

表格存储实时数据流

Stream的技术揭秘和应用场景

主讲人：木洛

提纲

- 表格存储简介
- Stream功能解析及应用
- 总结和展望

新挑战

应用开发新挑战

1. 服务化
2. 多元化海量数据
3. 数据高附加值的显现
4. 快速迭代

数据库新需求

1. 高可用
2. 多元化数据存储
3. 海量数据存储、高可靠、低成本
4. 在线存储和离线计算结合
5. 灵活的数据模型

传统关系型数据库与表格存储的对比

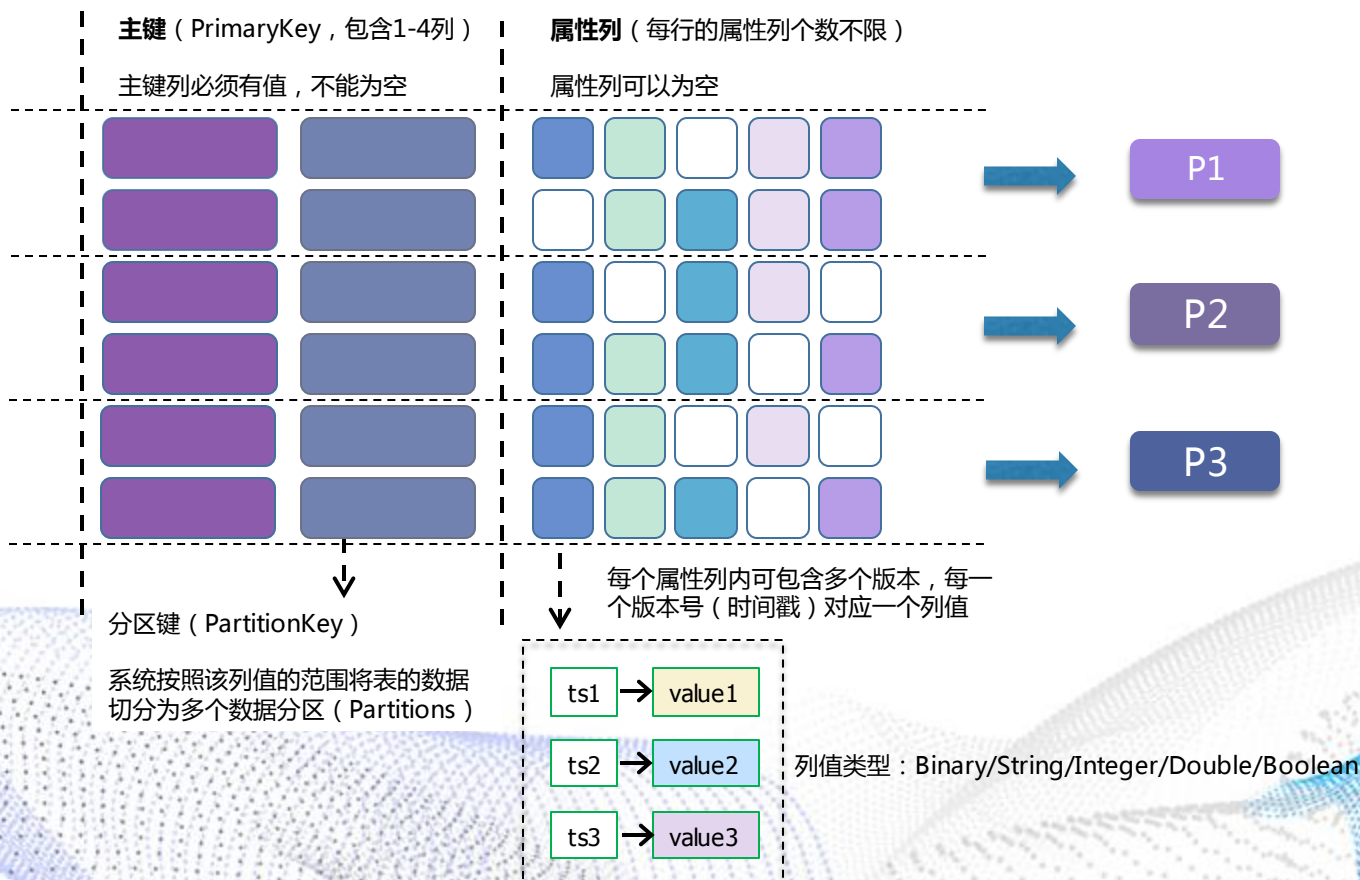
	传统关系型数据库	表格存储
容量	TB	PB
服务能力	十万TPS	百万TPS
分布式支持	分库分表	原生支持
售卖模式	按规格计费	按量计费
数据模型	关系型，强Schema	弱关系型，弱Schema

