Hadoop as a Service

VMware vCloud Automation Center

Table of Contents

1. Introduction

VMware vCloud Automation Center is an innovative self-service provisioning and lifecycle management solution that simplifies and automates deployments of infrastructure, multi-tier applications, desktop... and now any kind of IT service! It provides a secure portal where authorized administrators, developers or business users can request new IT services as well as manage specific cloud and IT resources based on their roles and privileges.

Hadoop as a service runs on top of the Big Data Extensions allows you to automate the deployment and management of Apache Hadoop and HBase on virtual environments such as vSphere.

Hadoop is designed to run on a large cluster of commodity servers and to scale to hundreds or thousands of nodes. Each disk, server, network link, and even rack within the cluster is assumed to be unreliable. This assumption allows the use of the least expensive cluster components consistent with delivering sufficient performance, including the use of unprotected local storage (JBODs).
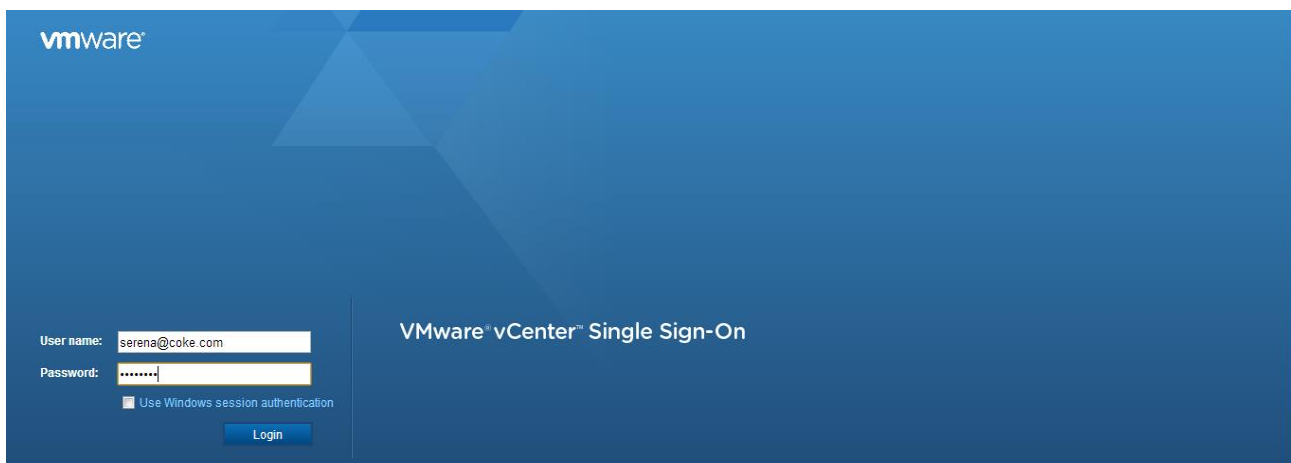
2. System Pre-requisites

1. Install and configure vCO 6.0
2. Install and Configure vCAC 6.0
3. Install and configure VMware vCO 6.0 with vCAC 6.0
4. Big Data Extensions configuration with vSphere 5.5

3. Set up

Follow the below steps to deploy the hadoop as a service in vCAC 6.0:

3.1 Request the Service as a Catalog Consumer

1. Log into vCAC Service Catalog (https://wdc-auroravcf-gen-dhcp85.eng.vmware.com/shell-ui-app/org/qe/) using SSO as a Catalog Consumer either tenant admin (tony@coke.com) or tenant user (serena@coke.com/password)

