TSUBAME利用講習会

平成22年度版

東京工業大学
学術国際情報センター

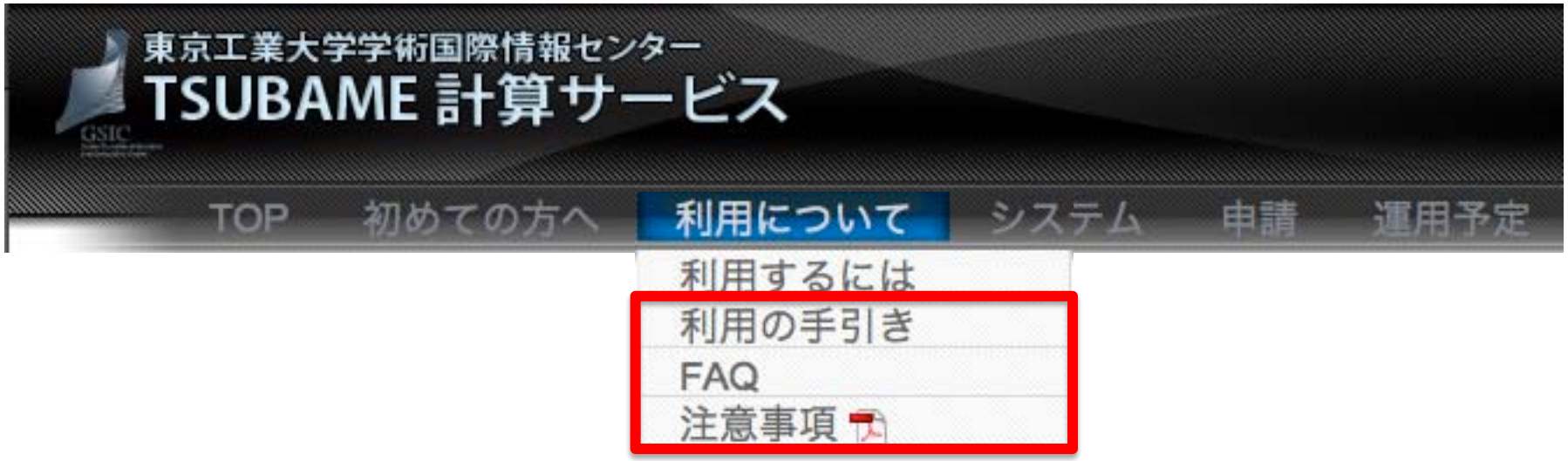
TSUBAME利用の情報源

GSIC研究用計算機システム>お知らせWeb

- http://www.gsic.titech.ac.jp/~ccwww/

TSUBAME利用の情報源

- http://www.gsic.titech.ac.jp/~ccwww/



- 「利用の手引き」「FAQ」「注意事項」は必ず目を通してください。

FAQおよび利用の手引き

- FAQ
  - センターシステム F.A.Q.
- 利用の手引
  - TSUBAME Grid Cluster 利用の手引
    - システム利用の手引
      - TSUBAME利用の手引き　日本語版、英語版
      - 高速バックアップ利用の手引き
    - アプリケーション利用の手引き
      - Fortranコンパイラ利用の手引き (html版)
      - Gaussian利用の手引き (html版)
      - Gaussian Linda利用の手引き (html版)

各文書のslaはinno, pinnoと読み替えてください

0. 質問窓口

- 先端研究施設共用促進事業トライアルユースの利用者は

  kyodo@gsic.titech.ac.jp
- 共同利用制度の有償利用の利用者は

  tsubame@gsic.titech.ac.jp

です。

- sodan@cc.titech.ac.jpは学内向けの相談窓口です。利用なさらないでください。

0.1 利用可能アプリケーション

- 商用アプリケーションは、PGIコンパイラとIntelコンパイラとGaussian 03/09のみ利用可能
- いわゆるFreewareは自由に利用できます。
  - サポート有りFreeware(/usr/apps/free*)
    - GSICでのSEによる一部サポートがあるものです。
  - サポート無しFreeware(/usr/apps/nosupport*)
    - GSICでのSEによるサポートがありません。

/usr/apps/free*

- free:

  acml3.0.0 condor fftw-3.1.2 gromacs-3.3.1 jdk1.5.0_11 nwchem4.7 povray3.5C utchem2004beta-a

  acml3.5.0 condor.pkg gamess ImageMagick6.2.7 lib openmpi1.1a2 povray3.6.1 xmgr4.1.2

  acml3.6.0 fftw-2.1.5 gromacs intel mpich2-1.0.4 povray sunstudio

  fftw3.1 gromacs3.3 j2sdk1.4.2_08 netlib povray3.1g tinker4.2
- free10:

  fftw gamess gromacs ImageMagick nwchem povray tinker utchem

/usr/apps/nosupport*

- nosupport:

  gnuplot-4.2.3 libxml2-2.6.28 php-4.4.7 php-5.2.2 python2.5.2 R-2.5.0 R-2.5.0_mpi R2.5.0_vltmpi R2.7.2_vltmpi tgc_formhis

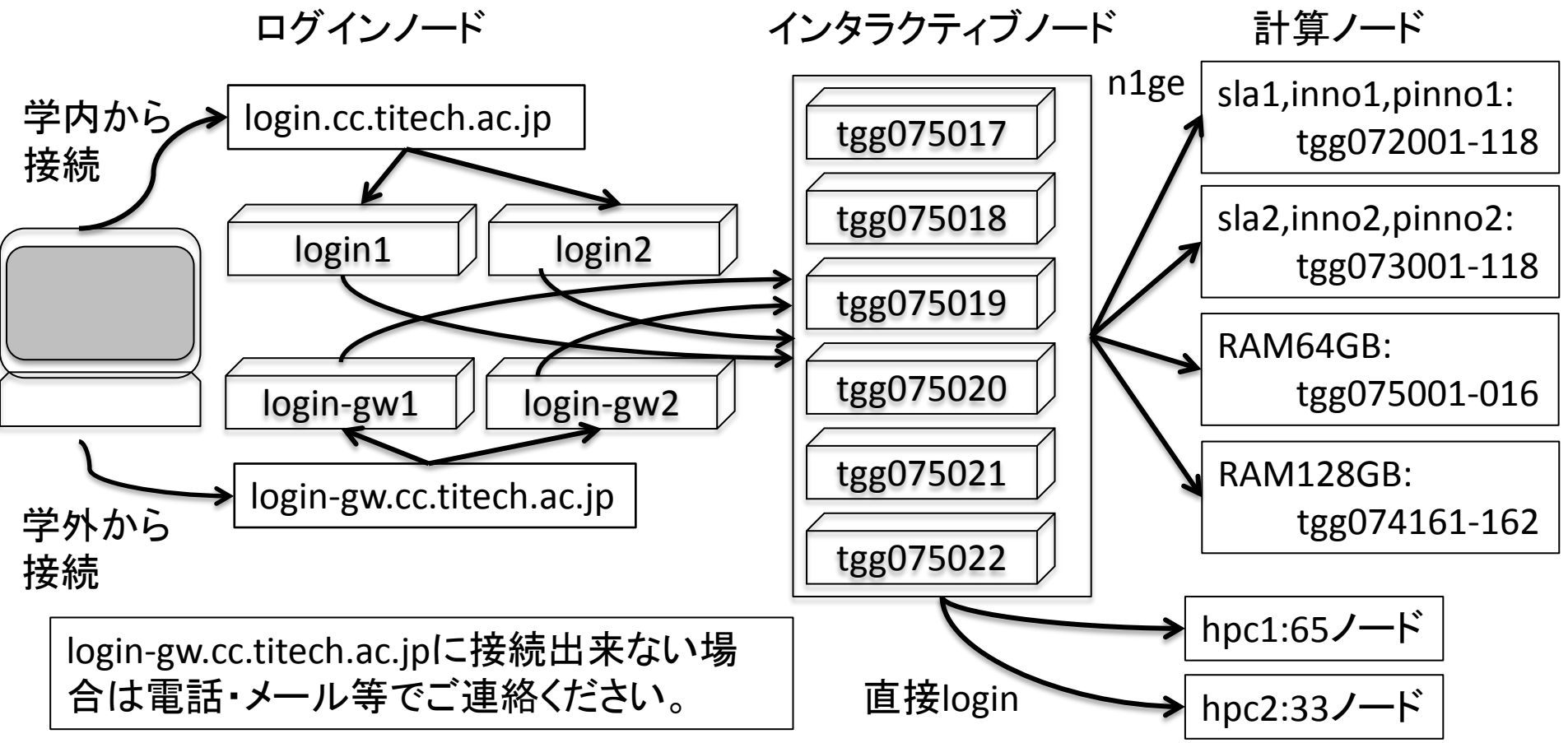
- nosupport10:

  abinit_mp glpk lam NVIDIA_CUDA_SDK oldpgi pgi php R sysnoise

  gfarm gnuplot libxml2 octave openmx pgi8 python ruby tgc_formhis

1. ログイン環境

- 会話処理(インタラクティブノード)もジョブ管理システムの制御下にあります。常に同じノードにログインできるとは限りません。



1.1 システムへのログイン

- SSHを利用して学外向けログインノードへ接続します。

  ssh ユーザ名@login-gw.cc.titech.ac.jp
- login-gwに接続出来ない場合は電話・メール等でご連絡ください。また、代替接続先としてmedia-o,media-sにアクセスください。
- 本日のように東工大に訪問されて端末室等で東工大内からアクセスする場合は

  ssh ユーザ名@login.cc.titech.ac.jp

1.1 システムへのログイン（２）

- ログインノードでの認証に成功すると自動的にインタラクティブノードにログインします。
- インタラクティブノードでは通常のUNIXの操作、プログラムのコンパイル、ジョブの投入、小規模プログラムの実行（デバッグ、プリポスト処理等：メモリーサイズ　4GB、並列数4 プロセス、実行経過時間３０分まで）が出来ます。

- ５並列以上のMPIジョブのデバッグはinno1, pinno1, inno2, pinno2, RAM64GB, RAM128GBのバッチキューをご利用ください。

1.1.2 ファイルの転送

- sftp, scpを利用してください。

  sftp ユーザ名@login-gw.cc.titech.ac.jp

  scp コピー元ユーザ名@login-gw.cc.titech.ac.jp:コピー先パス

例）

```
% sftp innoadm@login-gw.cc.titech.ac.jp
Connecting to login-gw.cc.titech.ac.jp...
Password:
sftp>
% scp  test.txt innoadm@login-gw.cc.titech.ac.jp:~/test
Password:
test.txt                                  100%    2     0.0KB/s   00:00
```

1.2 ファイルシステムとディレクトリ

以下のファイルシステムが用意されています。

- /ihome：共用促進事業専用、バックアップ有り、10Gb Ethernet接続ZFS、全ノード共有、強制容量制限無し、100GB程度に収めてください。
- /iwork：共用促進事業専用、バックアップ無し、10Gb Ethernet接続ZFS、全ノード共有、強制容量制限無し、100GB程度に収めてください。
- /work, /work2 ：学内共用、バックアップ無し、とInfiniBand（10Gbps）Lustre(並列ファイルシステム)、全ノード共有、強制容量制限1TB
- /archive ：学内共用、バックアップ無し、とInfiniBand（10Gbps）Lustre(並列ファイルシステム)、全ノード共有、強制容量制限2TB

1.2 ファイルシステムとディレクトリ

- /work, /work2, /archiveは新規に利用する際には，各自でユーザ名のサブディレクトリを作成してください.
- /work, /work2ストライピングで高速入出力が出来ますので、利用の手引きp.25「3.3.5. Lustre環境の変更」を参照してストライプのチューニングを行ってください。

  lfs help setstripe

  lfs help getstripe

  で設定方法が表示されます。

1.2 ファイルシステムとディレクトリ

- ホームディレクトリは/ihome/$LOGNAMEです。
- プログラムの実行ファイル等はワークディレクトリ/iworkに置くようにしてください。
- 東工大学内利用者向けと異なり、容量制限がかかっておりません。
  ただし、他の課題とディスク全体を共有しておりますので、これから作成する分は、各課題で100GBまでと自粛をお願いします。

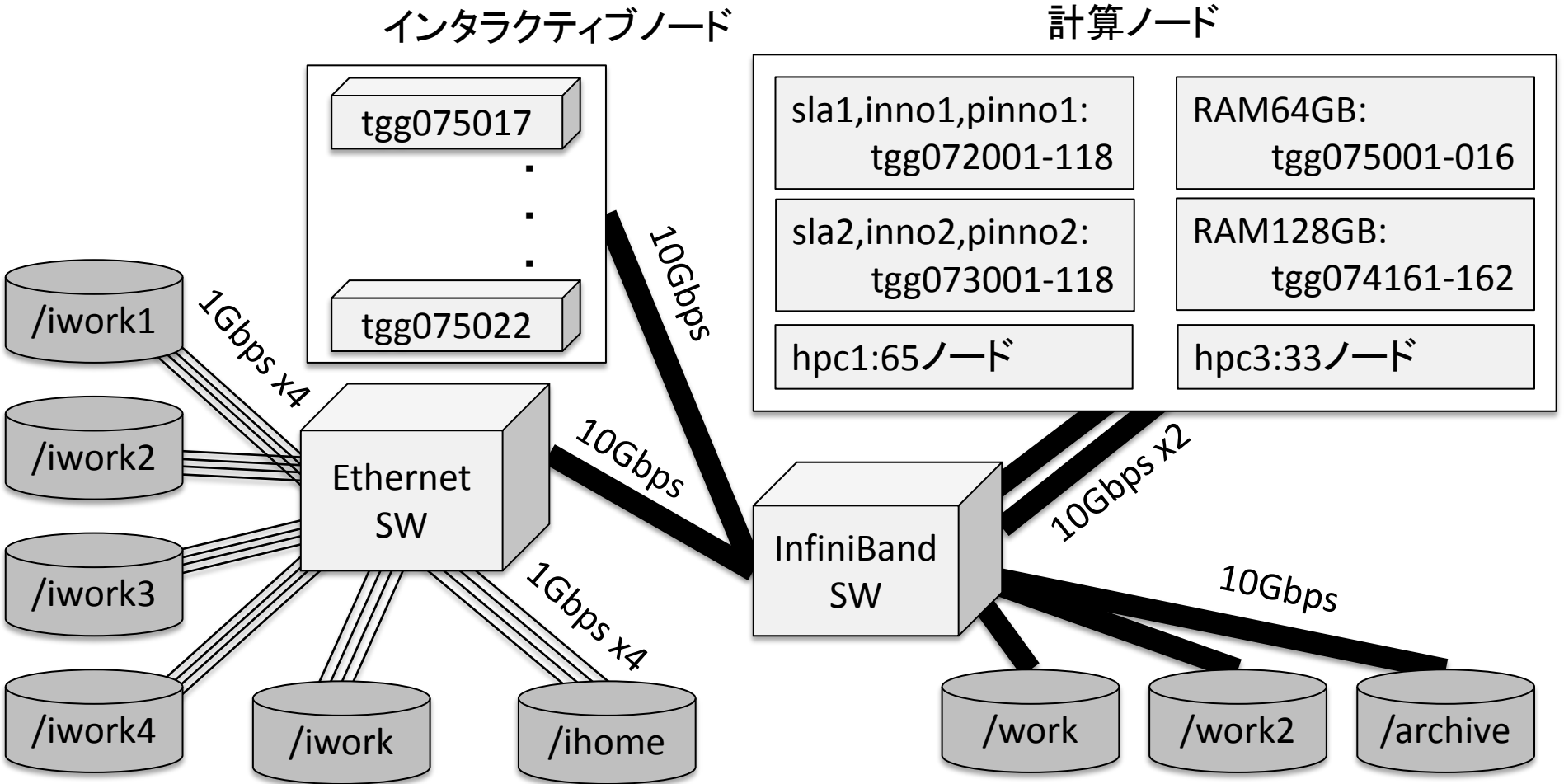
- 100GBを超える容量が必要な課題は、下記の作業領域をご利用いただけます。

```
tgt074187:/iwork1     15T   32K   15T   1% /iwork1_01
tgt074188:/iwork2     15T   32K   15T   1% /iwork2_01
tgt074189:/iwork3     15T   32K   15T   1% /iwork3_01
tgt074190:/iwork4     15T   32K   15T   1% /iwork4_01
```

  利用を希望される場合は、代表者の方が必要とする容量を kyoyo@gsic.titech.ac.jp までご連絡ください。各課題のグループ（IDの最初の５桁）名のディレクトリを上記のいずれかに作成します。

1.2ファイルシステムとディレクトリ

- /ihome, /iwork*は1Gbps x 4で接続に注意
- 大量に入出力する場合は/work、/work2を利用してください。



2. プログラムのコンパイル

- PGIのコンパイラが標準コンパイラとして利用可能です。
- この他、インテルコンパイラ、GNUコンパイラ（GCC）が利用可能です。
- 詳細は利用の手引きおよび以下のWebページを参照してください。
  http://www.gsic.titech.ac.jp/~ccwww/Announce/compiler.html
  このWebページでは、PGI、GCC、インテルコンパイラ、MPI標準環境(voltaire,MPI1.2準拠)、MPI2の利用(openmpiベース,PGI使用)、MPI2の利用(openmpi,gfortran使用)、MPI2の利用(openmpi,インテルコンパイラ)について解説してあります。
- なお、MPIに関しては可能な限りMPI標準環境を利用してください。
- PGI FortranでソースコードにSTOP文を含むものをコンパイルして作成した実行オブジェクトファイルの実行後に，標準エラー出力に「FORTRAN STOP」というメッセージが出力されることがありますが，実行およびその結果について影響はありません．また，このメッセージを抑止したい場合は，環境変数NO_STOP_MESSAGEを設定して下さい．なお，値は任意です．

3. ジョブ管理システムの利用法

3.1「qgroup -a」による課金グループの確認

- 「qgroup -a」コマンドにより課金グループとジョブ実行状態を確認します。
- GROUP_NAMEの列に表示されるのが課金グループです。
- 次のn1geコマンドで-gの後に指定するか、環境変数N1GE_GROUPに課金グループを設定してジョブを投入します。

3.1「qgroup -a」による課金グループの確認（2）

```
% qgroup –a
GROUP_NAME      TYPE   STATUS       MAX     NJOBS      PEND       RUN   CONSUME   AVAILABLE
-------------------------------------------------------------------------------------------
3S090205        SLA    OK        2880(h)   6(  11)   0(   0)   6(  11)      -    1432(h)

--

[ Running jobs ]
job-ID  pl et vc  rt(min) name       user       group    state start at    running queue        slots vslots
-------------------------------------------------------------------------------------------------------------

 323914  1  0  1     1440 Gaussian   innoadm    3S090205 r     11/16 11:59 inno2@tgg073015         4     4
 323916  1  0  1     2880 Gaussian   innoadm    3S090205 r     11/16 12:00 inno2@tgg073112         2     2
 323918  1  0  1     2880 Gaussian   innoadm    3S090205 r     11/16 12:01 inno2@tgg073033         2     2
 324017  1  0  1     2880 Gaussian   innoadm    3S090205 r     11/16 13:28 inno1@tgg072025         1     1

[ Pending jobs ]
job-ID  pl et vc  rt(min) name       user       group    state submit at    requested queue       slots vslots
-------------------------------------------------------------------------------------------------------------
 322184  1  0  1      30 OTHERS     innoadm    3S090205 qw    11/13 06:52 RAM64GB              256    256
 324474  1  0  1     120 OTHERS     innoadm    3S090205 qw    11/17 00:13 inno1                512    512
```

3.2「n1ge」によるジョブの投入

- n1geコマンドによりジョブを実行します。-helpで詳細な解説が得られます。
- どのキューが空いているかは、後述の「qstatus -sum」で確認してください。
- 利用可能な常設キューはinno1, inno2, pinno1, pinno2, RAM64GB,RAM128GBです。
- この他に事前予約制のhpc1、hpc3キューが利用可能です。
- なおinno1, pinno1, inno2, pinno2キュー名にtes2がついているものはTESLAを利用可能なノードのみから構成されたキューです。

3.2「n1ge」によるジョブの投入（2）

- 一例でMPIジョブを投入する場合のオプションを以下に示します。
- n1ge -pl 優先度 -noreq -q キュー名 –g 課金グループ -rt 実行上限時間（分） -mpi 全プロセス数:1ノードのプロセス数 -mem 1プロセス当たりのメモリ（GB） -N ジョブ名 プログラム名

3.2.1 n1geのオプション

n1ge –helpで参照可能

% n1ge -pl 2 -noreq -q inno1 -g 3S090205 -rt 1440 -mpi 16:8 -mem 3.3 -N testjob ./a.out

-pl : ジョブ投入時の優先度を指定します。1:通常、2:優先、3:最優先で、課金係数はそれぞれ1,2,4倍となります。トライアルユース採択課題はpl=2がデフォルトで強制適用されます、1は選べません。

-noreq 指定すると指定するとシステム障害等でのジョブ異常終了時のTSUBAMEシステム側での自動ジョブ再投入を抑制します。

-q キュー名　投入キュー名、TSUBAME外部利用では、inno1, pinno1, inno2, pinno2 RAM64GB, RAM128GBが利用できます。

inno1, pinno1, inno2, pinno2:最大118ノード、2.4GHz CPU, 32GBメモリ

RAM64GB:最大16ノード、2.6GHz CPU, 64GBメモリ

RAM128GB:最大2ノード、2.6GHz CPU, 128GBメモリ

-g 課金グループ、qgroup -aで確認ください。

環境変数N1GE_GROUPに課金グループを設定すると省略可能

-rt 実行時間上限（分単位）指定しない場合は、30分になります。

-mpi MPI並列数:ノード内並列数 プログラム名

-mem プロセスが利用するメモリの最大値（GB単位）、固定小数点指定可能

例）-mpi 4 -mem 3.3 : 4*3.3=13.2GBを合計で利用

-memについては利用の手引きを良く参照のこと。MPI並列では1プロセス当たりのメモリ使用量になります。

-N ジョブ名　指定しない場合は自動的にプログラム名に応じたものが付きます。

プログラム名 : 実行モジュールだけでなく、スクリプトも指定可能です。

3.3　利用可能なキュー

- 常設キュー

  interactive：会話処理を行っているノード、無償

  inno1, **pinno1**、**inno2**, **pinno2**：32GB共有メモリノード課金係数1（これに優先度課金係数2を乗じます）

  RAM64GB：64GB共有メモリノード課金係数2（同上）

  RAM128GB：128GB共有メモリノード課金係数4（同上）
- 予約制キュー

  hpc1, hpc3 それぞれ65ノード、33ノードを予約して利用できます。

  利用期間は約3日間(70時間)と約4日間(94時間)の2種類ですが、両方とも課金時間は50時間となります。

  hpc1:65ノードx50時間=3,250ノード時間

  hpc3:33ノードx50時間=1,650ノード時間

  の課金

（続）3.3　利用可能なキュー

- **inno1, pinno1, sla1** と **inno2, pinno2, sla2**、

のキューの実体はそれぞれ同一です。

- 各制度で利用可能なキューが異なります。
- **inno1, inno2**：共用促進事業、共同利用有償成果公開
- **pinno1, pinno2**：共同利用有償成果非公開
- **sla1, sla2**：東工大学内

3.4 バッチキューの状態の確認

- qstatus –sumによりバッチキュー毎の空きノード、空きCPUコア、空きメモリ、利用中ノード、停止ノード、利用不可ノード、合計ノードが出力されます。

```
--
QUEUE       FREE NODE   FREE CPU   FREE MEMORY  USED NODE  DOWN NODE  DISABLE NODE      TOTAL NODE
-------------------------------------------------------------------------------------------------
TOTAL          544   3634 CPU     12131 GB       464         3          79
 - bes1        107    845 CPU      2456 GB         9         0           0        116
 - bes1tes2     19     76 CPU       397 GB         1         0           0         20
 - bes2         74    277 CPU       902 GB        41         0           3        118
 - bes2tes2     38    152 CPU       504 GB        27         0           2         67
 - cs1         107    214 CPU      2456 GB         9         0           0        116
 - cs2          74    148 CPU       902 GB        41         0           3        118
 - hpc1          0      0 CPU         0 GB         0         0          67         67
 - novice       11    176 CPU       352 GB         0         0           0         11
 - RAM128GB      2     32 CPU       256 GB         0         0           0          2
 - RAM64GB      15    240 CPU       960 GB         0         0           1         16
 - sla1         41    656 CPU      1312 GB        76         0           1        118
 - sla1tes2     22    352 CPU       704 GB        37         0           1         60
 - sla2         18    288 CPU       576 GB        99         0           1        118
 - sla2tes2      9    144 CPU       288 GB        48         0           0         57
 - tsubasa       7     34 CPU        66 GB        76         3           0         86
```

オプション指定

- qstatus -sum -inno -ramとすると共用促進事業で利用可能なキューに限って出力されます。

% qstatus -sum -inno -ram

```
--
QUEUE       FREE NODE   FREE CPU   FREE MEMORY  USED NODE  DOWN NODE  DISABLE NODE  TOTAL NODE
-------------------------------------------------------------------------------------------
TOTAL           78   1248 CPU      3168 GB       173         0           3
 - inno1        43    688 CPU      1376 GB        74         0           1        118
 - inno2        18    288 CPU       576 GB        99         0           1        118
 - RAM128GB      2     32 CPU       256 GB         0         0           0          2
 - RAM64GB      15    240 CPU       960 GB         0         0           1         16
```

−inno を−pinnoとすると有償非公開制度での利用可能キューが表示されますが、空きノードなどの情報は−innoと同一です。

3.5 ジョブの削除

- qdel_slaまたはqdeleteコマンドでジョブを削除します。

(例) inno1で実行中ジョブ(Job-ID 1350)を削除する場合

% qdel_sla 1350

または

% qdelete -c sla 1350

^^^※キュークラス名

または

% qdelete -q inno1 1350

^^^^^※キュー名

Job-IDはqstat –u $LOGNAME等で確認してください。

利用上不明なことがありましたら、気兼ねなく、

- 先端研究施設共用促進事業トライアルユースの利用者は

  kyodo@gsic.titech.ac.jp
- 共同利用制度の有償利用の利用者は

  tsubame@gsic.titech.ac.jp

までお問い合わせください。