

# Azure HD-Insight (Provisioning + Discovering Web UI)

arnaud.nauwynck@gmail.com

Esilv 2024

# <https://portal.azure.com>

## Home page => Azure Services

The screenshot shows the Microsoft Azure portal home page. At the top, there is a header bar with a back arrow, forward arrow, refresh button, and a search bar containing "portal.azure.com/#home". The Microsoft Azure logo is on the left, and on the right are icons for Copilot, notifications (1), settings, help, and a search icon.

Below the header is a section titled "Azure services" featuring ten icons:

- Create a resource
- HDInsight clusters
- Subscriptions
- Storage accounts
- Virtual networks
- Resource groups
- Microsoft Entra ID
- Enterprise applications
- Kubernetes services
- More services (with a right-pointing arrow)

Under the "Resources" section, there are tabs for "Recent" (which is selected) and "Favorite". Below these tabs is a table with columns for "Name", "Type", and "Last Viewed". A single cube icon is shown, indicating "No resources have been viewed recently". There is a "View all resources" button at the bottom of this section.

At the bottom, there is a "Navigate" section with four links:

- Subscriptions (key icon)
- Resource groups (cube icon)
- All resources (grid icon)
- Dashboard (dash icon)

# <https://portal.azure.com>

Search service "hdinsight"

The screenshot shows the Microsoft Azure portal homepage with a search bar at the top containing the text "hdinsight". A red box highlights the search term in the search bar. Below the search bar, there are several sections: "Azure services" (with a "Create a resource" button and a "HDInsight clusters" icon), "Resources" (Recent and Favorite tabs, Name filter), and "Navigate" (Subscriptions, Resource groups, All resources, Dashboard). The main content area displays search results under "Services", "Marketplace", and "Documentation". The "Services" section shows "HDInsight clusters" (highlighted with a red box) and other items like "Azure HDInsight on AKS cluster pools (preview)" and "Azure HDInsight on AKS clusters (preview)". The "Marketplace" section lists "Azure HDInsight", "Azure HDInsight on AKS (preview)", "HDInsight Hadoop Monitoring", and "HDInsight Spark Monitoring". The "Documentation" section includes links to "Integration runtime - Azure Data Factory & Azure Synapse" and "Azure Hdinsight-Containers SDK for Java". At the bottom, there's a note "Searching all subscriptions." and a "Give feedback" link.

# List of Provisioned HDInsight clusters

← → ⌂ portal.azure.com/#browse/Microsoft.HDInsight%2Fclusters

Microsoft Azure Search resources, services, and docs (G+) Copilot arnaud.nauwynck@gmail.com DEFAULT DIRECTORY (ARNAUD)

Home > HDInsight clusters ⚙ ...

Default Directory (arnaudnauwynck@gmail.onmicrosoft.com)

+ Create ⚙ Manage view Refresh Export to CSV Open query Assign tags Delete

Filter for any field... Subscription equals all Resource group equals all Location equals all Add filter

Showing 0 to 0 of 0 records. No grouping List view

Name ↑↓	Cluster type ↑↓	Status ↑↓	Resource group ↑↓	Location ↑↓	Cluster Version ↑↓
---------	-----------------	-----------	-------------------	-------------	--------------------

 No HDInsight clusters to display

Create an HDInsight cluster to process massive amounts of data using popular open-source frameworks such as Hadoop, Spark, Hive, LLAP, Kafka, Storm, ML Services, and more.

[Create HDInsight cluster](#)

[Learn more about HDInsight ↗](#)

# Create a Cluster

<https://portal.azure.com/#create/Microsoft.HDInsightCluster>

The screenshot shows the 'Create HDInsight cluster' wizard in the Microsoft Azure portal. The 'Basics' tab is selected. In the 'Project details' section, under 'Subscription \*', a dropdown menu shows 'Pay-As-You-Go'. A red box highlights an error message: 'The subscription is not registered to use the 'Microsoft.HDInsight' resource provider. [Click here to register.](#)' Under 'Resource group \*', a dropdown menu shows '(New) test-hdi1' with a 'Create new' link below it. In the 'Cluster details' section, under 'Cluster name \*', a text input field contains 'cluster1'. A red box highlights an error message: 'There was an error trying to validate cluster name. Please retry again.' At the bottom, there are 'Review + create', '« Previous', and 'Next: Storage »' buttons.

Microsoft Azure

Search resources, services, and docs (G+)

Home > HDInsight clusters >

## Create HDInsight cluster

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

New to HDInsight? Get started with our [training 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 \*

Pay-As-You-Go

✖ The subscription is not registered to use the 'Microsoft.HDInsight' resource provider. [Click here to register.](#)

Resource group \*

(New) test-hdi1

Create new

**Cluster details**

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

Cluster name \*

cluster1

✖ There was an error trying to validate cluster name. Please retry again.

**Review + create**

« Previous

Next: Storage »

# Need Register Feature

The screenshot shows the Microsoft Azure Pay-As-You-Go Resource providers page. The top navigation bar includes the Microsoft Azure logo, a search bar, and various icons for Copilot, notifications, settings, help, and user profile.

The main content area displays the 'Pay-As-You-Go | Resource providers' section. On the left, there's a sidebar with links like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Security, Events, Cost Management, Cost analysis, Cost alerts, Budgets, Advisor recommendations, and Billing.

The main pane has a search bar with 'hdins' typed in, and buttons for Register, Unregister, Refresh, and Feedback. Below the search bar are filters for Status: All and Registration Policy: All.

A table lists resource providers:

Provider ↑	Status	Registration Policy
Microsoft.HDInsight	... Registering ⓘ	RegistrationRequired ⓘ

The 'Provider' column header is sorted in ascending order. The 'Status' and 'Registration Policy' columns are also visible. The 'Microsoft.HDInsight' row is highlighted with a red box around the provider name and the status cell.

# Select HDInsight, click "Register" ... wait few minutes

Home > Pay-As-You-Go

## Pay-As-You-Go | Resource providers

Subscription

Search: hdins

Provider: Microsoft.HDInsight Status: registered

Provider ↑	Status	Registration Policy
Microsoft.HDInsight	registered	RegistrationRequired 1

Overview Activity log Access control (IAM) Tags Diagnose and solve problems Security Events Cost Management

Register Unregister Refresh Feedback

Status : All Registration Policy : All

This screenshot shows the Azure Resource Providers page for the 'Pay-As-You-Go' subscription. The search bar at the top contains the text 'hdins'. A single result is listed in the table: 'Microsoft.HDInsight' is shown as 'registered' with a green checkmark icon. The registration policy is marked as 'RegistrationRequired' with a blue information icon.

# Create HDI - Step 1: Create Resource Group

Home > HDInsight clusters >

## Create HDInsight cluster ...

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

New to HDInsight? Get started with our [training 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 \*

Pay-As-You-Go

Resource group \*

[Create new](#)

### Cluster details

Name your cluster, pick a region, and choose a type.

Cluster name \*

Name \*

test-cluster-hdi1



OK

Cancel

Region \*

Cluster type \*

[Review + create](#)

[« Previous](#)

[Next: Storage »](#)

# Create HDI - Choose Cluster Type (=> Spark)

Home > HDInsight clusters >

## Create HDInsight cluster

your resources.

Subscription \*

Pay-As-You-Go

Resource group \*

(New) test-cluster-hdi1

[Create new](#)

### Cluster details

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

Cluster name \*

test-cluster-hdi1

Region \*

West Europe

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

[Review + create](#)

[« Previous](#)

[Next: Storage »](#)

## Select cluster type

### Hadoop

Petabyte-scale processing with Hadoop components like MapReduce, Hive (SQL on Hadoop), Pig, Sqoop and Oozie.

[Select](#)

### Spark

Fast data analytics and cluster computing using in-memory processing.

[Select](#)

### Kafka

Build a high throughput, low-latency, real-time streaming platform using a fast, scalable, durable, and fault-tolerant publish-subscribe messaging system.

[Select](#)

### HBase

Fast and scalable NoSQL database. Available with both standard and premium (SSD) storage options.

[Select](#)

### Interactive Query

Build Enterprise Data Warehouse with in-memory analytics using Hive (SQL on Hadoop) and LLAP (Low Latency Analytical Processing).

Note that this feature requires high memory instances.

[Select](#)

# Create HDI - Choose Spark Version available 3.3.0 as of 2024 (but latest spark=3.5.3)

Home > HDInsight clusters >

## Create HDInsight cluster

Cluster name \* test-cluster-hdi1 ✓

Region \* West Europe ▾

Availability zone ⓘ ▾

Cluster type \* **Spark** Change

Version \*

- Spark 3.3.0 (HDI 5.1) (selected)
- Spark 2.4 (HDI 4.0)
- Spark 3.1 (HDI 5.0)
- Spark 3.3.0 (HDI 5.1)

Cluster credentials

Enter new credentials that will be used to 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 HDI - Admin Username + Password

Home > HDInsight clusters >

## Create HDInsight cluster

Cluster name \*

 ✓

Region \*

 ✓

Availability zone ⓘ

 ✓

Cluster type \*

**Spark**  
[Change](#)

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 \* ⓘ

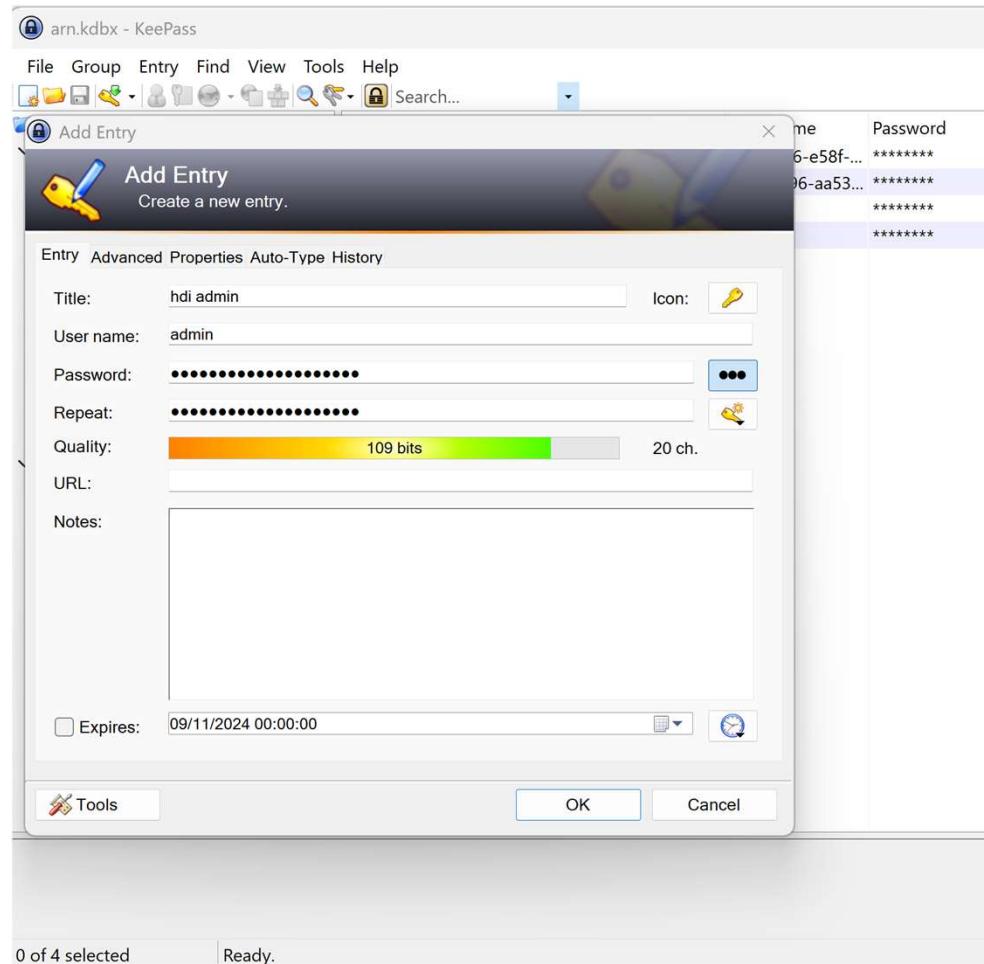
 sshuser

Use cluster login password for SSH

- ✓ Must be at least 10 characters in length.
- ✓ Must contain at least one digit.
- ✓ Must contain at least one uppercase character.
- ✓ Must contain at least one lowercase character.
- ✓ Must contain at least one non-alphanumeric character, except characters ' " ` \ ).
- ✓ Must not contain any three consecutive characters from cluster login username.
- ✓ Should not be a very common and weak password.

[Review + create](#) [« Previous](#) [Next: Storage »](#)

# Always Generate your Random Passwords (and use KeePass)



# Create HDI - Next : Storage

Home > HDInsight clusters >

## Create HDInsight cluster

Cluster name

Region \*

Availability zone

Cluster type \* **Spark**  
[Change](#)

Version \*

### Cluster credentials

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

Cluster login username \*

Cluster login password \*  ✓

Confirm cluster login password \*  ✓

Secure Shell (SSH) username \*

Use cluster login password for SSH

---

[Review + create](#)[« Previous](#)[Next: Storage »](#)

# change default Storage Type "ADLS Gen 2"

Home > HDInsight clusters >

## 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 \*

Primary storage account \*

Filesystem \* ⓘ

Enable secure channel

Azure Data Lake Storage Gen2

Azure Storage

Azure Data Lake Storage Gen2

### Identity

Select a user-assigned managed identity to represent the cluster for Azure Data Lake Gen2 Storage account access. Only identities with access to the selected storage account are listed. Assign the managed identity to the 'Storage Blob Data Owner' role on the storage account. [Learn More](#)

User-assigned managed identity \* ⓘ

[Review + create](#)

[« Previous](#)

[Next: Security + networking »](#)

# Create a Storage Account (in same Region, with type ADLSGen2)

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

Home > Storage accounts >

## Create a storage account

Select the subscription in which to create the new storage account. Choose a new or existing resource group to organize and manage your storage account together with other resources.

Subscription \* Pay-As-You-Go

Resource group \* (New) storage-test-hdi [Create new](#)

**Instance details**

Storage account name \* storagetesthd1

Region \* (Europe) West Europe [Deploy to an Azure Extended Zone](#)

Primary service Azure Blob Storage or Azure Data Lake Storage Gen 2

Performance \*  Standard: Recommended for most scenarios (general-purpose v2 account)  
 Premium: Recommended for scenarios that require low latency.

Redundancy \* Locally-redundant storage (LRS)

[Previous](#) [Next](#) [Review + create](#)

# Creating Storage Account...

Home >

 storagetesthdi1\_1731168750012 | Overview ⚡ ...

Deployment

Search × « Delete Cancel Redeploy Download Refresh

 Overview  Inputs  Outputs  Template

... Deployment is in progress

 Deployment name: storagetesthdi1\_1731168750012  
Subscription: Pay-As-You-Go  
Resource group: storage-test-hdi

Start time: 11/9/2024, 5:13:36 PM  
Correlation ID: 2c1144e2-1029-4a8c-a77b-e7e42a20cf57 

▴ Deployment details

Resource	Type	Status	Operation details
No results.			

# Created Storage Account

The screenshot shows the Microsoft Azure Storage account overview page for 'storagetesthd1'. The top navigation bar includes the Microsoft Azure logo, a search bar, and various icons for Copilot, notifications, settings, and help.

The main content area displays the storage account details under the 'Essentials' section:

Setting	Value	Performance	Replication	Account kind	Provisioning state	Created
Resource group (move)	<a href="#">storage-test-hdi</a>	Standard				
Location	: westeurope					
Subscription (move)	<a href="#">Pay-As-You-Go</a>					
Subscription ID	: 4cc76a76-e58f-4f11-bb52-c1cb81ac8cc0					
Disk state	: Available					
Tags (edit)	: Add tags					

The left sidebar lists various management options:

- Overview
- Activity log
- Tags
- Diagnose and solve problems
- Access Control (IAM)
- Data migration
- Events
- Storage browser
- Storage Mover
- Partner solutions
- Data storage
  - Containers
  - File shares
  - Queues
  - Tables
- Security + networking
- Data management

The 'Properties' tab is selected, showing sections for Blob service, Security, Networking, and other capabilities.

# Create HDI - Select your Storage Account (bug Azure ... need close+reenter HDI Form)

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 \*

Primary storage account \*  storagetesthdi1

Filesystem \*  storagetesthdi1

Enable secure channel (i)

**Identity**

Select a user-assigned managed identity to represent the cluster for Azure Data Lake Gen2 Storage account access. Only identities with access to the selected storage account are listed. Assign the managed identity to the 'Storage Blob Data Owner' role on the storage account. [Learn More](#)

User-assigned managed identity \*

# ... Need a Managed Identity

Home > HDInsight clusters >

## Create HDInsight cluster

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 Data Lake Storage Gen2

Primary storage account \*

storagetesthdi1

Filesystem \* ⓘ

test-cluster-hdi1-2024-11-09t16-19-16-813z

Enable secure channel ⓘ

### Identity

Select a user-assigned managed identity to represent the cluster for Azure Data Lake Gen2 Storage account access. Only identities with access to the selected storage account are listed. Assign the managed identity to the 'Storage Blob Data Owner' role on the storage account. [Learn More](#)

User-assigned managed identity \* ⓘ

No identities found with access to the selected storage account.

### Additional Azure Storage

Link additional Azure Storage accounts to the cluster.

No available items.

[Add Azure Storage](#)

[Review + create](#)

[« Previous](#)

[Next: Security + networking »](#)

# Managed Identity

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

## Managed Identities ...

Default Directory (arnaudnauwynck@gmail.onmicrosoft.com)

+ Create Manage view Refresh Export to CSV Open query Assign tags Delete

Filter for any field... Subscription equals all Resource group equals all Location equals all Add filter

Showing 0 to 0 of 0 records. No grouping

Name ↑↓	Type ↑↓	Resource group ↑↓	Location ↑↓	Subscription ↑↓
---------	---------	-------------------	-------------	-----------------

 No managed identities to display

Simplify access management for services running on Azure with automatically managed identities that are securely provisioned and scoped to specific Azure resources.

Create managed identity Learn more ↗

# Create Managed Identity

[Home](#) > [Managed Identities](#) >

## Create User Assigned Managed Identity

...

[Basics](#)   [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 \* ⓘ

Pay-As-You-Go

Resource group \* ⓘ

(New) rg-mgid

[Create new](#)

### Instance details

Region \* ⓘ

West Europe

Name \* ⓘ

mgid-test-hdi



[Previous](#)

[Next](#)

[Review + create](#)

# Managed Identity

Home > Microsoft.ManagedIdentity-20241109172434 | Overview >

 mgid-test-hdi | Overview ⚡ ☆ ...

