

Microsoft

Exam Questions DP-203

Data Engineering on Microsoft Azure



NEW QUESTION 1

- (Exam Topic 3)

You have two Azure Blob Storage accounts named account1 and account2?

You plan to create an Azure Data Factory pipeline that will use scheduled intervals to replicate newly created or modified blobs from account1 to account2?

You need to recommend a solution to implement the pipeline. The solution must meet the following requirements:

- Ensure that the pipeline only copies blobs that were created or modified since the most recent replication event.
- Minimize the effort to create the pipeline. What should you recommend?

- A. Create a pipeline that contains a flowlet.
- B. Create a pipeline that contains a Data Flow activity.
- C. Run the Copy Data tool and select Metadata-driven copy task.
- D. Run the Copy Data tool and select Built-in copy task.

Answer: A

NEW QUESTION 2

- (Exam Topic 3)

You develop data engineering solutions for a company.

A project requires the deployment of data to Azure Data Lake Storage.

You need to implement role-based access control (RBAC) so that project members can manage the Azure Data Lake Storage resources.

Which three actions should you perform? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

- A. Assign Azure AD security groups to Azure Data Lake Storage.
- B. Configure end-user authentication for the Azure Data Lake Storage account.
- C. Configure service-to-service authentication for the Azure Data Lake Storage account.
- D. Create security groups in Azure Active Directory (Azure AD) and add project members.
- E. Configure access control lists (ACL) for the Azure Data Lake Storage account.

Answer: ADE

Explanation:

References:

<https://docs.microsoft.com/en-us/azure/data-lake-store/data-lake-store-secure-data>

NEW QUESTION 3

- (Exam Topic 3)

You are designing a date dimension table in an Azure Synapse Analytics dedicated SQL pool. The date dimension table will be used by all the fact tables.

Which distribution type should you recommend to minimize data movement?

- A. HASH
- B. REPLICATE
- C. ROUND ROBIN

Answer: B

Explanation:

A replicated table has a full copy of the table available on every Compute node. Queries run fast on replicated tables since joins on replicated tables don't require data movement. Replication requires extra storage, though, and isn't practical for large tables.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-overview>

NEW QUESTION 4

- (Exam Topic 3)

You have an Azure Synapse Analytics workspace named WS1 that contains an Apache Spark pool named Pool1.

You plan to create a database named DB1 in Pool1.

You need to ensure that when tables are created in DB1, the tables are available automatically as external tables to the built-in serverless SQL pod.

Which format should you use for the tables in DB1?

- A. Parquet
- B. CSV
- C. ORC
- D. JSON

Answer: A

Explanation:

Serverless SQL pool can automatically synchronize metadata from Apache Spark. A serverless SQL pool database will be created for each database existing in serverless Apache Spark pools.

For each Spark external table based on Parquet or CSV and located in Azure Storage, an external table is created in a serverless SQL pool database.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-storage-files-spark-tables>

NEW QUESTION 5

- (Exam Topic 3)

You have an Azure Data Lake Storage account that contains a staging zone.

You need to design a dairy process to ingest incremental data from the staging zone, transform the data by executing an R script, and then insert the transformed

data into a data warehouse in Azure Synapse Analytics.

Solution: You use an Azure Data Factory schedule trigger to execute a pipeline that copies the data to a staging table in the data warehouse, and then uses a stored procedure to execute the R script.

Does this meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

If you need to transform data in a way that is not supported by Data Factory, you can create a custom activity with your own data processing logic and use the activity in the pipeline.

Note: You can use data transformation activities in Azure Data Factory and Synapse pipelines to transform and process your raw data into predictions and insights at scale.

Reference:

<https://docs.microsoft.com/en-us/azure/data-factory/transform-data>

NEW QUESTION 6

- (Exam Topic 3)

You have an Azure Synapse Analytics dedicated SQL pool named Pool1 that contains a table named Sales. Sales has row-level security (RLS) applied. RLS uses the following predicate filter.

```
CREATE FUNCTION Security.fn_securitypredicate(@SalesRep AS sysname)
    RETURNS TABLE
WITH SCHEMABINDING
AS
    RETURN SELECT 1 AS fn_securitypredicate_result
WHERE @SalesRep = USER_NAME() OR USER_NAME() = 'Manager';
```

A user named SalesUser1 is assigned the db_datareader role for Pool1.

A user named SalesUser1 is assigned the db_datareader role for Pool1. Which rows in the Sales table are returned when SalesUser1 queries the table?

- A. only the rows for which the value in the User_Name column is SalesUser1
- B. all the rows
- C. only the rows for which the value in the SalesRep column is Manager
- D. only the rows for which the value in the SalesRep column is SalesUser1

Answer: A

NEW QUESTION 7

- (Exam Topic 3)

You have an Azure Synapse Analytics dedicated SQL pool mat contains a table named dbo.Users.

You need to prevent a group of users from reading user email addresses from dbo.Users. What should you use?

- A. row-level security
- B. column-level security
- C. Dynamic data masking
- D. Transparent Data Encryption (TDD)

Answer: B

NEW QUESTION 8

- (Exam Topic 3)

You are designing an Azure Databricks cluster that runs user-defined local processes. You need to recommend a cluster configuration that meets the following requirements:

- Minimize query latency.
 - Maximize the number of users that can run queues on the cluster at the same time « Reduce overall costs without compromising other requirements
- Which cluster type should you recommend?

- A. Standard with Auto termination
- B. Standard with Autoscaling
- C. High Concurrency with Autoscaling
- D. High Concurrency with Auto Termination

Answer: C

Explanation:

A High Concurrency cluster is a managed cloud resource. The key benefits of High Concurrency clusters are that they provide fine-grained sharing for maximum resource utilization and minimum query latencies.

Databricks chooses the appropriate number of workers required to run your job. This is referred to as autoscaling. Autoscaling makes it easier to achieve high cluster utilization, because you don't need to provision the cluster to match a workload.

Reference:

<https://docs.microsoft.com/en-us/azure/databricks/clusters/configure>

NEW QUESTION 9

- (Exam Topic 3)

You plan to build a structured streaming solution in Azure Databricks. The solution will count new events in five-minute intervals and report only events that arrive

during the interval. The output will be sent to a Delta Lake table.
Which output mode should you use?

- A. complete
- B. update
- C. append

Answer: C

Explanation:

Append Mode: Only new rows appended in the result table since the last trigger are written to external storage. This is applicable only for the queries where existing rows in the Result Table are not expected to change.
<https://docs.databricks.com/getting-started/spark/streaming.html>

NEW QUESTION 10

- (Exam Topic 3)

You have two fact tables named Flight and Weather. Queries targeting the tables will be based on the join between the following columns.

Table	Column
Flight	ArrivalAirportID ArrivalDateTime
Weather	AirportID ReportDateTime

You need to recommend a solution that maximum query performance. What should you include in the recommendation?

- A. In each table, create a column as a composite of the other two columns in the table.
- B. In each table, create an IDENTITY column.
- C. In the tables, use a hash distribution of ArriveDateTime and ReportDateTime.
- D. In the tables, use a hash distribution of ArriveAirPortID and AirportID.

Answer: D

NEW QUESTION 10

- (Exam Topic 3)

You have an Azure Synapse Analytics dedicated SQL pool named pool1.

You plan to implement a star schema in pool1 and create a new table named DimCustomer by using the following code.

```
CREATE TABLE dbo.[DimCustomer](
    [CustomerKey] int NOT NULL,
    [CustomerSourceID] [int] NOT NULL,
    [Title] [nvarchar](8) NULL,
    [FirstName] [nvarchar](50) NOT NULL,
    [MiddleName] [nvarchar](50) NULL,
    [LastName] [nvarchar](50) NOT NULL,
    [Suffix] [nvarchar](10) NULL,
    [CompanyName] [nvarchar](128) NULL,
    [SalesPerson] [nvarchar](256) NULL,
    [EmailAddress] [nvarchar](50) NULL,
    [Phone] [nvarchar](25) NULL,
    [InsertedDate] [datetime] NOT NULL,
    [ModifiedDate] [datetime] NOT NULL,
    [HashKey] [varchar](100) NOT NULL,
    [IsCurrentRow] [bit] NOT NULL
)
WITH
(
    DISTRIBUTION = REPLICATE,
    CLUSTERED COLUMNSTORE INDEX
);
GO
```

You need to ensure that DimCustomer has the necessary columns to support a Type 2 slowly changing dimension (SCD). Which two columns should you add?
Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

- A. [HistoricalSalesPerson] [nvarchar] (256) NOT NULL
- B. [EffectiveEndDate] [datetime] NOT NULL
- C. [PreviousModifiedDate] [datetime] NOT NULL
- D. [RowID] [bigint] NOT NULL
- E. [EffectiveStartDate] [datetime] NOT NULL

Answer: AB

NEW QUESTION 14

- (Exam Topic 3)

You have a SQL pool in Azure Synapse.

You plan to load data from Azure Blob storage to a staging table. Approximately 1 million rows of data will be loaded daily. The table will be truncated before each daily load.

You need to create the staging table. The solution must minimize how long it takes to load the data to the staging table.

How should you configure the table? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Distribution:

	▼
Hash	
Replicated	
Round-robin	

Indexing:

	▼
Clustered	
Clustered columnstore	
Heap	

Partitioning:

	▼
Date	
None	

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Graphical user interface, application, table Description automatically generated

Box 1: Hash

Hash-distributed tables improve query performance on large fact tables. They can have very large numbers of rows and still achieve high performance.

Box 2: Clustered columnstore

When creating partitions on clustered columnstore tables, it is important to consider how many rows belong to each partition. For optimal compression and performance of clustered columnstore tables, a minimum of 1 million rows per distribution and partition is needed.

Box 3: Date

Table partitions enable you to divide your data into smaller groups of data. In most cases, table partitions are created on a date column.

Partition switching can be used to quickly remove or replace a section of a table. Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-partitio> <https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-distribu>

NEW QUESTION 17

- (Exam Topic 3)

You have an Azure data factory that connects to a Microsoft Purview account. The data factory is registered in Microsoft Purview.

You update a Data Factory pipeline.

You need to ensure that the updated lineage is available in Microsoft Purview.

What You have an Azure subscription that contains an Azure SQL database named DB1 and a storage account named storage1. The storage1 account contains a file named File1.txt. File1.txt contains the names of selected tables in DB1.

You need to use an Azure Synapse pipeline to copy data from the selected tables in DB1 to the files in storage1. The solution must meet the following requirements:

- The Copy activity in the pipeline must be parameterized to use the data in File1.txt to identify the source and destination of the copy.
- Copy activities must occur in parallel as often as possible.

Which two pipeline activities should you include in the pipeline? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

- A. If Condition
- B. ForEach
- C. Lookup
- D. Get Metadata

Answer: CD

NEW QUESTION 22

- (Exam Topic 3)

You have an Azure Factory instance named DF1 that contains a pipeline named PL1.PL1 includes a tumbling window trigger.

You create five clones of PL1. You configure each clone pipeline to use a different data source.

You need to ensure that the execution schedules of the clone pipeline match the execution schedule of PL1. What should you do?

- A. Add a new trigger to each cloned pipeline
- B. Associate each cloned pipeline to an existing trigger.
- C. Create a tumbling window trigger dependency for the trigger of PL1.
- D. Modify the Concurrency setting of each pipeline.

Answer: B

NEW QUESTION 25

- (Exam Topic 3)

