FUNDI COMPUTING RESOURCES

MPIFR, 11 FEB 2025, BONN

OUTLINE

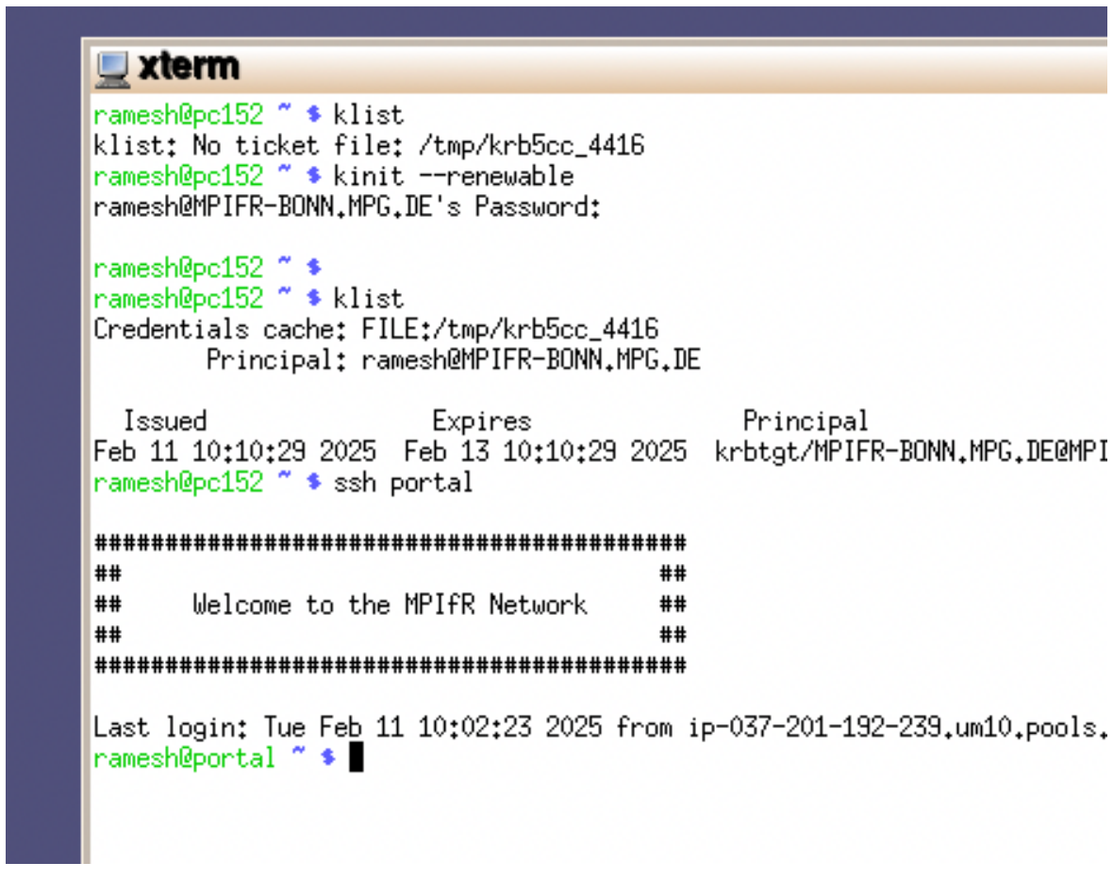
- Group Computing Resources
- HPC - Queue managers
- RISC or CISC computing?
- Some useful compute tips

