Total No. of Questions: 6] SEAT No:

P1578 Total No of Pages: 2

**BE/ Insem/Apr -255A**

**B.E.(Computer Engineering )**

**BIG DATA ANALYTICS**

**(2015 Pattern) (Semester - II) (Open Elective)**

**----------------------------------------------------------------------------------------------------------------**

**BDDA-MODEL ANSWER**

**Q1a). Define Big Data? List any three characteristics of Big Data. [05M]**

Big Data refers to large sets of complex data, both

structured and unstructured which traditional processing

techniques and/or algorithms are unable to operate on. It aims

to reveal hidden patterns and has led to an evolution from a

model-driven science paradigm into a data-driven science

paradigm.

Big Data refers to large sets of complex data, both

structured and unstructured which traditional processing

techniques and/or algorithms are unable to operate on. It aims

to reveal hidden patterns and has led to an evolution from a

model-driven science paradigm into a data-driven science

paradigm.

Big Data refers to large sets of complex data, both

structured and unstructured which traditional processing

techniques and/or algorithms are unable to operate on. It aims

to reveal hidden patterns and has led to an evolution from a

model-driven science paradigm into a data-driven science

paradigm.

* Big Data refers to large sets of complex data, both structured and unstructured which traditional processing techniques and/or algorithms are unable to operate on. It aims to reveal hidden patterns and has led to an evolution from a model-driven science paradigm into a data-driven science paradigm.
* Big Data is also data but with a huge size. Big Data is a term used to describe a collection of data that is huge in size and yet growing exponentially with time. In short such data is so large and complex that none of the traditional data management tools are able to store it or process it efficiently.
* Big data means huge amount of data, it is a collection of large datasets that cannot be processed using traditional computing techniques. Big Data is complex and difficult to store, Maintain or access in regular file system, big data becomes a complete subject, which involves different techniques, tools, and frameworks.

**Following are the Characteristics of Big Data.**

1. Volume

2. Velocity

3. Variety

4. Variability

**1. Volume:**

* The name Big Data itself is related to a size which is enormous. Size of data plays a very crucial role in determining value out of data. Also, whether a particular data can actually be considered as a Big Data or not, is dependent upon the volume of data. Hence, 'Volume' is one characteristic which needs to be considered while dealing with Big Data.

**2. Variety:**

* Variety refers to heterogeneous sources and the nature of data, both structured and unstructured. During earlier days, spread sheets and databases were the only sources of data considered by most of the applications. Nowadays, data in the form of emails, photos, videos, monitoring devices, PDFs, audio, etc. are also being considered in the analysis applications, this variety of unstructured data poses certain issues for storage, mining and analyzing data

**3. Velocity:**

* The term 'velocity' refers to the speed of generation of data. How fast the data is generated and processed to meet the demands, determines real potential in the data. Big Data Velocity deals with the speed at which data flows in from sources like business processes, application logs, networks, and social media sites, sensors,[Mobile](https://www.guru99.com/mobile-testing.html)devices, etc. The flow of data is massive and continuous.

**4. Variability:**

* This refers to the inconsistency which can be shown by the data at times, thus hampering the process of being able to handle and manage the data effectively.

**Q1 b. Differentiate between Data Warehouse and DBMS. [05M]**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr.No** | **Characteristics** | **Data Warehouse** | **DBMS** |
| 01 | Data Types | Relational data from transactional systems, operational databases and application | Structured, numerical, data, text, and dates organized in relational model |
| 02 | Purpose | Data stored for business intelligence batch reporting and data visualization | Transaction Processing |
| 03 | Data Capture | Data captured from multiple relational sources | Data captured from a single sources, such as transactional system |
| 04 | Data Normalization | Deformalized schemas; schemas on write | Uses normalized static schemas |
| 05 | Benefits | Historical data from many sources stored in one place; data is classified with user in mind for accessibility | Provide consistent data for critical business application |
| 06 | Data Quality | Curated data that is centralized and ready for use in BI and analytics | Data is organized and consistent |

