

HDI+Talena Resources Deployment Guide

June 2017

© 2017 Talena Inc. All rights reserved.

Talena, the Talena logo are trademarks of Talena Inc., registered in the U.S.

Other company and product names mentioned herein may be trademarks of their respective companies.

Release 3.0.0

June 2017

TABLE OF CONTENTS

GETTING STARTED WITH THE DEPLOYMENT	4
CONFIGURING THE IP ADDRESS IN TALENA GUI.....	7
FINDING THE IP ADDRESS TO SET UP HIVE FOR HDI CLUSTER	7
SETTING THE IP ADDRESS.....	10

Getting started with the deployment

This section describes the procedure of deploying HDI+Talena resources to your Azure subscription in five easy steps:

1. Install Azure CLI
2. Clone the repo from github.com
3. Modify the JSON parameters
4. Validate the deployment
5. Launch the deployment

To deploy HDI+Talena resources:

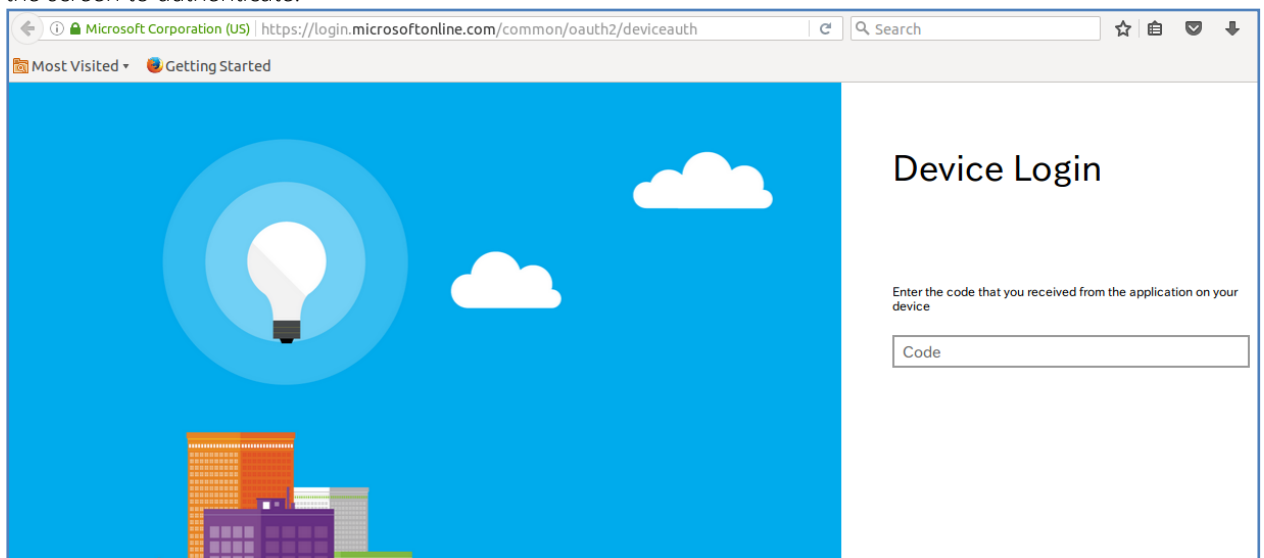
1. Install Azure CLI.

```
curl -L https://aka.ms/InstallAzureCli | bash
```

```
[azure@server4 ~]$ az login
To sign in, use a web browser to open the page https://aka.ms/devicelogin and
enter the code H6WR42QNV to authenticate.
```

```
[
  {
    "cloudName": "AzureCloud",
    "id": "7bd1ce62-b5d3-40ec-9792-07145f619a50",
    "isDefault": true,
    "name": "Pay-As-You-Go",
    "state": "Enabled",
    "tenantId": "8ca2403a-3a8f-4da8-9a9a-5dddd34ac0eb",
    "user": {
      "name": "companyname@company.com",
      "type": "user"
    }
  }
]
```

2. Open a Web browser, type the URL <https://aka.ms/devicelogin>, and then type the code given on the screen to authenticate:



