Scaling up ATLAS production system for the LHC Run 2 and beyond: project ProdSys2

M. Borodin, K. De, J. Garcia Navarro, D. Golubkov, A. Klimentov, T. Maeno and A. Vaniachine for the ATLAS Collaboration

Argonne National Laboratory, USA; Brookhaven National Laboratory, USA; IFIC, Valencia, Spain; IHEP, Protvino, Russia; NRC Kurchatov Institute, Russia; NRNU MEPhI, Russia; University of Texas, Arlington, USA

The Big Data processing needs of the ATLAS experiment grow continuously, as more data and more use cases emerge. For Big Data processing the ATLAS experiment adopted the data transformation approach, where software applications transform the input data into outputs. In the ATLAS production system, each data transformation is represented by a task, a collection of many jobs, submitted by the ATLAS workload management system (PanDA) and executed on the Grid.

Our experience shows that the rate of tasks submission grows exponentially over the years. To scale up the ATLAS production system for new challenges, we started the ProdSys2 project [1]. PanDA has been upgraded with the Job Execution and Definition Interface (JEDI) [2]. As patterns in ATLAS data transformation workflows are composed of many tasks, a scalable production system framework is needed for template definitions of the many-tasks workflows (Figure 1). These workflows are implemented in the Database Engine for Tasks (DEFT) [3] that generates individual tasks for processing by JEDI.

Figure 1: The Monte Carlo workflow is composed of many steps: generate or configure hard-processes, hadronize signal and minimum-bias (pileup) events, simulate energy deposition in the ATLAS detector, digitize electronics response, simulate triggers, reconstruct data, convert the reconstructed data into ntuples for physics analysis, etc. Outputs are merged and/or filtered as necessary. The real data are processed through the same workflow chain starting from the Reconstruction step.

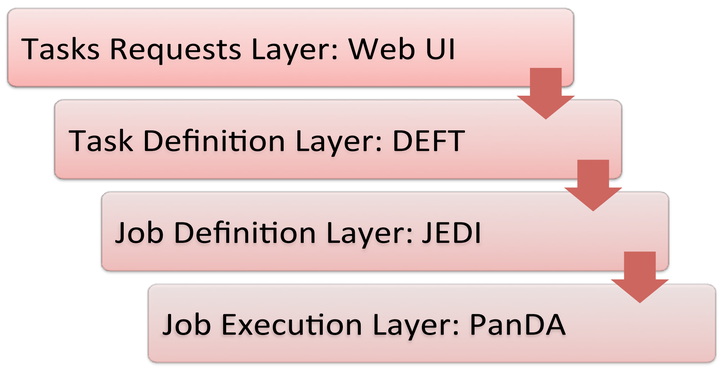

Figure 2: ProdSys2 components.

Figure 2 shows that ProdSys2 is implemented as independent components:
Web UI (Figure 3) for Production Managers and Users provides the interface for task and production request managing and monitoring at the higher level.
Database Engine for Tasks (DEFT): is responsible for formulating the tasks, chains of tasks and also task groups (production request), complete with all necessary parameters. It also keeps track of the state of production requests, chains and their constituent tasks.
Job Execution and Definition Interface (JEDI): is an intelligent component in the panda server to have capability for task-level workload management. Key part of it is 'dynamic' job definition, which highly optimizes resources usage in contrast to the 'static' ' job definition used in ProdSys1. Dynamic job definition in JEDI is also crucial for multi-core, HPC's and other new requirements.
The **BigPanda Monitor** (not shown) monitors all parts of the system.

