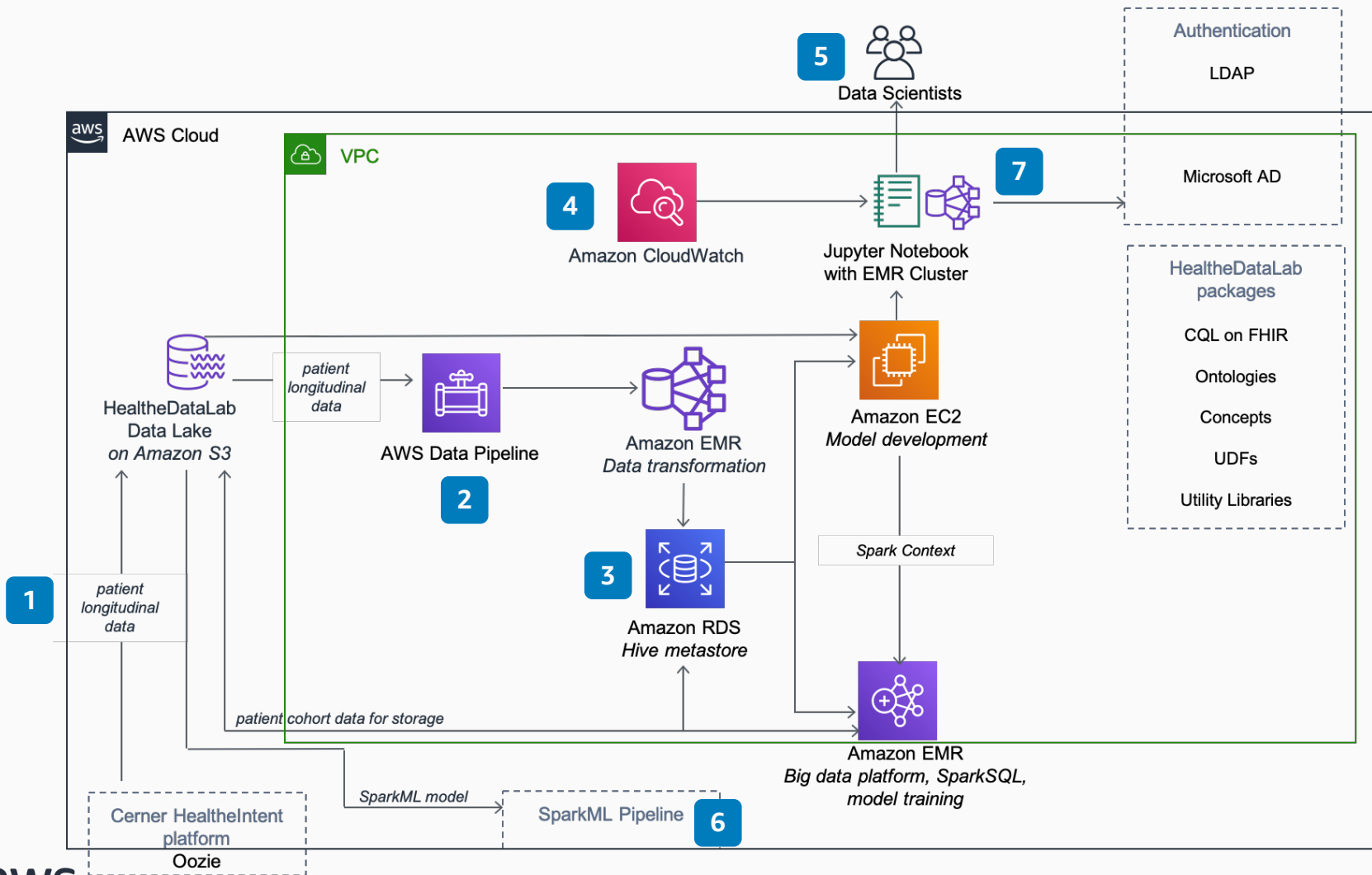


# Cerner HealthDataLab on AWS

Use Cerner HealthDataLab on AWS to develop machine learning models and algorithms for healthcare. For more information, see the [Overview of Cerner HealthDataLab on AWS](#) whitepaper.



- 1 Patient longitudinal data is delivered via the Cerner HealthIntent platform to the HealthDataLab data lake hosted on **Amazon Simple Storage Service (Amazon S3)**.
- 2 **AWS Data Pipeline** extracts the patient longitudinal data and launches an **Amazon EMR cluster** which processes it into metadata describing patient records, including sorting into records cohorts sharing comorbidities.
- 3 Cohort data is stored as a Hive metastore on **Amazon Relational Database Service (Amazon RDS)**.
- 4 For any new patient data, Amazon CloudWatch launches an **Amazon EC2 instance** triggering an algorithm and ML model development workflow.
- 5 Data scientists can interact with and analyze data using **Jupyter notebooks**. The attached **EMR cluster** allows model training and development.
- 6 Trained and verified models and algorithms are deployed using SparkML Pipeline.
- 7 **Amazon EMR clusters** host HealthDataLab packages and user-defined functions to support data scientist activities.

