DTCC

2017第八届中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2017

Flink技术栈及其适用场景

@时金魁

2017/05

华为杭研所

shijinkui@huawei.com

Who am i

- @时金魁
- Sohu -> Alibaba -> Huawei
- Work on：high performance computing / Spark / Flink

Flink概览

Streaming Compute Frameworks

APACHE STORM™

APACHE Spark™

samza

APACHE kafka™
A distributed streaming platform

JStorm

Amazon Kinesis Streams

Aliyun StreamCompute

Azure stream-analytics

HERON
-Twitter

Flink

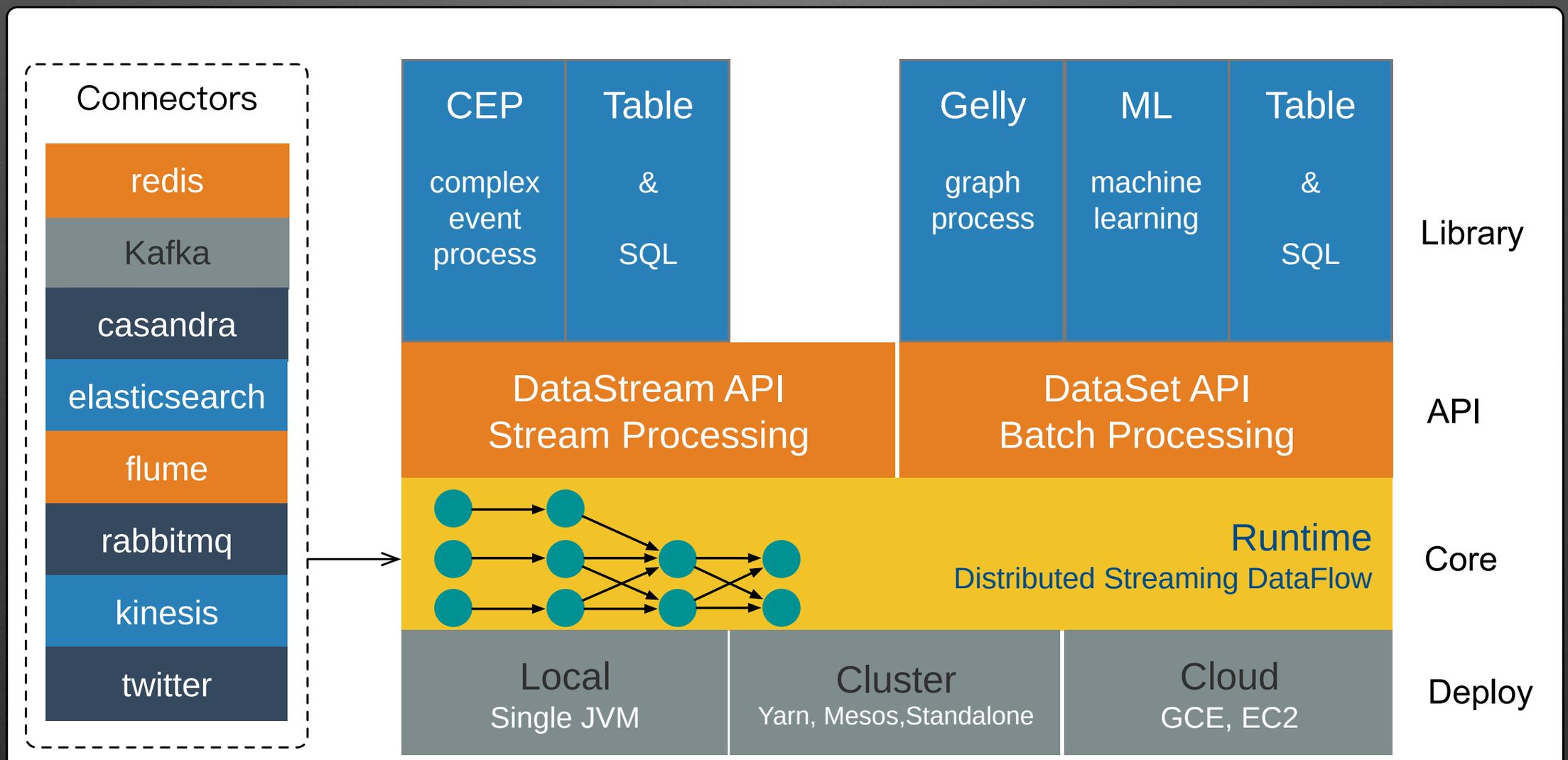
Apache Edgent (incubating) -IBM

akka

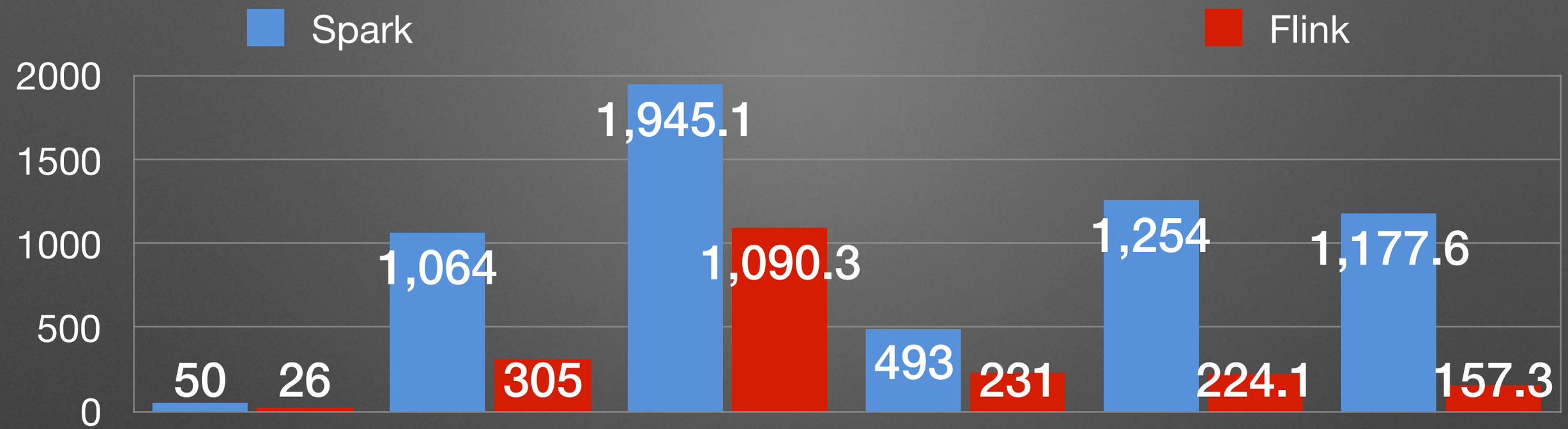
Apache Gearpump (incubating) -Intel

HUAWEI

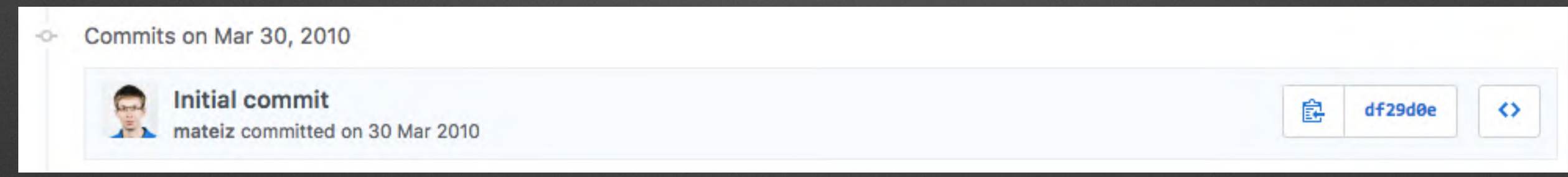
Flink



Open source community Status



Committer	Contributers	Commits	PR	Star	Fork



批处理：兴盛

流计算：兴起

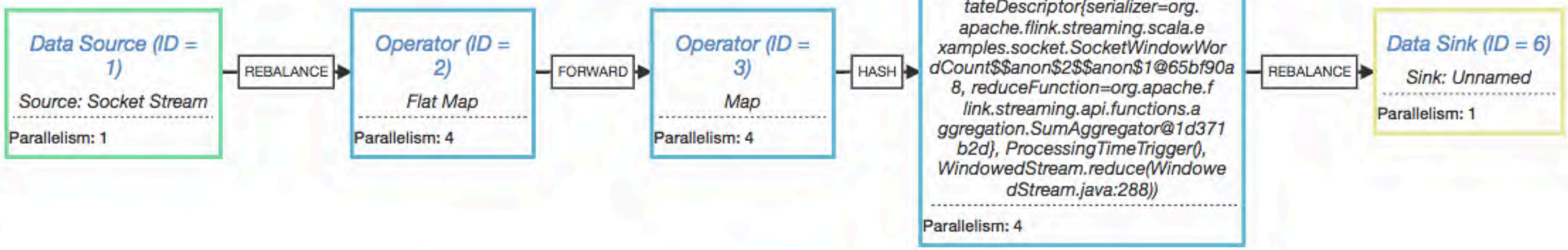
Flink example

```
case class WordWithCount(word: String, count: Long)

val windowCounts = env
      .socketTextStream(hostname, port, '\n')
      .flatMap { w => w.split("\\s") }
      .map { w => WordWithCount(w, 1) }
      .keyBy("word")
      .timeWindow(Time.seconds(5))
      .sum("count")

env.execute("Socket Window WordCount")
```

