

# 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 2022

JTの創薬研究におけるAWSの活用事例

- 自己紹介・会社紹介
- 創薬研究における計算機化学の利用について
- 計算機化学におけるAWSの活用について
  - ◆ Virtual Screening : CPU
  - ◆ Machine Learning : GPU
- 展望
- まとめ

- 自己紹介・会社紹介
- 創薬研究における計算機化学の利用について
- 計算機化学におけるAWSの活用について
  - ◆ Virtual Screening : CPU
  - ◆ Machine Learning : GPU
- 展望
- まとめ

# 自己紹介



篠原 妙子 (しのはら たえこ)

[taeko.shinohara@jt.com](mailto:taeko.shinohara@jt.com)

岡田 晃季 (おかだ あきとし)

[akitoshi.okada@jt.com](mailto:akitoshi.okada@jt.com)

医薬総合研究所 化学研究所

計算機化学業務

- Virtual Screening
- MM/MD/QM/MLを用いた化合物の  
各種パラメータ予測と提案
- 新規CADD技術 開拓・開発・実装・運用





# 会社概要

## Corporate Profile

ひと  
ときを、  
想う。 JT

## 事業展開

お客様に信頼される「JTならではのブランド」を生み出し、育て、高め続けていくために……

### たばこ事業 Tobacco Business



### 医薬事業 Pharmaceutical Business



### 加工食品事業 Processed Food Business



## 病気から世界の人々を救う「オリジナル新薬の創出」に挑み続けます

### 画期的な新薬創出のための研究開発

- 医薬総合研究所を中心に、主に「循環器・代謝」「免疫・炎症」「中枢」の領域で研究開発を推進
- 自社研究開発力の充実・強化
- 国内外大手製薬企業との連携



横浜リサーチセンター（神奈川県横浜市）



### 鳥居薬品（株）との協業

- JTが研究開発機能を、鳥居薬品（株）が製造・販売およびプロモーション機能を担う

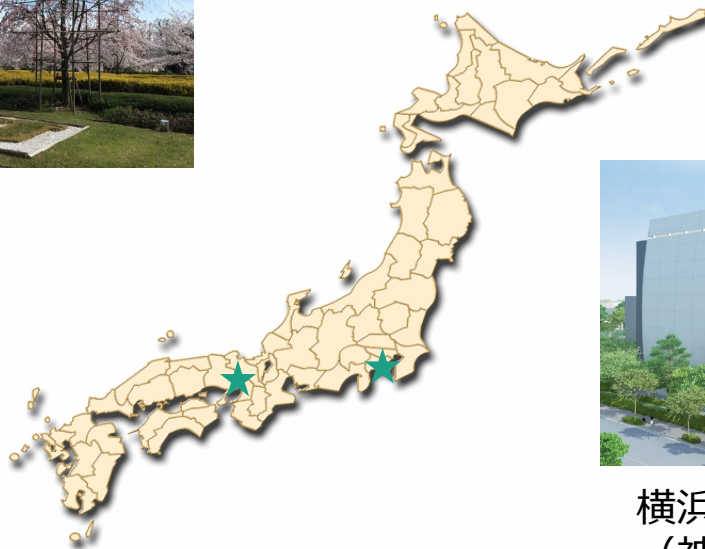


高リン血症治療剤「リオン錠250mg」（2014年5月発売）



医薬総合研究所  
(大阪府高槻市)

- ・ 化学研究所
- ・ 生物研究所
- ・ 薬物動態研究所
- ・ 生産技術研究所



横浜リサーチセンター  
(神奈川県横浜市)

- ・ 安全性研究所
- ・ 医薬探索研究所

- 自己紹介・会社紹介
- 創薬研究における計算機化学の利用について
- 計算機化学におけるAWSの活用について
  - ◆ Virtual Screening : CPU
  - ◆ Machine Learning : GPU
- 展望
- まとめ

どの段階でも計算(機)科学活用

★上市★



臨床試験

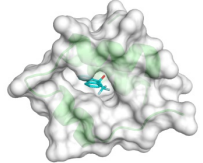


前臨床試験



○ 医薬品候補化合物の最適化

○ 医薬品候補化合物  
(ヒット・リード化合物) の探索



疾患・ターゲット探索



本日は青枠の計算(機)化学にフォーカス

# Hit to Leadにおける計算機化学利用例



## ■ Hit

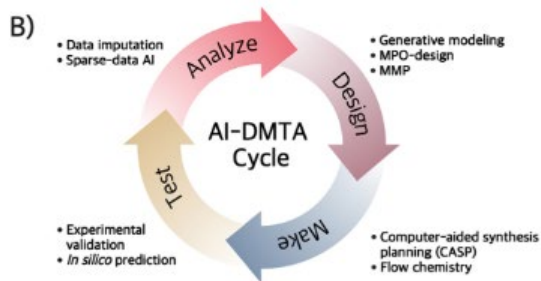
- ◆ 2D/3D Similarity Search
- ◆ (Giga) Docking

## ■ Optimization (to Lead)

- ◆ 活性予測：分子動力学計算，機械学習
- ◆ 物性予測：機械学習
- ◆ 化合物発生：SMILESベース機械学習，反応ベース発生

# 例：創薬における機械学習

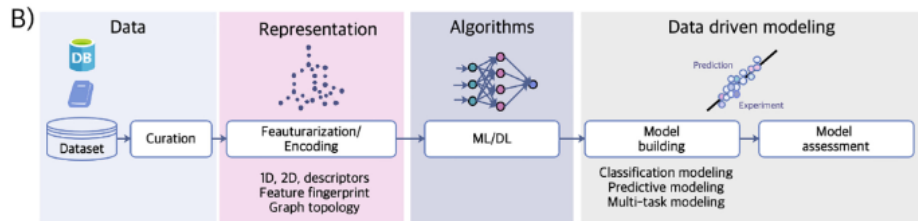
## DMTAサイクルの効率化



## 様々な物性の予測

A)

Absorption	Distribution	Metabolism	Excretion	Toxicity	PhysChem
Caco-2 permeability	Human serum albumin binding	Microsomal stability	CL <sub>int</sub>	Ames mutagenicity	Aq. solubility
Caco-2 efflux ratio	Plasma-protein binding	Hepatocyte stability	Terminal half-life	CYP inhibition	LogD
PAMPA permeability	Fraction unbound (fu)	Site of metabolism		Phaspholipidosis	Membrane affinity
Blood-brain barrier penetration	Volume of distribution (vdss)			hERG inhibition	pKa
	Brain/plasma ratio (Kp)			Drug-induced liver injury (DILI)	



Drug Discovery Today

## Enhancing preclinical drug discovery with artificial intelligence

R.S.K. Vijayan<sup>a</sup>, Jan Kihlberg<sup>b</sup>, Jason B. Cross<sup>a,\*</sup>, Vasanthanathan Poongavanam<sup>b,\*</sup>

<sup>a</sup>Institute for Applied Cancer Science, MD Anderson Cancer Center, Houston, TX, USA

<sup>b</sup>Department of Chemistry BMC, Uppsala University, Uppsala, Sweden

Drug Discov Today.

