

Life Cycle of Data



Life Cycle

Acquisition

- Acquiring Data from Various Data Sources.
- Raw data is collected from various logs, collected from various sources.

Aggregation:

- Raw data is gathered and expressed in a summary form for statistical analysis.
- average, minimum, maximum, sum, and count.

Life Cycle

Analysis:

- Analyzing the aggregated data to gain insights about particular resources or resource groups.
- Deriving Knowledge from data

Application:

- Applying this knowledge in building applications

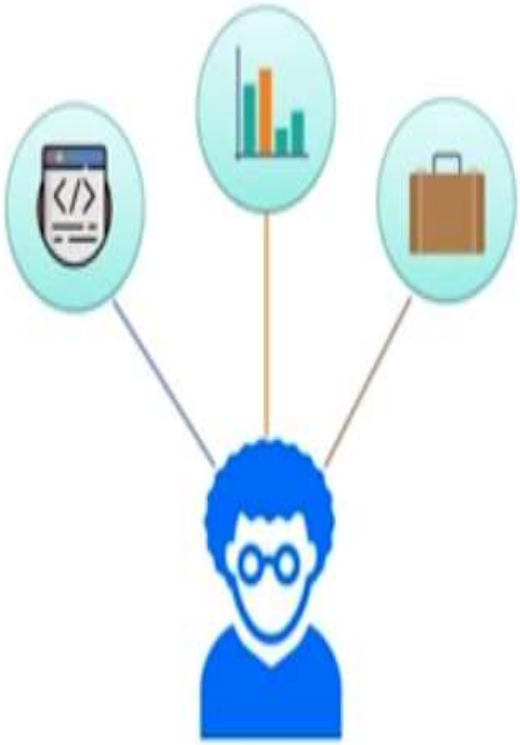
Fundamentals of Cloud Computing and Big Data

DATA SCIENCE, DATA MINING, ADVANTAGES ETC.

Data is generating a fast speed

- Every minute on Facebook: 510,000 comments are posted, 293,000 statuses are updated, and 136,000 photos are uploaded.
- 300 hours of vedio are uploaded to YouTube each minute.
- Instagram users upload over 100 million photos and videos overyday. That is 69,444 million posts every minute!
- Over 3.5 Billion Google searches are conducted worldwide each minute of everyday. That is **2 trillion searches per year** worldwide. That is over 40,000 search queries per second!

