Apache Doris MCP Server & Client 产品介绍

苏奕嘉 Apache Doris Committer 飞轮科技 资深解决方案架构师

ONO

目录

1. MCP 是什么?

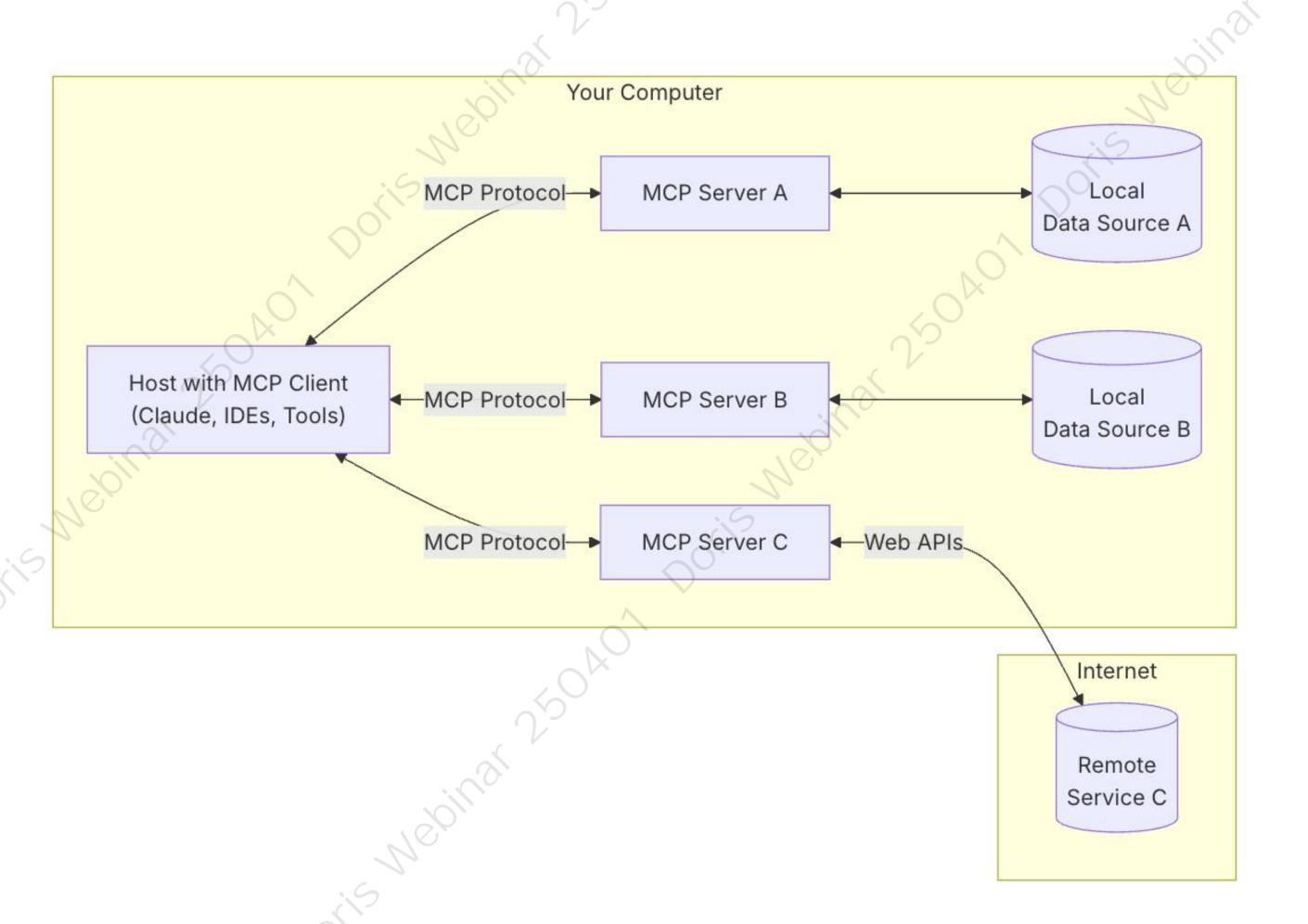
2. Apache Doris MCP 能力现状

3. Apache Doris MCP 能力展示

定义与定位

- MCP (Model Context Protocol): AI 大模型上下文拓展协议,旨在增强 AI 与数据系统的交互能力。
- Apache Doris MCP Server:基于大语言模型与Doris 实时数仓,提供自然语言交互、SQL 优化、智能 ETL 等能力的全链路数据分析工具。

