HT-Condor tutorial for HEP users (w/ TIDC facilities)

Felix Lee
NSTCCore & TIDC User Training Workshop

Computing resource
& accessible end point

- tidc-ui[01~04].grid.sinica.edu.tw (UI)
  - tidc-ui[01,03] : CentOS7 worker nodes (to be decommissioned)
    - tidc-ui02 is under maintenance...
  - tidc-ui04: Alma9 worker nodes
    - Will add tidc-ui05 in the future.
  - ssh UI, Condor scheduler, job submitter.
  - Accessible by ssh client with dicos account.
- tidc-arc6-1.grid.sinica.edu.tw (ARC CE)
  - Under to migration to Condor-CE with Alma9 (ongoing)
  - Grid CE.
  - Accessible by CMS Crab3 with grid certificate
  - There will be Condor-CE in the future: tidc-condor-cm.grid.sinica.edu.tw
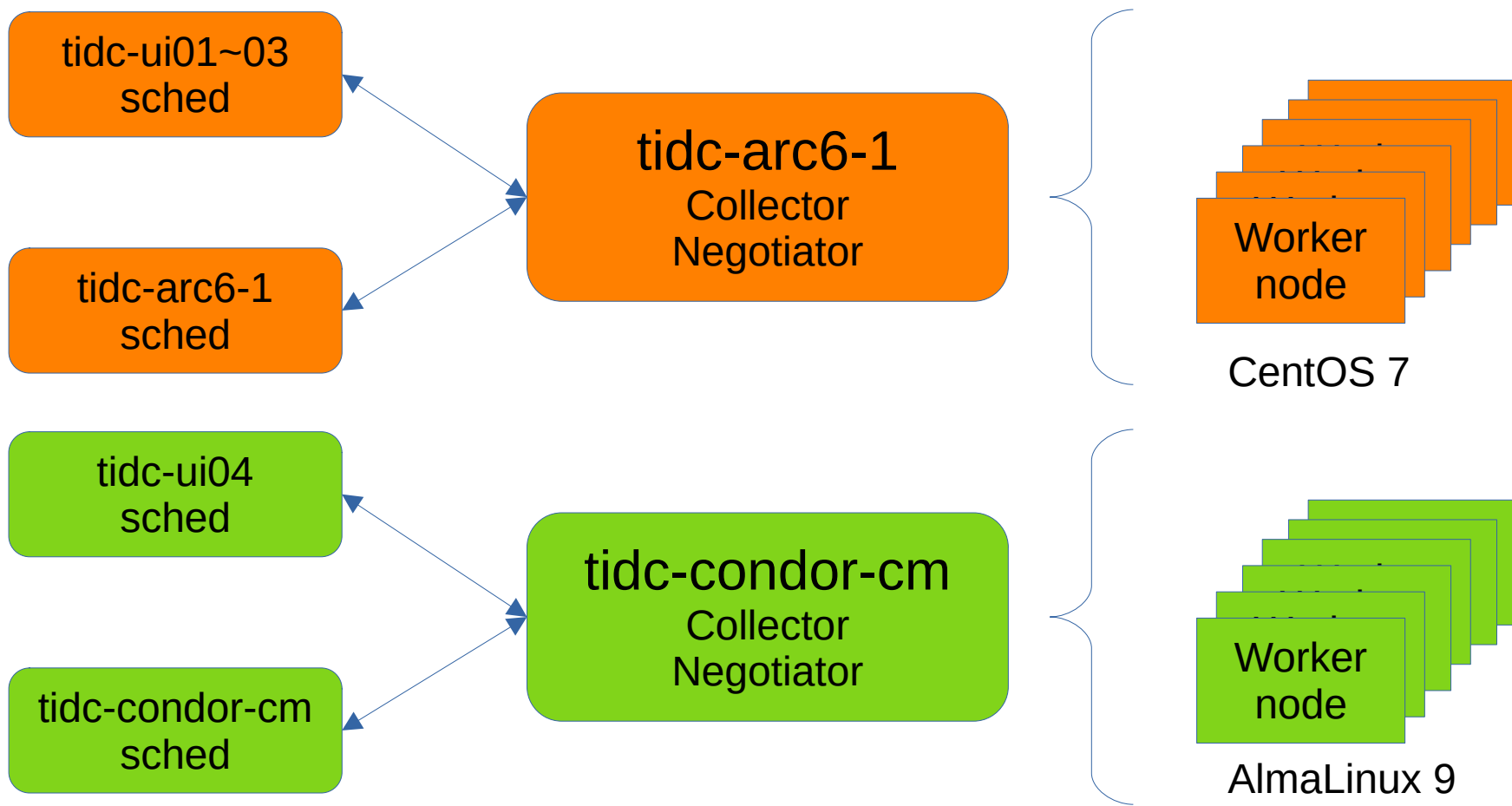
Storage Access

- EOS:
  - By xrootd:
    - root://tidc-smstor1.grid.sinica.edu.tw/eos/
  - By filesystem access. (fuse)
    - /eos
- Shared filesystem
  - /ceph/work/<group name>

