

A Legacy Executor For ATLAS Data Challenge 2

John Alan Kennedy Ludwig-Maximilians-Universität München, Germany



1. ATLAS Data Challenges

The ATLAS experiment at CERN will begin data taking in 2007 when the LHC is commissioned. From then on data is expected to be recorded at a rate of 2 PetaBytes per year with an additional 1 PetaByte of simulated data produced per year.

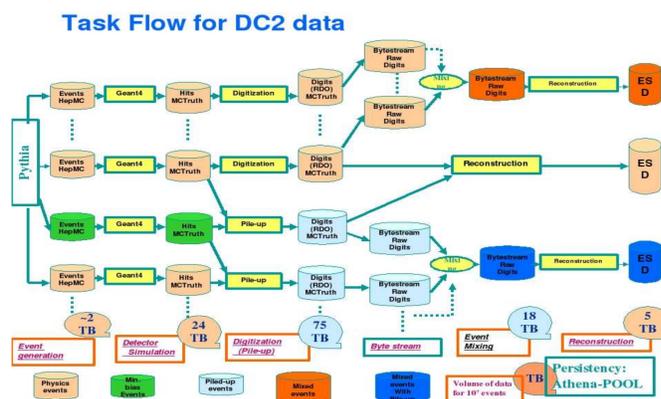
This huge data volume means that it is impossible to perform all the processing and analysis at CERN. ATLAS aims to utilise distributed computing resources in collaborating countries from all around the world.

To aid with the development of this distributed computing infrastructure a series of data challenges were started in 2002. The data challenges allow us to evaluate the ATLAS computing model, the full software suite, the data model and also to ensure the correctness of technical choices made for ATLAS computing.

2. Data Challenge 2

The second data challenge, DC2, began in summer 2004. DC2 focuses on the use of an automated production system and GRID software.

The task of producing simulated data for DC2 was split into several subtasks such as event generation, simulation and reconstruction.

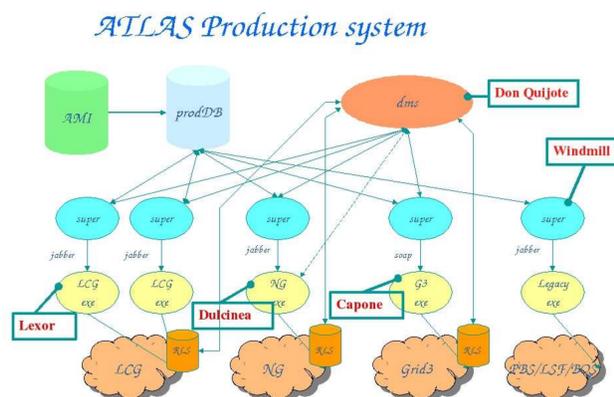


Different paths can be taken through the production depending on which final simulated data is required. This modularity is very useful and powerful within a production system.

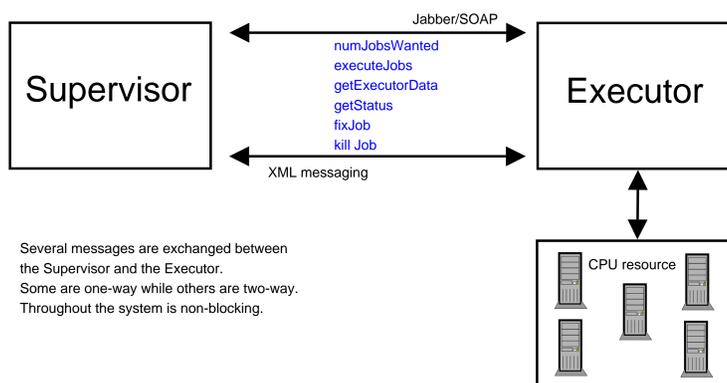
3. The Production System

- Supervisor - management of jobs between several GRIDs.
- Executor - interface to a GRID flavour or batch system.
- DMS - management of data between several GRIDs.

At the core of the production system is the supervisor - executor system. The coupling of a supervisor and an executor forms a bridge which joins together the computing power of the GRIDs and the job definition data and metadata associated with the production jobs.



The supervisor – executor communication is achieved using an XML dialog . Messages are passed via jabber or by use of web services.



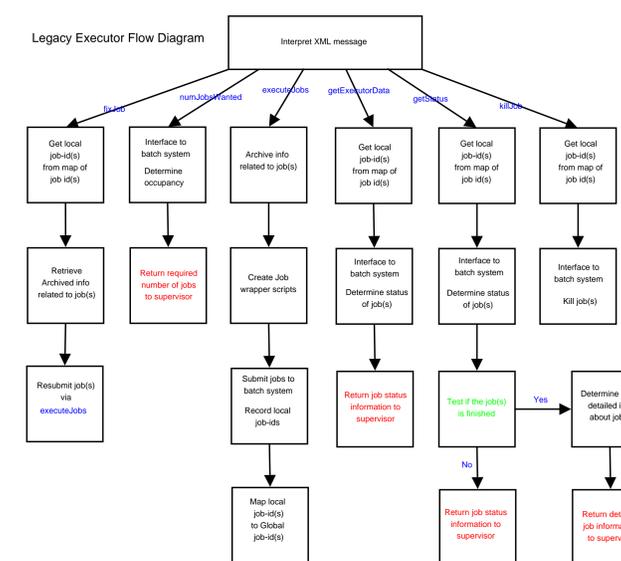
Several messages are exchanged between the Supervisor and the Executor. Some are one-way while others are two-way. Throughout the system is non-blocking.

Data Management in an inter GRID manner is a very important part of the production system. It is imperative that the individual GRID flavours can share files and datasets.

4. Legacy Executor

Despite DC2's strong commitment to using GRID systems a strong case was made for the development of an interface to standard batch systems, this is the Legacy Executor.

The Legacy Executor provides a fallback solution if problems occur with GRID based executors and also allows institutes which are not integrated into a GRID to contribute to DC2. The Legacy Executor may also be used to aid with testing.



Designed to be simplistic, the Legacy Executor makes use of common tools provided within the supervisor framework. The PBS, BQS and LSF batch systems are currently supported.

5. The Future for Legacy

The Legacy Executor was designed with simplicity in mind. Following this, several components of the production system do not exist for the Legacy Executor, for instance RLS. These may or may not be added in the future, but how GRID-like do we want our non-GRID solution to become?

The production of a simple non-GRID executor provides more flexibility to the production system and aids with testing.