

Blink:阿里新一代实时计算引擎

马国维

2017.4



促进软件开发领域知识与创新的传播



关注InfoQ官方信息
及时获取QCon软件开发者
大会演讲视频信息



扫码，获取限时优惠

ArchSummit
全球架构师峰会 2017 [深圳站]

2017年7月7-8日 深圳·华侨城洲际酒店

咨询热线：010-89880682

QCon

全球软件开发大会 [上海站]

2017年10月19-21日

咨询热线：010-64738142

Who am I?

✓ 2010 – 2017 Alibaba Search

- iStream
- Blink

✓ 2007 – 2010 Baidu Web Search

Outline

1 The Streaming Architecture

2 What is Flink

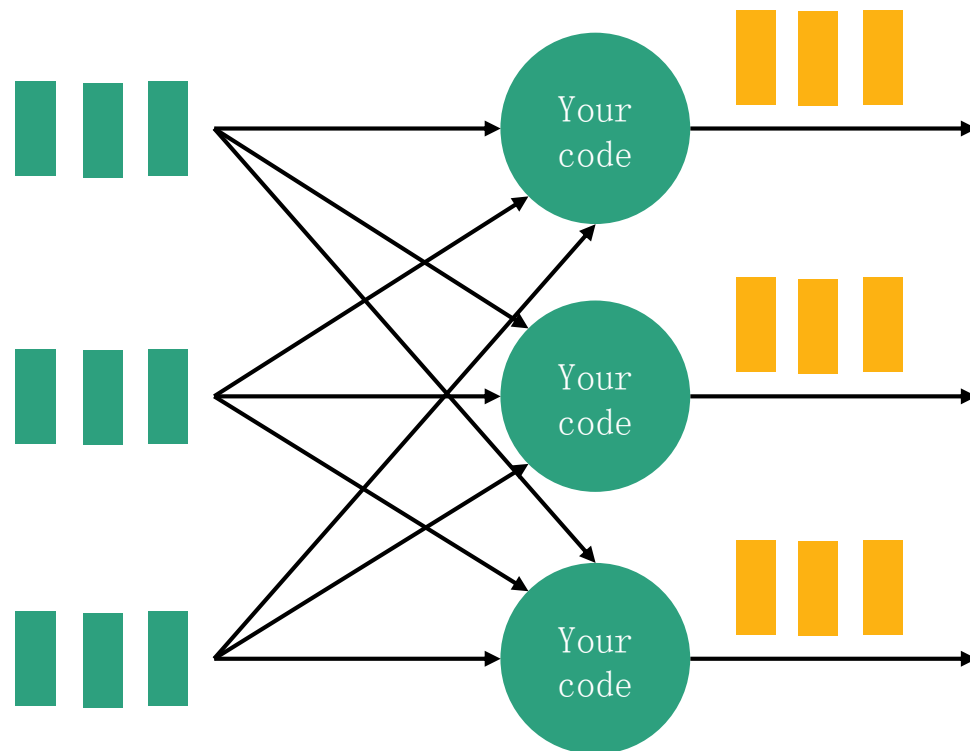
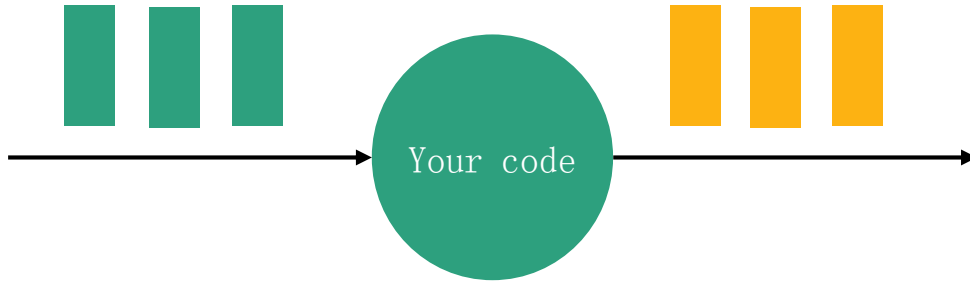
3 What is Blink

4 Future Plans

The streaming architecture

Part I

What is streaming?



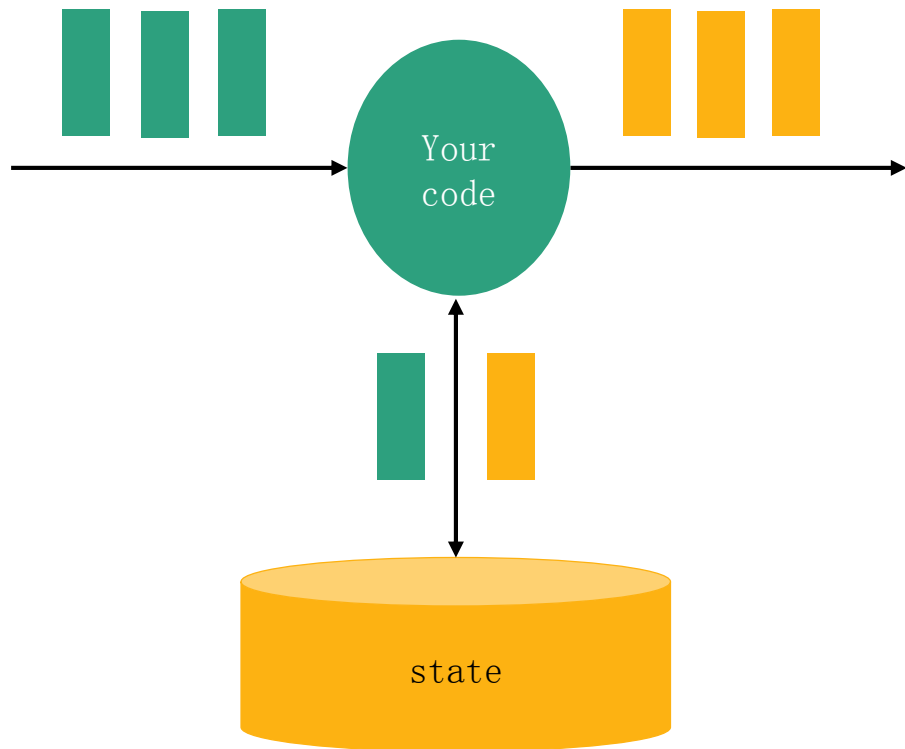
✓ What is streaming?