Resources

- Condor cluster
  - 768 cores(AMD EPYC 7713) : CentOS7
  - 3072 cores (AMD EPYC 9654, Genoa): Alma9
    - Another 1920 cores will be coming (AMD EPYC 9645, Turin)
- EOS storage
  - 649.99 TB
- Ceph filesystem
  - 3TB per group.
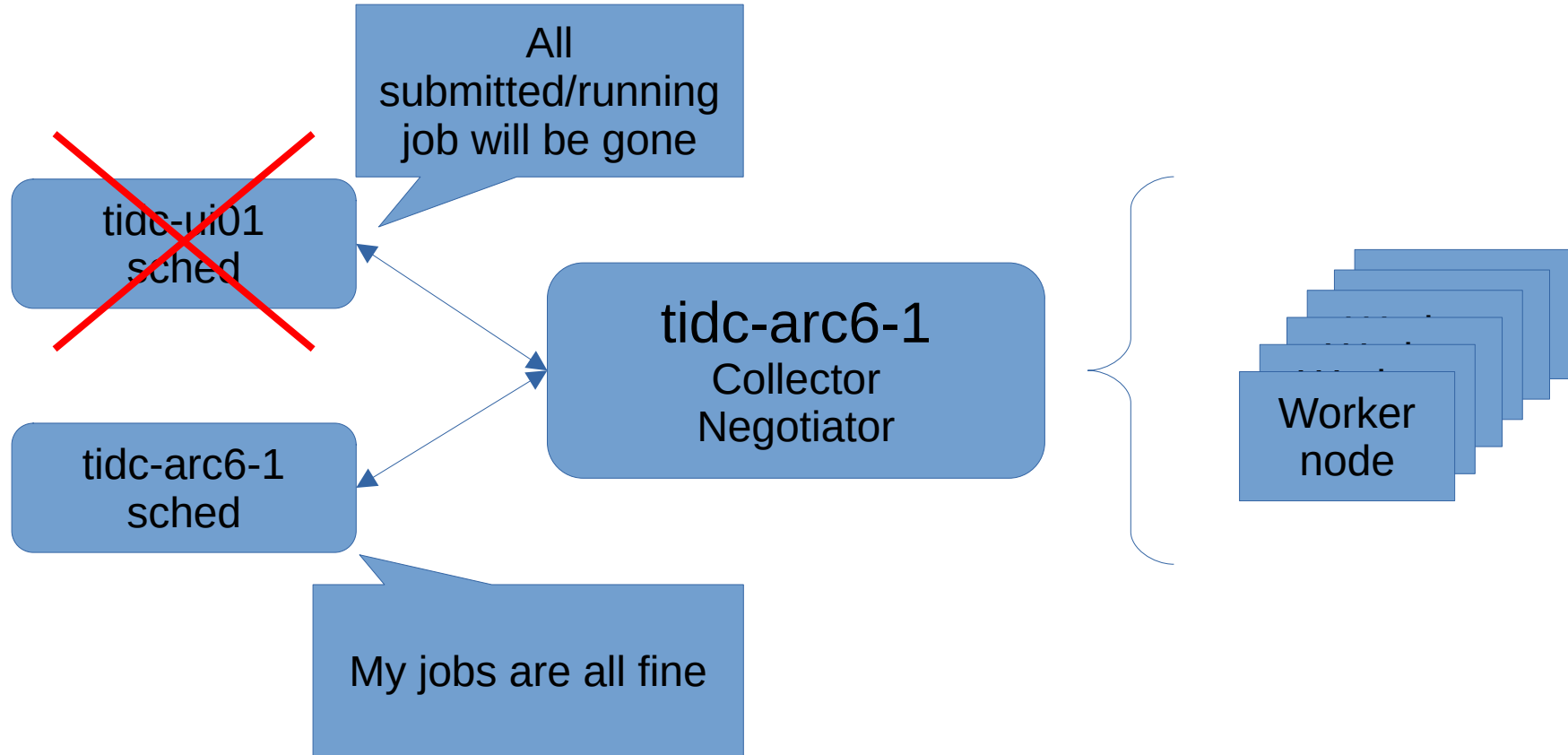   - Can be extended.
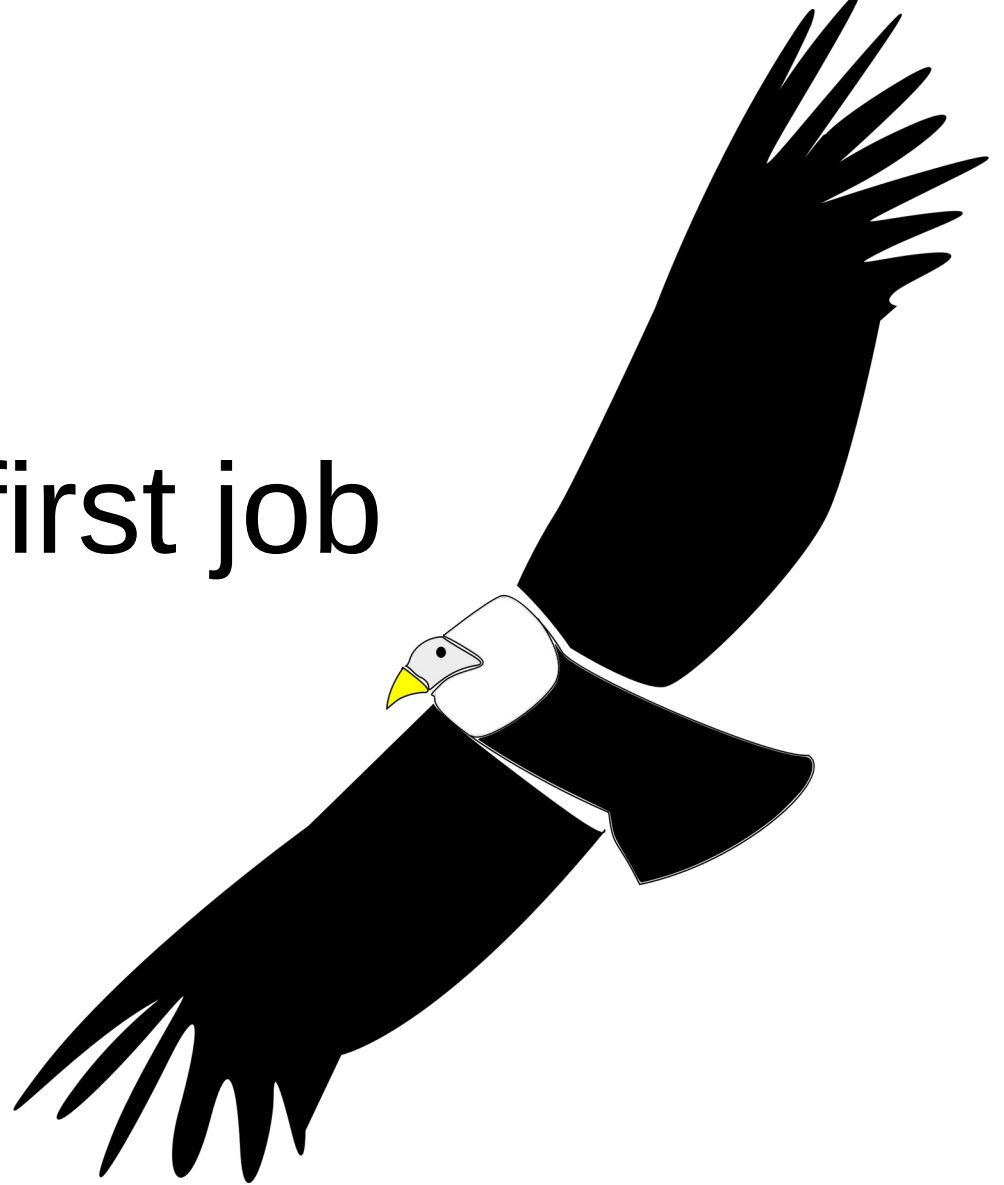
TIDC Condor cluster overview (cur)

