

HUDI 在字节跳动推荐系统中的实践

管梓越

目录

01

场景需求

02

设计选型

03

功能支持

04

性能调优

05

未来展望

01 场景需求

- **BigTable CDC**
- **特征工程**

场景需求



场景需求

BigTable CDC

- 获取宽表型存储的CDC
- 提供高效的OLAP查询
- 提供线上无感的数据同步
- 数据不规整（大小，格式）
- 需求不统一

特征工程

- Instance和label的拼接
- 提供高效IO减枝的访问
- 高维复杂数据（万列，嵌套类型，稀疏）
- 高吞吐近实时写入（百GB/s）
- EB级存储

02 设计选型

- 多种数据湖引擎
- MOR or COW
- 索引类型
- 计算引擎

设计选型



01 数据湖选型

Iceberg: 良好的数据抽象和优秀的接口设计

Hudi: 灵活的接口实现，全局索引，MOR

DeltaLake: 和spark强绑定

02 实时写入

COW or **MOR**

03 索引类型

Simple Bloom **HbaseIndex**

04 计算引擎

Spark or **Flink** | **RDD** API or **DataSource** API

03 功能支持

- MVCC
- Schema注册系统

MVCC



Schema

Schema Registry

- 原子变更
- 异地多活
- 多版本回溯
- Column property
- 异构自动同步
- Column序列编码

Pull Sync

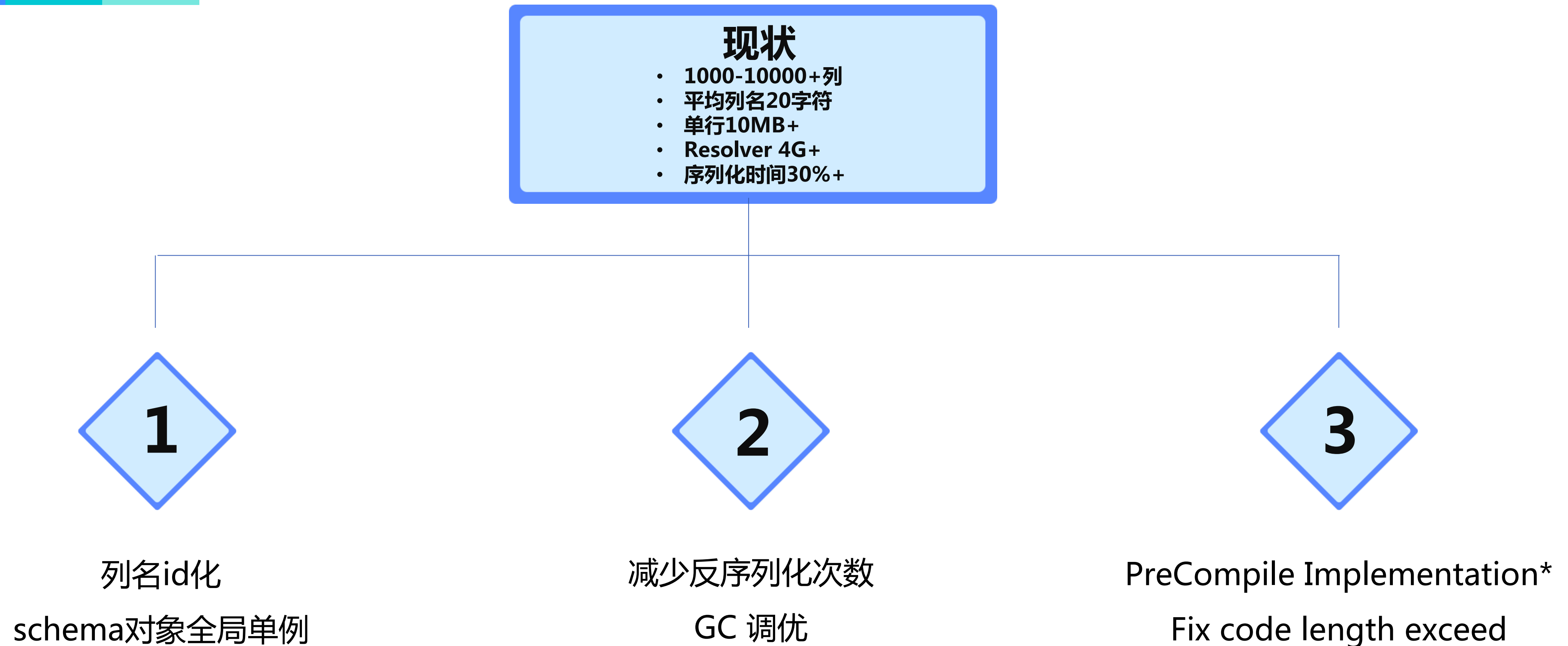
Application

Local
Cache

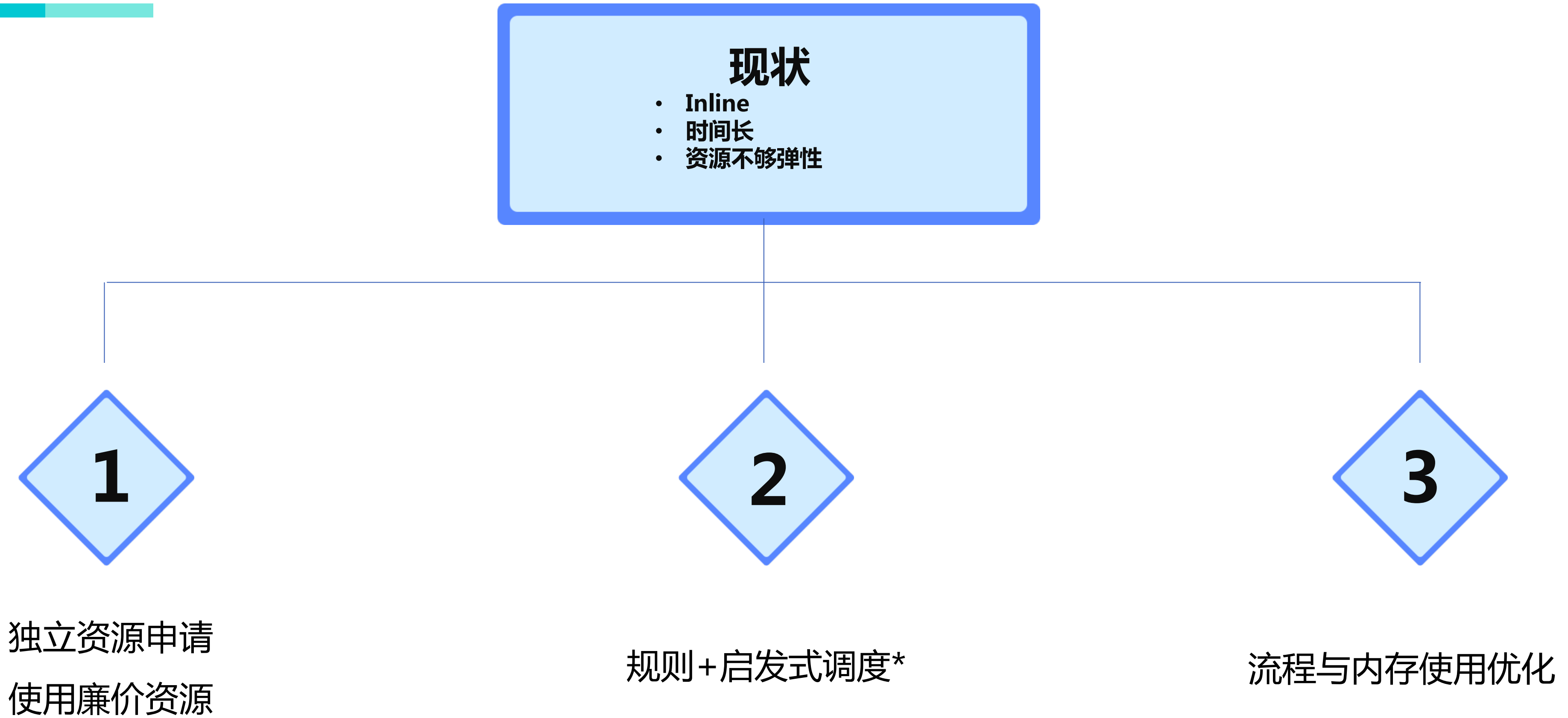
04 性能调优

- 序列化
- Compaction
- HDFS SLA
- 流程优化

序列化



Compaction



HDFS SLA



Hflush or Hsync



激进策略重试
超时切块/rollover



独立集群



流程优化



一些琐碎的流程优化与bug fix

- 避免rewrite操作
- 插件式的record size评估
- 基于row count的小文件评估
- 简易的adaptive execution 规避写入倾斜
- 自定义partitioner 优化shuffle
- Bulkinsert 索引bulkload
- Timeline缓存不一致更新
-

