Getting Started with Interactive SQL for Hadoop

John Russell

Alan Choi

Alex Behm

Why are you here?

- You have SQL skills, Hadoop interest.

- Data volume increasing rapidly.

- Can't do analytics fast enough with batch-oriented approaches.

- Your assignment: to unlock business value from Big Data.

Strata+Hadoop WORLD

What will you learn?

- SQL in Hadoop context.
  - HDFS storage layer.
  - File formats and interoperability.
  - Distributed queries.
- Big Data use cases and challenges for SQL.
  - Think bigger.
  - Unlearn OLTP habits.

Outline of the Tutorial

- Developer theme, "getting started" scope.

- Lessons using the Impala interactive SQL engine for Hadoop.

- Question and answer format:

  - With code samples.

  - Discuss Hadoop aspects.

  - Dive down a level to explore performance and scalability.

  - Follow along and experiment in Cloudera QuickStart VM.

Complementary to O'Reilly Book

- Impala in the Big Data ecosystem.

- High points of Impala's SQL.

- Developer-themed tasks.

- Billion-row tables.

- Query performance for joins.

- Date and time manipulation.

- Walkthroughs with real code.

- Discount code for tutorial attendees.