Condor scheduler

- The scheduler(submitter) is distributed
  - Each schedd manages their own jobs.
    - e.g. if tidc-ui01 crashes, all of your submitted and running jobs from tidc-ui01 will:
      - Be in held state, if the machine can be recovered.
        - You can rerun it, but the job will restart from fresh unless you do check-point by yourself.
      - All gone, if the machine can't be recovered.
  - So, please use Condor UI wisely, **please be gentle** with UI. :)

Condor scheduler

Your first job

Tutorial files

- /ceph/sharedfs/software/tutorial/condor/
  - Jobmission files:
    - condor.jdl
    - condor_queue_index2.jdl
    - condor_queue_random.jdl
  - Job script
    - test.sh
- Copy them to your home, if you wanna play with it.
  - `cp /ceph/sharedfs/software/tutorial/condor/*  ~/`

Preparing your condor job

- To submit a condor job, we need:
    - 1. Condor submission file
        - Handling files:
            - job executable, job standard out & error, input, output, condor log.
        - Defining job requirement:
            - CPU, memory, disk space.
    - 2. Job executable file. (binary code or script)
    - 3. Input file (optional)
    - 4. Output file (optional)

Job submission file (1/4)

```
executable   = /ceph/work/ASGC/felixlee/test.sh
arguments    = yes I do 1 2 3

output       = outputfile.$(ClusterId).$(ProcId).out
error        = errorfile.$(ClusterId).$(ProcId).out
log          = myexe.$(ClusterId).$(ProcId).log

request_cpus   = 1
request_memory = 1024
request_disk   = 10240

#should_transfer_files = yes

queue
```

- executable:
  - Where your executable is.
    - It can be binary or shell script
    - Remember to set executable permission:
      - chmod +x <your job file>

- arguments:
  - Your argument for executable.

- should_transfer_files:
  - yes/no/IF_NEEDED
  - Default: yes

Job submission file (2/4)

```
executable   = /ceph/work/ASGC/felixlee/test.sh
arguments    = yes I do 1 2 3

output       = outputfile.$(ClusterId).$(ProcId).out
error        = errorfile.$(ClusterId).$(ProcId).out
log          = myexe.$(ClusterId).$(ProcId).log

request_cpus   = 1
request_memory = 1024
request_disk   = 10240

#should_transfer_files = yes

queue
```

- output:
  - Job stdout file name
- error:
  - Job stderr file name
- log:
  - Condor job log file name

Job submission file (3/4)

```
executable   = /ceph/work/ASGC/felixlee/test.sh
arguments    = yes I do 1 2 3

output       = outputfile.$(ClusterId).$(ProcId).out
error        = errorfile.$(ClusterId).$(ProcId).out
log          = myexe.$(ClusterId).$(ProcId).log

request_cpus   = 1
request_memory = 1024
request_disk   = 10240

#should_transfer_files = yes

queue
```

- request_cpus:
  - Asking how many cores
- request_memory:
  - Asking memory in MB
- request_disk:
  - Asking disk in KB

Job submission file (4/4)

```
executable   = /ceph/work/ASGC/felixlee/test.sh
arguments    = yes I do 1 2 3

output       = outputfile.$(ClusterId).$(ProcId).out
error        = errorfile.$(ClusterId).$(ProcId).out
log          = myexe.$(ClusterId).$(ProcId).log

request_cpus   = 1
request_memory = 1024
request_disk   = 10240

#should_transfer_files = yes

queue
```

- queue [number]
  - Put job into queue.
    - If follows with numbers, it means queuing # of jobs.
      - e.g. queue 10
    - If no number is given, by default it means queuing one job.

More on job submission file

```
executable   = /ceph/work/ASGC/felixlee/test.sh
arguments    = yes I do 1 2 3

output       = outputfile.$(ClusterId).$(ProcId).out
error        = errorfile.$(ClusterId).$(ProcId).out
log          = myexe.$(ClusterId).$(ProcId).log

request_cpus   = 1
request_memory = 1024
request_disk   = 10240

#should_transfer_files = yes

queue
```

- Submission file supports variables
  - Useful embedded variables
    - $(ClusterId), $(ProcId)
  - You can also define your own variables.
    - MyIndex = "hello"
  - Useful macros:
    - $RANDOM_INTEGER(min, max[, step])
    - $INT(item-to-convert, format-specifier)

Playing with multiple jobs (1/4)

```
executable   = /ceph/work/ASGC/felixlee/test.sh

output       = outputfile.$(ClusterId).$(ProcId).out
error        = errorfile.$(ClusterId).$(ProcId).out
log          = myexe.$(ClusterId).$(ProcId).log

MyIndex = $(ProcId) * 10
arguments = $INT(MyIndex, %04d)

queue 4
```

- Queue multiple jobs with auto-generated arguments.
  - $INT() case.
  - The arguments will be:
    - 0000, 0001, 0002, 0003

Playing with multiple jobs (2/4)

```
executable   = /ceph/work/ASGC/felixlee/test.sh

output       = outputfile.$(ClusterId).$(ProcId).out
error        = errorfile.$(ClusterId).$(ProcId).out
log          = myexe.$(ClusterId).$(ProcId).log

arguments = $RANDOM_INTEGER(0, 100)

queue 4
```

- Queue multiple jobs with auto-generated arguments.
  - $RANDOM_INTEGER() case.

Playing with multiple jobs (3/4)

```
executable   = /ceph/work/ASGC/felixlee/test.sh

output       = outputfile.$(ClusterId).$(ProcId).out
error        = errorfile.$(ClusterId).$(ProcId).out
log          = myexe.$(ClusterId).$(ProcId).log

input = file1
arguments = -a -b 26
queue

input = file2
arguments = -c -d 92
queue
```

- "queue" can be also specified multiple times with different segments.

