--Data Engineering Challenge by David Ribeiro Ferreira

Please follow the steps to setup the solution in Azure Cloud environment. Also refer to FLOWCHART attached to github.

1 - Create a storage account (Azure Data Lake) to store the CSV file

2 –Create an Azure Databricks cluster (data-engineering-challenge.ipynb)

3 – Create an Azure SQL Server (SqlServerProvisoningTemplate)

4 – Create an Azure Data Factory Resource as below:

In Azure Portal create a Data Factory resource:

Interface gráfica do usuário, Aplicativo

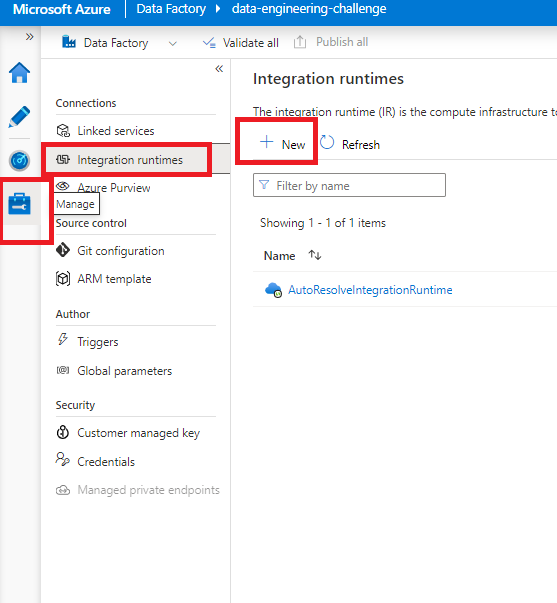
Descrição gerada automaticamente

Enter subscription information , create a new resource or use an existing one , inform name of for the resource, in our case is data-engineering-challenge

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Once the resource is created go to Manage – Integration Runtime and then NEW



Linha do tempo

Descrição gerada automaticamenteSelect Self-Hosted:

In network environment select Self-Hosted :

Interface gráfica do usuário, Texto, Aplicativo, Email

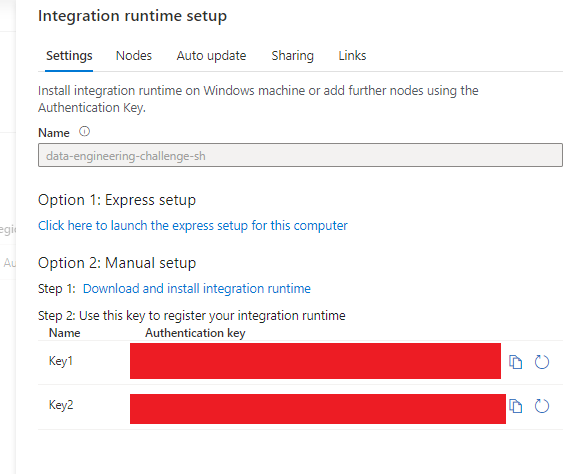
Descrição gerada automaticamente

Name and Create

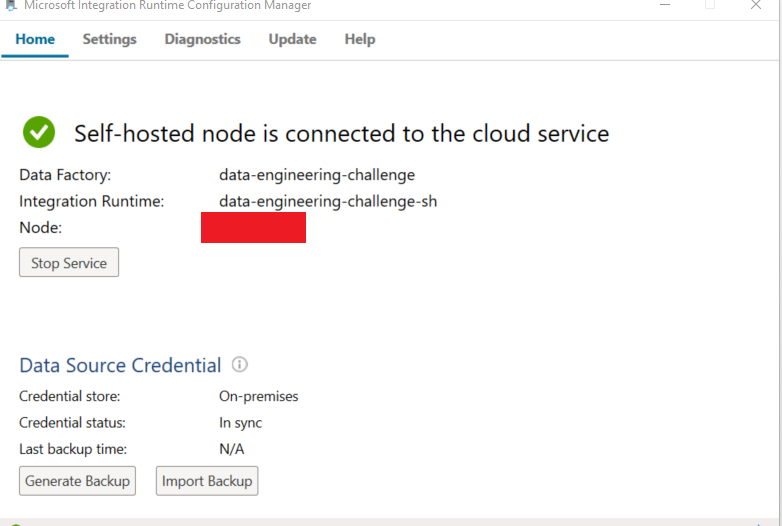
Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Click here to launch the express setup…



Download and configuration will start automatically



Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

Next step is to create a Linked server to read the csv file from localhost, in the file (git hub) linkek\_server.json it is reading the data from C:\data-eng-challenge":

1 - linked\_server.json

Script Pipeline:

2 – Pipeline.json

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

Script-Trigger

3 – Print Trigger:

Interface gráfica do usuário, Aplicativo, Email

Descrição gerada automaticamente