BROOKHAVEN NATIONAL LABORATORY

PanDA for ATLAS Distributed Computing in the Next Decade

F. H. Barreiro Megino, K. De, A. Klimentov, T. Maeno, P. Nilsson, D. Oleynik, S. Padolski, S. Panitkin, T. Wenaus
on behalf of ATLAS collaboration

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

CHEP2016, San Francisco, USA, Oct 10-14 2016

Introduction

- PanDA = Production and Distributed Analysis System
  - Designed to meet ATLAS production/analysis requirements for a data-driven workload management system capable of operating at LHC data processing scale
  - Performed well for ATLAS in the last decade including the LHC data taking period
- Many new components and features have been delivered to ATLAS before LHC Run2
  - DEFT, JEDI, Dynamic job definition, event service, new monitoring, and so on
- The system has revealed great improvements in LHC Run2 but still has issues to be addressed

Issues

BROOKHAVEN NATIONAL LABORATORY

- ➢ Inefficiency due to old resource partitioning based on geographical grouping of computing centers
- ➢ Suboptimal usages of non-traditional resources due to job-based workload management
- ➢ Incoherent implementations for various HPC workflows
- ➢ Overstretched architecture of the pilot to support non-traditional resources
- ➢ To leverage prediction capabilities for resource availability actively developed with recent computing technologies like machine learning
- ➢ Operational difficulties with new workflows due to job-centric visualization

System Evolution

Resource Consolidation

BROOKHAVEN NATIONAL LABORATORY

- Old MONARC hierarchical model
  - Tier1 + Tier2 centers: 10 Tier1 centers → 10 partitions
  - Rather static combination between Tier1 and Tier2 centers based on geographical grouping
  - Problematic when small partitions got high priority tasks → Complicated workload brokerage
- New model with no hierarchy
  - A single resource partition
  - Two new concepts
    - Nucleus : destination of output data
    - Satellite : processing jobs to produce output data
  - Formation of sub-partition (nucleus+satellites) based on static configuration and dynamic information on network quality between nuclei and satellites
  - Reliable Tier2 centers as nuclei in addition to Tier1 centers
- Details in CHEP16 talk