3. Clone the template repository available on github.com:

```
#:~/talena/azure$ git clone git@github.com:Talena/public-templates.git
Cloning into 'public-templates'...
X11 forwarding request failed on channel 0
remote: Counting objects: 49, done.
remote: Compressing objects: 100% (8/8), done.
remote: Total 49 (delta 0), reused 3 (delta 0), pack-reused 41
Receiving objects: 100% (49/49), 20.90 KiB | 0 bytes/s, done.
Resolving deltas: 100% (18/18), done.
Checking connectivity... done.

#:~/talena/azure$ cd public-templates/replication-template/
```

4. Modify the JSON template to enter values against the respective parameters as shown below:

```
{
  "artifactsBaseUrl": {
    "value":
"https://talenaahdistorageaccount.blob.core.windows.net/talenaahdi"
  },
  "firstName": {
    "value": "John"
  },
  "lastName": {
    "value": "Hopkins"
  },
  "company": {
    "value": "Talena"
  },
  "email": {
    "value": "jhopkins@talena-inc.com"
  },
  "phone": {
    "value": "+91 1234567890"
  },
  "title": {
    "value": "Senior Developer"
  },
  "licenseKey": {
    "value": "1234567890"
  },
  "clusterName": {
    "value": "talenaazurevd1"
  },
  "username": {
    "value": "talena"
  },
  "password": {
    "value": "example@123"
  },
  "clusterLoginUserName": {
    "value": "admin"
  },
  "clusterLoginPassword": {
    "value": "hdinsight@123"
  },
  "sharedKey": {
    "value": "abc123"
  },
  "location1": {
    "value": "West US"
  },
  "location2": {
    "value": "East US"
  }
}
```

5. Validate the deployment of the resources by assigning execute permissions to validate.sh

```
chmod +x validate.sh
```

```
#: ./validate.sh TalenaHDIRGVD1
```

6. Ensure that the validation succeeds with the following message:

```
"provisioningState": "Succeeded",  
"template": null,  
"templateLink": null,  
"timestamp": "2017-06-15T05:47:34.587951+00:00"
```

7. Launch the deployment of the resources:

```
chmod +x deploy.sh
```

```
#: ./deploy.sh TalenaHDIRGVD
```

8. Login to Azure portal, and verify your deployment status by going over to the **Resource Groups** option, and then click the **Deployments** option.

The screenshot shows the Microsoft Azure portal interface. On the left sidebar, the 'Deployments' option under the 'TalenaHDIRGVD1' resource group is highlighted with a red box. The main content area displays a table of deployments for the 'TalenaHDIRGVD1' resource group. The table has columns for 'DEPLOYMENT NAME', 'STATUS', 'TIMESTAMP', and 'DURATION'. The deployments listed include 'mainTemplate', 'gateway', 'hdi2', 'talena', 'scripts', 'hdi1', 'parentDeployment-0-1444368793748180a29dca2d22432', 'subDeployment-10-1444368793748180a29dca2d22432', 'subDeployment-6-1444368793748180a29dca2d22432', 'subDeployment-4-1444368793748180a29dca2d22432', 'vm2', 'subDeployment-8-1444368793748180a29dca2d22432', 'vm1', 'vm0', 'parentDeployment-0-c5c5148288774dc58a77253676afdfc', 'subDeployment-4-c5c5148288774dc58a77253676afdfc', 'subDeployment-10-c5c5148288774dc58a77253676afdfc', 'subDeployment-8-c5c5148288774dc58a77253676afdfc', 'subDeployment-6-c5c5148288774dc58a77253676afdfc', 'network', 'talenaazurev1-vnet', and 'vnet'. All deployments are marked as 'Succeeded'.

