



最新事例に学ぶ 創薬研究領域向け AWS クラウド活用セミナー 2023

創薬研究における AWSのHPC関連サービス紹介

石尾 千晶

アマゾンウェブサービスジャパン合同会社
ソリューション アーキテクト

自己紹介

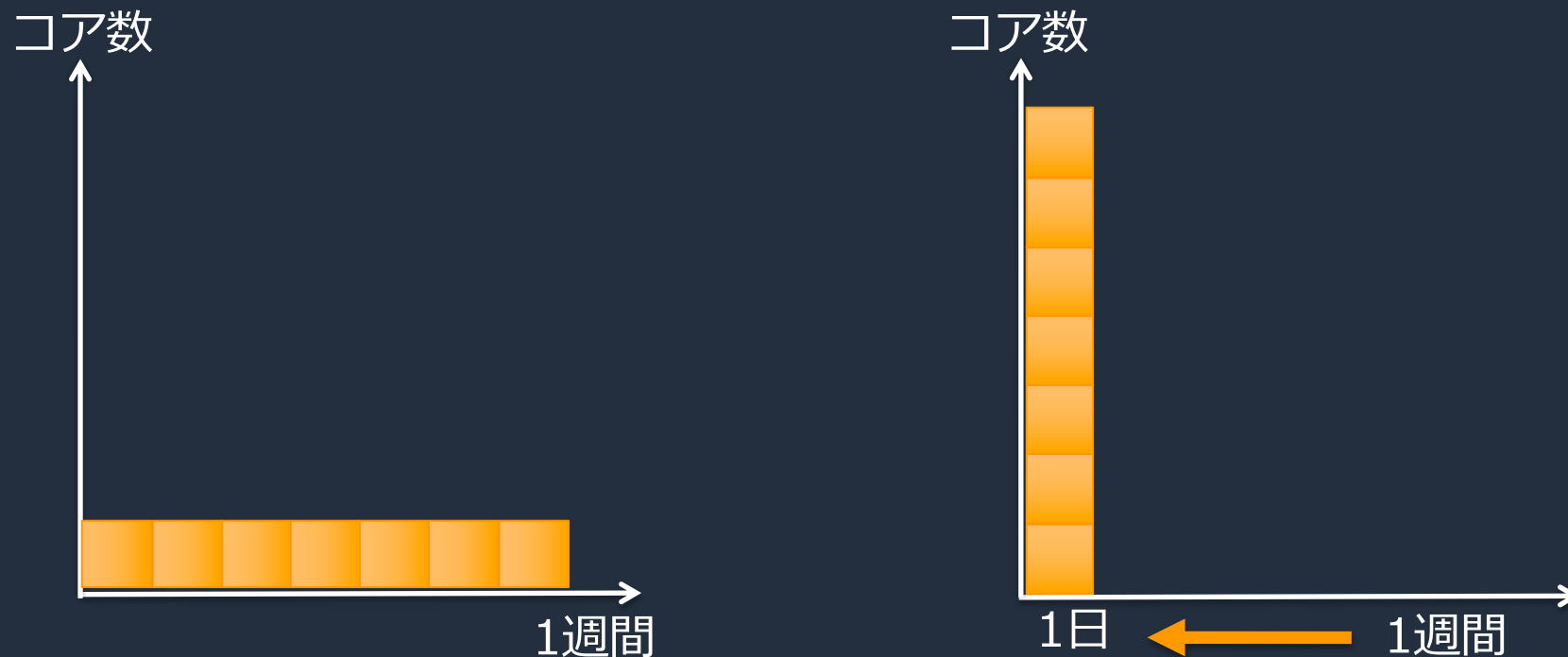
石尾 千晶 (Chiaki Ishio)

