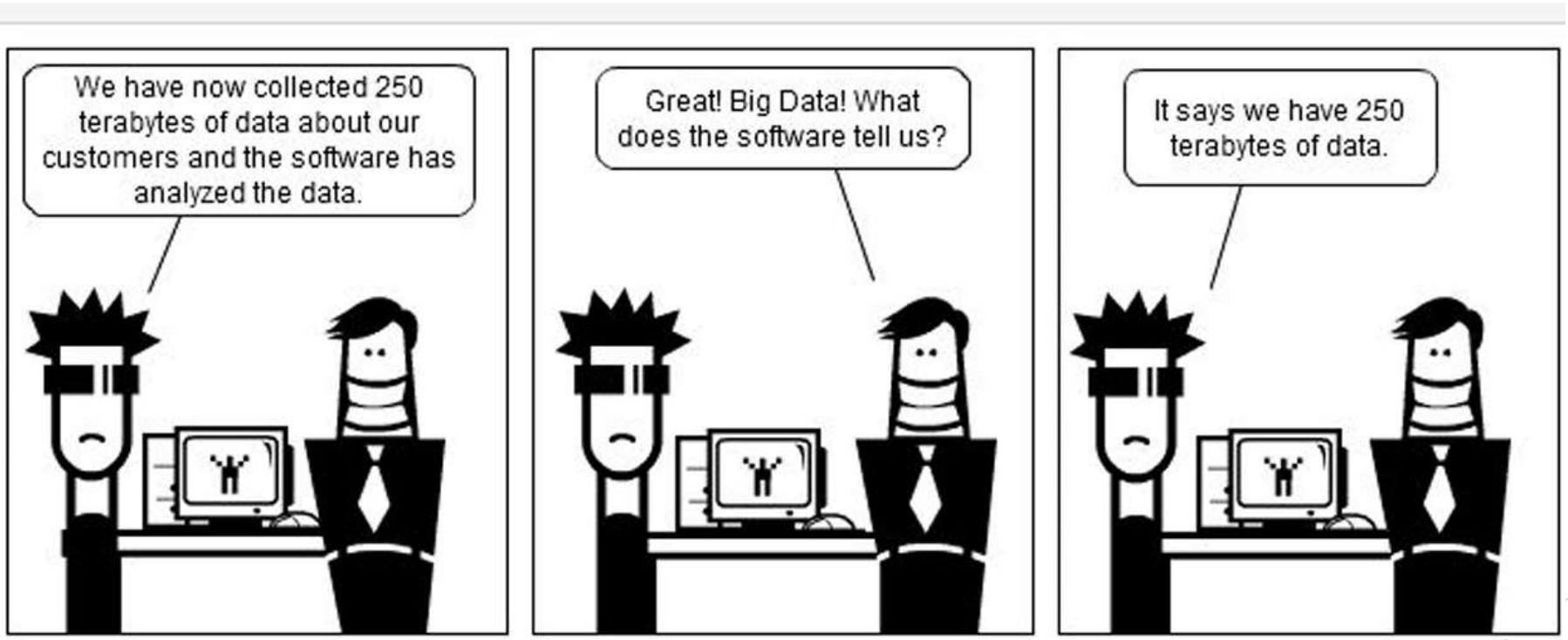
BIG DATA CHALLENGE



Types of Big Data

a) Structured

Structured is one of the types of big data and By structured data, we mean data that can be processed, stored, and retrieved in a fixed format.

It refers to highly organized information that can be readily and seamlessly stored and accessed from a database by simple search engine algorithms.

For instance, the employee table in a company database will be structured as the employee details, their job positions, their salaries, etc., will be present in an organized manner.

b) Unstructured

- Unstructured data refers to the data that lacks any specific form or structure whatsoever. This makes it very difficult and time-consuming to process and analyze unstructured data. Email is an example of unstructured data. Structured and unstructured are two important types of big data.

unstructured Data

Semi-structured

- Semi structured is the third type of big data.
- Semi-structured data pertains to the data containing both the formats mentioned above, that is, structured and unstructured data.
- To be precise, it refers to the data that although has not been classified under a particular repository (database), yet contains vital information or tags that segregate individual elements within the data .
- Thus we come to the end of types of data.