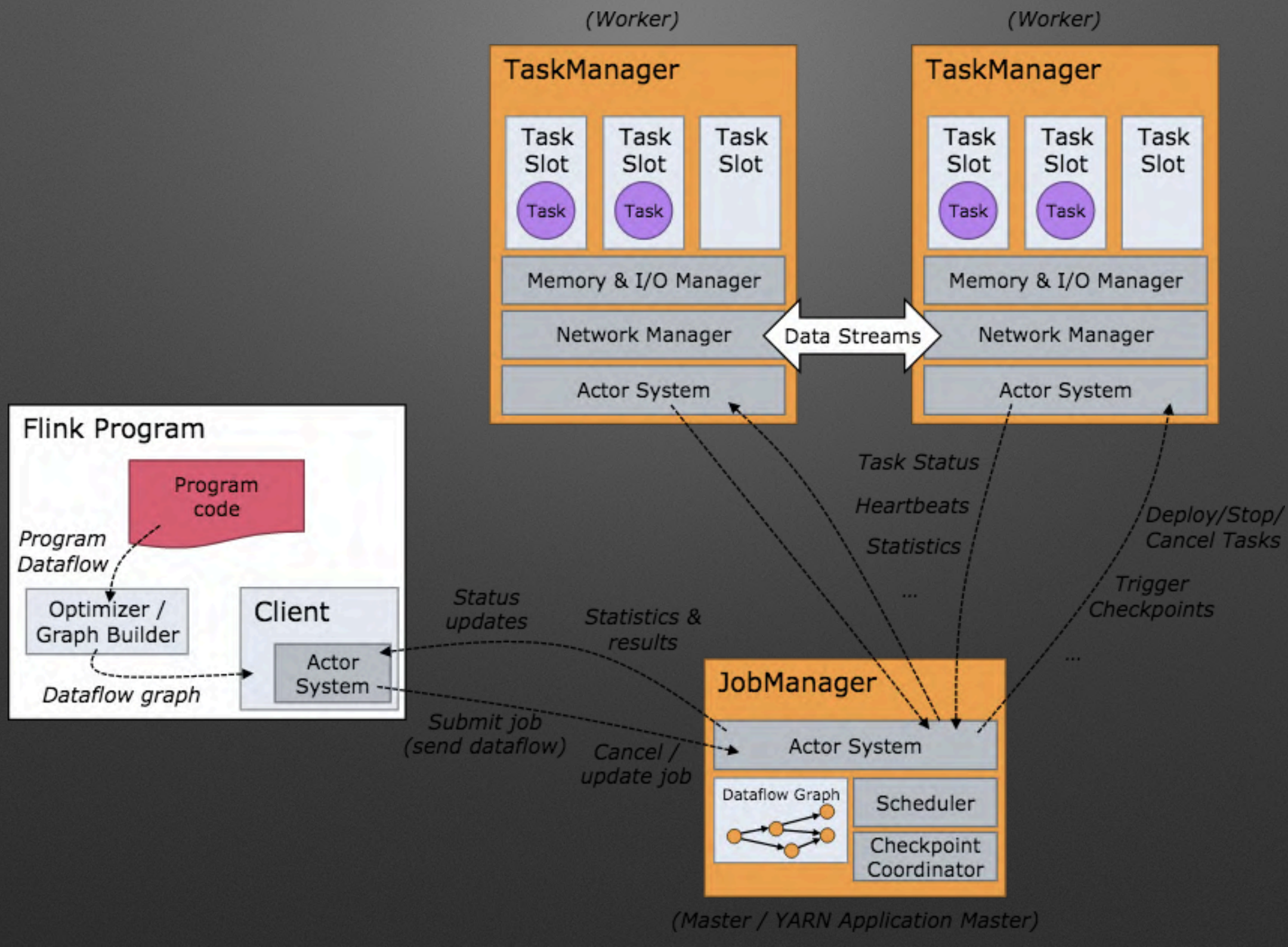
Data Source

Transform Function

Execution

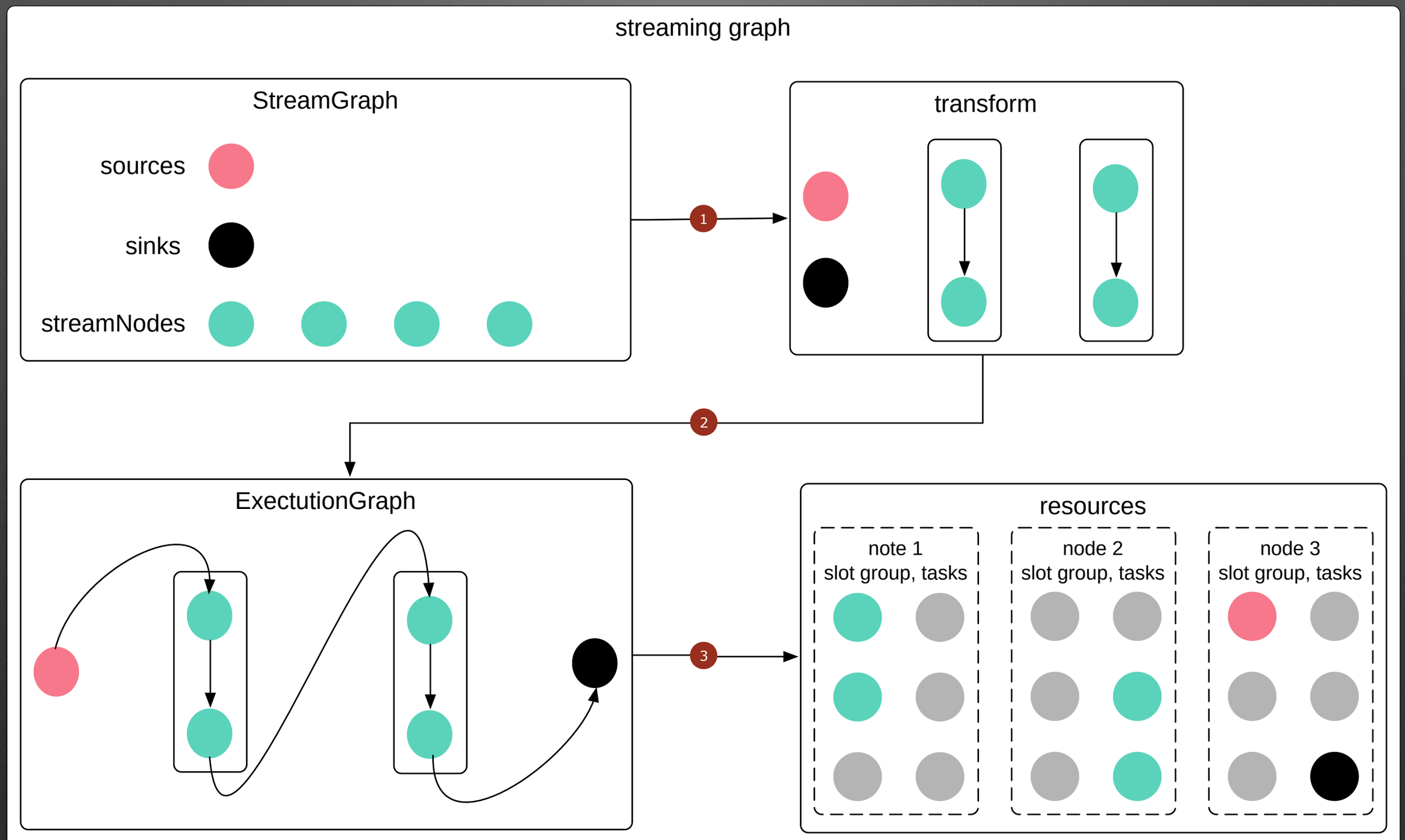


Flink architecture

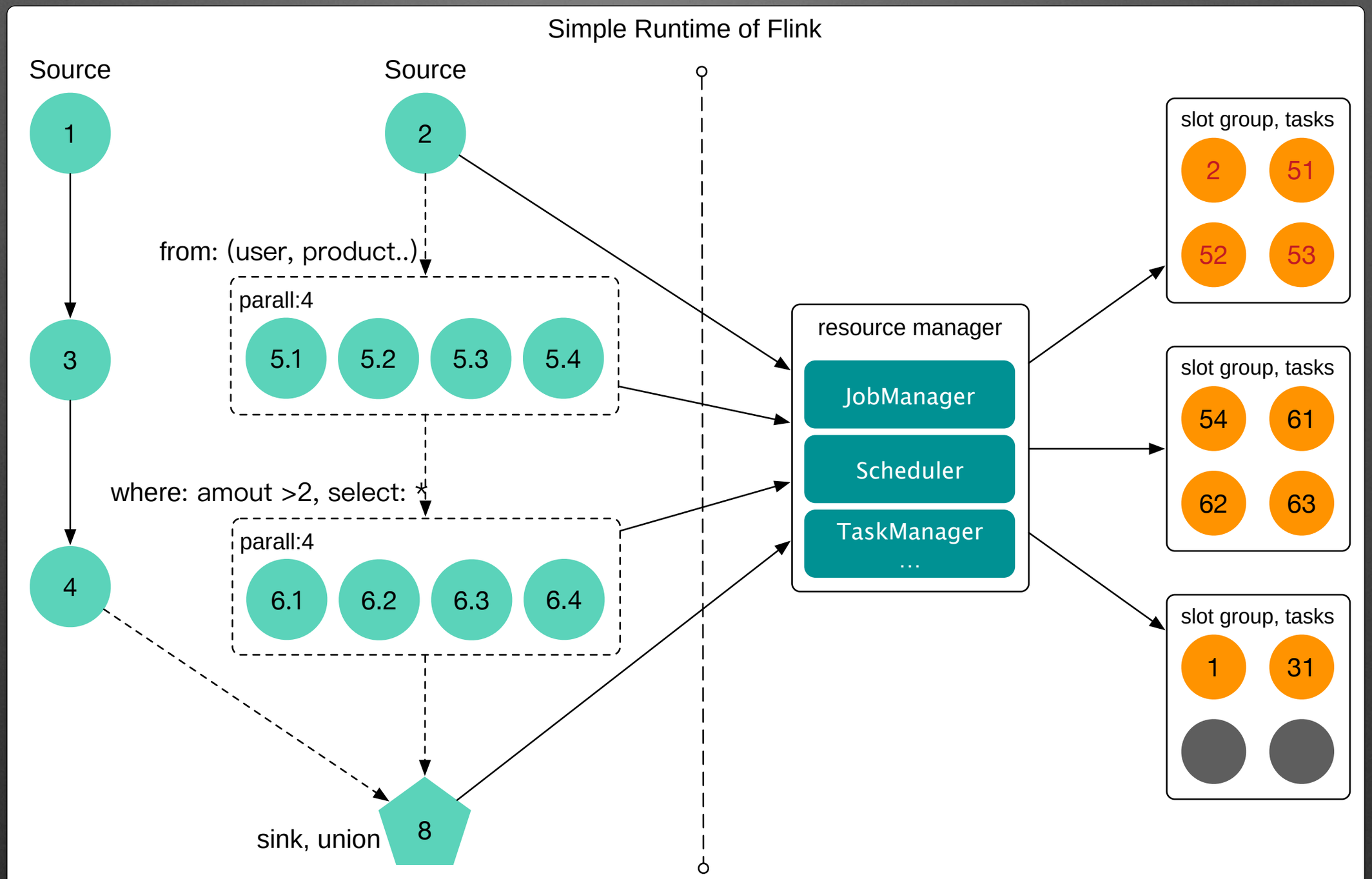


Flink Inside

build graph


streaming graph
StreamGraph
sources
sinks
streamNodes
1
transform
2
ExectutionGraph
3
resources
note 1
slot group, tasks
node 2
slot group, tasks
node 3
slot group, tasks

