



Practice Test 2

Completed on 17-June-2020



Attempt



Marks Obtained



Your score



Time Taken



Result

02

53 / 55

96.36%

00 H 07 M 16 S

Congratulations!
Passed

Share your Result with your friends



Domains wise Quiz Performance Report

No	Domain	Total Question	Correct	Incorrect	Unattempted	Marked as Review
1	Monitor and optimize data solutions	15	15	0	0	0
2	Manage and develop data processing	20	20	0	0	0
3	Implement data storage solutions	20	18	2	0	0
Total	All Domain	55	53	2	0	0

[Review the Answers](#)

Sorting by [All](#)

Question 1

Correct

Domain :Monitor and optimize data solutions

Your team has created a new Azure Data Factory environment. You have to analyse the pipeline executions. Trends need to be identified in execution duration over the past 30 days. You need to create a solution that would ensure that data can be queried from Azure Log Analytics.

Which of the following would you choose as the Log type when setting up the diagnostic setting for Azure Data Factory?

- A. ActivityRuns
- B. AllMetrics
- C. PipelineRuns
- D. TriggerRuns

Explanation:

Answer – C

Since you need to measure the pipeline execution, consider storing the data on pipeline runs.

The Microsoft documentation gives the schema of the log attributes for pipeline runs. Here there are properties for the start and end time for all activities that run within the pipeline

Pipeline-run log attributes

```
JSON

{
    "Level": "",
    "correlationId":"",
    "time":"",
    "runId":"",
    "resourceId":"",
    "category":"PipelineRuns",
    "level":"Informational",
    "operationName":"",
    "pipelineName":"",
    "start":"",
    "end":"",
    "status":"",
    "properties":
    {
        "Parameters": {
            "<parameter1Name>": "<parameter1Value>"
        },
        "SystemParameters": {
            "ExecutionStart": "",
            "TriggerId": "",
            "SubscriptionId": ""
        }
    }
}
```

Start and end time for all activities

Option A is incorrect since this will store the log for each activity execution within the pipeline itself.

Option B is incorrect since this will store all the metrics for the Azure Data Factory resource.

Option D is incorrect since this will store each trigger run for the Azure Data Factory resource.

For more information on monitoring Azure Data Factory, one can visit the below URL

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

Ask our Experts

Rate this Question?  

View Queries

[open ▾](#)

Question 2

Correct

Domain :Monitor and optimize data solutions

Your team has created a new Azure Data Factory environment. You have to analyse the pipeline executions. Trends need to be identified in execution duration over the past 30 days. You need to create a solution that would ensure that data can be queried from Azure Log Analytics.

Which of the following would you use as the storage location when setting up the diagnostic setting for Azure Data Factory?

- A. Azure Event Hub
- B. Azure Storage Account
- C. Azure Cosmos DB
- D. Azure Log Analytics

Explanation:

Answer – D

Since we have to query the logs via Log Analytics, we need to choose the storage option as Azure Log Analytics.

Since this is clearly mentioned as a requirement in the question, all other options are incorrect

For more information on monitoring Azure Data Factory, one can visit the below URL

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

[Ask our Experts](#)Rate this Question?  

View Queries

[open ▾](#)

Question 3

Correct**Domain :Manage and develop data processing**

You have to develop a solution that will make use of Azure Stream Analytics. The solution will perform data streaming and will also need a reference data store. Which of the following could be used as the input type for the reference data store?

- A. Azure Cosmos DB
- B. Azure Event Hubs
- C. Azure Blob storage
- D. Azure IoT Hub

Explanation:

Answer – C

You can use Azure Blob storage as an input type for the reference data

The Microsoft documentation mentions the following

Using reference data for lookups in Stream Analytics

06/21/2019 • 9 minutes to read •  +9

Reference data (also known as a lookup table) is a finite data set that is static or slowly changing in nature, used to perform a lookup or to augment your data streams. For example, in an IoT scenario, you could store metadata about sensors (which don't change often) in reference data and join it with real time IoT data streams. Azure Stream Analytics loads reference data in memory to achieve low latency stream processing. To make use of reference data in your Azure Stream Analytics job, you will generally use a [Reference Data Join](#) in your query.

Stream Analytics supports Azure Blob storage and Azure SQL Database as the storage layer for Reference Data. You can also transform and/or copy reference data to Blob storage from Azure Data Factory to use [any number of cloud-based and on-premises data stores](#).

Since this is clearly mentioned in the documentation, all other options are incorrect

For more information on using reference data, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-use-reference-data>

Ask our Experts

Rate this Question?  

[View Queries](#)

open ▾

Question 4

Correct

You have to develop a solution using Azure Stream Analytics. The stream will be used to receive Twitter data from Azure Event Hubs. The output would be sent to an Azure Blob storage account. The key requirement is to output the number of tweets during the last 3 minutes every 3 minutes. Each tweet must be counted only once. Which of the following would you use as the windowing function?

- A. A three-minute Session window
- B. A three-minute Sliding ion window
- C. A three-minute Tumbling window
- D. A three-minute Hopping window

Explanation:

Answer – C

The Tumbling window guarantees that data gets segmented into distinct time segments. And they do not repeat or overlap.

The Microsoft documentation mentions the following

Tumbling window

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.

