

使用 Apache Doris MCP 构建智能查询解决方案

苏奕嘉

Apache Doris Committer

飞轮科技 资深解决方案架构师



目录

1. Agent-Facing Analytics 时代变革

2. 面对 AI 新时代的挑战和应对

3. Doris MCP Server 设计与实现

4. Doris MCP Server 案例分享

Agent时代的到来

从传统BI到Agent-Facing Analytics的变革

传统BI时代

(2010-2020)



传统报表

- 线性、可预测工作流程
- 人工驱动分析过程
- 查询量相对有限
- 实时性要求宽松
- 预定义报表仪表盘

传统BI成熟期

Agent时代

(2025+)



智能Agent

- 自主决策、智能工作流
- AI驱动分析过程
- GB-TB级数据量查询
- 秒级毫秒级响应
- 动态查询实时分析

全面Agent化

2025

Agent革命年

核心变革特征

- 用户角色：从人类分析师到AI Agent
- 工作模式：从被动响应到主动决策
- 查询特征：从预定义到动态生成

- 性能要求：从宽松到极致优化
- 数据规模：从MB级到PB级
- 并发需求：从数十到数万+

- 技术架构：从单体到分布式
- 决策速度：从分钟级到毫秒级
- 智能程度：从工具到伙伴

Agent工作模式的本质差异

传统用户 vs AI Agent

传统用户模式

线性、串行处理流程

1. 业务分析师提出问题

明确业务需求，人工定义查询逻辑



2. 数据工程师编写SQL

手工编写和优化，预定义查询模式



3. 等待查询结果

分钟到小时级延迟，串行处理模式



核心特征

• 线性可预测工作流程 • 查询量有限 (<100/天) • 分钟-小时级响应时间

VS

AI Agent模式

动态、并行处理流程

单一用户请求

探索性查询

关联分析

实时计算

上下文查询

多维分析

动态聚合

指标计算

依赖分析

向量检索

关系挖掘

模式识别

趋势预测

数据处理特征

- 数据量：GB-TB级（大集团PB级）
- 并发度：数千甚至上万次并发查询 • 响应时间：秒级-毫秒级要求

核心特征

• 动态高并发工作流程 • 查询量巨大 (>1000/秒) • 毫秒级响应时间

关键指标对比分析

传统模式指标

查询量：< 100次/天 并发数：< 10
响应时间：分钟-小时级 数据量：MB-GB级

性能提升需求
1000x+

Agent模式指标

查询量：> 1000次/秒 并发数：数千+
响应时间：秒级-毫秒级 数据量：GB-PB级

目录

1. Agent-Facing Analytics 时代变革

2. 面对 AI 新时代的挑战和应对