2021 Nov 25;S1359-6446(21)00504-3.



- ✦ 膨大化するケミカルスペース
- ✦ 予測精度 対 計算時間
- ✦ 予測モデル開発検討における計算量

計算時間が  
ボトルネックに



クラウドサービスによる高速化

## ■ 幅広いサービス



Amazon EC2 (スケーラブルなコンピューティング)



Amazon SageMaker (機械学習に特化したフレームワークや  
スケーラブルなコンピューティング環境など)



AWS ParallelCluster (容易なHPCクラスター管理)



AWS Cost Explorer (タグ付けによる計算コスト管理, 予算アラート)

## ■ サポート



導入のご相談・ご支援



AWS Business support (チャットによるリアルタイム相談)

## ■ セキュリティ



AWS Trusted advisor (様々な経験的推奨項目を提案)



AWS Well-architected tool (MFAやcloud trail機能などの紹介と  
サポーターとのチェック)

- 自己紹介・会社紹介
- 創薬研究における計算機化学の利用について
- 計算機化学におけるAWSの活用について
  - ◆ Virtual Screening : CPU
  - ◆ Machine Learning : GPU
- 展望
- まとめ

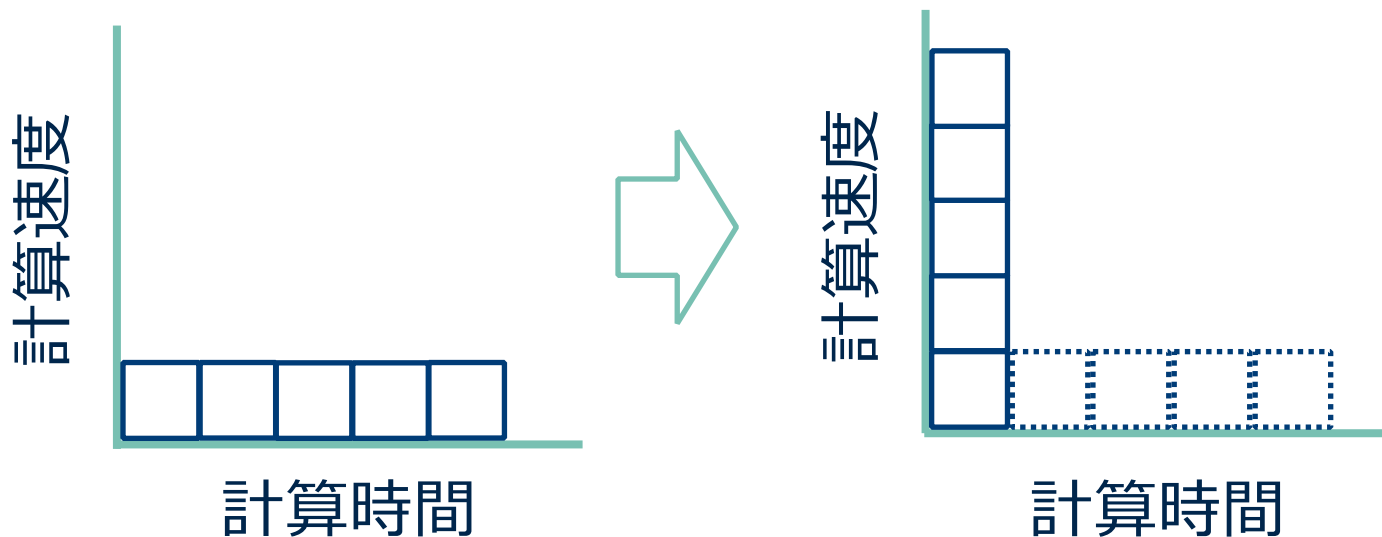
# クラウドプラットフォームによる計算効率化



コスト：調達費用・計画・期間，メンテナンス費用，**時間**

コンピューティング

- ◆ CPU：
- ◆ GPU：**必要なリソースを必要な時**に利用



# クラウドプラットフォームによる計算効率化



コスト：**調達費用・計画・期間**, **メンテナンス費用**, 時間

調達

費用：従量課金性

計画：フレキシブルに対応

期間：当日から

メンテナンス：無



**導入が容易**

## ✦ Hit Finding (VS)

- ◆ 構造発生
- ◆ 類似構造検索
- ◆ ドッキング

## ✦ Optimization

- ◆ 活性予測
- ◆ 物性予測

これらの比較的高コストな計算の効率化

## 其 CPU

- ◆ OMEGA 構造発生
- ◆ ROCS 類似構造検索
- ◆ Psi4 量子化学計算
- ◆ AutoDOCK ドッキング
- ◆ KNIME ワークフロー

## 其 GPU

- ◆ GROMACS 分子動力学計算
- ◆ AMBER 分子動力学計算
- ◆ fastROCS 類似構造検索
- ◆ Python 物性予測(機械学習)
  - ◆ DeepChem
  - ◆ DGL LifeSci
  - ◆ PyTorch

**前スライドの実施事項にツールが合致**

(サイトライセンスやOSSだと拡張が容易)

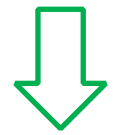
- 自己紹介・会社紹介
- 創薬研究における計算機化学の利用について
- 計算機化学におけるAWSの活用について
  - ◆ Virtual Screening : CPU, Storage
  - ◆ Machine Learning : GPU
- 展望
- まとめ

