Alibaba Group 阿里巴巴集团

# 大数据和人工智能计算

王绍翾（大沙）

2018.11

2018携程技术峰会 上海

Alibaba Group 阿里巴巴集团

王绍翾（花名“大沙”）
阿里巴巴资深技术专家
shaoxuan.wsx@alibaba-inc.com

北京大学
EECS

美国加州大学圣地亚哥分校
Computer Engineering

博通（Broadcom）
High-Perf Platform

脸书（Facebook）
Social Graph Storage

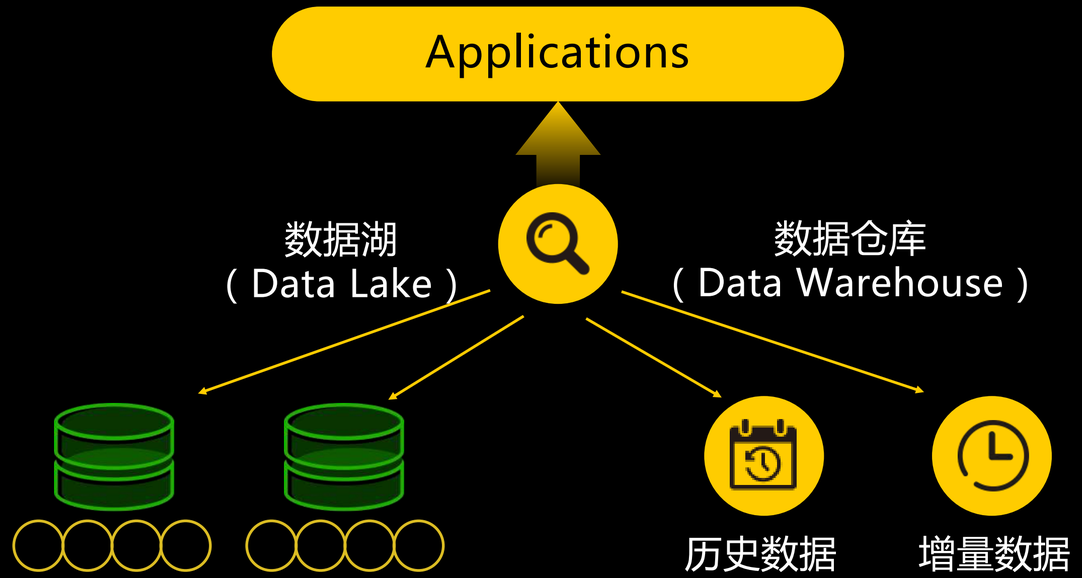
阿里巴巴
Real-Time Data Infra

负责大数据实时计算平台和算法平台

# 大数据计算的类型

批计算

Changing Query
Fixed Data



# 大数据计算的类型

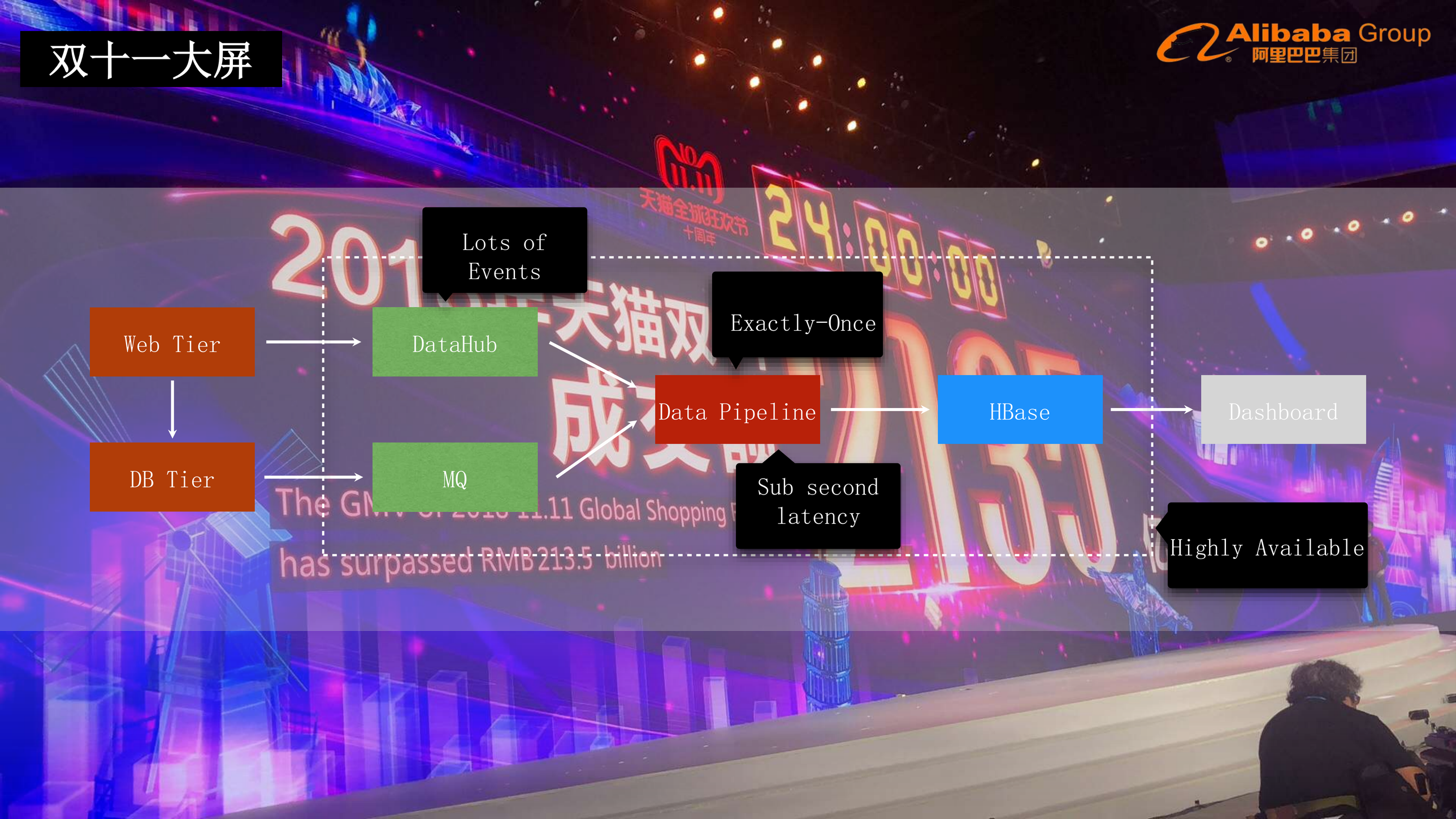
批计算

Changing Query
Fixed Data

流计算

Fixed Query
Changing Data

# 双十一大屏


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

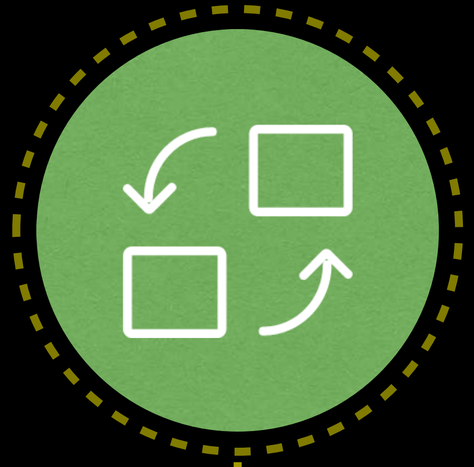
# 大数据计算的类型

批计算

Changing Query
Fixed Data

流计算

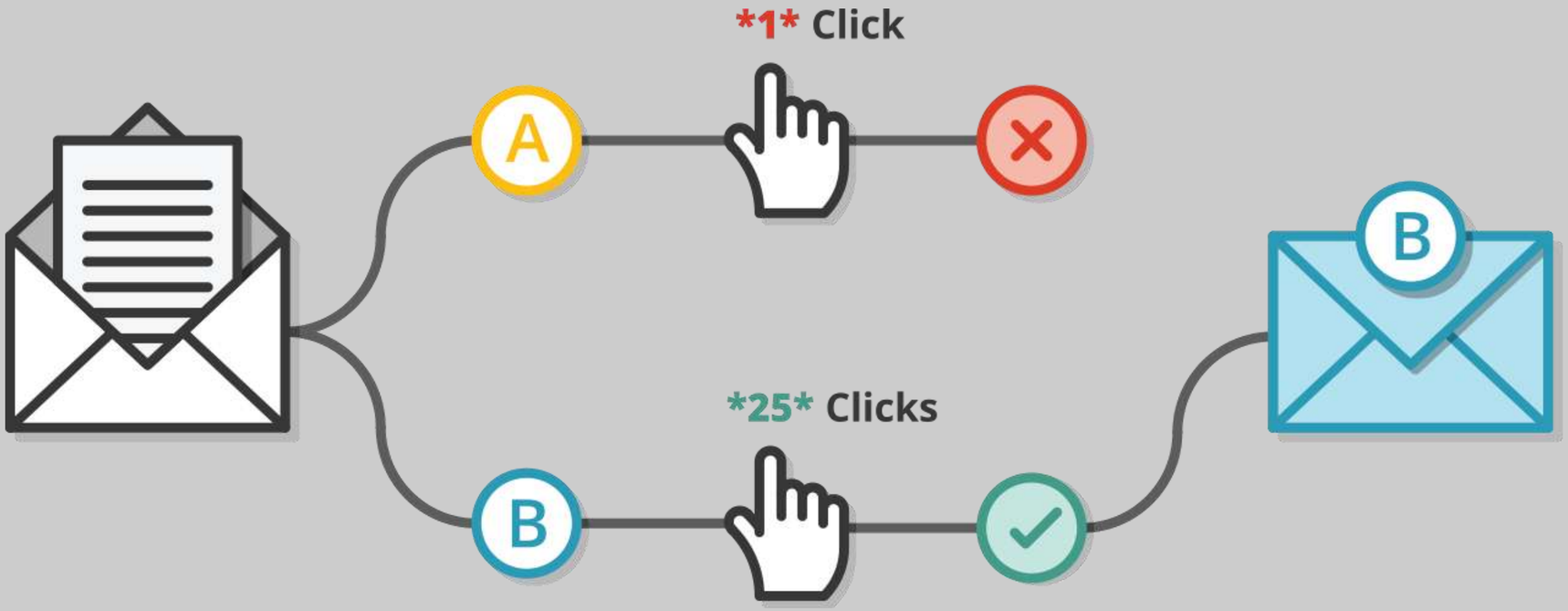
Fixed Query
Changing Data



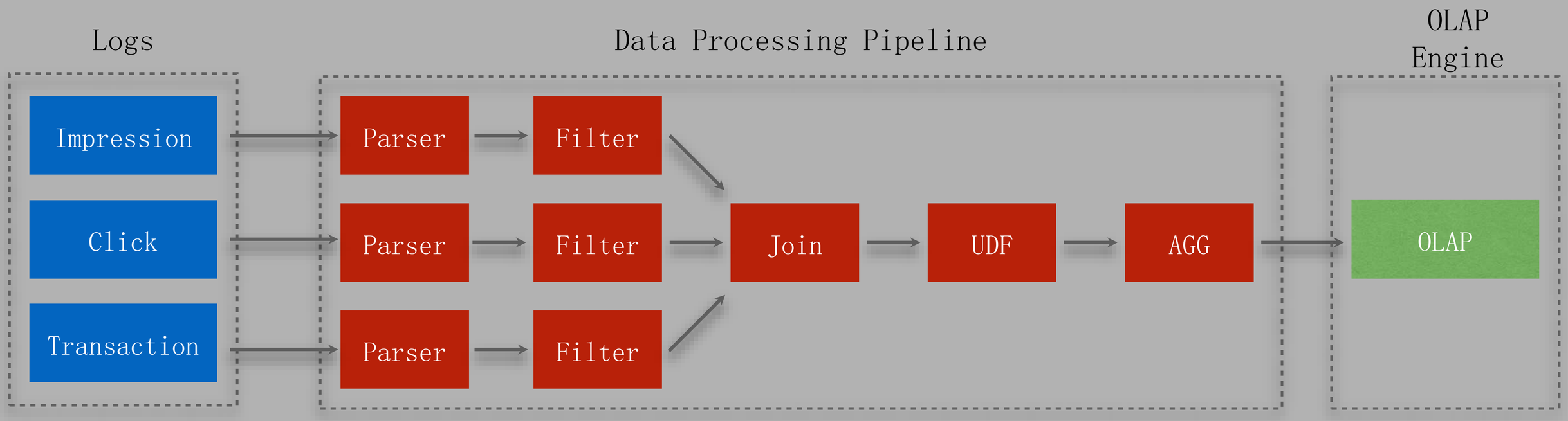
交互式分析

Changing Query
Changing Data

# A/B Testing



# A/B Testing


Logs
Data Processing Pipeline
OLAP Engine
Impression
Click
Transaction
Parser
Parser
Parser
Filter
Filter
Filter
Join
UDF
AGG
OLAP