3. Doris MCP Server 设计与实现

4. Doris MCP Server 案例分享

Agent时代数据访问挑战

并发、实时性、智能查询优化

并发性能指数级挑战

1

并发量增长趋势

传统BI: 10-100 QPS
智能客服: 1,000-10,000+ QPS
推荐系统: 数万+ QPS
需要MPP架构支撑

电商平台Agent集群

客服Agent + 推荐Agent + 库存Agent
价格Agent + 营销Agent + 风控Agent
数百个Agent同时访问数据库
传统架构无法支撑

技术挑战

- 单点瓶颈问题
- 资源争用冲突
- 查询队列拥堵
- 需要分布式无主架构

实时性要求新高度

2

响应时间要求

传统BI: 分钟-小时级
Agent应用: 秒级-毫秒级
实时决策: 亚秒级响应
任何延迟都影响体验

金融风控场景

实时交易行为分析 + 历史信用记录
当前账户状态 + 欺诈风险评估
毫秒级风险判断, 零容忍延迟
延迟即风险

技术要求

- 秒级数据写入延迟
- 高并发主键等值查询
- 实时流处理能力
- 需要行列混存架构

智能查询优化挑战

3

查询模式变化

传统: 预定义查询路径
Agent: 动态生成查询
高度不可预测性
需要智能优化

复杂查询特征

多表关联分析 + 复杂查询依赖图
多维度聚合计算 + 实时指标计算
探索性数据发现 + 上下文关联
传统优化器失效

解决方案

- CBO智能优化器
- 多种索引策略
- 向量化执行引擎
- 适应动态查询模式

数据规模爆炸式增长

传统BI: MB-GB级数据量 → Agent应用: GB-TB级实时处理 → 大型企业: PB级日增数据

关键数据: 单次Agent请求触发数十到数百个查询 | 中等企业需支持数千并发Agent | 每个Agent每秒发起数次到十几次查询

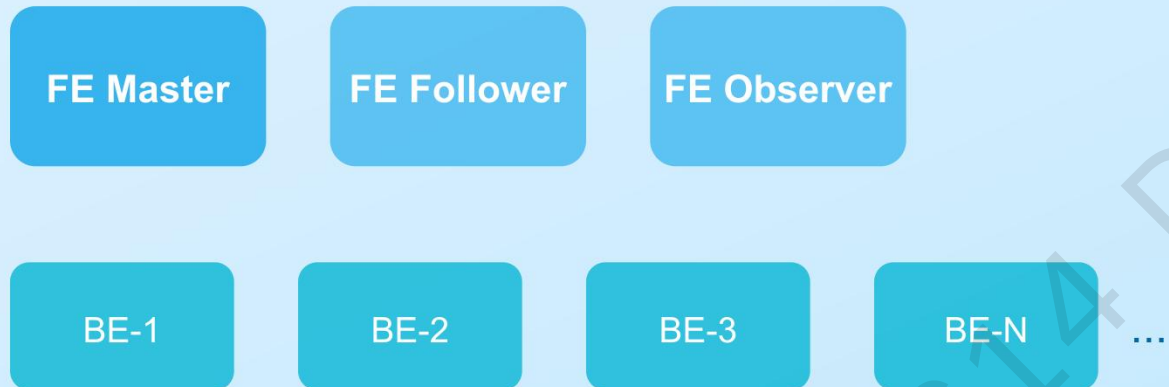
核心挑战: 传统数据库架构在Agent时代面临性能、实时性、智能化三重挑战, 急需新一代数据平台

Apache Doris 技术优势

专为Agent时代设计的智能数据平台

MPP分布式架构

无主节点设计，线性扩展并发能力



核心特性:

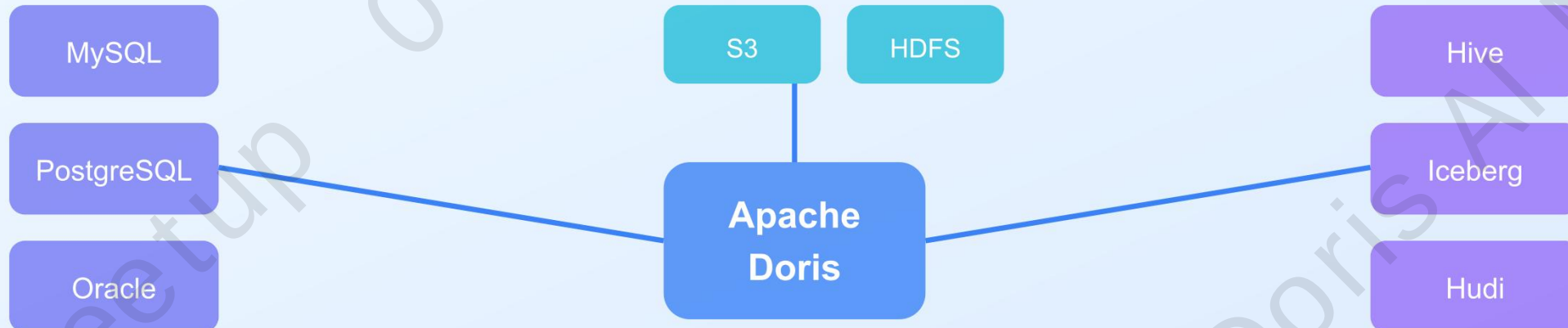
- 无主节点避免单点瓶颈
- 向量化执行引擎
- SIMD指令集优化
- 5-10倍性能提升
- 线性扩展能力

并发处理能力：数千到数万QPS

支持Agent高并发访问场景

Multi-Catalog统一数据访问

一套SQL接口访问所有数据源



智能查询优化：跨数据源查询自动优化

Agent获得完整数据视图，做出更准确决策

实时处理与向量检索

秒级数据写入，毫秒级查询响应



核心能力:

- 秒级数据写入延迟
- 低级查询响应
- 向量原生支持
- 复杂关联查询优化

应用场景：风控Agent毫秒级风险评估，推荐Agent实时特征计算

支持数万并发的高频点查操作

安全性与可观测性

企业级安全保障与全方位监控

细粒度权限控制

完整审计日志

SQL注入防护

Agent可观测性支持:

- Agent运行日志统一存储分析
- 性能指标智能分析
- 执行轨迹实时监控
- 异常行为快速定位

企业级安全：多层次认证 + 数据访问追踪 + 异常检测

为Agent自主决策提供安全可靠的数据访问保障

目录

1. Agent-Facing Analytics 时代变革

2. 面对 AI 新时代的挑战和应对

3. Doris MCP Server 设计与实现

4. Doris MCP Server 案例分享

MCP协议：AI应用的"USB-C接口"

Model Context Protocol - 连接AI与数据的标准化桥梁

就像USB-C统一了设备连接标准

笔记本

手机

平板

USB-C接口

显示器

硬盘

充电器

MCP为AI应用与数据源提供统一的协议规范

传统烟囱式开发模式

AI应用

MySQL接口
重复开发
高维护成本

PostgreSQL接口
重复开发
高维护成本

S3接口
重复开发
高维护成本

其他接口
重复开发
高维护成本

MySQL

PostgreSQL

S3存储

其他数据源

传统模式问题

- 重复开发，维护成本高 • 集成复杂，开发周期长
- 接口不统一，学习成本高 • 难以规模化部署
- 每增加一个数据源需要单独适配

MCP革命

MCP标准化开发模式

AI应用

Apache Doris MCP Server
统一数据访问接口 + 联邦查询能力

Apache Doris

MySQL

PostgreSQL

