# Question49

Case Study

Complete the Case Study

* Background

Company A develops inventory management software. Its flagship product allows employees to scan product barcodes in stores and have that data delivered to a central repository. A website allows supervisors to view the data sent to the repository.

* Inventory Data

Inventory data consists of the following fields:

* UPC
* Name
* Description
* Quantity
* Store ID
* Aisle Number
* Price
* Expiration Date
* Technical Solution

Inventory data is currently stored in two Microsoft SQL Server databases. One database resides in California, and the other database resides in New York. Over 200 terabytes (TB) of total data is stored across the two databases. The scanners submit inventory data to an application server over HTTPS. A service on the application server then analyzes the data and sends it to the databases.  
  
The new solution must allow processing of the inventory data in batches every hour. After the data is processed, it must be kept for at least two years. It must be stored in such a way that parallel queries can be run against the data.  
  
Business stakeholders must be able to graphically visualize the data without writing any code. Data engineers must be able to graphically visualize the data by using Python.

* Data Engineering Requirements

The data engineers at your company are familiar with C#, Python and SQL. Any recommended solution must take advantage of their existing skills.

## Question 49.1

You need to design a solution for storing the initial inventory data.  
  
Which resource should you use?

Complete the Case Study

* Background
* Inventory Data
* Technical Solution
* Data Engineering Requirements
* Question 1
* Question 2
* Question 3
* Question 4
* Question 5

Power BI

Azure SQL Data Warehouse

Azure Data Lake

Event Hub

## Question 49.2

You need to design a solution for analyzing the inventory data by using C# user-defined functions (UDFs) when it arrives.  
  
Which resource should you use?

Complete the Case Study

* Background
* Inventory Data
* Technical Solution
* Data Engineering Requirements
* Question 1
* Question 2
* Question 3
* Question 4
* Question 5

Power BI

Azure Data Lake Analytics

IoT Hub

Azure Databricks

## Question 49.3

You need to design a solution for storing the data long-term.  
  
Which resource should you use?

Complete the Case Study

* Background
* Inventory Data
* Technical Solution
* Data Engineering Requirements
* Question 1
* Question 2
* Question 3
* Question 4
* Question 5

Azure SQL Data Warehouse

Azure SQL Database

Azure Data Factory

Azure Databricks

## Question 49.4

You need to design a solution to allow data engineers to visualize the data.  
  
What should you use?

Complete the Case Study

* Background
* Inventory Data
* Technical Solution
* Data Engineering Requirements
* Question 1
* Question 2
* Question 3
* Question 4
* Question 5

Azure Databricks

Stream Analytics

Power BI

Azure Data Lake

## Question 49.5

You need to design a solution to allow business stakeholders to visualize the data.  
  
What should you use?

Complete the Case Study

* Background
* Inventory Data
* Technical Solution
* Data Engineering Requirements
* Question 1
* Question 2
* Question 3
* Question 4
* Question 5

Power BI

Stream Analytics

Azure Databricks

Azure Data Lake

# Question54

Case Study

Complete the Case Study

* Overview

You are the database administrator for your company. The company is a reseller of internet information about financial and distribution markets.  
  
The company is planning an investment into Azure and is looking for the right type of data platform to be able to consume data from all the various incoming data feeds. The feeds consist of Comma Separated Values (CSV) files and Parquet format files.   
  
The company wants to find the simplest way to ingest the data, transform it into a relational format, and make it readily available for consumers to view as well as ensuring efficient query latency.

## Question 54.1

You need to choose the data platform Azure to use as the initial destination of the incoming data feeds.  
  
Which platform should you use?

Complete the Case Study

* Overview
* Question 1
* Question 2
* Question 3
* Question 4

Azure Synapse Analytics

Azure Data Factory

Azure PolyBase

Azure Data Lake

## Question 54.2

You need to choose the solution to transform the data from the various formats and import the data into a relational platform. You must select the simplest implementation path.  
  
Which solution should you use?

Complete the Case Study

* Overview
* Question 1
* Question 2
* Question 3
* Question 4

PolyBase

Azure Synapse Analytics

Azure Data Factory

Azure Data Lake

## Question 54.3

The company wants to automate the copying of data feeds to Azure to gain more efficiency and bring the data to their users faster.  
  
