

基于TiDB 构建实时流计算

小红书 存储与中间件研发工程师 张亿皓



Part I – Why TiCDC ?



“走进来” & “走出去”

- 数据“走进来”：让用户“用上” TiDB
 - TiDB 生态提供了一系列数据高速导入和同步工具
 - Lightning, DM, ...
- 数据“走出去”：让用户“用好” TiDB
 - 通过 TiCDC 提供可扩展的数据分发能力
 - 可结合数仓业务、实时计算等，丰富使用场景



TiDB binlog 对比

- Kafka 单 partition 写入
- Drainer 扩展性
- 协议不通用，业务改造难度



Part II – How TiCDC ?

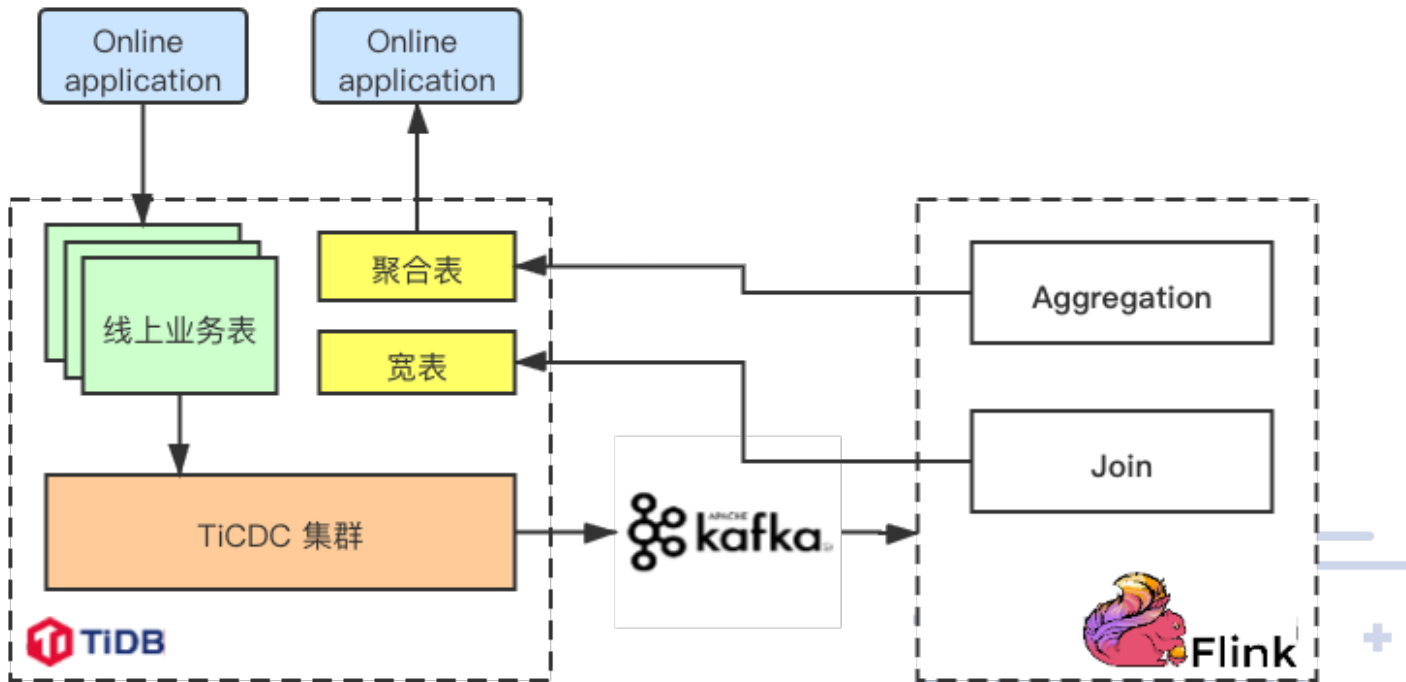


业务场景

- 内容审核
- 笔记标签推荐
- 增长审计



基于 TiDB 的实时流计算架构



参与的TiCDC开发

- 兼容输出 alibaba Canal 协议
- 基于 cdc Sink 接口，开发对应的 encoder
- 多多参与开源社区！

```
// EventBatchEncoder is an abstraction for events encoder
type EventBatchEncoder interface {
    // AppendResolvedEvent appends a resolved event into the batch
    AppendResolvedEvent(ts uint64) error
    // AppendRowChangeEvent appends a row changed event into the batch
    AppendRowChangeEvent(e *model.RowChangeEvent) error
    // AppendDDLEvent appends a DDL event into the batch
    AppendDDLEvent(e *model.DDLEvent) error
    // Build builds the batch and returns the bytes of key and value.
    Build() (key []byte, value []byte)
    // Size returns the size of the batch(bytes)
    Size() int
}
```

```
type CanalEventBatchEncoder struct {
    messages      *canal.Messages
    packet        *canal.Packet
    entryBuilder  *canalEntryBuilder
}
```



Part III – TiCDC Next ?



Problems with TiCDC (based on v4.0.2)

- Delete 语句非唯一键字段缺失 (软删除的改造)
- Update 语句更新前字段值缺失
- 多个 pk/uk 时 , pk/uk 的 partition 会失效

新版本 TiCDC , 提供更完善解决方案



Thank You !

