WLCG Information System Use Cases

Editor: Maria Alandes
Date: 15.09.2015
Version: 1.4
Contributors/Authors:

Alejandro Alvarez Ayllón (GFAL2/CERN), Julia Andreeva (Grid Monitoring/CERN), Alexey Anisenkov (ATLAS/CERN), Marian Babik (SAM/CERN), Stefano Belforte (CMS/INFN), Brian Bockelman (CMS/OSG), Chris Brew (GridPP), Stephen Burke (GridPP), Alessandro di Girolamo (ATLAS/CERN), Alessandra Forti (ATLAS/Manchester University), John Gordon (Accounting/RAL), Stephen Jones (GridPP), Edward Karavakis (Grid Monitoring/CERN), Maarten Litmaath (ALICE/CERN), Edward McMahon (GridPP), Stefan Roiser (LHCb/CERN), Gareth Roy (GridPP), Pablo Saiz (REBUS/CERN), Andrea Sciaba (CMS/CERN), Sam Skipsey (GridPP), Andrea Valassi (LHCb/CERN)

Table of Contents

1. Introduction

This document describes the existing use case of the WLCG Information System [1]. It is based on feedback collected from LHC experiments and WLCG activities like monitoring or accounting.

The WLCG Information system comprises different components like the BDII, GOCDB, OIM and a set of information providers. The interactions with the different components and how the information is actually obtained from the Information System and in what format (usually LDAP queries to get information described using the GLUE 1 schema) is also described.

The document also includes feedback collected from GridPP sites. Feedback from other sites belonging to WLCG could be added to the document in future versions if considered useful.
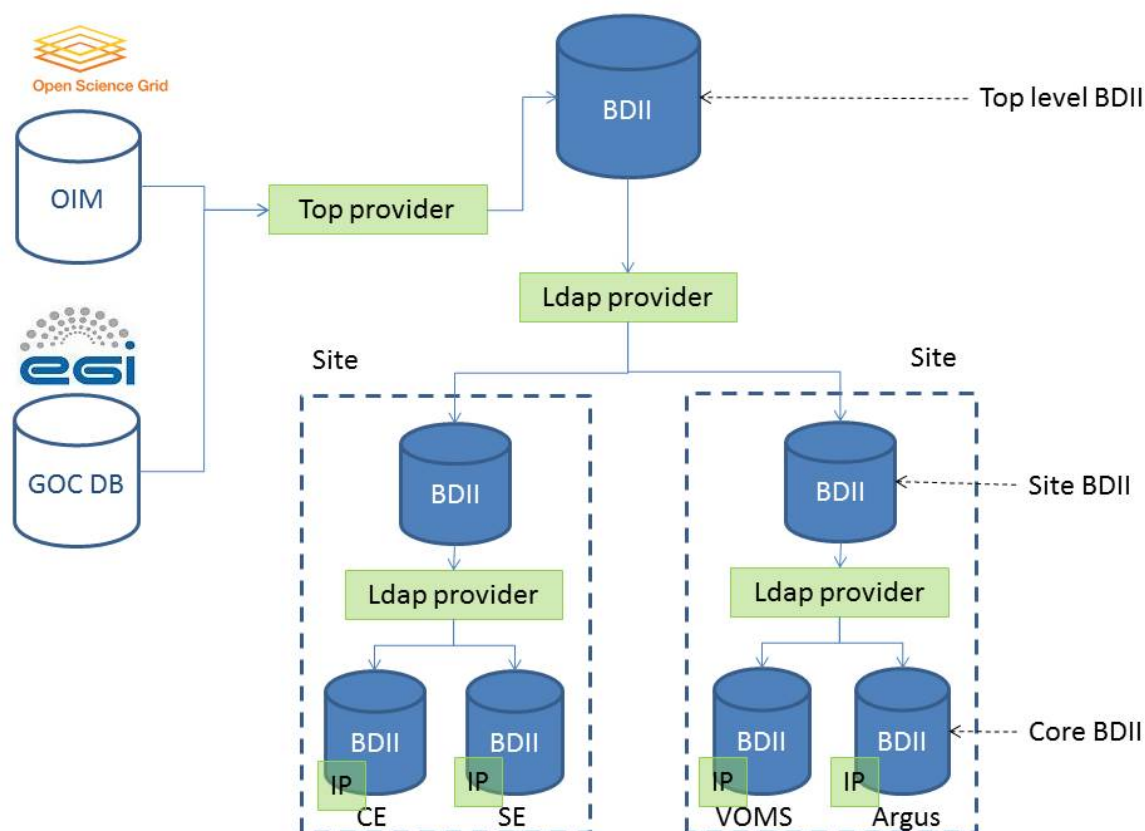
2. WLCG Information System Overview

Figure 1 - Information System components

The WLCG information system [1, 2] is a mission-critical component in the WLCG grid infrastructure. It provides detailed information about grid services which is needed for various different tasks. The grid information system has a hierarchical structure of three levels. The fundamental building block used in this hierarchy is the Berkley Database Information Index (BDII). Although the BDII has additional complexity, it can be visualized as an LDAP [3] database. The resource level or core BDII is usually co-located with the grid service and provides information about that service. Each grid site runs a site level BDII. This aggregates the information from all the resource level BDIIs running at that

site. The top level BDII aggregates all the information from all the site level BDIIs and hence contains information about all grid services. There are multiple instances of the top level BDII in order to provide a fault tolerant, load balanced service. The information system clients query a top level BDII to find the information that they require.

