Splout SQL

A web-latency SQL spout for Hadoop

DATASALT

The problem it solves

- When the output of a Hadoop process is big, how do you serve it live?

  Streaming it directly to a database is usually inefficient and error-prone. Instead, Splout SQL creates B-Trees in Hadoop and moves generated files atomically to a serving cluster.

- What database can you use to query data generated in Hadoop in a flexible way?

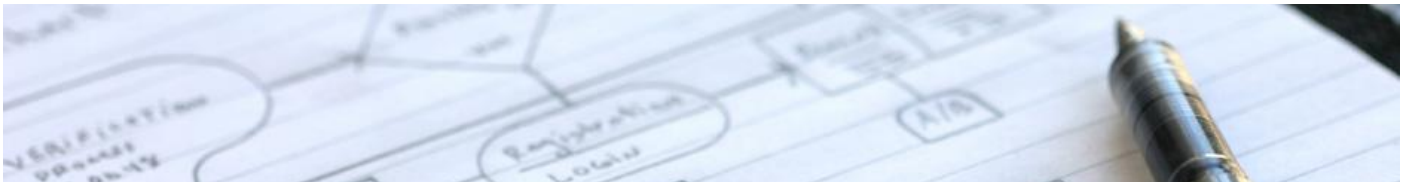
  There are a lot of key/values, but they force you to pre-calculate everything. There are other stores, but they lack fast SQL aggregation primitives. Splout SQL is both Hadoop-friendly and SQL.

Meet Splout SQL

- As **fast** as key/values such as Voldemort or ElephantDB, but **with SQL**.
- Query Hadoop-generated datasets **for the web** (high throughput, very low latency).

  Unlike Impala or Drill, which are for offline analytics.
- Deploy Hadoop-generated datasets **without pain**

  With command-line tools, no programming needed.
- Horizontal **scaling**, replication for **fail-over** & **Open Source**



Typical usages

1. *A website that calculates recommendations in Hadoop and wants to serve them live.*
2. *A website that calculates daily metrics from user's activity and wants to show timelines back to the users (Google-analytics-like apps).*
3. *An internal webapp for a big company, showing per-client / per-provider metrics that are pre-aggregated in Hadoop.*
4. *A banking / retailer mobile application to check your activity between arbitrary time periods.*
5. *An advertisement-tracking website that offers statistics for both advertisers and clients.*

At Datasalt we are looking forward to hear about your use case to help you integrating Splout SQL: info@datasalt.com