

PBSの利用方法

How to use PBS

22/Nov./2022 Hiroyo NISHIDE

Copy the directory for the course.

1. Copy `"/bio/course/bias2211"` to your home directory.
2. Move to `"~/bias2211/pbs_lec"`
3. List directory contents `"ls"`

```
$ cp -r /bio/course/bias2211 ~/
```

```
$ cd ~/bias2211/pbs_lec
```

```
$ ls
```

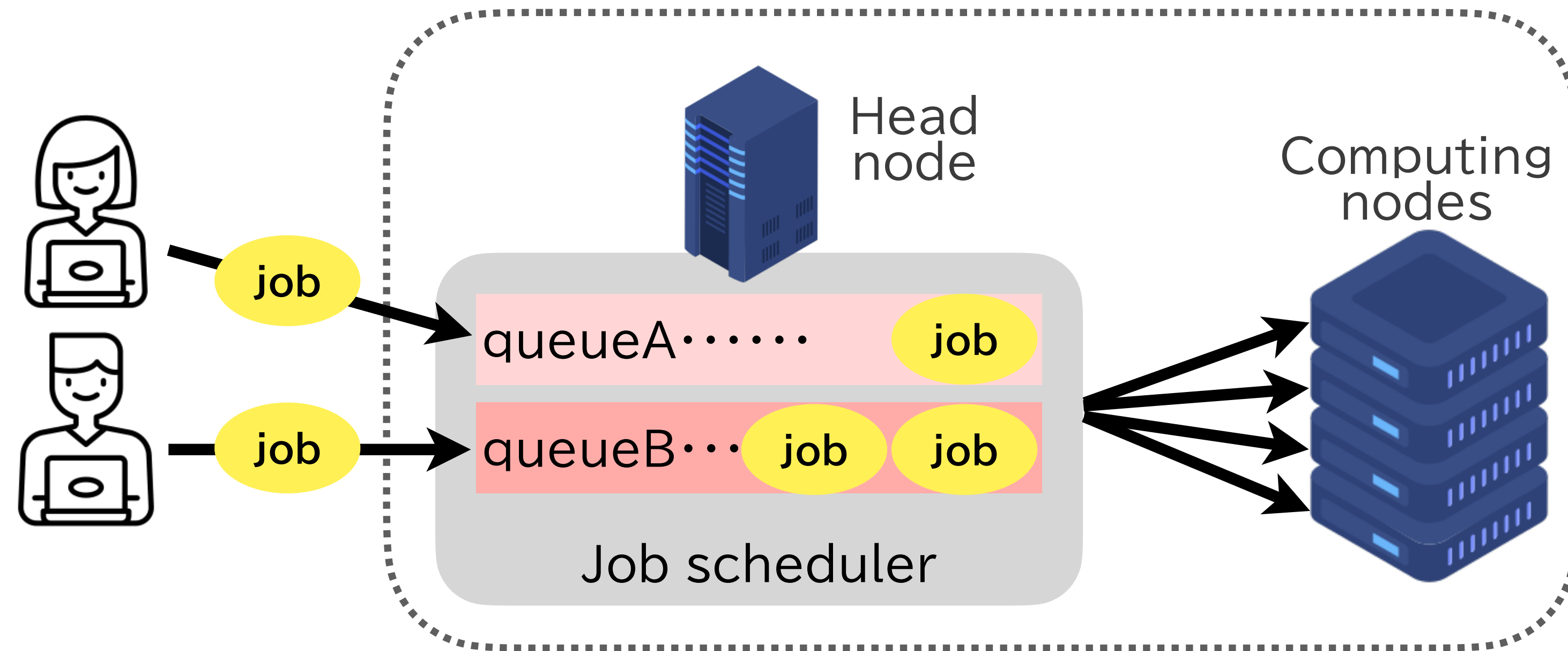
```
diamond.sh  sce_prot.fasta  spo.dmnd  test_qsub.sh
```

Job Management System: PBS

- 複数の人間が同じ計算機群を使う
- どのマシン/CPUが空いてるか？ どの計算を優先させるべきか？

ジョブ管理システム

- ▶ 計算機資源の割り当てを自動で行い効率を上げる
- ▶ ユーザはコマンドで親ノードにジョブを投げるだけ
- ▶ ジョブは「キュー」と呼ばれるリストに入って実行を待つ



ジョブ管理システム **PBS**

- PBSにおけるジョブ：
実行するコマンドやアプリケーションを記述したシェルスクリプト形式のタスク

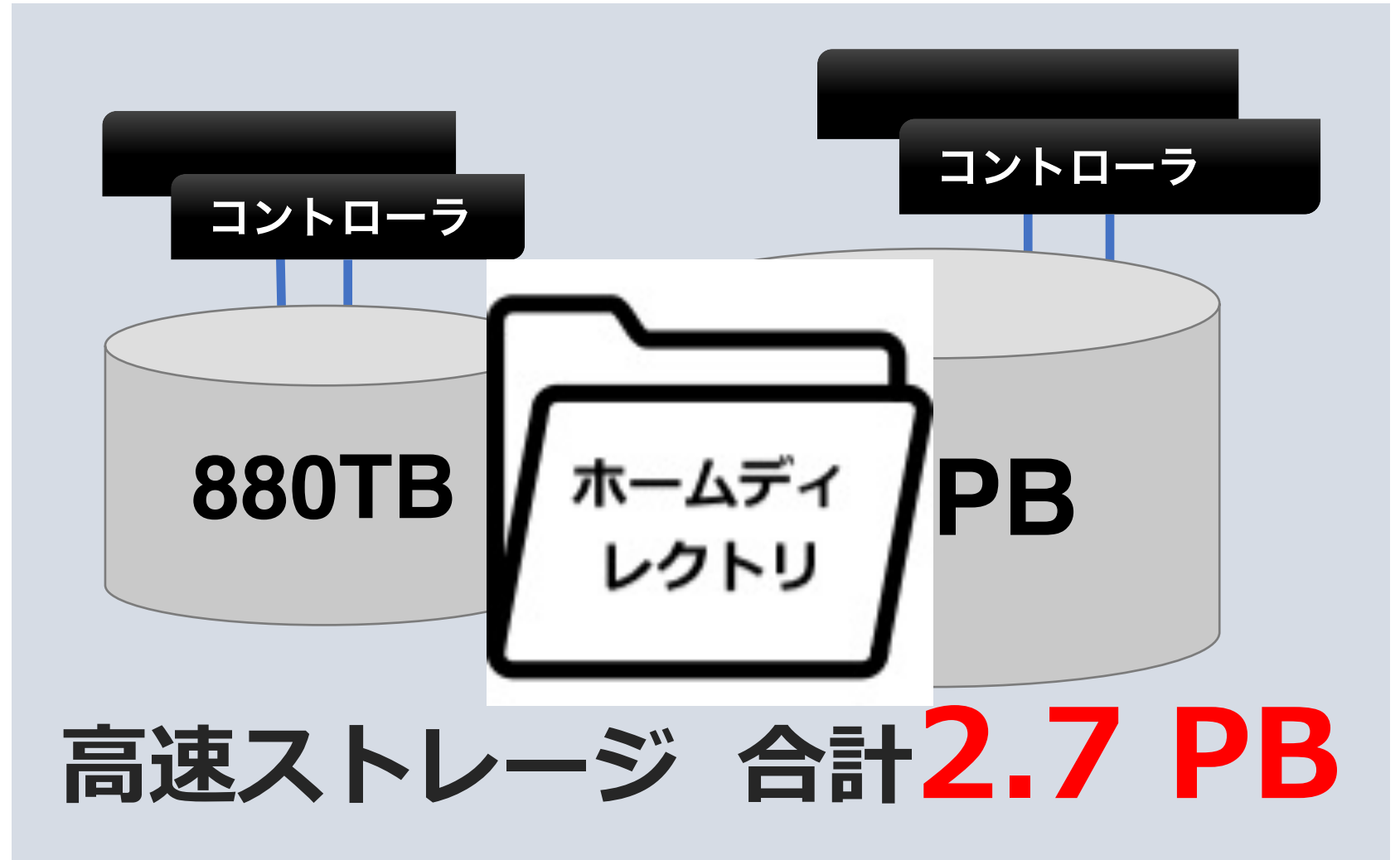
生物情報解析システム (bias5)