Managed Identity

Search Delete

Overview

Activity log

Access control (IAM)

Tags

Azure role assignments

Associated resources (preview)

JSON View

Essentials

Resource group	: <a href="#">rg-mgid</a>	Type	: Microsoft.ManagedIdentity/userAssignedIdentities
Location	: West Europe	Client ID	: ca4f2fcb-97d3-4b7a-9f3a-bdd7d6028c60
Subscription	: <a href="#">Pay-As-You-Go</a>	Object (principal) ID	: ea47109e-83a7-4904-81a4-997bc853f7c4
Subscription ID	: 4cc76a76-e58f-4f11-bb52-c1cb81ac8cc0		

> Settings

> Monitoring

> Automation

> Help

# Adding "Storage Blob Data Owner" Role to Managed Identity

Home > Storage accounts > storagetesthd1 | Access Control (IAM) >

## Add role assignment ...

Role Members • Conditions Review + assign

A role definition is a collection of permissions. You can use the built-in roles or you can create your own custom roles. [Learn more](#)

Job function roles Privileged administrator roles

Grant access to Azure resources based on job function, such as the ability to create virtual machines.

storage blo		Type : All	Category : All
Name ↑↓	Description ↑↓	Type ↑↓	
Defender CSPM Storage Data Scanner	Grants access to read blobs and files. This role is used by the data scanner of Defender CSPM.	BuiltInRole	
Defender for Storage Data Scanner	Grants access to read blobs and update index tags. This role is used by the data scanner of Defender for Storage.	BuiltInRole	
Storage Blob Data Contributor	Allows for read, write and delete access to Azure Storage blob containers and data	BuiltInRole	
Storage Blob Data Owner	Allows for full access to Azure Storage blob containers and data, including assigning POSIX access control.	BuiltInRole	
Storage Blob Data Reader	Allows for read access to Azure Storage blob containers and data	BuiltInRole	
Storage Blob Delegator	Allows for generation of a user delegation key which can be used to sign SAS tokens	BuiltInRole	

Showing 1 - 6 of 6 results.

[Review + assign](#)

[Previous](#)

[Next](#)

# Adding Role..

Home > Storage accounts > storagetesthd1 | Access Control (IAM) >

## Add role assignment ...

Role    **Members \***    Conditions    Review + assign

Selected role    Storage Blob Data Owner

Assign access to     User, group, or service principal  
                           Managed identity

Members    + Select members

Name	Object ID	Type
------	-----------	------

No members selected

Description

Optional

[Review + assign](#)

[Previous](#)

[Next](#)

## Select managed identities

⚠ Some results might be hidden due to your ABAC condition.

Subscription \*

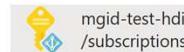
Pay-As-You-Go

Managed identity

User-assigned managed identity (2)

Select ⓘ

Search by name



mgid-test-hdi  
/subscriptions/4cc76a76-e58f-4f11-bb52-c1cb81ac8cc0/resourceGroups/rg-mgid/...

Selected members:



uamgid-test-hdi  
/subscriptions/4cc76a76-e58f-4f11-bb52-c1cb81ac8cc0/resourceGroups/r...

[Remove](#)

[Select](#)

[Close](#)

[Feedback](#)

# Create HDI - Managed Identity

Home > HDInsight clusters >

## Create HDInsight cluster

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 Data Lake Storage Gen2

Primary storage account \*

storagetesthdi1

Filesystem \* ⓘ

test-hdi-cluster1-2024-11-09t19-51-05-515z

Enable secure channel ⓘ

### Identity

Select a user-assigned managed identity to represent the cluster for Azure Data Lake Gen2 Storage account access. Only identities with access to the selected storage account are listed. Assign the managed identity to the 'Storage Blob Data Owner' role on the storage account. [Learn More](#)

User-assigned managed identity \* ⓘ

uamgid-test-hdi

### Additional Azure Storage

Link additional Azure Storage accounts to the cluster.

[Add Azure Storage](#)

[Review + create](#)

[« Previous](#)

[Next: Security + networking »](#)

# HDI.. Need Increase Subscription Cores Quota Limit

## Create HDInsight cluster ...

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.

✖ Sku family standarda0\_a7family quota is not enough, available value: 0, additional cores: 4;Sku family standardav2family quota is not enough, available value: 0, additional cores: 5; 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	A5 (2 Cores, 14 GB RAM), 0.24 EUR/hour	2	0.47 EUR
Zookeeper node	A1 v2 (1 Cores, 2 GB RAM), 0.00 EUR/hour (FR...	3	0.00 (FREE)
Worker node	A2m v2 (2 Cores, 16 GB RAM), 0.15 EUR/hour	1	0.15 EUR

You have reached your subscription's cores quota limit in West Europe. Please choose a different region or request billing support to increase your limit for West Europe.

Enable managed disk

[Review + create](#)

[« Previous](#)

[Next: Tags »](#)

# For Quota .. need Resource provider "Microsoft.Compute"

Home > Subscriptions > Pay-As-You-Go

Pay-As-You-Go | Resource providers

Subscription

Search: compute

Register Unregister Refresh Feedback

Provider ↑	Status	Registration Policy
Microsoft.ClassicCompute	NotRegistered	RegistrationRequired ⓘ
Microsoft.Compute	Registering ⓘ	RegistrationRequired ⓘ
Microsoft.ComputeSchedule	NotRegistered	RegistrationRequired ⓘ
Microsoft.HybridCompute	NotRegistered	RegistrationRequired ⓘ

Provider list:

- External services
- Payment methods
- Partner information
- Settings
  - Programmatic deployment
  - Resource groups
  - Resources
  - Preview features
  - Usage + quotas
  - Policies
- My permissions
- Resource providers
  - Deployments
  - Deployment stacks
  - Properties
  - Resource locks

# Quota vCPUs

Home > Subscriptions > Pay-As-You-Go

**Pay-As-You-Go | Usage + quotas** ...

Subscription

Search External services Payment methods Partner information Programmatic deployment Resource groups Resources Preview features Usage + quotas Policies My permissions Resource providers Deployments Deployment stacks Properties Resource locks

New Quota Request Refresh Download

You can now set up alerts for your Quota usage and receive notifications. Simply click on any Quota to create one. [Learn More](#).

**Recommended** To view and manage quotas across all your subscriptions from a central location, go to [Azure Quotas](#).

Search Provider : Compute Region : France Central Usage : Show all

Showing 1 to 100 of 176 records in 1 groups. Group by usage

Quota name	Region	Subscription	Current Usage ↓	Adjustable
No usage (Showing 100 of 176)				
Availability Sets	France Central	Pay-As-You-Go	0%	0 of 2,500
Total Regional vCPUs	France Central	Pay-As-You-Go	0%	0 of 10
Virtual Machines	France Central	Pay-As-You-Go	0%	0 of 25,000
Virtual Machine Scale Sets	France Central	Pay-As-You-Go	0%	0 of 2,500
Dedicated vCPUs	France Central	Pay-As-You-Go	0%	0 of 3,000
Cloud Services	France Central	Pay-As-You-Go	0%	0 of 2,500
Total Regional Spot vCPUs	France Central	Pay-As-You-Go	0%	0 of 3
⚠ Basic A Family v... (troubleshoot)	France Central	Pay-As-You-Go	0%	0 of 10

< Previous Page 1 of 2 Next >

# Support Request Answer: Quota Increased

Home > Help + support | All support requests >

## Support Request: Quota request for HDInsight

[New message](#) [Upload file](#) [Update contact info](#) [Advanced diagnostic information](#) [Change severity](#) [Close request](#) [Open in Mobile](#)

### Overview

Severity	C
Status	Open
Support request ID	2411090050000774
Created on	Sat, Nov 9, 2024, 11:30:01 PM
Created by	Arnaud Nauwynck
Contact method	Email: arnaud.nauwynck@gmail.com
Support plan	Basic
Advanced diagnostic information	Permission granted

[View All](#)

### Give Feedback

[Tell us about your experience with support request overview](#)

### Messages

### Issue details

 support@mail.support.microsoft.com

Sun, Nov 10, 2024, 11:46:07 AM

Hello Arnaud Nauwynck,

Warm Greetings!

I am glad to share an update to you that, your Quota request for HD Insight for 10 cores under Total Regional vCPUs in France Central Region on your subscription ID 4cc76a76-e58f-4f11-bb52-c1cb81ac8cc0 has been successfully completed.

Please verify and let me know if you see the changes. To verify your current quota, please see the steps at [Check Azure resource usage against limits | Microsoft Docs](#). Please be sure to select the "show all" filter to see the allocated quota you had requested.

Please do let me know if you need any further assistance from my end. I will be happy to assist.

Looking forward to your response.

# Quota ...

Home > Quotas

**Quotas | My quotas** ... X

Search X « » New Quota Request Refresh Download

i You can now set up alerts for your Quota usage and receive notifications. Simply click on any Quota to create one. [Learn More.](#) X

Search Provider : Azure HDInsight Subscription : Pay-As-You-Go Region : France Central Usage : Show all

Showing 1 to 8 of 8 records in 1 groups. Group by usage

Quota name	Region	Subscription	Current Usage ↓	Request adjustm...
<b>No usage (8)</b>				
Total Regional vCPUs	France Central	Pay-As-You-Go	<div style="width: 0%;"></div> 0% 0 of 10	<span>edit</span>
standardA0_A7Family	France Central	Pay-As-You-Go	<div style="width: 0%;"></div> 0% 0 of 0	<span>edit</span>
StandardAV2Family	France Central	Pay-As-You-Go	<div style="width: 0%;"></div> 0% 0 of 0	<span>edit</span>
StandardDsV2Family	France Central	Pay-As-You-Go	<div style="width: 0%;"></div> 0% 0 of 0	<span>edit</span>
StandardDV2Family	France Central	Pay-As-You-Go	<div style="width: 0%;"></div> 0% 0 of 0	<span>edit</span>
standardElv3Family	France Central	Pay-As-You-Go	<div style="width: 0%;"></div> 0% 0 of 0	<span>edit</span>
standardEv3Family	France Central	Pay-As-You-Go	<div style="width: 0%;"></div> 0% 0 of 0	<span>edit</span>
standardFFamily	France Central	Pay-As-You-Go	<div style="width: 0%;"></div> 0% 0 of 0	<span>edit</span>

# Create Cluster: 2xA5 + 3xA1+ 1xA2

Home > Create HDInsight cluster ...

Skus available: 2xA5 + 3xA1 + 1xA2

**SKU Quota Error:** Sku family standarda0\_a7family quota is not enough, available value: 0, additional cores: 4;Sku family standardav2family quota is not enough, available value: 0, additional cores: 5;. 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	A5 (2 Cores, 14 GB RAM), 0.31 EUR/hour	2	0.62 EUR
	Zookeeper node	A1 v2 (1 Cores, 2 GB RAM), 0.00 EUR/hour (FREE)	3	0.00 (FREE)
	Worker node	A2m v2 (2 Cores, 16 GB RAM), 0.20 EUR/hour	1	0.20 EUR

Enable managed disk

Enable autoscale [Learn More](#)

Total estimated cost/hour 0.82 EUR

Script actions

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

[Review + create](#) [« Previous](#) [Next: Tags »](#)

# Create HDI ...

[Home](#) > [HDInsight clusters](#) >

## Create HDInsight cluster

• • • •

### Security + networking

Minimum TLS version	1.2
Resource provider connection	Inbound
Encryption at rest	Disabled
Encryption in transit	Enabled
Encryption at host on temp data disk	Disabled

### Storage

Primary storage type	Azure Data Lake Storage Gen2
Primary storage account	anauwstorage2
Filesystem	test-hdi1-2024-11-11t13-36-32-609z
User-assigned managed identity	mgid-test-hdi
Additional Azure Storage	None
Data Lake Storage Gen1 access	Disabled

### Cluster configuration

Head	2 nodes, D12 v2 (4 Cores, 28 GB RAM)
Zookeeper	3 nodes, A2 v2 (2 Cores, 4 GB RAM)
Worker	1 nodes, D12 v2 (4 Cores, 28 GB RAM)

---

[Create](#)

[« Previous](#)

[Next »](#)

[Download a template for automation](#)

# "Deployment" object, in progress

Home >

## HDInsight\_2024-11-11T14.02.26.749Z | Overview

Deployment

Search X ⏪ Delete Cancel ⏫ Redeploy ⏴ Download ⏴ Refresh

Overview Inputs Outputs Template

Deployment is in progress

Deployment name: HDInsight\_2024-11-11T14.02.... Start time: 11/11/2024, 3:02:28 PM  
Subscription: Pay-As-You-Go Correlation ID: 49e8e75c-3872-4ec0-8788-7cd22ed9e49e  
Resource group: test-hdi

Deployment details

Resource	Type	Status	Operation details
an-test-hdi4	Microsoft.HDInsight/clu...	OK	Operation details

# HDInsight Resource, Status=Accepted (!= Running)

Microsoft Azure    Search resources, services, and docs (G+/-)    Copilot    Notifications (3)    Settings    Help    arnaud.nauwynck@gmail.com

Home >

## HDInsight clusters

Default Directory (arnaudnauwynck@gmail.onmicrosoft.com)

+ Create    Manage view    Refresh    Export to CSV    Open query    Assign tags    Delete

Filter for any field...    Subscription equals all    Resource group equals all    Location equals all    Add filter

Showing 1 to 1 of 1 records.

Name	Cluster type	Status	Resource group	Location	Cluster Version
an-test-hdi4	Spark	Accepted	test-hdi	France Central	5.1.3000.0

# HDInsight - Resource Overview

[Home](#) > [HDInsight clusters](#) >

**HDInsight clusters** <> X

Default Directory (arnaudnauwynck@gmail.onmicrosoft.com)

+ Create Manage view ...

Filter for any field...

Name	...
an-test-hdi4	...

**an-test-hdi4** HDInsight cluster

Search Delete Refresh Feedback

**Overview** JSON View

Activity log  
Access control (IAM)  
Tags  
Diagnose and solve problems  
Settings  
Monitoring  
Automation  
Help

Resource group ([move](#))  
[test-hdi](#)  
Status  
Azure VM configuration  
Location  
France Central  
Subscription ([move](#))  
[Pay-As-You-Go](#)  
Subscription ID  
4cc76a76-e58f-4f11-bb52-c1cb81ac8cc0

Learn More  
[Documentation](#)  
Cluster type, HDI version  
Spark 3.3 (HDI 5.1)  
Image version  
5.1.3000.0  
URL  
<https://an-test-hdi4.azurehdinsight.net>  
Cluster ID  
0c48ea68f3804fc8885eaa669c1217aa

Tags ([edit](#))  
[Add tags](#)

Overview Get started

Dashboards

 Ambari home	 Ambari views	 Zeppelin notebook	 Jupyter notebook
 Spark history server	 Yarn		

< Page 1 > of 1

# HDInsight - Quota "Core in use"

an-test-hdi4 | Quota limits

HDInsight cluster

Search Cluster size Feedback

Each of your subscriptions has a per-region quota on the number of cores that HDInsight clusters can consume. If you'd like to increase the core quota in a region, please [request billing support](#)

Subscription: Pay-As-You-Go

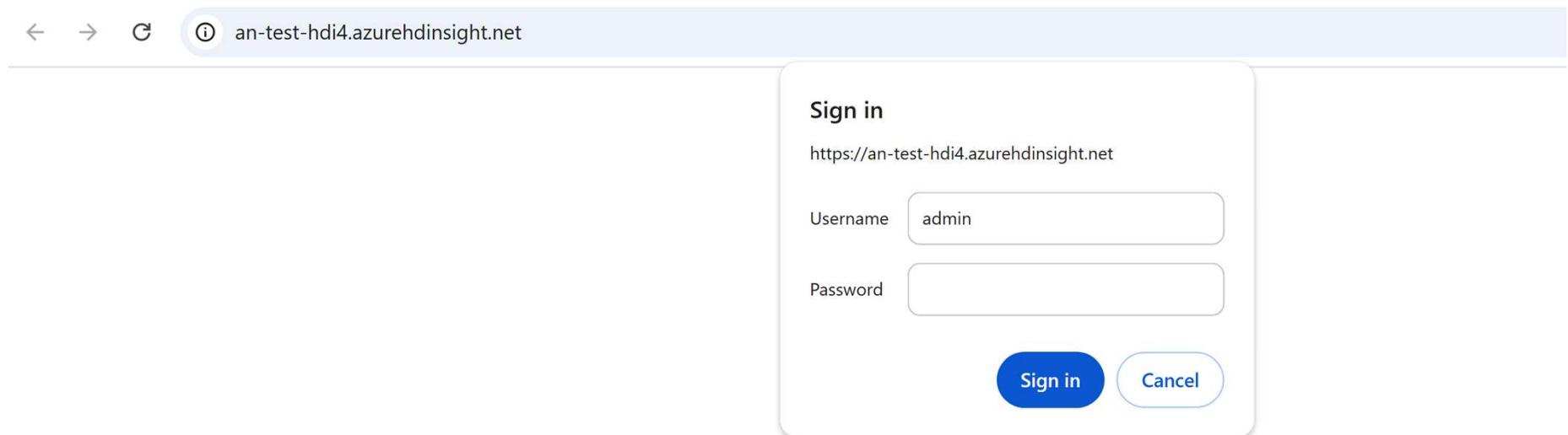
Filter by region: France Central

Quota name	Region	Cores in use	Available cores	Total cores
cores	francecentral	12	58	70
standardA0_A7Family	francecentral	0	0	0
StandardAV2Family	francecentral	0	30	30
StandardDsV2Family	francecentral	0	0	0
StandardDV2Family	francecentral	12	18	30
standardElv3Family	francecentral	0	0	0
standardEv3Family	francecentral	0	0	0
standardFFamily	francecentral	0	0	0

SSH + Cluster login  
Data Lake Storage Gen1  
Storage accounts  
Applications  
Script actions  
External metastores  
Properties

