



# Site Report – KISTI, Korea –

Geonmo Ryu  
KISTI GSDC



- Introduction of GSDC
- Network infrastructure of GSDC
- ALICE Tier-1
- CMS Tier-2
- ALICE & CMS Tier-3 Integration farm
- Recent News or issues at GSDC

## Korea Institute of Science and Technology Information (KISTI)

- Government-funded research institute founded in 1962 for National Information Services and Supercomputing
- National Supercomputing Center - Nurion  
( Top500's 15<sup>th</sup> supercomputer/2019 Jun)



## Global Science experimental Data hub Center (GSDC)

- Government-funded project, started in 2009 to promote Korean fundamental research through providing computing power and data storage
- Datacenter for data-intensive fundamental research



**New Supercomputing Building**



**GSDC Computing Room**



- ◎ All GSDC's computing resources moved to the new building
- ◎ All data replicated in a backup safely, then move to the new space
- ◎ Advancement of facility infrastructure  
(Better environment for computing resources: electric power distribution, constant temperature & humidity equipment etc.)

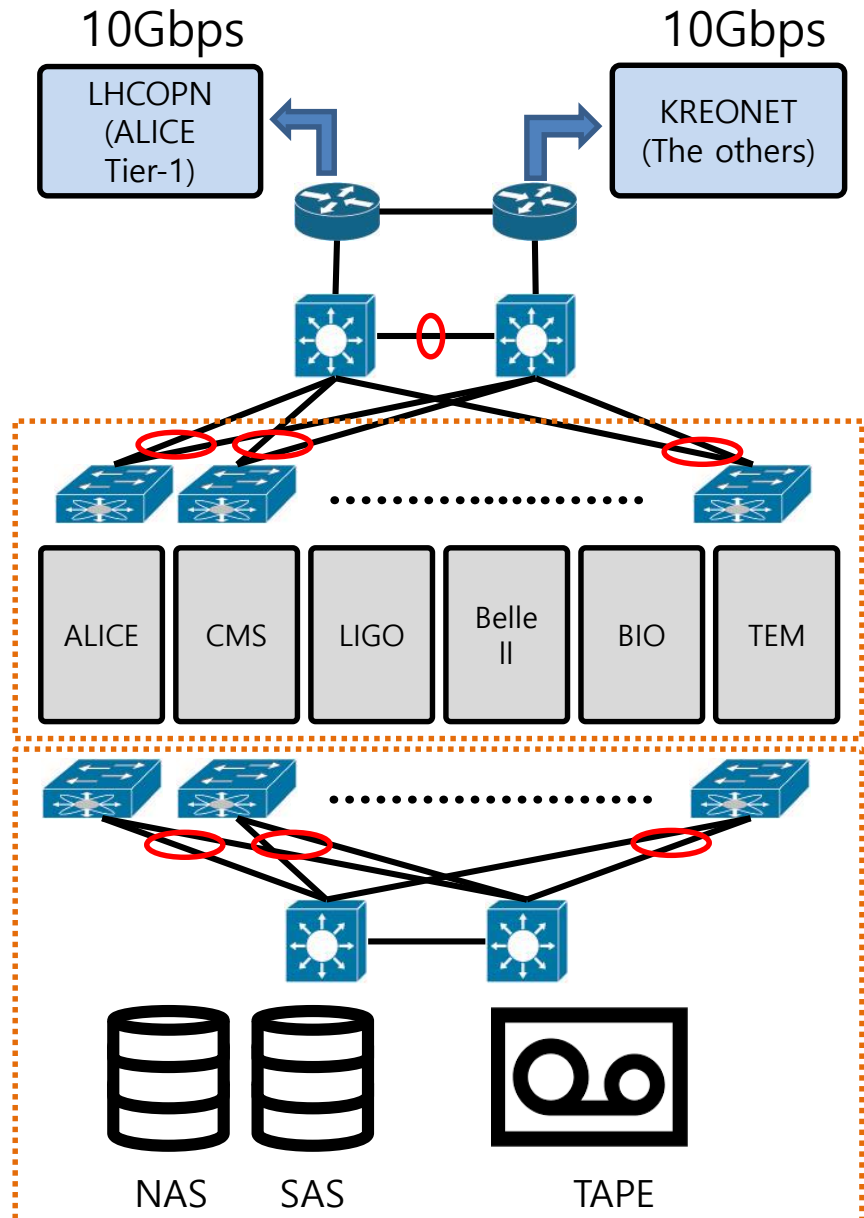
## • Network

### – External : 20Gbps

- Dedicated Network for ALICE Tier-1
  - LHCOPN 10Gbps
- KREONET for other experiments
  - 10Gbps

### – Internal : 80Gbps

- Network connectivity is redundant







## KISTI Tier-1 has been providing reliable and stable service

- Meet the pledge of 2019 (41kHS06 ↑)

