**Databricks Accelerator Run Book**

**Prerequisites:**

Azure resource group created with appropriate access

**Step 1: Create Databricks Cluster**

* Go to portal.azure.com and search for “databricks” in search box, select “Azure Databricks”

Graphical user interface, text, application

Description automatically generated

* Click on “Create” from top left corner

Graphical user interface, text, application, email

Description automatically generated

* Select the appropriate resource group, mention workspace name, select region, keep the pricing tier as standard(default). Click on “Review and create”
  + Resource group: as provided by the organization
  + Workspace name: could be any relevant name
  + Pricing tier: select based on the requirements

Graphical user interface, text, application

Description automatically generated

* After that simply click on “Create” button

Graphical user interface, text, application, email

Description automatically generated

* After the cluster is created click on “Launch Workspace”, this will launch the workspace

A picture containing text

Description automatically generated

**Step 2: Create ADLS Storage**

* Go to portal.azure.com and search for “storage” in search box, select “Storage accounts”

Graphical user interface, text, application, email

Description automatically generated

* Click on “Create” from top left corner to create a storage account

Graphical user interface, text, application, chat or text message

Description automatically generated

* Select the subscription, resource group, mention the storage account name and click on “Review and create”

Graphical user interface, text, application, email

Description automatically generated

* Once the storage account is created go to “Data storage” section on the left side and click on “Containers”

Graphical user interface, application

Description automatically generated

* Click on “Container”, then enter a “Name”, and click on “Create”

Graphical user interface, application

Description automatically generated

* After the container is created, try uploading some files using the “Upload” button

Graphical user interface, text, application, email

Description automatically generated

**Step 3: Register Azure active Directory Application**

* Go to portal.azure.com and search for “Azure active” and select “Azure Active Directory”

Graphical user interface, text, application

Description automatically generated

* Under “Manage” tab on the left, click on “App registration”

Graphical user interface, application

Description automatically generated

* Click on “New registration” to register the application

Graphical user interface, text, application, Word, chat or text message

Description automatically generated

* Add a “name” and click on “Register”

Graphical user interface, text, application, email

Description automatically generated

* After registering the application, go to “Manage” section on the left and select “Certificates and secrets”

Graphical user interface, text, application, chat or text message

Description automatically generated

* Click on “New client secret”, add a “Description” and click on “Add”

(Remember to copy the secret and keep it somewhere as you won’t be able to retrieve it later)

Graphical user interface, application

Description automatically generated

* Click on “overview” from the left panel, and copy the “Application ID” and “Directory ID”

Graphical user interface, text, application, email

Description automatically generated

**Step 4: Assign the Storage Blob Data Contributor Role to the ADLS Gen2 storage account**

* In azure portal go to the storage account that we created in step 2, and click on “Access Control (IAM)” from the left panel

Graphical user interface, text, application

Description automatically generated

* Click on “Add” and select “Add role assignment”

Graphical user interface, application

Description automatically generated

* Search for “storage blob”, select “Storage Blob Data Contributor” and click on “Next”

Graphical user interface, text, application, email

Description automatically generated

* Click on “Select members”, type in the azure active directory application that we created in step 3.1

A screenshot of a computer

Description automatically generated

* Click on “Review and assign”

Graphical user interface, text, application, email

Description automatically generated

**Step 5: Mount the ADLS to Databricks cluster**

* Go to databricks cluster, go to “Workspace” from left panel, select “Users” and then click on your username

Graphical user interface, text, application

Description automatically generated

* Right click under your name and then select “Create” and then “Notebook”

Graphical user interface, text, application, chat or text message

Description automatically generated

* Paste the below code in the notebook cell

//

configs =

{"fs.azure.account.auth.type": "OAuth","fs.azure.account.oauth.provider.type": "org.apache.hadoop.fs.azurebfs.oauth2.ClientCredsTokenProvider",

"fs.azure.account.oauth2.client.id": "<application-id>",

"fs.azure.account.oauth2.client.secret": “<secret>"),

"fs.azure.account.oauth2.client.endpoint": "https://login.microsoftonline.com/<directory-id>/oauth2/token"}

//

application-id: application ID copied in previous steps

directory-id: directory ID copied in previous steps

secret: secret copied in previous steps

Text, application

Description automatically generated with medium confidence

* Add the container name, storage account name and the mount name that we created in previous steps

dbutils.fs.mount(source = "abfss://<container-name>@<storage-account-name>.dfs.core.windows.net/", mount\_point = "/mnt/<mount-name>", extra\_configs = configs)

Graphical user interface, text, application

Description automatically generated

* Run the below command to see the files stored in the adls location

Graphical user interface, application, table

Description automatically generated

ADLS is successfully mounted to the Azure Databricks

**Step 6: Link Azure devops with Databricks**

* Go to portal.azure.com and search for “azure devops” and click on “Azure DevOps organizations”

