# Question14 - Develop batch processing solutions

Case Study

**Instructions**  
  
This case study contains 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.  
  
Note: You cannot go back or review questions of this type on the actual certification exam.

## Question 14.1-

You keep Azure Storage access keys in Azure Key Vault.  
  
You need to configure a reference to those keys from within Azure Databricks to enable secure access to Azure Blob Storage.  
  
Solution: You create a secret scope using the Databricks CLI (version 0.7.1 and above).  
  
Does this solution meet the goal?

Yes

No

## Question 14.2-

You keep Azure Storage access keys in Azure Key Vault.  
  
You need to configure a reference to those keys from within Azure Databricks to enable secure access to Azure Blob Storage.  
  
Solution: You create a secret scope using the Secrets API via the Azure Databricks 2.0/secrets/scopes/create endpoint.  
  
Does this solution meet the goal?

Yes

No

## Question 14.3-

You keep Azure Storage access keys in Azure Key Vault.  
  
You need to configure a reference to those keys from within Azure Databricks to enable secure access to Azure Blob Storage.  
  
Solution: You open Azure Databricks workspace from Azure portal, add #secrets/createScope to its URL, and fill in all the details to create the secret scope.  
  
Does this solution meet the goal?

Yes

No

# Question17 Develop batch processing solutions

Case Study

**Instructions**  
  
This case study contains 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.  
  
Note: You cannot go back or review questions of this type on the actual certification exam.

## Question 17.1-

You are a data engineer. You are developing a data ingestion solution that ingests data from large pipe-delimited text files in an Azure Data Lake Gen 1 storage account to Azure Data Warehouse.  
  
You need to load the data.  
  
Solution:  
  
You do the following:

* Create an external file format and external data source.
* Create an external table that uses the external data source.
* Load the data from the external table.

Does this solution meet the goal?

Yes

No

## Question 17.2-

You are a data engineer. You are developing a data ingestion solution that ingests data from large pipe-delimited text files in an Azure Data Lake Gen 1 storage account to Azure Data Warehouse.  
  
You need to load the data.  
  
Solution:  
  
You do the following:

* Create an Azure Databricks account and a linked server.
* Create an external table that points to the Azure Databricks account.
* Load the data by running the dbutils.fs.cp command.

Does this solution meet the goal?

Yes

No

## Question 17.3-

You are a data engineer. You are developing a data ingestion solution that ingests data from large pipe-delimited text files in an Azure Data Lake Gen 1 storage account to Azure Data Warehouse.  
  
You need to load the data.  
  
Solution:  
  
You do the following:

* Create an Azure Cosmos DB account and a linked server.
* Create an external table that points to the Azure Cosmos DB account.
* Load the data by running the BULK IMPORT statement.

Does this solution meet the goal?

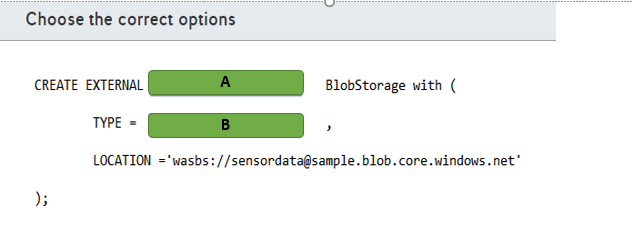
Yes

No

# Question88 - Develop batch processing solutions-

You are a data engineer for an Azure SQL Data Warehouse. You want to import data to a database from a large pipe-separated file in an Azure blob storage container.  
  
You need to create the connection to the container.  
  
How should you complete the T-SQL statement? To answer, select the appropriate code segments from the drop-down menus.

Choose the correct options



A)

1. DATA SOURCE
2. FILE FORMAT
3. TABLE

B)

1. BLOB\_STORAGE
2. HADOOP

# Question89 - Develop batch processing solutions – D -

You are a data architect for your company. Your team must analyze large CSV files from Azure blob storage daily. The team must be able to generate pie charts and bar charts without writing graphics code. The data engineers on the team know Python and SQL.  
  
