Table of Contents

Information System excerpts from 2012 TEG reports

Source⤢

Overview of the Information services:

The Grid information system enables users, applications and services to discover which services exist in a Grid infrastructure and retrieve information about their structure and state. A recent survey identified 6 main use cases:

- Service discovery;
- Installed Software;
- Storage Capacity (to identify dark data and for storage accounting);
- Batch system queue status (also required for pilot job submission);
- Configuration (to populate VO-specific configuration databases);
- Installed Capacity.

The information system a fully distributed hierarchical system based on OpenLDAP and it has been in continuous operation since the first LCG-2 release. It was designed and developed to overcome the limitations of previous implementations of the information system. Its main components are:

- A hierarchy of BDII services, publishing information via LDAP, which include: resource-level BDIIs, site-level BDIIs and top-level BDIIs;
- A schema describing the information (the GLUE schema) and its LDAP implementation;
- Information providers, which collect information at the resource level and send it to a resource-level BDII.

The information system infrastructure is reasonably stable, although the published information is sometimes unstable at the source. Overall, it meets the requirements of the WLCG production infrastructure, which has been achieved with a relatively small investment and support, especially when compared to other services.

A major achievement has been the agreement of a global information model (the GLUE schema) which took into account input from many diverse communities. Over time a substantial amount of operational experience has been accumulated.

The most critical issues for the information system are related to its stability and to the validity and the accuracy of the information.

Stability:

The information system was originally designed to show the current state of a Grid infrastructure for workload scheduling purposes. Given the highly dynamic nature of the distributed computing infrastructure, any issues in the underlying Grid service (instabilities of the service or its service- level BDII) or distributed infrastructure itself (site- or resource- level BDIIs disappearing, etc.), may cause information not to be consistently published, resulting in information instability. Today, the primary use cases of the information have changed to ones that require information stability, e.g. service discovery for operational tools. As such, this information instability is now seen as major issue. Typical symptoms are:

- Information that vanishes or becomes stale;
- Random output from queries (for example when more BDII nodes are behind an alias and one of them is not working well).

- Open issues include how to deal with disappearing services, for example to decide for how long service information has to be cached;

There is not yet a satisfactory way to use the BDII for ARC services (workarounds have to be deployed who are not generally applicable).

Information validity

The published information cannot be used if its format does not respect conventions. A first level of validation comes from the compliance with the GLUE schema: however, this still leaves a lot of margin for mistakes, in particular considering that the structure of the information is complex from a configuration point of view and it is not trivial for system administrators to configure, leaving sometimes room for different interpretations. A trivial example is multi-disk vs. multidisk for the GlueSEArchitecture. This is partially addressed by validation tools (gstat-validation and glue-validator) but they cannot prevent invalid information from being published. Such checks need to be improved and executed before the information is published;

The policy for publishing resources is also not clear, as there are uncertified sites not registered in GOCDB or OIM which are publishing services in the top BDII.

Information accuracy

Even if information is validated it can still be incorrect or too old. As the information system is not optimized for dynamic information, it was observed that the latency for a change in a dynamic value to propagate to a top BDII can be up to 10 minutes. This casts some doubts on the usefulness of dynamic information; Information providers are unreliable, in particular for storage information: very often the space usage numbers are obviously not realistic or are missing altogether. The published information should be better validated at the WLCG and the site level.

Other considerations

The BDII information should be more thoroughly certified and audited by WLCG. A strong push in this direction would come from a decision of the experiments to use the BDII information;

Retrieving information from the BDII is not straightforward due to the complexity of the LDAP schema and requires non-trivial code to be written: therefore better client tools to get the information should be provided;

It was suggested that we should rely more on NGI for provisioning of core services such as the BDII. EGI is aiming at 99% minimum availability for NGI-provided top-BDII services.

