

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

Mitsuaki Tsugo

Amazon Web Services Japan

自己紹介

名前：津郷 光明（つごう みつあき）

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

- ・ 課題の解決に向けたご支援
- ・ アーキテクチャ検討支援



このセッションの目的とゴール

Point 1

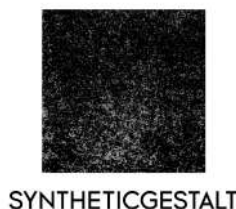
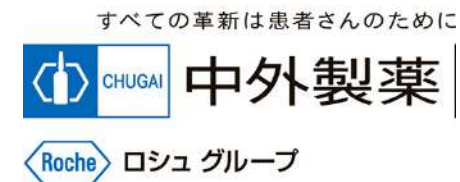
- 創薬研究領域においてAWSの利用が増えているユースケースの理解
 - HPC・ゲノム解析
 - 機械学習
 - データ管理・コラボレーション

Point 2

- 各ユースケースでの課題解決に有効なAWSサービスおよびソリューションを知っていただく

国内における製薬関連のお客様

(一部抜粋)



製薬バリューチェーンの各段階で DX を支援



創薬研究

- ✓ 共同研究基盤 (データ保管・転送など)
- ✓ HPC (ゲノム解析含めて)
- ✓ AI/ML 研究業務支援



臨床開発

- ✓ R&Dデータレイク
- ✓ IoT/モバイル (ePROなど)
- ✓ AI/ML 臨床開発業務支援 (CRO含めて)



製造

- ✓ スマートファクトリー
- ✓ GxP/CSV 対応支援
- ✓ MES on AWS
- ✓ SAP on AWS



営業・マーケティング

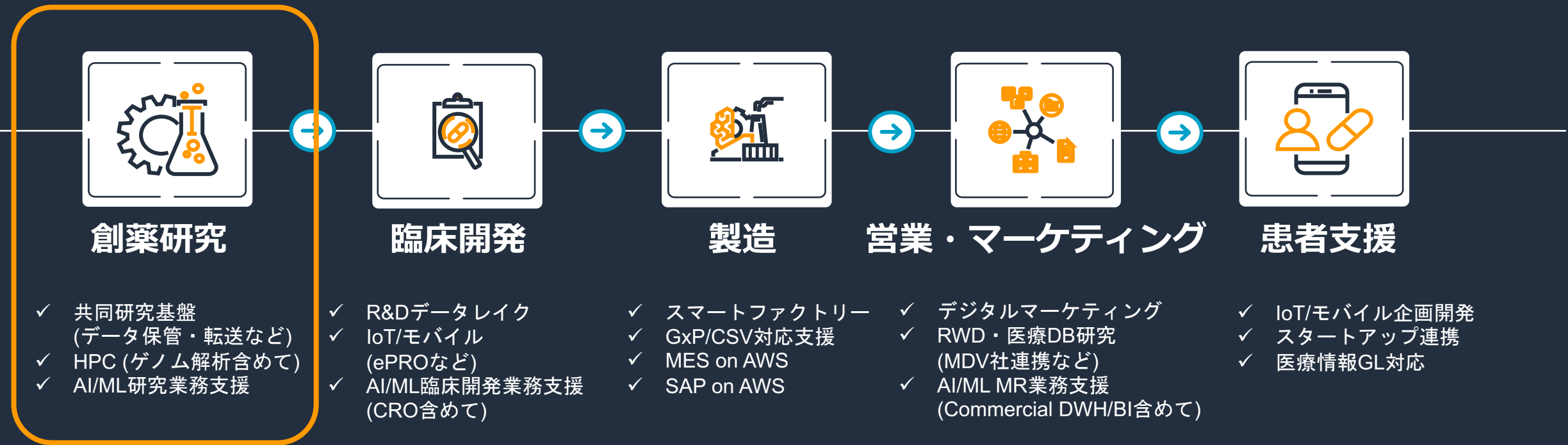
- ✓ デジタルマーケティング
- ✓ RWD・医療DB研究 (MDV社連携など)
- ✓ AI/ML MR 業務支援 (Commercial DWH/BI含めて)



