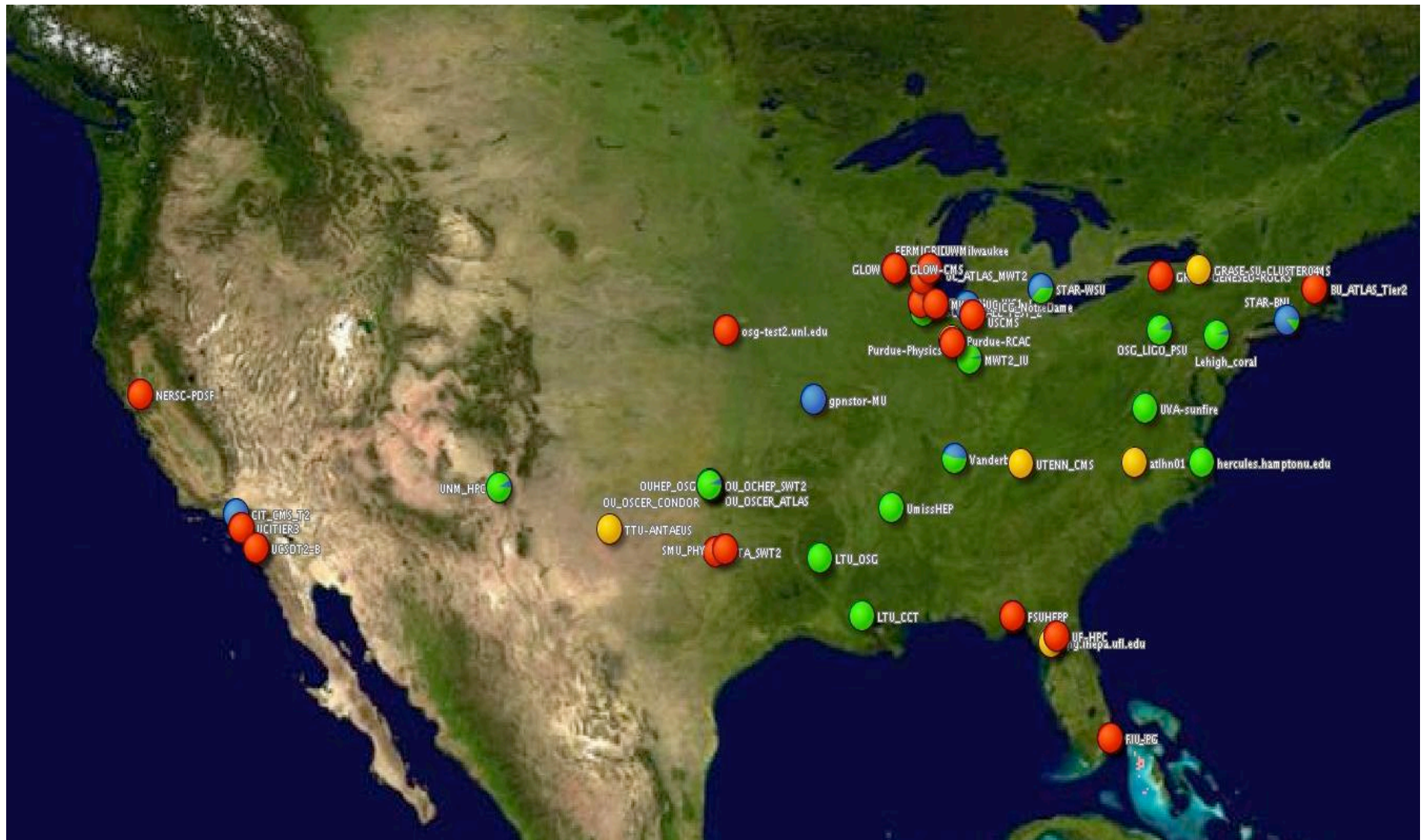
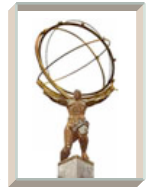


ATLAS MC Production on OSG with PanDA

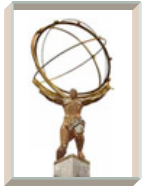
Yuri Smirnov
Brookhaven National Laboratory