共有メモリ型計算サーバ
3TB Mem 72 cores
bias5-smp.nibb.ac.jp

共有メモリ型計算サーバ
4TB Mem 80 cores
ldas-smp.nibb.ac.jp
smpls, smpm, smpl

共有メモリ型計算サーバ
527GB Mem 64 cores
diaf-smp1.nibb.ac.jp

共有メモリ型計算サーバ
527GB Mem 64 cores
diaf-smp2.nibb.ac.jp



small, medium
large, blast

分散処理計算機クラスタ
800cores 4.8GB Mem/core
bias5-node[01-20].nibb.ac.jp



ログインノード
768GB Mem
bias5.nibb.ac.jp



生物情報解析システム (bias5)

共有メモリ型計算サーバ

3TB Mem 72 cores
bias5-smp.nibb.ac.jp

共有メモリ型計算サーバ

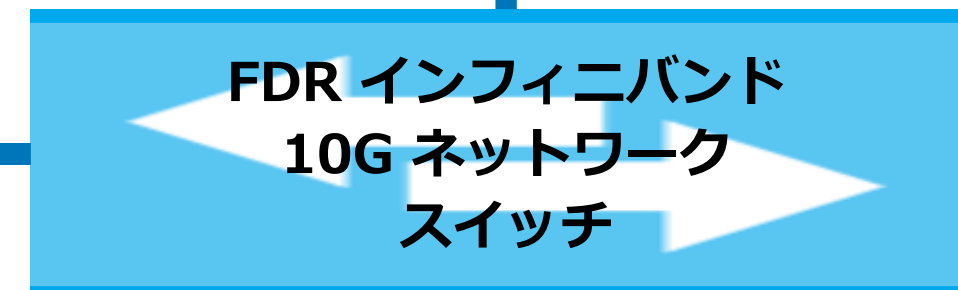
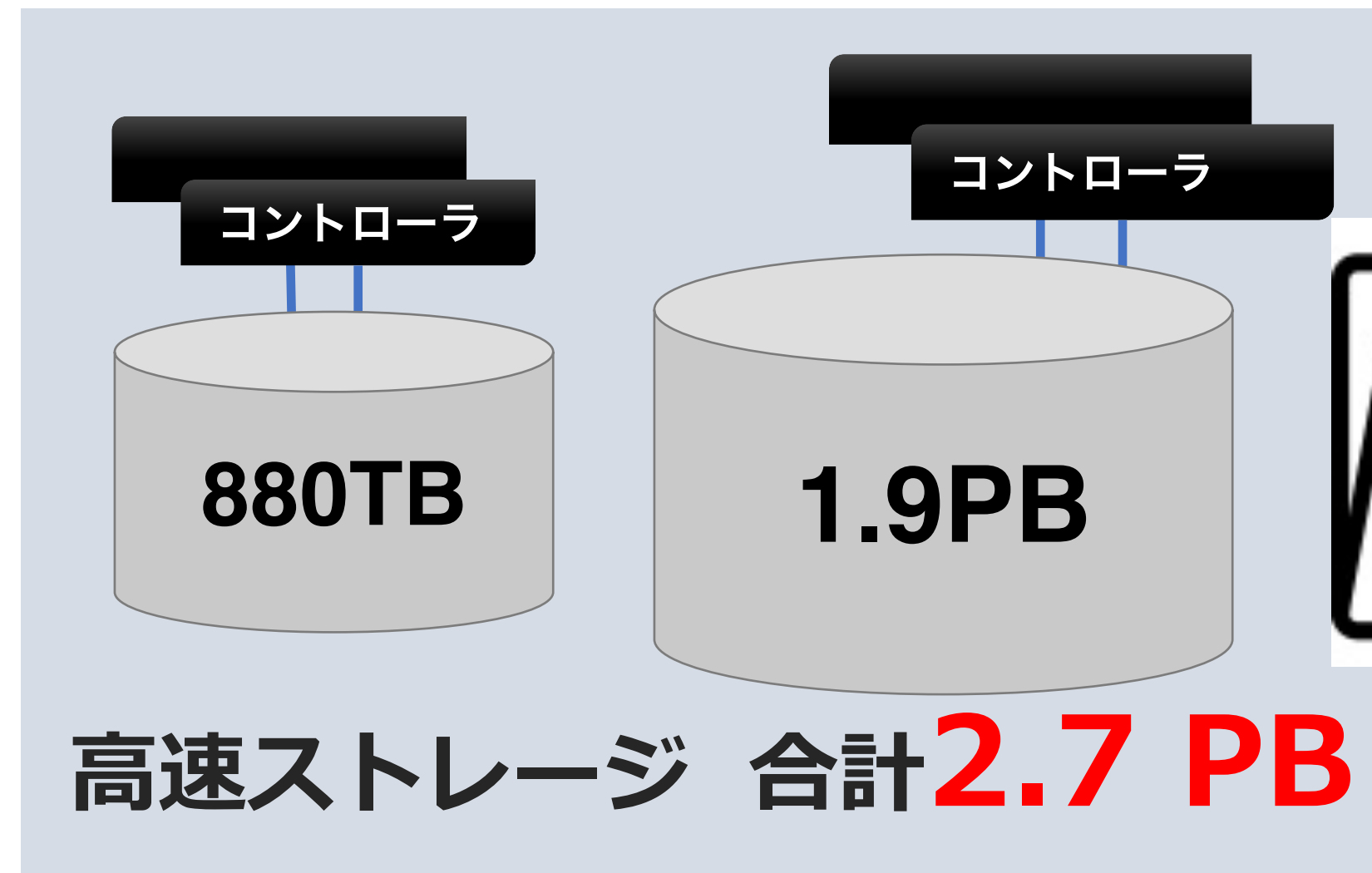
4TB Mem 80 cores
ldas-smp.nibb.ac.jp

共有メモリ型計算サーバ

527GB Mem 64 cores
diaf-smp1.nibb.ac.jp

共有メモリ型計算サーバ

527GB Mem 64 cores
diaf-smp2.nibb.ac.jp



Never do data analysis directly on bias5

Light testing, text manipulation, and data copying are exempt.

Always use PBS(**qsub** command) for analysis.