# 大数据计算的应用场景


数据计算 Data Lake/Warehouse Platform

提供完整的数据湖 / 数仓解决方案

Data Lake/Warehouse Solutions

工业大脑 ET Industrial Brain

工业大数据实时监控&预测赋能工业大脑解决方案

Industrial big data real time monitoring and prediction

Empower ET Industrial Brain Solutions

城市大脑 ET City Brain

支持视频流解析&结构化处理赋能城市大脑数据处理

Support video stream and structuring process

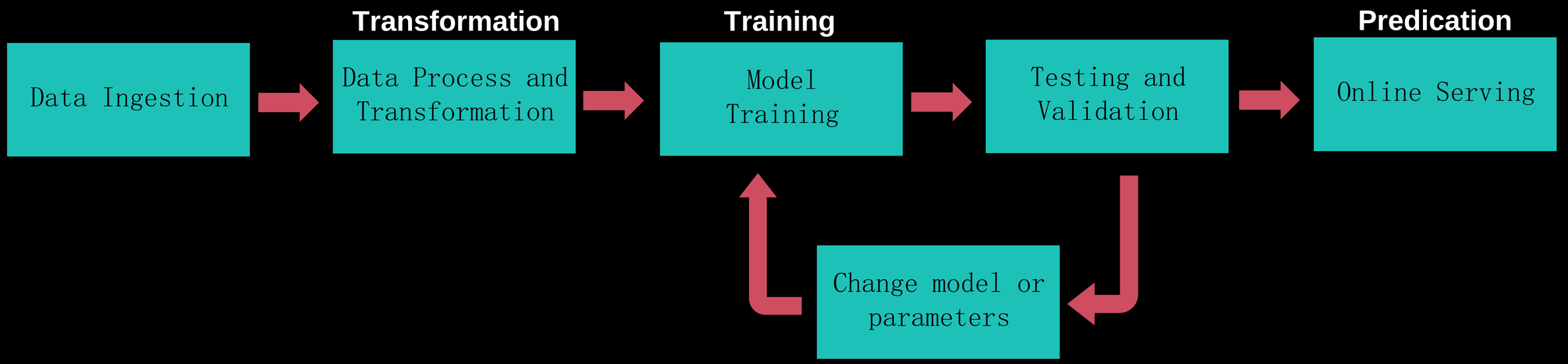
Empower ET City Brain data processing

IOT

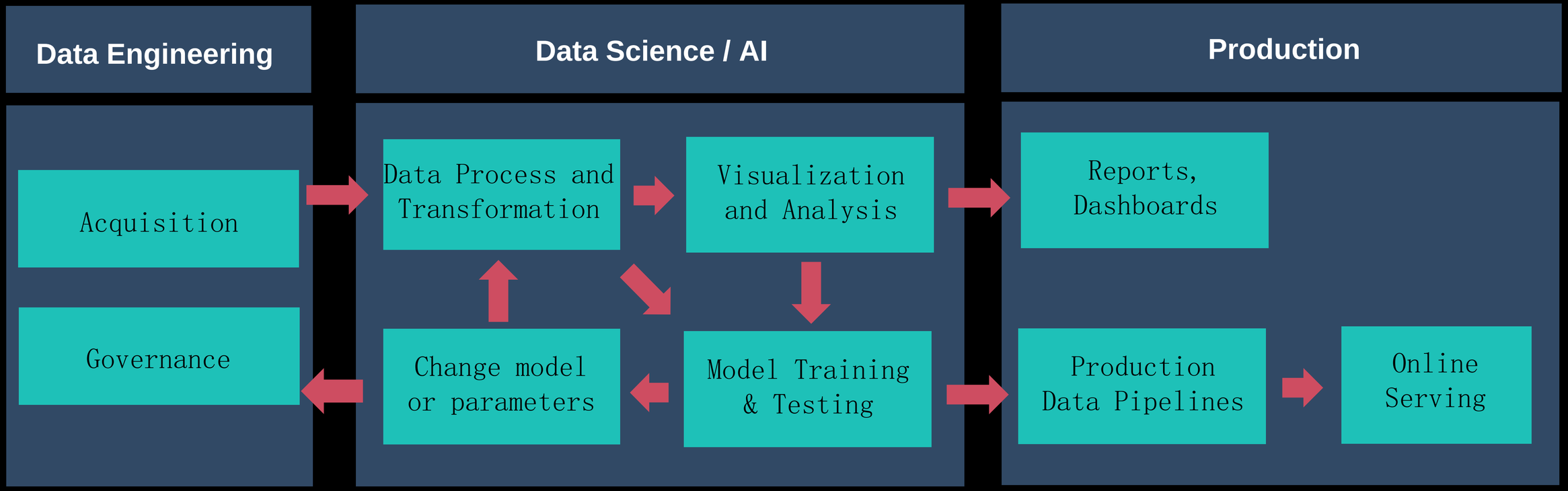
支持云上IOT数据处理实现云端计算一体化

Support IOT data process on Cloud

# 人工智能计算


Transformation
Training
Predication
Data Ingestion
Data Process and Transformation
Model Training
Testing and Validation
Online Serving
Change model or parameters

# 人工智能计算


Data Engineering
Data Science / AI
Production
Acquisition
Governance
Data Process and Transformation
Visualization and Analysis
Change model or parameters
Model Training & Testing
Reports, Dashboards
Production Data Pipelines
Online Serving

ALIYUN · 阿里云大数据服务

# 杭州大脑 交通数据实时监控

Traffic Data Real-Time Monitoring

10-28 19:21:05

30分钟内统计： 9

## 车辆监控
Vehicle Monitoring

| 视频源01 | 视频源10 | 视频源12 | 视频源23 |
|---|---|---|---|
| 浙A3002R | 浙AT1225 | 浙A61EFF | |
| 浙J3SD23 | 浙C6JF40 | 浙G7999M | |
| 浙A0M608 | 浙AZ1369 | 蒙JC1768 | |
| 浙AES048 | 浙A22ZV5 | 鄂A1Z501 | |
| 浙AM839A | 浙GX1886 | 浙ADD772 | |

## 地域信息
Area Information

## 流量监控
Traffic Volume Monitoring



## 监控数据
Vehicle Information

| | | 进入时间 | 进入位置 | 离开时间 | 离开位置 |
|---|---|---|---|---|---|
| 12 | 浙AB[illegible] | [illegible]:21 | [x:10,y:974,w:313,h:625] | 19:31:38 | [x:616,y:219,w:185,h:240] |
| 23 | 浙AHF320 | 18:17:25 | [x:341,y:1298,w:572,h:295] | 18:17:26 | [x:1027,y:572,w:170,h:222] |
| 10 | 浙AT1225 | 19:28:03 | [x:908,y:190,w:617,h:392] | 19:28:05 | [x:1524,y:919,w:383,h:266] |
| 10 | 浙AZ1369 | 19:27:49 | [x:715,y:2,w:288,h:128] | 19:27:52 | [x:1484,y:844,w:435,h:338] |
| [illegible] | [illegible] | [illegible]:27 | [illegible] | [illegible]:16 | [illegible] |

