Cloud DevOps Resources

Project Proposal

In collaboration with Jackpine Technologies, we will be developing assets for use in their primary product, CONS3RT, a cloud orchestration and service management software suite. The suite allows a user to build a system by picking an operating system and software that he wants the system to run. Systems can be run individually or combined with other systems to make up a scenario. CONS3RT supports multiple cloud services so the user decides which cloud provider he wants to deploy the system/scenario to.

1. <u>Vision and Goals Of The Project:</u>

Our main goal in one line: **Create useful assets that can be combined with other assets and included in systems through hassle-free setup of environments for applications.** In more detail, we will build new and useful assets for the Cons3rt library as well as improving existing assets if needed. Currently the assets library includes many useful applications but in cooperation with our mentor we have identified a few final products that would require new assets and modification of existing assets, the possible final products are:

- Hadoop Cluster
- LAMP stack (web server, database)
- Node.js Dev environment (IDE workstation, web server, database)

We will only focus on one project at a time and we will start with the Hadoop Cluster. We will build all required assets for a Cons3rt user to be able to set up such a cluster with minimal effort. Once deployed, the user will be able to run parallel processes for analyzing large amounts of data efficiently. It will be a multi-node cluster consisting of master and slave systems, and it will be

dynamically scalable to any number of machines. As of right now our focus is on delivering the Hadoop Cluster and only when all the moving parts of the system function as expected we will look into other projects, so by the end of the semester we will have added multiple useful assets to the Cons3rt library as a final product.

Users/Personas Of The Project

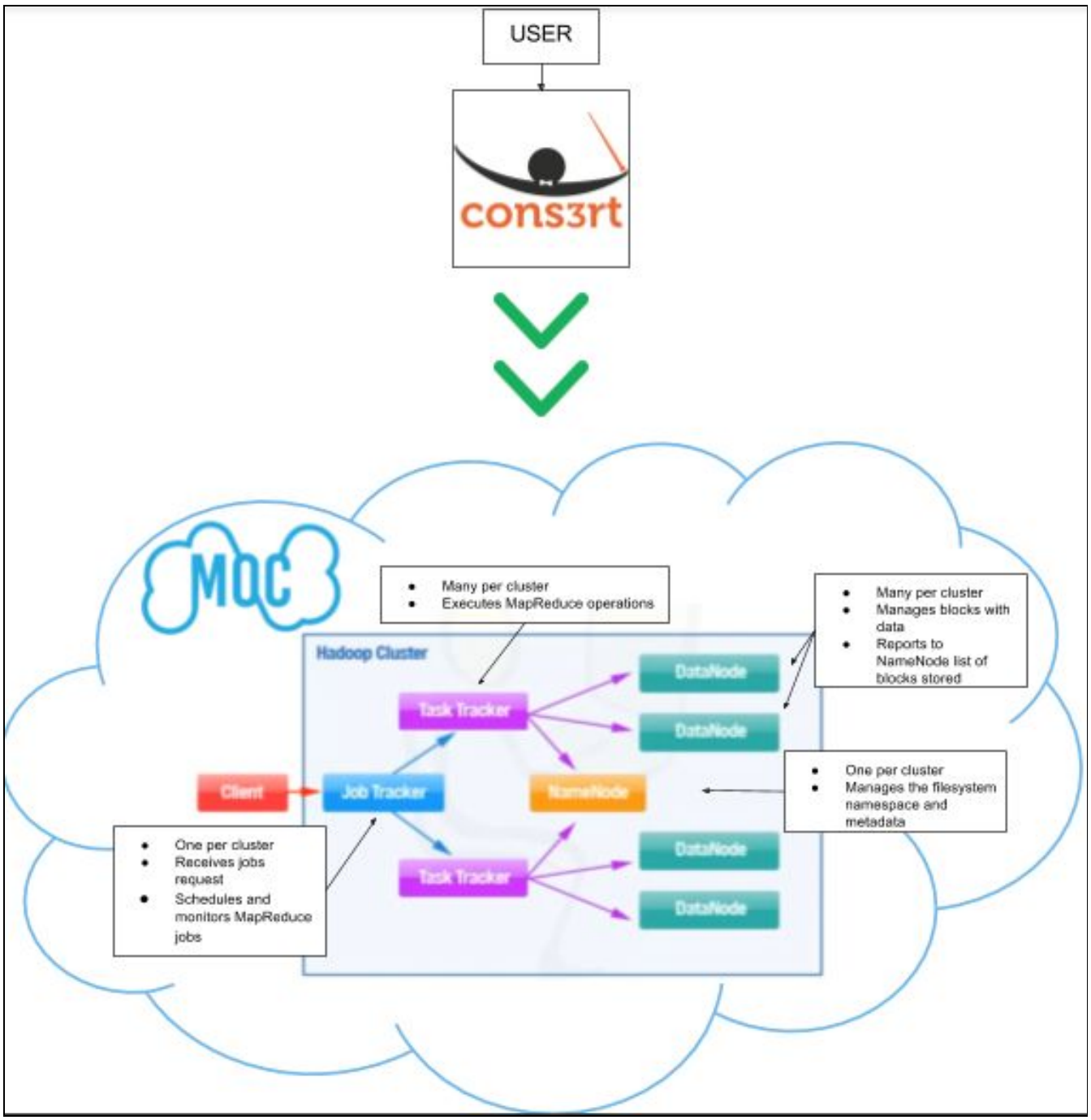
Hadoop is a open-source software library for reliable, scalable, distributed computing. A wide variety of companies and organizations currently use Hadoop for both research and production purposes.

