Azure Discovery Days 2019

Data Analytics & Near Real Time Intelligence with Azure

Scenario

In these labs, you will work through building a simple modern data estate that enables both analytic and near real-time data processing.

The scenario is as follows. You are working with an organization, Contoso, that came from a merger of two taxi companies (Yellow and Green). Contoso is planning to equip all their taxis with a device that sends information about each trip as the trip completes, and enables riders to submit an evaluation of the trip along with that upload.

The organization has years of historical trip data from the two taxi companies that merged. This data is broken down into multiple data sets with different structures. You will do data engineering work to ingest, clean, merge, and prepare this data for later analysis. You will store this data in a data mart and provide BI capabilities on top of the data.

You will then simulate a stream of incoming taxi trip data which includers rider feedback. You will enrich this stream of data in near real time with some analytics on the rider feedback, then send it on for both storage for later processing as well as to a real-time dashboard for trips where riders had a negative experience, so that Contoso can immediately respond to these riders.

By the end of the labs, you will have built a data analysis capability that includes both cold and hot paths, with historical and near real-time analytics running on the data.

Hands-On Lab Guide

In these hands-on labs, you will build a data estate modeled on a simple lambda architecture. This estate will include “cold” storage and analytics as well as a “hot” path of streaming data which will be operated on in near real-time. You will build both analytic as well as streaming dashboards.

The labs are detailed and prescriptive. Specific steps and technologies are described for you. If you are relatively new to Azure, we recommend you follow the steps carefully as documented. If you are more experienced with Azure, please feel free to adapt the labs and experiment on your own.

The labs use the Azure portal (https://portal.azure.com) as the primary tool. Some labs or sections may also show Azure Command-Line Interface (CLI) commands as an alternative to the graphical Azure portal. These can be executed in the Azure portal’s cloud shell, or locally after installing

the Azure CLI for your environment (see the Resources document). In the lab documents, the portal's cloud shell is used but the CLI commands are identical to those that should be executed in a local Azure CLI session.

IMPORTANT

The hands-on labs in this Discovery Day build on each other. Please work through the labs in numerical order. Lab 0 covers preparation and pre-requisites, and labs 1-4 are the actual lab modules.

In many of the lab docs, you will see screenshots showing how to do the various tasks. Many screenshots will show pre-filled values (such as host names, usernames, etc.). You do not need to imitate all the screenshots exactly: where appropriate, specify and use your own values for host names, usernames, deployment names, and so on.

Architecture

The overall architecture, broken down by lab, is shown below.