ASGC site report

Felix Lee
Academia Sinica Grid Computing Centre (ASGC), Taiwan

HEPiX Spring 2013

Resources update in ASGC

- Resource update since Q4 2012.
  - CPU: Dell C6220 * 16
    - It's 4 nodes in 2U enclosure, so, there are 64 nodes in total.
    - E5-2630L, hexa-core, 48GB memory, 1TB H.D.D
    - HEPSpec2k6: 220.42 per node (with HT)
  - Dell MD3660f + 2 * MD3060e(JBOD)
    - 3TB * 180 hard drives, 540TB raw capacity.
  - Promise Vtrak E830fD, 2 header, 18 JBOD.
    - 3TB * 240 hard drives, 720TB for each, 1.44 PB raw capacity in total.

Procurement plan

- We would focus on CPU and edge switch this year.
- Ongoing procurements
  - Purchasing another Dell C6220 * 14.
- Survey
  - 10GbE edge switch
    - Testing Dell PowerConnect 8164F.
  - E5-4600 series platform
    - 4 way system
    - Borrowed IBM x3750M4 from vendor for evaluation.
  - Supermicro server battery module solution.

WLCG service update - Grid

- All gLite components are in EMI-2 now
  - BDII, CREAM, DPM, worker node..
- DPM xrootd and https are now in production.
- Reconfiguring LVM binding for disk only instance(DPM)
  - For performance wise, we used to bind several RAIDs to be a huge and striping LVM volume, so we can get aggregated throughput, but it's also big size(120TB).
  - So, the failure can be also huge if any..
  - So, we decide to unbind LVM and reconfigure each RAID to be individual partition.
  - A lot of draining works to do..

WLCG service update - CASTOR

- Upgraded Oracle DB from 10G to 11G.
- Upgraded CASTOR.
  - 2.1.11-9 → 2.1.13-9
    - Suffered some SRM problems after CASTOR was upgraded to 2.1.13-9
    - The problem was caused by corrupted request id in DB.
      - The request id regressed and was reused to different files, this caused SRM suffering segmentation fault.
  - Very nice tape handling in 2.1.13-9.
    - Repack in 2.1.11-9 could only complete 4~8 LTO4 tapes per day in ASGC.
    - We can now repack 30+ LTO4 tape per day.
  - But, some issue with tape migration.
    - Some tape medium errors caused file being dropped by Castor, although it's not supposed to be corrupted file.
    - This only happened after we upgrade to 2.1.13-9, due to CASTOR implements new migration error handler since 2.1.13, which will drop corrupted files to avoid failed migration attempt, but it unfortunately got wrong action with our tape system...
    - We lost 55 files because of that..
    - The hotfix was immediately released by CASTOR team.

Network service update/issues

- Enable jumbo frame for our disk servers
  - Set frame size to be 9000 for most of our disk servers.
  - But it suffered error to everywhere including our local data center, until we adjusted frame size to be 9100 to our edge switches and router..
- Poor performance to most of EU sites.
  - Large packing drop between Chicago and Amsterdam router interface was observed.
  - Probably caused by hardware..., would try to upgrade interface

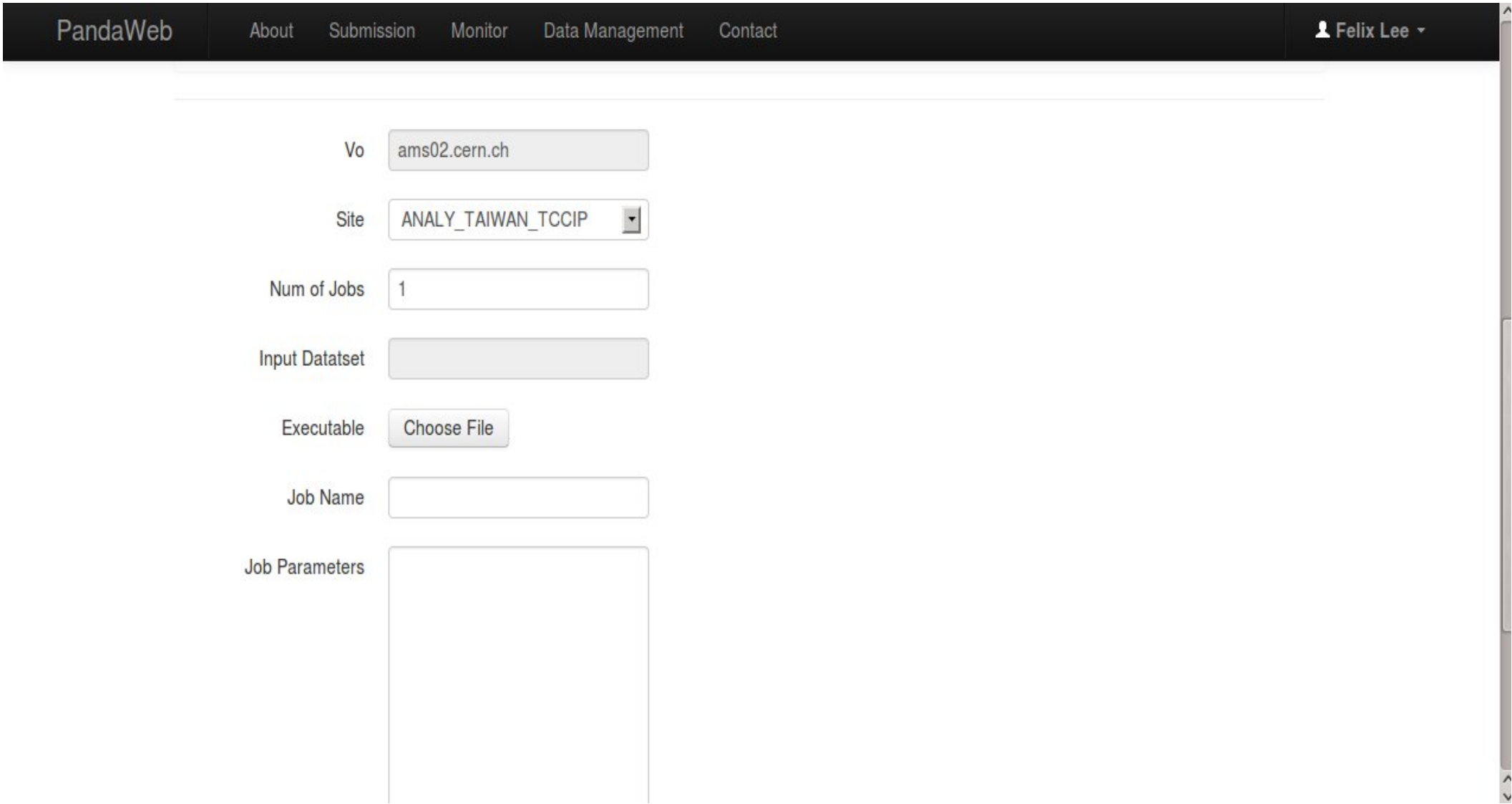
Cloud service update

- OpenNebula 3.8 → 3.8.1
  - 960 cores in HP BL460G6, max to 1920 VMs with HT
  - No change since 2012 Q3
- OpenStack Folsom (new deployment)
  - Deployed to Dell C6220.
  - 648 cores, max to 1296 VMs with HT.

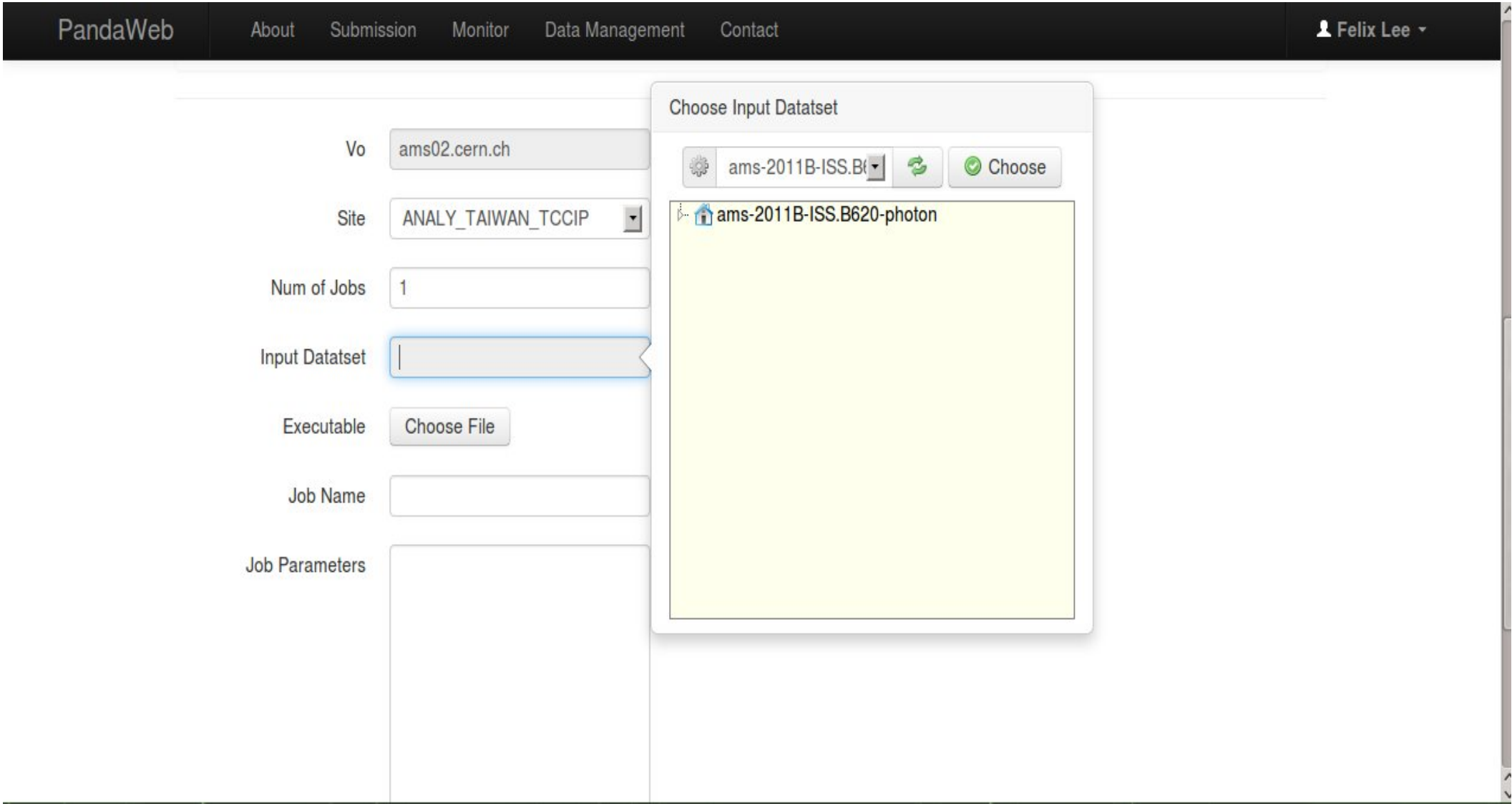
Recent activities else

- Integration of PanDA pilot framework.
  - Setup our own PanDA instance in Taiwan.
  - Developing a web UI for user job submission.
- Integrating also Atlas RUCIO as data management solution
  - User can also manage their data by RUCIO through web UI.
- Integrating Cloud resource into PanDA pilot framework.
  - No additional CE, no additional batch, pilot worker will go to VM then pull jobs directly.
- Developing and deploying single rack H/W solution

Panda Web – Define job



Panda Web – Define Input file



Panda Web – Manage jobs

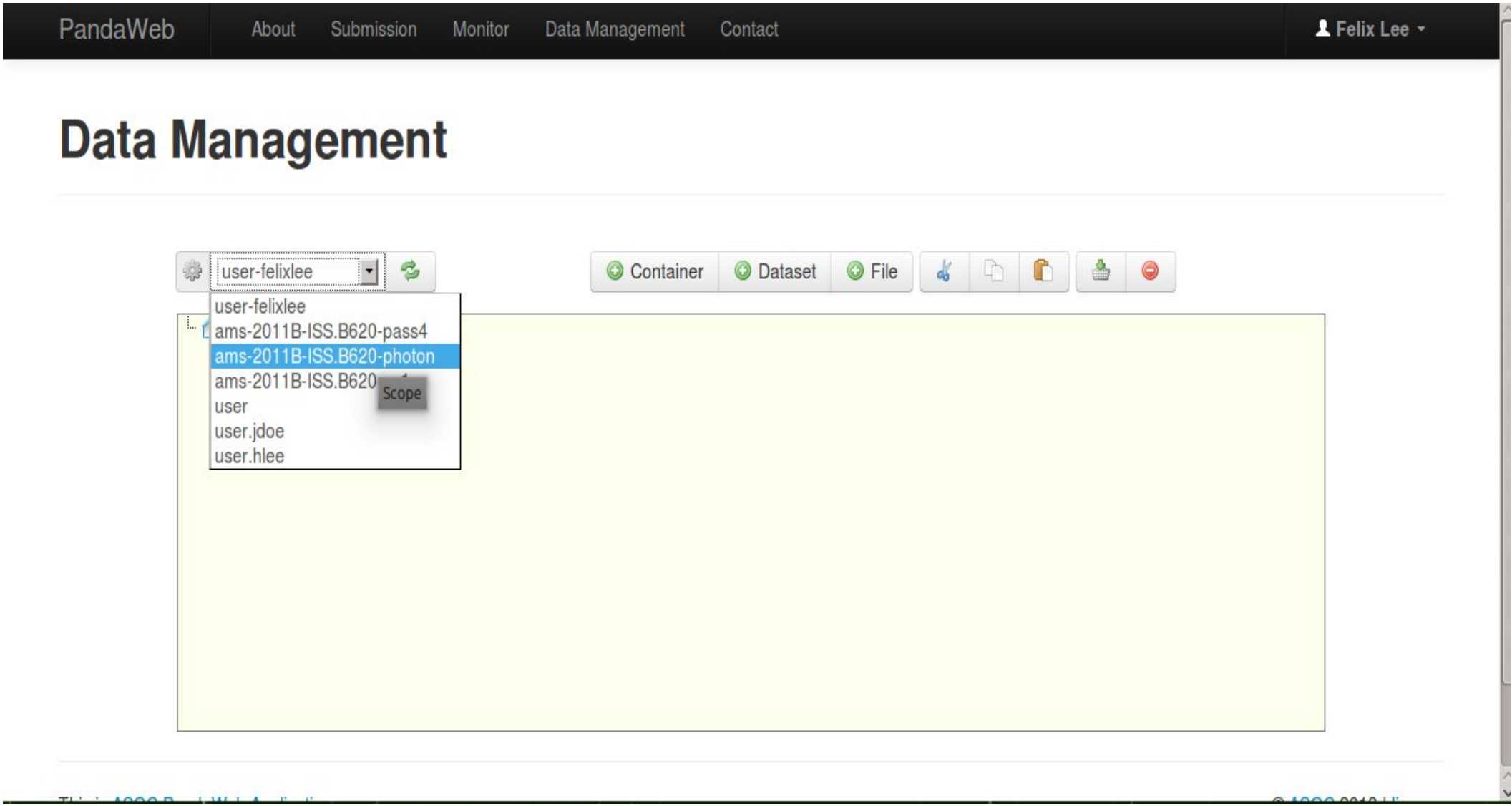
PandaWeb | About | Submission | Monitor | Data Management | Contact | Felix Lee

Display maximum 2000 rows

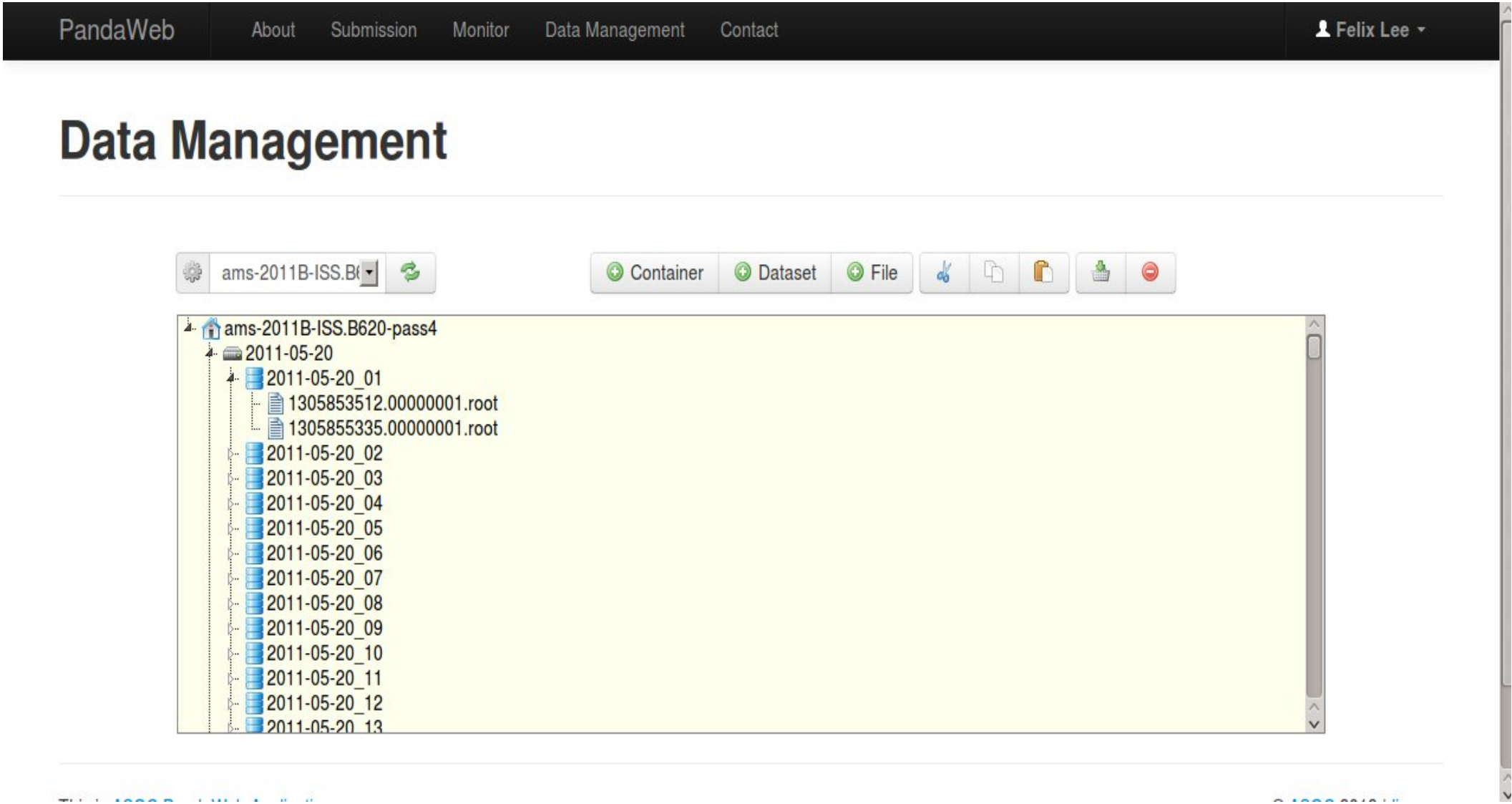
Select	PandaID	Job Name	User Name	Queue	Create Time	Status
	1471770742	el.B598-3_HPC_taiwan.872464...	Felix Dteam Lee	ANALY_TAIWAN_HPC	April 12, 2013	running
	1471770741	el.B598-3_HPC_taiwan.872464...	Felix Dteam Lee	ANALY_TAIWAN_HPC	April 12, 2013	running
	1471770740	el.B598-3_HPC_taiwan.872464...	Felix Dteam Lee	ANALY_TAIWAN_HPC	April 12, 2013	running
	1471770739	el.B598-3_HPC_taiwan.872464...	Felix Dteam Lee	ANALY_TAIWAN_HPC	April 12, 2013	running
	1471770738	el.B598-3_HPC_taiwan.872464...	Felix Dteam Lee	ANALY_TAIWAN_HPC	April 12, 2013	running
	1471770737	el.B598-3_HPC_taiwan.872464...	Felix Dteam Lee	ANALY_TAIWAN_HPC	April 12, 2013	running
	1471770736	el.B598-3_HPC_taiwan.872464...	Felix Dteam Lee	ANALY_TAIWAN_HPC	April 12, 2013	running
	1471770735	el.B598-3_HPC_taiwan.872464...	Felix Dteam Lee	ANALY_TAIWAN_HPC	April 12, 2013	running
	1471770734	el.B598-3_HPC_taiwan.872464...	Felix Dteam Lee	ANALY_TAIWAN_HPC	April 12, 2013	running
	1471770733	el.B598-3_HPC_taiwan.872464...	Felix Dteam Lee	ANALY_TAIWAN_HPC	April 12, 2013	running
	1471770732	el.B598-3_HPC_taiwan.872464...	Felix Dteam Lee	ANALY_TAIWAN_HPC	April 12, 2013	running
	1471770731	el.B598-3_HPC_taiwan.872464...	Felix Dteam Lee	ANALY_TAIWAN_HPC	April 12, 2013	running
	1471770730	el.B598-3_HPC_taiwan.872464...	Felix Dteam Lee	ANALY_TAIWAN_HPC	April 12, 2013	running
	1471770729	el.B598-3_HPC_taiwan.872464...	Felix Dteam Lee	ANALY_TAIWAN_HPC	April 12, 2013	running
	1471770728	el.B598-3_HPC_taiwan.872464...	Felix Dteam Lee	ANALY_TAIWAN_HPC	April 12, 2013	running

pandaweb.grid.sinica.edu.tw/monitor/1471770741/active/

Panda Web – Manage data



Panda Web – Manage data



Make use of PanDA framework

- What do we do with PanDA framework?
- Support AMS computing since 2012.
  - Whole node reco production jobs, MC jobs, user analysis.
- Also aiming applications other from HEP
  - Weather research and simulation, earth science, drug discovery.. blah blah...
  - In particular for MPI jobs.
    - A customized multiple-node MPI pilot factory, which submits pilot worker in different MPI size. e.g. 32, 64, 128...

