



創薬研究領域における AWSのHPCサービス活用について

Amazon Web Services Japan

中島 文博

2021/11/18

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

このセッションの目的と位置づけ

本イベントの全体アジェンダ

10:00 – 10:05 オープニング

10:05 – 10:35 「創薬研究領域におけるAWS最新事例のご紹介」

アマゾン ウェブ サービス ジャパン合同会社 ヘルスケア&ライフサイエンス事業開発部 マネージャー
片岡 勇人

10:35 – 11:05 「創薬研究領域におけるユースケースと関連するAWSソリューションのご紹介」

アマゾン ウェブ サービス ジャパン合同会社 技術統括本部
ソリューションアーキテクト
津郷 光明

11:05 – 11:20 「創薬研究領域におけるAWSのHPCサービス活用について」

アマゾン ウェブ サービス ジャパン合同会社 技術統括本部
ソリューションアーキテクト
中島 丈博

11:20 – 11:50 「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～」

アステラス製薬株式会社 開発研究部門 モダリティ研究所 ストラテジックスクリーニングサイエンス研究室
崎山 則征 様

本セッションではアステラス様でご利用頂いているAWS ParallelClusterと呼ばれるHPCクラスター環境についてご紹介いたします

AWS における HPC 関連サービス

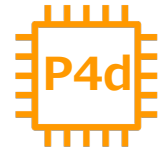
多様な HPC ワークロードに対応するための数多くのサービス

コンピューート

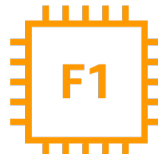
Amazon EC2



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

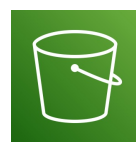


100 Gbps のネットワーク帯域

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



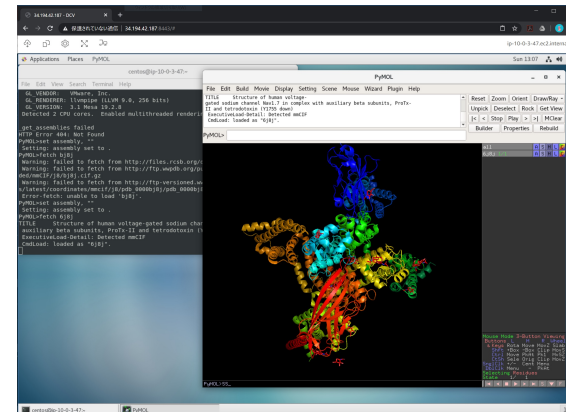
ストレージ内蔵の筐体を郵送することで S3 にデータを Import/Export

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
Slurm / SGE / Torque といったジョブスケジューラ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

AWS における HPC 関連サービス

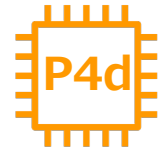
多様な HPC ワークロードに対応するための数多くのサービス

コンピューート

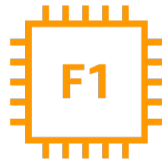
Amazon EC2



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



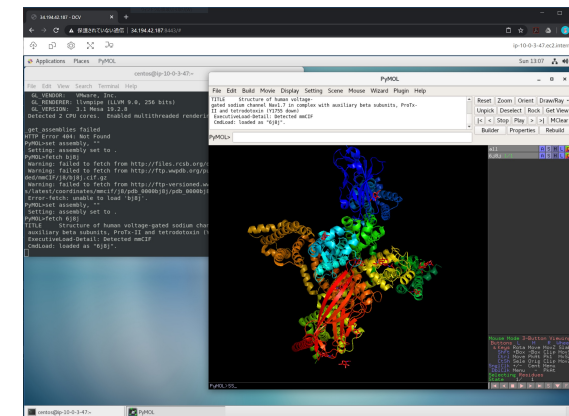
ストレージ内蔵の筐体を郵送することで S3 にデータを Import/Export

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



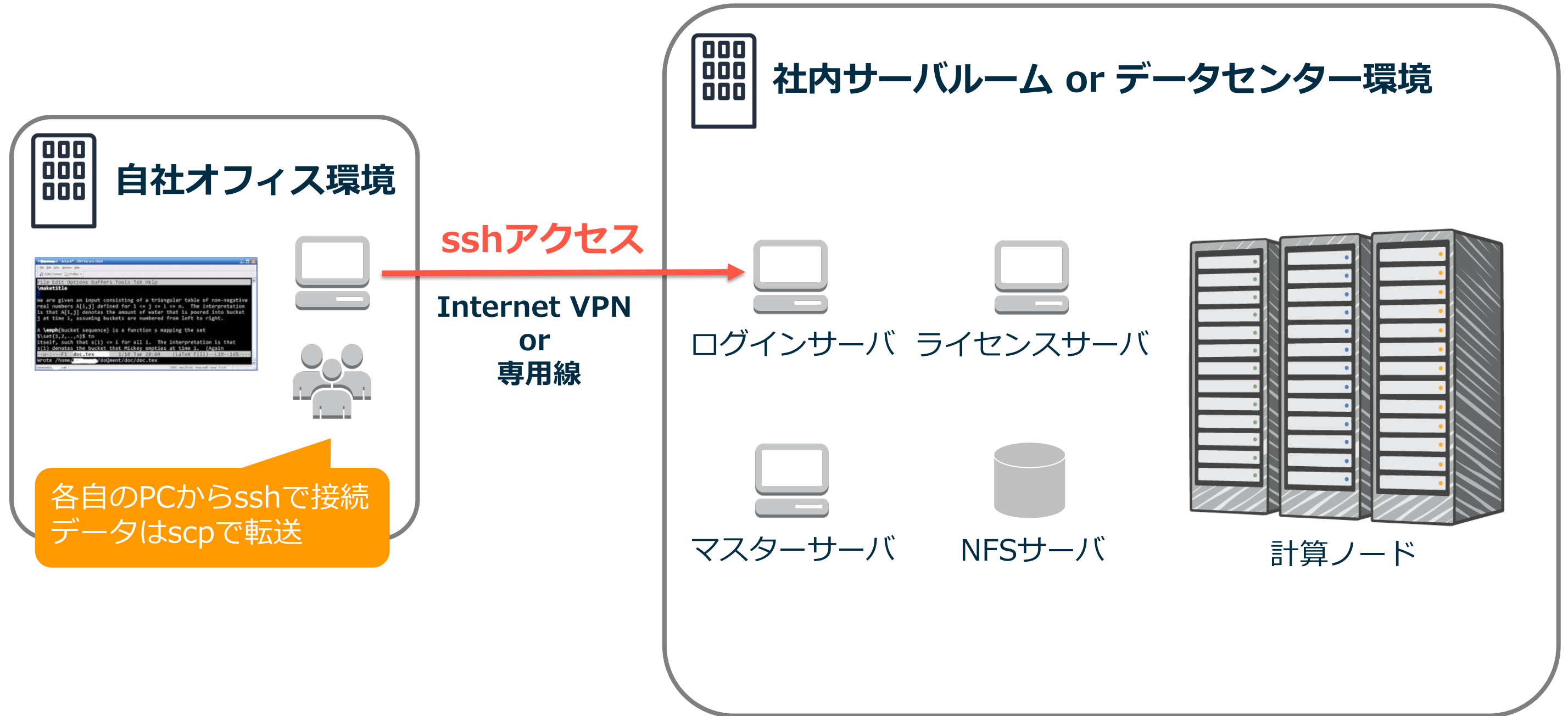
AWS上に HPC クラスタを自動で構築
Slurm / SGE / Torque といったジョブスケジューラに対応しており既存HPC環境からの移行が容易