**OR**

**Q2. a) Write Short Note on: Business Intelligence [05M]**

* Business intelligence may be defined as a set of mathematical models and analysis methodologies that exploit the available data to generate information and knowledge useful for complex decision-making processes.
* The advent of low-cost data storage technologies and the wide availability of Internet connections have made it easier for individuals and organizations to access large amounts of data. Such data are often heterogeneous in origin, content and representation, as they include commercial, financial and administrative transactions, web navigation paths, emails, texts and hypertexts, and the results of clinical tests, to name just a few examples.
* Their accessibility opens up promising scenarios and opportunities, and raises an enticing question: is it possible to convert such data into information and knowledge that can then be used by decision makers to aid and improve the governance of enterprises and of public administration

**Effective and Timely decision:**

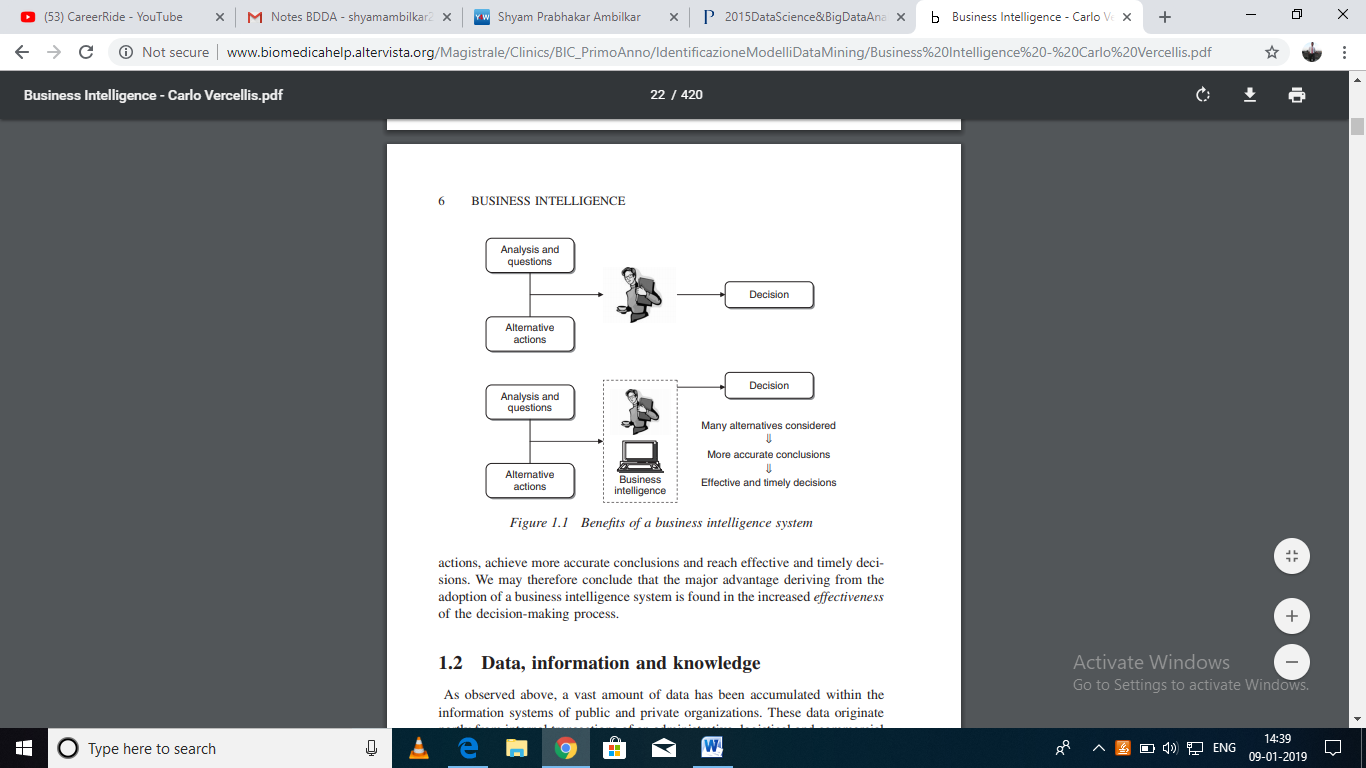
* In complex organizations, public or private, decisions are made on a continual basis. Such decisions may be more or less critical, have long- or short-term effects and involve people and roles at various hierarchical levels.
* The ability of these knowledge workers to make decisions, both as individuals and as a community, is one of the primary factors that influence the performance and competitive strength of a given organization.

