



DP- 200.examcollection.premium.exam.120q DP-200

DP-200

Implementing an Azure Data Solution

Version 3.0

Score: 800/1000

Version:

Time Limit: 120 Minutes

Implement data storage solutions

(30 questions)

Question 1

A company runs Microsoft SQL Server in an on-premises virtual machine (VM).

You must migrate the database to Azure SQL Database. You synchronize users from Active Directory to Azure Active Directory (Azure AD).

You need to configure Azure SQL Database to use an Azure AD user as administrator.

What should you configure?

- ☐ For each Azure SQL Database, set the Access Control to administrator.
- ☐ For each Azure SQL Database server, set the Active Directory to administrator.
- ☐ For each Azure SQL Database, set the Active Directory administrator role.
- ☐ For each Azure SQL Database server, set the Access Control to administrator.

Question 2

You have an Azure SQL data warehouse.

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

The external table has three columns.

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

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

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

- ☐ Option A
- ☐ Option B
- ☐ Option C
- ☐ Option D

Question 3

A company plans to use Azure Storage for file storage purposes. Compliance rules require:

A single storage account to store all operations including reads, writes and deletes
Retention of an on-premises copy of historical operations

You need to configure the storage account.

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

NOTE: Each correct selection is worth one point.

- ☐ Configure the storage account to log read, write and delete operations for service type Blob
- ☐ Use the AzCopy tool to download log data from \$logs/blob
- ☐ Configure the storage account to log read, write and delete operations for service-type table
- ☐ Use the storage client to download log data from \$logs/table
- ☐ Configure the storage account to log read, write and delete operations for service type queue

Question 4

You are a data engineer implementing a lambda architecture on Microsoft Azure. You use an open-source big data solution to collect, process, and maintain data. The analytical data store performs poorly.

You must implement a solution that meets the following requirements:

Provide data warehousing
Reduce ongoing management activities
Deliver SQL query responses in less than one second

You need to create an HDInsight cluster to meet the requirements.

Which type of cluster should you create?

- ☐ Interactive Query
- ☐ Apache Hadoop
- ☐ Apache HBase
- ☐ Apache Spark

Question 5

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

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

You have an Azure SQL database named DB1 that contains a table named Table1. Table1 has a field named Customer_ID that is varchar(22).

You need to implement masking for the Customer_ID field to meet the following requirements:

The first two prefix characters must be exposed.

The last four prefix characters must be exposed.

All other characters must be masked.

Solution: You implement data masking and use a random number function mask.

Does this meet the goal?

- ☐ Yes
- ☐ No

Question 6

You plan to create a dimension table in Azure Data Warehouse that will be less than 1 GB.

You need to create the table to meet the following requirements:

Provide the fastest query time.

Minimize data movement.

Which type of table should you use?

- ☐ hash distributed
- ☐ heap
- ☐ replicated
- ☐ round-robin

Question 7

You are developing a data engineering solution for a company. The solution will store a large set of key-value pair data by using Microsoft Azure Cosmos DB.

The solution has the following requirements:

Data must be partitioned into multiple containers.

Data containers must be configured separately.

Data must be accessible from applications hosted around the world.

The solution must minimize latency.

You need to provision Azure Cosmos DB.

- ☐ Cosmos account-level throughput.
- ☐ Provision an Azure Cosmos DB account with the Azure Table API. Enable geo-redundancy.
- ☐ Configure table-level throughput.
- ☐ Replicate the data globally by manually adding regions to the Azure Cosmos DB account.
- ☐ Provision an Azure Cosmos DB account with the Azure Table API. Enable multi-region writes.

Question 8

DRAG DROP

You manage a financial computation data analysis process. Microsoft Azure virtual machines (VMs) run the process in daily jobs, and store the results in virtual hard drives (VHDs.)

The VMs product results using data from the previous day and store the results in a snapshot of the VHD. When a new month begins, a process creates a new VHD.

You must implement the following data retention requirements:

Daily results must be kept for 90 days

Data for the current year must be available for weekly reports

Data from the previous 10 years must be stored for auditing purposes

Data required for an audit must be produced within 10 days of a request.

You need to enforce the data retention requirements while minimizing cost.

How should you configure the lifecycle policy? To answer, drag the appropriate JSON segments to the correct locations. Each JSON segment may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Code segments

Answer Area

```

{
  "version": "0.5",
  "rules": [
    {
      "name": "dataRetention",
      "type": "Lifecycle",
      "definition": {
        "actions": {
          "  ": {
            "  ": {daysAfterModificationGreaterThan": 365},
            "  ": {daysAfterModificationGreaterThan": 3650}
          },
          "  ": {
            "  ": {"daysAfterCreationGreaterThan": 90}
          }
        }
      }
    }
  ]
}

```

Question 9

DRAG DROP

You are developing a solution to visualize multiple terabytes of geospatial data.

The solution has the following requirements:

Data must be encrypted.

Data must be accessible by multiple resources on Microsoft Azure.

You need to provision storage for the solution.

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

Actions

Enable encryption on the Azure Data Lake using the Azure portal.

Add an access policy for the new Azure Data Lake account to the key storage container.

Create a new Azure Data Lake Storage account with Azure Data Lake managed encryption keys.

Select and configure an encryption key storage container.

Create a new Azure Data Lake Storage account with Azure Key Vault managed encryption keys.

Create a new Azure Data Lake Storage account with encryption disabled.

Answer Area



Question 10

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

Which two factors affect your costs when sizing the Azure SQL Database elastic pools? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

- ☐ maximum data size
- ☐ number of databases
- ☐ eDTUs consumption
- ☐ number of read operations
- ☐ number of transactions

Question 11

DRAG DROP

You are responsible for providing access to an Azure Data Lake Storage Gen2 account.

Your user account has contributor access to the storage account, and you have the application ID access key.

You plan to use PolyBase to load data into Azure SQL data warehouse.

You need to configure PolyBase to connect the data warehouse to the storage account.

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

Components

a database encryption key

an asymmetric key

an external data source

an external file format

a database scoped credential

Answer Area

Question 12

DRAG DROP

You develop data engineering solutions for a company. You must migrate data from Microsoft Azure Blob storage to an Azure SQL Data Warehouse for further transformation. You need to implement the solution.

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

Actions

Provision an Azure SQL Data Warehouse instance.

Connect to the Blob storage container by using SQL Server Management Studio.

Provision an Azure Blob storage container.

Run Transact-SQL statements to load data.

Connect to the Azure SQL Data Warehouse by using SQL Server Management Studio.

Build external tables by using Azure portal.

Build external tables by using SQL Server Management Studio.

Answer Area

Question 13

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?

- ☐ Azure Databricks
- ☐ Azure Traffic Manager
- ☐ Azure Resource Manager templates
- ☐ Ambari web user interface

Question 14

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.

- ☐ collection
- ☐ rows
- ☐ graph
- ☐ entities
- ☐ table

Question 15

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

You need to migrate the databases to Microsoft Azure by using a backup process of Microsoft SQL Server.

Which data technology should you use?

- ☐ Azure SQL Database single database
- ☐ Azure SQL Data Warehouse
- ☐ Azure Cosmos DB
- ☐ Azure SQL Database Managed Instance

Question 16

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?

- ☐ Install a standalone on-premises Azure data gateway at each location
- ☐ Install an on-premises data gateway in personal mode at each location
- ☐ Install an Azure on-premises data gateway at the primary location
- ☐ Install an Azure on-premises data gateway as a cluster at each location

Question 17

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

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

You have an Azure SQL database named DB1 that contains a table named Table1. Table1 has a field named Customer_ID that is varchar(22).

You need to implement masking for the Customer_ID field to meet the following requirements:

The first two prefix characters must be exposed.

The last four prefix characters must be exposed.

All other characters must be masked.

Solution: You implement data masking and use a credit card function mask.

Does this meet the goal?

- ☐ Yes
- ☐ No

Question 18

DRAG DROP

You have a table named SalesFact in an Azure SQL data warehouse. SalesFact contains sales data from the past 36 months and has the following characteristics:

Is partitioned by month
Contains one billion rows
Has clustered columnstore indexes

All the beginning of each month, you need to remove data SalesFact that is older than 36 months as quickly as possible.

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

Actions

Create an empty table named SalesFact_Work that has the same schema as SalesFact.

Drop the SalesFact_Work table.

