

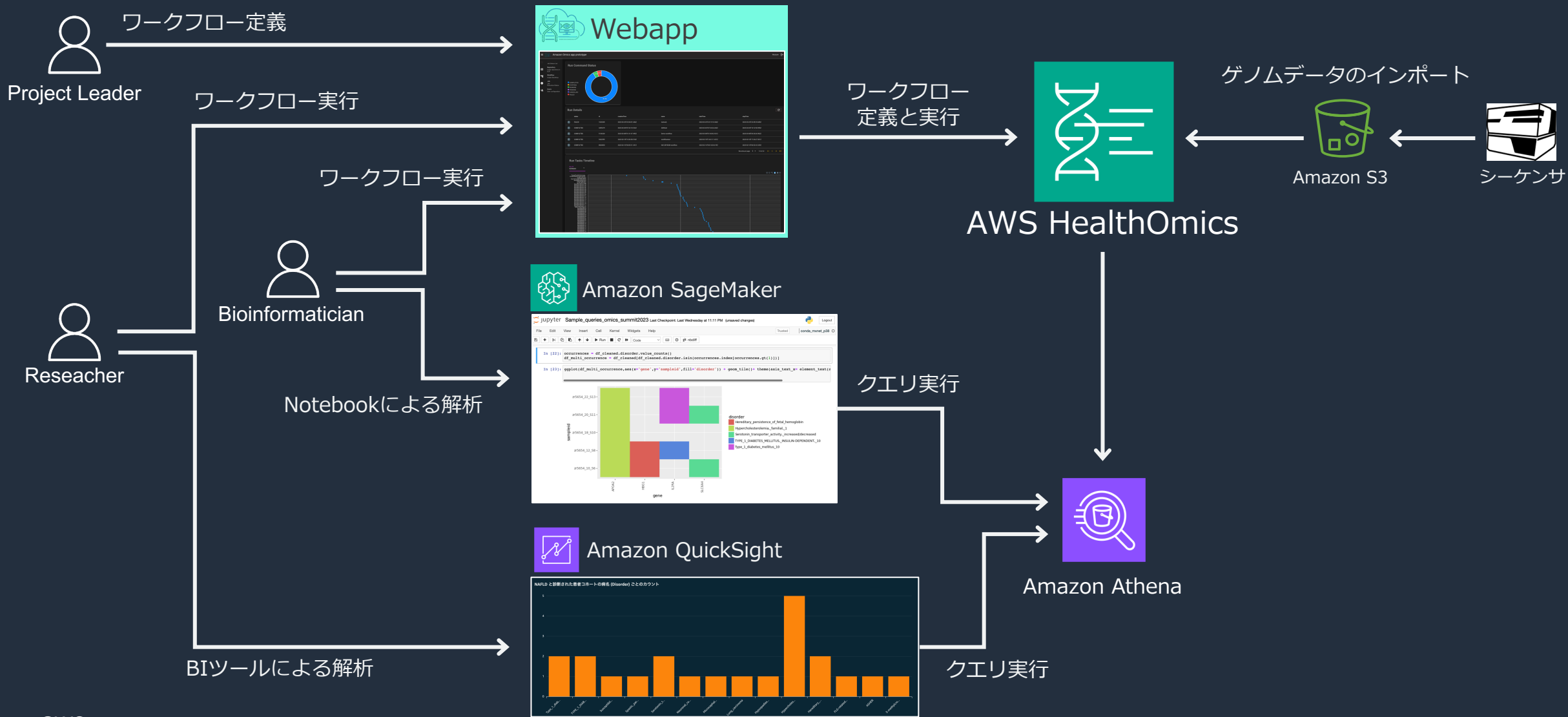
マルチオミクス解析 Demo

- ゲノムデータ解析処理を簡単実行 [AWS HealthOmics]
- 多様なユーザーによるオミクスデータ分析
- 試行錯誤のワークフロー実行

Amazon Web Services Japan

Demo Overview

WebappとAWS HealthOmicsによるゲノムデータの大規模計算とAWSサービスを利用した解析デモのご紹介



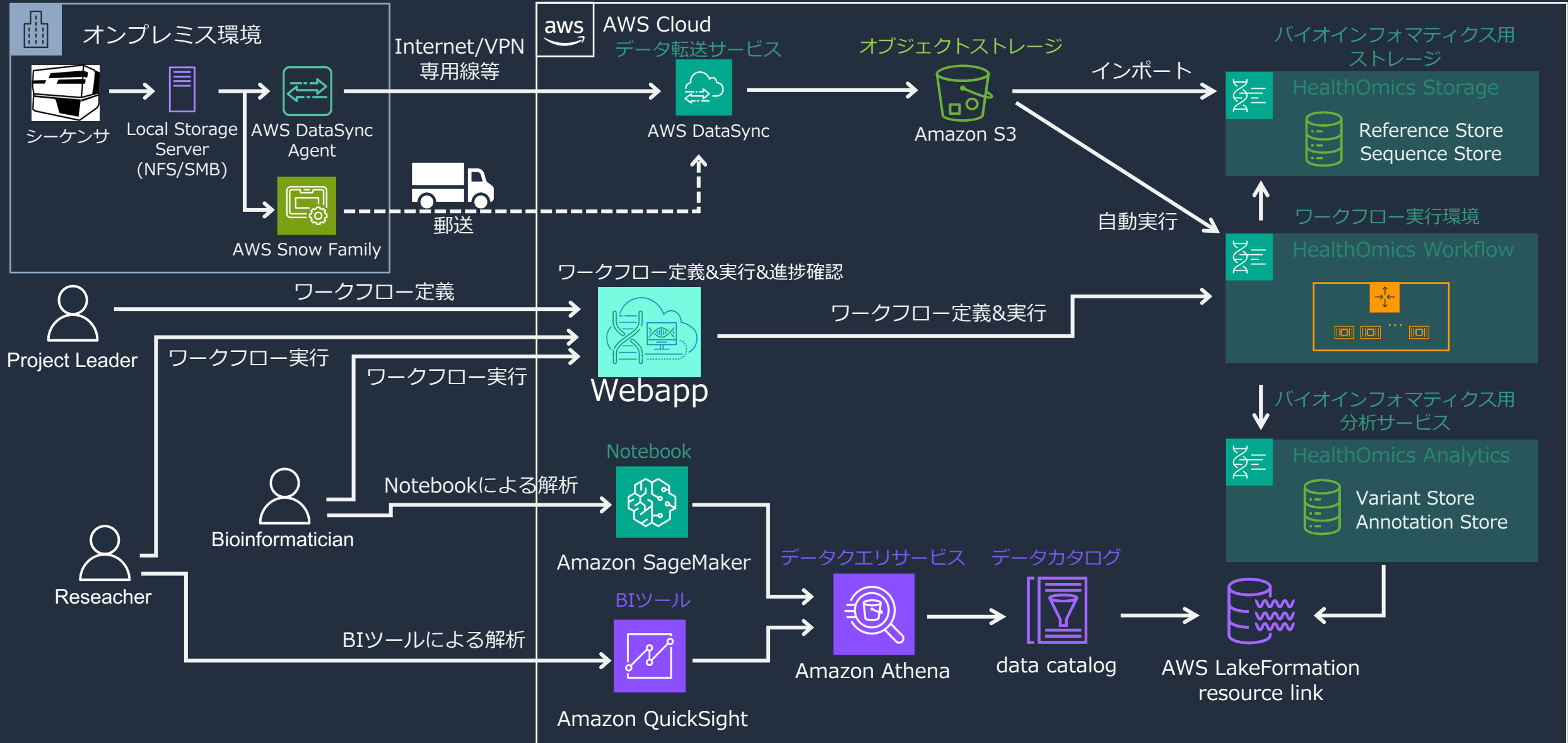
Agenda

1. オミクス解析における AWS 活用 **シチュエーション**
 - AWS HealthOmics WebApp 利用に関する全体像とデモスコープ
 - 想定ユーザーごとの WebApp 利用シチュエーション
 - Bioinformatician がユーザーの場合
 - Researcher がユーザーの場合
2. オミクス解析ワークフロー **実行**パート
 - AWS HealthOmics Webappを通して簡単実行
 - 入力となるオミクスデータ生成後の自動実行
