HPC User Group Meeting

September 21st, 2022

Jan Moren, SCDA

OIST

OKINAWA INSTITUTE OF SCIENCE AND TECHNOLOGY GRADUATE UNIVERSITY
沖縄科学技術大学院大学

TODAY

Ask SCDA

SCDA **Answers**

Storage Request System

We now have
a process!

Saion Update

New
system version!

New
module command!

New
GPU job settings!

The Bioinformatics Group Does it Again!

Refreshed Saion

Changes:

- OS updated to CentOS 7.9
- module system now same as Deigo ('ml' command now works)
- X11-related issues on the nodes are fixed
- Newer CUDA versions are available

Major Change:

We now limit GPUs per user

- 4/8 GPUs per user
- 36 CPU cores per user
- 7/2 Days per job

Refreshed Saion

- The lack of GPU limits was a problem with old Slurm versions
- Very few users have ever used more than 8 GPUs at once
- Most GPU jobs also need CPU
  - ➔ they will speed up with more CPU cores available than before

Major Change:

We now limit GPUs per user

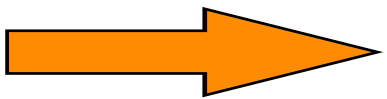
- 4/8 GPUs per user
- 36 CPU cores per user
- 7/2 Days per job

Ask SCDA

- IT section and SCDA have both grown
- We work differently, have different requirements

→ We need our own issue tracker

it-help@oist.jp



ask-scda@oist.jp

https://ask-scda.oist.jp

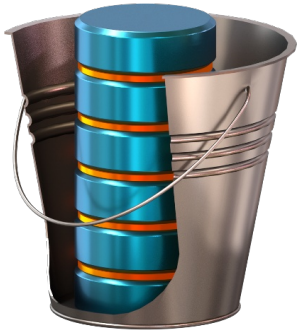
SCDA Answers

- Work in progress
- Right now: Desktop scientific computing
- In the future there will be more

https://scda-answers.oist.jp

Storage Requests

- Bucket is our main storage
  - Initial allocation is ~50TB
  - **Can be increased** if needed
- Specialized storage solutions sometimes possible
- Other storage is fixed
  — can not be increased



- Today requests are ad-hoc:
  - Requests are informal
  - very short timeframe
  - No concern about data management, data sources etc.

This is a problem

Misconceptions

Belief:

- We have spare storage
- adding storage is easy
- keeping data is free

Misconceptions

Belief:

- We have spare storage
- adding storage is easy
- keeping data is free

Reality:

- We can't buy "spare" storage
  - everything needs to have concrete justification
    - future projected need
    - new unit allocations
  - **Based on what you tell us**

Misconceptions

Belief:

- We have spare storage
- adding storage is easy
- keeping data is free

Reality:

- New storage is a large capital investment
  - Needs scientific justification
  - Acquisition follows OIST procurement rules and calendar

  → takes ~months to a year

Misconceptions

Belief:

- We have spare storage
- adding storage is easy
- keeping data is free

Reality:

- Storage is a consumable
  - Hardware has ~5 year lifetime
  - Needs maintenance
  - Must be replaced periodically
- Storage is difficult, critical
  - specialists, maintenance contracts are necessary

Misconceptions

Belief:		Reality:
• We have spare storage	→	• all capacity is accounted for - existing and future units
• adding storage is easy	→	• adding capacity takes months - need scientific justification
• keeping data is free	→	• keeping data is expensive - maintenance, replacement

Data Management

Data at scale is a research resource

Small data:

- Desktop computer
- Bench loupe
- Rotation student

→ simple to get
little planning
simple management

Big data:

- Cluster
- Electron microscope
- Project research group

→ complex, expensive
lots of planning
heavy management burden

A new Workflow

1.Fill in the storage request form:

`https://groups.oist.jp/scs/request-storage`

2. Wait for us to respond:

- Your request is easy to accommodate:
  → you get the allocation
- Your request needs system changes:
  → Discuss with us, wait for implementation
- Your request can't be fulfilled
  → Talk with us to find a different solution

Fill in the Form

- You need to be the unit leader to ask for storage
- You can't ask for more storage on the in-cluster systems `Flash` or `Work`
- We ask you three questions; tell us as much as you can, and as much you think is needed
- **Also please clean up and archive data** in your existing allocation

#1: Storage Amount

- How much do you need?
- When do you need it?
- Are there any specific conditions?

