

GoToGrid - A Web-Oriented Tool in Support to Sites for LCG Installations

G. Diez-Andino, O. Keeble, P. Nyczyk, A. Retico, M. W. Schulz, I. Tkachev, A. Usai.
CERN, Geneva, Switzerland *

Abstract

The installation and configuration of LCG middleware, as it is currently being done, is complex and delicate.

An "accurate" configuration of all the services of LCG middleware requires a deep knowledge of the inside dynamics and hundreds of parameters to be dealt with.

On the other hand, the number of parameters and flags that are strictly needed in order to run a working "default" configuration of the middleware is relatively small, due to the fact that the values to be set mainly deal with environment configuration and with a limited set of possible operation scenarios.

This "default" configuration appears to be the most suitable for sites joining LCG for the first time. The G2G system is aimed to support Site Administrators to easily perform such a configuration.

G2G combines the gathering of configuration information, provided by sites, with the dynamic adaptive creation of customized documentation and installation tools.

By using a web interface and being requested only for the relevant configuration information, site Administrators will be able to design the desired configuration of their own LCG site.

Site configuration data is collected and stored in a well defined format liable to be used as the interface to different configuration management tools.

BACKGROUND

The definition of production quality installation and configuration procedures for the LCG (LHC Computing Grid) Middleware, done during the last year, has revealed itself to be a significantly difficult task. This was mostly due to two main reasons: The first (logical) one is the inherent difficulty of the product itself, that is the result of the integration of several heterogeneous software components, each one with its own configuration language. The second (not so guessable) reason is the negative influence that the regular use made of the LCFGng fabric management tool as an integrated distribution system for the LCG releases has had on the quality of the existing configuration procedures.

When, in fact, the need for a generic installation procedure (not relying on any automated tool) arose, the existing LCFGng procedures were reverse-engineered in order to produce a set of generic installation guides for the differ-

ent node types [1]. The resulting procedures soon appeared to be somehow more complex and time-consuming than strictly necessary.

Namely, a surprising thing was the disproportion between the amount of local configuration information actually needed from the sites (sometimes less than five parameters), the number of required manual configuration steps (considerably higher), and the number of pages to be read by the Site Administrator (not less than thirty for a very simple node).

It was observed that the guides could be considerably simplified if some site specific information (e.g. the batch and storage system in use, the name and number of supported VOs) was known in advance. This would allow the guides to focus on the information relevant to the site and would eliminate coverage of scenarios not implemented locally.

Another issue affecting the readability of the guide arose indirectly from the lack of a generic configuration reference for the LCG software. In order to clarify more general aspects of middleware configuration, node-specific guides contained much information that would more naturally be centralised elsewhere, reducing the guides' readability and increasing the likelihood of administrators skipping important sections.

The need to refine the maintaining process of the installation documentation; the opportunity to have a central configuration repository for sites using the generic installation¹; the possibility to track site configuration changes; the desire to re-use existing documentation modules, in open order, were other relevant motivations that brought our group to design and start to develop the integrated system subject of this document.

SCOPE AND KEY FEATURES

The **GoToGrid** tool (*G2G* from now on) is a web-based application to support the generic installation and configuration of (LCG) grid middleware.

Relevant site-dependent configuration information is entered by Site Administrators and stored in a centrally managed database. The required configuration parameters are a considerably reduced subset of the entire space of LCG parameters. Their values, which the Site Administrator is requested for, mostly deal with site topology and organization (e.g. names, supported VOs, used batch systems) and not with the integration of middleware components

* This work has been partially funded by: "Istituto Nazionale di Fisica Nucleare", Roma, Italy; "Ministerio de Educacion y Ciencia", Madrid, Spain

¹For LCFGng-driven installations the site configuration history is currently maintained on a CVS server at CERN.

(e.g. port numbers, timeouts, flags, protocols...). Therefore, G2G does not define any new meta-configuration language to describe site local values, which are introduced into the system through simple forms (*Profile Manager*).