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)
```

Since this is clearly mentioned in the documentation, all other options are incorrect

For more information on stream analytics window functions, one can visit the below URL

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

Ask our Experts

Rate this Question?

View Queries

open

Question 5**Correct****Domain :Implement data storage solutions**

A company currently has an Azure SQL database. The company wants to create an offline exported copy of the database. This is so that users can work with the data offline when they don't have any Internet connection on their laptops. Which of the following are ways that can be used to create the exported copy? Choose 3 answers from the options given below

- A. Export to a BACPAC file by using Azure Cloud Shell and save the file to a storage account.
- B. Export to a BACPAC file by using SQL Server Management Studio.
Save the file to a storage account. ✓
- C. Export to a BACPAC file by using the Azure portal
- D. Export to a BACPAC file by using Azure PowerShell and save the file locally ✓
- E. Export to a BACPAC file by using the SqlPackage utility ✓

Explanation:

Answer – B,D and E

The Microsoft documentation mentions the different ways in which you can export a BACPAC file of a SQL database

Export an Azure SQL database to a BACPAC file

07/16/2019 • 5 minutes to read •  +11

When you need to export a database for archiving or for moving to another platform, you can export the database schema and data to a [BACPAC](#) file. A BACPAC file is a ZIP file with an extension of BACPAC containing the metadata and data from a SQL Server database. A BACPAC file can be stored in Azure Blob storage or in local storage in an on-premises location and later imported back into Azure SQL Database or into a SQL Server on-premises installation.

Considerations when exporting an Azure SQL database

- For an export to be transactionally consistent, you must ensure either that no write activity is occurring during the export, or that you are exporting from a [transactionally consistent copy](#) of your Azure SQL database.
- If you are exporting to blob storage, the maximum size of a BACPAC file is 200 GB. To archive a larger BACPAC file, export to local storage.

Option A is incorrect because there is no mention in the Microsoft documentation of being able to create a backup from Azure Cloud Shell

Option C is incorrect because even though you can create a backup using the Azure Portal, the backup won't be available locally.

For more information on SQL database export, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/sql-database/sql-database-export>

Is this page helpful?

 Yes  No

In this article

Considerations when exporting an Azure SQL database

Export to a BACPAC file using the Azure portal

Export to a BACPAC file using the SQLPackage utility

Export to a BACPAC file using SQL Server Management Studio (SSMS)

Export to a BACPAC file using PowerShell

Next steps

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 6

Correct

Domain :Manage and develop data processing

A company is planning on developing a solution that would do the following

Use an Azure HDInsight for data ingestion. The data would be ingested from a relational database hosted on another cloud service

Use an Azure Data Lake Storage account which would be used to store the processed data

Give the ability to users to download the processed data

Which of the following technology would you suggest using for the data ingestion process?

- A. Apache Sqoop 
- B. Apache DisCp
- C. Azure CLI
- D. Azure PowerShell

Explanation:

Answer – A

You can use Apache Sqoop to transfer data from relational data stores. The Microsoft documentation mentions the following

Use Apache Sqoop with Hadoop in HDInsight

04/12/2019 • 3 min. za čitanje •  +4

Overview ▾

Learn how to use Apache Sqoop in HDInsight to import and export data between an HDInsight cluster and an Azure SQL database.

Although Apache Hadoop is a natural choice for processing unstructured and semi-structured data, such as logs and files, there may also be a need to process structured data that is stored in relational databases.

[Apache Sqoop](#) is a tool designed to transfer data between Hadoop clusters and relational databases. You can use it to import data from a relational database management system (RDBMS) such as SQL Server, MySQL, or Oracle into the Hadoop distributed file system (HDFS), transform the data in Hadoop with MapReduce or Apache Hive, and then export the data back into an RDBMS. In this article, you are using a SQL Server database for your relational database.

Option B is incorrect since this tool can be used when migrating data from Apache Hadoop systems to Azure HDInsight

Options C and D are incorrect since these are tools used to interact with Azure based resources

For more information on using Apache Sqoop with Azure HD Insight, one can visit the below URL

<https://docs.microsoft.com/en/Azure/hdinsight/hadoop/hdinsight-use-sqoop>

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 7

Correct

Domain :Manage and develop data processing

A company is planning on developing a solution that would do the following

Use an Azure HDInsight for data ingestion. The data would be ingested from a relational database hosted on another cloud service

Use an Azure Data Lake Storage account which would be used to store the processed data

Give the ability to users to download the processed data

Which of the following would you use as a process for the solution?

- A. Apache DistCp
- B. Apache Kafka
- C. C#
- D. Apache Hive

Explanation:

Answer – D

Once the data has been ingested, you can use Apache Hive to process the data. The Microsoft documentation also mentions this.

Use Apache Sqoop with Hadoop in HDInsight

04/12/2019 • 3 min. za čitanje •  +4

Overview ▾

Learn how to use Apache Sqoop in HDInsight to import and export data between an HDInsight cluster and an Azure SQL database.

Although Apache Hadoop is a natural choice for processing unstructured and semi-structured data, such as logs and files, there may also be a need to process structured data that is stored in relational databases.

[Apache Sqoop](#) is a tool designed to transfer data between Hadoop clusters and relational databases. You can use it to import data from a relational database management system (RDBMS) such as SQL Server, MySQL, or Oracle into the Hadoop distributed file system (HDFS), transform the data in Hadoop with MapReduce or [Apache Hive](#), and then export the data back into an RDBMS. In this article, you are using a SQL Server database for your relational database.

Option A is incorrect since this tool can be used when migrating data from Apache Hadoop systems to Azure HDInsight

Option B is incorrect since this is used for streaming data

Option C is incorrect since this is a programming language

For more information on using Apache Sqoop with Azure HD Insight, one can visit the below URL

<https://docs.microsoft.com/en/Azure/hdinsight/hadoop/hdinsight-use-sqoop>

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 8

Correct

Domain :Manage and develop data processing

A company is planning on developing a solution that would do the following

Use an Azure HDInsight for data ingestion. The data would be ingested from a relational database hosted on another cloud service

Use an Azure Data Lake Storage account which would be used to store the processed data

Give the ability to users to download the processed data

Which of the following can be used to download the data?

- A. Apache Sqoop 
- B. MapReduce
- C. RevoScaleR
- D. Ambari Hive View

Explanation:

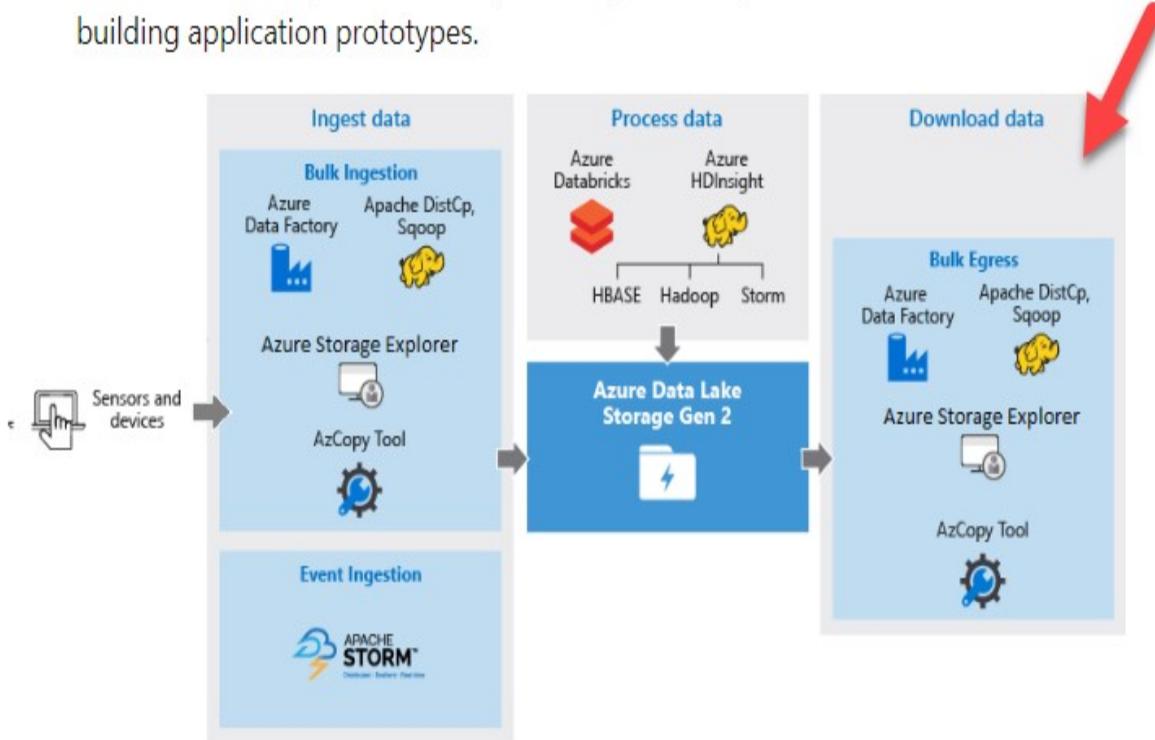
Answer – A

The Microsoft documentation provides the different tools you can use to download the data.

Download the data

You might also want to download or move data from Azure Data Lake Storage Gen2 for scenarios such as:

- Move data to other repositories to interface with your existing data processing pipelines. For example, you might want to move data from Data Lake Storage Gen2 to Azure SQL Database or on-premises SQL Server.
- Download data to your local computer for processing in IDE environments while building application prototypes.



Since this is clearly given in the Microsoft documentation, all other options are incorrect

For more information on downloading data from Azure Data Lake storage accounts, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-data-scenarios#download-the-data>

[Ask our Experts](#)Rate this Question?  [View Queries](#)[open ▾](#)**Question 9****Correct****Domain :Manage and develop data processing**

A company plans to store hundreds of files in an Azure Storage account and in Azure Data Lake Storage account. The files will be stored in the parquet format. A solution must be in place that would adopt the following requirements

Provide the ability to process the data every 5 hours

Give the ability for interactive data analysis

Give the ability to process data using solid-state drive caching

Make use of Directed Acyclic Graph processing mechanisms

Provide support for REST API calls for monitoring purposes

Ensure support for Python and Integration with Microsoft Power BI

Which of the following would you consider for the solution?

- A. Azure SQL Datawarehouse
- B. HDInsight Apache Storm cluster 
- C. Azure Stream Analytics
- D. HDInsight Spark cluster

Explanation:

Answer – B

All of these features are provided with the HDInsight Apache Storm cluster. The Microsoft documentation mentions the following

Why use Apache Storm on HDInsight?

Storm on HDInsight provides the following features:

- **99% Service Level Agreement (SLA) on Storm uptime:** For more information, see the [SLA information for HDInsight](#) document.
- Supports easy customization by running scripts against a Storm cluster during or after creation. For more information, see [Customize HDInsight clusters using script action](#).
- **Create solutions in multiple languages:** You can write Storm components in the language of your choice, such as Java, C#, and Python.
 - Integrates Visual Studio with HDInsight for the development, management, and monitoring of C# topologies. For more information, see [Develop C# Storm topologies with the HDInsight Tools for Visual Studio](#).
 - Supports the Trident Java interface. You can create Storm topologies that support exactly once processing of messages, transactional datastore persistence, and a set of common stream analytics operations.
- **Dynamic scaling:** You can add or remove worker nodes with no impact to running Storm topologies.
 - You must deactivate and reactivate running topologies to take advantage of new nodes added through scaling operations.
