UNIT 6

Q1) Explain big data analytics challenges in brief.  
**Data Volume: Managing and Storing Massive Amounts of Data  
Challenge:** The most apparent challenge with Big Data is the sheer volume of data being generated. Organizations are now dealing with petabytes or even exabytes of data, making traditional storage solutions inadequate. This vast amount of data requires advanced storage infrastructure, which can be costly and complex to maintain. **Solution:** Adopting scalable cloud storage solutions, such as[***Amazon S3***](https://www.geeksforgeeks.org/introduction-to-aws-simple-storage-service-aws-s3/)***,***[***Google Cloud Storage***](https://www.geeksforgeeks.org/cloud-storage-in-google-cloud-platform-gcp/)***, or***[***Microsoft Azure***](https://www.geeksforgeeks.org/what-is-microsoft-azure/)***,*** can help manage large volumes of data. These platforms offer flexible storage options that can grow with your data needs. Additionally, implementing data compression and deduplication techniques can reduce storage costs and optimize the use of available storage space.

**Data Variety: Handling Diverse Data Types  
Challenge:** Big Data encompasses a wide variety of data types, including structured data (e.g., databases), semi-structured data (e.g., XML, JSON), and unstructured data (e.g., text, images, videos). The diversity of [data types](https://www.geeksforgeeks.org/data-types-in-c/) can make it difficult to integrate, analyze, and extract meaningful insights. **Solution:** To address the challenge of data variety, organizations can employ data integration platforms and tools like Apache Nifi, Talend, or Informatica. These tools help in consolidating disparate data sources into a unified data model. Moreover, adopting schema-on-read approaches, as opposed to traditional schema-on-write, allows for more flexibility in handling diverse data types.

**Data Velocity: Processing Data in Real-Time  
Challenge:** The speed at which data is generated and needs to be processed is another significant challenge. For instance, [IoT](https://www.geeksforgeeks.org/internet-of-things-iot-gateways/)devices, social media platforms, and financial markets produce data streams that require real-time or near-real-time processing. Delays in processing can lead to missed opportunities and inefficiencies. **Solution:** To handle high-velocity data, organizations can implement real-time data processing frameworks such as[Apache Kafka](https://www.geeksforgeeks.org/apache-kafka/), [Apache Flink](https://www.geeksforgeeks.org/what-is-apache-flink/), or Apache Storm. These frameworks are designed to handle high-throughput, low-latency data processing, enabling businesses to react to events as they happen. Additionally, leveraging edge computing can help process data closer to its source, reducing latency and improving real-time decision-making.

**Data Veracity: Ensuring Data Quality and Accuracy  
Challenge:** With Big Data, ensuring the quality, accuracy, and reliability of data—referred to as data veracity—becomes increasingly difficult. Inaccurate or low-quality data can lead to misleading insights and poor decision-making. Data veracity issues can arise from various sources, including data entry errors, inconsistencies, and incomplete data. **Solution:** Implementing robust data governance frameworks is crucial for maintaining data veracity. This includes establishing data quality standards, performing regular data audits, and employing data cleansing techniques. Tools like **Trifacta, Talend Data Quality, and Apache Griffin** can help automate and streamline data quality management processes.

**Data Security and Privacy: Protecting Sensitive Information  
Challenge:** As organizations collect and store more data, they face increasing risks related to data security and privacy. High-profile data breaches and growing concerns over data privacy regulations, such as **GDPR** and **CCPA**, highlight the importance of safeguarding sensitive information. **Solution:** To mitigate security and privacy risks, organizations must adopt comprehensive data protection strategies. This includes implementing encryption, access controls, and regular security audits. Additionally, organizations should stay informed about evolving data privacy regulations and ensure compliance by adopting privacy-by-design principles in their data management processes.

