

DSpace Server Migration

Author: *Steve Thomas,
University of Adelaide Library*

Last Update: *13/06/2007 4:32 pm*

Abstract: *Describes the procedures for migrating a DSpace instance to a new server platform.*

Problem

Our installation of DSpace was initially done on a redundant Sun E450 running Solaris 8. As we moved quickly into production, it became apparent that we needed to migrate to a new platform, with faster and more robust hardware.

Early in 2007, we received funding for a new server (specifically an HP DL320s 3060 A1 AP Server with 1TB of storage and 4GB ram). This was installed by April, and we have since then been engaged in setup and installation. This report will summarise and describe the issues encountered.

Issues

The major hurdle was the differences between Solaris 8 and RHEL, the operating system chosen for the new server. While DSpace uses standard off-the-shelf open source software, differences in implementation exist and caused some confusion. In addition, the supported packages with RHEL are some several versions away from the latest release versions used on the existing Solaris box. The largest stumbling block was with PostgreSQL, which RHEL includes as version 7.4, while the latest version (used on Solaris) is 8.2.4. Additionally, DSpace requires that PostgreSQL 7 be compiled with specific options, and there is no way to tell whether the RHEL version included those options. And finally, since we wanted to use Perl for extraction and data massage, we needed the DBD::Pg module, which required a Postgres development kit.

Happily, all these issues were resolved (once understood) with the installation of the up to date Postgres packages, so were finally (by June) ready to proceed with the actual DSpace migration.

DSpace Server Migration

Installation

The first step was the installation of the DSpace software. This is well documented in the DSpace installation notes and need not be described here, except to mention that we also installed the mod_jk perl module to allow access to the DSpace database through Apache.

Configuration

Configuration is made trivial by the fact that DSpace is java-based. It is therefore possible to simply copy the dspace.war file from the old server to the new, and install in the tomcat webapps directory as per standard instructions. The dspace.cfg file is also copied. Then starting tomcat completes the configuration.

This allows us to continue to use the old server as a development box for configuration of the user interface, since we can trivially copy the dspace.war file to the new server once changes have been tested.

Migration

Migration of data turns out to be almost trivially easy, as follows:

1. Dump the DSpace database on the old server. We were already doing this nightly, using the following script:

```
#!/bin/sh

DATE=`date "+%Y%m%d%H%M"`
LD_LIBRARY_PATH=/usr/local/lib
export LD_LIBRARY_PATH

/usr/local/pgsql/bin/pg_dump \
    --verbose \
    --oids \
    --create \
    --format=c \
    --username=dspace \
    dspace \
> /data/dspace/dump/$DATE

exit;
```

2. Copy the dump to the new server.
3. Restore the DSpace database using the Postgres pg_restore function:

```
dropdb -U dspace dspace
```

DSpace Server Migration

```
pg_restore -C -d postgres -U dspace dumpfile
```

So far so good. But DSpace keeps the actual document files outside of the Postgres database, so it is also necessary to copy these across to the new server. We do this with rsync:

```
rsync -va \  
digital.library.adelaide.edu.au::home/uals/dspace/assetstore/ \  
/data/dspace/assetstore/
```

Similarly, we can use rsync to copy the (non-critical) parts of DSpace – log files and index files. Although it seems sensible to rebuild indexes on the new server using the index-all command.

Validation

Migration is not complete without validation: verification that all data was successfully transferred and that everything works as before.

Our validation plan includes the following steps:

1. visual comparison between the old and new systems, using specific, randomly chosen records;
2. using SQL queries to compare row counts for selected tables, e.g.:

```
dspace=> analyze verbose item;  
INFO:  analyzing "public.item"  
INFO:  "item": scanned 251 of 251 pages, containing 30029 live rows  
and 0 dead rows; 3000 rows in sample, 30029 estimated total rows
```

References

Include references as relevant.

DSpace System Documentation

<http://www.dspace.org/technology/system-docs/>