CHEP, Vancouver
September 5, 2007

Open Science Grid (OSG) map

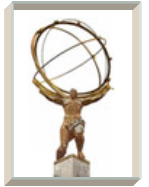


Introduction



- ❑ OSG includes many activities – see talk 42 by Ruth Pordes
- ❑ This talk – ATLAS MC production on OSG facilities
- ❑ OSG contributes >30% of ATLAS MC production for Computer System Commissioning (CSC) exercise
- ❑ OSG resources are organized as one large Tier 1 cloud within ATLAS:
 - ❑ BNL Tier 1 facility
 - ❑ Five regional Tier 2's (Tier 2's may provide multiple clusters):
NorthEast T2, MidWest T2, SouthWest T2, SLAC T2, GreatLakes T2
 - ❑ Resources at non-ATLAS sites (including Tier 3's)
 - ❑ We use PanDA system for ATLAS production and analysis
 - See talks 167-T. Maeno, 170- P. Nilsson

Service Organization



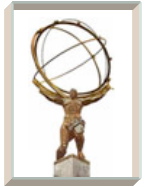
❑ Tier 1 and Tier 2 sites provide services:

- ❑ Batch systems (with local storage) through gatekeeper
- ❑ Global storage systems accessible through the ATLAS Distributed Data Management system, DQ2
- ❑ Other services required by ATLAS (networking, database...)
- ❑ Sites are responsible for QoS (site monitoring)
- ❑ Facilities integration and testing teams

❑ U.S. ATLAS operations teams:

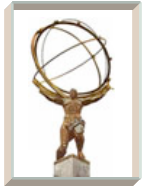
- ❑ Production operations team run daily shifts for managed production, regional production (U.S. groups), distributed analysis: YS (captain), Mark Sosebee, Nurcan Ozturk, Wensheng Deng, Barry Spurlock
- ❑ DDM operations team

ATLAS Production



- ❑ Computer System Commissioning (CSC)
 - ❑ Software integration and operations exercise
 - ❑ Started more than 1.5 years ago (since end of 2005)
 - ❑ Distributed (MC) production goals
 - Validation of Athena software using 'standard' samples
 - Continuous production of high statistics physics samples
 - To exercise widely distributed computing infrastructure
 - ❑ Distributed Data Management (DDM) goals
 - Worldwide data distribution and replication
 - Controlled flow of data $T0 \rightarrow T1 \rightarrow T2 \rightarrow T1$
- ❑ Bring sites and services to steady operations mode

Distributed Production Operations



❑ ATLAS is divided into three production grids

- ❑ EGEE, NDGF and OSG (10 Tier 1's, ~40 Tier 2's, +++)
- ❑ NDGF (Nordic) ~500 CPU's, ~60TB disk
- ❑ OSG (U.S.) ~2500 CPU's, ~386 TB disk
- ❑ EGEE ~3500 CPU's, ~358 TB disk

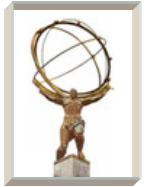
❑ Common features

- Task/Job definition, Job dispatch (supervisor), metadata (AMI)
- Access to data through ATLAS DDM system

❑ Differences

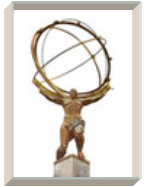
- Production software (executors), grid middleware
- Independent operations team for each grid
- Service architecture, storage systems...

PanDA



- ❑ PanDA is the distributed computing service infrastructure deployed in the U.S. and Canada (joined few weeks ago)
 - ❑ Includes all U.S. and Canadian Tier 1 and Tier 2's
 - ❑ Plus some opportunistic/shared sites
 - ❑ Works both with OSG (U.S.) and EGEE (Canada) middleware
- ❑ PanDA uses pilot jobs and provide a single task queue
 - ❑ Pilot jobs allow instant activation of highest priority tasks
 - ❑ Task queue brings familiar 'batch system' to distributed grids
- ❑ PanDA provides an integrated software and computing system for U.S./CA ATLAS sites, managing all production and user analysis activities

Dynamics of Production Operations



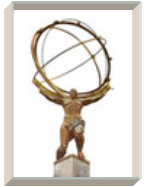
❑ First some nomenclature

- ❑ **Job** – atomic unit of execution ('n' events processed by ATHENA using a single CPU, real or virtual)
- ❑ **Task** – collection of jobs through a processing step (evgen, recon, pileup, digit...) for a physics topic (top, Higgs, SUSY...)
- ❑ **Dataset** – collection of similar files from a single task