- **Create streaming pipelines using multiple Azure services:** Storm on HDInsight integrates with other Azure services such as Event Hubs, SQL Database, Azure Storage, and Azure Data Lake Storage.

All of the other options are incorrect because they don't provide all of the capabilities

For more information on Apache Storm on Azure HDInsight, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/hdinsight/storm/apache-storm-overview>

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 10

Correct

Domain :Manage and develop data processing

You have to develop a solution which would perform the following activities

Ingest twitter-based data into Azure

Give the ability to visualize real-time twitter data

Which of the following would you use to implement this solution? Choose 3 answers from the options given below

- A. Make use of an Event Grid Topic
- B. Make use of Azure Stream Analytics to query twitter data from an Event Hub 
- C. Make use of Azure Stream Analytics to query twitter data from an Event Grid
- D. Have a Logic App in place that would send twitter data to Azure 
- E. Create an Event Grid subscription
- F. Create an Event Hub Instance 

Explanation:

Answer – B,D and F

There is an example in the Microsoft documentation which show cases how to use Azure Stream Analytics for processing of twitter data

Real-time Twitter sentiment analysis in Azure Stream Analytics

07/09/2019 • 13 minutes to read •  +13

Learn how to build a sentiment analysis solution for social media analytics by bringing real-time Twitter events into Azure Event Hubs. Then write an Azure Stream Analytics query to analyze the data and store the results for later use or create a [Power BI](#) dashboard to provide insights in real time.

Social media analytics tools help organizations understand trending topics. Trending topics are subjects and attitudes that have a high volume of posts on social media. Sentiment analysis, which is also called *opinion mining*, uses social media analytics tools to determine attitudes toward a product or idea.

Real-time Twitter trend analysis is a great example of an analytics tool because the hashtag subscription model enables you to listen to specific keywords (hashtags) and develop sentiment analysis of the feed.

Option A is incorrect because this is more of a messaging-based system

Options C and E are incorrect because the Event Grid service is used for event-based processing

For more information on the implementation, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-twitter-sentiment-analysis-trends>

Ask our Experts

Rate this Question?  

[View Queries](#)

open ▾

Question 11**Correct****Domain :Manage and develop data processing**

A company wants to pull data from an on-premise SQL Server and migrate the data to Azure Blob storage. The company is planning on using Azure Data Factory. Which of the following are steps that would be required to implement this solution?

- A. Create a new Azure Data Factory resource 
- B. Create a Virtual Private Network Connection from the on-premise network to Azure 
- C. Create a self-hosted integration runtime 
- D. Create a database master key
- E. Backup the database
- F. Configure the on-premise server to use an integration runtime

Explanation:

Answer – A ,B and C

First you have to create a Virtual Private Network Connection from the on-premise network to Azure. This is to ensure you have connectivity between your on-premise data centre and Azure.

Next create a new Azure Data Factory resource and then have a self-hosted integration runtime in Azure Data Factory.

Option D is incorrect because we don't need a database master key for this process

Option E is incorrect because we are using Azure Data Factory

Option F is incorrect because we need to configure the integration runtime in Azure Data Factory

For more information on how to copy data using Azure Data Factory, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/data-factory/tutorial-hybrid-copy-portal>

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 12

Correct

Domain :Manage and develop data processing

A company wants to integrate their on-premise Microsoft SQL Server data with Azure SQL database. Here the data must be transformed incrementally. Which of the following can be used to configure a pipeline to copy the data?

- A. Make use of the AzCopy tool with Blob storage as the linked service in the source
- B. Make use of Azure PowerShell with SQL Server as the linked service in the source
- C. Make use of Azure Data Factory UI with Blob storage as the linked service in the source
- D. Make use of .Net Data Factory API with Blob storage as the linked service in the source



Explanation:

Answer – C

You can make use of Azure Data Factory which makes use of Azure Blob storage. An example of this is also given in the Microsoft documentation

Move data from an on-premises SQL server to SQL Azure with Azure Data Factory

11/04/2017 • 8 minutes to read •  +9

This article shows how to move data from an on-premises SQL Server Database to a SQL Azure Database via [Azure Blob Storage](#) using the Azure Data Factory (ADF).

For a table that summarizes various options for moving data to an Azure SQL Database, see [Move data to an Azure SQL Database for Azure Machine Learning](#).

All other options are incorrect since you need to use the Azure Data Factory UI tool to develop a pipeline.

For more information on how to copy data using Azure Data Factory for on-premise SQL server, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/move-sql-azure-adf>

Ask our Experts

Rate this Question?  

[View Queries](#)

open ▾

Question 13

Correct

Domain :Monitor and optimize data solutions

A team currently managed Azure HDInsight cluster. The team spends quite a lot of time on creating and destroying clusters. They want to implement a solution that can be used to

deploy Azure HDInsight clusters with minimal effort. Which of the following can they implement for this requirement?

- A. Azure Databricks
- B. Azure Load Balancer
- C. Azure Resource Manager templates 
- D. Azure SQL data warehouse

Explanation:

Answer – C

You can use Azure Resource Manager templates to easily deploy resources in Azure. If you have to repeatable deploy the same infrastructure on Azure, you can make use of Azure Resource Manager templates.

Option A is incorrect since this is a separate Spark-based analytics platform

Option B is incorrect since this is used to distribute traffic to backend virtual machines

Option D is incorrect since this is data warehousing solution available on the Azure platform

For more information on the template format for creating HDInsight clusters, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/templates/microsoft.hdinsight/2018-06-01-preview/clusters>

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 14

Correct

A company wants to migrate a set of on-premise Microsoft SQL Server databases to Azure. They want to migrate the databases as a simple lift and shift process by using backup and restore processes.

Which of the following would they use in Azure to host the SQL databases?

- A. Azure SQL Database single database
- B. Azure SQL data warehouse
- C. Azure Cosmos DB
- D. Azure SQL Database managed instance 

Explanation:

Answer – D

For easy migration of on-premise databases, consider migrating to Azure SQL Database managed instance.

The Microsoft documentation mentions the following

What is Azure SQL Database managed instance?

11/27/2019 • 20 minutes to read •  +11

Managed instance is a new deployment option of Azure SQL Database, providing near 100% compatibility with the latest SQL Server on-premises (Enterprise Edition) Database Engine, providing a native [virtual network \(VNet\)](#) implementation that addresses common security concerns, and a [business model](#) favorable for on-premises SQL Server customers. The managed instance deployment model allows existing SQL Server customers to lift and shift their on-premises applications to the cloud with minimal application and database changes. At the same time, the managed instance deployment option preserves all PaaS capabilities (automatic patching and version updates, [automated backups](#), [high-availability](#)), that drastically reduces management overhead and TCO.

Option A is incorrect since this is a better option if you just want to host a single database on the Azure platform.

Option B is incorrect since this is data warehousing solution available on the Azure platform

Option C is incorrect since this is a NoSQL based database solution

For more information on Azure SQL Database managed instance, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/sql-database/sql-database-managed-instance>

Ask our Experts

Rate this Question?  

[View Queries](#)

open ▾

Question 15**Correct****Domain :Implement data storage solutions**

You have to design a Hadoop Distributed File System architecture. You are going to be using Microsoft Azure Data Lake as the data storage repository. You have to ensure the data repository has a resilient data schema.

Which of the following would you use to provide data access to clients?

- A. DataNode 
- B. NameNode
- C. PrimaryNode
- D. SecondaryNode

Explanation:

Answer – A

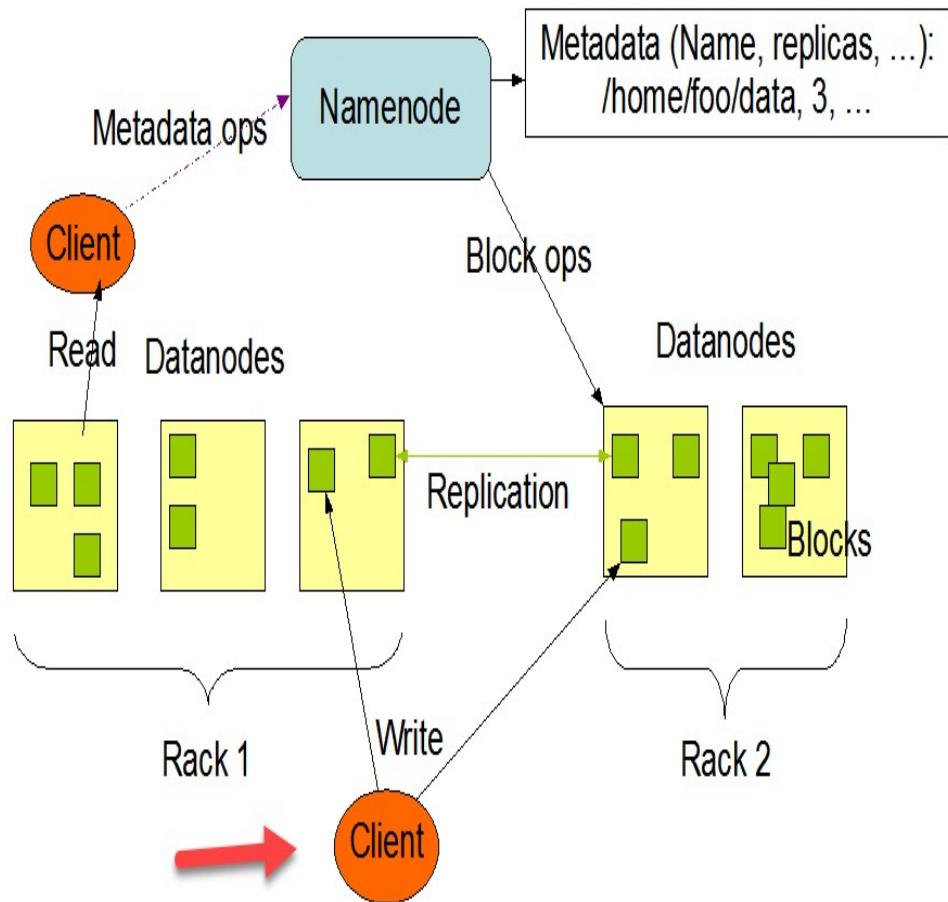
If you look at the architecture of the Hadoop Distributed File System, you will see that clients connect to the Data Nodes

The Hadoop documentation mentions the following

NameNode and DataNodes

HDFS has a master/slave architecture. An HDFS cluster consists of a single NameNode, a master server that manages the file system namespace and regulates access to files by clients. In addition, there are a number of DataNodes, usually one per node in the cluster, which manage storage attached to the nodes that they run on. HDFS exposes a file system namespace and allows user data to be stored in files. Internally, a file is split into one or more blocks and these blocks are stored in a set of DataNodes. The NameNode executes file system namespace operations like opening, closing, and renaming files and directories. It also determines the mapping of blocks to DataNodes. The DataNodes are responsible for serving read and write requests from the file system's clients. The DataNodes also perform block creation, deletion, and replication upon instruction from the NameNode.

HDFS Architecture



Since this is clear from the documentation, all other options are incorrect

For more information on HDFS design, one can visit the below URL

https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html#NameNode+and+DataNodes

View Queries

open ▾

Question 16

Correct

Domain :Implement data storage solutions

You have to design a Hadoop Distributed File System architecture. You are going to be using Microsoft Azure Data Lake as the data storage repository. You have to ensure the data repository has a resilient data schema.

