



創薬研究領域に関連する AWS 最新サービスご紹介

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
森下 裕介

森下裕介 / Morishita, Yusuke

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

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を担当



本セッションの概要

本セッションのゴール

技術的観点から以下の AWS サービスの概要を理解する

- 創薬領域において活用が進んでいる AWS サービス
- 日本たばこ産業株式会社様のセッションの理解に重要な AWS サービス

対象者

- 「これから AWS 上に創薬研究環境を構築したい」と検討中の研究者の方
- 「AWS 上に環境を構築はされているものの、もっと上手く活用したい」とお悩みの IT 部門の方

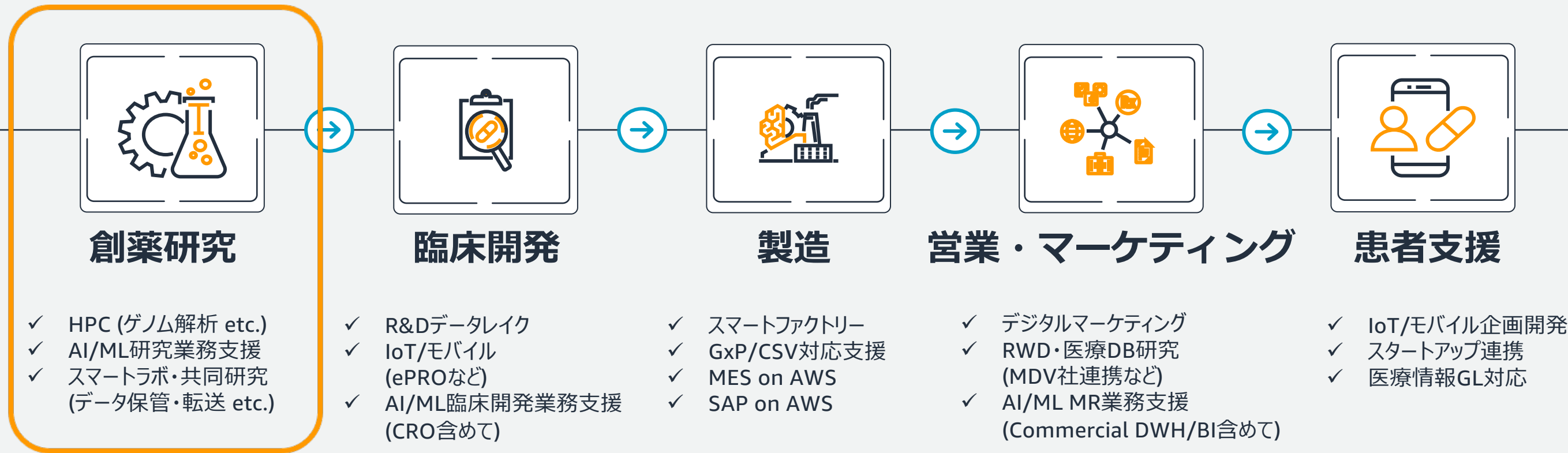
アジェンダ

- 最新ソリューションを交えた創薬領域で活用されている AWS サービスのご紹介
 - HPC
 - AI/ML
 - スマートラボ
- 日本たばこ産業様ご活用のサービスのご紹介



創薬領域で活用されている **AWS** サービスのご紹介

製薬バリューチェーンの各段階で DX を支援



本セッションでは **HPC** **AI/ML** **スマートラボ** の3カテゴリでご紹介します



創薬領域で活用されている **AWS** サービスのご紹介

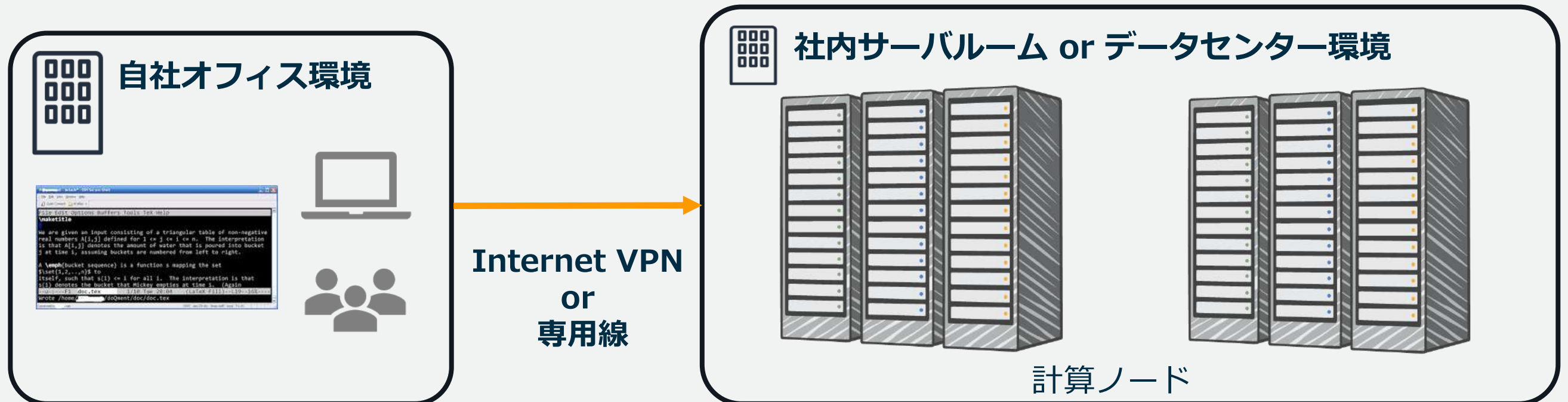
HPC

AI/ML

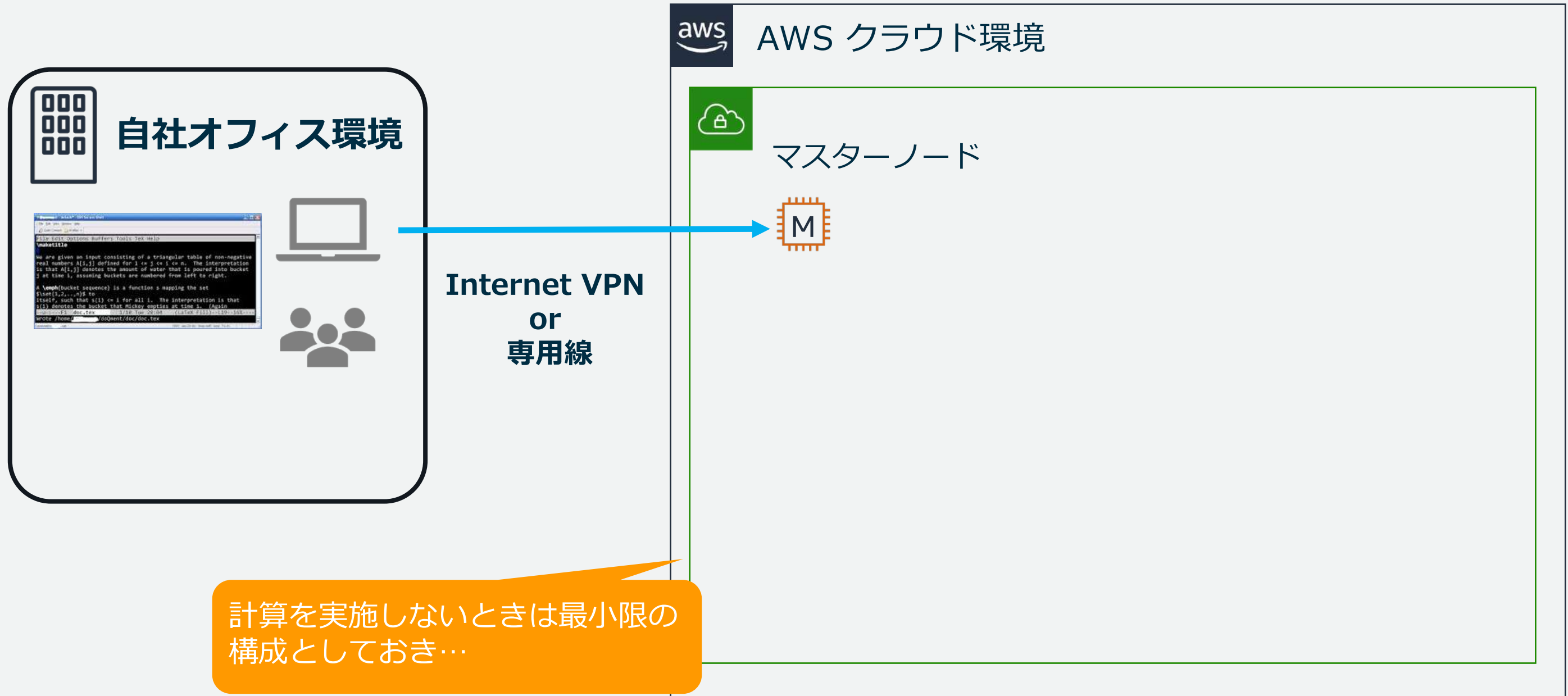
スマートラボ

従来の HPC クラスターの課題

- 台数が限られたサーバを共有するため、需要が増加する時期には**待ち時間が発生し**、繁忙期でない時期は、**リソースが無駄**になることもある
- サーバが使用されているか否かに関係なく、**維持費や保守管理の手間が発生**する
- 新規で構築する際は**調達に数ヶ月**かかることもある



