PROCEEDINGS OF SCIENCE

LCG/gLite BDII performance measurements.

Lev Shamardin[*†]
Scobeltsyn Institue of Nuclear Physics, Moscow State University
E-mail: shamardin@theory.sinp.msu.ru

Information system is a vital service for the contemporary grids. LCG and gLite use the Berkely Database Information Index as a core grid information system for resource allocation and planning, therefore its performance is critical for the operation of the whole grid. In this work a stress test and benchmarking application was developed to measure the performance of the BDII for the typical queries. We present the results of performance measurements for both test and production BDII servers in LHC Computing Grid, and discuss the limitations of current information system implementation.

XI International Workshop on Advanced Computing and Analysis Techniques in Physics Research
April 23-27 2007
Amsterdam, the Netherlands

[*]Speaker.

[†]The research was partially supported by INTAS-CERN Grant 2005-7509 and RFBR Grant 06-07-89199

 http://pos.sissa.it/

PoS(ACAT)014

1. Introduction

The core of the informational system in the LCG/gLite grid is a set of central BDII servers [1], [2]. BDII server provides information about available resources, their state and general access rules. On the low level BDII is using LDAP protocol to publish a hierarchy of grid resources in a custom schema [3]. The implementation in use is using OpenLDAP 2.x as a server [4].

The Berkeley Database Information Index (BDII) consists of two or more standard LDAP databases that are populated by an update process. Port forwarding is used to enable one database to serve data while the other is refreshing. The update process obtains LDIF from either doing an ldapsearch on LDAP URLs or by running a local script that generates LDIF. The LDIF is then inserted into the LDAP database. Options exist to update the the list of LDAP URLs from a web page and to use an LDIF file from a web page to modify the data before inserting it into the database [5].

The BDII performance has a significant influence on the grid efficiency. Failures in the BDII service cause chain reaction failures on data management and workload management software. The goal of this work is to measure the performance and scaling properties of the LCG/gLite BDII server.

2. Methodology

At the first stage of the investigations the logging of all queries on a production BDII server was performed during one week. This monitoring showed that the main consumers of the BDII services are the LFC utilities and the Resource Brokers. These two types of queries were chosen as the most representative queries for measuring the BDII performance.

The LFC utilities send two requests in a row to locate the Storage Elements. The filter expressions for these requests written in the LDAP notation [6] are:

```
(GlueSEUniqueID=<se_name>)
```

and

```
(&(GlueServiceURI=*<hostname>*)
  (GlueServiceType=srm_v1))
```

The results for such queries are several rows of LDAP data for each query.

The typical query filter expression from the gLite WMS service is

```
(|(objectClass=gluevoview)
  (|(objectClass=gluecesebind)
    (|(objectClass=gluece)
      (|(objectClass=gluecluster)
        (objectClass=gluesubcluster)))))
```

The result of such query is a huge dump of grid resources, with typical sizes of 15-20 megabytes.

The proposed testing method is the following. To simulate the simultaneous load from different nodes the LDAP queries are sent from a number of processes in parallel. Each of these processes runs a set of LFC-like and WMS-like queries. The number of queries in each type in a set is equal for all processes, but the queries in each process are shuffled in an own random sequence.

All querying processes are started at the same time. When all the processes finish the results are collected from all processes and analyzed.

The first implementation was written in Perl using the Net::LDAP package [7]. This implementation appeared to be amazingly slow and memory hungry. Even eight parallel processes were enough to overload the dual dual-core Opteron machine which hosted the performance testing suite and consume more then 1 GB of memory.

The second implementation was rewritten in C using openldap-client libraries [4]. This implementation of the test suite showed excellent performance an very low CPU and memory requirements. All results below were measured using this implementation.

3. Results

Two BDII servers were used in measurements. First server is a single-node BDII server lcg15.sinp.msu.ru for the Russian region of the EGEE grid. The second server is the central LCG/EGEE BDII server lcg-bdii.cern.ch which is implemented as a number of BDII nodes in the round robin.

Figure 1 shows the response times dependence from the number of sequential queries for different number of parallel queries. LFC-like queries are used. The results are measured for the lcg-bdii.cern.ch server.

It can be seen that there is almost no dependence of the response time from the number of sequential queries sent by the testing process. This means that it is not necessary required to perform measurements with the large sequences of requests.

