

先進ゲノム支援 情報解析講習会

遺伝研スパコンの利用方法・UNIX基本コマンド

谷沢靖洋（遺伝研・大量遺伝情報研究室）

講習の内容

- 遺伝研スパコンの概要
- 必要なソフトのインストール (事前準備の確認)
- スパコンへの接続 (事前準備の確認)
- 基本コマンド

遺伝研スパコン概要

国際塩基配列DBのための計算基盤・生命科学のための共同計算基盤
無償で利用可能（個人ゲノム解析区画や優先利用などは一部課金）

2025.3 リプレイス

合計約14,000 CPUコア

CPU最適化ノード（78台、総コア数13,186、総メモリ91TB）

メモリ最適化ノード（5台、総コア数960、総メモリ21TB）

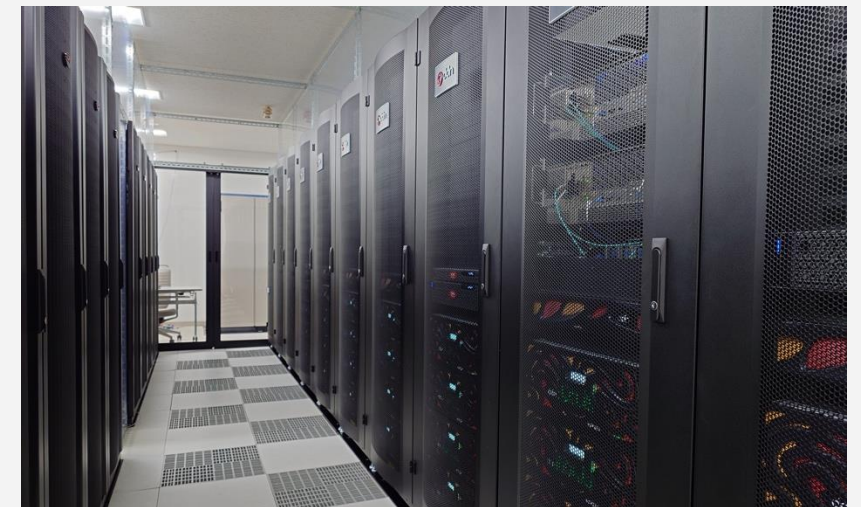
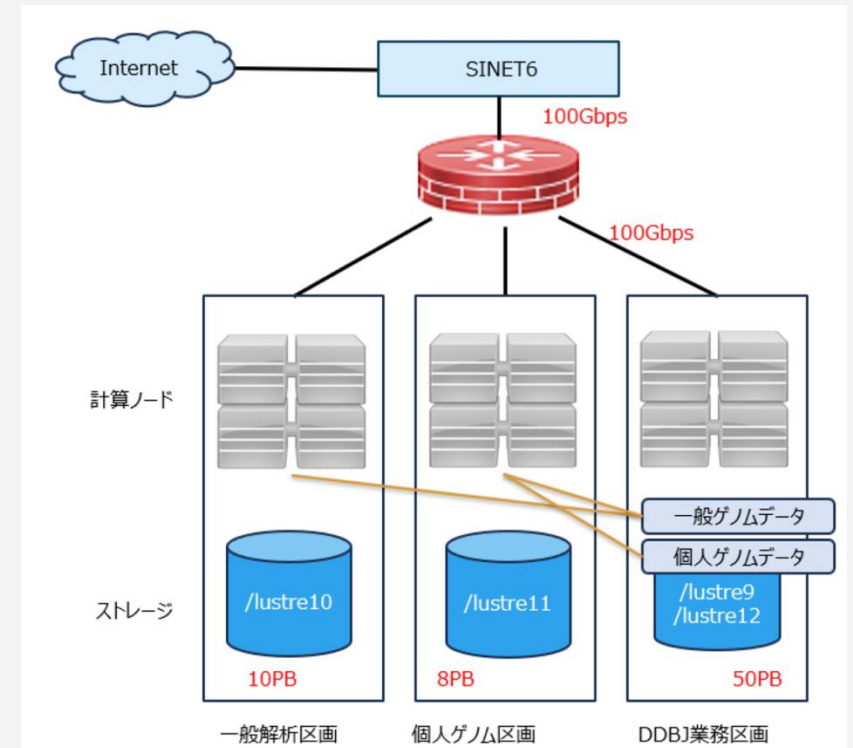
アクセラレータ最適化ノード（B200、L40S、PEZY-SC3）

ストレージ

一般計算用 10 PB / 個人ゲノム区画用 8 PB / DB用 50 PB

各ユーザーの初期ディスク割り当て 1 TB

大規模利用申請により拡張可能



<https://sc.ddbj.nig.ac.jp/guides/overview>

生命科学研究のためのソフトウェア・参照データ

ジョブスケジューラ slurm を利用してジョブの実行を行う

Apptainer (Singularity) コンテナを使い、インストール不要で様々な解析ツールが利用できる

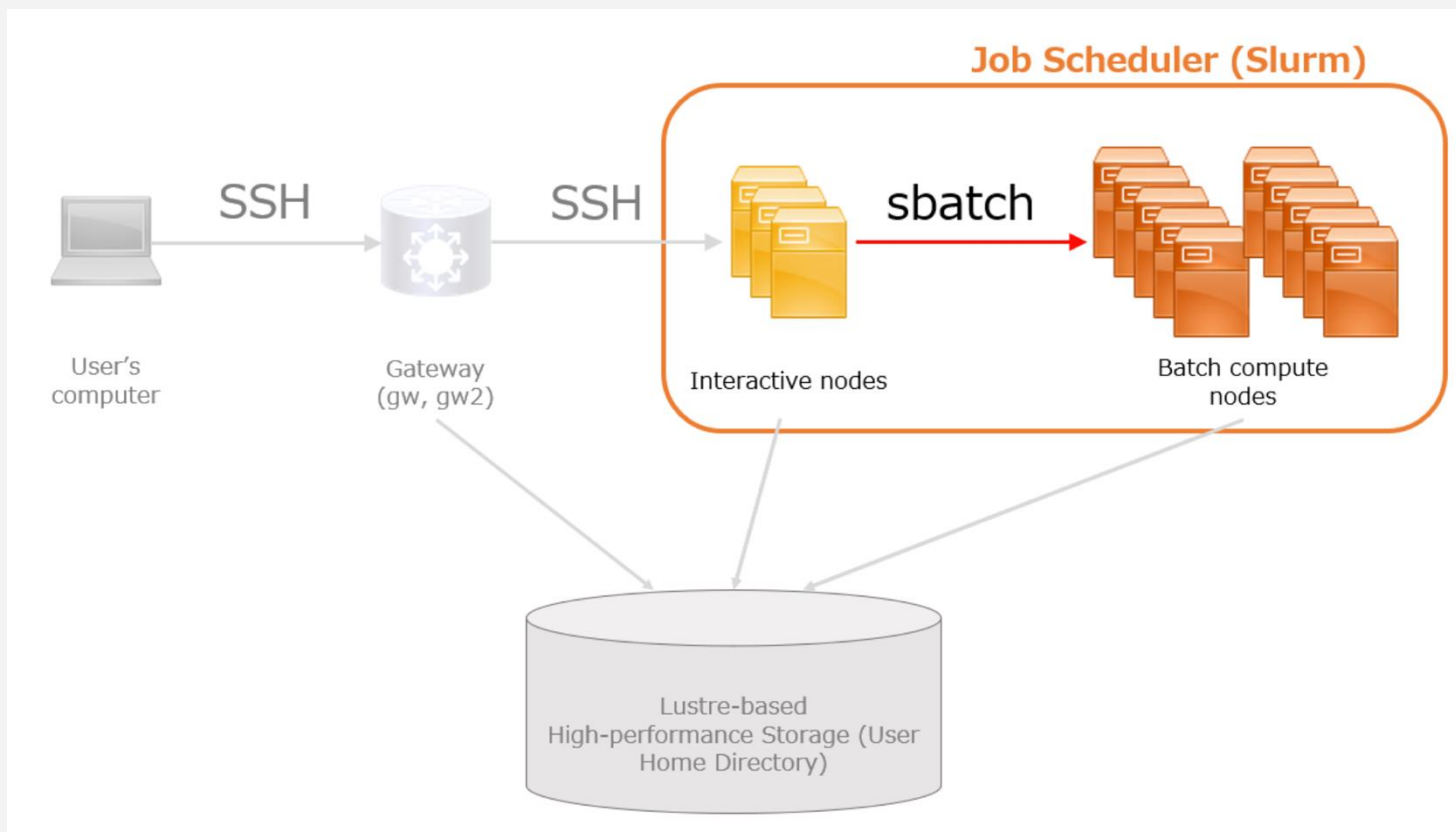
/usr/local/biotools に 2000 以上の解析ツールのコンテナが提供されている

Bioconda で提供されているツールをコンテナ化したもの

DDBJやNCBIで公開されている塩基配列のBLASTデータベースなど /usr/local/seq

DDBJ に登録された塩基配列データ（ゲノム、遺伝子、NGS生データ）
BioProject/BioSample の 生データ /usr/local/resources

基本的な使い方



gwノードにログインした後、インタラクティブノードに移動して作業

負荷の高い計算や長時間かかる計算は“sbatch”コマンドを用いてジョブを投入

必要なメモリ量やCPUを指定して計算リソースを確保する

ユーザーのデータ（ホームディレクトリ）はどのノードからも参照できる

講習の内容

- 遺伝研スパコンの概要
- 必要なソフトのインストール (事前準備の確認)
- スパコンへの接続 (事前準備の確認)
- 基本コマンド

本日の講習会で使うソフトウェア

- ・ ターミナルソフト

スパコンへの接続

- ・ SFTPクライアントソフト

ファイルの転送

- ・ テキストエディタ

ファイルの閲覧・編集、プログラミング

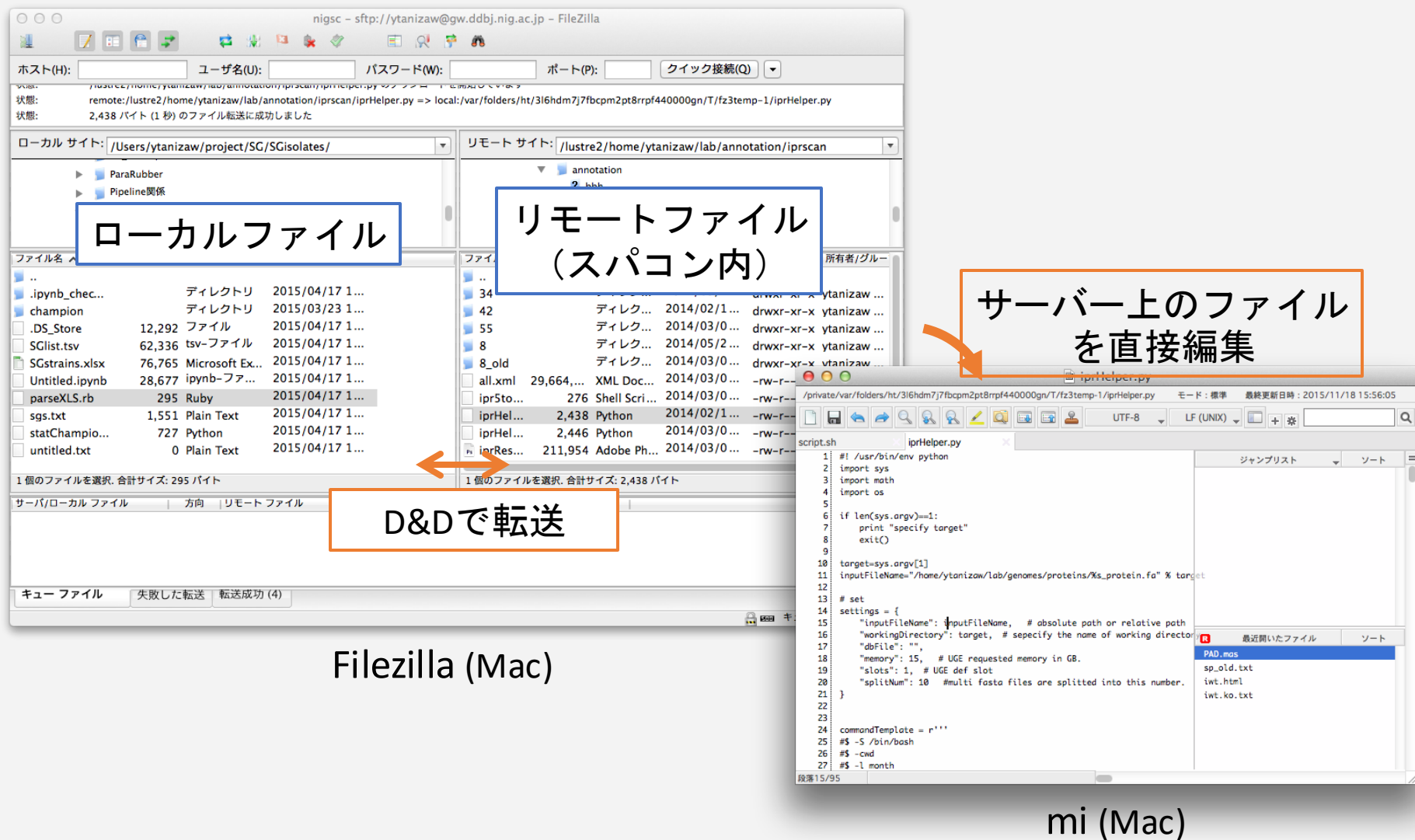
ターミナルソフト

- スパコンに接続するための必須ソフト
- Macでは「ターミナル(OS標準)」や「iTerm2」
- Windowsでは「PowerShell (OS標準)」 「TeraTerm」 など
- スパコンへの接続は SSH プロトコルを利用した安全な通信
- スパコンへの接続以外にも、コマンド操作による大量のファイルの一括処理や、自動処理に用いることが可能。

SFTPクライアントソフト

- スパコンとの間でファイルの転送を行う
- Macでは「Filezilla」「CyberDuck」など
- Windowsでは「WinSCP」「Filezilla」など
- ソフトによってはリモートファイルを直接閲覧・編集することも可能。

SFTPクライアントソフト



Filezilla (Mac)

mi (Mac)

テキストエディタ

- テキストファイルの閲覧・編集
- Macでは「テキストエディット」「mi」など
- Windowsでは「TeraPad」「EmEditor」など
- 高機能のものは、キーワードの自動補完や色分けなどプログラミングの支援機能も持つ

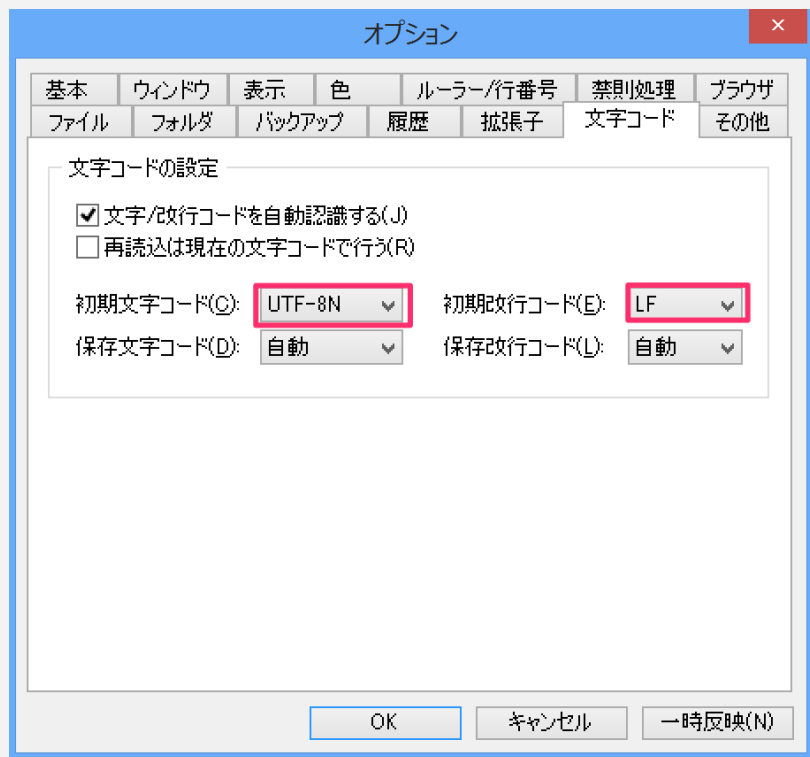
A screenshot of a code editor window with a dark theme. The window has several tabs at the top: 'tasks.py', 'memo', '#\$ -S /bin/bash', and 'fabfile.py'. The 'tasks.py' tab is active, showing Python code. The code includes imports for Celery and time, a Celery app configuration with a time limit, and three task functions: 'add', 'mul', and 'waitsum'. The 'waitsum' function includes a try-except block for handling 'SoftTimeLimitExceeded' exceptions. The code is color-coded: keywords in blue, strings in red, and comments in green.

```
1 from celeryapp import app
2 from celery.exceptions import SoftTimeLimitExceeded
3
4 import time
5 app.control.time_limit('waitsum',
6                         soft=30, hard=30, reply=True)
7
8 @app.task
9 def add(x, y):
10     print x, y, x + y
11     return x + y + y
12
13 @app.task
14 def mul(x, y):
15     return x * y
16
17 @app.task
18 def waitsum(x, y, z):
19     try:
20         time.sleep(z)
21         print x, y, z
22         return x + y
23     except SoftTimeLimitExceeded:
24         print "timeout during execution"
25
26 @app.task
27 def xsum(numbers):
28     time.sleep(10)
29     print "OK"
30     return sum(numbers)
```

