



Vendor: Microsoft

Exam Code: DP-100

Exam Name: Designing and Implementing a Data Science
Solution on Azure

Version: DEMO

Case Study 1

Overview

You are a data scientist in a company that provides data science for professional sporting events. Models will use global and local market data to meet the following business goals:

QUESTION 1

You need to implement a scaling strategy for the local penalty detection data.

Which normalization type should you use?

- A. Streaming
- B. Weight
- C. Batch
- D. Cosine

Answer: C

Explanation:

Post batch normalization statistics (PBN) is the Microsoft Cognitive Toolkit (CNTK) version of how to evaluate the population mean and variance of Batch Normalization which could be used in inference Original Paper.

In CNTK, custom networks are defined using the BrainScriptNetworkBuilder and described in the CNTK network description language "BrainScript." Scenario:

Local penalty detection models must be written by using BrainScript.

References:

<https://docs.microsoft.com/en-us/cognitive-toolkit/post-batch-normalization-statistics>

Case Study 2

Overview

You are a data scientist for Fabrikam Residences, a company specializing in quality private and commercial property in the United States. Fabrikam Residences is considering expanding into Europe and has asked you to investigate prices for private residences in major European cities. You use Azure Machine Learning Studio to measure the median value of properties. You produce a regression model to predict property prices by using the Linear Regression and Bayesian Linear Regression modules.

QUESTION 2

You need to select a feature extraction method.

Which method should you use?

- A. Mutual information
- B. Mood's median test
- C. Kendall correlation
- D. Permutation Feature Importance

Answer: C

Explanation:

In statistics, the Kendall rank correlation coefficient, commonly referred to as Kendall's tau coefficient (after the Greek letter τ), is a statistic used to measure the ordinal association between two measured quantities.

It is a supported method of the Azure Machine Learning Feature selection.

Scenario: When you train a Linear Regression module using a property dataset that shows data for property prices for a large city, you need to determine the best features to use in a model. You

can choose standard metrics provided to measure performance before and after the feature importance process completes. You must ensure that the distribution of the features across multiple training models is consistent.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/feature-selection-modules>

QUESTION 3

You are developing a hands-on workshop to introduce Docker for Windows to attendees.

You need to ensure that workshop attendees can install Docker on their devices.

Which two prerequisite components should attendees install on the devices? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Microsoft Hardware-Assisted Virtualization Detection Tool
- B. Kitematic
- C. BIOS-enabled virtualization
- D. VirtualBox
- E. Windows 10 64-bit Professional

Answer: CE

Explanation:

C: Make sure your Windows system supports Hardware Virtualization Technology and that virtualization is enabled.

Ensure that hardware virtualization support is turned on in the BIOS settings. For example:



E: To run Docker, your machine must have a 64-bit operating system running Windows 7 or higher.

References:

https://docs.docker.com/toolbox/toolbox_install_windows/

<https://blogs.technet.microsoft.com/canitpro/2015/09/08/step-by-step-enabling-hyper-v-for-use-on-windows-10/>

QUESTION 4

Your team is building a data engineering and data science development environment.

The environment must support the following requirements:

- support Python and Scala
- compose data storage, movement, and processing services into automated data pipelines
- the same tool should be used for the orchestration of both data engineering and data science
- support workload isolation and interactive workloads
- enable scaling across a cluster of machines

You need to create the environment.

What should you do?

- Build the environment in Apache Hive for HDInsight and use Azure Data Factory for orchestration.
- Build the environment in Azure Databricks and use Azure Data Factory for orchestration.
- Build the environment in Apache Spark for HDInsight and use Azure Container Instances for orchestration.
- Build the environment in Azure Databricks and use Azure Container Instances for orchestration.

Answer: B

Explanation:

In Azure Databricks, we can create two different types of clusters.

Standard, these are the default clusters and can be used with Python, R, Scala and SQL

High-concurrency

Azure Databricks is fully integrated with Azure Data Factory.

Incorrect Answers:

D: Azure Container Instances is good for development or testing. Not suitable for production workloads.

References:

<https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/data-science-and-machine-learning>

QUESTION 5

You plan to build a team data science environment. Data for training models in machine learning pipelines will be over 20 GB in size.

You have the following requirements:

- Models must be built using Caffe2 or Chainer frameworks.
- Data scientists must be able to use a data science environment to build the machine learning pipelines and train models on their personal devices in both connected and disconnected network environments.
- Personal devices must support updating machine learning pipelines when connected to a network.

You need to select a data science environment.

Which environment should you use?

- Azure Machine Learning Service
- Azure Machine Learning Studio

- C. Azure Databricks
- D. Azure Kubernetes Service (AKS)

Answer: A

Explanation:

The Data Science Virtual Machine (DSVM) is a customized VM image on Microsoft's Azure cloud built specifically for doing data science. Caffe2 and Chainer are supported by DSVM.

DSVM integrates with Azure Machine Learning.

Incorrect Answers:

B: Use Machine Learning Studio when you want to experiment with machine learning models quickly and easily, and the built-in machine learning algorithms are sufficient for your solutions.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview>

QUESTION 6

You are implementing a machine learning model to predict stock prices.

The model uses a PostgreSQL database and requires GPU processing.

You need to create a virtual machine that is pre-configured with the required tools.

What should you do?

- A. Create a Data Science Virtual Machine (DSVM) Windows edition.
- B. Create a Geo AI Data Science Virtual Machine (Geo-DSVM) Windows edition.
- C. Create a Deep Learning Virtual Machine (DLVM) Linux edition.
- D. Create a Deep Learning Virtual Machine (DLVM) Windows edition.
- E. Create a Data Science Virtual Machine (DSVM) Linux edition.

Answer: E

Explanation:Incorrect Answers:

A, C: PostgreSQL (CentOS) is only available in the Linux Edition.

B: The Azure Geo AI Data Science VM (Geo-DSVM) delivers geospatial analytics capabilities from Microsoft's Data Science VM. Specifically, this VM extends the AI and data science toolkits in the Data Science VM by adding ESRI's market-leading ArcGIS Pro Geographic Information System.

D: DLVM is a template on top of DSVM image. In terms of the packages, GPU drivers etc are all there in the DSVM image. Mostly it is for convenience during creation where we only allow DLVM to be created on GPU VM instances on Azure.

References:

<https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview>

Thank You for Trying Our Product

Braindump2go Certification Exam Features:

- ★ More than **99,900** Satisfied Customers Worldwide.
- ★ Average **99.9%** Success Rate.
- ★ **Free Update** to match latest and real exam scenarios.
- ★ **Instant Download** Access! No Setup required.
- ★ Questions & Answers are downloadable in **PDF** format and **VCE** test engine format.
- ★ Multi-Platform capabilities - **Windows, Laptop, Mac, Android, iPhone, iPod, iPad**.
- ★ **100%** Guaranteed Success or **100%** Money Back Guarantee.
- ★ **Fast**, helpful support **24x7**.



View list of all certification exams: <http://www.braindump2go.com/all-products.html>



Microsoft



ORACLE



CITRIX



JUNIPER
NETWORKS



EMC²
where information lives[®]

10% Discount Coupon Code: BDN2014