What storage system is appropriate? How can we provide it to you?

- Be concrete
- Include future need if you know
- **be as early as possible**

  → but we only increase when you're actually running out...

Specific Conditions:

- could be anything
- *no guarantee* that we can fulfill it

  → don't ask unless important

#2: Data Plan

- What kind of data, and how much?
- It comes from where? How will you use it?
- How long do you need to store it?
- What will happen with it at the end?

Determine data flow and needed connectivity. Storage retention estimations.

- What other systems are you going to connect with
- What kind of bandwidth, intermediate storage needed
- Storage is not free over time — need to know the time frame
- Do you know what to do with the data when project ends and people leave?

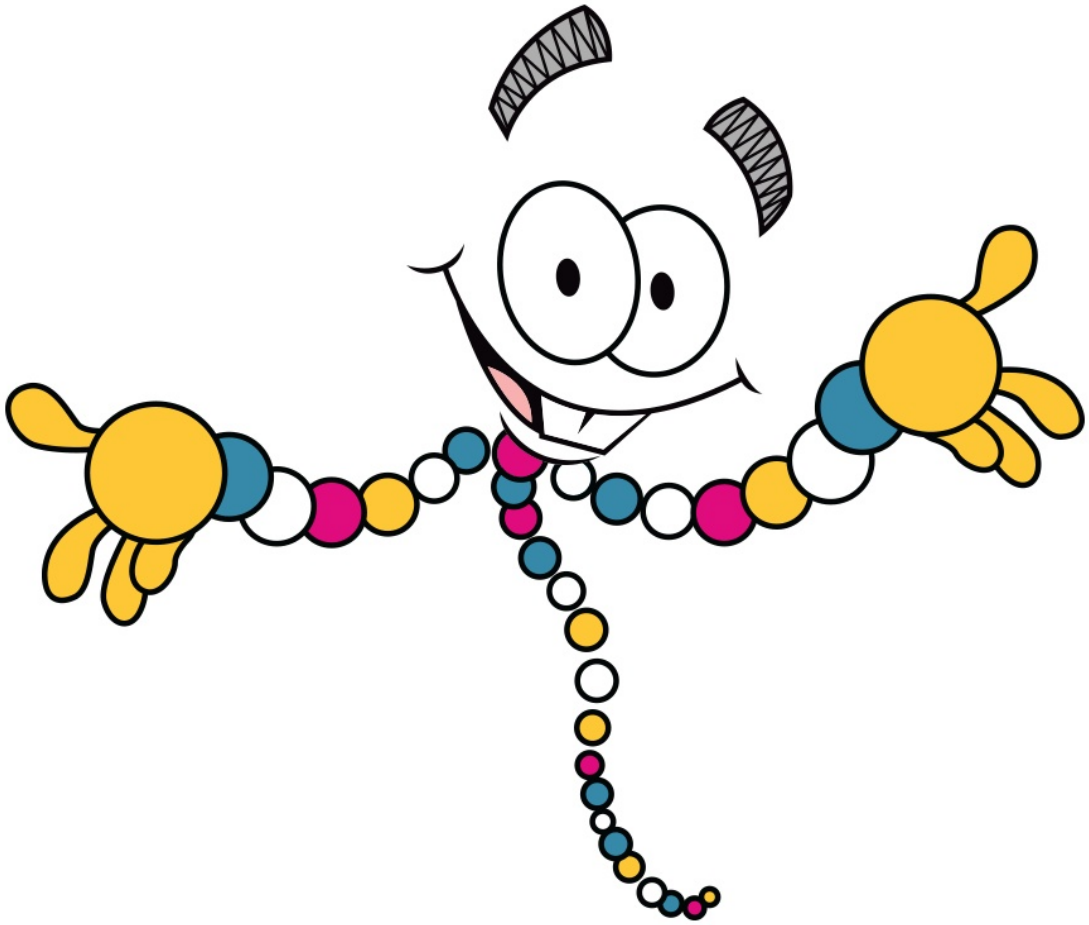
#3: Research Project

- Project title or description
- Ignore this for small or recurring request
- If you do have more info, you're very welcome to give it to us!

