**Support Isilon Hadoop Learning**

Table of Contents

[Introduction 4](#_Toc465935511)

[Isilon and Hadoop setup and integration 4](#_Toc465935512)

[Lab Overview 5](#_Toc465935513)

[Lab Overview 5](#_Toc465935514)

[Architecture 5](#_Toc465935515)

[Labs Scenario 6](#_Toc465935516)

[Access to Software 7](#_Toc465935517)

[Isilon License Keys 9](#_Toc465935518)

[Support for this lab 10](#_Toc465935519)

[Lab 1 - Ambari and HDP Deployments and Integrations 11](#_Toc465935520)

[Module 1 - Isilon Preparation for Hadoop Cluster Integration 11](#_Toc465935521)

[ Isilon networking configuration setup 11](#_Toc465935522)

[ Isilon Access Zone, User and Group and Base Directory requirement setup 11](#_Toc465935523)

[ Isilon HDFS configuration 11](#_Toc465935524)

[Module 2 - Ambari Hortonworks HDP Isilon Integration and install 11](#_Toc465935525)

[ Install Ambari; yum install ambari-server 11](#_Toc465935526)

[ Install Hortonworks Data Platform (HDP) 11](#_Toc465935527)

[ Integrate with Isilon 12](#_Toc465935528)

[ Test and validate the installation, services and jobs 12](#_Toc465935529)

[Module 3 - Test and validate Ambari & HDP operations and services 12](#_Toc465935530)

[ Run and validate jobs 12](#_Toc465935531)

[ Test and simulate operations and administrative tasks 12](#_Toc465935532)

[Module 4 - Kerberize Ambari HDP - Isilon Integration 13](#_Toc465935533)

[ Prepare MIT or Active Directory for Ambari Kerberos 13](#_Toc465935534)

[ Integrate Isilon into MIT or Active Directory 13](#_Toc465935535)

[ Kerberize and integrate HDP with MIT or Active Directory 13](#_Toc465935536)

[Module 5 - Test and validate Kerberized Ambari & HDP operations and services 13](#_Toc465935537)

[ Run and validate jobs 13](#_Toc465935538)

[ Test and simulate operations and administrative tasks 13](#_Toc465935539)

[Optional Tasks 13](#_Toc465935540)

[ Upgrade Ambari and HDP stacks 13](#_Toc465935541)

[ Upgrade Isilon OneFS 13](#_Toc465935542)

[ Break and troubleshoot components 13](#_Toc465935543)

[Lab 2 - Cloudera Manager and CDH Deployments and Integrations 14](#_Toc465935544)

[Module 1 - Isilon Preparation for Hadoop Cluster Integration 14](#_Toc465935545)

[ Isilon networking configuration setup 14](#_Toc465935546)

[ Isilon Access Zone, User and Group and Base Directory requirement setup 14](#_Toc465935547)

[ Isilon HDFS configuration 14](#_Toc465935548)

[Module 2 – Cloudera Manager with CDH and Isilon Integration and install 14](#_Toc465935549)

[ Install Cloudera Manager; 14](#_Toc465935550)

[ Install CDH Parcel 15](#_Toc465935551)

[o This requires the use of local parcels, use the webUI to add the local parcel repo to install service from; http://192.168.1.20/cdh5.8.2/ 15](#_Toc465935552)

[ Integrate with Isilon 17](#_Toc465935553)

[ Test and validate the installation, services and jobs 17](#_Toc465935554)

[Module 3 - Test and validate Cloudera Manager and CDH operations and services 17](#_Toc465935555)

[ Run and validate jobs 17](#_Toc465935556)

[ Test and simulate operations and administrative tasks 17](#_Toc465935557)

[Module 4 - Kerberize Cloudera - Isilon Integration 17](#_Toc465935558)

[ Prepare MIT or Active Directory for Ambari Kerberos 17](#_Toc465935559)

[ Integrate Isilon into MIT or Active Directory 17](#_Toc465935560)

[ Kerberize and integrate HDP with MIT or Active Directory 17](#_Toc465935561)

[Module 5 - Test and validate Kerberized Cloudera operations and services 17](#_Toc465935562)

[ Run and validate jobs 18](#_Toc465935563)

[ Test and simulate operations and administrative tasks 18](#_Toc465935564)

[Optional Tasks 18](#_Toc465935565)

