



Apache Flink

Zeppelin中玩转Flink + Hive

章剑锋 · 阿里巴巴 / 高级技术专家

Apache Flink China Meetup 上海
- 2019年09月07日



Apache Flink

My Journey



Pivotal™



2008

2010

2013

2014

2018

Data



APACHE



Apache Flink

Contents

目录 >>

1

Introduction

3

User Experience
in Zeppelin

2

Integration
between Hive &
Flink

4

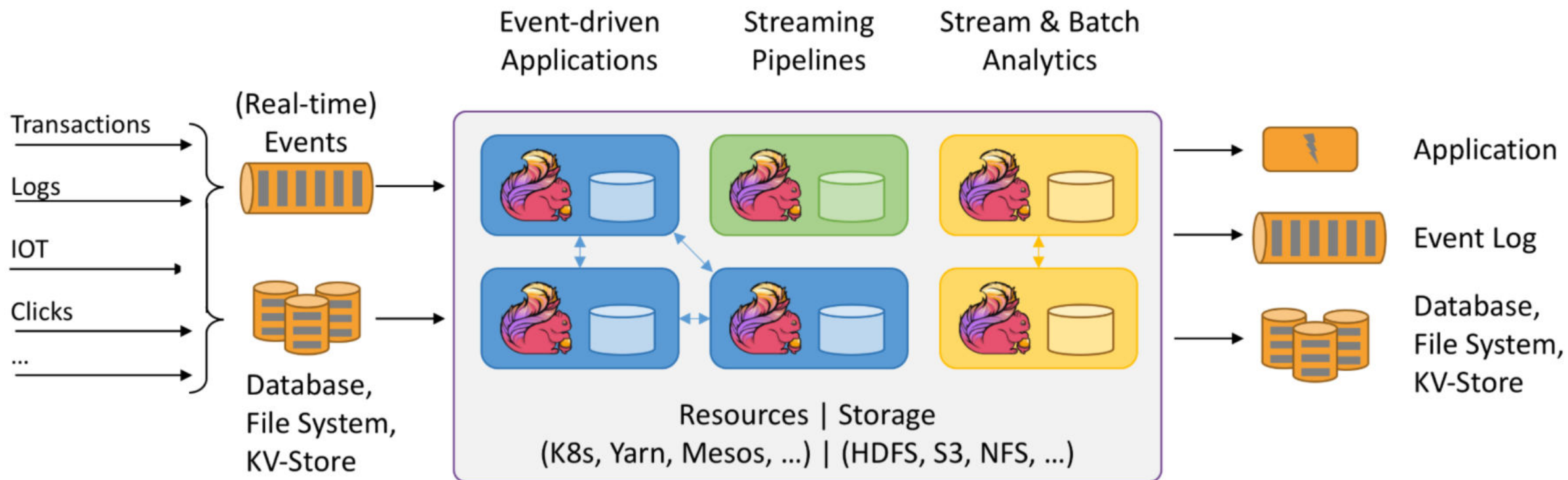
Demo



Apache Flink

PART 01

Introduction



Flink Use Cases



Apache Flink

- Event driven application
- Data Analytics application
- Data Pipeline application



Apache Flink

PART 02

Integration between
hive & flink



- Flink是最流行的流处理引擎之一
- Flink在批处理方面也展现了很大的潜力
- SQL是大数据处理中最重要的工具之一
- Flink SQL的不足
 - 缺少一个完善的Metadata持久化和管理系统
 - 缺少DDL
 - 缺少统一的Catalog接口
 - 缺少与Hive的互操作能力
 - ...



