

利用DSpace构建机构知识库

祝忠明
中国科学院国家科学图书馆兰州分馆
3.12-3.13
DLIB & OSS 2009 北京

提 纲

- DSpace的安装与运行环境
- DSpace的安装部署
- DSpace的本地化

一、DSpace的安装与运行环境

硬件平台

- 平台类型
 - PC服务器, 或
 - Unix/Linux服务器
 - 具体配置与拟建设IR系统在数据规模、保存、备份、性能等方面的需求相关。
- 常见配置:
 - PC服务器:
 - 主流双Intel Xeon处理器, 2~4GB 内存, 500G以上外存。
 - 若需数据离线备份, 考虑配置内置或外置磁带机(库)。
- 低配置考虑:
 - 单CPU主流配置PC服务器
 - 内存1G以上、500G以上外存

基础软件环境

- 操作系统:
 - 默认首选: Linux/Unix服务器操作系统
 - 推荐: Ubuntu 8 服务器版(免费), 或Red Hat Enterprise Linux(商业)。
 - 可选: Windows XP/2000系列操作系统(商业)
- 数据库管理系统: PostgreSQL 7.3+ (免费)
- Web应用服务器: Tomcat 4.0+ (免费)
- Java平台: J2SE 1.4+ (免费)
- ANT构建工具: ANT 1.5+

二、DSpace的安装部署

安装目录规划

- **JDK、ANT、Tomcat:**
 - 安装在/opt下
- **PostgreSQL的安装目录:**
 - /usr/local/pgsql
- **Dspace数据及配置文件安装目录:**
 - /dspace

安装JDK

- 从 java.sun.com 下载JDK
- 安装JDK

```
# chmod +x jdk-1_5_0_17-linux-i586.bin
# cp jdk-1_5_0_17-linux-i586.bin /opt
# ln -s /opt/jdk-1_5_0_17-linux-i586/bin /opt/java
```
- 设定相关环境变量

```
# echo export JAVA_HOME=/opt/java >> /etc/profile
在/etc/profile中加入或修改:
export PATH=$PATH:$JAVA_HOME/bin
```

安装ANT

- 从 ant.apache.org 下载ANT

```
# wget -c http://archive.apache.org/dist/ant/binaries/apache-ant-1.6.5-bin.tar.gz
```
- 安装ANT

```
# tar xvf apache-ant-1.6.5-bin.tar.gz -C /opt
# ln -s /opt/apache-ant-1.6.5/opt/ant
```
- 设定ANT_HOME环境变量

```
# echo export ANT_HOME=/opt/ant >> /etc/profile
在/etc/profile中加入或修改:
export PATH=$PATH:$ANT_HOME/bin
```

安装Tomcat

- 从 tomcat.apache.org 下载Tomcat

```
# wget -c http://apache.freelamp.com/tomcat/tomcat-5/v5.5.27/bin/apache-tomcat-5.5.27.tar.gz
```
- 安装Tomcat

```
# cd /opt
# tar xvf apache-tomcat-5.5.27.tar.gz -C /opt
# ln -s /opt/apache-tomcat-5.5.27/opt/tomcat
```
- 设定TOMCAT环境变量

```
# echo export CATALINA_HOME=/opt/tomcat >> /etc/profile
```

安装PostgreSQL（源码方式）（1）

- 创建postgres用户组及用户
 - `groupadd postgres`
 - `useradd -g postgres postgres`
- 从 www.postgresql.org 下载PostgreSQL源码包

```
# wget -c http://wwwmaster.postgresql.org/download/mirrors-ftp/source/v8.2.12/postgresql-8.2.12.tar.gz
```
- 安装PostgreSQL

```
# tar xvf postgresql-8.*.gz
# cd post*
# ./configure --prefix=/usr/local/pgsql
# make
# make install
```
- 设定环境变量(可选)
在/etc/profile中加入或修改:

```
export PATH=/usr/local/pgsql/bin:$PATH
```

安装PostgreSQL（源码方式）（2）

- 创建数据区目录

```
# mkdir /usr/local/pgsql/data
# chown postgres:postgres /usr/local/pgsql/data
# chmod 700 /usr/local/pgsql/data
```
- 创建数据库集群

```
# su - postgres
$ cd /usr/local/pgsql/bin
$ Antidb -D /usr/local/pgsql/data
```
- 启动PostgreSQL

```
$ ./pg_ctl -l pg_server_log -D /usr/local/pgsql/data start
```
- 创建dSPACE数据库

```
$ ./psql -c "CREATE USER dspace \
WITH CREATEDB PASSWORD 'xxxxxx'"
$ ./psql -U dspace -c "CREATE DATABASE dspace \
WITH OWNER=dspace ENCODING='UTF8'"
```

