# User Manual

This manual contains a step by step guide to set up the project source code of jupyter notebook to perform analysis and price prediction.

## Setup of Microsoft Azure Portal:

- ➔ Step 1 (Azure account and Subscription) :
  - ◆ Create an Azure Account and subscribe to ‘Pay-As-You-Go’ subscription to use Azure HDInsight service as it is not included in Azure Free Account.
- ➔ Step 2 (Resource Group) :
  - ◆ Create a Resource Group to create storage accounts and HDinsight clusters.
  - ◆ This is the reference to create resource group - Manage resource groups - Azure portal
- ➔ Step 3 (Storage Account) :
  - ◆ Create a storage account as a data lake storage gen 2 in the resource group created in step 2.
  - ◆ This is the reference to create data lake storage gen 2 account - Create a storage account for Azure Data Lake Storage Gen2 | Microsoft Docs
  - ◆

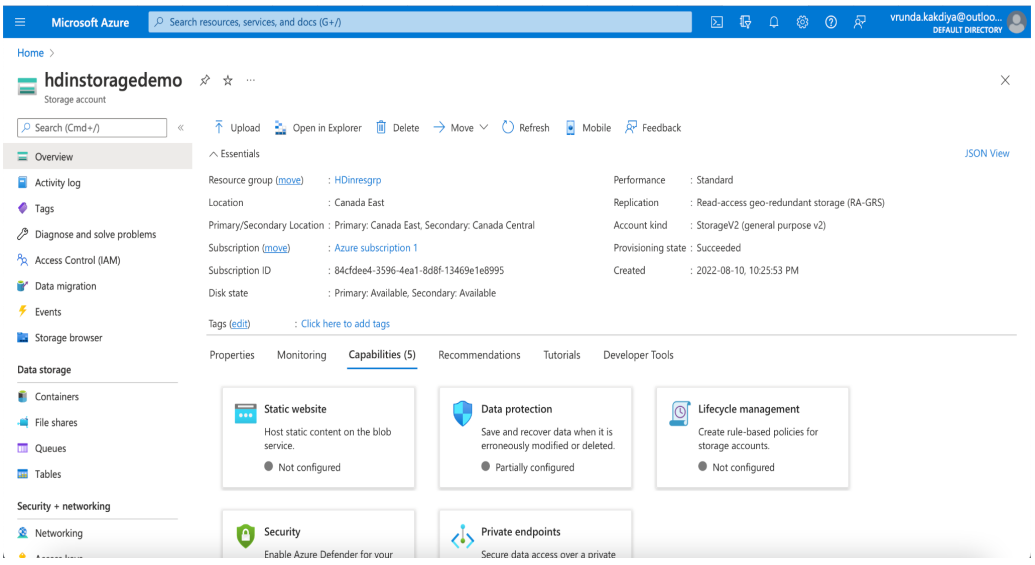


Fig. 1. Storage Account

- ➔ Step 4 (Upload the input file) :
  - ◆ Upload the input csv file in the blob container
- ➔ Step 5 (Create an Azure HDInsight cluster) :

- Create an Apache Spark 2.4 HDInsight cluster in a resource group. Select cores and nodes as per the requirement of the project.
- Reference for HDInsight creation part - [Quickstart: Create Spark cluster in HDInsight using Azure portal | Microsoft Docs]

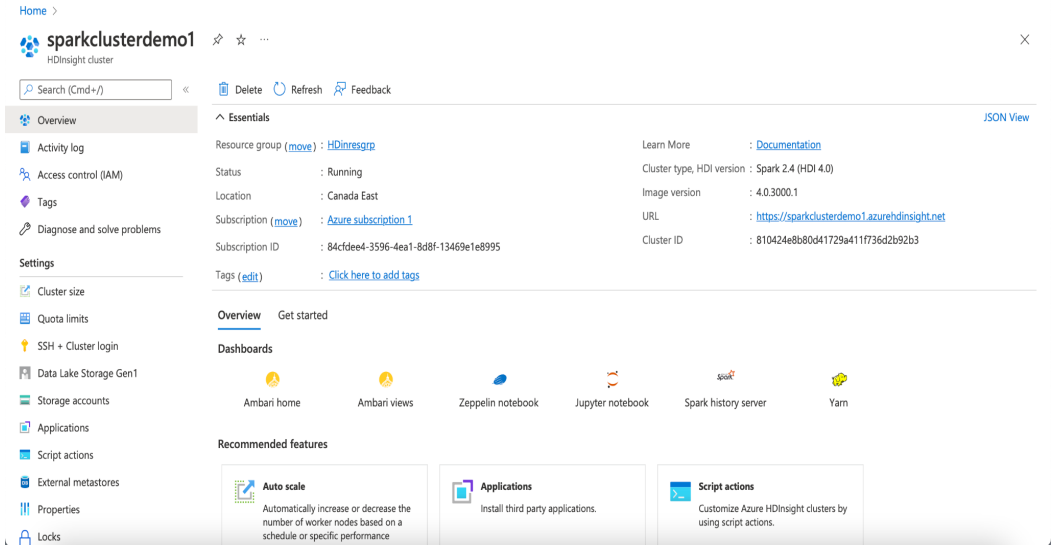


Fig 2. Azure HDInsight Cluster Creation

➔ Step 6 (Jupyter Notebook) :
- Open jupyter notebook and run the source code (car_price_pred.ipynb) attached with the document.

## Setup of Power BI:

➔ Install Power BI desktop application (To install please refer follow link : Downloads | Microsoft Power BI)

➔ Connect Power BI with Hdinsight spark cluster to load the cleaned data into power bi for the visualization (Reference : Power BI and Spark on Azure HDInsight; Step by Step Guide - RADACAD)

➔ Open pbix file attached along with this file to view the visualization of analysis.