

陽明生資所 Bio-cluster 種子訓練課程

Torque 基本操作

鄧偉華

wade@nchc.org.tw



財團法人國家實驗研究院

國家高速網路與計算中心

National Center for High-Performance Computing

Outline

- PBS 介紹
- PBS 常用指令
- PBS SCRIPT 撰寫
- 陽明生物計算軟體測試 SCRIPT 撰寫
 - MPICH
 - BLAST
 - SIM4
 - Clustal
 - Phylip
 - Velvet
 - Bowtie
 - SOAP

What is PBS

- 主要用在 **Job scheduling** 及資源管理，主要的功能為管理及分配計算 **job** 及其所需要的資源。
- 參考：<http://www.openpbs.org>

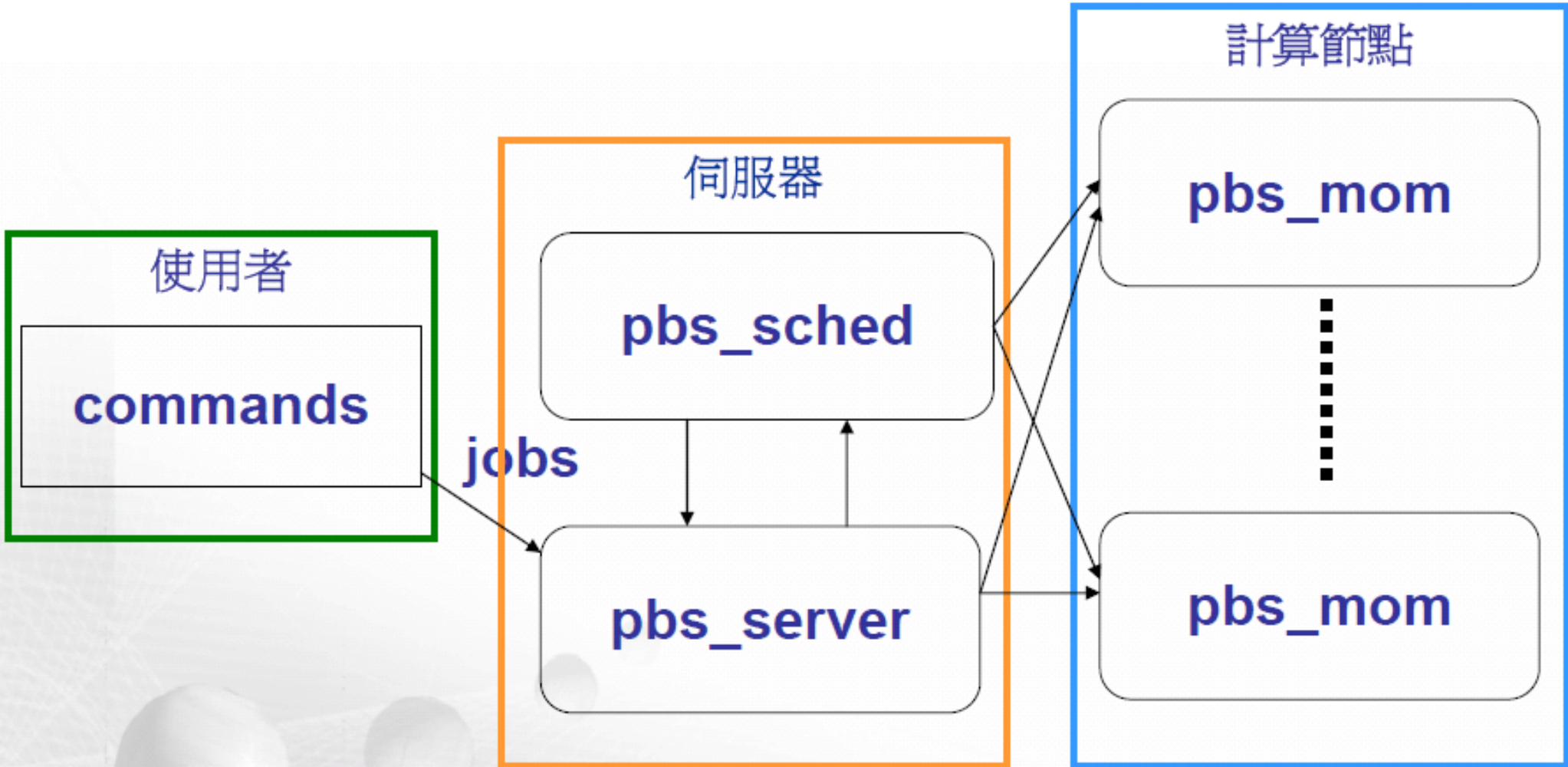
What is Torque

- Torque 是以 open PBS 為基礎建立的一套實作方法。
- 參考：
<http://www.globusconsortium.org/tutorial/ch2>

Why is Torque

- Cluster 內的 nodes 數太多。
- Cluster 內的 users 數太多。
- Cluster 內的 jobs 數太多。
- 使資源充份使用。
- 使資源公平使用。
- 管理資源：
 - 使用者：磁碟空間、執行時間、優先權、權限。
 - Cluster：切割資源、node 狀態、資源監控。
 - job：排程控制、負載平衡、錯誤恢復、狀態通知。

PBS 架構



PBS 常用指令

- 送出 JOB
- 查詢 JOB 狀態
- 管理 JOB
- 移除 JOB

查詢 nodes 是否正常啓動

- `pbsnodes -a`


```
wade@bio037:~/qsub$ pbsnodes
```

```
bio013
```

```
state = free
```

```
np = 1
```

```
ntype = cluster
```

```
status = opsys=linux,uname=Linux bio013 2.6.24-24-server #1 SMP Tue Aug 18
```

```
17:46:20 UTC 2009 i686,sessions=5231,nsessions=1,nusers=1,idletime=753208,totmem=8155868kb,availmem=8127532kb,physmem=2075308kb,ncpus=4,loadave=0.00,netload=355066597,state=free,jobs=,varattr=,rectime=1257890827
