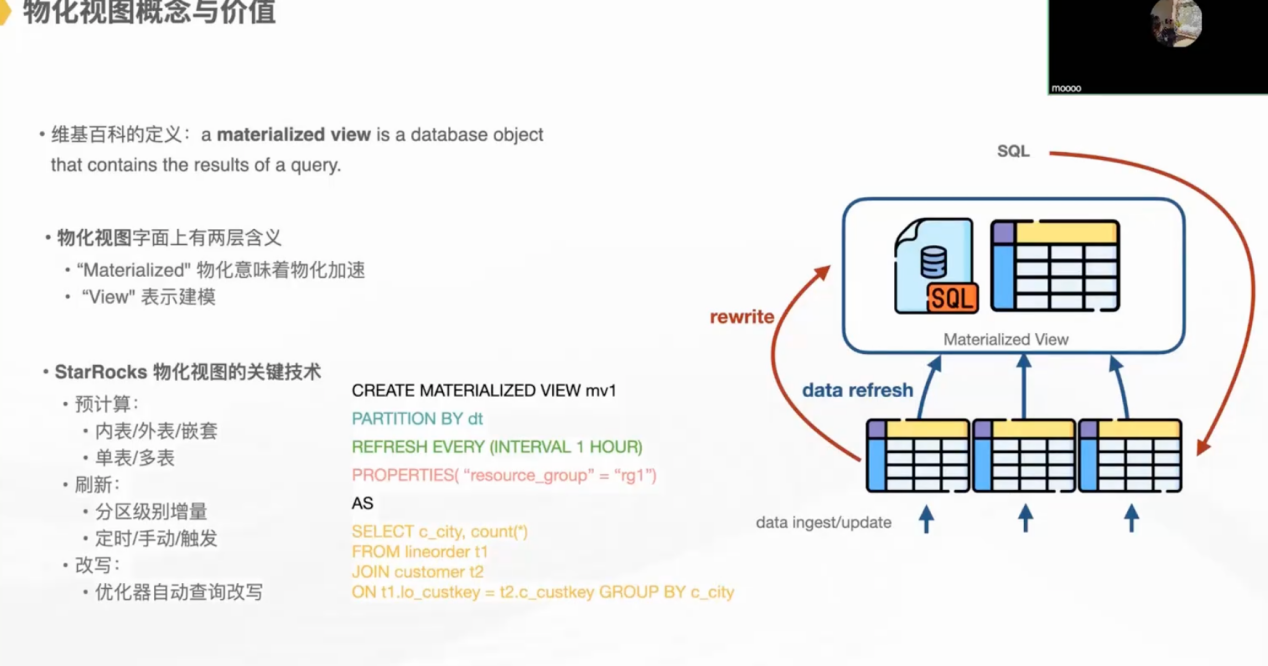
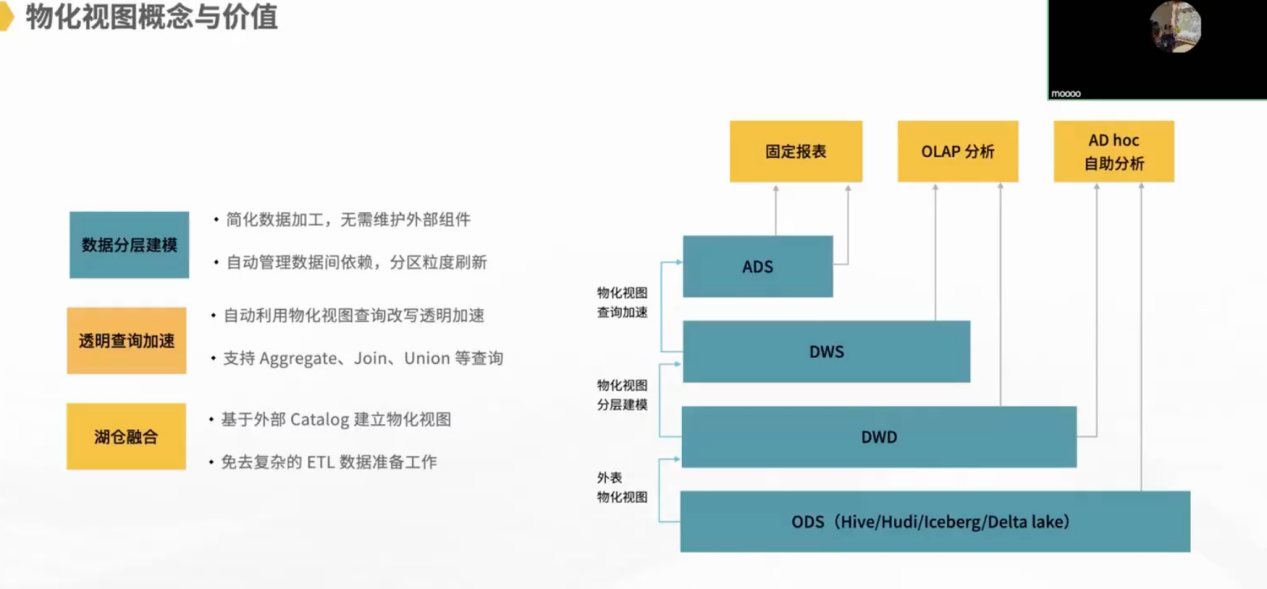
