Cluster Job Management with PBS

Andria

andria.arisal@lipi.go.id

Why?

- Resources are scarce
- Users want to have them all
- Administrator wants optimal resource utilization

- So... jobs should be managed

Outlines

- PBS Introduction
- Installation
- For Administrator
- For User

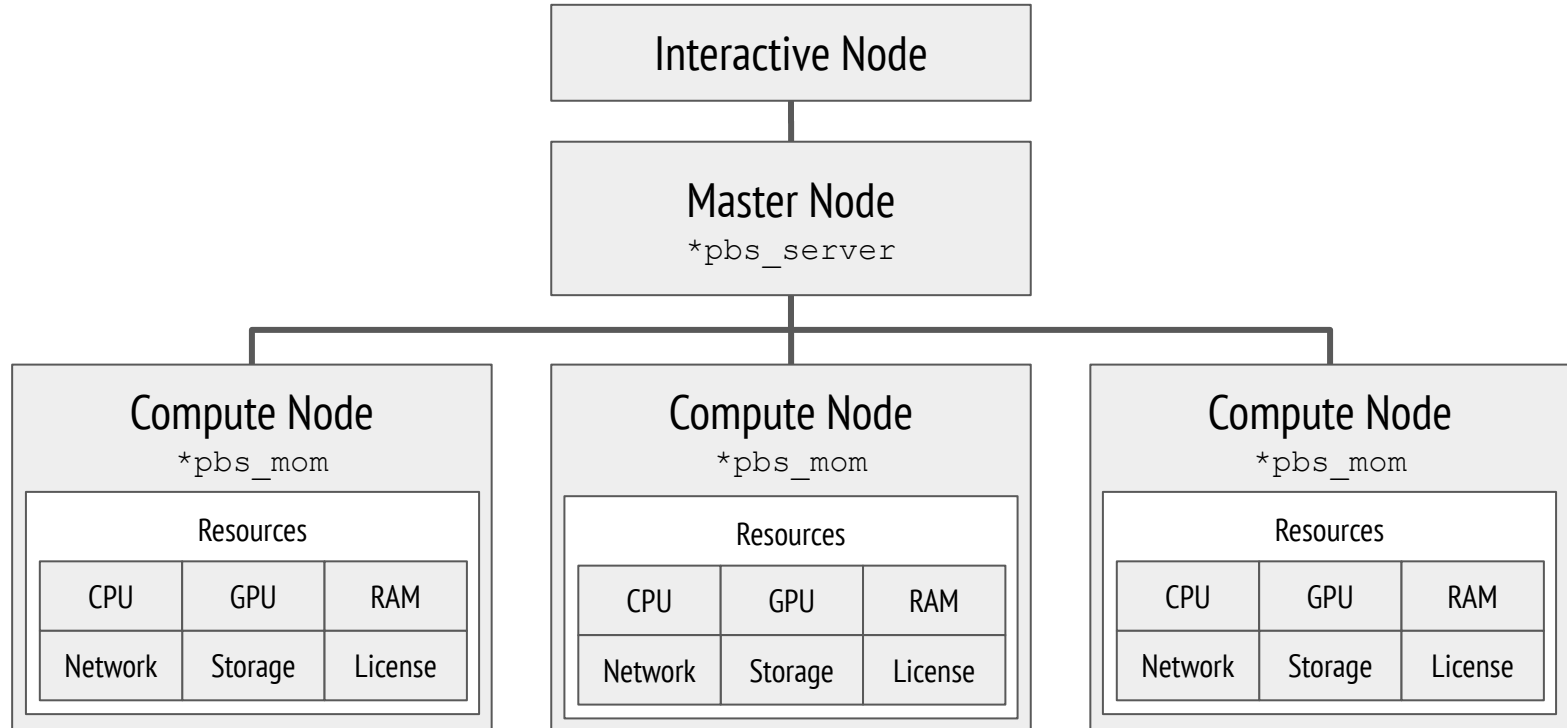# PBS Introduction

Portable Batch System

- Software for jobs scheduling
- Various fork
  - OpenPBS (no longer developed)
  - **TORQUE** (maintained by Adaptive Computing)
  - PBS Pro (commercially licensed by Altair Engineering)

TORQUE

- Terascale Open-source Resource and QUEue Manager
- Resource manager for managing batch jobs in distributed compute nodes

