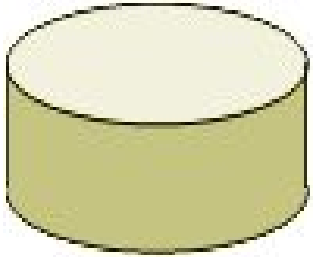
DATABASES



Unit 1 **Information storage**

Files

Content-based classification:

- **Plain text**
  - ➢ Used for storing sequences belonging to a determined character encoding (ASCII, Unicode, EBCDIC…)
  - ➢ Content can be managed via a text editor (Windows Notepad, Unix vi)
  - ➢ Examples: .txt (text), .csv (comma-separated values), .htm, .html (web page HTML code), .xml, .rss (markup languages)

- **Binary**
  - ➢ Binary-coded information prepared for application processing
  - ➢ Illegible by a text editor
  - ➢ Examples: .exe (executable files), .pdf, .docx, .xlsx (application documents), .jpg, .gif, .mp3, .avi, .mkv (media files), .dll (system files)

Plain text files

Internal organisation-based classification:

- **Sequential**. The data is written in consecutive locations. Accessing a piece of data implies running through each previous position

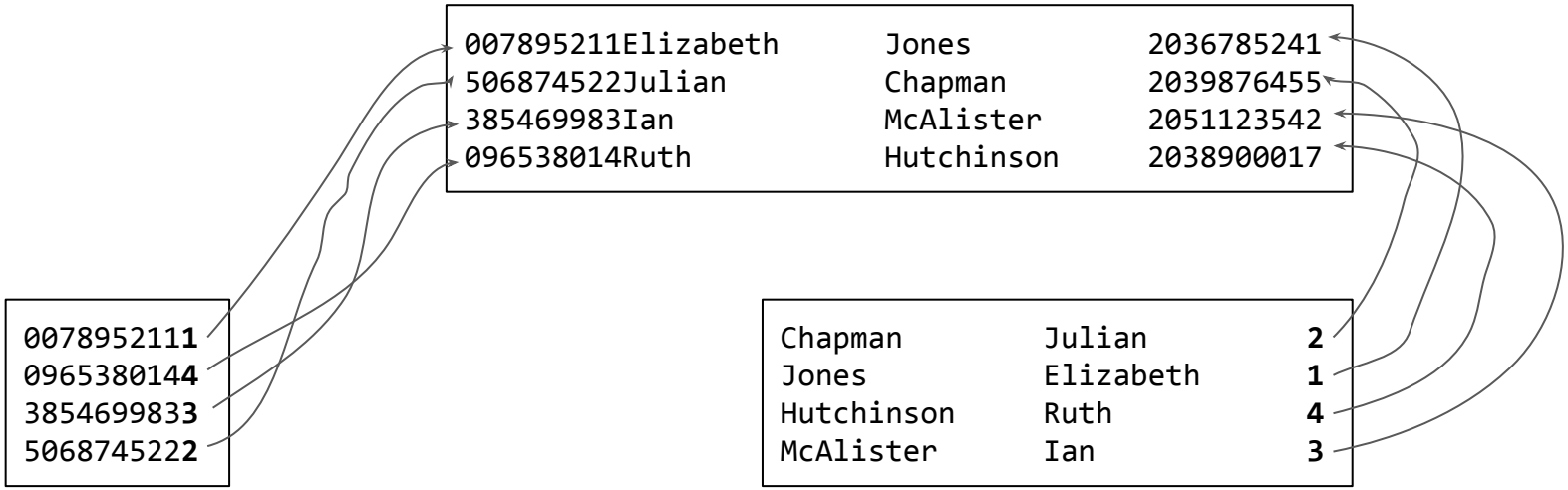
```
007895211#Elizabeth#Jones#2036785241$506874522#Julian#Chapman#2039876455$385469983#Ian
#McAlister#2051123542$096538014#Ruth#Hutchinson#2038900017%
```

- **Relative/Random/Direct access**. Each data item has a fixed size, so that it can be accessed directly.

```
007895211Elizabeth      Jones           2036785241
506874522Julian         Chapman         2039876455
385469983Ian            McAlister       2051123542
096538014Ruth           Hutchinson      2038900017
```

Plain text files

- **Indexed**. One or various external files sort the information in a direct access file based on one or various data items.



Database (DB)

Merriam-Webster dictionary:

”a usually large collection of data organized especially for rapid search and retrieval (as by a computer)”

C.J. Date (*An Introduction to Database Systems, 8th ed.*, 2004):

“A database is a collection of persistent data that is used by the application systems of some given enterprise”

Database (DB)

Data model. The architecture based on which data is stored and interrelated.

Types of databases (based on the underlying data model):

- **Hierarchical**. A relation of hierarchy is imposed upon the data. Obsolete.
- **Network**. More flexible and also more complex than the hierarchical model. Obsolete.
- **Relational**. Information is conceived as a set of entities and relations among them, all of them represented as bidimensional tables comprised of rows and columns. It has been the most popular data model since the 1970s. Examples: mySQL, SAP Sybase.
- **OO - Object-Oriented**. Emerging in the 1990s, it adapts the Object-Oriented paradigm into data modelling.
- **Object-Relational**. Hybrid model that incorporates OO elements into relational databases. Examples: Oracle, MS SQL Server, PostgreSQL, IBM DB2
- **Other**. Documental, Document-oriented, Multidimensional, Deductive

Database (DB)

Types of databases (based on information location):

- **Centralised**.
  The database exists in a single computer, typically the DB server.

- **Distributed**.
  The DB information is spread along different servers, usually geographically separated.

Database Management System (DBMS)

Abraham Silberschatz, Henry F. Korth & S. Sudarshan, *Database System Concepts, 6th ed.* (2011):

"A database-management system (DBMS) is a collection of interrelated data and a set of programs to access those data. The collection of data, usually referred to as the database, contains information relevant to an enterprise. The primary goal of a DBMS is to provide a way to store and retrieve database information that is both *convenient* and *efficient*"

Functions:

- Retrieving and modifying data in a user-transparent way
- Guaranteeing data integrity away from semantic inconsistencies
- Providing a programming language through which to interact with the data
- Supporting the Data Dictionary
- Solving concurrent access-related issues
- Managing transactions, executing related instructions as a whole
- Providing backup and restore utilities
- Providing security mechanisms to avoid illegal access

Database Management System (DBMS)

Components:

- **Data**. Stored in the operating system's file system.
- **Data access tools**.
  - A programming language with which to create, read, and modify the data.
  - A Data Dictionary to define and store metadata (information about the design of the database).
  - A CLI (Command Line Interface) through which to access these tools.
- **Utilities**. Additional tools for managing backups, statistics, scheduled tasks, users, access permissions...
- **GUI - Graphical User Interface**. A user-friendly alternative to the CLI.

Database Management System (DBMS)

(O)RDBMS ranking (db-engines.org, July 2017):

1.	Oracle	Oracle Corporation	Commercial
2.	MySQL	Oracle Corporation	Open Source
3.	MS SQL Server	Microsoft Corporation	Commercial
4.	PostgreSQL	PostgreSQL Global Development Group	Open Source
5.	DB2	IBM Corporation	Commercial
6.	MS Access	Microsoft Corporation	Commercial
7.	SQLite	D. Richard Hipp	Open Source
8.	Teradata	Teradata	Commercial
9.	SAP Adaptive Server	SAP Sybase	Commercial
10.	FileMaker	FileMaker Incorporated	Commercial
11.	MariaDB	MariaDB Corporation Ab	Open Source