- Unbounded data

✓ What is streaming process engine?

- The data process engine that is designed with infinite data set in mind

What is stateful streaming



✓ Computation *and* state

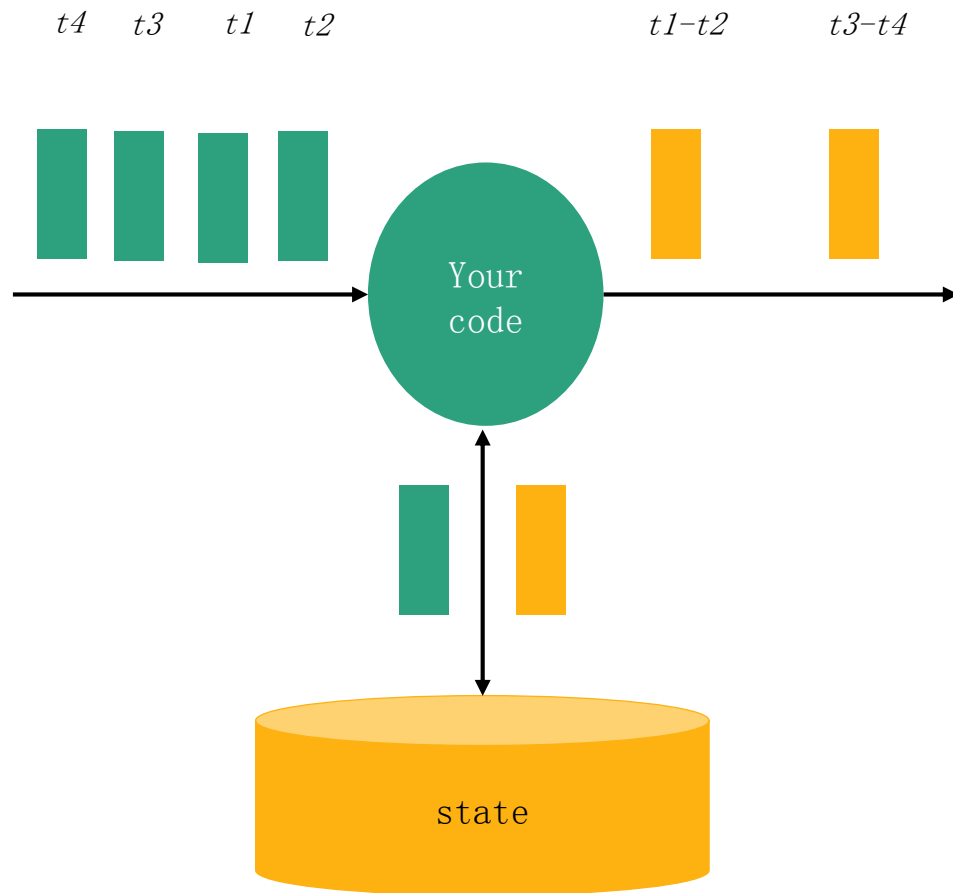
- E.g., counters, windows of past events, state machines, trained ML models