You have the following Azure Stream Analytics query.

```
WITH
```

```
step1 AS (SELECT *
           FROM input1
           PARTITION BY StateID
           INTO 10),
step2 AS (SELECT *
           FROM input2
           PARTITION BY StateID
           INTO 10)

SELECT *
INTO output
FROM step1
PARTITION BY StateID
UNION
SELECT * INTO output
FROM step2
PARTITION BY StateID
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No.
NOTE: Each correct selection is worth one point.

Statements	Yes	No
The query combines two streams of partitioned data.	<input type="radio"/>	<input type="radio"/>
The stream scheme key and count must match the output scheme.	<input type="radio"/>	<input type="radio"/>
Providing 60 streaming units will optimize the performance of the query.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: No
Note: You can now use a new extension of Azure Stream Analytics SQL to specify the number of partitions of a stream when reshuffling the data. The outcome is a stream that has the same partition scheme. Please see below for an example: WITH step1 AS (SELECT * FROM [input1] PARTITION BY DeviceID INTO 10), step2 AS (SELECT * FROM [input2] PARTITION BY DeviceID INTO 10) SELECT * INTO [output] FROM step1 PARTITION BY DeviceID UNION step2 PARTITION BY DeviceID Note: The new extension of Azure Stream Analytics SQL includes a keyword INTO that allows you to specify the number of partitions for a stream when performing reshuffling using a PARTITION BY statement.
Box 2: Yes
When joining two streams of data explicitly repartitioned, these streams must have the same partition key and partition count. Box 3: Yes
Streaming Units (SUs) represents the computing resources that are allocated to execute a Stream Analytics job. The higher the number of SUs, the more CPU and memory resources are allocated for your job.
In general, the best practice is to start with 6 SUs for queries that don't use PARTITION BY. Here there are 10 partitions, so 6x10 = 60 SUs is good.
Note: Remember, Streaming Unit (SU) count, which is the unit of scale for Azure Stream Analytics, must be adjusted so the number of physical resources available to the job can fit the partitioned flow. In general, six SUs is a good number to assign to each partition. In case there are insufficient resources assigned to the job, the system will only apply the repartition if it benefits the job.
Reference:
<https://azure.microsoft.com/en-in/blog/maximize-throughput-with-repartitioning-in-azure-stream-analytics/> <https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-streaming-unit-consumption>

NEW QUESTION 27

- (Exam Topic 3)
You are designing an Azure Synapse solution that will provide a query interface for the data stored in an Azure Storage account. The storage account is only accessible from a virtual network.
You need to recommend an authentication mechanism to ensure that the solution can access the source data.
What should you recommend?

- A. a managed identity
- B. anonymous public read access
- C. a shared key

Answer: A

Explanation:

Managed Identity authentication is required when your storage account is attached to a VNet. Reference:
<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/quickstart-bulk-load-copy-tsql-exa>

NEW QUESTION 30

- (Exam Topic 3)

You have an enterprise data warehouse in Azure Synapse Analytics named DW1 on a server named Server1. You need to determine the size of the transaction log file for each distribution of DW1. What should you do?

- A. On DW1, execute a query against the sys.database_files dynamic management view.
- B. From Azure Monitor in the Azure portal, execute a query against the logs of DW1.
- C. Execute a query against the logs of DW1 by using the Get-AzOperationalInsightsSearchResult PowerShell cmdlet.
- D. On the master database, execute a query against the sys.dm_pdw_nodes_os_performance_counters dynamic management view.

Answer: A

Explanation:

For information about the current log file size, its maximum size, and the autogrow option for the file, you can also use the size, max_size, and growth columns for that log file in sys.database_files.

Reference:

<https://docs.microsoft.com/en-us/sql/relational-databases/logs/manage-the-size-of-the-transaction-log-file>

NEW QUESTION 34

- (Exam Topic 3)

You have an Azure subscription that contains an Azure Synapse Analytics workspace named workspace1. Workspace1 contains a dedicated SQL pool named SQL Pool and an Apache Spark pool named sparkpool. Sparkpool1 contains a DataFrame named pyspark.df.

You need to write the contents of pyspark_df to a table in SQLPoolM by using a PySpark notebook. How should you complete the code? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Answer Area



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area



NEW QUESTION 36

- (Exam Topic 3)

You have an Azure Databricks workspace named workspace1 in the Standard pricing tier. Workspace1 contains an all-purpose cluster named cluster1. You need to reduce the time it takes for cluster1 to start and scale up. The solution must minimize costs. What should you do first?

- A. Upgrade workspace1 to the Premium pricing tier.
- B. Create a cluster policy in workspace1.
- C. Create a pool in workspace1.
- D. Configure a global init script for workspace1.

Answer: C

Explanation:

You can use Databricks Pools to Speed up your Data Pipelines and Scale Clusters Quickly.

Databricks Pools, a managed cache of virtual machine instances that enables clusters to start and scale 4 times faster.

Reference:

<https://databricks.com/blog/2019/11/11/databricks-pools-speed-up-data-pipelines.html>

NEW QUESTION 40

- (Exam Topic 3)

You use Azure Stream Analytics to receive Twitter data from Azure Event Hubs and to output the data to an Azure Blob storage account.

You need to output the count of tweets from the last five minutes every minute. Which windowing function should you use?

- A. Sliding
- B. Session
- C. Tumbling
- D. Hopping

Answer: D

Explanation:

Hopping window functions hop forward in time by a fixed period. It may be easy to think of them as Tumbling windows that can overlap and be emitted more often than the window size. Events can belong to more than one Hopping window result set. To make a Hopping window the same as a Tumbling window, specify the hop size to be the same as the window size.

Reference:

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-window-functions>

NEW QUESTION 44

- (Exam Topic 3)

You have an Azure Data Lake Storage Gen2 container that contains 100 TB of data.

You need to ensure that the data in the container is available for read workloads in a secondary region if an outage occurs in the primary region. The solution must minimize costs.

Which type of data redundancy should you use?

- A. zone-redundant storage (ZRS)
- B. read-access geo-redundant storage (RA-GRS)
- C. locally-redundant storage (LRS)
- D. geo-redundant storage (GRS)

Answer: B

Explanation:

Geo-redundant storage (with GRS or GZRS) replicates your data to another physical location in the secondary region to protect against regional outages.

However, that data is available to be read only if the customer or Microsoft initiates a failover from the primary to secondary region. When you enable read access to the secondary region, your data is available to be read at all times, including in a situation where the primary region becomes unavailable.

Reference:

<https://docs.microsoft.com/en-us/azure/storage/common/storage-redundancy>

NEW QUESTION 49

- (Exam Topic 3)

You have an Azure Synapse Analytics dedicated SQL pool that contains the users shown in the following table.

Name	Role
User1	Server admin
User2	db_datereader

User1 executes a query on the database, and the query returns the results shown in the following exhibit.

```
1  SELECT c.name,
2      tbl.name as table_name,
3      typ.name as datatype,
4      c.is_masked,
5      c.masking_function
6  FROM sys.masked_columns AS c
7  INNER JOIN sys.tables AS tbl ON c.[object_id] = tbl.[object_id]
8  INNER JOIN sys.types typ ON c.user_type_id = typ.user_type_id
9  WHERE is_masked = 1;
10
```

Results Messages

	name	table_name	datatype	is_masked	masking_function
1	BirthDate	DimCustomer	date	1	default()
2	Gender	DimCustomer	nvarchar	1	default()
3	EmailAddress	DimCustomer	nvarchar	1	email()
4	YearlyIncome	DimCustomer	money	1	default()

User1 is the only user who has access to the unmasked data.

Use the drop-down menus to select the answer choice that completes each statement based on the information presented in the graphic.

NOTE: Each correct selection is worth one point.

When User2 queries the YearlyIncome column, the values returned will be [answer choice].

a random number

the values stored in the database

XXXX

0

When User1 queries the BirthDate column, the values returned will be [answer choice].

a random date

the values stored in the database

XXXX

1900-01-01

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:
Graphical user interface, text, application, email Description automatically generated
Box 1: 0

The YearlyIncome column is of the money data type.
The Default masking function: Full masking according to the data types of the designated fields

> Use a zero value for numeric data types (bigint, bit, decimal, int, money, numeric, smallint, smallmoney, tinyint, float, real).

Box 2: the values stored in the database
Users with administrator privileges are always excluded from masking, and see the original data without any mask.
Reference:
<https://docs.microsoft.com/en-us/azure/azure-sql/database/dynamic-data-masking-overview>

NEW QUESTION 53

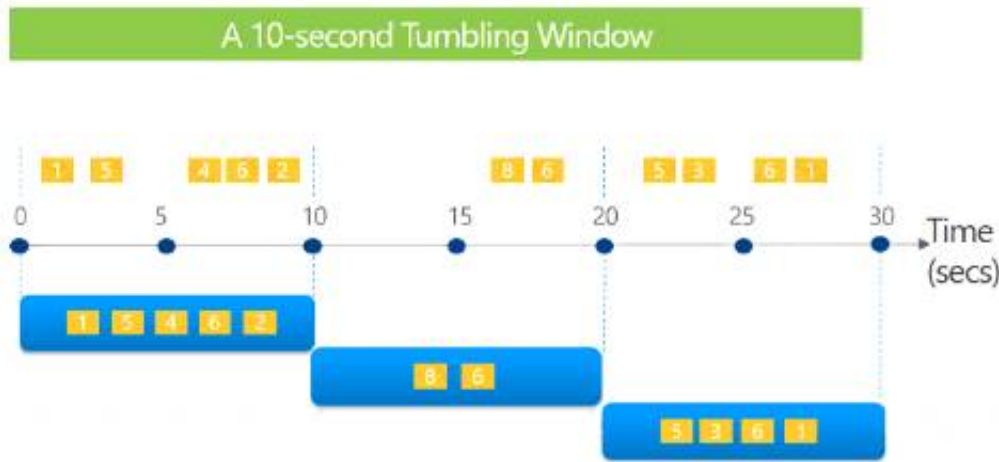
- (Exam Topic 3)
You are designing an Azure Stream Analytics job to process incoming events from sensors in retail environments. You need to process the events to produce a running average of shopper counts during the previous 15 minutes, calculated at five-minute intervals. Which type of window should you use?

- A. snapshot
- B. tumbling
- C. hopping
- D. sliding

Answer: B

Explanation:
Tumbling windows are a series of fixed-sized, non-overlapping and contiguous time intervals. The following diagram illustrates a stream with a series of events and how they are mapped into 10-second tumbling windows.

Tell me the count of tweets per time zone every 10 seconds



```
SELECT TimeZone, COUNT(*) AS Count
FROM TwitterStream TIMESTAMP BY CreatedAt
GROUP BY TimeZone, TumblingWindow(second,10)
```

Reference:
<https://docs.microsoft.com/en-us/stream-analytics-query/tumbling-window-azure-stream-analytics>

NEW QUESTION 58

- (Exam Topic 3)

You are monitoring an Azure Stream Analytics job.

You discover that the Backlogged Input Events metric is increasing slowly and is consistently non-zero. You need to ensure that the job can handle all the events. What should you do?

- A. Change the compatibility level of the Stream Analytics job.
- B. Increase the number of streaming units (SUs).
- C. Remove any named consumer groups from the connection and use \$default.
- D. Create an additional output stream for the existing input stream.

Answer: B

Explanation:

Backlogged Input Events: Number of input events that are backlogged. A non-zero value for this metric implies that your job isn't able to keep up with the number of incoming events. If this value is slowly increasing or consistently non-zero, you should scale out your job. You should increase the Streaming Units.

Note: Streaming Units (SUs) represents the computing resources that are allocated to execute a Stream Analytics job. The higher the number of SUs, the more CPU and memory resources are allocated for your job.

Reference:
<https://docs.microsoft.com/bs-cyrl-ba/azure/stream-analytics/stream-analytics-monitoring>

NEW QUESTION 61

- (Exam Topic 3)

You are designing a fact table named FactPurchase in an Azure Synapse Analytics dedicated SQL pool. The table contains purchases from suppliers for a retail store. FactPurchase will contain the following columns.

Name	Data type	Nullable
PurchaseKey	Bigint	No
DateKey	Int	No
SupplierKey	Int	No
StockItemKey	Int	No
PurchaseOrderID	Int	Yes
OrderedQuantity	Int	No
OrderedOuters	Int	No
ReceivedOuters	Int	No
Package	Nvarchar(50)	No
IsOrderFinalized	Bit	No
LineageKey	Int	No

FactPurchase will have 1 million rows of data added daily and will contain three years of data. Transact-SQL queries similar to the following query will be executed daily.

