

阿里云 SelectDB Serverless 驱动的 AI 可观测方案

Agent 时代的黑盒解药

从 Trace 追踪到 Cost 洞察，为 AI Agent 应用构建
端到端的可观测能力

2026/06

阿里云 SelectDB · AI 可观测



目录 / Agenda

01

AI 可观测：从锦上添花到生死线

痛点共鸣与行业共识

02

SelectDB 是谁 + 解药 Litefuse

全托管 Langfuse 兼容方案

03

解药 Serverless 弹性

写入尖刺 / 长尾查询 / 存储膨胀

04

Demo + 客户故事

眼见为实 + 行动召唤

AI 可观测

| Agent 时代：从锦上添花到生死线

LLMOps 三件套

| Agent 应用上线前必须解决的门槛性能力



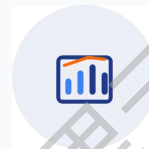
Trace 追踪

完整 Agent 调用链可视化，每次 LLM/工具/向量检索都被串成树



Cost 成本

Token 按用户、模型、业务标签实时聚合，直接影响 P&L



Eval 评估

模型版本升级、Prompt 改造，效果涨跌用数据说话

阿里云SelectDB 是谁

阿里云一方产品：为分析而生的实时湖仓，服务头部 AI 客户

解药① 托管 Litefuse

开箱即用的 AI 可观测方案

DataLens for AI 是什么

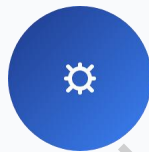
产品定位

| 基于 Litefuse 的全托管 AI 可观测平台



AI 可观测平台

专为 Agent / LLM 应用打造的全托管可观测服务，覆盖 Trace, Eval, Prompt 全链路



SelectDB 驱动

基于实时湖仓引擎，兼具高吞吐写入、毫秒级查询与向量检索能力