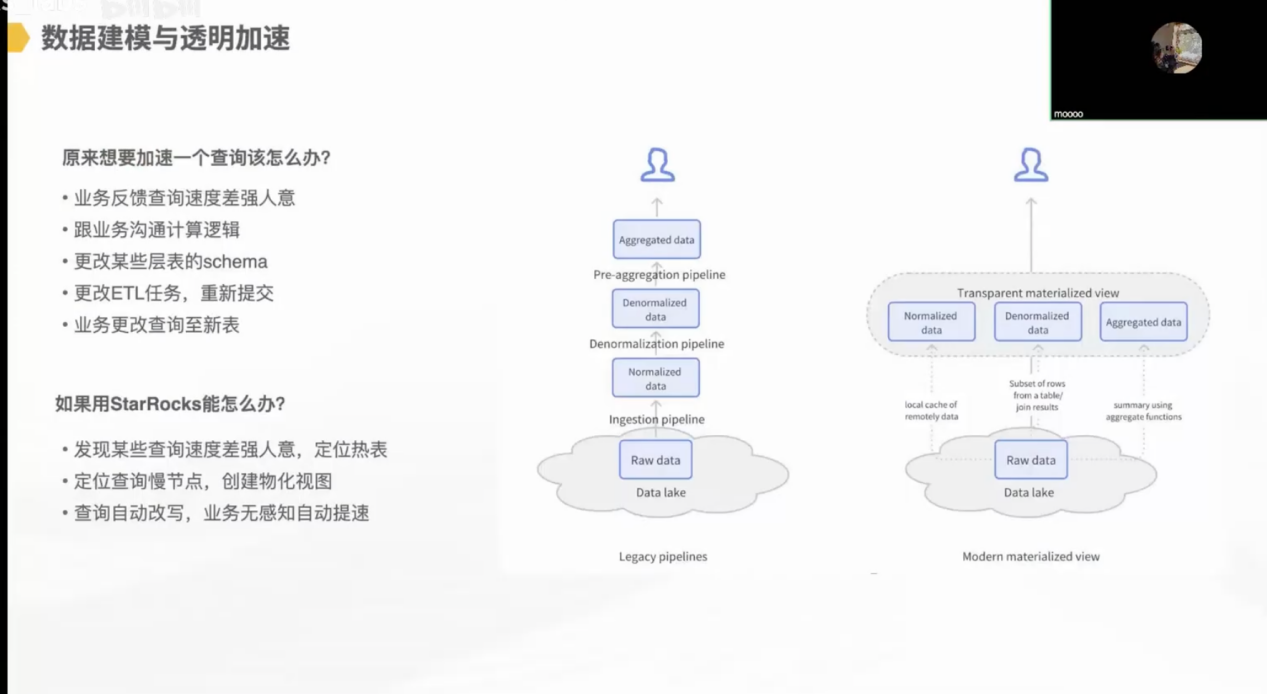
**物化视图答疑**