```
SELECT  
SupplierKey, StockItemKey, COUNT(*) FROM FactPurchase  
WHERE DateKey >= 20210101 AND DateKey <= 20210131  
GROUP By SupplierKey, StockItemKey
```

Which table distribution will minimize query times?

- A. round-robin
- B. replicated
- C. hash-distributed on DateKey
- D. hash-distributed on PurchaseKey

Answer: D

Explanation:

Hash-distributed tables improve query performance on large fact tables, and are the focus of this article. Round-robin tables are useful for improving loading speed.

Reference:
<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-distribu>

NEW QUESTION 62

- (Exam Topic 3)

You build a data warehouse in an Azure Synapse Analytics dedicated SQL pool.

Analysts write a complex SELECT query that contains multiple JOIN and CASE statements to transform data for use in inventory reports. The inventory reports will use the data and additional WHERE parameters depending on the report. The reports will be produced once daily.

You need to implement a solution to make the dataset available for the reports. The solution must minimize query times.

What should you implement?

- A. a materialized view
- B. a replicated table

- C. in ordered clustered columnstore index
- D. result set chaching

Answer: A

Explanation:

Materialized views for dedicated SQL pools in Azure Synapse provide a low maintenance method for complex analytical queries to get fast performance without any query change.

Note: When result set caching is enabled, dedicated SQL pool automatically caches query results in the user database for repetitive use. This allows subsequent query executions to get results directly from the persisted cache so recomputation is not needed. Result set caching improves query performance and reduces compute resource usage. In addition, queries using cached results set do not use any concurrency slots and thus do not count against existing concurrency limits.

Reference:

[https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/performance-tuning-materialized-](https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/performance-tuning-materialized-views) [https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/performance-tuning-result-set-cac](https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/performance-tuning-result-set-caching)

NEW QUESTION 63

- (Exam Topic 3)

You are monitoring an Azure Stream Analytics job by using metrics in Azure.

You discover that during the last 12 hours, the average watermark delay is consistently greater than the configured late arrival tolerance.

What is a possible cause of this behavior?

- A. Events whose application timestamp is earlier than their arrival time by more than five minutes arrive as inputs.
- B. There are errors in the input data.
- C. The late arrival policy causes events to be dropped.
- D. The job lacks the resources to process the volume of incoming data.

Answer: D

Explanation:

Watermark Delay indicates the delay of the streaming data processing job.

There are a number of resource constraints that can cause the streaming pipeline to slow down. The watermark delay metric can rise due to:

- Not enough processing resources in Stream Analytics to handle the volume of input events. To scale up resources, see [Understand and adjust Streaming Units](#).
- Not enough throughput within the input event brokers, so they are throttled. For possible solutions, see [Automatically scale up Azure Event Hubs throughput units](#).
- Output sinks are not provisioned with enough capacity, so they are throttled. The possible solutions vary widely based on the flavor of output service being used.

Reference:

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-time-handling>

NEW QUESTION 68

- (Exam Topic 3)

You use Azure Data Lake Storage Gen2.

You need to ensure that workloads can use filter predicates and column projections to filter data at the time the data is read from disk.

Which two actions should you perform? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

- A. Reregister the Microsoft Data Lake Store resource provider.
- B. Reregister the Azure Storage resource provider.
- C. Create a storage policy that is scoped to a container.
- D. Register the query acceleration feature.
- E. Create a storage policy that is scoped to a container prefix filter.

Answer: BD

NEW QUESTION 72

- (Exam Topic 3)

You have an Azure subscription that contains a Microsoft Purview account named MP1, an Azure data factory named DF1, and a storage account named storage. MP1 is configured

10 scan storage1. DF1 is connected to MP1 and contains 3 dataset named DS1. DS1 references 2 file in storage.

In DF1, you plan to create a pipeline that will process data from DS1.

You need to review the schema and lineage information in MP1 for the data referenced by DS1.

Which two features can you use to locate the information? Each correct answer presents a complete solution. NOTE: Each correct answer is worth one point.

- A. the Storage browser of storage1 in the Azure portal
- B. the search bar in the Azure portal
- C. the search bar in Azure Data Factory Studio
- D. the search bar in the Microsoft Purview governance portal

Answer: CD

Explanation:

➤ The search bar in the Microsoft Purview governance portal: This is a feature that allows you to search for assets in your data estate using keywords, filters, and facets. You can use the search bar to find the files in storage1 that are referenced by DS1, and then view their schema and lineage information in the asset details page12.

➤ The search bar in Azure Data Factory Studio: This is a feature that allows you to search for datasets, linked services, pipelines, and other resources in your data factory. You can use the search bar to find DS1 in DF1, and then view its schema and lineage information in the dataset details page. You can also click on the Open in Purview button to open the corresponding asset in MP13.

The two features that can be used to locate the schema and lineage information for the data referenced by DS1 are the search bar in Azure Data Factory Studio

and the search bar in the Microsoft Purview governance portal.

The search bar in Azure Data Factory Studio allows you to search for the dataset DS1 and view its properties and lineage. This can help you locate information about the source and destination data stores, as well as the transformations that were applied to the data.

The search bar in the Microsoft Purview governance portal allows you to search for the storage account and view its metadata, including schema and lineage information. This can help you understand the different data assets that are stored in the storage account and how they are related to each other.

The Storage browser of storage1 in the Azure portal may allow you to view the files that are stored in the storage account, but it does not provide lineage or schema information for those files. Similarly, the search bar in the Azure portal may allow you to search for resources in the Azure subscription, but it does not provide detailed information about the data assets themselves.

References:

- > What is Azure Purview?
- > Use Azure Data Factory Studio

NEW QUESTION 74

- (Exam Topic 1)

You need to design the partitions for the product sales transactions. The solution must meet the sales transaction dataset requirements.

What should you include in the solution? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Partition product sales transactions data by:

Sales date

Product ID

Promotion ID

Store product sales transactions data in:

An Azure Synapse Analytics dedicated SQL pool

An Azure Synapse Analytics serverless SQL pool

An Azure Data Lake Storage Gen2 account linked to an Azure Synapse Analytics workspace

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Sales date

Scenario: Contoso requirements for data integration include:

- > Partition data that contains sales transaction records. Partitions must be designed to provide efficient loads by month. Boundary values must belong to the partition on the right.

Box 2: An Azure Synapse Analytics Dedicated SQL pool

Scenario: Contoso requirements for data integration include:

- > Ensure that data storage costs and performance are predictable.

The size of a dedicated SQL pool (formerly SQL DW) is determined by Data Warehousing Units (DWU). Dedicated SQL pool (formerly SQL DW) stores data in relational tables with columnar storage. This format significantly reduces the data storage costs, and improves query performance.

Synapse analytics dedicated sql pool Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-overview-wha>

NEW QUESTION 79

- (Exam Topic 1)

You need to integrate the on-premises data sources and Azure Synapse Analytics. The solution must meet the data integration requirements.

Which type of integration runtime should you use?

- A. Azure-SSIS integration runtime
- B. self-hosted integration runtime
- C. Azure integration runtime

Answer: C

NEW QUESTION 81

- (Exam Topic 1)

You need to design an analytical storage solution for the transactional data. The solution must meet the sales transaction dataset requirements.

What should you include in the solution? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Table type to store retail store data:

▼

Hash

Replicated

Round-robin

Table type to store promotional data:

▼

Hash

Replicated

Round-robin

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Graphical user interface, text, application, table Description automatically generated
Box 1: Round-robin
Round-robin tables are useful for improving loading speed.
Scenario: Partition data that contains sales transaction records. Partitions must be designed to provide efficient loads by month.
Box 2: Hash
Hash-distributed tables improve query performance on large fact tables. Reference:
<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-distribu>

NEW QUESTION 82

- (Exam Topic 1)
You need to design a data retention solution for the Twitter feed data records. The solution must meet the customer sentiment analytics requirements.
Which Azure Storage functionality should you include in the solution?

- A. change feed
- B. soft delete
- C. time-based retention
- D. lifecycle management

Answer: D

Explanation:

Scenario: Purge Twitter feed data records that are older than two years.
Data sets have unique lifecycles. Early in the lifecycle, people access some data often. But the need for access often drops drastically as the data ages. Some data remains idle in the cloud and is rarely accessed once stored. Some data sets expire days or months after creation, while other data sets are actively read and modified throughout their lifetimes. Azure Storage lifecycle management offers a rule-based policy that you can use to transition blob data to the appropriate access tiers or to expire data at the end of the data lifecycle.
Reference:
<https://docs.microsoft.com/en-us/azure/storage/blobs/lifecycle-management-overview>

NEW QUESTION 87

- (Exam Topic 3)
You have an Azure Active Directory (Azure AD) tenant that contains a security group named Group1. You have an Azure Synapse Analytics dedicated SQL pool named dw1 that contains a schema named schema1.
You need to grant Group1 read-only permissions to all the tables and views in schema1. The solution must use the principle of least privilege.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.
NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.

Actions

Answer Area

Create a database role named Role1 and grant Role1 SELECT permissions to schema1.

Create a database role named Role1 and grant Role1 SELECT permissions to dw1.

Assign the Azure role-based access control (Azure RBAC) Reader role for dw1 to Group1.

Create a database user in dw1 that represents Group1 and uses the FROM EXTERNAL PROVIDER clause.

Assign Role1 to the Group1 database user.

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Step 1: Create a database role named Role1 and grant Role1 SELECT permissions to schema You need to grant Group1 read-only permissions to all the tables and views in schema1.

Place one or more database users into a database role and then assign permissions to the database role. Step 2: Assign Rol1 to the Group database user

Step 3: Assign the Azure role-based access control (Azure RBAC) Reader role for dw1 to Group1 Reference:

<https://docs.microsoft.com/en-us/azure/data-share/how-to-share-from-sql>

NEW QUESTION 89

- (Exam Topic 3)

You are creating an Azure Data Factory data flow that will ingest data from a CSV file, cast columns to specified types of data, and insert the data into a table in an Azure Synapse Analytics dedicated SQL pool. The CSV file contains columns named username, comment and date.

The data flow already contains the following:

- A source transformation
- A Derived Column transformation to set the appropriate types of data
- A sink transformation to land the data in the pool

You need to ensure that the data flow meets the following requirements;

- All valid rows must be written to the destination table.
- Truncation errors in the comment column must be avoided proactively.
- Any rows containing comment values that will cause truncation errors upon insert must be written to a file in blob storage.

Which two actions should you perform? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point

- A. Add a select transformation that selects only the rows which will cause truncation errors.
- B. Add a sink transformation that writes the rows to a file in blob storage.
- C. Add a filter transformation that filters out rows which will cause truncation errors.
- D. Add a Conditional Split transformation that separates the rows which will cause truncation errors.

Answer: BD





NEW QUESTION 91

- (Exam Topic 3)

You have an Azure data factory that has the Git repository settings shown in the following exhibit.

Git repository

Git repository information associated with your data factory. [CI/CD best practices](#)

 Edit  Overwrite live mode  Disconnect  Import resources

Repository type	Azure DevOps Git
Azure DevOps Account	
Project name	ADFDDeployDemo
Repository name	ADFDDeployDemo
Collaboration branch	main
Publish branch	adf_publish
Root folder	/
Last published commit	23b144ac4aa7daf16f2fe7c2ab0eb303a8e4ed65
Publish (from ADF Studio)	Enabled

Use the drop-down menus to select the answer choose that completes each statement based on the information presented in the graphic.

NOTE: Each correct answer is worth one point.

Answer Area

Changes to pipelines will be saved in Azure DevOps [answer choice].

every 20 seconds

every 20 seconds

when the pipeline is published

when the pipeline is saved

To publish changes by using Azure Data Factory Studio, the changes must first be saved in the [answer choice].

root folder

adf_publish branch

main branch

root folder

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

Changes to pipelines will be saved in Azure DevOps [answer choice].

every 20 seconds
every 20 seconds
when the pipeline is published
when the pipeline is saved

To publish changes by using Azure Data Factory Studio, the changes must first be saved in the [answer choice].

root folder
adf_publish branch
main branch
root folder

NEW QUESTION 95

- (Exam Topic 3)
You have an enterprise data warehouse in Azure Synapse Analytics.
You need to monitor the data warehouse to identify whether you must scale up to a higher service level to accommodate the current workloads
Which is the best metric to monitor?
More than one answer choice may achieve the goal. Select the BEST answer.

- A. Data 10 percentage
- B. CPU percentage
- C. DWU used
- D. DWU percentage

Answer: C

NEW QUESTION 96

- (Exam Topic 3)
You configure version control for an Azure Data Factory instance as shown in the following exhibit.

Connections

Linked services

Integration runtimes

Source control

Git configuration

ARM template

Parameterization template

Author

Triggers

Global parameters

Security

Customer managed key

Managed private endpoints

Git repository

Git repository information associated with your data factory. [CI/CD best practices](#)

Setting Disconnect

Repository type Azure DevOps Git

Azure DevOps Account CONTOSO

Project name Data

Repository name dwh_batchetl

Collaboration branch main

Publish branch adf_publish

Root folder /

Use the drop-down menus to select the answer choice that completes each statement based on the information presented in the graphic.
NOTE: Each correct selection is worth one point.

Azure Resource Manager (ARM) templates for the pipeline assets are stored in [answer choice].

/
adf_publish
main
Parameterization template

A Data Factory Azure Resource Manager (ARM) template named contososales can be found in [answer choice].

/
/contososales
/dwh_batchetl/adf_publish/contososales
/main

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Letter Description automatically generated

Box 1: adf_publish

The Publish branch is the branch in your repository where publishing related ARM templates are stored and updated. By default, it's adf_publish.

Box 2: / dwh_batchetl/adf_publish/contososales

Note: RepositoryName (here dwh_batchetl): Your Azure Repos code repository name. Azure Repos projects contain Git repositories to manage your source code as your project grows. You can create a new repository or use an existing repository that's already in your project.

Reference:

<https://docs.microsoft.com/en-us/azure/data-factory/source-control>

NEW QUESTION 97

- (Exam Topic 3)

You need to create an Azure Data Factory pipeline to process data for the following three departments at your company: Ecommerce, retail, and wholesale. The solution must ensure that data can also be processed for the entire company.

How should you complete the Data Factory data flow script? To answer, drag the appropriate values to the correct targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Values

all, ecommerce, retail, wholesale
dept=='ecommerce', dept=='retail', dept=='wholesale'
dept=='ecommerce', dept=='wholesale', dept=='retail'
disjoint: false
disjoint: true
ecommerce, retail, wholesale, all

Answer Area

```
CleanData
split(
    [ ]
    [ ]
    [ ] ~> SplitByDept@([ ])
)
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

The conditional split transformation routes data rows to different streams based on matching conditions. The conditional split transformation is similar to a CASE decision structure in a programming language. The transformation evaluates expressions, and based on the results, directs the data row to the specified stream.

Box 1: dept=='ecommerce', dept=='retail', dept=='wholesale'

First we put the condition. The order must match the stream labeling we define in Box 3. Syntax:

```
<incomingStream> split(
<conditionalExpression1>
<conditionalExpression2> disjoint: {true | false}
) ~> <splitTx>@(stream1, stream2, ..., <defaultStream>)
```

Box 2: discount : false

disjoint is false because the data goes to the first matching condition. All remaining rows matching the third condition go to output stream all.

Box 3: ecommerce, retail, wholesale, all Label the streams

Reference:

<https://docs.microsoft.com/en-us/azure/data-factory/data-flow-conditional-split>

NEW QUESTION 100

- (Exam Topic 3)

You are designing an Azure Synapse Analytics dedicated SQL pool.

You need to ensure that you can audit access to Personally Identifiable information (PII). What should you include in the solution?

- A. dynamic data masking
- B. row-level security (RLS)
- C. sensitivity classifications
- D. column-level security

Answer: C

Explanation:

Data Discovery & Classification is built into Azure SQL Database, Azure SQL Managed Instance, and Azure Synapse Analytics. It provides basic capabilities for discovering, classifying, labeling, and reporting the sensitive data in your databases.

Your most sensitive data might include business, financial, healthcare, or personal information. Discovering and classifying this data can play a pivotal role in your organization's information-protection approach. It can serve as infrastructure for:

- > Helping to meet standards for data privacy and requirements for regulatory compliance.
- > Various security scenarios, such as monitoring (auditing) access to sensitive data.
- > Controlling access to and hardening the security of databases that contain highly sensitive data.

Reference:

<https://docs.microsoft.com/en-us/azure/azure-sql/database/data-discovery-and-classification-overview>

NEW QUESTION 102

- (Exam Topic 3)

You are designing a highly available Azure Data Lake Storage solution that will include geo-zone-redundant storage (GZRS). You need to monitor for replication delays that can affect the recovery point objective (RPO). What should you include in the monitoring solution?

- A. availability
- B. Average Success E2E Latency
- C. 5xx: Server Error errors
- D. Last Sync Time

Answer: D

Explanation:

Because geo-replication is asynchronous, it is possible that data written to the primary region has not yet been written to the secondary region at the time an outage occurs. The Last Sync Time property indicates the last time that data from the primary region was written successfully to the secondary region. All writes made to the primary region before the last sync time are available to be read from the secondary location. Writes made to the primary region after the last sync time property may or may not be available for reads yet.

Reference:
<https://docs.microsoft.com/en-us/azure/storage/common/last-sync-time-get>

NEW QUESTION 105

- (Exam Topic 3)

From a website analytics system, you receive data extracts about user interactions such as downloads, link clicks, form submissions, and video plays. The data contains the following columns.

Name	Sample value
Date	15 Jan 2021
EventCategory	Videos
EventAction	Play
EventLabel	Contoso Promotional
ChannelGrouping	Social
TotalEvents	150
UniqueEvents	120
SessionWithEvents	99

You need to design a star schema to support analytical queries of the data. The star schema will contain four tables including a date dimension. To which table should you add each column? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

EventCategory:

▼

DimChannel

DimDate

DimEvent

FactEvents

ChannelGrouping:

▼

DimChannel

DimDate

DimEvent

FactEvents

TotalEvents:

▼

DimChannel

DimDate

DimEvent

FactEvents

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Table Description automatically generated

Box 1: DimEvent

Box 2: DimChannel

Box 3: FactEvents

Fact tables store observations or events, and can be sales orders, stock balances, exchange rates, temperatures, etc

Reference:

<https://docs.microsoft.com/en-us/power-bi/guidance/star-schema>

NEW QUESTION 106

- (Exam Topic 3)

You have an Azure Synapse Analytics dedicated SQL pool named pool1.

