Data Analysis Model

Power by Open Source Platform

ubuntu · Hadoop · HIVE · Power BI · aws · CLOUDERA

1M Data

Data Set:
01.Healthcare: [Record–46935]
02.Weather-history:[Record–4573]
03.World Demography
04.Census Tracts 2010:[Record-21]
05.Animal_Services_Intake_Data
06.Average_Daily_Traffic_Counts
07.Accidental_Durg_Related_Death
08.Popular_Baby_Names:[Record
09.Sales_Tax_Rates:[Record-1911]
10.Restaurants:[Record-1328]
11.Accidental_Durg_Related_Death
12.Census Tracts 2010:[Record-216]
13.Employees_Salary:[Record–824]
14.Customer_transactional_spending
15.Customer_Order:[Record–1000]
16.Employees_Salary:-[Record–824]

Web, App Cloud, Database, Excel, Notepad, etc.

1 Data Sourc
Data source is the primary location from where **data** comes. The **data source can** be a database, a dataset, a spreadsheet or even hard-coded **data**

2 Data Lake
A **data lake** is usually a single store of all enterprise **data**

Excel/CVS/Text File Format:
John,30,New York
Robert,25,DC

JSON, Web Application File Format : {
"employee":{
"name":"John", "age":30, "city":"New York" }
}

Text File/Sequence File
RC File/ORC File
AVRO File
Parquet/Binary File
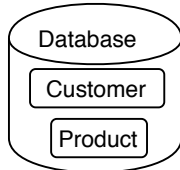
ETL/ELT, sqoop import-all-tables

3 Data Pipeline
a pipeline, also known as a data pipeline, is a set of data processing elements connected in series

4 Data Store
A *Data Store* is a connection to a store of data, whether the data is stored in a database or in one or more files

Linux O/S ,
Hadoop Big Data,
Hadoop File System(HDFS)
3 replication Datanode

Database
Customer
Product

Hive Database,
Data Warehouse

5 Data Warehouse
A **data warehouse** is a system that pulls together **data** from many different sources

6 Data Processing
Data processing is, generally, "the collection and manipulation of items of data to produce meaningful information

SELECT * FROM Customers;

SELECT Country FROM Customers;

SELECT * FROM Customers
WHERE Country='Mexico';

-**Top** 10 Revenue-
generating products
-**Total** revenue per product
- **Top** 10 Diseases
-**Weather Analysis**

7 Data Revenue
The statistic shows the *revenue* from the global big *data*

8 Data Visulization
Data visualization is the graphical representation of information and **data**

Version: v.001.2020