Graphical user interface, text, application, email

Description automatically generated

* Click on “My Azure DevOps Organizations”

Graphical user interface, text, application, email

Description automatically generated

* Create new organization and fill in the details and click on “Continue”

Graphical user interface, application

Description automatically generated

* Click on new project and create a new project by adding the name and click on “Create”

Graphical user interface, application, email

Description automatically generated

* Click on the “Repos” from the left panel and then click on “Initialize”

Graphical user interface, text, application, email

Description automatically generated

* Click on clone and then copy the URL

Graphical user interface, text

Description automatically generated

* Go to azure databricks workbook and click on “Revision history” from top right of the notebook, and then click on Git: Not linked

Graphical user interface, text, application, chat or text message

Description automatically generated

* Paste the link copied in previous steps in the “Link”, keep the branch as “main”, “Path in git Repo” will be automatically picked by the databricks, click on save

Graphical user interface, text, application

Description automatically generated

* Now go to “Repos” from the left panel, and you will see the repository name, click on it and you will see the files inside the Azure repository

Graphical user interface, text, application

Description automatically generated

Azure Repos is linked to Azure Databricks

**Step 7: Create build (CI) pipeline in Azure Devops**

* Go to azure devops and then go to “Pipelines” from the left panel and click on “Pipelines”, click on “Create Pipeline”

Graphical user interface

Description automatically generated with medium confidence

* Click on “Use the classic editor” at the end

Graphical user interface, text, application, email

Description automatically generated

* Select the repository and branch and click on “Continue”

Graphical user interface, text, application, email

Description automatically generated

* Click on “Empty job”

Graphical user interface, text, application

Description automatically generated

* Click on + icon to add task, search for publish build and select “Publish Build Artifacts” and click on add

Graphical user interface, text, application, email

Description automatically generated

* Click on the task that you have added, and then fill the details, please do mention the “Path to publish”

Graphical user interface, application

Description automatically generated

* Go to the “Triggers” tab from the top and then enable the continuous integration to trigger the pipeline on its own

Graphical user interface, application

Description automatically generated

* Click on “Save & queue” and then click on “Save”

Graphical user interface, application, Word

Description automatically generated

**Step 8: Create release (CD) pipeline in Azure Devops**

* Go to releases under pipelines from left panel and click on “New Pipeline”

Graphical user interface, text, application

Description automatically generated

* In select template option, go with “Empty job”

Graphical user interface, text, application

Description automatically generated

* Go to Tasks from top menu and add task in agent, search for databricks and then select “Databricks Deploy Notebook”
* Note: (Very Important): We first must install "Databricks Script Deployment Task by Data Thirst", then the displayed Databricks tasks will become available. This package is provided by 3rd party. Just click on - "Databricks Script Deployment Task by Data Thirst" & follow along, it will get installed.

Graphical user interface, application

Description automatically generated

* Click on the task added and then fill in the details:
  + Source files path: the file path to be sourced
  + Target files path: path where the notebooks will be pushed (/Shared)
  + Authentication Method: to get the bearer token go to databricks, Navigate to User Icon --> User Settings --> Access Tokens tab --> Generate New Token

Graphical user interface, text, application

Description automatically generated

* Navigate to the Pipeline tab. Select the Continuous deployment trigger on the artifact, then enable the continuous deployment trigger

Graphical user interface, application

Description automatically generated

**Step 9: Push and verify the changes from databricks to Azure devops**

* Go to azure databricks, under repos create a notebook that need to be pushed to azure devops, I have just added a comment for now

Graphical user interface, text, application

Description automatically generated

* Click on the master branch button prior to the notebook name, it will detect the latest changes, add the summary and click on “Commit and Push”

Graphical user interface, text, application, email

Description automatically generated

* Once pushed, go to azure devops and click on “Repos” from the left panel, you will see a message “You updated master just now”, and a button to create a pull request. Click on “Create a pull request”

Graphical user interface, text, application, chat or text message

Description automatically generated

* Approve the pull request first and then click on Complete

Graphical user interface

Description automatically generated with low confidence

* Go to pipeline and see the triggered build once the pull request is merged with main branch

Text

Description automatically generated

* After the build pipeline is complete, got to Release under Pipelines section and find the release pipeline

Graphical user interface, text, application, chat or text message, email

Description automatically generated

* Finally verify the notebooks in the output files location as mentioned while creating the release pipeline (in our case path: /Shared)

Graphical user interface, application

Description automatically generated

**References:**

<https://medium.com/geekculture/ci-cd-on-azure-databricks-using-azure-devops-82441386c6e8>

<https://levelup.gitconnected.com/continuous-integration-and-delivery-in-azure-databricks-1ba56da3db45>

<https://www.linkedin.com/pulse/ci-cd-azure-databricks-using-devops-deepak-rajak>

<https://medium.com/@bedatse/azure-devops-ci-cd-with-azure-databricks-and-data-factory-part-1-c05a44536a8e>