AZURE SYNAPSE STUDIO DEVELOPMENT TOOLS

RUSS LOSKI

INTRODUCTION

- Russ Loski
- Data Engineer
- Husband, dad and grandad
- russloski@sqlmovers.com
- www.sqlmovers.com
- https://twitter.com/sqlmovers
- https://www.linkedin.com/in/russlosk
- Slides and Code: https://github.com/rloski-public/Presentations/tree/main/Baton%20Rouge%202022

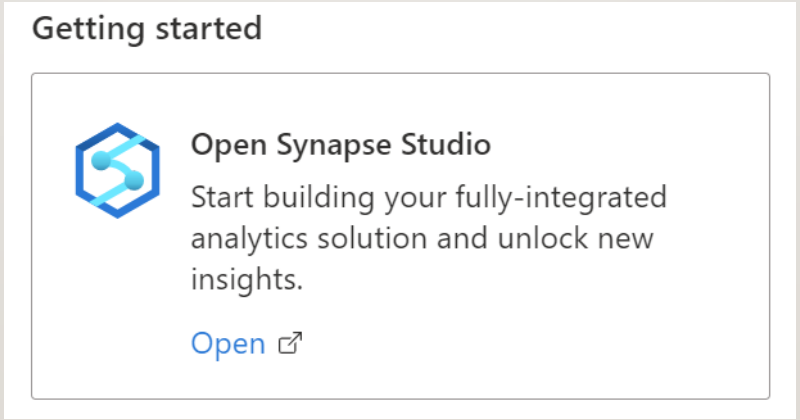
AGENDA

- Overview of interface
- Developing SQL
- Developing Spark

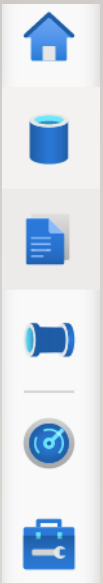
DATA IS THE NEW TREASURE

LAUNCHING AZURE SYNAPSE STUDIO

- https://web.azuresynapse.net/
- Alternatively you can log in to your Azure portal, find your Azure Synapse Analytics Workspace and click the Open Synapse Studio link