Opiis

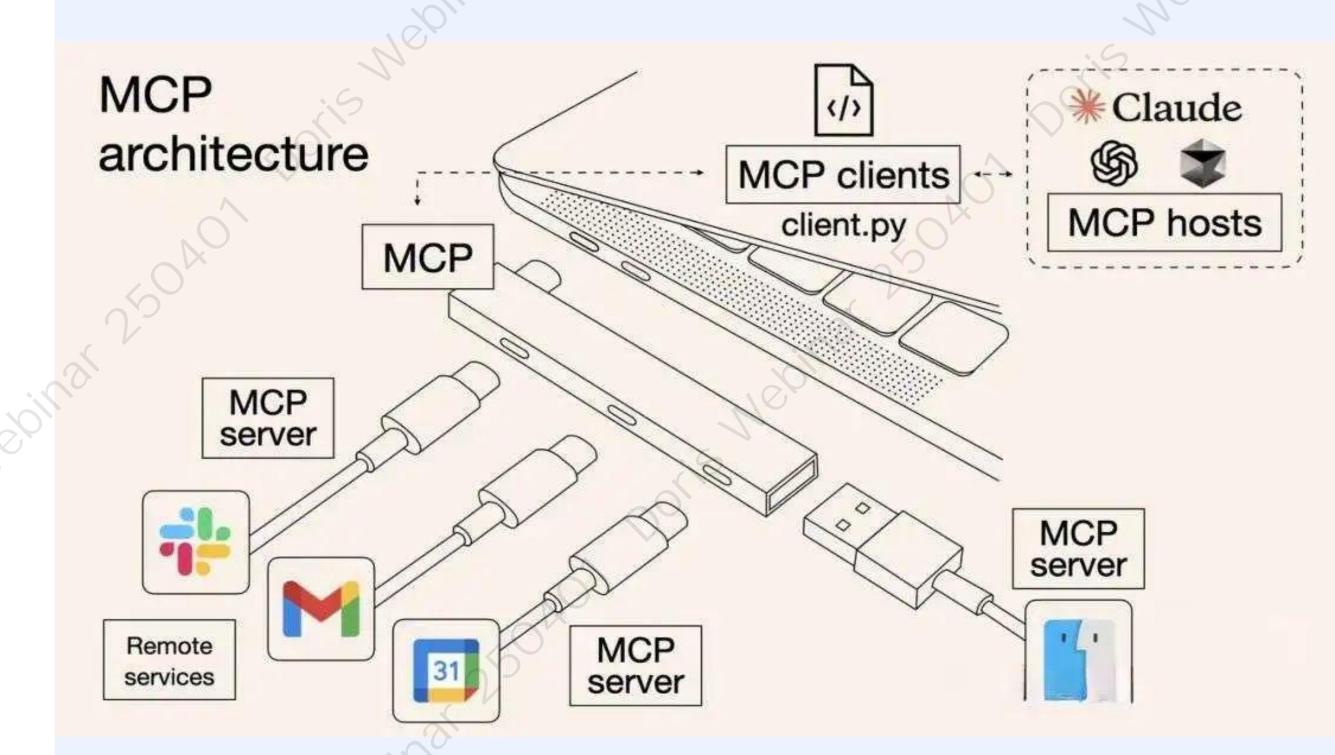


核心价值

Ne₁₀

- 零门槛分析: 非技术用户可通过自然语言直接操作数据。
- 模块化扩展:支持 NL2SQL、集群诊断、智能 ETL 等多场景能力按需扩展。

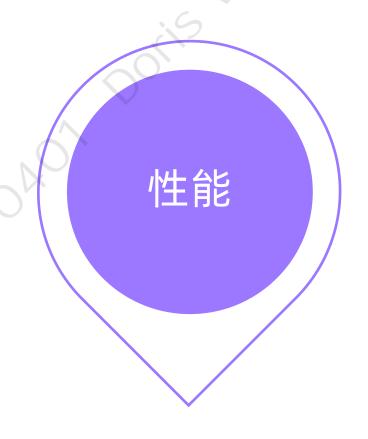
Opiis



Doris MCP Server 特有优势

高效

原生





数仓直出

深度融合

Doris 加持

极简适配

无论是使用 Doris 做 OLAP 引擎,还是做实时数仓,架 设即用,无需额外开发 基于 Doris 元数据库检索信息, 适配 Doris SQL 语法及特性, 非通用化 MCP