Flow of job execution by PBS

- Prepare shell script file

test_qsub.sh

```
#!/bin/bash
cd ${PBS_O_WORKDIR}

diamond blastp --db spo.dmnd --out sce.tab --outfmt 6 --query sce_prot.fasta
```

- **qsub** submits the job. **qsub** with the shell script file.

```
$ qsub test_qsub.sh
5812.bias5-adm
```

- **qstat** displays execution status

```
$ qstat
```

Practice: Submit your job

- List directory contents, and read "qsub_test.sh" file

```
$ ls  
$ less test_qsub.sh
```

- Submit the job. **qsub** with "qsub_test.sh"

```
$ qsub test_qsub.sh
```

- **qstat** displays execution status

```
$ qstat
```

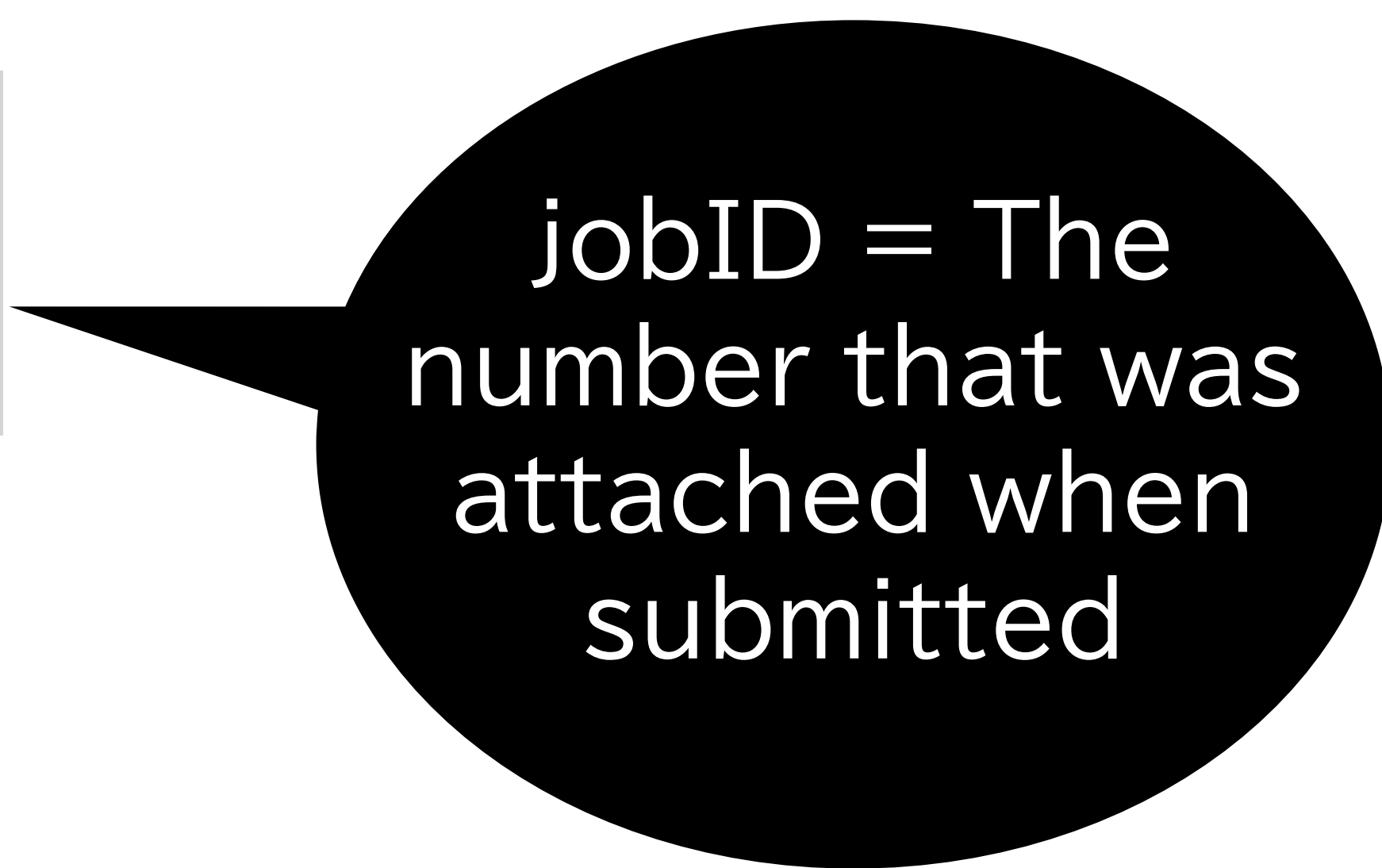

Practice: Show results

- List directory contents, and read "sce.tab" file (diamond results)

```
$ ls  
$ less sce.tab
```

- Read standard output stream file & standard error stream file

```
$ less test_qsub.ejobID  
$ less test_qsub.ojobID
```



jobID = The number that was attached when submitted

- **tracejob** : Confirm execution status

```
$ tracejob jobID
```

Submit PBS job : qsub

- PBSにおけるジョブ:
実行するコマンドやアプリケーションを記述したシェルスクリプト形式のタスク

```
$ qsub [option] スクリプトファイル
```

```
$ qsub diamond.sh  
5812.bias5-adm
```

- ↑ **5812** = **jobID**
- `bias5-adm` = our Head node

PBSジョブのためのシェルスクリプトファイル（ジョブファイル）

diamond.sh

```
#!/bin/bash
#PBS -q small
#PBS -l ncpus=4
#PBS -l mem=12gb

cd ${PBS_O_WORKDIR}

diamond blastp --threads ${NCPUS} --db /bio/db/diamond/swissprot/swissprot \
--outfmt 6 --query ./sce_prot.fasta --out ./sce2.tab --sensitive
```

←①シェルの指定

←② qsubオプション

←③ジョブ投入ディレクトリへの移動

←④実行するコマンド

① シェルの指定 : ジョブスクリプトのシェルの種類を指定

② qsubオプション

行の先頭を #PBS で始めると、qsubコマンドのオプション指示として処理される

③ ジョブ投入ディレクトリへの移動

qsubコマンドを実行したディレクトリに移動（変数 `${PBS_O_WORKDIR}` に qsub実行ディレクトリ名が自動で設定される）

これをしないとホームディレクトリにいる状態で続くコマンドを実行しようとする

④ 実行するコマンド : コマンド、パスの設定、変数のセット等スクリプト本体を記述

qsub options

\$ man qsub

- #PBS に続けてジョブスクリプト内に書いても、qsubコマンドに付けてもよい（コマンド付けの方が優先）

option	description	default
-q <i>queue_name</i>	Where the job is sent upon submission	small queue
-N <i>job_name</i>	Sets job's name to <i>job_name</i>	job file name
-e <i>file_name</i>	<i>file_name</i> to be used for the job's standard error stream	<i>job_name.ejobID</i>
-o <i>file_name</i>	<i>file_name</i> to be used for the job's standard output stream	<i>job_name.ojobID</i>
-j [oe, eo]	Whether and how to join the job's standard error and standard output streams. oe : Standard error and standard output are merged into standard output. eo : Standard error and standard output are merged into standard error.	
-J X-Y[:Z]	Declares that this job is an array job.	
-l <i>resource_list</i>	Allows the user to request resources and specify job placement.	
-v <i>variable_list</i>	Lists environment variables and shell functions to be exported to the job. EX) -v var1=10, "var2='A,B'"	
-S <i>shell_name</i>	Specifies the interpreter or shell path for the job script.	your login shell
-V	Declares that all environment variables and shell functions in the user's login environment where qsub is run are to be exported to the job.	

Queue Configuration

	分散処理用計算機クラス (800core) parallel distributed processing				共有メモリ型計算サーバ shared memory		
queue name	small (default)	medium	large	blast	smps	smpm	smpi
Characteristics of queue	short time/many parallel lines	mid-scale	long time	for BLAST nr/nt	middle memory	large memory	Max memory
Avairable hosts	bias5-node01~ bias5-node20	bias5-node01~ bias5-node20	bias5-node01~ bias5-node20	bias5-node01~ bias5-node20	bias5-smp ldas-smp diaf-smp1 diaf-smp2	bias5-smp ldas-smp	bias5-smp ldas-smp
Max. execution time / job	6h	72h	no limit	no limit	no limit	no limit	no limit
Max.num of running jobs / queue	no limit	no limit	no limit	no limit	16	6	2
Max. num of running cpus / queue	580	200	20	120	176	96	72
Max.num of running cpus / user	480						
Max.num of running cpus / user.queue	300	150	10	80	no limit	no limit	no limit
Default num of CPUs	1	1	1	20	1	1	1
Max. num of running cpus / host	40				72/80/64/64	72/80	
Max. memory usage / job	96GB	96GB	96GB	188GB	500GB	1TB	3TB
Default memory size / job	3.5GB	3.5GB	3.5GB	98GB	250GB	500GB	1500GB

- If you do not specify a queue, it will be submitted to the **small** (default) queue
- Jobs submitted in excess of the number of concurrent jobs will wait for execution.

