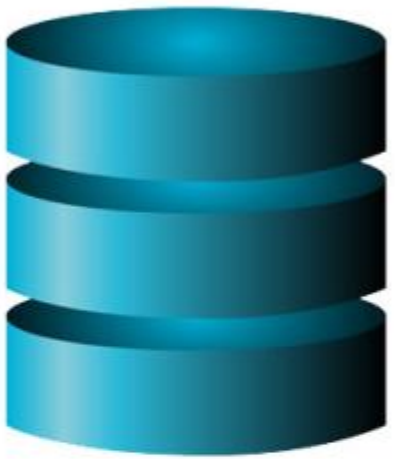




CODING ON DATA SCIENCE

MOHAMAD IRWAN AFANDI

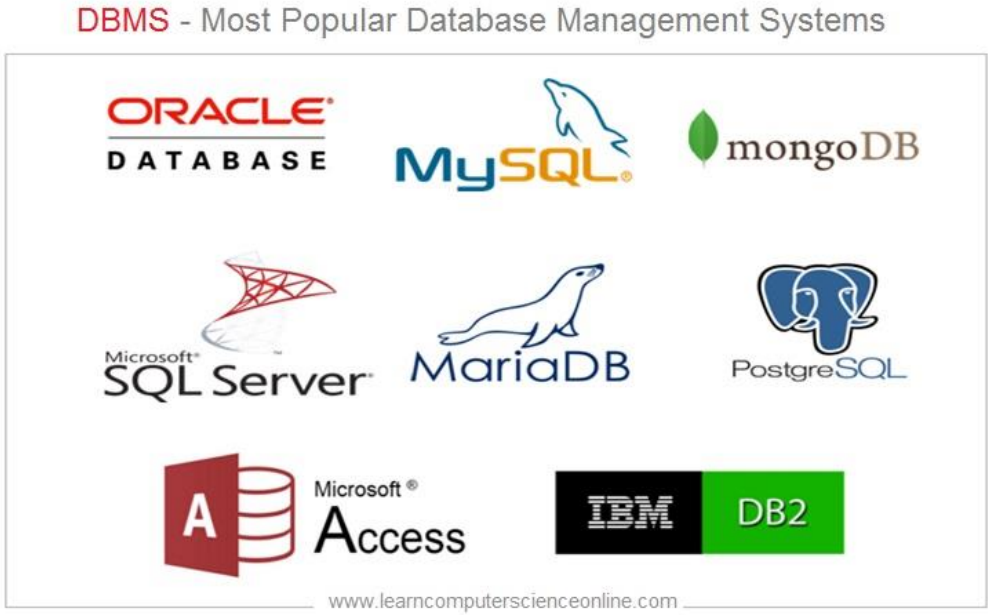
1

What is Database?

A database is a collection of information that is organized so it can accessed, managed and updated easily. Typically stored electronically in a computer system.

There are two types of database: **flat database** (single table structure) and **relational database** (multiple table structure).

For management the database, we used Database Management System **(DBMS).**



Why we need Database?

SIZE

How to handle thousand or million rows?

ACCURACY

How if someone entered Incorrect data or different format?

SCURITY

How to make someone can only access in certain of data?

REDUDANCY

How to handle copy of data in million rows?

