

CBI 学会 2022 年大会 スポンサーセッション

SS-02 アマゾン ウェブ サービス ジャパン合同会社

『創薬研究におけるクラウド活用の実際～第一三共でのデータ駆動型創薬化学研究基盤～』

15:30 - 15:50 (20 min)

AWS サービスを活用した High Performance Computing (HPC) 環境の構築

15:50 - 16:50 (60 min)

創薬研究におけるクラウド活用の実際

～ 第一三共でのデータ駆動型創薬化学研究基盤～

16:50 - 17:00 (10 min)

Q&A



AWS サービスを活用した High Performance Computing (HPC) 環境の構 築

Amazon Web Services
Japan

中島 丈博

2022/10/25

© 2022, Amazon Web Services, Inc. or its Affiliates.



自己紹介



名前：中島 丈博（なかじま たけひろ）

所属：アマゾン ウェブ サービス ジャパン
ソリューションアーキテクト

役割：製薬業界のお客様を中心に
クラウド利用に関する技術支援

- ・ご要望の具現化
- ・課題の解決に向けたご支援
- ・製薬業界のクラウド利用事例のご紹介

本セッションについて

SS-02 アマゾン ウェブ サービス ジャパン合同会社

『創薬研究におけるクラウド活用の実際～第一三共でのデータ駆動型創薬化学研究基盤～』

15:30 - 15:50 (20 min)

AWS サービスを活用した High Performance Computing (HPC) 環境の構築

15:50 - 16:50 (60 min)

創薬研究におけるクラウド活用の実際

～ 第一三共でのデータ駆動型創薬化学研究基盤～

16:50 - 17:00 (10 min)

Q&A

本セッションは
後半の第一三共様セッションの理解を深めて頂くことを目的に
第一三共様が構築された創薬研究基盤でご利用頂いている AWS サービス
についてご紹介いたします

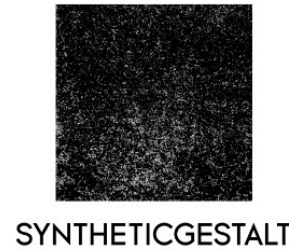
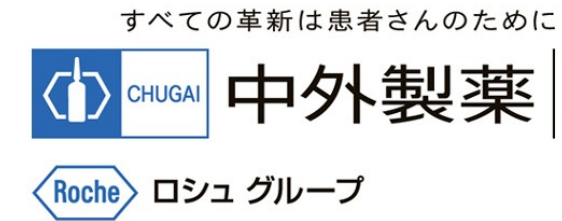
アジェンダ

- 創薬研究環境における AWS 活用
- 第一三共様の環境でご利用頂いている AWS サービス
- AWS からの支援内容

創薬研究環境における AWS 活用

国内における製薬関連のお客様

(一部抜粋)



国内における公的研究・ゲノミクス・予防・介護関連のお客様

(一部抜粋)



京都大学
KYOTO UNIVERSITY



国立循環器病研究センター
National Cerebral and Cardiovascular Center



TAKARA BIO INC.



XCOO [ténku:]

JMDC
● + × ◀

Genesis Healthcare



R:RUNNET

FiNC
Technologies



Moff

wiseman

国内における医療関連のお客様

(一部抜粋)



京都大学
KYOTO UNIVERSITY



藤田医科大学
FUJITA HEALTH UNIVERSITY



国立循環器病研究センター
National Cerebral and Cardiovascular Center



特定機能病院 / 地方独立行政法人 大阪府立病院機構
大阪国際がんセンター

AOI 国際病院 医療法人社団 葵会
AOI UNIVERSAL HOSPITAL



東京都済生会中央病院
TOKYO SAISEIKAI CENTRAL HOSPITAL



平成医療福祉グループ
HEISEI MEDICAL WELFARE GROUP



日本医師会 ORCA 管理機構



国内のユースケース（抜粋）



アステラス製薬株式会社様

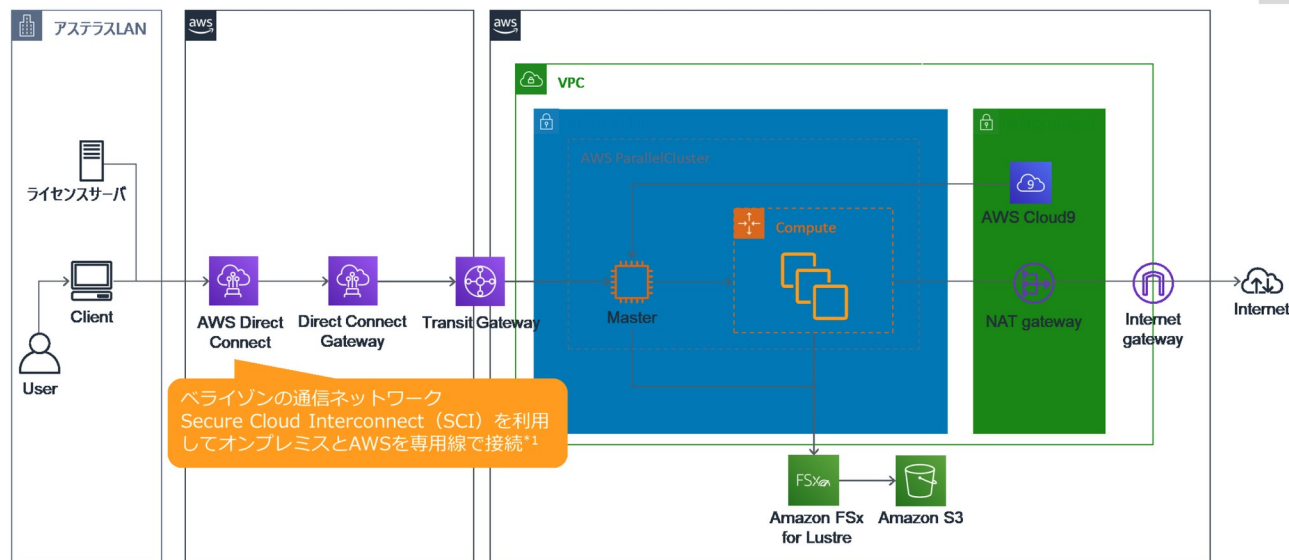
「アステラス製薬の創薬研究における AWS 活用事例

～クラウドによる機械学習ベースの Ultra-large scale virtual screening ～」



AL-Glideを使うためのAWS構成図

24



Transit Gateway等の構築は情報システム部、その他はAWSサポートを受けながら研究員が構築

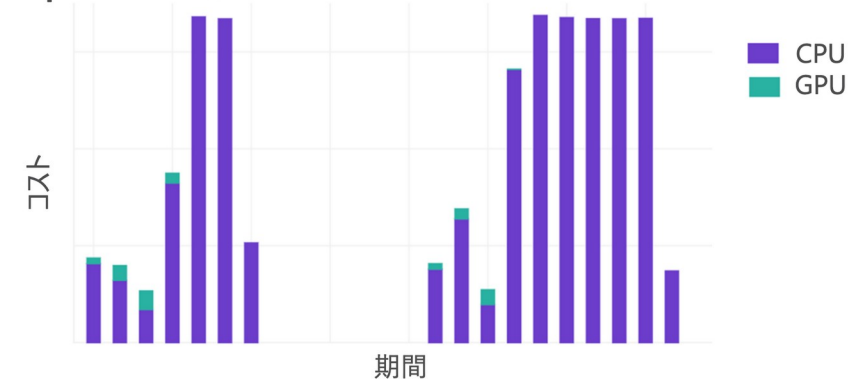
AWS 中島様ご提供

*1) <https://it.impress.co.jp/articles/-/15656>

ParallelClusterで計算終了まで効率的に計算資源を管理できた

25

□AWSのCost explorerで確認



研究員が計算前にインスタンスを起動する必要がなく、ParallelClusterが必要な時にインスタンスを起動し、リソースの無駄使いを防いでくれる。
特にGPUは必要な時にのみ起動することでコストを大幅に抑えられた。



<https://d1.awsstatic.com/local/health/20211118%20drug%20discovery%20EIB%20seminar%20session%204.pdf>



国内のユースケース（抜粋）



塩野義製薬株式会社様

AWS 上へのスケーラブルな HPC 環境の構築によりオンプレミス比で数倍から 10 倍以上の計算期間を短縮
数億の化合物規模のバーチャルスクリーニングを実現

HPC のユースケース

- ゲノム分析
- バーチャルスクリーニング
- タンパク質立体構造解析



シオノギは新中期経営計画『Shionogi Transformation Strategy 2030 (STS2030)』を策定しました。従来の製造業から、ヘルスケアサービスを提供する企業になるための変革には、デジタル変革は必須です。AWS の活用をさらに推進することにより、STS2030 を実現できると考えています