vmware®

VMware® vCenter™ Single Sign-On

User name: serena@coke.com
Password: ••••••••
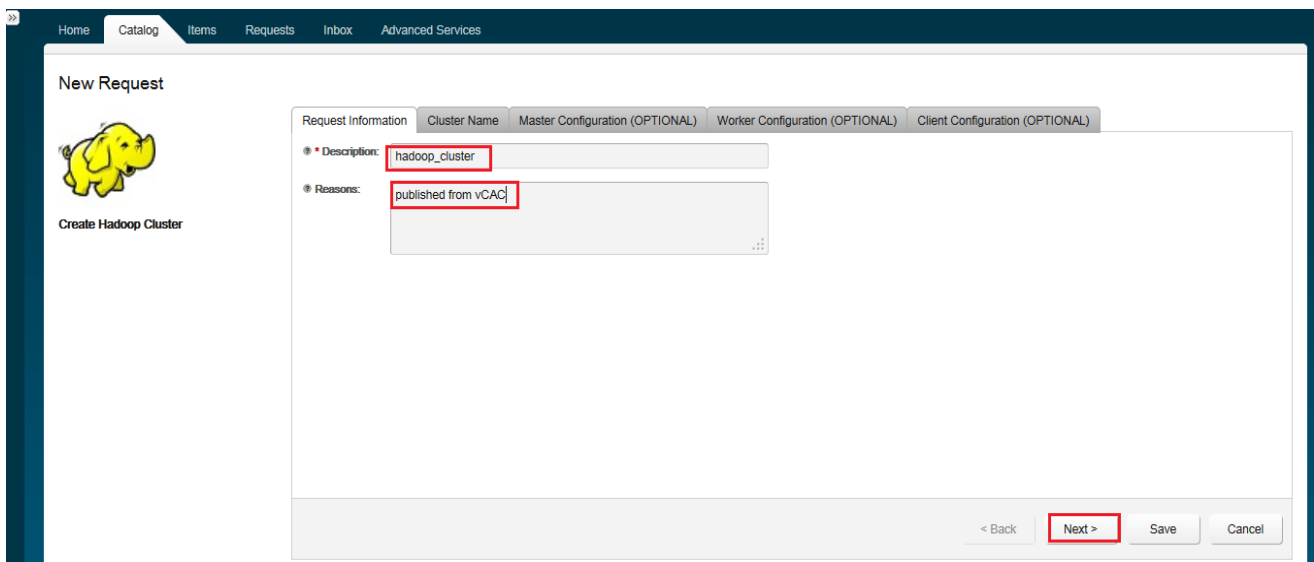☐ Use Windows session authentication
Login
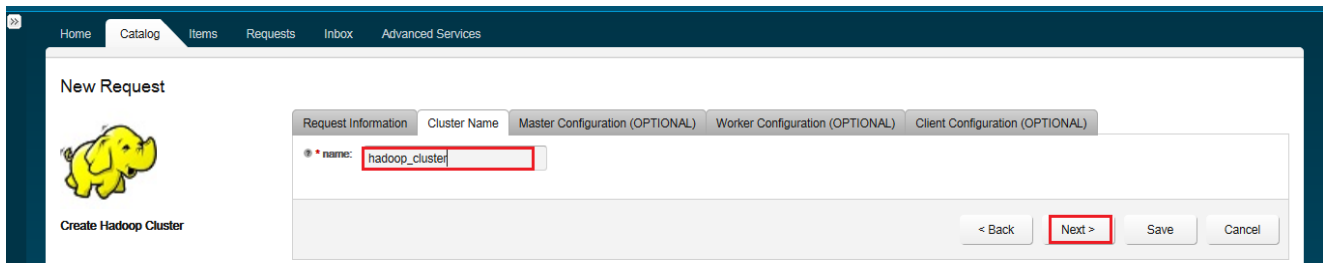
---

