DATA VISUALIZATION WITH POWER BI

MCQS FUNDAMENTALS OF DATA:

1][i]THE ABILITY TO ASK QUESTIONS AND LEARN NEW THINGS?

ANS: INFORMATION

[ii]THE APPLICATION OF INFORMATION WHICH IS MEASURED BY THE ABILITY TO DO THINGS?

ANS: KNOWLEDGE

[iii]THE ABILITY TO USE YOUR KNOWLEDGE AND EXPERIENCE TO MAKE GOOD DECISIONS AND JUDGEMENTS?

ANS: WISDOM

[iv]CONTEXTUAL, ORGANIZED AND VETTED DATA THAT CONVEY SOME SORT OF TREND OR PATTERN?

ANS: INFORMATION

[v]RAW AND UNSTRUCTURED FACTS, NUMBERS, OR FIGURES WHICH CONVEY A MESSAGE?

ANS: DATA

2]ALEX IS IN THE PROCESS OF CREATING A REPORT THAT DISPLAYS THE RESULTS OF A SURVEY. WHICH DATA TYPE BEST DESCRIBES THE DATA THAT ALEX IS DEALING WITH?

ANS: OBSERVATIONAL

3]A UNIVERSITY PROFESSOR COLLECTS INFORMATION ABOUT THE CLASSIFICATION OF HER STUDENTS AS FRESHMEN, SOPHOMORES, JUNIORS, SENIORS. THE DATA IS THEN DISPLAYED IN A BAR CHART. WHAT TYPE OF DATA IS THE UNIVERSITY PROFESSOR COLLECTING?

ANS: QUALITATIVE DATA

4][i]GETS A BACUP FOR ALL DATA WITHIN THE HARD DRIVE?

ANS: FULL BACKUP

[ii]GETS A BACKUP OF THE DATA GENERATED OR REVISED SINCE THE LAST BACKUP, REGARDLESS OF THE TYPE OF THE LAST BACKUP?

ANS: INCREMENTAL BACKUP

[iii]GETS A BACKUP OF DATA FOR THE PAST N YEARS?

ANS: DIFFERENTIAL BACKUP

[iv]GETS A BACKUP OF THE DATA GENERATED OR REVISED SINCE THE LAST FULL BACKUP?

ANS: DIFFERENTIAL BACKUP

5]DATA MIGRATION STEPS?

ANS: PLANNING, ANALYZING THE DATA, DESIGN, IMPLEMENTATION, TESTING, FINAL MIGRATION

6]WHICH CHARACERISTICS DO ALL DATA MIGRATION PROJECTS HAVE IN COMMON?

ANS: THEY ALL REQUIRE PERFORMING PROPER ETL MAPPING TO ENSURE CONSISTENCY AND COMPATIBILITY, THEY ALL BEGIN WITH DATA INGESTIONS AND CLEANSING OF THE DATA PRIOR TO INTEGRATION, THEY ALL REQUIRE IDENTIFYING SOURCE AND TARGET SYSTEMS.

7]RAVI WANTS TO CREATE A DATA VISUALIZATION TO SHOW WHICH PARTS OF HIS COMPANY WEBSITE ARE RECEIVING THE MOST CLICKS AND ARE BEING MOST VIEWED BY HIS VIEWERS. WHICH DATA VISUALIZATION WILL PROVIDE DAN WITH A VISUAL THAT IS EASY TO ASSIMILATE AND MAKE DECISIONS FROM?

ANS: HEAT MAP

8][i]NAME OF THE TABLE TO PULL DATA FROM?

ANS: FROM

[ii]CREATE GROUPS TO SUMMARIZE DATA?

ANS: GROUP BY

[iii]APPLIES FILTERING LOGIC TO LIMIT RECORDS IN YOUR RESULTS?

ANS: WHERE

[iv]DESCRIBES HOW YOU WANT YOUR DATA SORTED?

ANS: ORDER BY

[v]APPLIES FILTERING LOGIC TO YOUR GROUPS?

ANS: HAVING

MCQS DATA MANAGEMENT SYSTEMS:

1]SUPPOSE YOU HAVE A COMPANY OF 50 EMPLOYEES AND YOU ARE WRITING CODE FOR A VERY SPECIFIC TYPE OF PROGRAM. THERE ARE FIVE VENDORS THAT PROVIDE YOU THE GRAPHICAL SUPPORT AND YOUR TARGET CLIENT ARE VERY SMALL BUSINESS. WHICH STATEMENT WILL BE TRUE IN THE DOMAIN OF DATA MANAGEMENT?

ANS: YOU DONOT NEED A DATA MANAGEMENT PROGRAM AT ALL

2]WHAT WILL BE THE BEST APPROACH OF DATA MANAGEMENT FOR A MULTI-FACETED ENTERPRISE WITH MULTIPLE DOMAINS?

ANS: A COMBINATION OF TOP-DOWN AND MIDDLE-OUT APPROACHES

3]WHICH IS NOT A TYPE OF DATA THAT WOULD BE ENCOUNTERED IN AN ENTERPRISE?

ANS: LOGICAL

4]WHAT IS THE PRIMARY REASON FOR DATA INTEGRATION ACROSS DOMAINS?

ANS: **PROVIDE A UNIFIED SINGLE VERSION OF THE DATA**

5]WHICH CORRECTLY DESCRIBES THE DIFFERENCE BETWEEN STATIC AND DYNAMIC DATA?

ANS: **STATIC DATA DOESNOT REQUIRE UPDATING WHEREAS DYNAMIC DATA REQUIRES REGULAR UPDATING**

6]IN YOUR MLTI-DOMAIN ENTERPRISE WHERE THE PRIMARY FUNCTION IS OF A STOCK MARKET BROKER AND WHERE YOU NEED REAL-TIME DATA SYNCHRONIZATON WHAT WILL BE THE REQUIRED STYLE OF ARCHITECTURE NEEDED FOR THE DATA MANAGEMENT PROGRAM?

ANS: **HYBRID STYLE**

7]WHICH IS NOT A FUNCTION OF ENTITY RESOLUTION?

ANS: **DATA PROPAGATION**

8]IN YOUR ENTERPRISE WHERE YOU ARE PLANNING TO DEVELOP A PERFECTLY ALIGNED SYSTEM ACROSS ALL THE DOMAINS, WHICH FUNCTION WILL YOU DEEM AS NOT NECESSARY FOR BUILDING A TRULY ALIGNED SYSTEM?

ANS: **DATA QUALITY PROGRAM**

9]WHAT IS THE PROPER ORDER FOR THE LEVELS OF MANAGEMENT AND CONTROL FROM A MEASUREMENT OF MATURITY POINT OF VIEW?

ANS: **UNSTRUCTURED, STRUCTURED, MANAGED, OPTIMIZED**

10]WHICH IS NOT A GOAL OF METADATA MANAGEMENT?

ANS: **DATA REFERENCE**

MCQS DATA GOVERNANCE:

1]WHICH IS NOT THE KIND OFJOB THAT IS TO BE INCLUDED IN THE ROLE OF A DATA STEWARD?

ANS: **DECIDING WHICH INFORMATION IS TO BE PROVIDED TO WHICH INDIVIDUAL**

2]WHAT IS THE INTEGRAL MEANING OF DATA HARMONIZATION?

ANS: **DATA HARMONIZATON MEANS OBTAINING A SINGLE VERSION OF THE TRUTH**

3]WHICH CAN BE A MAJOR PROBLEM IN COMPANIES THAT HAVE A "BRING YOUR OWN DEVICE" POLICY?

ANS: **LEAKAGE OF INFORMATION WHEN THE DEVICE IS USED IN A PUBLIC NETWORK**

4]WHEN IMPLEMENTING YOUR DATA GOVERNANCE PRACTICE IN YOUR ORGANIZATION, WHICH IS THE ONE FACTOR THAT WILL NOT PLAY A MAJOR ROLE WHEN CONSIDERING THE IMPLEMENTATION?

ANS: **DATA QUALITY**

5]WHICH CAN BE CONSIDERED AN ADVANTAGE OF IT SYSTEMS IN PREVENTING SECURITY THREATS?

ANS: ACCESS CONTROL IS LIMITED TO A PARTICULAR SYSTEM

6]WHICH ARE THE TWO FACTORS CRITICAL TO AN ORGANIZATION WHEN CONSIDERING THE REGULATION OF DATA PRIVACY?

ANS: STORAGE NETWORKING INDUSTRY ASSOCIATION AND GENERAL DATA PROTECTION REGULATION

7]WHICH IS NOT A PRIMARY RULE FOR DATA VALIDATION?

ANS: GOVERNANCE PRACTICES

8]WHICH IS NOT TRUE ABOUT THE ENTITY RESOLUTION PROCESS?

ANS: APPLICATION OF BUSINESS AND DATA QUALITY RULES IS NOT A STEP IN ENTITY RESOLUTION

9]IN YOUR MULTI-DOMAIN ENTERPRISE WHERE THERE ARE MULTIPLE VENDORS AND MULTIPLE RESOURCES, WHICH WILL BE A KEY DECIDING FACTOR IN HOW YOU WILL MANAGE THE CRUD?

ANS: ENTERPRISE RESOURCE PLANNING

10]IN BUILDIN YOUR DATA GOVERNANCE PRACTICE, WHICH WILL NOT BE AN OBJECTIVE FOR THE GOVERNANCE PRACTICE?

ANS: DIVIDING THE BUSINESS

MCQS DATA QUALITY MANAGEMENT:

1]WHICH IS AN ESSENTIAL FACTOR TO ENSURE GOOD DATA QUALITY?

ANS: FIX DATA QUALITY ISSUES

2]WHICH IS NOT A DATA QUALITY AIMED PROJECT INITIATIVE?

ANS: DATA INTEGRATION

3]IN YOUR DATA DRIVEN ENTERPRISE, HOW WILL YOU ENSURE THAT GOOD QUALITY DATA IS BEING MAINTAINED AND RECONCILED ACROSS SYSTEMS?

ANS: ENSURE SOURCE TO TARGET CONSISTENCY

4]WHAT IS MEANT BY REFERENCE DATA?

ANS: DATA THAT IS USED FOR CLASSIFICATION OF OTHER DATA

5]WHAT ARE THE ADVANTAGES OF A GOOD DATA COMPLIANCE STRATEGY?

ANS: PROACTIVE ENVIRONMENT, ORGANIZATIONAL GROWTH AND CUSTOMER RELATIONSHIPS

6]WHICH IS PERFORMANCE MEASUREMENT LEVEL?

ANS: STRATEGIC MEASUREMENT

7]WHICH IS NOT A CONTINUOUS IMPROVEMENT METHOD IN DATA GOVERNANCE?

ANS: OVERLOOK ACHIEVABLE GOVERNANCE MATURITY MILESTONES

8]WHICH FUNCTION IS A MUST FOR DATA COMPLIANCE BUT NOT REQUIRED AT ALL FOR DATA MANAGEMENT AND DATA GOVERNANCE?