アマゾンウェブサービスジャパン合同会社
ソリューション アーキテクト

製薬業界のお客様向けに、
クラウド活用の技術支援をおこなっています



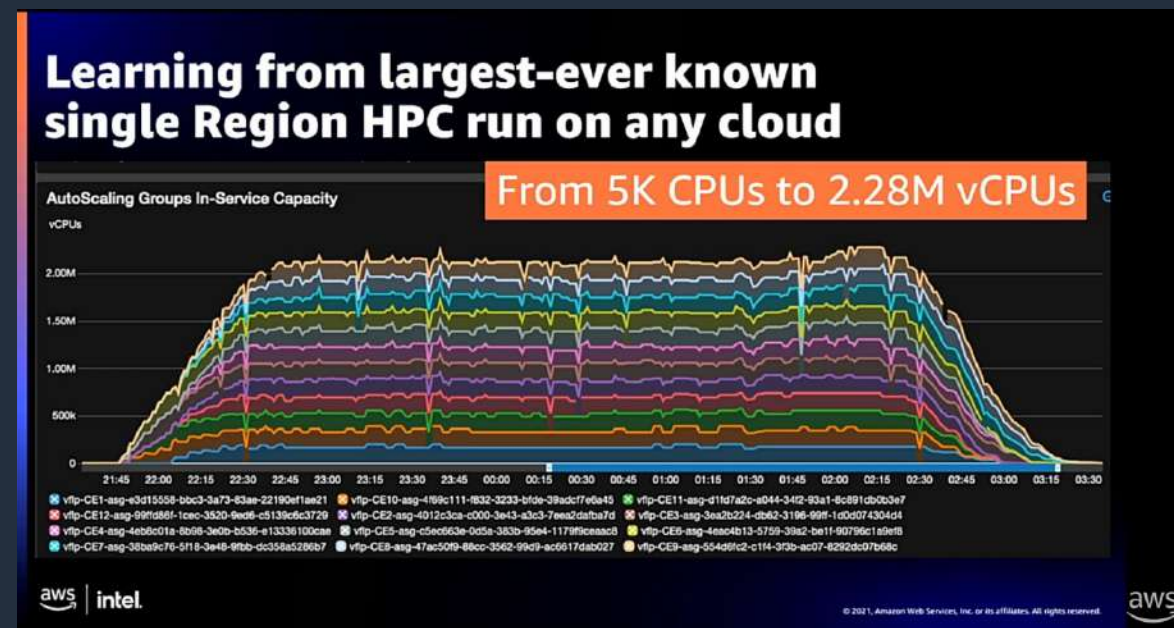
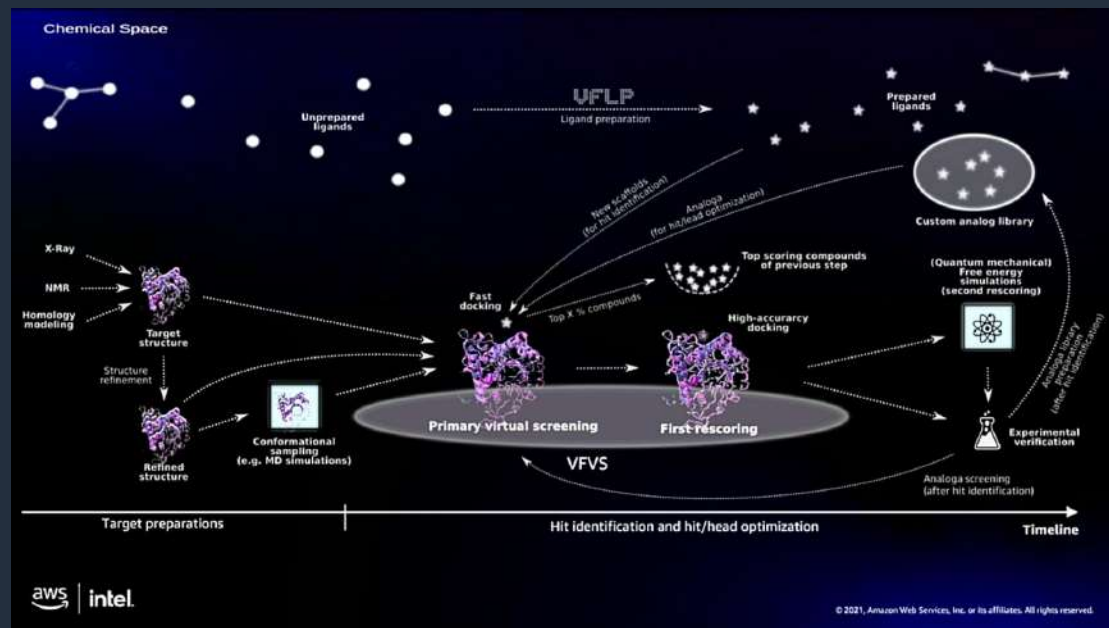
クラウド HPC の魅力： スケーラビリティの活用による計算時間短縮



従来は手持ちの限られたリソースで、逐次処理していたジョブも
AWSなら必要な台数、インスタンスを起動して、一斉処理。
しかも費用は「時間×台数」なのでどちらも同じ。

創薬領域での事例1 : Harvard Medical School 様

- 標的となるタンパク質と化合物の相互作用を予測するためのドッキングシミュレーション
- スケーラブルなドッキングシミュレーションを実現するために、VirtualFlow というソフトウェアを開発
- AWS Batch + Spot Instance の活用により高いコスト効率を実現
- **最大時 228万 vCPU を使用することで、数十億化合物規模のドッキングシミュレーションを数時間で完了**