✓ Result depends on history of stream

✓ Stateful stream processor gives the tools to manage state

- Recover, roll back, version, upgrade, etc

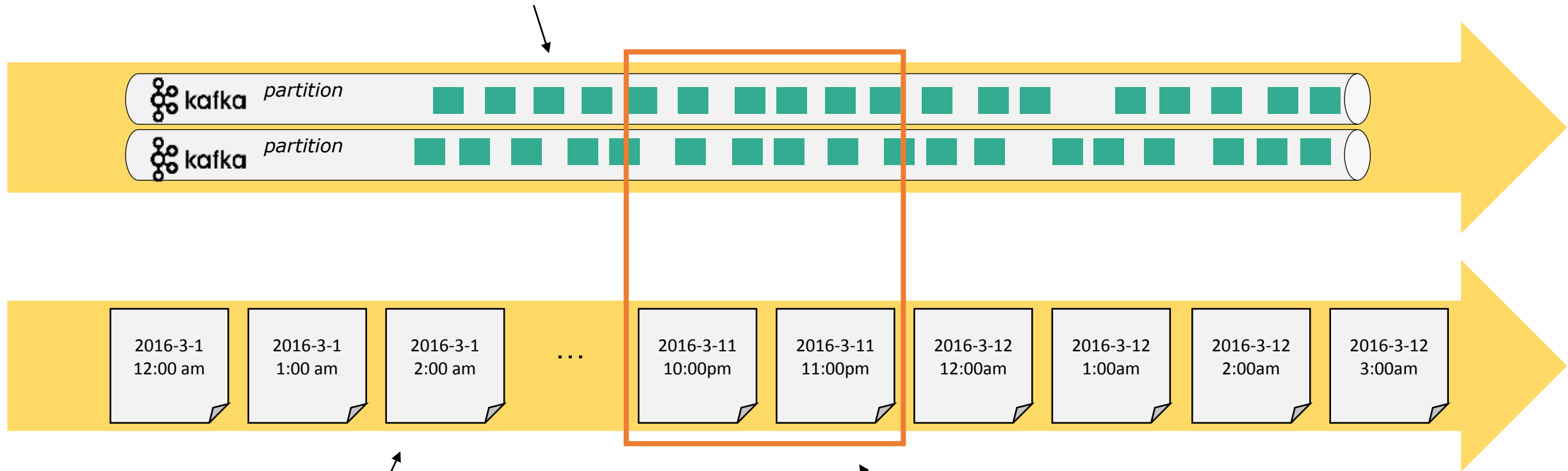
What is event-time streaming



- ✓ Data records associated with timestamps (time series data)
- ✓ Processing depends on timestamps
- ✓ Event-time stream processor gives you the tools to reason about time
 - E.g., handle streams that are out of order
 - Core feature is watermarks – a clock to measure event time

Streaming Subsumes Batch

Stream (low latency)



Stream (high latency)

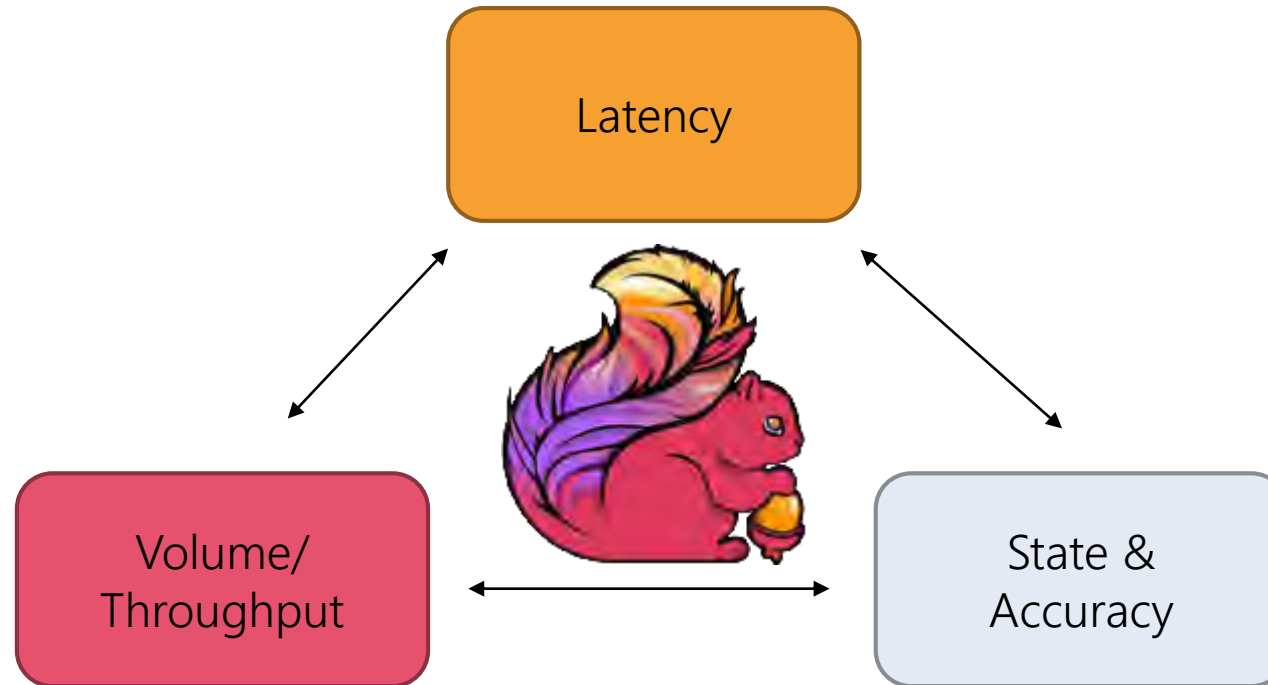
Batch
(bounded stream)

What is Flink?

