

# TSUBAME2.0 利用講習会

平成24年度版 (Rev.20120718)

東京工業大学

学術国際情報センター

# TSUBAME利用の情報源

[TSUBAME共同利用](#) > よくある質問、FAQ

一般的情報 (<http://www.gsic.titech.ac.jp/kyodou/FAQ#faq-list>)

Q. TSUBAME2.0 を利用するための、基本的な情報はどこを見ればよいですか。

A.

この共同利用のFAQは今後、充実予定です。

[TSUBAME計算サービス内のTSUBAME 2.0 利用の手引きHTML,PDF](#)

[TSUBAME計算サービス内のTSUBAME2.0利用ポータル利用の手引き\(PDF\)](#)

[TSUBAME計算サービス内の各種利用の手引き](#)

[TSUBAME計算サービス内のFAQ](#)

・ [TSUBAME2.0 を利用するための、基本的な情報はどこを見ればよいですか。](#)

「利用の手引き」「FAQ」は必ず目を通してください。

Q. FAQを見てもわからない場合には、どこへ連絡したらよいですか。

A.

TSUBAME共同利用 利用者は [kyodo\\_at\\_gsic](mailto:kyodo_at_gsic) へ、先端研究施設共用促進事業 利用者は [kyoyo\\_at\\_gsic](mailto:kyoyo_at_gsic) へご連絡ください。

(上記のメールアドレスの \_at\_ を半角の@へ変換した後に .titech.ac.jp を付けてください。)

先端研究施設共用促進事業の利用者は [kyoyo@gsic.titech.ac.jp](mailto:kyoyo@gsic.titech.ac.jp)  
共同利用制度の有償利用の利用者は [kyodo@gsic.titech.ac.jp](mailto:kyodo@gsic.titech.ac.jp)  
[soudan@cc.titech.ac.jp](mailto:soudan@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。

# 1.1 利用可能アプリケーション

- 学外利用の方が無償で利用可能なソフトウェア

- OS (SUSE Linux Enterprise Server 11 SP1)
- ジョブスケジューラ (PBS Professional)
- PGI CDK、Intel Compiler、TotalView Debugger等 開発環境
- Gaussian 09、GaussView 5、Gaussian用Linda
- Gaussian 03

利用可能ですが、正式サポートして  
おりません。ご確認の上、ご利用ください。

G03 の利用方法:

```
export g03root=/usr/apps/isv/gaussian/gaussian03.E01
source $g03root/g03/bsd/g03.profile
g03 <入力ファイル>
```

- TSUBAMEに導入済みのFree Software

- 上記以外の一般的な商用ソフトウェア

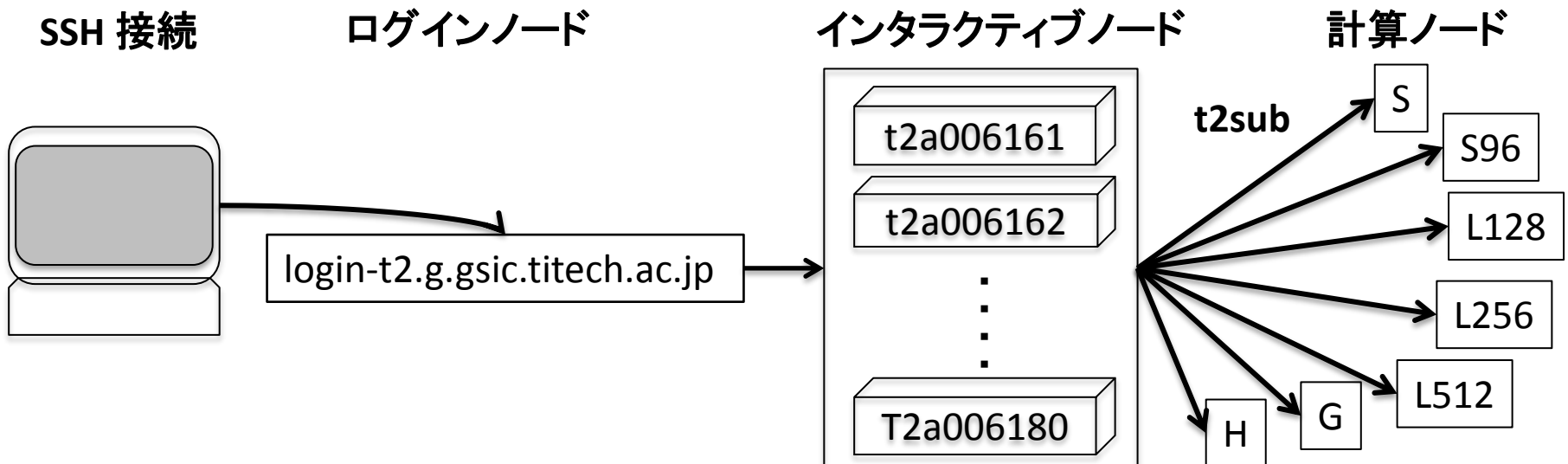
- 利用者がベンダーからライセンスを取得して利用者のディレクトリにインストールしてご利用ください。

