Automated Test Bed

Release 1.0

Getting Started Guide

Contents

[Introduction 3](#_Toc462410787)

[Pre-requisite 3](#_Toc462410788)

[Typical test setup topology 3](#_Toc462410789)

[Testbed setup 4](#_Toc462410790)

[Ansible based Hadoop installation-User Guide 4](#_Toc462410791)

[Pre-requisite 4](#_Toc462410792)

[User inputs before running the script on Ansible node: 4](#_Toc462410793)

[Running the Ansible script: 5](#_Toc462410794)

[Splunk Application setup 6](#_Toc462410795)

[Performance Monitoring 8](#_Toc462410796)

[Analytics application 8](#_Toc462410797)

[Main Dashboard 8](#_Toc462410798)

[Additional dashboards 8](#_Toc462410799)

[How to use Analytics App 8](#_Toc462410800)

[Performance details 9](#_Toc462410801)

[How to analyze the data. 10](#_Toc462410802)

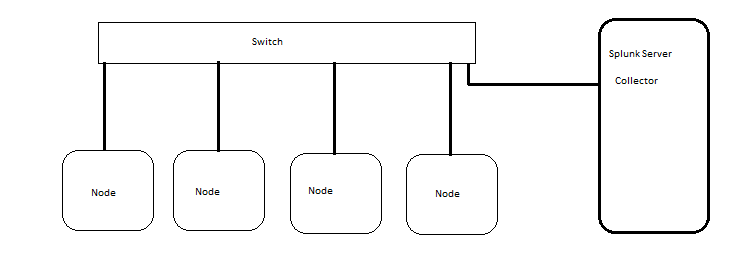
# Introduction

Automated test bed project is a set of scripts, and Splunk based custom application for collection and analysis of performance metrics of various components of system. This project includes scripts for automation of setting up of Hadoop and running Hadoop based workload in a test setup. The collected data can be analyzed for the performance impact of running large Hadoop workload, on servers, and switches.

## Pre-requisite

* Network infrastructure with switches is setup
* Bare metal servers are setup and Ubuntu Linux is installed on these servers.

## Typical test setup topology



## Setup Steps

Below is a summary of steps in involved in automated test setup for Hadoop based workload.

* Setup infrastructure specifically, switches, server (with Linux OS), VM with Splunk, A VM for running Ansible and collector
* Setup Hadoop by using Ansible scripts
* Build Splunk app by running build.py
* Install Splunk app on Splunk server and restart Splunk
* Configure collectd-client.conf with collectd server/ Splunk server address and update these on the nodes from where metrics are collected.
* Add a Splunk Job from application dashboard and monitor the performance counters. Repeat test multiple times to take reasonable error free values.

# Testbed setup

## Ansible based Hadoop installation-User Guide

Ansible scripts for Hadoop installation are located under Hadoop installation directory. Copy these scripts to Ansible node (Node which has Ansible installed).

### Pre-requisite

1. SSH key from Ansible node should be copied from Ansible node to the cluster nodes.

### User inputs before running the script on Ansible node:

1. /etc/ansible/hosts files has to be updated with list of nodes for each service.
2. Update the hosts file with list of IP addresses and hostnames
3. Update settings.cfg file with network and hostname.
4. Subnet has to be given as an input to prepare\_nodes.yaml
5. Cluster\_name and subnet has to be given as input for Hadoop\_install.yaml for mapr installation.

/etc/ansible/hosts file will have the following format.

[allnodes] #Should have all node IP addresses

192.168.0.46

192.168.0.66

[ansible-node]

192.168.0.12

[mapr-cldb]

192.168.0.46

192.168.0.66

[mapr-zookeeper]

192.168.0.46

[mapr-nfs]

192.168.0.46

[mapr-webserver]

192.168.0.46

[mapr-fileserver]

192.168.0.46

192.168.0.66

[mapr-jobtracker]

192.168.0.46

[mapr-resourcemanager]

192.168.0.46

[mapr-historyserver]

192.168.0.46

[mapr-nodemanager]

192.168.0.66

[mapr-tasktracker]

192.168.0.66

[squid3]

192.168.0.46

[ntp-server]

192.168.0.46

User should enter the list of IP’s where each service should run.