The BDIIs are populated with information by running information providers [4]. These are scripts which obtain information, format it as LDIF and print the result to standard out. These information providers can also be used to query other BDIIs which is how the hierarchy is built. The order in which these information providers are run is random.

The information in the information systems conforms to a schema called the GLUE schema [8]. The GLUE schema started as collaboration effort between European and US grid projects to facilitate interoperation between them. The Open Grid Forum (OGF) [5] is now responsible for the GLUE schema.

The information system is bootstrapped from the information registered in the Operations Databases of EGI and OSG grid infrastructures (GOCDB [6] and OIM [7]). When a site registers, it enters the URL for the site level BDII into the GOCDB/OIM. GOCDB/OIM then generates a list of LDAP URLs for all the sites in the grid and this is downloaded by the information provider running on the top level BDII. These URLs are then used to query all the site level BDII and the result is used to populate the top level BDII.

3. ALICE

3.1. Which information is needed from the Information System?

ALICE currently relies on information published by the resource BDII of CREAM CEs and on the site BDII of ARC CEs. In particular, the following information is used:

- Status of CEs (production vs anything else)
- Number of waiting jobs in the VOView
- Number of running jobs in the VOView

3.2. How is this information obtained from the Information System?

Most ALICE sites have a WLCG VOBOX on which local AliEn [9] services manage the workflow for their site. At such sites one such service then gathers the necessary information by querying the resource or site BDII at most once per minute using the Net::LDAP Perl module and getting the following GLUE 1 attributes:

- *GlueCEStateStatus*
- *GlueCEStateWaitingJobs*
- *GlueCEStateRunningJobs*

New jobs can be submitted to a CE when its status is Production and its numbers are below the site's thresholds determined from the ALICE site configuration LDAP service. For historical reasons a few more attributes are collected, but they do not matter for the workflow anymore, the assumption being

that an established ALICE site will have the corresponding properties set to values that are compatible with normal ALICE workflows:

- *GlueHostMainMemoryRAMSize*
- *GlueHostMainMemoryVirtualSize*
- *GlueCEPolicyMaxWallClockTime*

VOBOX admins and ALICE experts regularly make use of the *lcg-infosites* command to see what job numbers currently are published by the site's CE(s), while very occasionally the status of a CE may have to be checked through the *ldapsearch* command.

3.3.   What about sites that do not appear in the Information System?

Sites can also contribute to ALICE grid computing by setting up a private VOBOX with AliEn services that do not depend on the Information System. For example, the ALICE sites in OSG work like that. Their VOBOX either submits directly into their batch system or uses a local HTCondor installation that keeps track of the jobs and thereby knows how many are running and waiting. Such usage of HTCondor could also be envisaged for sites that currently depend on the resource or site BDII, but as long as the ALICE use cases can be served fairly reliably by the BDII, the latter approach is preferred, being the less complex option. Furthermore, under normal conditions there is no need to keep track of each and every pilot job individually. The BDII approach then puts less load on the CEs.

As of September 2015 a new AliEn module is being developed to handle the new HTCondor CE. Even though for EGI sites it will appear in the BDII, for ALICE it may end up being driven through a local HTCondor installation on the site's VOBOX, if that approach turns out to be easier to maintain.

4. ATLAS

4.1.   Which information is needed from the Information System?

ATLAS relies on various sources of information to populate its own information system called AGIS [10, 11, 12]. One of the key features of AGIS is that it makes a clear distinction between *resources provided by* and *resources used by* [13]. AGIS collects the site and services information on one side, and establishes a relationship between these resources to *ATLAS objects*, like *ATLAS Panda Resources* or *ATLAS Rucio Storage Elements*.

For what concerns the *resources provided by*, AGIS queries GOCDB, OIM, BDII and REBUS to get the relevant information for the resources allocated to ATLAS. It also queries VOMS (for VOMS roles) and PANDA Database (for dynamic PANDA resources properties).

AGIS also allows the addition and editing of information from other resources that are not available in neither GOCDB, OIM, BDII nor REBUS (e.g. full path of *Storage Areas* and their access rules, Squid configuration for Frontier, etc.)

In particular, AGIS needs the following information from GOCDB and OIM:

- Site and Services Information (list and type of services)

From BDII it needs (for each service published in GOCDB and OIM):

- CE service information for all CE flavours including HTCondor CEs
- Local queues and its properties

Storage information like Storage Areas used to be taken from BDII. This information was not published in a homogeneous way by all sites and this made it difficult to integrate it in AGIS automatically. In any case, ATLAS needed to add more information that was not available in the BDII (the full path), so it was decided to include all storage information manually in AGIS.