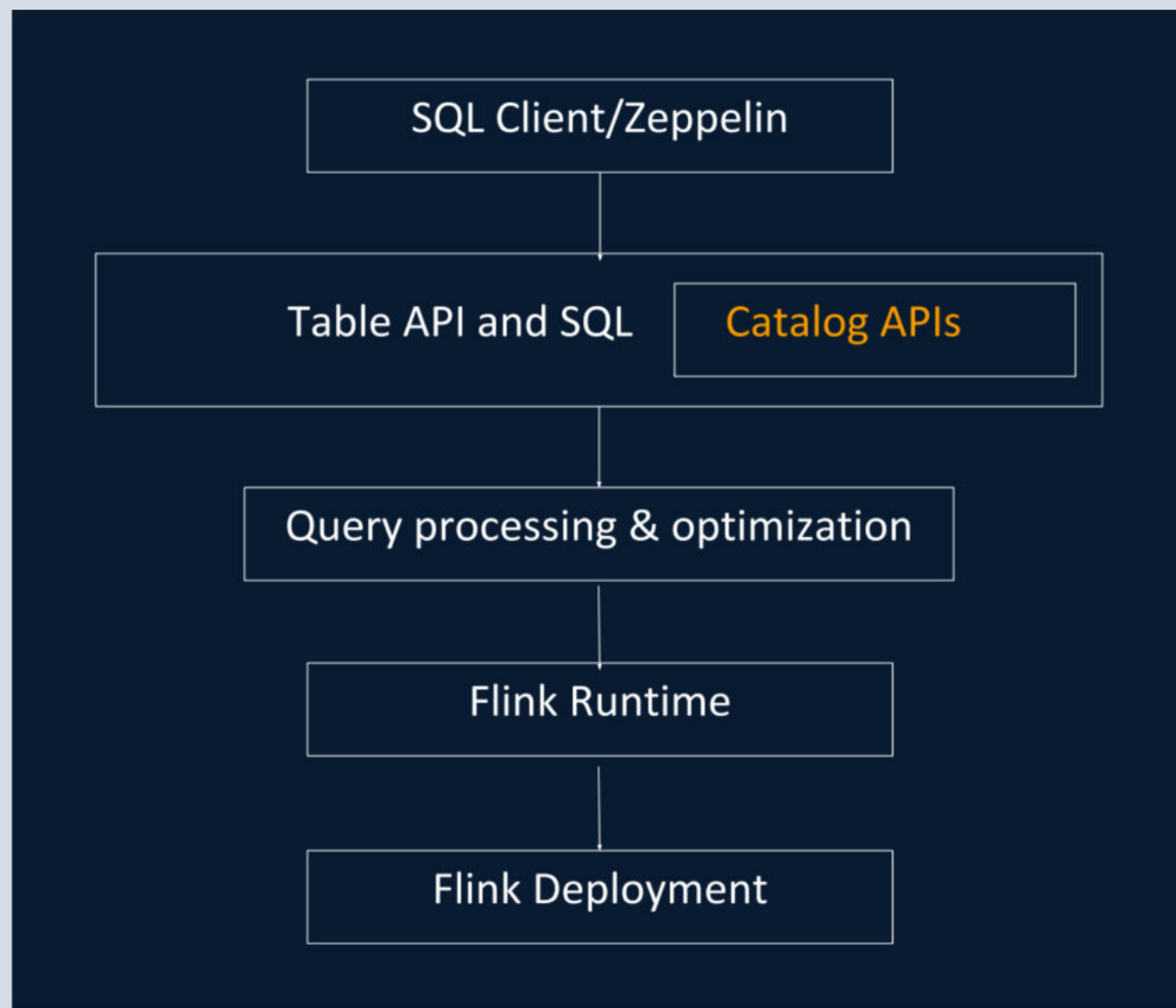
Goals

- 统一的Catalog接口
- 提供基于内存的和可持久化的Catalog实现
- 支持与Hive的互操作
 - 访问Hive元数据
 - 访问Hive表数据
- 支持Flink作为Hive的引擎（长期目标）



Benefits

- 增强可用性，支持更多场景
- 统一的引擎，提高效率
- 降低维护成本
- 吸引Hive用户迁移到Flink
- 进一步完善Flink生态





Hive Data Connector

- HiveTableFactory (实现TableSourceFactory与TableSinkFactory)
 - 与HiveCatalog关联
 - 对于Hive表
 - 读数据—>创建HiveTableSource—>创建HiveTableInputFormat
 - 写数据—>创建HiveTableSink—>创建HiveTableOutputFormat
 - 对于Generic表
 - 现有的service discovery机制



多版本支持

- 通过HiveShim兼容多版本的Hive
 - Hive-1.2.1
 - Hive-2.3.4
- 需要用户部署hive-exec.jar
- 可通过yaml文件指定hive-site.xml与Hive版本



