

Apache Doris

在金融壹账通的实践经验

兰后

平安金融壹账通 数据挖掘工程师

目录

01 指标平台介绍与应用

02 基于 Apache Doris 的指标架构升级

03 基于 Apache Doris 的指标应用性能调优实践

1 指标平台介绍与应用

指标平台：提供集指标构建、指标查询、指标治理一体的指标数据服务平台

助力企业经营数字化转型



指标口径、标准不统一

- 各个业务条线生成的报表堆积如山，不同分析工具生产的指标口径相互打架。



对指标生命周期进行管理

- 支持指标的录入、发布、派生、分享、可视化应用以及深度分析等，并结合管理制度，在使用的各个环节落实指标的管理规范。



指标计算重复、分析效率低

- “需求提出-数据探源与开发-报表制作-上线验收”的过程费时费力。



对指标进行归因关系分析

- 基于指标标准体系，帮助指标各层级找到过程中的问题，有效科学决策。

指标平台：多种指标类型+不同维度构建指标体系

指标： 通过数据量化业务，从而促进业务目标可描述、可衡量、可拆解

维度： 分析业务的视角，包括各层级员工常用视角、业务属性、行业特征、地域特征等



指标平台：为银行提供全面的指标分析与展示、智能归因分析，辅助经营决策

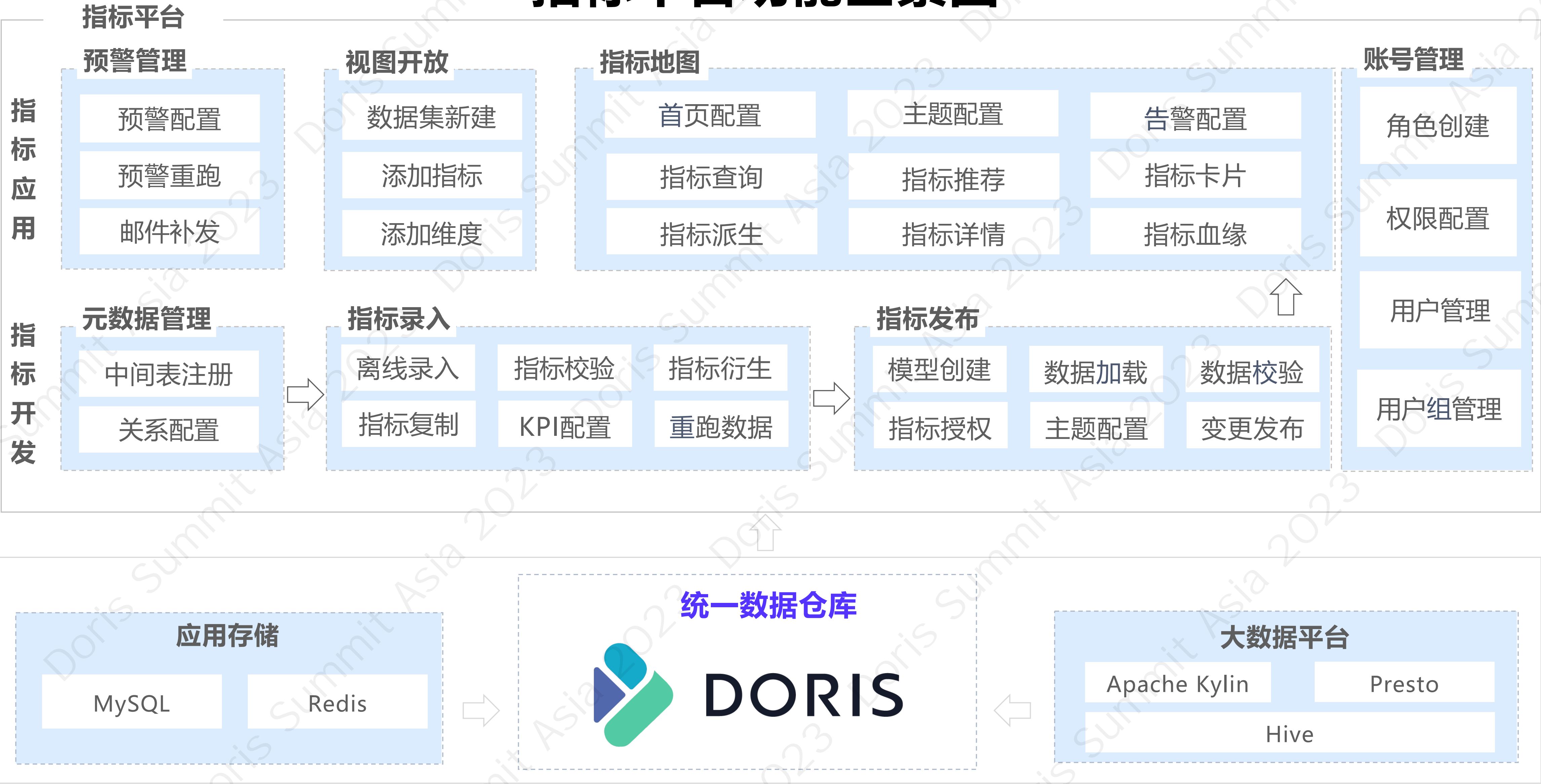


内容：支持更多指标类型，支持指标配置化接入，完善指标/维度价值管理。

BI：支持更多分析场景，持续优化操作体验，移动分析场景优化。

AI：融入更多智能算法，赋能更多数据洞察。

指标平台功能全景图



2 基于 Apache Doris 的指标架构升级

指标架构升级：引入 Apache Doris



指标架构1.0: Hadoop + Presto + Apache Kylin

痛点:

- Apache Kylin 无法满足更灵活的分析需求; 使用**成本高**
- Presto 负责明细查询下压, 但**性能不足, 并发不够**
- 维护多套组件, **运维成本高**

诉求:

- 复杂场景**多表关联**查询性能好
- 支持**高并发**
- **运维简单**, 方便扩展
- 易用性强, **开发简单方便**



OLAP 选型

OLAP	优点	缺点	是否开源	协议
ClickHouse	<ul style="list-style-type: none">• MPP• 单表查询性能强• 保留明细• 物化视图与索引	<ul style="list-style-type: none">• 多表 Join 支持差• 并发度低• 实时更新支持一般	开源	非标准协议
Apache Doris	<ul style="list-style-type: none">• MPP• 高并发和高吞吐的 Ad-hoc 查询• 多表 Join• 支持明细和聚合• 物化视图与索引• 实时+离线数据导入• Schema 在线变更• 支持一定程度更新	<ul style="list-style-type: none">• 社区相对年轻• 产品迭代较快	开源	兼容 MySQL 协议
TiDB	<ul style="list-style-type: none">• HTAP• 架构简单	<ul style="list-style-type: none">• OLAP 能力不足	开源	兼容 MySQL 协议
Greenplum	<ul style="list-style-type: none">• HTAP• 支持多级分区• 多表 Join	<ul style="list-style-type: none">• 单机性能瓶颈	开源	兼容 MySQL 协议