## 城市大脑

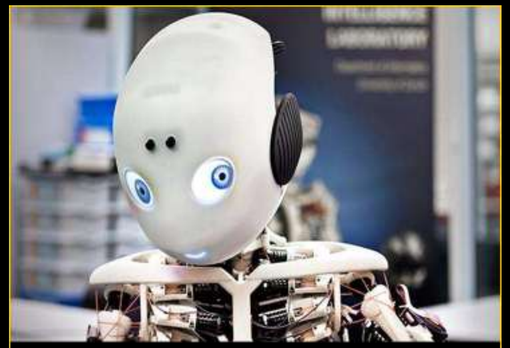
**杭州城市大脑使用阿里云流计算，实时监控交通状况，通过智能调度道路红绿灯、实时更新导航系统，优化城市道路拥堵情况。**


## 传媒视频广告植入效果分析
Intelligent access control solution

使用多模态AI技术，实时分析视频植入广告的露出节点和效果

## 智能客服解决方案
Intelligent customer service solution

辅助坐席客服人员了解客户意图，提供金牌话术应答，促进销售成单



## 外贸客服实时同声传译
Real-time simultaneous interpretation of foreign trade customer service

不同语言的语音和文本实时互译，比如外贸客服、出境旅游

## 传媒视频直播解决方案
Live broadcast solution

实时监控敏感图像，给出等级区分，准确率高达99.9%

## 舆情监控与分析
Live broadcast solution

全方位分析互联网舆论，实时监控舆论动态，追溯事件脉络

## 媒体内容安全解决方案
Smart metro solution

图片、视频，文字等多媒体的内容风险智能识别服务

## 传媒视频多模态关键人物识别
Key character recognition in video

使用多模态AI技术，可以快速的锁定目标人物出现的时间轴

## 智慧地铁解决方案
Smart metro solution

首个AI地铁之城，买票动动嘴、闸机能看脸、摄像头会数数


## 智能门禁/闸机解决方案
Public opinion monitoring and analysis

软硬件结合门禁方案，提升人员系统化管理的安全性与便捷性

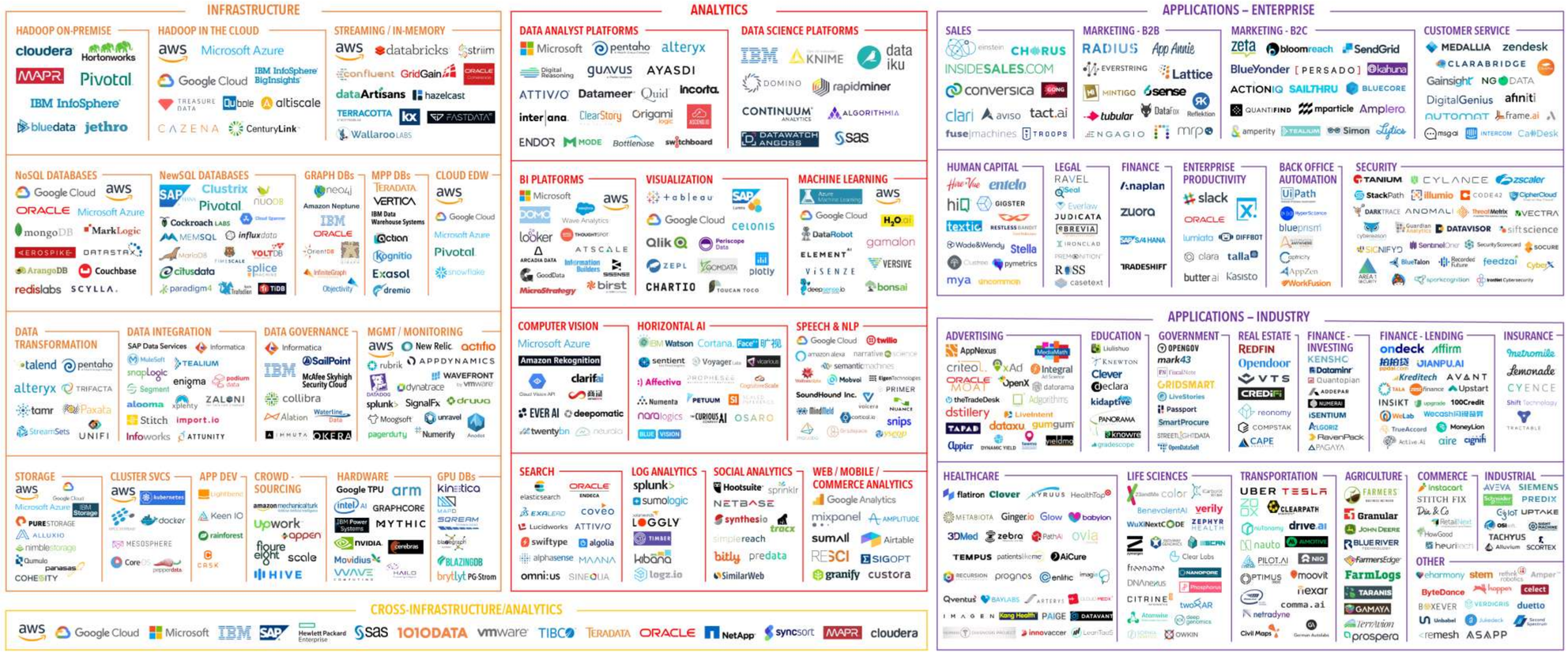
## 智能司法解决方案
Intelligent judicial solution