### Running the Ansible script:

1. Following script is used to install the pre-requisites

ansible-playbook prepare\_node.yaml --extra-vars="subnet=192.168.0.0/16"

1. Following script is used to install mapr packages

ansible-playbook hadoop\_install.yaml --extra-vars="subnet=192.168.0.0/16 cluster\_name=new1\_cluster"

1. Following script is used to add the license to the cluster

ansible-playbook add\_license.yaml --extra-vars=”path1=/home/cisco/req\_test cluster\_name=new1\_cluster”

1. Following script is used to validate the cluster after license is added

ansible-playbook validate.py

1. Following command is used cleanup/uninstall mapr packages on all the nodes

ansible-playbook hadoop\_cleanup.yaml

**Test Environment:**

These scripts are tested on a 2 node setup with Ubuntu as an operating system.

# Splunk Application setup

* Build custom Splunk application by running python script build.py. This script creates MonitoringSolutions.tar.gz
* Copy MonitoringSolutions.tar.gz to Splunk server.
* Untar the application under, /opt/splunk/etc/apps directory.
* Restart splunk. Below is sample commands and output

cd /opt/splunk/etc/apps

sudo tar -zxf MonitoringSolutions.tar.gz

sudo /opt/splunk/bin/splunk restart

splunkd 32516 was not running.

Stopping splunk helpers...

Done.

Stopped helpers.

Removing stale pid file... done.

splunkd is not running.

Splunk> Take the sh out of IT.

Checking prerequisites...

Checking http port [8000]: open

Checking mgmt port [8089]: open

Checking appserver port [127.0.0.1:8065]: open

Checking kvstore port [8191]: Checking configuration... Done.

Checking critical directories... Done

Checking indexes...

Validated: \_audit \_internal \_introspection \_thefishbucket history main summary

Done

open

Checking filesystem compatibility... Done

Checking conf files for problems...

Done

Checking default conf files for edits...

Validating installed files against hashes from '/opt/splunk/splunk-6.4.0-f2c836328108-linux-2.6-x86\_64-manifest'

All installed files intact.

Done

All preliminary checks passed.

Starting splunk server daemon (splunkd)...

Done

Waiting for web server at http://127.0.0.1:8000 to be available.

* Open Splunk URL from browser. Analytics application is listed on the left hand side.
* Click the application link and the dashboard comes up.
* Navigate the dashboard to start the test, view test metrics.

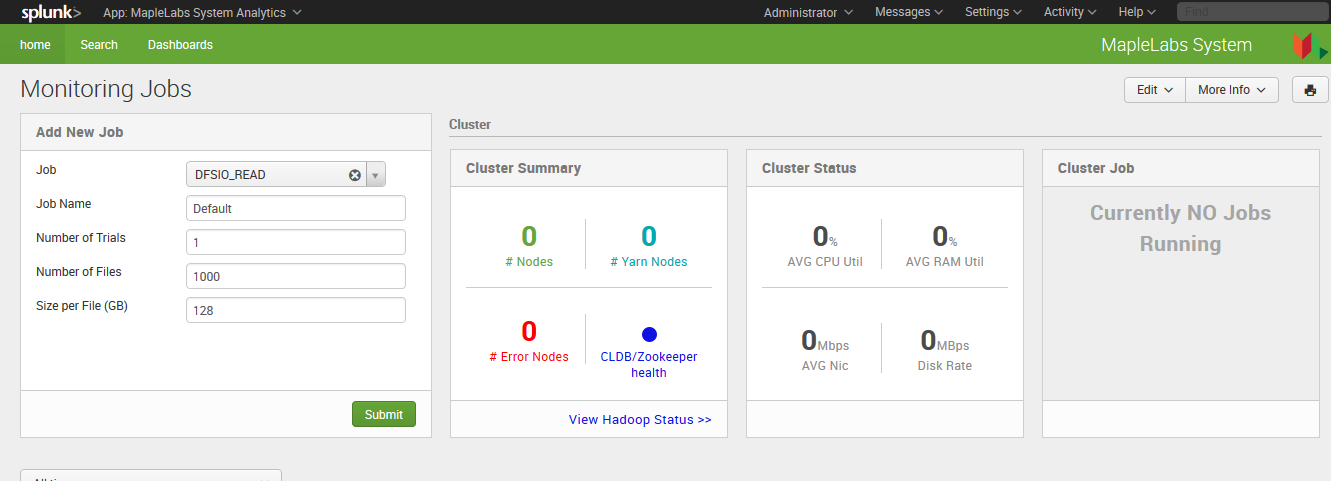
# Performance Monitoring

## Analytics application

Once you setup infrastructure, application and Splunk servers. Visit the Splunk application URL,

http://<splunk server url>

On the left side select Analytics application. This will load the main dashboard of the application.



## Main Dashboard

Main dashboard helps you to start a new job. Standard Hadoop jobs are prepopulated. This also has sections for providing information about the cluster summary and details, let of jobs running, a table consisting of job details.