**Effective decisions:**

1. The application of rigorous analytical methods allows decision makers to rely on information and knowledge which are more dependable. As a result, they are able to make better decisions and devise action plans that allow their objectives to be reached in a more effective way.
2. Indeed, turning to formal analytical methods forces decision makers to explicitly describe both the criteria for evaluating alternative choices and the mechanisms regulating the problem under investigation.
3. Furthermore, the ensuing in-depth examination and thought lead to a deeper awareness and comprehension of the underlying logic of the decision-making process.

**Timely decisions:**

1. Enterprises operate in economic environments characterized by growing levels of competition and high dynamism.
2. As a consequence, the ability to rapidly react to the actions of competitors and to new market conditions is a critical factor in the success or even the survival of a company

* 

**Figure: Benefits of business intelligence system**

**Q2. b). Explain the Terms:**

i) Semi Structured Data

ii) Quasi-structured data

**i) Semi Structured Data:**

* Semi-structured data is information that doesn’t reside in a relational database but that does have some organizational properties that make it easier to analyze. With some process you can store them in relation database (it could be very hard for some kind of semi structured data), but the semi structure exists to ease space, clarity or compute…
* Examples of semi-structured: CSV but XML and JSON documents are semi structured documents, NoSQL databases are considered as semi structured.
* Semi-structured data falls in the middle between structured and unstructured data. It contains certain aspects that are structured, and others that are not.
* **Advantages of Semi Structured Data:**

1. Programmers persisting objects from their application to a database do not need to worry about [object-relational impedance mismatch](https://en.wikipedia.org/wiki/Object-relational_impedance_mismatch), but can often serialize objects via a light-weight library.
2. Support for nested or hierarchical data often simplifies data models representing complex relationships between entities.
3. Support for lists of objects simplifies data models by avoiding messy translations of lists into a relational data model.

**ii) Quasi Structured Data:**

**Q3. A) Explain the example of Big Data innovation on the IT infrastructure. [05M]**

* Innovation is an iterative process aimed at the creation of new products , processes , knowledge , or services by the use of new or even existing knowledge
* Data-driven innovation entails exploitation of any kind of data in the innovation process to create value
* The emerging trend of big data-driven innovation is leading to the development of data-driven goods and services and can enable data-driven planning, data-driven marketing, and data-driven operations across all industrial sectors and domains.
* From the economic perspective, data as a non-rivalrous good or commons such as oil serves an *infrastructural resource* (from a functional perspective) that could be exploited simultaneously by many users or actors for different competing or complementary ends.
* The demand for data in this sense according to the OECD is driven primarily by downstream productive activities that require data as an input and, in fact, a non-trivial capital.
* The same authors assert that data resources may be used as input into a wide variety of goods, including private, public, and social goods.
* Big data-driven innovations are implicitly associated with a value chain model or more precisely a “virtual value chain” specifying how the data of interest will be gathered, organized, selected, transformed into products or services, and distributed.

**Q4. B) What are the market drivers related to Big Data? [05M]**

Business entails market, sales and financial side of things, there are different drivers are involve in Business. Following are the drivers involve in the Big Data,

1. Data Driven Initiatives:

1. Data driven Innovation

2. Data driven decision making

3. Data driven discovery

1. Data Science is competitive advantage
2. Sustained Processes
3. Cost advantages of commodity hardware and Open Source Software
4. Quick turnaround and less bench time
5. Automation to backfill redundant/mundane task
6. Optimize workforce to leverage high talent cost

**1. Data Driven initiatives:**

There are primarily categorized into three types:

**i. Data Driven Innovations:**

* Ability to drive innovation through those uber targeted data indicators.

**ii. Data driven decision making:**

* Data driven decision-making is the inherent ability of analytics to sieve through globs of data and identify the best path forward.
* Whether in terms of finding the best route to validating the current route and estimating the success/failure in current strategy.
* It takes decision making away from gut and focus on data backed reasoning for higher chances of success.

**iii. Data driven discovery:**

* Your data know a whole lot about you than you image. Having a discovery mechanism will help you understand hidden insights that were not visible through traditional means.

**2. Data Science as a competitive advantage:**

* Big data as a capability to add to their competitive advantage. With a proper data driven framework, businesses could build sustainable capabilities and further leverage these capabilities as a competitive edge.
* If businesses were able to master big data driven capabilities, businesses could use these capabilities to establish secondary source of revenues by selling it to other businesses.