2. Click on the **Catalog** tab

3. There are list of catalog services. Click on the **Request** Button for "**Create Hadoop Cluster**"
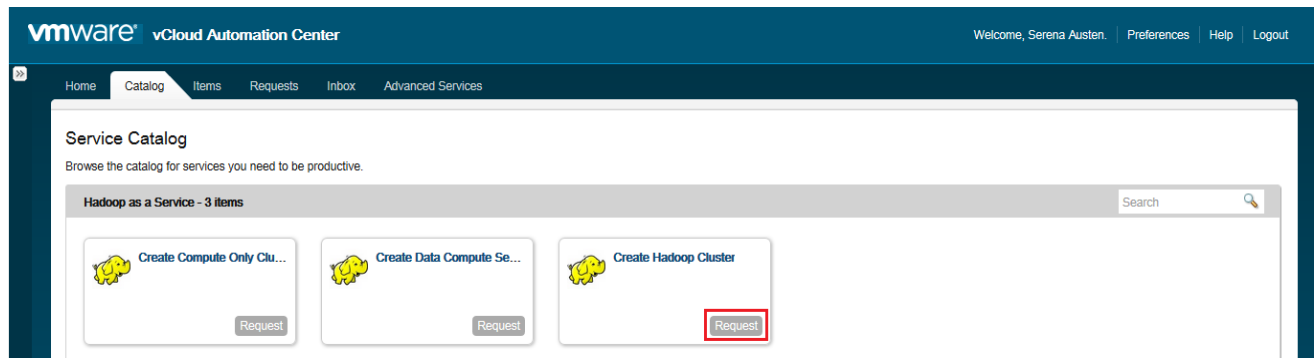
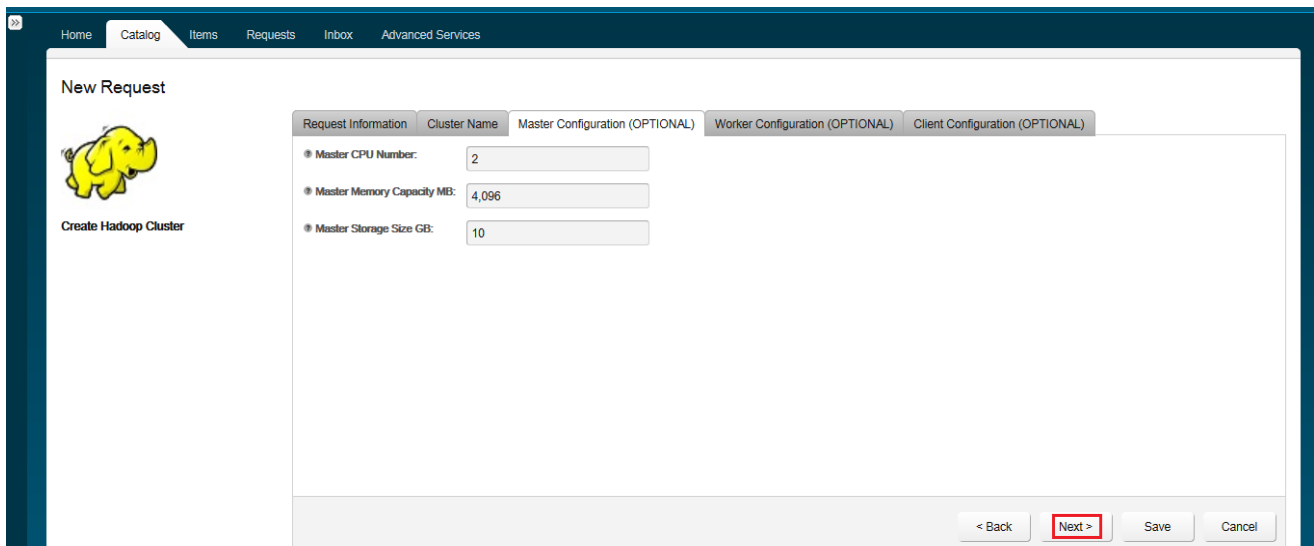4. Provide Description & Details and click **next**.

5. Enter the cluster name to be created and click **Next**

6. Review all the Master Configuration. If required you can change the default value as per your requirement. Click **Next**.

7. Review all the Worker Configuration. If required you can change the default value as per your requirement. Click **Next**.

8. Review all the Client Configuration. If required you can change the default value as per your requirement. Finally click **Submit**.
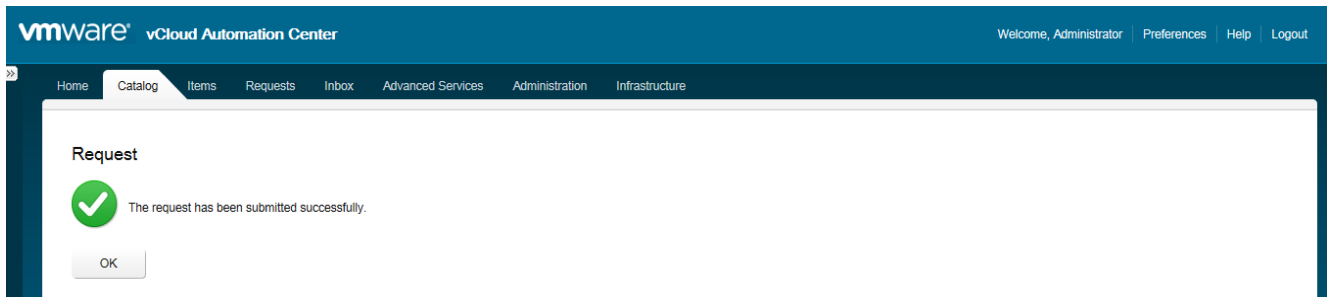
3.2 View the status of the application deployment

