Migration from TSUBAME3.0 to TSUBAME4.0

TSUBAME Computing Services, Global Scientific Information and Computing Center

2024-06-04

Table of contents

1. Introduction

This document is intended for existing TSUBAME 3.0 users, with an emphasis on the differences between the new TSUBAME 4.0 system and TSUBAME 3.0, to get you up and running quickly and easily.

1.1. TSUBAME4.0 architecture

TSUBAME4.0 architechure is described here.

2. Getting started, Login

2.1. Get an account

In order to use this system, it is necessary to register a user account.

Please refer Getting Accounts Page for details, as the procedure depends on your affiliation and program to apply.

The process of starting to use and obtaining an account is the same as for TSUBMA3.0.

2.2. Login

You need to upload the SSH public key to access the login node.

Please refer to TSUBAME portal User's Guide "SSH public key registration" for the operation of public key registration.

The login method is the same as for TSUBMA3.0.

2.3. Login nodes

To use this system, you must first log in to the login node; the terminal software used in TSUBAME 3.0 can be used without modification.

The login destination is connected to via SSH. File transfers also connect via SFTP.

```
login.t4.gsic.titech.ac.jp
```

The domain has been changed from login.t3.gsic.titech.ac.jp to login.t4.gsic.titech.ac.jp.

If you want to login to any login node, specify the following host name (FQDN)

```
login1.t4.gsic.titech.ac.jp
login2.t4.gsic.titech.ac.jp
```

The login nodes of TSUBAME3.0 were login0 and login1, while those of TSUBAME4.0 are login1 and login2.

2.3.1. Restrictions for heavy work in login nodes

As login nodes (login, login0, login1) are shared with many users at the same time, please do not execute programs which dominate CPU time. For parallel or long-time computation, please use compute nodes using qsub and qrsh commands, Followings are examples of judgment criteria. When the system administrator noticed your program is preventing others, it will be terminated even if it is permitted or not prohibited here,

Permitted operations

- File transfer, compression, decompression (e.g., scp, sftp, rsync, tar)
- Program compilation (If you consume lots of resources by parallel compilation etc., use compute nodes)

Prohibited operations

- Calculation using commercial applications, freeware, or programs made by yourself
- Execution of programs that exceeds 10 minutes (except for file transfer)
- Execution of parallel programs (including python and MPI)
- Execution of programs that consumes lots of memory
- Execution of lots of processes simultaneously (e.g., parallel compilation)
- Execution of server programs or auto-restart programs (e.g., VSCode Server, Jupyter Notebook)
- Other operations which use lots of CPU resources

The login nodes have 4 GB memory limit per process. The system administrator will terminate programs with excessive loads without prior notice. If you want to execute such operations, or you feel login nodes are too heavy, please use compute nodes as interactive jobs via job scheduler.

The rules for using login nodes and computation nodes are the same as in TSUBAME 3.0.

2.4. How to check TSUBAME points

You can check TSUBAME points with `t4-user-info group point` command as below.

```
$ t4-user-info  group point -g  TESTGROUP
gid     group_name                      deposit      balance
------------------------------------------------------------
xxxx  TESTGROUP                            5000     800000000
```

The TSUBAME point confirmation command has been changed from t3-user-info to t4-user-info.

3. Environment

3.1. Storage

In TSUBAME4.0, a home directory and two types of group disks (fast storage area and large storage area) are available.

The home directory and fast storage area are built on SSD shared storage, and the large storage area is built on HDD shared storage.

TSUBAME4.0	Storage	Mount point	Capacity	Filesystem
	High-speed storage area Home directory (SSD)	/gs/fs /home	372TB	Lustre
	Large-scale (Big) storage area Shared application deployment (HDD)	/gs/bs /apps	44.2PB	Lustre
	Local scratch area (SSD)	/local	1.92TB/node	xfs

The local scratch area is located on the NVMe SSD of each compute node and can be used for temporary files, etc. during the computation.

Info

The capacity of the available local scratch area is determined by the resources acquired.
The shared scratch area (BeeOND) that was available in TSUBAME3 has been discontinued. For details, see Appendx.4. Storage for details.

Resource type	Local scratch area (GB)
node_f	1920
node_h	960
node_q	480
node_o	240
gpu_1	240
gpu_h	120
cpu_160	96
cpu_80	48
cpu_40	24
cpu_16	9.6
cpu_8	4.8
cpu_4	2.4

3.2. Compute nodes

The compute node for TSUBAME 4.0 is a 4th generation AMD EPYC 9654 on the Zen4 architecture, with more than 6 times more cores per node than TSUBAME 3.0.

The compute node has 4 NVIDIA H100 Tensor Core GPUs.

