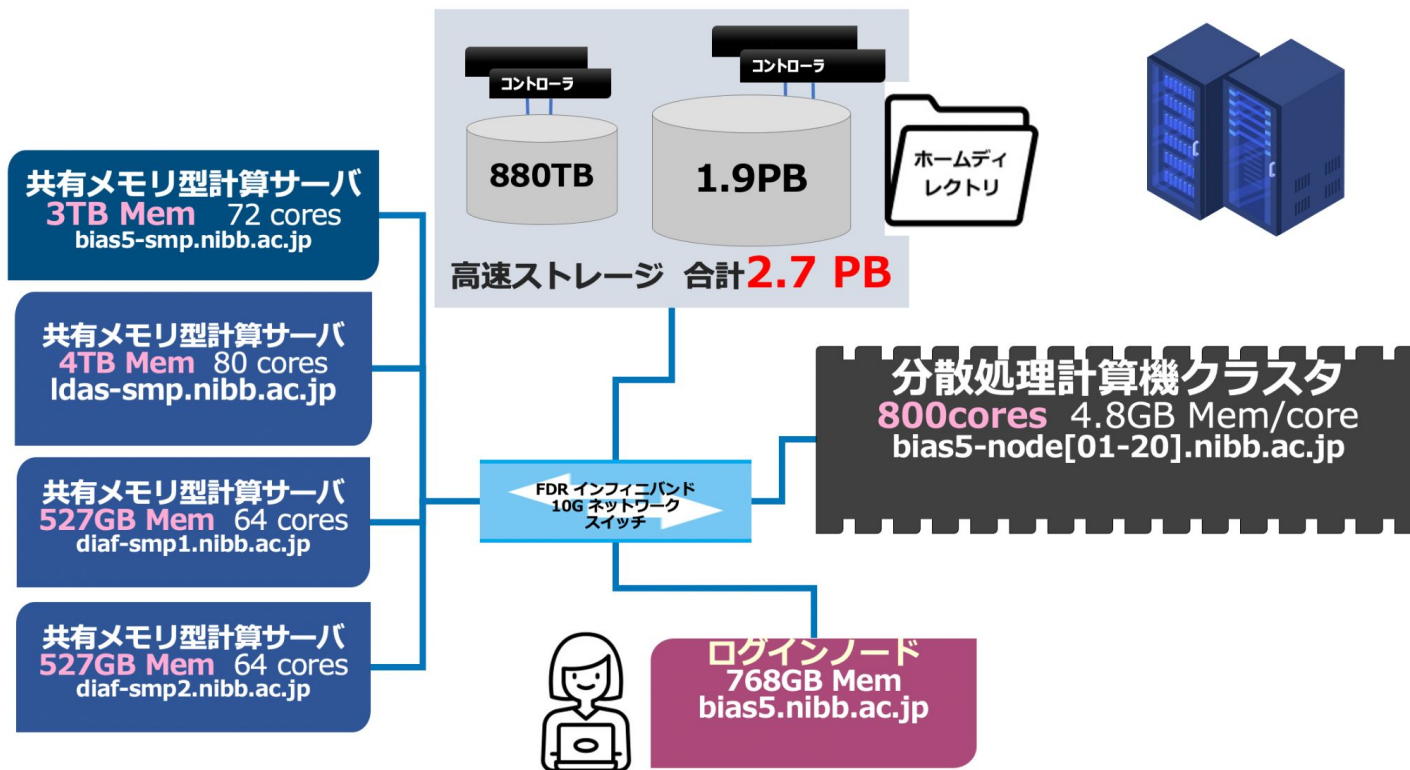


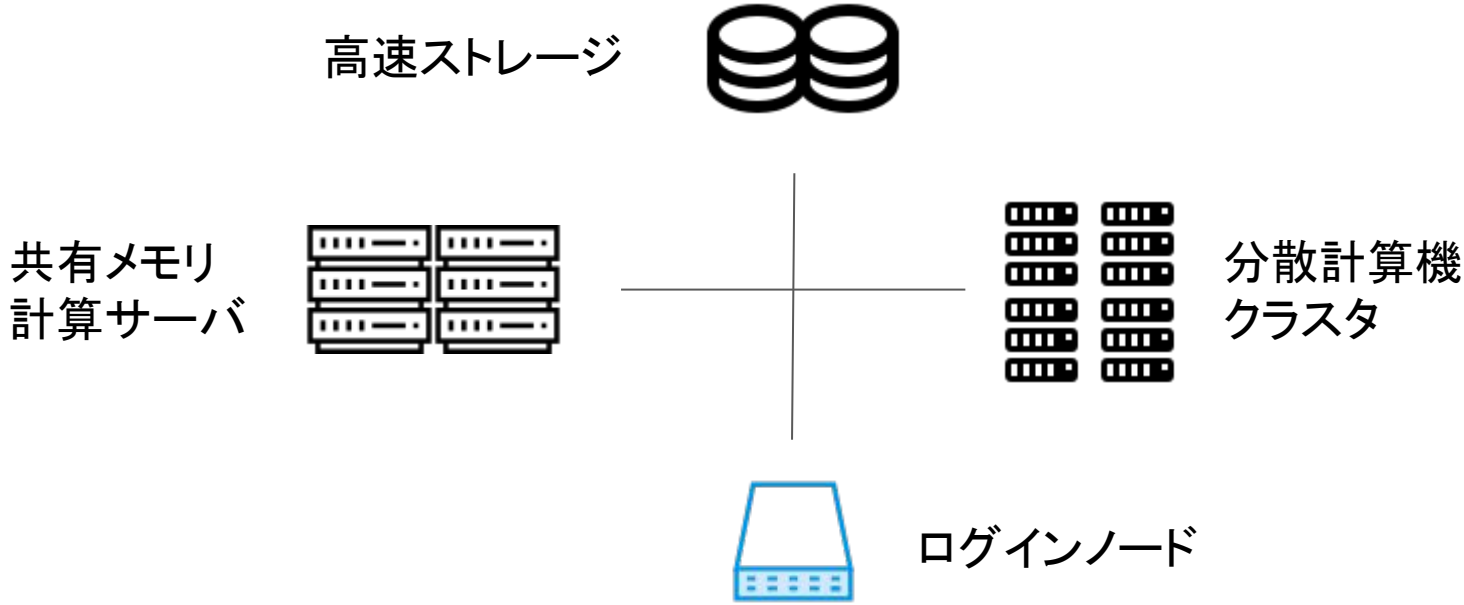
生物情報解析システム (BIAS5) の概要

データ統合解析室 中村

生物情報解析システム (BIAS5) の構成



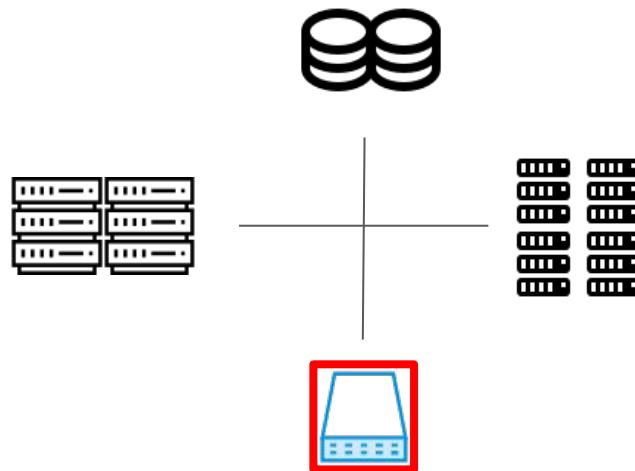
生物情報解析システム(BIAS5)の構成略図



ログインノード bias5-login