指标架构2.0：Apache Doris

数据：

- 使用 Duplicate 模型+物化视图
- 使用 RANGE 进行时间分区
- 指定 维度键 作为 KEY

应用：

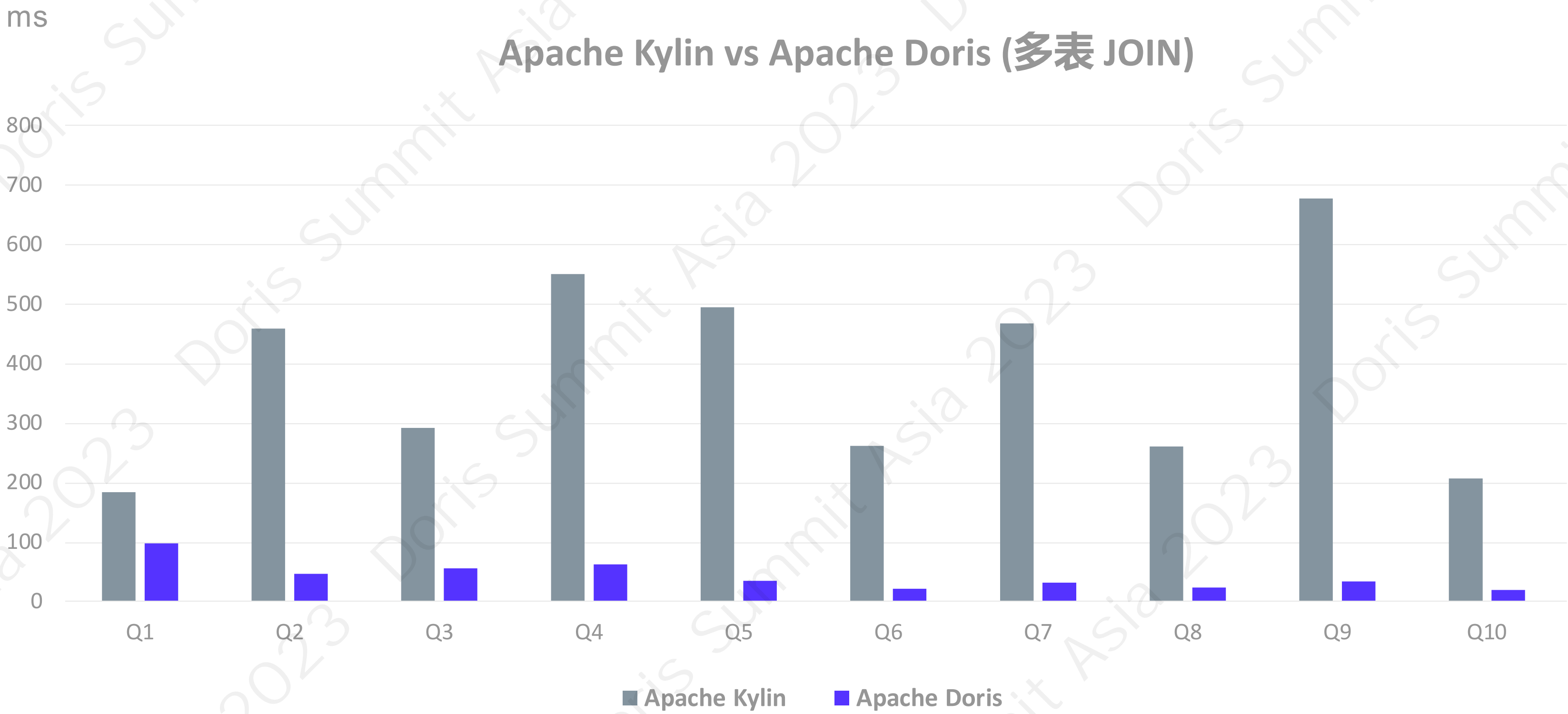
- 重写MySQL兼容的查询引擎
- 调整简化指标发布流程



指标架构2.0：Apache Doris

收益

- 支持更灵活的指标分析
- 高效查询分析性能
- 运维简单



3 基于 Apache Doris 的指标应用性能调优实践

1. 数据准备

1

建表：Duplicate + 时间分区

考虑到指标平台以多表关联场景为主，
我们选择明细模型作为数据模型。
KEY 值选择：指定维度键作为 KEY
指定 DT(INT) 字段进行 RANGE 方式进行。

