NAME :- DIVYANG BAGLA
PANEL :- C PC-33
BATCH :- C-3

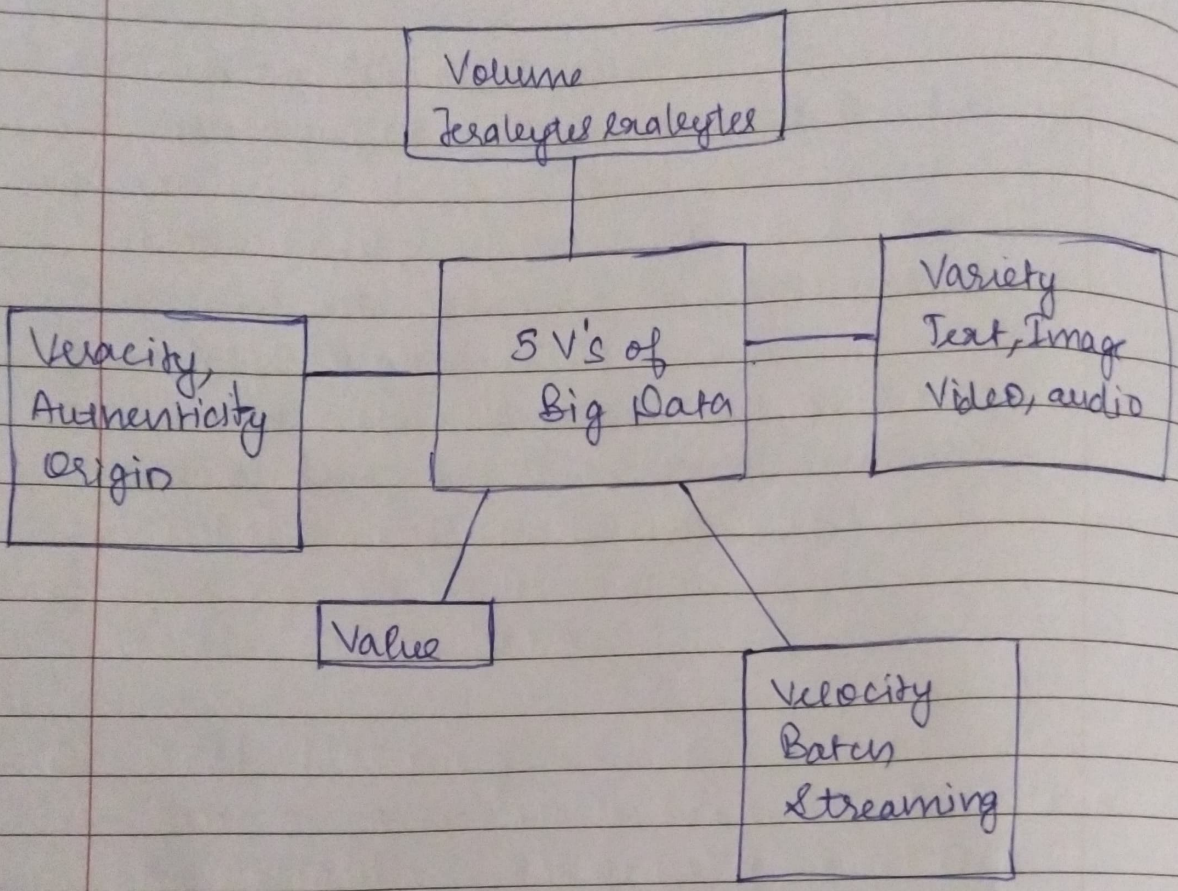
FINAL YEAR B.Tech

PAGE NO:

BDA

BDA THEORY ASSIGNMENT - 1

Q.1 Explain 5V's of Big data along with suitable diagram.