```

```
bio014
```

```
state = free
```

```
np = 1
```

```
ntype = cluster
```

```
status = opsys=linux,uname=Linux bio014 2.6.24-24-server #1 SMP Tue Aug 18
```

```
17:46:20 UTC 2009 i686,sessions=? 15201,nsessions=? 15201,nusers=0,idletime=753208,totmem=2205248kb,availmem=2180520kb,physmem=2075308kb,ncpus=4,loadave=0.00,netload=352801885,state=free,jobs=,varattr=,rectime=1257890829
```

```
bio015
```

```
state = free
```

```
np = 1
```

```
ntype = cluster
```

```
status = opsys=linux,uname=Linux bio015 2.6.24-24-server #1 SMP Tue Aug 18
```

```
17:46:20 UTC 2009 i686,sessions=? 15201,nsessions=? 15201,nusers=0,idletime=753213,totmem=2205260kb,availmem=2180744kb,physmem=2075308kb,ncpus=4,loadave=0.00,netload=351992936,state=free,jobs=,varattr=,rectime=1257890840
```

```
bio016
```

```
state = free
```

```
np = 1
```

```
ntype = cluster
```

```
status = opsys=linux,uname=Linux bio016 2.6.24-24-server #1 SMP Tue Aug 18
```

```
17:46:20 UTC 2009 i686,sessions=? 15201,nsessions=? 15201,nusers=0,idletime=753208,totmem=2204992kb,availmem=2179068kb,physmem=2075308kb,ncpus=4,loadave=0.00,netload=356194245,state=free,jobs=,varattr=,rectime=1257890853
```