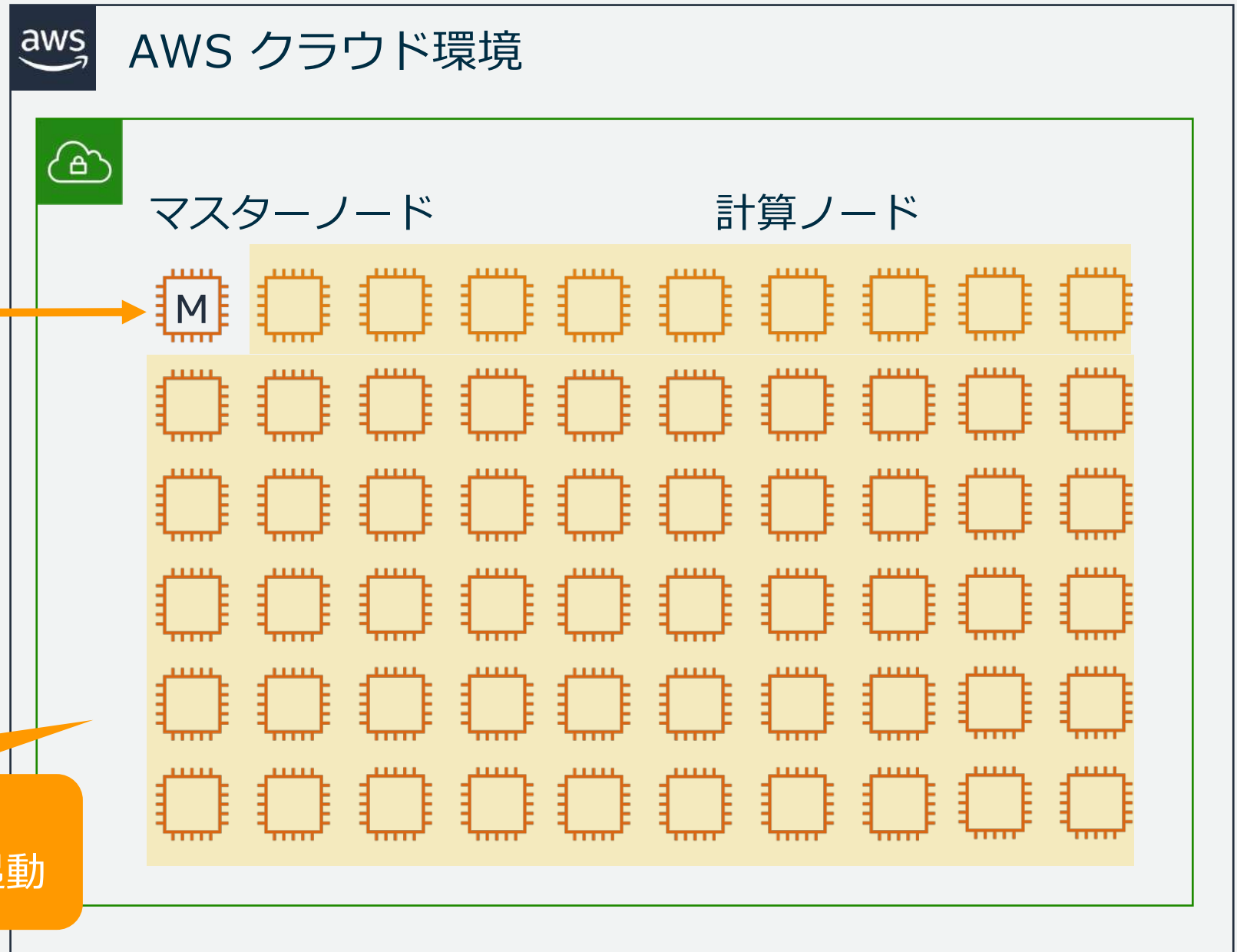
AWS を利用した基本的なシステム構成



AWS を利用した基本的なシステム構成

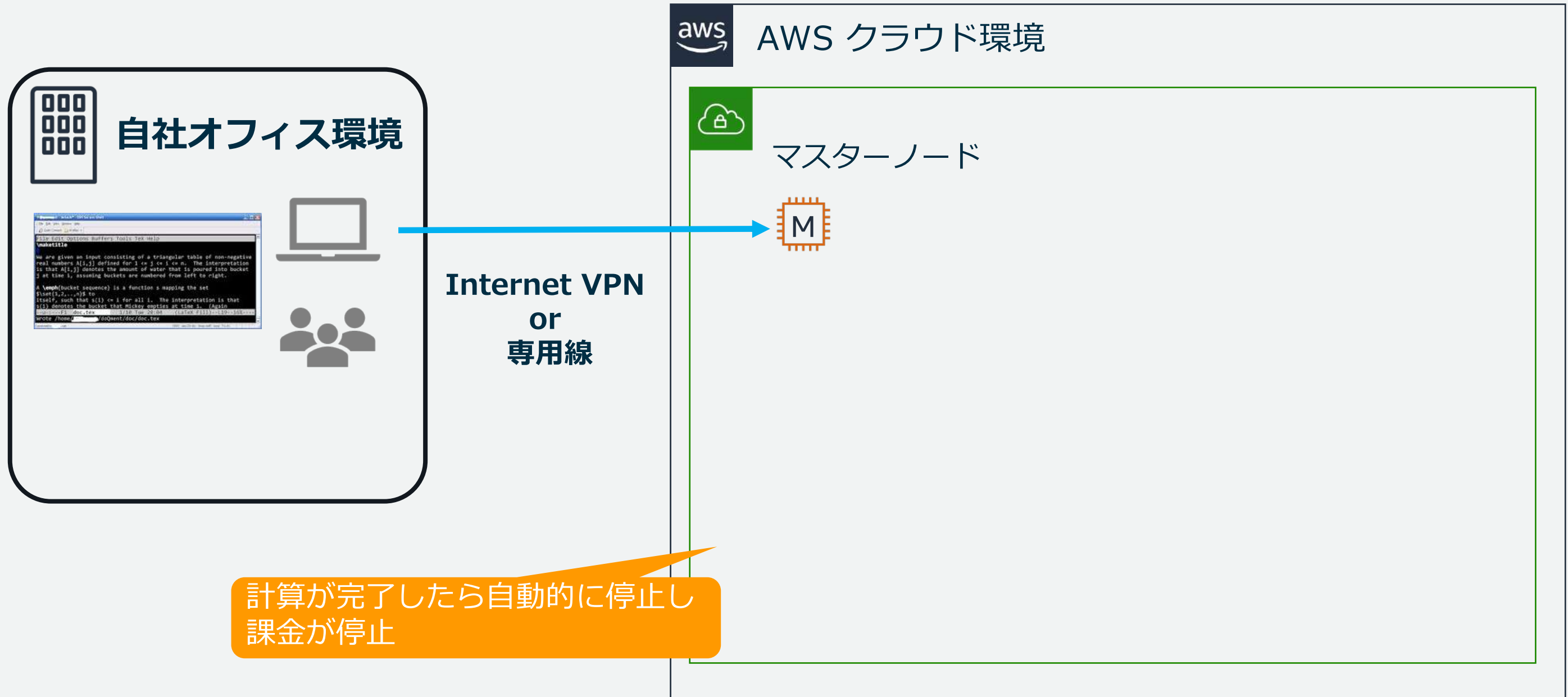


Internet VPN
or
専用線



計算を実施する時は、用途に合った
計算ノードを必要なスペックで台数だけ起動

AWS を利用した基本的なシステム構成



AWS における HPC 関連サービス

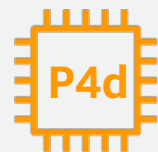
多様な HPC ワークロードに対応するための数多くのサービス

コンピューート

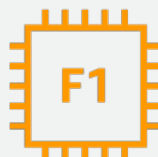
Amazon EC2



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



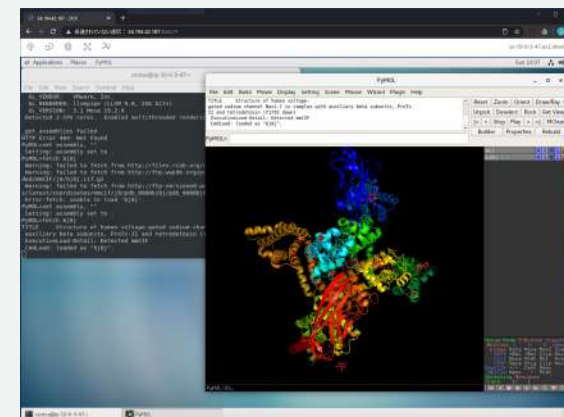
ストレージ内蔵の筐体を郵送することで S3 にデータを Import/Export

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 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 ParallelCluster



AWS上に HPC クラスタを自動で構築。ジョブスケジューラとして Slurm に対応しており、既存 HPC 環境からの移行が容易

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



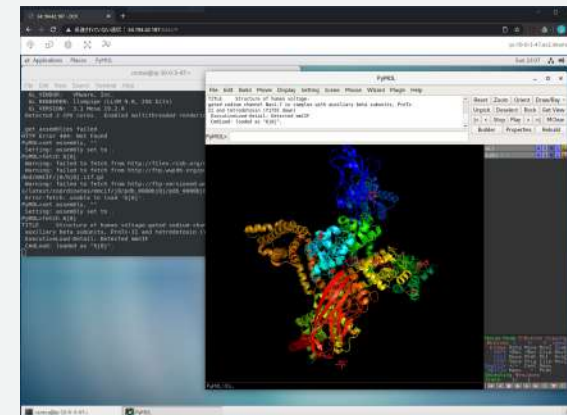
ストレージ内蔵の筐体を郵送することで S3 にデータを Import/Export

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 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 ParallelCluster



AWS上に HPC クラスタを自動で構築。ジョブスケジューラとして Slurm に対応しており、既存 HPC 環境からの移行が容易



Amazon EC2 (Elastic Compute Cloud)

必要なときに必要な計算リソースを確保可能な仮想サーバサービス

- 数分で起動し、秒単位の従量課金（一部タイプについては1時間単位）
- 独自の仮想化基盤 Nitro System により、仮想化オーバーヘッドを極小化
- ワークロードに応じて様々なインスタンスタイプを選択可能

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

アクセラレータの選択肢



Intel Xeon processor
(x86_64 arch)

AMD EPYC processor*
(x86_64 arch)

AWS Graviton Processor
(64-bit Arm arch)

NVIDIA GPU

Xilinx FPGA

M6i インスタンス
Ice Lake
最大時全コア 3.5 GHz 駆動

C6a インスタンス
AMD EPYC (Milan)
最大 3.6 GHz 駆動

C6g インスタンス
64bit Arm Neoverse N1ベース
Graviton2 CPU 搭載

P3 インスタンス
V100 GPU 搭載
P4d インスタンス
A100 GPU 搭載
G4 インスタンス
T4 GPU 搭載

F1 インスタンス
Virtex UltraScale+
VU9P 搭載

M5zn インスタンス
Cascade Lake
最大全コア 4.5 GHz 駆動



Amazon S3 (Simple Storage Service)

インフラ管理不要な容量無制限のストレージサービス

容量無制限

1ファイル最大5TB

低コスト

Standard: 0.025 USD/GB※1

～ S3 Glacier Deep Archive: 0.002 USD/GB ※1

(例: 1PBのデータをS3 Glacier Deep Archiveで1年保管 = 288万円/年) ※2

高い耐久性

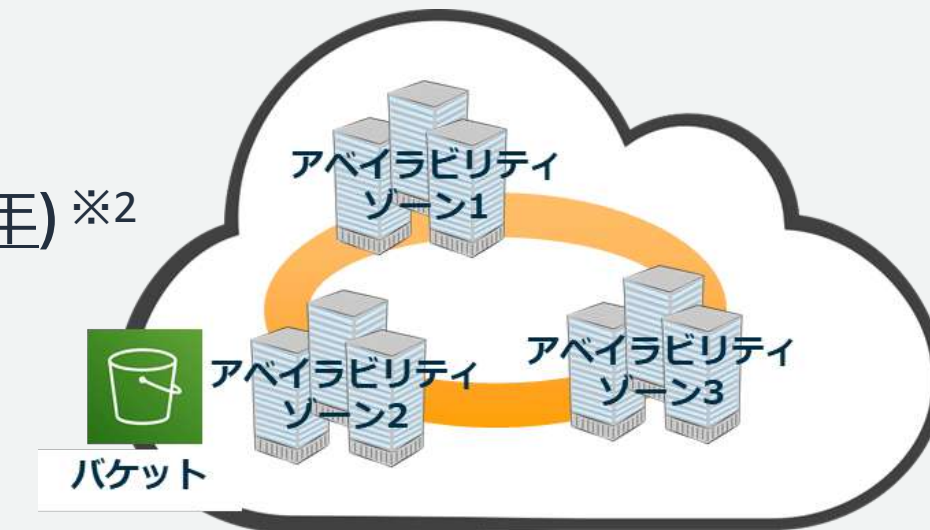
設計上の耐久性は99.9999999999%

HTTPSでアクセス

PUT/GET、その他メソッドで

スケーラブルで安定した性能

データ容量に依存しない性能

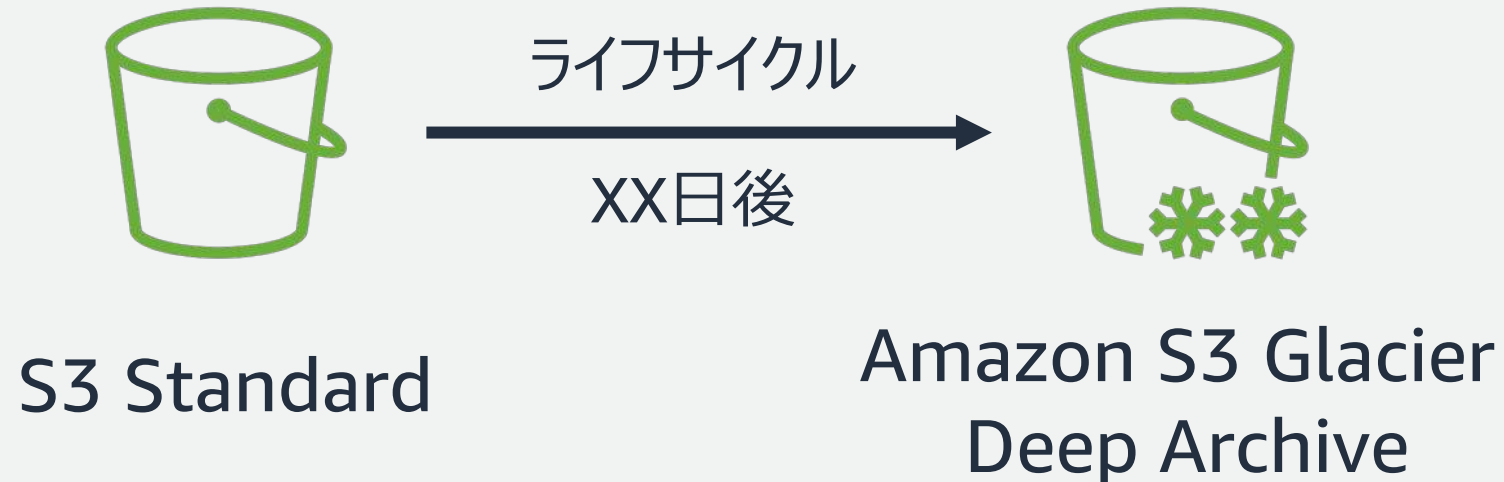


※1: 2022年4月現在のap-northeast-1での価格
価格はストレージクラスによって異なる
<https://aws.amazon.com/jp/s3/pricing/>

※2: 1ドル120円で計算

ストレージクラスとライフサイクル管理

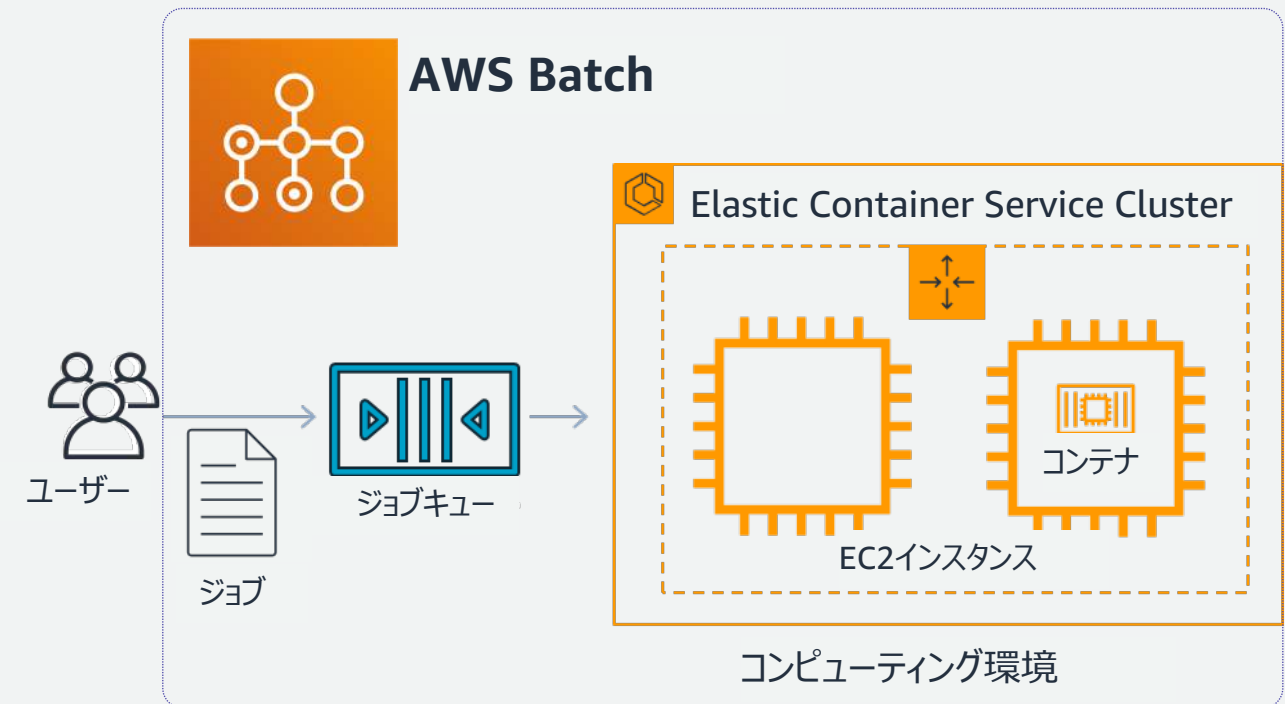
- データの利用頻度や可用性に応じて料金が異なる複数の**ストレージクラスをご用意**
 - S3 Standard: アクセス頻度の高いデータ向け、ミリ秒単位のアクセス
 - S3 Glacier Deep Archive: 長期アーカイブ向け、数時間で取出、最も安価なストレージクラス
- 格納データに対して、**ストレージクラス変更やデータ削除を自動で行うライフサイクル管理**



AWS Batch

大規模バッチ処理のため環境をフルマネージドで提供

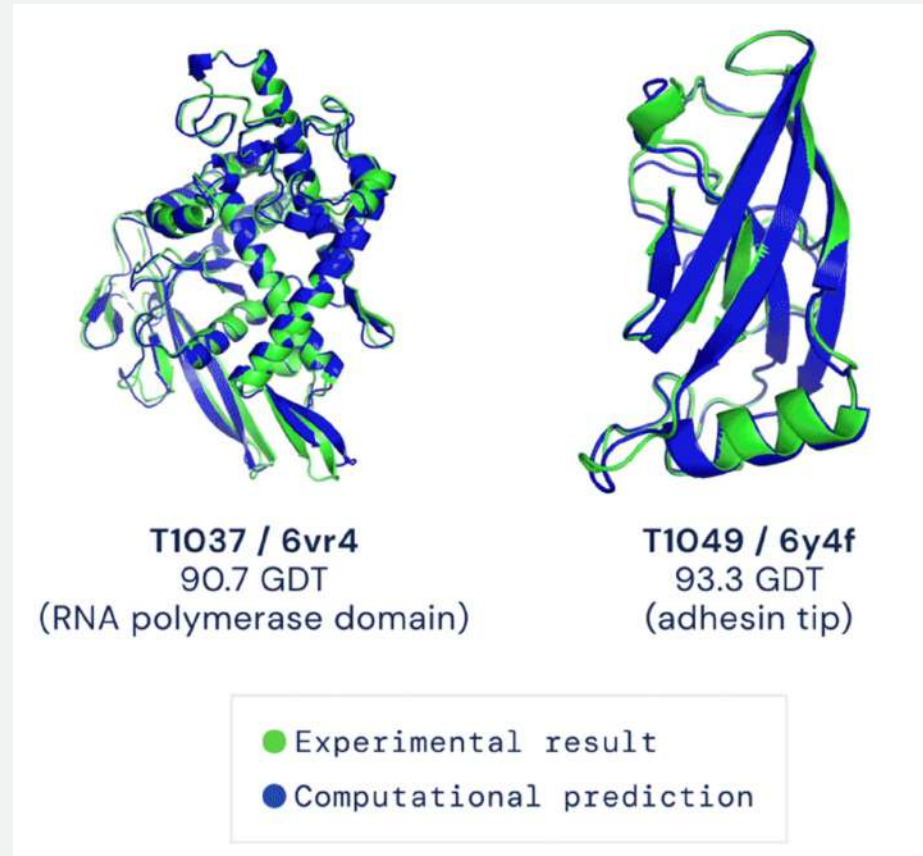
- AWS Batch がインスタンスの起動や停止を行うため、スケジューラや計算ノードなどの **管理が不要**
- ジョブは **Docker コンテナイメージ** を元に作成し、自動でスケールするコンピューティング環境で実行する
- コンピューティング環境ではインスタンスタイプや vCPU 数、スポットインスタンス利用有無などを任意に指定可能



コンテナイメージを用意するだけでスケーラブルな大規模バッチ処理環境が得られる

RoseTTAFold / AlphaFold on AWS Batch [New]

AWS Batchを用いた大規模なタンパク質立体構造予測

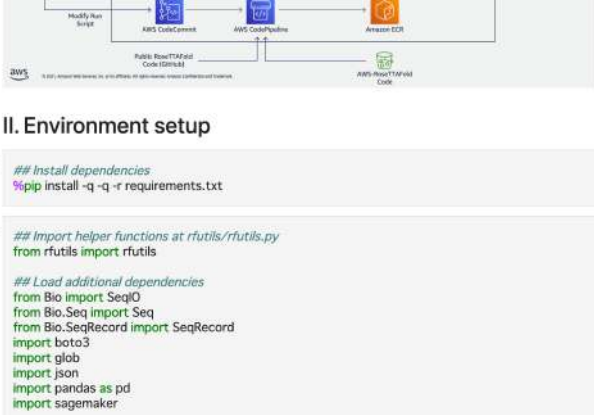


- AWS Batch を活用したスケーラブルなジョブ実行環境で RoseTTAFold / AlphaFold などのタンパク質立体構造予測を実行可能
- Amazon EC2 G4dn インスタンス (NVIDIA T4 GPU 搭載) を使用、RoseTTAFoldの元論文と同等またはそれ以上のパフォーマンスを発揮
- CloudFormation テンプレートが提供されており簡単に展開可能
 - 予測ジョブを実行するAWS Batchのコンピューティング環境
 - 構造予測ジョブを投入するJupyter Notebook
 - 構造予測のジョブ定義
 - etc.

RoseTTAFold 版: <https://aws.amazon.com/jp/blogs/news/predicting-protein-structures-at-scale-using-aws-batch/> (blog)
<https://github.com/aws-samples/aws-rosettafold> (GitHub)

AlphaFold2 版 : <https://github.com/aws-samples/aws-batch-architecture-for-alphafold> (GitHub)

RoseTTAFold / AlphaFold on AWS Batch の構成



```

II. Environment setup
In [ ]: ## Install dependencies
%pip install -q -r requirements.txt

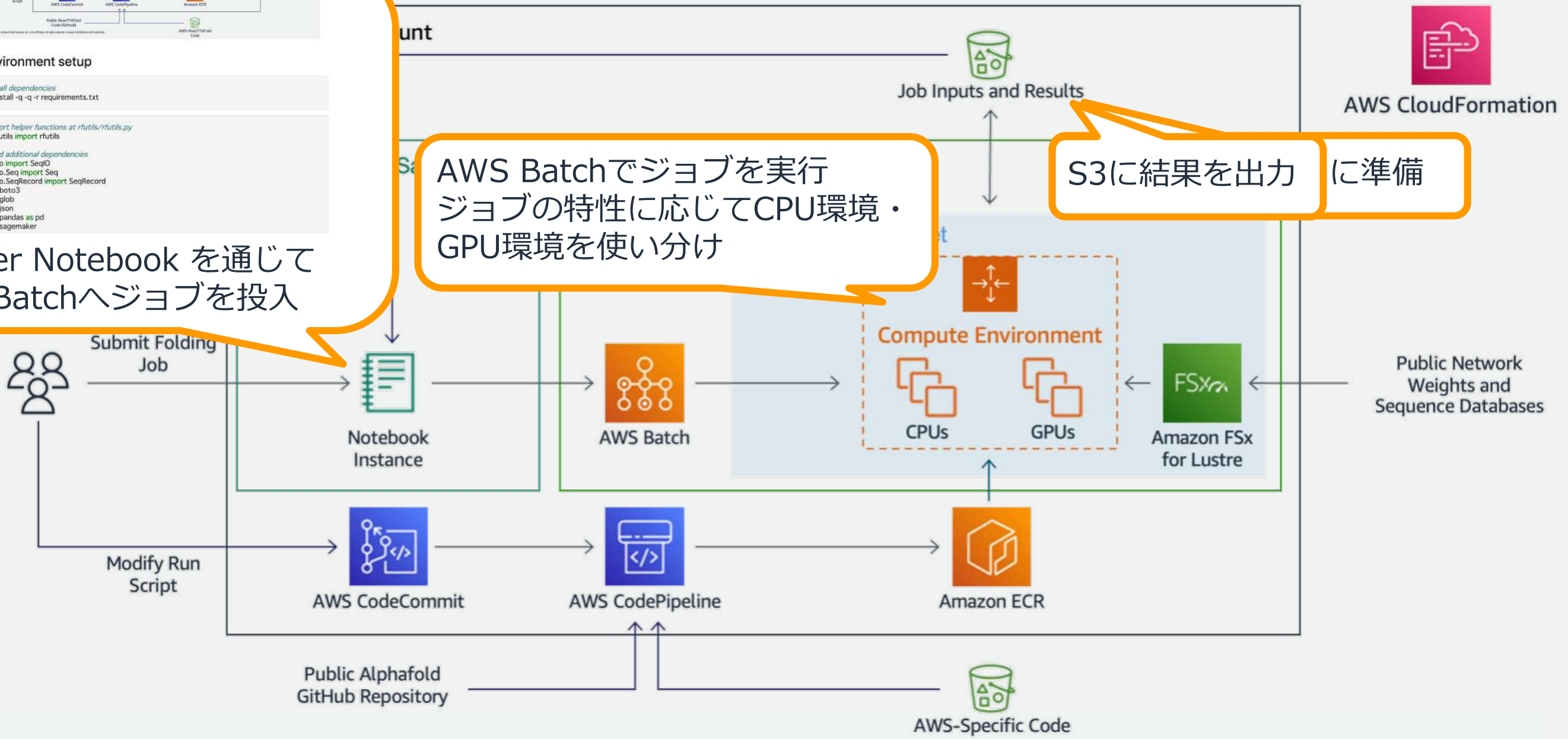
In [ ]: ## Import helper functions at rutils/rutils.py
from rutils import rutils

## Load additional dependencies
from Bio import SeqIO
from Bio.Seq import Seq
from Bio.SeqRecord import SeqRecord
import boto3
import glob
import json
import pandas as pd
import sagemaker
  
```

Jupyter Notebook を通じて
AWS Batchへジョブを投入

AWS Batchでジョブを実行
ジョブの特性に応じてCPU環境・
GPU環境を使い分け

S3に結果を出力
に準備



RoseTTAFold / AlphaFold on AWS Batch 特徴

- 特徴量生成や構造予測に他のアルゴリズムを容易に取り入れることが可能
 - MSA 生成と構造予測をそれぞれ別のジョブとして定義しているため
 - 例：MSA 生成アルゴリズムを hhblits から MMseqs2 に変える、など
- コスト例 (RoseTTAFold on AWS Batch、テンプレートと同一構成の場合)
 - 立ち上げたAWSリソースに対して課金、ソリューション自体の使用は無料
 - 維持費：\$760 / month
 - ジョブ：\$0.50 / job



