**Analyze Apache Logs and**

**Build your own web Analytics Dashboard with**

**Cloudera Solr and Hue**



**Walk-Through Guide**

Table of Contents

[1. About Cloudera 3](#_Toc471315022)

[2. Getting Started 3](#_Toc471315024)

[3. Logging into Cloudera interface 4](#_Toc471315025)

[4. Install the Solr Service 6](#_Toc471315026)

[5. Configure the Hue Service 12](#_Toc471315027)

[6. Restart the Hue Service 14](#_Toc471315028)

[7. Login to Hue Web UI 16](#_Toc471315029)

[8. Injecting Apache logs to Cloudera 22](#_Toc471315030)

[9. Configure Search Dashboard 23](#_Toc471315031)

[10. Summary 33](#_Toc471315032)

## About Cloudera

# This document will guide you through several use cases around installing and using cloudera services, interactively explore some real Apache Log data, Analyse Apache logs and build your own Web Analytics dashboard with Hadoop and Solr

Cloudera provides an enterprise-ready, open source distribution that includes Apache Hadoop and related projects. Cloudera Enterprise includes [CDH](http://www.cloudera.com/content/cloudera/en/products-and-services/cdh.html), the world’s most popular open source Hadoop-based platform, as well as advanced system management and data management. This massively scalable platform unites storage with an array of powerful processing and analytics frameworks and adds enterprise-class management, data security, and governance. Cloudera Enterprise includes core elements of Hadoop (HDFS, MapReduce, YARN) as well as HBase, Impala, Solr, Spark and more.

This Guide will step you through the Cloudera interface and how to successfully use powerful processing and analytics frameworks and adds enterprise-class management for your web server logs. This guide will utilize the sample Apache web server logs which is hidden behind the scenes during this evaluation.

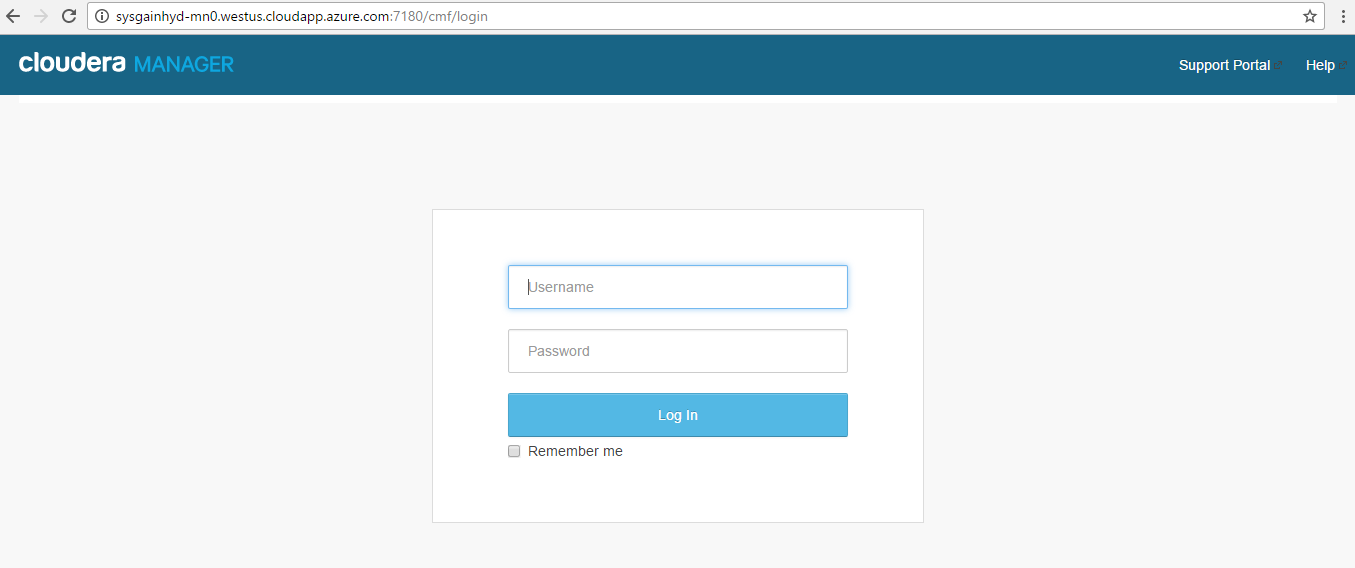
## Getting Started

You will receive an email with the actual ClouderaManagerUrl.

It will look something like:

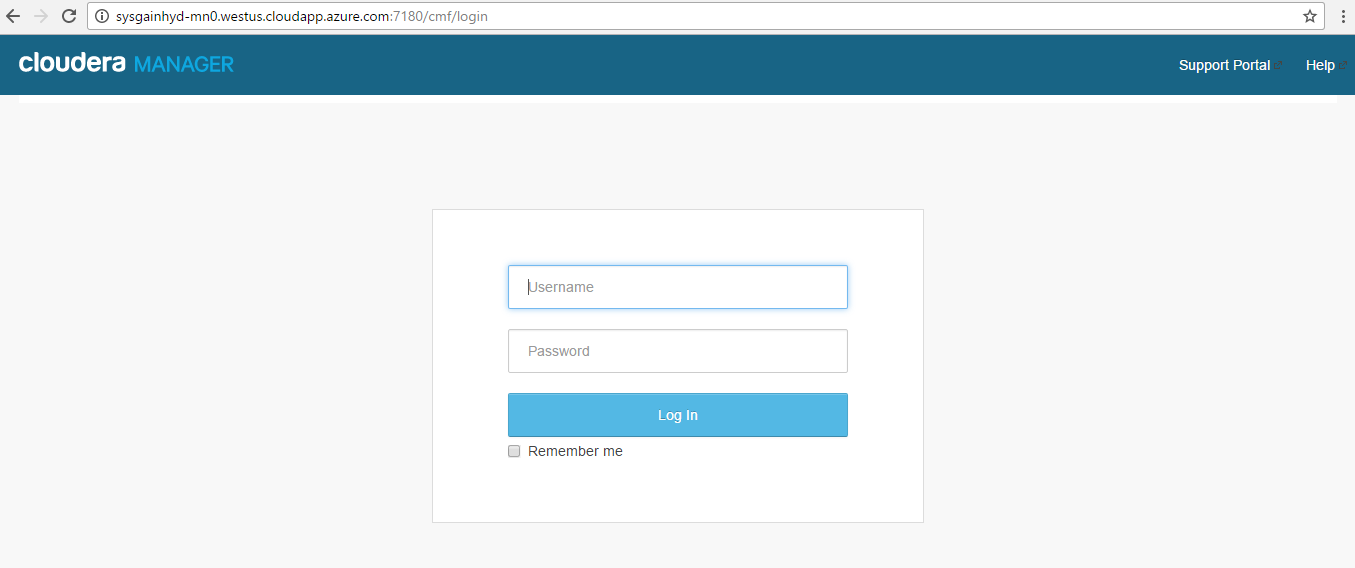
<http://clouderau3y4789zxy-mn0.westus.cloudapp.azure.com:7180/>

Enter that url into I.E. or Chrome and this screen will pop up.

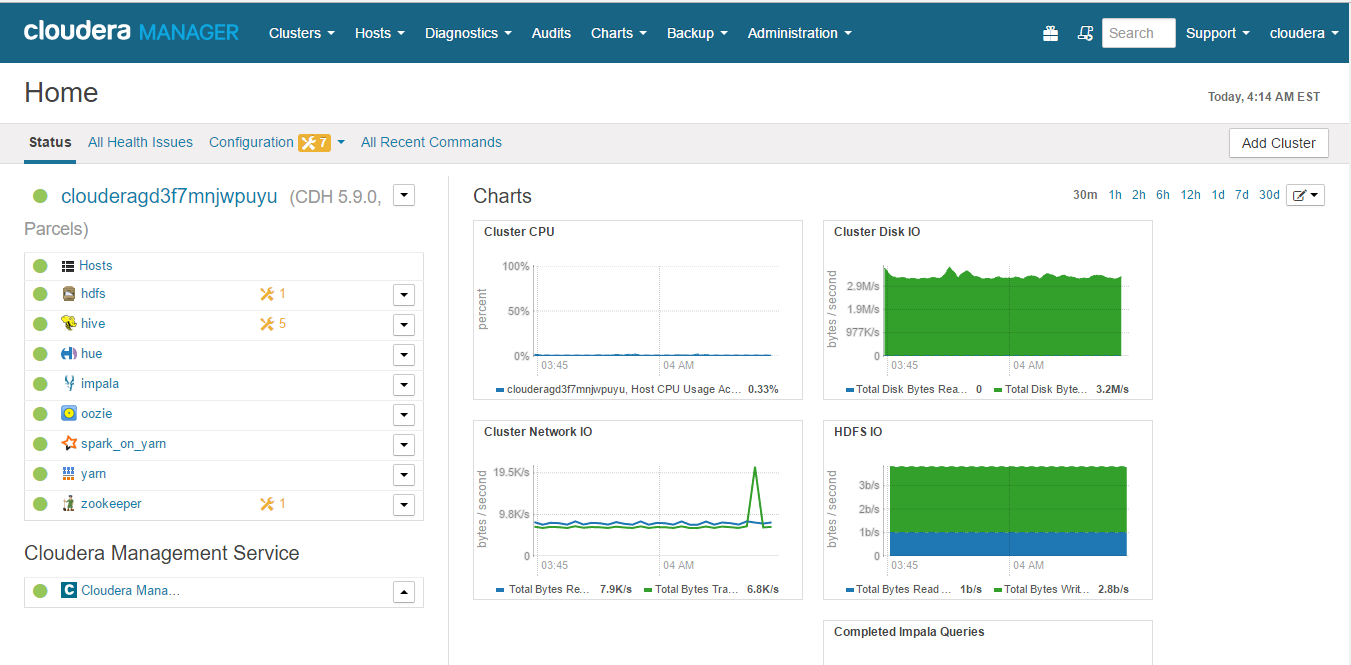


## Logging into Cloudera interface

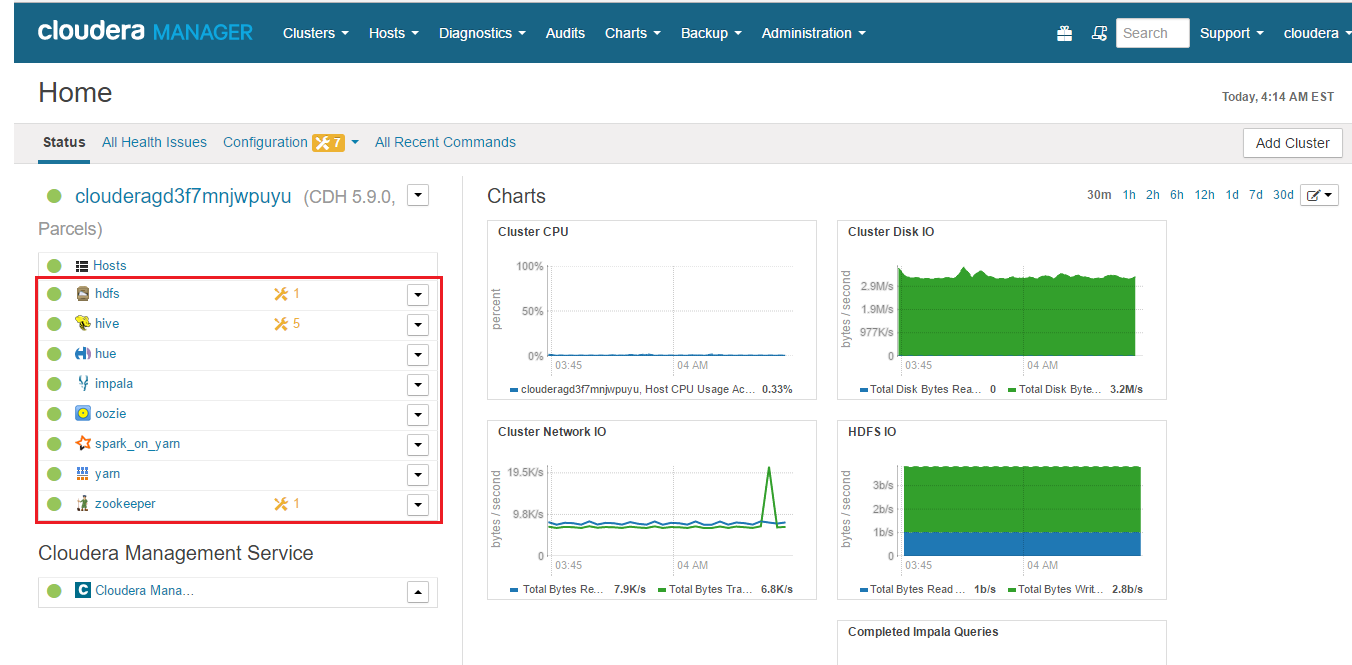
* Enter username: **cloudera**
* Enter password: **Cloudera@123**

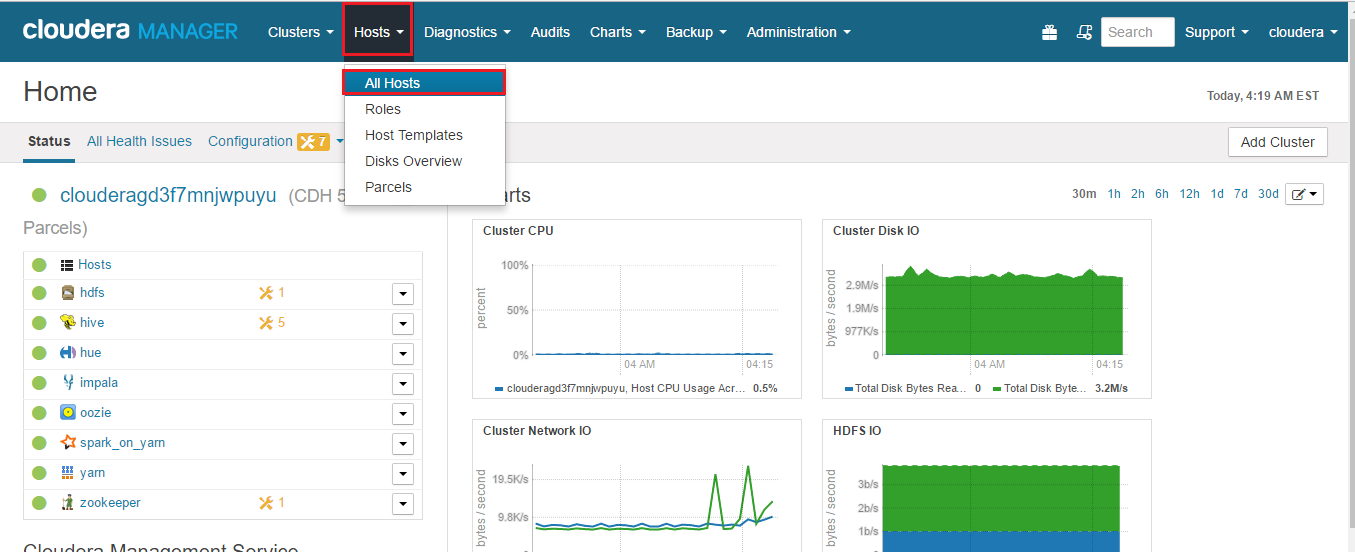


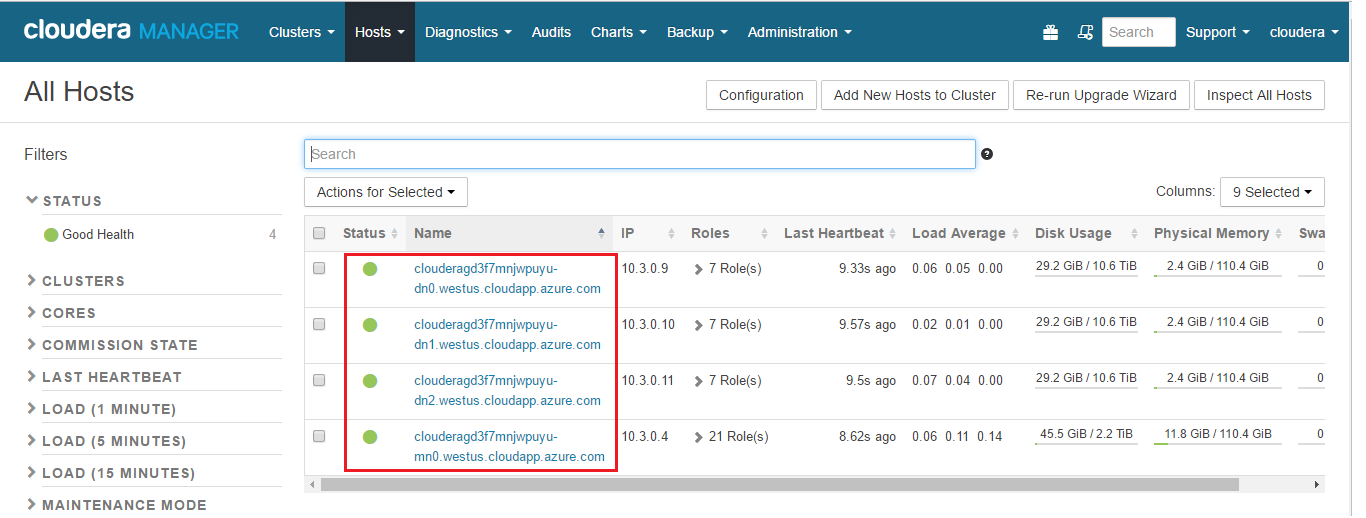
You are now logged into the Cloudera Manager



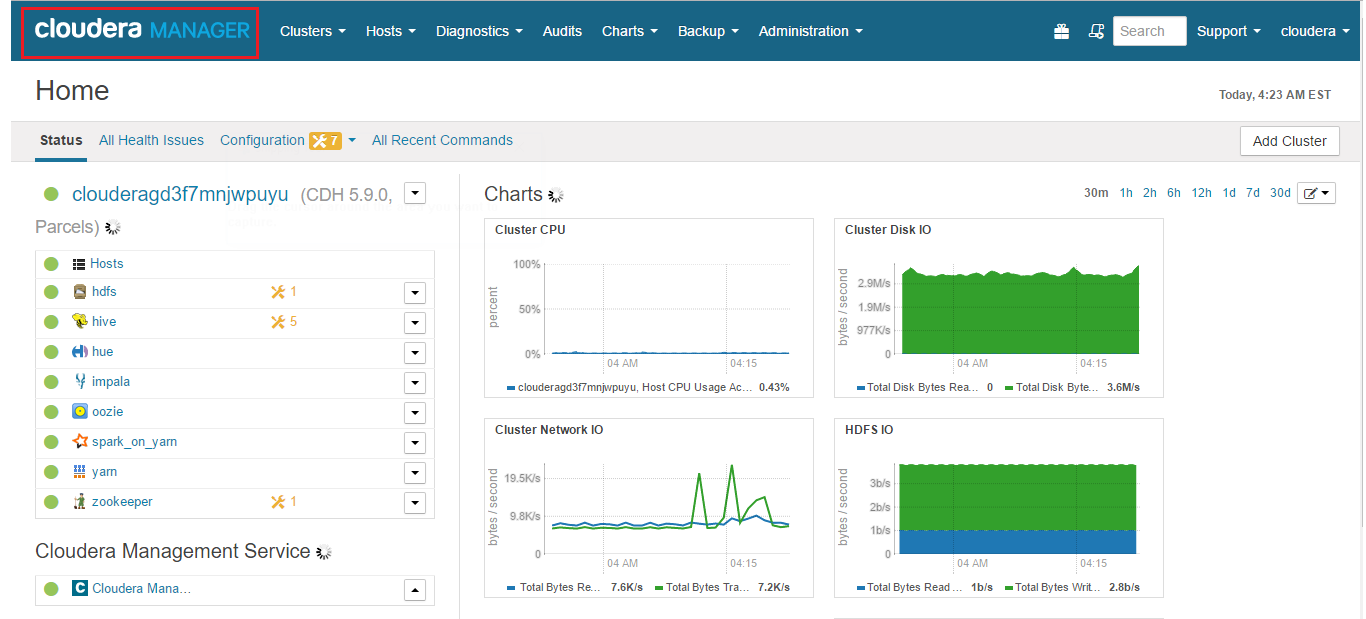
* On the left side of the Cloudera interface, you will see the services running:



* Click on Hosts and then select All Hosts in the Cloudera Manager as shown below:
* 
* You can observe the 1 Master Node and 3 Data Nodes running in the Cloudera Cluster - the screen should be similar to below



Click on Cloudera Manager as shown below to go back to the home screen



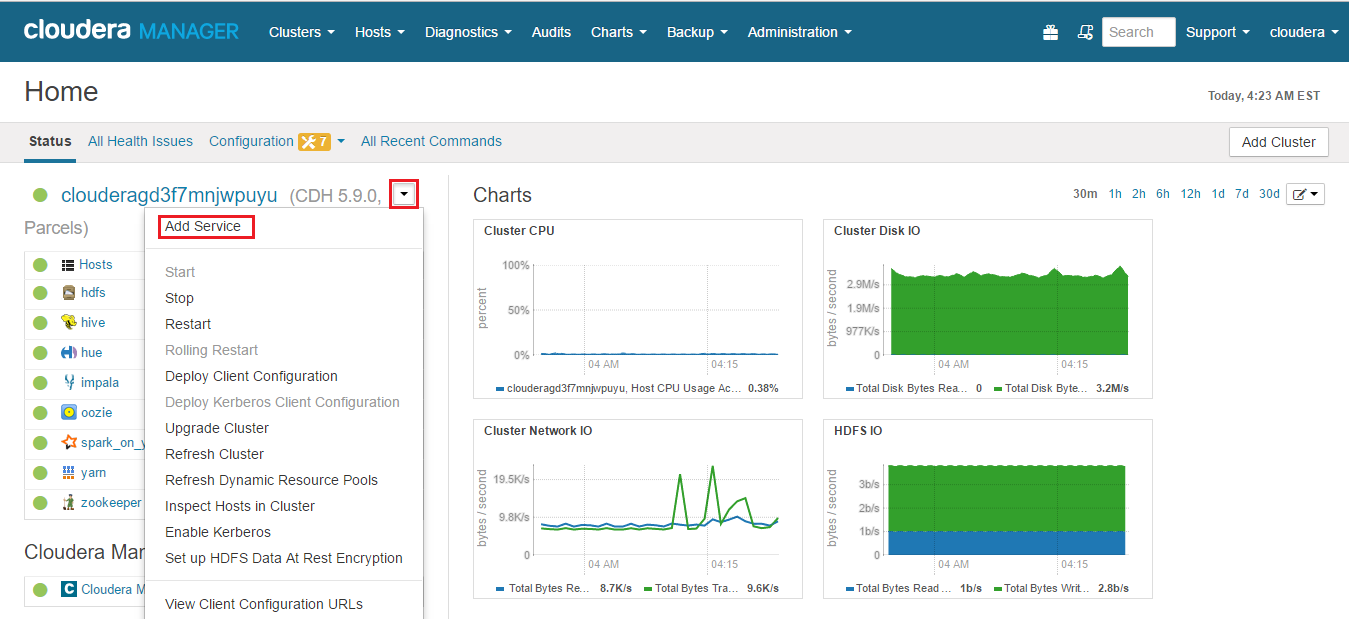
## Install the Solr Service

Cloudera Search is provided by the Solr service, which is included with CDH 5. Cloudera Search supports one instance of the Solr service on each host in a cluster. Using multiple Solr instances on a host is not supported.

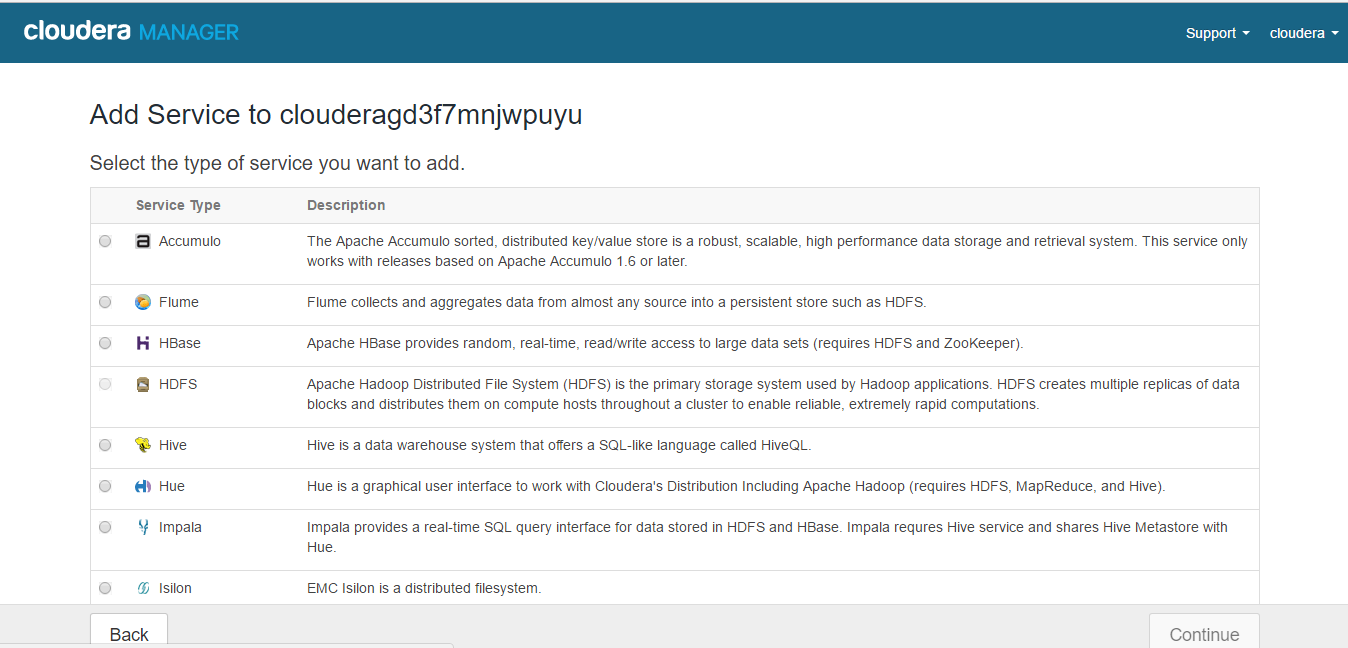
To add a service:

1. On the **Home** > **Status** tab, click  to the right of the cluster name and select **Add a Service**. A list of service types display. You can add one type of service at a time.

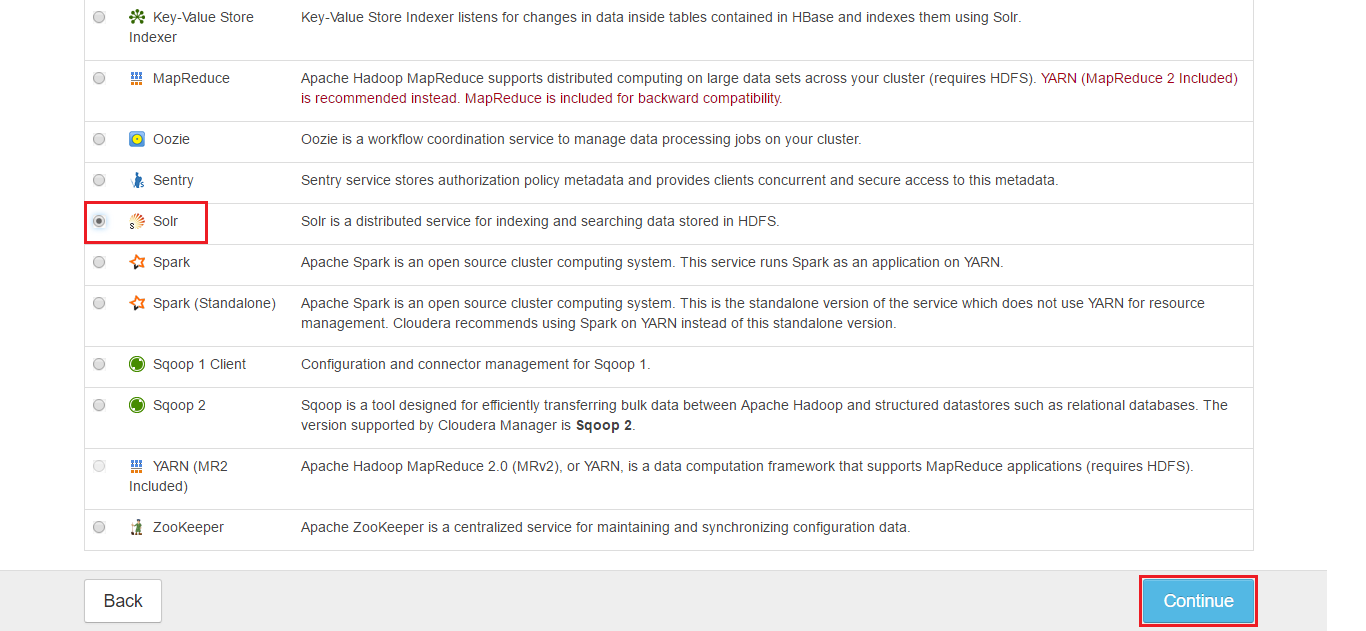
Click on the dropdown list beside the Cluster as shown in the figure below:



Click on Add Service.



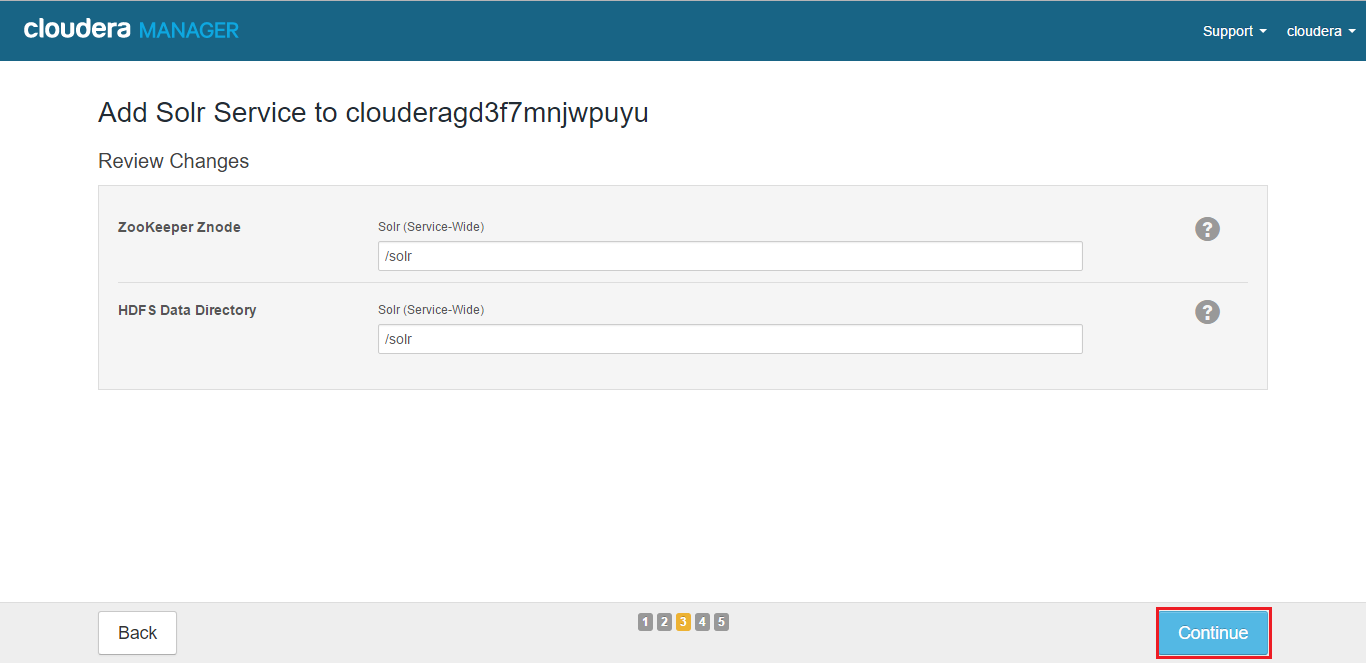
Scroll down to select the solr service. Select the Solr service as shown below and click on Continue**.**



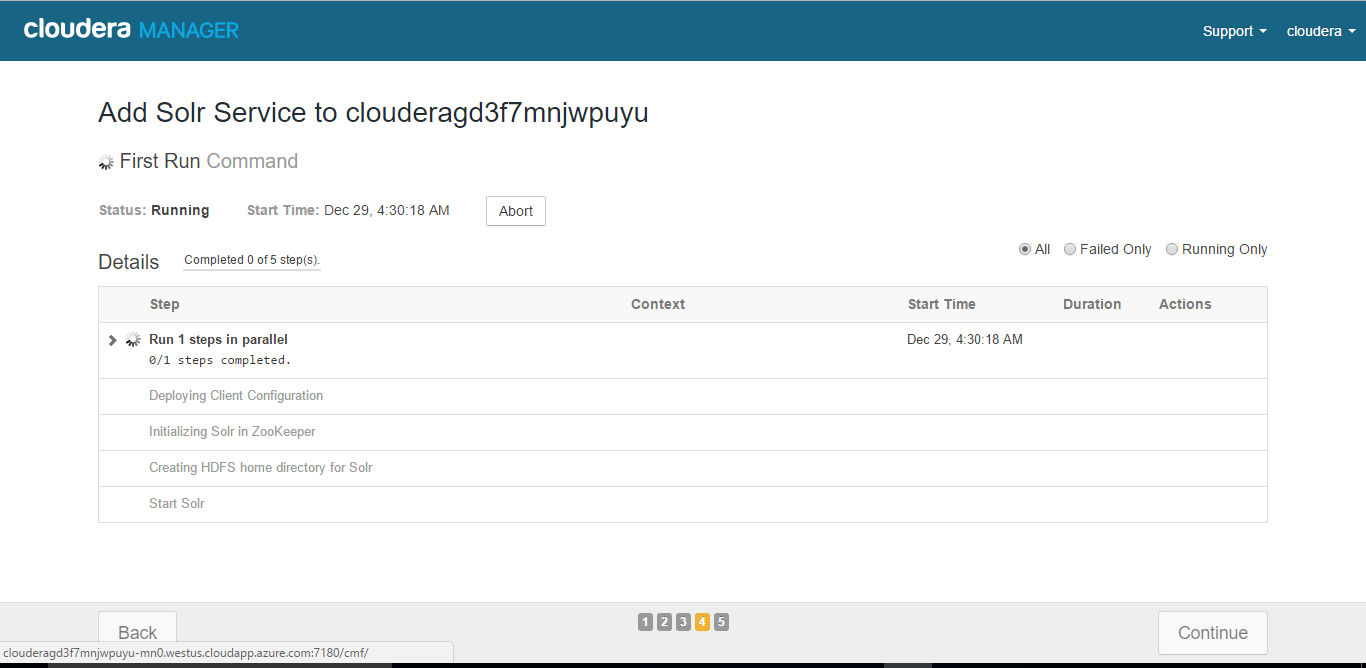
Cloudera Manager by default selects a host in the cluster to add Solr. Note down the FQDN of the host that the Solr service is deployed to, as shown below and click on Continue

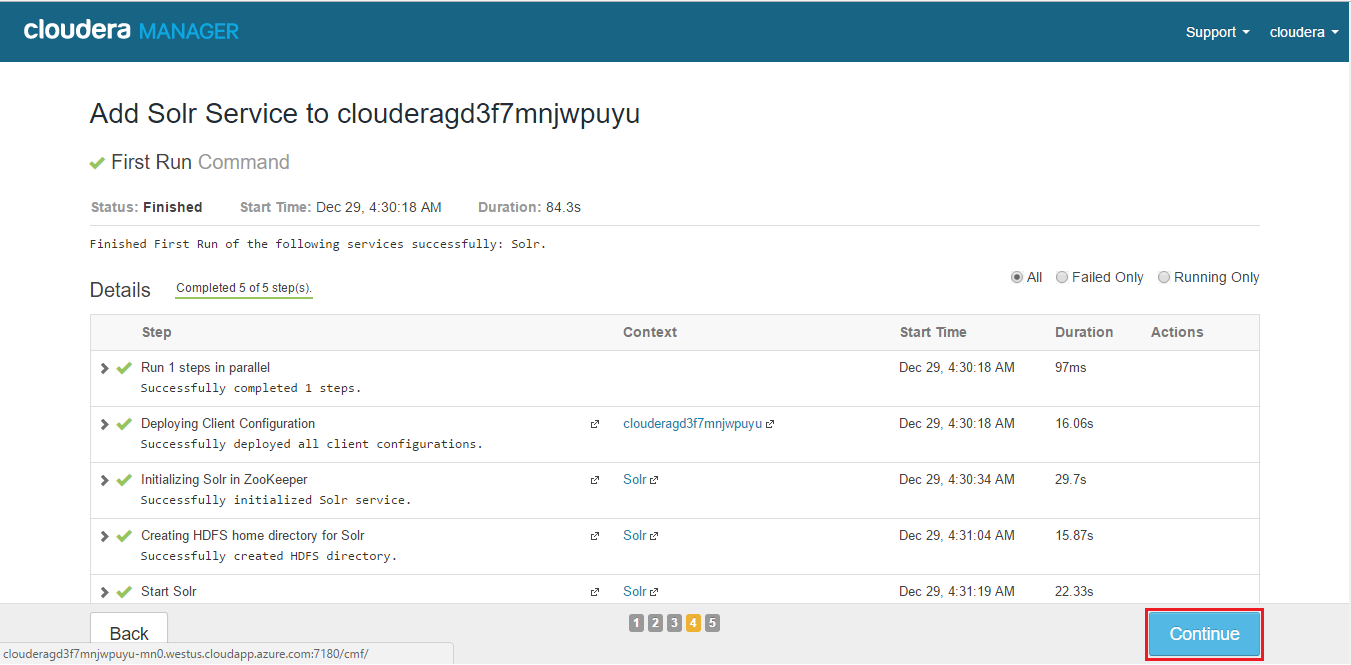


The wizard will automatically configure and start the dependent services and the Solr service.