S3存储

数据湖

MCP + Doris优势

- 标准化接口，一次开发处处运行
- 联邦查询，统一访问多数据源

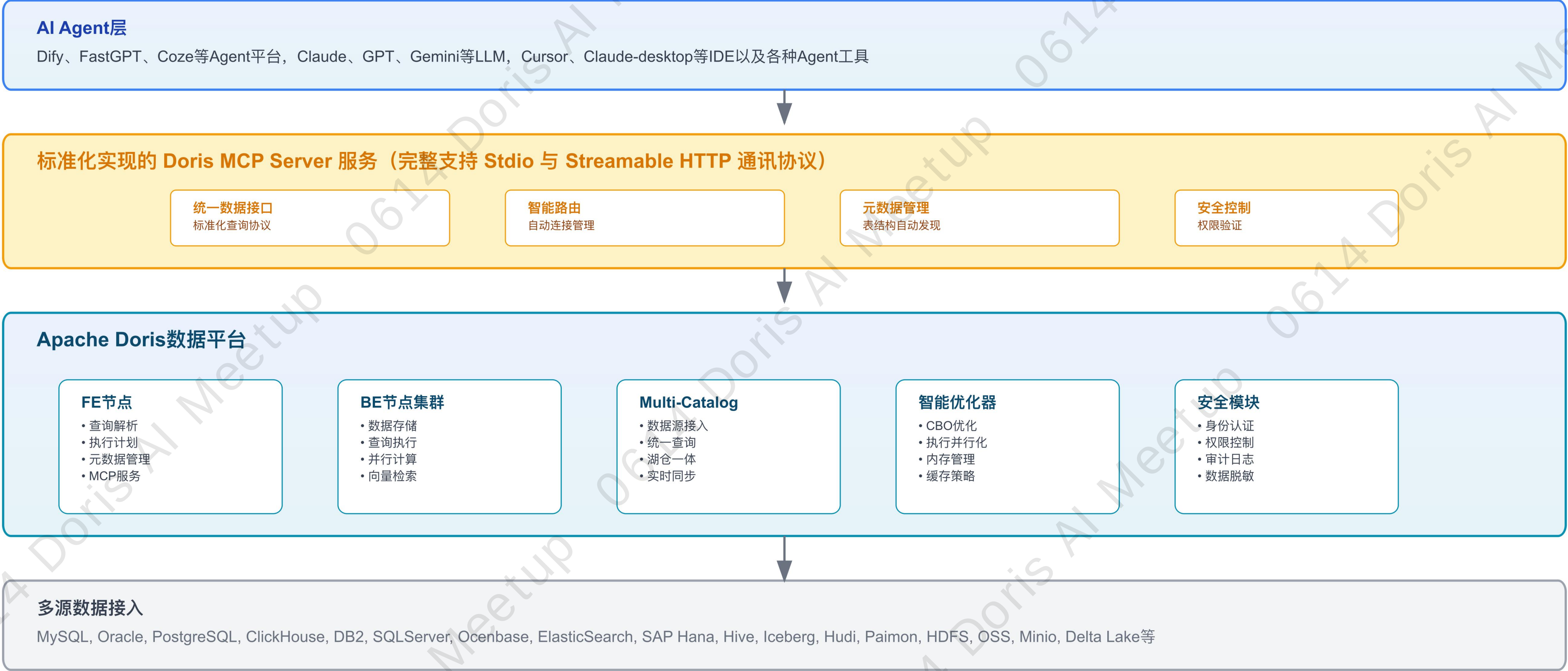
MCP + Apache Doris：完美结合的技术价值

标准化协议接口 + Multi-Catalog联邦查询 + 高性能MPP架构

让AI开发者通过一套API访问所有数据源，专注算法创新而非底层数据访问技术细节

Apache Doris MCP架构实现

专为Agent时代设计的智能数据平台



Doris MCP Server 应用场景

赋能企业数字化转型，驱动行业智能化变革

企

企业智能分析助手

CEO询问："上季度各产品线的收入增长趋势如何？"

AI Agent自动连接财务、销售、产品数据库

实时生成多维度分析报告和可视化图表

支持自然语言追问和深度钻取分析

核心优势

- 跨部门数据统一查询，打破数据孤岛
- 毫秒级响应，支持高管实时决策需求
- 智能SQL优化，处理千万级数据秒级返回

金

金融风控智能助手

风控专员问："这笔交易是否存在欺诈风险？"

AI实时检索用户行为、交易历史、设备指纹

毫秒级风险评分和异常行为识别

自动生成详细风险报告和处理建议

核心优势

- 实时风险评估，支持万笔/秒交易并发
- 多维度数据融合，提升风控准确性
- 向量检索技术，快速识别相似风险模式

商

电商智能推荐助手

用户询问："帮我找适合春节送礼的产品"

AI分析用户画像、购买历史、季节趋势

实时计算个性化推荐和价格优惠策略

智能配置营销活动和库存调配建议

核心优势

- 亿级商品库实时检索，个性化推荐精准
- 多源数据融合，提升转化率和用户满意度
- 弹性扩展架构，应对促销活动流量峰值

医

医疗智能诊断助手

医生询问："这个患者的症状可能对应哪些疾病？"

