

# Microsoft

## Exam Questions 70-774

Perform Cloud Data Science with Azure Machine Learning (beta)



**NEW QUESTION 1**

You are analyzing taxi trips in New York City. You leverage the Azure Data Factory to create data pipelines and to orchestrate data movement. You plan to develop a predictive model for 170 million rows (37 GB) of raw data in Apache Hive by using Microsoft R Server to identify which factors contribute to the passenger tipping behavior. All of the platforms that are used for the analysis are the same. Each worker node has eight processor cores and 26 GB of memory. Which type of Azure HDInsight cluster should you use to produce results as quickly as possible?

- A. Hadoop
- B. HBase
- C. Interactive Hive
- D. Spark

**Answer:** D

**Explanation:** References:

<https://azure.microsoft.com/en-gb/blog/general-availability-of-hdinsight-interactive-query-blazing-fast-data-war>

**NEW QUESTION 2**

Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.

You plan to create a predictive analytics solution for credit risk assessment and fraud prediction in Azure Machine Learning. The Machine Learning workspace for the solution will be shared with other users in your organization. You will add assets to projects and conduct experiments in the workspace.

The experiments will be used for training models that will be published to provide scoring from web services. The experiment for fraud prediction will use Machine Learning modules and APIs to train the models and will predict probabilities in an Apache Hadoop ecosystem.

End of repeated scenario.

You need to alter the list of columns that will be used for predicting fraud for an input web service endpoint. The columns from the original data source must be retained while running the Machine Learning experiment.

Which module should you add after the web service input module and before the prediction module?

- A. Edit Metadata
- B. Import Data
- C. SMOTE
- D. Select Columns in Dataset

**Answer:** D

**NEW QUESTION 3**

Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.

Start of repeated scenario

You plan to use Azure platform tools to detect and analyze food items in smart refrigerators. To provide families with an integrated experience for grocery shopping and cooking, the refrigerators will connect to other smart appliances, such as stoves and microwave ovens, on a LAN.

You plan to build an object recognition model by using the Microsoft Cognitive Toolkit. The object recognition model will receive input from the connected devices and send results to applications.

The training data will be derived from more than 10 TB of images. You will convert the raw images to the sparse format.

End of repeated scenario.

You need to ensure that a web service endpoint can receive image data and use an object recognition model to return the expected object and the confidence level of the model. The solution must minimize the effort required to generate the client code to access the web service.

Which resource should you use?

- A. the edX Data Science Learning Dashboard
- B. Azure Machine Learning Studio
- C. Cortana Intelligence Gallery
- D. the Data Science Virtual Machine

**Answer:** B

**NEW QUESTION 4**

You have a dataset that is missing values in a column named Column3. Column3 is correlated to two columns named Column4 and Column5.

You need to improve the accuracy of the dataset, while minimizing data loss. What should you do?

- A. Replace the missing values in Column3 by using probabilistic Principal Component Analysis (PCA).
- B. Remove all of the rows that have the missing values in Column4 and Column5.
- C. Replace the missing values in Column3 with a mean value.
- D. Remove the rows that have the missing values in Column3.

**Answer:** A

**NEW QUESTION 5**

The manager of a call center reports that staffing the center is difficult because the number of calls is unpredictable. You have historical data that contains information about the calls.

You need to build an Azure Machine Learning experiment to predict the number of total calls each hour. Which model should you use?

- A. Multiclass Logistic Regression
- B. Boosted Decision Tree Regression
- C. Decision Forest Regression

D. Poisson Regression

**Answer:** D

#### NEW QUESTION 6

Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.

You plan to create a predictive analytics solution for credit risk assessment and fraud prediction in Azure Machine Learning. The Machine Learning workspace for the solution will be shared with other users in your organization. You will add assets to projects and conduct experiments in the workspace.

The experiments will be used for training models that will be published to provide scoring from web services. The experiment for fraud prediction will use Machine Learning modules and APIs to train the models and will predict probabilities in an Apache Hadoop ecosystem.

You plan to configure the resources for part of a workflow that will be used to preprocess data from files stored in Azure Blob storage. You plan to use Python to preprocess and store the data in Hadoop.

You need to get the data into Hadoop as quickly as possible.