You need to recommend a solution for analyzing the file.  
  
What should you recommend?

Choose the correct answer

Azure Databricks

Azure Data Lake

Stream Analytics

Log Analytics

# Question90 - Develop batch processing solutions-

You are a data engineer for your company. You create an Azure Databricks account. You add code to a notebook cell to import data from a comma-separated-value (CSV) file named sensordata.csv into a folder named /tmp.  
  
You need to copy the data to an Azure Data Lake Storage Gen 1 account.   
  
Which command should you run?

Choose the correct answer

dbutils.fs.ls("copy /tmp/sensordata.csv abfss://samplefs@sample.dfs.core.windows.net/")

spark.read.json("abfss://tmp/sensordata.csv@sample.dfs.core.windows.net")

dbutils.fs.cp("file:///tmp/sensordata.csv", "abfss://samplefs@sample.dfs.core.windows.net/")

spark.conf.set("tmp/sensordata.csv", "abfss://samplefs@sample.dfs.core.windows.net/")

# Question91 - Develop batch processing solutions – D -

You are a data engineer for your company. You create an Azure Databricks account by using the Azure portal. You plan to ingest data from blob storage into Databricks. You import a notebook from Github.  
  
You need to create the next resource so that you can run code to ingest the data.  
  
What should you create next?

Choose the correct answer

Cosmos DB account

SQL Data Warehouse instance

Spark cluster

Master key

# Question93 - Develop batch processing solutions - D

You are implementing a data engineering solution for your company. Every Thursday night, a third-party company loads shipping data to an Azure blob storage container. This data is in the form of large comma-separated-value (CSV) files. You want to run analytics on that data to automatically generate bar charts that you can visualize. If the data looks good visually, you want to import that data into a SQL Data Warehouse database table.  
  
You need to schedule a job to generate the charts.  
  
What should you do?

Choose the correct answer

Create a scheduled WebJob.

Create a scheduled Azure Databricks job.

Create a scheduled runbook in Azure Automation.

Create a scheduled function.

# Question94 - Develop batch processing solutions-

You are a data engineer. You have an Azure Databricks account with an imported notebook. You also create an Azure data factory.  
  
You need to ensure that Data Factory can access Databricks.  
  
What should you create?

Choose the correct answer

An access policy on a blob storage container

A master key

A blob storage container

An access token

# Question95 - Develop batch processing solutions-

You are a data engineer. You use Azure Data Factory to copy and transform data from Azure blob storage to an on-premises server.  
  
You need to ensure that you can successfully copy data.  
  
Which two actions should you perform? Each correct answer presents part of the solution.

Choose the correct answers

Create a self-hosted integration runtime in Azure Data Factory UI.

Create an Azure integration runtime.

Install the self-hosted integration runtime on an Azure virtual machine (VM).

Install the self-hosted integration runtime on the local network.

Create an Azure-SSIS integration runtime.

# Question96 - Develop batch processing solutions-

You are migrating a corporate research analytical solution from an internal datacenter to Azure.  
  
200 TB of research data is currently stored in an on-premises Hadoop cluster. You plan to copy it to Azure Storage. Your internal datacenter is connected to your Azure Virtual Network (VNet) with Express Route private peering. The Azure Storage service endpoint is accessible from the same VNet.  
  
Corporate policy dictates that the research data cannot be transferred over public internet.  
  
You need to securely migrate the research data online.   
  
What should you do?

Choose the correct answer

Transfer the data using Azure Data Box Heavy devices.

Transfer the data using Azure Data Box Disk devices.

Transfer the data using Azure Data Factory in distributed copy (DistCopy) mode, with an Azure Data Factory self-hosted Integration Runtime (IR) machine installed in the on-premises datacenter.

Transfer the data using Azure Data Factory in native Integration Runtime (IR) mode, with an Azure Data Factory self-hosted IR machine installed on the Azure VNet.