患者支援

- ✓ IoT/モバイル企画開発
- ✓ スタートアップ連携
- ✓ 医療情報GL対応

製薬バリューチェーンの各段階で DX を支援



AWSを活用した近年の創薬研究領域でのユースケース

- HPCの活用・ゲノム解析
- 機械学習の活用
- データ管理・コラボレーション基盤の構築

HPC活用・ゲノム解析

機械学習の活用

データ管理・コラボレーション基盤

HPCの活用・ゲノム解析

HPC活用・ゲノム解析

機械学習の活用

データ管理・コラボレーション基盤



創薬研究におけるHPC活用

膨大な計算や処理のためハイパフォーマンスコンピューティングを利用

ゲノム解析



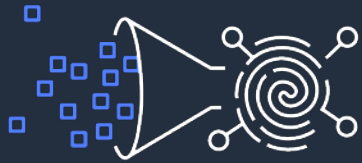
大規模バーチャルスクリーニング



クライオEM単粒子解析



これまでの HPC 環境の課題



限られたリソース・
長い待ち時間

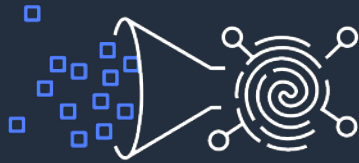


リソース不足・無駄



アプリケーションの
多様化・複雑な管理

これまでの HPC 環境の課題



限られたリソース・
長い待ち時間



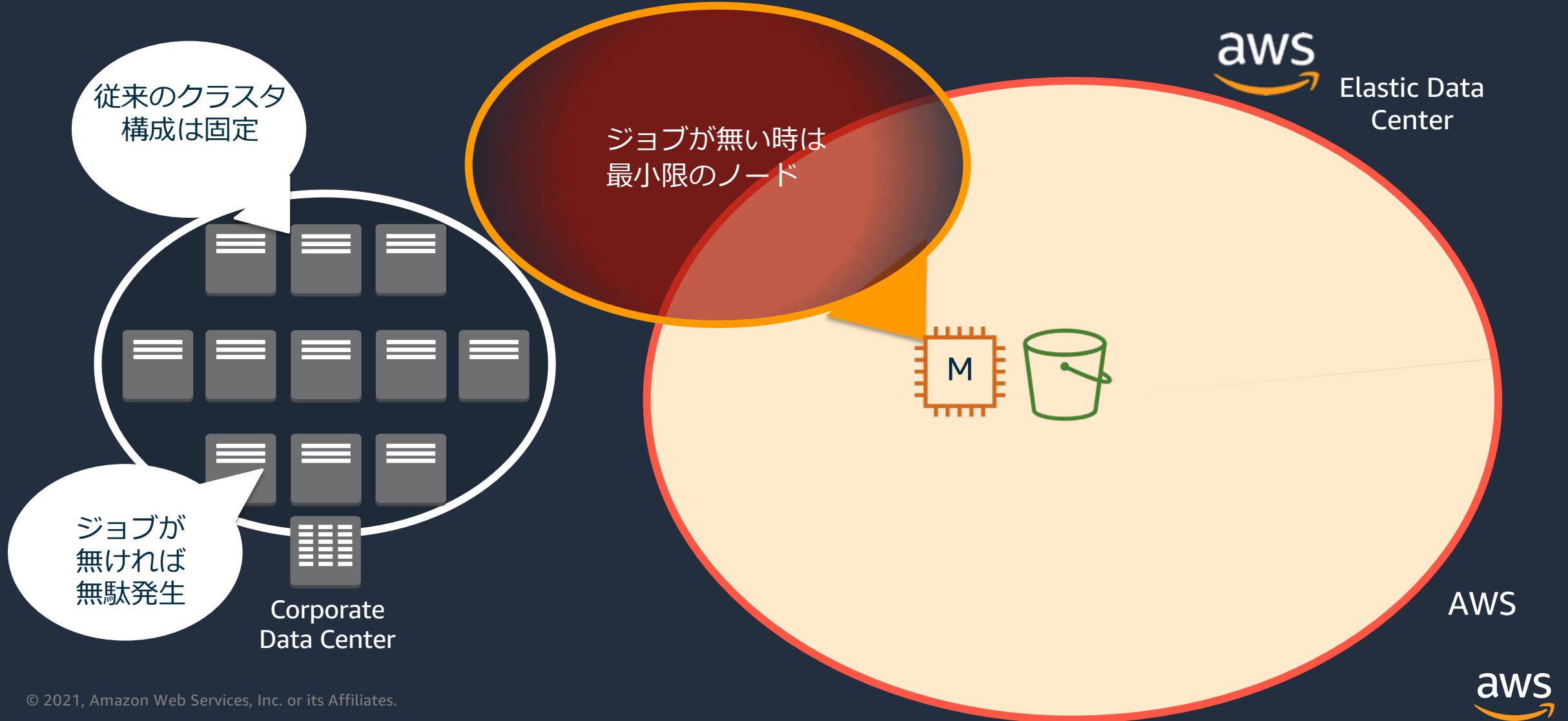
リソース不足・無駄



アプリケーションの
多様化・複雑な管理

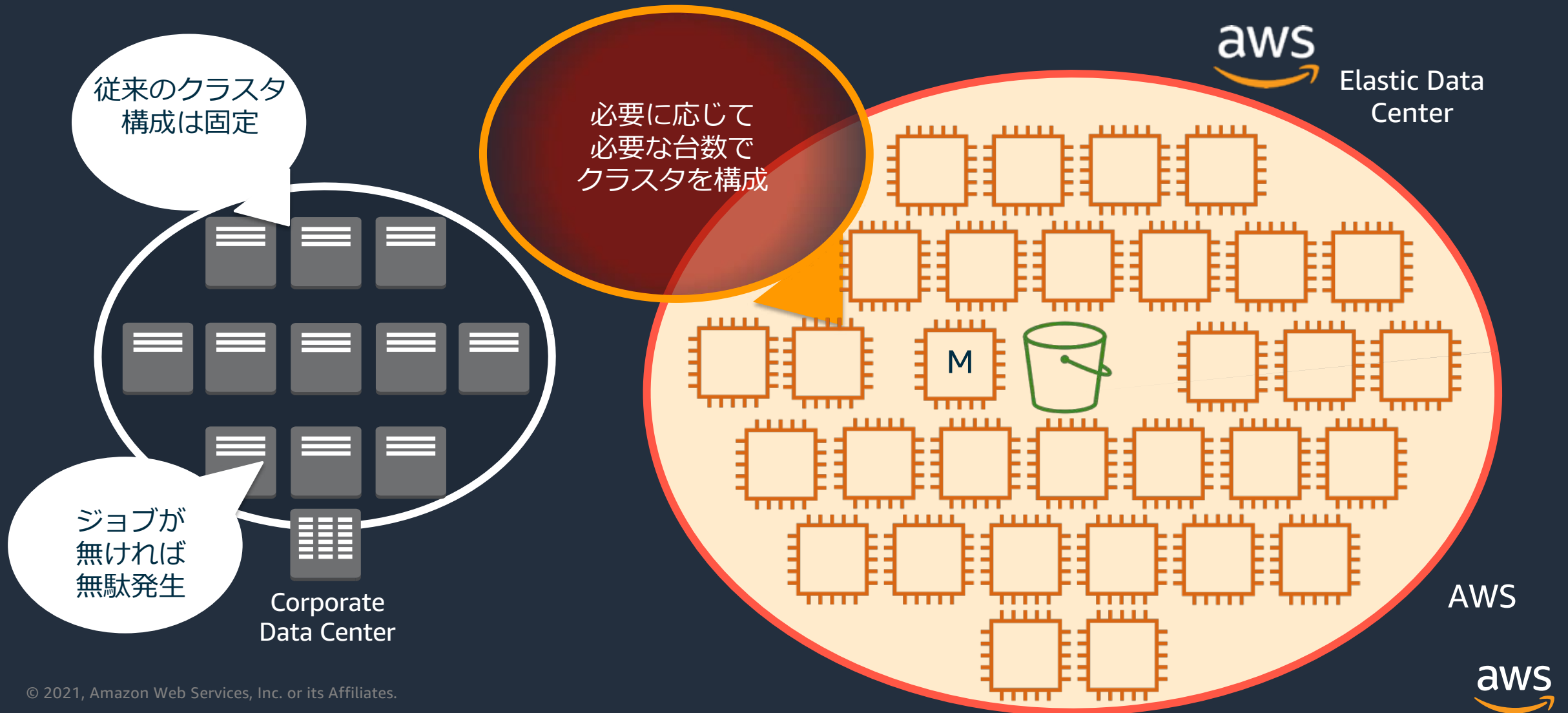
AWSなら、必要な時に必要なだけ利用可能

スケーラブルなリソースによりジョブ実行待ちの無いHPC環境を実現



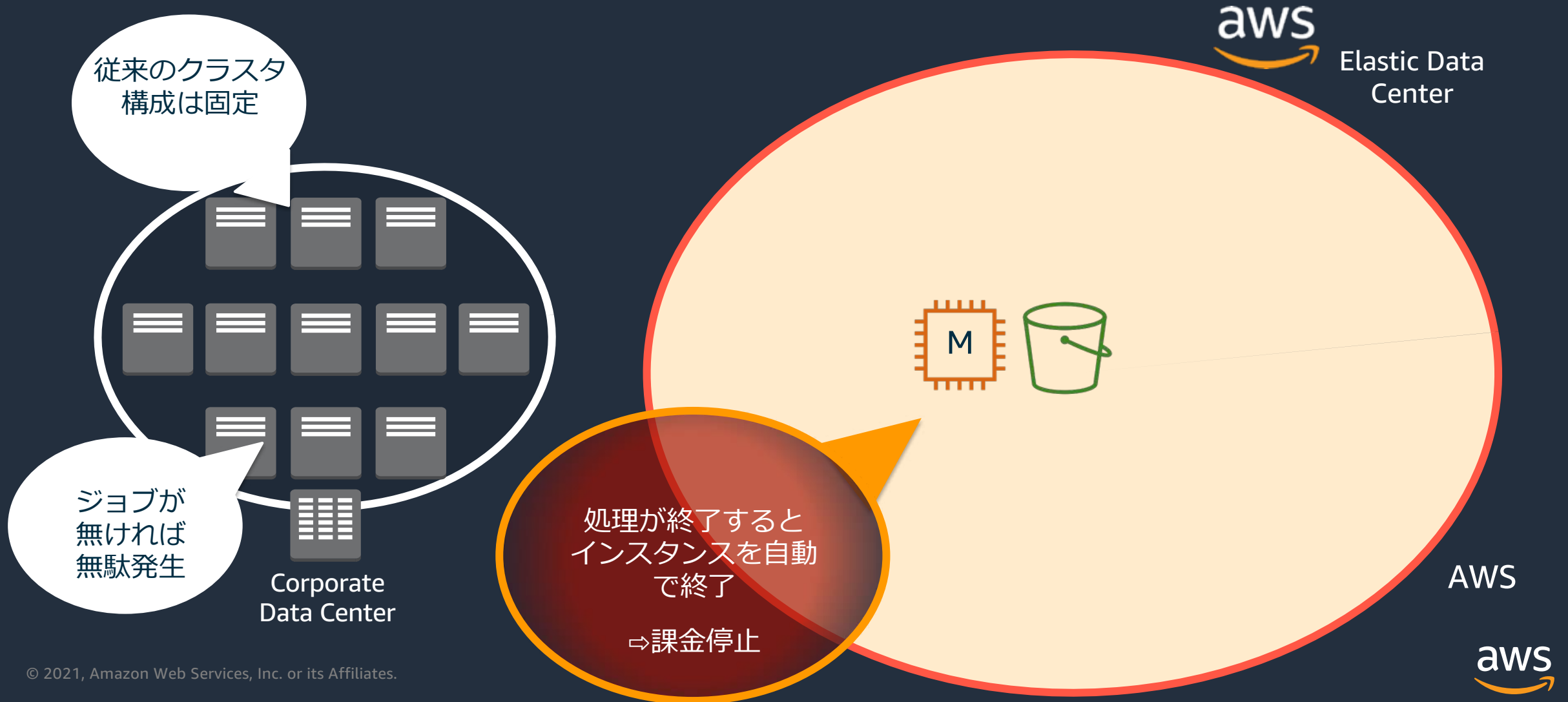
AWSなら、必要な時に必要なだけ利用可能

スケーラブルなリソースによりジョブ実行待ちの無いHPC環境を実現

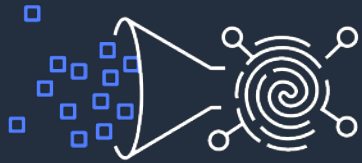


AWSなら、必要な時に必要なだけ利用可能

スケーラブルなリソースによりジョブ実行待ちの無いHPC環境を実現



これまでの HPC 環境の課題



限られたリソース・
長い待ち時間



リソース不足・無駄



アプリケーションの
多様化・複雑な管理

アプリケーションに合わせた構成のクラスタを構築可能

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て、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
- アクセラレータ
- ネットワーク
- インストールするソフトウェア

AWS で利用可能なCPU/アクセラレータ例



One size does not fit all!

AWSはやりたいことをご自身で実現する Self Service Platform

1つのサービスやツールでは自由度と実装コストの両立に限界

⇒ 複数のサービスを適材適所で組み合わせ、**やりたいことを最小の手間で実現する**

➡ **Building Block**



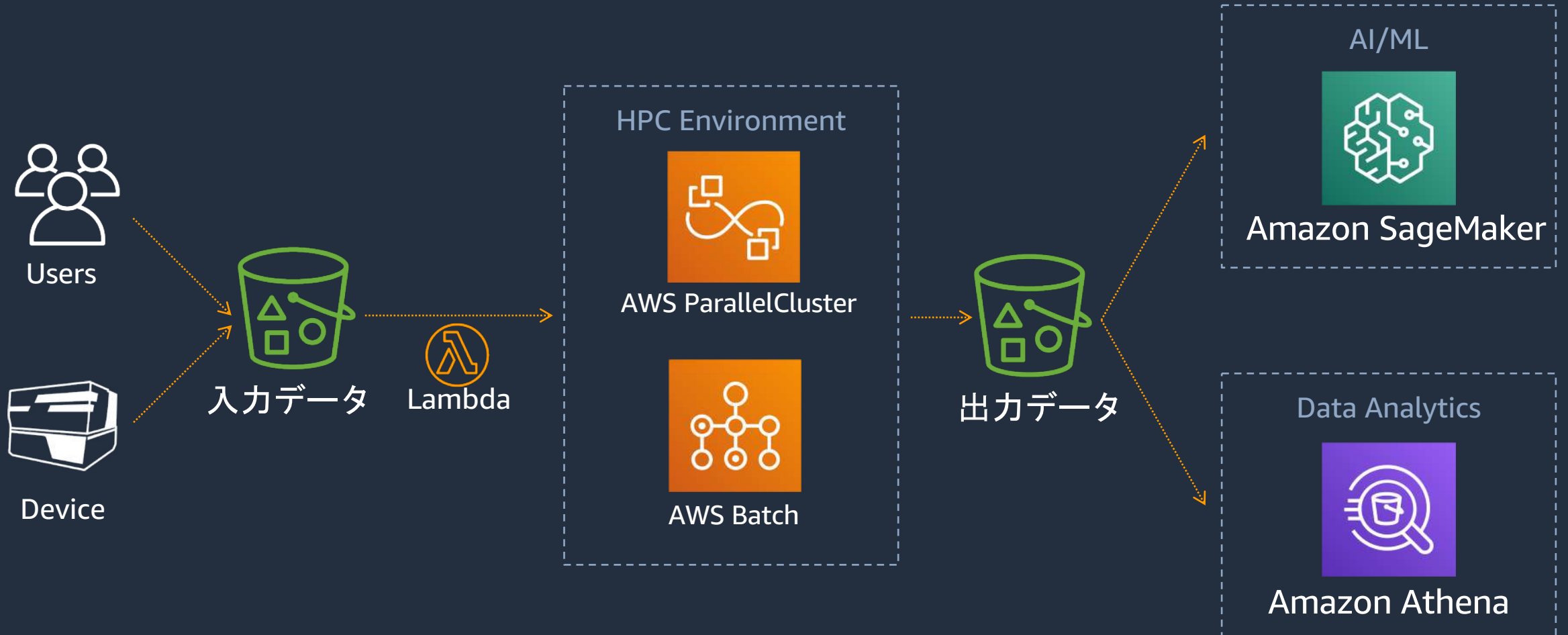
http://farm4.static.flickr.com/3514/3281353786_c1a130ff2e_b.jpg

- 価値創出にフォーカスできる
- 失敗や試行錯誤が容易
- リードタイムの短縮

例：データドリブンなHPC環境とデータ活用

データのアップロードをトリガーにHPC環境を展開し自動で処理を行う

更にS3のデータレイク化により大規模シミュレーション結果を機械学習環境で活用



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

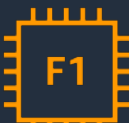
Amazon EC2



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



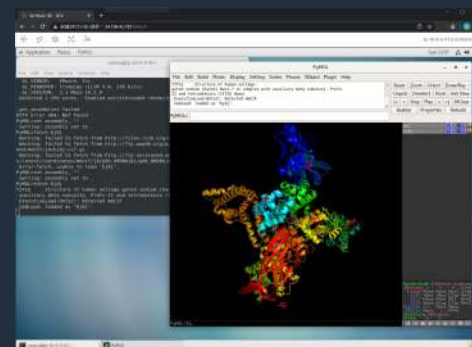
ストレージ内蔵の筐体を郵送することで S3 にデータを Import/Export

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

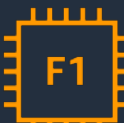
Amazon EC2



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



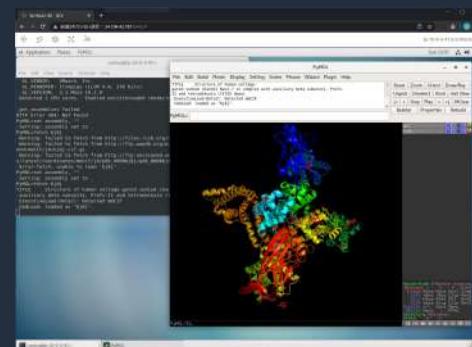
ストレージ内蔵の筐体を郵送することで S3 にデータを Import/Export