ユーザーがSSHコマンドで最初にログインするサーバです。

自分のファイルを見たり書いたりすることができます。ファイルの実体は高速ストレージ上にあります。



プログラムを実行する場合、ログインノード上で実行することは避けてください。
後に説明するジョブ管理システム(PBS)を使用し他の計算サーバに実行させてください。

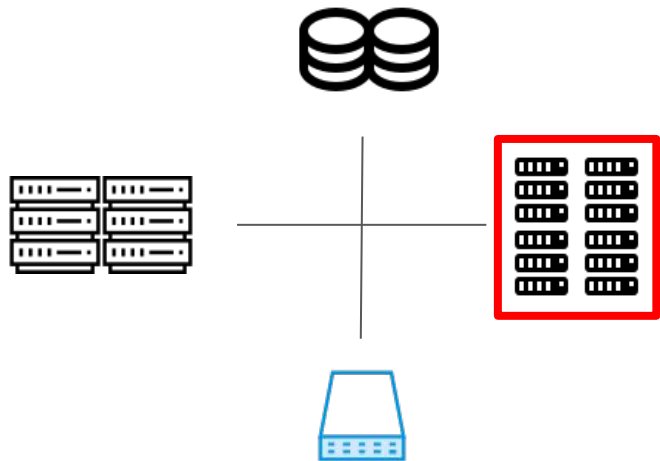
分散処理用計算機クラスタ bias5-node[01-20]