We will provide a quick and easy way for anyone that requires analysis of big data to deploy a Hadoop cluster in any cloud and run their application on the cluster. Examples would be corporations, whose size is between medium and large, since they will have large amounts of data to analyze. Applications that generate this kind of data would be social networking sites, financial data analytics, network traffic security analysis, etc.

2. Scope and Features Of The Project:

For the Hadoop Cluster:

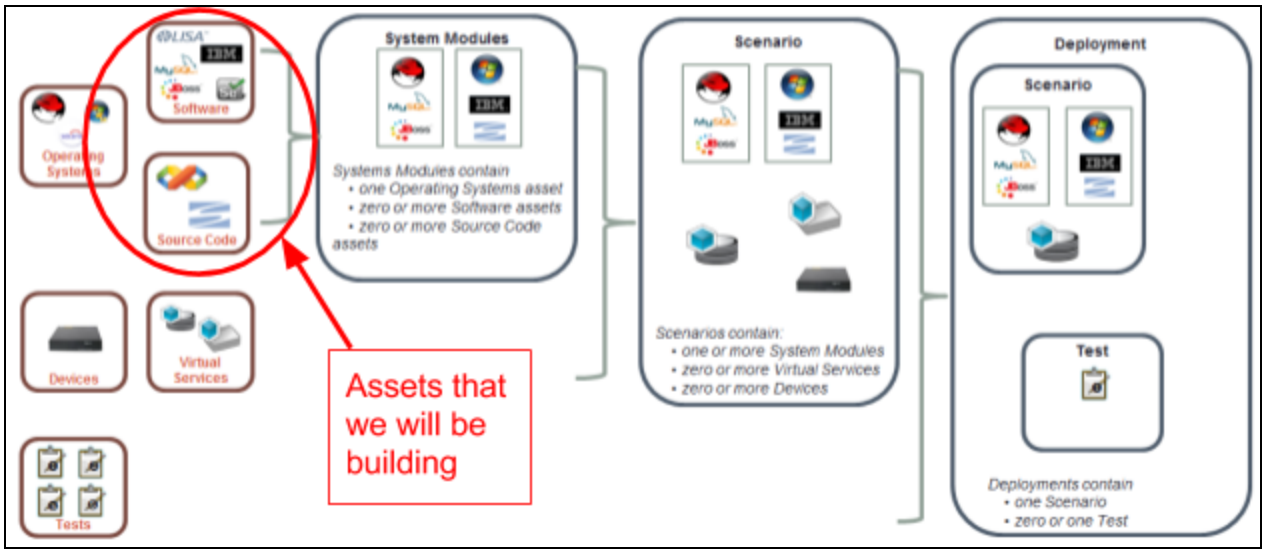
- Provide an easy accessibility to a Hadoop Cluster through CONS3RT
- Allow the cluster to be scalable, but also have limited resources, which will be configured by the user



- Scope of other project assets is not known at the moment

3. Solution Concept

Below is diagram that describes the building blocks of a deployment. We will primarily be building application software assets.

