



PULPO
BIG DATA

phadoop

Making Hadoop easy with Cloudera Manager

- Automated deployment and configuration
- Customizable Monitoring and reporting
- Effortless, robust troubleshooting
- Zero downtime maintenance

cloudera[®]
Ask Bigger Questions

Cluster Building Block

Scalable up to PentaByte (PB)

Data Fabrican

10GbE Switch (1-2) .48Port
10G SFP+/10GBase-T

Management Network

10GbE Switch (1-2) .48Port
10G SFP+/10GBase-T

Hadoop Name Nodes

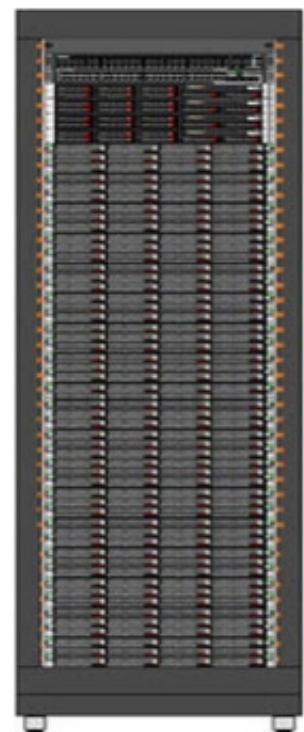
SuperServer (3) 1U DP Xeon
E5-2600 v3 Hardware RAID, Redundant Power

SuperRack

42U rack with Metered PDUs

Hadoop Data Nodes

TwinPro or FatWin(34) 2U-
4U Xeon E5-2600 v3, dual 10GbEer (3) 1U DP Xeon



CLOUDERA is a trademark of Cloudera, Inc

Cloudera Named a Visionary in 2016 Gartner Magic Quadrant for Data Warehouse Data Management Solutions for Analytics

Cloudera Manager

Makes it easy to manage Hadoop deployments of any scale in production. Quickly deploy, configure, and monitor your cluster through an intuitive UI - complete with rolling upgrades, backup and disaster recovery, and customizable alerting.

Feature

Deployment & Configuration

Cloudera Manager can deploy client configurations within the cluster, each applicable service has a Deploy Client Configuration action. This action does not necessarily deploy the client configuration to the entire cluster, it only deploys the client configuration to all the hosts that this service has been assigned to.



Cluster Health Alert

Cloudera Manager monitors the health of the services, roles, and hosts that are running in your clusters via health tests. The Cloudera Management Service also provides health tests for its roles.

Role-based health tests are enabled by default



// Custom Reporting

Reports Manager generates reports that provide an historical view into disk utilization by user, user group, and directory, processing activities by user and YARN pool, and HBase tables and namespaces. This role is not added in Cloudera Express.

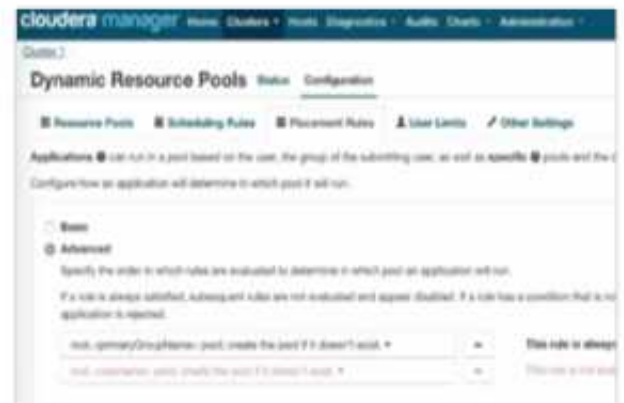


File	Download
Namespaces by User	
Current Disk Usage By User	CSV (1.5)
Current Disk Usage By Group	CSV (1.5)
Current Disk Usage By Directory	CSV (1.5)
Historical Disk Usage By User	CSV (1.5)
Historical Disk Usage By Group	CSV (1.5)
Historical Disk Usage By Directory	CSV (1.5)
Usage Data	
Usage Statistics	
Files with Large Modifications	
Overcommitted Directories	
Directories with Quotas	
Quota Alerts	
User Access (HDFS)	
File	Download
Namespaces by User	
Directory Access By Group	CSV (1.5)
Applications (HWT jobs included) (YARN HWT included)	Configure Aggregations

// Dynamic Resource Management

Resource management helps ensure predictable behavior by defining the impact of different services on cluster resources. The goals of resource management features are to:

- Guarantee completion in a reasonable time frame for critical workloads
- Support reasonable cluster scheduling between groups of users based on fair allocation of resources per group
- Prevent users from depriving other users access to the cluster



// Troubleshoot Query Performance

The Cloudera Management Service Host Monitor role performs health tests and collects host metrics to allow you to monitor the health and performance of the hosts.

// Backup and Disaster Recovery

Runtime state is what processes are running where, and what commands (for example, rebalance HDFS or execute a Backup/Disaster Recovery schedule or rolling restart or stop) are currently being executed. The runtime state includes the exact configuration files needed to run a process. When you select Start in the Cloudera Manager Admin Console, the server gathers up all the configuration for the relevant services and roles, validates it, generates the configuration files, and stores them in the database.

Create HDFS Replication

Source: HDFS-1 (Cluster 1 @ Dev Source)
Destination: HDFS-1 (Cluster 1)
Path: User/SalesHistory/2018

Use [Add Paths](#) to add more replication sources

Schedule: Immediate

Regular Expression-Based Path Exclusion ⓘ
[Add Exclusion](#)

Advanced Options

MapReduce Service: MAPREDUCE-1 (Cluster 1)
Scheduler Profile: Default
Run As: Default

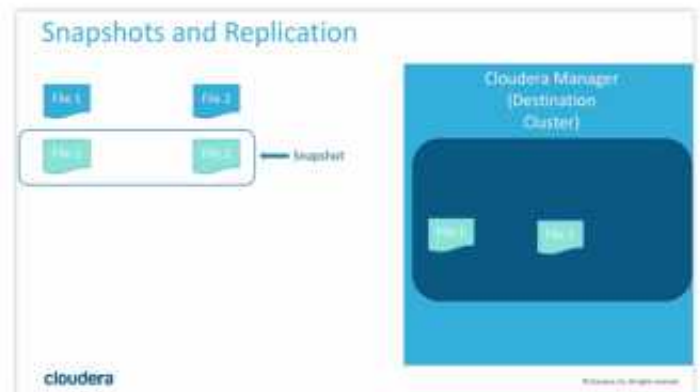
Log Path: Default

Maximum Map Slots: Leave empty for default (20)

Maximum Bandwidth: MB/s (per mapper) Leave empty for default (100)

// Snapshots for Efficient Backup

Runtime state is what processes are running where, and what commands (for example, rebalance HDFS or execute a Backup/Disaster Recovery schedule or rolling restart or stop) are currently being executed.



www.pulposystem.com

cloudera
Ask Bigger Questions