AWS re:Invent

NOV. 29 – DEC. 3, 2021 | LAS VEGAS, NV

HLC308

Accelerate pediatric research with ML and healthcare interoperability

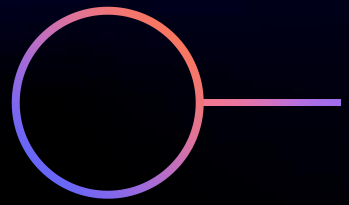
Ujjwal Ratan
Lead and Principal Machine learning SA,
Healthcare and Life Sciences
Amazon Web Services

Allison P. Heath, PhD
Director of Technology and Innovation
Center for Data-Driven Discovery in Biomedicine (D3b)
Children's Hospital of Philadelphia

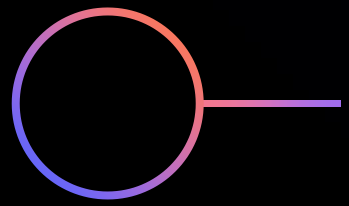
aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

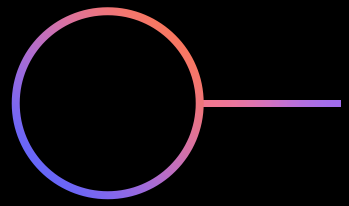
Agenda



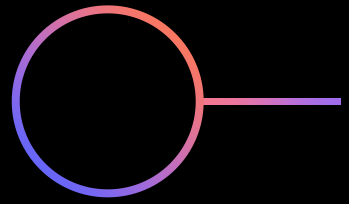
Genomics analysis for clinical research



CHOP case study



AWS for genomics



Looking ahead – AWS support for enabling interoperability

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

JAMA Network

SOURCE – EFFECT OF WHOLE GENOME SEQUENCING ON THE CLINICAL MANAGEMENT OF ACUTELY III INFANTS WITH SUSPECTED GENETIC DISEASE: A RANDOMIZED CLINICAL TRIAL

JAMA PEDIATRICS PUBLISHED ONLINE SEPTEMBER 27, 2021, DOI:10. 100 1/JAMAPEDIATRICS.202 1.3496

JAMA Pediatrics

Population

201 males, **153** females



Infants admitted to an intensive care unit with a suspected genetic disease and aged between 0 and 120 days

Mean age, 15 days (IQR, 7–32 days)



5 US academics centers and affiliated children's hospitals

Population

354 patients



176 early whole-genome sequencing testing

Whole-genome sequencing results returned 15 days after study enrollment



178 delayed whole-genome sequencing testing

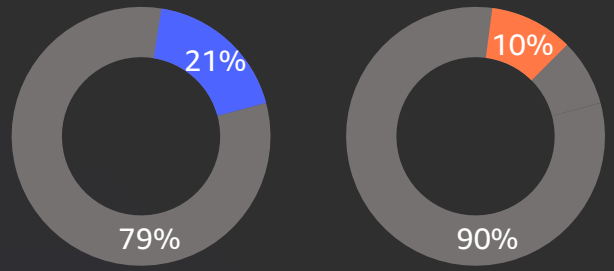
Whole-genome sequencing results returned 60 days after study enrollment

Primary outcome

Difference in the proportion of infants in the early and delayed groups who received a change of management (COM) 60 days after enrollment

Findings

The proportion of infants who received COM was significantly higher among infants receiving early whole-genome sequencing testing compared with delayed testing



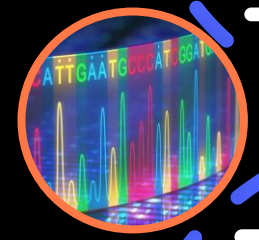
Proportion of infants with COM:
Early testing: 34 of 161 (21.1%)
Delayed testing: 17 of 165 (10.3%), P < .009

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Genomics is at the heart of the precision medicine research

Helping customers accelerate genomics at the intersection of biology and technology



Pharmacogenomics

Clinical genomics testing

Genetic risk scores

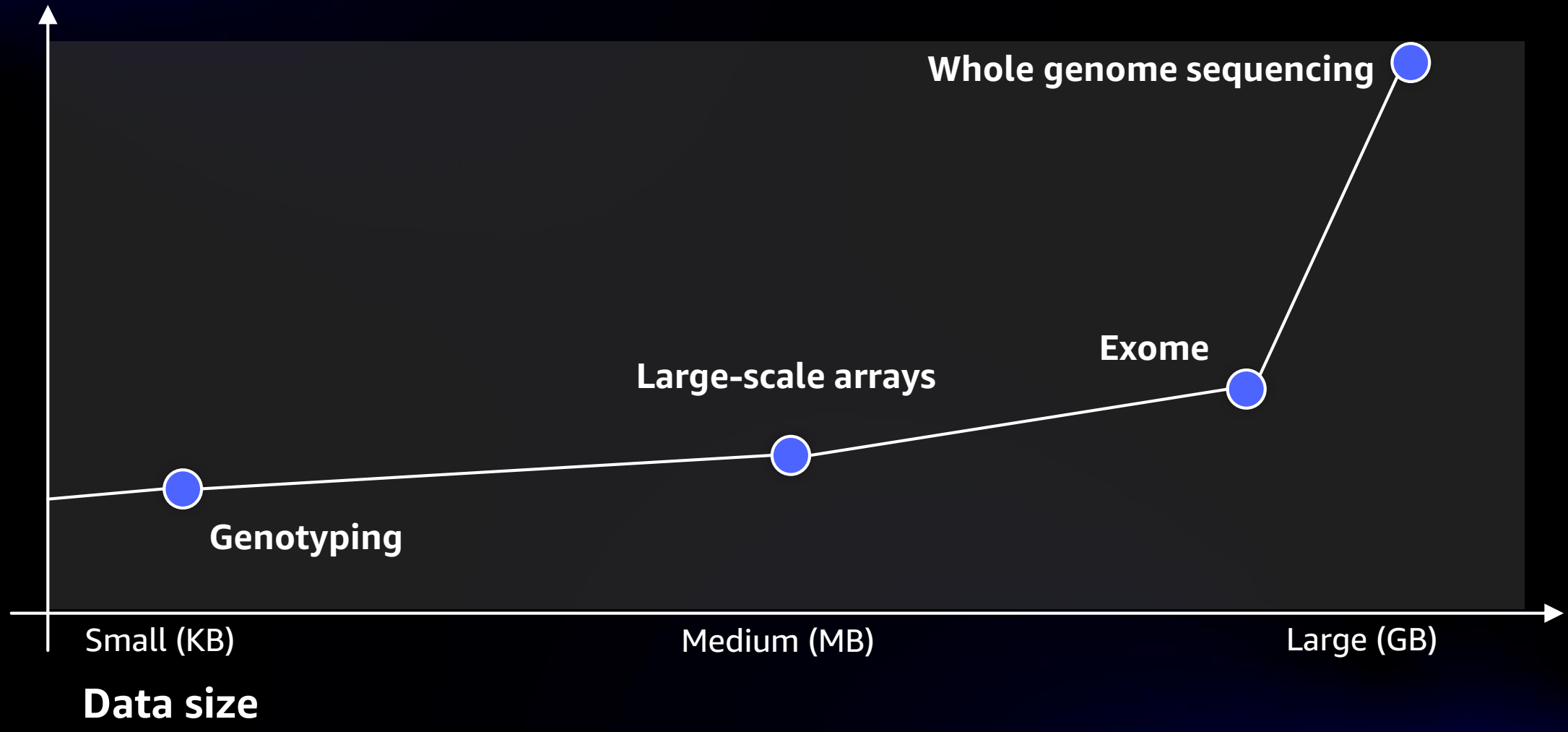
Induced pluripotent stem cells

CRISPR genome editing

Oncology testing

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Integrating large-scale clinical genomic data



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Introducing AWS for Health

Accelerating innovation from benchtop to bedside

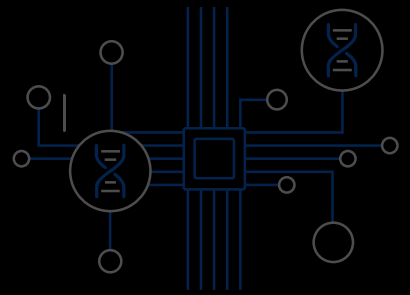
aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

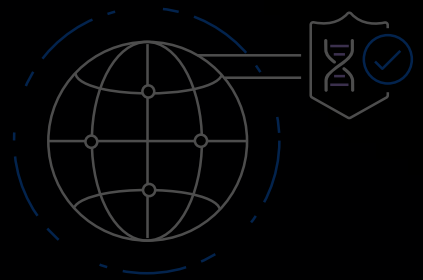
Solutions in genomics



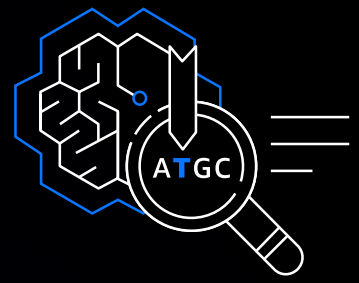
Data transfer and storage



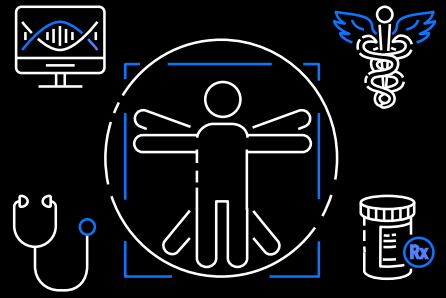
Workflow automation for secondary analysis



Data aggregation and governance



Interpretation and ML for tertiary analysis



Clinical translation

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Clinical genomics at the point of care

What is it?

Solution to find meaning in genomics and help individuals with diagnosis and treatment decisions

Why is it needed?