ANS: INTER-RELATIONS

9]WHICH IS A MAJOR SOLUTION TO ADDRESS THE DATA GOVERNANCE AND COMPLIANCE ISSUES?

ANS: KNOWING THE DATA

10]WHICH IS NOT AN ADVANTAGE OF DATA LAKEHOUSE?

ANS: PROVIDES AN ON-PREMISES DATA WAREHOUSE SOLUTION

MCQS TQ CLOUD:

1.COMPANY has signed a multi-year agreement to help a mobile telecom provider migrate its applications to the Cloud over a period of several years, as it seeks to implement a leaner operating model.Which activity should COMPANY execute first to provide the client with an optimal Cloud solution/landscape?
Ans: **define a strategy and roadmap**

2.An international fast-food chain is looking for opportunities to leverage the Cloud to gain insights into customer data, including location, and order patterns. The client has engaged COMPANY to help create customer-focused experiences, including promotions.
They need COMPANY's help with this campaign and to create a unique user experience for all their customers.
Which technology, when partnered with Cloud, would allow this to occur?
Ans : **Artificial intelligence**

3.A company offers customers access to servers storage and networking resources within their data center. What type of service does this represent?
Ans : **infrastructure as a service**.

4.Based on highly volatile seasonal workloads, a client is considering moving some workloads to the cloud, while retaining highly sensitive data and apps on-premise. Which deployment model would be the best option?

Ans. Private cloud

5.A large gambling company needs to be able to accept high volumes of customer wagers within short timeframes for high-profile sporting events. Additionally, strict laws prohibit any gambling activities outside of specially-licensed business zones. What is an example of an effective, elastic Cloud solution that can meet this client's needs?
Ans. A mobile app that only accept wagers based on the users location

6.An COMPANY client is eager to begin their cloud transformation but needs guidance on selecting the right approach to meet their specific needs and objectives. Which asset can help this client discover assess design and simulate end to end cloud migration solutions at scale
Ans: myNav

7. Which clouds service type generates the most revenue?
Ans: infrastructure as a service

8. Why is service level agreement (SLA) important?
Ans. It documents expectations of availability ,up time and security

9 what is an example of using the internet of things (IOT) to delivery in over to cloud based solutions to customers?
Ans: wearable technology that provides customers with on-the-spot personalized experience

10. The all India cricket council is experiencing sporadic activity on their ticketing websites depending on the announcement of maths schedules the required high speed crossing during peak times but otherwise website traffic is nominal which unique cloud feature seamlessly helps manage the increase and decrease in traffic?
Ans . Elasticity

11. What is a feature of passengers my now platform?
Ans. Helps asses architect and simulate cloud solution to determine the right cloud architecture based on our clients business needs

12.HealthyApp, a new COMPANY H&PS client, specializes in developing custom software for hospitals and other medical facilities. For the upcoming fiscal year, they want to move to a more agile and innovative software development model, in which their developers can focus on application features and functionality, rather than underlying environments and infrastructure. Which Cloud computing model would provide HealthyApp with the framework to help achieve this goal?
Ans . Platform as a service (paas)

13. The activities bikes during Black Friday sales and the holiday session put significant stress on a major retailers website to what feature of cloud computing would resolve this problem?
Ans: elasticity

14. Which statement accurately describes the concept of elasticity in cloud computing?
Ans :the degree to which a system can adopt to workload changes by provisioning and deprovisioning resources automatically

15. Which service is major commercial cloud platform?
Ans . Amazon Web services AWS

16. Sarah's business offer a service where multiple users applications are hosted on a server at the same time based on the scenario on which facet of cloud computing does saras business rely
Ans. Multitenancy

17. A healthcare client approaches COMPANY for help in defining their Cloud migration strategy. During the initial discussions, the client explains that there are very specific regulatory and other compliance requirements within their industry. Which Cloud deployment model would give this client complete control over their data and security?
Ans: private cloud

18.Most companies are using Cloud. However, many companies fail to achieve the anticipated benefits from their Cloud investments. Of the companies

COMPANY speaks with, what percentage realize the full benefits they expected to get from the Cloud?

Ans **75%**

19.The Ocean Medallion is a wearable device that connects virtually to various devices on board a cruise ship. When a guest is wearing the medallion, they can order a drink with their smartphone or open their cabin door automatically by standing in front of it.

20.Which emerging technology does Cloud combine with to enable this medallion's features?

Ans: **internet of things (IOT)**

21. THE COMPANY team is involved in helping a client in the transformation journey using cloud computing how is mine are beneficial to the team?

Ans. **It helps access architect and simulate cloud migration solutions**

22.COMPANY began its own Cloud journey in 2015, after our internal IT organization recognized the significant advantages to be gained in leveraging Cloud capabilities at scale. Today, which percentage of COMPANY's infrastructure and processes are in the Cloud?

Ans: **95%**

23. What is a characteristics of cloud computing?

Ans . **It allows multiple customers to share application while retaining data privacy**

24. What is the difference between Infrastructure as a Service (IaaS) and Software as a Service (SaaS)?

Ans .**Iaas is Cloud computing that provides hardware and software tools over the internet. SaaS is Cloud computing that provides database solutions over the internet,**

25. What is a characteristic feature of multi tenancy?

Ans . **Multiple users share or rent the same server**

26.An international fast-food chain is struggling to identify clear worldwide trends in customer preferences due to various countries' data being in different languages.How could a Cloud solution help the company address this problem?
Ans: **by creating cuisines that are specific to each geographical location**

27. Which benefits does the cloud provide to startup companies without access to large funding?
Ans. **The ability to pay as you go and create another ornaments with little upfront costs.**

28.Based on highly volatile seasonal workloads, a client is considering moving some workloads to the cloud, while retaining highly sensitive data and apps on-premise. Which deployment model would be the best option?
Ans . **Private cloud**

29.To help build credibility with a potential new client, an COMPANY Sales Lead shares the story of COMPANY's own journey to the Cloud. Which statement is true about COMPANY's own Cloud transformation journey?
Ans . **COMPANY has implemented our cloud solution at scale and delivered clear business values**

30.COMPANY is meeting with a well-known, global financial institution that specializes in private and commercial banking to discuss the migration of its sizable application ecosystem to the Cloud. The client has quickly become overwhelmed with the available choices, complexity, and security. Which tool helps the client self-assess their Cloud landscape to determine the best solution?

Ans . **MyNav**

31.An electronics company wants to use a type of Cloud computing architecture that would enable them to use Microsoft Azure as a platform, while storing and retrieving data in AWS.Which type of Cloud deployment would meet the company's business needs?
Ans: **hybrid cloud**

32 A Software Developer wants to add a new feature to an existing application operating in the Cloud, but only wants to pay for the computing time the code actually uses when it is called. Which term describes this feature of computing?
Ans: **serverless computing**

33. The activity spikes during Black Friday sales and the holiday season put significant stress on a major retailer's website. What feature of Cloud computing would resolve this problem?
Ans. **Elasticity**

34.A company offers customers access to servers storage and networking resources within their data center. What type of service does this represent?
Ans. **Infrastructure as a service (iaaS)**

35. Why is a Service Level Agreement (SLA) important?
Ans. **It documents expectations of avaliability, uptime, and security**.

36. What is a characteristic of cloud computing?
Ans: **it allows multiple customers to share application while retaining data privacy**

37.A healthcare client approaches COMPANY for help in defining their Cloud migration strategy. During the initial discussions, the client explains that there are very specific regulatory and other compliance requirements within their industry. Which Cloud deployment model would give this client complete control over their data and security?
Ans. **Private cloud**

38. COMPANY began its own Cloud journey in 2015, after our internal IT organization recognized the significant advantages to be gained in leveraging Cloud capabilities at scale. Today, which percentage of COMPANY's infrastructure and processes are in the Cloud?
Ans .**95%**

39. Which statement accurately describes the concept of elasticity in cloud computing?

Ans. The degree to which a system can adapt to workload changes by provisioning and deprovisioning resource automatically

40. In future where will Yan increasing majority of data come from?
Ans **. Individual users and devices**

41.COMPANY has signed a multi-year agreement to help a mobile telecom provider migrate its applications to the Cloud over a period of several years, as it seeks to implement a leaner operating model.Which activity should COMPANY execute first to provide the client with an optimal Cloud solution/landscape?
Ans **. Define a strategy and roadmap**

42. What would be an ideal scenario for using age computing solutions?
Ans **. A hospital department that perform high risk time sensitive surgeries.**

43. A leading pharmaceutical company is launching a major transformation of its entire cloud operations in order to more quickly and efficiently perform research and conduct clinical trials how can a cloud first strategy helped this client?
Ans: **by providing a data centric approach that offers more speed agility and security**

44.A Chief Technology Officer has seen the company grow significantly in the last three months. Historically, the application was hosted with on-premise hardware, but recently moved some components of the application to the Cloud. Which type of Cloud environment is being used?
Ans : **Hybrid cloud**

45. Which term describes a cloud provider allowing more than one company to to share or rent the same server?
Ans. **Multitenancy**

46. How does a cloud first strategy approach clients migration to the cloud?
Ans. **By bringing multiple services together to serve the client's business needs**

47. How has the covid-19 pandemic affected business relationship to cloud computing?
Ans. It has accelerated the urgency for a business to move to the cloud quickly

48. In which type of situation would it make sense to use it?
Ans. Where critical decision must be made on a split second basis.

49.Healthy App, a new COMPANY H&PS client, specializes in developing custom software for hospitals and other medical facilities. For the upcoming fiscal year they want to move to a more agile and innovative software development model in which their developers can focus on application features and functionality rather than underlying environments and infrastructure. Which Cloud computing model would provide Healthy App with the framework to help achieve this goal?
Ans . Platform as a service (Paas)

50. A company of customer access to services storage and networking resources within their data centres what type of service does this represent?
Ans. Infrastructure as a service (IAAs)

MCQS TQ DATA:
1. Which scenario best illustrates the implementation of data governance?

Ans. Drafting a press release based on that are extracted from the news of the day

2. How does COMPANY helps companies harness the power of data to achieve optimal business outcomes?

Ans. They develop data governance frameworks that go on to form the basis of the company data analytics strategy.

3. What is the main difference between structured and unstructured data?
Ans. Structured data is easier to such and analyse while and structure that are requires more effort to process

4. A client approaches COMPANY for a solution to bolate the large amounts of data it has collected over the years from various sources which solution would help the client pool their data together?
Ans . A cloud analytics solution

5. What does the DIKW pyramid represents?
Ans. It is a model used to represent the functional relationship between data information knowledge and wisdom.

6. What does data warehousing allows organisation to achieve?
Ans . Allows for analytics and reporting across the entire organisation

7. How to assign learning and artificial intelligence(AI) technologies help business use their enterprise data effectively?
Ans: they curate data quickly for multiple business scenarios

8. What is an advantage of storing data in data Lake without applying a specific schema to it initially?
Ans. It allows more flexibility to use the data in various innovative ways.