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

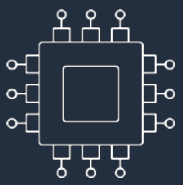


AWS上に HPC クラスタを自動で構築。Slurm / SGE / Torque といったジョブスケジューラに対応しており既存 HPC環境からの移行が容易

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



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

M5zn インスタンス

Cascade Lake

最大全コア 4.5 GHz 駆動

C5a インスタンス

EPYC Rome

最大 3.3 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

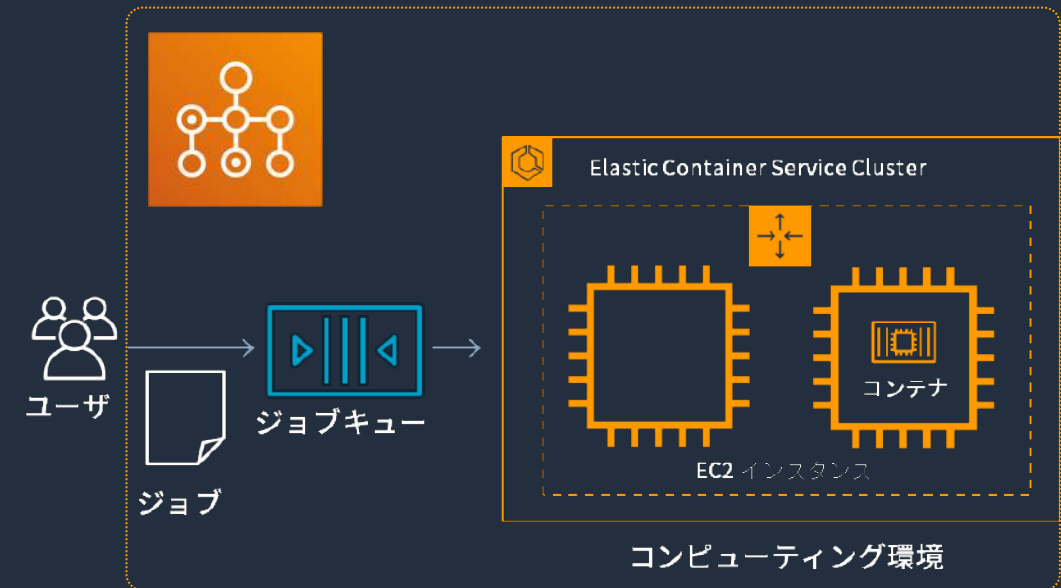
Virtex UltraScale+
VU9P 搭載



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

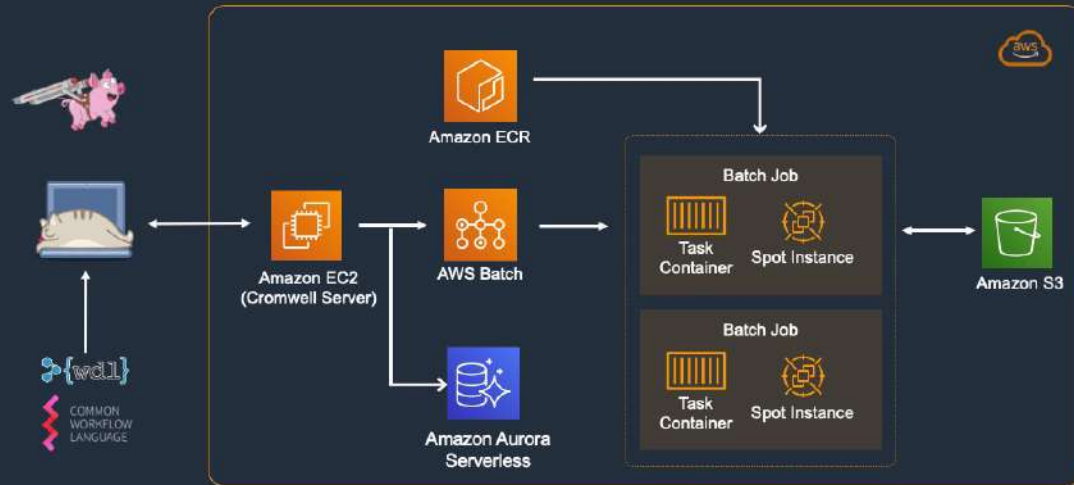


コンテナイメージを用意するだけでスケーラブルな大規模バッチ処理環境が得られる

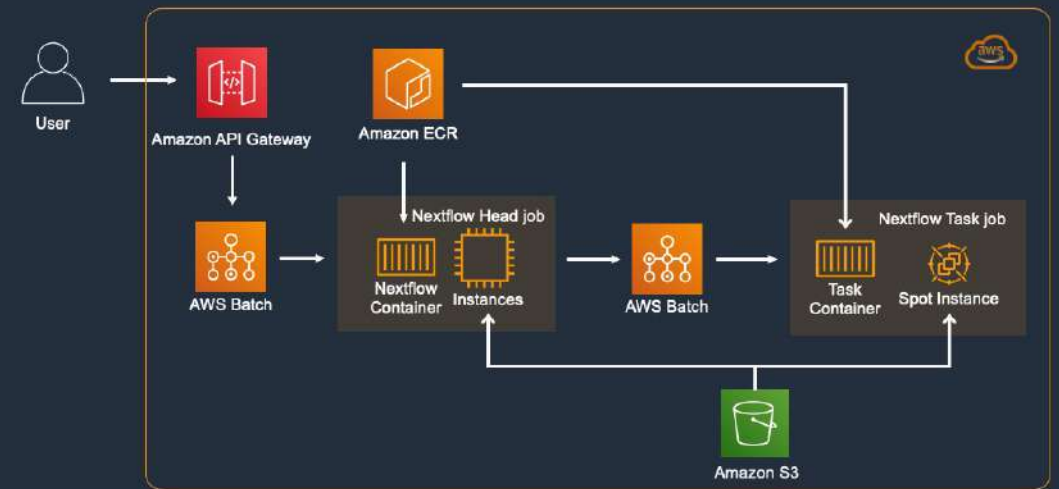
AWS上でのワークフローエンジン実行

複雑な工程や膨大なプログラムをワークフローエンジンで管理

Cromwell on AWS



Nextflow on AWS

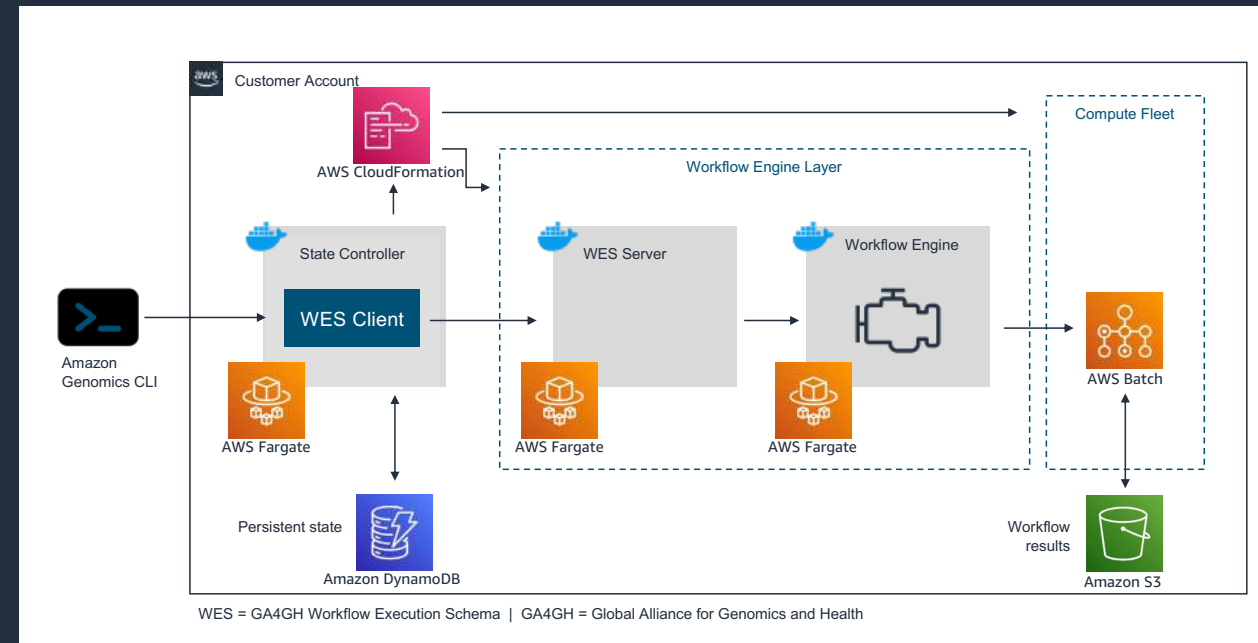


ゲノム解析に特化した環境構築支援ツール

Amazon Genomics CLIを利用して環境構築を自動化

✓ Amazon Genomics CLI

- ✓ ワークフローエンジンに対応したAWS環境を構築するOpen Source CLI
 - ✓ 構築の手間を削減
 - ✓ 再現性の確保
 - ✓ 計算量に合わせて自動拡張
- ✓ 選択可能なワークフローエンジン
 - ✓ Cromwell、miniWDL、Nextflow、GATK