Cloud with PanDA pilot

Condor-G/EC2

Check Queue

Submit VM
&
Shutdown VM

1. Download pilot script and proxy by globus-gass
2. Pass job status back to Condor

Pull jobs

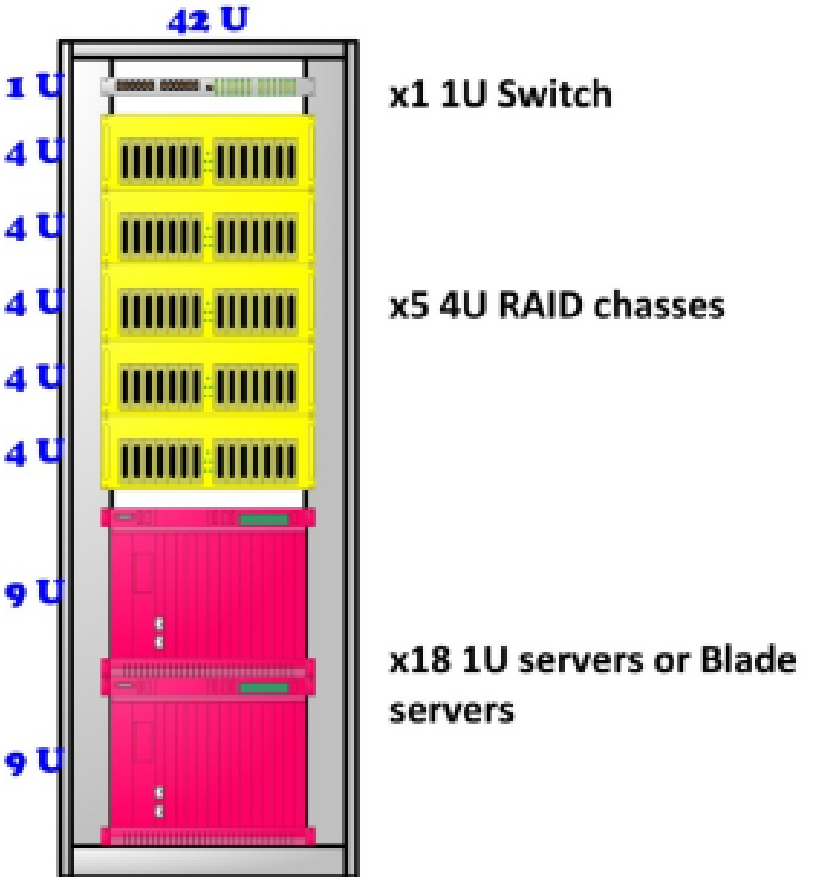
OpenNebula.org
The Open Source Toolkit for Cloud Computing

openstack™

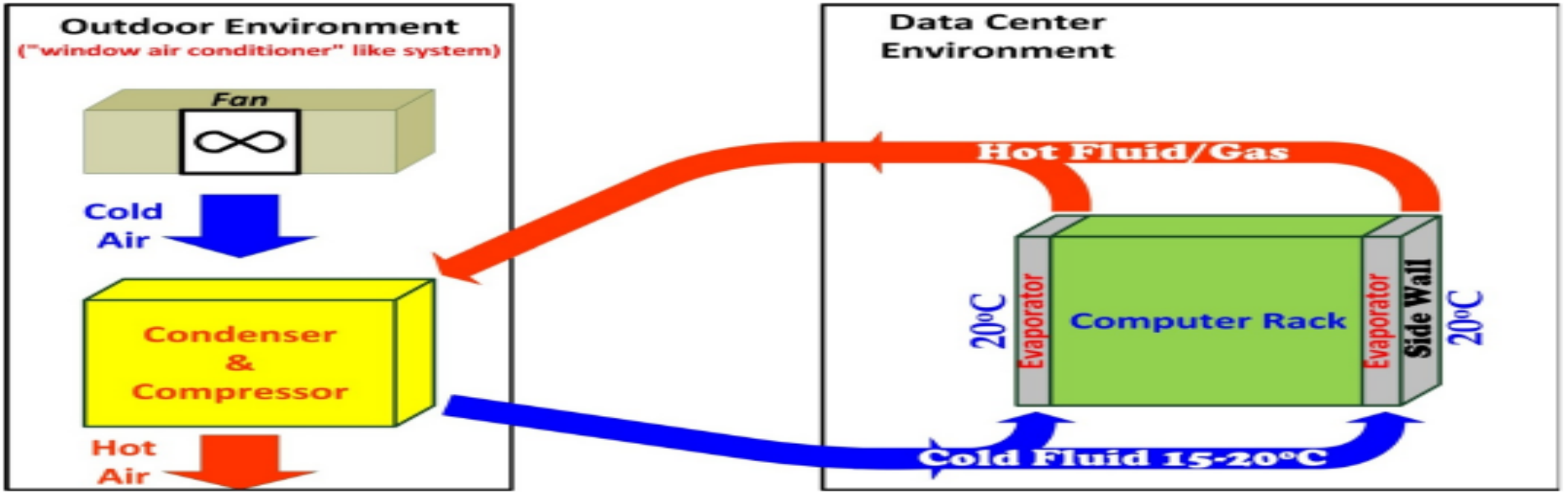
Create VM

Single rack project - concept

- 360 TB+ storage (4U, 24bay)
- 432 CPU cores(twins, 6cores)
- One switch
- 10kw to 15kw consumption
- A basic unit in DC
- But not all of labs/institutes have DC.
  - But they do need computing power
  - How about we put it into their lab or just a room with general air-condition available.
  - Can we make this happened?