Figure 2 shows the response times and response failures dependence from the number of parallel queries for the LFC-like queries for the lcg15.sinp.msu.ru BDII server.

Figure 3 shows the response times and response failures dependence from the number of parallel queries for the LFC-like queries for the lcg-bdii.cern.ch BDII server.

The results show that the response time dependence from the number of parallel processes is linear and that the BDII can handle a big amount of simultaneous queries even on a single node without any failures but the response times will increase significantly.

Figure 4 shows the response times dependence from the number of parallel queries for the WMS-like queries for the lcg15.sinp.msu.ru BDII server.

Figure 5 shows the response times dependence from the number of parallel queries for the WMS-like queries for the lcg-bdii.cern.ch BDII server.

Again, the response time dependence from the number of parallel processes is linear and that the BDII can handle a rather big amount of simultaneous queries even on a single node without any failures but the response times will increase significantly.

4. Conclusion

The measurements showed that the BDII performance scales linearly. High loads on the BDII servers may significantly increase response times, but nevertheless the queries will not time out or fail up to very high loads (more then 100 parallel queries per BDII node). It also appeared that the LDAP protocol and implementation for the grid informational service is not very efficient. The data transfer delay for the amount of data for WMS resources dump could be less then 5 seconds with a 100-Mbit network, but the best results for the BDII show more then 20 seconds. It also appeared that the failures in the grid tools relying on the BDII could be reduced if the timeouts for the BDII replies are increased.

Performance characteristics measured in this research were further used in the grid simulations [8]. The simulation shows that though the BDII could serve the required loads, it is used in an inefficient ways in current grids.

The author would like to thank Andrey Kiryanov, Victor Abramovsky, Alexander Kryukov, Andrey Demichev, Gregory Shpiz for useful discussions, and also Mirco Mazzukato and Jürgen Knobloch for their help with the organization of this research.

References

[1] M. Lamanna, *The LHC computing grid project at CERN*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **543** (2004), no. 1-2 1–6. Proceedings of the IXth International Workshop on Advanced Computing and Analysis Techniques in Physics Research.

[2] *gLite: Lightweight Middleware for Grid Computing*. http://glite.web.cern.ch/glite/.

[3] C. Donley, *LDAP Programming, Management, and Integration*. Manning Publicatoins, 1st ed., 2002.

[4] *OpenLDAP*. http://www.openldap.org/.

[5] L. Field, *Berkeley Database Information Index*, 2006. Available online, https://twiki.cern.ch/twiki//bin/view/EGEE/BDII.

[6] T. Howes, *RFC 2254 — The String Representation of LDAP Search Filters*. Netscape Communications Corp., 1997.

[7] *Perl-LDAP*. http://ldap.perl.org/.

[8] A. Demichev, A. Kryukov, L. Shamardin, and G. Shpiz, *Grid infrastructure analysis with a simple flow model*, . in these proceedings.

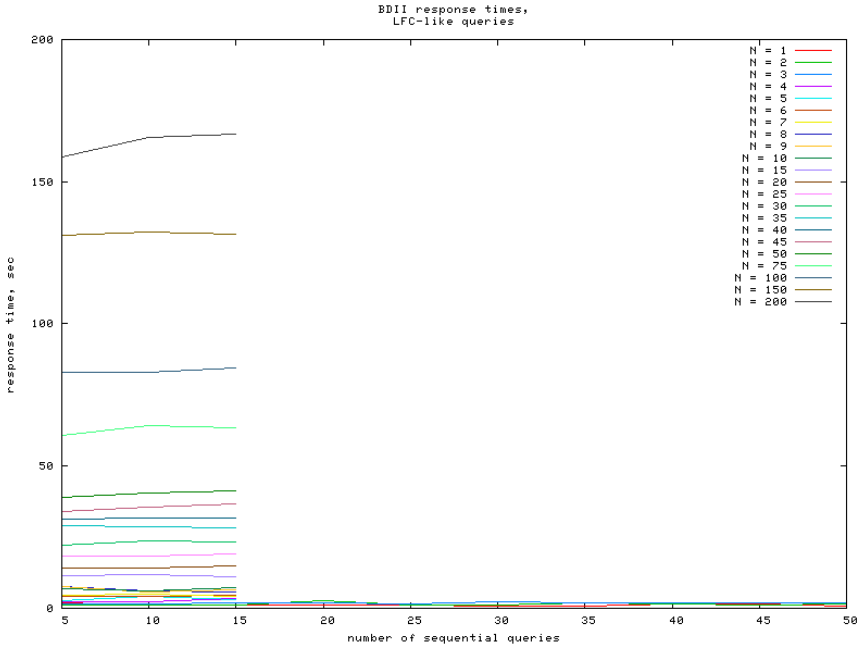

Figure 1: Dependence of the response times for LFC-like queries from the number of sequential queries for different number of parallel queries.