AWS Batch

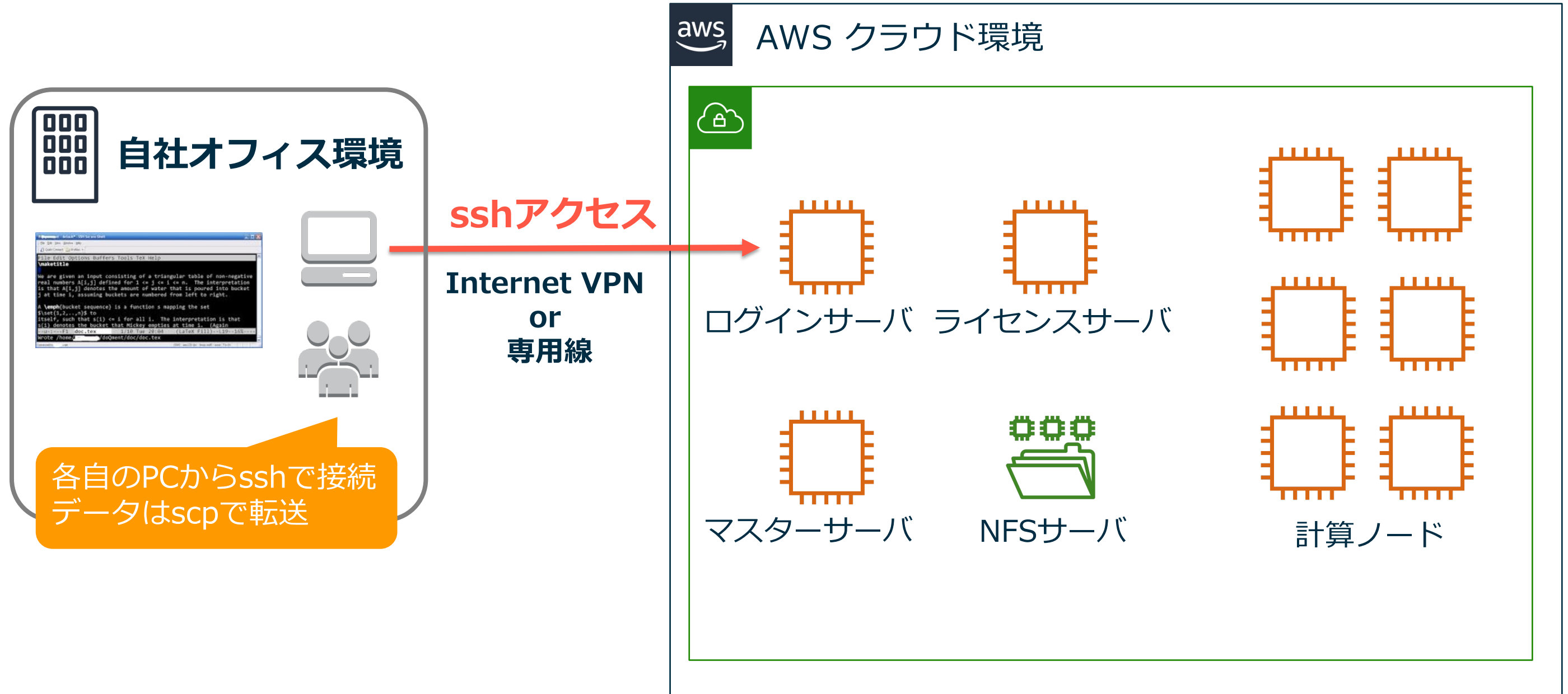


コンテナベースの大規模バッチジョブコンピューティング環境をフルマネージドで提供

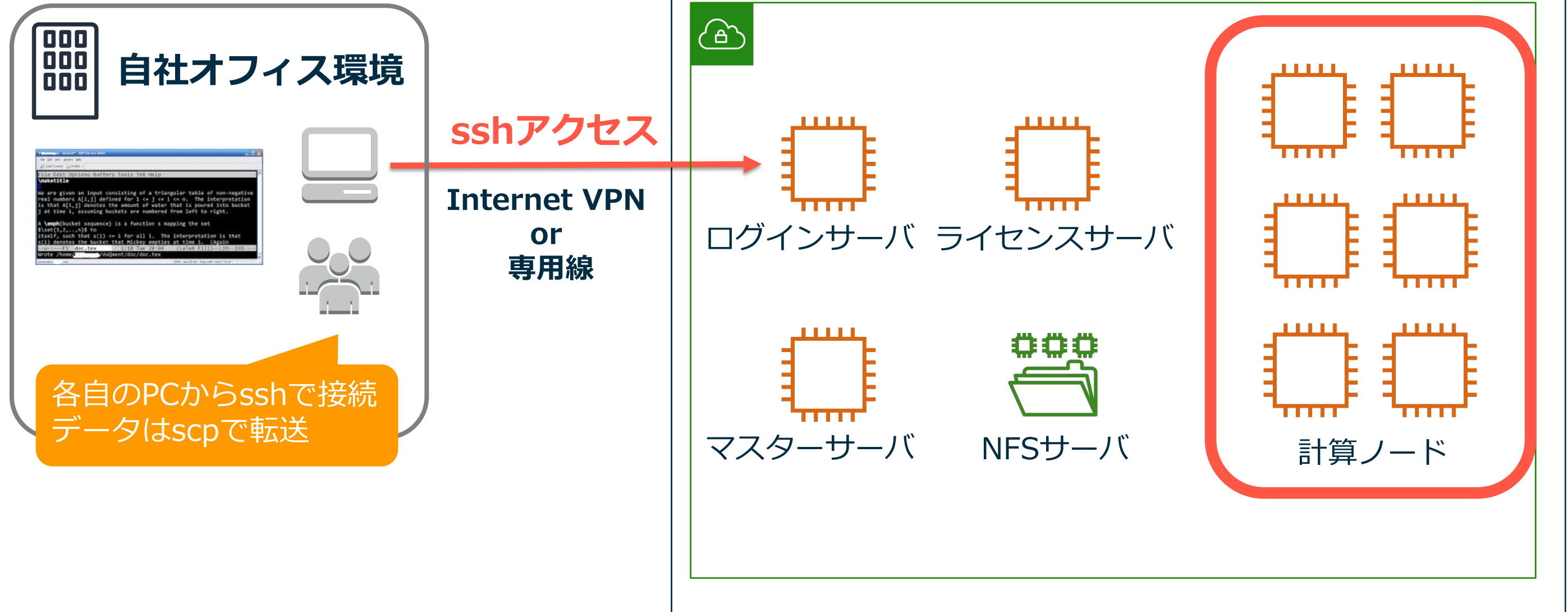
一般的なHPC環境



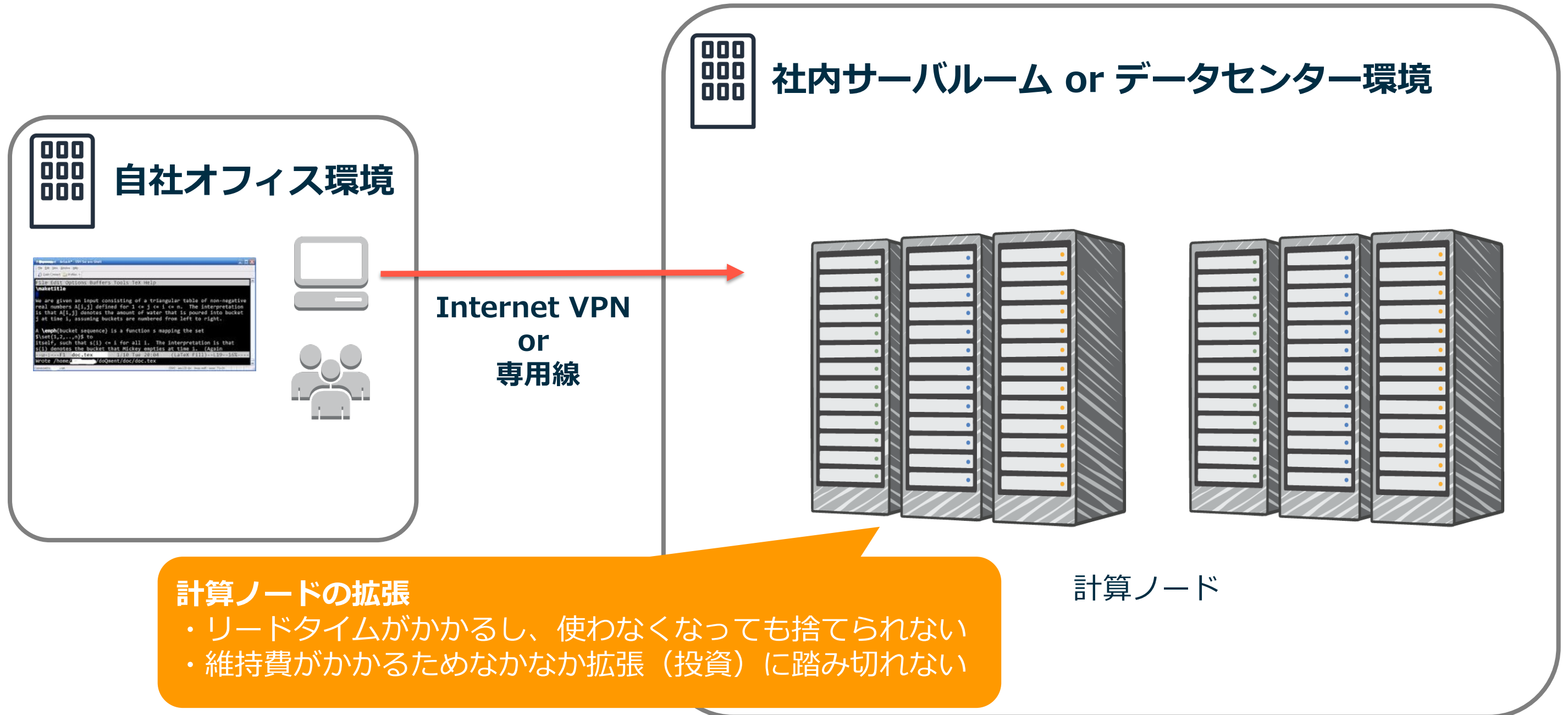
AWSを利用した基本的なシステム構成