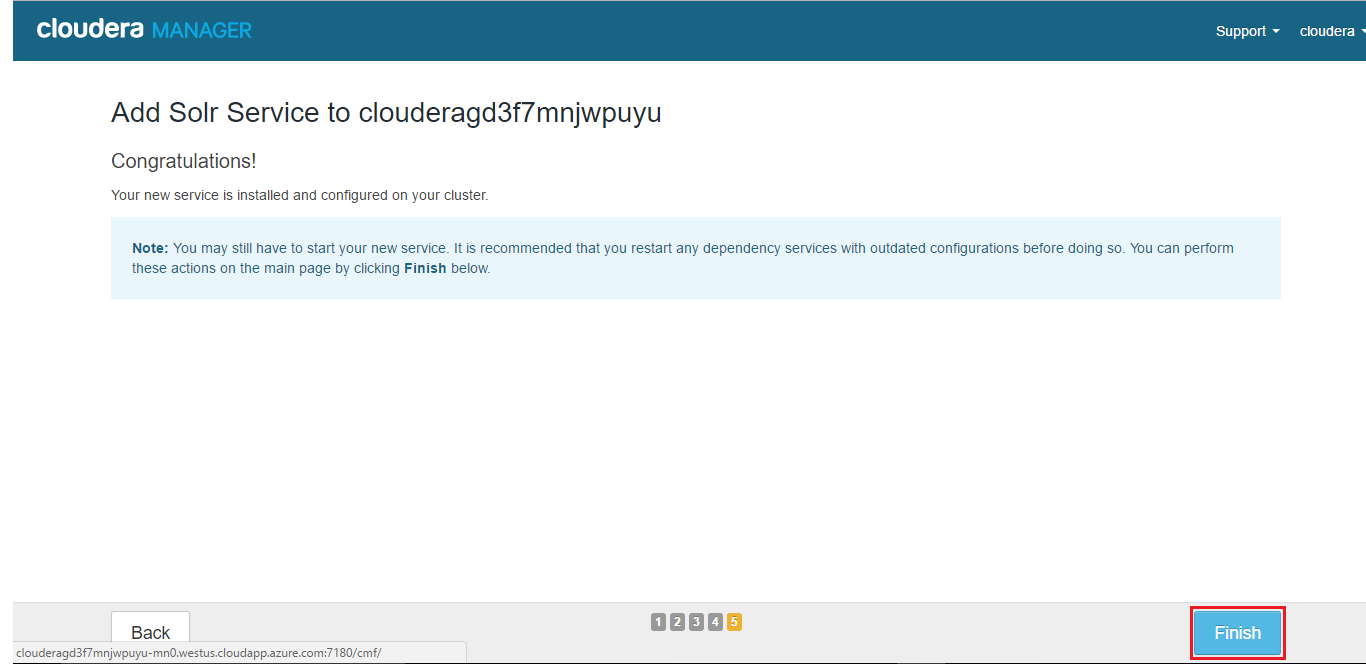


The service is started.

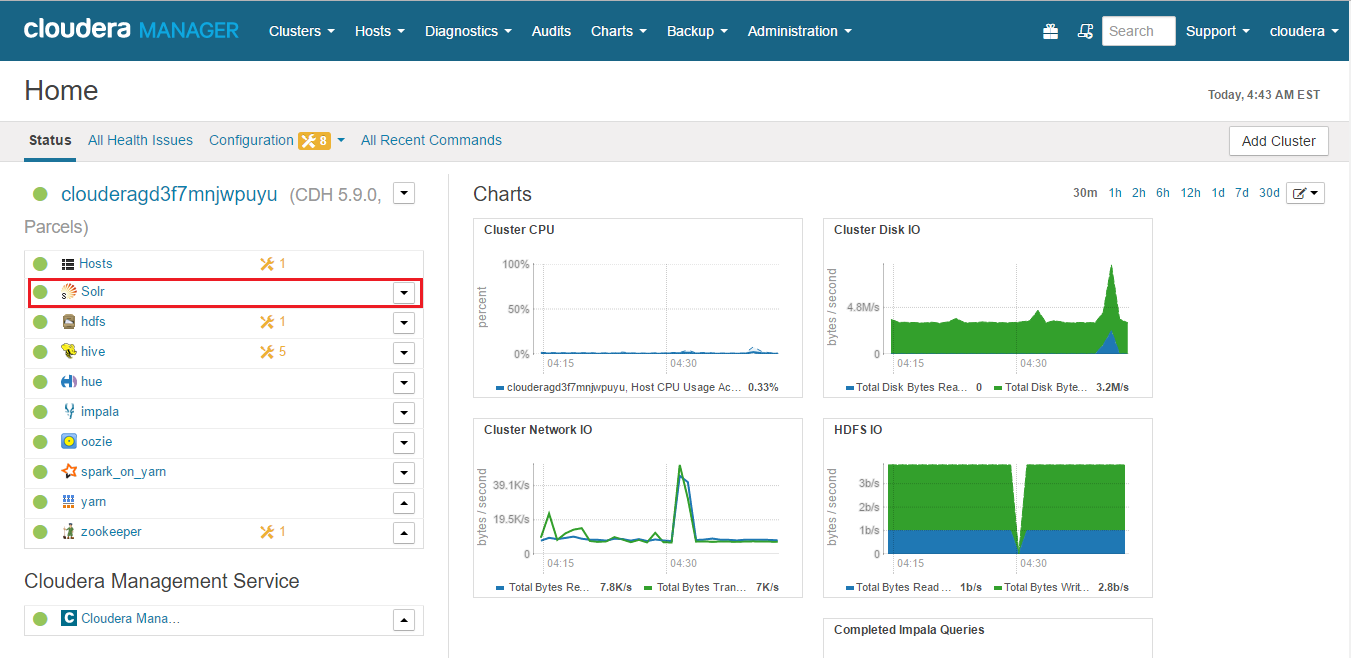




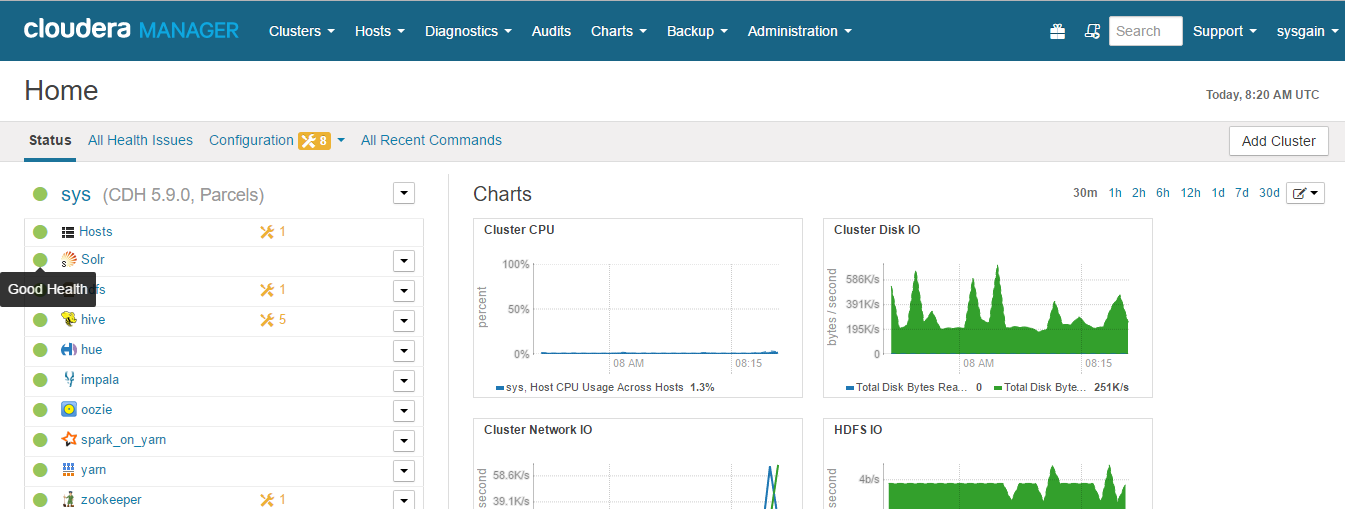
* Click **Continue** then click **Finish**.



* You are returned to the [Home](https://www.cloudera.com/documentation/enterprise/5-8-x/topics/cm_intro_home.html#xd_583c10bfdbd326ba--7f25092b-13fba2465e5--7f7c) page.
* Verify the new Solr service is started properly by checking the health status for the new service. If the Health Status is **Good**, then the service started properly.



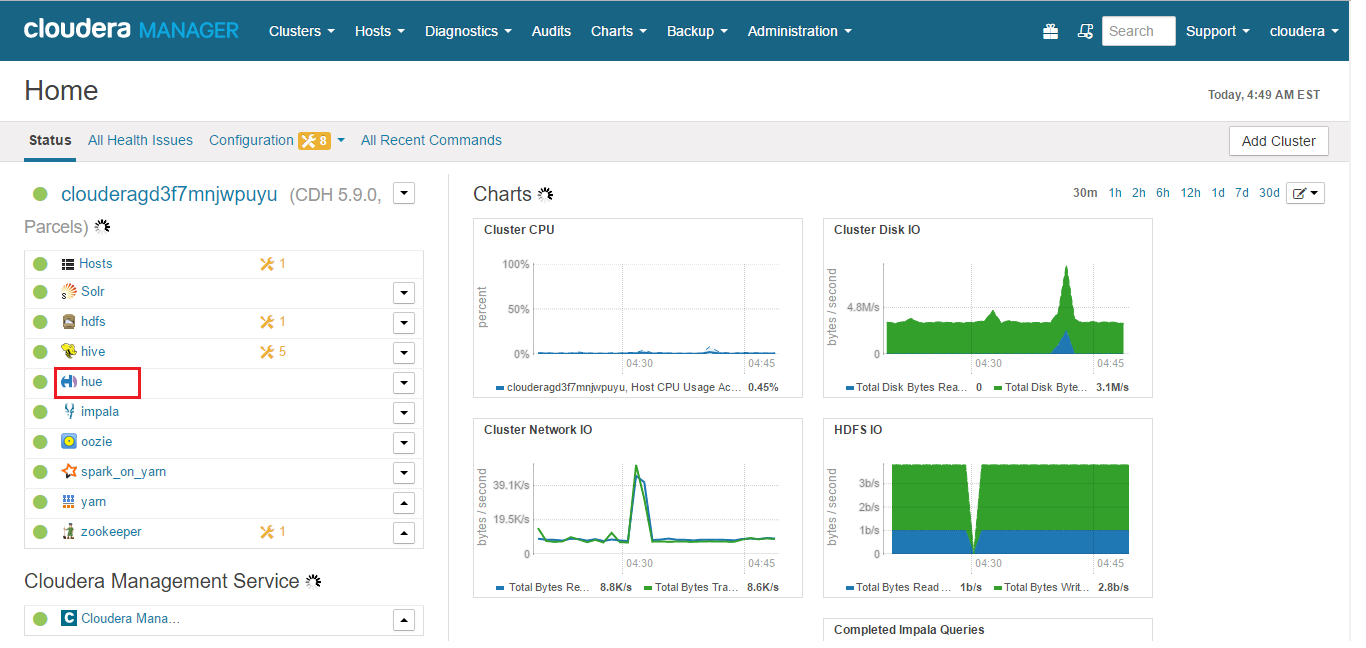
A green symbol indicates the Good Health status of the service



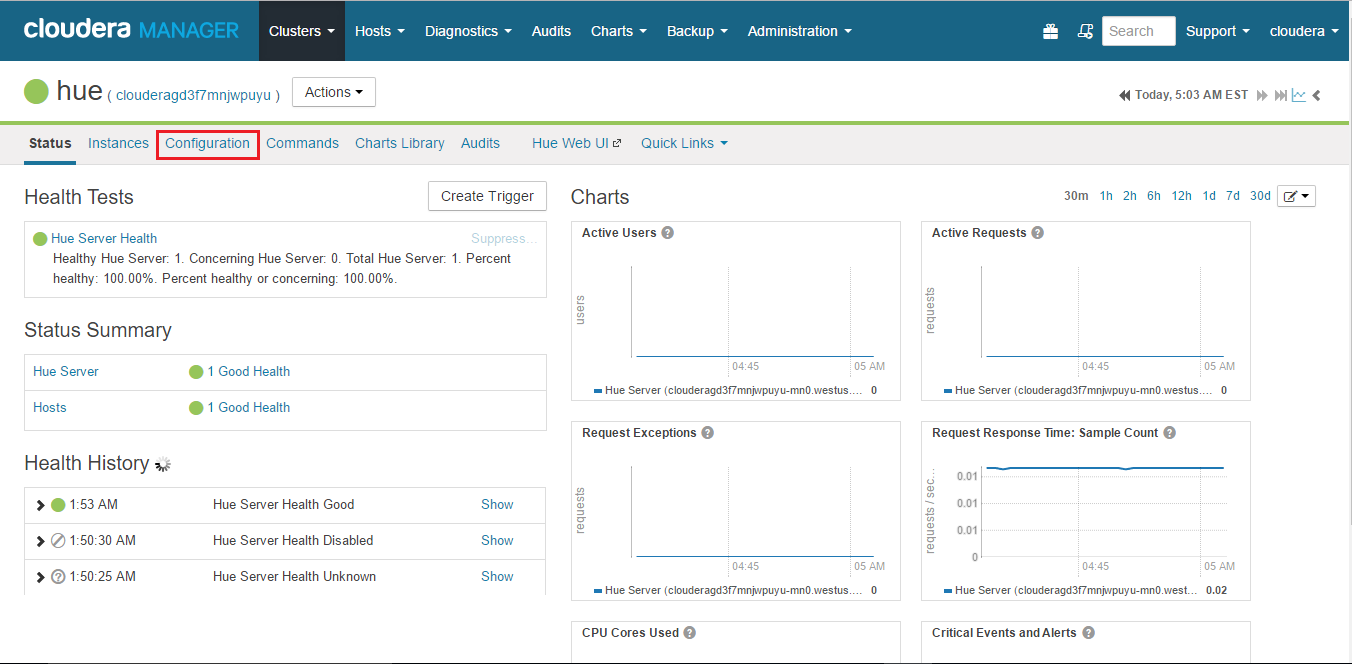
## Configure the Hue Service

From the Services select and click on Hue

Press Ctrl + Click on Hue as shown below:

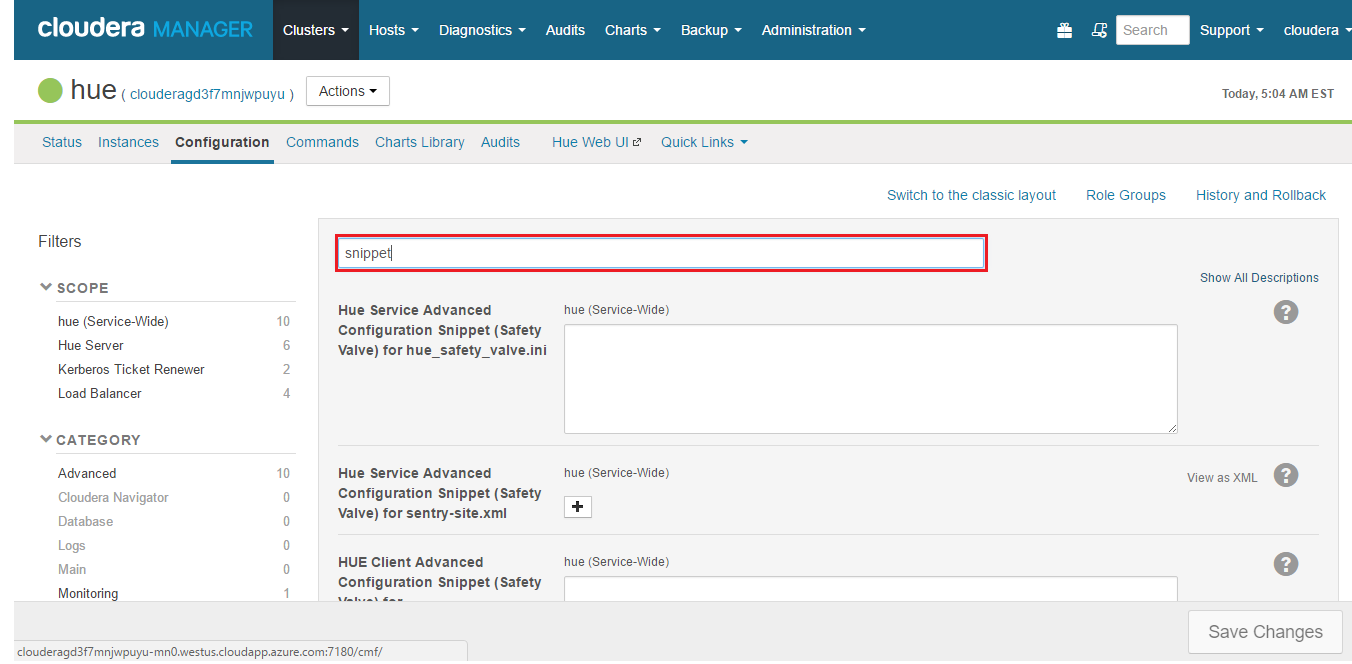


Click the **Configuration** tab.



Type the word "snippet" in the Search box.

A set of Hue advanced configuration snippet properties displays.



Add information about your Solr host to the **Hue Server Configuration Advanced Configuration Snippet for hue\_safety\_valve\_server.ini** property.

Copy and paste the below information in this property.