ATLAS also uses the HEP-SPEC06 and Logical CPUs information published in the REBUS Site Capacities view [14]. This is used by the ATLAS dashboard [15] to calculate resource utilisation. ATLAS does not need a precise number of the HEP-SPEC06 and Logical CPUs at the sites, it just needs a reasonable average core power value for each job slot provided by the site. The average core power is calculated dividing HEP-SPEC06/Logical CPUs. ATLAS measures the CPU and wall clock time of each job. For each site, ATLAS then calculates, multiplying by the average core power, the HEP-SPEC06 hours provided by the site to ATLAS. ATLAS Accounting dashboards for both computing and storage, take the pledges from REBUS. This information is also copied and cached in AGIS.

ATLAS Accounting dashboards for both computing and storage, take the pledges from REBUS. This information is also copied and cached in AGIS.

4.2. How is this information obtained from the Information System?
AGIS gets information from the BDII, REBUS, OIM and GOCDB.

AGIS queries the CERN top-level BDII (lcg-bdii.cern.ch) every 2 hours. The following GLUE 1 attributes are queried:

Site Information:
- *GlueSiteUniqueID*
- *GlueSiteName*
- *GlueSiteWeb*
- *GlueSiteLongitude*
- *GlueSiteLatitude*

CE and queue Information:
- *GlueCEUniqueID*
- *GlueCEName*
- *GlueCEStateStatus*
- *GlueCEInfoHostName*
- *GlueCEInfoGatekeeperPort*
- *GlueCEAccessControlBaseRule*
- *GlueCEInfoJobManager*
- *GlueCEImplementationName*
- *GlueCEInfoGRAMVersion*
- *GlueCEPolicyMaxWallClockTime*
- *GlueCEPolicyMaxCPUTime*
- *GlueClusterUniqueID*
- *GlueClusterName*

AGIS queries REBUS to collect the following information:

- Site HEP-SPEC06 and Logical CPUs to calculate average core power capacity metric: synchronized every 24 hours
- Federations topology: regional center, tier level, country associations: synchronized every 3 hours
- Pledges: synchronized on demand

AGIS collects information from OIM every 6 hours:

- List of sites and associated information (name, status, *emailContact*)
- List of services SE (SRM, GRIDFTP), CE, PerfSonars (Bandwidth, Latency) and associated information (endpoint, type, flavor, status, *is_monitored*, site)
- Site downtimes information (This is collected every 30 mins)

AGIS collects information from GOCDB every 6 hours:

- List of sites and associated information (name, status, emailContact, country, *certification_status*, *gocdb_pk*)
- List of services SE(SRM, GRIDFTP), CE (CE, CREAMCE, ARCCE), PerfSonars (Bandwidth, Latency) and associated information (endpoint, type, flavor, status, *is_monitored*, site relation, *gocdb_pk*)
- Site downtimes info (This is collected every 30 mins)

It should be noted that AGIS offers the possibility of manually inserting and modifying the information collected from all these sources through the AGIS web UI. AGIS encapsulates the specific logic of the different raw data providers (API, protocols, schemas, etc.) and uses high level collectors to populate its database. One of the main advantages is the possibility to replace, extend or remove the implementation of the raw data providers without affecting high level AGIS logic.

5. CMS

5.1. Which information is needed from the Information System?

5.1.1. Monitoring

The SAM tests rely on a VO feed file to have updated information on the CMS service topology. Currently, the CMS VO feed is generated by querying the BDII, SiteDB [17] and the glideinWMS factory configuration [18]. In particular, the BDII is queried to get the hostnames of the computing elements available to CMS at a site. It is planned to remove very soon the dependency on the BDII by extracting the CE endpoints only from the glideinWMS factory configuration. This will ensure perfect consistency among the CEs known to the CMS production system and the CEs to be tested by SAM; this includes also CEs that are under testing and not yet validated for production.

The CMS dashboard also uses the HEP-SPEC06 and Logical CPUs information published in REBUS to calculate the resource utilisation. REBUS capacities come from the BDII. However these plots are not used by CMS, which makes this use case irrelevant.

5.1.2. Workload management

Currently the BDII and other sources are queried to bootstrap the glideinWMS factory configuration for a site, while subsequent changes are manually applied based on information directly received from the sites. Information about HTCondor CEs in OSG comes from special collectors, as these CEs do not appear in the Information System.

It is considered useful to have a way to query the WLCG Information System to retrieve the list of CE endpoints accessible to CMS across the different infrastructure projects and their relevant properties. These include:

- Queue name, if applicable
- Number of cores, CPU and wall clock time limits

5.1.3. Data management

There are no dependencies from the WLCG Information System in the CMS data management system [19] (PhEDEx, AAA, DDM, storage monitoring, ASO) beyond the indirect dependency through FTS.

However, CMS needs to know the total available space (disk and tape) for all its storage endpoints and the quotas assigned to the various CMS groups, as the information is used for resource planning and for the DDM and transfer operations. Although quota information is very likely outside of the scope of the WLCG Information System, reliable information about the total available space would be useful. Currently this is directly communicated by the sites to the service operators and not queried from the BDII; it is then inserted in the DDM database, from which a text summary is generated and made available in the web.

