



v



Vizle

This PDF is generated automatically by **Vizle**.
Slides created *only for a few minutes* of your Video.



For the full PDF, please **Login to Vizle**.

<https://vizle.offnote.co> (Login via Google, top-right)

Stay connected with us:

Join us on **Facebook, Discord, Quora, Telegram**.

Plan and Implement Data Platform Resources

Testlet 1

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 questions in 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 in 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

Itware, Inc. is a renewable energy company that has a main office in Boston. The main office hosts a sales department and the primary datacenter for the company.

Physical Locations

Existing Environment

Itware has a manufacturing office and a research office in separate locations near Boston. Each office has its own datacenter and internet connection.

The manufacturing and research datacenters connect to the primary datacenter by using a VPN.

Network Environment

The primary datacenter has an ExpressRoute connection that uses both Microsoft peering and private peering. The private peering connects to an Azure virtual network named HubVNet.

Identity Environment

Itware has a hybrid Azure Active Directory (Azure AD) deployment that uses a domain named itwareinc.com. All Azure subscriptions are associated to the itwareinc.com Azure AD tenant.

Database Environment

The sales department has the following database workload:

- An on-premises named SERVER11 hosts an instance of Microsoft SQL Server 2012 and two 1-TB databases.
- A logical server named SalesSrv01A contains a geo-replicated Azure SQL database named SalesSQLDb1. SalesSQLDb1 is in an elastic pool named SalesSQLDb1Pool. SalesSQLDb1 uses database firewall rules and contained database users.
- An application named SalesSQLDb1App1 uses SalesSQLDb1.

The manufacturing office contains two on-premises SQL Server 2016 servers named SERVER2 and SERVER3. The servers are nodes in the same Always On availability group. The availability group contains a database named ManufacturingSQLDb1.

Database administrators have two Azure virtual machines in HubVnet named VM1 and VM2 that run Windows Server 2019 and are used to manage all the Azure databases.

Licensing Agreement

Litware is a Microsoft Volume Licensing customer that has License Mobility through Software Assurance.

Current Problems

SalesSQLDb1 experiences performance issues that are likely due to out-of-date statistics and frequent blocking queries.

Requirements

Planned Changes

Litware plans to implement the following changes:

- Implement 30 new databases in Azure, which will be used by time-sensitive manufacturing apps that have varying usage patterns. Each database will be approximately 20 GB.
- Create a new Azure SQL database named ResearchDB1 on a logical server named ResearchSrv01. ResearchDB1 will contain Personally Identifiable Information (PII) data.
- Develop an app named ResearchApp1 that will be used by the research department to populate and access ResearchDB1.
- Migrate ManufacturingSQLDb1 to the Azure virtual machine platform.
- Migrate the SERVER1 databases to the Azure SQL Database platform.

Technical Requirements

Litware identifies the following technical requirements:

- Maintenance tasks must be automated.
- The 30 new databases must scale automatically.
- The use of an on-premises infrastructure must be minimized.
- Azure Hybrid Use Benefits must be leveraged for Azure SQL Database deployments.
- All SQL Server and Azure SQL Database metrics related to CPU and storage usage and limits must be analyzed by using Azure built-in functionality.

Security and Compliance Requirements

Litware identifies the following security and compliance requirements:

- Store encryption keys in Azure Key Vault.
- Retain backups of the PII data for two months.
- Encrypt the PII data at rest, in transit, and in use.
- Use the principle of least privilege whenever possible.
- Authenticate database users by using Active Directory credentials.
- Protect Azure SQL Database instances by using database-level firewall rules.
- Ensure that all databases hosted in Azure are accessible from VM1 and VM2 without relying on public endpoints.

Business Requirements

Litware identifies the following business requirements:

- Meet an SLA of 99.99% availability for all Azure deployments.
- Minimize downtime during the migration of the SERVER1 databases.
- Use the Azure Hybrid Use Benefits when migrating workloads to Azure.
- Once all requirements are met, minimize costs whenever possible.