3. オミクスデータ **分析**パート
 - ノートブックでの分析 by Amazon SageMaker
 - ダッシュボードを使った分析 by Amazon QuickSight
4. オミクス解析ワークフロー **開発**パート

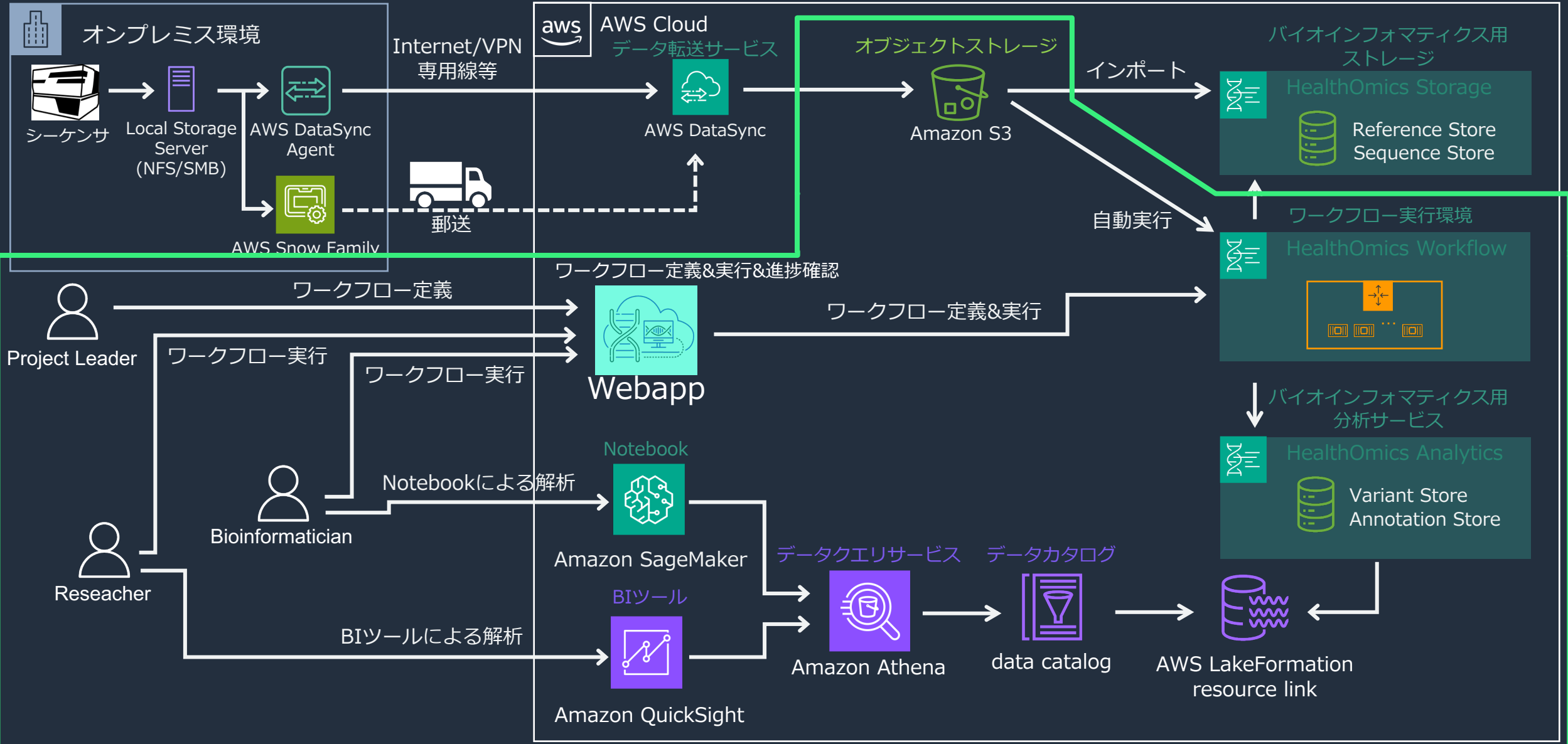
Agenda

1. オミクス解析における AWS 活用 **シチュエーション**
 - AWS HealthOmics WebApp 利用に関する全体像とデモスコープ
 - 想定ユーザーごとの WebApp 利用シチュエーション
 - Bioinformatician がユーザーの場合
 - Researcher がユーザーの場合
2. オミクス解析ワークフロー **実行**パート
 - AWS HealthOmics Webappを通して簡単実行
 - 入力となるオミクスデータ生成後の自動実行
3. オミクスデータ **分析**パート
 - ノートブックでの分析 by Amazon SageMaker
 - ダッシュボードを使った分析 by Amazon QuickSight
4. オミクス解析ワークフロー **開発**パート

