Using Grid5000

Himadri CS, Whisper

About me

- Studied Machine Intelligence in India
- Discovered passion for Operating Systems via an Outreachy internship
- Joined Whisper as an engineer to work on Frequency aware scheduling policy for the Linux Kernel
- Fell in love with the research life so now pursuing a PhD!

https://himadripandya.me/

Challenges for my experiments

- 5 different kinds of servers and one desktop machine
- Compile and install kernels with new scheduling policies on all of them
- Run 200+ phoronix benchmarks with all kernels on all machines
- Make traces for each kernel on every machine and analyze them
- Need days of compute time on each machine
- Generate terabytes of traces and results data

About Grid5000

- Grid'5000 is a large-scale and flexible testbed for experiment-driven research in all areas of computer science, with a focus on parallel and distributed computing including Cloud, HPC and Big Data and AI.
- 15000 cores, 800 compute-nodes grouped in homogeneous clusters, and featuring various technologies: PMEM, GPU, SSD, NVMe, 10G and 25G Ethernet, Infiniband, Omni-Path
- One can make custom OS images and play with them
- Tricky usage policy

https://www.grid5000.fr/w/Grid5000:Home

Grid5000 Usage Policy

- Time of your reservation matters:
  - During weekdays - Your jobs can start after 09:00 AM and must finish by 07:00 PM
  - During nights - Your jobs can start at 7:00 PM and must finish by 9:00 AM
  - During weekends - Your jobs can start at 7:00 PM on Friday and must finish by 9:00 AM on Monday

- Type of resources also matters:

  During weekdays - You should not use more than the equivalent of 2 hours on all the cores of the cluster during a given day

  ➔ Grenoble site has 4 Yeti nodes with 4 Intel Xeon Gold sockets with 16 cores per cpu
  So, Allowed usage = 2h * 4n * 4s * 16c = 512 units
  1. If you reserve 1 yeti for 8 hours then your usage = 8h * 1n* 4s * 16c = 512 units ✅
  2. If you reserve all 4 yetis for 2 hours each then your usage = 4h * 2n * 4s * 16c = 512 units ✅
  3. If you reserve 1 yeti for 6 hours and another yeti for 4 hours then
     your usage = (6h * 1n * 4s * 16c) + (4h * 1n * 4s * 16c) = 384 + 256 = 640 > 512 ❌

  During nights and weekends - You can reserve as many resources as you want and as many of them are available

https://www.grid5000.fr/w/Grid5000:UsagePolicy

Advance reservations & Changes in reservations

- People working at INRIA belong to the Silver privilege level:
  - We can make 2 advance reservations per site
  - We can submit a request for advance reservation maximum 48 hours before the reservation starts

- You can delete a reservation if needed

- Exceptions in usage policy:
  - One can extend a reservation 10 minutes before it ends, and for up to one hour
  - Night reservation can start early at 5:00 PM in the evening
  - Day reservation can start early at 12:00 AM at night

- You cannot shrink a reservation.

https://www.grid5000.fr/w/Grid5000:UsagePolicy#Special_cases

Things to keep in mind

- /home is hosted on frontend using NFS, hence it is slow
- Default homedir size on every site = 25GB, but you can request extension
- /home not synchronized across different sites, so you need a seperate copy of your OS images on each site
- Any data that is not on /home is lost when your node reservation expires
- One can reserve extra disks for more storage, request group storage
- Multiple people competing for same machines as conference deadlines approach

Reserving a node

- You can start playing with dahus (2 x Intel Xeon Gold 6130, 16 cores/CPU, 192 GiB) in the Grenoble cluster
  - Get any 2 dahus for 2 hours with the default OS image (debian11-x64-nfs):
    `oarsub -p "cluster='dahu'" -l host=2,walltime=2 -I`
  - Get a specific node (dahu-12) for 2 hours
    `oarsub -p "host='dahu-12.grenoble.grid5000.fr'" -l host=1,walltime=2 -I`

Making and deploying a custom OS image

- Reserve a specific node with the default environment
  `fgrenoble: oarsub -p "host='dahu-12.grenoble.grid5000.fr'" -l host=1,walltime=2 -I`
- Customize the environment on the node
  `dahu-12: apt update, apt upgrade, apt install …`
- Create the custom image from the node on the frontend
  `fgrenoble: tgz-g5k -m dahu-12 -f ~/custom_image.tgz`
- Make a new environment specification file for the custom image on the frontend
  `fgrenoble: kaenv3 -p debian11-x64-base -u hpandya > custom-environment.yaml`
- Add the new environment to list of available environments
  `fgrenoble: kaenv3 -a custom-environment.yaml`
- Deploy the new environment on the node
  `fgrenoble: kadeploy3 -m dahu-12 -e custom-environment`

Search for "environment creation" in the documentation
https://www.grid5000.fr/w/Advanced_Kadeploy#How_can_I_make_my_own_environment.3F

Deleting a reservation

- Use the command *oardel jobid:*
  - e.g.
    `fgrenoble:~$ oardel 2081205`

Extending a reservation

- Use the command *oarwalltime jobid +number_of_hours*:
  - e.g.
    ```
    fgrenoble:~$ oarwalltime 2081207 +1
    ```

Making an advance reservation

- You can request for multiple nodes in the same advance reservation request:
  - e.g. Reserve troll-4 and yeti-4 in the grenoble site for the weekend:
    ```
    fgrenoble:~$ oarsub -t exotic -t deploy -r "2022-02-11 19:00:00" -l
    {"host in ('troll-4.grenoble.grid5000.fr',
    'yeti-4.grenoble.grid5000.fr')"}/nodes=2,walltime=61:55:00
    ```
- At any moment, you can have 2 such reservations per site
- When the reservation starts, you can deploy your custom image on each node
- All nodes requested in an advance reservation appear under the same jobid, so you cannot delete just one of them

A few other things…

- How to check platform status - https://www.grid5000.fr/w/Status
- How to check if your usage is not violating the rules
  `usagepolicycheck -t`
- Scolding emails from the support staff on violation of usage policy
- g5k-users email group (users@lists.grid5000.fr)
- Responsive support staff (support-staff@lists.grid5000.fr)
- A more detailed presentation
  https://www.grid5000.fr/mediawiki/images/Grid5000.pdf