With insights from patient genetic codes, clinicians can make more accurate diagnoses and more effective treatment decisions faster



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

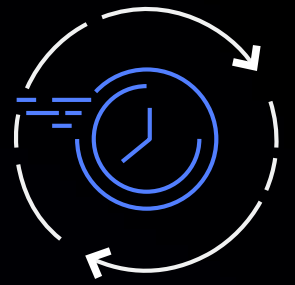
Challenges in adoption of clinical genomics



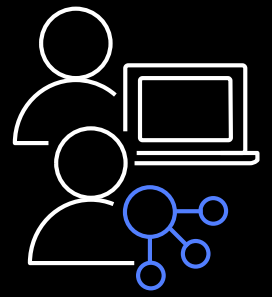
Security
and compliance



Reproducible and
interpretable results



Speed
to insights



Data sharing
and collaboration

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Children's Hospital of Philadelphia

Center for Data Driven Discovery in Biomedicine

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

D3b mission and vision: CHOP center of emphasis

Mission

The mission of the Center for Data-Driven Discovery in Biomedicine (D3b) is to build a transformative healthcare discovery ecosystem to provide innovative care for all children through collaborative, data-driven science

Vision

Breakthroughs for every child, every time, everywhere



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

D3b's team of subject matter experts

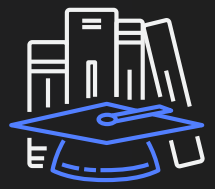
Children's Hospital of Philadelphia℠



Clinical research unit (CRU)



Biospecimen research unit (BRU)



Molecular diagnostic research unit (MDRU)



Preclinical laboratory research unit (PCLRU)



Bioinformatics unit (BIXU)



Advanced data applications and platform technologies unit (ADAPT)



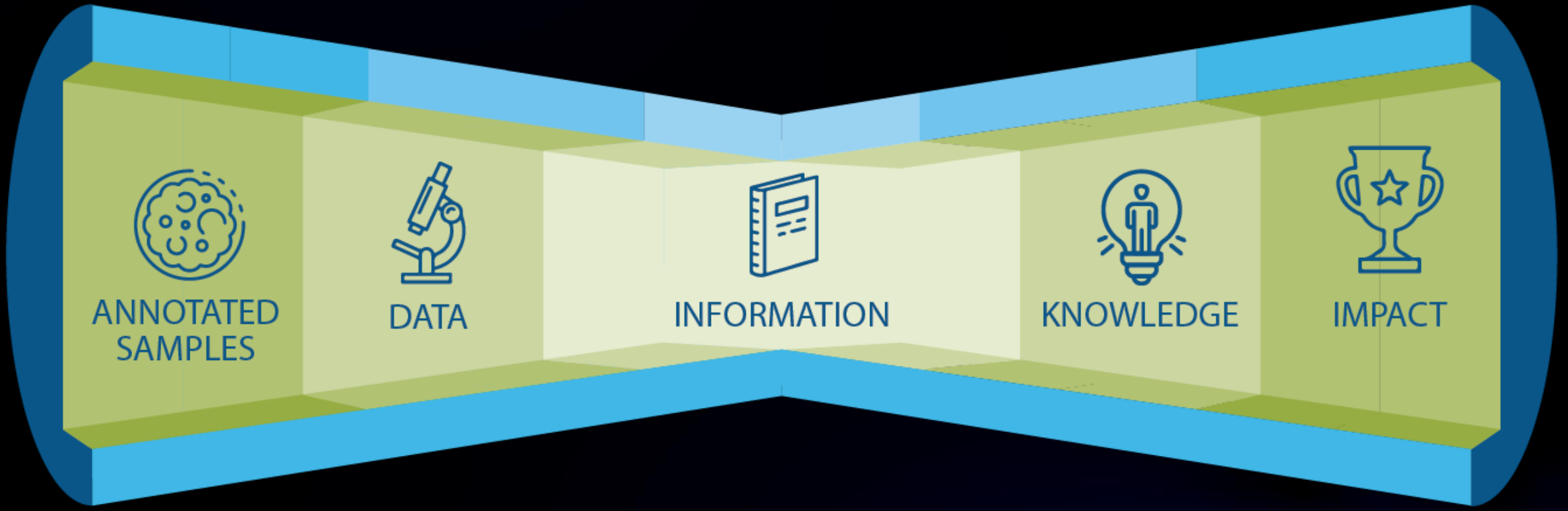
Strategic coordination unit (SCU)

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

D3b research methodology

Children's Hospital of Philadelphia℠



aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Children’s Brain Tumor Network (CBTN)

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The challenge

Within pediatric illnesses, brain tumors are the **leading cause of disease-related death in children** in the **US** and numerous other countries **around the world**



Photo courtesy of Ann & Robert H. Lurie Children's Hospital of Chicago

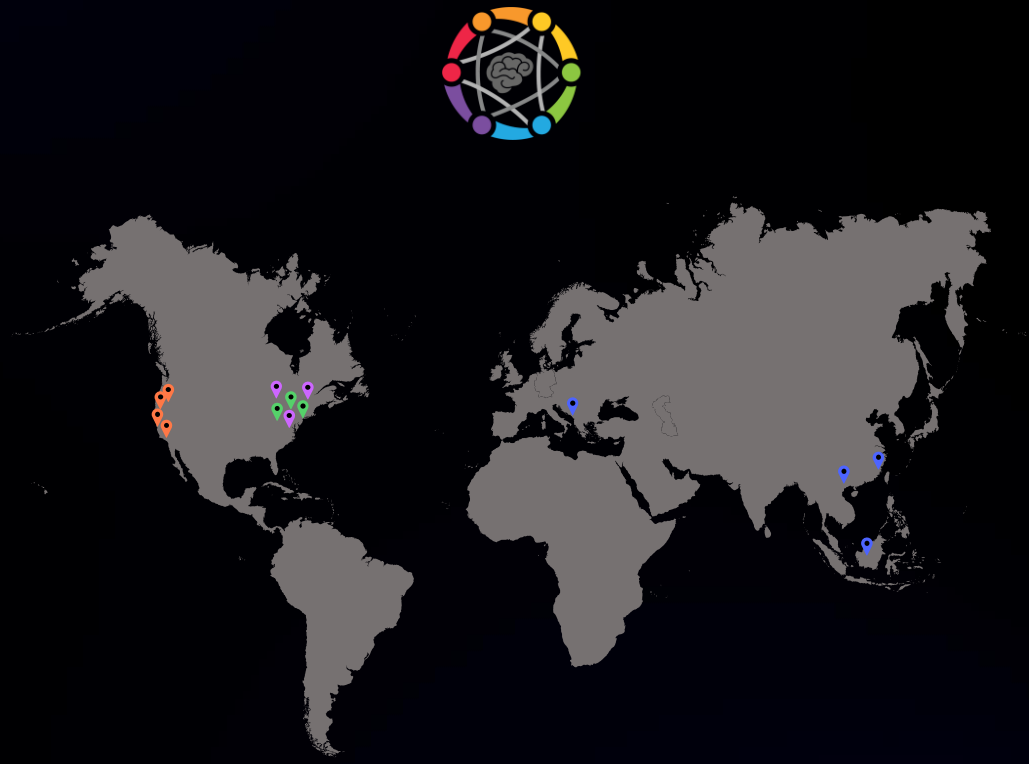
aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

CBTN member institutions

- Ann & Robert H. Lurie Children's Hospital of Chicago
- Dayton Children's Hospital
- UPMC Children's Hospital of Pittsburgh
- Weill Cornell Pediatric Brain & Spine Center
- University of Alabama at Birmingham

- Seattle Children's Hospital
- UCSF Benioff Children's Hospital
- Lucile Packard Children's Hospital Stanford
- UCSC Treehouse Childhood Cancer Initiative



The 18 CBTN member institutions (with another 3 onboarding) contribute disease and scientific expertise, develop precision-based clinical trials, enroll research subjects, and collect biospecimens, clinical data globally

- Joseph M. Sanzari Children's Hospital at Hackensack UMC
- Children's Hospital of Philadelphia
- Children's National Health System
- Wake Forest Baptist Health: Brenner Children's Hospital
- Johns Hopkins Medicine

- Meyer Children's Hospital
- The Beijing Tiantan Hospital Neurosurgery Center
- GeneBank/Beijing Genomics Institute
- The Hudson Institute for Medical Research

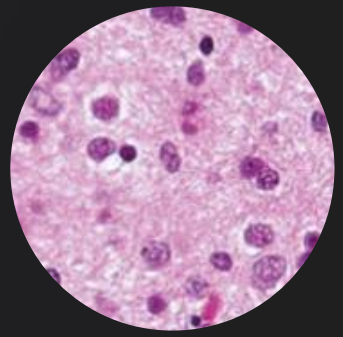
aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

2021 CBTN cohort tumor histology overview

The CBTN enrolls children and young adults with any brain tumor diagnosis from the most prevalent to the rarest and uniquely captures the evolution of tumor by following patients through their brain tumor journey

+4,200 participants across all brain tumor types



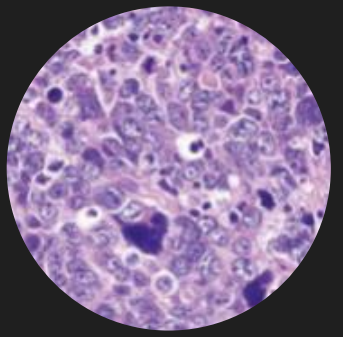
HGG

High-grade glioma/astrocytoma (WHO grade III/IV)

High-grade gliomas (HGG) in children nearly always result in a dismal prognosis. Although novel therapeutic approaches are currently in development, preclinical testing has been limited, due to a lack of pediatric-specific . . .

347 participants

47 projects supported



Medulloblastoma

Medulloblastoma

Medulloblastomas comprise the vast majority of pediatric embryonal tumors and by definition arise in the posterior fossa, where they constitute approximately 40% of all posterior fossa tumors. Other forms of embryonal tumors . . .

443 participants

33 projects supported

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

CBTN methodology



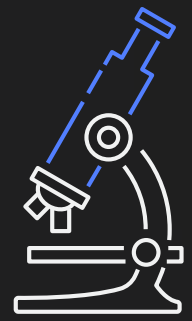
Patients

Enroll patients on to the CBTN research protocol and collect clinical data & biospecimens related to the brain tumor journey



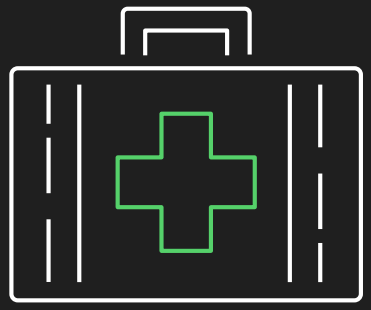
Data & models

From the collections, we generate large-scale, harmonized genomic datasets and preclinical models



Research

Using the clinical data, genomic data, and models, we conduct basic research & bioinformatics analysis to uncover disease mechanisms and therapeutics discoveries



Therapeutics

We take the findings in real time for patient-specific analysis on clinical trials, develop new molecularly driven clinical trials, and provide clinical decision support

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Impact of collaboration

Traditional research model	CBTN research model
Limited to local subject enrolled at a **single institution**	**More than 3,600** study participants enrolled at CBTN member institution
5–10 specimens collected monthly (expected)	**5–100 specimens** collected monthly (expected)
One project focusing on a single area of study	More than **200 CBTN discovery projects** across multiple brain tumor diagnoses
Data are **siloed or unavailable** until published in academic journals	Nearly **400 TB of data** are accessible in real-time through the CBTN's cloud-based platforms

VS.

