

Apache Doris 如何实现 盲测性能 10 倍提升

李昊鹏 Apache Doris PMC 成员

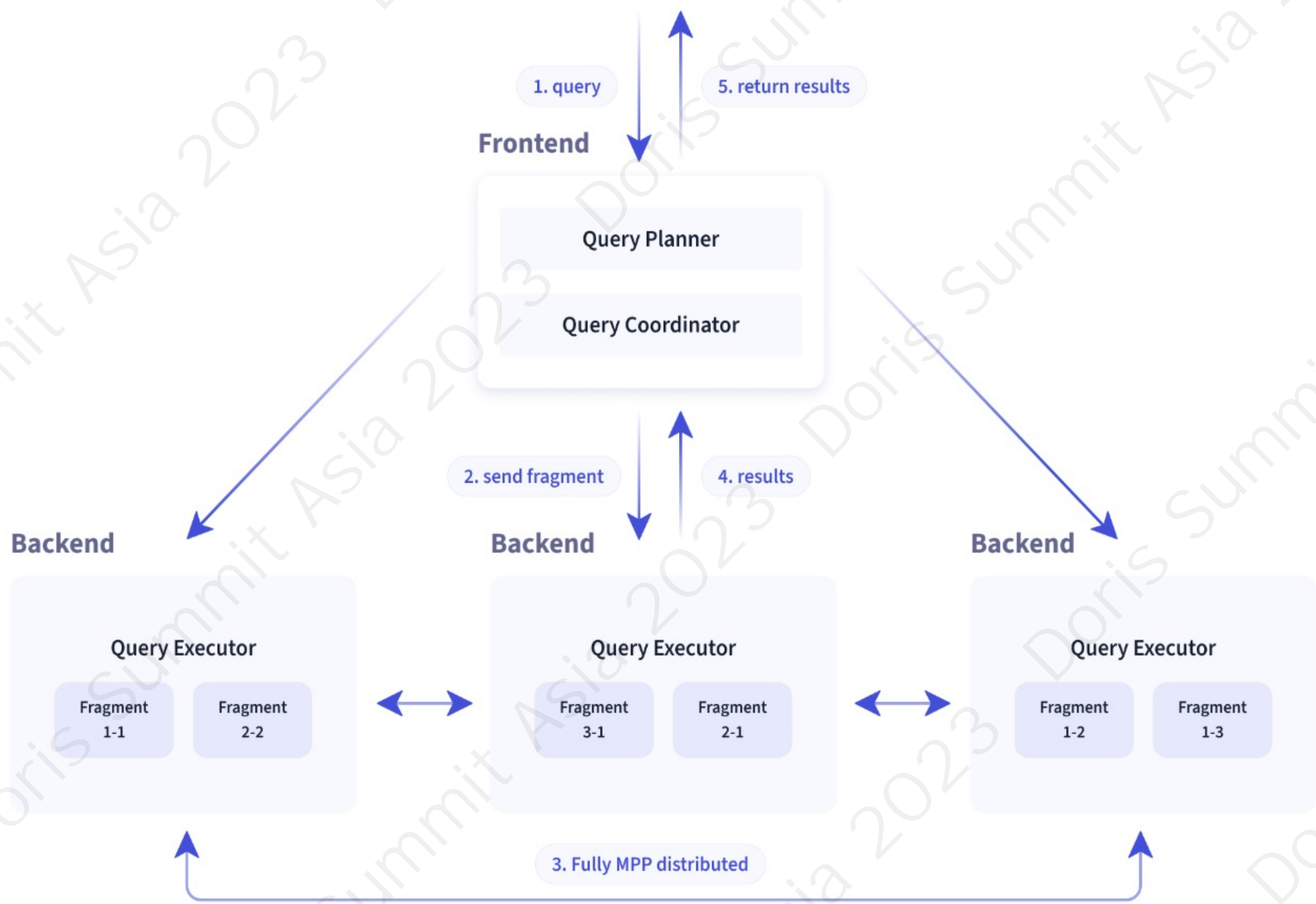
熊仲健 飞轮科技 资深技术专家

目录

1. Apache Doris 查询总体架构简介
2. Apache Doris 执行引擎演变
3. Apache Doris 执行引擎后续规划
4. Apache Doris 新优化器特性和架构
5. Apache Doris 新优化器总结、挑战和展望
6. Apache Doris 盲测性能提升
7. 欢迎加入 Apache Doris社区