Apache Flink

PART 03

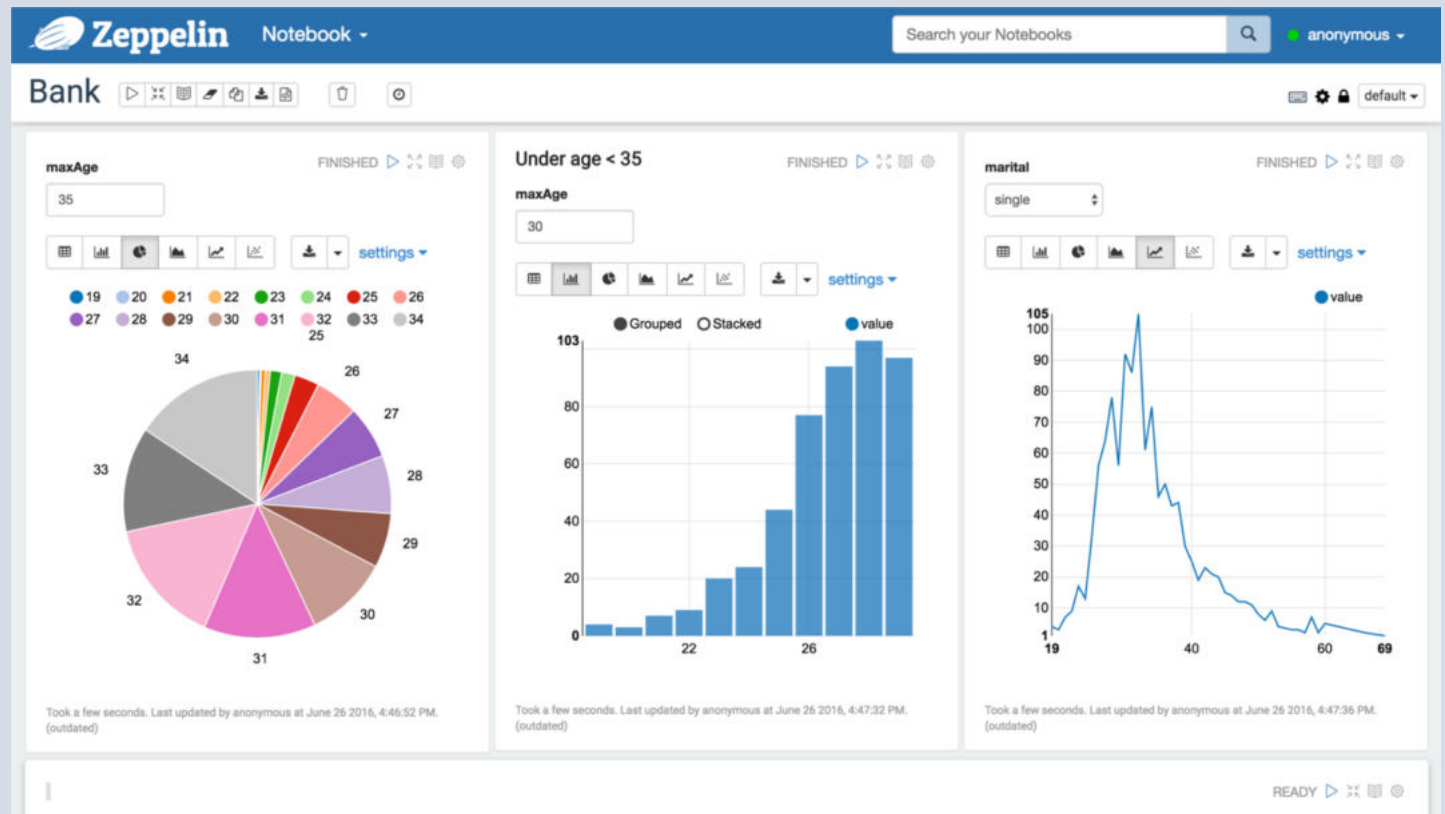
User Experience in Zeppelin

Apache Zeppelin



Apache Flink

Web-based notebook that enables data-driven, interactive data analytics and collaborative documents with SQL, Scala and more.



