AZURE FOR DATA SCIENCE RESEARCH TRAINING DESCRIPTION

Azure Data Science 101 - Storage and Azure Machine Learning

Azure Storage

Azure Storage is a set of services for storing data in the cloud. Of particular interest to researchers is Azure Blob Storage, which serves as a source of input and output for Azure data services. In this session, participants learn how to create storage accounts and storage containers, upload and download blobs using the cross-platform Azure Storage Explorer, and securely share data hosted in blob storage using shared-access signatures.

Azure Machine Learning

Azure Machine Learning is a cloud-based predictive-analytics service that offers a streamlined experience for data scientists of all skill levels. In this module, participants use the interactive Azure Machine Learning Studio to build, train, and score a machine-learning model. Then they put the model to work performing predictive analytics.

Azure Data Science 301 – Big Data

IoT and Azure Stream Analytics

Azure Stream Analytics is a service that enables researchers to query and analyze high-velocity data streaming from Internet-of-Things (IoT) devices and other data sources in real time. In this module, participants combine Azure Stream Analytics with Azure Event Hubs to perform real-time analytics on data emanating from simulated ATM machines.

Apache Spark for Azure HDInsight

Apache Spark is an open-source engine for analyzing large data sets extremely fast using clusters of inexpensive computers. Building and maintaining a Spark cluster is no easy task, but with Azure HDInsight, you can deploy a Spark cluster in minutes. In this session, participants deploy an HDInsight Spark cluster in the cloud. They then connect to it, use Jupyter notebooks to analyze a data set and apply structure to it, and incorporate the transformed data into a machine-learning model.

Azure Data Lake

Azure Data Lake allows researchers to aggregate of different shapes and sizes and from disparate data sources and process it in a platform-independent manner. Built on YARN, Azure Data Lake features seamless compatibility with Apache Hadoop, Apache Spark, and other popular big-data tools. In this module, participants migrate data from files and databases into a Data Lake Store, query it using U-SQL, and visualize the results with Power BI.

Azure High-Performance Computing (HPC)

With Azure, you can spin up high-performance computing clusters composed of virtual machines in minutes. In this module, students learn about HPC in Azure and deploy an HPC cluster that uses SLURM to distribute workloads between nodes. Then they use the cluster to perform the compute-intensive task of converting a batch of color images to grayscale.

Azure Batch and Azure Batch Shipyard

Azure Batch is a service that enables you to run batch processes on high-performance computing (HPC) clusters without incurring large capital expenditures to build out such clusters. Batch processes are ideal for handling compute-intensive tasks such as rendering videos and predicting the weather. Azure Batch Shipyard extends Azure Batch to perform batch processing in Docker containers. In this module, participants learn about Batch and Batch Shipyard and use them to convert text files to audio files.

Azure Data Science 501 - Advanced services and APIs

Microsoft Cognitive Services

Microsoft Cognitive Services is a set of cloud-based APIs for building intelligent applications. These APIs offer a range of capabilities, from recognizing faces in photos and videos to performing sentiment analysis on data generated by social media or left in feedback forums on Web sites. In this module, participants build a Web site for uploading photos and pass each uploaded photo to the Computer Vision API to generate captions and search metadata.

Azure Container Service

Containers are all the rage these days, and the most popular container platform in the world is Docker. The Azure Container Service (ACS) lets you host Docker apps in the cloud. It supports DC/OS and Docker Swarm for scaling to tens of thousands of containers, and it provides portability between cloud platforms. In this module, participants create a Docker image containing Python code that converts color photos to grayscale. Then they deploy the image to ACS and run the code to process a collection of photos.

Azure Functions

Azure Functions is a service that allows code written in a variety of languages, including C#, JavaScript, Python, and Bash, to run in the cloud in response to triggers such as timer ticks or blobs appearing in storage containers. It is the basis for writing microservices and nanoservices, and it is useful in research scenarios for decoupling actions from code executed in response to those actions. In this module, students learn about Azure Functions and write a function that, combined with Microsoft Cognitive Services, analyzes images uploaded to a blob container for adult content and segregates the images based on their content scores.

Microsoft Bot Framework

The Microsoft Bot Framework provides the tools developers need to build, connect, manage, and publish intelligent bots that interact naturally with users using a range of services. In this module, students create a bot using Visual Studio Code and the Microsoft Bot Framework. Then they interact with the bot using Skype — one of many popular services that bots built with the Microsoft Bot Framework can integrate with.