**3. Sustained Processed:**

* Data driven approach creates sustainable processes, which gives a huge endorsement to big data analytics strategy as a go for enterprise adoption.
* Randomness kills businesses and adds scary risks, while data driven strategy reduces the risk by bringing statistical models, which are measurable.

**4. Cost advantages of commodity hardware & open source software:**

* Cost advantage is music to CXO’s ears.
* How about the savings your IT will enjoy from moving things to commodity hardware and leverage more open source platforms for cost effective ways to achieve enterprise level computations and beyond.
* No more overpaying of premium hardware when similar or better analytical processing could be done using commodity and open source systems.

**5. Quick turnaround and less time bench:**

* Complex processes and communication gives you hard time connecting with someone who could get the task done.
* Things take forever long and cost fortunes with substandard quality.
* A good big data and analytics strategy could reduce the proof of concept time smoothly and substantially.
* It reduces the burden on IT and gets more high quality, fast and cost effective solutions baked.
* So, you will waste less time waiting for analysis / insights and more time digging through complex data, and use it for better insights and analyses which was never heard of before.

**6. Automation to backfill redundant/mundane tasks:**

* How about doing something to the 80% of time that is wasted in data cleaning and pre-processing.
* There is great deal of automation that could be take part and sky rocket enterprise efficiency. Less manual time spent on data prep and more time is spent on doing analysis that would have substantial ROI compared to mundane data preps and monotonous tasks.

**7. Optimize workforce to leverage high talent cost:**

* Big data & analytics strategy ensures current workforce is leveraged to its core in handling enterprise big data and also ensures right number of data scientists is involved with clearer sight to their contribution and their ROI.

**OR**

**Q4 A) Explain typical Big Data analytical architecture with suitable diagram. [05M]**

* There is need of workspace to Data Science projects which are basically built for experimenting with data, with flexible as well as agile data architectures.
* Numbers of organizations will possess data warehouses which give excellent support for reporting in traditional way and signified data analysis activities but problems arise when there is need of more robust analysis.
* For the purpose of data sources to be loaded into the data warehouse there is need that the data should be well understood, in structured format, and normalized with the suitable data type definitions.
* Even if such type of centralization leads to security, backup, and failover of highly critical data.
* It also indicates that the data should carry out effective pre-processing as well as checkpoints prior to entering in this of controlled environment, which does not allow its use in data exploration and iterative analytics.
* As a result of such level of control on the enterprise data warehouse (EDW),it is possible that some more local systems emerge in the role of departmental warehoused and local data marts which are created by business users for the purpose of accommodating their requirements of flexible analysis.

**Dashboards**

Departmental

**Data Science users**



**Reports**

**Reports**

**Alerts**

**Figure: Typical analytic Architecture**

* There may not be similar constraints regarding security and structure on their local data marts as of the main EDW and let users to implement some level of more in-depth analysis.
* Still such off system exists in isolation, usually are unsynchronized or not integrated with other types of data stores and also may not be backed up.
* For BI and reporting purposed data is acceded by more applications in the environment of enterprise in the data warehouse.
* These are considered as high priority operational processed which retrieve critical data feeds from the data warehoused and repositories.
* When this workflow ends analysts obtain data which is basically provisioned for their downstream analytics.
* It is not allowed for users to run custom or intensive analytics on production databases, analysts have to generate data extracts from the enterprise data warehouse (EDW) for the purpose of analysing data offline in R different local analytical tools.
* Number of such tools is limited to in-memory analytics on desktops analysing samples of data instead of whole population of a dataset.
* Since the base of these analyses is data extracts, they are located in a separate location and the outcomes of the analysis and any insights on the quality of the data or anomalies-rarely are sent to the main data repository
* The moving speed of data is slow in EDW and also the process of chanting data schema takes longer became the process of accumulation of new data sources take more time in the EDW due to the through validation and structuring process.

