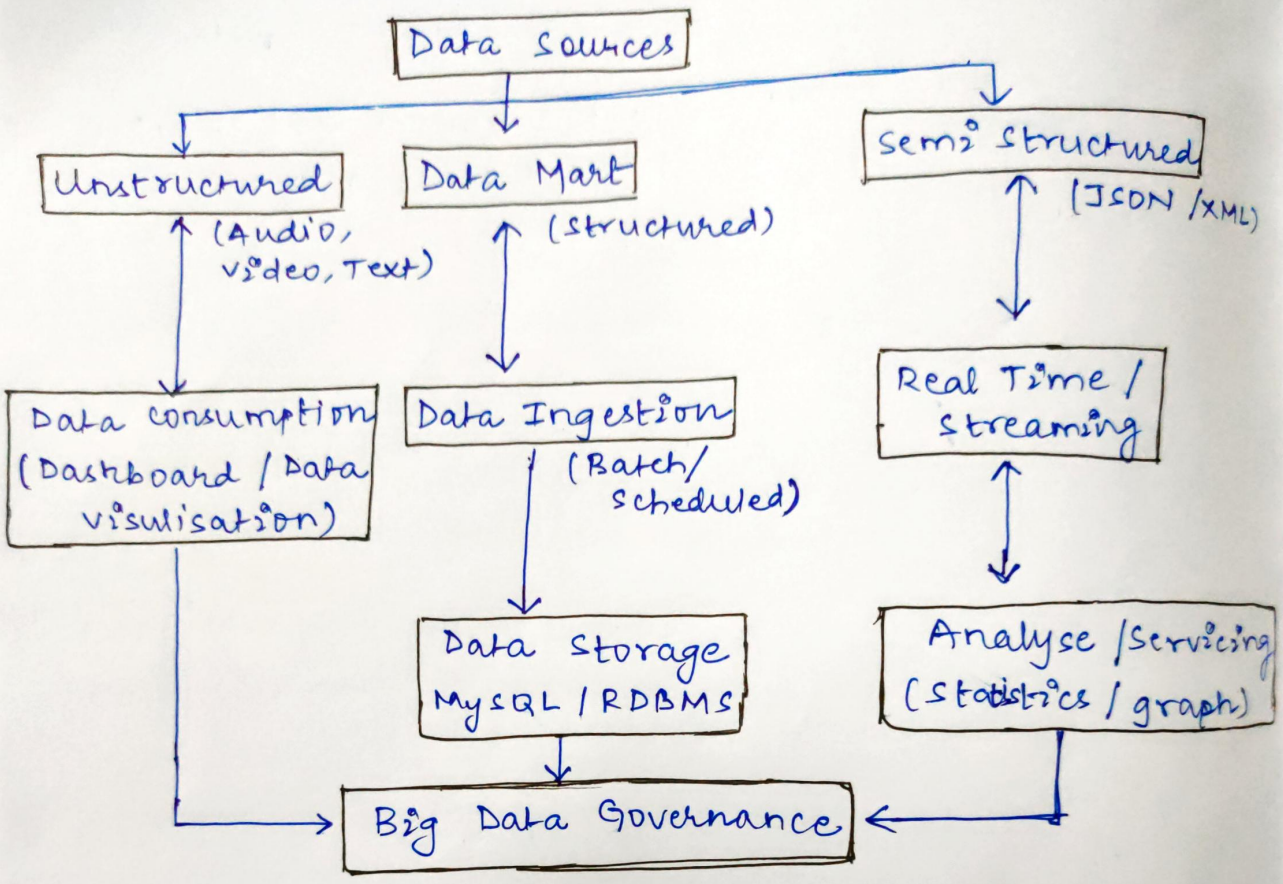
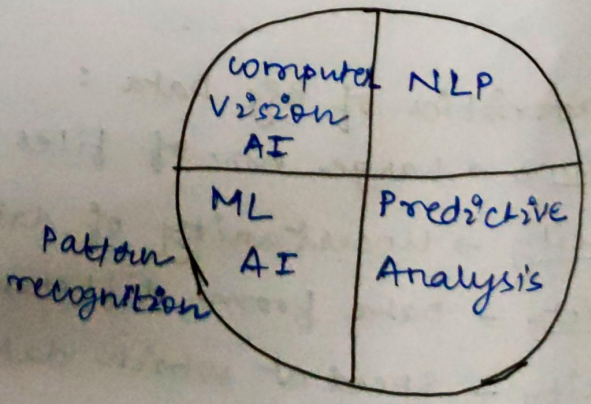
* Big Data Architecture :-



* components of Big Data :-

1. Data capture
2. Data storage
3. Data Processing
4. Data Visualisation



* Outliers are the data which ~~does~~ is not correct.

* component technologies of Big Data :-

① Apache Hadoop → Open source framework. Used in distributed computing.

② Apache Spark → Open source processing engine can be used with Hadoop.

③ Apache Flink → Open source stream processing framework. Used for live streaming of real time data.

④ Presto → Open source SQL Engine that supports interactive analysis on huge data set.

⑤ Druid → Open source analytical data storage ~~query~~ ~~bas~~ designed for query on event based data.

⑥ Map Reduce

⑦ Cloudera

⑧ IBM Big Insights

⑨ Hourton Works

⑩ Oracle Big Data Appliance.

* Features of Big Data :-

① Data Prepration → used during model construction but before data iteration.

② Data Exploration → visualing things based on given data ~~as~~ maybe using pictorial represent -ation.

③ Scalability → As size of data increases the resources should also be able to handle this.
For.e.g → millions users on social media at same time.

However, cost effectiveness and low power consumption should also be considered without compromising efficiency.

④ Supports for various types of analytics → We can use various analysis techniques such as graphs, tabular form, pie chart, star ratings, text ratings.

⑤ Version Control → With updation in technologies there should be a side by side adaption for the new technologies. Previous versions of code should work with new tech.

⑥ Data Management → Capture, store and using data with bringing security into picture. Basically managing the huge amount of data

⑦ Data Integration → collecting data from different sources and integrating them to retrieve some useful information.

⑧ Data Governance → The data should be available, reliable, accurate and easy to use. Thats how we can govern the data Access control, encryption and decryption.
Intrusion detection system or firewalls can be used to secure the big data.

⑨ Visualisation → Data driven environment.

Applications of big data :-

1. Customer shopping behaviour
2. Recomendation
3. Healthcare
4. Traffic control system

*Auditing and Analysis :-

1. Reduction in operational cost
2. Improved Decision Making
3. High customer retention.
4. High satisfaction rate
5. used in various fields like banking, manufacturing--

* Hadoop Features :-

→ Provides access to file system
→ Hadoop common package contain JAR files and scripts
→ Package also provides source code, documentation, contribution section.

* Hadoop Architecture :-