# Virtual Screening

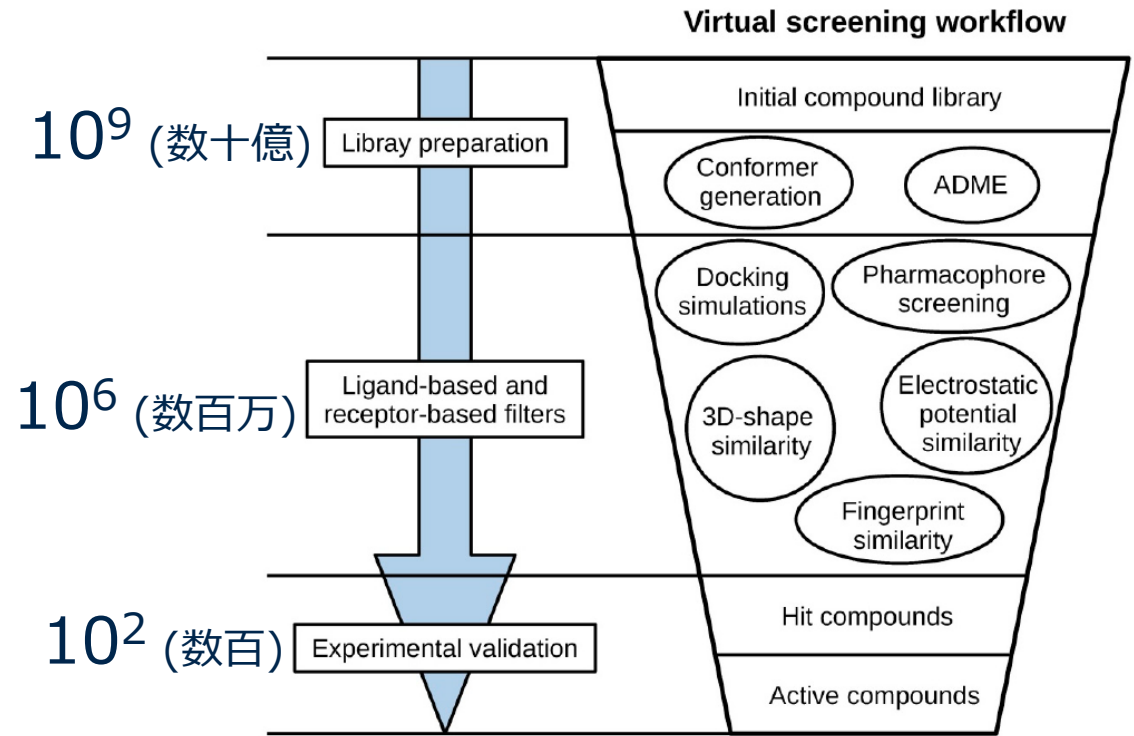


■ 数十億の化合物群から  
活性・物性が見込める  
化合物群の絞り込み

■ 上流の計算時間が膨大



AWSによる高速化

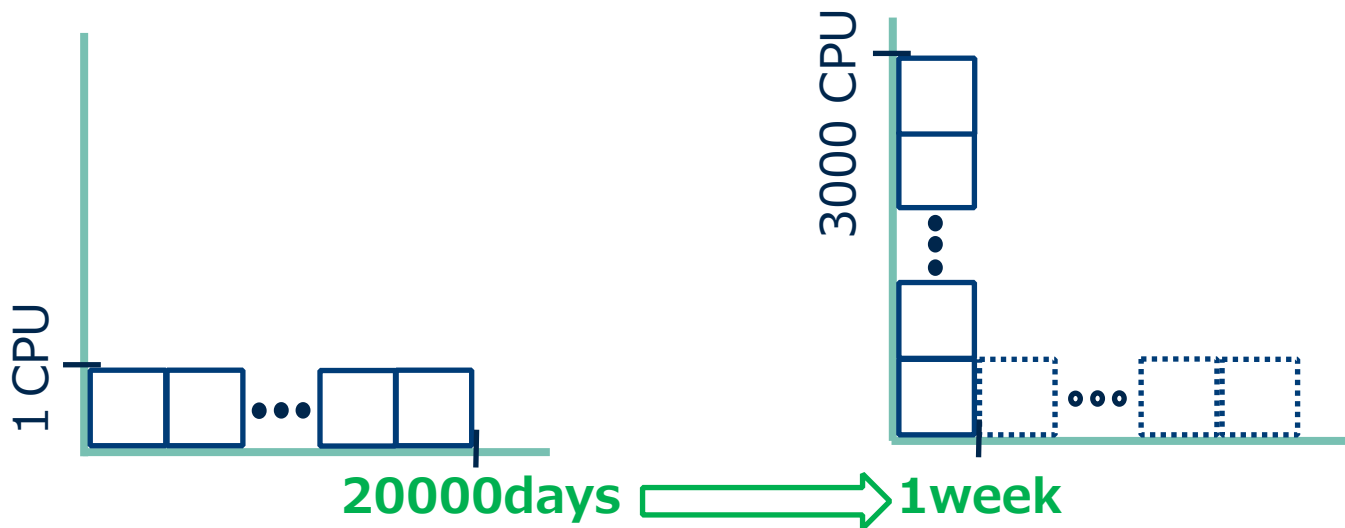


# オンプレマシンによるOMEGA事前調査

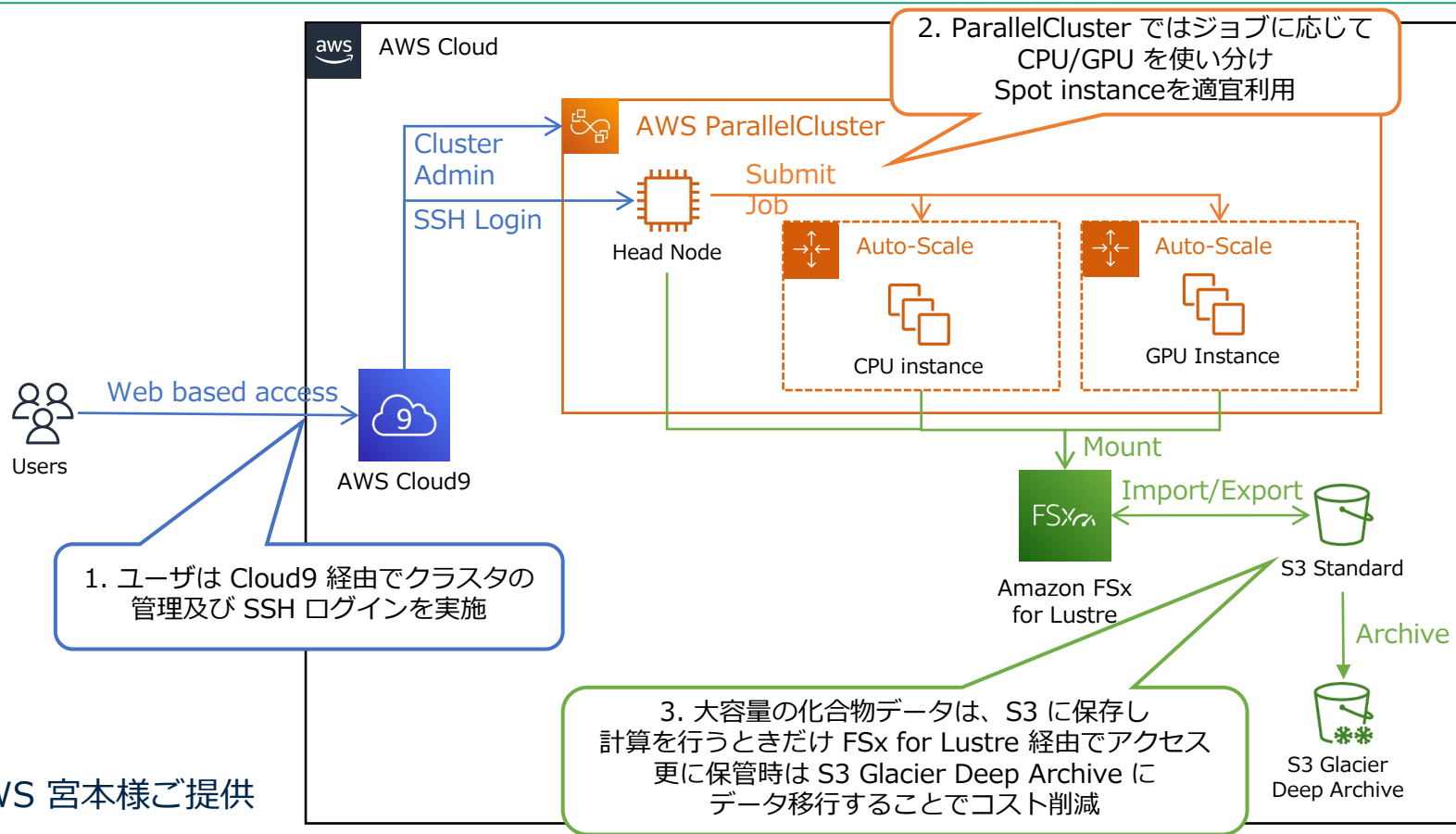


