Loading data into Neo4j using Kettle and Apache Beam (via Google Dataflow)

Rodrigo Haces - Customer Success Architect

June 16, 2020



Agenda

- What is?
  - Kettle
  - Apache Beam
  - Google Dataflow
- Demo

neo4j

What Is?

Kettle

Kettle is an open source data integration (ETL) platform with the ability to visually design your work. It has been around for over 17 years and in that time it became quite mature, stable, high performing and feature rich.

- Kettle is an open source data integration (ETL)
- Development via GUI
- Extensible
- Execution in multiple engines
- Full Neo4j integration (Kettle Neo4j Remix)

Apache Beam

An advanced unified programming model which implements batch and streaming data processing jobs that run on any execution engine

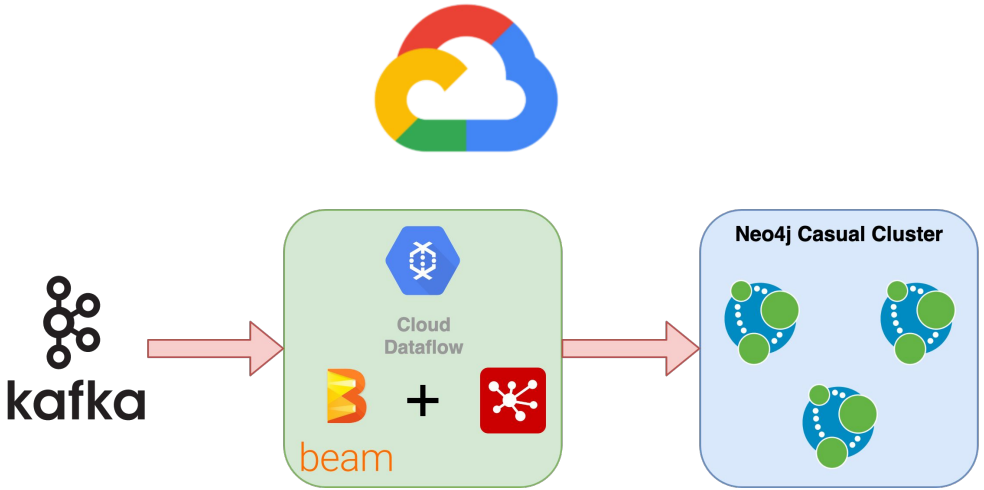
- Unified: Allows batch and stream processing
- Portable: Can be executed on multiple environments
- Extensible: Supports multiple SDKs, Runners, IO connectors and other libraries

Google Dataflow

Unified stream and batch data processing that's serverless, fast, and cost-effective

- Fully managed, no-ops execution environment
- Integration with Google Cloud Platform

Architecture



Demo

References

- Kettle Neo4j Remix
- The easy way to create Beam Pipelines (By Diethard Steiner)
- Apache Beam
- Google Dataflow


Thank You
neo4j