送出 JOB

- echo " 指令 " | qsub
 - ex : echo "sleep 60; date" | qsub

```
wade@bio037:~$ echo "sleep 60; date" | qsub
```

```
68.bio037
```

job ID

host name

送出 JOB

- qsub [script]
 - ex: qsub PBStest.pbs

```
wade@bio037:~/qsub$ qsub PBStest.pbs  
133.bio037
```

查詢 JOB 狀態

- qstat

```
wade@bio037:~/qsub$ qstat
Job id          Name          User          Time Use S Queue
-----
134.bio037     mytest       wade          0 R batch
135.bio037     mytest       wade          0 R batch
136.bio037     mytest       wade          0 R batch
137.bio037     mytest       wade          0 R batch
138.bio037     mytest       wade          0 R batch
139.bio037     mytest       wade          0 R batch
140.bio037     mytest       wade          0 R batch
141.bio037     mytest       wade          0 R batch
142.bio037     mytest       wade          0 R batch
143.bio037     mytest       wade          0 R batch
```

查詢 JOB 狀態

- `qstat -f`

```
wade@bio037:~/qsub$ qstat -f
```

```
Job Id: 187.bio037
```

```
Job_Name = mytest
```

```
Job_Owner = wade@bio037
```

```
job_state = R
```

```
queue = batch
```

```
server = bio037
```

```
Checkpoint = u
```

```
ctime = Wed Nov 11 06:01:46 2009
```

```
Error_Path = bio037:/home/wade/qsub/./PBStest.err
```

```
exec_host = bio016/0+bio015/0+bio014/0+bio013/0
```

```
Hold_Types = n
```

```
Join_Path = n
```

```
Keep_Files = n
```

```
Mail_Points = a
```

```
mtime = Wed Nov 11 06:01:46 2009
```

```
Output_Path = bio037:/home/wade/qsub/./PBStest.log
```

```
Priority = 0
```

```
qtime = Wed Nov 11 06:01:46 2009
```

```
Rerunnable = True
```

```
Resource_List.nodect = 4
```

```
Resource_List.nodes = 4
```

```
Variable_List = PBS_O_HOME=/home/wade, PBS_O_LANG=en_US.UTF-8,
```

```
  PBS_O_LOGNAME=wade,
```

```
  PBS_O_PATH=/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/b
```

```
in:/usr/games:/usr/lib/jvm/java-6-sun/bin:/opt/torque/bin:/opt/mpich2
```

```
/bin:/opt/mpiblast/bin:/opt/intel/Compiler/11.0/084/bin/ia32:/opt/torq
```

```
ue/bin:/opt/mpich2/bin:/opt/mpiblast/bin:/opt/sim4:/opt/autofact/scrip
```

```
ts:/opt/PHYLIP/exe:/opt/bowtie:/opt/soap1:/opt/velvet:/opt/clustalW,
```

```
  PBS_O_MAIL=/var/mail/wade, PBS_O_SHELL=/bin/bash, PBS_SERVER=bio037,
```

```
  PBS_O_HOST=bio037, PBS_O_WORKDIR=/home/wade/qsub, PBS_O_QUEUE=batch
```

```
comment = Job started on Wed Nov 11 at 06:01
```

```
etime = Wed Nov 11 06:01:46 2009
```

```
submit_args = mpi-4node.pbs
```

```
start_time = Wed Nov 11 06:01:46 2009
```

```
start_count = 1
```

移除 JOB

- `qdel [Job id]`

```
wade@bio037:~/qsub$ qstat
Job id                Name                User                Time Use  S Queue
-----
170.bio037            mytest              wade                0 W  batch
171.bio037            mytest              wade                0 W  batch
172.bio037            mytest              wade                0 W  batch
173.bio037            mytest              wade                0 W  batch
wade@bio037:~/qsub$ qdel 171
wade@bio037:~/qsub$ qstat
Job id                Name                User                Time Use  S Queue
-----
170.bio037            mytest              wade                0 W  batch
172.bio037            mytest              wade                0 W  batch
173.bio037            mytest              wade                0 W  batch
```