DEPLOYMENT NAME	STATUS	TIMESTAMP	DURATION
mainTemplate	Succeeded	6/15/2017, 12:13:59 PM	43 minutes 30 seconds
gateway	Succeeded	6/15/2017, 12:13:53 PM	42 minutes 36 seconds
hdi2	Succeeded	6/15/2017, 11:57:26 AM	26 minutes 10 seconds
talena	Succeeded	6/15/2017, 11:55:15 AM	24 minutes 2 seconds
scripts	Succeeded	6/15/2017, 11:55:03 AM	20 minutes 6 seconds
hdi1	Succeeded	6/15/2017, 11:45:39 AM	14 minutes 27 seconds
parentDeployment-0-1444368793748180a29dca2d22432	Succeeded	6/15/2017, 11:34:52 AM	24 seconds
subDeployment-10-1444368793748180a29dca2d22432	Succeeded	6/15/2017, 11:34:47 AM	16 seconds
subDeployment-6-1444368793748180a29dca2d22432	Succeeded	6/15/2017, 11:34:44 AM	10 seconds
subDeployment-4-1444368793748180a29dca2d22432	Succeeded	6/15/2017, 11:34:42 AM	10 seconds
vm2	Succeeded	6/15/2017, 11:34:41 AM	2 minutes 31 seconds
subDeployment-8-1444368793748180a29dca2d22432	Succeeded	6/15/2017, 11:34:39 AM	4 seconds
vm1	Succeeded	6/15/2017, 11:34:39 AM	2 minutes 29 seconds
vm0	Succeeded	6/15/2017, 11:34:17 AM	2 minutes 6 seconds
parentDeployment-0-c5c5148288774dc58a77253676afdfc	Succeeded	6/15/2017, 11:33:47 AM	29 seconds
subDeployment-4-c5c5148288774dc58a77253676afdfc	Succeeded	6/15/2017, 11:33:37 AM	12 seconds
subDeployment-10-c5c5148288774dc58a77253676afdfc	Succeeded	6/15/2017, 11:33:37 AM	12 seconds
subDeployment-8-c5c5148288774dc58a77253676afdfc	Succeeded	6/15/2017, 11:33:34 AM	9 seconds
subDeployment-6-c5c5148288774dc58a77253676afdfc	Succeeded	6/15/2017, 11:33:31 AM	6 seconds
network	Succeeded	6/15/2017, 11:32:05 AM	47 seconds
talenaazurev1-vnet	Succeeded	6/15/2017, 11:31:29 AM	3 seconds
vnet	Succeeded	6/15/2017, 11:31:02 AM	23 seconds

Configuring the IP address in Talena GUI

In a multi-region deployment setup, if an HDInsight cluster is connected to Talena Cluster using VPN connection, (where HDInsight clusters are setup in different geographical regions), then the host names of HDInsight cluster may not get resolved.

In this case, you need to first find out the IP address to set up Hive for HDI cluster and then configure it at appropriate places in the Talena UI. This section describes how you can do the following:

1. Find the IP address from the Azure Portal or HDInsight Cluster Terminal
2. Setting the IP address in Talena GUI

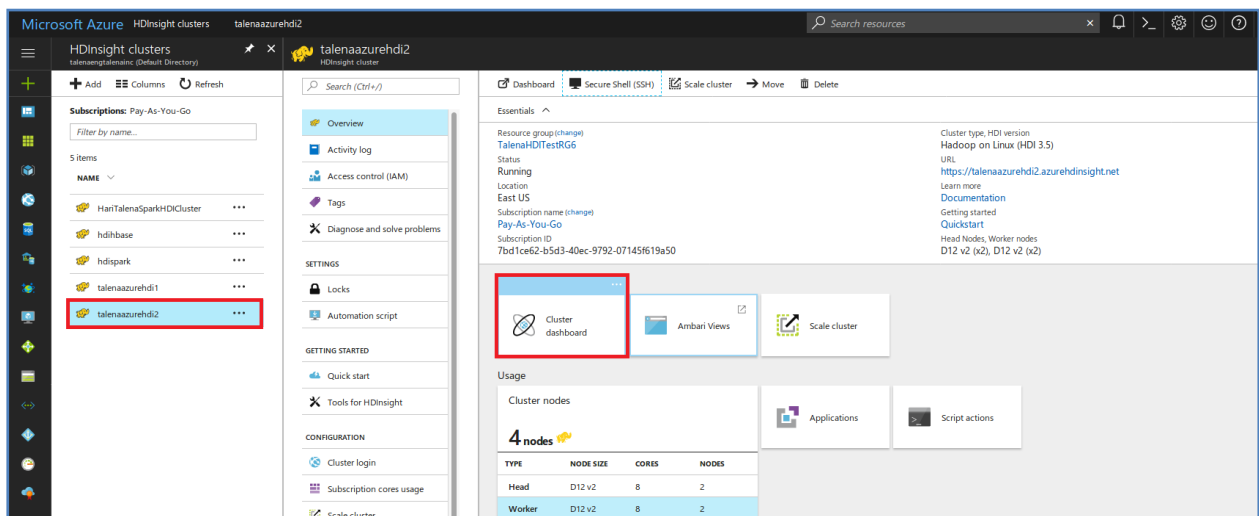
Finding the IP address to set up Hive for HDI cluster