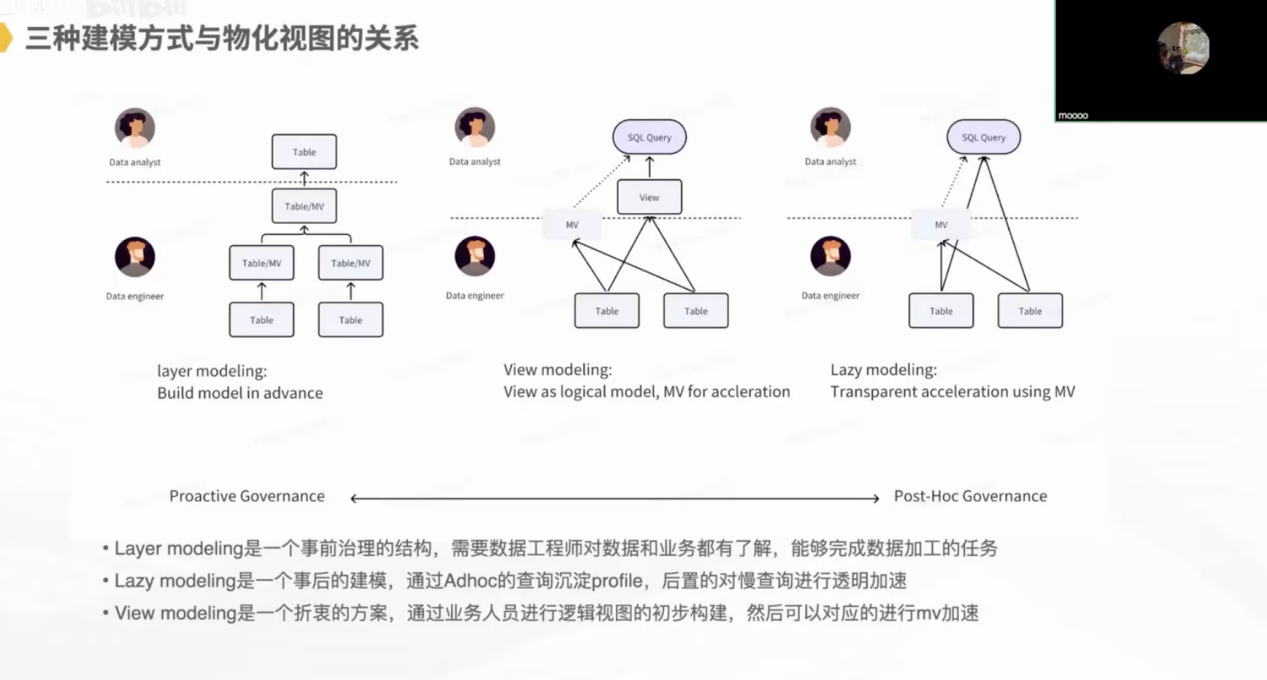
https://www.bilibili.com/video/BV1a94y1P7f4/?spm\_id\_from=333.337.search-card.all.click&vd\_source=a922512cabb051021f3d962645daa462