Data Science



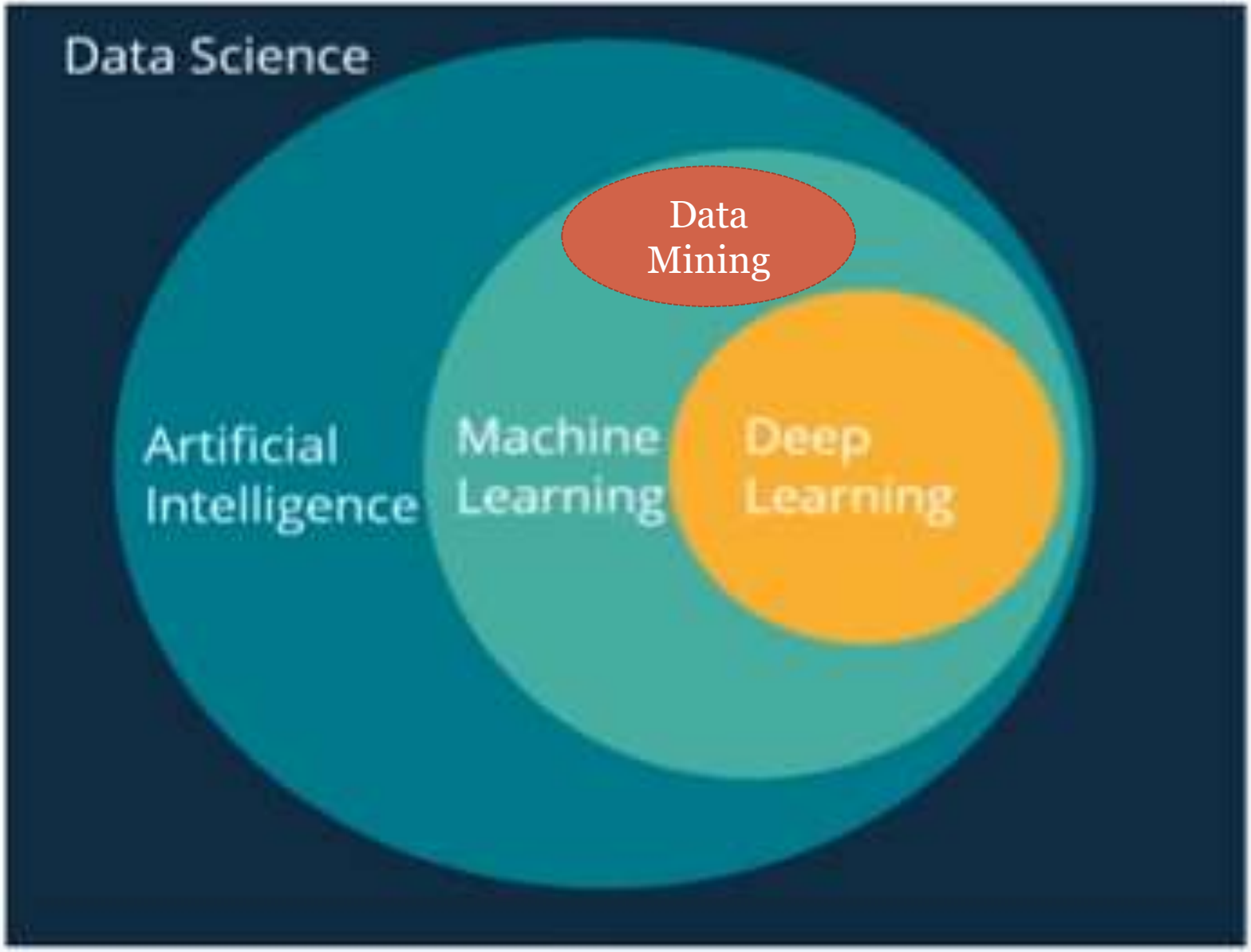
Programming + Statistics + Business

Data Science Domains

- Estimated by 2020, 1.7 MB of data will be created each second by each person over the earth.
- Results into Lots of data!!!!!!!!
- Gigabyte→Terabyte→Petabyte→Exabyte→ →Zettabyte→Yottabyte→…….etc

What is Data Science?

- Data Science is the process of extracting useful information (knowledge) from data by using different algorithms and techniques.


Data Science
Data Mining
Artificial Intelligence
Machine Learning
Deep Learning

Basic Concepts of AI/ML/DM/DL

- Definition of AI:

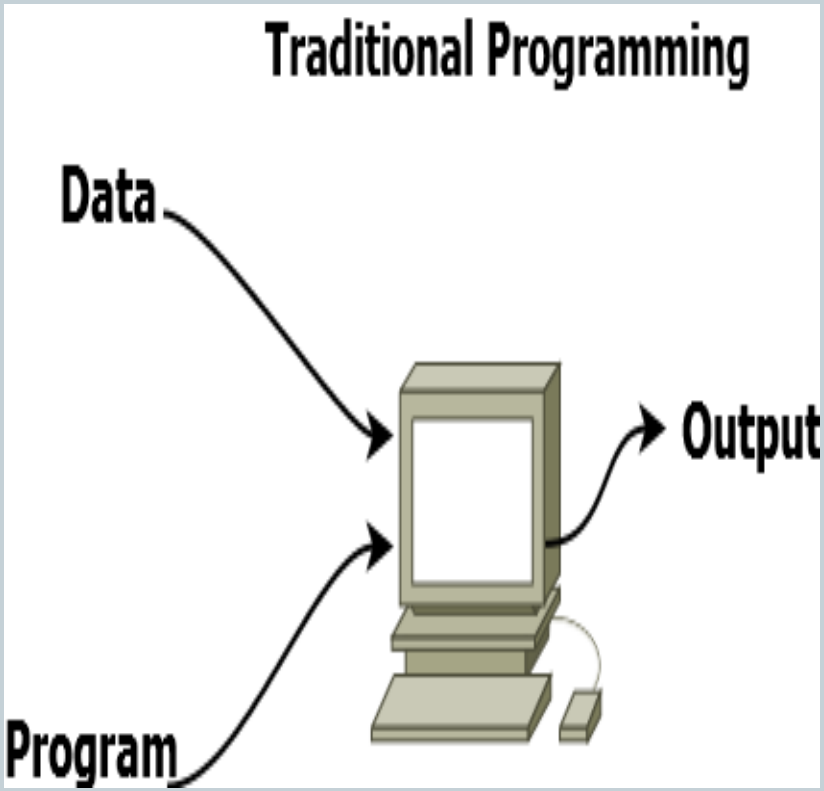
- **Artificial Intelligence is technique which enables machines to mimic human behaviour.**

