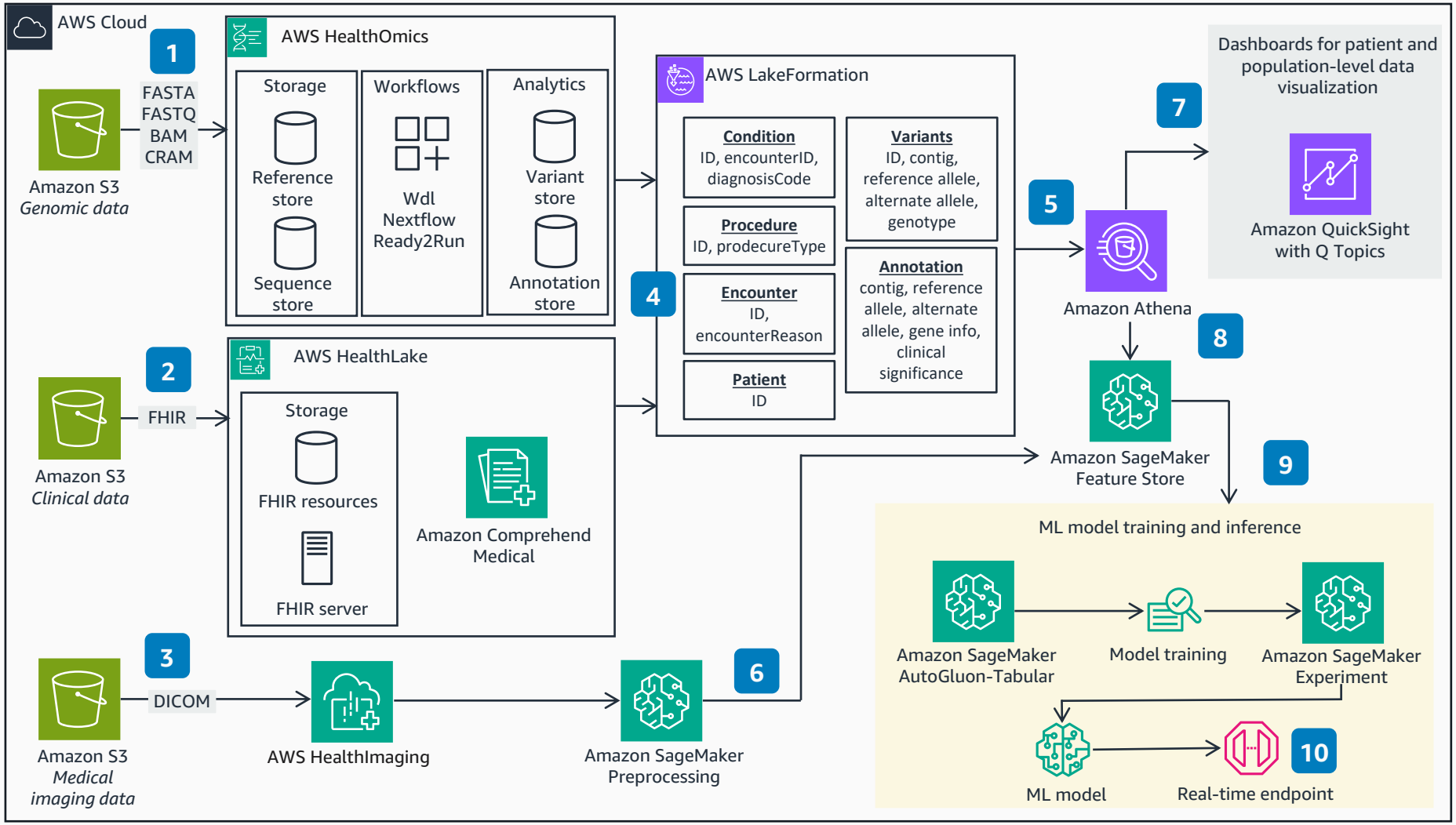
Guidance for Multi-Modal Data Analysis with AWS Health and ML Services

This architecture shows an end-to-end framework for integrating and analyzing genomic, clinical, and medical imaging data.



1. Ingest genomic data from **Amazon Simple Storage Service (Amazon S3)** or **Registry of Open Data on AWS (RODA)** to **AWS HealthOmics**. Use **HealthOmics Reference store** for reference genome data, such as Fast-All (FASTA), and **HealthOmics Sequence store** for sequence data, such as FASTQ, Binary Alignment Map (BAM), and Compressed Reference-oriented Alignment Map (CRAM). Use **HealthOmics Variant store** for VCF files and **HealthOmics Annotation store** for annotation files. To run private or Ready2Run workflows, use **HealthOmics Workflows**.
2. Ingest FHIR data to **AWS HealthLake**.
3. Ingest DICOM images to **AWS HealthImaging** and read into insight toolkit (ITK) image object in-memory through API calls.
4. View tables from **HealthOmics** and **HealthLake** as resources in **AWS Lake Formation**.
5. Query the tables with **Amazon Athena**.
6. Generate brain masking with the Medical Open Network for AI (MONAI) segmentation model. Use **Amazon SageMaker Preprocessing** to parallelize radiomic feature computation for each image representation.
7. Build visualization dashboards with **Amazon QuickSight**.
8. Store the multimodal feature set in **Amazon SageMaker Feature Store**.
9. Build and train ML models on multimodal features with **SageMaker AutoGluon-Tabular**.
10. Deploy the model as an endpoint for real-time inference.

Reviewed for technical accuracy July 19, 2023

AWS Reference Architecture