Concept

Concept

Creation	Submission	Execution	Finalization

`$ vi job.pbs`
 	
`#PBS -N myJob`
`#PBS -S /bin/sh`
`#PBS -l`
`nodes=1:ppn=2,walltime=240:00:0`
`0`
`#PBS -M user@lapan.go.id`
`#PBS -m ea`
`source ~/.bashrc`
`cd $HOME/work/dir`
`sh myBlast.sh -i -v`
	`$ qsub job.pbs`	`$ qstat`	`$ ls *.out`
			`$ ls *.err`

MUNGE

- MUNGE (MUNGE Uid 'N' Gid Emporium)
- Authentication service for creating and validating credentials
- Security contract between users on the cluster
- Used by TORQUE

Installation

0. Infrastructure

- Servers are installed with Operating System (CentOS 7)
- Network, firewall, passwordless-ssh, NFS, ... are configured

- **[root@* ~]# cat /etc/hosts**
  ```
  127.0.0.1        localhost.localdomain        localhost
  ::1              localhost.localdomain        localhost
  10.100.200.5     submitter.grid.lapan.go.id   submitter
  10.100.200.10    master.grid.lapan.go.id      master
  10.100.200.11    worker01.grid.lapan.go.id    worker01
  10.100.200.12    worker02.grid.lapan.go.id    worker02
  10.100.200.13    worker03.grid.lapan.go.id    worker03
  ```

LIPI

0. Infrastructure

```
                    ┌─────────────────────────────┐
                    │  submitter.grid.lapan.go.id │
                    │        10.100.200.5         │
                    └─────────────┬───────────────┘
                                  │
                    ┌─────────────┴───────────────┐
                    │   master.grid.lapan.go.id   │
                    │        10.100.200.10        │
                    └─────────────┬───────────────┘
            ┌─────────────────────┼─────────────────────┐
┌───────────────────────┐ ┌───────────────────────┐ ┌───────────────────────┐
│ worker01.grid.lapan.go.id │ │ worker02.grid.lapan.go.id │ │ worker03.grid.lapan.go.id │
│     10.100.200.11     │ │     10.100.200.12     │ │     10.100.200.13     │
└───────────────────────┘ └───────────────────────┘ └───────────────────────┘
```
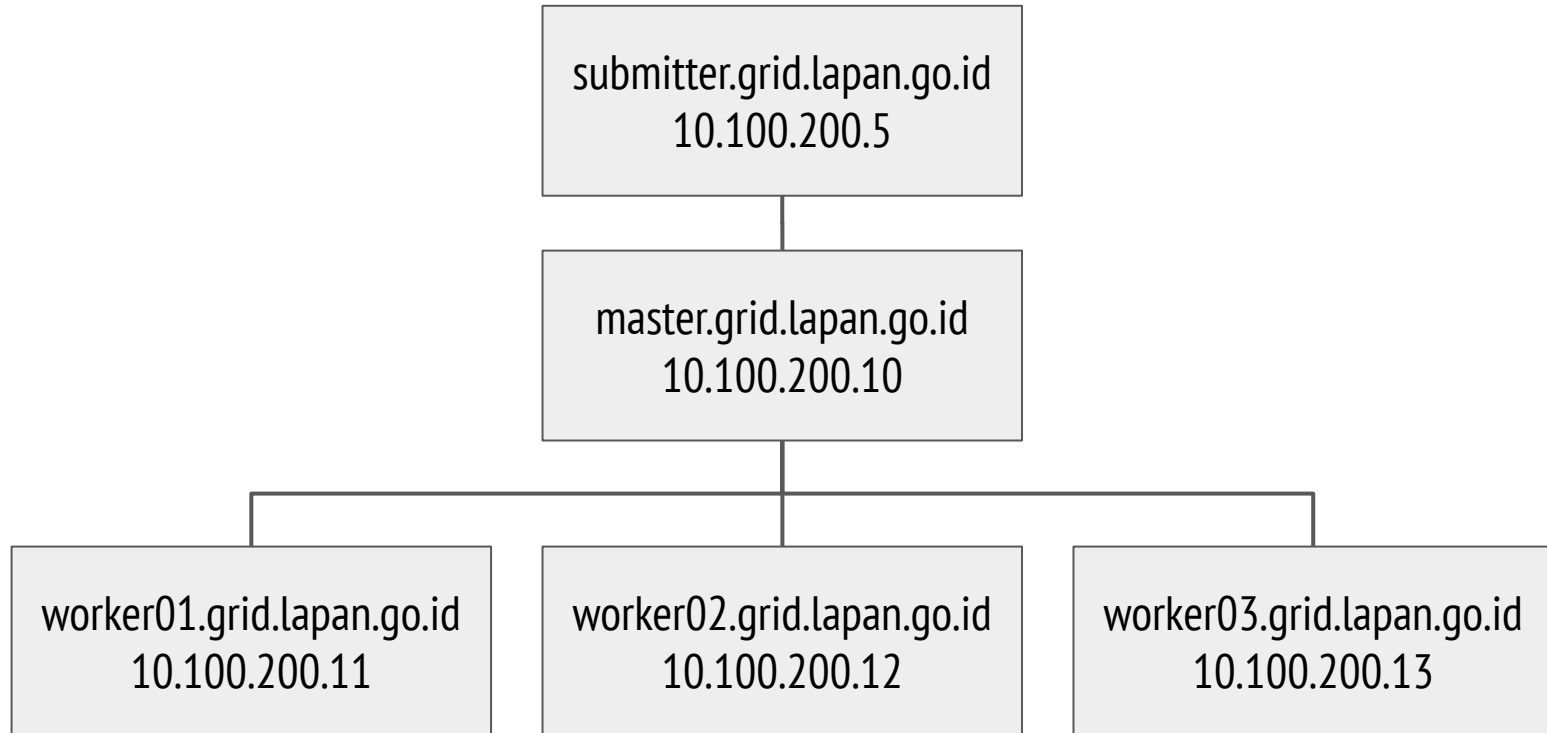
1. PBS Master (installation)

