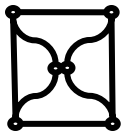
NETWORK DEPLOYMENT PLAN

MetaArchive of Southern Digital Culture Project



2005-03-11

Summary

This document describes the plan for deploying the Preservation Network of the MetaArchive of Southern Digital Culture Project. This deployment equates to Project Phase D1 in the project plan. Required steps and responsibilities are detailed for this phase, and notes concerning subsequent phases D2 – D4 are summarized. This document makes reference to the project plan and other project documents which provide more details about other aspects of the project.

Terminology

The MetaArchive project will establish a preservation network based on the LOCKSS software. Because this network will be configured distinctly and function separately from the original LOCKSS network, the constituent servers be referred to by a differentiating name: CLOCKSS, for *Collecting Lots of Copies Keeps Stuff Safe*. This slight name change is simply intended to help clarify when we are talking about the reconfigured and separate servers comprising the MetaArchive preservation network, as opposed to the standard LOCKSS servers comprising the LOCKSS network proper. The name is meant to be evocative of long term preservation.

Where

The four Development Sites identified in the project plan (Emory, FSU, GA Tech, and VA Tech) will be responsible for installing, configuring, and testing the initial CLOCKSS preservation nodes during the first portion of Project Phase D1 (Network Deployment). Louisville and Auburn may activate their nodes during this phase, but will not be required to do so until the next project phase (see timeline below).

What

The MetaArchive preservation network will ultimately be comprised of identically configured preservation node server pairs, each comprised of a Vault and Gate as described in the project plan. The Vaults are planned to include large Dell AX-100 storage arrays attached to Dell 1850 server units. The Gates are currently also specified as dedicated Dell 1850 servers. The Gates will have Red Hat Enterprise Server (RHES) version 4.0 Linux installed for the operating system, and the Vaults will have RHES 3.0 (because of driver compatibility issues with the AX-100 storage arrays). RHES 4.0 has hardened security features appropriate for the Gate functions. The Vault servers will have the CLOCKSS software installed, in addition to moderate security features.

Although the above describes the configuration of equipment ultimately planned for each node in the network, the full complement of equipment is not necessary for the deployment of the initial network and harvest. The minimum that is required for Phase D1 is that each of the Development sites establish a functioning CLOCKSS server as a Vault. The dedicated firewalls (the Gates) are not

required to accomplish the basic network configuration and initial harvesting activities slated for this phase of the project. In actuality, even the AX-100 storage units are not required for the Vaults to function, as the Dell 1850 units come with more than 70 GB of mirrored internal hard drive space. The AX-100 units are obviously needed for the project goal of a composite archive approaching 3 TB, but the initial development Vaults could simply be the Dell 1850 servers that we are ordering, or even (in a pinch) any PC that meets the minimum requirements to run the software (for those specifications, see: http://www.lockss.org/installing/installing.htm). The main consideration relevant to this phase of the project is how much space the collections targeted for the initial harvest will require. What each of the Development Sites actually brings up during Phase D1 (as opposed to subsequent phases) may therefore be seen as an issue separate from the key network deployment activity of bringing up a CLOCKSS server at each Development Site.

The core tasks associated with deploying the initial version of the network are the following, which each Development Site needs to accomplish according to the timetable below:

1. *Equipment Installation:* The equipment to comprise the Vault (with or without the AX-100 storage array) must be assembled.

2. *OS Installation:* Performing an installation of the hardened version of Linux on the server to become the local Vault.

3. *CLOCKSS Installation:* The reconfigured LOCKSS software (CLOCKSS) has to installed on the Vault. Will entail preparatory steps of downloading and rebuilding the LOCKSS daemon.

4. *Testing:* The resulting Vault must be tested by harvesting a first trial collection to make sure it is working.

5. *Initial Harvest:* The collections identified from the conspectus for the first harvest, must be aggregated according to the harvest plan.

All six Vaults comprising the primary preservation nodes of the project must be deployed by the conclusion of Project Phase D2 (Network Synchronization), including the AX-100 storage arrays. The Gate firewalls may also be deployed during this phase, or may be deployed in Project Phase D3 (Network Testing).

How

None of the core tasks described above has so far been specifically undertaken, although the project staff members obviously have experience in activities similar to many of these tasks (server OS installs, for example). Notes will be recorded on the project WIKI by the technical staff of the project as they begin work in earnest on the five core tasks described above. A separate notes page for each of these five tasks will be created and systematically built up for subsequent reference by other sites. The first three core tasks can potentially be investigated in parallel.

While the documentation for the first three core tasks is being completed, the other contributing technical staff should begin work, actively asking questions along the way of the individuals specifically investigating the relevant topics. The overall goal is to have all the requisite parts of the network functioning by the time we need to begin the initial harvest.

Who

The respective members of the Steering Committee will be responsible for ensuring that the required activities for the Network Deployment take place according to the timetable set forth below. Technical staff working on the project will be the ones who actually do much of the work in practice, but the Steering Committee members are the individuals responsible for ensuring that the work gets done in time to meet the project timeline.

The technical staff will comprise a project working group, will take responsibility for investigating the five core tasks listed on the previous page, and for documenting findings on WIKI pages associated with the respective core tasks. The Technical Working Group will include at least the following individuals: Thomas Robertson, Robert McDonald, Kyle Fenton, Johnny Healey, Curtis Carr, Larry Hansard, and additional technical hires at the partner sites as they become available.

When

May 20:	Documentation complete on how to configure the hardened Red Hat OS on the Dell 1850 units, and the details of the CLOCKSS configuration and installation
June 15:	Target for getting Development Site nodes operational
July 1:	Target for starting the initial harvest
August 1:	Drop-dead for getting Development Site nodes operational and starting the initial harvest

Notes on this document

This document was completed at the second MetaArchive All-Project Meeting held at Emory University on Friday 2005-03-11, by the Steering Committee and additional attending project participants. The document was last updated by Martin Halbert on 2005-03-14.