[desktop]

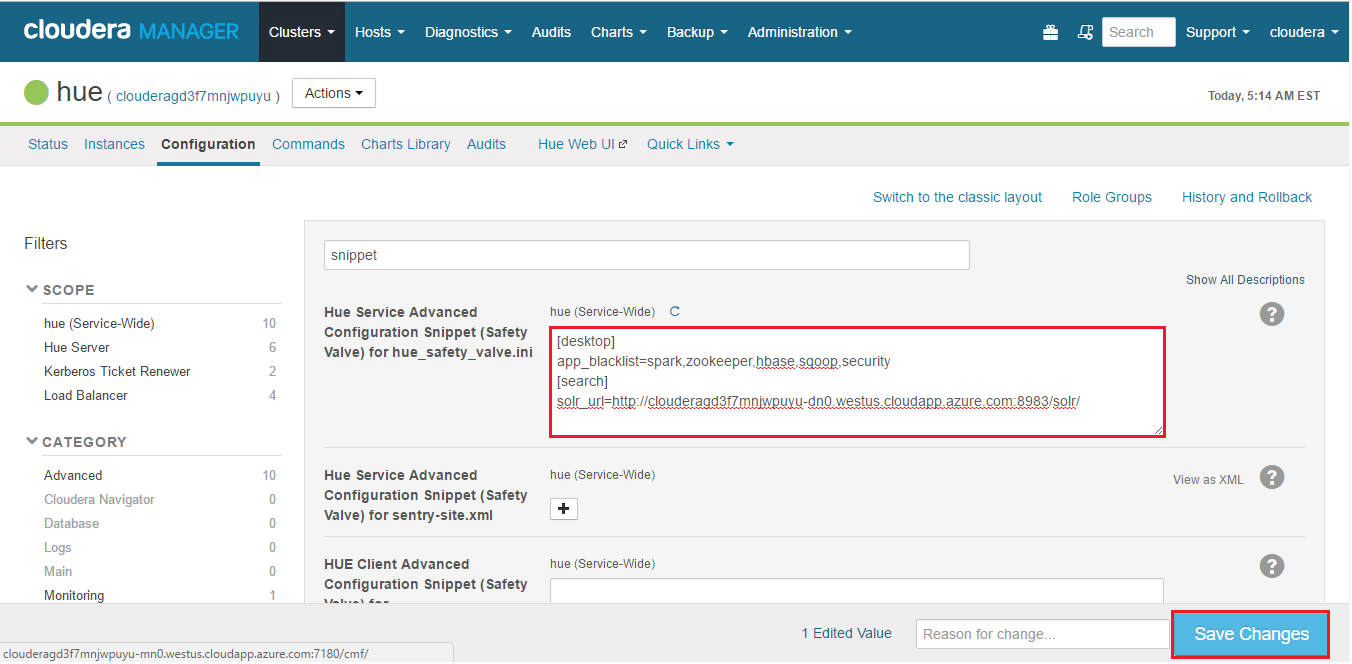
app\_blacklist=spark,zookeeper,hbase,sqoop,security

[search]

solr\_url=http:// clouderat5c8yt3-dn0.westus.cloudapp.azure.com:8983/solr/

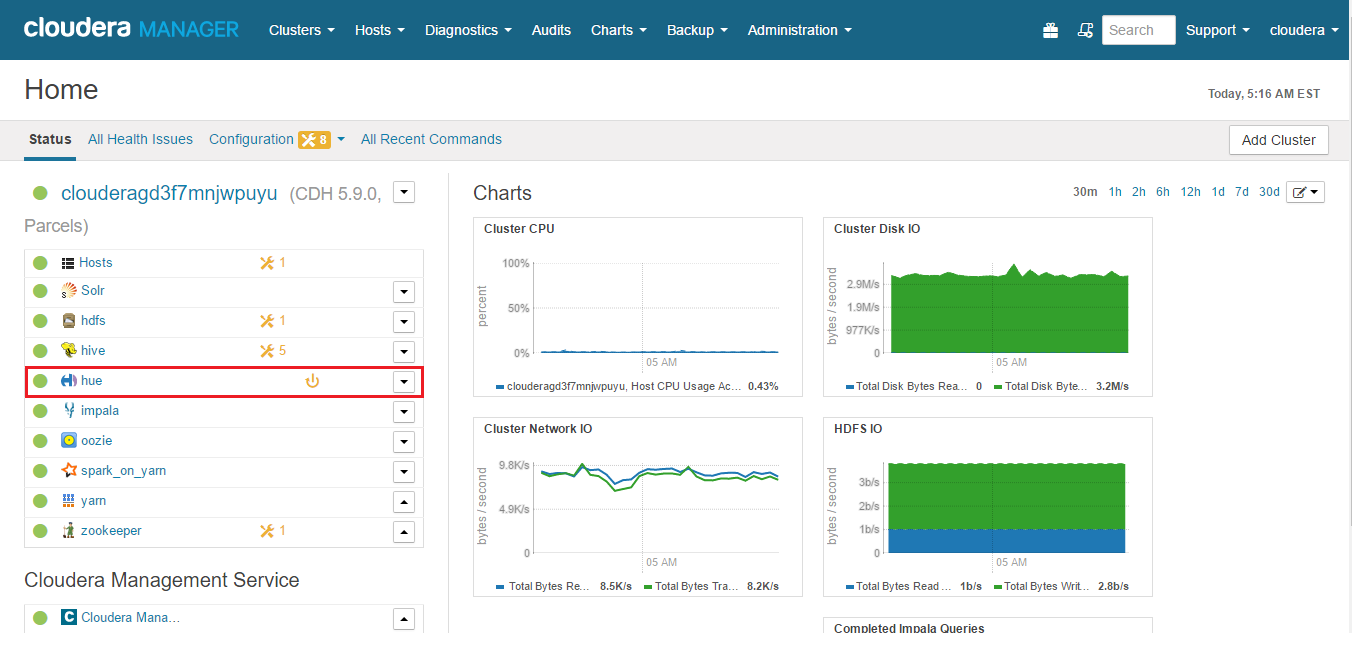
replace the url highlighted in Red with the Solr Url you received in your mail.

Click on Save Changes

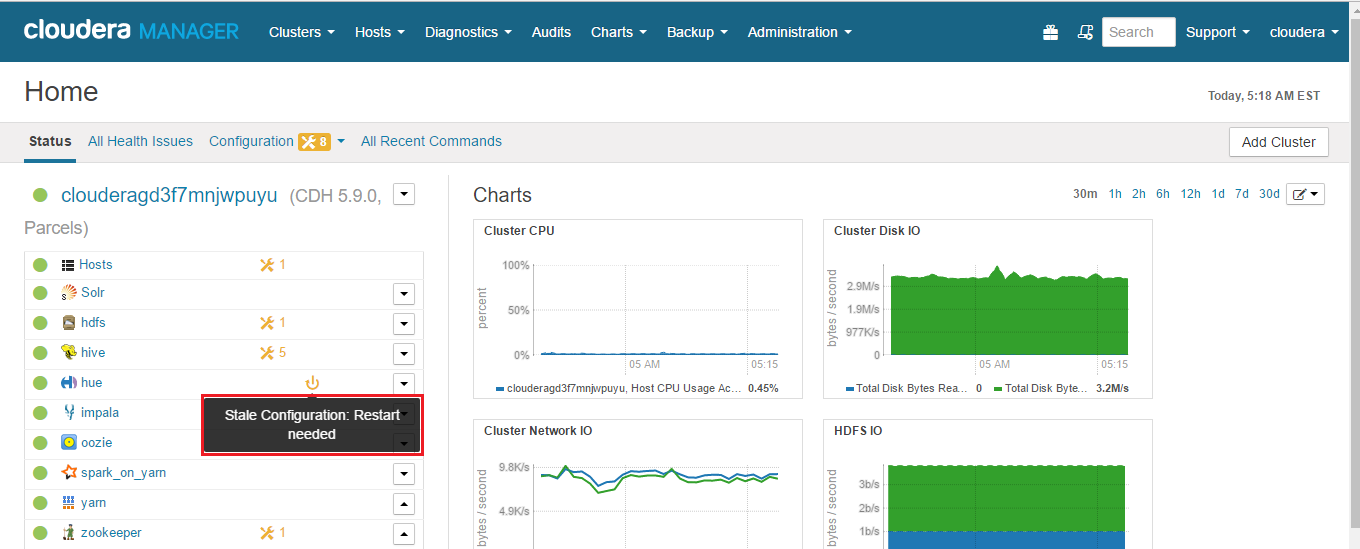


## Restart the Hue Service

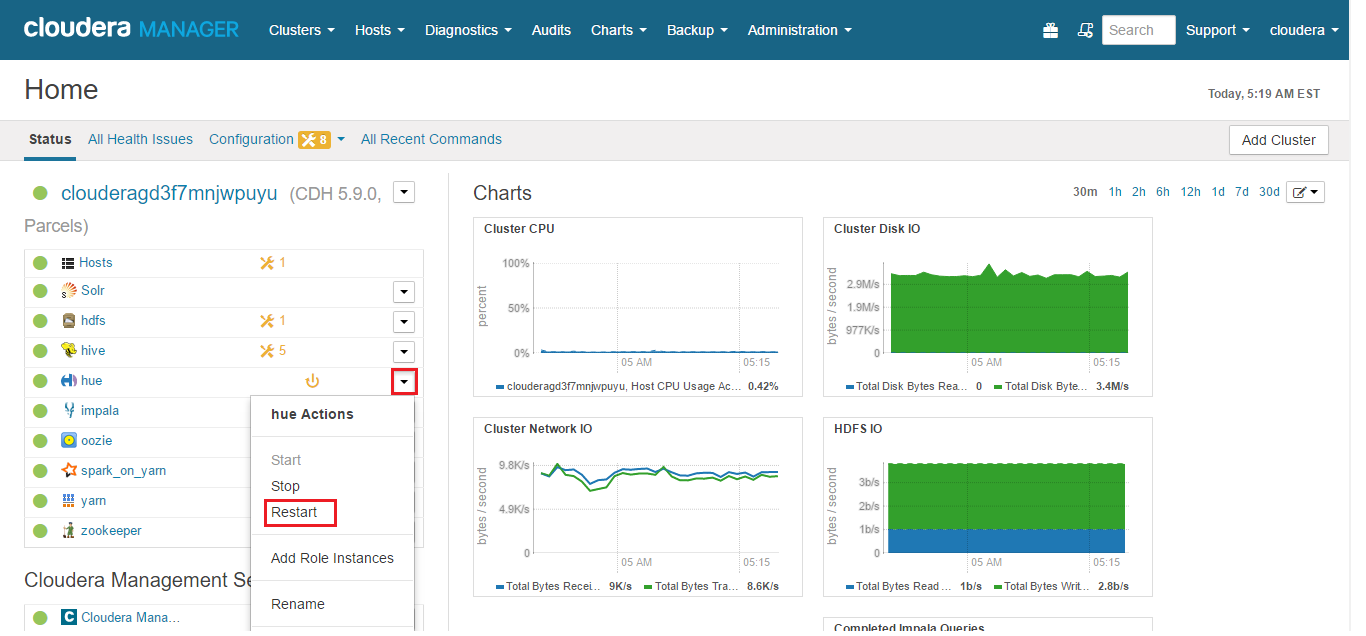
Select the Hue Service,  click  to the right of the Service name and select Restart



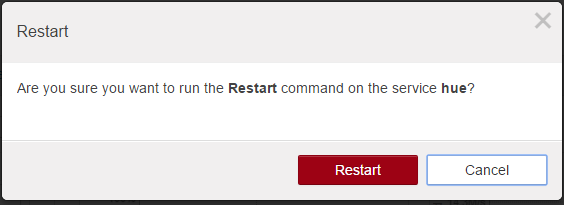
Hue service needs a restart as we made the configuration changes



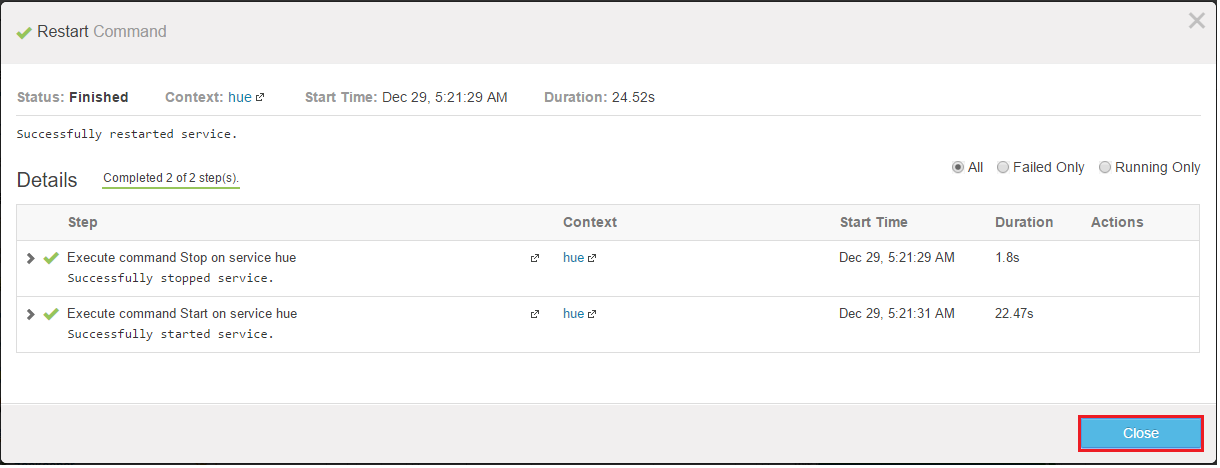
Select the dropdown list on Hue service, and click on Restart as shown below:



Click on Restart



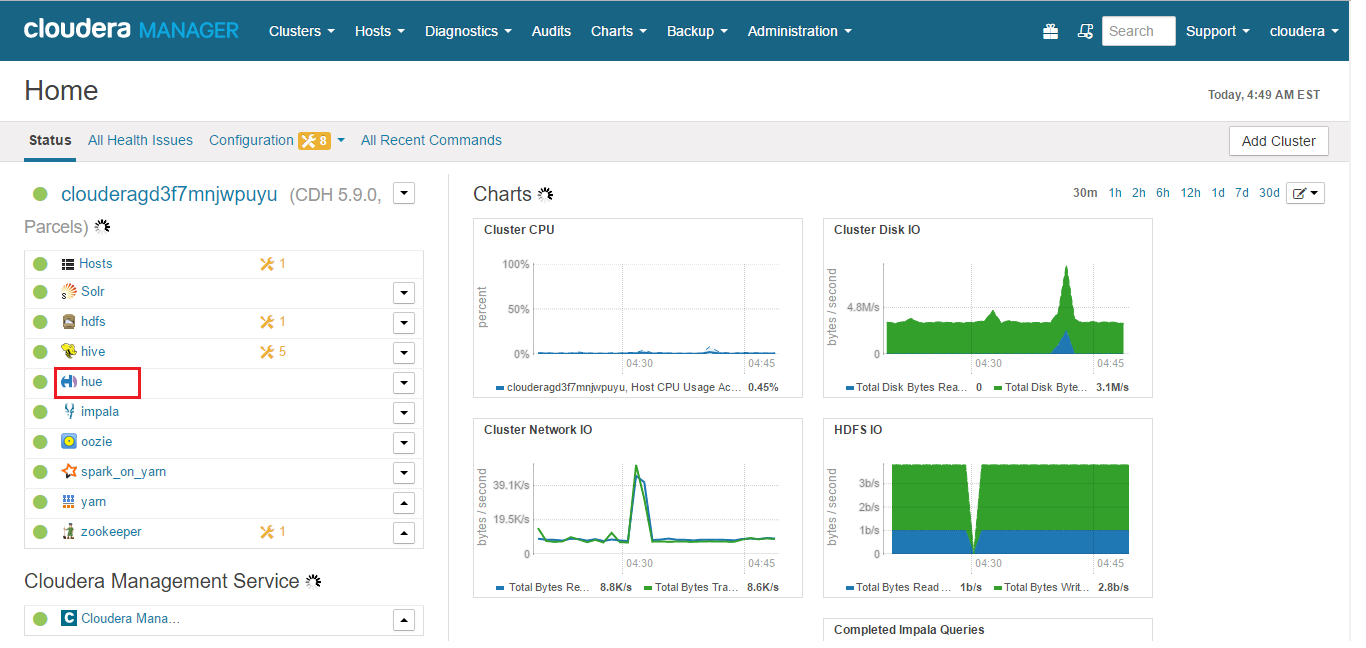
Wait for the service stop and start successfully and click on Close.



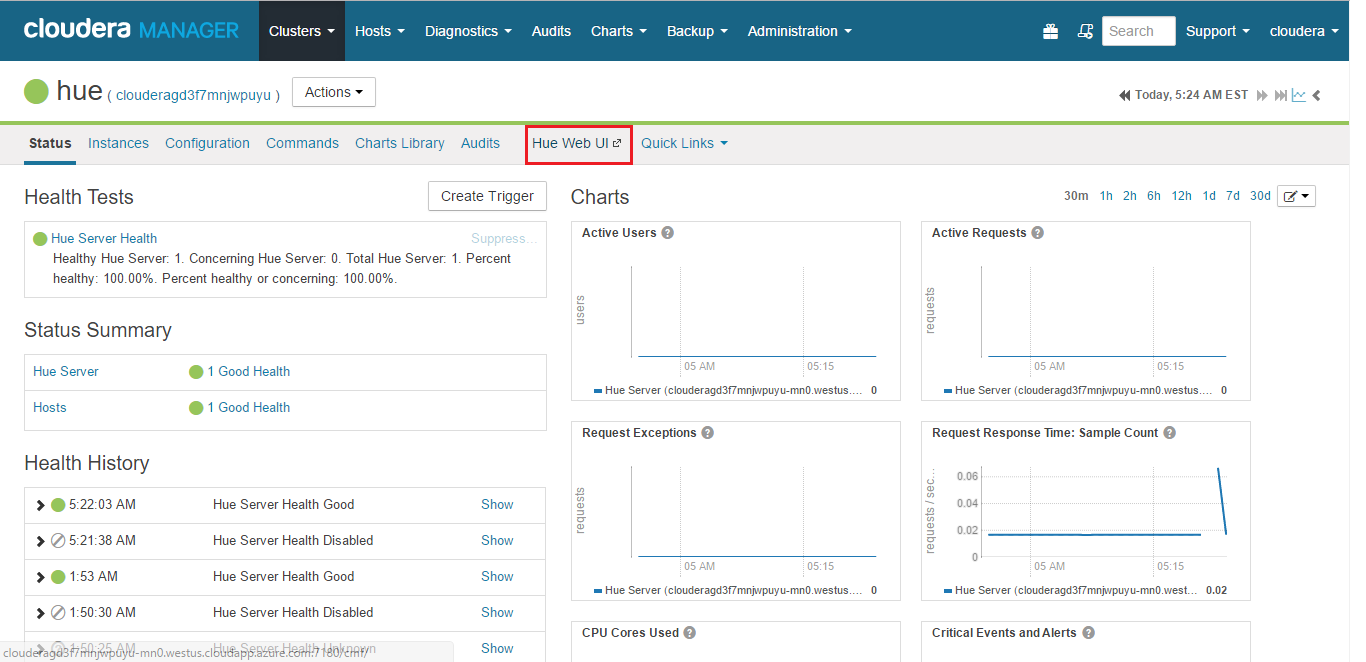
## Login to Hue Web UI

From the Services select Hue service

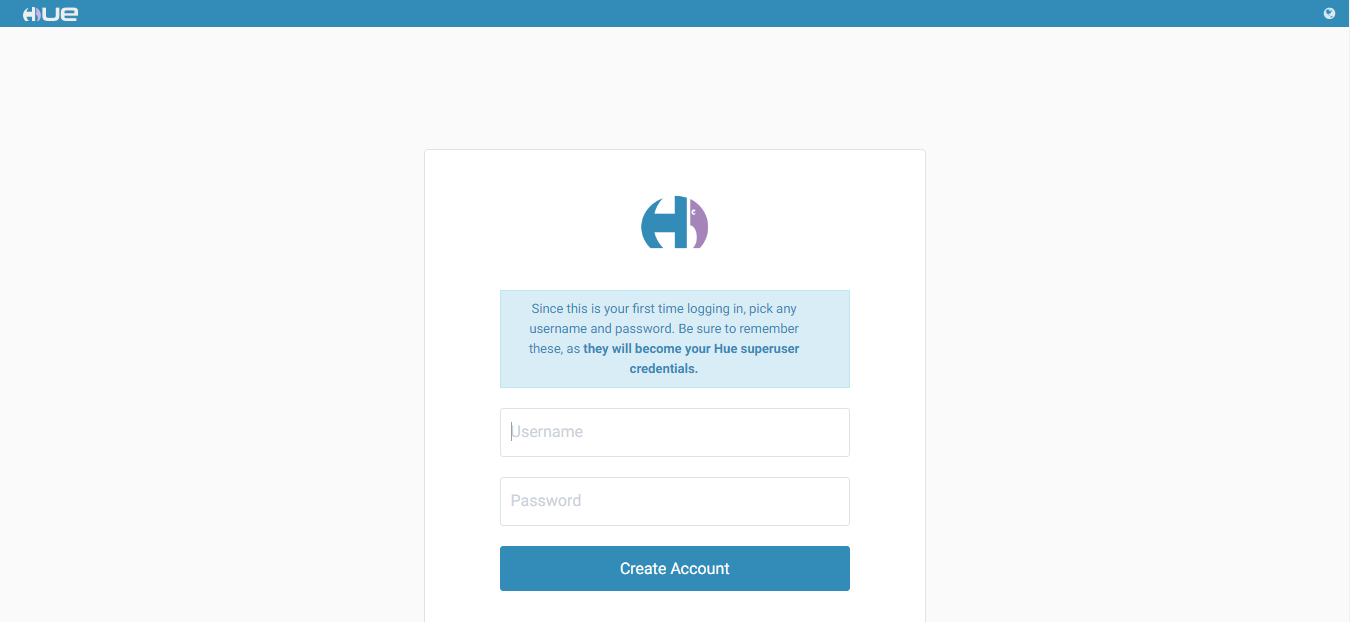
Press ctrl + click on Hue service



Click on Hue Web UI



You will be prompted with a Login screen as below:

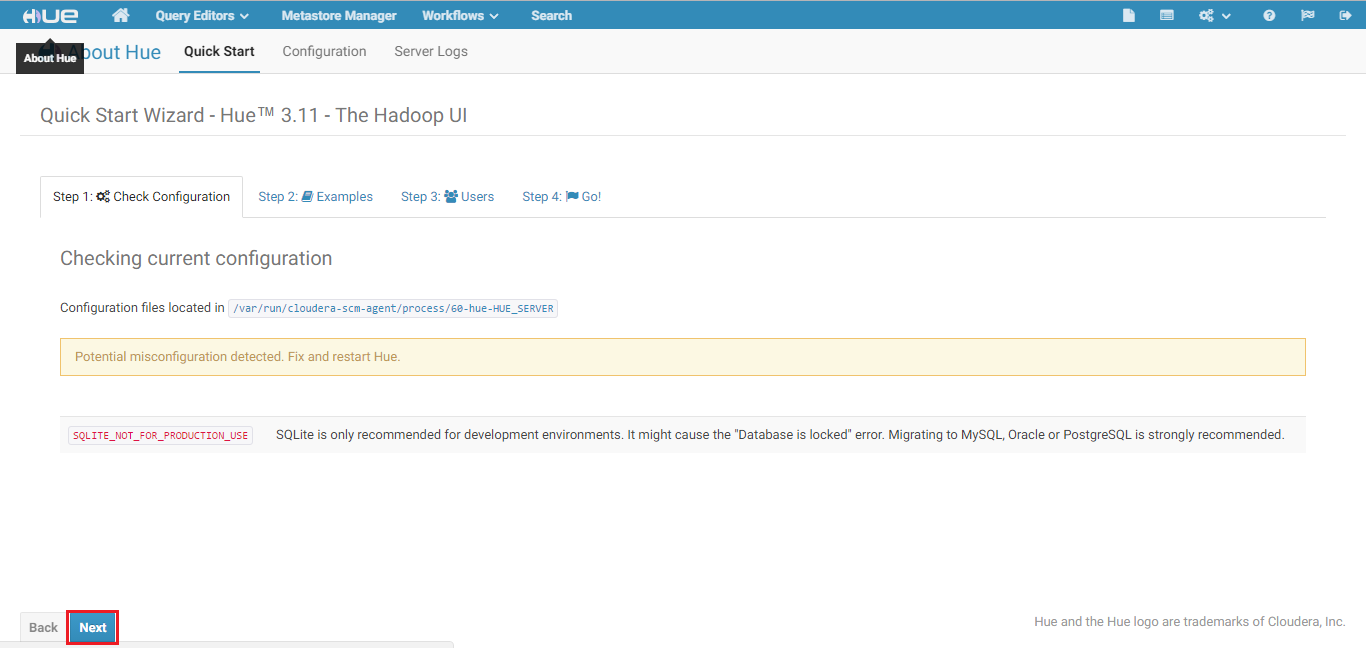


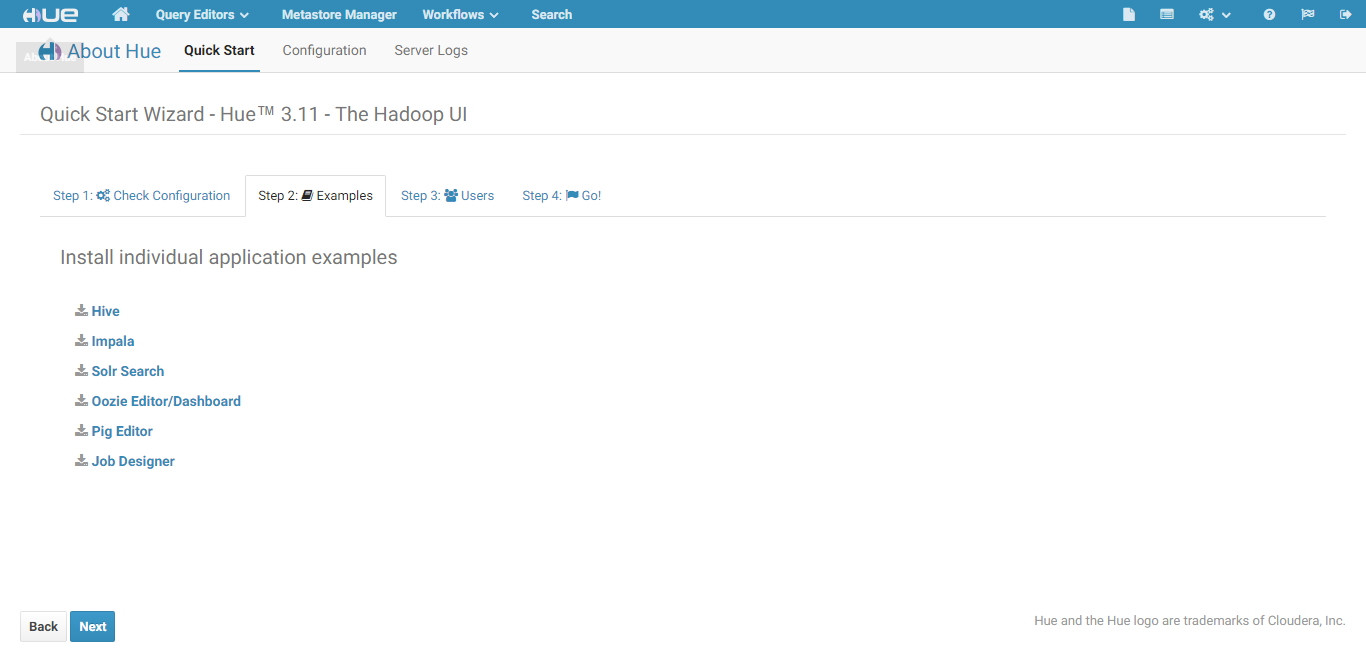
Create any username and password. Be sure to remember these, as **they will become your Hue superuser credentials.**

**Click on Create Account**

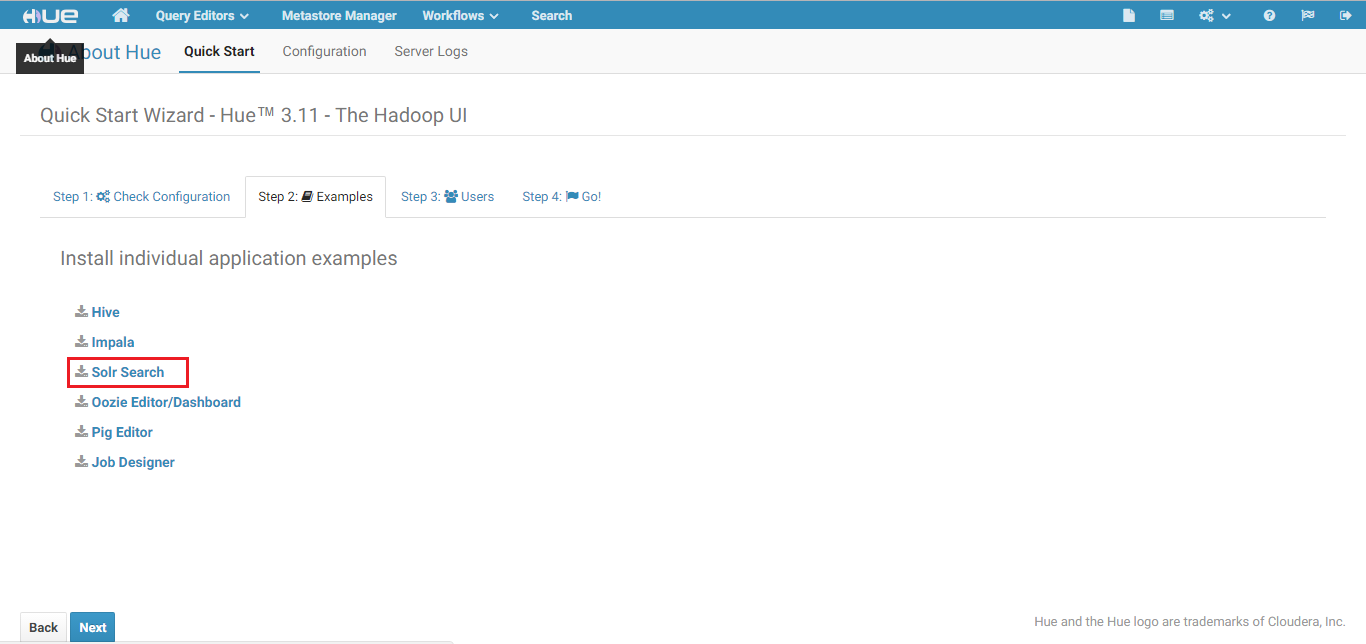
You will be prompted with a Quick start wizard as below:

Click on Next.

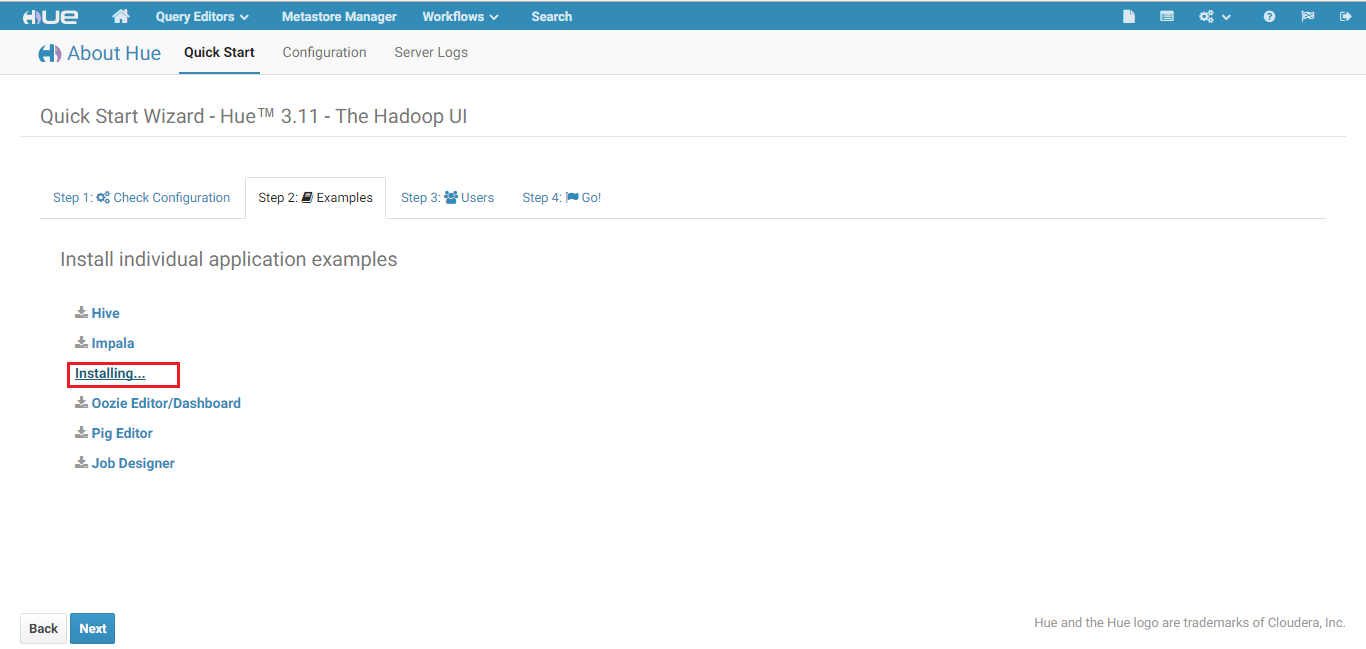




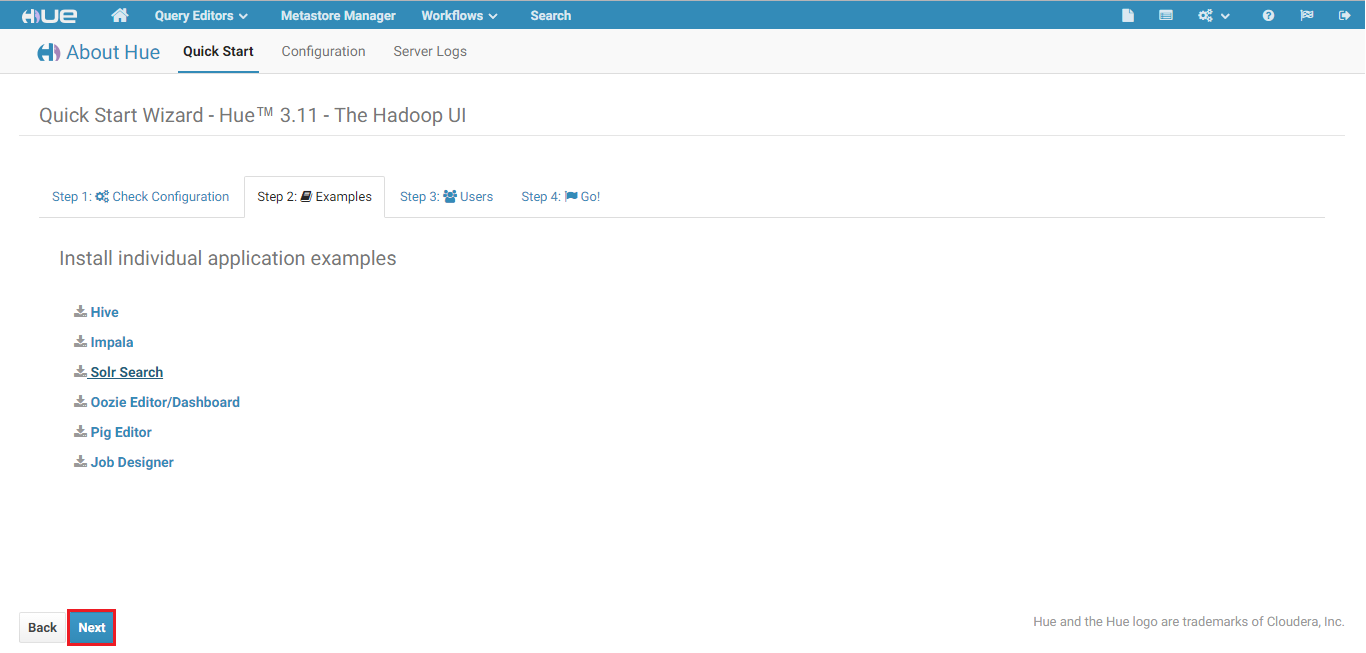
Select Solr Search to install Solr in Hue.



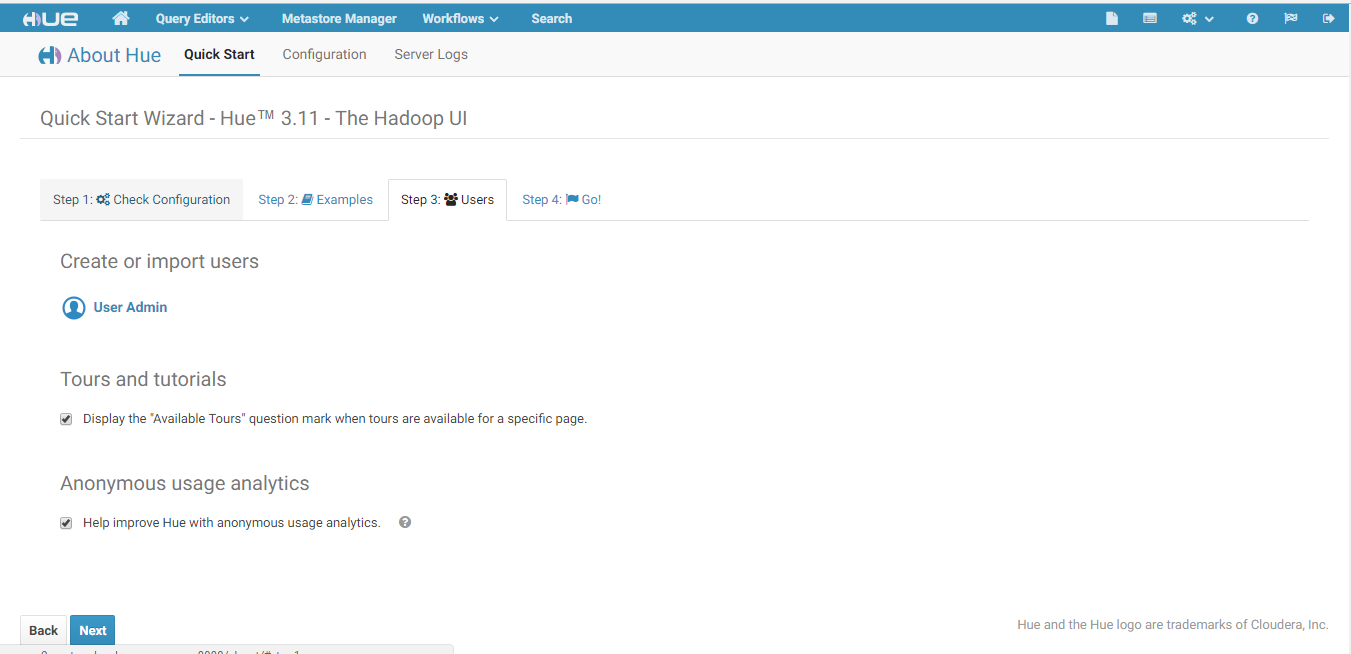
Wait for the Solr service complete installing.