Research studies provide vital information to better understand childhood brain tumors and support the development of clinical trials

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

CBTN milestones

2013
CBTN consortium launch

2014
CBTN biospecimen & clinical data portal launch

2016
Genomics Analytic Platform Launch: Cavatica

2017
PBTA launch (largest genomic cohort of childhood brain tumors without embargo)

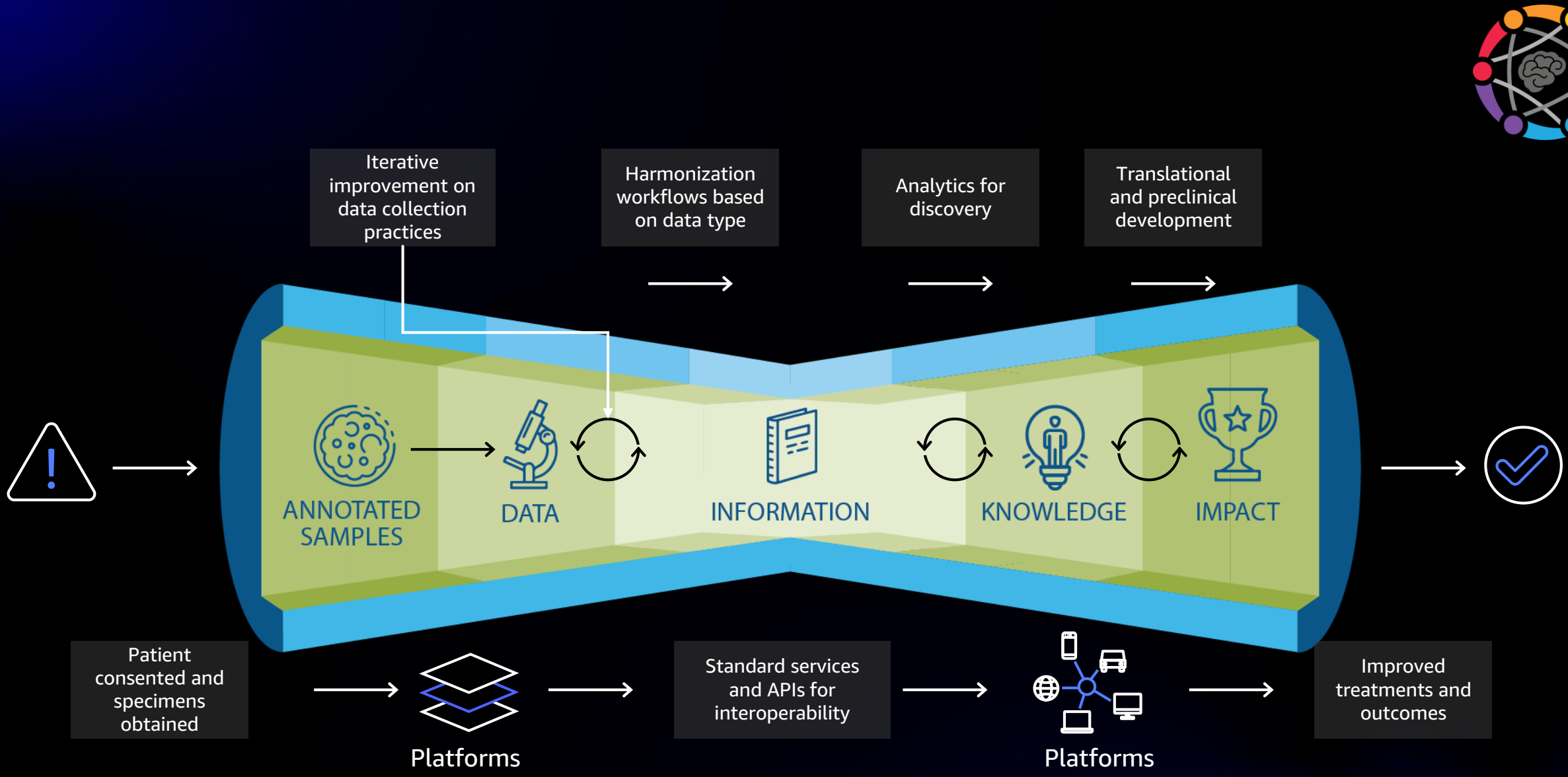
2018
Genomic & clinical data ecosystem: Kids First Data Resource Portal

2019
First per-subject clinical trial treatment clinical decision

2020
Expansion to Children's Brain Tumor Network

AWS-based data infrastructure since 2016

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

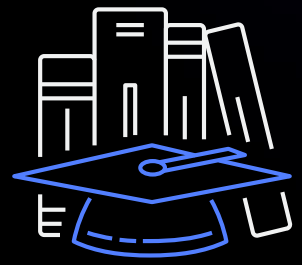


© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

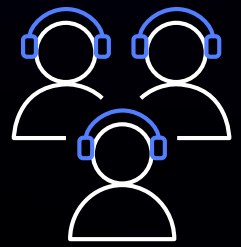
Data to impact at scale

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

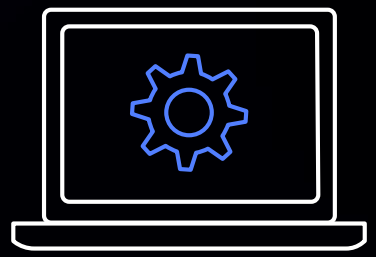
Design principles for scaling data to impact



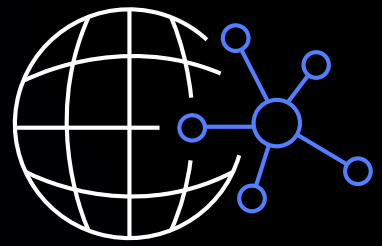
Transparent knowledge base



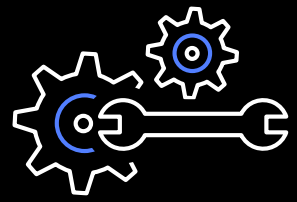
Maximize users & processors



Diversity of information



Connectivity

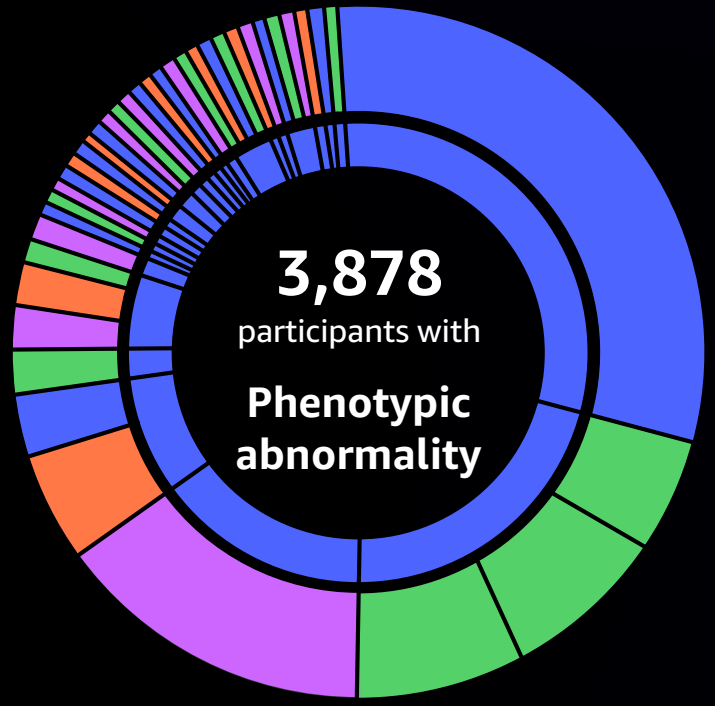


Frictionless flow

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Collaborative genomic infrastructure

