Data Processing using Amazon Managed Apache Flink

Real-time data processing is a critical component in modern data architectures, enabling businesses to extract valuable insights and make informed decisions as events unfold. Apache Flink, a powerful open-source stream processing framework, has emerged as a leading solution for handling real-time data streams.

In this module, we'll explore Amazon Managed Apache Flink, a fully managed service that simplifies the deployment and management of Apache Flink applications and leverage its Studio Notebooks powered by Apache Zeppelin to interactively query data streams in real time, and easily build and run stream processing applications using standard SQL, Python, and Scala. With a few clicks in the AWS Management console, you can launch a serverless notebook to query data streams and get results in seconds.

Apache Zeppelin provides Studio notebooks with a complete suite of analytics tools, including the following:

1. Data Visualization
2. Exporting data to files
3. Controlling the output format for easier analysis

With a notebook, you model queries using the Apache Flink Table API & SQL in SQL, Python, or Scala, or DataStream API in Scala. With a few clicks, you can then promote the Studio notebook to a continuously-running, non-interactive, Managed Service for Apache Flink stream-processing application for your production workloads.

Previous Next