安装dspace一文件及目录准备

- 创建dspace安装目录
 - # `mkdir /dspace`
- 下载dspace源码包
 - # `wget -c http://jaist.dl.sourceforge.net/sourceforge/dspace/dspace-1.4.2-source.tgz`
- 安装dspace
 - # `tar xzvf dspace-1.4.2-source.tgz`
 - # `ln -s dspace-1.4.2-source dspace`
 - # `cd dspace`

安装dspace一修改配置文件

- 编辑config/dspace.cfg

```
.....
dspace.dir = [dspace数据文件目录]
dspace.url = [访问dspace的URL网址]
dspace.hostname = [dspace主机名称]
dspace.name = [dspace名称]

db.username = [dspace数据库用户名]
db.password = [dspace数据库用户口令]

mail.server = [发送邮件服务器域名]
mail.server.username = [发送邮件的认证用户]
mail.server.password = [发送邮件认证用户口令]
mail.from.address = [表明邮件来源的地址]
feedback.recipient = [接收反馈信息的邮件地址]
mail.admin = [dspace管理员邮件地址]

Handle.prefix = [handle前缀]
.....
```

安装dspace一编译部署

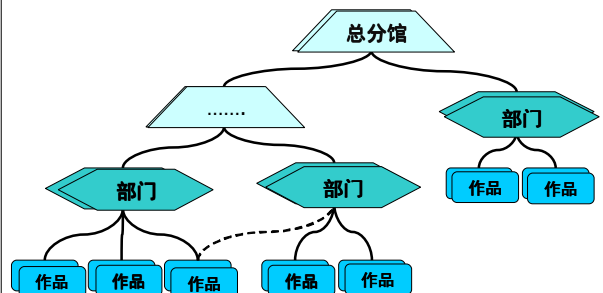
- 编译安装
 - # `ant fresh_install`
- Web部署
 - # `cp build/*.war /opt/tomcat/webapps/`
- 启动Tomcat
 - # `/opt/tomcat/bin/catalina.sh start`
- 访问dspace
 - `http://[dspace_url]:8080/dspace`

安装dspace一基本应用

- 创建管理员用户
 - # `cd /dspace/bin`
 - # `./create-administrator`
- 社群/专题规划创建
 - 访问: `http://[dspace_url]:8080/dspace/dspace-admin`
 - 以dspace管理员登录后, 即可进行社群 (communities) 和专题 (collections) 的创建。
- 提交内容
 - 至提交目标专题
 - 点击专题下的提交

三、DSpace的本地化

信息组织模型（映射）



界面信息的中文化

- 翻译Messages.properties文件
 - 位置: config/language-packs
 - 复制和重命名: Messages.zh_CN.properties
 - 将其中的“键/值 (key/value)”对中的值翻译为中文, 以完成大部分与用户界面有关的中文文化工作。
- 需要中文化处理的其他文件
 - config/input-forms.xml文件: 表单输入
 - config/emails下的所有文件: 各类邮件信息
 - jsp/help下的所有文件: 帮助信息

其他方面的中文支持处理

- 中文全文索引支持(config/dspace.cfg)

```
.....
search.analyzer = org.apache.lucene.analysis.cn.ChineseAnalyzer
.....
```
- Tomcat get请求正确支持中文参数的传递
 - 编辑/opt/tomcat/conf/server.xml

```
.....
<Connector port="8080" ..... URIEncoding="UTF-8"/>
.....
```
- 存在于java文件中的少量提示信息的中文化

WEB界面的调整和优化(1)

- DSpace用户界面主要通过jsp加CSS样式文件的方式实现。修改之前, 按照DSpace的建议, 把需要修改的jsp文件(含目录结构)复制到jsp/local目录下, 以便后续的升级管理
- DSpace为所有的页面提供了统一的布局模式:
 - 头部区: /jsp/layout/header-default.jsp
 - 底部区: /jsp/layout/footer-default.jsp
 - 方位栏: /jsp/layout/location-bar.jsp
 - 导航栏: /jsp/layout/navbar-default.jsp
- 修改以上jsp文件可改变所有使用默认布局的页面的效果
- 页面显示风格统一由CSS样式文件:
 - jsp/styles.css.jsp定义和控制。

