Flink Forward China 2018

公司：美团点评

职位：研究员

演讲者：鞠大升

Flink Forward China 2018

基于Flink的美团点评实时计算平台实践和应用

The Practice and Application of MTDP’s Realtime Compute Platform on Flink

美团点评 鞠大升 2018-12-10

Outline

- 介绍 Introduction
- 平台建设实践 Practice in Platform Construction
- 实时应用 The Realtime Applications
- 挑战&未来 Challenges and the Future Work

Outline

- 介绍 Introduction
- 平台建设实践 Practice in Platform Construction
- 实时应用 The Realtime Applications
- 挑战&未来 Challenges and the Future Work

关于我们 About us

美团

大众点评

美团外卖

猫眼电影

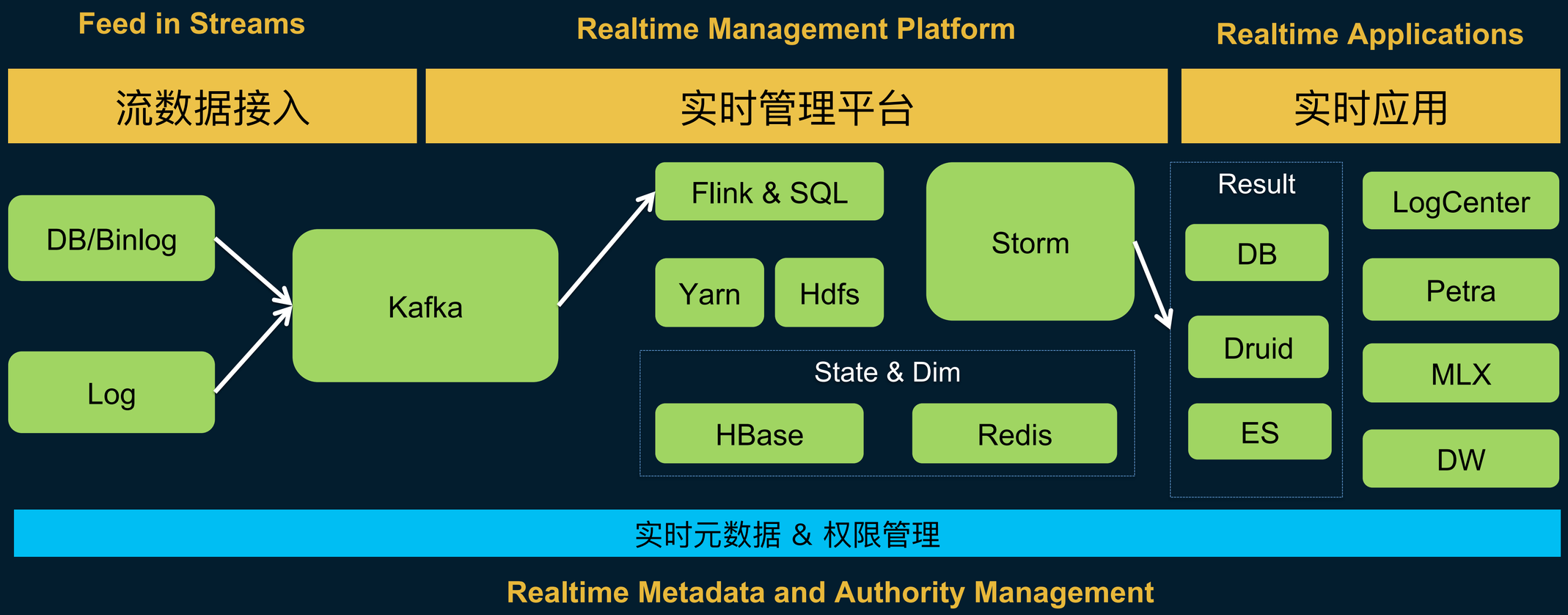
美团点评是中国领先的生活服务电子商务平台

MeituanDianping is China’s leading life service e-commerce platform

业务特点 Business Characteristics

- 多业务线、形态各异 **Multiple Service Lines with Heterogeneous Patterns**
- 涉及交易、链路长 **Involve Long Transaction Chains**
- 业务协同需求强 **Require Strong Business Collaboration**

平台架构 The Platform Architecture

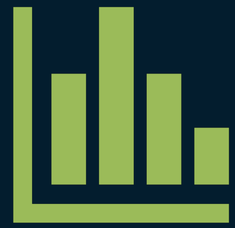


平台现状 The Current Status

10 thousand
Jobs

4 thousand
Machines

1000 billion
Messages/Day



10 million
Peak Messages/s

应用场景 Application Scenarios

- 风控 & 反爬虫 — **Risk Control and Anti-Crawling**
- 实时流量分析 — **Traffic Analysis in Realtime**
- 业务监控 — **Business Monitoring**
- B端应用 — **Browser Applications**
- 运营分析 — **Operations Analysis**

Outline

- 介绍 Introduction
- 平台建设实践 Practice in Platform Construction
- 实时应用 The Realtime Applications
- 挑战&未来 Challenges and the Future Work