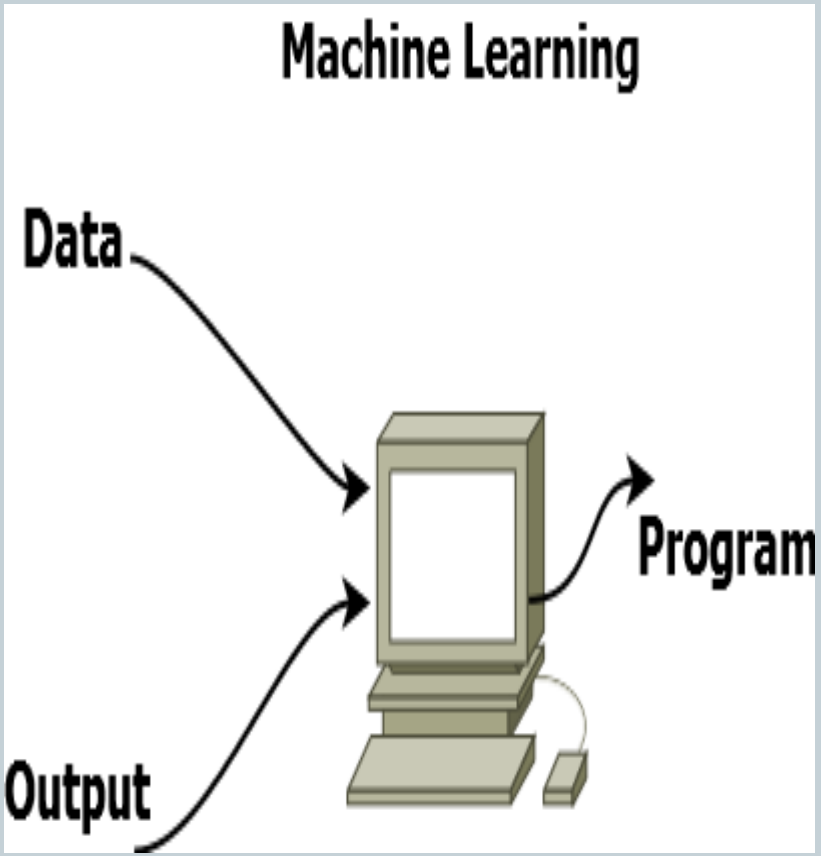
- **Ex. Robotics, AI-powered Chatbot**

Definition of ML:

- **Machine learning algorithms build programs based on historic data in order to make predictions or decisions to perform the task without being explicitly programmed.**

Basic Concepts of AI/ML/DM/DL





Example: Diabetic Patient Data

```
'preg','plas','pres', 'skin', 'test',
'mass','pedi','age','class
```

```
6,148,72,35,0,33.6,0.627,50,1
1,85,66,29,0,26.6,0.351,31,0
8,183,64,0,0,23.3,0.672,32,1
1,89,66,23,94,28.1,0.167,21,0
0,137,40,35,168,43.1,2.288,33,1
5,116,74,0,0,25.6,0.201,30,0
3,78,50,32,88,31.0,0.248,26,1
10,115,0,0,0,35.3,0.134,29,0
2,197,70,45,543,30.5,0.158,53,1
8,125,96,0,0,0.0,0.232,54,1
4,110,92,0,0,37.6,0.191,30,0
```

Basic Concepts of AI/ML/DM/DL

Definition of ML:

Machine Learning is a subset of AI technique which uses statistical methods to enables machines to build model for predictions which improve with experiences.

(Field of Data Science which does predictive analysis)

Basic Concepts of AI/ML/DM/DL

Definition of Deep Learning (DL):

Deep Learning is a subset of ML which make use of computation of multi-layer

Neural Network.

