

## **DP-200.27q**

Number: DP-200  
Passing Score: 800  
Time Limit: 120 min

## **DP-200**

### **Implementing an Azure Data Solution (beta)**

## Question Set 1

### QUESTION 1

You develop data engineering solutions for a company. The company has on-premises Microsoft SQL Server databases at multiple locations.

The company must integrate data with Microsoft Power BI and Microsoft Azure Logic Apps. The solution must avoid single points of failure during connection and transfer to the cloud. The solution must also minimize latency.

You need to secure the transfer of data between on-premises databases and Microsoft Azure.

What should you do?

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

**Correct Answer:** D

**Section:** [none]

**Explanation**

#### **Explanation/Reference:**

Explanation:

You can create high availability clusters of On-premises data gateway installations, to ensure your organization can access on-premises data resources used in Power BI reports and dashboards. Such clusters allow gateway administrators to group gateways to avoid single points of failure in accessing on-premises data resources. The Power BI service always uses the primary gateway in the cluster, unless it's not available. In that case, the service switches to the next gateway in the cluster, and so on.

References:

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

### QUESTION 2

You are a data architect. The data engineering team needs to configure a synchronization of data between an on-premises Microsoft SQL Server database to Azure SQL Database.

Ad-hoc and reporting queries are being overutilized the on-premises production instance. The synchronization process must:

- Perform an initial data synchronization to Azure SQL Database with minimal downtime
- Perform bi-directional data synchronization after initial synchronization

You need to implement this synchronization solution.

Which synchronization method should you use?

- A. transactional replication
- B. Data Migration Assistant (DMA)
- C. backup and restore
- D. SQL Server Agent job
- E. Azure SQL Data Sync

**Correct Answer:** E

**Section:** [none]

**Explanation**

#### **Explanation/Reference:**

Explanation:

SQL Data Sync is a service built on Azure SQL Database that lets you synchronize the data you select bi-directionally across multiple SQL databases and SQL Server instances.

With Data Sync, you can keep data synchronized between your on-premises databases and Azure SQL databases to enable hybrid applications.

Compare Data Sync with Transactional Replication

	Data Sync	Transactional Replication
Advantages	<ul style="list-style-type: none"><li>- Active-active support</li><li>- Bi-directional between on-premises and Azure SQL Database</li></ul>	<ul style="list-style-type: none"><li>- Lower latency</li><li>- Transactional consistency</li><li>- Reuse existing topology after migration</li></ul>
Disadvantages	<ul style="list-style-type: none"><li>- 5 min or more latency</li><li>- No transactional consistency</li><li>- Higher performance impact</li></ul>	<ul style="list-style-type: none"><li>- Can't publish from Azure SQL Database single database or pooled database</li><li>- High maintenance cost</li></ul>

References:

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

### QUESTION 3

An application will use Microsoft Azure Cosmos DB as its data solution. The application will use the Cassandra API to support a column-based database type that uses containers to store items.

You need to provision Azure Cosmos DB. Which container name and item name should you use? Each correct answer presents part of the solutions.

**NOTE:** Each correct answer selection is worth one point.

- A. collection
- B. rows
- C. graph
- D. entities
- E. table

**Correct Answer:** BE

**Section:** [none]

**Explanation**

**Explanation/Reference:**

Explanation:

B: Depending on the choice of the API, an Azure Cosmos item can represent either a document in a collection, a row in a table or a node/edge in a graph. The following table shows the mapping between API-specific entities to an Azure Cosmos item:

Cosmos entity	SQL API	Cassandra API	Azure Cosmos DB's API for MongoDB	Gremlin API	Table API
Azure Cosmos item	Document	Row	Document	Node or Edge	Item

E: An Azure Cosmos container is specialized into API-specific entities as follows:

Azure Cosmos entity	SQL API	Cassandra API	Azure Cosmos DB's API for MongoDB	Gremlin API	Table API
Azure Cosmos container	Collection	Table	Collection	Graph	Table

References:

<https://docs.microsoft.com/en-us/azure/cosmos-db/databases-containers-items>

#### QUESTION 4

A company has a SaaS solution that uses Azure SQL Database with elastic pools. The solution contains a dedicated database for each customer organization. Customer organizations have peak usage at different periods during the year.

You need to implement the Azure SQL Database elastic pool to minimize cost.

Which option or options should you configure?

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

**Correct Answer:** E

**Section:** [none]

**Explanation**

#### Explanation/Reference:

Explanation:

The best size for a pool depends on the aggregate resources needed for all databases in the pool. This involves determining the following:

- Maximum resources utilized by all databases in the pool (either maximum DTUs or maximum vCores depending on your choice of resourcing model).
- Maximum storage bytes utilized by all databases in the pool.

Note: Elastic pools enable the developer to purchase resources for a pool shared by multiple databases to accommodate unpredictable periods of usage by individual databases. You can configure resources for the pool based either on the DTU-based purchasing model or the vCore-based purchasing model.

References:

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

#### QUESTION 5

A company manages several on-premises Microsoft SQL Server databases.

You need to migrate the databases to Microsoft Azure by using a backup and restore process.

Which data technology should you use?

- A. Azure SQL Database single database
- B. Azure SQL Data Warehouse
- C. Azure Cosmos DB
- D. Azure SQL Database Managed Instance

**Correct Answer:** D

**Section:** [none]

**Explanation**

**Explanation/Reference:**

Explanation:

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.

