

Creating the Hadoop Cluster

Create HDInsight cluster

[Basics](#) [Storage](#) [Security + networking](#) [Configuration + pricing](#) [Review + create](#)

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

* Resource group

Select existing...

[Create new](#)

Cluster details

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

* Cluster name

Enter new cluster name

* Location

East US

* Cluster type

Select cluster type

* Version

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

☒

Review + create

« Previous

Next: Storage »

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 *	<input type="text" value="Azure Storage"/>
Selection method * ⓘ	<input checked="" type="radio"/> Select from list <input type="radio"/> Use access key
Primary storage account *	<input type="text" value="(New) myazurestorageacct"/> Create new
Container * ⓘ	<input type="text" value="mycontainer"/>

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](#)

Metastore settings

To preserve your Hive and/or Oozie metadata outside of this cluster, select a SQL database for this cluster.

SQL database for Hive ⓘ	<input type="text" value="Select a database, type to filter."/>
SQL database for Oozie ⓘ	<input type="text" value="Select a database, type to filter."/>
SQL database for Ambari ⓘ	<input type="text" value="Select a database, type to filter."/>

[Review + create](#)

« Previous

Next: Security + networking »

Create HDInsight cluster

✓ Validation succeeded.

[Basics](#) [Storage](#) [Security + networking](#) [Configuration + pricing](#) [Review + create](#)

Hadoop 2.7.3 (HDI 3.6)

<price> USD Total estimated cost/hour

This estimate does not include subscription discounts or costs related to storage, networking, or data transfer.

Basics

Subscription	Azure
Resource group	(new) myResourceGroup
Location	East US
Cluster name	(new) myHadoop
Cluster type	Hadoop 2.7.3 (HDI 3.6)
Cluster login username	admin
Secure Shell (SSH) username	sshuser
Use cluster login password for SSH	Enabled

Storage

Primary storage type	Azure Storage
Primary storage account	(new) myhadoopdistorage
Container	myhadoop-2019-09-27t22-00-07-427z
Additional Azure storage	None
Data Lake Storage Gen1 access	Disabled

Cluster configuration

Head	2 nodes, D12 v2 (4 Cores, 28 GB RAM)
Worker	4 nodes, D4 v2 (8 Cores, 28 GB RAM)