AWS HealthOmics Webapp 利用想定シチュエーション Overview



AWS HealthOmics Webapp 利用想定シチュエーション Overview 今回の実演デモ



想定ユーザーごとの利用シチュエーション

想定ユーザーごとの WebApp 利用シチュエーション

Bioinformatician がユーザーの場合

プロジェクトの
リーダーを想定


Project Leader

①ワークフローを定義



Webapp

AWS HealthOmics

②ワークフローを実行

バイオマティシアン
主にDryな研究を実施
PythonやRでインタラクティブ
に解析を実施されたい方を想定


Bioinformatician

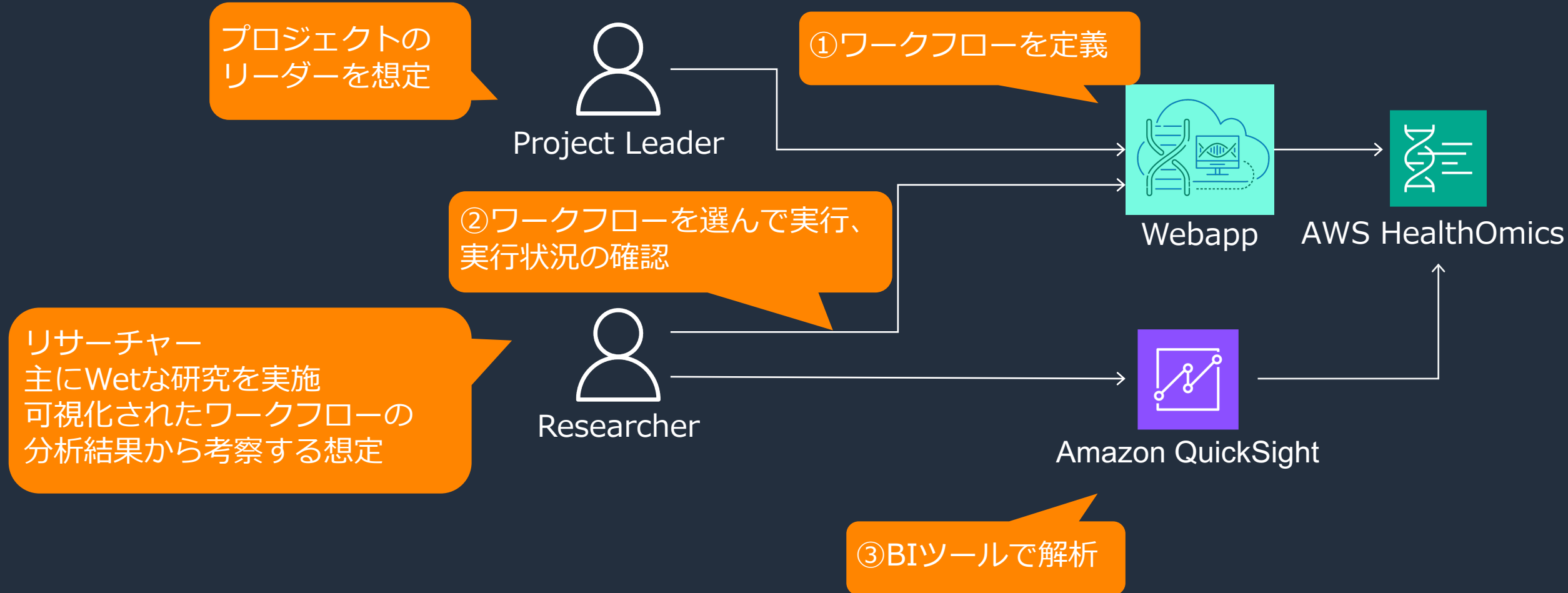


Amazon SageMaker

③Notebookで解析

想定ユーザーごとの WebApp 利用シチュエーション

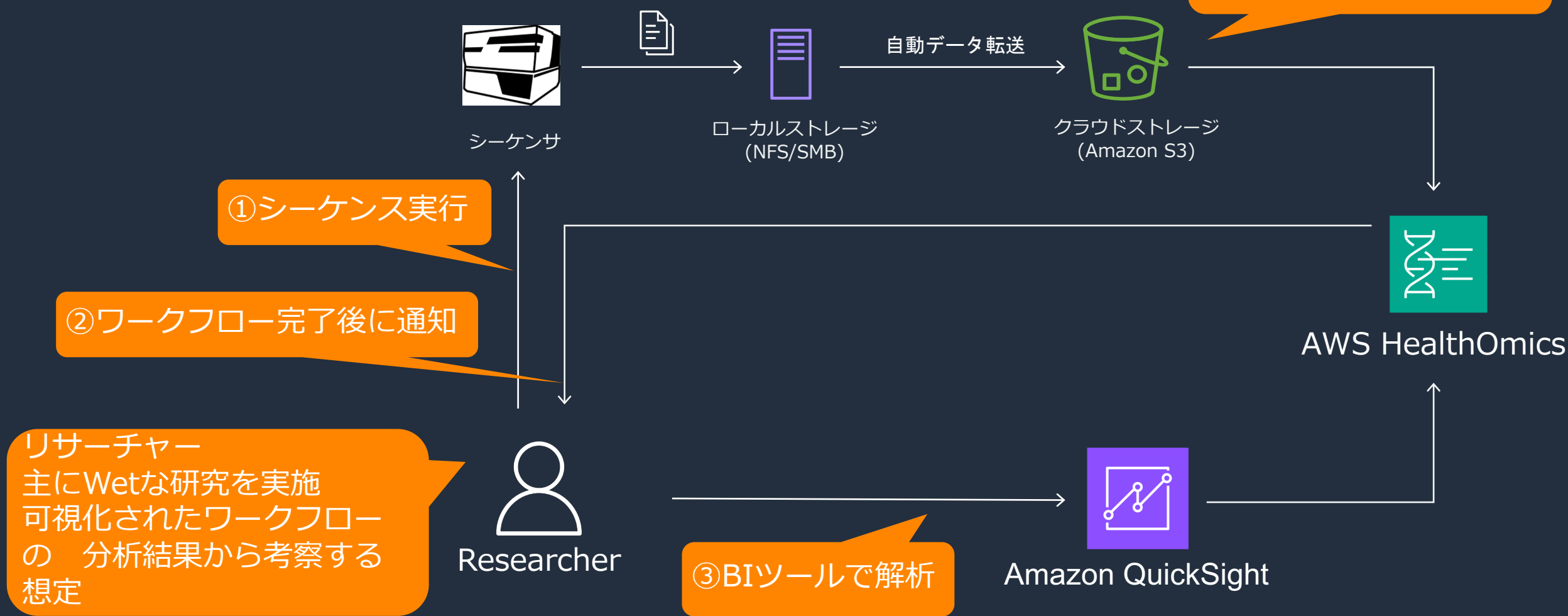
Researcher がユーザーの場合 A: Webapp のみでワークフロー実行・確認