You can find the IP address to set up Hive in two different ways:

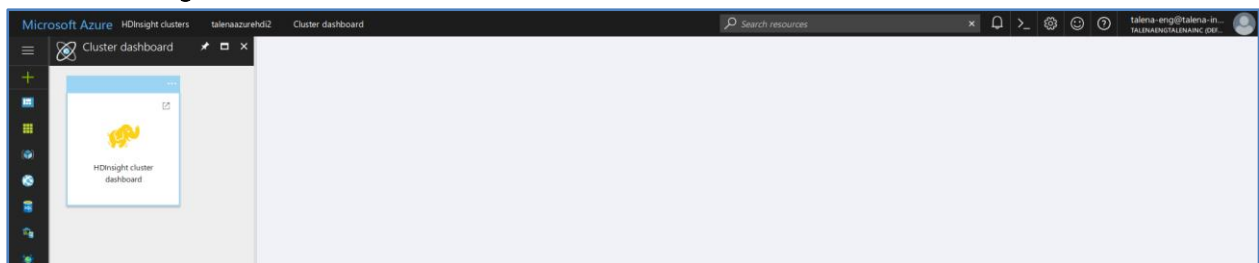
- From the Azure portal
- From HDInsight cluster terminal

From the Azure Portal

1. Login to the Azure portal and click the **HDInsight cluster** from the right pane, and then click the **Cluster Dashboard**.

