CRITICAL PATH INSTITUTE

Real-world evidence in the cloud: Tutorial on developing an end-to-end data and analytics pipeline using Amazon Web Services resources

Wes Anderson, Roopal Bhatnagar, Keith Scollick, Marco Schito, Ramona Walls, Jagdeep T. Podichetty

Abstract

In the rapidly evolving landscape of healthcare and drug development, the ability to efficiently collect, process, and analyze large volumes of real-world data (RWD) is critical for advancing drug development. This article provides a blueprint for establishing an end-to-end data and analytics pipeline in a cloud-based environment. The pipeline presented here includes four major components, including data ingestion, transformation, visualization, and analytics, each supported by a suite of Amazon Web Services (AWS) tools. The pipeline is exemplified through the CURE ID platform, a collaborative tool designed to capture and analyze real-world, off-label treatment administrations. By using services such as AWS Lambda, Amazon Relational Database Service (RDS), Amazon QuickSight, and Amazon SageMaker, the pipeline facilitates the ingestion of diverse data sources, the transformation of raw data into structured formats, the creation of interactive dashboards for data visualization, and the application of advanced machine learning models for data analytics. The described architecture not only supports the needs of the CURE ID platform, but also offers a scalable and adaptable framework that can be applied across various domains to enhance data-driven decision making beyond drug repurposing.

Read the publication in its entirety here.