

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

この資料に記載されている現在の計画、予想、戦略、想定に関する記述及びその他の過去の事実ではない記述は、アステラス製薬の業績等に関する将来の見通しです。これらの記述は経営陣の現在入手可能な情報に基づく見積りや想定によるものであり、既知及び未知のリスクと不確実な要素を含んでいます。様々な要因によって、これら将来の見通しは実際の結果と大きく異なる可能性があります。その要因としては、(i) 医薬品市場における事業環境の変化及び関係法規制の改正、(ii) 為替レートの変動、(iii) 新製品発売の遅延、(iv) 新製品及び既存品の販売活動において期待した成果を得られない可能性、(v) 競争力のある新薬を継続的に生み出すことができない可能性、(vi) 第三者による知的財産の侵害等がありますが、これらに限定されるものではありません。また、この資料に含まれている医薬品(開発中のものを含む)に関する情報は、宣伝広告、医学的アドバイスを目的としているものではありません。

Agenda

3

1 自己紹介と会社紹介

4 AWSを使った大規模計算の検証

2 Hit Identificationにおける
計算創薬ゲームチェンジ

5 今後

3 超大規模計算の実施における
課題と解決策

崎山 則征 (さきやま のりゆき)
博士(工学)

noriyuki-sakiyama@astellas.com

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

*In silico*創薬関連の業務の主担当

- ・抗体配列解析、構造予測、デザイン
- ・標的アセスメント
- ・Virtual screening
- ・分子動力学シミュレーション

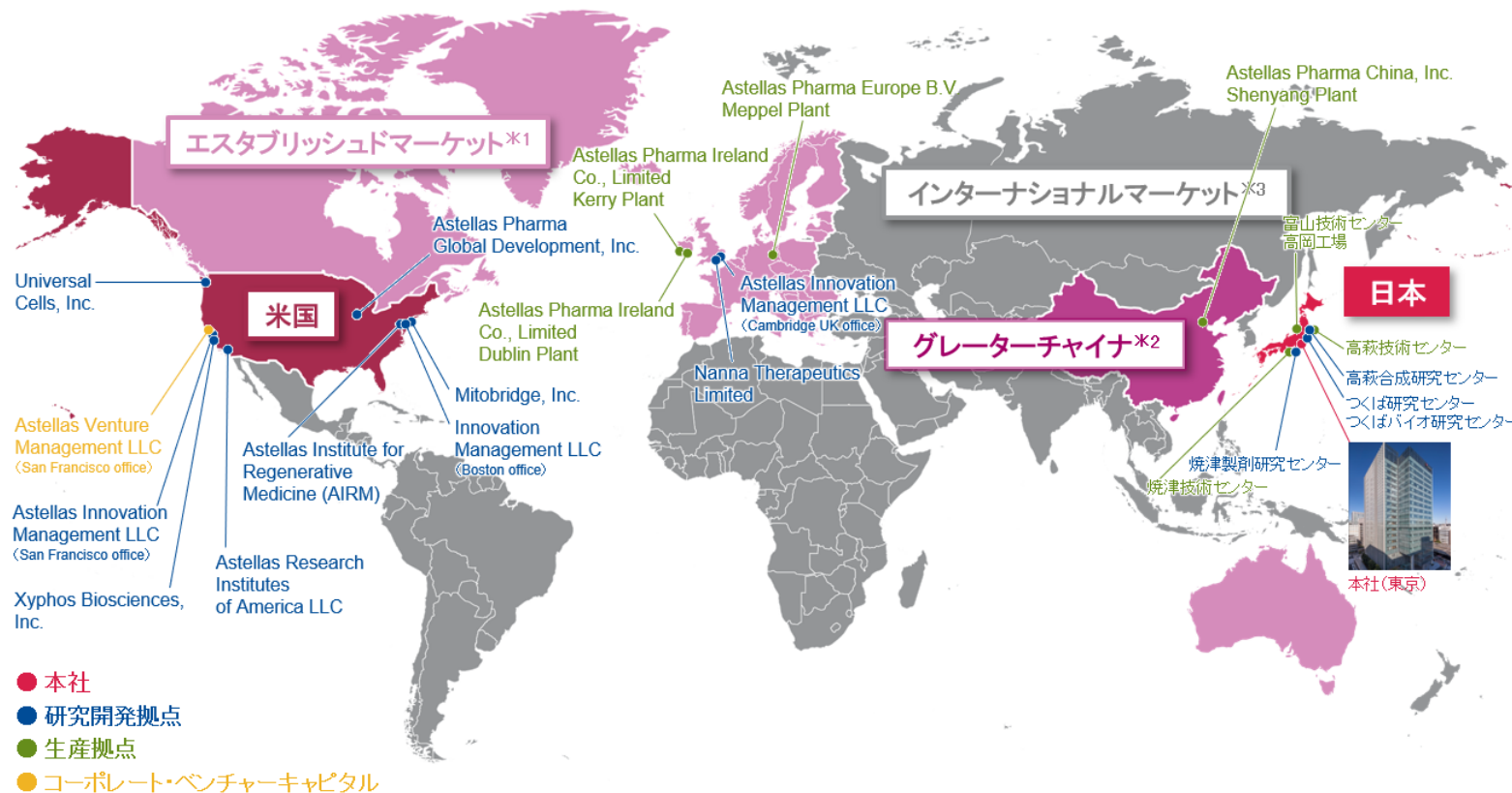
グローバル製薬企業として世界70カ国以上でビジネス展開



本社
(東京都中央区日本橋本町)

