DRTC-HP International Workshop on
Building Digital Libraries using DSpace
7th – 11th March, 2005
DRTC, Bangalore

Paper: D

An Introduction to DSpace Installation

ARD Prasad

Documentation Research and Training Centre
Indian Statistical Institute
Bangalore 560059

ard@drtc.isibang.ac.in

1. Introduction

This article attempts to explain configuration of DSpace in simple terms and does not cover Tomcat integration with Apache or Secure Socket layer (SSL). As DSpace is mostly configured in Linux, the following sections are applicable to Linux, in particular Redhat Linux. The other Linux flavors may have different directory structure; so some of the system files may reside in different directories than the ones mentioned here. For example, sendmail and postgreSQL files. This is typically the case with other Unix versions like HP-UX, Solaris, or AIX. Experienced Unix administrators should not find it difficult to load DSpace in any Unix environment. Although DSpace can be installed in MS-Windows operating system using Oracle or using postgreSQL in cygwin environment, it is strongly discouraged for a variety of good reasons.

2. About Linux

It is always a good idea to load full Linux, rather than choosing either server or workstation options. In most of the cases that I came across, either one or the other services or modules (like Java Database Connectivity - JDBC drivers for postgreSQL) would be missing, when Linux was loaded partially.

To load DSpace you require super user password. If you are new to Linux, take the help of your system administrator. As such DSpace installation is simple, however, it uses many Linux services like mail, postgreSQL, Domain Name Service (DNS) etc. Mail and other services are to be properly configured for a successful installation of DSpace.

3. Required Linux Software and Services

3.1 Mail

DSpace heavily uses mail in
        a. Registration of new members
        b. Login
        c. Sending notification to authors, reviewers, metadata validators etc
        d. Sending message to subscribers of any new additions to a collection

3.2 postgreSQL

This is the backend database of DSpace, which stores information about

        a. Communities
        b. Collections
        c. Members, their e-mail addresses and passwords in encrypted form
        d. E-groups (reviewers, metadata validators for each collection etc.)
        e. Metadata of digital items etc.

3.3 Domain Name Service (DNS)

The purpose of DNS is to get IP number for a given IP address. For example, if you try to access **drtc.isibang.ac.in**, it gets translated to 210.212.206.71. Of course, this is quite transparent to the end user. Most likely your system might have been using the main server as the DNS server of your organization. If you prefer a separate name for your DSpace server, you have to ask your administrator to make an entry of the new host name in the DNS configuration.

4. Mail configuration

In Linux you can use either *sendmail* or *postfix* as mail server. Here, a very rudimentary *sendmail* configuration is explained. All the Linux services will have daemons to run the services in the background. In case of mail, the daemon is */etc/rc.d/init.d/sendmail*. You

have to make sure that the sendmail daemon is invoked at the boot time. It can also be run using any of the following commands:

> service sendmail start
> service sendmail stop
> service sendmail restart (To stop and start)

In addition, you may have to tweak mail configuration file.

> *Step 1:* open */etc/mail/sendmail.mc*

> *Step 2:* comment the line which looks like the following
> **DAEMON_OPTIONS('Port=smtp,Addr=127.0.0.1, Name=MTA')dnl**
> This line should look line the following
> **dnl #DAEMON_OPTIONS('Port=smtp,Addr=127.0.0.1, Name=MTA')dnl**

> *Step 3:* If you are using an external mail server, uncomment the following line
> **dnl # define(`SMART_HOS', `smtp.your.provider')**
> It should like the following
> **define(`SMART_HOS', `drtc.isibang.ac.in')**
> **Note:** *drtc.isibang.ac.in* is mail server for drtc, you can replace it by your mail server.

> *Step 4:* As .mc file is macro file, you have to use the *m4* command to generate mail configuration file, using the following command
> **m4 /etc/mail/sendmail.mc > /etc/mail/sendmail.cf**
> If *m4* fails to generate *sendmail.cf* file, it is likely you have loaded *sendmail* but not *sendmail-cf*. To load *sendmail-cf*, search the Linux CDs, use *rpm* command to load *sendmail-cf* file.

> *Step 5:* **#service sendmail restart**

Note: Check sending and receiving mail from your system, to make sure that mail server is working properly. If your host system does not have domain name sendmail may not work, as mail uses Fully Qualified Domain Name (FQDN). If issue the following command
> **hostname**
The response should be **hostname.domainname**, if only hostname appears, it means your system is not assigned a domain name. Ask your system administrator to assign FQDN to your system. Normally, you can see the related error message in the */dspace/log/dspace.log* file.