Part 2

Flink - Streaming Compute Engine

Latency down to the milliseconds

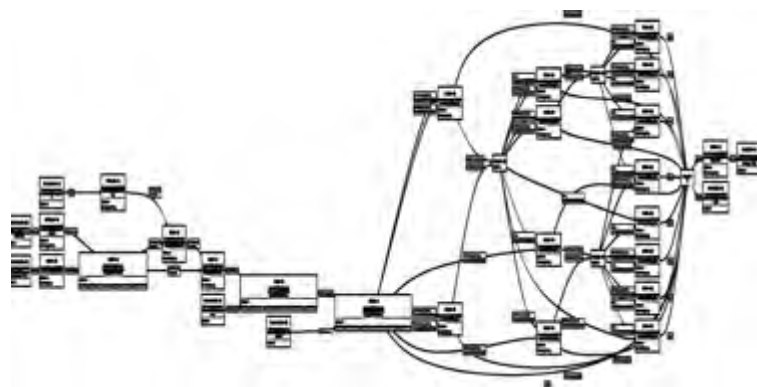


10s of millions evts/sec
for stateful applications

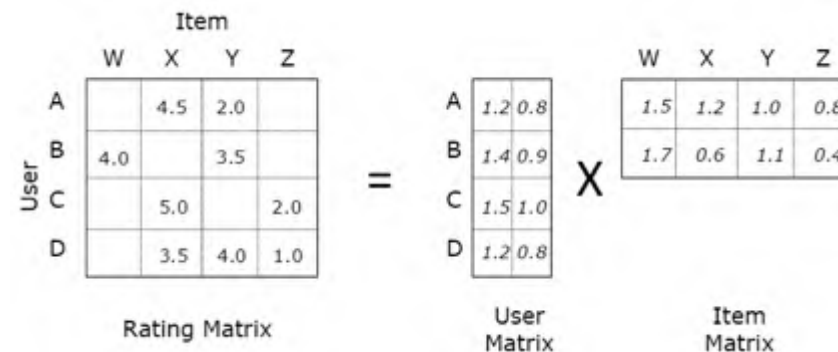
Exactly-once semantics
Event time processing

