AGIS: The ATLAS Grid Information System

A. Anisenkov [1], A. Di Girolamo[2], A. Klimentov[3], A. Senchenko[1] on behalf of the ATLAS Collaboration

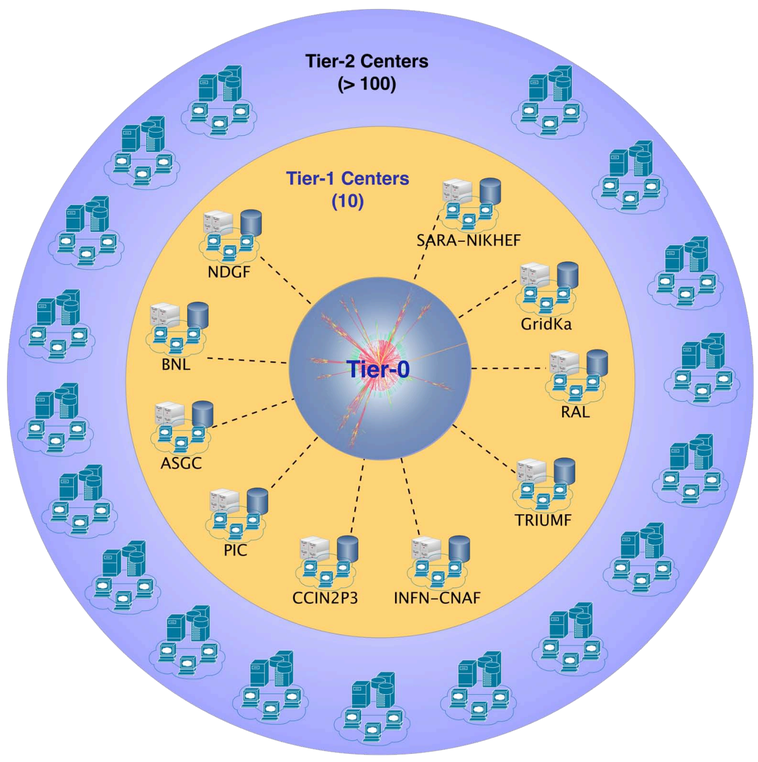
[1] Budker Institute of Nuclear Physics, Novosibirsk, Russia
[2] CERN, Geneva, Switzerland
[3] Brookhaven National Laboratory, Upton, USA

ATLAS Grid Information System (AGIS) designed to integrate configuration and status information about resources, services and topology of whole ATLAS Grid needed by ATLAS Distributed Computing (ADC) applications and services[1].

AGIS project goals

The ATLAS experiment[2] produces petabytes of data per year that needs to be distributed globally to sites worldwide according to the ATLAS computing model[1]. The ATLAS computing model is based on a worldwide Grid computing infrastructure that uses a set of hierarchical tiers. The root node of the complex tiered topology is Tier-0, located at CERN itself. After a primary data processing at CERN in a Tier-0 facility, raw data from the detector and reconstructed data products are replicated to 10 Tier-1 centers according to the ATLAS data distribution policy. In addition to the data taken from the detector, simulated data is produced on Grid resources worldwide, and this data must also be replicated. Tier-1 and associated underlying tiers are grouped into a cloud.
The ATLAS grid is composed by two independent sub-grids:

- **European Grid Infrastructure (EGI)[3]**
- **Open Science Grid (OSG)[4]**

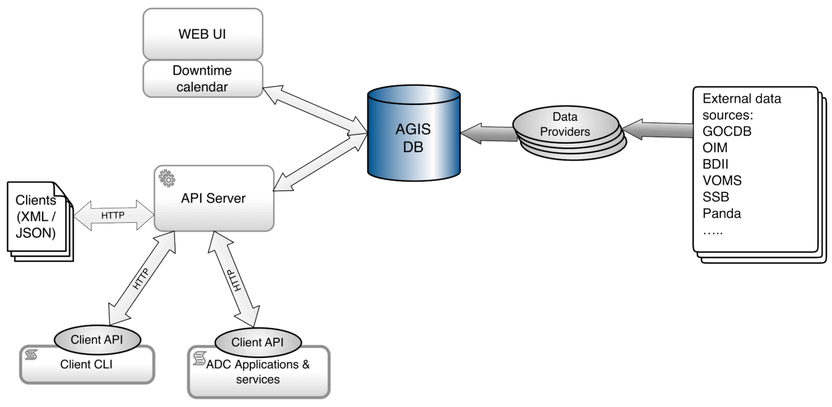


ADC applications[5] and services require the diversity of common information, configurations, parameters and quasi-static data stored in different sources. The difficulty faced by ADC applications is that ATLAS computing uses a variety of Grid infrastructures which have different information services, application interfaces, communication systems and policies. AGIS is designed to integrate information about resources, services and topology of the ATLAS grid infrastructure from various independent sources including BDII, GOCDB, the ATLAS data management system[6] and the ATLAS PanDA workload management system[7]. Being an intermediate middleware system between a client and external information sources, AGIS automatically collects and keeps data up to date, caching information required by and specific for ATLAS, removing the source as a direct dependency for clients but without duplicating the source information system itself.