Which three actions should you perform? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

- A. Create an Azure virtual machine (VM), and then configure MapReduce on the VM.
- B. Create an Azure HDInsight Hadoop cluster.
- C. Create an Azure virtual machine (VM), and then install an IPython Notebook server.
- D. Process the files by using Python to store the data to a Hadoop instance.
- E. Create the Machine learning experiment, and then add an Execute Python Script module.

**Answer:** BDE

#### NEW QUESTION 7

Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.

A travel agency named Margie's Travel sells airline tickets to customers in the United States.

Margie's Travel wants you to provide insights and predictions on flight delays. The agency is considering implementing a system that will communicate to its customers as the flight departure nears about possible delays due to weather conditions. The flight data contains the following attributes:

The weather data contains the following attributes: AirportID, ReadingDate (YYYY/MM/DD HH), SkyConditionVisibility, WeatherType, WindSpeed, StationPressure, PressureChange, and HourlyPrecip.

You need to use historical data about on-time flight performance and the weather data to predict whether the departure of a scheduled flight will be delayed by more than 30 minutes.

Which method should you use?

- A. clustering
- B. linear regression
- C. classification
- D. anomaly detection

**Answer:** C

**Explanation:** References:

<https://gallery.cortanaintelligence.com/Experiment/Binary-Classification-Flight-delay-prediction-3>

#### NEW QUESTION 8

Note: This question is part of a series of questions that use the same or similar answer choices. An answer choice may be correct for more than one question in the series. Each question is independent of the other questions in this series. Information and details provided in a question apply only to that question.

You need to transform the columns in a dataset. The resulting data must be mean centered and have a variance of 1. The solution must use a native module.

Which module should you use?

- A. Execute Python Script
- B. Import Data
- C. Edit Metadata
- D. Select Columns in Dataset
- E. Clean Missing Data
- F. Tune Model Hyperparameters
- G. Clip Values
- H. Normalize Data

**Answer:** H

#### NEW QUESTION 9

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 sections, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are working on an Azure Machine Learning experiment. You have the dataset configured as shown in the following table.

Model	Mean absolute error (MAE)
Boosted decision tree	.2
Relative absolute error (RAE)	.43

You need to ensure that you can compare the performance of the models and add annotations to the results. Solution: You connect the Score Model modules from each trained model as inputs for the Evaluate Model

module, and use the Execute R Script module.

Does this meet the goal?

- A. Yes
- B. No

**Answer:** A

**Explanation:** References:  
<https://msdn.microsoft.com/en-us/library/azure/dn905915.aspx>

#### NEW QUESTION 10

You plan to use the Data Science Virtual Machine for development, but you are unfamiliar with R scripts. You need to generate R code for an experiment. Which IDE should you use?

- A. XgBoost
- B. Rattle
- C. Vowpal Wabbit
- D. R Tools for Visual Studio

**Answer:** C

**Explanation:** References:  
<https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/provision-vm>

#### NEW QUESTION 10

You deploy Microsoft Data Management Gateway.  
You plan to use the Import Data module in Azure Machine Learning Studio to import data from an on-premises Microsoft SQL Server instance. Which operation can you perform?

- A. Write the data back to the on-premises SQL Server instance.
- B. Filter the data as the data is being read by using the Import Data module.
- C. Run a Transact-SQL query and use SQL views to filter the data as the data is being read.
- D. Access the on-premises SQL Server instance without using credentials, and then import the data.

**Answer:** D

#### NEW QUESTION 13

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 sections, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.  
You are designing an Azure Machine Learning workflow.  
You have a dataset that contains two million large digital photographs. You plan to detect the presence of trees in the photographs.  
You need to ensure that your model supports the following: Solution: You create an endpoint to the Computer vision API. Does this meet the goal?

- A. Yes
- B. No

**Answer:** B

#### NEW QUESTION 16

You are building a classification experiment in Azure Machine Learning.  
You need to ensure that you can use the Evaluate Model module the experiment.  
Which two actions should you perform? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

- A. Connect the input of the Score Model modules to the output of the Evaluate Model module.
- B. Connect the input of the Score Model modules to the output of the Train Model modules and the output Split Data modules.
- C. Connect the output of the Score Model modules to the input of the Evaluate Model module.
- D. Connect the output of the Score Model modules to the input of the Train Model modules and the input of the Split Data modules.