9. What is an example of big data?
Ans . Providing real time data feeds on millions of people with variable devices.

10. How do data platforms help business with their data governance needs?
Ans . By controlling security and accessability

11. Which technology is combined to make that are critical organisational asset?
Ans . Machine learning and artificial intelligence (AI)

12. Why is it important for companies to invest in building a complete data and analytics platform?

Ans . **It produces cultural support and alignment a growth mindset and new ideas and priorities to improve business process**

13. How does the implementation of an enterprise wide data and analytics strategy help organisation?

Ans. **By housing data in oneplus and converting it into easily consumerical information at high speed**

14. A mobile application delivers market predictions based on stock data from the stock market data platform knowing that the data platform can stream its data how should the application interact with the data platform to deliver predictions in real time?

Ans. **Interact programmatically through the data platform API (application programming interface)**

15. Yah healthcare provider approaches COMPANY to help them increase their efficiency through an advance data science platform what is the first step essential could take ?

Ans . **Train the client to independently manage their data**

16. What is an advantage of using fully integrated cloud based data analytics platform?

Ans . **It provides the computing power needed to convert raw data into meaningful information for decision makers**

17. Why is data driven analytics of interest to companies?

Ans . **It helps data analyst shape and analytics problem from a business problem**

18. Which example illustrate the idea of collecting data?

Ans. **Mining all product reviews from an online store for the processing**

19. What happens when a data set include records with missing data?

Ans . **Itna the data set which must then be discarded and recollected**

20. Which comprehensive approach should companies follow to trust and use that as a critical differentiater?
Ans . **Develop clear data governance policies with a strong data strategy**

21. A client approaches COMPANY for resolution to collect the large amount of data it has collected over the years from various sources which solution would help the client pool there that are together?
Ans. **A cloud analytics solution**

22. How does COMPANY help companies turnless meaningful data collection into an effective data strategy?
Ans . **By recommending the tool and platforms to drive the data analytics governance and process**

23. Where does change management play a major role in transforming client business into a data driven intelligence enterprise?
Ans. **Data governance and management**

24. What is the most important action that organisations should take with the data they captured about their customers?
Ans. **process and organise it into meaningful information.**

25 what differentiates data from information?
Ans. **Information is contextualized data is generic**

26. How does COMPANY help companies turnless meaningful data collection into an effective data strategy?
Ans. **by recommending the tools and platform to drive the data analytics governance and process**

27. What does COMPANY help business implement to ensure their data is trustworthy and reliable?
Ans . **Data governance and management policies**

28. What role does the data information knowledge wisdom DIKW pyramid plane managing a data driven project?

Ans. It serves as a linear model to show the various ways of extracting insights and value from all types of data

29. How should a common data source like social media comments be categorised?
Ans . Unstructured data

30. A client approach Assassin Safari solution to call at the large amount of data was collected over the years from various sources which solution would help the client pool their data together?
Ans. A cloud analytics solution

31. The department manager researches new data plans from the company and request the list of essential features which essential features should the list included?
Ans . Strong security centralisation programmetic access

32. How does a fully integrated data and analytics platform enable organisation to converted data into consumeable information and insight?
Ans. Buy creating analytics reports and build mission learning modules to refined predictive capabilities

33 . What is an advantage of using a fully integrated cloud based data analytics platform?
Ans. It provides the computing power needed to convert raw data into meaningful information for decision makers.

34. COMPANY is involved in the transformation of your business organisation into a data-driven and intelligent enterprise which ability does COMPANY need to ensure the organisation has become a successful business?
Ans . Together we organise and store different categories of all raw data for easy and fast retrieval

35. A department manager is planning a presentation on how data platforms work but need to simplify the language for the audience to better understand the main steps which three steps most accurately describe the managers approach?

Ans . **Why are data soils problem atticgather organise and share data**

36. What does data warehousing allow organisation to achieve?
Ans. **yellows for analytics and reporting across the entire organisation.**

37. What is the key object of data analysis?
Ans . **To find meaning in data and use it to make inform decisions**

38. Which technology is combined to make that a critical organisational at asset?
Ans . **Mission learning and artificial intelligence (AI)**

39. How should a company adopt a data-driven culture that will stick?
Ans. **use change management to transformer how the company things about data.**

40.COMPANY is helping a large retailer transform their online sales and services. A data analyst audits the client's customer journey and maps out the kind of data they need to collect at each touch point. How does this help the client achieve their objectives?
Ans. **By increasing their customer knowledge and leveraging that information to improve customer experience.**

41. Everyone with in a company needs to be aware of what data can do to improve business process and how to make it happen with concept is refer to in the statement above?
Ans . **Data literacy**

42. What is the next up and organisation should take after capturing and collecting data?
Ans . **Process and curate data**

43. What does data warehousing yellow organisation to achieve?
Ans . **Yellow for analytics and reporting across the entire organisation**

44. What is an example of big data?
Ans. **Providing real time data feeds on millions of people with wearable devices**

45. Why are data soils problematic?

Ans . **Enter the process of collecting deep actionable insights from organisational data and create a barrier to viewing company data holistically**

46 what function do regulations like the general data protection regulation GDPR serve?

Ans. **To enforce the protection of consumer data through criminal penalties against non complaint corporations**

47. What is an advantage of using a fully integrated cloud based data analytics platform?

Ans. **It provides the computing power needed to convert raw data into meaningful information for decision making.**

48. COMPANY processes and migrates data for an E-Commerce frame to billion advanced data platform how does this platform benefit the firm?

Ans . **Increases business returns**

49. The city of alums field is using electronic sensors to collect data on traffic patterns in pollution and crime this data is used to manage assets and the resources more efficiently and in real time which of these teams would better define the data collected from the sensors in this scenario?

Ans. **Big data**

50. A large hotel chain approaches Accentur for advice the company has a accumulated a vast amount of data during its years of business but it uncertain how to put the data to use how can a sensor best help the client?

Ans . **By helping the client development enterprise by data and analytics strategy**

51. What is the purpose of a company's data strategy?

Ans. **To collect store manager share and use data**

52 what should a company do to develop a better culture?

Ans. **Ensure that everyone in the company understands the importance of data and are there of what data can do to improve business processes**

53. A client approaches are essential for a solution to call at the large amount of data that has collected over the ear from various sources which solution would help the client will get that are together?
Ans. Proud analytics solution

54. How can the adoption of data platform simplify data governance for an organisation?
Ans. By checking the integrity of data that is currently stored in the organisations data lake

55. How does machine learning and artificial intelligence yeah AI technologies help business use their enterprise data effectively?
Ans. The capture all the data in real time or near real time.

MCQS DATA WAREHOUSE ESSENTIAL: CONCEPTS

1]IDENTIFY ESSENTIAL FEATURES OF STRATEGIC INFORMATION?

ANS: PRESERVES DATA INTEGRITY, TIME-VARIANT

2]WHICH OF THE FOLLOWING OLAP STATEMENTS ARE TRUE?

ANS: OLAP PROVIDES REAL-TIME ANALYTICAL CAPABILITY, OLAP IS A PART OF THE OVERALL DATA WAREHOUSE IMPLEMENTATION

3]WHICH STATEMENTS ABOUT CLOUD-BASED DATA WAREHOUSE ARE TRUE?

ANS: CLOUD-BASED DATA WAREHOUSE ARE SCALABLE, CLOUD-BASED DATA WAREHOUSES ARE ELASTIC

4]MATCH THE DATA WAREHOUSING SOLUTIONS WITH THEIR ASSOCIATED BENEFITS?

ANS: ON-CLOUD DATA WAREHOUSE-PROVIDES SCALABILITY, COST EFFECTIVE, OFFERS BETTER SPEED AND CONNECTIVITY. ON- PREMISE DATA WAREHOUSE-PREFERRED IN BANKINF OR GOVERNMENT DOMAIN, OFFERS ABSOLUTE CONTROL OVER SECURITY

5]WHICH PROCESSES INVOLVED IN A DATA WAREHOUSE PROJECT ARE IMPORTANT?

ANS: DATA CLEANING, ETL

6]MATCH THE DATA MODELLING STRATEGIES WITH THEIR FEATURES?

ANS: DENORMALIZATION-USED ON PREVIOUSLY NORMALIZED DATABASES TO INCREASE, ADDS REDUNDANT DATA, COMBINES TABLES. NORMALIZATION-APPLIES FORMAL RULES TO ENFORCE DEPENDENCIES, ENSURES THE DEPENDENCIES ARE PROPERLY ENFORCED, SPLITS TABLES

7]IDENTIFY SOME OF THE ESSENTIAL FEATURES WHICH DIFFERENTIATES A DATA WAREHOUSE FROM OLTP?

ANS: DATA WAREHOUSE PROVIDES PREDICTIVE ANALYTICAL CAPABILITIES, DATA WAREHOUSE STORES HISTORICAL DATA

8]WHICH OF THE FOLLOWING LOCAL AND GLOBAL WAREHOUSE STATEMENTS ARE TRUE?

ANS: WE CAN PROVISION A SINGLE GLOBAL REPOSITORY FOR A PARTICULAR DOMAIN, LOCAL DATA WAREHOUSE CANNOT BE ACCESSED GLOBALLY

9]IDENTIFY THE TERMINOLOGIES THAT WE GENERALLY USE IN DATA WAREHOUSING?

ANS: ETL, DIMENSIONAL MODEL

10]WHICH ARE ESSENTIAL TASKS THAT WE CAN EXECUTE TO FACILITATE BUSINESS INTELLIGENCE?

ANS: EXTRACT, LOAD

11]SELECT DATA MODELLING STRATEGIES THAT WE CAN ADOPT TO CREATE AN ER MODEL?

ANS: NORMALIZATION, DENORMALIZATION

12]IDENTIFY ESSENTIAL TABLE TYPES THAT WE CAN USE TO IMPLEMENT A STAR SCHEMA?

ANS: DIMENSION TABLE, FACT TABLE

13]WHICH STATEMENTS ABOUT SNOWFLAKE ARE TRUE?

ANS: A SNOWFLAKE SCHEMA IS AN EXTENSION OF THE STAR SCHEMA, APPLICATION BINARY INTERFACE ALLOWS US TO CONTEXTUALIZES CONTRACTS

14]CHOOSE THE CHARACTERISTICS OF WEIGHTED REPORTS?

ANS: WITH WEIGHTED REPORTS WE GET A MEANINGFUL SUBTOTAL AND TOTAL, WEIGHTED REPORTS MULTIPLY ALL THE FACTS BY WEIGHT BEFORE AGGREGATING

15]IDENTIFY THE ESSENTIAL LEVELS OF MANAGEMENT THAT REQUIRE STRATEGIC REPORTS?

ANS: STRATEGIC, MIDDLE MANAGEMENT LEVEL

16]IDENTIFY THE ESSENTIAL DIFFERENCES BETWEEN RDBMS AND DATA LAKES?

ANS: RDBMS DATABASES ARE TRANSACTION WHILE DATA LAKES ARE NOT TRANSACTIONAL, RDBMS DATABASES DEFINES FIXED SCHEMA WHILE DATA LAKE FOLLOWS NO SCHEMA DESIGN

17]IDENTIFY THE ESSENTIAL COMPONENTS OF AZURE DATA LAKE?

