



Panda Experience with DDM

Kaushik De

DDM Workshop, CERN

January 26, 2007

Background



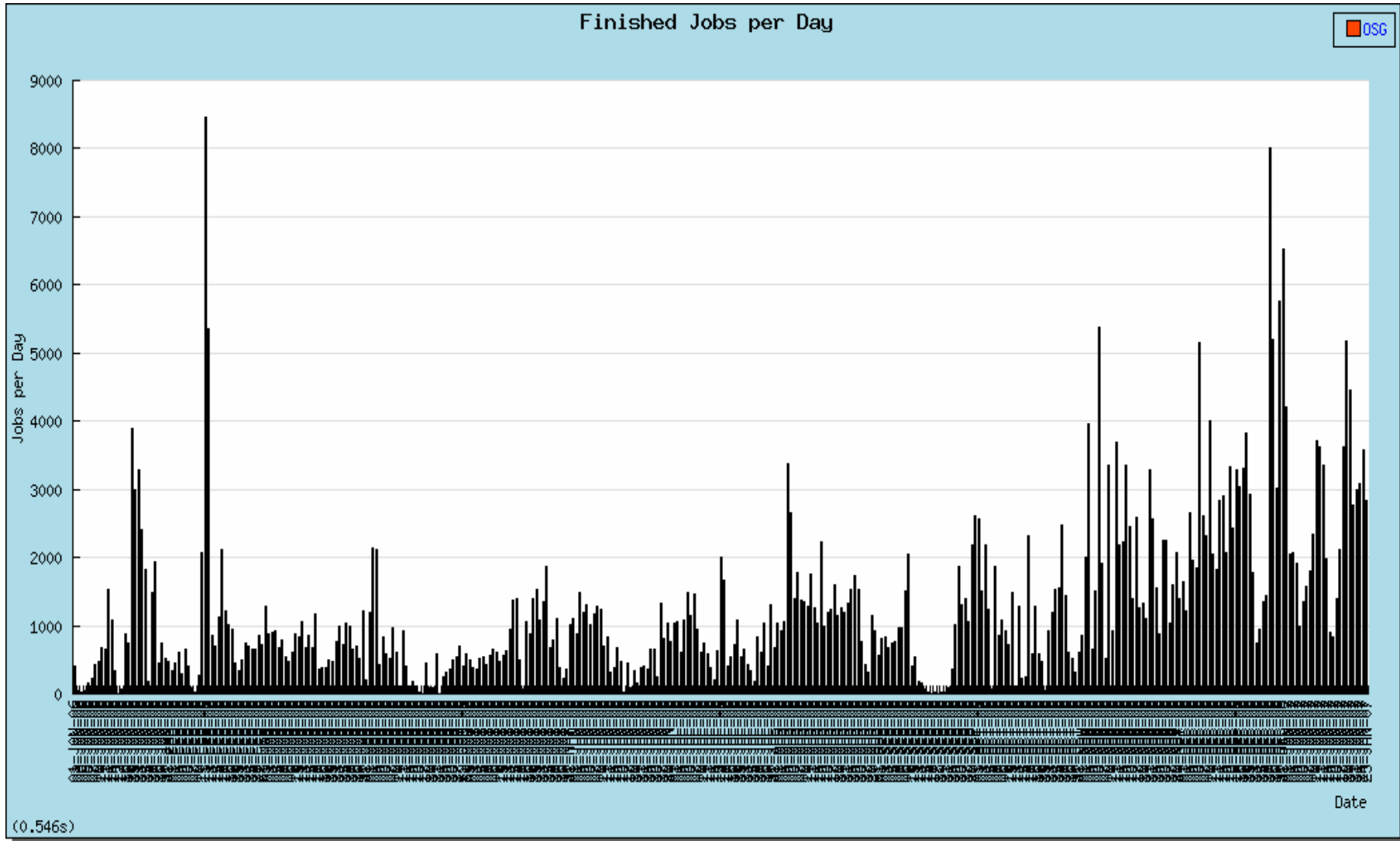
- ❑ This talk is **not** about how Panda is integrated with DDM (see many previous talks) – this is about DDM operations
- ❑ **DQ2 has been used in Panda production >14 months**
- ❑ Originally, Panda used a central catalogue at BNL
- ❑ Since ~Feb, 2006, Panda uses DQ2 catalogue at CERN
- ❑ Since ~June, 2006, Panda uses DQ2 version 0.2.x
- ❑ **All Panda file transfers are done using DQ2 subscriptions**
- ❑ **All Panda production sites run local site services**
- ❑ **Panda (Tadashi) helped to develop early dq2 user tools**
- ❑ Panda uses MySQL LRC implementation for file catalogues
- ❑ Not all sites use SRM/dcache (yet)

Good Experience



- ❑ Currently, BNL LRC has 2 million entries!
- ❑ Panda uses dozen other LRC's at Tier 2 sites
- ❑ DQ2 in continuous use for T1->T2 and T2->T1 transfers
- ❑ In addition to 'physics' datasets, Panda uses smaller datasets to transfer files through subscriptions – we have experience with creating >100k datasets in Panda
- ❑ Daily, DQ2 handles subscriptions for at least 5,000 files for Panda (with peaks of 50,000 files during reconstruction)
- ❑ Daily, DQ2 handles subscriptions for ~1,000 datasets to/from all Panda sites
- ❑ Error rate is much less than 1%

Panda Daily Job Rate in 2006



Still Learning



- ❑ Panda now fails jobs if 'transfers' to BNL fail – increasing latency and obscuring some problems
- ❑ How to find and solve problems with subscriptions (DQ2 monitoring has been very useful – but needs improvement)
- ❑ Optimization of subscription handling – when under heavy load of subscriptions (clearing backlogs, during recon jobs)
- ❑ Discovering and debugging rare missing files
- ❑ Better communication and collaboration with DQ2 developers (we do it well already, I think – but always strive for more)
- ❑ Customizations for Panda use (these were in original DQ2 design, but not used on EGEE) – LRC, http interface...

Still Work To Do



- ❑ Why do some transfers fail terminally (much less than 1%)?
- ❑ How to quickly discover site problems?
- ❑ How to improve subscription handling – under heavy load?
- ❑ Solving problem of too many open subscriptions (cannot scale indefinitely)
- ❑ Getting files timely and reliably from other T1 sites
- ❑ Better reporting and differentiation of errors (interactions between Panda, DQ2, LRC, FTS, dcache is complex)

Recent Work/Improvements



- ❑ Robust Pacman based installation procedures (will upgrade soon to 0.2.12 from 0.2.11)
- ❑ Increasing PFN limit from 256 to 512 characters (it was easier to change MySQL version and table structure than to convince users not to use pfn for all metadata)
- ❑ Removing poolFC dependency (soon)
- ❑ Various LRC schema and indexing improvements
- ❑ In conclusion – Panda has been active partner in DQ2 from the beginning, and invests heavily on DQ2 success