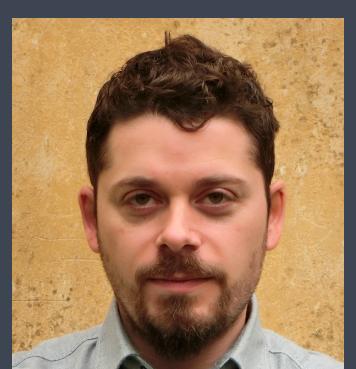




Malte
Sandstede



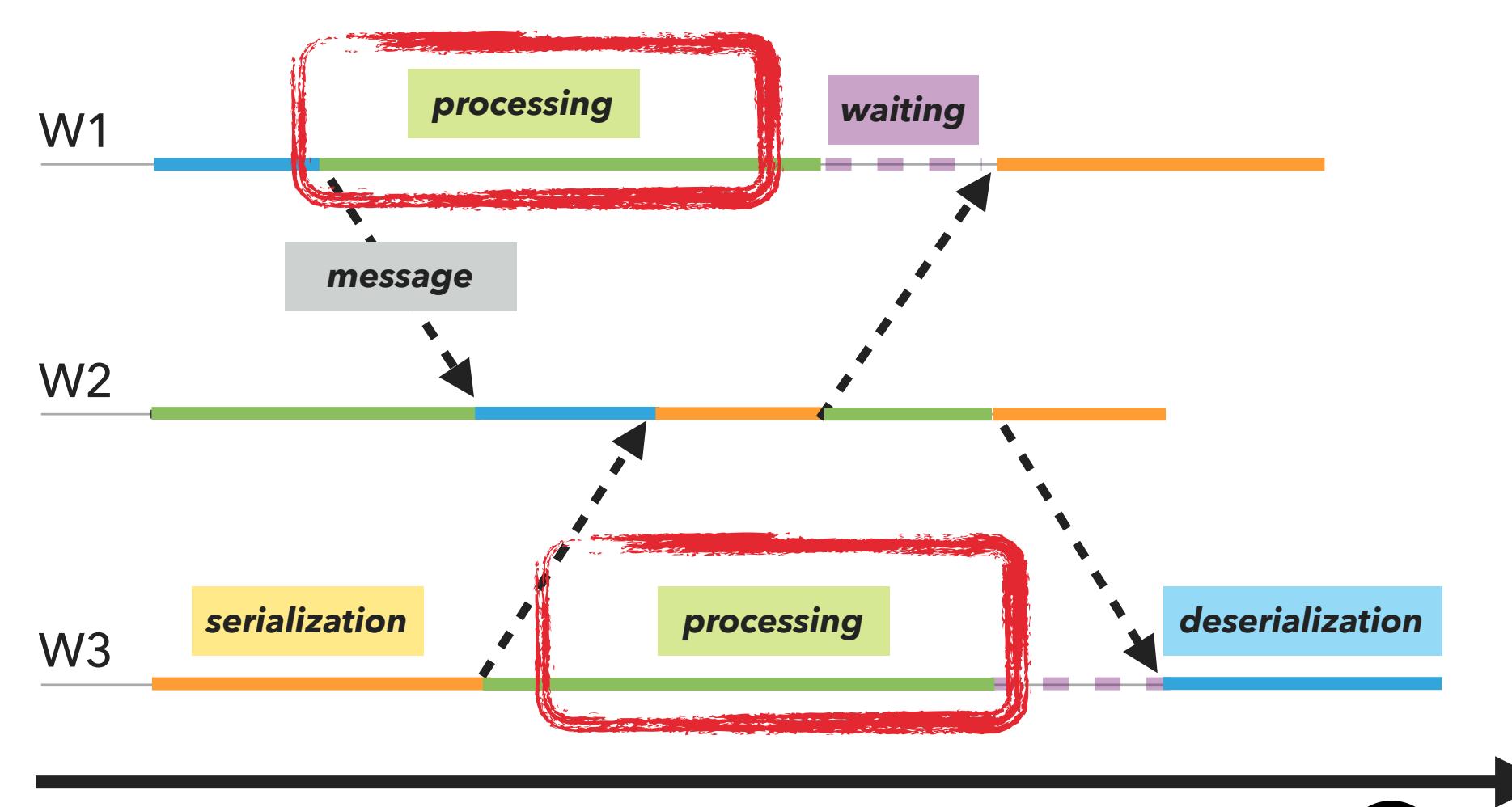
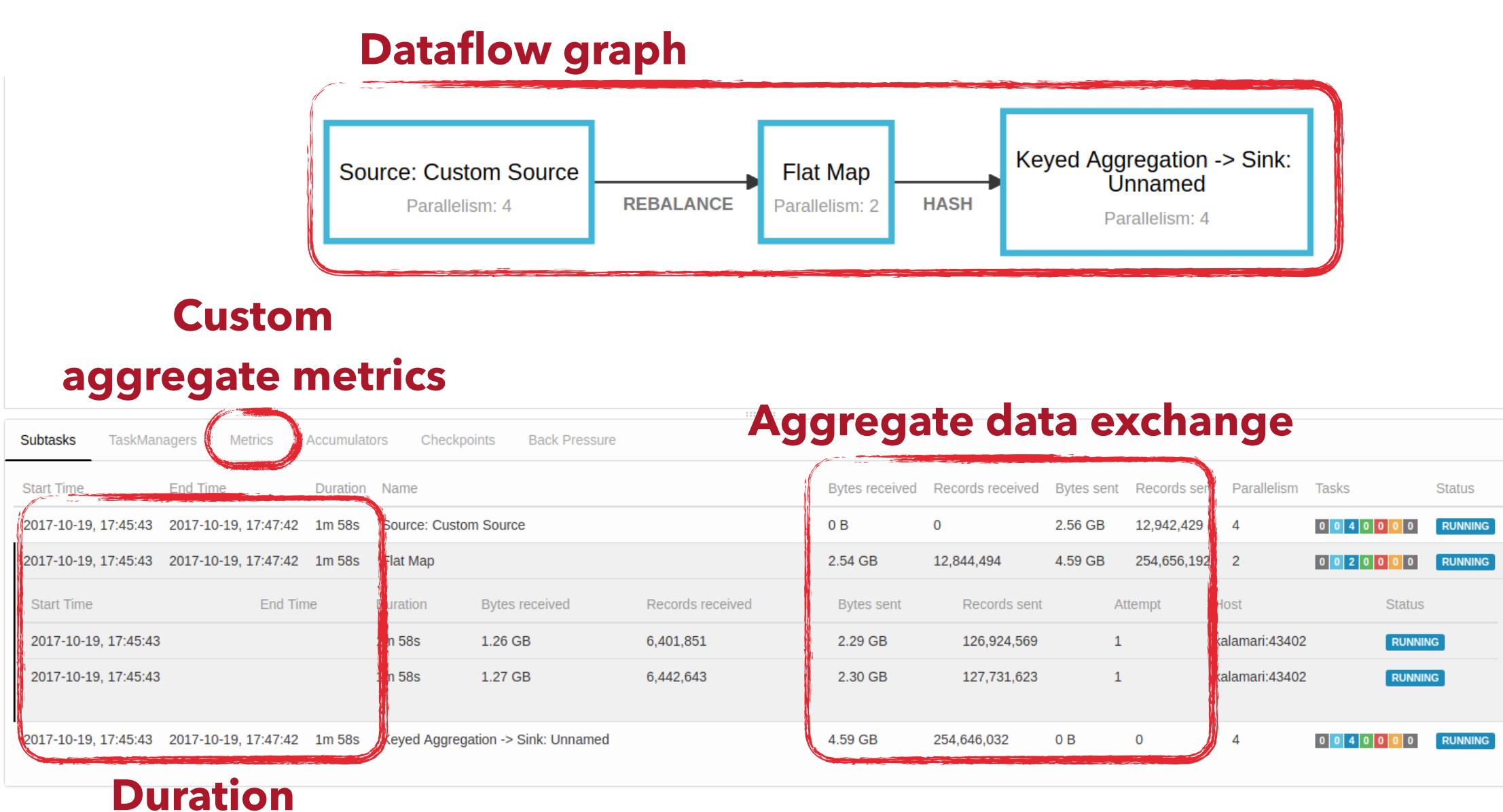
John
Liagouris