## 1.2 アプリケーション (Freeware編)

- いわゆるFreewareは自由に利用できます
  - <http://tsubame.gsic.titech.ac.jp/ja/applications> のFree softwareへ掲載予定
- 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。
  - fftw3.2.2 gamsess\_25Mar2010 gromacs\_4.0.7 libpng10  
lzma povray\_3.6.1 R\_2.11.1 teTeX tgif tinker\_5.1.08  
NVIDIA\_GPU\_Computing\_SDK(3.0, 3.1, 3.2, 4.0, 4.1)
- サポート無しFreeware(/usr/apps/free/nosupport)
  - GSICでのSEによるサポートがありません。
  - goto2\_noth Gromacs\_4.5.3 openmm\_2.0 sshfs\_2.2

# 2.1 利用法概略

- SSH (鍵認証) でリモートログイン  
ssh ユーザ名@login-t2.g.gsic.titech.ac.jp  
(東工大内からは、パスワード認証によるSSH接続も可)
- インタラクティブノードへログイン  
複数のインタラクティブノードのいずれかに割り振り
- 計算ノードでのジョブの実行



## 2.2 ファイルの転送

- (Linux) sftp, scpを利用してください。

```
sftp ユーザ名@login-t2.g.gsic.titech.ac.jp
```

```
scp コピー元ユーザ名@login-t2.g.gsic.titech.ac.jp:コピー先パス
```

例)

```
% sftp watanabe-t-bd@login-t2.g.gsic.titech.ac.jp
```

```
Connecting to login-t2.g.gsic.titech.ac.jp...
```

```
sftp>
```

```
% scp test.txt watanabe-t-bd@login-t2.g.gsic.titech.ac.jp:~/test
```

```
test.txt                100%  2  0.0KB/s  00:00
```

- (Windows/Mac 用) GUIクライアントソフトウェアとしてはWinSCPやCyberduckが使えます。

## 2.3 ファイルシステムとディレクトリ

### 利用可能ファイルシステムの概要

- 利用可能ストレージ容量（インタラクティブ／バッチ共通）
  - ホームディレクトリ            1利用者当たり    25GB
  - グループディスク(30Pt/1TB・1月)    課題当たり    上限30TB
  - /work0    Lustre FS
  - /data0    GPFS + テープシステム
- スクラッチ領域（ジョブ実行中のみ利用可能）
  - ローカル(SSD)            各ノード当たり    56～240GB
  - 並列ファイルシステム    共有領域            最大786TB

## 2.3 ファイルシステムとディレクトリ詳細 (1/2)

以下のファイルシステムが用意されています。

### 1. home

- ユーザーあたり 25GBを利用できます。
- 利用状況の確認は t2quota コマンドを使用します。

### 2. グループディスク

- 有償です。1TB・1月あたり 30TSUBAME ポイントです。容量上限は、1TB単位での月毎の申請制です。
- 使用量や上限量は、t2quota -g コマンドにて確認してください。

Disk Utilization (/work0):

	Block Limits				File Limits				
GROUP	KB	quota	limit	grace	files	quota	limit	grace	
t2ggsic-tw	3838018136*	1063004405	1073741824		-	128478	0	0	-

- /work0/グループ名/、または /data0/グループ名/ の下に、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。
- Lustreの設定方法は、利用の手引「4.6.2 Lustre」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。



