



「富岳」におけるOpen OnDemandの活用事例

中尾昌広（理化学研究所 計算科学研究センター）

背景

- HPCクラスタを用いるための前提知識が多く、利用するまでの学習コストが大きい
 - シェルなどのコマンドラインインタフェース
 - SSH鍵ペアの生成と公開鍵の登録
 - ジョブスケジューラ
- 計算ノード上でリモートデスクトップやJupyterLabなどのGUIアプリケーションを対話的に実行したい要望があるが、その手順は複雑
 - リモートデスクトップの設定や、SSHトンネリングの設定



```
top
Processes: 596 total, 2 running, 594 sleeping, 2418 threads      10:05:40
Load Avg: 2.15, 1.81, 1.63  CPU usage: 4.39% user, 3.8% sys, 92.51% idle
SharedLibs: 643M resident, 120M data, 40M linkedit.
MemRegions: 96880 total, 4270M resident, 425M private, 2994M shared.
PhysMem: 14G used (1893M wired, 364M compressor), 992M unused.
VM: 223T vsize, 4283M framework vsize, 160759(0) swapins, 248728(0) swapouts.
Networks: packets: 10719029/12G in, 2457839/525M out.
Disks: 8336373/93G read, 2912845/60G written.

PID    COMMAND    %CPU  TIME    #TH   #WQ   #PORT  MEM    PURG  CMPRS  PGRP
177    WindowServer 31.0  03:44:15 26    10    4438+  2247M  223M+  79M   177
0      kernel_task  6.0   01:53:35 484/8  0      0      2002M+ 0B     0B    0
78311  top          5.3   00:00.43 1/1    0      28     5761K  0B     0B    78311
78313  screencaptur 4.1   00:00.33 4      3      71     7618K  752K  0B    1803
1794   iTerm2      3.7   22:59.68 6      3      422-   226M+  13M-  83M   1794
```

新しいHPCクラスタを使う度に再学習が必要な場合もある

Open OnDemand <https://openondemand.org>

- HPCクラスタ用のWebポータルOpen OnDemandに着目
- オハイオ州立大学が中心となって開発
 - オープンソースソフトウェア
- WebブラウザからHPCクラスタの操作が可能
 - 他のソフトウェアのインストールは不要
 - ファイルのアップロード・ダウンロード、ジョブ投入と監視、CLIによるシェルアクセス
 - 計算ノード上でGUIアプリケーションを簡易に利用可能

HPC ASIA 2022のワークショップInternational Workshop on Arm-based HPC: Practice and Experience (IWAHPCE)の発表で Open OnDemandを知った (オーガナイザは佐藤三久先生)

“Onboarding Users to A64FX via Open OnDemand”, Aaron Jezghani et al.

Welcome to the supercomputer Fugaku

RIKEN R-CCS Center for Computational Science

Message of the day

Information

- Jul 24, 2023 Operation July 2023 Large-scale job execution period
- Jul 21, 2023 Operation Resource groups during the large scale job execution period
- Jul 19, 2023 Operation Occurrence of inaccessibility and poor response at login nodes and jobs due to a file system maintenance(vol0005)

Pending jobs

| | | | |
|-----------------------|---------------------|-----------------------|-------------------|
| fugaku-small 17867 | fugaku-large 915 | prepost-gpu1 12 | prepost-gpu2 0 |
| prepost-mem1 22 | prepost-mem2 3 | prepost-ondemand 0 | |

Fugaku Schedule

Today July 2023

| Sun | Mon | Tue | Wed | Thu | Fri | Sat |
|-----|-----|-------|-----|-----|-----|-------|
| 25 | 26 | 27 | 28 | 29 | 30 | Jul 1 |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 16 | 17 | 18 | 19 | 20 | 21 | 22 |
| 23 | 24 | 25 | 26 | 27 | 28 | 29 |
| 30 | 31 | Aug 1 | 2 | 3 | 4 | 5 |

Accounting (Updated at 2023/07/25 02:32:10 (JST))

| Group | Volume | Disk (GiB) | | | | Disk (inode) | | | | Resource (NH) | | | |
|----------|----------|------------|---------|---------|------|--------------|-------------|------------|------|---------------|--------|---------|------|
| | | Limit | Usage | Avail. | Rate | Limit | Usage | Avail. | Rate | Limit | Usage | Avail. | Rate |
| rccs-aot | /vol0400 | 5,120 | 608 | 4,512 | 11% | 1,500,000 | 265,259 | 1,234,741 | 7% | 527,360 | 81,666 | 445,693 | 5% |
| f-op | /vol0403 | 614,400 | 160,260 | 454,140 | 26% | 180,000,000 | 144,261,340 | 35,738,660 | 80% | - | - | - | - |
| ra030002 | /vol0403 | 5,120 | 1 | 5,119 | 0% | 1,500,000 | 2 | 1,499,998 | 0% | - | - | - | - |
| /home | /vol0400 | 20 | 5 | 15 | 25% | 200,000 | 11,619 | 188,381 | 5% | - | - | - | - |

Recently Used Apps

- Desktop
- OpenFOAM
- SCALE
- Jupyter

Passenger Apps

- Active Jobs
- Home Directory
- Gakunin RDM
- HPCI Storage
- Job Composer
- Fugaku Shell Access

Open OnDemandの導入

- 理研R-CCSでは2022年8月から「富岳」においてテスト運用、2023年5月から本番運用
- 設定ファイルなどはGitHubで公開しています
- https://github.com/RIKEN-RCCS/ondemand_fugaku
- Open OnDemandを「富岳」のジョブスケジューラ Fujitsu TCSに対応させた（「玄界」も同じ）
- ダッシュボードのカスタマイズ
- アプリケーションの追加（約50個）
 - HPCI共用ストレージとGakuNin RDMとのデータ共有機能
 - アプリケーションごとの設定ファイルの自動生成
- 運用状況について紹介

Welcome to the supercomputer Fugaku

RIKEN Center for Computational Science

Message of the day

Information

- Jul 24, 2023 Operation July 2023 Large-scale job execution period
- Jul 21, 2023 Operation Resource groups during the large scale job execution period
- Jul 19, 2023 Operation Occurrence of inaccessibility and poor response at login nodes and jobs due to a file system maintenance(vol0005)

Pending jobs

| | | | |
|-----------------------|---------------------|-----------------------|-------------------|
| fugaku-small 17867 | fugaku-large 915 | prepost-gpu1 12 | prepost-gpu2 0 |
| prepost-mem1 22 | prepost-mem2 3 | prepost-ondemand 0 | |

Fugaku Schedule

Today: July 2023

| Sun | Mon | Tue | Wed | Thu | Fri | Sat |
|-----|-----|-------|-----|-----|-----|-------|
| 25 | 26 | 27 | 28 | 29 | 30 | Jul 1 |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 16 | 17 | 18 | 19 | 20 | 21 | 22 |
| 23 | 24 | 25 | 26 | 27 | 28 | 29 |
| 30 | 31 | Aug 1 | 2 | 3 | 4 | 5 |

Accounting (Updated at 2023/07/25 02:32:10 (JST))

| Group | Volume | Disk (GiB) | | | | Disk (inode) | | | | Resource (NH) | | | |
|----------|----------|------------|---------|---------|------|--------------|-------------|------------|------|---------------|--------|---------|------|
| | | Limit | Usage | Avail. | Rate | Limit | Usage | Avail. | Rate | Limit | Usage | Avail. | Rate |
| rccs-aot | /vol0400 | 5,120 | 608 | 4,512 | 11% | 1,500,000 | 265,259 | 1,234,741 | 7% | 527,360 | 81,666 | 445,693 | 15% |
| f-op | /vol0403 | 614,400 | 160,260 | 454,140 | 26% | 180,000,000 | 144,261,340 | 35,738,660 | 80% | - | - | - | - |
| ra030002 | /vol0403 | 5,120 | 1 | 5,119 | 0% | 1,500,000 | 2 | 1,499,998 | 0% | - | - | - | - |
| /home | /vol0400 | 20 | 5 | 15 | 25% | 200,000 | 11,619 | 188,381 | 5% | - | - | - | - |

Recently Used Apps

- Desktop
- OpenFOAM
- SCALE
- Jupyter

Passenger Apps

- Active Jobs
- Home Directory
- Gakunin RDM
- HPCI Storage
- Job Composer
- Fugaku Shell Access

Fujitsu TCSへの対応 (1/2)

- Open OnDemand-2.0.27 (2022年6月リリース) が対応しているジョブスケジューラ
 - Slurm, PBS Pro, Torque, Grid Engineなど
- Fujitsu TCSに対応するためのコード変更 & Pull Request
 1. https://github.com/OSC/ood_core/pull/766
 2. <https://github.com/OSC/ondemand/pull/2194>
- **Open OnDemand-2.0.28 (2022年8月リリース) からFujitsu TCSもサポート**
 - リリースのタイミングの関係で、1. だけマージしてもらった
 - 1. だけでも、Fujitsu TCSは動作する
 - 2. はジョブの詳細表示に関するコードで、2.0.29で利用可能になった
 - 2.0.29以降も、富士通エンジニアの人と相談しつつ、細かなバグ修正をいくつか行っています
- 中尾昌広, 三浦信一, 山本啓二, ``スーパーコンピュータ「富岳」における HPCクラスタ用WebポータルOpen OnDemandの導入", 第186回HPC研究発表会, 2022年9月



Fujitsu TCSへの対応 (2/2)

SWoPP 2024

HPCシステム用ウェブポータルにおけるジョブスケジューラの抽象化

○谷澤 悠太・高橋 慧智・下村 陽一・滝沢 寛之 (東北大)

NQSVの
対応

- 様々なジョブスケジューラに対応するためのアダプタインタフェースが定義
- そのインタフェースで定義されている下記のメソッドをFujitsu TCS用の実装
- 実装言語はRubyで、コード量は約250行 (コメントを除く)

| | |
|---------------------|------------------------------|
| submit | ジョブの投入 |
| delete | ジョブの削除 |
| status | ジョブの状態を取得 |
| hold | ジョブのホールド |
| release | ホールドされたジョブの開放 |
| info | ジョブの情報を取得 |
| info_all | 全ジョブの情報を取得 |
| cluster_info | HPCクラスタのシステム情報を取得 |
| supports_job_arrays | バルクジョブのサポートの可否 |
| directive_prefix | ジョブスケジューラで用いられる接頭辞を取得 (#PJM) |

ダッシュボードの変更 (1/5)