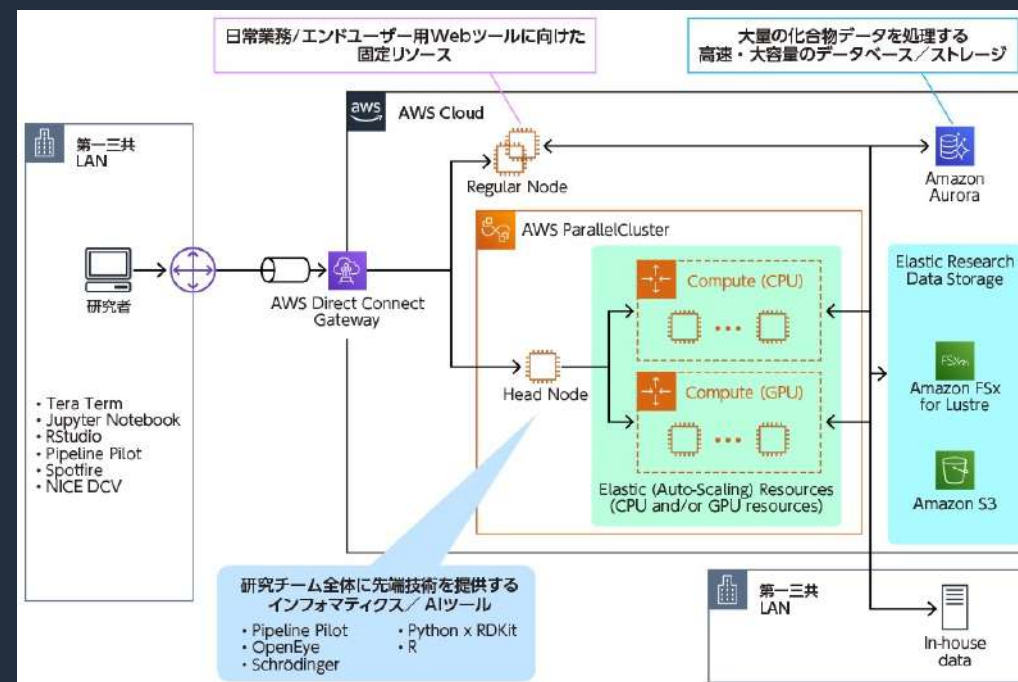
創薬領域での事例2：第一三共株式会社様

研究者のニーズを満たすプラットフォームを作る

- ①利用中の研究ツールやソリューションを使い続けられること
- ②計算処理ボリュームの変動に動的に対応できること

使い慣れた研究ツールや Web ツールからプラットフォームにアクセスできる

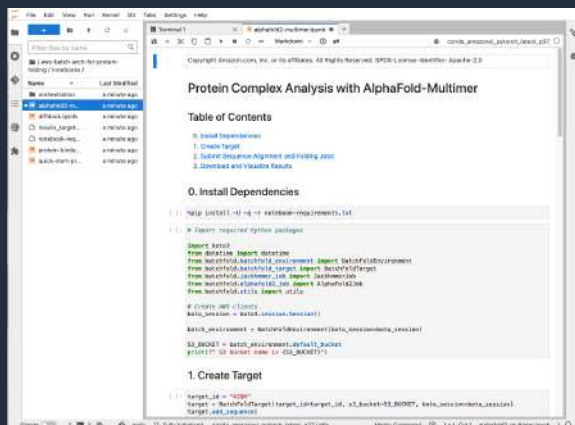
- 使用しているツールの例：
Pipeline Pilot, OpenEye, Schrödinger,
Python x RDKit, R



そのほかの創薬研究での HPC 活用例

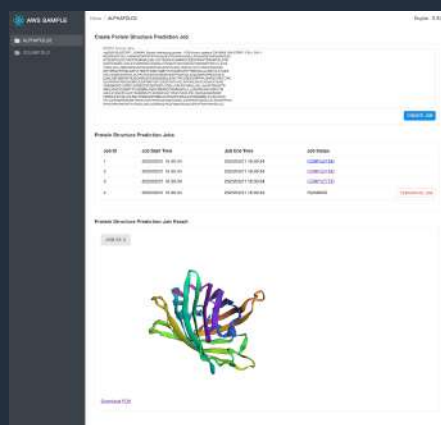
- 分子動力学シミュレーション、ドッキングシミュレーション、タンパク質構造予測など
- 目的に応じてアプリケーション、インターフェース、バックエンドの構成を変更可能