1 Apache Doris 查询架构简介

Apache Doris 查询架构简介



优化器

- 查询改写：基于规则 + 基于代价
- 查询优化：Cascades + DPHyper
- 统计信息 & 代价模型
- 分布式计划优化

执行引擎

- 基于列存的向量化执行引擎
- 新 Pipeline 并行执行引擎
- Adaptive Query Execution

2

Apache Doris 执行引擎的演变

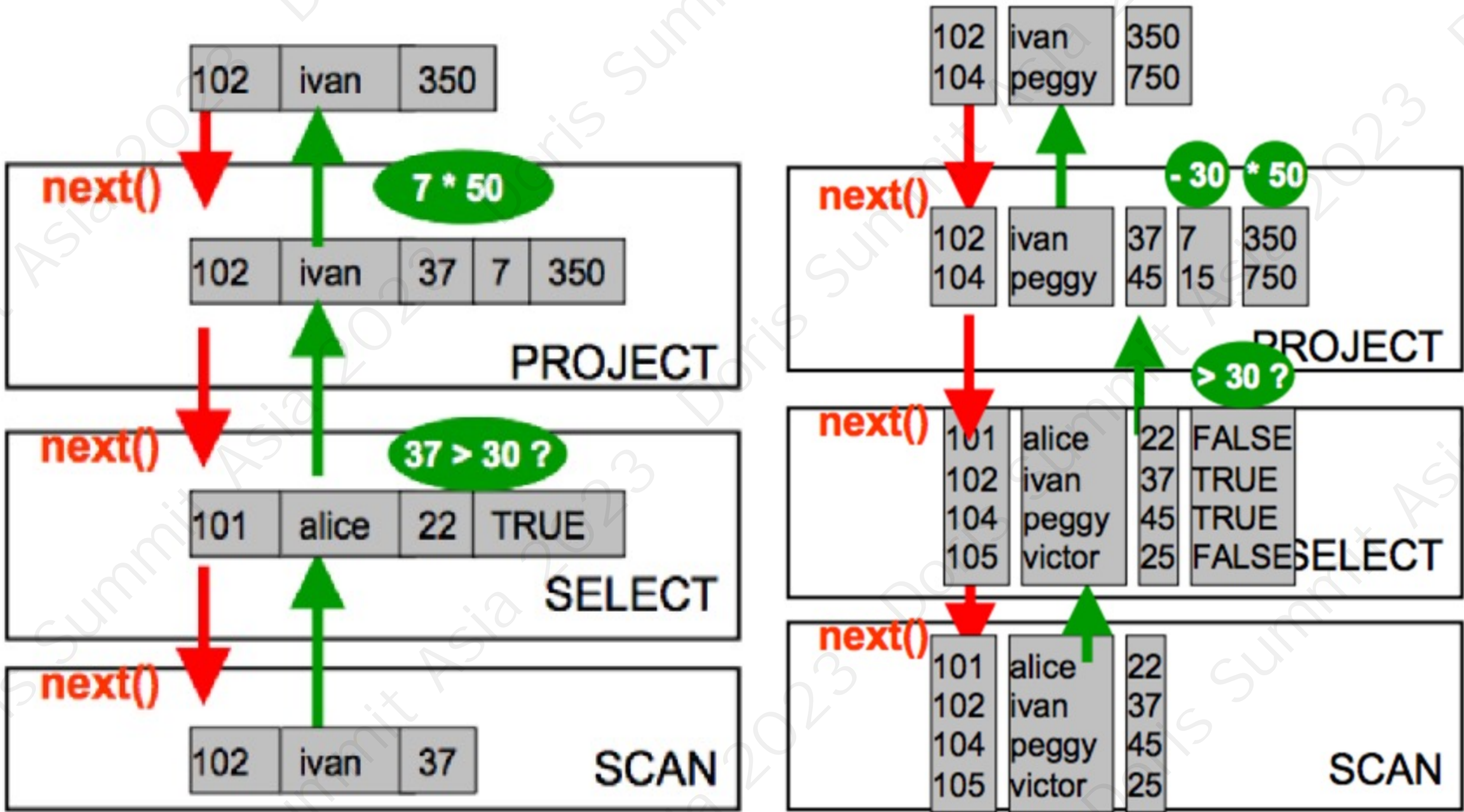
Apache Doris 执行引擎的进化



MPP -> 向量化 -> Pipeline : 多机->单核->多核

性能释放的三部曲

解决单核性能利器：向量化



向量化的意义

- Cache亲和度
