Overview of ASCR “Big PanDA” Project

PanDA Workshop @ UTA

Alexei Klimentov
Brookhaven National Laboratory

Main topics

- **Introduction**
  - PanDA in ATLAS
    - PanDA philosophy
    - PanDA's success
- **ASCR project "Next generation workload management system for Big Data" – *Big PanDA***
  - Project scope
  - Work packages
  - Status and plans
- **Summary and conclusions**

Alexei Klimentov
BNL/PAS

PanDA in ATLAS

- **The ATLAS experiment at the LHC - Big Data Experiment**
  - ATLAS Detector generates about 1PB of raw data per second – most filtered out
  - As of 2013 ATLAS Distributed Data Management System manages ~140 PB of data, distributed world-wide to 130 of WLCG computing centers
  - Expected rate of data influx into ATLAS Grid ~40 PB of data per year
  - Thousands of physicists from ~40 countries analyze the data
- **PanDA project was started in Fall 2005. Production and Data Analysis system**
  - Goal: An automated yet flexible workload management system (WMS) which can optimally make distributed resources accessible to all users
  - Originally developed in US for US physicists
- **Adopted as the ATLAS wide WMS in 2008 (first LHC data in 2009) for all computing applications**
- **Now successfully manages O(10E2) sites, O(10E5) cores, O(10E8) jobs per year, O(10E3) users**

Alexei Klimentov
BNL/PAS

PanDA Philosophy

- **PanDA Workload Management System design goals**
  - Deliver transparency of data processing in a distributed computing environment
  - Achieve high level of automation to reduce operational effort
  - Flexibility in adapting to evolving hardware, computing technologies and network configurations
  - Scalable to the experiment requirements
  - Support diverse and changing middleware
  - Insulate user from hardware, middleware, and all other complexities of the underlying system
  - Unified system for central Monte-Carlo production and user data analysis
    - Support custom workflow of individual physicists
  - Incremental and adaptive software development

PanDA's Success

- **PanDA was able to cope with increasing LHC luminosity and ATLAS data taking rate**
- **Adopted to evolution in ATLAS computing model**
- **Two leading HEP and astro-particle experiments (CMS and AMS) has chosen PanDA as workload management system for data processing and analysis. ALICE is interested in PanDA evaluation for Grid MC Production and LCF.**
- **PanDA was chosen as a core component of Common Analysis Framework by CERN-IT/ATLAS/CMS project**

PanDA was cited in the document titled "Fact sheet: Big Data across the Federal Government" prepared by the Executive Office of the President of the United States as an example of successful technology already in place at the time of the "Big Data Research and Development Initiative" announcement

Evolving PanDA for Advanced Scientific Computing

- **Proposal titled "Next Generation Workload Management and Analysis System for BigData" – Big PanDA was submitted to ASCR DoE in April 2012.**
- **DoE ASCR and HEP funded project started in Sep 2012.**
  - Generalization of PanDA as meta application, providing location transparency of processing and data management, for HEP and other data-intensive sciences, and a wider exascale community.
    - Other efforts
      - PanDA : US ATLAS funded project (Sep 3, talks)
      - Networking : Advance Network Services (ANSE funded project, S.McKee talk)
- **There are three dimensions to evolution of PanDA**
  - Making PanDA available beyond ATLAS and High Energy Physics
  - Extending beyond Grid (Leadership Computing Facilities, Clouds, University clusters)
  - Integration of network as a resource in workload management

Next Generation Workload Management and Analysis System for Big Data

Program Announcement title and number:
Scientific Collaboration at Extreme Scale - Lab 12-695

April 27, 2012

BROOKHAVEN NATIONAL LABORATORY
Alexei Klimentov, Lead PI
Group Leader, Physics Application Software
Physics Department
Upton, NY 11973
Telephone: (631) 344-7855
Fax Number: (631) 344-5078
E-Mail: aak@bnl.gov

Sergey Panitkin, Ph.D.
Senior Software Engineer
Physics Department
Upton, NY 11973
Telephone: (631) 344-7739
Fax Number: (631) 344-5078
E-Mail: panitkin@bnl.gov

Torre Wenaus, Ph.D.
Senior Physicist
Physics Department
Upton, NY 11973
Telephone: (631) 344-4755
Fax Number: (631) 344-5078
E-Mail: wenaus@bnl.gov

Dantong Yu, Ph.D.
Group Leader
Computing Science Center
Upton, NY 11973
Telephone: (631) 344-3042
Fax Number: (631) 344-5751
E-Mail: dtyu@bnl.gov

The University of Texas at Arlington
Kaushik De, Professor of Physics (kaushik@uta.edu)
Gergely Zaruba, Associate Professor of Computer Science Engineering
(zaruba@uta.edu)

Argonne National Laboratory
Alexandre Vaniachine, Software Engineer (vaniachine@anl.gov) (Collaborator)

Official Signing for Laboratory
Thomas Ludlam, Chair
Physics Department

“Big PanDA” Work Packages

- **WP1 (Factorizing the core): Factorizing the core components of PanDA to enable adoption by a wide range of exascale scientific communities (K.De)**
- **WP2 (Extending the scope): Evolving PanDA to support extreme scale computing clouds and Leadership Computing Facilities (S.Panitkin)**
- **WP3 (Leveraging intelligent networks): Integrating network services and real-time data access to the PanDA workflow (D.Yu)**
- **WP4 (Usability and monitoring): Real time monitoring and visualization package for PanDA (T.Wenaus)**

There are many commonalities with what we need for ATLAS

BigPanDA work plan

- **3 years plan**
  - Year 1. Setting the collaboration, define algorithms and metrics
  - Year 2. Prototyping and implementation
    - 2014 is ultimately important year
  - Year 3. Production and operations

"Big PanDA" project will help to generalize PanDA and to use it beyond ATLAS and HEP

- *While being careful not to allow distractions from ATLAS PanDA needs ie the effort has to be incremental over that required by ATLAS PanDA (which itself has a substantial todo list)*
- *We work very close with PanDA core software developers (Tadashi and Paul)*

Alexei Klimentov
BNL/PAS

“Big PanDA” first steps. Forming the team

- **Funding started Sep 1, 2012**
  - Hiring process was completed in May-June 2013. Development team is formed
    - *Sergey Panitkin* (BNL) 0.5 FTE starting from Oct 1, 2012
      - HPC and Cloud Computing
    - *Jaroslava Schovancova* (BNL) 1 FTE starting from June 3, 2013
      - PanDA core software
    - *Danila Olejnik* (UTA) 1 FTE starting from May 7, 2013
      - HPC and PanDA core software
    - *Artem Petrosyan* (UTA) 0.5 FTE starting from May 7, 2013
      - Intelligent networking and monitoring
    - There is a commitment from Dubna to support Artem and Danila for 1-2 more years

“Big PanDA”. Status. WP1

- **WP1 (Factorizing the core)**
  - Evolving PanDA pilot
    - Until recently the pilot has been ATLAS specific, with lots of code only relevant for ATLAS
      - To meet the needs of the Common Analysis Framework project, the pilot is being refactored
    - Experiments as plug-ins
      - Introducing new experiment specific classes, enabling better organization of the code
      - E.g. containing methods for how a job should be setup, metadata and site information handling etc, that is unique to each experiment
      - CMS experiment classes have been implemented
    - Changes are being introduced gradually, to avoid affecting current production

P.Nilsson ’s talk

“Big PanDA”. Status. WP1. Cont’d

- **WP1 (Factorizing the core)**
  - PanDA instance @EC2
    - Back to mySQL
    - VO independent
    - It will be used as a test-bed for non-LHC experiments
      - » PanDA Instance with all functionalities is installed and running at EC2. Database migration from Oracle to MySQL is finished. The instance is VO independent.
        - » LSST MC production is the first use-case for the new instance
    - Next step will be refactoring PanDA monitoring package

J.Schovancova ’s talk

“Big PanDA”. Status. WP2

- **WP2 (Extending the scope)**

Google Compute Engine (GCE) preview project

- Google allocated additional resources for ATLAS for free
  - ~5M cpu hours, 4000 cores for about 2 month, (original preview allocation 1k cores)

– Resources are organized as HTCondor based PanDA queue
  - Centos 6 based custom built images, with SL5 compatibility libraries to run ATLAS software
  - Condor head node, proxies are at BNL
  - Output exported to BNL SE

