≋sqlstream

Automated advertising service goes from 180 nodes and 3h delays
to 5 servers streaming analytics on the entire load, continuously and in real time.

AT A GLANCE

Company
Rubicon Project

Industry
Advertising exchange

Country
USA

Impact
1 Billion people see ads processed in real time by the Rubicon Project Exchange, across 1 million websites and over 20,000 apps.

Website
http://rubiconproject.com/

BACKGROUND

The Rubicon Project Exchange brings buyers and sellers closer together on a robust advertising technology platform. One of the largest cloud and Big Data computing systems in the world, Rubicon leverages over 50,000 algorithms and analyzes billions of data points in real time to deliver the best results for sellers and buyers. Their operations are impressive: 9 trillion bid requests per month, 5 million peak queries per second, and over 300 real-time data-driven decisions per transaction. It processes more data signals than search engines, counts more bids than the physical world's auction houses combined, and makes complex business decisions in tiny fractions of a second.

And since the Cloud is continually learning, the volumes of data flowing through Rubicon's pipes make it stronger and smarter- but also expensive to maintain, and slow to react.

rubicon
PROJECT

> Streaming data processing means a future where the computational model of 'big data' doesn't require data to hit disk before processing it.
>
> *MATTHEW BREEDLOVE, VP, TECHNICAL & DATA OPERATIONS*

NEEDS

With over 65,000+ CPUs and 5.0 Petabytes of storage working at 100 gigabits/second, Rubicon needed an efficient solution that would permit adaptive scaling without the added cost.

The requirements for the solution were:

Reducing or stalling infrastructure costs
- Low-latency, real-time analytics on a unique architecture for all real-time operations applications
- Continuous, real-time integration of historic data, application data, and existing databases
- Continuous updates of performance metrics to inform the requirements for production infrastructure (the previous solution took 3 hours and 180 nodes to generate metrics on service quality).

SOLUTION

SQLstream Blaze enabled Rubicon to build the most efficient real-time platform for streaming ingestion, continuous ETL, and streaming analytics in the cloud. The solution delivered:

- Performance capabilities that can be scaled up and scaled out to allow processing of up to 440B records/day – 6.7M rows/second in the peak hour – on a relative handful of servers
- Data ingest throughput at continuous running rate scaling up to 1.8M records/second (155B records/day) on one server. This is only limited by network bandwidth – input rates between 6 and 7.5Gbps were measured
- Input rates up to 1.8M rows/sec (155B records/day – simple ingest, unpack required fields and count, 60 partitions/topic, 14 topics; using 81% of a 24-core cpu)
- Recoverable ingest rate – recording the Kafka offsets for each topic/partition each minute – up to 1.2M rows/second (95B records/day, reading from 14 topics, 60 partitions/topic – using 49% of a 24-core cpu)
- Running ingest at a rate that is sustainable used 40% of a 24-core cpu for Kafka ingest alone.
- Continuous operation over a 24 hour period at continuous running rate of over 650K records/second – using 85% of a 24-core cpu.

RESULTS

From 180 nodes taking 3 hours to generate metrics on service quality and over 65,000 CPUs working on their real-time auction processes, Rubicon went down to using only 5 servers to run streaming analytics on their entire load (110B records/day)—now getting insights continuously and in real time.

sqlstream

1540 Market Street
San Francisco, CA 94102
www.sqlstream.com