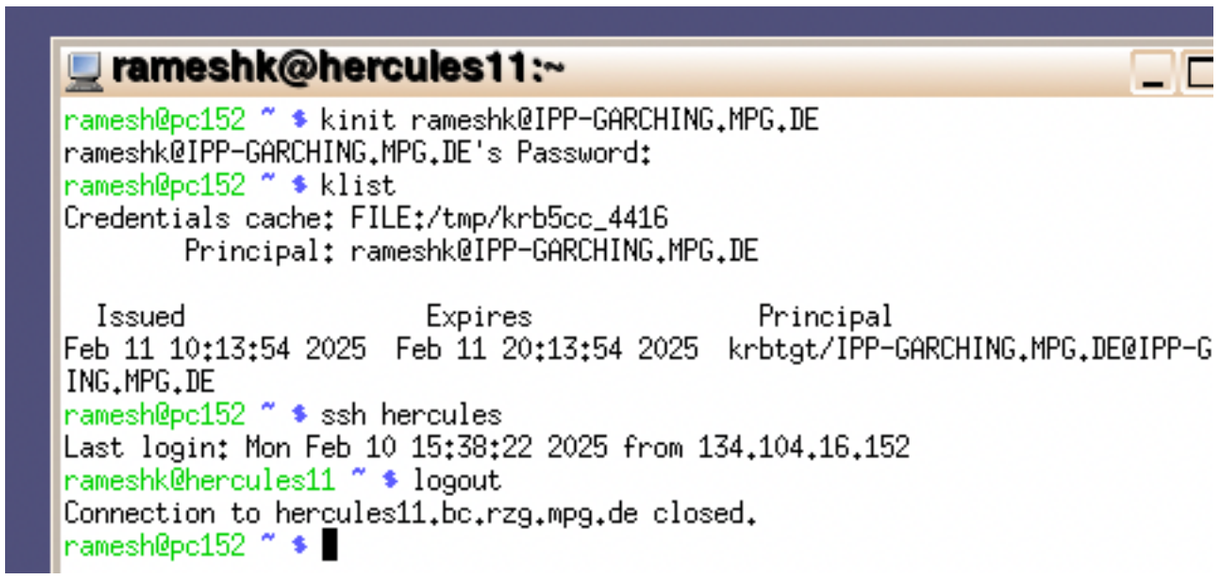
PULSAR GROUP SERVERS

dogmatix3/dogmatix4	AMD EPYC 9634 - 84-cores 512GB RAM, 2.25 - 3.7 GHz
dogmatix0/dogmatix1	Xeon 5118, 24 Cores 192GB, 2.3 GHz
instantmix	Intel Xeon 6126, 24 cores 324 GB RAM, 2.6 GHz
miraculix2	Xeon E5-2687W v3, 20 cores 128GB RAM, 3.1 GHz
miraculix	Intel Xeon X7550 - 64-cores 64GB RAM, 2GHz

COMPUTE INFRASTRUCTURE– KERBEROS

- Institute-wide services - portal/login
  - kerberos tickets + ssh





Renew ticket in the background, eg. in bash:
ramesh@pc152 ~ $ (for x in {0..60};do sleep 86400;kinit -R;done) >& /dev/null&

PULSAR GROUP STORAGE

/fpra/soft/01	25T	179G	25T	1 %	pstorage:/srv/storage_12/soft
/fpra/timing/01	128T	108T	20T	85 %	bigfoot1:/srv/storage_10/timing
/fpra/globc/01	128T	112T	16T	88 %	bigfoot2:/srv/storage_11/globc
/fpra/mkat/01	128T	101T	28T	79 %	bigfoot1:/srv/storage_12/mkat
/fpra/comiss/01	128T	82T	46T	64 %	bigfoot2:/srv/storage_12/comiss
/fpra/bband/01	128T	91T	37T	72 %	bigfoot1:/srv/storage_11/bband
/fpra/polmag/01	128T	125T	3.1T	98 %	bigfoot4:/srv/storage_12/polmag
/fpra/polmag/02	153T	133T	21T	87 %	bigfoot5:/srv/storage_12/polmag
/fpra/galc/01	128T	110T	19T	86 %	bigfoot4:/srv/storage_11/galc
/fpra/galc/02	153T	145T	7.9T	95 %	bigfoot5:/srv/storage_11/galc
/fpra/galc/03	153T	44T	110T	29 %	miraculix2:/srv/storage_10/galc
/fpra/galc/04	85T	15T	70T	18 %	miraculix2:/srv/storage_11/galc
/fpra/lofar/01	128T	122T	5.9T	96 %	bigfoot4:/srv/storage_10/lofar
/fpra/lofar/02	153T	147T	6.6T	96 %	bigfoot5:/srv/storage_10/lofar
/fpra/search/01	128T	40T	88T	32 %	bigfoot3:/srv/storage_10/search
/fpra/search/02	128T	18T	111T	14 %	bigfoot3:/srv/storage_11/search
/fpra/search/03	128T	57T	71T	45 %	bigfoot3:/srv/storage_12/search
/fpra/bursts/01	128T	125T	3.0T	98 %	bigfoot2:/srv/storage_10/bursts
/fpra/bursts/02	315T	19T	296T	6 %	pstorage:/srv/storage_10/bursts
/fpra/bursts/03	340T	114T	227T	34 %	pstorage:/srv/storage_11/bursts

- ~3PB storage
- 11 groups
- RAID6 + mirrored
- RAID6 only

COMPUTING – EFFELSBERG

EDGAR	30x AMD EPYC 9334 2x L40 nVidia GPUs
paf0/paf1 (psrdaq0/psrdaq1)	2x GTX Titan X + beegfs (655TB)

- EDGAR storage
  - /beegfsEDD/EDD_pipeline_data/production/pipeline_data/timing[1-5]
  - /beegfsEDD/EDD_pipeline_data/production/pipeline_data/search[1-5]
  - /beegfsEDD/EDD_pipeline_data/production/pipeline_data/baseband[1-5]
- 4PB storage
- 80GB/s throughput

COMPUTING – HERCULES CLUSTER

- Check https://fpra.mpifr-bonn.mpg.de/doku.php?id=computing
- Apply hercules account: https://selfservice.mpcdf.mpg.de/register/antrag.php?lang=en

2x login nodes	hercules11.bc.rzg.mpg.de hercules12.bc.rzg.mpg.de
32x compute nodes	2x Xeon 8268 CPUs 48cores @ 2.9 GHz 377 GB RAM
54x GPU nodes	2x Xeon 8268 CPUs 3x nVidia RTX 6000 48cores @ 2.9 GHz 377 GB RAM
25 Gb/s ethernet interconnects	

The image on the left shows the Hercules2 cluster as installed in the Max-Planck Computing and Data Facility (MPCDF), Garching. The GPU powering the cluster is displayed on top.