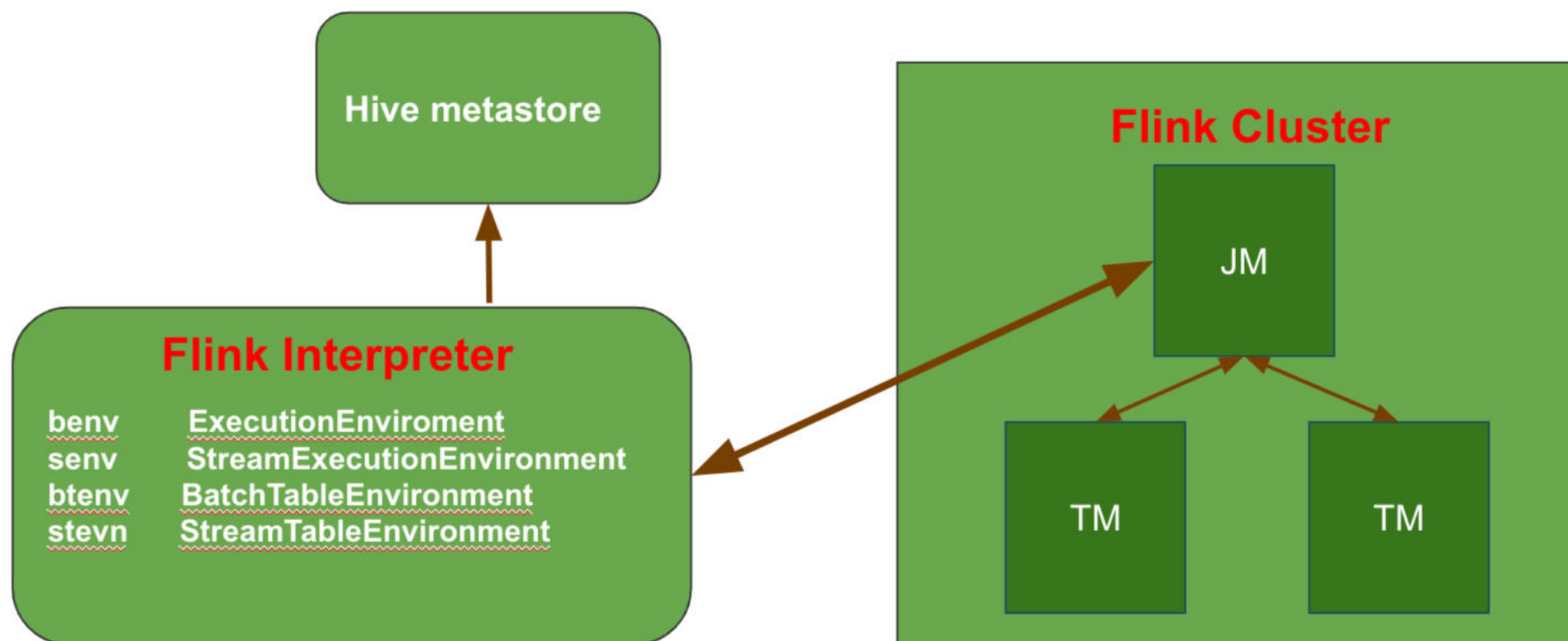


Flink Interpreter

- Scala. (Flink scala api)
- Python. (PyFlink)
- SQL. (Flink Sql, Hive)



Flink Interpreter





Apache Flink

PART 04

Demo



Flink Forward Asia

全球最大的 Apache Flink 官方会议

预计 2000+ 参会人员， 2019年11月28-30日 @北京国家会议中心

国内外一线厂商悉数参与

阿里巴巴、腾讯、字节跳动、intel、 DellEMC 、Uber、美团点评、Ververica ...



大会官网，查看更多



Flink Forward Asia

The largest Apache Flink official conference all over the world

Expected 2000+ participants

28th November 2019 - China National Convention Center @Beijing

Numbers of Top domestic and foreign companies

Alibaba, Tencent, Bytedance, Intel, DellEMC, Uber, Meituan Dianping, Ververica...



Official website

Apache Flink 社区微信公众号「Ververica」



Meetup动态 / Release 发布信息 / Flink 应用实践





Apache Flink

THANKS

Apache Flink China Meetup

▪ SHANGHAI