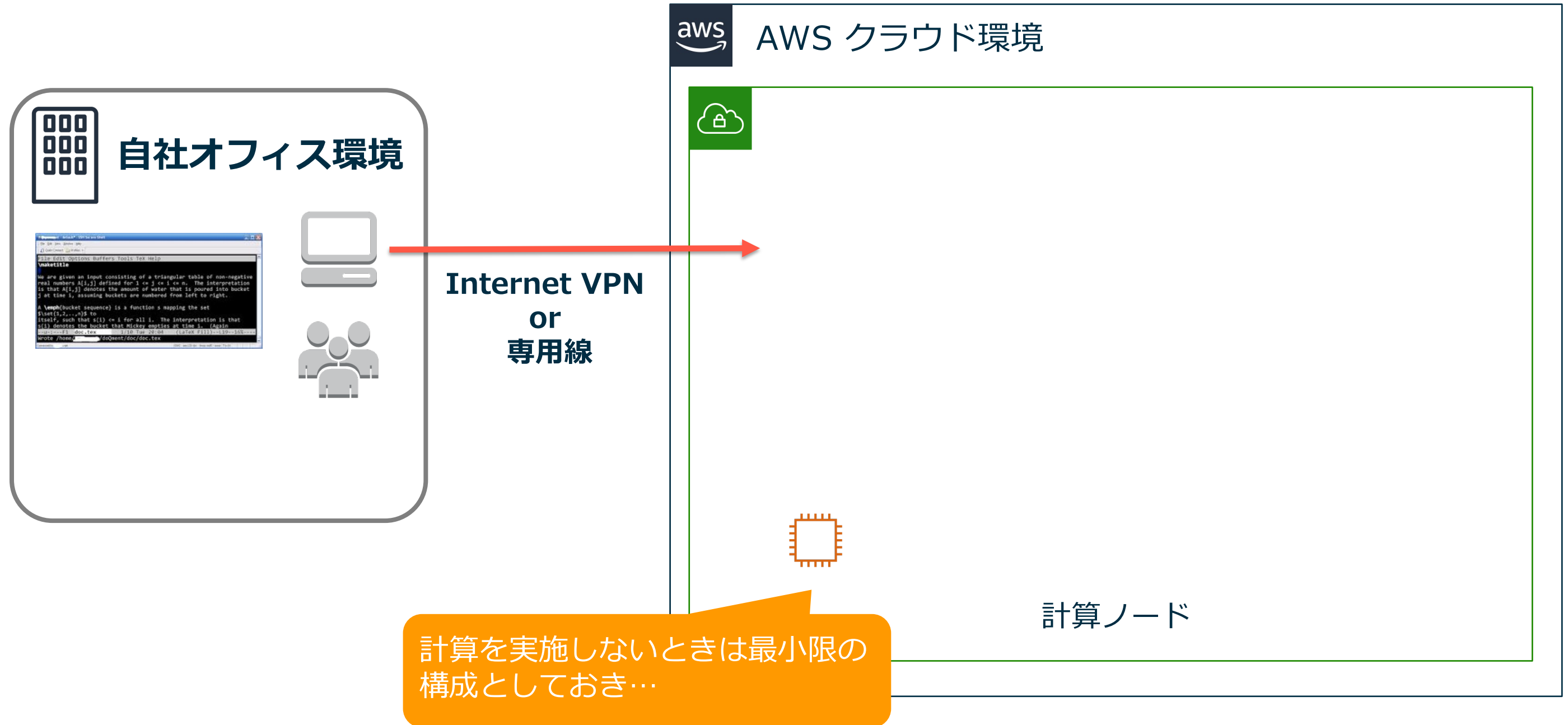
AWSを利用した基本的なシステム構成



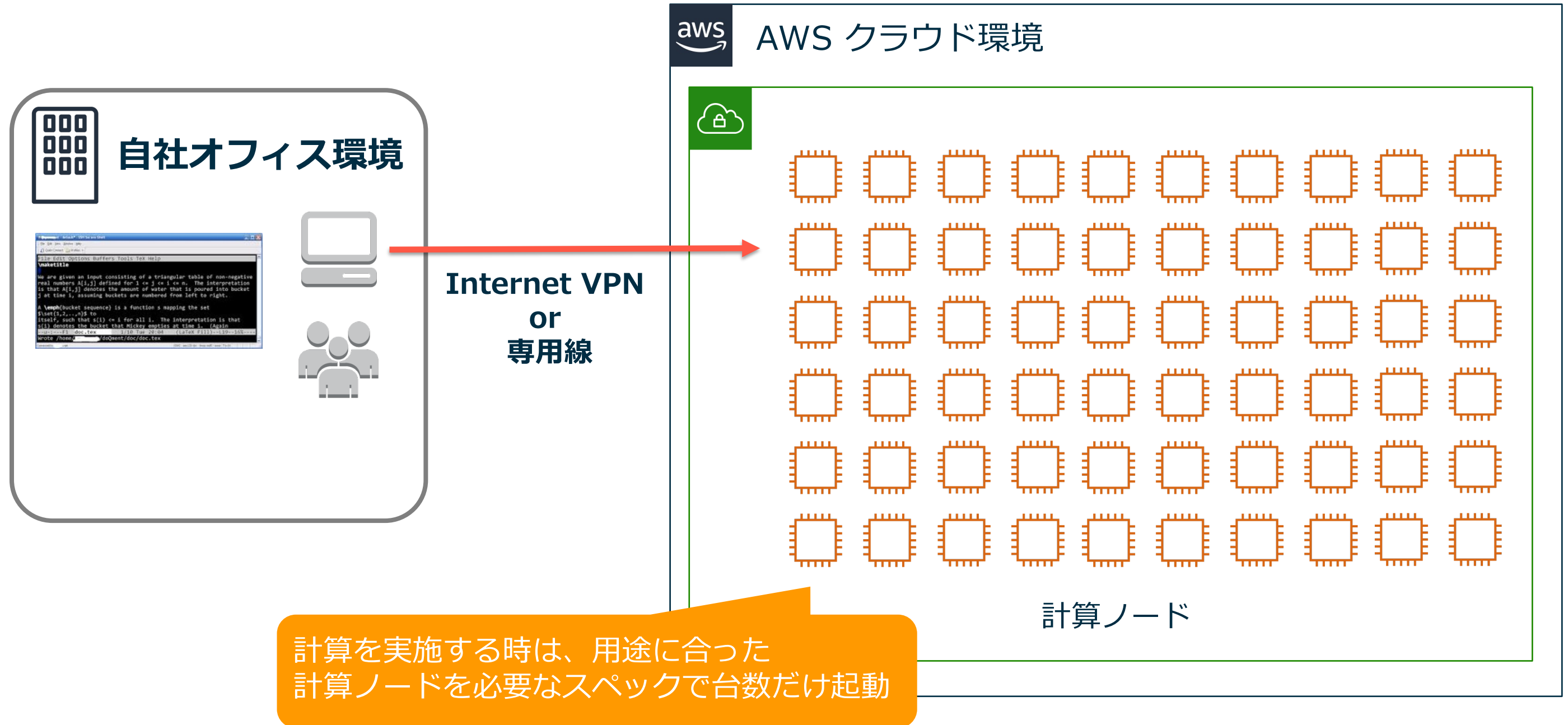
一般的なHPC環境の計算ノード



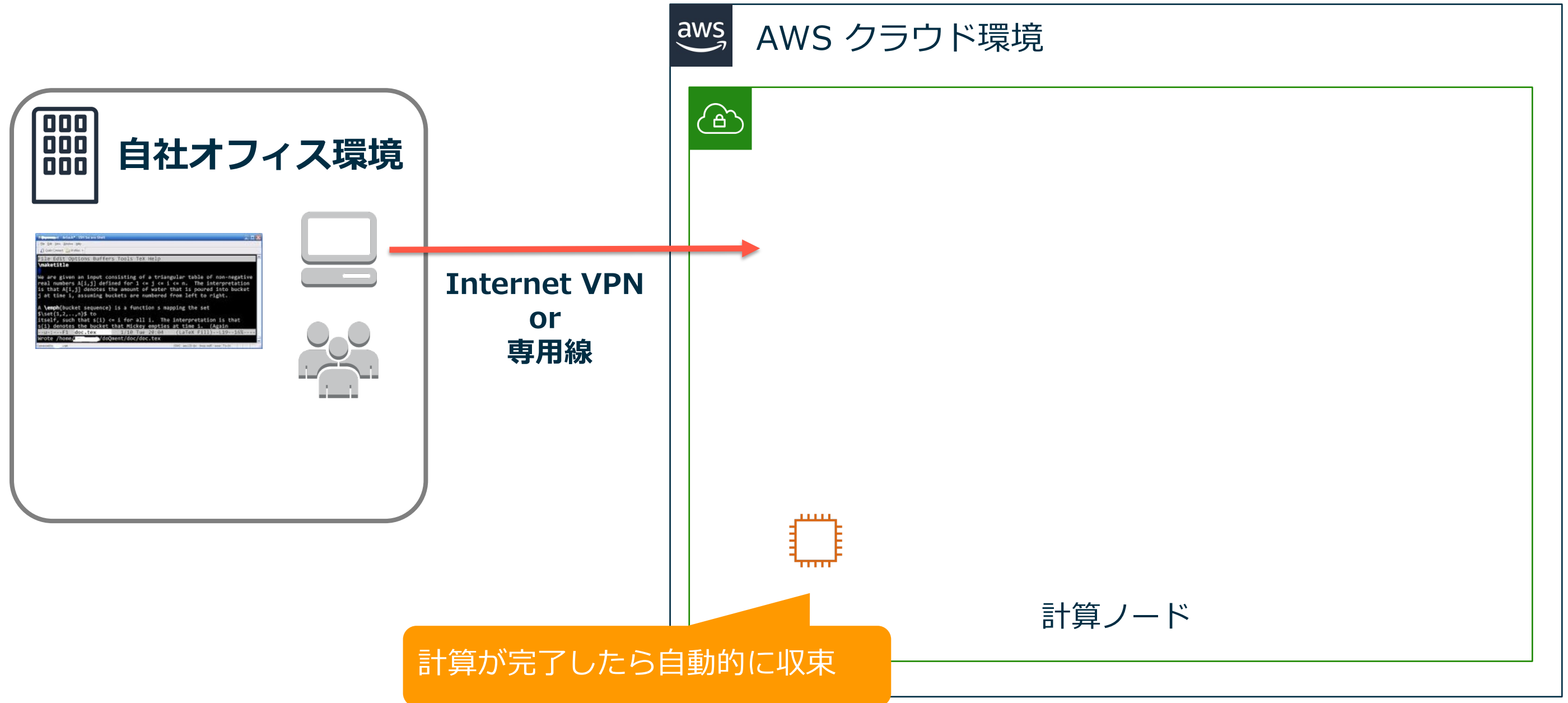
AWSを利用した基本的なシステム構成



AWSを利用した基本的なシステム構成

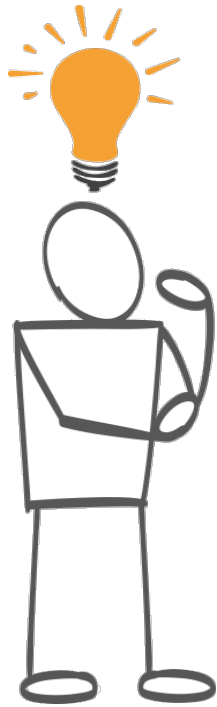


AWSを利用した基本的なシステム構成



拡張性の高いHPCのクラスター構築する際の悩み

- HPC アプリケーションの知識に加えて **AWS** や**インフラの知識**も必要
- 実際に**自動でスケールする**クラスター環境を作成するのは大変
- EC2 以外にもストレージなど、**AWS** の他のサービスも利用したい