Which of the following would be used to run operations on files and directories on the file system?

- A. DataNode
- B. NameNode
- C. PrimaryNode
- D. SecondaryNode

Explanation:

Answer – B

The file system namespace resides on the NameNode

The Hadoop documentation mentions the following

The File System Namespace

HDFS supports a traditional hierarchical file organization. A user or an application can create directories and store files inside these directories. The file system namespace hierarchy is similar to most other existing file systems; one can create and remove files, move a file from one directory to another, or rename a file. HDFS does not yet implement user quotas. HDFS does not support hard links or soft links. However, the HDFS architecture does not preclude implementing these features.

The NameNode maintains the file system namespace. Any change to the file system namespace or its properties is recorded by the NameNode. An application can specify the number of replicas of a file that should be maintained by HDFS. The number of copies of a file is called the replication factor of that file. This information is stored by the NameNode.

Since this is clear from the documentation, all other options are incorrect

For more information on HDFS design, one can visit the below URL

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 17

Correct

Domain :Implement data storage solutions

You have to design a Hadoop Distributed File System architecture. You are going to be using Microsoft Azure Data Lake as the data storage repository. You have to ensure the data repository has a resilient data schema.

Which of the following is used to perform block creation, deletion and replication?

- A. **DataNode** 
- B. **NameNode**
- C. **PrimaryNode**
- D. **SecondaryNode**

Explanation:

Answer – A

Here this is carried out by the Data Nodes.

The Hadoop documentation mentions the following

NameNode and DataNodes

HDFS has a master/slave architecture. An HDFS cluster consists of a single NameNode, a master server that manages the file system namespace and regulates access to files by clients. In addition, there are a number of DataNodes, usually one per node in the cluster, which manage storage attached to the nodes that they run on. HDFS exposes a file system namespace and allows user data to be stored in files. Internally, a file is split into one or more blocks and these blocks are stored in a set of DataNodes. The NameNode executes file system namespace operations like opening, closing, and renaming files and directories. It also determines the mapping of blocks to DataNodes. The DataNodes are responsible for serving read and write requests from the file system's clients. **The DataNodes also perform block creation, deletion, and replication upon instruction from the NameNode.**

Since this is clear from the documentation, all other options are incorrect

For more information on HDFS design, one can visit the below URL

https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html#NameNode+and+DataNodes

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 18

Correct

Domain :Monitor and optimize data solutions

A company wants to make use of Azure SQL Database with Elastic Pools. They have different customers who will have their own database in the pool. Each customer database has their own peak usage during different periods of the year. You need to consider the best way to implement Azure SQL database elastic pools to minimize costs. Which of the following is an option you would need to consider when configuring elastic pools?

- A. Number of transactions only
- B. eDTUs per database only
- C. Number of databases only
- D. CPU usage only
- E. eDTUs and maximum data size 

Explanation:

Answer – E

When you implement Elastic Pools using the DTU-based purchasing model, you have to consider both the eDTU's and the storage size for the databases.

The Microsoft documentation mentions the following

DTU-based purchasing model

- The eDTU price for an elastic pool includes a certain amount of storage at no additional cost. Extra storage beyond the included amount can be provisioned for an additional cost up to the max size limit in increments of 250 GB up to 1 TB, and then in increments of 256 GB beyond 1 TB. For included storage amounts and max size limits, see [Elastic pool: storage sizes and compute sizes](#).
- Extra storage for an elastic pool can be provisioned by increasing its max size using the [Azure portal](#), [PowerShell](#), the [Azure CLI](#), or the [REST API](#).
- The price of extra storage for an elastic pool is the extra storage amount multiplied by the extra storage unit price of the service tier. For details on the price of extra storage, see [SQL Database pricing](#).

Since this is clear from the documentation, all other options are incorrect

For more information on SQL database elastic pools, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/sql-database/sql-database-elastic-pool-scale>

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 19

Correct

Domain :Implement data storage solutions

A company needs to configure synchronization of data between their on-premise Microsoft SQL Server database and Azure SQL database. The synchronization process must include the following

Be able to perform an initial data synchronization to the Azure SQL Database with minimal downtime.

Be able to perform bi-directional synchronization after the initial synchronization is complete

Which of the following would you consider as the synchronization solution?

- A. Data Migration Assistant
- B. Backup and restore
- C. SQL Server Agent Job
- D. Azure SQL Data Sync

Explanation:

Answer – D

Azure SQL Data Sync can be used to synchronize data between the on-premise SQL Server and the Azure SQL database

The Microsoft documentation mentions the following

When to use Data Sync

Data Sync is useful in cases where data needs to be kept updated across several Azure SQL databases or SQL Server databases. Here are the main use cases for Data Sync:

- **Hybrid Data Synchronization:** With Data Sync, you can keep data synchronized between your on-premises databases and Azure SQL databases to enable hybrid applications. This capability may appeal to customers who are considering moving to the cloud and would like to put some of their application in Azure.
- **Distributed Applications:** In many cases, it's beneficial to separate different workloads across different databases. For example, if you have a large production database, but you also need to run a reporting or analytics workload on this data, it's helpful to have a second database for this additional workload. This approach minimizes the performance impact on your production workload. You can use Data Sync to keep these two databases synchronized.
- **Globally Distributed Applications:** Many businesses span several regions and even several countries/regions. To minimize network latency, it's best to have your data in a region close to you. With Data Sync, you can easily keep databases in regions around the world synchronized.

Option A is incorrect since this is just used to assess databases for the migration process

Option B is incorrect since this would just be the initial setup activity

Option C is incorrect since this is used to run administrative tasks on on-premise SQL databases

For more information on SQL database Sync, one can visit the below URL

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

View Queries

open ▾

Question 20

Correct

Domain :Monitor and optimize data solutions

A company has on-premise Microsoft SQL Server databases at several locations. The company wants to integrate the data in the databases with Microsoft Power BI and Microsoft Azure Logic Apps. You need to implement a solution that would avoid any single points of failure during the connection and transfer of data to the cloud. Latency must also be minimized. The transfer of data between the on-premise databases and Microsoft Azure must be secure. Which of the following would you implement for this requirement?

- A. Install a standalone on-premise Azure data gateway at each company location
- B. Install an on-premise data gateway in personal mode at each company location
- C. Install an Azure on-premise data gateway at the primary company location
- D. Install an Azure on-premise data gateway as a cluster at each location



Explanation:

Answer – D

If you need a high available solution, then you can install the on-premise data gateway as a cluster

The Microsoft documentation mentions the following

High-availability clusters for an on-premises data gateway

You can create high-availability clusters of gateway installations. The clusters help ensure that your organization can access on-premises data resources from cloud services like Power BI and PowerApps. Gateway admins use such clusters to avoid single points of failure when accessing on-premises data resources.

The gateway cloud service always uses the primary gateway in a cluster unless that gateway is not available. In that case, the service switches to the next available gateway in the cluster.

Since this is clear from the documentation, all other options are incorrect

For more information on high available clusters for the gateway, one can visit the below URL

<https://docs.microsoft.com/en-us/data-integration/gateway/service-gateway-high-availability-clusters>

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 21

Correct

Domain :Manage and develop data processing

You need to migrate data from an Azure Blob storage account to an Azure SQL Data warehouse. Which of the following actions do you need to implement for this requirement? Choose 4 answers from the options given below

- A. Provision an Azure SQL Data Warehouse instance 

- B. Connect to the Blob storage container via SQL Server Management Studio
- C. Create an Azure Blob storage container
- D. Run the T-SQL statements to load the data 
- E. Connect to the Azure SQL Data warehouse via SQL Server Management Studio 
- F. Build external tables by using Azure portal
- G. Build external tables by using SQL Server Management Studio 

Explanation:

Answer – A, D,E and G

You first need to create an Azure SQL Data Warehouse instance.

Then you need to connect to the data warehouse via SQL Server Management Studio

Then create external tables to the Azure Blob storage account

And then finally use T-SQL statements to load the data

This is also given as an example in github as part of the Microsoft documentation on loading data from Azure Blob to an Azure SQL data warehouse.

This tutorial uses PolyBase to load New York Taxicab data from a global Azure blob to Azure SQL Data Warehouse. The tutorial uses the [Azure portal](#) and [SQL Server Management Studio \(SSMS\)](#) to:

```
[!div class="checklist"]
```

- Create a data warehouse in the Azure portal
- Set up a server-level firewall rule in the Azure portal
- Connect to the data warehouse with SSMS
- Create a user designated for loading data
- Create external tables for data in Azure blob storage
- Use the CTAS T-SQL statement to load data into your data warehouse
- View the progress of data as it is loading
- Create statistics on the newly loaded data

Option B is incorrect because you can't connect to Blob storage from SQL Server Management Studio

Option C is incorrect because you already have the blob data in place

Option F is incorrect because you need to build the external tables in SQL Server Management Studio

For more information on the example, one can visit the below URL

<https://github.com/MicrosoftDocs/azure-docs/blob/master/articles/sql-data-warehouse/load-data-from-azure-blob-storage-using-polybase.md>

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 22

Correct

Domain :Implement data storage solutions

A company wants to implement a lambda architecture on Azure. They want the solution to have a data warehousing store. They want to reduce on-going management activities and ensure response to SQL queries are less than a second. Which of the following HDInsight cluster could be used to fulfil this requirement?

- A. Interactive Query
- B. Apache Hadoop
- C. Apache HBase
- D. Apache Spark 

Explanation:

Answer – D

To implement a lambda architecture, you need to implement Apache Spark on Azure HDInsight.

The Microsoft documentation mentions the following

Azure Cosmos DB: Implement a lambda architecture on the Azure platform

08/01/2019 • 10 minutes to read •  +3

Lambda architectures enable efficient data processing of massive data sets. Lambda architectures use batch-processing, stream-processing, and a serving layer to minimize the latency involved in querying big data.

To implement a lambda architecture on Azure, you can combine the following technologies to accelerate real-time big data analytics:

- [Azure Cosmos DB](#), the industry's first globally distributed, multi-model database service.
- [Apache Spark for Azure HDInsight](#), a processing framework that runs large-scale data analytics applications
- Azure Cosmos DB [change feed](#), which streams new data to the batch layer for HDInsight to process
- The [Spark to Azure Cosmos DB Connector](#)

Since this is clear from the documentation, all other options are incorrect

For more information on the lambda architecture, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/cosmos-db/lambda-architecture>

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 23**Correct****Domain :Implement data storage solutions**

Your company has an Azure Data Lake storage account. They want to implement role-based access control (RBAC) so that project members can manage the Azure Data Lake Storage resources. Which of the following actions should you perform for this requirement? Choose 3 answers from the options given below

- A. Ensure to assign Azure AD security groups to Azure Data Lake Storage ✓
- B. Make sure to configure end-user authentication to the Azure Data Lake Storage account
- C. Make sure to configure service-to-service authentication to the Azure Data Lake Storage account
- D. Create security groups in Azure AD and then add the project members ✓
- E. Configure Access control lists for the Azure Data Lake Storage account ✓

Explanation:

Answer – A, D and E

You can assign users and service principals, but the Microsoft documentation recommends giving Azure AD groups permissions for Azure Data Lake Storage account. For the storage account itself, you can manage the permissions via Access Control Lists

The Microsoft documentation mentions the following

What is the best way to apply ACLs?

Always use Azure AD security groups as the assigned principal in ACLs. Resist the opportunity to directly assign individual users or service principals. Using this structure will allow you to add and remove users or service principals without the need to reapply ACLs to an entire directory structure.) Instead, you simply need to add or remove them from the appropriate Azure AD security group. Keep in mind that ACLs are not inherited and so reapplying ACLs requires updating the ACL on every file and subdirectory.

Since this is clear from the documentation, all other options are incorrect

For more information on Azure Data Lake storage access control, one can visit the below URL

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

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 24

Correct

Domain :Implement data storage solutions

A company has an Azure SQL Database and an Azure Blob storage account. They want data to be encrypted at rest on both systems. The company should be able to use their own key.

Which of the following would they use to configure security for the Azure SQL Database?

- A. Always Encrypted
- B. Cell-level encryption
- C. Row-level encryption

Q D. Transparent data encryption 

Explanation:

Answer – D

Transparent Data Encryption is used to encrypt data at rest for Azure SQL Server databases.

The Microsoft documentation mentions the following

Transparent Data Encryption (TDE)

05/09/2019 • 11 minutes to read •  +7

APPLIES TO:  SQL Server  Azure SQL Database  Azure Synapse Analytics (SQL DW)  Parallel Data Warehouse

Transparent Data Encryption (TDE) encrypts SQL Server, Azure SQL Database, and Azure Synapse Analytics (SQL DW) data files, known as encrypting data at rest. You can take several precautions to help secure the database such as designing a secure system, encrypting confidential assets, and building a firewall around the database servers. However, in a scenario where the physical media (such as drives or backup tapes) are stolen, a malicious party can just restore or attach the database and browse the data. One solution is to encrypt the sensitive data in the database and protect the keys that are used to encrypt the data with a certificate. This prevents anyone without the keys from using the data, but this kind of protection must be planned in advance.

All other options are incorrect as they would not give the facility to encrypt data at rest for the entire database.

For more information on Transparent Data Encryption, one can visit the below URL

<https://docs.microsoft.com/en-us/sql/relational-databases/security/encryption/transparent-data-encryption?view=sql-server-ver15>

[Ask our Experts](#)Rate this Question?  [View Queries](#)

open ▾

Question 25**Correct****Domain :Implement data storage solutions**

A company has an Azure SQL Database and an Azure Blob storage account. They want data to be encrypted at rest on both systems. The company should be able to use their own key.

Which of the following would they use to configure security for the Azure Blob storage account?

- A. Azure Disk Encryption
- B. Secure Transport Layer Security
- C. Storage Account Keys
- D. Default Storage Service Encryption 

Explanation:

Answer – D

You can manage the encryption of data at rest for Azure storage accounts by using the default storage service encryption

The Microsoft documentation mentions the following

Customer-managed keys

You can choose to manage Azure Storage encryption at the level of the storage account with your own keys. When you specify a customer-managed key at the level of the storage account, that key is used to encrypt and decrypt all data in the storage account, including blob, queue, file, and table data. Customer-managed keys offer greater flexibility to create, rotate, disable, and revoke access controls. You can also audit the encryption keys used to protect your data.

Option A is incorrect since this is used for encrypting data at rest for Azure Virtual machines

Option B is incorrect since this is used to encrypt data in transit

Option C is incorrect since this is used for authorization to storage accounts

For more information on Storage Service Encryption, one can visit the below URL

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

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 26

Correct

Domain :Monitor and optimize data solutions

A company has a set of Azure SQL Databases. They want to ensure that their IT Security team is informed when any security related operation occurs on the database. You need to configure Azure Monitor while ensuring administrative efforts are reduced. Which of the following actions would you perform for this requirement? Choose 3 answers from the options given below

- A. Create a new action group which send email alerts to the IT Security team 
 - B. Make sure to use all security operations as the condition 
 - C. Ensure to query audit log entries as the condition
 - D. Use all the Azure SQL Database servers as the resource 
-

Explanation:

Answer – A, B and D

You can setup alerts based on all the security conditions in Azure Monitor. When any security operation is performed, an alert can be sent to the IT Security team.

Option C is incorrect since we need to monitor all security related events

For more information on alerts for Azure SQL Databases, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/sql-database/sql-database-insights-alerts-portal>

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 27

Correct

Domain :Manage and develop data processing

You need to deploy a Microsoft Azure Stream Analytics job for an IoT based solution. The solution must minimize latency. The solution must also minimize the bandwidth usage between the job and the IoT device. Which of the following actions must you perform for this requirement? Choose 4 answers from the options given below

- A. Ensure to configure routes 

- B. Create an Azure Blob storage container ✓
 - C. Configure Streaming Units
 - D. Create an IoT Hub and add the Azure Stream Analytics modules to the IoT Hub namespace ✓
 - E. Create an Azure Stream Analytics edge job and configure job definition save location ✓
 - F. Create an Azure Stream Analytics cloud job and configure job definition save location
-

Explanation:

Answer – A,B,D and E

There is an article in the Microsoft documentation on configuring Azure Stream Analytics on IoT Edge devices

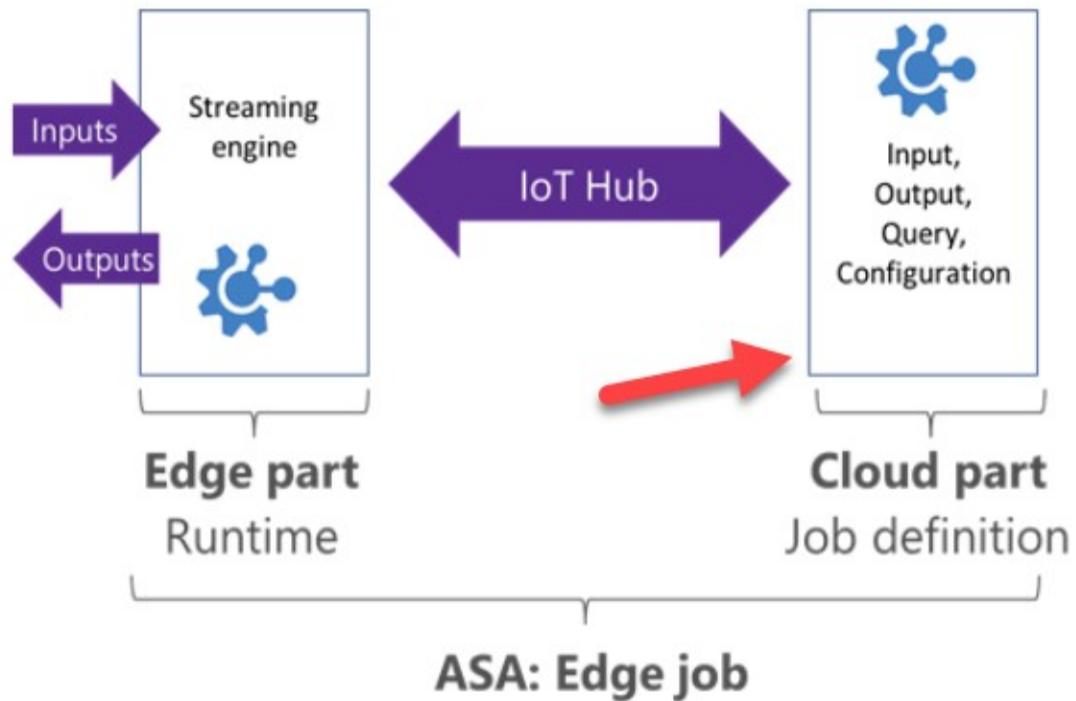
You need to have a storage container for the job definition

Installation instructions

The high-level steps are described in the following table. More details are given in the following sections.

Step	Notes
1 Create a storage container	Storage containers are used to save your job definition where they can be accessed by your IoT devices. You can reuse any existing storage container.

You also need to create a cloud part job definition



You also need to set the modules for your IoT edge device

Deployment ASA on your IoT Edge device(s)

Add ASA to your deployment

- In the Azure portal, open IoT Hub, navigate to **IoT Edge** and click on the device you want to target for this deployment.
- Select **Set modules**, then select **+ Add** and choose **Azure Stream Analytics Module**.
- Select the subscription and the ASA Edge job that you created. Click Save.

The screenshot shows the 'Set Modules' step in the Azure IoT Edge deployment process. It is the first step in a three-step wizard:

- Step 1: Add Modules (optional)**: This section is for deploying IoT Edge modules provided by Azure services. It currently supports Azure Machine Learning and Azure Functions. A button to "Import Azure Service IoT Edge Module" is available.
- Step 2: Specify Routes (optional)**: This section is for defining routes between modules.
- Step 3: Review Template (optional)**: This section allows reviewing the deployment template before saving.

In the main area, under "Stream Analytics - Edge Job", the "Subscription" is set to "Azure IoT client team subscription" and the "Edge job" is set to "JSASAEedge".

At the bottom, there are "Previous" and "Next" buttons, and a prominent blue "Save" button.

You also need to configure the Routes

Configure routes

IoT Edge provides a way to declaratively route messages between modules, and between modules and IoT Hub. The full syntax is described [here](#). Names of the inputs and outputs created in the ASA job can be used as endpoints for routing.

Since this is clear from the Microsoft documentation, all other options are incorrect

For more information on Stream Analytics on edge devices, one can visit the below URL

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

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 28

Correct

Domain :Monitor and optimize data solutions

A company has currently setup an Azure HDInsight cluster. The cluster is used to process log files generated from 100 web servers. Generally, around a few GB of log data is generated each day. All of the log files are stored in a single folder in Azure Data Lake Storage Gen 2. Which of the following are changes that can be carried out that can be used to improve the performance of the process? Choose 2 answers from the options given below

- A. Look at combining the daily log files into one file 
 - B. Look into moving the log files into folders 
 - C. Look to increasing the number of worker nodes
 - D. Look to increasing the value of the mapreduce.map.memory parameter
-

Explanation:

Answer – A and B

There is an article in the Microsoft documentation on improving the performance of Azure Data Lake Gen 2 storage