創薬領域で活用されている **AWS** サービスのご紹介

HPC

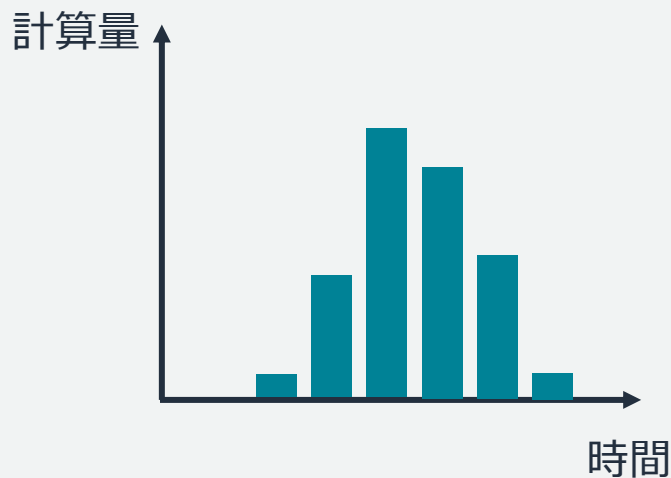
AI/ML

スマートラボ

従来の R&D 機械学習ワークロードの課題

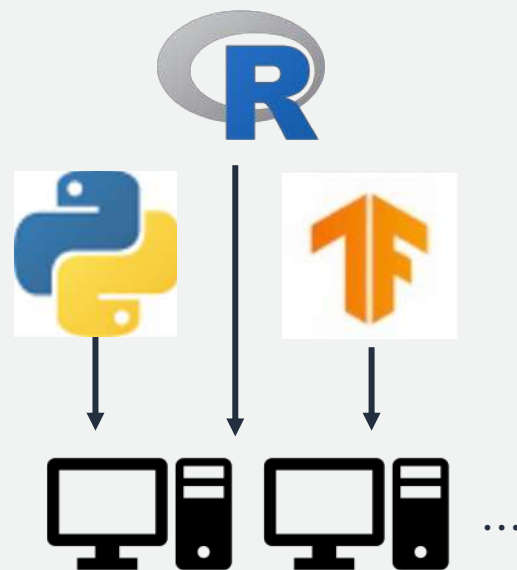
計算量の変化

実験期間中のみ
大量の計算が発生
→ 待ち時間・アイドル
タイムの発生



実験設定の手間

環境を変えるたびに
実験用設定や
環境構築を実施



実験内容の管理

実験を繰り返すうちに
データ量が膨大になり
結果の管理が煩雑に


























機密データの保護

R&D で利用するデータの
多くは機密性が高く
社外に出すのが躊躇われる




AWS の機械学習サービススタック

AI サービス: 機械学習の深い知識なしに利用可能

HEALTH AI  NEW Amazon HealthLake  Amazon Transcribe Medical  Amazon Comprehend Medical			INDUSTRIAL AI  NEW AWS Panorama + Appliance  NEW Amazon Monitron  NEW Amazon Lookout for Equipment  NEW Amazon Lookout for Vision			ANOMALY DETECTION  NEW Amazon Lookout for Metrics	CODE AND DEVOPS  NEW Amazon DevOps Guru  Amazon CodeGuru			
VISION  Amazon Rekognition	SPEECH  Amazon Polly  Amazon Transcribe <small>+Medical</small>		TEXT  Amazon Comprehend <small>+Medical</small>  Amazon Translate  Amazon Textract		SEARCH  Amazon Kendra	CHATBOTS  Amazon Lex	PERSONALIZATION  Amazon Personalize	FORECASTING  Amazon Forecast	FRAUD  Amazon Fraud Detector	CONTACT CENTERS  Contact Lens  Voice ID <small>For Amazon Connect</small>

ML サービス: 機械学習のプロセス全体を効率化するマネージドサービス

Amazon SageMaker  **SAGEMAKER STUDIO IDE**

Label data

NEW Aggregate & prepare data

NEW Store & share features

Auto ML

Spark/R

NEW Detect bias

Visualize in notebooks

Pick algorithm

Train models

Tune parameters

NEW Debug & profile

Deploy in production

Manage & monitor

NEW CI/CD

Human review

NEW: SageMaker JumpStart

NEW: Model management for edge devices

ML フレームワークとインフラストラクチャ: 機械学習の環境を自在に構築して利用

 TensorFlow  mxnet  PyTorch	 GLUON  Keras  learn  HOROVOD  DeepGraphLibrary	Deep Learning AMIs & Containers	GPUs & CPUs	Elastic Inference	Trainium	Inferentia	FPGA
--	--	---------------------------------	-------------	-------------------	----------	------------	------



Amazon SageMaker

機械学習のプロセス全体を効率化するマネージドサービス

Amazon SageMaker

準備

- SageMaker Ground Truth**
教師データをラベリング
- SageMaker Data Wrangler**
データを統合し準備
- SageMaker Processing**
ビルトインの Python, BYO R/Spark
- SageMaker Feature Store**
特徴量を保存、更新、取得、共有

開発

- SageMaker Studio Notebooks**
伸縮可能なコンピュートと共有機能を持つ Jupyter notebooks
- Built-in and Bring-Your-Own Algorithms**
数十種類の最適化されたアルゴリズム、または自分のアルゴリズムを持ち込む
- Local Mode**
ローカルマシンでテストしプロトタイプ
- SageMaker Autopilot**
可視性を備えた機械学習モデルの自動構築

学習とチューニング

- One-click training**
分散型インフラ管理
- SageMaker Experiments**
すべてのステップを記録し、整理し、比較
- Automatic Model Tuning**
ハイパーパラメータの最適化
- SageMaker Debugger**
学習時のデバッグ
- Managed Spot Training**
学習コストを 90% 削減

デプロイと管理

- One-click Deployment**
Fully managed, ultra low latency, high throughput
- Kubernetes & Kuberflow integration**
Kubernetesベースの機械学習を簡素化
- Multi-model endpoints**
インスタンスごとに複数のモデルをホストすることでコストを削減
- Model monitor**
デプロイされたモデルの精度を維持
- SageMaker Pipelines**
ワークフローのオーケストレーションと自動化

SageMaker Studio

Integrated development environment (IDE) for ML



Amazon SageMaker Studio

機械学習ワークロードのための統合開発環境 (IDE)

- 機械学習エンジニアやサイエンティスト向けに、モデル構築、学習、デプロイなどのそれぞれの作業ステップを実行できる Web ベースのグラフィカルな UI を提供
- SageMaker Studio Notebooks での学習・推論実行、各種 SageMaker の機能を GUI / コードから使用することが可能

The screenshot displays the Amazon SageMaker Studio interface. The main window shows a Jupyter Notebook with the following content:

```

xgboost_customer_churn.ipynb
conda_amazonei_mxnet_p27

• Have the predictor variable in the first column
• Not have a header row

But first, let's convert our categorical features into numeric features.

[ ]: model_data = pd.get_dummies(churn)
      model_data = pd.concat([model_data[Churn: True], model_data.drop([Churn: True], axis=1)])
      ...

And now let's split the data into training, validation, and test sets. This will help prevent us from overfitting the model, and allow us to test the models accuracy on data it hasn't already seen.

[ ]: train_data, validation_data, test_data = np.split(model_data.sample(frac=1, random_state=1234), [int(0.7 * len(model_data)), int(0.1 * len(model_data))])
      train_data.to_csv('train.csv', header=False, index=False)
      validation_data.to_csv('validation.csv', header=False, index=False)
      ...

Now we'll upload these files to S3.

[ ]: boto3.Session().resource('s3').Bucket(bucket).Object(os.path.join(prefix, 'train.csv')).upload(train_data.to_csv('train.csv', header=False, index=False))
      boto3.Session().resource('s3').Bucket(bucket).Object(os.path.join(prefix, 'validation.csv')).upload(validation_data.to_csv('validation.csv', header=False, index=False))
      ...
    
```

On the right side, there are two panels:

- Trial Component Chart:** A line chart showing training metrics (train_loss, test_loss) over 10 periods. The y-axis ranges from 0.0 to 0.4. The x-axis is labeled 'period'.
- Trial Component List:** A table listing trial components. It shows 10 rows selected, all with a status of 'Completed'.

Status	Experiment	Type	Trial	Trial ID
Completed	customer-churn-predi...	Training job	Trial-3	Tr...
Completed	customer-churn-predi...	Training job	Trial-2	Tr...
Completed	customer-churn-predi...	Training job	Trial-1	Tr...
Completed	customer-churn-predi...	Training job	Trial-0	Tr...

The bottom status bar shows: Mode: Command | Ln 1, Col 1 | xgboost_customer_churn.ipynb



Amazon SageMaker Canvas [New]

ノーコードで ML モデル構築・推論生成

- コードを書かずに ML モデルを構築したい人に向けた、**ノーコードで高精度な機械学習による予測処理**を実現できる SageMaker の新機能
- SageMaker と同じテクノロジーにより、データに基づいてモデルを自動的にトレーニング。内部で数百のモデルを作成し、最良の予測を出力
- **東京リージョン**含む 6 つのリージョンで使用可能 (2022 年 4 月現在)

The screenshot shows the SageMaker Canvas interface for building a model. The 'Build' tab is active, showing a 'Select a column to predict' dropdown set to 'Target'. A 'Value distribution' bar chart shows two categories, 1 and 0. The 'Model type' section recommends '2 category prediction' and shows a 'Quick build' button. Below, a table lists columns from the 'cancer_dataset_train.csv' dataset with their data types and correlations to the target.

Column name	Data type	Missing	Mismatched	Unique	Mean / Mode	Correlation to target
<input checked="" type="checkbox"/> worst texture	Numeric	0.00% (0)	0.00% (0)	454	17.7	-0.484
<input checked="" type="checkbox"/> worst symmetry	Numeric	0.00% (0)	0.00% (0)	447	0.24	-0.42
<input checked="" type="checkbox"/> worst smoothness	Numeric	0.00% (0)	0.00% (0)	375	0.15	-0.469

Below the table, there are histograms for selected columns: '_c0', 'worst texture', 'worst symmetry', 'worst smoothness', 'worst radius', and 'worst perimeter'.