想定ユーザーごとの WebApp 利用シチュエーション

Researcher がユーザーの場合 B: データ生成後に自動実行

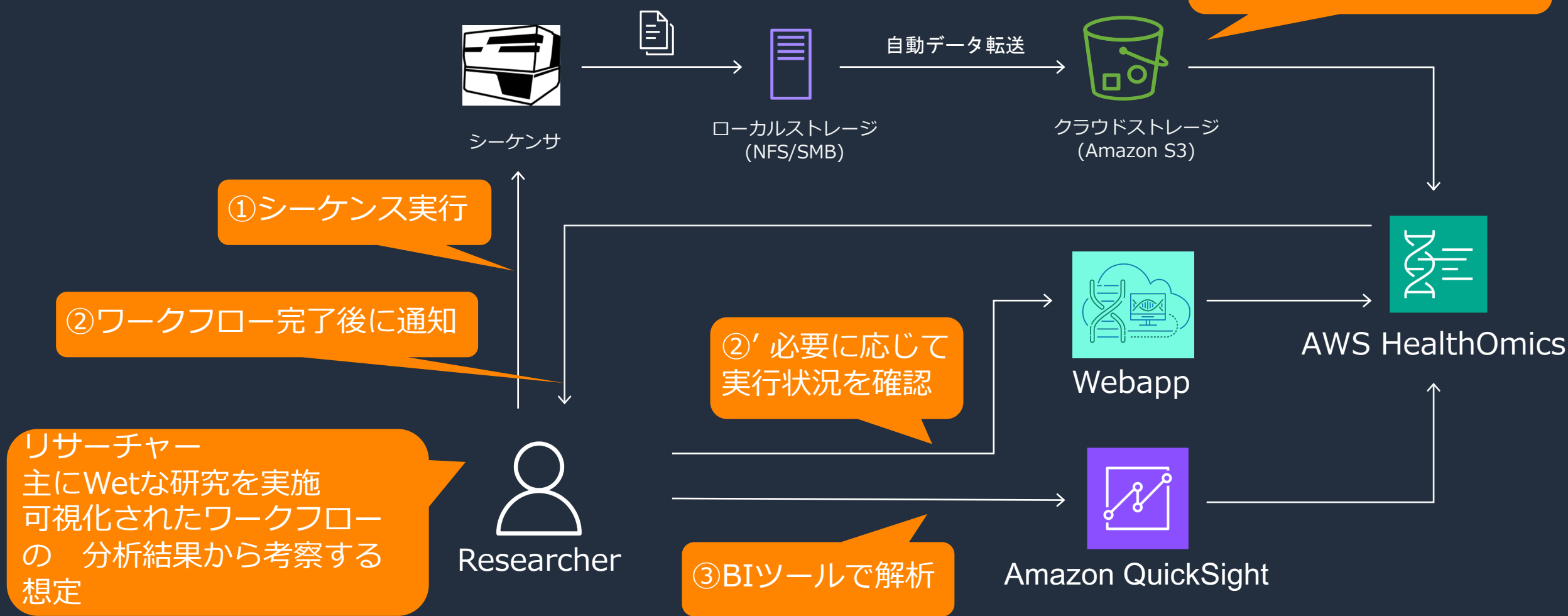
データアップロード後にワークフローを自動実行



想定ユーザーごとの WebApp 利用シチュエーション

Researcher がユーザーの場合 B: データ生成後に自動実行

データアップロード後に
ワークフローを自動実行



①シーケンス実行

②ワークフロー完了後に通知

②' 必要に応じて
実行状況を確認

③BIツールで解析

リサーチャー
主にWetな研究を実施
可視化されたワークフロー
の 分析結果から考察する
想定

Agenda

1. オミクス解析における AWS 活用シチュエーション
 - AWS HealthOmics WebApp 利用に関する全体像とデモスコープ
 - 想定ユーザーごとの WebApp 利用シチュエーション
 - Bioinformatician がユーザーの場合
 - Researcher がユーザーの場合
2. オミクス解析ワークフロー実行パート
 - AWS HealthOmics Webappを通して簡単実行
 - 入力となるオミクスデータ生成後の自動実行
3. オミクスデータ分析パート
 - ノートブックでの分析 by Amazon SageMaker
 - ダッシュボードを使った分析 by Amazon QuickSight
4. オミクス解析ワークフロー開発パート

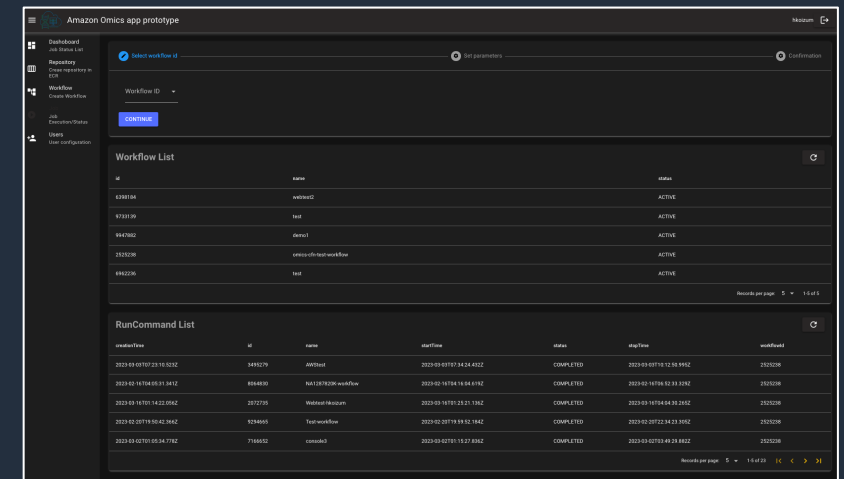
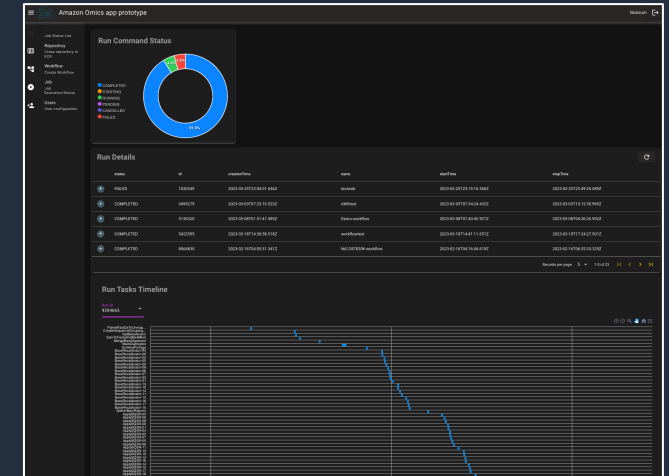
AWS HealthOmics Webapp のご紹介

AWS HealthOmics Webapp の概要



- バイオインフォマティクスに関わるバイオインフォマティシャンやリサーチャーのためのウェブアプリ
- AWS HealthOmics ワークフローをウェブアプリから操作可能
- 実行中のワークフローのステータスをチャートとテーブルで視覚化
- プロジェクトリーダーは、メンバー作成したりグループに追加したりして管理することが可能
- aws-samplesにオープンソースとして提供

GitHub - aws-samples/amazon-omics-webapp-integrations
<https://github.com/aws-samples/amazon-omics-webapp-integrations>



The screenshot shows the 'Workflow List' and 'RunCommand List' sections. The 'Workflow List' table has columns for ID, NAME, and STATUS. The 'RunCommand List' table has columns for workflow, ID, NAME, workflow, status, and start_time.