- カスタマイズが可能

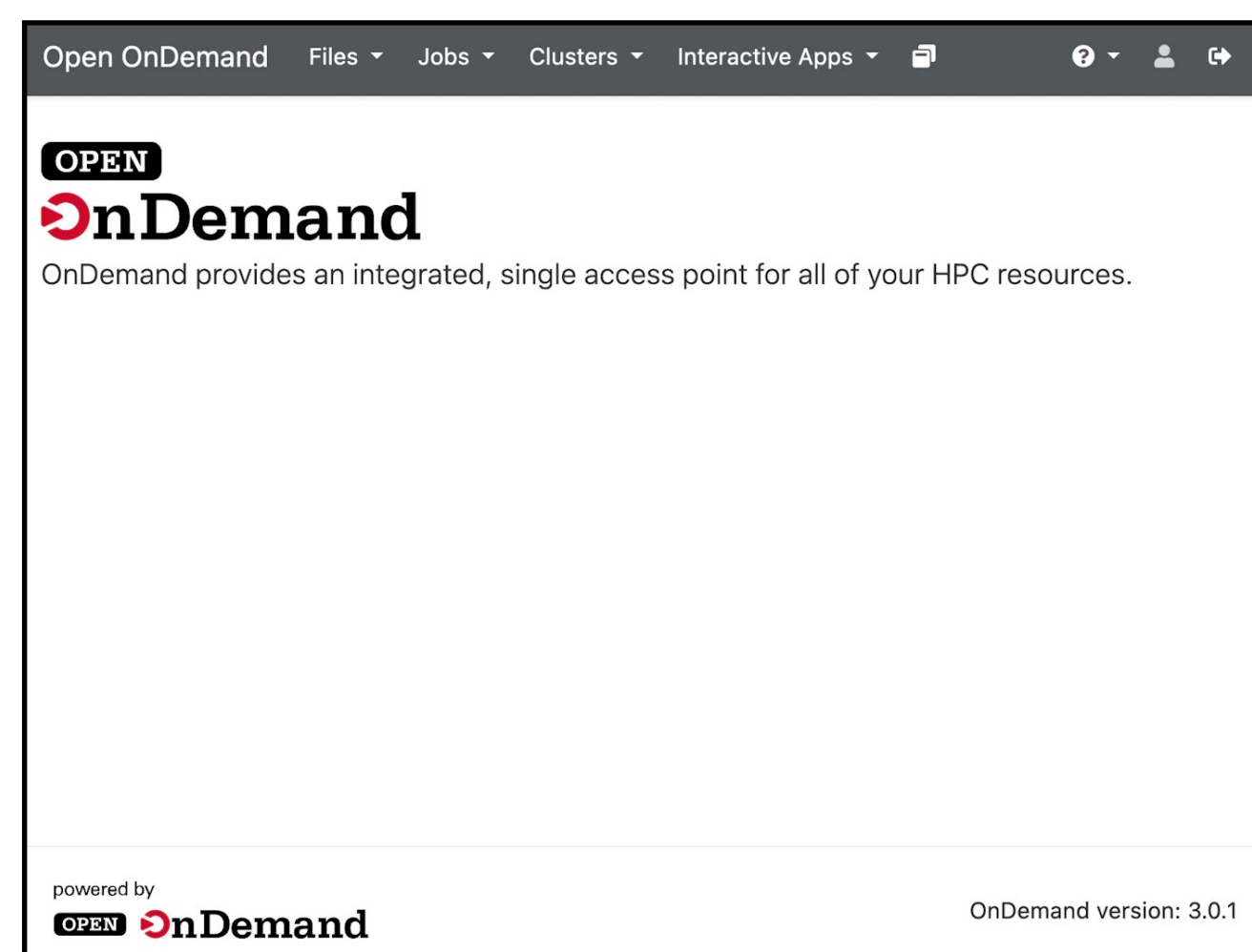
- (i) 既存のテンプレートを用いる

- ロゴの追加、アプリを種類ごとに表示

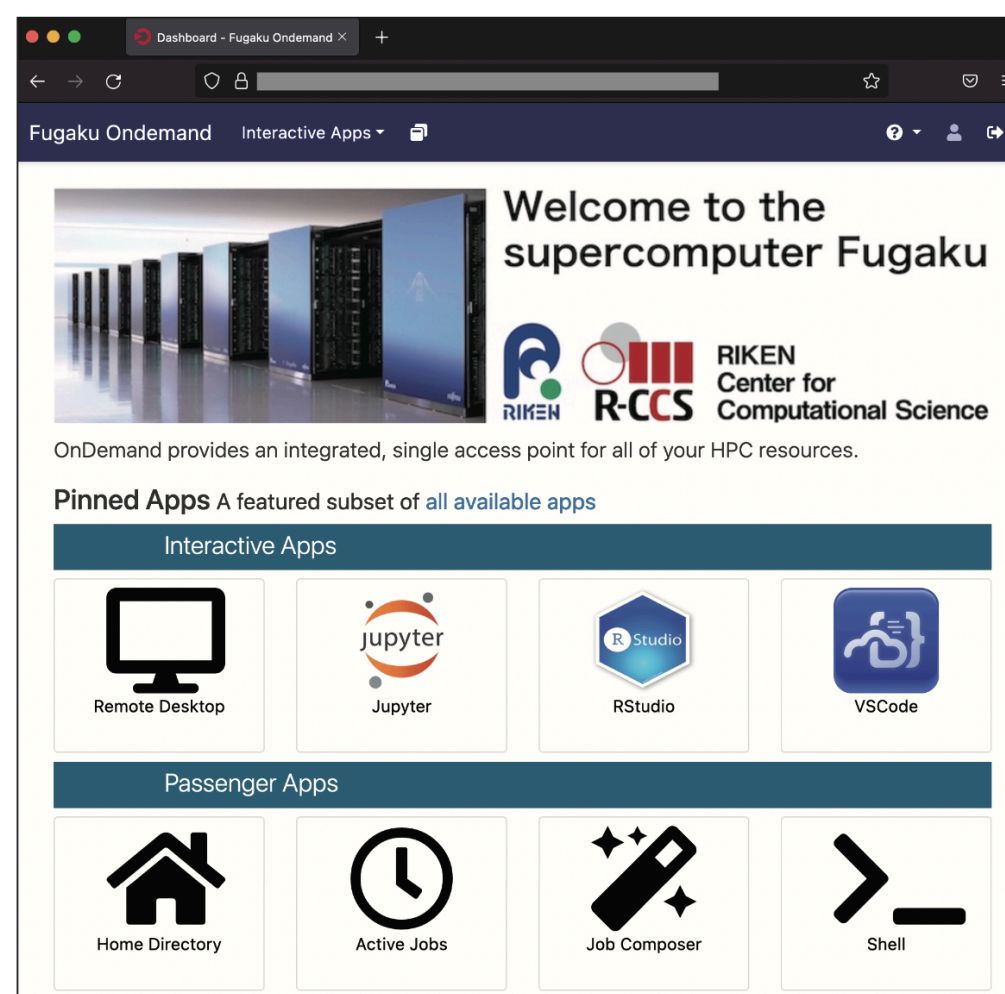
- (ii) テンプレートをeRubyで記述する

- ジョブを投入する際にユーザが欲しい情報を表示

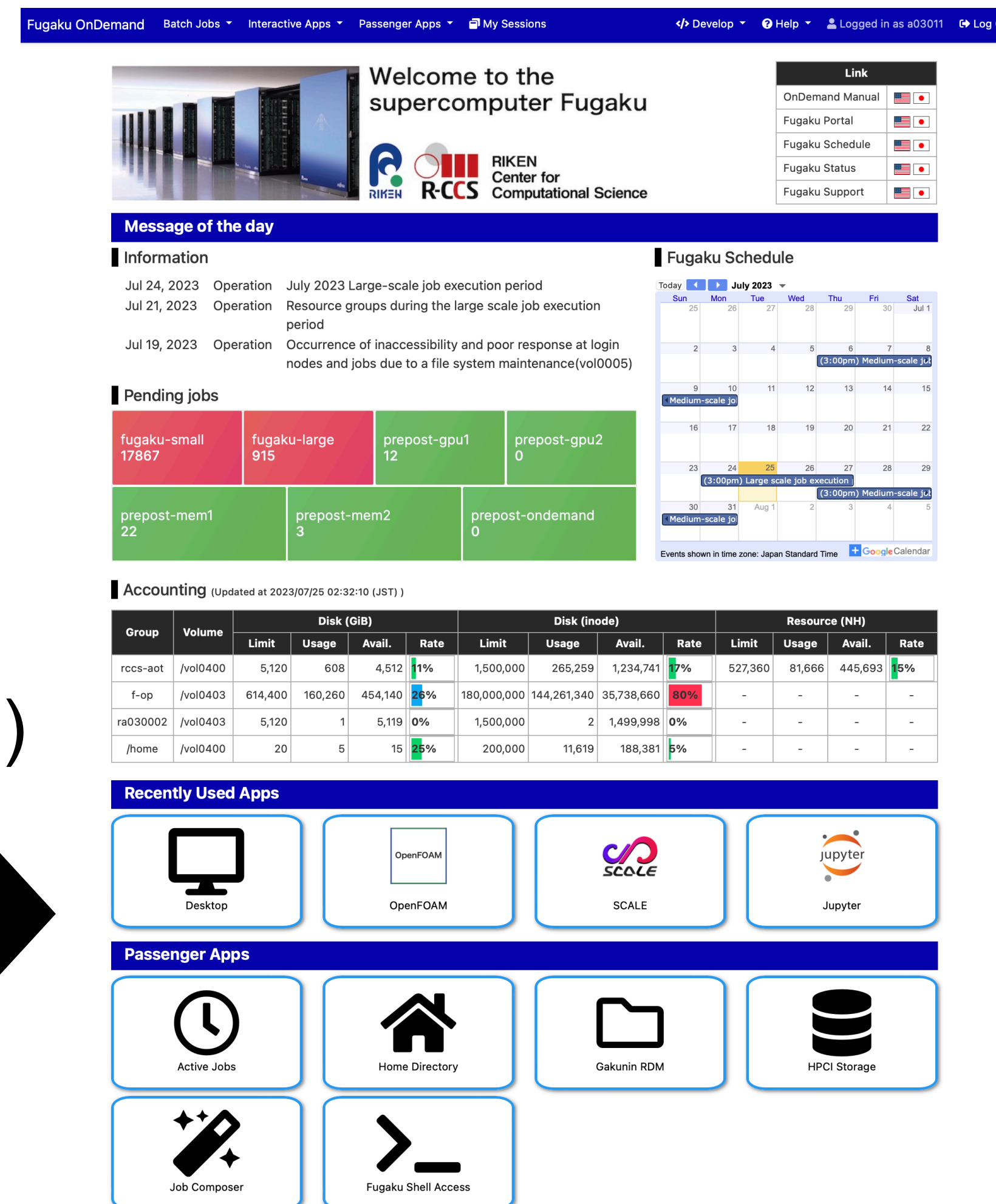
デフォルト



2022年9月ごろ



現在



ダッシュボードの変更 (2/5)

- A. 外部リンク (富岳マニュアルなど)
- B. 障害情報やオペレーション情報など
- C. 各キューの待ちジョブ数 (Grafanaを利用)
- D. 運用カレンダー (Googleカレンダーを利用)
- E. ユーザのディスクとバジレットの利用率
 - 他のアプリはナビゲーションバーから選択可能
- F. 最近使った計算ノードで動作するアプリ (最大4個)
 - ファイルのアップロードやジョブ監視など
- G. Open OnDemandサーバで動作するユーティリティ
 - ファイルのアップロードやジョブ監視など

現在

The screenshot shows the Fugaku OnDemand dashboard with the following annotations:

- A**: Points to the 'Link' menu in the top right corner, which includes links for OnDemand Manual, Fugaku Portal, Fugaku Schedule, Fugaku Status, and Fugaku Support.
- B**: Points to the 'Information' section, which displays operational messages such as 'July 2023 Large-scale job execution period' and 'Occurrence of inaccessibility and poor response at login nodes and jobs due to a file system maintenance(vol0005)'.
- C**: Points to the 'Pending jobs' section, which shows a bar chart of job counts for various queues: fugaku-small (17867), fugaku-large (915), prepost-gpu1 (12), prepost-gpu2 (0), prepost-mem1 (22), prepost-mem2 (3), and prepost-ondemand (0).
- D**: Points to the 'Fugaku Schedule' calendar, which shows a monthly view of job execution periods and maintenance events.
- E**: Points to the 'Accounting' table, which displays disk and resource usage statistics for various groups.
- F**: Points to the 'Recently Used Apps' section, which shows icons for Desktop, OpenFOAM, SCALE, and Jupyter.
- G**: Points to the 'Passenger Apps' section, which shows icons for Active Jobs, Home Directory, Gakunin RDM, HPCI Storage, Job Composer, and Fugaku Shell Access.

| Group | Volume | Disk (GiB) | | | | Disk (inode) | | | | Resource (NH) | | | |
|----------|----------|------------|---------|---------|------|--------------|-------------|------------|------|---------------|--------|---------|------|
| | | Limit | Usage | Avail. | Rate | Limit | Usage | Avail. | Rate | Limit | Usage | Avail. | Rate |
| rcs-aot | /vol0400 | 5,120 | 608 | 4,512 | 11% | 1,500,000 | 265,259 | 1,234,741 | 17% | 527,360 | 81,666 | 445,693 | 15% |
| f-op | /vol0403 | 614,400 | 160,260 | 454,140 | 26% | 180,000,000 | 144,261,340 | 35,738,660 | 80% | - | - | - | - |
| ra030002 | /vol0403 | 5,120 | 1 | 5,119 | 0% | 1,500,000 | 2 | 1,499,998 | 0% | - | - | - | - |
| /home | /vol0400 | 20 | 5 | 15 | 25% | 200,000 | 11,619 | 188,381 | 5% | - | - | - | - |

ダッシュボードの変更 (3/5)

```

<%-
require 'utils.rb'
info = dashboard_info("/system/ood/motd/info.txt")
-%>

:
<h4>Information</h4>
<table>
<%- info.each do |i| -%>
  <tr>
    <%- c = i.split(":") -%>
    <td><%= c[0] %></td><td><%= c[1] %></td><td><%= c[2] %></td>
  </tr>
<%- end -%>
</table>