**Q5 B) Compare Business Intelligence and Data Science [05M]**

* Nowadays to handle the various types of business problems, organization has to be more analytical and data driven
* Business drivers for advanced analytics

|  |  |  |
| --- | --- | --- |
| Sr.No | Business Driver | Example |
| 01 | Optimize business operations | Sales, Pricing, Profitability, efficiency |
| 02 | Identify Business risk | Customer churn, fraud, default |
| 03 | Predict new business opportunities | Upsell, cross-sell, best new customer prospects |
| 04 | Comply with laws or regulatory requirements | Anti-Money Laundering Fair Lending, Basel II-III, Sarbanes Oxley(SOX) |

* We can observe that there are four generalized categories of common business problems necessary for them to leverage advanced analytics for the purpose of creating competitive advantage.
* Instead of just working on standard reporting on their areas, it is possible for organizations to apply some advanced analytical techniques for the purpose of optimizing processes and get more value from the usual tasks.
* The first three examples are not concerned with new problems.
* Organizations have been trying to provide good service increase sales for many years.
* What exactly new advantage is the chance to combine advanced analytical techniques with big data so as to generate more impactful analyse for the various traditional problems.
* The last example is concerned with various emerging regulatory requirements
* There are number of compliance as well as regulatory laws present for decades but new more requirements are added year by year which leads to increase in complexity and data requirements for organizations.
* Laws which represent the AML (Anti-Money Laundering) and fraud preventions need s some more advanced analytical techniques for the purpose of comply with and manage properly.

**BI Vs. Data Science:**

* The four business drivers which we have discussed in previous section need a variety of analytical techniques to address them properly.
* There are number of ways which helps to compare these groups of analytical techniques.
* One way for the evaluation of the type of analysis being carried out is to observe the time horizon and the type of analytical approaches being used.
* BI usually provides reports dashboards and queries on business questions for the current period or in the past
* BI systems helps to simplify to answer questions regarding quarter-to date revenue, progress towards quarterly targets and known quantity of given product was sold in a prior quarter or year
* These question considered as closed-ended and explain current or past behaviour normally by the process of aggregating historical data and grouping it in some way
* BI offers hindsight and little insights and usually answers questions regarding “when” and “where” events occurred.
* When compared with BI it is found that Data Science like to use disaggregated data with a more forward looking exploratory techniques concentrating on analysing the present and enabling informed decisions about the future.
* Instead of aggregating historical data to search for quantity of product sold in the previous quarter it is possible for a team to employ Data Science techniques like time series analysis.
* Such techniques help to guess future product sales and revenue more precisely as compared to extending a simple trend line.
* Also Data Science considered as ore exploratory in nature and may like to refer scenario optimization for the purpose of dealing with more open-ended questions.
* This approach helps to get insights into current activity and foresight into events while usually concentrating on questions regarding “how” and “why” event occur.
* Where BI problems need highly structured data which has been organized in rows and columns for accurate reporting. Data Science projects mostly refer various kinds of data sources including large or unconventional datasets
* Based on the future goals of organizations it may prefer to board on a PI project if there is reporting dashboards creation o simple visualizations or it may prefer to board on Data Science projects if it required to do a more sophisticated analysis with datasets which are in the form of disaggregated or distinct.

|  |  |
| --- | --- |
| Predictive Analytics and Data Mining (Data Science) | |
| Typical techniques and data types | Optimization predictive modelling, forecasting, statistical analysis  Structured/Unstructured data many types of sources very large datasets |
| Common Questions  Data Science | What if?  What’s the optimal scenario for our business?  What will happen next? What if those trends continue? Why is this happening? |

|  |  |
| --- | --- |
| Business Intelligence | |
| Typical techniques and data types  Business Intelligence | Standard and ad hoc reporting dashboards alerts queries details on demand  Structured data traditional sources manageable datasets |
| Common Questions | what happened that queries?  How many units sold?  Where is the problem? In which situation? |

Past Time Future

**Figure: Comparing BI with Data Science**

**Q5. A) Explain Data Exploration and variable Selection in the Model Planning Phase. [5M]**

* In this Phase the data science team identifies candidate models to apply to the data for clustering, classifying, or finding relationships in the data depending on the goal of the project.
* It is during this phase that the team refers to the hypotheses developed in Phase 1, when they first became acquainted with the data and understanding the business problems or domain area.
* These hypotheses help the team frame the analytics to execute in Phase 4 and select the right methods to achieve its objectives.
* Some of the activities to consider in this phase include the following:
  + Assess the structure of the datasets. The structure of the data sets is one factor that dictates the tools and analytical techniques for the next phase.
  + Depending on whether the team plans to analyze textual data or transactional data, for example, different tools and approaches are required.
  + Ensure that the analytical techniques enable the team to meet the business objectives and accept or reject the working hypotheses.
  + Determine if the situation warrants a single model or a series of techniques as part of a larger analytic workflow. A few example models include association rules and logistic regression.
  + Other tools, such as Alpine Miner, enable users to set up a series of steps and analyses and can serve as a front-end user interface (UI) for manipulating Big Data sources in PostgreSQL.