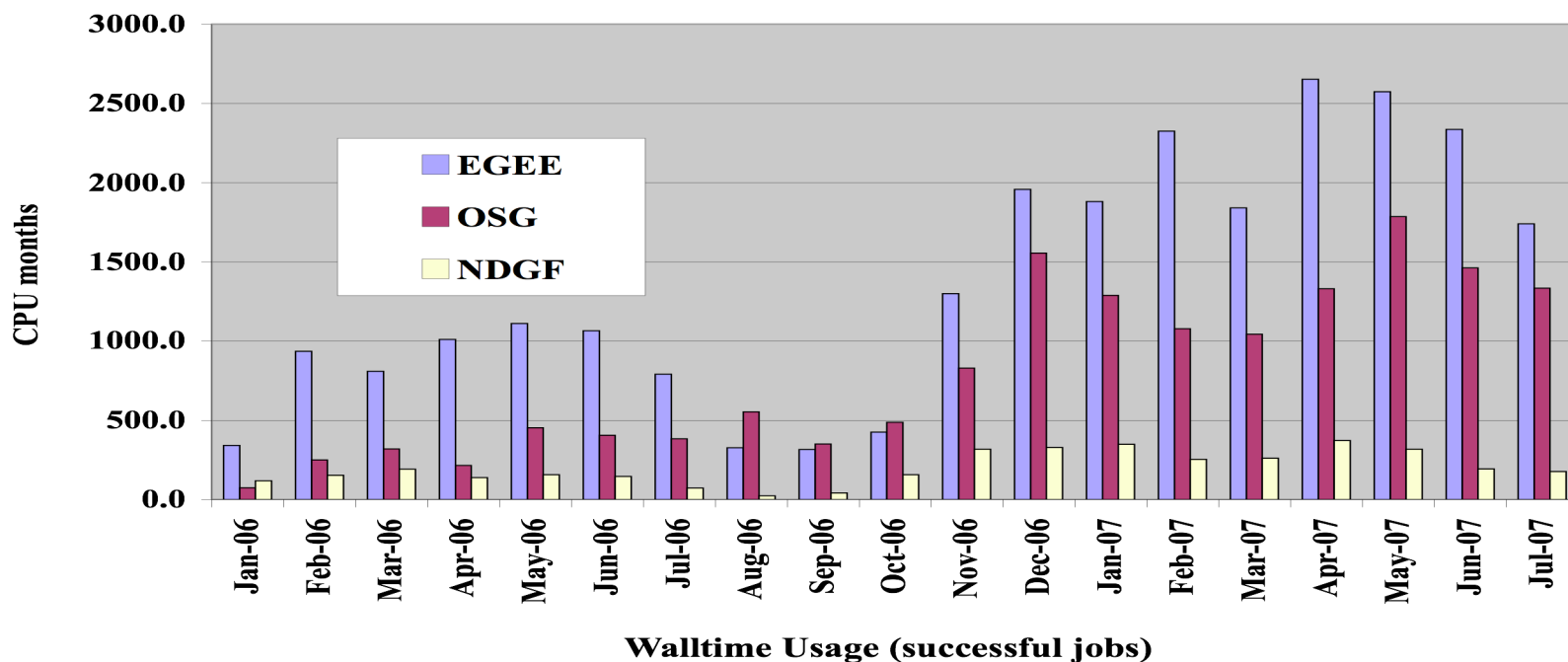
❑ MC Production steps

- ❑ Physics groups define tasks (assigned to a specific grid)
- ❑ Tasks are converted to jobs and stored centrally
- ❑ Shift teams manage successful execution of jobs
- ❑ DDM operations team monitors and manages data flow
- ❑ Physicists analyze data on grid

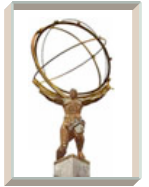
Resource Usage



- ❑ CPU and disk resources available to ATLAS rising steadily
- ❑ Production system efficiencies are steadily increasing
- ❑ But much more will be required for data taking
- ❑ Additional resources are coming online soon