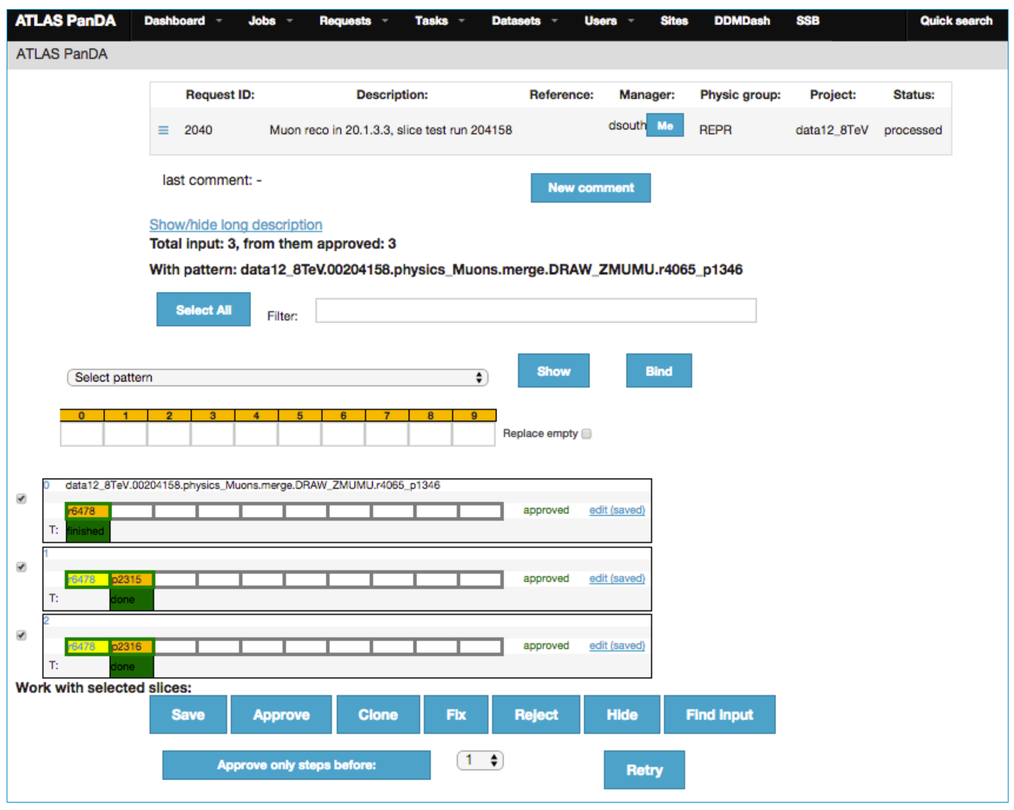

Figure 3: Web UI – user interface for manipulating and monitoring steps in production request.

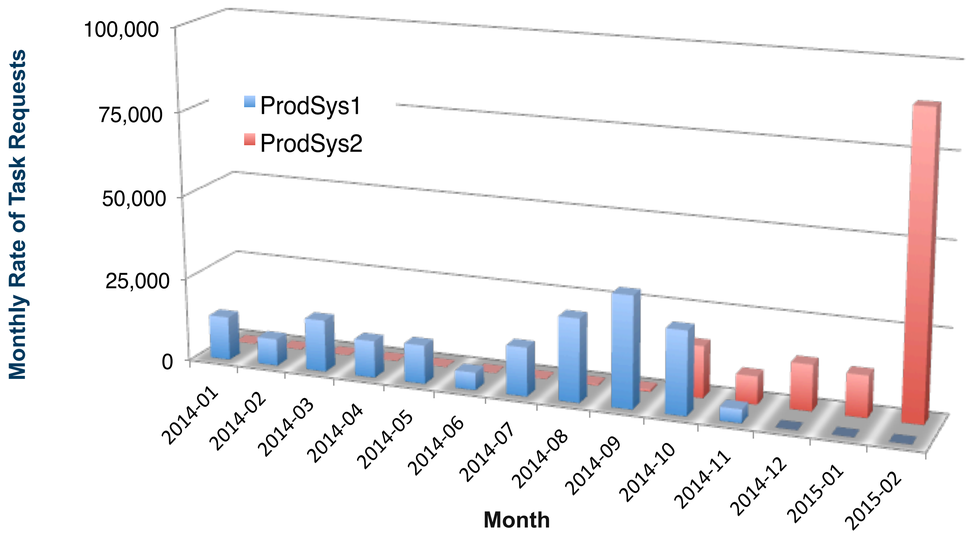

Figure 4: Monthly rate of production tasks.

DEFT is the new system responsible for:

- Submitting production request
  - Checking meta-data and parameters
  - Internal bookkeeping for production requests
- Prepare job and task description to run in JEDI
  - e.g, setting offset for new tasks extending the existing samples
- Production task handling: aborting, fixing
- Post-production
- Automatic handling via JIRA tickets

DEFT is designed to support all workflows used ProdSys1, and also many new workflows which would have been impossible or extremely difficult to manage in ProdSys1. The rate of task processed by ProdSys2 recently exceeded that of ProdSys1 (Figure 4).

Conclusions and next steps: The ATLAS production system fully satisfies the requirements of ATLAS data reprocessing, simulations, and production by physics groups. The LHC shutdown provided an opportunity for enhancing the production system, whilst retaining those core capabilities most valued by production managers. As the ATLAS experiment continues optimising the use of Grid computing resources in preparation for the LHC data taking in 2015, the next generation production system is integrated with other ATLAS Distributed Computing layers – ATLAS Metadata Interface [4] and Distributed Data Management system Rucio [5]. Major workflows are validated and in production for physics analysis and other ATLAS main activity areas - Trigger and Data Preparation.

Acknowledgements: We wish to thank all our colleagues who contributed to ATLAS Data processing activities. This work was funded in part by the U. S. Department of Energy, Office of Science, High Energy Physics and ASCR Contract No. DE-AC02-98CH10886 and Contract No. DE-AC02-06CH11357. NRC KI team work was funded by the Russian Ministry of Science and Education under Contract №14.Z50.31.0024.

References:

[1] D. Golubkov *et al.* "ATLAS Grid Data Processing: system evolution and scalability" 2012 *J. Phys.: Conf. Ser.* **396** 032049
[2] T. Maeno *et al.* "Evolution of the ATLAS PanDA workload management system for exascale computational science" 2014 *J. Phys.: Conf. Ser.* **513** 032062
[3] K. De *et al.* "Task Management in the New ATLAS Production System" 2014 *J. Phys.: Conf. Ser.* **513** 032078
[4] J. Fulachier *et al.* "Looking back on 10 years of the ATLAS Metadata Interface" 2014 *J. Phys.: Conf. Ser.* **513** 042019
[5] V. Garonne *et al.* "The ATLAS Distributed Data Management project: Past and Future" 2012 *J. Phys.: Conf. Ser.* **396** 032045