On the basis of the information introduced in the database, G2G provides added-value tools, such as ready-to-use configuration files pre-compiled with site's relevant values, customized installation scripts, individual installation guides and configuration examples dynamically created according to site actual settings. The supported installation uses "default" values for the majority of not site-specific parameters. In particular, this "default" configuration appears to be the most suitable for sites joining LCG for the first time. Post-installation changes in the configuration of not environmental parameter, such as their maintenance, is out of the scope of G2G and should then be done on the real configuration files (manually or by a third-party fabric management tool).

The installations of application software from the Virtual Organizations (E.g. proprietary software from the LHC Experiments) on the fabric node is as well out of the scope of G2G.

The G2G server is supposed to run at CERN and to be maintained, for the time being, by the Grid Deployment Team.

Although the "core business" of G2G is to provide documentation and installation tools mainly focused on the need of medium-small sites, the tool can be used at different levels from sites of different scale. *All* sites (large and small ones) can profit, for instance, of the G2G first-level orientation and help system (*Grid Assistant*), of the site profile editing tool (*Profile Editor*) and of the *Profile Description File*. *Small* sites can use the automatically generated configuration files and scripts to have a "default" installation done with a minimum effort. *Large* sites, finally, can take advantage of the output documentation and of the provided "default" installation kit (configuration files and scripts) in order to configure their own fabric management system. Furthermore, as a side-effect, the site configuration information stored in the database and the change logs are suitable to be queried by support people in case of wrong configuration.

A short introduction to the main G2G features is provided in the following sub-sections.

G2G Portal

The *G2G Portal* is the entry point for the G2G features. It is the hub from where the user can access the two main Front-End services (*Grid Assistant* and *Profile Manager*), trigger on the generation of the desired output products and, finally, get the desired output.

Furthermore this web page contains links to other relevant sites for installation operation and can be used to notify and stress relevant news.

Grid Assistant

The *Grid Assistant* is a G2G public service (no user registration needed to access). It is basically a static tree of predefined questions and answers the aim of which is to provide a first-level help for those Site Administrators, even being completely new to LCG Project, can use The Grid Assistant to start a site configuration from scratch.

The Administrator looking for hints to configure the site can start the interactive assistant and he is conducted through a sequence of some very basic questions like "How many nodes will your site be composed of?", "How many VOs do you want to support?", "Which ones?" and so on.

Once the user has chosen among the predefined answers and sent the form he is given back a document with a list of relevant mandatory requirements, recommended choices, suggestions and possibilities like "The suggested configuration of your LCG site is: 1 CE, 1 SE, and the other nodes as WNs"...

Profile Manager

The *Profile Manager* is a tool to create, modify, handle and delete site profiles. It allows the user to insert, query and modify data.

The access to the Profile Manager is limited to **registered users**. In order to create new profiles or to edit existing ones users need to log-in first.

After logging in, pop-up notifications with release-specific relevant news and warnings may be issued.

Site Administrators can then select the version of the Middleware they want to work with and create site profiles accordingly.

The amount of information that a Site Administrator is required for is neatly reduced with respect to the LCFGng-driven procedure. The parameters handled in the G2G information schema are (roughly) a subset of the content of the "site-cfg.h" LCFGng configuration file holding less than sixty parameters per site. This is because, as it was said before, G2G is not supposed to handle all possible configuration changes (local and functional ones) but limits its scope to the local and environmental site values.

Site Administrators can create as many profiles they want, and they can then choose to *activate* the whole of them or just a subset. To activate a profile means to trigger on the creation of the G2G output products (documentation and scripts) in support of the site installation according to the "active" profile.

Another interesting feature of the Profile Manager is the *Take Snapshot* option, by which users can "freeze" a particular revision of a given profile. When a snapshot is taken, the profiles goes into a "read only" state and subsequent changes will be saved in a new revision. No actions can be performed on "frozen" profile except the "activation" of the profile (or its deletion)².

²A more detailed description of the possible states of the profile can be found in [3]

The information inserted by the Site Administrator in the database via the Profile Manager will persist through backward compatible releases of G2G. That means that, in case of a new release of the system, all the information dealing with the existing profiles is silently migrated.

Configuration Repository

All the inserted site profiles are stored in a central database. The *Site Configuration Repository* is a component of the architecture that in a certain way can be seen as an overall feature of the G2G system.