SublimeText (Mac or Win)

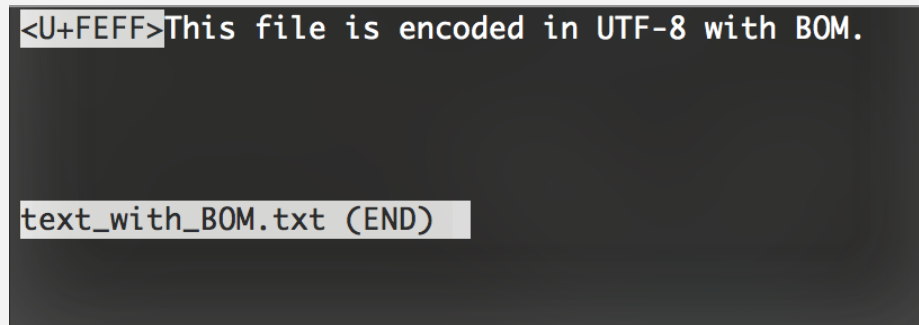
Windows での注意点

- 文字コードは UTF-8 (BOMなし、UTF-8Nであること)
- 改行コードは LF



TeraPadの設定画面

BOM付きUTF-8で作成したファイル



先頭に見えているのが BOM

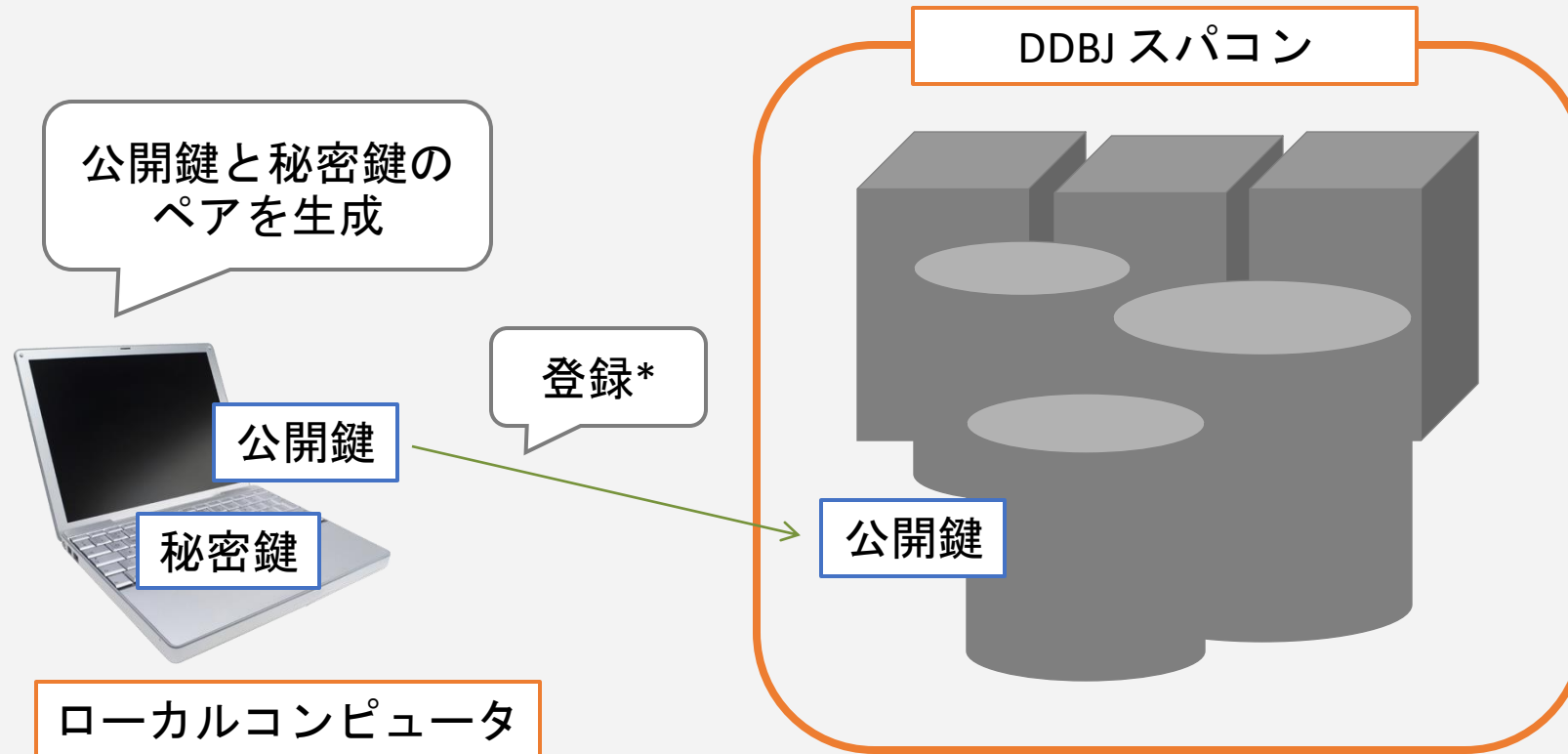
通常は単にUTF8といえはBOMなしのUTFを指す

講習の内容

- 遺伝研スパコンの概要
