

サイバーメディアセンター 大規模計算機システムの利用

大阪大学 情報推進部 情報基盤課

本日のプログラム

- I. システムのご紹介
- II. 利用方法の解説・実習
 - i. システムへの接続
 - ii. プログラムの作成・コンパイル
 - iii. ジョブスクリプトの作成
 - iv. ジョブスクリプトの投入
- III. 利用を希望する方へ

SX-ACE

NEC製のベクトル型スーパーコンピュータ



※今年度中にサービス終了予定

	ノード毎	1クラスタ (512ノード)	総合 (3クラスタ)
CPU数	1	512	1024
コア数	4	2048	4096
演算性能	276 GFLOPS	141 TFLOPS	282 TFLOPS
ベクトル 性能	256 GFLOPS	131 TFLOPS	262 TFLOPS
メモリ	64GB	32TB	64TB

OCTOPUS

NEC製のスカラ型クラスタシステム
構成の異なる4種類のノードで構成されている



	CPUノード	GPUノード	Xeon Phi ノード	大容量主記憶 搭載ノード	合計
CPU数	2	2	1	8	606
コア数	24	24	64	128	9624
演算性能	1.996 TFLOPS	23.196 TFLOPS	2.662 TFLOPS	8.192 TFLOPS	1.463 PFLOPS
メモリ	192GB	192GB	192GB	6TB	72.864TB
ノード数	236ノード	37ノード	44ノード	2ノード	319ノード

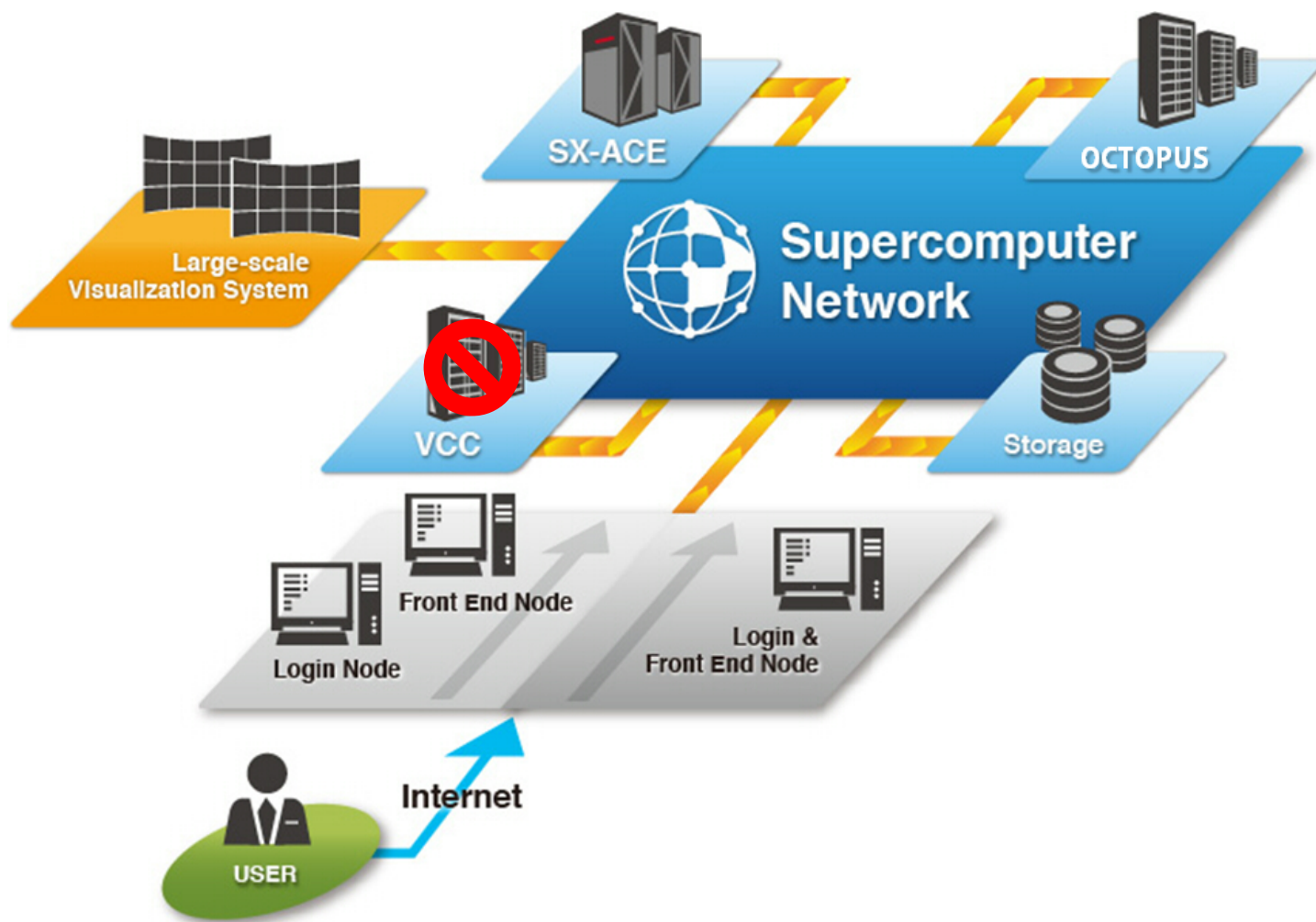
フロントエンド端末

プログラムのコンパイルや計算結果の確認を行うための
の**作業用端末**

フロントエンド端末から各計算機に対して
処理の実行を指示 ※詳細は後述

計算機自体へのログインは**原則禁止**(一部例外有)