- 2 login, 32 CPU and 54 GPU nodes
- 377GB RAM + 2x24-core Xeon 8268 per node
- 3x nVidia Quadro RTX 6000 in GPU nodes
- 1.28 PB storage, 10Gb/s link to Bonn
- 25Gb/s ethernet interconnect
- Slurm-based resource manager

- RTX 6000 w/ 24GB RAM
- 4608 compute cores
- 16.3 TFLOPS F32 performance
- 672GB/s memory bandwidth
- Active cooling

COMPUTING – HERCULES-GATEWAY DATA TRANSFERS

```
ramesh@pc152 ~ $ ssh hgw
ramesh@hgw ~ $ df
Filesystem                     1K-blocks       Used    Available Use% Mounted on
udev                          23931508         0    23931508   0% /dev
tmpfs                          4795684       1904     4793780   1% /run
/dev/mapper/vgsys-root                23058432    13157924     9900508  58% /
tmpfs                         23978400         0    23978400   0% /dev/shm
tmpfs                             5120         0        5120   0% /run/lock
tmpfs                         23978400         0    23978400   0% /sys/fs/cgroup
/dev/md0                      1011148      437588      504852  47% /boot
hc_mkfs                      421860999168  247846875136 174014124032  59% /media/mkfs
hercules_mandap               362924736512  239205281792 123719454720  66% /media/mandap
hercules_fs                   544387104768  250599202816 293787901952  47% /media/hercules
zsre:/homes/ramesh              699748352    14433280         685315072   3% /homes/ramesh
134.104.18.81:/srv/storage_10/timing 136720689152 115694475264  21026213888  85% /fpra/timing/01
ramesh@hgw ~ $
ramesh@hgw ~ $ cd /media/hercules/results/rameshk/
ramesh@hgw /media/hercules/results/rameshk $ cd /fpra/timing/01
ramesh@hgw /fpra/timing/01 $
```

COMPUTING – CONNECTION TO HERCULES

- Check https://fpra.mpifr-bonn.mpg.de/doku.php?id=computing
- Apply hercules account: https://selfservice.mpcdf.mpg.de/register/antrag.php?lang=en

ssh from MPIfR machines, eg. portal/pc152

```
ramesh@pc152 ~ $ ssh rameshk@hercules11.bc.rzg.mpg.de
```

Set up a tunnel (MPIfR eg. pc152)

```
ramesh@pc152 ~ $ ssh -L 5912:hercules11:5903 rameshk@hercules11.bc.rzg.mpg.de
```

Setting up VNC on hercules

```
rameshk@hercules11 ~ $ vncserver -geometry 1200x980 -depth 24 :3
```

Connect on pc152

```
ramesh@pc152 ~ $ vncviewer localhost:12
```

COMPUTING – OTHER CLUSTERS IN GARCHING

raven

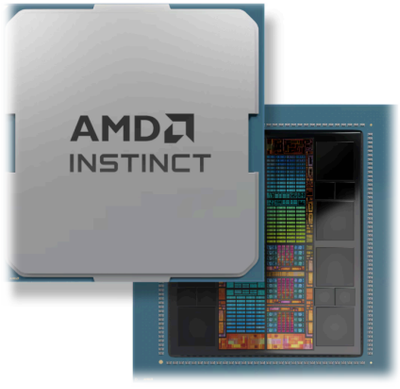
- 1592 nodes, 114624 cores
- 192 nodes (768 A100 GPUs)

viper

- 768 nodes, 98304 cores
- 512 GB/768GB/1TB RAM

viper-GPU

- 300 nodes
- 600 AMD APUs



AMD M1300A APU

- 14592 stream processors
- ~ 1Pflop single precision, 62 Tflops double precision
- 24-core AMD EPYC CPU, Zen4
- 128 GB GPU RAM

