

Steps:

1. Create HDInsight resource
2. Upload Notebook and Walmart.csv file in Azure Storage
3. Complete exercises on Notebook

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 *	Azure Storage
Selection method * ⓘ	<input checked="" type="radio"/> Select from list <input type="radio"/> Use access key
Primary storage account *	(New) sparklabmadamhdistorage Create new
Container * ⓘ	sparklabmadams-2021-09-27t16-45-30-129z ✓

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.

Account name

[Add Azure Storage](#)

Custom Ambari DB

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

Create HDInsight cluster ...

TLS

Select the minimum TLS version supported for your cluster. [Learn More](#)

Minimum TLS version ⓘ

1.2



Network settings

Resource provider connection ⓘ

Inbound



Connect this cluster to a virtual network. [Learn More](#)

Virtual network ⓘ



Encryption in transit

Configure encryption in transit settings. [Learn More](#)

☐

Enable encryption in transit ⓘ

Encryption at rest

Configure disk encryption settings. [Learn More](#)

☐

Provide your own key from key vault ⓘ

☐

Enable encryption at host on temp data disk ⓘ

Identity

Select a user-assigned service identity to represent your cluster for enterprise security package or disk encryption. [Learn More](#)

User-assigned managed identity ⓘ



[Review + create](#)

[« Previous](#)

[Next: Configuration + pricing »](#)

Create HDInsight cluster ...

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

Configure cluster performance and pricing. [Learn More](#)

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.

✖ There are not enough cores available to support the selected number of nodes. Please adjust the number of nodes selected, pick a different region, or open a support case to request additional HDInsight cores.
[View cores usage](#)
[Open an HDInsight quota increase support case](#)

+ Add application

Node type	Node size	Number of ...	Estimated cost/h...
Head node	D14 v2 (16 Cores, 112 GB RAM), 1.50 USD/hour ▾	2	2.99 USD
Zookeeper node	A2 v2 (2 Cores, 4 GB RAM), 0.13 USD/hour ▾	3	0.00 (FREE)
Worker node	D12 v2 (4 Cores, 28 GB RAM), 0.37 USD/hour ▾	<input type="text" value="4"/>	1.50 USD

You have reached your subscription's cores quota limit in East US. Please choose a different region or request billing support to increase your limit for East US. The value must be between 1 and 1.

☐ Enable autoscale
[Learn More](#)

Total estimated cost/hour 4.49 USD

Script actions

Use script actions to run custom PowerShell or Bash scripts on cluster nodes during cluster provisioning. [Learn about script actions](#)

+ Add script action

Create HDInsight cluster ...

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

Name ⓘ	Value ⓘ	Resource
<input type="text"/>	:	<input type="text"/> HDInsight cluster

Azure Blob Storage on IoT Edge ✎ ...

Microsoft



Azure Blob Storage on IoT Edge [Add to Favorites](#)

Microsoft

Create

Overview Plans Usage Information + Support Reviews

Azure Blob Storage module on IoT Edge enables Edge-local applications that use Azure Storage SDK to alternatively store the data locally on an Edge-local blob store. This module allows you to configure it to automatically upload the data from edge to Azure, and provides support for intermittent internet connectivity. It also allows you to configure it to automatically delete the data on edge.

Having edge-local blob storage would be invaluable in scenarios that require low latency access for local processing, e.g. local image or video processing before the processed data can be transferred to the cloud at infrequent intervals. The blob store location can be switched from the cloud to the edge with a simple change of connection string in code. Further, having Azure-consistent blob storage support on IoT Edge enables business logic of an app written to run on public cloud storage to be re-hosted on to the edge any time, while requiring no code changes in the application itself.

Azure Blob Storage on IoT Edge is a module designed for lightweight multi-OS deployment on top of the IoT Edge platform. This module places the edge-local blob store on