Flink China

# Flink: 构建下一代大数据处理引擎

巴真 @FlinkChina Meetup

Flink China | IT大咖说 知识共享平台

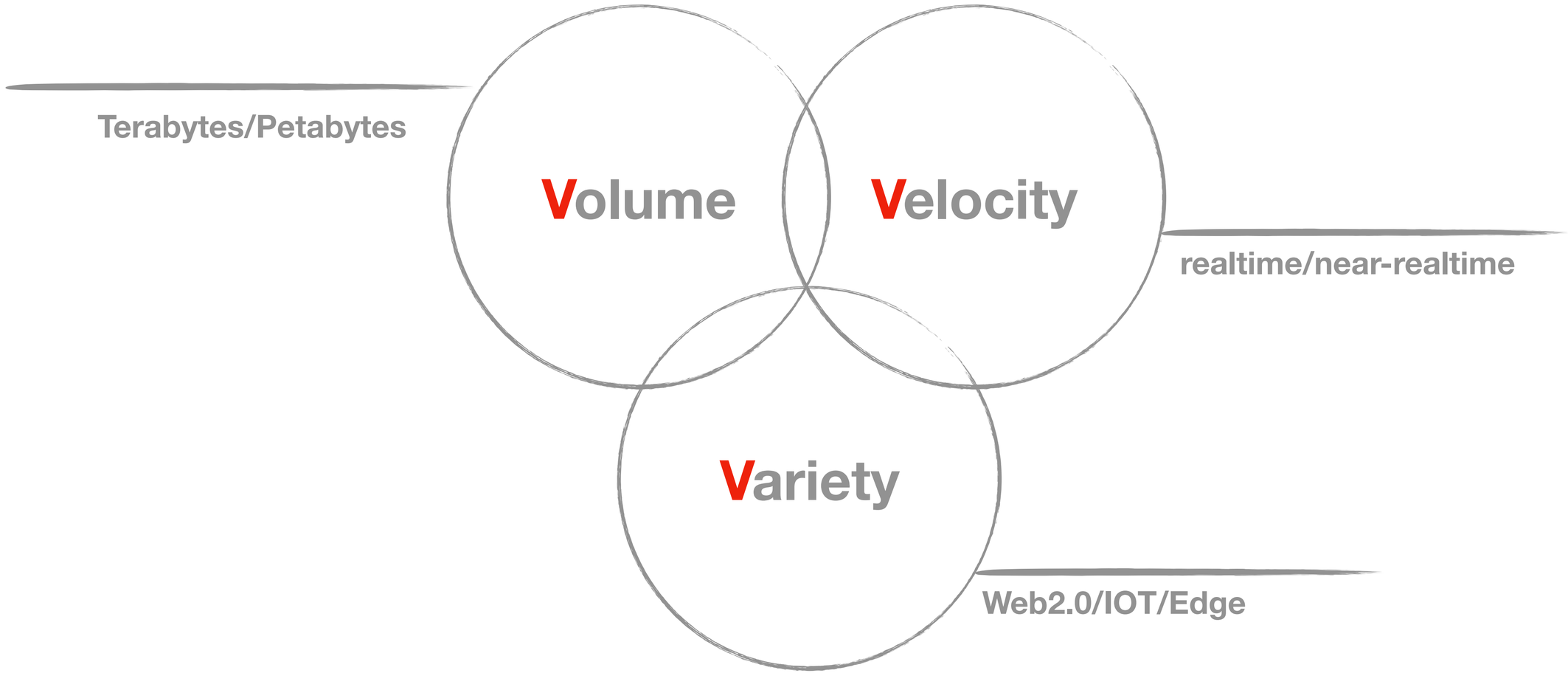
# Agenda

数据趋势

业界案例

阿里思考

Flink@阿里


Terabytes/Petabytes
Volume
Velocity
realtime/near-realtime
Variety
Web2.0/IOT/Edge