- 规避虚函数的调用
- 分支预测
- SIMD化

Apache Doris 老执行引擎的问题

多核利用能力弱

- 数据和执行线程强绑定，无法解耦
- 扩展性差，很难找到合适的并行设置

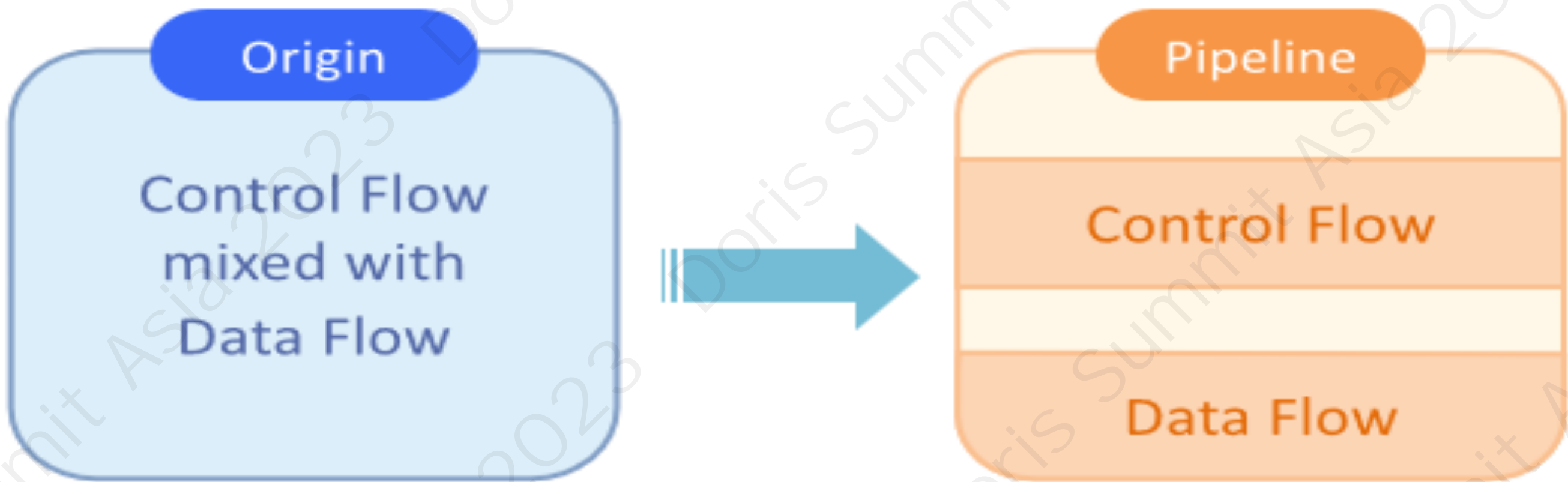
线程可控性差

- 线程池一旦打满，线程的假性死锁
- 线程切换开销较大
- 阻塞的算子会占用线程资源

资源管理能力弱

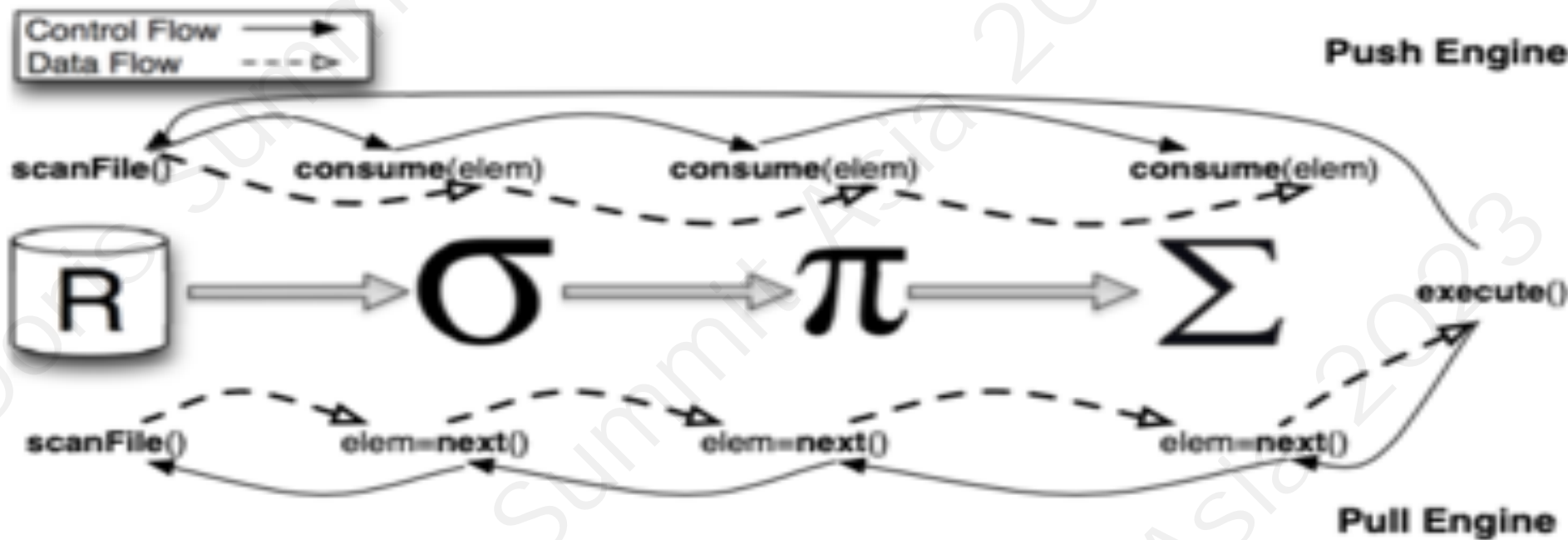
- 混合负载下的小查询饥饿问题
- 缺乏混合负载的内存和CPU的隔离能力

改变数据流的 Pipeline 执行引擎



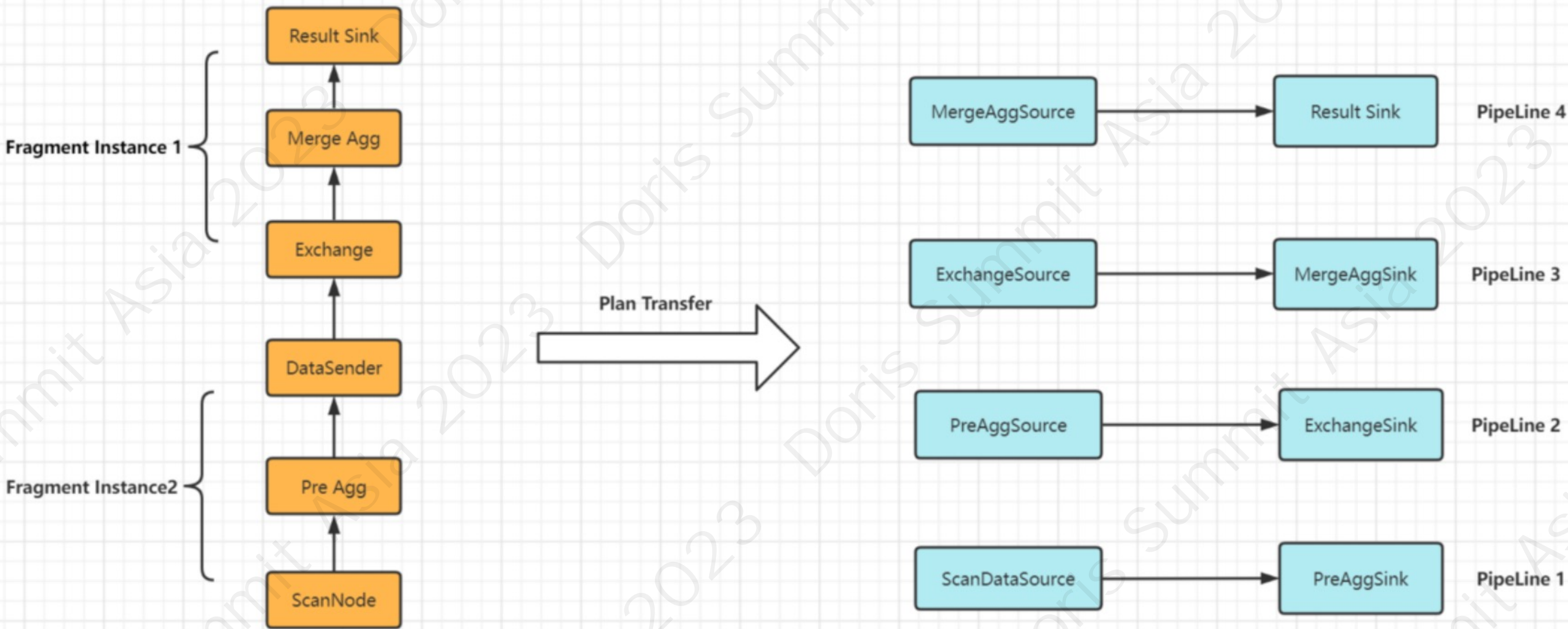
Pipeline 执行引擎