**Answer:** AB

#### NEW QUESTION 20

You need to use R code in a Transact-SQL statement to merge the repeating values 1 through 6 with Col1 in a table.  
Which statement should you use? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

## Answer Area

execute

▼

'Insert #MyData SELECT\* FROM #MyData1, #MyData2'  
sp\_execute\_external\_script  
sp\_execute\_remote  
sp\_executesql

```

@language = N'R'
, @script = N'
    df1 <- as.data.frame( array(1:6) );

    df2 <- as.data.frame( c(
        , df1 ));

    OutputDataSet <- df2'

, @input_data_1 = N' SELECT [Col1] from #MyData;
WITH RESULT SETS (( [Col2] int not null, [Col3] int not null ));

```

▼

Input\_Data\_1  
InputDataSet  
InputDataTable

**Answer:**

**Explanation:** References:

<https://docs.microsoft.com/en-us/sql/advanced-analytics/tutorials/rtsql-r-and-sql-data-types-and-data-objects>

### NEW QUESTION 23

You plan to use Azure Machine Learning to develop a predictive model. You plan to include an Execute Python Script module. What capability does the module provide?

- A. Outputting a file to a network location.
- B. Performing interactive debugging of a Python script.
- C. Saving the results of a Python script run in a Machine Learning environment to a local file.
- D. Visualizing univariate and multivariate summaries by using Python code.

**Answer:** D

**Explanation:** References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio/execute-python-scripts>

### NEW QUESTION 27

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 sections, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are working on an Azure Machine Learning experiment. You have the dataset configured as shown in the following table.

Model	Mean absolute error (MAE)
Boosted decision tree	.2
Relative absolute error (RAE)	.43

You need to ensure that you can compare the performance of the models and add annotations to the results. Solution: You connect the Score Model modules from each trained model as inputs for the Evaluate Model

module, and then save the results as a dataset.

Does this meet the goal?

- A. Yes
- B. No

**Answer:** A

**Explanation:** References:

<https://msdn.microsoft.com/en-us/library/azure/dn905915.aspx>

### NEW QUESTION 30

Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.

A travel agency named Margie's Travel sells airline tickets to customers in the United States.



Margie’s Travel wants you to provide insights and predictions on flight delays. The agency is considering implementing a system that will communicate to its customers as the flight departure nears about possible delays due to weather conditions. The flight data contains the following attributes:  
 The weather data contains the following attributes: AirportID, ReadingDate (YYYY/MM/DD HH), SkyConditionVisibility, WeatherType, WindSpeed, StationPressure, PressureChange, and HourlyPrecip.  
 You need to remove the bias and to identify the columns in the input dataset that have the greatest predictive power.  
 Which module should you use for each requirement? To answer, drag the appropriate modules to the correct requirements. Each module 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.

### Modules

Cross-validate Model

Evaluate Model

Filter and Sample

Filter Based Feature Selection Module

Parameter Sweep

Tune Model Hyperparameters

### Answer Area

Remove bias:

Module

Identify the columns that have the greatest predictive power:

Module

**Answer:**

**Explanation:** References:  
<https://gallery.cortanaintelligence.com/Experiment/Binary-Classification-Flight-delay-prediction-3>  
<https://msdn.microsoft.com/library/azure/038d91b6-c2f2-42a1-9215-1f2c20ed1b40>

**NEW QUESTION 34**  
 You have data about the following:  
 You need to predict whether a user will like a particular movie. Which Matchbox recommender should you use?

A. Item Recommendation  
 B. Related Items  
 C. Rating Prediction  
 D. Related Users

**Answer:** C

**Explanation:** References:  
<https://msdn.microsoft.com/en-us/library/azure/dn905970.aspx#RatingPredictionOptions>

**NEW QUESTION 36**  
 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 sections, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.  
 You are working on an Azure Machine Learning experiment. You have the dataset configured as shown in the following table.

Model	Mean absolute error (MAE)
Boosted decision tree	.2
Relative absolute error (RAE)	.43

You need to ensure that you can compare the performance of the models and add annotations to the results. Solution: You save the output of the Score Model modules as a combined set, and then use the Project Columns module to select the MAE.  
 Does this meet the goal?