One is with regard to the file size and the other is with regards to organizing the data into folders.

Structure your data set

When data is stored in Data Lake Storage Gen2, the file size, number of files, and folder structure have an impact on performance. The following section describes best practices in these areas.

File size

Typically, analytics engines such as HDInsight and Azure Data Lake Analytics have a per-file overhead. If you store your data as many small files, this can negatively affect performance. In general, organize your data into larger sized files for better performance (256MB to 100GB in size). Some engines and applications might have trouble efficiently processing files that are greater than 100GB in size.

Sometimes, data pipelines have limited control over the raw data which has lots of small files. It is recommended to have a "cooking" process that generates larger files to use for downstream applications.

Organizing time series data in folders

For Hive workloads, partition pruning of time-series data can help some queries read only a subset of the data which improves performance.

Those pipelines that ingest time-series data, often place their files with a very structured naming for files and folders. Below is a very common example we see for data that is structured by date:

Since this is clear from the Microsoft documentation, all other options are incorrect

For more information on performance tuning for Azure Data Lake storage, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-performance-tuning-guidance>

Ask our Experts

Rate this Question?  

View Queries

[open ▾](#)

Question 29

Correct

Domain :Monitor and optimize data solutions

A company is using an Azure SQL Data Warehouse Gen2. Users are complaining that performance is slow when they run commonly used queries. They do not report such issues for infrequently used queries. Which of the following should they monitor for find out the source of the performance issues?

- A. Cache used percentage 
- B. Memory percentage
- C. CPU percentage
- D. Failed connections

Explanation:

Answer - A

To check for issues on frequently used queries, you can look at the cache percentage used.

The Microsoft documentation mentions the following

How to monitor the Gen2 cache

09/06/2018 • 2 minutes to read •  +2

The Gen2 storage architecture automatically tiers your most frequently queried columnstore segments in a cache residing on NVMe based SSDs designed for Gen2 data warehouses. Greater performance is realized when your queries retrieve segments that are residing in the cache. This article describes how to monitor and troubleshoot slow query performance by determining whether your workload is optimally leveraging the Gen2 cache.

Since this is clear from the Microsoft documentation, all other options are incorrect

For more information on monitoring Gen2 cache, one can visit the below URL

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

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 30

Correct

Domain :Monitor and optimize data solutions

A company has implemented a real-time data analysis solution. This solution is making use of Azure Event Hub to ingest the data. The data is then sent to Azure Stream Analytics cloud job. The cloud job has been configured to use 100 Streaming Units. Which of the following two actions can be performed to optimize the performance of the Azure Stream Analytics job?

- A. Scale up the Streaming Units of the job 

- B. Make use of event ordering
 - C. Make use of Azure Stream Analytics user-defined functions
 - D. Implement query parallelization by partitioning the data input 
-

Explanation:

Answer – A and D

You can scale up the streaming units and also implement parallelization.

The Microsoft documentation mentions the following

Understand and adjust Streaming Units

10/28/2019 • 9 minutes to read •  +6

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. This capacity lets you focus on the query logic and abstracts the need to manage the hardware to run your Stream Analytics job in a timely manner.

To achieve low latency stream processing, Azure Stream Analytics jobs perform all processing in memory. When running out of memory, the streaming job fails. As a result, for a production job, it's important to monitor a streaming job's resource usage, and make sure there is enough resource allocated to keep the jobs running 24/7.

The SU % utilization metric, which ranges from 0% to 100%, describes the memory consumption of your workload. For a streaming job with minimal footprint, this metric is usually between 10% to 20%. If SU% utilization is low and input events get backlogged, your workload likely requires more compute resources, which requires you to increase the number of SUs. It's best to keep the SU metric below 80% to account for occasional spikes. Microsoft recommends setting an alert on 80% SU Utilization metric to prevent resource exhaustion. For more information, see [Tutorial: Set up alerts for Azure Stream Analytics jobs](#).

Partitions in sources and sinks

Scaling a Stream Analytics job takes advantage of partitions in the input or output. Partitioning lets you divide data into subsets based on a partition key. A process that consumes the data (such as a Streaming Analytics job) can consume and write different partitions in parallel, which increases throughput.

Since this is clear from the Microsoft documentation, all other options are incorrect

For more information on stream analytics parallelization and scaling of stream analytic jobs, one can visit the below URL

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

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

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 31

Correct

Domain :Manage and develop data processing

View Case Study

Which of the following should be used as a technology for the "Data Ingestion" layer?

- A. Azure Logic Apps
- B. Azure Data Factory 
- C. Azure Automation

Q D. Azure Functions

Explanation:

Answer – B

Since you are looking at a data pipeline process, you must consider using Azure Data Factory. This can connect to multiple sources. You can define a workflow or pipeline and it can also run SQL Server Integration Service packages.

The Microsoft documentation mentions the following

Azure Data Factory is the platform that solves such data scenarios. It is the *cloud-based ETL and data integration service that allows you to create data-driven workflows for orchestrating data movement and transforming data at scale*. Using Azure Data Factory, you can create and schedule data-driven workflows (called pipelines) that can ingest data from disparate data stores. You can build complex ETL processes that transform data visually with data flows or by using compute services such as Azure HDInsight Hadoop, Azure Databricks, and Azure SQL Database.

Since this is the perfect fit for the requirement, all other options are incorrect

For more information on Azure Data Factory, one can visit the below URL

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

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 32

Correct

Domain :Manage and develop data processing

[View Case Study](#)

Which of the following should be used as a technology for the "Storage" layer?

- A. Azure Data Lake Storage 
- B. Azure Blob Storage
- C. Azure Files
- D. Azure SQL Data warehouse

Explanation:

Answer – A

Azure Data Lake Storage fulfils all of the right aspects as being built for Big Data Analytics. It can also scale in terms of storage.

The Microsoft documentation mentions the following

Introduction to Azure Data Lake Storage Gen2

10/11/2019 • 4 minutes to read •  +4

Azure Data Lake Storage Gen2 is a set of capabilities dedicated to big data analytics, built on [Azure Blob storage](#). Data Lake Storage Gen2 is the result of converging the capabilities of our two existing storage services, Azure Blob storage and Azure Data Lake Storage Gen1. Features from [Azure Data Lake Storage Gen1](#), such as file system semantics, directory, and file level security and scale are combined with low-cost, tiered storage, high availability/disaster recovery capabilities from [Azure Blob storage](#).

Since this is the perfect fit for the requirement, all other options are incorrect

For more information on Azure Data Lake Storage, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-introduction>

Ask our Experts

Rate this Question?  

View Queries

[open ▾](#)

Question 33

Correct

Domain :Manage and develop data processing

[View Case Study](#)

Which of the following should be used as a technology for the "Prepare and Train" layer?

- A. HDInsight Apache Spark Cluster
- B. Azure Databricks 
- C. HDInsight Apache Storm Cluster
- D. Azure SQL Data warehouse

Explanation:

Answer – B

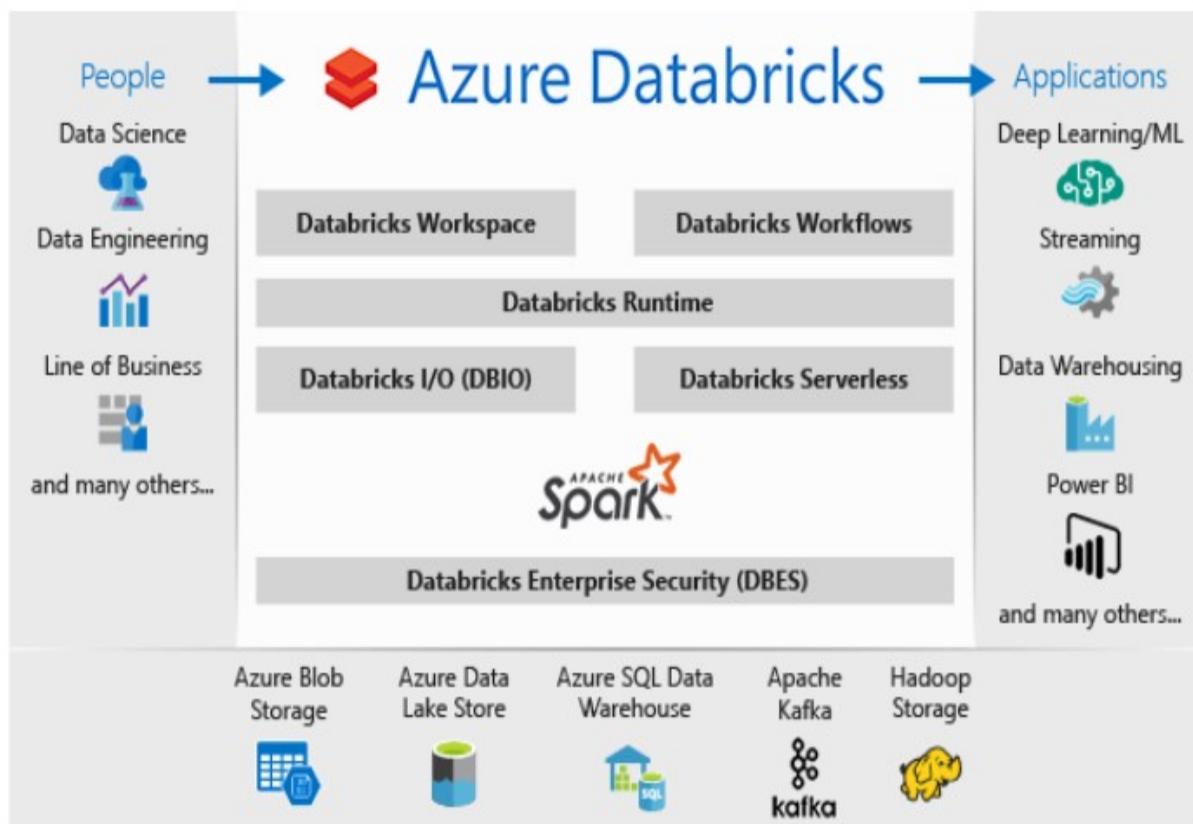
Azure Databricks is perfect for the Prepare and Train layer. Here you can perform interactive analysis using different programming languages.

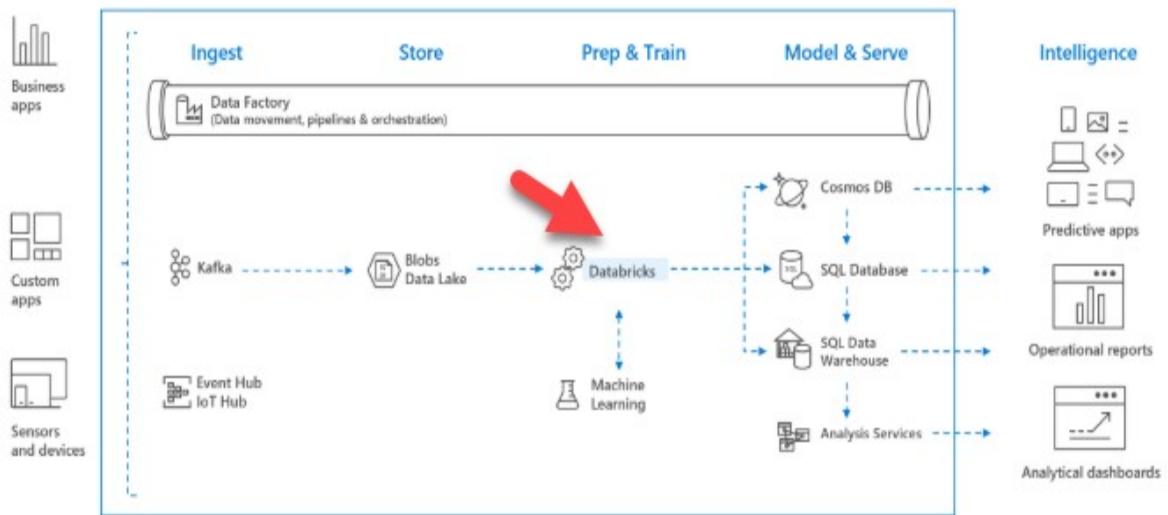
The Microsoft documentation mentions the following

What is Azure Databricks?

05/08/2019 • 3 minutes to read • 

Azure Databricks is an Apache Spark-based analytics platform optimized for the Microsoft Azure cloud services platform. Designed with the founders of Apache Spark, Databricks is integrated with Azure to provide one-click setup, streamlined workflows, and an interactive workspace that enables collaboration between data scientists, data engineers, and business analysts.





Since this is the perfect fit for the requirement, all other options are incorrect

For more information on Azure Databricks, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/azure-databricks/what-is-azure-databricks>

Ask our Experts

Rate this Question?

View Queries

open ▾

Question 34

Correct

Domain :Manage and develop data processing

View Case Study

Which of the following should be used as a technology for the "Model and Service" layer?

- A. HDInsight Apache Kafka cluster
- B. Azure SQL Data warehouse
- C. Azure Data Lake Storage

Q D. Azure Blob Storage

Explanation:

Answer – B

For columnar storage you can make use of Azure SQL data warehouse

The Microsoft documentation mentions the following

In a cloud data solution, data is ingested into big data stores from a variety of sources. Once in a big data store, Hadoop, Spark, and machine learning algorithms prepare and train the data. When the data is ready for complex analysis, SQL Analytics uses PolyBase to query the big data stores. PolyBase uses standard T-SQL queries to bring the data into SQL Analytics tables.

SQL Analytics stores data in relational tables with columnar storage. This format significantly reduces the data storage costs, and improves query performance. Once data is stored, you can run analytics at massive scale. Compared to traditional database systems, analysis queries finish in seconds instead of minutes, or hours instead of days.

The analysis results can go to worldwide reporting databases or applications. Business analysts can then gain insights to make well-informed business decisions.

Since this is the perfect fit for the requirement, all other options are incorrect

For more information on Azure SQL data warehouse, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-overview-what-is>

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 35**Correct****Domain :Monitor and optimize data solutions**

A company has setup several Azure HDInsight clusters. There were some errors reported while using these clusters. You have to recommend a solution that could be used to review the health of these clusters. Which of the following would you include in the recommendation?

- A. Azure Service Health
- B. Azure Automation
- C. Azure Log Analytics
- D. Application Insights

Explanation:

Answer – C

You can use Azure Monitor Logs which still stores logs in Azure Log Analytics to monitor your HDInsight clusters.

The Microsoft documentation mentions the following

Use Azure Monitor logs to monitor HDInsight clusters

08/05/2019 • 3 minutes to read •  +10

Learn how to enable Azure Monitor logs to monitor Hadoop cluster operations in HDInsight, and how to add an HDInsight monitoring solution.

[Azure Monitor logs](#) is a service in Azure Monitor that monitors your cloud and on-premises environments to maintain their availability and performance. It collects data generated by resources in your cloud and on-premises environments and from other monitoring tools to provide analysis across multiple sources.

Note

This article was recently updated to use the term Azure Monitor logs instead of Log Analytics. Log data is still stored in a Log Analytics workspace and is still collected and analyzed by the same Log Analytics service. We are updating the terminology to better reflect the role of [logs in Azure Monitor](#). See [Azure Monitor terminology changes](#) for details.

Option A is incorrect since this service is used to view the health of the underlying Azure services

Option B is incorrect since this service is used to provide automation on the Azure platform

Option D is incorrect since this service is used to view performance metrics for web applications

For more information on using Azure Log Analytics, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop-oms-log-analytics-tutorial>

Ask our Experts

Rate this Question?  

