PanDA & BigPanDA

Kaushik De

Univ. of Texas at Arlington

BigPanDA Workshop, CERN
October 21, 2013

Introduction

- ## PanDA genesis
  - WMS (Workload Management System) for ATLAS applications running on distributed computing resources for ~8 years
  - New computing paradigm in HEP
  - Provides access to WLCG sites with high efficiency for Central Production, Group Production, and User Production and Analysis
 - Extraordinary PanDA scalability at the LHC

- ## PanDA evolution
  - Success in ATLAS has sparked interest among other communities
  - Any organization with large data volumes, distributed resources and a large user base can benefit from PanDA WMS
  - During past year new PanDA collaborators, and new sources of funding have led to the question – what is the future of PanDA?

What can PanDA do?

- **PanDA can manage complex workflows for:**
    - Large data volume – hundreds of petabytes
    - Distributed resources – hundreds of computing centers worldwide
    - Collaborative work – thousands of scientific users
- **PanDA is built on two key concepts:**
    - Central job queue for distributed resources
    - Pilot based job execution environment

Jobs/week Completed by PanDA

Completed jobs
208 Weeks from Week 39 of 2009 to Week 39 of 2013
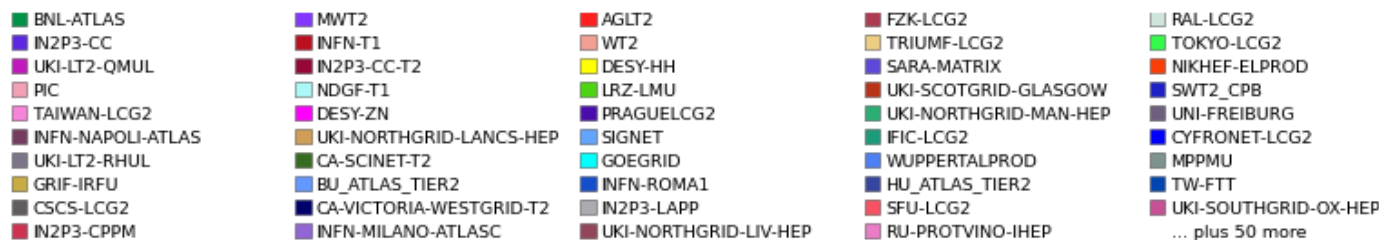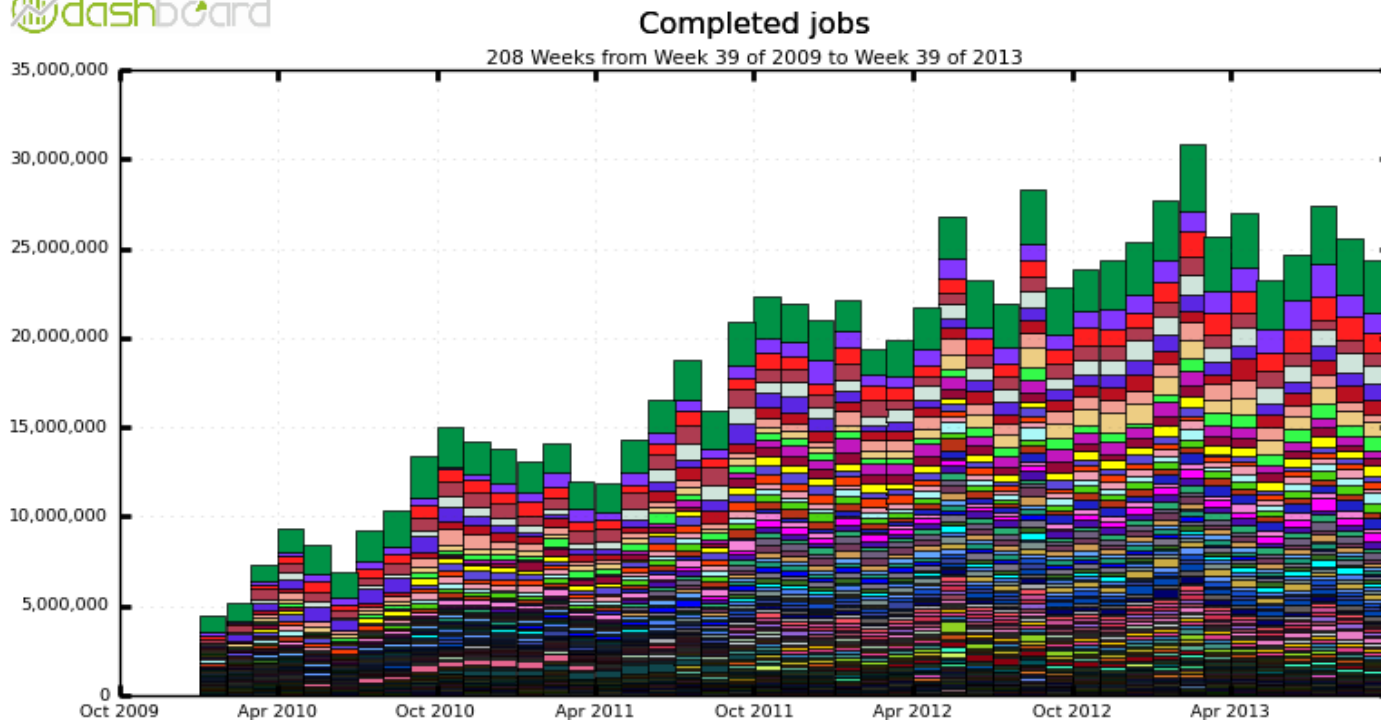
Maximum: 30,811,075 , Minimum: 0.00 , Average: 18,182,748 , Current: 24,387,364