Concurrent executions

	分散処理用計算機クラスター (800core)				共有メモリ型計算サーバ		
queue name	small (default)	medium	large	blast	smps	smpm	smpi
キューの特徴	短時間・並列多	中規模	長時間	for BLAST nr/nt	中メモリ	大メモリ	最大メモリ
Available hosts	bias5-node01～ bias5-node20	bias5-node01～ bias5-node20	bias5-node01～ bias5-node20	bias5-node01～ bias5-node20	bias5-smp ldas-smp diaf-smp1 diaf-smp2	bias5-smp ldas-smp	bias5-smp ldas-smp
最大同時実行ジョブ数 / キュー	no limit	no limit	no limit	no limit	16	6	2
最大同時使用CPU数 / キュー	580	200	20	120	176	96	72
最大同時実行CPU数 / ユーザ	480						
最大同時実行CPU数 / ユーザ/キュー	300	150	10	80	no limit	no limit	no limit
デフォルト使用CPU数 / ジョブ	1	1	1	20	1	1	1
最大同時使用CPU数 / ホスト	40				72/80/64/64	72/80	

smallキューでは...

- キューあたりの同時実行CPU数 **580**
- 1ユーザあたりの同時実行CPU数 **300**

キューをまたいだ1ユーザあたりの同時実行CPU数 **480**

Memory limitation

- Forced kill when job memory usage exceeds the specified amount of memory
 - Only small, medium, large and blast queue
- message:

=>> PBS: job killed: mem 3977304kb exceeded limit 3686400kb

	分散処理用計算機クラス (800core)			
queue name	small (default)	medium	large	blast
キューの特徴	短時間・並列多	中規模	長時間	for BLAST nr/nt
Avairable hosts	bias5-node01~ bias5-node20	bias5-node01~ bias5-node20	bias5-node01~ bias5-node20	bias5-node01~ bias5-node20
Max.memory usage / job	96GB	96GB	96GB	188GB
Default memory size / job	3.5GB	3.5GB	3.5GB	98GB

CPU数	総メモリ量
40/1台	198GB/1台

- small, medium, large キューの上限まで設定すると1台あたり2ジョブ
- blast キューの上限まで設定すると1台あたり1ジョブ

How to specify resources (memory, cpu..)

host	CPU	Memory
bias5-node01~20	40/1台	198GB/1台
bias5-smp	72	3TB
ldas-smp	80	4TB
diaf-smp[1, 2]	64/1台	527GB/1台

- qsub option **-l**

`-l resource_name=value[,resource_name=value ...]`

リソース名=値

複数つける場合は一行ずつ書くか、カンマで区切る `#PBS -l ncpus=8,mem=48GB`

`#PBS -l ncpus=8`

CPU数を指定

この値は変数 `#{NCPUS}` に自動でセットされるので、CPU数を指定できるアプリケーションにはこれを使うこと

`#PBS -l mem=48GB`

最大利用メモリ量を指定

キューの選び方について

- mediumキュー(~72h)が混んでいることが多いが、smallの制限時間である6h以内に終わっているジョブがほとんど
- まずはsmallキューに、デフォルト条件で投入してみて、時間やメモリ利用量がオーバーするようなら他のキュー利用を検討いただきたい
- blastキューは最大188GBまでメモリを使えるので、smp[s, m, l]キューが混んでいる場合はこちらを利用してみてください

PBSジョブのためのシェルスクリプトファイル（ジョブファイル）

diamond.sh

```
#!/bin/bash  
#PBS -q small  
#PBS -l ncpus=4  
#PBS -l mem=12gb  
  
cd ${PBS_O_WORKDIR}  
  
diamond blastp --threads ${NCPUS} --db /bio/db/diamond/swissprot/swissprot \  
--outfmt 6 --query ./sce_prot.fasta --out ./sce2.tab --sensitive
```

←①シェルの指定

←② qsubオプション

←③ジョブ投入ディレクトリへの移動

←④実行するコマンド

① シェルの指定 : ジョブスクリプトのシェルの種類を指定

② qsubオプション

行の先頭を #PBS で始めると、qsubコマンドのオプション指示として処理される

③ ジョブ投入ディレクトリへの移動

qsubコマンドを実行したディレクトリに移動（変数 `${PBS_O_WORKDIR}` に qsub実行ディレクトリ名が自動で設定される）

これをしないとホームディレクトリにいる状態で続くコマンドを実行しようとする

④ 実行するコマンド : コマンド、パスの設定、変数のセット等スクリプト本体を記述

Display status of PBS jobs, queues, or servers : **qstat**

```
$ qstat [option]
```

```
$ man qstat
```

<code>qstat -Q</code>	Display queue status
<code>qstat -Q <i>queue_name</i></code>	Display queue status (selected queue)
<code>qstat -Qf <i>queue_name</i></code>	Display queue status and conditions (selected queue)
<code>qstat</code>	Display all job status
<code>qstat -t</code>	Display all job status (job arrays, and subjobs)
<code>qstat -u <i>user_name</i></code>	Display job status (selected user)
<code>qstat -i -u <i>user_name</i></code>	Display job status (selected user, held or waiting jobs)
<code>qstat -f <i>jobID</i></code>	Display causes waiting to be executed (アレイジョブの場合は <i>jobID[]</i> と記述)
<code>qstat -r</code>	Display all running or suspended jobs
<code>qstat -nr1</code>	Display all running or suspended jobs with host name
<code>qstat -J</code>	Display all array jobs
<code>qstat -J -t -nr1</code>	Display all running array jobs with host name
<code>qstat -x</code>	Displays status information for finished

qstat

```
$ qstat
```

```
bias5-adm:
```

Job ID	Username	Queue	Jobname	SessID	NDS	TSK	Req'd Memory	Req'd Time	Status	Elap Time
362.bias5-adm	hiroyo	small	interpro	37276	1	1	3600mb	06:00	R	00:00
363.bias5-adm	hiroyo	small	interpro	145771	1	1	3600mb	06:00	R	00:00
364.bias5-adm	ideas	medium	dbupdate	145814	1	1	12gb	72:00	R	00:00
365.bias5-adm	sgiadm	small	testjob	145858	1	1	3600mb	06:00	R	00:00
366.bias5-adm	sgiadm	small	testjob	145902	1	1	3600mb	06:00	R	00:00
367.bias5-adm	nyacco	smps	myscript	145946	1	1	250gb	--	R	00:00
368[].bias5-a	hpe	blast	blastjob1	145990	1	1	3600mb	06:00	B	00:00
380.bias5-adm	sgiadm	small	TEST.sh	146033	1	1	8gb	06:00	R	00:00
381.bias5-adm	hogehe	smpm	Trinity_rn	--	1	1	500gb	--	Q	--