システム全体図



本日のプログラム

I. システムのご紹介

II. 利用方法の解説・実習

i. システムへの接続

ii. プログラムの作成・コンパイル

iii. ジョブスクリプトの作成

iv. ジョブスクリプトの投入

III. 利用を希望する方へ

利用方法の解説・実習

本項では初心者を対象に
大規模計算機システムの利用方法を解説します

途中、スパコンを利用した実演も行います

配布したアカウントは講習会後もしばらく利用可能です
ご自宅からでも接続できますのでご自由にお試してください

利用の流れ

ユーザー



フロントエンド端末



システムへの接続

計算機



プログラム作成

コンパイル

ジョブスクリプト
作成

ジョブスクリプト
投入

本日のプログラム

I. システムのご紹介

II. 利用方法の解説・実習

i. システムへの接続

ii. プログラムの作成・コンパイル

iii. ジョブスクリプトの作成

iv. ジョブスクリプトの投入

III. 利用を希望する方へ

システムへの接続

ログインはSSH (Secure Shell) 接続

Win: TeraTermなど, Mac: ターミナルを使用

接続先は SX-ACE : login.hpc.cmc.osaka-u.ac.jp
OCTOPUS : octopus.hpc.cmc.osaka-u.ac.jp

ユーザー



ログインサーバ



フロントエンド端末



SSH

複数台あり、ログインユーザの
少ない端末に自動で振り分ける

本日のプログラム

I. システムのご紹介

II. 利用方法の解説・実習

i. システムへの接続

ii. プログラムの作成・コンパイル

iii. ジョブスクリプトの作成

iv. ジョブスクリプトの投入

III. 利用を希望する方へ

プログラムの作成

計算機を利用するために、まずプログラムを作成する必要があります

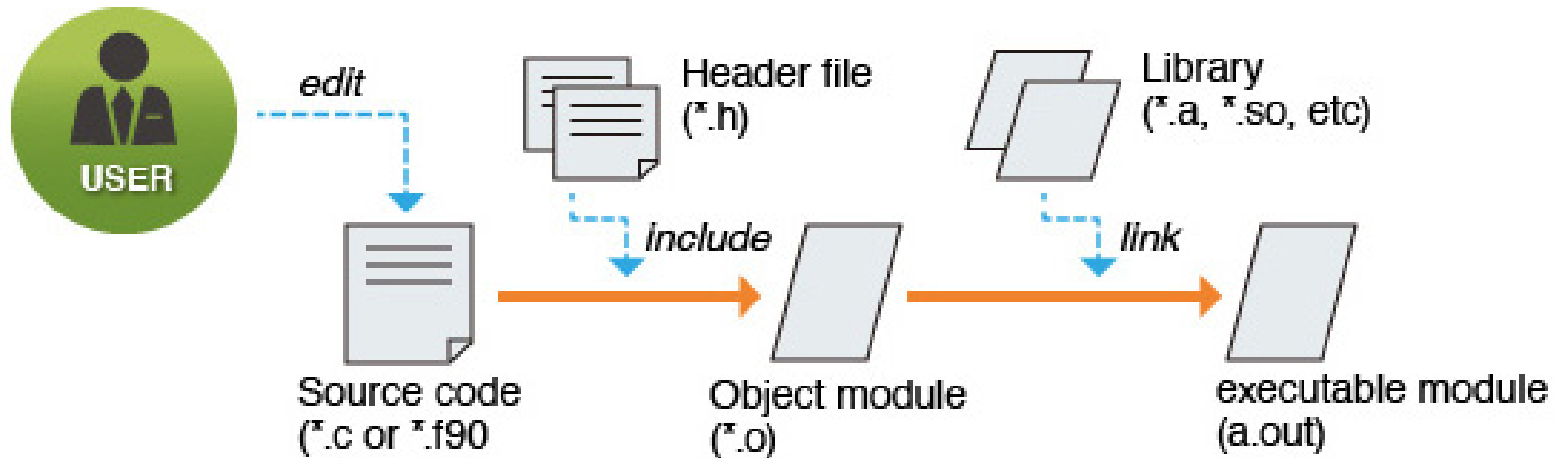
当センターの計算機で使用可能な主な言語

Fortran言語、C言語、C++言語

「プログラムの書き方」については
特に説明しません

コンパイル

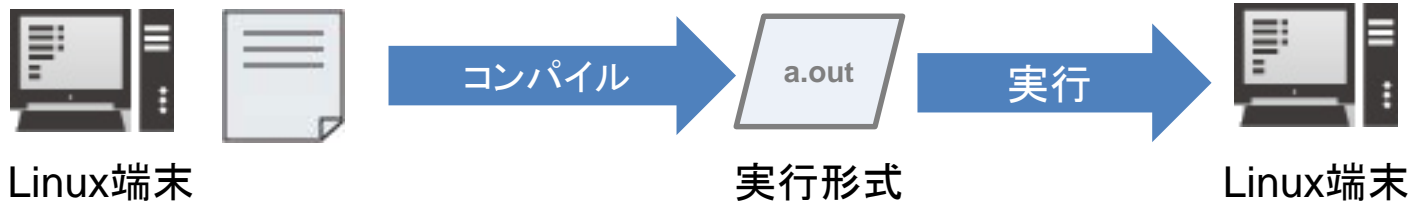