大規模な並列計算を行うサーバ群です。
ユーザーが直接ログインすることはありません。

使用するにはPBSを用いてジョブを投げること
になります。

CPUでの処理が多であるプログラムを使用
する場合におすすめです。

各ノードのアーカイブ容量はそれほどありません。
ログをノードに残す場合はご注意ください。



CPU: Intel Xeon Gold 6140
(2.3 GHz) 72 cores /
Memory : 3TB

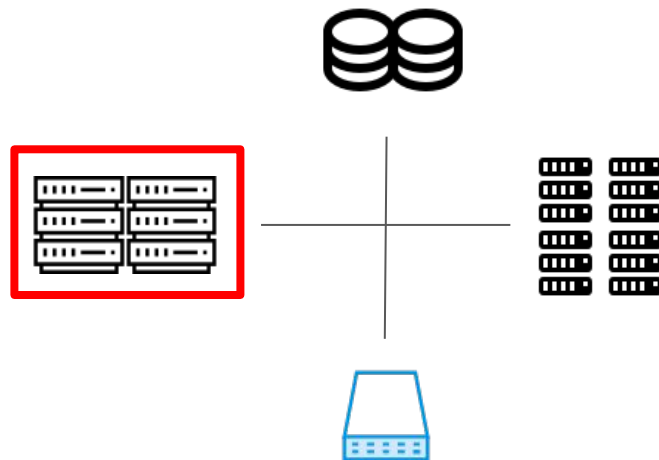
共有メモリ計算サーバ bias5-smp, ldas-smp

大規模なメモリを使用する計算を行うサーバ群です。ユーザーが直接ログインすることはありません。

使用するにはPBSを用いてジョブを投げることになります。

大きめのファイルを処理するようなプログラムを使用する場合におすすめです。

使用するメモリの大きさはPBSで指定できます。



bias5-smp

CPU: Intel Xeon (2.3 GHz) 72 cores

Memory : 3TB

ldas-smp

CPU: Intel Xeon (2.4 GHz) 80 cores

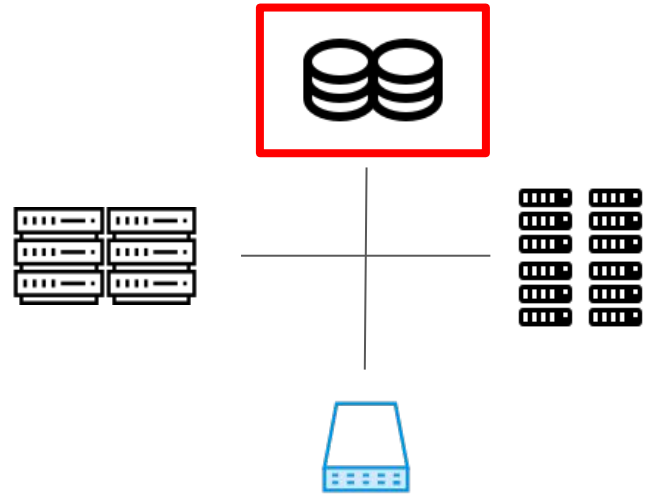
Memory: 4TB

高速ストレージシステム

ユーザーのホームディレクトリを始めとした、データを保存するストレージ群です。ユーザーが直接ログインすることではなく、ログインノードに入った時点でマウントされ使用可能になっています。

ストレージシステムはその用途別にファイルパスで区別され、それぞれにクォータ(容量制限)があります。(後述)

容量、クォータの変更についても受け付けております。データ統合解析室までお問い合わせください。



全体実効容量

880TB(A)+ 1.9PB (B)