## 50億化合物の3次元構造化

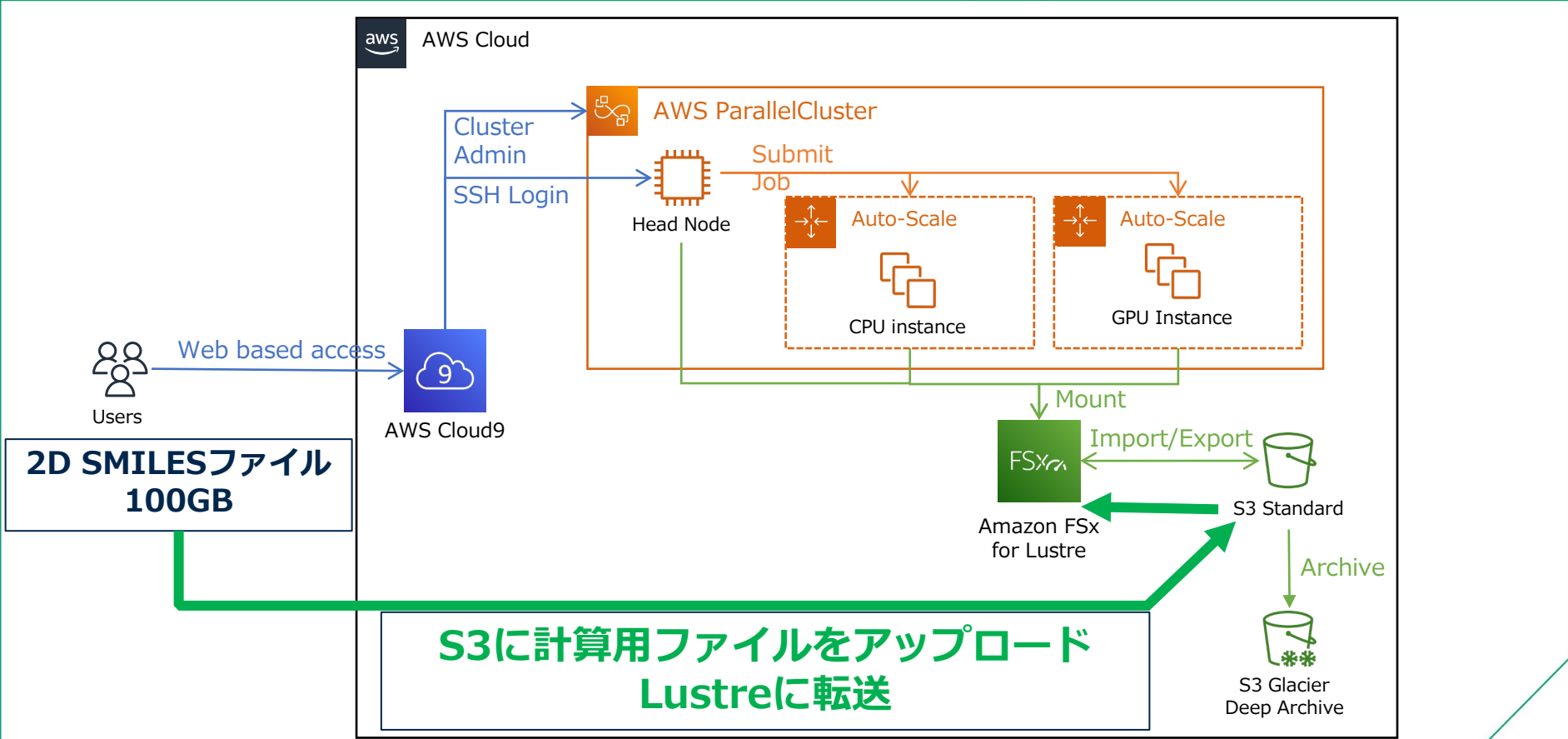
- $1.8 \times 10^9$  seconds (約20000 days) / 1CPU  
∴ 約20000 CPU / day or 約3000 CPU / week
- 化合物ライブラリの増大により, 計算時間が増大
- 一時的に $10^2 \sim 10^4$ 規模のCPUを利用できる環境が必要