**Data Integration: Combining Data from Multiple Sources  
Challenge:** Integrating data from various sources, especially when dealing with legacy systems, can be a daunting task. Data silos, where data is stored in separate systems without easy access, further complicate the integration process, leading to inefficiencies and incomplete analysis. **Solution:** [Data integration](https://www.geeksforgeeks.org/data-integration-in-data-mining/) platforms like [Apache Camel](https://www.geeksforgeeks.org/what-is-apache-camel/), **MuleSoft**, and **IBM DataStage** can help streamline the process of integrating data from multiple sources. Adopting a microservices architecture can also facilitate easier data integration by breaking down monolithic applications into smaller, more manageable services that can be integrated more easily.

**Data Analytics: Extracting Valuable Insights  
Challenge:** The ultimate goal of Big Data is to derive actionable insights, but the complexity of analyzing large, diverse datasets can be overwhelming. Traditional analytical tools may struggle to scale, and the lack of skilled data scientists can further hinder the ability to extract meaningful insights. **Solution:** Organizations should invest in advanced analytics platforms like[Apache Spark](https://www.geeksforgeeks.org/overview-of-apache-spark/), [Hadoop](https://www.geeksforgeeks.org/hadoop-an-introduction/), or [Google BigQuery](https://www.geeksforgeeks.org/google-cloud-platform-introduction-to-bigquery/), which are designed to handle large-scale data processing and analysis. Additionally, fostering a culture of data literacy and providing training for employees can help bridge the skills gap and empower teams to effectively analyze Big Data. **Data Governance: Establishing Policies and Standards  
Challenge:** As data becomes a critical asset, establishing effective data governance becomes essential. However, many organizations struggle with creating and enforcing policies and standards for data management, leading to issues with data consistency, quality, and compliance. **Solution:** Implementing a formal data governance framework is key to overcoming this challenge. This framework should define roles and responsibilities, establish data stewardship programs, and enforce data management policies. Tools like **Collibra, Alation, and Informatica’s**data governance suite can assist in creating and maintaining a robust data governance strategy.  
------------------------------------------**-------------------------------------------------------------------------------------------------------------------------------------  
q2) big data analytics in research**Big data analytics and data science are becoming the research focal point in industries and academia. Data science aims at researching big data and knowledge extraction from data.   
Applications of big data and data science include information science, uncertainty modeling, uncertain data analysis, machine learning, statistical learning, pattern recognition, data warehousing, and signal processing.   
Effective integration of technologies and analysis will result in predicting the future drift of events. The research issues pertaining to big data analysis are classified into three broad categories namely internet of things (IoT), cloud computing, bio inspired computing, and quantum computing.

**IoT for Big Data Analytics:   
I**nternet has restructured global interrelations, the art of businesses, cultural revolutions and an unbelievable number of personal characteristics. Currently, machines are getting in on the act to control innumerable autonomous gadgets via internet and create Internet of Things. Thus, appliances are becoming the user of the internet, just like humans with the web browsers. IoT is attracting the attention of recent researchers for its most promising opportunities and challenges. It has an imperative economic and societal impact for the future construction of information, network and communication technology.

**Cloud Computing for Big Data Analytics:**The development of virtualization technologies have made supercomputing more accessible and affordable. Computing infrastructures that are hidden in virtualization software make systems to behave like a true computer, but with the flexibility of specification details such as number of processors, disk space, memory, and operating system. The use of these virtual computers is known as cloud computing which has been one of the most robust big data technique. Big Data and cloud computing technologies are developed with the importance of developing a scalable and on demand availability of resources and data. Cloud computing harmonize massive data by on demand access to configurable computing resources through virtualization techniques.  
------------------------------------------**-------------------------------------------------------------------------------------------------------------------------------------  
q3)**  **Explain four big data use cases.**Customer Price Segmentation (Yield Management):This involves segmenting customers based on their price sensitivity to optimize pricing strategies. By understanding how different customer segments react to price changes, businesses can minimize revenue loss and maximize profits.

Predictive Maintenance:  
Big data analytics can predict when equipment or machinery is likely to fail, allowing for proactive maintenance. This prevents unexpected downtime, reduces repair costs, and ensures smooth operations.

Operational Efficiency:  
Analyzing data from various sources can identify areas where operations can be streamlined and made more efficient. This could involve optimizing supply chains, improving resource allocation, or reducing waste.