小林 博幸 氏
塩野義製薬株式会社
デジタルインテリジェンス部 部長

<https://aws.amazon.com/jp/solutions/case-studies/shionogi/>

国内のユースケース（抜粋）

日本たばこ産業株式会社様

「JTの創薬研究におけるAWSの活用事例」



Virtual Screening

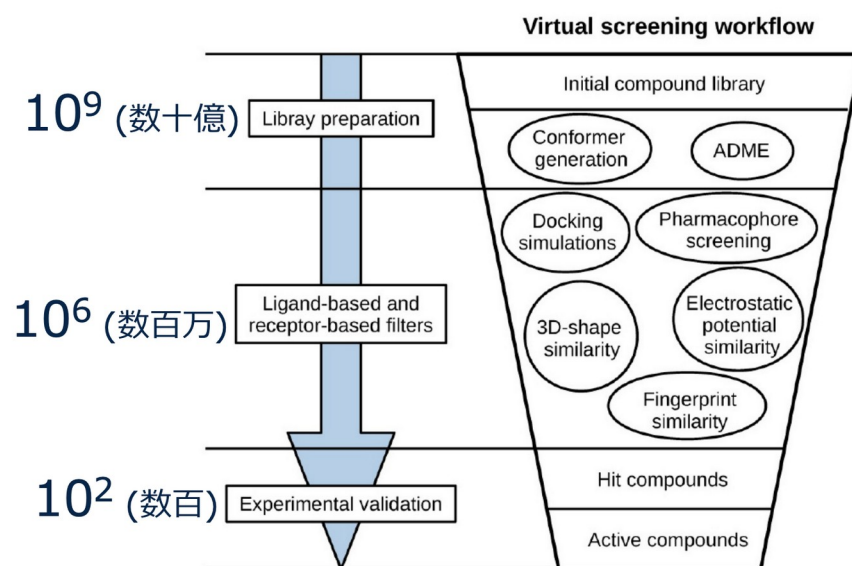


数十億の化合物群から
活性・物性が見込める
化合物群の絞り込み

上流の計算時間が膨大



AWSによる高速化

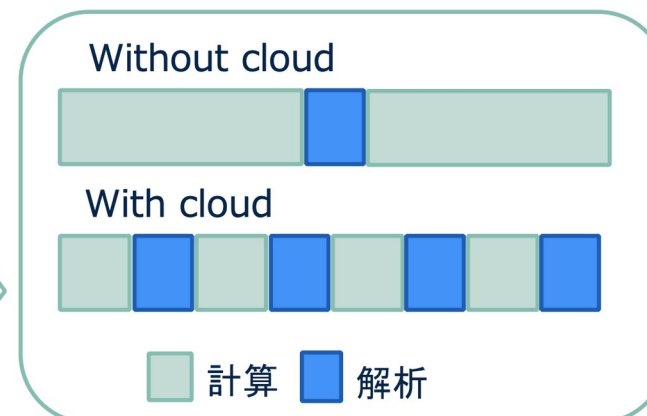
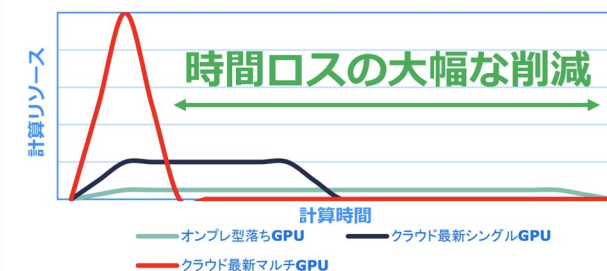
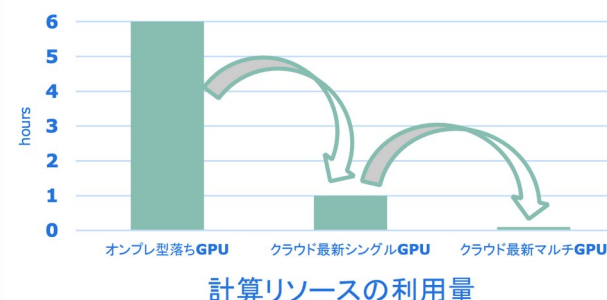


Int. J. Mol. Sci. 2019, 20(6), 1375

AWSを利用した DGL LifeSci



PDBBindの学習計算時間



- 計算サイクルの高速化
- 環境構築までのAWSサポート
- true human in the loop**

https://d1.awsstatic.com/local/health/20220421_session3.pdf

<https://youtu.be/6hLaEd4gyTA>



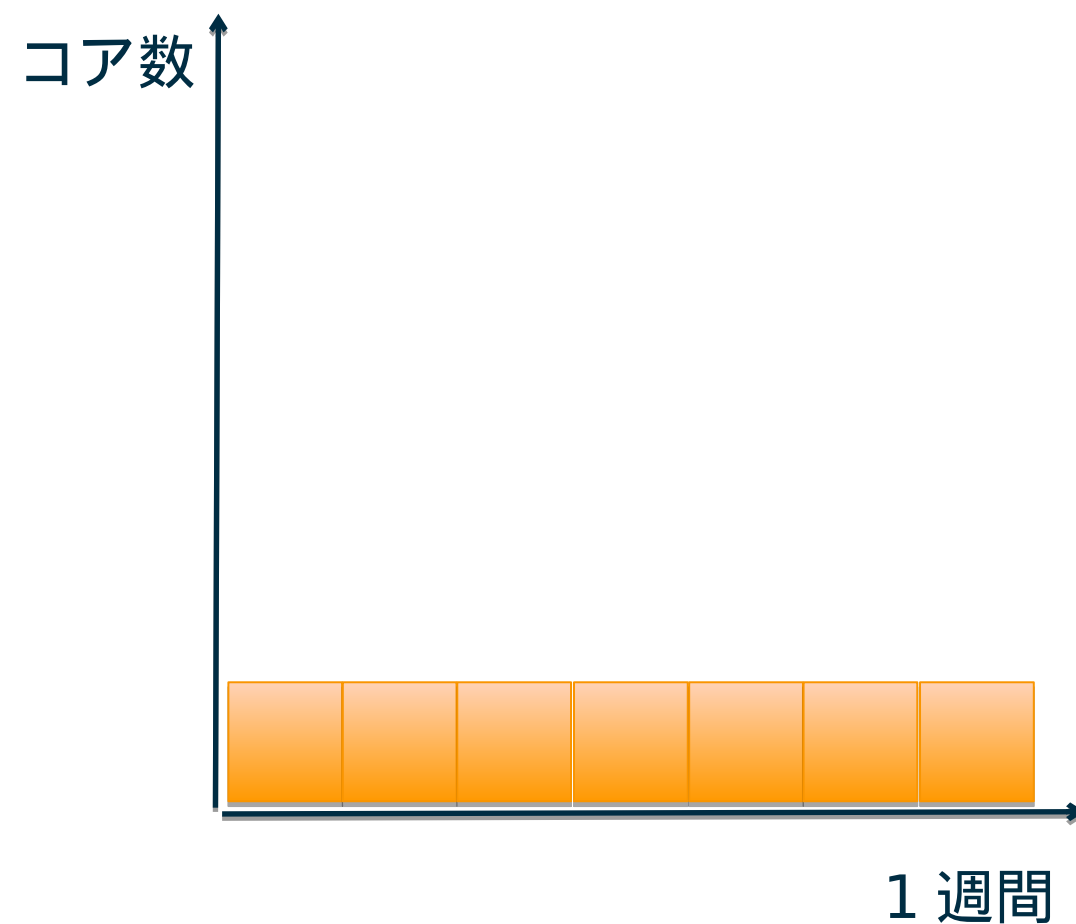
創薬研究環境における **HPC** の課題

- 多数のジョブをすぐに実行したい、短時間で完了したい
- 計算の用途の応じたクラスターを構成したい
- 日々変化する計算需要の対応が大変

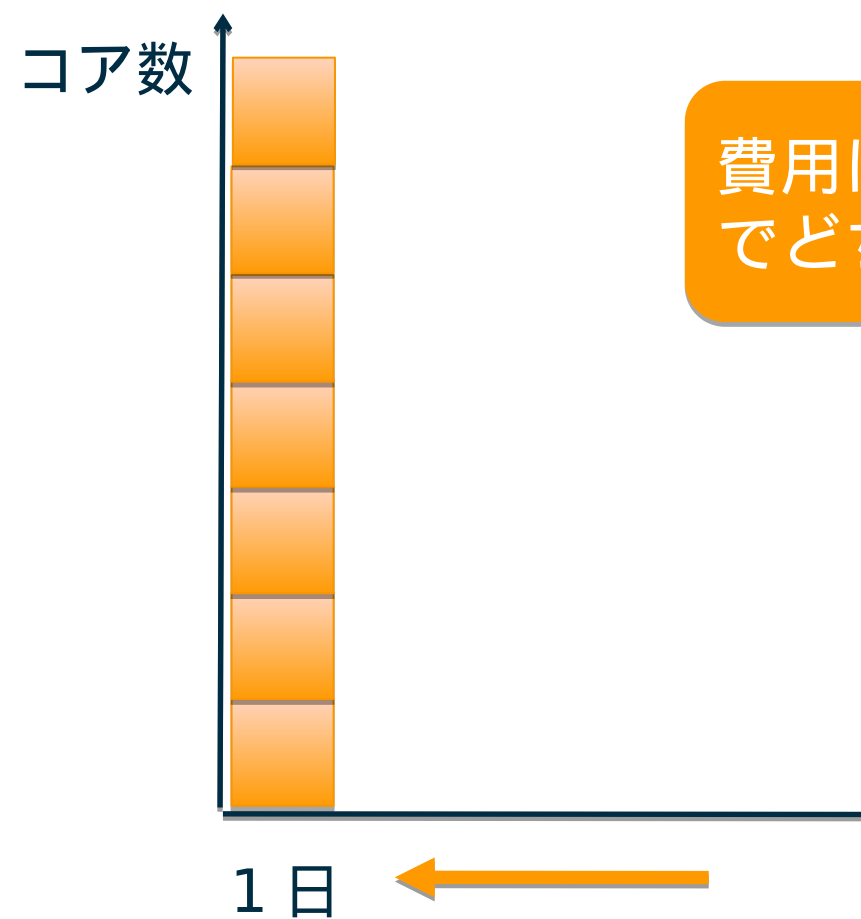


スケーラビリティの活用による計算時間短縮

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を
必要な EC2 インスタンス（仮想サーバー）を起動して一斉処理することが
可能