Current scale – 25 million jobs completed monthly at ~hundred sites

Recent Meetings on PanDA Evolution

- A series of recent meetings have been held to define the scope of future directions for PanDA -> BigPanDA

- Workshop at BNL June 4-6, 2013
  https://indico.bnl.gov/conferenceDisplay.py?confId=612

- Workshop at UTA September 3-4, 2013
  https://indico.cern.ch/conferenceTimeTable.py?confId=266191

- Focused meetings at ORNL, LBNL (EsNet) and others

- This meeting at CERN
  https://indico.cern.ch/conferenceDisplay.py?confId=277339

Scope of BigPanDA

- **What is BigPanDA?**
    - Is it improved PanDA + ProdSys2 after LS1 shutdown?
    - Is it PanDA for CMS/AMS?
    - Is it DoE funding for BigData usage of PanDA WMS?
    - Is it PanDA integration with networking?
    - Is it PanDA moving to HPC and other platforms?
    - Is it PanDA for non-HEP sciences (eg. BioInfomatics)?
    - Is it something else?

- **BigPanDA is all of the above – not just PanDA for BigData!**

- **At this meeting we will hear from new stakeholders in PanDA – the people who are at the forefront of BigPanDA**

- **We will also hear about complimentary visions**

The Growing PanDA Ecosystem

- ## ATLAS PanDA
  - US ATLAS, CERN, UK, DE, NorduGrid, Dubna, Protvino, OSG …

- ## ASCR BigPanDA
  - DoE funded project at BNL, UTA – 3 years

- ## ANSE PanDA
  - NSF funded network project - CalTech, Michigan, Vanderbilt, UTA

- ## HPC and Cloud PanDA

- ## Taiwan PanDA

- ## CMS PanDA

- ## AliEn PanDA

- ## (MegaPanDA) …

DoE ASCR Project

- **"Next Generation Workload Management and Analysis System for Big Data"**

- **3 year DoE ASCR funding**
  - Lead Institution: Brookhaven National Laboratory
  - Lead PI: Alexei Klimentov
  - Principal Investigators:
    - Brookhaven National Laboratory: Alexei Klimentov, Sergei Panitkin, Torre Wenaus, Dantong Yu
    - Argonne National Laboratory: Alexandre Vaniachine
    - The University of Texas at Arlington: Kaushik De, Gergely Zaruba
  - **New hires**
    - Jaroslava Schovancova @ BNL, Danila Oleynik @ UTA, Artem Petrosyan @ UTA, Mikhaill Titov @ UTA