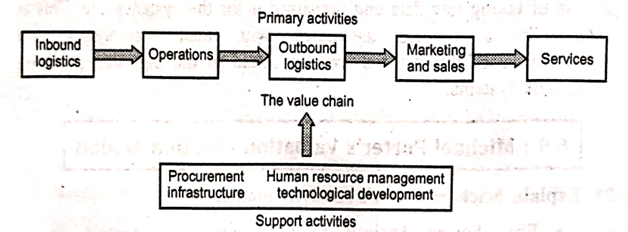
Data-Driven Decision Making:  
Big data helps in making informed business decisions by providing insights into market trends, customer behavior, and operational performance. This enables businesses to adapt quickly to changing conditions and stay ahead of the competition.

Cases of use of Big Data in Factories 4.0:  
**Improving warehouse processes:** By using sensors and portable devices, companies can improve operational efficiency by detecting human errors, performing quality controls, and determining optimal production or assembly routes.  
**Elimination of bottlenecks:** Big Data identifies variables that can affect performance, at no extra cost, guiding manufacturers in identifying the problem.  
**Predictive demand:** More accurate and meaningful predictions are possible through the visualization of activity through internal analysis (customer preferences) and external analysis beyond historical data, allowing the company to modify/optimize its product portfolio.  
**Predictive maintenance:** Data-fed sensors identify possible failures in the operation of machinery before it becomes a breakdown by identifying patterns. The system sends an alert to the equipment so that it can react in time.

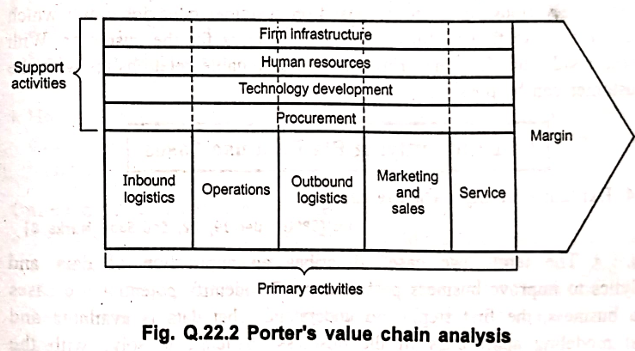
Finding the right use case is essential to the success of a data science project because it enables:

* Understanding the problem from an end-user perspective.
* Finding the right data-driven solution.
* Defining how to measure the project's success, which ensures that the solution adds business value.
* Going beyond data science in a technical sense and viewing the use case from a business perspective.

Big data is characterized by large and diverse datasets that grow rapidly over time and is utilized in machine learning, predictive modeling, and advanced analytics to address business challenges and support informed decision-making. In the context of Industry 4.0, big data plays a crucial role in smart manufacturing, enabling real-time decision-making, enhanced productivity, flexibility, and agility.   
------------------------------------------**-------------------------------------------------------------------------------------------------------------------------------------  
q4) Michael Porter's Analysis Model**the Five Forces Analysis provides input and perspective on an organization's competitiveness based on industry-wide and outside-in data. The five forces are:  
1. **Competitive jealousy:** This primarily focuses on the total number and volume of other organizations competing with the organization in question. It also includes the organizational size and its strategic directions.  
2. **Power of supplier:** This factor relates to the supplier's reputation, the area covered by the supplier, the various products and services they provide, and their capacity to bid on different products and services available.  
3. **Power of buyer:** This force helps analyze the "merchantry" by considering buyer choice and preferences, the volume of buyers, and their switching frequency and tendency.  
4. **Development of Products and Technologies:** This pertains to the quality and price of the various products and services offered. Exposure to marketplace allocation, market trends, and compliance risk, as well as legislative and government actions, are also included under product and technology development.  
5. **New market entrants:** This gives an idea about the barriers to newcomers or new entries into the market. Different geographical and cultural factors affecting "merchantry" are also discovered under this title.

