

AWS RE:INVENT

RE:CAP

JAPAN Jan 26, 2023

Amazon Omics のご紹介

中島 文博

Solutions Architect

Amazon Web Services Japan



自己紹介

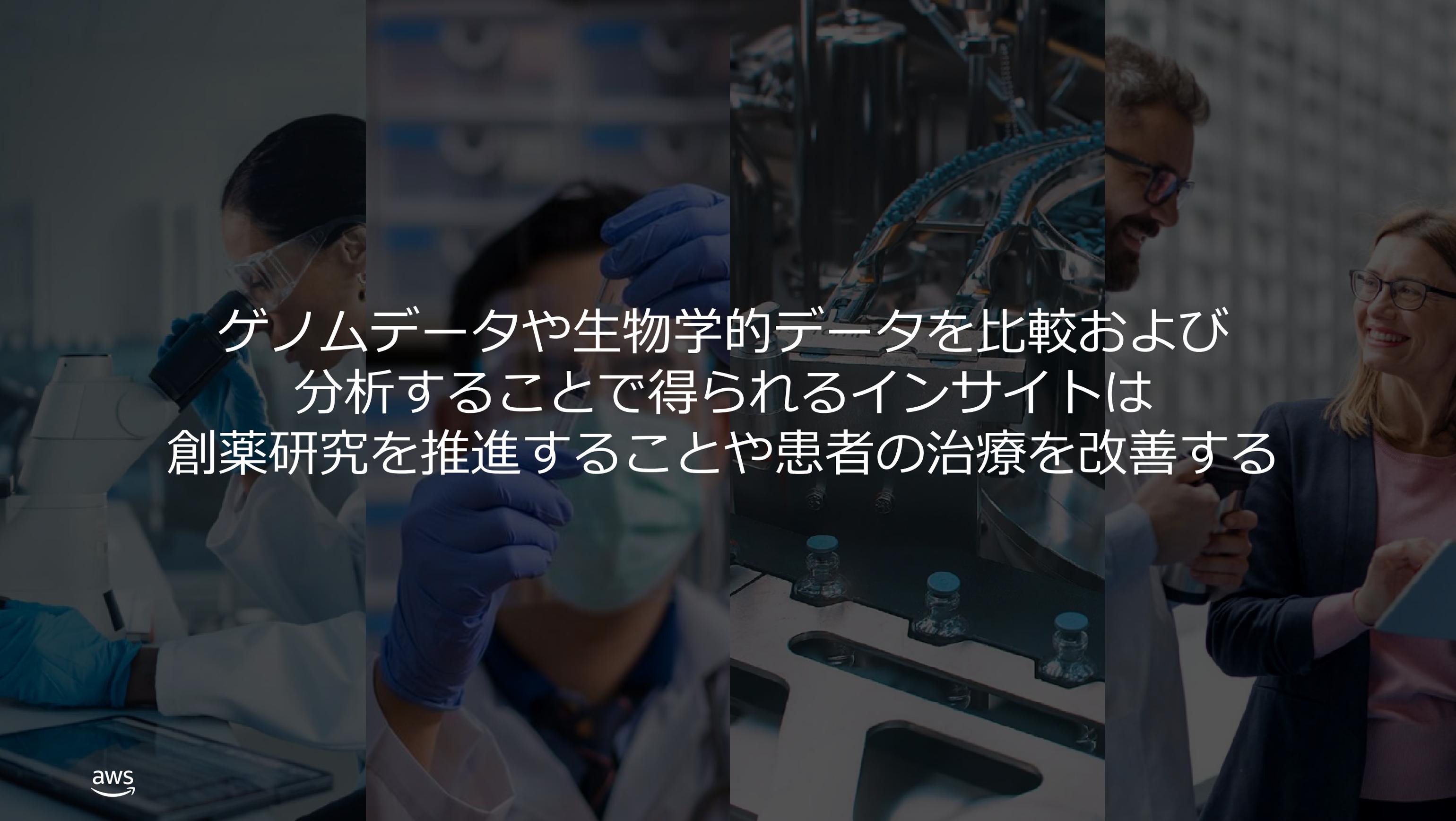


名前：中島 丈博（なかじま たけひろ）

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



ゲノムデータや生物学的データを比較および
分析することで得られるインサイトは
創薬研究を推進することや患者の治療を改善する

ゲノムデータや生物学的データを利用したデータ駆動型創薬における課題

ゲノムデータと生物学的データの比較・分析にはその規模や複雑さのため管理にコストがかかる

今後5年間で数千万の全ゲノムのシーケンスと保存

複数の専用ツールとワークフロー言語への対応

セキュリティ、プライバシー、コンプライアンス要件への対応が必須

Amazon Omics

ヘルスケアやライフサイエンス業界において
ゲノミクス、トランスクリプトミクスなどの
オミックスデータを**保存・検索・分析**し、
そのデータから**インサイトの取得、健康改善**
科学的発見を加速するための専用サービス