- **[root@master ~]# yum install torque-server torque-scheduler**

```
...
Installed:
  torque-scheduler.x86_64 0:4.2.10-10.el7
  torque-server.x86_64 0:4.2.10-10.el7
Dependency Installed:
  munge.x86_64 0:0.5.11-3.el7
  munge-libs.x86_64 0:0.5.11-3.el7
  torque.x86_64 0:4.2.10-10.el7
  torque-libs.x86_64 0:4.2.10-10.el7

Complete!
```

1. PBS Master (MUNGE configuration and service)

- **[root@master ~]# /usr/sbin/create-munge-key**
  Generating a pseudo-random key using /dev/urandom completed.
- **[root@master ~]# systemctl start munge**
- **[root@master ~]# systemctl enable munge**
- **[root@master ~]# systemctl status munge**
- munge.service - MUNGE authentication service
    Loaded: loaded (/usr/lib/systemd/system/munge.service; enabled; vendor preset: disabled)
    Active: active (running) since Mon 2019-01-28 08:55:33 WIB; 3s ago
      Docs: man:munged(8)
   Process: 5959 ExecStart=/usr/sbin/munged (code=exited, status=0/SUCCESS)
  Main PID: 5961 (munged)
    CGroup: /system.slice/munge.service
            └─5961 /usr/sbin/munged

1. PBS Master (TORQUE configuration)

- **`[root@master ~]# cat /var/lib/torque/server_name`**
  `master`
- **`[root@master ~]# cat /var/lib/torque/server_priv/nodes`**
  `worker01.grid.lapan.go.id np=28 general`

1. PBS Master (TORQUE authentication service)

- **[root@master ~]# systemctl start trqauthd**
- **[root@master ~]# systemctl enable trqauthd**
- **[root@master ~]# systemctl status trqauthd**
  - trqauthd.service - trqauthd
    Loaded: loaded (/usr/lib/systemd/system/trqauthd.service; enabled; vendor preset: disabled)
    Active: active (running) since Mon 2019-01-28 08:39:49 WIB; 17min ago
   Process: 5398 ExecStart=/usr/sbin/trqauthd (code=exited, status=0/SUCCESS)
  Main PID: 5399 (trqauthd)
    CGroup: /system.slice/trqauthd.service
            └─5399 /usr/sbin/trqauthd

1. PBS Master (PBS service)

- **[root@master ~]# systemctl start pbs_server**
- **[root@master ~]# systemctl enable pbs_server**
- **[root@master ~]# systemctl status pbs_server**
  - pbs_server.service - pbs-server
    Loaded: loaded (/usr/lib/systemd/system/pbs_server.service; enabled; vendor preset: disabled)
    Active: active (running) since Mon 2019-01-28 10:51:15 WIB; 1min 29s ago
    Process: 2030 ExecStart=/usr/sbin/pbs_server (code=exited, status=0/SUCCESS)
    Main PID: 2032 (pbs_server)
    CGroup: /system.slice/pbs_server.service
            └─2032 /usr/sbin/pbs_server

2. PBS Worker (installation)

- **[root@worker01 ~]# yum install torque-mom torque-client**

```
Installed:
  torque-client.x86_64 0:4.2.10-10.el7
  torque-mom.x86_64 0:4.2.10-10.el7

Dependency Installed:
  hwloc-libs.x86_64 0:1.11.8-4.el7
  munge.x86_64 0:0.5.11-3.el7
  munge-libs.x86_64 0:0.5.11-3.el7
  torque.x86_64 0:4.2.10-10.el7
  torque-libs.x86_64 0:4.2.10-10.el7

Complete!
```

