Evolution of the ATLAS PanDA Workload Management System for Exascale Computational Science

T. Maeno, K. De, A. Klimentov,
P. Nilsson, D. Oleynik, S. Panitkin,
A. Petrosyan, J. Schovancova,
A. Vaniachine, T. Wenaus, D. Yu
on behalf of the ATLAS collaboration

Brookhaven National Laboratory, USA
University of Texas at Arlington, USA
Argonne National Laboratory, USA

CHEP2013, Amsterdam, Netherlands, Oct 14 2013

Outline

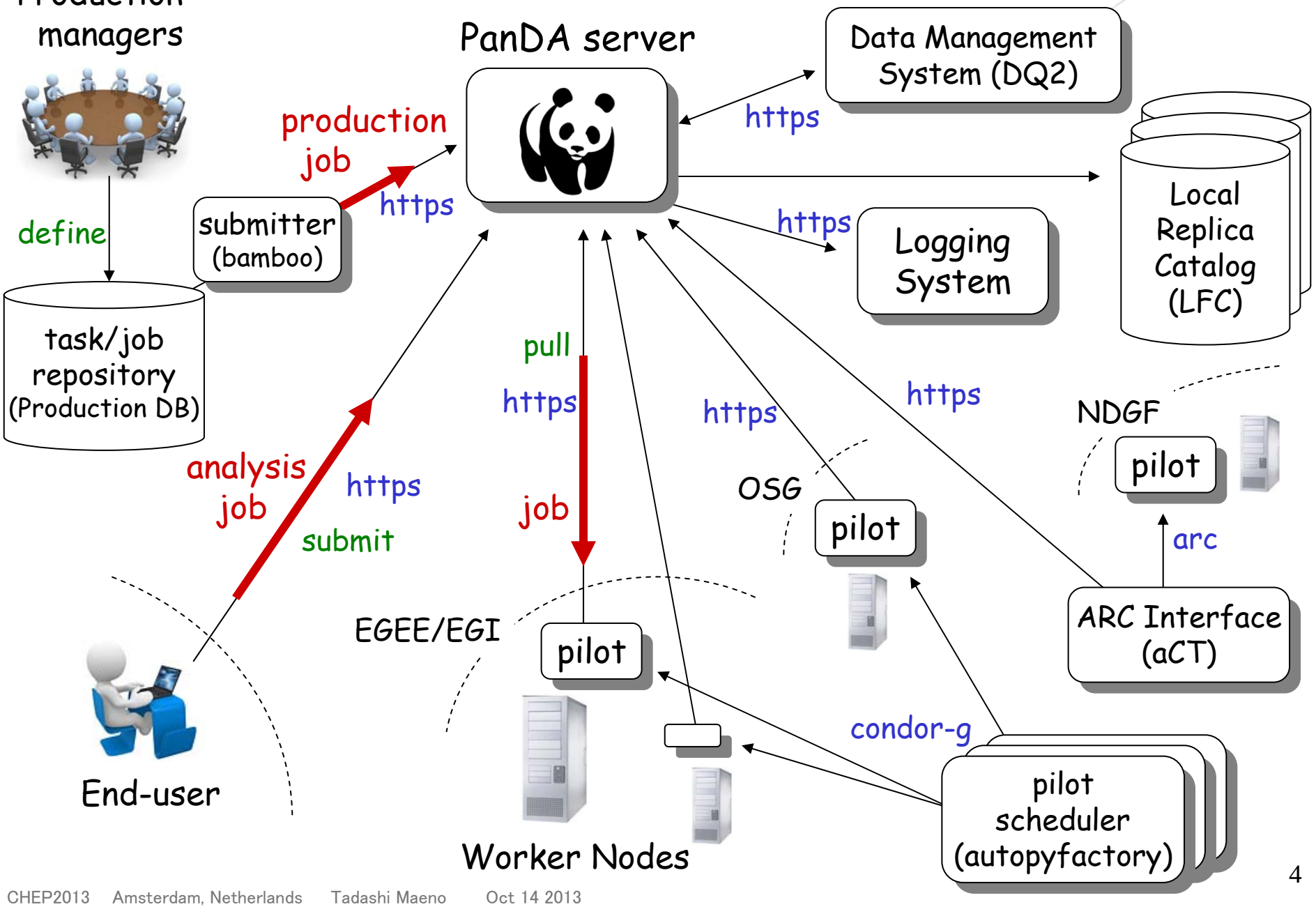
- Introduction
- Overview of the BigPanDA project
- Current status and plans
- Conclusions