References:

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

**QUESTION 6**

The data engineering team manages Azure HDInsight clusters. The team spends a large amount of time creating and destroying clusters daily because most of the data pipeline process runs in minutes.

You need to implement a solution that deploys multiple HDInsight clusters with minimal effort.

What should you implement?

- A. Azure Databricks
- B. Azure Traffic Manager
- C. Azure Resource Manager templates
- D. Ambari web user interface

**Correct Answer: C**

**Section: [none]**

**Explanation**

**Explanation/Reference:**

Explanation:

A Resource Manager template makes it easy to create the following resources for your application in a single, coordinated operation:

- HDInsight clusters and their dependent resources (such as the default storage account).
- Other resources (such as Azure SQL Database to use Apache Sqoop).

In the template, you define the resources that are needed for the application. You also specify deployment parameters to input values for different environments. The template consists of JSON and expressions that you use to construct values for your deployment.

References:

<https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop-create-linux-clusters-arm-templates>

**QUESTION 7**

You are the data engineer for your company. An application uses a NoSQL database to store data. The database uses the key-value and wide-column NoSQL database type.

Developers need to access data in the database using an API.

You need to determine which API to use for the database model and type.

Which two APIs should you use? Each correct answer presents a complete solution.

**NOTE:** Each correct selection is worth one point.

- A. Table API
- B. MongoDB API
- C. Gremlin API
- D. SQL API
- E. Cassandra API

**Correct Answer: BE**

**Section: [none]**

**Explanation**

**Explanation/Reference:**

Explanation:

B: Azure Cosmos DB is the globally distributed, multimodel database service from Microsoft for mission-critical applications. It is a multimodel database and supports document, key-value, graph, and columnar data models.

E: Wide-column stores store data together as columns instead of rows and are optimized for queries over large datasets. The most popular are Cassandra and HBase.

References:

<https://docs.microsoft.com/en-us/azure/cosmos-db/graph-introduction>

<https://www.mongodb.com/scale/types-of-nosql-databases>

### **QUESTION 8**

A company is designing a hybrid solution to synchronize data and on-premises Microsoft SQL Server database to Azure SQL Database.

You must perform an assessment of databases to determine whether data will move without compatibility issues. You need to perform the assessment.

Which tool should you use?

- A. SQL Server Migration Assistant (SSMA)
- B. Microsoft Assessment and Planning Toolkit
- C. SQL Vulnerability Assessment (VA)
- D. Azure SQL Data Sync
- E. Data Migration Assistant (DMA)

**Correct Answer: E**

**Section: [none]**

**Explanation**

**Explanation/Reference:**

Explanation:

The Data Migration Assistant (DMA) helps you upgrade to a modern data platform by detecting compatibility issues that can impact database functionality in your new version of SQL Server or Azure SQL Database. DMA recommends performance and reliability improvements for your target environment and allows you to move your schema, data, and uncontained objects from your source server to your target server.

References:

<https://docs.microsoft.com/en-us/sql/dma/dma-overview>

## Question Set 1

### QUESTION 1

**Note: This question is part of series of questions that present the same scenario. Each question in the series contain a unique solution. Determine whether the solution meets the stated goals.**

You develop a data ingestion process that will import data to a Microsoft Azure SQL Data Warehouse. The data to be ingested resides in parquet files stored in an Azure Data Lake Gen 2 storage account.

You need to load the data from the Azure Data Lake Gen 2 storage account into the Azure SQL Data Warehouse.

Solution:

1. Create an external data source pointing to the Azure storage account
2. Create an external file format and external table using the external data source
3. Load the data using the `INSERT...SELECT` statement

Does the solution meet the goal?

- A. Yes
- B. No

**Correct Answer: B**

**Section: [none]**

**Explanation**

**Explanation/Reference:**

Explanation:

You load the data using the `CREATE TABLE AS SELECT` statement.

References:

<https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-load-from-azure-data-lake-store>

### QUESTION 2

**Note: This question is part of series of questions that present the same scenario. Each question in the series contain a unique solution. Determine whether the solution meets the stated goals.**

You develop a data ingestion process that will import data to a Microsoft Azure SQL Data Warehouse. The data to be ingested resides in parquet files stored in an Azure Data Lake Gen 2 storage account.

You need to load the data from the Azure Data Lake Gen 2 storage account into the Azure SQL Data Warehouse.

Solution:

1. Create an external data source pointing to the Azure storage account
2. Create a workload group using the Azure storage account name as the pool name
3. Load the data using the `INSERT...SELECT` statement

Does the solution meet the goal?

- A. Yes
- B. No

**Correct Answer: B**

**Section: [none]**

**Explanation**

**Explanation/Reference:**

Explanation:

You need to create an external file format and external table using the external data source. You then load the data using the `CREATE TABLE AS SELECT` statement.

References:

<https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-load-from-azure-data-lake-store>

### QUESTION 3

You develop data engineering solutions for a company.

You must integrate the company's on-premises Microsoft SQL Server data with Microsoft Azure SQL Database. Data must be transformed incrementally.

You need to implement the data integration solution.

Which tool should you use to configure a pipeline to copy data?

- A. Use the Copy Data tool with Blob storage linked service as the source
- B. Use Azure PowerShell with SQL Server linked service as a source
- C. Use Azure Data Factory UI with Blob storage linked service as a source
- D. Use the .NET Data Factory API with Blob storage linked service as the source

**Correct Answer: C**

**Section: [none]**

**Explanation**

#### **Explanation/Reference:**

Explanation:

The Integration Runtime is a customer managed data integration infrastructure used by Azure Data Factory to provide data integration capabilities across different network environments.

