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 output. 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.

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.

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: