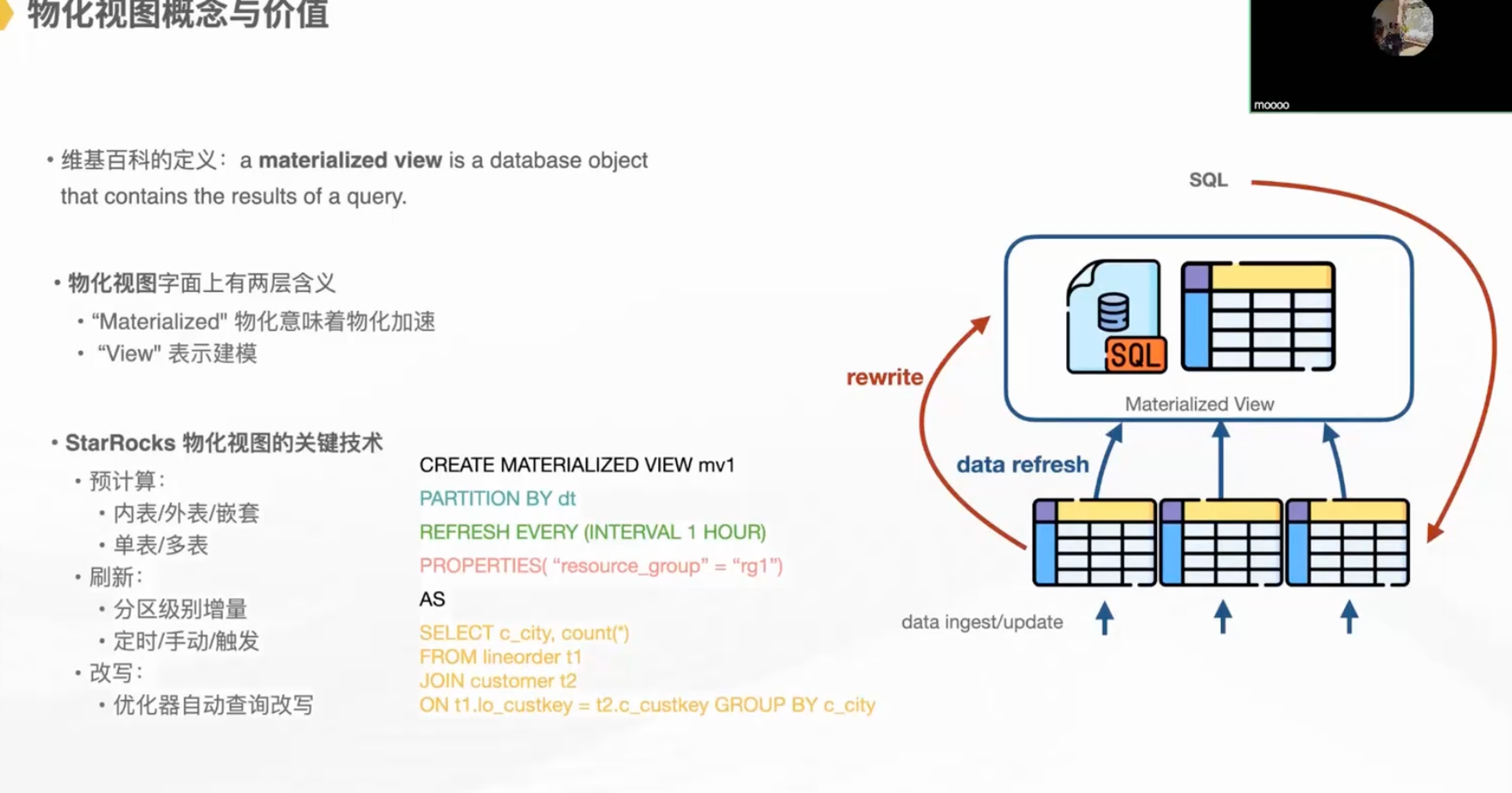
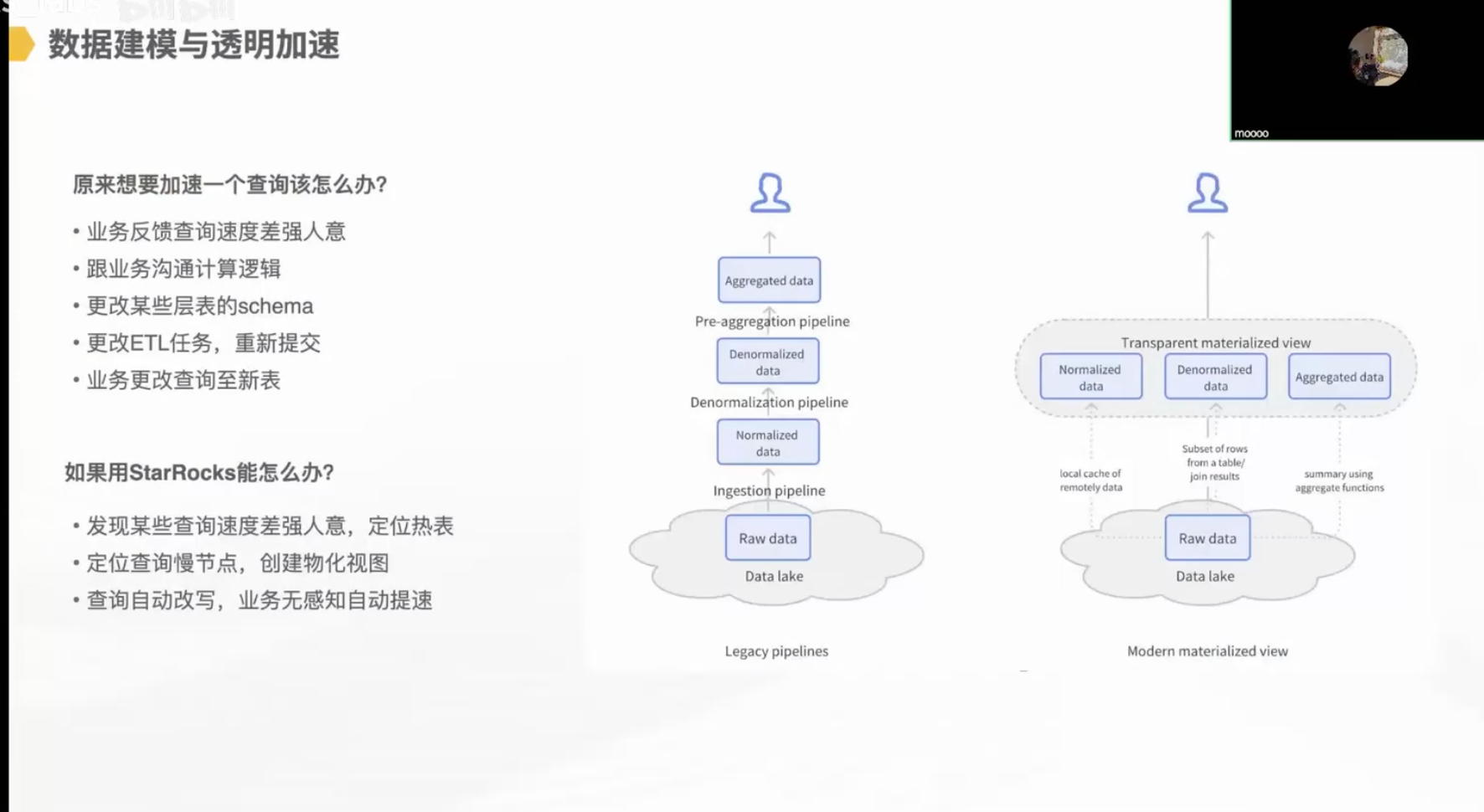