1. Log into vCAC Service Catalog (https://wdc-auroravcf-gen-dhcp85.eng.vmware.com/shell-ui-app/org/qe/) using SSO as a Catalog Consumer either tenant admin (tony@coke.com) or tenant user (serena@coke.com/password)
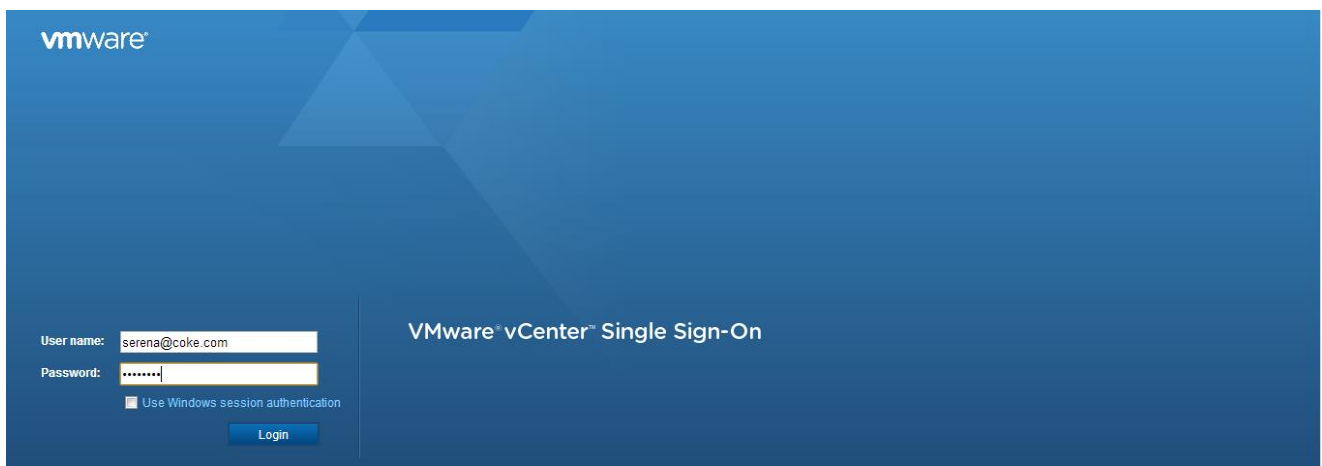
2. Click on the **Requests** tab.
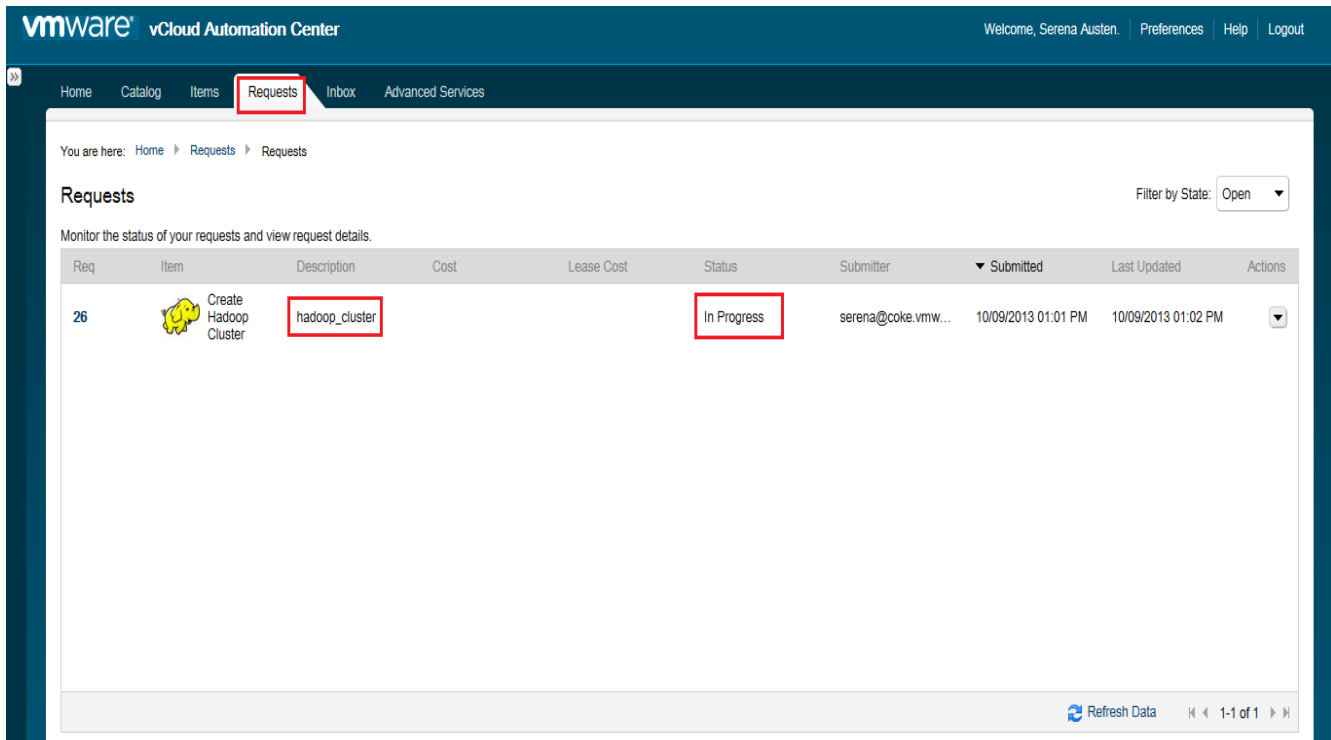
3. Wait for a while to complete the deployment.