Introduction

BROOKHAVEN NATIONAL LABORATORY

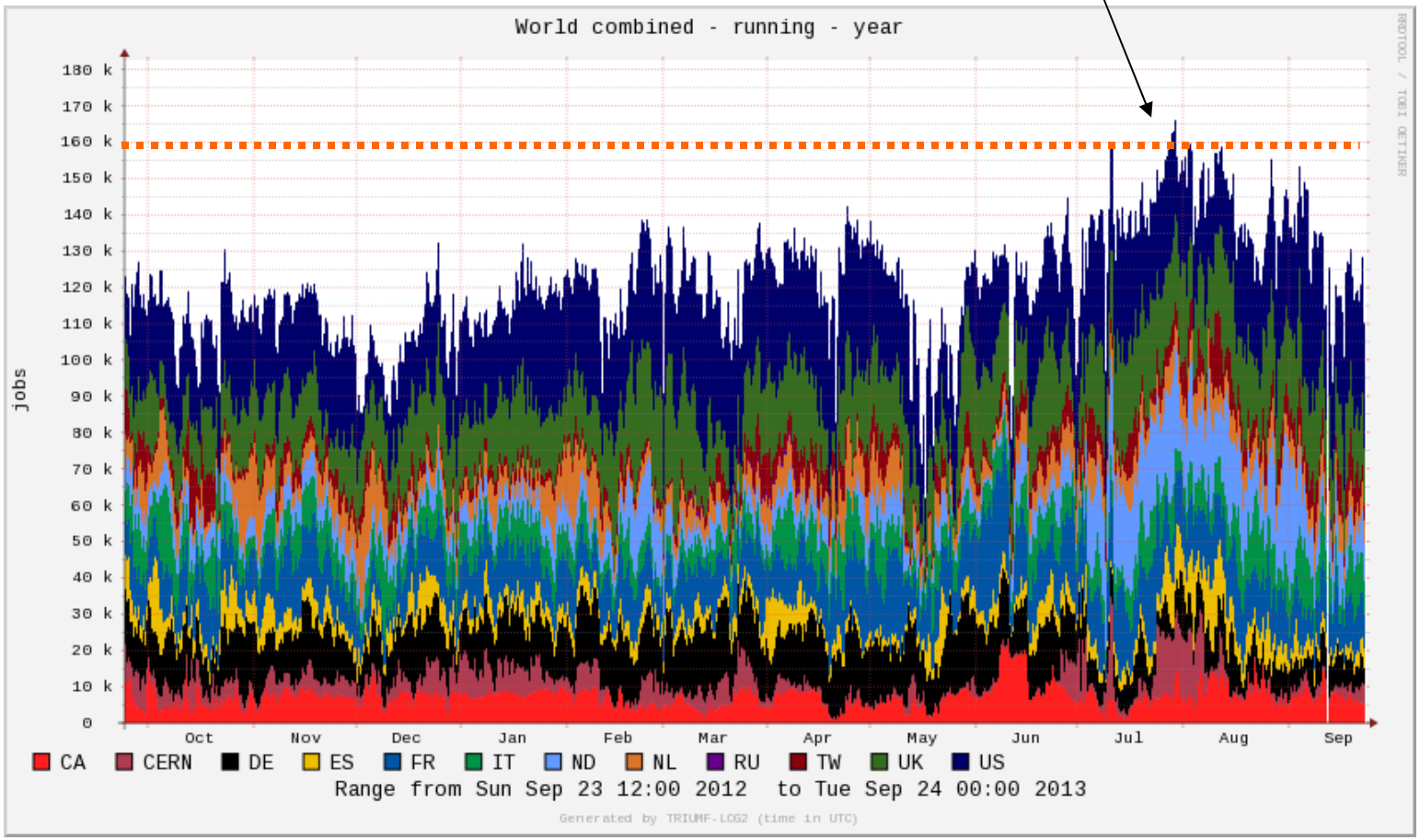
- PanDA = Production and Distributed Analysis Sytem
  - Designed to meet ATLAS production/analysis requirements for a data-driven workload management system capable of operating at LHC data processing scale
- Highly automated, low operational manpower, integrated monitoring system
- History
  - Aug 2005: Project started, Sep 2005: The first prototype, Dec 2005: Production in US ATLAS,
    2008: Adopted as the workload management system for the entire ATLAS collaboration, Recent: AMS and CMS have deployed own PanDA instances
- Performed well during LHC Run 1 for data processing, simulation and analysis, while actively evolving to meet rapidly changing physics needs
  - Successfully managing more than 100 sites, about 5 million jobs per week, and about 1500 users

