



# Clinical Trials Data Repository on AWS

## Customer Challenges:

Healthcare and Life Science organizations want to make their data and analytic tools more easily accessible across teams and departments to improve clinical operations and speed up clinical development while ensuring data is secured for only authorized users. Applying analytics, artificial intelligence, and machine learning (AI/ML) across disparate datasets can provide new insights that improve clinical efficiencies, boost profitability, and deliver high-quality patient services. While storing healthcare data is key to these initiatives, parsing and classifying the wide range of data types is challenging.

## AWS Health for Data:

AWS makes it easier for customers to ingest, classify, and securely share clinical datasets at scale across organizational boundaries, helping organizations uncover insights from disparate datasets to improve clinical operations and clinical development. Customers have numerous options to help ingest a variety of Healthcare and Life Science data through services such as AWS Data Exchange and Amazon Kinesis. AWS Glue makes classifying data quick and easy for customers while Amazon DataZone unlocks data across the organization securely with a simple to use portal.

## Top Benefits:



**Streamline third-party data consumption**



**Classify data quickly and easily**

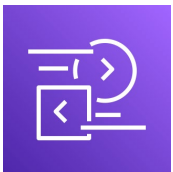


**Unlock data across the organization**



**Unify business intelligence**

## Featured AWS Services:



### AWS Data Exchange

Easily find, subscribe to, and use third-party data



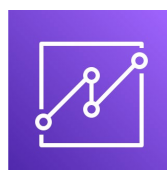
### Amazon Kinesis

Collect, process, and analyze real-time data streams



### Amazon DataZone

Unlock data across organizational boundaries with built-in governance

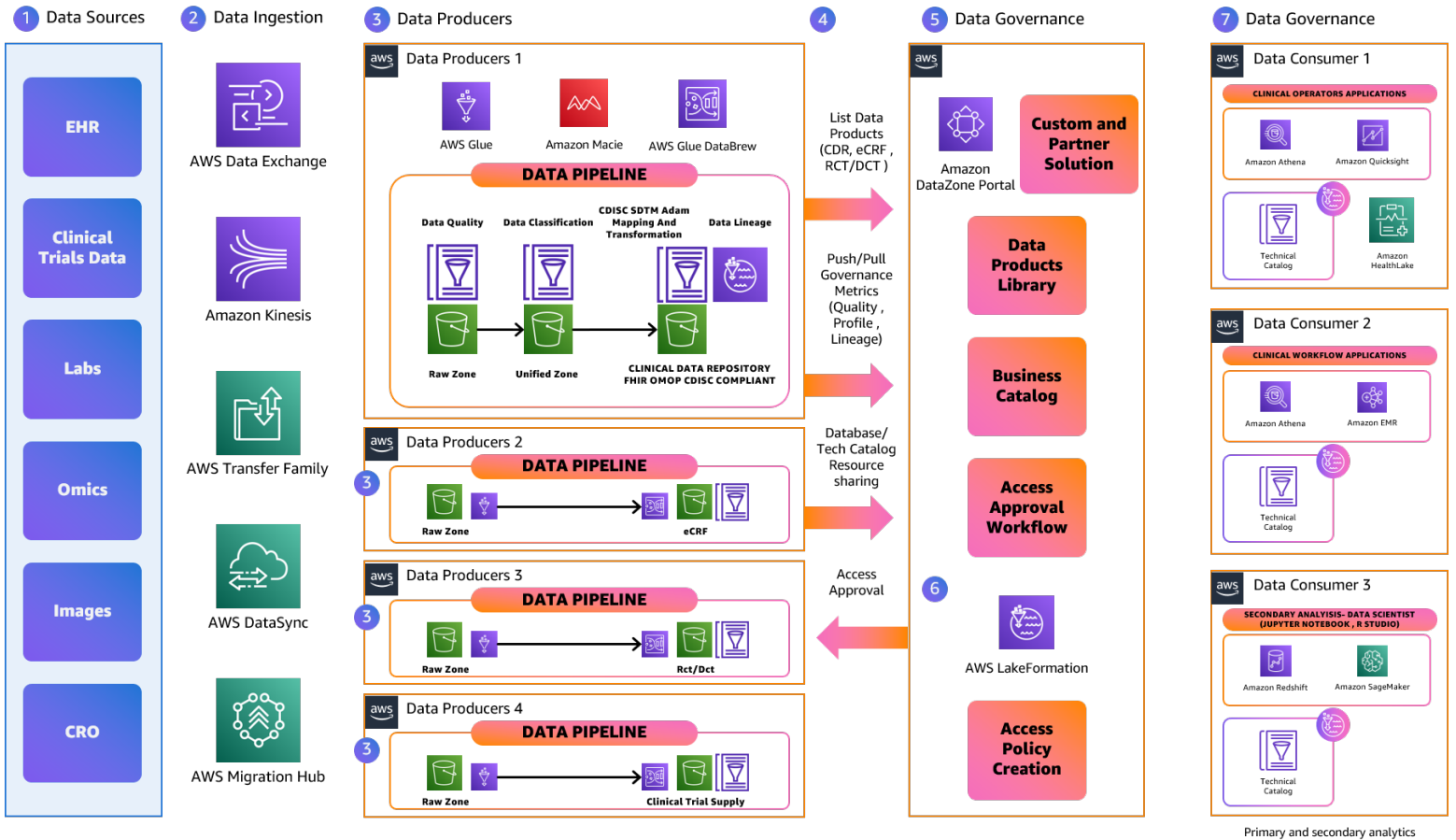


### Amazon QuickSight

Unified business intelligence at hyperscale



# Clinical Trials Data Architecture



1/ Diverse set of data sources. Intermediate data is stored in Amazon Simple Storage Service (Amazon S3).

2/ Ingested data is saved to the raw S3 layer automatically or on demand through a variety of AWS native services such as AWS Data Exchange, AWS Transfer Family, Amazon Kinesis, AWS DataSync, and AWS Migration Hub that help customers discover, plan, and orchestrate their migration plan.

3/ Data is cataloged using AWS Glue data catalog. Using fit-for-purpose data processing frameworks, the ingested data is cleaned, curated, transformed and stored in a purpose-built storage that can be published as a data product.

4/ Data products are published to the enterprise business catalog along with metadata (schema / data quality metrics/ lineage etc.). The schema is also shared with the central data catalog running on AWS Glue. AWS Lake Formation has a feature for sharing data catalogs across multiple accounts that can leveraged here. Access workflow receives requests from consumers and routes to the data owners for approval.

5/ Share, search, and discover data at scale across organizational boundaries with Amazon DataZone. Collaborate on data projects through a unified data analytics portal that gives you a personalized view of all your data while enforcing your governance and compliance policies.

6/ Amazon DataZone can manage access policies and permissions management for AWS LakeFormation managed AWS Glue tables and Amazon Redshift tables and views.

7/ Publish and subscribe workflow with access management within Amazon DataZone. Use the automated workflow to better secure data between producers and consumers and to ensure access to the right data for the right purpose.