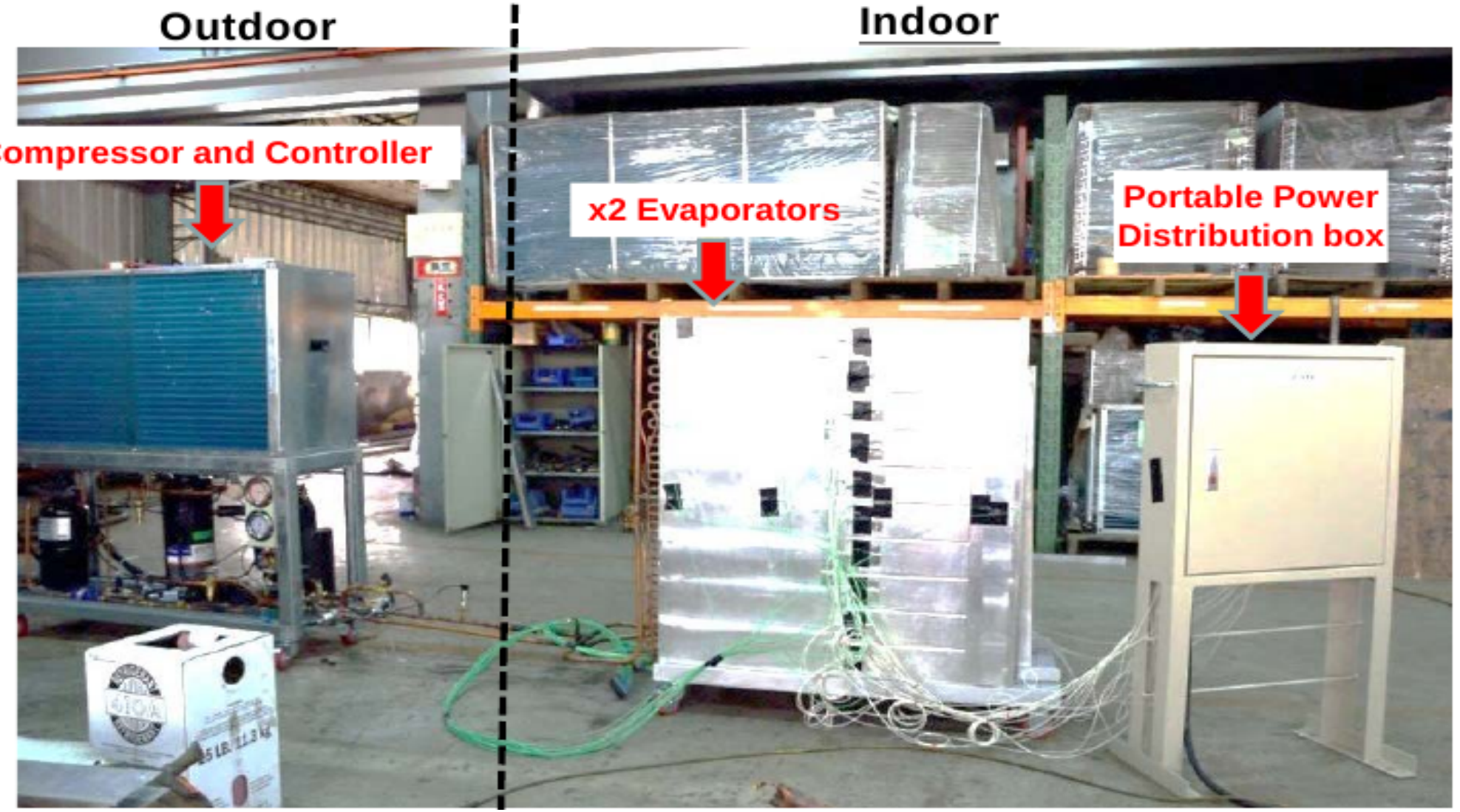


Single rack project – concept, the fanless rack

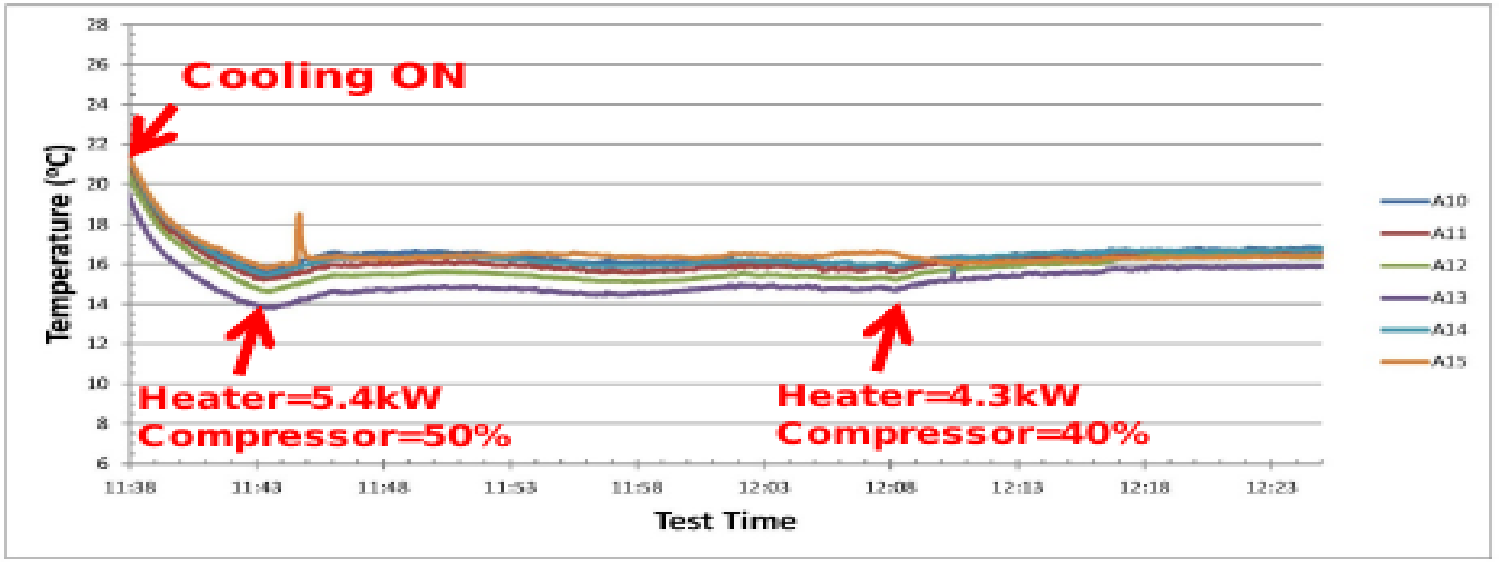


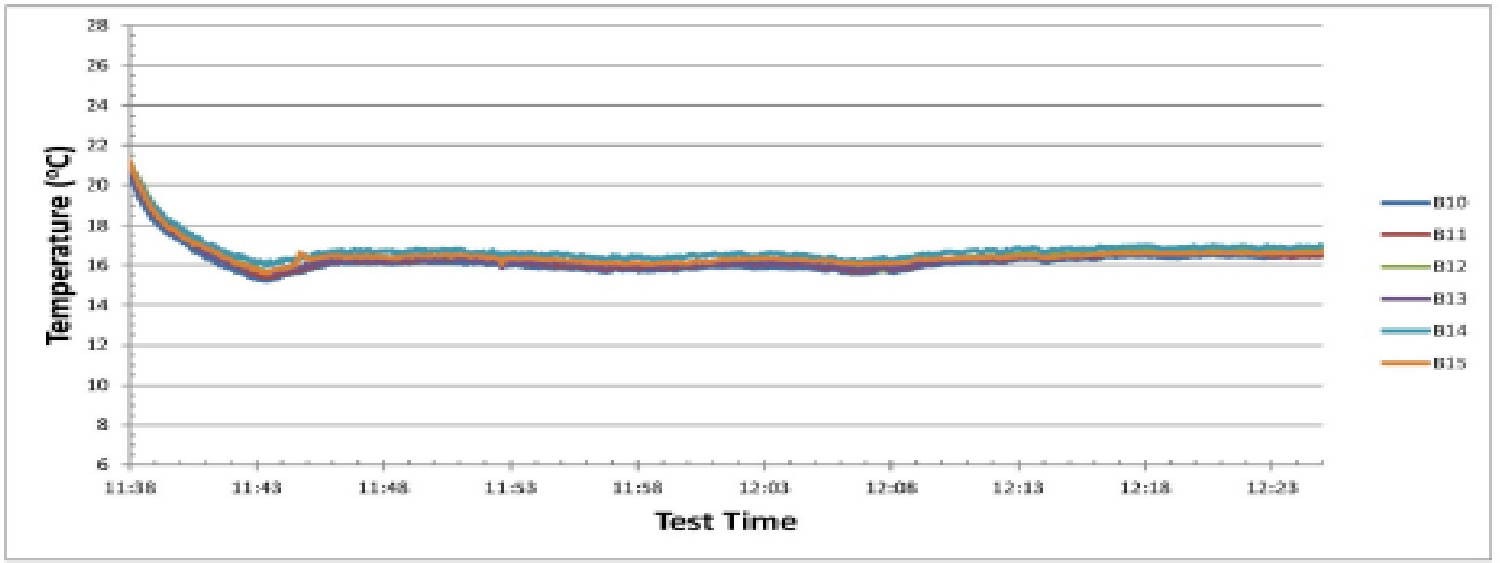
- Conductional cooling from inner parts to rack sidewalls.
- Use rack sidewalls as evaporators.
- Flow refrigerant through it.