Playing with multiple jobs (4/4)

```
executable   = /ceph/work/ASGC/felixlee/test.sh

output       = outputfile.$(ClusterId).$(ProcId).out
error        = errorfile.$(ClusterId).$(ProcId).out
log          = myexe.$(ClusterId).$(ProcId).log

queue input, arguments from (
  file1, -a -b 26
  file2, -c -d 92
)
```

- "queue" can be also tuple like:
  - queue [variable] from (
    -a -b 26
    -c -d 92
)

Job flavours (1/2)

```
executable   = /ceph/work/ASGC/felixlee/test.sh
arguments    = yes I do 1 2 3

output       = outputfile
error        = errorfile
log          = myexe.log

request_cpus   = 256
request_memory = 1024
request_disk   = 10240

#should_transfer_files = yes
+JobFlavour = "large"

queue
```

- Used to specify walltime.
  - To avoid resource abuse.

- Usage:
  - +JobFlavour = "keyword"

- If JobFlavour is missing, system will use "default"

Job flavours (2/2)

```
executable   = /ceph/work/ASGC/felixlee/test.sh
arguments    = yes I do 1 2 3

output       = outputfile
error        = errorfile
log          = myexe.log

request_cpus   = 256
request_memory = 1024
request_disk   = 10240

#should_transfer_files = yes
+JobFlavour = "large"

queue
```

- default
  - Walltime = 1 day
  - Cpu = 1
- short
  - Walltime = 3 days
- large
  - Walltime = 14 days
- devel
  - Walltime = 1 hour
- long_serial
  - Walltime = 14 days
  - Cpu = 1

Submitting and monitoring job

- To submit job(s):
  - condor_submit your_jdl_file

```
[felixlee@tidc-ui01 condor]$ condor_submit condor.jdl
Submitting job(s).
1 job(s) submitted to cluster 127.
```

- Monitoring job(s):
  - condor_q

```
[felixlee@tidc-ui01 condor]$ condor_q

-- Schedd: queue@tidc-ui01.grid.sinica.edu.tw : <202.140.187.218:9618?... @ 11/03/23 13:15:23
OWNER      BATCH_NAME     SUBMITTED    DONE    RUN      IDLE   TOTAL JOB_IDS
felixlee ID: 127       11/3   13:14                       1       1 127.0

Total for query: 1 jobs; 0 completed, 0 removed, 1 idle, 0 running, 0 held, 0 suspended
Total for felixlee: 1 jobs; 0 completed, 0 removed, 1 idle, 0 running, 0 held, 0 suspended
Total for all users: 4 jobs; 0 completed, 0 removed, 4 idle, 0 running, 0 held, 0 suspended
```

Handling files with condor 101

- Unlike slurm, when condor job reaches worker node, it won't land at directory where you submitted job from.
  - e.g. you run "condor_submit condor.jdl" from home directory, but condor job will not get executed from your home.
    - Instead, it will create another temporary working directory and run jobs from there.
  - So, recommend to "cd" to global working directory in your script whenever needed.
  - Also, if you don't make should_transfer_files, please put absolute path into executable for your job script.
- Let condor handling output doesn't always work somehow.
  - So, strongly recommend to handle output by your scripts.

Manage your jobs

More on monitoring (1/5)

- condor_q

```
[root@tidc-arc6-1 ~]# condor_q


-- Schedd: queue@tidc-arc6-1.grid.sinica.edu.tw : <202.140.187.104:25714?... @ 11/04/23 09:12:48
OWNER    BATCH_NAME       SUBMITTED    DONE    RUN     IDLE   TOTAL JOB_IDS
cmsuser ID: 173817   11/4  02:47       _       1       _        1 173817.0
cmsuser ID: 173821   11/4  03:08       _       1       _        1 173821.0
cmsuser ID: 173822   11/4  03:09       _       1       _        1 173822.0
cmsuser ID: 173824   11/4  03:10       _       1                1 173824.0
cmsuser ID: 173825   11/4  03:10       _       1                1 173825.0
cmsuser ID: 173826   11/4  03:11       _       1       _        1 173826.0
cmsuser ID: 173827   11/4  03:11       _       1       _        1 173827.0
cmsuser ID: 173931   11/4  04:37       _       _       1        1 173931.0
cmsuser ID: 173935   11/4  04:45       _       _       1        1 173935.0
cmsuser ID: 173937   11/4  04:55       _       _       1        1 173937.0
cmsuser ID: 173944   11/4  05:32       _       _       1        1 173944.0
cmsuser ID: 173946   11/4  05:43       _       _       1        1 173946.0
```

More on monitoring (2/5)

- condor_q

```
[root@tidc-arc6-1 ~]# condor_q


-- Schedd: queue@tidc-arc6-1.grid.sinica.edu.tw : <202.140.187.104:25714?... @ 11/04/23 09:12:48
OWNER    BATCH_NAME        SUBMITTED    DONE    RUN    IDLE   TOTAL  JOB_IDS
cmsuser ID: 173817   11/4  02:47       _       1      _        1   173817.0
cmsuser ID: 173821   11/4  03:08       _       1      _        1   173821.0
cmsuser ID: 173822   11/4  03:09       _       1      _        1   173822.0
cmsuser ID: 173824   11/4  03:10       _       1      _        1   173824.0
cmsuser ID: 173825   11/4  03:10       _       1      _        1   173825.0
cmsuser ID: 173826   11/4  03:11       _       1      _        1   173826.0
cmsuser ID: 173827   11/4  03:11       _       1      _        1   173827.0
cmsuser ID: 173931   11/4  04:37       _       _      1        1   173931.0
cmsuser ID: 173935   11/4  04:45       _       _      1        1   173935.0
cmsuser ID: 173937   11/4  04:55       _       _      1        1   173937.0
cmsuser ID: 173944   11/4  05:32       _       _      1        1   173944.0
cmsuser ID: 173946   11/4  05:43       _       _      1        1   173946.0
```

More on monitoring (3/5)