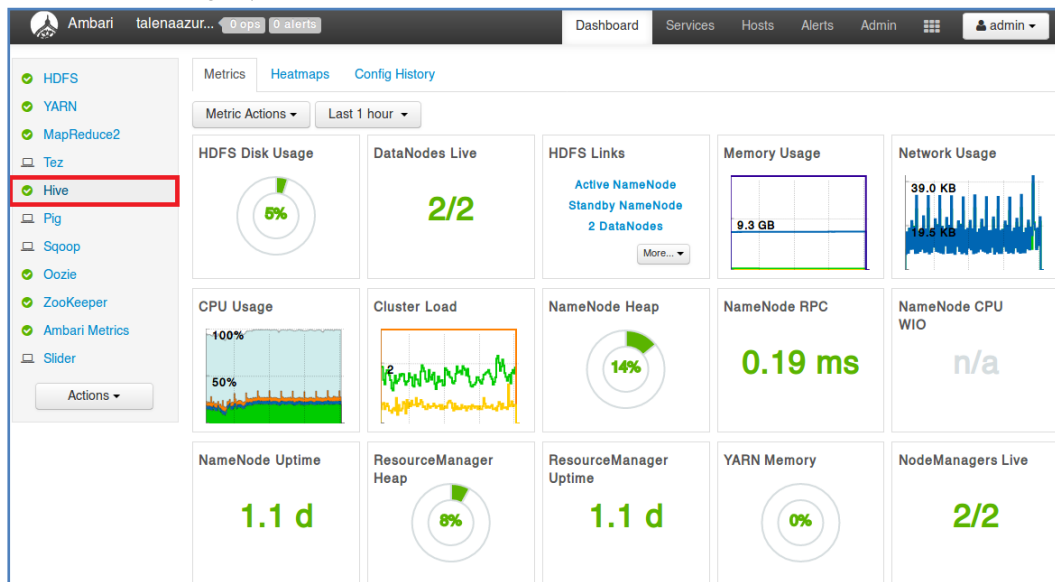


2. Click the **HDInsight Cluster dashboard**.

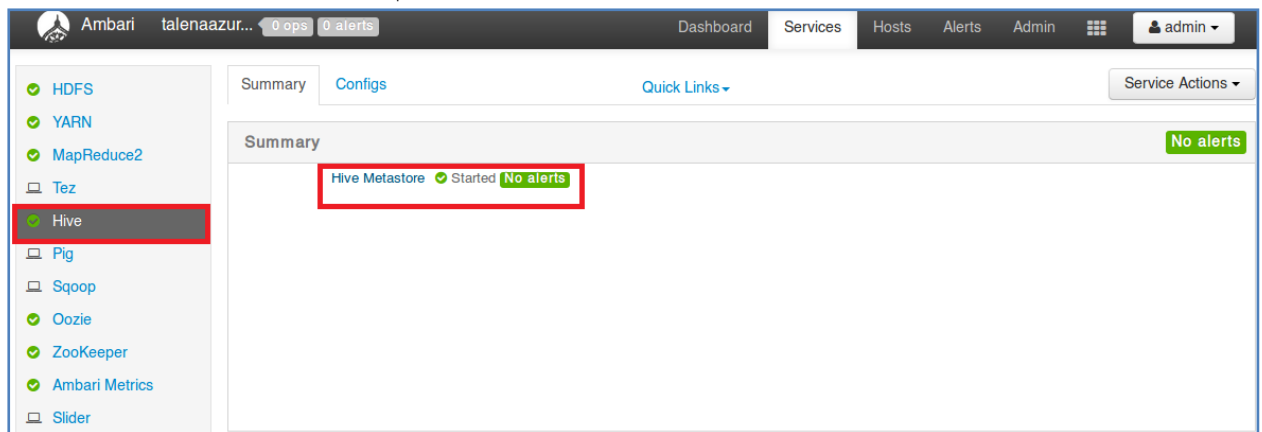


The Ambari dashboard will be displayed...

3. Click Hive in the right pane in the Ambari dashboard screen as shown in the screenshot below:



4. Click the Hive Metastore in the left pane as shown the screenshot below:



5. Scroll down to the bottom of the page and copy the **IP address** mentioned in the **Summary** section:

The screenshot displays the HDInsight cluster management interface for a cluster named `hn0-talena.slyy33ylzy2ebo1njb3ezldi3b.bx.internal.cloudapp.net`. The interface includes tabs for Summary, Configs, Alerts, and Versions, along with a Host Actions dropdown. The Components section lists various services like App Timeline Server, History Server, Hive Metastore, and their status (Started or Installed). The Host Metrics section shows real-time graphs for CPU Usage, Disk Usage, Load, Memory Usage, Network Usage, and Processes. The Summary section at the bottom provides host details, with the IP address `10.2.1.16` highlighted in a red box.

Components

- App Timeline Server / YARN (Started)
- History Server / MapReduce2 (Started)
- Hive Metastore / Hive (Started)
- HiveServer2 / Hive (Started)
- Metrics Collector / Ambari Metrics (Started)
- Active NameNode / HDFS (Started)
- Oozie Server / Oozie (Started)
- Active ResourceM... / YARN (Started)
- WebHCat Server / Hive (Started)
- Metrics Monitor / Ambari Metrics (Started)
- ZKFailoverContro... / HDFS (Started)

Host Metrics (Last 1 hour)

- CPU Usage: 100% (graph)
- Disk Usage: 931.3 GB (graph)
- Load: 6 (graph)
- Memory Usage: 13.9 GB (graph)
- Network Usage: 195.3 KB (graph)
- Processes: 200 (graph)

Summary

Hostname: `hn0-talena.slyy33ylzy2ebo1njb3ezldi3b.bx.internal.cloudapp.net`
IP Address: 10.2.1.16
Hack: `/default-rack`
OS: `ubuntu16 (x86_64)`
Cores (CPU): 4 (4)
Disk: 26.34GB/1166.09GB (2.26% used)
Memory: 27.48GB
Load Avg: 0.20
Heartbeat: a moment ago
Current Version: 3.5.4.0.7

From HDInsight cluster terminal

1. Login to the HDInsight cluster

```
$ ssh talena@talenaazurehdi2-ssh.azurehdinsight.net

Authorized uses only. All activity may be monitored and reported.

talena@talenaazurehdi2-ssh.azurehdinsight.net's password: *****

Welcome to Ubuntu 16.04.2 LTS (GNU/Linux 4.4.0-78-generic x86_64)
* Documentation:  https://help.ubuntu.com
* Management:    https://landscape.canonical.com
* Support:       https://ubuntu.com/advantage
Get cloud support with Ubuntu Advantage Cloud Guest:
http://www.ubuntu.com/business/services/cloud
34 packages can be updated.
23 updates are security updates.
```