**Porter's Value Chain Analysis**Porter's value chain concept states that there is a chain of events within a company, from the procurement of raw materials to the delivery of goods and post-sales service. Value chain analysis is an analytical framework that assists in identifying business activities that create value and competitive advantage.  
Most organizations engage in numerous activities, which can be classified as either primary or support activities.  
 

According to Porter, the **primary activities** are:  
1. **Inbound Logistics:** Involves relationships with suppliers and includes all activities required to receive, store, and disseminate inputs.  
2. **Operations:** These are all the activities required to transform inputs into outputs (products and services).  
3. **Outbound Logistics:** Includes all the activities required to collect, store, and distribute the output.  
4. **Marketing and Sales:** Activities that inform buyers about products and services, induce buyers to purchase them, and facilitate their purchase.     
5. **Service:** Includes all the activities required to keep the product or service working effectively for the buyer after it is sold and delivered.

The **secondary activities** (also called support activities) are:     
1. **Procurement:** The acquisition of inputs, or resources, for the firm.  
2. **Human Resource Management:** Consists of all activities involved in recruiting, hiring, training, developing, compensating, and (if necessary) dismissing or laying off personnel.  
3. **Technological Development:** Pertains to the equipment, hardware, software, procedures, and technical knowledge used in the firm's transformation of inputs into outputs.     
4. **Infrastructure:** Serves the company's needs and ties its various parts together. It consists of functions or departments such as accounting, legal, finance, planning, public affairs, government relations, quality assurance, and general management.  
  
------------------------------------------**-------------------------------------------------------------------------------------------------------------------------------------  
q5)** **big data value terminology** **Big data:** Big data is an umbrella term for datasets that cannot reasonably be handled by traditional computers or tools due to their volume, velocity, and variety. This term is also typically applied to technologies and strategies used to work with this type of data.

 **Batch processing:** Batch processing is a computing strategy that involves processing data in large sets. This is typically ideal for non-time-sensitive work that operates on very large sets of data. The process is started, and at a later time, the results are returned by the system.

 **Cluster computing:** Clustered computing is the practice of pooling the resources of multiple machines and managing their collective capabilities to complete tasks. Computer clusters require a cluster management layer which handles communication between the individual nodes and coordinates work assignment.

 **Data warehouse:** Data warehouses are large, ordered repositories of data that can be used for analysis and reporting. In contrast to a data lake, a data warehouse is composed of data that has been cleaned, integrated with other sources, and is generally well-ordered. Data warehouses are often spoken about in relation to big data, but typically are components of more conventional systems.

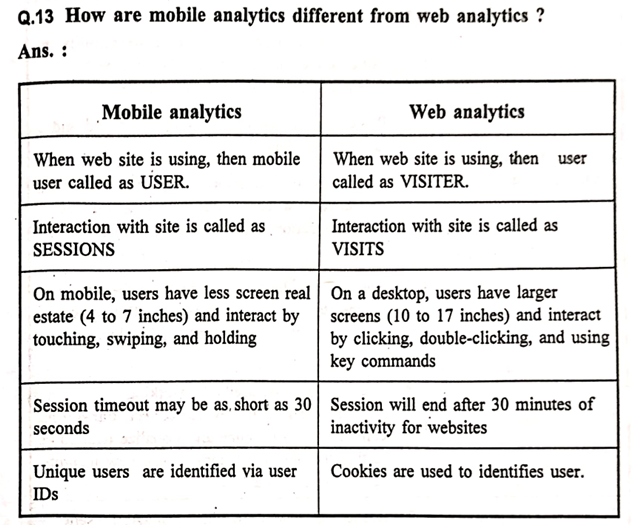
 **ETL:** ETL stands for extract, transform, and load. It refers to the process of taking raw data and preparing it for the system's use. This is traditionally a process associated with data warehouses, but characteristics of this process are also found in the ingestion pipelines of big data systems.  
------------------------------------------**-------------------------------------------------------------------------------------------------------------------------------------  
q6)**  **Explain big data impact on organizations.**Big Data and Data Science, with the help of wide statistics and analytics-helps, are becoming capable of describing and possibly predicting future events. They provide information regarding a wide variety of aspects and require varied skills and approaches.