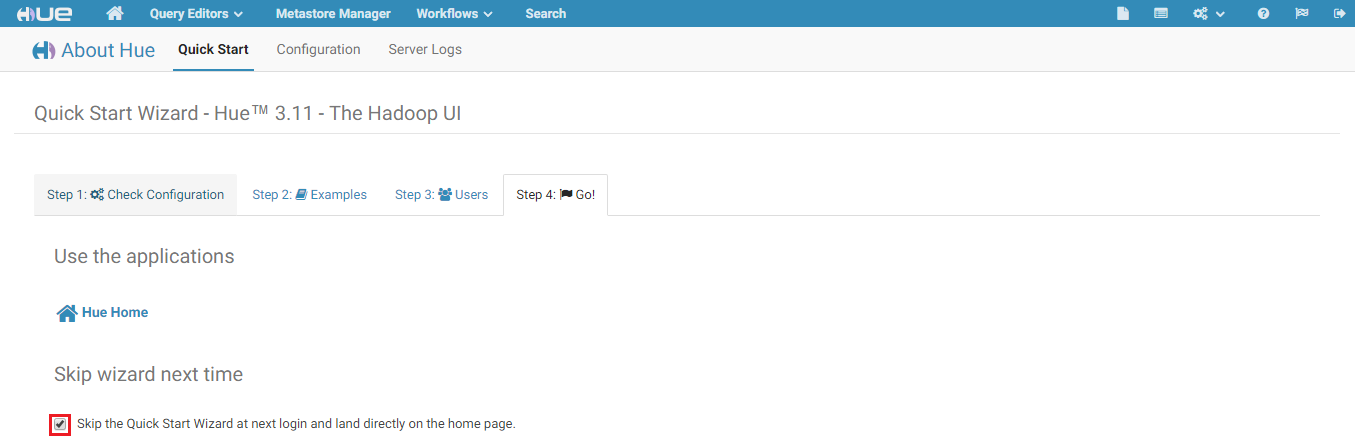
Click on Next



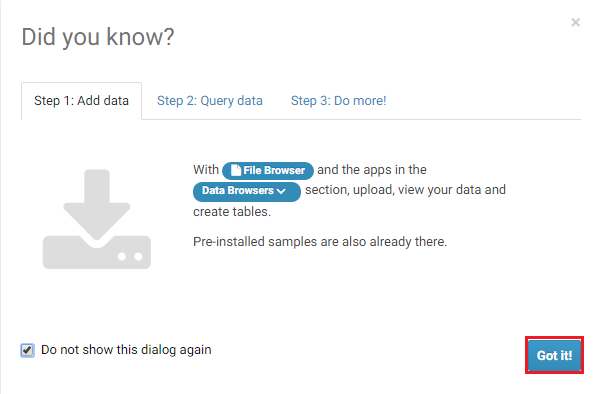
Click on Next



Click the checkbox “Skip the Quick Start Wizard at next login and land directly on the home page” and click Done.



You are prompted with a window to add Data. Click on Got it!



## Injecting Apache logs to Cloudera

You will receive an email with the actual ClouderaManagerUrl.

It will look something like:

clouderau3y4789zxy-mn0.westus.cloudapp.azure.com

SSH into the Cloudera Manager Node using the URL you received for SSH

Login with the below Credentials:

Username: cloudera

Password: Cloudera@123

From the console install the necessary packages and execute the commands one by one:

sudo pip install pyyaml ua-parser

sudo pip install pygeoip

wget <http://geolite.maxmind.com/download/geoip/database/GeoLiteCity.dat.gz>

gunzip GeoLiteCity.dat.gz

chmod 777 GeoLiteCity.dat

wget <https://raw.githubusercontent.com/romainr/hadoop-tutorials-examples/master/search/indexing/apache_logs.py>

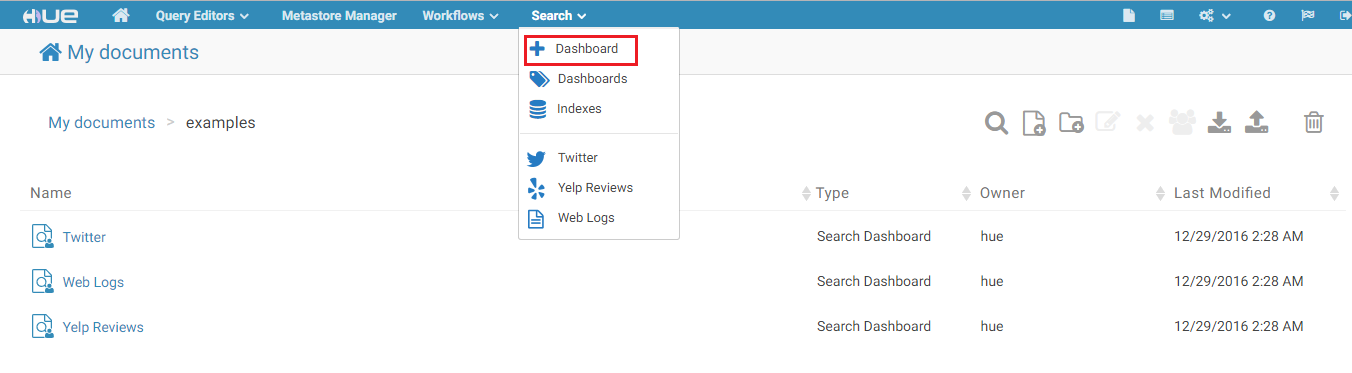
sudo chmod +x apache\_logs.py

wget <https://raw.githubusercontent.com/cloudera/hue/master/apps/search/examples/collections/solr_configs_log_analytics_demo/index_data.csv>

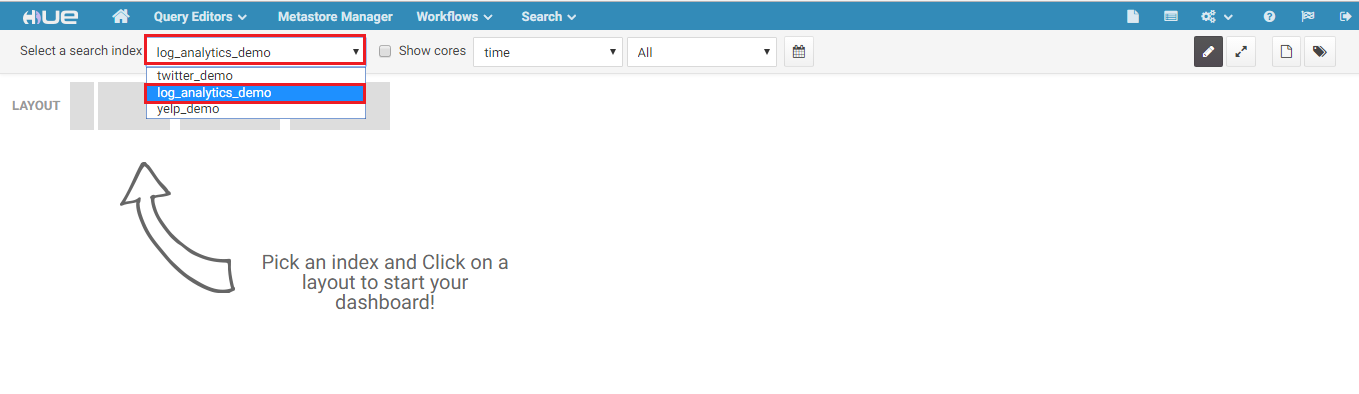
hdfs dfs -put index\_data.csv /user/cloudera/

## Configure Search Dashboard

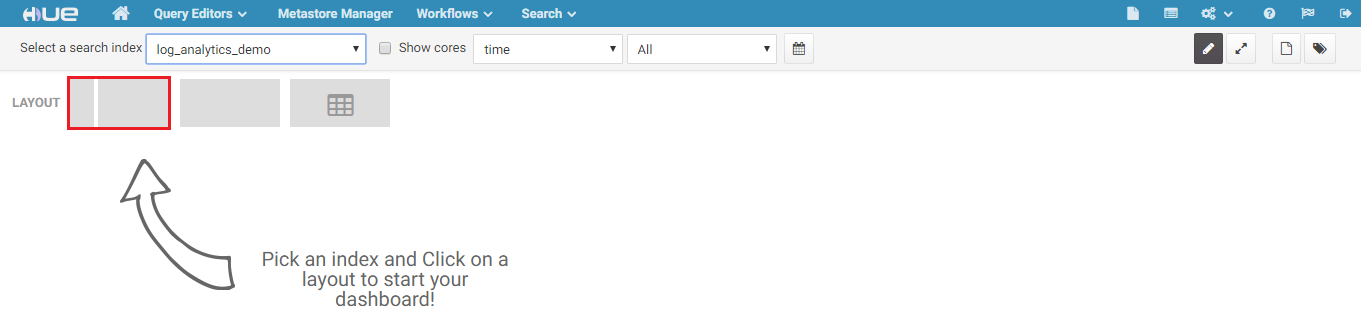
Select the search and select on Dashboard as shown below:



Select the search Index log\_analytics\_demo as shown below:

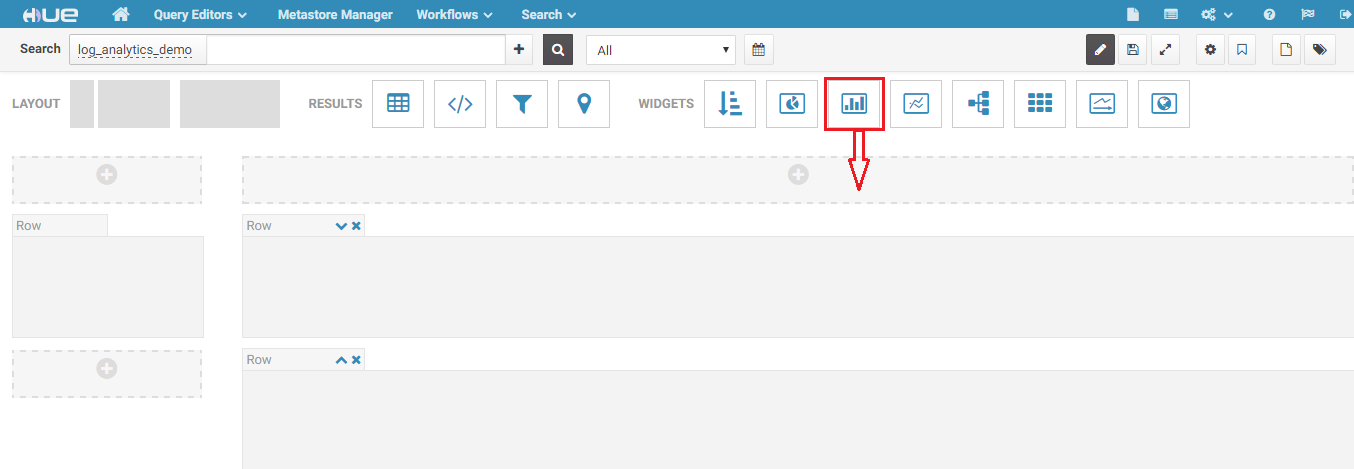


Select the layout as shown below:

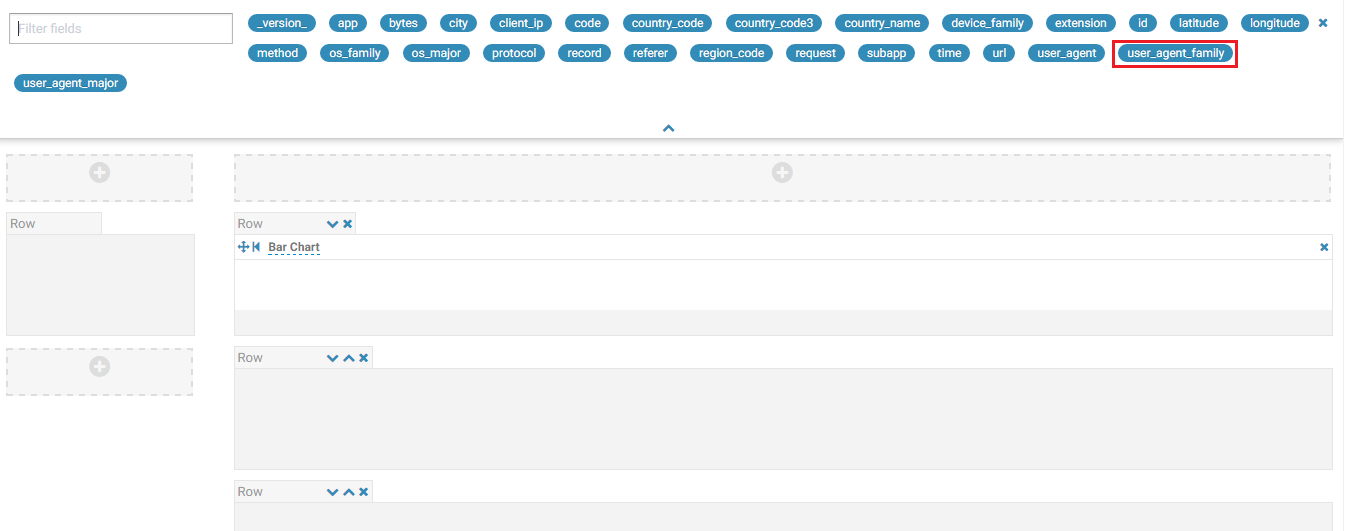


To analyze the browsers that your website users are using,

Select the bargraph as shown below and drag it to the Row



Select the attribute user\_agent\_family as shown below:



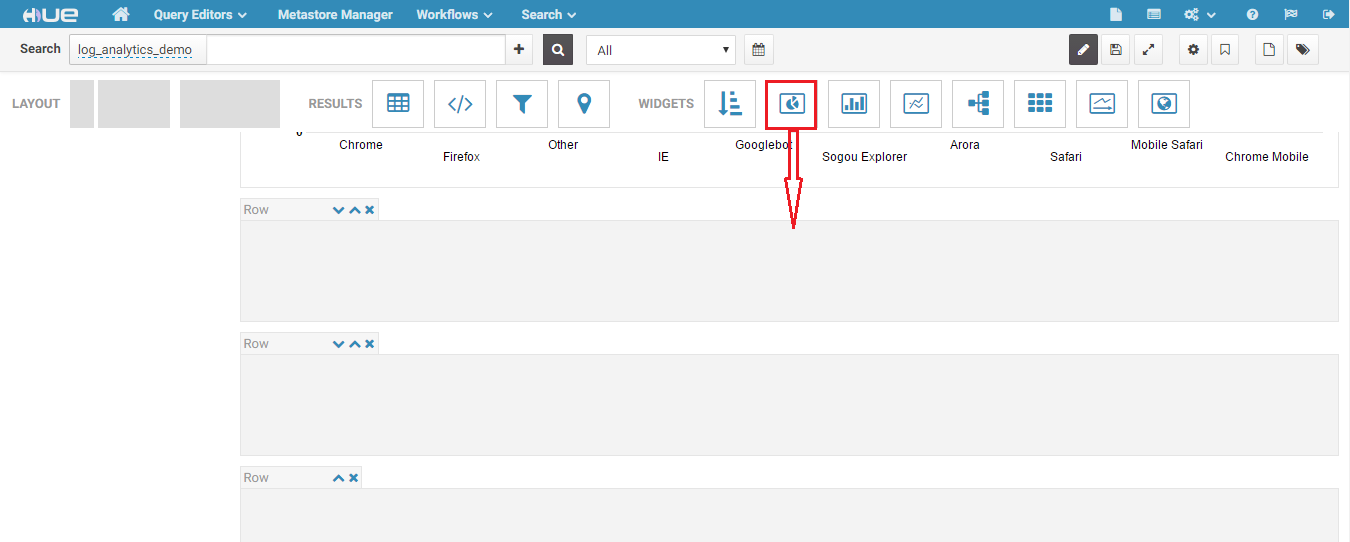
Hue spins the dashboard as a barchart indicating the browsers that the users are using to access your website.



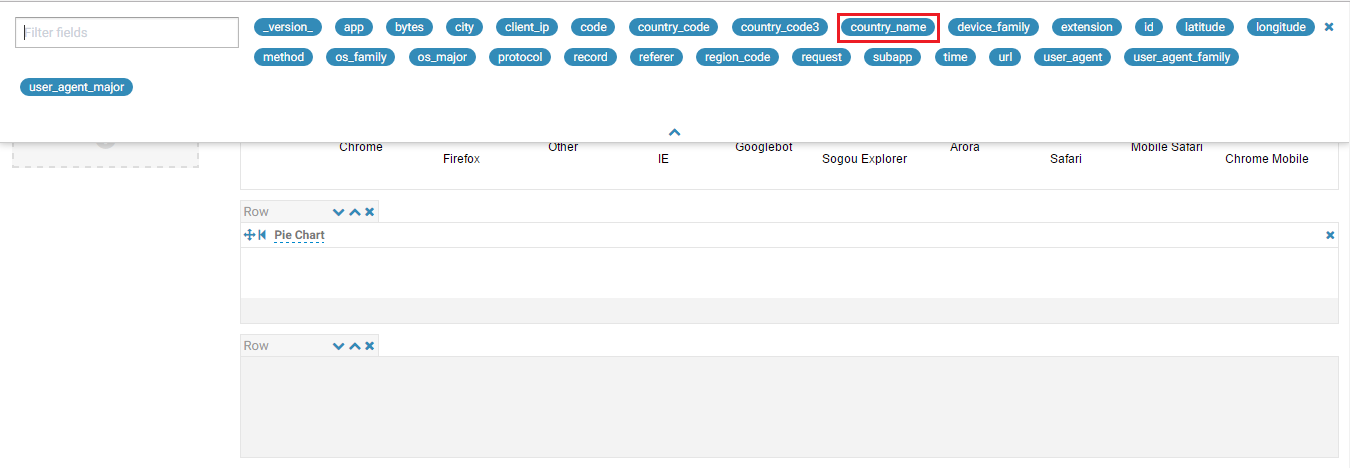
To add more analytics into your Dashboard, scroll down and add more.