# Volume



EB Total

PB Everyday

100B Event/Day

100M Events/sec

# Volume

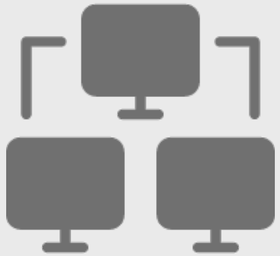

3亿用户

百万商家

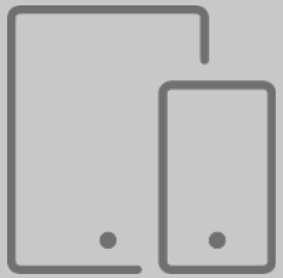
千亿成交

# Variety

IT化



网络化



移动化

万物互联

# Variety


AWS S3

阿里云OSS

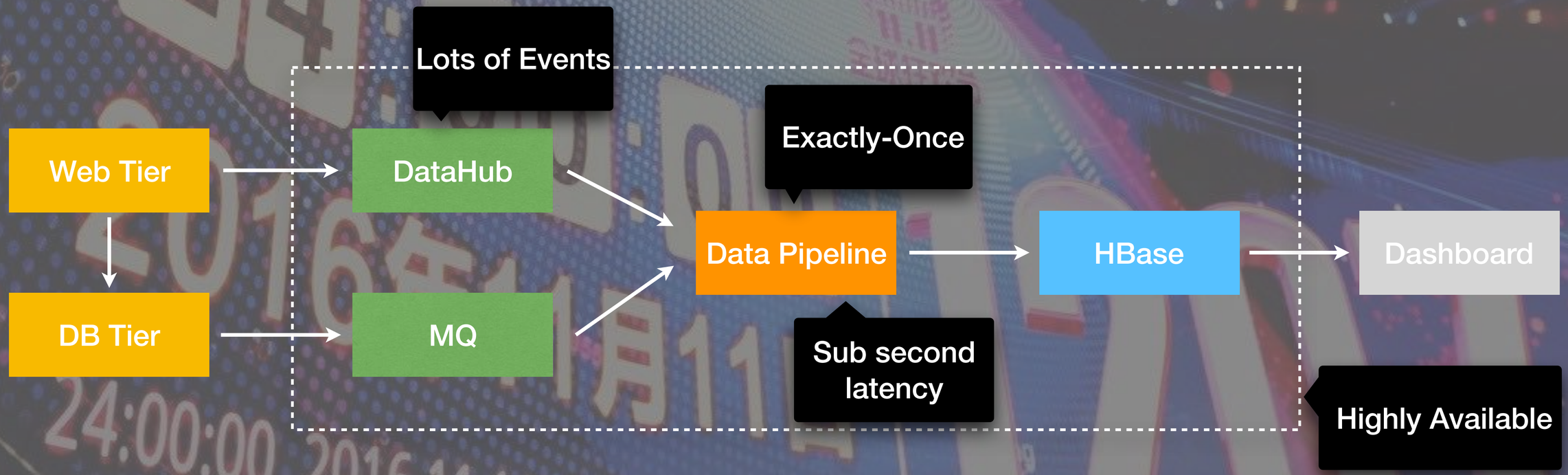
Semi-Structured Unstructured

# Velocity

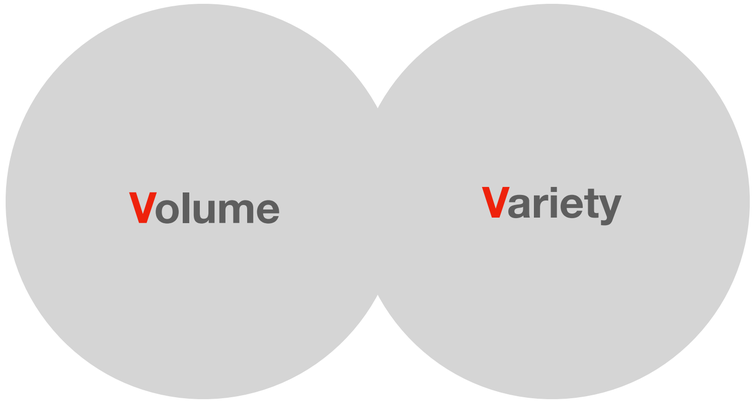
472M events/sec

Sub-second Latency

# Velocity


Lots of Events
Exactly-Once
Web Tier
DataHub
Data Pipeline
HBase
Dashboard
DB Tier
MQ
Sub second latency
Highly Available

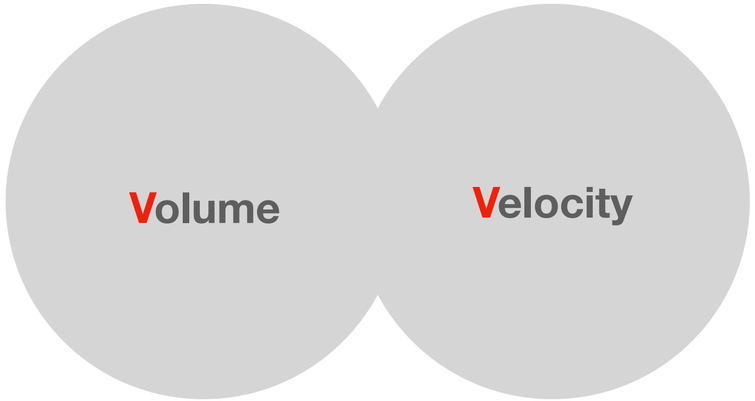
# 数据趋势




Flink China

hadoop + APACHE Spark™

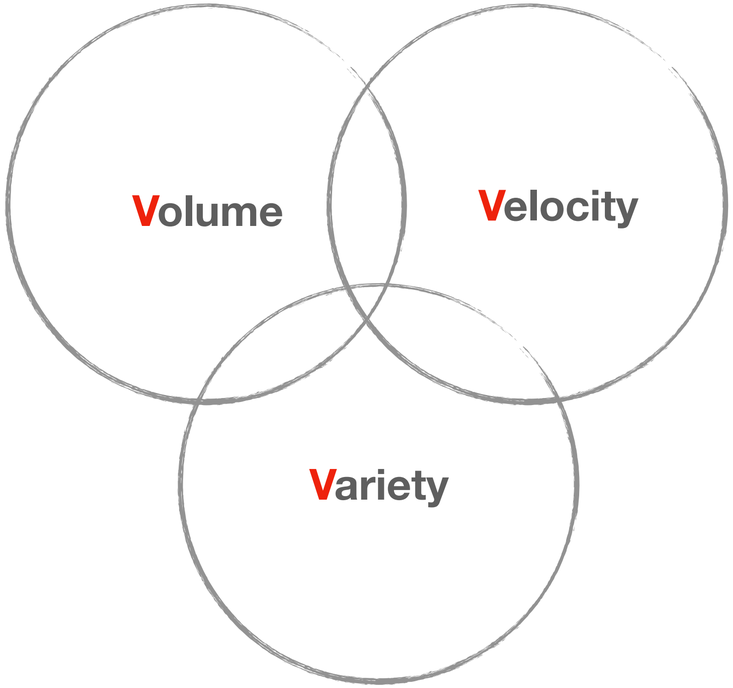
# 数据趋势




Flink China

APACHE STORM™ + Flink

# 数据趋势




?