ST2: Analyzing Performance Issues in Distributed Dataflows



BOSTON
UNIVERSITY

CONVENTIONAL PROFILING IGNORES DEPENDENCIES

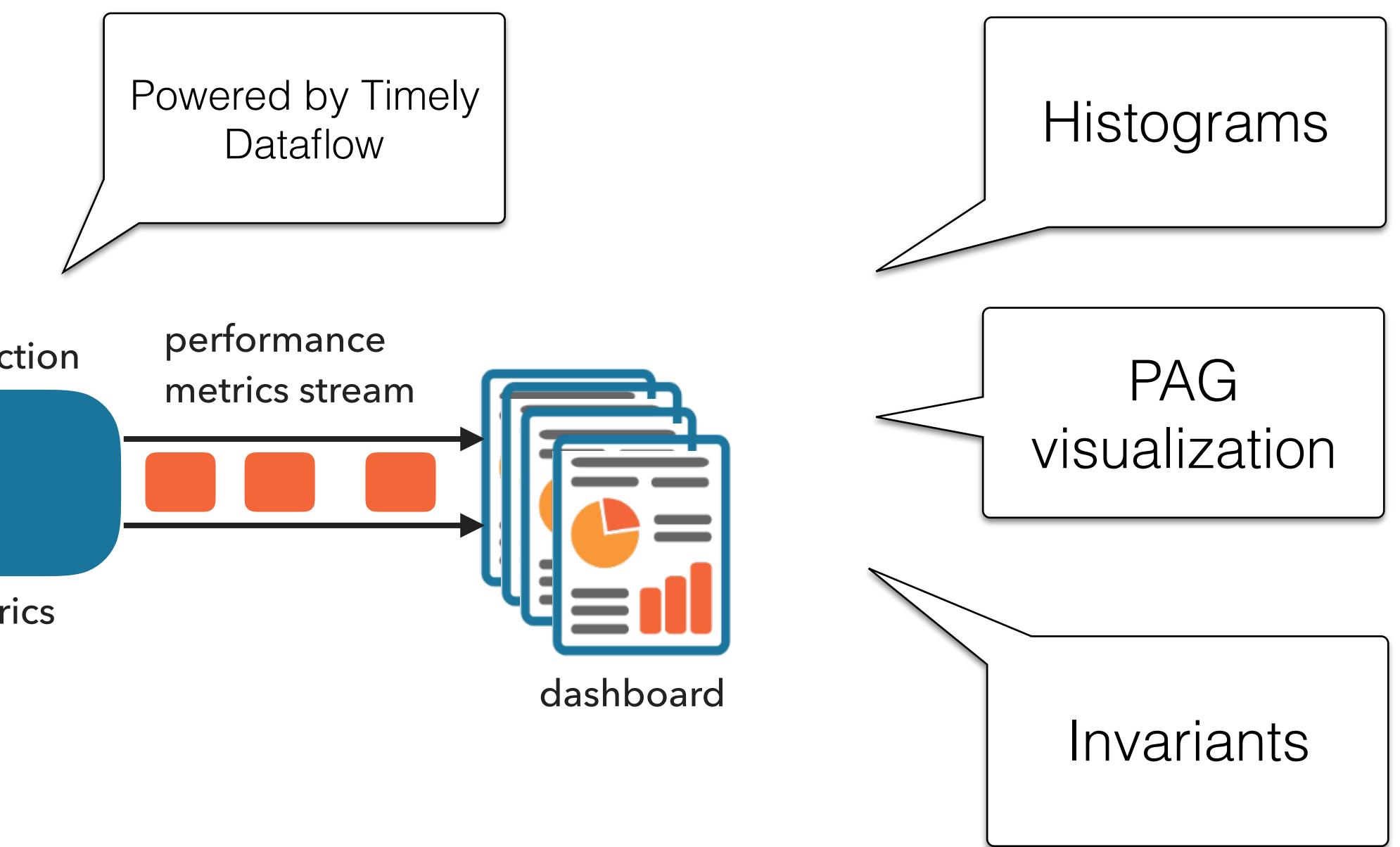
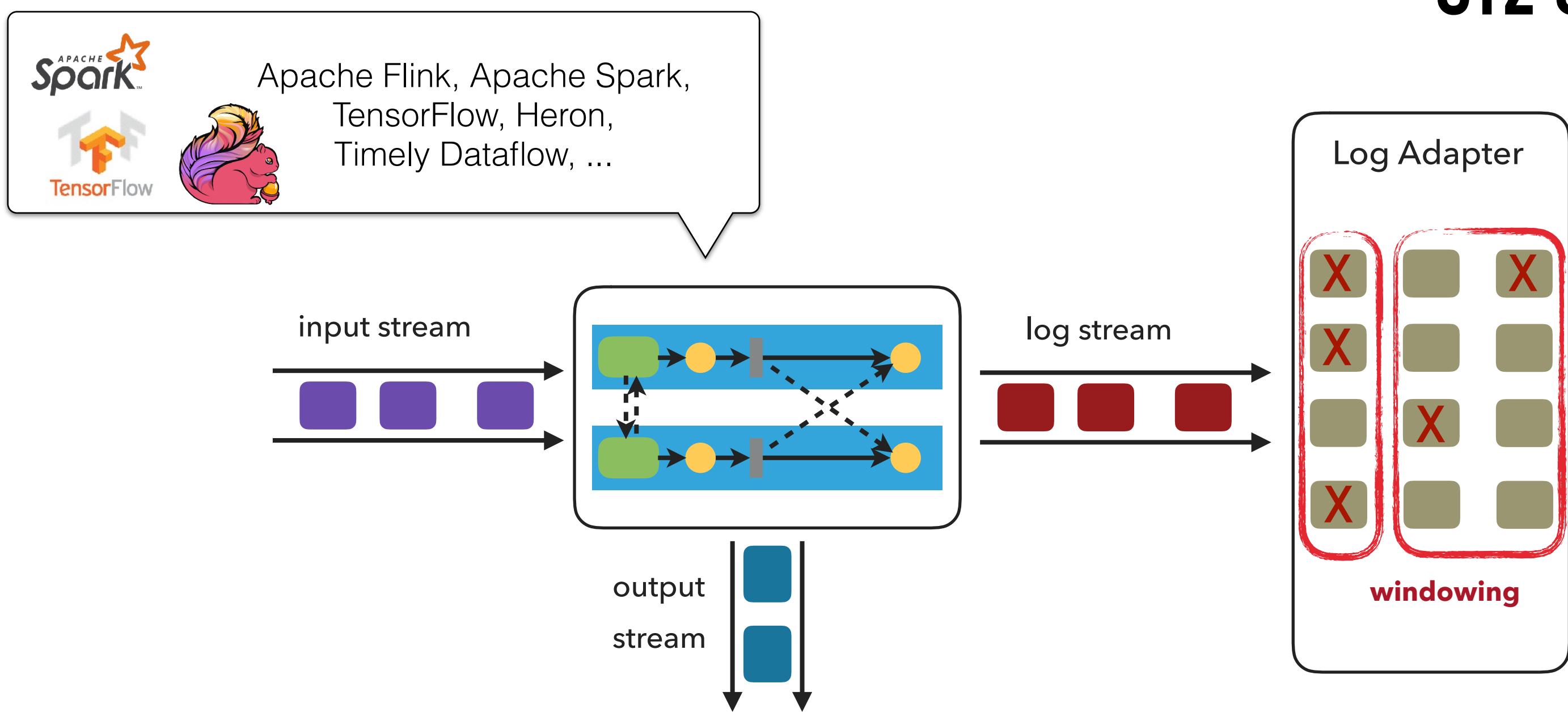