QUESTION 1
HOTSPOT

You are planning the migration of the SERVER1 databases. The solution must meet the business requirements.

What should you include in the migration plan? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Azure Database Migration Service pricing tier:

Standard 2-vCore
Standard 4-vCore
Premium 4-vCore

Required Azure resource:

A virtual network that has service endpoints
A VPN gateway
An Azure Logic app

Correct Answer:

Answer Area

Azure Database Migration Service pricing tier:

Standard 2-vCore
Standard 4-vCore
Premium 4-vCore

Required Azure resource:

A virtual network that has service endpoints
A VPN gateway
An Azure Logic app

Section: (none)
Explanation

Explanation/Reference:
Explanation:

Azure Database Migration service



Box 1: Premium 4-vCore

Scenario: Migrate the SERVER1 databases to the Azure SQL Database platform.
- Minimize downtime during the migration of the SERVER1 databases.

Premium 4-vCore is for large or business critical workloads. It supports online migrations, offline migrations, and faster migration speeds.

Incorrect Answers:

The Standard pricing tier suits most small- to medium- business workloads, but it supports offline migration only.

Box 2: A VPN gateway

You need to create a Microsoft Azure Virtual Network for the Azure Database Migration Service by using the Azure Resource Manager deployment model, which provides site-to-site connectivity to your on-premises source servers by using either ExpressRoute or VPN.

Reference:

<https://azure.microsoft.com/pricing/details/database-migration/>

<https://docs.microsoft.com/en-us/azure/dms/tutorial-sql-server-azure-sql-online>

QUESTION 2

HOTSPOT

You need to recommend the appropriate purchasing model and deployment option for the 30 new databases. The solution must meet the technical requirements and the business requirements.

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

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Purchasing model:

Deployment option:

Correct Answer:

Answer Area

Purchasing model:

Azure virtual machine reserved instances
 DTU
 vCore

Deployment option:

An Azure SQL Database managed instance
 A SQL Server Always On availability group

Section: (none)

Explanation

Explanation/Reference:

Explanation:

Box 1: DTU

Scenario:

- The 30 new databases must scale automatically.
- Once all requirements are met, minimize costs whenever possible.

You can configure resources for the pool based either on the DTU-based purchasing model or the vCore-based purchasing model.

In short, for simplicity, the DTU model has an advantage. Plus, if you're just getting started with Azure SQL Database, the DTU model offers more options at the lower end of performance, so you can get started at a lower price point than with vCore.

Box 2: An Azure SQL database elastic pool

Azure SQL Database elastic pools are a simple, cost-effective solution for managing and scaling multiple databases that have varying and unpredictable usage demands. The databases in an elastic pool are on a single server and share a set number of resources at a set price. Elastic pools in Azure SQL Database enable SaaS developers to optimize the price performance for a group of databases within a prescribed budget while delivering performance elasticity for each database.

Reference:

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

<https://docs.microsoft.com/en-us/azure/sql-database/reserved-capacity-overview>



Plan and Implement Data Platform Resources

Testlet 2

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 questions in 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 in 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

Contoso, Ltd. is a clothing retailer based in Seattle. The company has 2,000 retail stores across the United States and an emerging online presence.

The network contains an Active Directory forest named contoso.com. The forest is integrated with an Azure Active Directory (Azure AD) tenant named contoso.com. Contoso has an Azure subscription associated to the contoso.com Azure AD tenant.

Existing Environment

Transactional Data

Contoso has three years of customer, transaction, operational, sourcing, and supplier data comprised of 10 billion records stored across multiple on-premises Microsoft SQL Server servers. The SQL Server instances contain data from various operations systems. The data is loaded into the instances by using SQL Server Integration Services (SSIS) packages.

You estimate that combining all product sales transactions into a company-wide sales transactions dataset will result in a single table that contains 5 billion rows, with one row per transaction.