SLURM – HERCULES

- Simple Linux Utility for Resource Management

```
rameshk@hercules11 ~ $ sinfo -a
PARTITION     AVAIL  TIMELIMIT  NODES  STATE NODELIST
parallel.q       up 10-00:00:0      6    mix hc[201-206]
parallel.q       up 10-00:00:0      7  alloc hc[207,210-214,226]
parallel.q       up 10-00:00:0     17   idle hc[208-209,215-225,
long.q           up 10-00:00:0      6    mix hc[201-206]
long.q           up 10-00:00:0      7  alloc hc[207,210-214,226]
long.q           up 10-00:00:0     17   idle hc[208-209,215-225,
short.q*         up    4:00:00      6    mix hc[201-206]
short.q*         up    4:00:00      7  alloc hc[207,210-214,226]
short.q*         up    4:00:00     73   idle hc[208-209,215-225,
gpu.q            up 10-00:00:0     52   idle hcg[001-052]
gpu42cores.q     up 10-00:00:0     52   idle hcg[001-052]
gpu6cores.q      up 10-00:00:0     52   idle hcg[001-052]
interactive.q    up    6:00:00      4   idle hc[231-232],hcg[053
project.q        up 10-00:00:0      6    mix hc[201-206]
project.q        up 10-00:00:0      7  alloc hc[207,210-214,226]
project.q        up 10-00:00:0     73   idle hc[208-209,215-225,
rzgmon       inact       1:00      2   idle hercules[11-12]
rameshk@hercules11 ~ $
```

```
rameshk@hercules11 ~ $ squeue -a | head -20
             JOBID PARTITION     NAME     USER ST       TIME  NODES NODELIST(REASON)
          28982540    long.q  HP_tune dbhatnag  R 3-23:40:24      1 hc201
          29005692    long.q     3mhz    jjang  R    4:26:20      1 hc207
          28983750    long.q     bash   vishnu  R 3-21:04:09      1 hc226
          28992856 parallel. anj_try3    pdeni  R 1-23:33:21      5 hc[210-214]
29008889_[685-1000   short.q    chisq    pdeni PD       0:00      1 (AssocMaxJobsLimit)
   29009495_[0-39]   short.q     sims    pdeni PD       0:00      1 (AssocMaxJobsLimit)
          29008669   short.q     bash   jjawor  R    1:03:13      1 hc206
      29008889_684   short.q    chisq    pdeni  R       0:03      1 hc204
      29008889_683   short.q    chisq    pdeni  R       0:06      1 hc204
      29008889_681   short.q    chisq    pdeni  R       0:09      1 hc204
      29008889_682   short.q    chisq    pdeni  R       0:09      1 hc204
      29008889_680   short.q    chisq    pdeni  R       0:12      1 hc204
      29008889_679   short.q    chisq    pdeni  R       0:29      1 hc204
      29008889_678   short.q    chisq    pdeni  R       0:35      1 hc205
      29008889_677   short.q    chisq    pdeni  R       0:38      1 hc203
      29008889_676   short.q    chisq    pdeni  R       0:41      1 hc205
      29008889_675   short.q    chisq    pdeni  R       1:09      1 hc205
      29008889_674   short.q    chisq    pdeni  R       1:16      1 hc203
      29008889_673   short.q    chisq    pdeni  R       1:24      1 hc203
rameshk@hercules11 ~ $ squeue -a | wc -l
207
rameshk@hercules11 ~ $
```

sinfo/squeue/srun/sbatch/scancel/scontrol etc.

For details, see https://slurm.schedmd.com/job_array.html

SLURM – HERCULES PARTITIONS

short.q	default/86 nodes/4h-minmax/1n-per-job/370GB max
parallel.q	32 nodes/24hmin/240hmax/32n-per-job/370GB max
gpu.q	54 nodes/24hmin/240hmax/48cpu-node/370GB max
gpu42cores.q	54 nodes/24hmin/240hmax/42cpu-node/370GB max
gpu6cores.q	54 nodes/24hmin/240hmax/6cpu-node/370GB max
interactive.q	4 nodes/2h min/12h max/370GB max

For details see https://docs.mpcdf.mpg.de/doc/computing/clusters/systems/Radioastronomy.html

SLURM – SOME BASICS

```
rameshk@hercules11 ~/batchscripts $ cat 2021-11-23-02:41:59.0075
#!/bin/bash
#SBATCH --job-name=2021-11-23-02:41:59.0075
#SBATCH --output=/u/rameshk/batchlogs/2021-11-23-02:41:59.0075.log
#SBATCH --partition=short.q
#SBATCH --ntasks=1
#SBATCH --mem=40g
#SBATCH --time=02:00:00
#SBATCH --mail-type=FAIL,TIME_LIMIT --mail-user=ramesh.karuppusamy@gmail.com
mkdir -p /hercules/results/rameshk/0737allSP/UHF/2021-11-23-02:41:59
cd /hercules/results/rameshk/0737allSP/UHF/2021-11-23-02:41:59
singularity exec -B /u/rameshk -B /hercules/results/rameshk -B /mkfs/02/search /u/rameshk/dspsr_202111.sif dspsr -scloffs -O 2021-11-23-02
r.psh /mkfs/02/search/J0737-3039A/2021-11-23-02:41:59/4/816/2021-11-23-03:11:54.sf
singularity exec -B /u/rameshk -B /hercules/results/rameshk -B /mkfs/02/search /u/rameshk/dspsr_202111.sif pam -p -F --setnbin=512 -e FF 2
singularity exec -B /u/rameshk -B /hercules/results/rameshk -B /mkfs/02/search /u/rameshk/dspsr_202111.sif python /u/rameshk/getSPstats_cl
```

