

Glance Project: a database retrieval mechanism for the ATLAS detector

C. Maidantchik

COPPE, UFRJ, Brazil

F. F. Graef and K. K. Galvão

Escola Politécnica, UFRJ, Brazil

K. Pommès

CERN, Switzerland

Abstract. During the construction and commissioning phases of the ATLAS detector, data related to the installation, placement, testing and performance of the equipment are stored in relational databases. Each group acquires and saves information in different servers, using diverse technologies, data modeling and terminologies. Installation and maintenance during the experiment construction and operation depends on the access to this information, as well as imply in its update. The development of retrieval and update systems for each data set requires too much effort and high maintenance cost. The Glance system retrieves and inserts/updates data independently of the modeling and technology used for the storage, recognizes the repositories internal structure and guides the user through the creation of search and insertion interfaces. Distinct and spread data sets can be transparently integrated in one interface. Data can be exported/imported to/from various formats. The system handles many independent interfaces, which can be accessed by users or other applications at any time. This paper describes the Glance conception, its development and features. The system usage is illustrated with examples. Current status and future work are also discussed.

1. Introduction

The ATLAS Detector is constructed by an international collaboration and is assembled and tested at CERN. During the construction and commissioning activities, each group acquires data concerning, for example, equipment parts identification and performance, electronic configuration and tests results. The information is stored using different data modeling and technologies. Considering that the collaborators are geographically spread and have diverse cultures, distinct terminologies arise. As the detector subsystems are integrated, the collaborators need to retrieve and relate data sets stored in several sources. Data acquired and stored during the equipment tests commonly compose sets of measurements over the time, such as voltages, currents and temperatures. These data need to be transformed by calculating means and RMS values, which are more suitable for long term analyzes.

It would demand an unreasonable effort and time to define a single modeling structure that fulfills all the requirements for the whole collaboration. To build a specific retrieval tool, it

is necessary to know the details on how data was modeled and the technology used to store it. If each group creates their own system for each data set of interest, a small change in one repository may require adjustments in many programs, increasing the maintenance costs.

Each collaborator might need a specific view over the same data set. For example, the group related to one of the sub-detector's cryogenic needs to see the routing of the installed pipes that connect to their equipment. On the other side, the infrastructure group is interested only on the pipe placement, type and length, regardless of its connections.

Insertion and update of data in distributed repositories face similar problems. Full equipment traceability must be achieved in order to meet INB (Institute Nuclaire de Base) rules. Equipment installation, replacement and removal/waste have to be managed. This requires updates in more than one data set at a time. For example, to register one device installation it is necessary to insert its identity and specific properties in the Equipment database, the description of its electronic configuration in the Functional Positions database, and the configuration of the connectivity with other items in the Cables database. Since each data set has its own update mechanism, the insertion of this information should be done in three different steps, using three distinct solutions. This effort is multiplied by the number of equipment items that must be managed in the 15 years of the experiment operation.

Sunderraman et. al. [1], in a research on integrating data from distributed and heterogeneous sources, proposes an extension of the SQL language. Using his extension, multiple remote databases may be referred in the same query or update statement. He also implemented a Java API, which splits a global query into sub-queries corresponding to each data source referenced. The API merges the results received from the remote systems. However, this approach still requires that the user have detailed knowledge of how the repositories were modeled and of interface programming.

This paper presents the Glance Project, a system designed to address the aspects discussed above. Its concept and features for data retrieval, manipulation and insertion are presented. The system architecture is then discussed, followed by the results obtained by its usage.

2. The Glance Project

The Glance [2] is a single system that performs data retrieval and insertion/update in distinct and geographically spread repositories. The system connects to several storage technologies, such as Oracle (used by the main CERN database servers) [3], MySQL [4] and Microsoft SQL Server [5]. Glance automatically recognizes the internal structure of the databases. It allows the creation of customized Search Interfaces (SI) and Insertion Interfaces (II), without the necessity of previously knowing the data modeling of each data set.