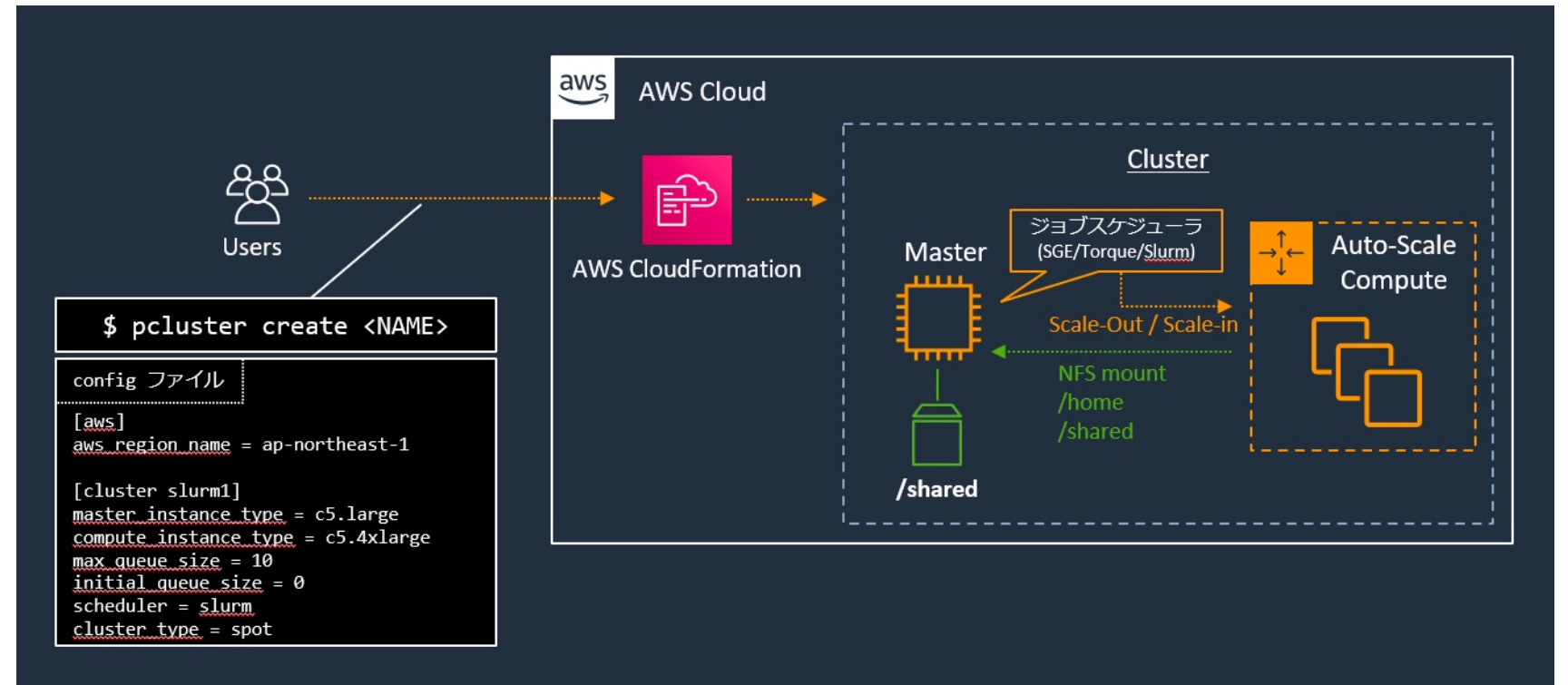
AWS ParallelCluster で解決！

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



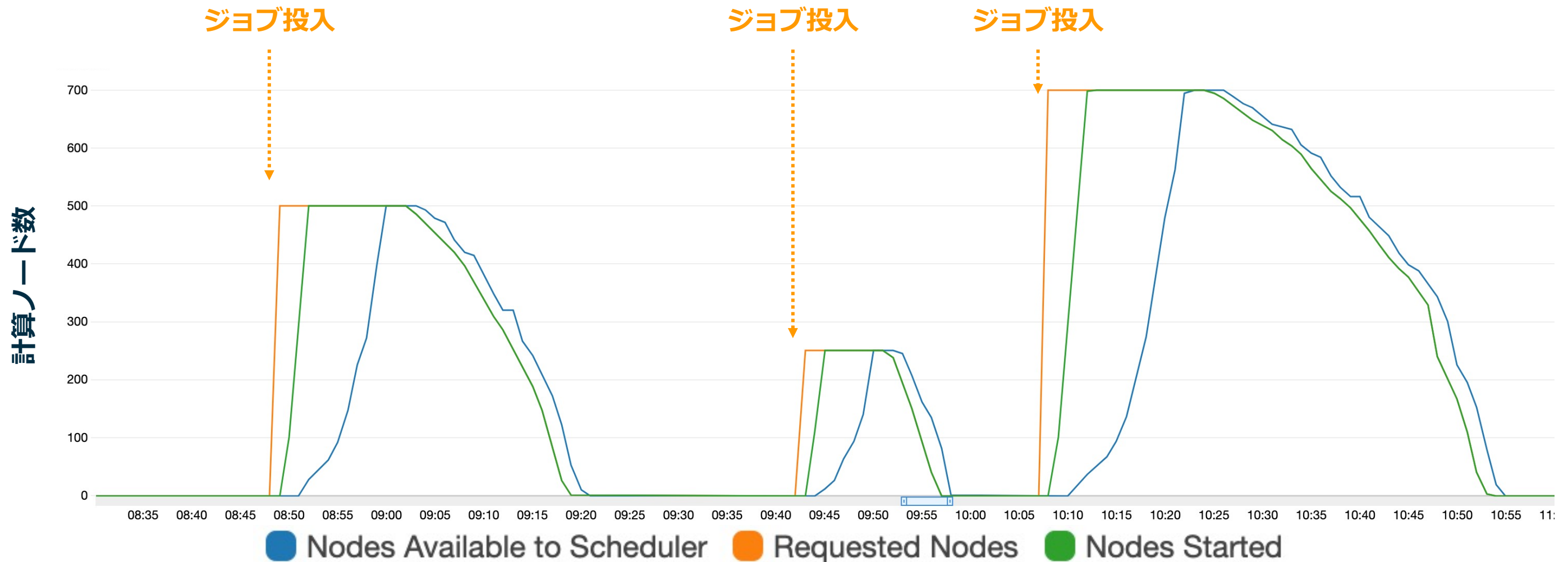
- 数コマンド操作で自動でスケールするクラスタ環境をセットアップ
- AWS の様々なサービスと連携
- 研究者が自分専用のクラスタ環境を作成することも可能
- オープンソースのジョブスケジューラを選択可能
 - SGE
 - Slurm
 - Torque