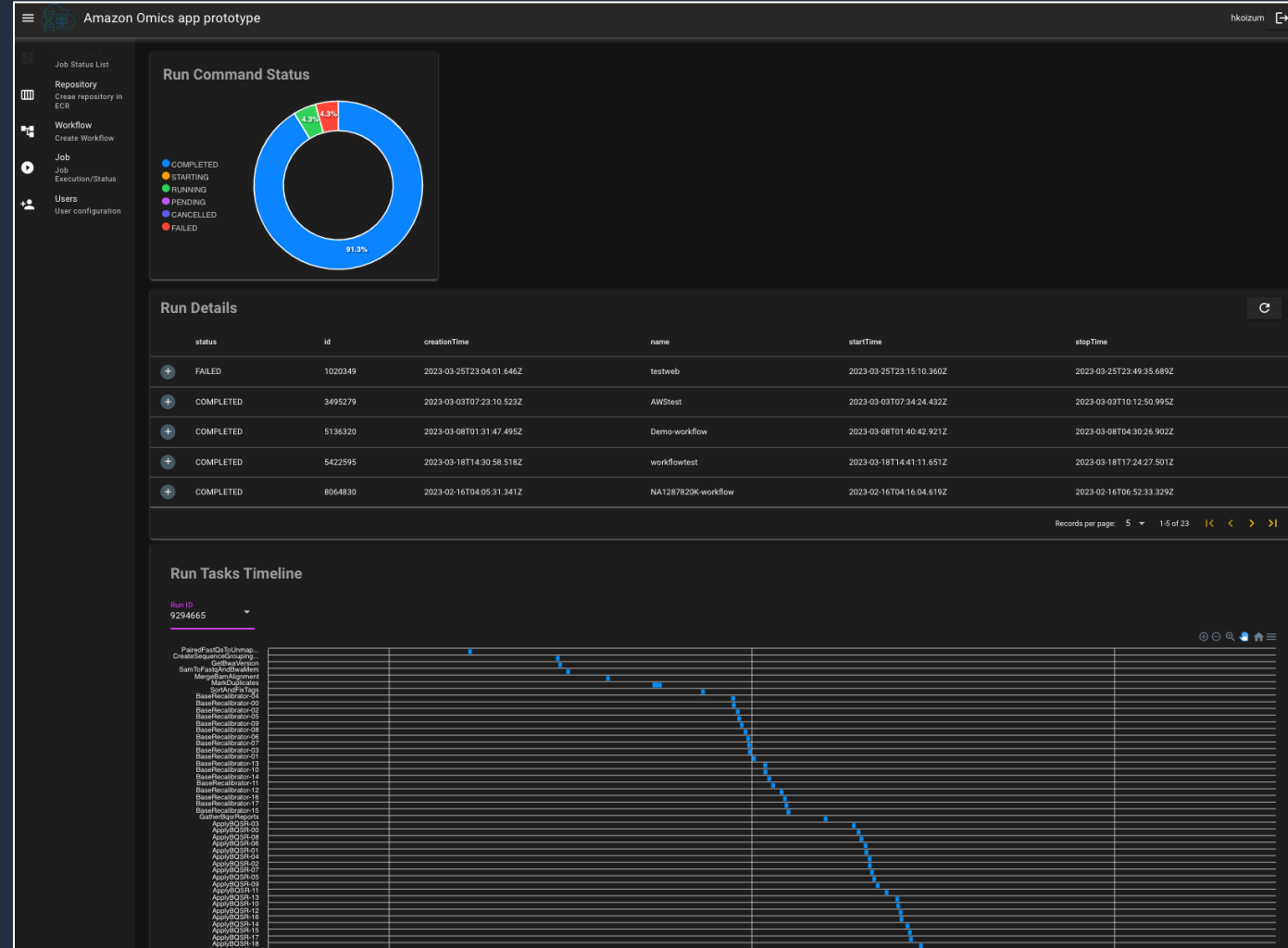
workflow	ID	NAME	workflow	status	start_time
2023-03-0710123116112	849279	testWorkflow	2023-03-0710123116112	COMPLETED	2023-03-0710123116112
2023-03-1610408131812	854880	NA128783W.workflow	2023-03-161041618161812	COMPLETED	2023-03-161041618161812
2023-03-1610114221062	207275	WebinarDemo.com	2023-03-161011312111312	COMPLETED	2023-03-161041618161812
2023-03-201910421362	623865	Test.workflow	2023-03-20191019101912142	COMPLETED	2023-03-20191019101912142
2023-03-0710105141712	716652	testWorkflow	2023-03-0710115171812	COMPLETED	2023-03-0710115171812

AWS HealthOmics Webapp の各機能



ダッシュボード

- 実行中のワークフローにおける Run Command ステータスのチャート
- Run Command の詳細を表示
- 各タスクをタイムラインで表示



AWS HealthOmics Webapp の各機能



リポジトリ

- ECRにコンテナリポジトリを作成し、ワークフロー実行環境のコンテナイメージをプッシュ可能
- コンテナリポジトリ一覧を表示

Amazon Omics app prototype

hkoizum

Dashboard
Job Status List

+ CREATE REPOSITORY

Repository List

repositoryArn	repositoryName	repositoryUri
arn:aws:ecr:	cdk-hkoizdev-container-assets-	*kr.ecr.us-east-1.amazonaws.com/cdk-hkoizdev-container-assets-1-
arn:aws:ecr:	gatk	:ecr.us-east-1.amazonaws.com/gatk
arn:aws:ecr:	genomes-in-the-cloud	:ecr.us-east-1.amazonaws.com/genomes-in-the-cloud
arn:aws:ecr:	cdk-hnb659fds-container-asset	:ecr.us-east-1.amazonaws.com/cdk-hnb659fds-container-assets
arn:aws:ecr:	webtest	:ecr.us-east-1.amazonaws.com/webtest

Records per page: 5 1-5 of 5

AWS HealthOmics Webapp の各機能



独自ワークフローの定義

- S3バケットにアップロードしたワークフローの定義ファイルを指定してワークフローを定義可能
- 必要なストレージ容量やワークフロー言語を指定
- パラメーター定義はJSONで定義したファイルをアップロード可能

Amazon Omics app prototype

hkoizum

Define workflow

Confirmation

Workflow Definition

Workflow name	Description	Storage capacity
Workflow definition in S3		Workflow language

Format: s3://bucket/prefix/object. File type must be .zip.

Parameters

Select your json file

Upload your json file for advanced parameters 0 (0.0B)

CONFIRMATION

AWS HealthOmics Webapp の各機能



ジョブ

- 定義済みワークフローを指定してジョブを実行可能
- パラメーターはJSONで定義したファイルをアップロード可能

The screenshot displays the Amazon Omics app prototype interface. The top navigation bar includes a menu icon, the text "Amazon Omics app prototype", and a user profile icon labeled "hkozum". The left sidebar contains navigation options: Dashboard, Job Status List, Repository (with a sub-option "Create repository in ECR"), Workflow (with a sub-option "Create workflow"), Job Execution/Status, and Users (with a sub-option "User configuration").

The main content area is divided into two sections:

- Workflow List:** A table with columns for id, name, and status. It contains five rows of data, all with a status of "ACTIVE".
- RunCommand List:** A table with columns for creationTime, id, name, startTime, status, stopTime, and workflowId. It contains six rows of data, all with a status of "COMPLETED".

Both tables include pagination controls at the bottom right, showing "Records per page: 5" and "1-5 of 5".

AWS HealthOmics Webapp の各機能



事前構築済みワークフローの実行

- 36 の事前定義済みのワークフロー（Ready2Run ワークフロー）の実行が可能

The screenshot shows the AWS HealthOmics App interface. The main content area displays a table titled 'Ready2Run Workflow List' with columns for id, name, status, publisher, estimatedDuration, and version. The table lists several workflows, including GATK-BP fq2bam, NVIDIA Parabricks Germline HaplotypeCaller WGS for up to 30X, scRNAseq with STARsolo, NVIDIA Parabricks BAM2FQ2BAM WGS for up to 50X, and AlphaFold for 601-1200 residues. A 'NEXT' button is visible at the bottom of the table.

id	name	status	publisher	estimatedDuration	version	
<input type="checkbox"/>	3768383	GATK-BP fq2bam	ACTIVE	Broad Institute	610	1.0
<input type="checkbox"/>	3021525	NVIDIA Parabricks Germline HaplotypeCaller WGS for up to 30X	ACTIVE	NVIDIA Corporation	75	1.0
<input type="checkbox"/>	2174942	scRNAseq with STARsolo	ACTIVE	nf-core	150	1.0
<input type="checkbox"/>	7112412	NVIDIA Parabricks BAM2FQ2BAM WGS for up to 50X	ACTIVE	NVIDIA Corporation	164	1.0
<input type="checkbox"/>	6094971	AlphaFold for 601-1200 residues	ACTIVE	DeepMind	675	1.0

GATK best practice pipelines

GATK BP germline fq2vcf for 30x

GATK BP FASTQ to BAM

GATK BP Germline BAM to VCF for 30x

GATK BP Somatic WES

GATK BP Somatic WGS

Single cell transcriptomic analysis

scRNAseq with STARsolo

scRNAseq with Kallisto + BUSStools

scRNAseq with Salmon
Alevin-fry + AlevinQC

Protein folding prediction

ESMFold for up to 800 residues

AlphaFold + MSA for
up to 600 residues

AlphaFold + MSA for
601-1200 residues



Element Biosciences

ElementBio Bases2Fastq for 2x75

ElementBio Bases2Fastq for 2x150

ElementBio Bases2Fastq for 2x300



Sentieon Inc.