Example

- **Let's take one of the wonderful application of these concepts:**

- **Easyday wants to find the best location where they will get higher customers sales.**
- **So that they can open a new store at that location**

Customer Location (Latitude and longitude)

```
1.083576,7.319176
11.12067,14.40678
23.71155,2.557729
24.169929999999997,32.02478
21.665779999999998,4.892855
4.693683999999998,12.34217
19.21191,-1.121366
4.230391,-4.441536
9.127130000000001,23.60572
0.4075031,15.29705
7.314846000000001,3.3093120000000003
```

Basic Concepts of AI/ML/DM/DL

Definition of DM:

Data mining is the process of discovering hidden patterns in large data sets using methods of machine learning, statistics, and database technology.

Data Mining

- **Data Mining is the:**
  - Discovery of useful, possibly unexpected, patterns in data
  - Non-trivial extraction of implicit, previously unknown and potentially useful information from data

Example

X: Input Data and Y:Output Data

X	Y
1	2
2	4
3	6
4	9
5	11
6	11
7	13
8	17
9	19

if X=10 then Y?

Example 1: Housing Data

```
Per_square_feet,price
150,6450
200,7450
250,8450
300,9450
350,11450
400,15450
600,18450
```

Relationship between input and output

- Y=mX+C //Algorithm, Program, Formula, Function

(**Formally called Model**)

- Univariate Input (x) and Mutli Variate Input (X={x1, x2,…}
- Independent Variables means X, Dependant Variable means Y
- Training Data
- Test Data
- Actual output and Predicted output

Basic Concepts of AI/ML/DM/DL

Input Data /Training Data / Historic Data (given to) Model

Model (applied on) Testing Data

Basic Concepts of AI/ML/DM/DL

- Formal Terms

- **Input (X) is called features**
- **Output (Y) is called Target, Label or Class**

- **Labeled Training Data means?**

Steps in Building a ML/AI/DM Application

- **Collect Data**
- **Prepare the Input Data:** cleanse, format, handling missing data
- **Analyze the Input Data:** Plotting, finding features etc.
- **Train the Program:** Run the designed algorithm on training data
- **Test the Algorithm:** See to what extent, the algorithm works on test data
- **Use it:** If testing is successful, use it with new data for which we need an answer.

Please note: 60-80% percent effort goes to first three steps.

Basic Concepts of AI/ML/DM/DL

- **How to handle missing value**

i) Delete the entire row having missing value (some info will be lost)

ii) Delete the column having missing value (more info will be lost

iii) Fill the best probability value at that place

iv) Mean value (avg) of the entire column, can be filled at the missing field.

Basic Concepts of AI/ML/DM/DL

- We can’t use text data (such as name, course name) directly.
- We apply some process techniques to use the text data by converting it into numeric eqivalent.

Basic Concepts of AI/ML/DM/DL

- We don’t use the model generated by machine.
- First the model is applied on test data.
- If we are satisfied by performance of model then we apply it.

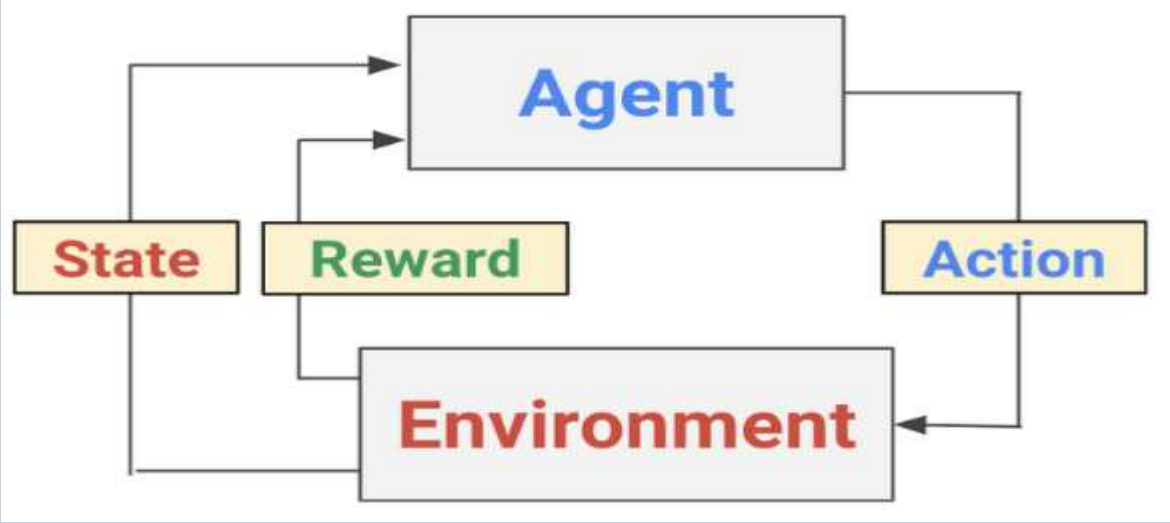
Types of ML that can be applied by DM techniques (Data Mining Tasks)

