Appendix C - Easy Installation of NetarchiveSuite

Contents

- Verify that you have all the needed software installed by installing the !QuickStart according to http://netarchivedk/suite/Quick_Start_Manual_316 eg in /home/test/netarchive by starting the Quickstart.

Below, you find other deploy examples (They have to be modfied to your environment).

- You can now create, run and browse according to the !QuickStart - or User Manual.

Examples of deploy configuration files

The following example of configuration file requires adaptation to your own system before use.

[deploy_distributed_example.xml](#)

The instance with two replicas divided over two physical locations Each physical locations contain several machines Bitarchive machines, harvester machine and viewerproxy machine Only one physical location has an administator machine, which contains the GUI application, the Bitarchive monitors, the HarvestJobManager, HarvestJobMonitor and the arc repository.

---

h3 How to add a harvester more on the same machine and set all to HIGHPRIORITY selective harvesting
Using eg deploy_examplexml

- Duplicate the existing harvester <applicationName> definition within <deployMachine>.

In the new duplicate harvester config, change all following duplicate values to new unique values within <deployMachine>:

- <applicationInstanceId>
- <common><jmx><port> and <rmiPort>
- <heritrix><guiport> and <jmxPort>
- <serverDir>harvester_high_2</serverDir>

and set

- <queuePriority>HIGHPRIORITY</queuePriority>

eg:

```
<applicationName name="dknetarkivetharvesterharvestingHarvestControllerApplication">
 <settings>
  <common>
   <applicationInstanceId>high2</applicationInstanceId>
   <jmx>
    <port>8112</port>
    <rmiPort>8212</rmiPort>
   </jmx>
  </common>
   <harvester>
    <harvesting>
     <queuePriority>HIGHPRIORITY</queuePriority> <heritrix>
      <guiPort>8192</guiPort> <!-- T: jmxPort to be modified by test (was 8093) --> <jmxPort>8193</jmxPort>
       <jmxUsername>controlRole</jmxUsername>
       <jmxPassword>R_D</jmxPassword>
     </heritrix>
     <serverDir>harvester_high_2</serverDir>
    </harvesting>
   </harvester>
  </settings>
 </applicationName>
```

How to configure which Heritrix report has to be uploaded in the metadata ARC file

Three settings properties control which heritrix reports are added to the metadata ARC file:

- **settingsharvesterharvestingmetadataheritrixFilePattern** is a java pattern that allows you select which files in the crawl dir (not recursively) to include in the metadata ARC.

- **settingsharvesterharvestingmetadatareportFilePattern** is also a java pattern that controls which subset of the files selected by heritrixFilePattern are to be considered as report files All the other files will be considered as setup files.

- **settingsharvesterharvestingmetadatalogFilePattern** is a third java pattern that controls which files in the logs subdirectory of the crawldir are to be added as log files to the metadata ARC.