Sentieon Germline DNaseq
FASTQ WGS & WES

Sentieon Somatic WGS & WES

Sentieon LongRead for PacBio HiFi

Sentieon LongRead for ONT

Sentieon Germline DNaseq
BAM WES & WGS



NVIDIA

NVIDIA Germline (DeepVariant)

NVIDIA Germline (Haplotype Caller)

NVIDIA BAM2FQ2BAM

NVIDIA FQ2BAM

NVIDIA Somatic WGS & WES

AWS HealthOmics Webapp の各機能



ユーザー管理 (admin group only)





- admin groupに所属するプロジェクトリーダーはユーザー管理が可能
- ユーザーをグループを指定して追加したり削除したりすることが可能

Amazon Omics app prototype

hkoizum

+ ADD USER

List User

name	status	email	group	action
hkoizum	CONFIRMED		admin	 
researcher1	CONFIRMED		bioinformatician	 

Records per page: 5 1-2 of 2

AWS HealthOmics Webapp のアーキテクチャ



データ生成後の自動実行

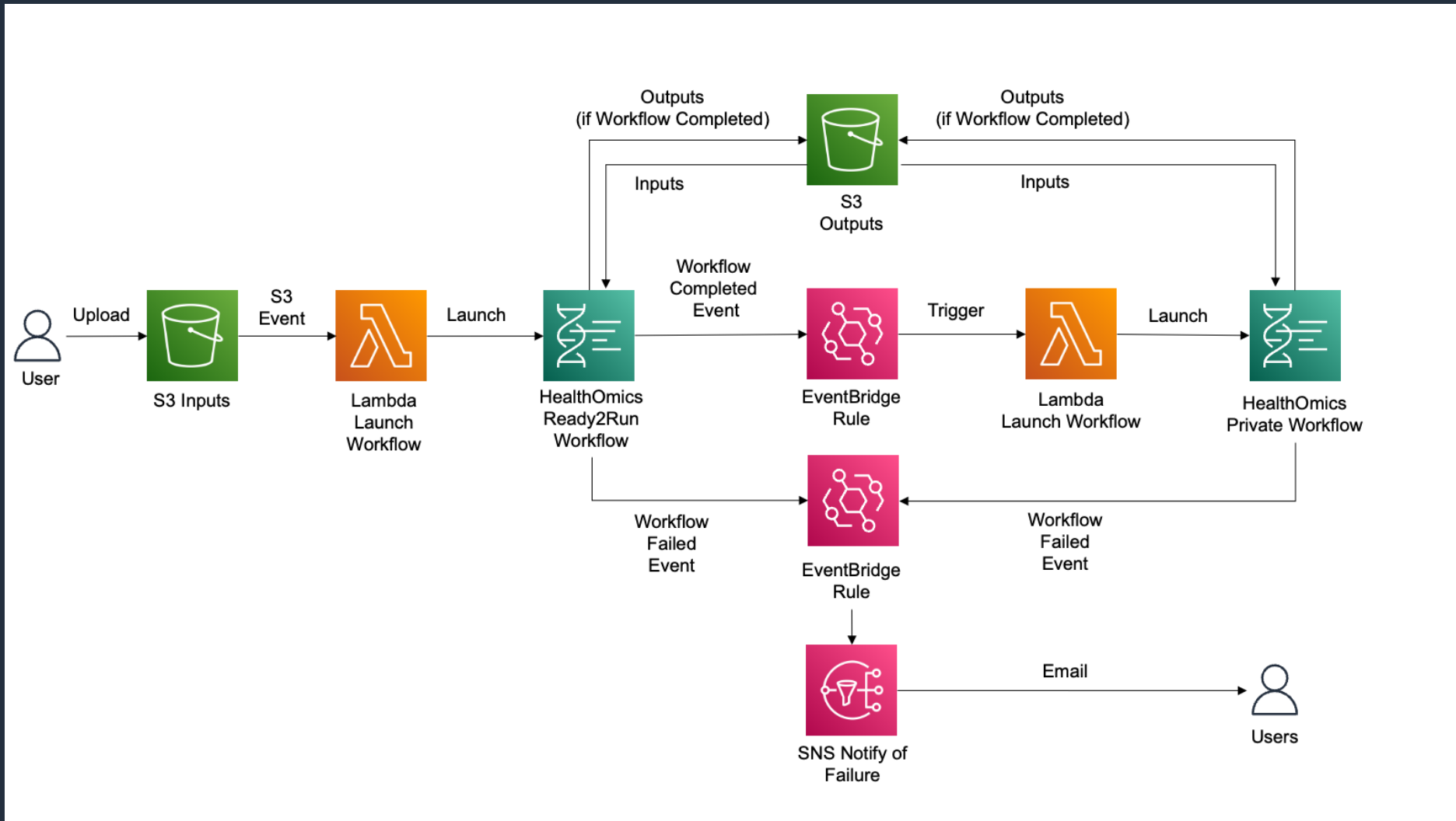
サービスを組み合わせることで便利な自動実行も可能に

例えば・・・

1. シーケンサからオミクスデータのファイルをローカルストレージに出力
2. ローカルストレージのファイルをクラウド上のストレージに自動転送
3. データ転送をトリガーに、変異解析ワークフローを自動実行
4. 変異解析ワークフローが完了したら、独自のワークフローを実行
5. 実行結果をストレージに保存したり、連携先のサービスに自動インポート
6. ワークフローが成功/失敗した場合はユーザーに通知



サンプル実装のアーキテクチャ



解説ブログ: <https://aws.amazon.com/jp/blogs/industries/designing-an-event-driven-architecture-for-bioinformatics-workflows-using-aws-healthomics-and-amazon-eventbridge/>
ソースコード: <https://github.com/aws-samples/aws-healthomics-eventbridge-integration>

Agenda

1. オミクス解析における AWS 活用シチュエーション
 - AWS HealthOmics WebApp 利用に関する全体像とデモスコープ
 - 想定ユーザーごとの WebApp 利用シチュエーション
 - Bioinformatician がユーザーの場合
 - Researcher がユーザーの場合
2. オミクス解析ワークフロー実行パート
 - AWS HealthOmics Webappを通して簡単実行
 - 入力となるオミクスデータ生成後の自動実行
3. オミクスデータ分析パート
 - ノートブックでの分析 by Amazon SageMaker
 - ダッシュボードを使った分析 by Amazon QuickSight
4. オミクス解析ワークフロー開発パート