ANS: SQL SERVER, DATA FACTORY

MCQS DATA WAREHOUSE ESSENTIAL: ARCHITECTURE FRAMEWORKS AND IMPLEMENTATION:

1]WHAT ARE SOME OF THE INTEGRATED COMPONENTS OF A DATA WAREHOUSE?

ANS: DATA STAGING, STORAGE

2]WHICH OF THE FOLLOWING STATEMENTS CORRECTLY DEFINES THE CHARACTERISTICS OF THE KIMBALL MODEL?

ANS: KIMBALL MODEL USES THE DIMENSIONAL MODEL, IN THE KIMBALL MODEL THE ANALYTICAL SYSTEMS CAN ACCESS THE DATA DIRECTLY

3]SPECIFY SOME OF THE ESSENTIAL LOGICAL COMPONENTS OF A DATA WAREHOUSE?

ANS: ENTITIES, ATTRIBUTES

4]WHAT ARE SOME OF THE IMPORTANT TASKS THAT WE NEED TO PERFORM TO IMPLEMENT A PHYSICAL MODEL FOR A GIVEN LOGICAL MODEL?

ANS: CREATE TABLES TO REPRESENT THE ENTITIES, CREATE FOREIGN KEYS TO ESTIBLISH THE RELATIONSHIPS AMONG OBJECTS

5]WHICH OF THE FOLLOWING DIMENSIONAL COMPONENTS DIFFERENTIATES A DIMENSIONAL MODEL FROM AN ER MODEL?

ANS: FAST TABLES, DIMENSIONAL TABLES

6]SPECIFY SOME OF THE OUTCOMES OF A DATA WAREHOUSE REALIZATION?

ANS: INTUITIVE DASHBOARDS, PREDICTIVE ANALYTICAL REPORTS

7]IDENTIFY ESSENTIAL TASKS INVOLVED IN AN ETL PROCESS?

ANS: EXTRACTING DATA FROM DIVERSIFIED SOURCES, TRANSFORMING EXTRACTED DATA

8]SELECT PROMINENT ETL TOOLS THAT WE CAN USE WITH A DATA WAREHOUSE IMPLEMENTATION?

ANS: POWERCENTER INFORMATICA, TALEND

9]WHAT ARE SOME OF THE ESSENTIAL TASKS THAT WE CAN EXECUTE USING TALEND?

ANS: ETL JOB DESIGNING, BUSINESS MODELLING

10]WHICH OF THE FOLLOWING COMPONENTS ARE PROVIDED BY TALEND TO DESIGN ETL JOBS?

ANS: TFILEINPUT, TMAP

MCQS MODERN DATA WAREHOUSES:

1]WHAT IS NOT A COMPONENT OF AN ON-PREMISES DATA WAREHOUSE?

ANS: PROCESSING SPACE

2]YOU HAVE AN ON-PREMISES DATA WAREHOUSE ALREADY INSTALLED IN YOUR ORGANIZATION. IN THE PERIOD FOLLOWING THE COVID PANDEMIC, YOUR BUSINESS STARTED TO GROW EXPONENTIALLY, AND YOU WERE TIRED OF ADDING MORE NODES TO THE PHYSICAL STORAGE OF THE DATA WAREHOUSE. YOU DECIDE TO MODERNIZE YOUR DATA WAREHOUSE AND MAKE IT CLOUD-BASED. WHAT MAJOR ADVANTAGE WILL YOU ACHIEVE BY MODERNIZING YOUR DATA WAREHOUSE?

ANS: SPEED UP TIME AND ANALYTICS

3]YOU HAVE INCORPORATED THE USAGE OF AMAZON REDSHIFT IN YOUR ORGANIZATION, AND YOU DON'T WANT YOUR DATA TO BE CORRUPTED BY PROCESSING. THEREFORE YOU WANT THE DATA TO BE STORED IN RAW FORMAT BEFORE THE PROCESSING IS DONE, WHICH IS A SERVICE OFFERED BY AMAZON REDSHIFT. WHAT IS THE KEY BENEFIT OF STORING DATA IN RAW FORMAT?

ANS: MNIMAL LOSS OF DATA

4]BIGQUERY IS A CLOUD-NATIVE WAREHOUSE THAT IS ALSO A FULLY MANAGED DATA WAREHOUSE. WHAT IS THE MAJOR ADVANTAGE OF BIGQUERY OVER AMAZON REDSHIFT THAT MAY BE A DECIDING FACTOR FOR THE SELECTION OF A DATA WAREHOUSE?

ANS: ACCESS CONTROL ALLOWS IMPROVED DATA SHARING

5]YOU HAVE A VERY WELL-RUN MANAGEMENT PROGRAM IN YOUR ORGANIZATION, WHICH IS VERY SECURE. YOU USE ANALYTICS FOR MAKING BIG DATA DECISIONS. IN RECENT MONTHS, YOU REALIZE THAT WHATEVER DECISIONS YOU ARE MAKING BASED ON THE ANALYTICS ARE PROVIDING TO BE WRONG AND HARMFUL FOR THE ORGANIZATION. AFTER CONSULTING WITH THE DATA ANALYTICS AND DATA STEWARDSYOU CONCLUDE THAT IT IS DUE TO POOR DATA QUALITY. WHAT SHOULD BE THE NEXT STEP OF ACTION?

ANS: INSTALL A FIREWALL

6]YOUR ORGANIZATION HAS AN EFFECTIVE SALES TEAM THAT IS BACKED UP BY ANALYTICS THAT HELP ACCLERATE THE PROCESS OF SALES FROM THE INITIAAL CONTACT. ONE OF THE PRIMARY REASONS FOR ITS EFFECTIVENESS IS A DATA INPUT TOOL THAT HASTENS THE PROCESS OF DATA ENTRY BY PROVIDING PRESET SUGGESTIONS. WHAT ARE THESE SUGGESTIONS COMMONLY KNOWN AS IN DATA SCIENCE TERMINOLOGY?

ANS: REFERENCE DATA

7]YOU HAVE AN AUTOMOBILE COMPANY THAT HELPS SORT OUT VEHICLES AND MAKE MONTHLY SALES VERSUS EXPENDITURE REPORTS. WHAT IS THE BEST WAY TO HANDLE THE DATA FOR CENTRALLY STORING IT?

ANS: BATCH PROCESSING

8]OTHER THAN GAINING REAL-TIME INSIGHTSOF THE DATA, WHAT IS ANOTHER MAJOR ADVANTAGE OF STREAMING ANALYTICS?

ANS: REAL-TIME DASHBOARDS

9]IN A CANDLESTICK CHART, YOU SEE THE SHARE PRICE OF YOUR COMPANY FALLING. YOU HAVE IMPLEMENTED ASTREAMING ANALYTICAL TOOL HELPS IN ANALYZING AND DASHBOARDING THE DATA AS IT IS PRODUCED. OU REALIZE THAT THE CANDLESTICKS PATTERN HAS CONSOLIDATED AND IS NOT RESPONDING WELL TO THE INFLUX OF DATA. WHAT IS THE SOURCE OF THIS PROBLEM?

ANS: SERVER DOWNTIME

10]YOU PLAN TO IMPLEMENT A MODERN DATA WAREHOUSE SOLUTION INTO YOUR ENTERPRISE. YOU HAVE UNDERSTOOD THE PROPER DATA MANAGEMENT AND GOVERNANCE ISSUES. YOU HAVE SET UP ALL YOUR DOMAINS AND DATA INGESTION. NOW YOU PLAN TO MAKE A CENTRAL REPOSITORY OF ALL YOUR FILES. WHAT SHOULD BE THE NEXT STEP FOR THE IMPLEMENTATION OF THE DATA WAREHOUSE SOLUTION?

ANS: SELECTION OF THE NATURE OF THE DATA

MCQS AZURE DATA BRICKS AND DATA PIPELINES:

1]THE REASON WHY AZURE DATABRICKS IS SO EASY TO USE IS BECAUSE IT IS UNIVERSAL AND IS INTEGRATED WITH MICROSOFTS SERVER FOR BETTER PARSING OF INFORMATION. WHICH PLATFORM HAS THE SAME SOURCE OF ORIGIN AS DATABRICKS, WHICH GIVES IT AN ANALYTICAL ADVANTAGEOVER OTHER PLATFORMS?

ANS: APACHE SPARK

2]PLACE THE STEPS FOR TYPICAL AZURE DATABRICKS WAREHOUSE IN THE CORRECT ORDER?

ANS: INGEST, STORE, PREP AND TRAIN, MODEL AND SERVE

3]WHAT IS THE ONE DIFFERENCE THAT SEPARATES THE MODEL OF THE SNOWFLAKE DATA WAREHOUSE FROM ALL THE OTHER DATA WAREHOUSE SOLUTIONS?

ANS: HYBRID MODEL

4]WHAT ARE THE MAJOR DISADVANTAGES OF SNOWFLAKE THAT MIGHT BE TROUBLESOME FOR A FEW COMPANIES THAT SEEK DATA CATEGORIZATION?

ANS: FEWER OPTIONS WITH GEOSPATIAL SPACE, NO OPTION FOR UN-STRUCTURED DATA

5]WHAT IS NOT A DESIGN COMPONENT OF A DATA PPIPELINE?

ANS: DATA INTEGRATION

6]WHAT ARE THE MAJOR ADVANTAGES OF A CLOUD WAREHOUSE SOLUTION OVER AN ON-PREMISES DATA WAREHOUSE SOLUTION?

ANS: LESS WORRY ABOUT STORAGE, LOW COST

7]WHILE SETTING UP AN INTEGRATED DATA PIPELINE FOR YOUR ENTERPRISE TO FACILITATE DATA INGESTION IN THE WAREHOUSE, WHAT SHOULD YOU PLACE MORE EMPHASIS ON FROM A BUSINESS PERSPECTIVE?

ANS: ANALYTICS AND BUSINESS INTELLIGENCE

8]WHAT ARE THE TWO DIFFERENT DATA PIPELINE TOOLS THAT ADDRESS SPECIFIC JOB ROLES?

ANS: DATA ENGINEERS AND ANALYSTS

9]IN YOUR ENTERPRISE, CUSTOMER INFORMATION NEEDS TO BE READILY AVAILABLE TO INDICATE WHETHER THE INFORMATION GIVEN BY CUSTOMERS IS VALID OR NOT, AND TO SEE THE POTENCY OF A POSITIVE DEAL. WHICH FACTORS WOULD YOU CONSIDER A PRIORITY WHEN SELECTING THE APPROPRIATE DATA PIPELINE TOOL?

ANS: BATCH-WISE vs. REAL TIME INGESTION AND ANALYTICS

10]SUPPOSE YOU HAVE A FULL-BLOWN DATA MANAGEMENT PROGRAM WITH A WELL-RUNNING DATA WAREHOUSE AND AN OPTIUM DATA PIPELINE TOOL THAT FACILITATES DATA TRANSFORMATION AND TRANSMISSION. WHILE MEASURING THE MATURITY OF THE DATA PIPELINE TOOL, WHAT WILL BE THE SOLE FACTOR THAT WILL DETERMINE THE EFFICIENCY OF THE DATA PIPELINE TOOL?

