DP-203 Q&As

Data Engineering on Microsoft Azure

Pass Microsoft DP-203 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.passapply.com/dp-203.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Microsoft Official Exam Center

- **Instant Download** After Purchase
- **100% Money Back** Guarantee
- **365 Days** Free Update
- **800,000+** Satisfied Customers

QUESTION 1

HOTSPOT

You develop a dataset named DBTBL1 by using Azure Databricks.

DBTBL1 contains the following columns:

1.

SensorTypeID

2.

GeographyRegionID

3.

Year

4.

Month

5.

Day

6.

Hour

7.

Minute

8.

Temperature

9.

WindSpeed 10.Other

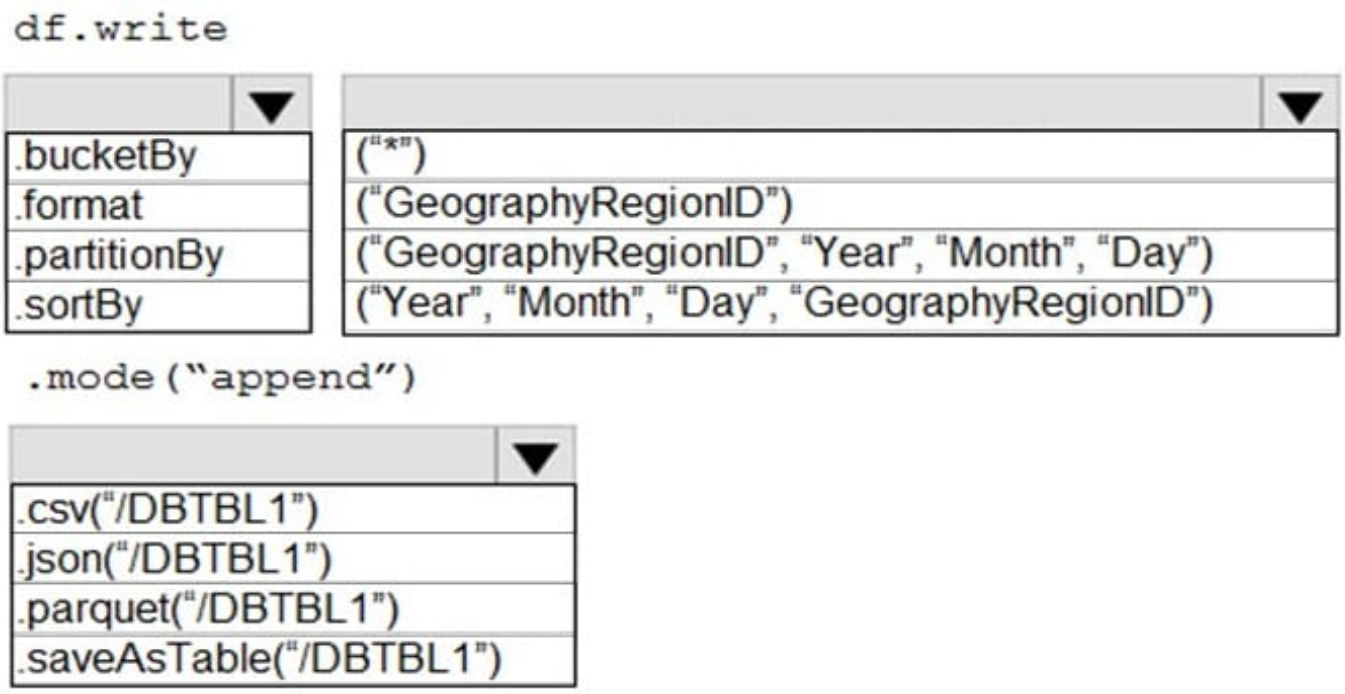
You need to store the data to support daily incremental load pipelines that vary for each GeographyRegionID. The solution must minimize storage costs.

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

NOTE: Each correct selection is worth one point.