You need to perform a monthly audit of SQL statements that affect sensitive data. The solution must minimize administrative effort.

What should you include in the solution?

- A. Microsoft Defender for SQL
- B. dynamic data masking
- C. sensitivity labels
- D. workload management

Answer: B

NEW QUESTION 111

- (Exam Topic 3)

You are designing a slowly changing dimension (SCD) for supplier data in an Azure Synapse Analytics dedicated SQL pool.

You plan to keep a record of changes to the available fields. The supplier data contains the following columns.

Name	Description
SupplierSystemID	Unique supplier ID in an enterprise resource planning (ERP) system
SupplierName	Name of the supplier company
SupplierAddress1	Address of the supplier company
SupplierAddress2	Second address line of the supplier company
SupplierCity	City of the supplier company
SupplierStateProvince	State or province of the supplier company
SupplierCountry	Country of the supplier company
SupplierPostalCode	Postal code of the supplier company
SupplierDescription	Free-text description of the supplier company
SupplierCategory	Category of goods provided by the supplier company

Which three additional columns should you add to the data to create a Type 2 SCD? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. surrogate primary key
- B. foreign key
- C. effective start date
- D. effective end date
- E. last modified date
- F. business key

Answer: CDE

Explanation:

Reference:

<https://docs.microsoft.com/en-us/sql/integration-services/data-flow/transformations/slowly-changing-dimension>

NEW QUESTION 112

- (Exam Topic 3)

You need to implement a Type 3 slowly changing dimension (SCD) for product category data in an Azure Synapse Analytics dedicated SQL pool.

You have a table that was created by using the following Transact-SQL statement.

```
CREATE TABLE [DBO].[DimProduct] (
    [ProductKey] [int] IDENTITY(1,1) NOT NULL,
    [ProductSourceID] [int] NOT NULL,
    [ProductName] [nvarchar] (100) NULL,
    [Color] [nvarchar] (15) NULL,
    [SellStartDate] [date] NOT NULL,
    [SellEndDate] [date] NULL,
    [RowInsertedDateTime] [datetime] NOT NULL,
    [RowUpdatedDateTime] [datetime] NOT NULL,
    [ETLAuditID] [int] NOT NULL
)
```

Which two columns should you add to the table? Each correct answer presents part of the solution.
 NOTE: Each correct selection is worth one point.

- A. [EffectiveScarcDate] [datetime] NOT NULL,
- B. [CurrentProduccCacegory] [nvarchar] (100) NOT NULL,
- C. [EffectiveEndDace] [dacecime] NULL,
- D. [ProductCategory] [nvarchar] (100) NOT NULL,
- E. [OriginalProduccCacegory] [nvarchar] (100) NOT NULL,

Answer: BE

Explanation:

A Type 3 SCD supports storing two versions of a dimension member as separate columns. The table includes a column for the current value of a member plus either the original or previous value of the member. So Type 3 uses additional columns to track one key instance of history, rather than storing additional rows to track each change like in a Type 2 SCD.

This type of tracking may be used for one or two columns in a dimension table. It is not common to use it for many members of the same table. It is often used in combination with Type 1 or Type 2 members.

Graphical user interface, application, email Description automatically generated



CustomerID	FirstName	LastName	CurrentEmail	OriginalEmail	CompanyName	InsertedDate	ModifiedDate
2	Keith	Harris	keith0@aw.com	keith0@aw.com	Progressive Sports	2021-03-20	2021-03-20
3	Donna	Carreras	donna0@aw.com	donna0@aw.com	A Bike Store	2021-03-20	2021-03-20

CustomerID	FirstName	LastName	CurrentEmail	OriginalEmail	CompanyName	InsertedDate	ModifiedDate
2	Keith	Harris	keith0@aw.com	keith0@aw.com	Progressive Sports	2021-03-20	2021-03-20
3	Donna	Carreras	dc3@aw.com	donna0@aw.com	A Bike Store	2021-03-20	2021-03-22

Reference:

<https://k21academy.com/microsoft-azure/azure-data-engineer-dp203-q-a-day-2-live-session-review/>

NEW QUESTION 114

- (Exam Topic 3)

You plan to implement an Azure Data Lake Gen2 storage account.

You need to ensure that the data lake will remain available if a data center fails in the primary Azure region. The solution must minimize costs.

Which type of replication should you use for the storage account?

- A. geo-redundant storage (GRS)
- B. zone-redundant storage (ZRS)
- C. locally-redundant storage (LRS)
- D. geo-zone-redundant storage (GZRS)

Answer: C

Explanation:

Locally redundant storage (LRS) copies your data synchronously three times within a single physical location in the primary region. LRS is the least expensive replication option

Reference:

<https://docs.microsoft.com/en-us/azure/storage/common/storage-redundancy>

NEW QUESTION 118

- (Exam Topic 3)

You have an Azure Databricks workspace and an Azure Data Lake Storage Gen2 account named storage1. New files are uploaded daily to storage1.

- Incrementally process new files as they are upkorage1 as a structured streaming source. The solution must meet the following requirements:
- Minimize implementation and maintenance effort.
- Minimize the cost of processing millions of files.
- Support schema inference and schema drift. Which should you include in the recommendation?

- A. Auto Loader
- B. Apache Spark FileStreamSource
- C. COPY INTO
- D. Azure Data Factory

Answer: D

NEW QUESTION 123

- (Exam Topic 3)

You are designing a dimension table in an Azure Synapse Analytics dedicated SQL pool.

You need to create a surrogate key for the table. The solution must provide the fastest query performance. What should you use for the surrogate key?

- A. a GUID column
- B. a sequence object
- C. an IDENTITY column

Answer: C

Explanation:

Use IDENTITY to create surrogate keys using dedicated SQL pool in AzureSynapse Analytics.

Note: A surrogate key on a table is a column with a unique identifier for each row. The key is not generated from the table data. Data modelers like to create surrogate keys on their tables when they design data warehouse models. You can use the IDENTITY property to achieve this goal simply and effectively without affecting load performance.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-identity>

NEW QUESTION 128

- (Exam Topic 3)

You have an Azure Data Lake Storage Gen 2 account named storage1.

You need to recommend a solution for accessing the content in storage1. The solution must meet the following requirements:

- > List and read permissions must be granted at the storage account level.
- > Additional permissions can be applied to individual objects in storage1.
- > Security principals from Microsoft Azure Active Directory (Azure AD), part of Microsoft Entra, must be used for authentication.

What should you use? To answer, drag the appropriate components to the correct requirements. Each component may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Components

Access control lists (ACLs)

Role-based access control (RBAC) roles

Shared access signatures (SAS)

Shared account keys

Answer Area

To grant permissions at the storage account level:

To grant permissions at the object level:

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Role-based access control (RBAC) roles

List and read permissions must be granted at the storage account level.

Security principals from Microsoft Azure Active Directory (Azure AD), part of Microsoft Entra, must be used for authentication.

Role-based access control (Azure RBAC)

Azure RBAC uses role assignments to apply sets of permissions to security principals. A security principal is an object that represents a user, group, service principal, or managed identity that is defined in Azure Active Directory (AD). A permission set can give a security principal a "coarse-grain" level of access such as read or write access to all of the data in a storage account or all of the data in a container.

Box 2: Access control lists (ACLs)

Additional permissions can be applied to individual objects in storage1. Access control lists (ACLs)

ACLs give you the ability to apply "finer grain" level of access to directories and files. An ACL is a permission construct that contains a series of ACL entries. Each ACL entry associates security principal with an access level.

Reference: <https://learn.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-access-control-model>

NEW QUESTION 131

- (Exam Topic 3)

You store files in an Azure Data Lake Storage Gen2 container. The container has the storage policy shown in the following exhibit.


```
{
  "rules": [
    {
      "enabled": true,
      "name": "contosorule",
      "type": "Lifecycle",
      "definition": {
        "actions": {
          "version": {
            "delete": {
              "daysAfterCreationGreaterThan": 60
            }
          },
          "baseBlob": {
            "tierToCool": {
              "daysAfterModificationGreaterThan": 30
            }
          }
        },
        "filters": {
          "blobTypes": [
            "blockBlob"
          ],
          "prefixMatch": [
            "container1/contoso"
          ]
        }
      }
    }
  ]
}
```

Use the drop-down menus to select the answer choice that completes each statement based on the information presented in the graphic.
NOTE: Each correct selection is worth one point.

The files are [answer choice] after 30 days:
 [
 deleted from the container
 moved to archive storage
 moved to cool storage
 moved to hot storage
]

The storage policy applies to [answer choice]:
 [
 container1/contoso.csv
 container1/docs/contoso.json
 container1/mycontoso/contoso.csv
]

- A. Mastered
- B. Not Mastered

Answer: A

Explanation: Graphical user interface, text, application Description automatically generated
Box 1: moved to cool storage
The ManagementPolicyBaseBlob.TierToCool property gets or sets the function to tier blobs to cool storage. Support blobs currently at Hot tier.
Box 2: container1/contoso.csv As defined by prefixMatch.
prefixMatch: An array of strings for prefixes to be matched. Each rule can define up to 10 case-senstive prefixes. A prefix string must start with a container name.
Reference: https://docs.microsoft.com/en-us/dotnet/api/microsoft.azure.management.storage.fluent.models.managementpoli

NEW QUESTION 134

- (Exam Topic 3)
You need to output files from Azure Data Factory.
Which file format should you use for each type of output? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Columnar format:

	▼
Avro	
GZip	
Parquet	
TXT	

JSON with a timestamp:

	▼
Avro	
GZip	
Parquet	
TXT	

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Parquet

Parquet stores data in columns, while Avro stores data in a row-based format. By their very nature, column-oriented data stores are optimized for read-heavy analytical workloads, while row-based databases are best for write-heavy transactional workloads.

Box 2: Avro

An Avro schema is created using JSON format. AVRO supports timestamps.

Note: Azure Data Factory supports the following file formats (not GZip or TXT).

- > Avro format
- > Binary format
- > Delimited text format
- > Excel format
- > JSON format
- > ORC format
- > Parquet format
- > XML format

Reference:

<https://www.datanami.com/2018/05/16/big-data-file-formats-demystified>

NEW QUESTION 135

- (Exam Topic 3)

You have an Azure Synapse Analytics dedicated SQL pool named Pool1. Pool1 contains a table named table1. You load 5 TB of data into table1. You need to ensure that columnstore compression is maximized for table1. Which statement should you execute?

- A. ALTER INDEX ALL on table1 REORGANIZE
- B. ALTER INDEX ALL on table1 REBUILD
- C. DBCC DBREINDEX (table1)
- D. DBCC INDEXDEFRAG (pool1,table1)

Answer: B

Explanation:

Columnstore and columnstore archive compression

Columnstore tables and indexes are always stored with columnstore compression. You can further reduce the size of columnstore data by configuring an additional compression called archival compression. To perform archival compression, SQL Server runs the Microsoft XPRESS compression algorithm on the data. Add or remove archival compression by using the following data compression types:

Use COLUMNSTORE_ARCHIVE data compression to compress columnstore data with archival compression.

Use COLUMNSTORE data compression to decompress archival compression. The resulting data continue to be compressed with columnstore compression. To add archival compression, use ALTER TABLE (Transact-SQL) or ALTER INDEX (Transact-SQL) with the REBUILD option and DATA COMPRESSION = COLUMNSTORE_ARCHIVE.

Reference: <https://learn.microsoft.com/en-us/sql/relational-databases/data-compression/data-compression>

NEW QUESTION 138

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You have an Azure Synapse Analytics dedicated SQL pool that contains a table named Table1. You have files that are ingested and loaded into an Azure Data Lake Storage Gen2 container named container1.

You plan to insert data from the files in container1 into Table1 and transform the data. Each row of data in the files will produce one row in the serving layer of Table1.

You need to ensure that when the source data files are loaded to container1, the DateTime is stored as an additional column in Table1.
 Solution: You use an Azure Synapse Analytics serverless SQL pool to create an external table that has an additional DateTime column.
 Does this meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Instead use the derived column transformation to generate new columns in your data flow or to modify existing fields.

Reference:

<https://docs.microsoft.com/en-us/azure/data-factory/data-flow-derived-column>

NEW QUESTION 139

- (Exam Topic 3)

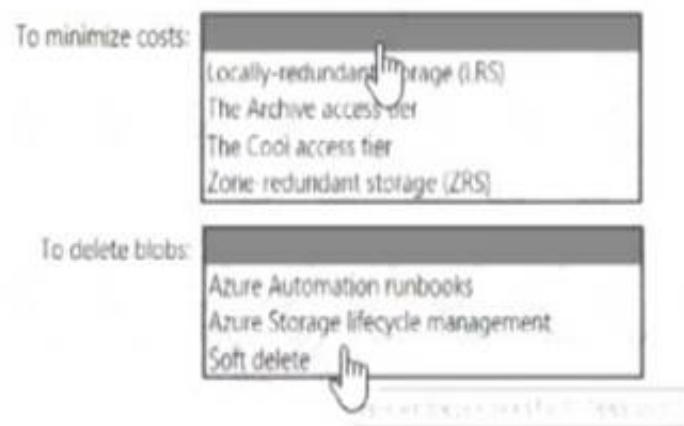
You have an Azure subscription.

You need to deploy an Azure Data Lake Storage Gen2 Premium account. The solution must meet the following requirements:

- Blobs that are older than 365 days must be deleted.
- Administrator efforts must be minimized.
- Costs must be minimized

What should you use? To answer, select the appropriate options in the answer area. NOTE Each correct selection is worth one point.

Answer Area



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

<https://learn.microsoft.com/en-us/azure/storage/blobs/premium-tier-for-data-lake-storage>

NEW QUESTION 141

- (Exam Topic 3)

You have an Azure Data Factory version 2 (V2) resource named Df1. Df1 contains a linked service. You have an Azure Key vault named vault1 that contains an encryption key named key1.