PanDA System



ATLAS Jobs Concurrently Running ’12 –’13

~ 160 k running jobs at peak



Overview of BigPanDA project

BigPanDA project

BROOKHAVEN NATIONAL LABORATORY

Evolving PanDA for Advanced Scientific Computing

- The interest in PanDA by other big data sciences provided the primary motivation to generalize the PanDA system
- A project to extend PanDA as meta application, providing location transparency of processing and data management, for HEP and other data-intensive sciences, and a wider exascale community
- 3 FTE for 3 years from 2012
- Three dimensions to evolution
  - Making PanDA available beyond ATLAS and High Energy Physics
  - Extending beyond Grid (Leadership Computing Facilities, Clouds, University clusters)
  - Integrating network as a resource in workload management

Work Plan

- 3 year plan
  - Year 1. Setting the collaboration, define algorithms and metrics
    - Hiring process was completed in June 2013
    - Development team is formed (3 FTE)
  - Year 2. Prototyping and implementation
  - Year 3. Production and operations
- 4 work packages
  - WP1 : Factorizing the core
  - WP2 : Extending the scope
  - WP3 : Leveraging intelligent networks
  - WP4 : Usability and monitoring

Work Packages 1/4

BROOKHAVEN NATIONAL LABORATORY

- WP1 (Factorizing the core)
  - Factorizing the core components of PanDA to enable adoption by a wide range of exascale scientific communities
  - To take the core components of PanDA and package them in an experiment neutral package
    - General components + customizable layers
    - The experiment specific layers as plug-ins + configuration files
  - Advanced features will have sensible defaults and can be turned on for demanding applications

Work Packages 2/4

- WP2 (Extending the scope)
  - Evolving PanDA to support extreme scale computing clouds and Leadership Computing Facilities
    - Historically HEP community used the Grid infrastructure for large scale production
      - E.g., Leadership Computing Facilities were not used very extensively
  - Adding extra resources
  - Further expansion of the potential user community and the resources available to them

Work Packages 3/4

BROOKHAVEN NATIONAL LABORATORY

- WP3 (Leveraging intelligent networks)
  - Many of the research efforts into dynamic network provisioning, quality of service and traffic management
  - Integrating network services and real-time data access to the PanDA workflow
    - Integrating these services within existing and evolving infrastructures
      - E.g., PanDA currently does not interface with any of the advanced network provisioning technologies available
    - Automating the discovery and use of such services transparently to the scientists
      - E.g., instead of exposing intelligent network services directly to scientists, PanDA could be updated to directly interface with them without any involvement of scientists

Work Packages 4/4

BROOKHAVEN NATIONAL LABORATORY

- WP4 (Usability and monitoring)
  - The PanDA monitoring has been identified to require a special effort for factorization and generalization
    - Each experiment has own workflow and visualization needs
  - To design a generic PanDA monitor browser view and skeleton from which experiment (ATLAS and other) browser views are derived customizations
  - To provide generic components and APIs for user communities to easily implement and customize their monitoring views and workflow visualizations