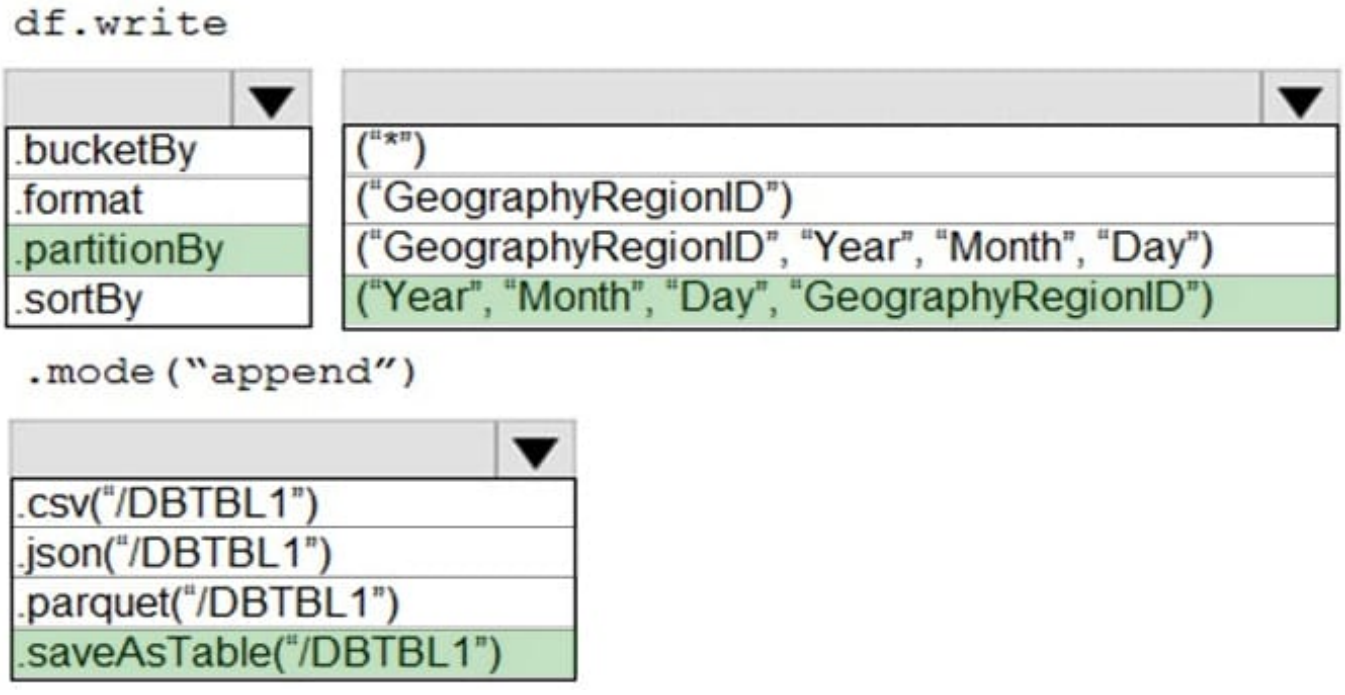
Hot Area:

Answer Area



Correct Answer:

Answer Area



Box 1: .partitionBy

Incorrect Answers:

.format:

Method: format():

Arguments: "parquet", "csv", "txt", "json", "jdbc", "orc", "avro", etc.

.bucketBy:

Method: bucketBy()

Arguments: (numBuckets, col, col..., coln)

The number of buckets and names of columns to bucket by. Uses Hive\\'s bucketing scheme on a filesystem.

Box 2: ("Year", "Month", "Day","GeographyRegionID")

Specify the columns on which to do the partition. Use the date columns followed by the GeographyRegionID column.

Box 3: .saveAsTable("/DBTBL1")

Method: saveAsTable()

Argument: "table_name"

The table to save to.

Reference:

https://www.oreilly.com/library/view/learning-spark-2nd/9781492050032/ch04.html

https://docs.microsoft.com/en-us/azure/databricks/delta/delta-batch

QUESTION 2

You need to implement the surrogate key for the retail store table. The solution must meet the sales transaction dataset requirements. What should you create?

A. a table that has an IDENTITY property

B. a system-versioned temporal table

C. a user-defined SEQUENCE object

D. a table that has a FOREIGN KEY constraint

Correct Answer: A

Scenario: Implement a surrogate key to account for changes to the retail store addresses.