2

运行 Python 脚本造数

生成多个 1~2G 的 CSV 文件。

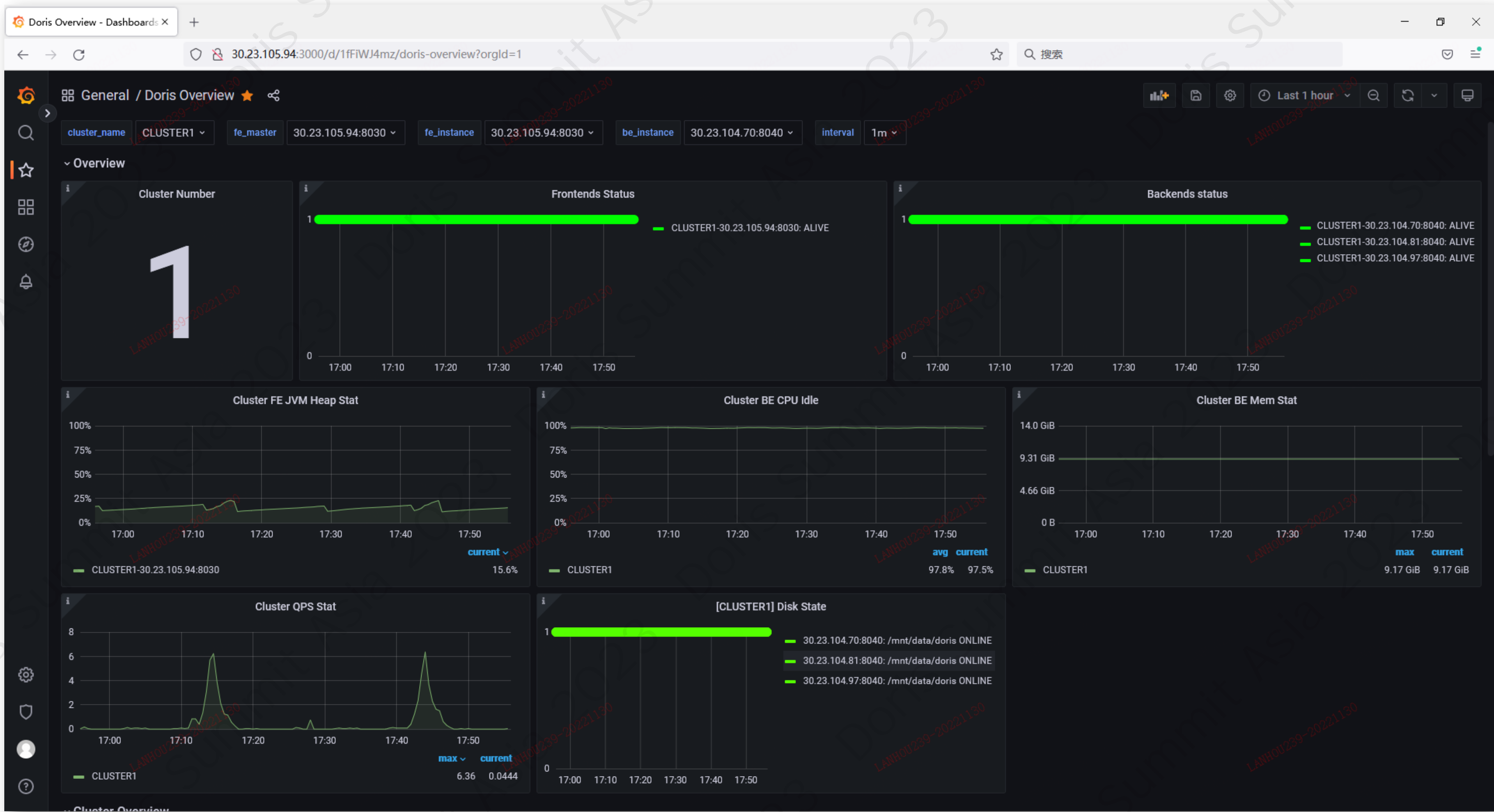
3

运行 Bash 脚本发送导入请求

基于 StreamLoad 方式，完成批量导入。
最终压缩比 3:1，速度 40MB/s。

表中文名	行数
汽融申请表	900,000,000
客户维度表	900,000,000
省市区维度表	74,210
银行机构维度表	4480
业务员维度表	4480

2. 部署集群监控：Grafana + Prometheus



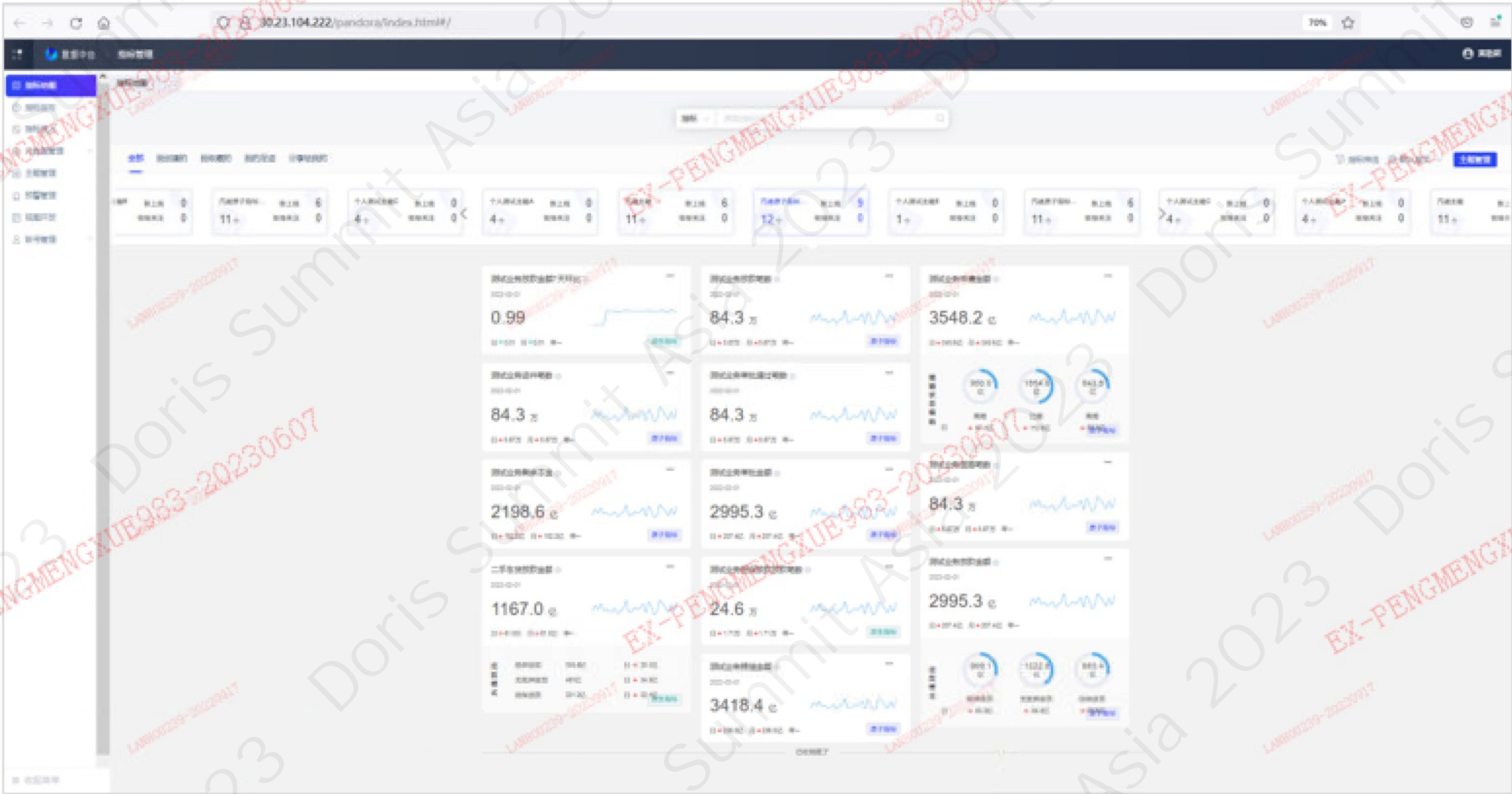
单页面加载性能瓶颈

问题发现：

- 多指标查询时，BE 节点因 OOM 宕机
- 单接口查询性能无法达到预期

解决方案：

- Colocation Join
- 调整数据模型



解决方案

1. Colocation Join

2. 调整数据模型

1. Colocation Join

获取指标趋势图

```
select sum(f.METRIC) as m, f.DT as DT from f
left join dim0
on f.FKEY = dim0.DKEY and f.DT = dim0.DT
left join dim1
on dim0.DKEY = dim1.DKEY and dim0.DT = dim1.DT
where f.DT in (20220103, ..., 20220201)
group by f.DT
```