## 2.3 ファイルシステムとディレクトリ詳細(2/2)

以下のファイルシステムが用意されています。(つづき)

### 3. 共有スクラッチ

- 並列ファイルシステム(Lustre)で構成された共有スクラッチ領域を利用できます。利用する際には、\$TMPDIRに共有スクラッチ領域のパスが設定されます。ジョブスクリプトの中で、作業領域のパスを指定することにより参照可能です。
- Gaussianでの利用法は、利用の手引「7.6 Gaussian」を参照してください。

### 4. ローカルスクラッチ

- SSDを高速スクラッチとして使用することができます。
- PBSのstagein/stageout を用いてご利用ください。利用方法は、利用の手引「4.6.4 ローカルスクラッチ」をご参照ください。
- SSDで使用できる容量は50GB前後ですので、Gaussian等の大きなファイルを使用するアプリケーションでは共有スクラッチ(Lustre)をご利用ください
- PBSが転送処理を行います。ファイル数が多い場合やファイルサイズが大きい場合などは転送に時間がかかる場合がありますので、ご使用になるプログラムの処理をご考慮の上、ご利用ください。

## 2.4 インタラクティブノードの利用範囲

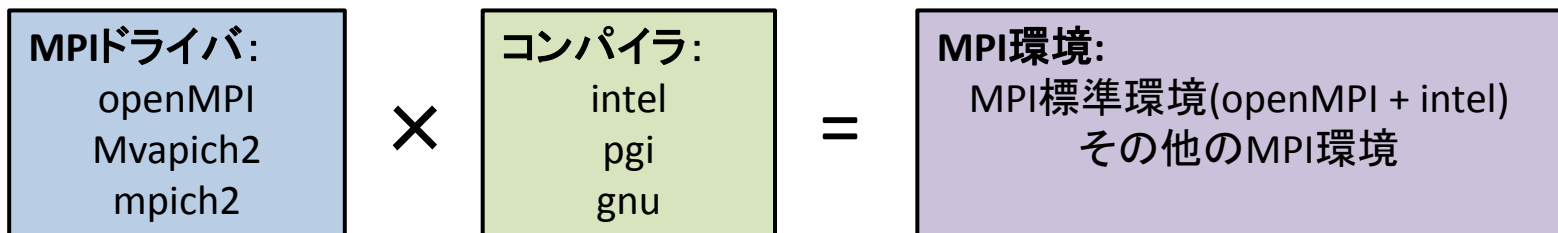
- インタラクティブノードでは、以下の制限の範囲内にて、通常のUNIXの操作、プログラムのコンパイル、ジョブの投入、小規模プログラムの実行が出来ます。
  1. ユーザーあたりの利用メモリ 4GB
  2. ユーザーあたりのプロセス数 4
  3. ユーザーあたりのディスク使用上限量  
/home=25GB, グループディスク = 0~30TB/group
  4. CPU連続利用時間は 30分
  5. 1時間利用がない場合は接続を切断
- それ以上のジョブのデバッグには2ノード・10分までの無償デバックジョブ(\*)や、有償のバッチキューをご利用ください。

(\*) HとGを除く各キューで可能 (Group ID を指定せずに投入する。)

# 3. プログラムのコンパイル

- TSUBAME2ではコンパイラは、intel, pgi, gnuの3種類が利用可能です。基本的情報は、利用の手引「4. プログラミングと環境」をご参照ください。
- MPI並列

MPI環境は、コンパイラとMPIドライバの組み合わせで幾種類かを選択して利用することが可能です。openMPIとintelコンパイラがMPI標準環境です。その他のMPI環境の利用方法は利用の手引「6.3. MPI環境」をご参照ください。



# 4. ジョブ管理システムの利用法

4.1 「id」によるTSUBAMEグループの確認

4.2 「t2group」による利用可能ポイントの確認

4.3 「t2sub」によるジョブの投入

4.4 「t2sub」によるジョブの投入例

4.5 「t2del」によるジョブの削除

# 4.1 「id」&「t2group」コマンド

- 「id」コマンドで、所属するグループを確認します。

```
> id
```

```
uid=10000XXXXX(watanabe-t-bd) gid=100(users) 1000000XXX(t2g-TSUBAME-Group)
```