https://docs.aws.amazon.com/ja_jp/sagemaker/latest/dg/canvas.html



創薬領域で活用されている **AWS** サービスのご紹介

HPC

AI/ML

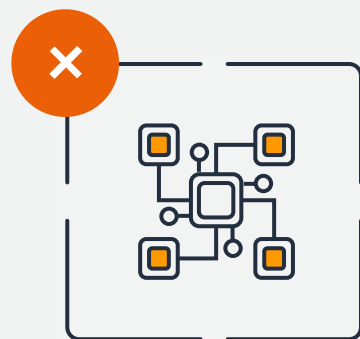
スマートラボ

創薬ラボにおけるよくある課題

データのサイロ化

各機器・各組織ごとに
データをバラバラに管理

社内外での安全なデータ
共有が困難



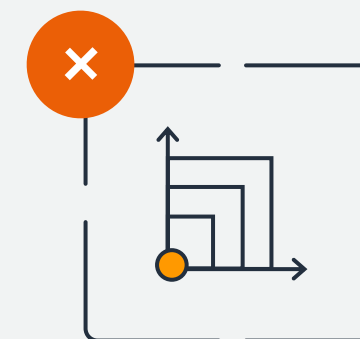
手作業での 反復作業

ルーティンワークの自動化
ができておらず実験に使う
時間が減少



研究展開の遅延

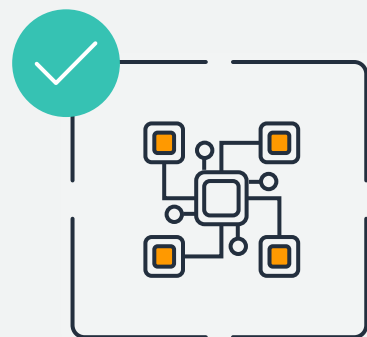
小規模な研究から
様々なデータを結合した
大規模な研究への
展開が遅い



スマートラボ化による解決

データの自動収集・統合

データを自動で収集・統合
社内外へ迅速かつ安全な
データの共有、コラボレーション
が容易に



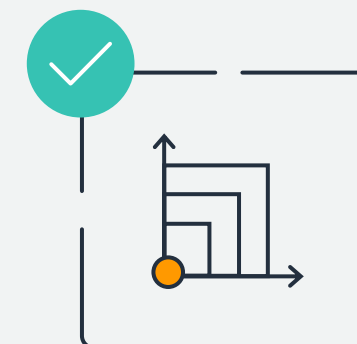
作業の自動化

ルーティンワークを自動化し
価値の高い作業に集中できる

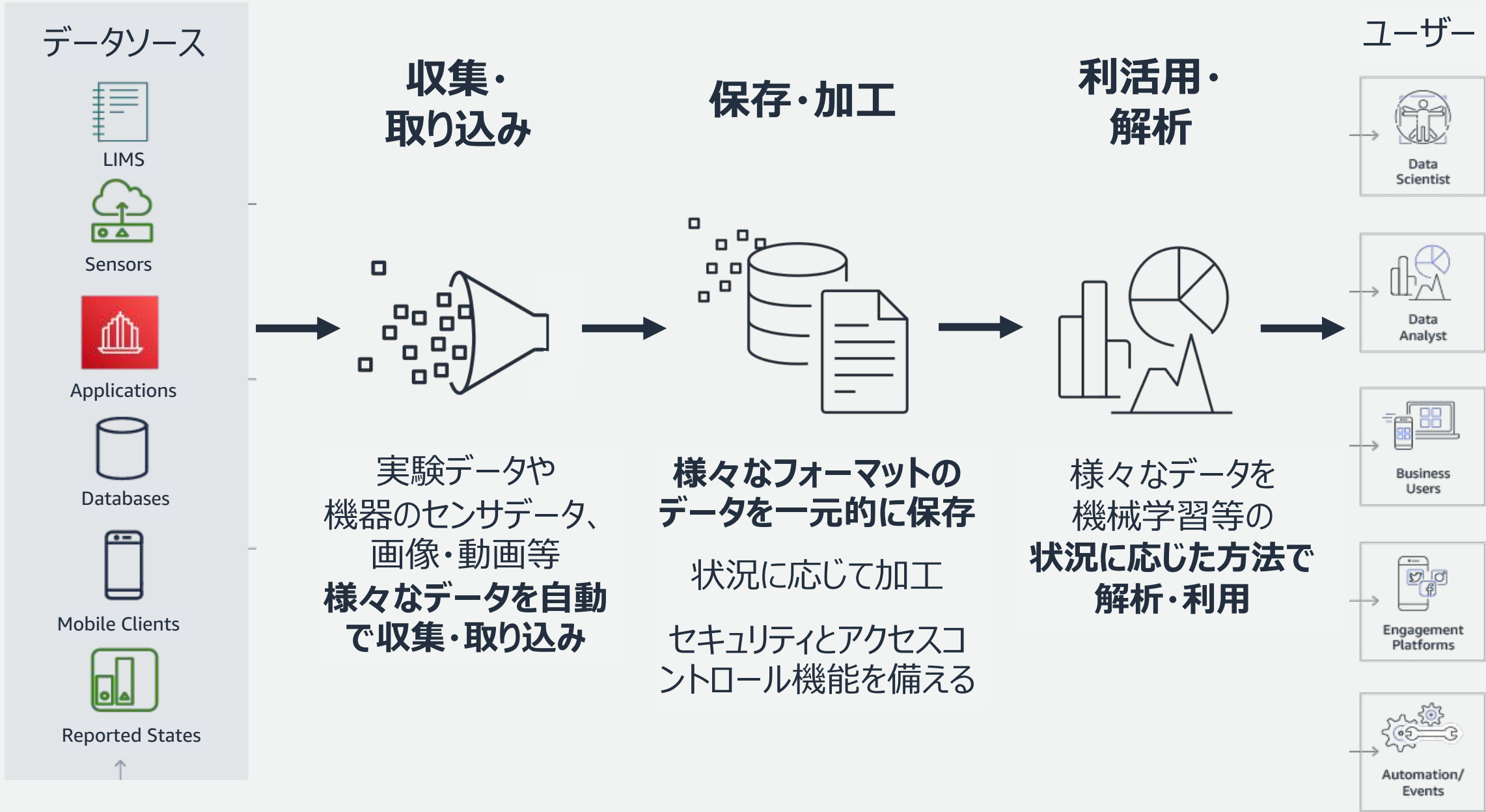


研究の展開の加速

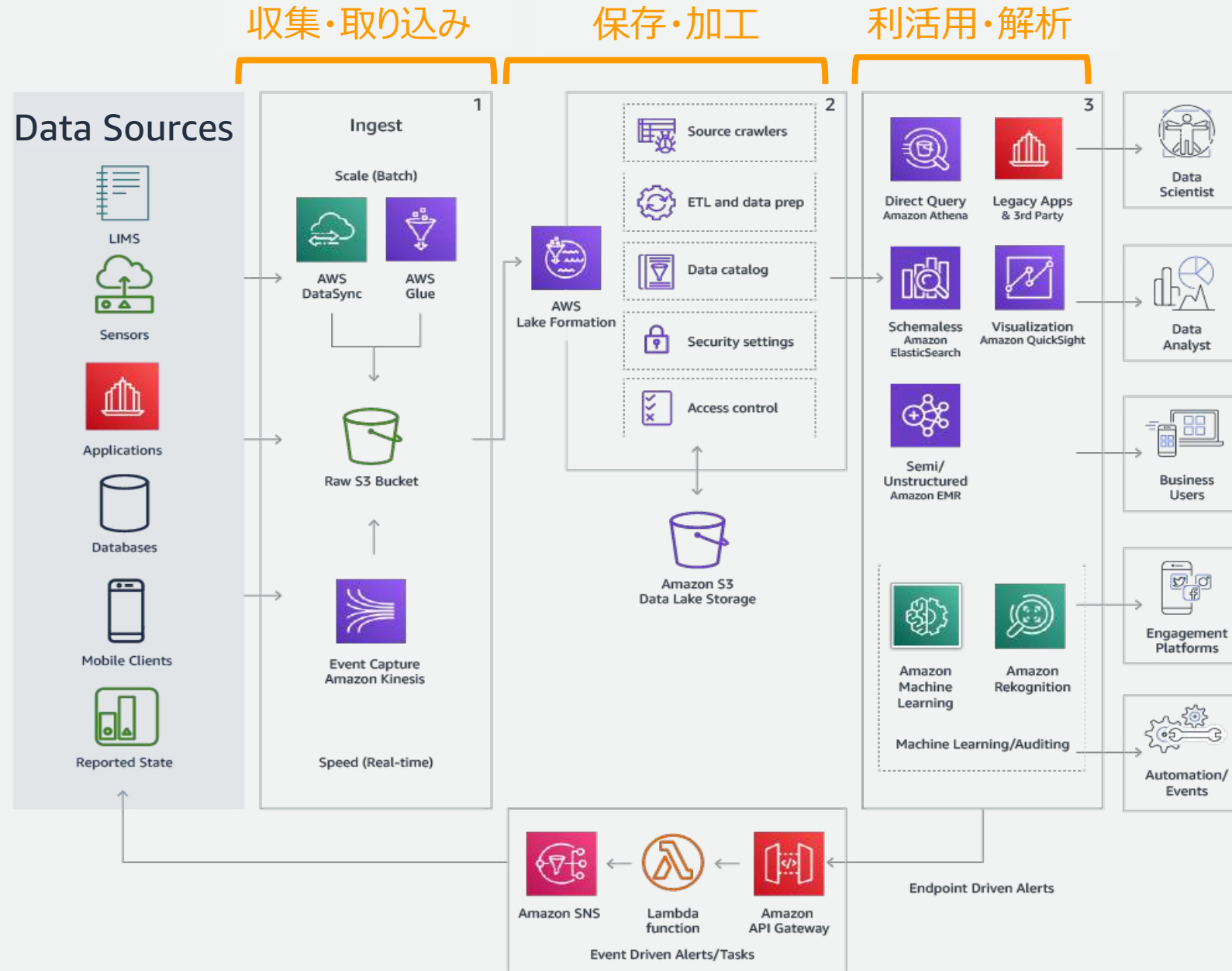
最適化されたライフサイ
クル管理とデータ統合に
より研究が加速



スマートラボ - データ統合

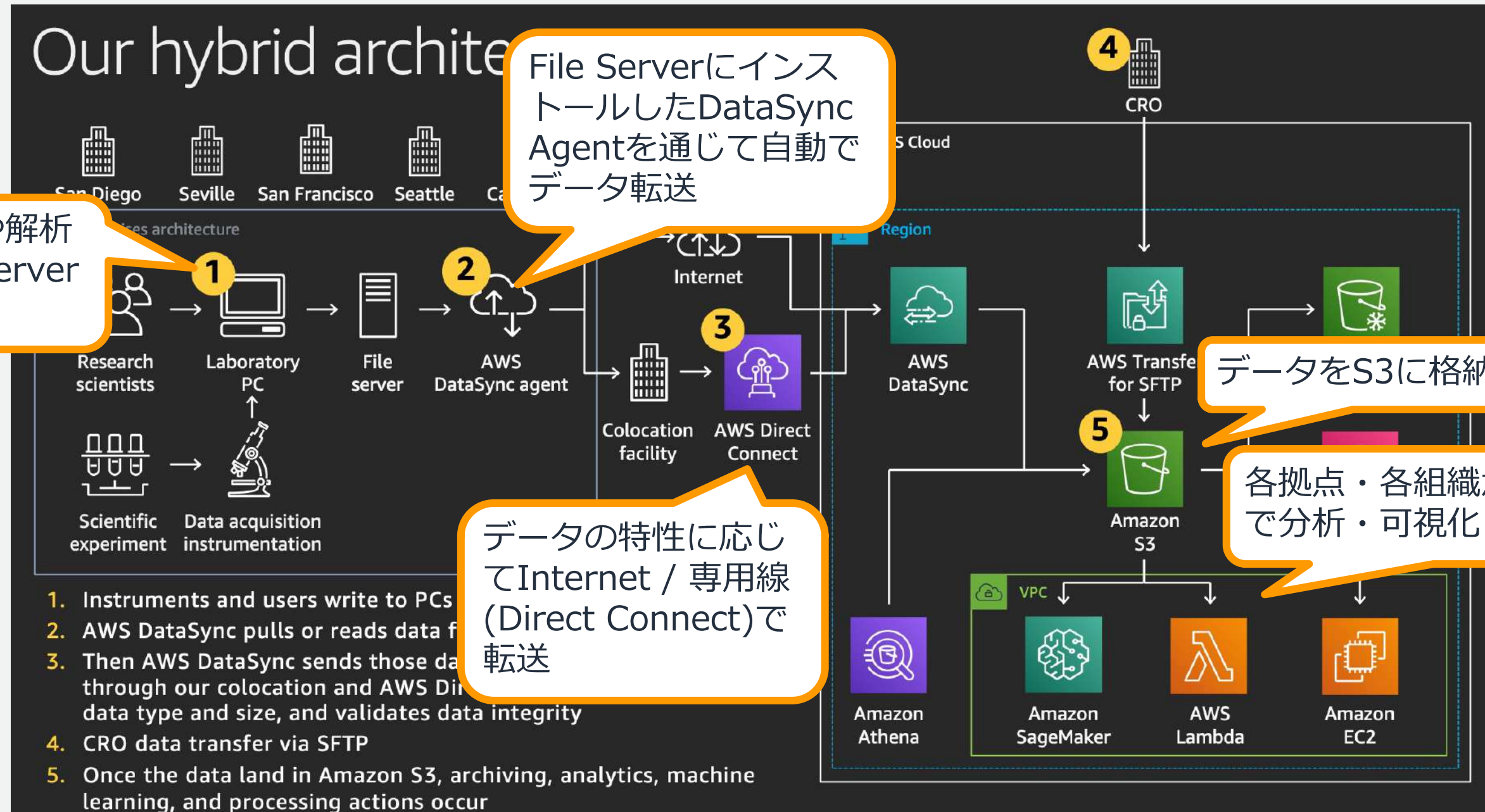


スマートラボ リファレンスアーキテクチャ



スマートラボ事例: セルジーン様

研究業務促進のための複数拠点のデータを統合





日本たばこ産業様ご活用サービスのご紹介

日本たばこ産業様ご活用サービスについて

セッションをご理解していただくために重要なサービスの概要をご説明

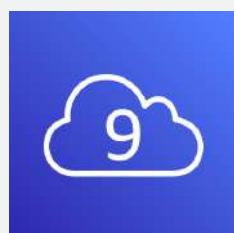
HPC 関連



AWS ParallelCluster
HPC クラスタ管理自動化



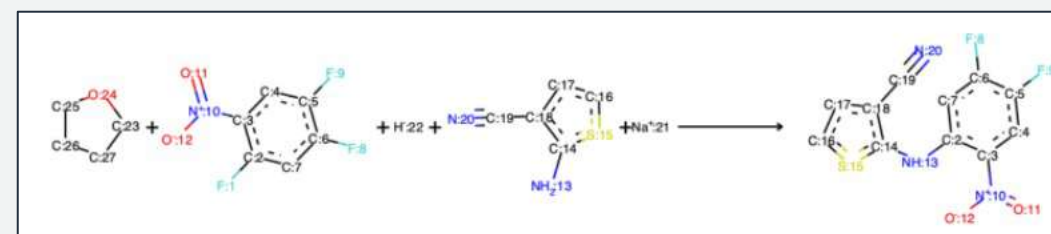
Amazon FSx for Lustre
分散ファイルシステム



AWS Cloud9
クラウドベースのIDE



AI/ML



DGL-LifeSci

ライフサイエンス領域向けの
GNNライブラリ

https://github.com/aws-labs/dgl-lifesci/tree/master/examples/reaction_prediction/rexgen_direct

日本たばこ産業様ご活用サービスについて

セッションをご理解していただくために重要なサービスの概要をご説明

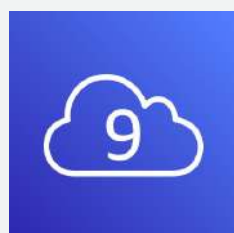
HPC 関連



AWS ParallelCluster
HPC クラスタ管理自動化



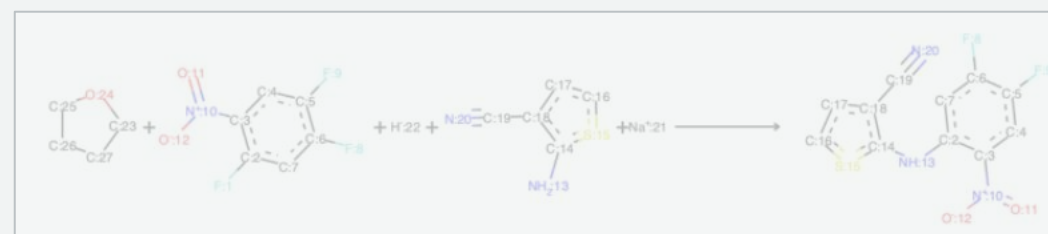
Amazon FSx for Lustre
分散ファイルシステム



AWS Cloud9
クラウドベースのIDE



AI/ML



DGL-LifeSci

ライフサイエンス領域向けの
GNNライブラリ

https://github.com/aws-labs/dgl-lifesci/tree/master/examples/reaction_prediction/rexgen_direct