創業	1923年 (2005年: アステラス製薬発足)
代表者	代表取締役社長CEO 安川 健司
事業内容	医薬品の研究、開発、製造および販売
連結売上収益	1兆2,495億円 (2021年3月期)
連結当期利益	1,206億円 (2021年3月期)
連結従業員数	15,455名 (2021年3月期)

2021年3月期



*1 エスタブリッシュドマーケット: 欧州、カナダ、オーストラリア *2 グレーターチャイナ: 中国、香港、台湾 *3 インターナショナルマーケット: ロシア、中南米、中東、アフリカ、東南アジア、南アジア、韓国等

2021年4月現在



□Seed discoveryに必要なケイパビリティを持った、人材豊富な研究室

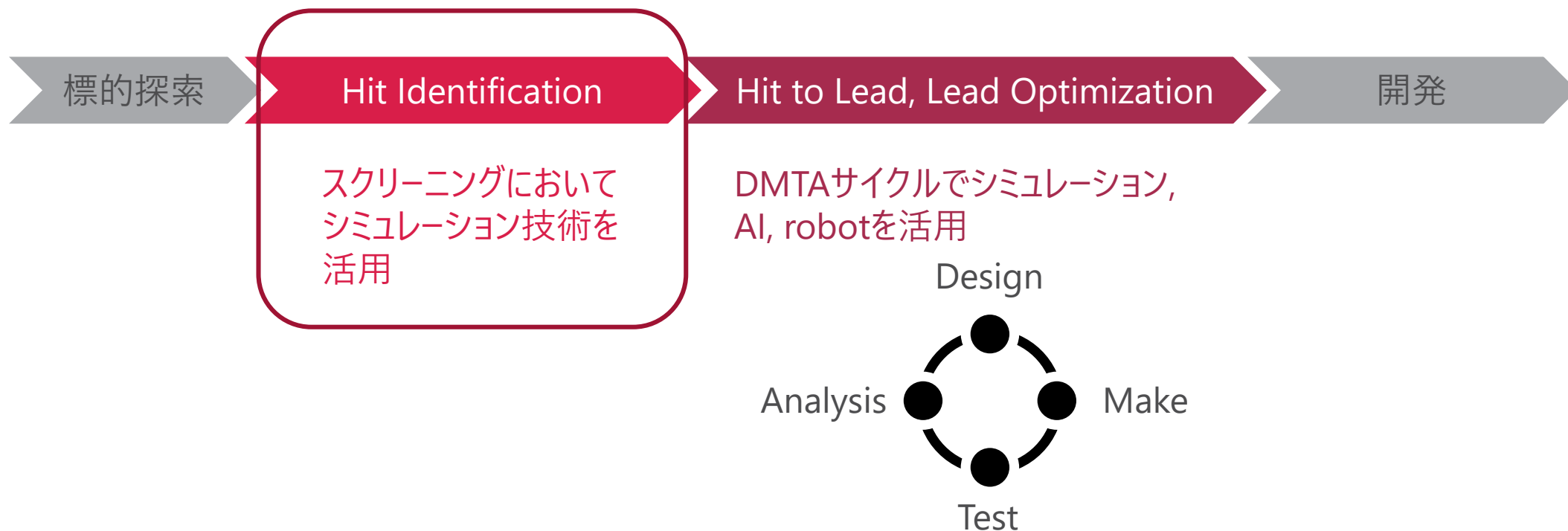
□タンパク質調製

□X-ray/Cryo-EMによる構造解析

□多種多様なアッセイ

□*in silico*解析 (標的アセスメント, virtual screening, 分子動力学シミュレーション等)

初期創薬過程においてシミュレーション/AI/ロボティクスを積極的に活用



Hit IdentificationにおけるAWS活用の取り組みをご紹介します。

これまでのアステラス製薬におけるVirtual screeningの取り組み

□ Ligand-based virtual screening(LBVS)

- 標的の立体構造が不明な場合にも活用でき、短時間で実施可能。
- リガンドの2D similarity searchやPharmacophore searchでHitを探索。

□ Structure-based virtual screening(SBVS)

