国家高能物理科学数据中心
National HEP Science Data Center

高能所計算中心
IHEP Computing Center

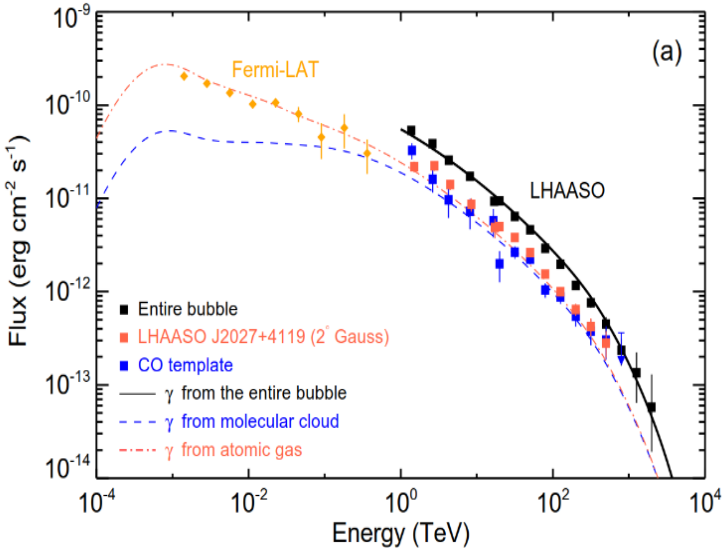
Integration of A Supercomputing Center in LHAASO Distributed Computing System

Qingbao HU, Xiaowei JIANG
IHEP Computing Center
2024-10-24

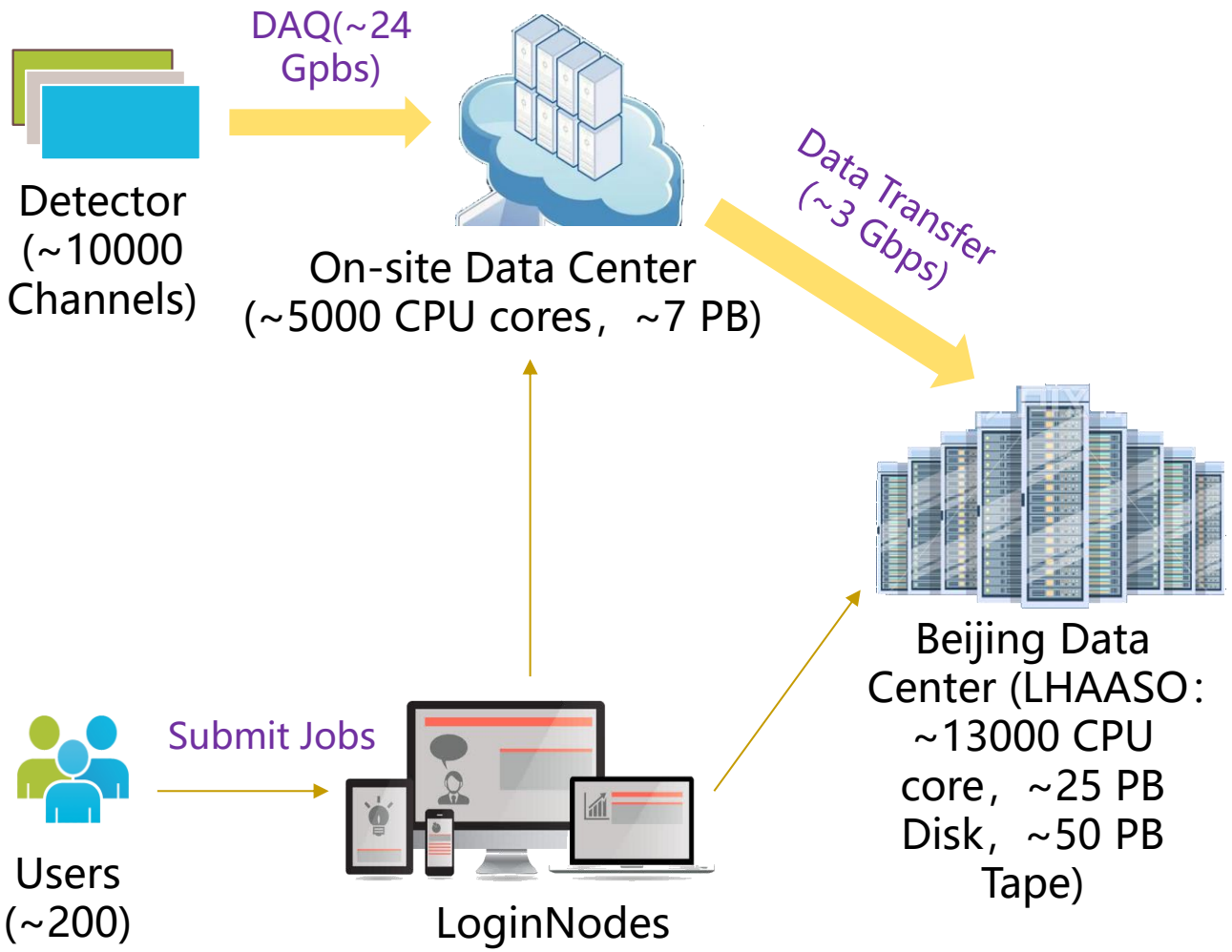
Outline

Large High Altitude Air Shower Observatory (LHAASO)

- World largest air shower array(with e, m, water Č detectors and Č telescope) for the high energy γ-astronomy and cosmic-ray physics
- Construction completed in 2023 and interesting results came out:
  - Highest γ-rays from the Milky Way： 2.5 PeV
  - 43 identified γ-rays sources up to ~1 PeV ➔ PeVatrons in the Milky Way
  - Energy spectrum of high energy γ-rays from the Crab Nebula as the standard candle
- International Collaboration : Countries/regions:5, members:~300