# Question97 - Develop batch processing solutions-

Your company is implementing a new Azure Databricks workspace.  
  
This environment should be able to access private data stored in Azure Data Lake Store Gen2. Credentials for Azure Databricks should be available in a secure way.  
  
You create a secret scope in the workspace and add the storage account key and a service principal secret.  
  
You need to configure a notebook in this workspace to read from Azure Data Lake Store Gen2.  
  
Which two actions should you perform? Each correct answer presents part of the solution.

Choose the correct answers

Generate a Shared Access Signature (SAS).

Read from the storage account using the RDD API.

Mount a filesystem using a service principal.

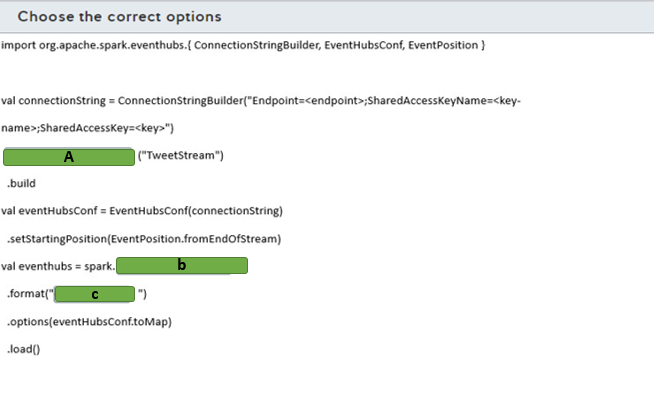
Configure the storage account key with spark.conf.set().

Configure the storage account key with spark.sparkContext.hadoopConfiguration().

# Question98 - Develop batch processing solutions-

Your team is implementing event processing to handle real-time streaming data from Twitter.  
  
You configure an Azure Event hub named TweetStream to ingest the streaming data. You use Azure Databricks to analyze the streaming data from Event Hub. You create a new workbook named TweetAnalysis.  
  
You need to configure the TweetAnalysis workbook to connect with TweetStream.  
  
How should you complete the notebook code? To answer, select the appropriate options from the drop-down menus.

Choose the correct options



A)

1. setEventHubName
2. SetNamespaceName
3. SetSasKeyName

B)

1. Read
2. ReadStream
3. WriteStream

C)

1. EventHubs
2. Kinesis
3. memory

# Question99 - Develop batch processing solutions-

You work as a data engineer in a company that uses Azure Data Factory for data pipelines.  
  
The company needs to connect with an on-premises database and move data periodically to an Azure SQL Database. The data pipeline is allowed to run at specific, fixed-size time intervals.  
  
You need to implement the Azure Data Factory component to connect with the on-premises database and use the appropriate pipeline execution.  
  
Which component and execution should you use? To answer, select the appropriate options from the drop-down menus.

Choose the correct options



A)

1. Activity
2. Linked Service
3. Pipeline

B)

1. Manual
2. Scehdual
3. Tumbling window

# Question100 - Develop batch processing solutions-

Your company hosts an enterprise resource planning (ERP) system with an on-premises SQL Server configured with SQL Server Integration Services (SSIS) packages to extract data from the ERP to an on-premises SQL Server Data Warehouse.  
  
You need to integrate the SSIS packages with Azure Data Factory by configuring the self-hosted integration runtime (IR) as a proxy for Azure-SSIS IR. You already created an Azure Blob storage for the integration.  
  
Which five actions should you perform? To answer, move the appropriate actions from the list of possible actions to the answer area and arrange them in any order.

Create a list in any order

Possible actions

Actions to perform

* Install the self-hosted IR in the on-premises data warehouse.
* Create a linked service in Azure Data Factory with an on-premises data warehouse.
* Create a linked service in Azure Data Factory with Azure Blob Storage.
* Create an Azure-SSIS IR in Azure Data Factory.
* Install the self-hosted IR in the on-premises SSIS.
* Set up the self-hosted IR as a proxy for your Azure-SSIS IR.
* Register the self-hosted IR with the authentication key.