**Data Exploration and Variable Selection:**

* Although some data exploration takes place in the data preparation phase, those activities focus mainly on data hygiene and on assessing the quality of the data itself.
* The objective of the data exploration is to understand the relationships among the variables to inform selection of the variables and methods and to understand the problem domain.
* As with earlier phases of the Data Analytics Lifecycle, it is important to spend time and focus attention on this preparatory work to make the subsequent phases of model selection and execution easier and more efficient.
* A common way to conduct this step involves using tools to perform data visualizations.

**Q5. B) Explain “Data Discovery” Phase of Data analytic lifecycle in detail.**

The first phase of the Data Analytics Lifecycle involves discovery .In this phase, the data science team must learn and investigate the problem, develop context and understanding, and learn about the data sources needed and available for the project. In addition, the team formulates initial hypotheses that can later be tested with data.

**1. Learning the Business domain:**

* Understanding the domain area of the problem is essential.
* In many cases, data scientists will have deep computational and quantitative knowledge that can be broadly applied across many disciplines.
* An example of this role would be someone with an advanced degree in applied mathematics or statistics. These data scientists have deep knowledge of the methods, techniques, and ways for applying heuristics to a variety of business and conceptual problems. Others in this area may have deep knowledge of a domain area, coupled with quantitative expertise.

**2. Resources:**

* As part of the discovery phase, the team needs to assess the resources available to support the project. In this context, resources include technology, tools, systems, data, and people.
* During this scoping, consider the available tools and technology the team will be using and the types of systems needed for later phases to operationalize the models.
* The skills and computing resources, it is advisable to take inventory of the types of data available to the team for the project. Consider if the data available is sufficient to support the project's goals.

**3. Framing the Problem:**

* Framing the problem well is critical to the success of the project. Framing is the process of stating the analytics problem to be solved. At this point, it is a best practice to write down the problem statement and share it with the key stakeholders.
* Each team member may hear slightly different things related to the needs and the problem and have somewhat different ideas of possible solutions.

**4. Identifying key stack holders:**

* Another important step is to identify the key stakeholders and their interests in the project.
* The team can identify the success criteria, key risks, and stakeholders, which should include anyone who will benefit from the project or will be significantly impacted by the project.

**5. Developing initial Hypothesis:**

* Developing a set of IHs is a key facet of the discovery phase. This step involves forming ideas that the team can test with data.
* Generally, it is best to come up with a few primary hypotheses to test and then be creative about developing several more.
* These IHs form the basis of the analytical tests the team will use in later phases and serve as the foundation for the findings
* The team should perform five main activities during this step of the discovery phase:

**i .Identify data sources:**

* Make a list of candidate data sources the team may need to test the initial hypotheses outlined in this phase.
* Make an inventory of the datasets currently available and those that can be purchased or otherwise acquired for the tests the team wants to perform.

**ii. Capture aggregate data sources:**

* This is for previewing the data and providing high-level understanding. It enables the team to gain a quick overview of the data and perform further exploration on specific areas.
* It also points the team to possible areas of interest within the data.

**iii. Review the raw data:**

* Obtain preliminary data from initial data feeds. Begin understanding the interdependencies among the data attributes, and become familiar with the content of the data, its quality, and its limitations.

**iv. Evaluate the data structures and tools needed:**

* The data type and structure dictate which tools the team can use to analyze the data. This evaluation gets the team thinking about which technologies may be good candidates for the project and how to start getting access to these tools.

**v. Scope the sort of data infrastructure needed for this type of problem:**

* In addition to the tools needed, the data influences the kind of infrastructure that's required, such as disk storage and network capacity.