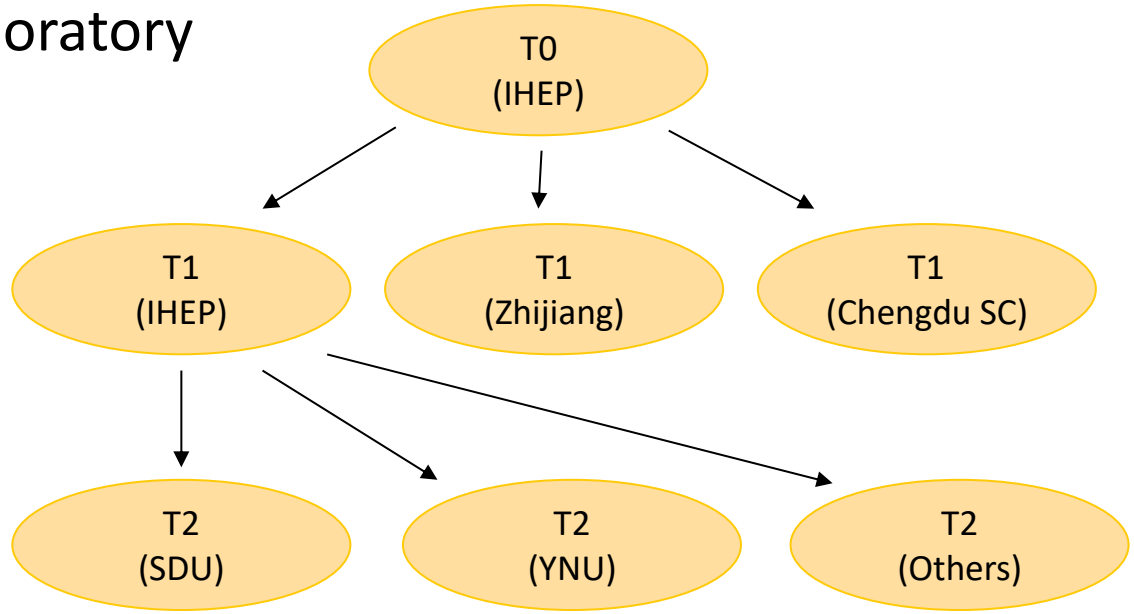
LHAASO Data Processing



- Resources
  - 30 PB Disk
  - 50 PB Tape
  - 18000 CPU cores
- Services
  - Computing, Disk storage, Tape storage, Network, Data Transfer
- ~11 PB increase up every year

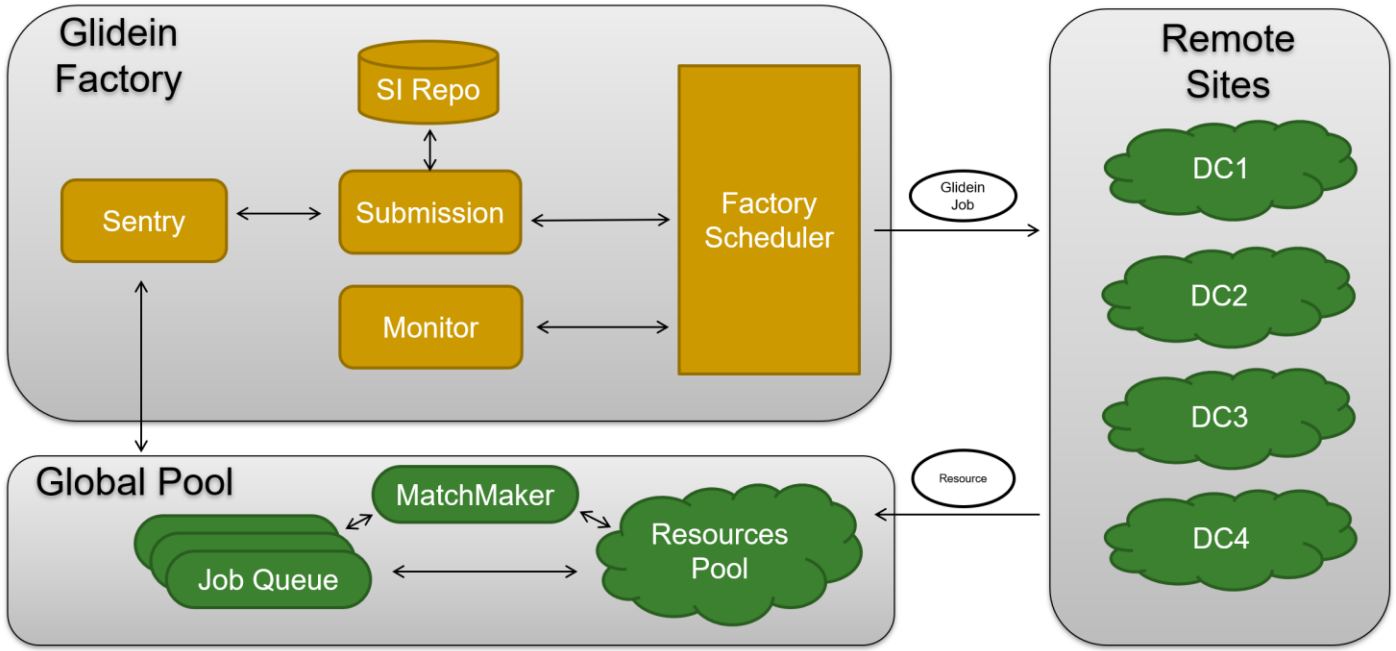
Distributed Computing Model for LHAASO

- Tier model (similar with WLCG's grid computing model)
  - T0: IHEP
  - T1: IHEP, Chengdu SC Center, Zhijiang Laboratory
  - T2: collaboration sites
- Systems and Services
  - Distributed computing system
  - Distributed storage system
  - Service and job environment
  - Authentication and Authorization
  - User interfaces (HepJob)