WEB界面的调整和优化(2)

- 改变某一个页面的布局效果:
 - 根据需求设计和定义header-XXX.jsp、footer-XXX.jsp或navbar-XXX.jsp等形式命名的页面文件, 存放在jsp/local/layout目录下
 - 在打算使用自定义页面布局文件的jsp文件中通过<dspace:layout>的相关属性来指定使用自定义的页面布局文件。如:
 - <dspace:layout style="myStyle" navbar="myNavbar".....>
 - 对应的jsp文件将通过包含和使用“header-myStyle.jsp”、“footer-myStyle.jsp”、和“navbar-myNavbar.jsp”文件来构造和显示相应的页面。
 - 方位栏(location bar)由于不包含实质性的可定制内容, 只提供了一些禁用(off)、是否支持链接(link/nolink)等选项。
 - 页面显示风格的定制则通过修改jsp/styles.css.jsp文件实现, 即根据需要定义特定的CSS类并在相应页面的jsp文件中进行调用即可。

元数据格式调整

- DSpace采用QDC元数据来支持对Item存档对象进行描述和组织
- 对于Community和Collection存档集合对象的也提供一些包含简单描述性信息的元数据。
- 也提供部分有关保藏、来源、权限政策等管理性元数据的支持
- 如需增加或调整元数据元素(编译部署前), 但增加的元素要在编辑或显示过程中使用, 需要修改相关程序文件
 - 修改: config/registries/dublin-core-types.xml

```
.....
<dc-type>
  <schema>dc</schema>
  <element>contributor</element>
  <qualifier>author</qualifier>
  <scope_note></scope_note>
</dc-type>
.....
```

编辑模板的定制(1)

- 元数据编辑模板的定制主要通过修改config/input-forms.xml来实现
- 是以Collection为基础的元数据编辑模板定制机制, 即需要事先已创建好collection
- input-forms.xml的结构

```
<form-map>
  <name-map collection-handle="default" form-name="traditional"/>
  <name-map collection-handle="22392" form-name="periodical"/>
  .....
</form-map>
```

 - 标签<form-map>定义并建立collection与form(即元数据数据编辑模板)之间的映射关系。
 - 为在映射表中出现的collection自动映射到默认的“traditional”模板。
 - 映射关系的定义中, 使用的是collection handle而不是collection的名称。

编辑模板的定制 (2)

```
<form-definitions>
  <form name="periodical">
    <page number="1">
      <field>
        <dc-element>Identifier</dc-element>
        <dc-qualifier></dc-qualifier>
        <repeatable>true</repeatable>
        <label>标识符</label>
        <input-type value-pairs-name="common_identifiers">qualdrop_value</input-type>
        <hint>输入条目的标识符</hint>
        <required>标识符的值不能为空。</required>
      </field>
      .....
    </page>
    .....
  </form>
  .....
</form-definitions>
.表单定义
```

编辑模板的定制 (3)

```
<form-value-pairs>
  <value-pairs value-pairs-name="common_identifiers" dc-term="identifier">
    <pair>
      <displayed-value>URI(统一资源标识符)</displayed-value>
      <stored-value>uri</stored-value>
    </pair>
    <pair>
      <displayed-value>ISBN</displayed-value>
      <stored-value>isbn</stored-value>
    </pair>
    .....
  </value-pairs>
  .....
</form-value-pairs>
```

- 为dropdown、qualdrop_value类型的输入框定义和安装下拉列表。

定时任务的设置(1)

```
;/var/spool/cron/root
.....
0 1 * * * (SCATALINA_HOME/bin/shutdown.sh >/dev/null) && \
  (cp -f $CATALINA_HOME/logs/catalina.out |
  $CATALINA_HOME/logs/catalina.out.yest) && \
  (cat /dev/null >$CATALINA_HOME/logs/catalina.out) && \
  ($CATALINA_HOME/bin/startup.sh >/dev/null)
10 1 * * * $DSPACE_DIR/bin/sub-daily > $DSPACE_DIR/sub-daily.log 2>&1
30 1 * * * $DSPACE_DIR/bin/filter-media > $DSPACE_DIR/log/filter-media.log
2>&1"
0 2 * * * $DSPACE_DIR/bin/stat-monthly
10 3 * * * $DSPACE_DIR/bin/stat-general
20 3 * * * $DSPACE_DIR/bin/stat-report-monthly
30 3 * * * $DSPACE_DIR/bin/stat-report-general
```