Amazon SageMaker Notebook



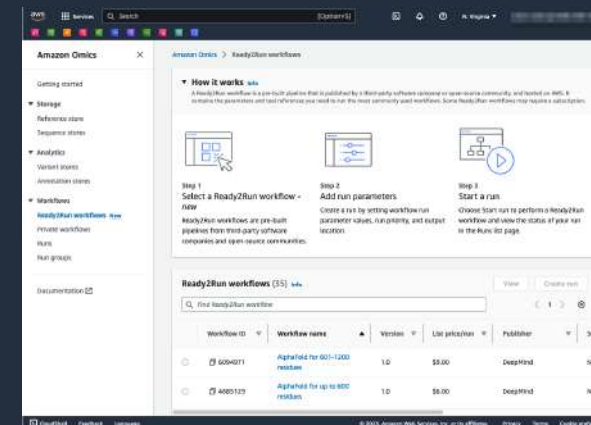
[AWS Batch Architecture for Protein Folding and Design](#)

目的に特化したウェブアプリ



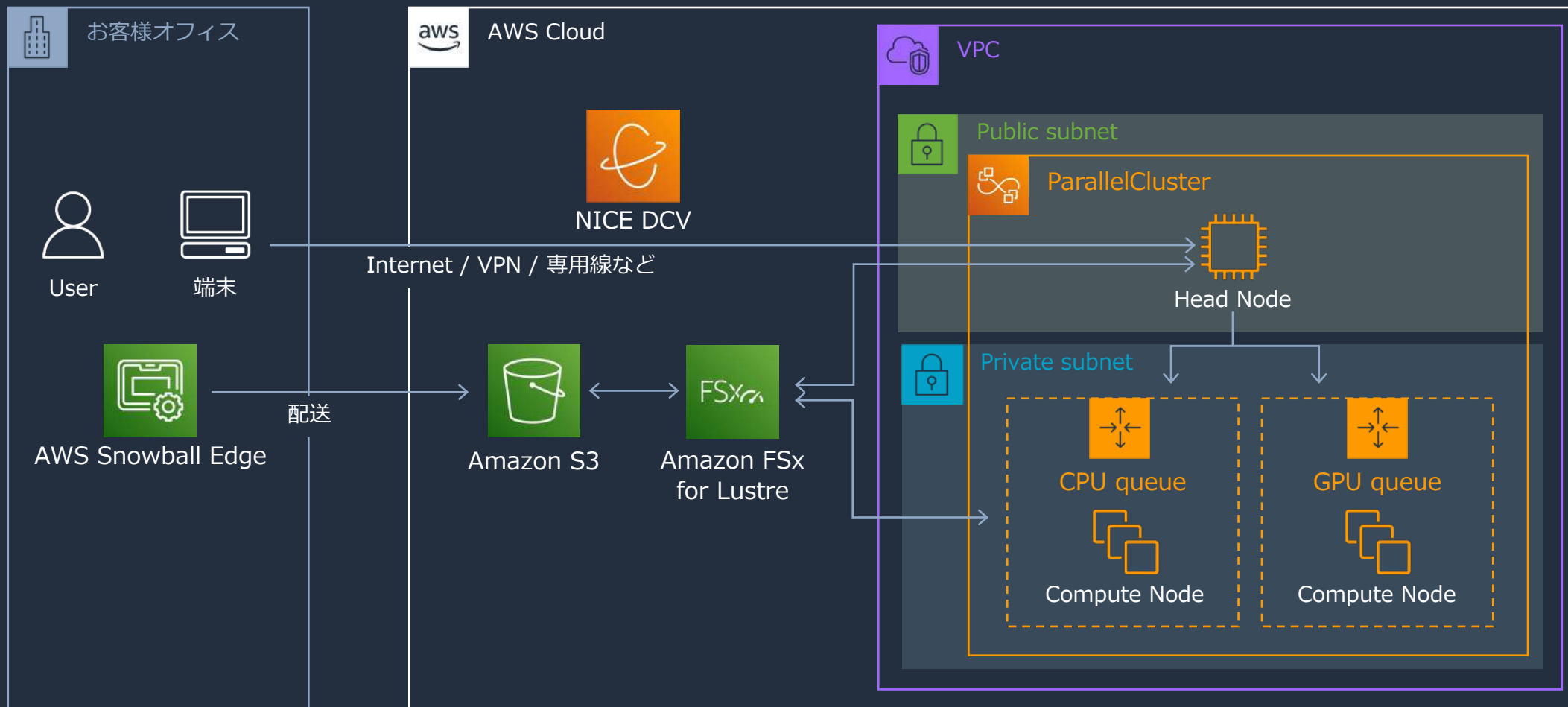
[AlphaFold2 Webapp on AWS](#)