ANS: RELIABILITY AND SCALABILITY.

MCQS DATA MAPPING:

1]WHICHOF THESE STATEMENTS ACCURATELY DESCRIBES THE UNIQUE MATCH JOIN MATCH MODEL IN THE MAP EDITOR?

ANS: LAST MATCH IS CONSIDERED AND PASSED TO THE OUTPUT, SET AS DEFAULT WHEN CONFIGURING AN EXPLICIT JOIN

2]YOU ARE BUILDING AN EXPRESSION IN MAP EDITOR FOR A COLUMN IN WHICH YOU WANT TO PAD THE ROW1.CUSTOMERID STRING WITH LEADING ZEROS UP TO MAXIMUM LENGTH OF 6 CHARACTERS. WHICH CODE CAN YOU USE TO ACCOMPLISH THIS FOR YOU?

ANS: STRING.FORMAT("$06d", row1.CustomerID)

3]YOU HAVE MAP EDITOR OPEN FOR A Tmap OBJECT FOR WHICH YOU ARE MAPPING DATABASES OBJECTS. YOU CLICK THE JOIN MODEL FOR A TABLE JOIN PROPERTY. WHICH OPTIONS APPEAR IN THE OPTIONS DIALOG THAT APPEARS?

ANS: INNER JOIN, LEFT OUTER JOIN

4]YOU ARE BUILDING A FILTER EXPRESSION IN MAP EDITOR FOR A COLUMN IN WHICH YOU WANT TO FILTER THE PRODUCT NAME, PRD.NAME TO EQUAL "TURBO WIDGETS", AND YOU WANT THE TRANSACTION QUANTITY, TX.QTY TO BE GREATER THAN 100. WHICH CODE CAN YOU USE TO ACCOMPLISH THIS FOR YOU?

ANS: prd.name.equals ("Turbo Widgets")&&tx.qty>100

5]WHICH COMMON FTP OPERATIONS ARE SUPPORTED BY TALEND OPEN STUDIO COMPONENTS AVAILABLE IN THE PALETTE?

ANS: **FILE EXIST, PUT, DELETE**

6]ON THE COMPONENT TAB DISPLAYED FOR THE Tsortrow COMPONENT, WHEN YOU CLICK TO ADD CRITERIA TO THE CRITERIA TABLE, TALEND AUTOMATICALLY POPULATES THE COLUMN VALUES WITH DEFAULTS. IN THE "SORT NUM OR ALPHA?" COLUMN TALEND HA SCHOOSEN NUM BY DEFAULT FOR CUSTOMER_ID AS DISPLAYED. WHICH OTHER VALUES ARE AVAILABLE FOR THE "SORT NUM OR ALPHA?" COLUMN WHEN YOU CLICK TO OPEN THE DROPDOWN LIST FOR THAT COLUMN?

ANS: **ALPHA, DATE**

7]YOU ARE USING A tExtractDelimitedFields COMPONENT TO SPLIT THE ADDRESS2 FIELD IN THE DELIMITED FILE AS DISPLAYED. WHAT MUST YOU SPECIFY THE FIELD SEPARATOR PROPERLY FOR THE tExtractDelimitedFields COMPONENT TO PROPERLY SPLIT THE ADDRESS2 FIELD?

ANS: **"," (A COMMA)**

8]WHEN CONFGURING THE PROPERTIES FOR A tReplace COMPONENT, YOU CAN OPTIONALLY CLICK THE ADVANCED MODE CHECKBOX. DOING SO ALLOWS YOU TO SPECIFY WHAT TYPE OF EXPRESSION AS THE PATTERN TO SEARCH FOR?

ANS: **REGULAR EXPRESSION**

9]WHEN CONFIGURING THE PROPERTIES FOR A tAggregateRow component. YOU ARE GOING TO GROUP BY THE CUSTOMER_ID FIELD IN ORDER TO AGGREGATE THE SALES ON A CUSTOMER_ID BASIS, SO THAT IN THE RESULTING OUTPUT FILE YOU WILL HAVE ONE ROW FOR EACH CUSTOMER_ID WITH AGGREGATED SALES FIGURES. WHICH FUNCTION VALUE MUST YOU CHOOSE WHEN CONGIGURING THE OPERATIONS TABLE FOR THIS tAggregateRow COMPONENT?

ANS: **SUM**

10]YOU ARE USING tNormalize COMPONENT TO NORMALIZE THE CATEGORY FIELD IN THE DELIMITED FILE AS DISPLAYED. WHAT MUST YOU SPECIFY AS THE ITEM SEPARATOR PROPERTY FOR THE tNormalize COMPONENT TO PROPERLY NORMALIZE THE CATEGORY FIELD?

ANS: **";" (A SEMI- COLON)**

MCQS INSTALLATION AND INTRODUCTION:

1]SELECT THE MAIN DEPENDENCY THAT HAS TO BE INSTALLED FOR TALEND TO BE INSTALLED?

ANS: **JAVA**

2]SELECT THE SUPPORTED Oss BY TALEND OPEN STUDIO?

ANS: **WINDOWS 64, MacOS**

3]SELECT THE MAIN PARTS OF THE DEFAULT UI OF TALEND OPEN STUDIO?

ANS: **PALETTE, REPOSITORY**

4]WHEN INSTALLING MYSQL RELATIONAL DATABASE TO BE USED WITH TALEND, SELECT THE CONFIGURATION TO BE PERFORMED ONCE THE INSTALLATION OF THE COMPONENTS IS COMPLETE?

ANS: MYSQL ROOT PASSWORD, MYSQL SERVER PORT

5]SELECT THE FOLDER THAT CONTAINS ALL THE PROJECT INFORMATION FOR A JOB THAT IS EXPORTED FROM TALEND STUDIO?

ANS: PROCESS

6]RANK THE STEPS REQUIRED TO FOT ANY JOB IN TALEND STUDIO?

ANS: CREATE JOB IN THE REPOSITORY, ADD COMPONENTS FROM PALETTE TO THE DESIGN SPACE, CONFIGURE THE COMPONENTS PROPERTIES, RUN THE JOB

7]SELECT THE CORRECT DESCRIPTION OF TALEND METADATA BRIDGE?

ANS: SYNCHRONIZE METADATA ACROSS DATA PIPELINES

8]MATCH EACH DESCRIPTION WITH THE VARIABLE IN TALEND STUDIO?

 i]USER DEFINED GLOBAL VARIABLES?

ANS: AD-HOC VARIABLES THAT CAN BE CONFIGURED IN JOBS

ii]STUDIO GLOBAL VARIABLES?

ANS: STUDIO PROVISIONS THROUGH COMPONENTS USED IN JOBS INTEGRATION

iii]JOB CONTEXTS VARIABLES?

ANS: EXECUTES JOBS WITH PARAMETERS FOR DIFFERENT ENVIRONMENTS.

MCQS TRANSFORMING DATA:

1]SELECT THE CORRECT DIFFERENCES BETWEEN XML ATTRIBUTES AND ELEMENTS?

ANS: ELEMENTS CAN HAVE MULTIPLE VALUES, ELEMENTS CAN CONTAIN TREE STRUCTURE

2]SELECT THE COMPONENT USED TO GENERATE AN XML FILE FROM A CSV FILE IN TALEND STUDIO?

ANS: tFileOutputXML

3]SELECT THE EXIT CODE VALUE THE SIGNIFIES THE SUCCESSFUL COMPLETION OF A JOB IN TALEND STUDIO?

ANS: 0

4]SELECT THE tMap COMPONENT DESCRIPTION IN TALEND STUDIO?

ANS: CONGREGATE INPUT DATA TO OUTPUT DATA

5]SELECT THE OPTION TO BE ENABLED TO ALLOW THE SPECIFICATION OF 2 SCHEMAS FOR AN XML INPUT FILE IN TALEND STUDIO?

ANS: ENABLE XPATH IN COLUMN "SCHEMA XPATH LOOP" BUT LOSE THE ORDER

6]IN ORDER TO GENERATE A COMPLEX XML FILE WHERE DATA IS SPECIFIED USING ATTRIBUTES OF ELEMENTS AND ELEMENTS TREES IN TALEND STUDIO, WHICH COMPONENT ALLOWS SUCH?

ANS: tAdvancedFileOutputXML

7]SELECT THE COMPONENT USED TO PERFORM LOOKUP DATA IN TALEND STUDIO?

ANS: tJoin

8]SELECT THE COMPONENT ACCESS A MySQL DATABASE IN TALEND STUDIO?

ANS: tMysqlinput

9]SELECT THE TOOL THAT ALLOWS SPECIFYING THE RELATION OF MULTIPLE TABLES AS DATA SOURCES WHEN READING DATA FROM A DATABASE AS INPUT I TALEND STUDIO?

ANS: SQL BUILDER

10]MATCH THE ATTRIBUTE VALUE WITH ITS ATTRIBUTE THAT WILL ONLY AD NEW RECORDS OR MODIFY EXISTING ONES WITHOUT MODIFYING THE TABLE STRUCTURE OR OTHER RECORDS ALREADY EXIST IN THE TABLE WHEN WRITING DATA TO A DATABASE TABLE IN TALEND STUDIO. TWO OPTIONS ARE INVALID?

i]ACTION ON DATA?

ANS: INSERT OR UPDATE

ii]ACTION ON TABLE?

ANS: DEFAULT

11]SELECT THE COMPONENT THAT ALLOWS UPDATING DATA IN A DATABASE IN TALEND STUDIO?

ANS: tMySQLRow

12]SELECT THE COMPONENTS AND CONCEPTS TO FACILITATES ACCEPTING AS INPUT MULTIPLE DATABASES IN TALEND STUDIO?

ANS: tMap, Lookup

13]SELECT THE COMPONENT USED TO COMBINE MULTIPLE DATABASE RECORDS TO A SINGLE RECORDS IN TALEND STUDIO?

ANS: tDenormalize.

MCQS Complex Data Types in Python: Working with Dictionaries & Sets in Python:

1] Consider a dictionary of names and ages set up as below:

Code Editor:

names_ages = {'John': 35, 'Jim': 45, 'Alice': 25}

 How would you look up Alice's age in this dictionary?

Ans: **names_ages['Alice']**

2] Consider a dictionary of names and ages set up as below:

 Code Editor:

names_ages = {'John': 35, 'Jim': 45, 'Alice': 25}

 What would the output be if you were to run this code?

 names_ages['Tim']

Ans: **KeyError: 'Tim'**

3] Consider a dictionary of names and ages set up as below:

 Code Editor:

names_ages = {'John': 35, 'Jim': 45, 'Alice': 25}

 and a second dictionary as below:

 Code Editor:

updated_names_ages = {'Ella': 29, 'John': 36}

 How would you update the names_ages dictionary with the values in updated_names_ages dictionary?