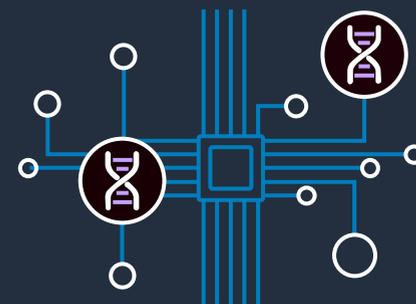
Amazon Omics: 特徴



**Multiomic and
multimodal
analysis**



**Population-level
scale**



**Fully-managed
bioinformatics
computation**



**Built-in security,
privacy, and
compliance**

※ ご利用可能なリージョン

米国東部 (バージニア北部)、米国西部 (オレゴン)、欧州 (アイルランド)、欧州 (ロンドン)
欧州 (フランクフルト)、およびアジアパシフィック (シンガポール)

Amazon Omics: 主な機能



ストレージ

ワークフロー

分析

Amazon Omics: ゲノムデータ分析の流れ



ストレージ

store raw sequencing data

GCTTGGTCA
GATAATGCA *fasta* →  Reference Store
ref. sequence

 *fastq*
bam
cram →  Sequence Store
raw sequence

ワークフロー

run analysis workflows

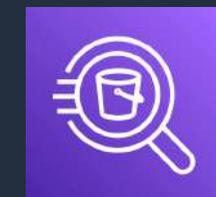
 {wdl}
Workflow Definition Language
 Nextflow
Omics Workflow