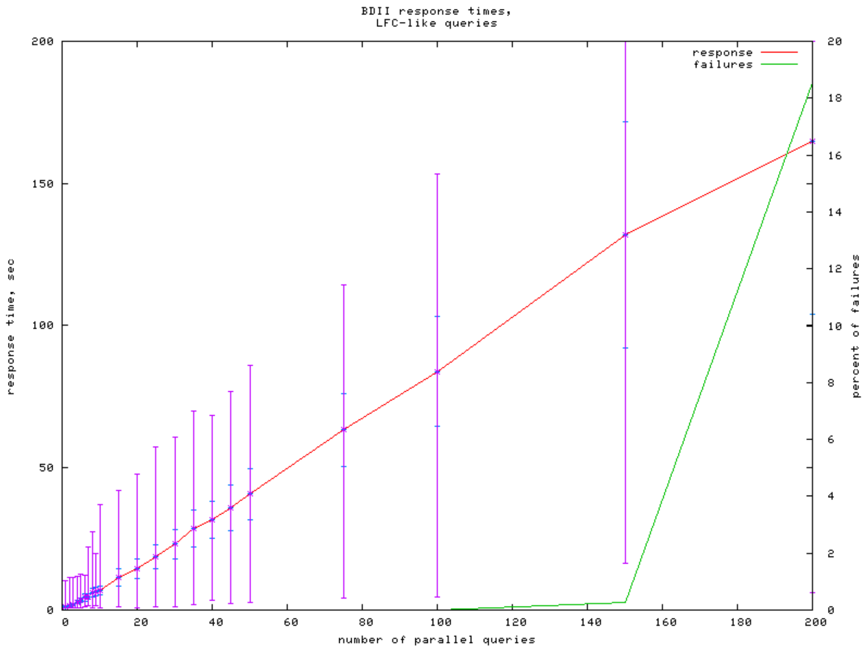

Figure 2: Dependence of the response times and response failures for LFC-like queries from the number of parallel queries for single node BDII. Error bars show the statistical error, ticks on error bars show results within the σ

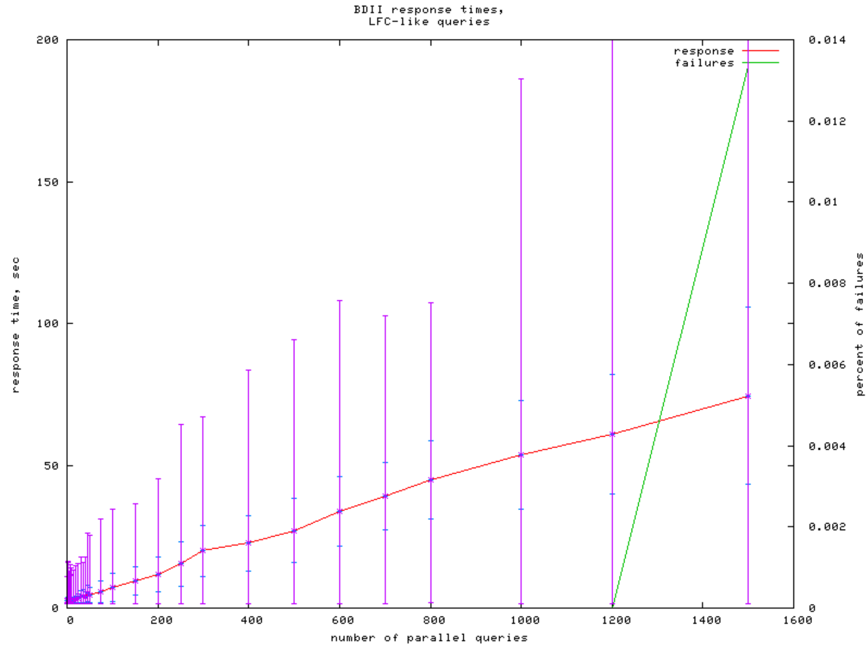

Figure 3: Dependence of the response times and response failures for LFC-like queries from the number of parallel queries for round robin BDII. Error bars show the statistical error, ticks on error bars show results within the σ