プログラムを「**機械が実行できる形式**」に変換すること



コンパイルの種類

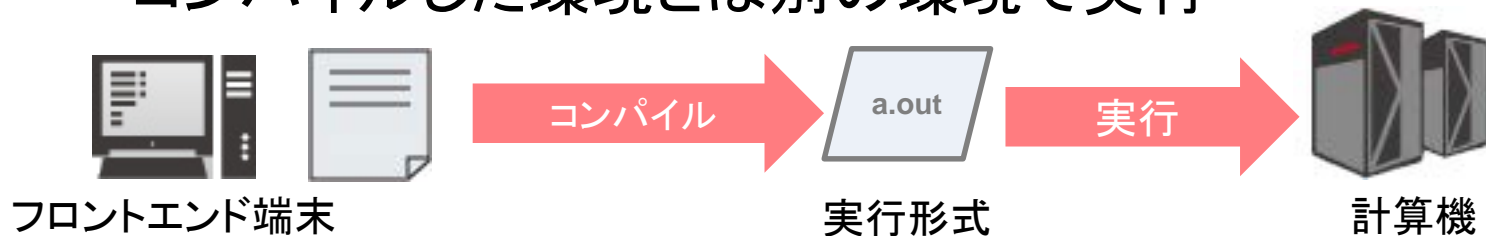
セルフコンパイル

コンパイルした環境と同じ環境で実行



クロスコンパイル

コンパイルした環境とは別の環境で実行



当センターでは「クロスコンパイル」を使用

コンパイルの方法

コンパイルを行う際のコマンド

	Fortran言語	C言語	C++言語
SXクロスコンパイラ (SX-ACE用)	sxf90	SXC++	
Intelコンパイラ (OCTOPUS用)	ifort	icc	icpc

コマンド例(SX-ACE用Fortranプログラム)

```
$ sxf90 program.f
```

→実行形式ファイル「a.out」が生成

コンパイルオプション

コンパイル時にオプションを指定することで
様々な機能を使用することが可能

```
$ sxf90 program.f -option
```

オプションの一例

-o [filename] : 実行形式のファイル名を指定

指定しない場合は「a.out」が出力

-Rn : 翻訳リスト出力 (nには0~5を指定)

最適化等によるプログラムの変形内容を出力

-ftrace : 簡易性能解析機能

ジョブスクリプトに“setenv F_FTRACE YES ”の指定が必要
プログラム実行後に解析ファイルを出力

コンパイルオプション(参考)

オプションの一例

-P [suboption] : 並列化オプション

並列化処理を使用する場合に指定

suboptionには、auto、openmp、multi等を指定可能

-C [suboption] : 最適化オプション

ベクトル化、最適化のレベル指定

suboptionには、hopt、vopt、vsafe、ssafe、debugを指定可

詳しい解説は下記の講習会にて行います

SX-ACE 高速化技法の基礎

2020年6月29日(月) 13:30 – 16:30

並列コンピュータ 高速化技法の基礎

2020年6月30日(火) 13:30 – 15:30

デモ1 (コンパイル)

1. 演習用プログラムの取得 (コピー)