- 標的の立体構造が既知の場合に活用でき、LBVSに比べて時間がかかる。
- リガンドを標的構造にドッキングし、ドッキングスコア等でHitを探索。

□ 数百万規模の自社、商用libraryを対象として実施

なぜこのタイミングでクラウドの活用が必要なのか？

Hit Identificationにおける計算創薬ゲームチェンジ

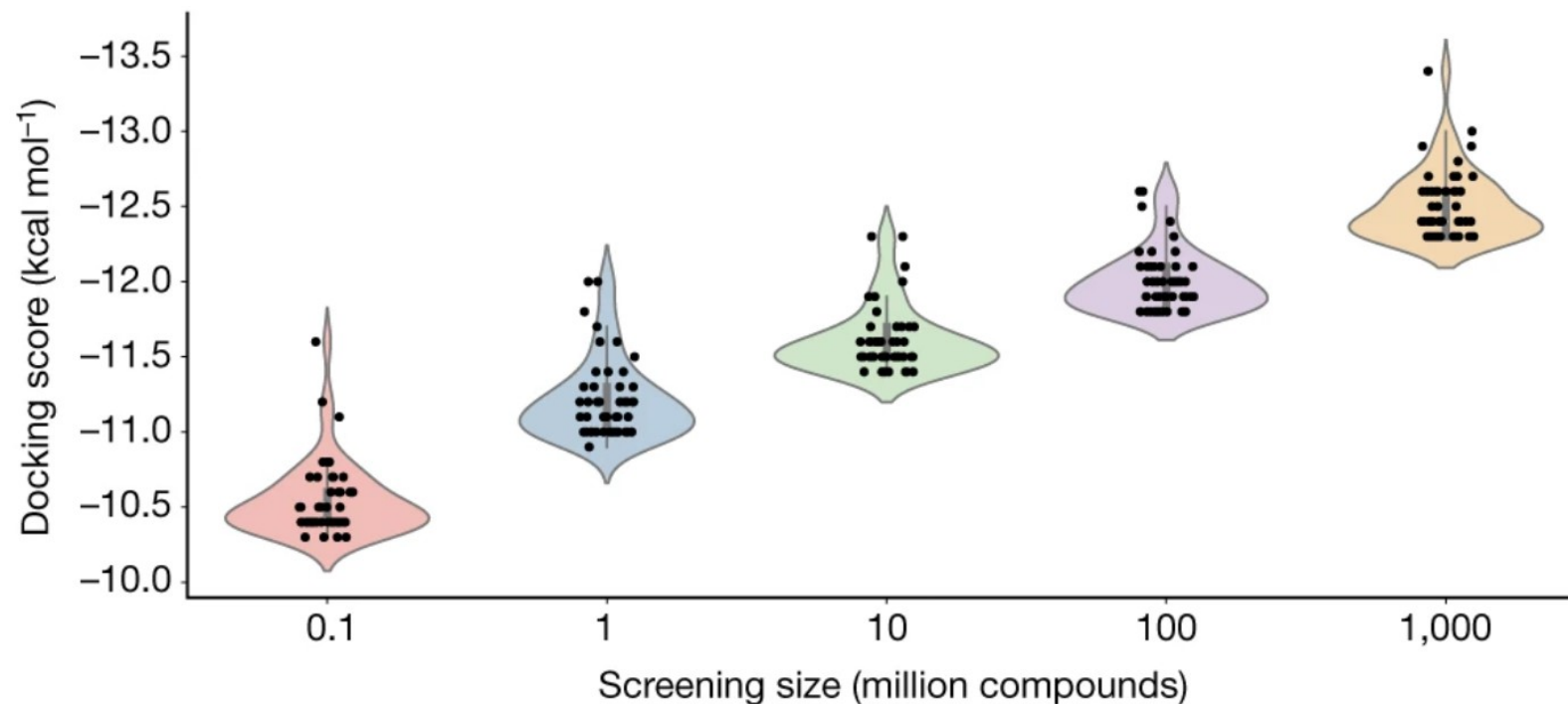
Ultra-large scale virtual screening実施例が増加

高活性化合物もHitしている。

文献DOI	報告年	標的	Library size	使用CPU数	活性値例
10.1038/s41586-019-0917-9	2019	AmpC D ₄ receptor	ZINC 9900万 ZINC 1億3800万	1500 cores	18.4nM
10.1038/s41586-020-2117-z	2020	KEAP1	Enamine 14億	8000 cores	114nM
10.1038/s41586-020-2027-0	2020	Melatonin receptor	ZINC 1億5000万	1500 cores	0.47nM
10.1016/j.isci.2020.102021	2021	SARS-CoV-2 proteins	Enamine 10億 ZINC 1000万	~16万 cores	-