分析

store and query variant data

GCTTGGACA
GATCATGCA *vcf*
gvcf →  Variant Store
variants

 *vcf*
gff
tsv →  Annotation Store
annotations

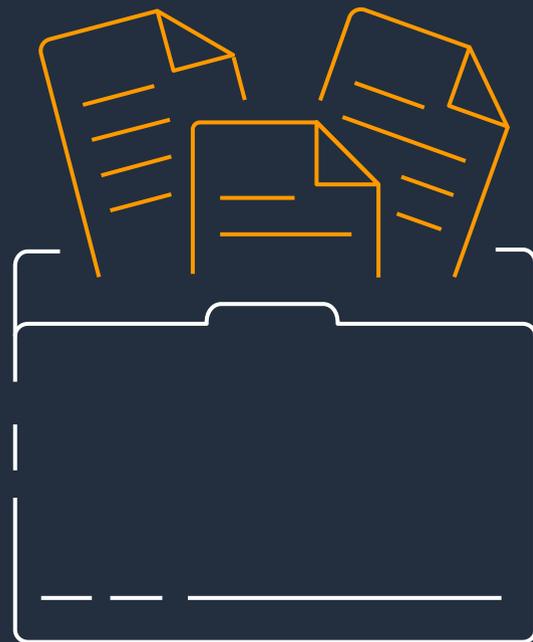


query and interpret
with Amazon Athena

Amazon Omics: ストレージ



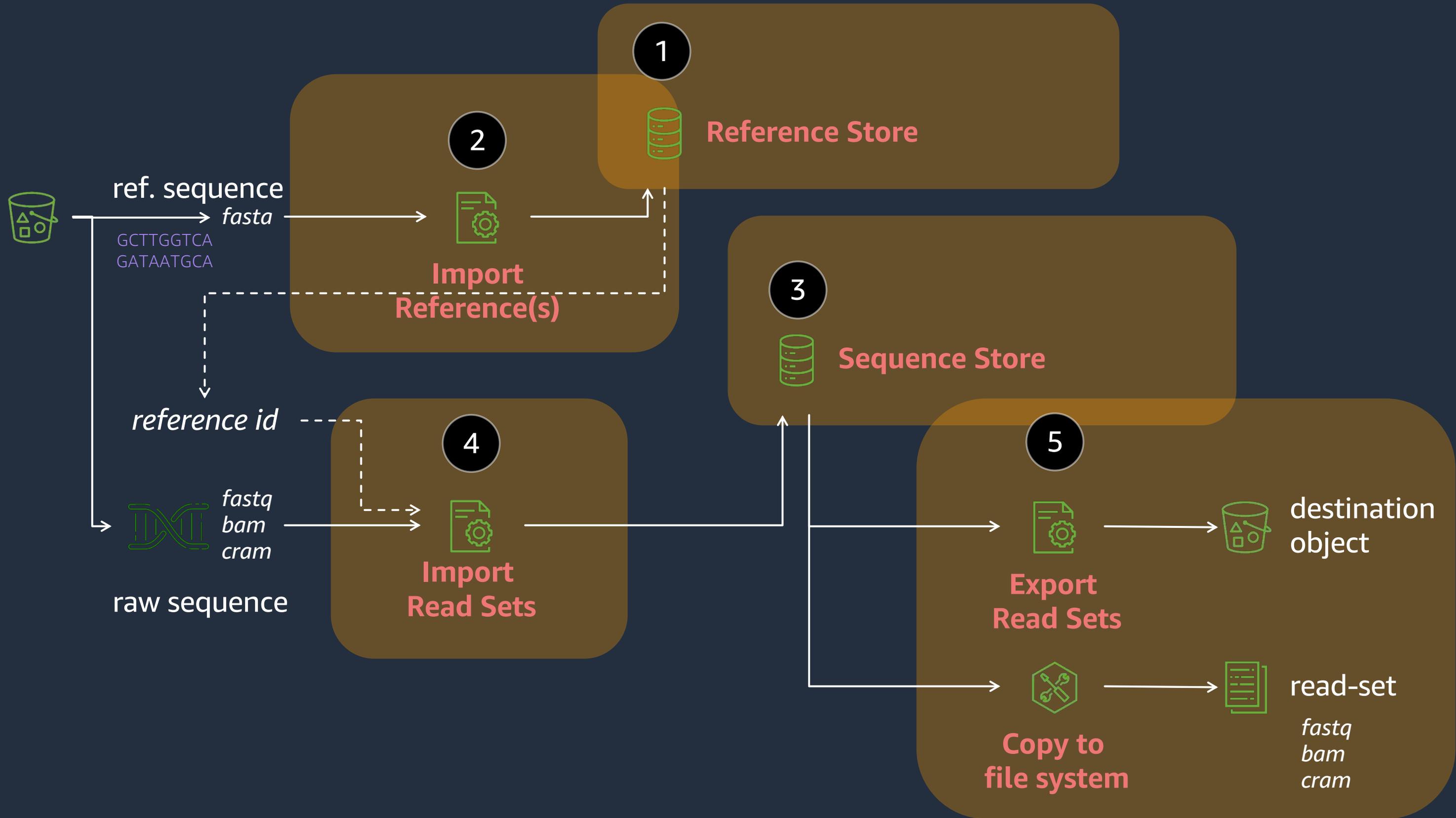
Discover data



Share data



**Low cost per
gigabase**



コンソール画面：リファレンスストア

Amazon Omics ×

開始方法

▼ ストレージ

- 参照ストア
- シーケンスストア

▼ 分析

- バリエーションストア
- 注釈ストア

▼ ワークフロー

- ワークフロー
- 実行
- 実行グループ

ドキュメント

Amazon Omics > 参照ストア

参照ストア Info

詳細 | インポート | タグ

[参照ゲノムをインポート](#)

詳細

名前 参照ストア	Amazon リソースネーム (ARN) arn:aws:omics:us-east-referenceStore/	作成時刻 December 19, 2022, 11:26 (UTC -5)
参照ストア ID		

参照ゲノム (2) Info

🔍 参照ゲノムを検索 < 1 >

名前	インポートステータス	説明	作成済み
GRCh-hg38	✔ Active	-	December 19, 2022, 14:33
hg38	✔ Active	Genome Reference Consortium Human Reference 38 (hg38)	December 19, 2022, 11:27

コンソール画面：シーケンスストア

Amazon Omics ×

Amazon Omics > シーケンスストア > MySequenceStore > 2684118187

2684118187 [Info](#) 削除

読み取りセット詳細

読み取りセット ID 2684118187	サブジェクト ID mySubject	整列 UNALIGNED	読み取りセット URI omics:/storage.us-east-1.amazonaws.com/9125833763/readSet/2684118187
読み取りセット名 HG00146	サンプル ID mySample	生成元 -	ステータス アクティブ
説明 -	ソースファイルタイプ FASTQ	参照名 GRCh-hg38	作成時刻 December 19, 2022, 14:59 (UTC -5)
シーケンスストア MySequenceStore	ベースの合計数 677717384	Amazon リソースネーム (ARN) arn:aws:omics:us-east-1:sequenceStore/9125833763/readSet/2684118187	

タグ (2) [Info](#) タグを管理

Key	Value
aws:omics:subjectId	mySubject
aws:omics:sampleId	mySample

開始方法

▼ ストレージ

- 参照ストア
- シーケンスストア

▼ 分析

- バリエーションストア
- 注釈ストア

▼ ワークフロー

- ワークフロー
- 実行
- 実行グループ

ドキュメント

データのエクスポート

- S3へのエクスポート : start-read-set-export-job API

```
$ aws omics start-read-set-export-job --sequence-store-id (sequence-store-id) \  
--destination (valid-s3-uri) \  
--role-arn (role-arn) \  
--sources readSetId=(read-set-id-1) readSetId=(read-set-id-2)
```

<https://docs.aws.amazon.com/omics/latest/dev/read-set-exports.html>

- ファイルシステムへのエクスポート : Omics Transfer Manager

```
import boto3  
from omics.transfer import ReferenceFileName, ReadSetFileName  
from omics.transfer.manager import TransferManager  
from omics.transfer.config import TransferConfig  
  
REFERENCE_STORE_ID = "<my-reference-store-id>"  
SEQUENCE_STORE_ID = "<my-sequence-store-id>"  
  
client = boto3.client("omics")  
manager = TransferManager(client)  
  
# Download all files for a reference.  
manager.download_reference(REFERENCE_STORE_ID, "<my-reference-id>")  
  
# Download all files for a read set to a custom directory.  
manager.download_read_set(SEQUENCE_STORE_ID, "<my-read-set-id>", "my-sequence-data")
```