You need to choose the Azure solution to move the data from on-premises to Azure.  
  
Which solution should you use?

Complete the Case Study

* Overview
* Question 1
* Question 2
* Question 3
* Question 4

Azure Data Factory

Azure Synapse Alalytics

Azure Data Lake

Azure Polybase

## Question 54.4

For the final step in the processing of the data, you need to select the right destination Azure repository for the transformed data.  
  
Which solution should you use for the final repository of the data?

Complete the Case Study

* Overview
* Question 1
* Question 2
* Question 3
* Question 4

Azure Polybase

Azure Data Lake

Azure Data Factory

Azure Synapse Analytics

# Question110

You manage an on-premises Microsoft SQL Server 2019 instance and an Azure SQL Database Managed instance.  
  
A data engineer is writing a query in an on-premises application that needs to join data stored in a database on the Azure SQL Database managed instance. This query will be used to submit data to an Event Hub every minute for batch processing.  
  
You need to recommend a solution to prepare the environment.  
  
What should you recommend?

Choose the correct answer

Create a master key on the Azure SQL Database managed instance.

Create a linked server on the on-premises instance.

Enable PolyBase on the Azure SQL Database managed instance.

Enable PolyBase on the on-premises instance.

# Question111

You are designing a solution that allows sports leagues to send scores to the cloud. The solution must allow streaming of over 5,000 HTTPS requests each day. However, the payload of each request is small.  
  
You need to choose a resource for initially receiving the data. Your solution must be cost effective.  
  
Which resource should you use?

Choose the correct answer

Data Factory

Data Lake

Event Hub

IoT Hub

# Question112

You are designing a big data batch processing and streaming solution.  
  
You need to choose the most appropriate resource for different scenarios.  
  
Which resources should you choose? To answer, drag the appropriate resource to each scenario. A resource may be used once, more than once, or not at all.

Drag and drop the answers

https://pts.measureup.com/web/instances/MUP/assets/images/DP-201/DP-201_64233/StreamAnalytics.png

https://pts.measureup.com/web/instances/MUP/assets/images/DP-201/DP-201_64233/Databricks.png

https://pts.measureup.com/web/instances/MUP/assets/images/DP-201/DP-201_64233/DataFactory.png

https://pts.measureup.com/web/instances/MUP/assets/images/DP-201/DP-201_64233/IoTHub.png

https://pts.measureup.com/web/instances/MUP/assets/images/DP-201/DP-201_64233/StreamAnalytics.png

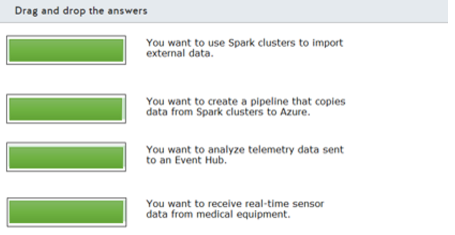
https://pts.measureup.com/web/instances/MUP/assets/images/DP-201/DP-201_64233/Databricks.png

https://pts.measureup.com/web/instances/MUP/assets/images/DP-201/DP-201_64233/DataFactory.png

https://pts.measureup.com/web/instances/MUP/assets/images/DP-201/DP-201_64233/IoTHub.png

https://pts.measureup.com/web/instances/MUP/assets/images/DP-201/DP-201_64233/SQLDataWarehouse.png

https://pts.measureup.com/web/instances/MUP/assets/images/DP-201/DP-201_64233/DataLake.png



# Question113

You are designing a big data streaming solution.  
  
You need to choose the most appropriate resource for different scenarios.  
  
Which Azure resource should you choose? To answer, drag the appropriate resource to each scenario. A resource may be used once, more than once, or not at all.