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

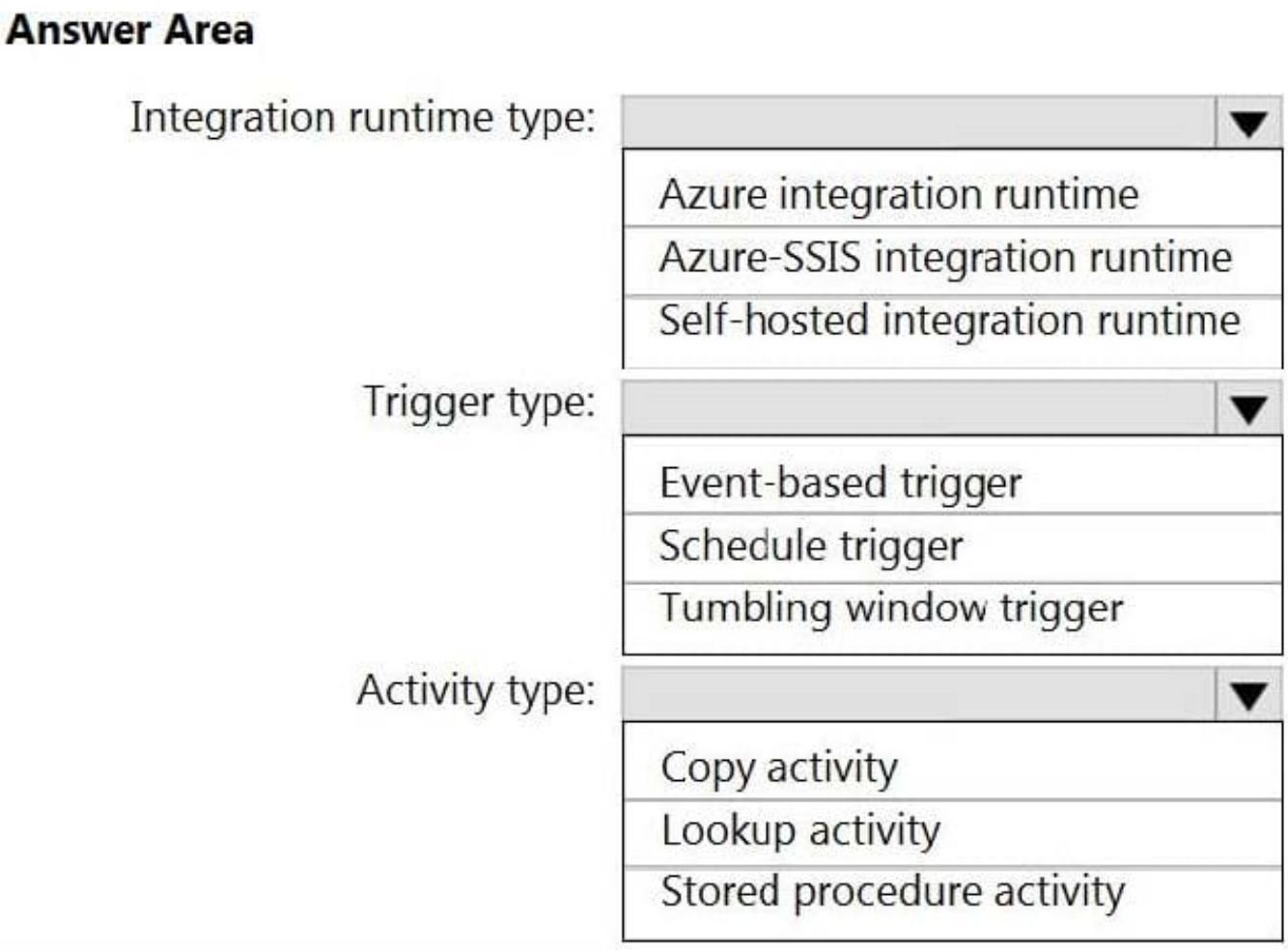
QUESTION 3

HOTSPOT

Which Azure Data Factory components should you recommend using together to import the daily inventory data from the SQL server to Azure Data Lake Storage? To answer, select the appropriate options in the answer area.