PBS 指令

- qsub
- qstat
- qdel
- qmgr
- qrls
- qsig
- qrun
- qselect
- hostn
- qhold
- qmove
- qstart
- qstop
- printtracking
- qdisable
- chk_tree
- pbsnodes
- pbs_track
- pbsdsh
- qalter
- qorder
- qrerun
- nqs2pbs
- printjob
- qenable
- printserverdb
- tracejob
- pbs-config
- qterm

PBS SCRIPT 撰寫

- PBS SCRIPT 規則
- PBS SCRIPT 範例

PBS SCRIPT 規則

- PBS 指令
 - 1個 #
- PBS 註解
 - 3個 #

PBS SCRIPT 範例

- `#!/bin/bash`
- `### Job 名稱`
- `#PBS -N mytest`
- `### 錯誤及輸出檔案名稱`
- `#PBS -e mytest.err`
- `#PBS -o mytest.log`
- `### 最長執行時間`
- `#PBS -l walltime=01:00:00`

PBStest.pbs

```
#!/bin/bash
### Job name
#PBS -N mytest
###=====
# 顯示目錄及時間資訊
echo Working directory is $PBS_O_WORKDIR
cd $PBS_O_WORKDIR
echo Running on host `hostname`
echo Time is `date`
echo Directory is `pwd`
# 執行檔案
date
```

輸出結果

- 標準錯誤輸出檔： [Job name] . e [Job ID]
- 標準輸出 (結果) 檔： [Job name] . o [Job ID]

```
wade@bio037:~/qsub$ ls
cpi                mpi-4node.pbs    mytest.e183     mytest.o182     PBStest-1.pbs
mpd.hosts          mytest.e181     mytest.e184     mytest.o183     PBStest.pbs
mpi-1node.pbs     mytest.e182     mytest.o181     mytest.o184
```

```
wade@bio037:~/qsub$ cat mytest.o184
Working directory is /home/wade/qsub
Running on host bio013
Time is Wed Nov 11 05:23:04 CST 2009
Directory is /home/wade/qsub
Wed Nov 11 05:23:04 CST 2009
wade@bio037:~/qsub$
```

PBStest-1.pbs

```
#!/bin/bash
### Job name
#PBS -N mytest
### output file
#PBS -e ./PBStest.err
#PBS -o ./PBStest.log
###=====
# 顯示目錄及時間資訊
echo Working directory is $PBS_O_WORKDIR
cd $PBS_O_WORKDIR
echo Running on host `hostname`
echo Time is `date`
echo Directory is `pwd`
# 執行檔案
sleep 30
date
```


輸出結果

```
wade@bio037:~/qsub$ cat PBStest.log
Working directory is /home/wade/qsub
Running on host bio013
Time is Wed Nov 11 04:18:00 CST 2009
Directory is /home/wade/qsub
Wed Nov 11 04:18:30 CST 2009
```

陽明生物計算軟體測試 SCRIPT 撰寫

- MPICH
- BLAST
- SIM4
- Velvet
- Bowtie
- SOAP

MPICH

- 檔名：mpi-4node.pbs

```
#!/bin/bash
### Job name
#PBS -N mytest
### output file
#PBS -e ./PBStest.err
#PBS -o ./PBStest.log
### how many nodes
#PBS -l nodes=4
#####
# 顯示目錄及時間資訊
echo Working directory is $PBS_O_WORKDIR
cd $PBS_O_WORKDIR
echo Running on host `hostname`
echo Time is `date`
echo Directory is `pwd`
# 執行檔案
mpdboot -n 23
mpiexec -n 10 ./cpi
mpdallexit
```

