# Azure Databricks Setup

Contents

[Azure Databricks Setup 1](#_Toc173227884)

[Create Azure Databricks Resource 2](#_Toc173227885)

[Set up access to storage account 6](#_Toc173227886)

[Create secret scope in ADB 6](#_Toc173227887)

[Create a key with the storage account access key 9](#_Toc173227888)

[Create a secret in your key vault to store the storage account key. 11](#_Toc173227889)

[Runtime 11](#_Toc173227890)

This walkthrough provides the steps required to set up a Databricks workspace in Azure, create a compute capability in that workspace, and then connect the workspace to the storage account created in assignment 1.

## Create Azure Databricks Resource

Marketplace 
Get Sta rted 
Service Providers 
Management 
Private Marketplace 
Private Offer Management 
My Marketplace 
Favorites 
Recently created 
Private products 
Categories 
Analytics (36) 
p Azure Data bricks 
Showing 1 to 20 of 44 results for Azure 
Azure Databricks 
Microsoft 
Azure Service 
Azure Databricks is the fast, easy and 
collaborative Apache Spark-based 
analytics platform. 
Create v 

Home > MarchMadness > Marketplace > Azure Databricks > 
Create an Azure Databricks workspace 
Basics Networking 
Project Details 
Advanced 
Tags 
Review + create 
Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and 
manage all your resources. 
Subscription * O 
Resource group 
Instance Details 
Workspace name * 
Region * 
Pricing Tier* O 
david.froslie Student Subscription 
MarchMadness 
Create new 
MarchMadnessADB 
East US 
Standard (Apache Spark, Secure with Azure AD) 

Create an Azure Databricks workspace 
Basics Networking Advanced 
Tags 
Deploy Azure Databricks workspace with Secure 
Cluster Connectivity (No Public IP) O 
Deploy Azure Databricks workspace in your own 
Virtual Network (VNet) 
Review + create 
C) Yes (S) No 
C) Yes @ No 

Create an Azure Databricks workspace 
e Validation Succeeded 
Basics Networking 
Summary 
Basics 
Workspace name 
Subscription 
Resource group 
Region 
Pricing Tier 
N etworking 
Advanced 
Tags 
Review + create 
MarchMadnessADB 
david.froslie Student Subscription 
MarchMadness 
East US 
standard 
Deploy Azure Databricks workspace with No 
Secure Cluster Connectivity (No Public IP) 
Deploy Azure Databricks workspace in 
your own Virtual Network (VNet) 
Advanced 
Enable Infrastructure Encryption 
No 
No 
Create 
< Previous 
Download a template for automation 

### Set up minimalistic compute (0.75 dbu / hour, or $0.30 / hour)

 Navigate to the Azure resource that you just created. Databricks operates in a workspace set that you go to be clicking on Launch Workspace.

Home > 
MarchMadnessADB 
Azure Databricks Service 
Overview 
ActMty log 
Access cmtrol (IAM) 
Tags 
Diagnose and solve problems 
Settings 
Virtual Network Peerings 
Encryption 
e' Networking 
Properties 
A Locks 
Tasks (preview) 
Export template 
Status 
MatchMadness 
Resource group . 
Location 
: East US 
Subscription 
Subscription ID 
: 6<24bfa-68d7-4316-a88d-OdOdabdd2f1d 
Tags (edit) 
• Add_tags 
x 
JSON View 
Managed Resource Group : 
: htt 
URL 
Pricing 
Launch Workspace 
Upgrade to Premium 

After logging in, navigate to the Compute page and click Create Compute. Create a Single node "cluster".

Microsoft Azure 
New 
Workspace 
C) Recents 
00 Data 
Workflows 
Compute 
SQL > 
Data Engineering 
Job Runs 
Data Ingestion 
Delta Live Tables 
Machine Learning 
Experiments 
Q Search data, notebooks, recents, and more... 
CTRL + p 
Compute 
All-purpose compute 
Job compute 
Q Filter compute you have access to 
pools 
Created by 
State 
Name 
Runtime 
Active mem... 
Active cores 
Active DBU 
Source 
MarchMadnessADB 
Creator 
david.froslie@ndus.edu v 
Create compute 
Notebooks 
No compute 
Create compute to run workloads from your notebooks and jobs. Leam more about best practices for 
compute configuration 
Create compute 