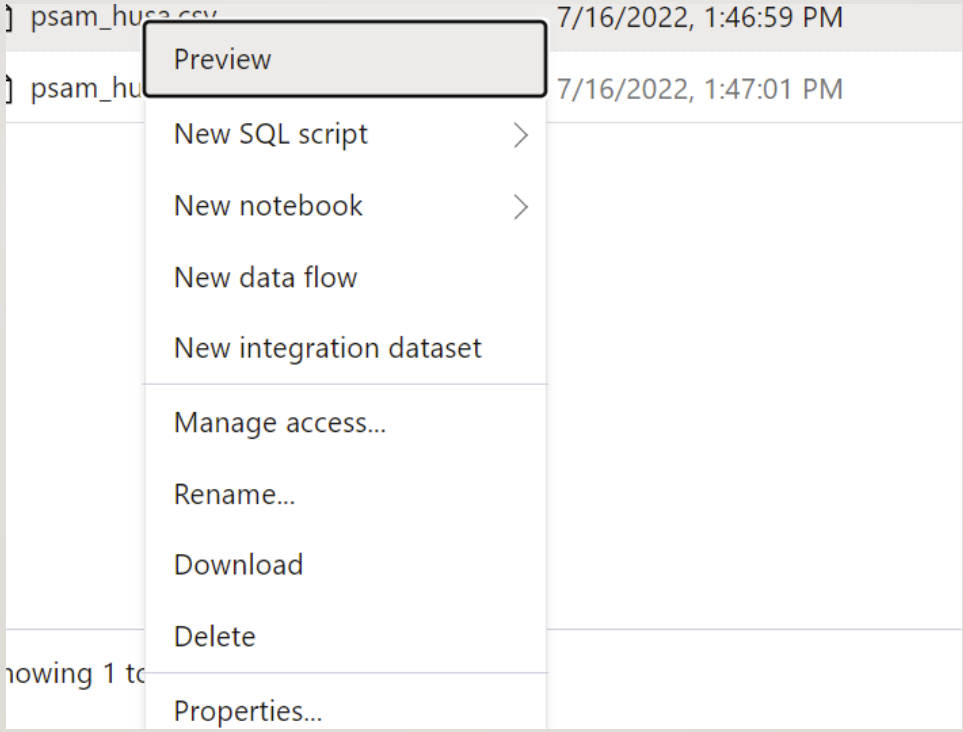


MAIN TABS



- Home
- Data
- Develop
- Integrate (Azure Data Factory)
- Monitor
- Manage

DATA – CSV FILE CONTEXT MENU



SQL SELECT FROM FILE

```
SELECT
    TOP 100 *
FROM
    OPENROWSET(
        BULK 'https://<account>.dfs.core.windows.net/<container>/census/PUMS/1-year/2017/Housing/psam_husa.csv',
        FORMAT = 'CSV',
   PARSER_VERSION = '2.0',
        HEADER_ROW = TRUE
    ) AS [result];
```

SQL DATA TYPES

```
SELECT
    *
 FROM
    OPENROWSET(
        BULK  'https://<account>.dfs.core.windows
.net/<container>/census/PUMS/1-
year/2017/Housing/psam_husa.csv',
        FORMAT = 'CSV',
   PARSER_VERSION = '2.0'
        , FIRSTROW = 2
    )
```

```
WITH (
        SerialNumber NVARCHAR(13) 2,
        DivisionCode NVARCHAR(1) 3,
        RegionCode NVARCHAR(1) 5,
        StateCode NVARCHAR(2) 6,
        NumberOfPersons int 10,
        NumberOfBedrooms int 16,
        NumberOfRooms int 38,
        FamilyIncome int 57
    ) AS [result];
```

