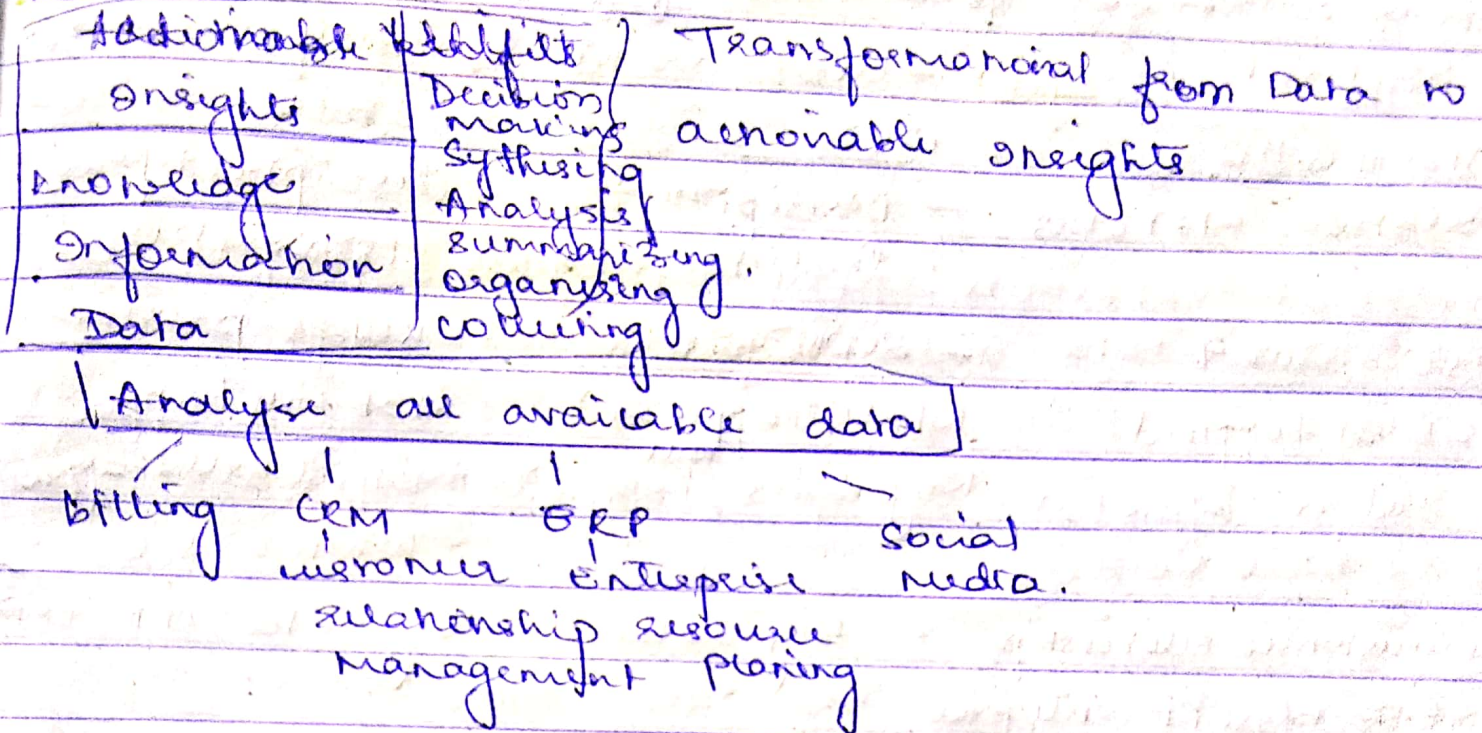


The term Business Intelligence (BI) refers to technologies, applications and practices for the measuring, collecting, analyzing, summarizing and organizing of business information. The purpose of Business Intelligence is to support better business decision making. Essentially, Business Intelligence systems are data-driven Decision Support Systems (DSS).

Unit - V

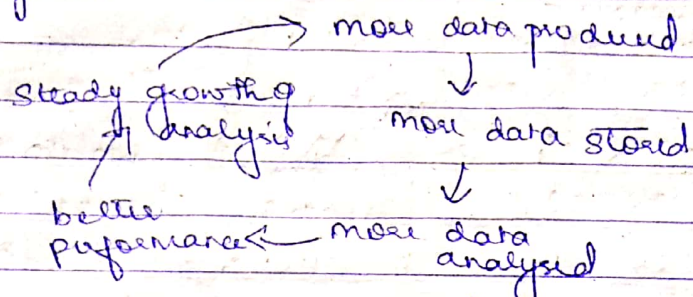


What is big data Analytics?