機械学習の活用

HPC活用・ゲノム解析

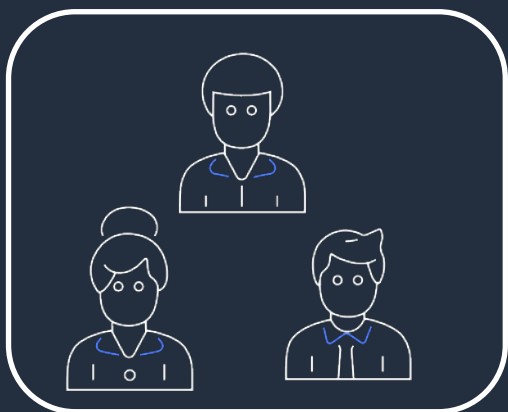
機械学習の活用

データ管理・コラボレーション基盤

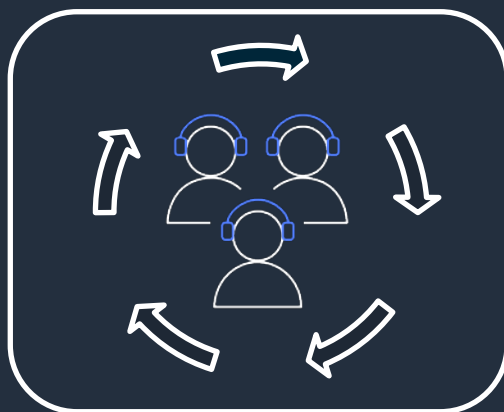
創薬研究における機械学習

生産性向上に向けて創薬研究領域においても機械学習技術を活用

個人差への対応・作業品質の安定



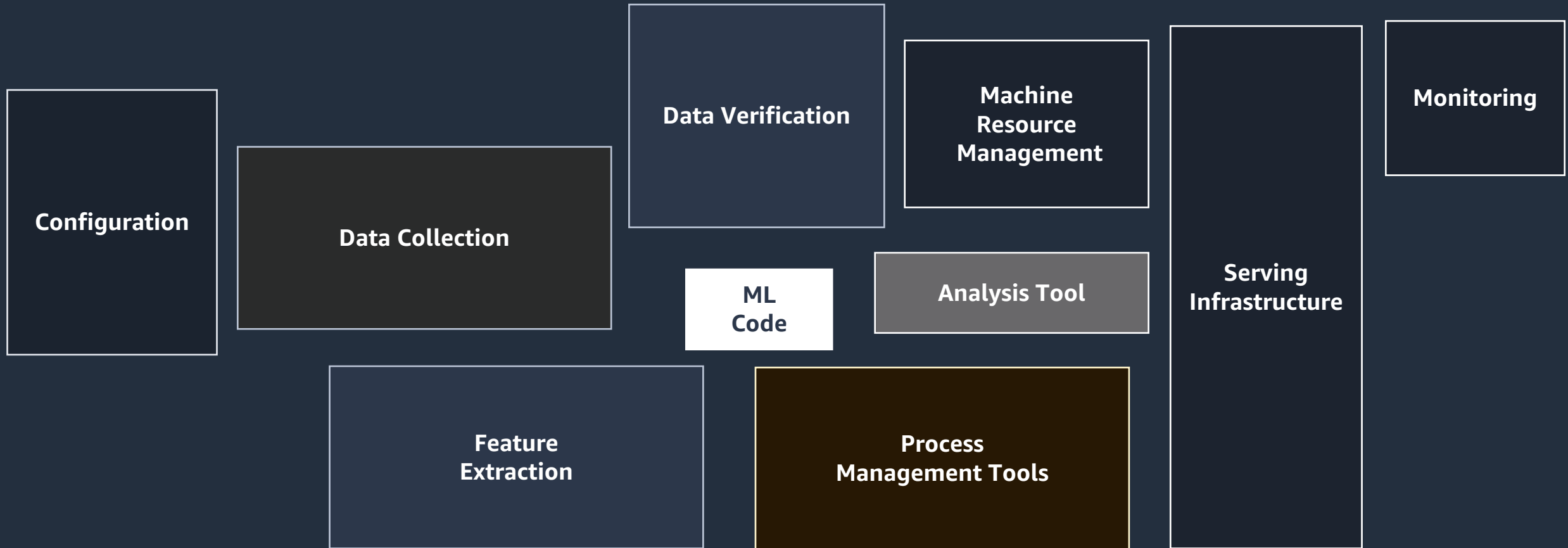
人の反復作業削減



複雑な関係性からの洞察



機械学習における課題



“Only a small fraction of real-world ML systems is composed of the ML code”

source: Hidden Technical Debt in Machine Learning Systems [D. Sculley, & al.] – 2015

<https://papers.nips.cc/paper/5656-hidden-technical-debt-in-machine-learning-systems.pdf>

機械学習における課題

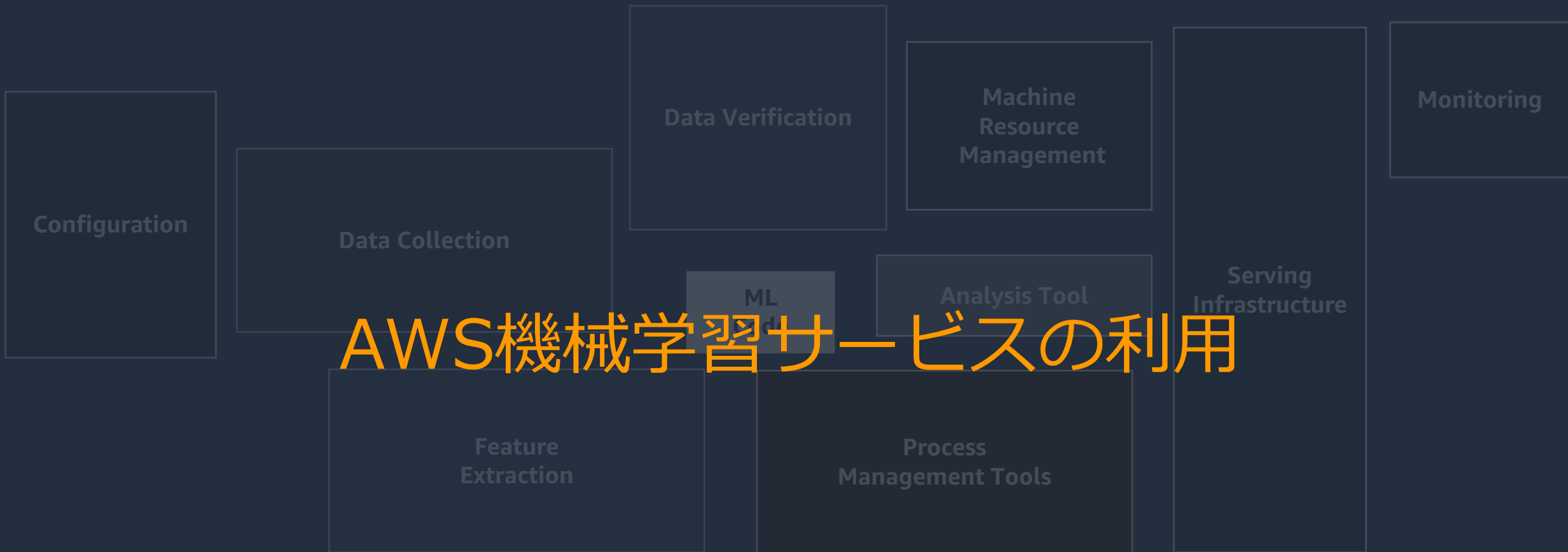


“Only a small fraction of real-world ML systems is composed of the ML code”

source: Hidden Technical Debt in Machine Learning Systems [D. Sculley, & al.] – 2015

<https://papers.nips.cc/paper/5656-hidden-technical-debt-in-machine-learning-systems.pdf>

機械学習における課題



AWS機械学習サービスの利用























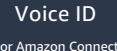
“Only a small fraction of real-world ML systems is composed of the ML code”

source: Hidden Technical Debt in Machine Learning Systems [D. Sculley, & al.] – 2015

<https://papers.nips.cc/paper/5656-hidden-technical-debt-in-machine-learning-systems.pdf>

AWS の機械学習スタック

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 +Medical		TEXT  Amazon Comprehend +Medical  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

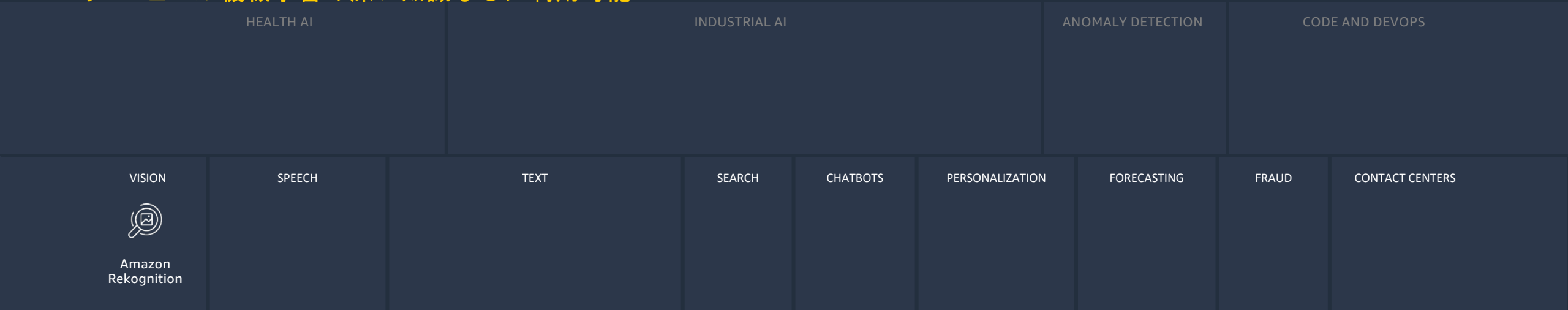
NEW: Model management for edge devices

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  Deep Learning AMIs & Containers	 GPU & CPUs  Elastic Inference	 Trainium  Inferentia	 FPGA
--	---	---	--	--