Rack cooling design prototype

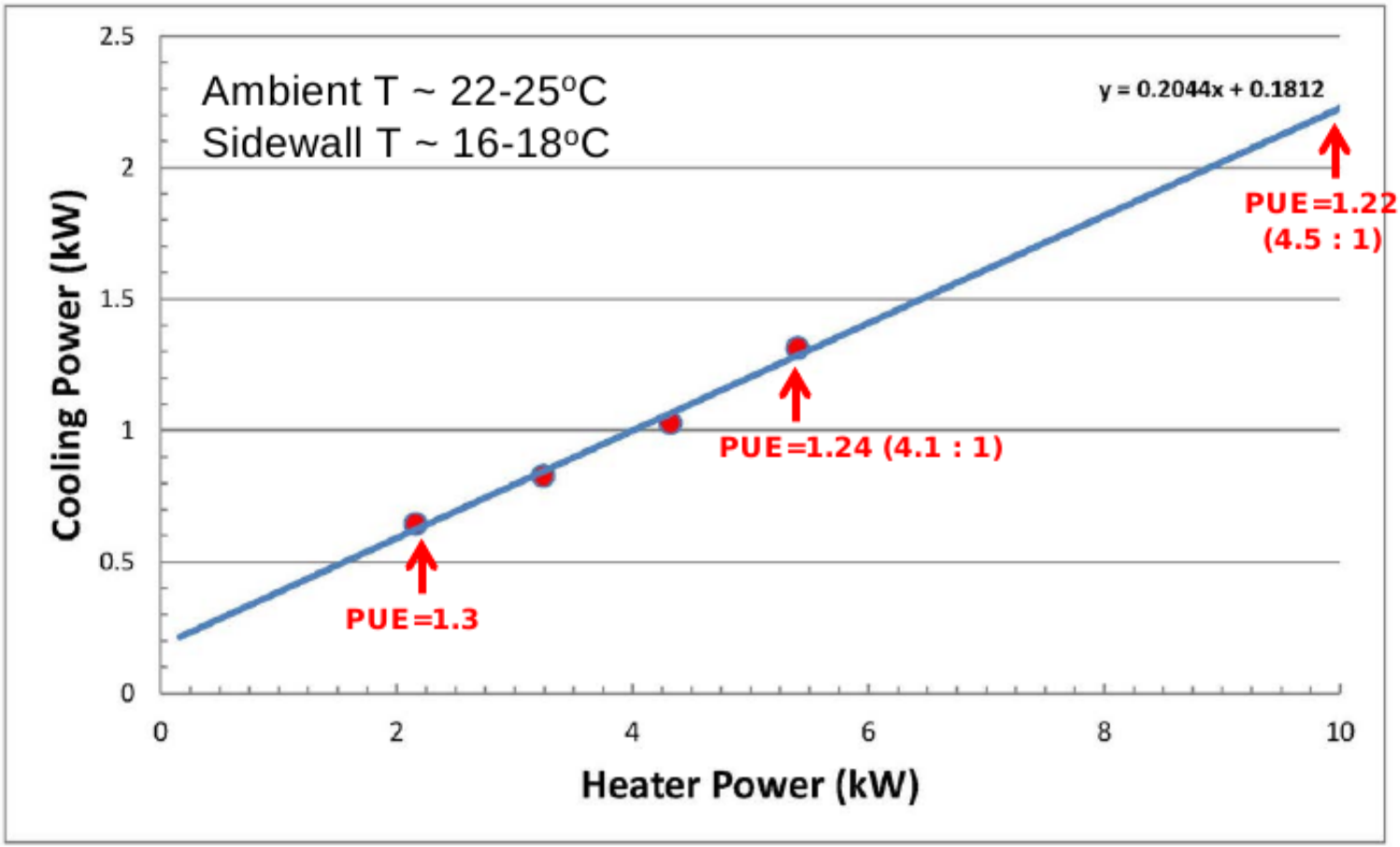


Fanless rack cooling PoC

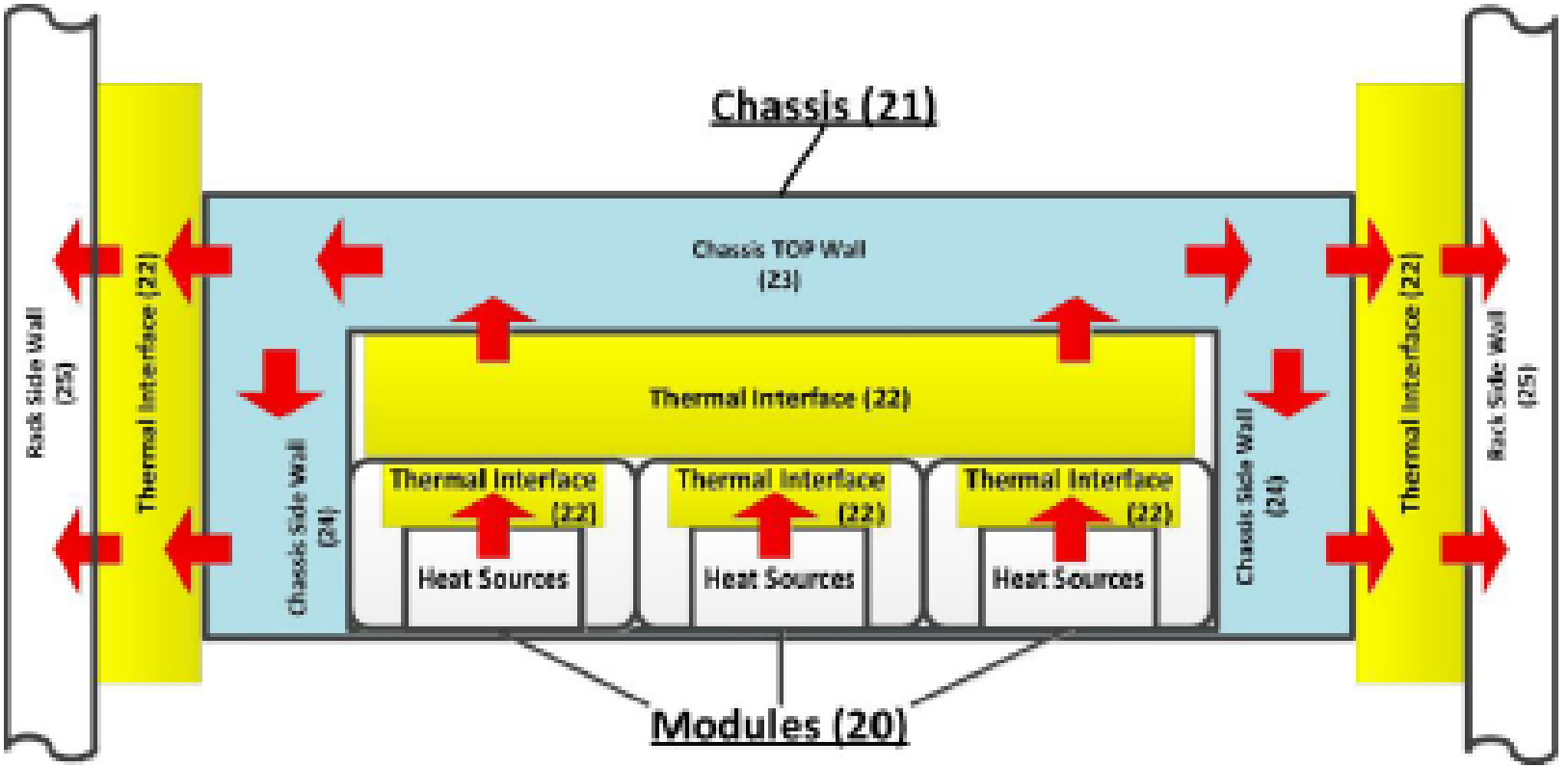




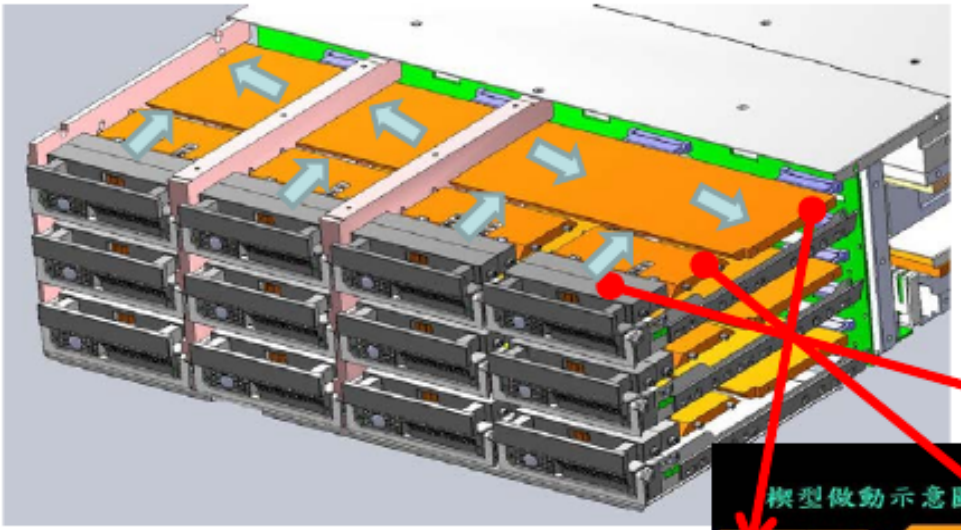
PUE of fanless rack



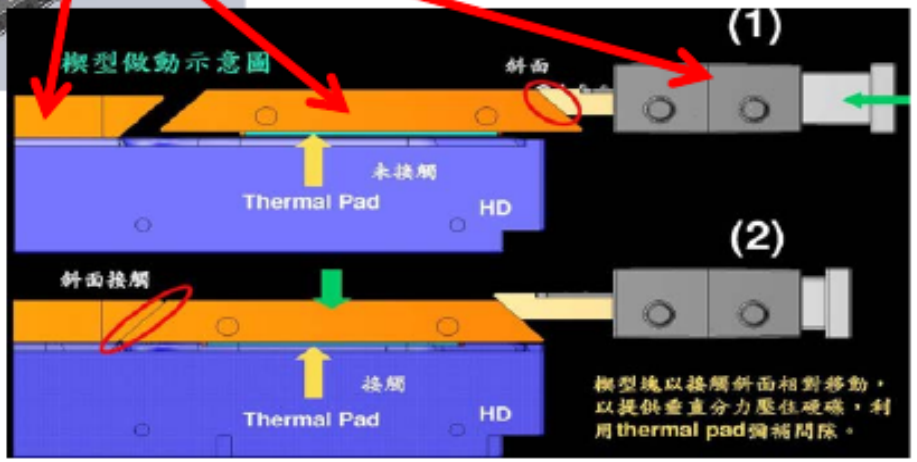
Fanless enclosure design concept



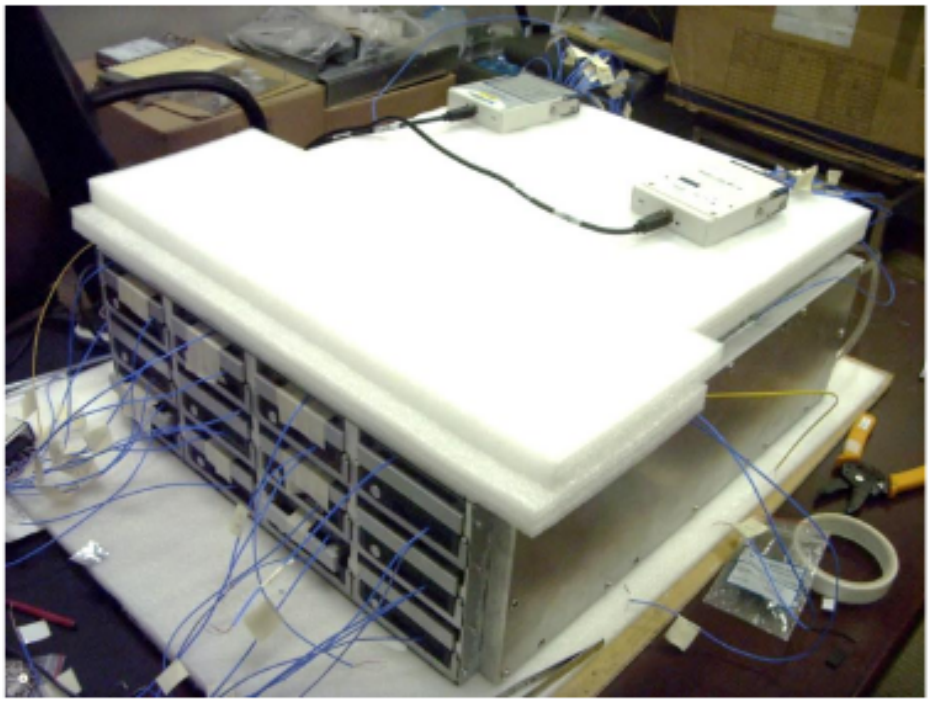
Fanless RAID sys design

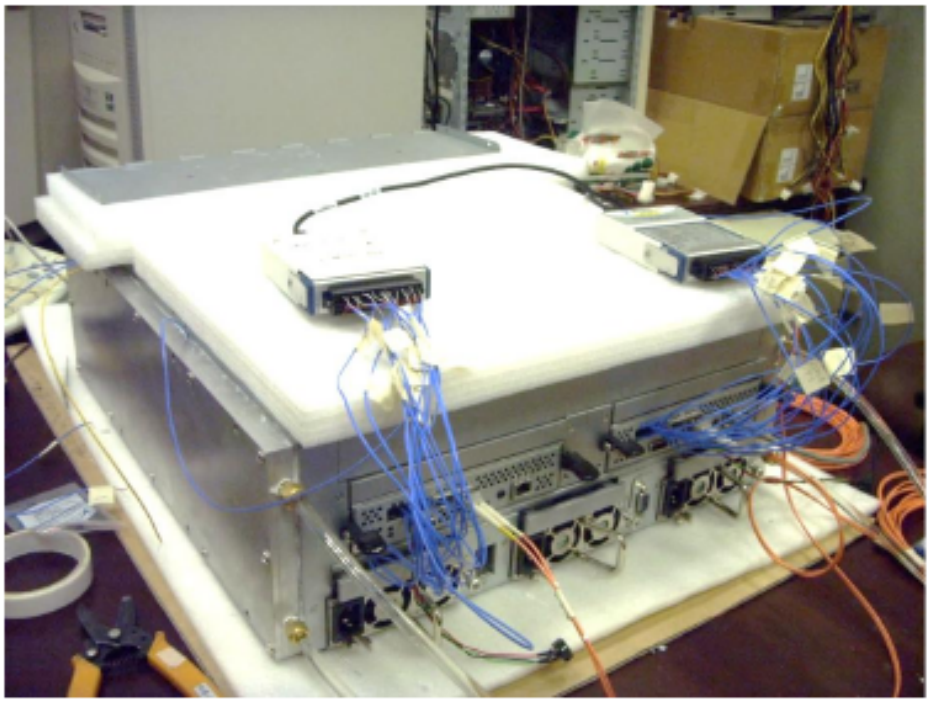


Thermal Flow