AWS における HPC 関連サービス (再掲)

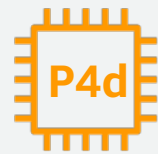
多様な HPC ワークロードに対応するための数多くのサービス

コンピューート

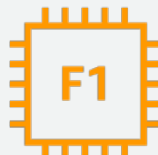
Amazon EC2



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



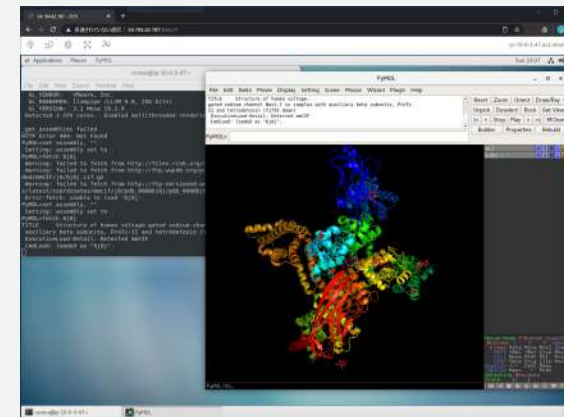
ストレージ内蔵の筐体を
郵送することで S3 にデータを
Import/Export

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 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 ParallelCluster



AWS上に HPC クラスタを
自動で構築。ジョブスケ
ジューラとして Slurm に対
応しており、既存 HPC 環
境からの移行が容易

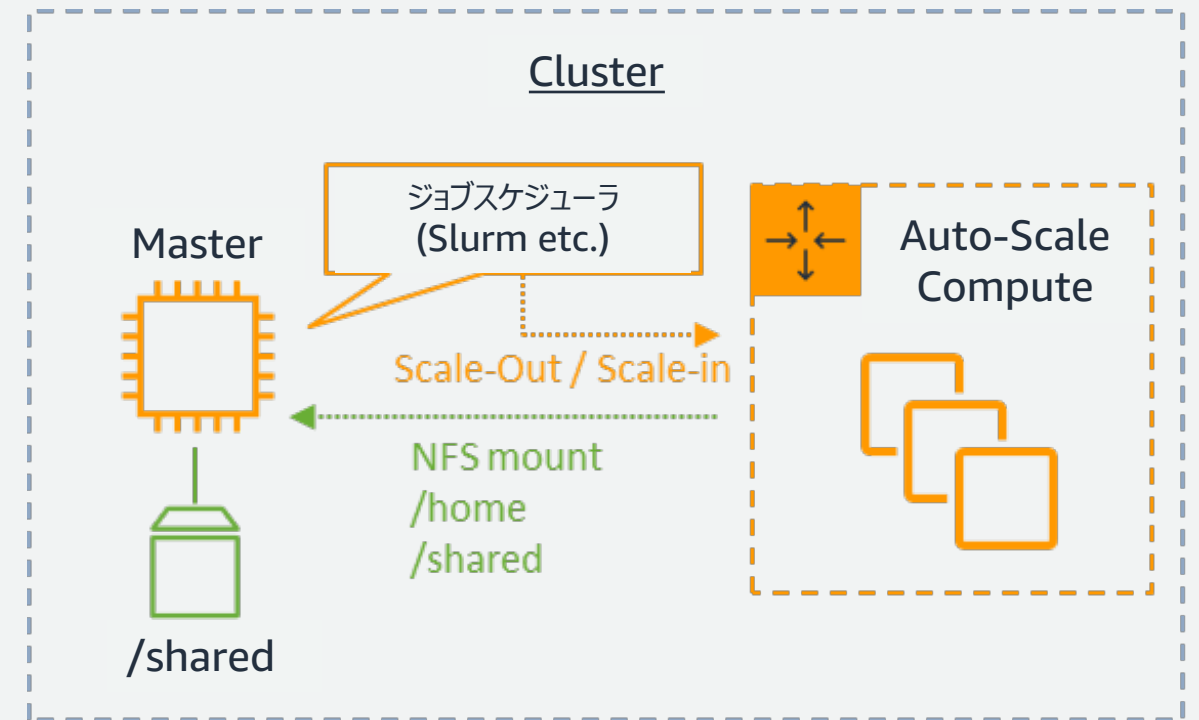


AWS ParallelCluster

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

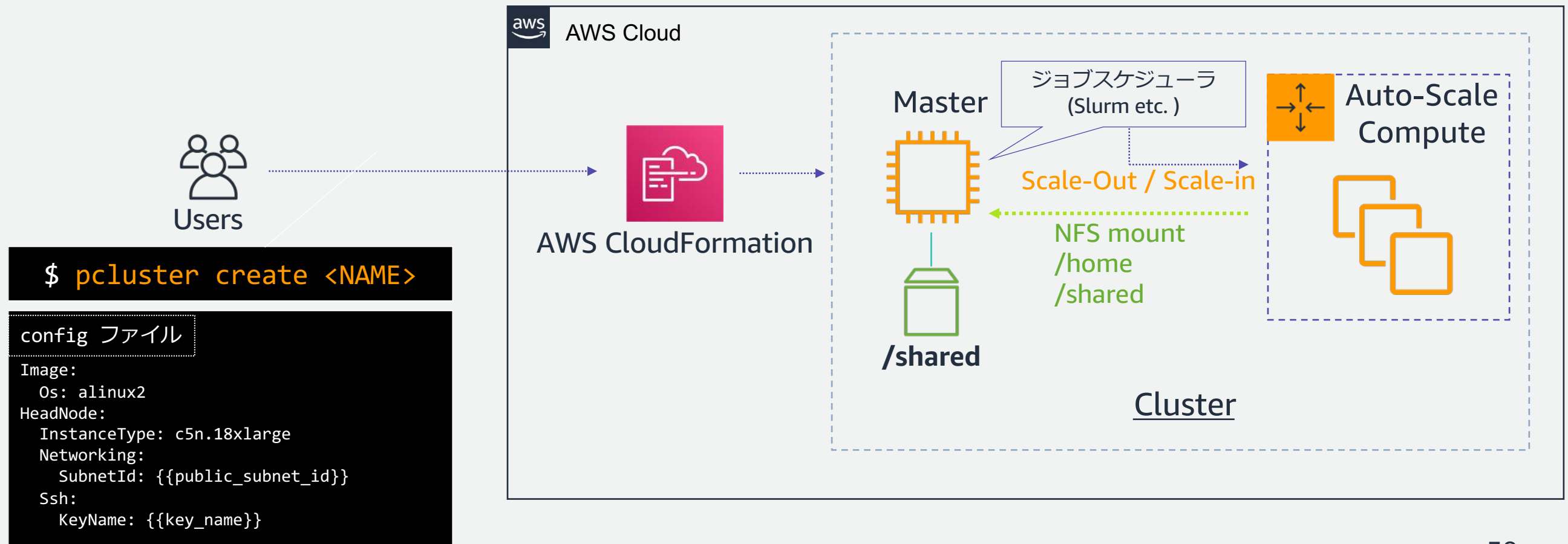
AWS ParallelCluster の特徴

- 既存の HPC 向けジョブスケジューラと Auto-Scaling を連携した環境を作成
 - Slurm に対応
- 少しいのコマンド操作でクラスタ作成可能
- MPI/NCCL 環境がセットアップ済みで、すぐに利用可能
- OS やネットワーク環境、ストレージ構成などを柔軟にカスタマイズ可能
- オープンソースプロジェクトであり、誰でもソースコードを入手可能: <https://github.com/aws/aws-parallelcluster>



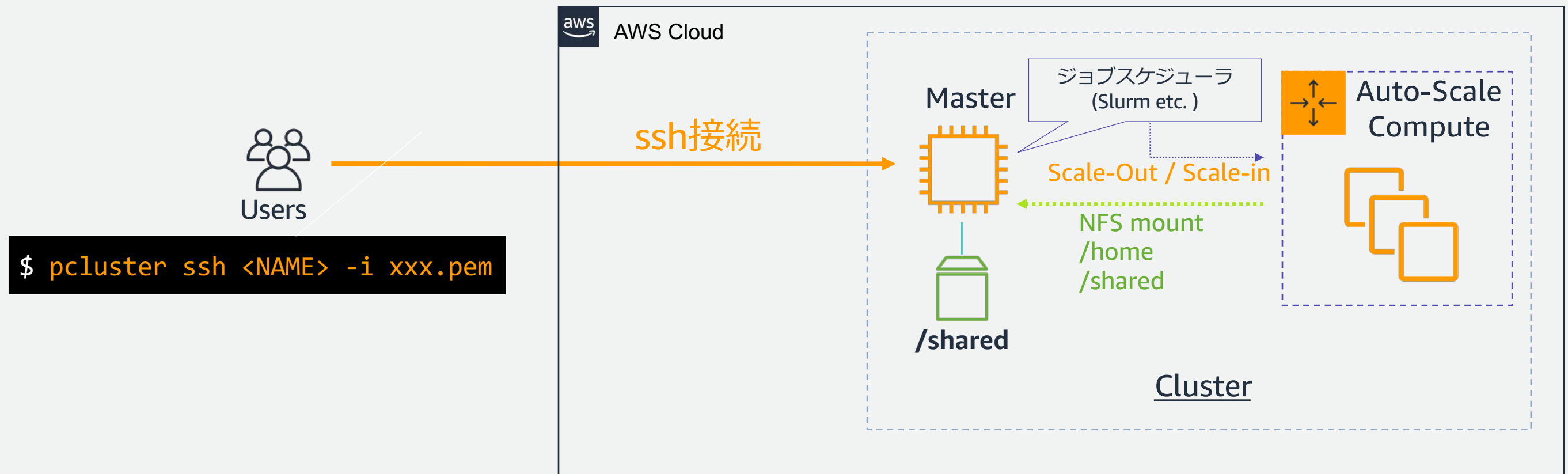
AWS ParallelCluster のご利用イメージ – クラスタ作成

- ローカル PC 等に ParallelCluster ソフトウェアをインストール
- config ファイルを記述し、pcluster create コマンドを実行することでジョブ投入に応じて Auto-Scale するクラスタ環境が自動的に作成される



AWS ParallelCluster のご利用イメージ – 接続

- pcluster ssh コマンドを実行することで作成されたクラスタに接続

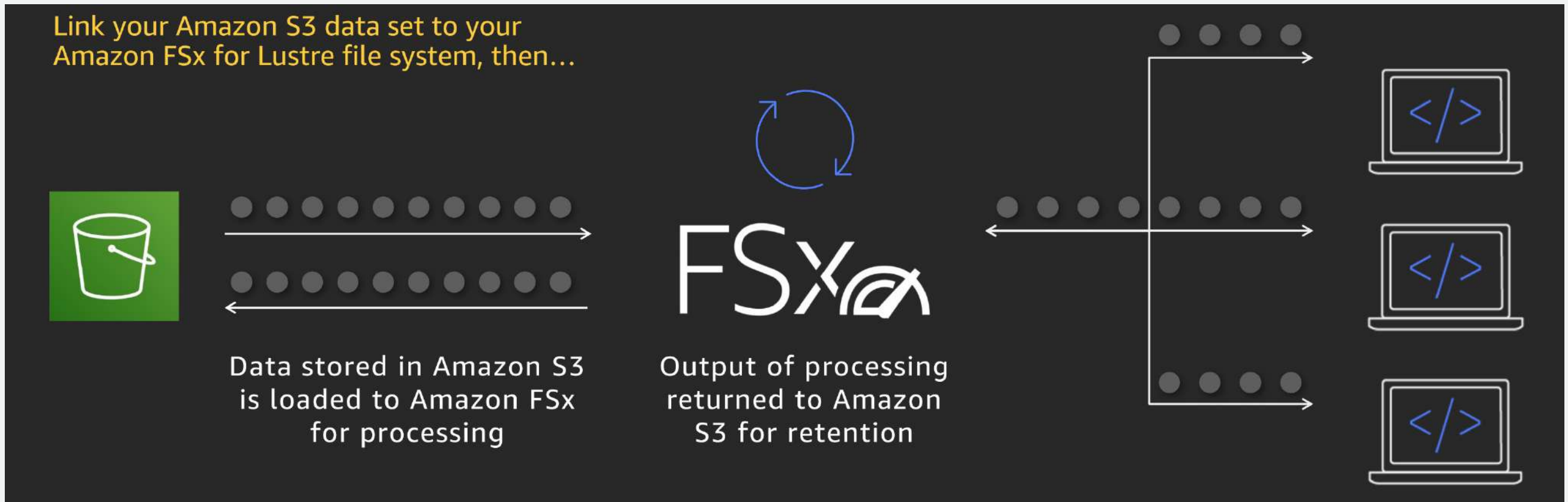


FSx

Amazon FSx for Lustre

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

- 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
- 高速な分散ファイルシステムである Lustre をフルマネージドで提供

