

After logging into your Azure Portal, click “Create a resource”. This will take you to the Marketplace where you can search for Azure HDInsight.

Microsoft Azure Search resources, services, and docs (G+)

Home > Create a resource >

Marketplace

Private Marketplace
Favorites
Recently created
Service Providers

Categories


- Get Started
- AI + Machine Learning
- Analytics
- Blockchain
- Compute
- Containers
- Databases
- Developer Tools
- DevOps
- Identity
- Integration

hdinsight

Pricing : All Operating System : All Publisher

Showing results for 'hdinsight'.

Showing 1 to 20 of 37 results.



Azure HDInsight


Microsoft

★★★★☆ 3.8 (163 ratings)

Azure Service

Cloud-based Big Data service. Apache Hadoop, Spark and other popular big data solutions.

Create



Unravel for Azure HDInsight

Unravel Data


No rating

Azure Application

A complete monitoring, tuning and troubleshooting tool for big data running on Microsoft Azure.

Price varies

Create



HFactory Tools (for HDInsight)

Ubeeko


No rating

Virtual Machine

The simplest way to build web services on top of HDInsight HBase

Bring your own license

Create



StreamSets Data Collector for HDInsight

StreamSets


No rating

Azure Application


Build batch & streaming ingest & process data workflows

Price varies


Create




Unravel for Azure HDInsight Application



KoçSistem Azure HDInsight Management



H2O.ai Sparkling Water for HDInsight



Unravel for Azure Databricks and Azure HDInsight

Fill in the required fields and make sure to select Spark as the Cluster type and pick the latest version.

Create HDInsight cluster ...

New to HDInsight? Get started with our [learning resources](#).

Create a managed HDInsight cluster. Select from Spark, Kafka, Hadoop, Storm, and more. [Learn more](#)

Project details

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

Subscription *

Azure subscription 1



Resource group *

(New) rg-sb-free-01



[Create new](#)

Cluster details

Name your cluster, pick a region, and choose a cluster type and version. [Learn more](#)

Cluster name *

sb-sparklab-042121



Region *

West US



Cluster type *

Spark

[Change](#)



Version *

Spark 2.4 (HDI 4.0)



Cluster credentials

Enter new credentials that will be used to administer or access the cluster.

Cluster login username * ⓘ

admin

Cluster login password *

.....



Confirm cluster login password *

.....



Secure Shell (SSH) username * ⓘ

sshuser

Use cluster login password for SSH



Create HDInsight cluster ...

Basics **Storage** Security + networking Configuration + pricing Tags Review + create

Select or create storage accounts that will be used for the cluster's logs, job input, and job output. Configure the cluster's access to these accounts, if needed.

Primary storage

Select or create a storage account that will be the default location for cluster logs and other output.

Primary storage type *	<div>Azure Storage</div>
Selection method * ⓘ	<input checked="" type="radio"/> Select from list <input type="radio"/> Use access key
Primary storage account *	<div>(New) sbsparklab042hdistorage</div> <div>Create new</div>
<div>Container * ⓘ</div>	<div>sb-sparklab-042121-2021-04-21t18-50-37-763z</div>

Data Lake Storage Gen1

Provide details for the cluster to access Data Lake Storage Gen1. The cluster will be able to access any Data Lake Storage Gen1 accounts that the chosen service principal has access to.

Data Lake Storage Gen1 access [Configure access settings](#)

Additional Azure Storage

Link additional Azure Storage accounts to the cluster.

[Add Azure Storage](#)

Custom Ambari DB

Use an external Ambari database for greater flexibility, control, and customization. [Learn More](#)

Since there is a maximum of 40 available cores, I tweaked the number of worker nodes to fit the limit. The data we are processing wasn't large so there weren't any issues. After configuring the cluster size, hit the create button and wait for completion.

Create HDInsight cluster ...


Node configuration

Configure your cluster's size and performance, and view estimated cost information.

The cost estimate represented in the table does not include subscription discounts or costs related to storage, networking, or data transfer.