従来の HPC 環境



AWS の HPC 環境

費用は「時間 × 台数」なのでどちらも同じ！

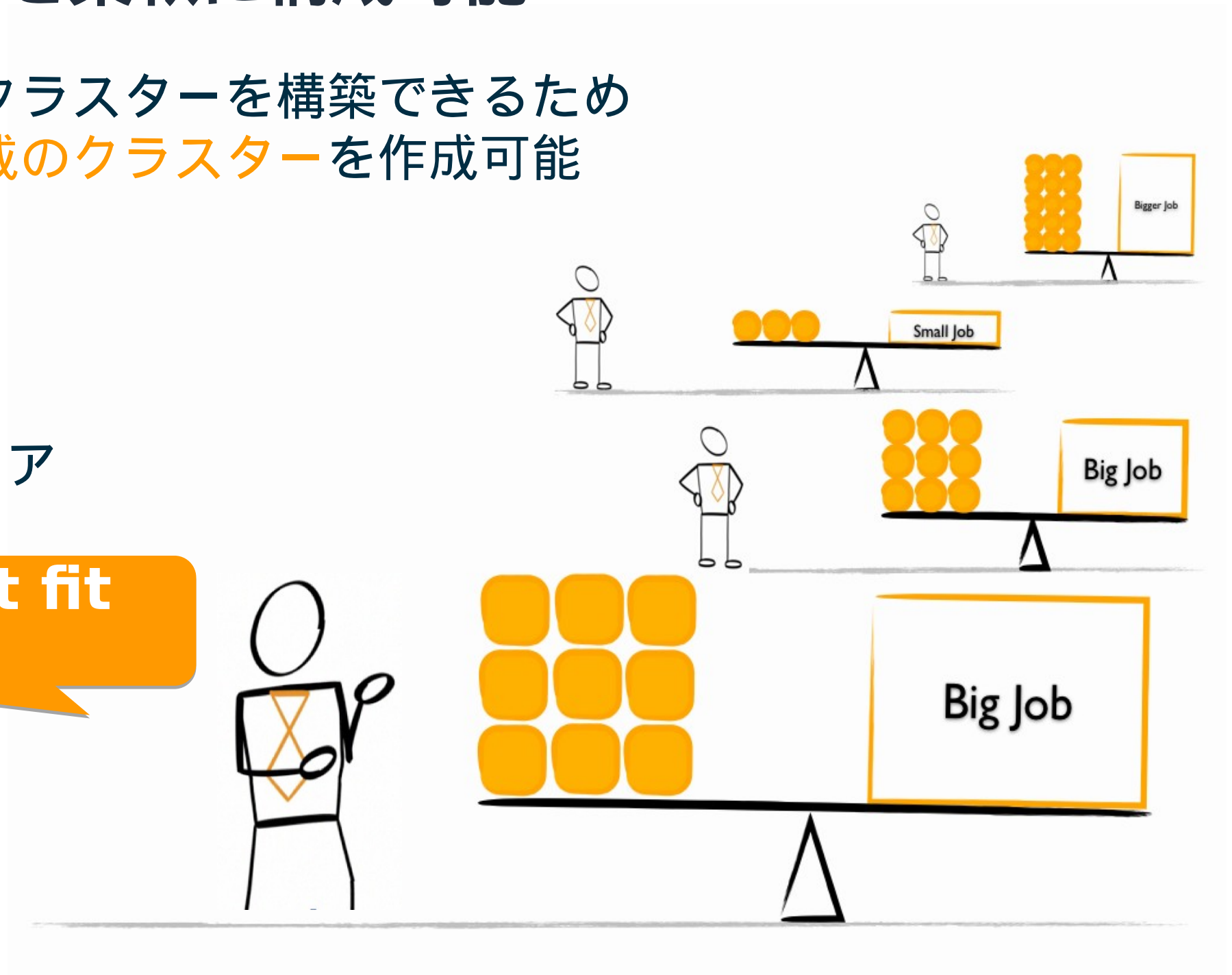


最適な HPC クラスタを柔軟に構成可能

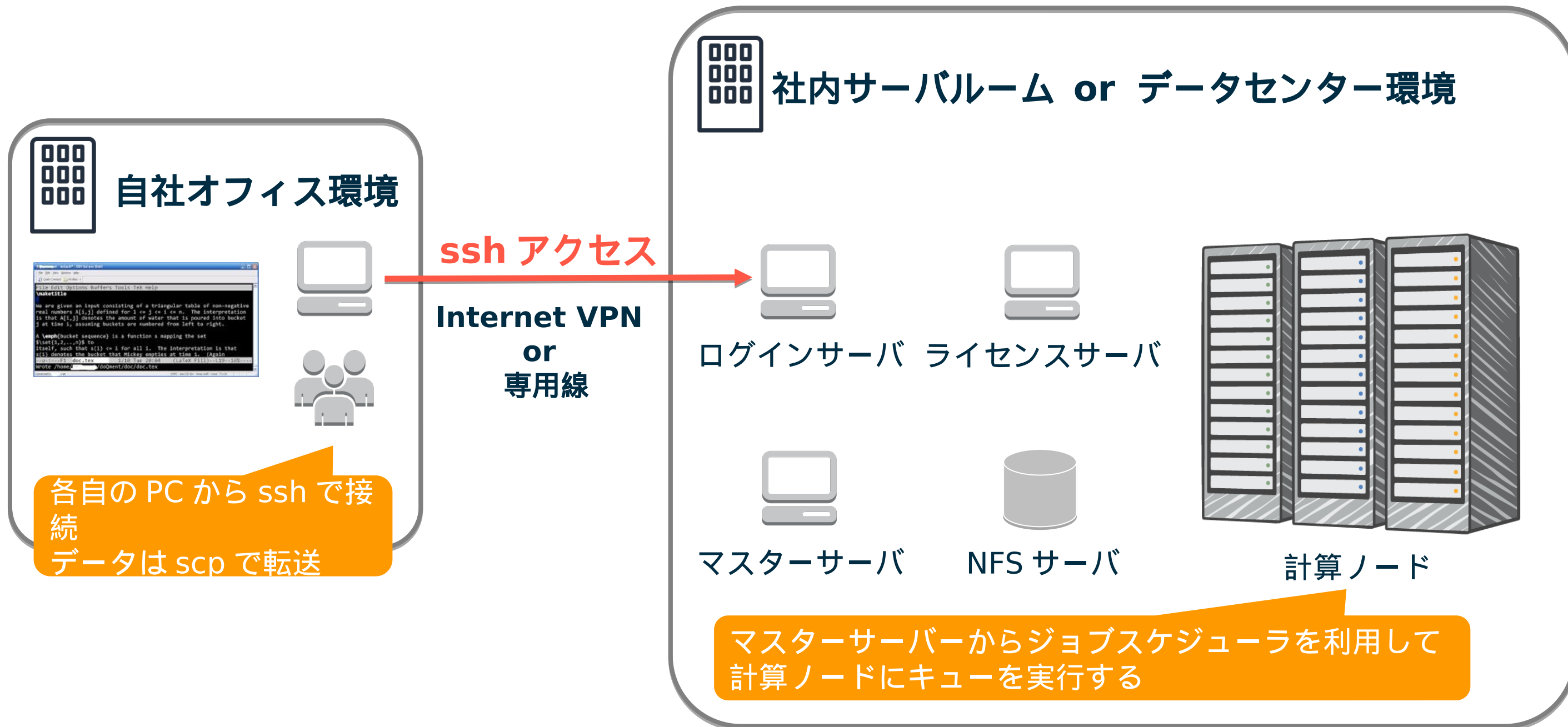
ユーザやタスク単位で専用のクラスタを構築できるため
要件や規模に合わせた最適構成のクラスタを作成可能

- CPU コア / メモリ
- ストレージ
- ネットワーク
- インストールするソフトウェア

One size does not fit
all !!