**OR**

**Q6. A) List any three commercial tools and any two free and open source tools for Model Building Phase. [05M]**

There are many tools available to assist in this phase, focused primarily on statistical analysis or data mining software. Common tools in this space include, but are not limited to, the following:

* **Commercial Tools:**
* **SAS Enterprise Miner:** allows users to run predictive and descriptive models based on large volumes of data from across the enterprise. It interoperates with other large data stores, has many partnerships, and is built for enterprise-level computing and analytics.
* **SPSS Modeler:** (provided by IBM and now called IBM SPSS Modeler) offers methods to explore and analyze data through a GUI.
* **Matlab:** provides a high-level language for performing a variety of data analytics, algorithms, and data exploration.
* **Alpine Miner:** provides a GUI front end for users to develop analytic workflows and interact with Big Data tools and platforms on the back end.
* **STATISTICA** and **Mathematica** are also popular and well-regarded data mining and analytics tools.
* **Free or Open Source tools:**
* **Rand PL/R:** R was described earlier in the model planning phase, and PL!R is a procedural language for PostgreSQL with R. Using this approach means that R commands can be executed in database.
* This technique provides higher performance and is more scalable than running R in memory.
* **Octave:** a free software programming language for computational modeling, has some of the functionality of Matlab. Because it is freely available, Octave is used in major universities when teaching machine learning.
* **WEKA:** is a free data mining software package with an analytic workbench. The functions created in WEKA can be executed within Java code.
* **Python**: python is a programming language that provides toolkits for machine learning and analysis, such as scikit-learn, numpy, scipy, pandas, and related data visualization using matplotlib.
* **SQL:** SQL in-database implementations, such as MADlib, provide an alternative to in-memory desktop analytical tools. MADiib provides an open-source machine learning library of algorithms that can be executed in-database, for PostgreSQL or Greenplum.

**Q6. B) Explain “Operationalize” Phase of Data analytic lifecycle in detail. [05M]**

* In the final phase, the team communicates the benefits of the project more broadly and sets up a pilot project to deploy the work in a controlled way before broadening the work to a full enterprise or ecosystem of users.
* The team scored the model in the analytics sandbox, represents the first time that most analytics teams approach deploying the new analytical methods or models in a production environment. Rather than deploying these models immediately on a wide-scale basis, the risk can be managed more effectively and the team can learn by undertaking a small scope, pilot deployment before a wide-scale rollout.
* This approach enables the team to learn about the performance and related constraints of the model in a production environment on a small scale and make adjustments before a full deployment.
* During the pilot project, the team may need to consider executing the algorithm in the database rather than with in-memory tools such as R because the run time is significantly faster and more efficient than running in-memory, especially on larger datasets.
* While scoping the effort involved in conducting a pilot project, consider running the model in a production environment for a discrete set of products or a single line of business, which tests the model in a live setting.
* This allows the team to learn from the deployment and make any needed adjustments before launching the model across the enterprise.
* Be aware that this phase can bring in a new set of team members- usually the engineers responsible for the production environment who have a new set of issues and concerns beyond those of the core project team.
* This technical group needs to ensure that running the model fits smoothly into the production environment and that the model can be integrated into related business processes. Part of the operationalizing phase includes creating a mechanism for performing on-going monitoring of model accuracy and, if accuracy degrades, finding ways to retrain the model.
* If feasible, design alerts for when the model is operating "out-of-bounds." This includes situations when the inputs are beyond the range that the model was trained on, which may cause the outputs of the model to be inaccurate or invalid.
* If this begins to happen regularly, the model needs to be retrained on new data.
* Often, analytical projects yield new insights about a business, a problem, or an idea that people may have taken at face value or thought was impossible to explore.
* Four main deliverables can be created to meet the needs of most stakeholders.

The key outputs for each of the main stakeholders of an analytics project and what they usually expect at the conclusion of a project.

**1. Business User:**

* Typically tries to determine the benefits and implications of the findings to the business.

**2. Project Sponsor:**

* Typically asks questions related to the business impact of the project, the risks and return on investment (ROI), and the way the project can be evangelized within the organization (and beyond).

**3. Project Manager:**

* Project Manager needs to determine if the project was completed on time and within budget and how well the goals were met.

**4. Business Intelligence Analyst:**

* Needs to know if the reports and dashboards he manages will be impacted and need to change.

**5. Data Engineer and Database Administrator (DBA):**

* Typically need to share their code from the analytics project and create a technical document on how to implement it.

**6. Data Scientist:**

* Needs to share the code and explain the model to her peers, managers, and other stakeholders.

**-------------------------------------------------THE END--------------------------------------------------**