Big data has significantly impacted organizations in the following ways:

* It has revealed details for defining and putting into numbers terms that are valuable, important, and successful. These details fuel the core of competitive wholesomeness.
* New big data sources and new wide capabilities of deep learning support higher loyalty in answers to organizational questions.
* Most organizations now understand that if they capture all the data streaming into their businesses, they can deploy big data analytics to gain significant leverage in understanding their customers, forecasting business trends, reducing operational costs, and realizing more profits. Regardless of industry, data analysis and visualization are becoming accessible and impactful in critical ways.
* Big data analytics helps to improve quality in industries where inconsistencies are hard to reduce.
* Big Data Analytics is beneficial for organizations in sectors like Travel and hospitality, Healthcare, Government, Retail, and more, which rely on agile and quick decisions to stay competitive.
* The high-performance analytics retrieved from Big Data analytics resources help organizations do things they have never thought of before.
* They can get quick results in a few seconds rather than days, which allows for quick reactions to key business challenges and questions.
* Big Data and Analytics are becoming closely intertwined and work together on unstructured data to get precise answers and uncover new growth opportunities.
* It helps optimize the allotment of sufficient sales resources in view of the best sales opportunities by the sales department. Identifying potential and important business accounts is an equally important task done by the sales team.
* To identify and confirm suppliers who are cost-effective and supply good quality products in a timely manner.
* To measure the device performance and process variance are the main indicators of manufacturing, processing or quality problems.

------------------------------------------**-------------------------------------------------------------------------------------------------------------------------------------  
q7) Explain in brief data analytics life cycle.**The data analytics lifecycle is described through the involvement of different stakeholders:

* **Business user (BI analyst):** Their main responsibility is to define the business process by identifying key performance indicators (KPIs) and metrics to measure these processes. Business users must have clarity on the set of questions and desired answers for making business decisions.
* **Data warehouse manager:** Their primary responsibility is to define, develop, and manage the data platform. New innovative technologies help the data warehouse manager broaden their responsibility using technologies like Hadoop, data federation, and in-memory computing. They are also involved in processing structured and unstructured data using new technologies.
* **Data Scientist:** Their responsibility is to mine unstructured and structured data from inside and outside the organization to discover new business insights. They are continuously searching for new data sources to fulfill the requirements of analytical insights for improving key business processes. The data scientist needs a working environment where they can collect, transform, combine, cross-examine, and visualize data to find hidden relationships and new insights across available data items.
* **Business Intelligent analyst:** Their responsibility is to identify, manage, present, and publish KPIs and metrics against which all business users will calculate and watch business success. Various dashboards and reports are developed by the BI analyst so that business users can run the business and discover new business insights based on real-time data.
* **Business user:** Finally, the analytical process cycles back to business users who utilize the reports and dashboards for their business operations and decision-making.

At the end of the lifecycle, the effectiveness of various decisions made by the business users is entirely dependent on the effective work done by the data scientist, BI analyst, and data warehouse manager.  
------------------------------------------**-------------------------------------------------------------------------------------------------------------------------------------  
q8)**   
****  
------------------------------------------**-------------------------------------------------------------------------------------------------------------------------------------  
q9) Mobile Analytics**Mobile analytics is the practice of measuring and analyzing data related to user behavior and application performance on mobile platforms, including mobile sites and mobile applications. It is the practice of collecting user behavior data and determining user intent from these metrics to drive attention, engagement, and conversion. Mobile analytics is similar to web analytics in identifying the unique customer and recording their usages. Analytics usually comes in the form of software that integrates into a company's existing websites and apps to capture, store, and analyze the data. Measuring critical KPIs (Key Performance Indicators) is always important for businesses.

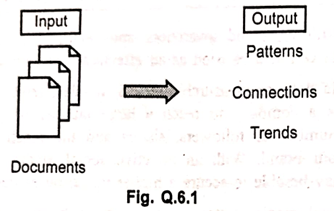
**How it Helps Organizations:**Mobile analytics helps organizations by providing insights into user behavior, which can be used to:

* Improve cross-channel marketing initiatives.
* Optimize the mobile experience for customers.
* Grow mobile user engagement and retention.