我们关注什么？ What we cares

- 引擎能力（精确计算、状态管理）
- 平台化（多租户、资源、权限）
- 效率（开发、调试、问题追查、调优、SQL）
- 高可靠（容灾、运维）
- 场景化应用（日志中心、Petra、MLX）

- **Capabilities of the Engine (Precise Calculation and State Management)**
- **Platformization (Multi-Tenant, Resources and Authorities)**
- **Efficiency (Development, Debugging, Tracing, Tuning and SQL)**
- **High Availability (Disaster Tolerance and Maintenance)**
- **Scenario Applications (Log Center, Petra and MLX)**

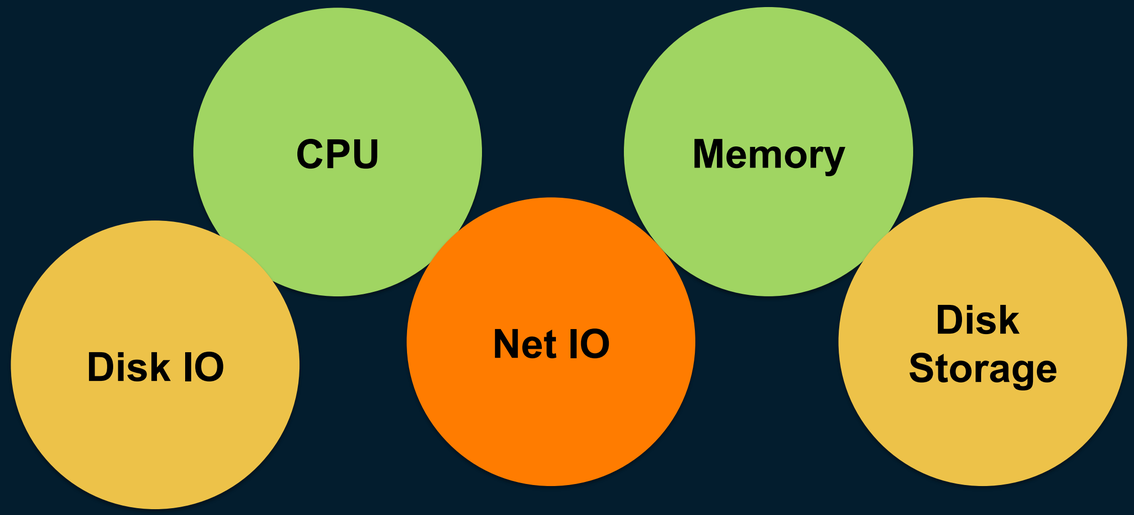
稳定性建设 How to be Stable

- 资源隔离
  - 离线机群/实时机群 — 物理隔离部署
  - 不同业务线 — Yarn标签隔离

Resource Isolation

The Cluster for Offline Jobs / The Cluster for Realtime Jobs - Physical Isolation