Copy the data to a new table by using CREATE TABLE AS SELECT (CTAS).

Truncate the partition containing the stale data.

Switch the partition containing the stale data from SalesFact to SalesFact_Work.

Execute a DELETE statement where the value in the Date column is more than 36 months ago.

Answer Area

Question 19

HOTSPOT

You are a data engineer. You are designing a Hadoop Distributed File System (HDFS) architecture. You plan to use Microsoft Azure Data Lake as a data storage repository.

You must provision the repository with a resilient data schema. You need to ensure the resiliency of the Azure Data Lake Storage. What should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Requirement	Node		
Provide data access to clients.	<table> <tr> <td data-bbox="1091 976 1289 1061">DataNode NameNode</td><td data-bbox="1297 976 1331 1061">▼</td></tr> </table>	DataNode NameNode	▼
DataNode NameNode	▼		
Run operations on files and directories of the file system.	<table> <tr> <td data-bbox="1091 1117 1289 1202">DataNode NameNode</td><td data-bbox="1297 1117 1331 1202">▼</td></tr> </table>	DataNode NameNode	▼
DataNode NameNode	▼		
Perform block creation, deletion, and replication.	<table> <tr> <td data-bbox="1091 1258 1289 1344">DataNode NameNode</td><td data-bbox="1297 1258 1331 1344">▼</td></tr> </table>	DataNode NameNode	▼
DataNode NameNode	▼		

Question 20

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

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

You have an Azure SQL database named DB1 that contains a table named Table1. Table1 has a field named Customer_ID that is varchar(22).

You need to implement masking for the Customer_ID field to meet the following requirements:

The first two prefix characters must be exposed.

The last four prefix characters must be exposed.

All other characters must be masked.

Solution: You implement data masking and use a credit card function mask.

Does this meet the goal?

- ☐ Yes
- ☐ No

Question 21

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?

- ☐ SQL Server Migration Assistant (SSMA)
- ☐ Microsoft Assessment and Planning Toolkit
- ☐ SQL Vulnerability Assessment (VA)
- ☐ Azure SQL Data Sync
- ☐ Data Migration Assistant (DMA)

Question 22

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.

- ☐ Table API
- ☐ MongoDB API
- ☐ Gremlin API
- ☐ SQL API
- ☐ Cassandra API

Question 23

DRAG DROP

You are developing the data platform for a global retail company. The company operates during normal working hours in each region. The analytical database is used once a week for building sales projections.

Each region maintains its own private virtual network.

Building the sales projections is very resource intensive and generates upwards of 20 terabytes (TB) of data.

Microsoft Azure SQL Databases must be provisioned.

Database provisioning must maximize performance and minimize cost

The daily sales for each region must be stored in an Azure SQL Database instance

Once a day, the data for all regions must be loaded in an analytical Azure SQL Database instance

You need to provision Azure SQL database instances.

How should you provision the database instances? To answer, drag the appropriate Azure SQL products to the correct databases. Each Azure SQL product may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Answer area

Azure SQL products	Database	Azure SQL product
Azure SQL Database elastic pools	Daily Sales	Azure SQL product
Azure SQL Database Premium	Weekly Analysis	Azure SQL product
Azure SQL Database Managed Instance		
Azure SQL Database Hyperscale		

Question 24

A company plans to use Azure SQL Database to support a mission-critical application.

The application must be highly available without performance degradation during maintenance windows.

You need to implement the solution.

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

NOTE: Each correct selection is worth one point.

- ☐ Premium service tier
- ☐ Virtual machine Scale Sets
- ☐ Basic service tier
- ☐ SQL Data Sync
- ☐ Always On availability groups
- ☐ Zone-redundant configuration

Question 25

HOTSPOT

You are developing a solution using a Lambda architecture on Microsoft Azure.

The data at rest layer must meet the following requirements:

Data storage:

Serve as a repository for high volumes of large files in various formats.

Implement optimized storage for big data analytics workloads.

Ensure that data can be organized using a hierarchical structure.

Batch processing:

Use a managed solution for in-memory computation processing.

Natively support Scala, Python, and R programming languages.

Provide the ability to resize and terminate the cluster automatically.

Analytical data store:

Support parallel processing.

Use columnar storage.

Support SQL-based languages.

You need to identify the correct technologies to build the Lambda architecture.

Which technologies should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Architecture requirement

Technology

Data storage

▼

Azure SQL Database
Azure Blob Storage
Azure Cosmos DB
Azure Data Lake Store

Batch processing

▼

HDInsight Spark
HDInsight Hadoop
Azure Databricks
HDInsight Interactive Query

Analytical data store

▼

HDInsight HBase
Azure SQL Data Warehouse
Azure Analysis Services
Azure Cosmos DB

Question 26

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

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

You have an Azure SQL database named DB1 that contains a table named Table1. Table1 has a field named Customer_ID that is varchar(22).

You need to implement masking for the Customer_ID field to meet the following requirements:

The first two prefix characters must be exposed.
The last four prefix characters must be exposed.
All other characters must be masked.

Solution: You implement data masking and use an email function mask.

Does this meet the goal?

- ☐ Yes
- ☐ No

Question 27

You plan to implement an Azure Cosmos DB database that will write 100,000 JSON every 24 hours. The database will be replicated to three regions. Only one region will be writable.

You need to select a consistency level for the database to meet the following requirements:

Guarantee monotonic reads and writes within a session.
Provide the fastest throughput.
Provide the lowest latency.

Which consistency level should you select?

- ☐ Strong
- ☐ Bounded Staleness
- ☐ Eventual
- ☐ Session
- ☐ Consistent Prefix

Question 28

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?

- ☐ transactional replication
- ☐ Data Migration Assistant (DMA)
- ☐ backup and restore
- ☐ SQL Server Agent job
- ☐ Azure SQL Data Sync

Question 29

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?

- ☐ Number of transactions only
- ☐ eDTUs per database only
- ☐ Number of databases only
- ☐ CPU usage only
- ☐ eDTUs and max data size

Question 30

DRAG DROP

Your company has on-premises Microsoft SQL Server instance.

The data engineering team plans to implement a process that copies data from the SQL Server instance to Azure Blob storage. The process must orchestrate and manage the data lifecycle.

You need to configure Azure Data Factory to connect to the SQL Server instance.

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

Actions

Configure a linked service to connect to the SQL Server instance.

From the on-premises network, install and configure a self-hosted runtime.

From the SQL Server, backup the database and then copy the database to Azure Blob storage.

Deploy an Azure Data Factory.

From the SQL Server, create a database master key.

Answer Area



Case Study (2 questions)

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 storage account. The storage must be available in the event of a regional disaster. The solution must minimize costs.

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 31

DRAG DROP

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

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

Actions

Parameterize deployment by using Azure Integration Runtime

Configure an Azure Logic App to deploy the deployment artifact

Configure Azure DevOps to deploy the deployment artifact

Create a deployment artifact containing an extracted Azure Resource Manager template

Parameterize deployment by using the Azure Resource Manager template parameter file

Create a deployment artifact containing a SQL Server Integration Services (SSIS) package

Answer Area

Question 32

DRAG DROP

You need to provision the polling data storage account.

How should you configure the storage account? To answer, drag the appropriate Configuration Value to the correct Setting. Each Configuration Value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Configuration values		Answer Area
	Setting	Configuration value
LRS		
GRS	Account type	
RA-GRS	Replication type	
Storage		
StorageV2		

Case Study (4 questions)

Case Study

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other question on this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study

To display the first question on this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview

General Overview

Litware, Inc, is an international car racing and manufacturing company that has 1,000 employees. Most employees are located in Europe. The company supports racing teams that compete in a worldwide racing series.

Physical Locations

Litware has two main locations: a main office in London, England, and a manufacturing plant in Berlin, Germany.

During each race weekend, 100 engineers set up a remote portable office by using a VPN to connect the datacentre in the London office. The portable office is set up and torn down in approximately 20 different countries each year.

Existing environment

Race Central

During race weekends, Litware uses a primary application named Race Central. Each car has several sensors that send real-time telemetry data to the London datacentre. The data is used for real-time tracking of the cars.

Race Central also sends batch updates to an application named Mechanical Workflow by using Microsoft SQL Server Integration Services (SSIS).