Using HDI .. Discovering the UI

# login Http Gateway (=Ambari)



# Still in progress => Ambari Alerts

← → ⌂ an-test-hdi4.azurehdinsight.net/#/main/dashboard/metrics

Ambari / Dashboard / Metrics

an-test-hdi4   admin

METRICS HEATMAPS CONFIG HISTORY METRIC ACTIONS LAST 1 HOUR

NameNode Heap: 17% 

HDFS Disk Usage: 8% 

NameNode CPU WIO: 0.4% 

DataNodes Live: 1/1

NameNode RPC: 0 ms

Memory Usage: No Data Available

Network Usage: No Data Available

CPU Usage: No Data Available

Cluster Load: No Data Available

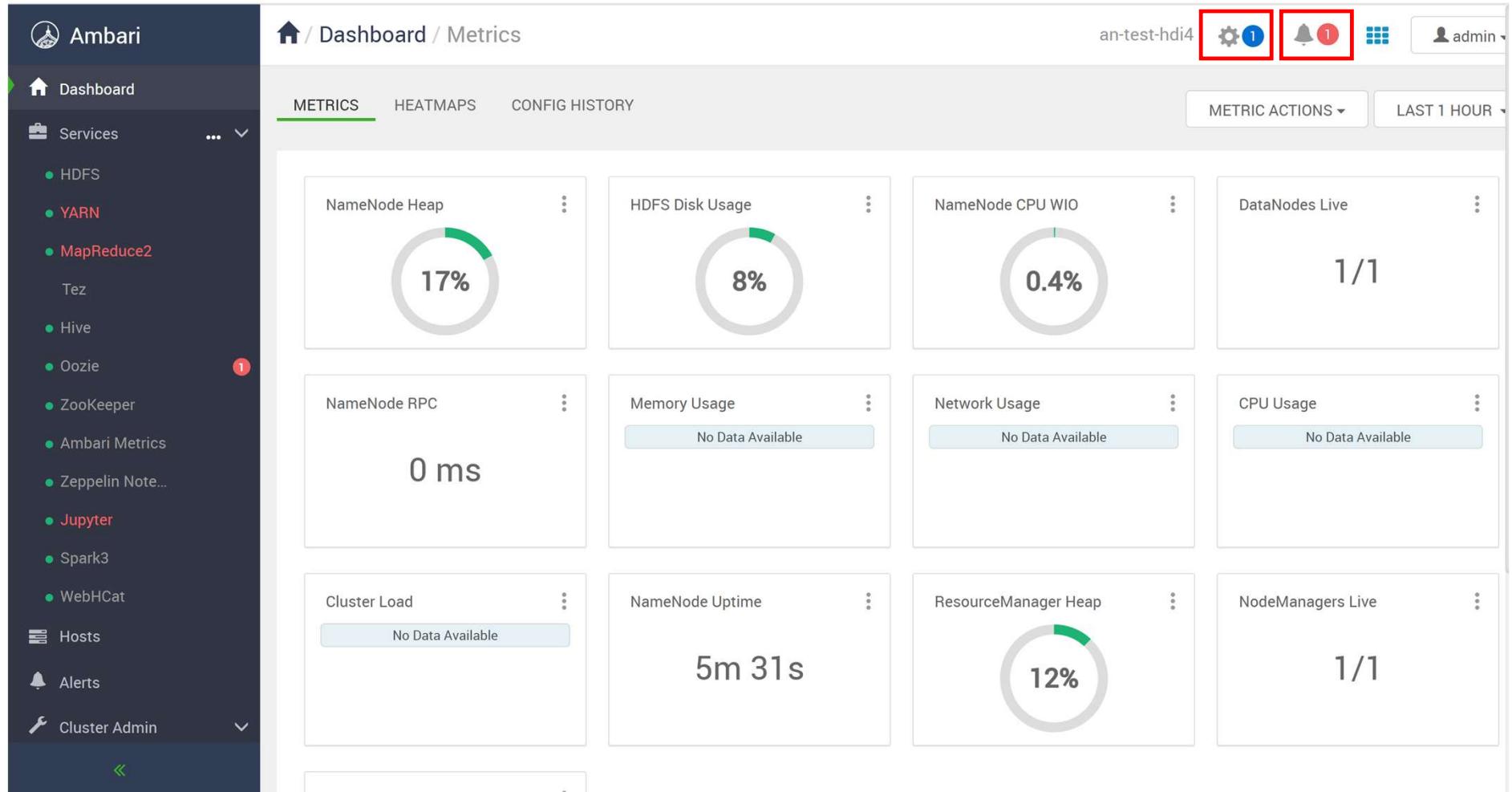
NameNode Uptime: 5m 31s

ResourceManager Heap: 12% 

NodeManagers Live: 1/1

Ambari Services: HDFS, YARN, MapReduce2, Tez, Hive, Oozie, ZooKeeper, Ambari Metrics, Zeppelin Note..., Jupyter, Spark3, WebHCat

Hosts, Alerts, Cluster Admin



# Ambari Scripts

The screenshot shows the Ambari Dashboard interface. At the top, the navigation bar includes 'Ambari', a home icon, 'Dashboard / Metrics', the host name 'an-test-hdi4', and user information 'admin'. A red box highlights the gear icon in the top right corner of the dashboard area. The main content area is titled 'Background Operations' and displays a list of 0 operations running. A dropdown menu shows 'ALL (10)'. Below this, a table lists seven completed operations with green status bars:

Operations	Status	User	Start Time	Duration
✓ Customizing host(s) with script: 'HeadNodeScriptsPostOperational.sh' from account hdiconfigactions.	100%	hdinsightwatchd@gday	15:21	1m 44s
✓ Stopping HA components on host hn1 from host hn1 : JUPYTER_MASTER	100%	hdinsightwatchd@gday	15:21	1s
✓ Starting HA components on host hn0 from host hn0 : HISTORYSERVER	100%	hdinsightwatchd@gday	15:20	6s
✓ Stopping HA components on host hn1 from host hn0 : JUPYTER_MASTER*	100%	hdinsightwatchd@gday	15:21	2s
✓ Update components on hosts	100%	hdinsightwatchd@gday	15:19	1m 38s
✓ Start components on host hn1-an-tes fihfl32hrenyczv02neic3dhb narv internal clouddann net	100%	internal	Today 15:15	4m 7s

At the bottom left, there is a checkbox for 'Do not show this dialog again when starting a background operation'. On the far right, a green 'OK' button is visible. The footer of the dashboard shows '2', '7m 18s', '23%', and '1/1'.

Exploring Ambari Services ...

Ambari > HDFS

Reminder ... No classical "HDFS"  
on HD-Insight !

=> using Azure Storage (AdlsGen2)  
+ Adapter class for Hadoop  
"VirtualFileSystem"

# Ambari > Services > HDFS

The screenshot shows the Ambari interface with the sidebar navigation open. The 'Services' section is selected, and 'HDFS' is highlighted with a red box. The main content area displays the 'Summary' tab for the HDFS service, showing various metrics and component status.

**Summary**

Components	Started ACTIVE NAMENODE	Started ZKFAILOVERCONTROLLER	Started STANDBY NAMENODE	Started ZKFAILOVERCONTROLLER
8m 55s NAMENODE UPTIME	11.1% 111.4 MB / 1004.0 MB NAMENODE HEAP	3/3 Live JOURNALNODES	0/0 Started NFSGATEWAYS	
DATANODES STATUS	1 Live	0 Dead	0 Decommissioning	
Service Metrics	BLOCKS			
n/a TOTAL FILES + DIRECTORIES	4 Total	0 Corrupt Replica	0 Missing	0 Under Replicated
	No pending upgrade UPGRADE STATUS	Not in safe mode SAFE MODE STATUS		
	0.00% 25.0 KB / 194.8 GB DISK USAGE (DFS USED)	8.22% 16.0 GB / 194.8 GB DISK USAGE (NON DFS USED)	91.78% 178.8 GB / 194.8 GB DISK REMAINING	

**Quick Links**

- hn1-an-tes.firebaseio.com (Active)
- NameNode UI
- NameNode Logs
- NameNode JMX
- Thread Stacks
- hn1-an-tes.firebaseio.com (Standby)
- NameNode UI
- NameNode Logs
- NameNode JMX
- Thread Stacks

# Ambari > HDFS > Configs

Ambari Services / HDFS / Configs

SUMMARY HEATMAPS CONFIGS METRICS

Version: 2

Config Group: Default (6)

SETTINGS ADVANCED

NameNode

NameNode directories: /hadoop/hdfs/namenode

NameNode Java heap size: 1GB (0 GB to 256 TB)

NameNode Server threads: 100 (1 to 200)

Minimum replicated blocks %: 0.01

DataNode

DataNode directories: /mnt/resource/hadoop/hdfs/data

DataNode failed disk tolerance: 0 (0 to 2)

DataNode maximum Java heap size: 1GB (0 GB to 256 TB)

DataNode max data transfer threads: 1024 (0 to 48000)

ACTIONS DISCARD SAVE

The screenshot shows the Ambari interface for managing HDFS configurations. The left sidebar lists various services, with 'HDFS' selected and highlighted by a red box. The main content area is titled 'Services / HDFS / Configs'. It displays two sections: 'NameNode' and 'DataNode', each with several configuration parameters. The 'CONFIGS' tab is active and highlighted with a red box. The 'ADVANCED' tab is also visible. The 'DISCARD' and 'SAVE' buttons are located at the bottom right.

# Ambari > HDFS > Configs > Advanced

The screenshot shows the Ambari interface for managing HDFS configurations. The left sidebar lists various services, with 'HDFS' selected and highlighted by a red box. The top navigation bar shows the current path: Services > HDFS > Configs. The main content area has four tabs: SUMMARY, HEATMAPS, CONFIGS (which is active and highlighted with a red box), and METRICS. Below these are two sub-tabs: SETTINGS and ADVANCED (also highlighted with a red box). The ADVANCED tab displays configuration settings for NameNode and DataNode.

**NameNode**

- NameNode hosts: hn0-an-tes.firebaseio.com and 1 other
- NameNode new generation size: 200 MB
- NameNode maximum new generation size: 200 MB
- NameNode permanent generation size: 128 MB
- NameNode maximum permanent generation size: 256 MB

**DataNode**

- DataNode host: wn1-an-tes.firebaseio.com
- DataNode directories permission: 750

**General**

DISCARD SAVE

# Ambari > HDFS > Configs > Advanced fs.defaultFS in core-site.xml

The screenshot shows the Ambari interface for managing HDFS configurations. The left sidebar lists various services like Dashboard, Services, HDFS, YARN, MapReduce2, Tez, Hive, Oozie, ZooKeeper, Ambari Metrics, Zeppelin Notebooks, Jupyter, Spark3, WebHCat, Hosts, Alerts, Cluster Admin, Stack and Versions, Service Accounts, and Service Auto Start. The 'HDFS' service is selected.

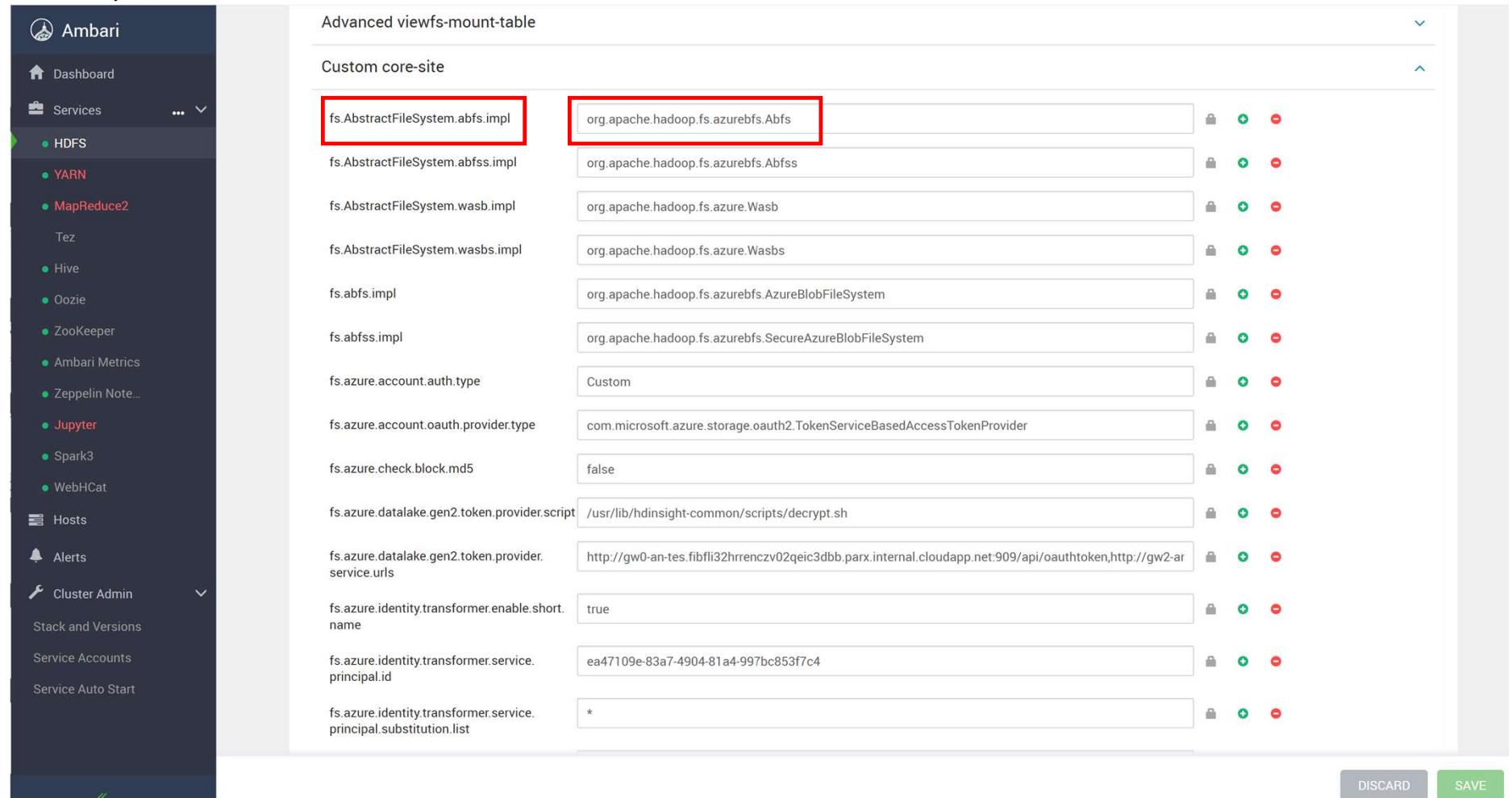
The main panel displays the 'Advanced core-site' configuration section. It includes fields for NFSGateway host (No host assigned), NFSGateway maximum Java heap size (1024 MB), NFSGateway dump directory (/tmp/hdfs-nfs), and Allowed hosts (\* rw). Below this, under 'Advanced core-site', there are several configuration properties:

- fs.azure.user.agent.prefix: APN/1.0 Azure HDInsight/5.1.3000.0.2409240625 MDH-Spark
- fs.defaultFS**: abfss://an-test-hdi4-2024-11-11t14-01-25-666z@teststoragehdifrance.dfs.core.windows.net (This field is highlighted with a red box)
- fs.gs.application.name.suffix: (GPN: Azure; version 1.0) HDInsight/5.1.3000.0.2409240625 Spark
- fs.gs.path.encoding: uri-path
- fs.gs.working.dir: /
- fs.s3a.fast.upload: true
- fs.s3a.fast.upload.buffer: disk
- fs.s3a.multipart.size: 67108864
- fs.s3a.user.agent.prefix: APN/1.0 Azure HDInsight/5.1.3000.0.2409240625 MDH-Spark
- fs.trash.interval: 0
- ha.failover-controller.active-standby-elector: zk.op.retries: 120

At the bottom right are 'DISCARD' and 'SAVE' buttons.

# Ambari > HDFS > Configs > Advanced custom hdfs-site.xml

Hook for support Azure Storage  
AdlsGen2 (abfs://...)  
in Hadoop



The screenshot shows the Ambari interface for managing HDFS configurations. The left sidebar lists various services like Dashboard, Services, HDFS, YARN, MapReduce2, Tez, Hive, Oozie, ZooKeeper, Ambari Metrics, Zeppelin Notebooks, Jupyter, Spark3, and WebHCat. The 'HDFS' service is selected. The main content area is titled 'Advanced viewfs-mount-table' under the 'Custom core-site' section. It displays a list of configuration properties with their current values. Two specific entries are highlighted with red boxes: 'fs.AbstractFileSystem.abfs.impl' and 'org.apache.hadoop.fs.azurebfs.Abfs'. Both of these have a value of 'org.apache.hadoop.fs.azurebfs.Abfs'. Other visible properties include fs.AbstractFileSystem.abfss.impl, fs.AbstractFileSystem.wasb.impl, fs.AbstractFileSystem.wasbs.impl, fs.abfs.impl, fs.abfss.impl, fs.azure.account.auth.type, fs.azure.account.oauth.provider.type, fs.azure.check.block.md5, fs.azure.datalake.gen2.token.provider.script, fs.azure.datalake.gen2.token.provider.service.urls, fs.azure.identity.transformer.enable.short.name, fs.azure.identity.transformer.service.principal.id, and fs.azure.identity.transformer.service.principal.substitution.list. At the bottom right are 'DISCARD' and 'SAVE' buttons.

Property	Value	Actions			
fs.AbstractFileSystem.abfs.impl	org.apache.hadoop.fs.azurebfs.Abfs				
fs.AbstractFileSystem.abfss.impl	org.apache.hadoop.fs.azurebfs.Abfss				
fs.AbstractFileSystem.wasb.impl	org.apache.hadoop.fs.azure.Wasb				
fs.AbstractFileSystem.wasbs.impl	org.apache.hadoop.fs.azure.Wasbs				
fs.abfs.impl	org.apache.hadoop.fs.azurebfs.AzureBlobFileSystem				
fs.abfss.impl	org.apache.hadoop.fs.azurebfs.SecureAzureBlobFileSystem				
fs.azure.account.auth.type	Custom				
fs.azure.account.oauth.provider.type	com.microsoft.azure.storage.oauth2.TokenServiceBasedAccessTokenProvider				
fs.azure.check.block.md5	false				
fs.azure.datalake.gen2.token.provider.script	/usr/lib/hdinsight-common/scripts/decrypt.sh				
fs.azure.datalake.gen2.token.provider.service.urls	http://gw0-an-tes.firebaseio.com/02qeic3dbb.firebaseio.cloudapp.net:909/api/oauth/token,http://gw2-ar				
fs.azure.identity.transformer.enable.short.name	true				
fs.azure.identity.transformer.service.principal.id	ea47109e-83a7-4904-81a4-997bc853f7c4				
fs.azure.identity.transformer.service.principal.substitution.list	*				

DISCARD SAVE

# Ambari > HDFS > Metrics

Ambari

Dashboard Services ...