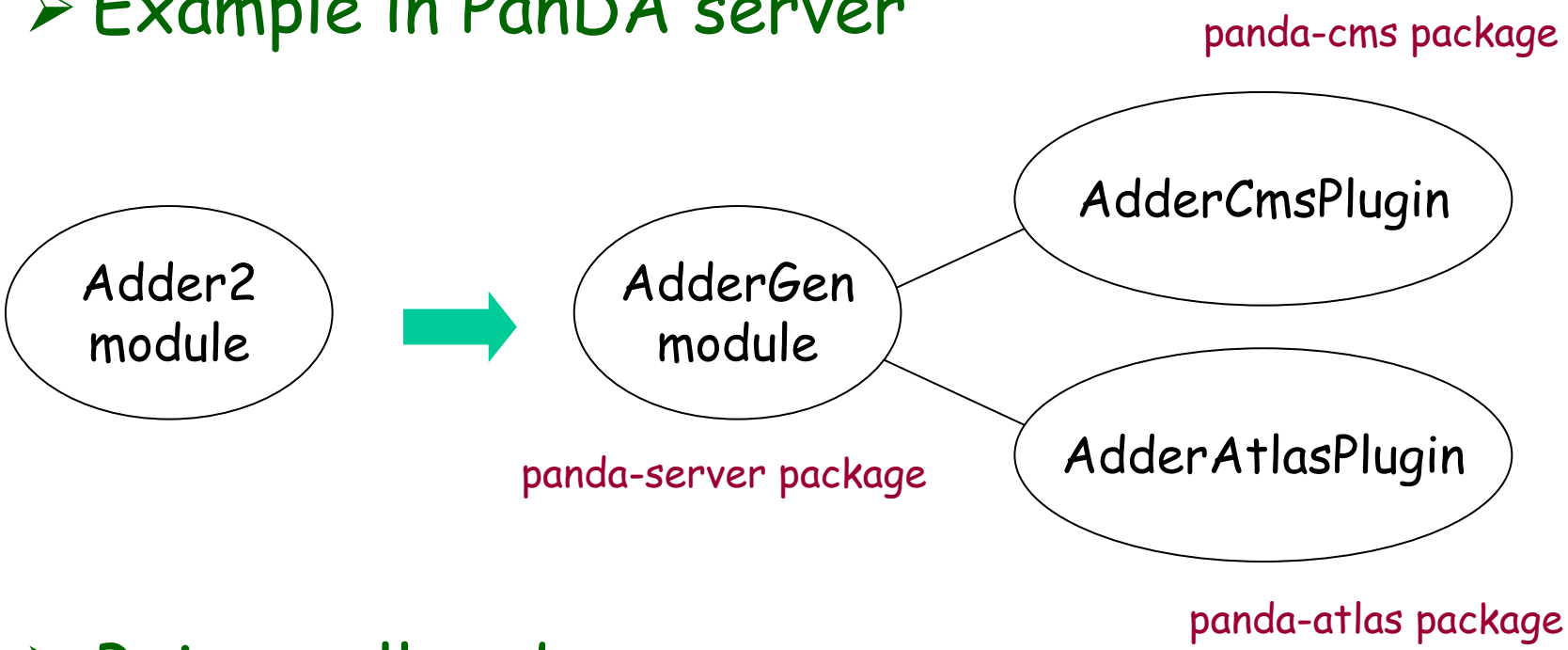
BigPanDA and ATLAS

- ATLAS remains the largest user community and provides the primary focus to PanDA development although PanDA is not only for ATLAS any longer
- Working very close with ATLAS PanDA developers
- One set of software packages for all experiments including ATLAS
  - Core packages + experiment-specific plug-ins
- The effort has to be incremental and coherent with many challenging developments for ATLAS

Current Status and Plans

Factorization of Core Components

- Example in PanDA server



- Being well underway
  - Decomposition of experimental code to plug-ins & separate packaging
  - See Paul Nilsson's poster as well