表格存储简介 - 特性



表格存储：基于共享存储的高性能、低成本、易扩展、全托管的半结构化数据存储平台，支撑互联网和物联网数据的高效计算与分析。

表格存储简介－数据模型



`SortedMap<PrimaryKey, SortedMap<ColumnName, List<Pair<ColumnValue, Timestamp>>>>`

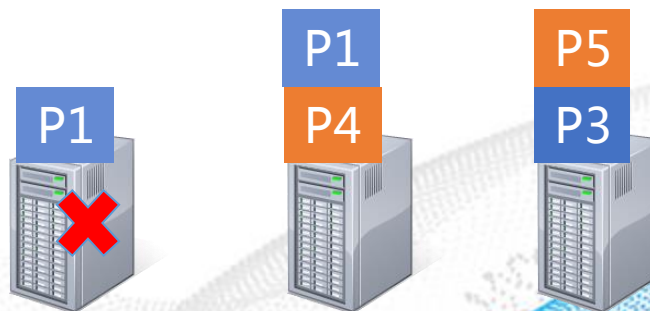
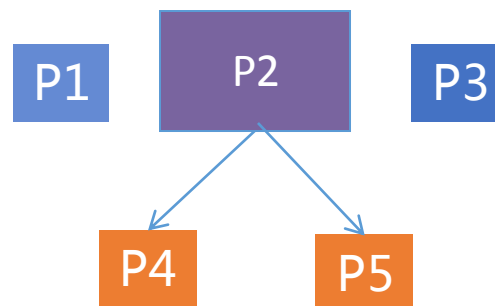
表格存储简介 – 高可用架构