- i) technology enabled Analytics - tools to analyse data & visualization techniques.
- ii) gaining deeper insight into business or know customer behaviour.
- iii) better decision making
- iv) handshake between 3 communities - IT, business users & data scientists

- i) size of i/r dataset, excess storage & processing capacity
- ii) moving from code to data

What is big data entries entries?



classification of analysis — first school of thought

i) — those who classify analytics based on operational basic advanced.

ii) analytics classified as analytics 1.0, 2.0, 3.0

i) a) Basic Analytics — historical data, basic visualisation are provided to improve business.

b) operationalised analytics: various operations performed

c) advanced analytics: predictive modelling, ^{prescriptive} modelling — for predictions of future

d) monetized Analytics: drive direct business reviews

ii) Analytics 1.0

— descriptive statistics

— 1950 — 2009

keys — what happened?
why did it happen?

— CRM (customer relationship management, ERP (enterprise resource planning)

— data was initially stored

— relational DB

2.0

— Descriptive + Predictive statistics

— 2005 — 2012

what will happen?
why did it happen?

— externally
— Hadoop, DB application & Hadoop cluster

3.0

— descriptive + predictive + prescriptive

— 2012 — present

what will happen
why it will happen
& when will it happen

— Big data, CRM, ERP

— both

— in memory analytics
— in DB processing & few ML techniques

greatest challenge that prevent business from capitalising on big data

it obtaining executive sponsorship for big data related activities

- i) getting business units to share data across organisations
- ii) finding right skills that can manage large amount of structured, unstructured, semi-structured data
- iii) determine approach to scale rapidly
- iv) deciding whether to use structured/unstructured / internal data for making decisions
- v) choosing right way to report findings & analysis
- vi) Determining what to do with insights found with big data.

Top challenge facing big data

- 1) Scale - i/p data analysing
- 2) Security - i/p security of data
- 3) Schema - should be flexible - dynamic schema
- 4) Continuous Availability - 24/7 availability
- 5) consistent data
- 6) Partition tolerant
- 7) Data Quality - Accuracy, completeness & timeliness

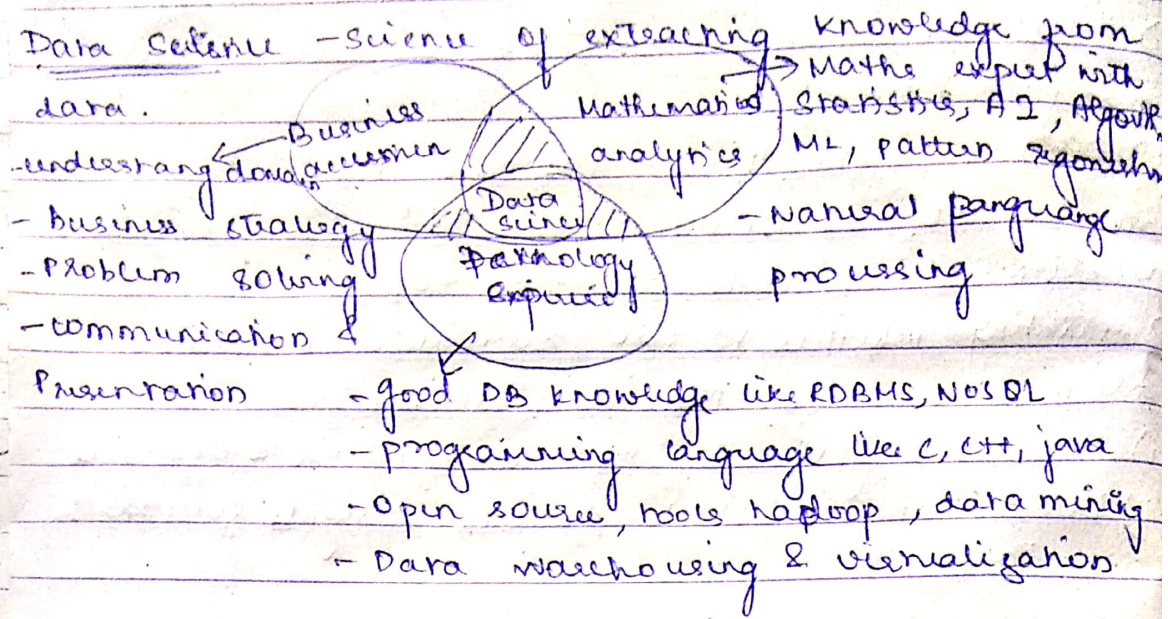
why is Big data Analytics important?

