

1、测试概述

Hash Join 功能能够支持 Cedar 所支持的所有数据类型的 join 操作，当前若使用该功能需要使用 hint 对两表或多个数据表进行 join 操作，用 hint 指定 join 类型（hash_join）的连接。本测试针对相关模块和可能错误点对 hash join 基本功能进行测试。

2、实验环境

单集群：在一台机器上部署 rs/ups/cs/ms

三集群：在三台机器上分别部署 rs/ups/cs/m

3、实验测试观点说明

- a. 通过 hint 指定某两张或多张表做 hash join。
- b. 对不同数据类型连接的支持：int 型、varchar 型、double 型、bool 型、timestamp 型、decimal 型。
- c. 不同数据类型的列之间相互进行连接查询操作。
- d. 对不同连接类型的支持：inner join、left join、right join、full join。
- e. 连接语句的位置：from 中、where 条件中。
- f. 连接时是否有过滤操作：在连接属性上过滤、在非连接属性上过滤、过滤操作在 where 中还是 from 中、单点过滤还是范围过滤、过滤操作是否涉及多表多属性、过滤操作中是否含有系统函数。
- g. 对于连接后的中间结果集是否还有其他操作：如 group by、order by、sum、count 等函数、系统函数。
- h. 嵌套子查询的支持：from 语句中，子查询作为中间表（join 左右表是否都是该种情况）、where 语句中，子查询作为一个中间数据集。
- i. 并行执行的支持：对相同数据表并行执行多个 Hash Join 操作。
- j. 每日合并时，Hash Join 执行的正确性问题。

4、实验结果

hash join 共发现 5 个 bug，现已全部解决。