<https://github.com/aws-labs/amazon-omics-tools>

Amazon Omics: ワークフロー



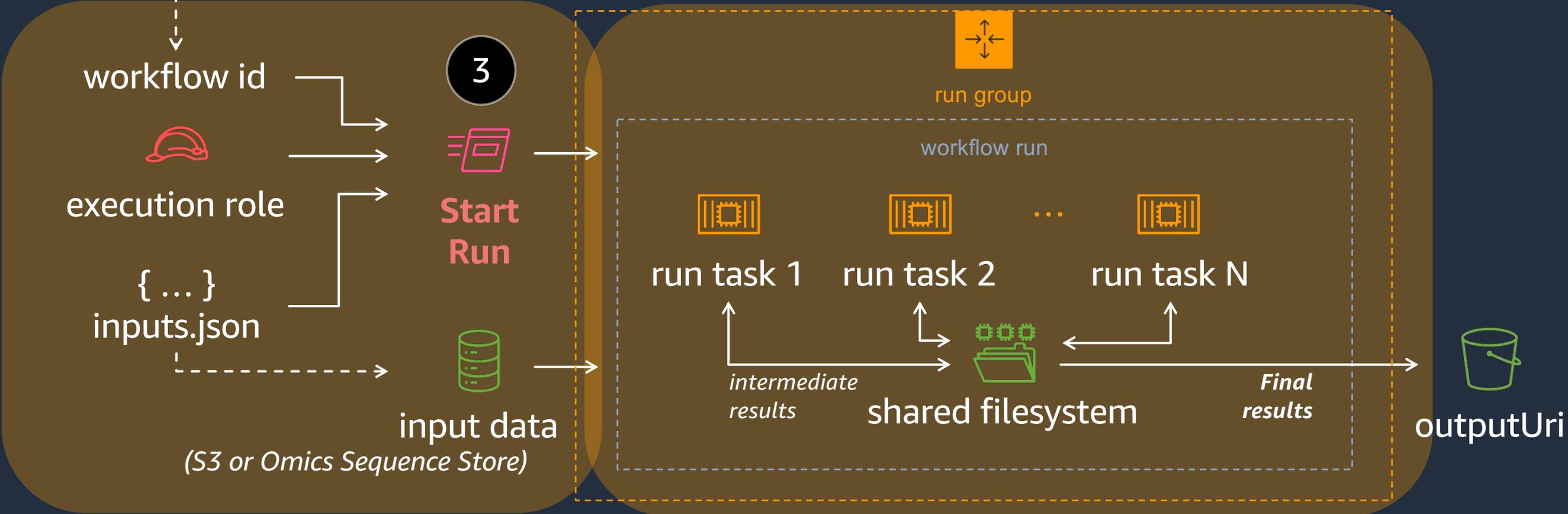
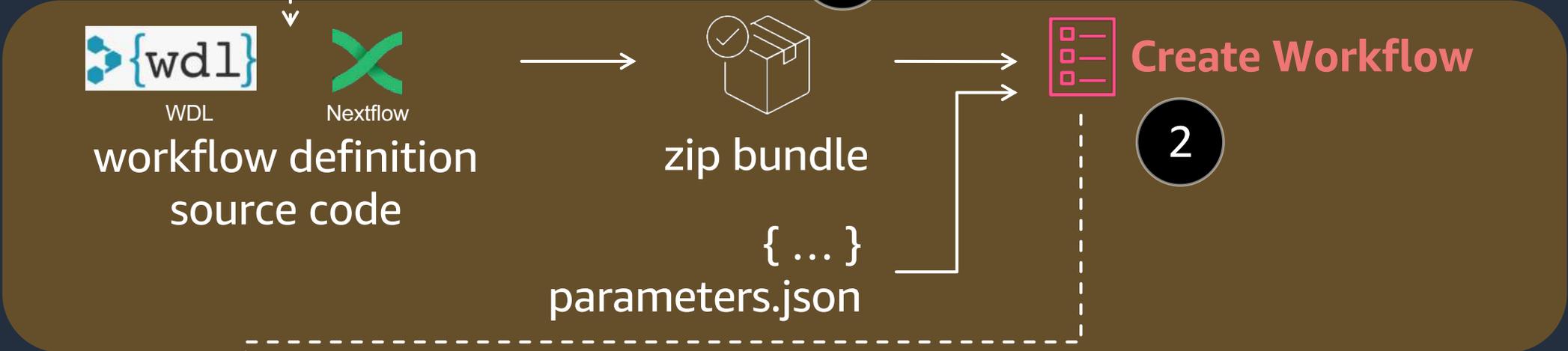
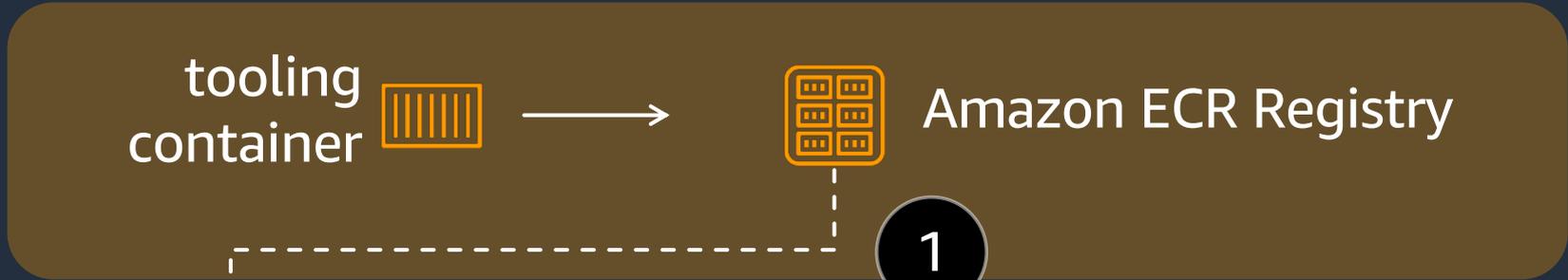
**Science not
infrastructure**