```

Information

| | | |
|--------------|-----------|-------------------------------------------------------------------------------------------------------------------|
| Jul 24, 2023 | Operation | July 2023 Large-scale job execution period |
| Jul 21, 2023 | Operation | Resource groups during the large scale job execution period |
| Jul 19, 2023 | Operation | Occurrence of inaccessibility and poor response at login nodes and jobs due to a file system maintenance(vol0005) |

Bのinfo.txtの内容は、cronを使って1日1回の頻度で、富岳ポータルサイトから取得している。

Eのディスク使用率とバジェット情報も基本的に同じ。

Accounting (Updated at 2023/07/25 02:32:10 (JST))

E

| Group | Volume | Disk (GiB) | | | | Disk (inode) | | | | Resource (NH) | | | |
|----------|----------|------------|---------|---------|------|--------------|-------------|------------|------|---------------|--------|---------|------|
| | | Limit | Usage | Avail. | Rate | Limit | Usage | Avail. | Rate | Limit | Usage | Avail. | Rate |
| rccs-aot | /vol0400 | 5,120 | 608 | 4,512 | 11% | 1,500,000 | 265,259 | 1,234,741 | 17% | 527,360 | 81,666 | 445,693 | 15% |
| f-op | /vol0403 | 614,400 | 160,260 | 454,140 | 26% | 180,000,000 | 144,261,340 | 35,738,660 | 80% | - | - | - | - |
| ra030002 | /vol0403 | 5,120 | 1 | 5,119 | 0% | 1,500,000 | 2 | 1,499,998 | 0% | - | - | - | - |
| /home | /vol0400 | 20 | 5 | 15 | 25% | 200,000 | 11,619 | 188,381 | 5% | - | - | - | - |

ダッシュボードの変更 (4/5)

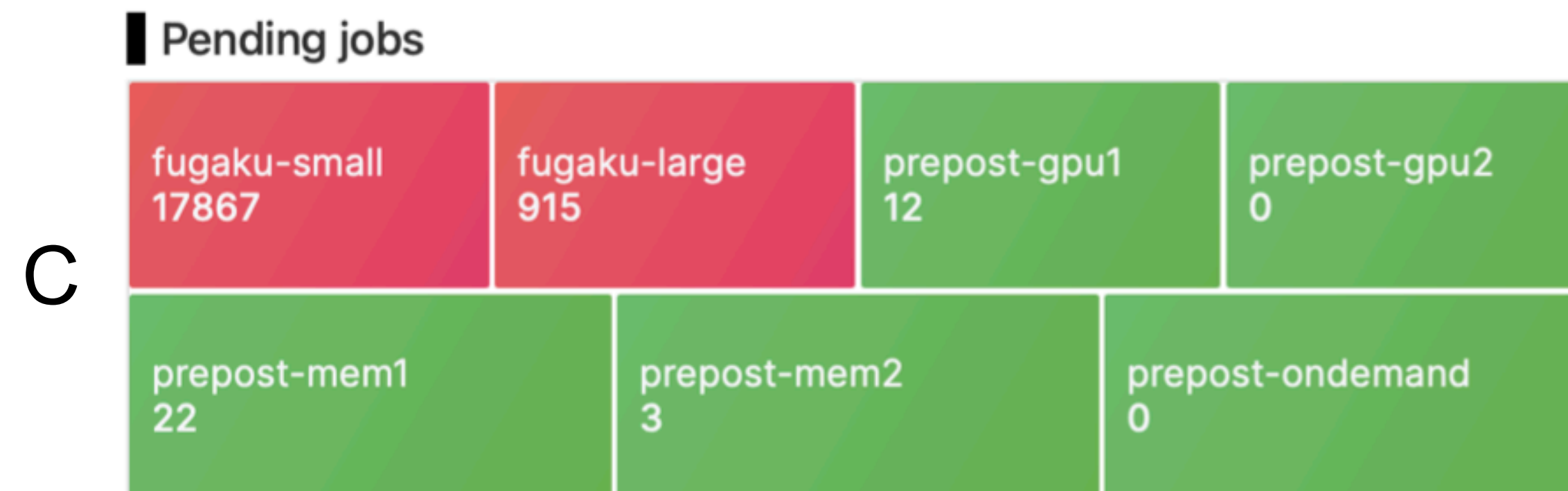
```
<%-
require 'time'

# URL of Dashboard
dashboard_url = "https://status.fugaku.r-ccs.riken.jp/d-solo/cf06d886-
e672-41d8-a587-85ccb32fce7e/
5a6f6KGM5b6F44Gh44K444On44OW5pWw?
orgId=1&theme=light&panelId=2"

# Get the current time as epoch milliseconds.
now = (Time.now.to_f * 1000).to_i

# Get time one month ago as epoch milliseconds
one_month_ago = now - (30 * 24 * 60 * 60 * 1000)
-%>
:
<h4>Pending Jobs</h4>
<iframe src="<%= dashboard_url %>&to=<%= now %>&from=<%=
one_month_ago %>" frameborder="0" width="100%" height="225px">
</iframe>
```

CとDはGrafanaとGoogleカレンダーが、元々グラフを外部のHTMLにポーティングする機能を提供している



Fugaku Schedule



ダッシュボードの変更 (5/5)

F. 最近使った計算ノードで動作するアプリ (最大4個)

- 既存のテンプレートの機能を利用
- 既存のテンプレートでは、アイコンをクリックすると前回実行したパラメータで再実行されてしまうため、テンプレートを変更し、異なるパラメータを設定できるようにした

G. Open OnDemandサーバで動作するPassenger Apps

- デフォルト + HPCI共用ストレージ接続アプリ + GakuNin RDM接続アプリなど

現在

The screenshot shows the Fugaku OnDemand dashboard with the following sections:

- Link**: OnDemand Manual, Fugaku Portal, Fugaku Schedule, Fugaku Status, Fugaku Support.
- Message of the day**: Information (July 24, 2023: July 2023 Large-scale job execution period; July 21, 2023: Resource groups during the large scale job execution period; July 19, 2023: Occurrence of inaccessibility and poor response at login nodes and jobs due to a file system maintenance(vol0005)), Fugaku Schedule (calendar view).
- Pending jobs**: fugaku-small (17867), fugaku-large (915), prepost-gpu1 (12), prepost-gpu2 (0), prepost-mem1 (22), prepost-mem2 (3), prepost-ondemand (0).
- Accounting** (Updated at 2023/07/25 02:32:10 (JST)):

| Group | Volume | Disk (GiB) | | | | Disk (inode) | | | | Resource (NH) | | | |
|----------|----------|------------|---------|---------|------|--------------|-------------|------------|------|---------------|--------|---------|------|
| | | Limit | Usage | Avail. | Rate | Limit | Usage | Avail. | Rate | Limit | Usage | Avail. | Rate |
| rccs-aot | /vol0400 | 5,120 | 608 | 4,512 | 11% | 1,500,000 | 265,259 | 1,234,741 | 17% | 527,360 | 81,666 | 445,693 | 15% |
| f-op | /vol0403 | 614,400 | 160,260 | 454,140 | 26% | 180,000,000 | 144,261,340 | 35,738,660 | 80% | - | - | - | - |
| ra030002 | /vol0403 | 5,120 | 1 | 5,119 | 0% | 1,500,000 | 2 | 1,499,998 | 0% | - | - | - | - |
| /home | /vol0400 | 20 | 5 | 15 | 25% | 200,000 | 11,619 | 188,381 | 6% | - | - | - | - |

- Recently Used Apps**: Desktop, OpenFOAM, SCALE, Jupyter.
- Passenger Apps**: Active Jobs, Home Directory, GakuNin RDM, HPCI Storage, Job Composer, Fugaku Shell Access.

アプリケーションの種類

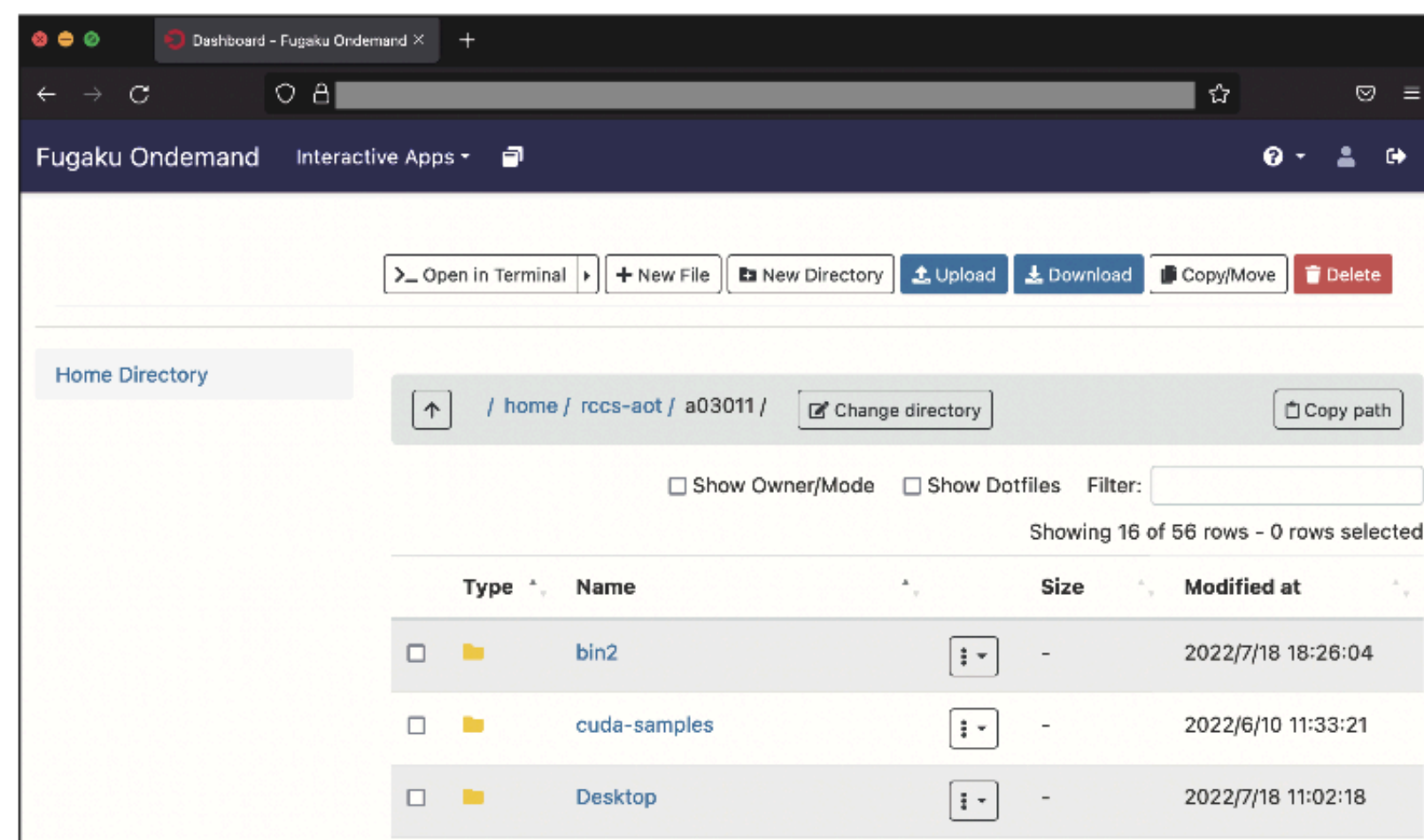
- Passenger Application : Open OnDemandがインストールされたWebサーバで動作するアプリケーション (ファイルのアップロードなど)
- Interactive Application : 計算ノード上で動作するアプリケーション
 - リモートデスクトップやJupyterLabなどの対話型GUIアプリケーション
 - 通常のバッチジョブ