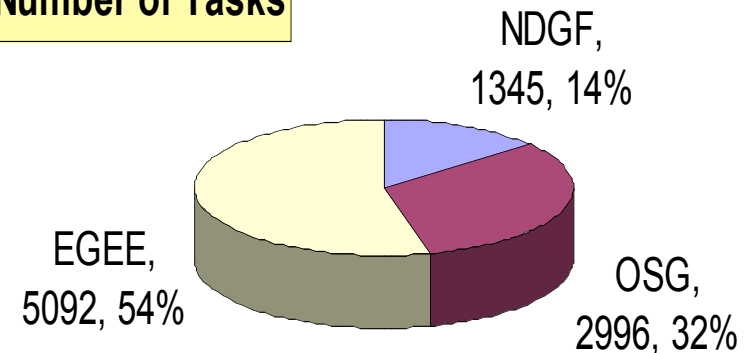


CSC Production Statistics

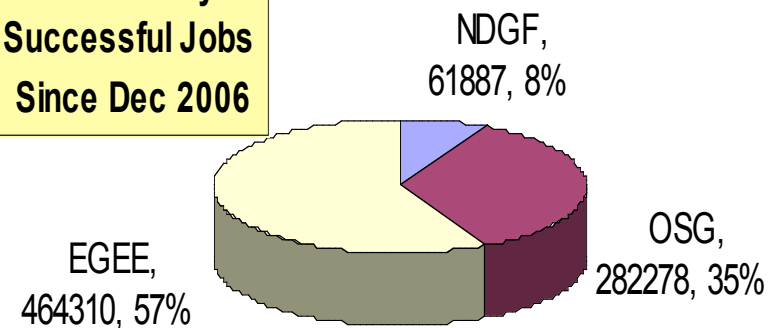


- ❑ Many hundreds of physics processes have been simulated
- ❑ Tens of thousands of tasks spanning two major releases
- ❑ Dozens of sub-releases (about every three weeks) have been tested and validated
- ❑ Thousands of 'bug reports' fed back to software and physics
- ❑ 50M+ events done from CSC12
- ❑ >300 TB of MC data on disk
- ❑ Impressive team effort!

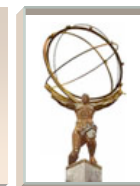
Number of Tasks



Walltime Days
Successful Jobs
Since Dec 2006



PanDA (US, Canada) Production Statistics



❑ PanDA has completed ~30M fully simulated physics events (simul+digit step), >30% of total central production

❑ Also successfully completed >15M single particle events

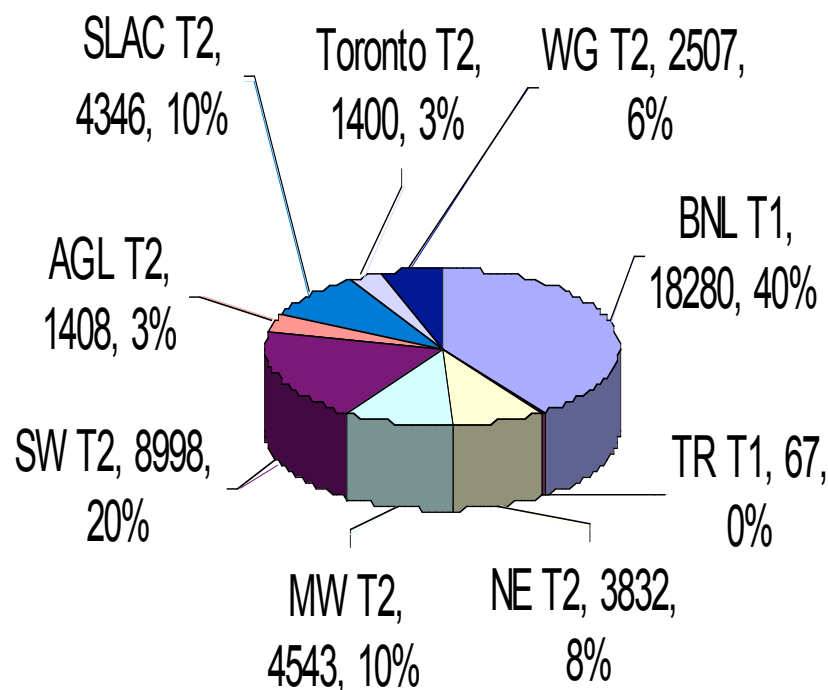
❑ Since November, all available CPU's occupied (ran out of jobs only for few days, plus few days of service outages)

❑ About 400 TB of original data stored at BNL T1 (includes data generated on other grids)

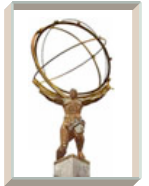
❑ Additional ~100 TB of replicas kept at U.S. ATLAS Tier 2 sites