Existing tools:

- ▶ report simple aggregated metrics
- ▶ operate offline
- ▶ ignore execution dependencies

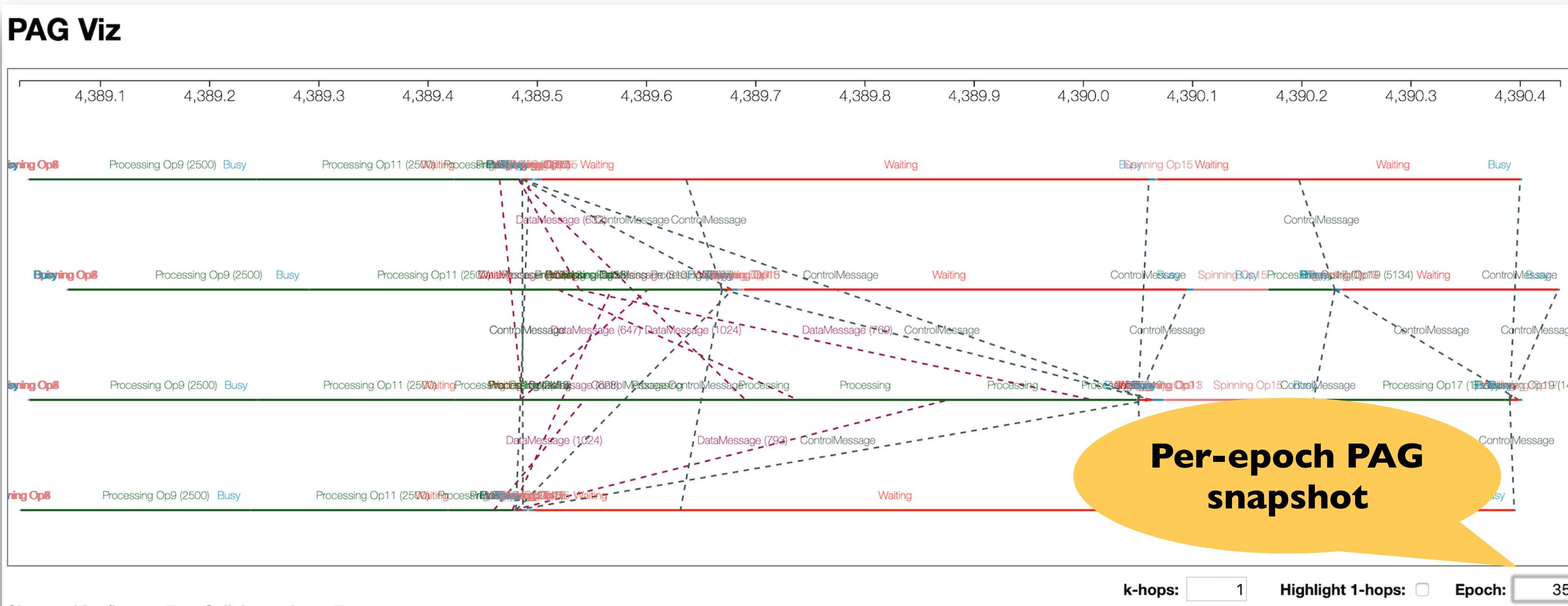
ST2 analyzes distributed dataflow performance at scale, takes dependencies into account, and provides feedback to running applications in real time.