Ultra-large scale libraryを使ったVirtual screeningの重要性

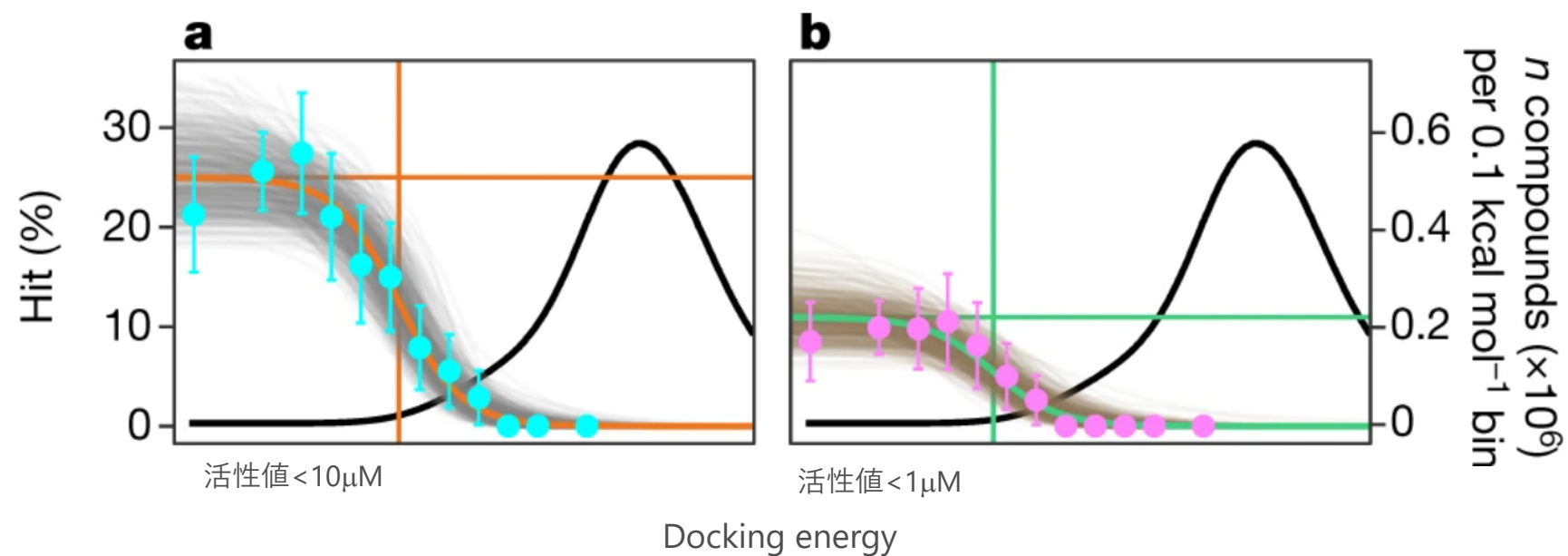
- Libraryサイズが大きくなるほど、ドッキングスコアの良い化合物の割合が向上する。



*1) <https://www.nature.com/articles/s41586-020-2117-z>から引用

Ultra-large scale libraryを使ったVirtual screeningの重要性

- Docking energy **上位**のHit rateは**22%~26%**である一方、**下位**ではHit rateが**急激に低下**する。



Docking energyとHit rateの関係

<https://www.nature.com/articles/s41586-019-0917-9>から引用

→ 超大規模なVirtual libraryを使ったVirtual screeningの実施が必要。

研究員が、

- ・数億単位のVirtual libraryを使ったVirtual screeningを、
 - ・コストを抑えつつ現実的な時間内で、
- 実施できること。

超大規模計算の実施における課題と解決策

Ultra-large scale virtual screeningを実施するための二つの方法

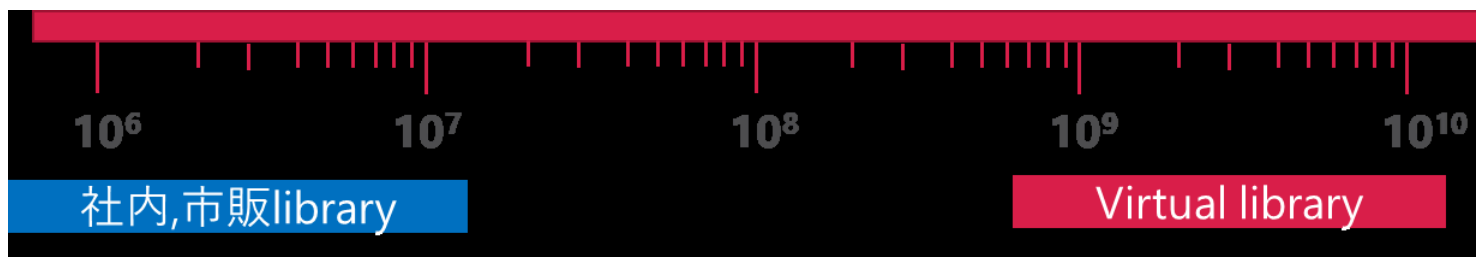
Approach	Brute Force(BF) based approach	Machine Learning(ML) based approach
Application	VirtualFlow ^{*1} , Giga docking ^{*2}	Active learning glide(AL-Glide) ^{*3}
Method	全化合物をドッキング	1. 少数のドッキング結果から予測モデルを構築し、全化合物のドッキングスコア予測 2. 上位化合物に対してドッキングによるリスコアリングを実施
Pros	Full Dockingのため取りこぼしがない	機械学習を活用するため、必要な計算リソースはBF-basedに比べて少ない
Cons	必要な計算リソースが膨大(数千CPU以上)	Full Dockingで上位に来る化合物の7割程度の回収率との報告