script for the queue

#!/bin/bash
#SBATCH --job-name=2021-11-23-02:41:59.0075
#SBATCH --output=/u/rameshk/batchlogs/2021-11-23-02:41:59.0075.log
#SBATCH —error=/u/rameshk/stderr/2021-11-23-02:41:59.0075.err
#SBATCH --partition=short.q
#SBATCH --ntasks=1
#SBATCH --mem=400
#SBATCH --time=02+00:00
#BATCH --mail-type=FAIL,TIME_LIMIT --mail-user=ramesh.karuppusang@goail.com
< stuff that you want to do>

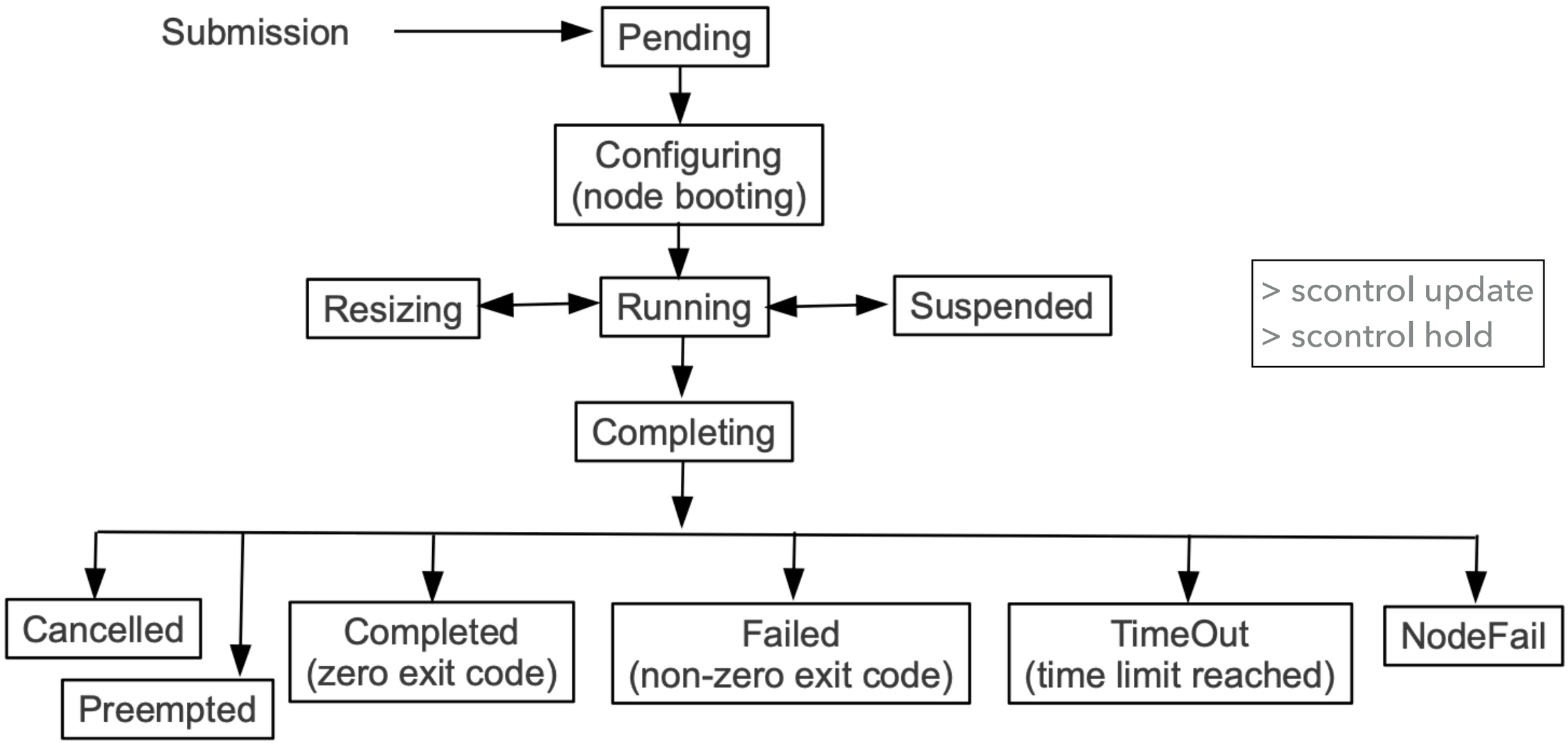
And finally, sbatch <script name>

interactive job

```
rameshk@hercules11 ~ $ salloc --time=01:00:00 -c 4 --mem-per-cpu=40G -p interactive.q
salloc: Pending job allocation 29025075
salloc: job 29025075 queued and waiting for resources
salloc: job 29025075 has been allocated resources
salloc: Granted job allocation 29025075
salloc: Waiting for resource configuration
salloc: Nodes hc231 are ready for job
rameshk@hercules11 ~ $ srun --pty bash
rameshk@hc231 ~ $

rameshk@hercules11 ~ $ srun --pty --nodes=2 --ntasks=1 --mem=16G --time=01:00:00 bash
srun: Warning: can't run 1 processes on 2 nodes, setting nnodes to 1
srun: job 29025074 queued and waiting for resources
srun: job 29025074 has been allocated resources
rameshk@hc206 ~ $
```

