Applied Databases Introduction

HIGHER DIPLOMA IN DATA ANALYTICS

What is data?

- Datum

  Single piece of information fact or statistic.

- Data

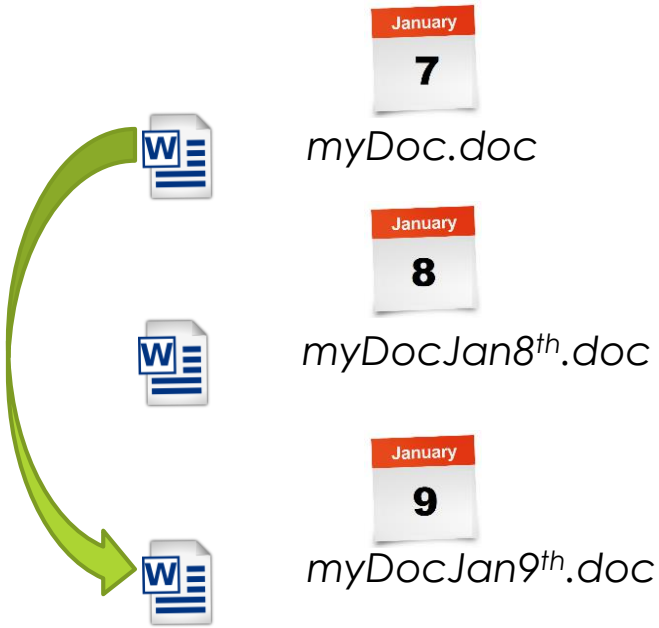
  A series of facts or statistics.

- Types of Data
  - Non digital information.
  - Digital Information
    - Active Digital Footprint
    - Passive Digital Footprint

Ever increasing data… per minute

- 120+ new professionals join
- 456,000 tweets sent
- 3.6 million searches
- 4.1 million videos watched
- 18 million forecast requests received

Database



- A Database is a collection of related data, organised in a way that data can be easily accessed, managed and updated.

Types of Database

- Relational Databases

- Non-Relational (NoSQL) Databases

Relational Databases

- A relational database consists of a set of tables used for storing data.
- A table is collection of related data
- Each table has a unique name and may relate to one or more other tables in the database through common values.

Relational Databases

- A table in a database is a collection of rows and columns.

  Tables are also known as entities or relations.

- A row contains data pertaining to a single item or record in a table.

  Rows are also known as records or tuples.

- A column contains data representing a specific characteristic of the records in the table.

  Columns are also known as fields or attributes.

Spreadsheets

	A	B	C	D	E
1	**PM**	**PM Phone**	**Project Name**	**Start Date**	**Budget**
2	Bill Jones	086 325 5689	OS Upgrade	Jan 1st 2018	15,000
3	Bill Jones	086 325 5689	Database Installation	Nov 23rd 2017	10,000
4	Mary Smith	086 325 5547	Payroll Modernisation	Jun 2nd 2017	5,000
5	Mary Smith	086 325 5547	HR GDPR Compliance	Nov 3rd 2016	5,500
6	Alan Murphy	087 558 6985	Car Park Repainting	Jul 1st 2015	4,450
7	Alan Murphy	087 558 6985	New Server Room Building	Mar 2nd 2016	25,200

Projects

Spreadsheets

	A	B	C	D	E
1	**Customer**	**Eircode**	**Product Bought**	**Date of Transaction**	**Amount**
2	Fred Jones	N37 T8P8	Philips 32" TV	Jan 10th 2019	500
3	Fred Jones	N37 R9Z8	iPhone	Dec 3rd 2017	600
4	Alice O'Neill	H91 K8F12	Philips 32" TV	Dec 18th 2018	500
5	Brian Collins	H92 L8L3	Dell Inspiron laptop	Nov 14th 2018	477.5
6	Orla McTiernan	H53 U3N9	Oral-B Electric Toothbrush	Apr 4th 2015	95.73

Orders

Database Schema

- A database consists of schemas, tables, views and other objects.
- A database schema represents the logical configuration of all or part of a database.
- It defines how the data, and relationships between the data, is stored.

Database Schema

- Two types of Schema:

  - Physical Schema

    Defines out how data is stored physically on a storage system in terms of files and indices.

  - Logical Schema

    Defines the logical constraints that apply to the stored data, the tables in the database and the relationships between them.

Logical Schema

- The Logical Schema is designed before the database is created.
- No data is contained in the logical schema.