[ Upgrade Ambari and HDP stacks 18](#_Toc465935566)

[ Upgrade Isilon OneFS 18](#_Toc465935567)

[ Break and troubleshoot components 18](#_Toc465935568)

[Reference Materials 19](#_Toc465935569)

[Conclusion 19](#_Toc465935570)

[Conclusion 19](#_Toc465935571)

Introduction

Isilon and Hadoop setup and integration

This lab provides the basic infrastructure to test and validate the procedures and configuration for deployment and operation of a Hadoop cluster against an Isilon OneFS cluster. The goal is to provide a baseline set of servers and services to allow the student to learn and develop the skills necessary to be successful in the deployment, operation or support of Isilon-Hadoop integrations. This lab does not provide step-by-step instructions for the student on how to accomplish these tasks, the primary purpose is to provide a lab space where the student has the required infrastructure and they can self-direct themselves and learn how to accomplish these tasks.

Lab Overview

Lab Overview

Architecture

The environment for this lab includes the following:

1 DC01 - AD / DNS Domain Controller Windows 2012 R2

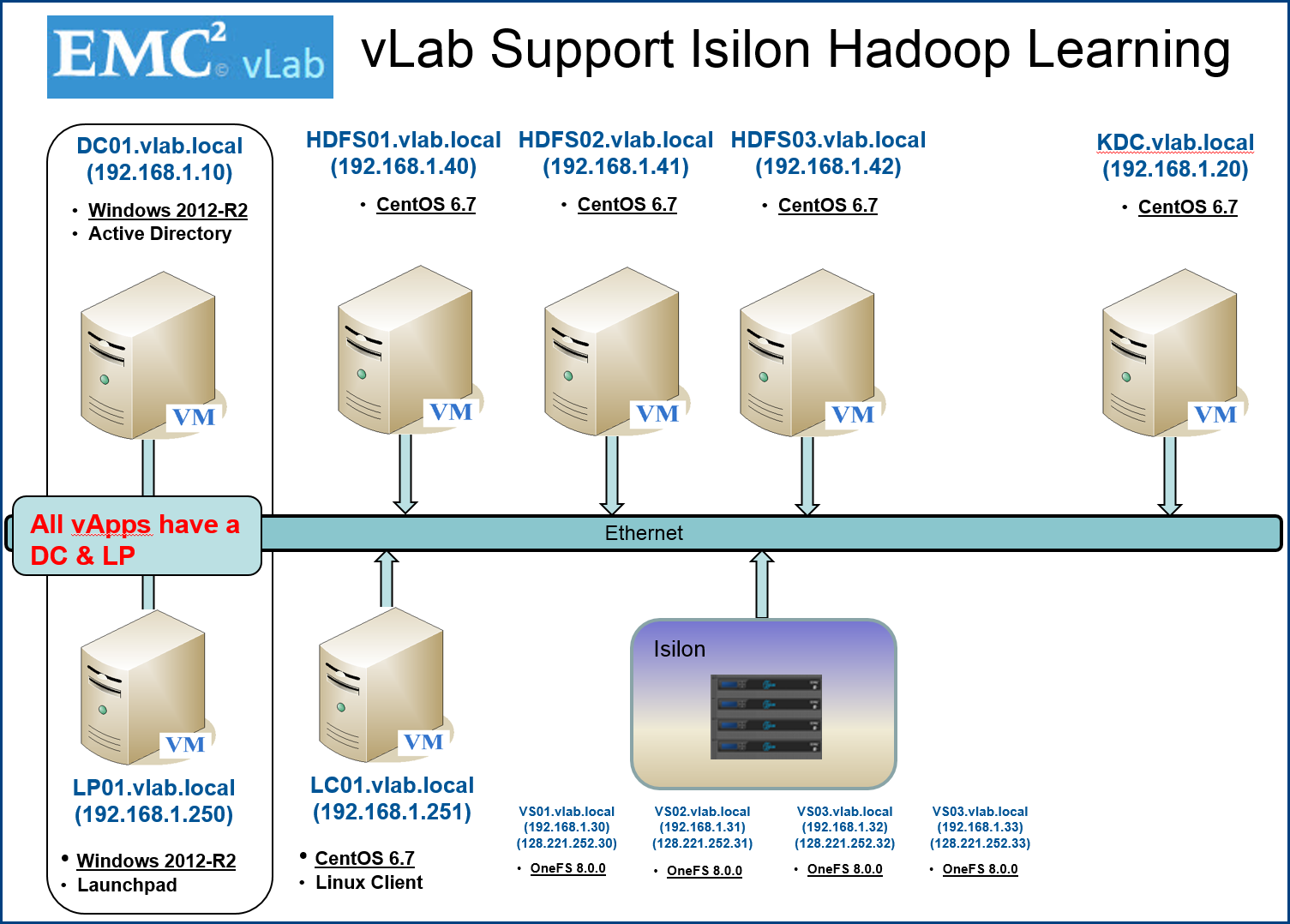
1 LP01 - Launchpad Windows 2012 R2

3 HDFS VMs - CentOS 6.7

1 LC01 - Linux Client CentOS 6.7

1 KDC - Linux CentOS 6.7 for use as MIT Kerberos server

**Note**: This lab guide covers the basic setup used to install and test Hadoop with Isilon storage.

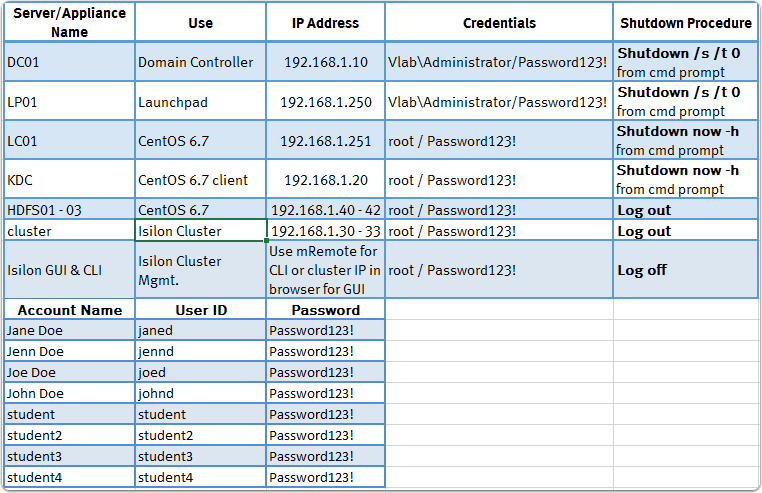


**Additional IP Pool range**

These IP ranges can be allocated to additional SmartConnect Pools on the Isilon cluster to create facilitate the use of Access Zones and multi-tenancy as needed

192.168.1.40 - 100

128.221.252.40 – 100



Labs Scenario

This lab is built to support the deployment and integration of two distinct Hadoop services; Hortonworks HDP or Cloudera CDH against an Isilon OneFS cluster initially in a non-Kerberized environment. But, on completion of deployment, the Hadoop cluster and Isilon can be Kerberized against an MIT KDC or Active Directory. The recommended approach is to execute Lab1 or Lab2 in newly deployed vlabs only.

