Insights Without Tradeoffs Using Structured Streaming

Michael Armbrust - @michaelarmbrust

Spark Summit East 2017

databricks

Parallelism

Split up the problem
to harness many
machines for computation

Complexity

Handle cross-machine
communication and failures

Developer Productivity

Quickly and concisely express common computations

Efficiency

Hand-tune code to minimize overheads and process the most data per cycle

Throughput

Process large historical repositories quickly

Latency

Up-to-date answers as new data arrives

Structured Streaming

Production Use Cases

Streaming at databricks

Collect logs and metrics from a variety of sources to ensure the security, availability and performance of our cloud platform.

Engineer Office Hours @ Databricks Booth

MY HOURS

TODAY 4:30 – 5:15

- Spark SQL
- Structured Streaming
- Databricks

OTHER ENGINEERS

TODAY 1:45 – 5:15

- Spark Core
- R
- Data Science
- ML
- GraphFrames, Deep Learning
- Databricks
- Spark SQL
- Structured Streaming

THURS 10:30 – 2:30

Thank you!

All code available at databricks.com/blog
and @michaelarmbrust