✓ 高可扩展性

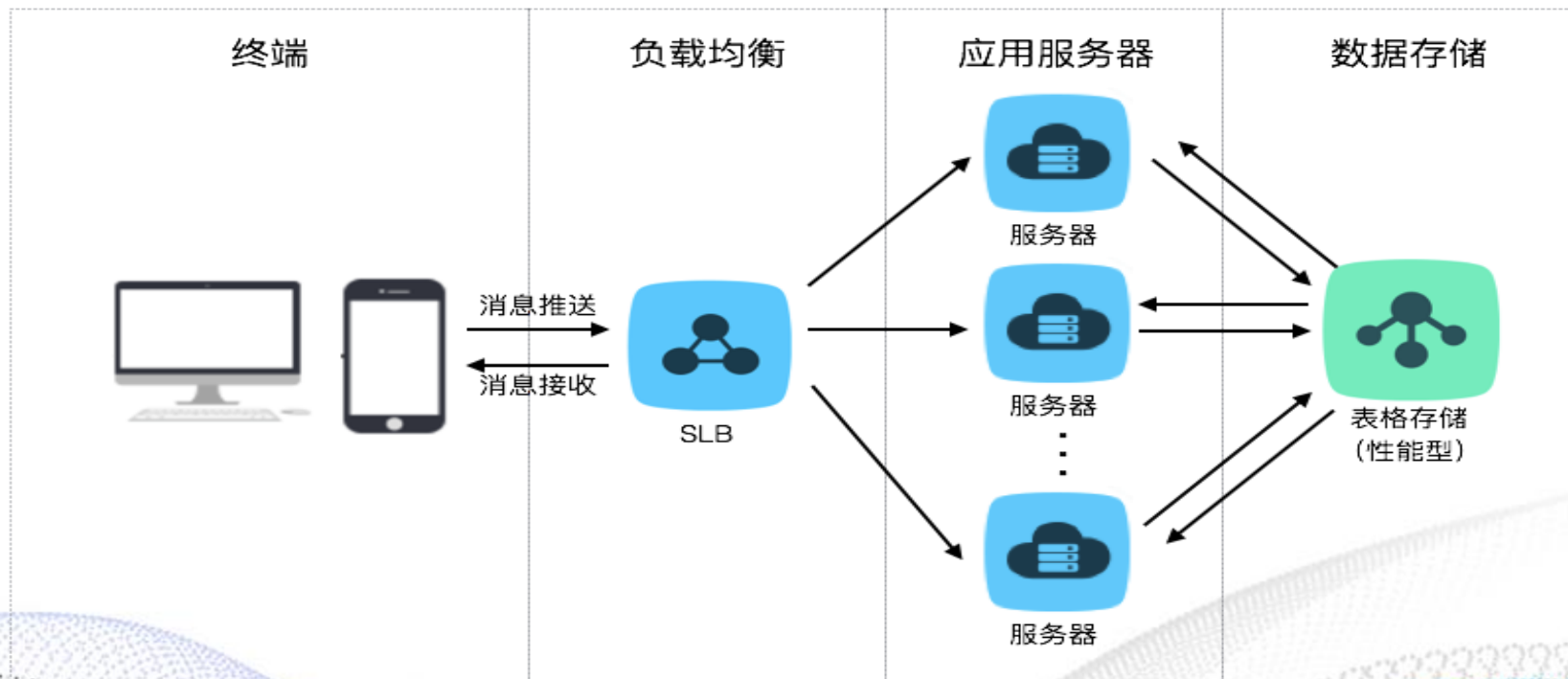
- ShareNothing架构
- 自动分区
- 动态负载均衡

✓ 故障自动恢复

- 分区自动迁移

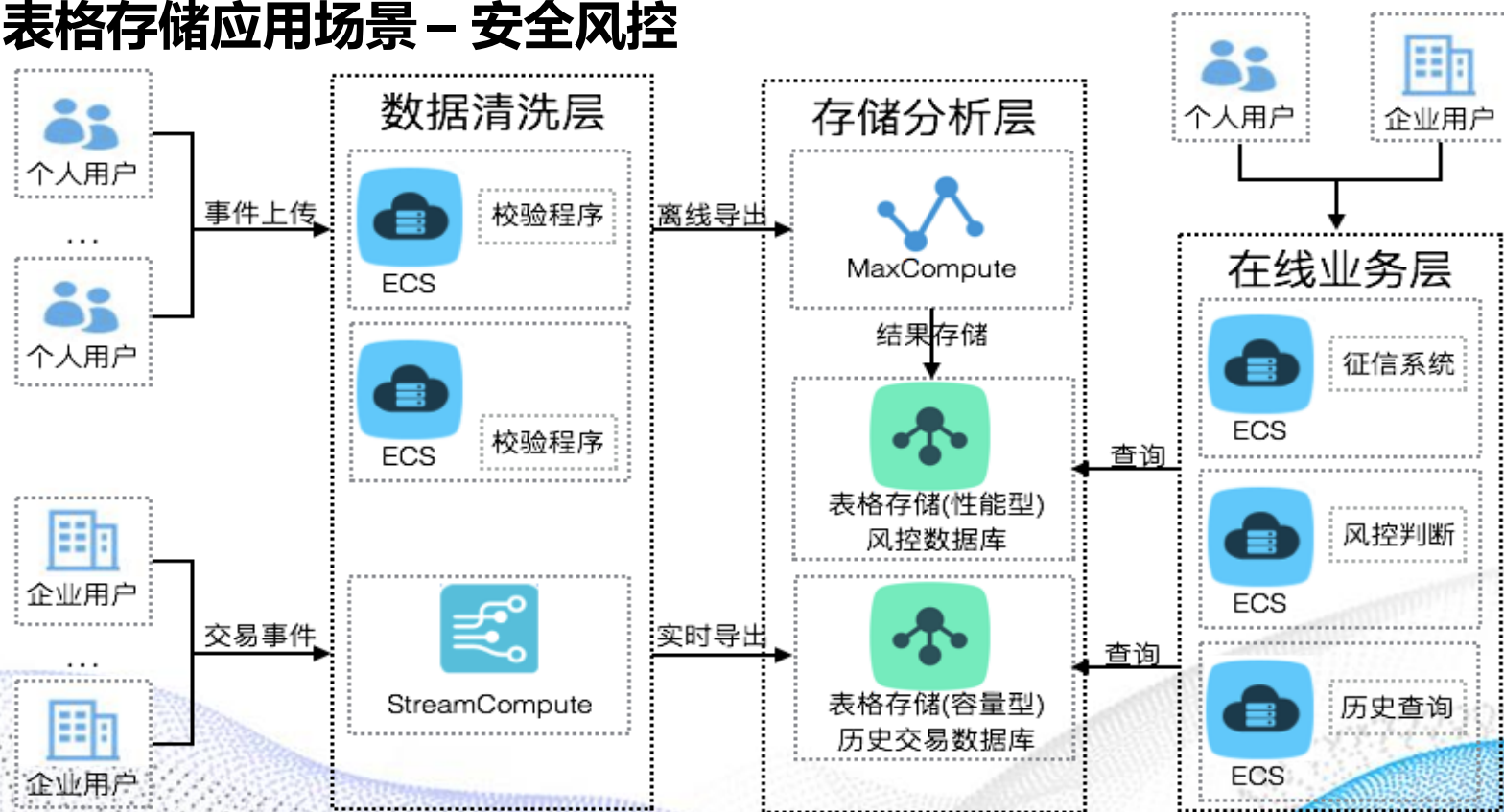


表格存储应用场景- 即时通讯



- 写扩散模型，很高的写入压力（十万TPS，峰值百万TPS）
- 低延迟的消息同步
- 海量消息存储
- 高可用