一般的な HPC 環境

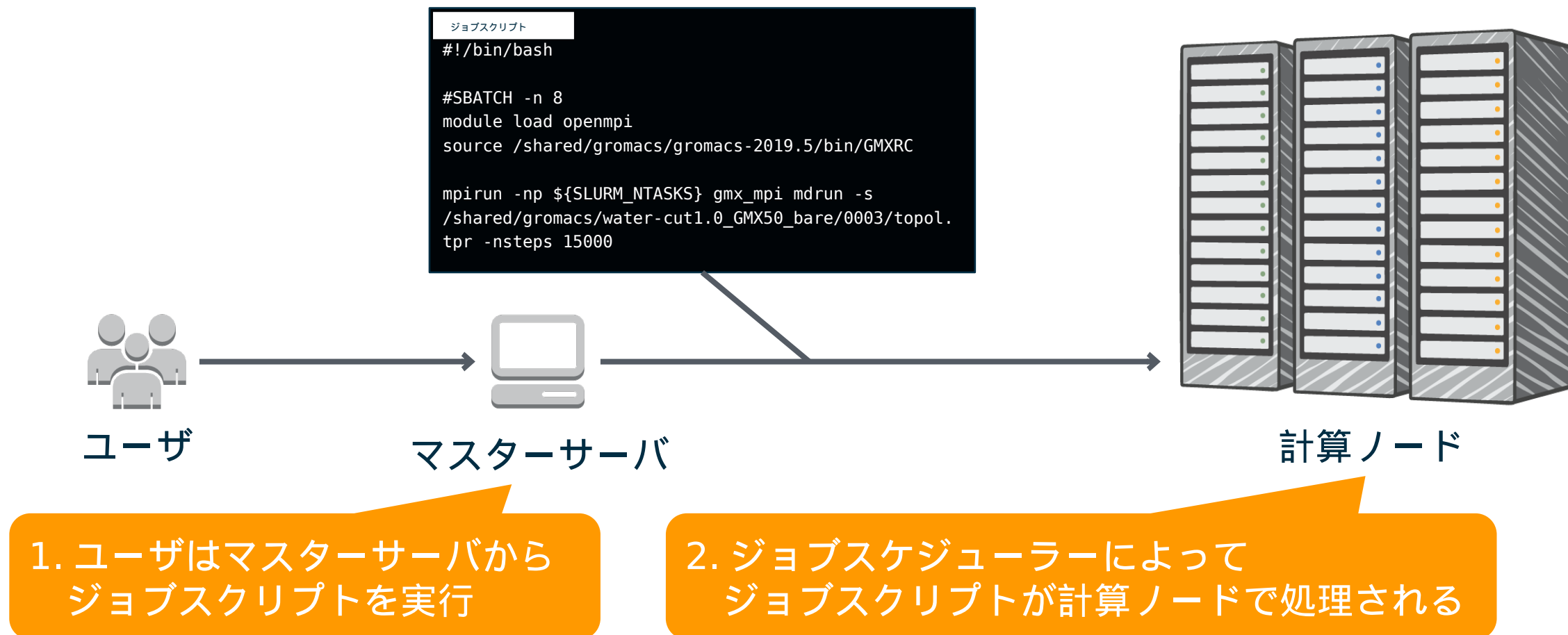


各自の PC から ssh で接続
データは scp で転送

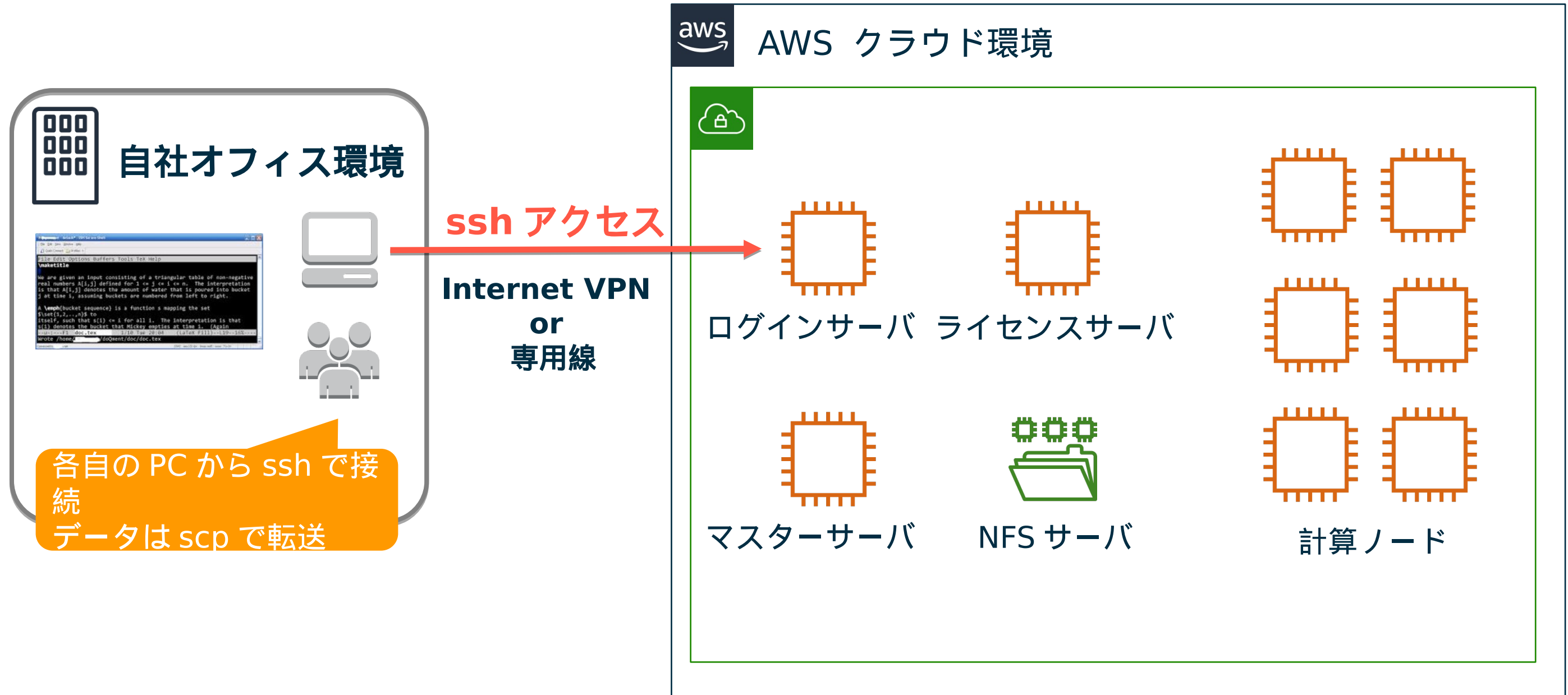
マスターサーバーからジョブスケジューラを利用して
計算ノードにキューを実行する

ジョブスケジューラとは

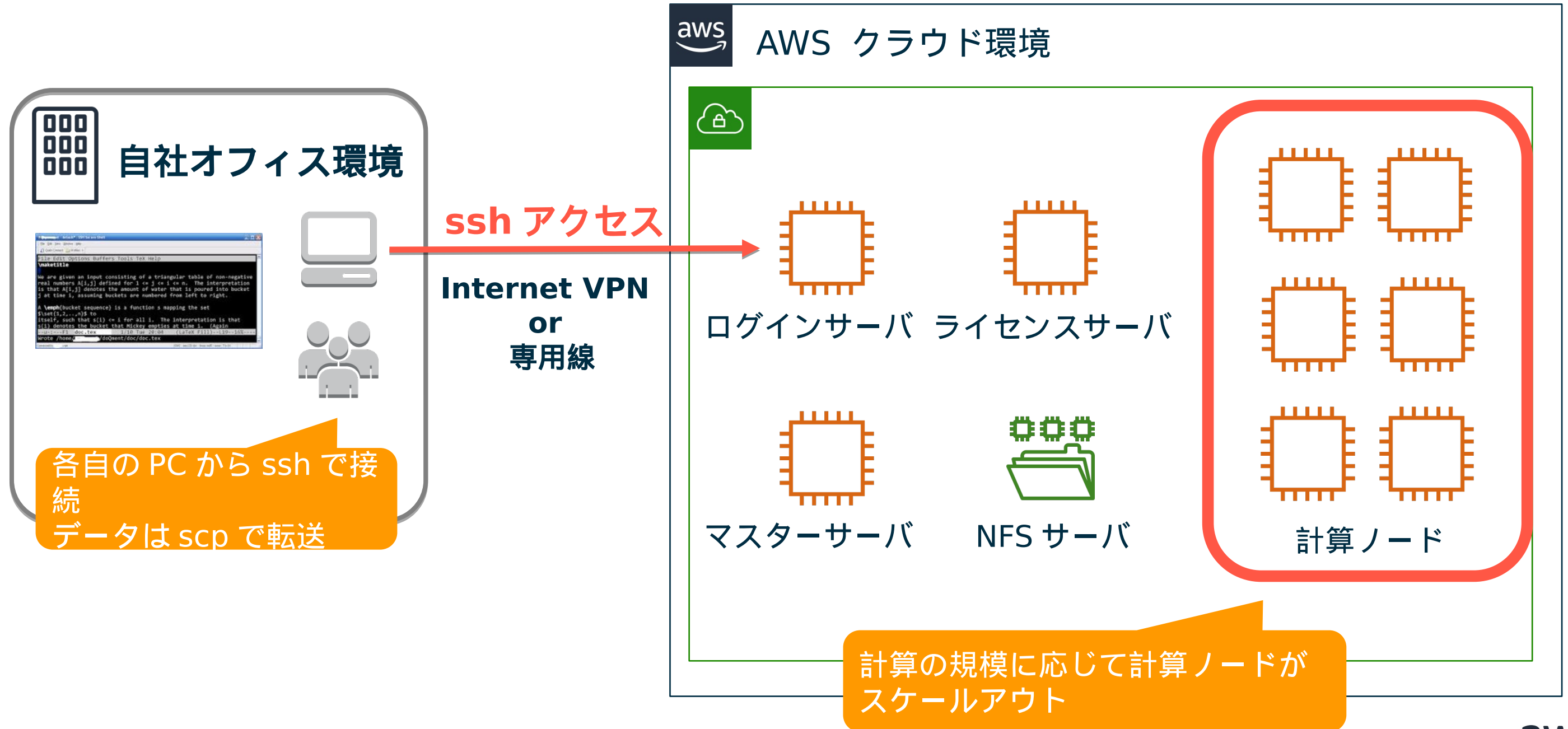
- HPC システム環境におけるコンピューターに投入されたジョブの起動や終了などの制御、監視を行うソフトウェア
- 代表的なソフトウェアとして Slurm がある