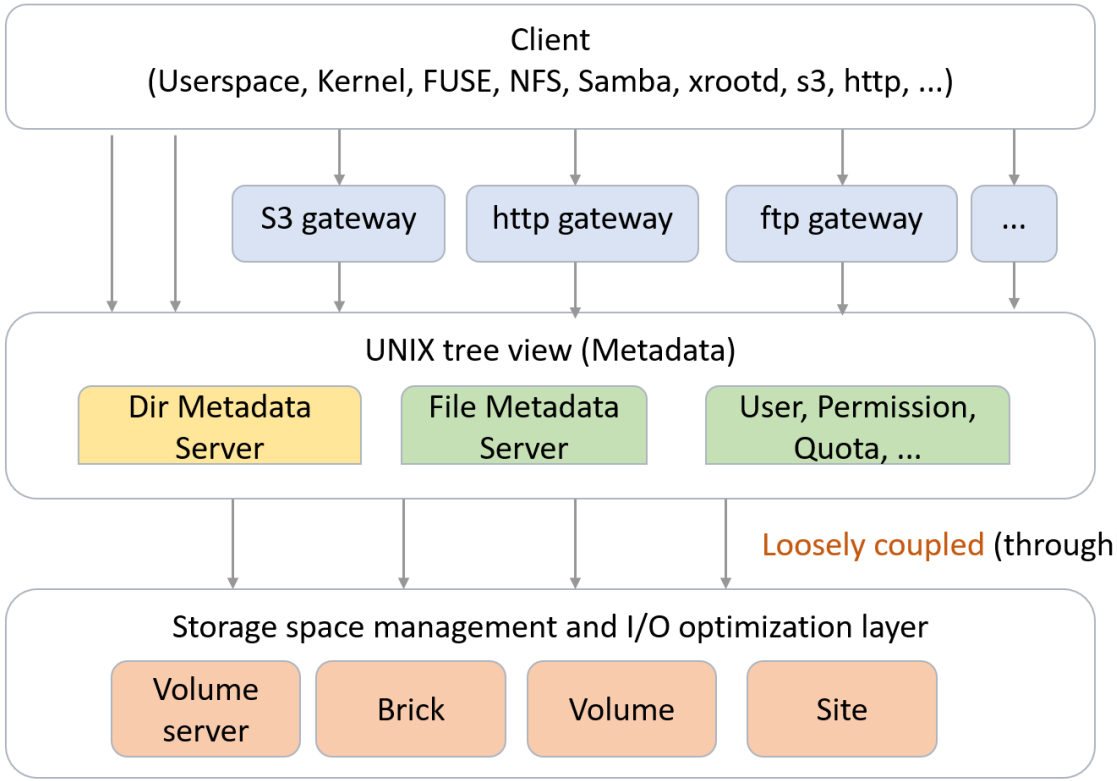
LHAASO Distributed Computing (1)

- Distributed Computing System
  - Find and publish the remote computing resources to global side
  - Dispatch the user job to remote worker node
- Components
  - Global Pool
  - Glidein Factory
  - Glidein Client Tool
- HTCondor-C is also used
  - Schedd server is different with glidein pool



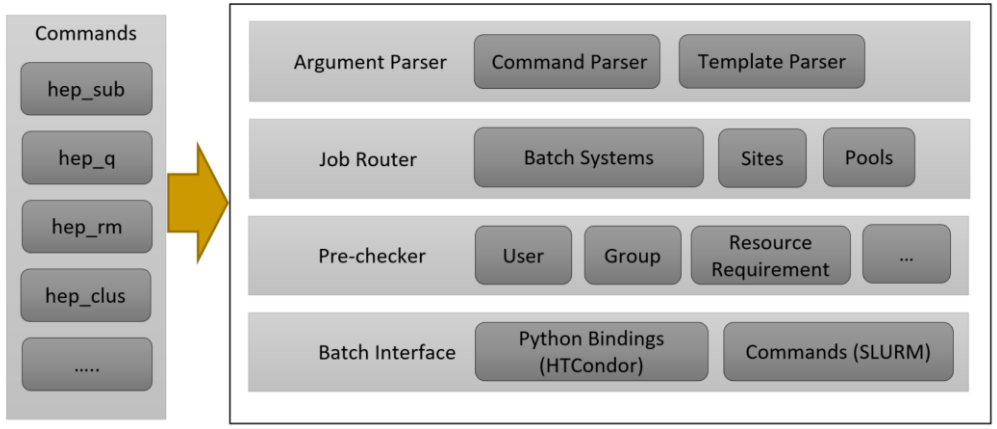
LHAASO Distributed Computing (2)