A. Yes  
 B. No

**Answer:** A

**Explanation:** <https://msdn.microsoft.com/en-us/library/azure/dn905915.aspx>

**NEW QUESTION 39**  
 You have data about the following:

- Users
- Movies

- User ratings of the movies

You need to predict whether a user will like a particular movie. Which Matchbox recommender should you use?

- A. Rating Prediction
- B. Related Users
- C. Item Recommendation
- D. Related Items

**Answer:** A

#### NEW QUESTION 43

You are building an Azure Machine Learning experiment.

You need to transform a string column that has 47 distinct values into a binary indicator column. The solution must use the One-vs-All Multiclass model.

Which module should you use?

- A. Select Column Transform
- B. Convert to Indicator Values
- C. Group Categorical Values
- D. Edit Metadata

**Answer:** B

#### NEW QUESTION 44

You are performing exploratory analysis of files that are encoded in a complex proprietary format. The format requires disk intensive access to several dependent files in HDFS.

You need to build an Azure Machine Learning model by using a canopy clustering algorithm. You must ensure that changes to proprietary file formats can be maintained by using the least amount of effort.

Which Machine Learning library should you use?

- A. MicrosoftML
- B. scikit-learn
- C. SparkR
- D. Mahout

**Answer:** C

#### NEW QUESTION 45

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 section, 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 ML experiment that contains an intermediate dataset. You need to explore data from the intermediate dataset by using Jupyter.

Solution: You add a Convert to CSV module to the Azure ML experiment and then open the module output in a new notebook.

Does this meet the goal?

- A. Yes
- B. No

**Answer:** A

#### NEW QUESTION 48

Note: This question is part of a series of questions that use the same or similar answer choices. An

answer choice may be correct for more than one question in the series. Each question is independent of the other questions in this series. Information and details provided in a question apply only to that question.

You need to change a column name to a friendly name. The solution must use a native module. Which module should you use?

- A. Normalize Data
- B. Select Columns in Dataset
- C. Import Data
- D. Edit Metadata
- E. Tune Model Hyperparameters
- F. Clean Missing Data
- G. Clip Values
- H. Execute Python Script

**Answer:** D

#### NEW QUESTION 53

You need to integrate code and formatted text into an Azure Machine Learning experiment that enables interactive execution.

What should you use?

- A. a Jupyter notebook
- B. Azure Stream Analytics
- C. an Execute Python Script module
- D. an Execute R Script module

**Answer:** A

**NEW QUESTION 55**

You have an Apache Spark cluster in Azure HDInsight. The cluster includes 200 TB in five Apache Hive tables that have multiple foreign key relationships. You have an Azure Machine Learning model that was built by using SPARK Accelerated Failure Time (AFT) Survival Regression Model (spark-survreg). You need to prepare the Hive data into a single table as input for the Machine Learning model. The Hive data must be prepared in the least amount of time possible.

What should you use to prepare the data?

- A. a Hive user-defined function (UDF)
- B. Spark SQL
- C. the GPU
- D. Java Mapreduce jobs

**Answer:** A

**NEW QUESTION 58**

Note: This question is part of a series of questions that use the same or similar answer choices. An answer choice may be correct for more than one question in the series. Each question is independent of the other questions in this series. Information and details provided in a question apply only to that question. You need to use only one percent of an Apache Hive data table by conducting random sampling by groups. Which module should you use?

- A. Execute Python Script
- B. Tune Model Hyperparameters
- C. Normalize Data
- D. Select Columns in Dataset
- E. Import Data
- F. Edit Metadata
- G. Clip Values
- H. Clean Missing Data

**Answer:** A

**Explanation:** References:

<https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/sample-data-hive>

**NEW QUESTION 62**

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 section, 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 ML experiment that contains an intermediate dataset. You need to explore data from the intermediate dataset by using Jupyter.

Solution: You add a Convert to ARFF module, and then add the Execute R Script module. Does this meet the goal?

- A. Yes
- B. No

**Answer:** B

**NEW QUESTION 67**

You are building an Azure Machine Learning solution for an online retailer.