Evolution

Even if proposing actual solutions is not in the scope of this document, it is worth reporting that there are already ideas of how to overcome the limitations of the current system. The information system mixes three types of information that have different properties/requirements. We will define this information to be structural data, meta data and state data. Structural data is information about the existence of sites and services etc. This contains information that is static through the lifetime of the service such as UniqueID and Type. As a result this information is never modified and changes very slowly. Meta data is all other information about a service/site that may be modified throughout its lifetime excluding state data. Typically such changes coincide with a service update. State data is information about the current state of the service that is transient. Such information is highly dynamic, such as the number of running jobs. Each type of information should be provided by a system optimized for that purpose. This would result in three systems that we will call the Service Catalogue, Meta Data Catalogue and a Service Monitoring System.

The Service Catalogue presents an interface for discovering services and should be restricted structural data. This should be a generic service for all infrastructure/VOs.

The Meta Data Catalogue should use the Service Catalogue to discover what services exist and contact them directly to obtain further information about them. This could be considered a generic service or a VO specific tool, such as the existing Configuration databases, when VOs, can annotate this information with their own naming and semantics.

The Service Monitoring System would carry the transient state information. A publish/subscribe model would be suitable for such a system.

Recomendations

I2.17 Improve the stability of the information services I2.18 Improve the validity of the information I2.19 Improve the accuracy of the information R2.4 In the short term, improve the stability of the Information System by deploying the cached BDII and the accuracy of the information via better validation tools In the long term, evolve the Information System by re-evaluating the usefulness of existing information and by refactoring the system into separate services for structural data, metadata and state data R7: OSG Operations is researching the best way forward with our Information Systems. Until this research is completed we will not make changes in the current IS in production. Upon completion of this research will we engage WLCG IS with our plans. R7: The information discovery system is deployed by a large number of EGI user communities for service discovery in workload distribution. Consequently, improvement of information accuracy from the information discovery through validation is a top-priority for EGI. EGI welcomes collaboration with WLCG to leverage on existing expertise in WLCG and EGI, and to avoid duplication of effort. A complementary ongoing EGI activity that is beneficial to WLCG is the improvement of failover configurations of the NGI information system servers. Top-BDII availability is monitored monthly since January 2012 and underperforming NGIs are requested to improve their configurations.

Operations and Tools TEG

The information system is used today very differently than in the past and at the same time its full potential is not reached due to significant information accuracy, validity and volatility issues. For some use cases, the Information System does not contain the information at all, for others it is felt that they are not able to provide reliable information and in any case contains no history. Sites prefer not to expose those information granting direct access to the batch systems since this is felt as too intrusive and anyway the interpretation of the numbers requires too much understanding of the site internals (dynamic fair-share is a clear example);

Experiments have no easy way to monitor tape systems. Also disk space monitoring is poor (numbers from the Information System are not always reliable);

Benchmarking. Normalisation of CPU data requires a reliable knowledge of the CPU power that must be measured with HEPSPEC06. This information is unreliable at several sites, as it has to be manually maintained by sites;

Short term

Improve the Information System via full deployment of the cached BDII and a strengthening of information validation. For this reason, in the short term it is recommended to encourage the full deployment of the cached BDII (which would greatly reduce the volatility by 1-2 orders of magnitude) and strengthen current validation tools, like Nagios probes, or develop new ones (which would improve the reliability of the information). This work should be followed up by a WLCG task force, which should also develop a strategy for improving those information providers relevant to WLCG.

Long term

Review the published information and possibly drop the unnecessary one; refactor the system into optimized tools focused on providing structural data (static), metadata and state data (transient). For a longer term WLCG should take into consideration which use cases are most relevant for the experiments today and in the foreseeable future (service discovery, experiment custom annotations, etc.).

The current technology should be revisited, possibly towards a separation of services depending on the type of information they produce (structural data, meta data and state data).

The goals to achieve are, on one side, the transition to a more robust and sustainable system, and on another side, to decrease the need by the experiments to develop parallel, possibly inconsistent systems.

WM TEG

See twiki

-- MariaALANDESPRADILLO - 2015-07-14

This topic: EGEE > ISTEG2012Summay
Topic revision: r1 - 2015-07-14 - MariaALANDESPRADILLO