Event Service
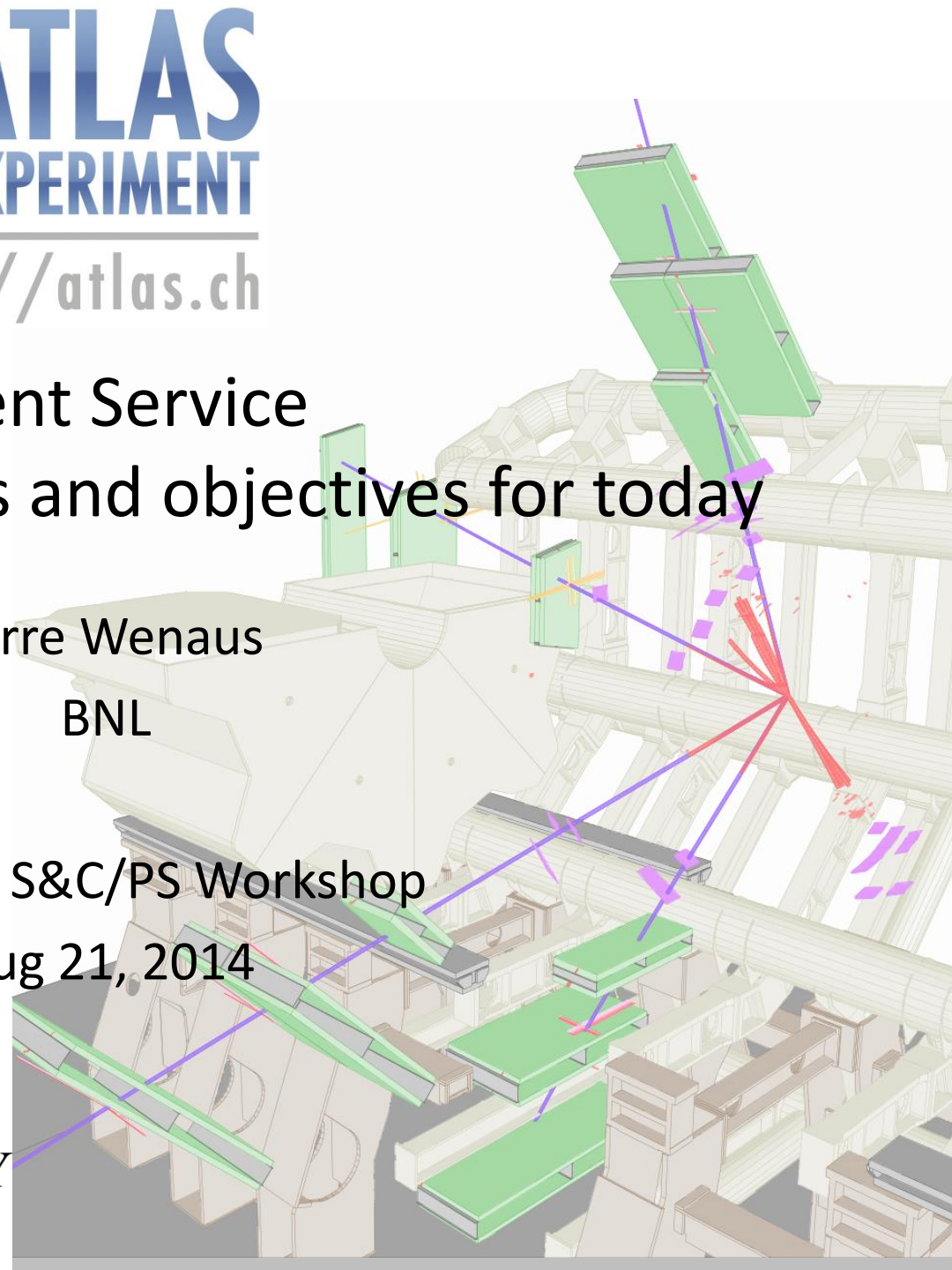## Intro, plans, issues and objectives for today

Torre Wenaus

BNL

US ATLAS S&C/PS Workshop

Aug 21, 2014

Event Service

- A new approach to event processing: stream events in near-continuous way, both at input and output, through a worker node that may have a very short (or a very long) lifetime

- Decouple processing from the chunkiness of files, and data locality considerations – potentially WAN latency insensitive

- Potential to make processing completely asynchronous from the flow of input data to the worker