Another use case is the aggregation of PhEDEx and perfSONAR data, which is already being taken care of by the WLCG Network and Transfer Metrics working group.

5.2. Which information is required and is missing from the Information System?

CMS would find extremely useful to have pledge information per site (this information is currently available in REBUS per federation). Pledges per site are necessary and sufficient for long term planning, although some sites may declare additional opportunistic resources available for a limited time.

More detailed information on the available capacity, reflecting variations with time, is nice to have but it is not required by CMS. If available, it would be useful as a troubleshooting tool, for example to investigate cases where the site utilisation is much lower than expected. However, for this purpose it is also possible to use pledge information and historical monitoring information.

Given the CMS does not require sites to publish their available capacity, calculating its total value over all CMS sites will not produce a meaningful information.

5.3. How is this information obtained from the Information System?

5.3.1. Monitoring

For the VO feed generation, an LDAP search is done on the top-level BDII (lcg-bdii.cern.ch) to find all the *GlueCEs* allowing as access control base rules *VO:CMS* or *VOMS:/cms/Role=production* and belonging to official CMS sites (those defined in the CMS SiteDB). The CE flavour is taken from *GlueCEImplementationName*.

5.3.2. Workload Management

The site entries in the factory configuration are bootstrapped using the following steps for all CMS sites:

1. Query the BDII to find all queues in a site that support the CMS VO and have reasonable wall clock limits (the minimum of *GlueCEPolicyMaxCPUTime* and *GlueCEPolicyMaxWallClockTime* should be long enough)

2. Extract hostname and queue name from the *GlueCEUniqueID* (which has the format *hostname:port/jobmanager-jm_type-queue*)

3. Manually fill in other fields (*entry_name*, *site_name*, GLIDEIN_Supported_VOs, *cmssite_name*)

4. If needed, fill by hand other properties (cores, WC / CPU limits, memory limits, etc.) if the values published in the BDII are found not to be correct.

The same procedure is followed to configure glideinWMS for other VOs as well.

6. LHCb

6.1. Which information is needed from the Information System?

LHCb relies on the following information published in the BDII:

- List of CEs
- Detailed properties of each CE

In particular, the most important pieces of information retrieved from the BDII are the Max CPU Time (in SI2K minutes) and SI2K CPU Scaling Reference attributes for each queue within a CE. This information is used by DIRAC [26] to calculate queue lengths in HEP-SPEC06 seconds as MaxCPU*SI2K*60/250. Queue lengths in HEP-SPEC06 are used within the DIRAC job submission framework, to ensure that jobs fit within the queues they are submitted to (LHCb uses only CPU time, not wall clock time, for job scheduling). This information is not used instead for monitoring purposes.

For each CE, the information can be visualized on the LHCb DIRAC Portal [27] from the Configuration Manager application (e.g. Resources -> Sites -> LCG -> LCG.CERN.ch -> CEs -> ce301.cern.ch -> Queues -> cream-lsf-grid_lhcb gives maxCPUTime=11520 and SI00=1000, as shown in Figure 2. For convenience, the same information is regularly extracted from the DIRAC Portal and displayed on a summary page listing all CEs [28].

It should be noted that LHCb relies only on the information published in the BDII (and re-published on LHCb DIRAC) for what concerns the capacities installed and available at sites. REBUS is also used in LHCb, but only for obtaining the pledged resources at the various sites. REBUS information is typically used as-is and without further processing, except for Tier-2 pledges, where the total pledge for a country without a single federation can only be obtained by computing the sum of all individual federations within that country (e.g. the pledged contribution from France is obtained as the sum of the pledges for CPPM-Marseille, GRIF-Paris, LAPP-Annecy and LPC-Clermont-Ferrand).

6.2. How is this information obtained from the Information System?

The BDII information is gathered by the DIRAC CE2CSAgent [29]. One instance of this agent runs on one node of the LHCb DIRAC infrastructure and is configured to query the BDII every 12 hours. The CE2CSAgent queries the BDII using *ldapsearch* [30] to get the following GLUE1 attributes of each CE:

- *GlueCECapability*
- *GlueCEImplementationName*
- *GlueCEPolicyMaxCPUTime*
- *GlueCEStateRunningJobs*
- *GlueCEStateStatus*
- *GlueCEUniqueID*
- *GlueChunkKey*
- *GlueClusterUniqueID*
- *GlueForeignKey*
- *GlueHostApplicationSoftwareRunTimeEnvironment*
- *GlueHostArchitecturePlatformType*
- *GlueHostBenchmarkSI00*
- *GlueHostOperatingSystemName*
- *GlueHostOperatingSystemRelease*
- *GlueHostOperatingSystemVersion*
- *GlueSEAccessProtocolType*
- *GlueSESizeFree*
- *GlueSEUniqueID*
- *GlueSiteLatitude*
- *GlueSiteLocation*
- *GlueSiteLongitude*
- *GlueSiteSysAdminContact*
- *GlueSiteUniqueID*
- *GlueSubClusterWNTmpDir*
- *GlueVOInfoPath*