Ans: **names_ages.update(updated_names_ages)**

4] A set in Python can contain which of the following data types?

Ans: **Floats, tuples, Strings**

5] Consider two sets of integers:

 Code Editor:

 set_1 = {2, 4, 6, 8}

set_2 = {1, 2, 5, 6, 7, 8}

 What operation would I run to get a result set with all of the elements from both sets?

Ans: set_1.union(set_2)

6] Consider a nested list of names and ages:

 Code Editor:

names_ages = [['John', 35], ['Jill', 38], ['Tim', 27]]

 How would you access Tim's age in this nested list?

Ans: names_ages[2][1]

7] Consider a nested list of names and ages:

 Code Editor:

names_ages = [['John', 35], ['Jill', 38], ['Tim', 27]]

 How would you convert this to a dictionary with names as the keys and ages as values?

Ans: dict(names_ages)

MCQS Complex Data Types in Python: Working with Lists & Tuples in Python:

1] Which of the following statements about Python lists are true?

Ans: Lists in python can have elements of different data types, Lists are ordered collections

2] If you wanted to insert an element at index 2 in a particular list named names_list what is the function that you would invoke?

Ans: names_list.insert(2,"john")

3] If you want to count the number of times the name "John" appears in the names_list what function would you invoke?

Ans: names_list.count("John")

4] If you wanted to sort the elements in the list names_list in alphabetical order which of the following statements in Python are valid?

Ans: names_list = sorted(names_list), names_list.sort()

5] Consider the list:

Code Editor:

some_list = ['a', 'b', 'c', 'd', 'e', 'f']

How do you slice this list to access the elements 'c', 'd'?

Ans: some_lst[2:4]

6] What will be the result of this slicing operation of the names_list?

 Code Editor:

names_list = ['John', 'James', 'Lily', 'Emily', 'Nina']

names_list[::2]

Ans: ['John','Lily','Nina']

7] What is the result of executing this code?

 Code Editor:

some_string = "Python"

a, b, c, d = some_string

Ans: **This is an error, "too many values to unpack"**

8] What is the result of executing this code?

 Code Editor:

city = 'Los Angeles'

city.find('x')

Ans: **-1**

9] Which of the following lines of code will print this string in reverse i.e. print out "olleH"

 Code Editor:

some_str = "Hello"

Ans: **some_str[::-1]**

10] All of the following statements are ways in which lists and tuples are similar. Which one of these is NOT true?

Ans: **Both once created cannot be updated**

11] All of the following statements are ways in which lists and tuples are different. Which one of these is true?

Ans: **A list can be changed once creating, a tupe is immutable and cannot be changed**

12] Which of the following are valid complex data types in Python?

Ans: **Lists, Dictionaries, sets**

MCQS Conditional Statements & Loops: If-else Control Structures in Python:

1] Consider three variables with values as shown:
a = 5
b = 10
c = "five"

What are the results of evaluating the conditional expression?
a == c
a >= b
not(a < b and a > b)

Ans: false, false, true

2] How is the body of an if-statement block syntactically represented in Python?

Ans: Using additional indentation from the left relative to the lines just before and after the block

3] What is the output for this code?

```python
if 'bin' in {'float': 1.2, 'bin': 0b010}:

    print('a')

    print('b')
print('c')
```

Ans: a b c

4] What is the output for this code?

```python
if None:

    print('Hi')
```

Ans: Nothing is printed – no output

5] Evaluate the expression provided. What does the following expression evaluate to?

'1' + '2' if '123'.isdigit() else '2' + '3'

Ans: **'12'**

6] What is the output of the code?

```python
a = [1, 'one', {2: 'two'}, 3]

b = len(a)

if b == 4:

  print('Length of this list is 4')

  if b == 5:

    print('Length of this list is 5')

  else:

    print(b)
```

Ans: **Length of this list is 4 4**

7] What is the value of b in the snippet of python code?

```python
a = "six"

b = (int(a), float(a))
```

Ans: **ValueError:invalid literal for int() with base 10: 'six'**

8] Consider the following snippet of Python code:

```python
a = "40.6 "

b = "60.4 "

c = a + b
```

What does c evaluate to?

Ans: **'40.6 60.4'**

9] What would the output of the following code snippet be?

```python
num_one = 76
```

num_two = 23.4

print("datatype of num_one:", type(num_one))

print("datatype of num_two:", type(num_two))

Ans: **datatype of num_one:<class 'int'>datatype of num_two:<class 'float'>**

10] What is the output of the code snippet below?

value = 4

a = str(value)

b = a + "^" + "2"

c = a + "^" + "3"

print(value, "+", b, "+", c)

Ans: **4 + 4 ^ 2 + 4 ^ 3**

11] What do the values of d[0], d[1], d[2], d[3] evaluate to after the execution of the Python code below?

new_list = ["Red", "Blue", "White", "Green"]

z = sorted(new_list)

d = list(z)

d[0], d[1], d[2], d[3] = d[3], d[2], d[1], d[0]

Ans: **"White","red","green","blue"**

12] What is the output of the program below?

var = "hi"

if(type(var) == int):

   print("Type of the variable is Integer")

elif(type(var) == float):

```
    print("Type of the variable is Float")

elif(type(var) == complex):

    print("Type of the variable is Complex")

else:

    print("Type of the variable is Unknown")
```

Ans: **Unknown**

13] What is the output of the program below?

```
total_classes = 100

attended_classes = 67

attendance = (attended_classes/total_classes)*100

if attendance >= 75:

    print ("You are eligible to appear for the test.")

else:

    print ("Sorry, you are ineligible to appear for the test.")
```

Ans: **sorry you are ineligible to appear for the test.**

MCQS Conditional Statements & Loops: The Basics of for Loops in Python:

1] Which of these Python data types can NOT be iterated through using for loops?

Ans: **int**

2] Given a variable my_dict which is a dictionary, consider you use it in a for loop in this manner:

```
for x in my_dict:

    print (x)
```

What are the contents printed out?

Ans: **The keys in the dictionary**

3] Which TWO of the following statements about for loops in Python are TRUE?

Ans: **They may have an associated else block, They can iterate over the elements in tuples, lists, and dictionaries**

4] Given the following code, what is the type of x which is printed out in each iteration?

my_list = [['tiger', 'lion', 'leopard'], ['camel', 'llama', 'alpaca'], ['zebra', 'donkey', 'wildebeest']]

for x in my_list:

   print(x)

Ans: **A list of strings**

5] What is the maximum value in the sequence x?

x = range(2, 14)

Ans: **13**

6] What is the correct value of x given the assignment shown?

x = list(range(-17, -7, 2))

Ans: **[-17, -15, -13, -11, -9]**

7] Which of the following function calls will generate the list below?
[10, 7, 4, 1, -2]

Ans: **list(range(10, -3, -3)).**

MCQS GET STARTED WITH PYTHON INTRODUCTION:

1] Which of the following is an open source distribution of the Python and R programming languages that uses the Conda package manager?

Ans: **Anaconda**

2] Which of the following terms best describes Jupyter notebooks that you can use to write Python code?

Ans: **Interactive, Can view results on the same screen, Browser-based**

3] As a new user of Python which version of Python should you use?

Ans: **Either version of python is fine but 3.7 should be preferred**

4] How can you execute shell commands on Jupyter notebook code cells?

Ans: **Prefix the shell command using ! i.e. !python-version**

5] What is the output of the following bit of code?

Code Editor:

```
 13 // 5?
```

Ans: **2**

6] Which of the following functions are valid built-in functions in Python?

Ans: **len(), print(), type()**

7] Which of the following commands are valid to store a numeric value of 2 in a variable named num_x?

Ans: **num_x = 2, num_x = 6-4**

8] Consider two Python variables initialized as shown below. Which of the following logical statements below will have a value of True?

Code Editor:

```
a = True
```

```
b = True
```

Ans: **a or b , a and b**

9] Consider this bit of Python code:

Code Editor:

num_1 = 10

num_2 = 20

num_3 = num_1

num_1 = 100

What is the final value stored in num_3?

Ans: **10**

10] If you want to increment the value stored in the num_1 variable by 10 which of the following Python statements are valid?

Ans: **num_1 = num_1 + 10, num_1 += 10**

11] Which of the following are valid data types in Python?

Ans: **float, int, bool**

12] What is the correct syntax for specifying multi-line strings in Python?

Ans: **"""This is a multi-line string"""**

13] Which of the following are valid string operations in Python?

Ans: **"Hello" * 3, "Hello" + "World"**

MCQS Functions in Python: Introduction

1] Which of the following statements about functions is true?

Ans: A function is defined using the "def" keyword, Function code is not executed when defined

2] Which of the following are valid function names in Python?

Ans: _function_name(), functionName(), _123_Function_Name()

3] Which of the following statements about functions is false?

Ans: functions cannot access variables which are declared outside the function

4] Consider a function definition which looks like this:

```
def some_function(a, b, c):

    print(a, b, c)
```

Which of the following function invocations are correct?

Ans: some_fuction(2,3,4), some_function(2,3,"Hello")

5] Consider the following bit of code. What will be the result of executing this bit of code?

```
x = 3

y = 4

def add(a, b):

  result = x + y

  print(result)

add(10, 20)
```

Ans: 7

6] Which of the following statement(s) about positional arguments to functions is/are true?

Ans: A function can accept any number of positional arguments, They can be of any data type- primitive or complex types

7] What is the default return value from a function when no return statement is specified?

Ans: **None**

8] Which of the following statement(s) about return values is/are false?

Ans: **A function can have just one return statement, A function has to have a return statement, A function with input arguments cannot have a return statement**

9] Which of the following statements(s) about the data types of return values is/are false?

Ans: **A return statement is mandatory in functions**

10] Which of the following are valid kinds of input arguments for Python functions?

Ans: **Positional arguments, Keyword arguments**

11] Which of the following are some of the advantages of using keyword arguments to invoke functions?

Ans: **Easier to maintain code since the value of each argument is clearly seen during invocation, Keyword arguments can be specified out of order**

12] What does this function definition indicate?

```python
def some_fn(a, b, c=True):

    print(a, b, c)
```

Ans: **a and b are required arguments, c is required, a and b are numeric arguments, c is boolean**

13] Which of the following function definitions allows the function to accept variable length arguments?

Ans: **some_fn(*args)**

14] Which of the following are valid types of arguments in Python?

Ans: Variable length arguments, Default arguments, Keyword arguments

15] Which of the following statement(s) about variable length arguments in Python is/are true?

Ans: Variable length positional arguments are passed into the function as a tuple, Variable length keyword arguments are passed into the function as a dictionary

MCQS Python - Introduction to NumPy for Multi-dimensional Data:

1] Which of these correctly match the following libraries in the Numpy ecosystem with what that library is used for?

i] Contains a collection of algorithms used for image processing?

Ans: Scikit-image

ii] Used to perform statistical operations?

Ans: Statsmodel

iii] Data visualization tool used for plotting 2-D graphs?