- 1) Supervised Learning
- 2) Unsupervised Learning
- 3) Reinforcement Learning
- 4) Semi-supervised Learning

- If experience data/Historic data has both X and Y→ Supervised Learning is applied
- If experience data has only X→ then unsupervised learning

Types of ML that can be applied by DM techniques

- Reinforcement learning – Based on feedback system reward penalty system



- Semi-supervised: Small section of data contain (X,Y) but majority of data contains X.

Supervised Learning

- **Supervised learning** is the Machine Learning /Data Mining task of learning a function that maps an input to an output based on example input-output pairs (Labeled Training Dataset).

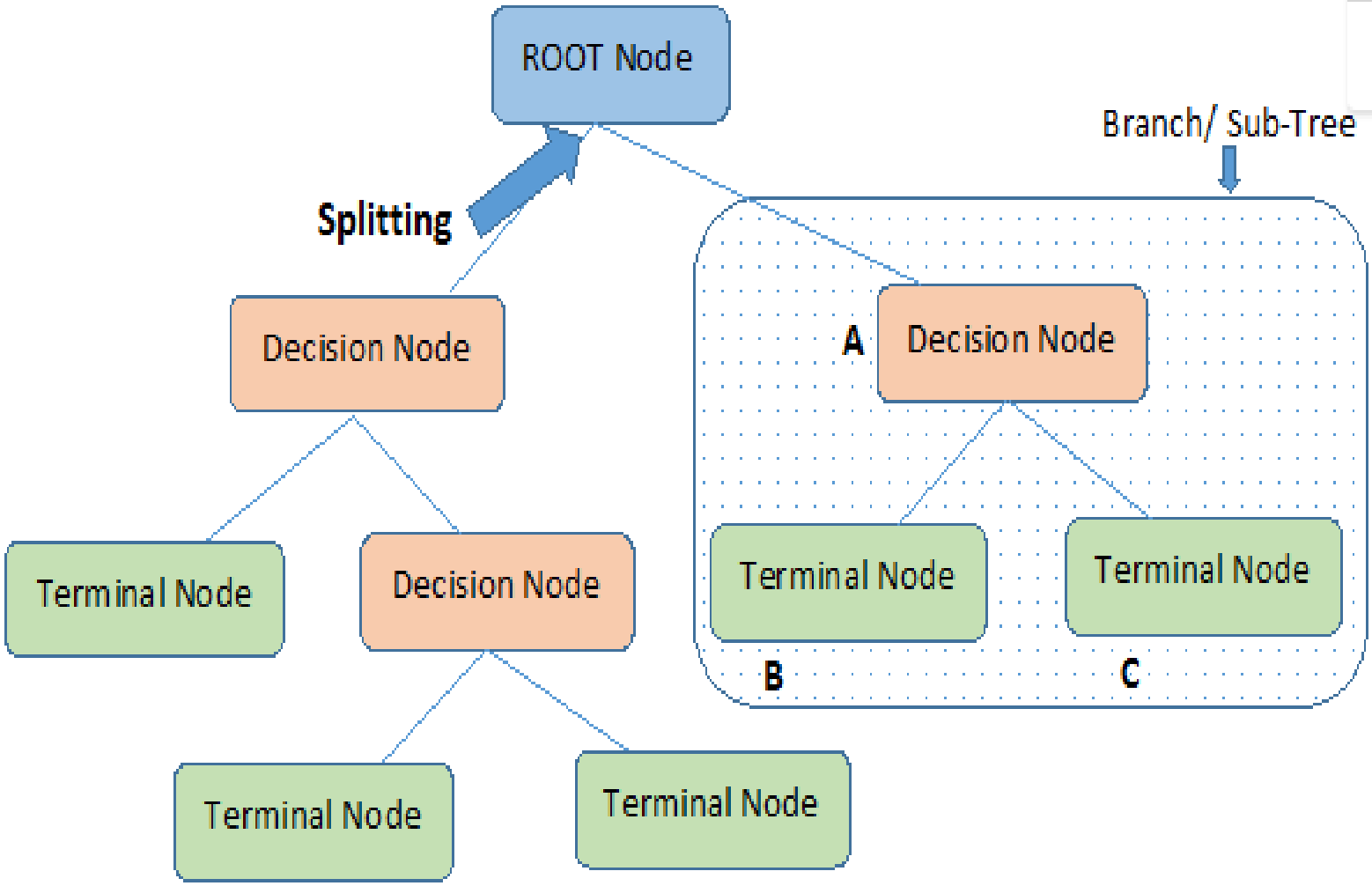
- Types of SL:
- A) Regression (ex. Linear Regression) (when we have continuous output (Y) values)
- B) Classification (ex. Logistic Regression) (when we have discrete output (Y) values/categorical values