- 必要なソフトのインストール (事前準備の確認)
- **スパコンへの接続 (事前準備の確認)**
- 基本コマンド

鍵認証

- スパコンへはパスワード認証ではなく、鍵交換による認証で接続する。

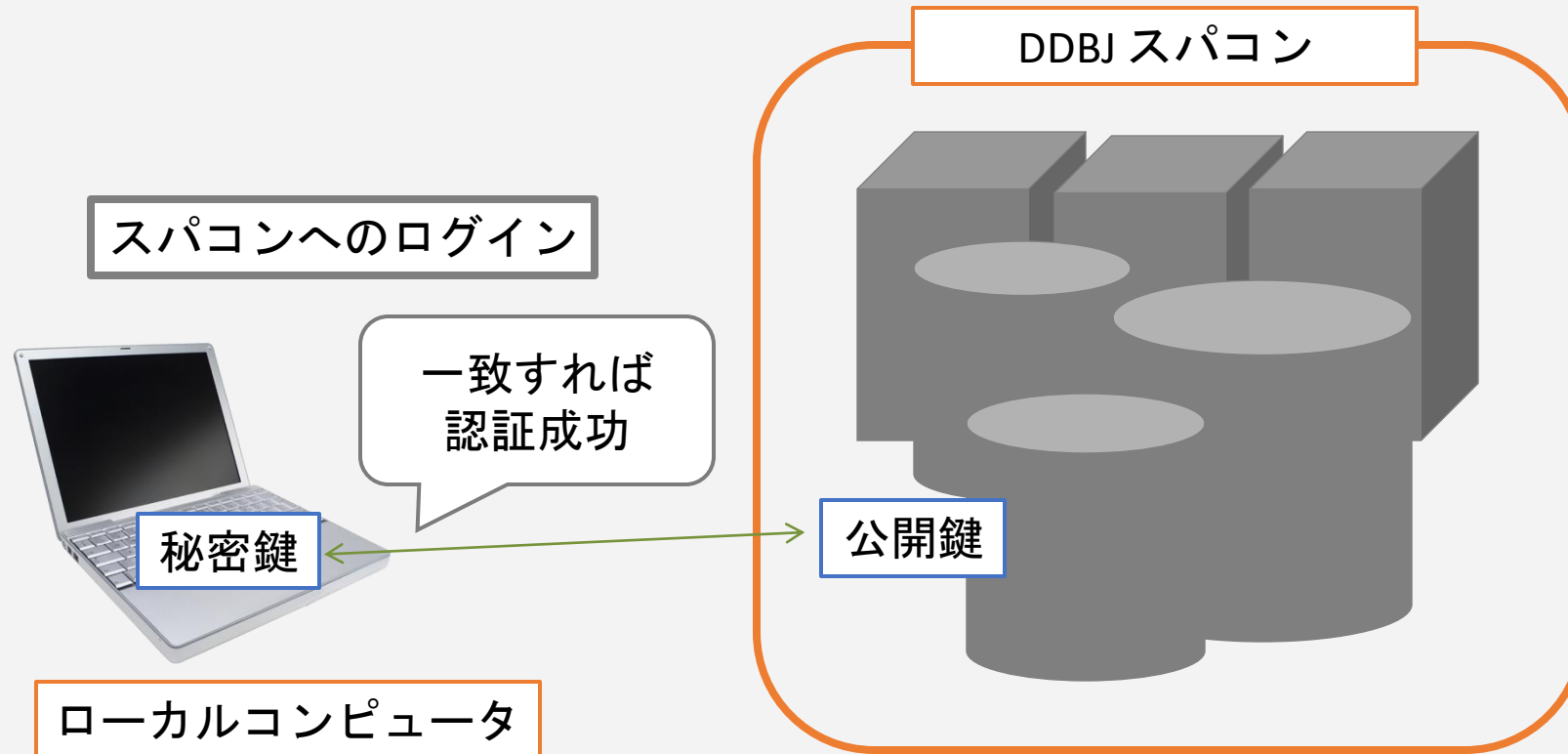


公開鍵にはパスフレーズをつけておくことを推奨

* 公開鍵の登録はスパコンアカウント申請時に行う。また、
<https://sc-account.ddbj.nig.ac.jp> でログインし、追加で公開鍵を登録することもできる。