4. Click on the **Items** tab. And click on the **hadoop_cluster**.
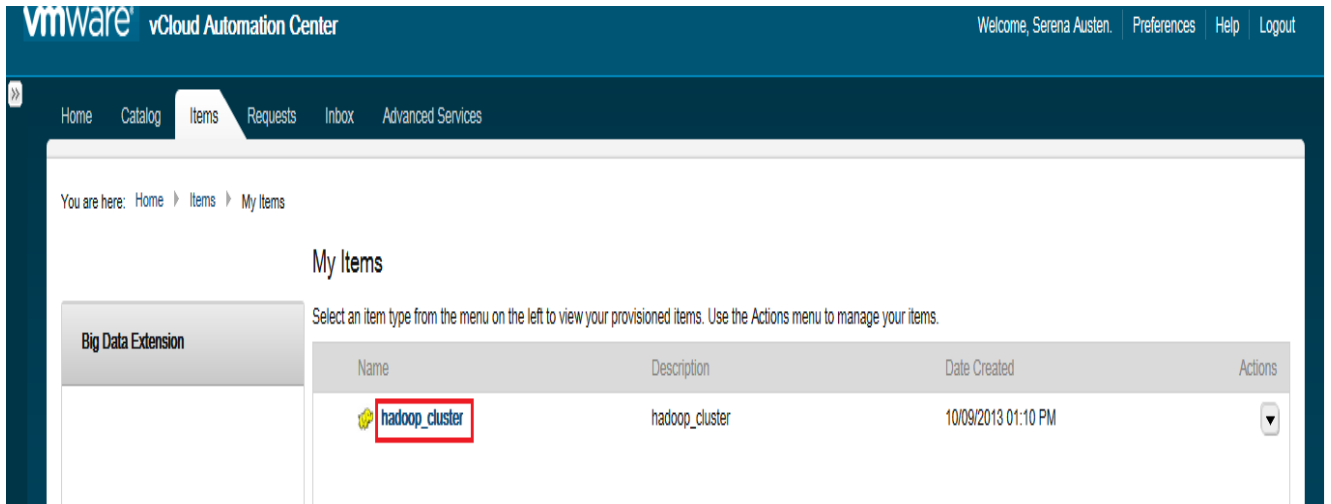
5. We can see the deployment status as RUNNING
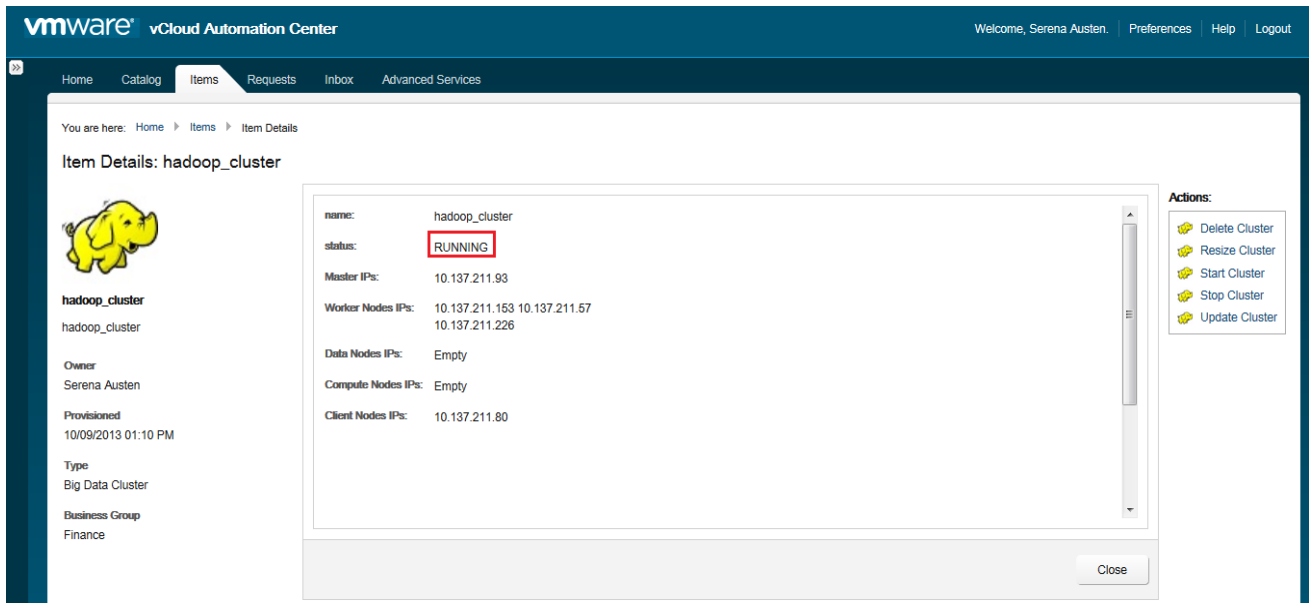
3.3 Launching the hadoop service

