**SoS, Stability and scalability of DM Services (2733)**

|  |  |
| --- | --- |
| Unique ID of the requirement | #2733 |
| Title of requirement | Stability and scalability of Data Management Services |
| Reference | <https://rt.egi.eu/guest/Ticket/Display.html?id=2733> |
| Priority | 5 |
| Submitter | Life Sciences Grid Community (LSGC VRC - <http://www.egi.eu/collaboration/LSGC.html>) |
| Status at TCB | In Clarification |
| RT tickets belonging to the requirement | -[#922 Management of many small files](https://rt.egi.eu/guest/Ticket/Display.html?id=922)  -[#919 Management of avalanches of job submission](https://rt.egi.eu/guest/Ticket/Display.html?id=919)  -[#916 DMS workload management](https://rt.egi.eu/guest/Ticket/Display.html?id=916) |
| Description of the requirement | -[#922 Management of many small files](https://rt.egi.eu/guest/Ticket/Display.html?id=922) The community may handle millions of small files, while the grid storage managers are usually mostly dealing with fewer larger files (sometimes completely preventing the registration of new files; often leading to file access low efficiency).  -[#919 Management of avalanches of job submission](https://rt.egi.eu/guest/Ticket/Display.html?id=919) Although the requirement reads about avalanches of jobs it highly related to SE and LFC systems. Due to the many jobs those perform various storage related operations many jobs do fail in general because of SE and LFC stability.  -[#916 DMS workload management](https://rt.egi.eu/guest/Ticket/Display.html?id=916) DMS is easily overloaded, which causes time-out to data access and many job failures. This kind of problem is frequent since many embarrassingly parallel application are capable of causing a large number of simultaneous file access to one or a few data servers. A mechanism is expected to avoid such failures due to server overload (e.g. these should be postponed and retried until success). |
| Related tickets in the EGI Helpdesk | Storage Elements are easily overloaded by VO activity: <https://gus.fzk.de/ws/ticket_info.php?ticket=68110>  <https://gus.fzk.de/ws/ticket_info.php?ticket=68471>  <https://gus.fzk.de/ws/ticket_info.php?ticket=68888>  LFC also has a max number of connections before it breaks: <https://gus.fzk.de/ws/ticket_info.php?ticket=60834> |
| Goals and objectives | Have more stable and scalable implementation of the following services:   * gLite SE * gLite LFC   (See version numbers below) |
| Impact | Failing jobs, lost files. |
| Affected services | Services of the LSGC VRC, i.e. sites supporting the biomed, vlemed and lsgrid VOs.  List of SE and LFC services supporting the VO (from Top-BDII at CERN):  **SEs (**StoRM, DPM, dCache), **LFC** |
| Acceptance criteria | It is possible that the current scalability and stability limits of the above listed SE and LFC services are consequence of incorrect and heterogeneous configurations of the sites. Therefore UCST asks technology providers to prepare performance documents that detail the scalability and stability limitations of the above listed SE and LFC services under typical configurations.  The LSGC VOs could use these documents to harmonise configurations across sites, and, if necessary, to request improvements in the SE and LFC software to reach higher scalability and stability levels. |

**Name:** Stability and scalability of Data Management Services

**Assessed Requirements**

The request essentially describes performance issues directly related to Storage Elements or SE’s in combination with the LFC. Performance degradation is observed either in conjunction with very small files or with an avalanche of job submissions. Customers are asking for advise on how to avoid performance degradation either by tuning configuration or enabling protection mechanisms within the storage elements.

**Executive Summary**

By design, the maximum data transfer rates of dCache and DPM scale with the number of installed disk servers. In order to satisfy the required number of file ‘open’ or ‘stat’ operations per second, the dimension of the name space database hardware needs to be chosen appropriately. For StoRM there is no general receipt as it can be operated on top of a variety of Posix file systems which consequently need to be dimensioned accordingly.

In any case, a storage element, independently of its brand, can be overloaded in different ways if not dimensioned correctly. If advice it needed for a particular storage element, customers should contact the storage software providers directly as only they can give the best advise and as ‘best practice’ for the configuration and dimensioning of storage elements may change over time. Most storage element providers offer annual workshops and tutorials and are present during EGI community events. Those events should be used to get in touch with them.

Even in cases where storage elements are configured and setup correctly for the average use case, there are always situations where peak requirements overload the storage system, degrading the overall performance. Right now only dCache provides protection against this effect by limiting the maximum number of transfers per storage node. For DPM similar protective measures are under discussion. For StoRM, it is up to the underlying storage system to protect itself against peak loads.

As the LFC is already tuned for maximum performance, peak loads can only be covered by choosing appropriately powerful hardware.

In case instabilities, other than reduced performance, is observed, the corresponding storage software provider should to be contacted directly as that is a deficiency or bug of the particular storage software.

**Efforts assessment**

N.A.

**Milestones and timelines**

N.A.

**Resources**

N.A.

**Risks**

N.A.

**Constraints**

N.A.

**Assumptions**

N.A.