## Additional dashboards

Select the link Dashboards to see other dashboards. These dashboards provides performance metrics for servers, switches and Hadoop job.

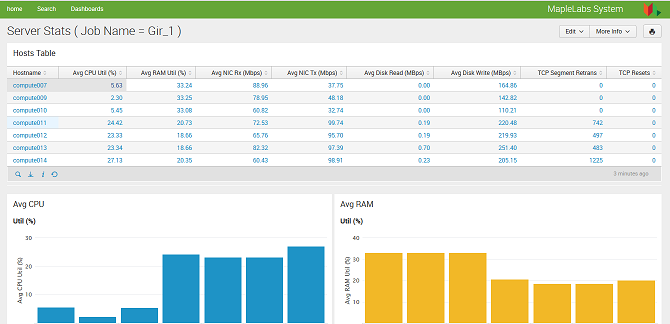
## How to use Analytics App

* Go to Main dashboard.
* Make sure no jobs are running on your test bed.
* Select specific job type. These are standard Hadoop jobs. Below are the job types.
  + Terragen
  + Terrasort
  + DFSIO Read
  + DFSIO write
* Provide a name to job.
* Number of trails to run
* Provide job specific configuration such as file size, number of files to create.
* Submit
  + Once job is submitted, job is executed on the test setup. The performance metrics are collected as the job executes.

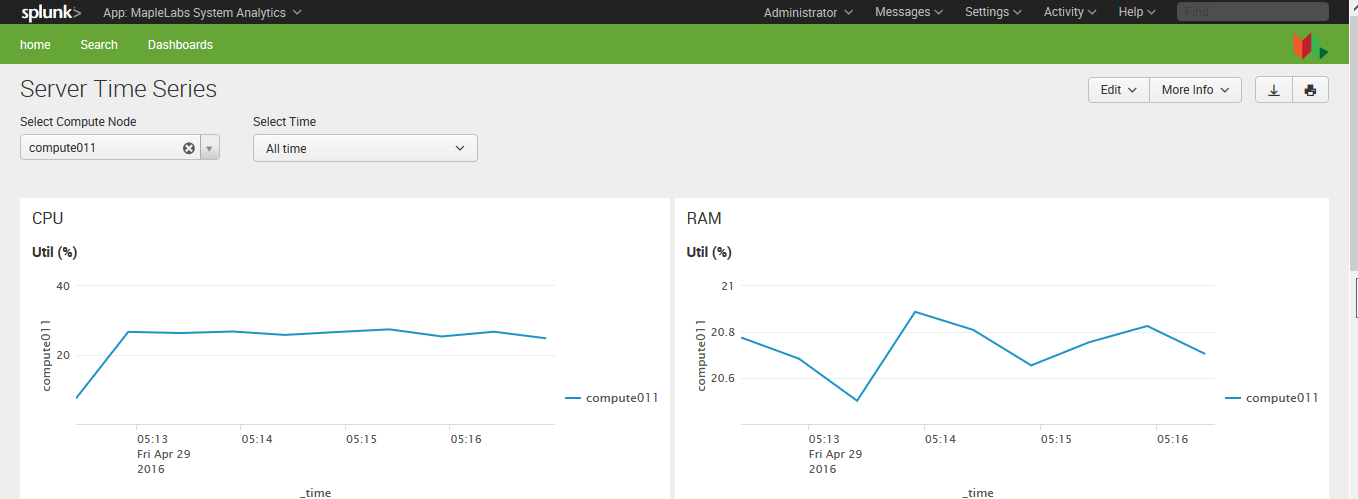
### Performance details

Job details table provides summary information about the jobs, job status, completion time, total maps, total reduces, average CPU utilization, Average RAM utilization, Average disk utilization, Average NIC RX and TX.

Select link server status to view the summary of all servers. Following is a sample server summary.



Click specific server to view server specific metrics. Below is a sample charts for a server.



Similarly select as switch to view switch performance metrics.

From the cluster summary section, click on view Hadoop stats to view Hadoop specific performance metrics.

## How to analyze the data.

* Start a Hadoop job as explained in previous section.
* View the cluster summary and cluster status.
  + If there are errors or non uniform utilization of CPU, memory or network, then view the status of servers, switches to make sure there is uniform load across all nodes.
  + View the server/switch parameters, specifically, CPU utilization, RAM utilization, Network utilization, switch buffer utilization to analyze which parameter is becoming bottle neck
  + Run the test multiple times to collect data and arrive at average values for benchmarking results.