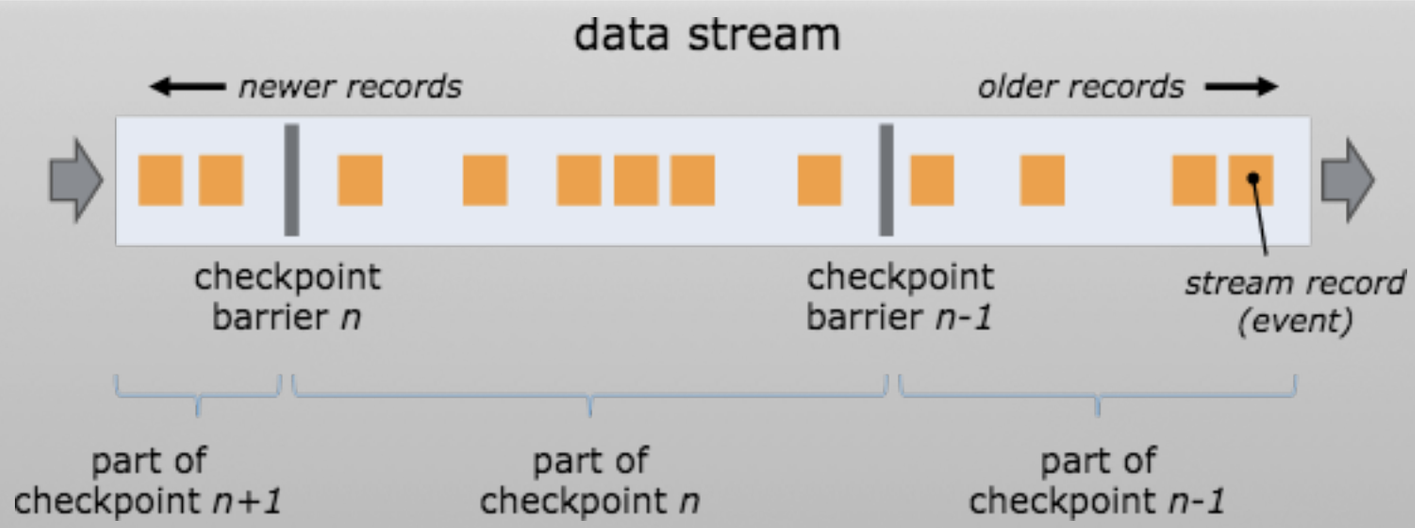
Runtime


Simple Runtime of Flink
Source
Source
1
2
from: (user, product..)
parall:4
5.1
5.2
5.3
5.4
3
where: amout >2, select: *
parall:4
6.1
6.2
6.3
6.4
4
sink, union
8
resource manager
JobManager
Scheduler
TaskManager
...
slot group, tasks
2
51
52
53
slot group, tasks
54
61
62
63
slot group, tasks
1
31

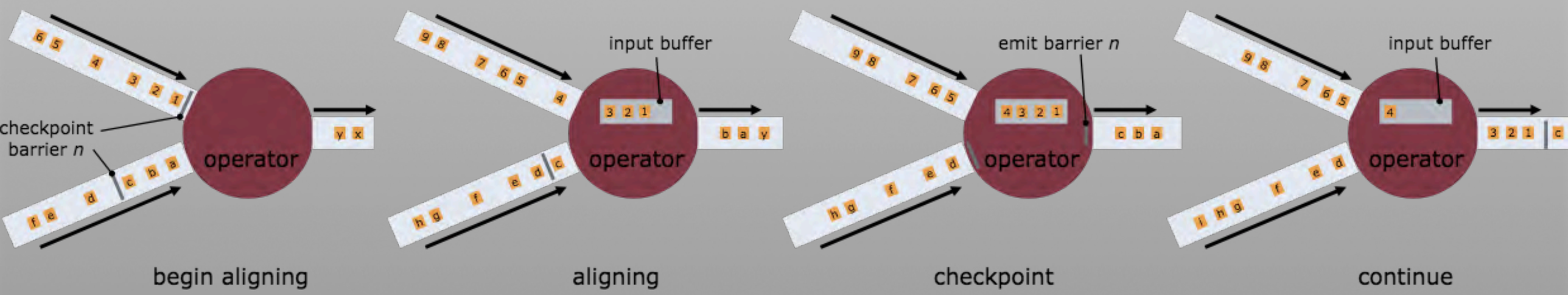
State backend and checkpoint

persist the running state:

1. memory(default 5M)
2. fs: support hdfs and dir
3. rocksDB: kv store rocksDB issue-1988 performance problem

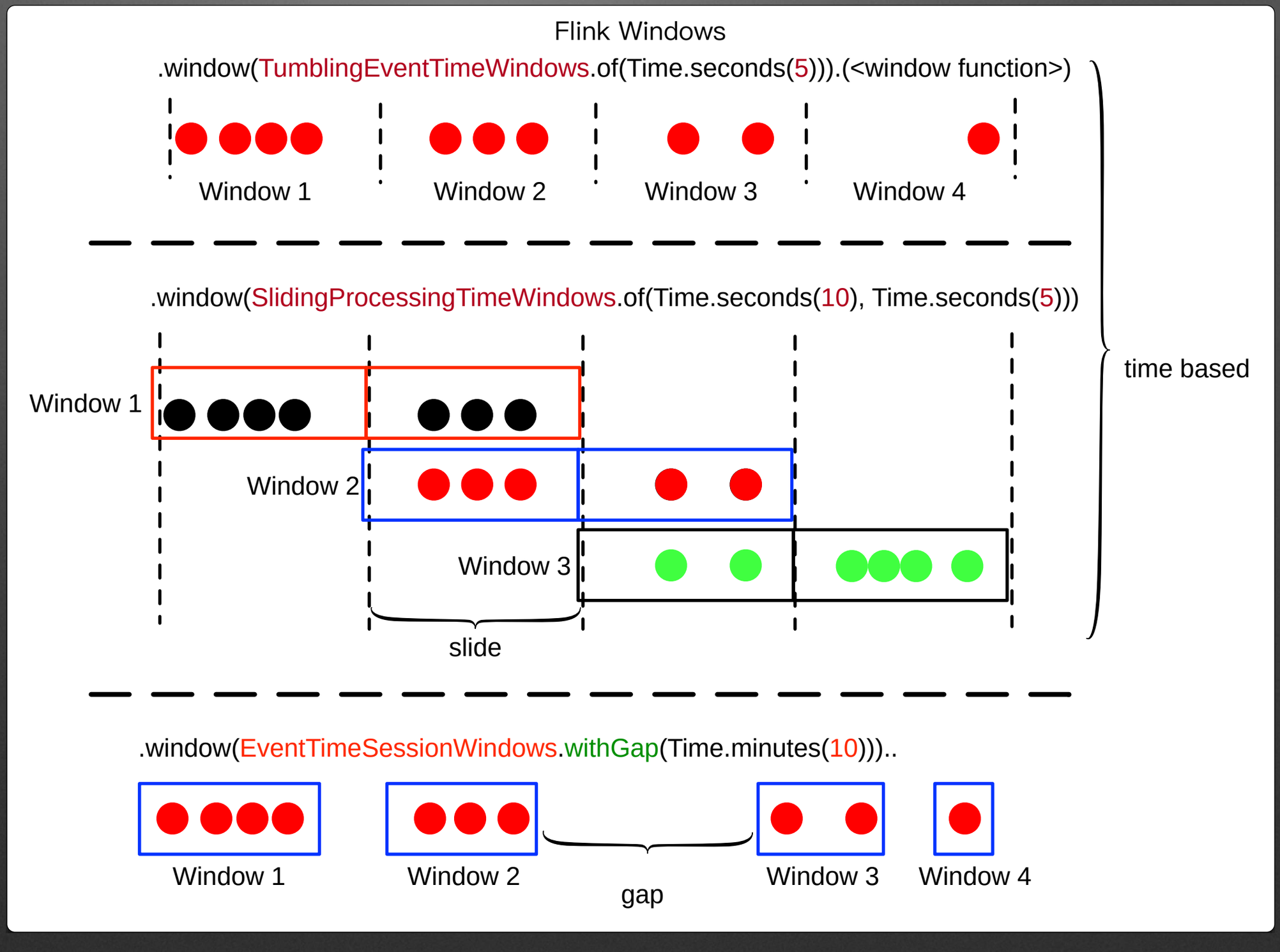


Picture is from Flink community.