- 1) Reactive Business Intelligence - data needed by business is provided to right person, at right time & right place.
- 2) Reactive Big data Analytics - Analytics is on static data which is huge amt of data & find insights
- 3) Proactive analysis - decision making process, text mining, statistical analysis, predictive modelling etc
- 4) Pro active Big data Analytics - filtering data on TB, PB and HB

Technologies needed to face challenge of Big data

- 1) abundant, storage, cheap
- 2) fast processor for processing large amt of data
- 3) open source, distributed platform like hadoop

- 4) Parallel processing, clustering, virtualization, high computing
- 5) cloud computing & other resource allocation strategies



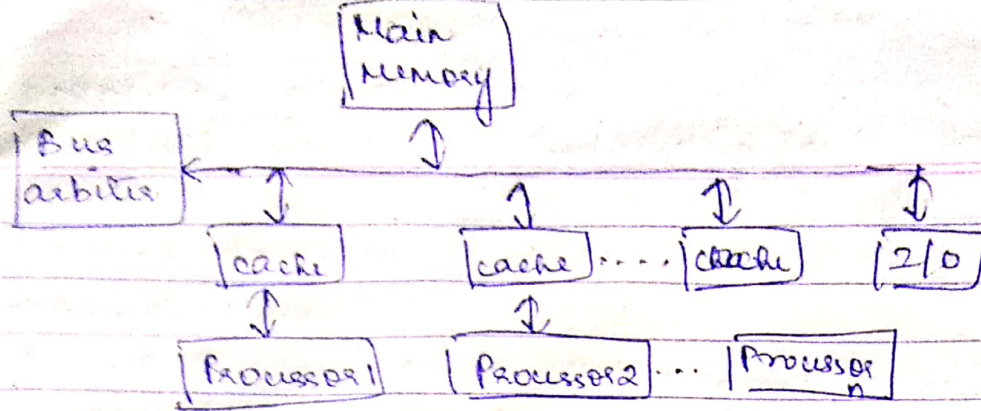
Data science is the process of collecting raw data, processing data, integrating data, analysis using different algorithms, presentation, communicating finding to stakeholders & make faster & better decision.

Responsibilities of Data scientist

- 1) Data Management - raw data is preprocessed and used for analysis. useful relationships are retrieved.
- 2) Analytical techniques : apply algorithms & models to understand data and interpret relationships among entities.
- 3) Business Analysis : insights are communicated with users / stakeholders

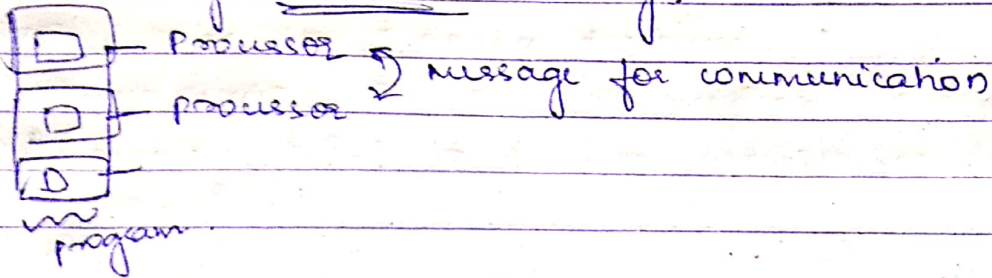
SMP Technologies used in Big data Env

- 1) On - Memory Analytics : Data is stored in RAM therefore it can be used when required advantages are faster access and high performance
- 2) On - Database processing : computations are done within database rather than loading it each time it is required like OLTP
- 3) Symmetric multiprocessor System (SMP)

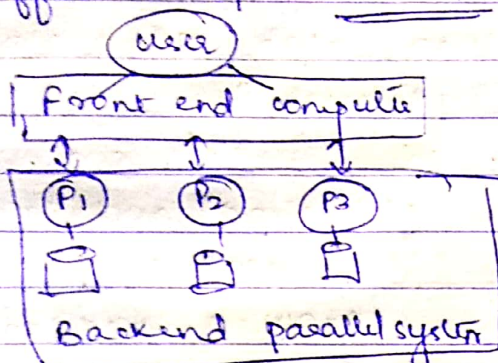


- is a tightly coupled system

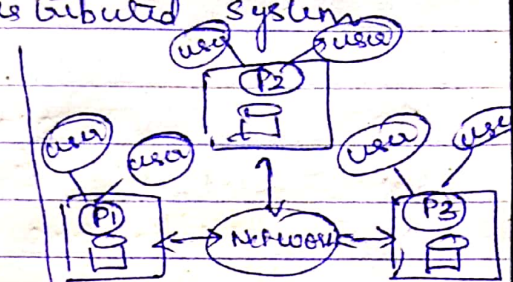
4) Massively Parallel Processing:



5) difference b/w Parallel & Distributed system



A common memory for all processors



each & every process have different memory & direct access to it

6) Shared Nothing Architecture

Shared memory (SM) - all processor share same memory

Shared disk (SD) - all processor share same disk

Shared nothing (SN) - nothing is shared.

CAP Theorem - consistency - reads history
 - Availability - read & write is always successful
 - Partition tolerance - system continues to work even if there is network partition

Eg: real time