ST2 OVERVIEW

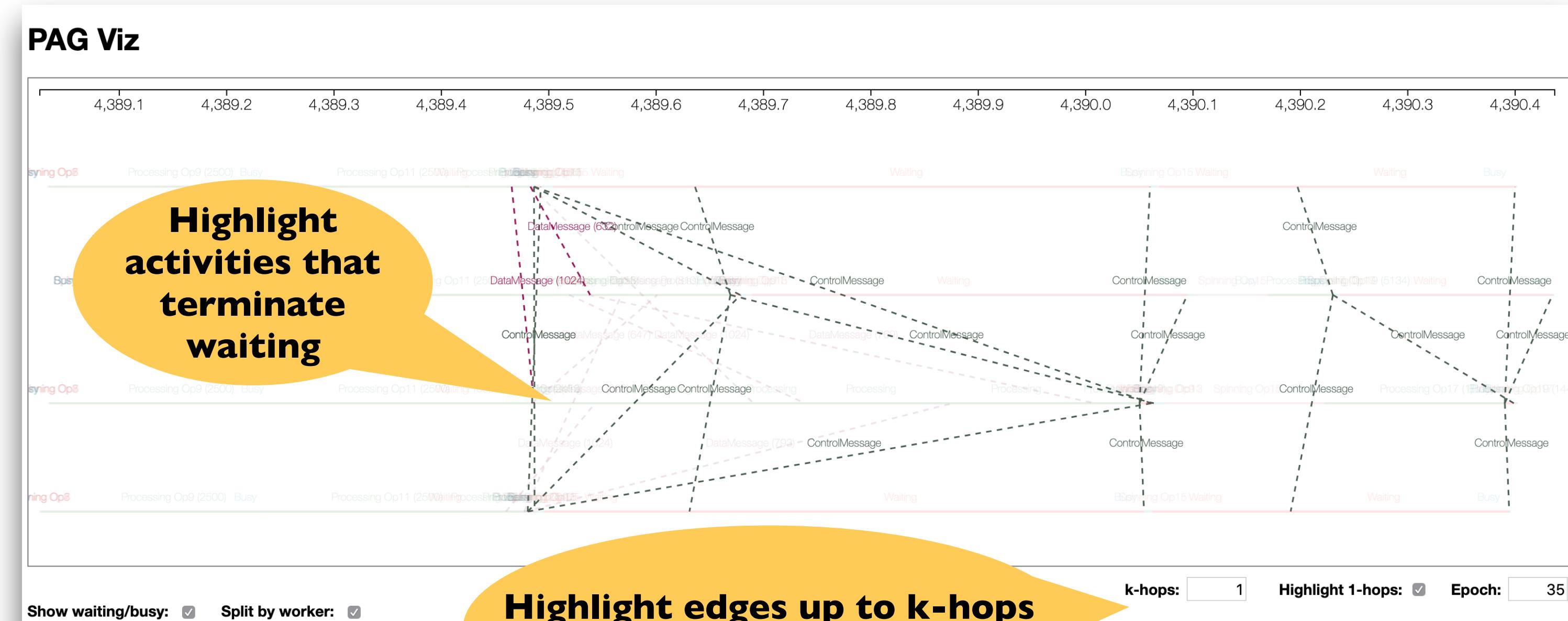


ST2 IN ACTION: PROFILING TPC-H QUERY 1 ON DIFFERENTIAL DATAFLOW