HDFS YARN MapReduce2 Tez Hive Oozie ZooKeeper Ambari Metrics Zeppelin Notebooks Jupyter Spark3 WebHCat Hosts Alerts Cluster Admin Stack and Versions Service Accounts

an-test-hdi4 0 0 admin

Home / Services / HDFS / Metrics

SUMMARY HEATMAPS CONFIGS METRICS ACTIONS LAST 1 HOUR

Metrics

NameNode GC count NameNode GC time NN Connection Load NameNode Heap

1 0.5 2 ms 4 2 1000 MB 500 MB

NameNode Host Load NameNode RPC Failed disk volumes Blocks With Corrupted Replicas

50 % 2 ms 0 0

Under Replicated Blocks HDFS Space Utilization

0 n/a

ACTIONS LAST 1 HOUR

Metrics

NameNode GC count

1  
0.5

NameNode GC time

2 ms

NN Connection Load

4  
2

NameNode Heap

1000 MB  
500 MB

NameNode Host Load

50 %

NameNode RPC

2 ms

Failed disk volumes

0

Blocks With Corrupted Replicas

0

Under Replicated Blocks

0

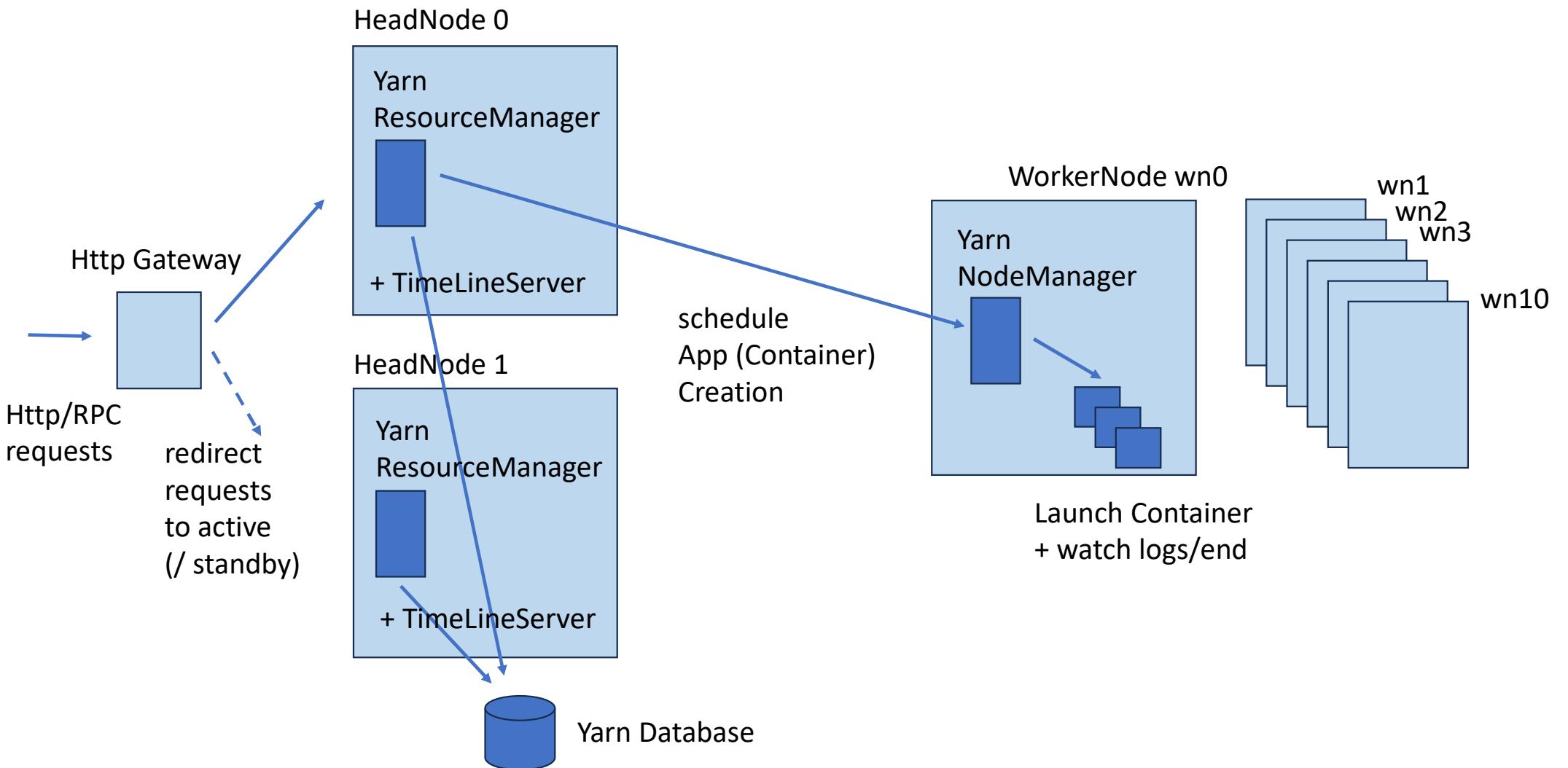
HDFS Space Utilization

n/a

ACTIONS LAST 1 HOUR

Ambari > Yarn

# Reminder Yarn Architecture



# Ambari > Yarn > Summary

Ambari

Dashboard

Services

HDFS

**YARN**

MapReduce2

Tez

Hive

Oozie

ZooKeeper

Ambari Metrics

Zeppelin Note...

Jupyter

Spark3

WebHCat

Hosts

Alerts

Cluster Admin

Stack and Versions

Service Accounts

Service Auto Start

◀

/ Services / YARN / Summary

SUMMARY HEATMAPS CONFIGS METRICS ACTIONS

an-test-hdi4 admin

**Summary**

Components

Started: TIMELINE SERVICE V1.5

Stopped: TIMELINE SERVICE V1.5

Started: STANDBY RESOURCEMANAGER

**Started: ACTIVE RESOURCEMANAGER**

1/1 Started NODEMANAGERS

2 Installed YARN CLIENTS

21m 38s RESOURCEMANAGER UPTIME

NODEMANAGERS STATUS

Active	Lost	Unhealthy	Rebooted
1	0	0	0

0 Decommissioned

212.7 MB / 910.5 MB RESOURCEMANAGER HEAP

Service Metrics

CONTAINERS

Allocated	Pending	Reserved
2	0	0

APPLICATIONS

3	2	0	1
---	---	---	---

Quick Links

hn0-an-tes.firebaseio.com (Standby)

ResourceManager UI

ResourceManager logs

ResourceManager JMX

Thread Stacks

hn1-an-tes.firebaseio.com (Active)

ResourceManager UI

ResourceManager logs

ResourceManager JMX

Thread Stacks

# Click on Quick Links "Resource Manager UI"

The screenshot shows the Ambari interface for the YARN service. The left sidebar lists various services like HDFS, YARN, and MapReduce2. The main content area displays the YARN Summary page with sections for Components, Node Managers Status, and Service Metrics. On the right side, there is a 'Quick Links' sidebar. The 'ResourceManager UI' link in this sidebar is highlighted with a red box. The URL for this link is <http://hn1-an-tes.firebaseio.com:8088>.

Ambari

Dashboard

Services

- HDFS
- YARN**
- MapReduce2
- Tez
- Hive
- Oozie
- ZooKeeper
- Ambari Metrics
- Zeppelin Note...
- Jupyter
- Spark3
- WebHCat

Hosts

Alerts

Cluster Admin

Stack and Versions

Service Accounts

Service Auto Start

◀

/ Services / YARN / Summary

SUMMARY HEATMAPS CONFIGS METRICS ACTIONS

Summary

Components

Started	Stopped
TIMELINE SERVICE V1.5	TIMELINE SERVICE V1.5

1/1 Started	2 Installed
NODEMANAGERS	YARN CLIENTS

21m 38s RESOURCEMANAGER UPTIME

NODEMANAGERS STATUS

1 Active	0 Lost	0 Unhealthy	0 Rebooted
----------	--------	-------------	------------

0 Decommissioned

212.7 MB / 910.5 MB RESOURCEMANAGER HEAP

Service Metrics

CONTAINERS

2 Allocated	0 Pending	0 Reserved
-------------	-----------	------------

APPLICATIONS

3	2	0	1
---	---	---	---

Quick Links

hn1-an-tes.firebaseio.com:8088

ResourceManager UI (highlighted)

ResourceManager logs

ResourceManager JMX

Thread Stacks

hn1-an-tes.firebaseio.com:8088

ResourceManager UI

ResourceManager logs

ResourceManager JMX

Thread Stacks

# <<baseurl>> /yarnui (=Webapp for Yarn ResourceManager)

← → ⌂ an-test-hdi4.azurehdinsight.net/yarnui/hn/cluster

Q ⭐ A ⋮

 All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Used Resources
3	0	2	1	2	<memory:3 GB, vCores:2>

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes
1	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:512, vCores:1>	<memory:25600, vCores:7>

Show 20 entries

ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Runn
application_1731334645766_0003	root	SparkSQL::10.0.0.16	SPARK		default	0	Mon Nov 11 15:23:10 +0100 2024	Mon Nov 11 15:23:10 +0100 2024	Mon Nov 11 15:23:35 +0100 2024	FINISHED	SUCCEEDED	N/A
application_1731334645766_0002	spark	Thrift JDBC/ODBC Server	SPARK		thriftsvr	0	Mon Nov 11 15:19:39 +0100 2024	Mon Nov 11 15:19:39 +0100 2024	N/A	RUNNING	UNDEFINED	1
application_1731334645766_0001	spark	Thrift JDBC/ODBC Server	SPARK		thriftsvr	0	Mon Nov 11 15:18:49 +0100 2024	Mon Nov 11 15:18:51 +0100 2024	N/A	RUNNING	UNDEFINED	1

Showing 1 to 3 of 3 entries

# Click on 1 Application (Stil Running OR Finished)

← → ⌂ an-test-hdi4.azurehdinsight.net/yarnui/hn/cluster

 All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Used Resources	
3	0	2	1	2	<memory:3 GB, vCores:2>	<memory:25

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes
1	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:512, vCores:1>	<memory:25600, vCores:7>

Show 20 entries

ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Runn Contai
application_1731334645766_0003	root	SparkSQL::10.0.0.16	SPARK		default	0	Mon Nov 11 15:23:10 +0100 2024	Mon Nov 11 15:23:10 +0100 2024	Mon Nov 11 15:23:35 +0100 2024	FINISHED	SUCCEEDED	N/A
application_1731334645766_0002	spark	Thrift JDBC/ODBC Server	SPARK		thriftsvr	0	Mon Nov 11 15:19:39 +0100 2024	Mon Nov 11 15:19:39 +0100 2024	N/A	RUNNING	UNDEFINED	1
application_1731334645766_0001	spark	Thrift JDBC/ODBC Server	SPARK		thriftsvr	0	Mon Nov 11 15:18:49 +0100 2024	Mon Nov 11 15:18:51 +0100 2024	N/A	RUNNING	UNDEFINED	1

Showing 1 to 3 of 3 entries

# /yarnui/hn/cluster/app/application\_<<id>>

example: a FINISHED Spark Application

Logged in as: dr.who

 Application application\_1731334645766\_0003

Application Overview

Cluster	User: root
About	Name: SparkSQL::10.0.0.16
Nodes	Application Type: SPARK
Node Labels	Application Tags:
Applications	Application Priority: 0 (Higher Integer value indicates higher priority)
NEW	YarnApplicationState: FINISHED
NEW SAVING	Queue: default
SUBMITTED	FinalStatus Reported by AM: SUCCEEDED
ACCEPTED	Started: Mon Nov 11 14:23:10 +0000 2024
RUNNING	Launched: Mon Nov 11 14:23:10 +0000 2024
FINISHED	Finished: Mon Nov 11 14:23:35 +0000 2024
FAILED	Elapsed: 24sec
KILLED	Tracking URL: http://hn1-an-tes.firebaseio.com/rsrcv02qeic3dbb.parx.internal.cloudapp.net:8088/proxy/application_1731334645766_0003/
Scheduler	Log Aggregation Status: SUCCEEDED
Tools	Application Timeout (Remaining Time): Unlimited
	Diagnostics:
	Unmanaged Application: false
	Application Node Label expression: <Not set>
	AM container Node Label expression: <DEFAULT_PARTITION>

Application Metrics

Total Resource Preempted: <memory:0, vCores:0>
Total Number of Non-AM Containers Preempted: 0
Total Number of AM Containers Preempted: 0
Resource Preempted from Current Attempt: <memory:0, vCores:0>
Number of Non-AM Containers Preempted from Current Attempt: 0
Aggregate Resource Allocation: 158966 MB-seconds, 60 vcore-seconds
Aggregate Preempted Resource Allocation: 0 MB-seconds, 0 vcore-seconds

Show 20 entries Search:

Attempt ID	Started	Node	Logs	Nodes blacklisted by the app	Nodes blacklisted by the system
appattempt_1731334645766_0003_000001	Mon Nov 11 15:23:10 +0100 2024	http://hn1-an-tes.firebaseio.com/rsrcv02qeic3dbb.parx.internal.cloudapp.net:30060	Logs 0	0	0

Showing 1 to 1 of 1 entries First Previous Next Last

# Yarn Application Internals : Attempt -> (AM) Master -> Containers

Admitedly : Yarn is over complex ... Think analogies with Kubernetes

Attempt => first try to launch, then maybe retries on failure

AM Master => idem "K8s Operator", but restricted to 1 Application

Container => mostly java apps (Docker images also supported after)

# Click on Application > Attempt "appattempt\_<<id>>\_000001"

Logged in as: dr.who



## Application application\_1731334645766\_0003

Cluster

- [About](#)
- [Nodes](#)
- [Node Labels](#)
- [Applications](#)
  - [NEW](#)
  - [NEW SAVING](#)
  - [SUBMITTED](#)
  - [ACCEPTED](#)
  - [RUNNING](#)
  - [FINISHED](#)
  - [FAILED](#)
  - [KILLED](#)

Scheduler

Tools

Application Overview

User:	root
Name:	SparkSQL::10.0.0.16
Application Type:	SPARK
Application Tags:	
Application Priority:	0 (Higher Integer value indicates higher priority)
YarnApplicationState:	FINISHED
Queue:	default
FinalStatus Reported by AM:	SUCCEEDED
Started:	Mon Nov 11 14:23:10 +0000 2024
Launched:	Mon Nov 11 14:23:10 +0000 2024
Finished:	Mon Nov 11 14:23:35 +0000 2024
Elapsed:	24sec
Tracking URL:	<a href="http://hn1-an-tes.firebaseio.com/rsrcv02qeic3dbb.parx.internal.cloudapp.net:8088/proxy/application_1731334645766_0003/">http://hn1-an-tes.firebaseio.com/rsrcv02qeic3dbb.parx.internal.cloudapp.net:8088/proxy/application_1731334645766_0003/</a>
Log Aggregation Status:	SUCCEEDED
Application Timeout (Remaining Time):	Unlimited
Diagnostics:	
Unmanaged Application:	false
Application Node Label expression:	<Not set>
AM container Node Label expression:	<DEFAULT_PARTITION>

Application Metrics

Total Resource Preempted:	<memory:0, vCores:0>
Total Number of Non-AM Containers Preempted:	0
Total Number of AM Containers Preempted:	0
Resource Preempted from Current Attempt:	<memory:0, vCores:0>
Number of Non-AM Containers Preempted from Current Attempt:	0
Aggregate Resource Allocation:	158966 MB-seconds, 60 vcore-seconds
Aggregate Preempted Resource Allocation:	0 MB-seconds, 0 vcore-seconds

Show 20 entries

Attempt ID	Started	Node	Logs	Nodes blacklisted by the app	Nodes blacklisted by the system
appattempt_1731334645766_0003_000001	Mon Nov 11 15:23:10 +0100 2024	<a href="http://hn1-an-tes.firebaseio.com/rsrcv02qeic3dbb.parx.internal.cloudapp.net:30060">http://hn1-an-tes.firebaseio.com/rsrcv02qeic3dbb.parx.internal.cloudapp.net:30060</a>	Logs	0	0

Showing 1 to 1 of 1 entries

First Previous  Next Last

# Application Attempt

Logged in as: dr.who

 Application Attempt appattempt\_1731334645766\_0003\_000001

Application Attempt Overview

Application Attempt State: FINISHED	
Started: Mon Nov 11 14:23:10 +0000 2024	
Elapsed: 24sec	
AM Container: container_1731334645766_0003_01_000001	
Node: N/A	
Tracking URL: <a href="#">History</a>	
Diagnostics Info:	
Nodes blacklisted by the application: -	
Nodes blacklisted by the system: -	

Total Allocated Containers: 3  
Each table cell represents the number of NodeLocal/RackLocal/OffSwitch containers satisfied by NodeLocal/RackLocal/OffSwitch resource requests.

	Node Local Request	Rack Local Request	Off Switch Request
Num Node Local Containers (satisfied by)	0		
Num Rack Local Containers (satisfied by)	0	0	
Num Off Switch Containers (satisfied by)	0	0	3

Show 20 entries Search:

Container ID	Node	Container Exit Status	Logs
No data available in table			

Showing 0 to 0 of 0 entries First Previous Next Last

# Application Attempt (finished) => click on "History"

Logged in as: dr.who

 Application Attempt appattempt\_1731334645766\_0003\_000001 Application Attempt Overview

Cluster

- About
- Nodes
- Node Labels
- Applications
- NEW
- NEW\_SAVING
- SUBMITTED
- ACCEPTED
- RUNNING
- FINISHED
- FAILED
- KILLED

Scheduler

Tools

Application Attempt State: FINISHED  
Started: Mon Nov 11 14:23:10 +0000 2024  
Elapsed: 24sec  
AM Container: container\_1731334645766\_0003\_01\_000001  
Node: N/A  
Tracking URL: [History](#) History  
Diagnostics Info:  
Nodes blacklisted by the application: -  
Nodes blacklisted by the system: -

Total Allocated Containers: 3  
Each table cell represents the number of NodeLocal/RackLocal/OffSwitch containers satisfied by NodeLocal/RackLocal/OffSwitch resource requests.

	Node Local Request	Rack Local Request	Off Switch Request
Num Node Local Containers (satisfied by)	0		
Num Rack Local Containers (satisfied by)	0	0	
Num Off Switch Containers (satisfied by)	0	0	3

Show 20 entries Search:

Container ID	Node	Container Exit Status	Logs
No data available in table			

Showing 0 to 0 of 0 entries First Previous Next Last

# equivalent Direct Link via "Tracking UI" column



Logged in as: dr.wh...

