



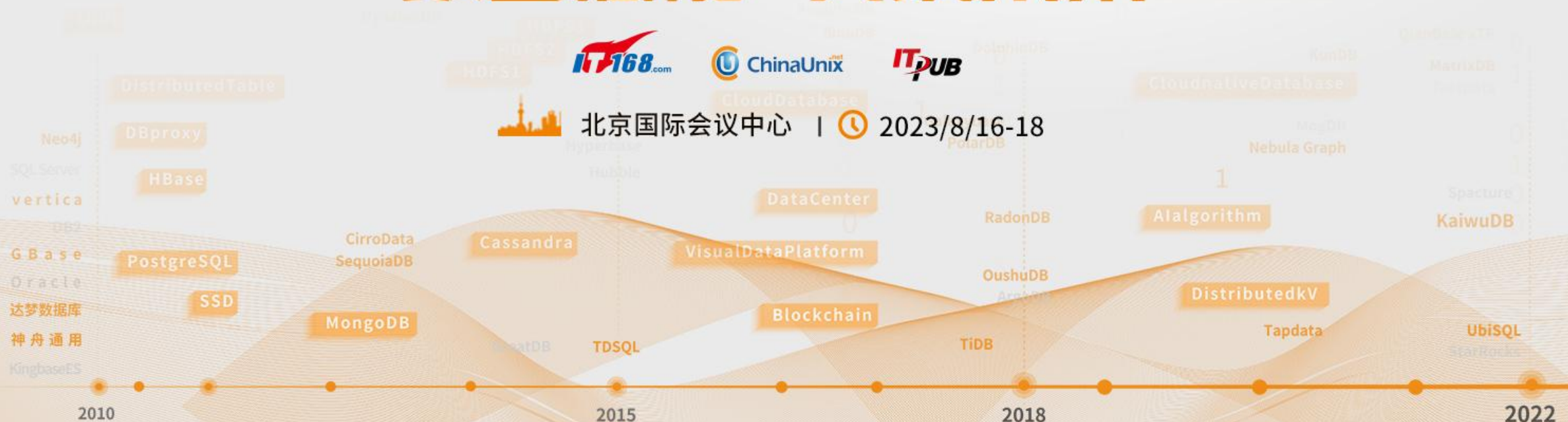
# 第十四届中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA

## 数智赋能 共筑未来



北京国际会议中心 | 2023/8/16-18



# 流批一体集成引擎赋能 CDC

字节跳动大数据技术专家

BitSail Committer, Apache Hudi PMC Member

李延加

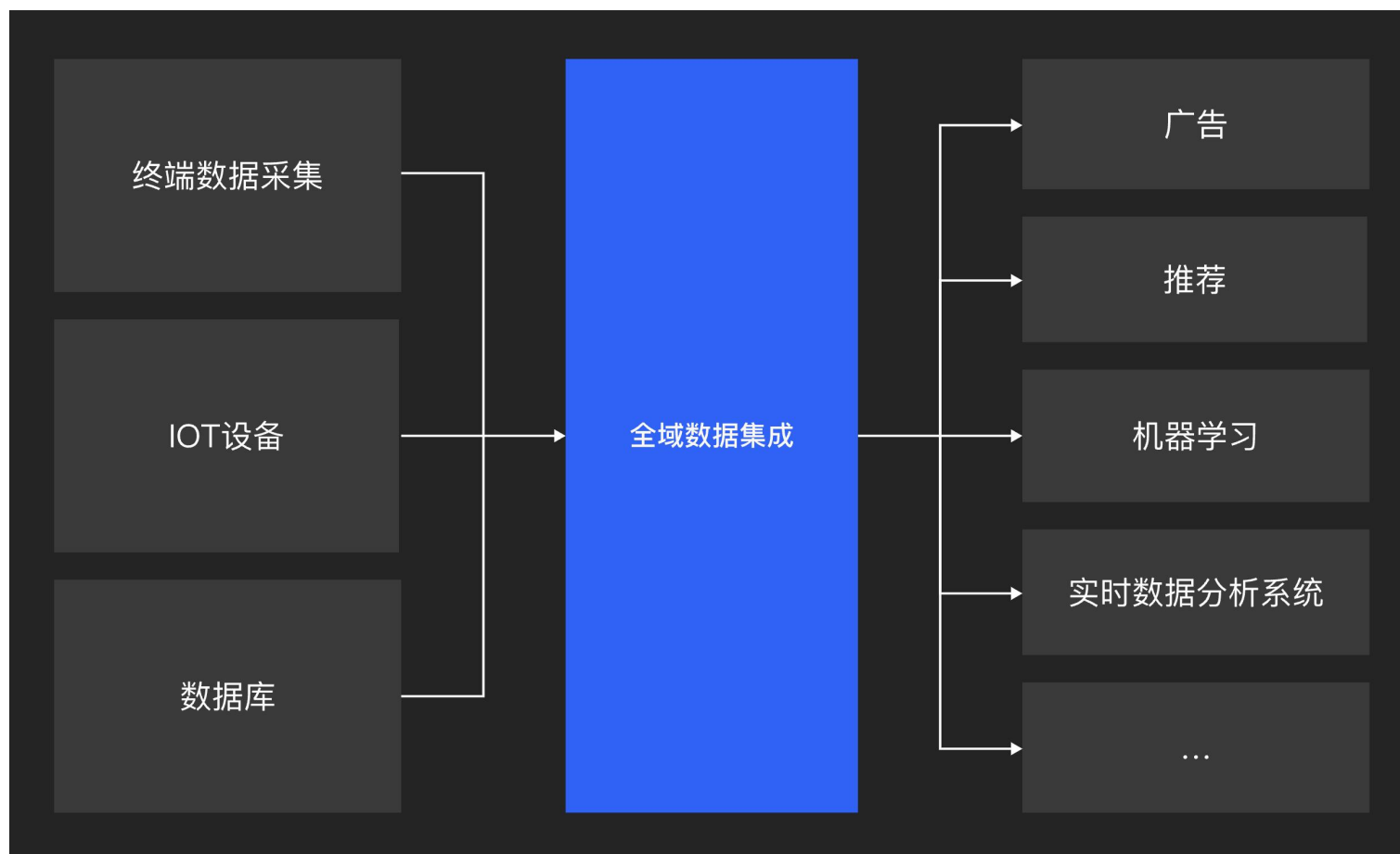
# 议程

1. BitSail 介绍
2. BitSail 亮点功能
3. BitSail 流批一体Connector详解
4. DataSail CDC能力介绍
5. 问答环节

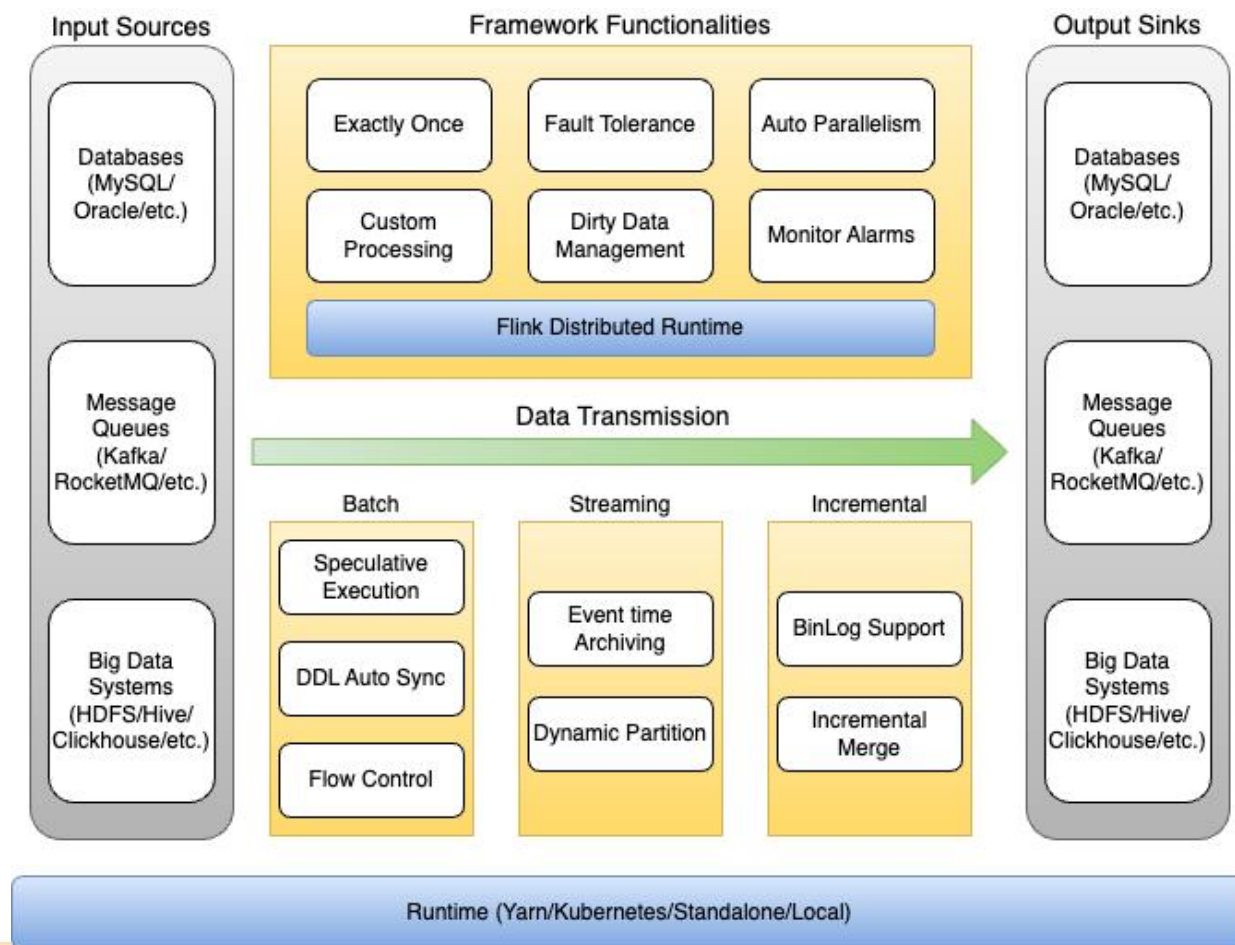
# 什么是数据集成

出于某种业务需要，需要把  
数据从A系统导入到B系统

将数据传输的复杂度从 $M*N$   
降低到 $M+N$



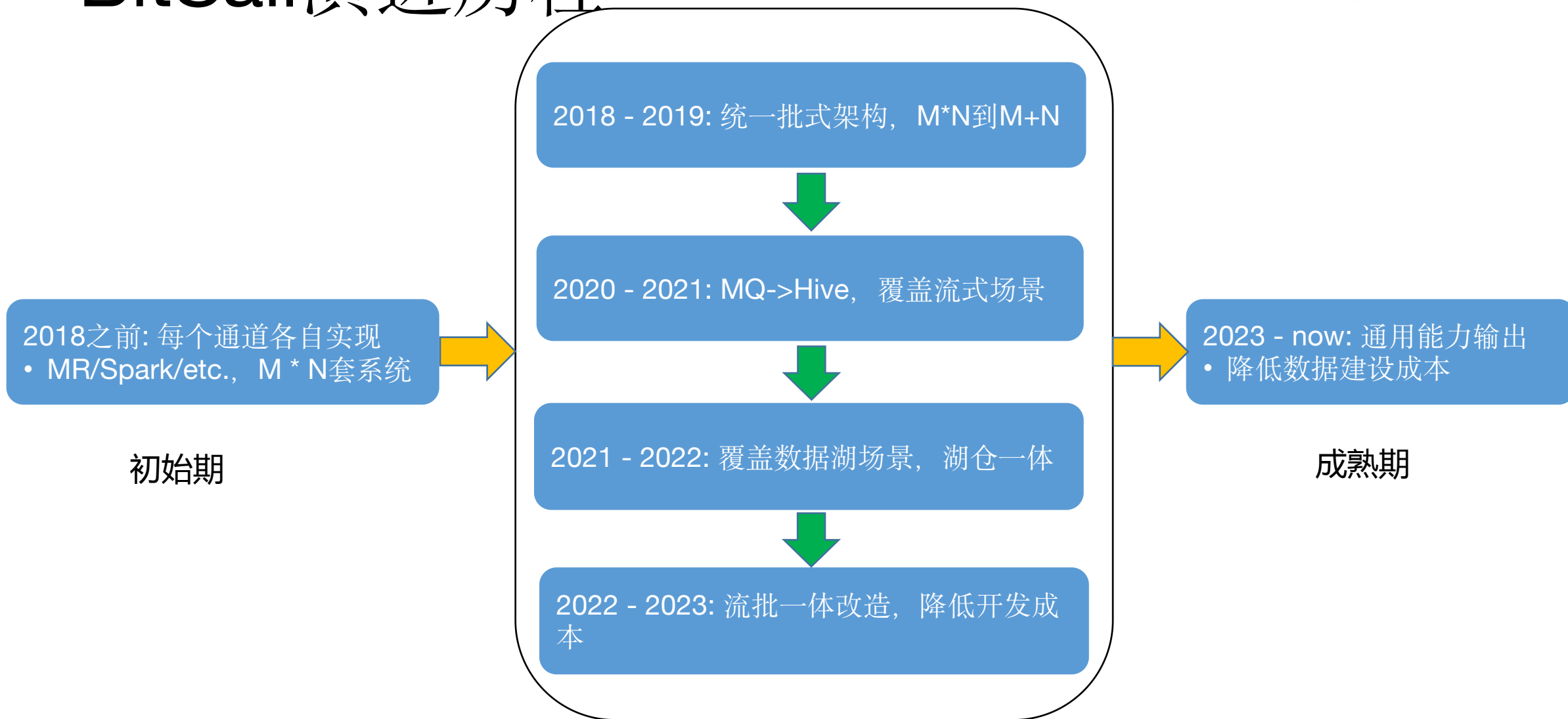
# BitSail介绍



DataSource	Sub Modules	Reader	Writer
Assert	-		✓
ClickHouse	-	✓	-
Doris	-		✓
Druid	-		✓
Elasticsearch	-		✓
Fake	-	✓	
FTP/SFTP	-	✓	
Hadoop	-	✓	✓
HBase	-	✓	✓
Hive	-	✓	✓
Hudi	-	✓	✓
LocalFileSystem	-	✓	
JDBC	MySQL		
	Oracle	✓	✓
	PostgreSQL		
	SqlServer		
Kafka	-	✓	✓
Kudu	-	✓	✓
LarkSheet	-	✓	
MongoDB	-	✓	✓
Print	-		✓
Redis	-		✓
RocketMQ	-		✓
SelectDB	-		✓