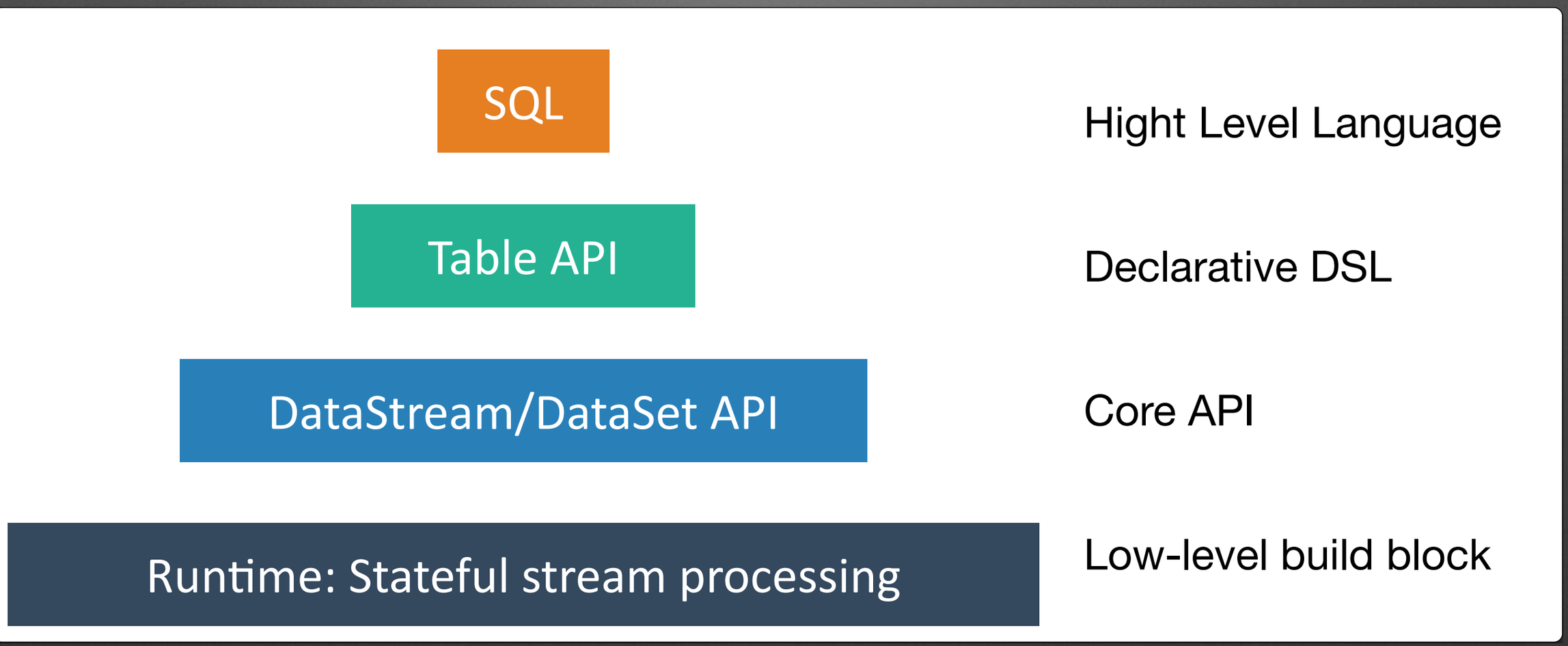


windows

- Tumbling Window
- Sliding Window
- Session Window
- Global Window



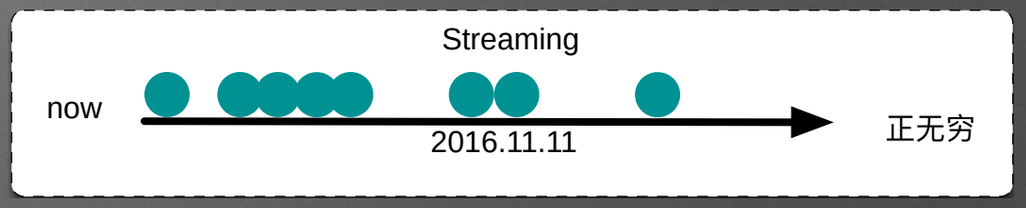
Stream SQL

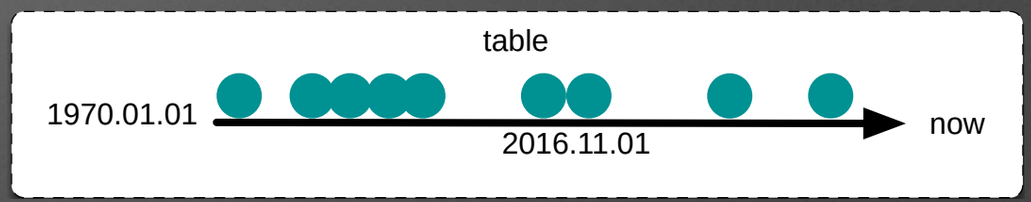

SQL
Table API
DataStream/DataSet API
Runtime: Stateful stream processing
Hight Level Language
Declarative DSL
Core API
Low-level build block

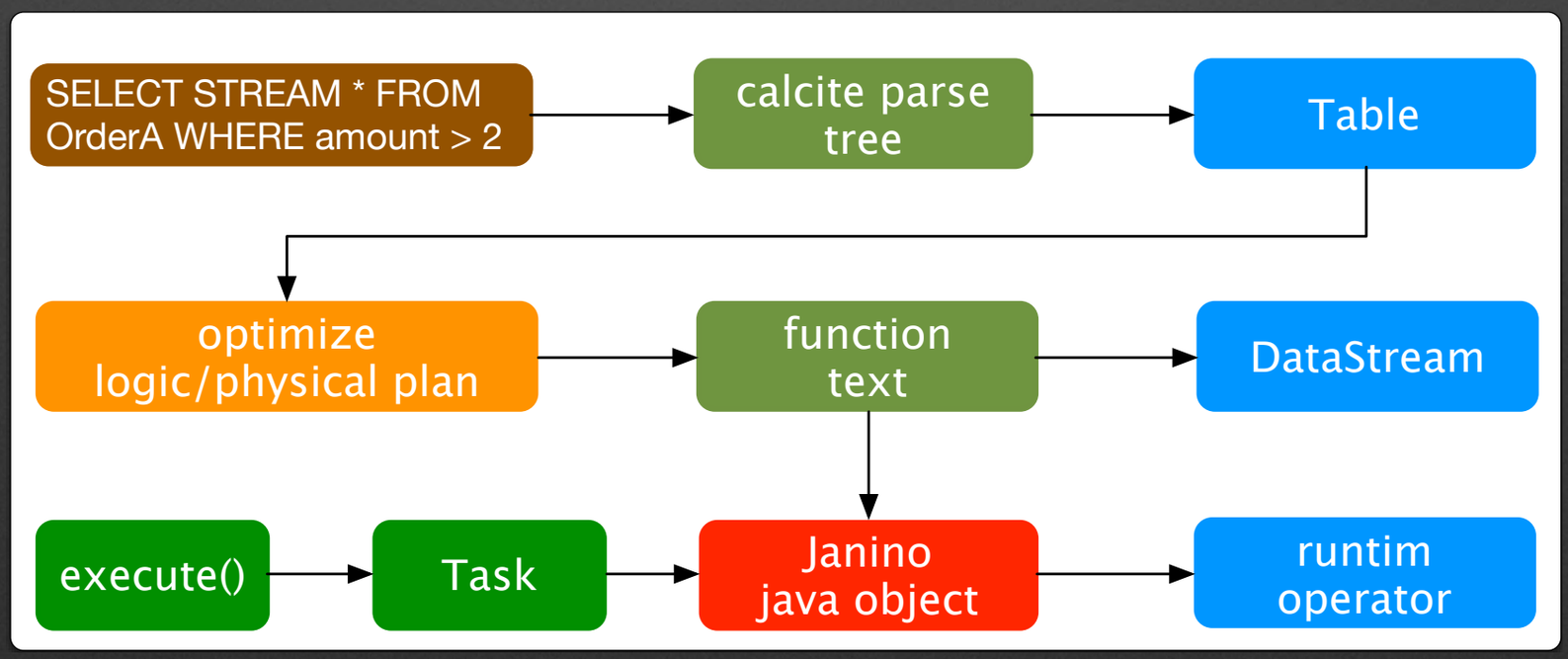
Calcite in Flink SQL

```
SELECT STREAM CEIL(rowtime TO HOUR) AS rowtime,
productId, COUNT(*) AS c, SUM(units) AS units
FROM Orders GROUP BY CEIL(rowtime TO HOUR), productId;
```