1. Log into vCAC Service Catalog (https://wdc-auroravcf-gen-dhcp85.eng.vmware.com/shell-ui-app/org/qe/) using SSO as a Catalog Consumer either tenant admin (tony@coke.com) or tenant user (serena@coke.com/password)
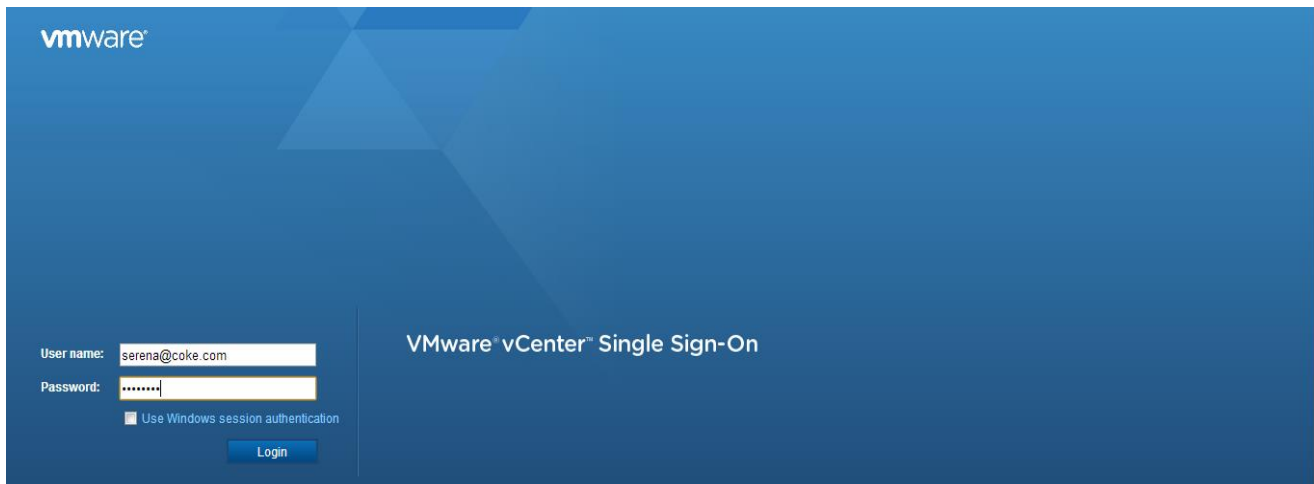
2. Click on the **Items** tab. And click on the **hadoop_cluster**.
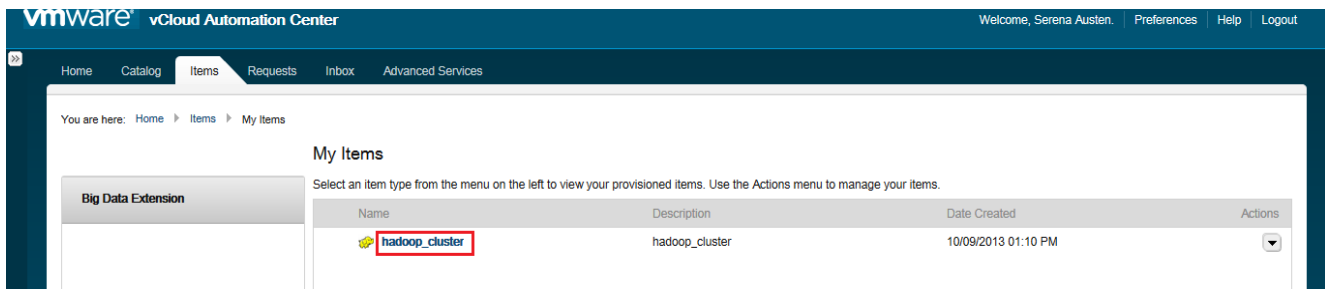