In particular, the following GLUE1 attributes are the most relevant to the calculation of queue lengths and job submission in the LHCb DIRAC framework:

- *GlueCEPolicyMaxCPUTime* for the Max CPU Time
- *GlueHostBenchmarkSI00* and *GlueCECapability.CPUScalingReferenceSI00* for the CPU Scaling Reference

```
LHCb-Prod [2015-07-24 12:44:27.411318]
   LogLevel = DEBUG
   DIRAC
   Systems
   Resources
      FileCatalogs
      StorageElements
      FTSEndpoints
      Sites
         LCG
            LCG.CERN.ch
               SE = CERN-FREEZER, CERN-RAW, CERN_M-DST, CERN-DST, CERN-USER, CERN-FAILOVER, CERN-disk, CERN-tape, CERN-RDST, CERN_MC_M-DST, CERN-HIST, CERN-ARCHIVE,
               CE = ce301.cern.ch, ce302.cern.ch, ce401.cern.ch, ce402.cern.ch, ce403.cern.ch, ce404.cern.ch, ce405.cern.ch, ce406.cern.ch, ce407.cern.ch, ce408.cern.ch
               Name = CERN-PROD
               Coordinates = 06.0458:46.2325
               Mail = grid-cern-prod-admins@cern.ch
               MoUTierLevel = 0
               CEs
                  ce301.cern.ch
                     wnTmpDir = .
                     architecture = x86_64
                     OS = ScientificCERNSLC_Carbon_6.4
                     SI00 = 1955
                     Pilot = False
                     CEType = CREAM
                     SubmissionMode = Direct
                     Queues
                        cream-lsf-grid_lhcb
                           maxCPUTime = 11520
                           SI00 = 1000
                           VO = lhcb
                           MaxTotalJobs = 1000
                           MaxWaitingJobs = 50
                        cream-lsf-grid_2nh_lhcb
                        VO = lhcb
                  ce302.cern.ch
                  ce401.cern.ch
                  ce402.cern.ch
                  ce403.cern.ch
```

Figure 2. CE queue information as displayed in the LHCb DIRAC Portal [27]

7. SAM

7.1. Which information is needed from the Information System?

There are two aspects to be considered: SAM test submission framework and SAM3 (availability calculation engine and UI). They both extract information about the set of services to be tested and evaluated from the VOFeed, which in its turn takes information from BDII and other sources, as explained in the previous sections.

In addition to the set of services, SAM [21] test submission framework requires the queue name which is not currently in the VOfeed (though is planned to be included there) and is being obtained querying the BDII.

7.2. How is this information obtained from the Information System?

The information is gathered by querying the BDII. It is done during the execution of every test and a dedicated SAM BDII instance is used for this purpose. This will go away as soon and the information about queues is included in the VOFeed.

Currently it is done the following way:

We first try to match resources using access control rule set to an FQAN that will be used to execute the test, e.g.

```
ldapsearch    -x    -H    ldap://sam-bdii.cern.ch:2170    -b    mds-vo-name=local,o=grid
"(&(objectClass=GlueCE)(GlueCEStateStatus=Production)(GlueCEInfoHostName=ce-iep-
grid.saske.sk)(GlueCEAccessControlBaseRule=VOMS:/atlas/Role=pilot))"
```

If this returns no result, then we try to match any resource, with access control set to a VO name, e.g.

```
ldapsearch    -x    -H    ldap://sam-bdii.cern.ch:2170    -b    mds-vo-name=local,o=grid
"(&(objectClass=GlueCE)(GlueCEStateStatus=Production)(GlueCEInfoHostName=ce-iep-
grid.saske.sk)(|(GlueCEAccessControlBaseRule=VOMS:/atlas)(GlueCEAccessControlBaseRule=VO:at
las)))"
```

8. REBUS

8.1. Which information is needed from the Information System?

There are two type of information published in REBUS:

- Pledges: they are manually reported by the WLCG federations once per year.
- Capacities: there are two types of capacities:
    - T1 and T2 capacities: used for the WLCG accounting reports [20]. They are pre-filled automatically from the accounting portal. In the case of T1, they can be manually edited.
    - Site Capacities: they are taken automatically from the information published in the BDII. For OSG sites, HEP-SPEC06 is taken from MyOSG where resource administrators enter manually this value.

8.2. How is this information obtained from the Information System?

The information is gathered by querying the top-level BDII (lcg-bdii.cern.ch) every hour, and saving the objects in a database. The objects queried are: *GlueCluster, GlueSubCluster, GlueService, GlueSE, GlueSA and GlueCE.*

In total, there are around 50 different attributes being stored, including *GlueCEInfoLRMSType, GlueCEInfoLRMSVersion, GlueCEImplementationVersion, GlueHostOperatingSystemName, GlueHostOperatingSystemRelease, GlueSubClusterLogicalCPUs, GlueSubClusterPhysicalCPUs, GlueSESizeFree, GlueSESizeTotal, GlueSETotalNearlineSize, GlueSEType and GlueSETotalOnlineSize*.