  F. H. Barreiro Megino : ATLAS World-cloud and networking in PanDA

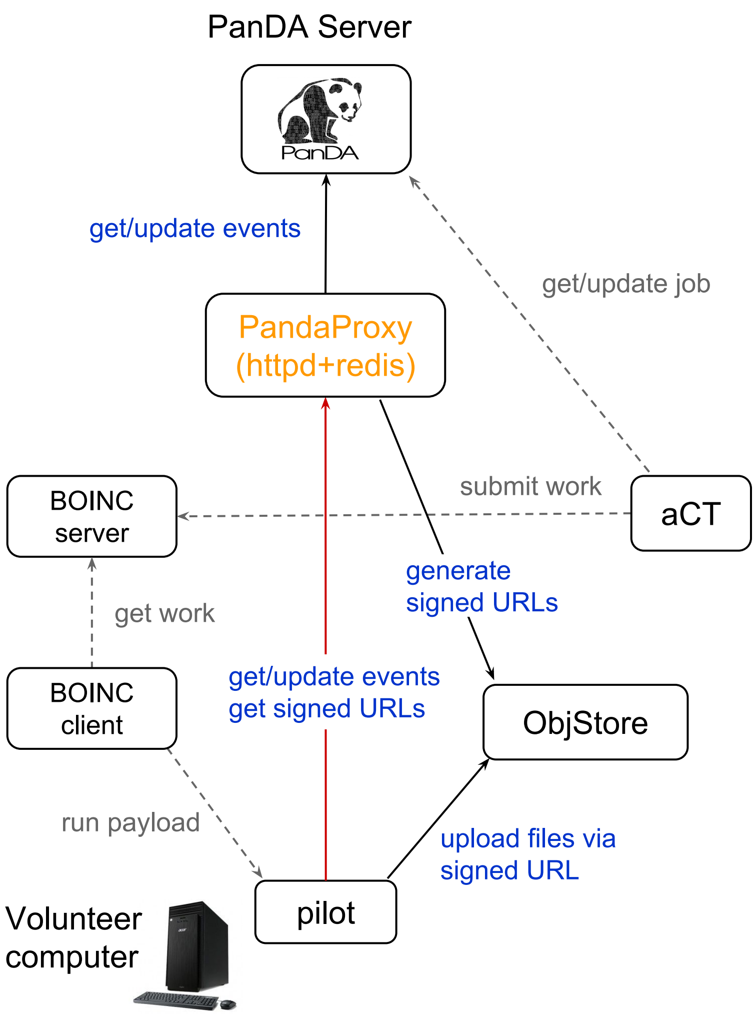
Intelligent Brokerage

BROOKHAVEN NATIONAL LABORATORY

- More intelligence to the brokerage based on
  - Job retry history
  - Network forecast
  - Cache hit rate
- Adding a new capability for workload provisioning
  - Passive workload assignment
    - Assigning jobs to resources once they become active
    - Good for traditional resources as a steady number of CPUs is usually available there except short disruptions
    - Latency too high for non-traditional resources like the ATLAS HLT farm as the number of available CPUs tends to ramp up and down immediately
  - Proactive workload assignment
    - Assigning jobs just before resources become available
    - Removing jobs just after resources become unavailable
    - Based on (quasi) real-time resource information
  - Workload assignment could trigger input data transfers
    - Should be smart enough for proactive assignment to minimize redundant data transfers

In-House Security

- To authenticate requests from the pilot running on special environment where standard X509-based auth is unavailable and/or suboptimal



- Using an internal secret token
  - Generated by the PanDA server per job or pilot scheduler to be kept in redis
  - Propagated to the pilot via job specification or VM contextualization
  - Sent to PandaProxy together with normal pilot requests for verification
- Use-cases
  - Volunteer computing
  - Off-grid
  - Commercial clouds with rapid spin-up

PanDA at HPC Centers

- Titan at OLCF
  - 18,688 nodes: 16 CPU cores
  - MPI jobs with multiple PanDA jobs
  - Data transfers with the DTN scheduler or pilot movers
  - Steady operations for continuous PanDA production job submission in backfill mode
    - 10k jobs per day on average (18k at peak)
  - Details in CHEP16 talk

    S. Panitkin : Integration of the Titan supercomputer at OLCF with the ATLAS Production System

- SuperMUC at LRZ
  - 3072 nodes: 28 CPU cores
  - MPI jobs with Event Service + ARC CE
  - Data transfers with ARC
  - Changed from short jobs to Event Service jobs due to frequent preemption

NERSC

- Edison/Cori at NERSC
  - 9,304 nodes: 68 CPU cores
  - MPI jobs with Event Service
  - Data transfers with 3rd party service or pilot movers
  - Low CPU efficiency to be addressed
  - Commonalities for coming Theta and Aurora at ALCF

- HPC2 at NRC-KI
  - 2560 nodes: 4 CPU cores
  - non-HEP : Next Generation Genome Sequencing (NGS)
  - No data transfers
  - Details in CHEP16 talk

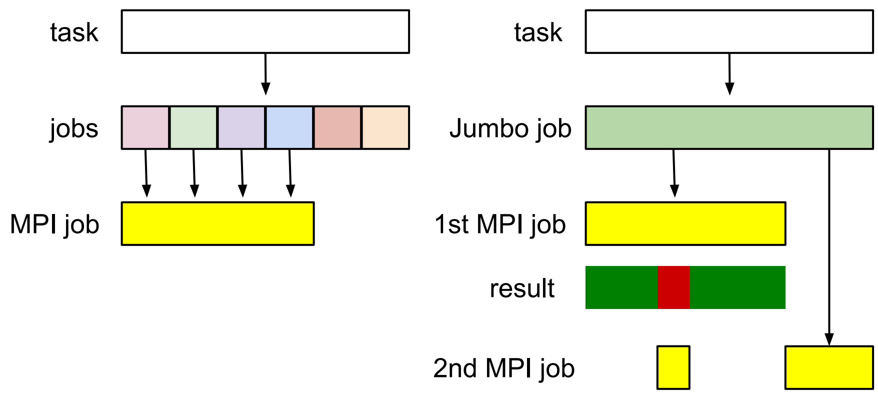
    A. Klimentov, R. Mashinistov : Using HEP Computing Tools, Grid and Supercomputers for Genome Sequencing Studies

- Many other HPCs ...