MPICH- 執行結果

- PBStest.log

```
wade@bio037:~/qsub$ cat PBStest.log
Working directory is /home/wade/qsub
Running on host bio016
Time is Wed Nov 11 05:53:26 CST 2009
Directory is /home/wade/qsub
pi is approximately 3.1415926535895626, Error is 0.00000000000002305
wall clock time = 5.581908
wade@bio037:~/qsub$ █
```

BLAST

```
#!/bin/bash
### Job name
#PBS -N mytest
### output file
#PBS -e ./PBStest.err
#PBS -o ./PBStest.log
###=====
# 顯示目錄及時間資訊
echo Working directory is $PBS_O_WORKDIR
cd $PBS_O_WORKDIR
echo Running on host `hostname`
echo Time is `date`
echo Directory is `pwd`
# 執行檔案
bl2seq -i nt.2655203 -j nt.3283410 -p blastn -o bl2seq.out
formatdb -i nt.ests -p F -o T -n test.nt
fastacmd -d test.nt -s 5986
blastall -p blastn -d test.nt -i test.ests -e 0.1 -o blastall.out
```

BLAST- 執行結果

- PBStest.log

```
wade@bio037:~/testdata/blast_test$ cat PBStest.log
Working directory is /home/wade/testdata/blast_test
Running on host bio013
Time is Wed Nov 11 06:33:27 CST 2009
Directory is /home/wade/testdata/blast_test
>gi|5986|emb|Z14321.1|Z14321 CEL11A10 Chris Martin sorted cDNA library Caenorhab
ditis elegans cDNA clone cm11a10 5'
CTTGAAGTTNTCGATAACAAATTTNTNATCCCAGAGAGAAAAGTGAACGAGGAGCCAATGTTTGCAGCGGANACGTTTAC
TCGATCGTNGGGCGTGGAACAGTTATCACTGGAAAGCTTGAGAGAGGAATTTTNAAGAGAGGAGATAAGATTGAAATTNT
TGGAGGAACAAAAGATGGAACCACTGTGAAATCAGTAATCTCTGGATTAGAATCCTTCCGTAAAACACTGTCGATCAAGCCG
AGCCTGGAGATCAATTGGNTGTGCTTCTTCGTGGTTTAGGNCCAAAAGATGTTTCGTAGAGGATGTGTTTTACTTCCACAA
GGNCACAAACATAAGGTCCTGATAAGGTTAAAGCTCAACTTTATGTGCTCAAAGAGAGC
```

BLAST- 執行結果

- 由 BLAST 所產生的結果：`blastall.out`


```

Query: 1   cttgaagttntcgataacaaattnatcccagagagaaaagtgaacgaggagccaatg 60
          |||
Sbjct: 1   cttgaagttntcgataacaaattnatcccagagagaaaagtgaacgaggagccaatg 60

Query: 61  tttgcagcgganacgtttactcgatcgtngggcgtggaacagttatcactggaaagcttg 120
          |||
Sbjct: 61  tttgcagcgganacgtttactcgatcgtngggcgtggaacagttatcactggaaagcttg 120

Query: 121 agagaggaattnaagagaggagataagattgaaatnttggaggaacaaaagatggaa 180
          |||
Sbjct: 121 agagaggaattnaagagaggagataagattgaaatnttggaggaacaaaagatggaa 180

Query: 181 ccactgtgaaatcagtaatctctggattagaatccttccgtaaaactgtcgatcaagccg 240
          |||
Sbjct: 181 ccactgtgaaatcagtaatctctggattagaatccttccgtaaaactgtcgatcaagccg 240

Query: 241 agcctggagatcaattggntgtgcttcttcgtggtttaggnccaaaagatggtcgtagag 300
          |||
Sbjct: 241 agcctggagatcaattggntgtgcttcttcgtggtttaggnccaaaagatggtcgtagag 300

Query: 301 gatgtgttttacttccacaaggncacaaacataagggtcactgataagggttaagctcaac 360
          |||
Sbjct: 301 gatgtgttttacttccacaaggncacaaacataagggtcactgataagggttaagctcaac 360

Query: 361 tttatgtgctcaaagagagc 380
          |||
Sbjct: 361 tttatgtgctcaaagagagc 380