In fact, if currently sites using LCFGng store their site-profile in a CVS server, there is no agreed convention on how to publish site info for sites using generic installation. This may be a drawback of the manual installation in terms of problem solving costs. Using the Site Configuration Repository the Site Administrator (or the support people) can easily keep track of occurring changes.

Logs of the activities done on the profiles are also kept in the database for configuration problem solving purposes.

Profile Description File

Another relevant general feature of the G2G system is the *Profile Description File*. When the user runs an “activate” or “create XML” command via the Profile Manager, an XML document is generated out of the site info on the database.

Within the G2G system this interface is used by back-end systems to generate customized guides and scripts. The flow of data from the database to the back-end subsystems, in other words, is split with an intermediate stage.

`database -> XML document -> application`

The reason why this intermediate stage has been defined in the data flow from the database to the back-end application is that, in this way, the gathering of information from the database is decoupled from its application. This interface can be seen as a “hook” in the system, a common data exchange format suitable to be used by internal subsystems, by future system extensions and possibly by third-party fabric management tools.

Furthermore, this document is a convenient “summary” of the site with a reasonably readable structure and it may be useful to an experienced person to have a “quick look” to the site in order to understand what’s going on.

Scripts and Guides on-demand

The core feature of G2G is the dynamic production of tools (configuration files, guides and installation scripts) to be used by Site Administrators in order to physically install their site with the minimum possible effort.

With this purpose, customized *configuration files* for all the middleware components are dynamically generated according to site-specific values present in the site profile. These files are there to be used both by the scripts, that will create them on the physical machines and by the guides

that will make highly specific configuration examples out of them.

Since the installation and configuration scripts can be seen as an implementation of the configuration guide, and, of course, with the guides they share the same site information, the task of generating customized scripts and guides could be logically seen as a single complex task for the system. Nevertheless the requirements for the two kinds of product are slightly different (so that that the production is done by two separate subsystems, as shown later in the “Architecture” section).

Particularly the *installation and configuration scripts* are focused on the minimization of the operations needed to Site Administrators for their use. They will install all the needed packages on the node writing appropriately the configuration files and running eventual post-installation steps. They will be mostly useful to administrators of medium-small sites with no particular need to go for different configuration than a “default” one.

The *Manual Installation and Configuration Guides* are the generic reference with a “step by step” description of the operations to be made to install and configure the site. They contain examples of the configuration files dynamically generated according to site specific values in the site profile. The focus in these documents, is on the minimization of logical branches due to different possible site configuration choices. All possible (supported) decisions are supposed to have been taken, at this stage, so that the final appearance of the guide is just a flat sequence of configuration steps to be made. The produced guides are in keeping with the needs of the administrator of a large site who is already prepared to do extra steps on his own, but wants a clear and simple recipe for what is supported.

A POSSIBLE USER SCENARIO

In this section we outline a possible user scenario where the G2G features above described are used by a “new-comer” Site Administrator, in order to perform a complete manual installation and configuration of his site³.

The LCG Site Administrator (SA from now on) gets access the G2G system through his own web browser. He needs some hints on how to configure her site. So he uses the **Grid Assistant** at this purpose.

After having made the opportune choices, SA is ready to go. Then he registers to the system, logs-in and starts using the **Profile Manager** tool to insert relevant configuration information.

At the end of this phase (that takes as long as SA’s experience allows) he decide **activate** the configuration in the database.

So he can take the scripts and the guide and start the physical installation and configuration of the site.