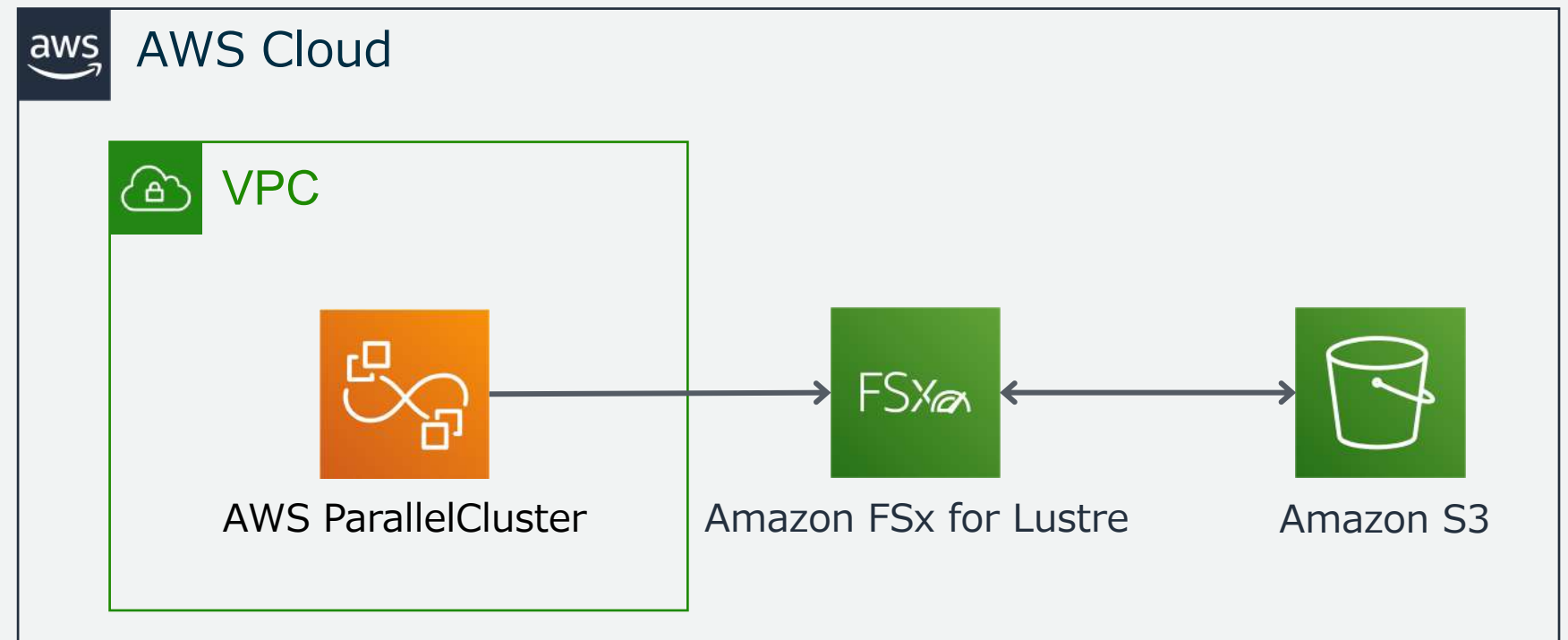


AWS ParallelClusterと統合されたAWSサービス

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

- AWS ParallelClusterのConfig ファイルに記述することで設定可能
- Amazon FSx for Lustre は Amazon S3 と連携可能
 - 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

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





AWS Cloud9

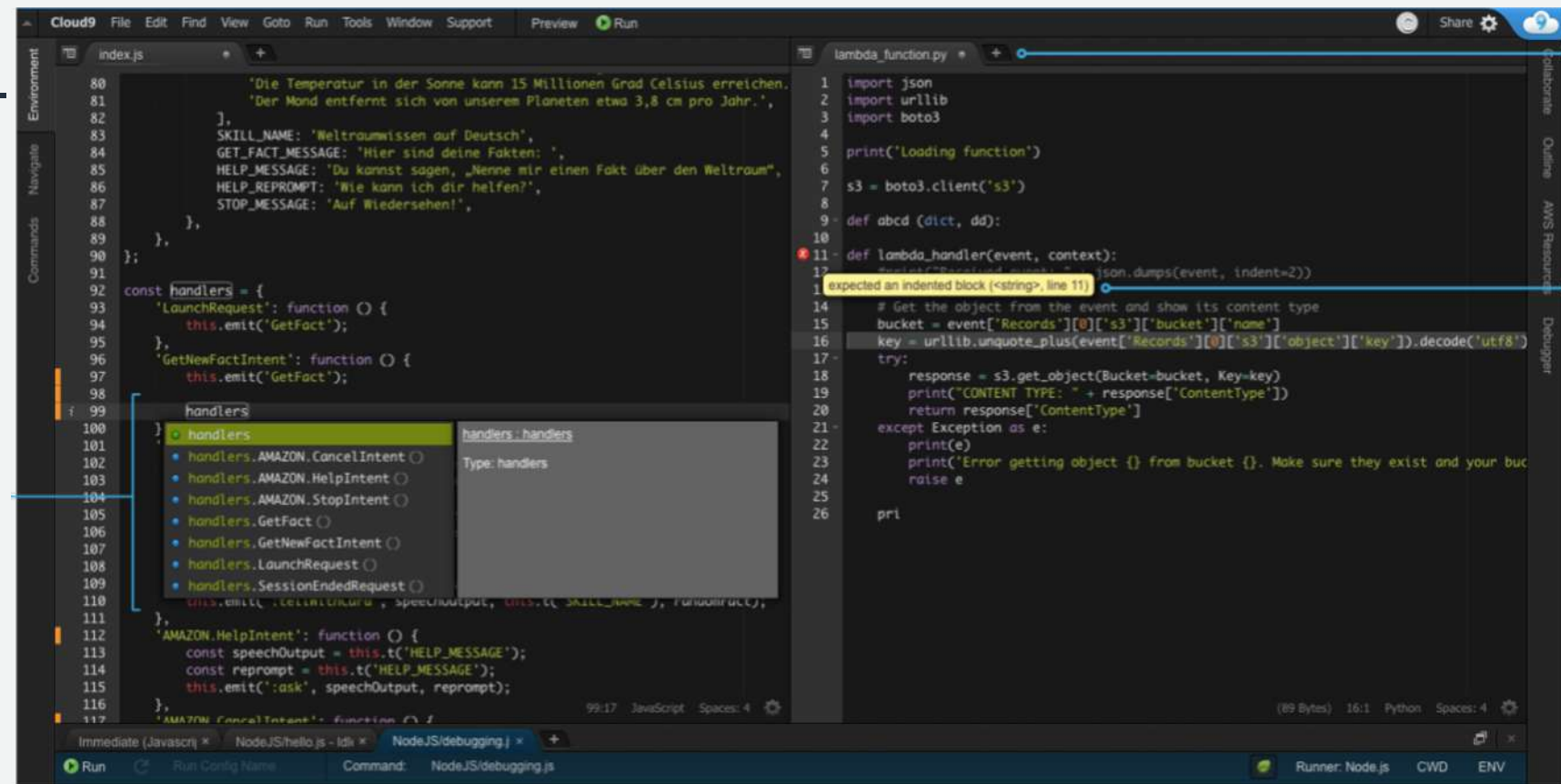
クラウドベースの統合開発環境 (IDE)

ブラウザだけでコードを記述、実行及びデバッグできる

JavaScript, Python, PHPなどのよく使用されるプログラミング言語の主要ツールを同梱

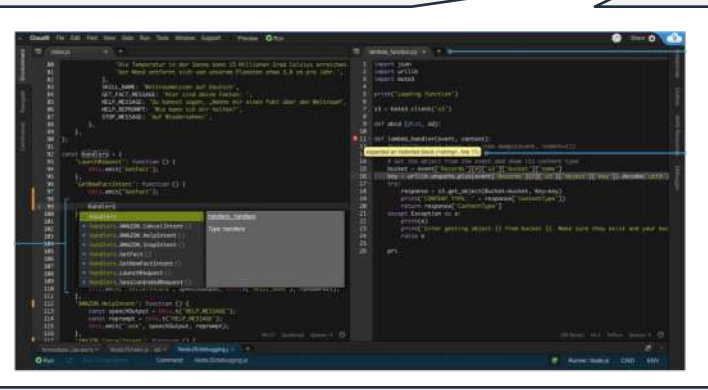
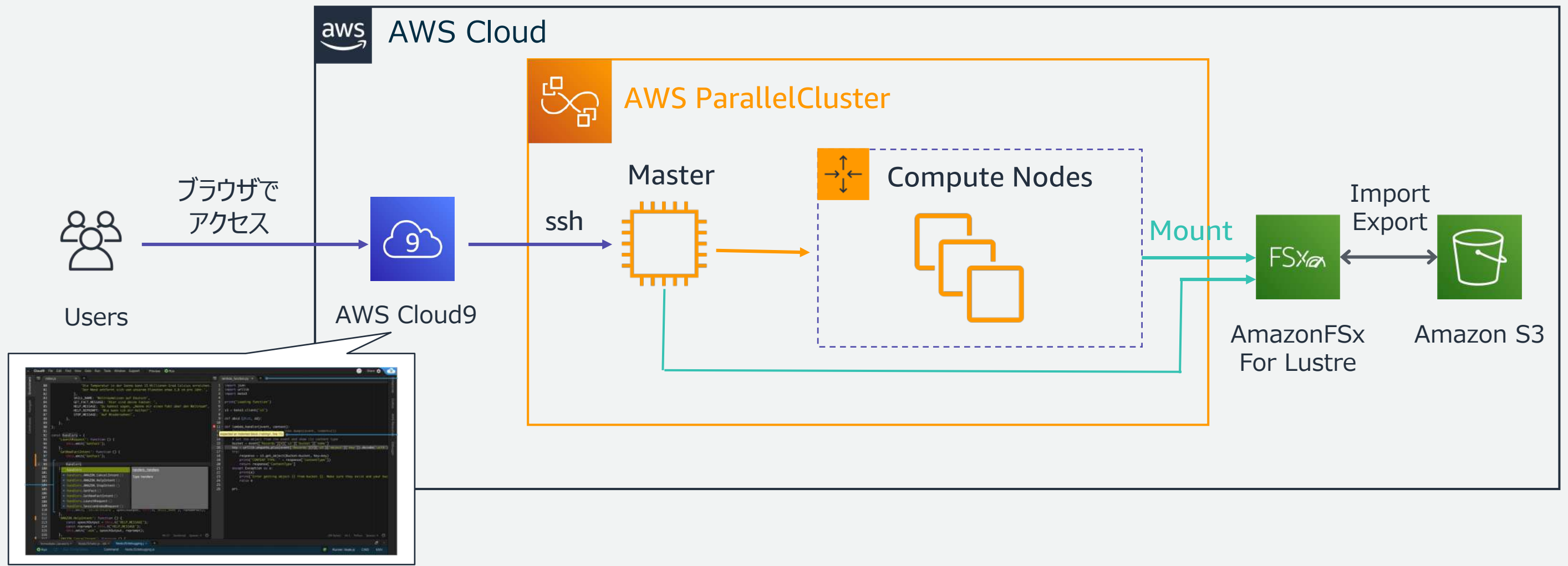
開発環境をチーム内で共有しペアプログラミング・コラボレーションが容易

EC2 インスタンスなど AWS サービスに直接ターミナルアクセス



AWS ParallelCluster への接続

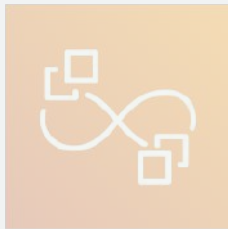
- AWS Cloud9 から AWS ParallelCluster のクラスタに接続、HPC ワークロードの実行