Supervised Learning Algorithms

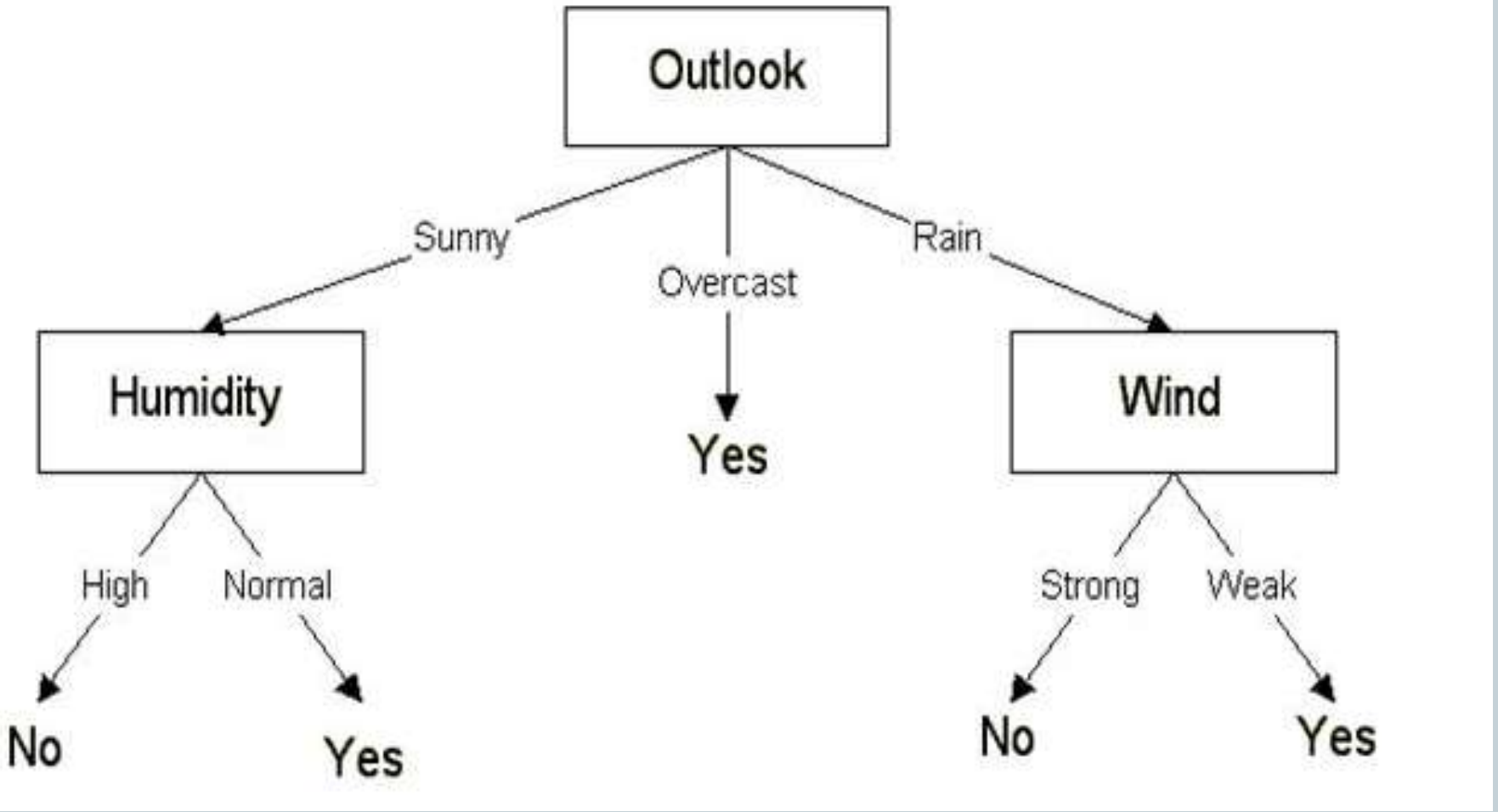
- Linear Regression
- Logistic Regression
- Decision Tree (ID3, C 4.5, CART)
- Naïve Bayes
- k-Nearest Neighbour (KNN)
- Support Vector Machine
- Back Propagation Neural Network