A linked service defines the information needed for Azure Data Factory to connect to a data resource. We have three resources in this scenario for which linked services are needed:

- On-premises SQL Server
- Azure Blob Storage
- Azure SQL database

Note: Azure Data Factory is a fully managed cloud-based data integration service that orchestrates and automates the movement and transformation of data. The key concept in the ADF model is pipeline. A pipeline is a logical grouping of Activities, each of which defines the actions to perform on the data contained in Datasets. Linked services are used to define the information needed for Data Factory to connect to the data resources.

References:

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

### QUESTION 4

#### HOTSPOT

A company runs Microsoft Dynamics CRM with Microsoft SQL Server on-premises. SQL Server Integration Services (SSIS) packages extract data from Dynamics CRM APIs, and load the data into a SQL Server data warehouse.

The datacenter is running out of capacity. Because of the network configuration, you must extract on-premises data to the cloud over https. You cannot open any additional ports. The solution must implement the least amount of effort.

You need to create the pipeline system.

Which component should you use? To answer, select the appropriate technology in the dialog box in the answer area.

**NOTE:** Each correct selection is worth one point.

**Hot Area:**



**Action****Technology**

Extract SQL data on-premises

Self-hosted integration runtime	✓
Azure-SSIS integration runtime	
Azure integration runtime	
Source	

Load SQL data warehouse

Self-hosted integration runtime	✓
Azure-SSIS integration runtime	
Azure integration runtime	
Sink	

**Correct Answer:****Action****Technology**

Extract SQL data on-premises

Self-hosted integration runtime	✓
Azure-SSIS integration runtime	
Azure integration runtime	
Source	

Load SQL data warehouse

Self-hosted integration runtime	✓
Azure-SSIS integration runtime	
Azure integration runtime	
Sink	

**Section: [none]****Explanation****Explanation/Reference:**

Explanation:

Box 1: Source

For Copy activity, it requires source and sink linked services to define the direction of data flow.

Copying between a cloud data source and a data source in private network: if either source or sink linked service points to a self-hosted IR, the copy activity is executed on that self-hosted Integration Runtime.

Box 2: Self-hosted integration runtime

A self-hosted integration runtime can run copy activities between a cloud data store and a data store in a private network, and it 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 on an on-premises machine or a virtual machine (VM) inside a private network.

References:

<https://docs.microsoft.com/en-us/azure/data-factory/create-self-hosted-integration-runtime>
**QUESTION 5**

You develop data engineering solutions for a company.

You need to ingest and visualize real-time Twitter data by using Microsoft Azure.

Which three technologies should you use? Each correct answer presents part of the solution.

**NOTE:** Each correct selection is worth one point.

- A. Event Grid topic
- B. Azure Stream Analytics Job that queries Twitter data from an Event Hub
- C. Azure Stream Analytics Job that queries Twitter data from an Event Grid
- D. Logic App that sends Twitter posts which have target keywords to Azure
- E. Event Grid subscription
- F. Event Hub instance

**Correct Answer:** BDF

**Section:** [none]

**Explanation**

**Explanation/Reference:**

Explanation:

You can use Azure Logic apps to send tweets to an event hub and then use a Stream Analytics job to read from event hub and send them to PowerBI.

References:

<https://community.powerbi.com/t5/Integrations-with-Files-and/Twitter-streaming-analytics-step-by-step/td-p/9594>

#### QUESTION 6

Each day, company plans to store hundreds of files in Azure Blob Storage and Azure Data Lake Storage. The company uses the parquet format.

You must develop a pipeline that meets the following requirements:

- Process data every six hours
- Offer interactive data analysis capabilities
- Offer the ability to process data using solid-state drive (SSD) caching
- Use Directed Acyclic Graph(DAG) processing mechanisms
- Provide support for REST API calls to monitor processes
- Provide native support for Python
- Integrate with Microsoft Power BI

You need to select the appropriate data technology to implement the pipeline.

Which data technology should you implement?

- A. Azure SQL Data Warehouse
- B. HDInsight Apache Storm cluster
- C. Azure Stream Analytics
- D. HDInsight Apache Hadoop cluster using MapReduce
- E. HDInsight Spark cluster

**Correct Answer:** B

**Section:** [none]

**Explanation**

**Explanation/Reference:**

Explanation:

Storm runs topologies instead of the Apache Hadoop MapReduce jobs that you might be familiar with. Storm topologies are composed of multiple components that are arranged in a directed acyclic graph (DAG). Data flows between the components in the graph. Each component consumes one or more data streams, and can optionally emit one or more streams.

Python can be used to develop Storm components.

References:

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

## Testlet 2

### Background

Proseware, Inc, develops and manages a product named Poll Taker. The product is used for delivering public opinion polling and analysis.

Polling data comes from a variety of sources, including online surveys, house-to-house interviews, and booths at public events.

### Polling data

Polling data is stored in one of the two locations:

- An on-premises Microsoft SQL Server 2019 database named PollingData
- Azure Data Lake Gen 2

Data in Data Lake is queried by using PolyBase

### Poll metadata

Each poll has associated metadata with information about the poll including the date and number of respondents. The data is stored as JSON.