Most queries targeting the sales transactions data will be used to identify which products were sold in retail stores and which products were sold online during different time periods. Sales transaction data that is older than three years will be removed monthly.

You plan to create a retail store table that will contain the address of each retail store. The table will be approximately 2 MB. Queries for retail store sales will include the retail store addresses.

You plan to create a promotional table that will contain a promotion ID. The promotion ID will be associated to a specific product. The product will be identified by a product ID. The table will be approximately 5 GB.

Streaming Twitter Data

The ecommerce department at Contoso develops an Azure logic app that captures trending Twitter feeds

referencing the company's products and pushes the products to Azure Event Hubs.

Planned Changes and Requirements

Planned Changes

Contoso plans to implement the following changes:

- Load the sales transaction dataset to Azure Synapse Analytics.
- Integrate on-premises data stores with Azure Synapse Analytics by using SSIS packages.
- Use Azure Synapse Analytics to analyze Twitter feeds to assess customer sentiments about products.

Sales Transaction Dataset Requirements

Contoso identifies the following requirements for the sales transaction dataset:

- Partition data that contains sales transaction records. Partitions must be designed to provide efficient loads by month. Boundary values must belong to the partition on the right.
- Ensure that queries joining and filtering sales transaction records based on product ID complete as quickly as possible.
- Implement a surrogate key to account for changes to the retail store addresses.
- Ensure that data storage costs and performance are predictable.
- Minimize how long it takes to remove old records.

Customer Sentiment Analytics Requirements

Contoso identifies the following requirements for customer sentiment analytics:

- Allow Contoso users to use PolyBase in an Azure Synapse Analytics dedicated SQL pool to query the content of the data records that host the Twitter feeds. Data must be protected by using row-level security (RLS). The users must be authenticated by using their own Azure AD credentials.
- Maximize the throughput of ingesting Twitter feeds from Event Hubs to Azure Storage without purchasing additional throughput or capacity units.
- Store Twitter feeds in Azure Storage by using Event Hubs Capture. The feeds will be converted into Parquet files.
- Ensure that the data store supports Azure AD-based access control down to the object level.
- Minimize administrative effort to maintain the Twitter feed data records.
- Purge Twitter feed data records that are older than two years.

Data Integration Requirements

Contoso identifies the following requirements for data integration:

- Use an Azure service that leverages the existing SSIS packages to ingest on-premises data into datasets stored in a dedicated SQL pool of Azure Synapse Analytics and transform the data.
- Identify a process to ensure that changes to the ingestion and transformation activities can be version-controlled and developed independently by multiple data engineers.

QUESTION 1

You need to design a data retention solution for the Twitter feed data records. The solution must meet the customer sentiment analytics requirements.

Which Azure Storage functionality should you include in the solution?

- A. time-based retention
- B. change feed
- C. lifecycle management
- D. soft delete

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:

Explanation:
 The lifecycle management policy lets you delete blobs, blob versions, and blob snapshots at the end of their lifecycles.

Scenario:

- Purge Twitter feed data records that are older than two years.
- Store Twitter feeds in Azure Storage by using Event Hubs Capture. The feeds will be converted into Parquet files.
- Minimize administrative effort to maintain the Twitter feed data records.

Incorrect Answers:

A: Time-based retention policy support. Users can set policies to store data for a specified interval. When a time-based retention policy is set, blobs can be created and read, but not modified or deleted. After the retention period has expired, blobs can be deleted but not overwritten.

Reference:

<https://docs.microsoft.com/en-us/azure/storage/blobs/storage-lifecycle-management-concepts>

QUESTION 2

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

What should you create?

- A. a table that has a FOREIGN KEY constraint
- B. a table that has an IDENTITY property
- C. a user-defined SEQUENCE object
- D. a system-versioned temporal table

Correct Answer: B

Section: (none)

Explanation

Explanation/Reference:

Explanation:
 Scenario: Contoso requirements for the sales transaction dataset include:
 • Implement a surrogate key to account for changes to the retail store addresses.