- condor_q -nobatch

  – ST(job state): R(running), I(Idling), H(Holding), C(Completed)

```
[root@tidc-arc6-1 ~]# condor_q -nobatch

-- Schedd: queue@tidc-arc6-1.grid.sinica.edu.tw : <202.140.187.104:25714?... @ 11/14/23 10:07:05
 ID         OWNER            SUBMITTED     RUN_TIME ST PRI SIZE    CMD
176545.0   cmsuser         11/12 17:14  1+16:52:31 R  0     9766.0 (gridjob)
176548.0   cmsuser         11/12 17:23  1+16:43:17 R  0     9766.0 (gridjob)
176561.0   cmsuser         11/12 18:00  1+16:06:27 R  0     9766.0 (gridjob)
176596.0   cmsuser         11/12 22:40  1+11:27:00 R  0     9766.0 (gridjob)
176715.0   cmsuser         11/13 08:05  1+02:01:08 R  0     9766.0 (gridjob)
176717.0   cmsuser         11/13 08:08  1+01:58:21 R  0     9766.0 (gridjob)
176720.0   cmsuser         11/13 08:11  1+01:55:33 R  0    12208.0 (gridjob)
176721.0   cmsuser         11/13 08:13  1+01:53:11 R  0     9766.0 (gridjob)
176722.0   cmsuser         11/13 08:14  1+01:52:51 R  0     9766.0 (gridjob)
```

More on monitoring (4/5)

- condor_q

Summary of your job status

More on monitoring (5/5)

- condor_q


```
[root@tidc-arc6-1 tmp]# condor_q

-- Schedd: queue@tidc-arc6-1.grid.sinica.edu.tw : <202.140.187.104:25714?... @ 10/25/23 06:03:31
OWNER    BATCH_NAME      SUBMITTED    DONE   RUN    IDLE   TOTAL JOB_IDS
cmsuser ID: 170916  10/24 09:44       _      1      _        1 170916.0
cmsuser ID: 171256  10/25 03:48              1      _        1 171256.0

Total for query: 2 jobs; 0 completed, 0 removed, 0 idle, 2 running, 0 held, 0 suspended
Total for all users: 2 jobs; 0 completed, 0 removed, 0 idle, 2 running, 0 held, 0 suspended
```

Summary of all jobs
(Yours and other user's)

Check jobs from all schedulers

- condor_q -global


```
[felixlee@tidc-ui01 condor]$ condor_q -global

-- Schedd: queue@tidc-arc6-1.grid.sinica.edu.tw : <202.140.187.104:9618?... @ 11/03/23 13:16:21
OWNER BATCH_NAME       SUBMITTED    DONE    RUN    IDLE   HOLD  TOTAL JOB_IDS

Total for query: 0 jobs; 0 completed, 0 removed, 0 idle, 0 running, 0 held, 0 suspended
Total for felixlee: 0 jobs; 0 completed, 0 removed, 0 idle, 0 running, 0 held, 0 suspended
Total for all users: 121 jobs; 0 completed, 0 removed, 25 idle, 96 running, 0 held, 0 suspended


-- Schedd: queue@tidc-ui01.grid.sinica.edu.tw : <202.140.187.218:9618?... @ 11/03/23 13:16:21
OWNER     BATCH_NAME       SUBMITTED    DONE    RUN    IDLE    HOLD   TOTAL JOB_IDS
felixlee ID: 127          11/3   13:14                  1             1 127.0

Total for query: 1 jobs; 0 completed, 0 removed, 1 idle, 0 running, 0 held, 0 suspended
Total for felixlee: 1 jobs; 0 completed, 0 removed, 1 idle, 0 running, 0 held, 0 suspended
Total for all users: 4 jobs; 0 completed, 0 removed, 4 idle, 0 running, 0 held, 0 suspended
```

Analyzing why job doesn't get running

- condor_q -analyze [job id]
  - Job id is the combination of ClusterId and ProcId.
    - e.g. 150.0
      - Where the "150" is ClusterId and "0" is ProcId.
      - The ProcId is serial integer when "queue" multiple jobs.
        - e.g. queue 4, we will get:
          - 150.0, 150.1, 150.2, 150.3
  - You can omit ProcId, it will query all jobs under the same ClusterId.

Analyzing why job doesn't get running

- condor_q -analyze [job id]


```
[root@tidc-arc6-1 ~]# condor_q -analyze 173931.0

-- Schedd: queue@tidc-arc6-1.grid.sinica.edu.tw : <202.140.187.104:25714?...
No successful match recorded.
Last failed match: Sat Nov  4 09:12:20 2023

Reason for last match failure: no match found

173931.000:  Run analysis summary ignoring user priority.  Of 6 machines,
      0 are rejected by your job's requirements
      0 reject your job because of their own requirements
      0 match and are already running your jobs
      0 match but are serving other users
      6 are able to run your job
```

The case when job is simply queuing.

Analyzing why job doesn't get running

- condor_q -analyze [job id]

```
[felixlee@tidc-ui01 condor]$ condor_q -analyze 135


-- Schedd: queue@tidc-ui01.grid.sinica.edu.tw : <202.140.187.218:9618?...
The Requirements expression for job 135.000 is

    (TARGET.Arch == "X86_64") && (TARGET.OpSys == "LINUX") && (TARGET.Disk >= RequestDisk) && (TARGET.Memory >= RequestMemory) &&
    (TARGET.Cpus >= RequestCpus) && ((TARGET.FileSystemDomain == MY.FileSystemDomain) || (TARGET.HasFileTransfer))



135.000:  Run analysis summary ignoring user priority.  Of 6 machines,
      6 are rejected by your job's requirements
      0 reject your job because of their own requirements
      0 match and are already running your jobs
      0 match but are serving other users
      0 are able to run your job

WARNING:  Be advised:
   No machines matched the jobs's constraints
```

The case when job is rejected by worker nodes.

Get more detailed job match analysis

- condor_q -better-analyze [job id]

Get more detailed job match analysis

- condor_q -better-analyze [job id]

```
[felixlee@tidc-ui01 condor]$ condor_q -better-analyze 135

-- Schedd: queue@tidc-ui01.grid.sinica.edu.tw : <202.140.187.218:9618?...
The Requirements expression for job 135.000 is

    (TARGET.Arch == "X86_64") && (TARGET.OpSys == "LINUX") && (TARGET.Disk >= RequestDisk) && (TARGET.Memory >= RequestMemory) &&
    (TARGET.Cpus >= RequestCpus) && ((TARGET.FileSystemDomain == MY.FileSystemDomain) || (TARGET.HasFileTransfer))

Job 135.000 defines the following attributes:

    FileSystemDomain = "tidc-ui01.grid.sinica.edu.tw"
    RequestCpus = 256
    RequestDisk = 10240
    RequestMemory = 1024

The Requirements expression for job 135.000 reduces to these conditions:

          Slots
Step    Matched   Condition
-----   --------  ---------
[0]          103  TARGET.Arch == "X86_64"
[1]          103  TARGET.OpSys == "LINUX"
[3]          103  TARGET.Disk >= RequestDisk
[5]          101  TARGET.Memory >= RequestMemory
[7]            0  TARGET.Cpus >= RequestCpus
[10]         103  TARGET.HasFileTransfer
```

The case when job is rejected by worker nodes.