定时任务的设置(2)

```
;/var/spool/cron/postgres
0 0 * * * vacuumdb -analyze dspace > /dev/null 2>&1
10 0 * * * pg_dumpall > /backup/dspace.backup
0 0 * * * tar czvf /backup/dspace.tar.gz /dSPACE
30 0 * * * (cp -f $PGDATA/pg_server_log |
  $PGDATA/pg_server_log.yest) && \
  (cat /dev/null >$PGDATA/pg_server_log)
```

其他

- tomcat、postgres服务自启动
- 普通用户身份运行tomcat及dSPACE服务：
 - jsvc、apache+tomcat、iptables等方式
 - iptables方式 (RedHat)：

```
# /sbin/iptables -t nat -A PREROUTING -p tcp --dport 80 -j REDIRECT --to-port 8080
# /sbin/iptables -t nat -A OUTPUT -p tcp -d $server_ip_address --dport 80 -j REDIRECT --to-port 8080
# /sbin/iptables-save > /etc/sysconfig/iptables
```

四、DSpace在Windows环境下的安装

软件准备

- 基础软件 (Windows版)
 - JDK 1.4+
 - Tomcat 5.0+
 - ANT 1.6+
 - PostgreSQL 8.0+
- 取得DSpace源码
 - DSpace 1.4.2为例

安装过程 (1)

- 安装JDK
 - 执行jdk安装程序
 - 设定系统环境变量
 - JAVA_HOME=C:\Program Files\Java\jdk1.5.0_17
 - CLASSPATH=%JAVA_HOME%\lib\dt.jar;%JAVA_HOME%\lib\tools.jar
 - PATH=.....;%JAVA_HOME%\bin
- 安装ANT
 - 解压ANT安装包至:C:\Program Files下
 - 设定环境变量
 - ANT_HOME=
 - PATH=.....;apache-ant-1.6.5

安装过程 (2)

- 安装Tomcat
 - 执行Tomcat安装程序, 按照提示安装
 - 修改conf/server.xml, 加入:
`<Connector port="8080" URIEncoding="UTF-8"/>`
 - 设定系统环境变量
 - CATALINA_HOME=CATALINA_HOME=C:\Program Files\Apache Software Foundation\Tomcat 5.5
- 安装PostgreSQL
 - 执行安装程序, 按提示安装

安装过程 (3)

- 安装DSpace
 - 创建dSPACE用户和dSPACE数据库
 - 通过PostgreSQL安装过程中安装的PgAdmin III来进行
 - 创建数据库用户dSPACE, 并使之具有创建数据库的权限
 - 创建数据库dSPACE, 使之成为dSPACE用户所拥有
 - 修改config/dSPACE.cfg
 - dSPACE.dir = [dSPACE数据文件目录, 比如 d:\dSPACE]*
 - dSPACE.url = [访问dSPACE的URL网址]*
 - dSPACE.hostname = [dSPACE主机名称]*
 - dSPACE.name = [dSPACE名称]*

 - db.username = [dSPACE数据库用户名]*
 - db.password = [dSPACE数据库用户口令]*

 - mail.server = [发送邮件服务器域名]*
 - mail.server.username = [发送邮件的认证用户]*
 - mail.server.password = [发送邮件的认证用户口令]*
 - mail.from.address = [表明邮件来源的地址]*
 - feedback.recipient = [接收反馈信息的邮件地址]*
 - mail.admin = [dSPACE管理员邮件地址]*

 - Handle.prefix = [handle前缀]*

安装过程 (4)

- 创建dSPACE安装目录: 比如d:\dSPACE
- 编译部署
 - 执行: ant fresh_install
 - 部署: 把编译后在build目录下生成的war文件拷贝至Tomcat的webapps目录下
 - 创建管理员用户:
 - 如在d:\dSPACE\bin下执行
 - *dsrun org.dSPACE.administer.CreateAdministrator*
- 启动Tomcat
- 访问: [http://\[dSPACE_url\]:8080/dSPACE](http://[dSPACE_url]:8080/dSPACE)

谢谢!