2.1. Data retrieval

When creating a Search Interface, Glance presents a list with few predefined databases, as shown in figure 1. Other repositories can be accessed by supplying the information necessary for locating and authenticating. The system then connects to the server and displays its internal structure, gradually increasing the level of details, so that the user can define the data set of interest.

Based on the selected data set, the description of a Search Interface is generated in XML format. It contains the necessary information to connect to the data source and the list of attributes with their types, names and association with database columns. This interface description may be customized by an advanced user in order to associate synonyms, description texts and lists of possible attribute's values to match specific terminologies. A display option can be set to determine whether an attribute should be available only for defining the search rules, only in the visualization of the results, or always. This description can be stored by the system

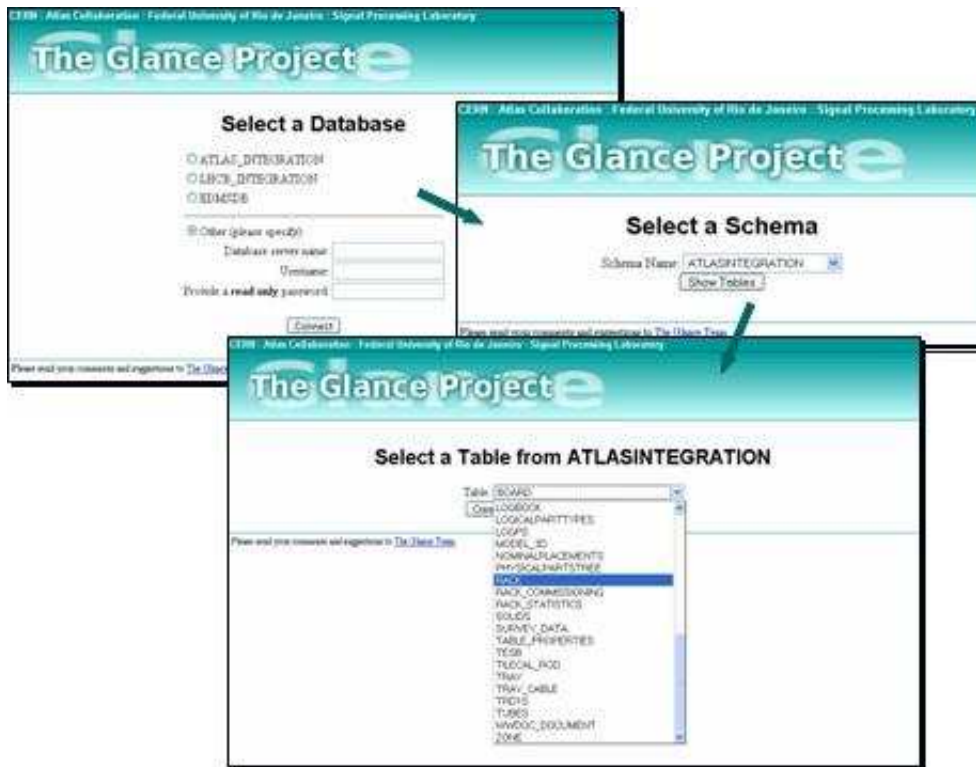


Figure 1. Creating a Search Interface.

for later use. A single system, therefore, handles several interfaces, permitting the creation of different views over the same data sets.

After creating a new Search Interface, or selecting from the list of the existing ones, the system retrieves its description and generates a corresponding HTML form. The user can define multiple search rules, which are composed by three elements. The first is an attribute, corresponding to a column in a table or a composition of various attributes in the data set. A select box presents all the options available. The second is an operator, to establish a comparison. Another select box shows different operators, depending on the type of the attribute previously selected. For example, for an alphanumeric attribute, the options can be "Contains", "Equal To" or "Different Than", while for a numeric attribute mathematical operators ("=", "<=", ">=", "<>", ">", "<") are presented. The third element is the reference value, specified in a text box. A list with options can be displayed, in order to guide the user in its selection. It can also be specified whether the results should match all rules or at least one of them.