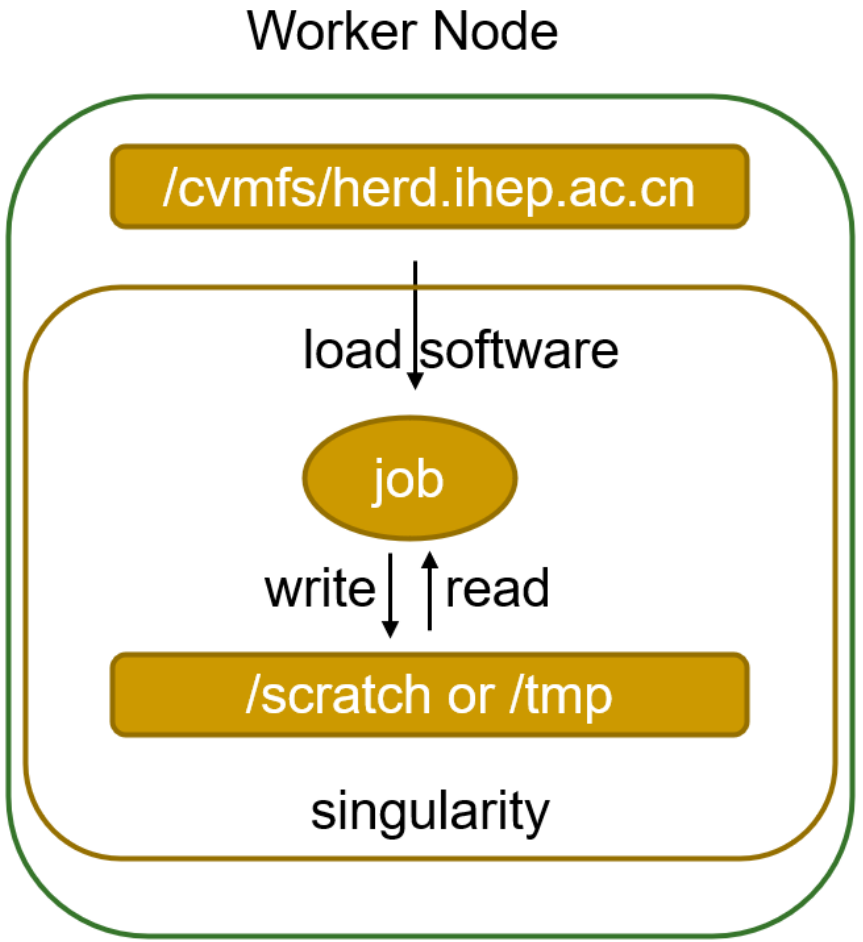
- Distributed Storage System
- XRootD Proxy
  - A central way to read/write data from EOS
- Particlefs or Ocloud
  - A unified file system above multiple sites
- Rucio
  - A data management system to manage data
- First step is still using xrootd proxy
  - Particlefs/Ocloud or Rucio would be the solution in future



LHAASO Distributed Computing (3)

- Authentication and Authorization – tokens
  - Kerberos token have been used in production
    - Local cluster (>10 years) and Dongguan data center
- Running Environment - Singularity
  - Container
- User Interface - HepJob



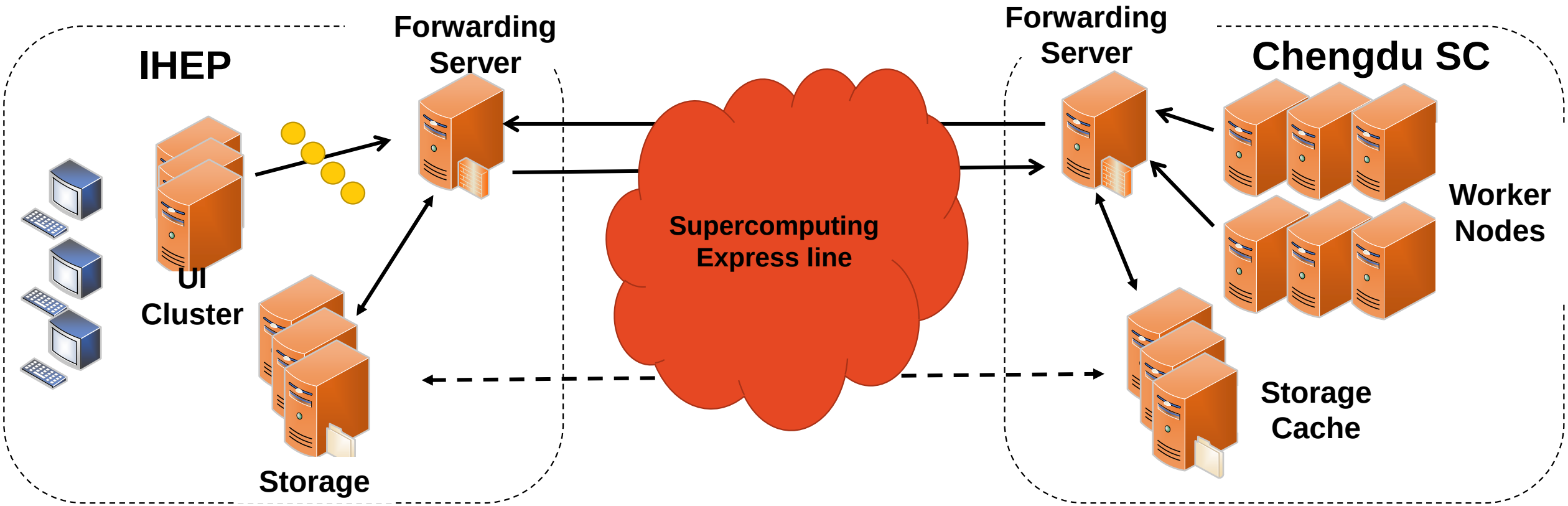


Trying to Integrate Chengdu Supercomputing Center

- Resources provided from Chengdu Supercomputing Center
  - >10,000 CPU cores and ~8PB Storage space
- The network link between Beijing and Chengdu was the big problem
  - Chengdu SC does not allow the inner servers to connect with outside
- A 10-Gbps network link was established
  - Cooperated with China Telecom Research Institute
  - Deployed the hardware VPNs on both of centers