SLURM – THE BATCH SYSTEM



SchedMD LLC
http://www.schedmd.com

HTCONDOR – EDGAR

- For Effelsberg observers

get an account, and ssh in to esched

ramesh@pc152 ~ $ ssh esched

```
ramesh@esched ~/htcondor $ cat submit_jobs.condor
# File: submit_jobs.condor

# Define the executable to run
Executable = /usr/bin/apptainer
transfer_executable = False

initialdir = /beegfsEDD/PULSAR_MIGRATION/0737Results/TEST

# Define the log, output, and error files
Log = job_$(Process).log
Output = job_$(Process).out
Error = job_$(Process).err

# Define the directories to process
arguments = exec -B /beegfsEDD /homes/ramesh/dspsr_202111.sif python /homes/ramesh/htcondor/getSPstats.py $(dir)

# Transfer the executable and input files
should_transfer_files = YES
when_to_transfer_output = ON_EXIT

# Queue a job for each directory
queue dir from directories.txt

ramesh@esched ~/htcondor $ _
```

ramesh@esched ~ $ condor_status

ramesh@esched ~ $ condor_submit submit_jobs.condor

ramesh@esched ~/htcondor $ cat directories.txt

/beegfsEDD/PULSAR_MIGRATION/0737Results/TEST

STORAGE – COLD

- Hierarchical Storage Management system
- arcsrv1 - ivdfile/ivdjob/rsync

- pushing files to tape
  - rsync -av –progress <src> <dest>
  - check staging area - /srv/ivdfs_03
  - ivdfile –list
  - ivdfile –migrate
  - ivdfile –release
- batch recalls
  - /bin/ls -1 *.dada > ~/flist
  - ivdfile –recall –if ~/flist

```
rameshk@hercules11:~
ramesh@arcsrv1 /srv/ivdfs_03/LEAP-BKUP/20110914_EFF_1310 $ ivdjob -l
No jobs running.
ramesh@arcsrv1 /srv/ivdfs_03/LEAP-BKUP/20110914_EFF_1310 $ ivdfile --list 2011-09-14-15:57:32_00000*.*
FileID  Status   FileName
 19003  online   2011-09-14-15:57:32_0000000000000000.000000.dada
 19004  online   2011-09-14-15:57:32_0000000800000000.000000.dada
 19005  online   2011-09-14-15:57:32_0000001600000000.000000.dada
 19006  online   2011-09-14-15:57:32_0000002400000000.000000.dada
 19007  online   2011-09-14-15:57:32_0000003200000000.000000.dada
 19008  online   2011-09-14-15:57:32_0000004000000000.000000.dada
 19009  online   2011-09-14-15:57:32_0000004800000000.000000.dada
 19010  online   2011-09-14-15:57:32_0000005600000000.000000.dada
 19011  online   2011-09-14-15:57:32_0000006400000000.000000.dada
 19012  online   2011-09-14-15:57:32_0000007200000000.000000.dada
 19013  online   2011-09-14-15:57:32_0000008000000000.000000.dada
 19014  online   2011-09-14-15:57:32_0000008800000000.000000.dada
 19015  online   2011-09-14-15:57:32_0000009600000000.000000.dada
 19016  offline  2011-09-14-15:57:32_0000010400000000.000000.dada
 19017  offline  2011-09-14-15:57:32_0000011200000000.000000.dada
 19018  offline  2011-09-14-15:57:32_0000012000000000.000000.dada
 19019  offline  2011-09-14-15:57:32_0000012800000000.000000.dada
ramesh@arcsrv1 /srv/ivdfs_03/LEAP-BKUP/20110914_EFF_1310 $ ivdfile --recall 2011-09-14-15:57:32_000000*.*
ramesh@arcsrv1 /srv/ivdfs_03/LEAP-BKUP/20110914_EFF_1310 $
```

```
ramesh@arcsrv1 /srv/ivdfs_03/LEAP-BKUP $ du -h --apparent-size /srv/ivdfs_03/LEAP-BKUP/
1.6T    /srv/ivdfs_03/LEAP-BKUP/20120130_EFF_1390
937G    /srv/ivdfs_03/LEAP-BKUP/20120107_EFF_1310
495G    /srv/ivdfs_03/LEAP-BKUP/20120129_EFF_1350
1.4T    /srv/ivdfs_03/LEAP-BKUP/20110914_EFF_1330
495G    /srv/ivdfs_03/LEAP-BKUP/20120129_EFF_1390
684G    /srv/ivdfs_03/LEAP-BKUP/20111006_EFF_1330
937G    /srv/ivdfs_03/LEAP-BKUP/20120107_EFF_1390
683G    /srv/ivdfs_03/LEAP-BKUP/20120108_EFF_1330
750G    /srv/ivdfs_03/LEAP-BKUP/20111002_EFF_1350
1.6T    /srv/ivdfs_03/LEAP-BKUP/20120130_EFF_1350
1.6T    /srv/ivdfs_03/LEAP-BKUP/20120130_EFF_1330
1.6T    /srv/ivdfs_03/LEAP-BKUP/20120130_EFF_1310
1.6T    /srv/ivdfs_03/LEAP-BKUP/20120130_EFF_1370
```