**物化视图答疑**

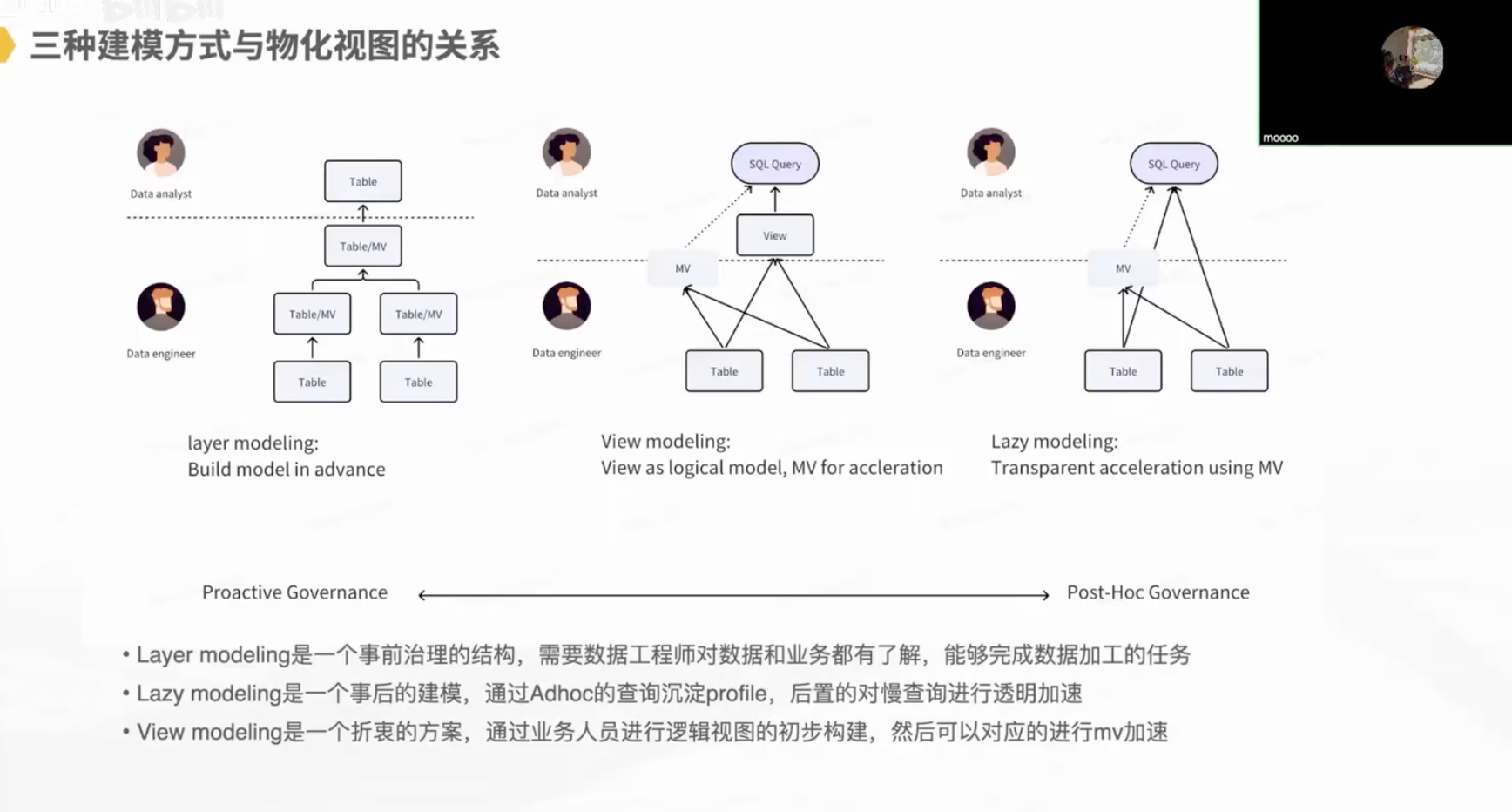
https://www.bilibili.com