Get more detailed job match analysis

- condor_q -better-analyze [job id]

```
[felixlee@tidc-ui01 condor]$ condor_q -better-analyze 135

-- Schedd: queue@tidc-ui01.grid.sinica.edu.tw : <202.140.187.218:9618?...
The Requirements expression for job 135.000 is

    (TARGET.Arch == "X86_64") && (TARGET.OpSys == "LINUX") && (TARGET.Disk >= RequestDisk) && (TARGET.Memory >= RequestMemory) &&
    (TARGET.Cpus >= RequestCpus) && ((TARGET.FileSystemDomain == MY.FileSystemDomain) || (TARGET.HasFileTransfer))

Job 135.000 defines the following attributes:

    FileSystemDomain = "tidc-ui01.grid.sinica.edu.tw"          All of your job requirements and conditions
    RequestCpus = 256
    RequestDisk = 10240
    RequestMemory = 1024

The Requirements expression for job 135.000 reduces to these conditions:

            Slots
Step     Matched   Condition
-----    --------  ---------
[0]           103  TARGET.Arch == "X86_64"
[1]           103  TARGET.OpSys == "LINUX"
[3]           103  TARGET.Disk >= RequestDisk          Detailed matching status
[5]           101  TARGET.Memory >= RequestMemory
[7]             0  TARGET.Cpus >= RequestCpus
[10]          103  TARGET.HasFileTransfer
```

Get more detailed job match analysis

- condor_q -better-analyze [job id]

```
[felixlee@tidc-ui01 condor]$ condor_q -better-analyze 135

-- Schedd: queue@tidc-ui01.grid.sinica.edu.tw : <202.140.187.218:9618?...
The Requirements expression for job 135.000 is

    (TARGET.Arch == "X86_64") && (TARGET.OpSys == "LINUX") && (TARGET.Disk >= RequestDisk) && (TARGET.Memory >= RequestMemory) &&
    (TARGET.Cpus >= RequestCpus) && ((TARGET.FileSystemDomain == MY.FileSystemDomain) || (TARGET.HasFileTransfer))

Job 135.000 defines the following attributes:

    FileSystemDomain = "tidc-ui01.grid.sinica.edu.tw"
    RequestCpus = 256
    RequestDisk = 10240
    RequestMemory = 1024

The Requirements expression for job 135.000 reduces to these conditions:

         Slots
Step   Matched  Condition
-----  -------- ---------
[0]        103  TARGET.Arch == "X86_64"
[1]        103  TARGET.OpSys == "LINUX"
[3]        103  TARGET.Disk >= RequestDisk
[5]        101  TARGET.Memory >= RequestMemory
[7]          0  TARGET.Cpus >= RequestCpus
[10]       103  TARGET.HasFileTransfer
```

Check the "Slots Matched" column, where the value is "0"

Get more detailed job match analysis

- condor_q -better-analyze [job id]

```
[felixlee@tidc-ui01 condor]$ condor_q -better-analyze 135

-- Schedd: queue@tidc-ui01.grid.sinica.edu.tw : <202.140.187.218:9618?...
The Requirements expression for job 135.000 is

    (TARGET.Arch == "X86_64") && (TARGET.OpSys == "LINUX") && (TARGET.Disk >= RequestDisk) && (TARGET.Memory >= RequestMemory) &&
    (TARGET.Cpus >= RequestCpus) && ((TARGET.FileSystemDomain == MY.FileSystemDomain) || (TARGET.HasFileTransfer))

Job 135.000 defines the following attributes:          Asking too many CPUs...

    FileSystemDomain = "tidc-ui01.grid.sinica.edu.tw"
    RequestCpus = 256
    RequestDisk = 10240
    RequestMemory = 1024

The Requirements expression for job 135.000 reduces to these conditions:

         Slots
Step    Matched   Condition            Check the "Slots Matched" column,
-----   --------  ---------            where the value is "0"
[0]         103   TARGET.Arch == "X86_64"
[1]         103   TARGET.OpSys == "LINUX"
[3]         103   TARGET.Disk >= RequestDisk
[5]         101   TARGET.Memory >= RequestMemory
[7]           0   TARGET.Cpus >= RequestCpus
[10]        103   TARGET.HasFileTransfer
```

More on condor_q

- condor_q -help

Delete your jobs

- condor_rm <your job id1> [<job id2> … <job idn>]
  - Delete your job by job id, where the job id can be specified multiple times:
    - `condor_rm 11 12 10`
  - Be aware of that, if you specific job id without ProcId, it means to delete all ProcId under the same ClusterId.
    - e.g. condor_rm 11 means to delete 11.0, 11.1,…, 11.x
    - And, condor_rm 11.0 means to delete only 11.0
- condor_rm -all
  - It will delete all of your jobs, use it carefully…
- condor_rm -help
  - More options on condor_rm

Other commands

- condor_release <job id>
  - Used when job is in hold state.
  - Usually, the jobs will be held by several reasons.
    - Schedd machine gets rebooted.
    - Worker encounters so problems.
    - You hold it by yourself with condor_hold.
  - The condor_release will get job restarted from fresh.
- condor_hold <job id>
  - Suspend your job, it can be resumed by condor_release.