The telemetry data is sent to a MongoDB database. A custom application then moves the data to databases in SQL Server 2017. The telemetry data in MongoDB has more than 500 attributes. The application changes the attribute names when the data is moved to SQL Server 2017.

The database structure contains both OLAP and OLTP databases.

Mechanical Workflow

Mechanical Workflow is used to track changes and improvements made to the cars during their lifetime.

Currently, Mechanical Workflow runs on SQL Server 2017 as an OLAP system.

Mechanical Workflow has a named Table1 that is 1 TB. Large aggregations are performed on a single column of Table 1.

Requirements

Planned Changes

Litware is the process of rearchitecting its data estate to be hosted in Azure. The company plans to decommission the London datacentre and move all its applications to an Azure datacentre.

Technical Requirements

Litware identifies the following technical requirements:

Data collection for Race Central must be moved to Azure Cosmos DB and Azure SQL Database. The data must be written to the Azure datacentre closest to each race and must converge in the least amount of time.

The query performance of Race Central must be stable, and the administrative time it takes to perform optimizations must be minimized.

The datacentre for Mechanical Workflow must be moved to Azure SQL data Warehouse.

Transparent data encryption (TDE) must be enabled on all data stores, whenever possible.

An Azure Data Factory pipeline must be used to move data from Cosmos DB to SQL Database for Race Central. If the data load takes longer than 20 minutes, configuration changes must be made to Data Factory.

The telemetry data must migrate toward a solution that is native to Azure.

The telemetry data must be monitored for performance issues. You must adjust the Cosmos DB Request Units per second (RU/s) to maintain a performance SLA while minimizing the cost of the RU/s.

Data Masking Requirements

During rare weekends, visitors will be able to enter the remote portable offices. Litware is concerned that some proprietary information might be exposed. The company identifies the following data masking requirements for the Race Central data that will be stored in SQL Database:

Only show the last four digits of the values in a column named SuspensionSprings.

Only Show a zero value for the values in a column named ShockOilWeight.

Question 33

HOTSPOT

Which masking functions should you implement for each column to meet the data masking requirements? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

ShockOilWieht:	Credit card Default Email Random number
SuspensionSprings:	Credit card Default Email Random number

Question 34

HOTSPOT

You are building the data store solution for Mechanical Workflow.

How should you configure Table1? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Table Type:	▼ Hash distributed Replicated Round-robin
Index type:	▼ Clustered Clustered columnstore Heap Nonclustered

Question 35

You need to build a solution to collect the telemetry data for Race Control.

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

NOTE: Each correct selection is worth one point.

Answer Area

	API:	Cassandra Gremlin MongoDB SQL Table
	Consistency level:	Eventual Session Strong

Question 36

On which data store you configure TDE to meet the technical requirements?

- ☐ Cosmos DB
- ☐ SQL Data Warehouse
- ☐ SQL Database

Manage and develop data processing

(33 questions)

Question 37

Note: This question is part of series of questions that present the same scenario. Each question in the series contains 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 Data Lake Gen 2 storage account
2. Create an external file format and external table using the external data source
3. Load the data using the CREATE TABLE AS SELECT statement

Does the solution meet the goal?

- ☐ Yes
☐ No

Question 38

HOTSPOT

You are implementing Azure Stream Analytics functions.

Which windowing function should you use for each requirement? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Segment the data stream into distinct time segments that repeat but do not overlap:

	▼
Hopping	
Session	
Sliding	
Tumbling	

Segment the data stream into distinct time segments that repeat and can overlap:

	▼
Hopping	
Session	
Sliding	
Tumbling	

Segment the data stream to produce an output only when an event occurs:

	▼
Hopping	
Session	
Sliding	
Tumbling	

Question 39

Note: This question is part of series of questions that present the same scenario. Each question in the series contains 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 CREATE TABLE AS SELECT statement

Does the solution meet the goal?

- ☐ Yes
- ☐ No

Question 40

A company uses Azure SQL Database to store sales transaction data. Field sales employees need an offline copy of the database that includes last year's sales on their laptops when there is no internet connection available.

You need to create the offline export copy.

Which three options can you use? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

- ☐ Export to a BACPAC file by using Azure Cloud Shell, and save the file to an Azure storage account
- ☐ Export to a BACPAC file by using SQL Server Management Studio. Save the file to an Azure storage account
- ☐ Export to a BACPAC file by using the Azure portal
- ☐ Export to a BACPAC file by using Azure PowerShell and save the file locally
- ☐ Export to a BACPAC file by using the SqlPackage utility

Question 41

HOTSPOT

You need to implement an Azure Databricks cluster that automatically connects to Azure Data Lake Storage Gen2 by using Azure Active Directory (Azure AD) integration.

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

NOTE: Each correct selection is worth one point.

Answer Area

Cluster Mode:

	▼
High Concurrency	
Premium	
Standard	

Advanced option to enable:

	▼
Azure Data Lake Storage Gen1 Credential Passthrough	
Table Access Control	

Question 42

DRAG DROP

You have data stored in thousands of CSV files in Azure Data Lake Storage Gen2. Each file has a header row followed by a property formatted carriage return (/r) and line feed (/n).

You are implementing a pattern that batch loads the files daily into an Azure SQL data warehouse by using PolyBase.

You need to skip the header row when you import the files into the data warehouse.

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

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

Actions

Create an external data source that uses the abfs location.

Create an external file format and set the First_Row option.

Create an external data source that uses the Hadoop location.

Create a database scoped credential that uses an OAuth2 token and a key.

Use CREATE EXTERNAL TABLE AS SELECT (CETAS) and create a view that removes the empty row.

Answer Area

Question 43

Note: This question is part of series of questions that present the same scenario. Each question in the series contains 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. Use Azure Data Factory to convert the parquet files to CSV files
2. Create an external data source pointing to the Azure storage account
3. Create an external file format and external table using the external data source
4. Load the data using the INSERT...SELECT statement

Does the solution meet the goal?

- ☐ Yes
- ☐ No

Question 44

HOTSPOT

You are designing a new Lambda architecture on Microsoft Azure.

The real-time processing layer must meet the following requirements:

Ingestion:

- Receive millions of events per second
- Act as a fully managed Platform-as-a-Service (PaaS) solution
- Integrate with Azure Functions

Stream processing:

- Process on a per-job basis
- Provide seamless connectivity with Azure services
- Use a SQL-based query language

Analytical data store:

- Act as a managed service
- Use a document store
- Provide data encryption at rest

You need to identify the correct technologies to build the Lambda architecture using minimal effort. Which technologies should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Architecture requirement	Technology
Ingestion	▼ HDInsight Kafka Azure Event Hubs HDInsight Storm HDInsight Spark
Stream processing	▼ Azure Stream Analytics HDInsight with Spark Streaming Azure Cosmos DB Change Feed Azure Analysis Services
Analytical Data Store	▼ Hive LLAP on HDInsight Azure Analysis Services Azure Cosmos DB SQL Data Warehouse

Question 45

You are creating a new notebook in Azure Databricks that will support R as the primary language but will also support Scala and SQL.

Which switch should you use to switch between languages?

- ☐ %<language>
- ☐ \[<language>]
- ☐ \(<language>)
- ☐ @<Language>

Question 46

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.

Action	Technology					
Extract SQL data on-premises	<table border="1"> <tr> <td>Self-hosted integration runtime</td><td rowspan="4">V</td></tr> <tr> <td>Azure-SSIS integration runtime</td></tr> <tr> <td>Azure integration runtime</td></tr> <tr> <td>Source</td></tr> </table>	Self-hosted integration runtime	V	Azure-SSIS integration runtime	Azure integration runtime	Source
Self-hosted integration runtime	V					
Azure-SSIS integration runtime						
Azure integration runtime						
Source						
Load SQL data warehouse	<table border="1"> <tr> <td>Self-hosted integration runtime</td><td rowspan="4">V</td></tr> <tr> <td>Azure-SSIS integration runtime</td></tr> <tr> <td>Azure integration runtime</td></tr> <tr> <td>Sink</td></tr> </table>	Self-hosted integration runtime	V	Azure-SSIS integration runtime	Azure integration runtime	Sink
Self-hosted integration runtime	V					
Azure-SSIS integration runtime						
Azure integration runtime						
Sink						

Question 47

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?

- ☐ Azure SQL Data Warehouse
- ☐ HDInsight Apache Storm cluster
- ☐ Azure Stream Analytics
- ☐ HDInsight Apache Hadoop cluster using MapReduce
- ☐ HDInsight Spark cluster

Question 48

HOTSPOT

A company plans to develop solutions to perform batch processing of multiple sets of geospatial data.

You need to implement the solutions.

Which Azure services should you use? To answer, select the appropriate configuration in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Scenario	Tool
Use a native client application to run interactive queries and batch processes.	HDInsight Tools for Visual Studio Hive View HDInsight REST API Azure Data Factory
Use a web browser to run interactive queries and batch processes.	HDInsight Tools for Visual Studio Hive View HDInsight REST API Azure PowerShell
Develop batch processing applications that use Azure HDInsight	HDInsight Tools for Visual Studio Hive View HDInsight REST API NoSQL database

Question 49

DRAG DROP

You develop data engineering solutions for a company.

You need to deploy a Microsoft Azure Stream Analytics job for an IoT solution. The solution must:

Minimize latency.

Minimize bandwidth usage between the job and IoT device.

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

Actions

Configure routes.

Create an Azure Data Lake Storage container.

Create an IoT Hub and add the Azure Stream Analytics module to the IoT Hub namespace.

Create an Azure Stream Analytics edge job and configure job definition save location.

Create an Azure Stream Analytics cloud job and configure job definition save location.

Create an Azure Blob storage container.

Configure Streaming Units.

Answer Area

Question 50

DRAG DROP

You have an Azure Data Lake Storage Gen2 account that contains JSON files for customers. The files contain two attributes named FirstName and LastName.

You need to copy the data from the JSON files to an Azure SQL data Warehouse table by using Azure Databricks. A new column must be created that concatenates the FirstName and LastName values.

You create the following components:

A destination table in SQL Data Warehouse

An Azure Blob storage container

A service principal

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

Actions

Write the results to Data Lake Storage.

Drop the data frame.

Perform transformations on the data frame.

Mount the Data Lake Storage onto DBFS.

Perform transformations on the file.

Read the file into a data frame.

Specify a temporary folder to stage the data.

Write the results to a table in SQL Data Warehouse.

Answer Area

Question 51**HOTSPOT**

A company is deploying a service-based data environment. You are developing a solution to process this data.

The solution must meet the following requirements:

Use an Azure HDInsight cluster for data ingestion from a relational database in a different cloud service

Use an Azure Data Lake Storage account to store processed data

Allow users to download processed data

You need to recommend technologies for the solution.

Which technologies should you use? To answer, select the appropriate options in the answer area.

Answer Area

Data process	Technology								
Ingest	<table border="1"> <tr> <td>RevoScaleR</td><td>☐</td></tr> <tr> <td>Apache Sqoop</td><td>☐</td></tr> <tr> <td>Apache DistCp</td><td>☐</td></tr> <tr> <td>Azure CLI</td><td>☐</td></tr> </table>	RevoScaleR	☐	Apache Sqoop	☐	Apache DistCp	☐	Azure CLI	☐
RevoScaleR	☐								
Apache Sqoop	☐								
Apache DistCp	☐								
Azure CLI	☐								
Process	<table border="1"> <tr> <td>Apache DistCp</td><td>☐</td></tr> <tr> <td>Apache Kafka</td><td>☐</td></tr> <tr> <td>C#</td><td>☐</td></tr> <tr> <td>Apache Hive</td><td>☐</td></tr> </table>	Apache DistCp	☐	Apache Kafka	☐	C#	☐	Apache Hive	☐
Apache DistCp	☐								
Apache Kafka	☐								
C#	☐								
Apache Hive	☐								
Download	<table border="1"> <tr> <td>Apache Sqoop</td><td>☐</td></tr> <tr> <td>MapReduce</td><td>☐</td></tr> <tr> <td>RevoScaleR</td><td>☐</td></tr> <tr> <td>Ambari Hive View</td><td>☐</td></tr> </table>	Apache Sqoop	☐	MapReduce	☐	RevoScaleR	☐	Ambari Hive View	☐
Apache Sqoop	☐								
MapReduce	☐								
RevoScaleR	☐								
Ambari Hive View	☐								

Question 52

DRAG DROP

You implement an event processing solution using Microsoft Azure Stream Analytics.

The solution must meet the following requirements:

Ingest data from Blob storage

Analyze data in real time

Store processed data in Azure Cosmos DB

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

Actions

Create a query statement with the ORDER BY clause.

Create a query statement with the SELECT INTO statement.

Configure Blob storage for a reference data JOIN clause.

Configure Azure Event Hub as input; select items with the TIMESTAMP BY clause

Set up Cosmos DB as the output.

Configure Blob storage as input; select items with the TIMESTAMP BY clause.

Answer Area



Question 53

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.

- ☐ Event Grid topic
- ☐ Azure Stream Analytics Job that queries Twitter data from an Event Hub
- ☐ Azure Stream Analytics Job that queries Twitter data from an Event Grid
- ☐ Logic App that sends Twitter posts which have target keywords to Azure
- ☐ Event Grid subscription
- ☐ Event Hub instance

Question 54

HOTSPOT

A company plans to analyze a continuous flow of data from a social media platform by using Microsoft Azure Stream Analytics. The incoming data is formatted as one record per row.

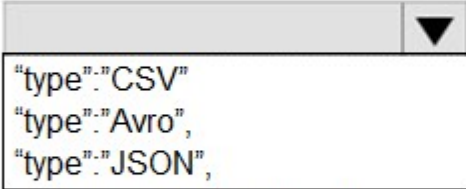
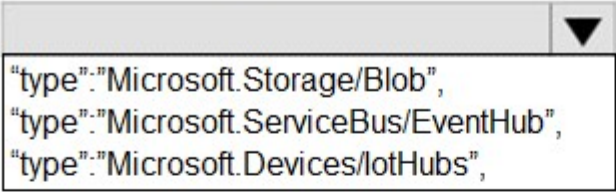
You need to create the input stream.

How should you complete the REST API segment? To answer, select the appropriate configuration in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

```

{
  "properties" : {
    "type" : "stream",
    "serialization" : {
      
    },
    "properties" : {
      "fieldDelimiter" : ". ",
      "encoding" : "UTF8"
    }
  },
  "datasource": {
    
  },
  "properties": {
    "serviceBusNamespace" : "sampleServiceBus",
    "sharedAccessPolicyName" : "SampleReceiver",
    "sharedAccessPolicyKey" : "<PolicyKey>"
    "eventHubName" : "sampleEventHub"
  },
  "compression": {
    "type" : "GZip"
  }
}

```

Question 55

You plan to perform batch processing in Azure Databricks once daily.

Which type of Databricks cluster should you use?

- ☐ job
- ☐ interactive
- ☐ High Concurrency

Question 56

DRAG DROP

Your company manages on-premises Microsoft SQL Server pipelines by using a custom solution.

The data engineering team must implement a process to pull data from SQL Server and migrate it to Azure Blob storage. The process must orchestrate and manage the data lifecycle.

You need to configure Azure Data Factory to connect to the on-premises SQL Server database.

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

Answer Area

Actions

Create an Azure Data Factory resource.

Configure a self-hosted integration runtime.

Create a virtual private network (VPN) connection from on-premises to Microsoft Azure.

Create a database master key on SQL Server.

Backup the database and send it Azure Blob storage.

Configure the on-premises SQL Server instance with an integration runtime.

Answer Area

Question 57

You have an Azure Storage account and an Azure SQL data warehouse by using Azure Data Factory. The solution must meet the following requirements:

Ensure that the data remains in the UK South region at all times.
Minimize administrative effort.

Which type of integration runtime should you use?

- ☐ Azure integration runtime
- ☐ Self-hosted integration runtime
- ☐ Azure-SSIS integration runtime

Question 58

Note: This question is part of series of questions that present the same scenario. Each question in the series contains 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?

- ☐ Yes
- ☐ No

Question 59

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?

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

Question 60

You need to develop a pipeline for processing data. The pipeline must meet the following requirements:

Scale up and down resources for cost reduction

Use an in-memory data processing engine to speed up ETL and machine learning operations.

Use streaming capabilities

Provide the ability to code in SQL, Python, Scala, and R

Integrate workspace collaboration with Git

What should you use?

