The Process of Mirroring of LAB Sources

Thirumal Ravula

2014-02-14

Contents

1. Introduction

The sources of all the labs are version controlled at IIIT, Hyderabad. These sources are backed up every night. Apart from this backup, it is thought to backup the same sources at a different location. Mirroring provides a mechanism to accomplish this task.

2. Objective

This document describes different approaches taken to provide the means to mirror the lab sources from IIIT, Hyderabad to IIT, Delhi. This document makes the case for one of the alternatives and provides the status of all the efforts made to mirror the sources at IIT, Delhi

3. Design Alternatives

Two design alternatives are provided and a case is made for one of the alternatives and the reasons for choosing one alternative over the other are also delineated.

3.1 Alternative I - PUSH from IIIT, Hyderabad

In this alternative, all the sources would be copied to a mirror machine - VM or a container. This machine should allow ssh access to the outside world so that the machine containing sources at IIIT, Hyderabad would copy the sources initially and later transfer only the modifications on a nightly basis to the mirror machine.

3.1.1 Considerations

rsync is the utility that is used to upload the files and this utility has the intelligence to upload only the deltas on a nightly basis.

The size of the sources is approximately 50 GB and with a 2Mbps link it would take days to transfer over the link. To enable faster copy of the sources initially, an external hard disk should be provided to the contact at IIT, Delhi that will contain:

- LAB Sources.
- a script to copy the contents of the hard disk to a storage location that would retain the timestamps and permissions of the files.
- README file containing the instructions and location of the LAB sources and the script.

A mirror machine - either a VM or container - should be created at IIT, Delhi. This machine should be accessible via the ssh port by the outside world. The contents of the external hard drive should be copied to a storage location and this location should be mounted during the startup of the mirror machine.

A script should run as part of cron job every night at IIIT, Hyderabad to push the changes incrementally to the mirror machine at IIT, Delhi.

3.1.2 Status

This section provides the status of the tasks.

The hard drive containing the lab sources, script that allows copying of the contents of the hard drive and README is given to the contact at IIIT, Delhi.

The creation of the mirror machine at IIT, Delhi is pending and therefore all other subsequent tasks.

3.2   Alternative II - PULL from BITBUCKET

In this alternative, all the lab sources will be uploaded to BitBucket - a web-based hosting service based on GIT. The repositories on BitBucket will be updated on a nightly basis by running a script as part of cron job on the machine containing sources at IIIT Hyderabad. Any mirror that will have to be created will only have to pull the files on a periodic basis from this repository-hosting site on the internet.

3.2.1 Considerations

An 'organization' user will be created in BitBucket who will own all the lab source repositories. This organization will have two types of groups - ADMIN and DEVELOPERS. For every mirror site, a user will be created who will be part of the 'DEVELOPER' group.

The repositories on the BitBucket are updated by running a nightly job on the source machine at IIIT, Hyderabad. This job will push all the GIT repositories. All the BZR repositories will be pushed using a command - dpush - provided by BZR to the GIT repository. All the SVN repositories are first transformed to GIT and then pushed.

A mirror machine - either a VM or a container - should be created with http/https ports open to the outside world. Using these protocols, the source files will be pulled by the mirror machine. For the very first time, all the source files will have to be pulled. After that, only the updates will be pulled on a periodic basis. This is accomplished by running a script as part of cron job on the mirror machine.

3.2.2 Status

This section provides the status of the tasks.

All the repositories are created on BitBucket - an internet repository-hosting site and the repositories are pushed on a periodic basis every night by a script running as part of cronjob on the sources machine at IIIT, Hyderabad.

All these repositories on the internet are part of an organization. Groups 'ADMIN' and 'DEVELOPER' are also created. 'iitdelhi' user is made part of the 'DEVELOPER' group.

The script that will pull the repositories to mirror machine from BitBucket is completed.

A mirror machine - either a VM or container - should be created. The script that pulls the sources from the BitBucket should be installed on the mirror machine and made part of a cron job.

3.3   Design Chosen

Alternative II is chosen as the mirroring option. This design allows setting up multiple mirror sites without any onus on IIIT, Hyderabad. Any mirror site will only have to plug-in the 'pull script' in the cron job which will pull the repositories from the internet based repository hosting site - BitBucket. In this alternative, the responsibility of IIIT, Hyderabad is to maintain a mirror on BitBucket.

4.   Create a Mirror

The following steps are necessary to create a mirror:
A.   A user on BitBucket should be added to Virtual Labs Organization as a 'developer'. This step should be performed by the VLEAD.

The following steps should be performed by the person creating a MIRROR:
B.   Clone the repository vlead/simo (https://github.com/vlead/simo) to a machine on which mirror is supposed to be created by running the command -
git clone https://github.com/vlead/mirror-agent.git, This will create a folder simo, change to this directory.
C.   Follow the steps listed in README_FORMIRRORING.md

5.   Conclusion

A mirror is already enabled on BitBucket now. To create multiple mirror sites, it is upto to the interested parties to create without any dependency on IIIT, Hyderabad.