- TSUBAMEグループは課題毎に設定されるt2gから始まる文字列
  - 課金グループ (ジョブ投入時に指定、計算機資源をTSUBAMEポイントとして管理)
  - Unix ファイルシステムのアクセス権管理
  - /work0 や /data0 などのグループディスク
- t2group コマンドにより、利用可能なTSUBAMEポイントが確認できます。

```
> t2group
```

```
Groups users t2g-XXXXXX of User watanabe-t-bd
```

```
-----Group-----|-----FixedQueue-----|-----  
MeasuredRateQueue----
```

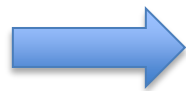
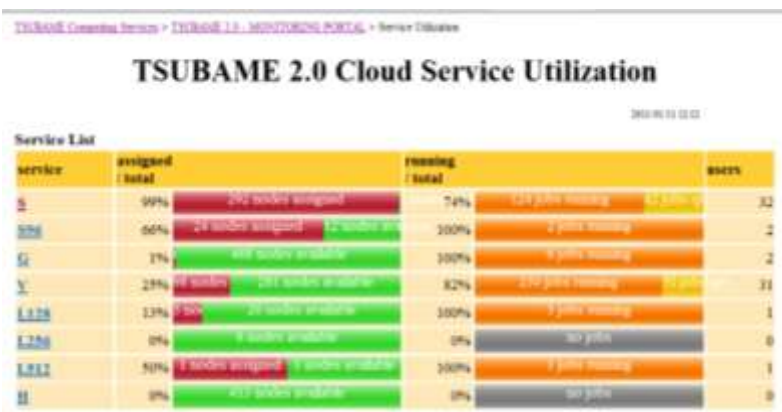
Name	Condition	Factor	Condition	units	Factor	Condition	sec
t2g-XXXXXX	used	4.0	stopped	0	1.0	used	<b>32400000</b>

これは、 $32400000 \div 3600 = 9000$  (TSUBAMEポイント) =  $3000$  (TSUBAMEポイント) × 3 (口)

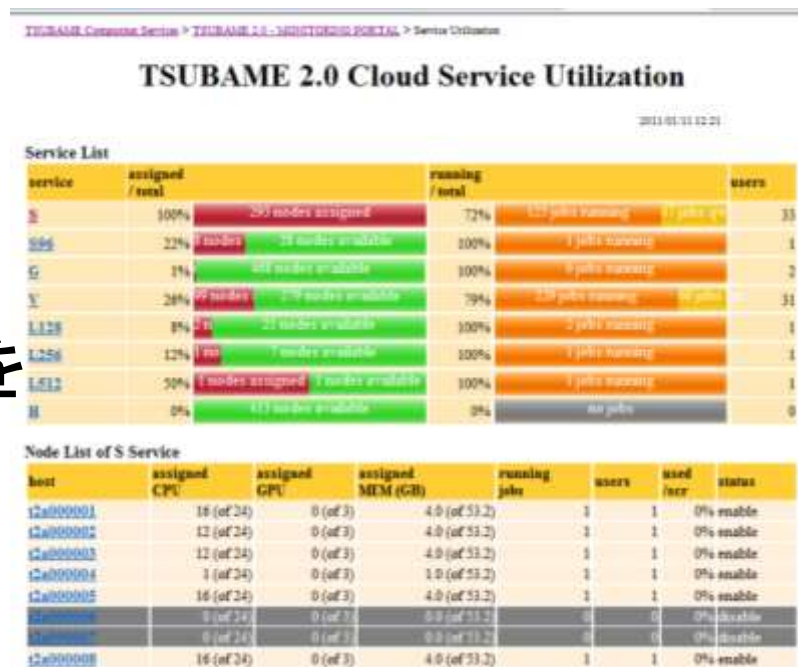
- TSUBAME利用ポータルで、より詳細な情報を確認できます。

# 4.2 「t2sub」によるジョブの投入

- t2sub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。
- 各キューの実行状況は下記URLをご参照ください。  
<http://mon.g.gsic.titech.ac.jp/summary/>



Sキューをクリック



## 4.3 「t2sub」によるジョブの投入例(1/2)

- シングルジョブを投入する場合

```
t2sub -q S -l select=1:mem=48gb -W group_list=t2g-XXXXX -l  
walltime=5:00:00 ./job.sh
```

```
#!/bin/sh  
cd $HOME/test  
./myprog < input_data
```