ASCR BigPanDA Work Plan

- **WP1 (Factorizing the core):**
  - Factorizing the core components of PanDA to enable adoption by a wide range of exascale scientific communities

- **WP2 (Extending the scope):**
  - Evolving PanDA to support extreme scale computing clouds and Leadership Computing Facilities

- **WP3 (Leveraging intelligent networks):**
  - Integrating network services and real-time data access to the PanDA workflow

- **WP4 (Usability and monitoring):**
  - Real time monitoring and visualization package for PanDA

WP1 – Factorizing the Core

- ## PanDA server factorizations
  - Modular structure to support multiple organizations
  - Ease of packaging new releases
  - Work ongoing

- ## PanDA pilot factorizations
  - Started >one year ago, in collaboration with CERN IT
  - Useful for Common Analysis Framework development

- ## DB backend factorizations
  - Oracle DB backend is used for ATLAS
  - MySQL backend now running on EC2 PanDA server

Future Packaging

➢ **Current**
- tarball made from the main branch
- All experiment-specific modules in the same package

➢ **Plans**
- tag + rpm
- Split to core and experiment packages
  - E.g., panda-server, panda-atlas, panda-cms, panda-ams, etc
- Specify package.module names in config files
  - No hard coding

From Tadashi Maeno

PanDA Pilot Refactorization

- Ongoing effort to improve overall code quality and efficiency

- Refactorization necessary for several reasons
  - Adaptation of PanDA Pilot for non-ATLAS use
  - Complex feature requests
  - Reduce/remove legacy code (in general: refactoring means reorganization, not just rewriting)
  - Reduce complexity to make it easier for code contributors

- Trivial vs non-trivial pilot refactorizations
  - Site Movers contain repeated code fragments (trivial)
  - Pilot code needs to support users other than ATLAS (non-trivial)

- "Only" 47k LOC, so not a hopeless effort..

Refactorization Status

- General pilot refactoring effort is progressing well

- New pilot versions have minor or major changes related to refactoring on basically every check-in

- PanDA Pilot is now effectively a "generic" pilot that is "not" ATLAS specific
  - NB 1: this means that the major changes are done, there are plenty of minor changes that remain (code relevant only for ATLAS is still present but is of little or no concern to other users)
  - NB 2: Some ATLAS specific code is deeply integrated with specific modules, namely the site movers and the DQ2 tracing code. Cannot easily be refactored

P. Nilsson   PanDA Workshop @ UTA September 2013

WP2 – Extending PanDA to HPC

- Extensive progress during past 6 months

- First demonstration: running ATLAS jobs on Google (GCE)

- Second demonstration: Joint project with ORNL – Oakridge Leadership Computing Facility

  - No show stoppers, basic functions working

  - Expect ATLAS payload running soon at Titan

PanDA batch queue on GCE II

- We ran for about 8 weeks (2 weeks were planned for scaling up)

- Very stable running on the Cloud side. GCE was rock solid.

- Most problems that we had were on the ATLAS side.