**On Completion of Lab1 or Lab2 it is suggested to redeploy a new vlab and start from a clean environment when moving on to the second lab.**

The focus of this lab is to provide a fenced in environment that allows you to configure the Isilon cluster for HDFS, Install one of two versions of HDFS. Test and validate the HDFS operation and services, integrate with MIT Kerberos or Active Directory, and test and validate the HDFS operations and services.

**Lab 1 Ambari and HDP Deployments and Integrations**

**Lab 2 Cloudera Manager and CDH Deployments and Integrations**

**As stated previously this lab does not include any specific tasks to complete, rather it is an open environment where you can undertake any installation tasks or integration you require to gain the knowledge and understanding of how to deploy Isilon and Hadoop. By providing this space the student can leverage any materials or guides they see as beneficial in learning how to deploy these services and tools. If you run into problems or issues, you can always just destroy and redeploy the vlab.**

**For each lab, we will provide a high level overview of the relevant tasks that need to be accomplished to complete the deployments.**

Access to Software

Since the basis of this lab is to provide a sandbox in which many different platforms and services can be deployed. We will leverage the Windows Launch Pad host as a jump box to the internet to obtain any software, packages to updates as needed.

The launch pad host will already contain a number of software packages; version of Ambari, HDP stacks, Cloudera Manager and CDH stacks that can be used in the vLab.

If newer or additional software packages are needed you can just download them to the Launch Pad host then copy them to any host with in the vlab as needed.

