Tokyo Tier-2 Site Report

Tomoe Kishimoto
ICEPP, The University of Tokyo

Apr. 25 2017
International Center for Elementary Particle Physics

R&D for ILC

MEG at PSI

TGC (KEK, Tokyo, TMU, Sinsyu, Nagoya, Kyoto, Kobe…)

DAQ (KEK, Sinsyu, Hiroshima-IT, Nagasaki-IAS)

High Level Trigger (KEK, TITeck, TITech, Waseda, Kobe…)

Muon TDC (KEK)

Solenoid (KEK)

SCT (KEK, Tsukuba, TITech, Ochanomizu, Kyusyu, Osaka…)

✅ Tokyo Tier2 is the the only WLCG site in ATLAS-Japan

HEPiX 2017

T. Kishimoto (ICEPP)
ICEPP regional analysis center

✓ Resource overview

- Support ATLAS VO in WLCG (Tier2) and provide ATLAS-Japan dedicated resources (local use)
- Hardwares are leased, and are replaced in every three years
- From Jan. 2016, **4th system** is running
  - ~10000 CPU cores including service instances and ~10 PB disk storage (T2 + local use)

**Single VO and uniform architecture**

✓ Operation team

- 5 university staffs + 2 SEs from company
Status in ATLAS

✓ Fraction of number of completed jobs

3840 CPU cores deployed
6144 CPU cores deployed

✓ Results in the last month:
- Production, 4.4% (Tier2) – 2.4% (All)
- Analysis, 4.7% (Tier2) – 2.9% (All)

← Good contributions

# of ATLAS–J authors ~ 100
# of ATLAS authors ~ 3000

✓ > 99% site availability has been achieved using the 4th system

Contains ambiguities on the multicore jobs
Slot allocation:
analysis : score prod : 8core prod = 20% : 20% : 60%
Recent update on CE and batch system

✓ Migration of the batch system is ongoing
  - Torque/Maui → HTCondor
  - In the last year, we deployed a small cluster of ARC-CE+HTCondor combination in production:

<table>
<thead>
<tr>
<th>CE</th>
<th>Batch system</th>
<th>CPU cores</th>
<th>Job types</th>
</tr>
</thead>
<tbody>
<tr>
<td>CREAM</td>
<td>Torque/Maui</td>
<td>2304 (96 WNs)</td>
<td>Single-core jobs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2304 (96 WNs)</td>
<td>Multi-core jobs</td>
</tr>
<tr>
<td>ARC</td>
<td>HTCondor</td>
<td>1536 (64 WNs)</td>
<td>Single- and Multi-core jobs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>(dynamic partitioning)</em></td>
</tr>
</tbody>
</table>

* Dynamic partitioning has been introduced in HTCondor
Static vs Dynamic partitioning

Static partitioning (Torque/Maui)

Idle CPUs due to either single- or multi-core jobs were not assigned

Dynamic partitioning (HTCondor)
CPU utilization

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>week1,2</td>
<td>week3,4</td>
<td>week1,2</td>
</tr>
<tr>
<td>Torque/Maui</td>
<td>98.8%</td>
<td>93.9%</td>
<td><strong>88.8%</strong></td>
</tr>
<tr>
<td>(Static partitioning)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HTCondor</td>
<td><strong>99.4%</strong></td>
<td><strong>98.0%</strong></td>
<td>94.8%</td>
</tr>
<tr>
<td>(Dynamic partitioning)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Test jobs (i.e. ops job) are overcommitted in HTCondor system

- Improvement of CPU utilization has been observed thanks to the dynamic partitioning.
- HTCondor is stable so far.
  - We plan to migrate all CPUs from Torque/Maui to HTCondor in the near future.
Recent update on database of SE

✔ Disk storage is managed by **DPM**
  - dpm-1.8.10-1
  - Currently, max 6.336 PB is available

✔ Database is **MySQL**
  - Community Edition 5.7.16 (Oracle)
  - DB size has reached 84 GB

✔ Previous configuration of SE

- **lcg-se01.icepp.jp**
  - 48 file servers
  - 132 TB disk array
    - RAID6
  - No redundancy in MySQL database..., risk of producing dark data
MySQL replication

✓ Semi-synchronous replication in MySQL has been implemented
  - Master server is replicated to slave server automatically
    → Can use slave server as new master server when a trouble occurs in master server
  - Daily backup from slave server (takes ~10 mins)
    → No impact on master server performances

- Fusion–IO ioDrive has been attached for database spaces to reduce time for maintenance
- “mysqlrplsync” command is used to check the consistency between master and slave data
- Binary log increases by 8GB per day
ATLAS data management monitor

Transfer efficiency: source is Tokyo

File deletion efficiency

Downtime for the database upgrade

✓ No issues have been observed after the database upgrade
International Network

✓ SINET5
- LHCONE peering for LA (100 Gbps) and London (20 Gbps) available
- LHCONE peering for TEIN in HongKong will be in place

✓ ICEPP⇒UTNet (campus network) was upgraded from 10 Gbps to 20 Gbps
- > 10 Gbps data transfers have been observed after the upgrade
Summary

✓ Tokyo Tier2 with the 4th system is running
  - Providing enough computing resources for ATLAS
  - > 99% site availability is achieved

✓ Migration from Torque/Maui to HTCondor is ongoing
  - Improvement of CPU utilization has been observed by introducing the Dynamic partitioning
    ‣ e.g. ~7% improvement in 2016/12/21 to 2016/12/31

✓ Redundancy in MySQL database has been implemented
  - Reduced the risk of producing dark data

✓ Bandwidth between ICEPP to WAN has been increased from 10 Gbps to 20 Gbps
Backup
Transfer volume