³for a more detailed view of the possible user interactions with the G2G system see [2]

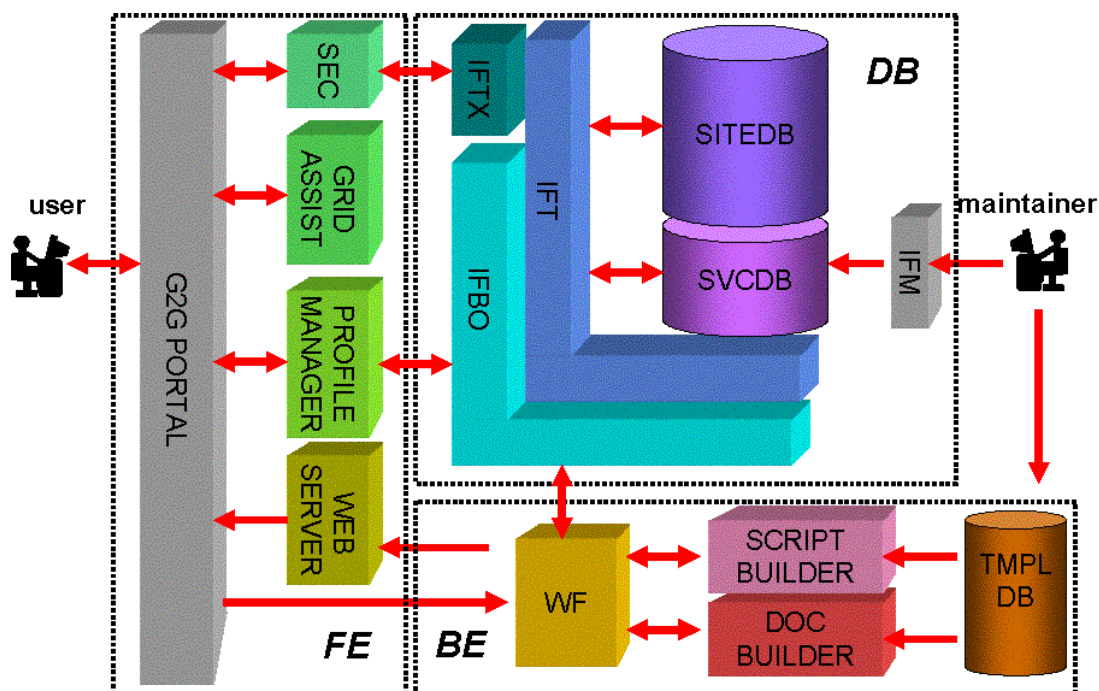


Figure 1: G2G Logical Architecture.

ARCHITECTURE

The logical architecture of the G2G system is shown in Fig. 1 In this section a short functional description is given for each one of the represented modules⁴.

FE - Front-End System

G2G Portal: Web site. Entry point to G2G features.

SEC: Registration, Authentication, Authorization.

Grid Assistant: First-level help and orientation interactive system.

Profile Manager: Profile Editing and Managing interactive tool.

TMP FS: Temporary storage for output products.

DB - Central Database

SITEDB: Site Information Repository.

IFTX,IFBO,IFT: Interface to Site Database, It decouples the internal, component-oriented, schema structure, from the presentation done by the Profile Manager, that is grid-node-oriented.

SVCDB: Service Configuration Repository. Internal System configuration. In particular this database contains the mapping between user-level requirements (e.g. a UI-RB-BDII node) and system requirements (the list

of all middleware components to install in order to have a UI-RB-BDII node).

IFM: Interface to Service DB. A set of tools to access and maintain the Service DB.

BE - Back-End System

WF: Work Flow manager. It decodes user orders using the configuration in SVCDB, retrieves site data from SITEDB, starts-up the appropriate Back-End engines (SCRIPT BUILDER and DOC BUILDER) to fulfill the order and finally publishes the results in TMP FS.

SCRIPT BUILDER: Engine for customized scripts generation.

DOC BUILDER: Engine for customized documents generation.

TMPLDB: Template Repository: It contains the templates of configuration files to be shared by SCRIPT BUILDER and DOC BUILDER.

REFERENCES

- [1] Grid Deployment Project - GIS Section , "LCG Manual Installation and configuration guide"
<http://www.cern.ch/grid-deployment/gis/release-docs/MIG-index.html>.
- [2] Antonio Retico , "G2G - Use Cases"
<http://www.cern.ch/antonio.retico/G2G/doc/UC.pdf>.
- [3] Antonio Retico , "G2G - UML Model"
<http://www.cern.ch/antonio.retico/G2G/model/>.

⁴for a more technical view of the architecture of the G2G system check the Class Diagrams in [3]