*1) <https://www.nature.com/articles/s41586-020-2117-z>

*3) <https://www.schrodinger.com/>

*2) <https://www.eyesopen.com/>

超大規模なvirtual libraryへのアクセスが可能になったがサーバーが追いつかない



以前はVirtual screeningに使う
Libraryは数十万～百万規模
→ 社内サーバーで計算可能



Enamine REAL*¹など合成可能性の高い大規模なVirtual libraryが利用可能に。
→ 社内サーバーで計算**困難**

Ultra-large scale virtual screening 実施可能なクラウド環境の構築

- BF-based VS: 超大規模な計算環境が必要(数千CPU)。
- ML-based VS: 数千もCPUは必要としないが、機械学習用に最新のGPUが望ましい。

技術検証段階でオンプレミスなサーバー構築は時間も費用も現実的ではない。



短時間での環境構築、最新のCPU/GPUが利用可能なクラウド環境を活用

Point

- ・セキュリティに問題がない。
- ・研究者が環境を構築でき、簡単に使う事が可能。
- ・オンプレミスに比べてコストが抑えられる。

Amazon Web Service(AWS)を活用することに決定

□セキュリティに問題がない。

- 社内でAWSが利用されており、社内ネットワークへの接続実績がある。

□研究者が環境を構築でき、簡単に使う事が可能。