AI检索病历、检查结果、医学知识库

快速匹配相似病例和诊断建议

生成个性化治疗方案和用药指导

核心优势

- 医疗数据安全访问，符合HIPAA等法规要求
- 多模态数据分析，整合文本、影像、检验数据
- 实时更新医学知识库，确保诊断建议准确性

目录

1. Agent-Facing Analytics 时代变革
2. 面对 AI 新时代的挑战和应对
3. Doris MCP Server 设计与实现
4. Doris MCP Server 案例分享

Dify + Doris MCP 构建 ChatBI

AI顶级开发平台 + 实时数仓工业界标准，颠覆传统数据探索模式

ChatBI 通讯架构：Streamable HTTP 通信模式



ChatBI 核心优势

自然语言交互

- 无需学习 SQL 语法
- 支持中文业务术语
- 多轮对话上下文理解
- 复杂查询意图识别
- 实时反馈和澄清

降低数据分析门槛 90%+

实时数据洞察

- 秒级查询响应
- 实时数据更新
- 动态仪表板生成
- 智能趋势预测
- 异常自动检测

决策效率提升 80%+

技术实现特性

SSE 通信模式

- HTTP 协议兼容性强
- 实时双向通信
- Web 应用集成友好
- 低延迟数据传输

智能 SQL 生成

- 自动表结构识别
- 复杂关联查询优化
- 查询性能自动调优
- 错误自动纠正

MCP 核心工具链

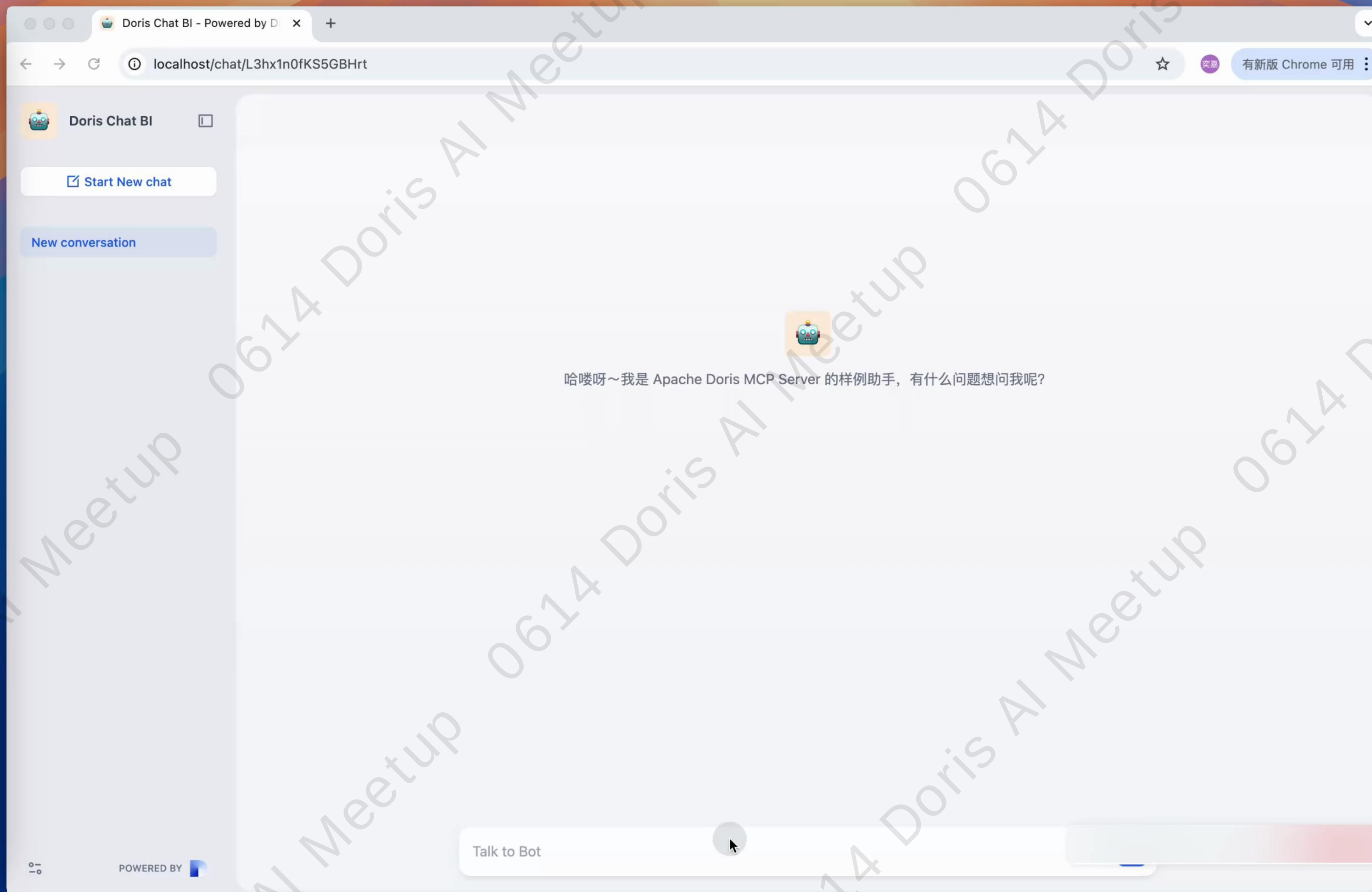
get_db_table_list → get_table_schema → exec_query → get_table_column_comments
元数据发现 → 结构理解 → 查询执行 → 语义增强

