

基于StarRocks和Paimon 打造湖仓分析新范式

 王日宇

StarRocks Committer

- 01 StarRocks湖仓分析的发展历程
 - 02 使用StarRocks+Paimon构建湖仓分析方案介绍
 - 03 使用StarRocks+Paimon构建湖仓分析方案的关键技术原理介绍
 - 04 使用StarRocks+Paimon的未来技术规划
- 

01 StarRocks湖仓分析的发展歷程

StarRocks湖仓分析的发展历程

方案一：增量 + 全量

面向批处理、面向 Spark



增量

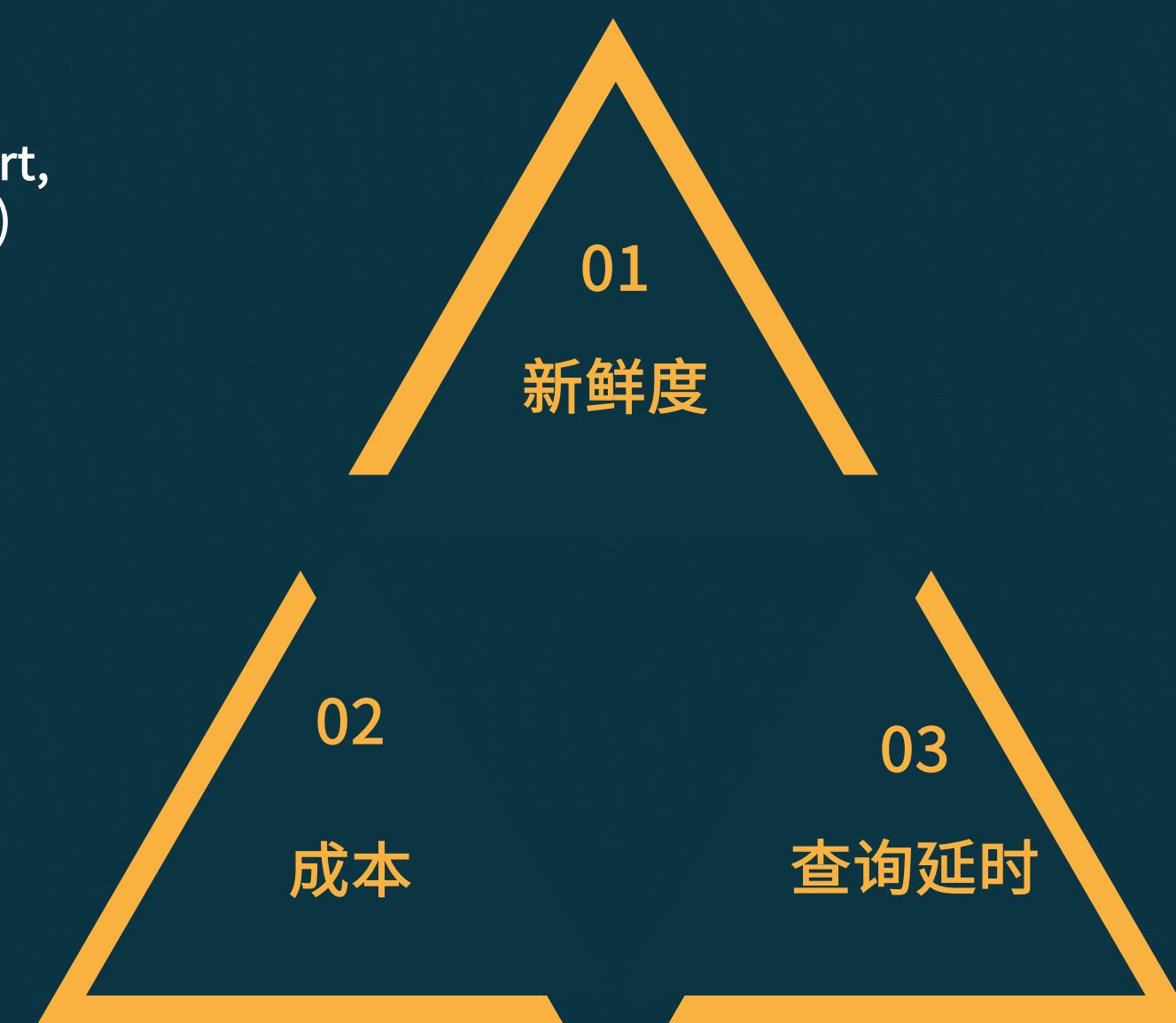
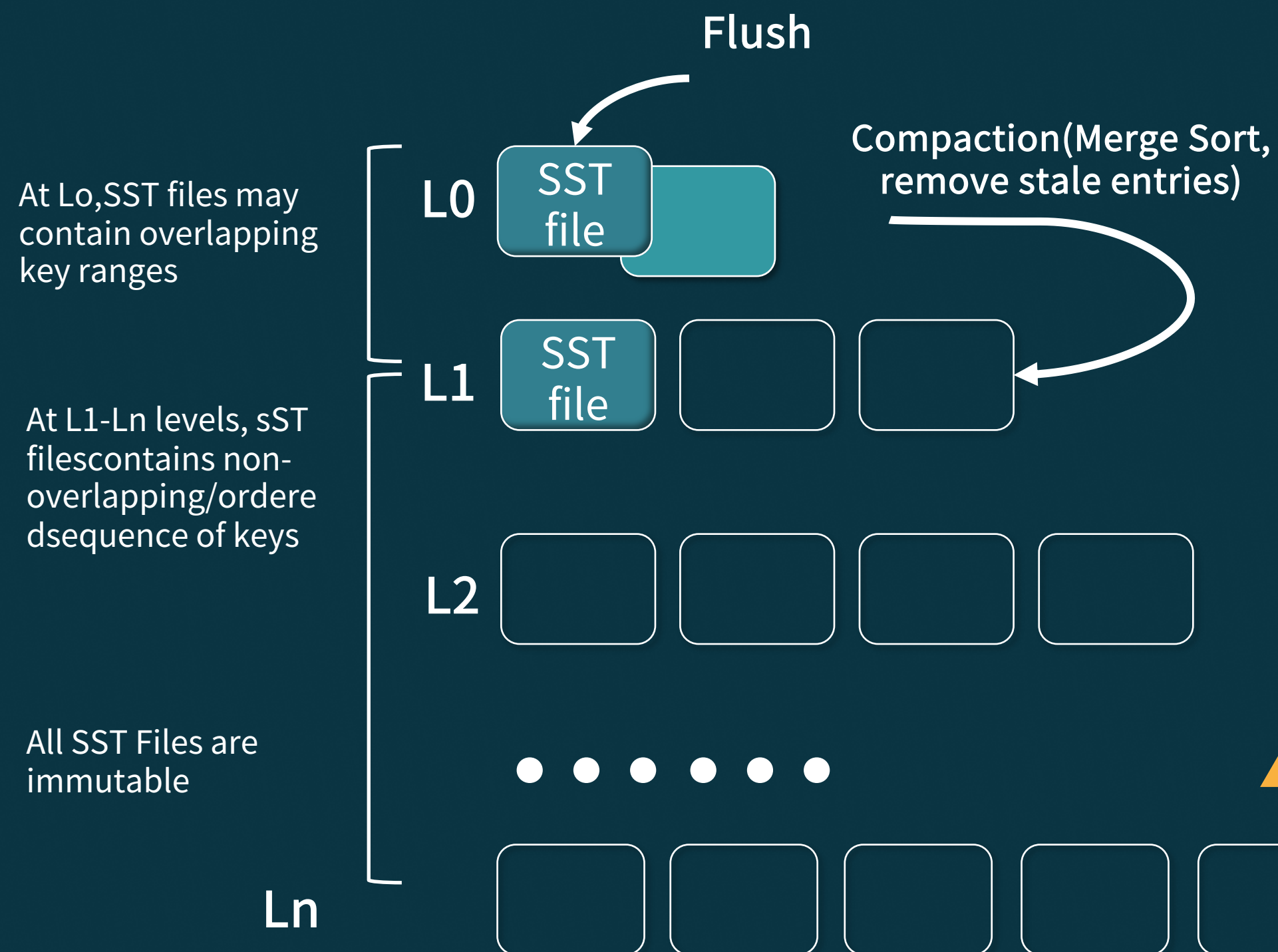
Position Delete File
Equality Delete File
(列存或行存)

