# Data analytics batch(22-8601)

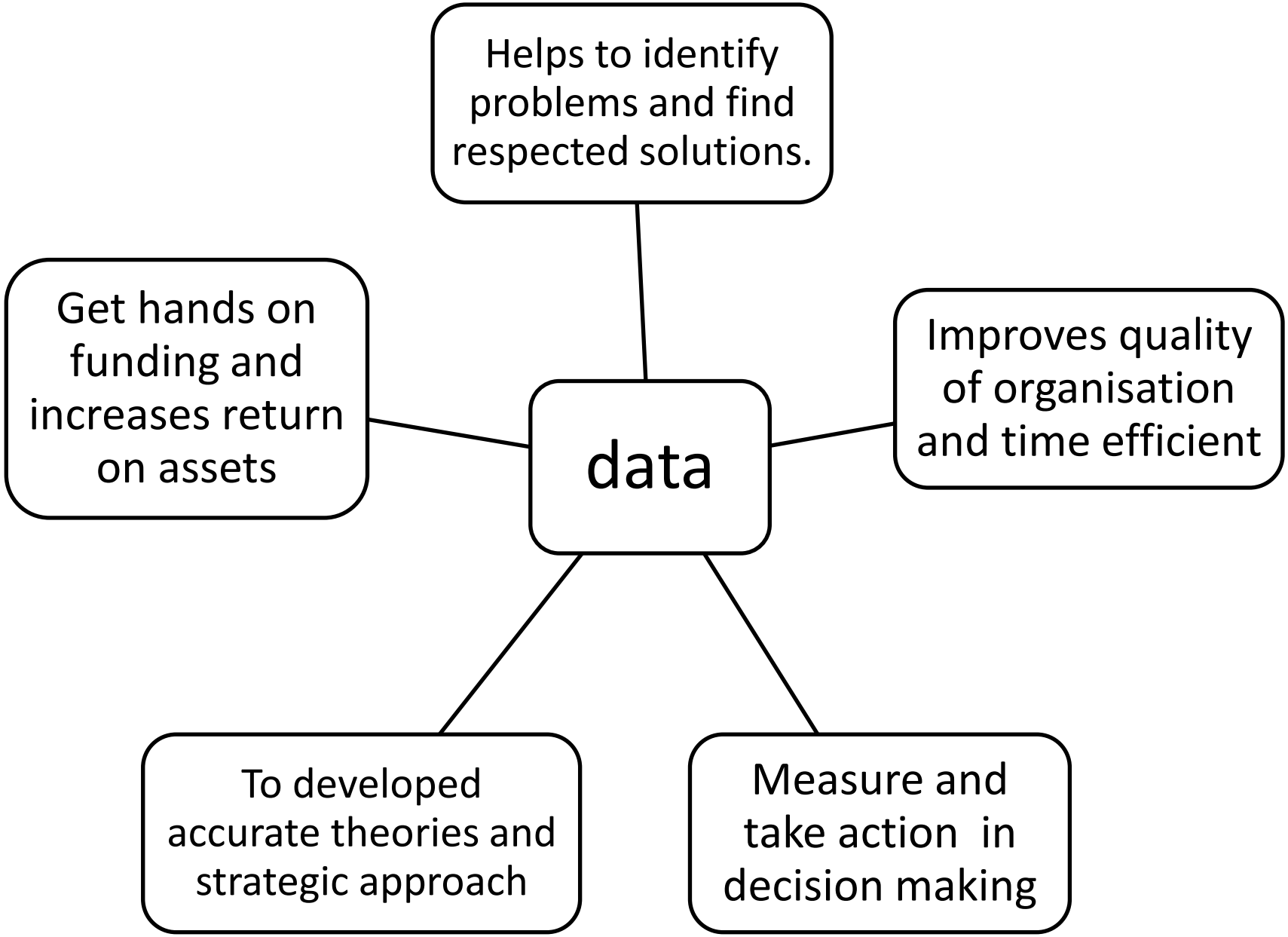
Assignment no. 01

Rahee d sutrave

# Q.1) differentiate between data and information

| details | Data | Information |
|---|---|---|
| Meaning | It refers as raw fact that have no specific meaning. | It refers as processed data that has purpose and meaning. |
| | Data is unrecognised and does independent on information. | Information is recognised and dependent on data. |
| Etymology | Word is derived from Latin word ‘DATUM’ means ‘something that is given’. | Word is derived from Latin word ‘INFORMATIO’ means ‘conception’. |
| Description | Data can be qualitative and quantitative. | Information can be ideas and inferences. |
| | It can be presented in tabular form, graph, pictorial representation. | It seen as language, ideas, and thoughts based on given data. |
| | Data is less significant and not enough to make decision. | Information is more significant and sufficient to help make decision in respected context. |
| Example | Score details of students in the class. | Average score of students in a particular subject based on data. |

# Q.2) How data is useful for us?



# Q.3) what is big data?

- Data set that is too large and complex to deal with traditional data processing application software.
- Large information body that could not comprehend when used only in similar amounts.