# 计算趋势

计算碎片化

模型多样化

Flink China

# BIG DATA LANDSCAPE 2017


数万Dataer

百万Job

在阿里，人人都是数据分析师

BI Engineer

AI Engineer

在阿里，BI工程师在转型AI工程师

## 典型案例分析

### 阿里业务

超大规模、超级复杂、计算多样、数据多样、用户众多

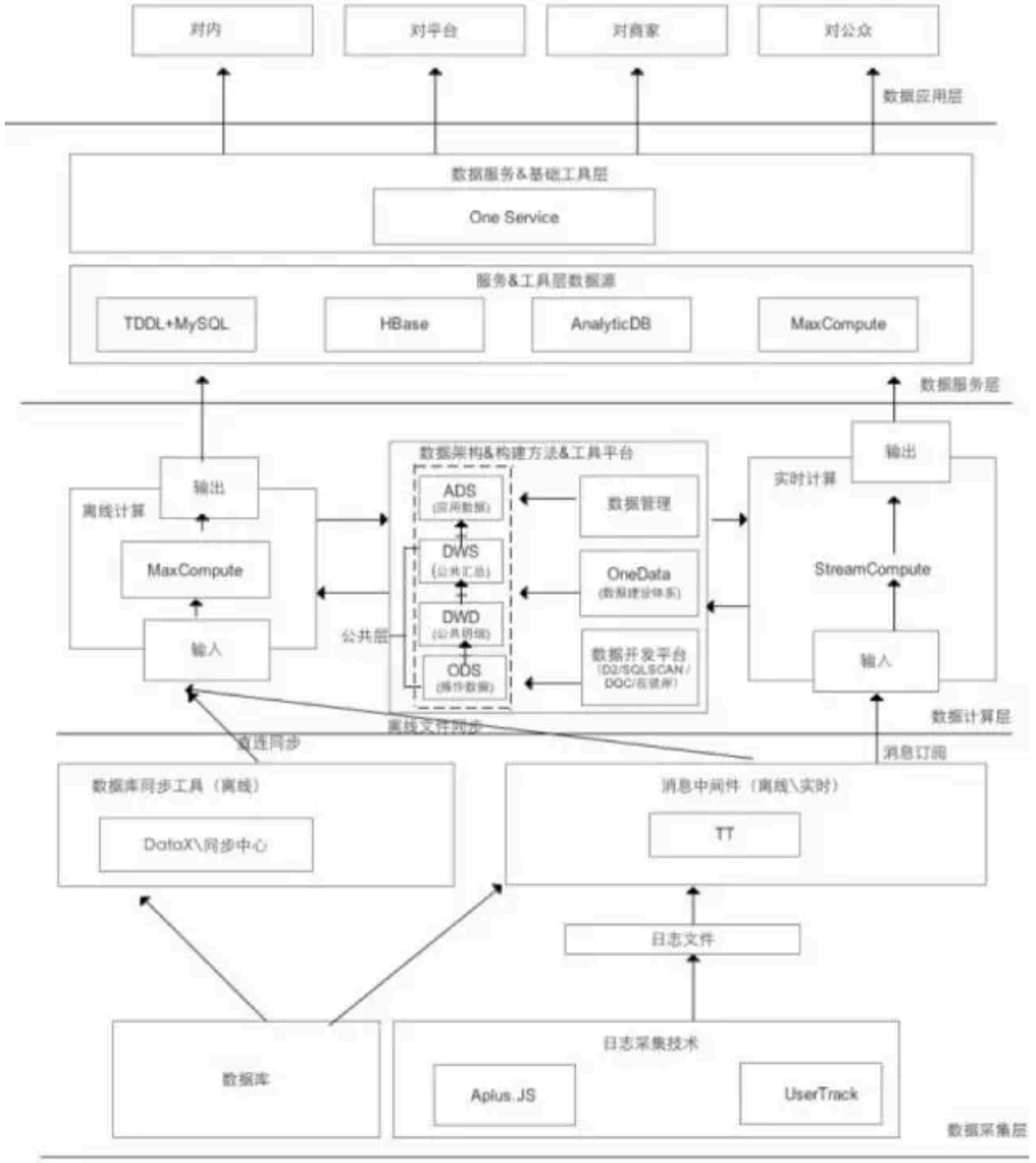
### 滴滴业务

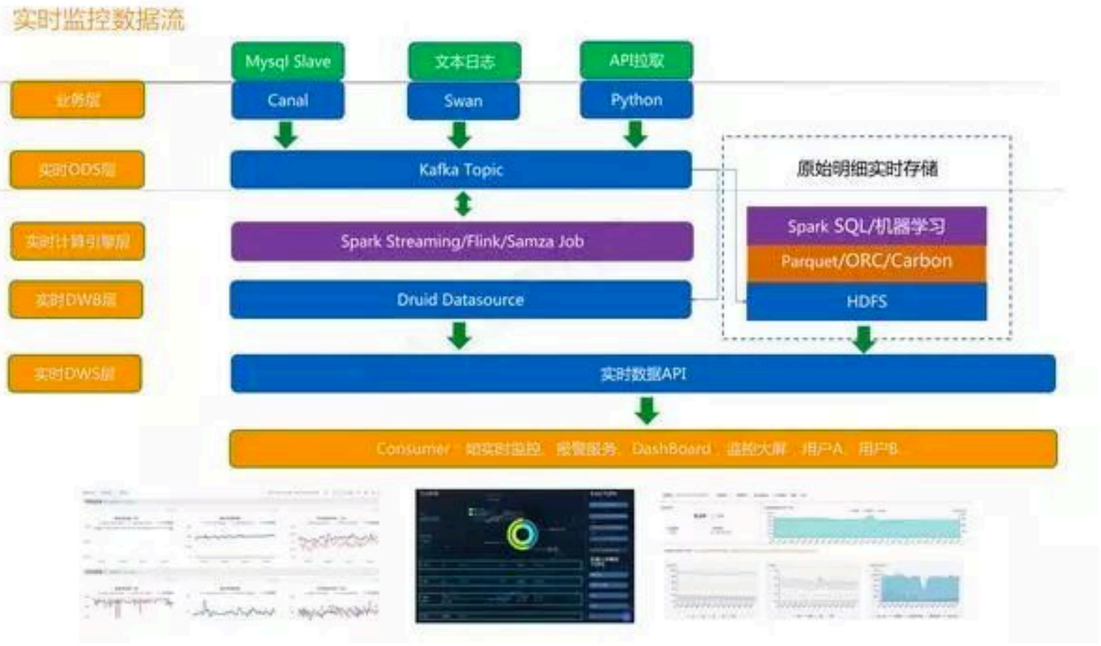
独角兽公司，爆发式增长、计算时效性强

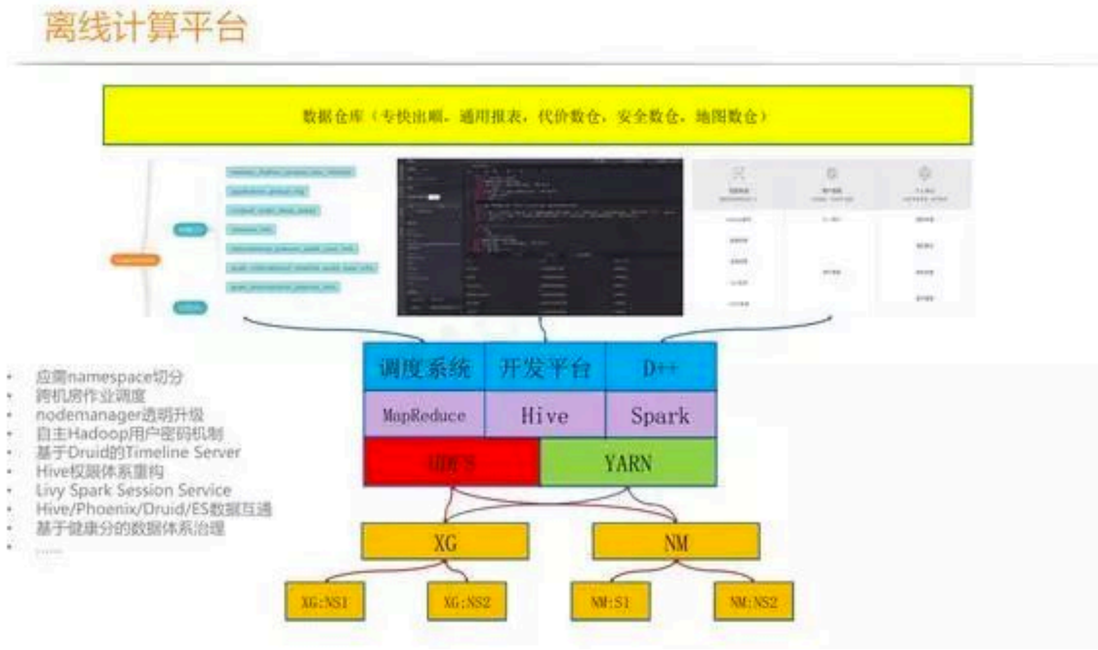
### 知乎业务

典型UGC、Web2.0业务

# 阿里业务

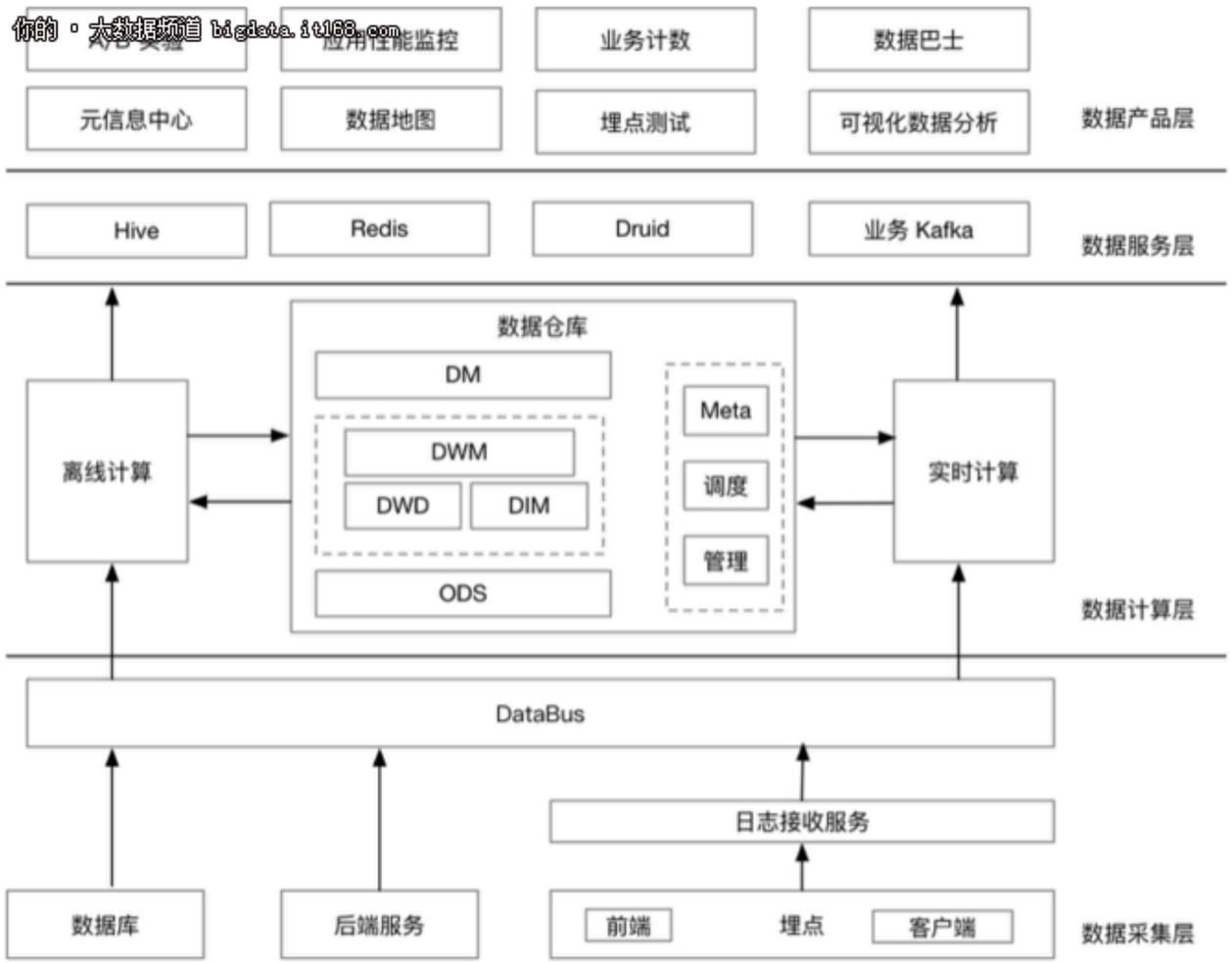

对内
对平台
对商家
对公众
数据应用层
数据服务&基础工具层
One Service
服务&工具层数据源
TDDL+MySQL
HBase
AnalyticDB
MaxCompute
数据服务层
数据架构&构建方法&工具平台
输出
离线计算
MaxCompute
输入
ADS
DWS
DWD
ODS
公共层
数据管理
OneData
数据开发平台
实时计算
输出
StreamCompute
输入
数据计算层
离线文件同步
消息订阅
数据库同步工具（离线）
DataX\同步中心
消息中间件（离线\实时）
TT
日志文件
数据库
日志采集技术
Aplus.JS
UserTrack
数据采集层


实时监控数据流
Mysql Slave
Canal
文本日志
Swan
API拉取
Python
业务层
实时ODS层
实时计算引擎层
实时DW8层
实时DWS层
Kafka Topic