AWS の機械学習スタック

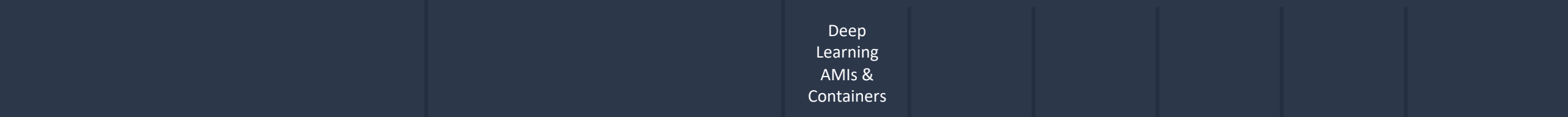
AI サービス: 機械学習の深い知識なしに利用可能



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

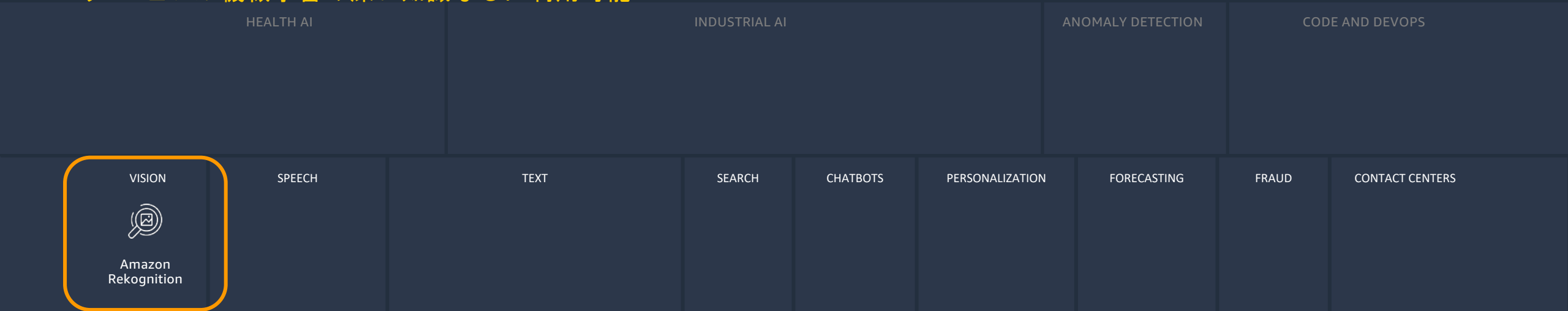


ML フレームワークとインフラストラクチャ: 機械学習の環境を自在に構築して利用



AWS の機械学習スタック

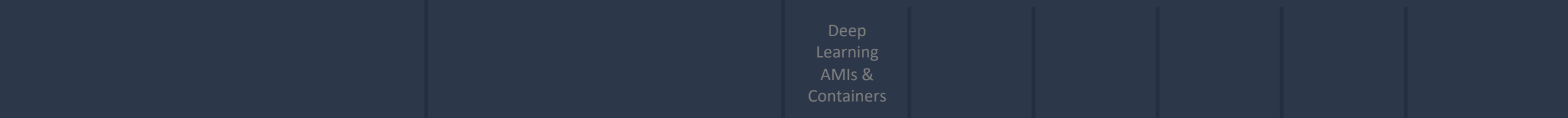
AI サービス: 機械学習の深い知識なしに利用可能



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



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



Amazon Rekognition

深層学習を用いた静止画/動画分析サービス

顔検出と分析



有名人の認識



物体の検出



シーンの検出



コンテンツのモデレーション



活動の検出



動線の検出

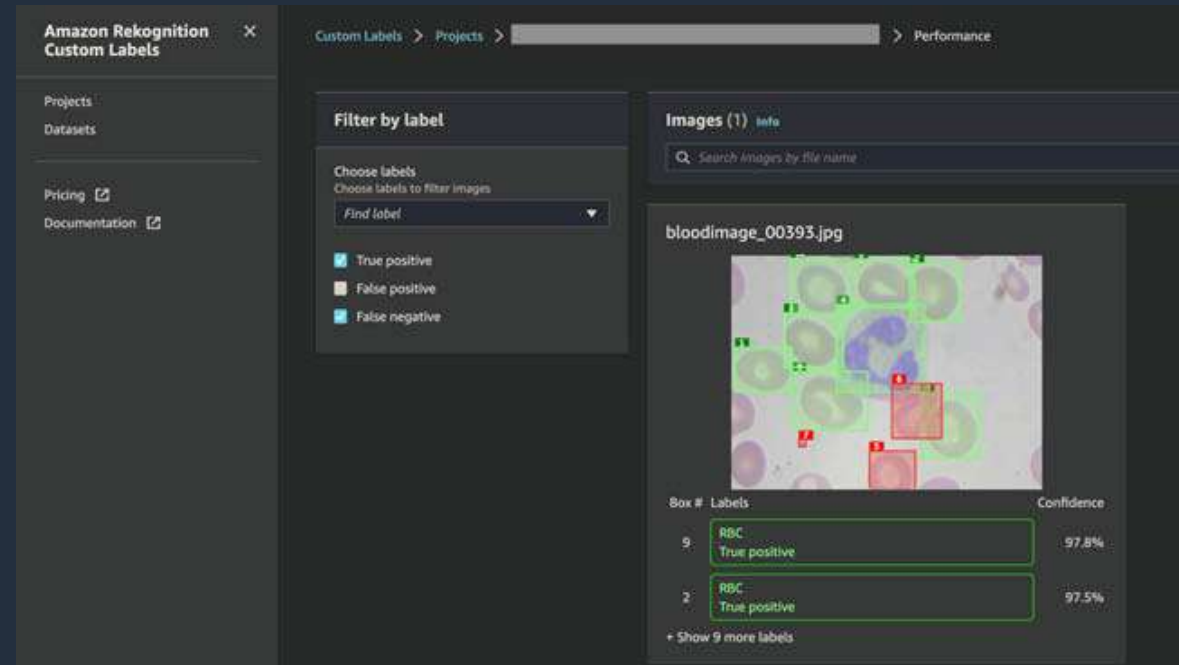


テキストの検出



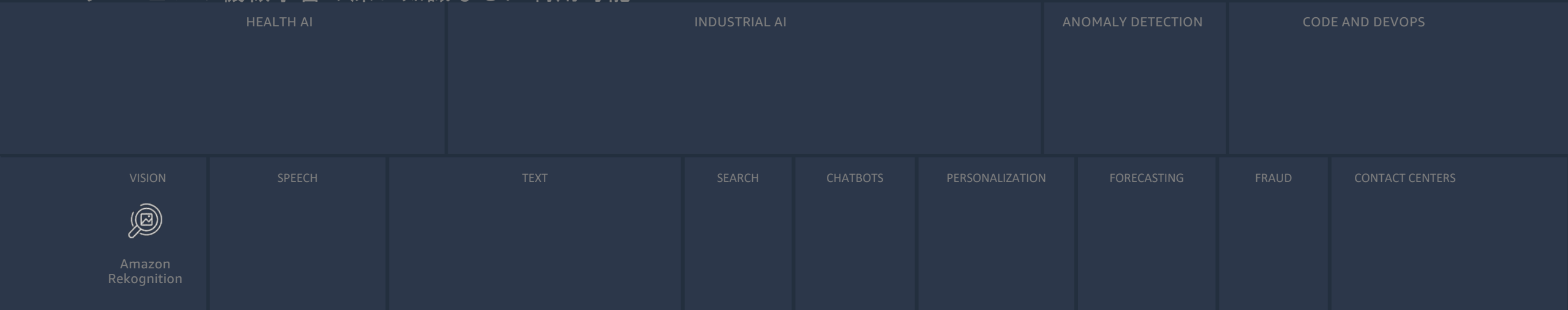
Amazon Rekognition custom Labels

- Amazon Rekognitionに独自の物体検出・シーン認識モデルを構築
 - 認識したい物体・シーンの画像を利用して用途に合った独自モデルが作成可能
- モデル作成にあたり、画像（データ）が少なくても済む
 - Amazon Rekognitionの学習済みモデルをベースにしている



AWS の機械学習スタック

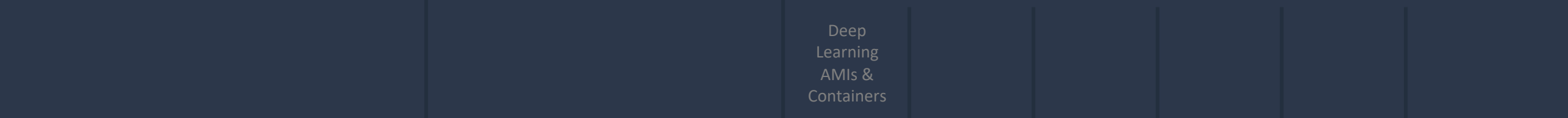
AI サービス: 機械学習の深い知識なしに利用可能



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



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



Amazon SageMaker でできること



準備

構築

トレーニング&チューニング

デプロイ&管理

機械学習のための Web ベース統合開発環境 (IDE)

モデルの自動構築・トレーニング

完全マネージドな
データ処理ジョブと
ラベリングの
ワークフロー

101011010
010101010
000011110

データ収集・準備

1クリックの
コラボレーションと
ビルトインの高性能な
アルゴリズム・モデル



機械学習
アルゴリズム構築

1クリックで
トレーニング



トレーニング環境の
セットアップ

デバッグ・最適化



モデルの
トレーニング・
デバッグ・
チューニング

可視化して
トラッキング・比較



実験管理

1クリックで
デプロイと
オートスケーリング



本番環境への
デプロイ

コンセプト
ドリフトの
自動検知



モデルの
監視

予測結果に
人手のレビュー



推論結果の検証
本番環境での
スケーリングと管理

オートスケーリング
によるコスト削減



Amazon SageMaker でできること



準備

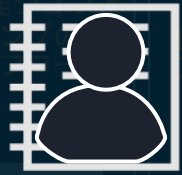
構築

トレーニング&チューニング

デプロイ&管理

機械学習のための Web ベース統合開発環境 (IDE)

モデルの自動構築・トレーニング



大規模なコラボレーション

コードの依存関係を追跡することなくノートブックを共有



簡単な実験管理

何千もの実験を整理、追跡、比較



自動モデル生成

コードを記述せずに、完全な可視性と制御を備えた正確なモデルを取得



高品質のMLモデル

エラーを自動的にデバッグし、モデルを監視し、高品質を維持

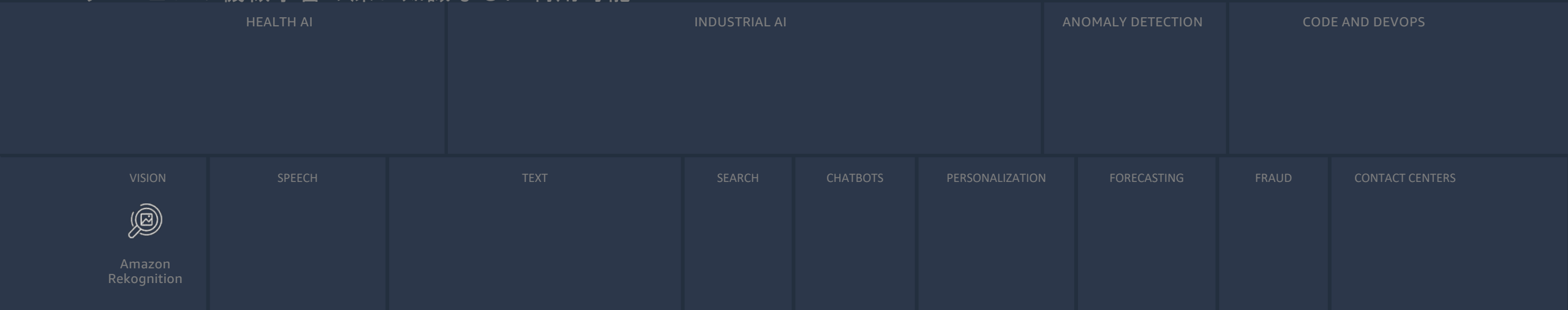


生産性の向上

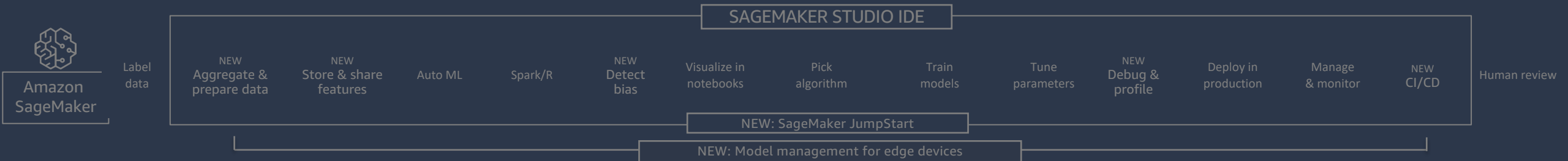
統合されたビジュアルインターフェイスでのコーディング、構築、トレーニング、展開、および監視

AWS の機械学習スタック

AI サービス: 機械学習の深い知識なしに利用可能



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



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

Deep Learning AMIs & Containers

AWS Deep Learning AMIs



深層学習アプリケーションをすばやく構築するために事前に構成された環境



Amazon EC2



AWS Deep Learning AMI

- 機械学習向け環境を事前インストール済み
 - NVIDIA CUDA, cuDNN, Anaconda Platformなど
- Deep Learning Frameworkも事前設定済み
 - TensorFlow, PyTorch, Apache MXNet, Chainer, Keras, Gluonなど
- 料金は選択したリソースの利用時間の課金のみ