❑ Canadian sites are now using PanDA for ATLAS production

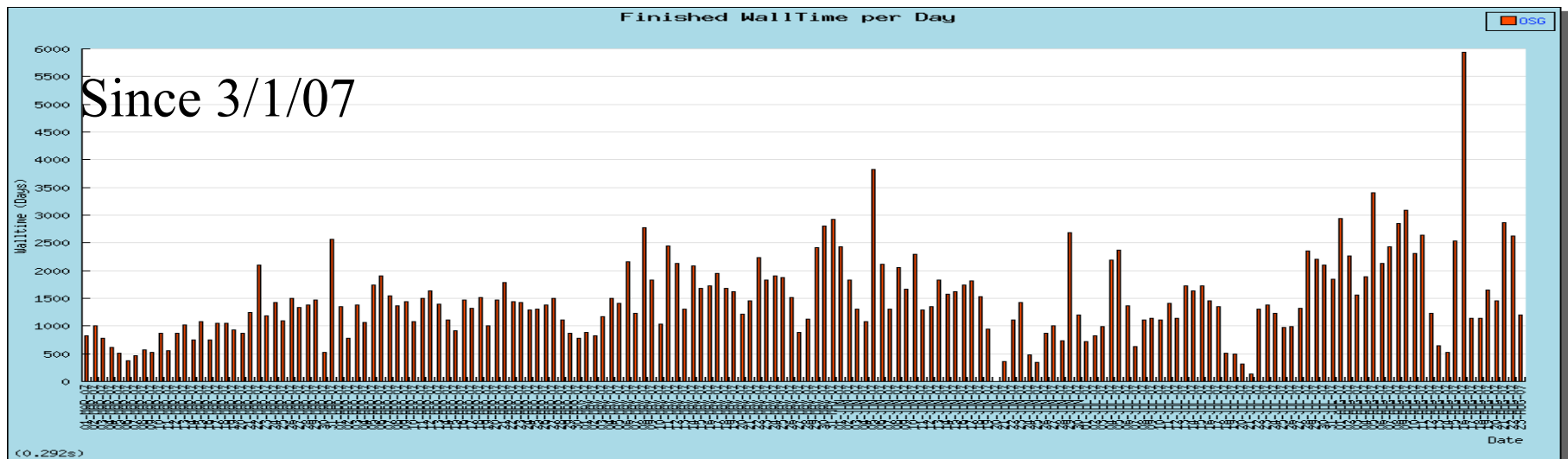
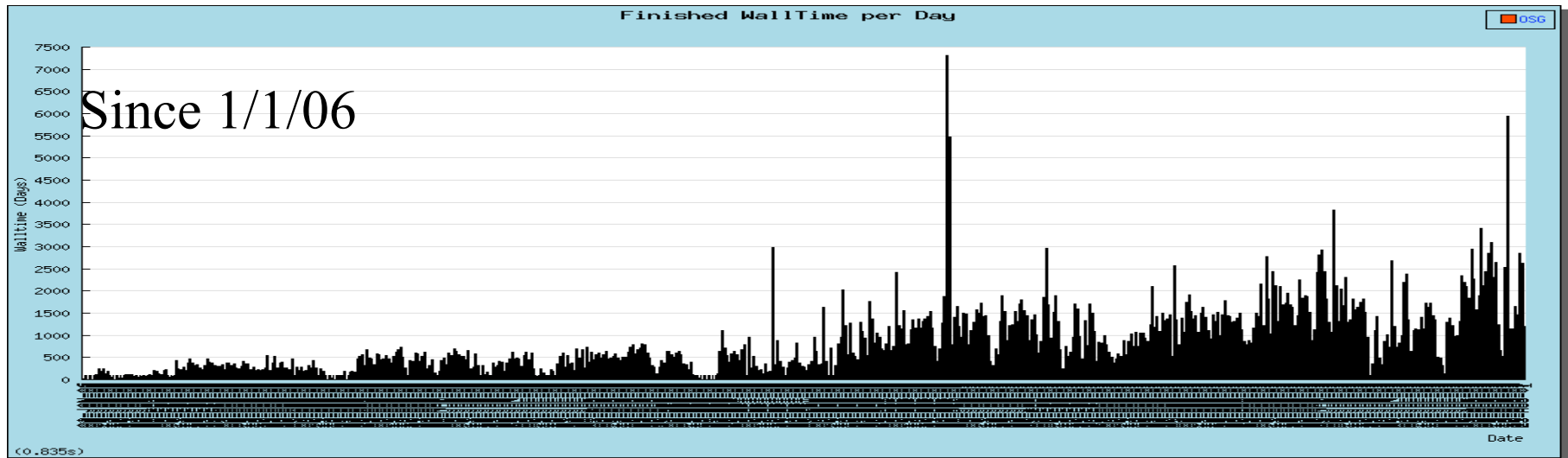
Walltime Usage by Successful Jobs in CPU days (Aug 1-Aug28)