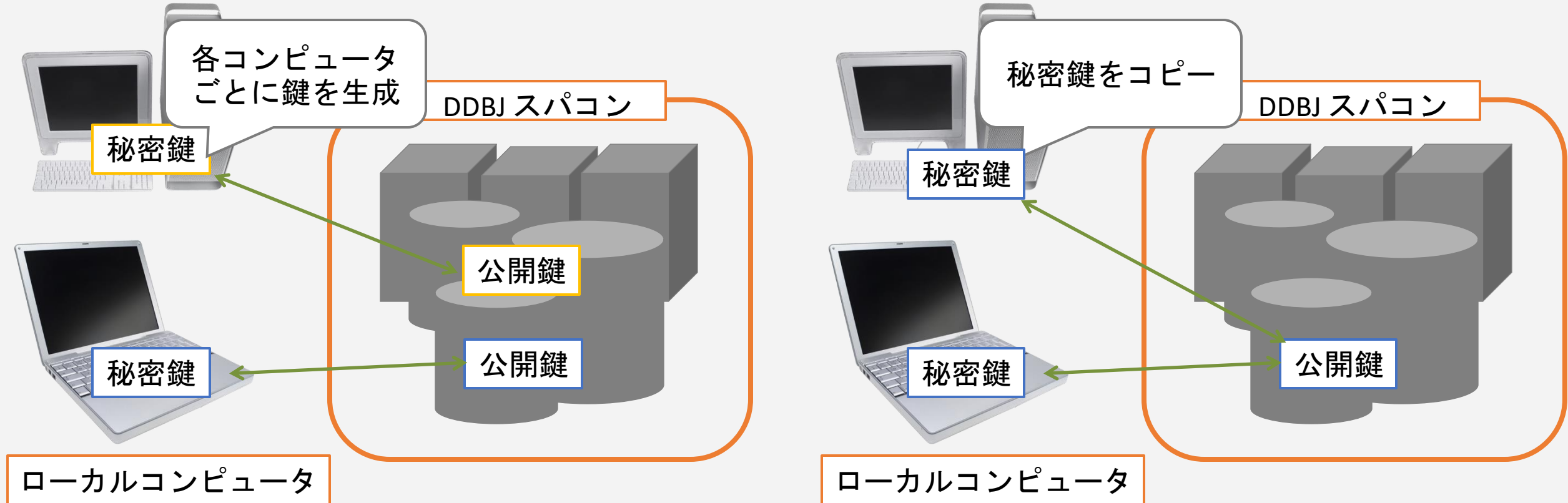
鍵認証

- スパコンへはパスワード認証ではなく、鍵交換による認証で接続する。



他の端末から接続するには？

↓ [方法1] それぞれのマシンで鍵ペアを生成し、公開鍵をスパコンに追加登録



[方法2] 最初に作った秘密鍵を別のマシンにコピーして使い回す ↑
(秘密鍵は他の人から読み取られる可能性のない場所におく必要がある)