将语音识别、OCR文本识别、同案类推和司法大数据融合

# 大数据智能计算


Business Intelligence
Artificial Intelligence
Big Data Infrastructure

# BIG DATA & AI LANDSCAPE 2018

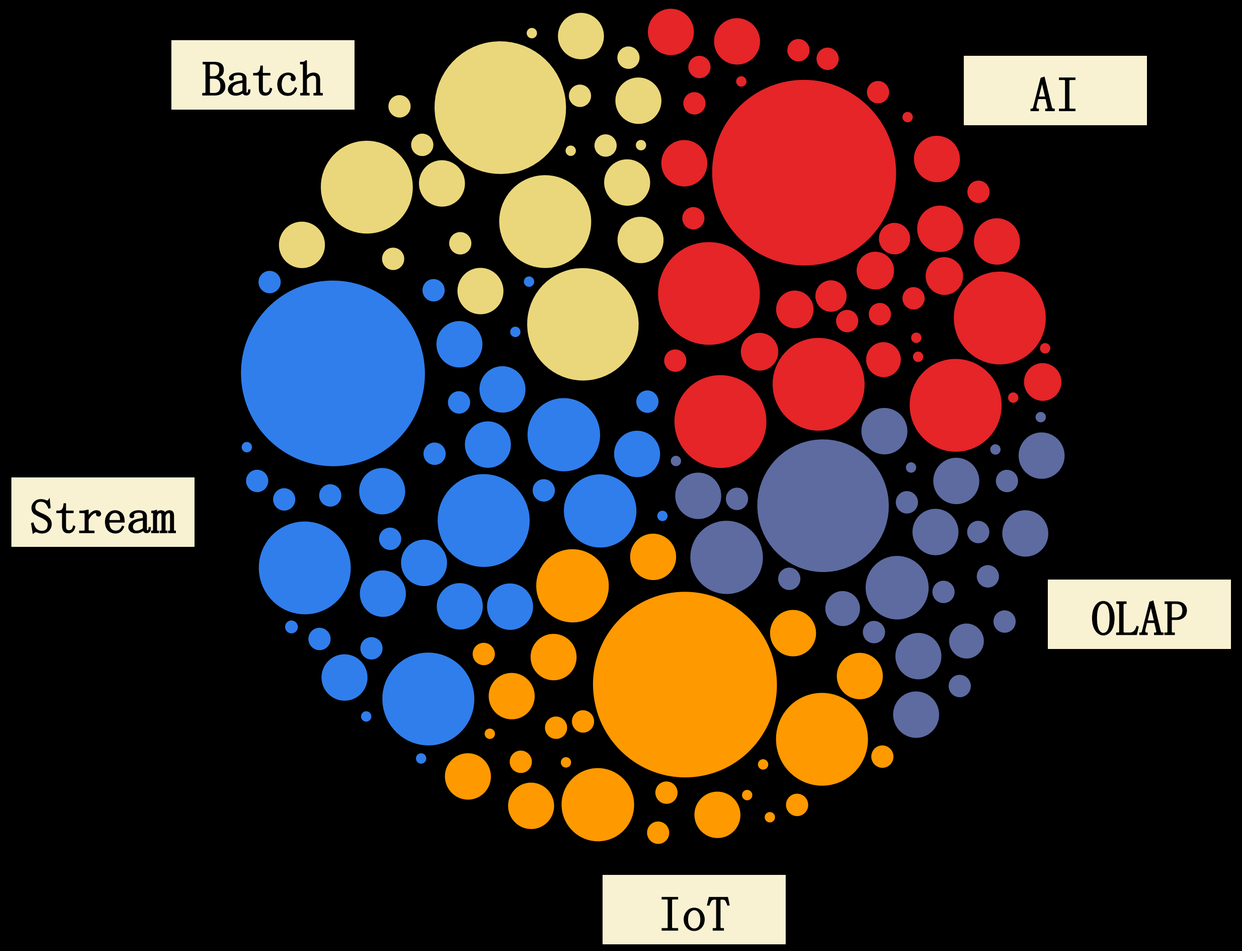






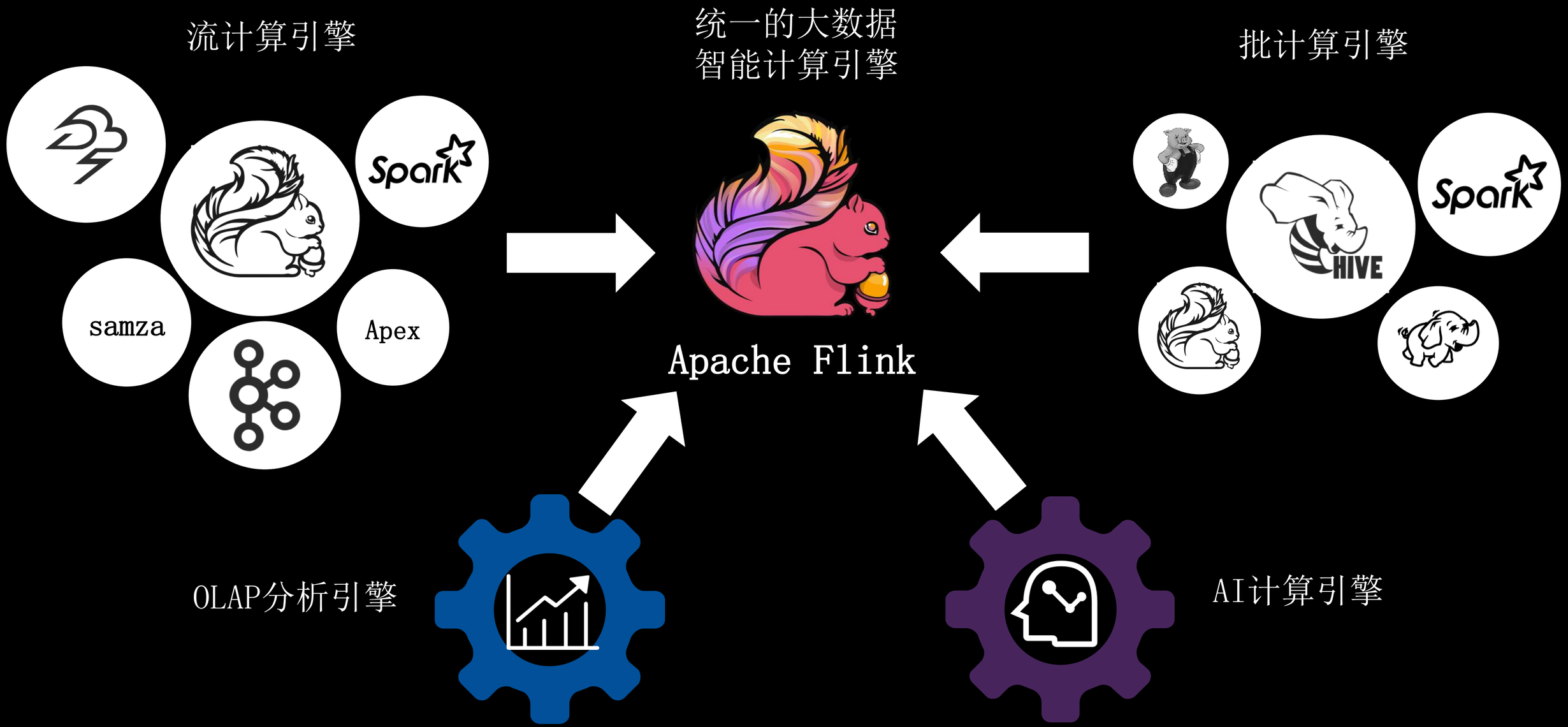

FIRSTMARK
EARLY STAGE VENTURE CAPITAL

# BigData & AI




能否用一套计算引擎解决所有问题

# 用Apache Flink构建大数据智能平台



# 阿里巴巴Blink

Apache Flink

Alibaba’s Improvements

**Blink1.0:**
high performance stream compute engine

**Blink2.0:**
a new unified high performance compute engine for complete data applications (including batch, stream, AI, and IoT)

# Blink1.0

Alibaba Group 阿里巴巴集团

## 天然的Flink纯流式计算的低延迟

## 大规模高并发部署的优化 2016

贡献社区

## 快速的容错

贡献社区

- Incremental Checkpoint 2016
- Fain-grained Recovery 2016
- Barrier Alignment Improvement 2017