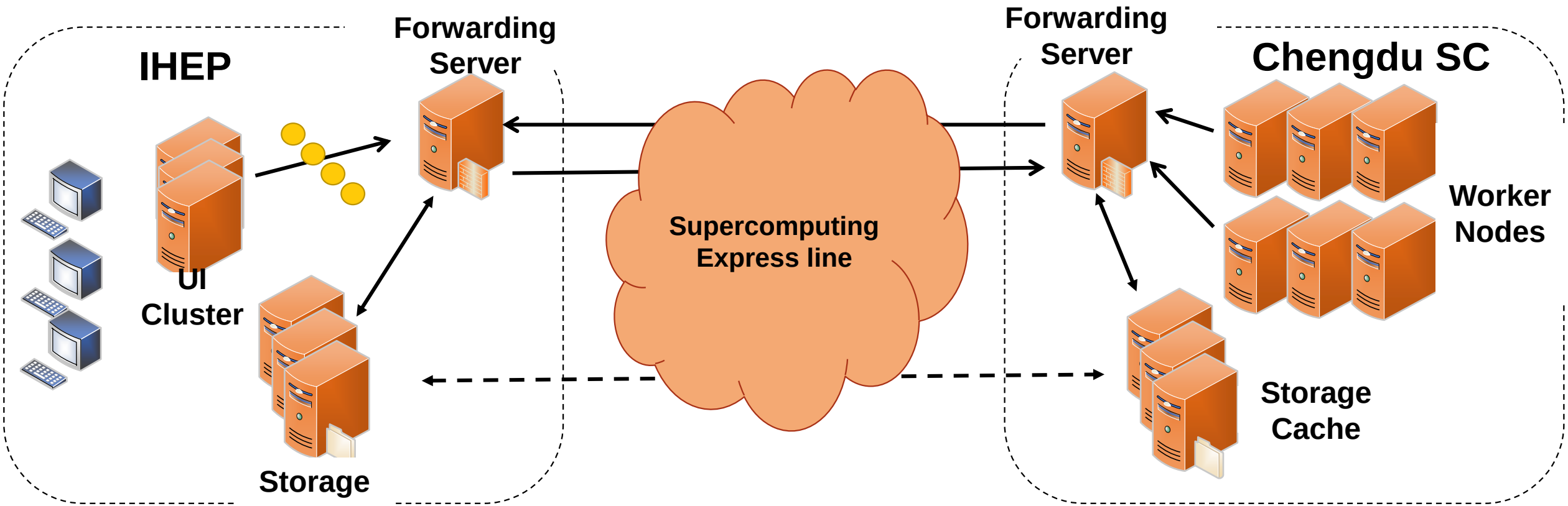
Trying to Integrate Chengdu Supercomputing Center

- Supercomputing Express line is built by China Telecom Research Institute
  - Aim to connect the supercomputing centers in China
  - This is the first application



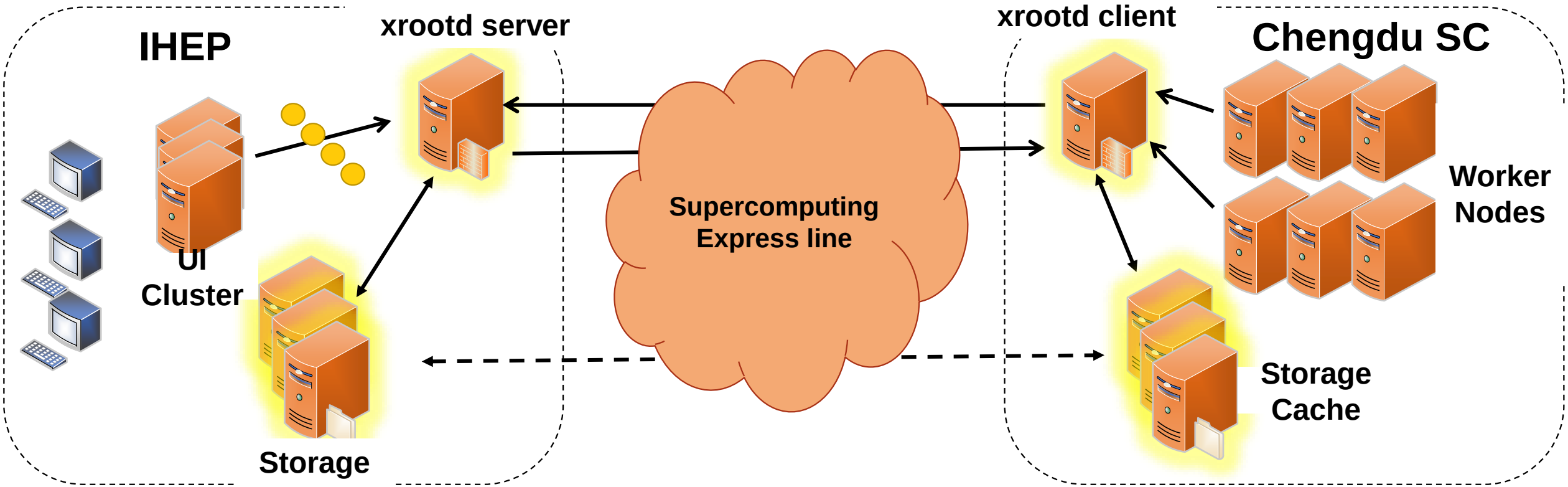
Trying to Integrate Chengdu Supercomputing Center

- All involved services are deployed on the forwarding servers
  - Forwarding servers are visible with each other via supercomputing express line
  - HTCondor for computing / Xrootd for data access