SOME VIEWS ON COMPUTING PARADIGMS

- RISC vs CISC computing

SOME VIEWS ON COMPUTING PARADIGMS

- RISC vs CISC computing

SOME USEFUL STUFF

- Dealing with small files - squashfs/tar
- Containers
- MPG's ChatGPT Bot: https://minervamessenger.mpdl.mpg.de
- VNCs and screen - persistent sessions
- Globus –> sharing data (up to 50TB)
- Keeper (1TB): https://keeper.mpdl.mpg.de/
- Edmond (1TB): https://edmond.mpg.de/
- GWDG (50GB): https://owncloud.gwdg.de/

SQUASHFS AND CO.

```
paf0:~> find /beegfsEDD/FUNDI/PULSAR_MIGRATION/titan0/MAGNETAR/1818-1607/20200314 | wc -l
5530
paf0:~> du -B MB /beegfsEDD/FUNDI/PULSAR_MIGRATION/titan0/MAGNETAR/1818-1607/20200314
6578MB  /beegfsEDD/FUNDI/PULSAR_MIGRATION/titan0/MAGNETAR/1818-1607/20200314
paf0:~> mksquashfs /beegfsEDD/FUNDI/PULSAR_MIGRATION/titan0/MAGNETAR/1818-1607/20200314 /beegfsEDD/
FUNDI/PULSAR_MIGRATION/titan0/MAGNETAR/1818-1607/20200314.sqfs
Parallel mksquashfs: Using 12 processors
Creating 4.0 filesystem on /beegfsEDD/FUNDI/PULSAR_MIGRATION/titan0/MAGNETAR/1818-1607/20200314.sqfs, block
size 131072.
[========|                                                                    ] 3057/54799   5%
```

```
paf0:~>mkdir chk
paf0:~> squashfuse /beegfsEDD/PULSAR_MIGRATION/titan0/MAGNETAR/1818-1607/20200314.sqfs ./chk
paf0:~> cd chk
paf0:~/chk> find . | wc -l
5530
paf0:~/chk> ls
single  SPpolnChk
paf0:~/chk> cd single/
paf0:~/chk/single> ls -l | head -5
-rw-r--r-- 1 pulsar psr  90580387 Jan 29  2021 add.zz
-rw-r--r-- 1 pulsar psr 106564387 Jan 29  2021 add.zz.pazi
-rw-r--r-- 1 pulsar psr     10509 Jan 29  2021 list.txt
-rw-r--r-- 1 pulsar psr   1068902 Mar 26  2020 pulse_0.ar
paf0:~/chk/single> cd
paf0:~> fusermount -u ./chk
paf0:~>
```

SCREEN – PERSISTENT SESSIONS

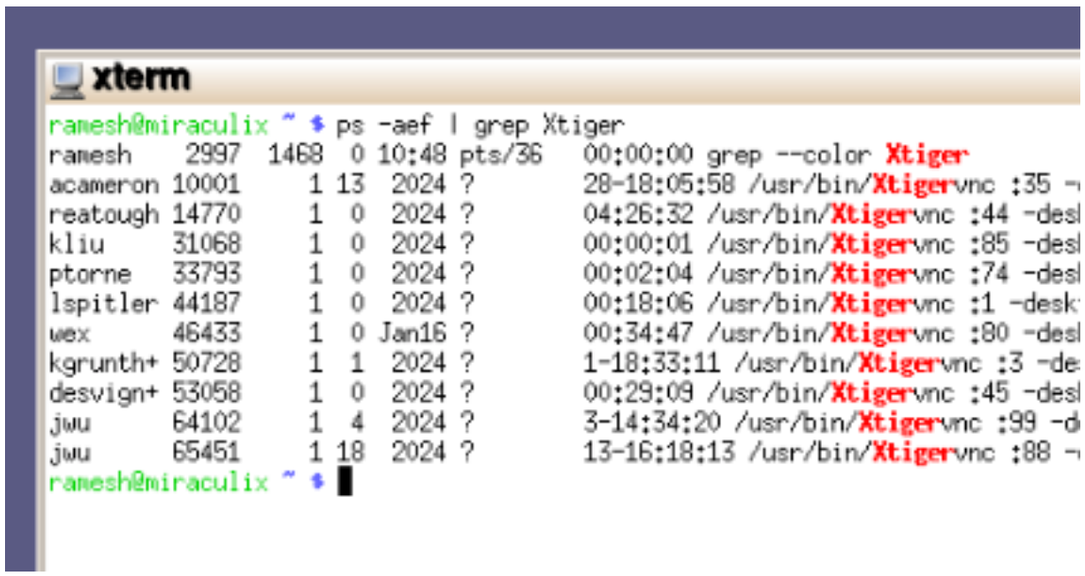
- Create a session: screen -S <session_name>
- Detach from a session: Ctrl-a d
- List running session : screen -ls
- Reattach to a running session: screen -r <session name>

