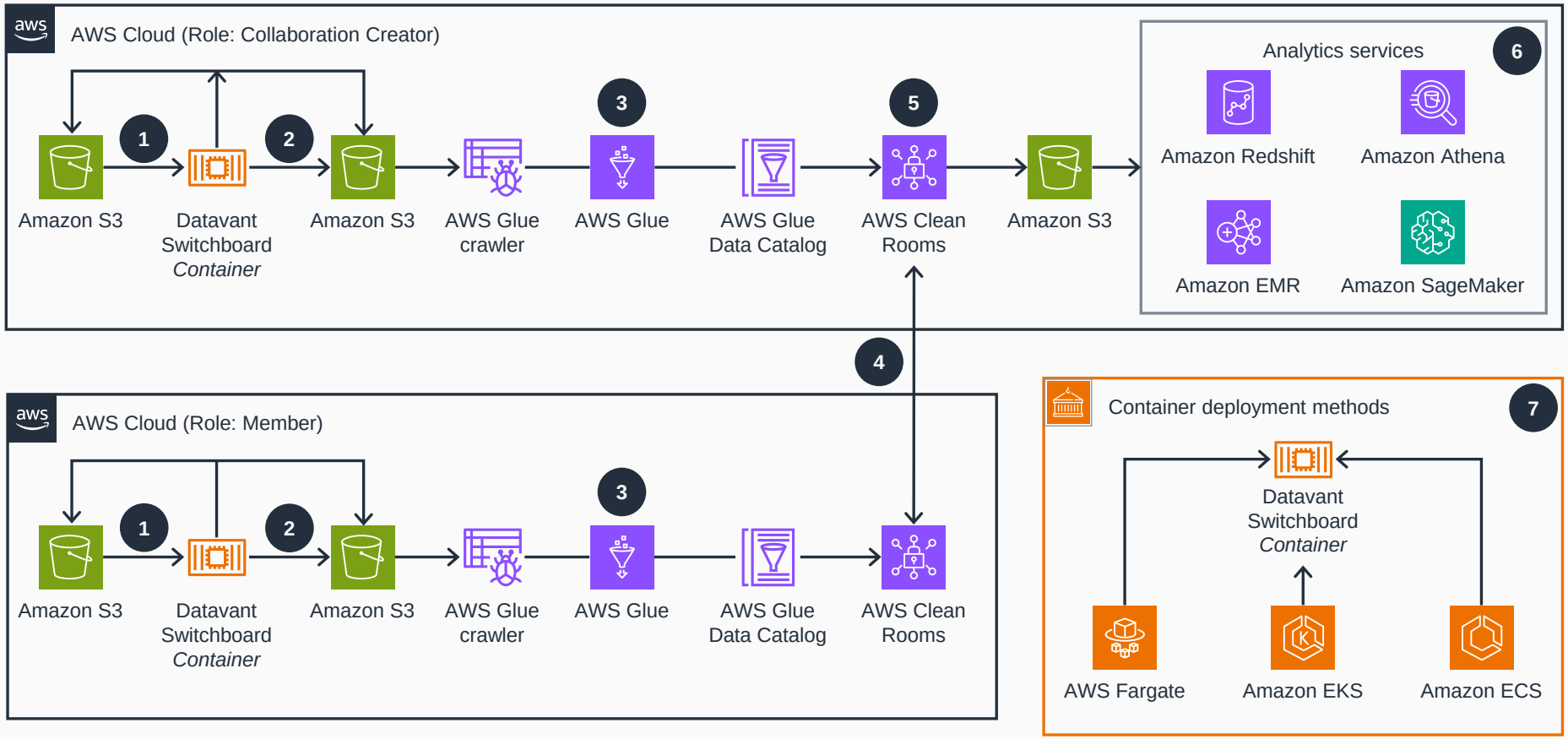
Guidance for Patient and Clinical Data Insights with Datavant and AWS Clean Rooms

This architecture diagram shows how to use Datavant's data deidentification and tokenization tools on AWS Clean Rooms to protect your healthcare and life sciences organization's private patient information. It also shows how to gain insights by replacing the data with encrypted tokens that cannot be reverse engineered.



1. Deidentify or tokenize data in an **Amazon Simple Storage Service (Amazon S3)** bucket using Datavant Switchboard (container). The container is deployed through supported container deployment methods, as detailed in Step 7.
2. Link the tokenized data with your fellow collaborators using the Datavant Switchboard (container) and store the output in an **Amazon S3** bucket.
3. Use an **AWS Glue** crawler to crawl the linked, tokenized data. Prepare the data source for collaboration with **AWS Glue** Data Catalog.
4. Instantiate **AWS Clean Rooms** and invite the member(s) to the collaboration to align on and implement analysis rules. Members can then associate configured tables from Data Catalog and use an **AWS Clean Rooms** service role to access their **AWS Glue** tables.
5. The member who is allowed to query uses Aggregate and List functions across tables in the collaboration. Results can be exported to **Amazon S3** for the member who is allowed to receive query results.
6. The member who receives query results can use analytics services, including **Amazon Redshift**, **Amazon Athena**, **Amazon EMR**, and **Amazon SageMaker**, to derive insights from the newly enriched dataset.
7. Datavant Switchboard container deployment methods include using **AWS Fargate**, **Amazon Elastic Kubernetes Service (Amazon EKS)**, and **Amazon Elastic Container Service (Amazon ECS)**.

Reviewed for technical accuracy November 20, 2024

AWS Reference Architecture