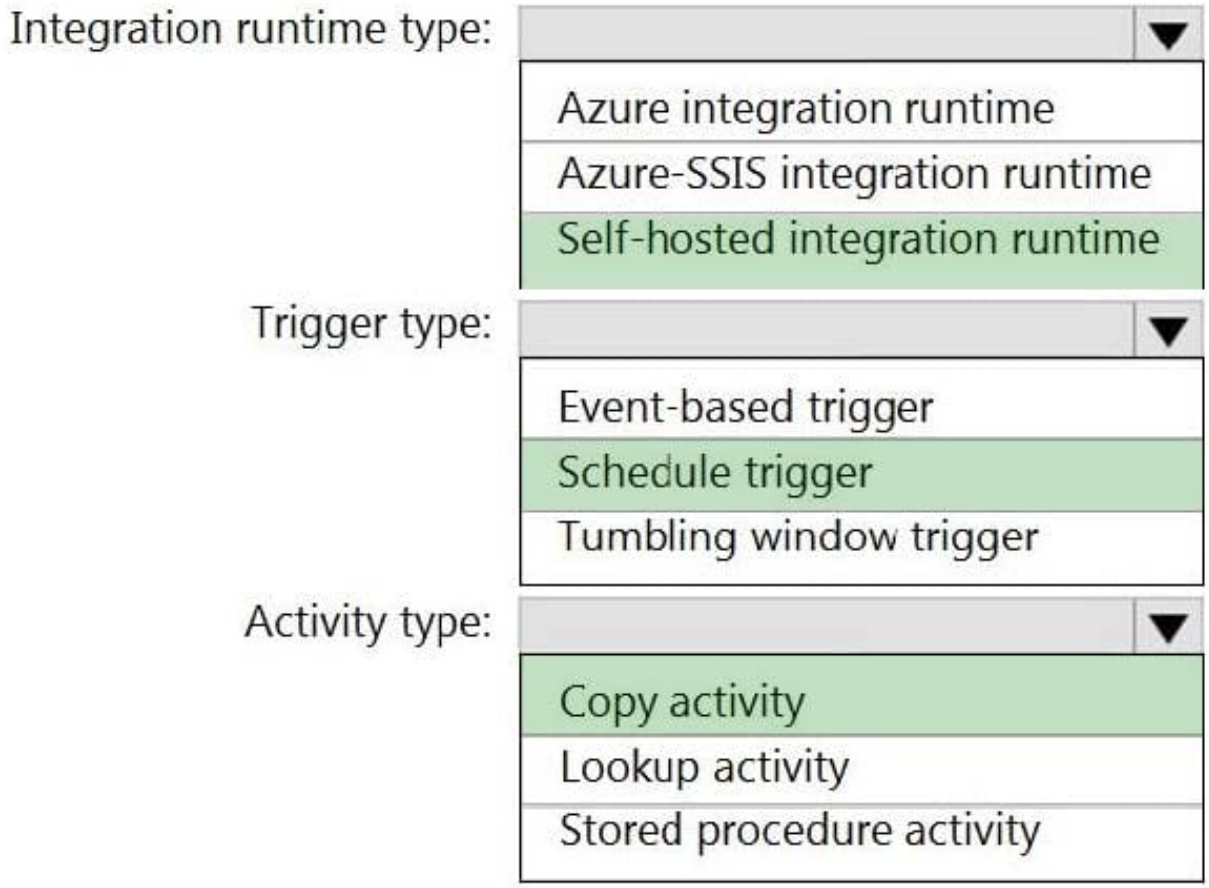
NOTE: Each correct selection is worth one point.

Hot Area:



Correct Answer:

Answer Area



Box 1: Self-hosted integration runtime

A self-hosted IR is capable of running copy activity between a cloud data stores and a data store in private network.

Box 2: Schedule trigger

Schedule every 8 hours

Box 3: Copy activity

Scenario:

Customer data, including name, contact information, and loyalty number, comes from Salesforce and can be imported into Azure once every eight hours. Row modified dates are not trusted in the source table.

Product data, including product ID, name, and category, comes from Salesforce and can be imported into Azure once every eight hours. Row modified dates are not trusted in the source table.

QUESTION 4

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.

Correct Answer: B

Hash-distribution improves query performance on large fact tables. Incorrect Answers:

A: Do not use a date column for hash distribution. All data for the same date lands in the same distribution. If several users are all filtering on the same date, then only 1 of the 60 distributions do all the processing work.

QUESTION 5

You have two Azure Data Factory instances named ADFdev and ADFprod. ADFdev connects to an Azure DevOps Git repository.

You publish changes from the main branch of the Git repository to ADFdev.

You need to deploy the artifacts from ADFdev to ADFprod.

What should you do first?

A. From ADFdev, modify the Git configuration.

B. From ADFdev, create a linked service.

C. From Azure DevOps, create a release pipeline.

D. From Azure DevOps, update the main branch.

Correct Answer: C

In Azure Data Factory, continuous integration and delivery (CI/CD) means moving Data Factory pipelines from one environment (development, test, production) to another.

Note:

The following is a guide for setting up an Azure Pipelines release that automates the deployment of a data factory to multiple environments.

1.

In Azure DevOps, open the project that\\'s configured with your data factory.

2.

On the left side of the page, select Pipelines, and then select Releases.

3.

Select New pipeline, or, if you have existing pipelines, select New and then New release pipeline.

4.

In the Stage name box, enter the name of your environment.

5.

Select Add artifact, and then select the git repository configured with your development data factory. Select the publish branch of the repository for the Default branch. By default, this publish branch is adf_publish.

6.

Select the Empty job template.

Reference: https://docs.microsoft.com/en-us/azure/data-factory/continuous-integration-deployment