PanDA-based GRID Workload Management

Maxim Potekhin
(presenting for BNL Physics Applications Group)
Brookhaven National Laboratory

OSG All Hands Meeting
March 2-5, 2009
LIGO Livingston Observatory

Open Science Grid

Panda Intro

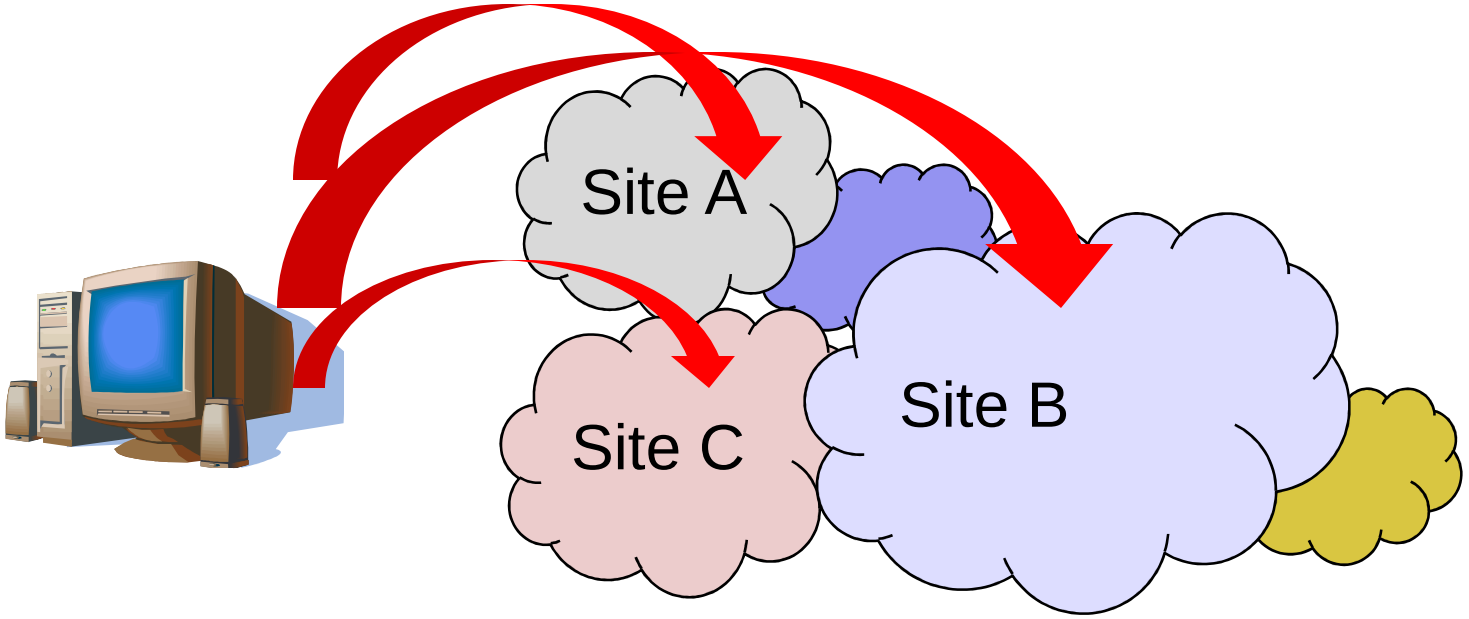
The Panda (Production ANd Distributed Analysis) system has been developed since summer 2005 to meet challenging requirements of ATLAS Collaboration for a large scale, data-driven workload management system for production and distributed analysis.

Since September 2006 Panda has also been a principal component of the US Open Science Grid (OSG) program in just-in-time (pilot-based) workload management. In October 2007 Panda was adopted by the ATLAS Collaboration as the sole system for distributed processing production across the Collaboration.

In addition to serving the needs of Atlas community, Panda has also been used by scientists from other disciplines, such as a group of researchers at National Institute of Health.

Since its commissioning, Panda has processed tens of millions of jobs on dozens of sites around the world. In addition to the production workflow, there are thousands of analysis jobs run daily.