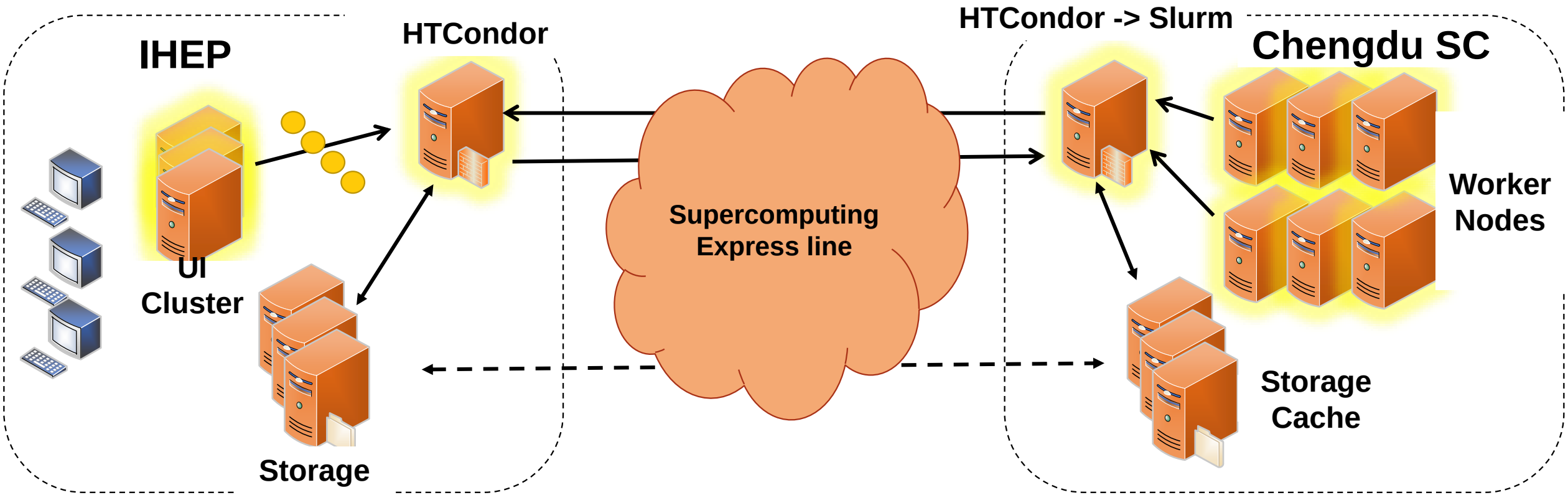
Data Access with XRootd

- The disk data storage is mounted on the xrootd server
  - Currently the data is manually transferred from IHEP to Chengdu SC via xrootd
- The file directories keep same structure on both sides



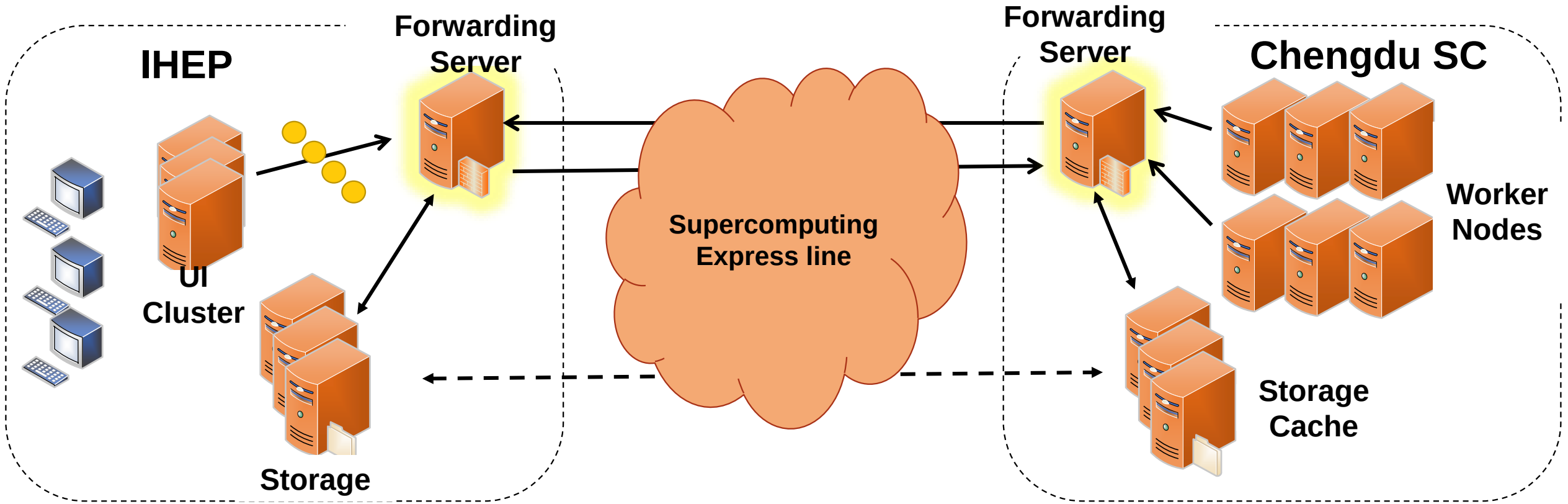
Computing Solution – HTCondor

- Job schedule: condor-c & condor-g
  - IHEP HTCondor -> Chengdu HTCondor -> Slurm



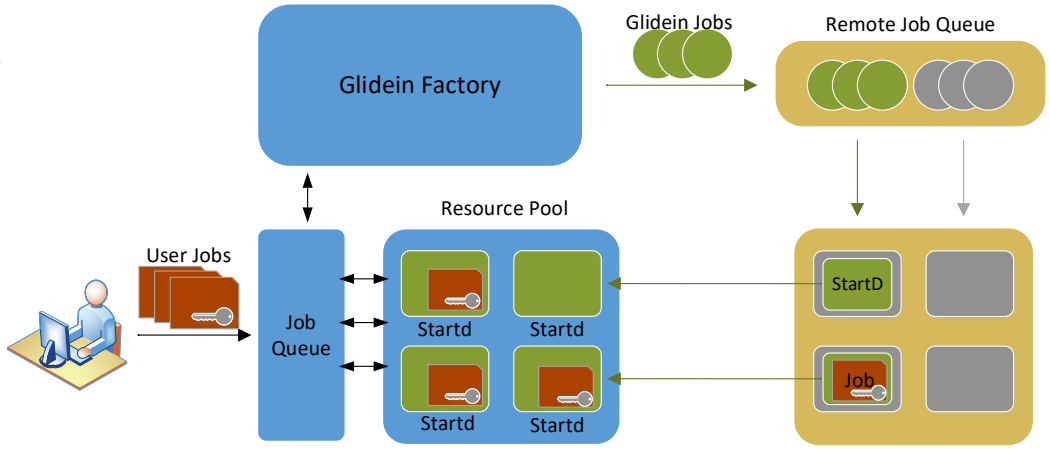
Software Synchronization

- Software is synchronized from IHEP side to Chengdu SC
  - /cvmfs is mounted on forwarding server, where the lhaaso software stores
  - The software files are transferred to the local storage on Chengdu SC forwarding server
  - The directories on Chengdu SC keep same with /cvmfs



