AliEn File Access Monitoring Service – FAMoS

Armenuhi Abramyan[1,2,3] **Narine Manukyan**[1,2,3]

[1] *A.I. Alikhanyan National Science Laboratory (Yerevan Physics Institute) Foundation, 2 Alikhanyan Brothers St., 0036, Yerevan, Armenia*
[2] *Armenian e-Science Foundation, 49 Komitas St., 375051 Yerevan, Armenia*
[3] *Department of Computer Systems and Informatics, State Engineering University of Armenia, 105 Teryan St., 0009 Yerevan, Armenia*
aabramya, nmanukya @mail.yerphi.am

Abstract

Hundreds of thousands of large volume data files are produced and simultaneously replicated in the processes of the data analysis and simulations performed in the Virtual Organization (VO) of CERN's ALICE experiment. As a result, the storage volume of Storage Elements of the ALICE VO sites get quickly replete, preventing further inflow of files to these sites.

To handle with this problem, a special service that would provide removal of the files from the Storage Elements on an organized and regular basis has to be developed.

The service should consist of three parts: i) Detailed monitoring of the data on the usage of files, ii) Determination of the order of the removal of files and iii) Development of the algorithm of the freeing of storage volumes on Storage Elements.

In this article, we describe our solution of the first part.

Keywords

File Popularity, Monitoring, Storage Element, Grid, AliEn, ALICE, VO, LHC, CERN.

1. Introduction

The aim of ALICE experiment [1], the one of 4 biggest LHC [2] experiments at CERN [3], is to explore the primordial state of matter that existed in the first instants of our Universe, immediately after the initial hot Big Bang. The ALICE detector has been built by a collaboration, which currently includes 1300 physicists and engineers from 120 Institutes in 35 countries. At its full operation, the ALICE experiment produces tens of Petabytes (10^{15} Bytes) of data annually, which thousands of scientists around the world need to access and analyze in the ALICE Virtual Organization (VO), built on the AliEn, ALICE Grid Environment [4].

As is required in Grids, the files created in the processes of the data analysis and simulations in ALICE are replicated and stored in Storage Elements (SE) of the AliEn high-performance computing sites (in total 70 sites) distributed all over the world. Since the storage capacities of the sites are limited, this leads to the repletion of the whole storage volume of ALICE VO in a time interval of the order of two months, blocking, thus, the inflow of newly produced files.

The solution of this problem via removal of certain portion of replicas looks as a three-step process:

1. Specification of the relevant characteristics (attributes) of the accesses to the files and continuous monitoring and accumulation of their values;
2. Construction of the algorithms, determining the order of the removal of files and/or their sets and the amount of files to be removed;
3. Implementation of the storage volume freeing agent-daemons working on the ALICE VO sites and removing the files on the base of the analysis of information gathered at the steps 1 and 2.

This paper describes our work on the 1st step. The File System ALICE VO as well as the types of file and file sets, subjects to replication, are described in Sections 2. A detailed description of developed ***File Access Monitoring Service*** (***FAMoS***) is given in Section 3.

2. File System and Data file and file set types of ALICE VO

2.1 File System of ALICE VO

ALICE distributed data storage system glues multiple types of storage systems and data transport protocols into one virtual file system called *File Catalogue (FC)* [5].

The AliEn *FC* Service manages petabytes of experimental data for over 60 disk Storage Elements (SEs) and for 10 tape SEs. At the time of writing this paper (19 June 2013) the total number of files stored by ALICE experiment in its SEs was 287580000 and the total size of these files was 19.57 PB.

Unlike real file systems, the *FC* does not own the files; it only keeps an association between the Logical File Name (*LFN*) and (possibly more than one) Physical File Names (*PFN*) on a file or mass storage system. Below we give an example of the *LFN* and *PFN* of a file named "testfile":

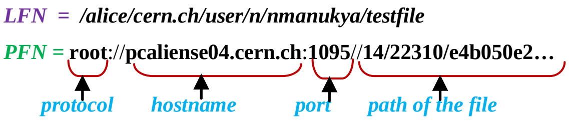
LFN = /alice/cern.ch/user/n/nmanukya/testfile
PFN = root://pcaliense04.cern.ch:1095//14/22310/e4b050e2…
protocol
hostname
port
path of the file

Figure1.The structure of *LFN* and *PFN*

In this example, the *LFN* is the address of the file in the AliEn *FC.* The *PFN* string contains the following fields:

- *protocol* to access the file remotely (e.g. gridFTP, root, rfio);

- *hostname* of the SE where the file is physically stored;
- *port* which is used by the *protocol* for the communication with SE;
- *path* shows the address of the file on the SE filesystem.