Etc.

Decision Tree



Decision Tree (Ex. Play Football or not)



Unsupervised Learning

- **Unsupervised learning** is the Machine Learning /Data Mining task to model the underlying structure or distribution in the data in order to learn more about the data.
- **Unsupervised learning is where you only have input data (X) and no corresponding output variables.**

Types of UL

- **Clustering**: A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior.

- **Association**: An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

Unsupervised Learning Algorithms

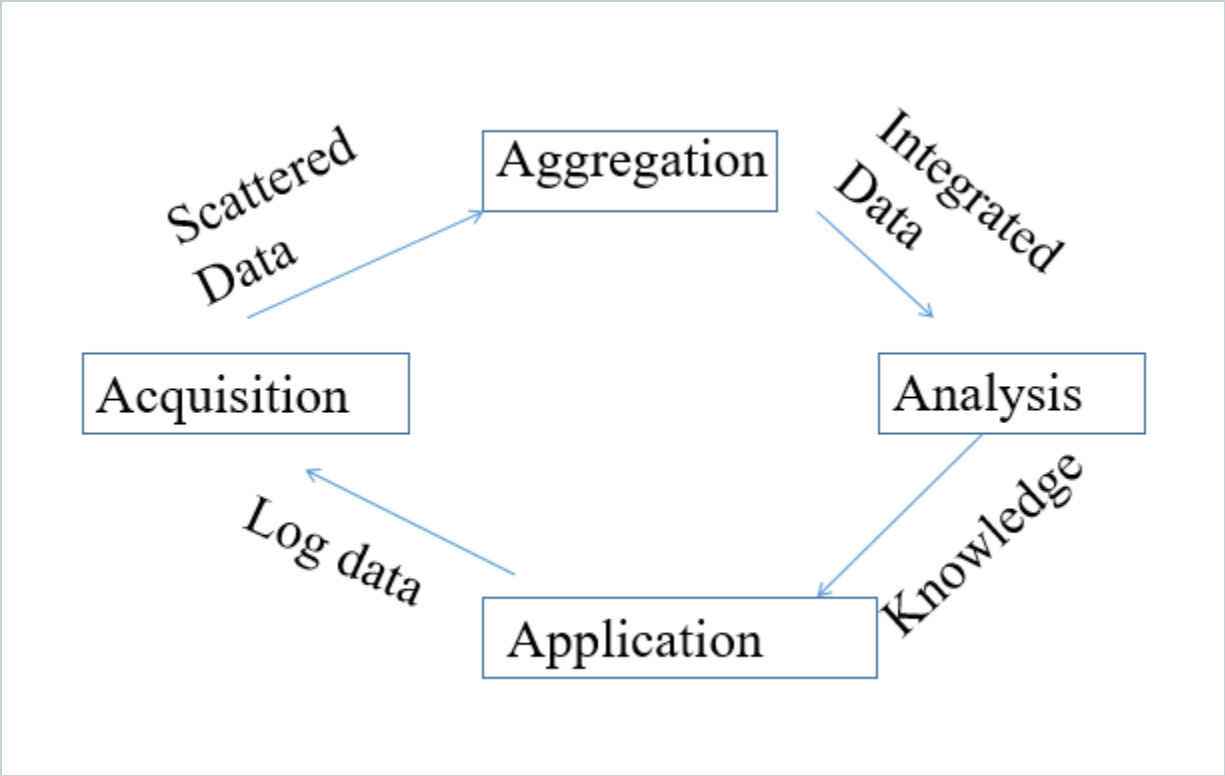
- K-Means Clustering (clustering)
- Apriori (associate rule)
- Association rule mining (association rule)
- Hierarchical Clustering (clustering)
- DBSCAN clustering (clustering)
- Other: Self Organizing Map (Neural Network)
  One-Class SVM