You need to encrypt Df1 by using key1. What should you do first?

- A. Add a private endpoint connection to vault 1.
- B. Enable Azure role-based access control on vault 1.
- C. Remove the linked service from Df1.
- D. Create a self-hosted integration runtime.

Answer: C

Explanation:

Linked services are much like connection strings, which define the connection information needed for Data Factory to connect to external resources.

Reference:

<https://docs.microsoft.com/en-us/azure/data-factory/enable-customer-managed-key> <https://docs.microsoft.com/en-us/azure/data-factory/concepts-linked-services>
<https://docs.microsoft.com/en-us/azure/data-factory/create-self-hosted-integration-runtime>

NEW QUESTION 145

- (Exam Topic 3)

You are building an Azure Analytics query that will receive input data from Azure IoT Hub and write the results to Azure Blob storage.

You need to calculate the difference in readings per sensor per hour.

How should you complete the query? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

```
SELECT sensorId,  
       growth = reading -  
           (reading) OVER (PARTITION BY sensorId  
                           LAG  
                           LAST  
                           LEAD  
                           (hour, 1))  
FROM input
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: LAG

The LAG analytic operator allows one to look up a “previous” event in an event stream, within certain constraints. It is very useful for computing the rate of growth of a variable, detecting when a variable crosses a threshold, or when a condition starts or stops being true.

Box 2: LIMIT DURATION

Example: Compute the rate of growth, per sensor: SELECT sensorId,
growth = reading
LAG(reading) OVER (PARTITION BY sensorId LIMIT DURATION(hour, 1)) FROM input

Reference:

<https://docs.microsoft.com/en-us/stream-analytics-query/lag-azure-stream-analytics>

NEW QUESTION 148

- (Exam Topic 3)

You have an Azure Synapse Analytics dedicated SQL pool named Pool1 and a database named DB1. DB1 contains a fact table named Table1. You need to identify the extent of the data skew in Table1. What should you do in Synapse Studio?

- A. Connect to the built-in pool and query sysdm_pdw_sys_info.
- B. Connect to Pool1 and run DBCC CHECKALLOC.
- C. Connect to the built-in pool and run DBCC CHECKALLOC.
- D. Connect to Pool1 and query sys.dm_pdw_nodes_db_partition_stats.

Answer: D

Explanation:

Microsoft recommends use of sys.dm_pdw_nodes_db_partition_stats to analyze any skewness in the data. Reference:
<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/cheat-sheet>

NEW QUESTION 153

- (Exam Topic 3)

You are designing an Azure Data Lake Storage solution that will transform raw JSON files for use in an analytical workload. You need to recommend a format for the transformed files. The solution must meet the following requirements:

- > Contain information about the data types of each column in the files.
- > Support querying a subset of columns in the files.
- > Support read-heavy analytical workloads.
- > Minimize the file size.

What should you recommend?

- A. JSON
- B. CSV
- C. Apache Avro
- D. Apache Parquet

Answer: D

Explanation:

Parquet, an open-source file format for Hadoop, stores nested data structures in a flat columnar format. Compared to a traditional approach where data is stored in a row-oriented approach, Parquet file format is more efficient in terms of storage and performance.

It is especially good for queries that read particular columns from a “wide” (with many columns) table since only needed columns are read, and IO is minimized.

Reference: <https://www.clairvoyant.ai/blog/big-data-file-formats>

NEW QUESTION 158

- (Exam Topic 3)

You have an Azure Synapse serverless SQL pool.

You need to read JSON documents from a file by using the OPENROWSET function.

How should you complete the query? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Answer Area

```
SELECT *
FROM OPENROWSET
(
    BULK
    'https://sourcedatalake.blob.core.windows.net/public/docs.json',
    FORMAT = 'JSON',
    FIELDTERMINATOR = '0x0b',
    FIELDQUOTE = '0x0b',
    ROWTERMINATOR = '0x0a'
)
WITH (jsondoc nvarchar(1024) NOT NULL) as jsonDocuments
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:
Answer Area

```
SELECT *
FROM OPENROWSET
(
    BULK
    'https://sourcedatalake.blob.core.windows.net/public/docs.json',
    FORMAT = 'JSON',
    FIELDTERMINATOR = '0x0b',
    FIELDQUOTE = '0x0b',
    ROWTERMINATOR = '0x0a'
)
WITH (jsondoc nvarchar(1024) NOT NULL) as jsonDocuments
```

NEW QUESTION 163

- (Exam Topic 3)

You have an Azure Synapse Analytics SQL pool named Pool1 on a logical Microsoft SQL server named Server1. You need to implement Transparent Data Encryption (TDE) on Pool1 by using a custom key named key1. Which five actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

Actions

- Enable TDE on Pool1.
- Assign a managed identity to Server1.
- Configure key1 as the TDE protector for Server1.
- Add key1 to the Azure key vault.
- Create an Azure key vault and grant the managed identity permissions to the key vault.

Answer Area



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Graphical user interface, text, application Description automatically generated
Step 1: Assign a managed identity to Server1
You will need an existing Managed Instance as a prerequisite.
Step 2: Create an Azure key vault and grant the managed identity permissions to the vault Create Resource and setup Azure Key Vault.
Step 3: Add key1 to the Azure key vault
The recommended way is to import an existing key from a .pfx file or get an existing key from the vault. Alternatively, generate a new key directly in Azure Key Vault.
Step 4: Configure key1 as the TDE protector for Server1 Provide TDE Protector key
Step 5: Enable TDE on Pool1 Reference:
<https://docs.microsoft.com/en-us/azure/azure-sql/managed-instance/scripts/transparent-data-encryption-byok-po>

NEW QUESTION 168

- (Exam Topic 3)
You have an Azure Storage account and a data warehouse in Azure Synapse Analytics in the UK South region. You need to copy blob data from the storage account to the data warehouse by using Azure Data Factory. The solution must meet the following requirements:
> Ensure that the data remains in the UK South region at all times.
> Minimize administrative effort.
Which type of integration runtime should you use?

- A. Azure integration runtime
- B. Azure-SSIS integration runtime
- C. Self-hosted integration runtime

Answer: A

Explanation:

IR type	Public network	Private network
Azure	Data Flow Data movement Activity dispatch	
Self-hosted	Data movement Activity dispatch	Data movement Activity dispatch
Azure-SSIS	SSIS package execution	SSIS package execution

Reference:
<https://docs.microsoft.com/en-us/azure/data-factory/concepts-integration-runtime>

NEW QUESTION 169

- (Exam Topic 3)
You have an Azure Synapse Analytics dedicated SQL Pool1. Pool1 contains a partitioned fact table named dbo.Sales and a staging table named stg.Sales that has the matching table and partition definitions.
You need to overwrite the content of the first partition in dbo.Sales with the content of the same partition in stg.Sales. The solution must minimize load times.
What should you do?
A. Switch the first partition from dbo.Sales to stg.Sales.
B. Switch the first partition from stg.Sales to db
C. Sales.
D. Update dbo.Sales from stg.Sales.
E. Insert the data from stg.Sales into dbo.Sales.

Answer: A

NEW QUESTION 173

- (Exam Topic 3)
You have an Azure subscription linked to an Azure Active Directory (Azure AD) tenant that contains a service principal named ServicePrincipal1. The subscription contains an Azure Data Lake Storage account named adls1. Adls1 contains a folder named Folder2 that has a URI of <https://adls1.dfs.core.windows.net/container1/Folder1/Folder2/>.
ServicePrincipal1 has the access control list (ACL) permissions shown in the following table.

Resource	Permission
container1	Access – Execute
Folder1	Access – Execute
Folder2	Access – Read

You need to ensure that ServicePrincipal1 can perform the following actions:

- Traverse child items that are created in Folder2.
- Read files that are created in Folder2.

The solution must use the principle of least privilege.

Which two permissions should you grant to ServicePrincipal1 for Folder2? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Access - Read
- B. Access - Write
- C. Access - Execute
- D. Default-Read
- E. Default - Write
- F. Default - Execute

Answer: DF

Explanation:

Execute (X) permission is required to traverse the child items of a folder.

There are two kinds of access control lists (ACLs), Access ACLs and Default ACLs. Access ACLs: These control access to an object. Files and folders both have Access ACLs.

Default ACLs: A "template" of ACLs associated with a folder that determine the Access ACLs for any child items that are created under that folder. Files do not have Default ACLs.

Reference:

<https://docs.microsoft.com/en-us/azure/data-lake-store/data-lake-store-access-control>

NEW QUESTION 177

- (Exam Topic 3)

You need to trigger an Azure Data Factory pipeline when a file arrives in an Azure Data Lake Storage Gen2 container.

Which resource provider should you enable?

- A. Microsoft.Sql
- B. Microsoft-Automation
- C. Microsoft.EventGrid
- D. Microsoft.EventHub

Answer: C

Explanation:

Event-driven architecture (EDA) is a common data integration pattern that involves production, detection, consumption, and reaction to events. Data integration scenarios often require Data Factory customers to trigger pipelines based on events happening in storage account, such as the arrival or deletion of a file in Azure Blob Storage account. Data Factory natively integrates with Azure Event Grid, which lets you trigger pipelines on such events.

Reference:

<https://docs.microsoft.com/en-us/azure/data-factory/how-to-create-event-trigger> <https://docs.microsoft.com/en-us/azure/data-factory/concepts-pipeline-execution-triggers>

NEW QUESTION 179

- (Exam Topic 3)

You plan to create a table in an Azure Synapse Analytics dedicated SQL pool.

Data in the table will be retained for five years. Once a year, data that is older than five years will be deleted. You need to ensure that the data is distributed evenly across partitions. The solution must minimize the amount of time required to delete old data.

How should you complete the Transact-SQL statement? To answer, drag the appropriate values to the correct targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Values	Answer Area
CustomerKey	CREATE TABLE [dbo].[FactSales]
HASH	(
ROUND_ROBIN	[ProductKey] int NOT NULL
REPLICATE	, [OrderDateKey] int NOT NULL
OrderDateKey	, [CustomerKey] int NOT NULL
SalesOrderNumber	, [SalesOrderNumber] nvarchar (20) NOT NULL
	, [OrderQuantity] smallint NOT NULL
	, [UnitPrice] money NOT NULL
)
	WITH
	(CLUSTERED COLUMNSTORE INDEX
	, DISTRIBUTION = Value ([ProductKey])
	, PARTITION ([Value] RANGE RIGHT FOR VALUES
	(20170101,20180101,20190101,20200101,20210101)
)
)

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: HASH

Box 2: OrderDateKey

In most cases, table partitions are created on a date column.

A way to eliminate rollbacks is to use Metadata Only operations like partition switching for data management. For example, rather than execute a DELETE statement to delete all rows in a table where the order_date was in October of 2001, you could partition your data early. Then you can switch out the partition with data for an empty partition from another table.

Reference:

<https://docs.microsoft.com/en-us/sql/t-sql/statements/create-table-azure-sql-data-warehouse> <https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/best-practices-dedicated-sql-pool>

NEW QUESTION 182

- (Exam Topic 3)

You have an enterprise data warehouse in Azure Synapse Analytics.

Using PolyBase, you create an external table named [Ext].[Items] to query Parquet files stored in Azure Data Lake Storage Gen2 without importing the data to the data warehouse.

The external table has three columns.

You discover that the Parquet files have a fourth column named ItemID.

Which command should you run to add the ItemID column to the external table?

- A. `ALTER EXTERNAL TABLE [Ext].[Items]
ADD [ItemID] int;`
- B. `DROP EXTERNAL FILE FORMAT parquetfile1;
CREATE EXTERNAL FILE FORMAT parquetfile1
WITH (
 FORMAT_TYPE = PARQUET,
 DATA_COMPRESSION = 'org.apache.hadoop.io.compress.SnappyCodec'
);`
- C. `DROP EXTERNAL TABLE [Ext].[Items]
CREATE EXTERNAL TABLE [Ext].[Items]
([ItemID] [int] NULL,
 [ItemName] nvarchar(50) NULL,
 [ItemType] nvarchar(20) NULL,
 [ItemDescription] nvarchar(250))
WITH
(
 LOCATION= '/Items/',
 DATA_SOURCE = AzureDataLakeStore,
 FILE_FORMAT = PARQUET,
 REJECT_TYPE = VALUE,
 REJECT_VALUE = 0
);`
- D. `ALTER TABLE [Ext].[Items]
ADD [ItemID] int;`

A. Option A

B. Option B

C. Option C

D. Option D

Answer: C

Explanation:

<https://docs.microsoft.com/en-us/sql/t-sql/statements/create-external-table-transact-sql>

NEW QUESTION 183

- (Exam Topic 3)

You have an Azure subscription that is linked to a hybrid Azure Active Directory (Azure AD) tenant. The subscription contains an Azure Synapse Analytics SQL pool named Pool1.

You need to recommend an authentication solution for Pool1. The solution must support multi-factor authentication (MFA) and database-level authentication.

Which authentication solution or solutions should you include in the recommendation? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

MFA:

▼

Azure AD authentication
Microsoft SQL Server authentication
Passwordless authentication
Windows authentication

Database-level authentication:

▼

Application roles
Contained database users
Database roles
Microsoft SQL Server logins

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:
 Graphical user interface, text, application, chat or text message Description automatically generated
 Box 1: Azure AD authentication
 Azure Active Directory authentication supports Multi-Factor authentication through Active Directory Universal Authentication.
 Box 2: Contained database users
 Azure Active Directory Uses contained database users to authenticate identities at the database level. Reference:
<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-authentication>

NEW QUESTION 184

- (Exam Topic 3)
 You are designing an Azure Synapse Analytics dedicated SQL pool.
 Groups will have access to sensitive data in the pool as shown in the following table.

Name	Enhanced access
Executives	No access to sensitive data
Analysts	Access to in-region sensitive data
Engineers	Access to all numeric sensitive data

You have policies for the sensitive data. The policies vary be region as shown in the following table.

Region	Data considered sensitive
RegionA	Financial, Personally Identifiable Information (PII)
RegionB	Financial, Personally Identifiable Information (PII), medical
RegionC	Financial, medical

You have a table of patients for each region. The tables contain the following potentially sensitive columns.