Spark Streaming/Flink/Samza Job
Druid Datasource
实时数据API
原始明细实时存储
Spark SQL/机器学习
Parquet/ORC/Carbon
HDFS
Consumer 如实时监控 报警服务 DashBoard 监控大屏 用户A 用户B


离线计算平台
数据仓库（专快出顺，通用报表，代价数仓，安全数仓，地图数仓）
应需namespace切分
跨机房作业调度
nodemanager透明升级
自主Hadoop用户密码机制
基于Druid的Timeline Server
Hive权限体系重构
Livy Spark Session Service
Hive/Phoenix/Druid/ES数据互通
基于健康分的数据体系治理
调度系统
开发平台
D++
MapReduce
Hive
Spark
YARN
XG
NM
XG:NS1
XG:NS2
NM:S1
NM:NS2

# 滴滴业务

# 知乎业务


你的 · 大数据频道 bigdata.it168.com
应用性能监控
业务计数
数据巴士
元信息中心
数据地图
埋点测试
可视化数据分析
数据产品层
Hive
Redis
Druid
业务 Kafka
数据服务层
数据仓库
DM
DWM
DWD
DIM
ODS
Meta
调度
管理
离线计算
实时计算
数据计算层
DataBus
日志接收服务
数据库
后端服务
前端
埋点
客户端
数据采集层

# 阿里在思考

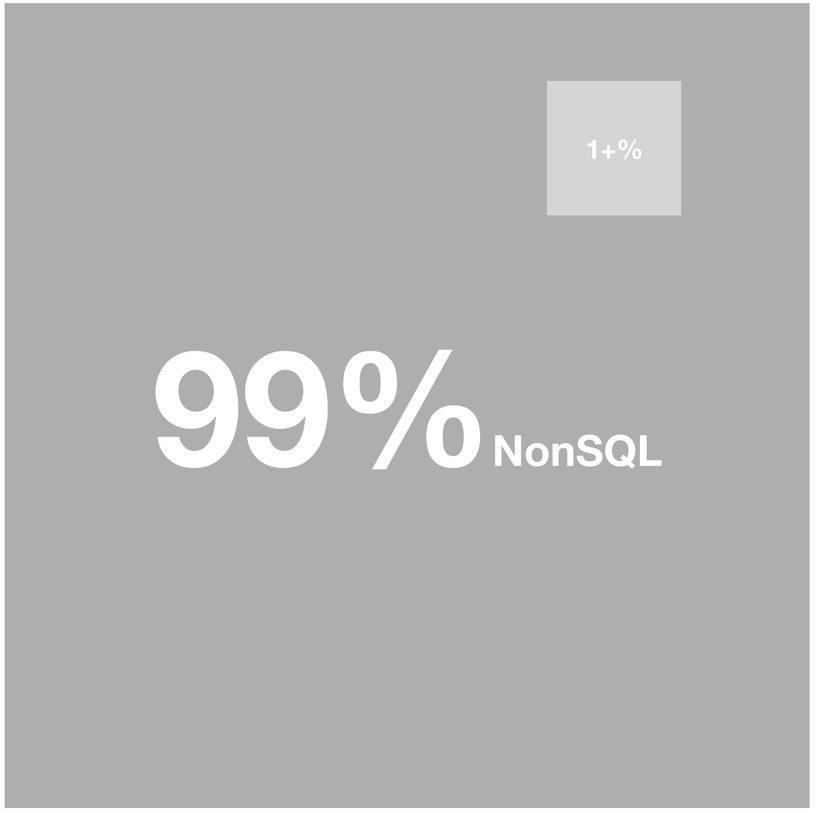
计算类型多，但核心模型可列举


描述方式多，但核心抽象可列举

人参考数据决策 -> 机器计算数据决策


计算类型多种多样，但入口业务仅此数种

Batch




Stream

# 阿里的思考

统一计算引擎: Batch/Stream/OLAP/ML/Graph...