Justify our budget requests. Show the scientific impact of HPC investment.

Goal: Show how this helps OIST mission

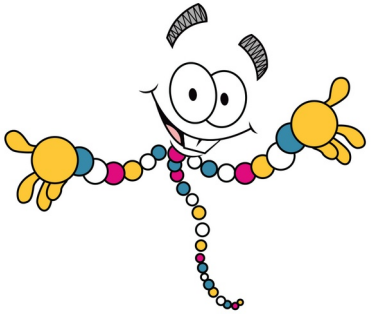
- Be brief
- not compulsory - use your own best judgement



OIST

Bioinformatics User Group

OIST's Bioinformatics User Group



Why?

- Targeted (bio)informatic **tools** (DebianMed collection, Other, AlphaFold, nanopore basecall)
- Access to wide range of curated bioinformatic **pipelines**: OIST's profile for Nextflow
- Overcome unit's storage limits for installation of **databases** (NCBI's nt, nr, etc.)

ssh deigo

```
## load modules
ml bioinfo-ugrp-modules
# load DebianMed
ml bioinfo-ugrp-modules DebianMed
```

```
# list Other softwares (= out of DebianMed distribution)
ml av Other
# list databases
ml av DB
```

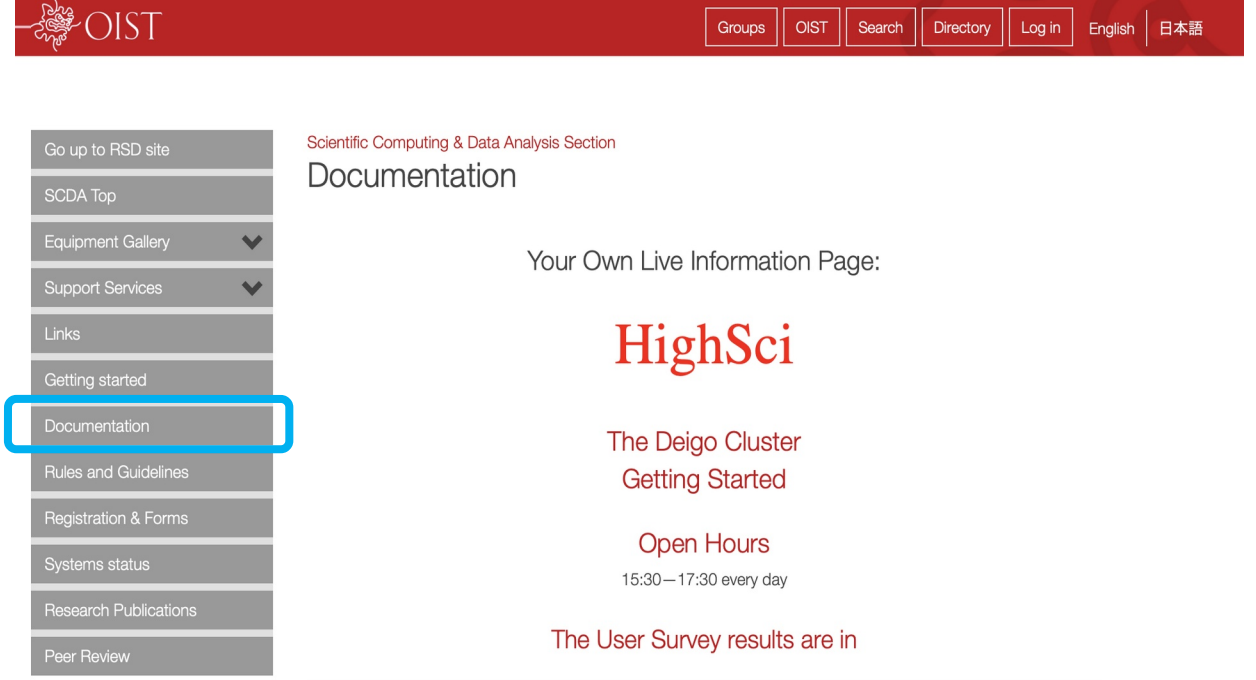
ssh saion

```
## load modules
module bioinfo-ugrp-modules
module available
```

(AlphaFold, Guppy, Rerio,…)

Where to find us?