Flink – Unified Compute Engine

Long batch pipelines



Machine Learning at scale



Streaming topologies

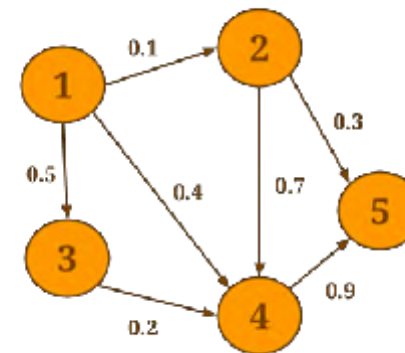
→ resource utilization

→ iterative algorithms



→ Low latency

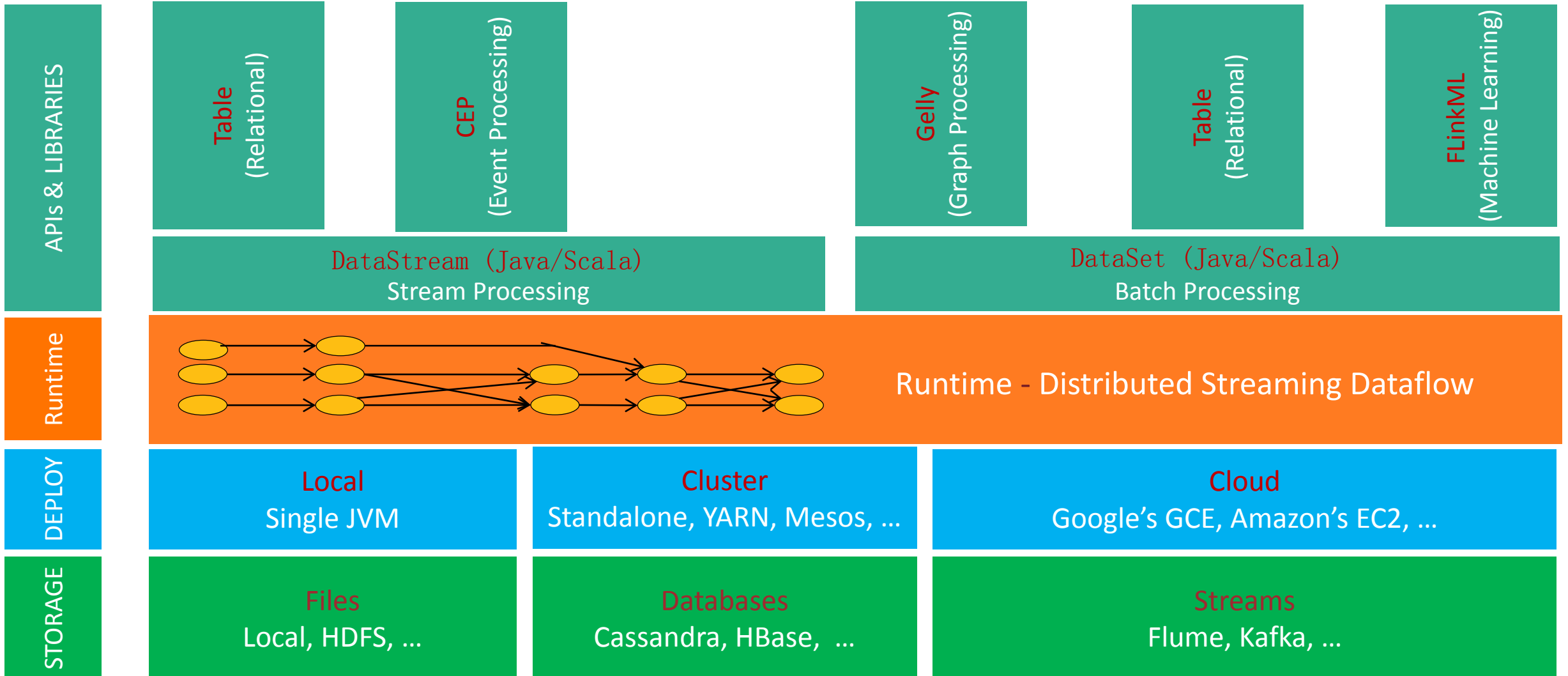
Graph Analysis



→ Mutable state

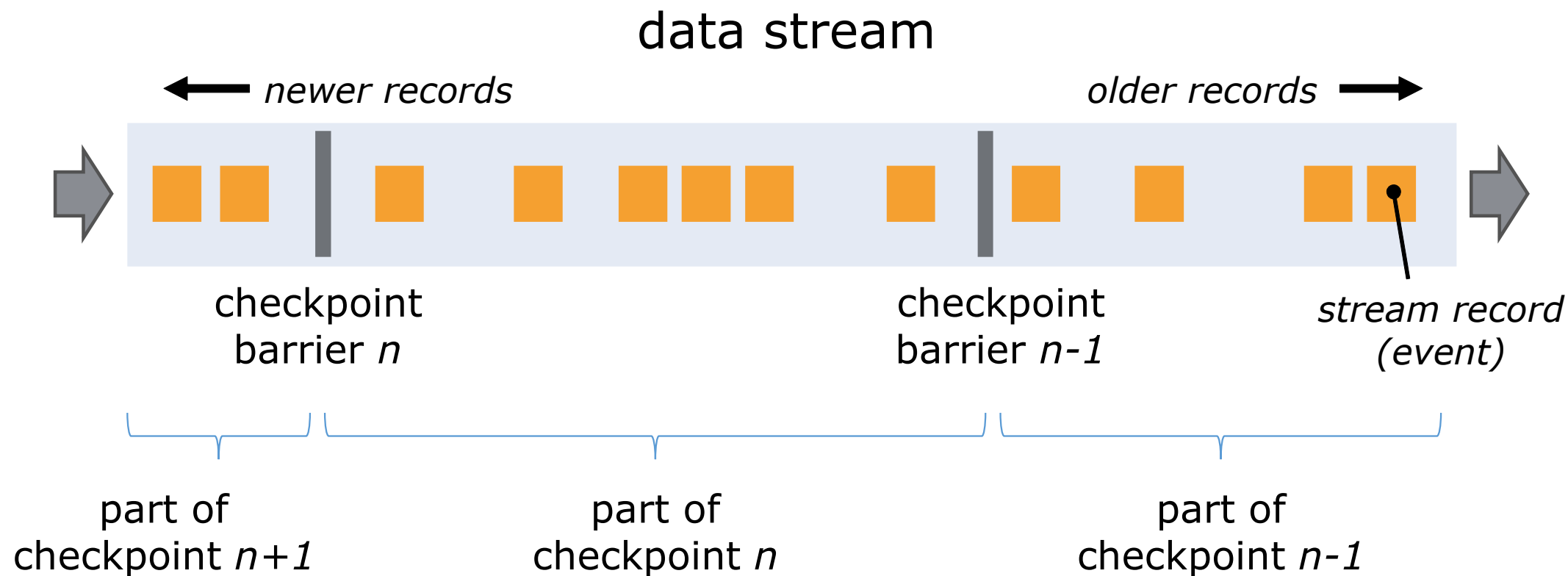
Flink

Flink Ecosystem

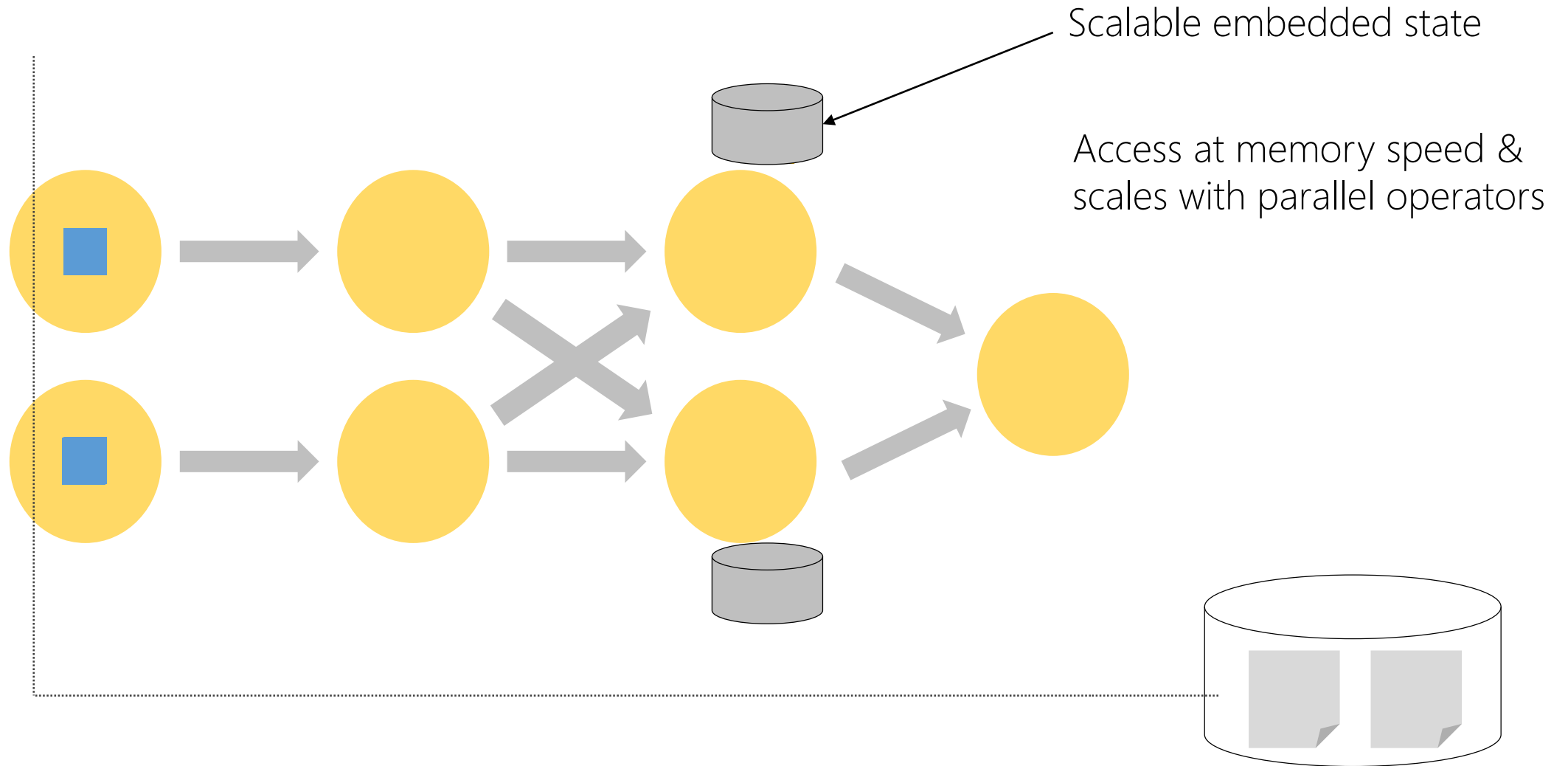


Checkpoint/Recovery

- Chandy-Lamport algorithm
- Periodic asynchronous consistent snapshots of application state
- Provide exactly-once *state* guarantees under failures



Stateful Steam Processing



What is Blink?

Part 3

Blink – Alibaba's version of Flink

✓ Looked into Flink two years ago

- best choice of unified computing engine
- a few of issues in flink that can be problems for large scale applications

✓ Started Blink project

- aimed to make Flink work reliably and efficiently at the very large scale at Alibaba

✓ Made various improvements in Flink runtime

- native run on yarn cluster
- failover optimizations for fast recovery
- incremental checkpoint for super large state
- async operator for high throughputs

✓ Working with Flink community to contribute back since last August

- several key designs
- hundreds of patches

Blink in Alibaba Production

- ✓ In production for almost one year
- ✓ More than 3000 nodes are running Blink
- ✓ The largest Blink cluster is more than 1000 nodes
- ✓ There are hundreds of production jobs supported by Blink
- ✓ Supported key online Service on last Nov 11th
 - The largest Blink job has 5000 concurrent, 10TB state, billions of QPS
 - Based on the Blink machine learning platform to significantly increase the transaction conversion

Blink Ecosystem in Alibaba

Alibaba Apps

Search

Recommendation

Ads

Ant

BI

Security

Blink

Machine Learning Platform (Porsche)