スパコン接続に関するトラブルシューティング & 質問

- パスフレーズ入力時には、入力された文字は画面には表示されません。
- 秘密鍵にパスフレーズを設定しなくても良いか？

設定することを推奨するが、自動化目的などの場合、設定しないこともある

- 秘密鍵の保存場所に注意

ホームディレクトリ以下 (他の人から見えない場所) に保存することを推奨
外付けドライブ等だとうまくいかない可能性あり

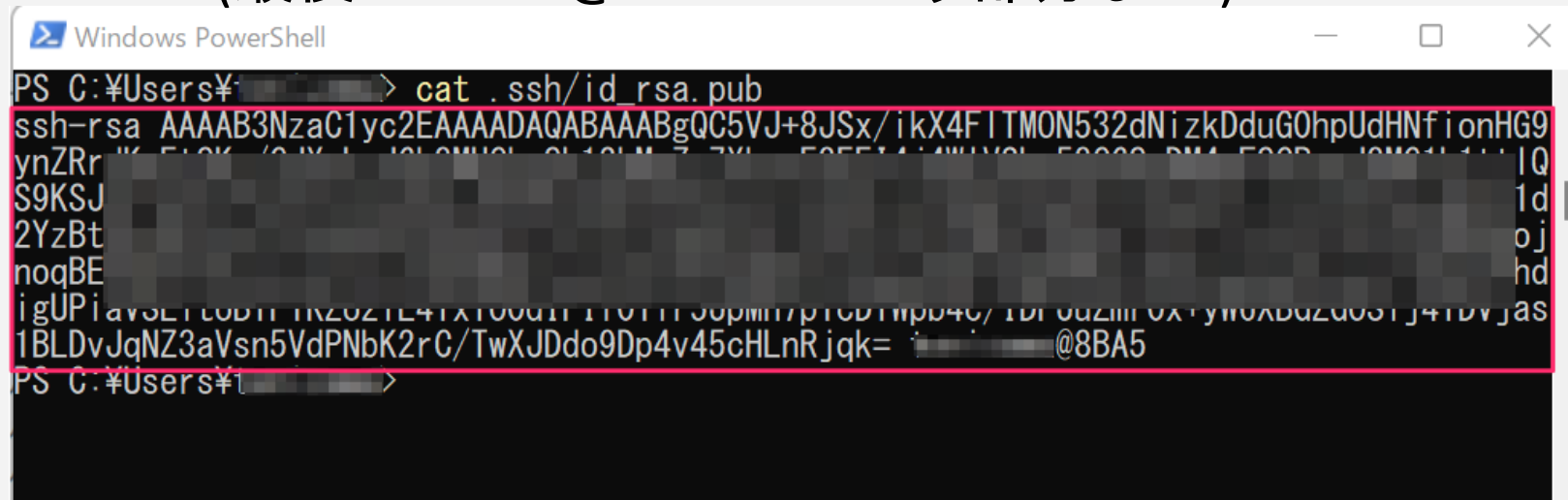
- アカウント名=ユーザー名、ホスト名=サーバー名 と考えてください。

使用するソフトによって使われる言葉が異なることがありますが、上記は同じ意味で使われています。

アカウント名は登録証に記載された名称、スパコン接続の場合、ホスト名は gw.ddbj.nig.ac.jp です。

スパコン接続に関するトラブルシューティング & 質問

- 所属機関によっては遺伝研スパコンへの接続が制限されている場合があります
ネットワーク担当者に依頼し、通信を許可してもらう
- 公開鍵はメール等で送っても良いか？
OK. 公開鍵から秘密鍵を生成することはできない (逆に秘密鍵から公開鍵を作成することはできるので、秘密鍵をメールやネットワーク経由で送ることは推奨されない)
- 公開鍵を保存するときは、先頭の"ssh-rsa"も含めてコピー&ペースト
(最後の xxxxx@xxxxxx という部分まで)



```
Windows PowerShell
PS C:\Users\¥> cat .ssh/id_rsa.pub
ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAQGC5VJ+8JSx/ikX4F1TMON532dNizkDduG0hpUdHNfionHG9
ynZRr...
S9KSJ...
2YzBt...
noqBE...
igUPiavSE1t0D11RZ0Z1E41X100d111011130p1m17p10D1np040/1D10u21m10X1y100X0u20031j41Dv1as
1BLDvJqNZ3aVsn5VdPNbK2rC/TwXJDdo9Dp4v45cHLnRjqk= @8BA5
PS C:\Users\¥>
```

講習の内容

- 遺伝研スパコンの概要
- 必要なソフトのインストール (事前準備の確認)
- スパコンへの接続 (事前準備の確認)
- 基本コマンド

遺伝研スパコンはLinuxで動いている

LinuxとUNIX

UNIX … 1960年代にAT&Tベル研究所で作られたサーバー用OS (商用)

MacOSはUNIXをベースにしている

Linux … Linus Torvalds氏によって作られたUNIX互換のOS (オープンソース)

RedHat系、Debian系などさまざまな系統 (ディストリビューション)がある

Windows Subsystem for Linux (WSL)をインストールするとWindows上でLinux環境を動作させることができる。

スパコンでは Debian系の ubuntu OS が使われている

遺伝研スパコンは共用の計算資源

ゲートウェイノードは、ログインやファイル転送専用。計算処理はログインノードで行うか、バッチジョブを投入して行うこと。

CPUやネットワーク帯域を独占しない

意図せず多数のCPUを使って計算を行うと他のユーザーのジョブに影響を与えることがある
外部サーバーへの通信が短時間に集中すると、相手方サーバーへの通信が遮断されてしまうことがある。