/video/BV1a94y1P7f4/?spm\_id\_from=333.337.search-card.all.click&vd\_source=a922512cabb051021f3d962645daa462



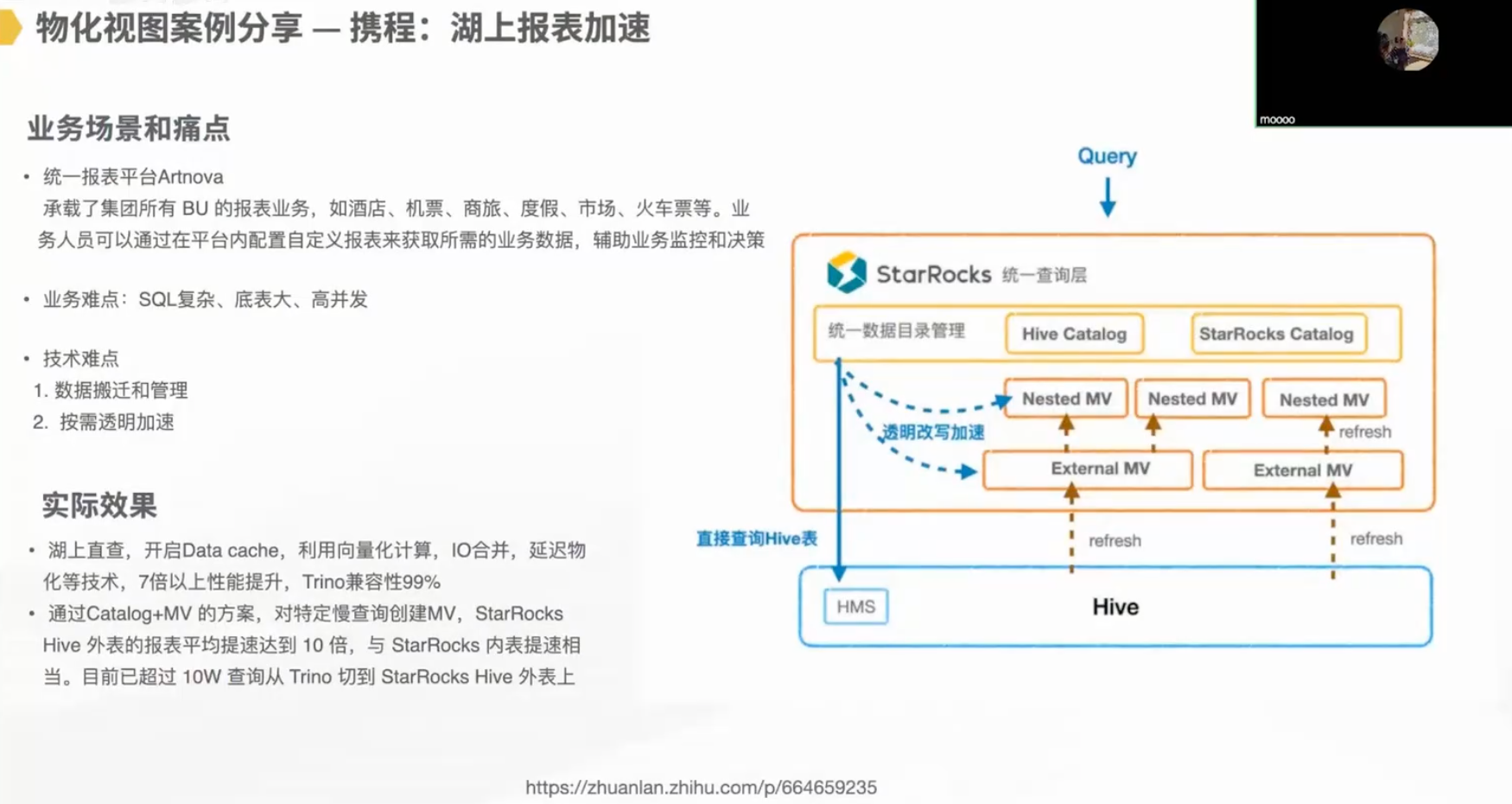