3. We can see the machine's IPs of name node and task tracker nodes
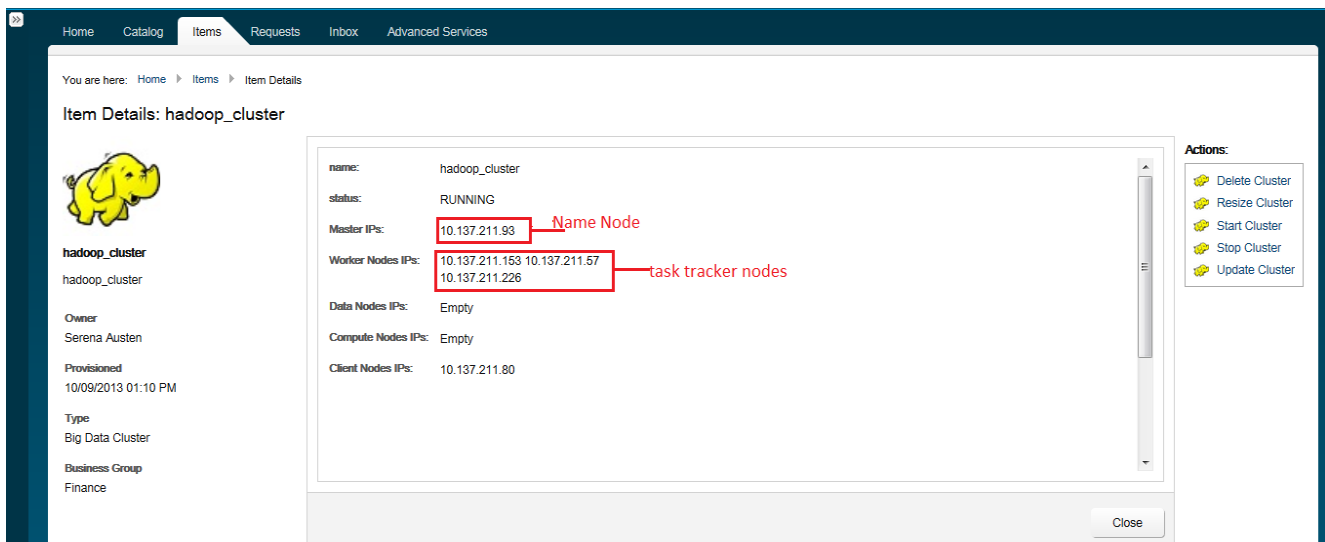
4. Open any browser of your choice and access the following links
   a. http://<nameNode_ip>:50070 for the Name node
      For example: http://10.137.211.93:50070
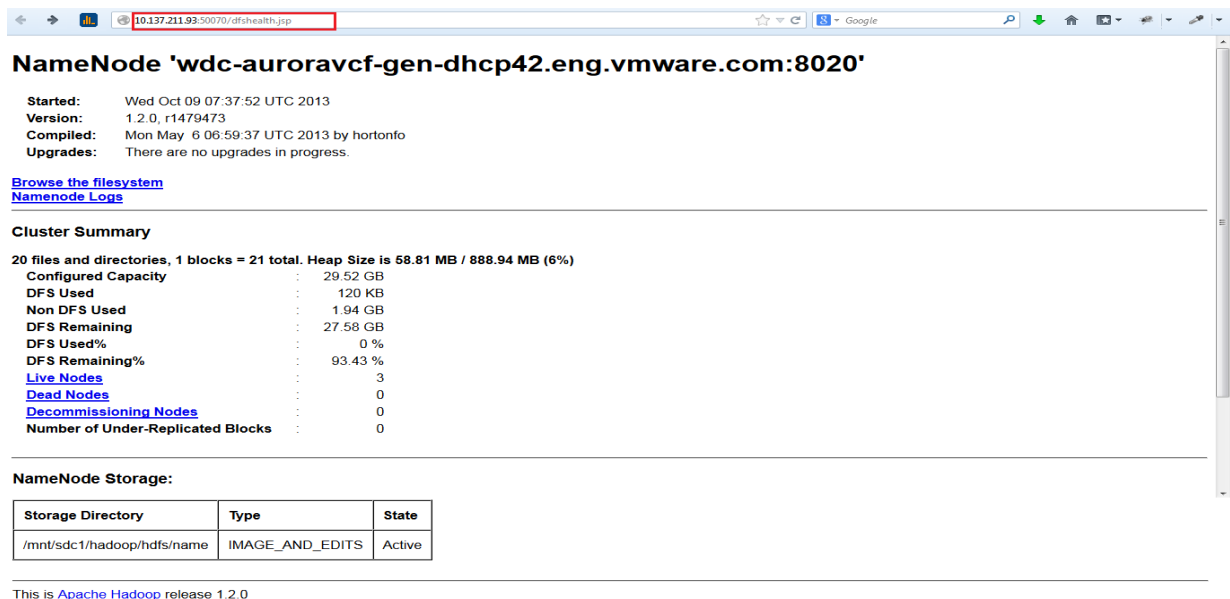
   b. http://<taskTracket_ip>:50060 for task tracker.

