Does the solution meet the goal?

A. Yes
B. No

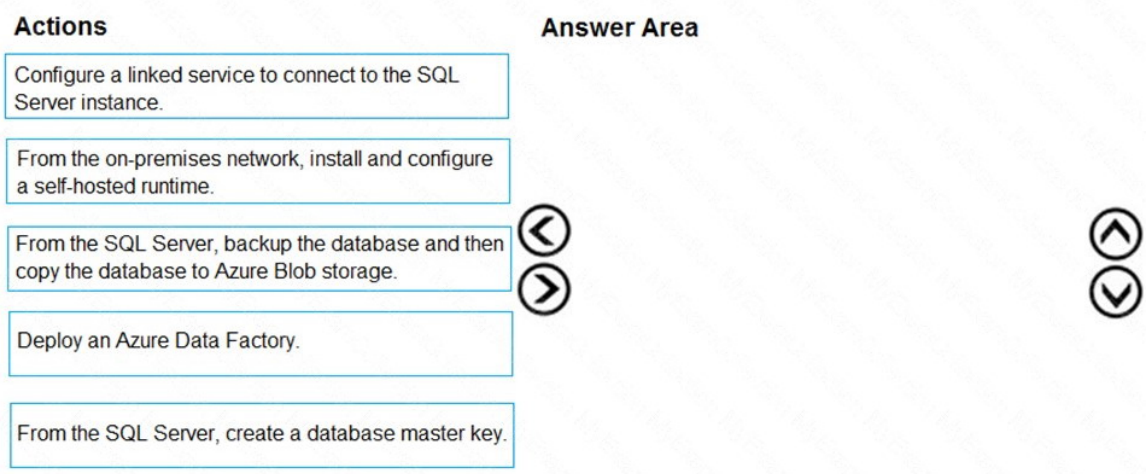
Correct Answer: B

QUESTION 51
DRAG DROP
Your company has on-premises Microsoft SQL Server instance.

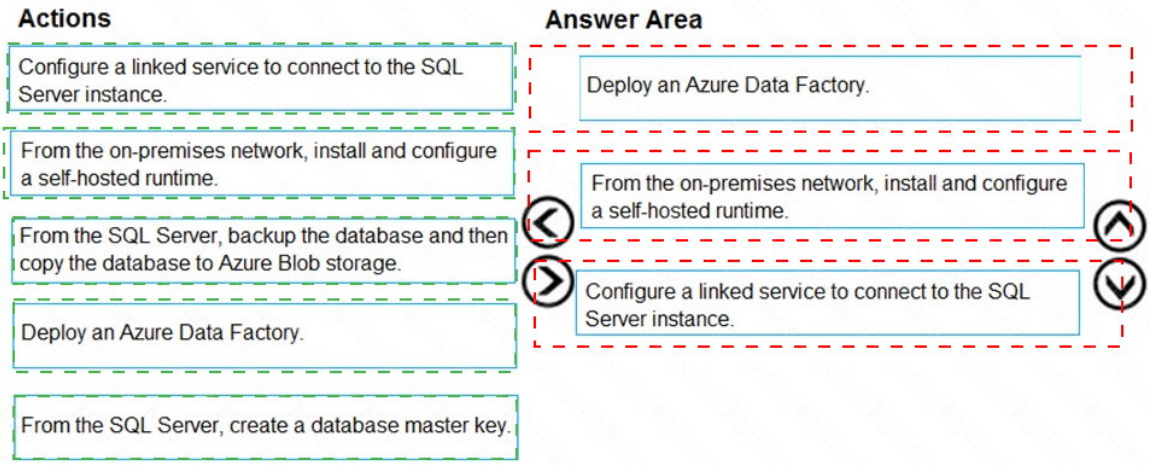
The data engineering team plans to implement a process that copies data from the SQL Server instance to Azure Blob storage. The process must orchestrate and manage the data lifecycle.

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.



Correct Answer:



QUESTION 52
Note: This question is part of series of questions that present the same scenario. Each question in the series contain a unique solution. Determine whether the solution meets the stated goals.