(例) \$ cp /sc/cmc/apl/kousyu/nyumon/sample.f ~/

2. sample.f をSX用にコンパイル

(例) \$ sxf90 sample.f

※文字入力時は [Tab]キーでの補完機能を活用してください

本日のプログラム

I. システムのご紹介

II. 利用方法の解説・実習

i. システムへの接続

ii. プログラムの作成・コンパイル

iii. ジョブスクリプトの作成

iv. ジョブスクリプトの投入

III. 利用を希望する方へ

計算機の利用方法

会話型（インタラクティブ利用）

コマンド等を通してコンピュータに直接命令し、リアルタイムで処理を実行

操作として手軽

一括処理型（バッチ利用）

コンピュータにまとめて処理を命令し実行

処理の命令が終われば、ログアウトしてもOK

会話型

原則として利用不可

旧SXでは会話型が利用可能だった

→現在稼働中の計算機では利用不可

ただし“会話型風”の機能はあり

フロントエンド端末での計算実行も禁止

基本的に「一括処理型」で利用

一括処理型

処理を「ジョブスクリプト」に記述

スクリプトに基づき計算機が処理を実行

SX-ACEでsx.outという
プログラムを実行したい

ジョブスクリプトを作成

送信

計算機が空き次第
実行指示

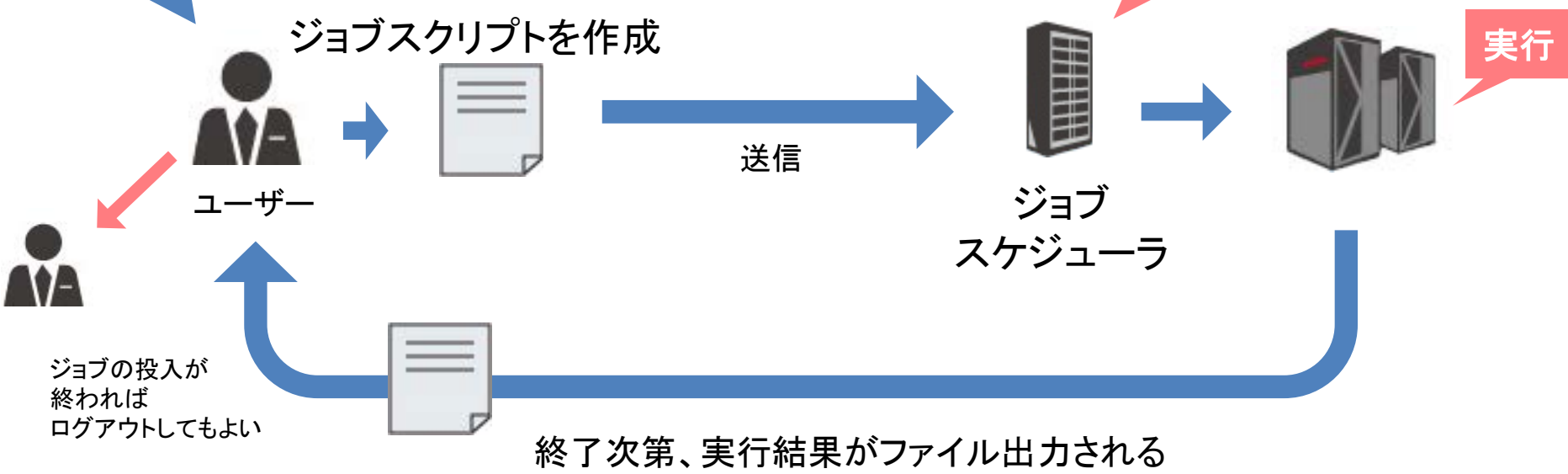
実行

ユーザー

ジョブ
スケジューラ

ジョブの投入が
終われば
ログアウトしてもよい

終了次第、実行結果がファイル出力される



ジョブスクリプト

ジョブスクリプトの構成

リソースや環境設定 : #PBSから始まるNQSオプション
計算機に実行させる処理の記述 : シェルスクリプト

ジョブスクリプトの例

```
#!/bin/csh
```

リソース、環境設定の指定

```
#PBS -q ACE
```

```
#PBS -l elapstim_req=1:00:00
```

```
cd $PBS_O_WORKDIR
```

```
./sx.out > result.txt
```

計算機に実行させる処理の記述

リソース、環境設定の指定

NQSIIオプション(以下)でリソースや環境の設定を行う

オプション	説明
#PBS -q	ジョブクラスを指定し、計算に使用する計算機やリソースを指定する
#PBS -l	使用する資源値
	elapstim_req : ジョブの経過時間
	memsz_job : 1ノードあたりのメモリ量
	cpunum_job : 1ノード当たりのCPU数
#PBS -m	計算の処理状態に変化が起きたときメール通知を行う
	a : ジョブが異常終了したとき
	b : ジョブが開始したとき
	e : ジョブが終了したとき
#PBS -M	メールの通知先アドレスを指定する
#PBS -v	環境変数の指定(setenvではなくこちらを使うことを推奨する)
#PBS -T	MPI 実行時に指定
	mpisx : MPI/SX 利用時
	intmpi : IntelMPI 利用時
#PBS -b	使用するノード数

必須!

ジョブクラス一覧(SX-ACE)

使用する計算機、リソースはジョブクラスで指定
NQSIIオプション「#PBS -q」の後に続けて記述

ジョブクラス	利用可能 経過時間	利用可能 最大Core数	利用可能 メモリ	同時利用可能 ノード数
ACE	120時間	1024Core (4Corex256ノード)	1.5TB (60GBx256ノード)	256ノード
DBG	20分	32Core (4Corex8ノード)	480GB (60GBx8ノード)	8ノード

ジョブクラス一覧(OCTOPUS)

ジョブクラス	利用可能 経過時間	利用可能 CPU数	利用可能 メモリ	同時利用 可能ノード数
OCTOPUS	120時間	3,072Core (24Corex128ノード)	24,320GB (190GBx128ノード)	128ノード
OCTPHI	120時間	2,048Core (64Corex32ノード)	6,080GB (190GBx32ノード)	32ノード
OCTMEM	120時間	256Core (128Corex2ノード)	12TB (6TBx2ノード)	2ノード
DBG	10分	24Core (24Corex1ノード)	190GB (190GBx1ノード)	1ノード

計算機に実行させる処理の記述

ファイルやディレクトリの実行・操作を記述
記述方法はシェルスクリプト

よく使用するNQSII 用の環境変数

\$PBS_O_WORKDIR : ジョブ投入時のディレクトリが設定される

処理の記述の最終行に改行を入れること！

⇒ 未入力の場合、その行のコマンドが実行されない

ジョブスクリプト解説

ジョブクラスの指定

```
#!/bin/csh
```

```
#PBS -q ACE
```

経過時間、メモリサイズの指定
コマ後にスペースを入れられないよう注意！

```
#PBS -l elapstim_req=1:00:00,memsz_job=60GB
```

```
cd $PBS_O_WORKDIR
```

ジョブ投入時のディレクトリへ移動

```
./a.out > result.txt
```

a.outを実行し、結果をresult.txtに出力する
(リダイレクション)

デモ2(ジョブスクリプト)