52 node X 32 core, 334 HS06/node

38 node X 32 core, 356 HS06/node

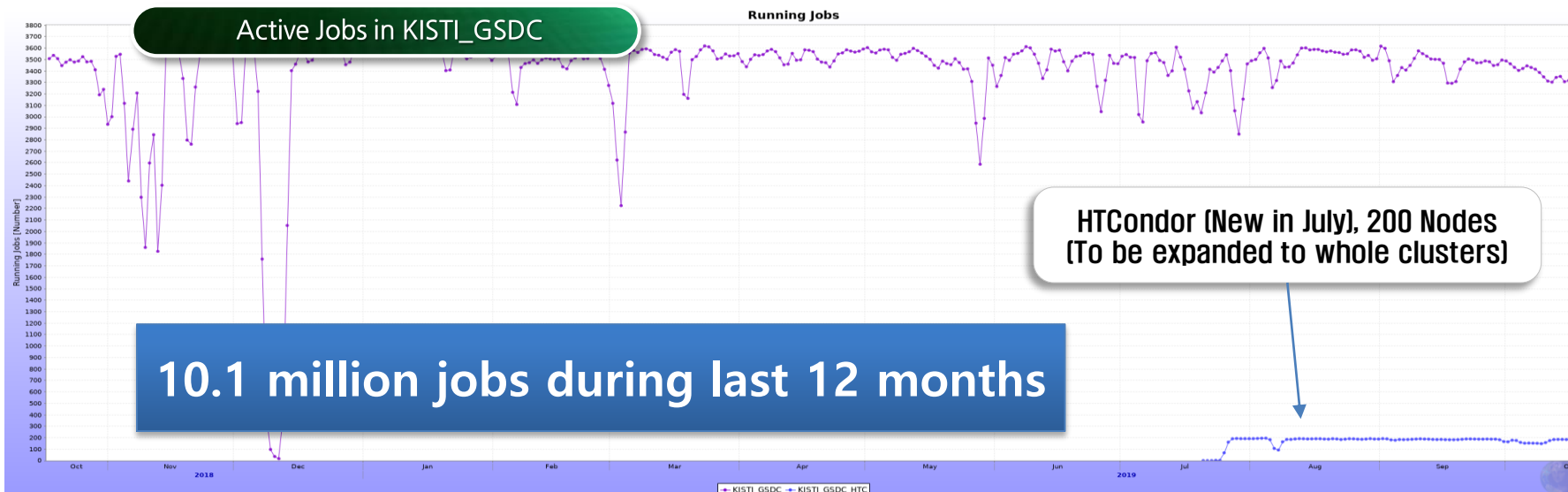
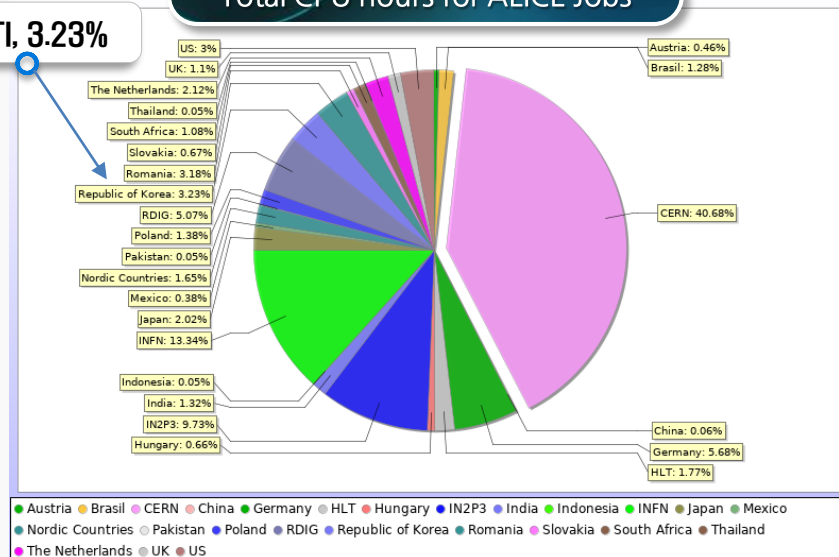
25 node X 40 core, 472 HS06/node

= 41 kHS06

- Max 3,880 concurrent jobs

Total CPU hours for ALICE Jobs

KISTI, 3.23%



## DISK (3,500 TB) / 2PB added in 2018

### DISK (SE2)

- XRootD based data handling (1,500 TB)
- 1 redirector + 5 storage nodes

Usage 1,036 TB, 72%

Read Availability 97.64%

Write Availability 96.79%

### DISK (EOS)

- EOS based data handling (2,000 TB)
- 2 MGMs (duplexing) + 5 FSTs (storage nodes)