耗时: 5 s 144 ms

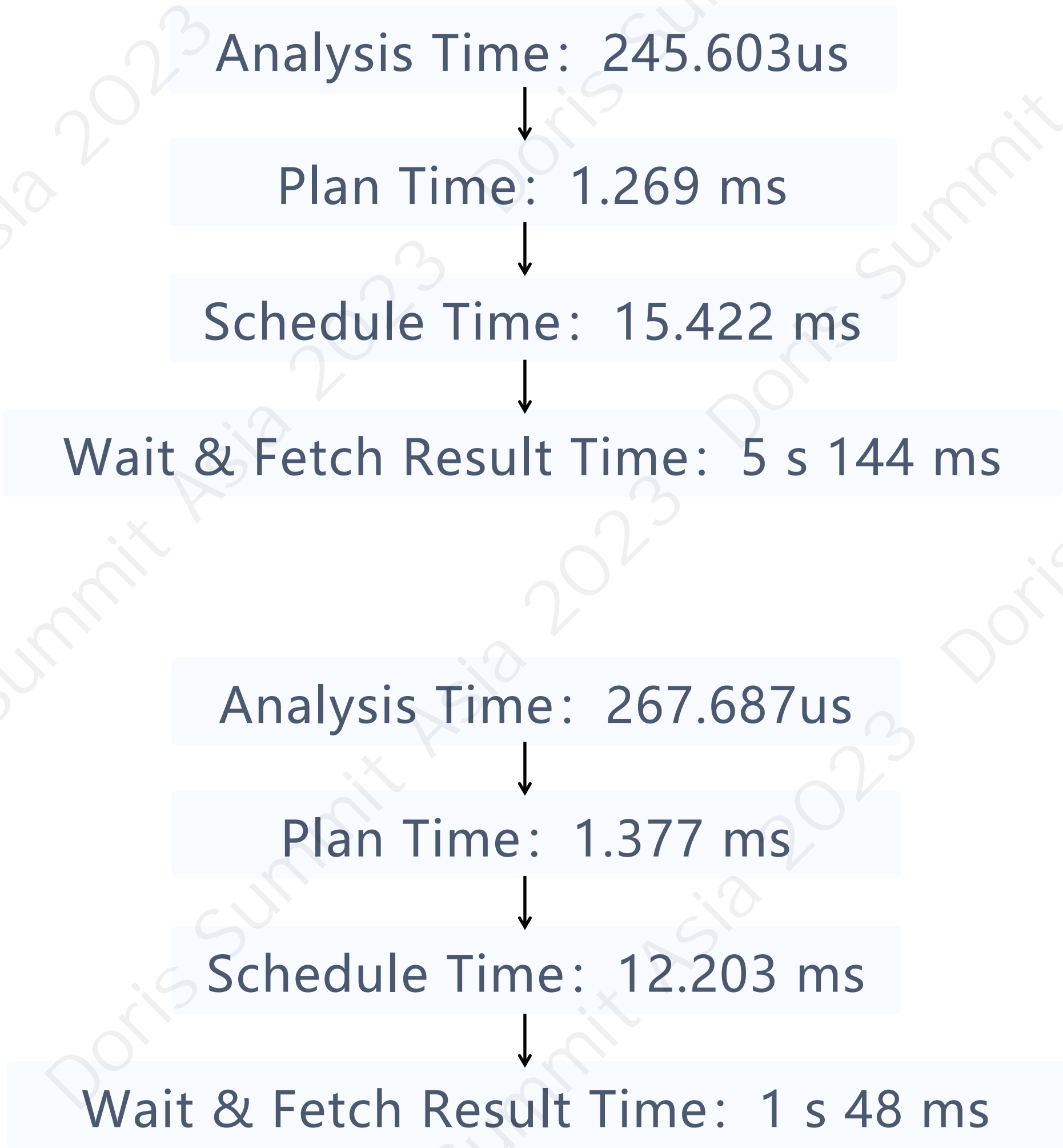
BroadCast Join

获取指标趋势图

```
select sum(f.METRIC) as m, f.DT as DT from f
left join dim0
on f.FKEY = dim0.DKEY and f.DT = dim0.DT
left join dim1
on dim0.DKEY = dim1.DKEY and dim0.DT = dim1.DT
where f.DT in (20220103, ..., 20220201)
group by f.DT
```