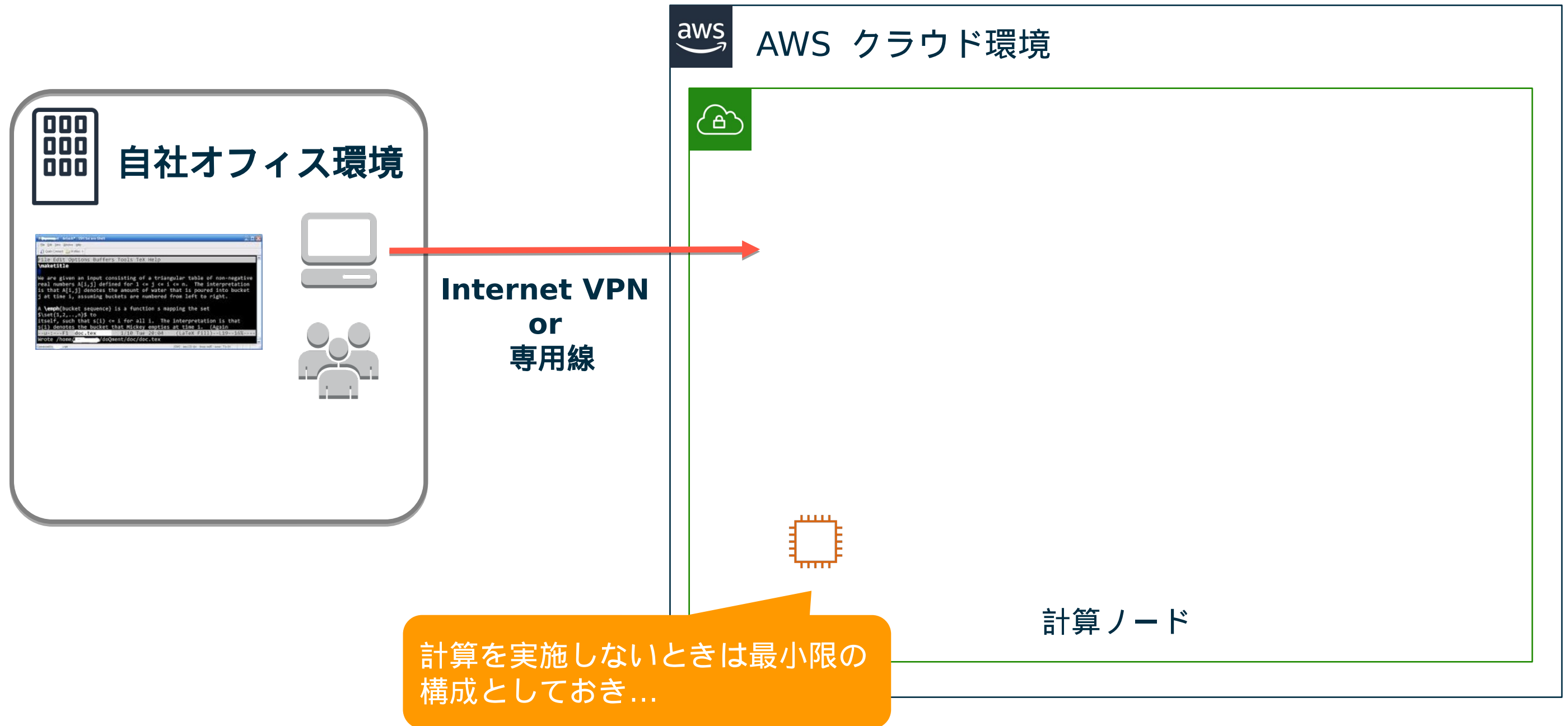
AWS を利用した基本的なシステム構成



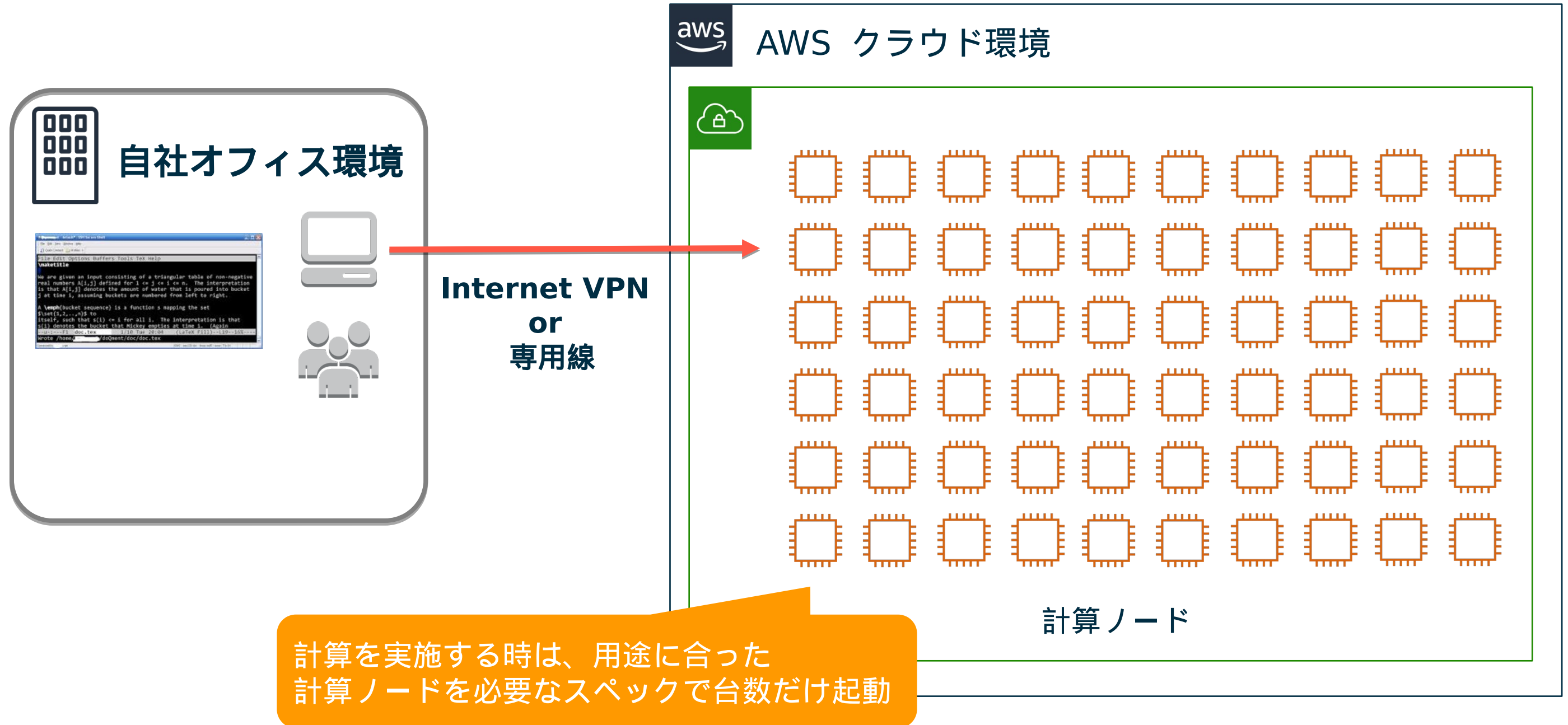
AWS を利用した基本的なシステム構成



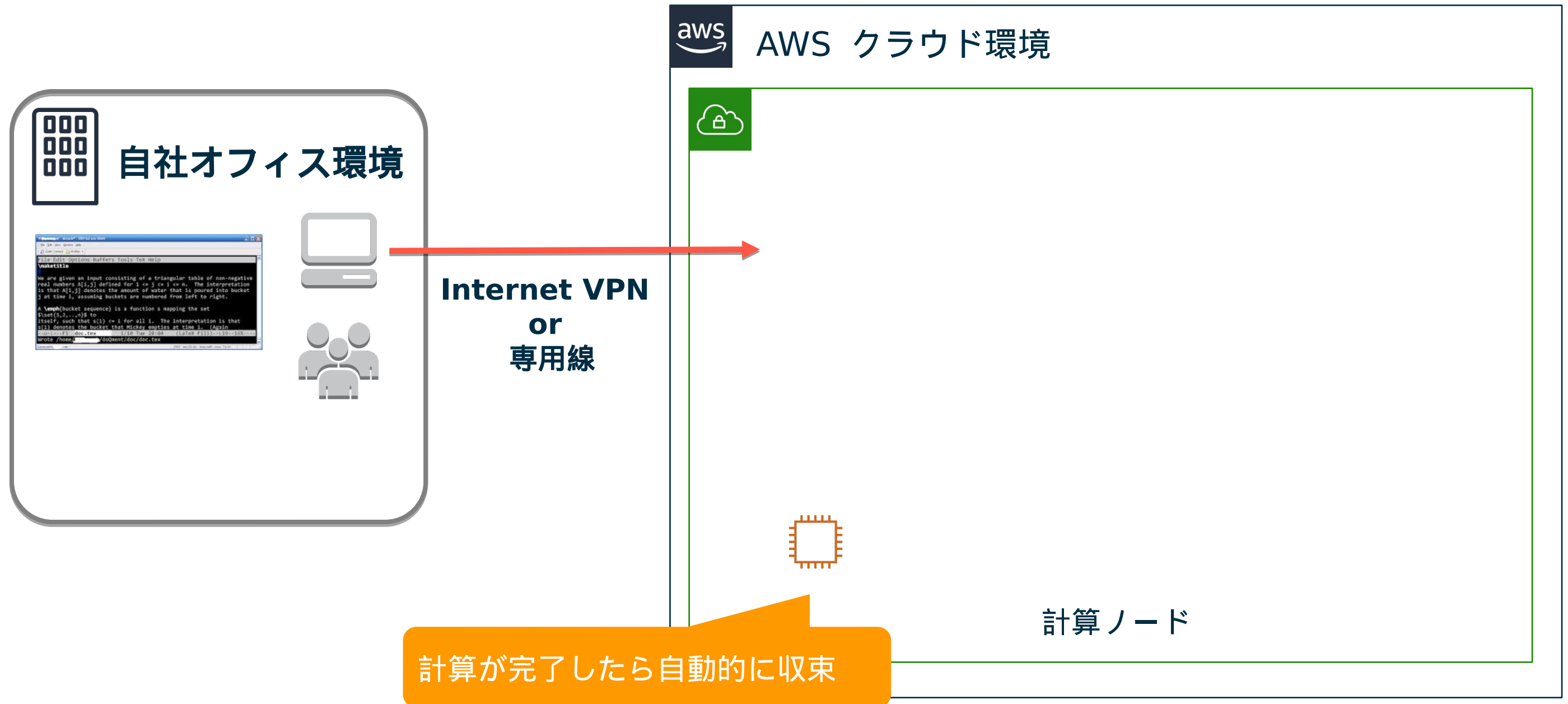
AWS を利用した基本的なシステム構成



AWS を利用した基本的なシステム構成



AWS を利用した基本的なシステム構成



AWS : 幅広いワークロードに対応する 200 以上のサービス

コンピューート



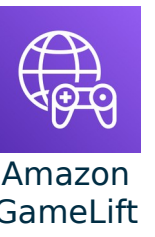
ネットワーク



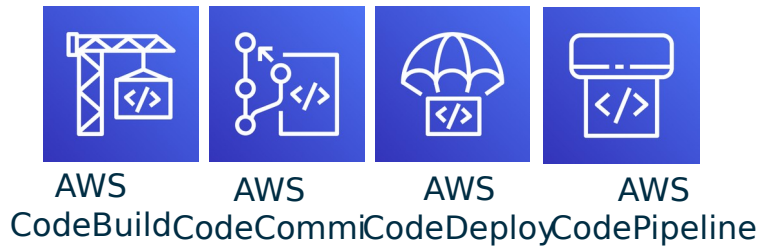
アナリティクス



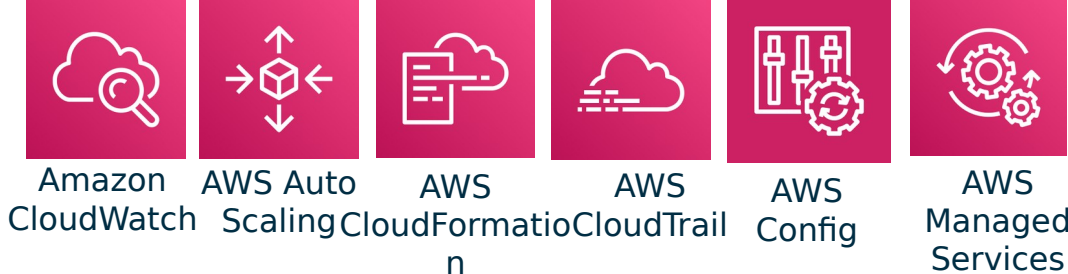
ゲーム



開発ツール



管理ツール



セキュリティ



ストレージ & 配信



アプリケーションインテグレーション



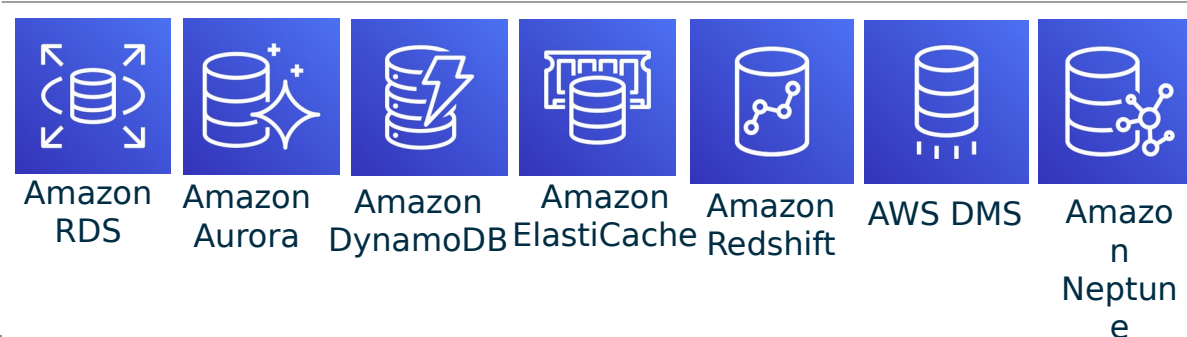
機械学習



モバイルサービス



データベース



IoT



※2022年10月時点

© 2022, Amazon Web Services, Inc. or its Affiliates.



AWS における HPC 関連サービス

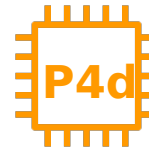
多様な HPC ワークロードに対応するための数多くのサービス

コンピューート

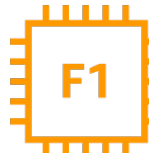
Amazon EC2



用途に応じて多様なインスタンスを利用可能な仮想サーバサービス



NVIDIA A100 GPU 搭載



Xilinx Virtex UltraScale+ 搭載



100 Gbps のネットワーク帯域

スポットインスタンスの活用で大幅なコスト減も可能

ストレージ

Amazon S3



高い耐久性と低コストでのデータ保管を実現するオブジェクトストレージ

Amazon FSx for Lustre



S3 連携可能な高速な分散ファイルシステムをフルマネージドで提供

AWS Snow ファミリー



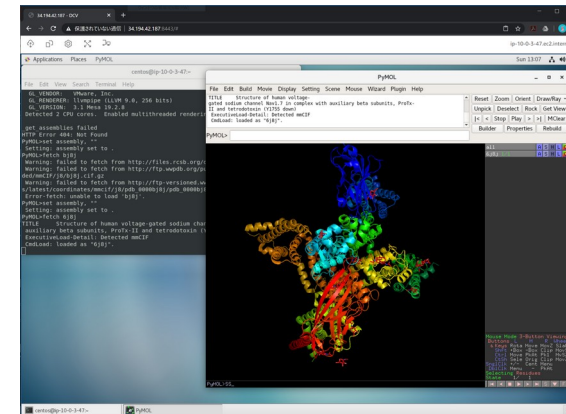
ストレージ内蔵の筐体を郵送することで S3 にデータを Import/Export

