Building an Azure Business Intelligence Solution End to End Hands on Workshop

SQL Saturday – Manchester 2018

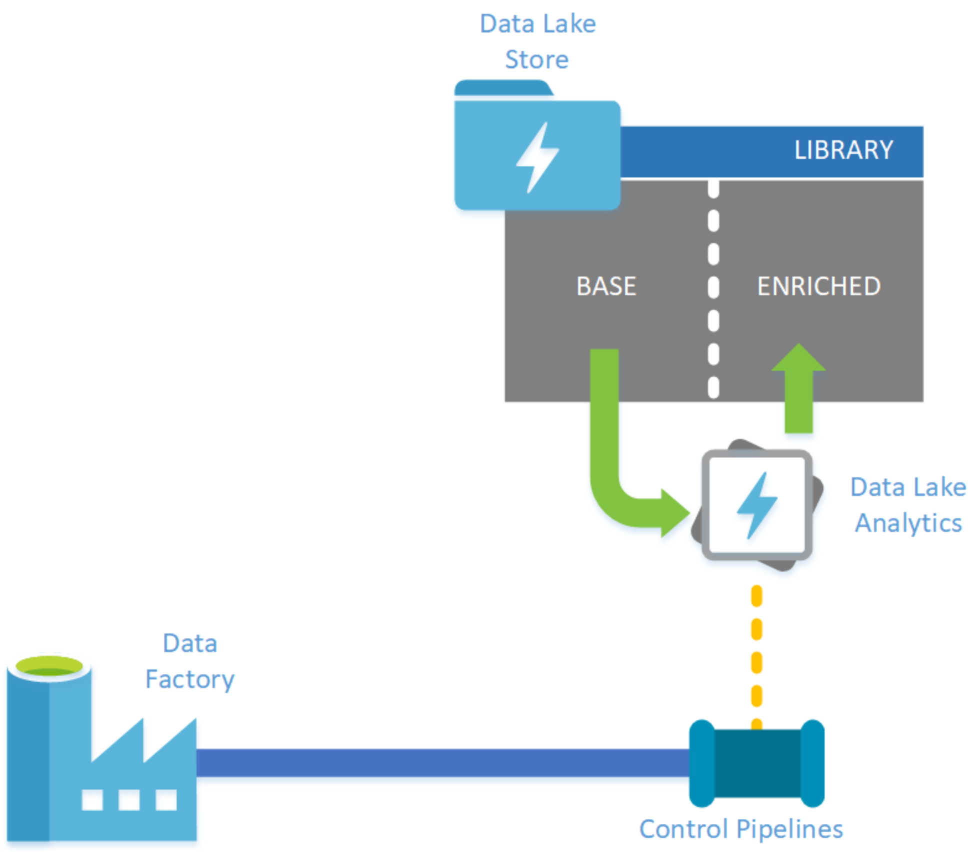
Paul Andrew Simon Whiteley Terry McCann

Lab Overview and Objective

Use a series of U-SQL scripts in Azure Data Lake Analytics to bring together and transform our RAW datasets both batch and streamed into Azure Data Lake Storage. Output a fact and dimension structure that can be visualised later.

Optional: Create stored procedures for the U-SQL scripts that can be called from your existing Azure Data Factory pipeline created in lab 1. This should be a downstream activity to the ingestion of the raw data copy activity.

This lab covers the below section of our overall architecture.



Steps

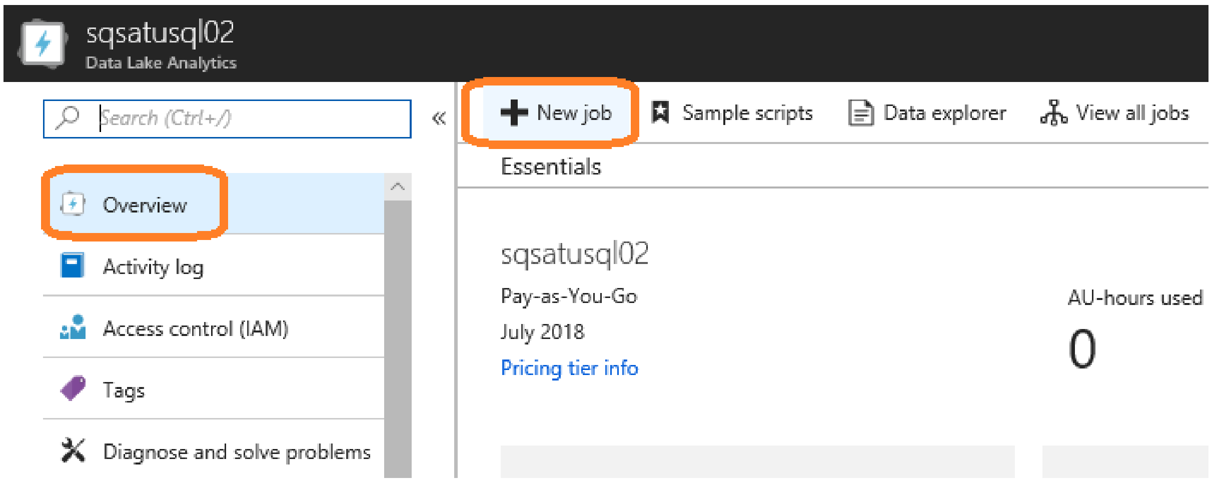
Narrative: U-SQL can be executed against Azure Data Lake Analytics using various tools and even debugged locally within a Visual Studio emulator. However, the simplest way is via the Azure portal using the browser blades.