Logical Schema

Patient Table
First_Name varchar(50)
Surname varchar(50)
Address varchar(200)
PPSN varchar(10)
Doctor varchar(50)
Doctor_Phone integer

Patient Table					
First_Name	**Surname**	**Address**	**PPSN**	**Doctor**	**Doctor_Phone**
John	Smyth	Athlone	7629913X	Dr. Jones	12345
Alan	Mulligan	Galway	9893333F	Dr. Murphy	88335
Fred	Collins	Castlebar	9898823W	Dr. Jones	12345

Logical Schema

Patient Table
First_Name varchar(50)
Surname varchar(50)
Address varchar(200)
PPSN varchar(10)
DoctorID integer

Doctor Table
DoctorID integer
Name varchar(50)
Phone integer

Patient Table				
First_Name	**Surname**	**Address**	**PPSN**	**DoctorID**
John	Smyth	Athlone	7629913X	100
Alan	Mulligan	Galway	9893333F	101
Fred	Collins	Castlebar	9898823W	100

Doctor Table		
DoctorID	**Name**	**Phone**
100	Dr. Jones	12345
101	Dr. Murphy	88335

Spreadsheet vs Database

	A	B	C	D	E	F
1	**First Name**	**Surname**	**Address**	**PPSN**	**Doctor**	**Doctor Phone**
2	John	Smyth	Athlone	7629913X	Dr. Jones	12345
3	Alan	Mulligan	Galway	9893333F	Dr. Murphy	88335
4	Fred	Collins	Castlebar	9898823W	Dr. Jones	12345

Patient Table

First_Name	Surname	Address	PPSN	DoctorID
John	Smyth	Athlone	7629913X	100
Alan	Mulligan	Galway	9893333F	101
Fred	Collins	Castlebar	9898823W	100

Doctor Table

DoctorID	Name	Phone
100	Dr. Jones	12345
101	Dr. Murphy	88335

Database Management System (DBMS)

- A Database Management System (DBMS) is software for creating and managing databases.
- The DBMS interacts with the user, the database itself, and other systems in order to store, retrieve and process data.

Database Management System (DBMS)

- The DBMS provides a centralized view of data that can be accessed by multiple users, from multiple locations, in a controlled manner.
- The DBMS can limit what data the end user sees, as well as how that end user can view the data, providing many views of a single database schema.

Employee Table
EID varchar(50)
Name varchar(50)
Salary varchar(200)
Next of Kin varchar(50)
Job Title varchar(50)

HR View

Employee Table
EID varchar(50)
Name varchar(50)
Salary varchar(200)
Next of Kin varchar(50)
Job Title varchar(50)

Project Manager View

Employee Table
EID varchar(50)
Name varchar(50)
Job Title varchar(50)

Database Management System (DBMS)

- The DBMS provides data independence, freeing users (and application programs) from knowing where or how the data is stored. Any changes in how or where the data is stored is completely transparent due to the DBMS.

- CRUD (Create, Read, Update, Delete) functions.

DBMS Functions

- Data Storage Management

- Security

- Backup and Recovery

DBMS Functions

- Transaction Management

- Debit Customer a/c
- Update Shipping Table
- Update Products Table
- Credit Store a/c

DBMS Functions

- Data integrity

Patient Table
First_Name varchar(50)
Surname varchar(50)
Address varchar(200)
PPSN varchar(10)
DoctorID integer

Doctor Table
DoctorID integer
Name varchar(50)
Phone integer

Patient Table

First_Name	Surname	Address	PPSN	DoctorID
John	Smyth	Athlone	7629913X	100
Alan	Mulligan	Galway	9893333F	101
Fred	Collins	Castlebar	9898823W	100
Mary	Connolly	Tuam	6789932A	200

Doctor Table

DoctorID	Name	Phone
100	Dr. Jones	12345
101	Dr. Murphy	88335
Dr. Kane	Dr. Kane	2314

DBMS Functions

- Concurrency

Advantages of DBMSs

- Controlling Redundancy

  Instead of each application having its own files with data stored multiple times, a centralised DBMS can store it once and allow many users to access it eliminating duplication.

	A	B	C	D	E	F
1	**First Name**	**Surname**	**Address**	**PPSN**	**Doctor**	**Doctor Phone**
2	John	Smyth	Athlone	7629913X	Dr. Jones	12345
3	Alan	Mulligan	Galway	9893333F	Dr. Murphy	88335
4	Fred	Collins	Castlebar	9898823W	Dr. Jones	12345

Patient Table

First_Name	Surname	Address	PPSN	DoctorID
John	Smyth	Athlone	7629913X	100
Alan	Mulligan	Galway	9893333F	101
Fred	Collins	Castlebar	9898823W	100

Doctor Table

DoctorID	Name	Phone
100	Dr. Jones	12345
101	Dr. Murphy	88335

Advantages of DBMSs

- Data Integrity
- Enforcement of Standards
- Backup and Recovery
- Security

Disadvantages of DBMSs

- Complexity
- Size
- Performance
- Higher impact of failure