2. Copy the IP address from the bottom the screen:

```
Welcome to HDInsight.
Last login: Wed Jun 14 18:54:35 2017 from 45.124.51.14
To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.

talena@hn0-talena:~$ nslookup hn0-
talena.siyy33ylzy2ebolnjb3ezldi3b.bx.internal.cloudapp.net
Server:      168.63.129.16
Address:     168.63.129.16#53



Name:      hn0-talena.siyy33ylzy2ebolnjb3ezldi3b.bx.internal.cloudapp.net
Address: 10.2.1.16
```

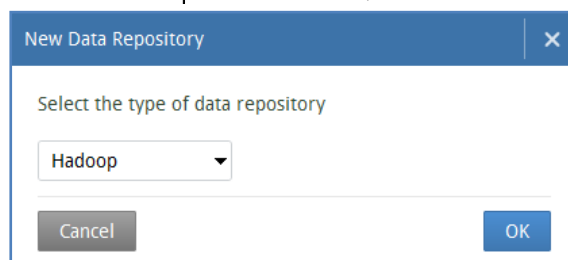
Setting the IP address

This section describes the procedure of setting the IP address in Talena GUI.

Once you get the IP address to set up Hive for HDI cluster, you need to access Talena UI and insert the IP address at the appropriate places.

To set IP address in Talena UI:

1. Login to Talena software
2. Click the **Main Menu**  > **System Setup** > **Data Repositories**.
3. On the **Data Repositories** screen, click the  icon. The following dialog box appears:



4. In the **New Data Repository** dialog box, select **Hadoop** from the **Select the type of data repository** drop-down list, and then click **OK**.
5. Type the following details:
 - a. **Data Repository Name:** Type a name for the data repository. The data repository name can include alphanumeric characters, numbers and/or special characters.
 - b. **Description:** A meaningful description of the data repository to help others identify it
6. In the **Applications** section, type the following:
 - a. **Primary Host:** The name or IP address of the machine on which the configuration directories related to the applications exist.
 - b. **Configuration Directory:** A list of one or more paths where configuration files related to the apps exist.
7. Click **Discover**. This action displays the **Credentials** dialog box that prompts you to type your login credentials.

8. In the **Credentials** dialog box, under **Authentication Mechanism** do one of the following:
 - Click the **Password** option button, type the login and password, and click **OK**
 - Click the **Private Key** option button, type the login id, copy-paste the private key, type the **passphrase** (if you have set the passphrase), and then click **OK**.

This action displays a list of all applications discovered in the data repository by Talena software. For example, HDFS, HBASE and/or HIVE.

If HIVE is auto-discovered, the following fields are automatically populated.

Field Name	Description	Example
Metastore Address	Stores the metadata for Hive tables and partitions in a relational database	t35vm1
Metastore Port	Service for accessing metadata about Hive tables and partitions	9083

In addition, Hive Server Address and Hive Server Port fields are also automatically populated depending upon the configuration properties in the hive-site.xml file. However, if you DO NOT want to activate masking and sampling for Hive, you must manually remove the values in the Hive Server Address and Hive Server Port fields from the Talena GUI.

9. Replace the values that are automatically populated in the **Hive Metastore Address** and **JDBC Driver Address** fields with the IP address that you got from the preceding section.

The screenshot shows the Talena GUI for configuring data repositories. On the left, a list of repositories includes 'talenahdi1-WestUS' and 'talenahdi2-EastUS'. The 'talenahdi2-EastUS' repository is selected. The main panel shows configuration details for this repository. At the top, there are fields for 'Data Repository Name' (talenahdi2-EastUS) and 'Description (optional)'. Below this, the 'Applications' section shows a 'Discover' button and a list of applications. The 'Hive' application is selected, and its configuration is displayed. It includes fields for 'Metastore Address' (10.2.1.16), 'Metastore Port' (9083), 'Hive Server Address' (10.2.1.16), and 'Hive Server Port' (10001). There is also a dropdown for 'Select a JDBC Driver' set to 'apache-hive-1.2.1'. At the bottom of the configuration panel, there are 'Remove' and 'Verify' buttons. The top right of the GUI shows the user 'admin'.

10. Click **Verify** to confirm the selection of your Hive data repository.
11. Set a data backup window and data recovery window, and then click **Save**.