- Light footprint on the WN; if it disappears we lose little

- WNs with very short lifetime: great for opportunistic resources

  - The sand filling up the "full" jar of rocks in HPCs

  - More examples of using others' excess capacity: Amazon spot market, volunteer computing (BOINC, created here)
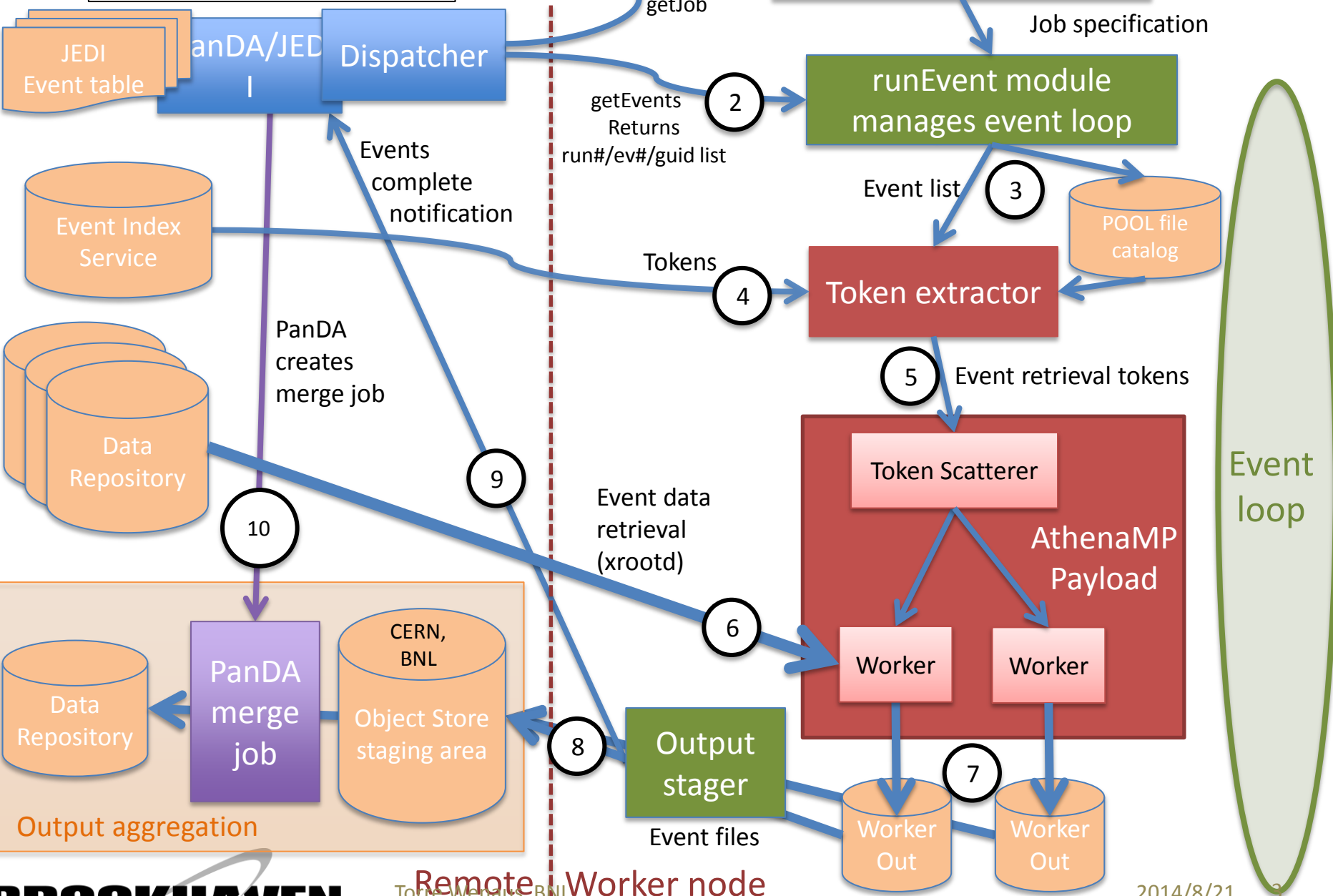
Event Service in US ATLAS

- A fruitful US-dominated collaboration in ATLAS
  - Across the labs
  - Across core and distributed software
  - Soon, across software and facilities as we deploy
- First target: simulation production
  - Relatively simple workflow, I/O and metadata
  - Relatively easy to port to target platforms like HPC
  - Gives the biggest gain: most of our cycles go to simu
- Beyond that: let's worry first about simulation production (but potential is there for less CPU-intensive workloads also, especially if we exploit asynchronous data delivery)

BROOKHAVEN

Status

- Transitioning from end-to-end testing to pre-production testing
- Formerly a best-effort activity, now a high profile one with important, dated deliverables
- Some of which are imminent
  - Opening of opportunity for Amazon spot market large scale deployment
  - Supercomputing 2014 demo sponsored by DOE ASCR, preliminary to new ASCR funding rounds
  - Rise of HPCs in ATLAS: lots capacity full of rocks but not sand
  - ATLAS@Home is up and running, but feeling the lack of ES