2.2 Data file and file set types in ALICE VO

The following data files are created and replicated in the processes of the ALICE experiment data of analysis and Monte Carlo simulations:

ESD (Event Summary Data) files contain detailed information about numerous characteristics of the registered events (trajectories of produced particles, interaction vertices, etc.). The ESDs are obtained in the processes of the event reconstruction from the raw data.

To speed up the data analysis work performed by users, the *ESD* files are filtered to smaller, AOD (Analysis Object Data) files.

In ALICE VO, each ESD and AOD is created in two copies stored on different SEs (in certain cases one creates not one but two replicas of ESDs and AODs). It is important to note that the replication is not performed file-by-file, but in the sets of ESD and AOD called *LHC Period*. The data contained in all ESDs and AODs of each of *LHC Periods* correspond to the investigation of a specific scientific problem.

3. Architecture of FAMoS

The ***FAMoS*** provides a facility to monitor the characteristics/attributes of the accesses to the files and to record in an organized manner the values of attributes to a special database (DB), called ***Accesses***.

The ***FAMoS*** includes the following components:

- ***Authen_ops*** **-** log file. Since all the accesses to the files within AliEn *FC* are authenticated by the AliEn *Authentication* (*Authen*) service [6], a plugin (called ***attributes***) has been developed and included in the *Authen* service in order to record the values of specified attributes (Section 3.1).
- ***Parser*** **–** Perl module, which reads the values of the attributes from *Authen_ops* log file and inserts them into ***Accesses*** DB;
- ***Accesses*** – MySQL database, consisting of 11 tables. It keeps the data for each access to each of files as well as summary data on the accesses to the file sets within one hour and one day time intervals (see Section 3.2). The summation of data for these intervals is provided by the following two Perl modules:
- ***HourlyCollector*** and ***DailyCollector*** - The details on the work of these modules are given in Section 4.

3.1 The attributes of file accesses

The values of the following attributes have to be monitored in order to have a detailed picture of usage of files in ALICE VO.

Attribute	Description
File name	The *LFN* of a file
SE name	Name of SE from where file was accessed
User name	Name of user by whom file was accessed
Access Time	Time and date when the file was accessed
Operation type	Read or write access
Operation result	Successful or failed access

Table 1. The list of attributes of the file accesses

3.2 The Accesses database

The values of the monitored attributes of the accesses to the files are accumulated in the ***Accesses*** DB, whose structure is presented in Figure 2.

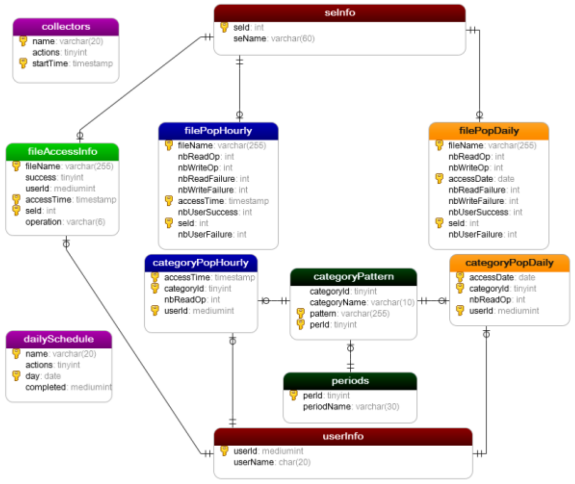

Figure 2. Structure of the A*ccesses* DB

The ***Acesses*** DB contains the following 11 tables:

- **collectors** and **dailySchedule** – to keep the information (start time, busyness state) used for the scheduling of the operations of ***FAMoS*** Perl modules (See Section 4);
- **userInfo** and **seInfo** - to keep the values of User name and SE name attributes;
- **cattegoryPattern** and **periods** – to keep the Period name, data types and other details of file sets (See Section 2.2);
- **fileAccessInfo** - serves as a buffer for the temporal keeping of the information obtained from *Authen_ops* log files;
- **filePopHourly** and **categoryPopHourly** – for permanent keeping of the information on the accesses to the files and file sets for 1 hour time interval;
- **filePopDaily** and **categoryPopDaily** – for permanent keeping of the information on the accesses to the files and file sets for 24 hour time interval.