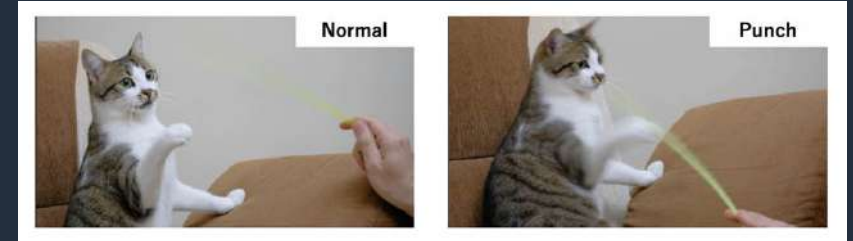


Amazon Rekognition custom Labelsを利用した行動検出

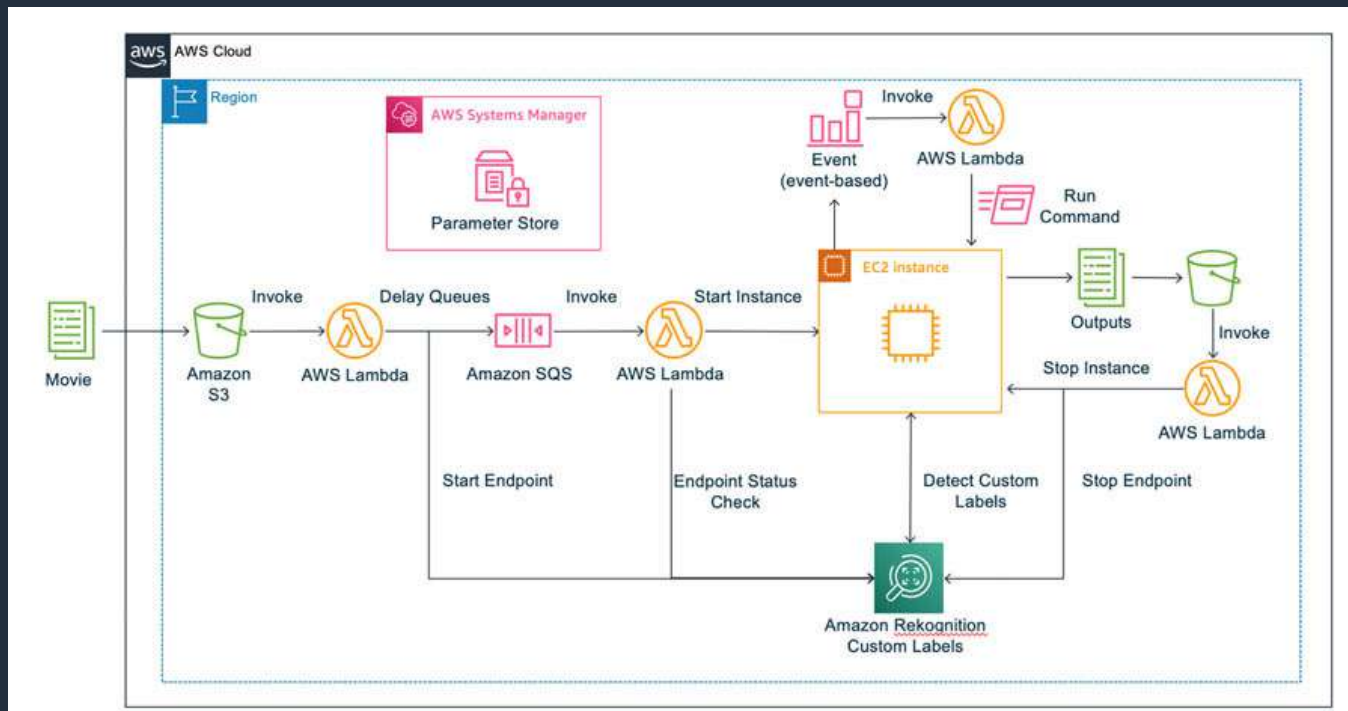
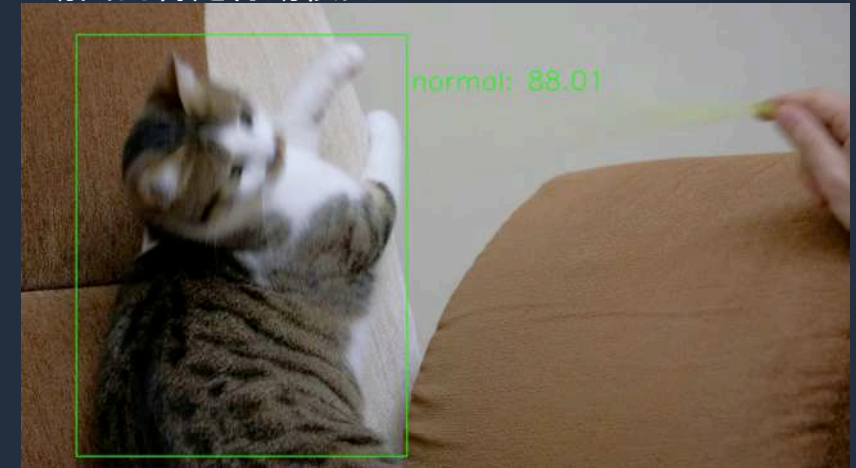
動画ファイルをアップロードするだけで、動画内の特定行動の検出までのプロセスを自動化

➤ イベント駆動アーキテクチャのため処理フロー制御など不要

画像分析モデルのトレーニング用画像



動画内特定行動検出

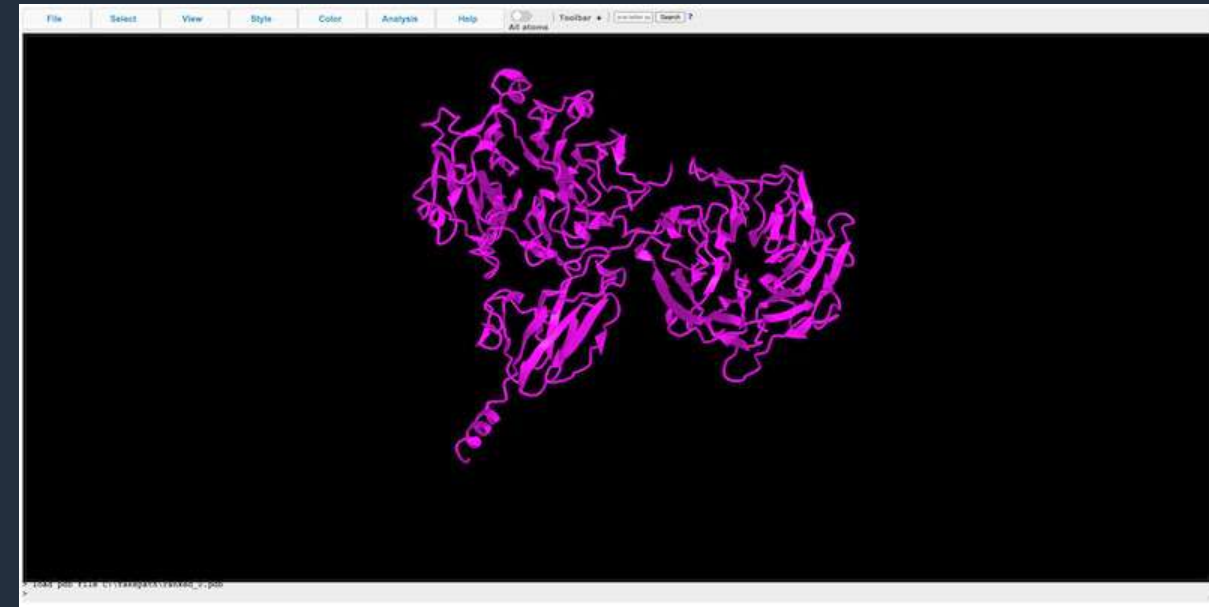
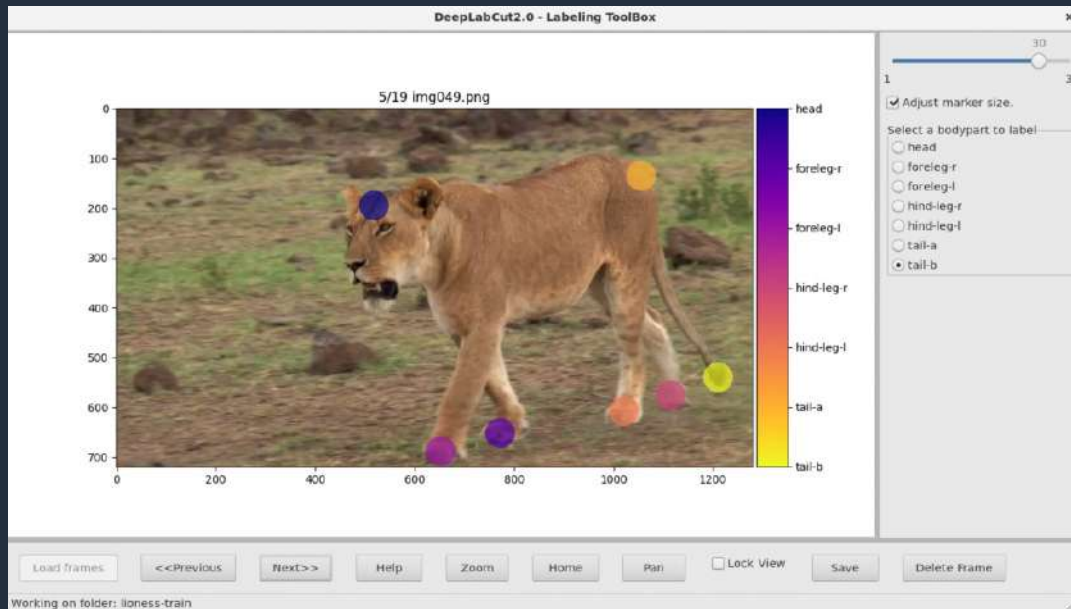


<https://aws.amazon.com/jp/blogs/news/detecting-playful-animal-behavior-in-videos-using-amazon-rekognition-custom-labels/>

AWS Deep Learning AMI を利用した環境構築

[AWS Deep Learning AMI を利用したDeepLabCut環境構築](#)

[Amazon EC2でのAlphaFold v2.0の実行](#)