# BitSail演进历程



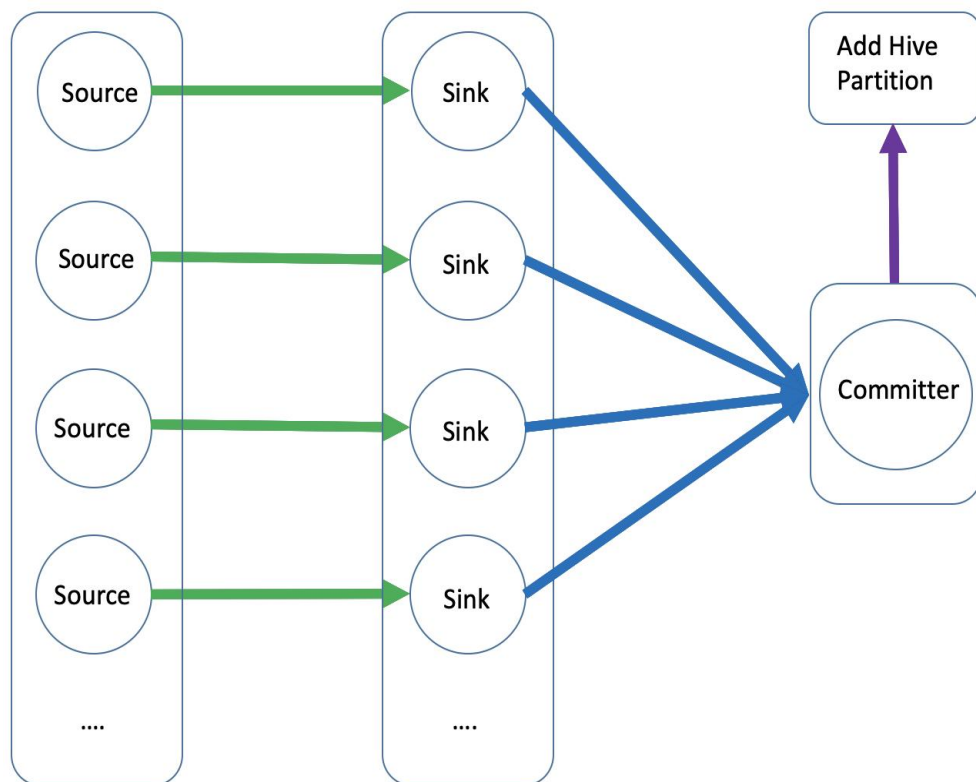
初始期

成熟期

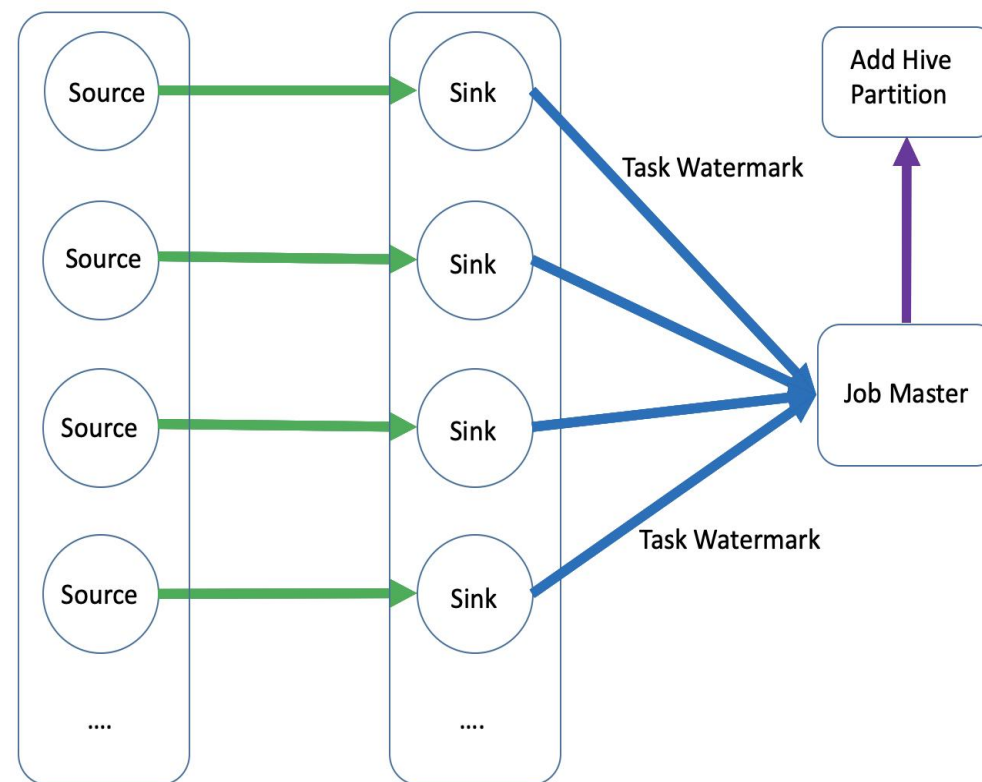
成长期

# BitSail亮点功能

# MQ2Hive写入流程优化



Shuffle



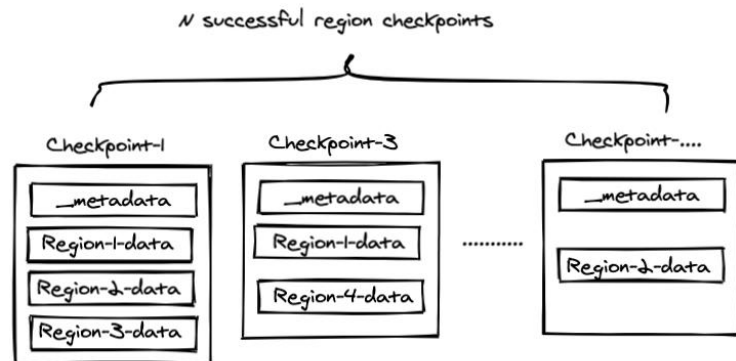
Pipelined



# MQ2Hive写入流程优化

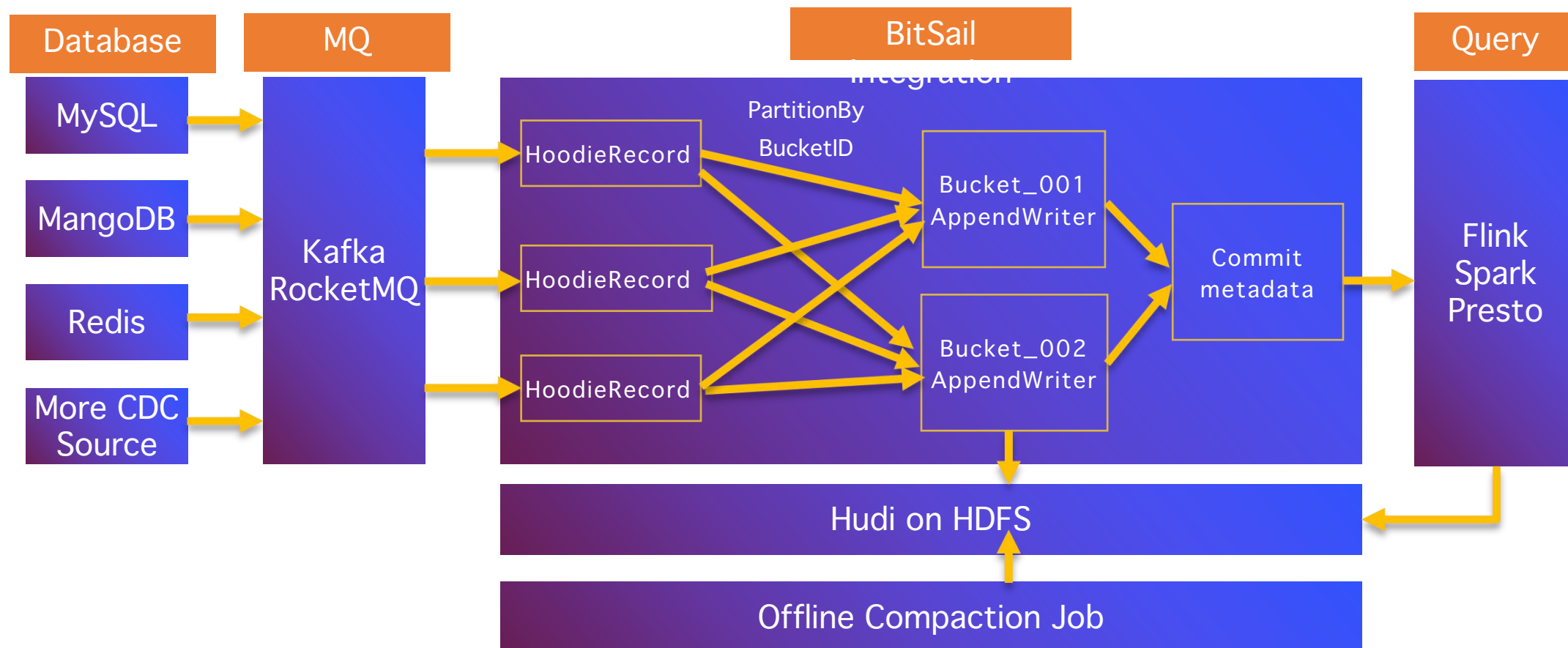
## Regional Checkpoint

- Pipeline链路每个task分别做Checkpoint
- 如果有task失败，不影响其他task
- 任务重启获取Checkpoint时，合并过去N个Checkpoint获得全集
- 大并发场景下任务稳定性大幅提升

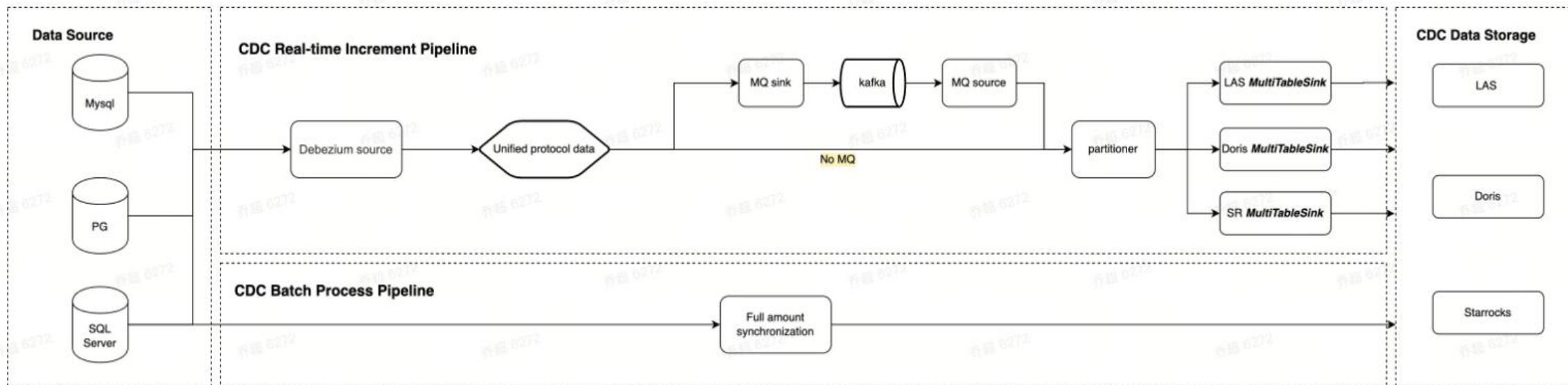


并发	单并发成功率	Global Checkpoint	Region Checkpoint
5000	99.99%	$\text{Math.pow}(99.99\%, 5000) = 60.65\%$	$\text{Math.pow}(1 - \text{Math.pow}(0.001\%, 2), 5000) = 99.9999\%$
10000	99.99%	$\text{Math.pow}(99.99\%, 10000) = 36.79\%$	$\text{Math.pow}(1 - \text{Math.pow}(0.001\%, 2), 10000) = 99.9999\%$