The search results may be displayed as HTML table, or downloaded as a XML or CSV file to be used with other programs. The system also generates an URL containing all the search parameters. With this URL, the same search can be repeated in the future retrieving the new results, but without the need to fill in the rules again.

2.2. Data manipulation: integration and operation

A Search Interface can describe related data sets located in distinct data sources. When executing a search, after retrieving the data from the various databases, a merge mechanism integrates the different data sets as described in the SI.

The system integrates with external programs to perform operations over the retrieved data.

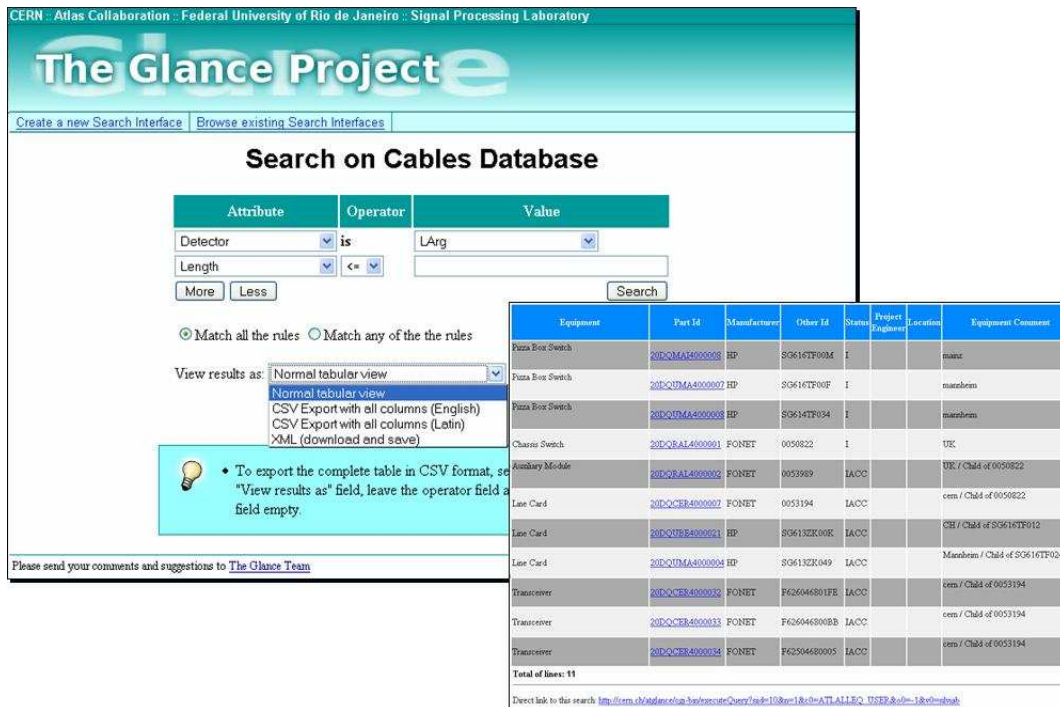


Figure 2. Search Interface.

Predefined tasks such as calculating mean values or standard deviation of numeric data, are executed by a specific program called by Glance after retrieving the data. Then, the results are delivered to the requestor (the user or another application).

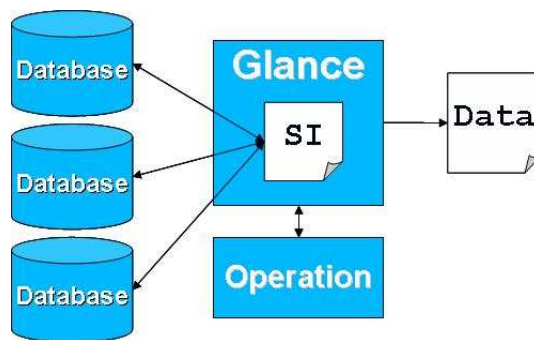


Figure 3. Data manipulation.