	TSUBAME3.0	TSUBAME4.0
Computing Unit	Compute node HPE SGI ICE-XA 540 nodes	Compute node HPE Cray XD665 240 nodes
Components (per node)		
CPU	Intel Xeon E5-2680 v4 2.4GHz x 2 Socket	AMD EPYC 9654 2.4GHz x 2 Socket
Cores/Threads	14cores / 28threads x 2CPU	96cores / 192threads x 2CPU
Memory	256GiB	768GiB (DDR5-4800)
GPU	NVIDIA TESLA P100 for NVlink-Optimized Servers x 4	NVIDIA H100 SXM5 94GB HBM2e x 4
SSD	2TB	1.92TB NVMe U.2 SSD
Interconnect	Intel Omni-Path HFI 100Gbps x 4	InfiniBand NDR200 200Gbps x 4

Info

TSUBAME4.0 calculation nodes are from r1n1 to r23n11. r: 1 to 23 n: 1 to 10 or 11

3.3. Job Scheduler

TSUBAME4.0 uses the Altair Grid Engine (AGE), the successor to the UNIVA Grid Engine (UGE) of TSUBAME3.0.

The resource types in TSUBAME4.0 are as follows.

The number of resource types has increased, and the number of cores available for each resource type has also increased.

Resource type	Physical CPU cores	Memory (GB)	GPUs	Local scratch area (GB)
node_f	192	768	4	1920
node_h	96	384	2	960
node_q	48	192	1	480
node_o	24	96	1/2	240
gpu_1	8	96	1	240
gpu_h	4	48	1/2	120
cpu_160	160	368	0	96
cpu_80	80	184	0	48
cpu_40	40	92	0	24
cpu_16	16	36.8	0	9.6
cpu_8	8	18.4	0	4.8
cpu_4	4	9.2	0	2.4

3.3.1. Subscription Job

TSUBAME4.0 introduces a "subscription" that allows quasi-exclusive use of computation nodes on a monthly basis.

Only intramural users and joint use (academic) users can use this service.

To submit a job under the subscription system, add -q prior. Other options are the same as the pay-as-you-go system.

```
$ qsub -q prior -g [TSUBAME group] SCRIPTFILE
```

Option	Description
-g	Specify the TSUBAME group name. Please add as qsub command option, not in script.
-q prior	Subscription job. Wait one hour at most until execution.

For more details about compute node subscription, check here.

> **Warning**
>
> Even if a job for the subscription group, note that if `-q prior` is not specified, the job will be processed as a pay-as-you-go job.

3.4. Software

3.4.1 Commercial application

The differences between commercial applications available in TSUBAME4.0 and TSUBAME3.0 can be found here.

Each application fee is required for the use of some commercial applications. For more details, please refer to Fare Overview **Commercial Applications (Partially charged in TSUBAME4.0)**.

3.4.2 Freesoft

The difference between the free software available for TSUBAME4.0 and TSUBAME3.0 can be found here.

3.4.3 Applications used in TSUBAME 3.0

TSUBAME4.0 and TSUBAME3.0 have different compilers, MPI, and various libraries, so they cannot be run as they are. It is necessary to recompile the program on TSUBAME4.0.

4. Data migration from TSUBAME3.0

The following is an example of copying data on TSUBAME3 to TSUBAME4.
All commands below are to be executed on TSUBAME3.
In the example, ux00000 should be read as your TSUBAME4 login name.

- Create SSH key pair on TSUBAME3

```
$ ssh-keygen -t ecdsa
```

- Register the created SSH public key to TSUBAME4 portal (Registration)

```
$ cat ~/.ssh/id_ecdsa.pub
```

- Confirm that you can SSH from TSUBAME3 to TSUBAME4

```
$ ssh ux00000@t4-login1
/home/1/ux00000
```

Note

There are two login nodes for TSUBAME4, which can be accessed from TSUBAME3 by t4-login1 and t4-login2.

- To move files from TSUBAME3 to TSUBAME4 home directory

```
$ rsync -az ./dir-to-send ux00000@t4-login1:/home/1/ux00000/dir-to-store
```

- When moving files from TSUBAME3 to TSUBAME4 group disk

```
$ rsync -az ./dir-to-send ux00000@t4-login1:/gs/bs/tgz-XXXXX/dir-to-store
```

Tips

TSUBAME4.0 group disk includes the high-speed storage area (SSD) and the large-scale (big) storage area. Details are described here.

Tips

If you want to use a non-default SSH key pair, specify the location as below.

```
$ rsync -az -e 'ssh -i /path/to/key' ./dir-to-send ux00000@t4-login1:/home/1/ux00000/dir-to-store
```

Revision History

Date	Change
2024/04/22	First edition