Observed phenotypes



29 studies

20,215 participants

8,249 families

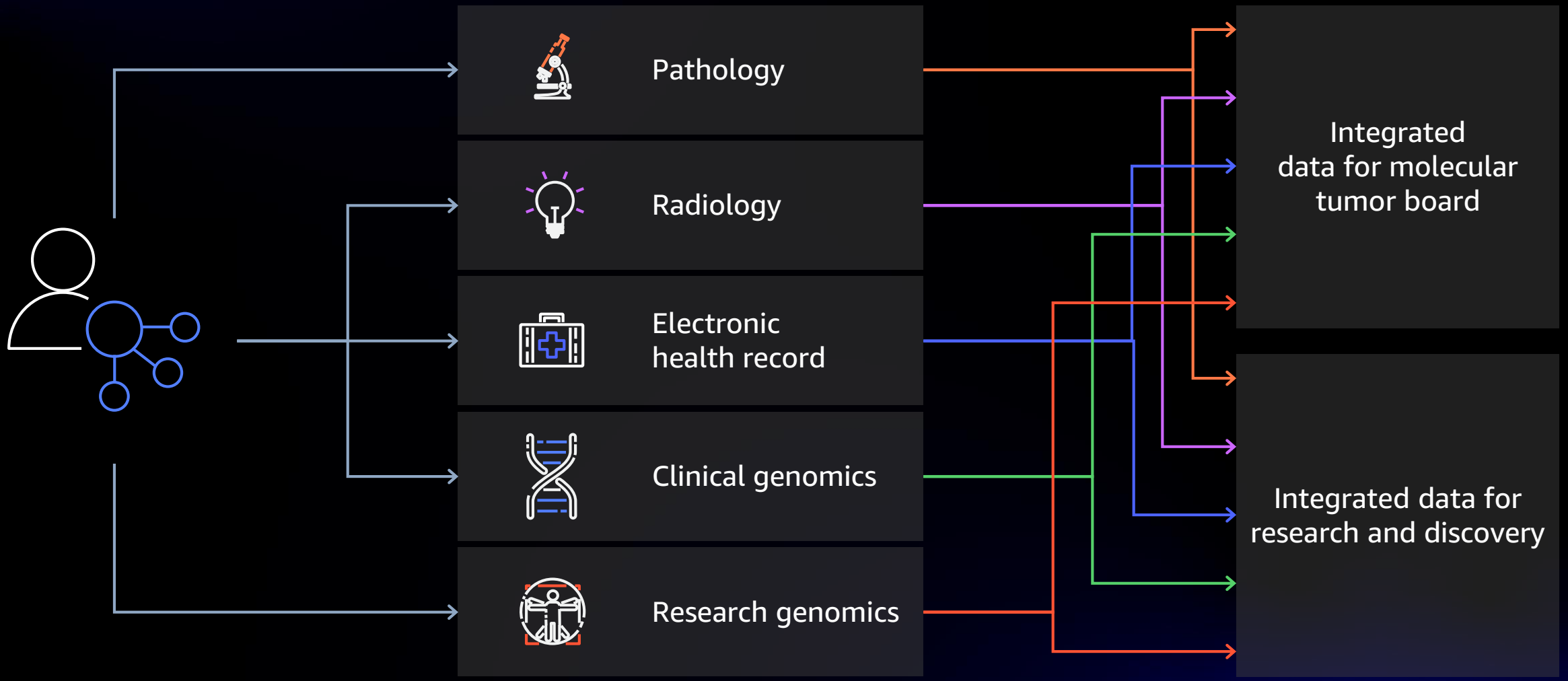
26,490 samples

127,952 files

1.44 PB size

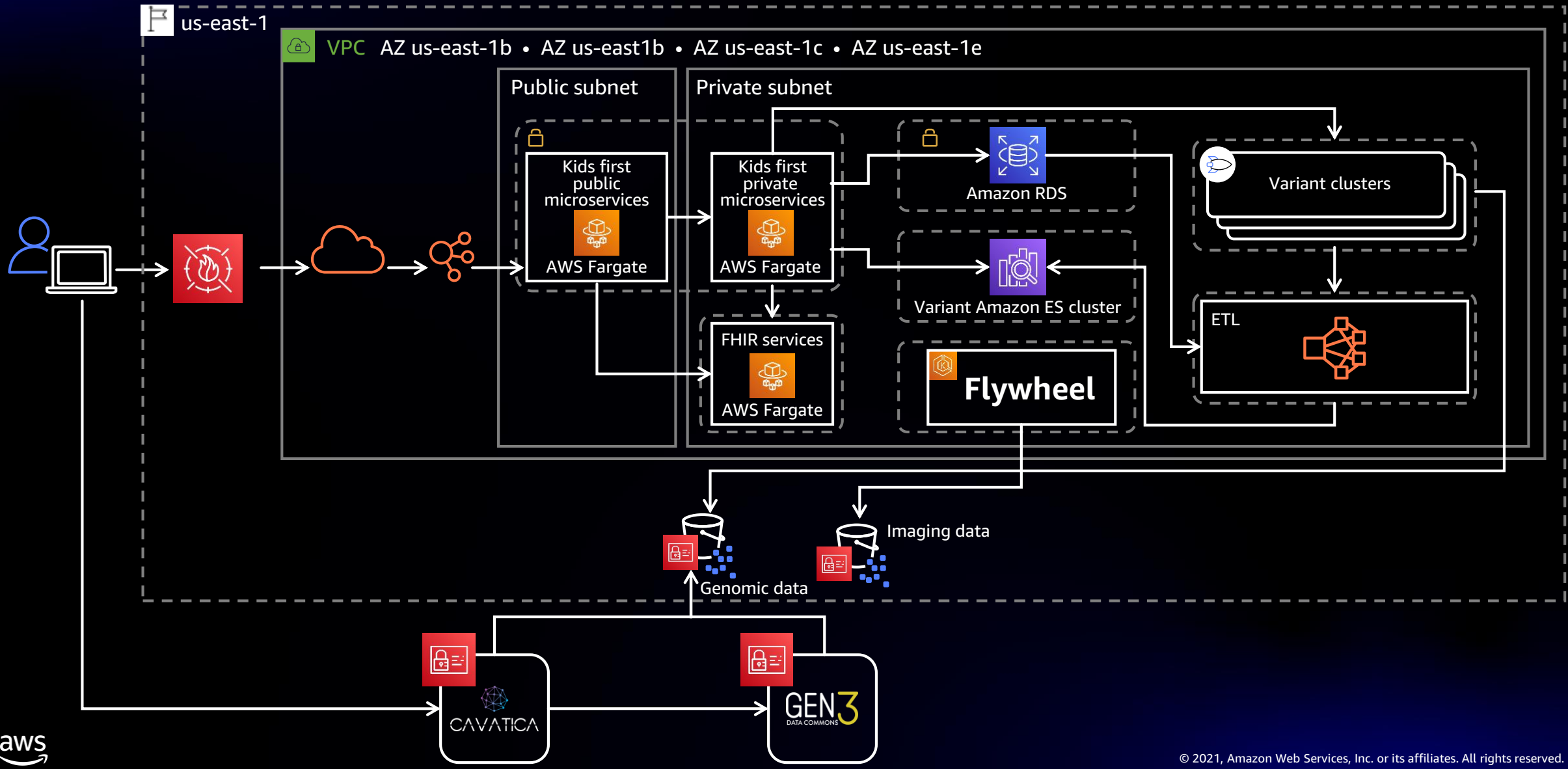
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data integration for impact



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Overarching architecture



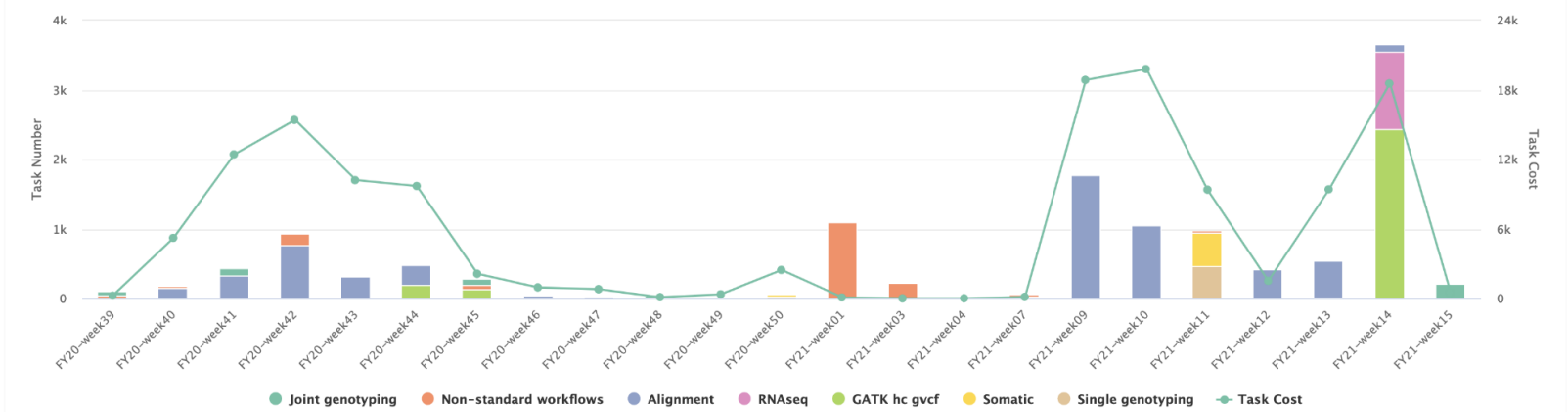
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data to information: scalable cloud pipelines

Overall Report

138465.78	13001	77758.785
Total Cost	Total Number	Total Hours

Tasks Plot | Summary | Detailed Table



*Standard workflows include: Alignment, Somatic, RNAseq, Joint genotyping, Single joint genotyping and GATK haplotypecaller gvcf.

Task_status	Total_number	Sum_cost	Min_cost	Median_cost	Mean_cost	Max_cost	Sum_hours	Min_hours	Median_hours	Mean_hours	Max_hours
COMPLETED	10451	130861.46	0	7.05	12.5214	139.64	72248.0986	0.0033	4.7392	6.913	47.4183
FAILED	746	5733.9	0.01	7.12	7.6862	102.03	3706.9858	0.0269	4.4858	4.9691	23.065
ABORTED	626	1870.42	0	0	2.9879	195.27	1803.7006	0.0022	1.6467	2.8813	92.5761
RUNNING	2	0	0	0	0	0	0	0	0	0	0
DRAFT	1176	0	0	0	0	0	0	0	0	0	0
TOTAL	13001	138465.78	0.01	14.17	23.1955	436.94	77758.785	0.0324	10.8717	14.7634	163.0594

