京东实时数据仓库开发实践

王日宇

2020年10月31日

京东

1 传统数据仓库面临的挑战

2 京东实时数据湖的探索和经验

3 Delta lake核心原理

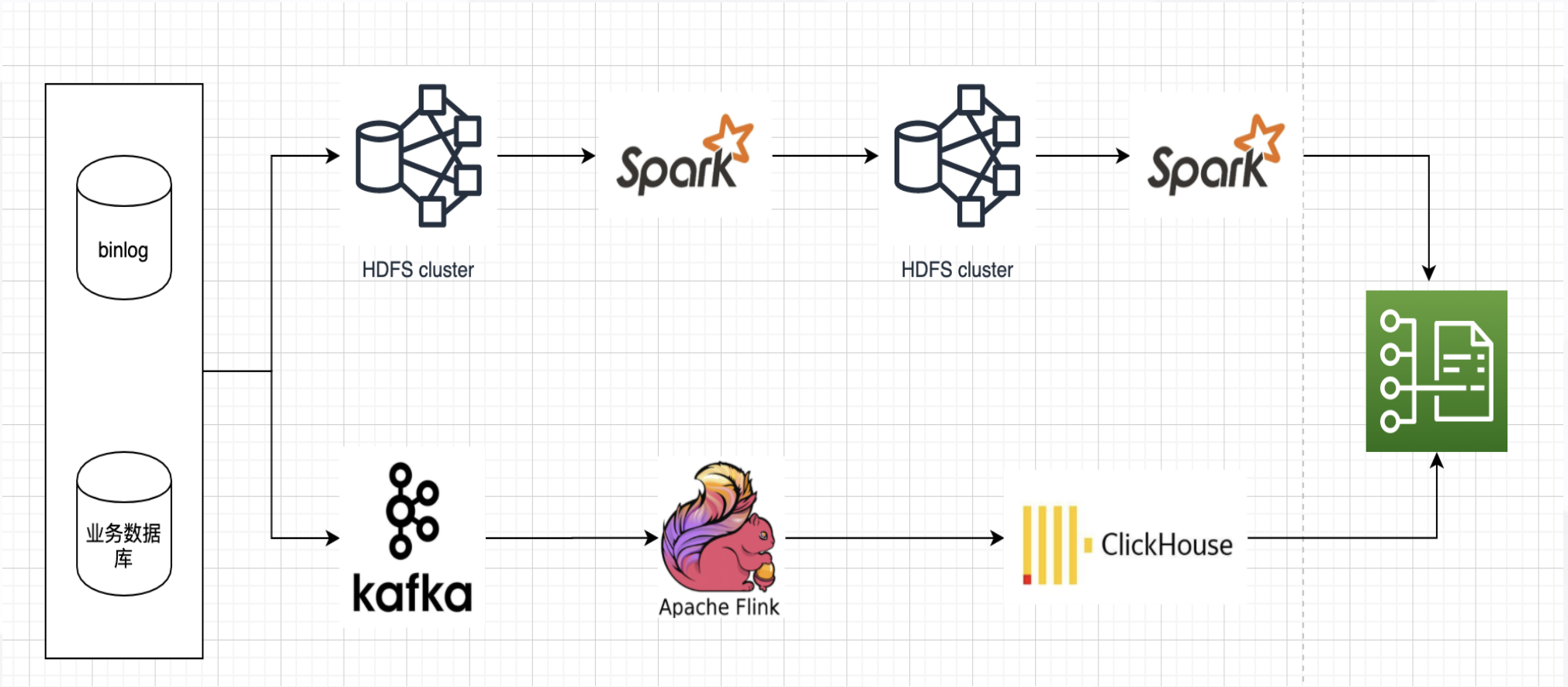
4 批流一体开发流程

5 优缺点和总结

京东

传统数据仓库面临的挑战

Lambda 架构：



传统数据仓库面临的挑战

1. ACID语义性无法保证
2. 离线入库有潜在的不可靠性
3. 细粒度的数据更新功能缺失
4. 数据流转路径复杂

京东

京东实时数据湖的探索和经验

自研方案 → 合并社区优秀经验

2020-10-28	Delta	Hudi	Iceberg
Open source time	2019-04-12	2018-11-06	2019-01-17
company	databricks	uber	netflix
watch	176	1.3k	83
star	2.8k	1.5k	772
fork	633	632	301
issues	205	645	343
PR	171	1465	1042
commits	476	1199	1145
contributors	76	122	102