**All Applications**

Cluster Metrics		Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Used Resources	Total Resources	Reserved Resources	Physical Mem Used %	Physical Vcores Used %									
Cluster Nodes Metrics		3	0	2	1	<memory:3 GB, vCores:2>	<memory:25 GB, vCores:7>	<memory:0 B, vCores:0>	13	0										
Scheduler Metrics		Active Nodes	Decommissioning Nodes	0	Decommissioned Nodes	0	Lost Nodes	0	Unhealthy Nodes	0	Rebooted Nodes	0	Shutdown Nodes							
Scheduler Type		Scheduling Resource Type	Capacity Scheduler	[memory-mb (unit=Mi), vcores]	Minimum Allocation	<memory:512, vCores:1>	Maximum Allocation	<memory:25600, vCores:7>	0	Maximum Cluster Application Priority	Scheduler Busy %									
application_1731334645766_0003	root	SparkSQL::10.0.0.16	SPARK	default	0	Mon Nov 11 15:23:10 +0100 2024	Mon Nov 11 15:23:10 +0100 2024	Mon Nov 11 15:23:35 +0100 2024	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A	N/A	0.0	0.0	History	0	
application_1731334645766_0002	spark	Thrift JDBC/ODBC Server	SPARK	thriftsvr	0	Mon Nov 11 15:19:39 +0100 2024	Mon Nov 11 15:19:39 +0100 2024	N/A	RUNNING	UNDEFINED	1	1	1536	-1	0	-1	12.0	6.0	ApplicationMaster	0
application_1731334645766_0001	spark	Thrift JDBC/ODBC Server	SPARK	thriftsvr	0	Mon Nov 11 15:18:49 +0100 2024	Mon Nov 11 15:18:51 +0100 2024	N/A	RUNNING	UNDEFINED	1	1	1536	-1	0	-1	12.0	6.0	ApplicationMaster	0

Showing 1 to 3 of 3 entries

First Previous 1 Next Last

**Scheduler Busy %**

0

Search:

of	Progress	Tracking UI	Blacklisted
ster	History	0	Nodes
	ApplicationMaster	0	
	ApplicationMaster	0	

First Previous 1 Next Last

# Application History... for type Spark = SparkUI

url: /sparkhistory/application\_<<id>>

The screenshot shows the Apache Spark UI interface. At the top, there is a browser-style header with back, forward, and refresh buttons, a search bar containing the URL 'an-test-hdi4.azurehdinsight.net/sparkhistory/history/application\_1731334645766\_0003/jobs/' highlighted with a red box, and other icons. Below the header is the main navigation bar with tabs: Jobs (selected), Stages, Storage, Environment, Executors, SQL / DataFrame, Data (Preview), Graph (Preview), and Diagnostic (Preview). To the right of the navigation bar, it says 'SparkSQL::10.0.0.16 application UI'. The main content area is titled 'Spark Jobs' with a question mark icon. It displays user information: 'User: root', 'Total Uptime: 30 s', and 'Scheduling Mode: FIFO'. There is also a link to 'Event Timeline'.

## Spark Jobs (?)

User: root

Total Uptime: 30 s

Scheduling Mode: FIFO

▶ Event Timeline

# Click on Application > Attempt > "logs"

Logged in as: dr.who

 **Application application\_1731334645766\_0003**

**Application Overview**

User:	root
Name:	SparkSQL::10.0.0.16
Application Type:	SPARK
Application Tags:	
Application Priority:	0 (Higher Integer value indicates higher priority)
YarnApplicationState:	FINISHED
Queue:	default
FinalStatus Reported by AM:	SUCCEEDED
Started:	Mon Nov 11 14:23:10 +0000 2024
Launched:	Mon Nov 11 14:23:10 +0000 2024
Finished:	Mon Nov 11 14:23:35 +0000 2024
Elapsed:	24sec
Tracking URL:	<a href="http://hn1-an-tes.firebaseio.com/rsrcv02qeic3dbb.parx.internal.cloudapp.net:8088/proxy/application_1731334645766_0003/">http://hn1-an-tes.firebaseio.com/rsrcv02qeic3dbb.parx.internal.cloudapp.net:8088/proxy/application_1731334645766_0003/</a>
Log Aggregation Status:	SUCCEEDED
Application Timeout (Remaining Time):	Unlimited
Diagnostics:	
Unmanaged Application:	false
Application Node Label expression:	<Not set>
AM container Node Label expression:	<DEFAULT_PARTITION>

**Application Metrics**

Total Resource Preempted:	<memory:0, vCores:0>
Total Number of Non-AM Containers Preempted:	0
Total Number of AM Containers Preempted:	0
Resource Preempted from Current Attempt:	<memory:0, vCores:0>
Number of Non-AM Containers Preempted from Current Attempt:	0
Aggregate Resource Allocation:	158966 MB-seconds, 60 vcore-seconds
Aggregate Preempted Resource Allocation:	0 MB-seconds, 0 vcore-seconds

Show 20 entries Search:

Attempt ID	Started	Node	Logs	Nodes blacklisted by the app	Nodes blacklisted by the system
<a href="#">appattempt_1731334645766_0003_000001</a>	Mon Nov 11 15:23:10 +0100 2024	<a href="http://hn1-an-tes.firebaseio.com/rsrcv02qeic3dbb.parx.internal.cloudapp.net:30060">http://hn1-an-tes.firebaseio.com/rsrcv02qeic3dbb.parx.internal.cloudapp.net:30060</a>	<a href="#">Logs</a>	0	0

Showing 1 to 1 of 1 entries First Previous  Next Last

# Yarn logs

# logs ... summary "stdout", "stderr"

an-test-hdi4.azurehdinsight.net/yarnui/jobhistory/logs/wn1-an-tes.fibfl32hrrenczv02qeic3dbb.pax.internal.cloudapp.net/port/30050/container\_1731334645766\_0003\_0...  

```
find . -maxdepth 5 -ls 1>> /mnt/resource/hadoop/yarn/log/application_1731334645766_0003/container_1731334645766_0003_01_000001/directory.info || :  
echo "broken symlinks(find -L . -maxdepth 5 -type l -ls):" 1>>/mnt/resource/hadoop/yarn/log/application_1731334645766_0003/container_1731334645766_0003_01_000001/directory.info || :  
find . -maxdepth 5 -type l -ls 1>>/mnt/resource/hadoop/yarn/log/application_1731334645766_0003/container_1731334645766_0003_01_000001/directory.info || :  
echo "Launching container"  
exec /bin/bash -c "$JAVA_HOME/bin/java -server -XX:+IgnoreUnrecognizedVMOptions --add-opens=java.base/java.lang=ALL-UNNAMED --add-opens=java.base/java.lang.invoke=ALL-UNNAMED --add-opens=java.base/java.lang.reflect=ALL-UNNAMED  
Log Type: prelaunch.err  
Log Upload Time: Mon Nov 11 14:23:36 +0000 2024  
Log Length: 0  
  
Log Type: prelaunch.out  
Log Upload Time: Mon Nov 11 14:23:36 +0000 2024  
Log Length: 100  
Setting up env variables  
Setting up job resources  
Copying debugging information  
Launching container  
  
Log Type: stderr  
Log Upload Time: Mon Nov 11 14:23:36 +0000 2024  
Log Length: 11303  
Showing 4096 bytes of 11303 total. Click here for the full log.  
/11 14:23:16 INFO YarnAllocator [main]: requested profile o doesn't exist, adding it  
24/11/11 14:23:16 INFO Configuration [main]: found resource resource-types.xml at file:/mnt/resource/hadoop/yarn/local/filecache/12/_spark_conf_.zip/_hadoop_conf__resource-types.xml  
24/11/11 14:23:16 INFO YarnAllocator [main]: Will request 2 executor container(s) for ResourceProfile Id: 0, each with 1 core(s) and 3456 MB memory. with custom resources: <memory:3456, vCores:1>  
24/11/11 14:23:16 INFO YarnAllocator [main]: Submitted 2 unlocated container requests.  
24/11/11 14:23:16 INFO ApplicationMaster [main]: Started progress reporter thread with (heartbeat : 1000, initial allocation : 200) intervals  
24/11/11 14:23:16 INFO YarnAllocator [Reporter]: Launching container container_1731334645766_0003_01_000002 on host wn1-an-tes.fibfl32hrrenczv02qeic3dbb.pax.internal.cloudapp.net for executor with ID 1 for ResourceProfile  
24/11/11 14:23:16 INFO YarnAllocator [Reporter]: Launching container container_1731334645766_0003_01_000003 on host wn1-an-tes.fibfl32hrrenczv02qeic3dbb.pax.internal.cloudapp.net for executor with ID 2 for ResourceProfile  
24/11/11 14:23:16 INFO YarnAllocator [Reporter]: Received 2 containers from YARN, launching executors on 2 of them.  
24/11/11 14:23:16 INFO ExecutorRunnable [ContainerLauncher-1]: Initializing service data for shuffle service using name 'spark_shuffle'  
24/11/11 14:23:16 INFO ExecutorRunnable [ContainerLauncher-0]: Initializing service data for shuffle service using name 'spark_shuffle'  
24/11/11 14:23:34 INFO YarnAllocator [dispatcher-event-loop-0]: Driver requested a total number of 0 executor(s), reset all resource profile ids  
24/11/11 14:23:34 INFO ApplicationMasterSAMEndpoint [dispatcher-event-loop-1]: Driver terminated or disconnected! Shutting down. hn0-an-tes.fibfl32hrrenczv02qeic3dbb.pax.internal.cloudapp.net:46825  
24/11/11 14:23:34 INFO ApplicationMasterSAMEndpoint [dispatcher-event-loop-0]: Driver terminated or disconnected! Shutting down. hn0-an-tes.fibfl32hrrenczv02qeic3dbb.pax.internal.cloudapp.net:46825  
24/11/11 14:23:34 INFO ApplicationMaster [dispatcher-event-loop-0]: Final app status: SUCCEEDED, exitCode: 0  
24/11/11 14:23:34 INFO metrics [main]: type=GAUGE, name=application_1731334645766_0003.applicationMaster.numContainersPendingAllocate, value=0  
24/11/11 14:23:34 INFO metrics [main]: type=GAUGE, name=application_1731334645766_0003.applicationMaster.numExecutorsFailed, value=0  
24/11/11 14:23:34 INFO metrics [main]: type=GAUGE, name=application_1731334645766_0003.applicationMaster.numExecutorsRunning, value=2  
24/11/11 14:23:34 INFO metrics [main]: type=GAUGE, name=application_1731334645766_0003.applicationMaster.numLocalityAwareTasks, value=0  
24/11/11 14:23:34 INFO metrics [main]: type=GAUGE, name=application_1731334645766_0003.applicationMaster.numReleasedContainers, value=0  
24/11/11 14:23:34 INFO metrics [main]: type=GAUGE, name=application_1731334645766_0003.applicationMaster.numContainersPendingAllocate, value=0  
24/11/11 14:23:34 INFO metrics [main]: type=GAUGE, name=application_1731334645766_0003.applicationMaster.numExecutorsFailed, value=0  
24/11/11 14:23:34 INFO metrics [main]: type=GAUGE, name=application_1731334645766_0003.applicationMaster.numExecutorsRunning, value=2  
24/11/11 14:23:34 INFO metrics [main]: type=GAUGE, name=application_1731334645766_0003.applicationMaster.numLocalityAwareTasks, value=0  
24/11/11 14:23:34 INFO metrics [main]: type=GAUGE, name=application_1731334645766_0003.applicationMaster.numReleasedContainers, value=0  
24/11/11 14:23:34 INFO ApplicationMaster [shutdown-hook-0]: Unregistering ApplicationMaster with SUCCEEDED  
24/11/11 14:23:35 INFO AMRMClientImpl [shutdown-hook-0]: Waiting for application to be successfully unregistered.  
24/11/11 14:23:35 INFO ApplicationMaster [shutdown-hook-0]: Deleting staging directory abfss://an-test-hdi4-2024-11-11t14-01-25-666z@teststoragehdifrance.dfs.core.windows.net/user/root/.sparkStaging/application_173133464576  
24/11/11 14:23:35 INFO ShutdownHookManager [shutdown-hook-0]: Shutdown hook called  
  
Log Type: stdout  
Log Upload Time: Mon Nov 11 14:23:36 +0000 2024  
Log Length: 0
```

# Click on stderr details "here"

← → ⌂ an-test-hdi4.azurehdinsight.net/yarnui/jobhistory/logs/wn1-an-tes.fibfli32hrrenczv02qeic3dbb.pax.internal.cloudapp.net/port/30050/container\_1731334645766\_0003\_0.



Application

Log Type: stderr  
Log Upload Time: Mon Nov 11 14:23:36 +0000 2024  
Log Length: 11303

SLF4J: Class path contains multiple SLF4J bindings.  
SLF4J: Found binding in [jar:file:/usr/hdp/5.1.6.7/spark3/jars/spark-streaming-kafka-0-10-assembly\_2.12-3.3.1.5.1.6.7.jar!/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: Found binding in [jar:file:/usr/hdp/5.1.6.7/spark3/jars/log4j-slf4j-impl-2.17.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: Found binding in [jar:file:/usr/hdp/5.1.6.7/hive\_warehouse\_connector/hive-warehouse-connector-assembly-2.1.0.5.1.6.7.jar!/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: See http://www.slf4j.org/codes.html#multiple\_bindings for an explanation.  
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]  
24/11/11 14:23:13 INFO SignalUtils [main]: Registering signal handler for TERM  
24/11/11 14:23:13 INFO SignalUtils [main]: Registering signal handler for HUP  
24/11/11 14:23:13 INFO SignalUtils [main]: Registering signal handler for INT  
24/11/11 14:23:13 WARN SparkConf [main]: The configuration key 'spark.yarn.access.hadoopFileSystems' has been deprecated as of Spark 3.0 and may be removed in the future. Please use the new key 'spark.kerberos.principal'  
24/11/11 14:23:13 INFO SecurityManager [main]: Changing view acls to: yarn,root  
24/11/11 14:23:13 INFO SecurityManager [main]: Changing modify acls to: yarn,root  
24/11/11 14:23:13 INFO SecurityManager [main]: Changing view acls groups to:  
24/11/11 14:23:13 INFO SecurityManager [main]: Changing modify acls groups to:  
24/11/11 14:23:13 INFO SecurityManager [main]: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(yarn, root); groups with view permissions: Set(); users with modify permissions: Set(); groups with modify permissions: Set()  
24/11/11 14:23:13 WARN NativeCodeLoader [main]: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
24/11/11 14:23:14 INFO ApplicationMaster [main]: ApplicationAttemptId: appattempt\_1731334645766\_0003\_00001  
24/11/11 14:23:14 WARN SparkConf [main]: The configuration key 'spark.yarn.access.hadoopFileSystems' has been deprecated as of Spark 3.0 and may be removed in the future. Please use the new key 'spark.kerberos.principal'  
24/11/11 14:23:14 INFO RequestHedgingRMFailoverProxyProvider [main]: Created wrapped proxy for [rm1, rm2]  
24/11/11 14:23:14 INFO YarnRMClient [main]: Registering the ApplicationMaster  
24/11/11 14:23:14 INFO RequestHedgingRMFailoverProxyProvider [main]: Looking for the active RM in [rm1, rm2]...  
24/11/11 14:23:14 INFO RequestHedgingRMFailoverProxyProvider [main]: Found active RM [rm2]  
24/11/11 14:23:14 INFO TransportClientFactory [netty-rpc-connection-0]: Successfully created connection to hn0-an-tes.fibfli32hrrenczv02qeic3dbb.pax.internal.cloudapp.net/10.0.0.16:46825 after 80 ms (0 ms)  
24/11/11 14:23:14 INFO ApplicationMaster [main]: Preparing Local resources  
24/11/11 14:23:15 INFO AbfsConfiguration [main]: AbfsClientRetryPolicy: default  
24/11/11 14:23:16 INFO ApplicationMaster [main]:  
=====Default YARN executor launch context:  
env:  
CLASSPATH -> {{PWD}}<CPS>{{PWD}}/\_spark\_conf\_<CPS>{{PWD}}/\_spark\_libs\_/\*<CPS>/usr/hdp/current/spark3-client/jars/\*<CPS>:/usr/hdp/current/spark3-client/jars/\*:/usr/lib/hdinsight-datalake/\*:/usr/hdp/SPARK\_DIST\_CLASSPATH -> :/usr/hdp/current/spark3-client/jars/\*:/usr/lib/hdinsight-datalake/\*:/usr/hdp/current/spark\_llap/\*:/usr/hdp/current/spark3-client/conf:/usr/hdp/current/hadoop-hdfs-client/lib/\*:  
SPARK\_YARN\_STAGING\_DIR -> abfs://an-test-hdi4-2024-11-11t14-01-25-666z@teststoragehdifrance.dfs.core.windows.net/user/root/.sparkStaging/application\_1731334645766\_0003  
SPARK\_USER -> root  
command:  
LD\_LIBRARY\_PATH=/usr/hdp/current/hadoop-client/lib/native:/usr/hdp/current/hadoop-client/lib/native/Linux-amd64-64:\$LD\_LIBRARY\_PATH\" \  
{JAVA\_HOME}/bin/java \  
-server \  
-Xmx3072m \  
'-XX:+IgnoreUnrecognizedVMOptions' \  
'--add-opens=java.base/java.lang=ALL-UNNAMED' \  
'--add-opens=java.base/java.lang.invoke=ALL-UNNAMED' \  
'--add-opens=java.base/java.lang.reflect=ALL-UNNAMED' \  
'--add-opens=java.base/java.io=ALL-UNNAMED' \  
'--add-opens=java.base/java.net=ALL-UNNAMED' \  
'--add-opens=java.base/java.nio=ALL-UNNAMED' \  
'--add-opens=java.base/java.nio.channels=ALL-UNNAMED'

# Using "Save As..." instead of click "here" for untruncated (10 Mega?)

Log Type: stderr  
Log Upload Time: Mon Nov 11 14:23:36 +0000 2024  
Log Length: 11303  
Showing 4096 bytes of 11303 total. Click here for the full log.

```
/11 14:23:16 INFO YarnAllocator [main]: Res...  
24/11/11 14:23:16 INFO Configuration [main]  
24/11/11 14:23:16 INFO YarnAllocator [main]  
24/11/11 14:23:16 INFO YarnAllocator [main]  
24/11/11 14:23:16 INFO ApplicationMaster [m...  
24/11/11 14:23:16 INFO YarnAllocator [Report...  
24/11/11 14:23:16 INFO YarnAllocator [Report...  
24/11/11 14:23:16 INFO YarnAllocator [Report...  
24/11/11 14:23:16 INFO ExecutorRunnable [Co...  
24/11/11 14:23:16 INFO ExecutorRunnable [Co...  
24/11/11 14:23:34 INFO YarnAllocator [dispa...  
24/11/11 14:23:34 INFO ApplicationMaster$AM...  
24/11/11 14:23:34 INFO ApplicationMaster$AM...  
24/11/11 14:23:34 INFO ApplicationMaster [d...  
24/11/11 14:23:34 INFO metrics [main]: type:  
24/11/11 14:23:34 INFO metrics [main]: type=
```