COMMON WORKFLOW LANGUAGE

Docker

CAVATICA

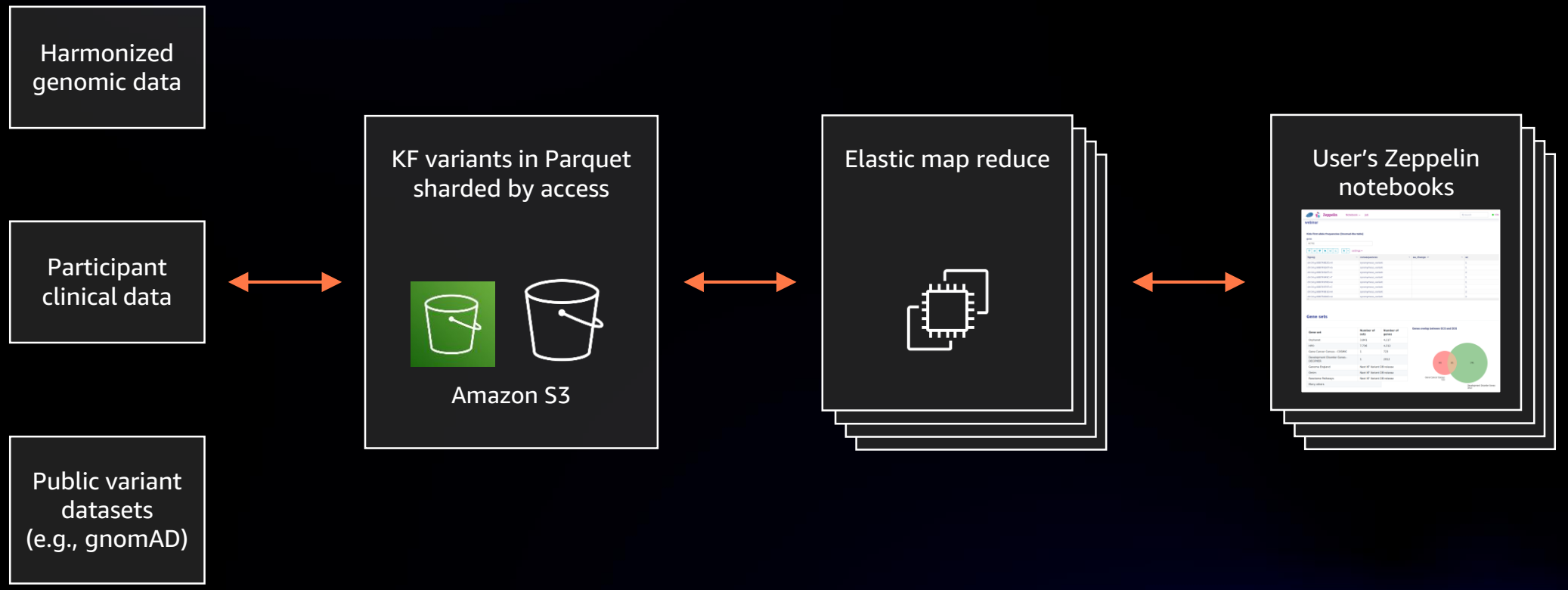
GitHub

github.com/kids-first

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

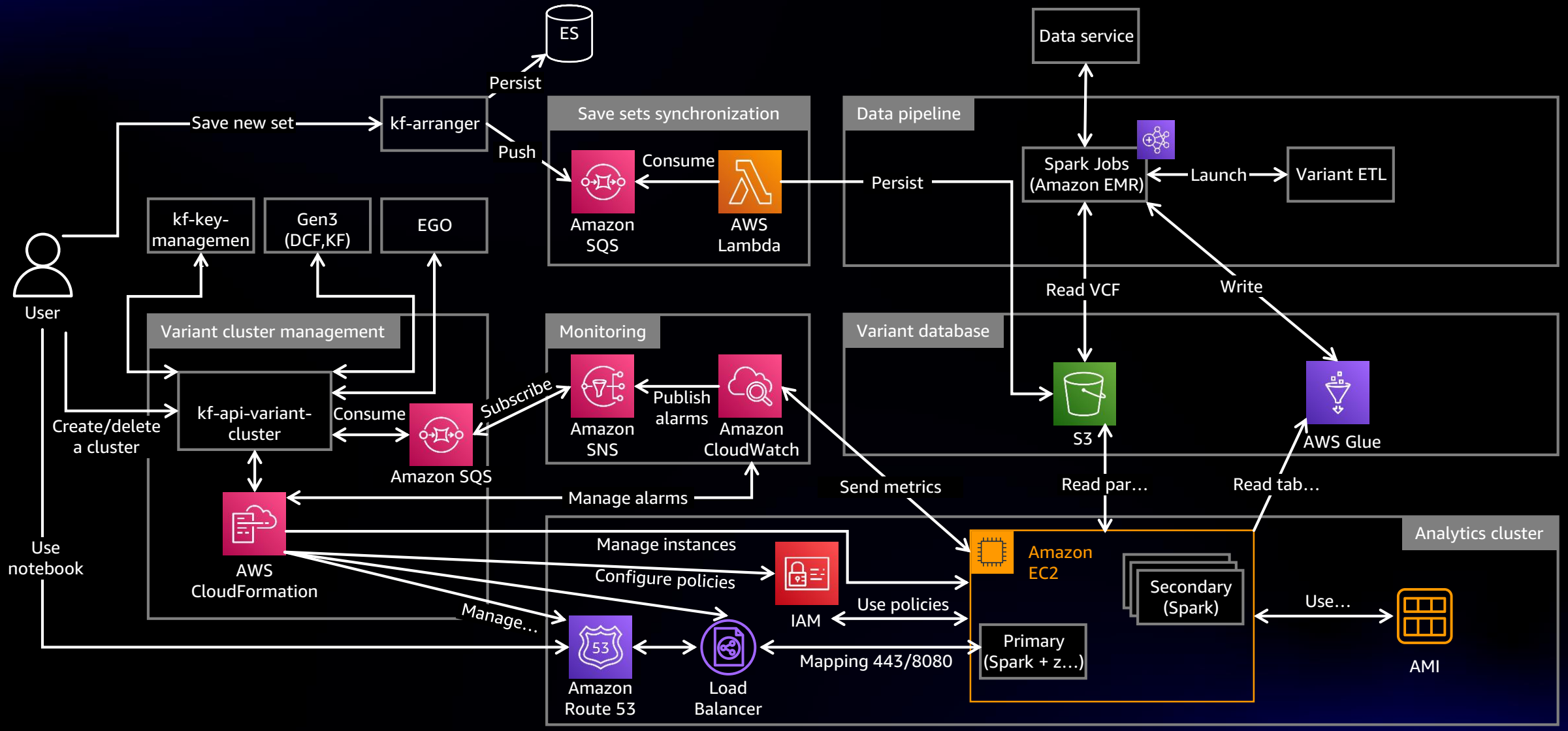
Information to knowledge

VARIANT WORKBENCH



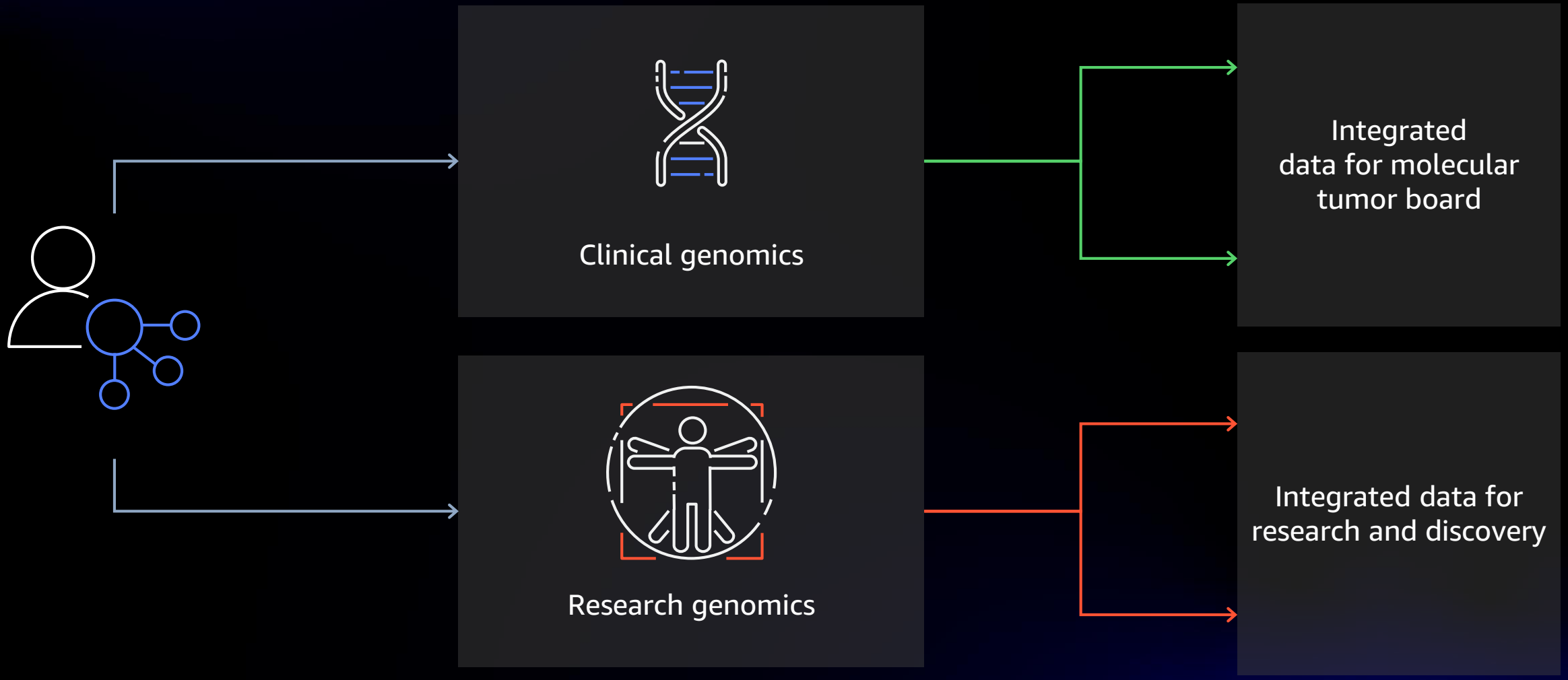
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Variant workbench infrastructure



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Clinical genomics in real time



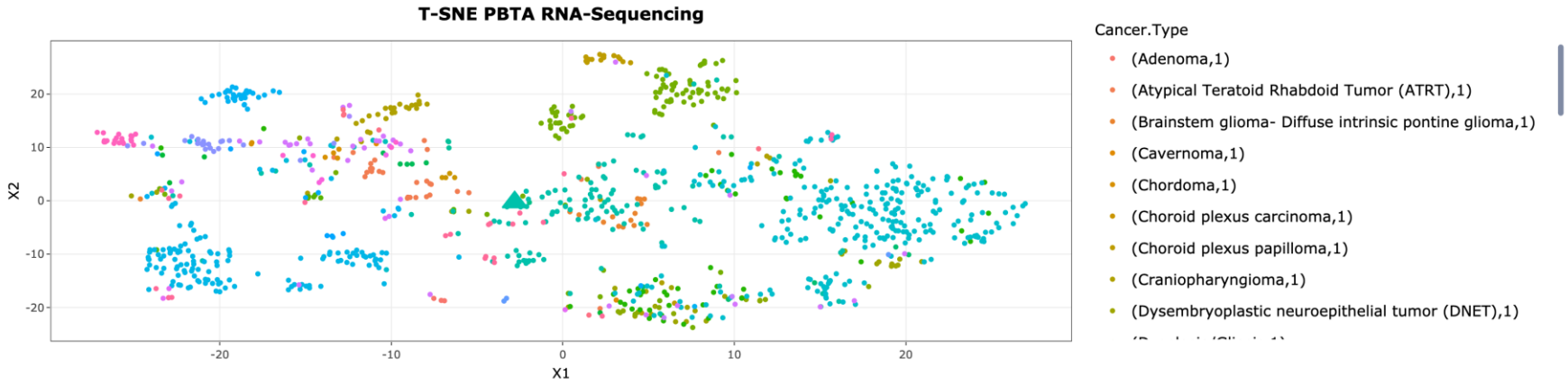
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Example report

Center for Data-Driven Discovery in Biomedicine

Patient ID	Report Date	Report Version
PT_TVJ5EGMR	2/27/20	V1.01

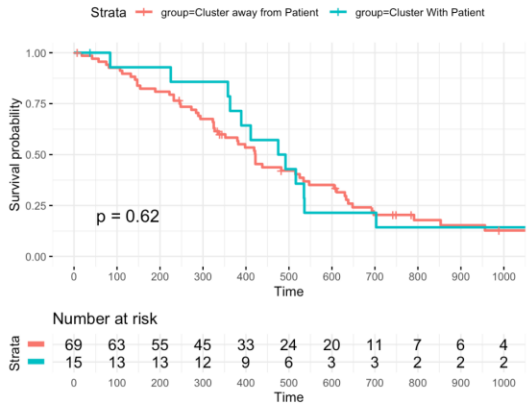
Nearest-Neighbor Analysis



Cancer.Type

- (Adenoma,1)
- (Atypical Teratoid Rhabdoid Tumor (ATRT),1)
- (Brainstem glioma- Diffuse intrinsic pontine glioma,1)
- (Cavernoma,1)
- (Chordoma,1)
- (Choroid plexus carcinoma,1)
- (Choroid plexus papilloma,1)
- (Craniopharyngioma,1)
- (Dysembryoplastic neuroepithelial tumor (DNET),1)