A surrogate key on a table is a column with a unique identifier for each row. The key is not generated from the table data. Data modelers like to create surrogate keys on their tables when they design data warehouse models. You can use the IDENTITY property to achieve this goal simply and effectively without affecting load performance.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-identity>

QUESTION 3

HOTSPOT

You need to design an analytical storage solution for the transactional data. The solution must meet the sales transaction dataset requirements.

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

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Table type to store retail store data:

Hash
Replicated
Round-robin

Table type to store promotional data:

Hash
Replicated
Round-robin

Correct Answer:

Answer Area

Table type to store retail store data:

Hash
Replicated
Round-robin

Table type to store promotional data:

Hash
Replicated
Round-robin

Section: (none)
Explanation

Explanation/Reference:
Explanation:

Box 1: Hash

Scenario:

Ensure that queries joining and filtering sales transaction records based on product ID complete as quickly as possible.

A hash distributed table can deliver the highest query performance for joins and aggregations on large tables.

Box 2: Round-robin

Plan and Implement Data Platform Resources

Question Set 3

QUESTION 1

You have 20 Azure SQL databases provisioned by using the vCore purchasing model.

You plan to create an Azure SQL Database elastic pool and add the 20 databases.

Which three metrics should you use to size the elastic pool to meet the demands of your workload? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. total size of all the databases
- B. geo-replication support
- C. number of concurrently peaking databases * peak CPU utilization per database
- D. maximum number of concurrent sessions for all the databases
- E. total number of databases * average CPU utilization per database

Correct Answer: ACE

Section: (none)

Explanation

Explanation/Reference:

Explanation:

CE: Estimate the vCores needed for the pool as follows:

For vCore-based purchasing model: $MAX(\lceil \text{Total number of DBs} \times \text{average vCore utilization per DB} \rceil, \lceil \text{Number of concurrently peaking DBs} \times \text{Peak vCore utilization per DB} \rceil)$

A: Estimate the storage space needed for the pool by adding the number of bytes needed for all the databases in the pool.

Reference:

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

QUESTION 2

DRAG DROP

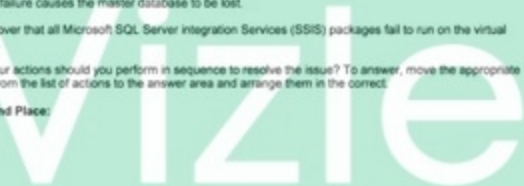
You have SQL Server 2019 on an Azure virtual machine that contains an SSISDB database.

A recent failure causes the master database to be lost.

You discover that all Microsoft SQL Server Integration Services (SSIS) packages fail to run on the virtual machine.

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

Select and Place:



Actions

Add a certificate to an Azure key vault

Enable Transparent Data Encryption (TDE)

Encrypt a copy of the master key by using the service master key

Turn on the TRUSTWORTHY property and the CLR property

Attach the SSISDB database

Open the master key for the SSISDB database

Answer Area

▶

▶

▶

▶

▶

Correct Answer:

Actions

Add a certificate to an Azure key vault

Enable Transparent Data Encryption (TDE)

Encrypt a copy of the master key by using the service master key

Turn on the TRUSTWORTHY property and the CLR property

Attach the SSISDB database

Open the master key for the SSISDB database

Answer Area

▶

▶

▶

▶

▶

Section: (none)

Explanation

Explanation/Reference:

Explanation:

Step 1. Attach the SSISDB database

Step 2. Turn on the TRUSTWORTHY property and the CLR property
If you are restoring the SSISDB database to an SQL Server instance where the SSISDB catalog was never created, enable common language runtime (CLR)

Step 3. Open the master key for the SSISDB database
Restore the master key by this method if you have the original password that was used to create SSISDB.
open master key decryption by password = 'LS1Setup' --Password used when creating SSISDB'

Alter Master Key Add encryption by Service Master Key

Step 4. Encrypt a copy of the mater key by using the service master key

Reference:

<https://docs.microsoft.com/en-us/sql/integration-services/backup-restore-and-move-the-osis-catalog>

QUESTION 3

You have an Azure SQL database that contains a table named factSales. FactSales contains the columns shown in the following table.

