Azure Data Platform 203

Question 1:

Scenario: Queen Consolidated was overtaken by Raymond Carson Palmer and re-branded as Palmer Technologies. Now that Ray is overseeing the operations at Palmer, Ray has decided to implement better applications.

You are working as a consultant with Palmer, and in a meeting with Ray and his IT team discussing Azure Data Factory. The team plans to use Azure Data Factory to prepare data to be queried by Azure Synapse Analytics serverless SQL pools.

Files will be initially ingested into an Azure Data Lake Storage Gen2 account as 10 small JSON files. Each file will contain the same data attributes and data from a subsidiary of Palmer. The team needs to move the files to a different folder and transform the data.

Required:

- Provide the fastest possible query times.

- Automatically infer the schema from the underlying files.

As the Azure expert, the team looks to you for advice on how they should configure the Data Factory copy activity with respect to the sink file type.

Which of the following should you advise them to use?

    A. **Parquet**
    B. **TXT**
    C. **CSV**
    D. **JSON**

Question 2:

Scenario: Stark Industries was founded by Howard Stark during the early twentieth century, a great pioneer in different types of technology and constantly helping the United States Armed Forces with different and innovative weapons. Some years after the death of Howard Stark, his son Tony took over the company.

Currently in the process of making many IT improvements at Stark Industries, you have been hired as a consultant to advise on several Microsoft Azure projects.

At the moment, the team is designing an Azure Stream Analytics job to process incoming events from sensors in retail environments.

Required:

- Process the events to produce a running average of shopper counts during the previous 15 minutes, calculated at five-minute intervals.

As the Azure expert, Tony expects you to advise the team on the best course of action.

Which type of window should you advise the team to use?

    A. **Sliding**
    B. **Hopping**

C. Snapshot
D. Tumbling

Question 3:

Identify the missing word(s) in the following sentence within the context of Microsoft Azure.

From a high level, the Azure Databricks service launches and manages Apache Spark clusters within your Azure subscription. Apache Spark clusters are groups of computers that are treated as a single computer and handle the execution of commands issued from notebooks.

In Databricks, the notebook interface … [?]

A. **Provides the fastest virtualized network infrastructure in the cloud.**
B. **Is the driver program.**
C. **Specifies the types and sizes of the virtual machines.**
D. **Pulls data from a specified data source.**

Question 4:

Within creating a notebook, you need to specify the pool that needs to be attached to the notebook that is, a SQL or Spark pool.

In order to bring data to a notebook, you have several options.

It is possible to load data from which of the following?

A. **Azure Blob Storage**
B. **Azure Data Factory**
C. **Azure Cosmos DB**
D. **Azure Data Lake Store Gen 2**
E. **SQL Pool**
F. **Azure File Storage**

Question 5:

Scenario: You are working on a project and are preparing to ingest data from a SQL Server database hosted on an on-premises Windows Server.

Which integration runtime is required for Azure Data Factory to ingest data from the on-premises server?

A. **Self-Hosted integration runtime**
B. **Azure-SSIS Integration Runtime**
C. **Azure Integration Runtime**
D. **On-demand HDInsight cluster**
E. **None of the listed options.**

Question 6:

There are some core concepts with which data engineers should be familiar. The data engineer will often work with multiple types of data to perform many operations using many scripting or coding languages that are appropriate to their individual organization.

Data in a relational database table is considered which of the following?

A. Unstructured
B. None of the provided options are correct
C. NoSQL
D. Semi-Strutured

Question 7:

Scenario: You are working in the sales department of a company and part of your role is to manage the storage of customer profile and sales data. A common request is to generate a list of *"the top 100 customers including name, account number and sales figures for a given time period"* or *"who are the customers within a given geographic region?"*

Is Azure Blob storage a good choice for this data?

A. Yes

B. No

Question 8:

Scenario: You are working on a project and have begun creating an Azure Data Lake Storage Gen2 account. Configuration of this account must allow for processing analytical data workloads for best performance.

Which option should you configure when creating the storage account?