Our goal by providing this flexibility is to provide the ability to install any version of these platforms as they are updated with newer releases.

Since the Linux hosts do not have direct connectivity to the internet, an internal yum repo has been provided on the KDC host (192.168.1.20) and can be used to obtain some software packages as required. If packages or versions of software are not present in the repo, follow the installation instruction for non-internet connected hosts when installing.

The following yum repos were created local to the lab to facilitate installs:

HDP-2.3.6.0 --- > Hortonworks HDP 2.3.6

HDP-UTILS-1.1.0.20 --- > Hortonworks Utilities 1.1.0.20

Ambari-2.2.0.0 --- > Ambari 2.2.0.0

cm-repo --- > Cloudera Manager 5.8.2

vlabrepo --- > CentOS 6

The Amabri and HDP repos are available at:

<http://192.168.1.20/hdp/HDP/centos6/2.x/updates/2.3.6.0/>

<http://192.168.1.20/hdp/HDP-UTILS-1.1.0.20/repos/centos6/>

A cloudera CDH parcel stack is also available at: <http://192.168.1.20/cdh5.8.2/>

(all are bookmarked in the chrome browser)

**These repos and parcels will allow the local installation of Ambari, HDP and Cloudera Manager without internet access to public repos.**

**It is suggested to use these repos for yum installs of these base software installs and then upgrade to later versions (this will require following the correct instruction to obtain offline upgrade packages as required).**

**Additional repos and parcels can be downloaded and installed to the KDC repo host as needed.**

A number of packages and installs have been downloaded to the LaunchPad host and are available from C:\software

Isilon License Keys

-------------------------------------------------------------------------------  
  
Software Key Duration: 90 days  
Software Key "Activate By" Date: 8/2/2017  
  
**Module: InsightIQ**  
Key: ISILO-GLUC0-O32JU-ZR0TW-VMULT  
  
**Module: Isilon for vCenter**  
Key: ISILO-SDHC0-0ROJU-NZJTW-BBIKT  
  
**Module: SmartConnect Advanced**  
Key: ISILO-HJXC0-H15JU-YTZTW-OKZLT  
  
**Module: SmartLock**  
Key: ISILO-BI1C0-J0RJU-4SVTW-QHLLT  
  
**Module: SmartPools**  
Key: ISILO-GAZB0-O0RKU-Z0XQW-VHLIT  
  
**Module: SmartQuotas**  
Key: ISILO-VA0B0-5SSKU-K0UQW-GAKIT  
  
**Module: SnapshotIQ**  
Key: ISILO-4KUC0-SY2JU-BQ0TW-ZFULT  
  
**Module: SyncIQ**  
Key: ISILO-XYLD0-5LDIU-ICFSW-GZ2IT  
  
**Module: HDFS for OneFS**  
Key: ISILO-HVOC0-PHOJU-YPATW-W4FKT  
  
**Module: ObjectIQ**  
Key: ISILO-YFID0-QXEIU-H5GSW-XO2IT  
  
**Module: Platform API**  
Key: ISILO-CROB0-KTGKU-3LAQW-RC4GT

**Module: SmartDedupe**  
Key: ISILO-ZTFD0-RBNJU-GJLSW-YQGKT

**Module: CloudPools**  
Key: ISILO-OAZD0-GSBIU-R0XSW-N50JT  
  
**Module: Swift**  
Key: ISILO-3ISD0-V00MU-CS2SW-2HSOT

Support for this lab

Since the goal of this lab is to provide an unconfigured lab with no direct instructions on how to setup and configure these services, no formal support exists for this lab. If you have questions or are looking for other users who leverage this lab for learning, the best place to find answer is via the community support page at:

<https://inside.dell.com/docs/DOC-231359>

Lab 1 - Ambari and HDP Deployments and Integrations

This lab is intended to help you understand the deployment methodology for deploying Hortonworks Ambari HDP against an Isilon cluster.

The modules listed below highlights the overall deployment strategy that need to be accomplished to successfully deploy and operate HDP. The bulleted tasks provide guidance around the high level configuration tasks that need to be accomplished.

## Module 1 - Isilon Preparation for Hadoop Cluster Integration

Prepare and configure OneFS for hadoop integration; using the Isilon implementation guides for hadoop setup the cluster for integration.