Key concept:

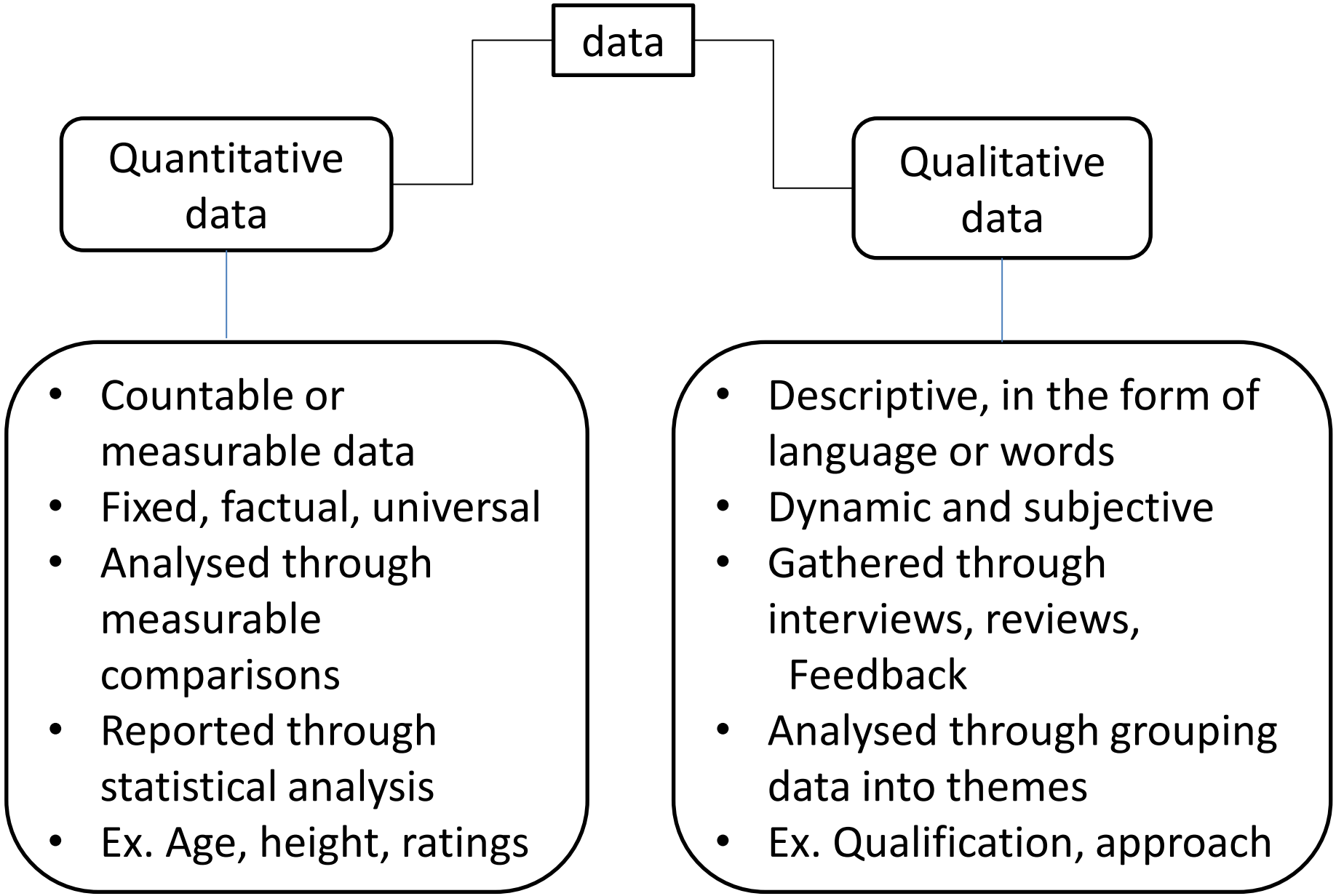
- Volume: huge amount of data
- Variety: different formats of data from various sources
- Value: extract useful data
- Velocity: high speed of accumulation of data
- Veracity: inconsistencies and uncertainty in data

- Relational database management system and desktop statistical software package use to visualizing, processing and analyzing big data.

# Q.4) differentiate between structured, semi structured, unstructured data?

| properties | Structured data | Semi structured data | Unstructured data |
|---|---|---|---|
| organisation | Well organised | Partially organised | Not organised |
| Flexibility and scalability | Schema dependent less flexible and difficult to scale | More flexible and simpler than structured data but less than unstructured data | No schema more flexible and more scalable |
| technology | Based on relational database | Based on XML/RDF | Based character and binary data |
| Transaction management | Matured and various concurrency techniques available | Transaction from DBMS and no concurrency available. | no transaction and concurrency available |
| Version management | As tuples, rows and tables | As tuples(finite set) | As whole data |
| example | Financial data, bar codes | Tweets by hashtag, folder by topics | Media logs, videos, audios |

# Q.5) what are quantitative and qualitative data?



Q.6)what are different V's in big data?

V's present in big data are as follows:

1. VALUE: usefulness of data in decision making, business value of data.
2. VOLUME: amount of data in terms of potential of terabytes or petabytes, or amount of data collected by source.
3. VARIETY: an expanding universe of data, types, source i.e. different forms of data used.
4. VELOCITY: how fast data is travelled or being processed.
5. VERACITY: data reliability and trust, verifying and validating data.

Q.7) name some popular tools used in big data?

Following tools are used for big data:

- Microsoft Excel
- Tableau
- SAS
- Apache Cassandra
- Apache Hadoop
- Apache Flink
- Apache storm
- Statwing
- KNIME
- MongoDB
- RapidMiner

Q.8) what are different types of data?