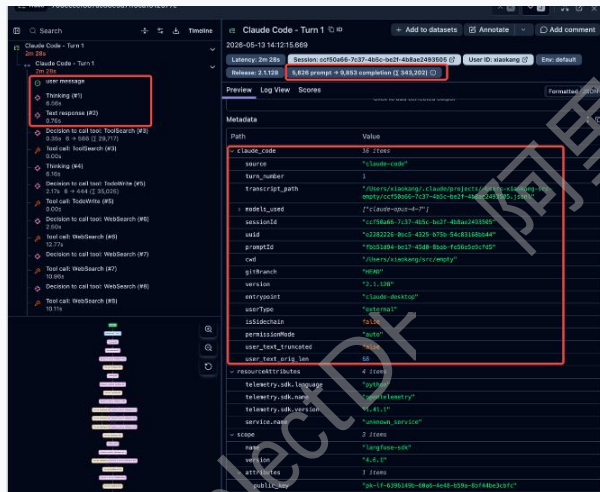


100% 兼容 Langfuse SDK

改一行 endpoint 配置即可从自建方案无缝迁移，零代码改造

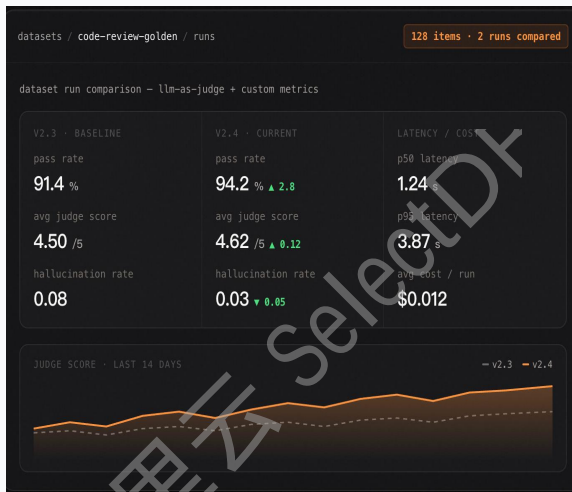
Litefuse 三大核心功能

| Trace · Eval · Prompt Management



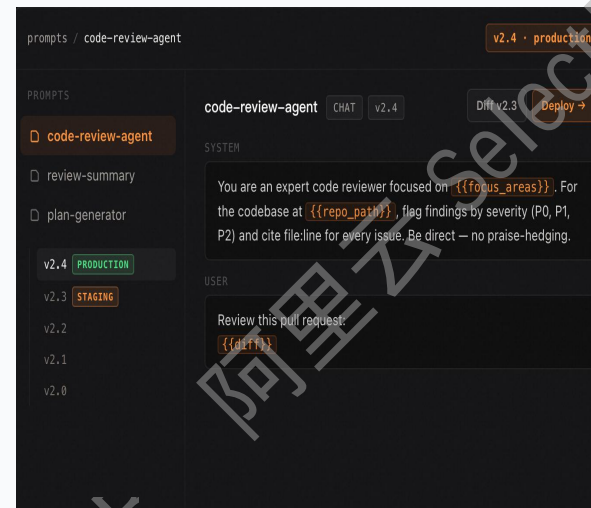
Tracing 全链路追踪

Agent 每一步可视化，嵌套 Trace 树展示 LLM 调用、工具执行、向量检索全过程



Evaluation 效果评估

LLM-as-Judge + 人工标注 + 数据集管理，Prompt 迭代效果量化可衡量



Prompt 版本管理

UI / SDK / API 三种方式管理 Prompt，版本追踪与生产链路自动关联

为什么选 DataLens for AI

vs 自建 Langfuse



-50%

运维复杂度降低

自建langfuse需 6 套组件（Web + Worker+ PG + Redis + Doris数仓+存储等），托管仅需 3 服务，免运维



Agent 领先

Agent 可观测能力领先

除 19+ 主流框架外，额外支持 Hermes / OpenClaw / ClaudeCode 等新一代 Agent 框架完整可观测，Langfuse 尚未覆盖



-80%

存储成本降低

存储架构优化 + 多轮对话数据压缩，热数据 SSD、冷数据沉入OSS，冷热分层，大幅优化长期存储成本

解药② 阿里云 SelectDB Serverless

弹性是底气：三大挑战与三件武器

AI 可观测的底层挑战

对数仓极限压力的三个维度



写入尖刺

用户一句话触发几十次调用，
峰值 10 倍尖刺，传统数据库按
峰值买浪费 90%



存储膨胀

99% 数据 7 天后变冷但不能删，