# BitSail实时数据湖架构



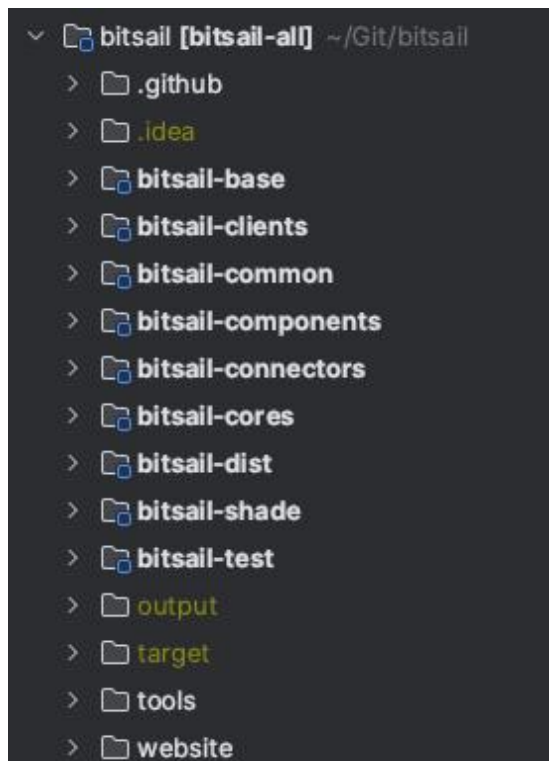
# BitSail CDC架构



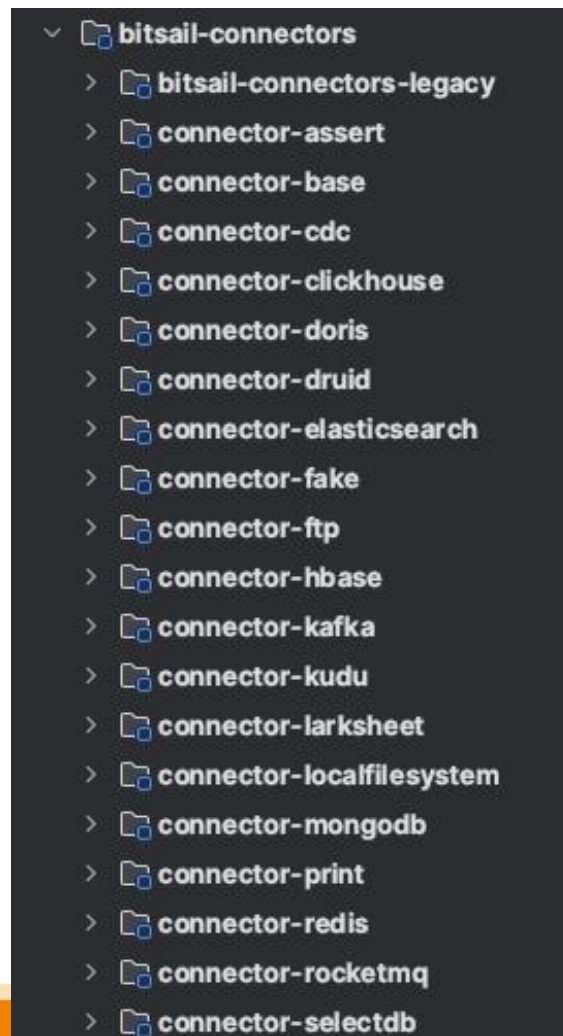
# BitSail Connector详解

# BitSail 代码结构

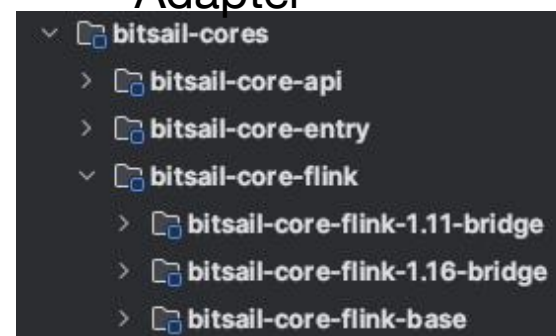
Root



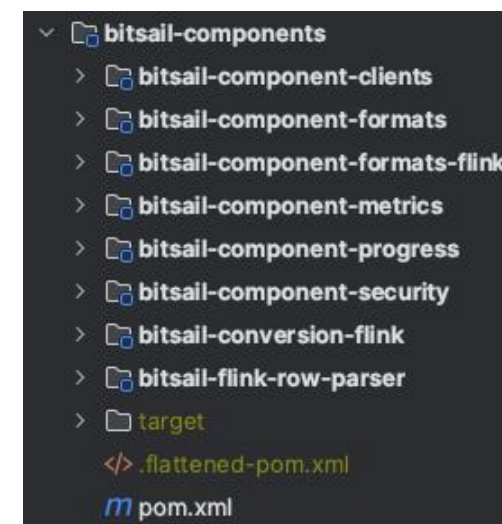
Connector



Engine  
Adapter



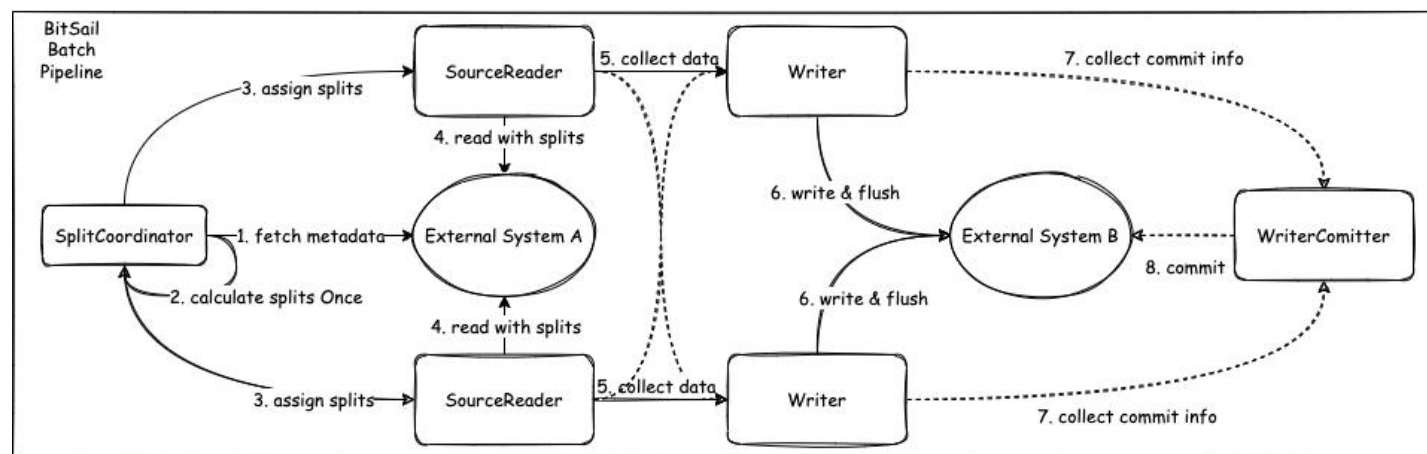
Components





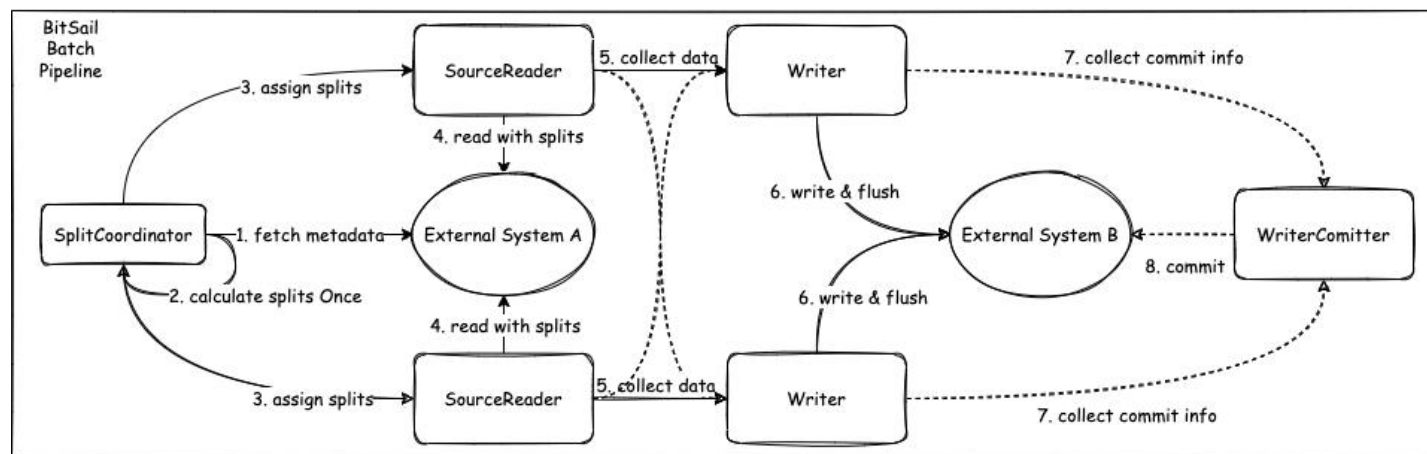
# BitSail Source 连接器交互流程

- **Source:** 数据读取组件的生命周期管理类，主要负责和框架的交互，构架作业，它不参与作业真正的执行；
- **SplitCoordinator:** 既然提到了Split，就需要有相应的组件去创建、管理Split；SplitCoordinator承担了这样的角色；
- **Split:** 数据读取分片；大数据处理框架的核心目的就是大规模的数据拆分成多个合理的Split；
- **SourceReader:** 真正负责数据读取的组件，在接收到Split后会对其进行数据读取，然后将数据传输给下一个算子；



# BitSail Sink连接器交互流程

- **Sink:** 数据写入组件生命周期的管理类，同样与框架进行交互，负责构建作业，它不参与到作业真正的执行；
- **Writer:** 主要负责将接收到的数据写到外部存储；
- **WriterCommitter(可选):** 用于对数据进行提交操作，如果对数据的准确性有要求，则依赖WriterCommitter来完成两阶段提交的操作；实现exactly-once的语义；