**Pay-as-you-go
pricing**



**Predictable cost
per workflow**



コンソール画面：ワークフロー

Amazon Omics ×

Amazon Omics > ワークフロー > 5905905

MyWDLWorkflow Info

[詳細](#) | [Run history](#) | [タグ](#)

[削除](#) [ダウンロード](#) [ワークフローを実行する](#)

Details

ワークフロー ID	タイプ	Amazon リソースネーム (ARN)
5905905	Private	 arn:aws:omics:us-east-1: :workflow/5905905
名前	ワークフロー定義ファイル	作成済み
MyWDLWorkflow	main.wdl	January 06, 2023, 11:21 (UTC -5)
説明	実行ストレージ容量 (TB)	ステータス
My WDL Workflow	-	 アクティブ
ワークフロー言語		
WDL		

パラメータ

名前	説明	必須
image	Optional ECR image	No
input_txt_file	Required input file	Yes

コンソール画面：実行

☰

2309281 is currently running...
選択 実行をキャンセル このワークフローの実行を停止します。

実行をキャンセル

0/0 0%

Amazon Omics > 実行 > 2309281

削除 runs.details.clone 再実行

Info

詳細 タスク タグ

詳細 runs.details.cloudWatch

実行 ID 2309281	実行ストレージキャパシティ 1.2 TB	Amazon リソースネーム (ARN) arn:aws:omics:us-ea: :run/2309281
実行名 -	ワークフロー ID 5905905	開始時刻 January 06, 2023, 11:23 (UTC -5)
実行グループ ID -	ワークフロー名 MyWDLWorkflow	停止時間 -
実行グループ名 -	ワークフロータイプ Private	ステータス 実行中
実行の優先度 -	出力 URI https://s3.console.aws.amazon.com/s3/buckets/ us-east-1/amazon-omics/output-files/	

パラメータ

名前	説明	値
input_txt_file	-	"s3://ast-1/amazon-omics/input-files/hello_world.txt"

コンソール画面：実行

☰

🔄 2309281 is currently running...
選択 実行をキャンセル このワークフローの実行を停止します。

1/1 0%

Amazon Omics > 実行 > 2309281

Info 削除 runs.details.clone 再実行

詳細 | **タスク** | タグ

タスク (1) Info
runs.runsTable.description

🔍 runs.runsTable.filteringPlaceholder < 1 > ⚙️

runs.runTasks.id ▼	名前	ステータス	runs.runTasks.logs ▼	runs.runTasks.runTime ▼	runs.runTasks.cpu	runs.runTasks.memory	runs.runTasks.start ▼	runs.runTasks.stop
7929516	TxtFileCopyTask	🔄 開始中	runs.runTasks.viewButtonText	-	1	1		