基本的な使い方

- まずログインノード bias5.nibb.ac.jp にログインする。
- ログインノードではプログラム作成などリソースを消費しない処理のみを行い、ジョブの実行は必ずPBSを通して行う。
- 分散処理クラスタ、および共有メモリサーバを使うには、それぞれの専用のキューにジョブをサブミットする。
- 利用法に関する詳しい情報は、専用ページ <https://biaswiki.nibb.ac.jp/menu> で確認する。

bias5へのログイン

bias5へのログインは、sshコマンドを使用して行います。

所内ネットワークのユーザーは、sshコマンドのみでログインできます。

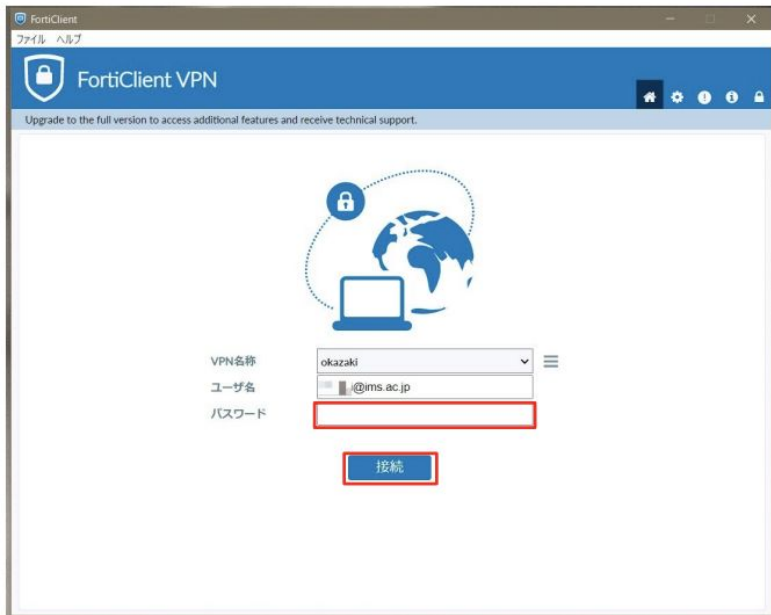
所外のユーザーは、VPNを使用して所内ネットワークに所属した後、sshコマンドを使用してログインを行います。

(所内のユーザーでも、所外からVPNを使用することによってbias5にログインを行うことができます)

VPNに関しては専用のクライアントを使用します。

VPNクライアント (FortiVPN / FortiCleint)

所外の方：
FortiVPN を使用します。



基生研の方：
所内ネットワークであればVPNを使用する必要はありません。
外部から接続するにはFortiClientを使用します。
(SAML認証が必要です)



bias5へのログイン

sshコマンドを使用してbias5にログインするには、以下のようにコマンドを実行します。

```
$ ssh user-account@bias5.nibb.ac.jp
```

“user-account”には、bias5上で使うアカウント名を入れます。
コマンドの実行後、パスワードを要求されますので、対応するパスワードを入力してください。

実行例:

```
$ ssh unixr@bias5.nibb.ac.jp
unixr@bias5.nibb.ac.jp's password:
Last login: xxx xxx xx xx:xx:xx 20xx from xxxxx.nibb.ac.jp
[unixr@bias5-login ~]$
```

bias5からのログアウト

bias5からログアウトするには、exitコマンドを実行します。

```
$ exit
```

ログアウト後は、bias5にログインする以前の端末に戻ります。

実行例:

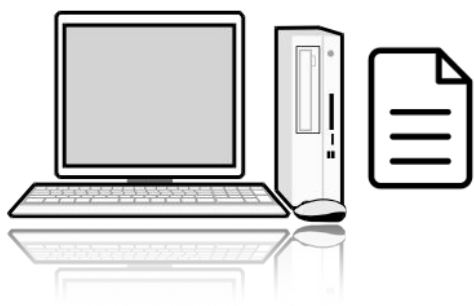
```
[unixr@bias5-login ~]$ exit  
logout
```

```
Connection to bias5.nibb.ac.jp closed. ←接続が解除されたメッセージが出る。
```

```
$
```

リモートファイルコピー (scp)