**Types of Mobile Analytics:**1. Advertising/Marketing Analytics:   
In today's marketplace, even if we develop an incredible app, the probability of it organically standing out among a million other apps is very low. The success of an app often hinges on whether marketing campaigns are able to attract the right types of users those that install, remain engaged, and contribute to -the financial components of the app.   
Partnering with one or more ad networks is one of the most common ways to market an app. In the olden days, this practice consisted of establishing a budget and permitting an ad network to arrange for our ad to be displayed on a variety of publisher websites and apps. If the campaign was successful, we would see an increase in installs, engagement, and financial metrics.   
Examples of common marketing analytics data that can be collected includes: Installs, Opens, Clicks, Purchases, Registrations, Content viewed, Level achieved, Shares and Custom events.

2. In-App Analytics   
Regardless of whether an app delivers content, sells products, or offers a gaming experience, in order to be successful, the app must satisfy the expectations of its users.   
Every app has one or more goals or objectives. In theory, apps are designed to enable users to achieve these objectives in the simplest manner possible.   
We may follow intuitive hunches or make marginally-educated guesses regarding user demographics and in-app behaviors, but with no user or in-app behavior data, knowing where to make improvements may as well be determined on the roulette wheel.   
In-app analytics is essentially "in-session" analytics what users are actually doing inside the app and how they are interacting with the app. This is where conversion funnel, pathway, and feature optimization is the primary focus. Although marketers sometimes get looped into this data, it is primarily used by product managers.

**Working of Mobile Analytics:**Most of the analytics tools need a library (an SDK) to be embedded into the mobile app's project code and at minimum an initialization code in order to track the users and screens.   
SDKs differ by platform so a different SDK is required for each platform such as iOS, Android, Windows Phone etc. On top of that, additional code is required for custom event tracking.   
With the help of this code, analytics tools track and count each user, app launch, tap, event, app crash or any additional information that the user has, such as device, operating system, version, IP address (and probable location).   
Unlike web analytics, mobile analytics tools don't depend on cookies to identify unique users since mobile analytics SDKs can generate a persistent and unique identifier for each device.   
The tracking technology varies between websites, which use either JavaScript or cookies and apps, which use a software development kit (SDK).   
Each time a website or app visitor takes an action, the application fires off data which is recorded in the mobile analytics platform.

**Difference from Social Media Analytics:**Mobile analytics involves measuring and analyzing data generated by mobile platforms and properties, such as mobile sites and mobile applications.   
With mobile analytics data, you can improve your cross-channel marketing initiatives, optimize the mobile experience for your customers, and grow mobile user engagement and retention.   
Social media analytics encompasses the collection, measurement, analysis, visualization and interpretation of digital data illustrating user behavior.   
Mobile analytics can provide more significant data and understanding than traditional web analytics.   
Mobile analytics do not only track the use of mobile apps, but also mobile web traffic. This combination of tracking mobile browsing and also offering a deeper understanding into user engagement with an app provide valuable insights how users react, interact and engage with different mobile features, pages and advertising.   
This approach of tracking provides feedback to developers, designers, advertisers and marketers to help them understand why users are or are not registering, buying or returning.  
Real time analytics are also key to understanding and improving user experience. These analytics focus on understanding user behaviour, instead of just providing a narrow set of metric data such as the amount of downloads.   
Web analytics offer great services, tracking the number of visits, recording how long they remained on a site and also providing information how they arrived at the site.   
However this information is starting to be diluted by the fact that even if some browser tabs are open, it doesn't mean that they are being engaged with.   
Or if a movie or TV show is being watched online, a web analytics can't know if a second screen is being engaged with at the same time. This doesn't make the web analytics irrelevant, but it is a thought that needs to be added to the equation.   
Social media analytics track IP addresses and the user agent, however with users working from different locations such as home, work, cafes, airports, etc., switching browsers for different reasons or clearing cookies suddenly a user can become anonymous.   
Mobile analytics face similar issues, a user might own several mobile devices, including tablets, alongside a PC and another PC at their workplace.   
Although there are a few advantages to keeping mobile users as they can connect through social authentication across several traditional web devices with their mobile devices. Mobile users also can clear their cookies, but it is not as common to reset their mobile identifiers.   
------------------------------------------**-------------------------------------------------------------------------------------------------------------------------------------  
q10) What is text mining? Draw and explain text mining architecture and its use. OR Explain text mining with example  
What is Text Mining?**  
Text mining can be broadly defined as a knowledge-intensive process in which a user interacts with a document collection over time by using a suite of analysis tools. Text mining seeks to extract useful information from data sources through the identification and exploration of interesting patterns.    
The simple input-output model for text mining (Figure Q.6.1) shows documents as the input, which undergo processing to produce output in the form of patterns, connections, and trends. This illustrates the core use of text mining: to transform unstructured text data into structured insights and discover valuable knowledge within document collections. The overall use of text mining, as implied by its definition and architecture, is to enable users to interact with large collections of text data and extract useful, pattern-based information for various applications.  
 **Text Mining Architecture and its Use:**  
**A diagram of a process