日本たばこ産業様ご活用サービスについて

セッションをご理解していただくために重要なサービスの概要をご説明

HPC 関連



AWS ParallelCluster
HPC クラスタ管理自動化



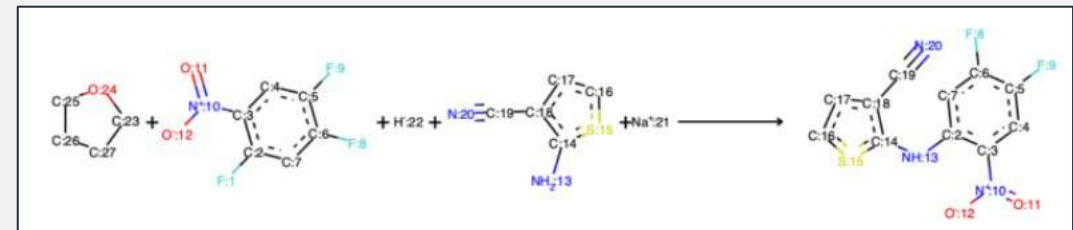
Amazon FSx for Lustre
分散ファイルシステム



AWS Cloud9
クラウドベースのIDE



AI/ML

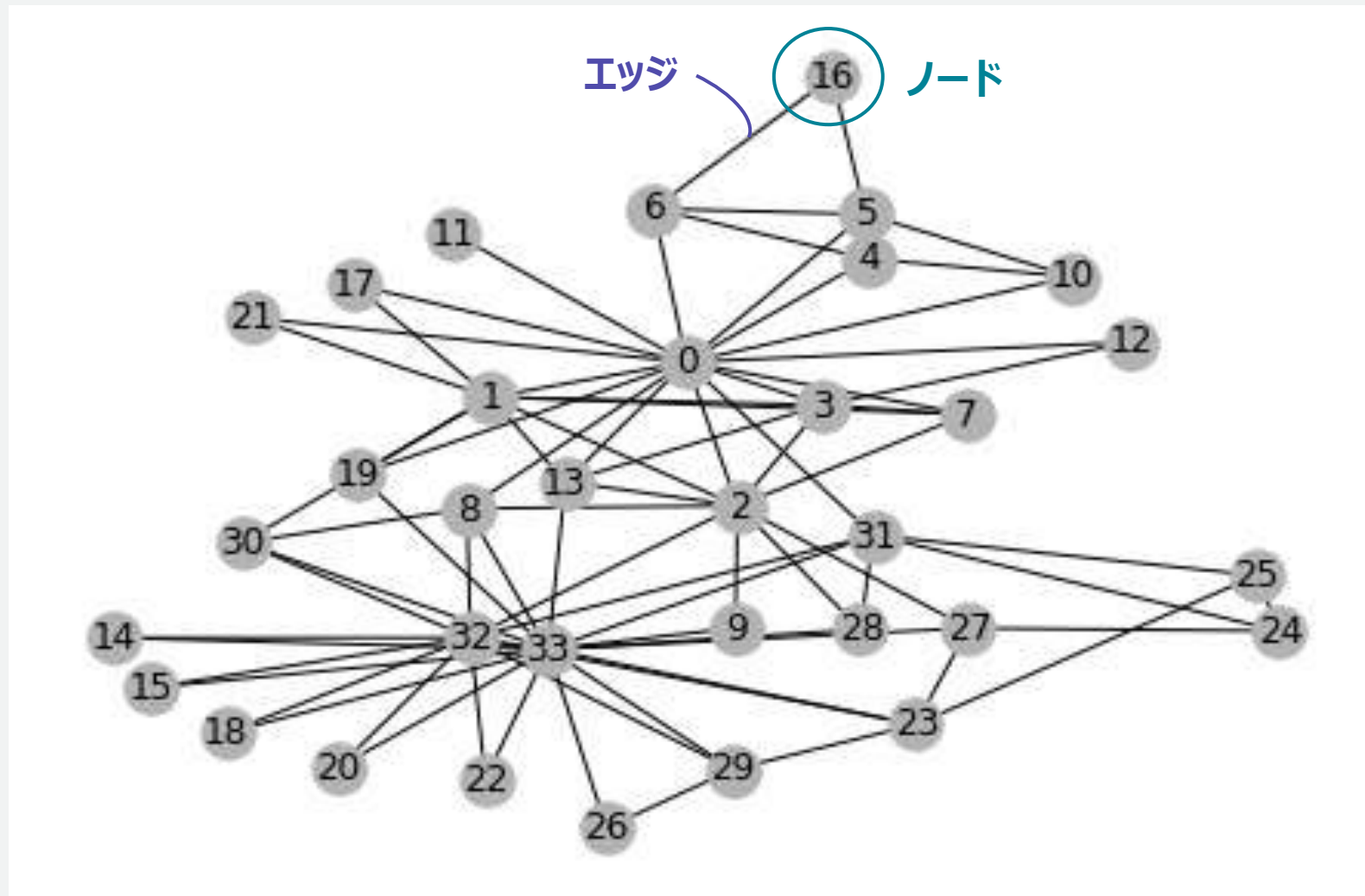


DGL-LifeSci
ライフサイエンス領域向けの
GNNライブラリ

https://github.com/aws-labs/dgl-lifesci/tree/master/examples/reaction_prediction/rexgen_direct

グラフデータ

ノード (頂点) とノード間を連結するエッジ (枝) で構成されるデータ構造



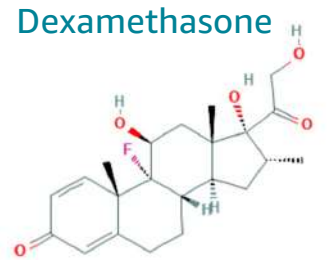
<https://julsimon.medium.com/a-primer-on-graph-neural-networks-with-amazon-neptune-and-the-deep-graph-library-5ce64984a276>

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

AWS におけるグラフデータに関連する研究例

AWS と米大学の共同研究で COVID-19 の治療を目的に Drug Repurposing Knowledge Graph (DRKG) を構築

How can I predict which drugs are good candidates for a disease?



amazon | science Latest news Research areas Conferences & events Publications

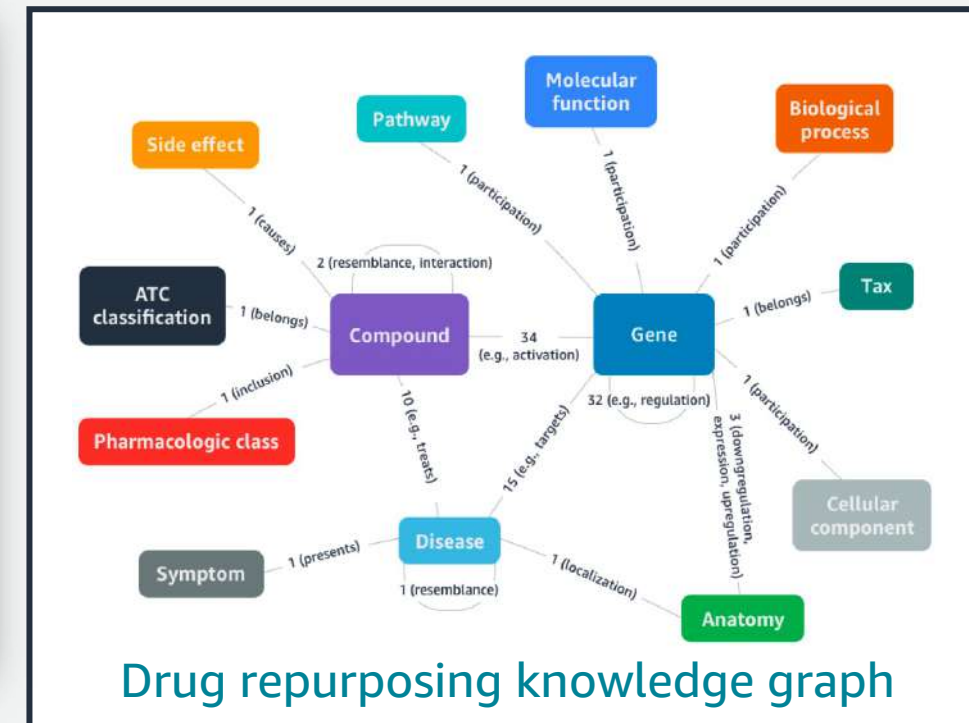
INFORMATION AND KNOWLEDGE MANAGEMENT

Amazon Web Services open-sources biological knowledge graph to fight COVID-19

Knowledge graph combines data from six public databases, includes machine learning tools.

By Cyrus Vahid
June 02, 2020

Drug name	Score	Ranking in top-100
Ribavirin	-0.21	0
Dexamethasone	-1.00	4
Colchicine	-1.08	8
Methylprednisolone	-1.16	16
Oseltamivir	-1.39	49
Deferoxamine	-1.51	87



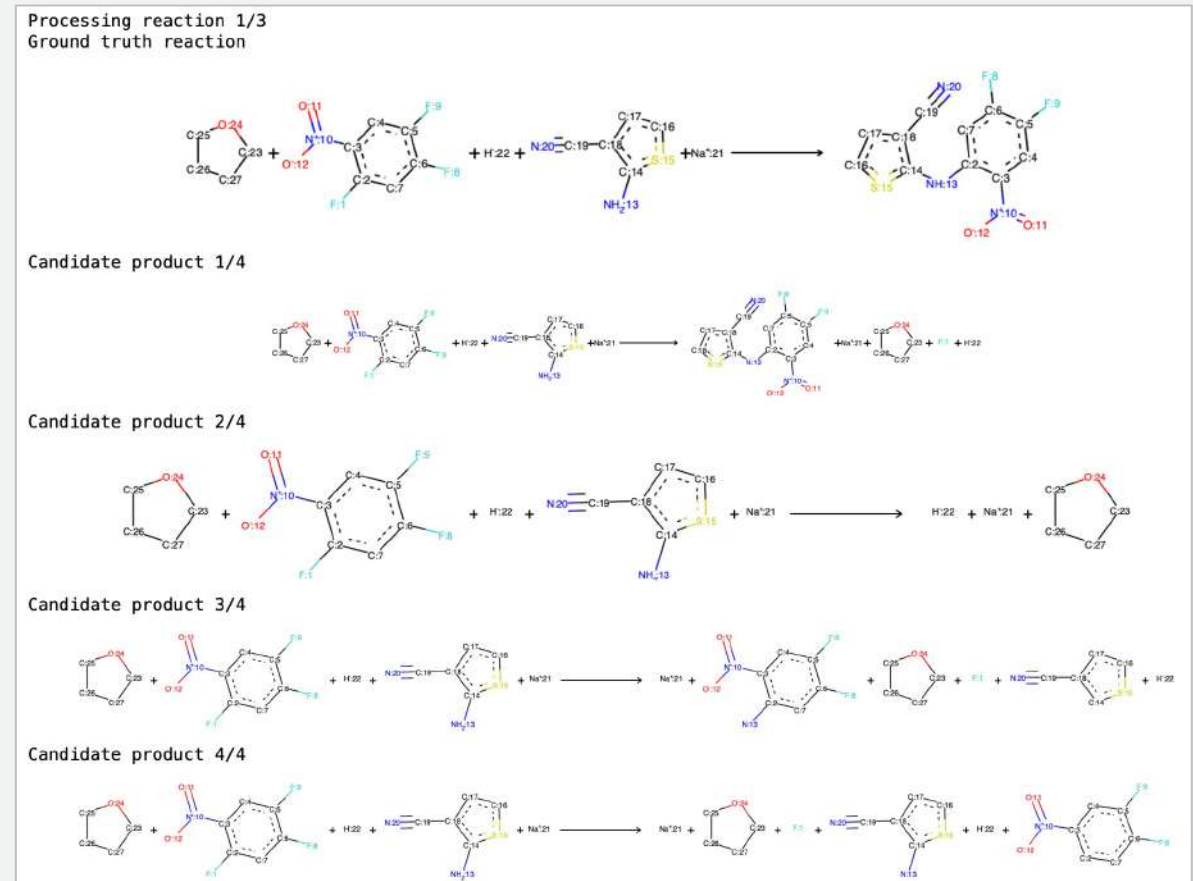
<https://github.com/gnn4dr/DRKG>
<https://arxiv.org/abs/2007.10261>



AWS ML Summit 2021 | Deep Graph Library: Deep Graph learning at scale: <https://www.youtube.com/watch?v=VmQkLro6UWo>

DGL-LifeSci

- ライフサイエンス領域でGNN（Graph Neural Network）を活用した様々なユースケースを実現するためのOSS
 - <https://github.com/awslabs/dgl-lifesci>
- ユースケース例
 - 分子生成
 - 分子物性予測
 - タンパク質-リガンド結合親和性予測
 - 反応予測
- 論文等で提案されている様々なアルゴリズム、モデルが実装されている



https://github.com/awslabs/dgl-lifesci/tree/master/examples/reaction_prediction/rexgen_direct



本セッションのまとめ

まとめ

- 最新ソリューションを交えた創薬領域で活用されている AWS サービスのご紹介
 - HPC
 - AI/ML
 - スマートラボ
- JT様ご活用のサービスの概要理解
 - AWS ParallelCluster, DGL-LifeSci, etc.
 - **具体的な使い方についてはJT様のセッションでご紹介いただきます**

最後に

- AWS は、今回紹介したサービス以外にも、創薬の様々な場面でご活用いただける多数のサービスを提供しております。
- また、お客様の構築サポートやアーキテクチャレビュー、ハンズオンなどの技術支援を行なっています。
- 今回紹介した AWS サービスのご利用や、AWSを利用した環境の構築をご検討いただく際は是非ご相談ください。

お問い合わせはこちらから ↓

<https://aws.amazon.com/jp/contact-us/sales-support/>



Thank you!