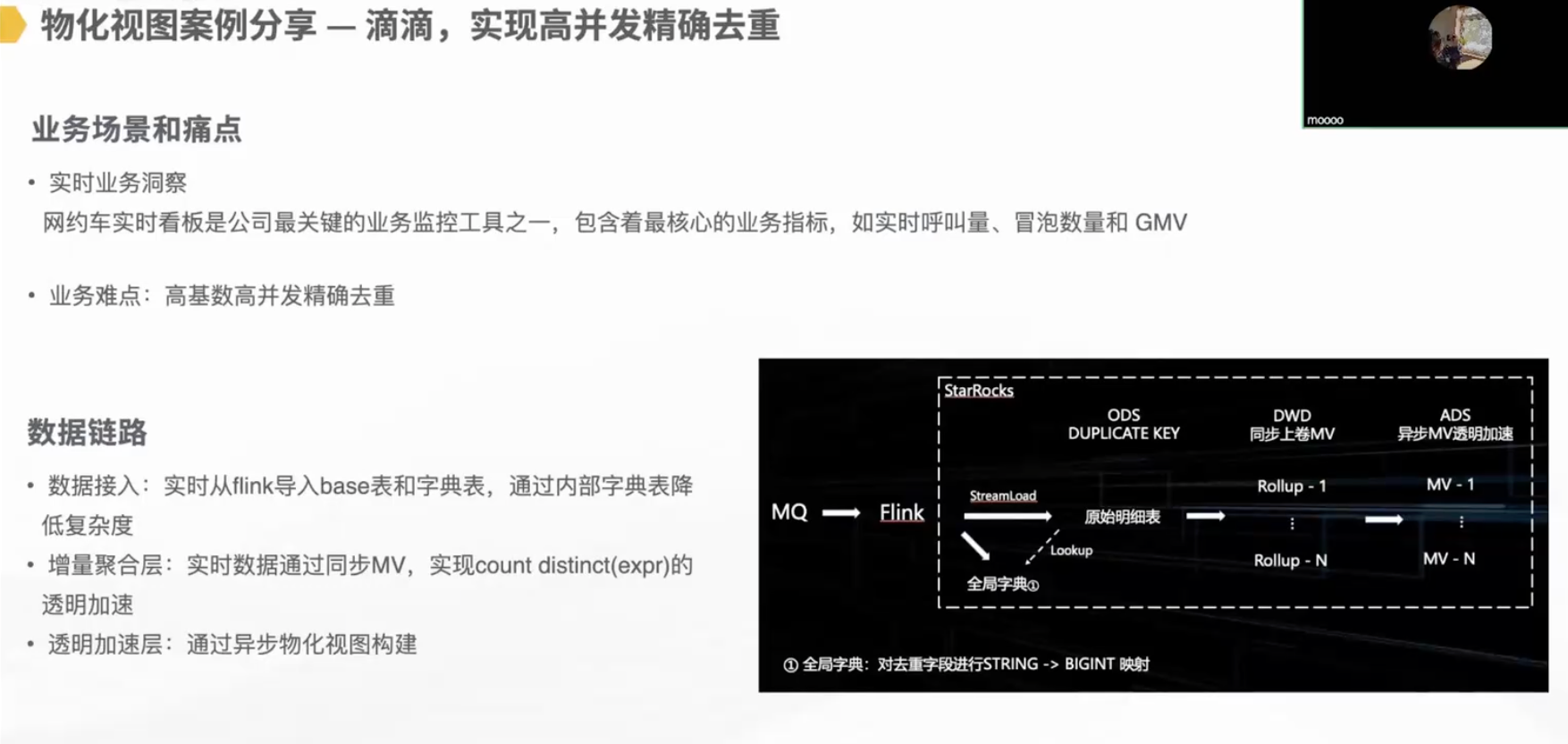




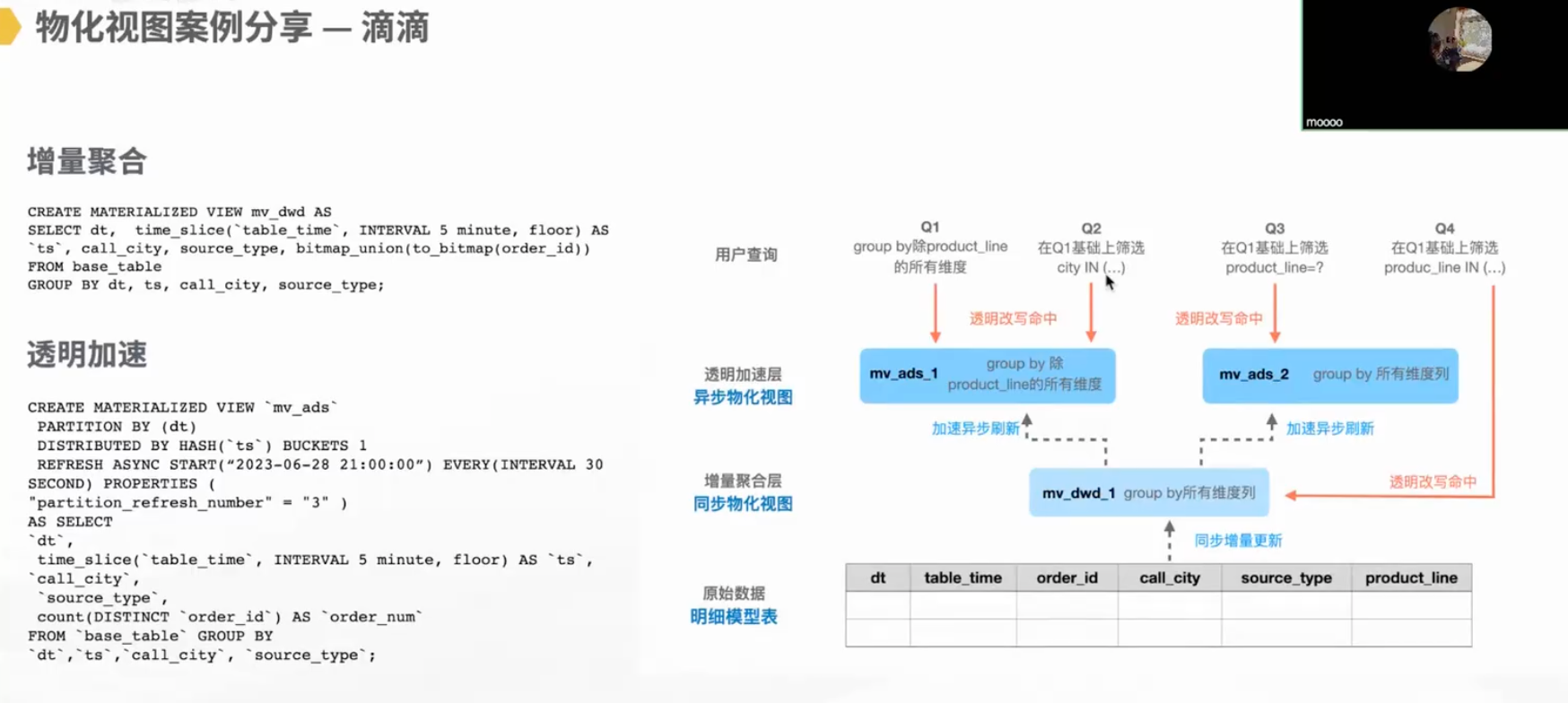