Different Service Lines –Label-Based Resource Isolation in Yarn



稳定性建设 How to be Stable

- 故障容灾
  - Job Manager HA
  - 作业自动拉起
  - Flink Kafka异常重试
  - 多机房容灾
  - 流热备

Fault Tolerance

Job Auto-Reboot

Retry on Exception for Flink Kafka

Multi-Datacenter for Disaster Recovery

Hot Standby for Streaming Systems

稳定性建设 How to be Stable

- 监控报警
  - 作业状态报警
  - 处理延迟报警
  - 自定义Metrics报警

Alarm Monitoring

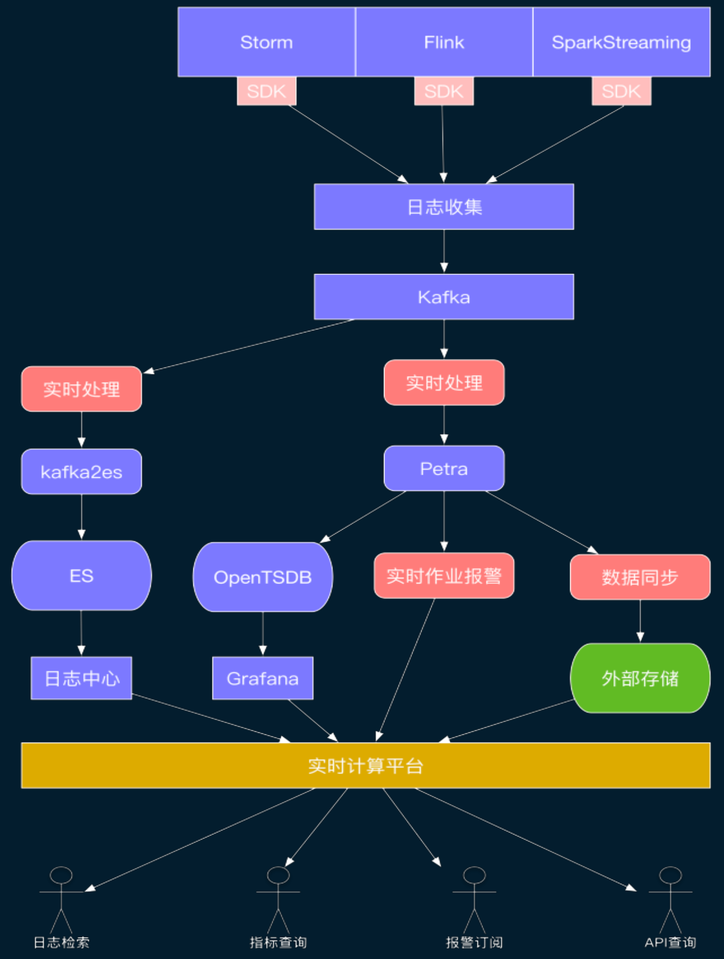
Job Status Alarm

Processing Delay Alarm

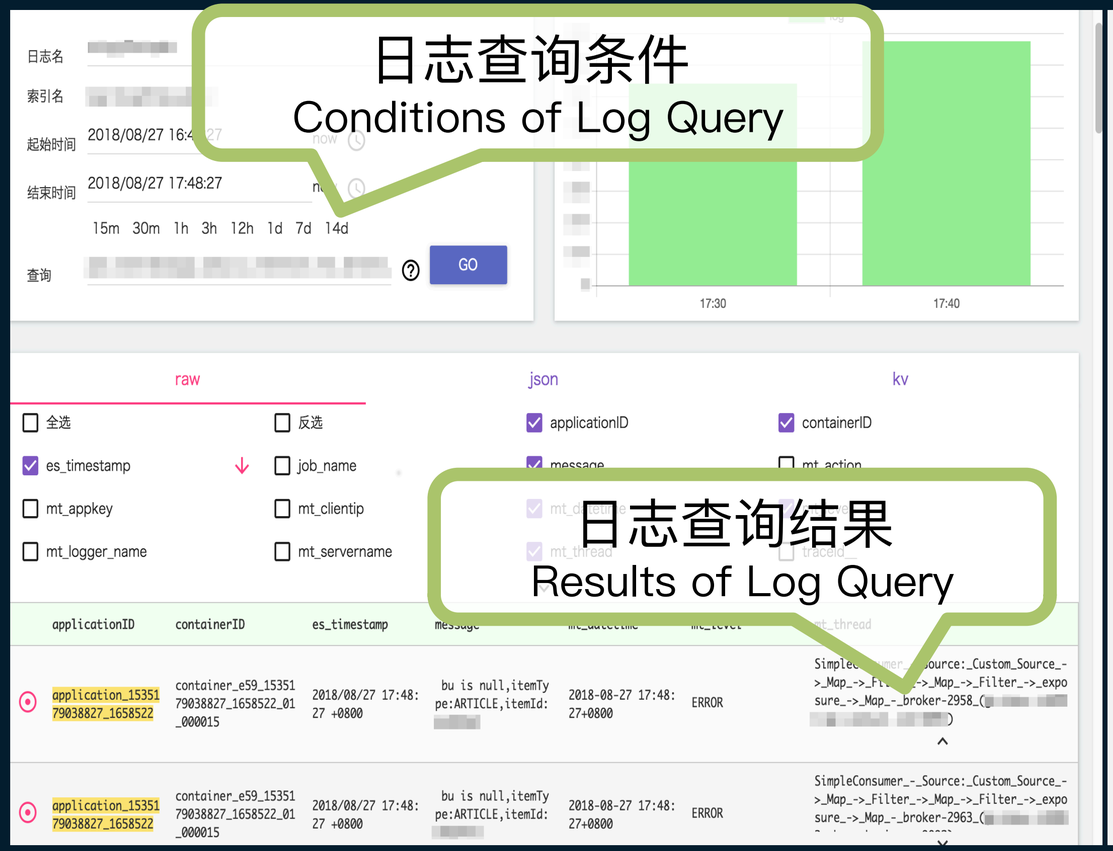
Custom Metrics Alarm

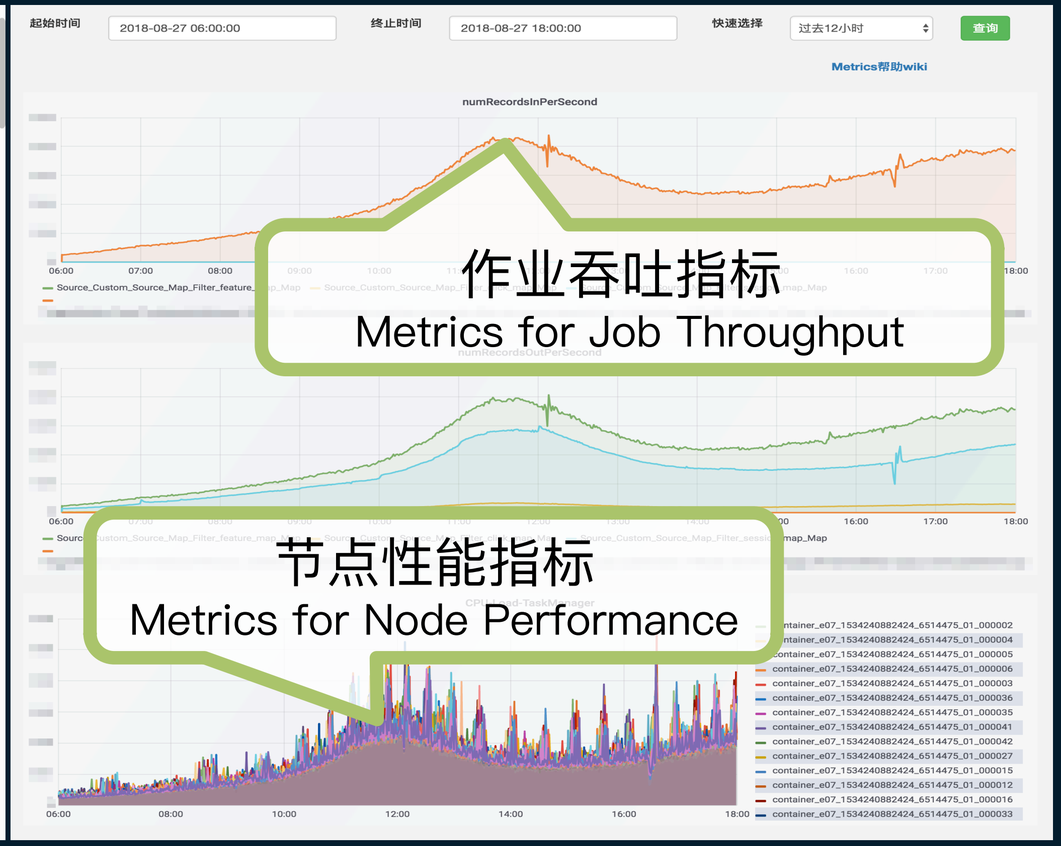
调优诊断 Tuning & Debugging

- 统一的日志收集和检索
  Unified Log Collection and Retrieval
- 统一指标收集和查询
  Unified Metrics Collection and Querying
- 基于指标的可配置报警
  Configurable Alarm on Metrics

