Lab 9 – Bringing it all Together

Technologies showcased: Azure SQL DW, SQL Server Management Studio

Author: Joshuha Owen

Date: 1/5/2018

Table of Contents

[Summary 3](#_Toc502930738)

[Pre-requisites 4](#_Toc502930739)

[Scenario 4](#_Toc502930740)

[Part 1 – Explore Data in the Azure SQL Data Warehouse 4](#_Toc502930741)

## Summary

In this Lab Module we will be exploring our data warehouse now that our Azure Data Factory has done a load of the data from our source systems (Azure Databases, Rest Data, and Blob Data merged via Hive). We want to verify we have data and also run a few queries to see how the data all relates.

## Pre-requisites

* Azure Subscription with rights to use/deploy Azure services, and X of Azure credit
* Azure SQL Data Warehouse, loaded with data from previous lab modules
* SQL Server Management Studio

## Scenario

|  |  |  |
| --- | --- | --- |
| Part 1 – Explore Data in the Azure SQL Data Warehouse | | |
| **Scenario** | | |
| In Lab Modules 6 though 8 we ran our Azure Data Factory pipeline to completely load our data warehouse. We’ll now connect SQL Server Management Studio to our database and do some verification. | | |
| **Commentary / Notes** | **Click Steps & ‘Bits’** | **Screenshots** |
|  | 1. Open SQL Server Management Studio. |  |
| Note the deployment script in Lab Module 1 creates the SQL Server name using the define prefix + SQL + a 4-character hash. To view your server name, visit your Resource Group created by the deployment script on <https://portal.azure.com>. | 1. You should be prompted to connect to a database server. If not, click the File menu -> Connect Object Explorer. 2. Fill in the server name and sqlusername and password created in Lab Module 1. |  |
|  | 1. Click the File Menu -> Open -> File. |  |
|  | 1. Browse to the Lab Module 9 folder from the downloaded Lab material and open the Module9query.sql file. |  |
|  | 1. You should see all the queries we will be running as part of this lab module. |  |
|  | 1. Highlight each query one at a time and run the query. Each step here will assume one query was run and discuss the purpose in the notes. To run use the Execute Icon or F5. |  |
| This query returns a count of all the flights in our FlightFact table. If you did not modify the package in Lab Module 2 this should contain flight data for December 2016. The AirlinePerformance-OLTP contains date for all of 2016 but the SSIS package only copied data for December 2016 to the ODS. | 1. Highlight and run Query 1 – Fact Count |  |
| This query returns the top 100 airports by originating airport on the flight. | 1. Highlight and run Query 2 – Busiest Origin Airports |  |
| This query returns the top 20 most used aircraft models by manufacturer. The DimAircraft dimension was our original set of FAA files from the Amazon S3 account that was later merged together via Hive. | 1. Highlight and run Query 3 – Most Used Planes |  |
| This query returns the Average delay for each day in December at MSP along with the weather data source from our REST source. Remember, this weather data was randomly generated. The query uses the origin airport as delays tend to affect outgoing flights more. | 1. Highlight and run Query 4 – Average Delay by Day in MSP and Weather |  |

**IMPORTANT: AVOID INCURRING EXTRA CHARGES BY PAUSING YOUR SUBSCRIPTION RESOURCES**