1) Volume :-
- Volume indicates huge volume/quantity of data.
- To distance/determine the value of data, size of data plays a crucial role. Large volume of data is Big data

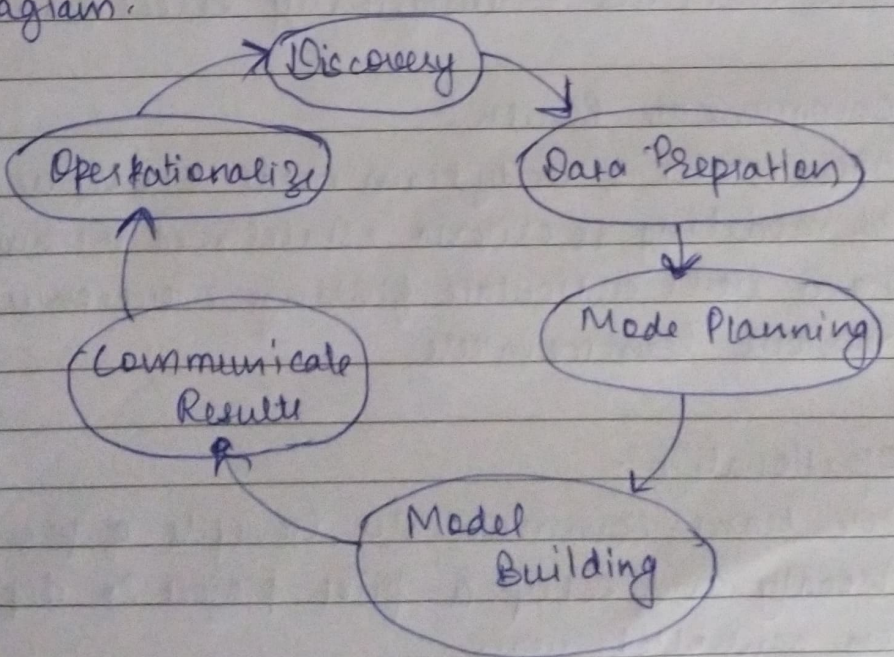
2) Velocity :-
- Velocity refers to high speed of accumulation of data
- In Big data, velocity data flows in from sources like machines, networks, social media etc.
- There is a massive & continous flow of data.

3) Variety :-
- It refers to nature of data that is structured, semi-structured and unstructured data.
- It refers heterogenous sources.

4.) Veracity :-
- It refers to inconsistences and uncertainity in data. that is data which is available can sometimes get messy.
- Big data is also available due to multitude of data.

5) Value :-
- After having the 4V's into account there comes one more V which stands out for value.
- Data itself is of no use or importance but it needs to be converted into something valuable to extract informations.

Q2 Elaborate all stages in data analytics life with neat diagram.



a) Discovery :-
- The data science team learn learn and investigate the problem
- Develop context and understanding.
- Get available data sources

b.) Data preparation :-
- Steps to explore, preprocess and condition data prior to modeling and analysis
- Tools :- Hadoop, Alpine, Open Refine etc.

c) Model ~~planning~~ building :-
- Team develops datasets for testing, training & production purposes
- Team also considers if existing tools will suffice for running the models
- Tools :- Octave, WEKA.

d) Model Planning :-
- Team explores data to learn about relationships b/w variables & subsequently select features.

e) Communicate Results :-
- After model is ready, team needs to compare outcomes of modeling to criteria established for success & failure.
- Team will articulate findings & outcomes to various team members.

f.) Operationalize :-
- The team communicates benefits of project more broadly and setup a pilot project to deploy work in controlled way.

PAGE NO.:

Q3 Explain BASE property for No SQL and ACID property for Relational Database?

a) NO SQL relies upon a softer model known as BASE model.

1) Basically available :- guarantees availability of data.
2) Soft state :- The state of the system could change over time.
3) Eventually consistency :- The system could change/ eventually become consistent.

b) RDMS follows ACID :-

1) Atomicity :- An all or nothing approach.
2) Consistency :- The transaction must meet all protocols.
3) Isolation :- No transaction has access to others.
4) Durability :- commited transaction is preserved via backups.

Q4 CAP theorem states that only two out of these properties C, A and P are possible to achieve if applied. make use of any example and explain it using diagram.

a) Consistency : like C in ACID.
b) Availability :- Resources should be always available.
c) Partition tolerance :- No single point of failure.