- 更加灵活的并发控制，解决查询导致的线程膨胀问题
- 更加高效的CPU资源的利用能力，降低线程调度产生的开销
- 自适应的并发控制的能力

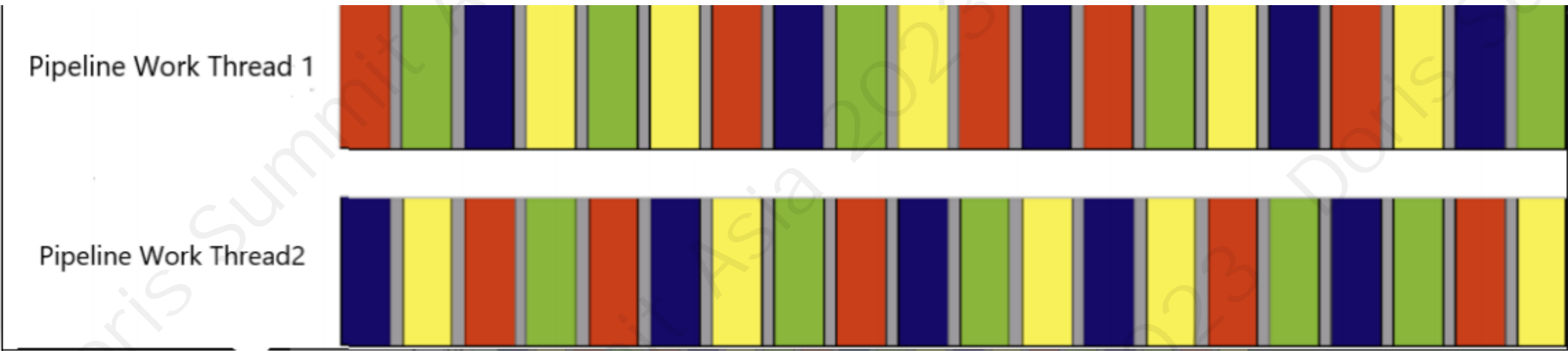


充分利用多核的 Pipeline 执行引擎

SQL: SELECT SUM(LO_EXTENDEDPRICE * LO_DISCOUNT) AS revenue FROM lineorder_flat

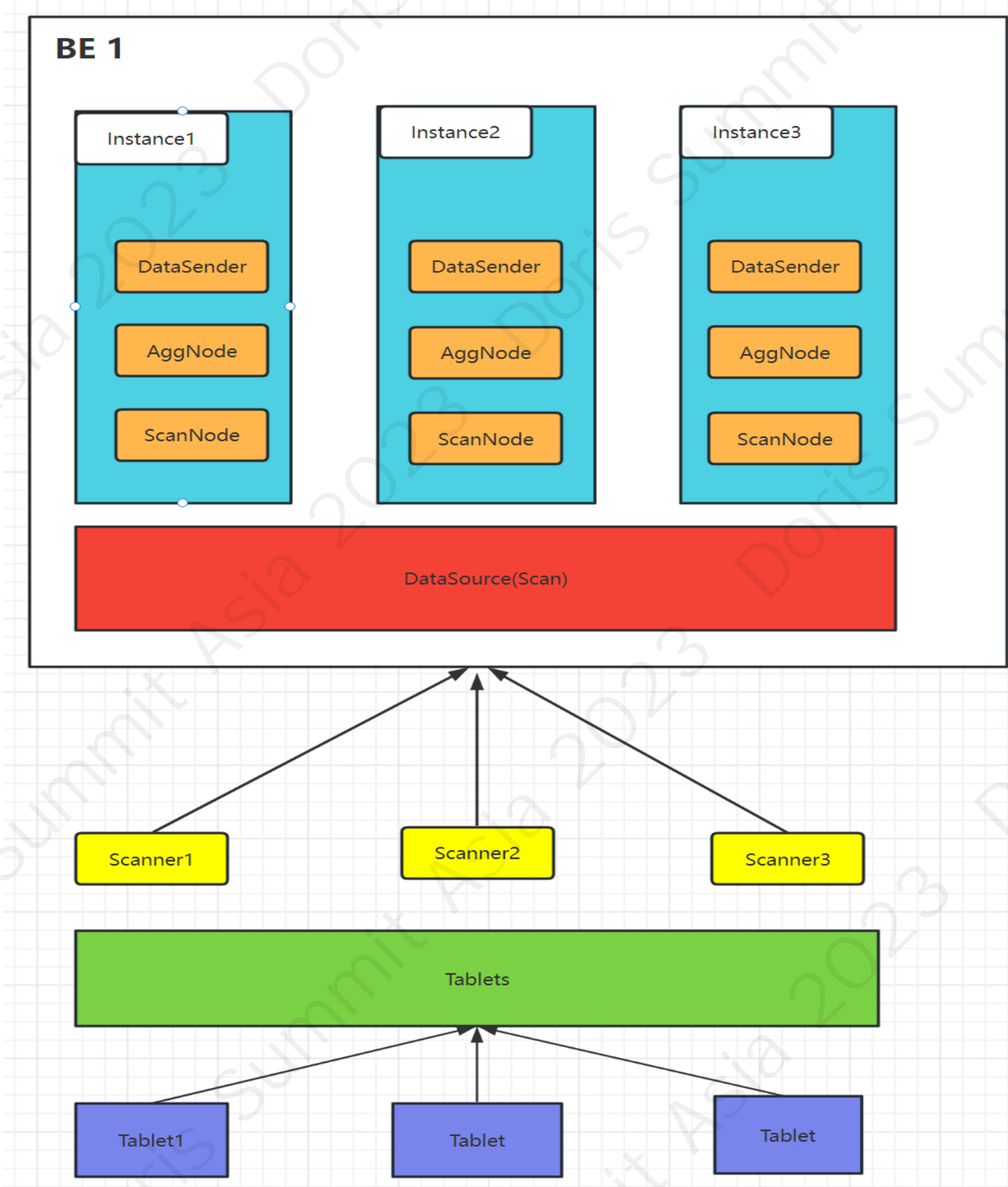


• 查询任务的拆解



• 多核执行查询任务调度

自适应的数据并行



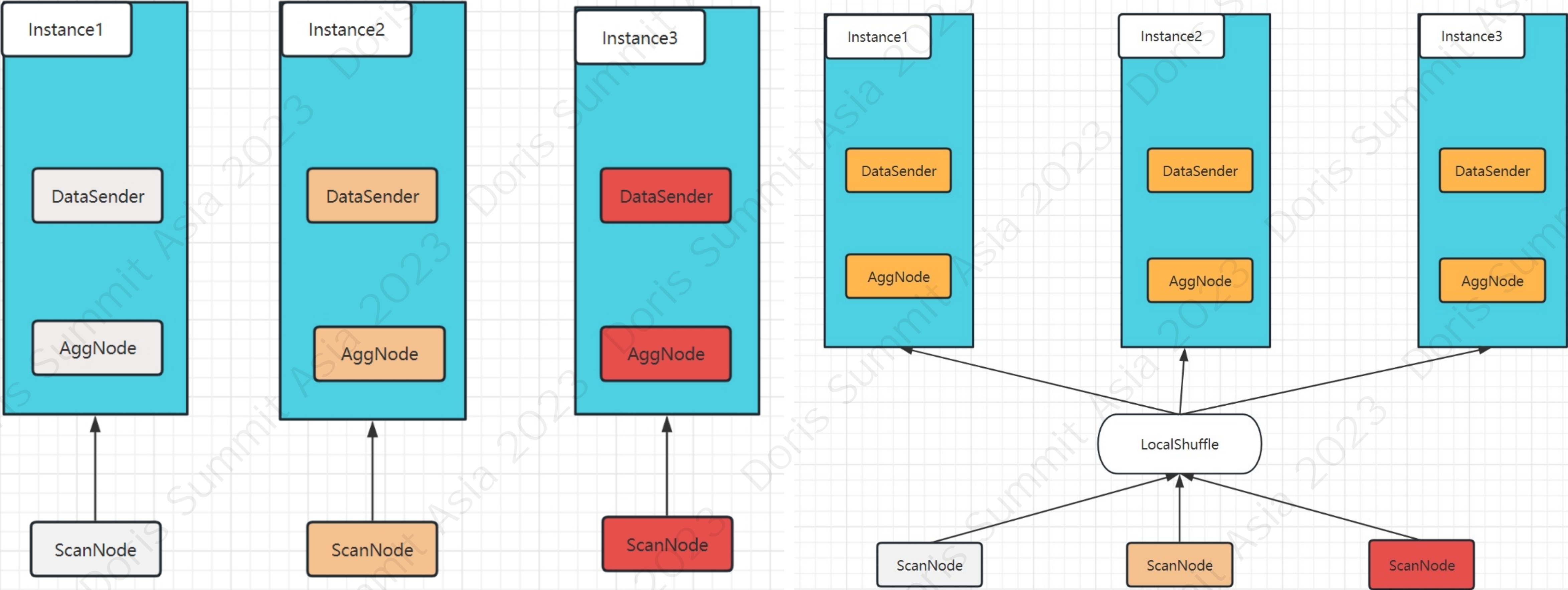
经典的 Fork Join 模型

- 解决建表数据分配不均的问题
- 多核并行化计算执行
- 多核并行化数据扫描

3

Apache Doris 执行引擎后续规划

更合理的数据并行能力



基于 Local Shuffle 的数据再分布策略

4 Apache Doris 2.0 新优化器特性和架构

优化器是分析数据库/大数据引擎的核心组件

数据库大脑

选出好计划

避免差计划

优化查询性能
减少资源消耗

*Once Velox standardizes the Execution Engine, the remaining differences will be in the **Optimizer** and Execution Runtime, which are the **true differentiating points between engines**.*