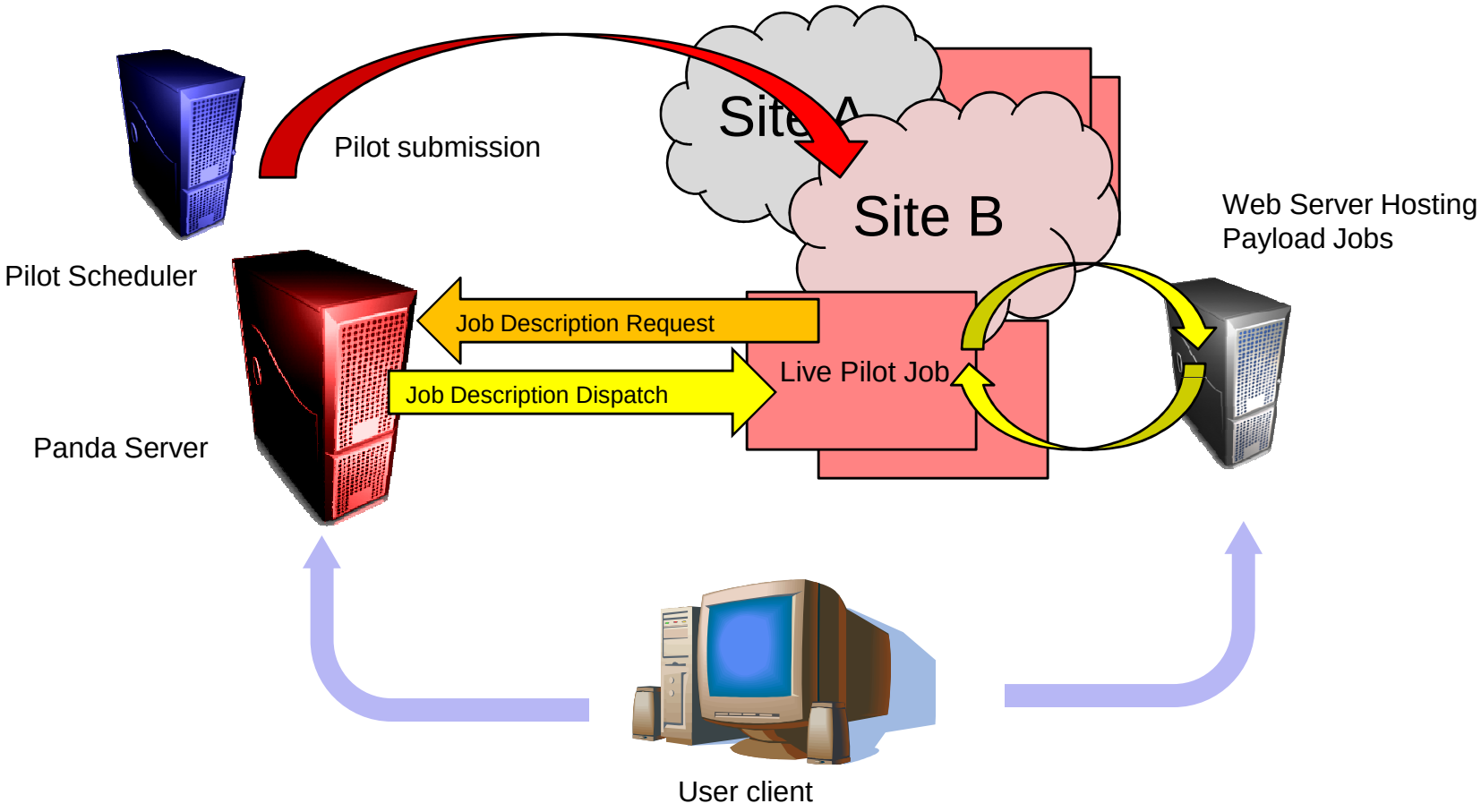
Direct Job Submission (without Panda)



Disadvantages:

- need to interface and manage diverse and heterogeneous processing resources
- absence of a system-wide view of job status and progress
- lack of uniform and integrated data management
- hard to control latencies and failure modes inherent in generic in job submission (critical for analysis)
- etc…

Panda Pilot-based job management: the concept



(…next slide)