AWS ParallelClusterの紹介と簡単な使い方の例が下記のBlogで紹介されています
<https://aws.amazon.com/jp/blogs/opensource/aws-parallelcluster/>

AWS ParallelCluster によるスケーリング (参考例)

AWS ParallelCluster を利用することにより計算ノードが 0 台から数百台規模まで需要に応じてスケーリングするクラスタ環境を作成することが可能



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

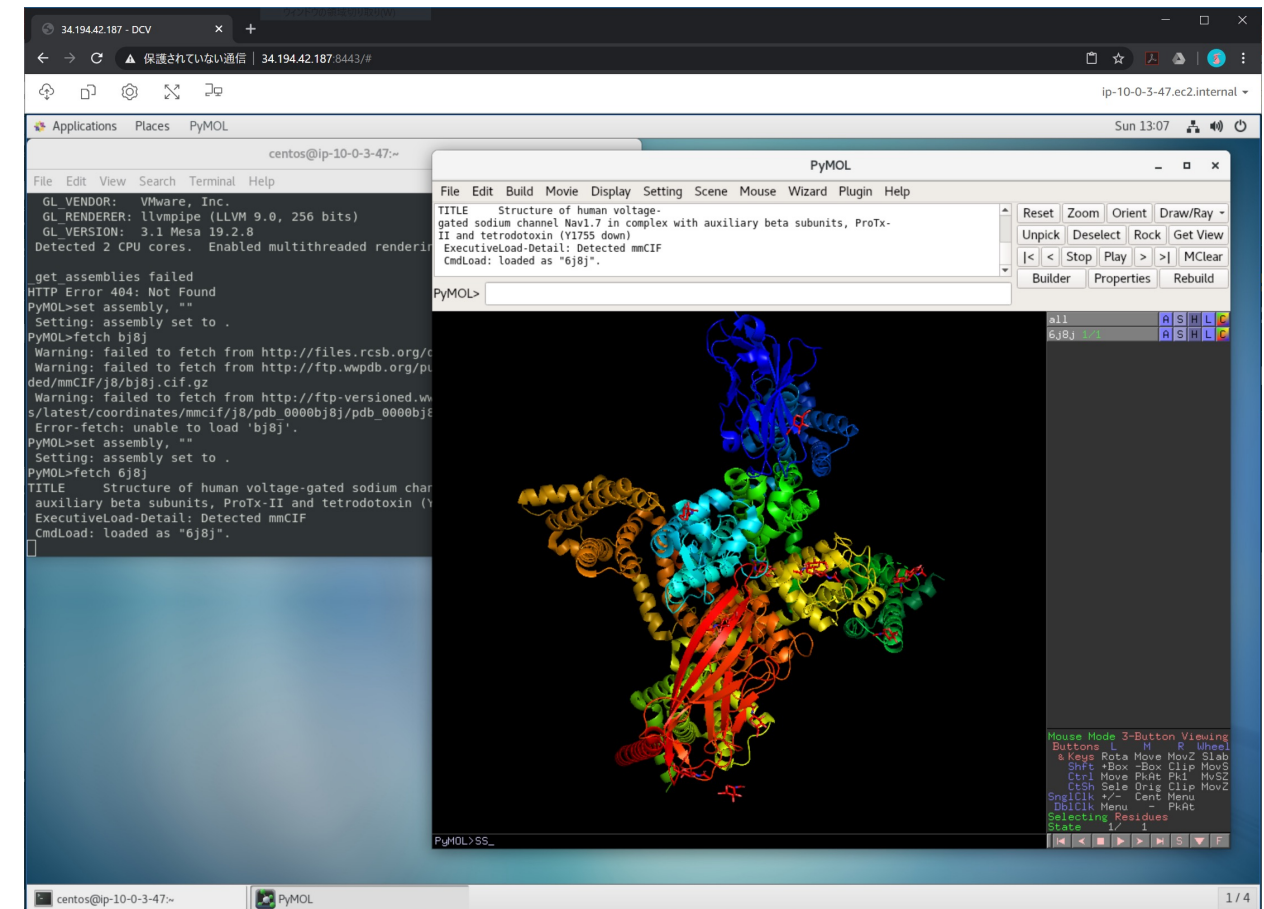
- ParallelCluster 2.5.0 以降、Amazon Linux 2, CentOS7, CentOS8, Ubuntu 18.04 に対応
- Config ファイルに記述することで設定可能

```
[cluster default]
dcv_settings = dcv1

[dcv dcv1]
enable = master
```

