Vidyavardhini's College of Engineering and Technology

Department of Artificial Intelligence & Data Science

AY: 2023-24

Class:	BE	Semester:	VII
Course Code:		Course Name:	BDA

Name of Student:	BARI ANKIT VINOD
Roll No. :	61
Assignment No.:	01
Title of Assignment:	
Date of Submission:	
Date of Correction:	

Evaluation

Performance Indicator	Max. Marks	Marks Obtained
Demonstrated knowledge	10	4
Legibility	5	3
Completeness and timely submission	5	2
Total	20	9

Performance Indicator	Exceed Expectations (EE)	Meet Expectations (ME)	Below Expectations (BE)
Demonstrated Knowledge	8-10	5-8	1-4
Legibility	4-5	2-3	1
Completeness and Timely submission	4-5	2-3	1

Checked by

Name of Faculty : Ms. Sweety Patil

Signature :

Date : 12/8/25

Assignment No. - 1

A

1) Amazon is one of the world's largest e-commerse platforms, handling massive volumes of user and product data. To stay competitive, Amazon leverages Big data technologies to personalized, customer experience, optimize supply chains and improve business decisions. Identify and explain how each of the 5Vs of big data volume, velocity, verity, value is applied in Amazon's operations. Support your answer with relevant examples.

→ (i) Volume - Refers to the massive amount of data generated & stored.
- Amazon processes petabytes of data daily from millions of users, product and transactions across the globe.
- Eg. every click, search, purchase and product review generates data.

(ii) Velocity - Refers to the speed at which data is generated, collected and analyzed.
- Amazon requires real-time or near-real-time processing of data to offer fast reccommendations & updates.
- Eg. when user search for a product.

(iii) Varity - Refers to the different types of data and sources of data (structured, semi-structured, unstructured)
- Amamazon handles a wide varity of data types : structured data, semi-structured data, unstructured.
- Eg. Voice queries from Alexa, product images.

(iv) Value - Refers to turning data into meaningfull insights or actions.
- Amazon uses data to improve customer exeperience, optimize logistics and boost sales.
- Eg. Amazon Prime's predictive shopping model helps items.

(v) Veracity - Refers to the trustworthiness and accuracy of data.
- Amazon must ensure the data is clean, realiable and secure to maintain user trust and operations.
- Eg. Uses AI and ML models to detect and remove fakes.

Q. 2> A data analytics company is planing to implement a big d
solution to store, process and analyze large volumes of sho
and unstructured data collected from various sources like mob
apps, social media and IoT sensors. They want a scalable, fa
tolerant system using open-source technologies. Apply the hadoop components the company should use for data storage, processing, resource, management and data access.

→ (i) Data storage : HDFS (Hadoop distributed file system)

Purpose : stores large volumes of data across a distributed cluster.

fault-tolerant (replicates data blocks across a distributed cluster)

scalable (can add more nodes to store more data)

Use cases · Stores raw data from mobile apps, social media feeds.

(ii) Data processing : Apache mapreduce or apache spark

Apache mapReduce - Traditional Hadoop processing engine.

Processes large data sets in a sets in a batch mode using a map and reduce approach.

Analyze user behaviour from app data, sentiments analysis from social media or real-time analytics on sensor data.

(iii) Resource management - YARN (Yet Another Resource Negotiator)

Purpose : Manages cluster resources and schedules jobs.

decouples resources management from data processing.

supports multiple processing engines (mapReduce, spark etc)

Allocates resources efficiently across processing jobs (eg. spark jobs analyzing social media data v.s. IoT sensor streams.)

(iv) Data access : Apache Hive, apache pig, Apache HBase

Apache Hive - SQL-like interface for querying structured data in HDFS.

Apache Pig - data flow scripting language for transforming data and analyzing large datasets.

Apache HBase - NoSQL database build on HDFS for real-time read / write access.