To analyze the user traffic coming from various countries,

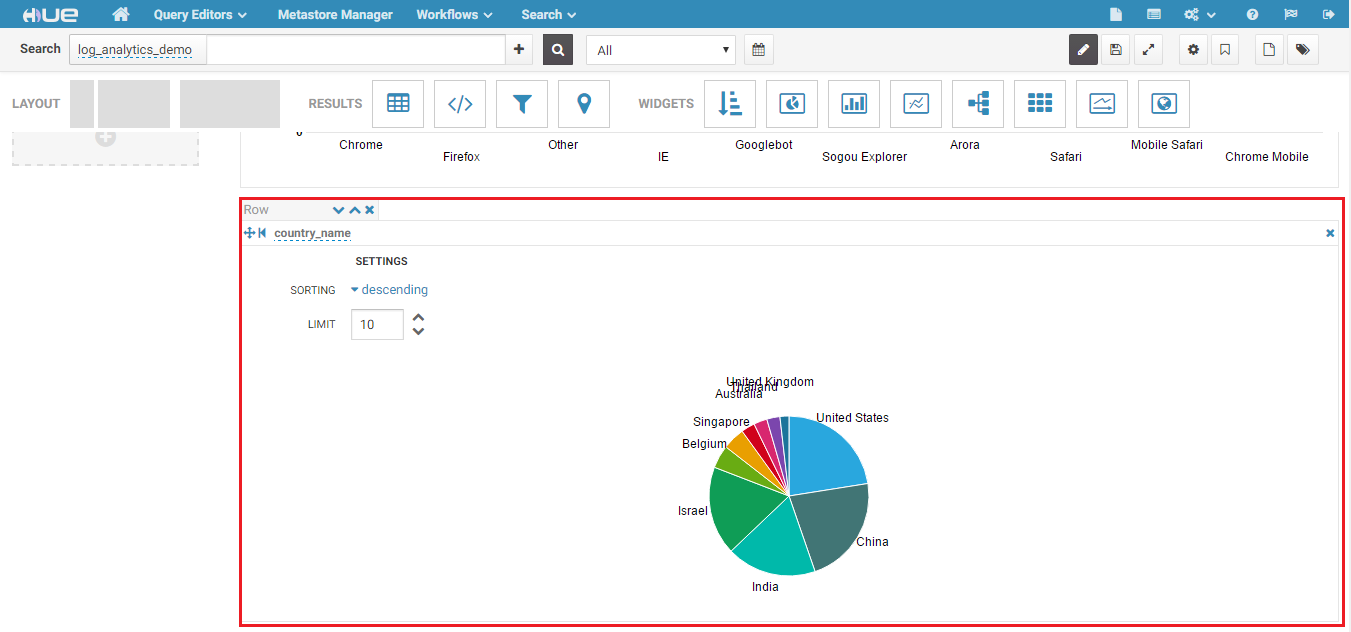
Drag the Pie Chart into the Row as shown below:



Select the country name attribute



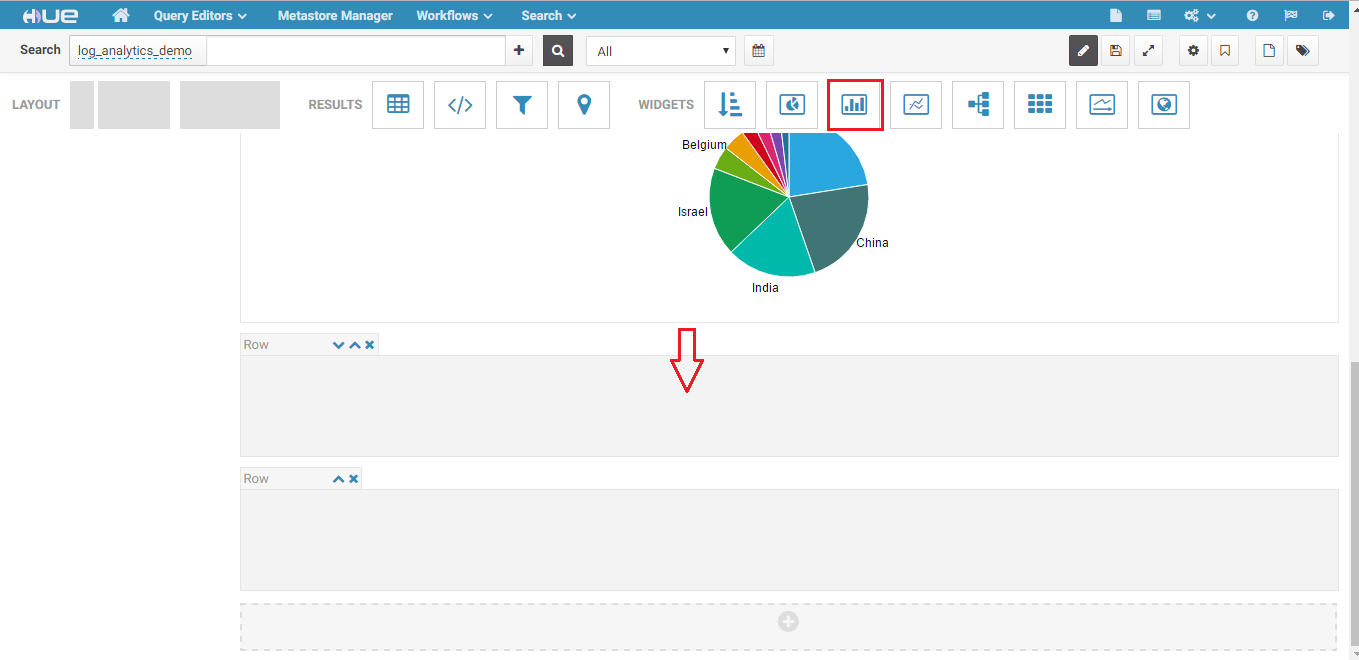
You can observe the piechart showing the analysis of the users accessing your website from various countries.



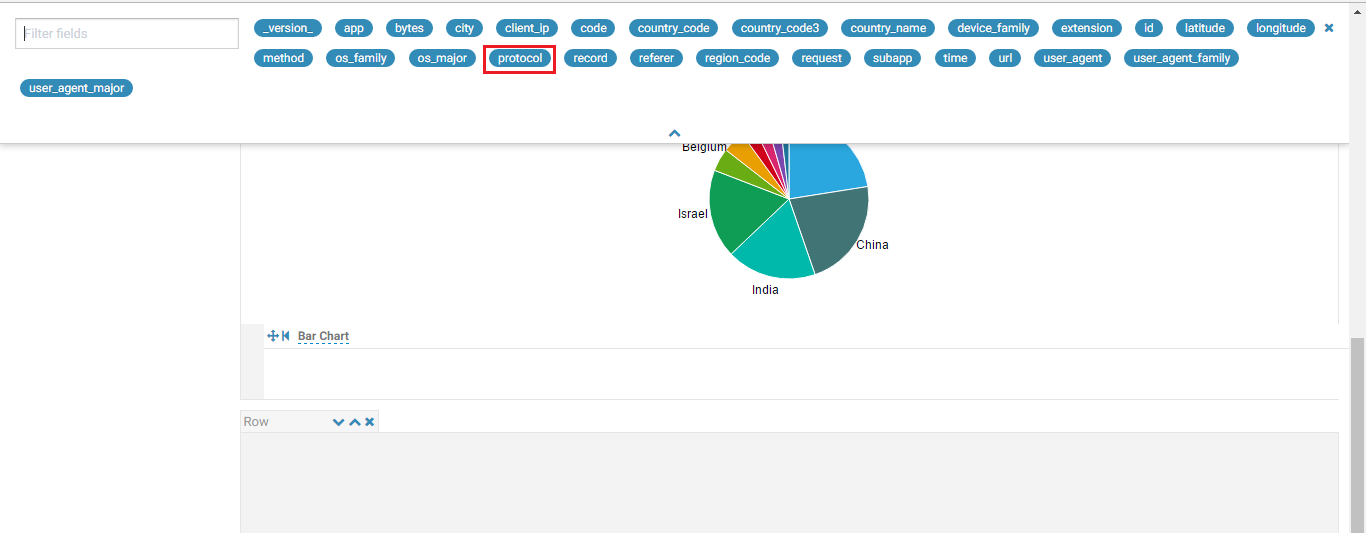
To add more analytics into your Dashboard, scroll down and add more.

To analyze the protocol used by the users accessing your website,

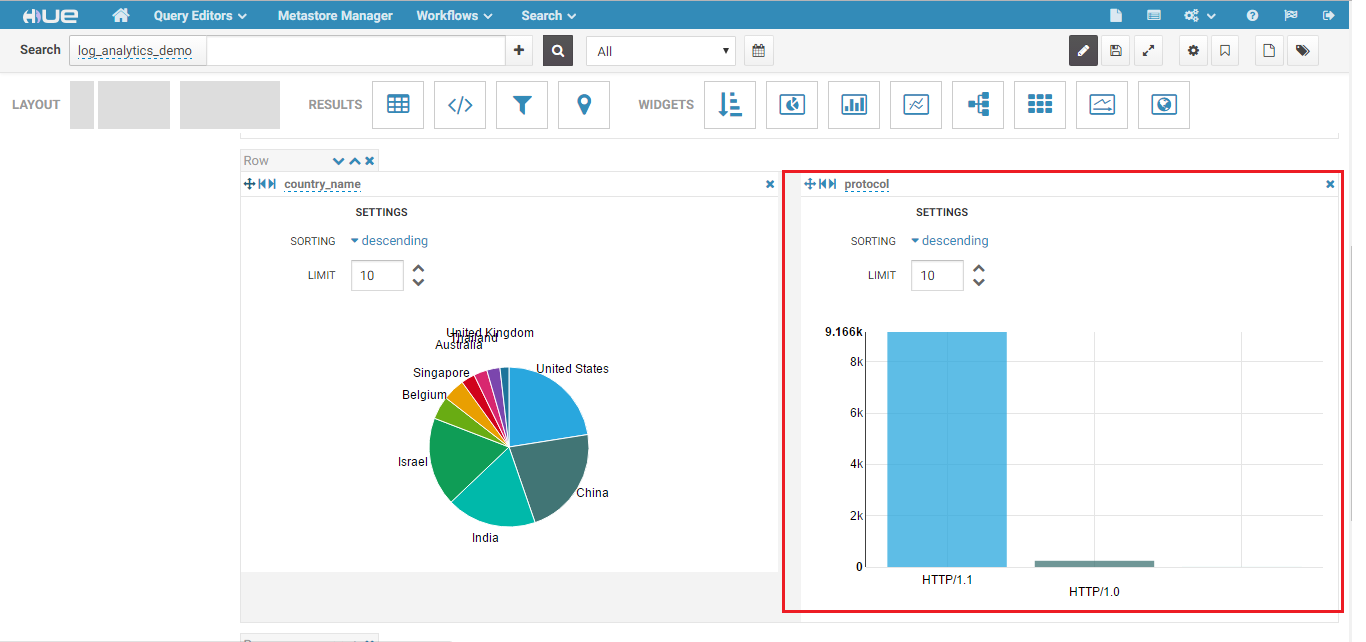
Drag the Bar Chart into the Row as shown below:



Select the protocol attribute



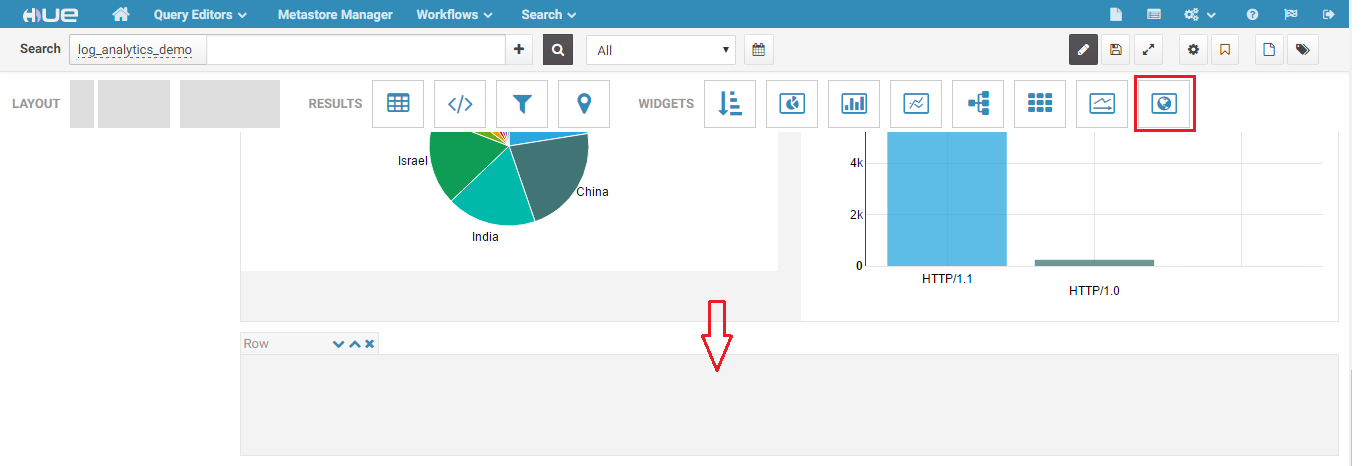
The barchart indicates the HTTP protocol used by the users accessing your website



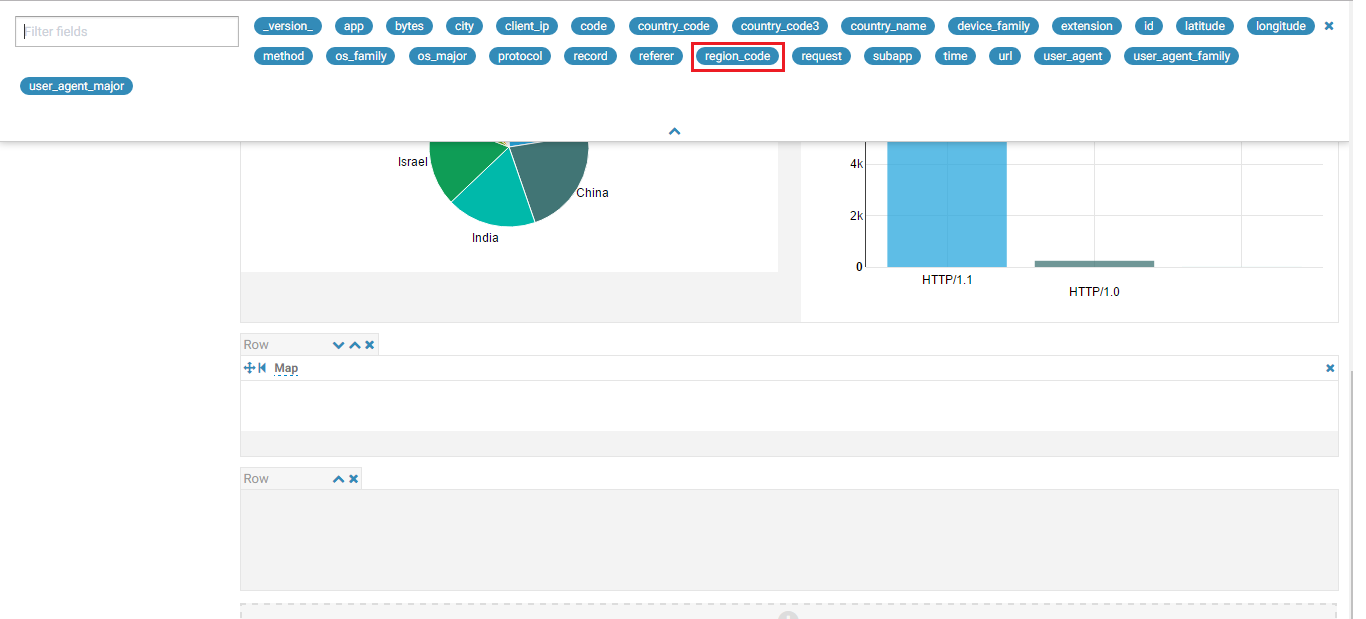
To add more analytics into your Dashboard, scroll down and add more.

To analyze the Operating System used by the users accessing your website,

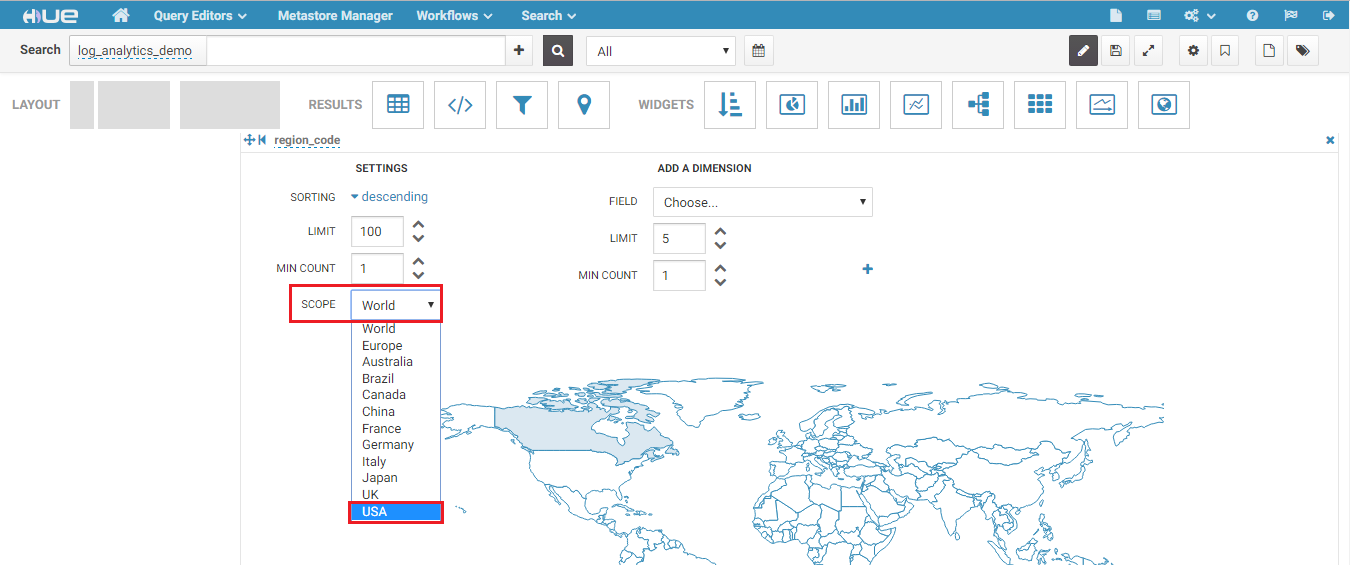
Drag the Map into the Row as shown below:



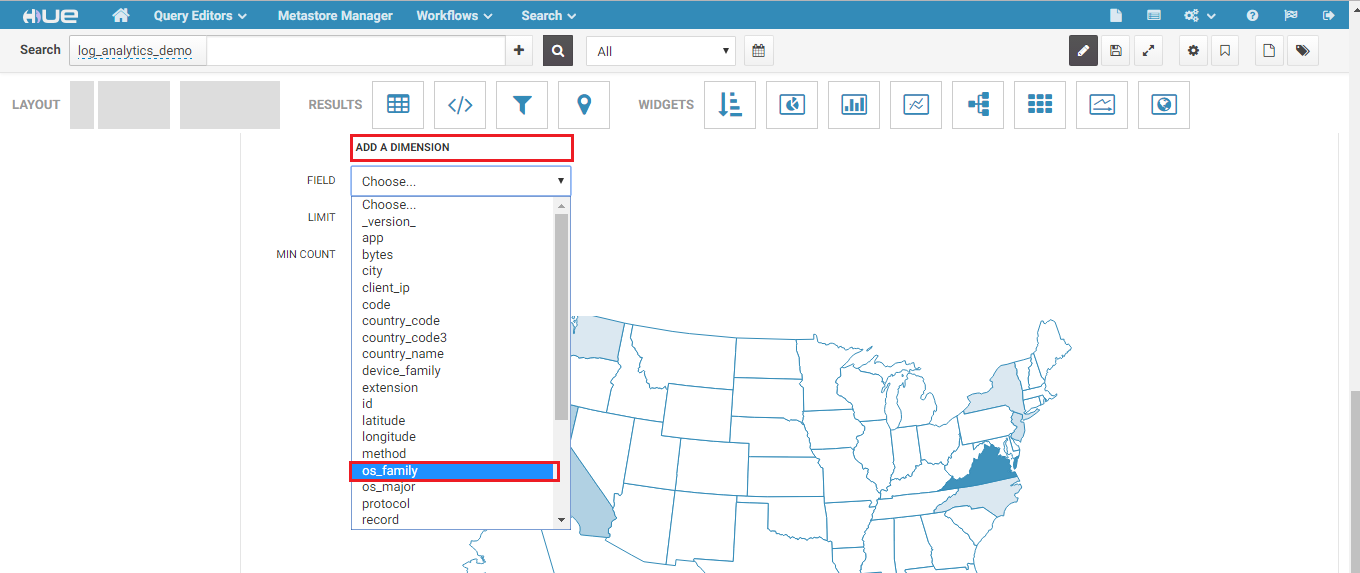
Select the region\_code attribute



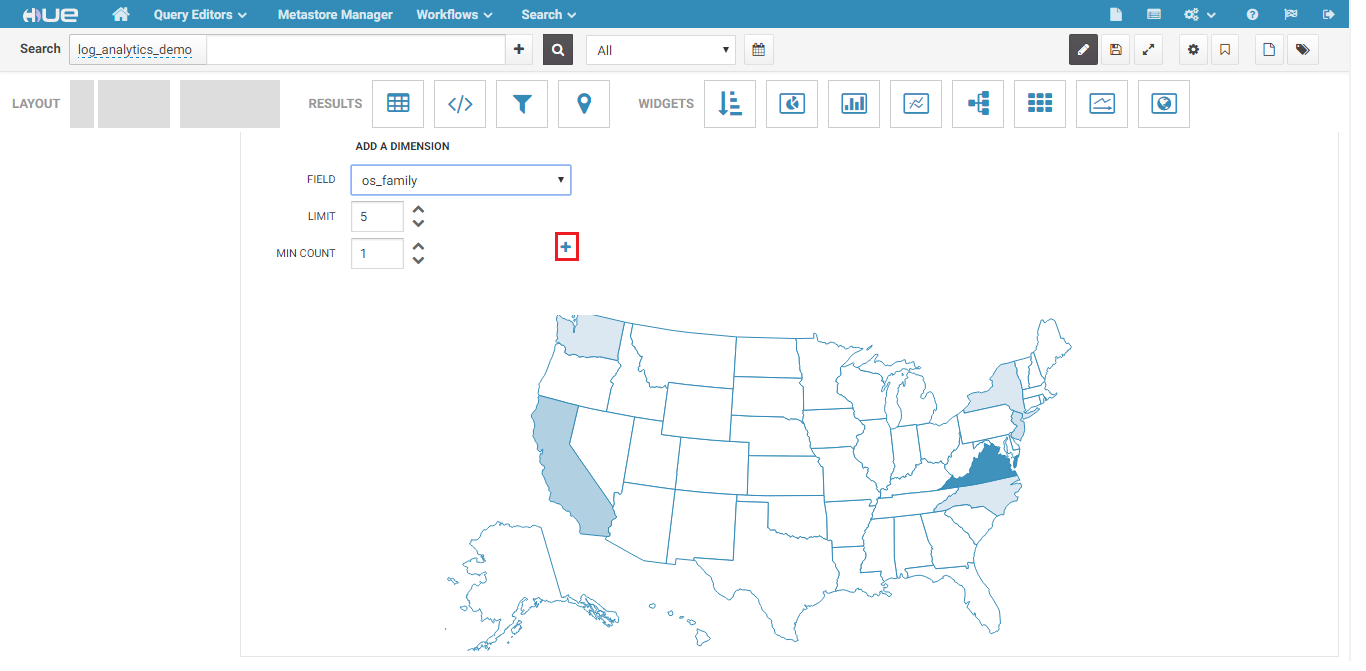
Select scope as USA as shown below:



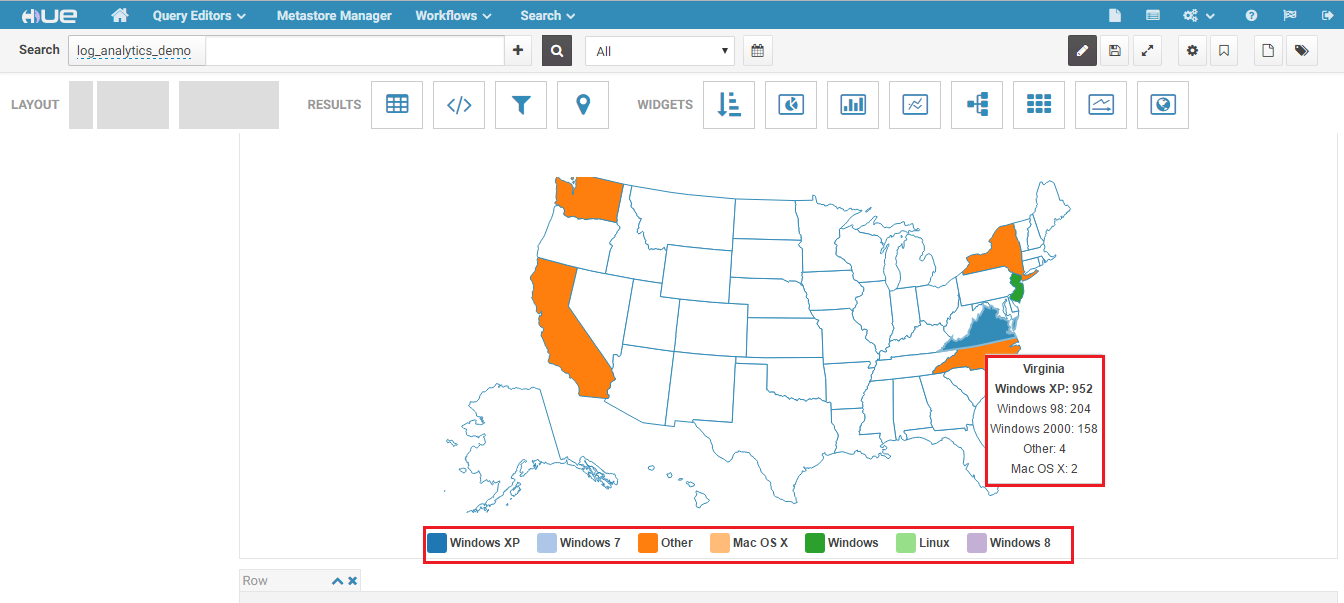
Add a Dimension as shown below and select os\_family attribute from the drop down list:



Click on + Button as shown below:



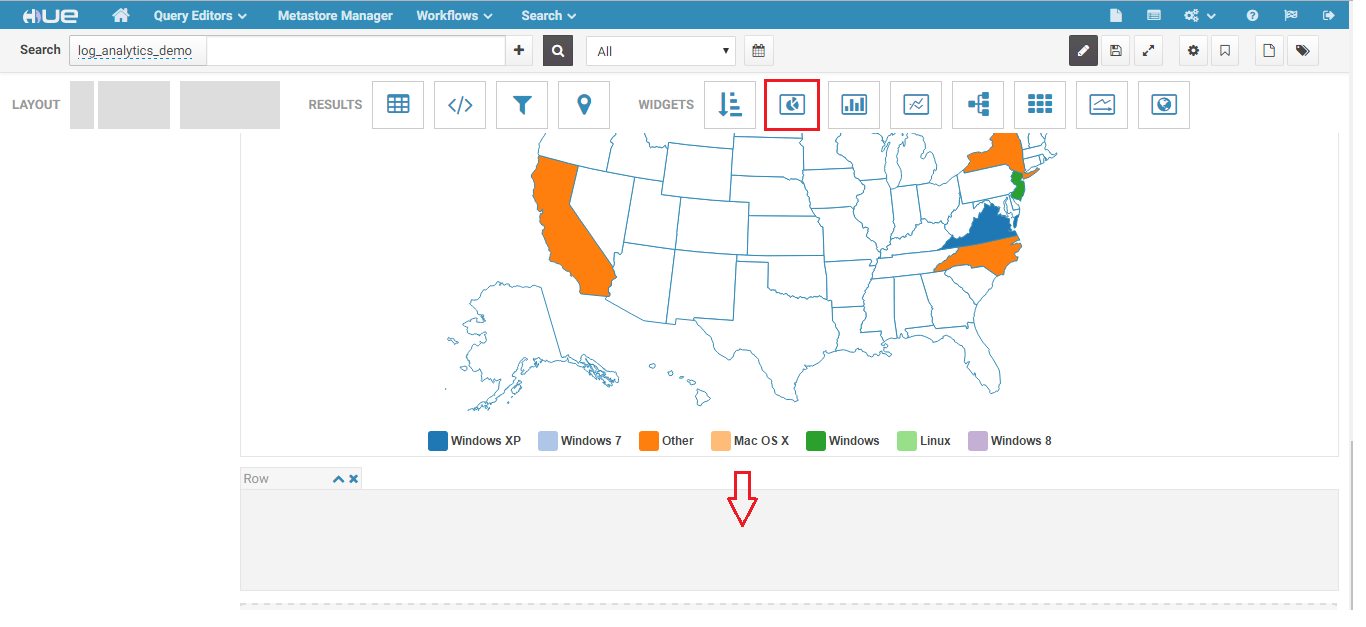
Point your mouse to the color shaded regions and you can observe the statistics of operating systems used by the Users accessing your website when you scroll on each region.



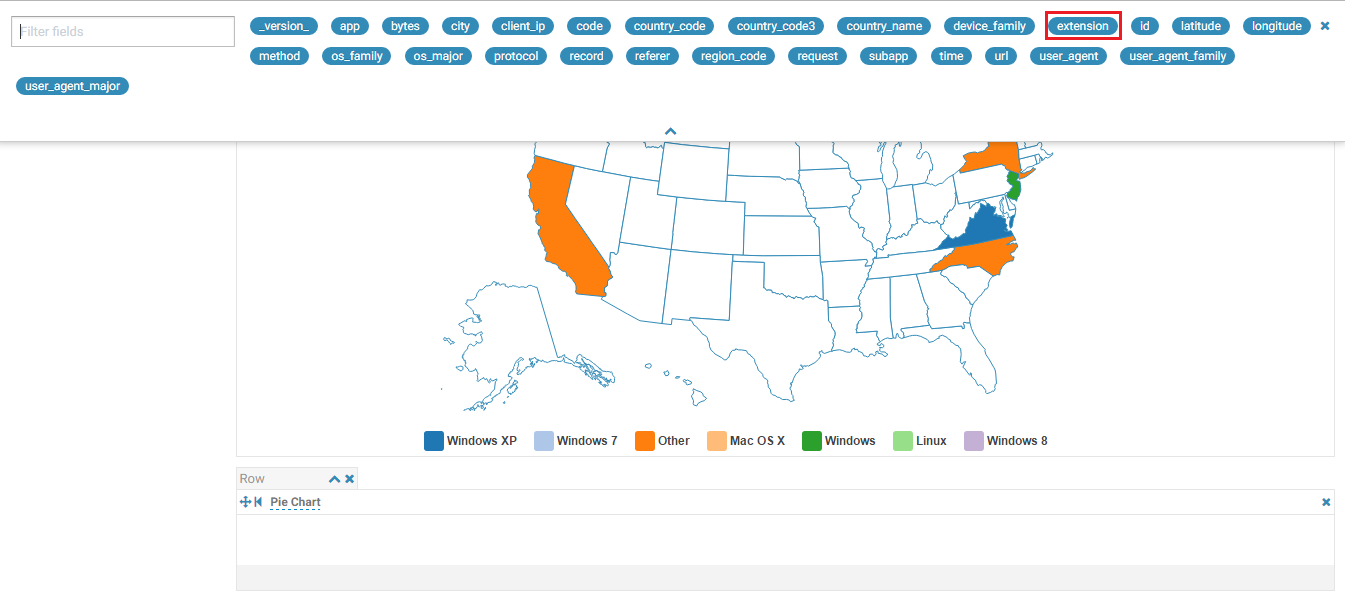
To add more analytics into your Dashboard, scroll down and add more.

To analyze the type of webpages are accessed by the users accessing your website,

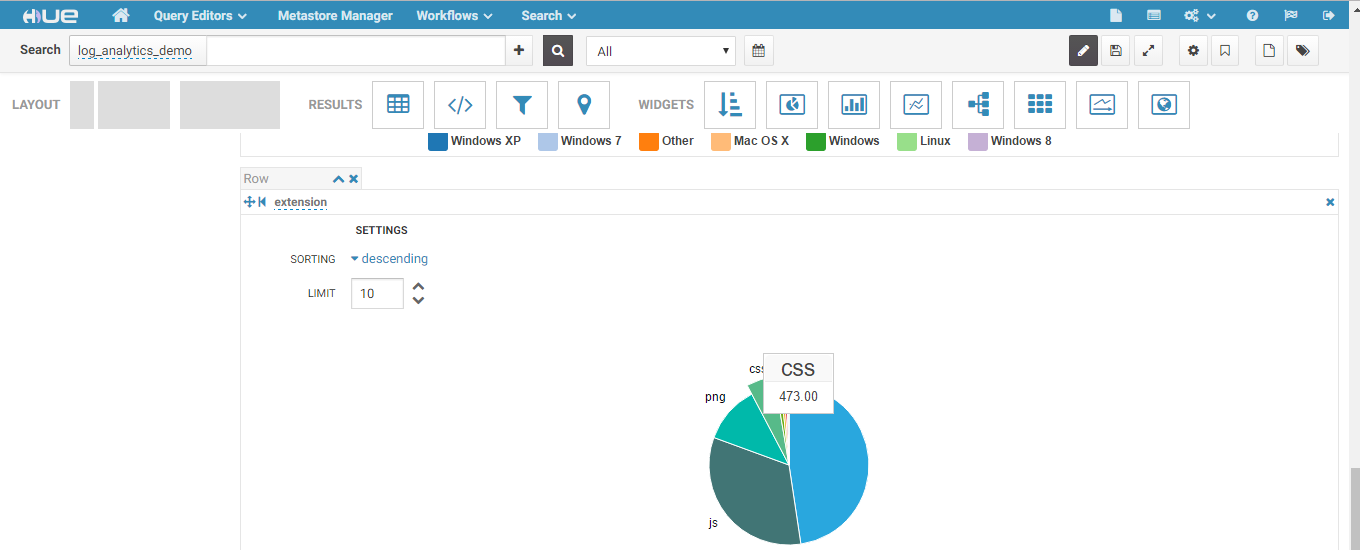
Drag the Pie Chart into the Row as shown below:



Select extension from the attributes.

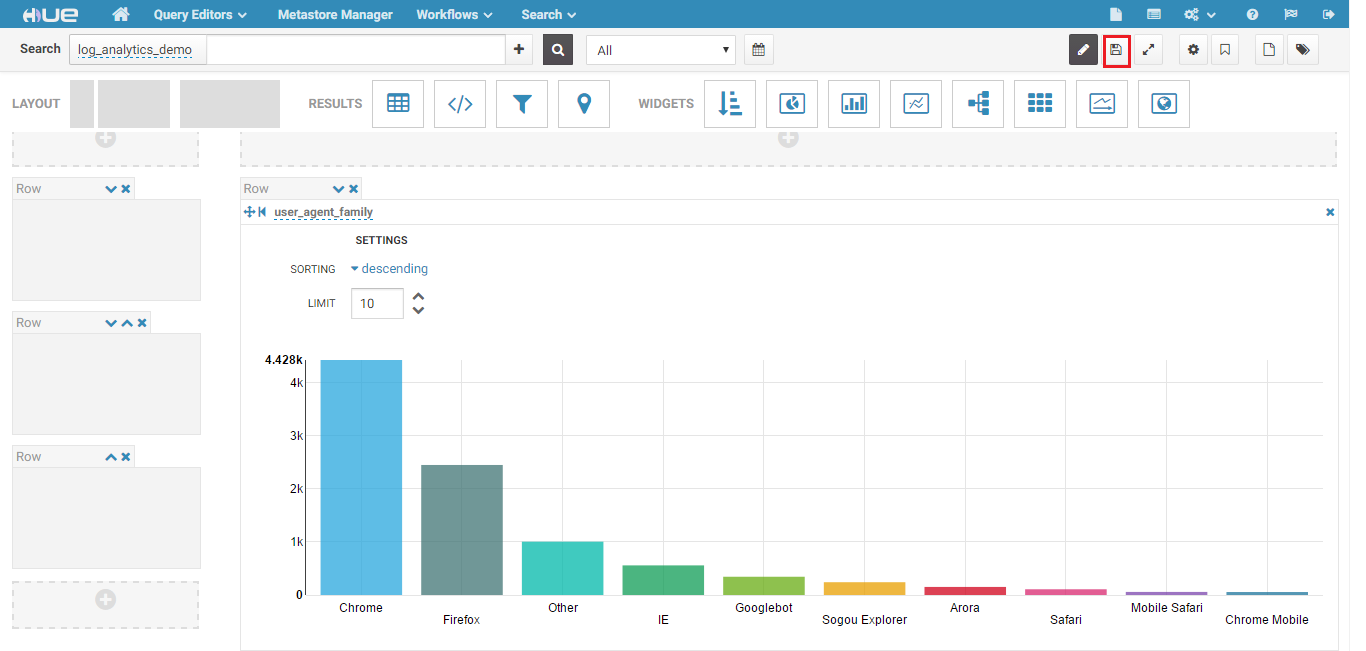


You can observe the number of users accessing the css pages, js pages, png images etc.



Once you are done with the analysis of your apache logs, you can save this search dashboard.

Scroll up and click on save Icon as shown below:



## Summary

This was a simple guide to navigate a novice user through the basic screens related to Cloudera Solr search, and Hue Dash boards.

If you want to evaluate this product against your own company data, please contact sales and we can arrange it.

Thank you for your time.