rowtime	productId	c	unit
11:00:00	20	3	12
11:00:00	3	2	3
12:00:00	4	4	44
..	..	..	..







Flink SQL最新进展

- 最新支持
  1. 支持Project、Filter、Union、UDAGG等基本能力
  2. Row-Window（over-window）
  3. Group-Window
  4. Calcite于1.12.0版本正式支持HOP/SESSION/TUMBLE，Flink目前已合入并支持

- 进展中
  1. join

     流流join、流表join等特性都已经有PR正在进行中
  2. 其他SQL能力支持

     包括sort/limit、inner-query、distinct等特性都有相应PR正在进行
  3. dynamic table

     流表对偶性，提供批和流无损转化的能力，需引入Retraction机制 ——FLINK-6047

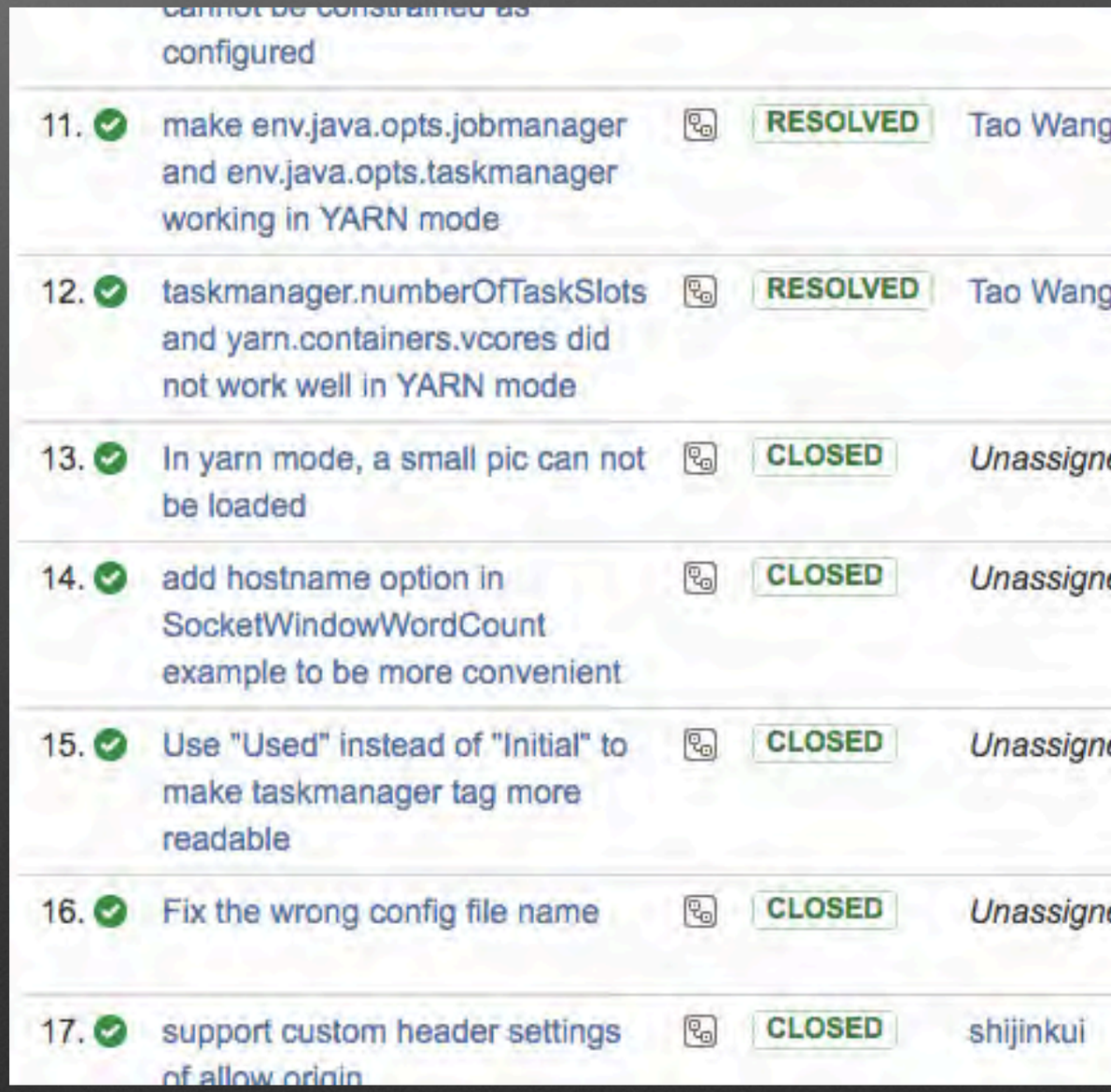
our work plan

- Security enhance
- Stream SQL develop
- performance optimization

Security enhance

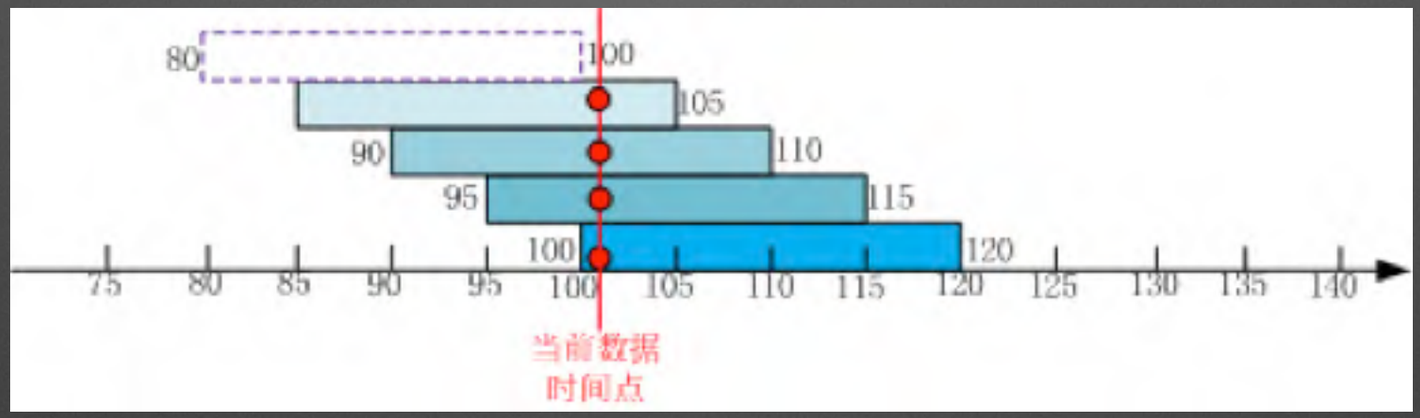
- auth：client and jm, jm and tm, tm and tm, with hadoop/zk/kafka
- security transfer: SSL enable
- web security header
- zk ACL

total fixed: nearly one hundred issues



slide window

window(SlidingEventTimeWindows.of(Time.seconds(20), Time.seconds(5)))