Drag and drop the answers

https://pts.measureup.com/web/instances/MUP/assets/images/DP-201/DP-201_64236/StreamAnalytics(2).png

https://pts.measureup.com/web/instances/MUP/assets/images/DP-201/DP-201_64236/SQLDataWarehouse(2).png

https://pts.measureup.com/web/instances/MUP/assets/images/DP-201/DP-201_64236/StreamAnalytics(2).png

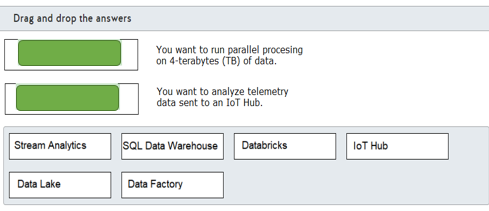
https://pts.measureup.com/web/instances/MUP/assets/images/DP-201/DP-201_64236/SQLDataWarehouse(2).png

https://pts.measureup.com/web/instances/MUP/assets/images/DP-201/DP-201_64236/Databricks(2).png

https://pts.measureup.com/web/instances/MUP/assets/images/DP-201/DP-201_64236/IoTHub(2).png

https://pts.measureup.com/web/instances/MUP/assets/images/DP-201/DP-201_64236/DataLake(2).png

https://pts.measureup.com/web/instances/MUP/assets/images/DP-201/DP-201_64236/DataFactory(2).png



# Question115

You are helping design a big data batch processing solution for long-running batch jobs. You need to select a batch processing technology that supports relational data storage and a batch job pricing model. The data store should support Azure Active Directory (Azure AD) authentication.  
  
Which should you choose?

Choose the correct answer

Azure Databricks

Azure SQL Data Warehouse

Azure Data Lake Analytics

HDInsight

# Question116

You are designing a data processing solution that will process data from several diverse on-premises and cloud-based data stores.  
  
You need to identify a data load solution to load data from the sources into an Azure SQL Data Warehouse.  
  
What should you use?

Choose the correct answer

Azure Data Factory

Oozie on HDInsight

SQL Server Integration Services (SSIS)

Azure Data Migration Assistant

# Question117

You are designing a batch processing solution that will process large quantities of data daily and load the results into an Azure SQL Data Warehouse store.  
  
You need to ensure high availability for your solution. You want to minimize costs and management overhead.  
  
Which two actions should you perform? Each correct answer presents a complete solution.

Choose the correct answers

Deploy multiple instances in different regions configured for failover with the same batch and storage accounts in each region.

Deploy multiple instances in different regions configured for failover with different batch and storage accounts in each region.

Deploy multiple instances in different regions configured as a shared workload with different batch and storage accounts in each region.

Deploy multiple instances in the same region configured for failover with different batch and storage accounts for each instance.

Deploy multiple instances in different regions configured as a shared workload with the same batch and storage accounts in each region.

**Explanation**

You should configure your solution based on one of these two models:

* Deploy multiple instances in different regions configured for failover with different batch and storage accounts in each region.
* Deploy multiple instances in different regions configured as a shared workload with different batch and storage accounts in each region.

You need to deploy multiple instances in two or more regions, either in a failover cluster configuration or as a shared workload, to ensure high availability. This requires you to configure different accounts in each region.  
  
You should not choose a solution that uses the same account across multiple regions. This is not supported as a solution.  
  
You should not deploy a solution that relies on a single region. This does not provide high availability in case of a regional failure.  
  
**References**  
  
[Design your application for high availability](https://docs.microsoft.com/en-us/azure/batch/high-availability-disaster-recovery)

# Question118

You are using an Azure Databricks interactive cluster for batch processing. The cluster is configured for auto-termination.  
  
You need to ensure that cluster configuration remains available for at least 180 days after termination, including cluster permissions.  
  
What should you do?

Choose the correct answer

Configure the cluster for autostart.

Pin the cluster.

Clone the cluster after termination.

Manually terminate the cluster.

# Question119

You are designing a Spark batch job to process daily log activity. The job should be scheduled to run each day. The job should display status information on the company intranet when it runs.  
  
You need to choose the appropriate technologies to meet the requirements.  
  
To answer, drag the appropriate technologies to each requirement. A technology may be used once, more than once, or not at all.