自分のローカル端末から bias5 にファイルを送信する場合、scp コマンドを使用します。
これによりネットワーク経由でファイルをやり取りすることができます。



bias5.nibb.ac.jp

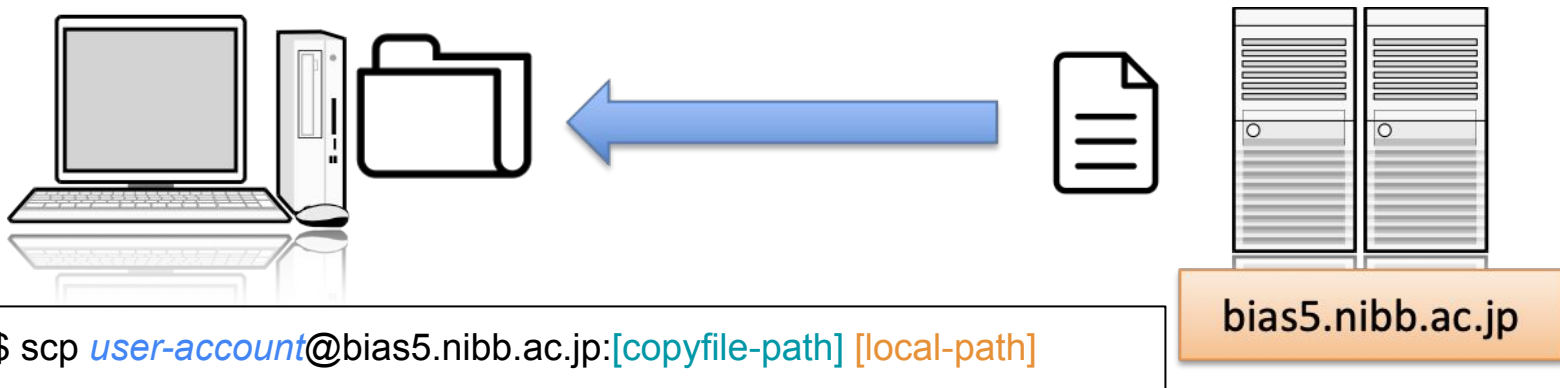
```
$ scp [copyfile-path] user-account@bias5.nibb.ac.jp:[file-path]
```

例: `$ scp file1.dat user-account@bias5.nibb.ac.jp:temp/data/`

上記コマンドを実行すると、ローカルのカレントディレクトリにある file1.dat というファイルを、bias5 のホームディレクトリの temp/data/file1.dat にコピーする。

リモートファイルコピー (scp)

先ほどとは逆に、bias5から自分のローカルにファイルを送信することも、scpコマンドを使用します。どこからscpコマンドを実行するのかということと、ファイルのコピー元 / コピー先の記載に注意してください。



例: `$ scp user-account@bias5.nibb.ac.jp:temp/data/result.dat ~/Desktop/result`

上記コマンドを実行すると、bias5のホームディレクトリの temp/data/result.dat というファイルを、自分のローカルの ~/Desktop/result/result.dat にコピーする。

クォータについて

- ホームディレクトリにはクォータ(制限)がかけられています。デフォルトは3TBです。
- 一時的に大量のディスクを使う場合はクォータがかかっていないスクラッチ領域 (/scratch/ユーザ名) を使用してください。ただし、スクラッチ領域のファイルは30日後に消去されます。
- バックアップやアーカイブとして長期的に保存したいファイルはSave領域に保存してください。
- 現在の使用量を確認したい場合、以下のコマンドを実行してください。

```
$ mmlsquota gpfsA:home
```

容量について

名称	パス	1ユーザーあたりの容量	主な用途
ホームディレクトリ	<code>/home/<i>user_name</i></code>	3TB(最大6TB)	ログイン時のホーム
スクラッチ領域	<code>/scratch/<i>user_name</i></code>	制限なし	一時作業用(30日後に自動的に消去)
Save領域	<code>/home/<i>user_name</i>/save</code>	6TB	データ保管用

ディスク領域拡張申請により、ホームディレクトリを6TBまで増やすことができます。

ジョブ管理システム (PBS)

ジョブ管理システムとは、多数の処理要求(ジョブ)を受け付けて管理し、CPUやメモリなどの計算資源を適切に割り当てて順次実行させる仕組みです。

bias5では、ジョブ管理システムとして PBS (Portable Batch System)が導入されています。

基本的に、すべての計算はPBSを通して実行します。

ログインノードでの実行は行わないようにしてください。

計算の規模や種類によって複数のキューが用意されています。これらのキューを適切に選ぶことで、作業を効率よく進めることができます。

(PBSに関しては3つ目の講義で詳しく行います)