Usage 923 TB, 46%

Read Availability 91.21%

Write Availability 83.35%

### TAPE

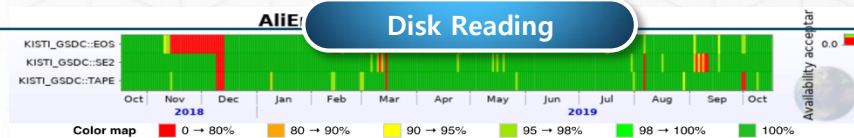
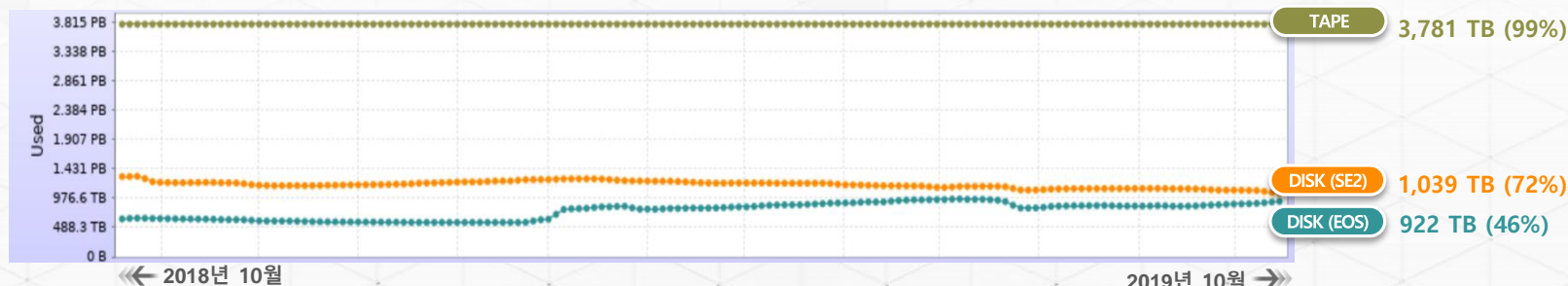
- XRootD & GPFS based data handling (3,800 TB)
- Tape (3,200 TB) + Disk Cache Pool (600 TB)

Usage 3,781 TB, 99%

Read Availability 98.28%

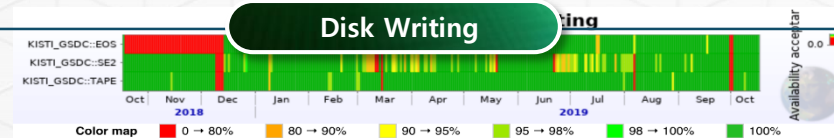
Write Availability 98.23%

## Storage status for last 12 months



### Disk Reading

Statistics						
Link name	Data		Individual results of reading tests			Overall Availability
	Starts	Ends	Successful	Failed	Success ratio	
KISTI_GSDC::EOS	17 Oct 2018 03:08	17 Oct 2019 03:31	8099	766	91.36%	91.21%
KISTI_GSDC::SE2	17 Oct 2018 02:56	17 Oct 2019 03:20	8669	212	97.61%	97.64%
KISTI_GSDC::TAPE	17 Oct 2018 02:59	17 Oct 2019 03:23	8725	150	98.31%	98.28%



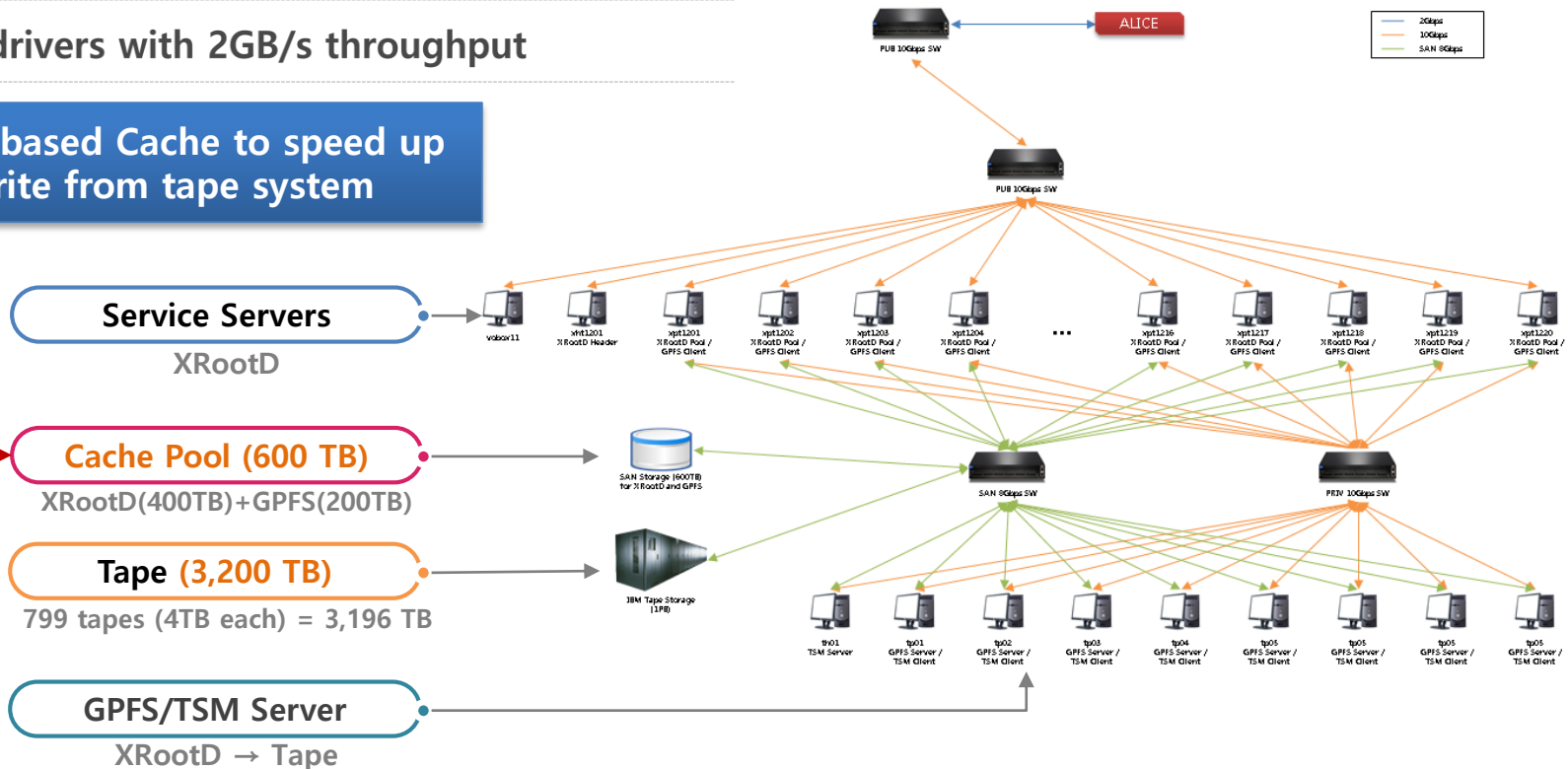
### Disk Writing

Statistics						
Link name	Data		Individual results of writing tests			Overall Availability
	Starts	Ends	Successful	Failed	Success ratio	
KISTI_GSDC::EOS	17 Oct 2018 03:08	17 Oct 2019 03:30	7278	1444	83.44%	83.35%
KISTI_GSDC::SE2	17 Oct 2018 02:55	17 Oct 2019 03:20	8453	283	96.76%	96.79%
KISTI_GSDC::TAPE	17 Oct 2018 02:59	17 Oct 2019 03:22	8577	154	98.24%	98.23%



- 3,200TB TAPE + 600TB Disk cache
- 3,781TB ALICE raw data stored
- 3,471TB RUN2 data transferred
- 8 tape drivers with 2GB/s throughput

Using Disk-based Cache to speed up the read/write from tape system



Tape storage elements														
Cache Pool														
SE Name	AliEn SE	AliEn name	Tier	Size	Used	Free	Usage	No. of files	Type	Size	Used	Free	Usage	Version
1. KISTI_GSDC - TAPE	ALICE::KISTI_GSDC::TAPE		1	387.2 TB	3.781 PB	0	999.9%	2,859,053	FILE	384.6 TB	304.6 TB	80.02 TB	79.2%	Xrootd v4.10.0
Total				387.2 TB	3.781 PB	0		2,859,053		384.6 TB	304.6 TB	80.02 TB		

## Keeping top most quality of services

	Reliability		Availability	
	Overall in 2018	Last 6 Months	Overall in 2018	Last 6 months
ALICE	99.9%	99.5%	98%	99.5%

 Monthly target of WLCG : 97%

Participating in WLCG operation meeting every week,  
closely collaborating with WLCG members

### Monthly Availability/ Reliability (%)

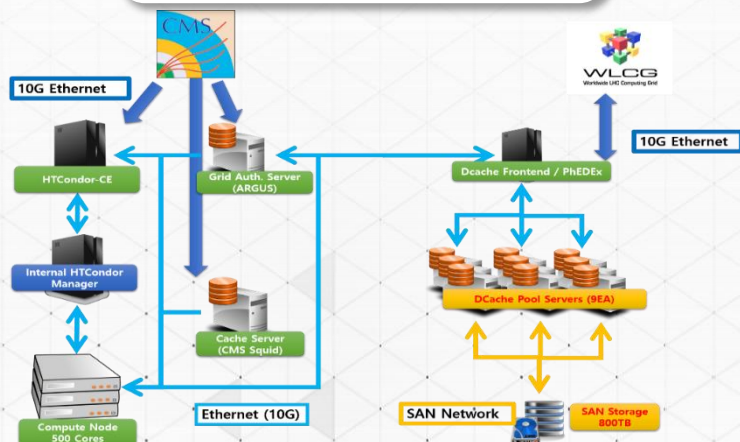
Service	Apr. 2019	May. 2019	Jun. 2019	July. 2019	Aug. 2019	Sep. 2019
Availability	100	100	100	99	98	100
Reliability	100	100	100	99	98	100

$$\bullet \text{ Reliability} = \frac{T_{up}}{T_{up} + (T_{DOWN} - T_{SCHED\,DOWN})}$$

$$\bullet \text{ Availability} = \frac{T_{up}}{T_{up} + T_{DOWN}}$$

- Movement of all GSDC's computing and storage resources to a new building (better environment)

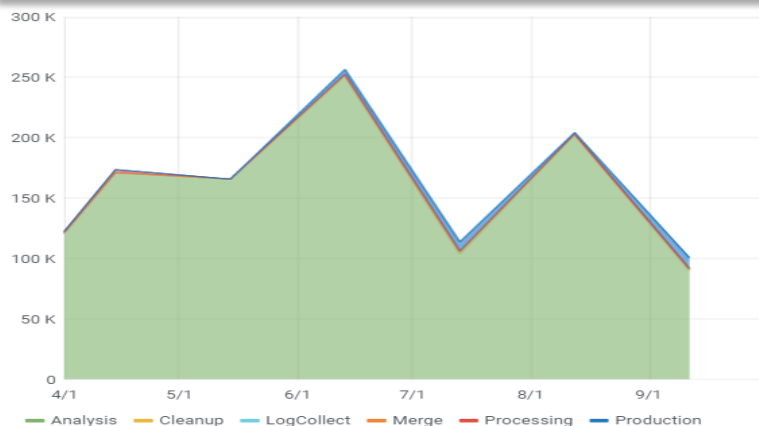
## CMS Tier-2 Infrastructure



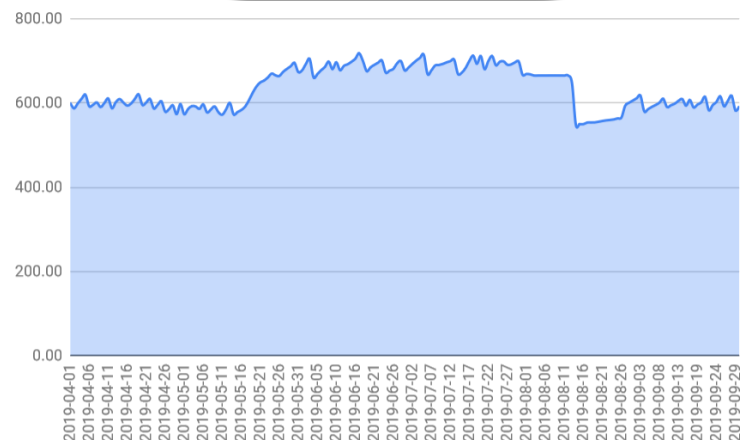
- Computing : Logical **1,000 Cores (10 kHS06)**
- About 1% of Total CMS Tier-2 computing resource

## Job Activities

1.07 million jobs during last 6 months



## Storage Usage



- Disk **800 TB (Usage 74.07%)**

## Data Transmission

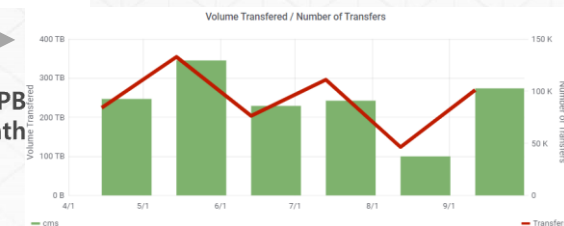


← ○ KISTI Tier-2 Data transmission

- Tier-0 link : 1
- Tier-1 link : 14
- Tier-2 link : 48
- Tier-3 link : 3

## Data Traffic

Total : 1.437PB  
Average : 239.5TB /month



	Reliability		Availability	
	Overall in 2018	Last 6 Months	Overall in 2018	Last 6 months
CMS	98.27%	98.83%	95.59%	98.83%

 Monthly target of WLCG : 95%

## Monthly Availability/ Reliability [%]

Service	Apr. 2019	May. 2019	Jun. 2019	July. 2019	Aug. 2019	Sep. 2019
Availability	100	99	99	99	98	98
Reliability	100	99	99	99	98	98

Federation Availability Reliability

CERN-PROD	99%	99%
HU-HGCC-T2	99%	99%
T2_US_Florida	99%	99%
T2_US_Nebraska	99%	99%
T2_US_Purdue	99%	99%
T2_US_Wisconsin	99%	99%
BR-SP-SPRACE	98%	98%
ES-CMS-T2	98%	98%
FI-HIP-T2	98%	98%
FR-GRIF	98%	99%
FR-IN2P3-CC-T2	98%	99%
FR-IN2P3-IPHC	98%	99%
KR-KISTI-GSDC-02	98%	99%
T2_US_Caltech	98%	98%
UA-Tier2-Federation	98%	100%
PK-CMS-T2	97%	97%
T2_US_MIT	97%	97%
TW-CMS-T2	97%	97%