Amazon Omics (Ready2Run Workflows)

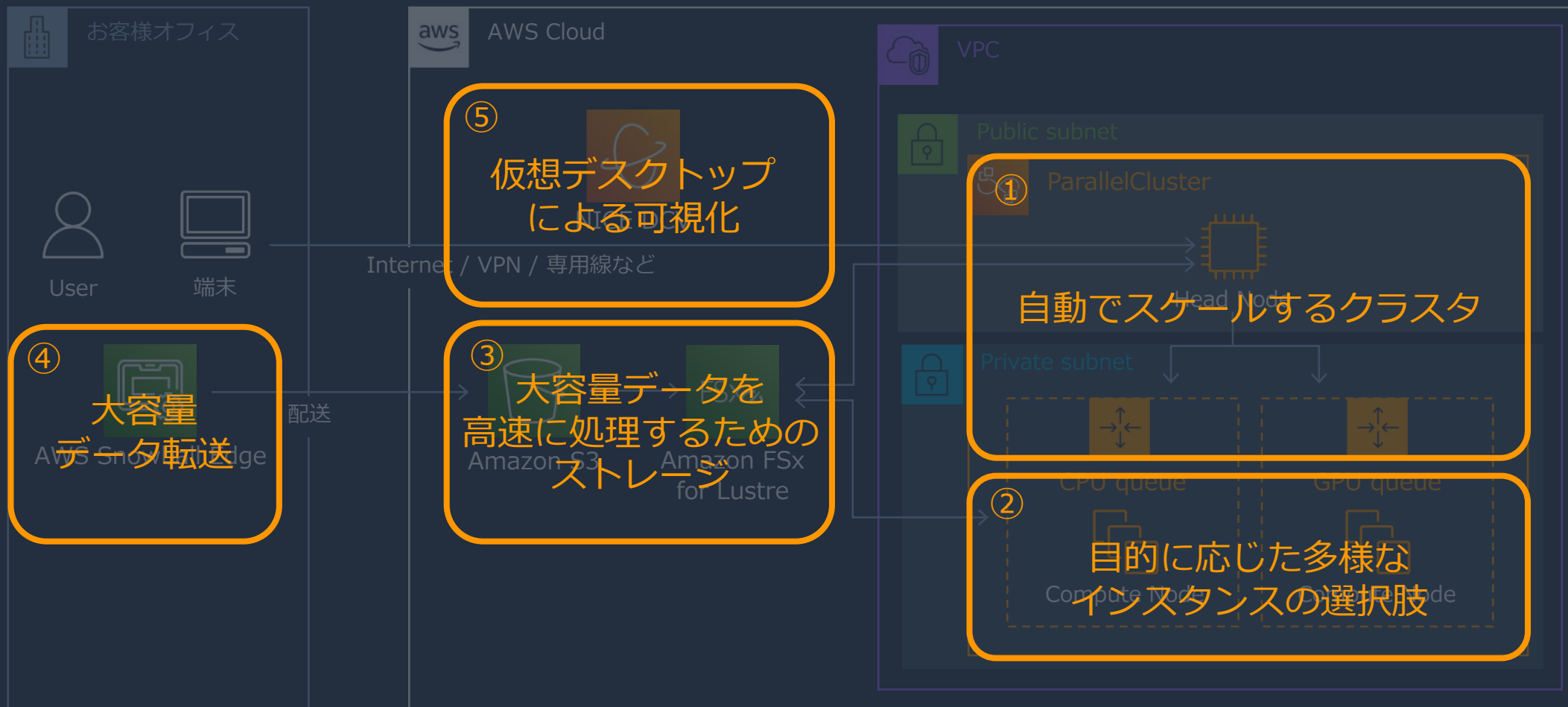


[Amazon Omics の新機能紹介ブログ](#)