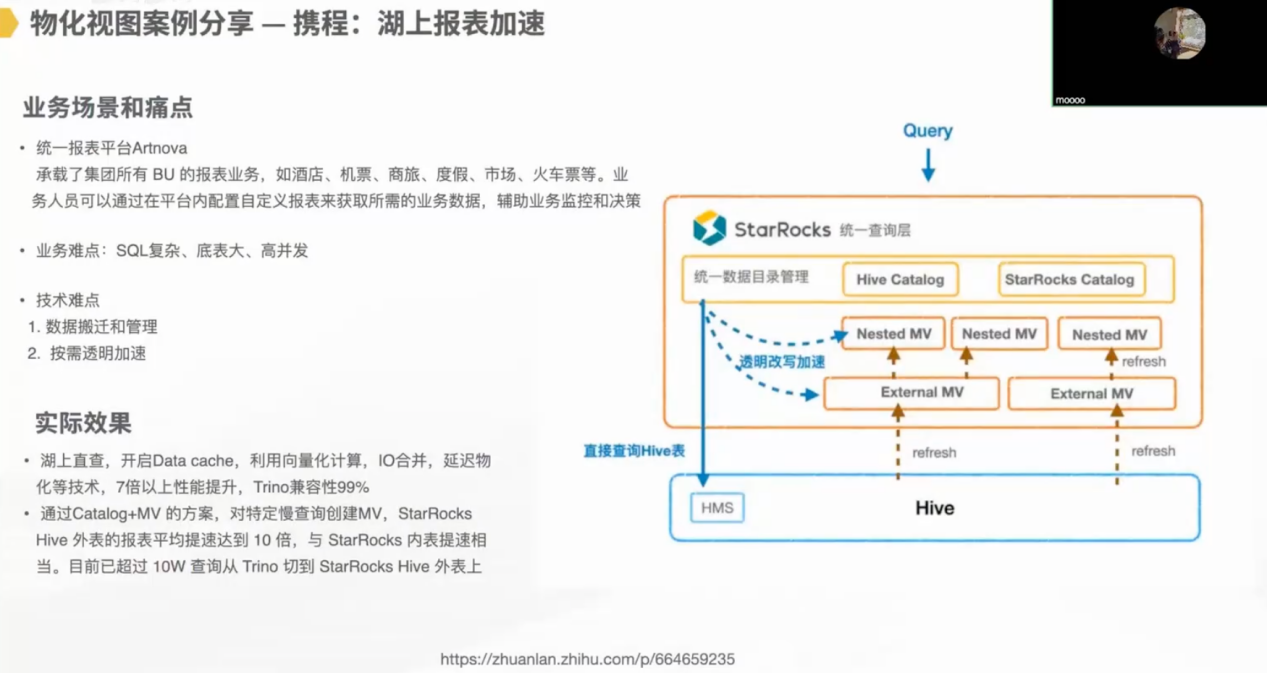




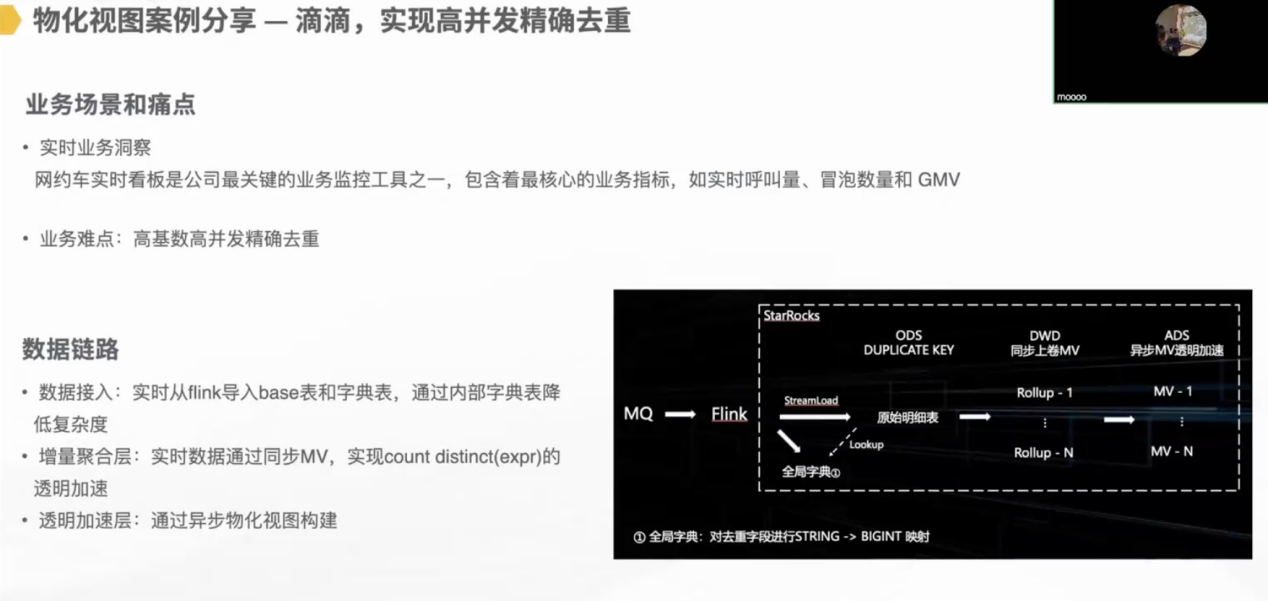










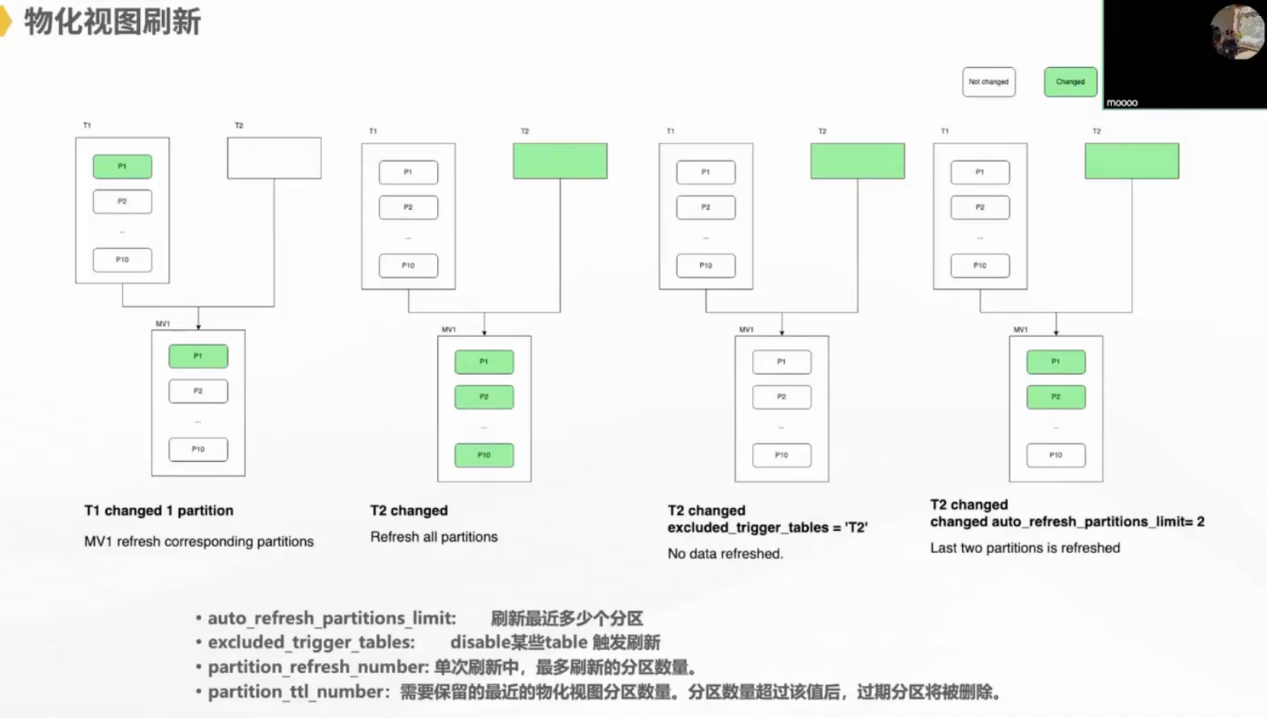


同步物化视图，跟导入是在一个树里的，导入数据的同时，构建同步物化视图，类似索引







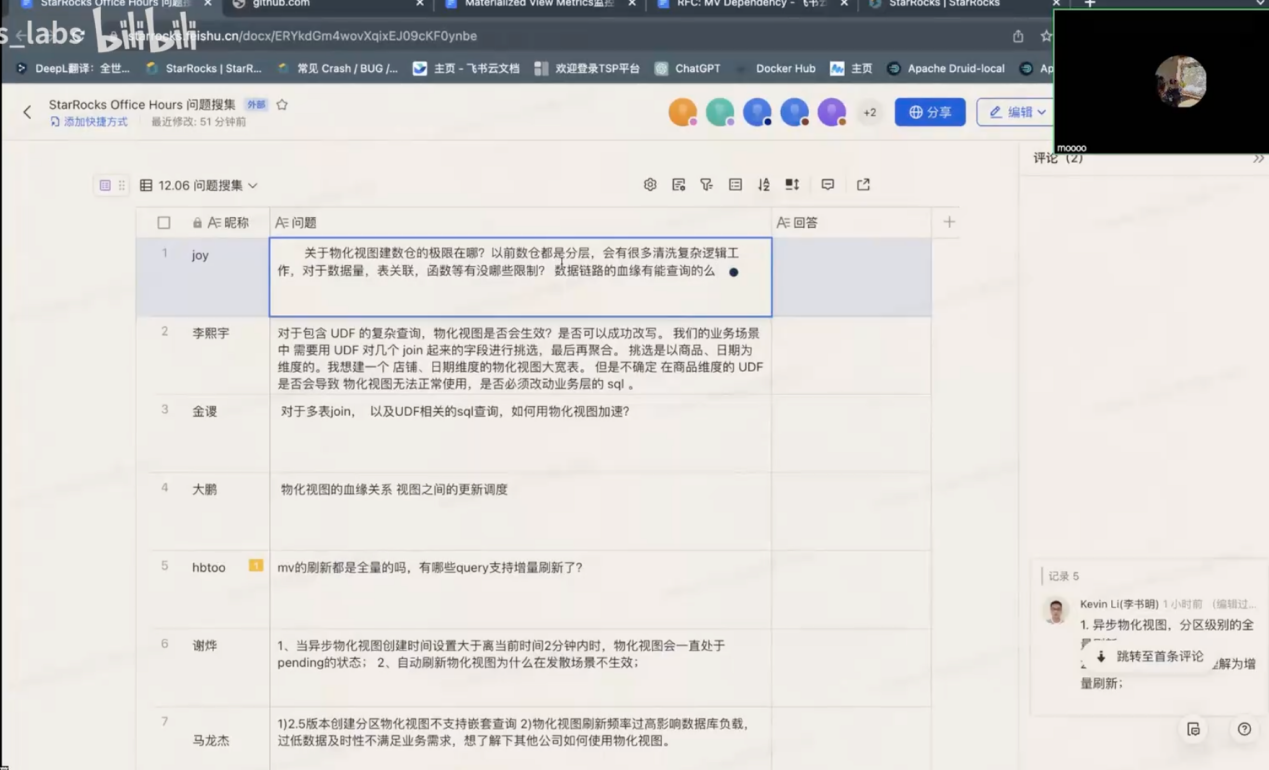


异步物化视图能做到分区级别的增量刷新，下游感知上游变动的分区，只刷新必要的分区。

参考：分区刷新策略。纯增量筛选计划中。

湖上加速

直接查湖比treenode快3倍，data cache打开能做到6倍，物化视图比直接用presto快30~40倍，跟查询相关，特别是聚合，物化视图对聚合优化很大



血缘规划在做，借助开源工具catlinte

udf正在做，改写不成功，有udf从物化视图查数据可以，

Udf为什么不能改写，因为udf是用户自己开发的，不能保证函数是幂等的，第一次调用跟第二次调用效果一样，改写不能够保证语义等价的，后面增加一些语法进行标注，标注完可以做，参考pg的udf标注