Table API

SQL

DataStream API

DataSet API

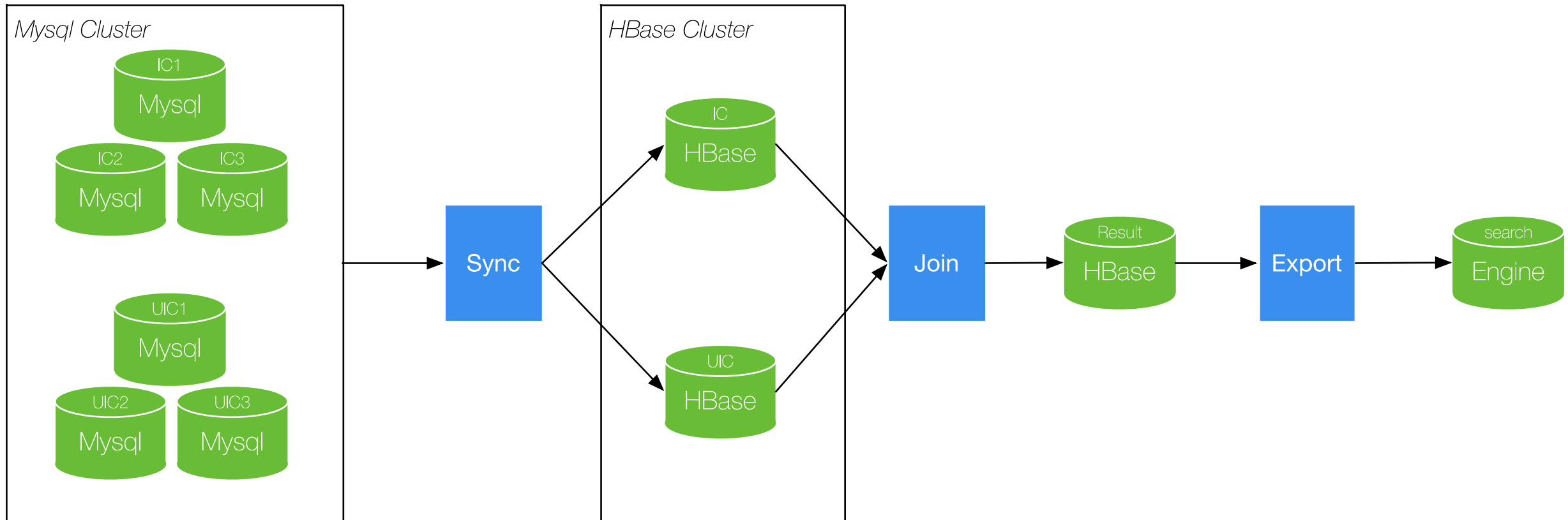
Runtime Engine

Hadoop

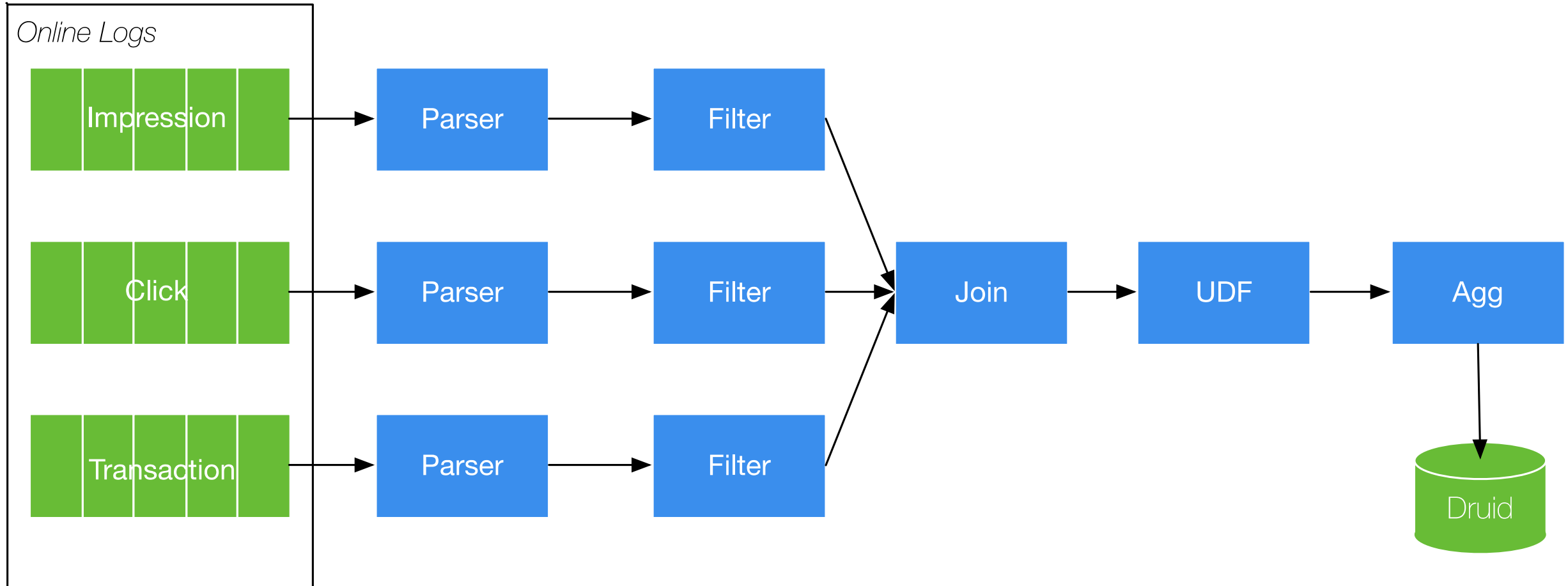
YARN (Resource Management)

HDFS (Persistent Storage)