# HPCシステム構成



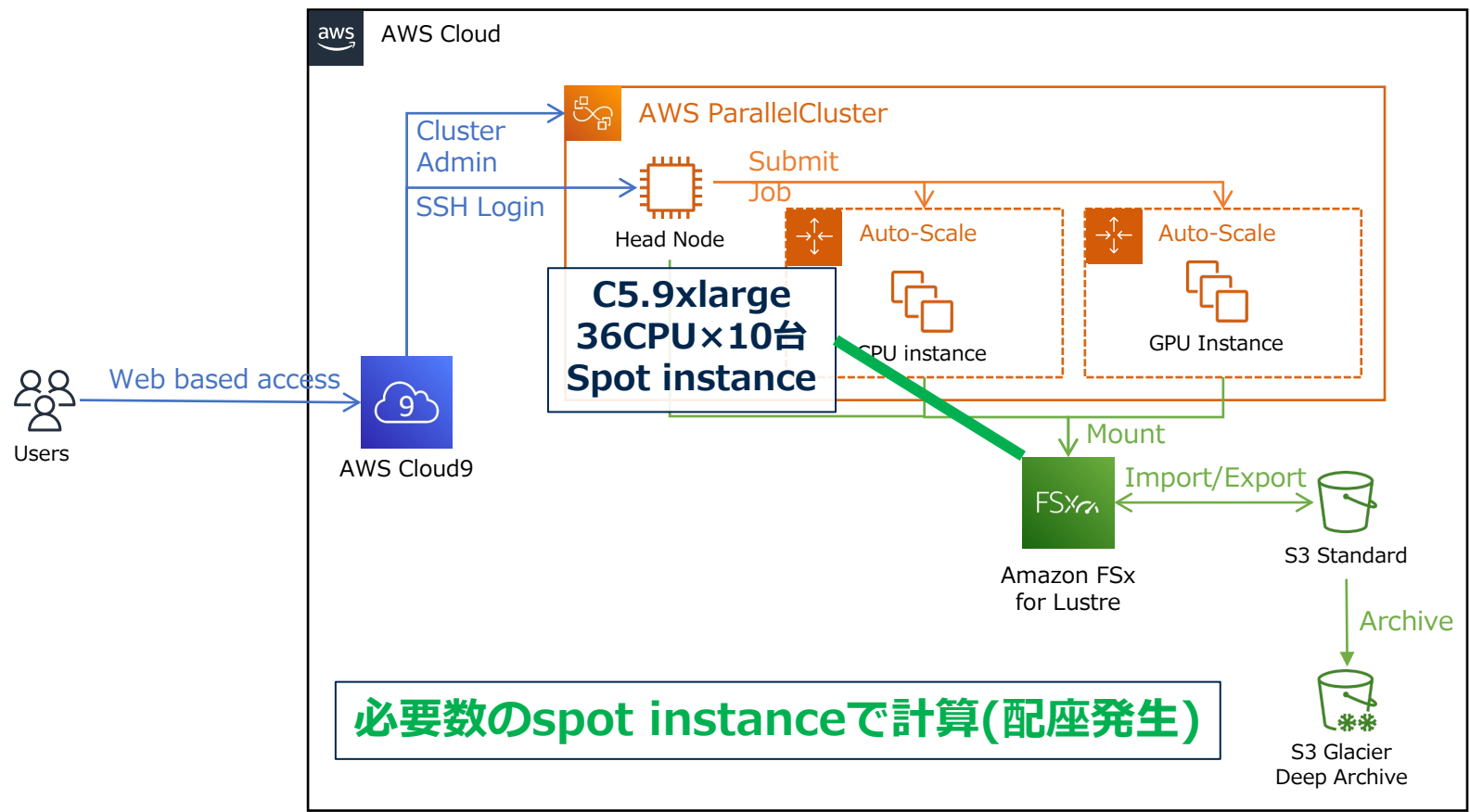
# 計算フロー



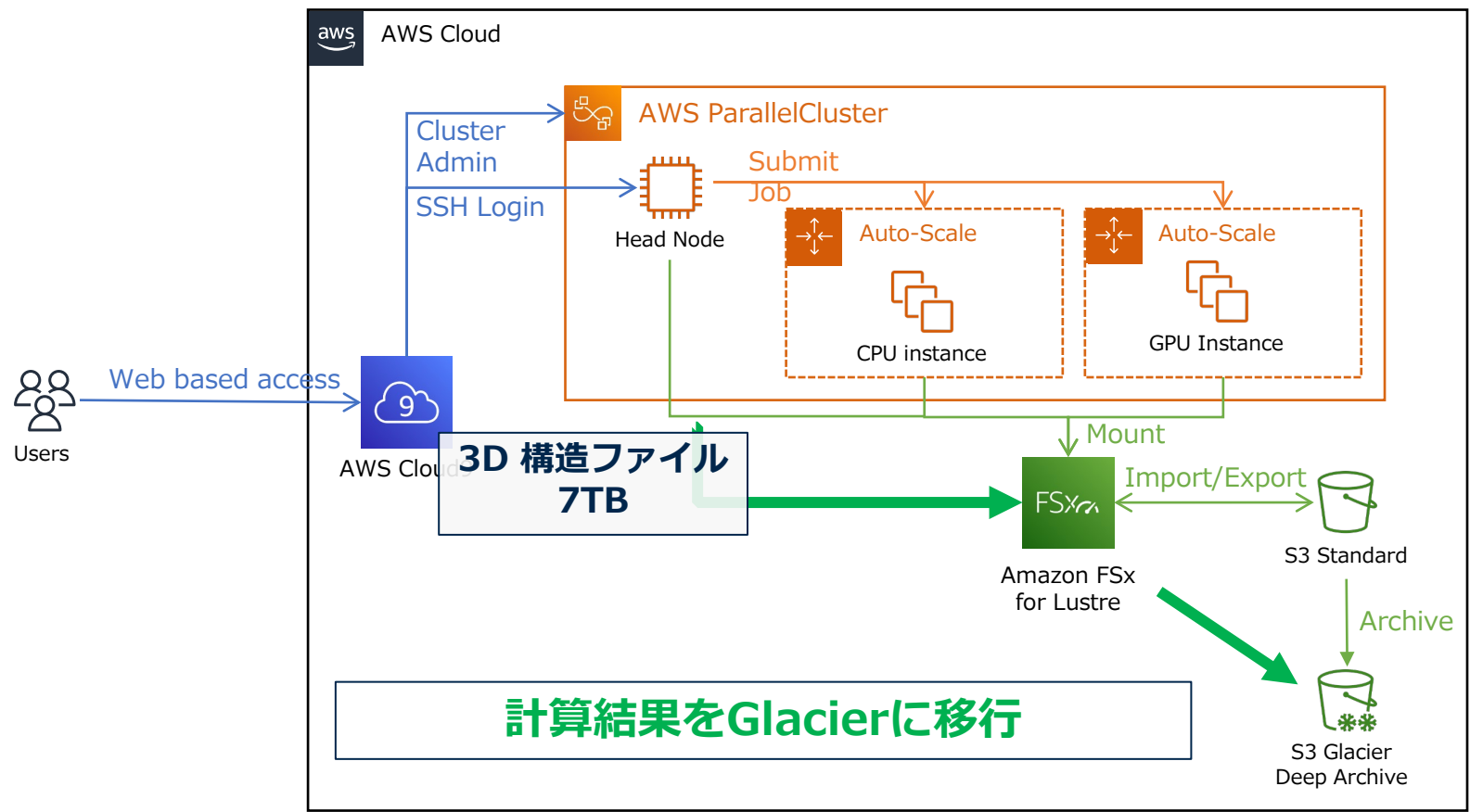
**S3に計算用ファイルをアップロード  
Lustreに転送**

**2D SMILESファイル  
100GB**

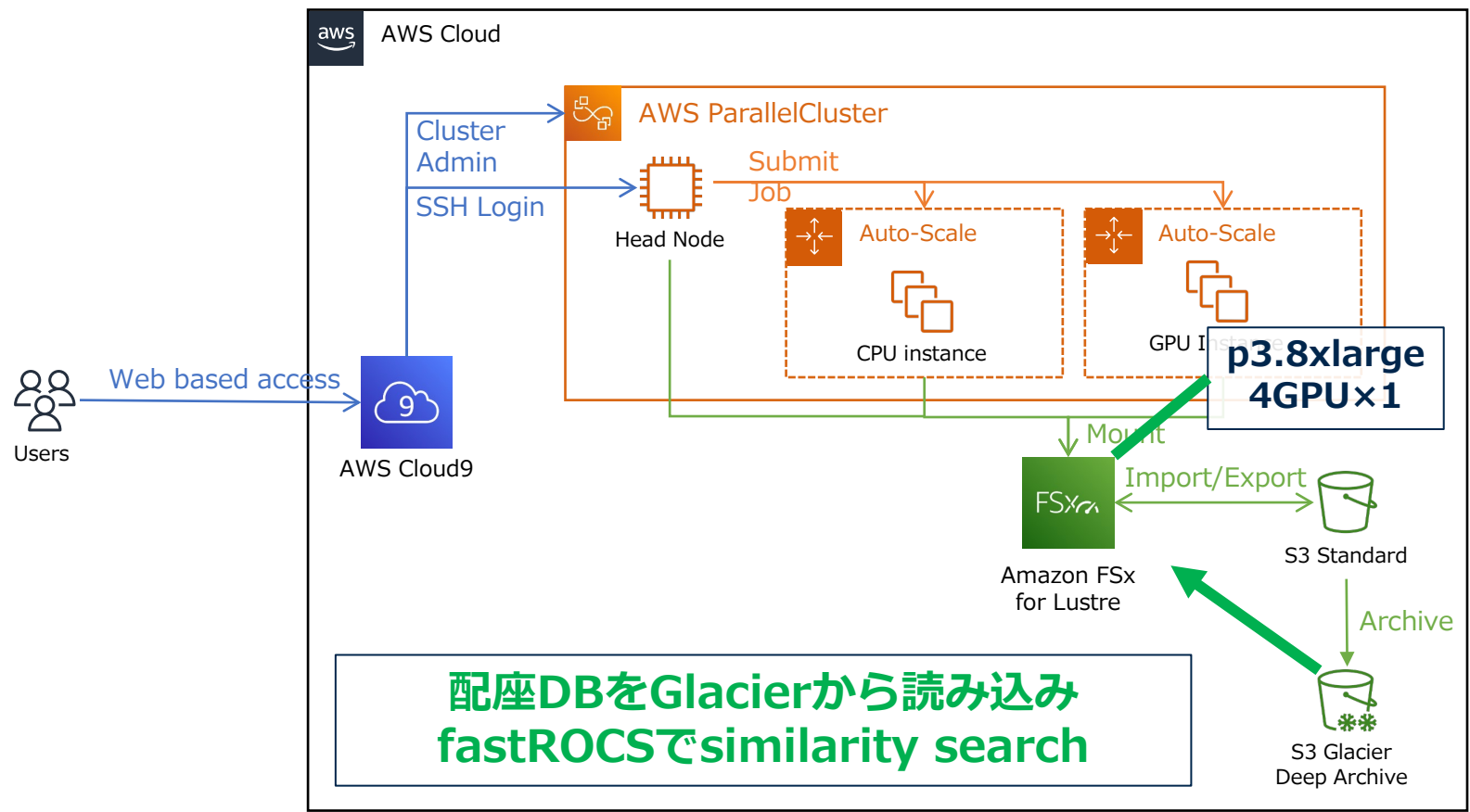
# 計算フロー



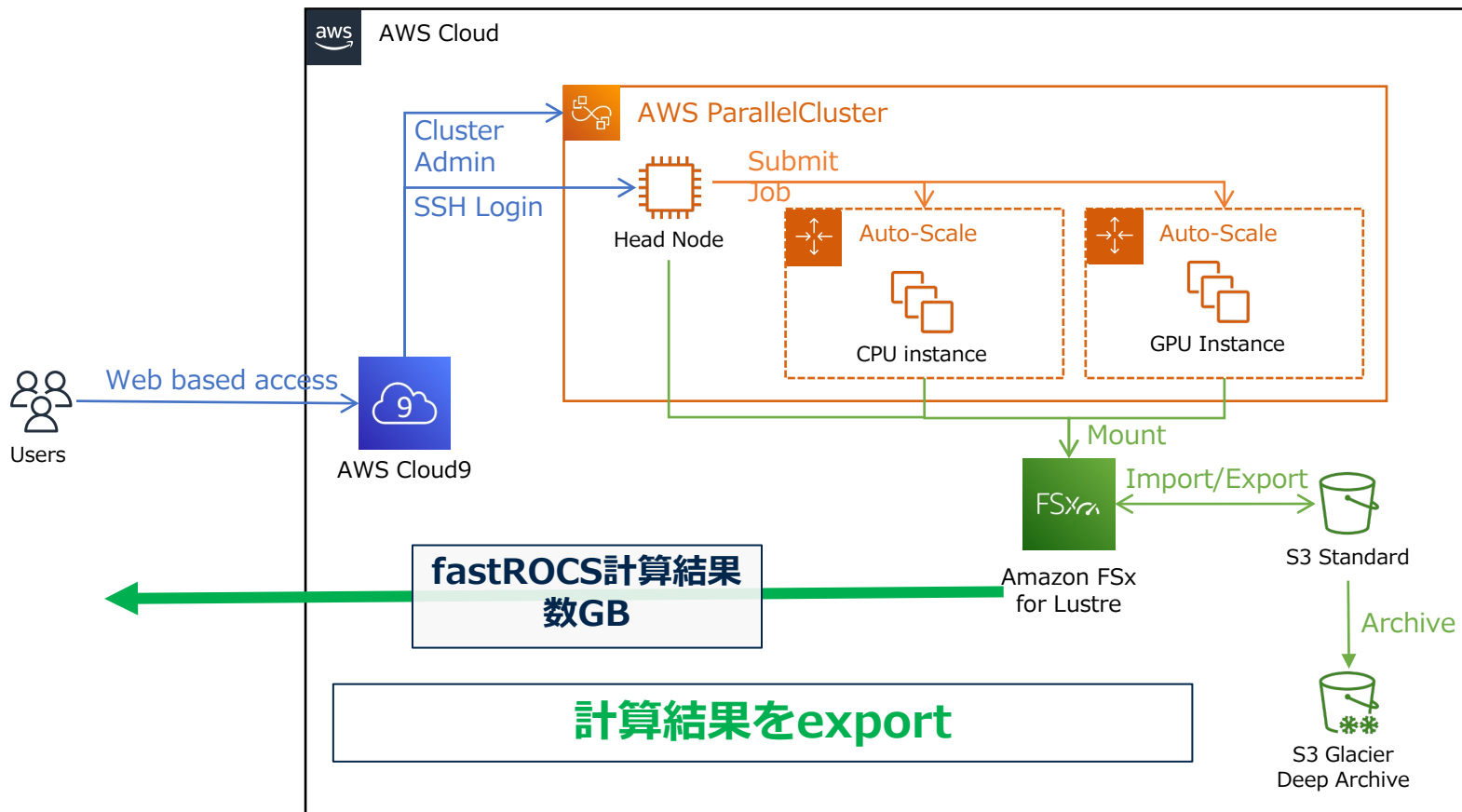
# 計算フロー



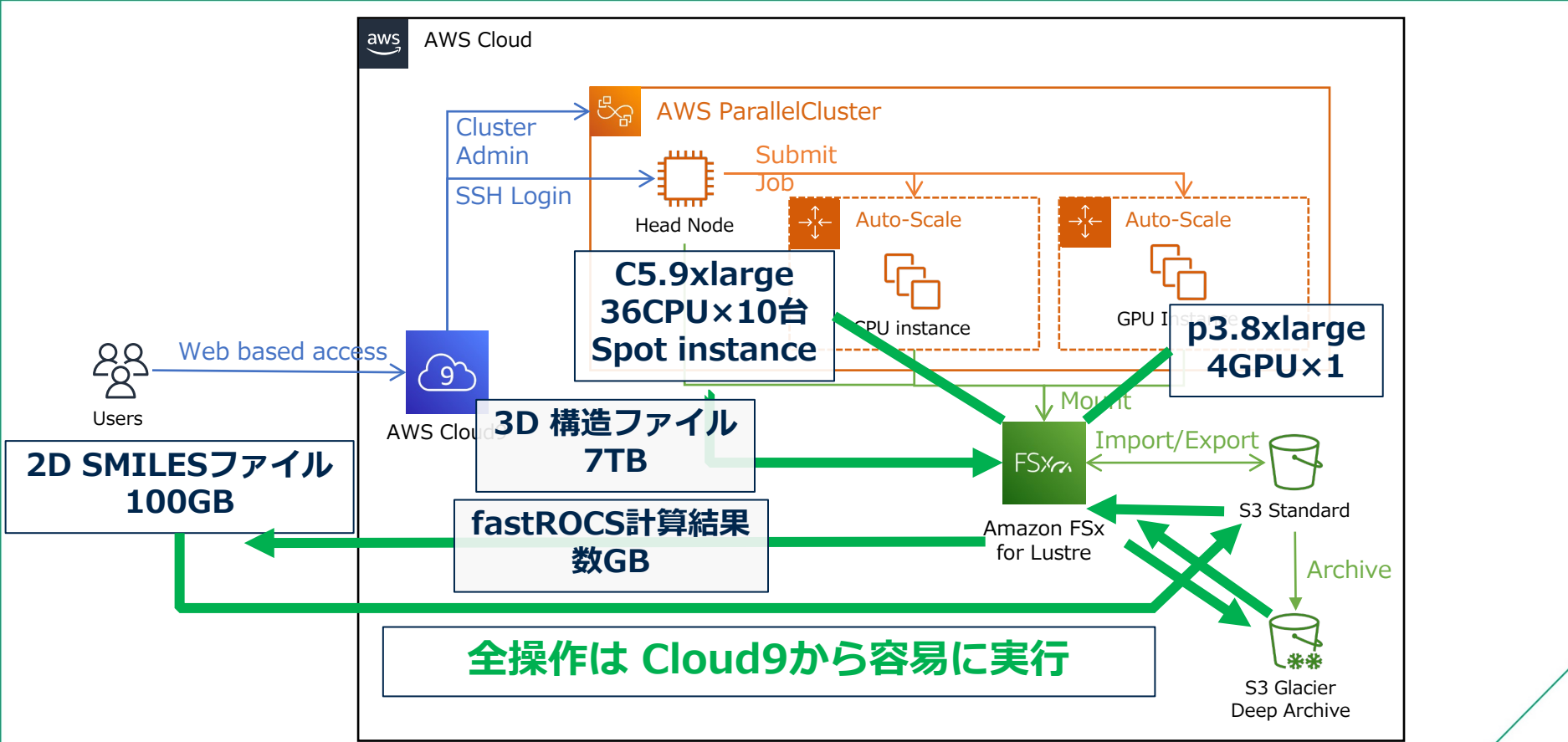
# 計算フロー



# 計算フロー



# 計算フロー



# AWSを利用したOMEGA計算結果



## コンピューティング

- 速度：中規模( $10^2$ )のCPU数を **spot instance**で確保し**半月**で実施※
- コスト：**約\$3000**
- サポート：ParallelClusterのhandson, Slurm利用のサポート

※本計算時は20億を計算対象とした

## ストレージ

- 安価：使用頻度に合わせた保存形態により**\$7/月**で**7TB**を保存