痛点：流数据会被分配到20/5=4个不同的窗口中，数据在内存中保存了4份。当窗口大小/滑动周期非常大时，冗余现象非常严重。

others

- exactly-once
- fault-tolerance
- stateful
- gelly
- FlinkML
- storm compatible API
- RM: yarn, mesos, kubernetes, docker

use case

- Real-time risk control
- ~~ETL~~
- Real-time analyze
- Anti-fraud
- cloud service

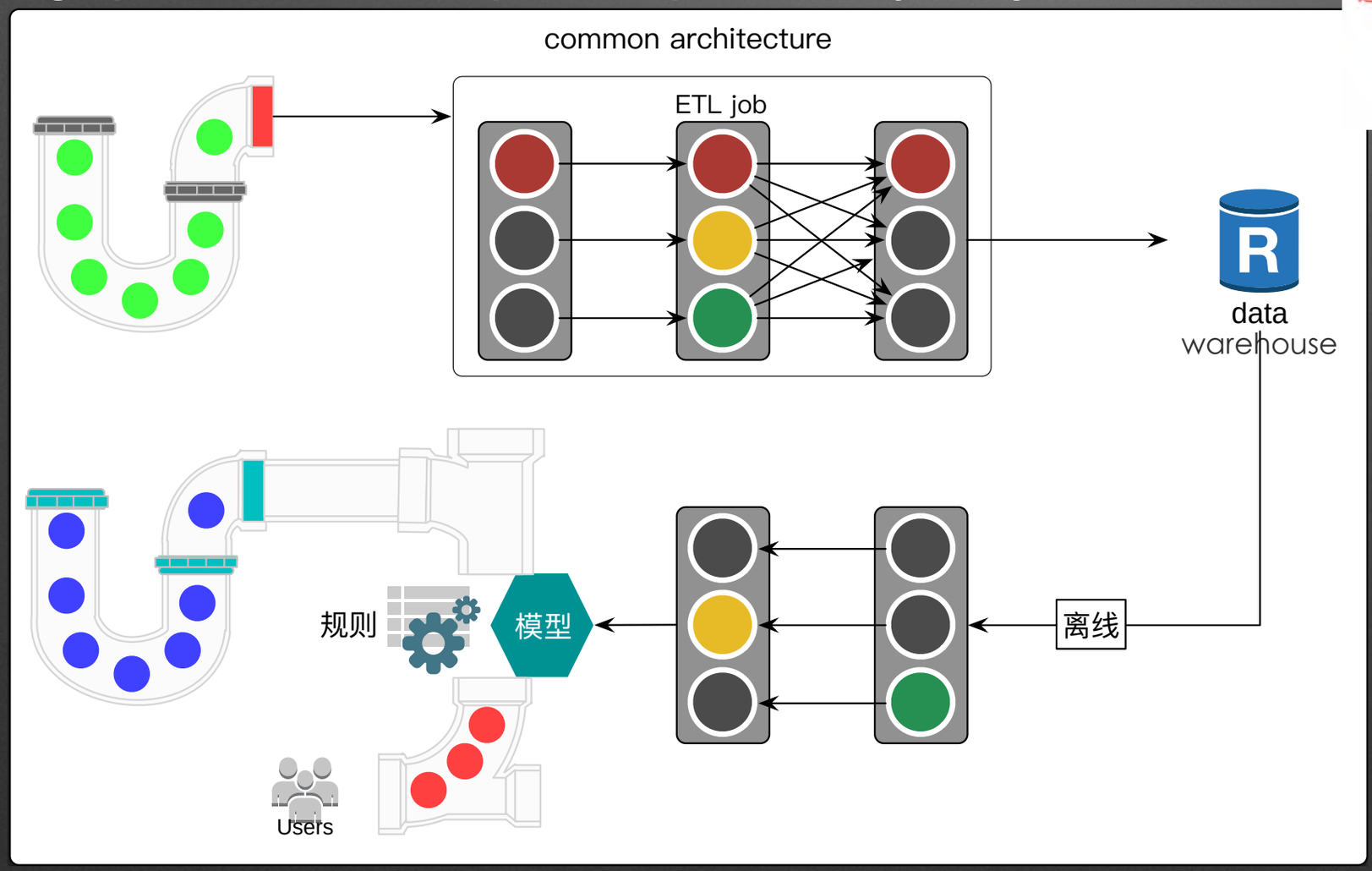
Real-time risk control

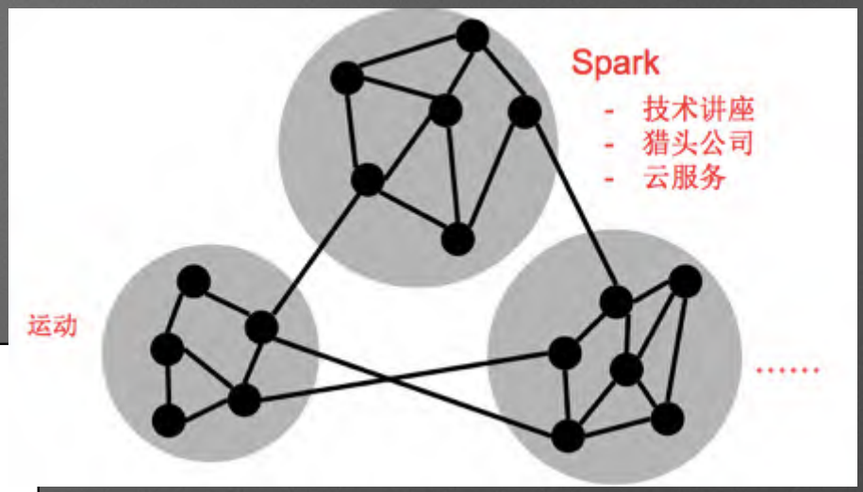
- millisecond level（less than 100ms）
- nearly million events per second
- rules：Stream API, Window API, SQL
- response：

API Gateway:
-> user connection
-> request(source data)
-> Flink process(distributed)
-> response(sink data)
-> routing to origin connection
-> user connection, flush data

Anti-fraud

- based on the rules
- based on the offline analyze model
- graph framework: GraphX, Graphite, neo4j, Gelly





- community detection：
  - GN
  - Label Propagation Algorithm
  - KCore Subgraph
  - Local Expansion
  - Particle Competition
  - Game Theory
- PageRank

