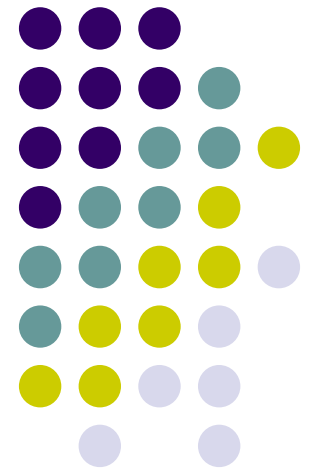


cluster
search

um ano de administração



Período experimental



- Ausência de políticas
- Todos utilizadores com os mesmos “direitos”
- Trabalhos longos, mais de 2 meses
- Elevada taxa de ocupação
- Dificuldade em reservar recursos

Período proto-produção



- Sistema de gestão
 - Maui escalonador
 - Torque gestor de recursos
- Definição de políticas de utilização

Maui: Fairshare



- Definição de grupos de utilizadores
- Quotas de utilização por grupos
- Limitação de acesso a recursos
- Prioridade com base na utilização histórica

Maui: Backfill



- Otimização de recursos disponíveis
- Escalonamento de trabalhos com menor prioridade
- Garantida a execução de trabalhos com maior prioridade na sua ordem

Maui: Reservas



- Utilizador agenda sessão de trabalho
- Administrador reserva recursos para utilizador

Maui – Config (1)



Fairshare use of the cluster requires a system level definition of a time slot, which can be defined as 1 week.

FSPOLICY	DEDICATEDPS
FSINTERVAL	24:00:00
FSDEPTH	7
FSDECAY	1
FSWEIGHT	1
FSUSERWEIGHT	1
QOSWEIGHT	1
XFACTORWEIGHT	1
QUEUEWEIGHT	1
TARGETQUEUEWEIGHT	1

Maui – Config (2)



Groups that have been defined (the list of users in each group is defined at Unix level)

GROUPCFG[impe]	FSTARGET=25+	MAXPS=9000000,28800000	QTTARGET=288	MAXIJOB=25
GROUPCFG[jade]	FSTARGET=12+	MAXPS=4320000,28800000	QTTARGET=600	MAXIJOB=12
GROUPCFG[amalandro]	FSTARGET=8+	MAXPS=2880000,28800000	QTTARGET=900	MAXIJOB=8
GROUPCFG[biocnat]	FSTARGET=8+	MAXPS=2880000,28800000	QTTARGET=900	MAXIJOB=8
GROUPCFG[metform]	FSTARGET=4+	MAXPS=1440000,28800000	QTTARGET=1800	MAXIJOB=4
GROUPCFG[sig]	FSTARGET=3+	MAXPS=1080000,28800000	QTTARGET=2400	MAXIJOB=3
GROUPCFG[neuroci]	FSTARGET=2+	MAXPS=720000,28800000	QTTARGET=3600	MAXIJOB=2
GROUPCFG[fisac]	FSTARGET=12+	MAXPS=4320000,28800000	QTTARGET=600	MAXIJOB=12
GROUPCFG[fiscris]	FSTARGET=10+	MAXPS=3600000,28800000	QTTARGET=720	MAXIJOB=10
GROUPCFG[fismr]	FSTARGET=8+	MAXPS=2880000,28800000	QTTARGET=900	MAXIJOB=8
GROUPCFG[anatum]	FSTARGET=8+	MAXPS=2880000,28800000	QTTARGET=900	MAXIJOB=8
GROUPCFG[outros]	FSTARGET=1+	MAXPS=360000		MAXIJOB=1

Maui – Config (3)



```
#
# ***** Backfill Policies *****

BACKFILLPOLICY  BESTFIT
BACKFILLMETRIC  PROCS
RESERVATIONPOLICY  CURRENTHIGHEST
BYPASSWEIGHT    1000

#
# ***** QOS definitions *****

QOSCFG[urgente]      PRIORITY=1000    QFLAGS=PREEMPTOR    MAXPS=8640000
QOSCFG[nao_urgente]  PRIORITY=0       QFLAGS=PREEMPTTEE
```



Torque: Filas

- Selecção de recursos baseada em propriedades
 - gen: nós com bi-processador
 - gpu: nós com bi-processador dual-core e placa gráfica
 - io: nós com FC à SAN



Torque: Filas

- Encaminhamento de trabalhos baseado em walltime
 - tlongos: $wt > 20d$
 - tmedios: $20h < wt < 20d$
 - tcurtos: $30m < wt < 20h$
 - tinteractivos: $wt < 30m$
- Ajuste periódico de recursos associados às filas

Torque – Config



```
# Create and define queue default
#
create queue default
set queue default queue_type = Route
set queue default route_destinations = erro
set queue default route_destinations += ttestes
set queue default route_destinations += tcurtos
set queue default route_destinations += tmedios
set queue default route_destinations += tlongos
set queue default enabled = True
set queue default started = True
#
# Create and define queue tcurtos
#
create queue tcurtos
set queue tcurtos queue_type = Execution
set queue tcurtos Priority = 100
set queue tcurtos max_running = 80
set queue tcurtos resources_max.walltime = 20:00:00
set queue tcurtos resources_min.walltime = 00:30:01
set queue tcurtos resources_default.walltime = 20:00:00
set queue tcurtos enabled = True
set queue tcurtos started = True
```

Vantagens



- Controlo na utilização de recursos
- Redução de trabalhos muito longos
- Disponibilidade de recursos em ciclos de 20h



Problemas

- Suspensão de trabalhos
- Compatibilizar utilização heterogénea
 - Instalação de software: NAMD, FFTW, VMD, Intel MKL, MPICH, etc
 - Partilha de recursos entre trabalhos
- Sistema de ficheiros NFS
 - Degradação de desempenho com muito I/O
 - Instabilidade de funcionamento do cluster
- Cortes de energia eléctrica



Período produção

- Instalação de UPS de 20 KVA
- Redundância de nó frontal
- Utilização da SAN
 - GPFS
- Ligação à Grid EGEE ?