AWS 上での HPC 構成例

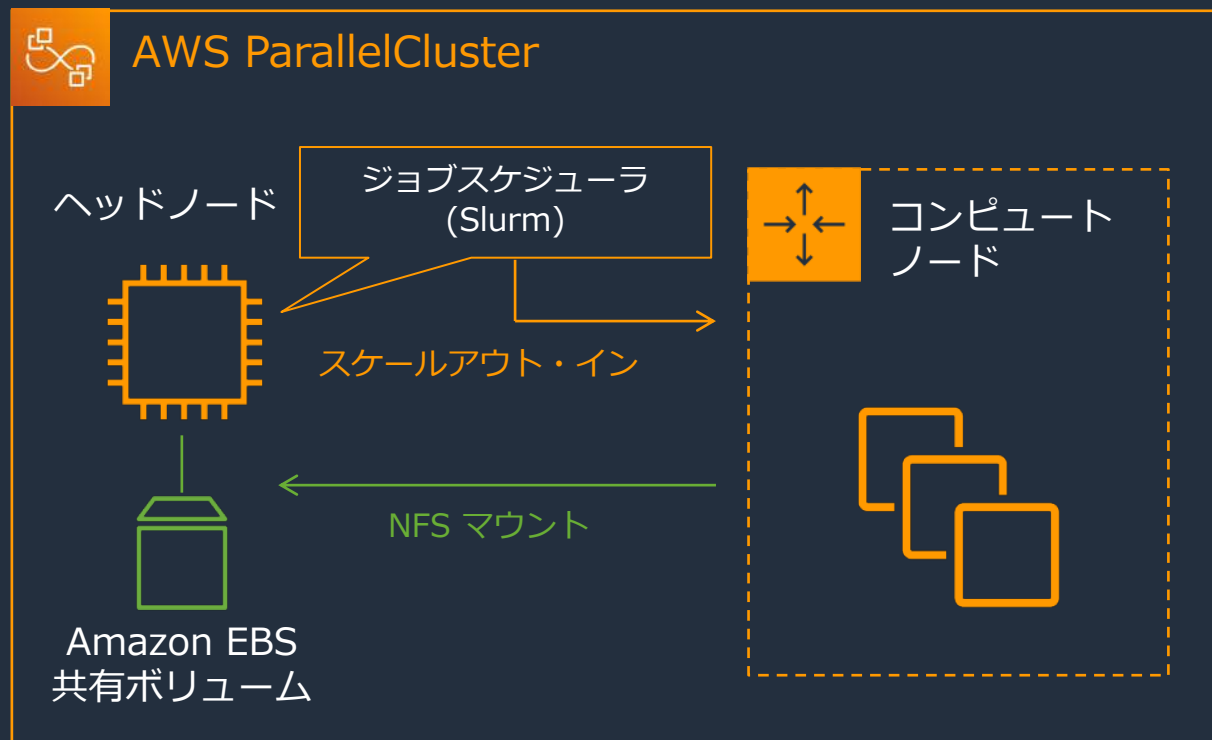


AWS 上での HPC 構成例



① 自動でスケールするクラスタ : AWS ParallelCluster

数コマンド操作でジョブ投入に応じて自動でスケールするクラスタを AWS 上に構築可能な AWS 公式のオープンソースソフトウェア



① 自動でスケールするクラスタ : AWS ParallelCluster

コマンドラインツールに加え、ユーザーの AWS アカウント内で展開可能なウェブツールも提供

コマンドラインツールでクラスタを作成・管理

```
$ pcluster create-cluster <NAME>
```

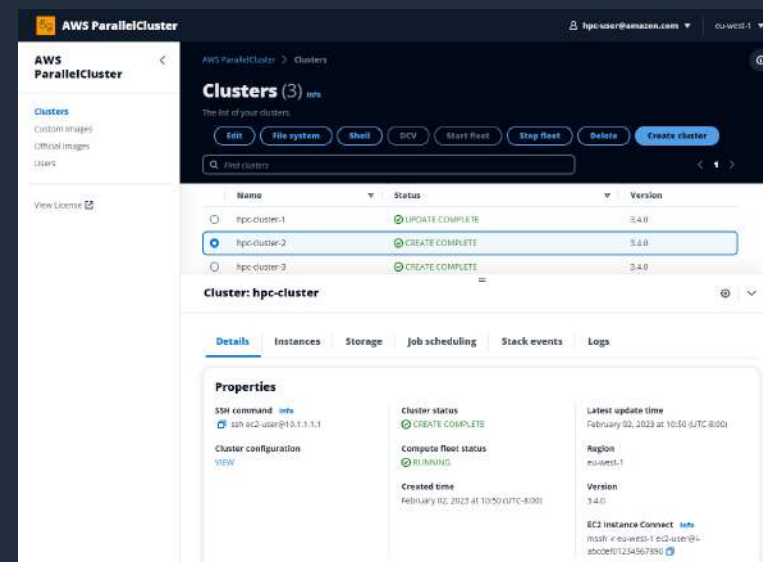
設定ファイル例 (YAML 形式)

```
HeadNode:
  InstanceType: c6g.large
Scheduling:
  Scheduler: slurm
SlurmQueues:
- Name: queue0
  ComputeResources:
  - Name: queue0-c6g16xlarge
    MinCount: 0
    MaxCount: 40
    InstanceType: c6g.16xlarge
```

or

ウェブツールでクラスタを作成・管理

AWS ParallelCluster UI



② 目的に応じた多様な選択肢 : Amazon EC2

高性能計算向けインスタンスタイプの例

高性能 CPU の選択肢



Intel Xeon processor
(x86_64 arch)