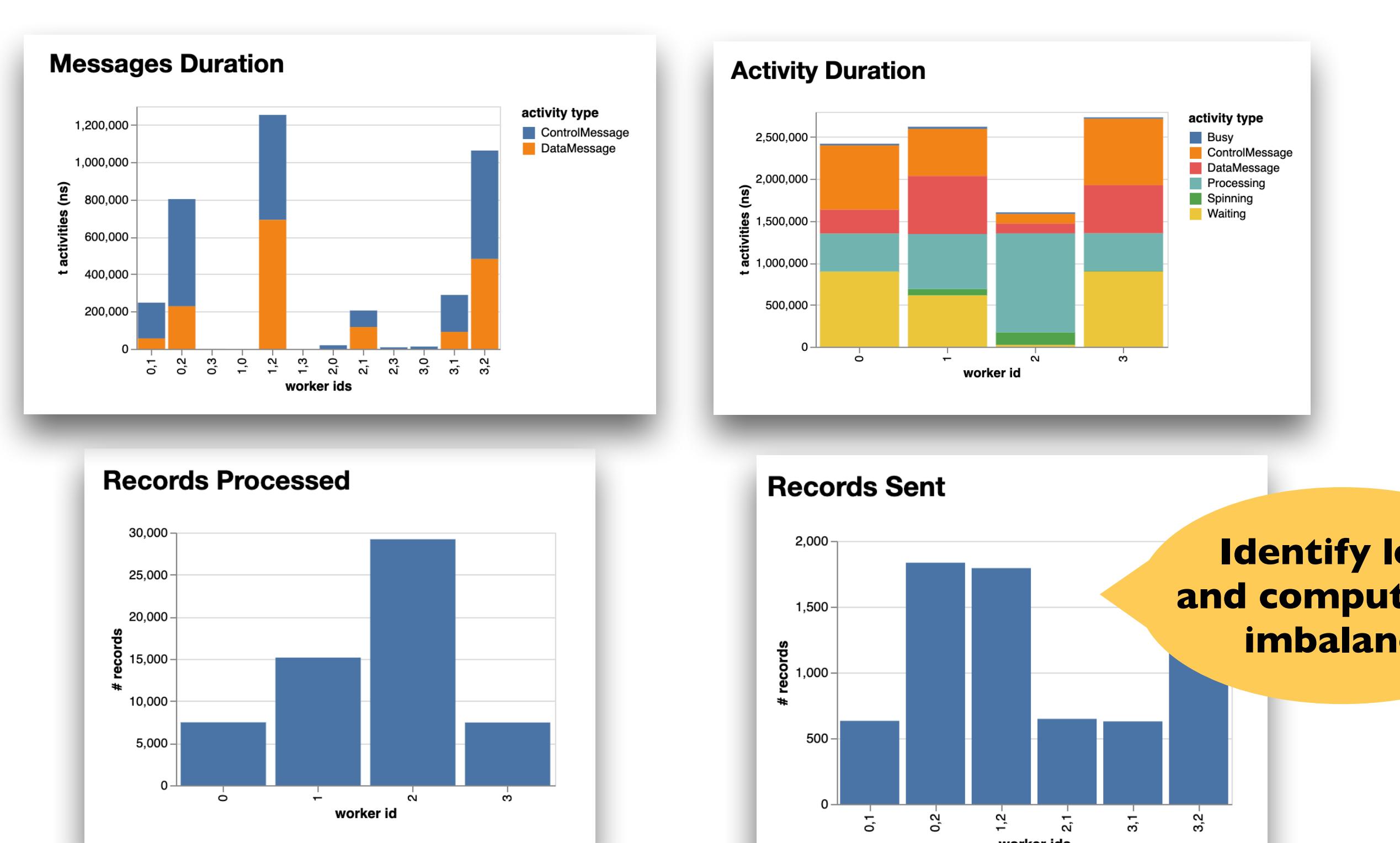
Program Activity Graph



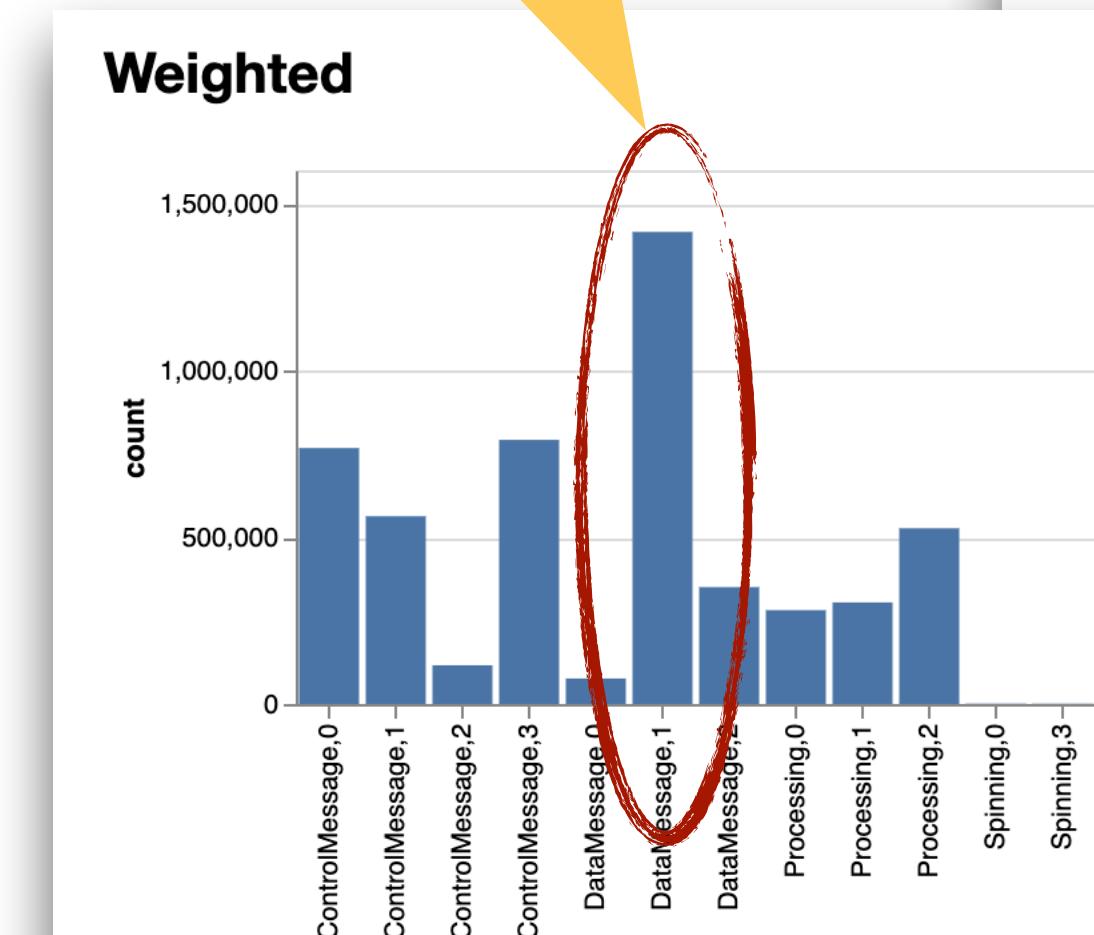
Path Highlighting