Opiis

不同阶段可选配不同 LLM 完成加工,最重要还是 Doris 在大规模数据检索下的超强性能

服务极简、交互极简,使用标准 SSE 协议通讯,自带可视化 Client 端,开箱即用

目录

1. MCP 是什么?

2. Apache Doris MCP 能力现状

3. Apache Doris MCP 能力展示

已实现功能







NL2SQL(自然语言转SQL)

用户输入自然语言 → 系统智能转换为 SQL → 执行并返回可视化结果

Mejo,

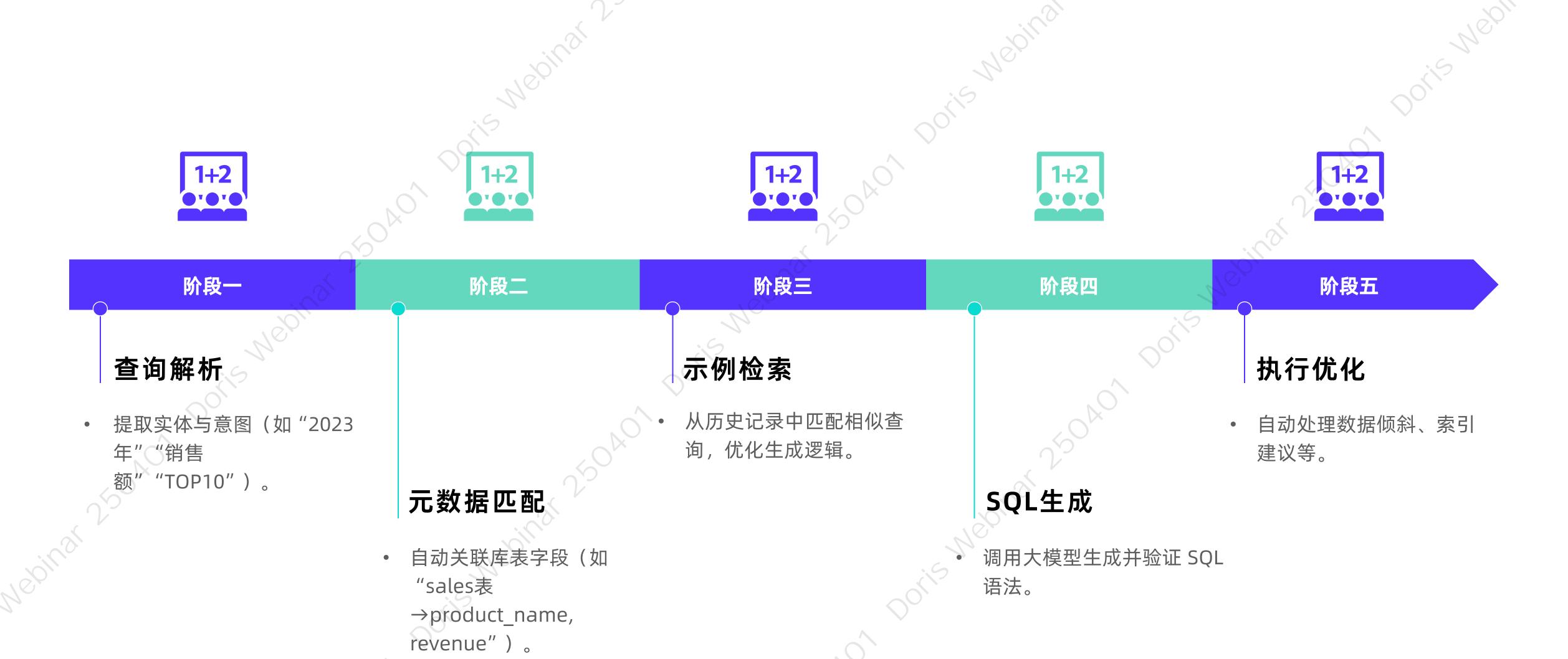
SQL自动优化分析

自动优化SQL并给出建议,防止未优化 SQL导致资源占用过多

MCP Client (配套客户端)

无需特定工具使用 Server, 自带 SSE 支持协议的客户端, 开箱即用

Doris MCP NL2SQL 流程详解



Doris MCP NL2SQL 底层要点



元数据预处理

加速查询响应 支持增量刷新

OORIS



元数据智能检索

多级识别, 避免透传或重复获取



智能缓存

成功反馈的结果自动缓存,自动路由



错误自愈

自动重试、SQL 语法修正、 降级处理

Doris MCP NL2SQL 能力亮点



零训练-开箱即用

- 支持预训练集拓展检索
- 支持无标注库表字段业务分析



跨库智能检索

- 支持多 DB 跨库检索(非跨组件检索)
- 支持检索内容自动加入训练集(越用响应越快、检索越准确)



