aws

Life Sciences Data Collection, Storage, and Processing

# Life Sciences Data Collection, Storage, and Processing: Architecture Diagrams

Table of Contents

Life Sciences Data Collection, Storage, and Processing

Publication date: **July 20, 2022 ([Diagram history](#))**

This architecture diagram helps you learn how to transfer life sciences data files to the cloud and provide data access using Amazon Web Services (AWS).

Life Sciences Data Collection, Storage, and Processing Diagram
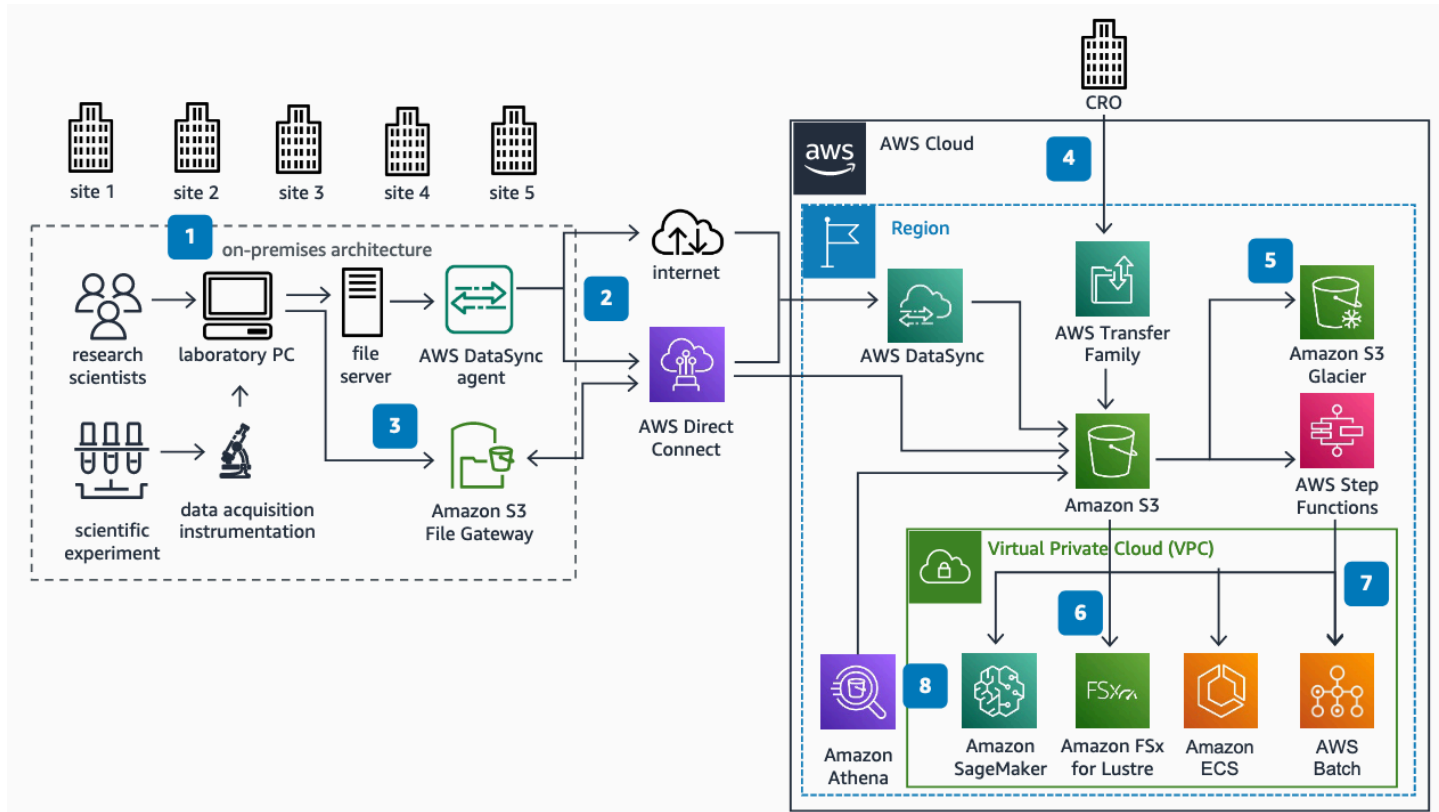
1. A lab technician runs an experiment and results are written to a folder on an on-premises file server. An AWS DataSync task is set up to sync the data from local storage to a bucket in Amazon Simple Storage Service (Amazon S3).

2. Data is transferred to AWS Cloud either through the internet, or through a low-latency direct connection that avoids the internet, such as AWS Direct Connect.

3. On-premises researchers analyze data in Amazon S3 in existing bioinformatics tools by using Network File System (NFS) or Server Message Block (SMB) through Amazon S3 File Gateway.

4. Partnering entities like a contract research organization (CRO) can upload study results to Amazon S3 by using AWS Transfer Family for FTP, SFTP, or FTPS.

5. You can optimize storage by writing instruments that run data to an S3 bucket configured for infrequent access. Identify your S3 storage access patterns to optimally configure your S3 bucket lifecycle policy and transfer data to Amazon S3 Glacier.

6. Using Amazon FSx for Lustre, data is made accessible to high performance computing (HPC) on the cloud for genomics, imaging, and other intensive workloads to provide a low millisecond-latency shared file system.

7. Research HPC workloads are orchestrated on the cloud with AWS Step Functions and AWS Batch, for flexible central processing unit (CPU) and graphics processing unit (GPU) computing on Amazon Elastic Compute Cloud (Amazon EC2) instances or Amazon Elastic Container Service (Amazon ECS) containers.

8. Machine learning is conducted with a common artificial intelligence and machine learning (AI/ML) toolkit that uses Amazon SageMaker AI for feature engineering, data labeling, model training, deployment and ML operations. Amazon Athena is used for flexible SQL queries with existing tools.

Download editable diagram

To customize this reference architecture diagram based on your business needs, download the ZIP file which contains an editable PowerPoint.

Create a free AWS account

Sign up now

Sign up for an AWS account. New accounts include 12 months of AWS Free Tier access, including the use of Amazon EC2, Amazon S3, and Amazon DynamoDB.

Further reading

For additional information, refer to

- AWS Architecture Icons
- AWS Architecture Center
- AWS Well-Architected

Diagram history

To be notified about updates to this reference architecture diagram, subscribe to the RSS feed.

Change	Description	Date
Initial publication	Reference architecture diagram first published.	July 20, 2022

> ⓘ **Note**
>
> To subscribe to RSS updates, you must have an RSS plugin enabled for the browser you are using.