```

SIM4

```
#!/bin/bash
### Job name
#PBS -N mytest
### output file
#PBS -e ./PBStest.err
#PBS -o ./PBStest.log
###=====
# 顯示目錄及時間資訊
echo Working directory is $PBS_O_WORKDIR
cd $PBS_O_WORKDIR
echo Running on host `hostname`
echo Time is `date`
echo Directory is `pwd`
# 執行檔案
sim4 query_test.txt reference.fasta
1,1
Top
```

SIM4- 執行結果

- PBStest.log

```
Working directory is /home/wade/testdata/sim4_test
Running on host bio013
Time is Wed Nov 11 06:47:22 CST 2009
Directory is /home/wade/testdata/sim4_test

seq1 = query_test.txt, 25 bp
seq2 = reference.fasta (scaffold000001), 4743 bp

1-25 (2326-2350) 100%

1,1 All
```

velvet

```
#!/bin/bash
### Job name
#PBS -N mytest
### output file
#PBS -e ./PBStest.err
#PBS -o ./PBStest.log
###=====
# 顯示目錄及時間資訊
echo Working directory is $PBS_O_WORKDIR
cd $PBS_O_WORKDIR
echo Running on host `hostname`
echo Time is `date`
echo Directory is `pwd`
# 執行檔案
velveth directory 17 -fasta -short F_reads.fasta
velvetg directory -exp_cov 3 -min_contig_lgth 100
1,1 Top
```

velvet- 執行結果

- 產生新的目錄：`directory`
- 目錄檔案：
 - `contigs.fa`
 - `Graph2`
 - `LastGraph`
 - `Log`
 - `PreGraph`
 - `Roadmaps`
 - `Sequences`
 - `stats.txt`

velvet

- PBStest.log

```
Initial node count 58018
Removed 0 null nodes
Concatenation over!
58018 nodes left
Read coherency...
Identifying unique nodes
Done, 10224 unique nodes counted
Trimming read tips
Renumbering nodes
Initial node count 58018
Removed 0 null nodes
Confronted to 0 multiple hits and 0 null over 0
Read coherency over!
Starting pebble resolution...
Preparing to correct graph with cutoff 0.200000
Computing read to node mapping array sizes
Computing read to node mappings
Estimating library insert lengths...
Done
Computing direct node to node mappings
Scaffolding node -50000
Scaffolding node -40000
Scaffolding node -30000
Scaffolding node -20000
Scaffolding node -10000
Scaffolding node 0
Scaffolding node 10000
Scaffolding node 20000
Scaffolding node 30000
Scaffolding node 40000
Scaffolding node 50000
Cleaning memory
Deactivating local correction settings
Pebble done.
Concatenation...
Renumbering nodes
Initial node count 58018
Removed 0 null nodes
Concatenation over!
Writing into stats file directory/stats.txt...
Writing into graph file directory/LastGraph...
Final graph has 58018 nodes and n50 of 44, max 929, total 2507449, using 612458/8080777 reads
wade@bio037:~/testdata/velvet_test/velvet_test$
```


bowtie

```
#!/bin/bash
### Job name
#PBS -N mytest
### output file
#PBS -e ./PBStest.err
#PBS -o ./PBStest.log
###=====
# 顯示目錄及時間資訊
echo Working directory is $PBS_O_WORKDIR
cd $PBS_O_WORKDIR
echo Running on host `hostname`
echo Time is `date`
echo Directory is `pwd`
# 執行檔案
bowtie -f -m 1 -l 25 -n 2 gl_ref ./bowtie_test_reads.txt output.txt
1,1 Top
```