长尾保留 90 天+ 是最大成本黑
洞



多模查询

明细/聚合/全文/向量四种查询
模式叠加同一份数据，传统方
案需三套引擎

阿里云 SelectB Serverless 三件武器

三个挑战，三件武器，一一对应



秒级弹性

→ 写入尖刺

写入压力来秒级扩，回落秒级缩，无需容量规划，只为实际用量付费



冷热分层

→ 存储膨胀

热数据存 SSD，冷数据自动沉 OSS，存储成本更低



统一引擎

→ 多模查询

OLAP + 倒排索引 + 向量检索
一个引擎搞定，无需维护三套系统

阿里云 SelectB Serverless 版：存算分离+自动弹性

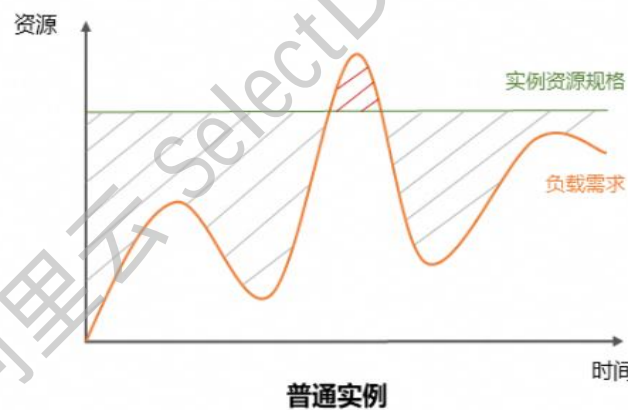
计算秒级弹性伸缩
多计算组共享数据、多读多写
读写分离、在离线业务隔离



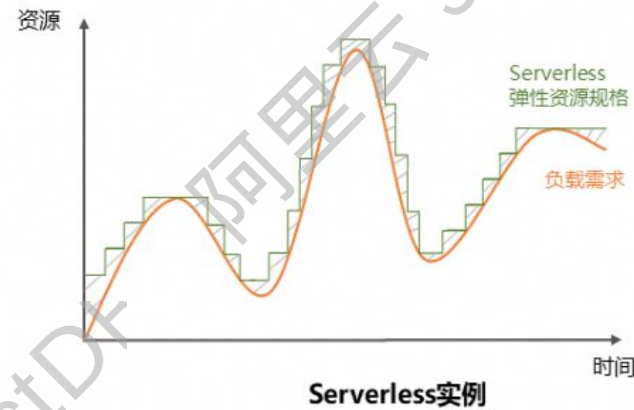
缓存秒级弹性伸缩
缓存默认与定制策略
缓存平滑处理，支持预热与迁移



存储实时弹性伸缩
存储空间隔离
存储访问优化与限速

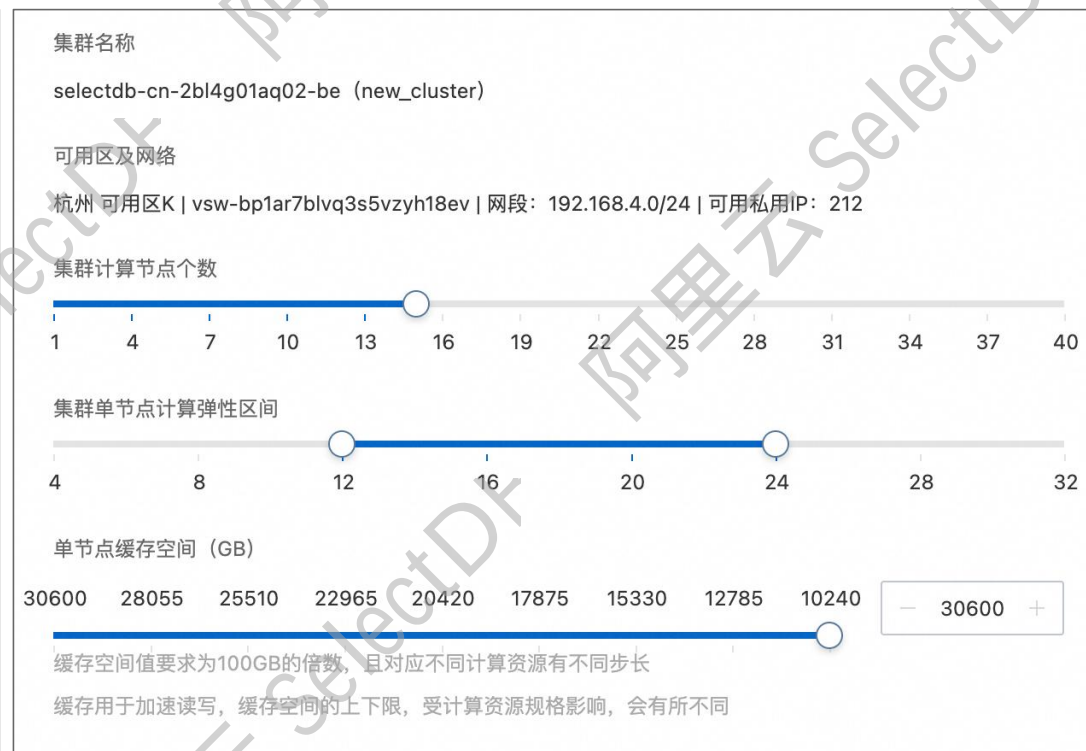
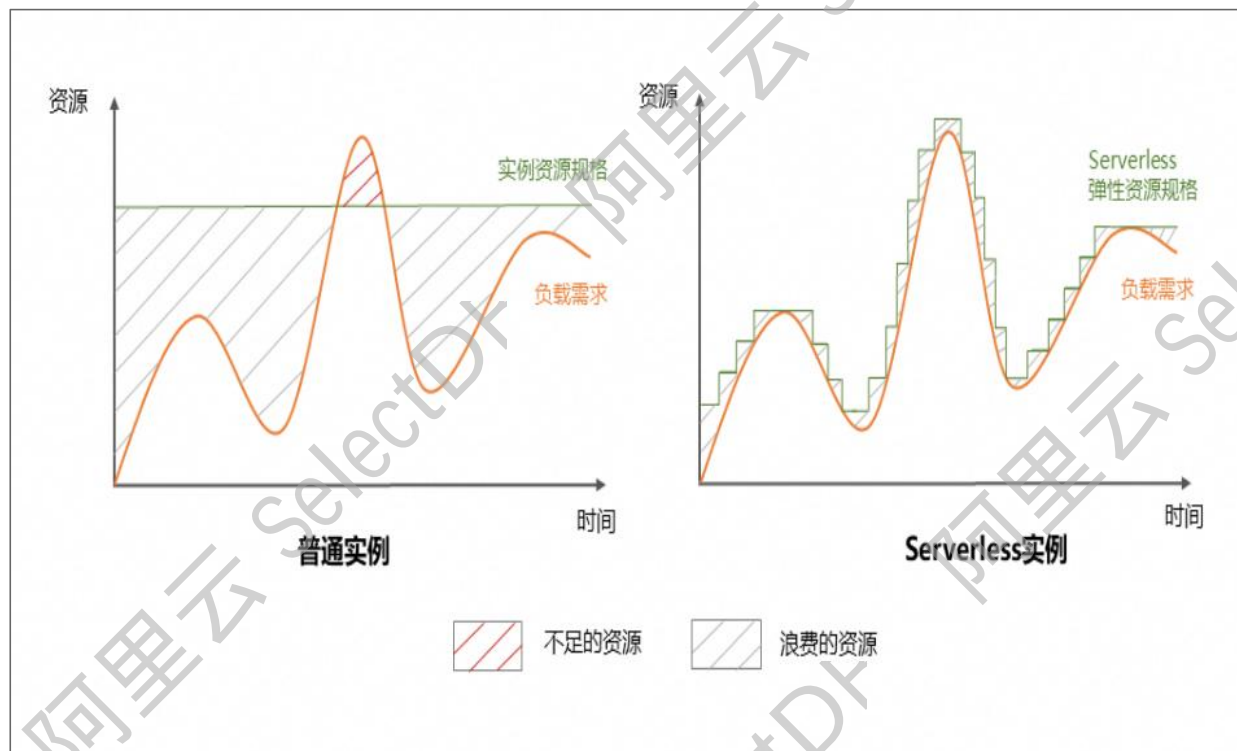


不足的资源



浪费的资源

阿里云 SelectB Serverless 版：自动弹性+成本可控



客户故事与行动召唤

真实数据验证产品价值

10亿+

日均 Trace

30%

成本降低

5x

性能提升

"某头部大模型公司：日均 trace 写入超 10 亿条，
存储成本降低 30%，长文本检索性能提升 5 倍"



Demo Time

I 5 分钟快速体验:基于阿里云SelectDB Serverless 的LLM/Agent 可观测方案完整链路



谢谢

扫码开通阿里云 SelectDB (Litefuse 免费公测)



扫码体验