统一抽象方式：Unified SQL & API

统一BI+AI引擎：数据清洗&数据训练

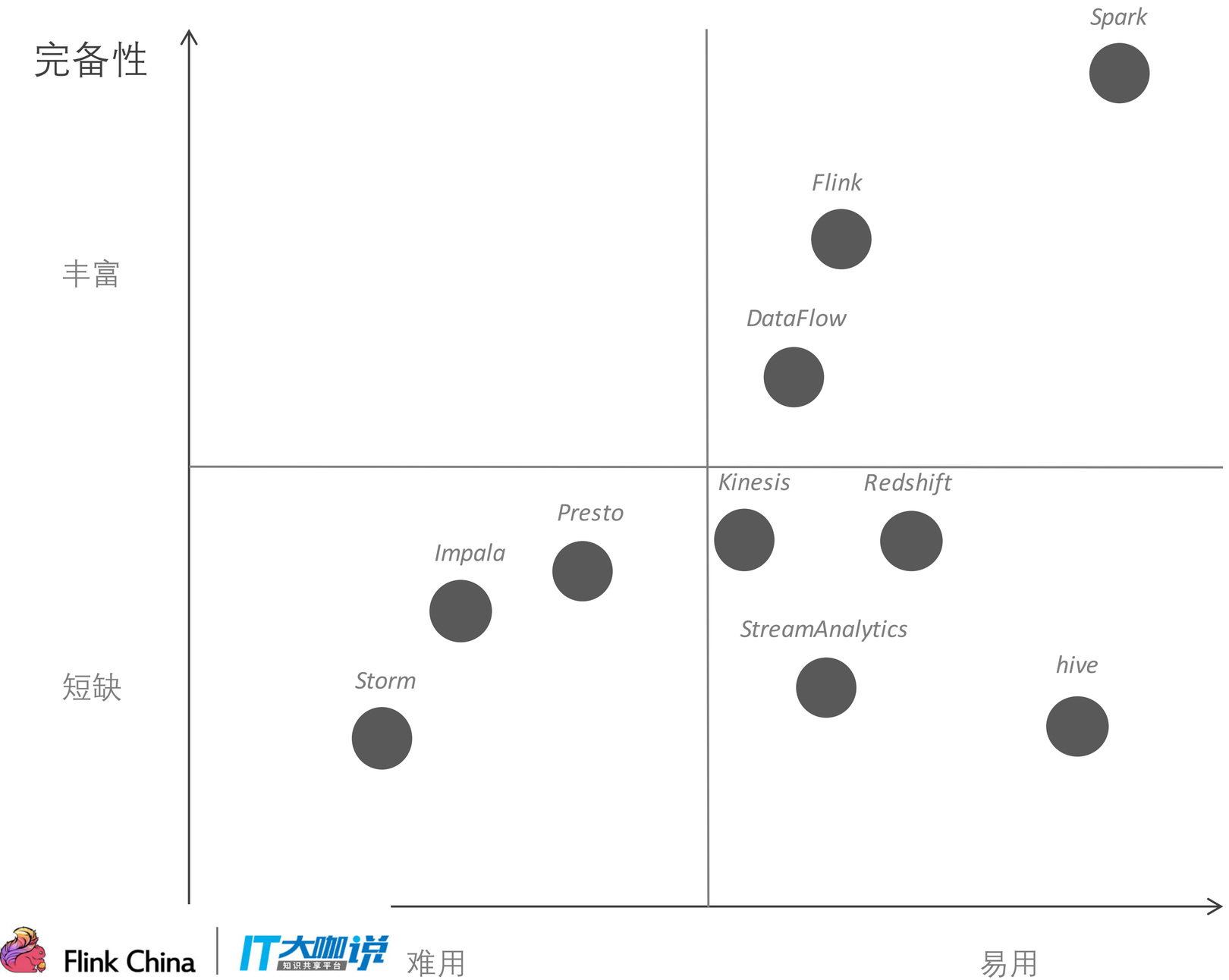
# 阿里的思考

如何构建下一代大数据处理引擎?


第一代


第二代

第三代




## Spark优势明显

功能完备: Batch/Stream/ML

简单易用: API丰富/文档较丰富/生态对接

社区活跃: 运营出彩、大数据最活跃社区

## 云产品系列

产品化: 产品化成熟，易于上手

稳定性: 稳定性好，商业托管

社区小: 产品封闭，不易形成社区

## 开源软件

社区活跃: 产品开放，易于形成社区

稳定性差: Bug多，不稳定

产品粗糙: 产品化初级，上手较难


Spark期望一套软件覆盖主要计算模型，但实际覆盖不完整

**21%**用户认为Spark Streaming在功能(集中在窗口)和时延(亚秒)等比不上Flink，增量流式业务考虑使用Flink

**17%**用户认为Spark ML部分落后，包括提供更多算法、对接TF，部分业务迁移到TF框架运行

**25%**的用户认为当前缺乏好用、内置的上层平台，包括开发界面、工作流调度，用户使用Spark同样需要重新搭建平台系统

稳定性/调优/排错 仍未解决

**31%** 用户吐槽Spark集群不稳定，经常性OOM导致业务产出不稳定

另外，几乎同样客户群体(说明都是深入生产使用Spark用户)都认为Spark作业排错、调优困难，易用性不够

中文资料/社区严重缺乏，未能形成有效组织

**30%**用户吐槽当前文档、案例过少，特别在调优、排错方面，用户往往不知所措

用户同时认为相关中文资料相比更少，时效性也落后英文社区太多

中美语言差异导致中国市场更加空白

# Flink: 下一代流式处理系统

——为什么Flink要比Storm/Spark更加优秀?