-- Andy Pavlo 《Database Review 2022 and 2023 Prediction》

Apache Doris 1.X 旧优化器的局限

基于过程优化

- 基于过程
- 扩展性差
- 维护成本高

优化手段有限

- 优化手段匮乏
- 人工介入调优
- 调优成本高

基础设施缺失

- 不支持统计信息
- 缺少代价模型，无法基于代价优化
- 没有维护数据分布/序等物理属性，无法生成高效分布式计划
- 运行时优化能力薄弱

Apache Doris 2.0 新优化器特性

规则框架&丰富的规则库

- 规则框架替代过程框架
- 扩展性好，可维护性高
- 丰富的优化规则

统计信息和代价模型

- 支持自动/手动，支持全量/采样
- 丰富的统计信息类型
- 精准的统计信息推导
- 持续迭代的代价模型

基于代价的查询改写框架

- 精准评估改写规则是否应该实施
- 代价改写规则不断扩充中

物化CTE支持

- 减少公共重复计算
- 公共Filter提取下压
- 公共Runtime Filter提取下压
- CTE Sender分离优化

Cascades + DPHyper 查询优化框架

- 兼顾计划质量 & 枚举效率
- 维护物理属性，优化分布式计划
- 支持大规模复杂SQL的查询优化

增强的运行时优化

- 增强的Runtime Filter(RF)下压支持
- 基于统计信息的RF自适应裁剪
- Partition TopN窗口函数计划优化

Apache Doris 2.0 新优化器优化流程



Apache Doris 新优化器查询改写

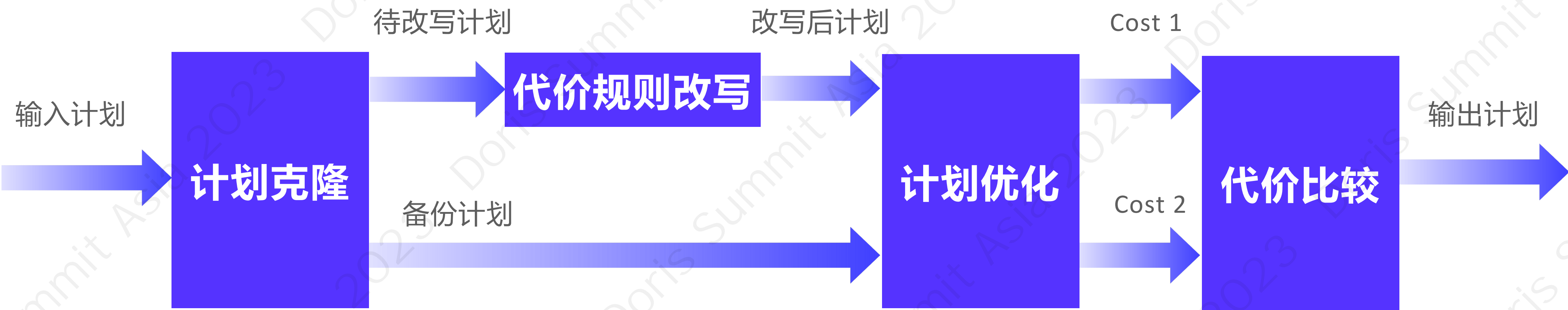
规则改写

- 谓词下推
- 常量折叠
- 分区/列裁剪
- 外连接消除
- 谓词推导
- 子查询提升
- Limit下压