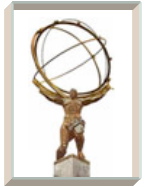
PanDA Daily Waltime Usage



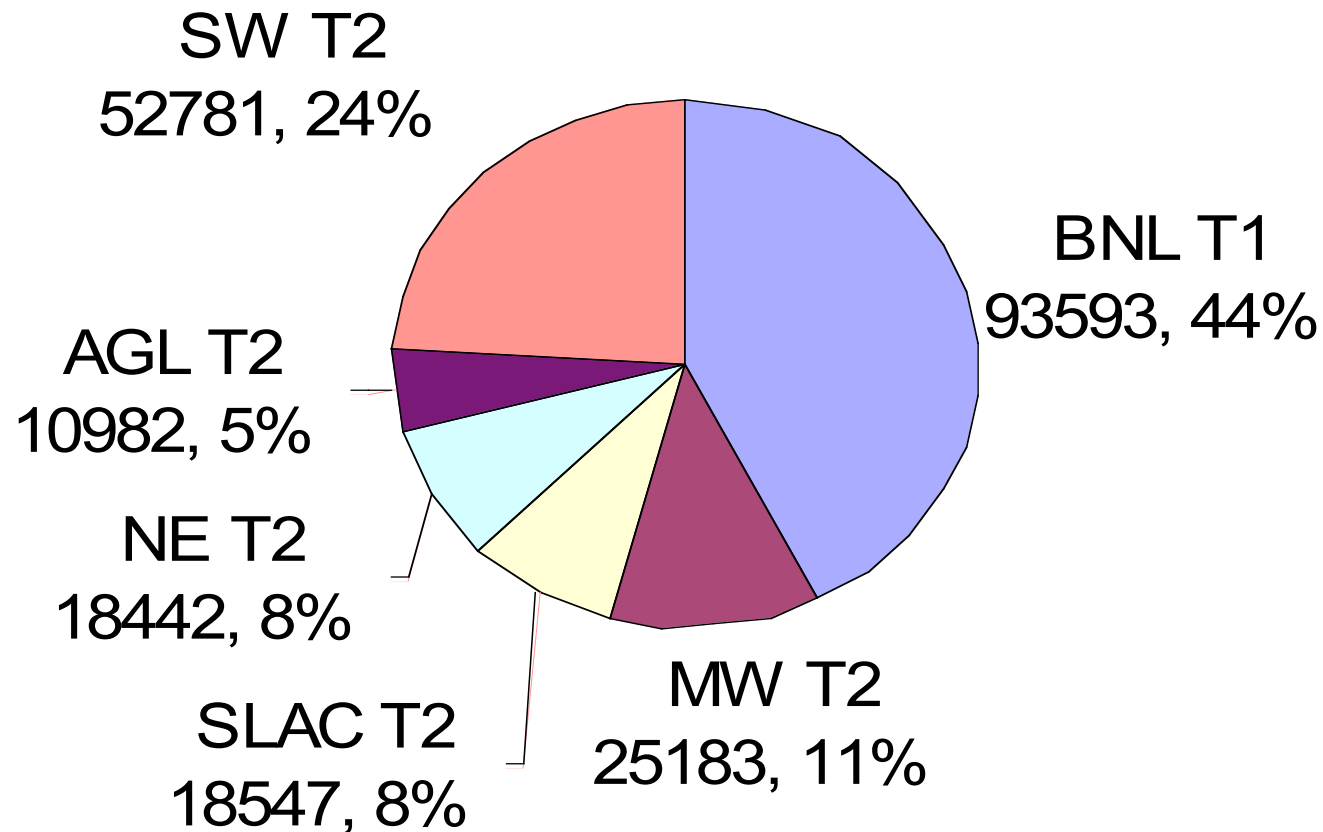
Walltime (Days)