The vlab Isilon cluster is currently running: OneFS v8.0.0.1

OneFS 8.0.0.2 & Patch 180735 and OneFS 8.0.1.0 are downloaded and available in C:\Software on the Launchpad host if needed.

### Isilon networking configuration setup

### Isilon Access Zone, User and Group and Base Directory requirement setup

It is suggested to use the Isilon Hadoop Tools (always get latest)

<https://github.com/Isilon/isilon_hadoop_tools>

### Isilon HDFS configuration

## Module 2 - Ambari Hortonworks HDP Isilon Integration and install

Setup and install Ambari and then deploy the Hortonworks Data Platform with Isilon integration.

### Install Ambari; yum install ambari-server

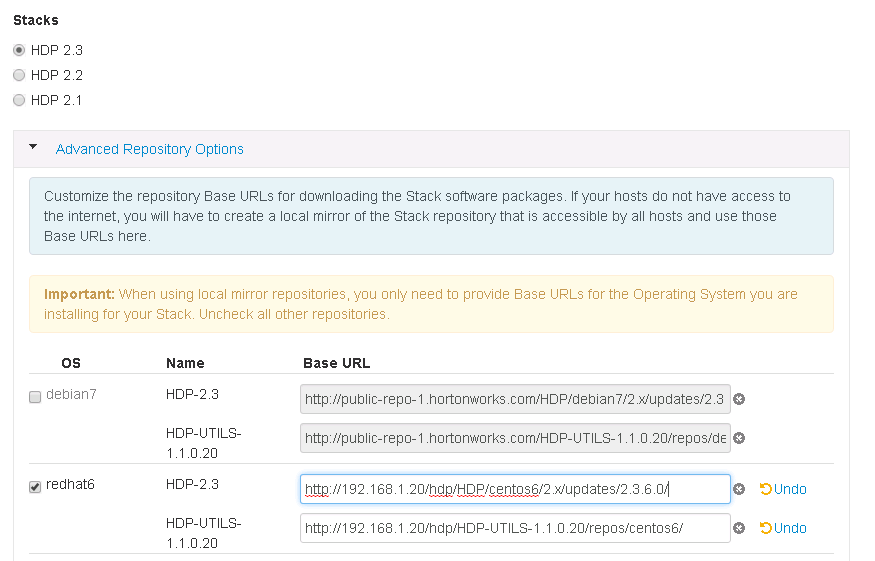
### Install Hortonworks Data Platform (HDP)

When installing the HDP stacks, the local repos should be used. Change the public hdp repos to point to the internally hosted 2.3 stack.

Set the repos to:

<http://192.168.1.20/hdp/HDP/centos6/2.x/updates/2.3.6.0/>

<http://192.168.1.20/hdp/HDP-UTILS-1.1.0.20/repos/centos6/>



### Integrate with Isilon

### Test and validate the installation, services and jobs

## Module 3 - Test and validate Ambari & HDP operations and services

Having deployed and integrated HDP with Isilon we can test and validate hadoop jobs and services.

### Run and validate jobs

* + Read and write files via hdfs
  + PI/Teragen/Teravalidate/Terasort

### Test and simulate operations and administrative tasks

## Module 4 - Kerberize Ambari HDP - Isilon Integration

Having deployed an operational HDP Hadoop cluster against Isilon we can then Kerberize the cluster to provide secure Hadoop services with Isilon. We have the option of deploying kerberos based on a MIT KDC or an Active Directory. Follow the documentation depending which kerberos provide you intend to deploy and integrate.

### Prepare MIT or Active Directory for Ambari Kerberos

### Integrate Isilon into MIT or Active Directory

### Kerberize and integrate HDP with MIT or Active Directory

Module 5 - Test and validate Kerberized Ambari & HDP operations and services

On completion of Kerberizing the Hadoop cluster we can now test and review the operations and behavior of Kerberized jobs and access against an Isilon cluster.

### Run and validate jobs

### Test and simulate operations and administrative tasks

## Optional Tasks

### Upgrade Ambari and HDP stacks

### Upgrade Isilon OneFS

### Break and troubleshoot components

Lab 2 - Cloudera Manager and CDH Deployments and Integrations

This lab is intended to help you understand the deployment methodology for deploying Cloudera’s CDH stack against an Isilon cluster.

The modules listed below highlights the overall deployment strategy that need to be accomplished to successfully deploy and operate Cloudera Manager and CDH. The bulleted tasks provide guidance around the high level configuration tasks that need to be accomplished.