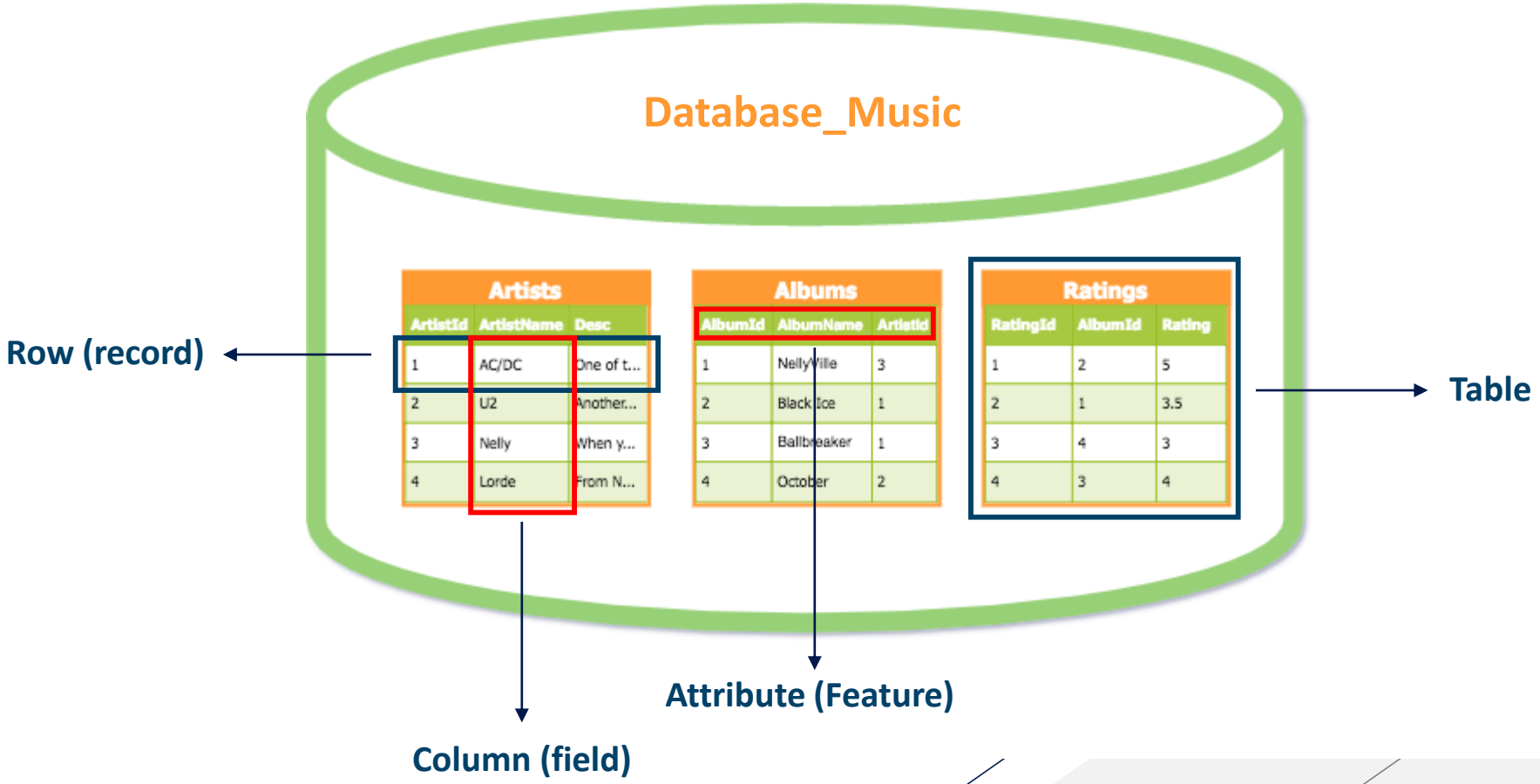
IMPORTANCE

How if the electricity suddenly goes of but you haven't saved your work yet?

OVERWRITING

How to handle if there are 100 people update the data at the same time?

Essential Table and Attribute Information



SQL (Structured Query Language)

SQL is a domain-specific language used in programming and designed for **managing data** held in *a relational database management system*, or for **stream processing** in *a relational data stream management system.*

Have you ever heard NoSQL? What is the different?

SQL used in Relational Data, NoSQL used in Not Relational Data (tabular, for example : Json file).

SQL COMMAND

DDL
(Data Definition Language)

1. Create
2. Alter
3. Drop
4. Rename
5. Truncate
6. Comment

DML
(Data Manipulation Language)

1. Select
2. Insert
3. Update
4. Delete
5. Marge
6. Call
7. Explain Plan
8. Lock Table

DCL
(Data Control Language)

1. Grant
2. Revoke

TCL
(Transaction Control Language)

1. Commit
2. Rollback
3. Savepoint
4. Set Transaction

SQL Pattern:

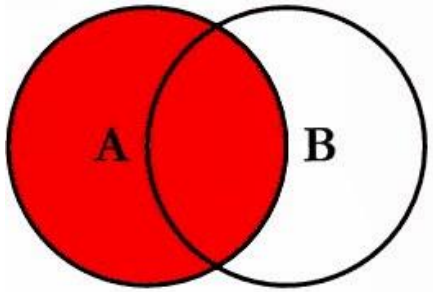
Create table tb_name();

Select column_name **from** tb_name **where** condition **group by** column_name **having** condition **order by** column_name;