适配多源 LLM

- 深度适配 DeepSeekV3、R1,支持主流LLM
- 支持本地 Ollama、vLLM、MLX 等框架

Opiis



改造简易高自由度

- 支持 Prompt 拓展词自定义
- 支持正则自定义库表先后检索排序

Doris MCP SQL Auto Optimization 能力亮点



Apache Doris MCP Server - SQL AutoOpt

错误 SQL 补充优化

加快开发进度,提升开发效率,错误语法校验,标点自动调整

复杂 SQL 简易优化

根据 Profile、Explain、业务释义、元数据信息等,优化复杂查询,节省集群资源,加速查询效率

历史 SQL 释义说明

前人留存的冗长 SQL 书写不规范、业务历史悠久,短时间难以理解和改造,助力快速理解历史 SQL 含义

Doris MCP Client

- SSE 原生适配支持
- 前后端分离部署
- · 多语言支持(中文/English)
- 现支持 NL2SQL、SQL-AutoOpt,后续将持续加入 新增功能
- 可移植性强,采用 Vue.js 开发,可嵌入至任意项目内

Apache Doris MCP 客户端 基于MCP协议的智能分析工具 ◎ LLM 配置管理 ※ SQL 智能优化分析 ─ NL2SQL 自然语言查询 提交您的SQL查询,系统将自动分析性能瓶 颈,提供优化建议,并生成改进后的SQL语 可视化建议。 管理配置 系统状态 服务器地址 192.168.31.13:3000 ☆ SQL Optimization **Apache Doris MCP Client** Intelligent Analysis Tool Based on MCP Protocol ☆ SQL Intelligent Optimization LLM Configuration Management NL2SQL Natural Language Query Use natural language to describe your data Submit your SQL query, and the system will Customize the large language models (LLM) needs, and the system will automatically automatically analyze performance used for different processing stages, and convert it into an SQL query and execute it, bottlenecks, provide optimization select appropriate models and parameter while providing business analysis and suggestions, and generate improved SQL configurations for different tasks. visualization suggestions. statements. Manage Configurations System Status **Connection Status** 192.168.31.13:3000 **Server Address**

目录

1. MCP 是什么?

2. Apache Doris MCP 能力现状

3. Apache Doris MCP 能力展示



目录 1. MCP 是什么? 2. Apache Doris MCP 能力现状 3. Apache Doris MCP 能力展示

Apache Doris MCP Server 现有优化



NL2SQL

- 联邦跨源检索查询
- 多行业预制分析模板切换
- 多源文件内容检索增强业务理解能力
- 自适应 Doris 权限管控体系

Opiis

• RAG知识库的结合



SQL-AutoOpt

- 物化视图、联邦查询等特性自动应用
- CBO、Pipeline 引擎结合自动 Hint
- WorkloadGroup 资源自适配
- Profile、Explain 多版本精细化解读

Apache Doris MCP Server 未来规划

集群诊断

通过智能化的诊断机制,系统能够自动收集并分析关键指标,识别潜在风险,提供专业级的诊断报告和修复建议

智能 ETL

通过 LLM 赋能实现"低代码/零代码"的 ETL流程构建, 该功能将覆盖数据集成全生命周期

数据智能治理

融合AI技术与数据管理最佳实践,打造全方位、自动化的数据治理解决方案



集群运行状态诊断

异常宕机状态诊断

高负载优化建议

Workload Group 自动路由



自然语言定义ETL

智能数据映射

自适应转换逻辑

异常数据智能处理

自动调优与监控



智能元数据管理

自动化数据质量监控

智能数据分级分类

合规风险智能识别

数据价值评估

一些说明

- 单组件无法完成极高准确度的 NL2SQL 能力,需要配合数仓建模、数据治理、元数据统一管理等方案并行操作
- Apache Doris MCP Server & Client 目前处于初期完善阶段

DORIES ME

•

Mejo,

250AO

开源与热点问题答疑

• Q1: 是否开源

· Q2: 迭代速度

• Q3: NL2SQL 准确度如何

• Q4: 是否支持其他数据来源

• Q5: 硬件环境标准

Nejo,

Q6:元数据缓存污染如何处理?

