

PRACTICAL :- 1

Task-1:

Q :- How is big data used in cricket?

Strategy Development: Teams use data to formulate strategies. By analysing historical match data, pitch conditions, and opponent strategies, teams can develop tactics that are more likely to succeed. For instance, data might show that certain bowlers perform better under specific conditions or that a particular batting order is more effective.

Player Scouting and Recruitment: Teams use data to scout and recruit players. By analysing performance data from various leagues and tournaments, teams can identify emerging talents and make informed decisions about player acquisitions. This is particularly relevant in T20 cricket, where teams often scout players from different countries to bring fresh talent into their squads.

Performance Analysis: Big data helps analyse players' performance in detail. By collecting and analysing vast amounts of data on players' batting and bowling statistics, match conditions, and opposition strengths, teams can identify strengths and weaknesses. For example, data can reveal a player's performance against different types of bowlers or in various weather conditions.

Player Development: Coaches and analysts use detailed performance data to identify areas for improvement and tailor training programs to address specific weaknesses or enhance strengths.

Q :- Explain 4Vs (Volume, Velocity, Variability, and Variety) characteristics in this company's data.

Volume: Cricket generates large volumes of data from every match, including ball-by-ball statistics, player performance metrics, and historical data. With numerous matches played across various leagues and tournaments worldwide, the volume of data can be substantial.

Velocity: In T20 cricket, data is generated at a high velocity. The fast-paced nature of T20 matches means that data must be collected and analysed in real-time. Fast processing of this data is crucial for making timely decisions, such as adjusting strategies or providing live updates to fans.

Variety: Cricket data comes in various forms, including structured data, unstructured data, and semi-structured data. The variety of data types allows for a comprehensive analysis of matches and player performances.

Variability: Cricket data can be highly variable due to factors such as different playing conditions, varying pitch types, and diverse playing styles. Managing this variability involves adjusting data analysis methods to account for different factors that influence match outcomes and player performances.

Task-2:

Study the Drivers of Big Data and summarize your understanding.

1. Increased use of Cloud Computing:

The increased use of cloud computing has significantly fuelled the growth of big data by offering scalable storage and processing power, enabling organizations to manage and analyse vast datasets efficiently. Cloud platforms provide elastic storage solutions, high-performance computing resources, and advanced data management tools that accommodate the enormous volumes and complexity of big data. Additionally, the global accessibility and collaborative features of cloud computing facilitate real-time data sharing and analysis, while cost-effective scaling and integrated analytics services streamline operations and support advanced insights. This combination of flexibility, cost-efficiency, and powerful capabilities makes cloud computing a key driver in the expansion and utilization of big data.

Task-3:

Study the Data Formats given below and summarize your understanding.

1. CSV (Comma-Separated Values): CSV is a plain text format that uses commas to separate values. It is widely used for storing tabular data in a simple, human-readable format.
PlayerName,Team,Matches,Runs,Average,StrikeRate Virat Kohli,India,96,7547,53.41,89.08
Steve Smith,Australia,95,7318,61.00,56.98

2. JSON (JavaScript Object Notation): JSON is a lightweight data-interchange format that is human readable and easy to parse. It is commonly used for transmitting structured data between a server and a web application. This example is a JSON string: {"PlayerName": "Virat Kohli", "Team": "India", "Matches": 96, "Runs": 7547}

3. Parquet: Parquet is a columnar storage file format that is highly optimized for analytics. It provides efficient compression and encoding techniques, making it suitable for big data processing frameworks like Apache Spark and Hadoop.

4. Avro: Avro is a binary data serialization format that provides schema evolution and rich data structures. It is widely used in Big Data processing systems as it enables efficient data storage and exchange between different programming languages.

5. Key-Value Pair: Key-Value Pair is a simple data format where data is stored as a collection of key value pairs. It is commonly used in NoSQL databases and distributed computing frameworks to store and process data efficiently based on key lookups.

Task-4

Explore the impact of BIG DATA in the following listed field and summarize your study, according to the questions given below:

- 1) Big Data in Healthcare.**
- 2) Big Data in Railway System.**
- 3) Big Data in Agriculture.**

Answer the following questions**1) How data inputs help in Big Data do based Customer value analytics?**

Data inputs are essential for Big Data-based customer value analytics as they provide a comprehensive view of customer behaviors and preferences through transaction history, demographics, browsing patterns, and social media interactions. By analyzing these diverse data sources, businesses can segment customers more effectively, predict future purchasing trends, personalize marketing efforts, and enhance overall customer experience. This data-driven approach enables companies to tailor their strategies to individual needs, optimize engagement, and ultimately increase customer satisfaction and loyalty.

2) How does Big Data help in Railway Reservation System?

Big Data enhances railway reservation systems by optimizing various aspects of ticketing and operations. It allows for dynamic pricing and better demand forecasting by analyzing historical booking patterns and real-time data, which helps in adjusting prices and seat availability accordingly. Real-time data processing improves the accuracy of schedule management and train tracking, reducing delays and improving service reliability. Additionally, Big Data enables personalized recommendations and targeted promotions based on passenger behavior and preferences, leading to a more efficient and customer-centric reservation experience. Overall, it streamlines operations, enhances user experience, and maximizes resource utilization.

3) Why does Big Data offer the potential to transform the medicine and healthcare system?

Big Data has the potential to transform medicine and healthcare by providing deep, actionable insights through the analysis of vast and diverse datasets, including electronic health records, genetic information, medical imaging, and wearable device data. This capability enables personalized medicine by tailoring treatments to individual genetic profiles and health histories, improves predictive analytics for early diagnosis and preventive care, and enhances operational efficiencies by optimizing resource allocation and patient management. Moreover, Big Data accelerates medical research and drug development by integrating and analyzing large volumes of clinical and research data, leading to more effective treatments and a more responsive healthcare system.

4) Describe ways of usages of Big Data analytics in Agriculture.

Big Data analytics transforms agriculture by enabling precision farming, where data from sensors, drones, and satellites optimize resource use and enhance crop management. It supports accurate yield prediction, disease and pest management, and efficient supply chain operations by analyzing historical and real-time data. Additionally, it helps farmers adapt to climate changes by providing insights into weather patterns and their effects on crops. These advancements lead to improved productivity, reduced waste, and more sustainable farming practices.