You develop data engineering solutions for a company.

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

- Scale to minimize costs
- Be monitored for cluster performance

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

Solution: Monitor cluster load using the Ambari Web UI.

Does the solution meet the goal?

A. Yes
B. No

Correct Answer: B
Explanation:
Ambari Web UI does not provide information to suggest how to scale.

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

References:
https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop-oms-log-analytics-tutorial
https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop-manage-ambari

QUESTION 53
The data engineering team manages Azure HDInsight clusters. The team spends a large amount of time creating and destroying clusters daily because most of the data pipeline process runs in minutes.

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

What should you implement?

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

Correct Answer: C
Explanation:

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

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

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

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

QUESTION 54
A company has a SaaS solution that uses Azure SQL Database with elastic pools. The solution contains a dedicated database for each customer organization. Customer organizations have peak usage at different periods during the year.

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

Which option or options should you configure?

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

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

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

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

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

QUESTION 55
Each day, company plans to store hundreds of files in Azure Blob Storage and Azure Data Lake Storage. The company uses the parquet format.

You must develop a pipeline that meets the following requirements:

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

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

Which data technology should you implement?

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

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

Python can be used to develop Storm components.

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

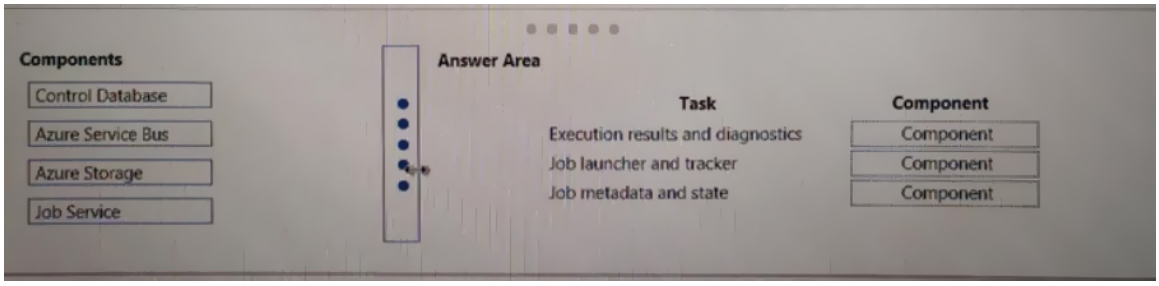
QUESTION 56
DRAG DROP
Your company uses Microsoft Azure SQL Database configure with Elastic pool. You use Elastic Database jobs to run queries across all databases in the pod.

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

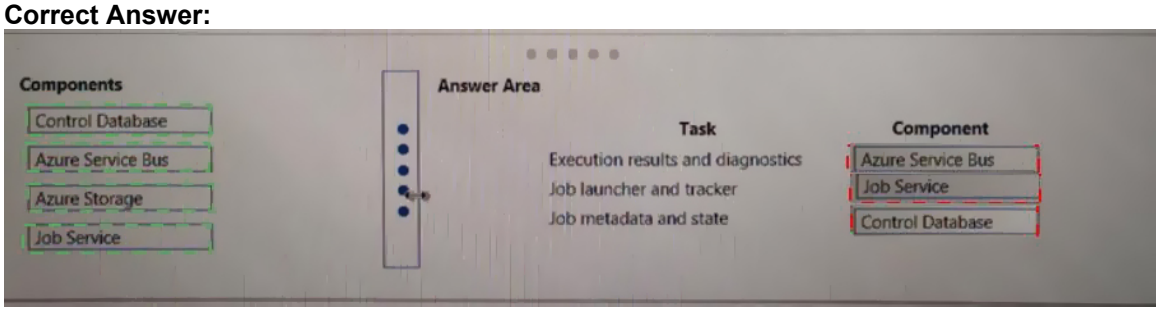
You need to determine the component responsible for running job service tasks.

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.



Correct Answer:



QUESTION 57
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