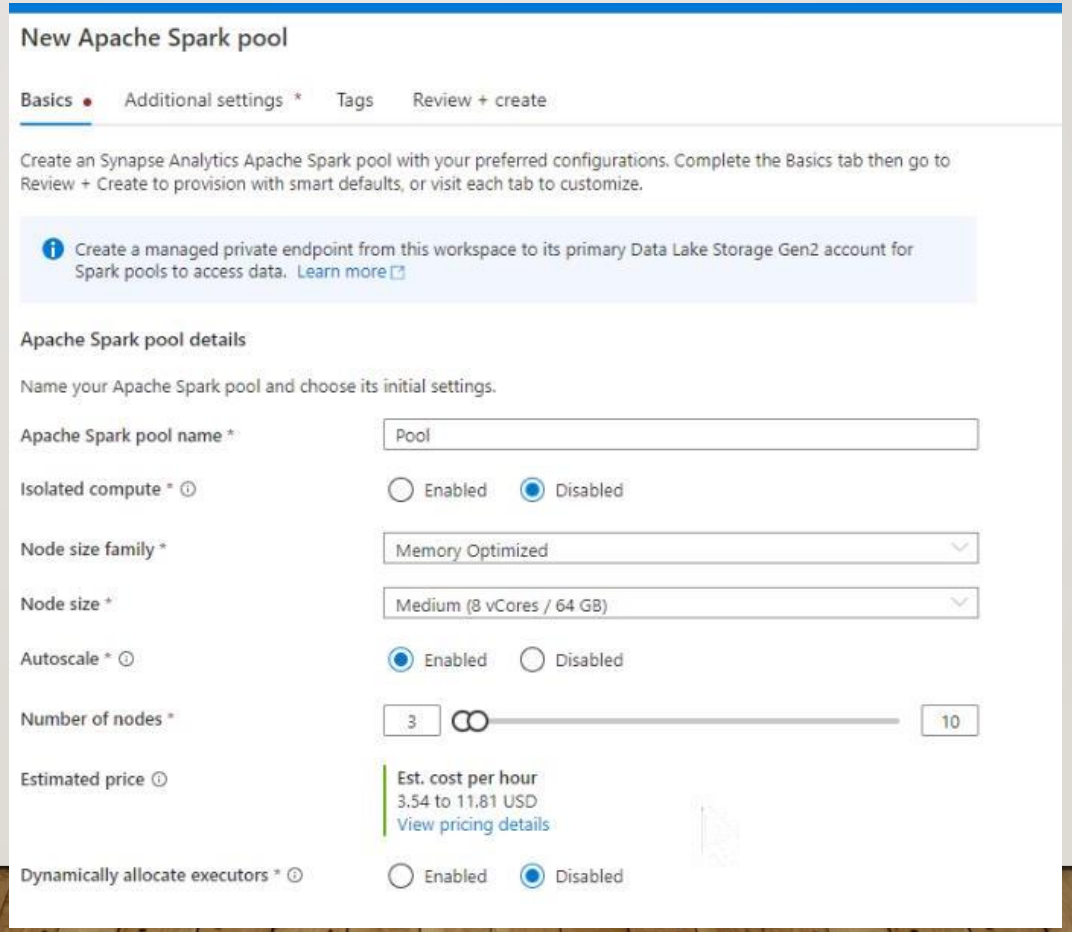
SERVERLESS SQL OBJECTS

- CREATE VIEW …
- CREATE EXTERNAL FILE FORMAT
- CREATE EXTERNAL DATA SOURCE
- CREATE EXTERNAL TABLE

SPARK

- Distributed system
- What is Apache Spark - Azure HDInsight | Microsoft Docs

CONFIGURE SPARK CLUSTER



DEVELOP - MARKDOWN

Demonstration

This is code to demonstrate how to work with the **Azure Synapse Studio** notebooks. You can get more help at Create, develop, and maintain Synapse notebooks in Azure Synapse Analytics.

We will examine the following:

1. Markdown cells Markdown for Jupyter notebooks cheatsheet
2. Working with the file magic commands

- `# Demonstration`
- `This is code to demonstrate how to work with the **Azure Synapse Studio** notebooks.`
- `You can get more help at [Create, develop, and maintain Synapse notebooks in Azure Synapse Analytics](https://docs.microsoft.com/en-us/azure/synapse-analytics/spark/apache-spark-development-using-notebooks).`
- `We will examine the following:`
- `1. Markdown cells [Markdown for Jupyter notebooks cheat sheet](https://www.ibm.com/docs/en/watson-studio-local/1.2.3?topic=notebooks-markdown-jupyter-cheatsheet)`
- `1. Working with the file magic commands`

