

Real-time Analytics with High Cardinality Dataset

Author: Ning Shi

Claps: 73

Date: Dec 19, 2018

How do you build a real-time analytics system that deduplicates over a million events a second with cardinality in the billion range, and also updates over a million counters a second? How do you scale such a system to keep up with the exponential growth of workload year over year? This is not a moonshot for us. We have solved this very problem and it is the type of tough engineering problem we solve in-house every day at Klaviyo.

We are starting a series called “[Counting](#)” to share our journey toward a scalable real-time analytics system. In this series, we will cover how we historically increased efficiency and scaled Cassandra counters. Following that, we will share how we scaled this complex system for exponential growth leveraging new stream processing technologies such as Kafka and Flink.