### Phone-based polling

#### Security

- Phone-based poll data must only be uploaded by authorized users from authorized devices
- Contractors must not have access to any polling data other than their own
- Access to polling data must set on a per-active directory user basis

#### Data migration and loading

- All data migration processes must use Azure Data Factory
- All data migrations must run automatically during non-business hours
- Data migrations must be reliable and retry when needed

#### Performance

After six months, raw polling data should be moved to a lower-cost storage solution.

#### Deployments

- All deployments must be performed by using Azure DevOps. Deployments must use templates used in multiple environments
- No credentials or secrets should be used during deployments

#### Reliability

All services and processes must be resilient to a regional Azure outage.

#### Monitoring

All Azure services must be monitored by using Azure Monitor. On-premises SQL Server performance must be monitored.

### QUESTION 1

You need to ensure that phone-based polling data can be analyzed in the PollingData database.

How should you configure Azure Data Factory?

- A. Use a tumbling schedule trigger
- B. Use an event-based trigger
- C. Use a schedule trigger
- D. Use manual execution

**Correct Answer: C**

**Section: [none]**  
**Explanation**

**Explanation/Reference:**

Explanation:

When creating a schedule trigger, you specify a schedule (start date, recurrence, end date etc.) for the trigger, and associate with a Data Factory pipeline.

Scenario:

All data migration processes must use Azure Data Factory

All data migrations must run automatically during non-business hours

References:

<https://docs.microsoft.com/en-us/azure/data-factory/how-to-create-schedule-trigger>

### Testlet 3

#### Overview

##### Current environment

Contoso relies on an extensive partner network for marketing, sales, and distribution. Contoso uses external companies that manufacture everything from the actual pharmaceutical to the packaging.

The majority of the company's data reside in Microsoft SQL Server database. Application databases fall into one of the following tiers:

Applications	Tier	Replication	Notes
Internal Contoso	1	Yes	
Internal Contoso	2	SQL Data Sync	Data Sync between
Internal Partner	3	Yes	Replicate to Partn
External Contoso	4,5,6	Yes	
External Partner	7,8	No	Partner managed
Internal Distribution and Sales	9	Yes, once ingested at branches	Data ingested from branches
External Distribution and Sales	10	Yes, once ingested at Contoso main office	Data is ingested from sources

The company has a reporting infrastructure that ingests data from local databases and partner services. Partners services consists of distributors, wholesales, and retailers across the world. The company performs daily, weekly, and monthly reporting.

##### Requirements

Tier 3 and Tier 6 through Tier 8 application must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

- Tier 1 internal applications on the premium P2 tier
- Tier 2 internal applications on the standard S4 tier

The solution must support migrating databases that support external and internal application to Azure SQL Database. The migrated databases will be supported by Azure Data Factory pipelines for the continued movement, migration and updating of data both in the cloud and from local core business systems and repositories.

Tier 7 and Tier 8 partner access must be restricted to the database only.

In addition to default Azure backup behavior, Tier 4 and 5 databases must be on a backup strategy that performs a transaction log backup eve hour, a differential backup of databases every day and a full back up every week.

Back up strategies must be put in place for all other standalone Azure SQL Databases using Azure SQL-provided backup storage and capabilities.

##### Databases

Contoso requires their data estate to be designed and implemented in the Azure Cloud. Moving to the cloud must not inhibit access to or availability of data.

Databases:

Tier 1 Database must implement data masking using the following masking logic:

Data type	Masking requirement
A	Mask 4 or less string data type characters
B	Mask first letter and domain
C	Mask everything except characters at the beginning and end

Tier 2 databases must sync between branches and cloud databases and in the event of conflicts must be set up for conflicts to be won by on-premises databases.

Tier 3 and Tier 6 through Tier 8 applications must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of a server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

- Tier 1 internal applications on the premium P2 tier
- Tier 2 internal applications on the standard S4 tier

## Reporting

### Security and monitoring

#### Security

A method of managing multiple databases in the cloud at the same time is must be implemented to streamlining data management and limiting management access to only those requiring access.

#### Monitoring

Monitoring must be set up on every database. Contoso and partners must receive performance reports as part of contractual agreements.

Tiers 6 through 8 must have unexpected resource storage usage immediately reported to data engineers.

The Azure SQL Data Warehouse cache must be monitored when the database is being used. A dashboard monitoring key performance indicators (KPIs) indicated by traffic lights must be created and displayed based on the following metrics:

Metric	Description
A	Low cache hit %, high cache usage %
B	Low cache hit %, low cache usage %
C	High cache hit %, high cache usage %

Existing Data Protection and Security compliances require that all certificates and keys are internally managed in an on-premises storage.

You identify the following reporting requirements:

- Azure Data Warehouse must be used to gather and query data from multiple internal and external databases
- Azure Data Warehouse must be optimized to use data from a cache
- Reporting data aggregated for external partners must be stored in Azure Storage and be made available during regular business hours in the connecting regions
- Reporting strategies must be improved to real time or near real time reporting cadence to improve competitiveness and the general supply chain
- Tier 9 reporting must be moved to Event Hubs, queried, and persisted in the same Azure region as the company's main office
- Tier 10 reporting data must be stored in Azure Blobs

## Issues

Team members identify the following issues:

- Both internal and external client application run complex joins, equality searches and group-by clauses. Because some systems are managed externally, the queries will not be changed or optimized by Contoso
- External partner organization data formats, types and schemas are controlled by the partner companies
- Internal and external database development staff resources are primarily SQL developers familiar with the Transact-SQL language.
- Size and amount of data has led to applications and reporting solutions not performing are required speeds
- Tier 7 and 8 data access is constrained to single endpoints managed by partners for access
- The company maintains several legacy client applications. Data for these applications remains isolated from other applications. This has led to hundreds of databases being provisioned on a per application basis

## QUESTION 1

You need to process and query ingested Tier 9 data.