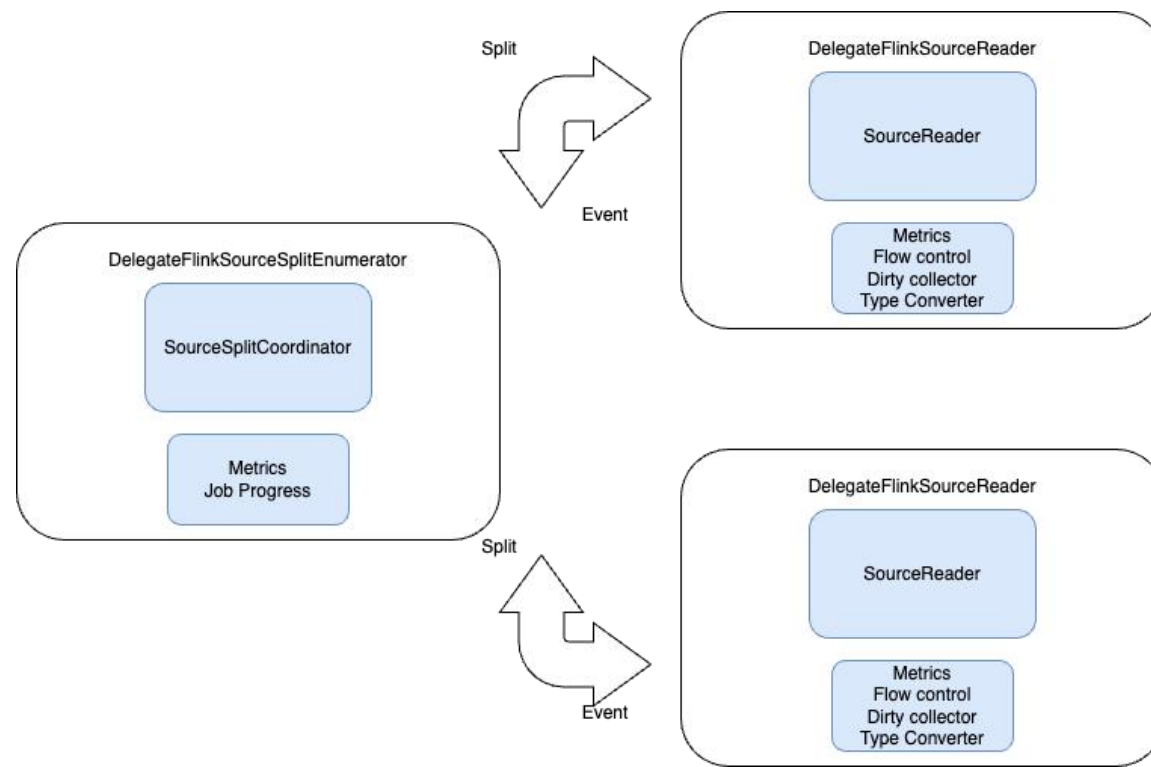
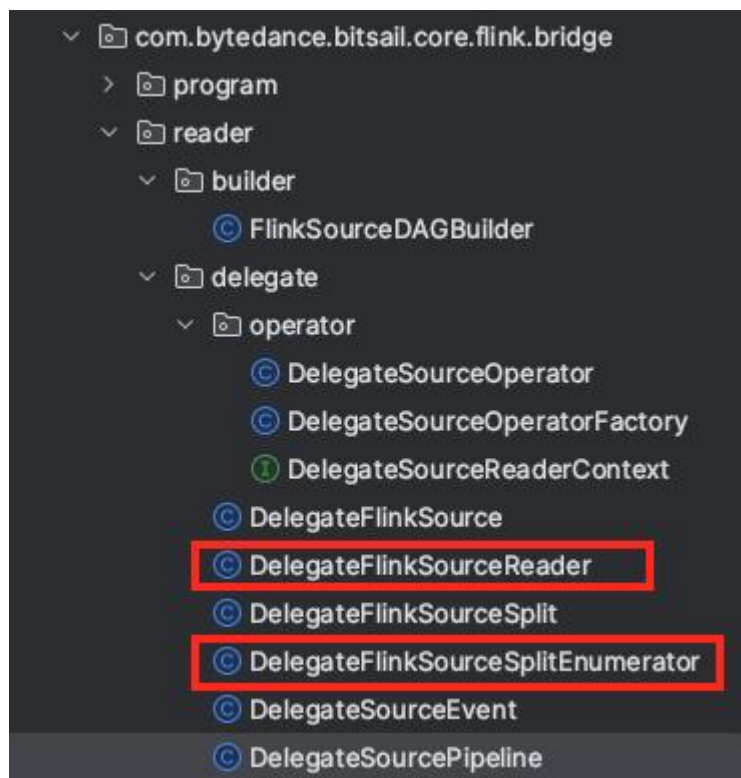


# BitSail Source 与引擎交互流程

DTCC 2023

第十四届中国数据库技术大会  
DATABASE TECHNOLOGY CONFERENCE CHINA 2023

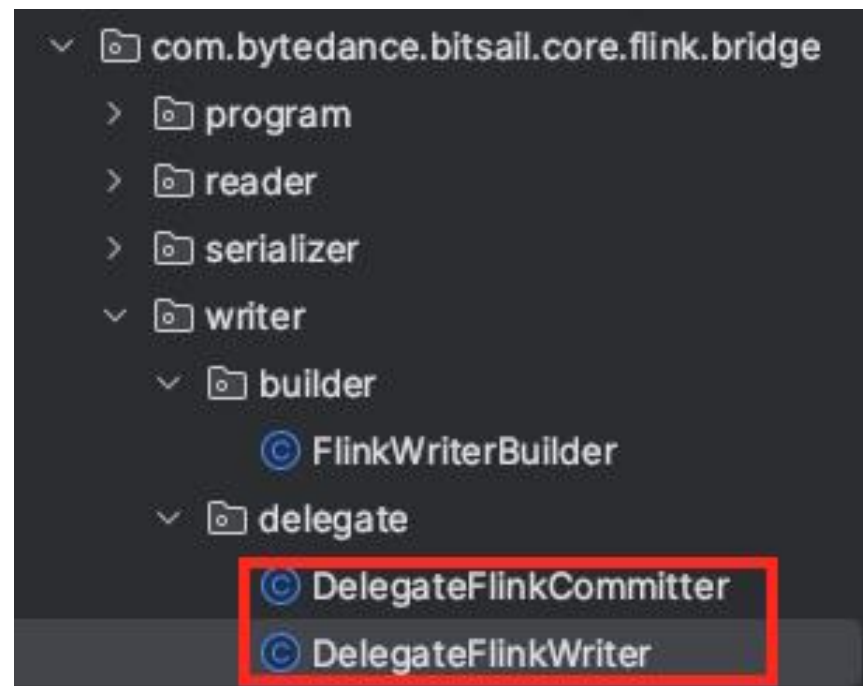
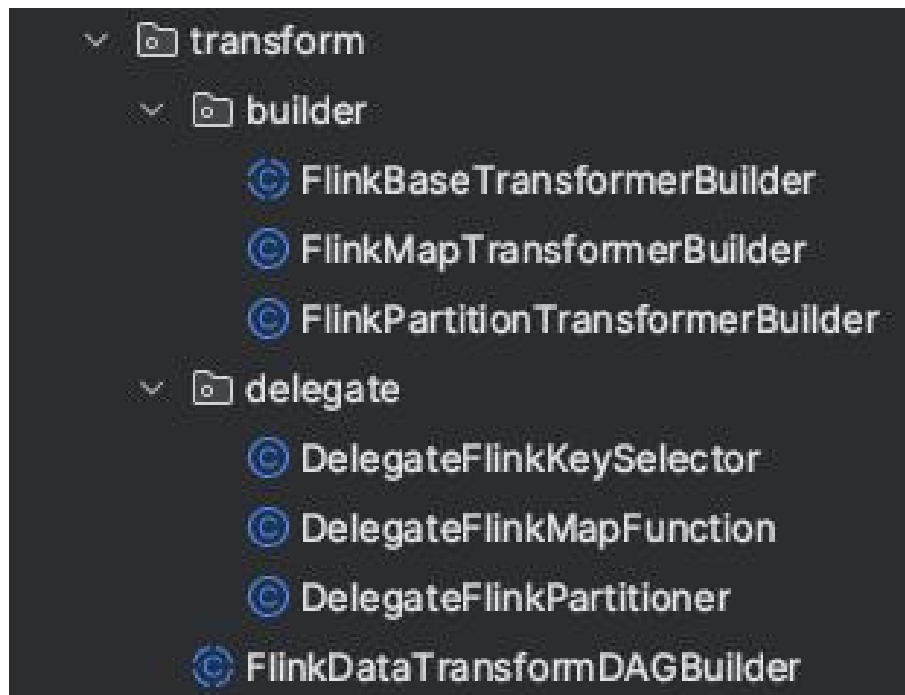
- DAGBuilder负责DAG图生成，DelegateXXX负责和引擎之间的接口翻译