Panda Pilot-based job management: the concept

Panda's Pilot Framework for Workload Management

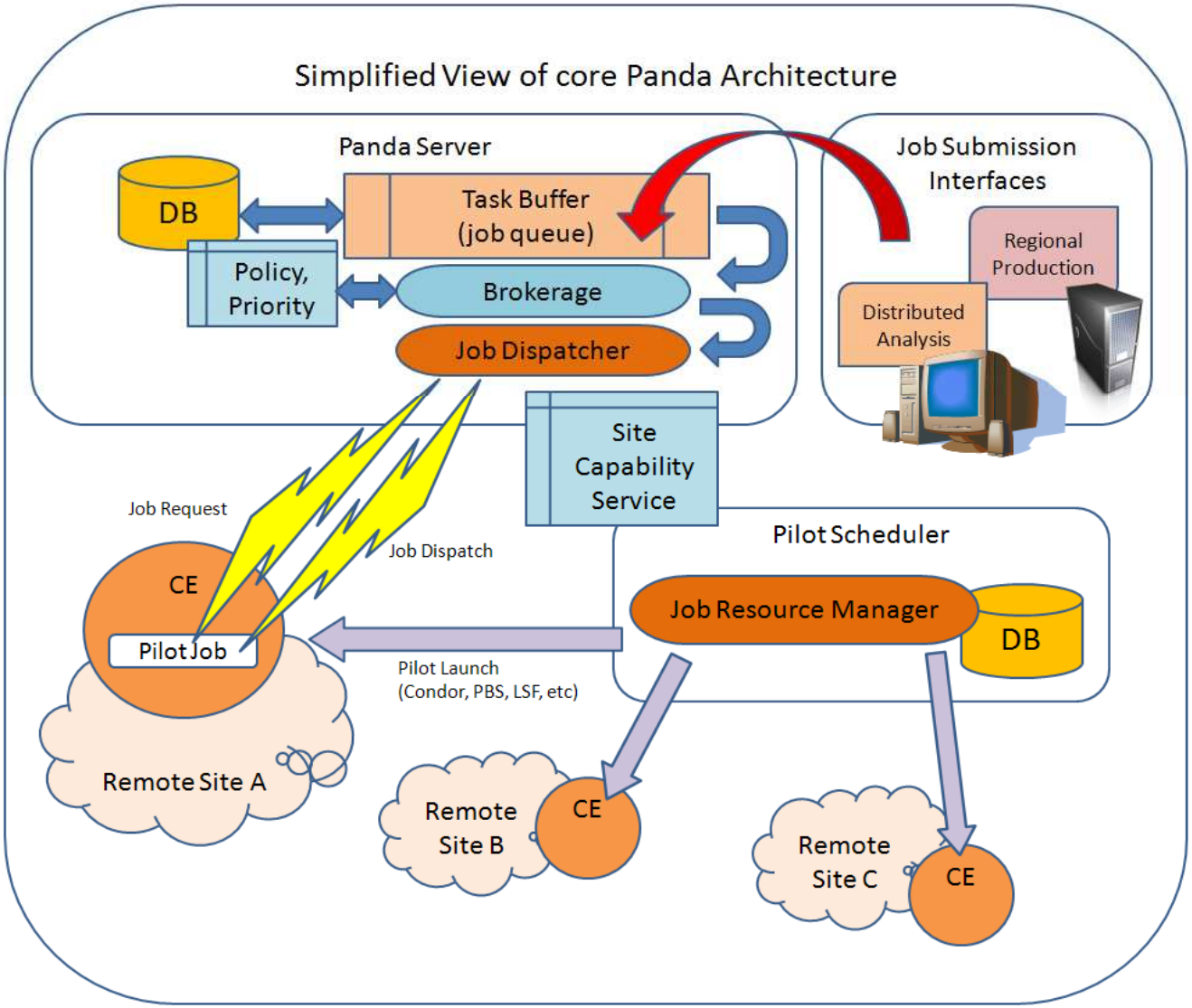
- Workload jobs are assigned to successfully activated and validated Pilot Jobs (lightweight processes which probe the environment and act as a 'smart wrapper' for payload jobs), based on Panda-managed brokerage criteria. This 'late binding' of workload jobs to processing slots prevents latencies and failure modes in slot acquisition from impacting the jobs, and maximizes the flexibility of job allocation to resources based on the dynamic status of processing facilities and job priorities.
- The pilot also encapsulates the complex heterogeneous environments and interfaces of the grids and facilities on which Panda operates. The users do not need to concern themselves with intricacies of Grid interface – Panda presents them with a unified mode of access to Grid resources.

Job Submission

Jobs are submitted via a client interface where the jobs sets, associated datasets, input/output files etc can be defined. Jobs received by the Panda server are placed in the job queue, and the brokerage module to prioritizes and assigns work based on job type, available resources, data location and other criteria. The payload is not stored on the Panda server - it is defined as a URL from which it can be retrieved, thus improving the scalability and ease of management by the users. To communicate with the Panda and payload servers, the Pilot needs to be capable of outbound HTTP connectivity from the Worker Node on which it is run.