5.   postgreSQL Configuration

postgreSQL daemon is */etc/rc.d/init.d/postgresql*. Its configuration files are in the directory */var/lib/pgsql/data*. However, if *postgres* was not initiated before, the required files might not have been generated. To generate the required files, you have to run the *postgres* daemon. Once the required files are generated, you have to tweak the following files - */var/lib/pgsql/data/postgresql.conf* and *pg_hba.conf* in the same directory.

Also make sure whether JDBC drivers for postgreSQL are installed. The JDBC driver files should be copied to *dspace-source/lib* directory

6.   Java

Some Unix installation have java compiler loaded. However, it is a good idea to download Java compiler from Sun site. Also make sure the environment variable PATH is set to the Java directory and precedes over the regular path. For example, if Sun Java is in */opt/java,* your environment should be as follows:

> **PATH=/opt/java/bin:$PATH**
> **export $PATH**

7. Ant

Apache Ant is similar to *'make'* of C programs. It is used to generate *.jar* or *.war* file(s). It is essential to have *build.xml* file in the directory where you run *ant*. In case of DSpace, it generates two *.war* files viz. *dspace.war* and *dspace-oai.war* and both should be copied to the *tomcat/webapps* directory.

8. Tomcat

Java Server Pages (JSP) is similar to ASP (Active Server Pages) or PHP (Hypertext Preprocessor), a server side scripting language for Web interface. As DSpace is written in JSP, we require a servlet container. Tomcat is one of the many available servlet engines. The others are JBoss, Jetty, Resin, J2EE. JBoss is believed to be a more professional and powerful software. DSpace can be run directly on Tomcat without using the Apache web server.

9. DSpace Installation

Follow all the steps enumerated in DSpace documentation.

10. Troubleshooting

The following are broad tips if you run into rough weather.

10.1 *At the time of Installation*
1. If the system responds **"ant not found"** or **"javac not found"**. It could mean that ant or java are not in your path, check **/dspace/.bash_profile** and modify PATH values correctly.

2. Any java related problems might have been caused by not setting JAVA_HOME and JAVA_OPTS properly. Or, you do not have the right java compiler. Also make sure that the **$PATH=/opt/java/bin:$PATH**, so that even if you have another java compiler, the one loaded in **/opt** directory gets priority.

   And re-login to dspace, so that the modifications to environment variables like PATH are effective. You can check any environment variable with the following command:

   > **echo $PATH or**
   > **echo $JAVA_HOME**
   > **echo $JAVA_OPTS**

3. While running **'ant fresh_install'** if you get SQL related errors, it could mean either
   a. postgreSQL configuration was not done properly or
   b. the JDBC drivers have not been copied into the **lib** directory of your **DSpace source** directory
   Note: This is by far the most common problem, while installing DSpace

10.2 *While running DSpace*

Once installation is complete, open DSpace in the Web browser. If you do not see DSpace screen, check whether Tomcat is running. For example: **http://drtc.isibang.ac.in:8080** should show the Tomcat screen and **http://drtc.isibang.ac.in:8080/dspace** should show DSpace screen.

In all likelihood **8080 port** is already being used by some other service or another Tomcat is running on the same port. In which case, you have to choose another port for your tomcat. This can be achieved by replacing 8080 and the related ports in the **tomcat/conf/server.xml** file.

If you see the DSpace running, it is a good idea to check whether it is sending mail to newly registered members. If you get **"Internal System Error"** message, it could mean that the mail server is not configured properly. Also check in the **dspace.cfg** file, if the **mail.server** variable is properly set.

If DSpace is sending mail, but the link given in the mail is not accessible on the browser means, in the **dspace.cfg** file, the host name is not specified properly. It may be having **https**, though you are not using SSL.

If you get **"Internal System Error"** while creating communities, it could mean postgreSQL was not configured properly.

10.3   LOG Files

"Internal System Error" is too generic an error, it might have been caused by any configuration lapse. To locate the exact problem, go through the tomcat/logs/catalina.out file and also */dspace/log/dspace.log* file

11.   Check on the LAN/Internet

If everything is working fine, try to access DSpace from another system on your LAN and also from another system on Internet. If DSpace is working on your LAN but not from outside, you are behind a firewall. You may have to ask your administrator to open 8080 port in the firewall.

12.   Further help

You can join the DSpace Mailing Lists mentioned at DSpace home page: **http://www.dspace.org**.   You can also join DRTC DSpace mailing list at **http://drtc/isibang.ac.in/dlrg**

13.   Sites for downloading the required files:

Ant:	*http://ant.apache.org*
Tomcat:	*http://jakarta.apache.org*
Java:	*http://java.sun.com/j2see*
DSpace:	*http://www.sourceforge.net/projects/dspace*
DSpace Homepage:	*http://www.dspace.org*