AI-generated content may be incorrect.  
1. Preprocessing tasks:** This includes all the routines, processes, and methods required to prepare data for the text mining system's core knowledge discovery operations. Preprocessing generally involves converting information from each original data source into a canonical format before applying feature extraction methods. These methods are used to create a new collection of documents fully represented by concepts.  
2. **Core Mining operations:** These are considered the heart of a text mining system and include activities such as pat-tern discovery, trend analysis, and incremental knowledge discovery algorithms.  
3. **Presentation Layer Components:** This layer includes the Graphical User Interface (GUI) and functionalities for pattern Browse, as well as access to the query language. Visualization tools and user-facing query editors and optimizers also fall under this architectural category. These components are crucial for users to interact with the text mining system and interpret the results.     
4. **Refinement techniques:** These methods are used to filter redundant information and cluster closely related data. Refinement techniques in a given text mining system aim to present a full, comprehensive suite of operations, including suppression, ordering, pruning, and generalization.  
------------------------------------------**-------------------------------------------------------------------------------------------------------------------------------------  
q11)** **social media analytics helps in value creation in several ways**Social media contains a wealth of data that is invaluable to marketers. Social media data, made up of collected information from social networks that shows how users engage with, view, or share content, can be tracked and analyzed to inform a successful marketing strategy.

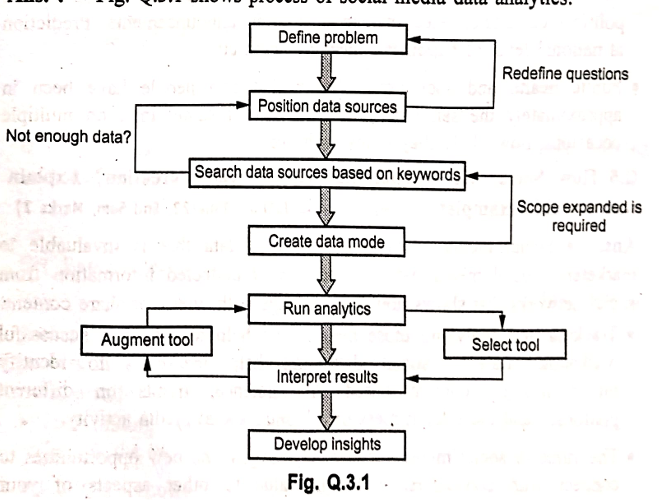
Here's how social media analytics helps in value creation:

* Social data helps marketers to identify high-performing content based on audience engagement, trends on different platforms, and the overall effectiveness of a brand's social media activity.
* The value of social media expands beyond just new opportunities to connect with consumers. It brings value to other aspects of your company, including learning more about your company's target audience, building brand awareness, and retaining existing customers. Social media can also be used as an effective customer service tool.
* Social media delivers measurable results in sales, leads, and branding. It enables a company to reach a large number of people at a low cost. The number of followers, shares, and the reach can impact the value of your brand. With an effective social media plan in place, a company can potentially achieve a higher valuation at the time of sale.
* Another social media strategy is to create a hashtag personal to your brand or one that encourages followers to create their own content as a way to increase engagement and ideally sales.