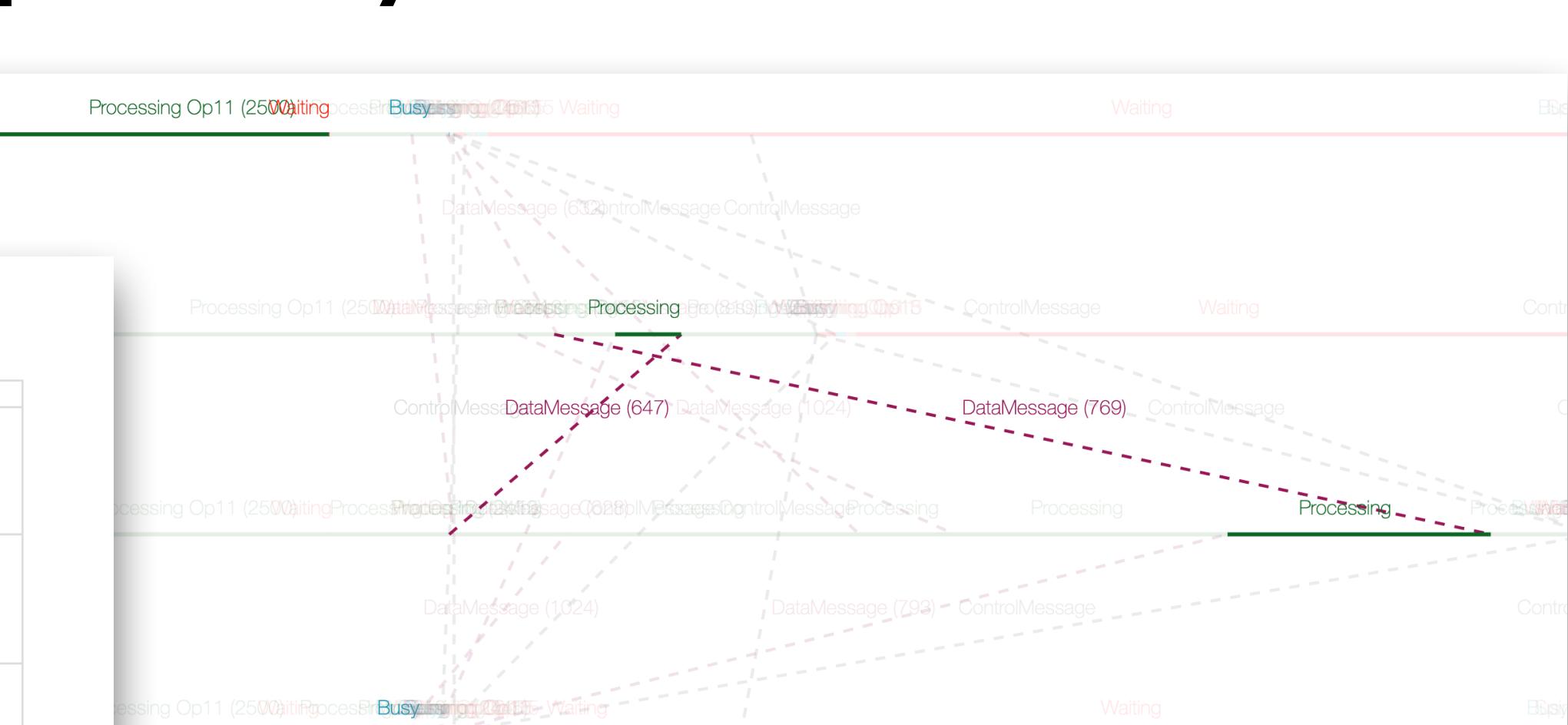
Aggregate Metrics



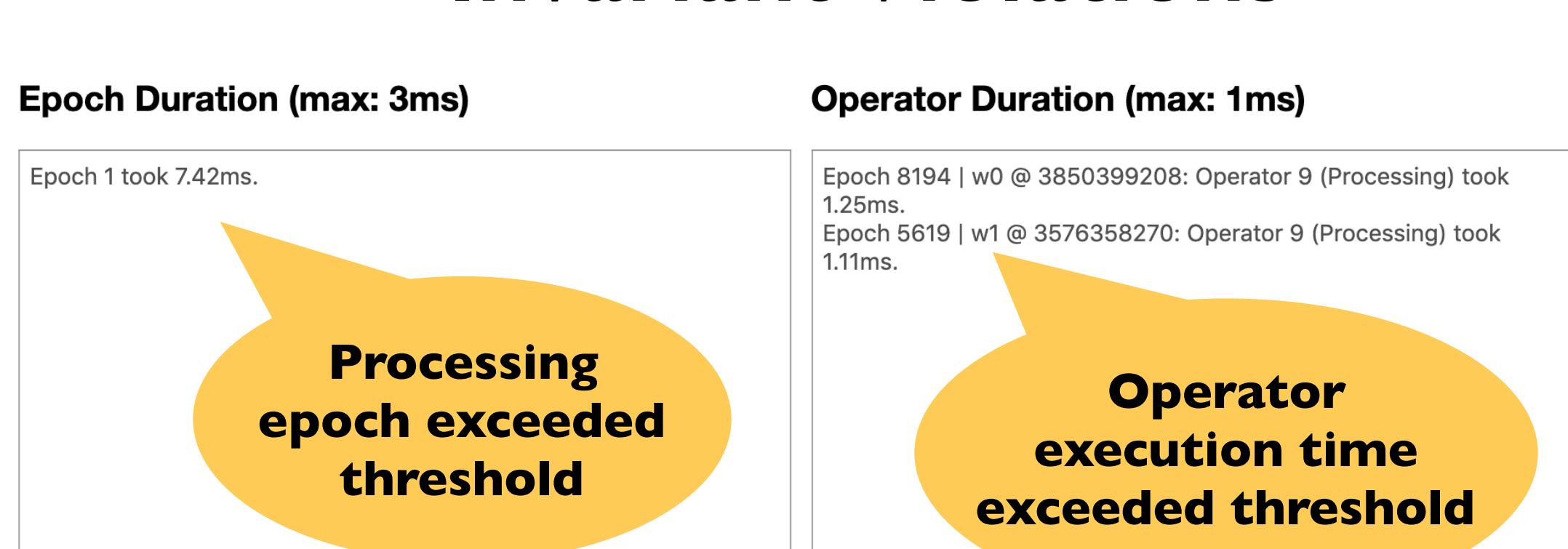
Identify causes of imbalance and bottlenecks



Dependency-based Metrics



Invariant Violations



Message duration exceeded threshold

ST2 Performance