– Work on capturing the GCE setup in Puppet

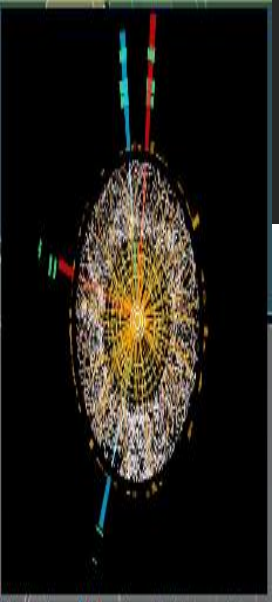
- Transparent inclusion of cloud resources into ATLAS Grid
- The idea was to test long term stability while running a cloud cluster similar in size to Tier 2 site in ATLAS
- Intended for CPU intensive Monte-Carlo simulation workloads
- Planned as a production type of run. Delivered to ATLAS as a resource and not as an R&D platform.
- We also tested high performance PROOF based analysis cluster

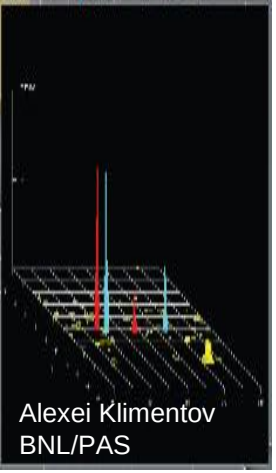
S.Panitkin’s talk

“Big PanDA”. Status. WP2. Cont’d

PanDA project on ORNL Leadership Computing Facilities

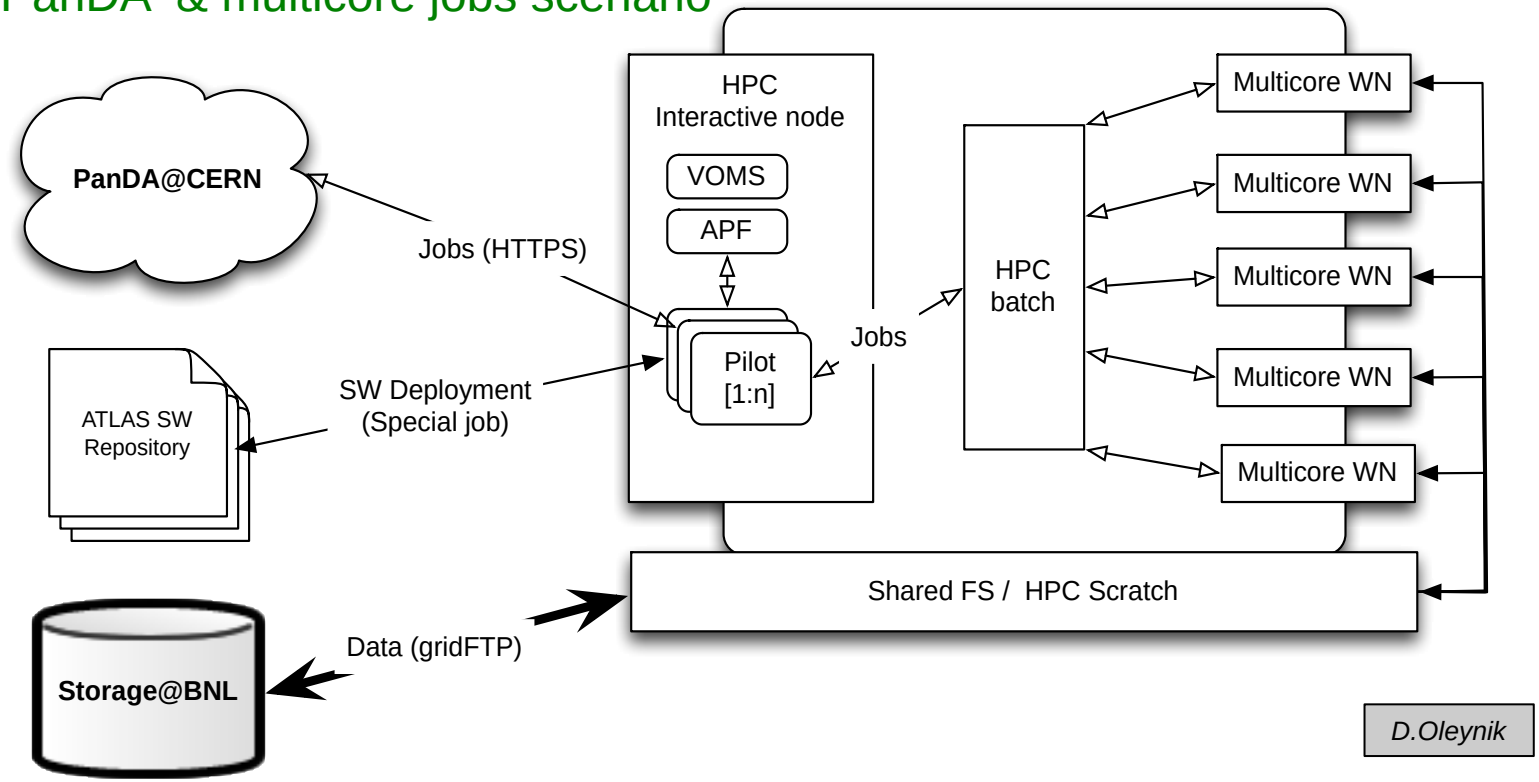
- Get experience with all relevant aspects of the platform and workload
  - job submission mechanism
  - job output handling
  - local storage system details
  - outside transfers details
  - security environment
  - adjust monitoring model
- Develop appropriate pilot/agent model for Titan
- MC generators will be initial use case on Titan
  - Collaboration between ANL, BNL, ORNL, SLAC, UTA, UTK
  - Cross-disciplinary project - HEP, Nuclear Physics , High-Performance Computing