- ☐ HDInsight Spark Cluster
- ☐ Azure Stream Analytics
- ☐ HDInsight Hadoop Cluster
- ☐ Azure SQL Data Warehouse
- ☐ HDInsight Kafka Cluster
- ☐ HDInsight Storm Cluster

Question 61

HOTSPOT

You develop data engineering solutions for a company.

A project requires an in-memory batch data processing solution.

You need to provision an HDInsight cluster for batch processing of data on Microsoft Azure.

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

NOTE: Each correct selection is worth one point.

Answer Area

New-AzureStorageContainer
New-AzureRmHDInsightClusterConfig
New-AzureRmHDInsightCluster

-Name \$clusterName -Context \$defaultStorageContext
\$objectConfig = New-Object "System.Collections.Generic.Dictionary" '2[System.String,System.String]"
&objectConfig.Add

"spark"
"hadoop"

, "2.3")

(

New-AzureRmHDInsightCluster
New-AzureRmHDInsightClusterConfig
New-AzureStorageContainer

-ResourceGroupName \$resourceGroupName'
-ClusterName \$clustername'
-Location \$location `
-ClusterSizeInNodes \$clusterSizeInNodes `
-ClusterType

"spark"
"hadoop"
"HBase"
"Storm"

,

-OSType \$clusterOS'
-Version \$clusterVersion'
-ComponentVersion \$objectConfig'
-HttpCredential \$httpCredential'
-DefaultStorageAccountName
"\$defaultStorageAccountName.blob.core.windows.net" `
-DefaultStorageAccountKey \$defaultStorageAccountKey'
-DefaultStorageContainer \$clusterName'
-SshCredential \$sshCredentials

Question 62

DRAG DROP

Your company plans to create an event processing engine to handle streaming data from Twitter.

The data engineering team uses Azure Event Hubs to ingest the streaming data.

You need to implement a solution that uses Azure Databricks to receive the streaming data from the Azure Event Hubs.

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

Actions**Answer Area**

Create and configure a Notebook that consumes the streaming data.

Import data from Blob storage.

Use Environment variables to define the Apache Spark connection.

Configure an ODBC or JDBC Connector.

Deploy the Azure Databricks service.

Deploy a Spark cluster and then attach the required libraries to the cluster.

**Question 63**

You are developing a solution that will stream to Azure Stream Analytics. The solution will have both streaming data and reference data.

Which input type should you use for the reference data?

- ☐ Azure Cosmos DB
- ☐ Azure Event Hubs
- ☐ Azure Blob storage
- ☐ Azure IoT Hub

Question 64

DRAG DROP

You develop data engineering solutions for a company.

A project requires analysis of real-time Twitter feeds. Posts that contain specific keywords must be stored and processed on Microsoft Azure and then displayed by using Microsoft Power BI. You need to implement the solution.

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

Actions	Answer Area
Create an HDInsight cluster with the Hadoop cluster type.	
Create a Jupyter Notebook.	
Run a job that uses the Spark Streaming API to ingest data from Twitter.	
Create a Runbook.	
Create an HDInsight cluster with the Spark cluster type.	
Create a table.	
Load the hvac table into Power BI Desktop.	

Question 65

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

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

Which windowing function should you use?

- ☐ a five-minute Session window
- ☐ a five-minute Sliding window
- ☐ a five-minute Tumbling window
- ☐ a five-minute Hopping window that has one-minute hop

Question 66

Note: This question is part of series of questions that present the same scenario. Each question in the series contains 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?

- ☐ Yes
☐ No

Question 67

HOTSPOT

A company plans to use Platform-as-a-Service (PaaS) to create the new data pipeline process. The process must meet the following requirements:

Ingest:

Access multiple data sources.

Provide the ability to orchestrate workflow.

Provide the capability to run SQL Server Integration Services packages.

Store:

Optimize storage for big data workloads

Provide encryption of data at rest.

Operate with no size limits.

Prepare and Train:

Provide a fully-managed and interactive workspace for exploration and visualization.

Provide the ability to program in R, SQL, Python, Scala, and Java.

Provide seamless user authentication with Azure Active Directory.

Model & Serve:

Implement native columnar storage.

Support for the SQL language.

Provide support for structured streaming.

You need to build the data integration pipeline.

Which technologies should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Architecture requirement

Technology

Ingest

▼

logic apps
Azure Data Factory
Azure Automation

Store

▼

Azure Data Lake Storage
Azure Blob storage
Azure Files

Prepare and Train

▼

HDInsight Apache Spark cluster
Azure Databricks
HDInsight Apache Storm cluster

Model and Serve

▼

HDInsight Apache Kafka cluster
Azure SQL Data Warehouse
Azure Data Lake Storage

Question 68

DRAG DROP

You are creating a managed data warehouse solution on Microsoft Azure.

You must use PolyBase to retrieve data from Azure Blob storage that resides in parquet format and load the data into a large table called FactSalesOrderDetails.

You need to configure Azure SQL Data Warehouse to receive the data.

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

Actions**Answer Area**

Create an external file format to map the parquet files.

Load the data to a staging table.

Create the external table FactSalesOrderDetails.

Enable Transparent Data Encryption.

Create an external data source for Azure Blob storage.

Create a master key on database.

Configure PolyBase to use Azure Blob storage.



Question 69

Note: This question is part of series of questions that present the same scenario. Each question in the series contains 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 a remote service binding pointing to the Azure Data Lake Gen 2 storage account
2. Create an external file format and external table using the external data source
3. Load the data using the CREATE TABLE AS SELECT statement

Does the solution meet the goal?

- ☐ Yes
- ☐ No

Case Study (2 questions)

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 storage account. The storage must be available in the event of a regional disaster. The solution must minimize costs.

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 70

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

How should you configure Azure Data Factory?

- ☐ Use a tumbling schedule trigger
- ☐ Use an event-based trigger
- ☐ Use a schedule trigger
- ☐ Use manual execution

Question 71

HOTSPOT

You need to ensure that Azure Data Factory pipelines can be deployed. How should you configure authentication and authorization for deployments? To answer, select the appropriate options in the answer choices.

NOTE: Each correct selection is worth one point.

Answer Area**Security requirement**

Authorization

Authentication

Technology

RBAC	✓
DAC	
MAC	
Claims	

Service Principal	^
Kerberos	
Certificate-based	
Bearer Token	

Case Study (3 questions)

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 databases
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 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 72

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

Which integration runtime should you use?

- ☐ self-hosted integration runtime
- ☐ Azure-SSIS Integration Runtime
- ☐ .NET Common Language Runtime (CLR)
- ☐ Azure integration runtime

Question 73

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.

- ☐ Azure Notification Hub
- ☐ Transact-SQL statements
- ☐ Azure Cache for Redis
- ☐ Apache Kafka statements
- ☐ Azure Event Grid
- ☐ Azure Stream Analytics

Question 74

HOTSPOT

You need set up the Azure Data Factory JSON definition for Tier 10 data.

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

NOTE: Each correct selection is worth one point.

Answer Area

Data factory component

Value

Connector

connection string	V
linked service name string	
gateway connection string	
data store name string	

Data movement activity

Azure SQL Data Warehouse	V
Azure Files	
Azure Blob	
Azure SQL Database	

Case Study (1 questions)

Case Study

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other question on this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study

To display the first question on this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview

General Overview

Litware, Inc, is an international car racing and manufacturing company that has 1,000 employees. Most employees are located in Europe. The company supports racing teams that compete in a worldwide racing series.

Physical Locations

Litware has two main locations: a main office in London, England, and a manufacturing plant in Berlin, Germany.

During each race weekend, 100 engineers set up a remote portable office by using a VPN to connect the datacentre in the London office. The portable office is set up and torn down in approximately 20 different countries each year

Existing environment

Race Central

During race weekends, Litware uses a primary application named Race Central. Each car has several sensors that send real-time telemetry data to the London datacentre. The data is used for real-time tracking of the cars.

Race Central also sends batch updates to an application named Mechanical Workflow by using Microsoft SQL Server Integration Services (SSIS).

The telemetry data is sent to a MongoDB database. A custom application then moves the data to databases in SQL Server 2017. The telemetry data in MongoDB has more than 500 attributes. The application changes the attribute names when the data is moved to SQL Server 2017.

The database structure contains both OLAP and OLTP databases.

Mechanical Workflow

Mechanical Workflow is used to track changes and improvements made to the cars during their lifetime.