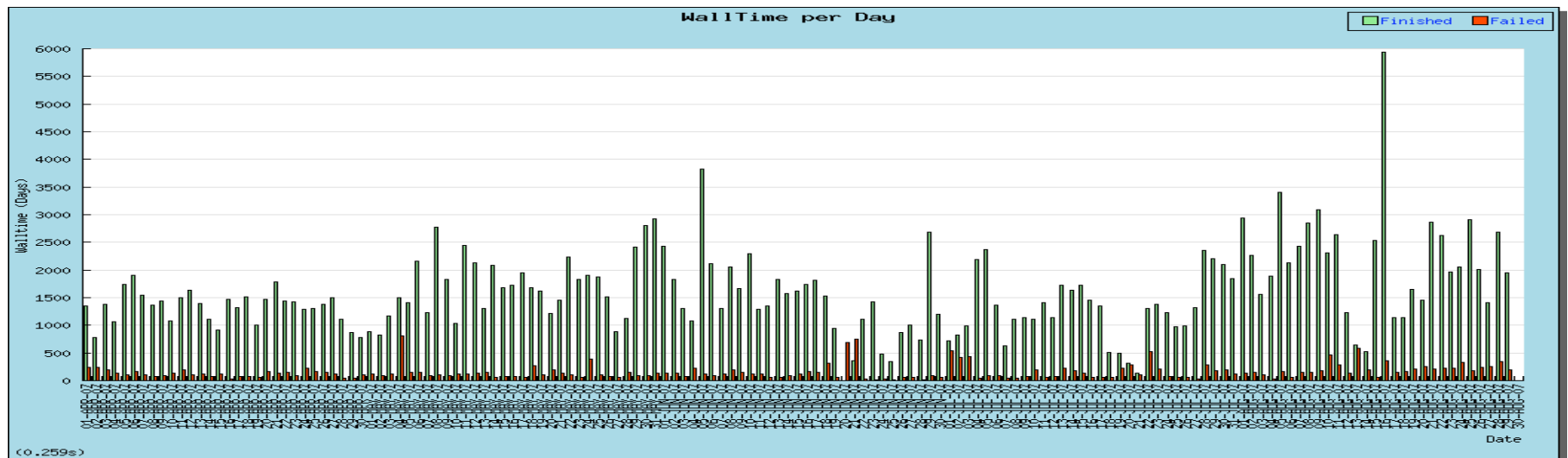
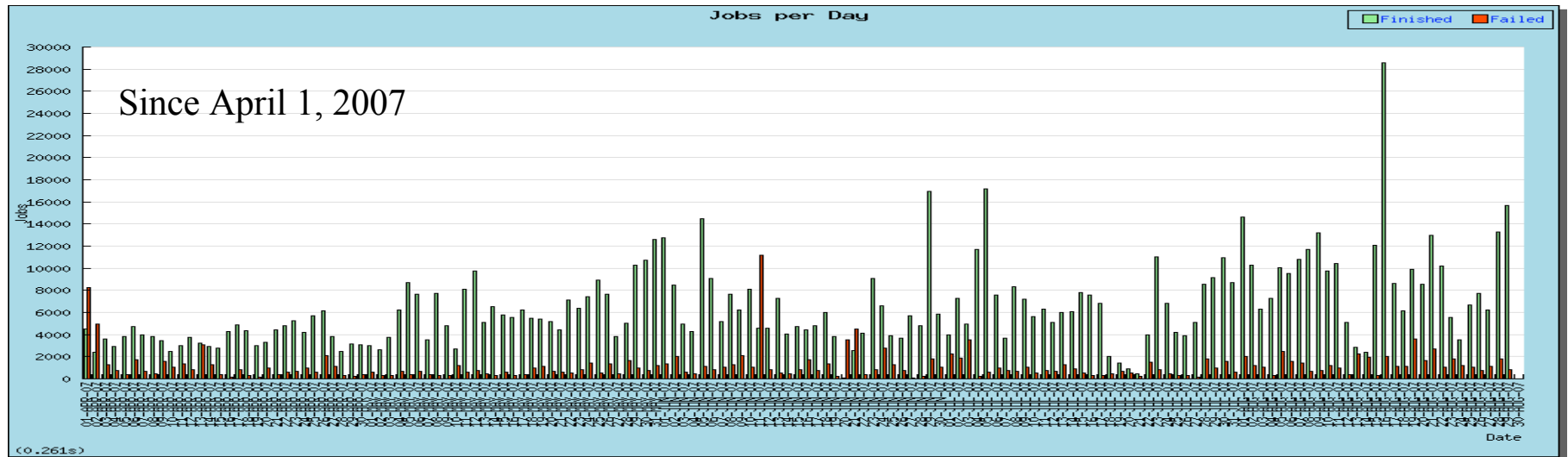
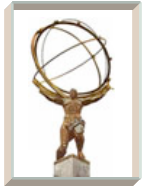
ATLAS Production in U.S. Cloud



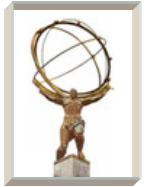
Walltime Usage (CPU days) since Jan 2006



Job/Walltime Efficiencies for Panda



Job/Walltime Efficiencies for Panda



❑ Since January 1, 2006

- ❑ 1950890 production jobs completed successfully
- ❑ 495413 production jobs failed
- ❑ Overall job efficiency is 79.74 %
- ❑ Overall walltime efficiency is 80.04%