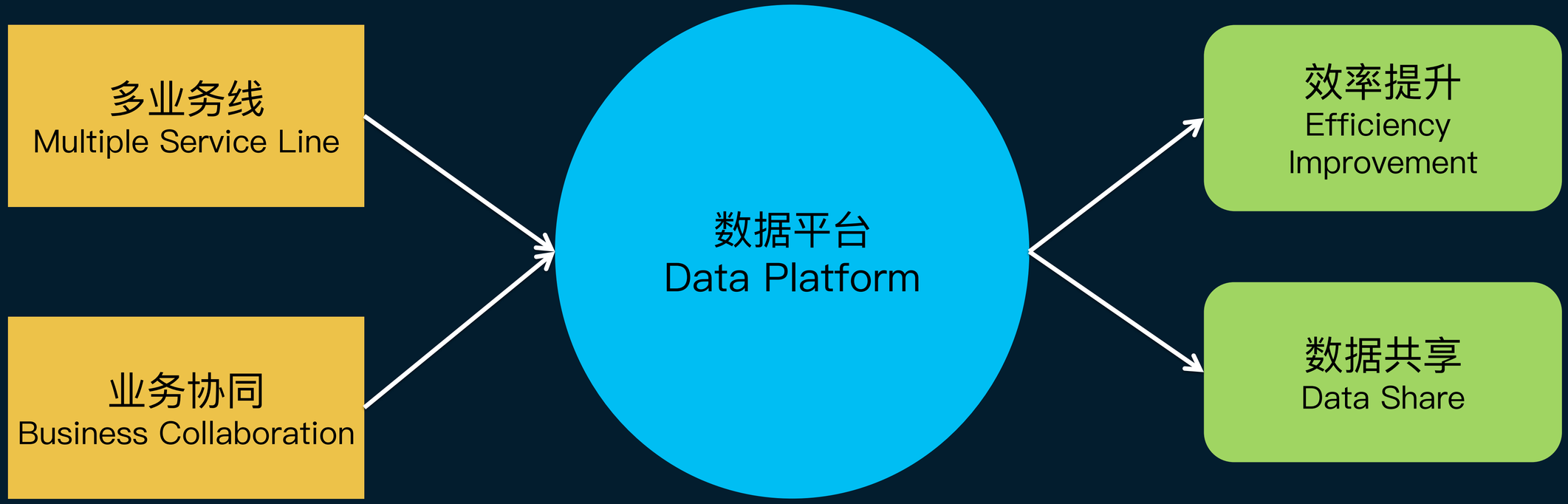


调优诊断 Tuning & Debugging

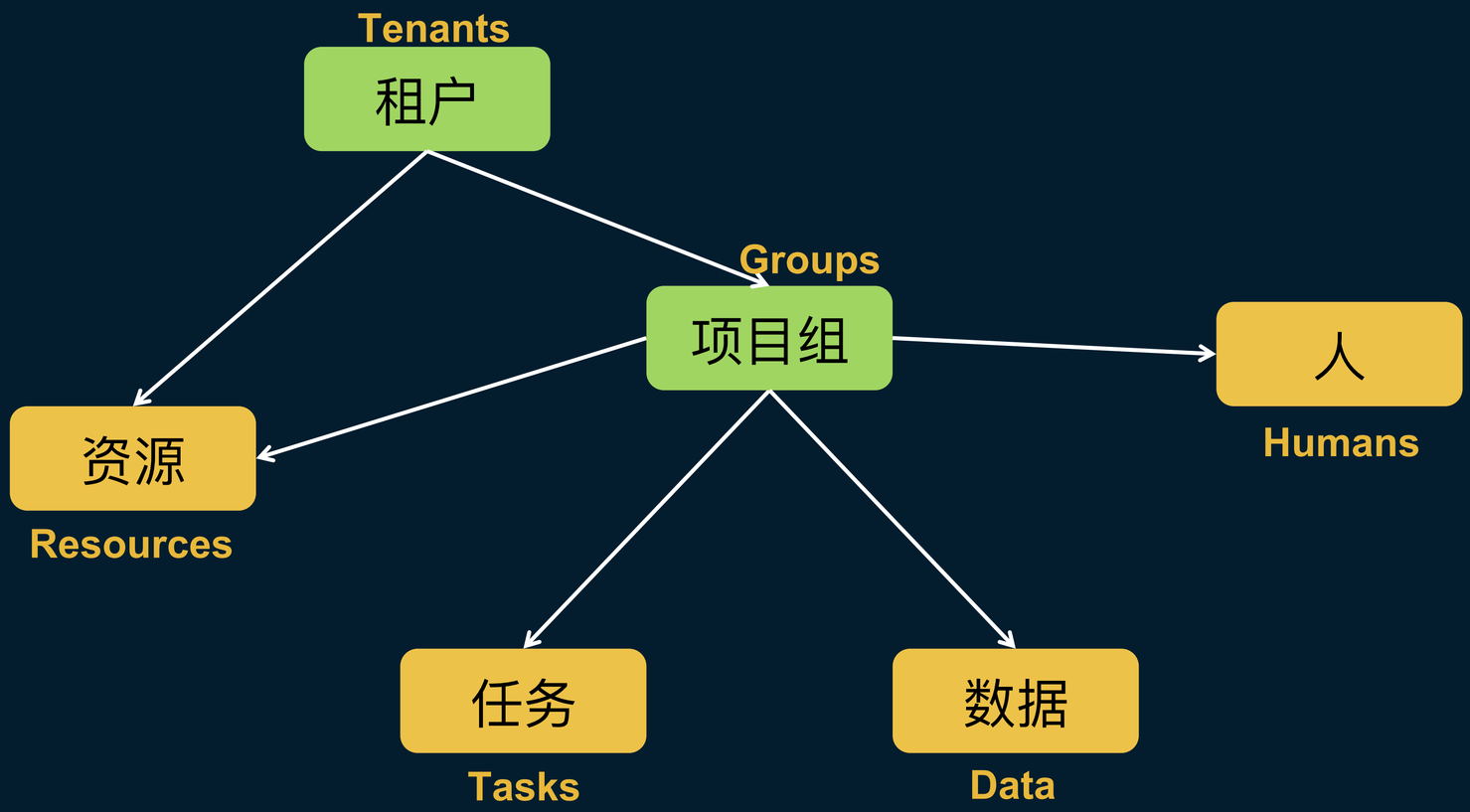




为什么需要平台化？ Why we need Platform?