# BitSail Transformer和Sink

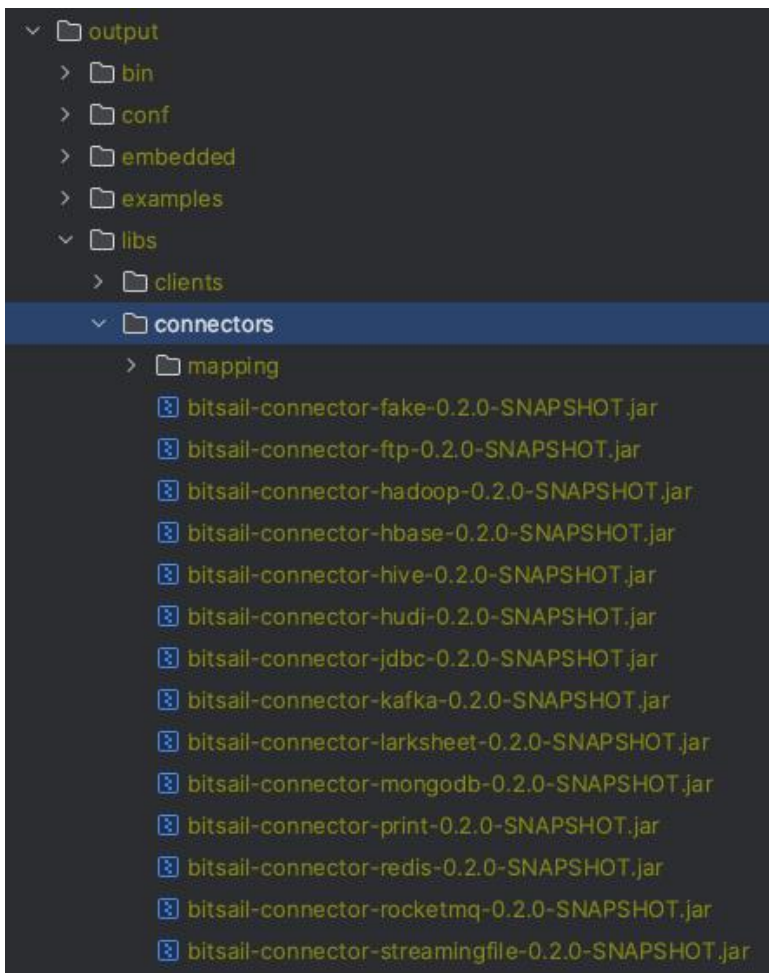
BitSail支持轻量级的Transform，主要场景有数据加密，过滤，和数据shuffle重分布。

Sink connector和Source类似，通过代理的模式将Flink Checkpoint机制对接BitSail，并实现两阶段提交





# 连接器包动态加载



遇到的问题：

Connector数量太多，每次提交任务jar包太大，并且有jar包冲突。

解决方案：

每个connector将会被独立打包在提交任务时，跟进当前任务所需要的connector，动态加载jar包提交任务



# DataSail CDC能力介绍

# 业界CDC相关项目

CDC ( Change data capture )：通过捕获数据库变更日志 ( binlog ) 的方式，将数据库的更新数据同步到外部数据源。相比于批式变更同步，CDC变更同步有着延迟更低，对线上数据库影响更小的优点。

常见开源 CDC 方案比较

	Flink CDC	Debezium	DataX	Canal	Sqoop	kettle	Oracle Goldengate
CDC 机制	日志	日志	查询	日志	查询	查询	日志
增量同步	✓	✓	✗	✓	✓	✗	✓
断点续传	✓	✓	✗	✓	✗	✗	✓
全量同步	✓	✓	✓	✗	✓	✓	✓
全量+增量	✓	✓	✗	✗	✓	✗	✓
架构	分布式	单机	单机	单机	分布式	分布式	分布式
Transformation	★★★★★	★★	★★	★★	★★	★	★
生态	★★★★★	★★★	★★★	★★★	★★	★★	★★★



Apache Flink



TiDB | COMMUNITY



Apache Flink

# 为什么自己做？

1. Source connector无法解决所有的问题，CDC需要端到端的解决方案

- DDL自动同步

2. 更灵活的部署方案

- 同时支持带MQ和不带MQ的部署方式，用户可以根据自身情况选择

3. 和Flink引擎解耦，未来更多拓展空间

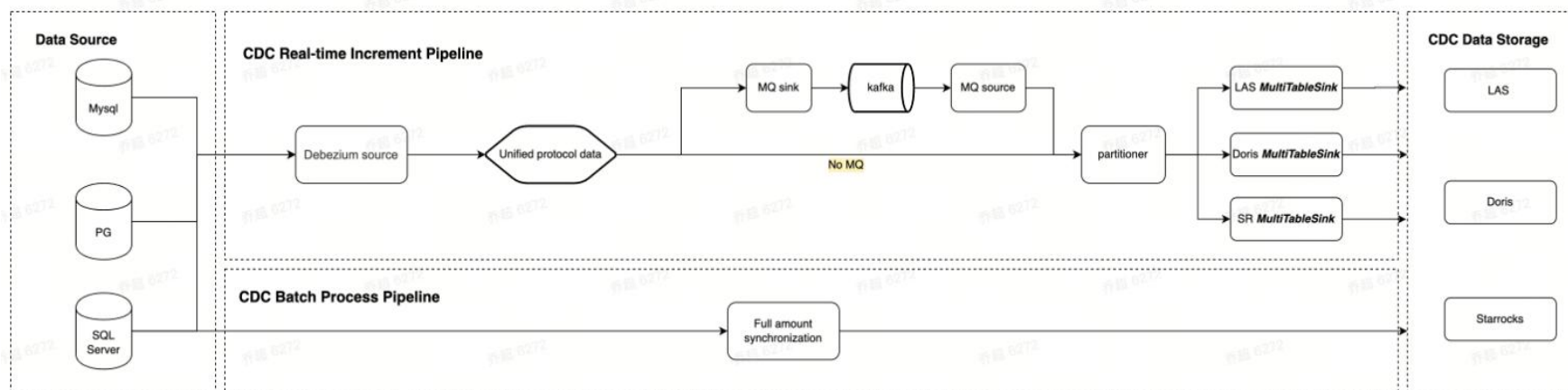
- CDC解决方案不与引擎绑定，未来可以根据业界技术发展切换底层引擎

4. 深入底层优化

- 企业级应用，未来对性能的诉求会越来越高，留足迭代优化的空间

# DataSail CDC架构

- DataSail: 火山引擎全域数据集成产品，提供了多源数据集成，统一数据管理，智能管理运维等功能，**集成引擎使用开源的BitSail。**
- DataSail CDC解决方案：整库迁移支持通过批量创建同步任务，如把一个Mysql DB库内所有表一并上传到下游数仓中；整库CDC支持一键同步一个库内的多张表，先进行全量同步，后进行CDC实时同步