Name	Data type
SalesID	Int
Product	Int
Total Number	Numeric(8,4)
Tax Number	Numeric(8,4)
SalesRep	Varchar(30)

FactSales has 6 billion rows and is loaded nightly by using a batch process.

Which type of compression provides the greatest space reduction for the database?

- A. page compression
- B. row compression
- C. columnstore compression
- D. columnstore archival compression

Correct Answer: D

Section: (none)

Explanation

Explanation/Reference:

Explanation:

Columnstore tables and indexes are always stored with columnstore compression. You can further reduce the size of columnstore data by configuring an additional compression called archival compression.

Note: Columnstore — The columnstore index is also logically organized as a table with rows and columns, but the data is physically stored in a column-wise data format.

Incorrect Answers:

B: Rowstore — The rowstore index is the traditional style that has been around since the initial release of SQL Server.

For rowstore tables and indexes, use the data compression feature to help reduce the size of the database.

Reference:

<https://docs.microsoft.com/en-us/sql/relational-databases/data-compression/data-compression>

QUESTION 4

You have a Microsoft SQL Server 2019 database named DB1 that uses the following database-level and instance-level features.

- Clustered columnstore indexes



- Automatic tuning
- Change tracking
- PolyBase

You plan to migrate DB1 to an Azure SQL database.

What feature should be removed or replaced before DB1 can be migrated?

- A. Clustered columnstore indexes
- B. PolyBase
- C. Change tracking
- D. Automatic tuning

Correct Answer: B
Section: (none)

Explanation

Explanation/Reference:

Explanation:

This table lists the key features for PolyBase and the products in which they're available.

Feature	SQL Server (Beginning with 2016)	Azure SQL Database	Azure Synapse Analytics	Project Data Warehouse
Query Hadoop data with Transact-SQL	Yes	No	No	Yes
Import data from Hadoop	Yes	No	No	Yes
Export data to Hadoop	Yes	No	No	Yes
Query, import from, export to Azure HDInsight	No	No	No	No
Push-down query computations to Hadoop	Yes	No	No	Yes
Import data from Azure Blob storage	Yes	Yes	Yes	Yes
Export data to Azure Blob storage	Yes	No	Yes	Yes
Import data from Azure Data Lake Store	No	No	Yes	No
Export data to Azure Data Lake Store	No	No	Yes	No
Run PolyBase queries from Microsoft BI tools	Yes	No	Yes	Yes

Incorrect Answers:

C: Change tracking is a lightweight solution that provides an efficient change tracking mechanism for applications. It applies to both Azure SQL Database and SQL Server.

D: Azure SQL Database and Azure SQL Managed Instance automatic tuning provides peak performance and stable workloads through continuous performance tuning based on AI and machine learning.

Reference:

<https://docs.microsoft.com/en-us/sql/relational-databases/polybase/polybase-versioned-feature-summary>

Answer Area

Backup type:

Full and log backups only
 Full backup only
 Log backup only

Backup option:

WITH CHECKSUM
 WITH NOINIT
 WITH UNLOAD

Section: (none)

Explanation

Explanation/Reference:

Explanation:

Box 1: Full and log backups only

Make sure to take every backup on a separate backup media (backup files). Azure Database Migration Service doesn't support backups that are appended to a single backup file. Take full backup and log backups to separate backup files.

Box 2: WITH CHECKSUM

Azure Database Migration Service uses the backup and restore method to migrate your on-premises databases to SQL Managed Instance. Azure Database Migration Service only supports backups created using checksum.

