Islamic University of Technology (IUT)

Organization of Islamic Cooperation (OIC)

Department of Electrical and Electronic Engineering (EEE)

COURSE NO : **EEE 4416**
LAB NO : **08 (Part A)**
TOPIC : **DATA HANDLING**

DATA HANDLING

Data handling refers to the process of collecting, organizing, managing, processing, and analyzing data to extract useful information or make decisions. In simple terms, it is everything you do to work with data — from getting it to cleaning it to using it for analysis or modeling.

Some of the key Steps in Data Handling are as follows.

i. **Data Collection:** Gathering raw data from various sources (e.g., sensors, surveys, databases, MRI, X-ray).
ii. **Data Storage:** Storing it in appropriate formats like .csv, .mat, .DICOM, databases, etc.
iii. **Data Cleaning:** Removing or correcting errors, handling missing or inconsistent values.
iv. **Data Transformation:** Converting data into usable formats, normalizing, scaling, or encoding variables.
v. **Data Analysis:** Using statistics, data mining techniques, or machine learning to draw conclusions or make predictions.
vi. **Data Visualization:** Presenting data using charts, graphs, and plots to reveal patterns or insights.
vii. **Data Exporting:** Saving the processed data or results for reporting or further use.

Each step in data handling can be done using specific tools, programming languages, or platforms. It is summarized below.

i. **Data Collection:**

Task	Tools / Languages
Manual data entry	Excel, Google Forms
Web scraping	Python (BeautifulSoup, Scrapy), R (rvest)
API-based data collection	Python (requests, JSON), JavaScript, R
Sensor/IoT data	Python, C/C++, MATLAB (Data Acquisition Toolbox)
Databases	SQL, Python (sqlite3, SQLAlchemy), R (DBI)

ii. **Data Storage:** Different types of data like audio, images, and MRI scans — require specialized formats and tools for efficient storage, retrieval, and processing.

Asif Newaz
Lecturer, EEE, IUT

Source	Format	Tools / Languages
Flat files	.csv, .txt	Excel, Python, R, MATLAB
Relational databases	.sql	MySQL, SQLite, Oracle
Cloud storage		AWS S3, Google Cloud Storage, Azure Blob
Audio	.wav, .mp3, .flac	MATLAB, Audacity, Python (librosa)
Image	.png, .jpg, .bmp, .gif, .tiff	Python (OpenCV, PIL), MATLAB, R, Photoshop
MRI	.DICOM, .NIfTI	Python (pydicom, nibabel), MATLAB, 3D Slicer

iii. **Data Processing:** Python (pandas, re [regexp], scikit-learn), R, MATLAB, Excel, Scala, PySpark (Big data), GeoPandas (geospatial data).

iv. **Data Analysis:**

Task	Tools / Languages
Descriptive statistics	Python (NumPy, pandas), R, MATLAB, Excel
Inferential statistics	R, Python (SciPy, statsmodels)
Machine learning	Python (scikit-learn), R (caret), MATLAB
Deep learning	Python (TensorFlow), PyTorch, MATLAB
Big data processing	Apache Spark (PySpark), Hadoop, Dask

v. **Data Visualization:**

Tool / Library	Features
Python (matplotlib, seaborn, plotly)	Flexible, publication-ready
R (ggplot2, plotly)	Interactive plots
MATLAB	Built-in plots, engineering-focused
Tableau, Power BI	Drag-and-drop interactive dashboards
Excel	Easy charts and pivot tables

Data Types

As you have seen so far, data can come from different sources such as image data, audio data, tabular data, etc. Depending on the type of data, it requires different handling schemes and processing tools.

In data handling and analytics, data is categorized by its structure, which determines how it's stored, processed, and analyzed. It can be divided into three distinct categories. They are as follows.

i. Structured data
ii. Unstructured data
iii. Semi-structured data

1. Structured Data

Structured data is **organized in a predefined schema** — like rows and columns in tables. It's easily searchable and fits neatly into relational databases. They have a tabular format. They are easy to store and query. Examples:

- Excel spreadsheets
- SQL databases (MySQL, Oracle)
- CSV files
- Sensor logs with time stamps

2. Unstructured Data

Unstructured data has **no fixed format or schema.** It's often rich in information but harder to process and analyze. Requires preprocessing and specialized tools (NLP, computer vision). Examples:

- Text documents (PDF, DOC), Emails, Social media posts
- Images (JPG, PNG)
- Videos (MP4)
- Audio (MP3)
- Scanned medical images (MRI, DICOM)

3. Semi-Structured Data

Semi-structured data **doesn't fit into a strict table** but still has some organizational properties**, like tags or key-value pairs.** They are not as rigid as structured data. Easier to parse than unstructured data. Flexible but with enough metadata for interpretation. Examples:

- JSON
- XML
- YAML
- HTML
- Log files
- NoSQL databases (MongoDB)

Asif Newaz
Lecturer, EEE, IUT