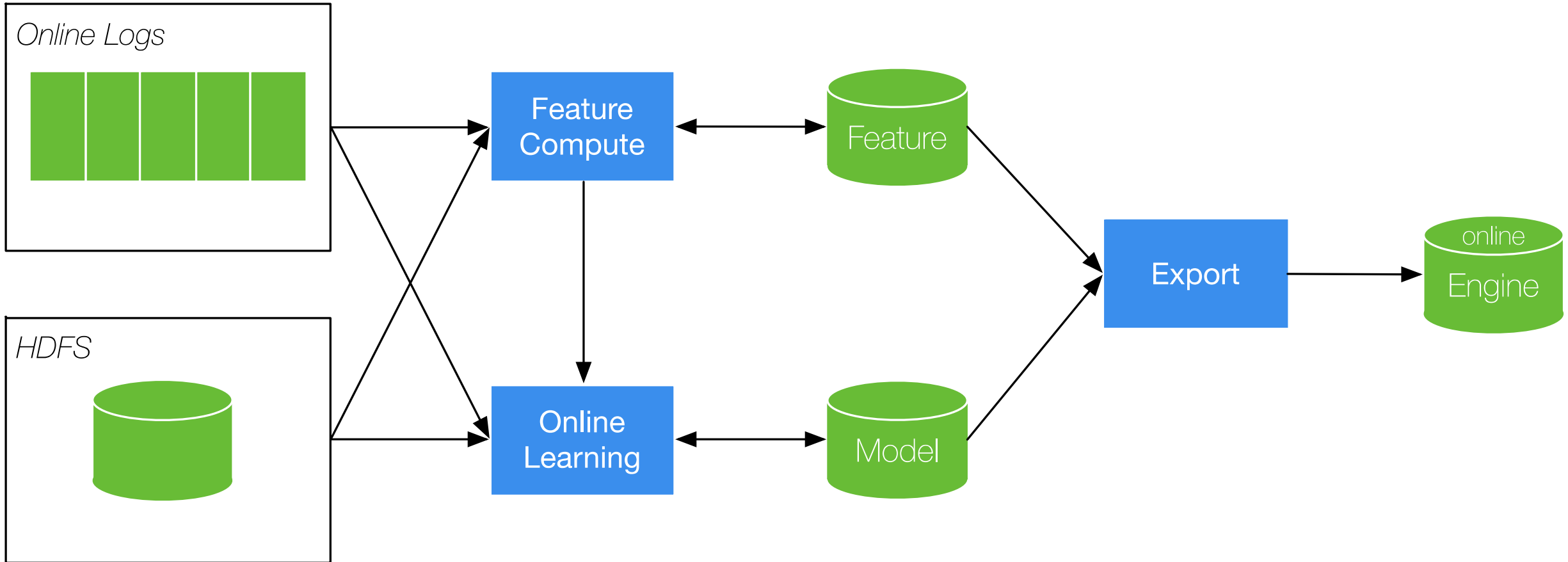
Use Case — Search Index Build & Update



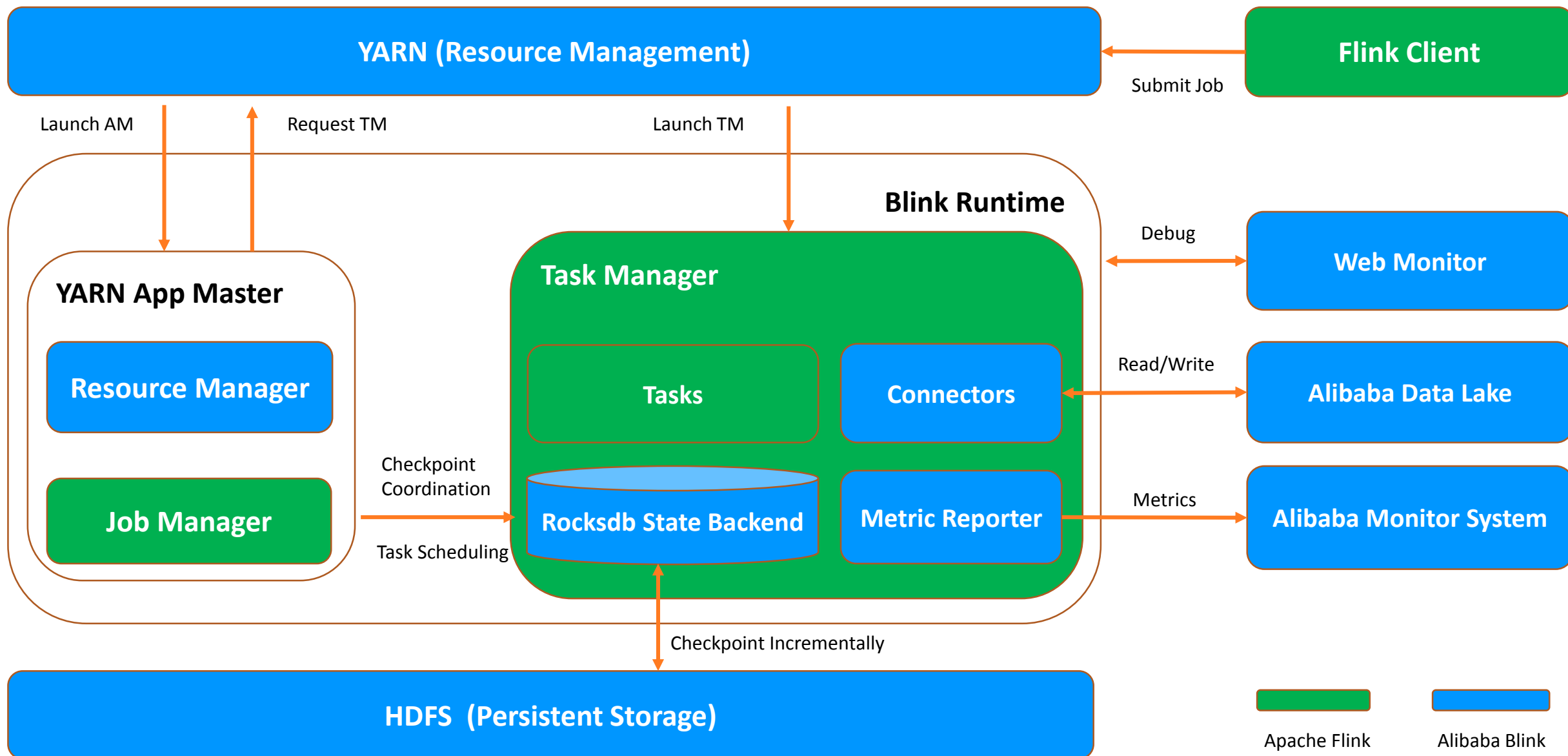
Use Case — Realtime A/B Test



Use Case — Online Machine Learning



Blink Architecture



Improvements to Flink Runtime

- ✓ Native integration with Resource Management for dynamic resource allocation and more larger scale
- ✓ Performance Improvements
 - Incremental Checkpoint
 - Asynchronous Operator
- ✓ Failover Optimization
 - Fine-grained Recovery for Task Failures
 - Allocation Reuse for Task Recovery
 - Non-disruptive JobManager Failures via Reconciliation

Future Plans

Section 4

Future Plans

- ✓ Blink is already popular in the streaming scenarios
 - more and more streaming applications will run on blink
- ✓ Make batch applications run on production
 - increase the resource utilization of the clusters
- ✓ Blink as Service
 - Alibaba Group Wide
- ✓ Cluster is growing very fast
 - cluster size will double
 - thousands of jobs run on production

THANKS

----- Q&A Section -----