应用成果

开发效率：传统 BI 报表开发周期从数周缩短至数分钟，开发人员可专注业务逻辑而非 SQL 调优

用户体验：业务人员可直接通过自然语言获取数据洞察，无需依赖技术人员支持

技术价值：标准化 MCP 接口大幅降低系统集成复杂度，提供 Web 级别的交互体验



Cursor + Doris MCP 智能开发环境

主流 AI IDE + Doris MCP，无需频繁切换工具和查看文档，一站式对答开发体验

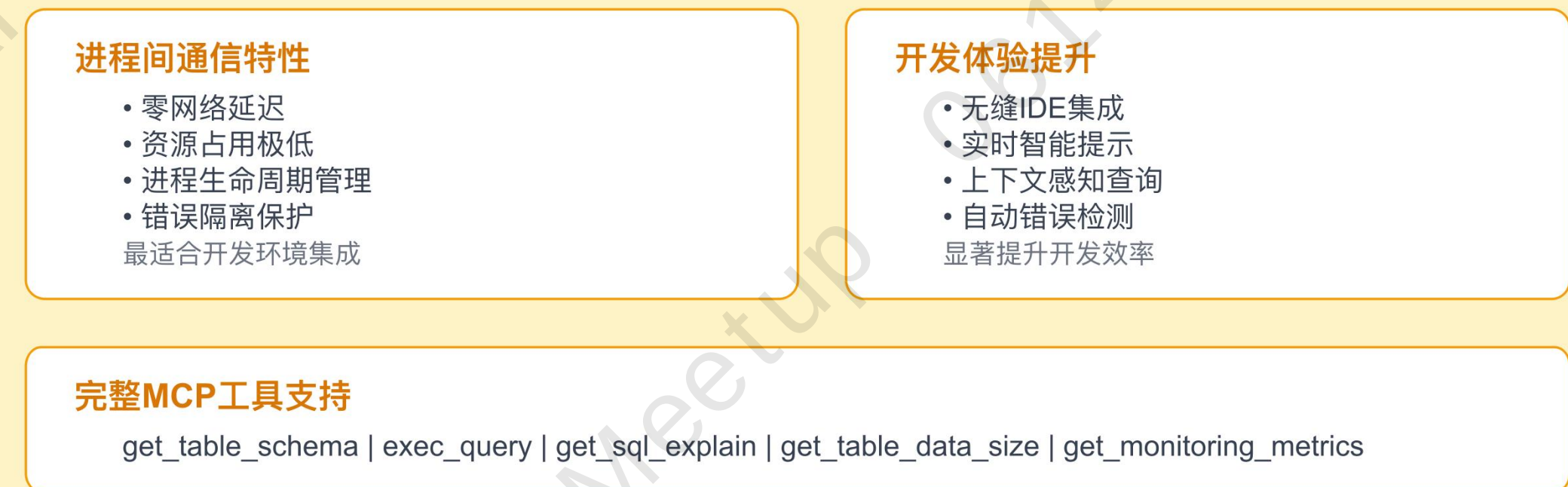
IDE 开发环境：Stdio 通信模式



开发场景应用



Stdio 通信优势

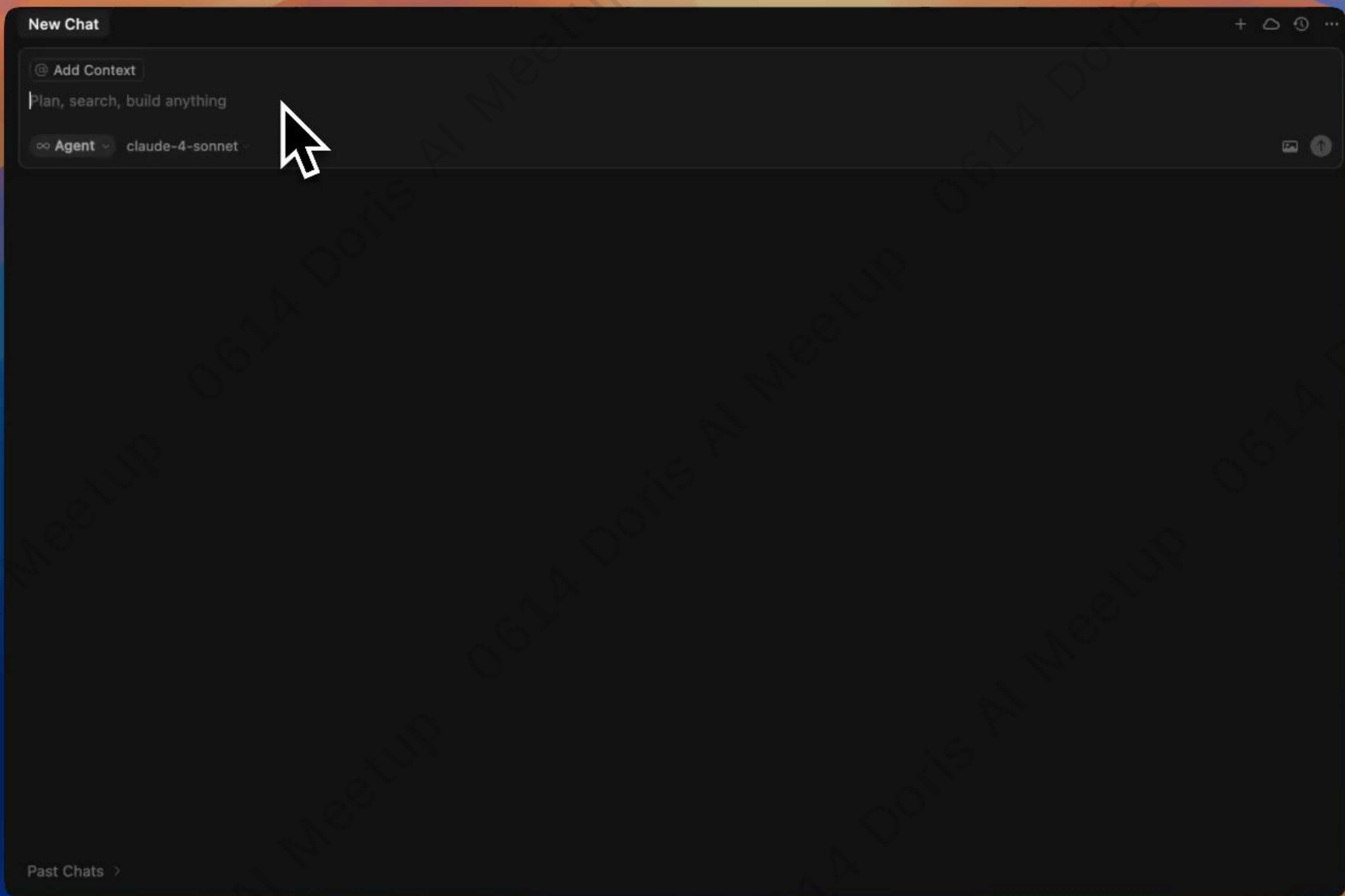


开发效率革命

效率提升: 数据库相关开发效率提升70%+, 无需频繁切换工具和查看文档

错误减少: 实时数据验证和智能提示, 显著降低因数据理解错误导致的Bug

学习成本: 新团队成员快速熟悉数据库结构, 降低项目上手门槛



FastGPT + Doris MCP 构建企业级 RAG 知识库

基于 MCP 协议实现多模态数据统一访问的智能知识检索系统

技术架构设计

用户交互层：自然语言问答 | 多轮对话 | 上下文理解

FastGPT 引擎：语义理解 | RAG 检索 | 答案生成 | workflow 编排

MCP 协议层：统一数据接口 | 标准化访问 | 安全认证

Apache Doris：向量检索 | 联邦查询 | 实时数据 | 多源融合

多元数据源：文档库 | 数据库 | 向量库 | API服务 | 实时流

核心应用场景

1 企业文档智能检索

跨部门文档、技术手册、业务流程的统一智能搜索与问答

2 智能客服知识支撑

实时关联产品信息、用户数据，提供精准的客服支持

3 业务数据智能分析

结合结构化数据与非结构化文档，生成深度业务洞察

4 研发知识管理

技术文档、代码注释、项目经验的智能检索与复用

核心技术价值

多模态数据融合

- 向量数据与结构化数据统一查询
- 语义检索与精确过滤相结合
