Characteristics of Big data :-

(i) Volume - Files are large (in no. of files & sizes)

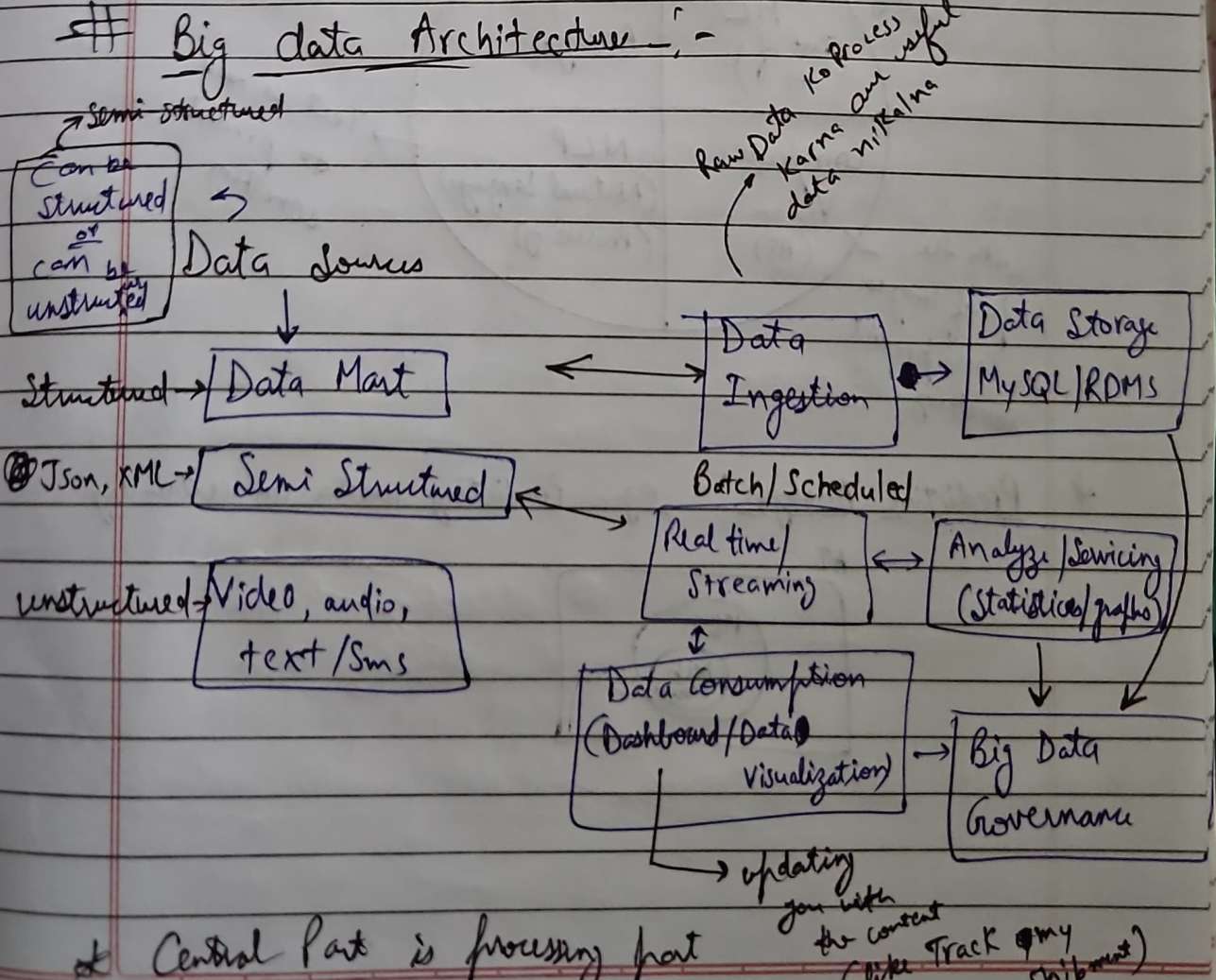
(ii) Varacity - Uncertainity of data.

(iii) Variety - Data is stored from different sources. Data can be video, text, etc.

(iv) Velocity - The speed with which data is collected.

(v) Value - The data that is to be collected is useful.