2.3. Data insertion and update

The Insertion Interfaces allow the registration of new records in a data set, as well as the update of existent records. In each interface, a table is presented like in figure 4, composed by the attributes of the corresponding data set. The user fills in the new data and can add lines to the table according to the amount of information to be inserted. Attribute values can be propagated to a set of selected lines. It is also possible to copy values from an entire line to others. As an alternative, the data can be imported from CSV and Excel files. The information

is validated according to the attribute types and other restrictions previously specified in the interface description. Some examples of these restrictions are: a numeric value should be inside a specified range, an alphanumeric field has a maximum length, among others.

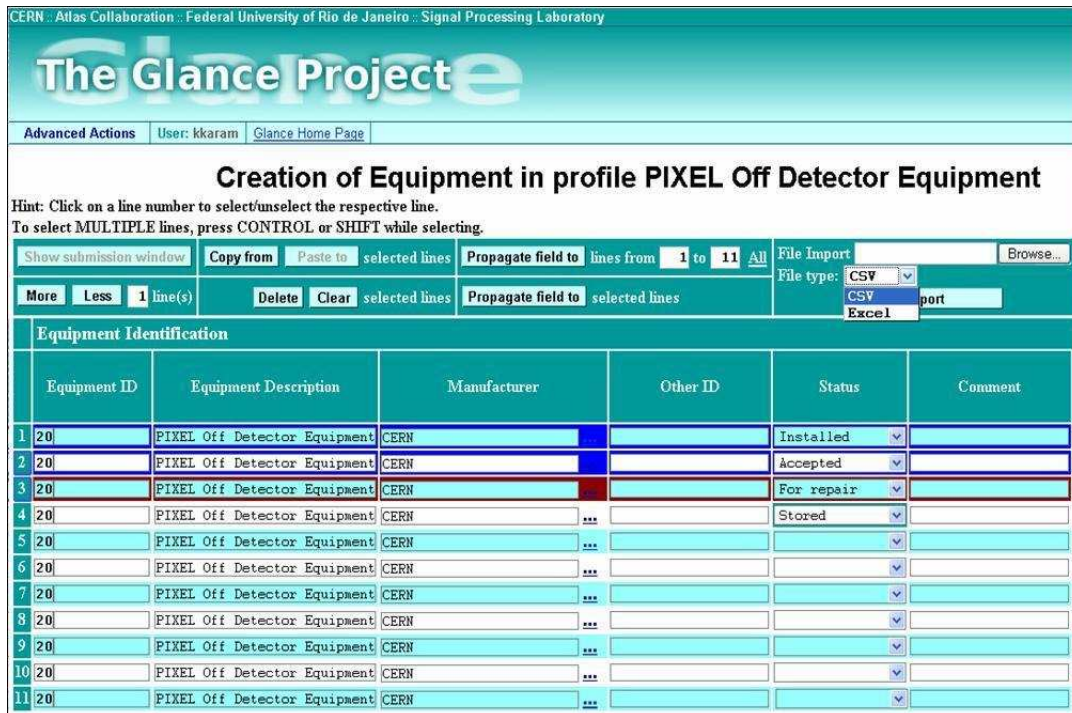


Figure 4. Insertion Interface.

3. The system architecture

The system is composed of a set of components written in C++, each one implementing a specific feature. Each component is a separate program, and is constructed using the same basic architecture.

The programs communicate with the web server via the Common Gateway Interface (CGI), receiving the user's input and generating the results of each step. The GNU CgiCC library [6] is used to implement this interface.

The connection to the databases is done through connectors. Each connector interfaces a specific database technology. The system can currently retrieve data from Oracle, using the OCCI library [7], as well as MySQL and Microsoft SQL Server, using an ODBC [8] driver for each one. New database technologies can be added by implementing other connectors and does not interfere with the feature specific logic of the components.