Survival



Treatment/Outcomes - Genomically Similar Patients

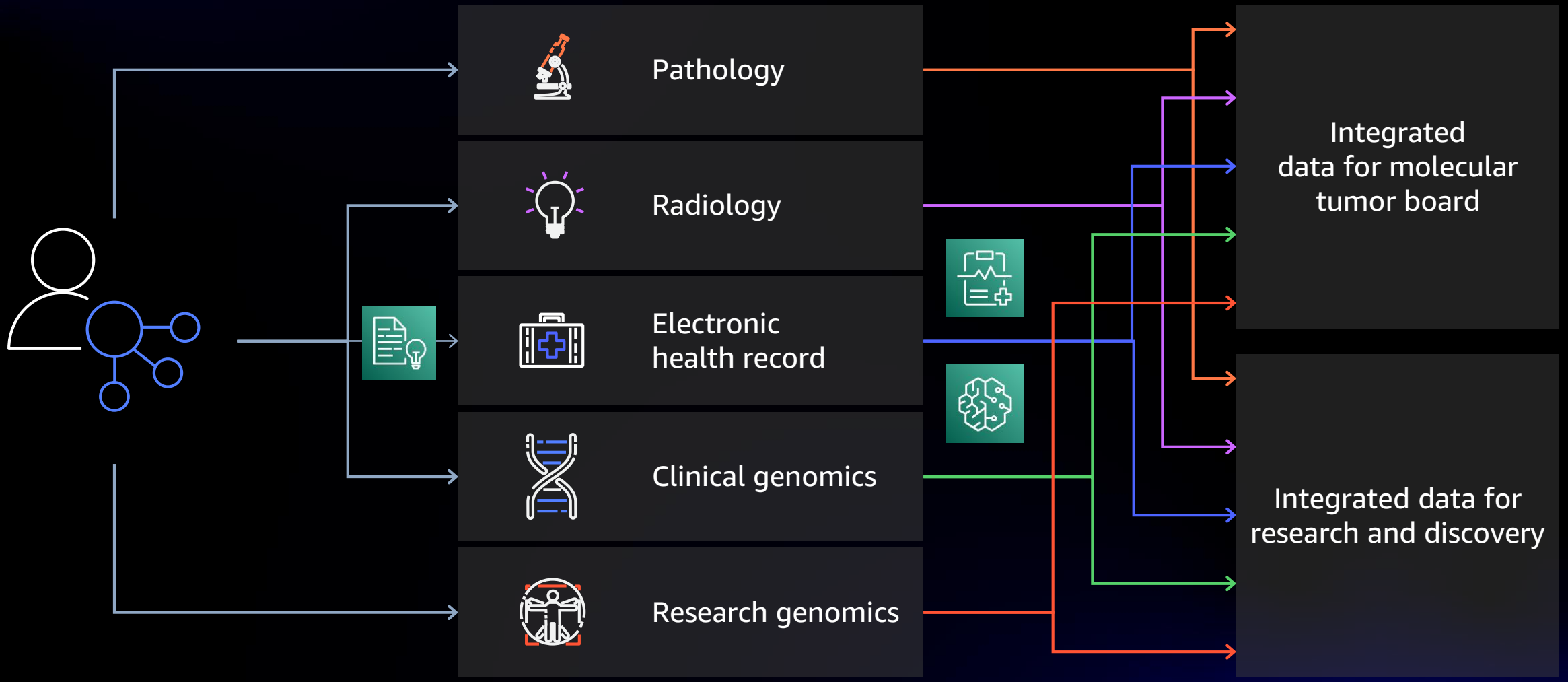
Show 10 entries Search:

Sample.ID	Cancer.Type	AGE	ETHNICITY	RACE	TUMOR_SITE
7316-1059-T-351287	High-grade glioma/astrocytoma (WHO grade III/IV)	13	Reported Unknown	White	Basal Ganglia, Ventricles
7316-1763-T-365898	High-grade glioma/astrocytoma (WHO grade III/IV)	10	Not Hispanic or Latino	White	Temporal Lobe, Thalamus, Ventricles
7316-1774-T-366086	High-grade glioma/astrocytoma (WHO grade III/IV)	6	Hispanic or Latino	White	Temporal Lobe
7316-1937-	High-grade glioma/astrocytoma (WHO	8	Not Hispanic or	Asian	Thalamus

Showing 1 to 10 of 15 entries Previous 1 2 Next

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Using AWS services for impact



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why AWS Cloud?

WHY IS THE CLOUD IMPACTFUL?

Accelerating data flows from data generation to community availability

- Utilized wide portfolio of services to handle clinical, genomic, and imaging workflows
- Rapidly scaled while maintaining different views for research and clinical purposes
- Data available in more “real time” for advanced downstream analytics

Empowering small infrastructure and engineering teams

- Critical in the space of rare/pediatric diseases
- Addressed security and compliance requirements for biomedical research data

Interoperating with commercial and academic partners on AWS

- Cavatica (Seven Bridges), Flywheel, Gen3 (University of Chicago)

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS for genomics: Clinical interpretation of genomic data using machine learning

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon SageMaker

MOST COMPLETE, END-TO-END ML SERVICE

Integrated workbench

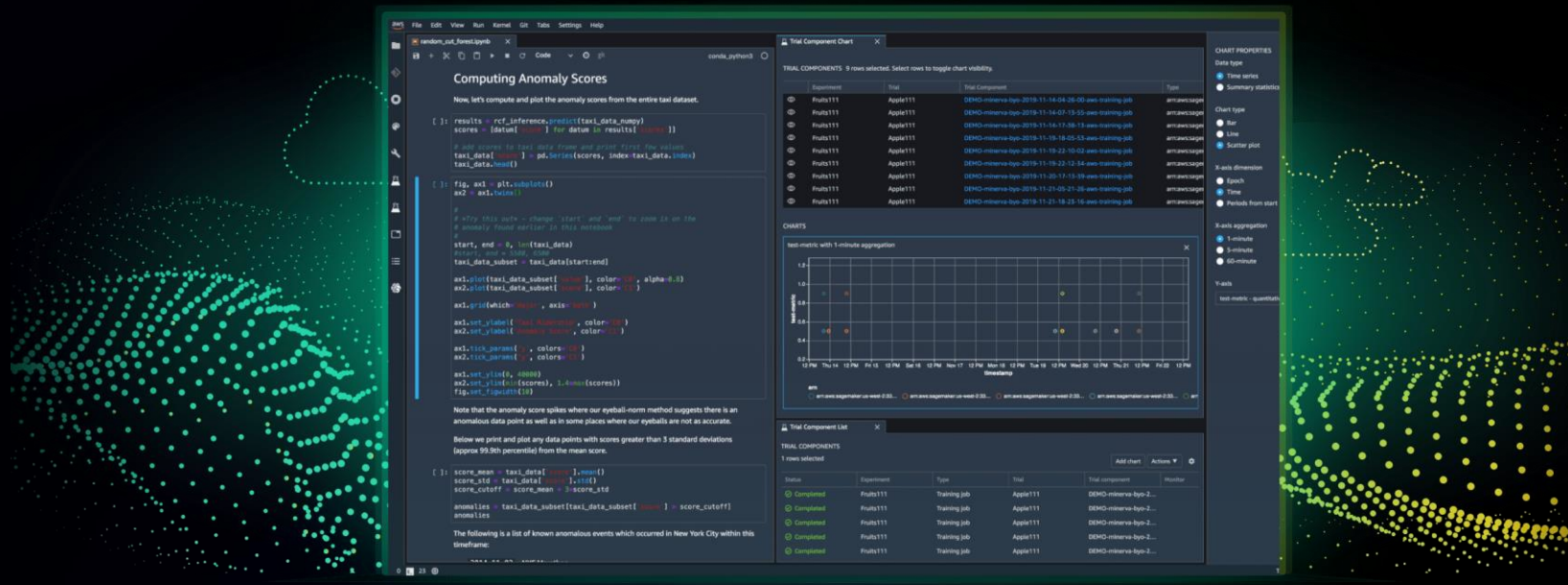
Capabilities designed specifically for ML, data preparation, experiment management, and workflows

Managed infrastructure

Designed for ultra low latency and high throughput, automatic scaling, and distributed training

Managed tooling

Purpose-built from the ground up to work together, including auto ML, collaboration, debugger, profiler, bias analyzer, and explainability



https://aws.amazon.com/sagemaker

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon SageMaker overview

Amazon SageMaker

PREPARE

SageMaker Ground Truth
Label training data for ML

SageMaker Data Wrangler NEW
Aggregate and prepare data for ML

SageMaker Processing
Built-in Python, BYO R/Spark

SageMaker Feature Store NEW
Store, update, retrieve, and share features

SageMaker Clarify NEW
Detect bias and understand model predictions

BUILD

SageMaker Studio notebooks
Jupyter notebooks with elastic compute and sharing

Built-in and bring your-own algorithms
Dozens of optimized algorithms or bring your own

Local mode
Test and prototype on your local machine

SageMaker Autopilot
Automatically create ML models with full visibility

SageMaker JumpStart NEW
Prebuilt solutions for common use cases

TRAIN & TUNE

Managed training
Distributed infrastructure management

SageMaker Experiments
Capture, organize, and compare every step

Automatic model tuning
Hyperparameter optimization

Distributed training libraries NEW
Training for large datasets and models

SageMaker Debugger NEW
Debug and profile training runs

Managed spot training
Reduce training cost by 90%

DEPLOY & MANAGE

Managed deployment
Fully managed, ultra low latency, high throughput

Kubernetes & Kubeflow integration
Simplify Kubernetes-based ML

Multi-model endpoints
Reduce cost by hosting multiple models per instance

SageMaker Model Monitor
Maintain accuracy of deployed models

SageMaker Edge Manager NEW
Manage and monitor models on edge devices

SageMaker Pipelines NEW
Workflow orchestration and automation

SageMaker Studio
Integrated development environment (IDE) for ML

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

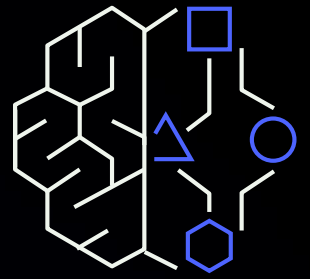
Amazon SageMaker distributed training libraries

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Challenges with large datasets and models



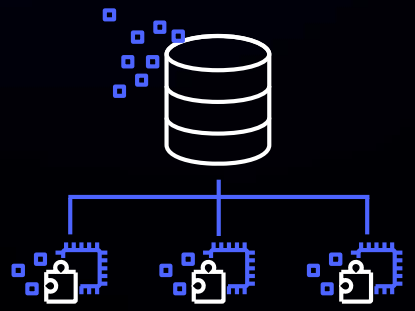
Large datasets take a long time to train, creating a bottleneck