2. PBS Worker (configuration)

- `[root@worker01 ~]# cat /var/lib/torque/server_name`
  master.grid.lapan.go.id
- `[root@worker01 ~]# cat /var/lib/torque/mom_priv/config`
  # Configuration for pbs_mom.
  $pbsserver master.grid.lapan.go.id
  $logevent 0x1ff

- `[root@master ~]# scp /etc/munge/munge.key worker01:/etc/munge/munge.key`
- `[root@worker01 ~]# chmod 400 /etc/munge/munge.key`

2. PBS Worker (MUNGE configuration and service)

- **[root@worker01 ~]# systemctl start munge**
- **[root@worker01 ~]# systemctl enable munge**
- **[root@worker01 ~]# systemctl status munge**
- munge.service - MUNGE authentication service
    Loaded: loaded (/usr/lib/systemd/system/munge.service; enabled; vendor preset: disabled)
    Active: active (running) since Mon 2019-01-28 08:55:33 WIB; 3s ago
      Docs: man:munged(8)
   Process: 5959 ExecStart=/usr/sbin/munged (code=exited, status=0/SUCCESS)
  Main PID: 5961 (munged)
    CGroup: /system.slice/munge.service
            └─5961 /usr/sbin/munged

2. PBS Worker (TORQUE authentication service)

- **[root@worker01 ~]# systemctl start trqauthd**
- **[root@worker01 ~]# systemctl enable trqauthd**
- **[root@worker01 ~]# systemctl status trqauthd**
  - trqauthd.service - trqauthd
    Loaded: loaded (/usr/lib/systemd/system/trqauthd.service; enabled; vendor preset: disabled)
    Active: active (running) since Mon 2019-01-28 08:39:49 WIB; 17min ago
   Process: 5398 ExecStart=/usr/sbin/trqauthd (code=exited, status=0/SUCCESS)
  Main PID: 5399 (trqauthd)
    CGroup: /system.slice/trqauthd.service
            └─5399 /usr/sbin/trqauthd

2. PBS Worker (PBS service)

- **[root@master ~]# systemctl start pbs_mom**
- **[root@master ~]# systemctl enable pbs_mom**
- **[root@master ~]# systemctl status pbs_mom**
  - pbs_mom.service - pbs-mom
     Loaded: loaded (/usr/lib/systemd/system/pbs_mom.service; enabled; vendor preset: disabled)
     Active: active (running) since Mon 2019-01-28 10:35:21 WIB; 14min ago
    Process: 2240 ExecStart=/usr/sbin/pbs_mom (code=exited, status=0/SUCCESS)
   Main PID: 2241 (pbs_mom)
     CGroup: /system.slice/pbs_mom.service
             └─2241 /usr/sbin/pbs_mom

3. PBS Submitter (installation)

- **[root@submitter ~]# yum install torque-client**

```
Installed:
  torque-client.x86_64 0:4.2.10-10.el7

Dependency Installed:
  hwloc-libs.x86_64 0:1.11.8-4.el7
  munge.x86_64 0:0.5.11-3.el7
  munge-libs.x86_64 0:0.5.11-3.el7
  torque.x86_64 0:4.2.10-10.el7
  torque-libs.x86_64 0:4.2.10-10.el7

Complete!
```

3. PBS Submitter (configuration)

- `[root@rumah ~]# cat /var/lib/torque/server_name`
  `master.grid.lapan.go.id`

- `[root@master ~]# scp /etc/munge/munge.key submitter:/etc/munge/munge.key`
- `[root@submitter ~]# chmod 400 /etc/munge/munge.key`

3. PBS Submitter (MUNGE configuration and service)

- **[root@submitter ~]# systemctl start munge**
- **[root@submitter ~]# systemctl enable munge**
- **[root@submitter ~]# systemctl status munge**
- munge.service - MUNGE authentication service
    Loaded: loaded (/usr/lib/systemd/system/munge.service; enabled;
vendor preset: disabled)
    Active: active (running) since Mon 2019-01-28 08:55:33 WIB; 3s ago
      Docs: man:munged(8)
   Process: 5959 ExecStart=/usr/sbin/munged (code=exited,
status=0/SUCCESS)
  Main PID: 5961 (munged)
    CGroup: /system.slice/munge.service
            └─5961 /usr/sbin/munged