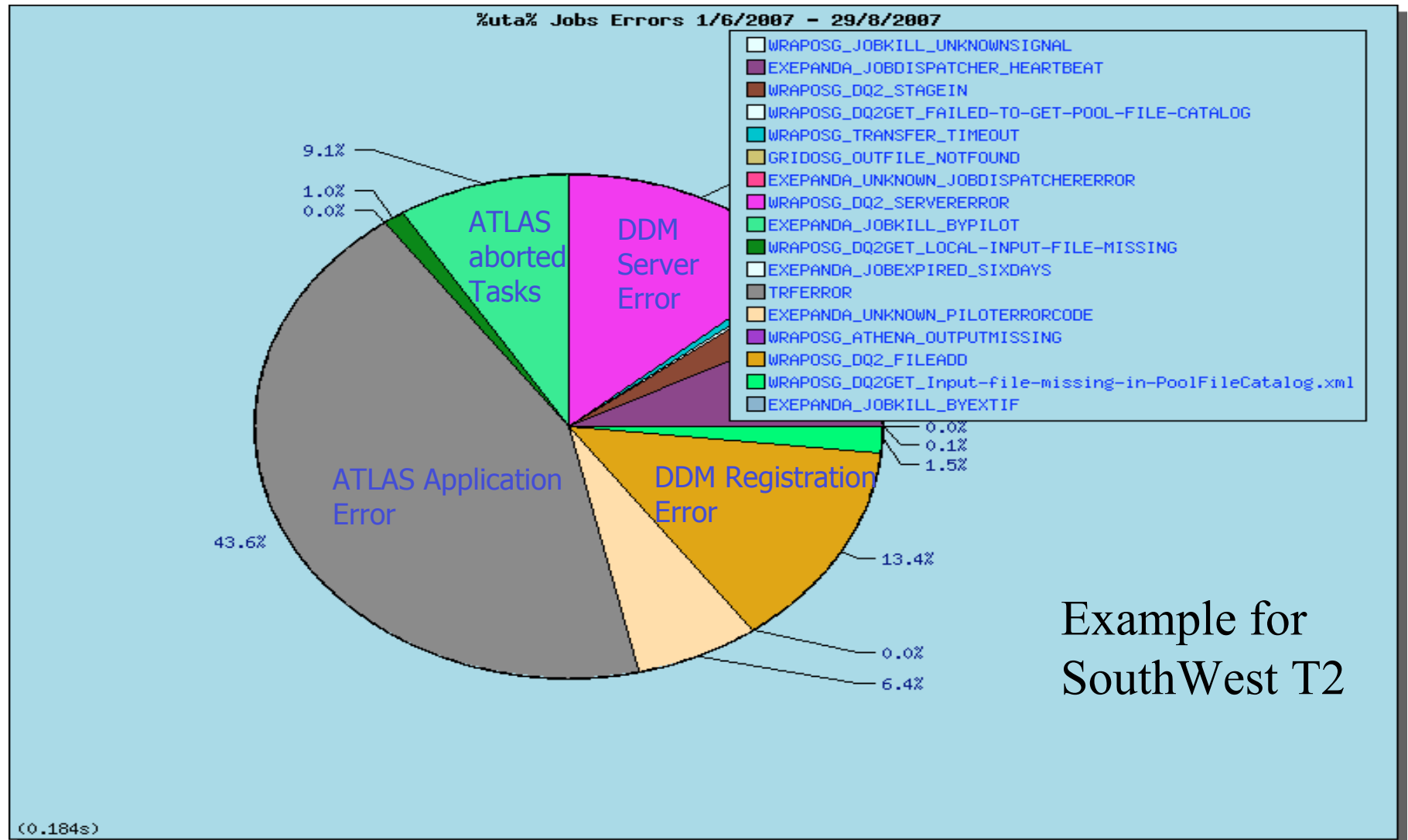
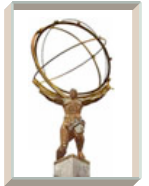
❑ Since April 1, 2007

- ❑ 1069453 successful jobs
- ❑ 191750 failed jobs
- ❑ 84.79% job efficiency
- ❑ 90.61% walltime efficiency

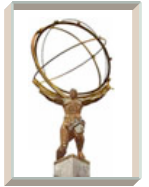
❑ During the last week

- ❑ 104032 jobs finished (only 18532 failed)
- ❑ Average per day – 14861 successful jobs

Job Failure Error Analysis



Conclusion



- ❑ ATLAS computing systems are in good shape – through long running commissioning (CSC) exercise
- ❑ 2006 – successful transition into operations mode
- ❑ 2007 – high volume production and DDM operations
- ❑ Successful distributed MC production for physics studies
- ❑ But many new challenges to come
- ❑ OSG production has achieved robust operations model
- ❑ PanDA works well
- ❑ Overall, progressing well towards full readiness for LHC data