Which two options should you use? Each correct answer presents part of the solution.

**NOTE:** Each correct selection is worth one point.

- A. Azure Notification Hub
- B. Transact-SQL statements
- C. Azure Cache for Redis
- D. Apache Kafka statements
- E. Azure Event Grid
- F. Azure Stream Analytics

**Correct Answer:** EF

**Section:** [none]

**Explanation**

### Explanation/Reference:

Explanation:

Event Hubs provides a Kafka endpoint that can be used by your existing Kafka based applications as an alternative to running your own Kafka cluster.

You can stream data into Kafka-enabled Event Hubs and process it with Azure Stream Analytics, in the following steps:

- Create a Kafka enabled Event Hubs namespace.
- Create a Kafka client that sends messages to the event hub.
- Create a Stream Analytics job that copies data from the event hub into an Azure blob storage.

Scenario:

Internal Distribution and Sales	9	Yes, once ingested at branches	Data ingested from Contoso branches
---------------------------------	---	--------------------------------	-------------------------------------

Tier 9 reporting must be moved to Event Hubs, queried, and persisted in the same Azure region as the



company's main office

References:

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

## QUESTION 2

You need to set up Azure Data Factory pipelines to meet data movement requirements.

Which integration runtime should you use?

- A. self-hosted integration runtime
- B. Azure-SSIS Integration Runtime
- C. .NET Common Language Runtime (CLR)
- D. Azure integration runtime

**Correct Answer: A**

**Section: [none]**

**Explanation**

**Explanation/Reference:**

Explanation:

The following table describes the capabilities and network support for each of the integration runtime types:

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

Scenario: The solution must support migrating databases that support external and internal application to Azure SQL Database. The migrated databases will be supported by Azure Data Factory pipelines for the continued movement, migration and updating of data both in the cloud and from local core business systems and repositories.

References:

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

## Testlet 1

### Overview

#### Current environment

Contoso relies on an extensive partner network for marketing, sales, and distribution. Contoso uses external companies that manufacture everything from the actual pharmaceutical to the packaging.

The majority of the company's data reside in Microsoft SQL Server database. Application databases fall into one of the following tiers:

Applications	Tier	Replication	Notes
Internal Contoso	1	Yes	
Internal Contoso	2	SQL Data Sync	Data Sync between d
Internal Partner	3	Yes	Replicate to Partner
External Contoso	4,5,6	Yes	
External Partner	7,8	No	Partner managed
Internal Distribution and Sales	9	Yes, once ingested at branches	Data ingested from C branches
External Distribution and Sales	10	Yes, once ingested at Contoso main office	Data is ingested from sources

The company has a reporting infrastructure that ingests data from local databases and partner services. Partners services consists of distributors, wholesales, and retailers across the world. The company performs daily, weekly, and monthly reporting.

#### Requirements

Tier 3 and Tier 6 through Tier 8 application must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

- Tier 1 internal applications on the premium P2 tier
- Tier 2 internal applications on the standard S4 tier

The solution must support migrating databases that support external and internal application to Azure SQL Database. The migrated databases will be supported by Azure Data Factory pipelines for the continued movement, migration and updating of data both in the cloud and from local core business systems and repositories.

Tier 7 and Tier 8 partner access must be restricted to the database only.

In addition to default Azure backup behavior, Tier 4 and 5 databases must be on a backup strategy that performs a transaction log backup eve hour, a differential backup of databases every day and a full back up every week.

Back up strategies must be put in place for all other standalone Azure SQL Databases using Azure SQL-provided backup storage and capabilities.

#### Databases

Contoso requires their data estate to be designed and implemented in the Azure Cloud. Moving to the cloud

must not inhibit access to or availability of data.

Databases:

Tier 1 Database must implement data masking using the following masking logic:

Data type	Masking requirement
A	Mask 4 or less string data type characters
B	Mask first letter and domain
C	Mask everything except characters at the beginning and end

Tier 2 databases must sync between branches and cloud databases and in the event of conflicts must be set up for conflicts to be won by on-premises databases.

Tier 3 and Tier 6 through Tier 8 applications must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of a server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

- Tier 1 internal applications on the premium P2 tier
- Tier 2 internal applications on the standard S4 tier

## Reporting

### Security and monitoring

#### Security

A method of managing multiple databases in the cloud at the same time is must be implemented to streamlining data management and limiting management access to only those requiring access.

#### Monitoring

Monitoring must be set up on every database. Contoso and partners must receive performance reports as part of contractual agreements.

Tiers 6 through 8 must have unexpected resource storage usage immediately reported to data engineers.

The Azure SQL Data Warehouse cache must be monitored when the database is being used. A dashboard monitoring key performance indicators (KPIs) indicated by traffic lights must be created and displayed based on the following metrics:

Metric	Description
A	Low cache hit %, high cache usage %
B	Low cache hit %, low cache usage %
C	High cache hit %, high cache usage %

Existing Data Protection and Security compliances require that all certificates and keys are internally managed in an on-premises storage.