3. PBS Submitter (TORQUE authentication service)

- **[root@submitter ~]# systemctl start trqauthd**
- **[root@submitter ~]# systemctl enable trqauthd**
- **[root@submitter ~]# systemctl status trqauthd**
- trqauthd.service - trqauthd
  Loaded: loaded (/usr/lib/systemd/system/trqauthd.service; enabled; vendor preset: disabled)
  Active: active (running) since Mon 2019-01-28 08:39:49 WIB; 17min ago
  Process: 5398 ExecStart=/usr/sbin/trqauthd (code=exited, status=0/SUCCESS)
  Main PID: 5399 (trqauthd)
  CGroup: /system.slice/trqauthd.service
          └─5399 /usr/sbin/trqauthd

for Administrator

Creating Queue (queue definition)

- **[root@master ~]# vi qmgr.input**

  ```
  #
  # Create queues and set their attributes.
  #
  #
  # Create and define queue admin
  #
  create queue admin
  set queue admin queue_type = Execution
  set queue admin enabled = True
  set queue admin started = True
  ```

Creating Queue (queue definition)

- **[root@master ~]# vi qmgr.input**

```
...
create queue public
set queue public queue_type = Execution
set queue public max_user_queuable = 4
set queue public resources_max.nodes = 16
set queue public resources_max.cput = 48:00:00
set queue public resources_max.walltime = 72:00:00
set queue public resources_default.neednodes = general
set queue public resources_default.nodes = 4
set queue public resources_default.walltime = 24:00:00
set queue public max_user_run = 2
set queue public enabled = True
set queue public started = True
```

Creating Queue (server attributes definition)

- **[root@master ~]# vi qmgr.input**

  ```
  ...
  set server scheduling = True
  set server acl_host_enable = False
  set server acl_hosts = master.grid.lapan.go.id
  set server acl_hosts += master
  set server managers = grid@mail.lapan.go.id
  set server managers += root@master.grid.lapan.go.id
  set server default_queue = default
  set server log_events = 511
  set server mail_from = adm
  set server scheduler_iteration = 600
  set server node_check_rate = 150
  set server tcp_timeout = 6
  ```

LIPI

Creating Queue (server attributes definition)

- **[root@master ~]# vi qmgr.input**

  ```
  ...
  set server default_node = worker
  set server node_pack = False
  set server job_stat_rate = 45
  set server poll_jobs = True
  set server mom_job_sync = True
  set server mail_domain = never
  set server kill_delay = 10
  set server submit_hosts = submitter.grid.lapan.go.id
  set server allow_node_submit = True
  set server next_job_number = 7739
  set server moab_array_compatible = True
  set server nppcu = 1
  ```

Creating Queue (queue definition loading)

- `[root@master ~]# qmgr < qmgr.input`

Monitoring Queue

- **`[root@master ~]# qstat -f -Q`**

  ```
  ...
  Queue: public
      queue_type = Execution
      max_user_queuable = 4
      total_jobs = 0
      state_count = Transit:0 Queued:0 Held:0 Waiting:0 Running:0 Exiting:0 Complete:0
      resources_max.cput = 48:00:00
      resources_max.nodes = 16
      resources_max.walltime = 72:00:00
      resources_default.neednodes = worker
      resources_default.nodes = 4
      resources_default.walltime = 24:00:00
      mtime = 1548983368
      resources_assigned.nodect = 0
      max_user_run = 2
      enabled = True
      started = True
  ```

Controlling & Monitoring PBS

- `[root@master ~]# pbs_server [options]`
- `[root@master ~]# qterm`
- `[root@master ~]# qmgr`

- `[root@master ~]# pbsnode [options]`

Controlling Queue

- `[root@master ~]# qstat`
- `[root@master ~]# qhold`
- `[root@master ~]# qrls`
- `[root@master ~]# qalter`
- `[root@master ~]# qdel`

- `[root@master ~]# qstart`
- `[root@master ~]# qstop`
- `[root@master ~]# qrun`
- `[root@master ~]# qselect`
- `[root@master ~]# qorder`

for User

PBS Job script

- **[root@submitter ~]# vi myJob.pbs**

  ```
  #!/bin/bash

  #PBS -N myJob
  #PBS -q public
  #PBS -l nodes=1
  #PBS -l walltime=1:00:00
  #PBS -k oe

  date
  hostname
  sleep 30
  date
  ```

Managing Job

- `[root@master ~]# qstat`
- `[root@master ~]# qhold`
- `[root@master ~]# qrls`
- `[root@master ~]# qalter`
- `[root@master ~]# qdel`