05 未来展望

- 产品化
- 生态对接
- 成本优化
- 性能优化
- 存储语义

未来展望

1

产品化

用户友好编程

可运维性

简化调优

2

生态对接

Flink

跨语言跨框架格式

通用访问

内部生态完善

3

成本优化

存储冷热分层

序列化优化

混部潮汐计算

优化compaction方式

4

性能优化

快设备

向量化

新格式索引

流程重构

5

存储语义

增量触发

Mutate

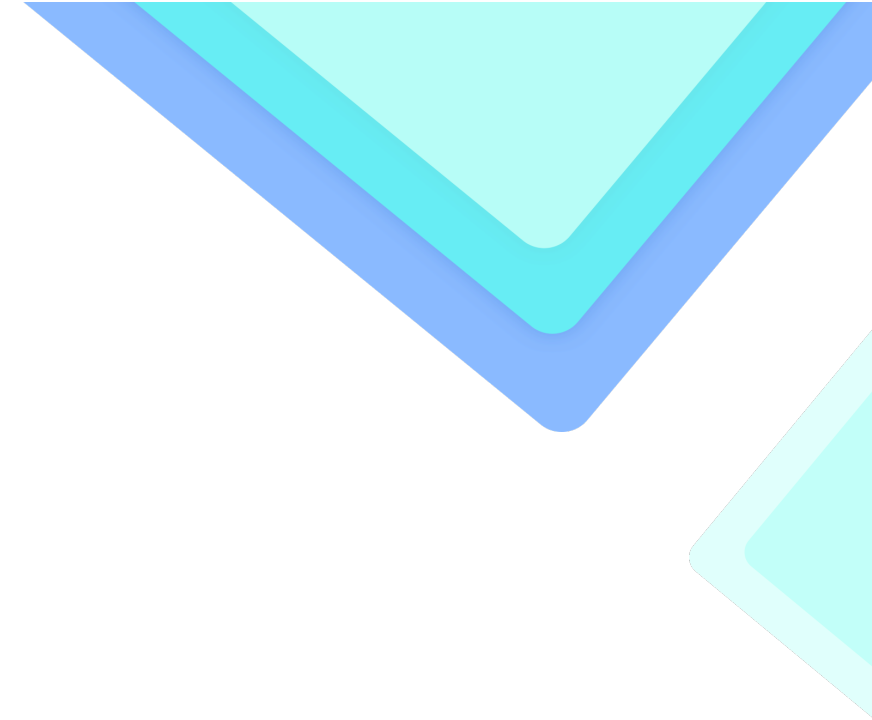
点查与scan

数据重整

Q&A



Q&A Time



We are hiring!

字节跳动推荐架构团队

- 负责抖音、今日头条、西瓜视频等产品的推荐架构的设计和开发，保障系统稳定和高可用；
- 负责在线服务、离线数据流性能优化，解决系统瓶颈，降低成本开销；
- 抽象系统通用组件和服务，建设推荐中台、数据中台，支撑新产品快速孵化以及为ToB赋能；
- 设计和实现高并发、高吞吐的服务框架、RPC框架，为业务提供快速构建服务以及高性能在线serving能力；
- 实现灵活可扩展的高性能存储系统和计算模型，打通离在线数据流，构建统一的数据中台，支持推荐/搜索/广告；

团队目前招聘以下岗位：

- 大数据开发工程师
深入了解大数据生态组件的原理
- 存储研发工程师
熟悉rocksdb/Hbase, 熟悉分布式存储
- 推荐/搜索/广告相关推荐架构工程师、后端开发工程师
- 深度学习框架研发
- devops/研发效能/编译优化
- 网络通信组件/rpc开发
- 运维工程师

工作地点：北京/上海/杭州/新加坡/山景城

欢迎自荐&推荐，岗位相关问题欢迎私戳微信
或将简历投递至邮箱

guanziyue.gzy@bytedance.com



THANKS.

 **ByteDance** 字节跳动