アイコンをクリックして
アプリケーションを起動する

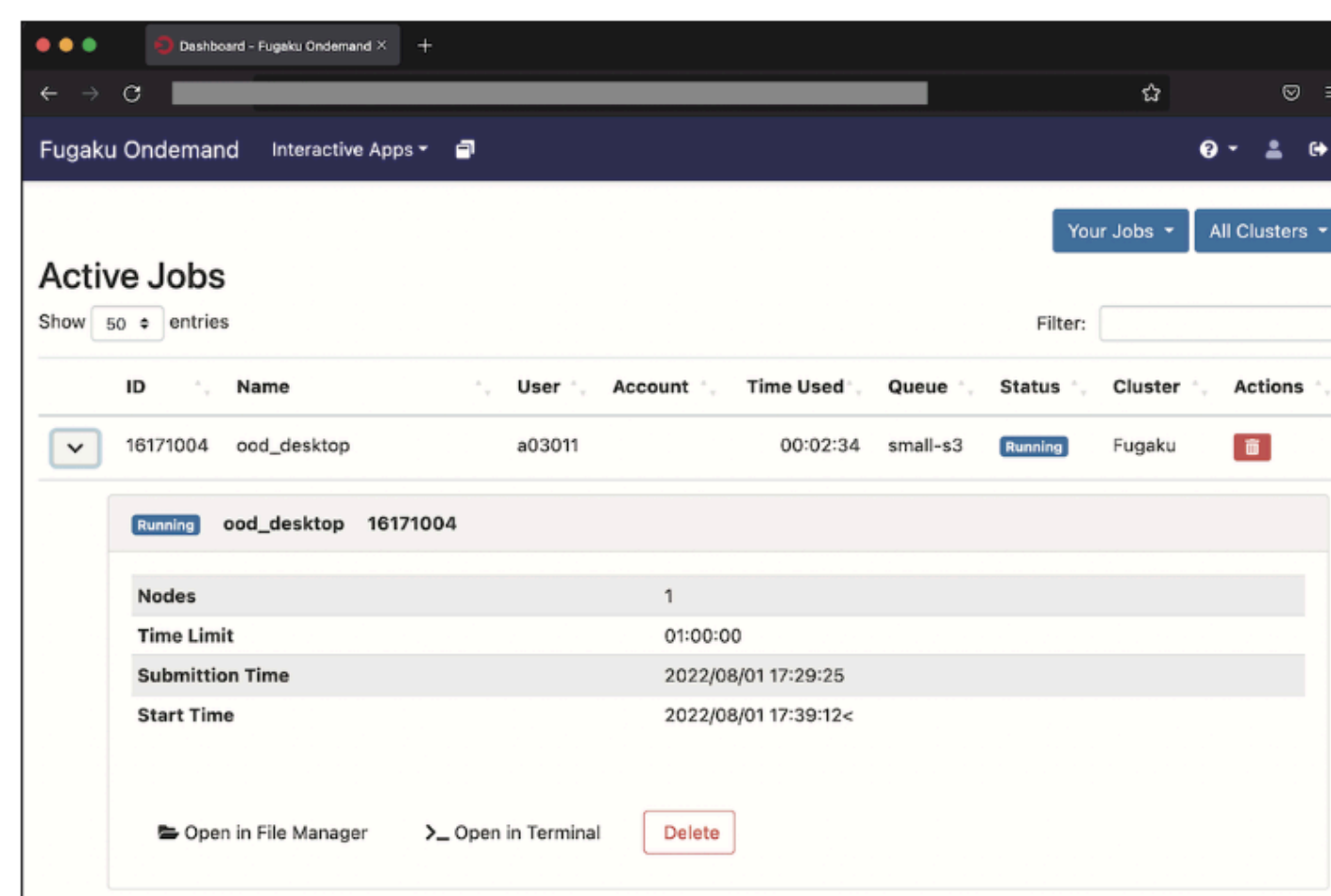
The screenshot shows the Fugaku OnDemand portal interface. At the top, there is a navigation bar with links for Fugaku OnDemand, Batch Jobs, Interactive Apps, Passenger Apps, My Sessions, and a user profile section. Below the navigation bar is a "Welcome to the supercomputer Fugaku" banner with the RIKEN R-CCS logo. A "Link" table provides quick access to various resources. The main content area is divided into several sections: "Message of the day", "Information" (with a table of recent operations), "Pending jobs" (with a bar chart showing job counts for different resource groups), "Fugaku Schedule" (with a calendar view), and "Accounting" (with a table of resource usage). At the bottom, there are sections for "Recently Used Apps" and "Passenger Apps", each containing icons for various applications like Desktop, OpenFOAM, SCALE, Jupyter, Active Jobs, Home Directory, Gakunin RDM, HPCI Storage, Job Composer, and Fugaku Shell Access.

| Group | Volume | Disk (GiB) | | | | Disk (inode) | | | | Resource (NH) | | | |
|----------|----------|------------|---------|---------|------|--------------|-------------|------------|------|---------------|--------|---------|------|
| | | Limit | Usage | Avail. | Rate | Limit | Usage | Avail. | Rate | Limit | Usage | Avail. | Rate |
| rccs-aot | /vol0400 | 5,120 | 608 | 4,512 | 11% | 1,500,000 | 265,259 | 1,234,741 | 7% | 527,360 | 81,666 | 445,693 | 15% |
| f-op | /vol0403 | 614,400 | 160,260 | 454,140 | 26% | 180,000,000 | 144,261,340 | 35,738,660 | 80% | - | - | - | - |
| ra030002 | /vol0403 | 5,120 | 1 | 5,119 | 0% | 1,500,000 | 2 | 1,499,998 | 0% | - | - | - | - |
| /home | /vol0400 | 20 | 5 | 15 | 25% | 200,000 | 11,619 | 188,381 | 5% | - | - | - | - |

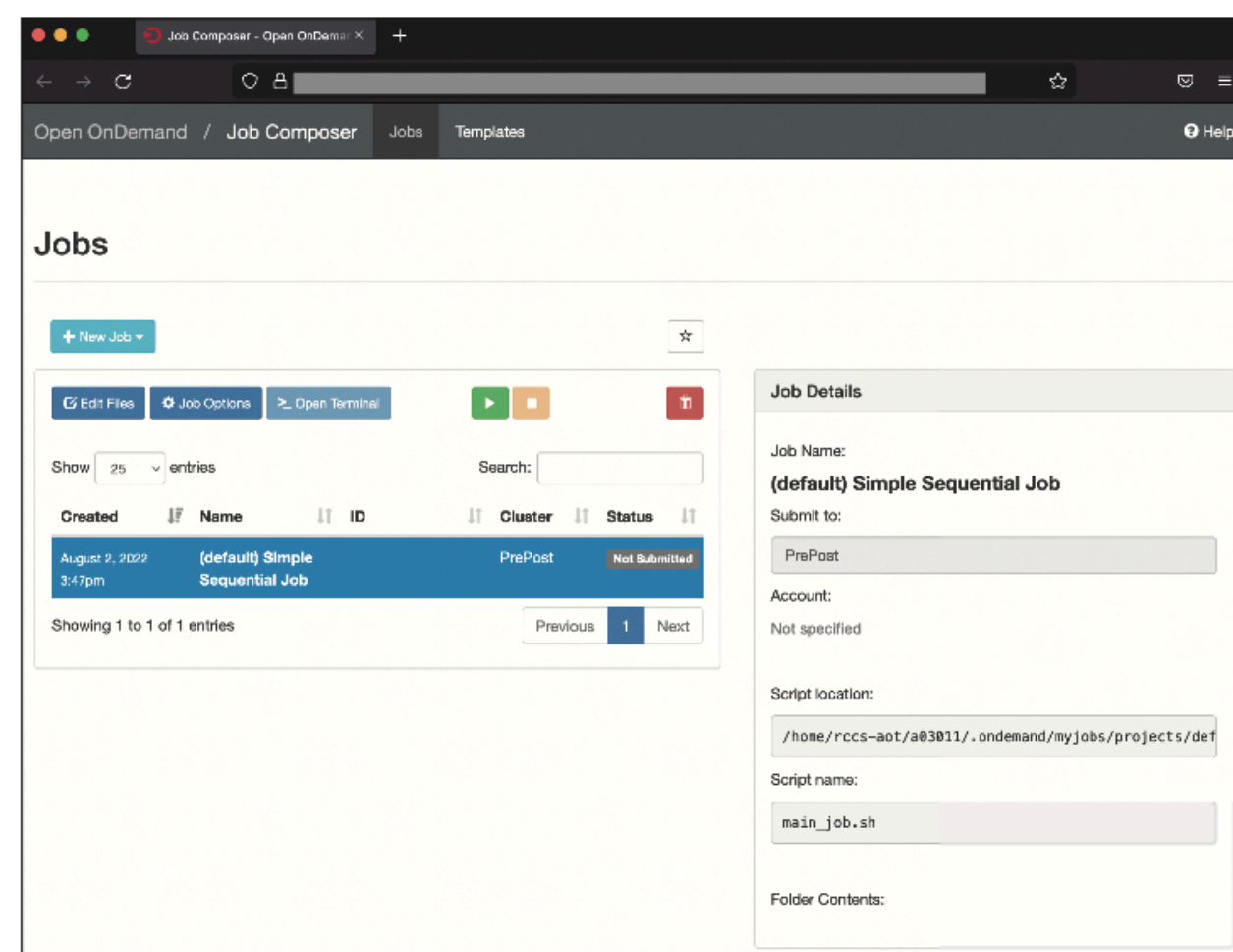
デフォルトのPassenger Apps



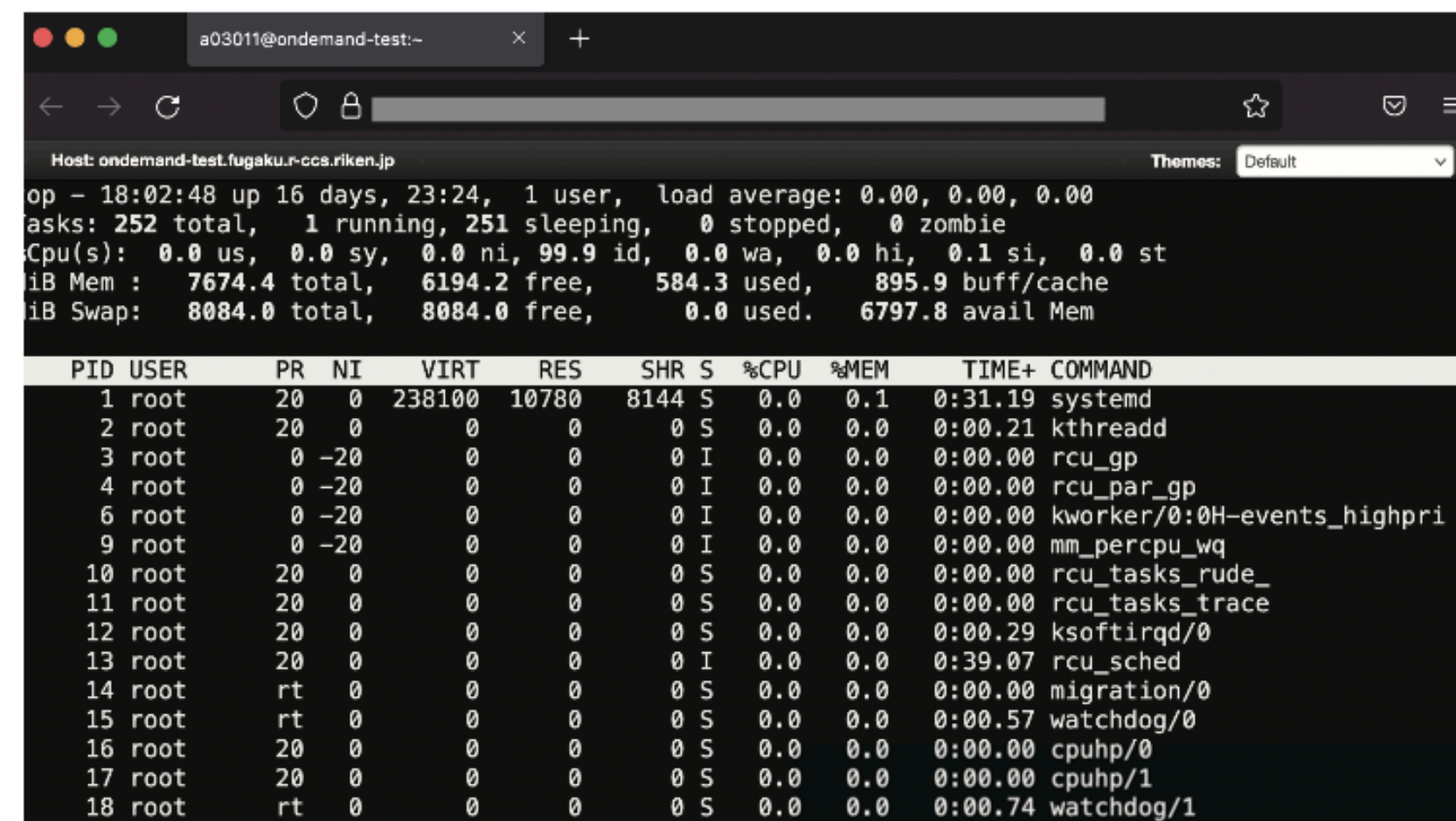
Home Directory



Active Jobs



Job Composer



Shell

- Home Directory : ファイルのアップロード・ダウンロード・編集が可能。rcloneと組み合わせることでAmazon S3などの外部ストレージとの連携も可能。