Federation Availability Reliability

DE-DESY-RWTH-CMS-T2	96%	96%
GR-Ioannina-HEP	96%	97%
UK-London-Tier2	96%	98%
T2_US_UCSD	95%	95%
UK-SouthGrid	95%	95%
AT-HEPHY-VIENNA-UIBK	94%	94%
IT-INFN-T2	94%	94%
IN-INDIACMS-TIFR	93%	94%
BE-TIER2	92%	95%
CH-CHIPP-CSCS	92%	93%
CN-IHEP	91%	97%
RU-RDIG	91%	93%
PT-LIP-LCG-Tier2	87%	87%
EE-NICPB	86%	88%
PL-TIER2-WLCG	82%	84%
TR-Tier2-federation	80%	80%
T2-LATINAMERICA	78%	78%

- Documentation about HTCondor-CE and APEL on UMD
  - CHEP poster
    - <https://doi.org/10.1051/epjconf/201921408020>
  - “Ansible” tutorial code using vagrant
    - [https://github.com/geommo/htcondorce\\_on\\_vagrant](https://github.com/geommo/htcondorce_on_vagrant)

Open Access

Issue	EPJ Web Conf. Volume 214, 2019 23 <sup>rd</sup> International Conference on Computing in High Energy and Nuclear Physics (CHEP 2018)
Article Number	08020
Number of page(s)	7
Section	T8 - Networks & facilities
DOI	<a href="https://doi.org/10.1051/epjconf/201921408020">https://doi.org/10.1051/epjconf/201921408020</a>
Published online	17 September 2019

EPJ Web of Conferences 214, 08020 (2019)  
<https://doi.org/10.1051/epjconf/201921408020>

## Establishment of new WLCG Tier Center using HTCondor-CE on UMD middleware

Geommo Ryu\* and Seo-Young Noh

Korea Institute of Science and Technology Information, 245 Daehak-ro, Yuseong-gu, Daejeon, 34141, Korea

\* e-mail: [geommo@kisti.re.kr](mailto:geommo@kisti.re.kr)

Published online: 17 September 2019

geommo / htcondorce\_on\_vagrant

Unwatch 1 Star 0 Fork 0

Code Issues 0 Pull requests 0 Projects 0 Wiki Security Insights Settings

No description, website, or topics provided.

Manage topics

47 commits 1 branch 0 releases 1 contributor

Branch: master New pull request Create new file Upload files Find file Clone or download

File	Commit Message	Latest commit
dummy_apel_log	Add script for dummy log.	5 days ago
files	Add apel configuration.	6 days ago
invendir	Fix wrong machie group.	11 days ago
templates	Add bdii configuration.	5 days ago
test	Add order and limit.	5 days ago
vars	Move variables to vars dir.	5 days ago
00_install_role.sh	Patch for other sites.	5 days ago
01_hosts.yml	Patch for other sites.	5 days ago
02_epel.yml	Update ansible before middleware installation.	11 days ago
03_updateAnsible.sh	Align script's order.	10 days ago
04_middleware.yml	Move signing_policy from htcondor-ce to middleware.	5 days ago
05_htcondor-ce.yml	Add bdii configuration.	5 days ago
06_htcondor.yml	Move variables to vars dir.	5 days ago
07_custom.yml	Add seprate file to sub dirs.	7 days ago
08_apel.yml	ScallFactor to 1.0	5 days ago
README.md	Update README.md	5 days ago
ansible.cfg	change ansible roles.	last month
create-all.sh	Add apel configuration.	6 days ago
extensions.cnf	Add some feature to replace configuration for remote LRMS.	10 days ago
remote_install_rootca.sh	Add stricky option for scp command.	5 days ago
setup_rootca.sh	change ansible roles.	last month



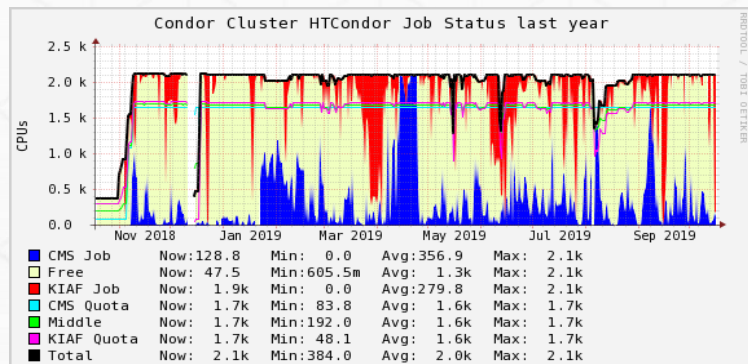
## Dedicated to KoALICE & KCMS community

- CPU 1,028 cores, Storage 1,604TB
- Integrated at Nov. 2018
- Bi-weekly user meeting (vidyo, each group)