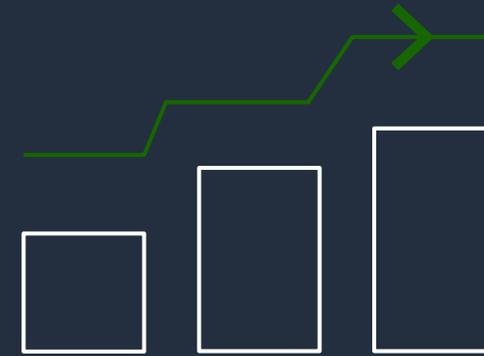
Amazon Omics: 分析



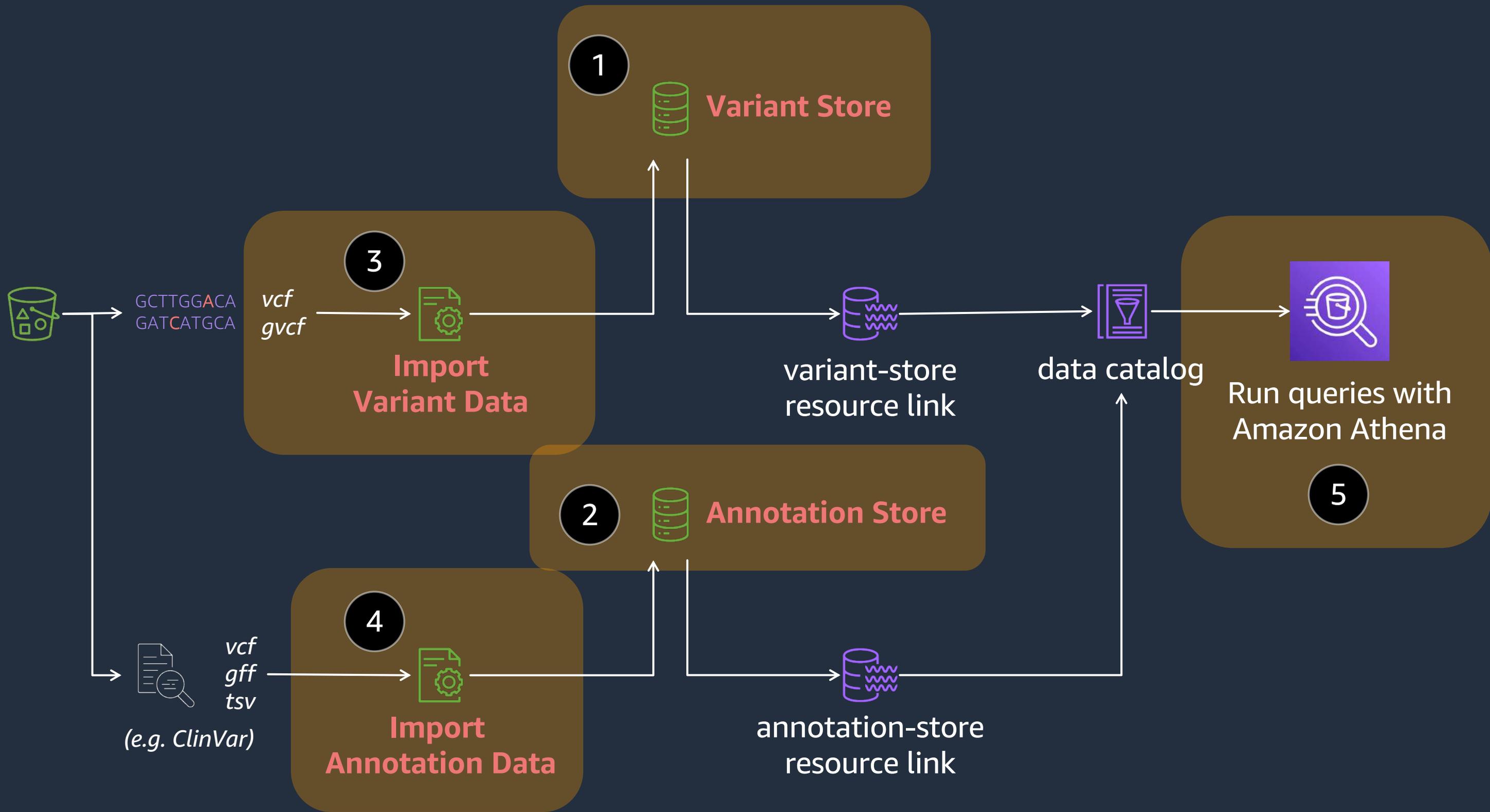
**Ingests variant
and annotation
data**



**Allows
fine-grained
access control**



**Scales
with you**



コンソール画面：バリエーションストア

Amazon Omics ×

Amazon Omics > バリエーションストア > variantstore_20230106

variantstore_20230106 Info

[アクション](#) [許可を管理](#) [データを表示](#)

Details

名前 variantstore_20230106	参照ゲノム GRCh-hg38	Amazon リソースネーム (ARN) arn:aws:omics:us-east-1: :variantStore/variantstore_20230106
説明 -	KMS キー AWS が所有するキー	作成時刻 January 06, 2023, 13:07 (UTC -5)
ストア ID 9adf3dd30a3c	ステータス ✔ アクティブ	更新日時 January 06, 2023, 13:15 (UTC -5)
データサイズ 2577218 Bytes		

[インポート](#) | [タグ](#)

インポートジョブ (1) Info

[キャンセル](#)

< 1 > [設定](#)

ジョブ ID	インポートステータス	作成済み	完了済み
4ead2306-7f3e-4afa-b994-83d14e5095ab	✔ 完了済み	January 06, 2023, 13:11	January 06, 2023, 13:15

コンソール画面：アノテーションストア



Amazon Omics > 注釈ストア > annotationstore_20230106

annotationstore_20230106 [Info](#)

アクション ▼

許可を管理 [🔗](#)

データを表示 [🔗](#)

Details

名前

annotationstore_20230106

説明

-

ストア ID

1b1cd68eaf53

データサイズ

55501479 Bytes

参照ゲノム

[GRCh-hg38](#)

KMS キー

AWS が所有するキー

ステータス

🟢 アクティブ

Amazon リソースネーム (ARN)

[🔗](#) arn:aws:omics:us-east-1:annotationStore/annotationstore_20230106

作成時刻

January 06, 2023, 17:00 (UTC -5)

最終アクセス日

January 06, 2023, 17:30 (UTC -5)

[インポート](#)

[スキーマ](#)

[タグ](#)

インポートジョブ (1) [Info](#)

ストアは VCF ソースタイプで作成されます。インポートされるすべてのデータは、同じ形式である必要があります。

🔍 インポートジョブを検索

< 1 > ⚙️

ジョブ ID



インポートステータス



作成済み (UTC+3:30)



完了済み (UTC+ 3:30)