1. 演習用スクリプトを取得してください

(例)\$ cp /sc/cmc/apl/kousyu/nyumon/sample.nqs ~/

2. sample.nqsを元にSX-ACE用のジョブスクリプトを作成してください

(例) \$ emacs sample.nqs -nw

ジョブクラスはDBGを使用してください

本日のプログラム

I. システムのご紹介

II. 利用方法の解説・実習

i. システムへの接続

ii. プログラムの作成・コンパイル

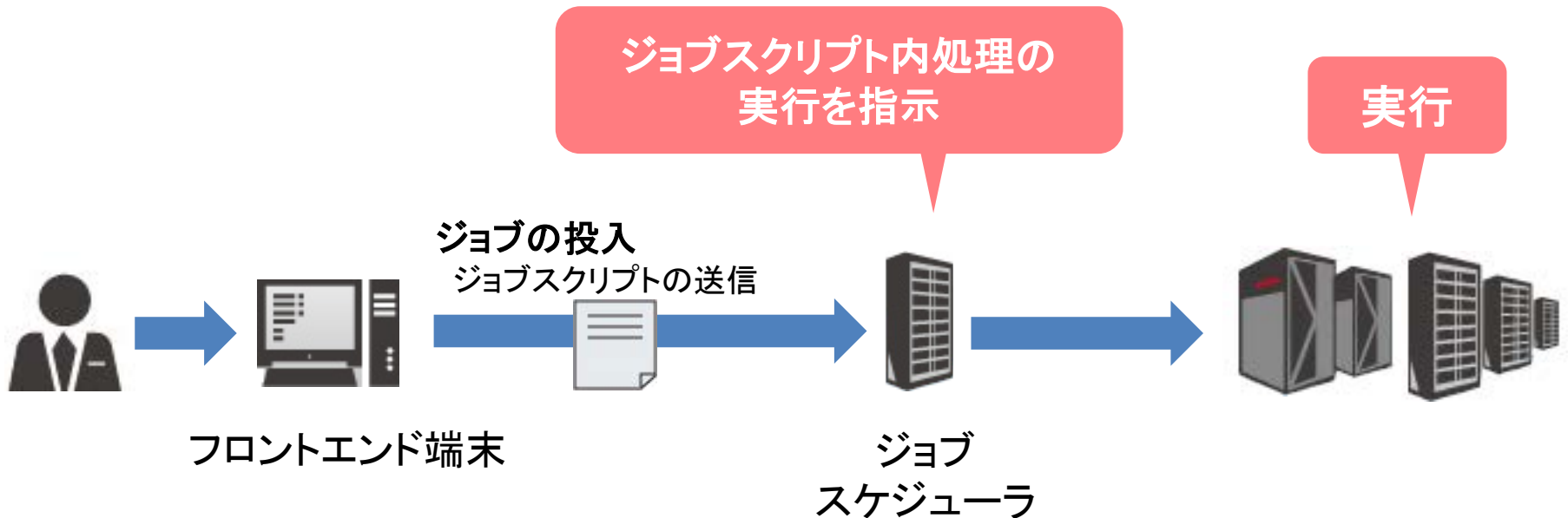
iii. ジョブスクリプトの作成

iv. ジョブスクリプトの投入

III. 利用を希望する方へ

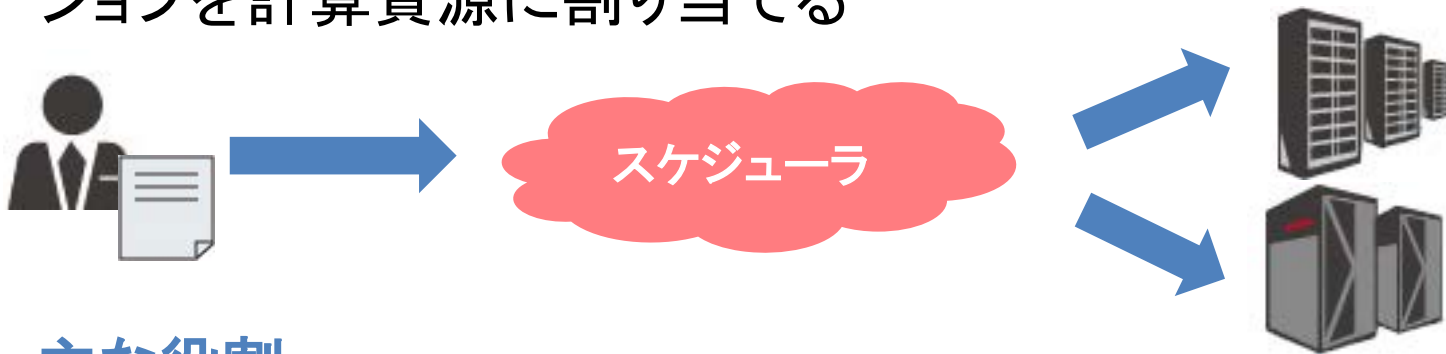
実行までの流れ

ジョブスクリプトは**ジョブスケジューラ**が受け付ける
ジョブスケジューラが各計算機にジョブの実行を指示



スケジューラとは

あらかじめ管理者によって設定された資源割当ポリシーに従い、ジョブを計算資源に割り当てる



主な役割

クラスタを構成する計算機(ノード)の静的情報※を把握

※ディスク容量、メモリ容量、CPU性能、etc

ノード毎の資源使用率を定期的に監視、管理

ユーザより実行したいジョブ要求を受信

ジョブを実行するのに適切なノードを選定

ジョブ実行に伴う入出力データのファイル転送

スケジューラとは

当センターではバックフィル型を採用

特徴

ジョブの実行開始時間のマップを作成する

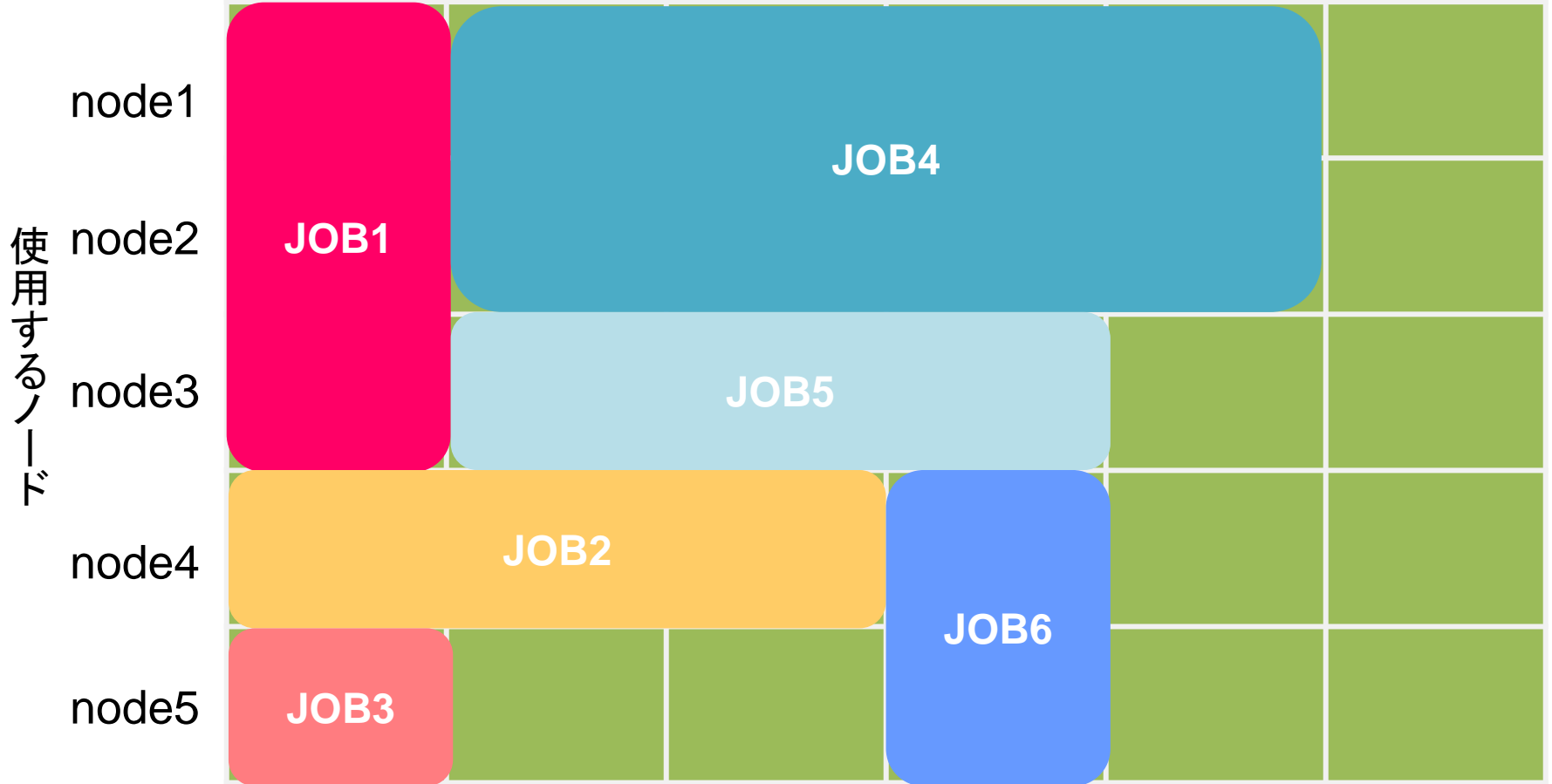
マップに載れば、実行開始時間と経過時間が保障される

実行中は指定したリソースを占有して割当てる

スケジューラのイメージ

ジョブが実行を開始する時間

0:00 1:00 2:00 3:00 4:00 5:00 6:00



ジョブの投入方法

フロントエンド端末からジョブを投入
コマンド

```
$ qsub [ジョブスクリプトファイル]
```

(参考) 特殊な投入方法

リクエスト連携: 順番通りにジョブを実行したい場合に利用

```
$ qsub [JobScript1] [JobScript2] ...
```

※ 順番無く複数のジョブを同時に投入する場合は
上記のようにしないよう注意

投入済みジョブの確認方法

ジョブの状態を確認することが可能

コマンド

\$ **qstat**

実行結果

RequestID	ReqName	UserName	Queue	STT	Memory	CPU	Elapse
12345.cmc	nqs-test	a61234	ACE	RUN	8.72G	830.66	208

ジョブの状態

待ち状態では「QUE」 実行が始まると「RUN」となる。

実行時間

CPU : 実際にジョブが消費した時間
複数CPU指定の場合は、全CPUを累積表示

Elapse : ジョブが実行されてからの経過時間

投入済みジョブの確認方法

ジョブの予約状況の確認することが可能
コマンド

\$ **sstat**

実行結果

RequestID	ReqName	UserName	Queue	Pri	STT	PlannedStartTime
12345.cmc	nqs-test	a61234	ACE	-1.5684/ -1.5684	ASG	2015-06-16 00:01:23

状態監視

実行時刻が決まると「ASG」表示になる。

混雑具合や優先度により、「実行時間の決定」までの待ち時間が異なるが、一旦実行時間が決定されるとその時刻にジョブ実行が始まる。

実行開始時刻

システムメンテナンスやトラブル時は再スケジュールされることをご了承ください。

投入済みジョブの操作方法

ジョブのキャンセル

コマンド

```
$ qdel [RequestID]
```

実行結果

```
$ qdel 12345.cmc
```

```
Request 12345.cmc was deleted.
```

実行結果の確認方法

実行結果や実行エラーは指定しない限り「標準出力」となる

標準出力はジョブスクリプト名.oリクエストID
標準エラー出力はジョブスクリプト名.eリクエストID
というファイル名で自動出力される

catやlessコマンドでファイルの内容を出力し確認

```
$ cat nqs.o12345
```

※リダイレクション(./a.out > result.txt)を使った場合は、そちらも確認

標準出力／標準エラー出力の容量制限

⇒ 100MB以上出力したい場合はリダイレクション(>)

デモ3 (ジョブスクリプトの投入)