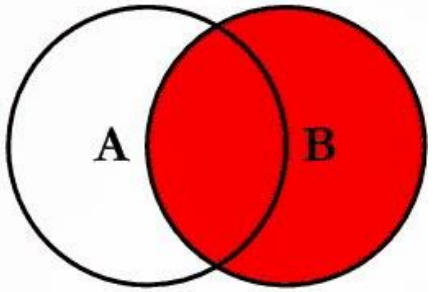
Update tb_name **set** column_name = value **where** condition;

Delete from tb_name **where** condition

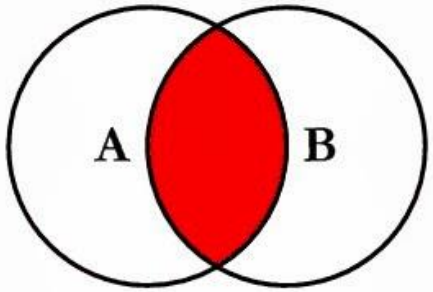
SQL JOINS



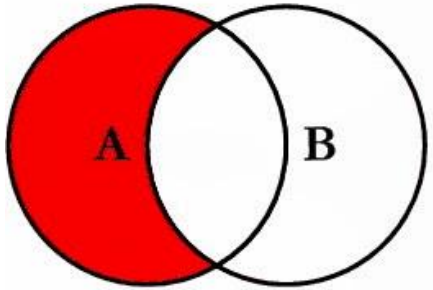

```
SELECT <select_list>
FROM TableA A
LEFT JOIN TableB B
ON A.Key = B.Key
```



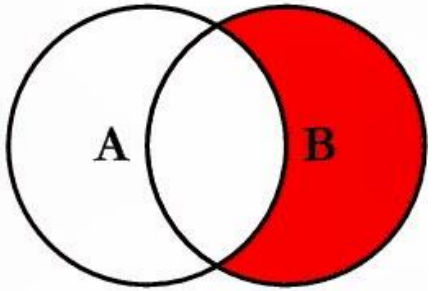

```
SELECT <select_list>
FROM TableA A
RIGHT JOIN TableB B
ON A.Key = B.Key
```



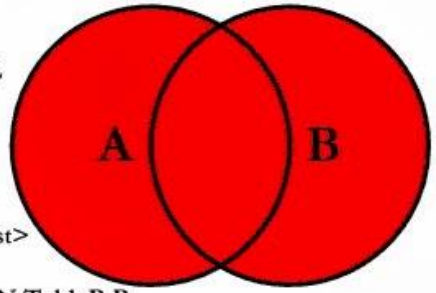

```
SELECT <select_list>
FROM TableA A
INNER JOIN TableB B
ON A.Key = B.Key
```



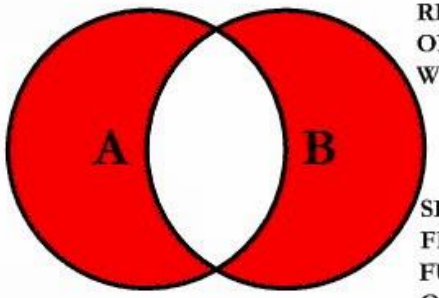

```
SELECT <select_list>
FROM TableA A
LEFT JOIN TableB B
ON A.Key = B.Key
WHERE B.Key IS NULL
```




```
SELECT <select_list>
FROM TableA A
RIGHT JOIN TableB B
ON A.Key = B.Key
WHERE A.Key IS NULL
```




```
SELECT <select_list>
FROM TableA A
FULL OUTER JOIN TableB B
ON A.Key = B.Key
```




```
SELECT <select_list>
FROM TableA A
FULL OUTER JOIN TableB B
ON A.Key = B.Key
WHERE A.Key IS NULL
OR B.Key IS NULL
```

Are you bored?

Let’s start coding

2

What is Python?

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace.

What can python do:

- Web development (server-side API, Django)
- Software development (create workflows)
- Can connect with database (psycopg2)
- Handle big data and perform complex mathematics
- System scripting
- Machine learning

Why we use Python?