A. **On the Basic tab, set the Performance option to Standard.**

B. **On the Basic Tab, set the Performance option to ON.**

C. **On the Networking tab, set the Hierarchical Namespace to ON.**

D. **On the Advanced tab, set the Hierarchical Namespace to Enabled.**

Question 9:

Linux foundation *Delta Lake* is an open-source storage layer for Spark that enables relational database capabilities for batch and streaming data. By using Delta Lake, you can implement a *data lakehouse* architecture in Spark to support SQL_based data manipulation semantics with support for transactions and schema enforcement. The result is an analytical data store that offers many of the advantages of a relational database system with the flexibility of data file storage in a data lake.

The benefits of using Delta Lake in Azure Databricks include which of the following? (Select four)

A. **Data versioning and time travel.**

B. **Support for ACID transactions.**

C. **The underlying data for Delta Lake tables is stored in JSON format**

D. **Support for batch and streaming data**

E. **You can select, insert, update, and delete rows of data in the same way you would in a relational database system.**

Question 10:

What is a lambda architecture and what does it try to solve?

A. **An architecture that splits incoming data into two paths - a batch path and a streaming path. This architecture helps address the need to provide real-time processing in addition to slower batch computations.**

B. **An architecture that defines a data processing pipeline whereby microservices act as compute resources for efficient large-scale data processing.**

C. **None of the listed options.**

D. **An architecture that employs the latest Scala runtimes in one or more Databricks clusters to provide the most efficient data processing platform available today.**

Question 11:

Scenario: Stark Industries was founded by Howard Stark during the early twentieth century, a great pioneer in different types of technology and constantly helping the United States Armed Forces with different and innovative weapons. Some years after the death of Howard Stark, his son Tony took over the company.

Currently in the process of making many IT improvements, you have been hired as a consultant to advise on several Microsoft Azure projects.

Stark has an Azure Data Lake Storage Gen2 account that contains a JSON file for customers. The file contains two attributes named FirstName and LastName.

The current project requires the team to copy the data from the JSON file to an Azure Synapse Analytics table by using Azure Databricks. A new column must be created that concatenates the FirstName and LastName values.

The team has created the following components:

- A destination table in Azure Synapse

- An Azure Blob storage container

- A service principal

The team has also created a list of actions they are considering to take in their Databricks notebook:

a. Mount the Data Lake Storage onto DBFS.

b. Use Staging (as a Blob)

c. Read file to DFrame

d. Perform Ops on DFrame

e. Write the results to Data Lake Storage.

f. Mount DBFS

g. Write to DWH from Stage

h. Perform transformations on the data frame.

As the Azure expert, Tony looks to you to guide the team on which actions to take, and the order to take them.

Which actions should you recommend the team perform in sequence in the Databricks notebook?

A. **f → c → d → b → g**

B. **h → c → d → g → b**

C. **d → a → c → f → g**

D. **c → a → f → h**

Question 12:

Monitoring is a key part of any mission-critical workload. It helps to proactively detect and prevent issues that might otherwise cause application or service downtime.

You can monitor Azure Stream Analytics jobs by using which of the following? (Select all that apply)

A. **Alerts on issues in applications or services**

B. **Predictive dashboards that show expected service and application health status**

C. **Diagnostic logs**

D. **An activity log for each running job**

E. **Real-time dashboards that show service and application health trends**

Question 13:

Scenario: You are working at OZcorp which is a multi-million dollar company run by Mayor Norman Osborn. Profits from the company are used to fund Norman's operatives, such as a police task force. You have been hired by OZcorp as a Microsoft Azure Expert.

At the moment, the team is designing the folder structure for an Azure Data Lake Storage Gen2 container where OZcorp users will query data by using a variety of services including Azure Databricks and Azure Synapse Analytics serverless SQL pools. The data will be secured by subject area. Most queries will include data from the current year or current month.

As the Azure expert, which folder structure should you recommend to support fast queries and simplified folder security?