You identify the following reporting requirements:

- Azure Data Warehouse must be used to gather and query data from multiple internal and external databases
- Azure Data Warehouse must be optimized to use data from a cache
- Reporting data aggregated for external partners must be stored in Azure Storage and be made available during regular business hours in the connecting regions
- Reporting strategies must be improved to real time or near real time reporting cadence to improve competitiveness and the general supply chain
- Tier 9 reporting must be moved to Event Hubs, queried, and persisted in the same Azure region as the company's main office
- Tier 10 reporting data must be stored in Azure Blobs

### Issues

Team members identify the following issues:

- Both internal and external client application run complex joins, equality searches and group-by clauses. Because some systems are managed externally, the queries will not be changed or optimized by Contoso
- External partner organization data formats, types and schemas are controlled by the partner companies
- Internal and external database development staff resources are primarily SQL developers familiar with the Transact-SQL language.
- Size and amount of data has led to applications and reporting solutions not performing as required speeds
- Tier 7 and 8 data access is constrained to single endpoints managed by partners for access
- The company maintains several legacy client applications. Data for these applications remains isolated from other applications. This has led to hundreds of databases being provisioned on a per application basis

### QUESTION 1

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

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

You need to configure data encryption for external applications.

Solution:

1. Access the Always Encrypted Wizard in SQL Server Management Studio
2. Select the column to be encrypted
3. Set the encryption type to Randomized
4. Configure the master key to use the Windows Certificate Store
5. Validate configuration results and deploy the solution

Does the solution meet the goal?

- A. Yes
- B. No

**Correct Answer: B**

**Section: [none]**

**Explanation**

**Explanation/Reference:**

Explanation:

Use the Azure Key Vault, not the Windows Certificate Store, to store the master key.

Note: The Master Key Configuration page is where you set up your CMK (Column Master Key) and select the key store provider where the CMK will be stored. Currently, you can store a CMK in the Windows certificate store, Azure Key Vault, or a hardware security module (HSM).

Always Encrypted

## Master Key Configuration

Introduction  
Column Selection  
**Master Key Configuration**  
Validation  
Summary  
Results

Help

To generate a new column encryption key, a column master key must be selected to protect it. The column master key is stored outside of the database.

Select column master key:

Auto generate column master key

Select the key store provider

☐ Windows certificate store

☒ Azure Key Vault

You are signed in as sstein@microsoft.com. [Change user](#)

Select a subscription to use:

Select an Azure Key Vault:

AeKeyVault

< Previous Next > Cancel

References:

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

### QUESTION 2

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

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

You need to configure data encryption for external applications.

Solution:

1. Access the Always Encrypted Wizard in SQL Server Management Studio
2. Select the column to be encrypted
3. Set the encryption type to Deterministic
4. Configure the master key to use the Windows Certificate Store
5. Validate configuration results and deploy the solution

Does the solution meet the goal?

- A. Yes
- B. No

**Correct Answer:** B  
**Section:** [none]  
**Explanation**

**Explanation/Reference:**

Explanation:

Use the Azure Key Vault, not the Windows Certificate Store, to store the master key.

Note: The Master Key Configuration page is where you set up your CMK (Column Master Key) and select the key store provider where the CMK will be stored. Currently, you can store a CMK in the Windows certificate store, Azure Key Vault, or a hardware security module (HSM).

Always Encrypted

## Master Key Configuration

Introduction  
Column Selection  
**Master Key Configuration**  
Validation  
Summary  
Results

Help

To generate a new column encryption key, a column master key must be selected to protect it. The column master key is stored outside of the database.

Select column master key:

Auto generate column master key

Select the key store provider

☐ Windows certificate store

☒ Azure Key Vault

You are signed in as sstein@microsoft.com. [Change user](#)

Select a subscription to use:

Select an Azure Key Vault:

AeKeyVault

< Previous Next > Cancel

References:

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

**QUESTION 3**

**Note:** This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some questions sets might have more than one correct solution, while others might not have a correct solution. After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You need to configure data encryption for external applications.

Solution:

1. Access the Always Encrypted Wizard in SQL Server Management Studio

2. Select the column to be encrypted
3. Set the encryption type to Deterministic
4. Configure the master key to use the Azure Key Vault
5. Validate configuration results and deploy the solution

Does the solution meet the goal?

- A. Yes
- B. No

**Correct Answer: A**

**Section: [none]**

**Explanation**

**Explanation/Reference:**

Explanation:

We use the Azure Key Vault, not the Windows Certificate Store, to store the master key.

Note: The Master Key Configuration page is where you set up your CMK (Column Master Key) and select the key store provider where the CMK will be stored. Currently, you can store a CMK in the Windows certificate store, Azure Key Vault, or a hardware security module (HSM).

Always Encrypted

## Master Key Configuration

Introduction  
Column Selection  
**Master Key Configuration**  
Validation  
Summary  
Results

Help

To generate a new column encryption key, a column master key must be selected to protect it. The column master key is stored outside of the database.

Select column master key:

Auto generate column master key

Select the key store provider

☐ Windows certificate store

☒ Azure Key Vault

You are signed in as sstein@microsoft.com. [Change user](#)

Select a subscription to use:

Select an Azure Key Vault:

AeKeyVault

< Previous Next > Cancel

References:

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

## Question Set 1

### QUESTION 1

**Note: This question is part of series of questions that present the same scenario. Each question in the series contain a unique solution. Determine whether the solution meets the stated goals.**

You develop data engineering solutions for a company.

A project requires the deployment of resources to Microsoft Azure for batch data processing on Azure HDInsight. Batch processing will run daily and must:

- Scale to minimize costs
- Be monitored for cluster performance

You need to recommend a tool that will monitor clusters and provide information to suggest how to scale.

