DAITSS 1.2 Installation

Introduction

DAITSS (Dark Archive in the Sunshine State) is a digital preservation system written in Java at the Florida Center for Library Automation (FCLA). DAITSS processes digital content for long-term archiving and preservation. This guide describes the installation process for the binary distribution of DAITSS version 1.2.

System Requirements

DAITSS requires Java 5 or later, and is meant to run on a Unix or Linux operating system. The installation has been tested on both the Red Hat Enterprise Linux 4 and Ubuntu version 6.10 LTS Linux distributions; we expect other Linux systems to support DAITSS. The installation script requires the `bash` shell and makes use of the standard Unix utilities `tar`, `gz`, `sed` and `awk`.

This guide does not address the use of back-end storage management modules such as the Tivoli Storage Manager. Refer to the DAITSS 1.2 Operations Manual for more information.

This DAITSS distribution is installed in a minimal configuration, and requires the following software:

- MySQL Version 4.x or later, with InnoDB data table support. It must be installed on the same host the runs DAITSS.
- The Sun Java Virtual Machine (JVM), version 5 or later. The JVM will be present if you have installed Sun's JRE, J2SE or J2EE systems.
- A working Mail Transfer Agent (MTA) such as Sendmail or Postfix.

Optional Programs

DAITSS can make use of additional programs for verifying the integrity and migratability of certain formats (a process called normalization); these programs are not distributed and their use has been disabled in the distributed `daitss.properties` configuration file. See the operations manual for details on enabling these programs. In particular, it is strongly recommended that you download and use the `clamav` virus checker. All are open source projects available, at the time of this writing, from the following locations:

- Clam AntiVirus, a GPL virus scanner for Unix systems.
- Ghostscript, a PostScript interpreter and toolkit.
- ffmpeg is a set of multimedia encoders, decoders, muxers and demuxers.
- mencoder, part of the `mplayer` suite of programs, is a simple movie encoder designed to encode video to various formats supported by `mplayer`.
- lqt_transcode, part of `libquicktime` project, handles quicktime/AVI video formats.

Installation Process

Quick Installation

As root, unpack the distributed tar file, cd to the created directory, and run the install script, answering the configuration questions:

```
bash$ tar xzvf daitss-1.2-binary.tar.gz

bash$ cd daitss-1.2-binary

bash$ ./install
```

Installation Details

This section describes the actions taken by the `install` script. The `install` program proceeds in two steps. The first step checks the system and collects configuration information from you, and is followed by prompt to begin the install. No changes are made to your system until you have a chance to review your entries, and explicitly start the install.

The install script first checks that it is running as `root`. The next check makes sure that the `mysql`, `rmiregistry` and the `java` programs are in your PATH environment. Further, the `java` program must be the Sun Java virtual machine, version 5 or higher.

If these basic requirements are met, the install script prompts for the install location (you may press enter to accept the default values):

```
Ready to configure the three installation directories.

Please enter the DAITSS home directory (default: /opt/daitss):
```

The DAITSS home directory will contain executables, scripts and configuration data; it requires 10 MBytes of free space.

```
Please enter the directory for processing DAITSS data (default:
/var/daitss):
```

The above directory is the location for the temporary storage and processing of documents to be archived. Our experience has been that as many as 3500 files totaling 60 GBytes can be processed in one day. Another way to estimate your storage requirements would be to consider the number of packages received; packages are self-contained entities such as a dissertations and are typically comprised of several document files. We have found packages average 48 MBytes, with a small amount of additional overhead.

```
Please enter the DAITSS log directory (default: /var/log/daitss):
```

The log directory roughly mirrors the layouot of the DAITSS data directory.

The account running DAITSS processes will need write access to these directories and files.

The next part of the `install` script sets up the `MySQL` database. As an example of database storage requirements, The Florida Digital Archive (FDA) project requires 1.9 GBytes disk space for its database files.

```
Collecting information for setting up the DAITSS MySQL database.
You will need administrative access.
NOTE: if this database exists, all the data in it will be deleted.

Please enter the DAITSS database name (default: daitss):
```

Please note that this part of the install will remove any pre-existing data for the specified database.

```
The DAITSS database will be configured for a user named
'daitss_user'.
Please enter the password for this user (default: top secret):
```

The DAITSS system requires a database username and password in the installed configuration file, `daitss.properties`. This MySQL account is granted access only from `localhost`.

```
Please enter a destination e-mail address for DAITSS reports
(default: daitss@you.edu):
```

This email address will be used to send DAITSS reports to repository administrators.

```
Please review the information you have entered:

System software installs to: /opt/daitss
Files will be processed in: /var/daitss
Log files will stored in: /var/log/daitss

DAITSS database name: daitss
DAITSS daitss_user password: top secret

Reports will mailed to: daitss@you.edu

Would you like to change any of the above information? (Y/n):
```

Review the information and enter `Y` if you would like to change your installation settings. Otherwise, press `N` to proceed with the installation.

At this point the install process will perform the following actions:

- Attempt to create the unix `daitss` group using the `addgroup` command.
- Create (or recreate, if it exists) a MySQL DAITSS database and grant access to the `daitss_user` user account for accessing it from `localhost`.
- Make directories in the DAITSS data and log directories.
- Create configuration files, copy the java programs and install scripts to the DAITSS home directory.
- Set permissions on the new directories limiting write access to members of the `daitss` group

```
The scripts ingest, prep, and rc.daitss have been installed
into /opt/daitss//bin. These scripts should be put on
your DAITSS users' path.

The program rc.daitss needs to be executed once before
the DAITSS programs can be run. The script rc.daitss starts
the rmiserver program, and a DAITSS OID-server program as
background processes.

We recommend that rc.daitss be copied to your system start-up
directory (normally /etc/init.d) and made part of your
standard boot up sequence. One way to do this is to execute
the following sequence of commands:

cp /opt/daitss//bin/rc.daitss /etc/init.d/daitss
ln -s /etc/init.d/daitss /etc/rc0.d/K91daitss
ln -s /etc/init.d/daitss /etc/rc1.d/K91daitss
ln -s /etc/init.d/daitss /etc/rc2.d/S91daitss
ln -s /etc/init.d/daitss /etc/rc3.d/S91daitss
ln -s /etc/init.d/daitss /etc/rc4.d/S91daitss
ln -s /etc/init.d/daitss /etc/rc5.d/S91daitss
ln -s /etc/init.d/daitss /etc/rc6.d/K91daitss

Would you like to start the DAITSS background processes now? (y/N):
```

Post Install Actions

Please see the DAITSS 1.2 Operations Manual to create DAITSS accounts and preservation policies. At minimum, it is necessary to create a contact, account, project, and default preservation rule to begin with DAITSS.