Open link in new tab  
Open link in new window  
Open link in incognito window  
Save link as...  
Copy link address  
Inspect

e:/mnt/resource/hadoop/yarn/local/filecache/12/\_spark\_conf\_.: ResourceProfile Id: 0, each with 1 core(s) and 3456 MB memory ts.  
h (heartbeat : 1000, initial allocation : 200) intervals  
4645766\_0003\_01\_000002 on host wn1-an-tes.fibfli32hrrenczv02qe:  
4645766\_0003\_01\_000003 on host wn1-an-tes.fibfli32hrrenczv02qe:  
nching executors on 2 of them.  
ata for shuffle service using name 'spark\_shuffle'  
ata for shuffle service using name 'spark\_shuffle'  
tal number of 0 executor(s), reset all resource profile ids  
terminated or disconnected! Shutting down. hn0-an-tes.fibfli3:  
terminated or disconnected! Shutting down. hn0-an-tes.fibfli3:  
SUCCEEDED, exitCode: 0  
3.applicationMaster.numContainersPendingAllocate, value=0  
3.applicationMaster.numExecutorsFailed, value=0  
3.applicationMaster.numExecutorsRunning, value=2  
3.applicationMaster.numLocalityAwareTasks, value=0  
3.applicationMaster.numReleasedContainers, value=0  
3.applicationMaster.numContainersPendingAllocate, value=0  
24/11/11 14:23:34 INFO metrics [main]: type=GAUGE, name=application\_1731334645766\_0003.applicationMaster.numExecutorsFailed, value=0

# Click on Application > Attempt > "node"

Logged in as: dr.who

**hadoop**

## Application application\_1731334645766\_0003

**Application Overview**

User:	root
Name:	SparkSQL::10.0.0.16
Application Type:	SPARK
Application Tags:	
Application Priority:	0 (Higher Integer value indicates higher priority)
YarnApplicationState:	FINISHED
Queue:	default
FinalStatus Reported by AM:	SUCCEEDED
Started:	Mon Nov 11 14:23:10 +0000 2024
Launched:	Mon Nov 11 14:23:10 +0000 2024
Finished:	Mon Nov 11 14:23:35 +0000 2024
Elapsed:	24sec
Tracking URL:	<a href="http://hn1-an-tes.firebaseio.com/02qeic3dbb.parx.internal.cloudapp.net:8088/proxy/application_1731334645766_0003/">http://hn1-an-tes.firebaseio.com/02qeic3dbb.parx.internal.cloudapp.net:8088/proxy/application_1731334645766_0003/</a>
Log Aggregation Status:	SUCCEEDED
Application Timeout (Remaining Time):	Unlimited
Diagnostics:	
Unmanaged Application:	false
Application Node Label expression:	<Not set>
AM container Node Label expression:	<DEFAULT_PARTITION>

**Application Metrics**

Total Resource Preempted:	<memory:0, vCores:0>
Total Number of Non-AM Containers Preempted:	0
Total Number of AM Containers Preempted:	0
Resource Preempted from Current Attempt:	<memory:0, vCores:0>
Number of Non-AM Containers Preempted from Current Attempt:	0
Aggregate Resource Allocation:	158966 MB-seconds, 60 vcore-seconds
Aggregate Preempted Resource Allocation:	0 MB-seconds, 0 vcore-seconds

Show 20 entries

Attempt ID	Started	Node	Logs	Nodes blacklisted by the app	Nodes blacklisted by the system
<a href="#">appattempt_1731334645766_0003_000001</a>	Mon Nov 11 15:23:10 +0100 2024	<a href="http://hn1-an-tes.firebaseio.com/02qeic3dbb.parx.internal.cloudapp.net:30060">http://hn1-an-tes.firebaseio.com/02qeic3dbb.parx.internal.cloudapp.net:30060</a>	Logs 0	0	0

Showing 1 to 1 of 1 entries

First Previous  Next Last

# Yarn NodeManager "wn1"

← → ⌂ an-test-hdi4.azurehdinsight.net/yarnui/wn1-an-tes.fibfli32hrrenczv02qeic3dbb.pax.internal.cloudapp.net/port/30060/node

Logged in as: dr.who



## NodeManager information

NodeManager information	
ResourceManager	
NodeManager	
Node Information	
List of Applications	
List of Containers	
Tools	
Total Vmem allocated for Containers	52.50 GB
Vmem enforcement enabled	false
Total Pmem allocated for Container	25 GB
Pmem enforcement enabled	false
Total VCores allocated for Containers	7
Resource types	memory-mb (unit=Mi), vcores
NodeHealthyStatus	true
LastNodeHealthTime	Mon Nov 11 15:10:08 UTC 2024
NodeHealthReport	
NodeManager started on	Mon Nov 11 14:16:01 UTC 2024
NodeManager Version:	3.3.4.5.1.6.7 from 494a08c4dd5f2680ed7a73b610bea0f9569b7dd5 by root source checksum e86c1be7db4cbdb06fec12e6ecd9827 on 2024-09-24T10:13Z
Hadoop Version:	3.3.4.5.1.6.7 from 494a08c4dd5f2680ed7a73b610bea0f9569b7dd5 by root source checksum de3c5cee9cc5a24aff9c3ca0474643 on 2024-09-24T10:11Z

# click on "Scheduler"

The screenshot shows the Hadoop Cluster Metrics interface. On the left, there is a sidebar with a yellow elephant icon and the word "hadoop". Below it is a navigation menu under "Cluster" with the following items:

- About
- Nodes
- Node Labels
- Applications
  - NEW
  - NEW\_SAVING
  - SUBMITTED
  - ACCEPTED
  - RUNNING
  - FINISHED
  - FAILED
  - KILLED
- Scheduler** (This item is highlighted with a red rectangular box.)
- Tools

The main content area displays three sections: "Cluster Metrics", "Cluster Nodes Metrics", and "Scheduler Metrics".

**Cluster Metrics**

Apps Submitted	Apps Pending
3	0

**Cluster Nodes Metrics**

Active Nodes	Standby Nodes
1	0

**Scheduler Metrics**

Scheduler Type: Capacity Scheduler [mem]

Show 20 entries

ID	User	Name
application_1731334645766_0003	root	SparkSQL:

# Yarn Scheduler Queue

**hadoop** Logged in as: dr.who

**NEW, NEW\_SAVING, SUBMITTED, ACCEPTED, RUNNING, FINISHED, FAILED, KILLED**

**Scheduler**

**Cluster Metrics**

Apps Submitted	0	Apps Pending	2	Apps Running	1	Apps Completed	2	Containers Running	<memory:3 GB, vCores:2>	Used Resources	<memory:25 GB, vCores:7>	Total Resources	<memory:0 B, vCores:0>	Reserved Resources	13	Physical Mem Used %	0	Physical Vcores Used %
----------------	---	--------------	---	--------------	---	----------------	---	--------------------	-------------------------	----------------	--------------------------	-----------------	------------------------	--------------------	----	---------------------	---	------------------------

**Cluster Nodes Metrics**

Active Nodes	1	Decommissioning Nodes	0	Decommissioned Nodes	0	Lost Nodes	0	Unhealthy Nodes	0	Rebooted Nodes	0	Shutdown Nodes	0
--------------	---	-----------------------	---	----------------------	---	------------	---	-----------------	---	----------------	---	----------------	---

**Scheduler Metrics**

Scheduler Type	Capacity Scheduler	Scheduling Resource Type	[memory-mb (unit=Mi), vcores]	Minimum Allocation	<memory:512, vCores:1>	Maximum Allocation	<memory:25600, vCores:7>	Maximum Cluster Application Priority	0	Scheduler Busy %	0
----------------	--------------------	--------------------------	-------------------------------	--------------------	------------------------	--------------------	--------------------------	--------------------------------------	---	------------------	---

**Dump scheduler logs** 1 min ▾

**Application Queues**

Legend: Capacity **Used** Used (over capacity) Max Capacity Users Requesting Resources Auto Created Queues

- Queue: root 12.0% used
- Queue: thriftsvr 24.0% used
- Queue: default 0.0% used

Show 20 entries Search: ▾

ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU Vcores	Allocated Memory MB	Allocated GPUs	Reserved CPU Vcores	Reserved Memory MB	Reserved GPUs	% of Queue	% of Cluster	Progress	Tracking UI	Blacklisted Nodes
application_1731334645766_0002	spark	Thrift JDBC/ODBC Server	SPARK	thriftsvr	0	N/A	Mon Nov 11 15:19:39 +0100 2024	Mon Nov 11 15:19:39 +0100 2024	N/A	RUNNING	UNDEFINED	1	1	1536	-1	0	0	-1	12.0	6.0	<input type="button" value="ApplicationMaster"/>	0	
application_1731334645766_0001	spark	Thrift JDBC/ODBC Server	SPARK	thriftsvr	0	N/A	Mon Nov 11 15:18:49 +0100 2024	Mon Nov 11 15:18:51 +0100 2024	N/A	RUNNING	UNDEFINED	1	1	1536	-1	0	0	-1	12.0	6.0	<input type="button" value="ApplicationMaster"/>	0	

Showing 1 to 2 of 2 entries First Previous  Next Last

**Aggregate scheduler counts**

Total Container Allocations(count)	5	Total Container Releases(count)	3	Total Fulfilled Reservations(count)	0	Total Container Preemptions(count)	0
------------------------------------	---	---------------------------------	---	-------------------------------------	---	------------------------------------	---

**Last scheduler run**

Mon Nov 11 15:27:31 +0000 2024	Time	Allocations(count - resources)	0 - <memory:0, vCores:0>	Reservations(count - resources)	0 - <memory:0, vCores:0>	Releases(count - resources)	0 - <memory:0, vCores:0>
--------------------------------	------	--------------------------------	--------------------------	---------------------------------	--------------------------	-----------------------------	--------------------------

**Last Preemption**

N/A	Time	Container Id	N/A	Node Id	N/A	Queue
-----	------	--------------	-----	---------	-----	-------

**Last Reservation**

N/A	Time	Container Id	N/A	Node Id	N/A	Queue
-----	------	--------------	-----	---------	-----	-------

**Last Allocation**

N/A	Time	Container Id	N/A	Node Id	N/A	Queue
-----	------	--------------	-----	---------	-----	-------

**Last Release**

Mon Nov 11 14:23:35 +0000 2024	Time	Container Id	wn1-an-tes.firebaseio.com:32288	Node Id	root.default	Queue
container_1731334645766_0003_01_000001						

# (Admin) Edit Yarn Queue

The screenshot shows the Ambari interface for managing YARN queues. On the left, the navigation bar includes links for Dashboard, Services (HDFS, YARN, MapReduce2), Tez, Hive, Oozie, ZooKeeper, Ambari Metrics, Zeppelin Notebooks, Jupyter, Spark3, WebHCat, Hosts, Alerts, Cluster Admin, Stack and Versions, and Service Accounts. The YARN link is highlighted.

The main page title is "YARN Queue Manager". It displays three queues: "root (100%)", "default (50%)", and "thriftsvr (50%)". A message says "Click on a queue to the left for details." Below this is a "Scheduler" configuration section with fields for Maximum Applications (10000), Maximum AM Resource (33 %), Node Locality Delay (0), Calculator (Default Resource Calculator), Queue Mappings, Queue Mappings Override (Disabled checked), and Additional Configs (yarn.scheduler.capacity.sched, yarn.scheduler.capacity.default, yarn.scheduler.capacity.sched, yarn.scheduler.capacity.sched).

A red box highlights the "Views" button in the top right corner, which is currently expanded to show "YARN Queue Manager" (highlighted with a red box), "Hive View 2.0", and "Tez View".

# filter "FINISHED" Applications

 **FINISHED Applications**

The screenshot shows the Hadoop Cluster Metrics interface. On the left, a sidebar menu lists cluster metrics: Apps Submitted (3), Apps Pending (0), Apps Running (2), Apps Completed (1), Containers Running (2), Used Resources (<memory:3 GB, vCores:2>), and Total Resources (<memory:25 GB, vCores:7>). Below these are sections for Cluster Nodes Metrics, Scheduler Metrics, and a detailed table of completed applications.

**Scheduler Metrics** table:

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:512, vCores:1>	<memory:25600, vCores:7>

**Applications** table:

ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers
application_1731334645766_0003	root	SparkSQL::10.0.0.16	SPARK		default	0	Mon Nov 11 15:23:10 +0100 2024	Mon Nov 11 15:23:10 +0100 2024	Mon Nov 11 15:23:35 +0100 2024	FINISHED	SUCCEEDED	N/A

A red box highlights the "FINISHED" status in both the Scheduler Metrics and Applications tables.

# filter "RUNNING" Applications

 **RUNNING Applications**

**Cluster Metrics**

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Used Resources	Total Resc
3	0	2	1	2	<memory:3 GB, vCores:2>	<memory:25 GB, vCores:7>

**Cluster Nodes Metrics**

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes
1	0	0	0

**Scheduler Metrics**

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:512, vCores:1>	<memory:25600, vCores:7>

Show 20 entries

ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCore
application_1731334645766_0002	spark	Thrift JDBC/ODBC Server	SPARK		thriftsvr	0	Mon Nov 11 15:19:39 +0100 2024	Mon Nov 11 15:19:39 +0100 2024	N/A	RUNNING	UNDEFINED	1	1
application_1731334645766_0001	spark	Thrift JDBC/ODBC Server	SPARK		thriftsvr	0	Mon Nov 11 15:18:49 +0100 2024	Mon Nov 11 15:18:51 +0100 2024	N/A	RUNNING	UNDEFINED	1	1

Showing 1 to 2 of 2 entries

# click a RUNNING Yarn Application

 **Application application\_1731334645766\_0002**

**Kill Application**  

User:	spark
Name:	Thrift JDBC/ODBC Server
Application Type:	SPARK
Application Tags:	
Application Priority:	0 (Higher Integer value indicates higher priority)
<b>YarnApplicationState:</b>	<b>RUNNING: AM has registered with RM and started running.</b>
Queue:	thriftsvr
FinalStatus Reported by AM:	Application has not completed yet.
Started:	Mon Nov 11 14:19:39 +0000 2024
Launched:	Mon Nov 11 14:19:39 +0000 2024
Finished:	N/A
Elapsed:	1hrs, 23mins, 37sec
Tracking URL:	<a href="http://hn1-an-tes.firebaseio.com/rsrcv02qeic3dbb.parx.internal.cloudapp.net:8088/proxy/application_1731334645766_0002/">http://hn1-an-tes.firebaseio.com/rsrcv02qeic3dbb.parx.internal.cloudapp.net:8088/proxy/application_1731334645766_0002/</a>
Log Aggregation Status:	NOT_START
Application Timeout (Remaining Time):	Unlimited
Diagnostics:	
Unmanaged Application:	false
Application Node Label expression:	<Not set>
AM container Node Label expression:	<DEFAULT_PARTITION>

Total Resource Preempted: <memory:0, vCores:0>  
Total Number of Non-AM Containers Preempted: 0  
Total Number of AM Containers Preempted: 0  
Resource Preempted from Current Attempt: <memory:0, vCores:0>  
Number of Non-AM Containers Preempted from Current Attempt: 0  
Aggregate Resource Allocation: 7705015 MB-seconds, 5016 vcore  
Aggregate Preempted Resource Allocation: 0 MB-seconds, 0 vcore-seconds

Show 20	entries			
Attempt ID	Started	Node	Logs	N
<a href="#">appattempt_1731334645766_0002_000001</a>	Mon Nov 11 15:19:39 +0100 2024	<a href="http://wn1-an-tes.firebaseio.com/rsrcv02qeic3dbb.parx.internal.cloudapp.net:30060">http://wn1-an-tes.firebaseio.com/rsrcv02qeic3dbb.parx.internal.cloudapp.net:30060</a>	Logs	0

# RUNNING Application > Attempt

 **Application Attempt appattempt\_1731334645766\_0002\_000001**

**Cluster**

- About
- Nodes
- Node Labels
- Applications
  - NEW
  - NEW SAVING
  - SUBMITTED
  - ACCEPTED
  - RUNNING
  - FINISHED
  - FAILED
  - KILLED
- Scheduler

**Tools**

Total Allocated Containers: 1  
Each table cell represents the number of NodeLocal/RackLocal/OffSwitch containers satisfied by NodeLocal/RackLocal/OffSwitch resource requests.

	Node Local Request	Rack Local Request
Num Node Local Containers (satisfied by)	0	
Num Rack Local Containers (satisfied by)	0	0
Num Off Switch Containers (satisfied by)	0	1

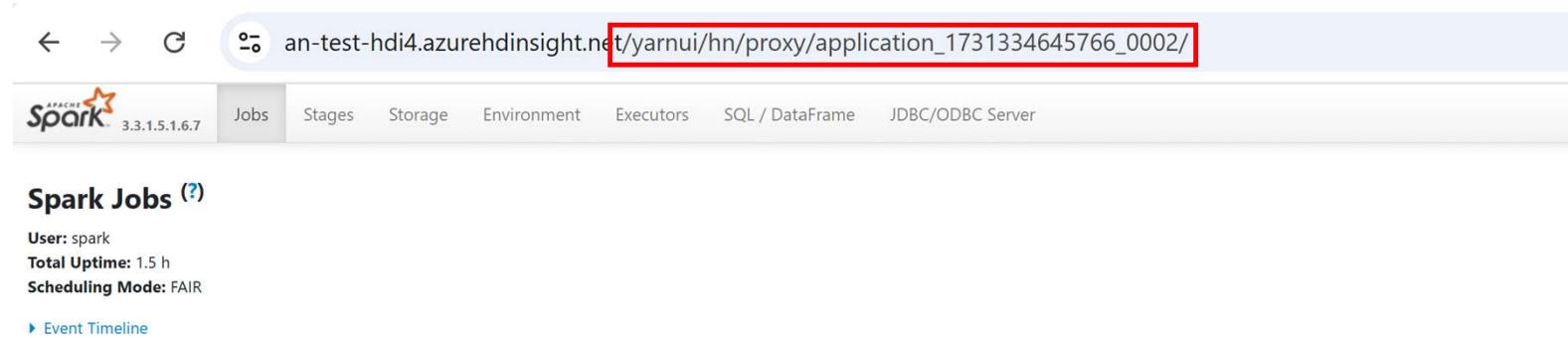
Show 20 ▾ entries

Container ID	Node	Cont
container_1731334645766_0002_01_000001	http://wn1-an-tes.firebaseio.com:30060	0

Showing 1 to 1 of 1 entries

when Applications are still running  
=> "/proxy" redirect to real process  
instead of "History" server

url: /yarnui/hn/proxy/application\_<<id>>



The screenshot shows the Apache Spark UI interface. At the top, there is a navigation bar with icons for back, forward, and search, followed by the URL 'an-test-hdi4.azurehdinsight.net/yarnui/hn/proxy/application\_1731334645766\_0002/'. Below the URL is a menu bar with tabs: Jobs (which is selected), Stages, Storage, Environment, Executors, SQL / DataFrame, and JDBC/ODBC Server. The main content area is titled 'Spark Jobs' with a question mark icon. It displays the following information: User: spark, Total Uptime: 1.5 h, and Scheduling Mode: FAIR. There is also a link to 'Event Timeline'.