Large models or models with many parameters can exceed the memory capacity of a single GPU

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

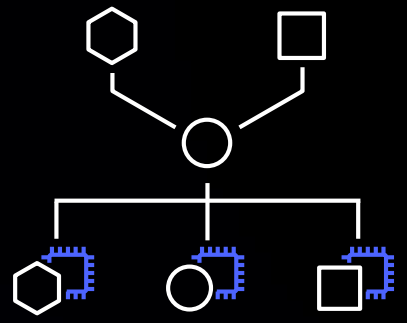
Solutions are hard to implement



Data parallelism

Existing solutions not optimized for AWS

Requires extensive experimentation



Model parallelism

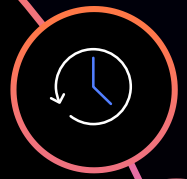
Bigger GPUs are more expensive and have memory limits

Requires expertise in model partitioning

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

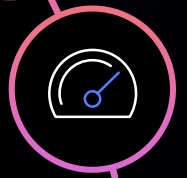
Amazon SageMaker distributed training libraries

The fastest and easiest way to train large deep learning models



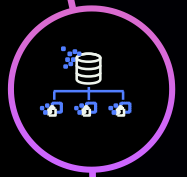
Reduced training time

Reduce training time by 25% with synchronization across GPUs



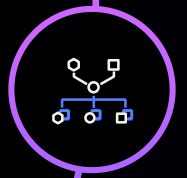
Optimized for AWS

Achieve near-linear scaling efficiency with data parallelism designed for AWS



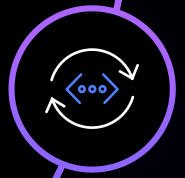
Support for popular ML framework APIs

Reuse existing APIs, such as Horovod, without custom training code



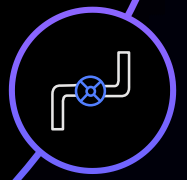
Automatic and efficient model partitioning

Avoid experimentation with automated model profiling and partitioning



Minimal code change

Implement model parallelism with fewer than 10 lines of code change



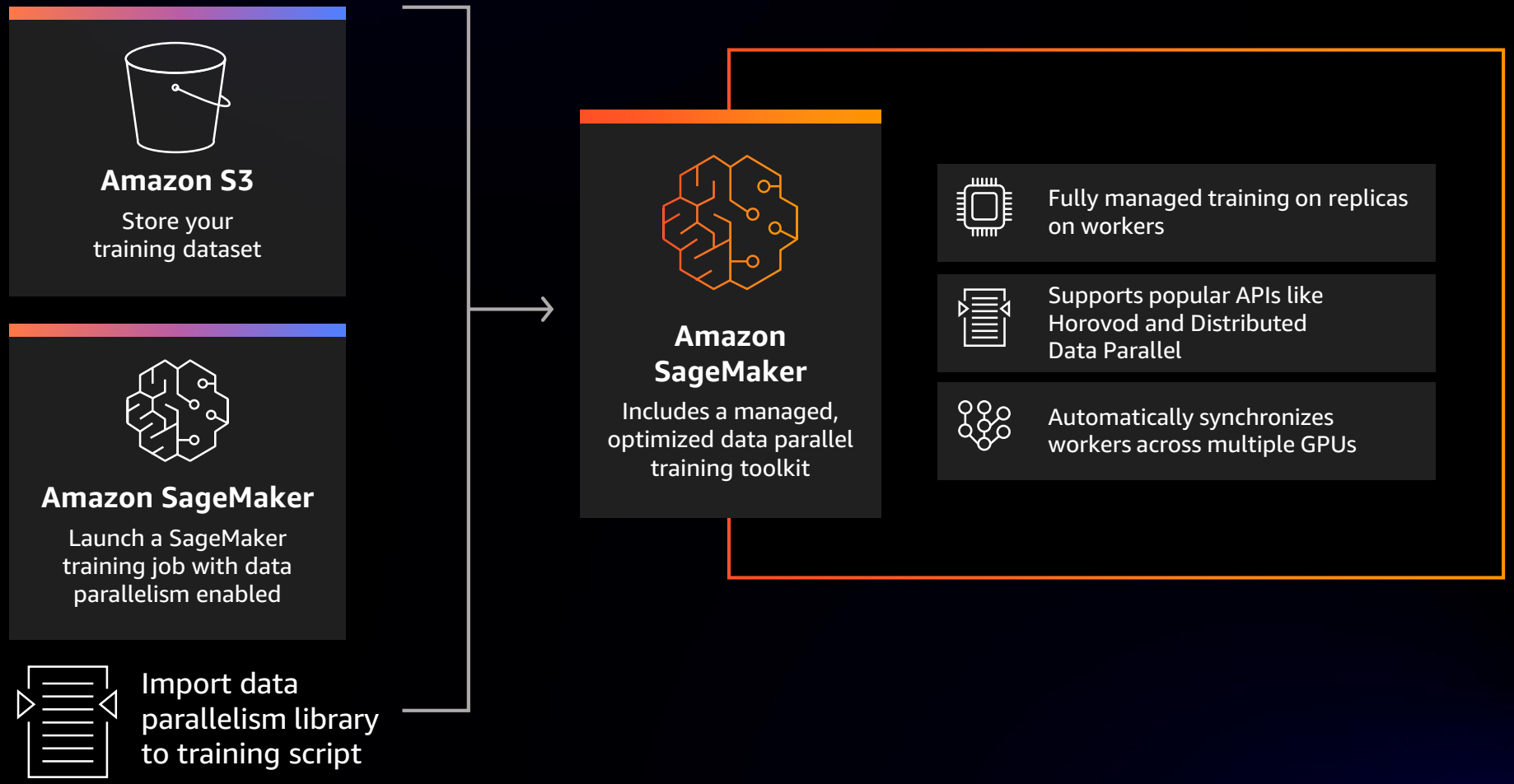
Efficient pipelining

Maximize resource usage with pipelining of microbatches that keeps all GPUs active

aws

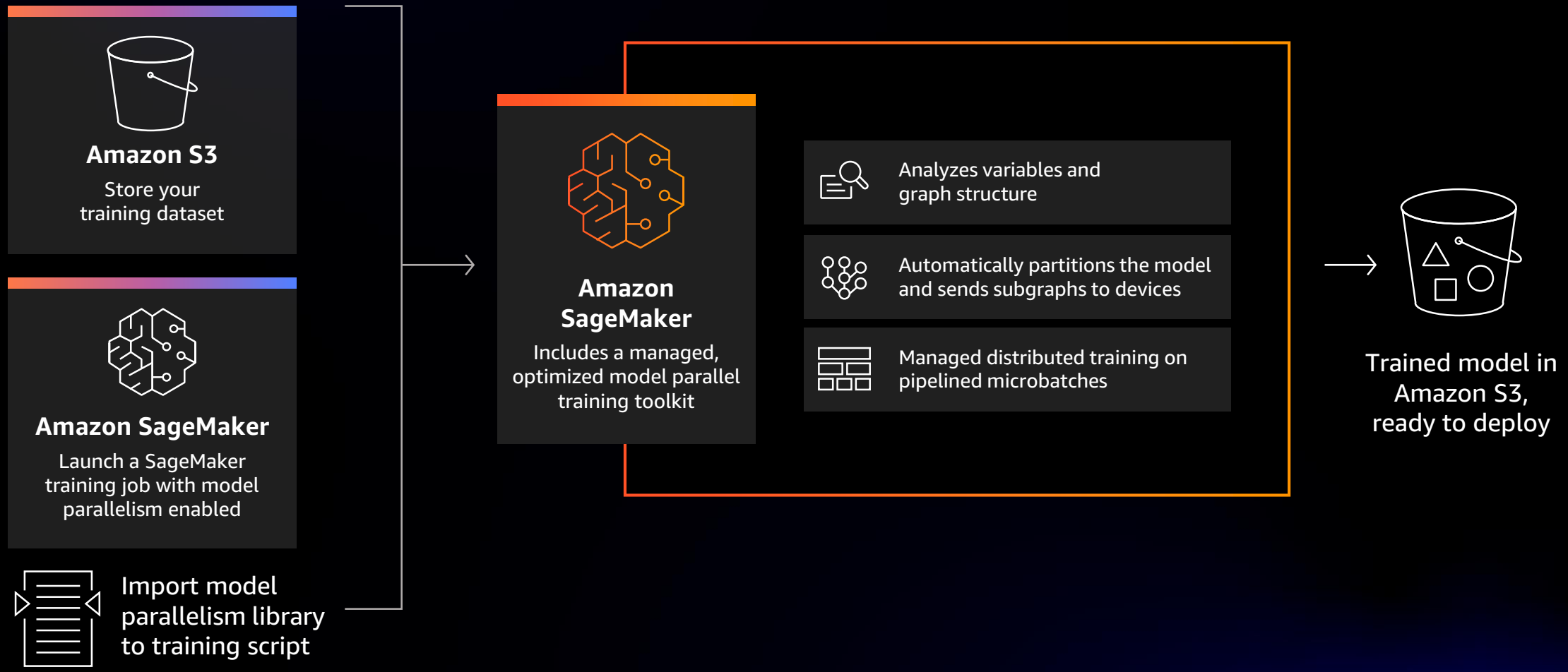
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

How it works: Data parallelism library



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

How it works: Model parallelism library



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Performance: Model parallelism library

GPT-2 LARGE, TENSORFLOW, P3.16XLARGE, 2 PARTITIONS, SAMPLES/SECOND

Number of GPUs	Without `modelparallel`	With `modelparallel`
2	Out of memory; training fails	15.3
4	Out of memory; training fails	24.0

BERT LARGE, TENSORFLOW, P3.16XLARGE, 2 PARTITIONS, SAMPLES/SECOND

Number of GPUs	Without `modelparallel`	With `modelparallel`
8	208.8	264.0 (1.25x faster)

T5-3B, PYTORCH, P3DN.24XLARGE, 8 PARTITIONS, SAMPLES/SECOND

Number of GPUs	Without `modelparallel`	With `modelparallel`
8	Out of memory; training fails	18.8

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Extracting clinical entities from genetic reports and literature

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

80% of the patient care information is “trapped” in free-form medical text



Insurance claims



History and physical



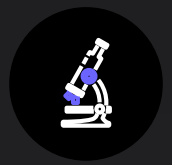
Visit summaries



Progress notes



Operative notes



Pathology reports



Medical charts



Diagnostic imaging reports



Discharge notes



And many more

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon Comprehend Medical distills a complex process into a simple API call