Name	Sensitive data	Description
CardOnFile	Financial	Debit/credit card number for charges
Height	Medical	Patient’s height in cm
ContactEmail	PII	Email address for secure communications

You are designing dynamic data masking to maintain compliance.
 For each of the following statements, select Yes if the statement is true. Otherwise, select No.
 NOTE: Each correct selection is worth one point.

Statements	Yes	No
Analysts in RegionA require dynamic data masking rules for [Patients_RegionA].	<input type="radio"/>	<input type="radio"/>
Engineers in RegionC require a dynamic data masking rule for [Patients_RegionA], [Height]	<input type="radio"/>	<input type="radio"/>
Engineers in RegionB require a dynamic data masking rule for [Patients_RegionB], [Height]	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Text Description automatically generated

Reference:

<https://docs.microsoft.com/en-us/azure/azure-sql/database/dynamic-data-masking-overview>

NEW QUESTION 185

- (Exam Topic 3)

You are designing a security model for an Azure Synapse Analytics dedicated SQL pool that will support multiple companies. You need to ensure that users from each company can view only the data of their respective company. Which two objects should you include in the solution? Each correct answer presents part of the solution

NOTE: Each correct selection it worth one point.

- A. a custom role-based access control (RBAC) role.
- B. asymmetric keys
- C. a predicate function
- D. a column encryption key
- E. a security policy

Answer: AE

Explanation:

Reference:

<https://docs.microsoft.com/en-us/sql/relational-databases/security/row-level-security> <https://docs.microsoft.com/en-us/azure/synapse-analytics/security/synapse-workspace-access-control-overview>

NEW QUESTION 190

- (Exam Topic 3)

You have a SQL pool in Azure Synapse.

You discover that some queries fail or take a long time to complete. You need to monitor for transactions that have rolled back.

Which dynamic management view should you query?

- A. sys.dm_pdw_request_steps
- B. sys.dm_pdw_nodes_tran_database_transactions
- C. sys.dm_pdw_waits
- D. sys.dm_pdw_exec_sessions

Answer: B

Explanation:

You can use Dynamic Management Views (DMVs) to monitor your workload including investigating query execution in SQL pool.

If your queries are failing or taking a long time to proceed, you can check and monitor if you have any transactions rolling back.

Example:

```
-- Monitor rollback SELECT
SUM(CASE WHEN t.database_transaction_next_undo_lsn IS NOT NULL THEN 1 ELSE 0 END), t.pdw_node_id,
nod.[type]
FROM sys.dm_pdw_nodes_tran_database_transactions t
JOIN sys.dm_pdw_nodes nod ON t.pdw_node_id = nod.pdw_node_id GROUP BY t.pdw_node_id, nod.[type]
```

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-manage-monit>

NEW QUESTION 195

- (Exam Topic 3)

You plan to create a dimension table in Azure Synapse Analytics that will be less than 1 GB. You need to create the table to meet the following requirements:

- Provide the fastest Query time.
- Minimize data movement during queries. Which type of table should you use?

- A. hash distributed
- B. heap
- C. replicated
- D. round-robin

Answer: C

Explanation:

A replicated table has a full copy of the table accessible on each Compute node. Replicating a table removes the need to transfer data among Compute nodes before a join or aggregation. Since the table has multiple copies, replicated tables work best when the table size is less than 2 GB compressed. 2 GB is not a hard limit.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/design-guidance-for-replicated-tab>

NEW QUESTION 199

- (Exam Topic 3)

You have an Apache Spark DataFrame named temperatures. A sample of the data is shown in the following table.

Date	Temp
...	...
18-01-2021	3
19-01-2021	4
20-01-2021	2
21-01-2021	2
...	...

You need to produce the following table by using a Spark SQL query.

Year	JAN	FEB	MAR	APR	MAY
2019	2.3	4.1	5.2	7.6	9.2
2020	2.4	4.2	4.9	7.8	9.1
2021	2.6	5.3	3.4	7.9	9.5

How should you complete the query? To answer, drag the appropriate values to the correct targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Values

Answer Area

CAST

COLLATE

CONVERT

FLATTEN

PIVOT

UNPIVOT

```
SELECT * FROM (
  SELECT YEAR(Date) Year, MONTH(Date) Month, Temp
  FROM temperatures
  WHERE date BETWEEN DATE '2019-01-01' AND DATE '2021-08-31'
)
  (
    AVG ( (Temp AS DECIMAL(4, 1)))
  )
  FOR Month in (
    1 JAN, 2 FEB, 3 MAR, 4 APR, 5 MAY, 6 JUN,
    7 JUL, 8 AUG, 9 SEP, 10 OCT, 11 NOV, 12 DEC
  )
)
ORDER BY Year ASC
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Text Description automatically generated

Box 1: PIVOT

PIVOT rotates a table-valued expression by turning the unique values from one column in the expression into multiple columns in the output. And PIVOT runs aggregations where they're required on any remaining column values that are wanted in the final output.

Reference:

<https://learnsql.com/cookbook/how-to-convert-an-integer-to-a-decimal-in-sql-server/> <https://docs.microsoft.com/en-us/sql/t-sql/queries/from-using-pivot-and-unpivot>

NEW QUESTION 204

- (Exam Topic 3)

A company uses Azure Stream Analytics to monitor devices.

The company plans to double the number of devices that are monitored.

You need to monitor a Stream Analytics job to ensure that there are enough processing resources to handle the additional load.

Which metric should you monitor?

- A. Early Input Events
- B. Late Input Events
- C. Watermark delay
- D. Input Deserialization Errors

Answer: A

Explanation:

There are a number of resource constraints that can cause the streaming pipeline to slow down. The watermark delay metric can rise due to:

- > Not enough processing resources in Stream Analytics to handle the volume of input events.
- > Not enough throughput within the input event brokers, so they are throttled.
- > Output sinks are not provisioned with enough capacity, so they are throttled. The possible solutions vary widely based on the flavor of output service being used.

Reference:
<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-time-handling>

NEW QUESTION 206

- (Exam Topic 3)

You have an Azure data factory named ADF1.

You currently publish all pipeline authoring changes directly to ADF1.

You need to implement version control for the changes made to pipeline artifacts. The solution must ensure that you can apply version control to the resources currently defined in the UX Authoring canvas for ADF1.

Which two actions should you perform? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Create an Azure Data Factory trigger
- B. From the UX Authoring canvas, select Set up code repository
- C. Create a GitHub action
- D. From the Azure Data Factory Studio, run Publish All.
- E. Create a Git repository
- F. From the UX Authoring canvas, select Publish

Answer: DE

Explanation:

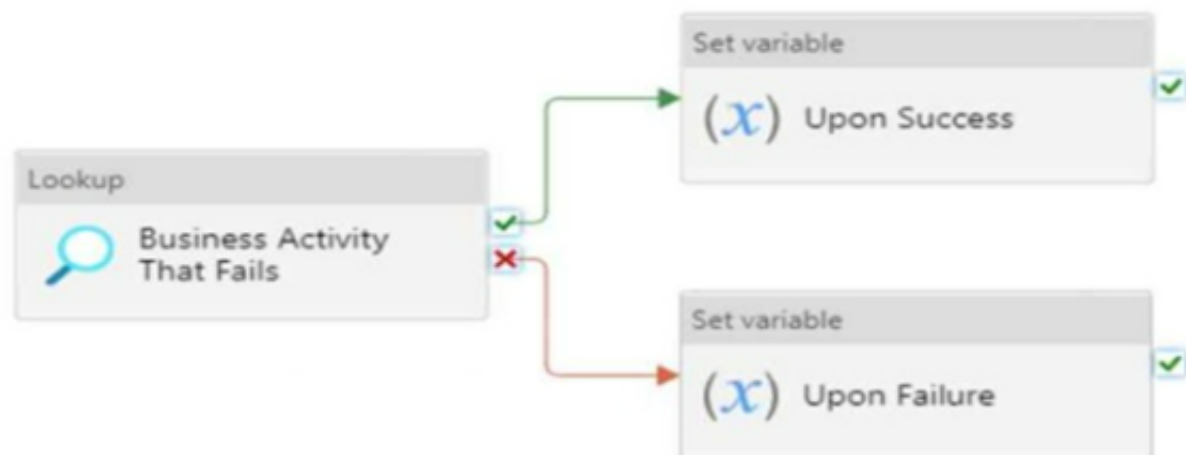
Reference:

<https://docs.microsoft.com/en-us/azure/data-factory/source-control>

NEW QUESTION 209

- (Exam Topic 3)

You have the Azure Synapse Analytics pipeline shown in the following exhibit.



You need to add a set variable activity to the pipeline to ensure that after the pipeline's completion, the status of the pipeline is always successful. What should you configure for the set variable activity?

- A. a success dependency on the Business Activity That Fails activity
- B. a failure dependency on the Upon Failure activity
- C. a skipped dependency on the Upon Success activity
- D. a skipped dependency on the Upon Failure activity

Answer: A

Explanation:

A failure dependency means that the activity will run only if the previous activity fails. In this case, setting a failure dependency on the Upon Failure activity will ensure that the set variable activity will run after the pipeline fails and set the status of the pipeline to successful.

NEW QUESTION 212

- (Exam Topic 3)

You have an Azure subscription that contains an Azure Synapse Analytics dedicated SQL pool. You plan to deploy a solution that will analyze sales data and include the following:

- A table named Country that will contain 195 rows
- A table named Sales that will contain 100 million rows
- A query to identify total sales by country and customer from the past 30 days

You need to create the tables. The solution must maximize query performance.

How should you complete the script? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Answer Area

```
CREATE TABLE [dbo].[Sales]
(
    [OrderDate] date NOT NULL
,   [CustomerId] int NOT NULL
,   [CountryId] int NOT NULL
,   [Total] money NOT NULL
)
WITH
(
    DISTRIBUTION = HASH([CustomerId])
    CLUSTERED COLUMNSTORE INDEX
    REPLICATE
    ROUND_ROBIN
)
CREATE TABLE [dbo].[Country]
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

```
CREATE TABLE [dbo].[Sales]
(
    [OrderDate] date NOT NULL
,   [CustomerId] int NOT NULL
,   [CountryId] int NOT NULL
,   [Total] money NOT NULL
)
WITH
(
    DISTRIBUTION = HASH([CustomerId])
    CLUSTERED COLUMNSTORE INDEX
    REPLICATE
    ROUND_ROBIN
)
CREATE TABLE [dbo].[Country]
```

NEW QUESTION 214

- (Exam Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this scenario, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You have an Azure Storage account that contains 100 GB of files. The files contain text and numerical values. 75% of the rows contain description data that has an average length of 1.1 MB.

You plan to copy the data from the storage account to an Azure SQL data warehouse. You need to prepare the files to ensure that the data copies quickly.

Solution: You modify the files to ensure that each row is more than 1 MB. Does this meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

Instead modify the files to ensure that each row is less than 1 MB. References:

<https://docs.microsoft.com/en-us/azure/sql-data-warehouse/guidance-for-loading-data>

NEW QUESTION 218

- (Exam Topic 3)

You have a Microsoft SQL Server database that uses a third normal form schema.
 You plan to migrate the data in the database to a star schema in an Azure Synapse Analytics dedicated SQL pool.
 You need to design the dimension tables. The solution must optimize read operations.
 What should you include in the solution? to answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Transform data for the dimension tables by:

Maintaining to a third normal form
 Normalizing to a fourth normal form
 Denormalizing to a second normal form

For the primary key columns in the dimension tables, use:

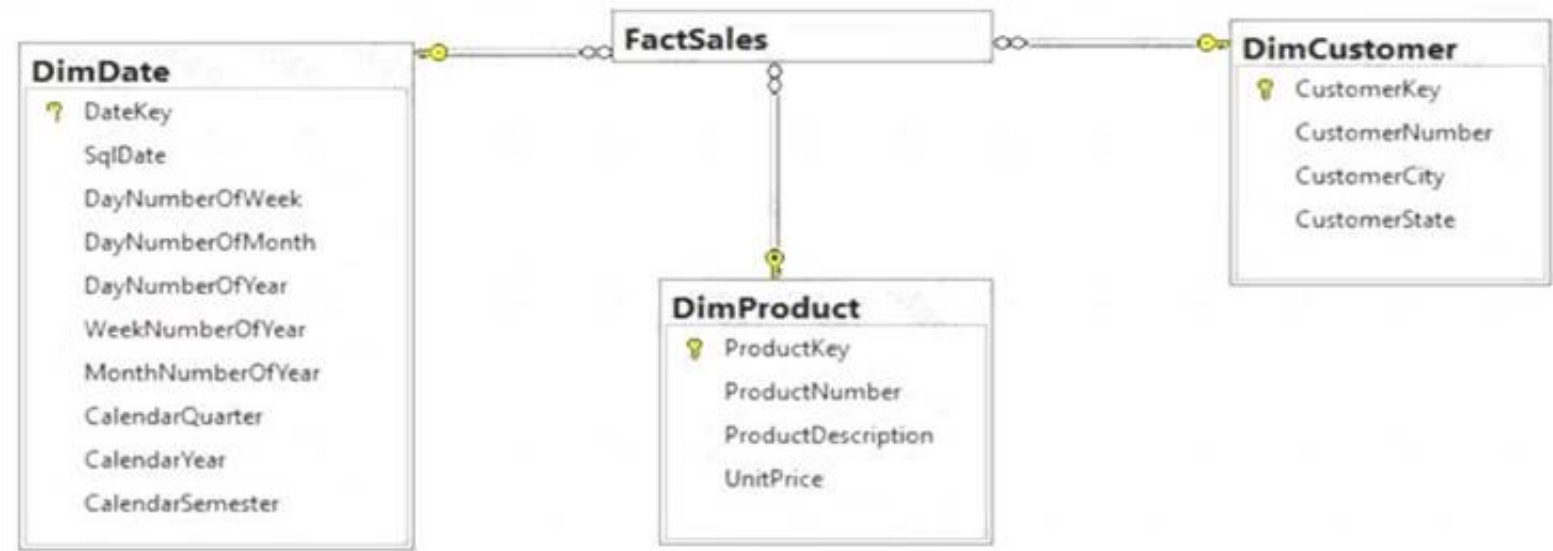
New IDENTITY columns
 A new computed column
 The business key column from the source sys

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Text, table Description automatically generated
 Box 1: Denormalize to a second normal form
 Denormalization is the process of transforming higher normal forms to lower normal forms via storing the join of higher normal form relations as a base relation. Denormalization increases the performance in data retrieval at cost of bringing update anomalies to a database.
 Box 2: New identity columns
 The collapsing relations strategy can be used in this step to collapse classification entities into component entities to obtain at dimension tables with single-part keys that connect directly to the fact table. The single-part key is a surrogate key generated to ensure it remains unique over time.
 Example:
 Diagram Description automatically generated



Note: A surrogate key on a table is a column with a unique identifier for each row. The key is not generated from the table data. Data modelers like to create surrogate keys on their tables when they design data warehouse models. You can use the IDENTITY property to achieve this goal simply and effectively without affecting load performance.
 Reference:
<https://www.mssqltips.com/sqlservertip/5614/explore-the-role-of-normal-forms-in-dimensional-modeling/> <https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-identity>

NEW QUESTION 223

- (Exam Topic 3)
 You have an Azure subscription that contains an Azure Synapse Analytics workspace named ws1 and an Azure Cosmos D6 database account named Cosmos1
 Costmos1 contains a container named container 1 and ws1 contains a serverless1 SQL pool.
 you need to ensure that you can Query the data in container by using the serverless1 SQL pool. Which three actions should you perform? Each correct answer presents part of the solution NOTE: Each correct selection is worth one point.