4. How does the FAMoS work?

The dynamics of ***FAMoS*** is provided by its Perl modules:

Parser – starts every 1 minute and does the following:

- gets the *startTime* of its task (if any) from **collectors** table;
- parses the data for the [*startTime*, *startTime+1hour*] time interval from the *Authen_ops* log files;
- inserts the parsed information into **fileAccessInfo** table;
- inserts a task for ***HourlyCollector*** module into **collectors** table;

- updates the value of *startTime* (by setting *startTime+1hour*) of ***Parser*** task of the **collectors** table.

HourlyCollector **-** starts every 2 minutes and does the following:

- gets the *startTime* of its tasks (if any) from **collectors** table;
- aggregates the data from **fileAccessInfo** table for the [*startTime, startTime+1hour*] interval;
- summarizes the number of accesses to the files by the attributes of these accesses and inserts the summarized data into **filePopHourly** table;
- extracts the *Period* name from the name of file (*LFN*) and if the *Period* name does not exist in the **periods** table, then inserts the *Period* name into the **periods** table;
- defines the *regular expressions* for AOD and ESD sets of files of that *Period* and inserts these *regular expressions* into the **cattegoryPattern** table;
- groups the files into sets by using these *regular expressions*;
- summarizes the number of accesses to the file sets by the attributes of these accesses and inserts the summarized data into **categoryPopHourly** table;
- deletes the data for the [*startTime, startTime+1hour*] interval from **fileAccessInfo** table;
- inserts a task for ***DailyCollector*** module into **dailySchedule** table;
- deletes its task from **collectors** table;

DailyCollector – starts every 30 minutes and does the following:

- gets the *startDate* of its tasks (if any) from **dailySchedule** table;
- aggregates the data from **filePopHourly** and **categoryPopHourly** tables for the *startDate*;
- summarizes the number of accesses to the files and file sets by the attributes of these accesses and inserts the summarized data into **filePopDaily** and **categoryPopDaily** tables, respectively;
- deletes its task from **dailySchedule** table.

The frequency of the starting of FAMoS modules depends on the number of records on file accesses in Authen operations log files.

CONCLUSION

The developed ***File Access Monitoring Service*** allows gathering all the information, necessary for the ordering of file and file set replicas with the aim to remove a part of them from the storage elements of ALICE VO. The functionality of ***FAMoS*** has been checked in the simulation experiments on a test server. ***FAMoS*** software is included in the AliEn version v2-21. A part of FAMoS, which records the details on file accesses (attributes plugin, see Section 3) has been deployed on AliEn central servers on 6th of August, 2013. After some analysis of the monitored data the other part of FAMoS will be deployed on the AliEn central servers. The work on the gathering and analysis of the statistics on file accesses is in progress.

ACKNOWLEDGEMENTS

We are grateful to Latchezar Betev and Pablo Saiz who suggested this work for ALICE VO and guided us in our studies of authentication and data file classification in AliEn.

Continuous discussions with Ara Grigoryan helped us to deepen our understanding of the storage volume repletion problem and of the ways to its solution.

Many details on the types of ALICE data have been clarified in the communications with our colleagues from ANSL and CERN: Vardanush Papikyan, Arsen Hayrapetyan, Costin Grigoras and Miguel Martinez Pedreira.

The details of the replica removal strategies used in CERN's CMS experiment [7] have been communicated to us by Domenico Giordano.

Our work was supported by the Grant of the World Federation of Scientists, within the Armenian H.E.P. National Scholarship Programme.

REFERENCES

[1] The ALICE experiment - http://aliweb.cern.ch

[2] The LHC (Large Hadron Collider) - http://home.web.cern.ch/about/accelerators/large-hadron-collider

[3] The CERN- http://home.web.cern.ch

[4] The AliEn (ALICE Environment on the Grid) - http://alien2.cern.ch

[5] P. Buncic, A. Peters, P.Saiz, The AliEn system, status and perspectives. Computing in High Energy and Nuclear Physics, 24-28 March 2003, La Jolla, California - http://www.slac.stanford.edu/econf/C0303241/proc/papers/MOAT004.PDF

[6] Jianlin Zhu et al, Enhancing the AliEn Web Service Authentication. 2011 J. Phys.: Conf. Ser. 331 062048 - http://iopscience.iop.org/1742-6596/331/6/062048/pdf/1742-6596_331_6_062048.pdf

[7] The CMS experiment - http://cms.web.cern.ch