View Queries

[open](#) ▾

Question 36

Correct

Domain :Monitor and optimize data solutions

A company wants to make use of Azure Data Lake Gen 2 storage account. This would be used to store Big Data related to an application. The company wants to implement logging. They decide to create an Azure Automation runbook which would be used to copy events. Would this fulfil the requirement?

- A. Yes
- B. No 

Explanation:

Answer – B

You need to make use of Azure Data Lake storage diagnostics for this purpose

For more information on Azure Data Lake Gen 1 storage diagnostics, one can visit the below URL

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

Ask our Experts

Rate this Question?  

View Queries

[open](#) ▾

Question 37

Correct

Domain :Monitor and optimize data solutions

A company wants to make use of Azure Data Lake Gen 2 storage account. This would be used to store Big Data related to an application. The company wants to implement logging. They decide to use the information that is stored in Azure Active Directory reports. Would this fulfil the requirement?

- A. Yes
- B. No

Explanation:

Answer – B

You need to make use of Azure Data Lake storage diagnostics for this purpose

For more information on Azure Data Lake Gen 1 storage diagnostics, one can visit the below URL

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

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 38

Correct

Domain :Monitor and optimize data solutions

A company wants to make use of Azure Data Lake Gen 2 storage account. This would be used to store Big Data related to an application. The company wants to implement logging. They decide to configure Azure Data Lake Storage diagnostics to store the logs and metric data in a storage account.

Would this fulfil the requirement?

A. Yes 

B. No

Explanation:

Answer – A

Yes, this is the right approach

The Microsoft documentation mentions the following

Accessing diagnostic logs for Azure Data Lake Storage Gen1

03/26/2018 • 6 minutes to read •  +4

Learn to enable diagnostic logging for your Azure Data Lake Storage Gen1 account and how to view the logs collected for your account.

Organizations can enable diagnostic logging for their Azure Data Lake Storage Gen1 account to collect data access audit trails that provides information such as list of users accessing the data, how frequently the data is accessed, how much data is stored in the account, etc. When enabled, the diagnostics and/or requests are logged on a best-effort basis. Both Requests and Diagnostics log entries are created only if there are requests made against the service endpoint.

For more information on Azure Data Lake Gen 1 storage diagnostics, one can visit the below URL

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

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 39

Correct

Domain :Manage and develop data processing

[View Case Study](#)

Which of the following can be used to process and query the ingested data for the Tier 9 data?

- A. Azure Notification Hubs
- B. Apache Cache for Redis
- C. Azure Functions
- D. Azure Stream Analytics 

Explanation:

Answer – D

One way is to use Azure Stream Analytics. The Microsoft documentation mentions the following

Process data from your event hub using Azure Stream Analytics

07/09/2019 • 3 minutes to read • 

The Azure Stream Analytics service makes it easy to ingest, process, and analyze streaming data from Azure Event Hubs, enabling powerful insights to drive real-time actions. This integration allows you to quickly create a hot-path analytics pipeline. You can use the Azure portal to visualize incoming data and write a Stream Analytics query. Once your query is ready, you can move it into production in only a few clicks.

Option A is incorrect since this is a Notification service

Option B is incorrect since this is a cache service

Option C is incorrect since this is a serverless compute service

For more information on Azure Stream Analytics, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/event-hubs/process-data-azure-stream-analytics>

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 40

Correct

Domain :Manage and develop data processing

[View Case Study](#)

The Azure Data Factory instance must meet the requirements to move the data from the On-premise SQL Servers to Azure. Which of the following would you use as the integration runtime?

- A. Self-hosted integration runtime 
 - B. Azure-SSIS Integration runtime
 - C. .Net Common Language Runtime
 - D. Azure Integration runtime
-

Explanation:

Answer – A

The self-hosted integration runtime can be used to move data between on-premise data stores to Azure cloud data stores.

The Microsoft documentation mentions the following

Create and configure a self-hosted integration runtime

06/18/2019 • 22 minutes to read •  +14

The integration runtime (IR) is the compute infrastructure that Azure Data Factory uses to provide data-integration capabilities across different network environments. For details about IR, see [Integration runtime overview](#).

A self-hosted integration runtime can run copy activities between a cloud data store and a data store in a private network. It also can dispatch transform activities against compute resources in an on-premises network or an Azure virtual network. The installation of a self-hosted integration runtime needs an on-premises machine or a virtual machine inside a private network.

Since this is clearly mentioned in the Microsoft documentation, all other options are incorrect

For more information on self-hosted runtime environments, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/data-factory/create-self-hosted-integration-runtime>

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 41

Correct

Domain :Implement data storage solutions

View Case Study

The data for the external applications needs to be encrypted at rest. You decide to implement the following steps

Use the Always Encrypted Wizard in SQL Server Management Studio

Select the column that needs to be encrypted

Set the encryption type to Randomized

Configure the master key to be used from the Windows Certificate Store

Confirm the configuration and deploy the solution

Would these steps fulfil the requirement?

A. Yes

B. No 

Explanation:

Answer – B

As per the documentation, the encryption type needs to set as Deterministic when enabling Always Encrypted

Column Selection

Click **Next** on the **Introduction** page to open the **Column Selection** page. On this page, you will select which columns you want to encrypt, [the type of encryption](#), and [what column encryption key \(CEK\)](#) to use.

Encrypt **SSN** and **BirthDate** information for each patient. The **SSN** column will use deterministic encryption, which supports equality lookups, joins, and group by. The **BirthDate** column will use randomized encryption, which does not support operations.

Set the **Encryption Type** for the **SSN** column to **Deterministic** and the **BirthDate** column to **Randomized**. Click **Next**.

The screenshot shows the 'Column Selection' step of the 'Always Encrypted' wizard. The left sidebar lists steps: 'Introduction', 'Column Selection' (which is selected and highlighted in blue), 'Master Key Configuration', 'Validation', 'Summary', and 'Results'. The main area has a search bar 'Search column name...' and a checkbox 'Apply one key to all checked columns:' followed by a dropdown menu showing 'CEK_Auto1 (New)'. Below this is a table with columns: Name, State, Encryption Type, and Encryption Key. The table lists columns from the 'dbo.Patients' table: PatientId, SSN, FirstName, LastName, MiddleName, and StreetAddress. The 'SSN' column is checked in the 'State' column. A red arrow points to the 'Encryption Type' dropdown for the 'SSN' row, which is currently set to 'Deterministic'. A context menu is open over this dropdown, showing options: 'Choose Type...', 'Deterministic' (which is highlighted in blue), and 'Randomized'.

Name	State	Encryption Type	Encryption Key
dbo.Patients			
PatientId			
SSN	checked	Deterministic	CEK_Auto1 (New)
FirstName			
LastName			
MiddleName			
StreetAddress			

For more information on implementing Always Encrypted, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/sql-database/sql-database-always-encrypted>

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 42

Correct

Domain :Implement data storage solutions

View Case Study

The data for the external applications needs to be encrypted at rest. You decide to implement the following steps

Use the Always Encrypted Wizard in SQL Server Management Studio

Select the column that needs to be encrypted

Set the encryption type to Deterministic

Configure the master key to be used from the Windows Certificate Store

Confirm the configuration and deploy the solution

Would these steps fulfil the requirement?

A. Yes 

B. No

Explanation:

Answer – A

Yes, this is the right series of steps to implement Always Encrypted

For more information on implementing Always Encrypted, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/sql-database/sql-database-always-encrypted>

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 43

Correct

Domain :Implement data storage solutions

[View Case Study](#)

The data for the external applications needs to be encrypted at rest. You decide to implement the following steps

Use the Always Encrypted Wizard in SQL Server Management Studio

Select the column that needs to be encrypted

Set the encryption type to Deterministic

Configure the master key to be used from Azure Key vault

Confirm the configuration and deploy the solution

Would these steps fulfil the requirement?

A. Yes

B. No

Explanation:

Answer – B

As per the case study, all keys and certificates need to be managed in on-premise data stores

For more information on implementing Always Encrypted, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/sql-database/sql-database-always-encrypted>

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 44

Incorrect

Domain :Implement data storage solutions

[View Case Study](#)

Which of the following should you use as the masking function for Data type whizlabA?

- A. Custom Text 
- B. Default 
- C. Email
- D. Random number

Explanation:

Answer – B

As per the case study, below is the requirement for the Data type

For Data type whizlabA – Mask 4 or less string data type characters

You can use the "Default" masking function for this requirement

The Microsoft documentation mentions the following

Masking

Function Masking Logic

Default Full masking according to the data types of the designated fields

- Use XXXX or fewer Xs if the size of the field is less than 4 characters for string data types (nchar, ntext, nvarchar).
- Use a zero value for numeric data types (bigint, bit, decimal, int, money, numeric, smallint, smallmoney, tinyint, float, real).
- Use 01-01-1900 for date/time data types (date, datetime2, datetime, datetimeoffset, smalldatetime, time).
- For SQL variant, the default value of the current type is used.
- For XML the document <masked/> is used.
- Use an empty value for special data types (timestamp table, hierarchyid, GUID, binary, image, varbinary spatial types).

Since this is clear from the Microsoft documentation, all other options are incorrect

For more information on dynamic data masking, one can visit the below URL

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

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 45

Correct

Domain :Implement data storage solutions

View Case Study

Which of the following should you use as the masking function for Data type whizlabB?

- A. Custom Text
- B. Default
- C. Email
- D. Random number

Explanation:

Answer C

As per the case study, below is the requirement for the Data type

For Data type whizlabB – Expose the first letter and mask the domain

You can use the "Email" masking function for this requirement

The Microsoft documentation mentions the following

Email Masking method, which exposes the first letter and replaces the domain with XXX.com using a constant string prefix in the form of an email address.

aXX@XXXX.com

Since this is clear from the Microsoft documentation, all other options are incorrect

For more information on dynamic data masking, one can visit the below URL

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

Ask our Experts

Rate this Question?  

View Queries

open ▼

Question 46

Incorrect

Domain :Implement data storage solutions

[View Case Study](#)

Which of the following should you use as the masking function for Data type whizlabC?

- A. Custom Text ✓
- B. Default
- C. Email
- D. Random number ✗

Explanation:

Answer - A

As per the case study, below is the requirement for the Data type

For Data type whizlabC – Mask everything except characters at the beginning and the end

You can use the "Custom Text" masking function for this requirement

The Microsoft documentation mentions the following

Custom text **Masking method, which exposes the first and last characters** and adds a custom padding string in the middle. If the original string is shorter than the exposed prefix and suffix, only the padding string is used.
prefix[padding]suffix

Masking Field Format		
Custom text		
Exposed Prefix	Padding String	Exposed Suffix
3 ✓	X*X*X	2 ✓

Since this is clear from the Microsoft documentation, all other options are incorrect