bowtie- 執行結果

- output.txt

```
1379_372_66_F3 + scaffold00001 length=4743 2325 TTCTCAGTCAGTCTTCATGGCAGCA IXXXXXXXXXXXXXXXXXXXXX
II 0
1380_574_1082_F3 + scaffold00001 length=4743 502 TCCTCCACCGCACCTCCTCTCTCTC IXXXXXXXXXXXXXXXX
XXXXXXXXXX 0
1380_770_1324_F3 - scaffold00001 length=4743 378 AAGAGGAGAGAGCCGAGAGACGACT IXXXXXXXXXXXXXXXX
XXXXXXXXXX 0
1380_1073_552_F3 + scaffold00001 length=4743 843 CACCCAGACTGCGTTCCTTCAACAC IXXXXXXXXXXXXXXXX
XXXXXXXXXX 0
1380_1389_821_F3 - scaffold00001 length=4743 4304 TTCACATGCACACGTCAACACCCA IXXXXXXXXXXXXXXXX
XXXXXXXXXX 0 23:A>T,24:A>T
1380_1879_1685_F3 + scaffold00001 length=4743 2081 TCCGAAATATCCTGCGGATCTTATT IXXXXXXXXXXXXXXXX
XXXXXXXXXX 0 2:T>C
```

bowtie- 執行結果

- PBStest.err

```
wade@bio037:~/testdata/bowtie_test/bowtie_test$ cat PBStest.err  
Reported 6 alignments to 1 output stream(s)
```

- PBStest.log

```
wade@bio037:~/testdata/bowtie_test/bowtie_test$ cat PBStest.log  
Working directory is /home/wade/testdata/bowtie_test/bowtie_test  
Running on host bio013  
Time is Wed Nov 11 07:13:03 CST 2009  
Directory is /home/wade/testdata/bowtie_test/bowtie_test
```

soap

```
#!/bin/bash
### Job name
#PBS -N mytest
### output file
#PBS -e ./PBStest.err
#PBS -o ./PBStest.log
###=====
# 顯示目錄及時間資訊
echo Working directory is $PBS_O_WORKDIR
cd $PBS_O_WORKDIR
echo Running on host `hostname`
echo Time is `date`
echo Directory is `pwd`
# 執行檔案
soap -a soap_test_reads.txt -d reference.fasta -o output.txt -s 10 -r 0
```

soap- 執行結果

- output.txt

```
1379_372_66_F3 TTCTCAGTCAGTCTTCATGGCAGCA hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh 1 a 25 + scaffold000001 2326 0
1380_574_1082_F3 TCCTCCACCGCACCTCCTCTCTCTC hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh 1 a 25 + scaffold000001 503 0
1380_770_1324_F3 AAGAGGAGAGAGCCGAGAGACGACT hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh 1 a 25 - scaffold000001 379 0
1380_1073_552_F3 CACCCAGACTGCGTTCCTTCAACAC hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh 1 a 25 + scaffold000001 844 0
1380_1389_821_F3 TTTCACATGCACACGTCAACACCCA hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh 1 a 25 - scaffold000001 4305 2 A->0T40 A->1T40
1380_1879_1685_F3 TCCGAAATATCCTGCCGATCTTATT hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh 1 a 25 + scaffold000001 2082 1 T->2C40
```

soap- 執行結果

- PBStest.log

```
Working directory is /home/wade/testdata/soap_test/soap_test
Running on host bio013
Time is Wed Nov 11 07:31:12 CST 2009
Directory is /home/wade/testdata/soap_test/soap_test
Start at: Wed Nov 11 07:31:12 2009

Load in 1 db seqs, total size 4743 bp. 0 secs passed
total_kmers: 1048576
Create seed table. 0 secs passed
Single read alignment:
Query: soap_test_reads.txt Reference: reference.fasta Output: output.txt
104 reads finished. 0 secs passed
Total number of aligned reads: 6 (5.8%)
Done.
Finished at Wed Nov 11 07:31:12 2009
Total time consumed: 0 secs
```