M6i インスタンス

Ice Lake
最大時全コア 3.5 GHz 駆動

M5zn インスタンス

Cascade Lake
最大全コア 4.5 GHz 駆動



AMD EPYC processor
(x86_64 arch)

Hpc6a インスタンス

EPYC Milan
HPC特化

M6a インスタンス

EPYC Milan
最大 3.3 GHz 駆動



AWS Graviton Processor
(64-bit Arm arch)

C7g インスタンス

64bit Arm Neoverse V1ベース
AWS Graviton3 CPU 搭載

Hpc7g インスタンス

AWS Graviton3 CPU 搭載
HPC特化

アクセラレータの選択肢



NVIDIA GPU

P5 インスタンス*

H100 GPU 搭載

P4d インスタンス

A100 GPU 搭載

G5 インスタンス

A10G GPU 搭載

G4 インスタンス

T4 GPU 搭載



AMD Xilinx FPGA

F1 インスタンス

Virtex UltraScale+
VU9P 搭載

③ 高速分散ファイルシステム : Amazon FSx for Lustre

- 高速な分散ファイルシステムの Lustre をフルマネージドで提供
- AWS ParallelCluster と連携させ、クラスターに自動マウントが可能
- 階層型ストレージの機能もあり、S3 と透過的にデータの import/export が可能

Amazon S3 に格納されたデータが
Amazon FSx for Lustre ファイルシステムに
ロードされて処理される



④ 大容量データの転送：AWS Snow ファミリー

- ハードウェアアプライアンスを配送することで、オンプレミス – クラウド間のデータ移行を高速化
- データの自動暗号化や不正開封防止筐体によるセキュリティ確保
- 一度に大量のデータを送ることができるが、配送に時間や手間がかかる点には注意が必要