- 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 <CLUSTER_NAME>
```



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

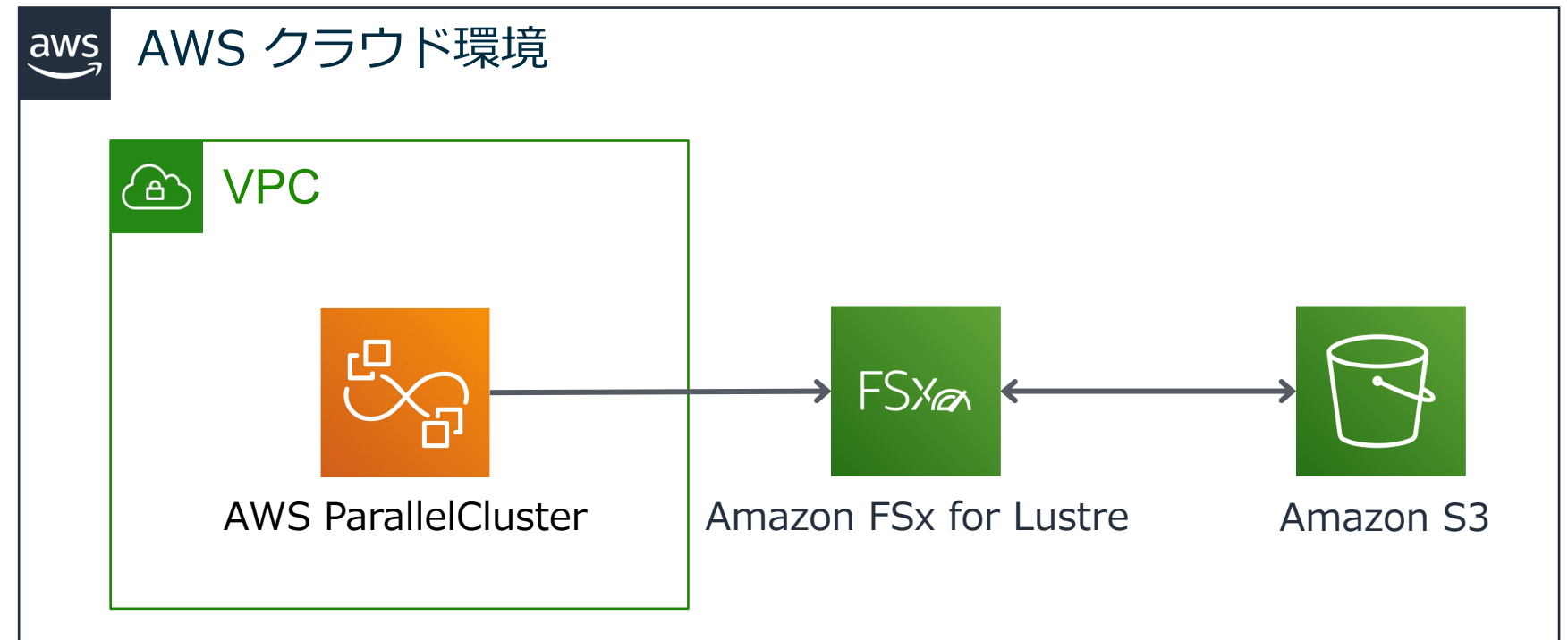
<https://aws.amazon.com/jp/blogs/opensource/deploy-hpc-cluster-remote-visualization-single-step-parallelcluster/>

AWS ParallelClusterと統合されたAWSサービス

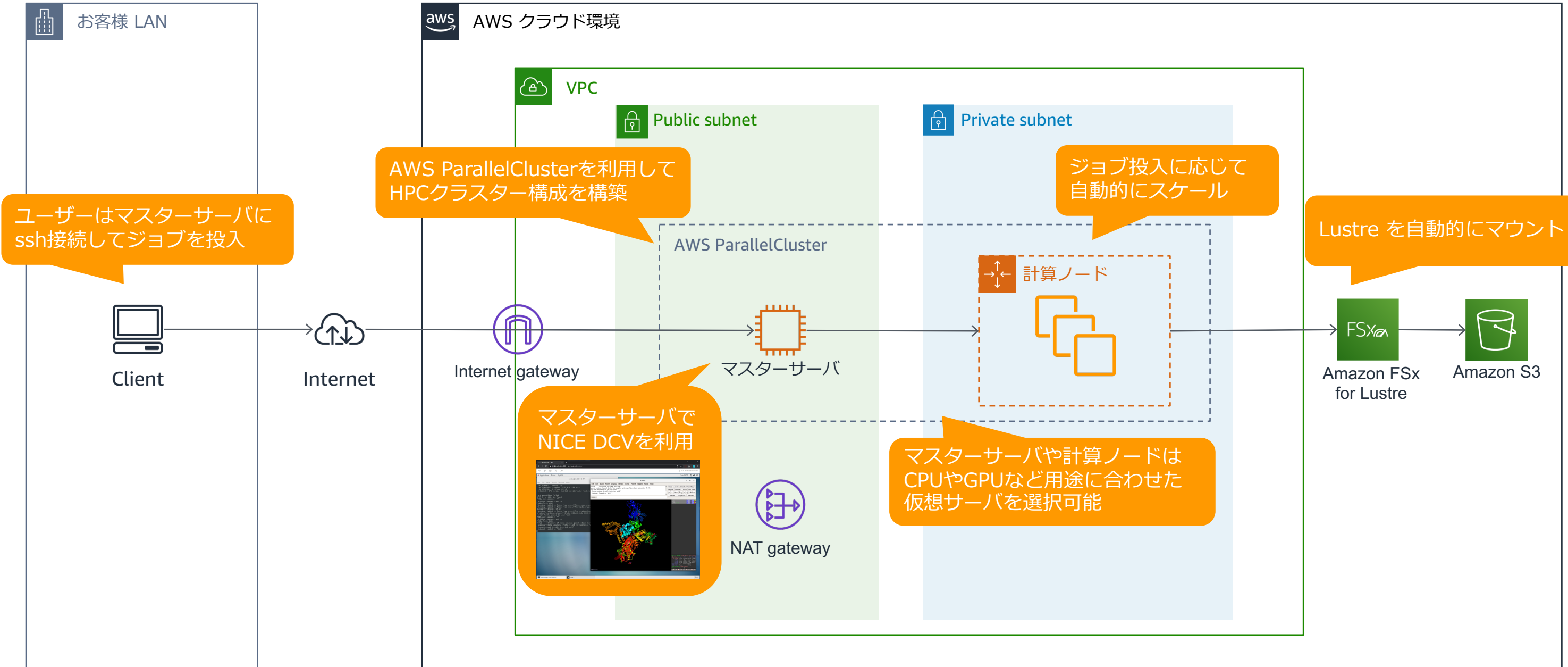
Amazon FSx for Lustre との統合による高速な外部ストレージ環境との接続

- 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