Enhancement of Event Service

- Fine grained (event-level) workload partitioning
  - Allowing jobs to be revoked in the middle of processing with minimized losses
- The old implementation assumed a modest number of events per job (~1k)
  - Good for preemptable resources
    - Dynamic fragmentation of jobs
  - Not good for large HPC resources
    - Preference for a large number of events in one go
    - Combination of multiple jobs to a single MPI job
      → Complicated workload management and bookkeeping

- New feature : Jumbo jobs allowing workloads to be tailored to any sizes of MPI jobs
- CHEP16 talk for Event Service
  T. Wenaus, V. Tsulaia : Production Experience with the ATLAS Event Service



Monitoring Evolution

BROOKHAVEN NATIONAL LABORATORY

- Rise of data visualization such as plots, histograms, and task chain diagrams
- Adding predictive analytics for the expected task completion time and comparison with actual task computation progress
- Intensive use of Redis cache and page preloading to improve user experiences
- Quick access to the data in external monitoring systems: Kibana, Agis, Rucio, dashboard
- Current development activities
  - Oracle side data aggregation
    - Supports for free search queries is mandatory in monitoring
    - On-demand data aggregation triggered by a typical query takes minutes to process 6 months historical data with 2M jobs per day
    - Data aggregation on Oracle server side with an advanced data layout strategy successfully demonstrated to reduce the query processing time from ~1 min (30k jobs limit, last 12 hours jobs) to ~10 sec (without limit, for last 12 hours = ~800k jobs)
  - SSO, VOMS, IGTF authentication
    - To introduce an information access control layer
    - Ability to execute commands directly from monitoring pages
    - Future implementation of user-centric contents

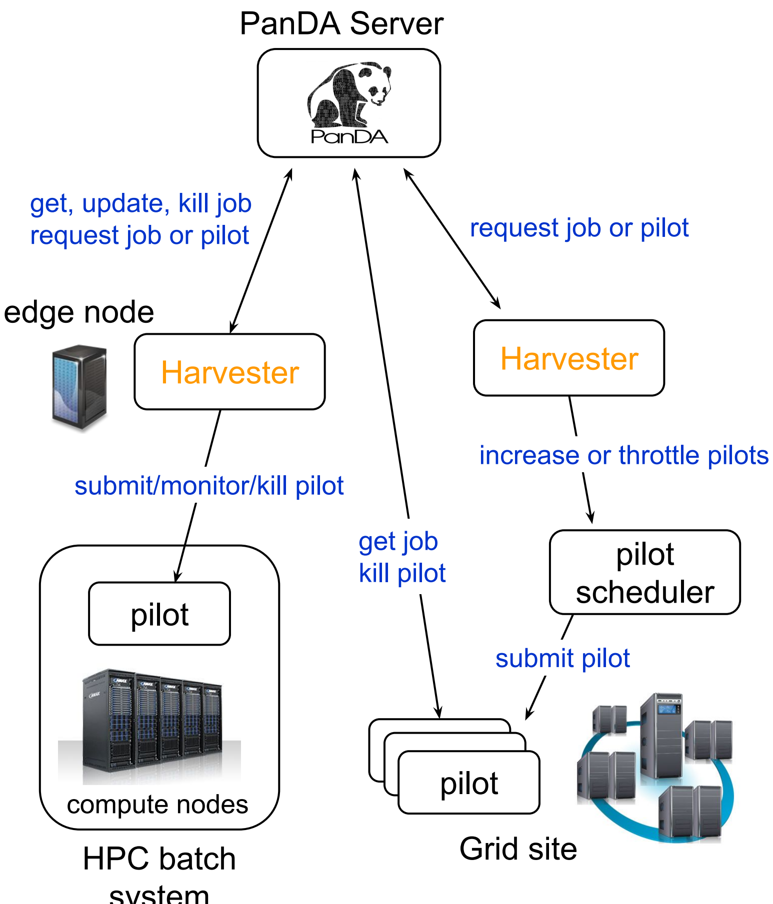
Pilot 2.0

BROOKHAVEN NATIONAL LABORATORY

- New PanDA Pilot Project launched in April 2016
  - Maintenance difficulties of aged PanDA Pilot
  - Overstretched architecture due to new features and workflows
  - Long-term project to span the next couple of years
  - An almost complete rewrite except some recent developments
  - Involvement of developers both within core PanDA and BigPanDA teams as well as from external teams
- Currently the project is in the design stage
  - Development for a MiniPilot system
    - A fully working pilot script for developers to test new components
    - Future evolution into a SimplePilot to new PanDA users for a rapid introduction
  - Git-based testing framework
    - A pull request into the Pilot Git repository to trigger a verification sequence including unit tests
  - Test implementations of the component model being evaluated with all workflows

