

Apache Hudi 0.8.0 版本发布, Flink 集成有重大提升以及支持并行写

迁移指南

- 如果从 0.5.3 以下版本迁移, 请检查这个版本后面的其他版本的升级说明。
- 如果需要升级到 0.8 版本, 请参阅 0.6.0 版本的升级指南, 因为本版本没有引入新的表版本 (table versions)
- HoodieRecordPayload接口不建议使用现有方法, 而推荐使用新方法, 该方法还允许我们在运行时传递属性。
鼓励用户从不建议使用的方法中迁移出来, 因为这些方法将从0.9.0中删除。

Flink 集成

自从 [0.7.0 版本首次支持 Hudi Flink Writer](#) 以来, Hudi 社区在改善 Flink/Hudi 集成方面取得了很大的进展, 包括重新设计具有更好性能和可伸缩性的 Flink writer 管道、Flink 支持写 MOR 表、批量读取 COW&MOR 表、流式读取 MOR 表以及使用 Flink SQL 读写 (source and sink) Hudi 的连接器。在 0.8.0 版本中, 用户可以使用 Flink 1.11+ 的所有这些特性。

请参阅 [RFC-24](#) 了解 Flink writer 的更多实现细节, 同时可以参考[这个](#)来快速在 Flink 中使用 Hudi。

并行写支持

根据许多用户的要求, 现在 Hudi 通过乐观并发控制支持多个 writers 到同一个 Hudi 表。Hudi 支持文件级 OCC, 也就是说, 对于发生在同一个表上的任意2个提交(或写入者), 如果没有对相同的文件进行修改, 那么两个写入操作都可以成功。这个特性目前还处于实验阶段, 需要ZooKeeper或HiveMetastore来获取锁。

请参阅 [RFC-22](#) 了解更多实现细节, 同时可以参考[这个](#)来了解 Hudi Tables 的并发控制。

写端的改进

- Flink writer client 支持 InsertOverwrite ;
- Java writer client 支持 CopyOnWriteTable。

查询端的改进

- 支持使用 Spark Structured Streaming 读取 Hudi 表 ;
- Metadata table 的性能有所提升 ;
- Clustering 的性能有所提升。

其他

<https://github.com/apache/hudi/releases/tag/release-0.8.0>

Apache Hudi 0.8.0 Releases Note : <http://hudi.apache.org/releases.html>

本博客文章除特别声明，全部都是原创！
原创文章版权归过往记忆大数据（[过往记忆](#)）所有，未经许可不得转载。
本文链接: [【】](#)（）