Other Types of Mining

- **Text mining**: application of data mining to textual documents
  - cluster Web pages to find related pages
  - cluster pages a user has visited to organize their visit history
  - classify Web pages automatically into a Web directory
- **Graph Mining**:

  Deal with graph data

Life Cycle of Data



Life Cycle

Acquisition

- Acquiring Data from Various Data Sources.
- Raw data is collected from various logs, collected from various sources.

Aggregation:

- Raw data is gathered and expressed in a summary form for statistical analysis.
- average, minimum, maximum, sum, and count.

Life Cycle

Analysis:

- Analyzing the aggregated data to gain insights about particular resources or resource groups.
- Deriving Knowledge from data

Application:

- Applying this knowledge in building applications

Big Data Advantages

- Big data analysis derives innovative solutions. Big data analysis helps in understanding and targeting customers. It helps in optimizing business processes.
  ➡It helps in improving science and research.
  ➡It improves healthcare and public health with availability of record of patients.
  ➡It helps in financial tradings, sports, polling, security/law enforcement etc.

Big Data Advantages

➡Any one can access vast information via surveys and deliver anaswer of any query.
➡Every second additions are made.
➡One platform carry unlimited information

- One of the biggest advantages of Big Data is predictive analysis. Big Data analytics tools can predict outcomes accurately, thereby, allowing businesses and organizations to make better decisions, while simultaneously optimizing their operational efficiencies and reducing risks.

- By harnessing data from social media platforms using Big Data analytics tools, businesses around the world are streamlining their digital marketing strategies to enhance the overall consumer experience. Big Data provides insights into the customer pain points and allows companies to improve upon their products and services.

- Being accurate, Big Data combines relevant data from multiple sources to produce highly actionable insights. Almost 43% of companies lack the necessary tools to filter out irrelevant data, which eventually costs them millions of dollars to hash out useful data from the bulk. Big Data tools can help reduce this, saving you both time and money.

- Big Data analytics could help companies generate more sales leads which would naturally mean a boost in revenue. Businesses are using Big Data analytics tools to understand how well their products/services are doing in the market and how the customers are responding to them. Thus, the can understand better where to invest their time and money.

- With Big Data insights, you can always stay a step ahead of your competitors. You can screen the market to know what kind of promotions and offers your rivals are providing, and then you can come up with better offers for your customers. Also, Big Data insights allow you to learn customer behavior to understand the customer trends and provide a highly ‘personalized’ experience to them.

Big Data Disadvantages

- Following are the drawbacks or **disadvantages of Big Data**:
  ➡Traditional storage can cost lot of money to store big data.
  ➡Lots of big data is unstructured.
  ➡Big data analysis violates principles of privacy.
  ➡It can be used for manipulation of customer records.
  ➡It may increase social stratification.

Big Data Disadvantages

- ➡Big data analysis is not useful in short run. It needs to be analyzed for longer duration to leverage its benefits.
- ➡Big data analysis results are misleading sometimes.
- ➡Speedy updates in big data can mismatch real figures.

Types of Big Data

- **Social Networks (human-sourced information)**:
  - Social Networks: Facebook, Twitter, Tumblr etc.
  - Blogs and comments
  - Personal documents
  - Pictures: Instagram, Flickr, Picasa etc.
  - Videos: Youtube etc.
  - Internet searches
  - Mobile data content: text messages
  - User-generated maps
  - E-Mail

Types of Big Data

- **Traditional Business systems (process-mediated data)**:

Data produced by Public Agencies

- Medical records

Data produced by businesses

- Commercial transactions
- Banking/stock records
- E-commerce
- Credit cards

Types of Big Data

- **Internet of Things (machine-generated data)**:

Data from sensors

Fixed sensors

Home automation

Weather/pollution sensors

Traffic sensors/webcam

Scientific sensors

Security/surveillance videos/images

Mobile sensors (tracking)

Mobile phone location

Cars

Types of Big Data

Data from computer systems

- Logs
- Web logs