- A. Enable Azure Synapse Link for Cosmos1
- B. Disable the analytical store for container1.
- C. In ws1. create a linked service that references Cosmos1
- D. Enable the analytical store for container1
- E. Disable indexing for container1

Answer: ACD

NEW QUESTION 227

- (Exam Topic 3)
 You have an Azure Synapse Analytics dedicated SQL pool.
 You need to monitor the database for long-running queries and identify which queries are waiting on resources Which dynamic management view should you use for each requirement? To answer, select the appropriate options in the answer area.
 NOTE; Each correct answer is worth one point.

Answer Area

Monitor the database for long-running queries:

▼

sys.dm_pdw_exec_requests
sys.dm_pdw_sql_requests
sys.dm_pdw_exec_sessions

Identify which queries are waiting on resources:

▼

sys.dm_pdw_waits
sys.dm_pdw_lock_waits
sys.resource_governor_workload_groups

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

Monitor the database for long-running queries:

▼

sys.dm_pdw_exec_requests
sys.dm_pdw_sql_requests
sys.dm_pdw_exec_sessions

Identify which queries are waiting on resources:

▼

sys.dm_pdw_waits
sys.dm_pdw_lock_waits
sys.resource_governor_workload_groups

NEW QUESTION 228

- (Exam Topic 3)

You need to build a solution to ensure that users can query specific files in an Azure Data Lake Storage Gen2 account from an Azure Synapse Analytics serverless SQL pool.

Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.

Actions

Answer Area

Create an external file format object

Create an external data source

Create a query that uses Create Table as Select

Create a table

Create an external table



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Graphical user interface, text, application, email Description automatically generated

Step 1: Create an external data source

You can create external tables in Synapse SQL pools via the following steps:

- CREATE EXTERNAL DATA SOURCE to reference an external Azure storage and specify the credential that should be used to access the storage.
 - CREATE EXTERNAL FILE FORMAT to describe format of CSV or Parquet files.
 - CREATE EXTERNAL TABLE on top of the files placed on the data source with the same file format.
- Step 2: Create an external file format object

Creating an external file format is a prerequisite for creating an external table. Step 3: Create an external table
 Reference:
<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-tables-external-tables>

NEW QUESTION 229

- (Exam Topic 3)

You are designing an application that will store petabytes of medical imaging data

When the data is first created, the data will be accessed frequently during the first week. After one month, the data must be accessible within 30 seconds, but files will be accessed infrequently. After one year, the data will be accessed infrequently but must be accessible within five minutes.

You need to select a storage strategy for the data. The solution must minimize costs.

Which storage tier should you use for each time frame? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

First week:	▼
	Archive
	Cool
	Hot

After one month:	▼
	Archive
	Cool
	Hot

After one year:	▼
	Archive
	Cool
	Hot

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

First week: Hot

Hot - Optimized for storing data that is accessed frequently. After one month: Cool

Cool - Optimized for storing data that is infrequently accessed and stored for at least 30 days.

After one year: Cool

NEW QUESTION 234

- (Exam Topic 3)

You are monitoring an Azure Stream Analytics job.

The Backlogged Input Events count has been 20 for the last hour. You need to reduce the Backlogged Input Events count.

What should you do?

- A. Drop late arriving events from the job.
- B. Add an Azure Storage account to the job.
- C. Increase the streaming units for the job.
- D. Stop the job.

Answer: C

Explanation:

General symptoms of the job hitting system resource limits include:

➤ If the backlog event metric keeps increasing, it's an indicator that the system resource is constrained (either because of output sink throttling, or high CPU).

Note: Backlogged Input Events: Number of input events that are backlogged. A non-zero value for this metric implies that your job isn't able to keep up with the number of incoming events. If this value is slowly increasing or consistently non-zero, you should scale out your job: adjust Streaming Units.

Reference:

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-scale-jobs> <https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-monitoring>

NEW QUESTION 238

- (Exam Topic 3)

You are processing streaming data from vehicles that pass through a toll booth.

You need to use Azure Stream Analytics to return the license plate, vehicle make, and hour the last vehicle passed during each 10-minute window.

How should you complete the query? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.


```
WITH LastInWindow AS
(
    SELECT
        (Time) AS LastEventTime
    FROM
        Input TIMESTAMP BY Time
    GROUP BY
        (minute, 10)
)
SELECT
    Input.License_plate,
    Input.Make,
    Input.Time
FROM
    Input TIMESTAMP BY Time
    INNER JOIN LastInWindow
    ON DATEDIFF(minute, Input, LastInWindow) BETWEEN 0 AND 10
    AND Input.Time = LastInWindow.LastEventTime
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Graphical user interface, text, application Description automatically generated

Box 1: MAX

The first step on the query finds the maximum time stamp in 10-minute windows, that is the time stamp of the last event for that window. The second step joins the results of the first query with the original stream to find the event that match the last time stamps in each window.

Query:

```
WITH LastInWindow AS (
SELECT
MAX(Time) AS LastEventTime FROM
Input TIMESTAMP BY Time GROUP BY
TumblingWindow(minute, 10)
) SELECT
Input.License_plate, Input.Make, Input.Time
FROM
Input TIMESTAMP BY Time INNER JOIN LastInWindow
ON DATEDIFF(minute, Input, LastInWindow) BETWEEN 0 AND 10 AND Input.Time = LastInWindow.LastEventTime
```

Box 2: TumblingWindow

Tumbling windows are a series of fixed-sized, non-overlapping and contiguous time intervals.

Box 3: DATEDIFF

DATEDIFF is a date-specific function that compares and returns the time difference between two DateTime fields, for more information, refer to date functions.

Reference:

https://docs.microsoft.com/en-us/stream-analytics-query/tumbling-window-azure-stream-analytics

NEW QUESTION 243

- (Exam Topic 3)

You are building an Azure Stream Analytics job that queries reference data from a product catalog file. The file is updated daily. The reference data input details for the file are shown in the Input exhibit. (Click the Input tab.)

Input Details

products

Test Delete

Container

Create new Use existing

refdata

Path pattern

product.csv

Date format

YYYY/MM/DD

Time format

HH

Event serialization format *

CSV

Delimiter

comma (,)

Encoding

UTF-8

Save

If the chosen resource and the stream analytics job are located in different regions, you will be billed to move data between regions.

The storage account container view is shown in the Refdata exhibit. (Click the Refdata tab.)

refdata

Container

Search (Ctrl + /) Upload Add Directory Refresh Rename Delete

Overview Access Control (IAM) Settings

Access policy Properties Metadata

Authentication method: Access key (Switch to Azure AD User Account)

Location: refdata / 2020-03-20

Search blobs by prefix (case-sensitive)

Name

[..]

product.csv

You need to configure the Stream Analytics job to pick up the new reference data.
What should you configure? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Path pattern:

{date}/product.csv

{date}/{time}/product.csv

product.csv

*/product.csv

Date format:

MM/DD/YYYY

YYYY/MM/DD

YYYY-DD-MM

YYYY-MM-DD

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:
Graphical user interface, application, table Description automatically generated
Box 1: {date}/product.csv
In the 2nd exhibit we see: Location: refdata / 2020-03-20

Note: Path Pattern: This is a required property that is used to locate your blobs within the specified container. Within the path, you may choose to specify one or more instances of the following 2 variables:
{date}, {time}
Example 1: products/{date}/{time}/product-list.csv
Example 2: products/{date}/product-list.csv
Example 3: product-list.csv
Box 2: YYYY-MM-DD
Note: Date Format [optional]: If you have used {date} within the Path Pattern that you specified, then you can select the date format in which your blobs are organized from the drop-down of supported formats.
Example: YYYY/MM/DD, MM/DD/YYYY, etc. Reference:
<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-use-reference-data>

NEW QUESTION 248

- (Exam Topic 3)
You have an Azure Synapse Analytics dedicated SQL pool.
You need to ensure that data in the pool is encrypted at rest. The solution must NOT require modifying applications that query the data.
What should you do?

- A. Enable encryption at rest for the Azure Data Lake Storage Gen2 account.
- B. Enable Transparent Data Encryption (TDE) for the pool.
- C. Use a customer-managed key to enable double encryption for the Azure Synapse workspace.
- D. Create an Azure key vault in the Azure subscription grant access to the pool.

Answer: B

Explanation:
Transparent Data Encryption (TDE) helps protect against the threat of malicious activity by encrypting and decrypting your data at rest. When you encrypt your database, associated backups and transaction log files are encrypted without requiring any changes to your applications. TDE encrypts the storage of an entire database by using a symmetric key called the database encryption key.
Reference:
<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-overviewmana>

NEW QUESTION 251

- (Exam Topic 3)
You have an Azure data factory.
You execute a pipeline that contains an activity named Activity1. Activity1 produces the following output.

```
{
  ...
  "dataRead": 1208,
  "dataWritten": 1208,
  "filesRead": 1,
  "filesWritten": 1,
  "sourcePeakConnections": 3,
  "sinkPeakConnections": 2,
  "copyDuration": 13,
  "throughput": 0.147,
  "effectiveIntegrationRuntime": "AutoResolveIntegrationRuntime (West Central US)",
  "usedDataIntegrationUnits": 4,
  "reportLineageToPurview": {
    "status": "Succeeded",
    "durationInSeconds": "4"
  }
}
```

For each of the following statements select Yes if the statement is true. Otherwise, select No. NOTE: Each correct selection is worth one point.

Answer Area		
Statements	Yes	No
Activity1 is a Copy activity.	<input type="radio"/>	<input type="radio"/>
Activity1 is executed by using a self-hosted integration runtime.	<input type="radio"/>	<input type="radio"/>
The data factory that executed the pipeline is connected to Microsoft Purview.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

Statements

Activity1 is a Copy activity.

Activity1 is executed by using a self-hosted integration runtime.

The data factory that executed the pipeline is connected to Microsoft Purview.

Yes

No

☒☐☒☐☐☒

NEW QUESTION 252

- (Exam Topic 3)

You have two fact tables named Flight and Weather. Queries targeting the tables will be based on the join between the following columns.

Table	Column
Flight	ArrivalAirportID ArrivalDateTime
Weather	AirportID ReportDateTime

You need to recommend a solution that maximizes query performance. What should you include in the recommendation?

- A. In the tables use a hash distribution of ArrivalDateTime and ReportDateTime.
- B. In the tables use a hash distribution of ArrivalAirportID and AirportID.
- C. In each table, create an identity column.
- D. In each table, create a column as a composite of the other two columns in the table.

Answer: B

Explanation:

Hash-distribution improves query performance on large fact tables.

NEW QUESTION 254

- (Exam Topic 3)

You have an enterprise-wide Azure Data Lake Storage Gen2 account. The data lake is accessible only through an Azure virtual network named VNET1.

You are building a SQL pool in Azure Synapse that will use data from the data lake.

Your company has a sales team. All the members of the 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.

You plan to load data to the SQL pool every hour.

You need to ensure that the SQL pool can load the sales data from the data lake.

Which three actions should you perform? Each correct answer presents part of the solution. NOTE: Each area selection is worth one point.

- A. Add the managed identity to the Sales group.
- B. Use the managed identity as the credentials for the data load process.
- C. Create a shared access signature (SAS).
- D. Add your Azure Active Directory (Azure AD) account to the Sales group.
- E. Use the shared access signature (SAS) as the credentials for the data load process.
- F. Create a managed identity.

Answer: ADF

Explanation:

The managed identity grants permissions to the dedicated SQL pools in the workspace.

Note: Managed identity for Azure resources is a feature of Azure Active Directory. The feature provides Azure services with an automatically managed identity in Azure AD

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/security/synapse-workspace-managed-identity>

NEW QUESTION 258

- (Exam Topic 3)

You are designing an enterprise data warehouse in Azure Synapse Analytics that will store website traffic analytics in a star schema.

You plan to have a fact table for website visits. The table will be approximately 5 GB.

You need to recommend which distribution type and index type to use for the table. The solution must provide the fastest query performance.

What should you recommend? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Distribution:

Hash
Round robin
Replicated

Index:

Clustered columnstore
Clustered
Nonclustered

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Hash

Consider using a hash-distributed table when: The table size on disk is more than 2 GB.

The table has frequent insert, update, and delete operations. Box 2: Clustered columnstore

Clustered columnstore tables offer both the highest level of data compression and the best overall query performance.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-distribu> <https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-index>

NEW QUESTION 261

- (Exam Topic 3)

You are building a database in an Azure Synapse Analytics serverless SQL pool. You have data stored in Parquet files in an Azure Data Lake Storage Gen2 container. Records are structured as shown in the following sample.