Medical Named Entity and Relationship Extraction (NERe API)

Protected Health Information Identification (PHId API)*

Entities

Medication
Medical condition
Test, treatments, and procedures
Anatomy
Protected Health Information (PHI)

Relationship extraction

Medication and dosage
Test and result
Many more

Entity traits

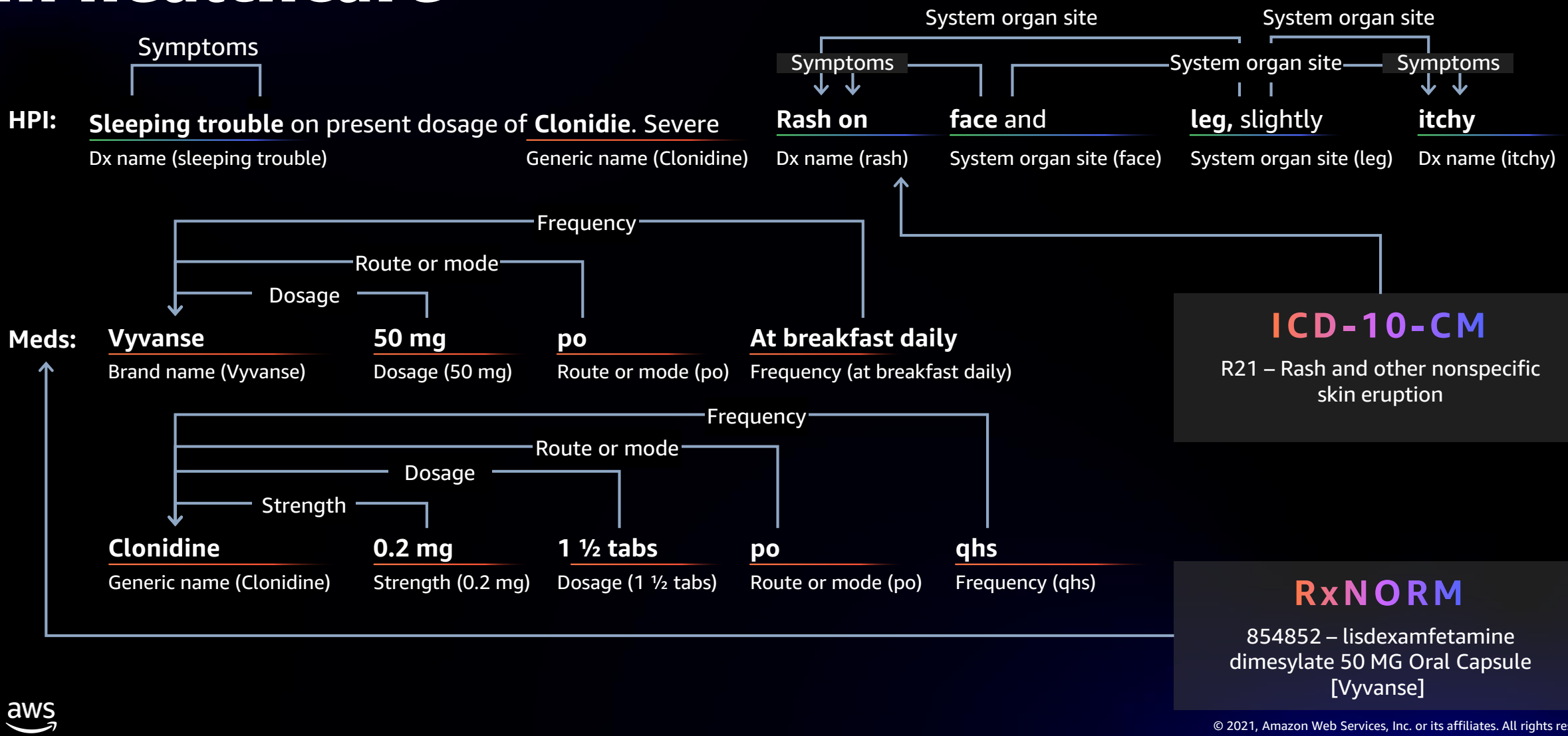
Negation
Temporality
Diagnosis, sign, or symptom

Service is HIPAA-eligible and "stateless," no customer data stored

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Medical ontology linking is critical in healthcare



ICD-10-CM

R21 – Rash and other nonspecific skin eruption

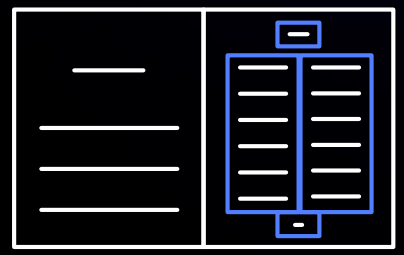
RxNORM

854852 – lisdexamfetamine dimesylate 50 MG Oral Capsule [Vyvanse]

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

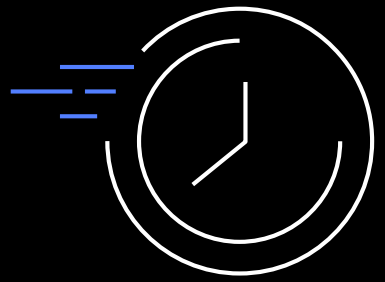
Benefits of Amazon Comprehend Medical



Extract medical information quickly and accurately



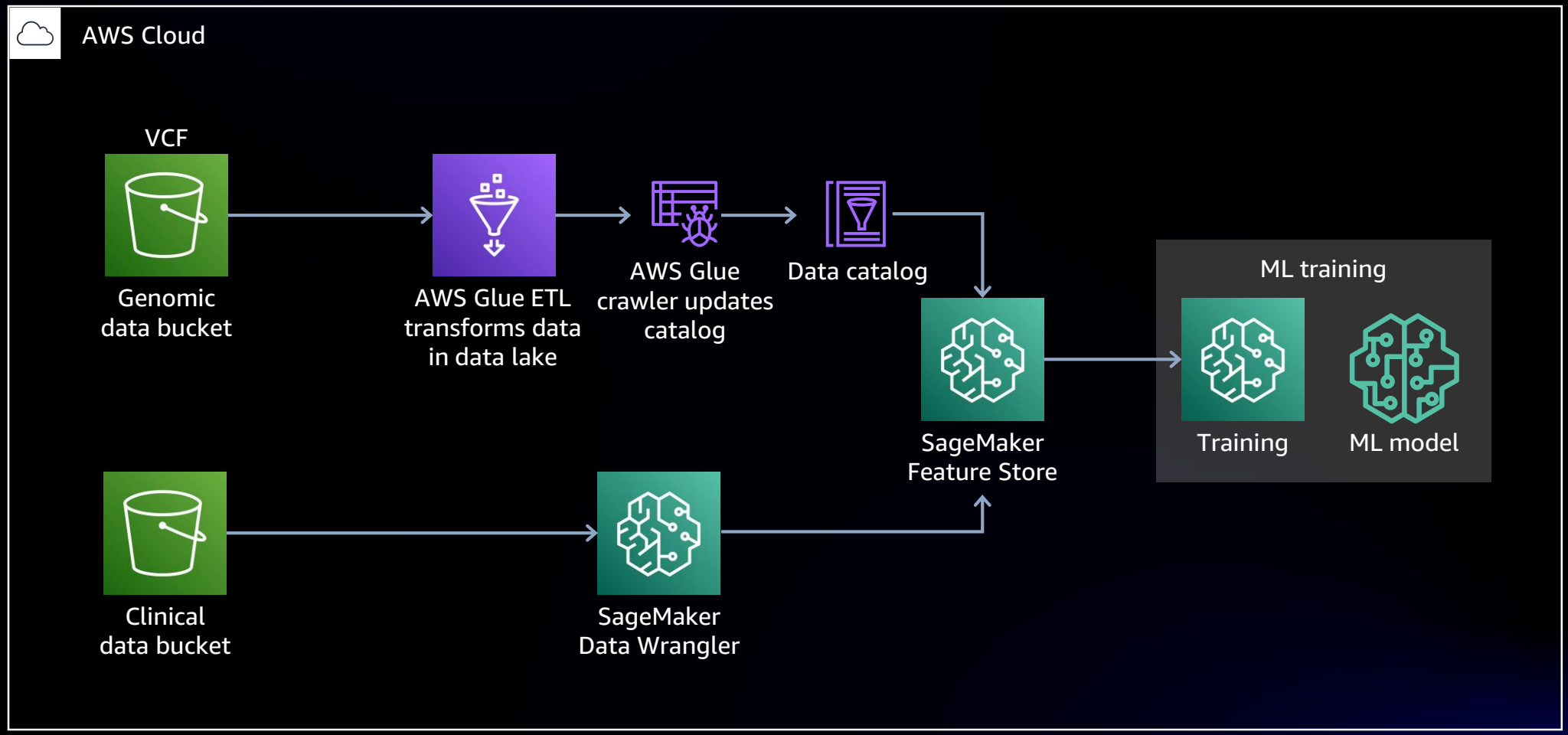
HIPAA-eligible (to help protect patient information)



Faster processing (25x for simple small charts, 375x for rare disease and oncology charts) at lower cost

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Combining genomics and clinical data for multimodal learning



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Supporting healthcare interoperability on AWS

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon HealthLake

A HIPAA-eligible service that enables healthcare providers, health insurance companies, and pharmaceutical companies to securely store, transform, query, and analyze health data at petabyte scale



IMPORT

Quickly and easily import medical records, including clinical notes, lab reports, and more



STORE

Stored in the AWS Cloud in a secure, compliant, and auditable way



TRANSFORM

Tag and index unstructured data using specialized ML models



QUERY & SEARCH

Powerful query and search capabilities to ask questions of the data



ANALYZE

Understand relationships in the data with integrated analytics and ML capabilities

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Purpose-built and HIPAA-eligible services

AI/ML SERVICES FOR HEALTHCARE AND LIFE SCIENCES USE CASES



Amazon HealthLake

Store, transform, query, and analyze health data in minutes



Amazon Comprehend Medical

Understand medical context with advanced text analytics



Amazon Transcribe Medical

Automatically convert medical speech to text



Amazon Textract

Easily extract text and data from virtually any medical document



Amazon Rekognition

Automate medical image and video analysis

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Visit our other healthcare sessions

HLC304 (chalk talk): See the bigger picture with machine learning for multimodal data

HLC307 (workshop): Using Fast Healthcare Interoperability Resources (FHIR) on AWS

HLC302 (builders' session): Building a HIPAA-eligible virtual care solution on AWS

HLC306 (chalk talk): Healthcare data center migration: Best practices and approaches

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Ujjwal Ratan
ujjwalr@amazon.com

Allison P. Heath, PhD
heathap@email.chop.edu

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.