- コストシミュレーション：pricing calculatorにより容易に想定が可能

The logo for the AWS pricing calculator, featuring the word 'aws' in white lowercase letters with the Amazon smile logo underneath, followed by the text 'pricing calculator' in a lighter white font, all on a dark blue background.

aws pricing calculator

# クラウドサービス(ORION)との違い



	自社AWS環境	ORION
費用	○	△
フレキシビリティ	○	△
操作性	△	◎

## ⌘ 費用

- ◆ Pros: オンプレ構築により計算自体の費用は微削減
- ◆ Cons: 構築にかかるリソース

## ⌘ フレキシビリティ

- ◆ Pros: ソフトウェアの細かい引数まで設定可能
- ◆ Cons: 特になし

## ⌘ 操作性

- ◆ Pros: Cloud9を利用したオンプレターミナルのような入力
- ◆ Cons: GUIと比較すると操作性は劣る

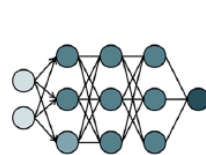
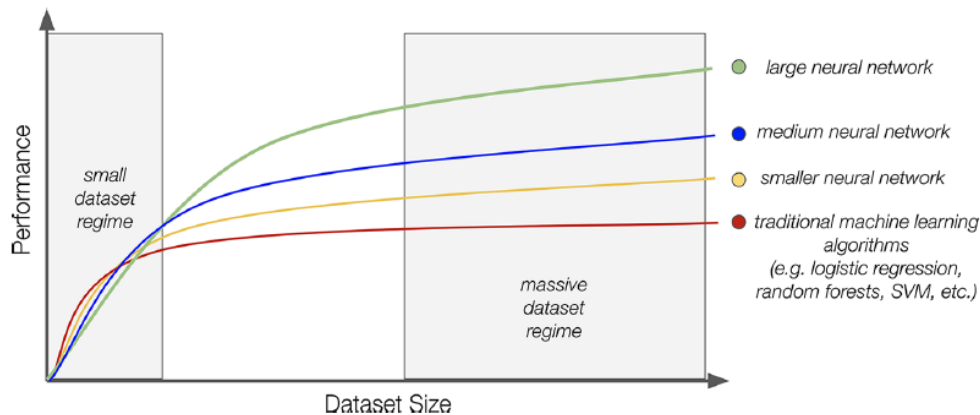
# クラウドサービス(ORION)との違い



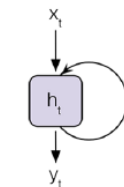
	自社AWS環境	ORION
費用	○	△
フレキシビリティ	○	△
操作性	△	◎

- どちらもクラウド(AWS)の利用により計算自体は(超)高速
- 計算の細かさによりサービスの使い分けが可能
- 将来発展としてVirtual Screening as a Service への展開

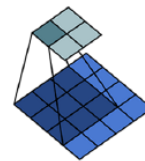
- 自己紹介・会社紹介
- 創薬研究における計算機化学の利用について
- 計算機化学におけるAWSの活用について
  - ◆ Virtual Screening : CPU, Storage
  - ◆ Machine Learning : GPU
- 展望
- まとめ



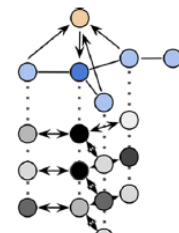
A. feed-forward neural network (NN)