AWS HealthOmics を利用した解析

AWS Blog のシナリオをもとに作成

Part 2: Amazon Omics を使用したゲノムデータの保管と分析のエンドツーエンド自動化 | Amazon Web Services ブログ

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

・ 利用データ

- AWS Data Exchange で利用可能な Ovation Dx 非アルコール性脂肪肝疾患 (NAFLD) の サンプル データセット (5 サンプル)
- ClinVar のバリエーションアノテーション

・ このシナリオで得られるインサイト

- すべてのNAFLD患者が家族性高コレステロール血症のリスクを高める APOA2 遺伝子のバリエーションを持っていること
- NAFLD患者の何人かは Type-1 Diabetes (1型糖尿病) に罹患しやすくなるバリエーションも持っていること

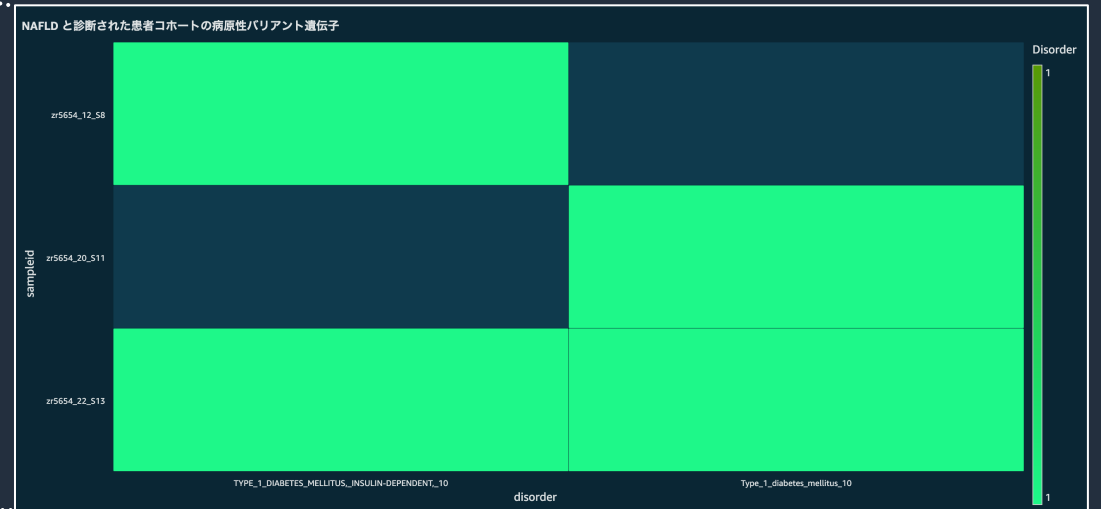
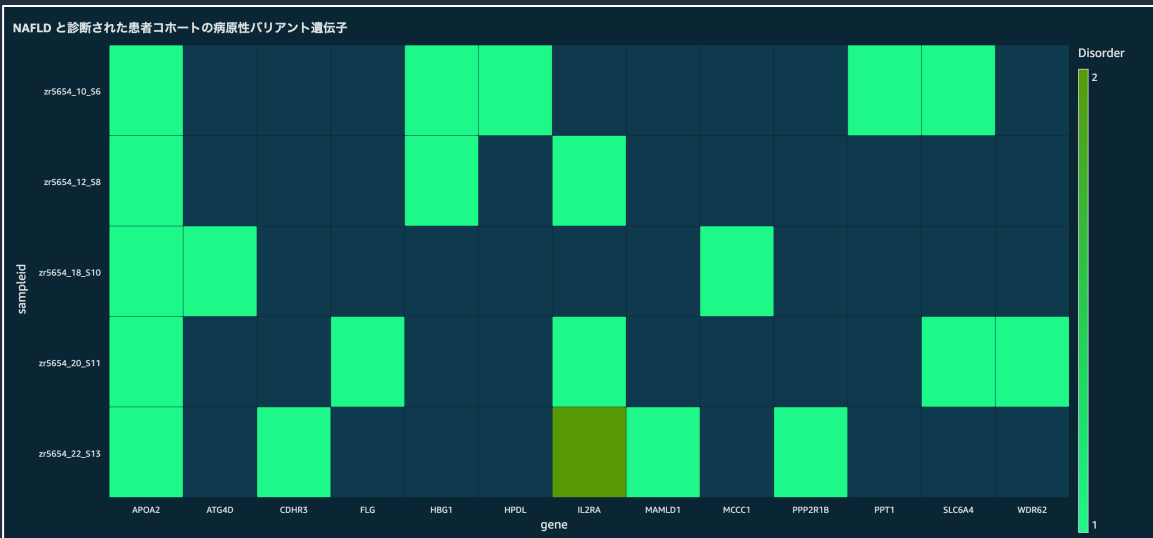
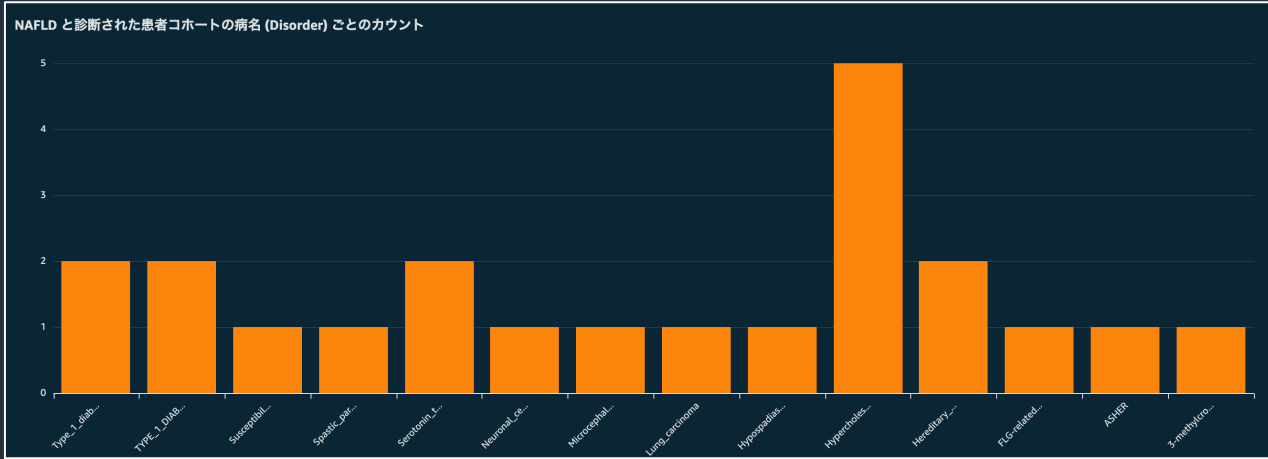
補足: HealthOmics Analytics のメリット



Amazon SageMaker Notebook を利用した解析



Amazon QuickSight を利用した解析



Agenda

1. オミクス解析における AWS 活用シチュエーション
 - AWS HealthOmics WebApp 利用に関する全体像とデモスコープ
 - 想定ユーザーごとの WebApp 利用シチュエーション
 - Bioinformatician がユーザーの場合
 - Researcher がユーザーの場合
2. オミクス解析ワークフロー実行パート
 - AWS HealthOmics Webappを通して簡単実行
 - 入力となるオミクスデータ生成後の自動実行
3. オミクスデータ分析パート
 - ノートブックでの分析 by Amazon SageMaker
 - ダッシュボードを使った分析 by Amazon QuickSight
4. オミクス解析ワークフロー開発パート