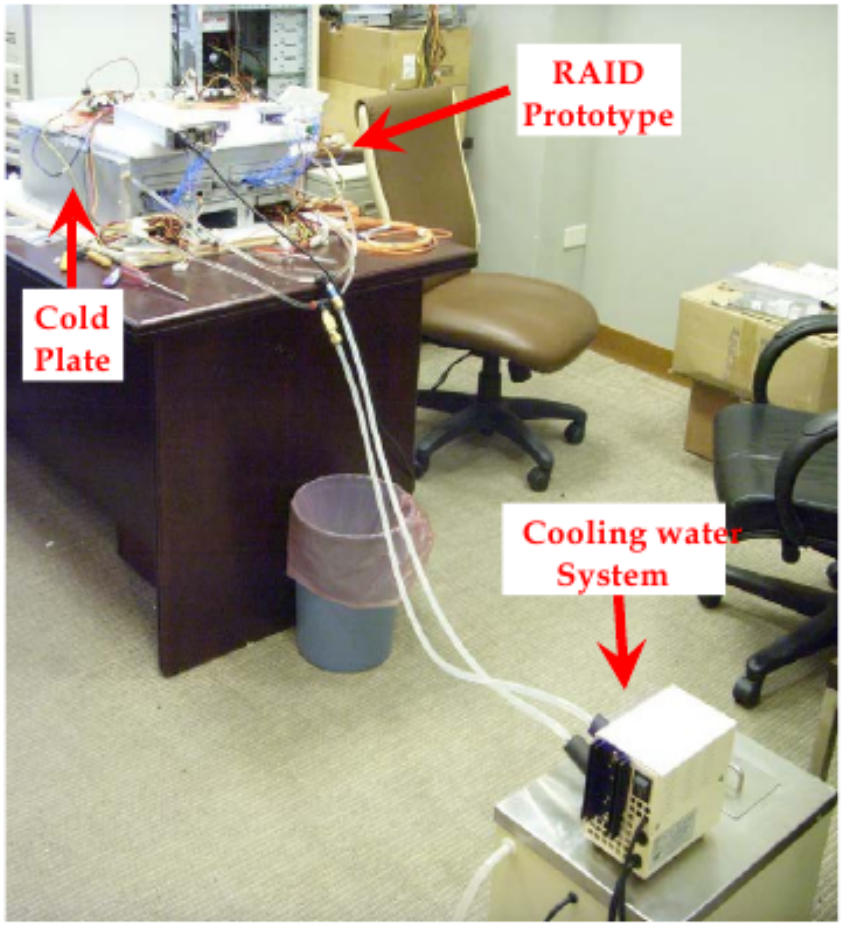
Fanless RAID sys prototype



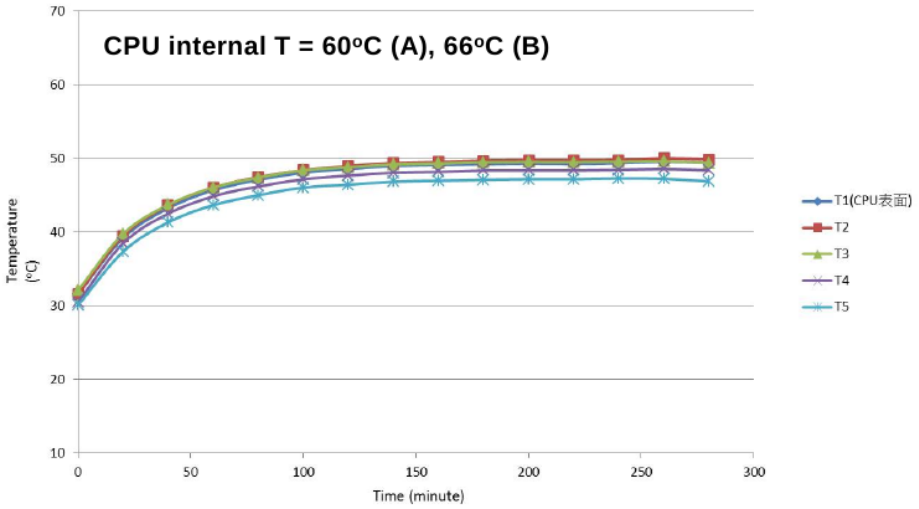


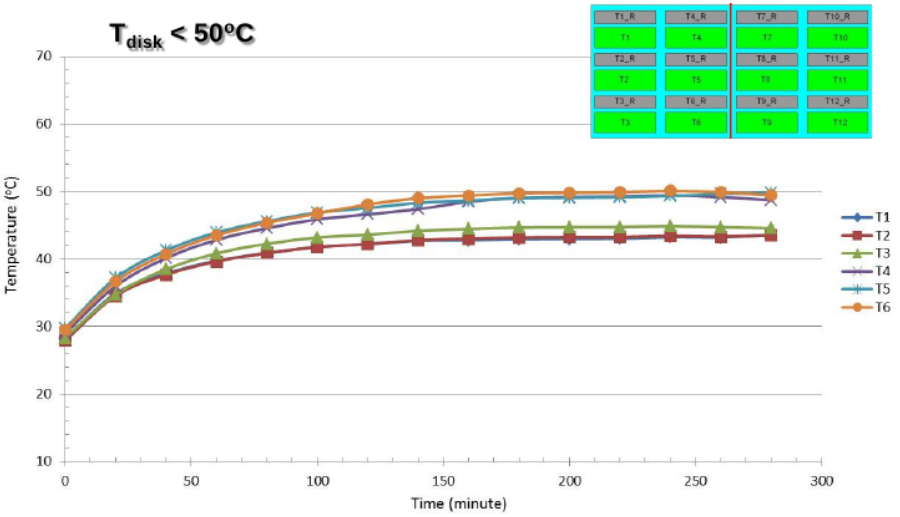
Fanless RAID sys PoC

- Simply use small water cooling system.
- Keep chassis sidewall to 25~28 ℃.



Fanless RAID sys design prototype





Future works

- Single rack project.
  - Design fanless server chassis
  - Put everything together.
  - Battery module also for storage.
- Network
  - Upgrade router interface between Chicago and Amsterdam.
  - Upgrade 2.5Gb backup link to 10Gb.
- PanDA/Grid
  - Finding more user/applications
  - Improving our UI by getting more user feedback.
- Cloud.
  - Improving scheduler, workflow, VM lifetime, VM traceability.
  - Looking for global filesystem for VM instance, maybe Ceph RBD.
  - Publish accounting info to APEL.

Thanks a lot!