○ job ID

○ Status of job

R : Running

B : Array jobs

Q : Waiting

Eqw : Error

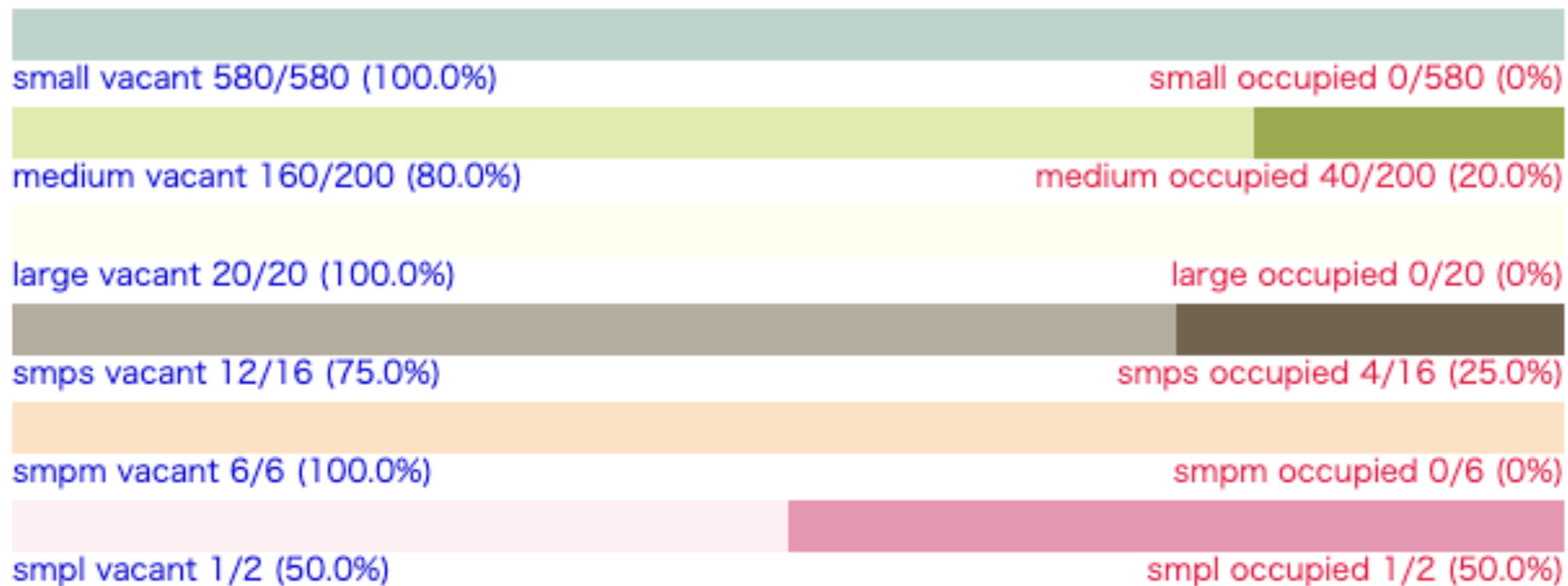
Node/Queue status (VPN required)

- Nodes

<http://bias5.nibb.ac.jp:8091/job/index3.html>

- Queues

<http://bias5.nibb.ac.jp:8091/job/index2.html>



node	status	#job	#Mem	Memory	run
node Total		78/800	1332/3768		
catm	free	0/32	0/504		
cats1	state-un	0/24	0/94		
cats2	free	0/24	0/94		
diaf-smp1	free	0/64	0/503		
diaf-smp2	free	48/64	500/503		
ldas-smp	free	0/80	0/4029		
node01	free	20/40	188/188		
node02	free	1/40	60/188		
node03	free	3/40	180/188		
node04	free	3/40	180/188		
node05	free	3/40	180/188		
node06	free	3/40	180/188		
node07	free	3/40	180/188		
node08	free	2/40	120/188		
node09	free	0/40	0/188		
node10	job-busy	40/40	64/188		
node11	free	0/40	0/188		
node12	free	0/40	0/188		
node13	offline	0/40	0/188		
node14	free	0/40	0/188		
node15	free	0/40	0/188		
node16	free	0/40	0/188		
node17	free	0/40	0/188		
node18	free	0/40	0/188		
node19	free	0/40	0/188		
node20	free	0/40	0/188		
smp	free	40/72	500/3023		
catl	free	0/32	0/1009		

キューの選び方について2

- qstatや稼働状況を見てリソース利用量を調整し、空いている/即実行されるキューにジョブを投入しましょう
- 分割できるものは分割してアレイジョブにし、全体の処理時間を短縮しましょう

Delete jobs

```
$ qdel jobID
```

- Delete by job ID
- Only your job

```
$ qdel 5812
```

Select jobs

```
$ qselect [option]
```

<code>qselect -u <i>user_name</i></code>	Display job IDs (selected user)
<code>qselect -u <i>user_name</i> -q <i>queue_name</i></code>	Display job IDs (selected user and queue)
<code>qselect -s Q -u <i>user_name</i></code>	Display waiting job IDs (selected user)
<code>qselect -s R -u <i>user_name</i></code>	Display running job IDs (selected user)

Bulk deletion selected jobs

```
$ qdel `qselect option`
```

- qdel + qselect
- `` `` backquote → Replaced by the execution result of the command written in `` ``
- **Bulk delete all my jobs.**

```
$ qselect -u my_user_name  
$ qdel `qselect -u my_user_name`
```

.e file and .o file

- Standard ERROR output *job_name.ejobID*
- Standard output *job_name.ojobID*
- These are kept on the node where the job was run and sent to the qsubbed machine after the job is finished.
- For some reason these files may not be able to be sent.
- If you would like to see them, please email support@nibb.ac.jp with your jobID.

Practice2: Submit your job and Display job status

```
$ less diamond.sh
```

diamond.sh

```
#!/bin/bash
```

←①シェルの指定

```
#PBS -q small
```

```
#PBS -l ncpus=2
```

```
#PBS -l mem=6gb
```

←② qsubオプション

```
cd ${PBS_O_WORKDIR}
```

←③ジョブ投入ディレクトリへの移動

```
diamond blastp --threads ${NCPUS} --db /bio/db/diamond/swissprot/swissprot \
```

```
--outfmt 6 --query ./sce_prot.fasta --out ./sce2.tab --sensitive ←④実行するコマンド
```

② qsubオプション

small キュー

2CPUを確保

6GBメモリを確保

Practice2: Submit your job and Display job status

1. Read "diamond.sh" with less command
2. Submit job
3. Display your job status, display all job status
4. Display log message with tracejob

```
($ less diamond.sh)  
$ qsub diamond.sh  
$ qstat -u your_user_name  
$ qstat  
$ ls  
$ tracejob job_ID
```

Print log messages for a PBS job : **tracejob**

```
$ tracejob [option] jobID
```

- You can only see the logs on the same day.

```
resources_used.walltime=  
resources_used.mem=
```

- It shows the time required, total amount of memory used, etc.

```
$ tracejob 6278
```