表格存储应用场景 – 安全风控



- 实时风险判断（低延迟读）
- 海量历史交易存储
- 离线计算结果导出到在线存储
- 灵活的数据模型
- 高可用

表格存储应用场景 – 时序数据



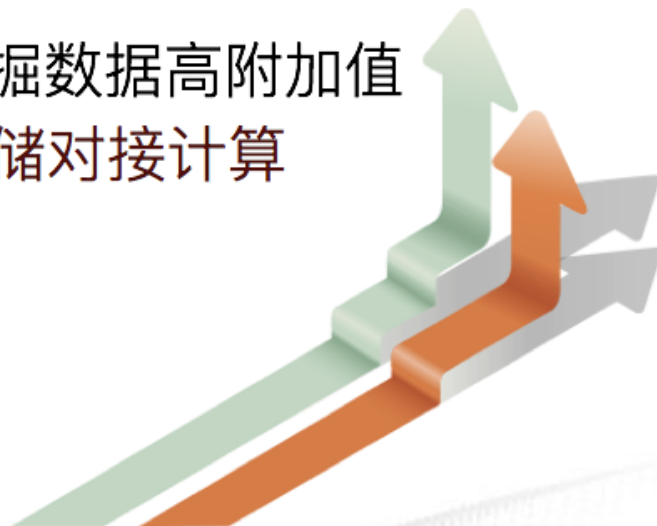
1. 高写入吞吐量 (百万TPS)
2. 写多读少 (90%以上比例是写)
3. 低读取延迟
4. 数据具有时效性
5. 数据基数大 (PB, 日增亿级)
6. 实时分析与离线计算

