RANKA
DATE / /
PAGE

Data Essentials

* Data, Information, Knowledge, And Wisdom
  - Raw and unstructured facts
  - contextualized, organized trend or pattern
  - Ability to solve a problem
  - Decision making and judgments

* Sources of Data
  - Experimental Data - clinical trials
  - Observational Data - recorded through observations
  - compiled Data - Manual data aggreated overtime and managed to database.
  - Simulated Data - result of studying certain behaviors
  - Human Generated Data - generated through mobilephones, Web, audio, socialmedia
  - Machine Generated Data - Digital Devices, IoT
  - Business Generated Data - stock exchange, Banks etc.

* common Types of Data

2main types
* Quantitative Data - measurable such as numbers and values, conclusive
* Qualitative Data - Not measurable; descriptive. (unstructured data)
* Geospatial Data - Earth data, GPS system
* Digital data - images, Audio, video, Web based
* Data and from Documentation and scripts - HTML, opentext document.

* Data Formats
  - Structured Data - Data having a defined format, model, structured
    Easy to search and analyze Ex: csv, XIS
  - Semi-structured Data - Data having flexible schema
    schema maybe descriptive, partial or evolving Ex: XML, JSON, Gmail
  - Unstructured Data - No structure, Format or model Ex: Pdf, Jpeg, mp4

* Data terminologies
  - Primary Data - collected by a person or group by themselves.
  - Secondary Data - collected by a third party

Data Analytics - Exploring Data

Data science - uses scientific methods tools to extract knowledge and insights from the data.

RANKA
DATE / /
PAGE

* Database - Repository of data

Data Analytics - process/technology used to explore data

Data Aggregation - procedures to gather data

Metadata - details about data

Time series - analysis studying temporal patterns.

* Anonymized Data - Individuals identities are kept secret

Ex: Feedback form

Augmented Reality - Data analytics utilizing ML and NLP

Data literacy - reading and communicating data as information

File format - xls, .csv

Data science -

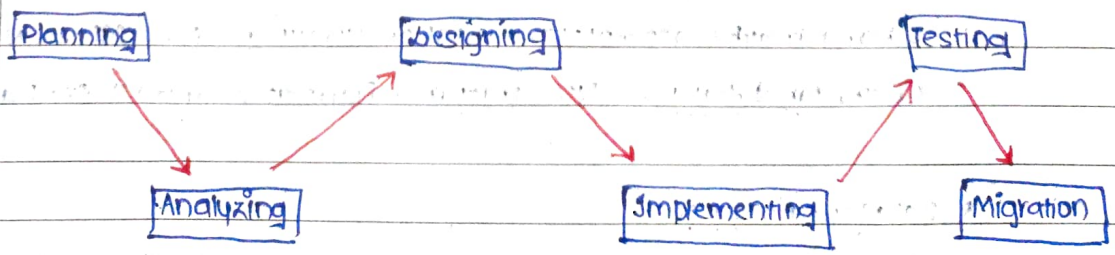
Deep learning - Neural network based ML Algorithms

AI - The ability of machines to demonstrate human like behavior

Data Storage and Backup

Full Backup	Differential Backup	Incremental Backup
Backup for all data	Backup of revised of last backup	only revisions or changes to previous
requires high space	medium to high storage	small storage
slowest speed	Fast speed	fastest speed
Lots of duplicate data	Intermediate amount of duplicate data	little or no duplicates

Data Migration : Transfer of data from one storage system to another



RANKA
DATE / /
PAGE

Planning

Identifying stakeholders

Identify data

Risk, Mitigation and Backup strategy

Analyzing

Explore the data

determine impact, migration cut-off point

Create a data dictionary

Design

Develop "source to target" mapping

Determine ETL Transformations

Implementation

configure tools

Write migration scripts

Testing

Deploy all required tools

create test plan with accurate data coverage

design a migration validation engine

Migration

Execute migration steps

Execute proper planning

go live including migrat

ETL

Extract	Transform	Load
retrieves and verifies data	processes and organizes extracted data	moves transfo data to reposit

Data visualization - representing data in the form of charts, graphs and maps.