- Active Jobs : ジョブの監視

- Job Composer : ジョブの投入(使いづらいため無効化して、代わりにアプリを開発した。次の発表で紹介)

- Shell : Webターミナル

アプリケーションの追加 (1/2)

- 計算ノード上で動作するアプリケーションの管理方法
 - 「富岳」にインストール済のアプリケーション
 - ・ spack or moduleといった管理ツールで管理されている
 - ・ Open OnDemandが生成するジョブスクリプト内でspack or moduleコマンドを呼び出す
 - 「富岳」に未インストールのアプリケーション (リモートデスクトップなど)
 - ・ Singularityを使ってコンテナイメージを作成
 - ・ Open OnDemandが生成するジョブスクリプト内でsingularityコマンドを呼び出す

アプリケーションの追加 (2/2)

- GUI

| Category | Application |
|-------------|----------------------------------------------------------------------------------------------------------------------------------------|
| Development | Remote Desktop, JupyterLab, MATLAB*, VSCode, Terminal |
| Profiler | NVIDIA Visual Profiler*, NVIDIA Nsight Compute*, NVIDIA Nsight Systems*, Vampir* |
| Viewer | AVS/Express*, C-Tools, GaussView*, Gnuplot, GrADS, ImageJ, OVITO, Paraview, PyMOL, SALMON view, Smokeview, VESTA, VMD, VisIt, XCrySDen |
| Workflow | WHEEL |

Singularityで管理されたアプリケーション ※ 商用アプリケーション

- バッチジョブ

| Category | Application |
|------------------------------|------------------------------------------------------------------------------|
| Climate | SCALE |
| Computer Aided Engineering | FDS, FFVHC-ACR, FrontFlow (blue/X), FrontISTR, OpenFOAM (Foundation/OpenCFD) |
| Condensed Matter Physics | ALAMODE, AkaiKKR, HΦ, mVMC, OpenMX, PHASE/0, Quantum Espresso, SALMON |
| Molecular Dynamics | GENESIS, GROMACS, LAMMPS, MODYLAS, PIMD |
| Quantum Chemistry | ABINIT-MP, Gaussian*, NTCChem, SMASH |
| Quantum Simulation | braket |
| Experimental Data Processing | KIERTÄÄ |

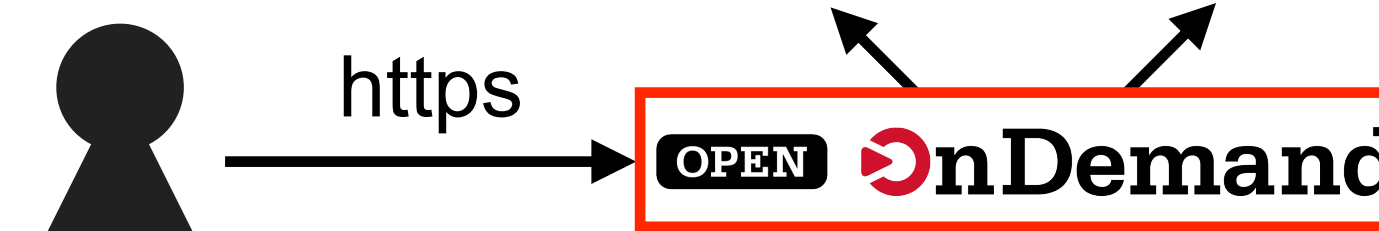
Webフォーム管理 (1/3)

- 理研R-CCSは「富岳」だけでなくプリポスト環境も提供
- プリポスト環境は可視化やデータ変換等を行うため、GPUや大容量メモリを搭載している
- **Open OnDemandサーバは共通**
- 「富岳」のCPUはARMアーキテクチャなのに対し、プリポスト環境のCPUはx86_64アーキテクチャ
 - そのため、アプリケーションのSingularityコンテナはアーキテクチャ毎に用意している
- 「富岳」とプリポスト環境の各リソースグループと設定項目の最大値は右表の通り
- プリポスト環境では1台に複数のジョブが実行可能なので、CPUコア数などを指定 (ただし1ノードのみ)

「富岳」



プリポスト環境



| Queue | Time (hours) | Nodes |
|--------------|--------------|--------|
| fugaku-small | 72 | 384 |
| fugaku-large | 24 | 12,288 |

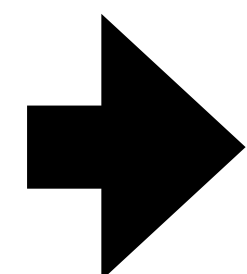
| Queue | Time (hours) | CPU Cores | Memory (GB) | GPUs |
|------------------|--------------|-----------|-------------|------|
| prepost-gpu1 | 3 | 72 | 186 | 2 |
| prepost-gpu2 | 24 | 36 | 93 | 2 |
| prepost-mem1 | 3 | 224 | 6,045 | - |
| prepost-mem2 | 24 | 56 | 1,511 | - |
| prepost-ondemand | 720 | 8 | 32 | - |

Webフォーム管理 (2/3)

- アプリケーション毎にWebフォームを設定する必要があり、管理が面倒
- /var/www/ood/apps/sys/[アプリケーション]/form.yml.erb
- 各アプリケーションでform.yml.erbはほぼ同じなので自動生成する

form.yml.erb (変更前)

```
gpu1_memory:  
  label: Memory (10 - 186 GB)  
  widget: number_field  
  value: 10  
  min: 10  
  max: 186  
  step: 1  
  required: true  
:
```



form.yml.erb (変更後)

```
<%-  
  require 'utils.rb'  
-%>  
<%= gpu1_memory(10) %>  
:
```

アプリケーションによって
異なる数値は引数で設定

utils.rb

```
def gpu1_memory(min = 5, max = 186)  
  $attr <<<<"EOF"  
  gpu1_memory:  
    label: Memory (#{min} - #{max} GB)  
    widget: number_field  
    value: #{min}  
    min: #{min}  
    max: #{max}  
    step: 1  
    required: true  
  EOF  
end
```

変更前は300行程度 x 50個だったが、
変更後は30行程度 x 50個に短縮できた

Webフォーム管理 (3/3)

fugaku-small キュー

Desktop
This app will launch an [Xfce](#) desktop.

Queue
fugaku-small

Group
rccs-aot

Elapsed time (1 - 72 hours)
1

Number of nodes (1 - 384)
1

Total number of processes (1 - 18,432)
1

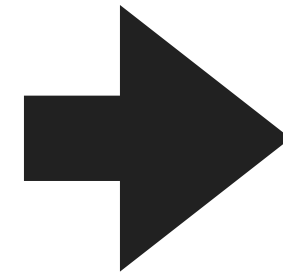
Total number of processes \leq Number of nodes x 48.

Execution mode
Normal

Please refer to the manual for details ([English](#) or [Japanese](#)).

Email (You will receive an email when it starts)

Launch



prepost-gpu1 キュー

Desktop
This app will launch an [Xfce](#) desktop.

Queue
prepost-gpu1

Elapsed time (1 - 3 hours)
1

Number of CPU cores (1 - 72)
1

Required memory (10 - 186 GB)
10

Number of GPUs (0 - 2)
0

Email (You will receive an email when it starts)

Launch

表示される項目が動的に変化する

Webフォーム管理の細かい話 (1/2)

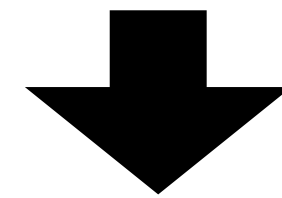
- 2025年1月にリリースされたOpen OnDemand 4.0から、新しい機能としてdata-labelが追加
<https://discourse.openondemand.org/t/easier-dynamic-label-changes/3241> (この機能は私がリクエストした)

```
node_type:  
  widget: select  
  options:  
    - [ 'small', 'small', data-max-cores: 4, data-label-cores: 'Number of Cores (1-4)' ]  
    - [ 'large', 'large', data-max-cores: 16, data-label-cores: 'Number of Cores (1-16)']  
  
cores:  
  widget: "number_field"  
  min: 1
```

Node Type

small

Number of Cores (1-4)



Node Type

large

Number of Cores (1-16)

Webフォーム管理の細かい話 (2/2)

- 以前のOpen OnDemandで似たようなことを行う場合

```
node_type:
  widget: select
  options:
    - [ 'small', 'small', data-hide-large-cores: true ]
    - [ 'large', 'large', data-hide-small-cores: true ]

small_cores:
  widget: "number_field"
  title: "Number of cores (1 - 4)"
  min: 1
  max: 4

large_cores:
  widget: "number_field"
  title: "Number of cores (1 - 16)"
  min: 1
  max: 16
```

- ラベル毎にWidgetを定義していたので、設定ファイルが非常に長くなっていた
- 運用後に設定変更することが難しい
 - 例えば、各コアの最小値の設定を2に変更した後に、node_typeでsmallを選択して、ジョブを投入しようとしても、投入できない
 - large_coresで前回に実行したコアの値（1）がキャッシュされているから。smallが選択されているとlarge_coresは隠れているので、問題特定が難しい

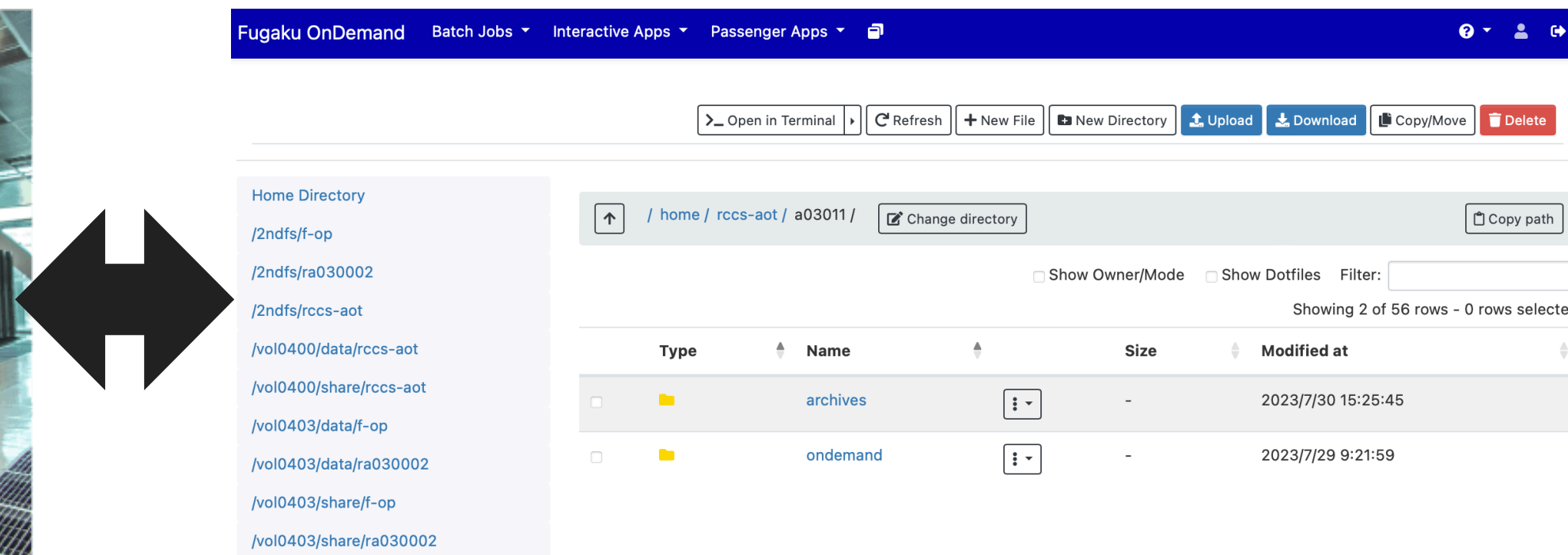
これらの問題が4.0で解決した！！

HPCI共用ストレージとGakuNin RDM (1/4)