Drag and drop the answers

https://pts.measureup.com/web/instances/MUP/assets/images/DP-201/DP-201_64374/DP-201-034c.jpeg

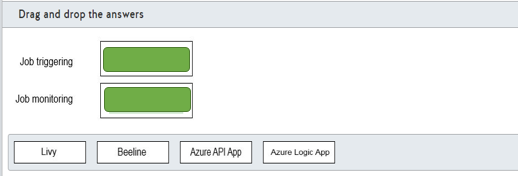
https://pts.measureup.com/web/instances/MUP/assets/images/DP-201/DP-201_64374/DP-201-034d.jpeg

https://pts.measureup.com/web/instances/MUP/assets/images/DP-201/DP-201_64374/DP-201-034c.jpeg

https://pts.measureup.com/web/instances/MUP/assets/images/DP-201/DP-201_64374/DP-201-034d.jpeg

https://pts.measureup.com/web/instances/MUP/assets/images/DP-201/DP-201_64374/DP-201-034a.jpeg

https://pts.measureup.com/web/instances/MUP/assets/images/DP-201/DP-201_64374/DP-201-034b.jpeg



# Question120

You are designing an Azure Data Factory pipeline that will be used for processing data. The pipeline must support processing data that is stored in general-purpose standard Azure storage. The environment for data processing should be created on-demand and should be removed when processing is complete.  
  
You need to recommend the appropriate transformation activity.  
  
What should you recommend?

Choose the correct answer

Data Lake Analytics U-SQL activity

HDInsight Pig activity

Databricks Python

Databricks Notebook

# Question121

You are designing a data analytics solution for use with large data sets. The solution should use Azure Notebooks and support Notebook scheduling. The solution needs to provide visualization through Power BI visualization. The solution should also support cluster automation to run scheduled jobs on a new cluster and terminate the cluster when processing is complete.  
  
Which solution should you recommend?

Choose the correct answer

Azure Databricks

Azure Batch

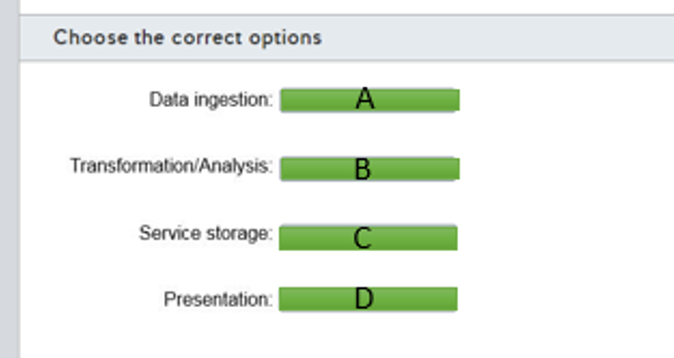
Azure Machine Learning (ML) Studio

Azure Stream Analytics

# Question122

You are developing a batch processing solution that collects input from IoT device events for analysis and presentation.  
  
You need to identify the appropriate technologies for each of the remaining process steps.  
  
What technology should you use for each process step? To answer, select the appropriate options from the drop-down menus.

Choose the correct options



A)

1. HDInsight Kafka
2. Azure Databricks
3. Cosmos DB
4. Power Bi

B)

1. HDInsight Kafka
2. Azure Databricks
3. Cosmos DB
4. Power Bi

C)

1. HDInsight Kafka
2. Azure Databricks
3. Cosmos DB
4. Power Bi

D)

1. HDInsight Kafka
2. Azure Databricks
3. Cosmos DB
4. Power Bi

# Question123

You want to test a batch processing solution that supports ingested data through Azure Data Factory batches, performs data analysis, and stores the result in Azure SQL Data Warehouse.  
  
You need to suggest a solution that performs Spark analyses and creates and deletes clusters on-demand.  
  
What should you use?

Choose the correct answer

Azure Databricks

Azure Kafka in HDInsight

Azure Data Lake Storage

Azure Cosmos DB

# Question124

You are designing a batch processing solution that uses Azure Data Lake Storage as its data store. You estimate that the solution must support at least 4000 IOPS.  
  
You need to select the appropriate disk type of your solution. You want to minimize costs.  
  
What should you choose?

Choose the correct answer

Standard SSD

Standard HDD

Premium SSD

Ultra SSD

# Question125

You are asked to implement a batch processing system in Azure that will support R and Spark SQL. The users of the system will also require connected notebook support. Support for fast cluster startup times and autoscaling is also required.  
  
You need to choose a batch processing solution to meet the requirements.  
  
Which batch processing solution should you use?

Choose the correct answer

Azure Data Lake Analytics

HDInsight

Azure Synapse Analytics

Azure Databricks