- 物化CTE
-

代价改写

- 窗口函数消除子查询
- Set算子Distinct推导
- Distinct下压
- OR-expansion
-

Apache Doris 新优化器代价改写框架



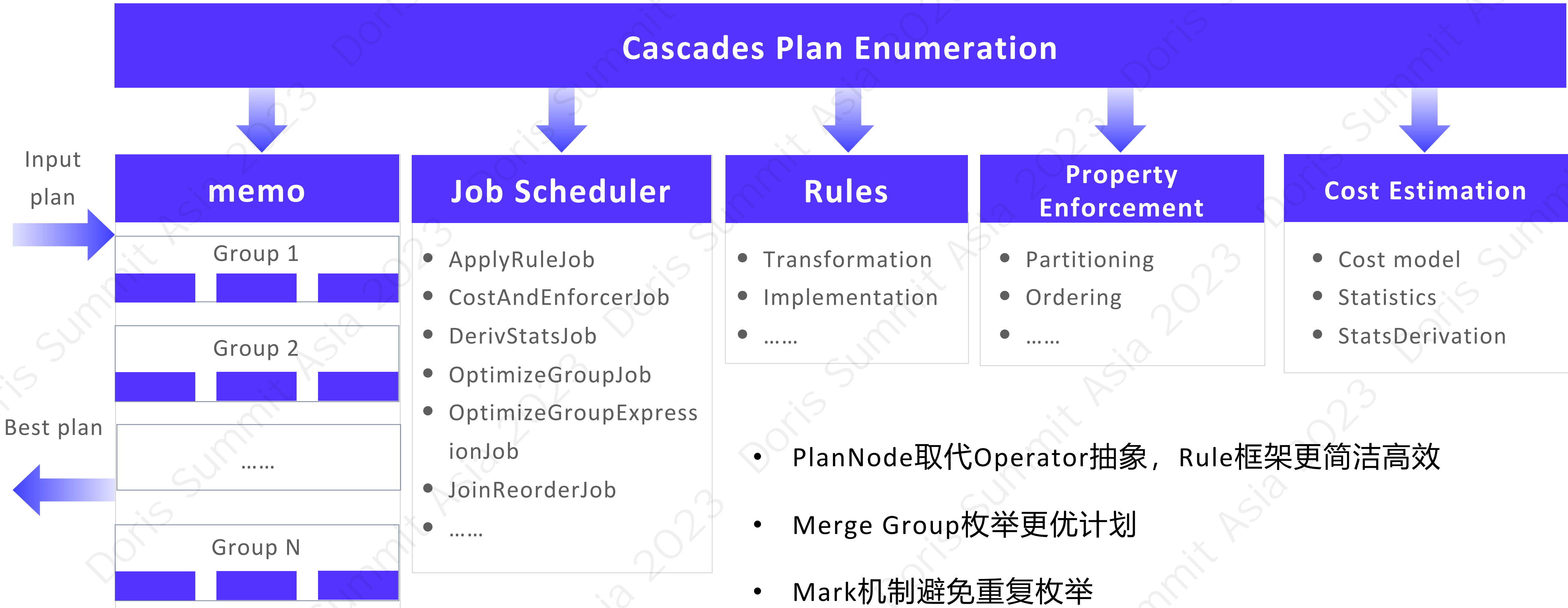
```
select sum(l_extendedprice) / 7.0 as avg_yearly
from lineitem, part
where
  p_partkey = l_partkey
  and p_brand = 'Brand#23'
  and p_container = 'MED BOX'
  and l_quantity < (
    select 0.2 * avg(l_quantity)
    from lineitem
    where l_partkey = p_partkey);
```

TPC-H Q17
子查询窗口函数代价改写

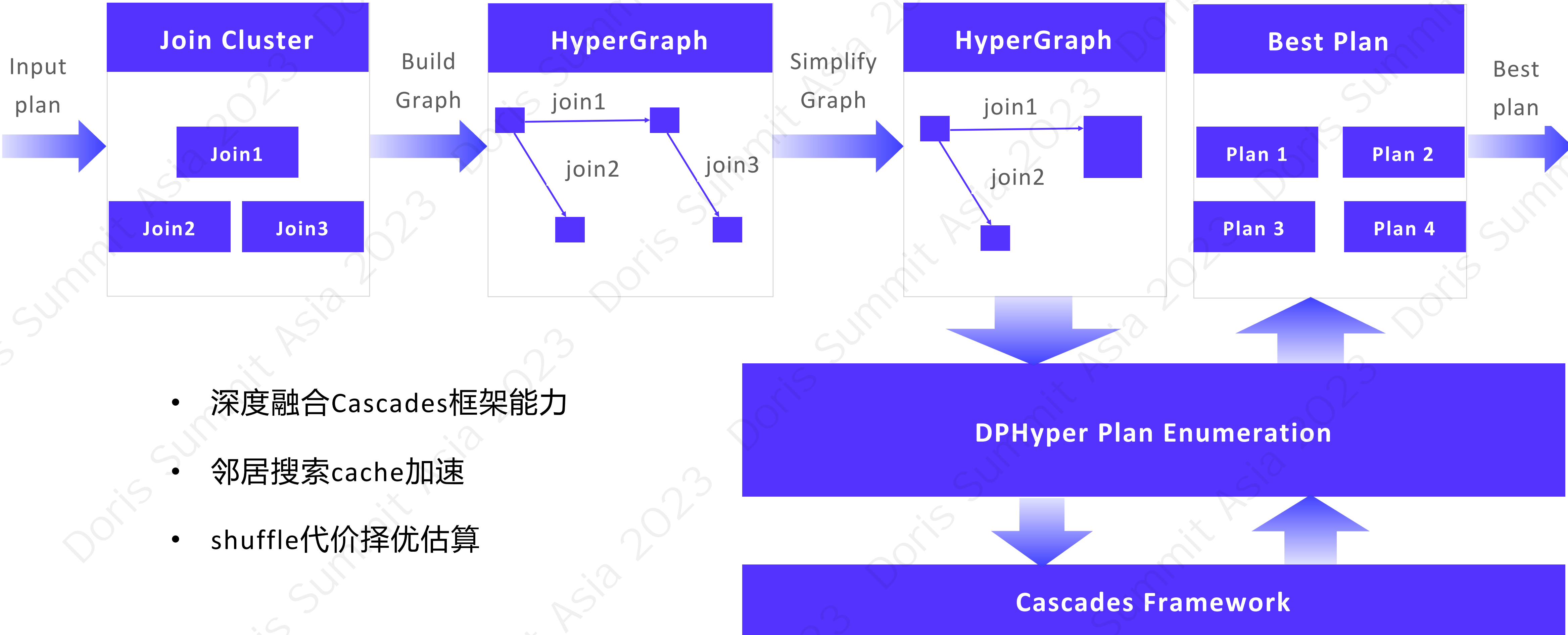
```
select sum(l_extendedprice) / 7.0 as avg_yearly
from (
  select l_quantity,
         l_extendedprice,
         0.2 * avg(l_quantity) over (partition by l_partkey) as val
  from lineitem, part
  where p_partkey = l_partkey
        and p_brand = 'Brand#23'
        and p_container = 'MED BOX') v
where l_quantity < val;
```