可視化

NICE-DCV



GPU アクセラレーションに対応し、インタラクティブなアプリケーションに適したデスクトップ仮想化



管理自動化

AWS ParallelCluster



AWS 上に HPC クラスタを自動で構築
オープンソースのジョブスケジューラ Slurm に対応しており
既存 HPC 環境からの移行が容易

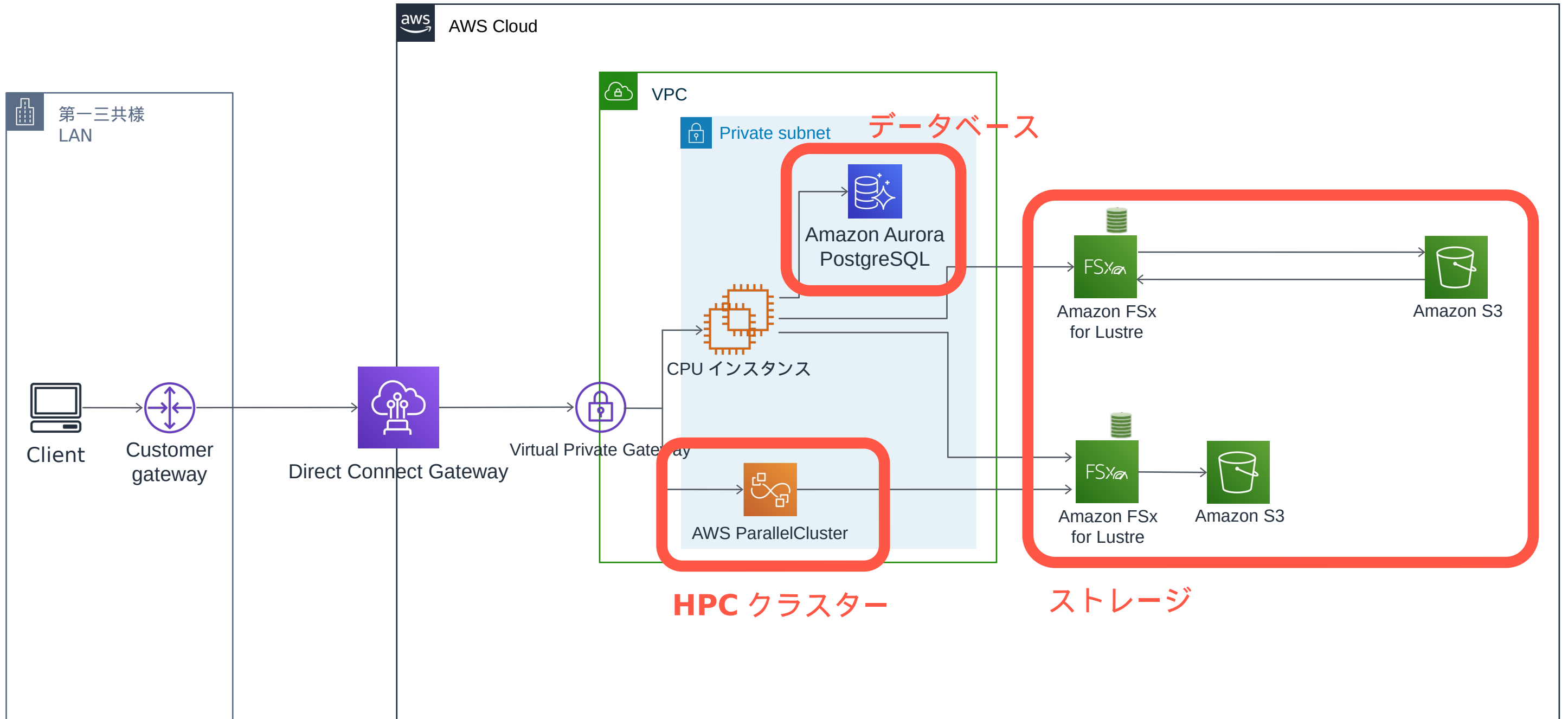
AWS Batch



コンテナベースの大規模バッチジョブコンピューティング環境をフルマネージドで提供

第一三共様の環境でご利用頂いている AWS サービス

第一三共様 データ駆動型創薬化学研究基盤 構成図

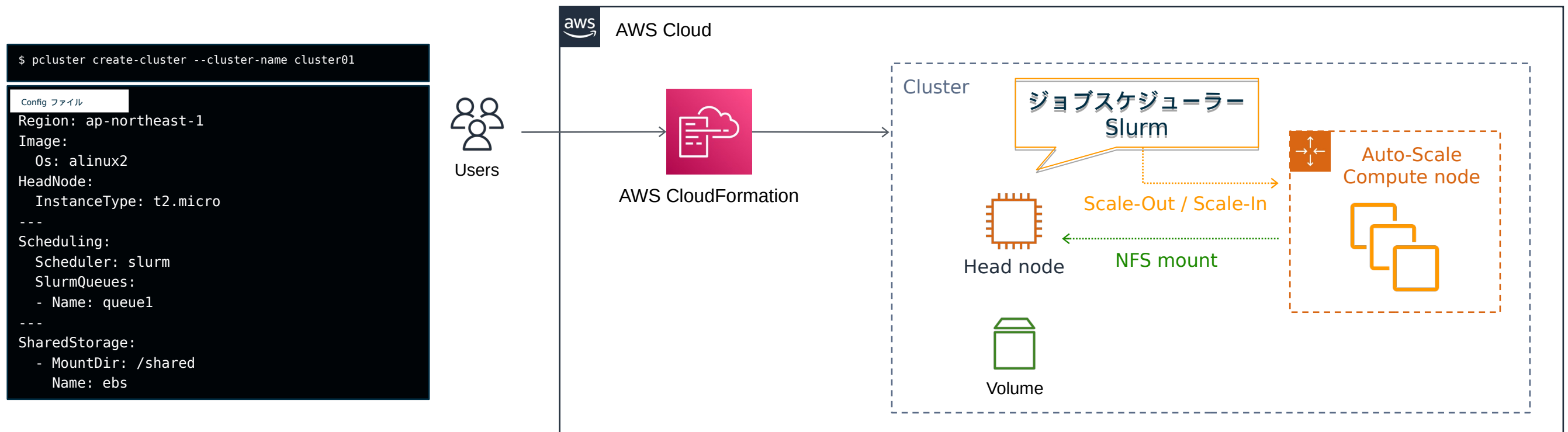


AWS ParallelCluster



ジョブ本数を監視して計算ノードを自動でスケールさせる クラスターを簡単に構成

- 数コマンド操作で自動でスケールするクラスタ環境をセットアップ
- AWS の様々なサービスと連携
- 研究者が自分専用のクラスタ環境を作成することも可能
- オープンソースのジョブスケジューラ Slurm を利用可能



<https://aws.amazon.com/jp/hpc/parallelcluster/>

© 2022, Amazon Web Services, Inc. or its Affiliates.



AWS ParallelCluster と統合された AWS サービス

NICE-DCV との統合によるリモート可視化

ブラウザからも利用可能なリモートデスクトップアプリケーションである NICE-DCV との統合により可視化などの処理が容易に

- ParallelCluster 2.5.0 以降、Amazon Linux 2, CentOS7, CentOS8, Ubuntu 18.04 に対応
- Config ファイルに記述することで設定可能

HeadNode:

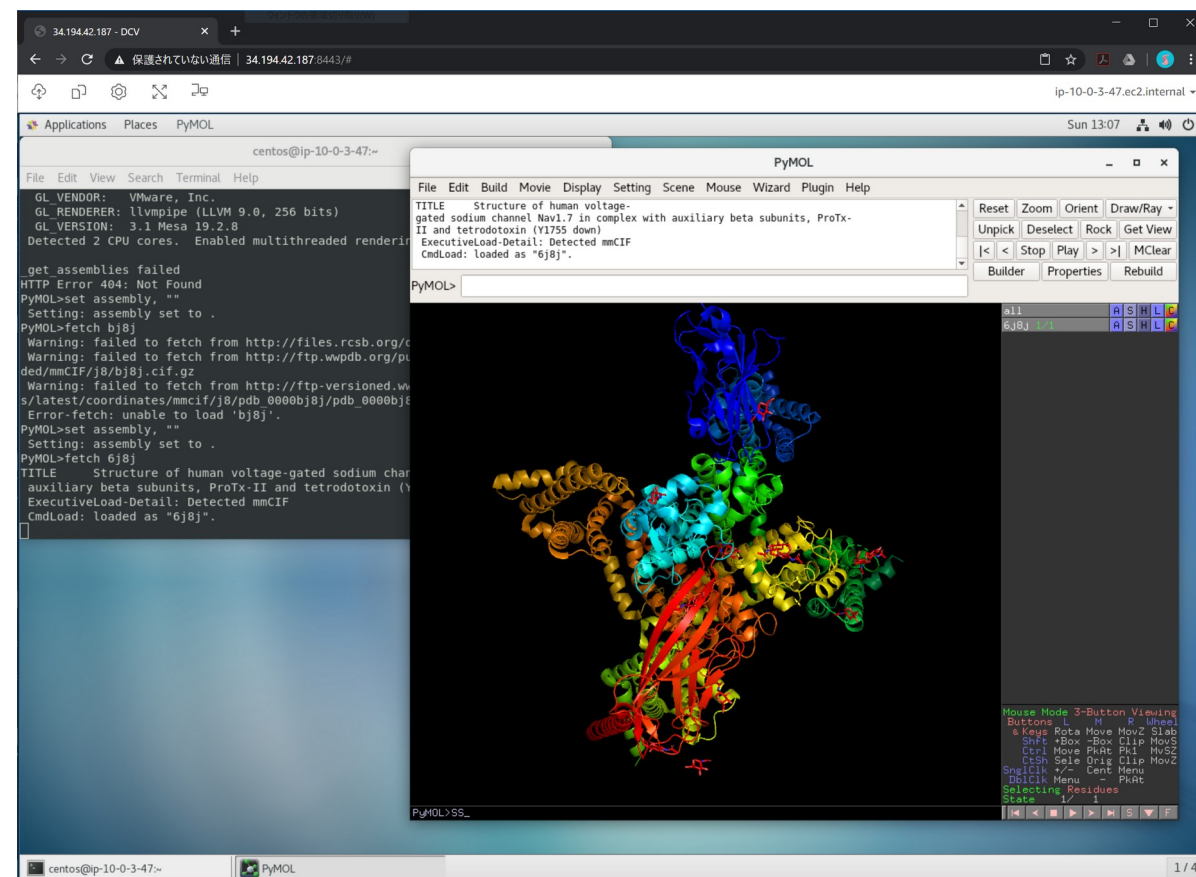
Dcv:

Enabled: true

AllowedIps: 172.31.0.0/16

- クラスタ起動後、dcv サブコマンドによりアクセス用 URL を生成

```
$ pcluster dcv-connect -n <CLUSTER-NAME> --key-path
```



<https://aws.amazon.com/jp/hpc/dcv/>

https://docs.aws.amazon.com/ja_jp/parallelcluster/latest/ug/pcluster.dcv-connect-v3.html

© 2022, Amazon Web Services, Inc. or its Affiliates.