Integrating these two blocks there is the Glance Engine. It is responsible for parsing the XML description of the interfaces, handling the database connections and queries, and communicating with the web server. The adequate database connector is chosen based on the information on the interface XML description. The DOM parser from the Xerces-C++ library [9] is used for interpreting with the XML files.

The specific feature of each component is then built on top of the Glance Engine. Figure 5 illustrates this architecture.

The execution of each of the system's task involves the interaction of a series of components. The results generated by each component gives the user the options of the next possible steps,

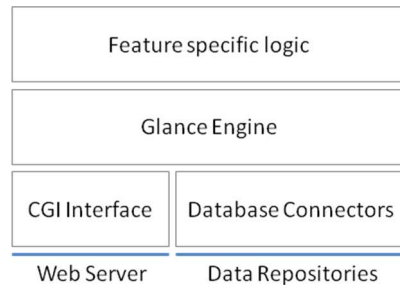


Figure 5. Architecture of the system's components.

each one corresponding to a different component. The actual interaction among the components is therefore based on the user's choices.

The typical component interaction when using an existing SI is illustrated in Figure 6. The first component lists the existing SIs. The next one interprets the selected SIs description and creates a corresponding HTML form for the user to fill the search parameters. The last one takes the parameters and combines the SIs XML description in order to query the database and display the results in the appropriate format, as well as integrating the data if more than one database is being queried at the same time.



Figure 6. Typical process of executing an existing interface.

For the process of creating a SI, the first component presents the user a list of known databases or accepts the directions for a new one. The second is the responsible for presenting the database structure, and iterates, gradually increasing the level of details displayed. When the user selects the data set of interest, the next component comes in action. It recognizes the structure of the data set and generates a corresponding SI description. It also offers the option to open the XML description and customize it according to the users' needs. This process is illustrated in Figure 7.

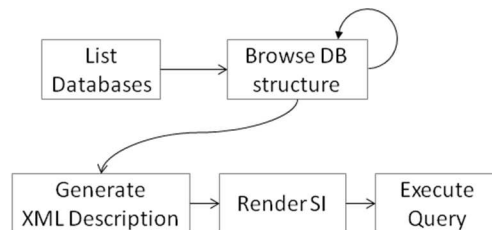


Figure 7. Typical process of creating a Search Interface.

The execution of an Insertion Interface is similar to the SI. First a list of existing IIs is displayed, and the second component is used to render the II to a HTML form. Depending on the task being performed, an additional component may be used to retrieve existing data to be modified. Another component may be used if the data is to be imported from a file supplied by

the user. Finally, the last component submits the data to the corresponding repositories. Figure 8

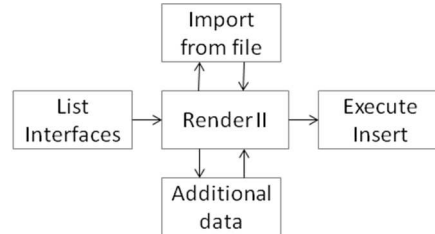


Figure 8. Components interaction when executing an Insertion Interface.

4. Results

The Glance is used by the ATLAS collaboration, accessing diverse data sources inside this group. The Equipment Database contains records from more than 92000 items located in the experimental cavern. This data is accessed via 10 Search Interfaces. Each SI provides a view over a subset of the data, divided by equipment types and the various ATLAS subsystems. Examples are: Inner Detector Equipment, LArg Equipment, Support Structures, among others. About 4000 devices were inserted in this database during the first semester of 2007, with the support of 10 Insertion Interfaces created with the Glance system.

The Functional Positions database stores information describing the location of each equipment item and its electronic configuration. This source contains data referring to about 18000 items, such as racks, crates, boards, sensors, among others. This information is accessed via 17 Glance Search Interfaces. Each SI concentrates on a specific view over the whole repository. For example, one SI allows retrieval of information concerning the equipment displacement history. Other SIs are associated to subsets of the items, dependent on the device types: connectors, channels, and others.