- AWS JAPANの技術者の方から手厚いサポートがある。ライフサイエンス分野での実績も多数ある。

□オンプレミスに比べてコストが抑えられる。

- ParallelClusterで計算資源を自動・効率的に活用できる。
- Spot instanceを利用することで安価に使う事ができる。

AWSを使った大規模計算の検証

今回はVirtualFlowとAL-Glideの検証を行った

□BF-based VS

→ VirtualFlowの環境構築

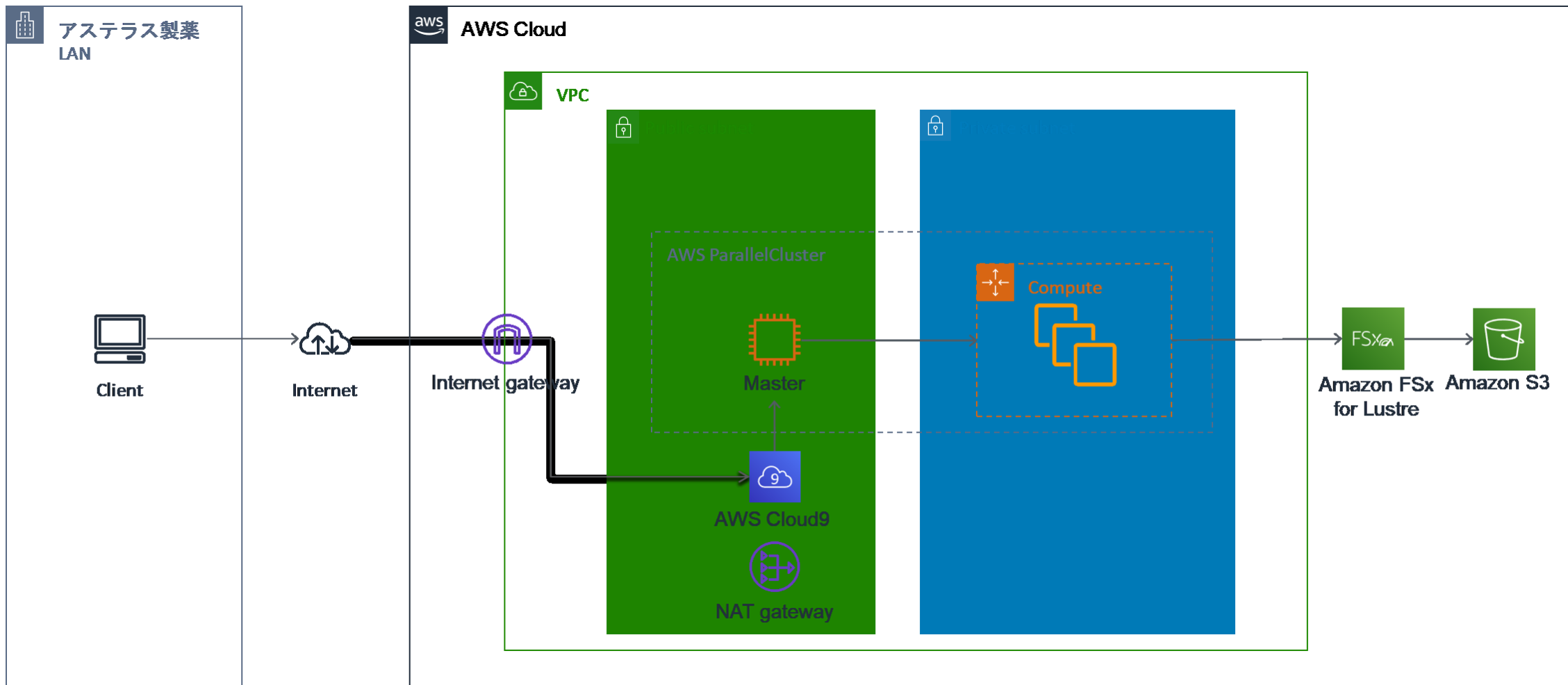
□ML-based VS

→ AL-Glideの環境構築

Point

- ・ParallelClusterで計算資源を有効活用できるかどうか
- ・Spot instanceが実用的かどうか
- ・研究員が簡単に使うことができるかどうか

VirtualFlowを使うためのAWS構成図



AWSサポートを受けながら研究員が環境構築実施



□ VirtualFlowのスケラブルな稼働を確認。

- ParallelClusterが正常に機能し、必要なだけの計算資源を自動管理できたことを確認。

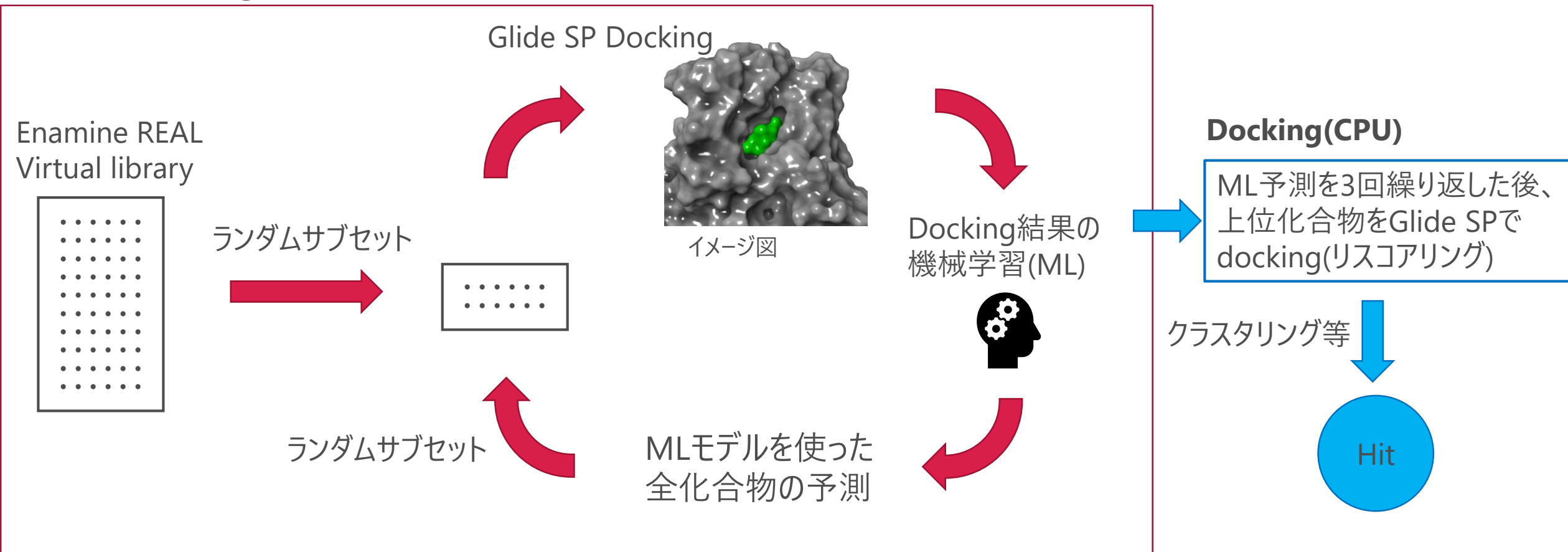
□ スモールサイズ(数百万)のLibraryによる検証を実施。

- スモールサイズのせいか、ライブラリ分割に問題があった。
1個しか含まれないサブセットや1000個含まれるサブセットなど
- スモールサイズの計算結果から見積もると、想定以上のコストとなった。