WORKING

It works in different Platforms (Windows, Mac, Linux, etc)

SINTAX

Simple and similar to the English Language

EFFICIENCY

Write programs with fewer lines then other

RUNNING

Code can be executed as soon as it is written, means that the prototype can be very quick

IMPORTANCE

Python can be treated in a procedural, object oriented or functional way

DATA TYPE

Boolean

True / False

```
if (number % 2) = 0:
        even = True
else:
        even = False
```

Numbers

Integers, Floats, Fractions and Complex Numbers

```
a = 5
b = 7.3
c = 2 + 3j
```

Strings

Sequences of Unicode Characters

```
s = “This is a string”
```

Bytes & ByteArray

Contain Single Bytes

```
b = ‘A\nB\nC’
```

Lists

Ordered sequences of values

```
a = [ 1, 2.2, “Python”]
```

Tuples

Ordered immutable sequences of values

```
t = [ 2, “Tuple”, “95”]
```

*use bracket ()
not square bracket []

Sets

Unordered bags of values

```
week = {‘Mon’, ‘Tue’,
‘Wed’, ‘Thu’, ‘Fri’, ‘Sat’,
‘Sun’}
```

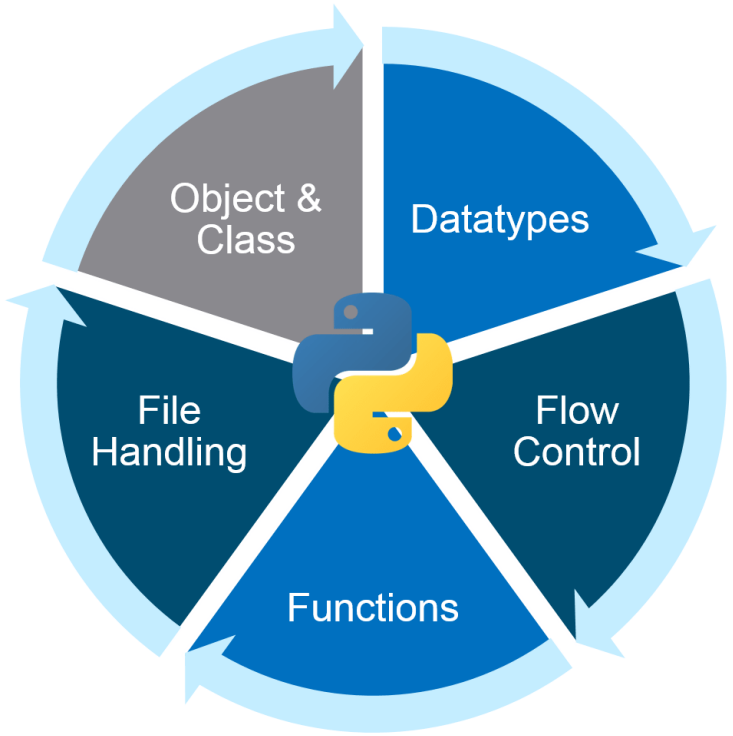
Dictionaries

Unordered bags of key-value pairs

```
d = {‘value’:5, ‘key’:125}
```

*lists are mutable, tuples are immutable

FUNDAMENTAL



FUNDAMENTALS

Datatypes	Flow Control	Functions	File Handling	Object & Class
Numbers Strings Lists Dictionaries	If Else For While Continue	Definition Function Call Docstring Return	Reading Writing Editing	Variables Functions

Need stretching?

Let’s start coding

3

What is Shell Scripting?

A shell script is a computer program designed to be run by the Unix/Linux shell which could be one this : ***bourne, C, Korn and GNUBourne-Again Shell.*** A shell is a command-line interpreter and typical operations performed by shell scripts include file manipulation, program execution, and printing text.