Big data Architecture :-



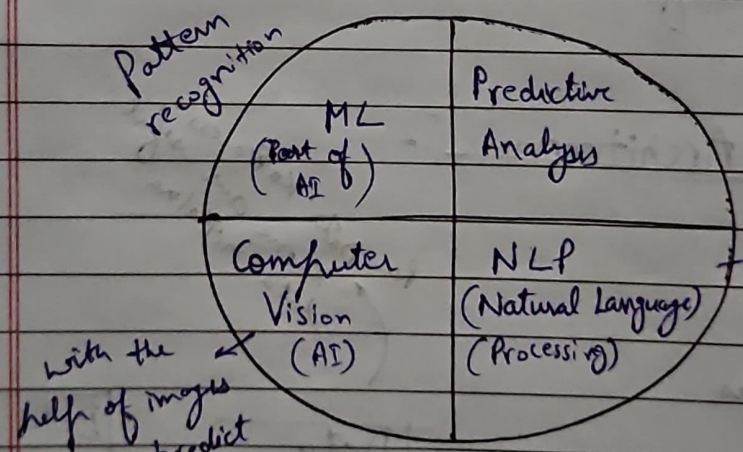
* Central Part is processing part

→ Attribute is also called ~~Feature~~ Feature in ML

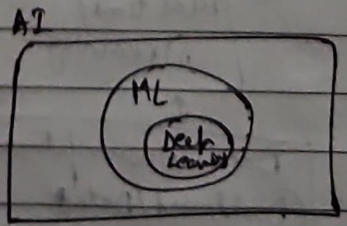
Technology Components :-

(i) Data Capture
(ii) Data Storage
(iii) Data Processing
(iv) Data visualization

* Imputation – Missing values/data ko fill karna, like mode bhar diya.
↓
Highest Frequency.



* Predictive Analysis :- eg. Predicting weather on certain day



PAGE NO.
DATE / /

Technology :-

(i) Apache Hadoop - It is an open-source framework.
eg. 10 years ago the facebook followed.
It enables the distribution of large scale dataset & it is flexible.

(ii) Apache Spark :- This can be used with Apache Hadoop.
It is an open-source processing engine.

(iii) Apache Flink :- It is an open-source stream processing framework.
It is High analytics
It is user friendly API

(iv) Presto :- It is an open source sequal engine that ~~suffers~~ supports interactive analyses on Huge dataset.

(v) Druid :- It is an open source analytical data storage design for queries on event based data eg Log files.
(Transaction)

Other Technology used :-
(i) Map Reduce
(ii) Cloud Era
(iii) Horton Works
(iv) IBM Big Insights
(v) Oracle Big data Appliance.

★ Features of Big Data :-

(i) **Data Preparation :-** It is used before the visualization model.

* It is used during model construction.

(ii) **Data Exploration :-** Visualize insights through pictorial representation.

ex Stock Market

(iii) **Scalability** – Efficient Energy Construction.

It should use less Network layer.

(iv) **Supports** for various types of analytics.
↳ Graph, chart, reviews

(v) **Version Control :-**

For every versions, the checks are done.
Previous code should be compatible with new version.

(vi) **Data Management :-**

↳ storing, using data, also bringing Security with cost effective way.

(vii) **Data Integration :-** Collecting different data and integrating them.

(viii) **Big Data Governance :-** Accurate, usable, reliable

↓

Data should be

PAGE NO.
DATE / /

★ Features of Big Data :-

(i) **Data Preparation** :- It is used before the iterative model.

* It is used during model construction.

(ii) **Data Exploration** :- Visualize insights through pictorial representation.

ex Stock Market

(iii) **Scalability** – Efficient Energy Construction.

It should use less Network layer.

(iv) Supports for various types of analytics.
↳ Graph, chart, reviews

(v) **Version Control** :-

For every versions, the checks are done.
Previous code should be compatible with new version.

(vi) **Data Management** :-

↳ storing, using data, also bringing security with cost effective way.

(vii) **Data Integration** :- Collecting different data sets and integrating them.

(viii) **Big Data Governance** :- Accurate, usable, reliable
↓
Data should be

(iv) Visualization

Applications :-

(i) Monitor User Behaviour

(ii) Recommendation :-

Audits :-

* Advantages of using the big data in auditing system -

1. Reduction in operational cost.
2. Improved decision making.
3. High customer retention.
4. Higher satisfaction rate.
5. Banking → Manufacturing

* If the data is real-time then the growth of the organisation would be very high and sustainability would be high.

* In Big data to give us most accurate results, we should embed

- AI
- Automation

] → Accuracy will be maximum.