Today

- Survey the technical status, emphasis on outstanding issues and near/midterm plans
- Then, in part 2, focusing on the early deployment targets
  - Preproduction on the grid
  - Amazon spot market
  - HPCs
  - Emerge from today with timelines and milestones
- In a splinter at the end of the day, continue discussion on biggest outstanding technical question mark: ES on HPCs
  - We have a concept, 'Yoda', that we want to move from concept to defined development activity today

# Concept for MPI-based event service extension for HPCs
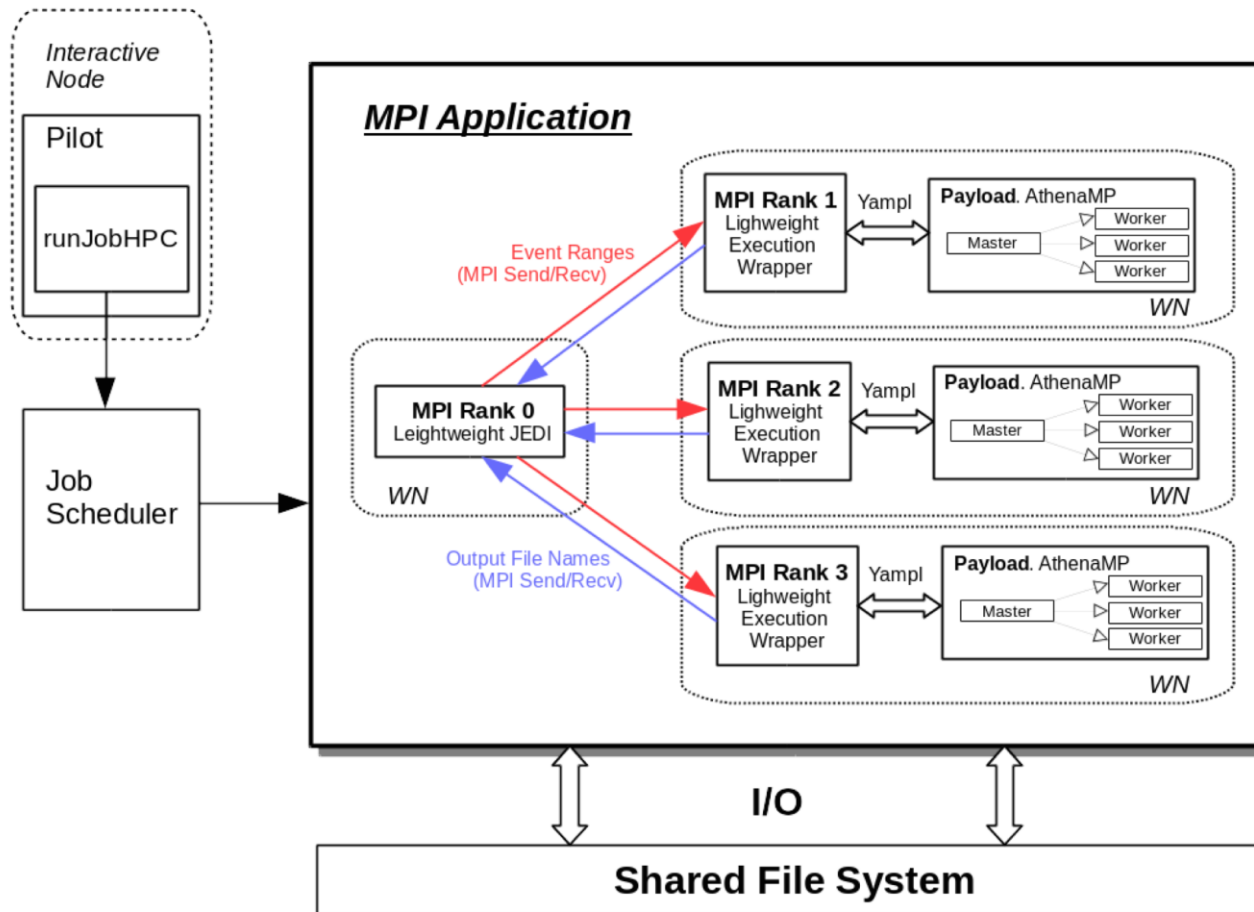
Supplemental – SC14 demo

Granular data processing on HPCs using an Event Service

Objectives and Impact

- Demonstrate a new fine grained approach to HEP event processing

- Agile & efficient in exploiting potentially short-lived resources: HPC hole-filling, spot market clouds, volunteer computing