Apache Flink

+


Alibaba's Improvements

=

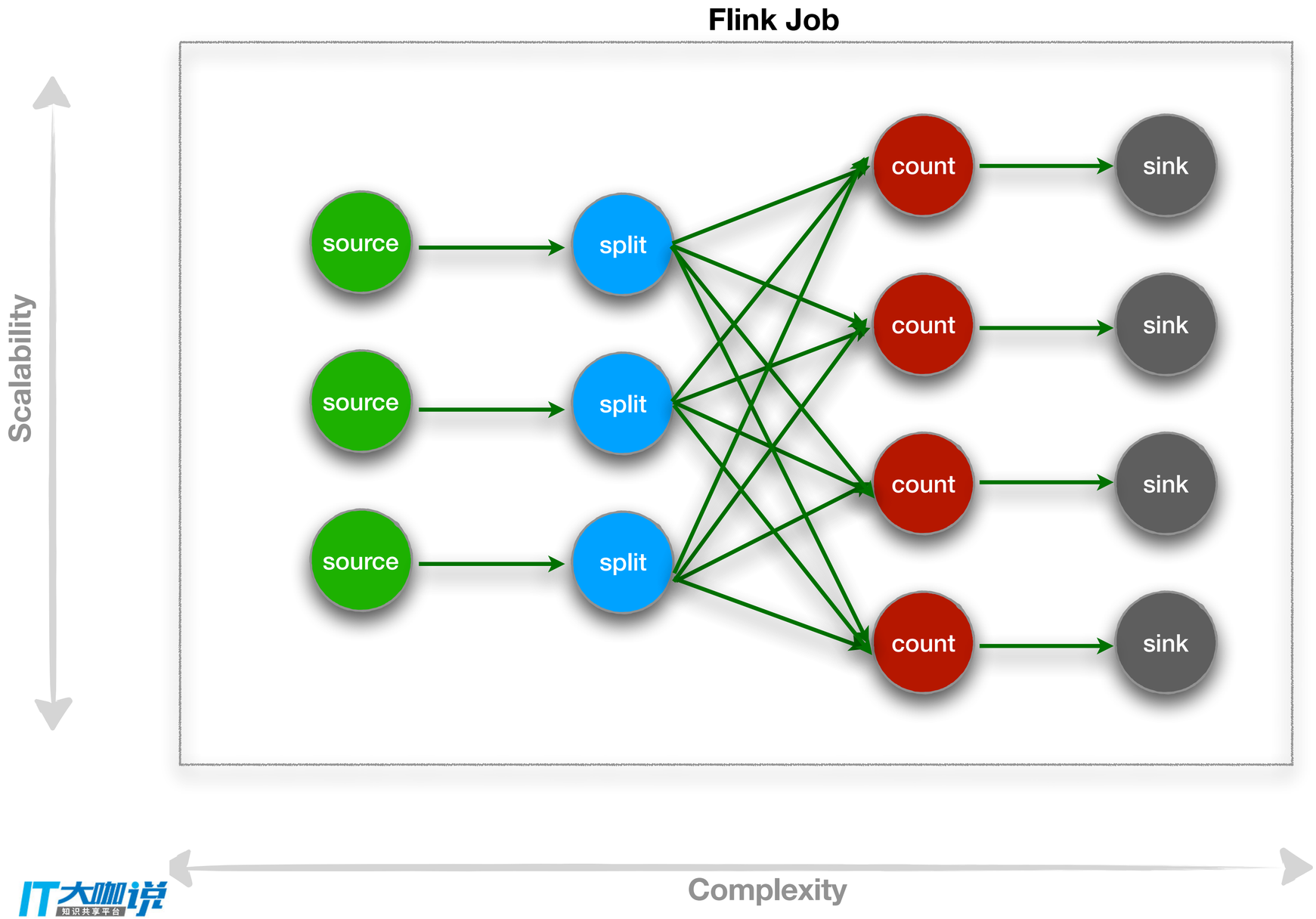
Alibaba Blink

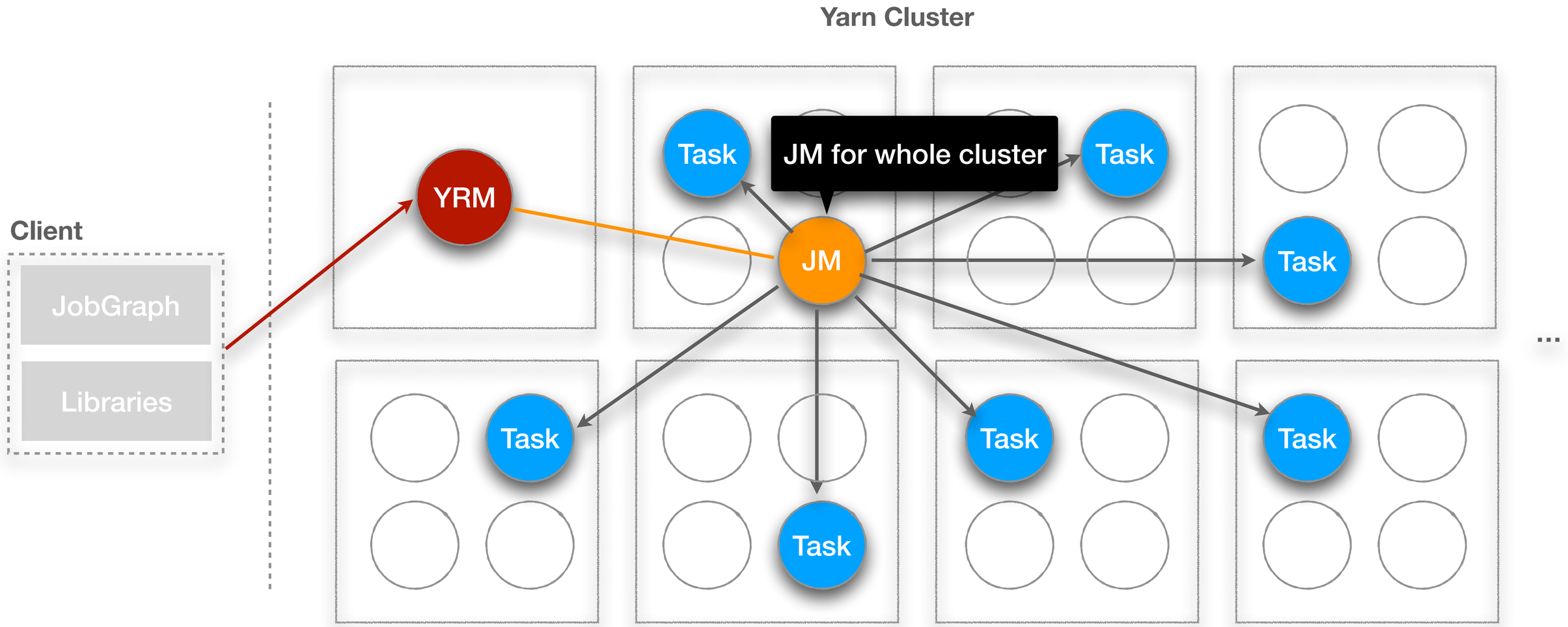
Alibaba Blink Platform

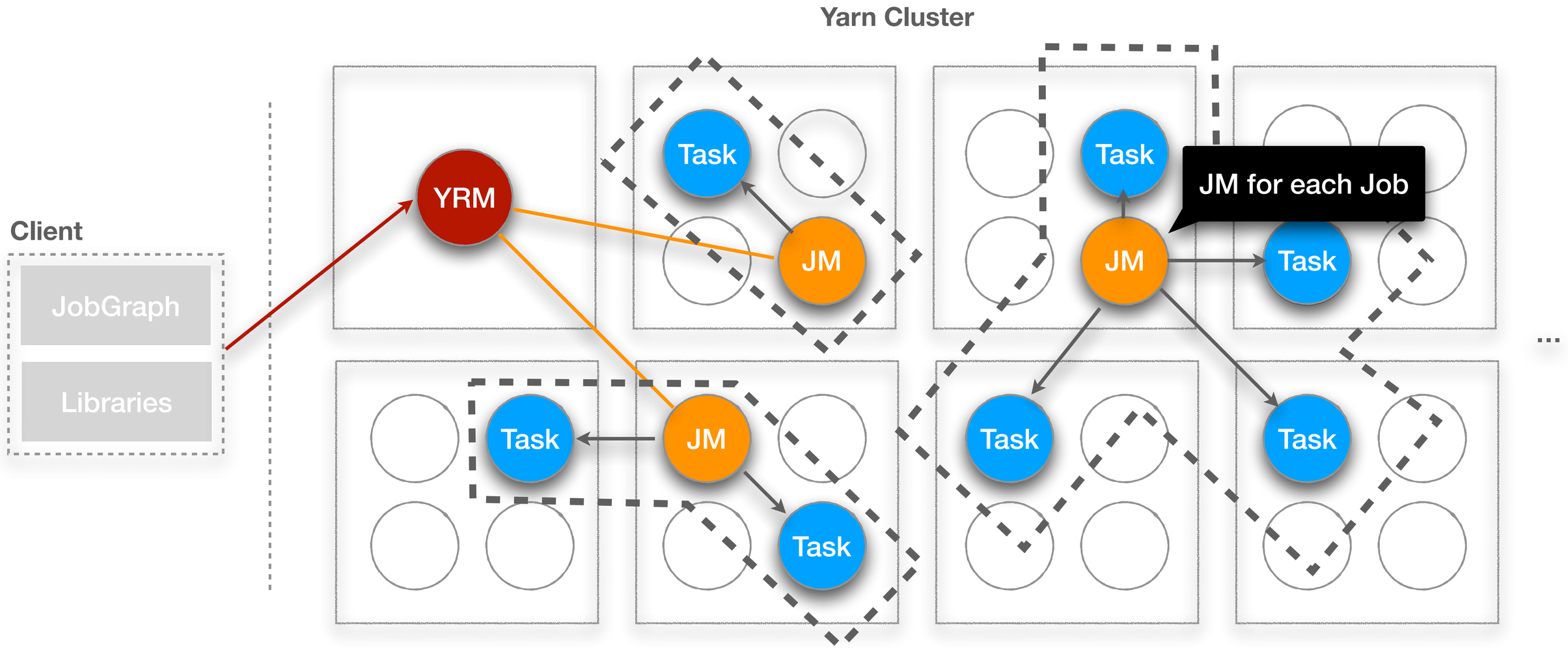
Scalability

Unified


API&SQL


Flink Job
source
source
source
split
split
split
count
count
count
count
sink
sink
sink
sink
Scalability
Complexity


Yarn Cluster
Client
JobGraph
Libraries
YRM
JM for whole cluster
JM
Task
Task
Task
Task
Task
Task
Task
...


Yarn Cluster
Client
JobGraph
Libraries
YRM
JM
Task
Task
JM for each Job
JM
Task
Task
JM
Task
Task
Task
...