## Module 1 - Isilon Preparation for Hadoop Cluster Integration

Prepare and configure OneFS for hadoop integration; using the Isilon implementation guides for hadoop setup the cluster for integration.

### Isilon networking configuration setup

### Isilon Access Zone, User and Group and Base Directory requirement setup

It is suggested to use the Isilon Hadoop Tools (always get latest)

<https://github.com/Isilon/isilon_hadoop_tools>

### Isilon HDFS configuration

## Module 2 – Cloudera Manager with CDH and Isilon Integration and install

Setup and install Cloudera Manager and then deploy the CDH stack with Isilon integration.

### Install Cloudera Manager;

* Download the installer cloudera-manager-installer.bin (also on C:\software on LP host) <http://archive.cloudera.com/cm5/installer/5.8.2.3/>
* Copy to the to be installed CM host from LaunchPad host
* chmod u+x cloudera-manager-installer.bin
* Install using ./cloudera-manager-installer.bin --skip\_repo\_package=1

Access the Cloudera Manager admin page as instructed by the installer.

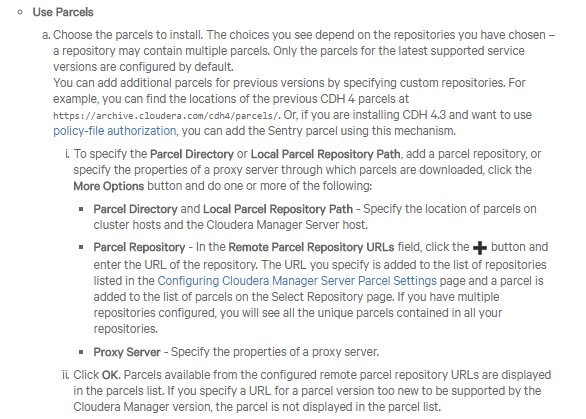


An excellent reference to installing CM with local repos and parcels:

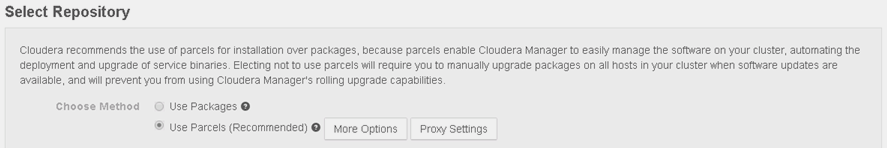
<http://www.cloudera.com/documentation/enterprise/latest/topics/installation_installation.html#concept_qpf_2d2_2p>

### Install CDH From Parcels

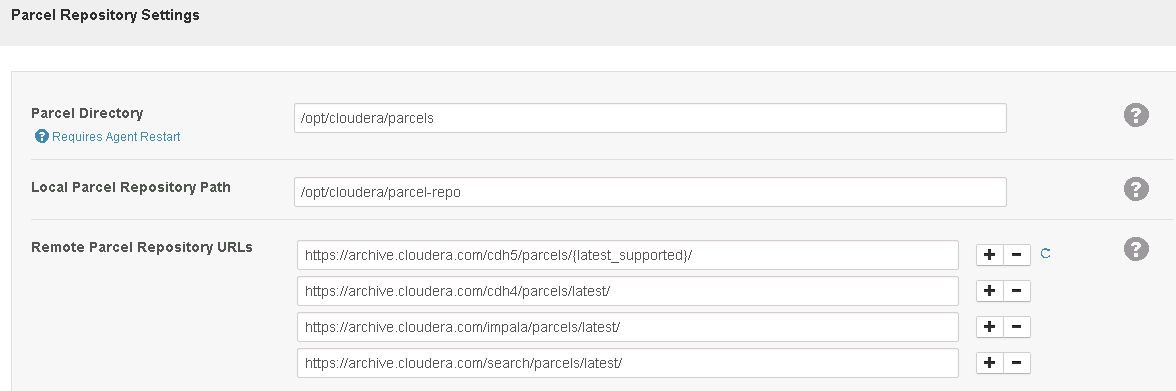
### This requires the use of local parcels, use the webUI to add the local parcel repo to install service from; <http://192.168.1.20/cdh5.8.2/>



When selecting a repo, select the more option.