# RUNNING Application > AM Master > Container

 **Container container\_1731334645766\_0002\_01\_000001**

<b>Cluster</b>	Container State: RUNNING
About	Exit Status: 0
Nodes	Node: http://wn1-an-tes.firebaseio.com:30060
<u>Node Labels</u>	Priority: 0
Applications	Started: Mon Nov 11 14:19:39 +0000 2024
NEW	Elapsed: 1hrs, 28mins, 21sec
NEW_SAVING	Resource: 1536 Memory, 1 Vcores
SUBMITTED	Logs: <a href="#">Logs</a>
ACCEPTED	Diagnostics:
RUNNING	
FINISHED	
FAILED	
KILLED	
Scheduler	
Tools	

# Container > logs



- ▼ ResourceManager
  - [RM Home](#)
- ▶ NodeManager
- ▶ Tools

## Local Logs:

```
directory.info : Total file length is 6504 bytes.  
launch_container.sh : Total file length is 5742 bytes.  
prelaunch.err : Total file length is 0 bytes.  
prelaunch.out : Total file length is 100 bytes.  
stderr : Total file length is 8283 bytes.  
stdout : Total file length is 0 bytes.
```

## Logs for container\_1731334645766\_0002\_01\_000001

UI Experience for browsing Log in Yarn ...

( and not Search by Text or Regexp,  
no selection by Time Range  
... nothing)



Ambari > Oozie

# Ambari > Oozie > Summary

The screenshot shows the Ambari web interface with the following details:

- Left Sidebar (Services):** Includes Dashboard, Services (HDFS, YARN, MapReduce2, Tez, Hive, Oozie), ZooKeeper, Ambari Metrics, Zeppelin Note..., Jupyter, Spark3, WebHCat, Hosts, and Alerts. The "Oozie" service is highlighted with a red box.
- Top Bar:** Shows the Ambari logo, user "admin", and a green "ACTIONS" button.
- Page Title:** / Services / Oozie / Summary
- Summary Tab:** The "SUMMARY" tab is selected and highlighted with a red box. The "CONFIGS" tab is also present.
- Summary Section:** Displays the status of components:
  - Components: Components, Status, Count
  - Started: OOZIE SERVER (2)
  - Installed: OOZIE CLIENTS (3)
- Quick Links:** Includes links to Oozie Web UI for two hosts: hn0-an-tes.firebaseio.com (Active) and hn1-an-tes.firebaseio.com (Active).

# Ambari > Oozie > Quick Link "Oozie Web UI"

The screenshot shows the Ambari web interface with the following details:

- Left Sidebar:** Shows the navigation menu with various services listed: Dashboard, Services (HDFS, YARN, MapReduce2, Tez, Hive, Oozie), ZooKeeper, Ambari Metrics, Zeppelin Note..., Jupyter, Spark3, WebHCat, Hosts, and Alerts.
- Header:** Displays the Ambari logo, the host name "an-test-hdi4", user "admin", and notification counts (0 for both).
- Page Title:** / Services / Oozie / Summary
- Summary Tab:** The "SUMMARY" tab is selected, while "CONFIGS" is also present.
- Summary Section:** Contains a "Summary" heading and a "Components" table:

Component	Status
OOZIE SERVER	Started
OOZIE SERVER	Started
OOZIE CLIENTS	3 Installed
- Quick Links Section:** A sidebar on the right lists "Quick Links" with two entries:
  - hn0-an-tes.firebaseio.com (Active)
  - hn1-an-tes.firebaseio.com (Active)

The link "Oozie Web UI" is highlighted with a red box.

# Oozie UI

The screenshot shows the Oozie Web Console interface. At the top, there is a navigation bar with back, forward, and search icons, followed by the URL "an-test-hdi4.azurehdinsight.net/oozie/?user.name=admin". To the right of the URL are a magnifying glass icon, a star icon, a purple circle with a white letter "A", and a "Relaunch to update" button with three dots.

The main header has a "Documentation" link and a "Oozie Web Console" title. Below the header is a navigation menu with tabs: "Workflow Jobs" (which is selected), "Coordinator Jobs", "Bundle Jobs", "System Info", "Instrumentation", and "Settings".

Under the menu, there is a filter section with buttons for "All Jobs", "Active Jobs" (which is highlighted in red), "Done Jobs", and "Custom Filter". To the right of the filter section, it says "Server version [5.2.1.5.1.6.7]".

The main content area displays a table with the following columns: Job Id, Name, Status, R..., User, Group, Created, Started, Last Modified, and Ended. There are no rows of data in the table.

At the bottom of the page, there is a footer with navigation icons (back, forward, search) and a page number indicator "Page 1 of 1".

Oozie .. no job launched yet,

need "workflow.xml" file  
+ submit (not from UI !)

Ambari > Spark

# Ambari > Spark > Summary

The screenshot shows the Ambari web interface. The left sidebar has a dark theme with a list of services: Dashboard, Services (with HDFS, YARN, MapReduce2, Tez, Hive, Oozie, ZooKeeper, Ambari Metrics, Zeppelin Note..., Jupyter, Spark3, WebHCat, Hosts, Alerts, Cluster Admin). The 'Spark3' item is highlighted with a red box. The main content area shows the 'Services / Spark3 / Summary' page. The top navigation bar includes 'an-test-hdi4', a gear icon (0), a bell icon (0), a grid icon, and a user dropdown for 'admin'. Below the navigation is a green 'ACTIONS' button. The main content is divided into sections: 'SUMMARY' (highlighted with a red box) and 'CONFIGS'. The 'SUMMARY' section contains a 'Summary' card with a bell icon (0) and a 'Components' card. The 'Components' card shows the following status: SPARK3 HISTORY SERVER (Started, 1/1 Live, LIVY FOR SPARK3 SERVER), SPARK3 THRIFT SERVERS (2/2 Live), and SPARK3 CLIENTS (3 Installed). To the right is a 'Quick Links' panel with a link to 'Spark3 History Server UI'. The bottom of the sidebar has a 'Cluster Admin' section.

# Ambari > Spark > Quick Link > "Spark History Server UI"

The screenshot shows the Ambari web interface. The left sidebar has a dark theme with various service links. The main area shows the 'Services / Spark3 / Summary' page. The top navigation bar includes the cluster name 'an-test-hdi4', user 'admin', and a green 'ACTIONS' button. The main content area has tabs 'SUMMARY' (selected) and 'CONFIGS'. The 'Summary' section shows the status of components: 'Started SPARK3 HISTORY SERVER', '1/1 Live LIVY FOR SPARK3 SERVER', '2/2 Live SPARK3 THRIFT SERVERS', and '3 Installed SPARK3 CLIENTS'. To the right, a 'Quick Links' box contains a single item, 'Spark3 History Server UI', which is highlighted with a red border.

# Spark History Server

url : /sparkhistory

an-test-hdi4.azurehdinsight.net/sparkhistory/

APACHE Spark 3.3.1.5.1.6.7 History Server

Event log directory: abfss://hdp/spark3-events

Last updated: 2024-11-11 17:02:59

Client local time zone: Europe/Paris

Search:

Version	App ID	App Name	Started	Completed	Duration	Spark User	Last Updated	Event Log
3.3.1.5.1.6.7	application_1731334645766_0003	SparkSQL::10.0.0.16	2024-11-11 15:23:04	2024-11-11 15:23:34	30 s	root	2024-11-11 15:23:34	<button>Download</button>

Showing 1 to 1 of 1 entries

Show incomplete applications

Click on SparkHistory > Application  
=> see Spark UI

(same URL as Tracker URL from Yarn)

The screenshot shows the Apache Spark History Server UI. The top navigation bar includes links for 'Jobs', 'Stages', 'Storage', 'Environment', 'Executors', 'SQL / DataFrame', 'Data (Preview)', 'Graph (Preview)', and 'Diagnostic (Preview)'. On the right side of the header, there are icons for search, star, user profile, and 'Relaunch to update'. The main content area is titled 'Spark Jobs' with a question mark icon. It displays system information: User: root, Total Uptime: 30 s, and Scheduling Mode: FIFO. Below this, there is a link to 'Event Timeline'.

## Spark Jobs [\(?\)](#)

User: root

Total Uptime: 30 s

Scheduling Mode: FIFO

[Event Timeline](#)

# Spark History Server

## Download "eventlog" Files

an-test-hdi4.azurehdinsight.net/sparkhistory/ Relaunch to update

**History Server**  
3.3.1.5.1.6.7

Event log directory: abfss://hdp/spark3-events

Last updated: 2024-11-11 17:02:59

Client local time zone: Europe/Paris

Search:

Version	App ID	App Name	Started	Completed	Duration	Spark User	Last Updated	Event Log
3.3.1.5.1.6.7	application_1731334645766_0003	SparkSQL::10.0.0.16	2024-11-11 15:23:04	2024-11-11 15:23:34	30 s	root	2024-11-11 15:23:34	<button>Download</button>

Showing 1 to 1 of 1 entries  
Show incomplete applications

download  
"evenlog" File  
( keep long,  
send by email..)

# Spark History Server

## click "Show Incomplete Applications"

an-test-hdi4.azurehdinsight.net/sparkhistory/

APACHE Spark 3.3.1.5.1.6.7 History Server

Event log directory: abfss://hdp/spark3-events

Last updated: 2024-11-11 17:02:59

Client local time zone: Europe/Paris

Search:

Version	App ID	App Name	Started	Completed	Duration	Spark User	Last Updated	Event Log
3.3.1.5.1.6.7	application_1731334645766_0003	SparkSQL::10.0.0.16	2024-11-11 15:23:04	2024-11-11 15:23:34	30 s	root	2024-11-11 15:23:34	<button>Download</button>

Showing 1 to 1 of 1 entries

[Show incomplete applications](#)

# Spark History > Incomplete Applications

← → ⌂ an-test-hdi4.azurehdinsight.net/sparkhistory/?showIncomplete=true

APACHE  3.3.1.5.1.6.7 History Server

Event log directory: abfss://hdp/spark3-events

Last updated: 2024-11-11 17:08:11

Client local time zone: Europe/Paris

Search:

Version	App ID	App Name	Started	Spark User	Last Updated	Event Log
3.3.1.5.1.6.7	application_1731334645766_0002	Thrift JDBC/ODBC Server	2024-11-11 15:19:29	spark	2024-11-11 17:08:11	<button>Download</button>
3.3.1.5.1.6.7	application_1731334645766_0001	Thrift JDBC/ODBC Server	2024-11-11 15:18:37	spark	2024-11-11 17:07:20	<button>Download</button>

Showing 1 to 2 of 2 entries

[Back to completed applications](#)

Notice: Spark History may discover "eventlog" Files for RUNNING apps, with a delay of ~5 minutes

Ambari > Zeppelin

# Ambari > Zeppelin > Summary

The screenshot shows the Ambari web interface. On the left, the sidebar lists various services: Dashboard, Services (HDFS, YARN, MapReduce2, Tez, Hive, Oozie, ZooKeeper, Ambari Metrics), Zeppelin Note... (highlighted with a red box), Jupyter, Spark3, WebHCat, Hosts, and Alerts. In the center, the main content area displays the 'Services / Zeppelin Notebook / Summary' page. It has tabs for SUMMARY (selected) and CONFIGS. The SUMMARY tab shows a 'Summary' section with a 'Components' table. The table has one row for 'ZEPPELIN NOTEBOOK' with status 'Started'. To the right is a 'Quick Links' panel containing a link to 'Zeppelin UI', which is also highlighted with a red box. The top right corner shows user information for 'admin'.

Ambari

Dashboard

Services

- HDFS
- YARN
- MapReduce2
- Tez
- Hive
- Oozie
- ZooKeeper
- Ambari Metrics
- Zeppelin Note... (highlighted)
- Jupyter
- Spark3
- WebHCat

Hosts

Alerts

/ Services / Zeppelin Notebook / Summary

SUMMARY CONFIGS ACTIONS ▾

Summary

Components

Started ZEPPELIN NOTEBOOK

Quick Links

Zeppelin UI (highlighted)

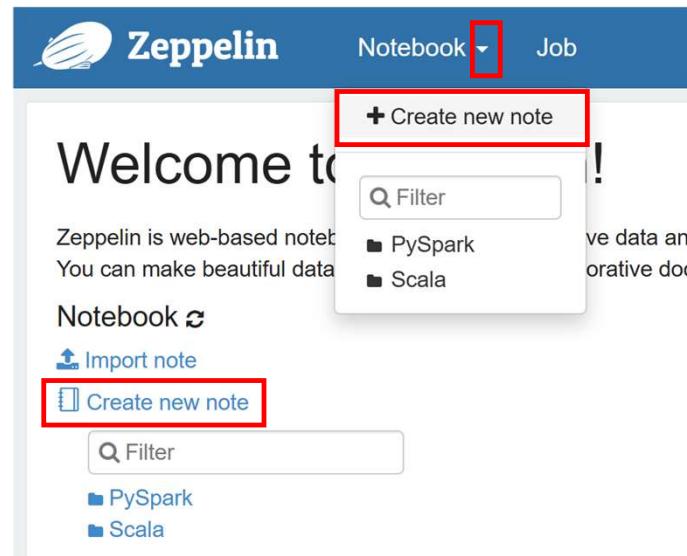
an-test-hdi4 0 0 0 admin ▾

# Zeppelin

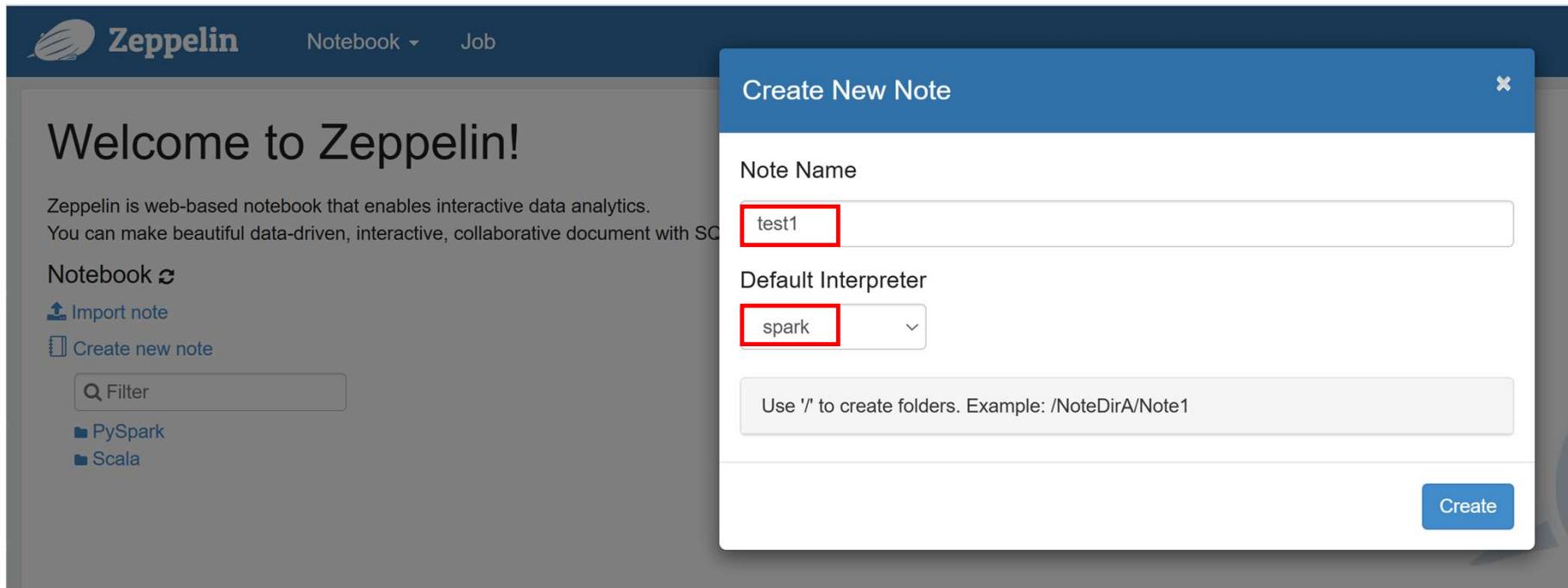
url: /zeppelin/#

The screenshot shows the Zeppelin web interface. At the top, there is a browser header with a back arrow, forward arrow, refresh button, and a URL bar containing "an-test-hdi4.azurehdinsight.net/zeppelin/#/" which is highlighted with a red box. To the right of the URL bar are search, star, and user icons, along with a "Relaunch to update" button. Below the header is a blue navigation bar with the Zeppelin logo, "Notebook", "Job", "Search" (with a magnifying glass icon), and "anonymous". The main content area has a white background. On the left, there's a sidebar titled "Notebook" with options like "Import note", "Create new note", a "Filter" search bar, and categories for "PySpark" and "Scala". The main content area features a large, stylized graphic of a dirigible or airship on the right side. In the center, the text "Welcome to Zeppelin!" is displayed in a large, bold font. Below it, a paragraph reads: "Zeppelin is web-based notebook that enables interactive data analytics. You can make beautiful data-driven, interactive, collaborative document with SQL, code and even more!". To the right of this text are sections for "Help" (link to documentation) and "Community" (links to mailing list, issues tracking, and GitHub). A small "A" icon is visible in the top right corner of the slide.

# Zeppelin > New Notebook



# Create Zeppelin Notebook (Interpreter=Spark)



# Zeppelin > Cell > "Shift-Enter" to execute

The screenshot shows the Zeppelin web interface. At the top, there is a browser header with the URL `an-test-hdi4.azurehdinsight.net/zeppelin/#/notebook/2KE7ZCW7T`. Below the header, the Zeppelin logo is on the left, followed by 'Notebook' and 'Job' dropdowns, a search bar, and a user status 'anonymous'. On the right, there are buttons for 'Relaunch to update' and more options. The main area is titled 'test1'. It contains a single cell with the following content:

```
spark.sql("select 1").show(10, false)
```

Below the code, it says 'Started a few seconds ago.' To the right of the cell, there is a progress bar labeled 'RUNNING 0%' with a red border around it. At the bottom right of the cell area, there is a 'READY' status with a small icon.