耗时: 1 s 48 ms

Colocation
Join



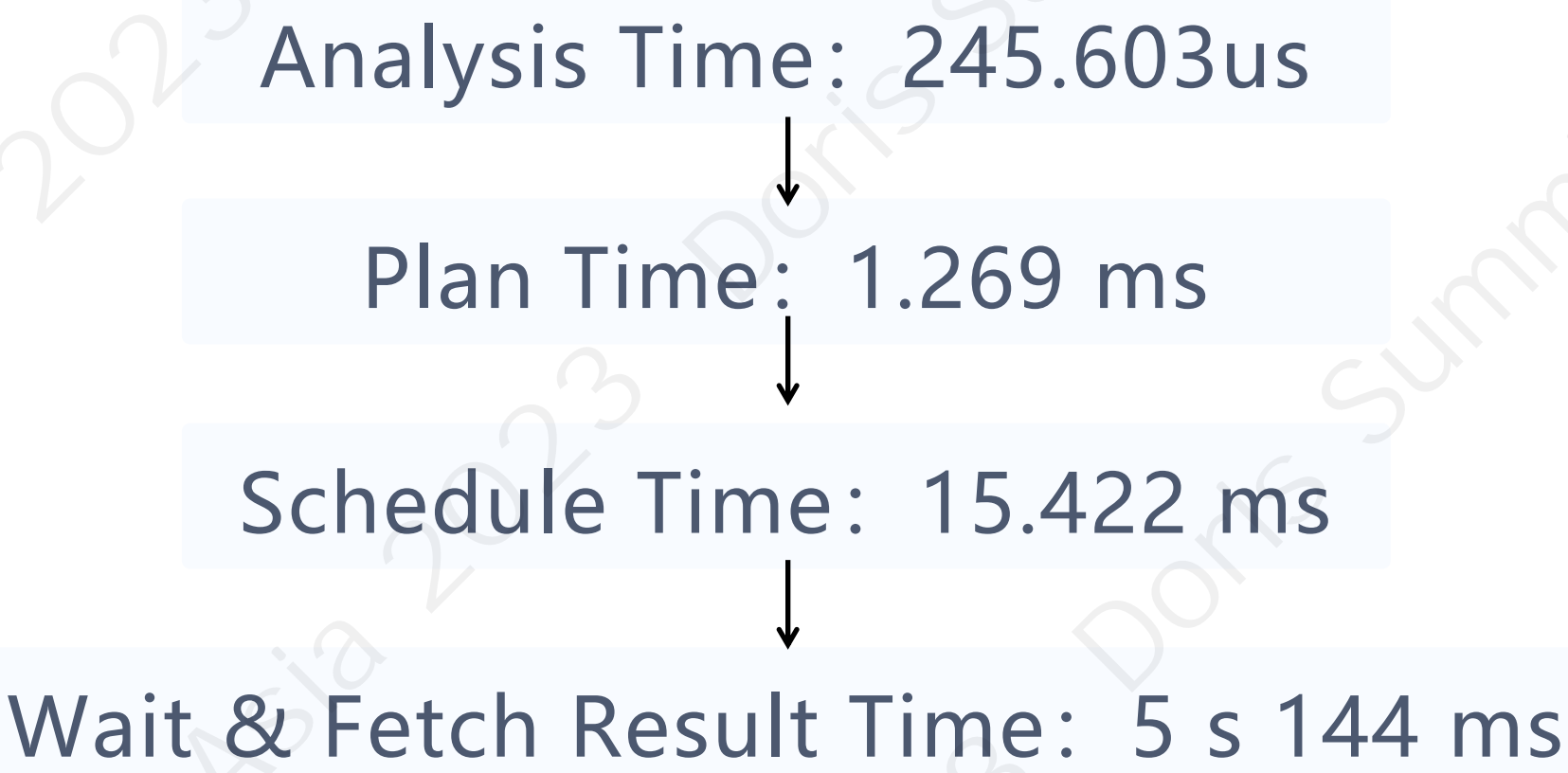
2. 调整数据模型：SQL 执行计划

获取指标趋势图

```
select sum(f.METRIC) as m, f.DT as DT from f
left join dim0
on f.FKEY = dim0.DKEY and f.DT = dim0.DT
left join dim1
on dim0.DKEY = dim1.DKEY and dim0.DT = dim1.DT
where f.DT in (20220103, ..., 20220201)
group by f.DT
```