    Nilsson P., *Next Generation PanDA Pilot for ATLAS and Other Experiments*

VO-independent PanDA Server

- PanDA server with MySQL database backend
  - Multiple backend support
    - Oracle or MySQL can be selected in config file
    - Code is being merged to the main branch
- PanDA server @ Amazon EC2
  - Serves experiments which don't want to maintain own PanDA server
  - MySQL backend

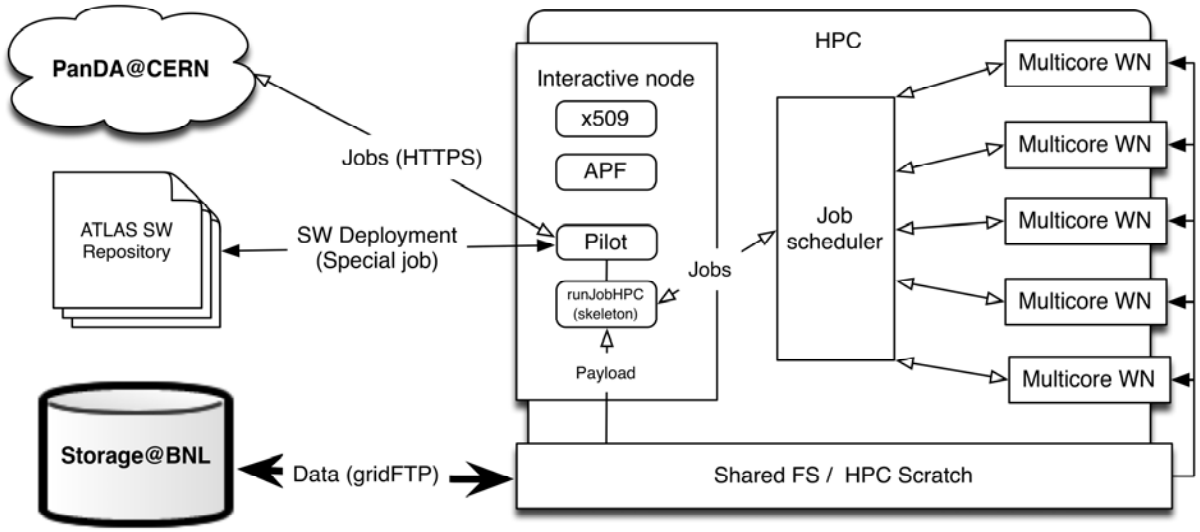
Google Compute Engine (GCE) preview project

- Google allocated resources for ATLAS for free
  - ~5M cpu hours, 4000 cores for ~2 months
- Transparent inclusion of cloud resources into ATLAS grid
  - Resources were organized as HTCondor-based PanDA queue
- Delivered to ATLAS as a production resource and not as an R&D platform
- See Sergey Panitkin's talk for details

  Panitkin S., *ATLAS Cloud Computing R&D*

PanDA + HPC

- In collaboration with ORNL Leadership Computing Facility
  - Get experience with all relevant aspects of the platform and workload
    - job submission, security, storage, monitoring, etc
  - Developing appropriate pilot/agent model for Titan



Leveraging Network

- Synchronized with two other efforts
  - Integration of FAX (Federated Xrootd for ATLAS) with PanDA
  - ANSE project
- Three layered software architecture
  - Collector
    - Collecting network performance information from various sources such as perfsonar, Grid sites status board, FAX, etc
  - AGIS (ATLAS Grid Information System)
    - Information pool
  - Calculator
    - Calculating weights which the brokerage take into account for site selection
- The site selection algorithm in the PanDA brokerage has been improved to consider cost metrics of FAX

Conclusions

- The PanDA system played a key role during LHC Run 1 data processing, simulation and analysis with great success, while actively evolving to meet rapidly changing physics needs
- The interest in PanDA by other big data sciences provided the primary motivation to generalize the PanDA system
- The BigPanDA project gives us a great opportunity to evolve PanDA beyond ATLAS and HEP
- Progress in many areas : networking, VO independent PanDA instance, cloud computing, HPC
- Strong interest in the project from several experiments and scientific centers to have a collaborative project