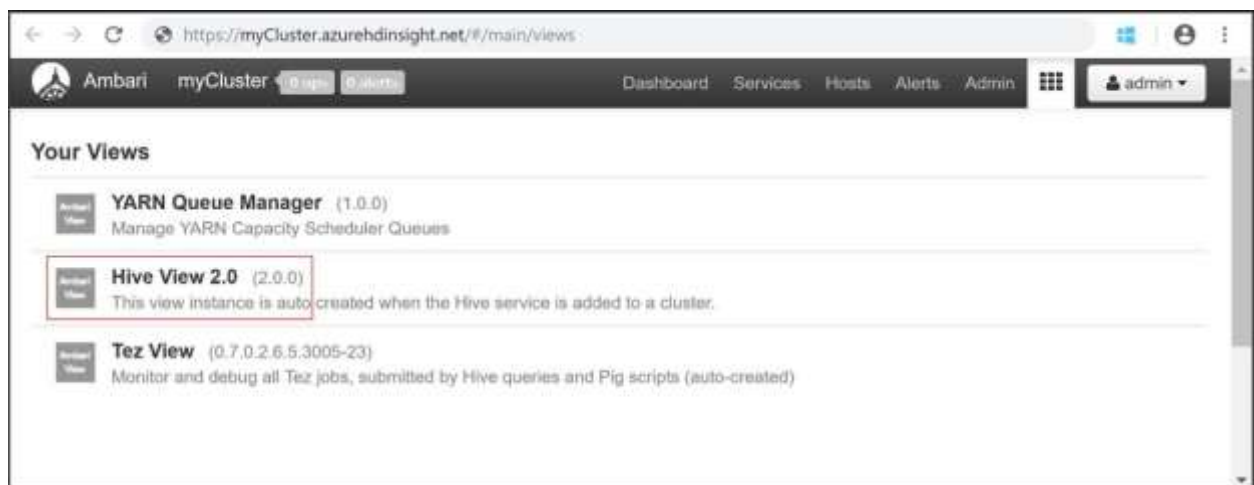
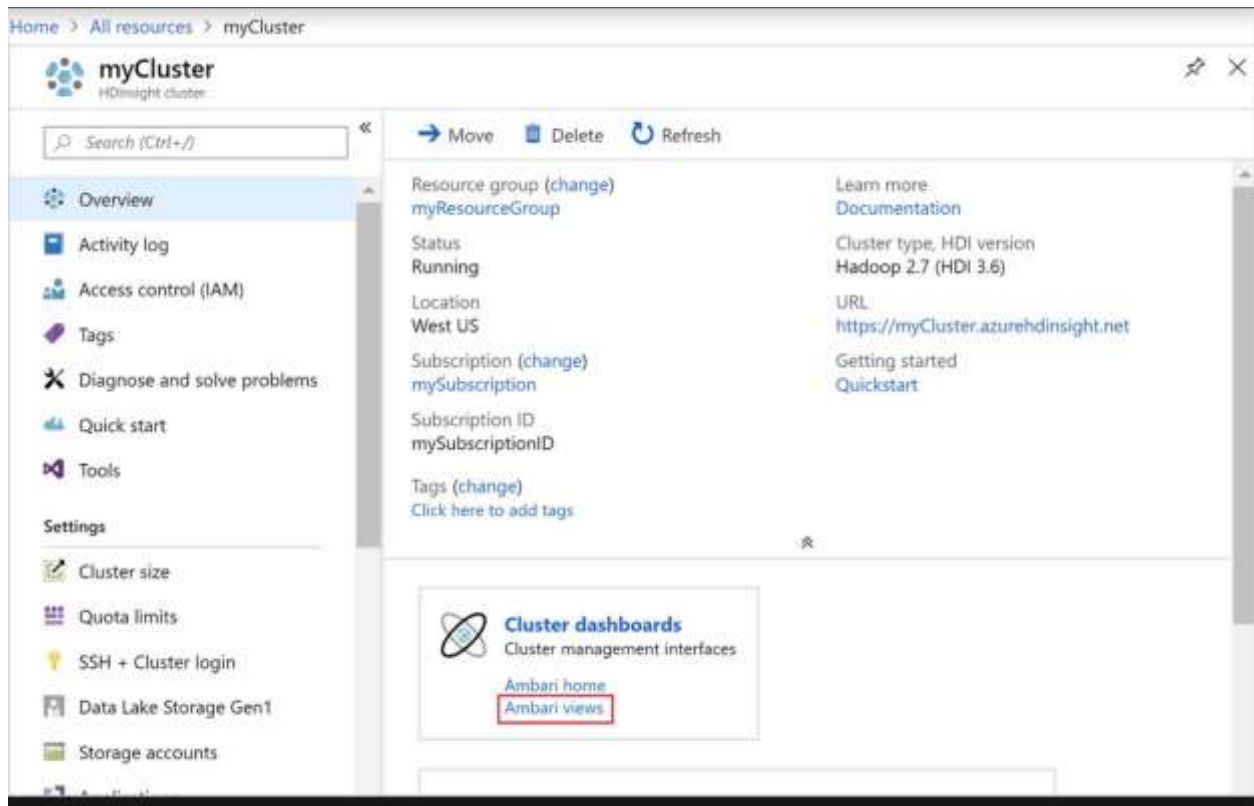
Create

« Previous

Next »


[Download a template for automation](#)