The scripts executed will transform our RAW Adventure Works OLTP datasets into an OLAP structure and output them into our ENRICHED layer. In this architecture ENRICHED can be considered as our Data Warehouse layer.

1. Execute the provided U-SQL scripts against your Azure Data Lake Analytics service.

    a. Via the Azure Portal navigate to your Azure Data Lake Analytics service deployed in lab 1.

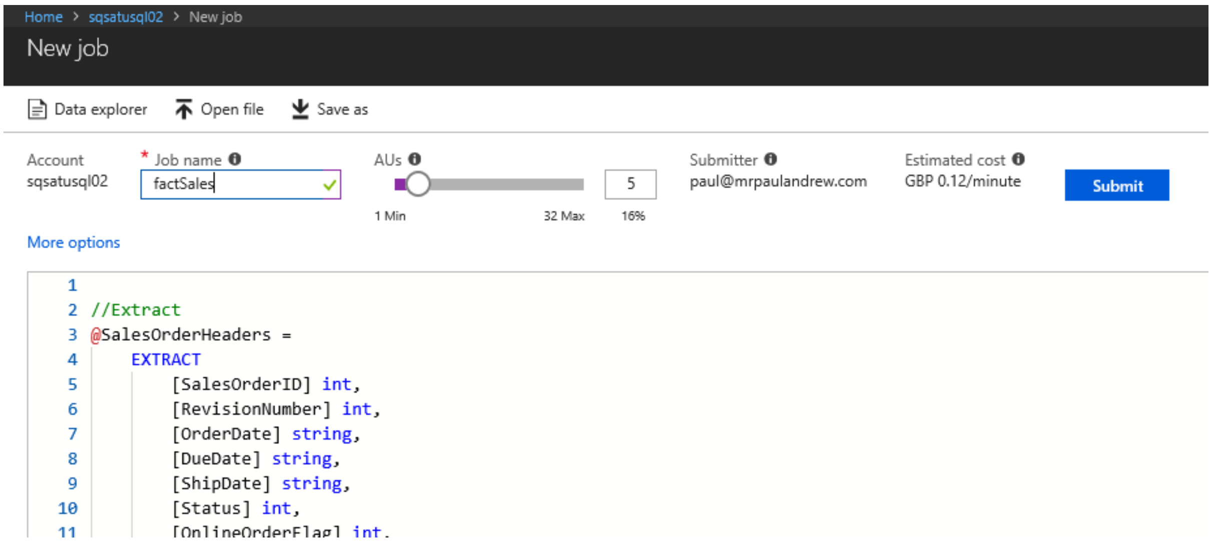
    b. In the presented blades go to **Overview** and click **+ New Job**.



    c. In the GitHub repository directory for this lab open the contents of the file **fact.Orders.usql** in any text editor.

    d. Copy the contents of the file into the Azure Portal new job canvas.

    e. Give the job a suitable name and click **Submit**.



Account sqsatusql02 · Job name factSales · AUs 5 (1 Min, 32 Max, 16%) · Submitter paul@mrpaulandrew.com · Estimated cost GBP 0.12/minute

```

//Extract
@SalesOrderHeaders =
    EXTRACT
        [SalesOrderID] int,
        [RevisionNumber] int,
        [OrderDate] string,
        [DueDate] string,
        [ShipDate] string,
        [Status] int,
        [OnlineOrderFlag] int,
```

f. Repeat steps b to e for the remaining U-SQL scripts.
   - dim.Address.usql
   - dim.Customers.usql
   - dim.Product.usql
   - fact.ProductSales.usql

g. After each data has completed you can view the Data Lake job vertex execution graph to verify the correct datasets have flowed through.

