Demo Questions

Microsoft DP-100 Exam

Designing and Implementing a Data Science Solution on Azure (beta)

Thank you for downloading DP-100 Exam PDF

Question #1 *Topic 1*

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.
Each correct selection is worth one point.
NOTE:

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

Correct Answer: *CE*
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 #2*Topic 1*

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?

- A. Build the environment in Apache Hive for HDInsight and use Azure Data Factory for orchestration.

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

Correct Answer: *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 #3*Topic 1*

DRAG DROP -
You are building an intelligent solution using machine learning models.
The environment must support the following requirements:
☞ Data scientists must build notebooks in a cloud environment
☞ Data scientists must use automatic feature engineering and model building in machine learning pipelines.
☞ Notebooks must be deployed to retrain using Spark instances with dynamic worker allocation.
Notebooks must be exportable to be version controlled locally.
▪
You need to create the environment.
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:

Correct Answer: *Explanation*
Step 1: Create an Azure HDInsight cluster to include the Apache Spark Mlib library
Step 2: Install Microsot Machine Learning for Apache Spark
You install AzureML on your Azure HDInsight cluster.
Microsoft Machine Learning for Apache Spark (MMLSpark) provides a number of deep

learning and data science tools for Apache Spark, including seamless integration of Spark Machine Learning pipelines with Microsoft Cognitive Toolkit (CNTK) and OpenCV, enabling you to quickly create powerful, highly-scalable predictive and analytical models for large image and text datasets.
Step 3: Create and execute the Zeppelin notebooks on the cluster
Step 4: When the cluster is ready, export Zeppelin notebooks to a local environment. Notebooks must be exportable to be version controlled locally.
References:
https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-zeppelin-notebook https://azuremlbuild.blob.core.windows.net/pysparkapi/intro.html

Question #4*Topic 1*

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?

- A. Azure Machine Learning Service
- B. Azure Machine Learning Studio
- C. Azure Databricks
- D. Azure Kubernetes Service (AKS) A

Correct Answer: *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 #5*Topic 1*

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. E

Correct Answer: *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