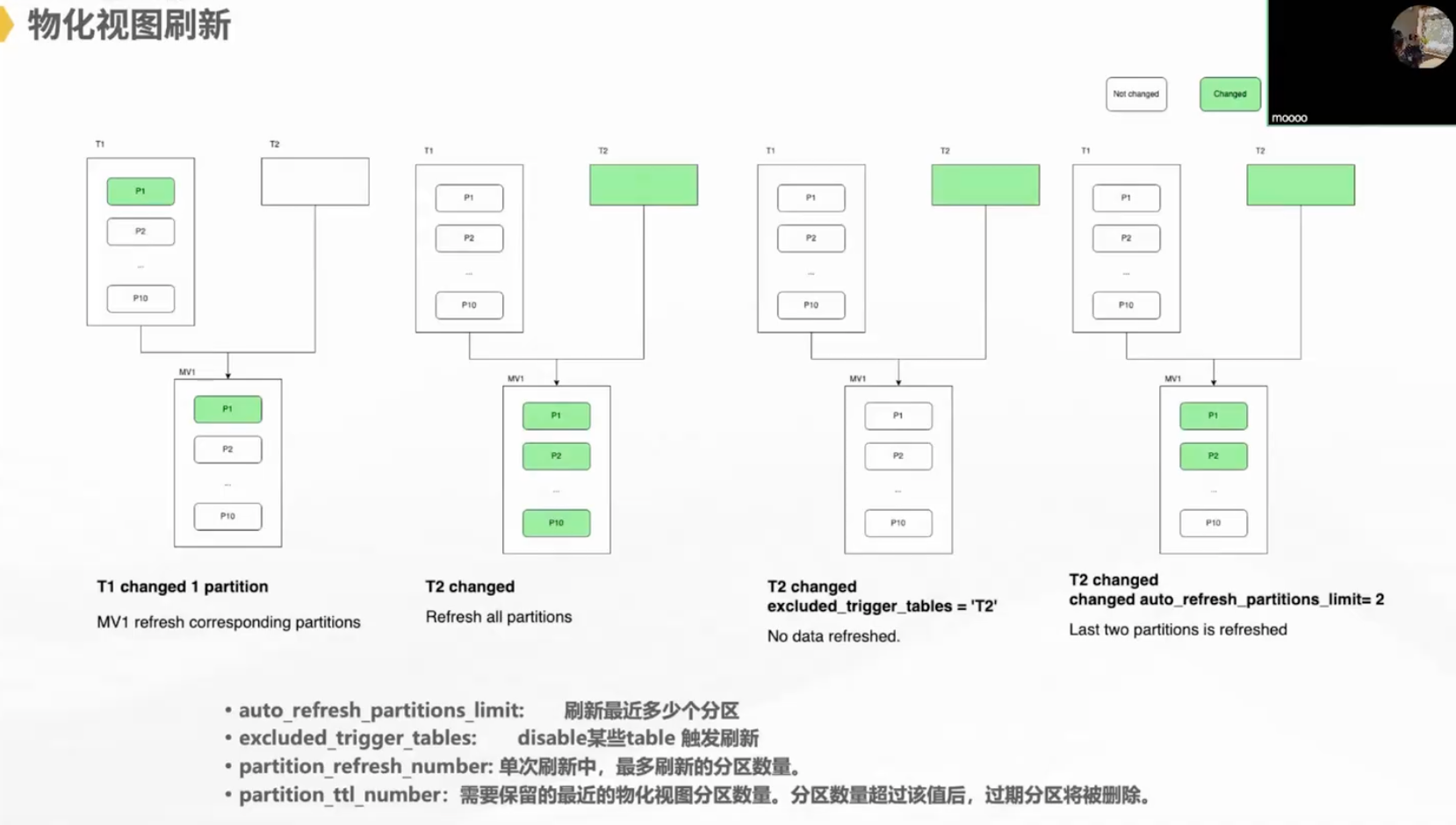


同步物化视图，跟导入是在一个树里的，导入数据的同时，构