System Architecture

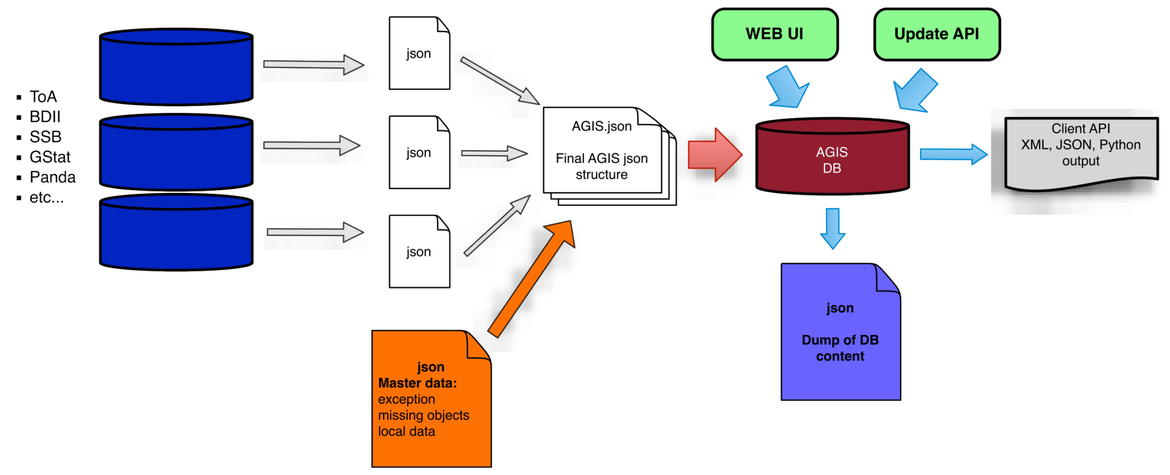
AGIS Architecture is based on the classic client-server computing model. Oracle DBMS is chosen as a database backend. The system provides various interfaces: API, WEB, CLI - to retrieve and to manage data.

AGIS uses Django framework[8] as a high-level web application framework written in Python. Object Relation Mapping technique built in Django framework allows to operate and to access data in terms of high level models avoiding direct dependence of relational database system used.



To automatically populate the database a set of agents runs on the main AGIS server. One of the operation requirements recently addressed to the information system is the possibility to track the changes and the functionality to fully reproduce the database content from previously saved state.

To meet this demand the data populating approach has been improved by introducing intermediate level in the data collecting procedure. New approach assumes that data collected from external source is saved into a temporary storage before actually being inserted into the database. It allows to overwrite specific values and control the data coming in AGIS. As the data interchange format the JSON structure is chosen.



Provided Information

There are two types of information stored in AGIS:

- **data retrieved from external sources (TiersOfATLAS, GOCDB, BDII, myOSG, PanDA databases, GStat source, etc),**
- **data managed within AGIS**

AGIS stores all information concerning ATLAS topology: clouds, regional centers, sites specifics, such as geography, time zone, geo coordinates, etc. It also stores resources and services information – LFC, FTS, CE, SRM, Squid, Frontier services and others.
Key examples of information stored in AGIS:

- **ATLAS topology (clouds, centres, sites and sites specifics)**
- **site's resources and services information, status and its description**
- **site information and configuration**
- **data replication sharing and pairing policy**
- **list of activities and its properties (Functional Tests, data distribution, etc.)**
- **global configuration parameters needed by ADC applications**
- **user related information (privileges, roles, account info) for ADC applications**
- **downtime information about ATLAS sites, list of affected services**
- **site blacklisting data**

Status and Plans

AGIS is developed to provide a single portal and information cache for application developers and interactive users. AGIS is currently in production for the ATLAS experiment.
Many AGIS services, in particular the API functionality, are starting to be actively used in the production applications. Web interfaces such as a site downtime calendar and ATLAS topology viewers are widely used by shifters and data distribution experts.
New external information sources will be supported in the near future. So, more ADC applications are expected to use AGIS as the main information system.
AGIS design and basic principles included into the architecture allows to use its core part for several HEP experiments. Russian Federation sites are considering AGIS as an information system to be used for RF cloud after new Tier-1 facilities will be commissioned.

References

1. *The ATLAS Computing Model, Adams, D ; Barberis, D ; Bee, C P ; Hawkings, R ; Jarp,S ; Jones, R ; Malon, D ; Poggioli, L ; Poulard, G ; Quarrie, D et al. - ATLAS Collaboration.,ATL-SOFT-2004-007; ATL-COM-SOFT-2004-009; CERN-ATL-COM-SOFT-2004-009;CERN-LHCC-2004-037-G-085.- Geneva : CERN, 2005 - 37*
2. *G. Aad et al., "The ATLAS Experiment at the CERN Large Hadron Collider," JINST 3 (2008) S08003.*
3. *European Grid Infrastructure (EGI) Web page, http://www.egi.eu/*
4. *Open Science Grid (OSG) Web page, http://www.opensciencegrid.org/*
5. *A.Anisyonkov, A.Klimentov, D.Krivashin, M.Titov. ATLAS Distributed Computing, XXII International Symposium on Nuclear Electronics & Computing (NEC 2009) (Institute for Nuclear Research and Nuclear Energy, Varna, Bulgaria, 2009), pp. 4. September 7-14, 2009; Proceedings of the Symposium; E10,11-2010-22, Dubna, 2010 - p. 45-49*
6. *M. Branco et al., "Managing ATLAS data on a petabyte-scale with DQ2," J. Phys. Con f.Ser. 119 (2008) 062017*
7. *T. Maeno, "PanDA: Distributed production and distributed analysis system for ATLAS," J. Phys. Conf. Ser. 119 (2008) 062036.*
8. *Django project web site, http://www.djangoproject.com*

Poster presented at the 19th International Conference on Computing in High Energy and Nuclear Physics (CHEP), New York, USA, 21-25 May 2012.