## 性能提升

部分贡献社区

- Async Operator 2016
- Credit Based Flow Control 2017
- Load Auto Balance 2017

## 主导制定 Flink SQL 语义

贡献社区

- Dynamic Table 2016-2017
- Retraction 2016-2017

## 完善 Flink SQL 功能

贡献社区

- Aggregation，Join，Window 2017
- 跑通全部TPCH/TPCDS Query 2018

## 性能提升

部分贡献社区

- 大量的Query Optimization 2017-2018

## 资源配置自动化 2018

# Blink2.0: unified compute engine for complete data applications

Alibaba Group 阿里巴巴集团

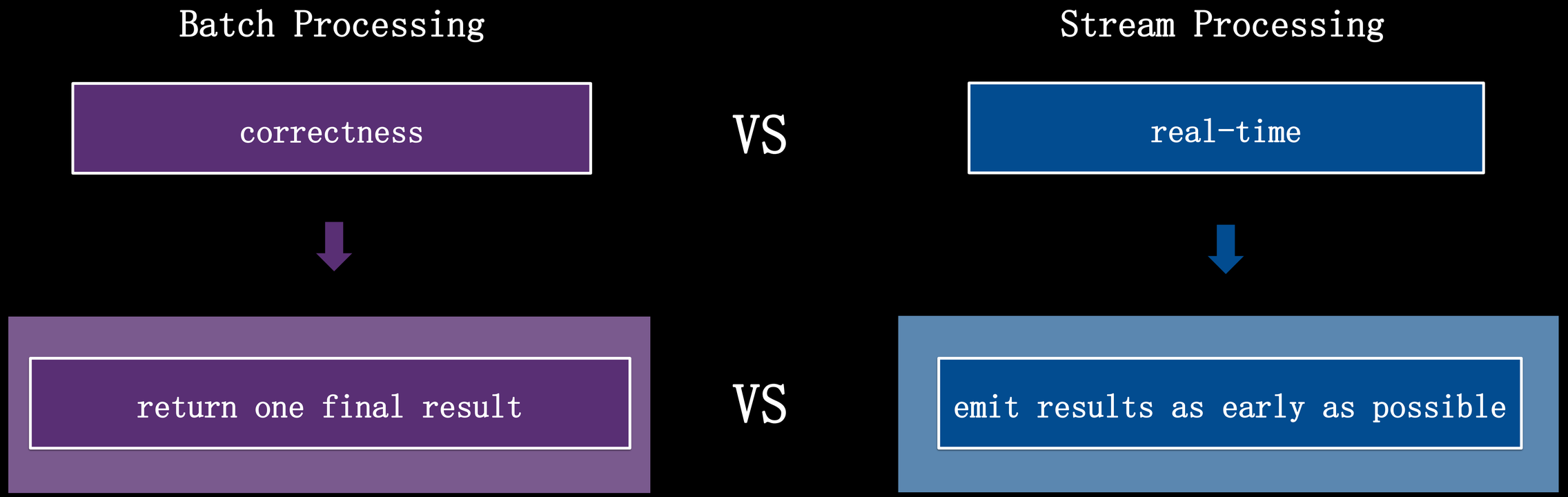
API的统一:
same query, same results

Query Processing的统一:
unified query optimization and query execution framework

应用的统一:
switch between batch processing and streaming seamlessly

# 批计算和流计算API的统一

Alibaba Group 阿里巴巴集团




in stream processing, it emits intermediate results, and keeps refining the results to ensure correctness

# 批计算和流计算API的统一

Alibaba Group 阿里巴巴集团

WHAT & HOW: results are calculated ----- Can be fully described by SQL

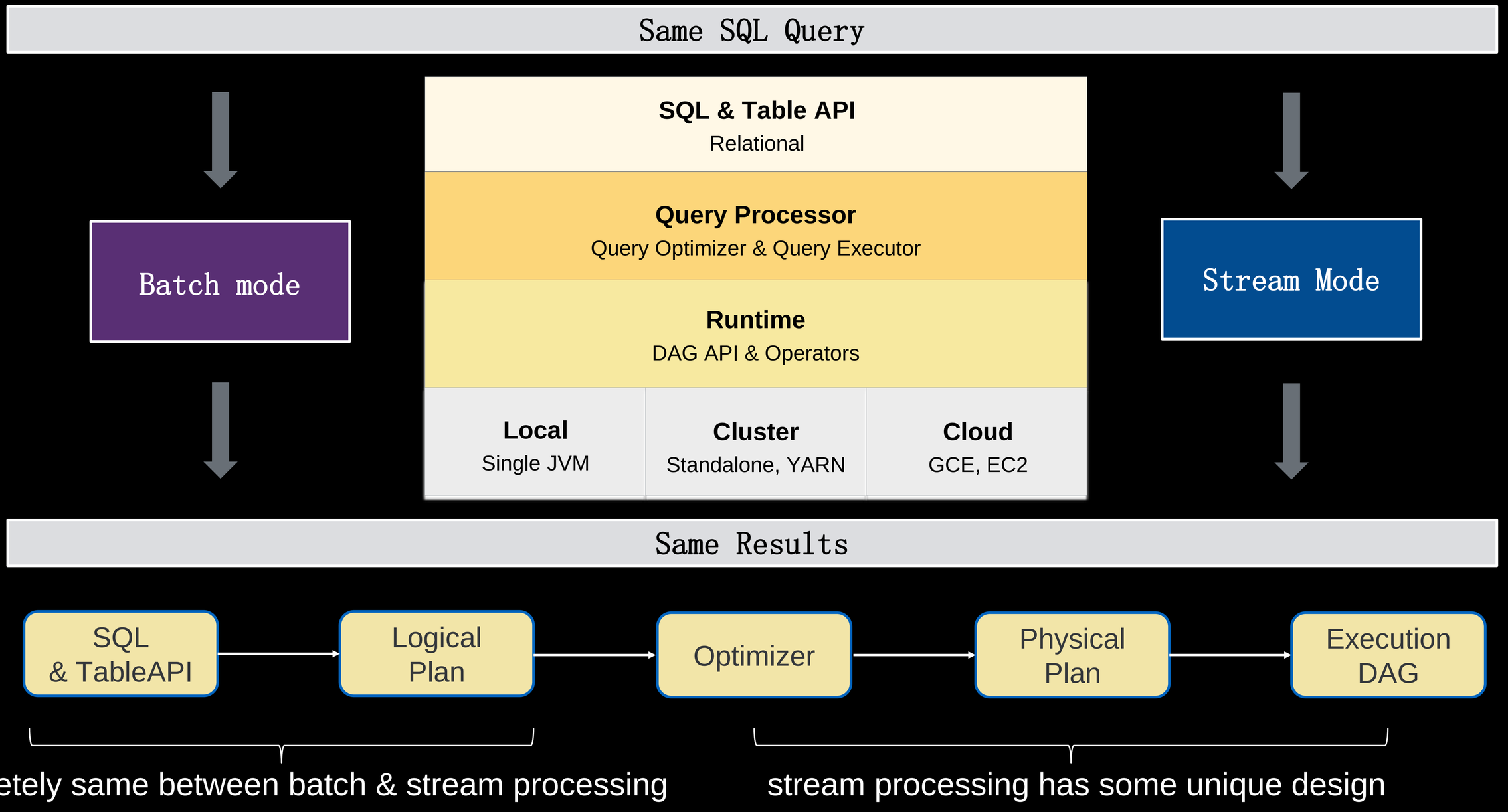
WHEN: to emit a (intermedia) result ----- Does not affect business logic