- Data flows utilize cloud data repositories with no data locality or prestaging requirements

- Minimize use of costly storage in favor of strongly leveraging powerful networks
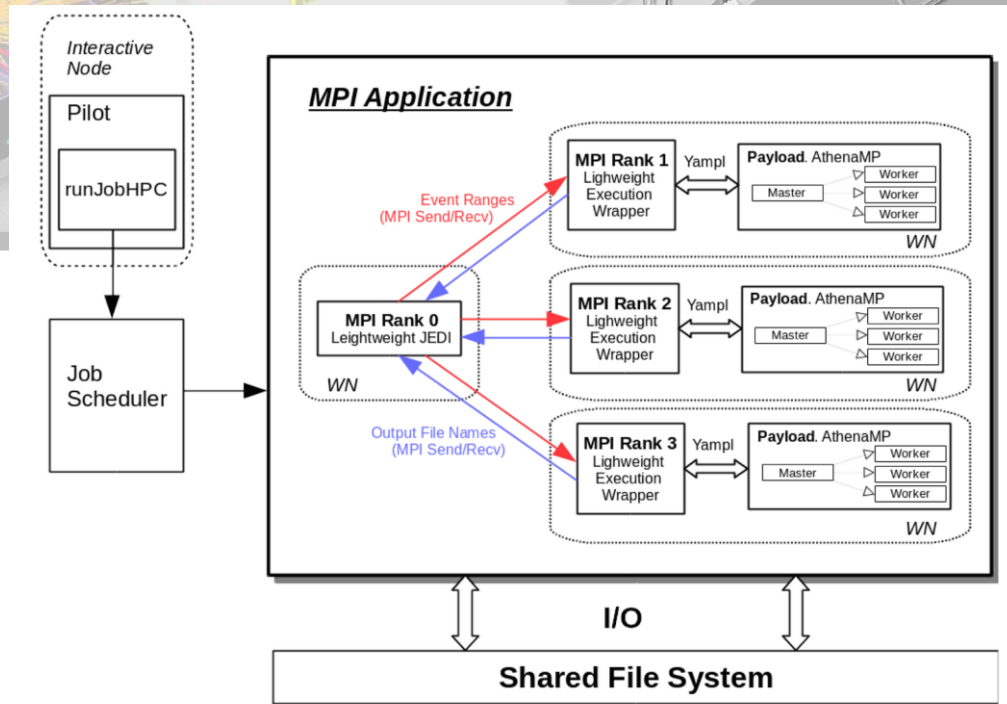
Schematic of the 'Yoda' extension of PanDA's Jedi dynamic job manager into HPCs as an MPI based service. Being implemented by the BNL, LBNL and UT Arlington based team that has implemented the event service and ported PanDA to HPC (ORNL Titan). The Yoda work's first target is NERSC HPCs.

Features

- Applicable to any scientific processing that can be finely partitioned

- ASCR leverage: HPCs, powerful network, cloud data stores, exascale workload management (ASCR supported BigPanDA, 1.3 ExaBytes processed in '13)

- BNL, LBNL, ANL, ORNL, PNNL involved

Technical Achievements

- Successful full chain test of Event Service on real ATLAS simulation workloads; entering pre-production testing

- Data flows to cloud object stores, currently CERN, adding BNL

- PanDA now HPC capable: running at OLCF Titan

- Now extending PanDA's new Jedi dynamic job manager into a mini MPI version as an event service manager for HPCs – 'Yoda'

BROOKHAVEN

Why the ES for an ASCR demo

- The US ATLAS core software teams at 3 DOE labs BNL, ANL, LBNL for the past ~year have been collaborating on developing an 'Event Service' (ES) for ATLAS, leveraging their ATLAS core expertise:
    - BNL - distributed computing, workload management (PanDA)
    - ANL - event I/O
    - LBNL - multiprocessing core framework
- The ES is a new approach to HEP event processing: fine grained, agile in exploiting potentially short-lived opportunistic resources (such as HPC hole-filling), heavily leveraging excellent networking, and designed to integrate well with data repositories 'in the cloud' (e.g. object stores)
- It leverages ASCR strengths: HPCs, powerful networking, 'virtual data centers' (whatever they are, the ES can make good use of them)
- It leverages a project ASCR is already paying for: BigPanDA, extending PanDA beyond ATLAS to HPCs, intelligent networking and the Exascale community
- It leverages an existing collaboration between four labs -- in addition to the three mentioned, ORNL participates through their (ALICE-directed) participation in the BigPanDA project, and PNNL (Belle II) interested also

BROOKHAVEN