For more information on dynamic data masking, one can visit the below URL

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

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 47

Correct

Domain :Implement data storage solutions

View Case Study

You need to implement the following requirement as per the case study

The Application access for Tier 7 and 8 must be restricted to the database only

Which of the following steps would you implement for this requirement? Choose 3 answers from the options given below

- A. Use Azure PowerShell to create a database firewall rule
- B. Configure the setting of “Allow Azure Services to Access Server” to  Disabled
- C. Configure the setting of “Allow Azure Services to Access Server” to Enabled
- D. Create a database firewall rule from the Azure portal
- E. Create a server firewall rule from the Azure portal 
- F. Use Transact-SQL to create a database firewall rule 

Explanation:

Answer – B, E and F

You can set database and firewall rules to restrict access to the server and the database

The Microsoft documentation mentions the following

Server-level IP firewall rules

These rules enable clients to access your entire Azure SQL server, that is, all the databases within the same SQL Database server. The rules are stored in the *master* database. You can have a maximum of 128 server-level IP firewall rules for an Azure SQL Server.

You can configure server-level IP firewall rules by using the Azure portal, PowerShell, or Transact-SQL statements.

- To use the portal or PowerShell, you must be the subscription owner or a subscription contributor.
- To use Transact-SQL, you must connect to the SQL Database instance as the server-level principal login or as the Azure Active Directory administrator. (A server-level IP firewall rule must first be created by a user who has Azure-level permissions.)

Database-level IP firewall rules

These rules enable clients to access certain (secure) databases within the same SQL Database server. You create the rules for each database (including the *master* database), and they're stored in the individual database.

You can only create and manage database-level IP firewall rules for master and user databases by using Transact-SQL statements and only after you configure the first server-level firewall.

Also ensure to set the "Allow Azure Services to Access Server" is set to Disabled so that no other service can access the database

Options A and D are incorrect since you can only create a database firewall rule via Transact-SQL

Option C is incorrect since the setting "Allow Azure Services to Access Server" should be set to Disabled.

For more information on server and database rules for Azure SQL databases, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/sql-database/sql-database-firewall-configure>

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 48

Correct

Domain :Monitor and optimize data solutions

View Case Study

You have to implement logging for monitoring the data warehousing solution. Which of the following would you log?

- A. RequestSteps
- B. DmsWorkers
- C. SQLRequests 
- D. ExecRequests

Explanation:

Answer – C

Since the SQL requests would affect the cache, these requests need to be monitored

The Microsoft documentation mentions the following on caching

Cache hit and used percentage

The matrix below describes scenarios based on the values of the cache metrics:

	High Cache hit percentage	Low Cache hit percentage
High Cache used percentage	Scenario 1	Scenario 2
Low Cache used percentage	Scenario 3	Scenario 4

Scenario 1: You are optimally using your cache. [Troubleshoot](#) other areas which may be slowing down your queries.

Scenario 2: Your current working data set cannot fit into the cache which causes a low cache hit percentage due to physical reads. Consider scaling up your performance level and rerun your workload to populate the cache.

Scenario 3: It is likely that your query is running slow due to reasons unrelated to the cache. [Troubleshoot](#) other areas which may be slowing down your queries. You can also consider [scaling down your instance](#) to reduce your cache size to save costs.

Scenario 4: You had a cold cache which could be the reason why your query was slow. Consider rerunning your query as your working dataset should now be in cached.

Since this is the ideal metric to monitor, all other options are incorrect

For more information on monitoring the cache, one can visit the below URL

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

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 49

Correct

Domain :Monitor and optimize data solutions

View Case Study

You need to fulfil the below requirement of the case study

"Applications with Tiers 6 through 8 must ensure that unexpected resource storage usage is immediately reported to IT data engineers."

Which of the following would you implement for this requirement?

- A. An alert rule that would be used to monitor CPU percentage for the database and then alert the IT Engineers
- B. An alert rule that would be used to monitor CPU percentage for the elastic pool and then alert the IT Engineers
- C. An alert rule that would be used to monitor storage percentage for the database and then alert the IT Engineers
- D. An alert rule that would be used to monitor storage percentage for the elastic pool and then alert the IT Engineers



Explanation:

Answer – D

Since the requirement asks for monitoring the storage, we should set this. Also, since the databases are going to be part of an elastic pool, we need to set it to monitor the percentage for the entire elastic pool.

For more information on working with alerts, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/azure-monitor/platform/alerts-metric>

[Ask our Experts](#)Rate this Question?  

View Queries

[open ▾](#)

Question 50

Correct**Domain :Implement data storage solutions**

You have to access Azure Blob Storage from Azure Databricks using secrets stored in a key vault. You already have the storage account, the blob container and Azure key vault in place.

You decide to implement the following steps

Add the secret to the storage container

Create a Databricks workspace and add the access keys

Access the blob container from Azure Databricks

Would these steps fulfil the requirement?

- A. Yes
- B. No

Explanation:

Answer – B

You need to add the secret to Azure Key vault and add the secret scope to the Databricks workspace

For more information on accessing Azure Blob storage from Azure Databricks using Azure Key vault, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/azure-databricks/store-secrets-azure-key-vault>

[Ask our Experts](#)Rate this Question?  

View Queries

[open ▾](#)

Question 51

Correct**Domain :Implement data storage solutions**

You have to access Azure Blob Storage from Azure Databricks using secrets stored in a key vault. You already have the storage account, the blob container and Azure key vault in place.

You decide to implement the following steps

Add the secret to the key vault

Create a Databricks workspace and add the secret scope

Access the blob container from Azure Databricks

Would these steps fulfil the requirement?

- A. Yes 
- B. No

Explanation:

Answer – A

Yes, this would fulfil the requirement. The Microsoft documentation mentions the following

Tutorial: Access Azure Blob Storage from Azure Databricks using Azure Key Vault

07/19/2019 • 5 minutes to read • 

This tutorial describes how to access Azure Blob Storage from Azure Databricks using secrets stored in a key vault.

In this tutorial, you learn how to:

- ✓ Create a storage account and blob container
- ✓ Create an Azure Key Vault and add a secret
- ✓ Create an Azure Databricks workspace and add a secret scope
- ✓ Access your blob container from Azure Databricks

For more information on accessing Azure Blob storage from Azure Databricks using Azure Key vault, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/azure-databricks/store-secrets-azure-key-vault>

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 52

Correct

Domain :Implement data storage solutions

You have to access Azure Blob Storage from Azure Databricks using secrets stored in a key vault. You already have the storage account, the blob container and Azure key vault in place.

You decide to implement the following steps

Add the secret to the key vault

Create a Databricks workspace and add the access keys

Access the blob container from Azure Databricks

Would these steps fulfil the requirement?

A. Yes

B. No

Explanation:

Answer – B

You are supposed to add a secret scope to the Databricks workspace and not the access keys

For more information on accessing Azure Blob storage from Azure Databricks using Azure Key vault, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/azure-databricks/store-secrets-azure-key-vault>

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 53

Correct

Domain :Manage and develop data processing

A company has created an Azure Data Lake Gen 2 storage account. They want to ingest data into the storage account from various data sources.

Which of the following can they use to ingest data from a relational data store?

- A. Azure Data Factory 
 - B. AzCopy Tool
 - C. Azure Event Hubs
 - D. Azure Event Grid
-

Explanation:

Answer – A

You can use Azure Data Factory for this requirement

The Microsoft documentation mentions the following

Relational data

You can also source data from relational databases. Over a period of time, relational databases collect huge amounts of data which can provide key insights if processed through a big data pipeline. You can use the following tools to move such data into Data Lake Storage Gen2.

Here's a list of tools that you can use to ingest relational data.

Tool	Guidance
Azure Data Factory	Copy Activity in Azure Data Factory

Since this is clearly mentioned in the Microsoft documentation, all other options are incorrect

For more information on data lake storage scenarios, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-data-scenarios>

Ask our Experts

Rate this Question?  

View Queries

open ▾

Question 54

Correct

Domain :Manage and develop data processing

A company has created an Azure Data Lake Gen 2 storage account. They want to ingest data into the storage account from various data sources.

Which of the following can they use to ingest data from a local workstation?

- A. Azure Data Factory
- B. AzCopy Tool 
- C. Azure Event Hubs
- D. Azure Event Grid

Explanation:

Answer – B

You can use the AzCopy tool for this requirement

The Microsoft documentation mentions the following

Ad hoc data

This represents smaller data sets that are used for prototyping a big data application. There are different ways of ingesting ad hoc data depending on the source of the data.

Here's a list of tools that you can use to ingest ad hoc data.

Data Source	Ingest it using
Local computer	Storage Explorer AzCopy tool
Azure Storage Blob	Azure Data Factory AzCopy tool
	DistCp running on HDInsight cluster

Since this is clearly mentioned in the Microsoft documentation, all other options are incorrect

For more information on data lake storage scenarios, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-data-scenarios>

Ask our Experts

Rate this Question?  

View Queries

open 

Question 55

Correct

Domain :Manage and develop data processing

A company has created an Azure Data Lake Gen 2 storage account. They want to ingest data into the storage account from various data sources.

Which of the following can they use to ingest data from log data stored on web servers?

- A. Azure Data Factory 
- B. AzCopy Tool
- C. Azure Event Hubs
- D. Azure Event Grid

Explanation:

Answer – A

You can use Azure Data Factory for this requirement

The Microsoft documentation mentions the following

Web server log data (upload using custom applications)

This type of dataset is specifically called out because analysis of web server log data is a common use case for big data applications and requires large volumes of log files to be uploaded to Data Lake Storage Gen2. You can use any of the following tools to write your own scripts or applications to upload such data.

Here's a list of tools that you can use to ingest Web server log data.

Tool	Guidance
Azure Data Factory	Copy Activity in Azure Data Factory

For uploading web server log data, and also for uploading other kinds of data (e.g. social sentiments data), it is a good approach to write your own custom scripts/applications because it gives you the flexibility to include your data uploading component as part of your larger big data application. In some cases this code may take the form of a script or simple command line utility. In other cases, the code may be used to integrate big data processing into a business application or solution.

Since this is clearly mentioned in the Microsoft documentation, all other options are incorrect

For more information on data lake storage scenarios, one can visit the below URL

<https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-data-scenarios>

Ask our Experts

Rate this Question?  

[View Queries](#)

open ▾

[Finish Review](#)

Certification

Cloud Certification

Java Certification

PM Certification

Big Data Certification

Company

Become Our Instructor

Support

Discussions

Blog

Business

Follow us

© Copyright 2020. Whizlabs Software Pvt. Ltd. All Right Reserved.