9. GFAL2

9.1. Which information is needed from the Information System?

GFAL 2 [22] has also a dependency on the BDII:

- BDII lookups can be disabled in all cases with DBDII:ENABLE=false, or modifying the corresponding configuration file (i.e. */etc/gfal2.d/bdii.conf*)
- However, if the BDII lookups are enabled and the SURL is not a full SURL, the BDII will be contacted

9.2. How is this information obtained from the Information System?

The information is gathered by querying the top-level BDII (lcg-bdii.cern.ch) using *ldapsearch* and getting the following GLUE 1 attributes:

- *GlueServiceVersion*
- *GlueServiceEndpoint*
- *GlueServiceType*

10. IT-SDC C5 Report

10.1. Which information is needed from the Information System?

IT-SDC group at CERN generates weekly reports that are presented to the CERN IT department every Friday in the so called C5 meeting [23]. An example of the report is provided in Figure 3.

All the installed capacity information by OS, CE type, batch system and storage type is obtained by querying the top-level BDII (lcg-bdii.cern.ch).

10.2. How is this information obtained from the Information System?

The information is gathered from the REBUS database, which, as mentioned in the previous section, reads it from the top-level BDII (lcg-bdii.cern.ch).

11. WLCG Google Earth Dashboard

11.1. Which information is needed from the Information System?

WLCG Google Earth Dashboard [24] relies on the BDII to obtain the site coordinates.

11.2. How is this information obtained from the Information System?

The information is gathered by querying the top-level BDII (lcg-bdii.cern.ch) every hour using the python equivalent of *ldapsearch* and getting the following GLUE 1 attributes: *GlueSiteUniqueID*, *GlueSiteLatitude* and *GlueSiteLongitude.*

This information is combined with the coordinates collected from the ALICE LDAP service.

12. APEL Accounting System

12.1. Which information is needed from the Information System?

The standard APEL [25] Usage Record for a job contains two time duration fields *CpuDuration*, the CPU time used by the job, and *WallDuration*, the elapsed time taken by the job. Both of these are in the units reported by the batch system that ran the job.

The UR also contains Fields to describe the units of time reported: *ServiceLevelDescription* which is the type of benchmark e.g. HEP-SPEC06 and *ServiceLevel* which is the numeric value of the benchmark.