A. **/{YYYY}/{MM}/{DD}/{SubjectArea}/{DataSource}/{FileData}_{YYYY}_{MM}_{DD}.csv**

B. **/{SubjectArea}/{DataSource}/{DD}/{MM}/{YYYY}/{FileData}_{YYYY}_{MM}_{DD}.csv**

C. **/{DD}/{MM}/{YYYY}/{SubjectArea}/{DataSource}/{FileData}_{YYYY}_{MM}_{DD}.csv**

D. **/{SubjectArea}/{DataSource}/{YYYY}/{MM}/{DD}/{FileData}_{YYYY}_{MM}_{DD}.csv**

Question 14:

Azure Storage provides a REST API to work with the containers and data stored in each account. Client libraries can save a significant amount of work for app developers because the API is tested and it often provides nicer wrappers around the data models sent and received by the REST API. Microsoft has Azure client libraries that support a number of languages and frameworks.

Which are Azure Storage supported languages and frameworks? (Select all that apply)

A. **Python**

B. **C#**

C. **Node.js**

D. **Go**

E. **.NET**

F. **Java**

Question 15:

Scenario: You are determining the type of Azure service needed to fit the following specifications and requirements:

Data classification: Semi-structured because of the need to extend or modify the schema for new products

Operations:

- Customers require a high number of read operations, with the ability to query many fields within the database.

- The business requires a high number of write operations to track its constantly changing inventory.

Latency & throughput: High throughput and low latency.

Transactional support: Because all of the data is both historical and yet changing, transactional support is required.

Which would be the best Azure service to select?

A. **Azure Cosmos DB**

B. **Azure Blob Storage**

C. **Azure Route Table**

D. **Azure SQL Database**

E. **Azure Queue Storage**

Question 16:

You can natively perform data transformations with Azure Synapse pipelines code free using the Mapping Data Flow task. Mapping Data Flows provide a fully visual experience with no coding required. Your data flows will run on your own execution cluster for scaled-out data processing.

Data flow activities can be operationalized via which of the following? (Select four)

A. **Integrate hub**

B. **Data Factory scheduling**

C. **Monitor hub**

D. **Control capabilities**

E. **Data hub**

F. **Monitoring capabilities**

G. **Flow capabilities**

H. **Manage hub**

Question 17:

When queries are submitted, a dedicated SQL pool query optimizer tries to determine which access paths to the data will result in the least amount of effort to retrieve the data required to resolve the query. It is a cost-based optimizer, and compares the cost of various query plans, and then chooses the plan with the lowest cost.

Statistics in a serverless SQL pool have the same objective of using a cost-based optimizer to choose an execution plan that will execute the fastest. How it creates the statistics is different.

True or False: In a serverless SQL pool, if statistics are missing, the query optimizer creates statistics on entire tables in the query predicate or join condition to improve cardinality estimates for the query plan.

A. **True**
B. **False**

Question 18:

In Azure Synapse Studio, the Data hub is where you access which of the following? (Select three)

A. **Data flows**

B. **SQL serverless databases**

C. **Notebooks**

D. **SQL scripts**

E. **Provisioned SQL pool databases**

F. **Pipeline canvas**

G. **External data sources**

H. **Power BI**

I. **Activities**

J. **Master Pipeline**

Question 19:

Scenario: You are working as a consultant at Avengers Security. At the moment, you are consulting with Tony, the lead of the IT team and the topic of discussion is Azure security.

Tony is asking the development team to enable the Delta Lake feature which will allow him to retrieve data from previous versions of a table.

Which of the following should you recommend to the team to employ?

    A. **Spark Structured Streaming**

    B. **Time Travel**

    C. **Hindsight**

    D. **Catalogue Tables**

Question 20:

How do you create a DataFrame object? (Select all that apply)

    A. **Use the DF.create() syntax**

    B. **Execute createOrReplaceObject()**

    C. **Use the createDataFrame() function**

    D. **Introduce a variable name and equate it to something like myDataFrameDF = spark.createDataFrame(data)**

Question 21:

Scenario: Your teammate is using import statements for transferring data between a dedicated SQL and Spark pool. Another teammate believes this is not necessary.

