

Set up Databricks

Monday, September 4, 2023 3:02 PM

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 Started

Service Providers

Management

My Marketplace

Categories


Analytics (36)

Azure Databricks

Microsoft

Azure Service

Azure Databricks is the fast, easy and collaborative Apache Spark-based analytics platform.

Create 

[Home](#) > [MarchMadness](#) > [Marketplace](#) > [Azure Databricks](#) >

Create an Azure Databricks workspace

Basics

Networking

Advanced

Tags

Review + create

Project Details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription *

david.frosolie Student Subscription

Resource group *

MarchMadness

[Create new](#)

Instance Details

Workspace name *

MarchMadnessADB

Region *

East US

Pricing Tier *

Standard (Apache Spark, Secure with Azure AD)

Create an Azure Databricks workspace

Basics

Networking

Advanced

Tags

Review + create

Deploy Azure Databricks workspace with Secure Cluster Connectivity (No Public IP) ⓘ

☐ Yes

☒ No

Deploy Azure Databricks workspace in your own Virtual Network (VNet)

☐ Yes

☒ No

Create an Azure Databricks workspace ...

✓ Validation Succeeded

Basics Networking Advanced Tags **Review + create**

Summary

Basics

Workspace name	MarchMadnessADB
Subscription	david.froslie Student Subscription
Resource group	MarchMadness
Region	East US
Pricing Tier	standard

Networking

Deploy Azure Databricks workspace with Secure Cluster Connectivity (No Public IP)	No
Deploy Azure Databricks workspace in your own Virtual Network (VNet)	No

Advanced

Enable Infrastructure Encryption	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

Search < Delete

Overview < Essentials JSON View

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Settings

Virtual Network Peerings

Encryption

Networking

Properties

Locks

Automation

Tasks (preview)

Export template

Status : Active

Managed Resource Group : databricks-rg-MarchMadnessADB-y3g7Zlaerc4

Resource group : MarchMadness

Location : East US

Subscription : david.froslie Student Subscription

Subscription ID : 6c124bfa-68d7-4316-a88d-0d0dabdd2f1d

URL : https://adb-1955915938806155.15.azuredatabricks.net

Pricing Tier : Standard (Apache Spark, Secure with Azure AD) (Click to change)

Tags (edit) : Add tags

Launch Workspace

Upgrade to Premium

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

Microsoft Azure | databricks Search data, notebooks, recents, and more... CTRL + P MarchMadnessADB david.froslie@ndus.edu

New

Workspace

Recents

Data

Workflows

Compute

SQL >

Data Engineering

Job Runs

Data Ingestion

Delta Live Tables

Machine Learning

Experiments

Compute

All-purpose compute Job compute Pools

Filter compute you have access to Created by

Create compute

State Name Runtime Active mem... Active cores Active DBU ... Source Creator Notebooks

No compute

Create compute to run workloads from your notebooks and jobs. Learn more about best practices for compute configuration

Create compute

Froslie, David's Cluster [✎](#)

☐ Multi node ☒ Single node

Access mode [?](#)

Single user access [?](#)

Single user [▼](#)

Froslie, David [▼](#)

Performance

Databricks runtime version [?](#)

Runtime: 12.2 LTS (Scala 2.12, Spark 3.3.2) [▼](#)

☒ Use Photon Acceleration [?](#)

Node type [?](#)

Standard_DS3_v2 14 GB Memory, 4 Cores [▼](#) [?](#)

☒ Terminate after 30 minutes of inactivity [?](#)

Tags [?](#)

Add tags

Key Value Add

> Automatically added tags

▶ Advanced options

Summary

1 Driver 14 GB Memory, 4 Cores

Runtime 12.2.x-scala2.12

Photon Standard_DS3_v2 1.5 DBU/h

UI | [JSON](#)

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



Search

Delete

Overview

Activity log

Access control (IAM)

Tags

Settings

Virtual Network Peerings

Essentials

Status : Active

Resource group : [MarchMadness](#)

Location : East US

Subscription : [david.froslie Student Subscription](#)

Subscription ID : 6c124bfa-68d7-4316-a88d-0d0dabdd2f1d

Tags (edit) : [Click here to add tags](#)

Managed Resource Group : [databricks-rg-MarchMadnessADB-yl3g72fasrcz4](#)

URL : [https://adb-1955915938806155.15.azuredatabric...](#)

Pricing Tier : standard

[JSON View](#)

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>

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 Vault URI for the DNS Name).

HomePage / Create Secret Scope

Create Secret Scope

[Cancel](#) [Create](#)

A store for secrets that is identified by a name and backed by a specific storage type. [Learn more](#)

Scope Name [?]

MarchMadnessScope

Manage Principal [?]

All Users

Azure Key Vault [?]

DNS Name

https://marchmadnesskv.vault.azure.net/

Resource ID

/subscriptions/6c124bfa-68d7-43b6-88d1-0d01b1d42614/resourceGroups/marchmadnesskv/providers/Microsoft.KeyVault/vaults/marchmadnesskv

The secret scope named MarchMadnessScope has been added.

Manage secrets in this scope in Azure KeyVault with manage principal = users

OK

Home > MarchMadness > MarchMadnessKV

MarchMadnessKV | Properties

Key vault

Search

Save Discard changes Refresh

- Overview
- Activity log
- Access control (IAM)
- Tags
- Diagnose and solve problems
- Access policies
- Events
- Objects
 - Keys
 - Secrets
 - Certificates
- Settings
 - Access configuration
 - Networking
 - Microsoft Defender for Cloud
 - Properties
 - Locks

Name	MarchMadnessKV
Skus (Pricing tier)	Standard
Location	eastus
Vault URI	https://marchmad
Resource ID	/subscriptions/6c1
Subscription ID	6c124bfa-68d7-43
Subscription Name	david.froslic Stude
Directory ID	ec37a091-b9a6-47
Directory Name	North Dakota Univ
Soft-delete	Soft delete has bee
Days to retain deleted vaults	90
Purge protection	<input checked="" type="radio"/> Disable purge p <input type="radio"/> Enable purge p

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 | Access keys ☆ ...

Storage account

Search

Set rotation reminder Refresh

- 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

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

key1 Rotate key

Last rotated: 10/9/2022 (0 days ago)

Key

..... Show

Connection string

..... Show

key2 Rotate key

Last rotated: 10/9/2022 (0 days ago)

Key

..... Show

Connection string

..... Show

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

Home > MarchMadnessKV | Secrets >

Create a secret ...

Upload options	Manual
Name *	marchmadstore-key
Secret value *
Content type (optional)	
Set activation date	<input type="checkbox"/>
Set expiration date	<input type="checkbox"/>
Enabled	Yes No
Tags	0 tags

Runtime

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

Endpoints	Table service	https://assign1storage.table.core.windows.net/
Locks		
Monitoring	Data Lake Storage	
Insights	Resource ID	/subscriptions/6c124bfa-68d7-4316-a88d-0d0
Alerts	Data Lake Storage	https://assign1storage.dfs.core.windows.net/

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/"
```