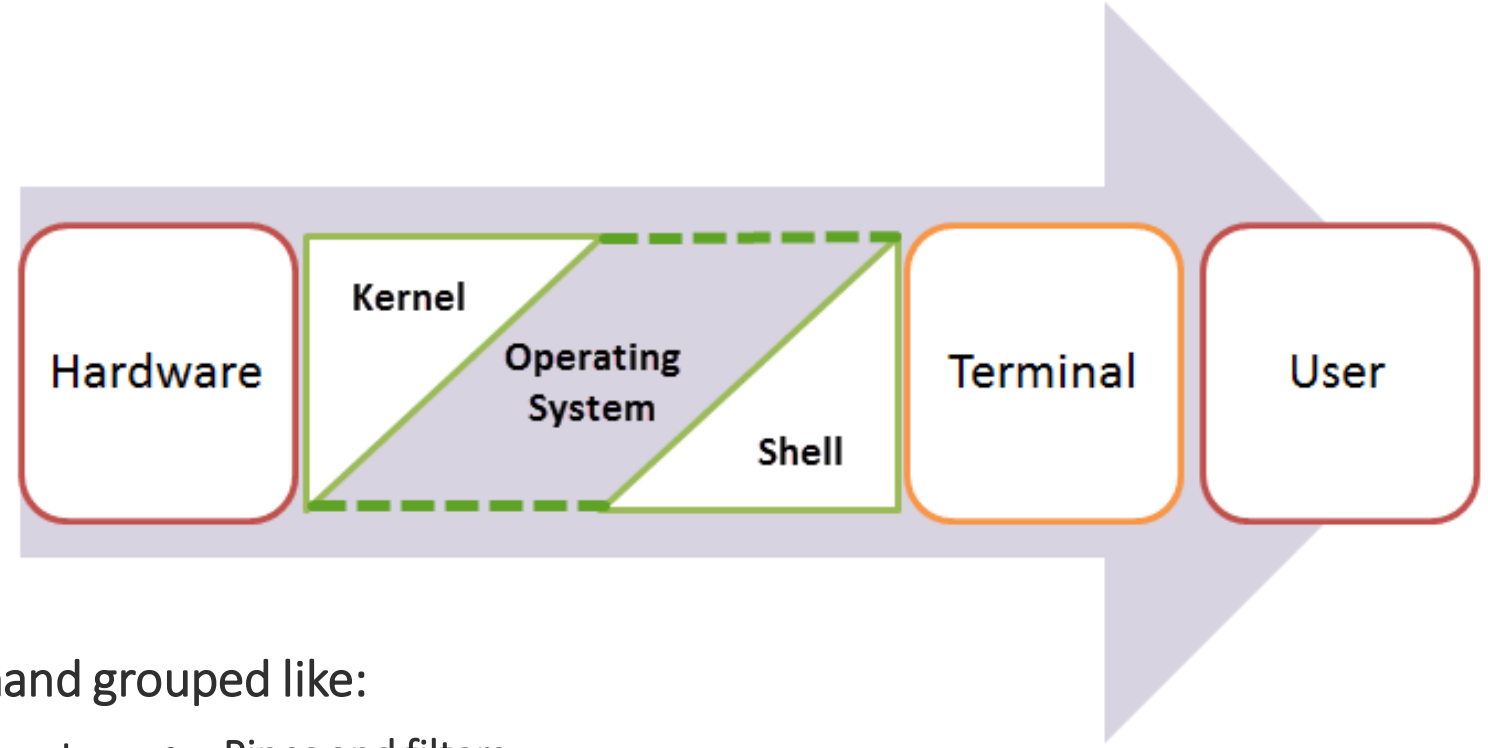
What Shell Script does, especially in Data Science:

- For communicate with server
- Running model on server
- Scripting instruction, like cron-job

PowerShell

Terminal

How Shell works?



Shell Script Command grouped like:

- File management
- Directories
- File Permission
- Pipes and filters
- Environment
- Vi editor, etc

BASIC COMMAND

Management File

- **ls** → show list the file and directories stored in the current dirrectory
- **ls-l** → same with ls but more detail (user action)
- **vi** file_name → create/edit file
- **cat** file_name → see the content of file
- **wc** file_name → counting the number of line, word and character in a file
- **cp** file_name copy_filename → to make a copy file
- **mv** old_file new_file → to rename the file
- **rm** file_name → to delete the file

Directories

- **cd** ~ → go to home directory
- **cd** ~username → go to other user's home dir
- **pwd** → to see your current location
- **mkdir** dir_name → crite new directory
- **mkdir** /dir_name/ new_dirname → to create new directory inside the dir_name
- **rm** dir_name → delete directory
- **mv** old_file new_file → to rename the file
- **cd** dir_name → go to dir_name
- **cd** /dir_a/dir_b/dir_c → go to dir_c
- **cd** ../ → back to previous dir
- **mv** old_dir new_dir → rename the directory

Thank you …

Reference

https://medium.com/omarelgabrys-blog/database-introduction-part-1-4844fada1fb0

http://www.sqlcourse.com/intro.html

https://searchdatamanagement.techtarget.com/definition/NoSQL-Not-Only-SQL

https://desktop.arcgis.com/en/arcmap/latest/manage-data/tables/essential-table-and-attribute-information-vocabulary.htm

https://javarevisited.blogspot.com/2012/11/how-to-join-three-tables-in-sql-query-mysql-sqlserver.html

https://www.w3schools.com/python/python_intro.asp

https://www.edureka.co/blog/python-programming-language

https://www.tutorialspoint.com/unix/index.htm

Access my video for the full version