表格存储应用场景 – 特点

使用表格存储的应用场景的特点

1. 弱关系型数据
2. 海量数据存储
3. 高并发访问
4. 低存储成本（容量型）
5. 低延迟访问（性能型）

挖掘数据高附加值
存储对接计算



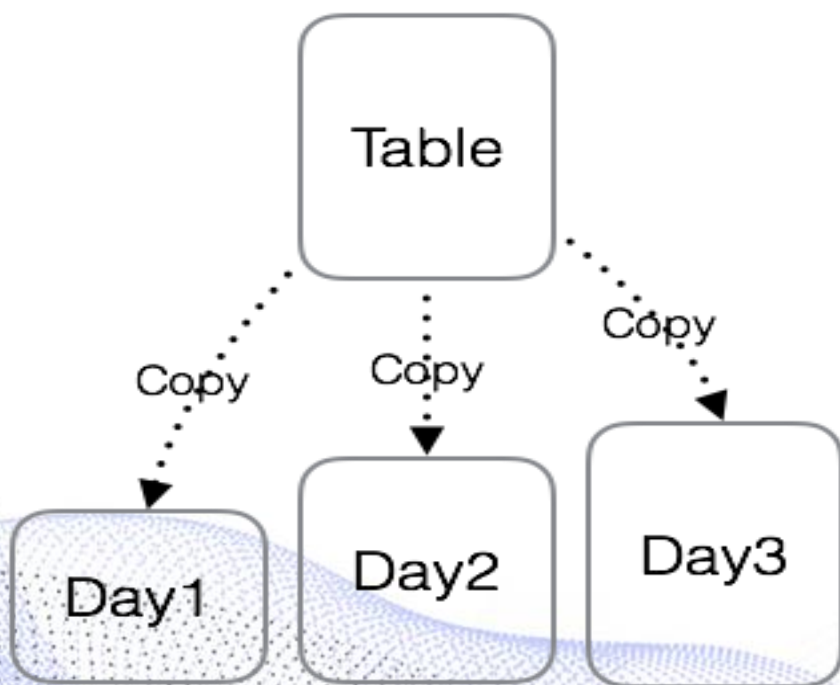
大数据时代数据存储与计算的闭环



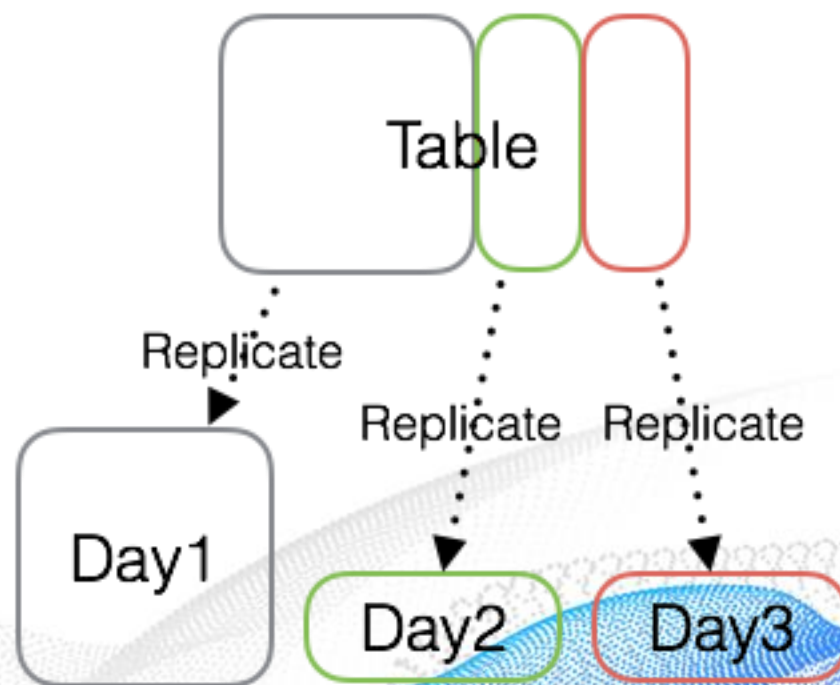
在当前大数据技术的发展背景下，没有一个系统能够满足所有的存储和计算的需求。因此，打通各个系统之间的数据通道，让数据在在线存储、实时分析、离线计算中高速流动，形成闭环，是打造大数据平台、挖掘数据价值的关键一环。