今回のPoCではVirtualFlowの稼働確認で終了とし、Ultra-large scale libraryへの適用は見送った。

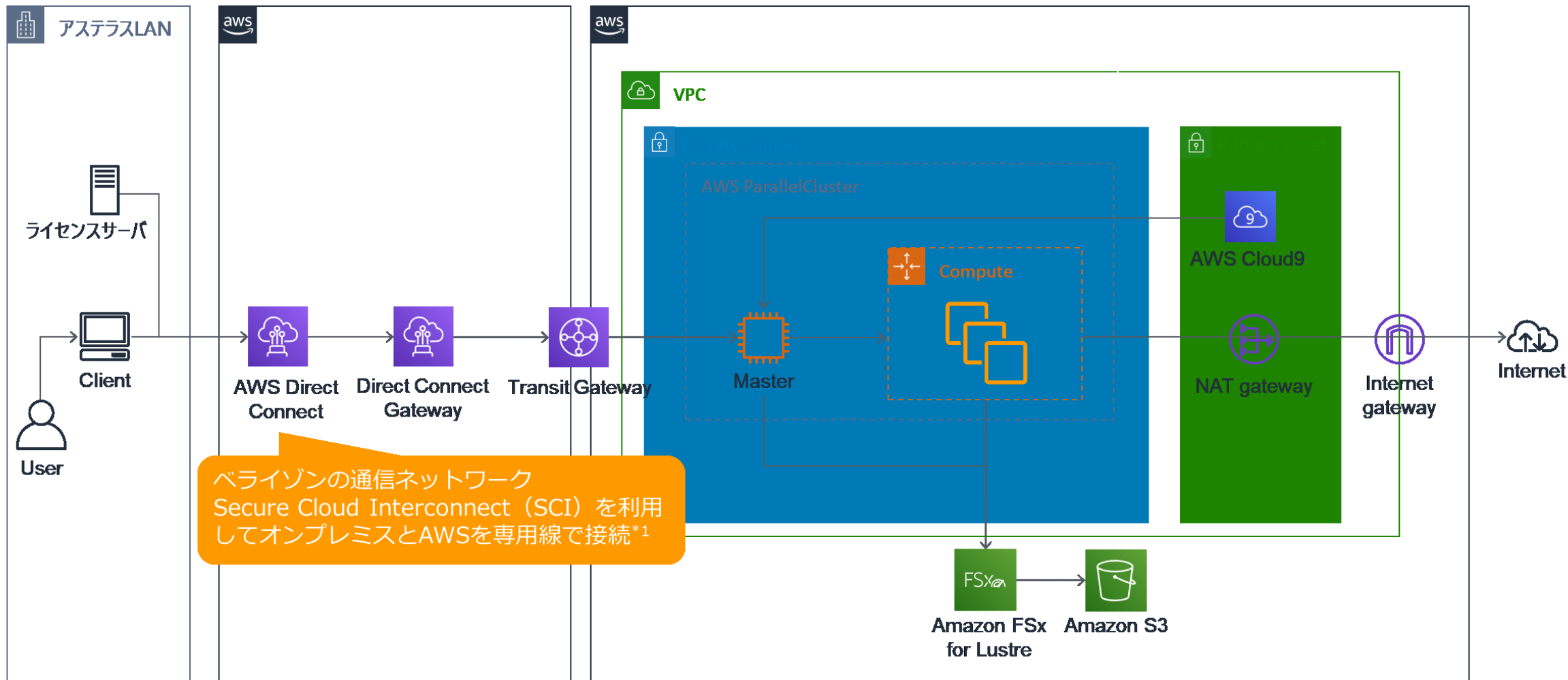
AL-Glideによるvirtual screening概要

Machine Learning(予測: CPU, 学習: GPU)