Currently, Mechanical Workflow runs on SQL Server 2017 as an OLAP system.

Mechanical Workflow has a named Table1 that is 1 TB. Large aggregations are performed on a single column of Table 1.

Requirements

Planned Changes

Litware is the process of rearchitecting its data estate to be hosted in Azure. The company plans to decommission the London datacentre and move all its applications to an Azure datacentre.

Technical Requirements

Litware identifies the following technical requirements:

Data collection for Race Central must be moved to Azure Cosmos DB and Azure SQL Database. The data must be written to the Azure datacentre closest to each race and must converge in the least amount of time.

The query performance of Race Central must be stable, and the administrative time it takes to perform optimizations must be minimized.

The datacentre for Mechanical Workflow must be moved to Azure SQL data Warehouse.

Transparent data encryption (TDE) must be enabled on all data stores, whenever possible.

An Azure Data Factory pipeline must be used to move data from Cosmos DB to SQL Database for Race Central. If the data load takes longer than 20 minutes, configuration changes must be made to Data Factory.

The telemetry data must migrate toward a solution that is native to Azure.

The telemetry data must be monitored for performance issues. You must adjust the Cosmos DB Request Units per second (RU/s) to maintain a performance SLA while minimizing the cost of the RU/s.

Data Masking Requirements

During rare weekends, visitors will be able to enter the remote portable offices. Litware is concerned that some proprietary information might be exposed. The company identifies the following data masking requirements for the Race Central data that will be stored in SQL Database:

Only show the last four digits of the values in a column named SuspensionSprings.

Only Show a zero value for the values in a column named ShockOilWeight.

Question 75

What should you include in the Data Factory pipeline for Race Central?

- ☐ a copy activity that uses a stored procedure as a source
- ☐ a copy activity that contains schema mappings
- ☐ a delete activity that has logging enabled
- ☐ a filter activity that has a condition

Manage data security

(8 questions)

Question 76

You plan to use Microsoft Azure SQL Database instances with strict user access control. A user object must:

Move with the database if it is run elsewhere
Be able to create additional users

You need to create the user object with correct permissions.

Which two Transact-SQL commands should you run? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- ☐ ALTER LOGIN Mary WITH PASSWORD = 'strong_password';
- ☐ CREATE LOGIN Mary WITH PASSWORD = 'strong_password';
- ☐ ALTER ROLE db_owner ADD MEMBER Mary;
- ☐ CREATE USER Mary WITH PASSWORD = 'strong_password';
- ☐ GRANT ALTER ANY USER TO Mary;

Question 77

DRAG DROP

A company uses Microsoft Azure SQL Database to store sensitive company data. You encrypt the data and only allow access to specified users from specified locations.

You must monitor data usage, and data copied from the system to prevent data leakage.

You need to configure Azure SQL Database to email a specific user when data leakage occurs.

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

Actions

In Auditing, enable **Auditing**.

Configure the service to create alerts for threat detections of type **Data Exfiltration**.

In Firewalls and virtual networks, enable **Allow access to Azure services**.

Enable advanced threat protection.

Configure the service to send email alerts to security@contoso.com

Answer Area

Question 78

DRAG DROP

You manage security for a database that supports a line of business application.

Private and personal data stored in the database must be protected and encrypted.

You need to configure the database to use Transparent Data Encryption (TDE).

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

Actions**Answer Area**

Create a database encryption key using a certificate generated with the master key.

Create a certificate and then create the master key using a password.

Set the context to the master database.

Create a master key using a password.

Set the context to the company database.

Enable encryption.

Question 79

DRAG DROP

You plan to create a new single database instance of Microsoft Azure SQL Database.

The database must only allow communication from the data engineer's workstation. You must connect directly to the instance by using Microsoft SQL Server Management Studio.

You need to create and configure the Database. Which three Azure PowerShell cmdlets should you use to develop the solution? To answer, move the appropriate cmdlets from the list of cmdlets to the answer area and arrange them in the correct order.

Azure PowerShell cmdlets

New-AzureRmSqlElasticPool

New-AzureRmSqlServerFirewallRule

New-AzureRmSqlServer

New-AzureRmSqlServerVirtualNetworkRule

New-AzureRmSqlDatabase

Answer Area

Question 80

HOTSPOT

Your company uses Azure SQL Database and Azure Blob storage.

All data at rest must be encrypted by using the company's own key. The solution must minimize administrative effort and the impact to applications which use the database.

You need to configure security.

What should you implement? To answer, select the appropriate option in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Service	Encryption at rest
Azure SQL Database	always encrypted cell-level encryption row-level security transparent data encryption
Azure Storage	Azure disk encryption secure transport layer security (TLS) storage account keys default storage service encryption

Question 81

You develop data engineering solutions for a company.

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

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

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

NOTE: Each correct selection is worth one point.

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

Question 82

DRAG DROP

You manage the Microsoft Azure Databricks environment for a company. You must be able to access a private Azure Blob Storage account. Data must be available to all Azure Databricks workspaces. You need to provide the data access.

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

Actions	Answer Area
Upload a certificate	
Add secrets to the scope	
Use Blob Storage access key	
Create a secret scope	
Configure a JDBC connector	
Mount the Azure Blob Storage container	

Question 83

HOTSPOT

You develop data engineering solutions for a company. An application creates a database on Microsoft Azure. You have the following code:

```
private static readonly string endpointUrl =
    ConfigurationManager.AppSettings["EndPointUrl"];

private static readonly SecureString authorizationKey =
    ToSecureString(ConfigurationManager.AppSettings
    ["AuthorizationKey"]);

var client = new DocumentClient(new Url(endpointUrl),
    authorizationKey);

Database database = await client.CreateDatabaseAsync(
    new Database
    {
        Id = <DatabaseName>
    });
```

Which database and authorization types are used? To answer, select the appropriate option in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Component	Technology		
Azure database type	<table><tr><td>Azure Cosmos DB Azure SQL Database files Blob</td><td>▼</td></tr></table>	Azure Cosmos DB Azure SQL Database files Blob	▼
Azure Cosmos DB Azure SQL Database files Blob	▼		
Key type	<table><tr><td>resource token Master key certificate</td><td>▼</td></tr></table>	resource token Master key certificate	▼
resource token Master key certificate	▼		

Case Study (1 questions)

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 storage account. The storage must be available in the event of a regional disaster. The solution must minimize costs.

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 84

HOTSPOT

You need to ensure polling data security requirements are met.

Which security technologies should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Context	Security technology		
SQL Server	<table><tr><td>Azure Active Directory user Domain Active Directory user Managed Identity</td><td>✓</td></tr></table>	Azure Active Directory user Domain Active Directory user Managed Identity	✓
Azure Active Directory user Domain Active Directory user Managed Identity	✓		
PolyBase	<table><tr><td>Database scoped credential Database encryption key Application role</td><td>✓</td></tr></table>	Database scoped credential Database encryption key Application role	✓
Database scoped credential Database encryption key Application role	✓		

Case Study (5 questions)

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 databases
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 backup every week.

Backup 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 must be implemented to streamline 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 at 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 85

HOTSPOT

You need to mask tier 1 data. Which functions should you use? To answer, select the appropriate option in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area**Data type****A****Masking function**

custom text	☐
default	☐
email	☐
random number	☐

B

custom text	☐
default	☐
email	☐
random number	☐

C

custom text	☐
default	☐
email	☐
random number	☐

Question 86

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?

- ☐ Yes
- ☐ No

Question 87

DRAG DROP

You need to set up access to Azure SQL Database for Tier 7 and Tier 8 partners.

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

Actions

Connect to the Database and use Azure PowerShell to create a database firewall rule

Set the Allow Azure Services to Access Server to Disabled

In the Azure portal, create a database firewall rule

In the Azure portal, create a server firewall rule

Connect to the database and use Transact-SQL to create a database firewall rule

Set the Allow Azure Services to Access Server setting to Enabled

Answer Area

Question 88

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?

- ☐ Yes
☐ No

Question 89

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?

- ☐ Yes
☐ No

Monitor and optimize data solutions

(20 questions)

Question 90

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.

A company uses Azure Data Lake Gen 1 Storage to store big data related to consumer behavior.

You need to implement logging.

Solution: Configure Azure Data Lake Storage diagnostics to store logs and metrics in a storage account.