Incorrect Answers:

NOINIT Indicates that the backup set is appended to the specified media set, preserving existing backup sets. If a media password is defined for the media set, the password must be supplied. NOINIT is the default.

UNLOAD

Specifies that the tape is automatically rewound and unloaded when the backup is finished. UNLOAD is the default when a session begins.

Reference:

<https://docs.microsoft.com/en-us/azure/dms/know-issues-azure-sql-db-managed-instance-online>

QUESTION 7

DRAG DROP

You have a resource group named App1Dev that contains an Azure SQL Database server named DevServer1. DevServer1 contains an Azure SQL database named DB1. The schema and permissions for DB1 are saved in a Microsoft SQL Server Data Tools (SSDT) database project.

You need to populate a new resource group named App1Test with the DB1 database and an Azure SQL Server named TestServer1. The resources in App1Test must have the same configurations as the resources in App1Dev.

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.

Select and Place:

Actions

Change the Active Directory Admin on TestServer1

Change the server name and related variables in the templates

From the database project, deploy the database schema and permissions

Add IP addresses to the firewall

From the Azure portal, export the Azure Resource Manager templates

From the Azure portal, deploy the templates.

Answer Area



Correct Answer:

Actions

Change the Active Directory Admin on TestServer1

Change the server name and related variables in the templates

From the database project, deploy the database schema and permissions

Add IP addresses to the firewall

From the Azure portal, export the Azure Resource Manager templates

From the Azure portal, deploy the templates.

Answer Area

From the Azure portal, export the Azure Resource Manager templates

Change the server name and related variables in the templates

From the Azure portal, deploy the templates

From the database project, deploy the database schema and permissions

Section: (none)

Explanation

Explanation/Reference:

**QUESTION 8
HOTSPOT**

You have an Azure Synapse Analytics dedicated SQL pool named Pool1 and an Azure Data Lake Storage Gen2 account named Account1.

You plan to access the files in Account1 by using an external table.

You need to create a data source in Pool1 that you can reference when you create the external table.

How should you complete the Transact-SQL statement? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

```
CREATE EXTERNAL DATA SOURCE sourcel
WITH
{ LOCATION = 'https://account1.
    .core.windows.net',
    PUSHDOWN = ON
    TYPE = BLOB_STORAGE
    TYPE = HADOOP
}
```

Correct Answer:

Answer Area

```
CREATE EXTERNAL DATA SOURCE sourcel
WITH
{ LOCATION = 'https://account1.
    .core.windows.net',
    PUSHDOWN = ON
    TYPE = BLOB_STORAGE
    TYPE = HADOOP
}
```

Section: (none)

Explanation

Answer Area

```

▼ applications
CREATE EXTERNAL TABLE
CREATE TABLE
CREATE VIEW
WITH (
    LOCATION = 'applications/',
    DATA_SOURCE = applications_ds,
    FILE_FORMAT = applications_file_format
)
AS
SELECT id, [address_housenumber] as addressnumber, [address_line1]
as addressline1
FROM
▼ (BULK 'https://contoso1.dfs.core.windows.net/
applications/year**/*.parquet',
CROSS APPLY
OPENROWSET
OPTIMIZE
FORMAT = 'PARQUET') AS [r]
GO
  
```

Section: (none)

Explanation:

Explanation/Reference:

Explanation:

Box 1: CREATE EXTERNAL TABLE

An external table points to data located in Hadoop, Azure Storage blob, or Azure Data Lake Storage. External tables are used to read data from files or write data to files in Azure Storage. With Synapse SQL, you can use external tables to read external data using dedicated SQL pool or serverless SQL pool.