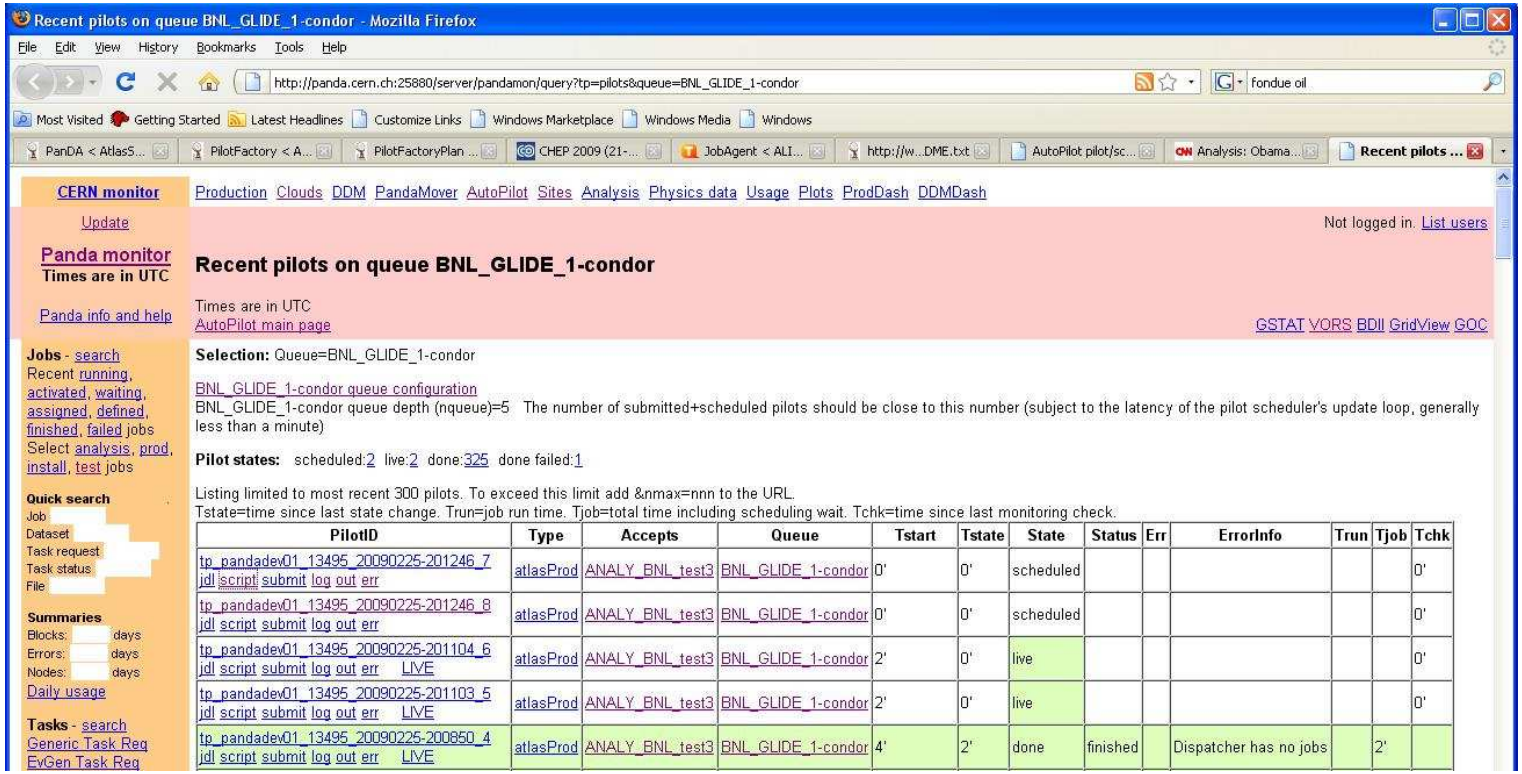
Open Science Grid

Panda Architecture



Panda Monitoring

Panda monitoring system is a separate component based on the Apache server, which allows the users and operators to have a comprehensive view of many aspects of the job progress through the system and gives them the capability to “drill down” into job execution detail.



Panda Pilot-based job management

Submission of Pilot Jobs

Panda makes extensive use of Condor (particularly Condor-G) as a Pilot job submission infrastructure of proven capability and reliability. Pilots are submitted via Pilot Schedulers (Generators), which are typically run by administrators of the Virtual Organization wishing to submit jobs to Panda. Submission rate is regulated by the number of job requests queued in the server, thus eliminating creation of unused pilots and waste of resources.

Other principal design features

• Through a system-wide job database, a comprehensive and coherent view of the system and job execution is afforded to the users.

•Integrated data management is based on the concept of 'datasets' (collections of files), and movement of datasets for processing and archiving is built into the Panda workflow. Asynchronous and automated pre-staging of input data minimizes data transport latencies

•Panda is based on the industry-standard Apache server and therefore renders itself to well understood performance tuning and scalability enhancing procedures. Its security is based on standard Grid tools (such as X.509 certificate proxy-based authentication and authorization)

Summary

Panda presents a coherent, homogeneous interface to distributes Grid resources to the user, in both production and analysis situation. It mitigates effects of job submission latency and isolates the user from many failure mode that may exist in Grid job submission scenario. In addition, it provides integrated data movement capabilities and extensive monitoring tools to users and operators. It has proved itself as a stable and scalable system, capable of addressing computing needs of a large and global organization, as well as of smaller research teams.