- 实时数据与历史文档融合分析
- 支持文本、图像、音频多模态

标准化接口优势

- MCP 协议统一数据访问标准
- 无需为每个数据源单独开发
- 支持热插拔式数据源扩展
- 降低系统集成复杂度

智能检索能力

- 语义理解与意图识别
- 上下文感知的多轮对话
- 自动关联相关数据源
- 结果排序与相关性评分

企业级特性

- 细粒度权限控制与安全审计
- 数据访问日志与合规管理
- 高可用架构与容灾备份
- 水平扩展与负载均衡

开发效率提升

- 快速集成，减少开发周期
- 可视化 workflow 编排
- 丰富的预设模板与插件
- 灵活的自定义配置

成本与运维优势

- 统一平台降低运维成本
- 减少重复数据存储
- 自动化运维与监控
- 开源生态，避免厂商锁定

通过 FastGPT + Doris MCP 构建的企业级 RAG 知识库，实现了多模态数据的统一访问与智能检索

Coze + Doris MCP 实现 Doris 智能远端运维

基于字节跳动扣子平台构建的数据库运维助手，可通过各类 IM 工具及机器人实现智能运维

智能运维架构

运维人员：自然语言指令 | 钉钉/飞书集成 | 移动端支持

Coze 智能引擎：工作流编排 | 多轮对话 | 意图识别 | 知识库管理

插件生态：运维脚本 | 告警处理 | 自动化工具 | 监控集成

MCP 协议层：数据接口标准化 | 安全认证 | 实时通信

Apache Doris：运维数据存储 | 性能监控 | 日志分析 | 元数据管理

基础设施：Doris 集群 | 监控系统 | 日志收集 | 告警通知

核心功能模块

性能监控

- 集群状态实时监控
- 查询性能趋势分析
- 资源使用率告警

故障诊断

- 自动错误日志分析
- 故障根因追踪
- 智能问题定位

自动优化

- 索引建议生成
- 分区策略优化
- SQL 查询调优

容量规划

- 数据增长趋势预测
- 存储容量规划
- 扩容建议生成

运维操作

- 一键重启服务
- 配置参数调整
- 备份恢复管理

报告生成

- 运维日报/周报
- 性能分析报告
