

KYLIN-5323 tableIndex 回答 select star 增强实现 Design

Dev Design

问题描述

kylin4升级kylin5问题，kylin4原来的明细查询可以通过basecuboid的回答查询，kylin5需要做兼容处理

背景

总体设计思路：

步骤一： kylin4中的basecuboid翻译成tableindex

步骤二： kylin5增强tableindex来回答select *。

kylin4的行为现状

场景一：

表A: a、 b、 c、 d、 e、 f、 g列

构建model01: 表A: a、 b、 c、 d、 e、 f为dimension

构建cube1: 选择model01，选择表A: a、 b、 c、 d为dimension

Select * from A 查询显示a、 b、 c、 d列

场景二：

表A: a、 b、 c、 d、 e、 f、 g列

构建model01: 表A: a、 b、 c、 d、 e、 f为dimension

构建cube1: 选择model01，选择表A: a、 b、 c、 d为dimension，e、 f为measure

Select * from A 查询显示a、 b、 c、 d、 e、 f列，但是只有a、 b、 c、 d列有数据，e、 f为mock数据

场景三：

表A: a、b、c、d、e、f列

构建model_01: 表A: a、b、c列

构建model_02: 表A: d、e列

构建cube1: 选择model_01, 表A: a、b、c列

构建cube2: 选择model_02, 表A: d、e列

Select * from A 显示no realization

补充: 跟踪代码发现select * 解析为a、b、c、d、e列

kylin5现状

Select * 会被翻译成table的所有column, 不能用部分column的tableindex回答select * 查询。

目的

kylin5需要支持只有部分column的tableindex回答select *。

实施方案

步骤一: kylin4中的basecuboid翻译成tableindex

步骤二: kylin5中增加开关kylin.query.use-tableindex-answer-select-star.enabled(项目级配置, 默认为false)——使得tableindex来回答select *。

具体实现:

- calcite注册表的schema的时候根据table扫描对应的model, 然后遍历model对应的dataflow.getAllColumns()筛选一下table对应的column, 在OLAPTable中的sourceColumns只显示构建的model暴露出来的column。
- 在tableindex的match的时候, 针对多余的unmatch columns与dataflow.getAllColumns()进行匹配, 如果有则直接remove, 则在tablescan的时候填充一下缺失的列; 如果没有匹配上则返回index unmatched。
- 针对2匹配到index中的unmatch columns size进行计算cost, 记录unmatch columns size, 选择回答索引的时候, 能选择unmatch columns size最小的tableindex回答。

tableScan: mock: project有的列则直接获取, 没有的列 mock: 1 as 1_dummy_XXX_XX

优化的策略：

Agg

project filedlist

tableScan (根据filedlist在tableScan去找，如果有的话直接写名称，如果没有的话就dummy)