Ans: Matplotlib

2] Let's say you have imported the numpy package as np and you want to assign the variable "x" with a 3 by 2 array of type integer, all of whose values are 1. Which of these commands will you use to do so?

Ans: x=np.ones((3,2),dtype=np.int32)

3] What will be the value stored in the variable y after we have executed the following code

import numpy np

y=np.arange(2,4,0.5)

Ans: [2, 2.5, 3, 3.5]

4] Let's say you have imported the numpy package as np and you want to print the first 2000 natural numbers in the form of an array and you want all 2000 of the numbers to be visible on screen when printing (including the number 2000). Which of these commands would you use to do so?

Ans: np.set_printoptions(threshold=np.nan) print(np.arange(1,2001))

5] What would be the output of the following code:

import numpy as np

x=np.array([[1,2] , [3,4]])

y=np.array([[5,6],[7,8]])

z=(x*y)

z

Ans: [[5,12], [21,32]]

6] What would be the output of the following section of code:

Import numpy as np

x=np.array([4,6,2,8])

np.median(x)

Ans: 5

7] Which of these slicing operations can be used to quickly get the reversed contents of a numpy array called "array"?

Ans: array[::-1]

8] Match the following features of the numpy nditer function mentioned here with the correct Boolean category

i] False

Ans: By default, the nditer object returns arrays that can be written on, The nditer can accept only one dimensional arrays as input

ii]True

Ans: Using this function, we can iterate through each of the individual elements of the array passed as an input argument

9] Which of these statements regarding the ravel() object in Python are true?

Ans: The ravel function reduces a multi-dimensional array to a single dimensional array, ravel() belongs to the numpy library and can be used on any object that can be parsed.

MCQS Python - Advanced Operations with NumPy Arrays:

1] Let's say you have a two dimensional numpy array called "twod" and you want to split it row-wise into two equal halves. Then, which of these numpy functions would you call on it to do so?

Ans: vsplit(twod,2)

2] Some of the features of digital images in Numpy are given below. Which of these are true?

Ans: In numpy, images can be represented as a 3D matrix where the first two dimensions represent the pixels in the image that are arranged in the form of a grid and the third dimension specifies the number of channels for the image, A digital image is a multidimensional array and every pixel in a digital image is represented by a number

3] Let's say you have an image that you have split into two equal halves along the x axis. You have stored these two halves of the original image in the variables x1 and x2 respectively. Which numpy function would you use to combine these two halves to reconstruct the original image?

Ans: concatenate((x1,x2),axis=1)

4] Let's say you have a numpy array called "array_1" and you initialize "another array called "array_2" with the help of a following command:

array_2 = array_1.view()

Match the following statements about "array_2" with the correct Boolean value

i]True

Ans: The base for array_2 points to the same object as array_1, array_1 and array_2 contain the same elements

ii]False

Ans: array_2 points to the same object as array_1, If we re-assign array_2, then we will end up re-assigning array_1 as well and change its contents

5] Let's say you have a numpy array called "array_3" and you initialize "array_4" with the help of a following command:

array_4 = array_3.copy()

Match the following statements about "array_4" with the correct Boolean value

i]False

Ans: If we re-assign array_4, then we will end up re-assigning array_3 as well and change its contents, If we change a single element of array_4, then the corresponding element in array_3 changes too, Changing the shape of array_4 will change the shape of array_3 as well

ii]True

Ans: array_3 and array_4 contain the same elements

6] Let's say you have a 1-D numpy array called "cubes" consisting of the cubes of the numbers 1,2,3 and so on till 10. What would be the value of the array:

cubes [[[4, 5], [1, 2]]]

Ans: [[125, 216] , [8, 27]]

7] Some of the features of Pandas is given below. Which of these are true?

Ans: A particular column of a pandas dataframe can be referenced by its column header, The column header of a Pandas dataframe can be treated in the same way as the index label of a numpy array

8] Let's say you imported numpy as np and you have initialized a 1-D array of integers called "array". What would np.all (x < 50) return?

Ans: This function would return a true boolean value if all the entries in your array are less than 50 and false otherwise

9] Let's say you have a Pandas dataframe called "phone_data" which contains the data of various phones released in 2018 and their prices. It has the following three columns:

"manufacturer", "phone name" and " price".

You want only the names of all the phones that are priced more than 10,000. Which of these commands can be used to print these values?

Ans: phone_data[phone_data['price'] > 10000]['phone name']

10] What are the conditions under which broadcasting can take place between two elements in Numpy?

Ans: A smaller array can be broadcast on a larger array only when the corresponding dimensions of the two arrays being operated upon are compatible i.e. when the corresponding dimensions are equal or one of the two dimensions is 1, Broadcasting works when at least one of the elements is a scalar

11] Match the following statements about broadcasting with the correct Boolean value:

i]False

Ans: The array [[1, 2] , [3, 4]] and the scalar 10 are incompatible with broadcasting

ii]True

Ans: The scalar 10 and the scalar 20 are compatible with broadcasting, The array [[1, 2] , [3, 4]] and the array [1, 2 ,3] are incompatible with broadcasting, The array [[1, 2] , [3, 4]] and the array [[1], [2]] are compatible with broadcasting.

MCQS Python - Introduction to Pandas and DataFrames:

1] Which of these correctly match the following libraries in the Numpy ecosystem with what that library is used for?

i] Used to perform statistical operations

Ans: Statsmodel

ii] Data visualization tool used for large datasets

Ans: Bokeh
iii] Specifically meant for machine learning, data mining, and data analysis

Ans: Scikit-learn

2] Which of these statements related to the Pandas Series object are true?

Ans: Pandas Series object is similar to a Python list, Once we create a Pandas Series object, an index representing the positions for each of the data points is automatically created for the list

3] In the following Python code, typing which Python command will give the user the CEO of Facebook?

import pandas as pd

companies_ceo = {

'Amazon' : 'Jeff Bezos'

'Apple' : 'Tim Cook',

'SpaceX': 'Elon Musk'

'Facebook': 'Mark Zuckerberg'

'Netflix': 'Reed Hastings'

}

companies_ceo_series= pd.Series(companies_ceo)

Ans: companies_ceo_series[3], companies_ceo_series['Facebook']

4] Match the following statements related to Pandas DataFrames with the correct boolean values.

i]True

Ans: All the data within a particular column in a Pandas DataFrame must be of the same data type

ii]False

Ans: Once a Pandas DataFrame has been created, it is not possible to add a new column to this DataFrame, Data in different columns of a Pandas DataFrame cannot be of different data types

5] In the following Python code, typing what command will create a DataFrame called "companies_ceo" whose first column has all the entries of the 'companies' list and whose second column has all the entries of the 'ceo' list, with the column names as the names of the respective variables?

import pandas as pd

companies = {

'Amazon'

'Apple'

'SpaceX'

'Facebook'

'Netflix'

        }

ceo = {

'Jeff Bezos'

'Tim Cook',

'Elon Musk'

'Mark Zuckerberg'

'Reed Hastings'

  }

Ans: companies_ceo_tuple = list (zip(companies, ceo)) companies_ceo = pd.dataframe(companies_ceo_tuple, columns=['companies', 'ceo'])

6] Let's say you have a pandas DataFrame called "frame" and you want to export this DataFrame along with its index as a CSV file called "data_frame" located in the datasets folder of our workspace.

Ans: frame.to_csv('datasets/data_frame.csv')

7] Let's say you have a pandas DataFrame called "panda" which has 8 rows in total and you want to remove the last row from this DataFrame. Which of these Python commands would you use to do so?

Ans: panda.drop(panda.index[7])

8] Let's say you have saved a dataset in a pandas DataFrame called "dataset" which has tons of records and you only want to access the details of the records in only the 5th, 8th and 14th index. Which of these Python commands can you use to do so?

Ans: dataset.loc[5,8,14], dataset.loc[[5,8,14],:]

9] Match the following statements related to the iloc indexer in Pandas with the correct boolean values.

i] False

Ans: The column headers can be passed as input arguments in the form of a string to the iloc function without any errors, When we pass 2:6 as input argument to the iloc function, we get all details of the records located in the second index all the way up to the 5th index of the DataFrame

ii]True

Ans: The iloc indexer is similar to the loc indexer and can be used to access records located at a particular index in a Pandas DataFrame

10] Match the following statements related to the concept of multiIndex in Pandas with the correct Boolean values.

i]True

Ans: MultiIndex is useful when we have large datasets where using numeric indexes to refer to each record is unintuitive, MultiIndex lets the user effectively store and manipulate higher dimensional data in a 2-dimensional tabular structure

ii]False

Ans: The MultiIndex for a row is some composite key made up of exactly one column

11] Which of these statements related to the pivot function in Pandas is true?

Ans: The combination of the row index and the column header must be unique in order to generate a pivot table, The Pivot function summarizes the details of each column in a DataFrame

12] What happens when we call the stack () function on a Pandas DataFrame?

Ans: It will create a new DataFrame such that a single row in the original DataFrame is stacked into multiple rows in the new DataFrame depending on the number of columns for each row in the original DataFrame

MCQS Python - Manipulating & Analyzing Data in Pandas DataFrames:

1] Consider the following Python code. What command would you use to iterate through the "companies_ceo" DataFrame and print the list of all the CEOs in this DataFrame?

import pandas as pd

companies = {

   'Company' : ['Facebook', 'Apple', 'Amazon', 'Netflix'],

       'CEO' : ['Mark Zuckerberg', 'Jeff Bezos', 'Tim Cook', ','Reed Hastings'],

       }

companies_ceo = pd.DataFrame(companies)

Ans: for row in companies_ceo.iterrows(): print(row[1]), for row in companies_ceo.itertuples(): print(row.CEO

2] Which of the following formats does Pandas not support natively when exporting the contents of a Dataframe?

Ans: JPEG

3] Let's say you have created a Pandas DataFrame called "unsorted" and you want to sort the contents of this DataFrame column wise in alphabetical order of the header name. Then, which function would you call on the "unsorted" DataFrame to do so?

Ans: unsorted.sort_index(axis=1)

4] Match the following functions that you can call on a Pandas DataFrame correctly with what they do

i] All the rows which contain a NaN value in any cell of that row are removed

Ans: **.dropna()**

ii] Every cell in the Dataset which has a NaN value will be replaced with 0

Ans: **.fillna(0)**

iii] Returns a Boolean array containing true or false values and returns the value in a cell as true if it contains NaN

Ans: **.isnull()**

iv] Returns a Boolean array containing true or false values and returns the value in a cell as true if it does not contain NaN

Ans: **.notnull()**

5] Match the following statements related to the .xs function in Pandas DataFrame with their correct Boolean values.?

i] False

Ans: **The .xs function cannot be used to return a cross section of columns**

ii]True

Ans: **By default, the .xs function only takes a look at values in the first level index, The .xs function is used when our Pandas DataFrame makes use of a MultiIndex**

6] Let's say you have imported Python as pd and have instantiated two DataFrames called "frame_1" and "frame_2" with the exact same schema. What command will you use to combine these two DataFrames into a single DataFrame and make sure that the combined DataFrame has its own unique index?