(参考)キューの設定

キュー	small	medium	large	blast	smps	smpm	smpi
ジョブの特徴	短時間・並列多	中規模	長時間	BLAST nr/ntのみ	中メモリ	大メモリ	最大メモリ
使用	計算ノード (01-20)	計算ノード (01-20)	計算ノード (01-20)	計算ノード (01-20)	共有メモリ サーバ	共有メモリ サーバ	共有メモリ サーバ
最大実行時間	6時間	72時間	無制限	72時間	無制限	無制限	無制限

より詳しくは→ https://biaswiki.nibb.ac.jp/menu/index.php/ジョブ管理システム_PBSを用いた計算機の利用

BIAS5で利用可能なバイオ関連ソフトウェア(一部)

次世代シーケンサー解析用ソフトウェア

マッピング, RNA-Seq解析, アセンブラ

Bowtie2, HISAT2, BWA,
StringTie, Trinity-RNAseq, salmon,
Velvet, ABySS, AllPaths-LG

ユーティリティ

samtools, bamtools, BEDtools, cutadapt,
SRA-toolkit

遺伝子予測 GeneMark, Augustus, BRAKER

ゲノムアライメント lastz, MUMmer, BLAT

マルチプルアライメント
ClustalW, Muscle, MAFFT

系統樹解析 Phylip, PhyML, MrBayes

モチーフ解析 InterProScan, HMMER, MEME

データベース検索 DBGET

統合配列解析 EMBOSS, BLAST+, Diamond,

より詳しくは → <https://biaswiki.nibb.ac.jp/menu/index.php/ソフトウェア>

利用可能なバイオ関連データベース(一部)

- BLAST用データベースへのパス: `/bio/db/blast/db`
- diamond用データベースへのパス: `/bio/db/diamond/db`
- フラットファイルへのパス: `/bio/db/ideas`

データベース名	説明	フォーマット	更新
Genbank, EMBL	核酸塩基配列	DBGET	定期
gss, HTGS, dbsts, patnt, env_nt	核酸塩基配列	FASTA, BLAST	定期
RefSeq(Nuc+Pep)	核酸塩基配列+タンパク質アミノ酸配列	DBGET	日々
NCBI nr-nt	非冗長核酸塩基配列	FASTA, BLAST	定期

より詳しくは → <https://biaswiki.nibb.ac.jp/menu/index.php/分子生物学データベース>

ログインノード

ホスト名 : bias5-login (ssh時には bias5.nibb.ac.jp でログイン可能)

機種 : HP ProLiant DL360 Gen10

CPU: Intel Xeon Gold 6138 (2.1 GHz) 32 cores Memory: 768 GB

主な用途: ユーザーログイン / ジョブの実行 / プログラムの作成

ログインして利用するが、プログラムはqsub経由で実行



分散処理用計算機クラスター

ホスト名 bias5-node[01-20] Total: 20 nodes, 800 cores

機種: HPE Apollo r2800

CPU: Intel Xeon Gold 6138 (2.0GHz) 40 cores/node

Memory: 192GB/node (4.8GB/core)

主な用途:

大規模な並列計算

直接ログインをせず qsub 経由で利用

キュー: small, medium, large



共有メモリ計算サーバ

ホスト名 bias5-smp

機種: HP ProLiant DL560 Gen10

CPU: Intel Xeon Gold 6140 (2.3 GHz) 72 cores / Memory : 3TB

主な用途:

大きなメモリを使う計算

直接ログインせずqsub 経由で利用

キュー:smps, smpm, smp1



高速ストレージシステム

DDN SFA7700X

実効容量 880TB(A)+ 1.9PB (B)

並列分散ファイルシステム / GPFS

主な用途:

ホームディレクトリ領域 (クォータ制限あり)

スクラッチ領域 (クォータ制限なし)

共通データベース プロジェクト領域 (制限に関しては応相談)



大容量データ解析システム(LDAS)

共有メモリ計算サーバ

ホスト名 ldas-smp

HP ProLiant DL980 G7

CPU: Intel Xeon (2.4 GHz) 80 cores

Memory: 4TB (50GB/core)

主な用途:

大きなメモリを使う計算

直接ログインをせず qsub 経由で利用

キュー:

smps, smpm, smp1