- openMP/Pthreadによる並列

```
t2sub -q S -l select=1:ncpus=12:mem=48gb  
-W group_list=t2g-XXXXX -l walltime=5:00:00 ./job.sh
```

```
#!/bin/sh  
export OMP_NUM_THREADS=12  
export NCPUS=12  
#  
cd $HOME/test  
./myprog < input_data
```

- MPIジョブ

```
t2sub -q S -l select=4:ncpus=12:mpiprocs=12:mem=48gb -l place=scatter  
-W group_list=t2g-XXXXX -l walltime=5:00:00 ./mpitest.sh  
(1ノードに12プロセスで48GBメモリ) が4つ = 4ノード48並列
```

```
#!/bin/sh  
export PATH=/usr/apps/openmpi/1.4.2/pgi/bin:$PATH  
export LD_LIBRARY_PATH=/usr/apps/openmpi/1.4.2/pgi/lib:$LD_LIBRARY_PATH  
#  
cd $HOME/test  
mpirun -n 48 -hostfile $PBS_NODEFILE ./myprog < input_data
```

## 4.3 「t2sub」によるジョブの投入例(2/2)

- ジョブは(投入ディレクトリでなく)home ディレクトリで実行される。

`cd $PBS_O_WORKDIR`                      ジョブ投入ディレクトリにて実行する。

```
#!/bin/sh
cd $PBS_O_WORKDIR
./myprog < input_data
```

- デバックジョブなど(課金グループを省略)

```
t2sub -q S -l select=2:ncpus=12:mpiprocs=12:mem=48gb -l place=scatter
```

```
-l walltime=0:10:00 ./mpitest.sh
```

2ノード、10分までに制限されるが、無償

- よく使う t2sub のオプション

`-W group_list=t2g-XXXXX`

課金グループの指定(有償ジョブの場合、必須)

`-l walltime=24:00:00`

計算時間上限の指定(省略時は1時間)

`-et {0|1|2}`

時間延長オプションの指定(デフォルトは0)

`-p {0|1|2}`

ジョブ優先度の指定(デフォルトは0)



## 4.4 「t2del」によるジョブの削除

- 投入したジョブを、終了を待たず削除するには t2del コマンドを使用します。

```
> t2del 147.t2zpbs03
```

- 実行状態のジョブを削除する場合は、-W force オプションを追加します。キュー待ち状態のジョブも削除可能です。

```
> t2sub -q -S -l select=1:ncpus=8 ./job.sh
```

```
1234.t2zpbs01
```

```
> t2del -W force 1234.t2zpbs01
```

- Job-IDはt2stat 等で確認してください。

# 5. 利用可能計算資源

キュー名	ノード数	CPUコア数	GPU	メモリ	SSD	ネットワーク	課金係数
S	300	<b>12コア</b> Intel Xeon 2.93GHz 6コアx2	M2050 <b>3GPU</b>	<b>54GB</b>	120GB	80Gbps QDR IB x 2	1.0
S96	41			<b>96GB</b>	240GB		1.2
G	480	<b>4コア</b> Intel Xeon 2.93GHz 6コア相当		<b>22GB</b>	120GB		0.5
L128F	10	<b>32コア</b> Intel Xeon 2.0GHz 8コアx4	<b>M2070 4GPU</b>	<b>128GB</b>	480GB	40Gbps QDR IB x 1	2.0
L128	10		S1070 2GPU	<b>128GB</b>	480GB	40Gbps QDR IB x 1	2.0
L256	8			<b>256GB</b>			4.0
L512	2			<b>512GB</b>			8.0

# 5.1 バッチキューー一覧

- インタラクティブノード
  - i: インタラクティブ専用ノード ([login-t2.g.gsic.titech.ac.jp](http://login-t2.g.gsic.titech.ac.jp))
- バッチキューー(計算ノード)
  - [S] ノード占有系: 12 CPUコア、3 GPU(M2050)のノード利用
  - [L] ノード占有系: 32 CPUコア、2 GPU(S1070)のノード利用
  - [L128F] : 32 CPUコア、4 GPU(M2070)のノード利用
  - [G] GPU系: 4CPUコア、3GPUのノード利用
  - [H] 予約系: Thinノードをノード数、期間を予約して利用
  - [X] ノード占有系: Hキューーに予約が無い場合に、Sキューー相当となる。