# Zeppelin -> Spark Job

← → ⌂ an-test-hdi4.azurehdinsight.net/zeppelin/#/notebook/2KE7ZCW7T

Search A Relaunch to update :

 Zeppelin Notebook Job Search anonymous default

**test1** ▶ ✖ ☰ 🖨️ 📎 ⬇️ ⬆️ ✖️ 🔍 trash

```
spark.sql("select 1").show(10, false)
```

+---+  
|1 |  
+---+  
|1 |  
+---+

Took 51 sec. Last updated by anonymous at November 11 2024, 5:18:01 PM.

View in Spark web UI SPARK JOB FINISHED ▶ ✖ ☰ 🖨️

READY ▶ ✖ ☰ 🖨️

# Zeppelin - Spark Job launched by "local" (not visible in yarnui) or by livy

← → ⌂ an-test-hdi4.azurehdinsight.net/sparkhistory/?showIncomplete=true

APACHE  3.3.1.5.1.6.7 History Server

Event log directory: abfss://hdp/spark3-events

Last updated: 2024-11-11 17:24:42

Client local time zone: Europe/Paris

Search:

Version	App ID	App Name	Started	Spark User	Last Updated	Event Log
3.3.1.5.1.6.7	local-1731341851191	spark-shared_process	2024-11-11 17:17:28	zeppelin	2024-11-11 17:23:12	<button>Download</button>
3.3.1.5.1.6.7	application_1731334645766_0002	Thrift JDBC/ODBC Server	2024-11-11 15:19:29	spark	2024-11-11 17:24:12	<button>Download</button>
3.3.1.5.1.6.7	application_1731334645766_0001	Thrift JDBC/ODBC Server	2024-11-11 15:18:37	spark	2024-11-11 17:24:22	<button>Download</button>

Showing 1 to 3 of 3 entries

[Back to completed applications](#)

# Zeppelin Spark Job in SparkUI

← → ⌂ an-test-hdi4.azurehdinsight.net/sparkhistory/history/local-1731341851191/jobs/ 🔍 ☆ A Relaunch to update :

APACHE Spark 3.3.1.5.1.6.7 Jobs Stages Storage Environment Executors SQL / DataFrame Data (Preview) Graph (Preview) Diagnostic (Preview) spark-shared\_process application UI

## Spark Jobs (?)

User: zeppelin  
Total Uptime:  
Scheduling Mode: FAIR  
Completed Jobs: 2

▶ Event Timeline

▼ Completed Jobs (2)

Page: 1 1 Pages. Jump to 1 . Show 100 items in a page. Go

Job Id (Job Group) ▾	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
1 (zeppelin anonymous 2KE7ZCW7T paragraph_1731341812437_990007050)	Started by: anonymous <a href="#">show at &lt;console&gt;:23</a>	2024/11/11 16:23:06	0.1 s	1/1	1/1
0 (zeppelin anonymous 2KE7ZCW7T paragraph_1731341812437_990007050)	Started by: anonymous <a href="#">show at &lt;console&gt;:23</a>	2024/11/11 16:17:59	2 s	1/1	1/1

Page: 1 1 Pages. Jump to 1 . Show 100 items in a page. Go

Jupyter

# Jupyter

## url :/jupyter

The screenshot shows the Jupyter Notebook interface. At the top, there is a browser-style header with back, forward, and refresh buttons, followed by the URL "an-test-hdi4.azurehdinsight.net/jupyter/tree". To the right of the URL are search and star icons. Below the header is a blue banner with the text "UPDATE Read [the migration plan](#) to Notebook 7 to learn about the new features and the actions to take if you are using extensions - Please note that updating to Notebook 7 might break some of your extensions." In the main area, there is a logo for "jupyter" and a navigation bar with tabs for "Files", "Running", and "Clusters". The "Files" tab is selected. Below the tabs, there is a message "Select items to perform actions on them." On the left, there is a file list with a checkbox, a dropdown menu, and a folder icon. On the right, there are buttons for "Upload", "New", and "Edit". Below the file list, a message says "The notebook list is empty." There are also buttons for "Name", "Last Modified", and "File size".

Ambari > Hive View  
(NOT RECOMMENDED, prefer Spark)

# Ambari > Hive View 2

The screenshot shows the Ambari Hive View 2.0 interface. On the left, there's a sidebar with links to Dashboard, Services (HDFS, YARN, MapReduce2), and a three-dot menu. The main area has tabs for QUERY, JOBS, TABLES, SAVED QUERIES, UDFs, and SETTINGS. Below these is a worksheet titled "Worksheet1". In the top right, there are status icons for "an-test-hdi4" (0 errors, 0 warnings) and a user dropdown for "admin". A red box highlights the "Views" icon in the top right corner of the main content area. A dropdown menu is open under "Views", listing "YARN Queue Manager", "Hive View 2.0" (which is also highlighted with a red box), and "Tez View". A green "NEW TABLE" button is visible on the far right of the dropdown.

# Ambari > Hive View (page loading...)

The screenshot shows the Ambari interface with the title "Hive View 2.0". The left sidebar lists various services: Dashboard, Services (HDFS, YARN, MapReduce2, Tez, Hive, Oozie, ZooKeeper, Ambari Metrics, Zeppelin Note..., Jupyter, Spark3, WebHCat), and Hosts. The top right corner shows the host name "an-test-hdi4", a gear icon with "0", a bell icon with "0", and a user icon for "admin". The main content area is titled "HIVE SERVICE CHECKS" and displays three green bars, each with a checkmark icon and a service name: "HDFS Check", "ATS Check", and "HIVE Check".

# Ambari > Hive View

The screenshot shows the Ambari Hive View 2.0 interface. The top navigation bar includes the Ambari logo, a home icon, the text "/ Hive View 2.0", user information for "an-test-hdi4" and "admin", and several status icons. Below the header, the main area is titled "HIVE" and contains a "Worksheet1" tab. The interface features a "QUERY" tab, a "JOBS" tab, a "TABLES" tab, a "SAVED QUERIES" tab, a "UDFs" tab, and a "SETTINGS" tab. A "NOTIFICATIONS" section is also present. On the left, a sidebar menu lists various services: Dashboard, Services (HDFS, YARN, MapReduce2), Tez, Hive, Oozie, ZooKeeper, Ambari Metrics, Zeppelin Note..., Jupyter, Spark3, WebHCat, Hosts, Alerts, Cluster Admin, Stack and Versions, and Service Accounts. The central workspace displays a "DATABASE" dropdown set to "default", a "Browse" button, and a large text input field labeled "1". To the right, a sidebar shows the "default" database selected, with "Tables(1)" listed and "hivesampletable" highlighted. At the bottom of the workspace, there are buttons for "Execute", "Save As", "Insert UDF", and "Visual Explain".

# HiveView > Execute SQL Query

The screenshot shows the Hive View 2.0 web interface. At the top, there's a header with a home icon, the text "/ Hive View 2.0", user information for "an-test-hdi4" and "admin", and notification counts (0). Below the header is a navigation bar with tabs: QUERY (selected), JOBS, TABLES, SAVED QUERIES, UDFs, and SETTINGS. There are also "+ NEW JOB" and "+ NEW TABLE" buttons, as well as a NOTIFICATIONS section.

The main area is titled "HIVE". It features a "Worksheet1 \*" tab and a "+" button to create new worksheets. Under "DATABASE", the "default" database is selected. To the right, a sidebar shows the "default" database with one table, "Tables(1)", named "hivesamptable". A search bar for tables is also present.

In the central query editor, the following SQL command is typed:

```
1 select 1
```

Below the query editor are buttons for "Execute" (which is checked), "Save As", "Insert UDF", and "Visual Explain".

The bottom navigation bar includes "RESULTS" (selected), "LOG", "VISUAL EXPLAIN", and "TEZ UI". The results pane shows a single row with the value "1".

(ADMIN) Ambari > Hosts

# Ambari > Hosts (admin)

The screenshot shows the Ambari interface for managing hosts. The left sidebar contains a navigation menu with items like Dashboard, Services, and Hosts, where 'Hosts' is highlighted with a red box. The main content area is titled 'Hosts' and displays a table of host information. The table columns include Name, IP Address, Rack, Cores, RAM, Disk Usage, Load Avg, Versions, and Components. The hosts listed are:

Name	IP Address	Rack	Cores	RAM	Disk Usage	Load Avg	Versions	Components
hn0-an-tes.firebaseio.com	10.0.0.16	/default-rack	4 (4)	27.41GB	Low	1.76	HDInsight-5.1.6.7	24 Components
hn1-an-tes.firebaseio.com	10.0.0.15	/default-rack	4 (4)	27.41GB	Medium	1.18	HDInsight-5.1.6.7	19 Components
wn1-an-tes.firebaseio.com	10.0.0.4	/default-rack	4 (4)	27.41GB	Low	0.59	HDInsight-5.1.6.7	7 Components
zk0-an-tes.firebaseio.com	10.0.0.8	/default-rack	2 (2)	3.83GB	High	0.49	HDInsight-5.1.6.7	4 Components
zk2-an-tes.firebaseio.com	10.0.0.9	/default-rack	2 (2)	3.83GB	Medium	0.72	HDInsight-5.1.6.7	4 Components
zk3-an-tes.firebaseio.com	10.0.0.6	/default-rack	2 (2)	3.83GB	Low	0.54	HDInsight-5.1.6.7	4 Components

Items per page: 10 1 - 6 of 6

Hosts : 2 Head Nodes (hn0, hn1) + 3 Zookeeper Nodes (zk0, zk1, zk2) + N (>=1) Worker Nodes (wn0, wn1, ...)

# Ambari > Hosts > wn1 Summary

Ambari

Dashboard Services ...

HDFS YARN MapReduce2 Tez Hive Oozie ZooKeeper Ambari Metrics Zeppelin Notebooks Jupyter Spark3 WebHCat Hosts Alerts Cluster Admin Stack and Versions Service Accounts

HOST ACTIONS ▾ LAST 1 HOUR ▾

SUMMARY CONFIGS ALERTS 0 VERSIONS

Components

Status	Name	Type	Action
✓	DataNode / HDFS	Slave	...
✓	Metrics Monitor / Ambari Metrics	Slave	...
✓	NodeManager / YARN	Slave	...
✓	Hive Client / Hive	Client	...
✓	Oozie Client / Oozie	Client	...
✓	Spark3 Client / Spark3	Client	...
✓	Tez Client / Tez	Client	...

Host Metrics

CPU Usage

Disk Usage

Load

Memory Usage

Summary

Hostname: wn1-an-tes.firebaseio.com  
IP Address: 10.0.0.4  
Rack: /default-rack  
OS: ubuntu18 (x86\_64)

an-test-hdi4 admin

LAST 1 HOUR ▾

100% 50%

279.3 GB 186.2 GB 93.1 GB

18.6 GB 9.3 GB

150 100 50

58.5 KB 39.0 KB 19.5 KB

(ADMIN) Ambari > Users & Groups

# (ADMIN) Ambari > Login > Manage Amari

The screenshot shows the Ambari Metrics dashboard. The left sidebar lists services: HDFS, YARN, MapReduce2, Tez, and Hive. The main area displays four metrics: NameNode Heap (15%), HDFS Disk Usage (8%), NameNode CPU WIO (0.1%), and DataNodes Live (1/1). The top right shows user 'admin' with a red box around the 'Manage Ambari' link in the dropdown menu.

Ambari

Dashboard / Metrics

an-test-hdi4

admin

METRICS HEATMAPS CONFIG HISTORY

NameNode Heap: 15%

HDFS Disk Usage: 8%

NameNode CPU WIO: 0.1%

DataNodes Live: 1/1

About

**Manage Ambari**

Settings

Sign out

# Ambari > Admin > Users

The screenshot shows the Ambari Admin / Users interface. The left sidebar has links for Dashboard, Cluster Management, Cluster Information, Versions, Remote Clusters, Users (which is selected and highlighted with a green bar), and Views. The main content area has tabs for USERS (selected and highlighted with a red box) and GROUPS. It shows a table with columns: Username, Display Name, Role, Status, Type, Group, and Actions. Two users are listed:

Username	Display Name	Role	Status	Type	Group	Actions
admin	admin	Ambari Administrator	Active	Local	-	
hdinsightwatchdog	hdinsightwatchdog	Ambari Administrator	Active	Local	-	

At the top right, there are links for 'an-test-hdi4' and 'admin'. The URL in the address bar is 'an-test-hdi4.azurehdinsight.net/views/ADMIN\_VIEW/2.7.3/INSTANCE/#/userManagement'.

# Ambari > Admin > Groups

The screenshot shows the Ambari web interface with the following details:

- Header:** Ambari logo, Admin / Users, User 'admin' (an-test-hdi4).
- Left Sidebar:** Dashboard, Cluster Management (with a dropdown arrow), Cluster Information, Versions, Remote Clusters, Users (selected), Views.
- Current View:** Admin / Users, Groups tab selected (highlighted with a red box).
- Table Headers:** Group name, Type, Members, Actions.
- Table Content:** NO GROUPS TO DISPLAY.
- Buttons:** ADD GROUPS.

(ADMIN) Ambari > Metrics > Grafana

# (ADMIN) Ambari Metrics > Grafana

The screenshot shows the Ambari Metrics Summary page. On the left, a sidebar lists various services: Dashboard, Services (HDFS, YARN, MapReduce2, Tez, Hive, Oozie, ZooKeeper), Ambari Metrics (selected and highlighted with a red box), Zeppelin Note..., Jupyter, Spark3, WebHCat, Hosts, and Alerts. The main content area displays a summary of components: METRICS COLLECTOR (Started) and GRAFANA (Started). It also shows 6/6 live METRICS MONITORS. On the right, there is a 'Quick Links' section with a 'Grafana' button, which is also highlighted with a red box.

an-test-hdi4.azurehdinsight.net/#/main/services/AMBARI\_METRICS/summary

an-test-hdi4 admin

SUMMARY CONFIGS METRICS ACTIONS

Summary

Components

Started METRICS COLLECTOR

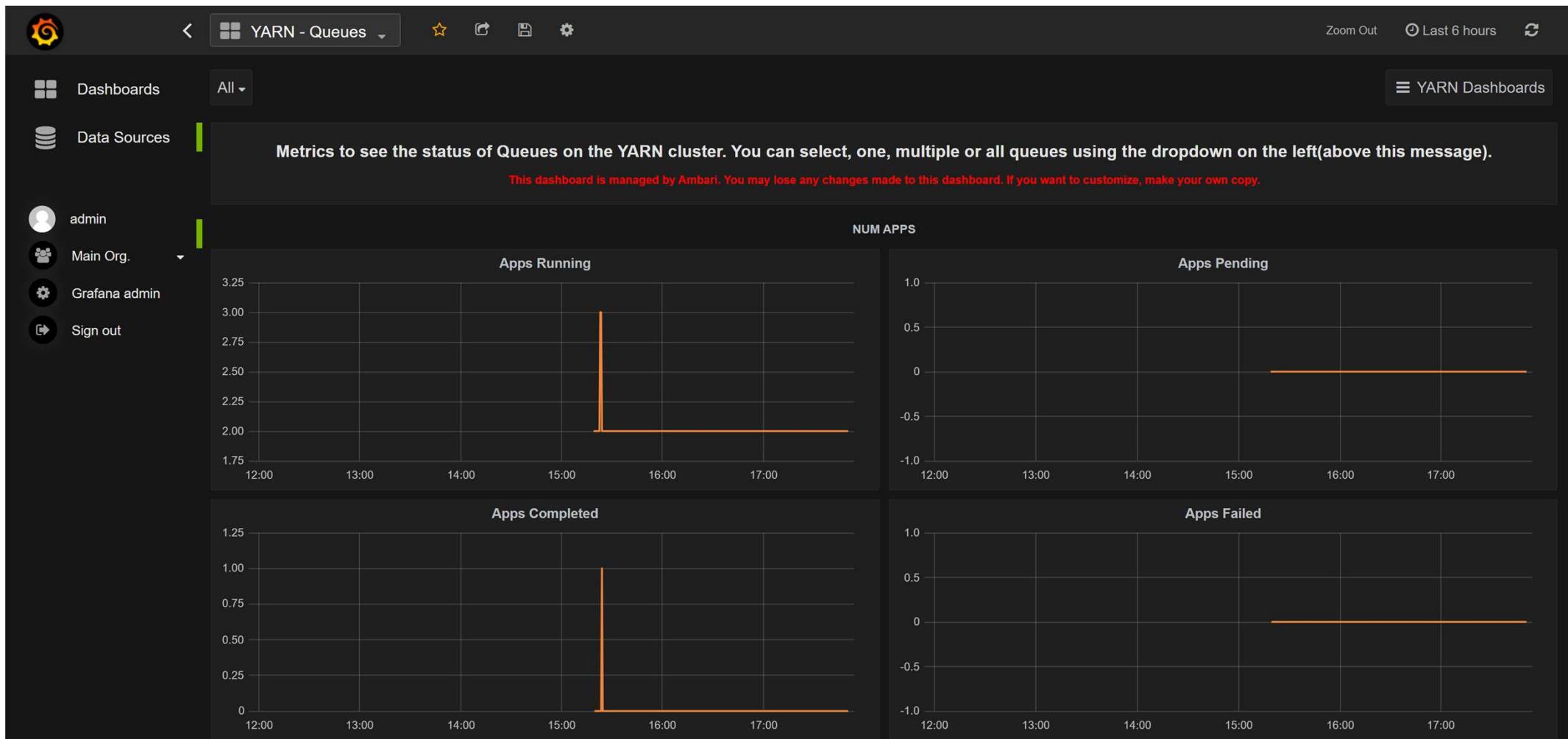
Started GRAFANA

6/6 Live METRICS MONITORS

Quick Links

Grafana

# Grafana



# Conclusion

HD Insight is a ready to use all-in-one Hadoop environment.

Hadoop would be extremely difficult to setup yourself on your PC (ok on linux, not windows).

Hadoop Yarn is overly complex.

Kubernetes is Better than Hadoop Yarn.

And Unfortunately, there is NO standard "pre-package" all-in-one environment for Spark on k8s + Jupyter + ...

But, with a custom environment,

Spark on Kubernetes is Better/Cheaper/Easier than Spark on HD-Insight.

Questions ?

arnaud.nauwynck@gmail.com