```
[fsx fs]
shared_dir = /fsx
storage_capacity = 3600
export_path = s3://bucket/folder
import_path = s3://bucket
```



AWS ParallelCluster を利用した一般的な HPC 構成例



AWS ParallelCluster を実際に利用を検討する中で出てくる課題

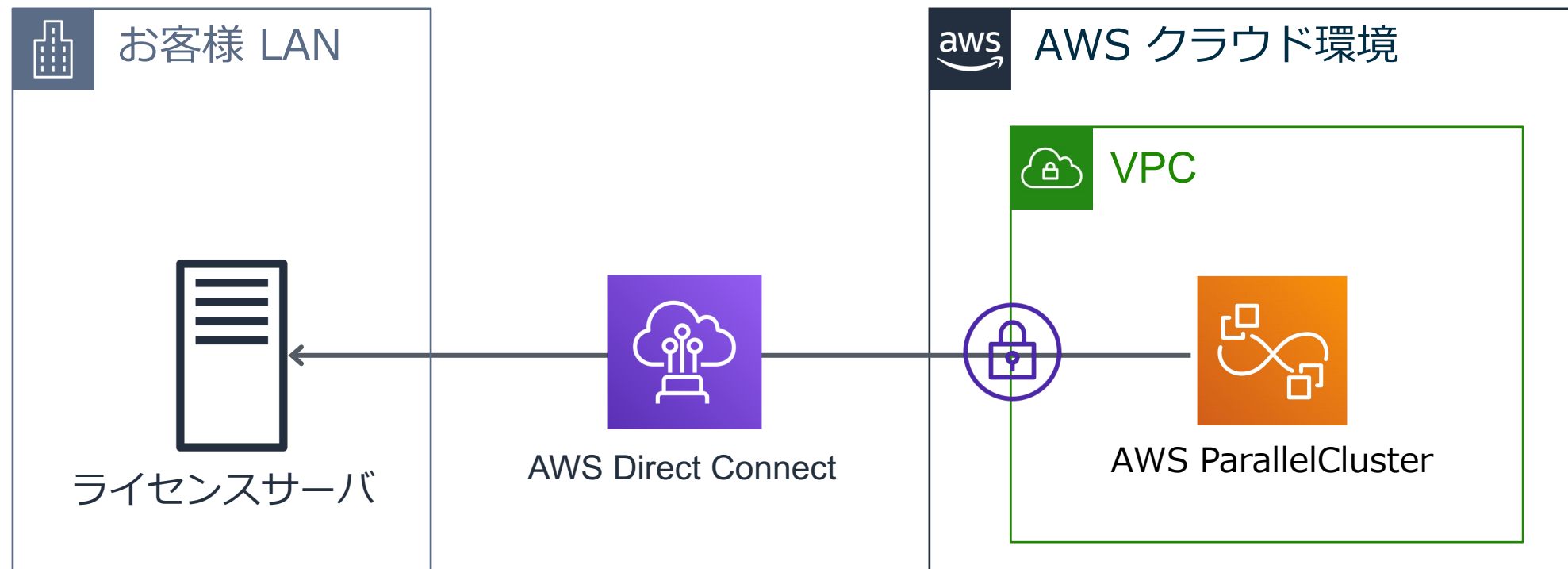
- ネットワーク
 - 社内ネットワークからの接続
 - 社内ライセンスサーバーとの接続
- 計算ノードのコスト効率化



ネットワークの課題

- 社内ネットワークからの接続・社内ライセンスサーバーとの接続

AWS Direct Connectと呼ばれる専用線接続を利用することでインターネット経由ではなく専用線経由でAWS環境にアクセス可能

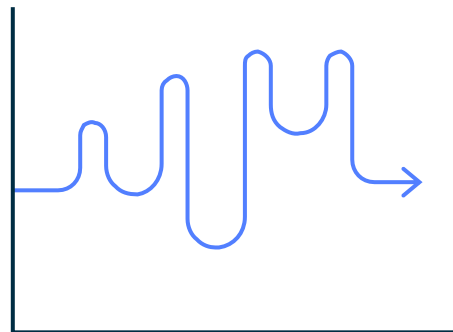


計算ノードのコスト効率化の課題

Amazon EC2の購入オプションを活用することでコストを抑えることが可能

オンデマンドインスタンス

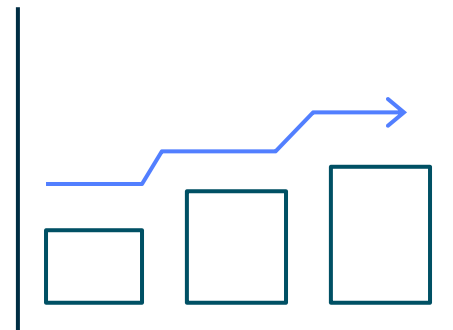
長期コミットなし、
使用分への支払い
(**秒単位/時間単位**)
Amazon EC2の定価



スパイクするような
ワークロード

リザーブドインスタンス (RI)

1年/3年の長期コミットに
応じた**大幅な**
ディスカウント価格



一定の負荷の見通しがあり、長期コミット
できるワークロード

Savings Plans

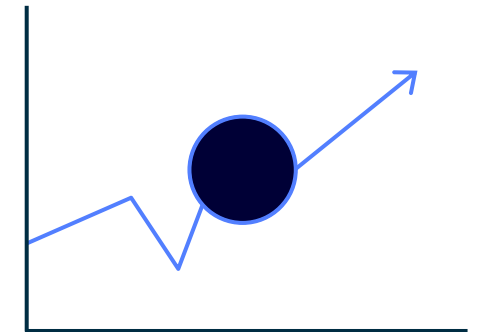
RIと同等のディスカウントに加え、**優れた柔軟性**
を提供



一定の負荷の見通しがあり、長期コミット
可能なとき
コンピュータ能力を
柔軟に変更できる

スポットインスタンス

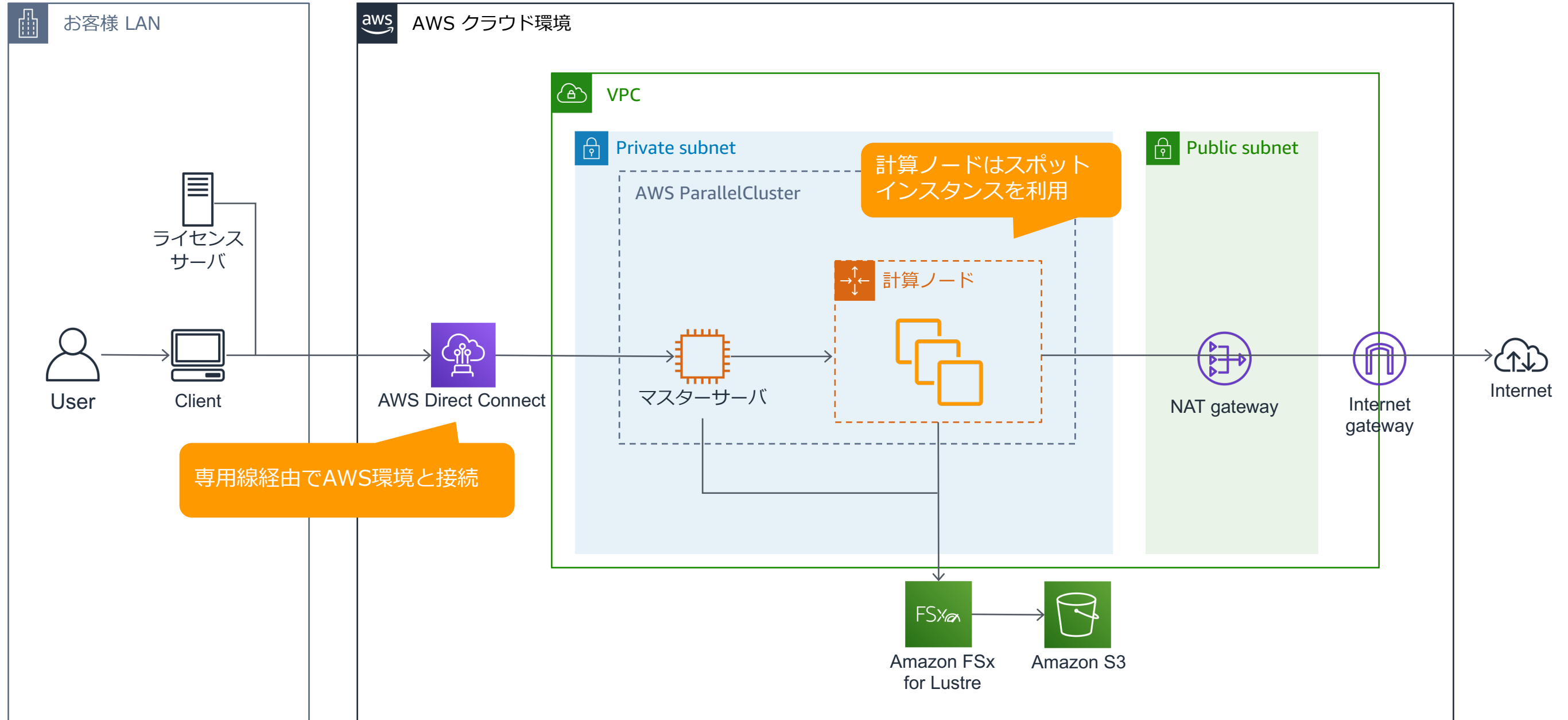
Amazon EC2の空き
キャパシティを活用し
最大90%の値引き
中断あり



中断に強く、様々なイン
スタンスタイプを活用
できるワークロード

EC2インスタンスとしての性能に違いはありません！

各課題を考慮したAWS ParallelClusterのHPC構成例



アステラス製薬様へのご支援内容

1. HPC on AWS のご紹介
2. AWS ParallelCluster のハンズオン実施
3. PoC 環境の検討
4. PoC 環境の構築サポート
5. 定例会での課題相談



AWSを利用したHPC環境構築をご検討頂く際はご相談頂けますと幸いです

まとめ

- AWS ParallelCluster を利用することで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可能
- お客様LAN環境とAWS環境を専用線で接続することで、セキュアな接続方式を構築可能
- 専用線を利用することでライセンスサーバーとの接続も可能
- スポットインスタンスの利用により計算ノードのコストを抑えることが可能

具体的なユースケースはアステラス製薬様のセッションでご紹介頂きます

Thank you!
ご清聴ありがとうございました