### Job Processed

	2017	2018(~Oct)	2018(Nov.~)	2019(~Jun.)	2019.(Jul.~)
KoALICE	1,974,480*	336,922	113,915	432,974	256,402
KCMS	3,093,847	4,642,578	1,324,121	2,531,687	460,151

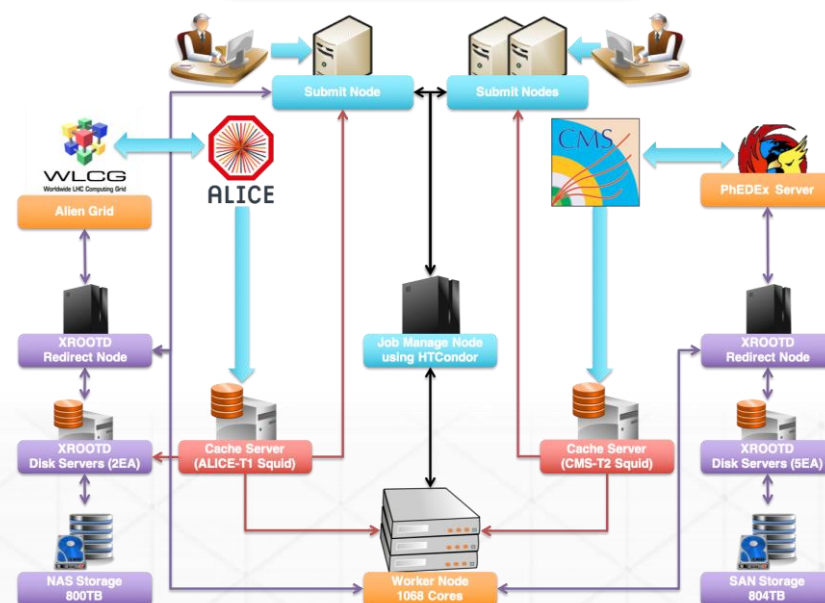
\* Jobs has been more than usual due to the simulation



### Storage Usage

	Total	Used	Free	Usage
KoALICE	800 TB	467.0 TB	333.0 TB	58.37%
KCMS	804 TB	661.8 TB	142.2 TB	82.32%