Open Apache Hive View




HIVE

 QUERY

 JOBS

 TABLES

 SAVED QUERIES

 UDFs

 SETTINGS

Worksheet1 *



DATABASE

Select or search database/schema

 default

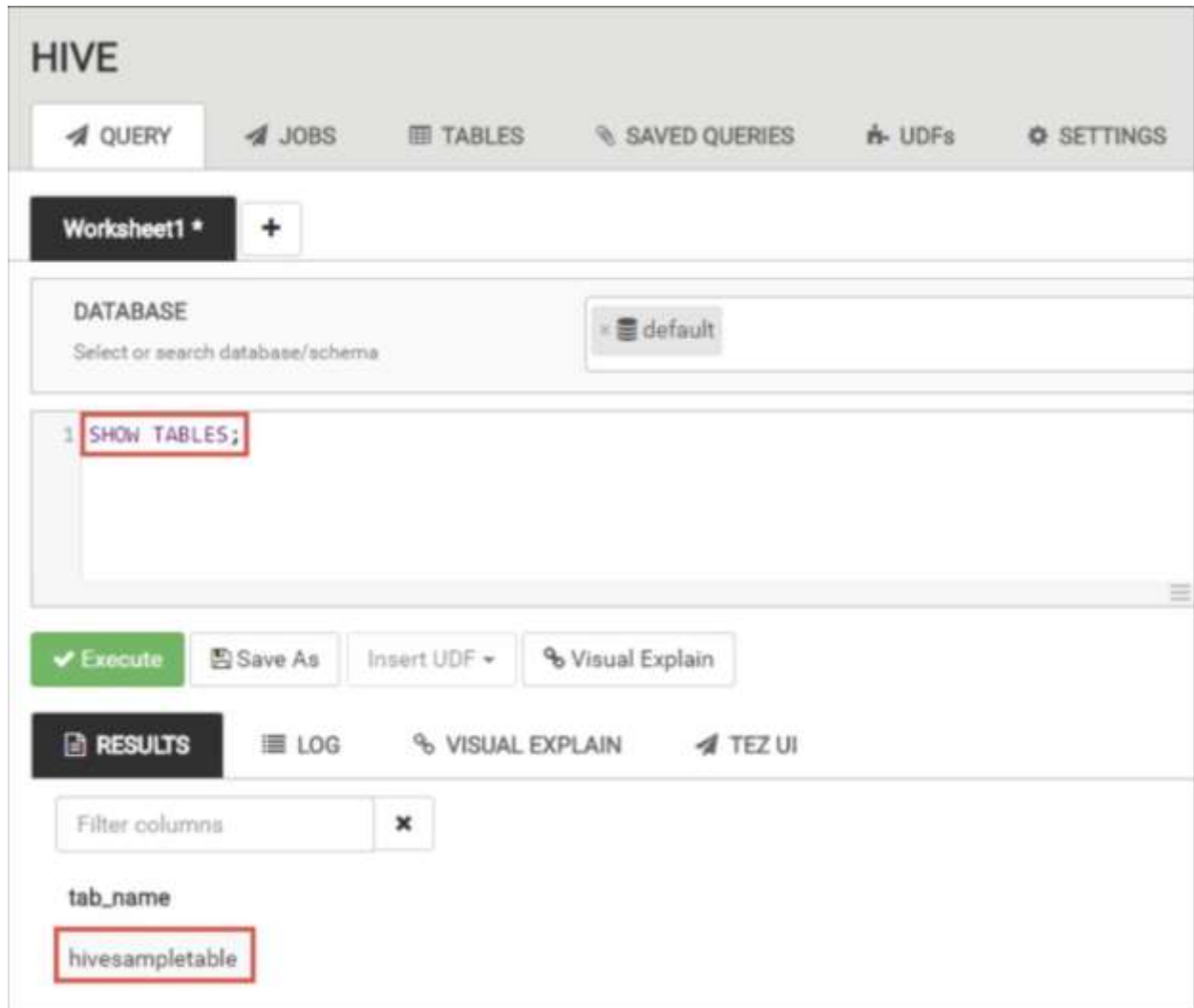
1. `SHOW TABLES;`

 Execute

 Save As

Insert UDF ▾

 Visual Explain



The screenshot shows the Hive web interface. At the top, there's a navigation bar with tabs: QUERY, JOBS, TABLES, SAVED QUERIES, UDFs, and SETTINGS. Below this, a 'Worksheet1' tab is active. The main area has a 'DATABASE' dropdown set to 'default'. A query editor contains the text '1 SHOW TABLES;'. Below the editor are buttons for 'Execute', 'Save As', 'Insert UDF', and 'Visual Explain'. A 'RESULTS' tab is selected, showing a table with one column 'tab_name' and one row 'hivesampletable'. A 'Filter columns' input is also visible.

HIVE

QUERY JOBS TABLES SAVED QUERIES UDFs SETTINGS

Worksheet1 +

DATABASE
Select or search database/schema default

1 SHOW TABLES;

✓ Execute Save As Insert UDF Visual Explain

RESULTS LOG VISUAL EXPLAIN TEZ UI

Filter columns x

tab_name
hivesampletable

Run the following query

```
SELECT * FROM hivesampletable;
```