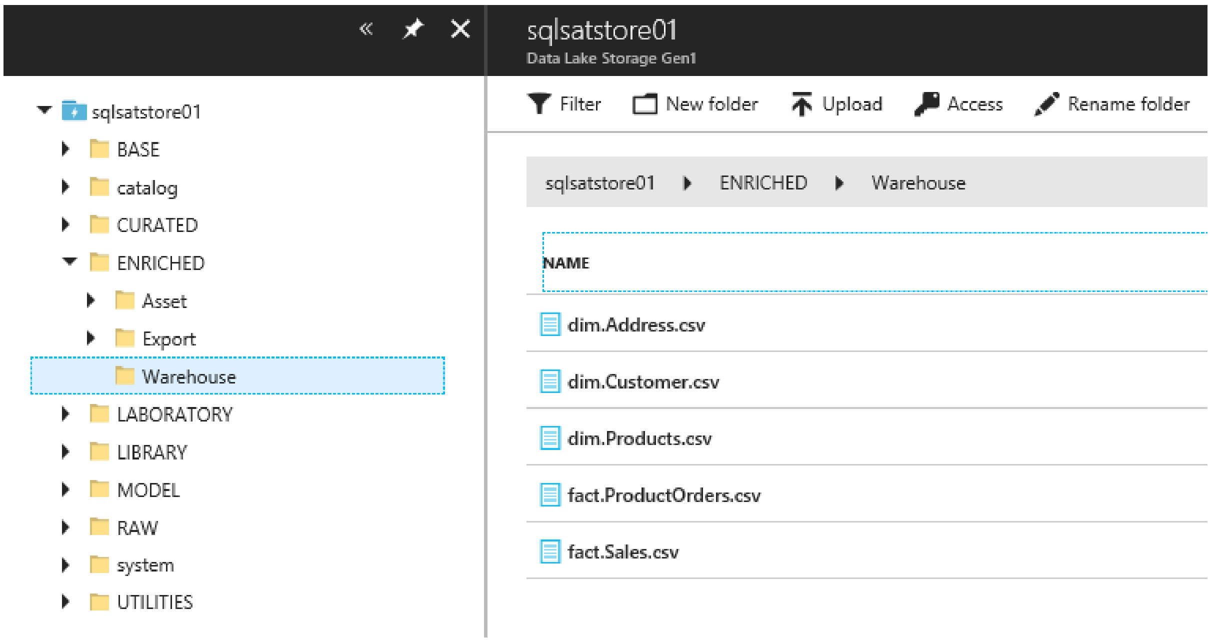
2. Verify the output files produced.

   a. Navigate to your Data Lake Storage deployed and setup in Lab 1.

   b. Go to **Data Explorer** under the **Data Lake Storage** section.

   c. Browse to your defined output directory: (ENRICHED\Warehouse).

   d. Confirm you have two fact datasets and



dim.Address.csv
dim.Customer.csv
dim.Products.csv
fact.ProductOrders.csv
fact.Sales.csv

That concludes the main part of lab 2.

To recap, we've executed a series of U-SQL scripts against our Azure Data Lake Analytics services to bring together and transform our RAW datasets, both batch loaded and streamed.

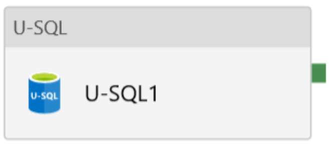
The output of the scripts has given us a Warehouse layer which can now be presented and visualised.

Optional Steps

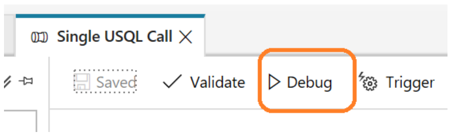
Narrative: In production we will want to automate the execution of our U-SQL code as part of our wider solution orchestration and control flow. In the next steps we'll do this using Azure Data Lake Analytics stored procedures with a U-SQL file called from Azure Data Lake Storage.

3. Create a U-SQL stored procedure to process all facts and all dimensions.
   a. Navigate to the Demo's directory in the GitHub repository. (\Demos\D2 - Data Lake and USQL\ETL\)
   b. Within the ETL solution you'll find U-SQL scripts to:
      i. Create an Azure Data Lake Analytics database.
      ii. Create fact and dim schema's.
      iii. A template to create each of the required stored procedures.
   c. Copy the contents of the U-SQL scripts already used to create the Enriched datasets and wrap them in the U-SQL DDL syntax.
   d. Deploy the stored procedures, via the Azure Portal or another tool.
4. Connect the U-SQL stored procedures to an Azure Data Factory activity.
   a. Create a U-SQL file with the following contents to call the stored procedures and place the file in a suitable location in your Azure Data Lake Storage.

```
USE AdventureWorksDW;
[fact].[CreateAllFacts]();
[dim].[CreateAllDims]();
```



   b. Within your Azure Data Factory pipeline add a U-SQL activity downstream to your Copy activity created.
   c. As part of setting up the activity a new linked service will be required to Azure Data Lake Analytics. This will also need the service principal credentials used for Azure Data Lake Storage in lab 1.
   d. Set the activity to use the U-SQL placed in your storage.
   e. Debug to pipeline to test the end to end process.


Single USQL Call
Saved
Validate
Debug
Trigger