底层刷新频率不是很高，可以用触发式刷新，否则任务过多会排队

**物化视图层级过多，刷新会占用资源影响io**

资源组：软隔离

3.1开始，mv刷新支持算子落盘，内存消耗可以控制在相对比较低的水平

硬隔离：开发中

**除了mv透明加速层，其他查询加速？**

mv透明加速大部分问题能解。query cache



**mv的刷新**

如果是分区的mv，可以按照分区级别进行更新，目前内部、hive、iceberg的mv都可以做到这一点。

**分区mv**

base是按秒分区，mv希望按天分区，支持分区映射，partition by跟上函数来支持，datetrunc支持，day不支持。

**MV定时任务没有执行完，堆积了怎么处理的？**

内部进行了任务的合并，不会造成堆积

**MV执行的监控**

infomation\_schema库中查看task和taskrun表能看到物化视图的执行的任务

**删除重建导致mv失效**

3.2当物化视图被设置为inactive，后台会再找base表，重新构建

**分区没分区更精确化的耗时**

目前只是记录的最后一个开始和结束时间

**滴滴的例子下层同步物化视图和上层异步物化视图刷新频率不一样**

同步物化视图目前不支持嵌套，异步物化视图可以嵌套

**实时的mv与离线数据的合并**

比如PK表提供实时能力，Broker Load 进来或者外表进来离线数据，上面做个view，再向上提供服务。

实际场景，比如直播间的数据，关播之后第二天做离线回刷个别字段。left join用coalote join去优化。实时数据的离线覆盖，相当于既查实时又查离线的。异步物化视图做实时表，剔除掉离线的数据。

**mv发版方式**

除了先drop掉再发布外，还支持spark构建mv，将新的mv与老的mv进行交换。

**mv无法构建依赖**

刷新频率不高，可以尝试做成触发式刷新，刷新频率过高会导致系统负载过重任务一直排队。

**MV查询改写**

改写规则，有一些tip，比如过于复杂的查询不建议做到一层里，先做一层Join，上面再根据不同的维度做agg

**大宽表不同字段聚合一起mv会比较复杂，但是分开管理有比较麻烦**

可以stream load加工好聚合

insert overwrite 分区替换，insert into，追加模型，数据会追加上去，如果是主键模型，两者都有覆盖语义，insert into可以实现行级更新，insert overwrite先写临时分区，比insert into性能好，create table as 比insert into性能也好一些，但是如果5分钟调度一次insert into对cpu毛刺特别高。

insert overwrite全量计算进行覆盖，但是不想全量计算怎么办？微批调度。

flink写Pk列表的时候毛刺、内存都比较严重，pk表开了主键索引物化之后，问题好了很多。

mv是否可以append而不是overwrite掉数据，有场景，看怎么整合功能里去。

**hive表更新过程中mv还没来得及刷新，查询会不会走到mv查询到旧数据？**

mv刷新的时候有个参数控制check数据一致性，它的一致性与直接访问catalog hive表的一致性是一样的。

如果设置了check数据一致性，mv数据已经过时，那就不会改写到mv，会基于base表重计算。

**支持trino sql创建物化视图吗?**

可以全局设置

mv命中了哪些？

在审计日志里可以查到，或者改写生效里