 This configuration will use 38 of 40 available cores in the East US region.
[View cores usage](#)

+ Add application

Node type	Node size	Number of ...	Estimated cost/h...
Head node	E8 V3 (8 Cores, 64 GB RAM), 0.66 USD/hour	2	1.31 USD
Zookeeper node	A2 v2 (2 Cores, 4 GB RAM), 0.13 USD/hour	3	0.00 (FREE)
Worker node	E8 V3 (8 Cores, 64 GB RAM), 0.66 USD/hour	2 	1.31 USD

☐ Enable autoscale
[Learn More](#)

Total estimated cost/hour 2.62 USD

sbsparklab042hdistorage | Storage Explorer (preview)

Storage account

Search (Ctrl+/)

- Overview
- Activity log
- Tags
- Diagnose and solve problems
- Access Control (IAM)
- Data migration
- Storage Explorer (preview)

Search

- BLOB CONTAINERS
 - sb-sparklab-042-2021-04-22t05-49
- FILE SHARES
- QUEUES
- TABLES

Upload Download Open New Folder

NAME	ACCESS TIER	ACCESS TIER LAST MODIFIED
PySpark		
Scala		
PySpark		
Scala		

Upload blob

sb-sparklab-042-2021-04-22t05-49-12-318z/HdiNotebooks

Files ⓘ

"walmart_stock.csv" "Spark DataFrames Proje..."

☐ Overwrite if files already exist

Advanced

Upload

sbsparklab042hdistorage | Storage Explorer (preview)

Storage account

- Overview
- Activity log
- Tags
- Diagnose and solve problems
- Access Control (IAM)
- Data migration
- Storage Explorer (preview)

Search

- BLOB CONTAINERS
 - sb-sparklab-042-2021-04-22t05-49
- FILE SHARES
- QUEUES
- TABLES

Upload Download Open New Folder

NAME	ACCESS
PySpark	
Scala	
PySpark	
Scala	
Spark DataFrames Project Exercise.ipynb	
walmart_stock.csv	

sb-sparklab-042-2021-04-22t05-49-12-318z/HdiNotebooks

Files ⓘ

Select a file

☐ Overwrite if files already exist

Advanced

Upload

Current uploads

Dismiss: Completed All

Spark DataFrames Proje...	✓ 14 KiB / 14 KiB	...
walmart_stock.csv	✓ 88 KiB / 88 KiB	...

Insert the Jupyter Notebook and csv files into the "HdiNotebooks" folder.



Files

Running

Clusters

Select items to perform actions on them.



PySpark



Scala



Spark DataFrames Project Exercise.ipynb



walmart_stock.csv

Return to the HDInsight Cluster and find Jupyter Notebook under the dashboards header in “Overview”. This is launch Jupyter Notebook in your browser where can see the files you uploaded earlier.

SparkLabResourceGroup

Resource group

Search (Ctrl+ /)



Overview

Activity log

Access control (IAM)

Tags

Events

Settings

Deployments

Security

Policies

Properties

Essentials

Subscription (change) : [Azure subscription 1](#)

Subscription ID : 122ee70b-d020-413a-a237-2904fe1e45a0

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

Filter for any field...

Type == all

Location == all

+ Add filter

Showing 1 to 2 of 2 records. ☐ Show hidden types ⓘ

☐ Name ↑↓

☐ sb-sparklab-042121

☐ sbsparklab042hdistorage

Depl

Loca

Type ↑

Storage

Are you sure you want to delete "SparkL



Warning! Deleting the "SparkLabResourceGroup" resource group is irreversible action you're about to take can't be undone. Going further will delete the group and all the resources in it permanently.

TYPE THE RESOURCE GROUP NAME:

sparklabresourcegroup

AFFECTED RESOURCES

There are 2 resources in this resource group that will be deleted.

Name	Type	Location
sb-sparklab-042121	HDInsight cluster	East US
sbsparklab042hdistorage	Storage account	East US

Be sure to delete the cluster once the tasks are completed to avoid unnecessary charges.