45d81ab2-7069-4bdc-bd2f-6a52a1f52d5d

🟢 完了済み

January 06, 2023, 17:26

January 06, 2023, 17:30

コンソール画面：Amazon Athena クエリエディタ

Amazon Athena > クエリエディタ

エディタ | 最近のクエリ | 保存したクエリ | 設定 ワークグループ version3

データ 🔄 < クエリ1 : ✕ 🟢 クエリ2 : ✕ + ▼

データソース: AwsDataCatalog

データベース: omicsdb

テーブルとビュー 作成 ▼ ⚙️

🔍 テーブルとビューをフィルタリング

▼ テーブル (1) < 1 >

- variant_20230106
 - importjobid string
 - contigname string
 - start bigint
 - end bigint
 - names array<string>
 - referenceallele string
 - alternatealleles array<string>
 - qual double
 - filters array<string>
 - splitfrommultiallelic boolean

クエリ 1 : ✕ 🟢 クエリ 2 : ✕

```
1 SELECT * FROM "omicsdb"."variant_20230106" limit 10;
```

SQL 行 1、列 1 ☰ 🔍 ⚙️

もう一度実行する Explain キャンセル クリア 作成 ▼ 🔴 クエリ結果を再利用する (60分前まで)

🟢 完了済み キュー内の時間: 249 ミリ秒 実行時間: 10.715 秒 スキャンしたデータ: 2.42 MB

結果 (10) 📄 コピー 結果をダウンロード

🔍 行を検索 < 1 > ⚙️

#	importjobid	contigname	start	end	names	referenceallele	alternatealleles	qual
1	4ead2306-7f3e-4afa-b994-83d14e5095ab	chr1	10000	10391		T	[<NON_REF>]	
2	4ead2306-7f3e-4afa-b994-83d14e5095ab	chr1	257666	277717		G	[<NON_REF>]	
3	4ead2306-7f3e-4afa-b994-83d14e5095ab	chr1	277717	277800		A	[<NON_REF>]	
4	4ead2306-7f3e-4afa-b994-83d14e5095ab	chr1	277800	277801		G	[<NON_REF>]	
5	4ead2306-7f3e-4afa-b994-83d14e5095ab	chr1	277801	277833		G	[<NON_REF>]	
6	4ead2306-7f3e-4afa-b994-83d14e5095ab	chr1	10391	10397		C	[<NON_REF>]	

Amazon Omics: ゲノムデータ分析の流れ (再掲)



