Digital Data

Objective:

Recording types of data and various file formats. Identifying data sources. Handling traditionally to start with at a small scale.

Types of Digital Data

1. Structured Data
2. Semi-structured Data
3. Unstructured Data

Various File Formats (Data Formats)

1. Spreadsheet (.odt .xls .xsls)
2. CSV/TSV
3. XML
4. JSON
5. Configuration File (.ini)
6. Properties file (.properties)
7. MS Access (.mdb .accdb)
8. Oracle Database Dump (.dmp)
9. YAML (https://docs.ansible.com/ansible/latest/reference_appendices/YAMLSyntax.html)
10. BSON (https://docs.mongodb.com/manual/reference/bson-types/) (http://bsonspec.org/)

Identifying Data Sources

Scenario-based question:

You are at the university library. You see a few students browsing through the library catalog on a kiosk. You observe the librarians busy at work issuing and returning books. You see a few students fill up the feedback form on the services offered by the library. Quite a few students are learning using the e-learning content.

Think for a while on the different types of data that are being generated in this scenario. Support your answer with logic.

Handling traditionally to start with at a small scale.

Guidelines:

- Explore GUI and CLI both options for the selected tool
- Note down / Think of challenges involved

Tasks

1. Given the spreadsheet file convert it into a csv
2. Import a csv into MySQL database table
3. Write a computer program to read records from database and generate data file.
   a. XML
   b. JSON
4. Import XML/JSON file into another database/table. I.e MS Access. Oracle, etc.
5. Export database dump for data migration/archival
6. Validate/Map data types across different database systems when migrating from one to another
7. Represent Data Cube and perform operations. OLAP - Data Warehouse
8. Generate pdf report.

Reference/Courtesy:

Textbook.
Big Data and Analytics – Seema Acharya and Subhashini Chellappan – Wiley India
Chapter 1. Types of Digital Data

Documented By:

    Jigar M. Pandya

    https://www.linkedin.com/in/jigar-pandya

Document Last updated: 28th Jun, 2020.