PanDA Evolution: Strategy, Priorities, and Next Steps

Tadashi Maeno (BNL)
On behalf of PanDA Collaboration

Mar 11 2026, ePIC Software & Computing Weekly Meeting

PanDA Strategic Planning for the LS3–HL-LHC Era

- Strategic planning meeting held at UMass in Jan
  - Joint meeting with the REDWOOD all-hands
  - Goal: define a realistic evolution plan for PanDA during LS3
- Focus
  - Ensure readiness for HL-LHC and EIC computing requirements
  - Review architecture, infrastructure, and emerging technologies
- Key discussion areas
  - Core architecture and modularization
  - AI/ML/LLM integration
  - GPU and heterogeneous resource support
  - Cross-experiment sustainability

Core Architecture and Modularization

- Drivers for Architectural Evolution
  - Integration of ePIC streaming workflows
  - Support for AI/ML/LLM capabilities
- Modularization priorities
  - Decouple core PanDA functionality from experiment-specific logic
  - Move shared features into the core framework
- Deployment model
  - ATLAS's gradual transition from VM-based deployments → Kubernetes
    - Align with current infrastructure used by most experiments
  - Fully-automated CI/CD
- Development environment
  - Revisit the "mini-PanDA" concept
    - Push-button playground with easy installation
    - Support for local development and rapid testing

AI/ML/LLM and Agent Integration

- Backend architecture
  - Maintain current DB / Django / REST backend services
- Interface and core evolution
  - Add agent-based front-ends using MCP tools
  - AI-driven agents as first-class citizens
- Expected benefits
  - Lower barrier for non-expert users
  - Improved monitoring and diagnostics
  - Enhanced reasoning, automation, error mitigation, workflow orchestration, …
- Open issues
  - Long-term funding model for LLM usage remains unresolved
  - Data governance considerations when sending data to external services

GPU Integration

- Current status
  - ~10 ATLAS GPU queues currently deployed
- Infrastructure goals
  - Expand support for diverse GPU hardware models
  - Handle different CUDA versions more systematically
- Scheduling improvements
  - Improve job brokerage through better hardware awareness:
    - Propagate worker-node metadata to CRIC
    - Alternatively use Pilot-based hardware maps
- Planning
  - Define a GPU development roadmap by end of 2026
  - Based on requirements from potential users
  - Includes support for long-running jobs

Cross-Experiment Support

- Strategic importance
  - Cross-experiment capability is important for future funding
- Current challenge
  - Limited personnel resources
- Operational principle
  - Maintain a strict separation between:
    - Core PanDA functionality
    - Experiment-specific components
- Goal
  - Ensure clear alignment between:
    - Funding sources
    - Personnel effort
    - Experiment mandates

Outlook

- LS3 as the key transition window
  - Use LS3 to evolve PanDA toward the scale and diversity expected in HL-LHC and EIC
  - Extend the system to support new computing paradigms and user communities
- Strategic directions
  - Strengthen modular architecture to support multiple experiments sustainably
  - Integrate heterogeneous computing resources including GPUs and future accelerators
  - Explore AI-assisted operations and agents to improve system usability and core functions
  - Enable new workflow models, including streaming and AI-driven workloads
- Collaboration and sustainability
  - Maintain a clear separation between core PanDA development and experiment-specific components
  - Align development priorities with available funding and personnel
  - Expand cross-experiment collaboration to ensure long-term sustainability
- Goal
  - Keep PanDA a central, flexible workload management system for HL-LHC, EIC, and emerging data-intensive science