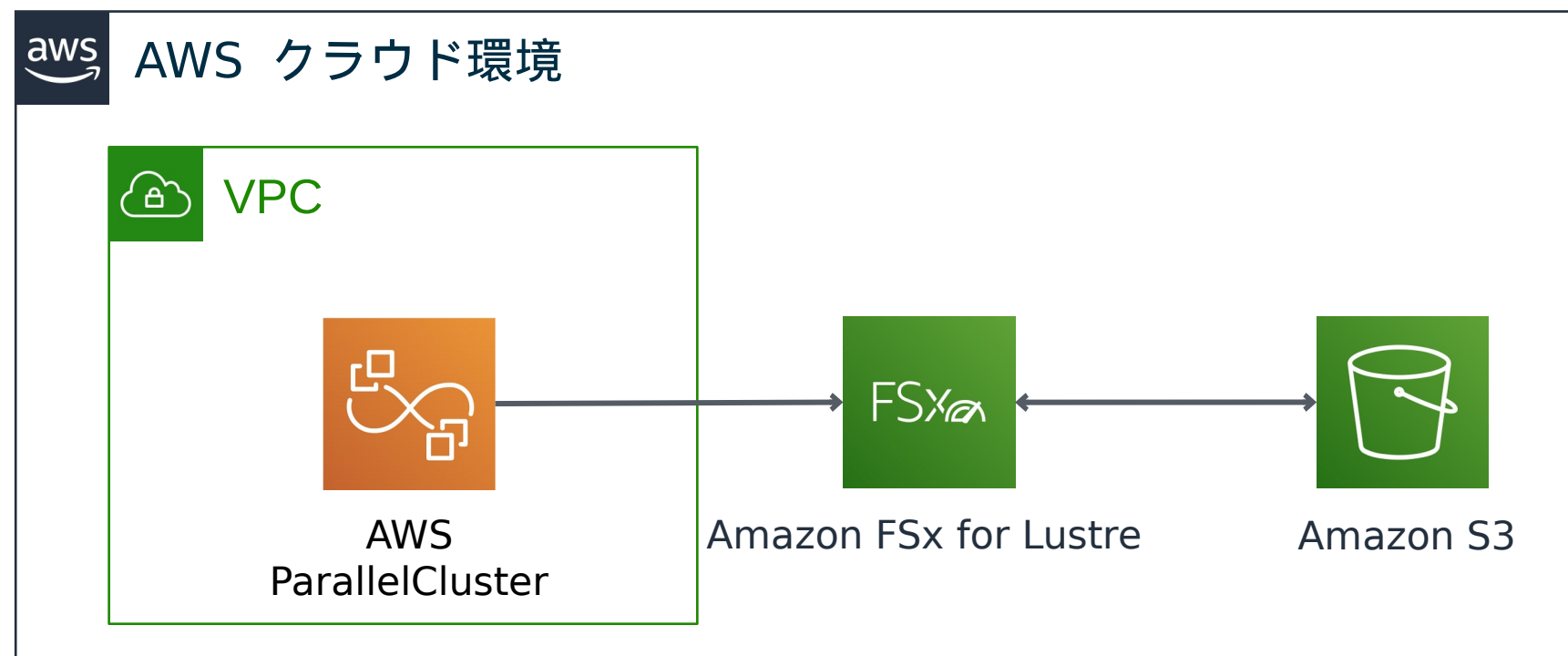


AWS ParallelCluster と統合された AWS サービス

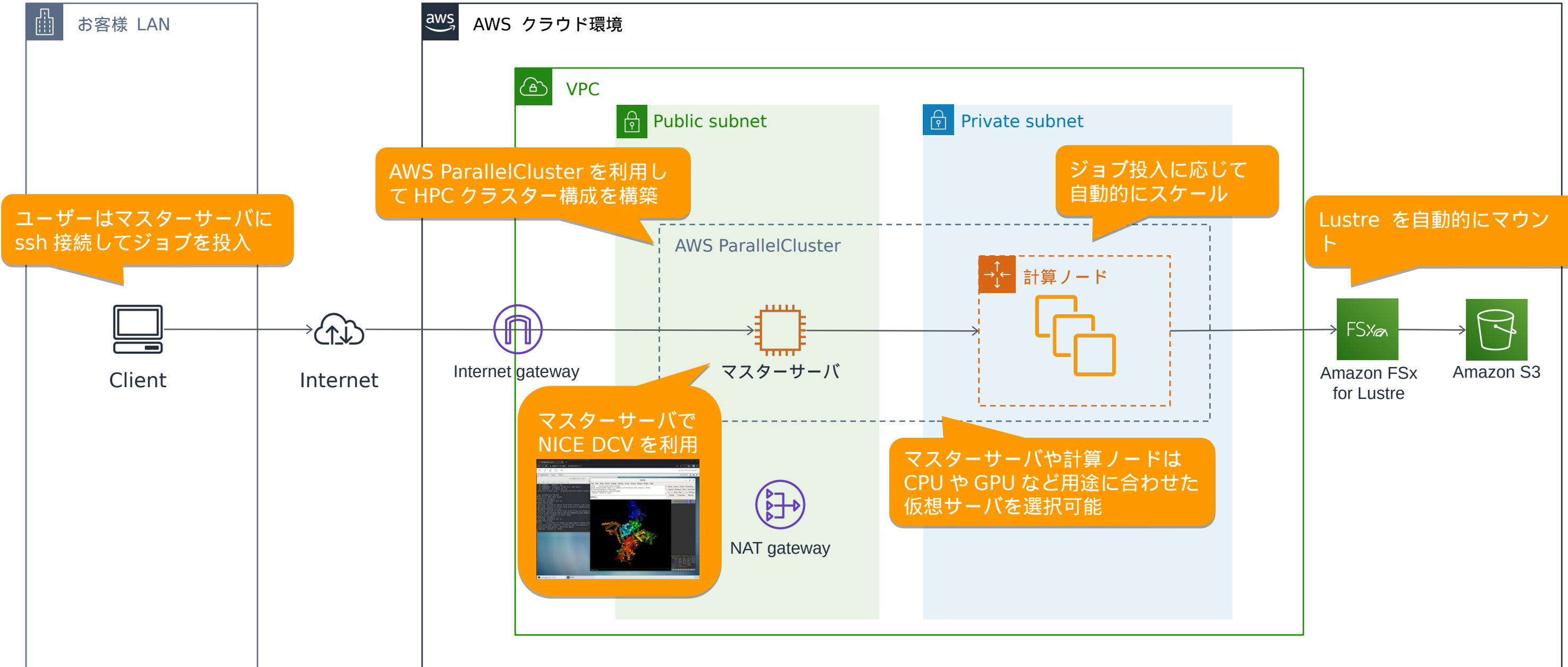
Amazon FSx for Lustre との統合による高速な外部ストレージ環境との接続

- Lustre は多くのスーパーコンピュータで利用される高性能な分散ファイルシステム
 - 15年以上の実績があり、Top100 スーパーコンピュータの60%が使用
- Amazon FSx for Lustre は Lustre の構築や運用を AWS が管理するフルマネージドサービス
- Amazon S3 と連携可能
- Config ファイルに記述することで設定可能

```
SharedStorage:  
- MountDir: /mnt/fsx  
  Name: fsx  
  StorageType: FsxLustre  
  FsxLustreSettings:  
    FileSystemId : <FS-ID>
```

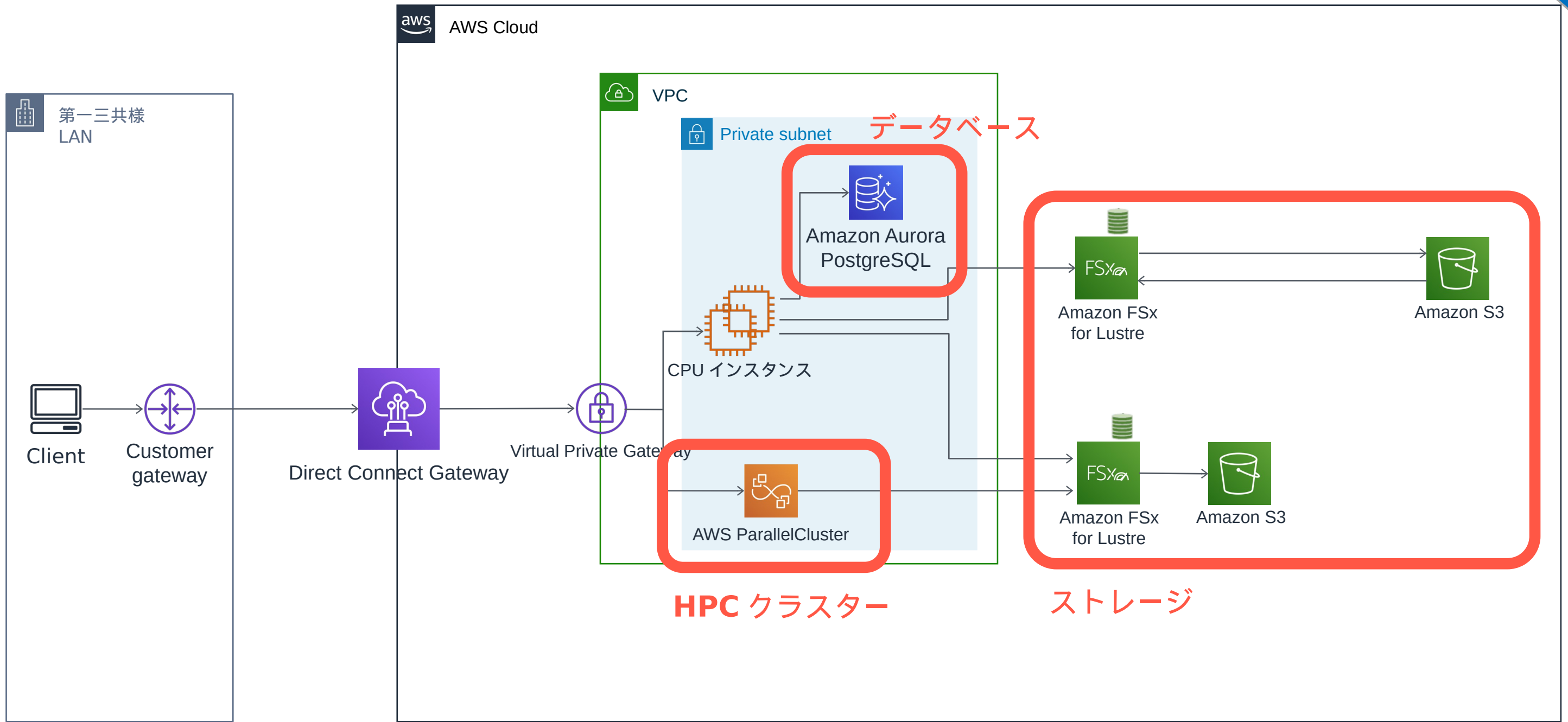


AWS ParallelCluster を利用した一般的な HPC 構成例



第一三共様 データ駆動型創薬化学研究基盤 構成図

再掲



Amazon Relational Database Service (RDS)