Does the solution meet the goal?

- ☐ Yes
- ☐ No

Question 91

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

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

You have a container named Sales in an Azure Cosmos DB database. Sales has 120 GB of data. Each entry in Sales has the following structure.

```
{  
  orderId: number,  
  OrderDetailId: number,  
  ProductName: string,  
  other information that might vary...  
}
```

The partition key is set to the OrderId attribute.

Users report that when they perform queries that retrieve data by ProductName, the queries take longer than expected to complete.

You need to reduce the amount of time it takes to execute the problematic queries.

Solution: You create a lookup collection that uses ProductName as a partition key and OrderId as a value.

Does this meet the goal?

- ☐ Yes
- ☐ No

Question 92

Note: This question is part of series of questions that present the same scenario. Each question in the series contains 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?

- ☐ Yes
- ☐ No

Question 93

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?

- ☐ Azure HDInsight .NET SDK
- ☐ Azure HDInsight REST API
- ☐ Ambari REST API
- ☐ Azure Log Analytics
- ☐ Ambari Web UI

Question 94

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

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

You have a container named Sales in an Azure Cosmos DB database. Sales has 120 GB of data. Each entry in Sales has the following structure.

```
{  
  OrderId: number,  
  OrderDetailId: number,  
  ProductName: string,  
  other information that might vary...  
}
```

The partition key is set to the OrderId attribute.

Users report that when they perform queries that retrieve data by ProductName, the queries take longer than expected to complete.

You need to reduce the amount of time it takes to execute the problematic queries.

Solution: You create a lookup collection that uses ProductName as a partition key.

Does this meet the goal?

- ☐ Yes
- ☐ No

Question 95

HOTSPOT

You are implementing automatic tuning mode for Azure SQL databases.

Automatic tuning is configured as shown in the following table.

Option	Server level	Database level
Force Plan	Inherited	Inherited
Create Index	Inherited	Inherited
Drop Index	Inherited	Inherited

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

NOTE: Each correct selection is worth one point.

Answer area

Statements	Yes	No
Force Plan for the database is ON.	☐	☐
Create Index for the database is ON.	☐	☐
Drop Index for the database is ON.	☐	☐

Question 96

You have an Azure SQL database that has masked columns.

You need to identify when a user attempts to infer data from the masked columns.

What should you use?

- ☐ Azure Advanced Threat Protection (ATP)
- ☐ custom masking rules
- ☐ Transparent Data Encryption (TDE)
- ☐ auditing

Question 97

HOTSPOT

You need to receive an alert when Azure SQL Data Warehouse consumes the maximum allotted resources.

Which resource type and signal should you use to create the alert in Azure Monitor? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Resource type:	▼ Resource group SQL server SQL data warehouse Subscription
Signal:	▼ CPU used Data IO percentage DWU limit DWU used

Question 98

HOTSPOT

A company is planning to use Microsoft Azure Cosmos DB as the data store for an application. You have the following Azure CLI command:

```
az cosmosdb create --name "cosmosdbdev1" --resource-group "rgdev"
```

You need to minimize latency and expose the SQL API. How should you complete the command? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area**Parameter****Value**

--default-consistency-level

Strong	V
Session	
Eventual	
Bounded staleness	

--kind

Parse	V
MongoDB	
GlobalDocumentDB	

Question 99

Note: This question is part of series of questions that present the same scenario. Each question in the series contains 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?

- ☐ Yes
- ☐ No

Question 100

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.

A company uses Azure Data Lake Gen 1 Storage to store big data related to consumer behavior.

You need to implement logging.

Solution: Configure an Azure Automation runbook to copy events.

Does the solution meet the goal?

- ☐ Yes
- ☐ No

Question 101

You configure monitoring for a Microsoft Azure SQL Data Warehouse implementation. The implementation uses PolyBase to load data from comma-separated value (CSV) files stored in Azure Data Lake Gen 2 using an external table.

Files with an invalid schema cause errors to occur.

You need to monitor for an invalid schema error.

For which error should you monitor?

- ☐ EXTERNAL TABLE access failed due to internal error: 'Java exception raised on call to HdfsBridge_Connect: Error [com.microsoft.polybase.client.KerberosSecureLogin] occurred while accessing external file.'
- ☐ EXTERNAL TABLE access failed due to internal error: 'Java exception raised on call to HdfsBridge_Connect: Error [No FileSystem for scheme: wasbs] occurred while accessing external file.'
- ☐ Cannot execute the query "Remote Query" against OLE DB provider "SQLNCLI11": for linked server "(null)", Query aborted- the maximum reject threshold (0 rows) was reached while reading from an external source: 1 rows rejected out of total 1 rows processed.
- ☐ EXTERNAL TABLE access failed due to internal error: 'Java exception raised on call to HdfsBridge_Connect: Error [Unable to instantiate LoginClass] occurred while accessing external file.'

Question 102

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?

- ☐ Azure Log Analytics log search query
- ☐ Ambari REST API
- ☐ Azure Monitor Metrics
- ☐ HDInsight .NET SDK
- ☐ Azure Log Analytics alert rule query

Question 103

Note: This question is part of series of questions that present the same scenario. Each question in the series contains 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?

- ☐ Yes
- ☐ No

Question 104

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.

A company uses Azure Data Lake Gen 1 Storage to store big data related to consumer behavior.

You need to implement logging.

Solution: Use information stored in Azure Active Directory reports.

Does the solution meet the goal?

- ☐ Yes
- ☐ No

Question 105

Contoso, Ltd. plans to configure existing applications to use Azure SQL Database.

When security-related operations occur, the security team must be informed.

You need to configure Azure Monitor while minimizing administrative effort.

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

NOTE: Each correct selection is worth one point.

- ☐ Create a new action group to email alerts@contoso.com.
- ☐ Use alerts@contoso.com as an alert email address.
- ☐ Use all security operations as a condition.
- ☐ Use all Azure SQL Database servers as a resource.
- ☐ Query audit log entries as a condition.

Question 106

Your company uses several Azure HDInsight clusters.

The data engineering team reports several errors with some applications using these clusters.

You need to recommend a solution to review the health of the clusters.

What should you include in your recommendation?

- ☐ Azure Automation
- ☐ Log Analytics
- ☐ Application Insights

Question 107

HOTSPOT

You have a new Azure Data Factory environment.

You need to periodically analyze pipeline executions from the last 60 days to identify trends in execution durations. The solution must use Azure Log Analytics to query the data and create charts.

Which diagnostic settings should you configure in Data Factory? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Log type:

	▼
ActivityRuns	
AllMetrics	
PipelineRuns	
TriggerRuns	

Storage location:

	▼
An Azure event hub	
An Azure Storage account	
Azure Cosmos DB	
Azure Log Analytics	

Question 108

DRAG DROP

Your company uses Microsoft Azure SQL Database configured with Elastic pools. You use Elastic Database jobs to run queries across all databases in the pool.

You need to analyze, troubleshoot, and report on components responsible for running Elastic Database jobs.

You need to determine the component responsible for running job service tasks.

Which components should you use for each Elastic pool job services task? To answer, drag the appropriate component to the correct task. Each component may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Answer Area

Components	Task	Component
Control Database	Execution results and diagnostics	
Azure Service Bus	Job launcher and tracker	
Azure Storage	Job metadata and state	
Job Service		

Question 109

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

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

You have a container named Sales in an Azure Cosmos DB database. Sales has 120 GB of data. Each entry in Sales has the following structure.

```
{  
  orderId: number,  
  OrderDetailId: number,  
  ProductName: string,  
  other information that might vary...  
}
```

The partition key is set to the OrderId attribute.

Users report that when they perform queries that retrieve data by ProductName, the queries take longer than expected to complete.

You need to reduce the amount of time it takes to execute the problematic queries.

Solution: You change the partition key to include ProductName.

Does this meet the goal?

- ☐ Yes
- ☐ No

Case Study (1 questions)

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 storage account. The storage must be available in the event of a regional disaster. The solution must minimize costs.

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 110

HOTSPOT

You need to ensure phone-based polling data upload reliability requirements are met. How should you configure monitoring? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Setting	Value
---------	-------

Metric	<table><tr><td>FileCount</td><td></td></tr><tr><td>BlobCapacity</td><td></td></tr><tr><td>FileCapacity</td><td></td></tr></table>	FileCount		BlobCapacity		FileCapacity	
FileCount							
BlobCapacity							
FileCapacity							

Aggregation	<table><tr><td>Avg</td><td></td></tr><tr><td>Sum</td><td></td></tr></table>	Avg		Sum	
Avg					
Sum					

Case Study (2 questions)

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 databases
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 backup every week.

Backup 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 must be implemented to streamline 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 111

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?

- ☐ extended events for average storage percentage that emails data engineers
- ☐ an alert rule to monitor CPU percentage in databases that emails data engineers
- ☐ an alert rule to monitor CPU percentage in elastic pools that emails data engineers
- ☐ an alert rule to monitor storage percentage in databases that emails data engineers
- ☐ an alert rule to monitor storage percentage in elastic pools that emails data engineers

Question 112

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?

- ☐ RequestSteps
- ☐ DmsWorkers
- ☐ SqlRequests
- ☐ ExecRequests

Case Study (3 questions)

Case Study

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other question on this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study

To display the first question on this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview

General Overview

Litware, Inc, is an international car racing and manufacturing company that has 1,000 employees. Most employees are located in Europe. The company supports racing teams that compete in a worldwide racing series.

Physical Locations

Litware has two main locations: a main office in London, England, and a manufacturing plant in Berlin, Germany.

During each race weekend, 100 engineers set up a remote portable office by using a VPN to connect the datacentre in the London office. The portable office is set up and torn down in approximately 20 different countries each year.

Existing environment

Race Central

During race weekends, Litware uses a primary application named Race Central. Each car has several sensors that send real-time telemetry data to the London datacentre. The data is used for real-time tracking of the cars.

Race Central also sends batch updates to an application named Mechanical Workflow by using Microsoft SQL Server Integration Services (SSIS).

The telemetry data is sent to a MongoDB database. A custom application then moves the data to databases in SQL Server 2017. The telemetry data in MongoDB has more than 500 attributes. The application changes the attribute names when the data is moved to SQL Server 2017.

The database structure contains both OLAP and OLTP databases.

Mechanical Workflow

Mechanical Workflow is used to track changes and improvements made to the cars during their lifetime.

Currently, Mechanical Workflow runs on SQL Server 2017 as an OLAP system.

Mechanical Workflow has a named Table1 that is 1 TB. Large aggregations are performed on a single column of Table 1.

Requirements

Planned Changes

Litware is the process of rearchitecting its data estate to be hosted in Azure. The company plans to decommission the London datacentre and move all its applications to an Azure datacentre.

Technical Requirements

Litware identifies the following technical requirements:

Data collection for Race Central must be moved to Azure Cosmos DB and Azure SQL Database. The data must be written to the Azure datacentre closest to each race and must converge in the least amount of time.

The query performance of Race Central must be stable, and the administrative time it takes to perform optimizations must be minimized.

The datacentre for Mechanical Workflow must be moved to Azure SQL data Warehouse.

Transparent data encryption (TDE) must be enabled on all data stores, whenever possible.

An Azure Data Factory pipeline must be used to move data from Cosmos DB to SQL Database for Race Central. If the data load takes longer than 20 minutes, configuration changes must be made to Data Factory.

The telemetry data must migrate toward a solution that is native to Azure.

The telemetry data must be monitored for performance issues. You must adjust the Cosmos DB Request Units per second (RU/s) to maintain a performance SLA while minimizing the cost of the RU/s.

Data Masking Requirements

During rare weekends, visitors will be able to enter the remote portable offices. Litware is concerned that some proprietary information might be exposed. The company identifies the following data masking requirements for the Race Central data that will be stored in SQL Database:

Only show the last four digits of the values in a column named SuspensionSprings.
Only Show a zero value for the values in a column named ShockOilWeight.

Question 113

Which two metrics should you use to identify the appropriate RU/s for the telemetry data? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- ☐ Number of requests
- ☐ Number of requests exceeded capacity
- ☐ End to end observed read latency at the 99th percentile
- ☐ Session consistency
- ☐ Data + Index storage consumed
- ☐ Avg Troughput/s

Question 114

You are monitoring the Data Factory pipeline that runs from Cosmos DB to SQL Database for Race Central.

You discover that the job takes 45 minutes to run.

What should you do to improve the performance of the job?

- ☐ Decrease parallelism for the copy activities.
- ☐ Increase that data integration units.
- ☐ Configure the copy activities to use staged copy.
- ☐ Configure the copy activities to perform compression.

Question 115

What should you implement to optimize SQL Database for Race Central to meet the technical requirements?

- ☐ the sp_update stored procedure
- ☐ automatic tuning
- ☐ Query Store
- ☐ the dbcc checkdb command

Manage and troubleshoot Azure data solutions

(5 questions)

Question 116

DRAG DROP

A company builds an application to allow developers to share and compare code. The conversations, code snippets, and links shared by people in the application are stored in a Microsoft Azure SQL Database instance. The application allows for searches of historical conversations and code snippets.

When users share code snippets, the code snippet is compared against previously share code snippets by using a combination of Transact-SQL functions including SUBSTRING, FIRST_VALUE, and SQRT. If a match is found, a link to the match is added to the conversation.

Customers report the following issues:

Delays occur during live conversations

A delay occurs before matching links appear after code snippets are added to conversations

You need to resolve the performance issues.

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

NOTE: Each correct selection is worth one point.

Technologies	Answer Area	
	Issue	Technology
columnstore index	Delays in conversations	
non-durable table	Delays in match links	
meterialized view		
memory-optimized table		

Question 117

A company has a real-time data analysis solution that is hosted on Microsoft Azure. The solution uses Azure Event Hub to ingest data and an Azure Stream Analytics cloud job to analyze the data. The cloud job is configured to use 120 Streaming Units (SU).

You need to optimize performance for the Azure Stream Analytics job.

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

NOTE: Each correct selection is worth one point.

- ☐ Implement event ordering
- ☐ Scale the SU count for the job up
- ☐ Implement Azure Stream Analytics user-defined functions (UDF)
- ☐ Scale the SU count for the job down
- ☐ Implement query parallelization by partitioning the data output
- ☐ Implement query parallelization by partitioning the data input

Question 118

You manage a Microsoft Azure SQL Data Warehouse Gen 2.

Users report slow performance when they run commonly used queries. Users do not report performance changes for infrequently used queries.

You need to monitor resource utilization to determine the source of the performance issues.

Which metric should you monitor?

- ☐ Cache used percentage
- ☐ Local tempdb percentage
- ☐ DWU percentage
- ☐ CPU percentage
- ☐ Data IO percentage

Question 119

You manage a process that performs analysis of daily web traffic logs on an HDInsight cluster. Each of the 250 web servers generates approximately 10 megabytes (MB) of log data each day. All log data is stored in a single folder in Microsoft Azure Data Lake Storage Gen 2.

You need to improve the performance of the process.

Which two changes should you make? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

- ☐ Combine the daily log files for all servers into one file
- ☐ Increase the value of the mapreduce.map.memory parameter
- ☐ Move the log files into folders so that each day's logs are in their own folder
- ☐ Increase the number of worker nodes
- ☐ Increase the value of the hive.tez.container.size parameter

Question 120

You implement an Azure SQL Data Warehouse instance.

You plan to migrate the largest fact table to Azure SQL Data Warehouse. The table resides on Microsoft SQL Server on-premises and is 10 terabytes (TB) in size.

Incoming queries use the primary key Sale Key column to retrieve data as displayed in the following table:

SaleKey	CityKey	CustomerKey	StockItemKey	InvoiceDateKey	Quantity	UnitPrice	TotalExcludingTax
49309	90858	70	69	10/22/13	8	16	128
49313	55710	126	69	10/22/13	2	16	32
49343	44710	234	68	10/22/13	10	16	160
49352	66109	163	70	10/22/13	4	16	64
49448	65312	230	70	10/22/13	8	16	128
49646	85877	271	70	10/24/13	1	16	16
49798	41238	288	69	10/24/13	1	16	16

You need to distribute the large fact table across multiple nodes to optimize performance of the table.

Which technology should you use?

- ☐ hash distributed table with clustered ColumnStore index
- ☐ hash distributed table with clustered index
- ☐ heap table with distribution replicate
- ☐ round robin distributed table with clustered index
- ☐ round robin distributed table with clustered ColumnStore index