### ALICE&CMS Tier-3 system



### User Accounts

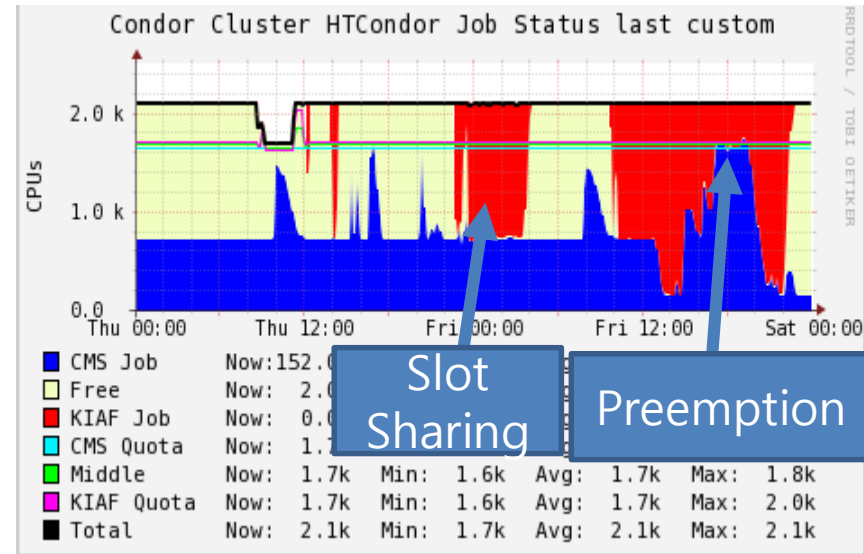
#### KoALICE

Univ.	Inha	Pusan	Yonsei	Total
#	13	6	8	27

#### KCMS

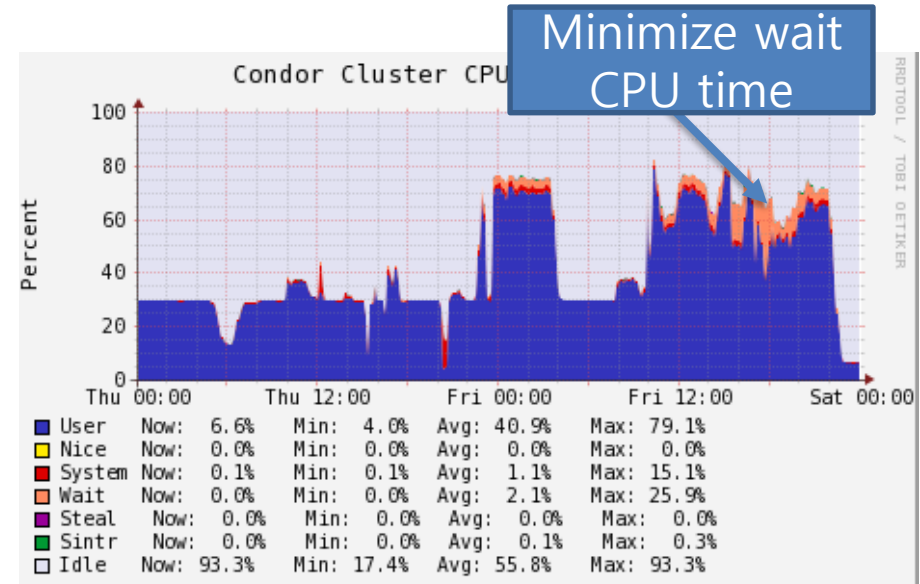
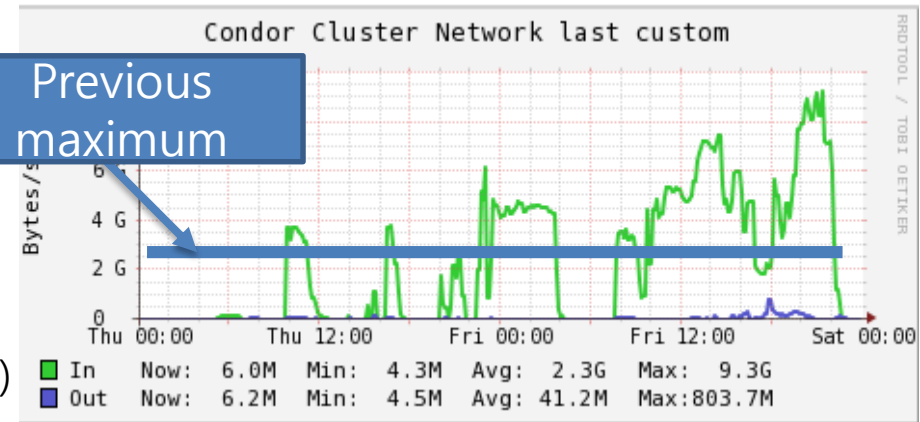
Univ.	KNU	KU	SNU	SKKU	UOS	KHU	HYU	JNU	SJU	Total
#	19	13	19	11	19	2	8	7	4	102

- Group Quota
  - Slot sharing (400 vs 1656) and preemption
- Static and dynamic slots
  - Divided the resource into three slots 2:1:1
    - Hyper-threading enabled 2 CPU systems must have a multiple of four cores.
  - 3x the number of slot preemption when negotiator cycle time.
- Singularity Container
  - To solve OS difference
    - SL6 vs CentOS7



```
NUM_SLOTS = 3
NUM_SLOTS_TYPE_1 = 1
SLOT_TYPE_1 = 50%
SLOT_TYPE_1_PARTITIONABLE = TRUE
NUM_SLOTS_TYPE_2 = 1
SLOT_TYPE_2 = 25%
SLOT_TYPE_2_PARTITIONABLE = TRUE
NUM_SLOTS_TYPE_3 = 1
SLOT_TYPE_3 = 25%
SLOT_TYPE_3_PARTITIONABLE = TRUE
```

- Problem occurs
  - Problem with data access speed through WN sharing
- Solution
  - Abandon access system using XRootD due to lack of machine ( Only 2 machines available for XRootD Server )
  - To acquire maximum performance of our NAS system, NFS directory is manually mounted for each WN machine
    - Data was migrated to a single volume but distributed across NAS devices
- Current Status
  - Peak read performance close to the speed limit 80Gbps (40 x 2 racks) of ToR switches ( 9.3GB/s = 74.4Gbps )



- Issue 1. NAS + SAN storage for backend
  - Problem : Manage 604TB of SAN storage and 200TB of NAS storage as a single XRootD volume
  - Current : Can be integrated by configuration but disables xattr functionality
    - Trial #1 : Used nfs\_xattrs(FUSE for NFS with xattr) but it was too slow to operate
- Issue 2. User Access Control
  - Problem : Using "acc.authdb" feature solves problems with xrdfs or xrdcp. However, problems can arise with fuse mounts through xroodfs
  - Current : Created a directory for each user and automounted their storage repository directory under that directory. This allows users to access only their own directory, so there is no possibility of accidentally deleting or changing other users' data.
  - Limitation : However, difficulties exist in using public spaces.

- KISTI-GSDC operates WLCG ALICE Tier-1, CMS Tier-2 and ALICE & CMS Tier-3 centers.
- ALICE Tier-1 and CMS Tier-2 meet the service levels requested by the WLCG.
- Since CREAM CE is scheduled to be terminated, the introduction of HTCondor-CE is in progress or completed.
- A small achievement was achieved by integrating separate clusters using only HTCondor.
- There are still many problems to solve. We need the help and feedback of other Asian centers to solve our problems quickly.