Solution: Monitor cluster load using the Ambari Web UI.

Does the solution meet the goal?

- A. Yes
- B. No

**Correct Answer: B**

**Section: [none]**

**Explanation**

**Explanation/Reference:**

Explanation:

Ambari Web UI does not provide information to suggest how to scale.

Instead monitor clusters by using Azure Log Analytics and HDInsight cluster management solutions.

References:

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

<https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop-manage-ambari>

### QUESTION 2

**Note: This question is part of series of questions that present the same scenario. Each question in the series contain a unique solution. Determine whether the solution meets the stated goals.**

You develop data engineering solutions for a company.

A project requires the deployment of resources to Microsoft Azure for batch data processing on Azure HDInsight. Batch processing will run daily and must:

- Scale to minimize costs
- Be monitored for cluster performance

You need to recommend a tool that will monitor clusters and provide information to suggest how to scale.

Solution: Monitor clusters by using Azure Log Analytics and HDInsight cluster management solutions.

Does the solution meet the goal?

- A. Yes
- B. No

**Correct Answer: A**

**Section: [none]**

**Explanation**



**Explanation/Reference:**

Explanation:

HDInsight provides cluster-specific management solutions that you can add for Azure Monitor logs. Management solutions add functionality to Azure Monitor logs, providing additional data and analysis tools. These solutions collect important performance metrics from your HDInsight clusters and provide the tools to search the metrics. These solutions also provide visualizations and dashboards for most cluster types supported in HDInsight. By using the metrics that you collect with the solution, you can create custom monitoring rules and alerts.

References:

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

**QUESTION 3**

**Note: This question is part of series of questions that present the same scenario. Each question in the series contain a unique solution. Determine whether the solution meets the stated goals.**

You develop data engineering solutions for a company.

A project requires the deployment of resources to Microsoft Azure for batch data processing on Azure HDInsight. Batch processing will run daily and must:

- Scale to minimize costs
- Be monitored for cluster performance

You need to recommend a tool that will monitor clusters and provide information to suggest how to scale.

Solution: Download Azure HDInsight cluster logs by using Azure PowerShell.

Does the solution meet the goal?

- A. Yes
- B. No

**Correct Answer: B**

**Section: [none]**

**Explanation**

**Explanation/Reference:**

Explanation:

Instead monitor clusters by using Azure Log Analytics and HDInsight cluster management solutions.

References:

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

**QUESTION 4**

A company has a Microsoft Azure HDInsight solution that uses different cluster types to process and analyze data. Operations are continuous.

Reports indicate slowdowns during a specific time window.

You need to determine a monitoring solution to track down the issue in the least amount of time.

What should you use?

- A. Azure Log Analytics log search query
- B. Ambari REST API
- C. Azure Monitor Metrics
- D. HDInsight .NET SDK
- E. Azure Log Analytics alert rule query

**Correct Answer: B**

**Section: [none]**

## Explanation

### Explanation/Reference:

Explanation:

Ambari is the recommended tool for monitoring the health for any given HDInsight cluster.

Note: Azure HDInsight is a high-availability service that has redundant gateway nodes, head nodes, and ZooKeeper nodes to keep your HDInsight clusters running smoothly. While this ensures that a single failure will not affect the functionality of a cluster, you may still want to monitor cluster health so you are alerted when an issue does arise. Monitoring cluster health refers to monitoring whether all nodes in your cluster and the components that run on them are available and functioning correctly.

Ambari is the recommended tool for monitoring utilization across the whole cluster. The Ambari dashboard shows easily glanceable widgets that display metrics such as CPU, network, YARN memory, and HDFS disk usage. The specific metrics shown depend on cluster type. The "Hosts" tab shows metrics for individual nodes so you can ensure the load on your cluster is evenly distributed.

References:

<https://azure.microsoft.com/en-us/blog/monitoring-on-hdinsight-part-1-an-overview/>

## QUESTION 5

You manage a solution that uses Azure HDInsight clusters.

You need to implement a solution to monitor cluster performance and status.

Which technology should you use?

- A. Azure HDInsight .NET SDK
- B. Azure HDInsight REST API
- C. Ambari REST API
- D. Azure Log Analytics
- E. Ambari Web UI

**Correct Answer:** E

**Section:** [none]

**Explanation**

### Explanation/Reference:

Explanation:

Ambari is the recommended tool for monitoring utilization across the whole cluster. The Ambari dashboard shows easily glanceable widgets that display metrics such as CPU, network, YARN memory, and HDFS disk usage. The specific metrics shown depend on cluster type. The "Hosts" tab shows metrics for individual nodes so you can ensure the load on your cluster is evenly distributed.

The Apache Ambari project is aimed at making Hadoop management simpler by developing software for provisioning, managing, and monitoring Apache Hadoop clusters. Ambari provides an intuitive, easy-to-use Hadoop management web UI backed by its RESTful APIs.

References:

<https://azure.microsoft.com/en-us/blog/monitoring-on-hdinsight-part-1-an-overview/>

<https://ambari.apache.org/>

## Testlet 2

### Overview

#### Current environment

Contoso relies on an extensive partner network for marketing, sales, and distribution. Contoso uses external companies that manufacture everything from the actual pharmaceutical to the packaging.