京东

京东实时数据湖的探索和经验

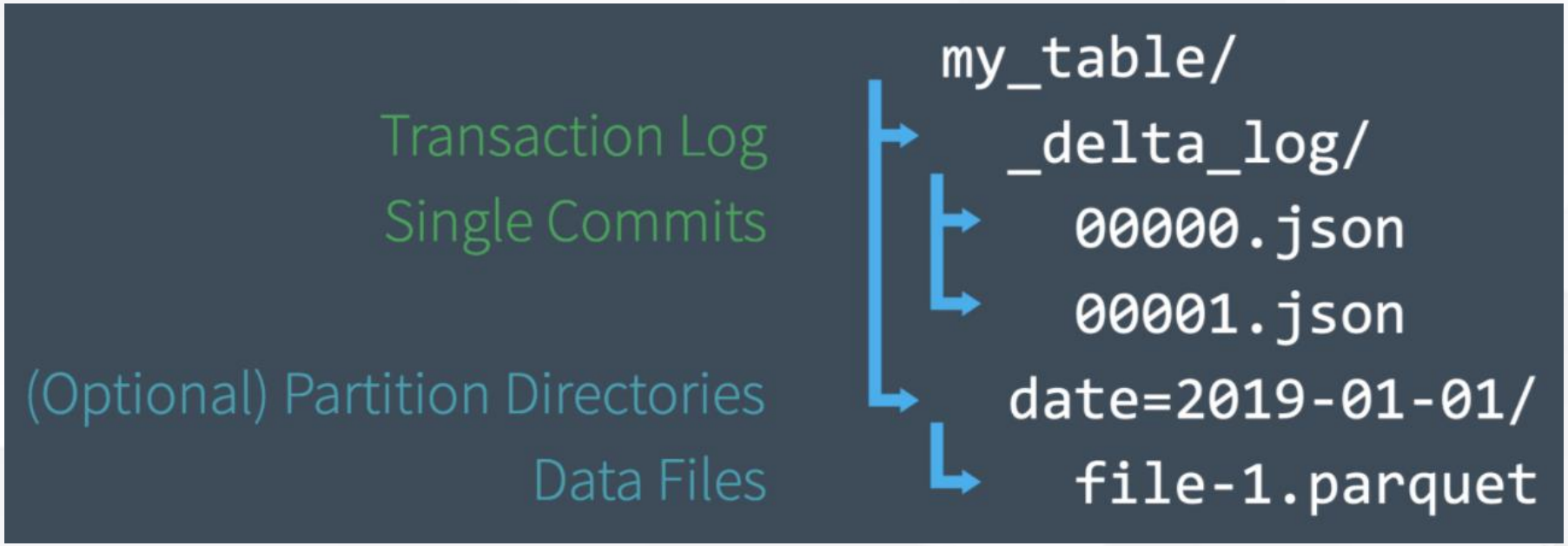
	Delta	Hudi	Iceberg
ACID	Y	Y	Y
Time Travel	Y	Y	Y
MVCC	Y	Y	Y
Schema Evolution	Y	Y	Y
Update/Delete	Y	Y	N
Streaming sink	Spark	DeltaStreamer(Spark)	Spark
Streaming source	Spark	N	N(WIP for Spark/Flink)
engine support	Spark/Presto/Hive	Spark/Presto/Hive	Spark/Presto

京东

Delta lake核心原理

Delta Lake is an open-source storage layer that brings ACID transactions to Apache Spark™ and big data workloads.