- 健康检查报告

应用场景与技术特性

自然语言运维

- "帮我检查昨天的慢查询情况"
- "分析一下今天的负载峰值原因"
- "为订单表创建合适的分区策略"
- "生成本月的集群健康报告"

工作流自动化

- Coze 工作流编排复杂运维任务
- 故障自动恢复流程设计
- 定时巡检任务自动执行
- 异常告警联动处理机制

企业集成优势

- 无缝接入企业 IM (钉钉/飞书/企微)
- 与现有运维工具生态整合
- 权限与企业安全策略统一
- 多团队协作与知识共享

MCP 协议优势

- 标准化数据接口，降低集成复杂度
- 支持实时数据流式传输
- 安全的身份认证与权限控制
- 可扩展的插件化架构设计

智能化能力

- 基于历史数据的预测性维护
- 异常模式自动识别与学习
- 运维知识库自动构建
- 上下文感知的智能决策

效率提升价值

- 降低运维门槛，减少专业技能依赖
- 7x24小时自动化监控与响应
- 运维经验沉淀与团队协作优化
- 从被动响应转向主动预防

通过 Coze + Doris MCP 构建的智能运维系统，实现了自然语言驱动的数据库运维与管理

构建Agent时代的智能数据未来

统一标准 · 智能高效 · 安全可靠 · 面向未来

Apache Doris + Doris MCP Server

Agent时代数据平台的最佳选择

开启AI应用开发新纪元，让数据智能触手可及，让创新不再受限于技术壁垒

1 技术先进性

- MPP架构，毫秒级查询响应
- 向量检索，智能数据发现
- Multi-Catalog，湖仓一体
- 实时计算，秒级数据更新
- 弹性扩展，按需伸缩

2 标准化优势

- MCP协议统一接入标准
- 降低集成成本80%以上
- 跨平台兼容，生态丰富
- 开发效率显著提升
- 维护成本大幅降低

3 商业价值

- 提升决策效率和准确性
- 大幅降低数据平台成本
- 加速业务数字化转型
- 增强企业竞争优势
- 创造可持续增长价值

拥抱Agent时代，引领数据智能

立即体验Apache Doris MCP解决方案

开启您的智能数据平台之旅

Thanks !