HOW: to refine the results ----- Can be solved by SQL engine

**ANSI SQL can Describe Stream Processing**

**可以用SQL统一流和批的计算**

# Blink2.0 统一了SQL Engine的QP



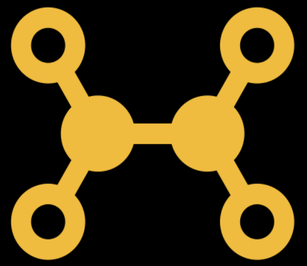
# Optimizations for Batch Processing in Blink2.0


Execution Optimizations

Query Optimizations

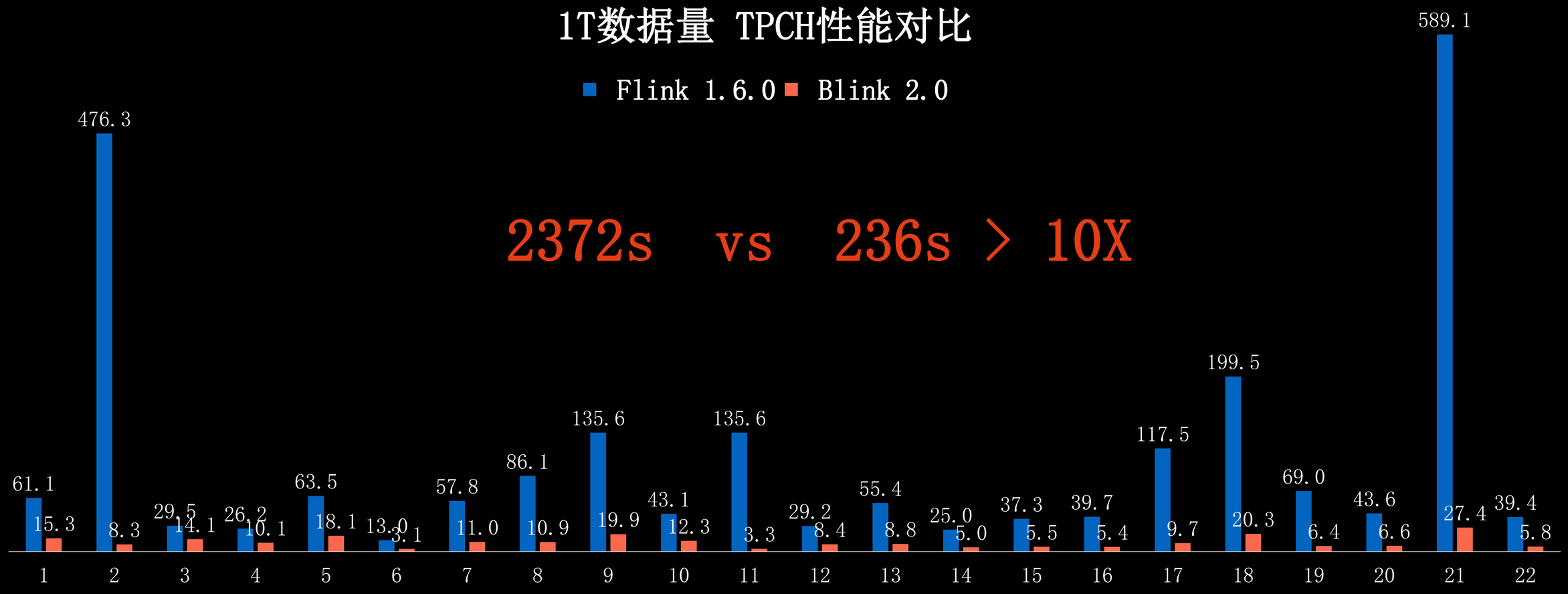
Customizable Job Scheduling

Batch Friendly Failover



Various Shuffle Service

# Optimizations for Batch Processing in Blink2.0



Next Steps

## Grand Unification of Data Processing

Switch between batch processing and streaming seamlessly

## Flink Machine Learning/AI

PyFlink, TableAPI, TensorFlow, Julia, Flink ML Improvements