解析パイプラインをどの環境で開発するか

パイプライン環境	Pros	Cons	ツール/サービス例
オンプレミス環境	<ul style="list-style-type: none"> • 手軽さ • パイプライン実装の柔軟性 (ワークフロー言語, シェルスクリプト, etc) 	<ul style="list-style-type: none"> • 計算リソース/ストレージの制限 →処理時間の長期化や解析データサイズ制限 	<ul style="list-style-type: none"> • ローカルPC • オンプレミスの開発用サーバー
クラウドのコンピューティングリソース	<ul style="list-style-type: none"> • スケーラビリティ • パイプライン実装の柔軟性 (ワークフロー言語, シェルスクリプト, etc) • 環境構築と削除の高速化 	<ul style="list-style-type: none"> • 学習コスト • 非定型ワークフロー(試行錯誤の段階)での柔軟性 	<ul style="list-style-type: none"> • クラウドサービス (e.g. AWS ParallelCluster, AWS Batch, AWS HealthOmics)
クラウドのワークフローサービス	<ul style="list-style-type: none"> • スケーラビリティ • GUIベースで実装 • 環境構築と削除の高速化 • 失敗タスクからのリトライやユニットテスト (シングルタスクテスト) 	<ul style="list-style-type: none"> • 学習コスト • タスクによる環境制限の可能性(実行時間、パッケージ容量など) • パラメーター受け渡しの工夫(リトライ処理) 	<ul style="list-style-type: none"> • AWS Step Functions

AWSのスケラブルなゲノム解析環境



AWS ParallelCluster

オープンソースの クラスタ管理ツール

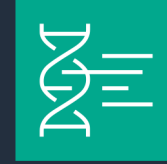
- ジョブスケジューラと連動し、ジョブに応じたインスタンス起動・停止が可能
- 既存ジョブスケジューラ環境からの移行が容易



AWS Batch

AWS独自のフルマネージドな ジョブスケジューラ

- ジョブをコンテナとして定義することで、ジョブに応じたインスタンス起動・停止が可能
- 実行処理をコンテナ化している場合は移行が容易

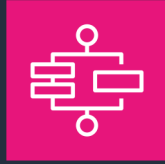


AWS HealthOmics

ゲノム解析特化型 ソリューション

- ゲノム用ストレージ、解析環境、分析環境を提供し、オミクスデータからインサイトを得ることが可能
- 事前定義済みワークフローの提供
- WDL, Nextflow, CWLのワークフローからの移行が容易

AWSのワークフローサービスで分析パイプライン構築



AWS Step Functions

実験的な試行錯誤ワークフロー

- 250以上のAWSサービスと統合されており、ワークフローから各サービスのAPIアクションを実行可能
- 特定タスク(State)をテスト可能
*TestState API*1*
- 途中失敗したタスクからの再実行が可能
Redrive



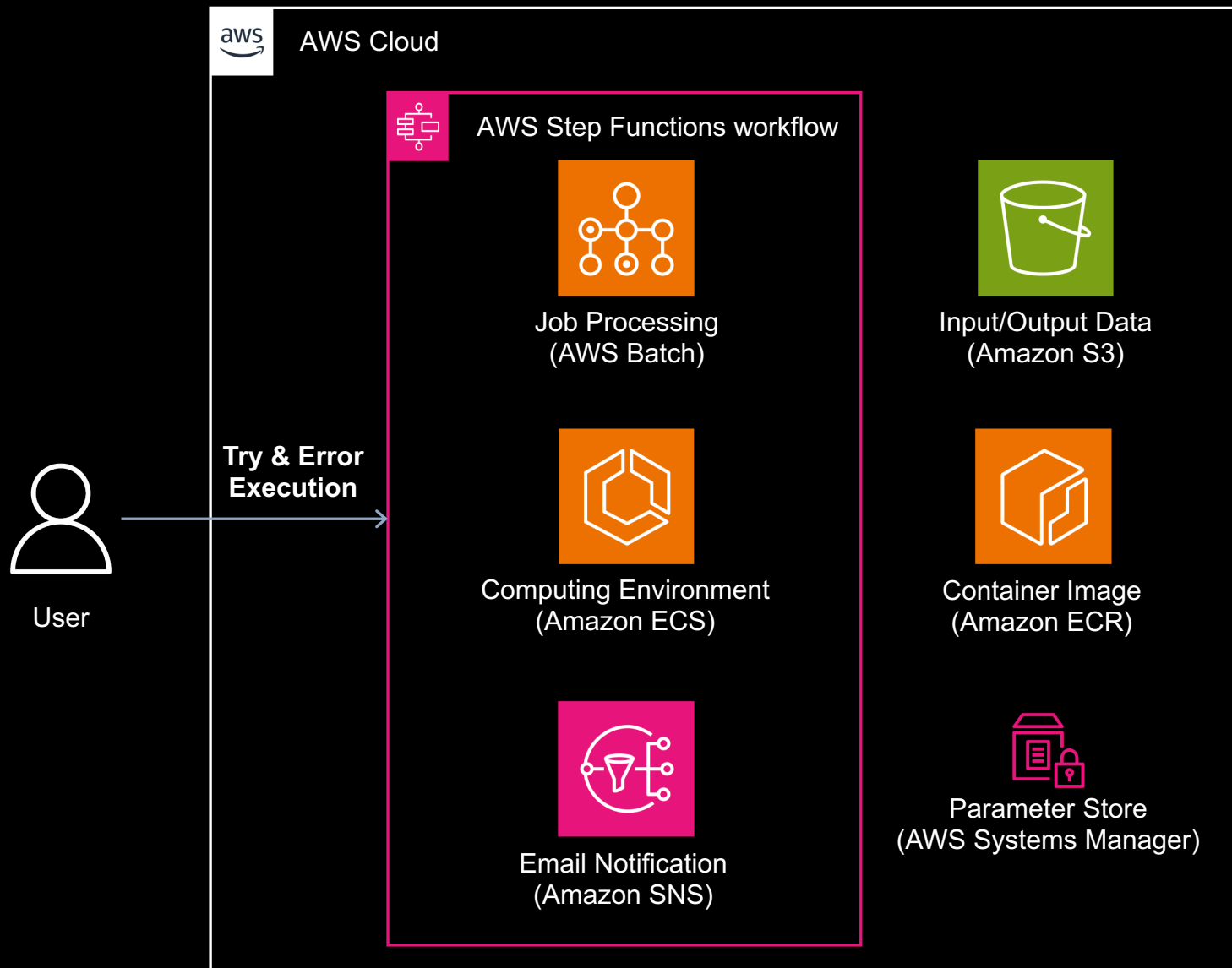
AWS HealthOmics

定型ワークフロー

- WDL, Nextflow, CWLのワークフローが実行可能
Private workflows
- 事前定義済みワークフローも実行可能
Ready2Run workflows

*1: TestState API does not support '.sync' integration patterns

Expression analysis with AWS Step Functions



Demo scenario : Variant calling

- Reference: hg38
- Pipeline :
QC -> Alignment -> Sort ->
MarkDuplicates -> Index -> Variant
calling
- Test single task : **TestState API**
- Retry workflow from failure:
Redrive

Demo environment deployment

- **sst/ion (Pulumi)**, not AWS CDK
<https://github.com/sst/ion>



Thank you!