ParallelClusterでスケラブルにCPU, GPUインスタンスを自動管理。

AL-Glideを使うためのAWS構成図

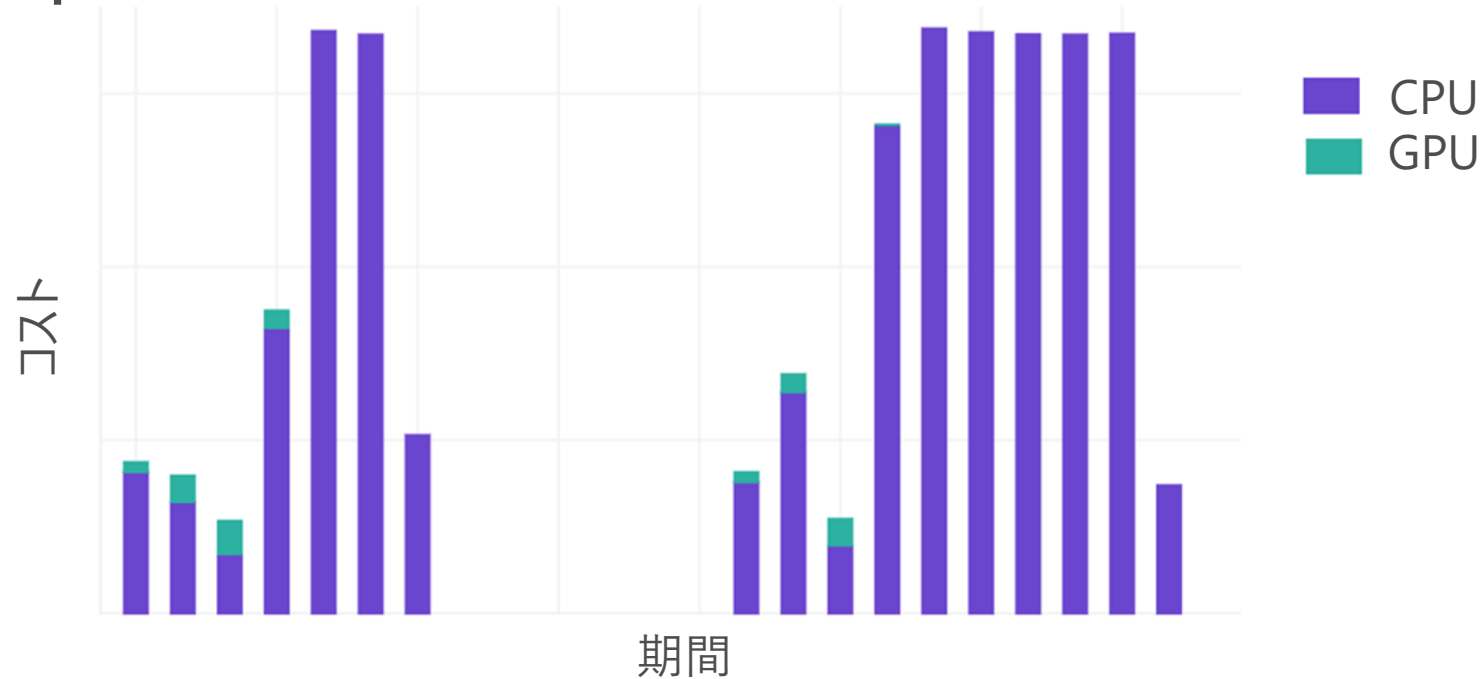


Transit Gateway等の構築は情報システム部、その他はAWSサポートを受けながら研究員が構築



ParallelClusterで計算終了まで効率的に計算資源を管理できた

□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。

□AL-Glideの稼働を確認。

- 社内のライセンスサーバーを参照し、正常な稼働を確認。
- ParallelClusterにより、必要なだけの計算資源を自動管理できたことを確認。

□Enamine REAL 約6.5億を対象とした検証を実施。

- 条件検討を複数回実施し、Schrödinger様から頂いた情報と大きく乖離しない時間内にGlide SPによるリスコアリングまで完了することを確認。
- 各論文記載の通り、小規模ライブラリと比較して、億単位ライブラリでのドッキングスコアの向上も確認。

□ ParallelClusterで計算資源を有効活用できるかどうか

- CPU/GPUインスタンスの併用も問題なく活用できた。

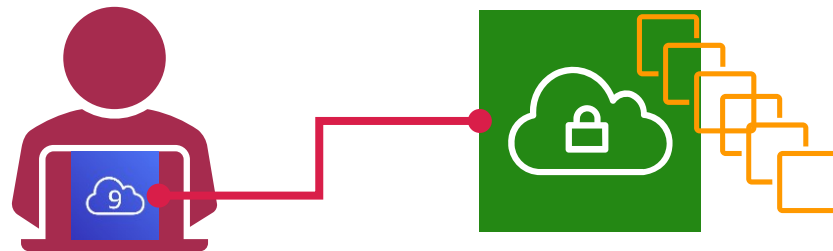
□ Spot instanceが実用的かどうか

- 検証段階でSpot instance切断が確認できたのは数回。
- 自動リスタートで最後まで完走(1回の計算を短時間で終わるような調整はした方が良い)。

□ 研究員が簡単に使うことが出来るかどうか

- ParallelClusterのインスタンス変更などもConfigに書き込むだけなので簡単。

AWS Cloud9上でアプリケーション
(VirtualFlow, AL-Glide)を実行



ParallelClusterが適切にインスタンスを
起動し、不要になればシャットダウン

研究員が、

- ・今回構築したAWS上の環境を使いこなし、
- ・数億単位のVirtual libraryを使ったVirtual screeningを実施し、化合物を見つける。

Amazon Web Service Japan 合同会社

Schrödinger 株式会社

アステラス製薬株式会社 情報システム部

アステラス製薬株式会社 ストラテジックスクリーニングサイエンス研究室

ご清聴ありがとうございました