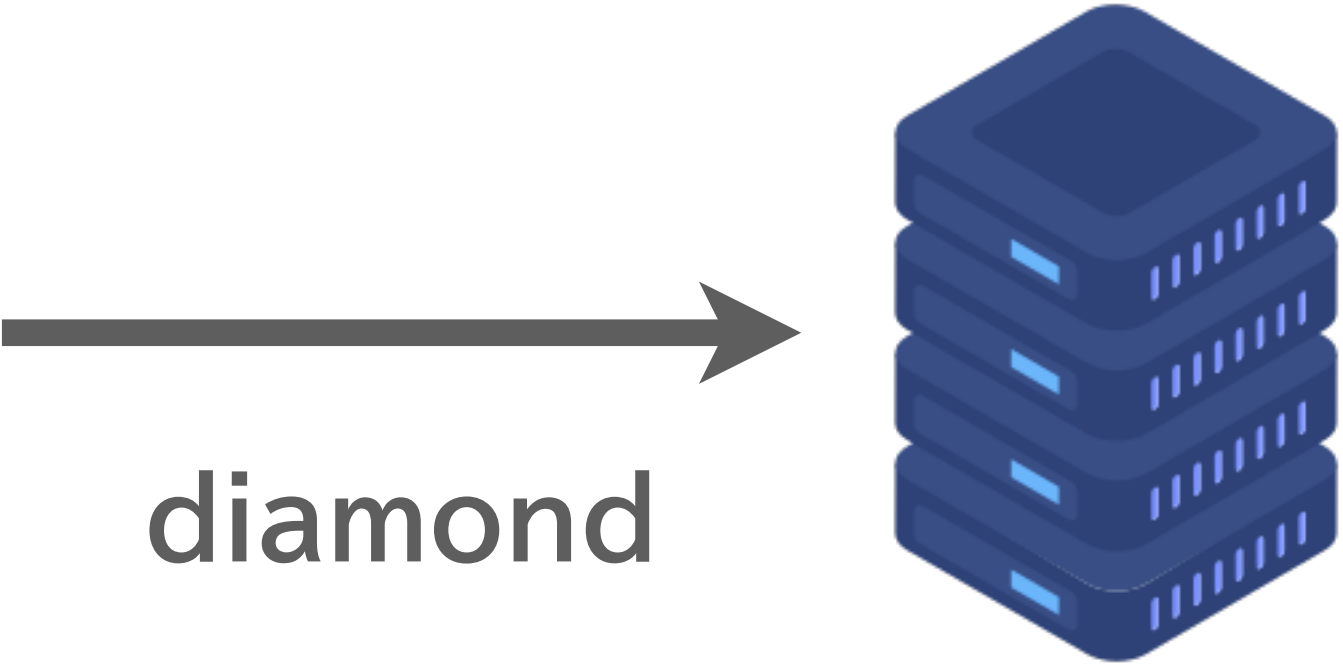
```
Job: 6278.bias5-adm
```

```
06/13/2018 11:26:30 L Considering job to run  
06/13/2018 11:26:30 S enqueueing into small, state 1 hop 1  
06/13/2018 11:26:30 S Job Queued at request of sgi@bias5-login, owner = sgi@bias5-login, job name = TB, queue = small  
06/13/2018 11:26:30 S Job Run at request of Scheduler@bias5-adm on exec_vnode (bias5-node10:mem=3145728kb:ncpus=4)  
06/13/2018 11:26:30 S Job Modified at request of Scheduler@bias5-adm  
06/13/2018 11:26:30 L Job run  
06/13/2018 11:36:30 S Obit received momhop:1 serverhop:1 state:4 substate:42  
06/13/2018 11:36:30 S Exit_status=0 resources_used.cput=00:00:00  
resources_used.mem=4000kb resources_used.ncpus=4 resources_used.vmem=357328kb resources_used.walltime=00:10:00
```

- tracejob can reveal the cause of an error.
- If you want to know the tracejob results of an old job, please email support@nibb.ac.jp with your jobID.

Array job

```
>spo:NP_001018179.1 hydroxymethylbilane synthase (predicted) [Schizosaccharomyce
MPSCTSFPIGTRKSKLAVIQSEI IREELEKHYPHLEFP IISRDTIGDEILSKALFEFKRQLAKSLWTRELEALLVTNQCR
ILVHSLKDLPEMPDGMVIACIPKRSCPLDAIVFKAGSHYKTVDLPPGSVVGTSIRRRALLARNFPHLRVFDIRGNVG
TRLAKLDAPDSQFDCLVLAAGLFRGLKDRIAQMLTAPFVYAVGQALAVEVRADDKEMIEMLKPLQHQETLYACLAE
RALMKRLQGGCAIPIGVQTDVLAISNSSYRISLLGTVLSADGLRAAFGNAEAVVSSEEEAEELGITVALALLKNGAGPIL
EEHQRSSDSEESLKNY
>spo:NP_001018181.1 poly(A) polymerase Cid14 [Schizosaccharomyces pombe 972h-]
MGKKSVSFNRRNYYKRNKNERTEPLRRIFKNDKPSKFKSKRKEKDKNSDAYDEMMLNNFTLLDQEEPVEIGSKSRND
NDSEGIRDKGGVEISNKNDPYIQFGKADPLEPLEKPDLPPEEAIKRGEP TILLGIPKREGRKTNPVHDKAVENNSDFIKFD
WNSDEDEDSVSNKSKNNE SLKSSKNEIPGFMQRGRFFHEANEKSDSNRKRKRQAYELDSQSCPWHRQYKVEREVSRI
FHQDILHFIDYITPTPEEHAVRKTLSRINQAVLQKWPDVSLYVFGSFETKLYLPTSDLDLVIISPEHHYRGTKKDMFVL
AHHLKKLKLASEVQVITANVPIIKFVDPLTKVHVDISFNQPGGLKTCVVNGFMKKYPALRPLVIIKHFLNMRALNEV
FLGGLSSYAIIVCLVVSFLQLHPRLSTGSMREEDNFGVLLLEFLELYGKQFYDVGIAVHNGGFYFSKMKMGWLKPNQPY
LLSIQDPVDFQNDVSKSSRGLLRVKATFANGFDLLTSKLYALASRIEREGVNRVKDFPSILSTILSVDEGVRQHREHMLK
CYKNNPVPLEPLVEVDALASIDVDKLPQDVGLQYVEDESDETDAAKDDLKVNES IETNGHENFQKQALTSTGEQSS
SNSRANPSKLFNISSDSEDEVPIIEDTTASDEESRAKKIRKRF
>spo:NP_001018183.2 pre-ribosomal protein Loc1 (predicted) [Schizosaccharomyces
MVKKSKPKNQIRVEDLDLPLKLNSTSKNPQTKIQKGGKKGKIFAETKDDLQNILNQVTEYELDDKIKSKLQVAHEREAVFS
KQSDRKISNNKADKKTGRKNEK
>spo:NP_001018187.2 dual specificity protein kinase Lkh1 [Schizosaccharomyces
MHSKRRRNHAPDWQDFYKNGVPQEVIVIEDSASPRLTPLNPPFVSHVQLQSFVPPQPPSSSPSTGTVAVPINGANAV
YPSTNSVSLPQSYDPWLDANGVVPLPHDVASHPSYMQSPTS YHACSNNQSPFPHSHHPLHNPLPVSCQPVLRRPPVPQ
VPSHWYPVSLPSPNLPHQPI SKPPVIPNLPKLQVHPNRLPHPIHNHPYSSPTSYPPLCPATYCPSNPPQLAPATAIAPS
SQSSQHKS VNYSVTPSSINNHTAVPLSPTLAVWLPMTQPTFQPPSANVYQPASNANQVITPVVISDYRPPKRRKRAAWPP
YKKVDRVNVVVDHTAFDPSTFDDDDGHYKVPVNSK FANRYTVVRL LGHGTGFKVIQCYDQSTGRHCAIKVTRAI PKYR
EASLIELRVLQTI AHS DPTNENKCIQLRDYFDYRKHCIVTDLFGWSVDFLKNNNYIPFPLKHIQMLSQQLFKSVAF LH
SLGLVHTDLK PENVLLVSNASRTIRLPYRNYSQKVLNSCEIRLIDFGSATFEDEYHSSVVSTRHYRAPEIILGLGWSYPC
DVWSIGCILVELFTGQALFQTHEDESEHL CMMEKILGPFDRNMISRSRSTSQRFFKSDGKVRYPLSNTPKKSINYLQSLQT
LEQIFAVSSPEVALLDLKLVFVYDPKRRITAKEALWHPFFTQPISSNL
>spo:NP_001018191.1 homologous-pairing protein 2 [Schizosaccharomyces pombe 972h
MAKAKEVKAKPIKGEAEKLVYEYLRKTNRPYSATDV SANLKNVSKQVAQKALEQLRDTGLIHGKLYGKQSVFVCLQDD
LAAATPEELAEMEKQIQELKDEVS VVKTYKEKCIELQALNNSLSPA EIREKIQSIDKEIEETSSKLES LRNGTVKQISK
EAMQKTDKNYDFAKKGFSNRKMFYDLWHLITDSLENPKQLWEKLGFE TEGPIDLN
>spo:NP_001018193.1 Rho family GTPase Rho3 [Schizosaccharomyces pombe 972h-]
MSSCFGSKKKPIYRKIVILGDGAAGKTSLLNVFTKGYFPQVYEPTIFENYIHDIFVDGNSIELSLWDTAGQEEYDQLRSL
SYSDTHVIMICFAVDSRDSLENVITKWLPEVSSNCPGVKLVLVALKCDLRGADEEQVDHSKIIDYEEGLAAAKKINAVRY
LECSAKLNRGVNEAFTEAARVALAAQPRGTDGADESHGTGCIIA
>spo:NP_001018195.1 UBA domain protein Mud1 [Schizosaccharomyces pombe 972h-]
MNNLTPENIRQTILATPFLNLRIRTEFPQLAAVLNDPNAFATTWQSINASQLLQIPSSYSMGMPSFSEDDLFDVEVQRR
IEEQIRQNAV TENMQSAIENHPEVFGQVYMLFVNVEINGHKVKAFVDSGAQATILSADCAEKCGLTRLLDTRFQGVAKGV
GMAKILGCVHSAPLKIGDLYLPCRFTVIEGRDVMMLGLDMLRRYQACIDLENNVLR IHGKEIPFLGESEIPKLLANVEP
SANAHGLGIEPASKASASSPNPQSGTRLGTKESVAPNNEGSSNPPSLVNPPTDPGLNSKIAQLVSMGFDPLEAAQALDAA
NGDL DVAASFLL
>spo:NP_001018218.1 SMN family protein Yip12 [Schizosaccharomyces pombe 972h-]
MPSKRKRNPLOYQTSGLDEETNQRSAFPQIDNNSASESLEYDIPLDGLDYLATVREEARKLVFVAARREPETRETIPL
RKLEIEAGKKS FDFPFLRYLLNIIDKEGERLEQYMESSLSDASILPKNLQQRVYIEHKAPCWA IAVVDLATVLEI LESL
SSWLEKDAIDLQSQWIFCFYKLPPELLNGEDISTLRSVLKSLRSTHSPALQMSASALQAVLVYRYGQKDLFQT
>spo:NP_001018219.2 ubiquitin-protein ligase E3 (predicted) [Schizosaccharomyce
MDSPLLEVLQVQVEKLI SPSLRFILAYFTHRYPRFLLRAYNSFDGIYLLVKLLLEKSQLKKNATSVERRFQLKRVIA
VRDSSIIAEFPQESSESATSLNGIDVLKKLFLTYCIPYLLKCESLTTVKENHTAVSILSLQARDKQK GALSVFYSKIKI
LLVRLKILHFVFR LIRKSN TYLQWLYLLYALGKTPYTNLADHILRQRVIY NVENIHSRKLISTREKSSLLTSIADHSM
EGFLIIQLIDWWQSNNYESH LKKEVAFTELAPPKLPFEINVSTTDICKICGEKIKNPAVLSTGFVFCYPCIQVWLQRH
PFKCPVTNLELSRKGESFWRLMI
```



