# FLINK

Another data stream framework!

# What is Flink?

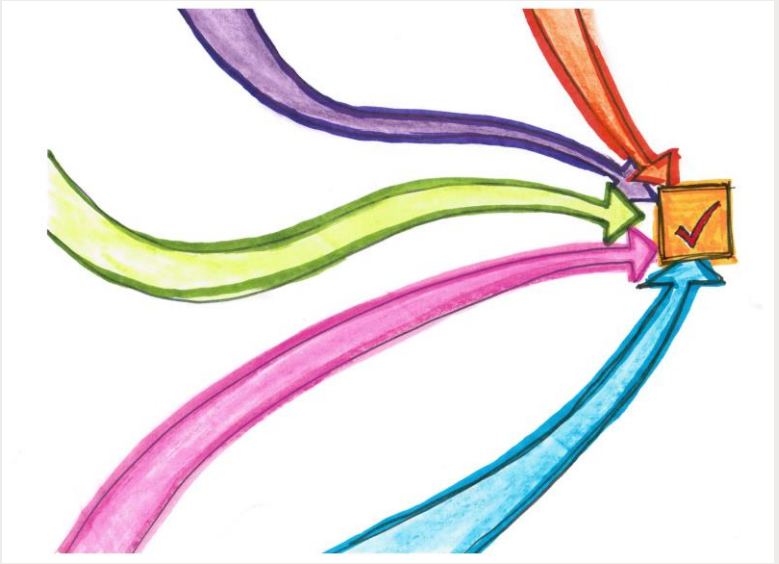

- German for quick and nimble
- Another stream processing engine – most similar to Storm
- Can run on standalone cluster, or on top of YARN or Mesos
- Highly scalable (1000's of nodes)
- Fault-tolerant
  - *Can survive failures while still guaranteeing exactly-once processing*
  - *Uses "state snapshots" to achieve this*
- Up & coming quickly

# Flink vs. Spark Streaming vs. Storm

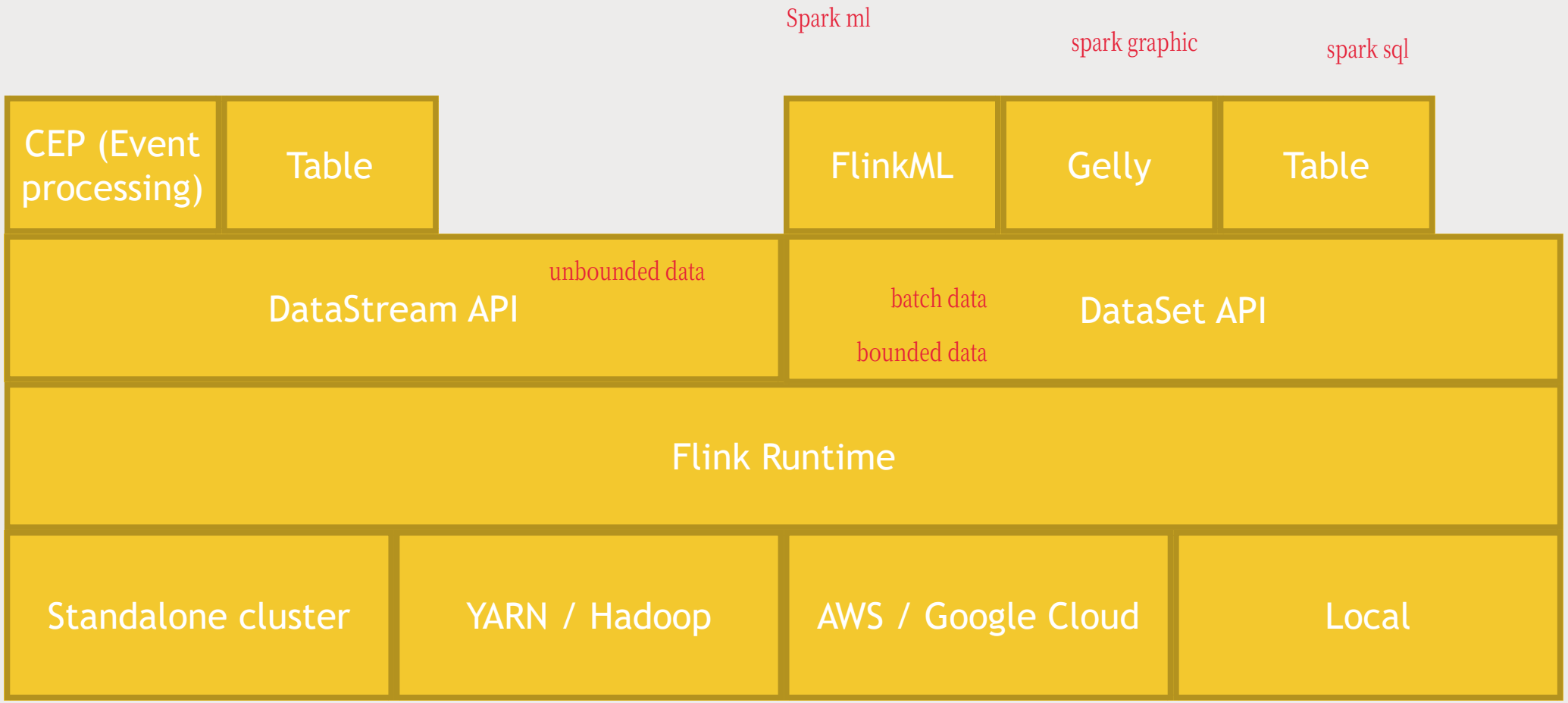
- Flink's faster than Storm
- Flink offers "real streaming" like Storm (but if you're using Trident with Storm, you're using micro-batches)
- Flink offers a higher-level API like Trident or Spark, but while still doing real-time streaming
- Flink has good Scala support, like Spark Streaming
- Flink has an ecosystem of its own, like Spark
- Flink can process data based on event times, not when data was received
  - *Impressive windowing system*
  - *This plus real-time streaming and exactly-once semantics is important for financial applications*
- But it's the youngest of the technologies

# All three are converging it seems

- Spark Streaming's "Structured Streaming" paves the way for real event-based streaming in Spark
- Becomes more a question of what fits best in your existing environment



# Flink architecture



# Connectors

- HDFS
- Cassandra
- Kafka
- Others
  - *Elasticsearch, NiFi, Redis, RabbitMQ*

# Let’s Fiddle with Flink