netarchive.dk

Setup of database, JMS broker, and FTP server

Søren V. Carlsen (svc@kb.dk)

Database setup (1)

- NetarchiveSuite can be used with three kinds of databases
  - External Derby
    - *dk.netarkivet.harvester.datamodel.DerbyClientSpecifics*
  - Embedded Derby (default)
    - *dk.netarkivet.harvester.datamodel.DerbyEmbeddedSpecifics*
  - MySQL (5.0.27+)
    - *dk.netarkivet.harvester.datamodel.MySQLSpecifics*

Database setup (2)

- Backup of the NetarchiveSuite database is only done automatically each 24 hours, if Embedded Derby is chosen. Backup is initiated at the *backupInitHour*, and dumps the database to a directory named DB-Backup-<timestamp>
- If other database is chose, you (or your) sysop must do the backup yourself

Configuring the database URL

- **The url setting (settings.harvester.datamodel.database.url)**
  - **Default URL: "jdbc:derby:harvestdefinitionbasedir/fullhddb" (Assumes, that the harvestdefinitionbasedir/fullhddb.jar has been unzipped)**
  - **Alternate urls:**
    - **(for external Derby) jdbc:derby://<hostname>:<port>/fullhddb**
    - **(for MySQL) jdbc:mysql://<hostname>/fullhddb?user=userName&password=thePassword**

Using embedded Derby as Database

- To use embedded Derby as database
  - cd <installation-dir>/harvestdefinitionbasedir
  - unzip fullhddb.jar
  - Done!

Using external Derby as Database

- Requirements
  - derbynet-X.jar in the classpath (jar file not yet part of distribution; but it is in our svn repository)
- Starting the server
  - Go to databasedir (where fullhddb.jar is unfolded)
  - java org.apache.derby.drda.NetworkServerControl start -p <port>
- Stopping the server
  - Go to databasedir (where fullhddb.jar is unfolded)
  - java org.apache.derby.drda.NetworkServerControl stop -p <port>

Using MySQL as database(1)

- Requirements
  - Download a mysql-connector-java-5.0.X-bin.jar from http://dev.mysql.com/downloads/connector/j/5.0.html
  - Add mysql-connector-java-5.0.X-bin.jar to classpath of the applications accessing the database (HarvestDefinitionApplication, HarvestTemplateApplication)

Using MySQL as database(2)

- Create the initial database
  - cd scripts/sql/
  - mysqladmin create fullhddb
  - mysql fullhddb < createfullhddb.mysql
- Insert the default harvest template
  - cd NetarchiveSuite installation dir.
  - java HarvestTemplateApplication create default_orderxml harvestdefinitionbasedir/order_templates/ default_orderxml.xml
  - Shell-script to do this

Configuring the JMS-broker

- Requirements
  - Sun MQ 4.1 installed(downloaded from https://mq.dev.java.net/downloads.html (Open Message Queue binaries)
  - Installed in $IMQ_HOME
- Setting maxNumActiveConsumers to 20
  - Append line "imq.autocreate.queue.maxNumActiveConsumers=20" to $IMQ_HOME/var/instances/imqbroker/props/config.properties
- Changing JMS listening port (default 7676) – cf. Setting settings.common.jms.port
  - Edit line "imq.portmapper.port=7676" in $IMQ_HOME/lib/props/broker/default.properties
- More Info: Installation Manual, appendix A

Selecting a RemoteFile plugin

- Larger lumps of data (e.g. > 10 mb) are transmitted between applications using RemoteFiles.
- We can choose between
  - HTTPRemoteFile
    - RemoteFile plugin needed:
      - Class: dk.netarkivet.common.distribute.HTTPRemoteFile
      - Needs additional settings for: port
    - uses http as transport protocol
    - Requires designated HTTP ports for all NetarchiveSuite applications sending or receiveing RemoteFiles.
  - FTPRemoteFile
    - uses FTP as transport protocol
    - requires the installation of one or more FTP servers, normally one per location
    - Remotefile plugin needed:
      - Class: dk.netarkivet.common.distribute.FTPRemoteFile
      - Needs additional settings for: serverName, serverPort, userName, userPassword, retries

Configuring out FTP-servers

- Choosing FTPRemoteFile as our RemoteFile plugin requires that the FTPservers allow
  - Existing files in the server to be overwritten to allow append-operations. The proftpd setting for this is ”AllowOverWrite on”
  - That upload to FTPserver can be restarted. The proftpd setting for this is ”AllowStoreRestart on”.

It may also be necessary to alter the number of concurrent connections to the FTPserver.The proftpd setting for this is ”MaxInstances <number of concurrent instances>”.

More information in Installation Manual, appendix A