Data flow and data handling

EOS space (1/3)

- How to access EOS via local cluster?
  - It can be accessed by xrootd tool with xrootd url (w/o Grid proxy, it's **Read-only**)
    - Xrootd url: root://tidc-smstor1.grid.sinica.edu.tw//<eos path>
    - Xrdcp
      - `xrdcp   <xrootd url>  .`
    - C++ or python Root API
      - std::unique_ptr<TFile> myFile(TFile::Open("root://tidc-smstor1.grid.sinica.edu.tw//eos/cms/store/user/felixlee/file.root"));
  - Or simply by generic Unix file operation. (fuse mount)
    - `ls  /eos/cms/store/`
    - `cp  /eos/cms/store/user/felix/file.root   .`
- Current EOS directory structure:
  - /eos/cms/store/data   -- <CMS production data>
  - /eos/cms/store/mc   -- <CMS mc production data>
  - /eos/cms/store/user   -- <CMS user data>

EOS space (2/3)

- Unix file operation with fuse:

```
[felixlee@tidc-ui01 condor]$ cp /eos/cms/store/data/Run2016B/MET/AOD/21Feb2020_ver2_UL2016_HIPM-v1/230000/91574BBC-89FB-BC49-8DBB-40FCDA42125
6.root /tmp/
[felixlee@tidc-ui01 condor]$ file /tmp/91574BBC-89FB-BC49-8DBB-40FCDA421256.root
/tmp/91574BBC-89FB-BC49-8DBB-40FCDA421256.root: ROOT file Version 61409 (Compression: 109)
```

- xrdcp:

```
[felixlee@tidc-ui01 condor]$ xrdcp root://tidc-smstor1.grid.sinica.edu.tw//eos/cms/store/data/Run2016B/MET/AOD/21Feb2020_ver2_UL2016_HIPM-v1/
230000/91574BBC-89FB-BC49-8DBB-40FCDA421256.root /tmp/
231109 08:34:31 1186531 cryptossl_X509CreateProxy: EEC certificate has expired
[752.6MB/752.6MB][100%][===============================================][125.4MB/s]
```

EOS space (3/3)

- Please bear in mind
  - Read only access is only available within TIDC facilities whether by xrdcp or unix file operation.
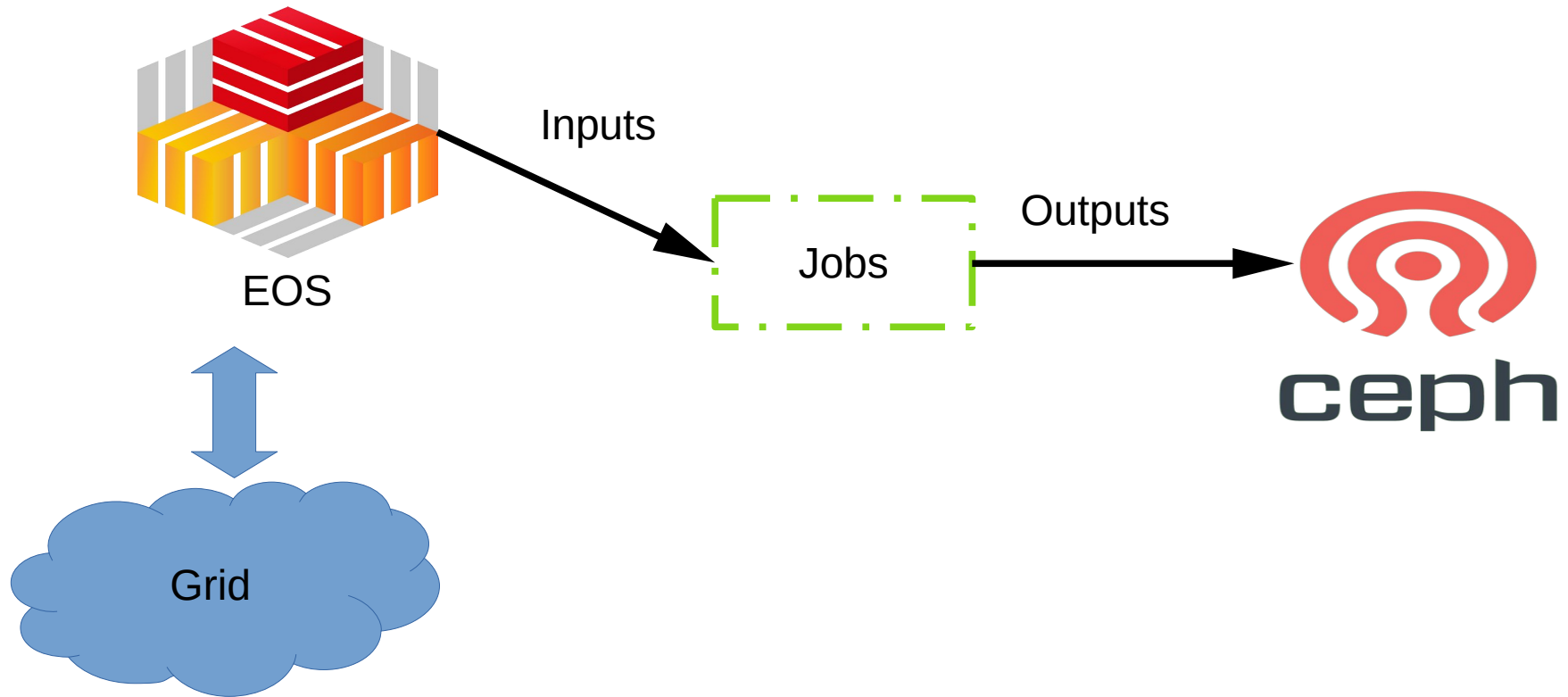   - You may also leverage scp or sftp via tidc-ui. e.g.:
      `scp tidc-ui01.grid.sinica.edu.tw:/eos/<xxx>/<xx>/myfile.root  ~/`
  - If you wanna access EOS outside TIDC facility or writable permission, you will need grid certificate and CMS VO.
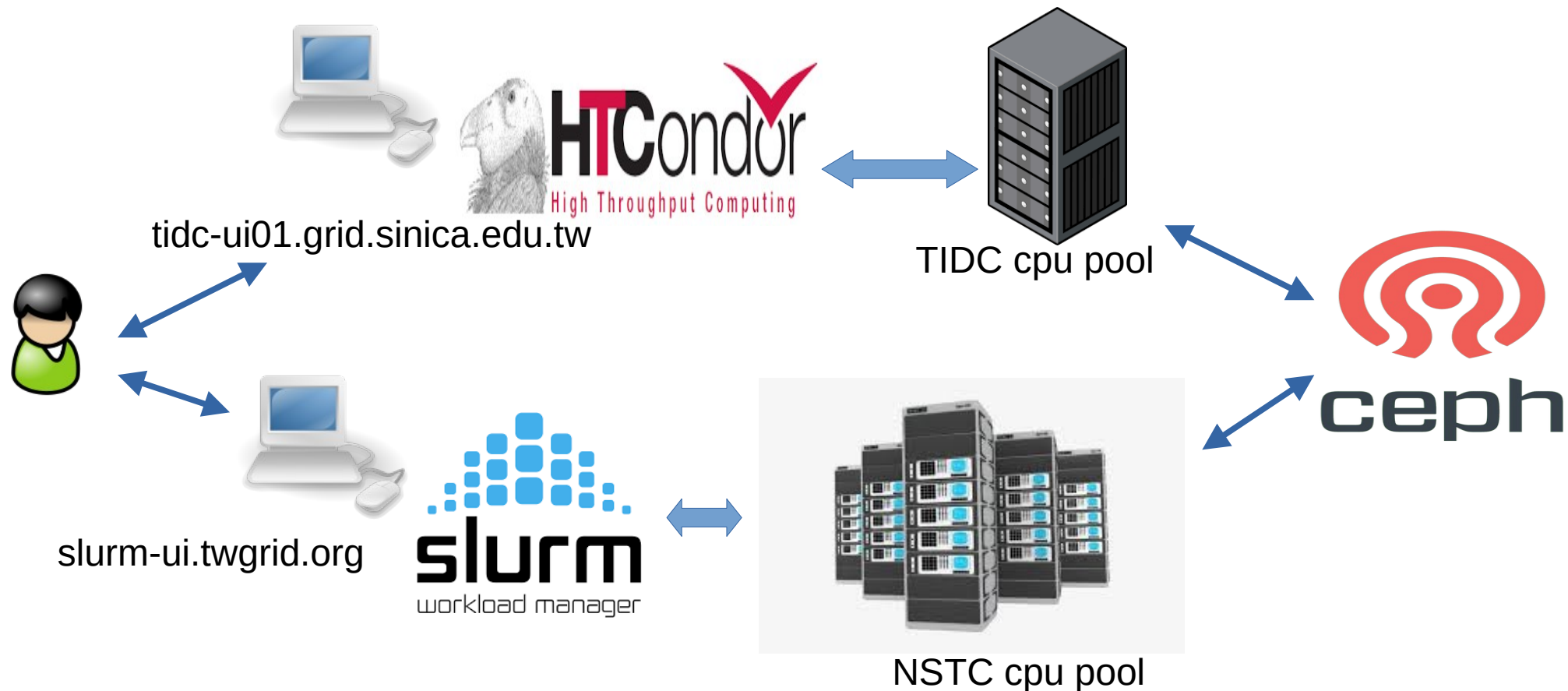
Ceph space