Compute New compute > UI preview 
Froslie, David's Cluster 
o 
Multi node Single node 
Send feedback 
Access mode O 
Single user 
v 
Performance 
Single user access O 
Froslie, David 
Summary 
1 Driver 
Runtime 
Photon 
UI I USQN 
14 GB Memory, 4 Cores 
12.2 x-scala2.12 
standard DS3 v2 
1.5 D8U/h 
Databricks runtime version O 
Runtime: 12.2 ITS (Scala 2.12, spark 3.3.2) 
IJse Photon Acceleration O 
Node type O 
e Terminate after 
Tags O 
Add tags 
Key 
30 
14 GB Memory. 4 Cores 
minutes of inactivity O 
> Automatically added tags 
Advanced options 
Create compute 
Cancel 

## Set up access to storage account

Use the key vault and storage account created in assignment 1 to connect the Databricks instance to the storage account. This is done in multiple steps:

1. Create a "secret scope" in Databricks that points to the key vault.
2. Create a secret in the key vault that has the access key to the storage account.
3. At run-time, specify the secret scope and the secret in the Spark configuration. This will enable access to the storage account.

### Create secret scope in ADB

 Follow the steps in to create the secret scope in Databricks: <https://learn.microsoft.com/en-us/azure/databricks/security/secrets/secret-scopes#--create-an-azure-key-vault-backed-secret-scope>

Home > 
MarchMadnessADB 
Azure Databricks Service 
Overview 
Activity log 
Access control (IA M) 
Tags 
Settings 
Virtual Network Peerinas 
Delete 
A Essentials 
Status 
Resource group 
Location 
Subscription 
Subscription ID 
: Active 
MarchMadness 
: East US 
: day.id:fæsl.ie-Student-Suhscip.tiQn 
: 6c124bfa-68d7-4316-a88d-OdOdabdd2f1d 
Managed Resource Group 
URL 
Pricing Tier 
JSON View 
. databricks-rg-MarchMadnessADB-yEq72fasrcz4 
• bttps://adb-1955915938806155.15.azuredatabric... 
. standard 

Substitute the URL shown in previous screenshot for <databricks-instance> in the following to navigate to the Create Secret Scope page: [https://<databricks-instance>#secrets/createScope](#secrets/createScope)

Fill in the values similar to what is shown below. The Azure Key Vault DNS and Resource ID can be found on the Key Vault Properties page as shown in the next screen shot (note - use Vaul URI for the DNS Name).

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

### Create a key with the storage account access key

Retrieve the Key from key 1 under Access keys for the storage account or create a Shared access signature (SAS) key and copy the resulting key.

Home > MarchMadness > marchmadstore 
marchmadstore I Access keys 
Storage account 
Search 
Overview 
Activity log 
Tags 
Diagnose and solve problems 
Access Control (IAM) 
Data migration 
Events 
Storage browser 
Data storage 
Containers 
File shares 
Queues 
Tables 
Security + networking 
Networking 
Access keys 
Shared access signature 
Encryption 
C) Set rotation reminder CD Refresh 
Access keys authenticate your applications' requests to this storage account. Keep your keys in a secure location like Azure 
Key Vault, and replace them often with new keys. The two keys allow you to replace one while still using the other. 
Remember to update the keys with any Azure resources and apps that use this storage account. 
Learn more about managing storage account access keys 
Storage account name 
marchmadstore 
keyl Rotate key 
Last rotated: 10/9/2022 (O days ago) 
Key 
Connection string 
key2 Rotate key 
Last rotated: 10/9/2022 (O days ago) 
Key 
Connection string 
Show 
Show 
Show 
Show 

### Create a secret in your key vault to store the storage account key.

Home > MarchMadnessKV I Secrets > 
Create a secret 
Upload options 
Name* G) 
Secret value* O 
Content type (optional) 
Set activation date (D 
Set expiration date O 
Enabled 
Tags 
Manual 
marchmadstore-key 
Yes 
O tags 
No 

## Runtime

Use the secret in the API call to connect the storage for ADB.

A screenshot of a computer

Description automatically generated

Extract the storage end point from the storage account in the Azure portal as shown above. Use the scope and key that you set up earlier in this process.

storage\_end\_point = "assign1storage.dfs.core.windows.net"

my\_scope = "MarchMadnessScope"

my\_key = "assign1-key"

spark.conf.set(

    "fs.azure.account.key." + storage\_end\_point,

    dbutils.secrets.get(scope=my\_scope, key=my\_key))

# Replace the container name (assign-1-blob) and storage account name (assign1storage) in the uri.

uri = "abfss://assign-1-blob@assign1storage.dfs.core.windows.net/"