my.fa

Array job

A single script file performs the same task independently

```
>spo:NP_001018181.1 poly(A) polymerase Cid14 [Schizosaccharomyces pombe 972h-]  
MGKKSVSFNRNNYKRRKNERTEPLRRIFKNDKPSKFKSKRKEKDKNSDAYDEMLLNNNFLLDQEEPVEIGSKKSRND  
NDSEGIRDKGGVEISNKNDPYIQFGKADPLEPLEKPDLPPEEAIKRGEPTILLGIPKREGRKTNPVHDKAVENNSDFIKFD  
WNSDEDEDSVSNKSKNNESLKKSSKNEIPGFMQRGRFHEANEKSDSNRKRKRQAYELDSQSCPWHRQYKVEREVSRI  
FHQDILHFIDYITPTPEEHAVRKTLSRINQAVLQKWPDVSLYVFGSFETKLYLPTSDDLVIISPEHHYRGTKKDMFVL  
AHHLKKLKLASEVQVITANVPIIKFVDPLTKVHVDISFNQPGGLKTCVNVNGFMKKYPALRPLVIIKHFLNMRALNEV  
FLGGLSSYAIVCLVVSFLQLHPRSTGSMREEDNFGVLLLEFLELYGKQFYDAVGIAVHNGGFYFSKMKMGWLKPNQPY  
LLSIQDPVDFQNDVSKSSRGLLRVKATFANGFDLLTSKLYALASRIEREGVNRVKDFPSILSTILSVDEGVRQHREHMLK  
CYKNNPVPLEPLVEVDALASIDVDKLPQDVGLQYVEDESDSDTDAKDDLKFNES IETNGHENFQKQALTSTGEOSS  
SNSRANPSKLFNISSDSEDEVPIIEDTTASDEESRAKKIRKRF
```

my.1.fa

```
>spo:NP_001018179.1 hydroxymethylbilane synthase (predicted) [Schizosaccharomyce  
MPSCTSFPIGTRKSKLAVIQSEIIREELEKHYPHLEFPIISRDTIGDEILSKALFEFKRQLAKSLWTRELEALLVTNOCR  
ILVHSLKLDLPSEMPDGMVIACIPKRSCPLDAIVFKAGSHYKTVADLPPGSVVGTSIRRRALLARNFPHLRFVDIRGNVG  
TRLAKLDAPDSQFDCVLAAAGLFRGLKDRIAQMLTAPFVYAVGQALAVEVRADDKEMIEMKPLQHQTLYACLAE  
RALMKRLQGGCAIPIGVQTDVLAISNSSYRISLLGTVLSADGLRAAFGNAEAVVSSEEEAEELGITVALALLKNGAGPIL  
EEHQSSDSEESLKNY  
>spo:NP_001018183.2 pre-ribosomal protein Loc1 (predicted) [Schizosaccharomyces  
MVVKKSKPKNQIRVEDLDLPLKLNSTKNPQTKIQKGGKKKGIKIFAETKDDLQNILNQVTYELDDKIKSKLQVAHEREAVFS  
KQSDRKISNNKADKKTGRKNEKK  
>spo:NP_001018187.2 dual specificity protein kinase Lkh1 [Schizosaccharomyces
```