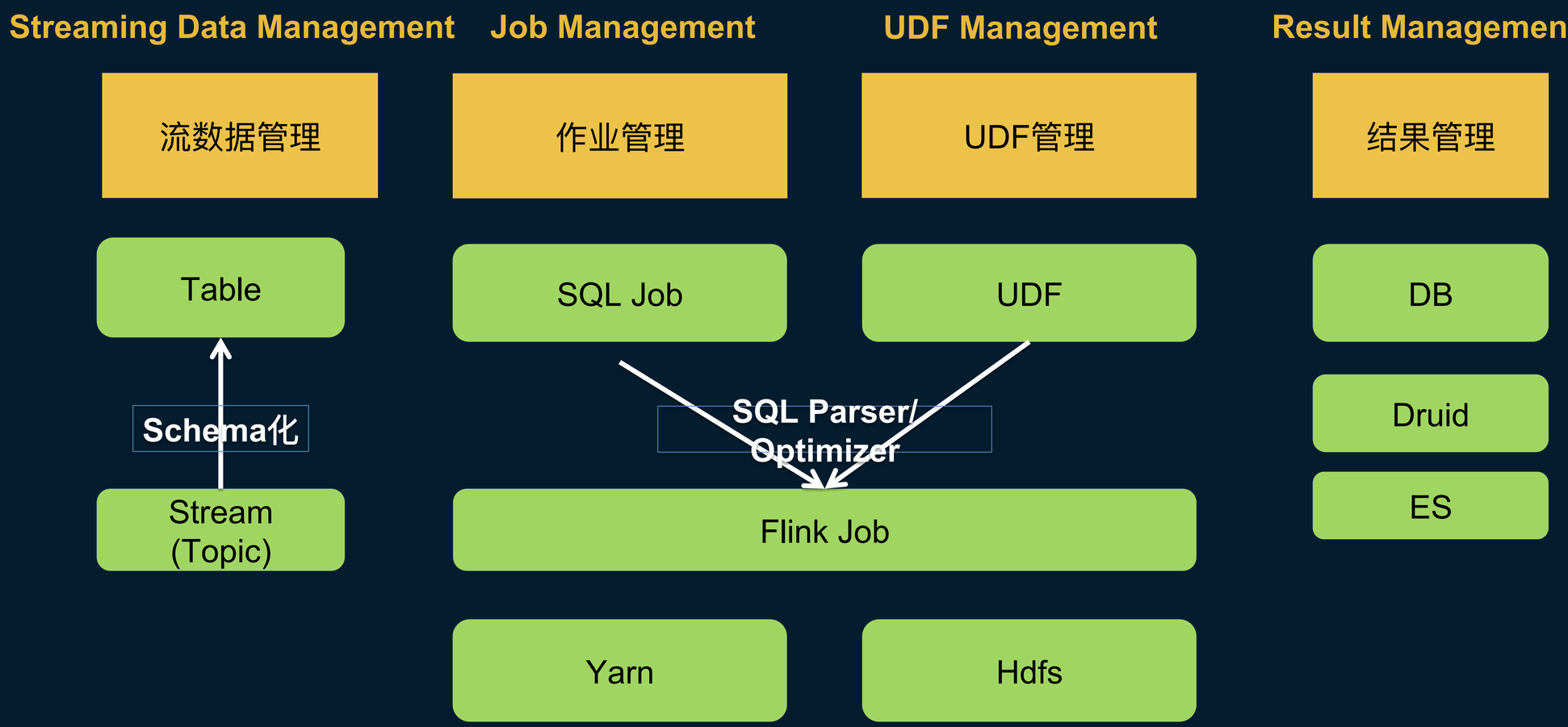
平台化建设 The Platform Construction



平台化建设 The Platform Construction

- 租户体系 The Tenant System
- 资源管理 Resource Management
- 任务&数据管理 Task & Data Management
- 权限管理 Authority Management

SQL化 Embrace SQL



SQL化 Embrace SQL

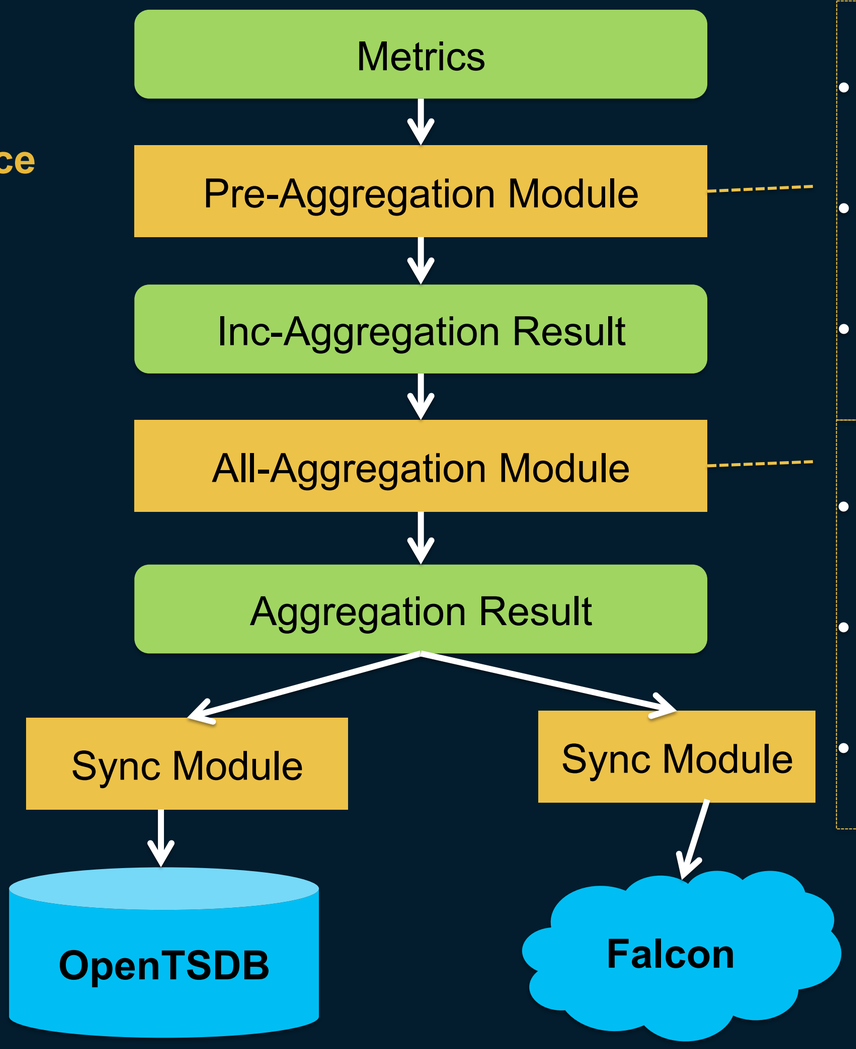
- 流Schema化和Table管理 — Introduce Schemas in Streams and Table Management
- SQL语义的丰富度 — The Diversity of SQL
- UDF扩展 — UDF Extensions
- 执行优化 — Execution Optimization