Cert&Auth with Kerberos Tokens

- Kerberos tokens for jobs and other services
  - Kerberos token is using in LHAASO local cluster
  - Access data from xrootd server with Kerberos
  - Possible for other services



- Current solution is to transfer token credential as a normal input file

Submit Side

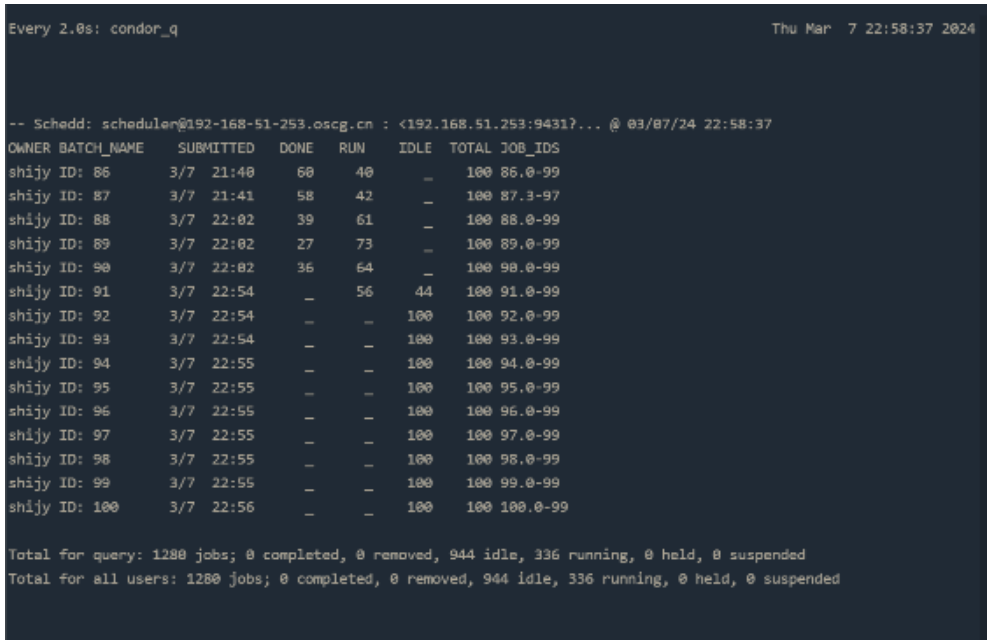
```
transfer_input_files = /tmp/krb5cc_10634
+HepJob_KRB5CCNAME = "krb5cc_10634"
```

Execute Side

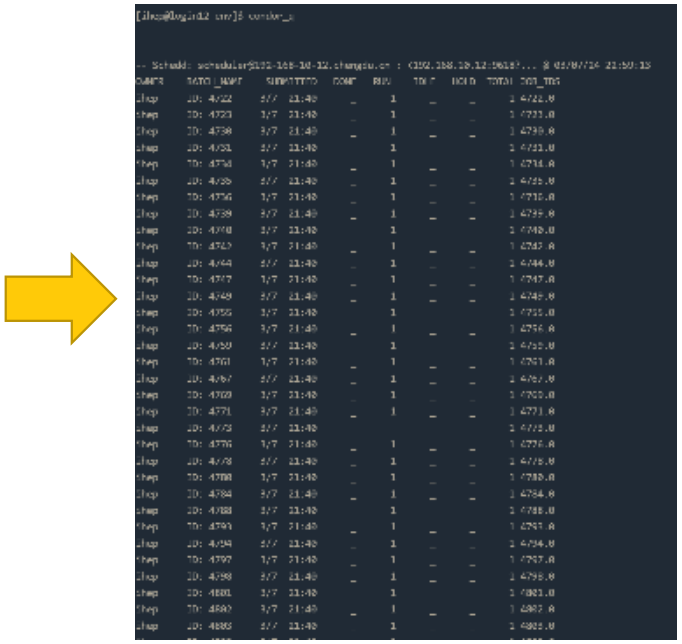
```
$ ls /var/lib/condor/execute/dir_6412/
condor_exec.exe  _condor_stderr  _condor_stdout  krb5cc_10634  tmp  var
```

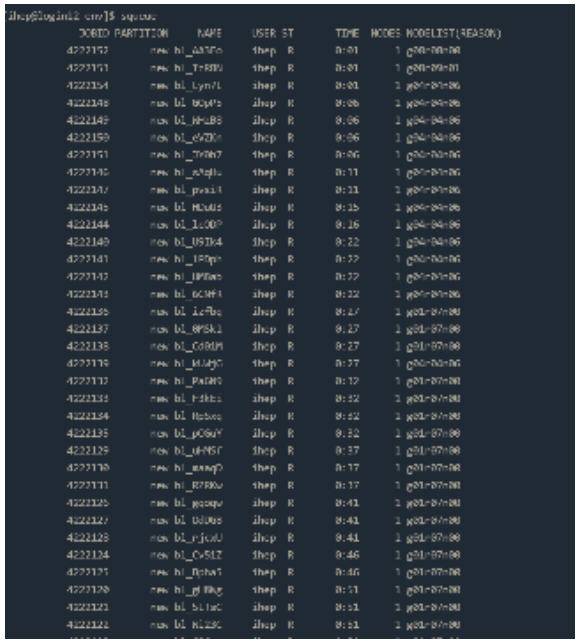
Job Test

- Job schedule: condor-c & condor-g
  - IHEP HTCondor -> Chengdu HTCondor -> Slurm