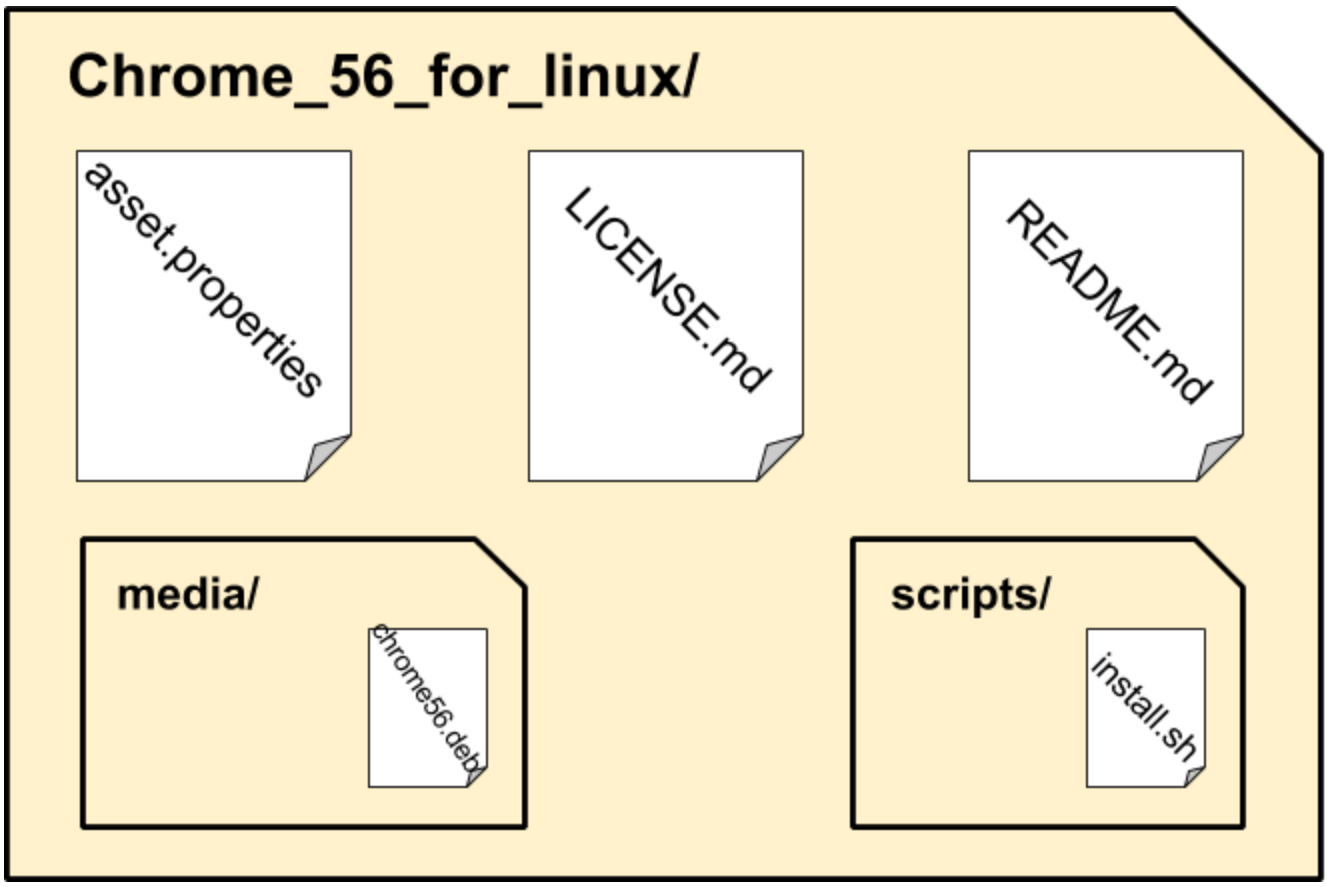


Global Architectural Structure Of the Project:

Software asset structure:

- The asset's main directory
  - Asset.properties
  - README.md
  - LICENSE.md
  - media(sub directory)
    - install media files
  - scripts(sub directory)
    - install.sh (Linux)
    - install.py (Linux)

Each asset that we build is it's own directory. Following is an example of a simple Google Chrome installer asset (built for a user that wants Chrome pre installed on a system):



asset.properties includes definitions for the Cons3rt system:

```
# Google Chrome Browser Asset properties file

assetType=software
softwareAssetType=Application
name=Chrome Browser v56 for Linux
description=Google Chrome Browser
installScript=install.sh
licenseFile=LICENSE.md
documentationFile=README.md
applicationOsFamilyType=LINUX
itarRestricted=false
instanceLimit=0
vendor=Google
softwareVersion=56
applicationArchitecture=X64
applicationBits=BITS64
pocEmail=david92@bu.edu
pocPhone=8574157238
pocName=David Asbjornsson
pocOrganization=Boston University
```

LICENCE.md includes a link to a licence page, apache licence for example.