For example: http://10.137.211.153:50060

tracker_wdc-auroravcf-gen-dhcp102.eng.vmware.com:localhost/127.0.0.1:42700 Task Tracker Status

Version: 1.2.0, r1479473
Compiled: Mon May 6 06:59:37 UTC 2013 by hortonfo

Running tasks

Task Attempts	Status	Progress	Errors

Non-Running Tasks

Task Attempts	Status

Tasks from Running Jobs

Task Attempts	Status	Progress	Errors

Local Logs

Log directory

This is Apache Hadoop release 1.2.0

4. Summary

Hadoop as a service utilizes industry's leading Cloud-enabled development and process automation platform, VMware **vCenter Orchestrator** and Cloud-enabled self-service provisioning solution available with integration of the VMware **vCloud Automation Center**.

5. Benefits

5.1 Single User Experience for Infrastructure, Desktop and Application Services

In addition to support for heterogeneous infrastructure, vCAC now offers a single catalog for publishing and consuming application services. Users can now browse the same catalog to request and provision single or multi-node applications, just like they have for infrastructure and desktop services.

5.2 Extensible to any IT Service

In addition to out-of-the-box services, customers can now more easily extend vCAC to publish any kind of IT service to the common service catalog. Whether its **Storage-as-a-Service**, **Backup-as-a-Service**, or something as simple as letting users add capacity to their e-mail account, the new Advanced Service Designer lets service architects design rich user forms and provisioning workflows in a matter of minutes. And like pre-built services, custom services leverage the same entitlement and approval policy engines, enabling organizations to use a consistent governance layer.

5.3 Business Management for the Cloud
1. Provides visibility into the cost and usage of on-premise virtual infrastructure and public cloud infrastructure, including benchmarking capabilities
2. Supports "what-if" cost analysis to determine the best infrastructure type and placement options
3. Includes capacity, cost, and budget analysis capabilities for proactive planning

5.4 Improvements in Infrastructure as a Service
1. Support for OpenStack
2. Log into vSphere machines from the service catalog
3. Integration with vCloud Networking & Security (vCNS)

5.5 Improved Administration Capabilities
1. Support for LDAP services.
2. Improvements in multi-tenancy
3. Improved compatibility with VMware Site Recovery Manager

5.6 Hadoop as a service release provides the following benefits:

1. **Support for Major Hadoop Distributions.** Big Data Extensions includes support for Apache Hadoop, Cloudera, Greenplum, Hortonworks, MapR, and Pivotal. HBase, Pig, and Hive are also supported. The Big Data Extensions virtual appliance includes Apache Hadoop 1.2. Customers can easily upload distributions of their choice and configure Big Data Extensions to deploy their preferred distributions.

2. **Quickly Deploy, Manage, and Scale Hadoop Clusters.** Big Data Extensions enables the rapid deployment of Hadoop clusters on VMware vSphere. You can quickly deploy, manage, and scale Hadoop nodes using the virtual machine as a simple and elegant container. Big Data Extensions provides a simple deployment toolkit that can be accessed through VMware vCenter Server to deploy a highly available Hadoop cluster in minutes using the Big Data Extensions user interface.
3. **Graphical User Interface Simplifies Management Tasks.** The Big Data Extensions plug-in, a graphical user interface integrated with vSphere Web Client, lets you easily perform common Hadoop infrastructure and cluster management administrative tasks.
4. **Elastic Scaling Lets You Optimize Cluster Performance and Resource Utilization.** Elasticity-enabled clusters start and stop virtual machines automatically and dynamically to optimize resource consumption. Elasticity is ideal in a mixed workload environment to ensure that high priority jobs are assigned sufficient resources. Elasticity adjusts the number of active compute virtual machines based on configuration settings you specify.