Outline

- 介绍 Introduction
- 平台建设实践 Practice in Platform Construction
- 实时应用 The Realtime Applications
- 挑战&未来 Challenges and the Future Work

Petra

- 实时指标聚合服务
  Realtime Metrics Aggregation Service
  - Event-time
    事件时间
  - Multi-dimension
    多维度
  - Compound indicator
    复合指标
  - Exactly Once
    精确一次语义
  - Low latency
    低延迟

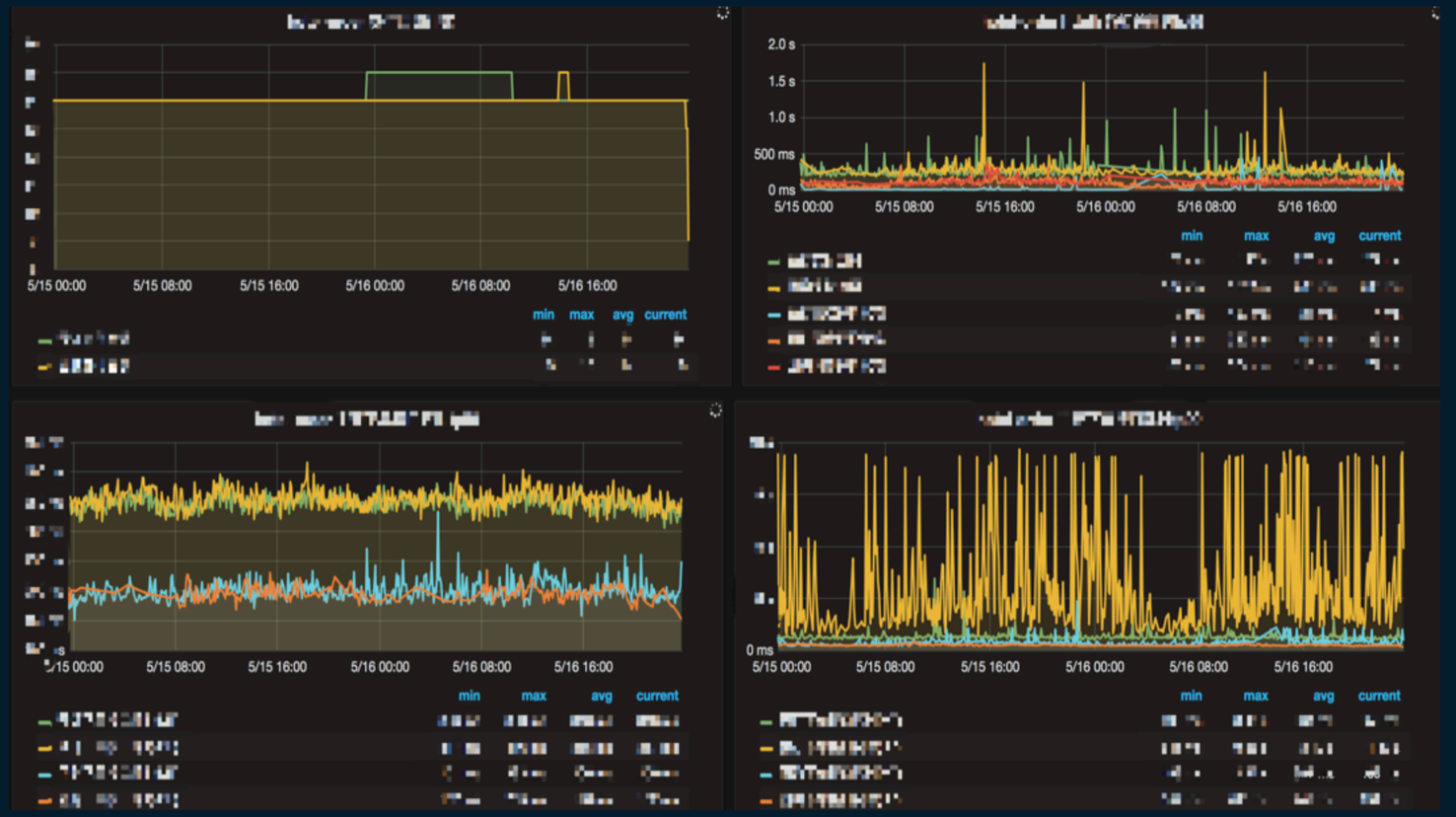


- Exactly Once
  精确一次语义
- Window Aggr based on process-time
  基于处理时间的窗口聚合
- Inc-Aggregation Result
  增量式聚合结果

- Exactly Once
  精确一次语义
- Window Aggr based on event-time
  基于事件时间的窗口聚合
- All-Aggregation Result
  全量式聚合结果

