

Installation Overview

- [Contents](#)
 - [Audience](#)
 - [Limitations](#)
- [Installation Overview](#)

Contents

The first part describes the functionality of the deploy software and how it can be used. This involves a description of how to run this module, the required and optional arguments, and the functionality of the scripts generated.

The second part describes the configuration file used by the deploy software, both in structure, content and examples. This also describes the requirements and limitations of Deploy.

The third part describes the different possible installation scenarios.

The fourth part describes the means of deployment, which includes description of how to obtain and install required libraries, how to install the software on separate machines.

Finally, the starting, stopping and monitoring of the system is described. This part is useful for those who want to go beyond the limitations inherent in the deploy software.

Some parts of NetarchiveSuite require external software to run. This software is described in appendix A.

This manual does not explain the configuration of the applications themselves (see the [Configuration Manual](#) for this), how to extend the functionality of the system (see the development project for this), or how to use the running system (see the [User Manual](#) for this).

Audience

The intended audience of this manual is system administrators who will be responsible for the actual installation of NetarchiveSuite, as well as technical personnel responsible for proper operation of NetarchiveSuite. Knowledge of Unix system administration is required, and some familiarity with XML and Java is an advantage.

Limitations

Even though the NetarchiveSuite software is developed in Java, and therefore is mostly platform independent, we do have a couple of external calls to the Unix *sort* command. The parts of our software using this external command therefore only run on Linux/Unix, or Windows with Cygwin installed. The parts in question are:

- The `dk.netarkivet.common.webinterface.GUIApplication`, if the `sitesection dk.netarkivet.viewerproxy.webinterface.QASiteSection` is used
- The `dk.netarkivet.harvester.indexserver.IndexServerApplication`

Specifically the following methods all use an external call to the Unix `sort()` command:

- `FileUtils#sortCrawlLog`
 - Used in
 - `dk.netarkivet.harvester.indexserver.CrawlLogIndexCache`,
 - `dk.netarkivet.viewerproxy.webinterface.Reporting`
- `FileUtils#sortCDX()` (only used in `dk.netarkivet.harvester.indexserver.CrawlLogIndexCache`)
- `dk.netarkivet.harvester.indexserver.CDXIndexCache#sortFile()`
- `dk.netarkivet.viewerproxy.LocalCDXCache#getIndex()`

The only part of NetarchiveSuite to have been tested under Windows is the `BitarchiveApplication`. It is therefore highly recommended that all other applications are used only in Linux environments.

Installation Overview

Using NetarchiveSuite's Deploy utility, the steps required to configure and start a webarchive are

1. Determine the target architecture - ie how many machines you will be using, their locations, their operating systems and which applications should run on each machine.
2. Configure the required machines, the required external software (see Appendices) and any relevant firewalls.
3. Unpack NetarchiveSuite.zip in a directory on a linux machine.
4. Create the `config.xml` file which describes the architecture and any custom settings. This will also specify your `environmentName` (e.g. `MY_WEBARCHIVE`).
5. Modify the other configuration files (logging and security properties) if necessary.

6. Run the Deploy utility. This will create a sub-directory MY_WEBARCHIVE with all the deploy scripts and configuration files you need.
7. Run the install scripts, then the start scripts. You should now have a running netarchivesuite installation.