<https://aws.amazon.com/jp/blogs/news/life-sciences-dlc-env/>

<https://aws.amazon.com/jp/blogs/news/run-alphafold-v2-0-on-amazon-ec2/>

データ管理・コラボレーション基盤

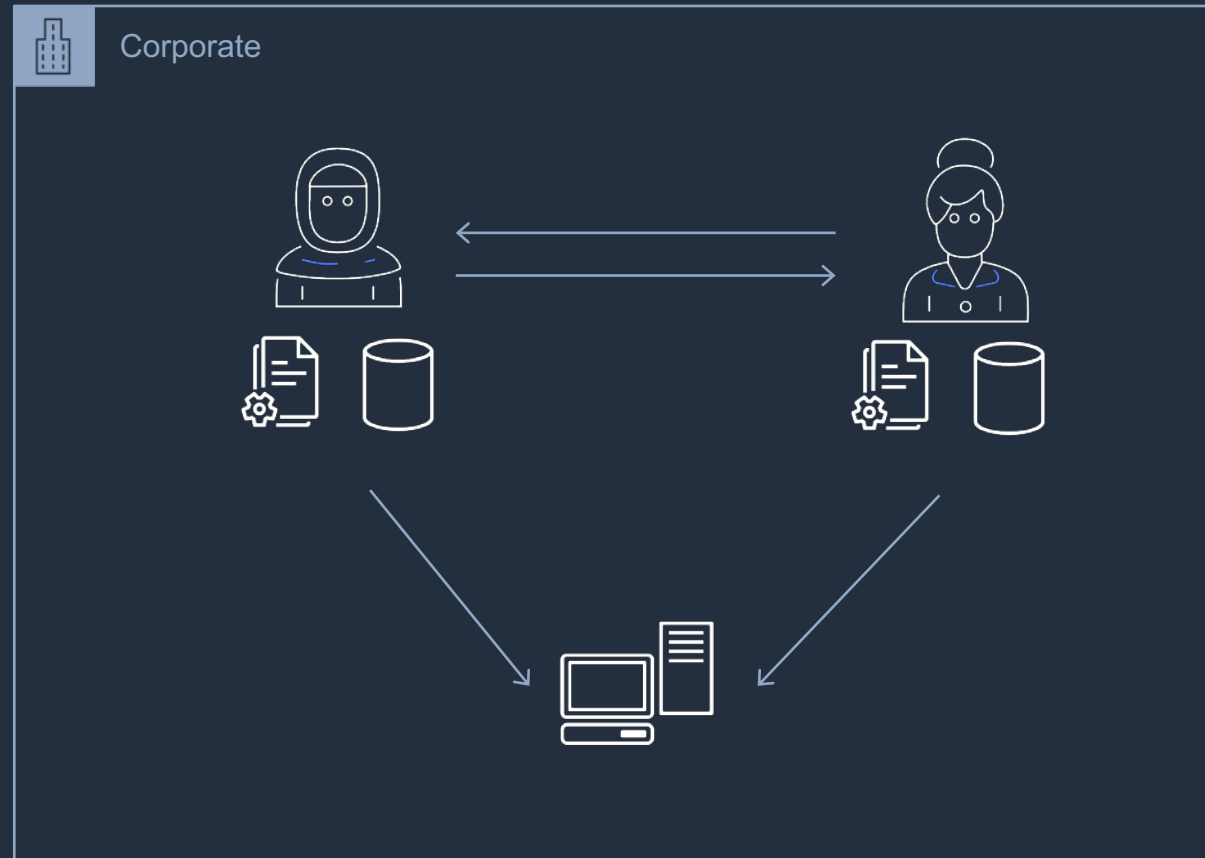
HPC活用・ゲノム解析

機械学習の活用

データ管理・コラボレーション基盤

これまでのHPCの活用や機械学習の活用

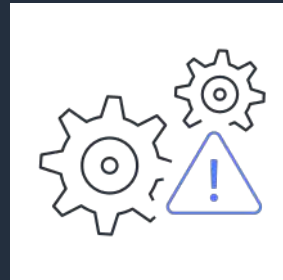
計算環境や各種データを共有しながら研究



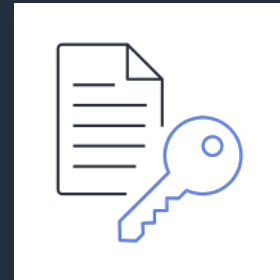
共同研究・コラボレーションにおける課題



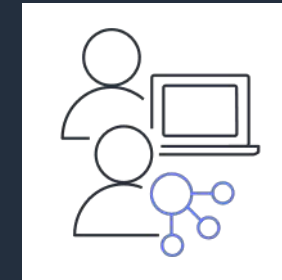
膨大なデータが分散・サイロ化



再現性



ガバナンス統制・セキュリティ



管理コスト

本日はご紹介するデータ管理・コラボレーション基盤に関するAWSサービス

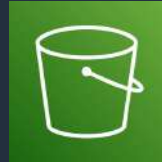
膨大なデータが分散・サイロ化への対応



AWS DataSync



AWS Snowcone



Amazon Simple Storage
Service (Amazon S3)

ガバナンス・セキュリティ・
再現性・管理コストへの対応

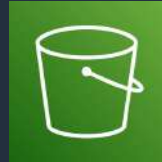


Service Workbench on AWS

本日はご紹介するデータ管理・コラボレーション基盤に関するAWSサービス

膨大なデータが分散・サイロ化への対応

ガバナンス・セキュリティ・
再現性・管理コストへの対応



Amazon Simple Storage
Service (Amazon S3)

Amazon Simple Storage Service (Amazon S3)



- **容量無制限、サイジング不要**

1オブジェクトあたり5TBまで、オブジェクト数は無制限

- **自動的に3箇所以上のAZ*へ隔地保管**

- **設計上のデータ耐久性は、99.999999999%**

- **データ容量に依存しない、スケーラブルで安定した性能**

※ AZは物理的に距離の離れたデータセンター群

※ オブジェクトは自動的に3ヶ所以上のAZへ隔地保存

Amazon Simple Storage Service (Amazon S3)



- 暗号化キーによるサーバー側の暗号化
- IAM、バケットポリシー、S3 アクセスポイントなどによる細かなアクセス制御が可能
- アクセスログも取得

本日はご紹介するデータ管理・コラボレーション基盤に関するAWSサービス

膨大なデータが分散・サイロ化への対応



AWS DataSync



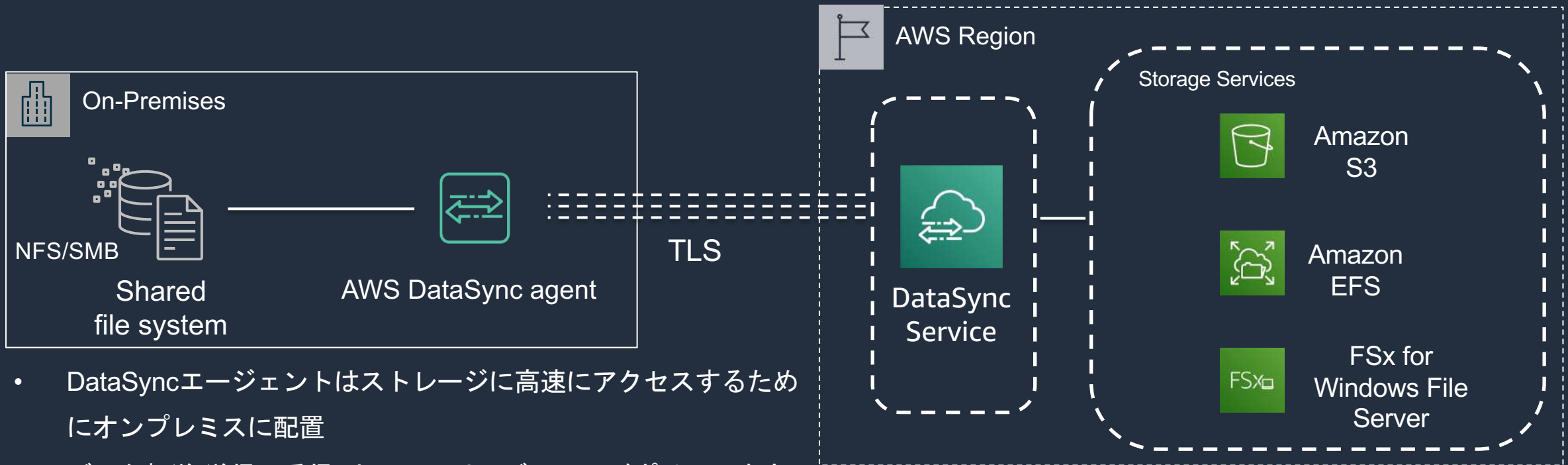
AWS Snowcone

ガバナンス・セキュリティ・
再現性・管理コストへの対応

AWS DataSync



オンプレミスストレージとクラウドのデータ転送をシンプルに、自動で、高速に実行する
オンライン転送サービス



- DataSyncエージェントはストレージに高速にアクセスするためにオンプレミスに配置
- データ転送(送信・受信)はAWSのサービスエンドポイントを介して実行
- 通信時はDataSync用のプロトコルを利用して高速に実施
- マネジメントコンソールやCLI/SDKで管理できる

AWS Snowcone

物理的なエッジコンピューティング、エッジストレージ、およびデータ転送デバイス
研究データを貯め、クラウドに転送することが可能

- 重さは約 2 キログラム
- 8 テラバイトの使用可能ストレージ
- 2 CPU, 4 GB compute
- 長さ 約 2.5 センチ、高さ約 7.5 センチ
- AC 電源またはオプションのバッテリーで稼働可能
- 東京リージョンに対応！！

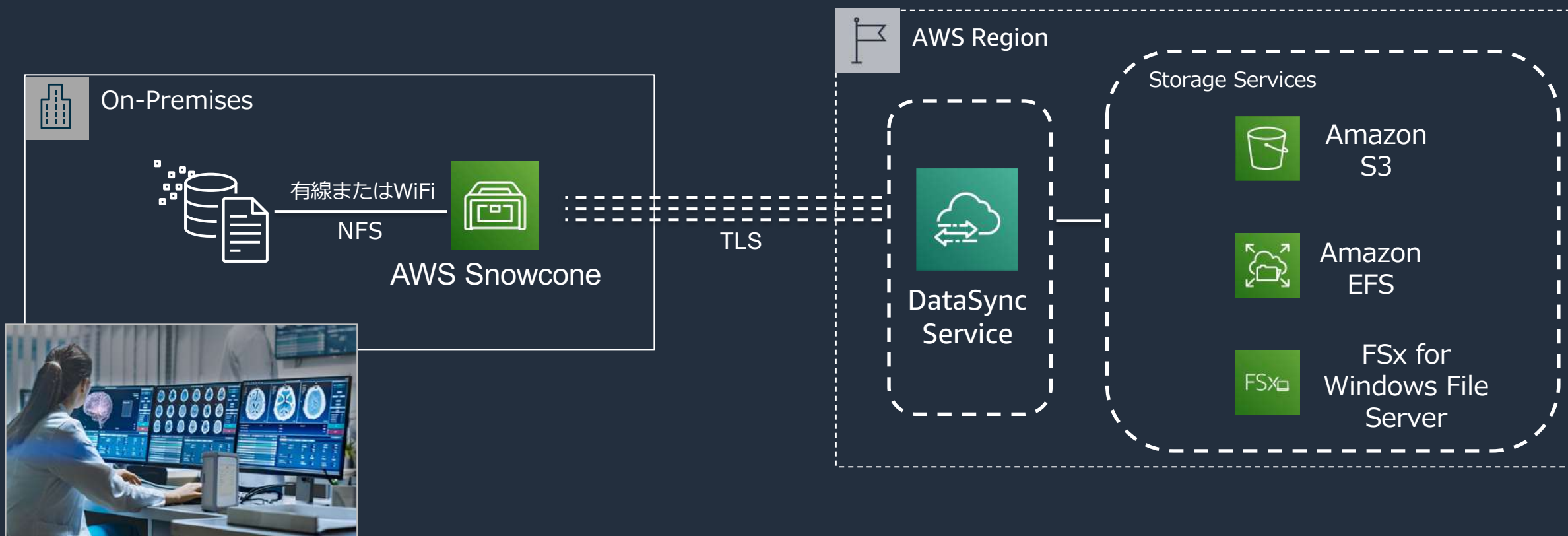


Use cases

デスクトップ、データセンター、メッセージバッグ、自動車などのさまざまな環境、およびドローンと連結させて使用可能。インダストリアルIoT、ヘルスケアIoT、コンテンツ配信、データ移行、ロジスティクスなど。