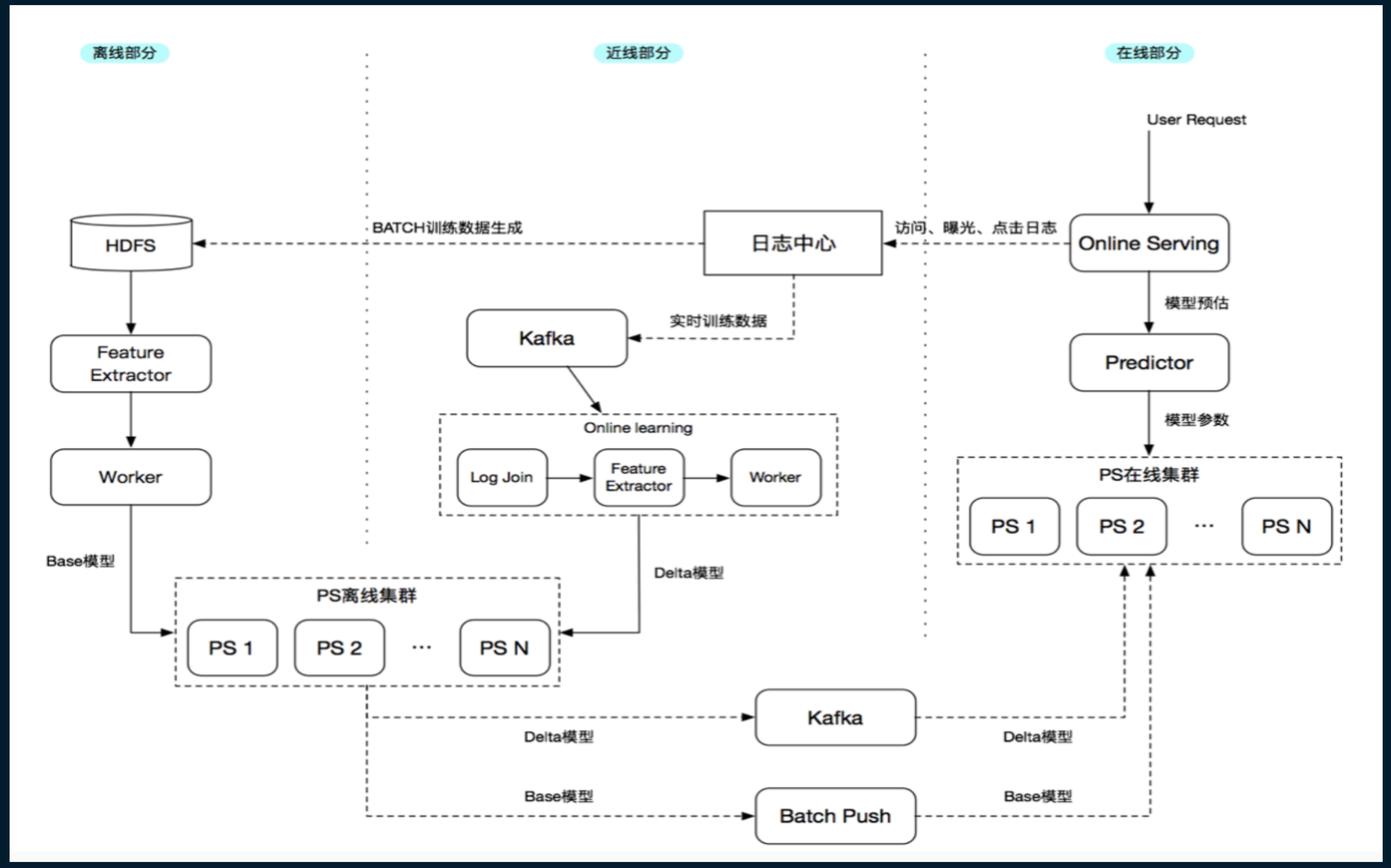
Petra

- Exactly Once 保障
  Exactly Once Guarantee
- 数据倾斜
  Data Skew
- 晚到数据处理
  Processing Late Data



MLX

- 机器学习平台
  A Machine Learning Platform
  - 近线训练部分
    Training Only
  - 流Join
    Stream Join
  - 大窗口
    Large Window



Outline

- 介绍 Introduction
- 平台建设实践 Practice in Platform Construction
- 实时应用 The Realtime Applications
- 挑战&未来 Challenges and the Future Work

我们的用户满意吗？Are Users Satisfied?

- 引擎能力和表意能力不足以支撑业务要求
- 效率（开发、调试、问题追查、热点机器）不尽如人意
- 高可靠性（全链路监控、容灾）达不到业务要求
- 数据质量（丢失率、延迟指标）低

The Engine Is NOT Powerful and Expressive Enough for the Business.

Low Efficiency (Development, Debugging and Tracing)

Unavailability (Link Monitoring and Disaster Tolerance)

Low Data Quality (Loss Rate and Delay)

Future Work

- 依托平台化，解决效率问题
  Improve Efficiency Based on the Platform
- 依托实时数仓建设，提升引擎能力和表意能力
  Improve the Power and Expressive Richness Based on the Realtime Data Warehouse
- 依托B端、监控业务建设，提升可靠性
  Improve the Reliability Based on the Browser and Monitoring Businesses

THANKS