The multi-plateform data storage system

VIP Project - ANR-09-COSI-03 - Milestone 2.2.2

Baptiste Grenier <bgrenier@maatg.com> - maatG
Jérôme Revillard <jrevillard@maatg.com> - maatG

Version 1.1, January 11, 2011

Revision table	
V1.0	Initial release
V1.1	Update of some command lines arguments

1. Overview

The goal of the data storage system is to ease an automatic management of a local DPM used as a cache.

The data storage system inspects recursively a directory of the local DPM to ensure that a policy on files' replicas quantity is applied.

The policy consist of enforcing a number of replica for a file:
- if a file doesn't have enough replicas some will be created until the targeted replica count (minimum-replicas option) is reached ;
- if the limit-replicas-number option is set then the demon will delete the exceeding replicas.
- if the cleanup-threshold value is set and is lower than the space used the application will try to free some space by deleting already well replicated replicas..

The application will analyze recursively a directory of the local DPM, and for each file it will check the number of replicas (by asking the LFC) and then it will apply the policy.

The application will use the DPM certificate/key to access the local files, and then it will ask MyProxy for a proxy of the file's owner. This proxy will allow the data management daemon to interact with the grid on behalf of the file's owner.

If a file does not have any reference in the LFC(s) (ie is orphan) it will be deleted from the local DPM.

Detected errors (file not accessible, no more proxy for this user...) will be logged and sent in a summary email at the end of the check.

2. Policies

The current daemon policies are defined by:
- the minimum-replicas number which define the targeted replica number (with a hard limit of one replicas per SE)
- the limit-replicas-number flag which will make the daemon delete replicas over `the minimum-replicas number`

3. Packaging

The daemon can be provided as a tarball or as a RedHat x86_64 RPM.

The RPM version installs the files like this:
- the dependencies are located in the /usr/lib/data-management-daemon directory ;
- the configuration files are located in /etc/data-management-daemon ;
- the init scipt is deployed as /etc/init.d/data-management-daemon ;
- the documentation is deployed as /usr/share/doc/data-management-daemon ;
- the binaries are deployed in the /usr/bin directory.

4. Using the daemon

The application can be launched manually, as a standard CLI or like a unix daemon.

The application reads /etc/data-management-daemon/dpmwatcher.properties, be sure to adjust it.
The mail server configuration is done in /etc/data-management-daemon/mailer.properties.
The logging configuration is done in /etc/data-management-daemon/log4j.properties.

1. CLI arguments

```
-h -help: help
--debug: show debugging output
-v --verbose: verbose output
--version: print version information and exit
-H --dpm-host: hostname of the local DPM
-m --minimum-replicas: minimum number of replicas
-p --monitored-path: path to monitor
-t --cleanup-threshold: cache cleanup threshold
-l --limit-replicas-number: limit replicas number (to minimum-replicas)
```

2) Daemon arguments

```
-H --dpm-host: hostname of the local DPM
-m --minimum-replicas: minimum number of replicas
-p --monitored-path: path to monitor
-t --cleanup-threshold: cache cleanup threshold
-l --limit-replicas-number: limit replicas number (to minimum-replicas)
```

The init script reads the daemon configuration in the /etc/data-management-daemon/daemon.config file and uses these configuration options as arguments to call the daemon.

5. Requirements

- a running JVM (>= 1.5)
- should run on a DPM server
- access to the the DPM's certificate and key
- the DPM user certificate should be allowed by MyProxy to access all the proxies.

The daemon uses a gLite DM OpenSource Java library for interacting with the DPM's DPNS and with the LFC. It uses also the JSAGA connector when it's possible.
This library, developed by maatG in the context of the VIP project, could be found at http://project-mgt.maatg.fr/projects/glite-data-management.

.