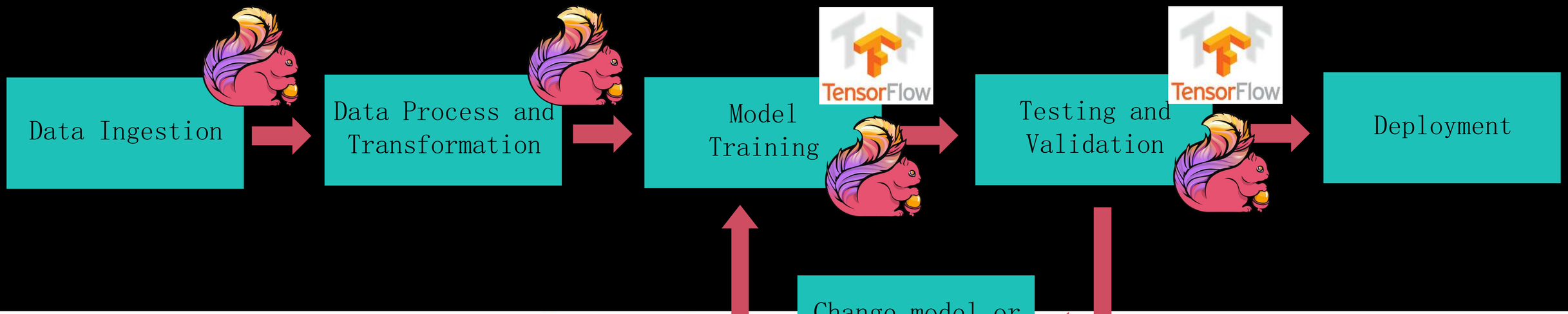


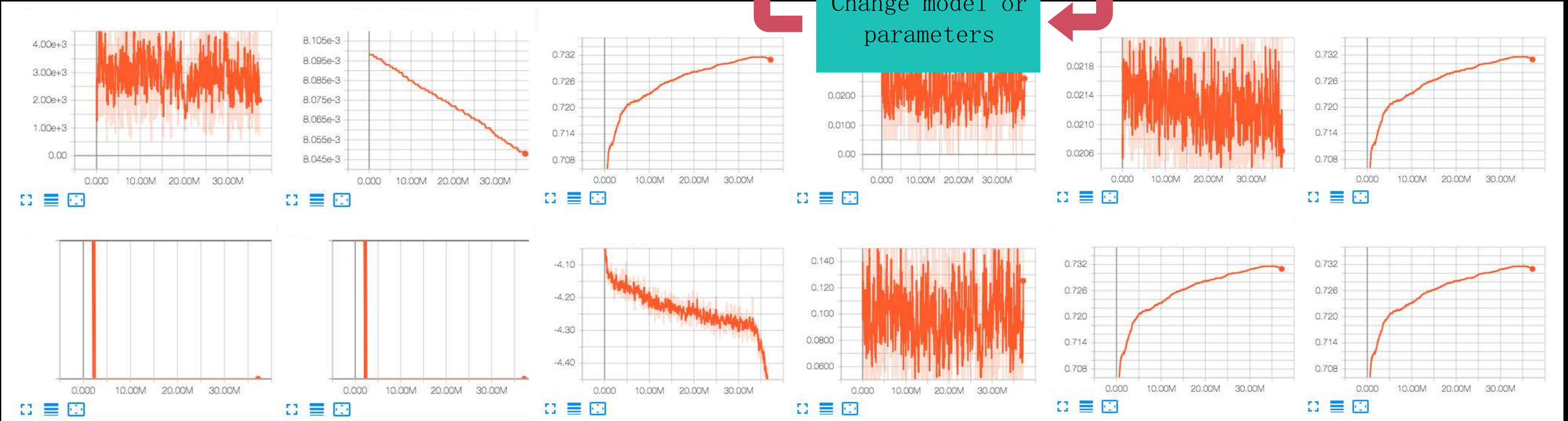

## Ecosystem

Hive, Zeppelin, Jupyter, Livy

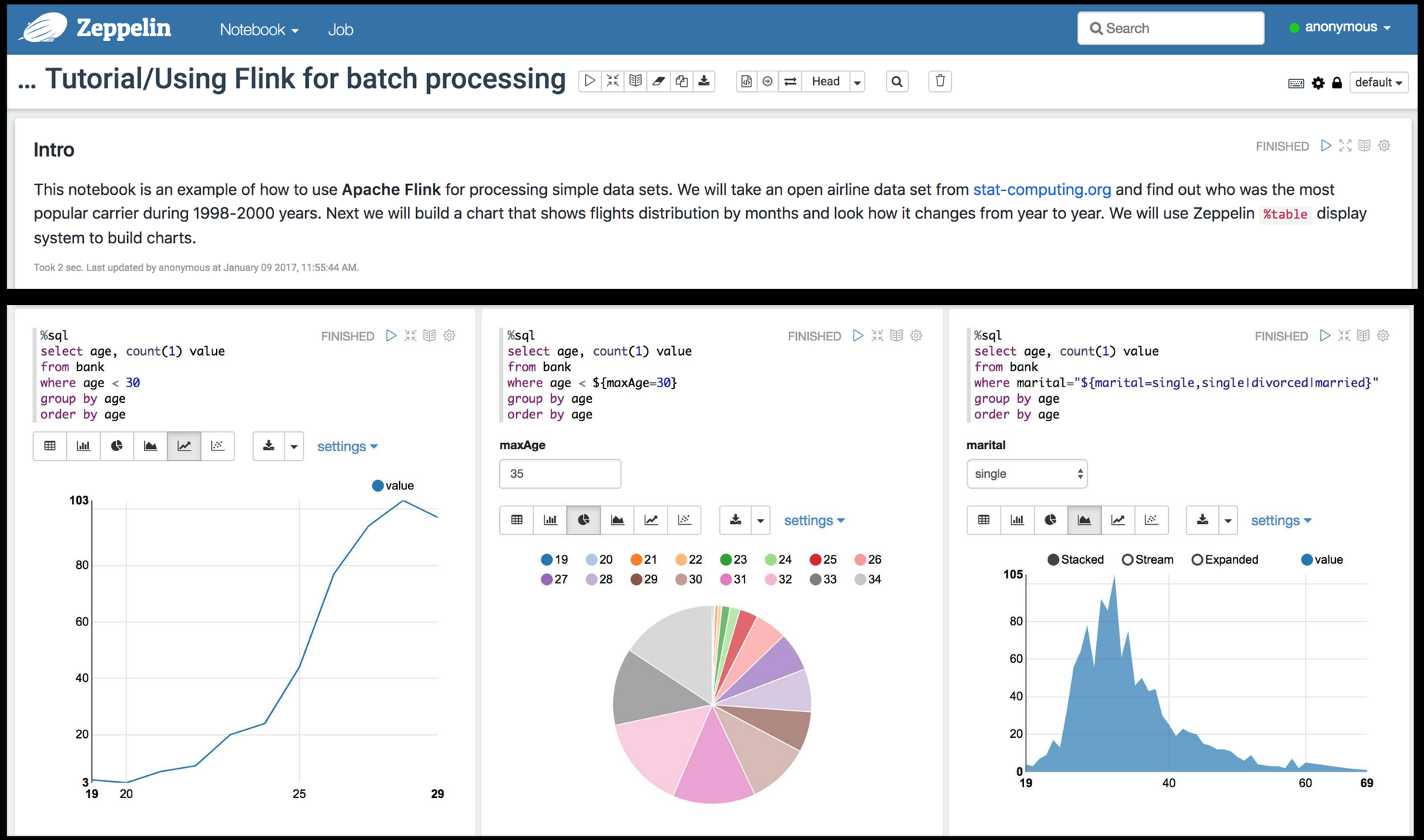

# TensorFlow on Flink

Alibaba Group 阿里巴巴集团



# Flink on Zeppelin



Flink Forward China - 北京 - 12.20

Alibaba Group 阿里巴巴集团

FLINK FORWARD

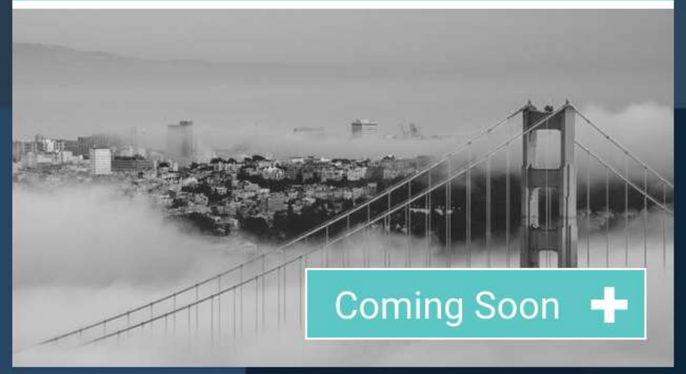
Home | Location

# The Apache Flink® Conference

Stream Processing | Event Driven | Real Time

**Flink Forward - Beijing**
December 20-21, 2018


**Flink Forward - San Francisco**
April 1-2, 2019




Flink China社区大群

扫一扫群二维码，立刻加入该群。

Alibaba Group
阿里巴巴集团

本PPT来自2018携程技术峰会

更多技术干货，请关注“携程技术中心”微信公众号