Scalability&Complex

**FLIP-6: Deployment and Process Model**

解决大规模部署+云计算隔离问题

**FLIP-12: Asynchronous IO**

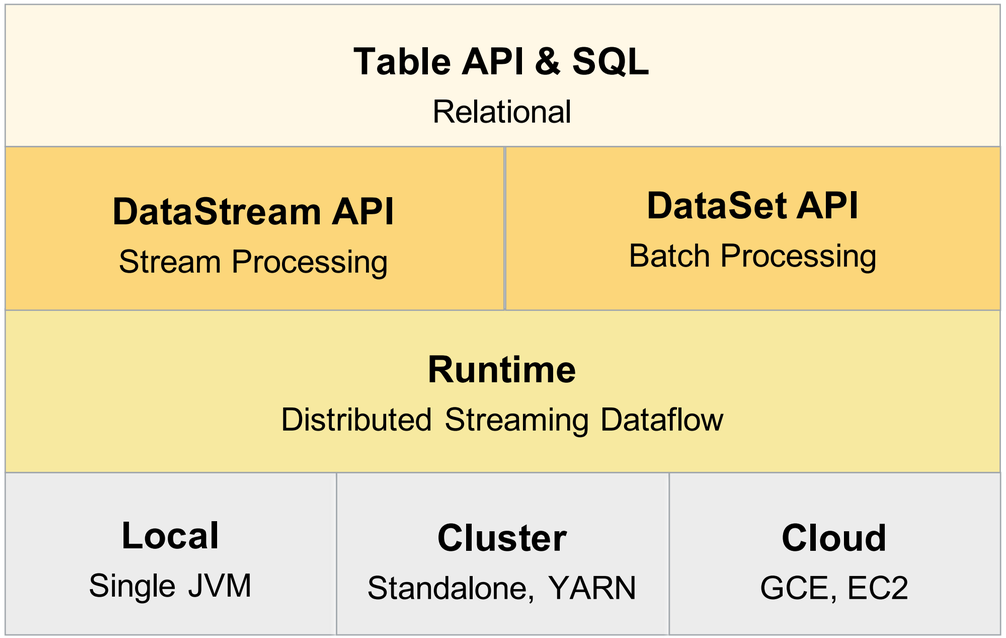
解决流计算时延问题

**Incremental Checkpoints**

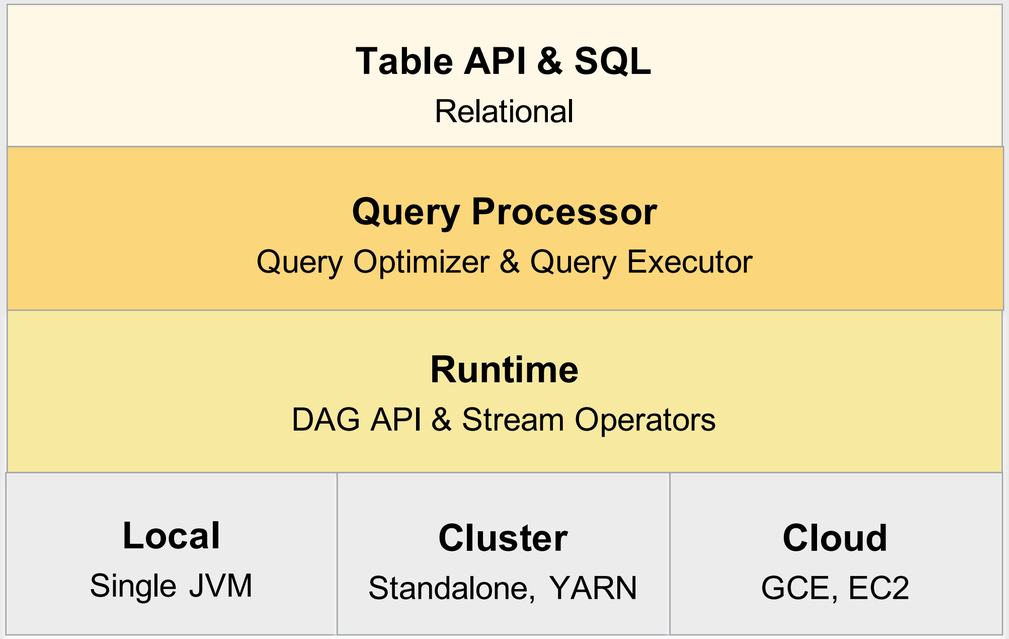
减少CP磁盘IO，解决流计算时延问题

**Dynamic Balance&Local-Global AGG**

减少Data Skew，避免木桶效应



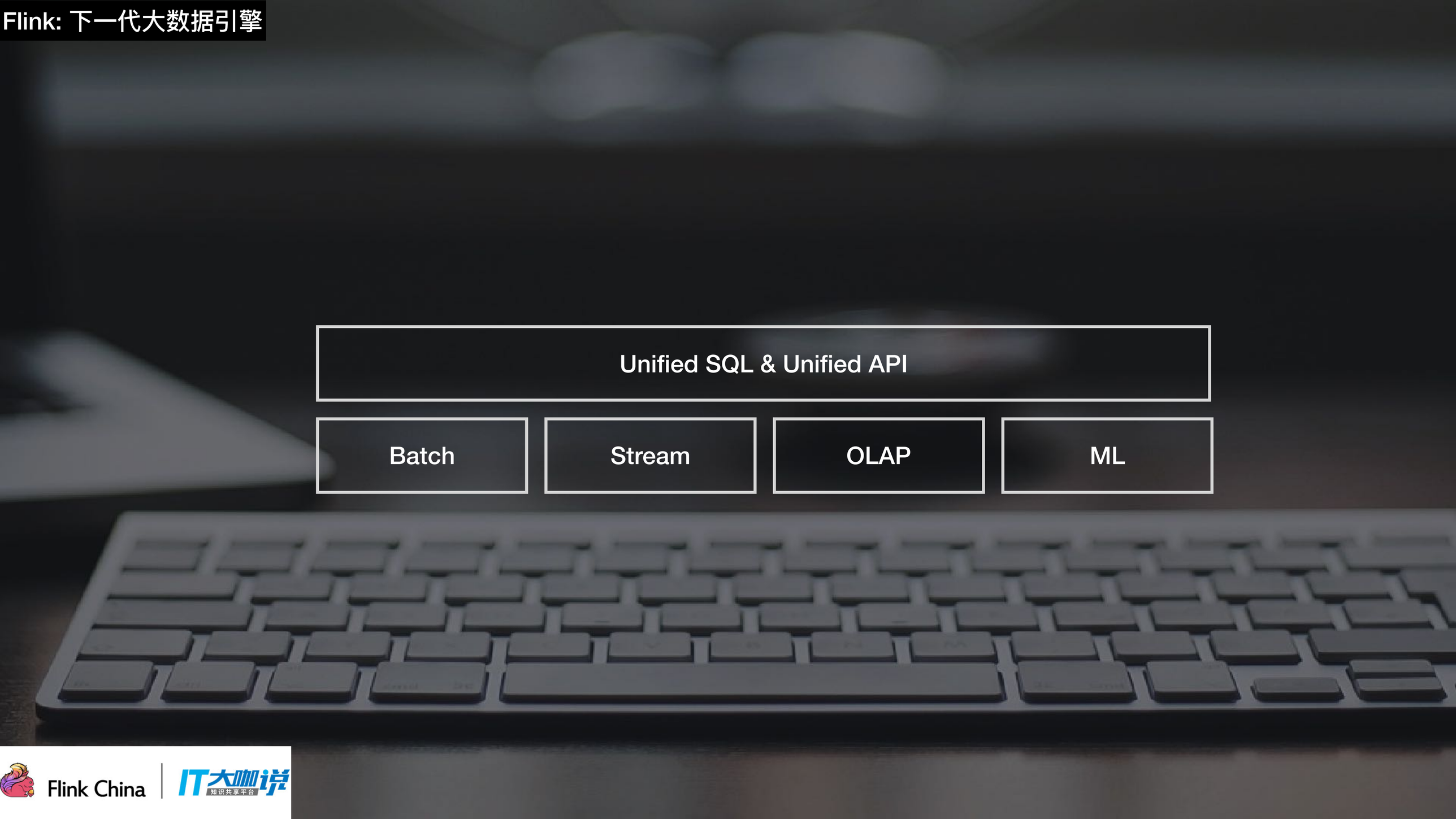

Flink Stack




Blink Stack

Unified Engine Unified API

# Flink: 下一代大数据引擎

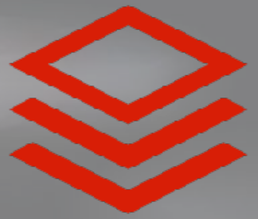


# Flink: 下一代大数据引擎

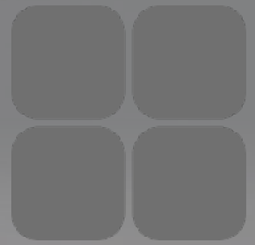