When a customer selects a product, you need to recommend products that the customer might like to purchase at the same time. The recommendation should be based on what other customers purchased when they purchased the same product.

Which model should you use?

- A. Collaborative filtering
- B. Boosted Decision Tree Regression model
- C. Two-Class boosted decision tree
- D. K-Means Clustering

**Answer:** A

**NEW QUESTION 69**

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 section, 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 ML experiment that contains an intermediate dataset. You need to explore data from the intermediate dataset by using Jupyter.

Solution: In Azure Mt Studio, you use the Save as dataset option, and then open the output in a new notebook. Does this meet the goal?

- A. Yes
- B. No

**Answer:** A

**NEW QUESTION 72**

You have an Azure Machine Learning environment. You are evaluating whether to use R code or Python.

Which three actions can you perform by using both R code and Python in the Machine Learning environment? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

- A. Preprocess, cleanse, and group data.
- B. Score a training model.



- C. Create visualizations.
- D. Create an untrained model that can be used with the Train Model module.
- E. Implement feature ranking.

Answer: ABC

NEW QUESTION 77

Note: This question is part of a series of questions that use the same or similar answer choices. An answer choice may be correct for more than one question in the series. Each question is independent of the other questions in this series. Information and details provided in a question apply only to that question. You need to remove rows that have an empty value in a specific column. The solution must use a native module. Which module should you use?

- A. Execute Python Script
- B. Tune Model Hyperparameters
- C. Normalize Data
- D. Select Columns in Dataset
- E. Import Data
- F. Edit Metadata
- G. Clip Values
- H. Clean Missing Data

Answer: H

Explanation: References:  
<https://blogs.msdn.microsoft.com/azuredev/2017/05/27/data-cleansing-tools-in-azure-machine-learning/>

NEW QUESTION 79

Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series. You plan to create a predictive analytics solution for credit risk assessment and fraud prediction in Azure Machine Learning. The Machine Learning workspace for the solution will be shared with other users in your organization. You will add assets to projects and conduct experiments in the workspace. The experiments will be used for training models that will be published to provide scoring from web services. The experiment for fraud prediction will use Machine Learning modules and APIs to train the models and will predict probabilities in an Apache Hadoop ecosystem. You finish training the model and are ready to publish a predictive web service that will provide the users with the ability to specify the data source and the save location of the results. The model includes a Split Data module. Which two actions should you perform to convert the Machine Learning experiment to a predictive web service? To answer, drag the appropriate actions to the correct targets. Each action 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.

Actions

Click Set Up Web Service for the training experiment.

Configure a web service endpoint for input and output, and then specify the parameters.

Remove the Split Data module.

Replace the Machine Learning algorithm and the train model by using a saved training model.

Save the trained model.

Answer Area

First action:

Action

Second action:

Action

Answer:

Explanation: References:  
<https://docs.microsoft.com/en-us/azure/machine-learning/studio/convert-training-experiment-to-scoring-experim>

NEW QUESTION 84

You are building an Azure Machine Learning experiment.

You are preparing the output of a Boosted Decision Tree Regression module. You add a Normalize Data module to the experiment. You need to ensure that the range of the transformation method produces an output on a scale of -1 to 1. Which transformation method should you use?

- A. MinMax
- B. TanH
- C. Logistic
- D. Zscore
- E. LogNormal

**Answer:** D

#### **NEW QUESTION 85**

You have an Azure Machine Learning experiment. You discover that a model causes many errors in a production dataset. The model causes only few errors in the training data. What is the cause of the errors?

- A. overfitting
- B. generalization
- C. underfitting
- D. a simple predictor

**Answer:** A

#### **NEW QUESTION 90**

.....

## Thank You for Trying Our Product

### We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questions and Answers in PDF Format

### 70-774 Practice Exam Features:

- \* 70-774 Questions and Answers Updated Frequently
- \* 70-774 Practice Questions Verified by Expert Senior Certified Staff
- \* 70-774 Most Realistic Questions that Guarantee you a Pass on Your First Try
- \* 70-774 Practice Test Questions in Multiple Choice Formats and Updates for 1 Year

**100% Actual & Verified — Instant Download, Please Click**  
**[Order The 70-774 Practice Test Here](#)**