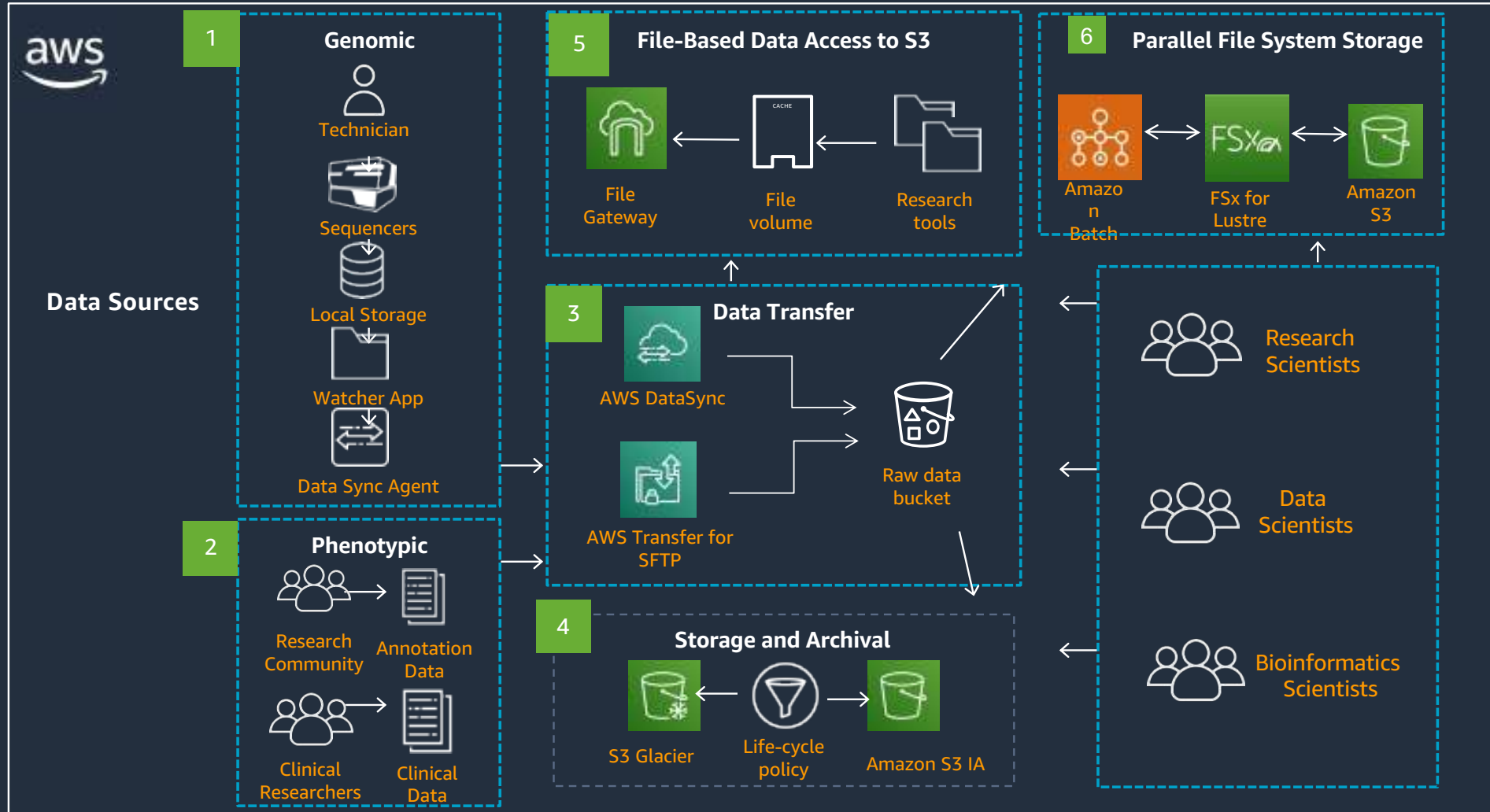
<https://aws.amazon.com/jp/snowcone/>

AWS Snowcone を利用したデータ連携例



創薬研究: ゲノムデータの転送・保管・解析

AWS Reference Architecture



本日はご紹介するデータ管理・コラボレーション基盤に関するAWSサービス

膨大なデータが分散・サイロ化への対応

ガバナンス・セキュリティ・
再現性・管理コストへの対応



Service Workbench on AWS

Service Workbench on AWS

研究者向けのコンピューティング環境をWebポータルとして提供可能なクラウドソリューション

- **ユーザーがオンデマンドで環境を構築可能**
 - 許可されたAWS環境をテンプレートとしてユーザーに提供
 - AWS環境を数分で立ち上げ可能
- **ユーザーによるAWSコンソール操作は不要**
 - ユーザーはService Workbench on AWSが提供するWebポータルサイトからAWS環境にアクセス可能
 - Webポータルサイトを利用頂くことでクラウドの経験がないユーザーも利用しやすい
- **社内外のチーム間で安全にコラボレーション**
 - Webポータルサイト経由で社外にAWS環境を提供可能
 - プロジェクト単位でコストを可視化可能

Galileo - Gateway (demo/us-east-1) Anthony Logout

Research Workspaces 5 Create Research Workspace

All Available Pending Errored Terminated

AVAILABLE

SageMaker-CVD19-ML1
Created 22 hours ago by Anthony 06094dae-6d4b-4ecd-8648-6c00c9ded2d8

Connections View Detail Terminate

HTTP Connections

SageMaker Notebook Connect

COVID-19 modeling using SageMaker

Owner	Anthony
Studies	3
Project	project-1
Restricted CIDR	72.21.198.66/32
Workspace Type	SageMaker Notebook-V1.0.0

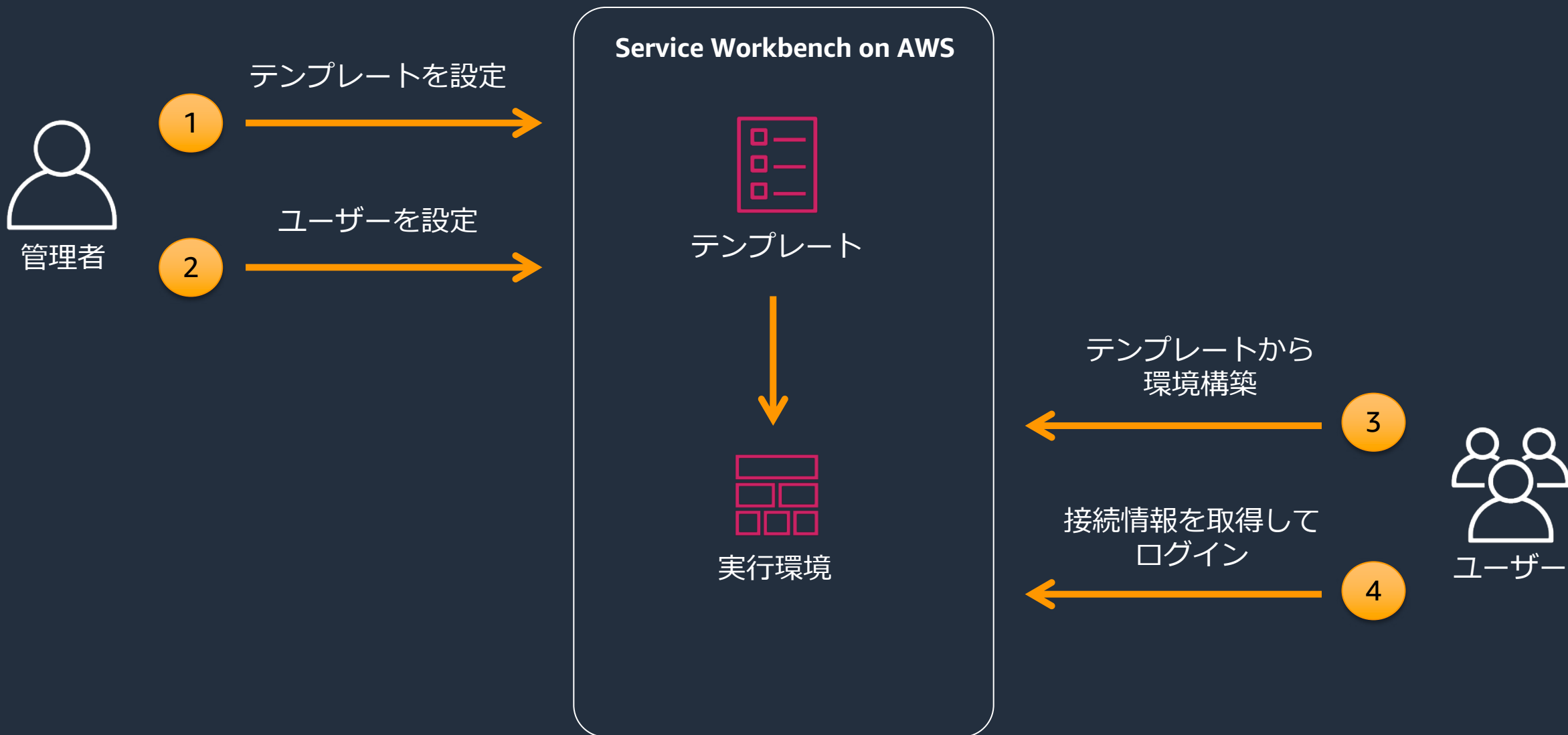
\$0.66
YESTERDAY'S COST

AVAILABLE

Hail-v2-EMR-001
Created 22 hours ago by Anthony 6d3b867e-ebac-49bc-a72d-f2e26f836662

Connections View Detail Terminate

Service Workbench on AWS ご利用の流れ



まとめ

Point 1

- 創薬研究領域においてAWSの利用が増えているユースケースの理解
 - HPC・ゲノム解析
 - 機械学習
 - データ管理・コラボレーション

Point 2

- 各ユースケースでの課題解決に有効なAWSサービスおよびソリューションを知っていただく

さいごに

- AWSでは、さまざまなユースケースやビジネスシーンにおいてご活用頂けるサービスが提供されております。
- サービスを組み合わせることにより様々な事が実現可能です。
- ハンズオンコンテンツやトレーニングもご用意しております。
- 案件のご相談等ございましたら、ぜひお声がけ頂ければ幸いです。

Thank you!