README.md includes information about the asset, prerequisites and exit codes.

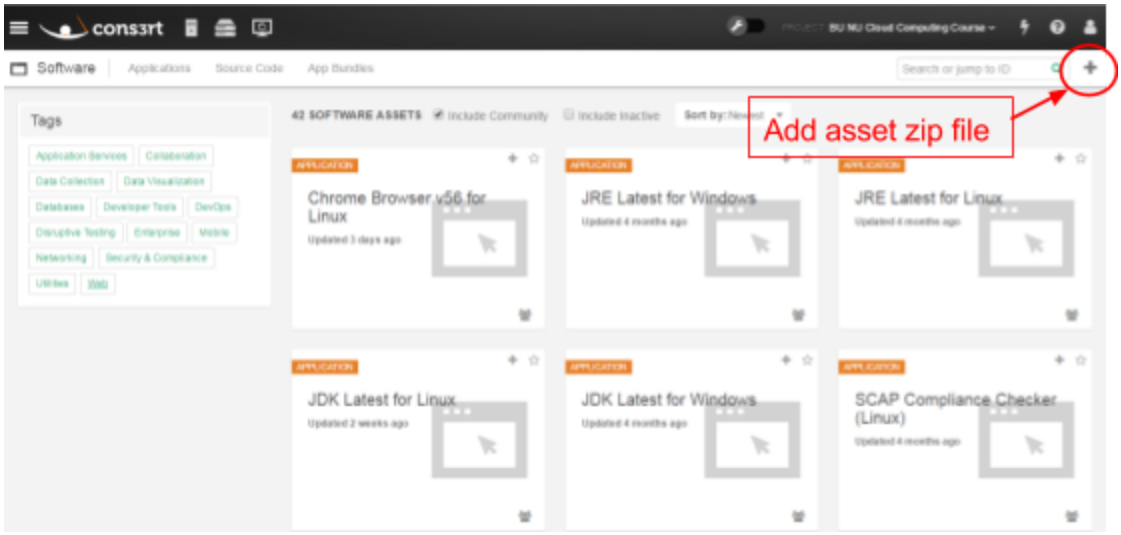
media/ is a sub directory that includes the actual installer file: .zip, .tar, .dep etc.

scripts/ is a sub directory that includes the install script, written in Python or Bash, that directs the Cons3rt system and allows it to install the file located in the media directory with specific modifications etc.

```
#!/bin/bash

# Created by David Asbjornsson  (2/22/17)

########################## INSTALLATION ##########################

# Locate the media directory
mediaDir=${ASSET_DIR}/media
# Get installer file name
chromeInstaller=$(ls ${mediaDir})
# Install Chrome using the .deb file
dpkg -i "${mediaDir}/${chromeInstaller}"

######################### EXIT CODES #########################

exitCode=0

echo "Exiting with code ${exitCode}"
exit #{exitCode}
```

The script files are the main component in each asset and a majority of our time will go into building and debugging these files. Note, this script file is for a very simple Chrome installer asset, the script files will get more complex.

From this point we zip the main asset directory and add it to the Cons3rt library through their web portal:



4. Acceptance Criteria

Minimum acceptance criteria would change for every asset the we develop. Currently, since we know we are doing the Hadoop Cluster for one of them, a minimal acceptance criteria would be a scenario of a master machine and multiple slave machines running the Hadoop software. The machines would be homogeneous and running the same version of the operating system, most likely the latest version of Ubuntu, which is 16.04.1. In addition, the hardware on the system will be the same, most likely intel x86_64 architecture. Also, we will have only two nodes, a master and a slave. Once we are able to do that, it should be fairly easy to scale to more slave machines.

5. Release Planning

Release #1 (Week 5):

A simple asset

Create a simple installer asset that can be deployed on a system

Release #2 (Week 7):

Hadoop Cluster

Build assets in order to deploy a multi-system Hadoop cluster

Release #3 (Week 9):

Another Useful Asset

Release #4 (Week 11):
Another Useful Asset

Release #5 (Week 13):
Another Useful Asset