EGI-InSPIRE

TOOLS FOR LOCATING FILES IN THE LFC RELATED TO A USER OR A SE

(TNA3.4)

Document identifier:	SE Decommission.doc
Date:	**10/8/2011**
Activity:	**NA3**
Lead Partner:	**EGI.eu**
Document Status:	**DRAFT**
Dissemination Level:	**PUBLIC**
Document Link:	https://documents.egi.eu/document/

Abstract

The aim of this document is to provide information about a tool for easing the process of locating the files in a LFC that are related to a specific SE in order to obtain the rightmost information to warn affected users or to execute automatically replication actions.

I. COPYRIGHT NOTICE

II. DELIVERY SLIP

	Name	Partner/Activity	Date
From	I. Blanquer, G. Borges	UPV, LIP/NA3	10/8/2011

III. DOCUMENT LOG

Issue	Date	Comment	Author/Partner
1	11/10/2010	First version	I. Blanquer / UPV

IV. APPLICATION AREA

This document is a public internal report of the EGI-InSPIRE project.

V. DOCUMENT AMENDMENT PROCEDURE

Amendments, comments and suggestions should be sent to the authors. The procedures documented in the EGI-InSPIRE "Document Management Procedure" will be followed: https://wiki.egi.eu/wiki/Procedures

VI. TERMINOLOGY

A complete project glossary is provided at the following page: http://www.egi.eu/about/glossary/.

1 DESCRIPTION OF THE PROBLEM

The selective reallocation of specific files is a complex task that VO managers have to deal when a SE becomes full or a SE needs to be decommissioned. Files registered in the catalogue stored in the affected SE need to be moved or replicated in a different SE. To do so, VO managers have to identify:

1. The files located in the affected SE. this involves the SURLs but especially the LFNs, which are often the only meaningful identifiers.
2. The owners of such files and their contact details.

It will be important to have a tool that could create reports of LFNs stored in a specific SE, registered in a specific LFC and belonging to a specific VO. Those reports should include, along with the information of the LFNs, the DNs of the users owning files and the size of the files. Those reports could be either exhaustive or comprehensive.

In an ideal scenario, it could be used to automatically warn users and send reminders and to replicate affected files in a different SE. The replication logic could be different for the LFNs of files that are stored only on the affected SE or the LFNs of files that have additional replicas somewhere else.

2 Tools available

2.1 LFC Commands

LFC commands provide a simple Command Line Interface to the data stored in a LFC. LFC commands can query the catalogue and extract the information about the LFN, SE, owner and other metadata. However, most of the LFC commands work on a single interaction basis, so time required for processing a large LFC could be intractable.

By using the "`lfc-ls -lR`" command provides a recursive list of LFNs, but replicas, which reflect the location of a file in a SE, are obtained through the `lcg-lr` command. Therefore, the location of the affected files is is normally performed by contacting the LFC site manager, who queries directly the database.

The identification of the owner of a file may be obtained through the output of the first command (`lfc-ls -lR`) by filtering the third column (uid). Then DNs can be obtained from uids through

command "`lfc-listusrmap –uid`".

Replications can be performed directly using the "`lcg-rep –d dest_host file`". Again, this is a lengthy process.

2.2 LFC APIs

LFC also exposes its functionality through a set of APIs that provide additional features. These APIs can be used to reduce processing time in some of the operations listed above.

- `lfc_listreplicax`. This API enables obtaining the list of all the SURLs stored in a given SE and registered in the LFC. It does not differentiate among different VOs that could be hosted in the same LFC. This function provides SURLs and unique file ids (common to different replicas of a file, but in a different format as the GUID) of the files.
- `lfc_getgrpbynam`. This API provides the identifier of a given VO in the metadata table of the LFC. This could be used to later on filter the SURLs.
- `lfc_statr`. This API provides the gid (VO identifier), uid (Owner identifier) and files size, among other information, for each SURL.
- `lfc_getpath`. This API provides the registration LFN (not the alias) for a given replica.
- `lfc_getusrbyuid`. This API provides the DN of a user identified through a LFC uid.
- `lfc_addreplica`. This API enables replicating a file in a different SE by providing the guid or a file id structure that contains the unique LFN file id and the SE where the file is stored.

3 Using LFCBrowse

A tool has been developed to ease this process. This tool can be downloaded from http://wiki.healthgrid.org/Biomed-Shifts:Index#Monitoring_tools and enables locating the affected files and users by means of the LFC API.

3.1 Usage Syntax

The tool is Command Line-based and enables listing different kinds of information about the files stored in a SE. The syntax comprises one compulsory filed (the affected SE) and eight optional arguments.

```
usage syntax: ./LFCBrowseSE <SE> [--vo <VO>] [--sfn] [--lfn] [--dn] [--size] [--guid] [--summary][--rep
        <DstSE>]
 SE: affected SE from files will be listed
 --vo VO: optional, only list files belonging to the specified VO
 --sfn: optional, lists all the SURLs of the files affected.
 --lfn: optional, lists all the LFNs of the files affected.
 --dn: optional, displays the DNs of the owners of all the files affected.
 --size: optional, displays the size of the files affected.
 --guid: optional, lists the GUIDs of the files affected.
 --summary: optional, displays a summary of the different users with files affected and the total
        storage space occupied
 --quiet: optional, it does not display the progress
--rep <DstSE>: optional, replicates matching instances on the specified SE
```

The “vo” option enables filtering by VO, listing only the files that belong to a specific VO. This option is necessary when dealing with LFC servers that store files from more than one VO. The `lfc_listreplicax` API lists all the files belonging to a specific SE, regardless the VO of the credentials of the user executing the command.

The options `--sfn`, `--lfn`, `--dn`, --size and `--guid` provide additional information about each if the entries. Since the information is obtained from different invocations of the APIs, the performance is not the same if only the SURLs are requested (very quick) or other information such as LFNs or DNs are specified. Since the process could be lengthy, progress is shown if the `--quiet` option is not indicated. The information can be aggregated in a summary that lists all the space consumed in the SE by each one of the users involved.

Finally, option --rep enables specifying a destination SE for making the replicas of the files.

4 Additional Requirements

The main drawback for the performance of this tool is the need for multiple interactions for obtaining the metadata of each entry. If both the information of the `lfc_filereplica` structure provided by `lfc_listreplicax` the information of the structure `lfc_filestatg` provided by `lfc_lstatr` and the path provided by lfc_getpath for the same fileid could be provided, the performance would improve very much, strongly reducing the traffic handled by the LFC server. This feature has been requested in the requirement https://rt.egi.eu/rt/Ticket/Display.html?id=2712.