my.2.fa

```
>spo:NP_001018187.2 dual specificity protein kinase Lkh1 [Schizosaccharomyces  
MHSLKRRRNHAPDWQDFYKNGVQEVIVIEDSASPRLTPNLPPFVSVHQLQSFVPPQPPSSSSPSTTGTVAVPINGANAV  
YPSTNSVSLPQSYDPWLDANGVVPLPHDVASHPSYMQSPTSYHACSNNQSPFPHSHHPPLHNPLPVSCQPVLRRPPVPQ  
VP SHWYPVSLPSPNLPHQPI SKPPVI PNLPKLQVHPNRLPHPIHNHPYSSPTSYPPLCPATYCPNPPQLAPATAIAPS  
SQSSQHKS VNYSVTPSSINNHTAVPLSPTLAVWLPMTQPTFPQPSANVYQPASNANQVITPVVISIDYRPPKRRKRAAWPP  
YKKVDRVNVVVDHTTAFDPSTFDDDDGHYKVVVNSKFNRYTVVRLGHGTFGKVIQCYDQSTGRHCAIKVTRAI PKYR  
EASLIELRVLQTIASHDPTNENKCIQLRDYFDYRKHICIVTDLFGWSVDFLKNNNYIPFPLKHIQMLSQQLFKSVAFLH  
SLGLVHTDLKPENVLVSNASRTIRLPYRNYSQKVLNSCEIRLIDFGSATFEDEYHSSVVSTRHYRAPEIILGLGWSYPC  
DVWSIGCILVELFTGQALFQTHEDSEHLCMMEKILGPFDRNMISRSSRTSQRFFKSDGKVRYP LSNTPKKSINYLQSLQT  
LEQIFAVSSPEVALLD LLLKVFVYDPKRRITAKEALWHPFFTQPISSNL
```

my.3.fa

```
>spo:NP_001018191.1 homologous-pairing protein 2 [Schizosaccharomyces pombe 972h  
MAKAKEVKAKPIKGEAEKLVYEYLKRTNRPYSATDVSANLKNVSVKQVAQKALEQLRDTGLIHGKLYGKQSVFVCLQDD  
LAAATPEELAEEMEKIQELKDEVSVVKTYKEKCIELQALNNSLSPA EIREKIQSIDKEIEETSSKLESLRNGTVKQISK  
EAMQKTDKNYDFAKKGF SNRKKMFYDLWHLITDLENPKQLWEKLGFEFTEGPIDLN  
>spo:NP_001018195.1 UBA domain protein Mud1 [Schizosaccharomyces pombe 972h-]  
MNNLTPENIRQITILATPFLNLRIRTEFPQLAAVLNDPNAFATTWQ SINASQLLQIPSSYSMGMPSFSEDDLFDVEVQRR  
IEEQIRQNAV TENMQSAIENHPEVFGQVYMLFVNVEINGHKVKAFVDSGAQATILSADCAEKCGLTRLLDTRFQGVAKGV  
GMAKILGCVHSAPLKI GDLYLPCRFTVIEGRDVMLLGLDMLRRYQACIDLENNVLR IHGKEIPFLGESEIPKLLANVEP  
SANAHGLGIEPASKASASSPNPQSGTRLGTKESVAPNNEGSSNPPSLVNPPTDPGLNSKIAQLVSMGFDPLEAAQALDAA  
NGDLLVAASFLL
```

my.4.fa

diamond



Array job

- qsub option `-J X-Y[:Z]`
- `X` is the 1st index, `Y` is the upper bound on the indices and `Z` is the stepping factor
- `index => ${PBS_ARRAY_INDEX}`

```
#!/bin/sh
#PBS -J 1-100
#PBS -l ncpus=8
cd ${PBS_O_WORKDIR}

diamond blastx --threads ${NCPUS} --db /bio/db/diamond/nr/nr --outfmt 6 \
--query ./my.${PBS_ARRAY_INDEX}.fa --out ./out.${PBS_ARRAY_INDEX}.tab --sensitive
```

- Run diamond 100 times independently. 100 jobs are submitted.
- Need to have a continuous numbered file.

Array job on qstat : *jobID[]*

- Display only array job (`-J`)

```
$ qstat -J
```

Job id	Name	User	Time Use	S	Queue
5162[].bias5-adm	diaarray	my_name	0	B	small

- Display array job with subjobs (`-t`)

```
$ qstat -t
```

Job id	Name	User	Time Use	S	Queue
5162[].bias5-adm	diaarray	my_name	0	B	small
5162[1].bias5-adm	diaarray	my_name	00:00:00	R	small
5162[2].bias5-adm	diaarray	my_name	00:00:00	R	small
5162[3].bias5-adm	diaarray	my_name	00:00:00	R	small

Use Modules on PBS jobs

- Describe operating conditions in job scripts

```
source /etc/profile.d/modules.sh  
module load program_name/version
```

```
#!/usr/bin/bash  
#PBS -q blast  
#PBS -l ncpus=40  
#PBS -l mem=188GB  
source /etc/profile.d/modules.sh  
module load Trinityrnaseq/2.14.0  
cd ${PBS_O_WORKDIR}  
Trinity --verbose --max_memory 188G --seqType fq \  
  --samples_file rnaseq_data.txt --CPU ${NCPUS} \  
  --output trinity_out --min_kmer_cov 2
```

qsub under various conditions

- conda

```
#!/bin/bash
#PBS -q medium
source ${HOME}/miniconda3/etc/profile.d/conda.sh
conda activate base
cd ${PBS_O_WORKDIR}
python3 my_script.py
conda deactivate
```

←①conda.shをロード

←②conda環境を起動

←③conda環境を終了

- singularity

```
#!/bin/bash
#PBS -q small
cd ${PBS_O_WORKDIR}
source /etc/profile.d/modules.sh
module load singularity
singularity exec /bio/container/BUSCO/5.1.3/busco_5.1.3--pyhdfd78af_0 \
busco -i my.fa -m tran -o b_result -l eukaryota_odb10
```

When the job fails...

PBS Error:

- status Eqw
- Not executed with status Q

`qstat -f` or `tracejob`

```
$ qstat -f jobID  
$ tracejob jobID
```

Program Error:

See `.e` file and `.o` file

”I cannot find `.e` file and `.o` file...”

Please email support@nibb.ac.jp with your jobID.

PBS help

```
$ man command_name
```

```
$ man qsub
```

```
$ man qstat
```

```
$ man tracejob
```

wiki

- BIAS wiki

<https://biaswiki.nibb.ac.jp/menu/index.php>

Advanced information

- MPI, mpicc
- AlphaFold2 installed



The screenshot shows the homepage of the BIAS wiki. At the top, there are navigation links for 'メインページ' (Main Page) and '議論' (Discussion). The main title is '生物情報解析システム' (Bioinformatics Analysis System). Below the title, the NIBB logo is displayed, followed by the text '基礎生物学研究所 生物機能解析センター 情報管理解析室'. A table of contents is provided, listing sections: 1 概要 (Overview), 2 障害・障害復旧情報 (Incidents and Restoration Information), 3 停止予定 (Stoppage Schedule), 4 システム構成 (System Configuration), 5 利用手続き (Usage Procedures), and 6 使い方 (Usage). The '概要' section is expanded, showing a list of updates and system specifications. A warning box indicates that users must have 'UNIXコマンドの知識' (UNIX command knowledge) to use the system. The '概要' section includes the following information:

- 生物情報解析システム Version5 は、2018年6月から運用中です。
 - 「共有メモリ型計算サーバ」メインメモリ 3TB (メモリが多く必要な解析に)
 - 「分散処理用計算機クラスタ」800コア (CPUを沢山使って分散処理したい解析に)
- 追加の計算機
 - 「共有メモリ型計算サーバ」メインメモリ 4TB
 - 「共有メモリ型計算サーバ」メインメモリ 520GB 2台
- 基礎生物学研究所内の方はもちろん、所外の研究者の方にもお使いいただけます。ユーザーアカウント
- 利用資格などは内規をご覧ください。 [生物情報解析システム:利用内規](#)

システムをお使いいただくには「UNIXコマンドの知識」があることが前提になります。デスクトップ

- 「利用者説明会資料」 2018年6月開催 (アクセスにはVPN接続が必要です)
 - 生物情報解析システム Ver. 5 利用者説明会