- 代价比较后仅当代价优于原计划才确定改写，否则不改写
- 代价规则改写后仍然需要进行纯规则改写迭代

基于经典 Cascades 的优化框架 (计划质量)



基于 DPHyper 的优化框架 (枚举效率)



Apache Doris 新优化器其他增强特性

统计信息和代价模型

- 统计信息收集
 - 支持自动/手动收集
 - 支持全量/采样收集
- 更丰富的统计信息类型
 - ndv, avgSize, numNulls, etc
- 更精准的统计信息推导
 - Filter/Having/GroupBy/Join/Window Function
- 持续迭代的代价模型
 - broadcast join基于build/probe代价调优
 - size代替rows的shuffle代价调优

运行时优化增强

- Runtime Filter (RF)下压计划增强
 - 支持下压SetOperation
 - 支持下压子查询
 - 支持下压CTE Sender
 - 支持级联下压
 - 支持多发
 - 支持传播
- 基于统计信息的RF自适应裁剪
 - 基于统计信息的rf过滤率估计
 - 无效rf自动识别 & 裁剪
- Partition TopN窗口函数计划优化
 - 节点本地分窗口计算topN + 二阶段全排
 - topN后减少二阶段计算数据量

物化CTE支持

- 减少公共重复计算逻辑
- 目前基于规则，未来基于代价
- 公共Filter/RF提取下压物化CTE
- 物化CTE发送端分离优化
 - 独立的发送端Partition/Projection/Filter/RF

5

Apache Doris 新优化器总结和展望

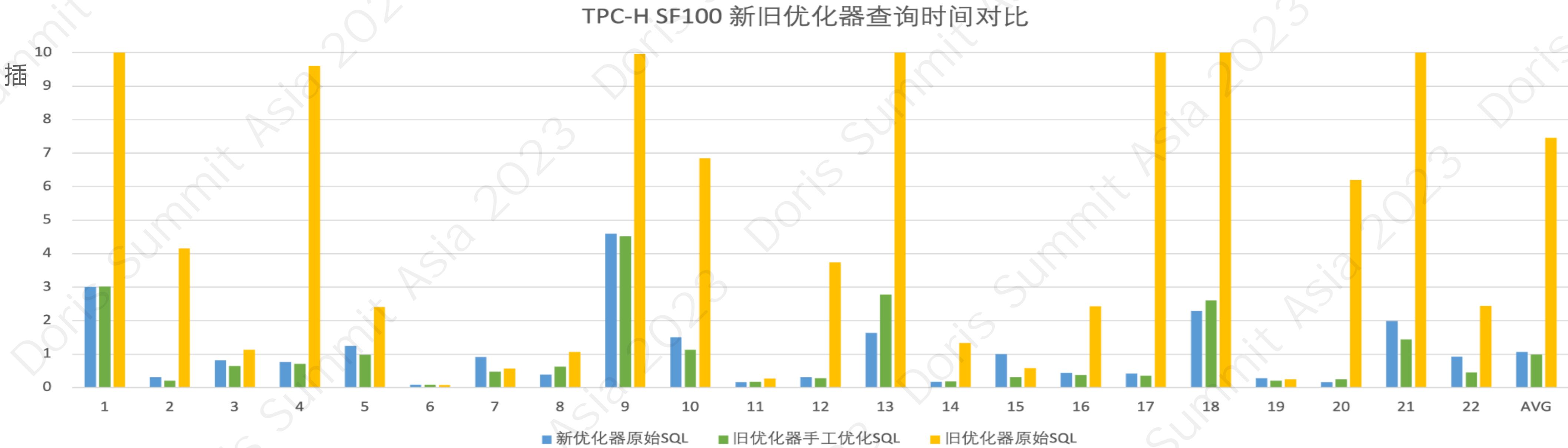
Apache Doris 新优化器性能总结、挑战和展望

总结	挑战	展望
<ul style="list-style-type: none">• 6大增强特性• 更易用，更稳定，更强大• 支持场景更丰富	<ul style="list-style-type: none">• 代价维度简单 & 估算不准• 优化规则应用顺序和策略不优• 超大规模复杂计划的枚举质量和枚举效率不优• 数据独立和均匀假设broken• 运行时自适应策略不优	<ul style="list-style-type: none">• 新代价模型• RBO规则框架完善 & RBO规则扩充 & 高阶代价改写规则扩充• 持续优化Cascades/DpHyper框架• 基于直方图的数据倾斜处理• 基于统计信息的RuntimeFilter自适应

6 Apache Doris 盲测性能 10 倍提升

Apache Doris 盲测性能 10 倍提升

- TPC-H SF100: 3BE环境
- 新优化器原始SQL性能表现与旧优化器手工优化SQL相近
- 原始SQL下新优化器对比旧优化器10倍性能提升





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>