delta数据表：数据文件和事务日志



京东

Delta lake核心原理

Transaction log的内容：


```
{"commitInfo":{"timestamp":1600071805932,"operation":"STREAMING UPDATE","operationParameters":{"outputMode":"Append","queryId":"a144cf0b-ede0-4ebb-82c4-0725f0b47f28","epochId":"1071"},"
Append":true,"operationMetrics":{"numRemovedFiles":"0","numOutputRows":"22912","numOutputBytes":"8731453","numAddedFiles":"8"}}}
{"txn":{"appId":"a144cf0b-ede0-4ebb-82c4-0725f0b47f28","version":1071,"lastUpdated":1600071805931}}
{"protocol":{"minReaderVersion":1,"minWriterVersion":2}}
{"metaData":{"id":"0c628905-315e-4266-a794-9437f6e4a766","format":{"provider":"parquet","options":{}},"schemaString":"{\"type\":\"struct\",\"fields\":[{\"name\":\"id\",\"type\":\"long\"
data\":{}}]}","partitionColumns":[],"configuration":{},"createdTime":1595846262428}}
{"add":{"path":"part-00000-53154cb7-664c-49ed-a9be-0e86b36427a9-c000.snappy.parquet","partitionValues":{},"size":1107391,"modificationTime":1600071805648,"dataChange":true}}
{"add":{"path":"part-00001-9c8b68c6-47aa-4ae4-bac5-6e1a2e1d12a1-c000.snappy.parquet","partitionValues":{},"size":1209554,"modificationTime":1600071805614,"dataChange":true}}
{"add":{"path":"part-00002-7d2a67cd-e153-48b9-9233-68a20b89fad1-c000.snappy.parquet","partitionValues":{},"size":1075446,"modificationTime":1600071805653,"dataChange":true}}
{"add":{"path":"part-00003-7c14441b-12c9-4c75-8941-8c7bd2000caf-c000.snappy.parquet","partitionValues":{},"size":1040188,"modificationTime":1600071805938,"dataChange":true}}
{"add":{"path":"part-00004-baaac8e1-41b1-456c-a7e1-0e4ac96aa588-c000.snappy.parquet","partitionValues":{},"size":1030761,"modificationTime":1600071805484,"dataChange":true}}
{"add":{"path":"part-00005-2c21d6d5-a767-469d-a876-40e4834ffb0d-c000.snappy.parquet","partitionValues":{},"size":1163016,"modificationTime":1600071805619,"dataChange":true}}
{"add":{"path":"part-00006-340f1673-f0d3-4f48-9852-8cb12cc68a46-c000.snappy.parquet","partitionValues":{},"size":1006224,"modificationTime":1600071805647,"dataChange":true}}
{"add":{"path":"part-00007-5a1f75f1-b76f-4cf1-82c1-c976ca2a6dc2-c000.snappy.parquet","partitionValues":{},"size":1098873,"modificationTime":1600071805624,"dataChange":true}}
```

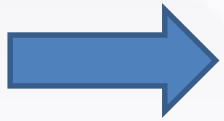

1. Commit的基本信息：when，who，how
2. 涉及到的具体的文件路径和统计信息
3. 表的Metadata信息，字段名，字段类型，文件格式，配置属性等

京东

Delta lake核心原理

比如一个表日志的变更历史如下：

000000.json
000001.json
000002.json
…
000010.json
000010.checkpoint.parquet
000011.json
000012.json
_last_checkpoint



Delta数据表读取流程：

1. 使用_last_checkpoint找到最近的checkpoint文件
2. 找到checkpoint版本之后的json log文件
3. 合并所有json log和checkpoint log的记录，得到数据表在该版本状态下包含哪些具体的数据文件

京东

Delta lake的特点

- ✓ 支持批流读写
- ✓ 提供ACID语义
- ✓ Update/delete的支持
- ✓ 历史版本回溯和审计
- ✓ 抽象存储接口
- ✓ 查询性能提升

京东

批流一体开发流程

京东

优缺点和总结

1. 有很多优秀特性是闭源的，如直接使用SQL进行版本回溯，DFP，Z-Ordering等
2. 小文件和历史文件的清理
3. Connector的支持
4. 计算引擎和使用方式的支持

感谢您的时间

Thanks.

京东