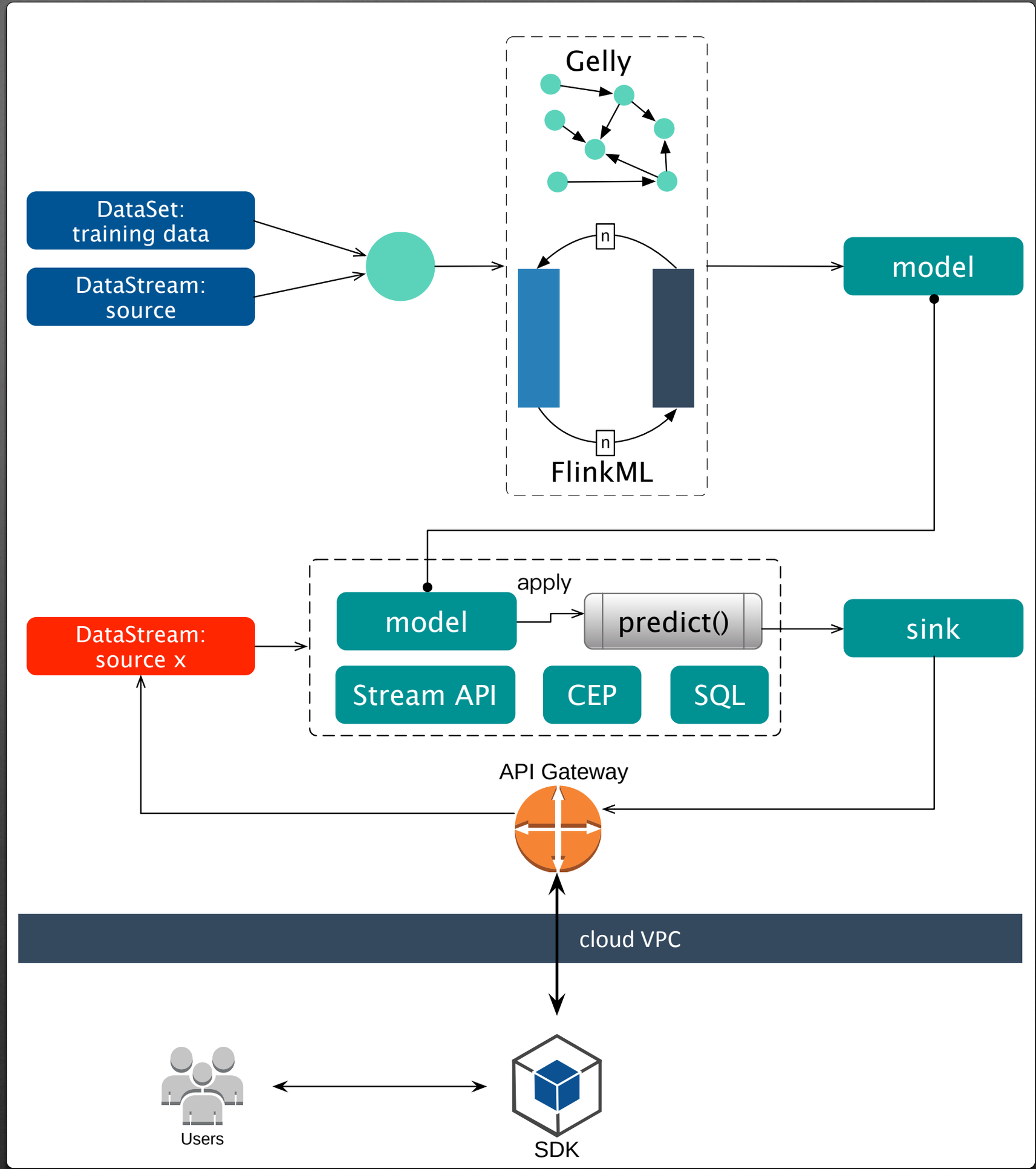
cloud service

- AWS: Kinesis — StreamSQL
- Aliyun: StreamCompute — Serverless
- Azure: stream-analytics — Mesos/Kubernetes

HUAWEI

What I image like:

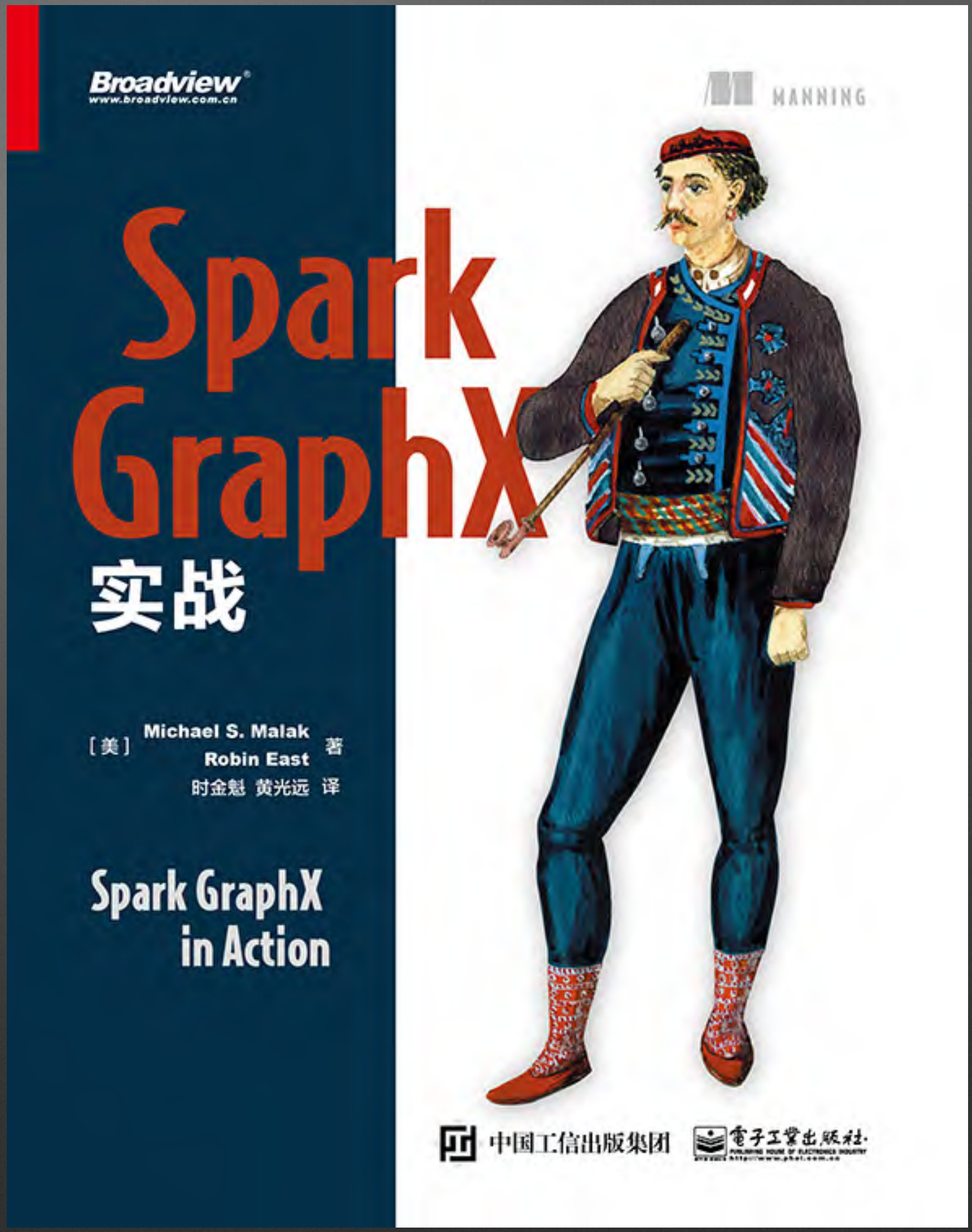
1. offline model used online
2. API Gateway
3. cloud hidden the detail
4. Stream ML and Graph



My translated book

下午4点茶歇，签售

地点：博文视点展位



Welcome to Huawei.

1. UED
2. Scala developer

shijinkui@huawei.com

HUAWEI

THANKS

SequeMedia 盛拓传媒 IT168.com ITPUB ChinaUnix.net