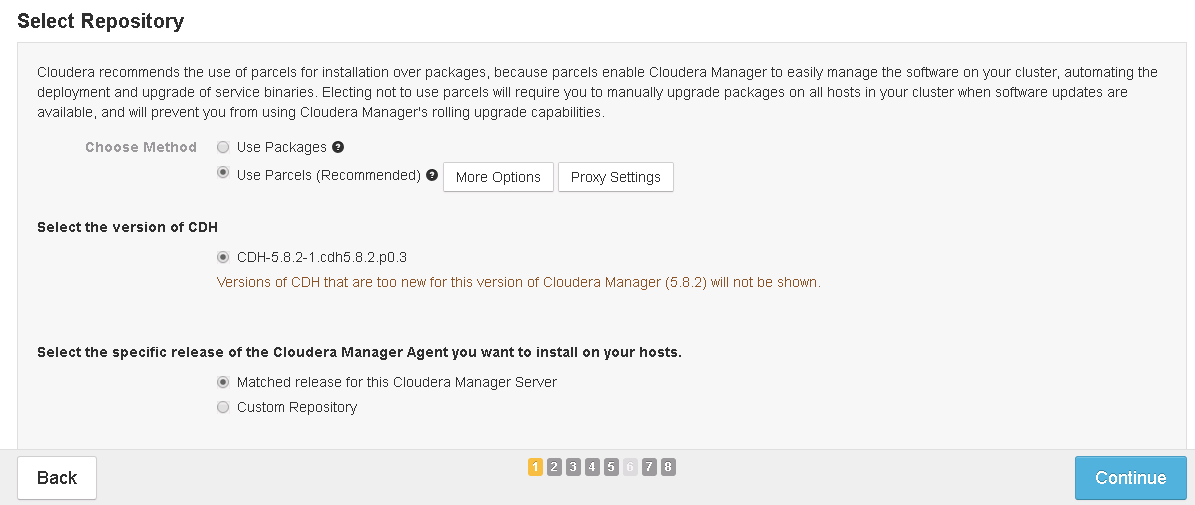
Remove all the existing repos.



Add the local repo: <http://192.168.1.20/cdh5.8.2/>



The version should now reflect the local repo parcel.



### Integrate with Isilon

### Test and validate the installation, services and jobs

## Module 3 - Test and validate Cloudera Manager and CDH operations and services

Having deployed and integrated CDH with Isilon we can test and validate hadoop jobs and services.

### Run and validate jobs

* + Read and write files via hdfs
  + PI/Teragen/Teravalidate/Terasort

### Test and simulate operations and administrative tasks

## Module 4 - Kerberize Cloudera - Isilon Integration

Having deployed an operational Cloudera Hadoop cluster against Isilon we can then Kerberize the cluster to provide secure Hadoop services with Isilon. We have the option of deploying kerberos based on a MIT KDC or an Active Directory. Follow the documentation depending which kerberos provide you intend to deploy and integrate.

### Prepare MIT or Active Directory for Ambari Kerberos

### Integrate Isilon into MIT or Active Directory

### Kerberize and integrate HDP with MIT or Active Directory

Module 5 - Test and validate Kerberized Cloudera operations and services

On completion of Kerberizing the Hadoop cluster we can now test and review the operations and behavior of Kerberized jobs and access against an Isilon cluster.

### Run and validate jobs

### Test and simulate operations and administrative tasks

## Optional Tasks

### Upgrade Ambari and HDP stacks

### Upgrade Isilon OneFS

### Break and troubleshoot components

Reference Materials

The Isilon Hadoop InfoHub

<https://community.emc.com/docs/DOC-39529>

Hortonworks

<http://hortonworks.com/>

Cloudera

<http://www.cloudera.com/>

Community Support for this lab

<https://inside.dell.com/docs/DOC-231359>

Isilon Hadoop Tools

https://github.com/Isilon/isilon\_hadoop\_tools

Conclusion

Conclusion

The unstructured nature and flexibility of this vlab is designed to provide maximum usability and learning opportunity to the student. By providing all the required infrastructure and a base configuration on all the hosts, the student has the ability to learn at their own pace and implement any deployment methodology they wish. The goal by not providing step by step instructions is to encourage the individual to gain deeper knowledge by learning how to implement these tools using any and all available resources. This represents a unique lab experience to provide maximum flexibility in learning the deployment and integration of Hadoop with Isilon.