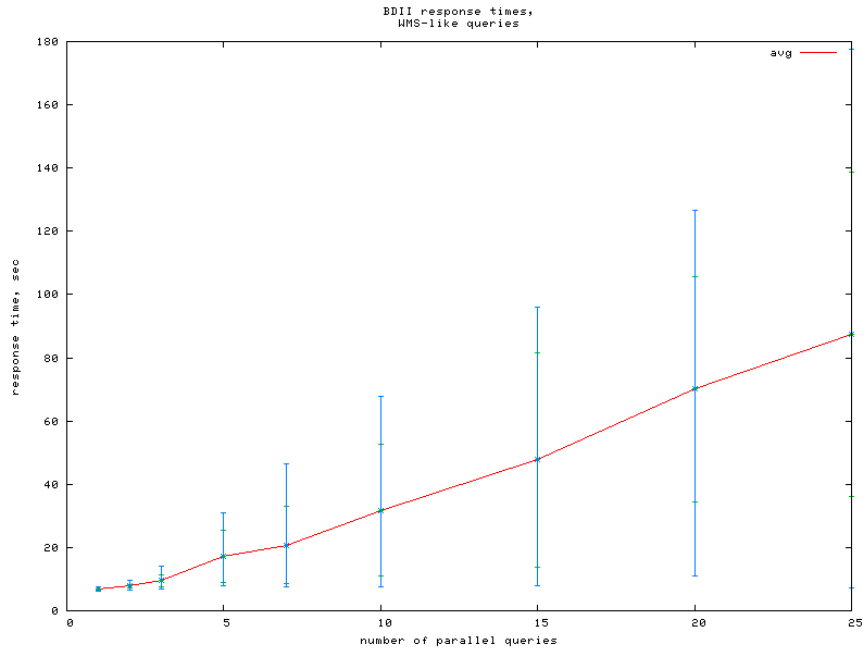

Figure 4: Dependence of the response times and response failures for WMS-like queries from the number of parallel queries for single node BDII. Error bars show the statistical error, ticks on error bars show results within the σ

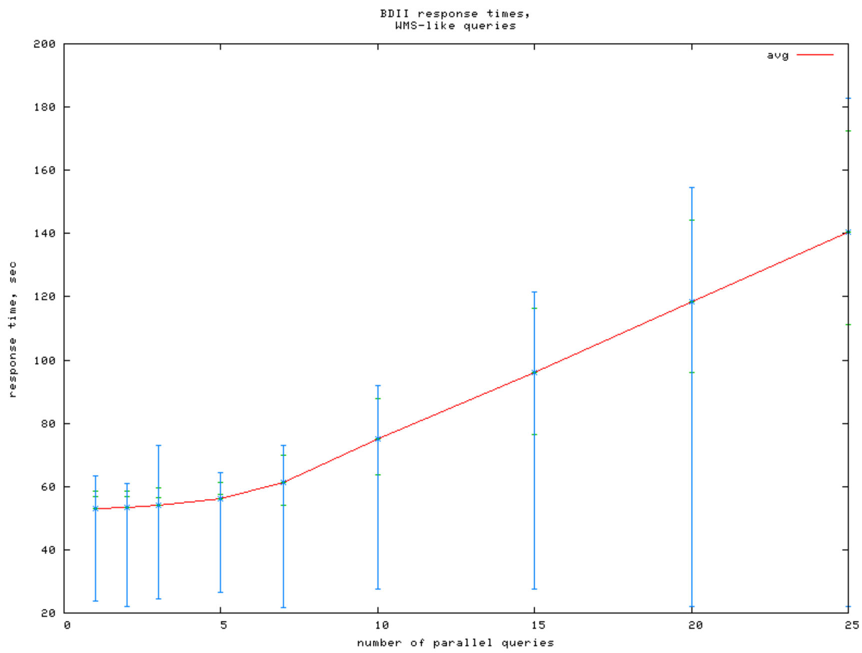

Figure 5: Dependence of the response times and response failures for WMS-like queries from the number of parallel queries for round robin BDII. Error bars show the statistical error, ticks on error bars show results within the σ