B. recurrent neural network (RNN)



C. convolutional neural network (CNN)



D. graph neural network (GNN)

適切な機械学習手法は  
データサイズに依存

Life Science分野における  
様々な機械学習手法

**↑データサイズ = ↑計算時間**

本発表ではGNNに注目

## DGL LifeSci on AWS

- GNNモデリングを容易に実施可能なライブラリ
- AWSで提供しているオープンソースのプロジェクト, クラウドとの連携

### 検討実施項目

- データの秘匿化⇒セキュリティ
- クラウドリソースの効率化⇒速度向上
- モデリング⇒パラメータチューニング



<https://www.dgl.ai/>

## DGL LifeSci on AWS

- GNNモデリングを容易に実施可能なライブラリ
- AWSで提供しているオープンソースのプロジェクト, クラウドとの連携

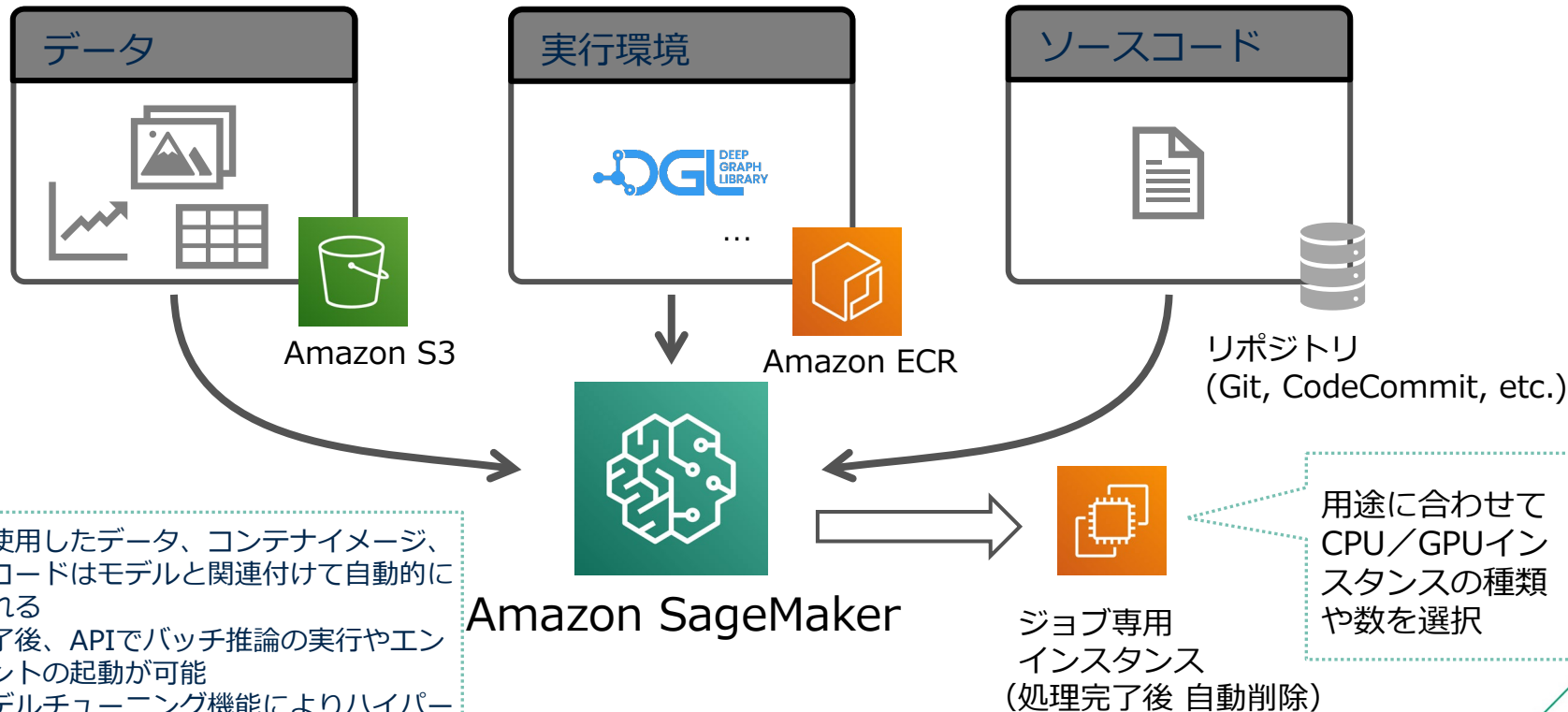
### 検討実施項目

- ✓ データの秘匿化⇒グラフ情報のシャッフリング
- ✓ クラウドリソースの効率化⇒GPUインスタンスの効率的な利用
- ✓ モデリング⇒SageMakerによるシステマティックなチューニング



# Amazon SageMaker の学習ジョブ

必要なリソースを自由に組み合わせて学習ジョブを実行可能



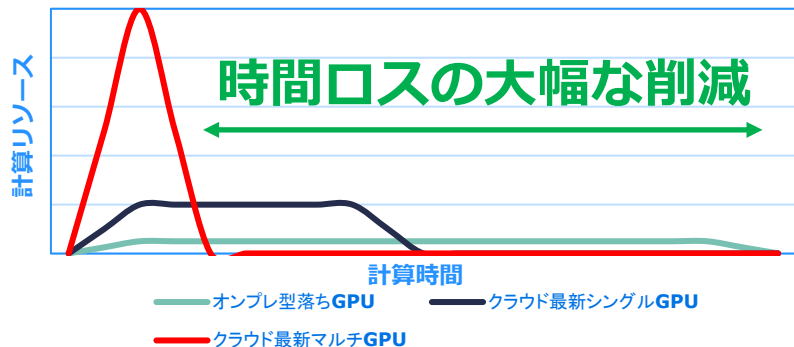
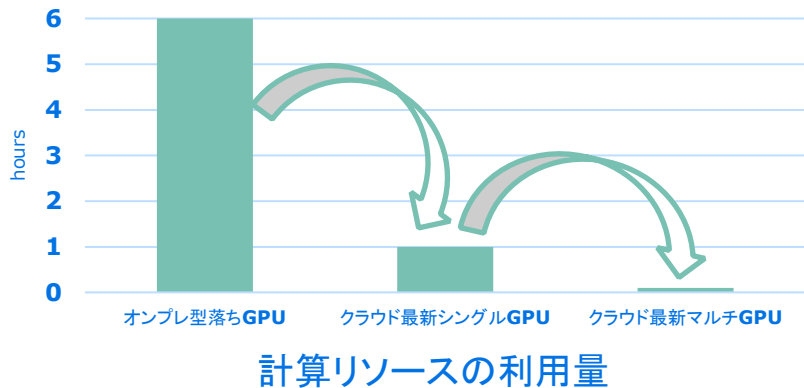
- 学習に使用したデータ、コンテナイメージ、ソースコードはモデルと関連付けて自動的に記録される
- 学習完了後、APIでバッチ推論の実行やエンドポイントの起動が可能
- 自動モデルチューニング機能によりハイパーパラメータの最適化なども可能

用途に合わせてCPU/GPUインスタンスの種類や数を選択

# AWSを利用した DGL LifeSci



## PDBBindの学習計算時間



Without cloud



With cloud



■ 計算 ■ 解析

- ⌘ 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

- 自己紹介・会社紹介
- 創薬研究における計算機化学の利用について
- 計算機化学におけるAWSの活用について
  - ◆ Virtual Screening : CPU, Storage
  - ◆ Machine Learning : GPU
- 展望
- まとめ

## ■ Multi Parameter Optimization

活性・物性予測における二つの課題

- ① 計算時間 → クラウド計算が直接的に有効な問題であり  
CPU計算高速化が解決の一例
- ② 予測精度 → クラウド計算を活用し  
最適なモデリング手法の探索を高速に実施

- ✦ 計算機化学は創薬において幅広く利用されている
- ✦ 創薬ターゲットの難化により
  - ◆ ケミカルスペースが膨大化
  - ◆ 機械学習を利用した効率化
- ✦ 従来の計算サーバーでは追いつけない計算量が必要
- ✦ クラウドの拡張性により解決できる課題として本日は二例を紹介

- アマゾン ウェブ サービス ジャパン合同会社
- オープンアイ・ジャパン株式会社
- 日本たばこ産業株式会社 医薬総合研究所 化学研究所

ご清聴ありがとうございました