- We ran computationally intensive jobs

    - Physics event generators, Fast detector simulation,, Full detector simulation

- Completed 458,000 jobs, generated and processed about 214 M events
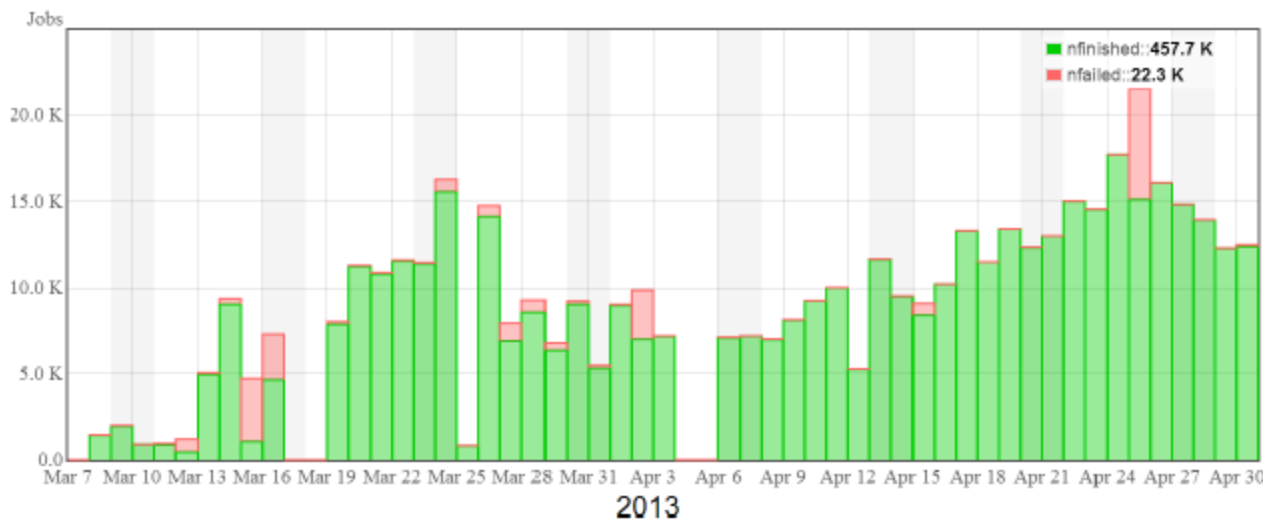
From
Sergey Panitkin

PanDA Backfill

- PanDA has potential to *generate* 300 M Titan hours in 2014 and again in 2015. Estimated to represent between $10 M to $15 M per year worth of computing.

- Jobs are *naturally parallel* (no longer *embarrassed!*) and can be short. Can backfill *arbitrary* amount of Titan queue.

- Exploring whether PanDA pilot can receive more extensive information concerning schedule than available via "qstat".

- Multiple "payloads" now at varying stages of functionality. "Payloads" need VALIDATION of performance and correctness.

- Testing possibility of remotely-compiled binaries running with associated shared libraries provided by CERN Virtual Machine Filesystem (CVMFS), presumably only appropriate for backfill jobs. This can greatly reduce the effort for validation. Need to check possible performance penalties and scaling.

- Question: ATLAS code validation? Timescale?

From Ken Read

Available PanDA Payloads on Titan

- Multiple shovel-ready HEP and NP codes potentially available for Titan backfill.

- These codes now run on Titan compute nodes:
  - ROOT
  - Geant3, Geant4
  - AliRoot
  - CL-SHASTA (see below)
  - and more...

- CVMFS can deliver code (and binaries) on the external login nodes (using FUSE).

- For the longer term, jobs which use a node fully (multi-threaded with some GPU acceleration) and scale well, may be very competitive at the Leadership Level, especially considering the scientific importance.