Syntax:

```

CREATE EXTERNAL TABLE [ database_name schema_name table_name | schema_name table_name ]
table_name
[ ( <column_definition> [ ,...n ] ) ]
WITH (
    LOCATION = 'folder_or_filepath',
    DATA_SOURCE = external_data_source_name,
    FILE_FORMAT = external_file_format_name
  )
  
```

Box 2: OPENROWSET

When using serverless SQL pool, CETAS is used to create an external table and export query results to Azure Storage Blob or Azure Data Lake Storage Gen2.

Example:

```

AS
SELECT decennialTime, stateName, SUM(population) AS population
FROM
OPENROWSET(BULK 'https://azureopendatastorage.blob.core.windows.net/censusdatacontainer/release/
us_population_countyyear**'.parquet,
FORMAT=PARQUET) AS [r]
  
```

```
GROUP BY decennialTme, stateName
GO
```

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-tables-external-tables>

QUESTION 12

You have an Azure Synapse Analytics Apache Spark pool named Pool1.

You plan to load JSON files from an Azure Data Lake Storage Gen2 container into the tables in Pool1. The structure and data types vary by file.

You need to load the files into the tables. The solution must maintain the source data types.

What should you do?

- A. Load the data by using PySpark.
- B. Load the data by using the `OPENROWSET` Transact-SQL command in an Azure Synapse Analytics serverless SQL pool.
- C. Use a Get Metadata activity in Azure Data Factory.
- D. Use a Conditional Split transformation in an Azure Synapse data flow.

Correct Answer: B

Section: (none)

Explanation

Explanation/Reference:

Explanation:

Serverless SQL pool can automatically synchronize metadata from Apache Spark. A serverless SQL pool database will be created for each database existing in serverless Apache Spark pools. Serverless SQL pool enables you to query data in your data lake. It offers a T-SQL query surface area that accommodates semi-structured and unstructured data queries.

To support a smooth experience for in place querying of data that's located in Azure Storage files, serverless SQL pool uses the `OPENROWSET` function with additional capabilities.

The easiest way to see to the content of your JSON file is to provide the file URL to the `OPENROWSET` function, specify `csv FORMAT`.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/query-json-files>

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/query-data-storage>

QUESTION 13

You are designing a date dimension table in an Azure Synapse Analytics dedicated SQL pool. The date dimension table will be used by all the fact tables.

Which distribution type should you recommend to minimize data movement?

- A. HASH
- B. REPLICATE
- C. ROUND_ROBIN

Correct Answer: B

Section: (none)

Explanation

Explanation/Reference:
Explanation:

A replicated table has a full copy of the table available on every Compute node. Queries run fast on replicated tables since joins on replicated tables don't require data movement. Replication requires extra storage, though, and isn't practical for large tables.

Incorrect Answers:

C: A round-robin distributed table distributes table rows evenly across all distributions. The assignment of rows to distributions is random. Unlike hash-distributed tables, rows with equal values are not guaranteed to be assigned to the same distribution.

As a result, the system sometimes needs to invoke a data movement operation to better organize your data before it can resolve a query.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-distribute>

QUESTION 14
HOTSPOT

From a website analytics system, you receive data extracts about user interactions such as downloads, link clicks, form submissions, and video plays.

The data contains the following columns:

Name	Sample value
Date	15 Jan 2021
EventCategory	Videos
EventAction	Play
EventLabel	Contoso Promotional
ChannelGrouping	Social
TotalEvents	150
UniqueEvents	120
SessionsWithEvents	99

You need to design a star schema to support analytical queries of the data. The star schema will contain four tables including a date dimension.

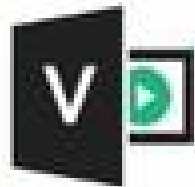
To which table should you add each column? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:



This PDF is generated automatically by **Vizle**.
Slides created *only for a few minutes* of your Video.



For the full PDF, please **Login to Vizle**.

<https://vizle.offnote.co> (Login via Google, top-right)

Stay connected with us:

Join us on **Facebook, Discord, Quora, Telegram**.