## 5.2 課金について

- インタラクティブノードは無償
- 各キュー(HとGを除く)で2ノード、10分までのデバックジョブを投入可能
  - Group ID を指定せずに投入する。
- グループディスクはディスク課金
- 課金 = 使用ノード数 × 経過時間
  - プロセス数には無関係
  - 標準エラー出力(ex. OTHERS.e\*)に概算値を表示
- 課金係数(Sキューを基準 1.0 とする。)

S96	1.2,	L128	2.0,	L256	4.0
L512	8.0,	H	1.25,	G	0.5
- 時間延長オプション(S, S96, L128, L256, L512, G 共通)  
-et 0 (24hまで,default), -et 1 (48hまで), -et 2 (96hまで)

# 5.3 確保したノードへの直接ログイン

- S系、L系、G系、H系キューにて可能

右枠内のスリープジョブ(dummy.sh)をバッチのSキューに投入します。

```
1 時間スリープの例
$ cat dummy.sh
#!/bin/sh
sleep 3600
```

```
$ t2sub -W group_list=t2gxxxx -l select=1:ncpus=4:mem=40gb -q S dummy.sh
```

ステータスを確認して、実行にはいっていたら  
ジョブ番号から使用できるマシンを確認します。

このt2a000174マシンにログインできます。

```
$ ssh t2a000174
```

作業が終わりましたら、exitして

```
$ t2del 61092.t2zpbs01
```

とすれば、数秒後にジョブが終了いたします。

```
$ t2stat
Job id      Name      User      Time Use S Queue
-----
61092.t2zpbs01 OTHERS    watanabe-t-bd  0 Q

$ t2stat
Job id      Name      User      Time Use S Queue
-----
61092.t2zpbs01 OTHERS    watanabe-t-bd  00:00:00 R S
```

実行状態へ

```
$ t2stat -n1
Job ID      Username Queue   Req'd Req'd Elap
Jobname    SessID NDS TSK Memory Time S Time
-----
61092.t2zpbs01 watanabe S      OTHERS 27221 1 2 40gb 00:00 R 00:00 t2a000174/0*2
```

## 5.4.1 ノード占有系：Sキュー・Lキュー

- Sキュー：12CPUコア, 3GPU, 54GBメモリを持つノードを利用
  - 多数CPUまたはGPUによる並列性や、I/O(ディスク・通信)性能が必要なジョブ向け
  - ノード内のジョブ混在は起こらない。
  - 確保したノードへの直接ログインも可能。
- 大容量メモリが必要なジョブには、S96, L128, L256, L512キュー
  - 数字はメモリ容量(GB)
  - Sに比べ1.5倍、2倍...の課金
  - L系はMedium/Fatノードなので、CPUが多く、GPUが古い

## 5.4.2 予約系：Hキュー

- 予約した期間ノードを占有して利用
  - 1000CPUコアレベルの並列性が必要なジョブ向け
  - Webから日程・ノード数を予約
  - バッチキューを介さない利用も可
    - Sleep ジョブを投入しなくても、同じグループに属するユーザーはログイン可能。
  - 柔軟な予約が可能
    - ノード数は16以上自由、期間は一日単位で最大7日

## 5.4.3 GPU系：Gキュー

- ノードあたり3GPU+4CPUコアを

- GPUジョブに適している

- 以下のようなノードに見える

- 4CPUコア
- 3GPU
- 22GBメモリ

- 残りの計算資源を仮想マシンで、別キューに提供

- 従量制課金、Sに比べ0.5倍 (お買い得)

- GPU講習会、GPUコンピューティング研究会

元々のノード構成

- 12 CPUコア
- 3 GPU
- 54 GBメモリ

残りの計算資源は？

- 8 CPUコア
- 32GBメモリ



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

[kyoyo@gsic.titech.ac.jp](mailto:kyoyo@gsic.titech.ac.jp)

- 共同利用制度の有償利用の利用者は

[kyodo@gsic.titech.ac.jp](mailto:kyodo@gsic.titech.ac.jp)

までお問い合わせください。