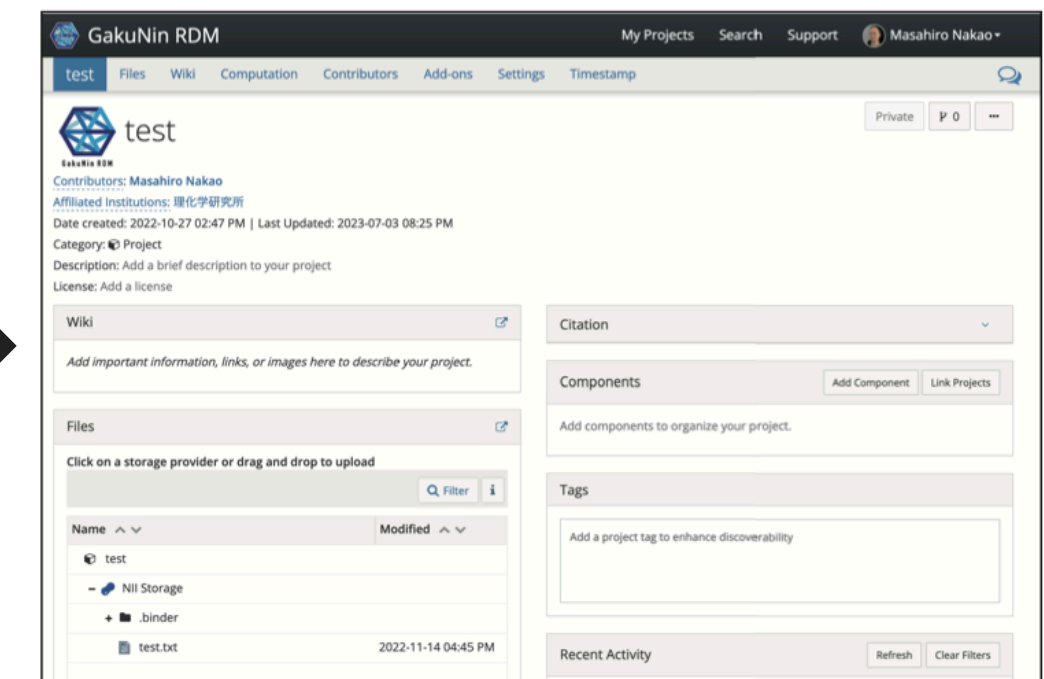
- 外部ストレージであるHPCI共用ストレージとGakuNin RDMとのデータ共有機能の開発
- これまではコマンドラインなどで行ってきた「富岳」と上の2つの外部ストレージとの間のデータ転送をOpen OnDemandを通して行うことが可能になった
- 注：Open OnDemand 3.0からrcloneを利用した外部ストレージ（Amazon S3など）とのデータ転送が可能になっているが、rcloneは上の2つの外部ストレージには対応していない



HPCI共用ストレージ



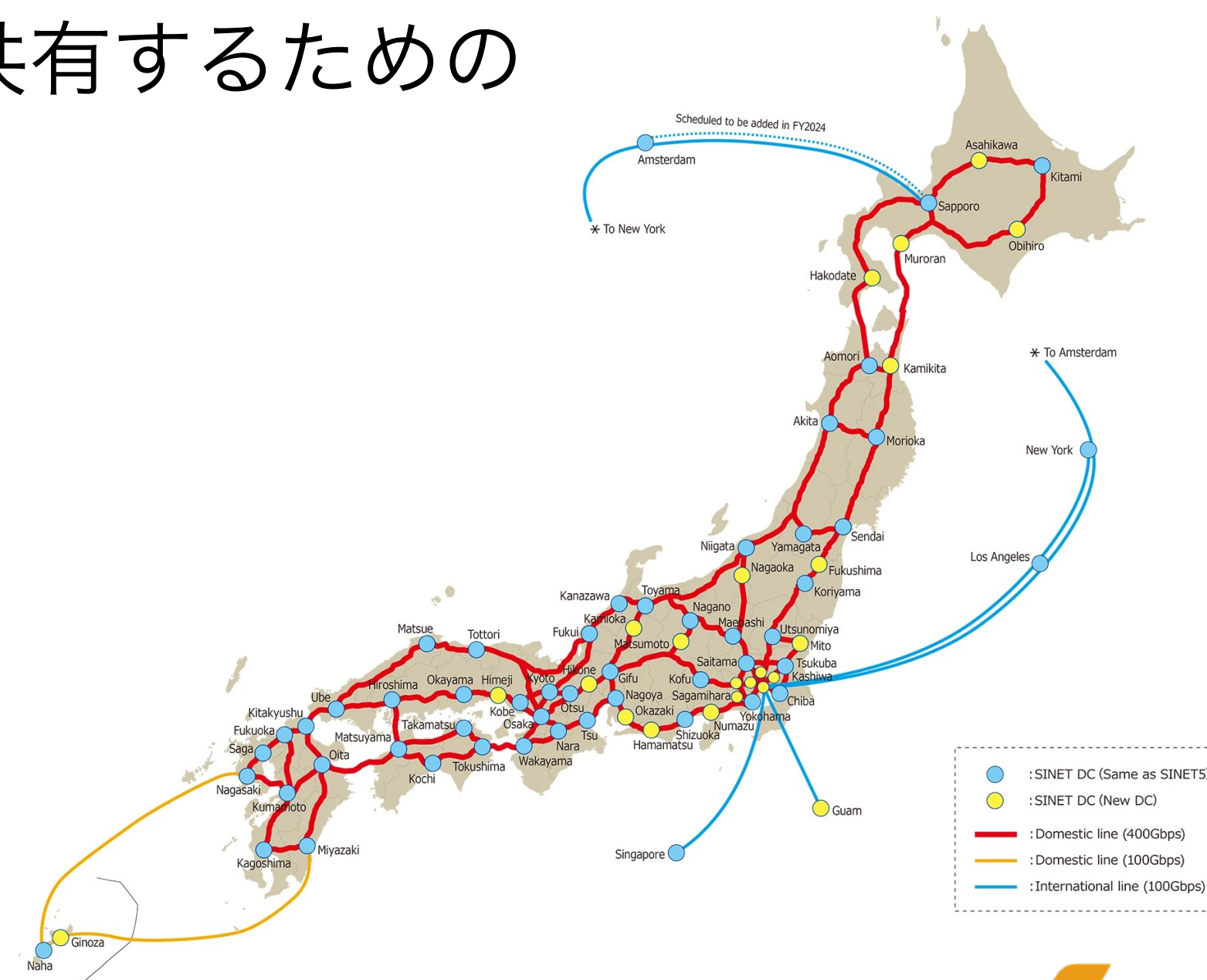
Open OnDemand on Fugaku



GakuNin RDM

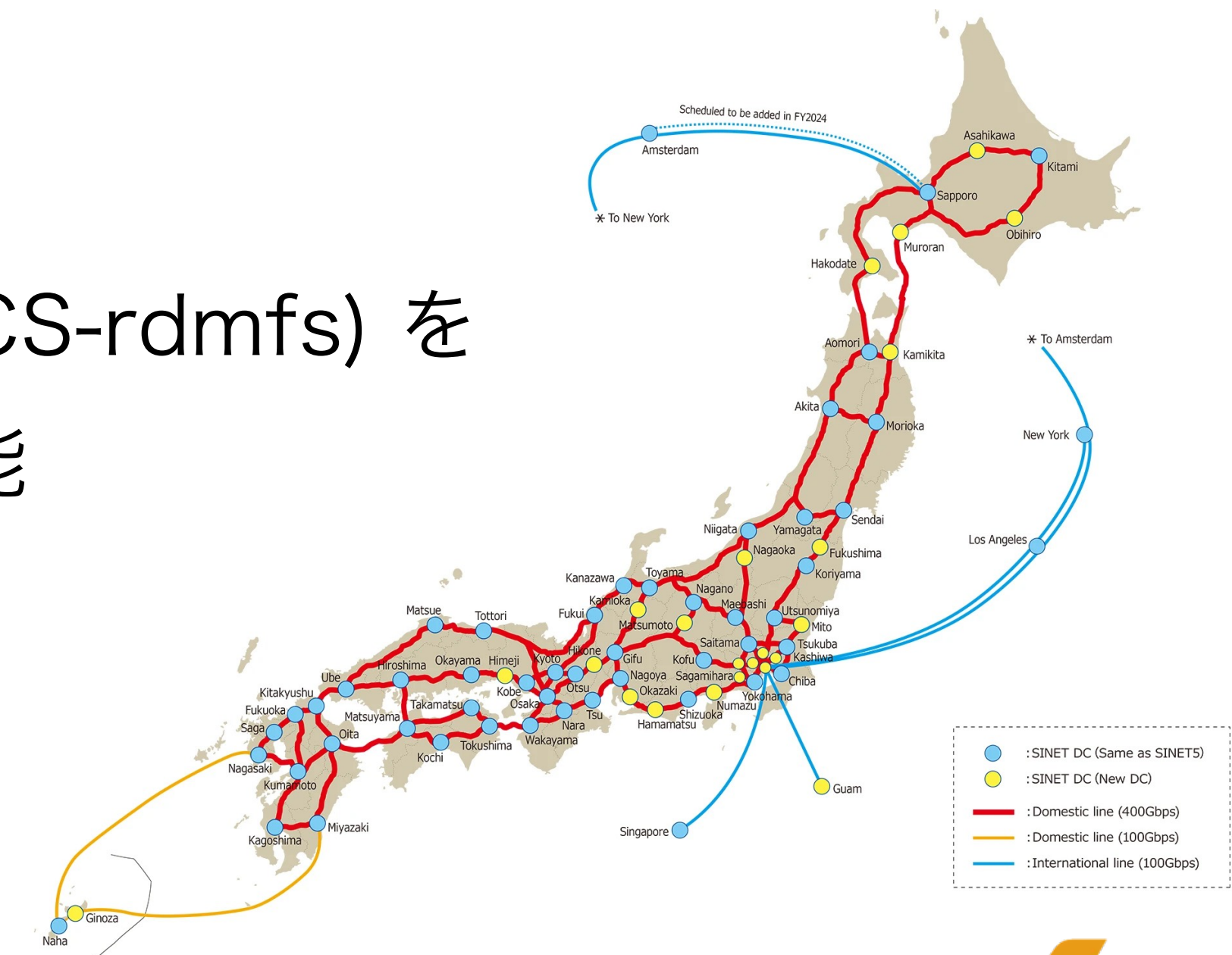
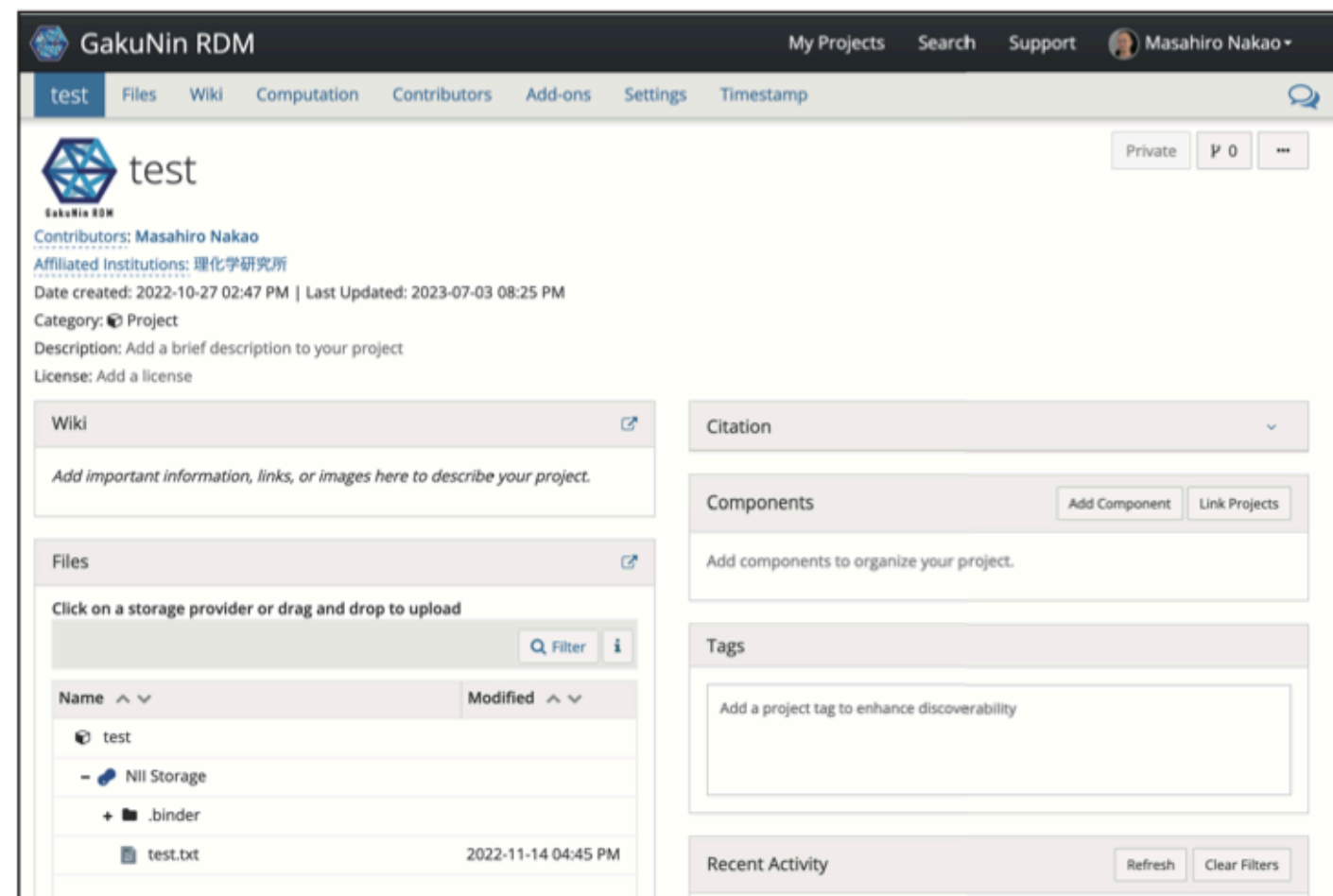
HPCI共用ストレージとGakuNin RDM (2/4)