DEVELOP – MAGIC

- %lsmagic
- %fs ls
- %fs head
- %%sql

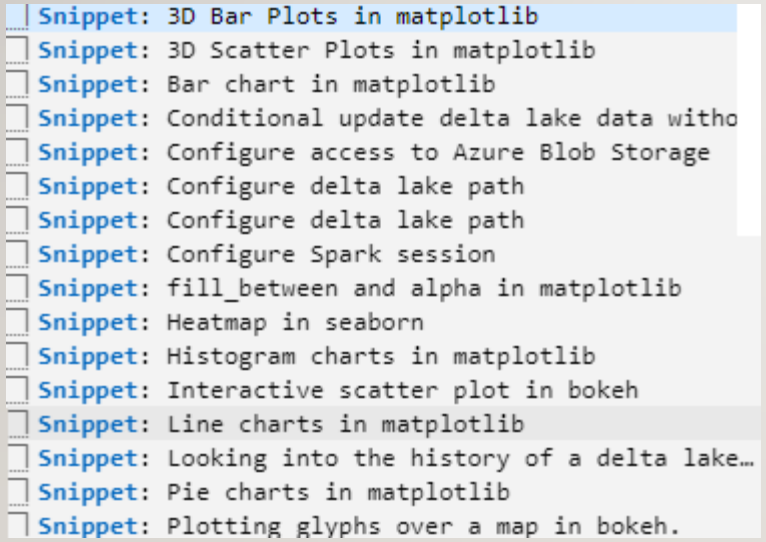
DEVELOP - MSSPARKUTIL

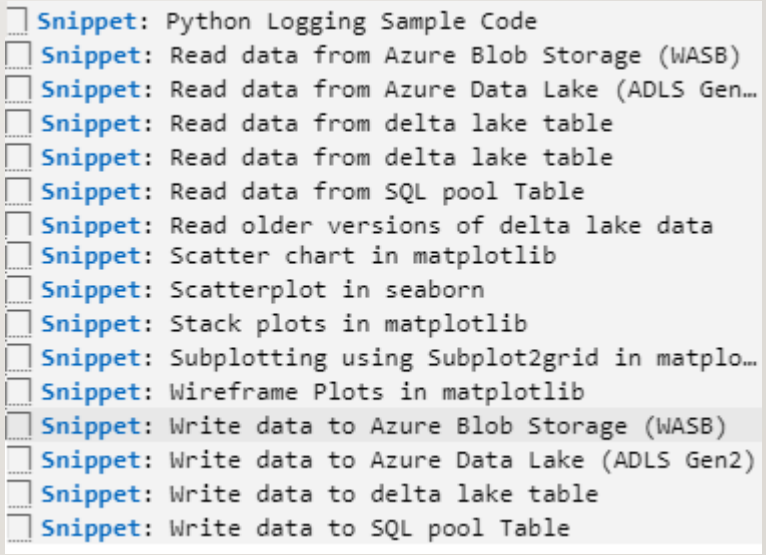
```
from notebookutils import mssparkutils

folder = "abfss://<container>@<
accountname>.dfs.core.windows.net/census/PUMS/1-year/2017/Housing/"

files = mssparkutils.fs.ls(folder)

for file in files:

    print(file.name, file.isDir, file.isFile, file.path, file.size)
```

DEVELOP – CONNECT TO FILE

```
df = spark.read.load('abfss://<containername>@<accountname>.dfs.core.windows.net/census/PUMS/1-year/2017/Housing/psam_husa.csv', format='csv'

## If header exists uncomment line below

, header=True

)

df.printSchema()
```

SNIPPETS


Snippet: 3D Bar Plots in matplotlib
Snippet: 3D Scatter Plots in matplotlib
Snippet: Bar chart in matplotlib
Snippet: Conditional update delta lake data witho
Snippet: Configure access to Azure Blob Storage
Snippet: Configure delta lake path
Snippet: Configure delta lake path
Snippet: Configure Spark session
Snippet: fill_between and alpha in matplotlib
Snippet: Heatmap in seaborn
Snippet: Histogram charts in matplotlib
Snippet: Interactive scatter plot in bokeh
Snippet: Line charts in matplotlib
Snippet: Looking into the history of a delta lake…
Snippet: Pie charts in matplotlib
Snippet: Plotting glyphs over a map in bokeh.


Snippet: Python Logging Sample Code
Snippet: Read data from Azure Blob Storage (WASB)
Snippet: Read data from Azure Data Lake (ADLS Gen…
Snippet: Read data from delta lake table
Snippet: Read data from delta lake table
Snippet: Read data from SQL pool Table
Snippet: Read older versions of delta lake data
Snippet: Scatter chart in matplotlib
Snippet: Scatterplot in seaborn
Snippet: Stack plots in matplotlib
Snippet: Subplotting using Subplot2grid in matplo…
Snippet: Wireframe Plots in matplotlib
Snippet: Write data to Azure Blob Storage (WASB)
Snippet: Write data to Azure Data Lake (ADLS Gen2)
Snippet: Write data to delta lake table
Snippet: Write data to SQL pool Table

RESOURCES

- How to use Synapse notebooks - Azure Synapse Analytics | Microsoft Docs
- SQL scripts in Synapse Studio - Azure Synapse Analytics | Microsoft Docs
- Markdown for Jupyter notebooks cheatsheet - IBM Documentation
- Index of /programs-surveys/acs/data/pums/2017/1-Year (census.gov)

CONTACT

- Russ Loski
- russloski@sqlmovers.com
- www.sqlmovers.com
- https://twitter.com/sqlmovers
- https://www.linkedin.com/in/russloski

- Slides and Code: https://github.com/rloski-public/Presentations/tree/main/Baton%20Rouge%202022