The Cable database stores information describing the connectivity between the equipment items. One important SI exports the data from 55600 cables in an interval of time of less than 30 seconds. Another Search Interface displays information for the pairs of connected cables.

The different substructures of the detector are being integrated and tested in the final operation site. During this phase, databases are set up to describe performed tests. The Glance is used in the retrieval of information from the Tile Calorimeter's Detector Control System (DCS) database and the Racks Commissioning database. The DCS monitors the voltage, current and temperature on several points of the Tile Calorimeter and registers the measurements. Glance retrieves these measurements and, through the use of an external program, calculates the mean value and standard deviation over the time. Glance makes these non-physics data available, in XML format, to the DCS Web System [10], an application specialized in the TileCal DCS monitoring.

A Web interface was created to display alarms details from the Detector Security System (DSS). It shows information about the alarms, input sensors and actions implemented by this system. This interface is based on the Glance Search Interfaces to retrieve data from the DSS database.

5. Conclusions

The international collaboration that builds the ATLAS detector stores the data in separate databases. The Glance project addresses the major necessities for retrieving and modifying

data in this environment. By displaying the database structure, then automatically generating an adequate search interface for the selected data set, the system allows any collaborator to retrieve data without knowing the details of the data model or the programming tools necessary to interface the database.

The system integrates data coming from different repositories, and can use external programs to perform operation over the retrieved information in order to display it in a way that will be useful for the collaboration. When inserting or modifying data, the system transparently takes the necessary steps for submitting information to the adequate repositories, allowing the users to concentrate on their tasks.

The system addresses the diversities of terminologies by allowing the customization of the attribute names and descriptions, and by displaying a list of possible options to be chosen by the user, instead of typing the values.

As the interfaces are described in a high level format, a single system handles many independent retrieval, insertion and modification mechanisms. Therefore the maintenance is reduced to a single code base. The chosen technologies for implementing the system are widely supported, what allows the system to be maintained over the whole lifetime of the experiment.

The Glance system currently has interfaces that are used to retrieve data from the ATLAS Technical Coordination databases [11], the Tile Calorimeter's DCS, the Detector Safety System. Its also used to retrieve, insert and modify data for the Equipment Database.

The next steps for the Glance project are to improve the recognition of relationship among the data sets, to detect changes on the database modeling based on the interfaces' description and automatic adjustments, to improve the data integration supporting other kinds of repositories that not necessarily are relational databases. Researches are also being done in the areas of data integration and performance optimizations.

References

- [1] Sunderraman R, Dogdu E, Madiraju P and Malladi L 2005 *ACM Southeast Regional Conference, Proceedings of the 43rd annual Southeast regional conference*
- [2] Maidantchik C, Graef F F, Galvão K K and Pommès K 2007 The Glance Project URL <http://atglance.web.cern.ch/atglance/>
- [3] CERN 2007 Database Infrastructure Services URL <http://service-oracle.web.cern.ch/service-oracle/>
- [4] MySQL 2007 The MySQL open source database URL <http://www.mysql.com/>
- [5] Microsoft 2007 Microsoft SQL Server URL <http://www.microsoft.com/sql/default.msp>
- [6] GNU 2007 Gnu cgicc library URL <http://www.gnu.org/software/cgicc/>
- [7] Oracle 2007 Oracle C++ Call Interface URL <http://www.oracle.com/technology/tech/oci/occi/index.html>
- [8] Microsoft 2007 ODBC - Open Database Connectivity Overview URL <http://support.microsoft.com/kb/110093>
- [9] Foundation T A S 2007 Xerces C++ parser URL <http://xerces.apache.org/xerces-c/>
- [10] Maidantchik C, Ferreira F G and Graef F F 2007 The DCS Web System URL <http://tcws.web.cern.ch/tcws/DCS/>
- [11] Pommès K, Molina-Perez J, Galvão K K and Malyukova I 2007 ATLAS Database Installation & Integration URL <http://atlas.web.cern.ch/Atlas/TCOORD/Activities/Installation/Database/index.html>