人気のある 6 つのデータベースエンジンの選択による管理された
リレーショナルデータベースサービス

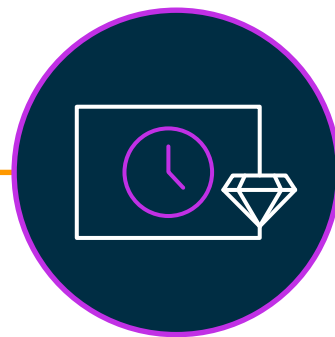
Amazon
Aurora



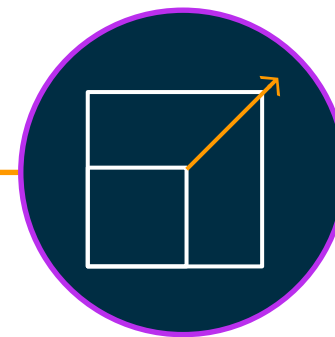
ORACLE®



インフラストラクチャの
プロビジョニングやデータ
ベースのインストール
メンテナンスは不要



マルチ AZ データレプリケー
ション、自動バックアップ、
スナップショット、自動フェ
イルオーバー



数クリックでデータベースの
コンピューとストレージを
拡張可能
アプリケーションの
ダウンタイムは最小限



SSD ストレージと保証された
プロビジョンド I/O
保存時や転送中のデータ暗号化

Amazon Aurora PostgreSQL の RDKit Extension サポート

コンピュータで化合物情報を扱うケモインフォマティクス分野で用いられる代表的なオープンソースのライブラリ

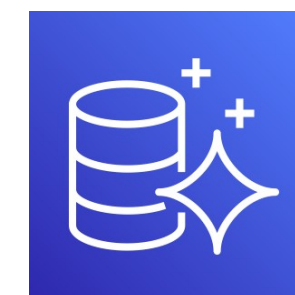


Open-Source Cheminformatics
and Machine Learning

現在 RDKit version 3.8 をサポート

```
database=> create extension rdkit;
CREATE EXTENSION
database=> \dx

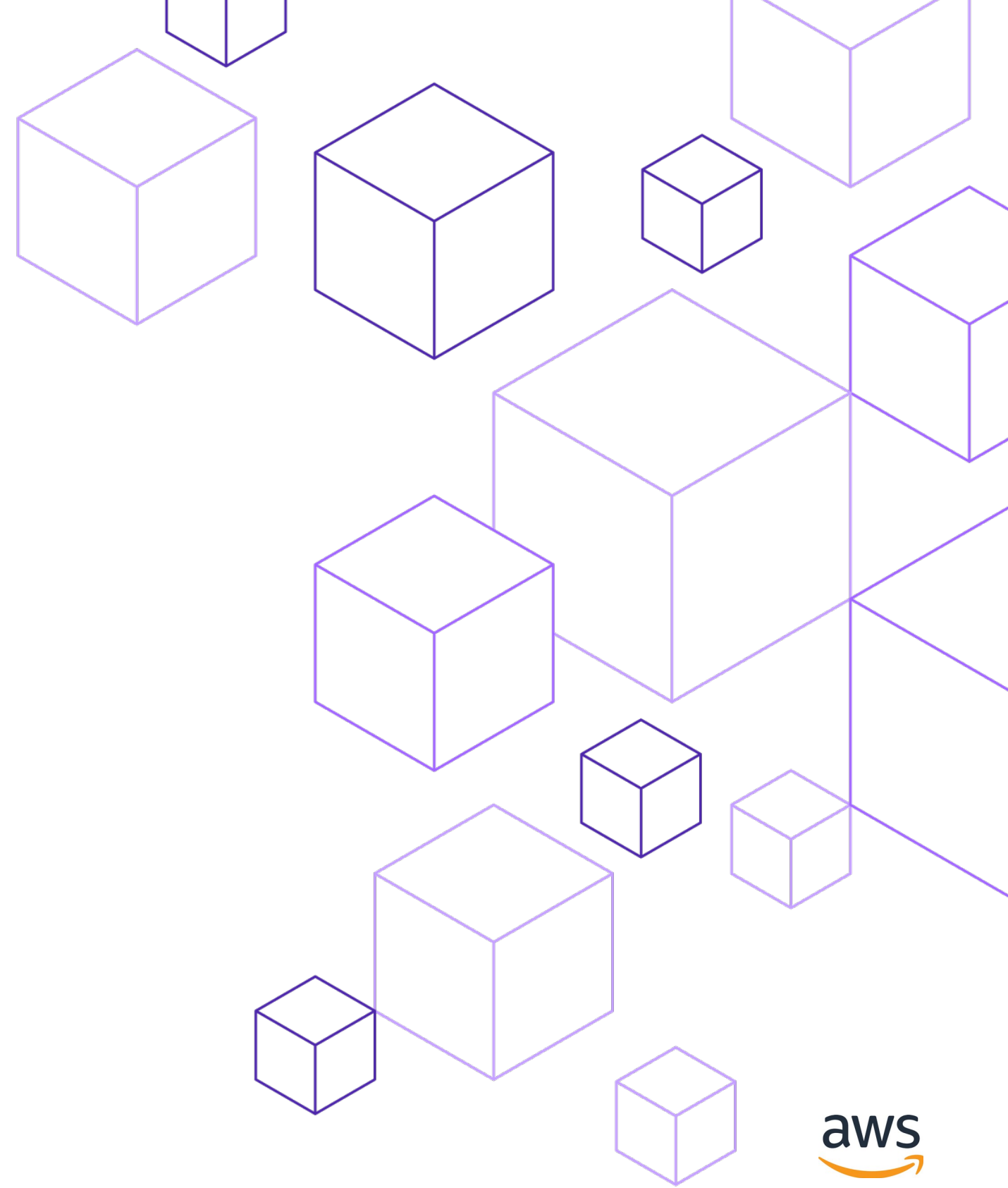
                List of installed extensions
  Name  | Version | Schema  | Description
-----+-----+-----+-----
plpgsql | 1.0     | pg_catalog | PL/pgSQL procedural language
rdkit   | 3.8     | public   | Cheminformatics functionality for
```



Amazon Aurora
PostgreSQL

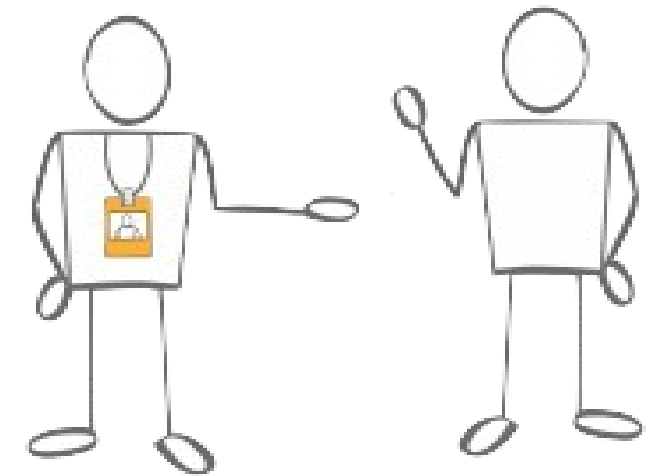
<https://aws.amazon.com/jp/about-aws/whats-new/2020/09/amazon-aurora-postgresql-supports-rdkit-extension/>

AWS からの支援内容



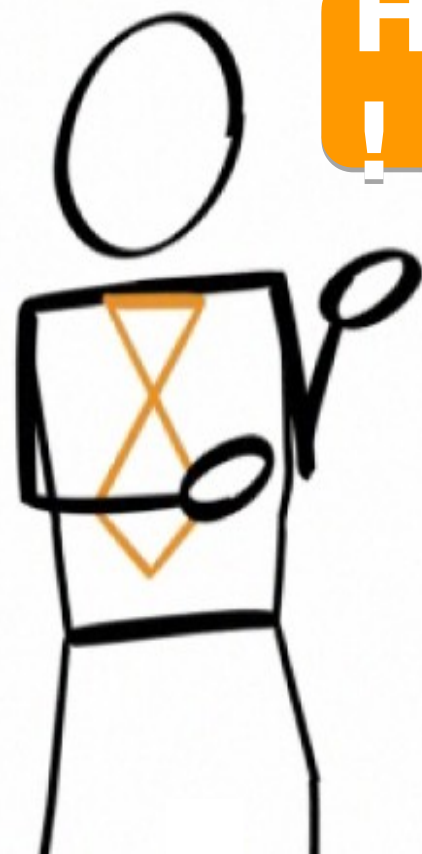
AWS からの支援内容

- AWS サービスのご紹介とハンズオン実施
- アーキテクチャ支援
- PoC 環境の検討
- クラスタ運用に関する検討支援
- AWS サービスにおける機能要望のフィードバック対応

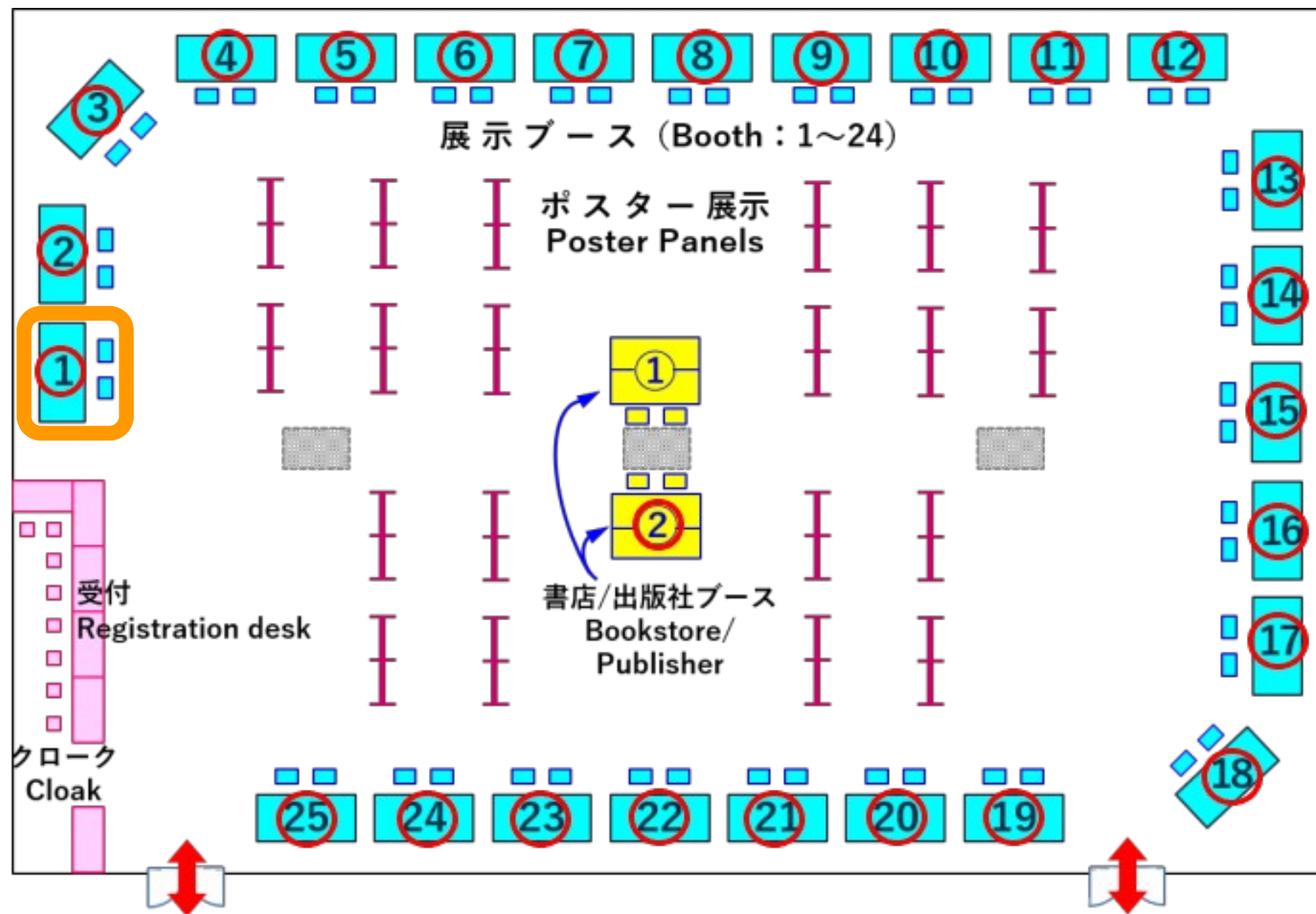


AWS を利用した創薬研究環境の構築をご検討頂く際はご相談頂けますと幸いです

AWS ブースございます！



Here!



<https://cbi-society.org/taikai/taikai22/exhibition.html#layout>

Thank you!
ご清聴ありがとうございました