From Ken Read

HPC Testing at ORNL LCF
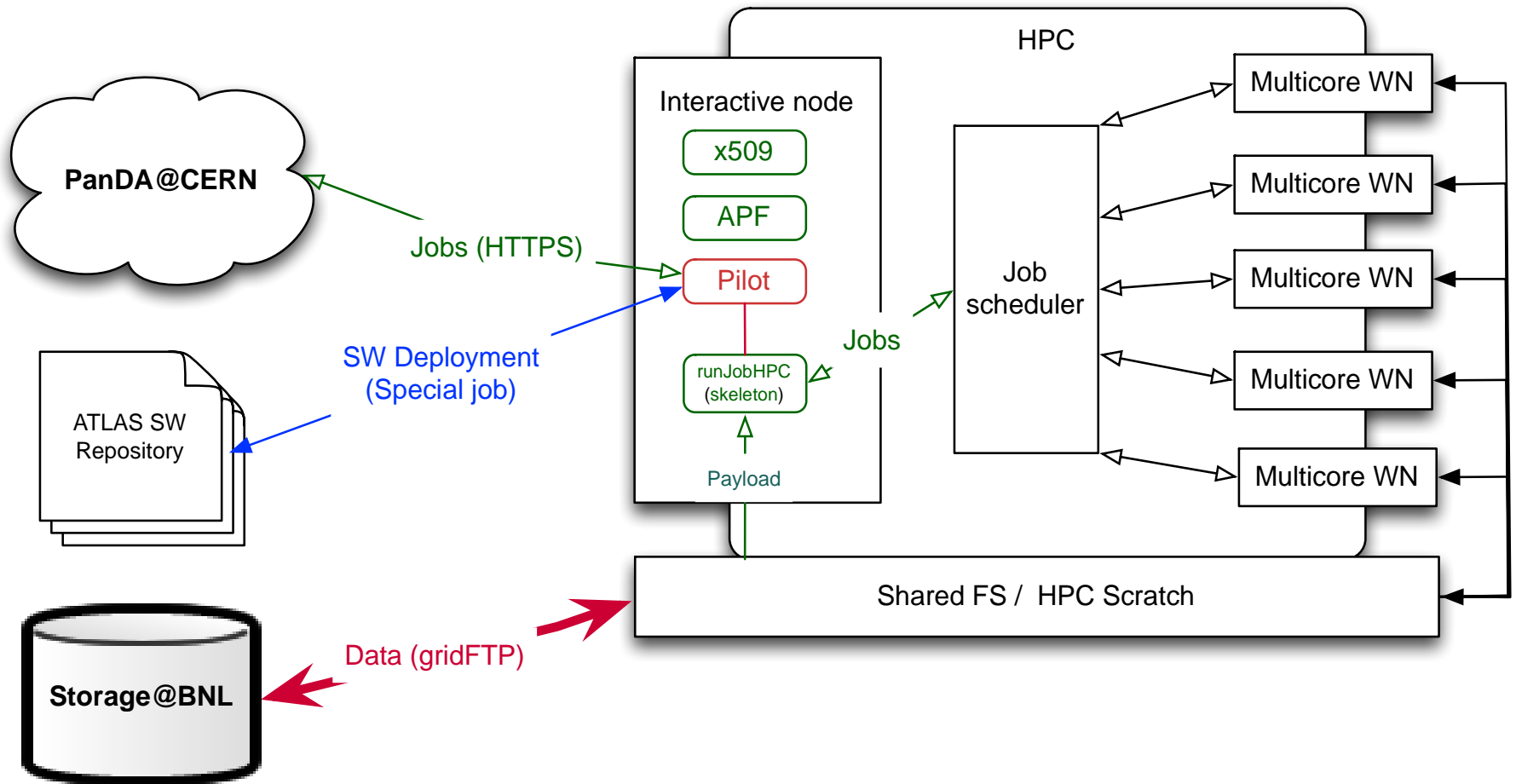
From Danila Oleynik

WP3 – Intelligent Networking

- **Much recent progress – Armen's talk later in the morning**
    - Network data automatically imported to PanDA
    - Code for initial use cases ready for testing

- **Meeting with ESNET two weeks ago**
    - Action items being prepared for joint activities

- **ASCR and ANSE projects are working together**
    - Rich community expertise
    - Network Element (NE) will soon be an equal partner to CE and SE

WP4 - Monitoring

- New framework for rapid development of new monitoring

- Common code for PanDA, ProdSys2

- Easy packaging and deployment for other experiments

- Newest area of development – pages coming soon

Synergies

- ## ATLAS PanDA
  - Core PanDA effort, primarily for ATLAS, but benefits the entire PanDA ecosystem

- ## ASCR
  - While ASCR funding is for use beyond ATLAS, clear benefit to ATLAS in core development, sustainability, networking
  - ASCR team works closely with ATLAS PanDA team

- ## ANSE
  - Networking – 50% of benefits/directed to ATLAS

- ## CERN/European/Asian PanDA efforts
  - A lot of new developments, new partners

Conclusion

- Baby PanDA is growing up to become BigPanDA