On SCDA's Documentation:
https://groups.oist.jp/scs/documentation

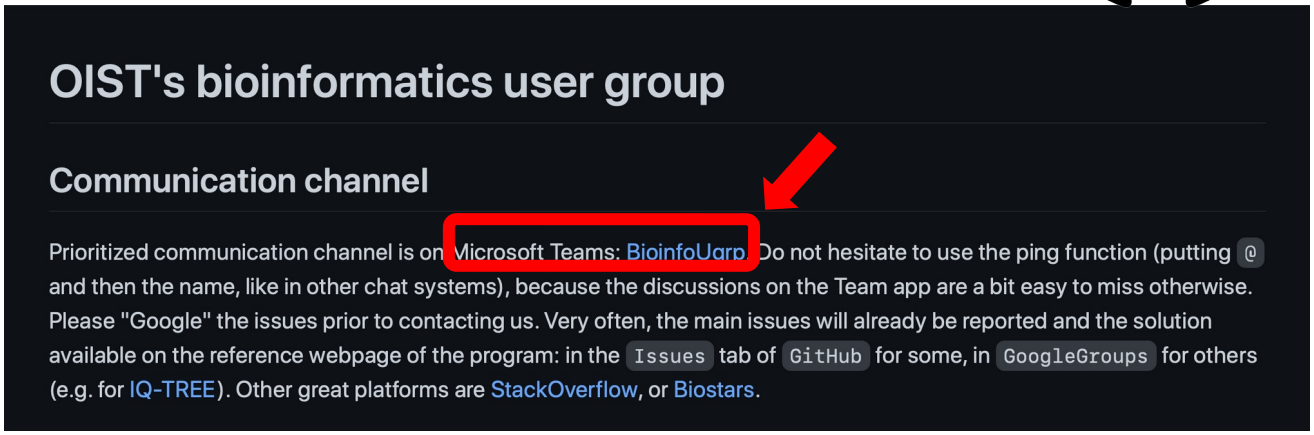


Repository GitHub:
https://github.com/oist/BioinfoUgrp

OIST's BioinfoUgrp

How to contact us?

Via Microsoft Teams
(looking at alternatives..)



SOON

Bimonthly meetings of the BioinfoUgrp

What: afternoon of Q&A

When: November-December

To be announced on Teams (?)

This and That

Comspace

- Share data between units at OIST
- Restricted data for a subgroup of lab members
- 5TB/unit
- GUI management

New Software

Geneious Prime now available

Copyright Infringement

A researcher was caught sharing movies using the OIST systems

Storage is *only* for research data!

- no copyrighted data
- no PC backups
- no personal files
- no unit administrative documents

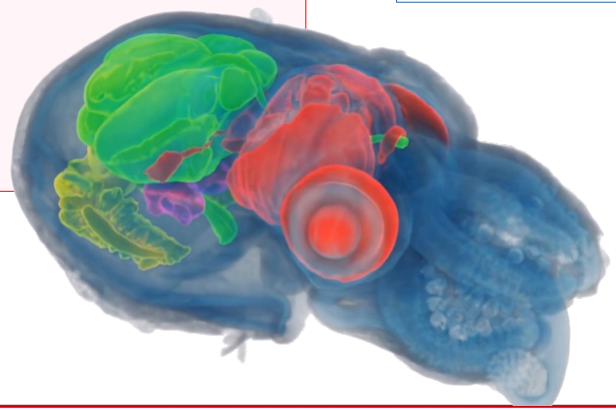
Software Week!

https://tida.oist.jp/announcements/software-week-back-oct-4th-oct-7th

Learn about your software!

One-on-one help and advice directly from the vendor!

Amira	Tue 4th, Wed 5th
Matlab	Wed 5th
COMSOL	Fri 7th



Introduction to HPC and Scientific Computing

Tuesday 4th and Thursday 6th
13-15, C700 or Zoom

Advanced Blender!

Wednesday 5th
13-15, C700 or Zoom