DataSail：全量增量一体化数据同步

Doris/Starrocks：向量化查询应对多表Join、聚合分析

LAS：湖仓一体分析

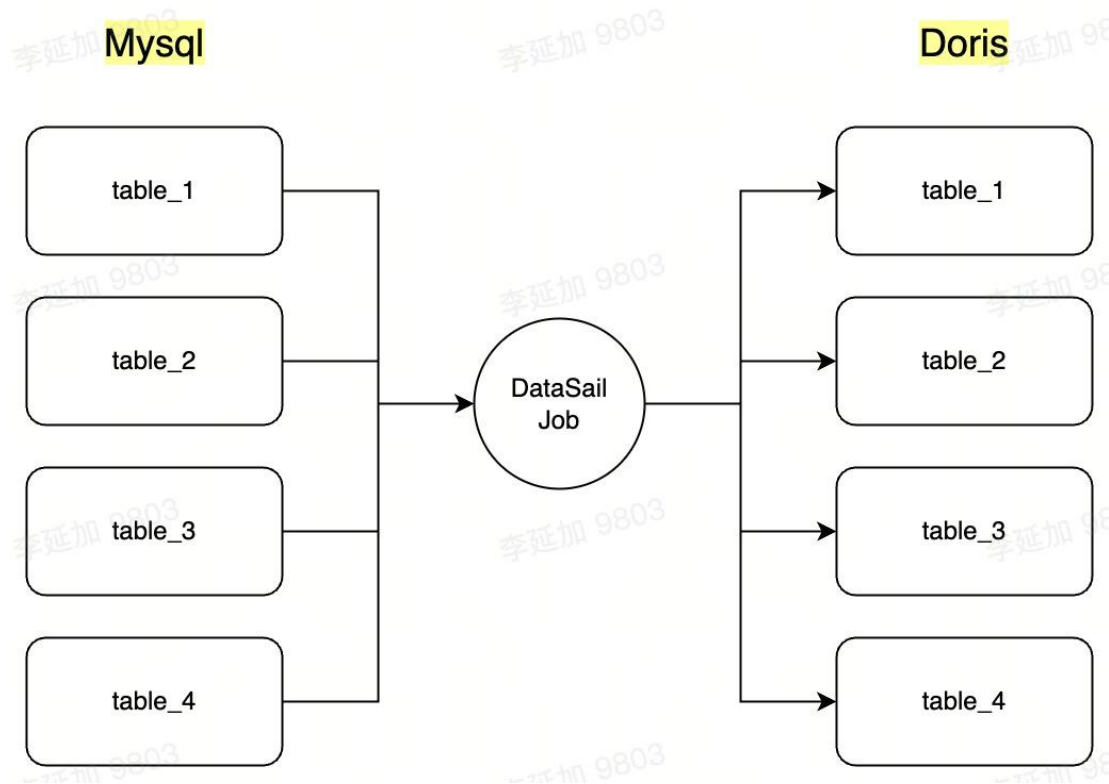
# 整库同步

## • 痛点

- 一张表对应一个任务，一个库需要上百个任务
- 目标端建表耗时耗力，易出错

## • 我们的方案

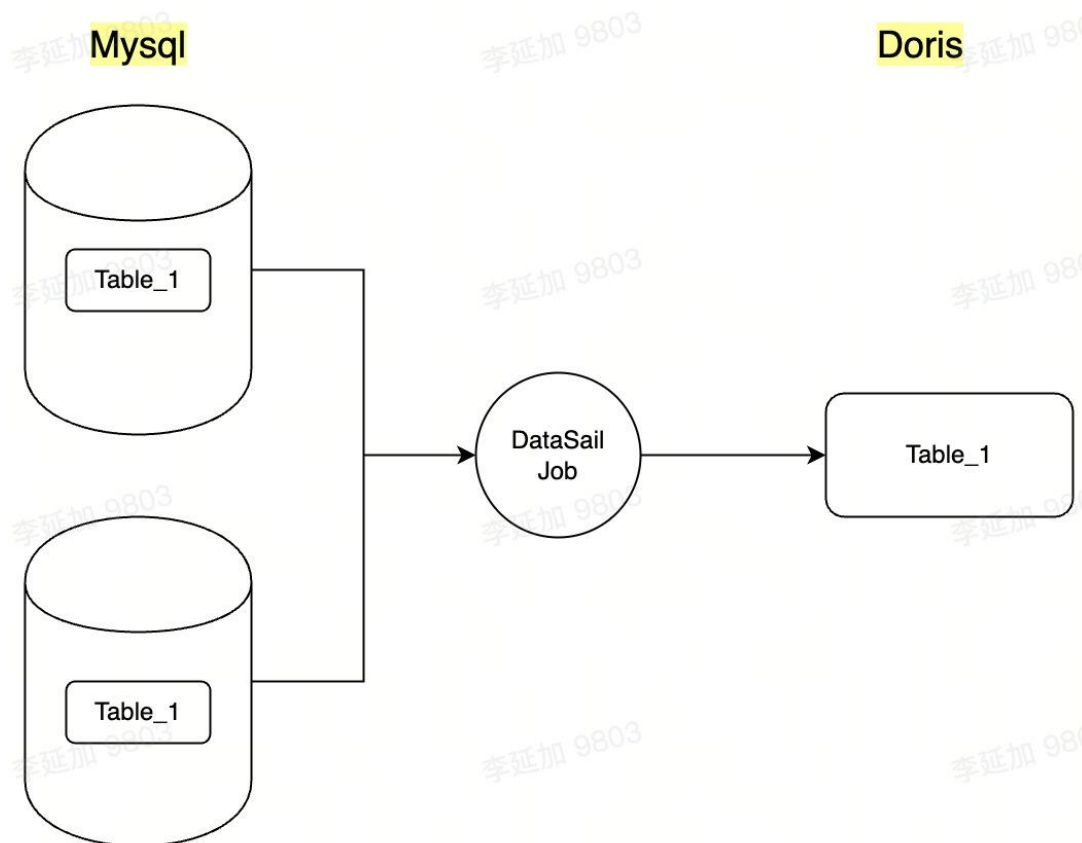
- 产品层调度整体流程，引擎单个任务写入整库数据到目标端
- 自动建表，自动字段映射





# 分库分表同步

- 痛点
  - 基于业务逻辑分表，无法导入数仓中的同一张表
  - 规则多变
- 我们的方案
  - 批式单任务并发读取多库
  - 流式Binlog基于正则过滤表名
  - 自定义N张表直接入仓



# 乱序数据处理

- 痛点
  - Mysql切主，长事务等问题导致
  - Binlog读取任务重试，导致数据回退
  - 中间缓存MQ扩容导致Binlog顺序难以保证
- 我们的方案
  - 严格递增的排序字段，适用于所有CDC数据源
    - 数据源原生字段，例如mysql的filename和binlog position
    - DataSail采集任务生成的全局递增字段
  - 下游数仓配置排序字段，保证新数据不被旧数据覆盖
    - 例如Hudi的precombine字段，doris的sequence字段

# DataSail CDC未来计划

- 2023Q4~2024Q1
  - 自动DDL同步，支持Create Table和Add Column
  - 支持单并发写入多张表，优化资源利用率
  - 当前接入更多CDC数据源

# THANKS

TDDL

DistributedTable

DBproxy

HBase

PostgreSQL

SSD

MongoDB

Cassandra

GreatDB

Hyperbase

Hubble

DataCenter

VisualDataPlatform

Blockchain

ArgoDB

Distributed

DatabaseKernel

TemporalData

CloudnativeData

AIalgorithm