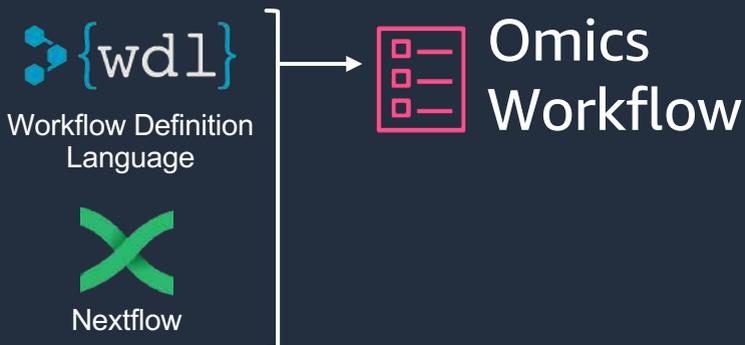
ストレージ

store raw sequencing data



ワークフロー

run analysis workflows



分析

store and query variant data



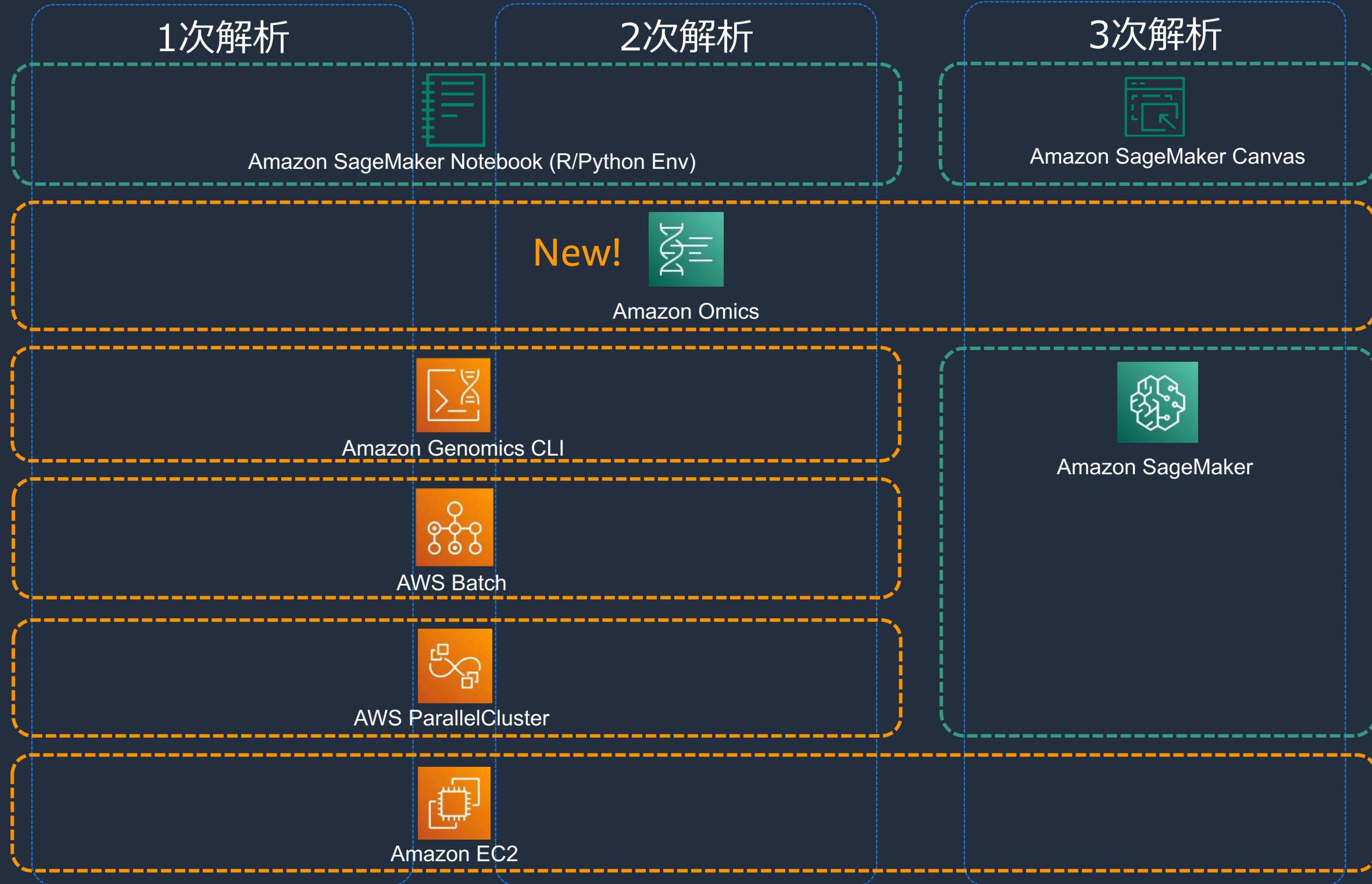
query and interpret
with Amazon Athena

ゲノム解析フェーズにおけるAWSサービスのマッピング

Managed



Self-service



Amazon Omics と Amazon Genomics CLI の使い分け



Amazon Omics

HIPAA 適格の大規模解析環境を
フルマネージドで
独立した AWS サービスとして提供

- オミックスデータから洞察を得るための環境をフルマネージドサービスとして提供
- データの保存、検索、分析機能
- HIPAA 適格
- マルチモーダル解析のためのアナリティクス機能
- [ワークフロー] サービスが対応するインスタンスタイプやストレージ容量からある程度規模の大きな分析向け



Amazon Genomics CLI

AWS サービスを組み合わせた高い柔軟性をもつ
バイオインフォマティクスワークフロー
実行環境を数分で AWS 上に構築

- AWS サービスを IaC (CDK) で抽象化し、研究者が CLI で操作するクラウド解析環境を手軽に手に入れることを支援する OSS ツール
- 規制準拠は別途対応が必要
- CLI 利用環境でのローカル実行とのシームレスな連携
- [ワークフロー] インスタンスタイプなどは背後の AWS サービスの制限に従うため、検証やテスト用途の規模の小さな解析から大規模解析まで幅広く対応

Amazon Omics: まとめ

- Amazon Omics は、ヘルスケアおよびライフサイエンスを担う組織を支援するフルマネージドなサービス
- 主な機能は3つ
 1. ストレージ：VCF、GFF3、TSV/CSV などのゲノミクスデータを効率的かつ低コストで保存、取得、整理、共有することが可能
 2. ワークフロー：WDL または Nextflow で定義されているワークフローを使用してゲノミクスデータを処理および分析可能
 3. 分析：バリアントデータとアノテーションデータの保存と分析が可能
- ゲノム解析における1次解析から3次解析までの利用を想定

AWS re:Invent 2022

[NEW LAUNCH!] Transform omics data into insights with Amazon Omics

AIM219

Introducing Amazon Omics

Aaron Friedman, PhD (he/him)

Principal Product Manager, Amazon Omics
AWS



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

<https://www.youtube.com/watch?v=46SLOjuMM6o>

Amazon Omics: 関連ブログ

Amazon Omics のご紹介 – ゲノムデータおよび生物学的データを大規模に保存、クエリ、分析するための専用サービス

<https://aws.amazon.com/jp/blogs/news/introducing-amazon-omics-a-purpose-built-service-to-store-query-and-analyze-genomic-and-biological-data-at-scale/>

Part 1: Amazon Omics のご紹介 – セキュアかつ大規模に生物配列データから洞察を得る

<https://aws.amazon.com/jp/blogs/news/part-1-introducing-amazon-omics-from-sequence-data-to-insights-securely-and-at-scale/>

Part 2: Amazon Omics を使用したゲノミクスデータの保管と分析のエンドツーエンド自動化

<https://aws.amazon.com/jp/blogs/news/automated-end-to-end-genomics-data-storage-and-analysis-using-amazon-omics/>

Thank you!