The majority of the company's data reside in Microsoft SQL Server database. Application databases fall into one of the following tiers:

Applications	Tier	Replication	Notes
Internal Contoso	1	Yes	
Internal Contoso	2	SQL Data Sync	Data Sync between data
Internal Partner	3	Yes	Replicate to Partner
External Contoso	4,5,6	Yes	
External Partner	7,8	No	Partner managed
Internal Distribution and Sales	9	Yes, once ingested at branches	Data ingested from Contoso branches
External Distribution and Sales	10	Yes, once ingested at Contoso main office	Data is ingested from multiple sources

The company has a reporting infrastructure that ingests data from local databases and partner services. Partners services consists of distributors, wholesales, and retailers across the world. The company performs daily, weekly, and monthly reporting.

#### Requirements

Tier 3 and Tier 6 through Tier 8 application must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

- Tier 1 internal applications on the premium P2 tier
- Tier 2 internal applications on the standard S4 tier

The solution must support migrating databases that support external and internal application to Azure SQL Database. The migrated databases will be supported by Azure Data Factory pipelines for the continued movement, migration and updating of data both in the cloud and from local core business systems and repositories.

Tier 7 and Tier 8 partner access must be restricted to the database only.

In addition to default Azure backup behavior, Tier 4 and 5 databases must be on a backup strategy that performs a transaction log backup every hour, a differential backup of databases every day and a full back up every week.

Back up strategies must be put in place for all other standalone Azure SQL Databases using Azure SQL-provided backup storage and capabilities.

#### Databases

Contoso requires their data estate to be designed and implemented in the Azure Cloud. Moving to the cloud must not inhibit access to or availability of data.

Databases:

Tier 1 Database must implement data masking using the following masking logic:

Data type	Masking requirement
A	Mask 4 or less string data type characters
B	Mask first letter and domain
C	Mask everything except characters at the beginning and end

Tier 2 databases must sync between branches and cloud databases and in the event of conflicts must be set up for conflicts to be won by on-premises databases.

Tier 3 and Tier 6 through Tier 8 applications must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of a server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

- Tier 1 internal applications on the premium P2 tier
- Tier 2 internal applications on the standard S4 tier

## Reporting

### Security and monitoring

#### Security

A method of managing multiple databases in the cloud at the same time is must be implemented to streamlining data management and limiting management access to only those requiring access.

#### Monitoring

Monitoring must be set up on every database. Contoso and partners must receive performance reports as part of contractual agreements.

Tiers 6 through 8 must have unexpected resource storage usage immediately reported to data engineers.

The Azure SQL Data Warehouse cache must be monitored when the database is being used. A dashboard monitoring key performance indicators (KPIs) indicated by traffic lights must be created and displayed based on the following metrics:

Metric	Description
A	Low cache hit %, high cache usage %
B	Low cache hit %, low cache usage %
C	High cache hit %, high cache usage %

Existing Data Protection and Security compliances require that all certificates and keys are internally

managed in an on-premises storage.

You identify the following reporting requirements:

- Azure Data Warehouse must be used to gather and query data from multiple internal and external databases
- Azure Data Warehouse must be optimized to use data from a cache
- Reporting data aggregated for external partners must be stored in Azure Storage and be made available during regular business hours in the connecting regions
- Reporting strategies must be improved to real time or near real time reporting cadence to improve competitiveness and the general supply chain
- Tier 9 reporting must be moved to Event Hubs, queried, and persisted in the same Azure region as the company's main office
- Tier 10 reporting data must be stored in Azure Blobs

### Issues

Team members identify the following issues:

- Both internal and external client application run complex joins, equality searches and group-by clauses. Because some systems are managed externally, the queries will not be changed or optimized by Contoso
- External partner organization data formats, types and schemas are controlled by the partner companies
- Internal and external database development staff resources are primarily SQL developers familiar with the Transact-SQL language.
- Size and amount of data has led to applications and reporting solutions not performing as required speeds
- Tier 7 and 8 data access is constrained to single endpoints managed by partners for access
- The company maintains several legacy client applications. Data for these applications remains isolated from other applications. This has led to hundreds of databases being provisioned on a per application basis

### QUESTION 1

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

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

You need to implement diagnostic logging for Data Warehouse monitoring.

Which log should you use?

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

**Correct Answer: C**

**Section: [none]**

**Explanation**

**Explanation/Reference:**

Explanation:

Scenario:

The Azure SQL Data Warehouse cache must be monitored when the database is being used.

Metric	Description
A	Low cache hit %, high cache usage %
B	Low cache hit %, low cache usage %
C	High cache hit %, high cache usage %

References:

<https://docs.microsoft.com/en-us/sql/relational-databases/system-dynamic-management-views/sys-dm-pdw-sql-requests-transact-sq>

## QUESTION 2

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

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

You need setup monitoring for tiers 6 through 8.

What should you configure?

- A. extended events for average storage percentage that emails data engineers
- B. an alert rule to monitor CPU percentage in databases that emails data engineers
- C. an alert rule to monitor CPU percentage in elastic pools that emails data engineers
- D. an alert rule to monitor storage percentage in databases that emails data engineers
- E. an alert rule to monitor storage percentage in elastic pools that emails data engineers

**Correct Answer:** E

**Section:** [none]

**Explanation**

**Explanation/Reference:**

Explanation:

Scenario:

Tiers 6 through 8 must have unexpected resource storage usage immediately reported to data engineers.

Tier 3 and Tier 6 through Tier 8 applications must use database density on the same server and Elastic pools in a cost-effective manner.