

## ATLAS Distributed Analysis (ADA)

D Adams, K Harrison, RWL Jones, A Soroko, A Tan  
(Contact: Roger.Jones@lancaster.ac.uk)

The ATLAS distributed analysis system (ADA<sup>1</sup>) faces the challenge of supporting distributed users, data and processing. It must support all analysis activities that can be sensibly distributed including Monte Carlo production, data reconstruction and the extraction of summary data from all types of event samples. It is also required to work across three current major Grid deployments: LCG<sup>2</sup>, Grid3<sup>3</sup> and NorduGrid<sup>4</sup> and the future systems into which these evolve.

A prototype system has been created based on the GANGA<sup>5</sup> user interface and the ATLAS production<sup>6</sup> and DIAL<sup>7</sup> job management systems. Work is also underway to incorporate middleware from EGEE<sup>8</sup> in the context of the LCG ARDA<sup>9</sup> project.

The architecture is shown in figure 1.

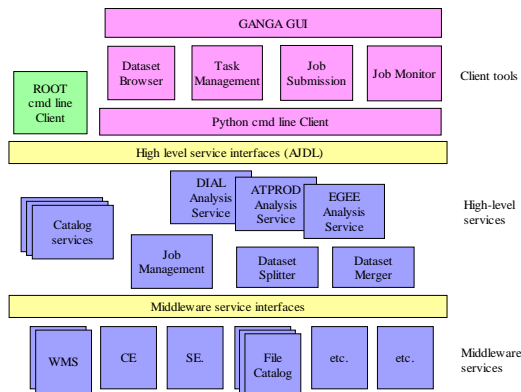


Figure 1: The ADA layered architecture

In addition to the usual identification of middleware services, the system is further decomposed into client and high-level service layers. A key component of the system is the Analysis Job Definition Language, AJDL that specifies the types of objects used to define the interfaces

for the high-level services. The specifications of these objects are extensible to support different requirements within ATLAS and to allow extension to other experiments or realms. High-level services are typically constructed using generic parts of the objects to facilitate reuse of these services.

The separation between client tools and high-level services also allows for independent development of these components.

The content and nature of the AJDL components is under active investigation and is driven by the requirements of both clients and high-level services. The components identified thus far include:

- Dataset describing the data input to and output from processing.
- Transformation describing the operation acting on a dataset. This is decomposed into an installable application and the task providing the user-specified configuration
- Job which allows monitoring and control of an activity performed on or off the grid. The typical example is the application of a transformation to an input dataset to produce and output dataset.

A sophisticated set of client tools and user interface are provided by GANGA using a python binding to AJDL and a python bus to tie the pieces together. In the initial implementation, production activities are carried out by an analysis service based on the ATLAS production system which provides separate executors for each grid flavour. A

separate analysis service based on DIAL provides interactive response for traditional analysis activities like the filling of histograms. A service based on the EGEE middleware will presumably be more scalable and robust and will eventually displace these systems.

The catalogue services are hosted in the ATLAS Metadata Interface system, AMI<sup>10</sup>.

Both high-level and middleware services are initially deployed as independent persistent web services. It is foreseen to move to a grid service model once standards (OGSI, WSRF, ...) have been established.

1 <http://www.usatlas.bnl.gov/ADA>

2 <http://lcg.web.cern.ch/LCG/>

3 <http://www.ivdgl.org/grid2003/>

4 <http://www.nordugrid.org/>

5 <http://ganga.web.cern.ch/>

6 <http://atlas.web.cern.ch/Atlas/GROUPS/SOFTWARE/OO>

7 <http://www.usatlas.bnl.gov/~dladams/dial>

8 <http://egee-intranet.web.cern.ch/egee-intranet>

9 <http://lcg.web.cern.ch/LCG/peb/arda>

10 <http://isnpx1158.in2p3.fr:8180/AMI/index.html>