如何让数据流动起来？

数据复制

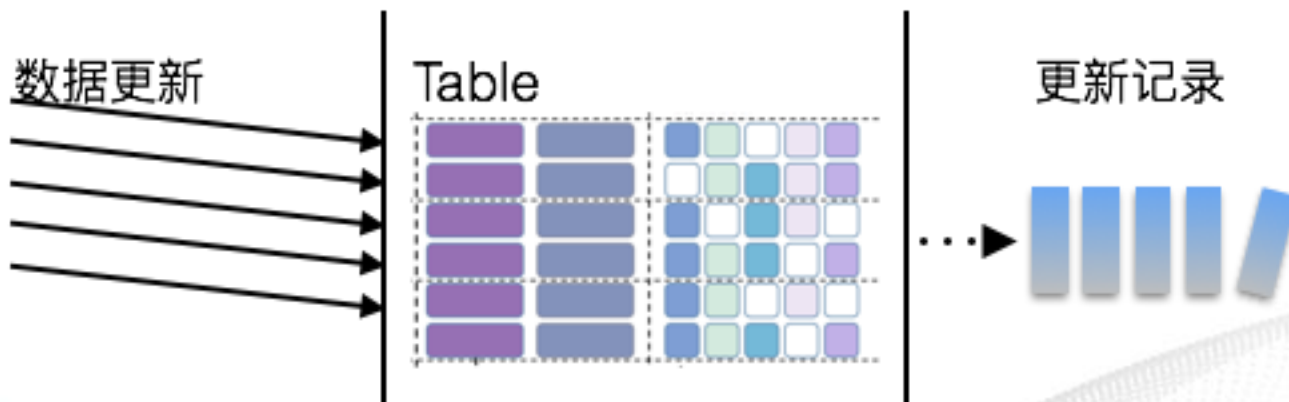


增量同步



表格存储Stream – Stream是什么

Stream能够实时捕获表内数据的变化，返回一个有序的行级的数据变更记录。其赋予了表格存储内数据流动的能力，是打造闭环的关键技术。



表格存储Stream – 设计目标

实时性

- 秒级同步

正确性

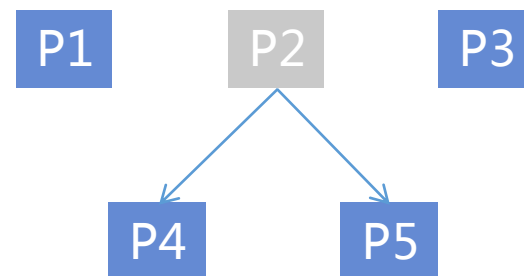
- 最终一致

高吞吐

- 快生产快消费

表格存储Stream – 实现原理

P1	R0	R1	R2	R3	R4	R5
P2	R0	R1	R2	R3	R4	R5
P3	R0	R1	R2	R3	R4	R5
P4	R6	R7	R8	R9	R10	R11
P5	R6	R7	R8	R9	R10	R11

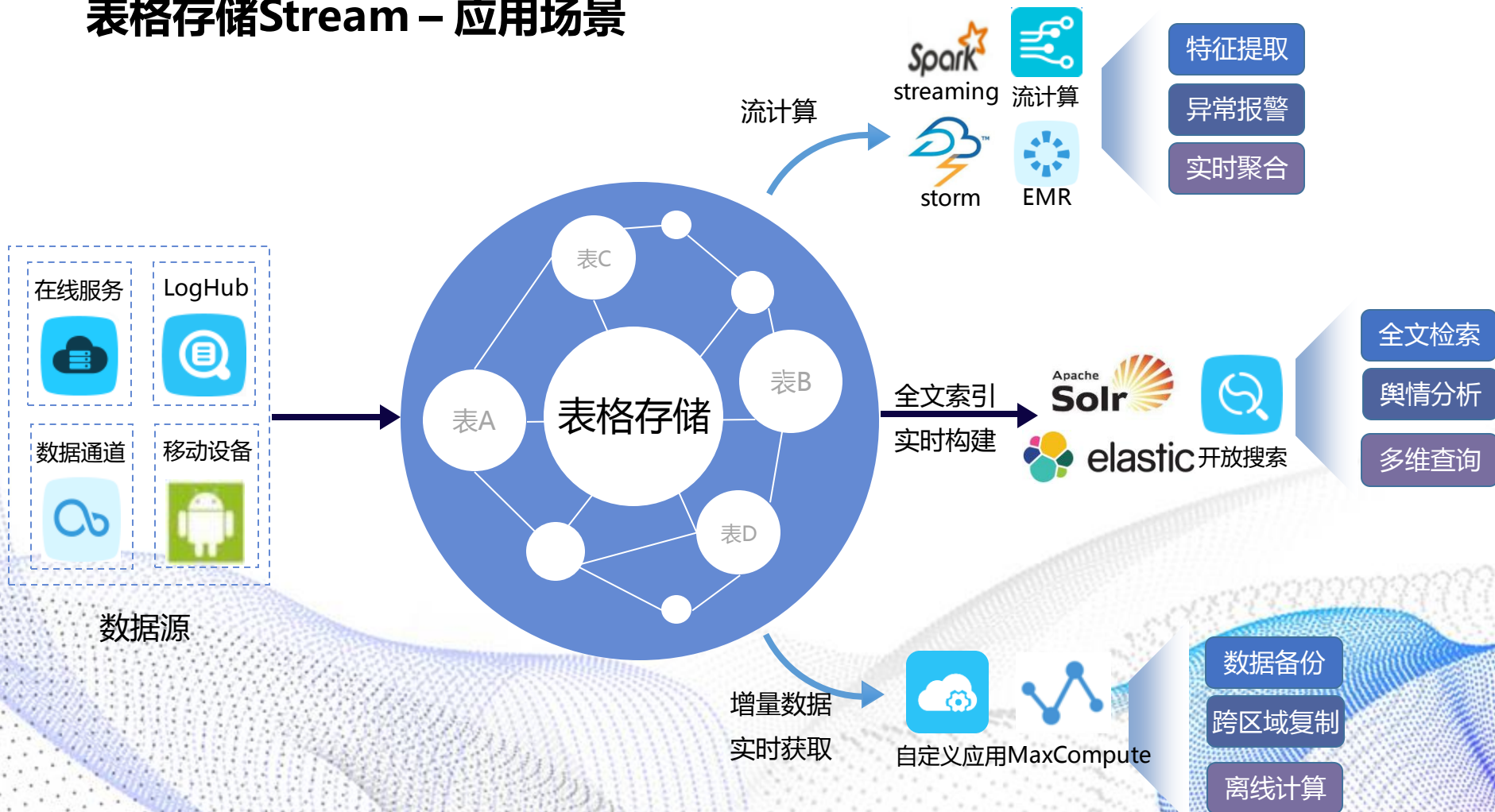


- Stream数据源自分区级的CommitLog
- Stream构建了一个树状的CommitLog数据消费体系

表格存储Stream – 如何使用Stream

- Stream Client
 - 分布式消费客户端
 - 能够自定义数据消费行为
- Stream Reader (DataX , CDP)
 - 与MaxCompute的无缝对接，能够实现T+1实时的数据增量同步
- 对接StreamCompute (即将发布)
 - 无缝对接流计算

表格存储Stream – 应用场景



总结和展望

表格存储Stream的未来规划

- 对接函数计算 (Function Compute)
- 对接开源生态：Storm、Spark Streaming等
- 全托管的数据实时同步服务
- 拓展更多的使用场景

欢迎加入钉钉交流群



扫一扫群二维码，立刻加入该群。



2017云栖大会·成都峰会
THE COMPUTING CONFERENCE

阿里云

云栖社区
yq.aliyun.com

飞天·智能

APSARA INTELLIGENCE

2017云栖大会·成都峰会

5月23日 成都世纪城天堂洲际大酒店