- HPCI (High Performance Computing Infrastructure) とは、「富岳」と全国の大学や研究機関が保有する様々なHPCクラスタなどを高速ネットワークSINETで結んだシステム
- HPCI共用ストレージとは、HPCIの各組織で研究データを高速に共有するための大規模データ共有基盤であり、理研R-CCS (兵庫県神戸市) と東京大学情報基盤センター (千葉県柏市) が共同で運用
- ファイルシステムにはGfarmが採用されている
- 通常の利用方法
 1. jwt-serverへアクセスして、アクセストークンをjwt-serverに格納する
 2. jwt-agentコマンドを実行し、パスフレーズを入力する
 3. mount.hpcciコマンドを実行すると、HPCI共用ストレージにfusermountが行える



HPCI共用ストレージとGakuNin RDM (3/4)

- GakuNin RDMは研究データ管理サービスであり、NIIが開発と運営を行っている
- Open OnDemandと同様にWebブラウザを介したデータ転送やデータ分析を行うためのアプリケーション（JupyterLabなど）を提供している
- GakuNin RDMもSINETと接続している
- NIIが公開しているスクリプト（<https://github.com/RCOSDP/CS-rdmfs>）を用いると、コマンドラインでGakuNin RDMにfusermountが可能



HPCI共用ストレージとGakuNin RDM (4/4)

- Open OnDemandが提供する手順を用いて開発 (RubyのSinatraを利用)
- <https://osc.github.io/ood-documentation/latest/tutorials/tutorials-passenger-apps.html>

The screenshot shows the Fugaku OnDemand dashboard. At the top, there's a navigation bar with 'Fugaku OnDemand', 'Batch Jobs', 'Interactive Apps', 'Passenger Apps', and 'My Sessions'. Below this is a 'Welcome to the supercomputer Fugaku' message with logos for RIKEN, RCCS, and the Center for Computational Science. A 'Message of the day' section follows, along with an 'Information' section detailing system operations. A 'Pending jobs' section shows a grid of job status tiles for 'fugaku-small', 'fugaku-large', 'prepost-gpu1', 'prepost-gpu2', 'prepost-mem1', 'prepost-mem2', and 'prepost-ondemand'. An 'Accounting' table is also present, showing disk and resource usage. At the bottom, there are 'Recently Used Apps' and 'Passenger Apps' sections, with 'GakuNin RDM' and 'HPCI Storage' highlighted in a red box.

The screenshot shows the 'HPCI Storage' interface. It includes a heading 'HPCI Storage' and a note: 'Please issue a proxy certificate with the HPCI Certificate Issuing System before mounting the HPCI storage.' Below this is a table with columns: 'Action', 'HPCI ID', 'Elapsed time (up to 168 hours)', 'Passphrase', and 'Mount Path'. The 'Action' column contains a 'mount' button. The 'Elapsed time' column has a dropdown menu set to '12'.

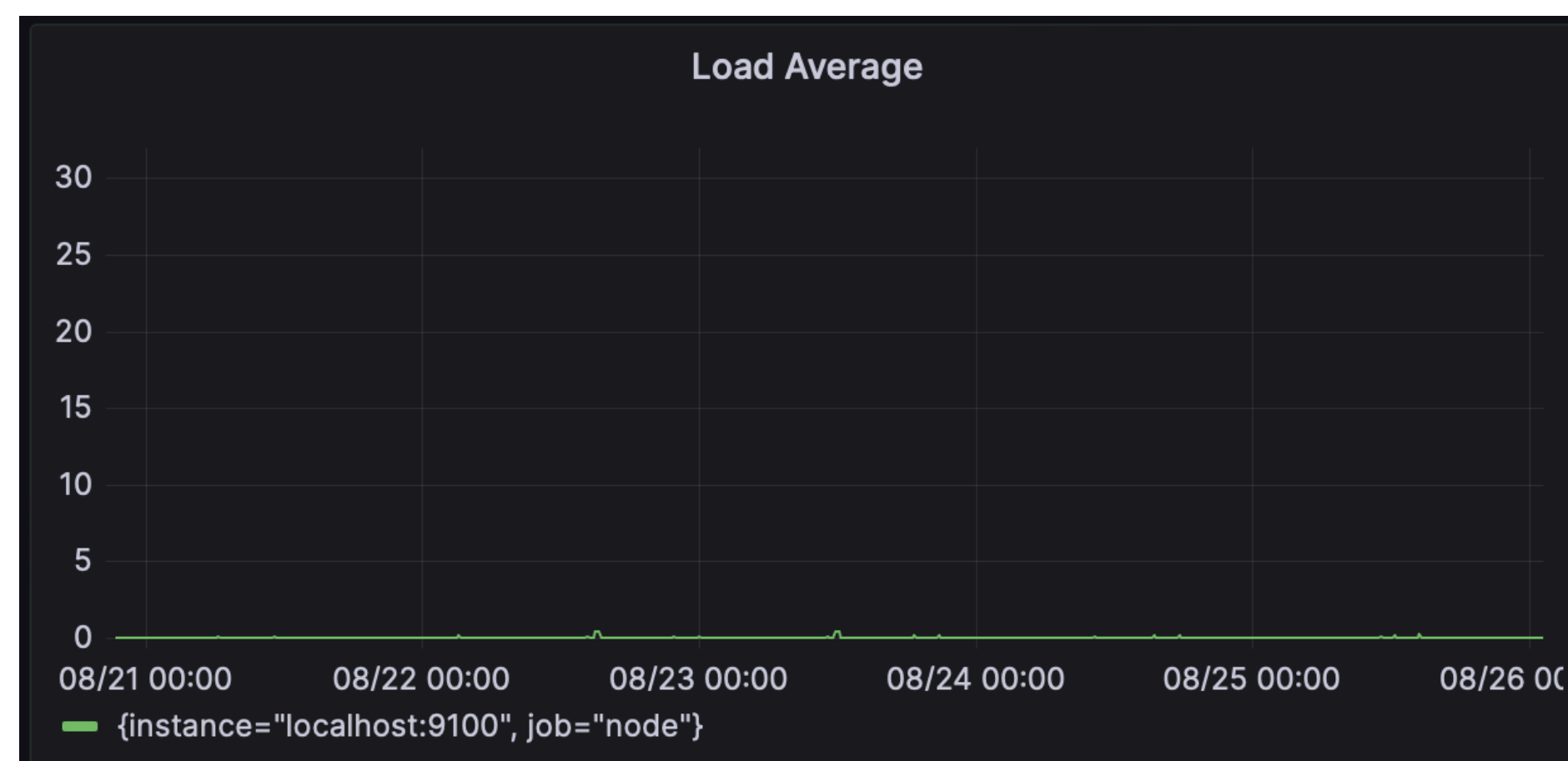
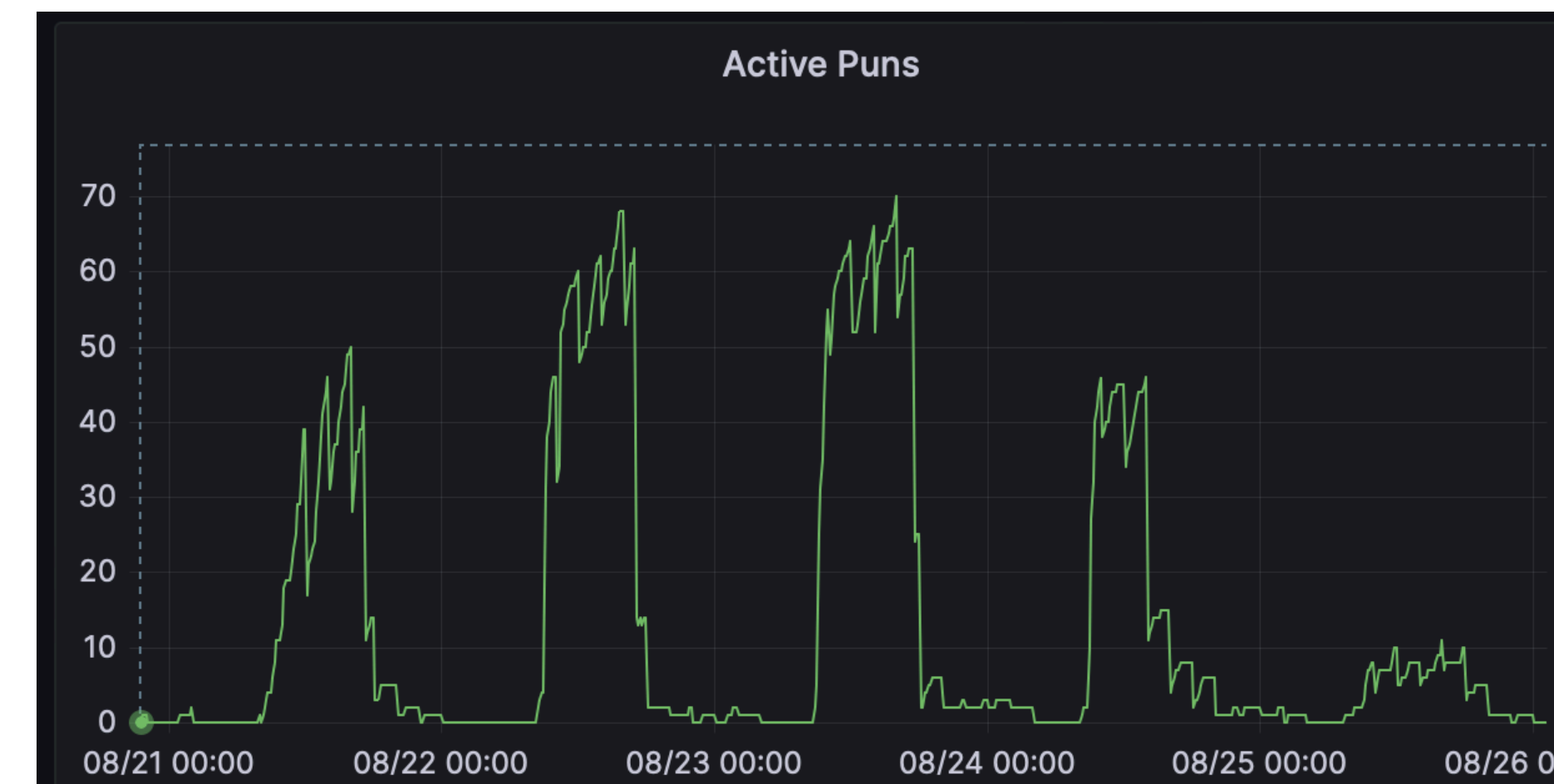
The screenshot shows the 'GakuNin RDM' interface. It features a table with columns: 'Available Actions', 'MOUNT_PATH', 'RDM_NODE_ID', and 'RDM_TOKEN'. The 'Available Actions' column contains a 'mount' button. The other columns are empty input fields.