```
{
  "id": 123,
  "address_housenumber": "19c", "address_line": "Memory Lane", "applicant1_name": "Jane", "applicant2_name": "Dev"
}
```

The records contain two applicants at most.

You need to build a table that includes only the address fields.

How should you complete the Transact-SQL statement? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

▼ applications

CREATE EXTERNAL TABLE
CREATE TABLE
CREATE VIEW

```
WITH (
  LOCATION = 'applications/',
  DATA_SOURCE = applications_ds,
  FILE_FORMAT = applications_file_format
)
AS
SELECT id, [address_housenumber] as addresshousenumber, [address_line1] as addressline1
FROM
  (BULK 'https://contosol.dfs.core.windows.net/applications/year=*/,*.parquet',
  CROSS APPLY
  OPENJSON
  OPENROWSET
  FORMAT='PARQUET') AS [r]
GO
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: CREATE EXTERNAL TABLE

An external table points to data located in Hadoop, Azure Storage blob, or Azure Data Lake Storage. External tables are used to read data from files or write data to files in Azure Storage. With Synapse SQL, you can use external tables to read external data using dedicated SQL pool or serverless SQL pool.

Syntax:
CREATE EXTERNAL TABLE { database_name.schema_name.table_name | schema_name.table_name | table_name }
(<column_definition> [,...n]) WITH (
LOCATION = 'folder_or_filepath', DATA_SOURCE = external_data_source_name, FILE_FORMAT = external_file_format_name
Box 2. OPENROWSET
When using serverless SQL pool, CETAS is used to create an external table and export query results to Azure Storage Blob or Azure Data Lake Storage Gen2.
Example: AS
SELECT decennialTime, stateName, SUM(population) AS population FROM
OPENROWSET(BULK
'https://azureopendatastorage.blob.core.windows.net/censusdatacontainer/release/us_population_county/year=*/
FORMAT='PARQUET') AS [r]
GROUP BY decennialTime, stateName GO
Reference:
<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-tables-external-tables>

NEW QUESTION 262

- (Exam Topic 3)
You have an Azure Data Lake Storage Gen2 container.
Data is ingested into the container, and then transformed by a data integration application. The data is NOT modified after that. Users can read files in the container but cannot modify the files.
You need to design a data archiving solution that meets the following requirements:

- > New data is accessed frequently and must be available as quickly as possible.
- > Data that is older than five years is accessed infrequently but must be available within one second when requested.
- > Data that is older than seven years is NOT accessed. After seven years, the data must be persisted at the lowest cost possible.
- > Costs must be minimized while maintaining the required availability.

How should you manage the data? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point

Five-year-old data:

Delete the blob.

Move to archive storage.

Move to cool storage.

Move to hot storage.

Seven-year-old data:

Delete the blob.

Move to archive storage.

Move to cool storage.

Move to hot storage.

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:
Box 1: Move to cool storage Box 2: Move to archive storage
Archive - Optimized for storing data that is rarely accessed and stored for at least 180 days with flexible latency requirements, on the order of hours.
The following table shows a comparison of premium performance block blob storage, and the hot, cool, and archive access tiers.

	Premium performance	Hot tier	Cool tier	Archive tier
Availability	99.9%	99.9%	99%	Offline
Availability (RA-GRS reads)	N/A	99.99%	99.9%	Offline
Usage charges	Higher storage costs, lower access, and transaction cost	Higher storage costs, lower access, and transaction costs	Lower storage costs, higher access, and transaction costs	Lowest storage costs, highest access, and transaction costs
Minimum storage duration	N/A	N/A	30 days ¹	180 days
Latency (Time to first byte)	Single-digit milliseconds	milliseconds	milliseconds	hours ²

Reference:

<https://docs.microsoft.com/en-us/azure/storage/blobs/storage-blob-storage-tiers>

NEW QUESTION 266

- (Exam Topic 3)

You plan to create an Azure Data Lake Storage Gen2 account

You need to recommend a storage solution that meets the following requirements:

- Provides the highest degree of data resiliency
- Ensures that content remains available for writes if a primary data center fails

What should you include in the recommendation? To answer, select the appropriate options in the answer area.

Answer Area

Replication mechanism:

- Change feed
- Zone-redundant storage (ZRS)
- Read-access geo-redundant storage (RA-GRS)
- Read-access geo-zone-redundant storage (RA-GRS)

Failover process:

- Failover initiated by Microsoft
- Failover manually initiated by the customer
- Failover automatically initiated by an Azure Automation job

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Graphical user interface, text, application Description automatically generated

Availability : "Microsoft recommends RA-GZRS for maximum availability and durability for your applications."

Failover: "The customer initiates the account failover to the secondary endpoint. " <https://docs.microsoft.com/en-us/azure/storage/common/storage-disaster-recovery-guidance?toc=/azure/storage/>

<https://docs.microsoft.com/en-us/answers/questions/32583/azure-data-lake-gen2-disaster-recoverystorage-acco.h>

NEW QUESTION 267

- (Exam Topic 3)

You are building a data flow in Azure Data Factory that upserts data into a table in an Azure Synapse Analytics dedicated SQL pool.

You need to add a transformation to the data flow. The transformation must specify logic indicating when a row from the input data must be upserted into the sink.

Which type of transformation should you add to the data flow?

- A. join
- B. select
- C. surrogate key

D. alter row

Answer: D

Explanation:

The alter row transformation allows you to specify insert, update, delete, and upsert policies on rows based on expressions. You can use the alter row transformation to perform upserts on a sink table by matching on a key column and setting the appropriate row policy

NEW QUESTION 270

- (Exam Topic 3)

You have an Azure subscription that contains an Azure SQL database named DB1 and a storage account named storage1. The storage1 account contains a file named File1.txt. File1.txt contains the names of selected tables in DB1.

You need to use an Azure Synapse pipeline to copy data from the selected tables in DB1 to the files in storage1. The solution must meet the following requirements:

- The Copy activity in the pipeline must be parameterized to use the data in File1.txt to identify the source and destination of the copy.
- Copy activities must occur in parallel as often as possible.

Which two pipeline activities should you include in the pipeline? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

- A. If Condition
- B. ForEach
- C. Lookup
- D. Get Metadata

Answer: BC

Explanation:

Lookup: This is a control activity that retrieves a dataset from any of the supported data sources and makes it available for use by subsequent activities in the pipeline. You can use a Lookup activity to read File1.txt from storage1 and store its content as an array variable1.

ForEach: This is a control activity that iterates over a collection and executes specified activities in a loop. You can use a ForEach activity to loop over the array variable from the Lookup activity and pass each table name as a parameter to a Copy activity that copies data from DB1 to storage11.

NEW QUESTION 272

- (Exam Topic 3)

You plan to develop a dataset named Purchases by using Azure databricks Purchases will contain the following columns:

- ProductID
- ItemPrice
- lineTotal
- Quantity
- StoreID
- Minute
- Month
- Hour
- Year
- Day

You need to store the data to support hourly incremental load pipelines that will vary for each StoreID. the solution must minimize storage costs. How should you complete the rode? To answer, select the appropriate options In the answer area.

NOTE: Each correct selection is worth one point.

df.write

.bucketBy

.partitionBy

.range

.sortBy

("*")

("StoreID","Hour")

("StoreID","Year","Month","Day","Hour")

.mode("append")

.csv("/Purchases")

.json("/Purchases")

.parquet("/Purchases")

.saveAsTable("/Purchases")

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: partitionBy

We should overwrite at the partition level. Example: df.write.partitionBy("y","m","d") mode(SaveMode.Append)

parquet("/data/hive/warehouse/db_name.db/" + tableName) Box 2: ("StoreID", "Year", "Month", "Day", "Hour", "StoreID") Box 3: parquet("/Purchases")

Reference:

<https://intellipaat.com/community/11744/how-to-partition-and-write-dataframe-in-spark-without-deleting-partiti>

NEW QUESTION 276

- (Exam Topic 3)

You have an Azure Synapse Analytics dedicated SQL pool that contains a table named Sales.Orders. Sales.Orders contains a column named SalesRep.

You plan to implement row-level security (RLS) for Sales.Orders.

You need to create the security policy that will be used to implement RLS. The solution must ensure that sales representatives only see rows for which the value of the SalesRep column matches their username.

How should you complete the code? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Answer Area

```
CREATE SCHEMA Security;

GO

CREATE FUNCTION Security.tvf_securitypredicate(@SalesRep AS nvarchar(50))

    RETURNS TABLE

WITH 

AS

    RETURN SELECT 1 AS tvf_securitypredicate_result

WHERE @SalesRep = USER_NAME();

GO

CREATE SECURITY POLICY SalesFilter
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

```
CREATE SCHEMA Security;

GO

CREATE FUNCTION Security.tvf_securitypredicate(@SalesRep AS nvarchar(50))

    RETURNS TABLE

WITH 

AS

    RETURN SELECT 1 AS tvf_securitypredicate_result

WHERE @SalesRep = USER_NAME();

GO

CREATE SECURITY POLICY SalesFilter
```

NEW QUESTION 277

- (Exam Topic 3)

You have an enterprise data warehouse in Azure Synapse Analytics named DW1 on a server named Server1. You need to verify whether the size of the transaction log file for each distribution of DW1 is smaller than 160 GB.

What should you do?

- A. On the master database, execute a query against the sys.dm_pdw_nodes_os_performance_counters dynamic management view.
- B. From Azure Monitor in the Azure portal, execute a query against the logs of DW1.
- C. On DW1, execute a query against the sys.database_files dynamic management view.

D. Execute a query against the logs of DW1 by using theGet-AzOperationalInsightSearchResult PowerShell cmdlet.

Answer: A

Explanation:

The following query returns the transaction log size on each distribution. If one of the log files is reaching 160 GB, you should consider scaling up your instance or limiting your transaction size.

```
-- Transaction log size SELECT
instance_name as distribution_db, cntr_value*1.0/1048576 as log_file_size_used_GB, pdw_node_id
FROM sys.dm_pdw_nodes_os_performance_counters WHERE
instance_name like 'Distribution_%'
AND counter_name = 'Log File(s) Used Size (KB)'
```

References:

<https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-manage-monitor>

NEW QUESTION 278

- (Exam Topic 3)

Note: The question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it As a result these questions will not appear in the review screen. You have an Azure Data Lake Storage account that contains a staging zone.

You need to design a dairy process to ingest incremental data from the staging zone, transform the data by executing an R script, and then insert the transformed data into a data warehouse in Azure Synapse Analytics.

Solution: You use an Azure Data Factory schedule trigger to execute a pipeline that executes a mapping data flow. and then inserts the data into the data warehouse.

Does this meet the goal?

- A. Yes
- B. No

Answer: B

NEW QUESTION 279

- (Exam Topic 3)

You use Azure Stream Analytics to receive Twitter data from Azure Event Hubs and to output the data to an Azure Blob storage account.

You need to output the count of tweets during the last five minutes every five minutes. Each tweet must only be counted once.

Which windowing function should you use?

- A. a five-minute Session window
- B. a five-minute Sliding window
- C. a five-minute Tumbling window
- D. a five-minute Hopping window that has one-minute hop

Answer: C

Explanation:

Tumbling window functions are used to segment a data stream into distinct time segments and perform a function against them, such as the example below. The key differentiators of a Tumbling window are that they repeat, do not overlap, and an event cannot belong to more than one tumbling window.

References:

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-window-functions>

NEW QUESTION 283

- (Exam Topic 3)

You have a C# application that process data from an Azure IoT hub and performs complex transformations. You need to replace the application with a real-time solution. The solution must reuse as much code as possible from the existing application.

- A. Azure Databricks
- B. Azure Event Grid
- C. Azure Stream Analytics
- D. Azure Data Factory

Answer: C

Explanation:

Azure Stream Analytics on IoT Edge empowers developers to deploy near-real-time analytical intelligence closer to IoT devices so that they can unlock the full value of device-generated data. UDF are available in C# for IoT Edge jobs

Azure Stream Analytics on IoT Edge runs within the Azure IoT Edge framework. Once the job is created in Stream Analytics, you can deploy and manage it using IoT Hub.

References:

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-edge>

NEW QUESTION 284

- (Exam Topic 3)

You have an Azure Synapse Analytics dedicated SQL pool.

You need to Create a fact table named Table1 that will store sales data from the last three years. The solution must be optimized for the following query operations:

Show order counts by week.

- Calculate sales totals by region.
- Calculate sales totals by product.

- Find all the orders from a given month. Which data should you use to partition Table1?

A. region
B. product
C. week
D. month

Answer: D

Explanation:

Table partitions enable you to divide your data into smaller groups of data. In most cases, table partitions are created on a date column.

Benefits to queries

Partitioning can also be used to improve query performance. A query that applies a filter to partitioned data can limit the scan to only the qualifying partitions. This method of filtering can avoid a full table scan and only scan a smaller subset of data. With the introduction of clustered columnstore indexes, the predicate elimination performance benefits are less beneficial, but in some cases there can be a benefit to queries.

For example, if the sales fact table is partitioned into 36 months using the sales date field, then queries that filter on the sale date can skip searching in partitions that don't match the filter.

Note: Benefits to loads

The primary benefit of partitioning in dedicated SQL pool is to improve the efficiency and performance of loading data by use of partition deletion, switching and merging. In most cases data is partitioned on a date column that is closely tied to the order in which the data is loaded into the SQL pool. One of the greatest benefits of using partitions to maintain data is the avoidance of transaction logging. While simply inserting, updating, or deleting data can be the most straightforward approach, with a little thought and effort, using partitioning during your load process can substantially improve performance.

Reference:

<https://learn.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-partitio>

NEW QUESTION 288

.....

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questions and Answers in PDF Format

DP-203 Practice Exam Features:

- * DP-203 Questions and Answers Updated Frequently
- * DP-203 Practice Questions Verified by Expert Senior Certified Staff
- * DP-203 Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- * DP-203 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

100% Actual & Verified — Instant Download, Please Click
[Order The DP-203 Practice Test Here](#)