Ans: **pd.concat([frame_2, frame_1], ignore_index = True), pd.concat([frame_1, frame_2], ignore_index = True)**

7] The 'how' argument in the Pandas merge function allows us to specify what kind of join operation we want to perform on the given Pandas DataFrames. What are the valid values that we can give for this argument?

Ans: **left, inner, outer, right**

8] Some statements related to working with SQL Databases in Python are given below. Match them with their correct Boolean values.

i] False

Ans: **Once we have created a table, we can use sqlite3's .execute() function to recreate the same table with the same table name so that we have duplicates of a table**

ii]True

Ans: **The sqlite3 library in Python allows us to create Databases on our local file system, All the changes that we make to an SQL database on a Jupyter notebook by connecting with it, will be committed to the database only after we execute sqlite3's .commit() function**

MCQS Big Data Concepts: Big Data Essentials:

1] Which statements are true about unstructured data?
Ans: **Web pages, video files, and audio files are examples of unstructured data, Unstructured data is very often linked to structured data. An example is how X-ray images at a hospital are linked to patient IDs or health card numbers**

2] Which statements accurately describe the differences between big data and data warehousing?

Ans: **Data warehouses only handle structured data (relational or non-relational), whereas big data can handle structured, un-structured, or semi-structured data, While only DBMS compatible data are stored in data warehouses, all kinds of data including transactional data, social media data (including audio and video), machinery data, or any DBMS data can be stored and managed using big data technologies**

3] Which statement about parallel or distributed computing is true?

Ans: **Distributed computing can allow an application on one machine to leverage processing power, memory, or storage on another machine**

4] Which statement about horizontal and vertical scaling is true?

Ans: **Horizontal scaling is typically the easiest scaling option**

5] Match each Hadoop component with its respective layer in the Hadoop ecosystem. One layer will not be used.

i]Hive

Ans: **Data access layer**

ii] MapReduce

Ans: **Data processing layer**

iii] HDFS

Ans: **Data storage layer**

iv] ZooKeeper

Ans: **Data management layer**

6] Which statements are correct about HDFS?

Ans: **HDFS provides high throughput access to application data by providing the data access in parallel, HDFS provides a fault-tolerant storage layer for Hadoop and its other components**

7] What are the benefits of migrating from Hadoop to the cloud?

Ans: **Better scalability, Long-term cost savings, Better collaboration, Easy access and resource availability**

8] What are the differences between Hadoop and cloud computing?

Ans: **Cloud computing focuses on on-demand, scalable, and adaptable service models, while Hadoop is all about extracting value out of volume, variety, and velocity, Cloud computing constitutes various computing concepts. This naturally involves a large number of computers that are usually connected through a real-time communication network. Hadoop, on the other hand, is a framework that uses simple programming models to process large data sets across clusters of computers**

9] What are the most important features of HDFS?

Ans: **Scalability, Replication, Distributed storage, High availability**

10] Which statement is true about in-memory storage systems?

Ans: **Data storage in an in-memory database is reliant on random access memory (RAM)**

MCQS Big Data Concepts: Getting to Know Big Data:

1] What could be accomplished by using big data technologies?

Ans: **New product development and optimized offerings, Cost reductions**

2] What are some examples of big data sources?

Ans: **Email, Social media, Open data, Sensor data**

3] Four of the seven characteristics of big data are listed. Match each characteristic with its description. One description will not be used?

i] Volume

Ans: **The amount of data that exists**

ii] Velocity

Ans: **The speed at which data is processed and becomes accessible**

iii] Veracity

Ans: **Making sure the data is accurate, which requires processes to keep bad data from accumulating in your systems**

iv] Variety

Ans: **The different types of data from XML to video to SMS**

4] Which statements are true about unstructured data?

Ans: **Unstructured data is information that does not have a predefined data model, Common examples of unstructured data include audio, video files, or No-SQL databases**

5] What are the main deliverables of big data?

Ans: **Text/image analytic, Multivariate analysis, Predictive models**

6] What are the most important advantages of big data, according to the International Institute for Analytics (IIA)?

Ans: **Big data leads to cost reductions, Big data enables faster, better decision making, Big data helps to identify what customers need and to introduce new products and services accordingly**

7] What are some of the main business domains that use big data tools today?

Ans: **Aviation industry, Credit scoring agencies, E-commerce industry, Transportation industry**

8] Which statements are correct about how Netflix utilizes big data?

Ans: **Netflix has screenshots of scenes people might have viewed repeatedly, the associated ratings, and the number of searches and the search topics,**

Netflix uses what is known as the big data recommendation algorithm to suggest TV shows and movies based on a user's preferences

9] What has Amazon been able to achieve by utilizing big data?

Ans: **Gathering of information on what each customer is likely to purchase based on what other people with similar interests have purchased, Gathering of data on the search patterns of its customers**

10] What are the main challenges that companies experience with big data?

Ans: **Unfamiliarity with big data and confusing it with traditional methods, Integrating data from a variety of sources, Data security issues, Unprecedented data growth.**

MCQS Spark for High-speed Big Data Analytics:

1] What are some components of Apache Spark?

Ans: **GraphX, Spark SQL**

2] Which statements are true about resilient distributed datasets (RDDs) and directed acyclic graphs (DAGs)?

Ans: **RDD is an immutable (read-only), fundamental collection of elements or items that can be operated on many devices at the same time (parallel processing), Compared to MapReduce that creates a graph in two stages, Apache Spark can create DAGs that contain many stages**

3] As Spark usage grew at Uber, users encountered an increasing number of issues. What were some of those issues/challenges?

Ans: **Multiple Spark versions, Multiple compute clusters**

4] What are some examples of metrics that Alibaba measures by utilizing Spark?

Ans: **Connected components, Degree distribution**

5] What are some advantages that Spark provides to modern healthcare providers?

Ans: Behind the scenes distributed execution, A user-friendly API, Convenient workflow fulfillment

6] Which statement is correct about how Spark and Hadoop are different?

Ans: The Hadoop MapReduce model provides a batch engine, hence it is dependent on different engines for other requirements, whereas Spark performs batch, interactive, machine learning and streaming all in the same cluster

7] What are some predominant industries that use Spark today?

Ans: Media and entertainment industry, Finance industry

8] What are some characteristics of Spark that help improve performance?

Ans: Lazy loading behavior, Cache appropriately

9] What are the three API types that are compatible with Spark?

Ans: RDD, DataFrame, DataSet

10] What are some of the most important best practices when it comes to using Apache Spark?

Ans: Proper tuning, Using the right level of parallelism, Joining a large and a medium size RDD.

Techniques for Big Data Analytics:

1] What are the biggest challenges associated with traditional data analytics?

Ans: Scalability, consistency, reliability, efficiency, and maintainability

2] Place the layers of big data analytics architecture in the correct order from the bottom to the top.

Ans: Data monitoring, Data security, Data storage, Data processing, Data query, Data visualization

3] What are the parameters of data ingestion?

Ans: Data size, Data format, Data frequency, Data velocity

4] What are some ways in which big data processing can be performed?

Ans: Stream processing, Batch processing

5] Which statement about data storage systems is correct?

Ans: The Hadoop distributed file system (HDFS) is the primary data storage system used by Hadoop applications

6] Which are the main components of the big data architecture?

Ans: The data model, Big data security, Big data analytics

7] Which are the main reasons for using batch processing?

Ans: To join tables in relational databases, To run complex algorithms on large datasets which require access to the entire batch

8] Which is correct about stream processing?

Ans: Stream processing provides analytical insights before the data storage stage

9] Which statement is true about the Lambda architecture?

Ans: Data that enters the system is dispatched to two layers in the Lambda architecture: the batch layer and the speed layer, The Lambda architecture provides fault-tolerance against possible hardware failures and human errors

10] Which statement is true about the Kappa architecture?

Ans: The Kappa architecture uses stream processing to manage data flows through a single path

MCQS Data Silos, Lakes, & Streams Introduction:

1] Which of the following is a characteristic of a data silo?

Ans: **Data may be in a raw, native format and not useful unless processed, Data is not easily accessible using common tools, Data is stored in isolation and cannot be combined with other sources**

2] Which of the following are valid data types that can be stored in a data lake?

Ans: **Unstructured data, Structured data, Semi-structured data**

3] Which of the following is not a characteristic of a data lake?

Ans: **Data is not searchable easily**

4] Which of the following are challenges involved in designing and building data lakes?

Ans: **Data lakes need to work with different data types and sparse and incomplete data, Data lakes need to be able to support a huge volume of data, Data lakes need to maintain data security and compliance**

5] Which of the following are valid differences between a traditional relational database and a data warehouse?

Ans: **A database supports ACID properties and a data warehouse does not, A data warehouse is optimized for read access, a database is optimized for read as well as write access**

6] Which of the following statements about data lakes and data warehouses are true?

Ans: **Data warehouses hold fairly structured data optimized for analysis, Data lakes need to maintain security and ensure compliance of the data stored within it, Data lakes promote shared data stewardship**

7] Which of the following is not an example of a data stream?

Ans: **Census data stored in a database**

8] Which of the following is not a valid service used to ingest data into the AWS cloud?

Ans: **Amazon Athena**

9] Which of the following correctly defines AWS Glue?

Ans: **A single catalog which indexes data from multiple sources to make it searchable**

10] Which of the following AWS services can be used to visualize data stored in a data lake on AWS?

Ans: **Amazon QuickSight**

MCQS Data Engineering Getting Started:

1] Select the benefits of a distributed system?

Ans: **fault tolerance, concurrency, scalability**

2] Arrange the following ETL processing steps in order from the top?

Ans: **ingest data from source, message brokering, streaming data engine, long-term storage and analytics**

3] Select the characteristics of a NoSQL data store?

Ans: **horizontal scaling, cluster-friendly, dynamic schema**

4] Match the data management category with its description?

i] data warehousing, transformation, extraction

Ans: **ETL**

ii] standardized data, static information

Ans: **Reference Data Management**

iii] dashboards and real-time results

Ans: **Visualization and Analytics**

iv] organizational data

Ans: **Master Data Management**

5] Match the ETL process with its description?

i] format and representation shift

Ans: **Transform**

ii] importing data for computation

Ans: **Load**

iii] selecting raw data

Ans: **Extract**

6] Where does the library of job components reside in the Talend Open Studio UI?

Ans: **Palette**

7] What high level model is used to get a project overview for ETL jobs in Talend Open Studio?

Ans: **Business Model**

8] Put the following AI hierarchy steps in pyramid order from the bottom up?

Ans: **ETL, Data Exploration, Aggregation, Machine Learning, Deep Learning**

9] Reducing the number of fields in the output is an example of what type of partitioning?

Ans: **column-based**

10] Match the data storage model approach with its descriptions.

i] Normalization

Ans: Less Redundancy, Standardized

ii] Star Schema

Ans: Fact Tables, Dimension Tables

11] Select the features common to interactive reporting tools.

Ans: sorting, filtering, drilling down

END