建同步物化视图，类似索引







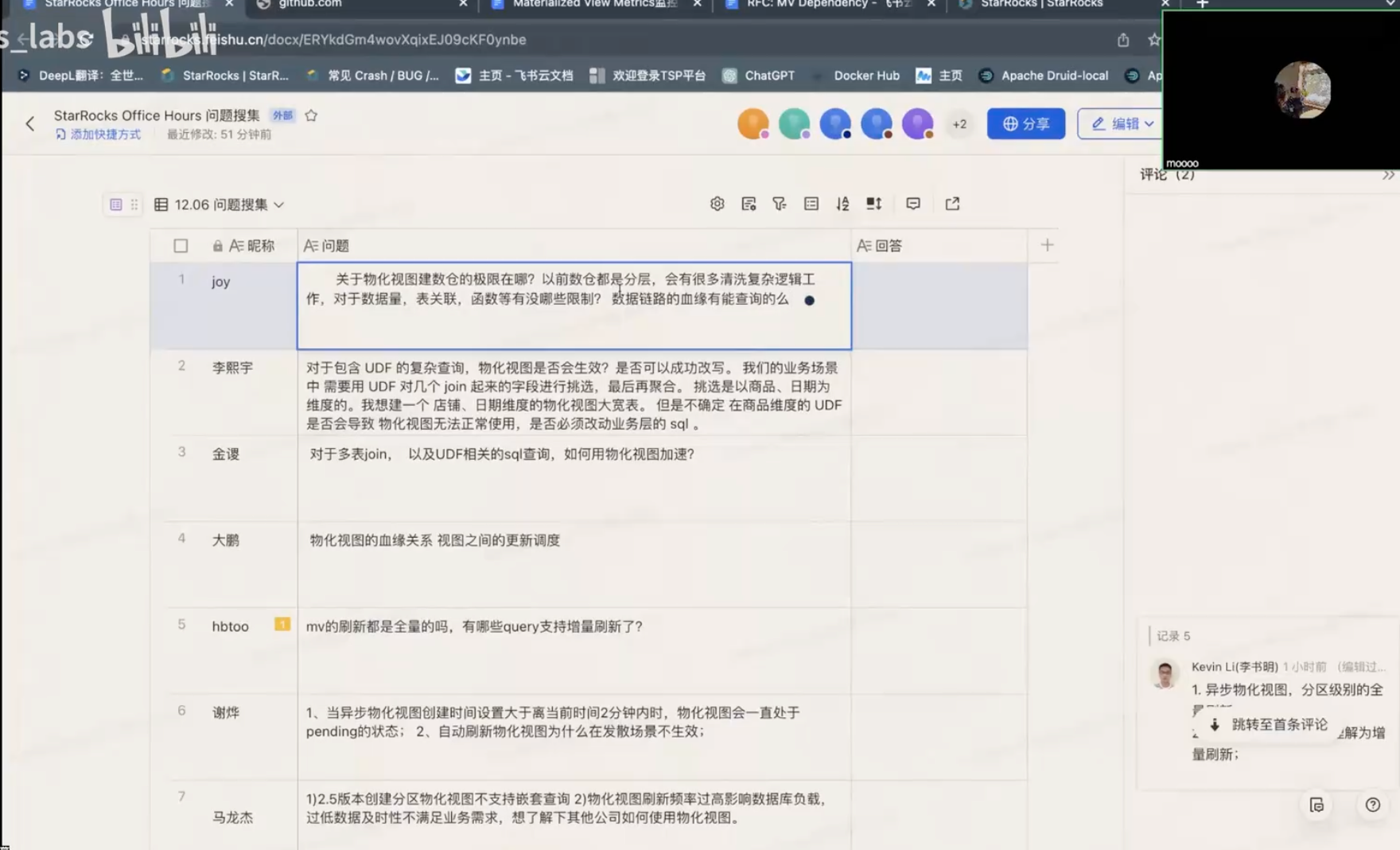


异步物化视图能做到分区级别的增量刷新，下游感知上游变动的分区，只刷新必要的分区。

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。