# Types of Digital Data

# STRUCTURED DATA

- Structured data is organised in semantic chunks with similar entities grouped together to form relations or classes.
- Characteristics:
- Data is stored in the forms of rows and columns.
- Data resides in fixed fields within a record or file.
- Definition format and meaning of data is explicitly known.
- Attributes in a group are the same.

# Where Does Structured data come from?

- Consists of fully described sets.
- Has clearly defined categories and sub categories.
- Is placed neatly in rows and columns.
- Data coming from databases such as access oltp systems and spreadsheets.

# Working with Structured data

- Storage: Both Defined an used defined types.
- Scalability: Is not generally an issue.
- Security: Ensuring security is easy.
- Operations: Can be easily updated and deleted.

# Hassle Free Retrieval

- Indexing & Searching , Mining of Data and BI works well with structured data.

# Unstructured Data

- Does not conform to any data model.
- Cannot be stored in the forms of rows and columns.
- Not in any particular format or sequence.
- Not easily usable by a program.
- Does not follow any rules or semantics.

# Unstructured Data- Bitmap Objects, Textual Objects.

- Web Pages
- Memos
- Videos
- Images
- Word document
- Power point
- Reports
- Surveys
- A lot of unstructured data is also noisy text such as emails, chats, SMS texts.

# How to store Unstructured data

- Storage Space : Difficult to store, a lot of storage space is required.
- Storing in RDBMS in form of BLOB’s and LOBs.
- Scalability – As data grows scalability becomes an issue.
- Security, update and delete operations and Index based retrieval all are a issue in this kind of data.

# Challenges of Unstructured Data.

- Changing Formats: Unstructured data may be converted to formats which are easily stored and searched. IBM is working on how to provide solutions.
- Developing new hardware:
- Storing in RDBMS Blobs and Lobs
- Storing in XML
- CAS – Content Addressable Storage (Organized files based on their meta data and assigns a unique name to every object.

# UIMA (unstructured Information Management Arch)

- Breaking of documents into separate words.
- Grouping and classifying according to taxonomy.
- Detecting parts of speech grammar and synonyms
- Detecting relationships between various elements.

# Semi Structured Data

- Email
- XML
- TCP/IP packets
- Binary executables.
- Integration of Data from various sources.