耗时：5 s 144 ms

雪花模型

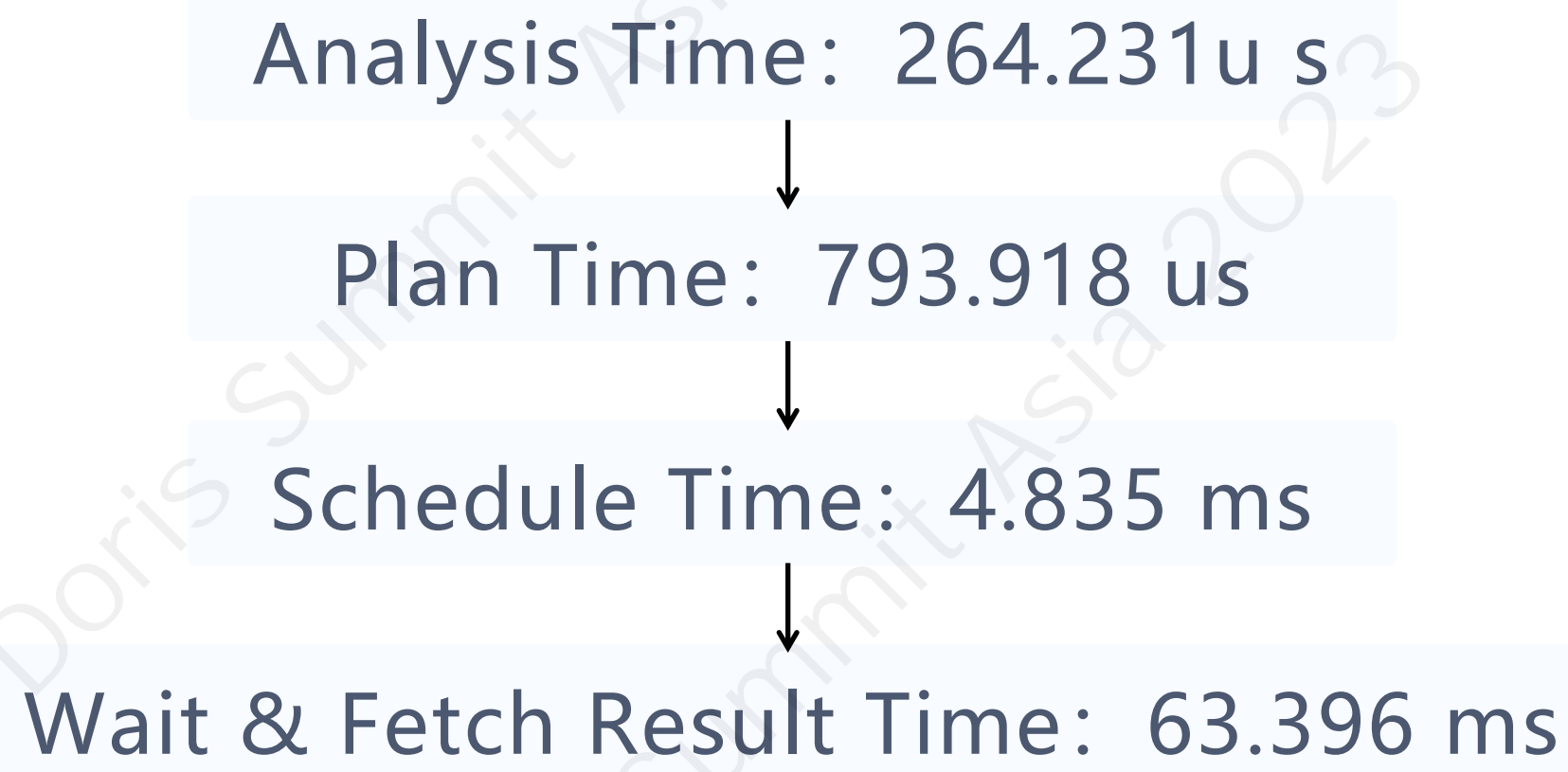


获取指标趋势图

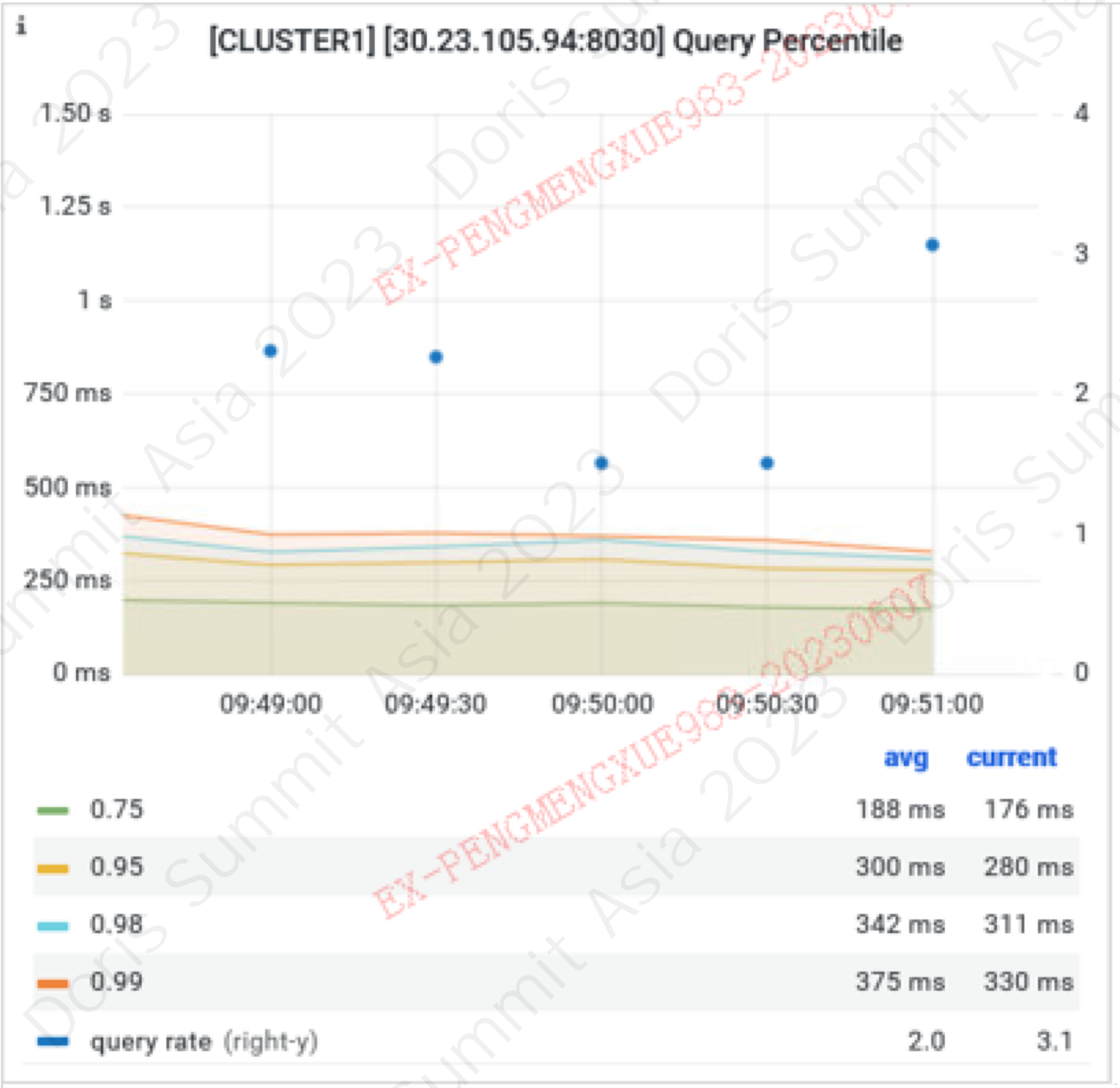
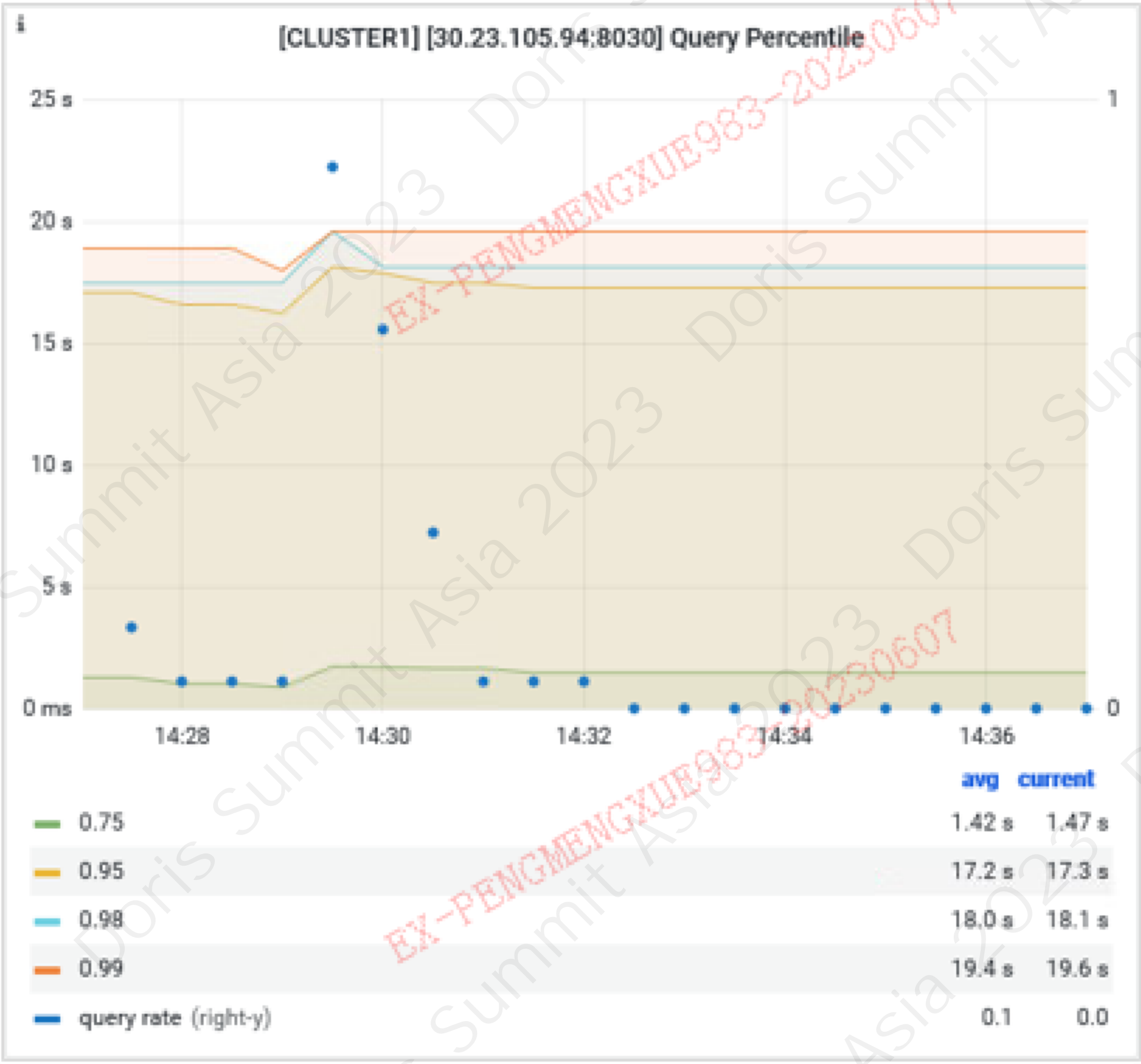
```
select sum(f_flat.METRIC) as m, f_flat.DT as DT from
f_flat
left join dim1
on f_flat.DKEY = dim1.DKEY and f_flat.DT = dim1.DT
where f_flat.DT in (20220103, ..., 20220201)
group by f_flat.DT
```

