**Executing Spark Code on Databricks Azure**

Prerequisites:

* Create an Azure account

1. Go to [www.azure.microsoft.com/en-us/free/](http://www.azure.microsoft.com/en-us/free/)

Graphical user interface, website

Description automatically generated

1. Click “Start free” and login with a Microsoft account (create one if you do not have one)
2. You will then have to create a free Azure account. You will have to enter your name, phone number, address, and credit card information. (You are signing up for the free tier, but they still require a card on file)

Graphical user interface, application, Teams

Description automatically generated

A screenshot of a computer

Description automatically generated

You now have a Microsoft Azure Account and have access to Azure Databricks

**Starting a Databricks Workspace**

1. Login to the Azure Portal [www.portal.azure.com/#home](http://www.portal.azure.com/#home)
2. Search and select the Azure Databricks service

Graphical user interface, text, application, email

Description automatically generated

1. Create a new workspace by clicking the ”Create” button

Graphical user interface, text, application, email

Description automatically generated

1. When creating your workspace, you must choose a resource group, workplace name, region, and pricing tier. You can create a new resource group if you do not have one created by clicking the “Create new” button. Then click “Review + create”

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

1. Review your workspace settings, then press “Create”

Graphical user interface, text, application, email

Description automatically generated

1. Once the workspace is created, head to that workspace and click “Launch Workspace”

Graphical user interface, text, application, email

Description automatically generated

1. You should now be on the Azure Databricks Workspace dashboard. Before we can work in a notebooks, we have to start up a cluster. Click “Create a cluster”. Alternatively, you can use the side bar and click on “Compute”. Then, click “Create Compute”

Graphical user interface, application

Description automatically generated

1. Set up your cluster. For development purposes, use the “Single Node” option. Then, click “Create Cluster”

Graphical user interface, application

Description automatically generated

1. Once the cluster is up and running, you can now go back to the dashboard and create a Notebook. Give it a name, default language, and a cluster for it to run on

Graphical user interface, application, email

Description automatically generated

1. You now have a notebook in which you can write spark code on. SparkSession and SparkContext are already initialized, along with a HiveMetastore. You may also upload data from the dashboard into the HiveMetastore which can then be queried using SparkSQL.

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, application

Description automatically generated