+

全量

Data File
(Parquet / ORC)

方案二：LSM (RocksDB, HBase, Clickhouse)



StarRocks湖仓分析的发展历程

湖仓重大特性

Hive/ES/MySQL 外表
StarRocks外表

Catalog数据目录
Iceberg/Hudi/Delta Lake外表
JNI Connector
外表物化视图
Json/Map/Struct类型

Paimon外表
JNI Connector复杂类型支持
算子落盘Spill
Trino兼容性
外表物化视图分区刷新
Hive/Iceberg数据写入

StarRocks 1.x

StarRocks 2.x

StarRocks 3.x

向量化执行引擎
CBO优化器
Runtime Filter

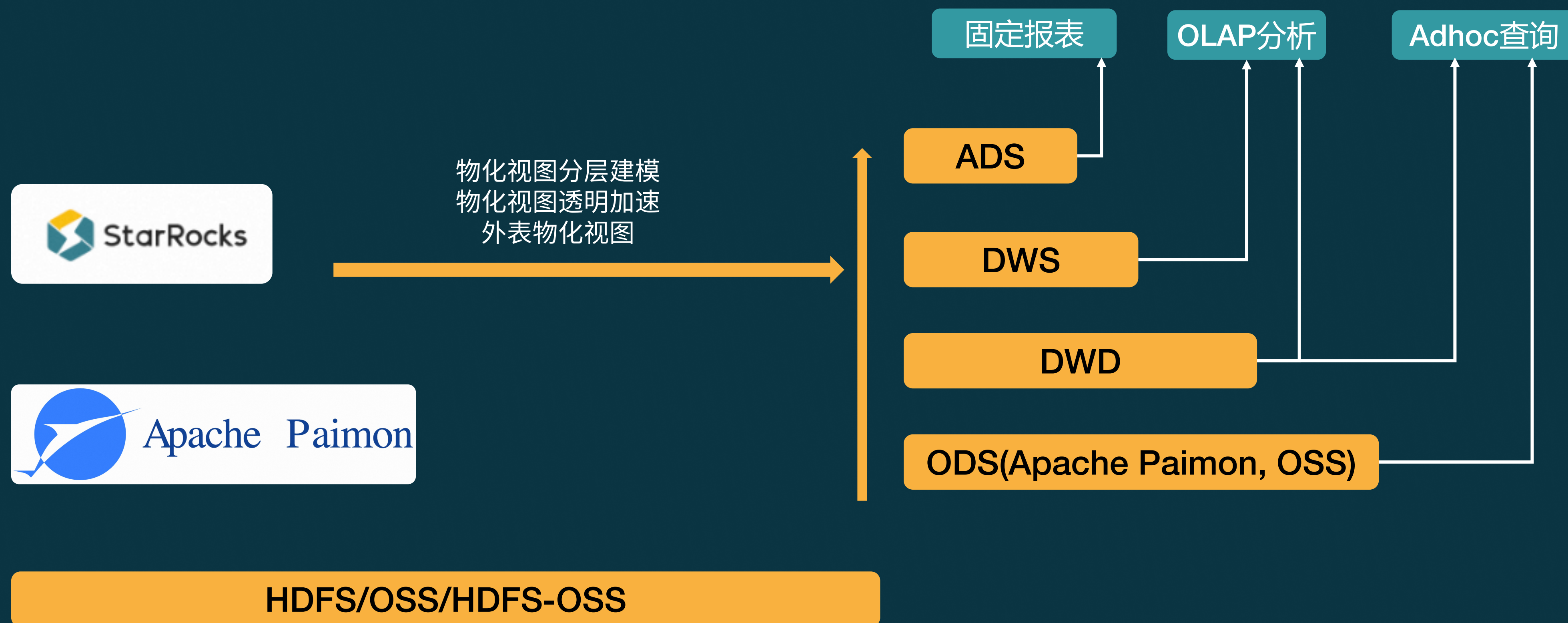
IO性能优化
延迟物化
Pipeline执行引擎
直方图统计信息

Data Cache
JNI Connector 性能优化

底层性能持续优化

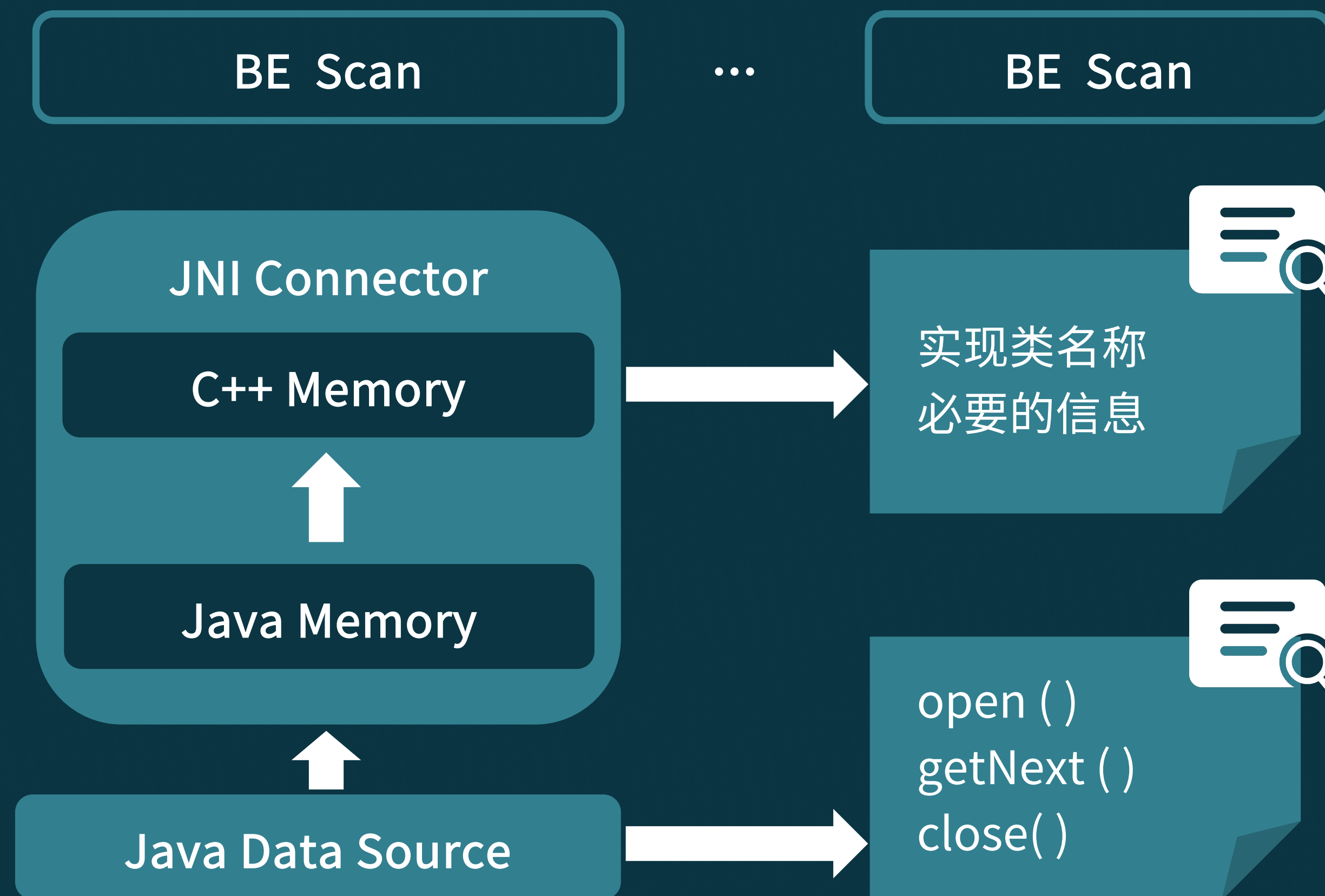
02 使用StarRocks+Paimon构建湖 仓分析方案介绍

使用StarRocks+Paimon构建湖仓分析方案介绍



03 使用StarRocks+Paimon构建湖仓分析方案的关键技术原理介绍

使用StarRocks+Paimon构建湖仓分析方案的关键技术原理介绍



JNI Connector

- 1 快速接入各类Java数据源，无需考虑数据转换
- 2 提供简单易用的Java接口
- 3 已支持Hudi MOR Table, Paimon Table
- 4 支持Struct, Map, Array复杂类型
- 5 BE代码零侵入，不需要考虑C++具体实现

源码地址: <https://github.com/StarRocks/starrocks/pull/9716> (2022年8月)

使用StarRocks+Paimon构建湖仓分析方案的关键技术原理介绍

1 定长字段存储 Null indicator:
| 1-byte boolean | 1-byte boolean | 1-byte boolean | ... |

2 数据 | 4-bytes INT | 4-bytes INT | 4-bytes INT | ... |

3 变长字段存储 Null indicator:
| 1-byte boolean | 1-byte boolean | 1-byte boolean | ... |

4 起始地址 | offset of row 0 | offset of row 1 | offset of row 2 | ... |

5 数据 | (length of row 0)-bytes | (length of row 1)-bytes | ... |

实现
例子

Paimon 读取支持: <https://github.com/StarRocks/starrocks/pull/24626>
Hudi MOR读取支持: <https://github.com/StarRocks/starrocks/pull/6780>

使用StarRocks+Paimon构建湖仓分析方案的关键技术原理介绍



dt=2023/11/13

dt=2023/11/14

dt=2023/11/15

dt=2023/11/16

dt=2023/11/17

对近期时间分区的数据
创建物化视图



dt=2023/11/15

dt=2023/11/16

dt=2023/11/17

查询湖数据，包括冷热数据

select * from <paimon> where dt = 2023/11/13-17 自动改写湖数据查询



Scan MV

Scan Paimon

UNION

业务场景

数据具有冷热特性，
想要加速近期查询

- 物化视图指定 TTL：
partition_ttl=3 month

冷热分区合并

- 数据物化加速：MV 物化近 3 个月数据，历史数据存储在 Paimon 中
- 自动查询改写：无需修改查询，优化器自动改写

04 使用StarRocks+Paimon的 未来技术规划

使用StarRocks+Paimon的未来技术规划

完善使用StarRocks对Paimon 表的分析特性

1

支持Append Only
表类型

2

优化date/datetime类型
的处理效率

3

物化视图支持Paimon
分区刷新

4

使用native reader
加速查询

5

支持
列统计信息

6

支持
元数据缓存

7

支持time travel和
snapshot查询

8

支持Paimon外表的
sink能力

感谢聆听!

Thanks for listening !



扫码入群交流



StarRocks 公众号