耗时：63.396 ms

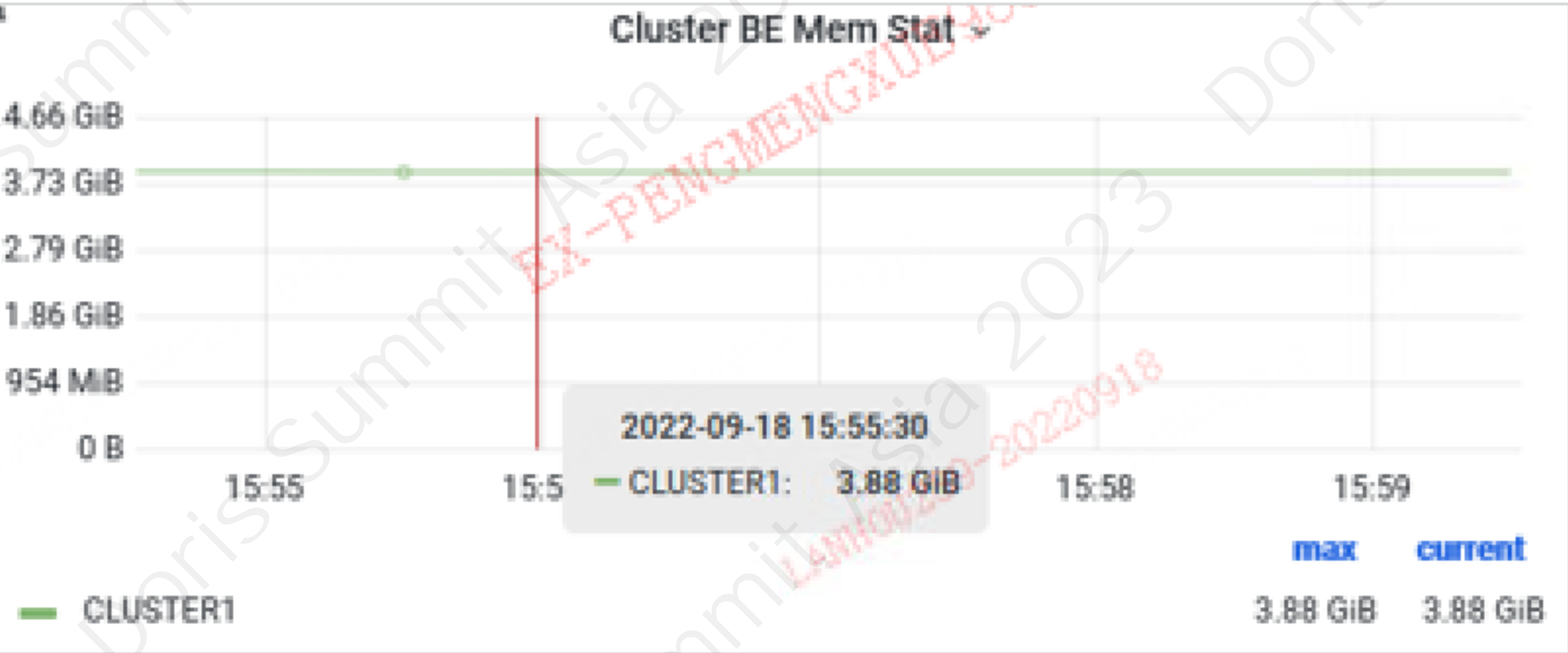
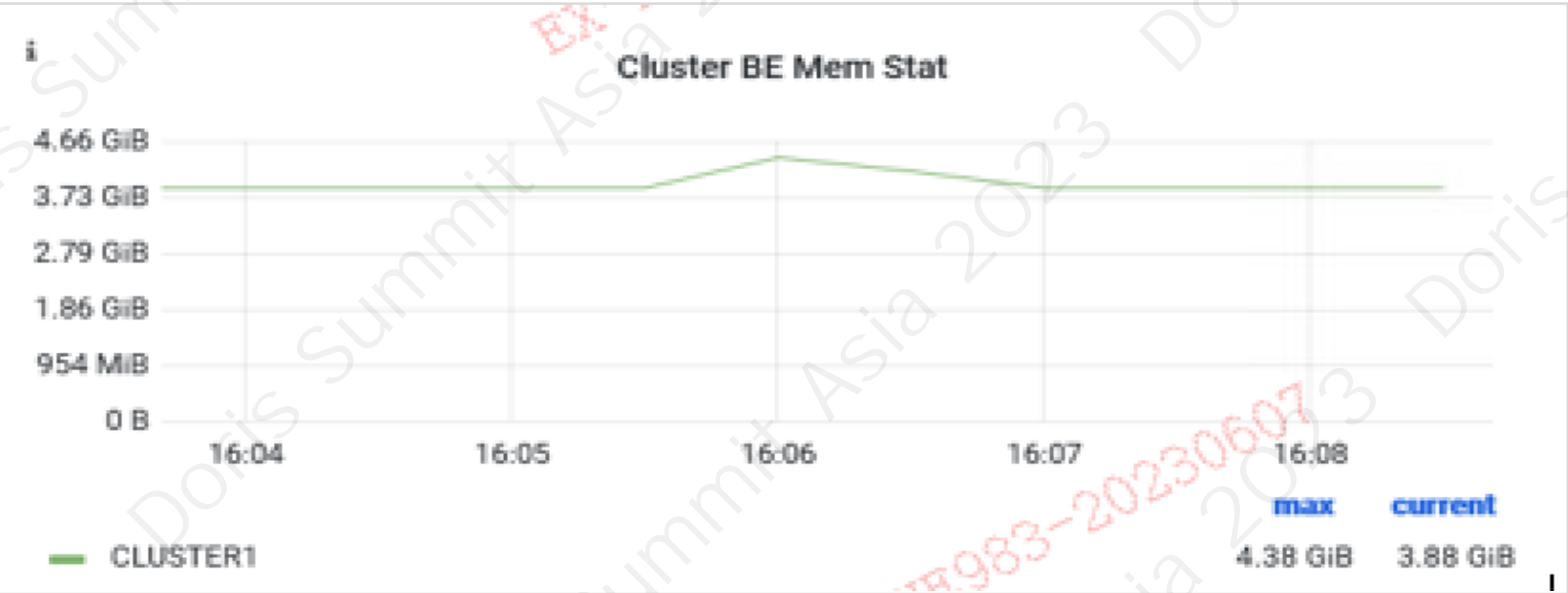
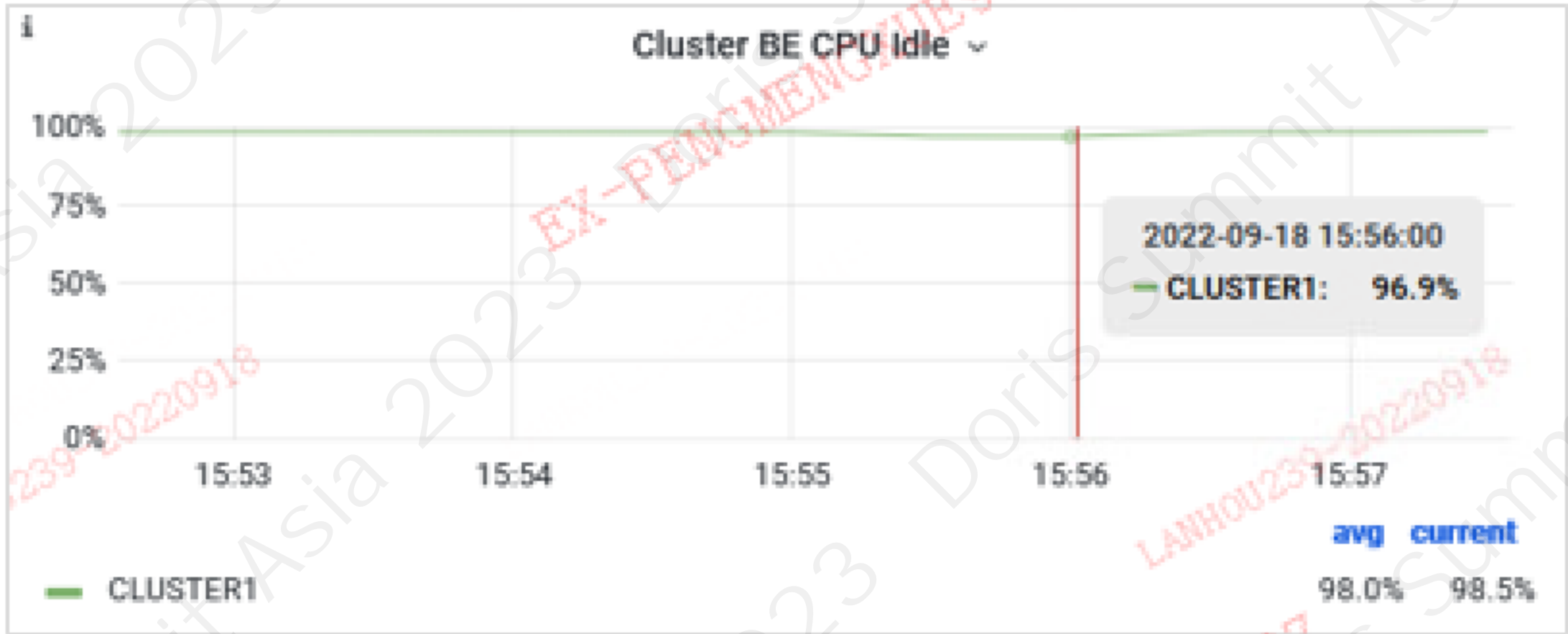
大宽表