- /ceph/work/<group>/ -- <group directory, 3TB free>
- It's accessible by generic Unix file operation.

```
[felixlee@tidc-ui01 condor]$ echo "yes yes yes" > /ceph/work/ASGC/felixlee/test1.txt
[felixlee@tidc-ui01 condor]$ cat /ceph/work/ASGC/felixlee/test1.txt
yes yes yes
[felixlee@tidc-ui01 condor]$
```
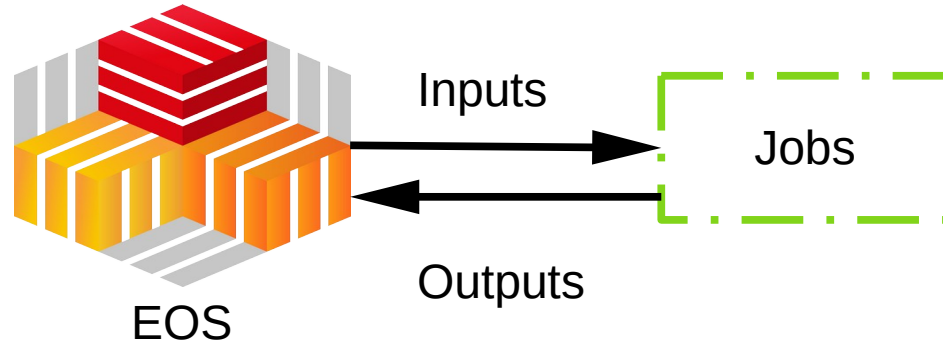
Typical data flow with job(1/2)

Typical data flow with job (2/2)

tidc-ui01.grid.sinica.edu.tw

TIDC cpu pool

slurm-ui.twgrid.org

NSTC cpu pool

ceph

EOS data flow only

- This work flow needs to define new EOS space outside Grid.
  - **Better not** to write back to /eos/cms/store/user directly, because it would disturb CMS's own accounting system.
    - Those data will become dark data, and will be purged routinely.
- Defining a private EOS space for local usage is feasible, but needs to get consensus on:
  - Capacity, Quota, ACL, ETC.

Example Job with data handling

```bash
#!/bin/bash
echo "Change to working directory: /ceph/work/<group name>"
cd /ceph/work/ASGC/felixlee/
pwd
echo "Za Warudo!"
sleep 9
echo "my argument: $@"

### EOS data access, path: /eos/cms
cp /eos/cms/store/data/Run2016B/MET/AOD/21Feb2020_ver2_UL2016_HIPM-
v1/230000/91574BBC-89FB-BC49-8DBB-40FCDA421256.root .
ls -l 91574BBC-89FB-BC49-8DBB-40FCDA421256.root
file 91574BBC-89FB-BC49-8DBB-40FCDA421256.root

### Ceph filesystem access, path: /ceph/work/<group name>
ls -l /ceph/work/ASGC/felixlee/91574BBC-89FB-BC49-8DBB-40FCDA421256.root

echo "job finished"
```

Wrap-up (access end-point)

- UI:
  - Condor + TIDC worker nodes
    - tidc-ui[01~04].grid.sinica.edu.tw
  - Slurm + NSTC worker nodes
    - slurm-ui.twgrid.org
- Storage:
  - TIDC EOS
    - /eos/cms/store/
    - root://tidc-smstor1.grid.sinica.edu.tw//eos/cms/store
  - Ceph
    - /ceph/work/<group name>

Wrap-up (Useful commands)

- condor_submit <job file>
  - Submit condor job by job file
- condor_rm <job id>
  - Delete condor job by id
- condor_q
  - Query job status
- condor_q -analyze
  - Briefly check why job doesn't get running.
- condor_q -better-analyze
  - Check why job doesn't get running with more information.
- xrdcp <EOS URL>
  - Download data from EOS.

Tutorial files

- /ceph/sharedfs/software/tutorial/condor/
  - Jobmission files:
    - condor.jdl
    - condor_queue_index2.jdl
    - condor_queue_random.jdl
  - Job script
    - test.sh
- Copy them to your home, if you wanna play with it.
  - `cp  /ceph/sharedfs/software/tutorial/condor/*  ~/`