AWS Snowball Edge

Storage Optimized の場合、
約 22 kg の筐体に 100 TB の HDD を搭載

AWS Snowcone

約 2 kg と軽量の筐体に 8TB の HDD を搭載

AWS DataSync エージェントを実行でき
オンライン転送用環境としても利用可能

※ AWS Snowball Edge/AWS Snowcone 共に東京リージョンで利用可能



AWS Snowball Edge



AWS Snowcone

⑤ 仮想デスクトップによる計算結果の可視化 : NICE DCV

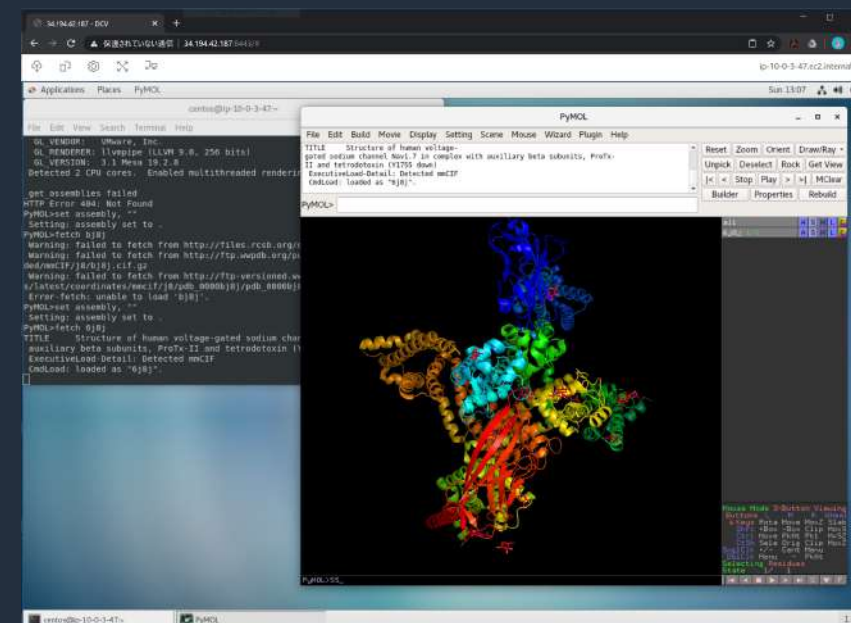
クラウド上のデスクトップ画面をストリーミングするためのソフトウェア

特徴

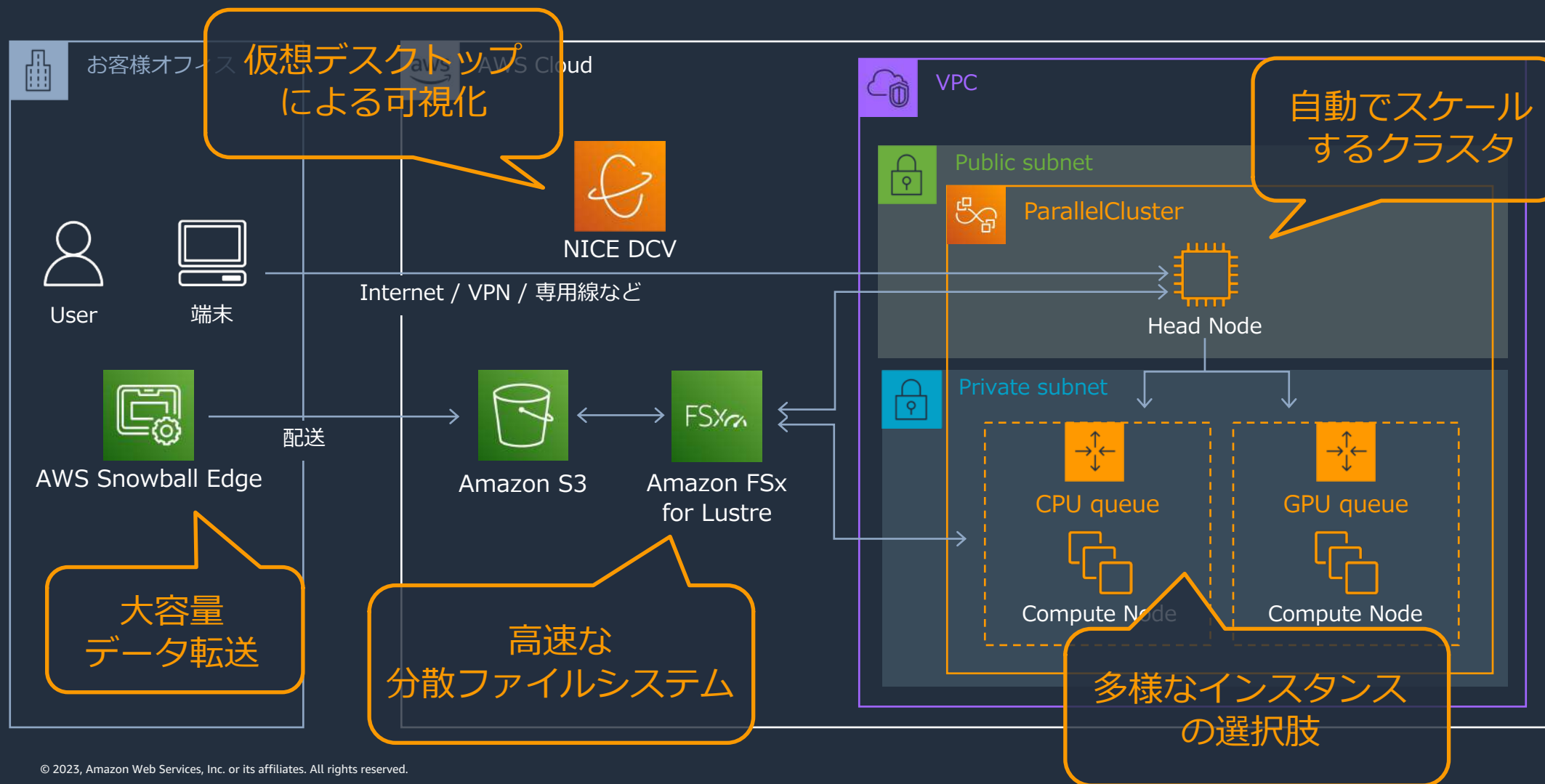
- 専用のプロトコルによる高速かつスムーズなストリーミング
- サーバはWindows、Linuxの両方に対応
- GPUにも対応し、G4dn/G4adインスタンス等を利用することでより高速な描画が可能
- ネイティブクライアントの他、HTML5対応ブラウザからも利用可能

コスト

- Amazon EC2で利用する場合はライセンスコスト無しで利用可能 (EC2以外での利用は有償)



まとめ：AWS 上での HPC 構成例





Thank you!