# Question101 - Develop batch processing solutions-

Your team manages a data pipeline in Azure Data Factory that is configured with an on-premises SQL server and an Azure SQL database as linked services.  
  
The data pipeline will be used to incrementally copy data from an on-premises SQL server table named Customers to an Azure SQL database. The Customers table was created with the T-SQL statement shown in the exhibit.  
  
You implement a watermark approach to load delta data from the Customers table to Azure SQL Database. You create the Watermark table and a stored procedure to update the watermark table in the on-premises table.  
  
You need to implement the activities in the data pipeline to incrementally copy data.  
  
Which four actions should you perform? To answer, move the appropriate actions from the list of possible actions to the answer area and arrange them in any order.

Create a list in any order

Possible actions

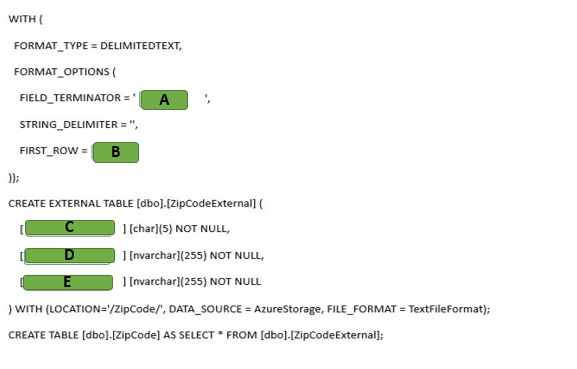
Actions to perform

* Look up the new watermark value in the Customers table from the PersonID field.
* Look up the old watermark value from the Customers table.
* Copy the delta data using the watermark values.
* Look up the new watermark value in the Customers table from the LastModifyTime field.
* Execute a stored procedure to update the watermarks in the Watermark table.
* Look up the old watermark value from the Watermark table.

# Question102 - Develop batch processing solutions-

You use a data pipeline in Azure Data Factory to move data from Azure Blob Storage to an Azure SQL Data Warehouse using PolyBase.  
  
The data is stored in a CSV file named zipCode.csv and has a structure as shown in the exhibit. The CSV file is stored in Azure Blob storage in a path named ZipCode inside a blob storage container.  
  
You need to complete the T-SQL query to ingest the data from Azure Blob Storage with PolyBase.  
  
How should you complete the code? To answer, select the appropriate options from the drop-down menus.

Choose the correct options



A)

1. ,
2. .
3. |

B)

1. 1
2. 2
3. 3

C)

1. City
2. Country
3. Zip code

D)

1. City
2. Country
3. Zip code

D)

1. City
2. Country
3. Zip code

# Question103 - Develop batch processing solutions-

You have an Azure Synapse Analytics SQL pool. You need to create two Azure Data Factory (ADF) pipelines to load data into the SQL pool.

* A pipeline to migrate data from SQL Server Analysis Services (SSAS)
* A pipeline for a daily incremental load from an Azure SQL Database

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

|  |  |  |
| --- | --- | --- |
| **Statement** | **Yes** | **No** |
| You can use the SSAS data source in an ADF Copy Activity. |  |  |
| You can implement the incremental load from the Azure SQL Database by using the change tracking feature combined with an ADF Copy Activity. |  |  |
| An ADF Copy Activity can invoke the Polybase feature to load the Azure Synapse Analytics SQL pool. |  |  |

# Question104 - Develop batch processing solutions - D

You are working with Azure Data Factory.   
  
Sales data from two regions must be imported into an Azure Synapse Analytics SQL pool. The data is stored in two CSV files.  
  
You have the following requirements:

* All data from the CSV files must be stored in a single destination table.
* Duplicate records must be inserted into the destination table.

You need to implement a mapping data flow to import the data.  
  
Which data flow transformation should you use?

Choose the correct answer

Join

Union

Aggregate

Lookup