**Example:**  
For example, Oreo's #PlayWithOreo Instagram campaign is mentioned. This campaign resulted in followers taking their own pictures featuring the ways they play with Oreo cookies. This gave followers who love Oreo a chance to interact with the brand, and followers purchased the product in order to participate.

Finding a way to adapt your company's product or services into a visually appealing, fun, and entertaining social media campaign is presented as a good way to gain extra exposure and more deeply connect with your followers, as well as increase sales. Showing a potential acquirer that your company can evolve and use new forms of social media to consistently expand sales can make your company more appealing and lead to a higher valuation.     
------------------------------------------**-------------------------------------------------------------------------------------------------------------------------------------  
q12)** **What do you mean social media analytics? OR Describe process of social media data analytics.**Social Media Analytics deals with development and evaluation of tools and frameworks to collect, monitor, analyze, summarize, and visualize social media data   
Buyer's perspective about various brands and businesses can be statistically analyzed to extract various insights required for decision making from a very large amount of unstructured and semi structured social media data.   
In social media, the two sources of information are the content (images, audios, customer feedbacks, product reviews, videos, bookmarks, sentiments etc.) generated by users and the relationships between the entities of network (people, organizations, products etc.).   
The social media analytics can be categorized into two parts: Content-based analytics and Structure-based analytics.   
In content-based analytics, analytics is performed on the content posted by the users on the social media platforms. Such content is of high volume, unstructured, noisy and dynamic nature.   
To extract insights from such data, the text, audio and video analytics techniques can be applied. The data processing challenges are addressed by the big data technologies.   
In structure-based analytics, the focus is on the structural attributes of the social network. Insights are extracted from the relationships of the entities

**Process of Social Media Data Analytics**

  
Social Media Analytics as a part of social analytics is the process of gathering data from stakeholder conversations on digital media and processing into structured insights leading to more information-driven business decisions and increased customer centrality for brands and businesses.   
Data analysis is the set of activities that assist in transforming raw data into insight, which in turn leads to a new base of knowledge and business value.   
In other words, data analysis is the phase that takes filtered data as input and transforms that into information of value to the analysts.   
Many different types of analysis can be performed with social media data. The data analysis step begins once we know what problem we want to solve and know that we have sufficient data that is enough to generate a meaningful result.

Applications of Social Media  
**Retail companies:** To harness their brand awareness, improve services, enhance advertising/marketing strategies, and identify influencers.   
**Finance:** To determine market sentiment and utilize news data for trading purposes.   
**Government and public officials:** For monitoring public perception on political candidates, election campaigns, and announcements. It can also be used for predicting national levels of happiness, unemployment, etc.   
**Public health and sociology:** Used in research questions, such as determining the likelihood of two people knowing each other if they have been in approximately the same geographic locale, at the same time, on multiple occasions.  
------------------------------------------**-------------------------------------------------------------------------------------------------------------------------------------  
Q13) Social Network Analysis (SNA)**Social Network Analysis (SNA) is the mapping and measuring of relationships and flows between people, groups, organizations, computers, URLs, and other connected information or knowledge entities. The term "social network" was introduced by Barnes in 1954.     
SNA is the study of social relations among a set of actors. The methods of data collection in network analysis are aimed at collecting relational data in a reliable manner. Data collection is typically carried out using standard questionnaires and observation techniques that aim to ensure the correctness and completeness of network data.     
Social network analysis is based on the assumption of the importance of relationships among interacting units. The social network perspective encompasses theories, models, and applications that are expressed in terms of relational concepts or processes.     
In the network, the nodes are the people and groups, while the links show relationships or flows between the nodes. SNA provides both a visual and a mathematical analysis of human relationships.   The advantage of social network analysis is that, unlike many other methods, it focuses on interaction. Network analysis allows us to examine how the configuration of the network influences how individuals and groups, organizations, or system functions.