When is it unnecessary to use import statements for transferring data between a dedicated SQL and Spark pool?

    A. **None of the listed options.**

    B. **Import statements are not needed since they are pre-loaded with the Azure Synapse Studio integrated notebook experience.**

    C. **Use token-based authentication.**

    D. **It is always necessary to use import statements for transferring data between a dedicated SQL and Spark pool.**

    E. **Use the PySpark connector.**

Question 22:

While Agile, CI/CD, and DevOps are different, they support one another.

Which is best described by:

"Focuses on culture highlighting roles that emphasize responsiveness."

    A. **SDLC**

    B. **DevOps**

    C. **Agile**

    D. **CI/CD**

Question 23:

Scenario: The company you work at is a financial services firm, and can only have account managers allowed to access a customer's social insurance number, phone numbers or other personal identifiable information. It is imperative to distinguish the role of an account manager versus the manager of the account managers.

Which type of security would typically be best used in for this scenario?

A. **Random Data security**
B. **Column-level security**
C. **Table-level security**
D. **Row-level security**

Question 24:

Scenario: Iceberg Lounge is Gotham's coolest night club and Penguin's pad for operations. It's a high-end nightclub in Gotham which acts as a legit forefront of his illegal activities.

You have been hired as a contractor for Iceberg Lounge and you are consulting on various IT functions. Oswald Cobblepot runs the show.

Today, the team is working on an enterprise-wide Azure Data Lake Storage Gen2 account. The data lake is accessible only through an Azure virtual network named VNET1.

Oswald wants the team to build a SQL pool in Azure Synapse that will use data from the data lake. All the members of the Iceberg Lounge sales team are in an Azure Active Directory group named Sales. POSIX controls are used to assign the Sales group access to the files in the data lake. The team plans to load data to the SQL pool every hour.

- Required:
  Ensure that the SQL pool can load the sales data from the data lake.

The team has assembled some actions being considered to meet the requirement which are shown below. As the Azure expert, Oswald looks to you for advice of the correct actions to take.

Which of the following actions should you recommend they perform? (Select three)

A. **Use the snared access signature (SAS) as the credentials for the data load process.**

B. **Add your Azure Active Directory (Azure AD) account to the Sales group.**

C. **Create a managed identity.**

D. **Create a shared access signature (SAS).**

E. **Add the managed identity to the Sales group.**

F. **Use the managed identity as the credentials for the data load process.**

Question 25:

Azure Synapse Analytics allows you to create, control, and manage resource availability when workloads are competing. This allows you to manage the relative importance of each workload when waiting for available resources.

To facilitate faster load times, you can create a workload classifier for the load user with … [?]

A. the "importance" set to above_normal or High.

B. the "rank" set to 10.

C. the "urgency" set to 10.

D. the "priority" set to 1.

Question 26:

Scenario: The Red Lion National Bank is an offshore firm managed by Felix Manning and Wilson Fisk. The bank is migrating to Microsoft Azure and has hired you as an Expert Consultant to oversee their IT team and to assist as needed.

Felix is instructing a developer to drop a managed catalogue table that contains Delta Lake data.

If they drop the table, what will happen?

A. The table metadata and data files will be deleted.

B. The table metadata will remain in the catalogue, but the data files will be deleted.

C. The table metadata will be removed from the catalogue, but the data files will remain intact.

D. The table metadata will be removed from the catalogue, the data files will be retained in bulk storage as orphaned files.

Answers:
Question 1: A
Question 2: B
Question 3: B
Question 4: A, C, D, E, F
Question 5: A
Question 6: B
Question 7: B
Question 8: D
Question 9: A, B, D, E
Question 10: A
Question 11: A
Question 12: A, C, D, E
Question 13: D

Question 14: A, B, C, D, E, F
Question 15: A
Question 16: B, D, F, G
Question 17: B
Question 18: B, E, G
Question 19: B
Question 20: C, D
Question 21: B
Question 22: B
Question 23: B
Question 24: C, E, F
Question 25: A
Question 26: A