1. 作成したジョブスクリプトを使用してジョブを投入

```
$ qsub sample.nqs
```

2. 投入したジョブの状態を確認

```
$ sstat
```

```
$ qstat
```

3. 結果ファイルの確認

```
$ cat sample.nqs.o12345
```

```
$ cat sample.nqs.e12345
```

※リダイレクション(./a.out > result.txt)を使った場合は、そちらも確認

より高度な利用に向けて

利用の参考になるWebページ

サイバーメディアセンター 大規模計算機システム Webページ
<http://www.hpc.cmc.osaka-u.ac.jp/system/manual/>

利用方法

<http://www.hpc.cmc.osaka-u.ac.jp/system/manual/>

FAQ

<http://www.hpc.cmc.osaka-u.ac.jp/faq/>

お問い合わせ

http://www.hpc.cmc.osaka-u.ac.jp/support/contact/auto_form/

研究成果

<http://www.hpc.cmc.osaka-u.ac.jp/researchlist/>

より高度な利用に向けて

本日以降に実施予定の講習会

講習会名	日時	場所
SX-ACE 高速化技法の基礎	6月29日(月) 13:30 - 16:30	オンライン配信
並列コンピュータ高速化技法の基礎 (OCTOPUS向け)	6月30日(火) 13:30 - 15:30	オンライン配信
並列プログラミング入門(MPI)	7月3日(金) 13:30 - 17:00	オンライン配信

本日のプログラム

- I. システムのご紹介
- II. 利用方法の解説・実習
 - i. システムへの接続
 - ii. プログラムの作成・コンパイル
 - iii. ジョブスクリプトの作成
 - iv. ジョブスクリプトの投入
- III. 利用を希望する方へ**

利用を希望する方へ

本センターの大規模計算機システムは
どなたでも**利用可能**です！

大学院生

教員

研究者

大阪大学

他大学

民間企業

利用負担金が必要になります

利用負担金



SX-ACE

共有利用



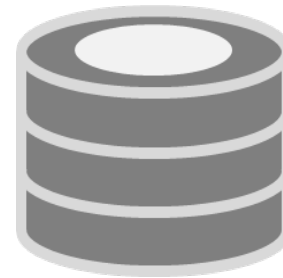
SX-ACE

占有利用



OCTOPUS

共有利用



ディスク
容量追加
オプション
(1TB単位)

SX-ACE利用者の方は500GB
OCTOPUSの利用者の方は1TB
無償で利用可能です

計算機の提供方法

共有利用

「ノード時間」,単位で
ノードを利用

利用者全員で一定数のノードを共有

大規模なノード間並列を試せる
「待ち時間」が発生する

占有利用

「年度/月」単位で
ノードを利用

他の利用者のグループとノードを共有しない

大規模なノード間並列は試し
難い
「待ち時間」が発生しない

「ノード時間」とは

$$\text{ノード時間} = \text{計算に使用するノード数} \times \text{計算時間(単位:時間)}$$

(例)

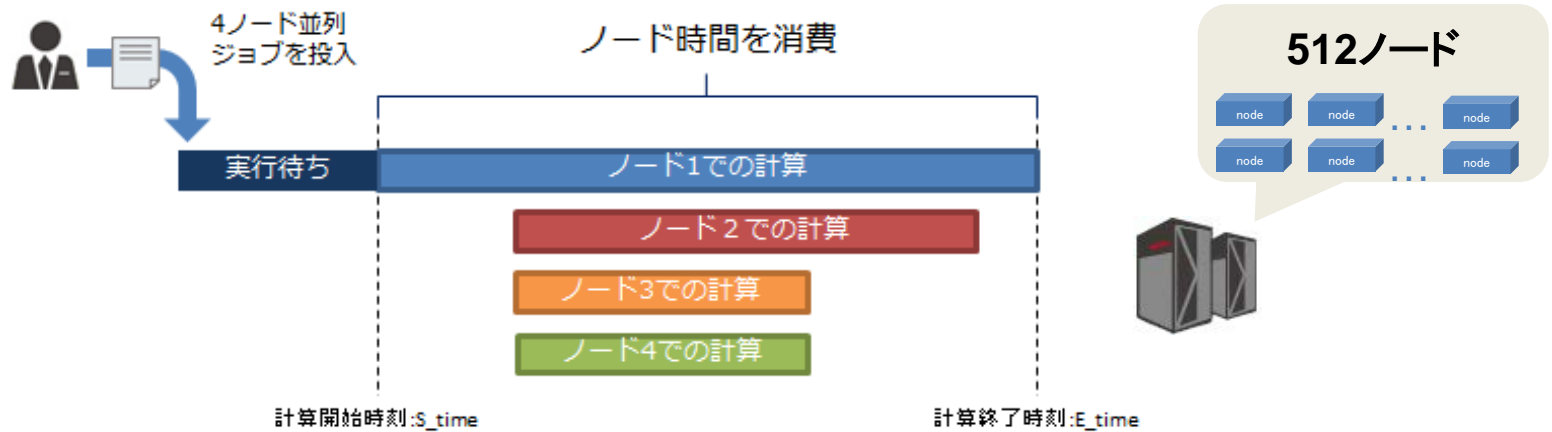
1ノードで3時間の計算	→	3ノード時間消費
30ノードで5時間の計算	→	150ノード時間消費
100ノードで1時間の計算	→	100ノード時間消費
1ノードで100時間の計算	→	100ノード時間消費