2. 调整数据模型：查询性能

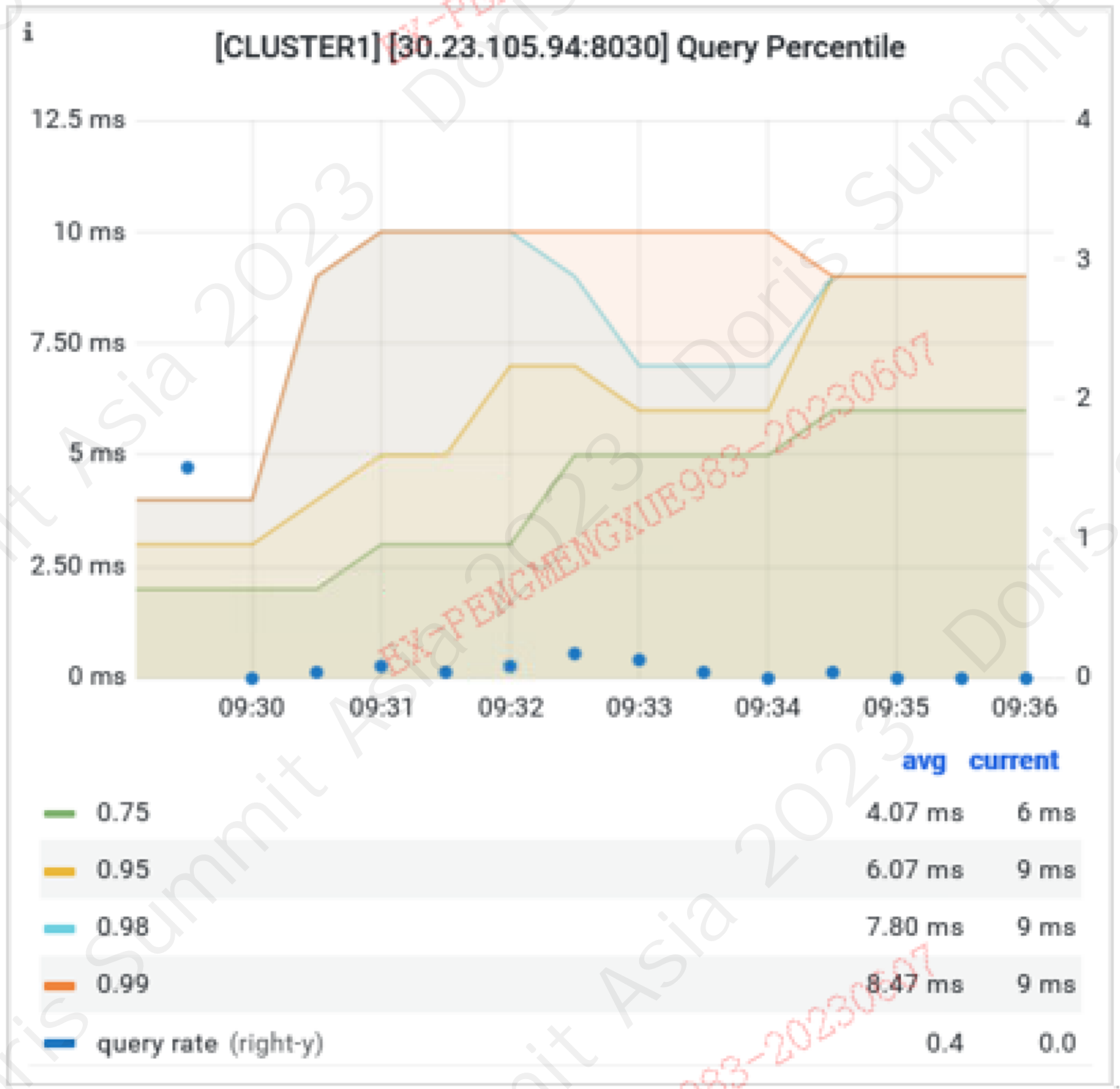


2. 调整数据模型：服务器压力



总结与思考

- 1 在启用缓存前提下，单用户访问，页面满足 4s 内加载完成，接口满足 600ms 内返回。并发场景下，最优 TPS 达到 300，CPU、内存、磁盘和 IO 满足 80% 以下。
- 2 经评估，3台 8C 16G 的 BE 节点至少可以缓存上万指标（4条SQL/指标）。
- 3 对于数据量大的多表关联，借助 AuditLoader 收集高频 SQL，构建宽表替代星型模型。



未来规划

- 1 整合 Flink CDC + IceBerg + Apache Doris 构建实时分析
- 2 探索多表关联下的查询优化，比如构建多表物化视图
- 3 将中台其他产品迁入 Apache Doris



获取更多社区动态与最佳实践

Apache Doris 官方平台:

- Apache Doris 官网: doris.apache.org
- Apache Doris GitHub: github.com/apache/doris/

获取更多峰会资料:

- Doris Summit 峰会官网: doris-summit.org.cn
- Doris Summit 峰会回放: <https://space.bilibili.com/1196172099/channel/collectiondetail?sid=1824324>