Harvester

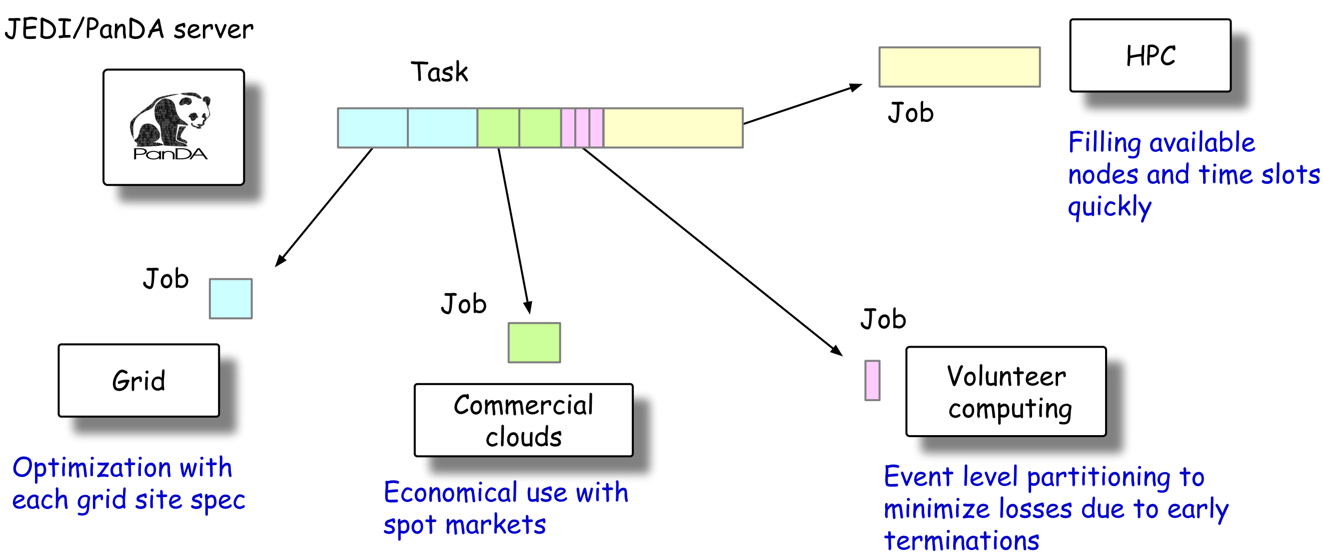
- A resource-facing service between PanDA server and collection of pilots
  - Propagating information or requests between PanDA server and resource managers such as batch systems and pilot schedulers



- Objectives
  - To add a capability for timely optimization of CPU allocation among various resource types
  - To provide a commonality layer in bringing coherence to HPC implementations
  - To have better integration between PanDA system and resources for new workflows
- Developments actively in progress with wider collaboration
  - First prototype for NERSC/Titan after CHEP

Future Plans

- Proactive control of the network to optimize workflows and dataflows
- New components in production
- Automation of the system based on prediction for resource availability and the expected completion time for each task
- More efficient and economical use of traditional and new computing resources



Summary

BROOKHAVEN NATIONAL LABORATORY

- PanDA has performed well for ATLAS in the last decade including the LHC Run 1 and Run2 data taking periods
- New components and features have been delivered to ATLAS
- Many developments and challenges to come while steadily running for LHC Run 2

Highlights

BROOKHAVEN NATIONAL LABORATORY

- PanDA has performed well for ATLAS in the last decade including the LHC Run 1 and Run2 data taking periods
- New components and features have been delivered to ATLAS
  - Resource consolidation
  - Intelligent brokerage
  - In-House security
  - PanDA on HPCs
  - Enhancement of event service
  - Monitoring evolution
- Many developments and challenges to come while steadily running for LHC Run 2
  - Pilot 2.0
  - Harvester
  - Network provisioning
  - Automation based on prediction capabilities
  - More optimal use of computing resources