“Big PanDA”. Status. WP2. Cont’d

PanDA & multicore jobs scenario



PanDA/OLCF meeting in Knoxville. Aug 9

- *PanDA deployment at OLCF was discussed and agreed, including AIMS project component*
- *Cyber-Security issues were discussed both for the near and longer term.*
- *Discussion with OLCF Operations*
- *Payloads for TITAN CE (followed by discussion in ATLAS)*

“Big PanDA”. Status. WP2. Cont’d

ATLAS PanDA Coming to Oak-Ridge Leadership Computing Facilities

- ATLAS uses sophisticated Production ANd Data Analysis (PanDA) workload management system to optimize data production and availability on the GRID.
- Project underway now to setup and tailor PanDA agent at OLCF.
- This pioneers connection of Titan to the LHC/OSG GRID.

Slide from Ken Read

“Big PanDA”. Status. WP3

WP3 (Leveraging intelligent networks)

- **Network as resource**
  - Optimal site selection should take network capability into account
  - Network as a resource should be managed (i.e. provisioning)
- **WP3 is synchronized with two other efforts**
  - US ATLAS funded, primarily integrate FAX with PanDA
  - ANSE funded (Shawn’s talk)

Networking database tables schema is finalized (for source-destination matrix) and implemented. Networking throughput performance and P2P statistics collected by different sources such perfsonar, Grid sites status board, Information systems are continuously exported to PanDA database. Tasks brokering algorithm for discussion during this workshop.

A.Petrosyan talk

Summary and Conclusions

- **The ATLAS experiment Distributed Computing and Software performance was a great success in LHC Run 1**
  - The challenge how to process and analyze the data and produce timely physics results was substantial, but at the end resulted in a great success
- **PanDA WMS and team played a vital role during LHC Run 1 data processing, simulation and analysis**
- **ASCR gave us a great opportunity to evolve PanDA beyond ATLAS and HEP and to start "Big PanDA " project**
- **Project team was set up.**
- **Progress in many areas : networking, VO independent PanDA instance, cloud computing, HPC**
  - The work on extending PanDA to Leadership Class Facilities has a good start
  - Large scale PanDA deployments on commercial clouds are already producing valuable results
- **Strong interest in the project from several experiments (disciplines) and scientific centers to have a joined project.**
- **Plan of work and coherent PanDA software development for discussion in this workshop**

Acknowledgements

- **Many thanks to K.De, B.Kersevan, P.Nilsson, D.Oleynik, S.Panitkin, K.Read for slides and materials used in this talk**