When central APEL receives UR from a site it produces a number of summaries by date, VO, User, CE, etc. In these summaries as well as the sums of the original CPU and wall clock time numbers there are normalised versions obtained by normalising each jobs times by the benchmark and summing.
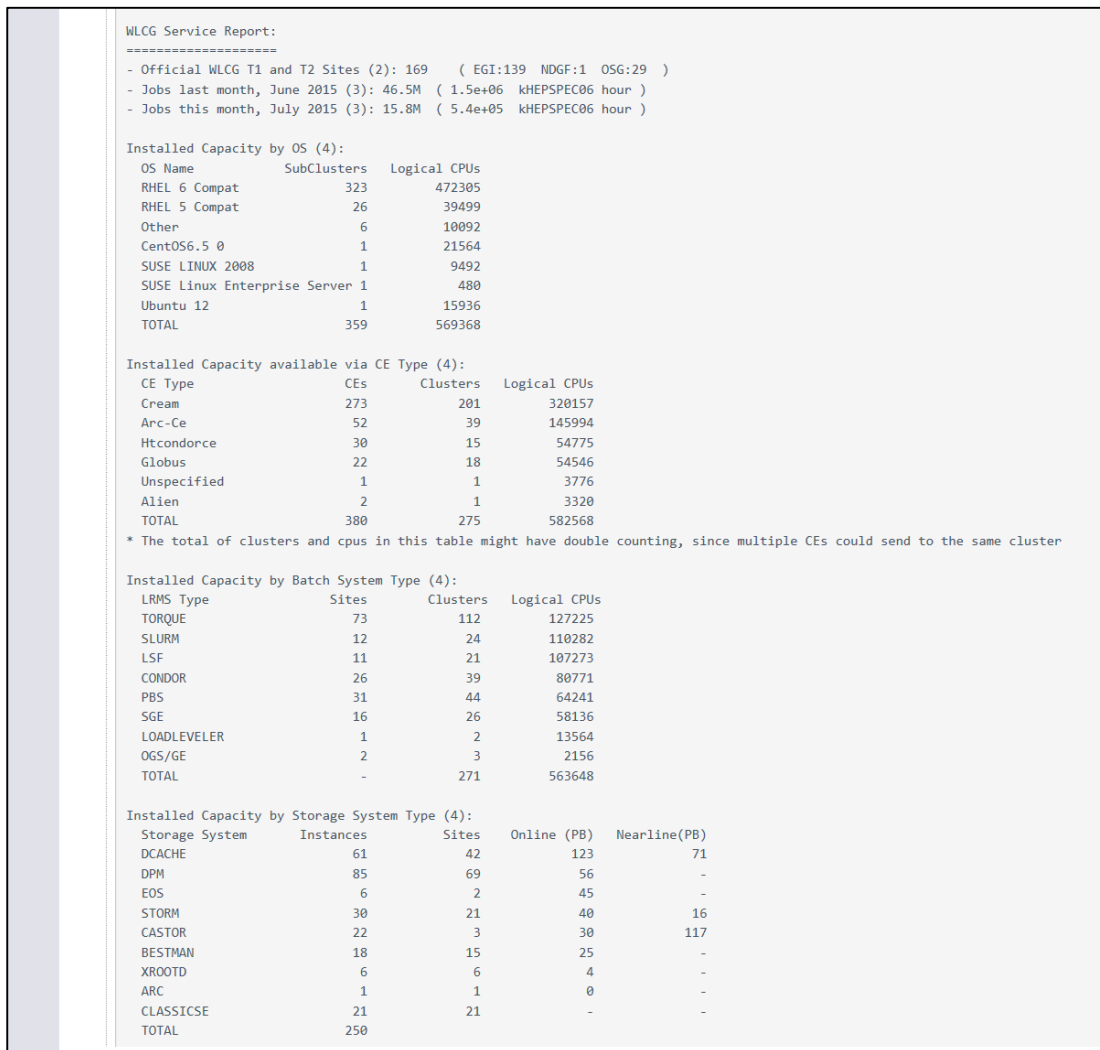
```
WLCG Service Report:
====================
- Official WLCG T1 and T2 Sites (2): 169   ( EGI:139  NDGF:1  OSG:29 )
- Jobs last month, June 2015 (3): 46.5M  ( 1.5e+06  kHEPSPEC06 hour )
- Jobs this month, July 2015 (3): 15.8M  ( 5.4e+05  kHEPSPEC06 hour )

Installed Capacity by OS (4):
  OS Name            SubClusters   Logical CPUs
  RHEL 6 Compat          323          472305
  RHEL 5 Compat           26           39499
  Other                    6           10092
  CentOS6.5 0              1           21564
  SUSE LINUX 2008          1            9492
  SUSE Linux Enterprise Server 1        480
  Ubuntu 12                1           15936
  TOTAL                  359          569368

Installed Capacity available via CE Type (4):
  CE Type           CEs      Clusters   Logical CPUs
  Cream             273        201         320157
  Arc-Ce             52         39         145994
  Htcondorce         30         15          54775
  Globus             22         18          54546
  Unspecified         1          1           3776
  Alien               2          1           3320
  TOTAL             380        275         582568
* The total of clusters and cpus in this table might have double counting, since multiple CEs could send to the same cluster

Installed Capacity by Batch System Type (4):
  LRMS Type        Sites    Clusters   Logical CPUs
  TORQUE            73        112         127225
  SLURM             12         24         110282
  LSF               11         21         107273
  CONDOR            26         39          80771
  PBS               31         44          64241
  SGE               16         26          58136
  LOADLEVELER        1          2          13564
  OGS/GE             2          3           2156
  TOTAL              -        271         563648

Installed Capacity by Storage System Type (4):
  Storage System    Instances    Sites    Online (PB)   Nearline(PB)
  DCACHE               61          42         123            71
  DPM                  85          69          56             -
  EOS                   6           2          45             -
  STORM                30          21          40            16
  CASTOR               22           3          30           117
  BESTMAN              18          15          25             -
  XROOTD                6           6           4             -
  ARC                   1           1           0             -
  CLASSICSE            21          21           -             -
  TOTAL               250
```

Figure 3 - CERN IT-SDC weekly C5 report

Most sites publish their accounting data via the APEL client which runs at the site to produce accounting usage records and send them (via message brokers) to the central APEL repository. The APEL client obtains the HEP-SPEC06 benchmark values for jobs from the BDII values by the site's CEs (or more correctly the *CE Subclusters*).

Accounting of OSG sites is done through the Gratia system [32]. Data is extracted from the centrally-operated OSG Gratia instance and sent to APEL.

APEL also uses the EGI Message Broker Network to send messages from sites to server. The current production message brokers are discovered by querying the BDII.

12.2. How is this information obtained from the Information System?

The information is gathered by the APEL client querying the top-level BDII (lcg-bdii.cern.ch) by default using *ldapsearch*. It is also possible to query a site BDII to get the information from the local site. In both cases, it gets the following GLUE 1 attributes:

- If there is a value published for *CPUScalingReferenceSI00* then it uses this.
- If not, then it uses *GlueHostBenchmarkSI00*.

In order to get the information for the message brokers, the top-level BDII (lcg-bdii.cern.ch) is also queried [31] and the following GLUE 1 attributes are used:

- *GlueServiceUniqueID*
- *GlueServiceEndpoint*
- *GlueServiceDataValue*

13.   Gridpp Sites

Sites belonging to the GridPP computing infrastructure [33] have also provided input that is considered relevant for the scope of this document.

- The Information System is considered useful for service discovery. It is queried by minor VO managers and users, especially using the *lcg-infosites* command.
- The Information System has proven to be useful during migration campaigns.
- Even if LHC VOs decide not to use it anymore, it will still be deployed by sites for non LHC VOs.
- In many cases, sys admins and smaller VOs managers hold multiple roles, so they are both consumers and providers of the Information System.
- The Information System contains too much information that is not used.
- Storage information in the Information System is more reliable than the one provided by the Computing Elements and it is in fact used by other tools.
- Supporting both GLUE 1 and GLUE 2 is a burden for sites and the motivation or benefits for doing this are not clear.
- It is not clear how the Information System will integrate future technologies such as cloud resources.
- It is useful to have WLCG information provided by one single service since this makes it easier to find it. On the other hand, there are objections about having everything in the BDII. Different types of unrelated information are mixed, e.g. Endpoint for service discovery vs Average HEP-SPEC06 values, static vs dynamic, etc.
- The maintenance of the static LDIF files and of the dynamic information providers without a configuration tool is painful. It is still labor intensive due to the size of the required LDIF files. Moreover, the existing code is not well documented and it is difficult to understand exactly which parts are the ones creating the LDIF files.
- Sites can't benefit of concepts like GLUE 1.3 *subclusters*, since there are no clients able to use them.

14. References

1 WLCG Information System: http://gridinfo.web.cern.ch/

2 Grid Deployment Experiences: The path to a production quality LDAP based grid information system. L. Field, M.W. Schulz http://cds.cern.ch/record/865688/files/p723.pdf

3 OpenLDAP: http://www.openldap.org/

4 Information providers: https://twiki.cern.ch/twiki/bin/view/EGEE/ISproviders

5 Open Grid Forum: https://www.ogf.org

6 GOCDB: http://goc.egi.eu/

7 OIM: https://oim.grid.iu.edu

8 GLUE WG: https://redmine.ogf.org/projects/glue-wg

9 ALIEN: http://alien2.cern.ch/

10 AGIS: http://atlas-agis.cern.ch/agis/

11 ATLAS Grid Information System. A. Anisenkov, A. Klimentov, R. Kuskov, T. Wenaus http://iopscience.iop.org/article/10.1088/1742-6596/331/7/072002/pdf;jsessionid=3AB5C06C6F50086D2D87AF99E2DDAC1B.c1

12 AGIS: The ATLAS Grid Information System. A. Anisenkov, S. Belov, A. Di Girolamo, S.Gayazov, A. Klimentov, D. Oleynik, A. Senchenko on behalf of the ATLAS collaboration http://iopscience.iop.org/article/10.1088/1742-6596/396/3/032006/pdf

13 AGIS: Evolution of Distributed Computing Information System for ATLAS A. Anisenkov, A. Di Girolamo, M. Alandes, E. Karavakis on behalf of the ATLAS Collaboration. https://cds.cern.ch/record/2014648/files/ATL-SOFT-PROC-2015-006.6.pdf

14 REBUS Site Capacities: http://rebus/apps/capacities/sites/

15 ATLAS dashboard: http://dashb-atlas-job.cern.ch

16 MyOSG: https://myosg.grid.iu.edu/about

17 SiteDB: marshalling people and resources available to CMS. S. Metson, D. Bonacorsi, M. Dias Ferreira, R. Egeland. http://iopscience.iop.org/article/10.1088/1742-6596/219/7/072044/pdf

18 glideinWMS: http://www.uscms.org/SoftwareComputing/Grid/WMS/glideinWMS/doc.prd/index.html

19 CMS Data Management System. M. Giffels, Y. Guo, V. Kuznetsov, N. Magini, T. Wildish. http://iopscience.iop.org/article/10.1088/1742-6596/513/4/042052/pdf

20 WLCG Accounting Reports: http://wlcg.web.cern.ch/collaboration/reporting/accounting

21 SAM: http://wlcg-sam.cern.ch/

22 GFAL 2: http://dmc.web.cern.ch/projects/gfal-2/documentation

23 IT C5 meeting: http://information-technology.web.cern.ch/about/meeting/c5

24 WLCG Google Earth Dashboard: http://dashb-earth.cern.ch/dashboard/

25 APEL: https://wiki.egi.eu/wiki/APEL

26 CPU Normalisation:
https://github.com/DIRACGrid/DIRAC/blob/integration/WorkloadManagementSystem/Client/CPUNormalization.py

27 LHCb DIRAC Portal: https://lhcb-portal-dirac.cern.ch/DIRAC

28 DIRAC queue information: http://lhcbproject.web.cern.ch/lhcbproject/Operations/queues.html

29 DIRAC CE2CSAgent :
https://github.com/DIRACGrid/DIRAC/blob/integration/ConfigurationSystem/Agent/CE2CSAgent.py

30 DIRAC ldapsearch queries: https://github.com/DIRACGrid/DIRAC/blob/integration/Core/Utilities/Grid.py

31 APEL code to get message brokers from BDII:
https://github.com/apel/ssm/blob/643e87b6850a8a3bf14669635d4404c1e11048b2/ssm/brokers.py

32 Open Science Grid Accounting: http://gratiaweb.opensciencegrid.org/gratia/

33 GridPP: http://www.gridpp.ac.uk/