「ノード時間」の注意点



ノード内で使用するコアを限定しても、ノード時間は変わりません

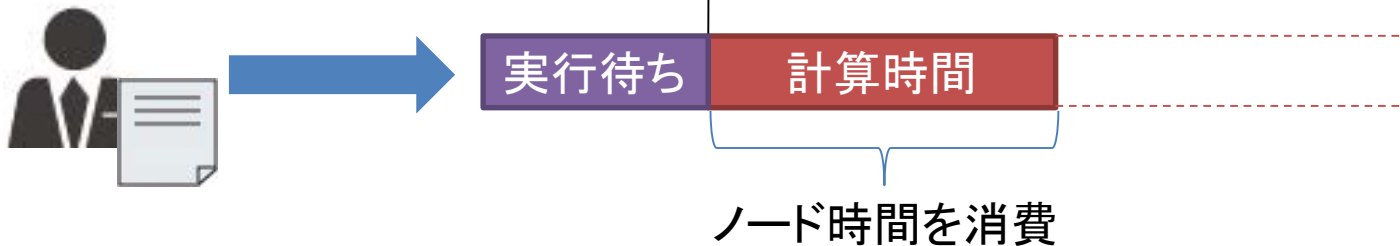
「ノード時間」の注意点



ノード時間は4ノード × (計算終了時間 - 計算開始時間)です

「ノード時間」の注意点

```
#!/bin/csh
#PBS -q DBG
#PBS -l memsz_job=1GB,elapstim_req=1:00:00
```



消費するノード時間は、実際かかった計算時間のみです

スケジューラのイメージ

ジョブが実行を開始する時間

0:00 1:00 2:00 3:00 4:00 5:00 6:00

使用するノード



「OCTOPUSポイント」とは

「OCTOPUS」への申請で全てのノードを自由に使用可能とすることを目的に導入された制度です。



申請



OCTOPUS



汎用CPU ノード

GPU ノード

Xeon Phi ノード

大容量主記憶搭載ノード

「OCTOPUSポイント」とは

OCTOPUSポイントの消費量は
以下の計算式から算出されます

使用したノード時間 × 消費係数 × 季節係数

「消費係数」について

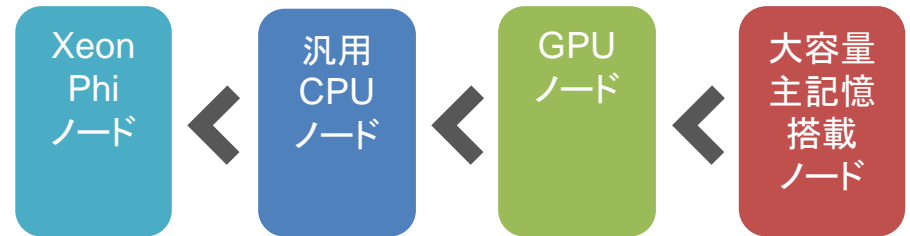
使用したノード時間 × 消費係数 × 季節係数

消費係数

各ノード群の消費電力を元に設定

ノード群	消費係数
CPUノード	0.0520
GPUノード	0.2173
Xeon Phiノード	0.0418
大容量主記憶搭載ノード	0.3703

同じノード時間を使用しても、
OCTOPUSポイントの消費量は異なる



「季節係数」について

使用したノード時間 × 消費係数 × 季節係数

季節係数

前年度の利用率を元に
0を超える1以下の値を設定

ノード群	4~ 6月	7~ 9月	10~ 12月	1~ 3月
CPUノード			1	
GPUノード			1	
Xeon Phiノード	0.7	0.7	1	1
大容量主記憶 搭載ノード	0.6	0.6	1	1

(例)

2018年度4月～6月の
Xeon Phiノード群の利用率が30%程度



2019年度4月～6月の
Xeon Phiノード群の季節係数を0.5に設定

「OCTOPUS」ポイントとは

$$\text{消費OCTOPUSポイント} = \text{使用ノード時間} \times \text{消費係数} \times \text{季節係数}$$

(例)

- ・CPUノードを10ノード並列実行で3時間使用(季節係数:1)

$$10 \times 3 \times 0.0520 \times 1 = 1.560 \rightarrow \mathbf{1.56\text{ポイント消費}}$$

- ・GPUノードを10ノード並列実行で3時間使用(季節係数:1)

$$10 \times 3 \times 0.2137 \times 1 = 6.519 \rightarrow \mathbf{6.519\text{ポイント消費}}$$

「OCTOPUS」ポイントとは

$$\text{消費OCTOPUSポイント} = \text{使用ノード時間} \times \text{消費係数} \times \text{季節係数}$$

(例)

- ・Xeon Phiノードを10ノード並列実行で3時間使用(季節係数:1)
 $10 \times 3 \times 0.0418 \times 1 = 1.254 \rightarrow \mathbf{1.254}$ ポイント消費
- ・Xeon Phiノードを10ノード並列実行で3時間使用(季節係数:0.5)
 $10 \times 3 \times 0.0418 \times 0.5 = 6.519 \rightarrow \mathbf{0.627}$ ポイント消費

まずは試用制度をお試ください

3カ月間 以下の資源をご提供

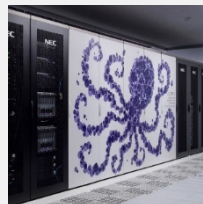
無料



SX-ACE

共有利用

500 ノード時間
+ ディスク 500GB



OCTOPUS

共有利用

26 OCTOPUSポイント
+ ディスク 1TB

500 ノード時間

119 ノード時間

622 ノード時間

70 ノード時間

汎用
CPU
ノード

Xeon
Phi
ノード

GPU
ノード

大容量
主記憶
搭載
ノード

※季節係数が1の場合

利用申請方法

大規模計算機システムの利用申請は
随時受け付け中です！

申請は年度単位(4月から翌年3月まで)です
申請はWEBフォームから受け付けています

詳細は下記のページをご覧ください！

一般利用(学術利用) <http://osku.jp/u094>

試用制度による利用 <http://osku.jp/e029>

大規模計算機システムに関するご質問は

大阪大学 情報推進部 情報基盤課
研究系システム班

system@cmc.osaka-u.ac.jp

までお気軽にご連絡ください！