Semi structured data : JSON files

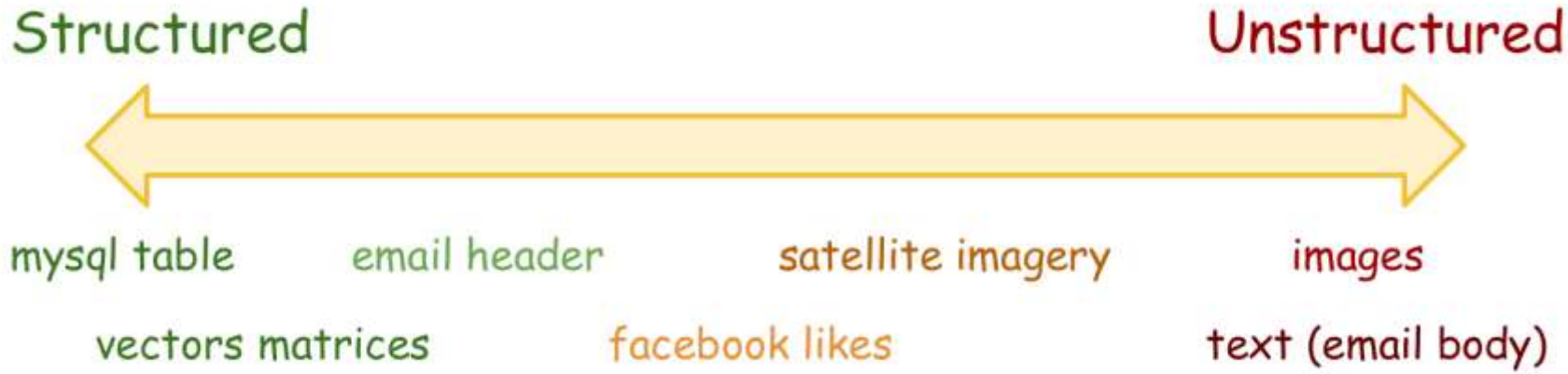
```
## Document 1 ##
{
   "customerID": "103248",
   "name":
   {
      "first": "AAA",
      "last": "BBB"
   },
   "address":
   {
      "street": "Main Street",
      "number": "101",
      "city": "Acity",
      "state": "NY"
   },
   "ccOnFile": "yes",
   "firstOrder": "02/28/2003"
}
```

Structured Data

- RELATIONAL DATABASES
- Characterastics
  - Data is stored in the form of rows and column. Eg : database
  - Data resides in fixed fields within a record or file
  - Similar entities are grouped together to form relations or classes
  - Entities in the same group have same attributes
  - Easy to access and query, so data can be easily used by other programs
  - Data elements are addressable, so efficient to analyse and process

Structured to Un structured

- Unstructured ≈ requires processing to get what is of interest
- Feature extraction used to turn unstructured into structured
- Near infinite amounts of potential features in unstructured data


Structured
Unstructured
mysql table
email header
satellite imagery
images
vectors matrices
facebook likes
text (email body)

INDUSTRY EXAMPLES OF BIG DATA

- HELTHCARE
- MANUFACTURING
- BANKING
- LETS FIND OUT WHICH ARE THE OTHER AREAS??????

SCHEMA ON WRITE-SCHEMA ON WRITE

- **Approaches to storing and managing data**
- **Schema-on-write** and **schema-on-read** are two different approaches to storing and managing data. Schema-on-write means that the schema, or structure, of the data, is defined when the data is written to the database. Schema-on-read means that the schema is defined when the data is read from the database.
- Structured data is typically stored using schema-on-write. This is because the schema is known in advance and can be used to optimize the storage and performance of the data. Unstructured data is typically stored using schema-on-read. This is because the schema is not known in advance and may need to be changed frequently.
- Which approach is best for a particular application depends on the specific needs of the application.