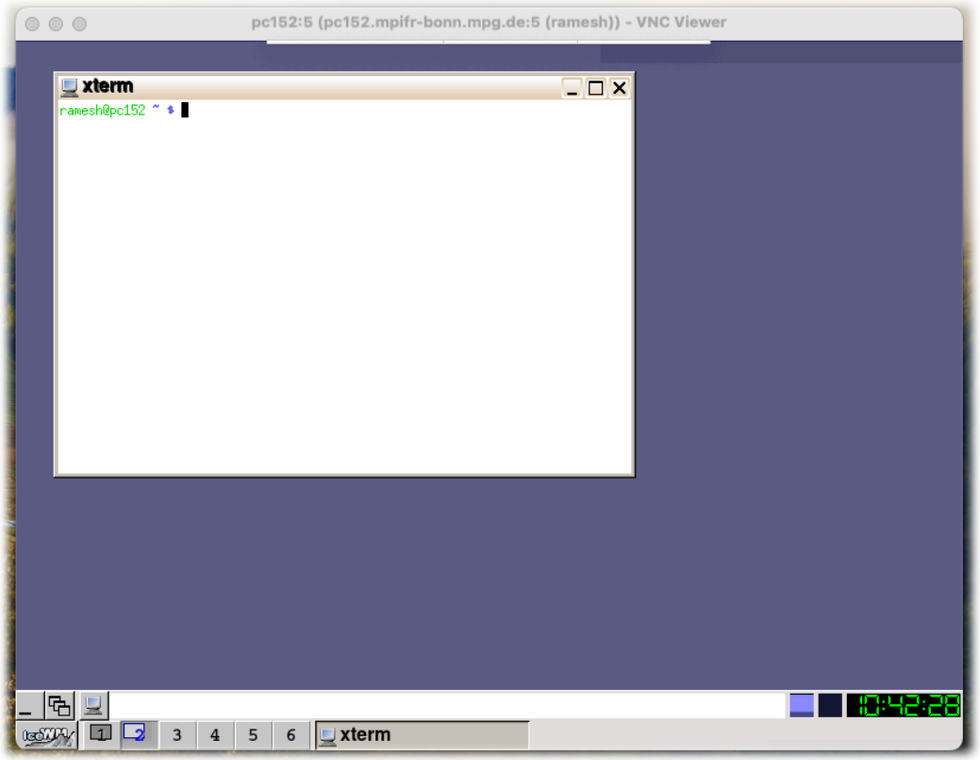
```
ramesh@pc152 ~ $ screen -ls
No Sockets found in /run/screen/S-ramesh.
ramesh@pc152 ~ $ screen -S testscr
ramesh@pc152 ~ $
ramesh@pc152 ~ $ screen -ls
There is a screen on:
        1896313.testscr (02/10/2025 10:22:03 AM)        (Attached)
1 Socket in /run/screen/S-ramesh.
<Ctrl-a d> <–- detach from the current screen session
```

VNC – PERSISTENT SESSIONS

- Create a session: vncserver
- Detach from a session: hit the **X** button
- Reattach to a running session: vncviewer host:<display-number>
- Purge a running session: vncserver -kill :<display_number>
- For hercules, see: https://docs.mpcdf.mpg.de/doc/computing/software/vnc.html

```
ramesh@pc152 ~ $ vncserver -geometry 800x600 -depth 24 :5
New Xtigervnc server 'pc152.mpifr-bonn.mpg.de:5 (ramesh)' on port 5905 for display :5. .....
ramesh@pc152 ~ $ vncserver -kill :5
ramesh@pc152 ~ $ cat .vnc/xstartup
#!/bin/sh
xrdb $HOME/.Xresources
xsetroot -solid grey
xterm -geometry 80x24+10+10 -ls -title "$VNCDESKTOP Desktop" &
vncconfig --nowin&
/usr/bin/icewm-session
```





QUESTIONS/HANDS-ON?