- 必要事項を入力した後、前ページで説明したコマンドが実行される
- マウント先へのリンクをクリックすると、ファイルのアップロードなどを行う既存のアプリが起動する

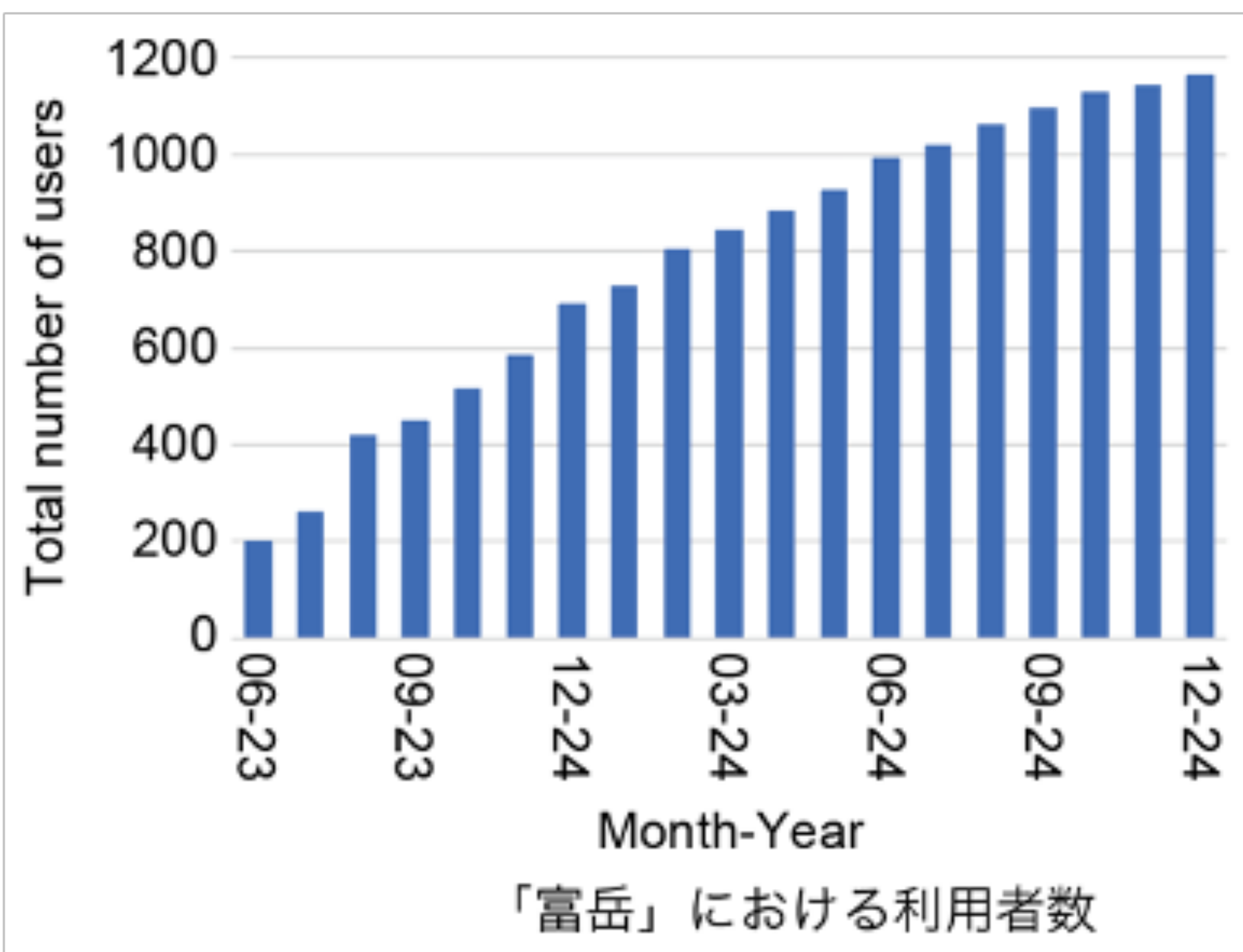
The screenshot shows the Fugaku OnDemand file manager interface. It displays a directory listing for the path '/home/rcs-aot/a03011/'. The listing includes columns for 'Type', 'Name', 'Size', and 'Modified at'. The files listed are 'bin2', 'cuda-samples', and 'Desktop'. The interface also shows navigation buttons like 'Open in Terminal', 'New File', 'New Directory', 'Upload', 'Download', 'Copy/Move', and 'Delete'.

負荷について

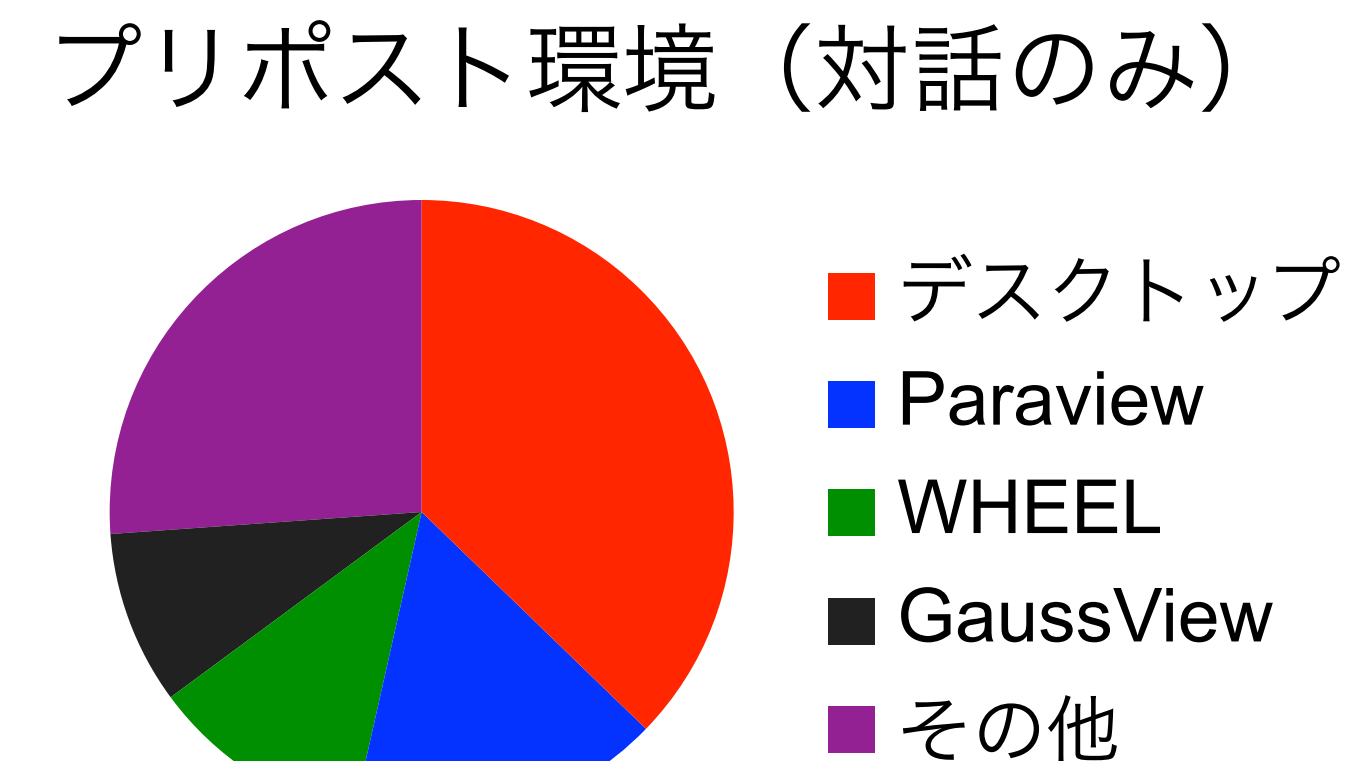
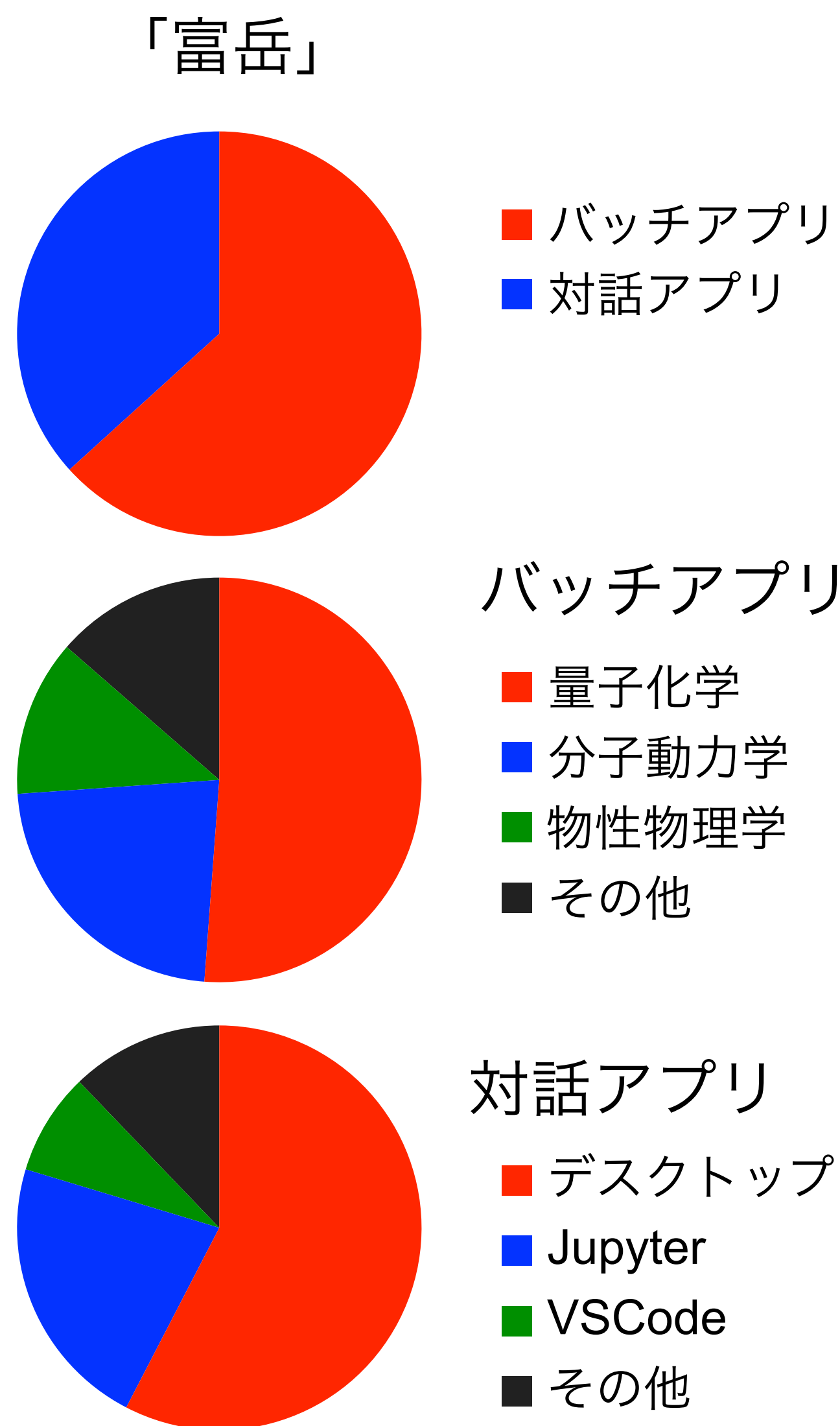
- 2023年8月21日～25日に、高校生を対象としたスーパーコンピューティングコンテストを「富岳」を使って開催
- 昨年まではSSH + ターミナルを使っていたが、2023年はOpen OnDemandを採用
 - PuTTYやWinSCPの利用、鍵の生成や登録の説明がなくなった。説明がOS非依存になった。
- 最大70名の同時利用だったが、問題なかった
- Intel Xeon Gold 6338 x 2、256GB、100GbEのサーバ



運用状況



現在の富岳のアカウント数は4900くらい。
 上のグラフは、累積数なので、削除済のユーザも含まれている



2023年6月～2024年12月の
アプリケーション起動数

まとめと今後の予定

- まとめ

- HPCクラスタが持つ計算資源を簡易に利用可能にするOpen OnDemandを「富岳」で運用
- Fujitsu TCSへの対応、ダッシュボードの変更、アプリケーションの追加、Webフォーム管理コスト低減、HPCI共用ストレージとGakuNin RDMとのデータ共有機能の開発
- 数十人規模でも1台で問題ない

- 今後の予定

- 単一のインターフェースから複数の計算資源を活用できる環境を構築するため、複数のHPCクラスタとクラウドとを統合的に扱えるアグリゲータポータルをOpen OnDemandをベースに開発