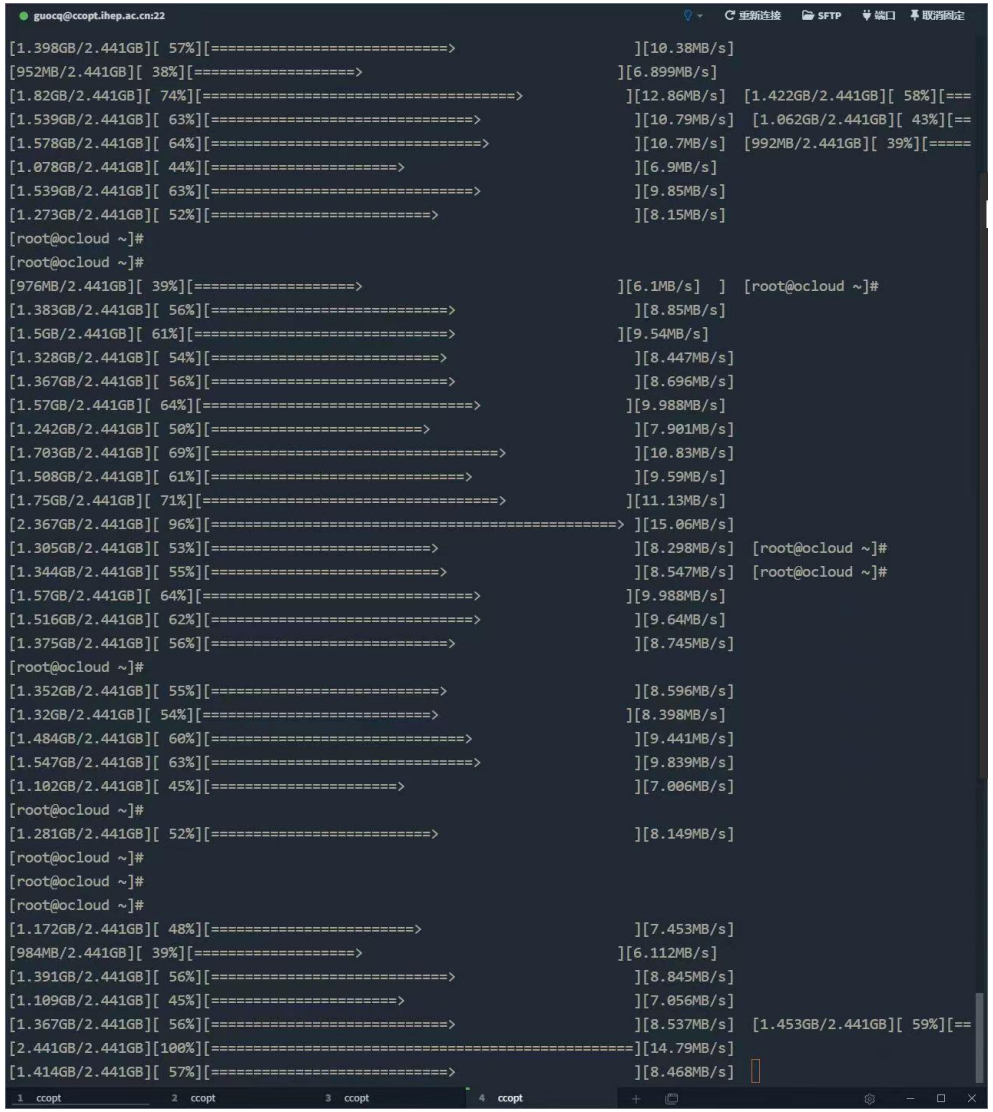
```
Every 2.0s: condor_q                                                    Thu Mar  7 22:58:37 2024


-- Schedd: scheduler@192-168-51-253.oscg.cn : <192.168.51.253:9431?... @ 03/07/24 22:58:37
OWNER BATCH_NAME     SUBMITTED   DONE   RUN    IDLE  TOTAL JOB_IDS
shijy ID: 86       3/7  21:40     60     40       _    100 86.0-99
shijy ID: 87       3/7  21:41     58     42       _    100 87.3-97
shijy ID: 88       3/7  22:02     39     61       _    100 88.0-99
shijy ID: 89       3/7  22:02     27     73       _    100 89.0-99
shijy ID: 90       3/7  22:02     36     64       _    100 90.0-99
shijy ID: 91       3/7  22:54      _     56      44    100 91.0-99
shijy ID: 92       3/7  22:54      _      _     100    100 92.0-99
shijy ID: 93       3/7  22:54      _      _     100    100 93.0-99
shijy ID: 94       3/7  22:55      _      _     100    100 94.0-99
shijy ID: 95       3/7  22:55      _      _     100    100 95.0-99
shijy ID: 96       3/7  22:55      _      _     100    100 96.0-99
shijy ID: 97       3/7  22:55      _      _     100    100 97.0-99
shijy ID: 98       3/7  22:55      _      _     100    100 98.0-99
shijy ID: 99       3/7  22:55      _      _     100    100 99.0-99
shijy ID: 100      3/7  22:56      _      _     100    100 100.0-99

Total for query: 1280 jobs; 0 completed, 0 removed, 944 idle, 336 running, 0 held, 0 suspended
Total for all users: 1280 jobs; 0 completed, 0 removed, 944 idle, 336 running, 0 held, 0 suspended
```





Data Test

- Data Transfer: xrootd
  - Transfer speed is above 1GB/s



Summary

- LHAASO is building the distributed computing system to handle the big scale of data processing
  - Supercomputing Center can contribute computing and storage
- LHAASO is trying to integrate the Chengdu Supercomputing Center to the distributed computing system
  - Network solution: Supercomputing Express line
  - Computing solution: HTCondor-C & HTCondor-G
  - Data solution: Xrootd
- Problem and Next Plan
  - To develop a real-time data transfer service and deploy it on Supercomputing side
  - Investigate a proxy way to use glidein to publish the worker nodes from SC to Central Glidein Pool
  - The network solution is too heavy now and needs a better one

Thanks!

Q&A