Declarative

Optimizable

Understandable



Stable

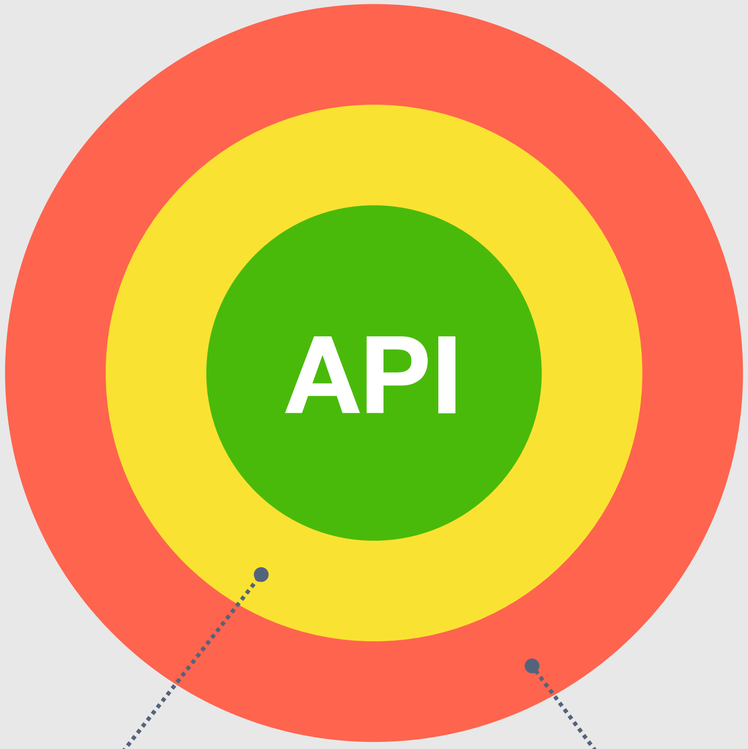


Unify

Just API?


SQL/API Only Solves Part of the Problem

We Provide More!




**Blink Platform**
Development
Debugging
Deployment
WebConsole
…

**Ecosystem**
Metadata Management
Data Quality
…

开源计划

(约)一个季度内Blink分支贡献给Apache Flink

Flink Forward China

**12月20-21日 @北京 国家会议中心**

亚洲区第一场Flink官方大会，预计3000+参会人员

**Flink中文社区大牛&公司悉数到场**

华为、腾讯、滴滴、饿了么、阿里等共同举办

**Call For Keynotes&Sponsors**

召集keynotes&Sponsors

# Thanks