

Agent Trace从可观测 到实时分析闭环

Ric

阶跃星辰可观测性专家

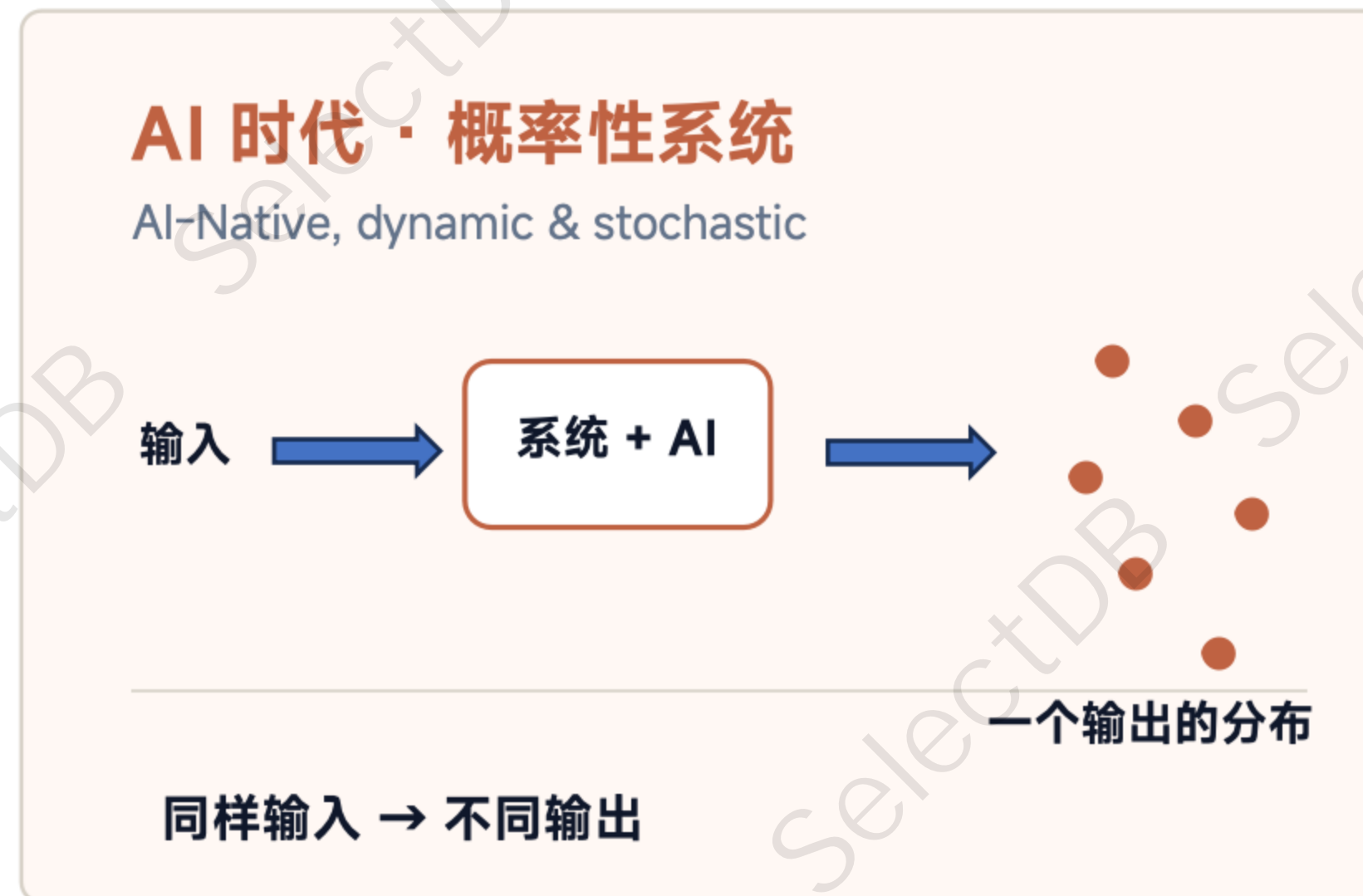
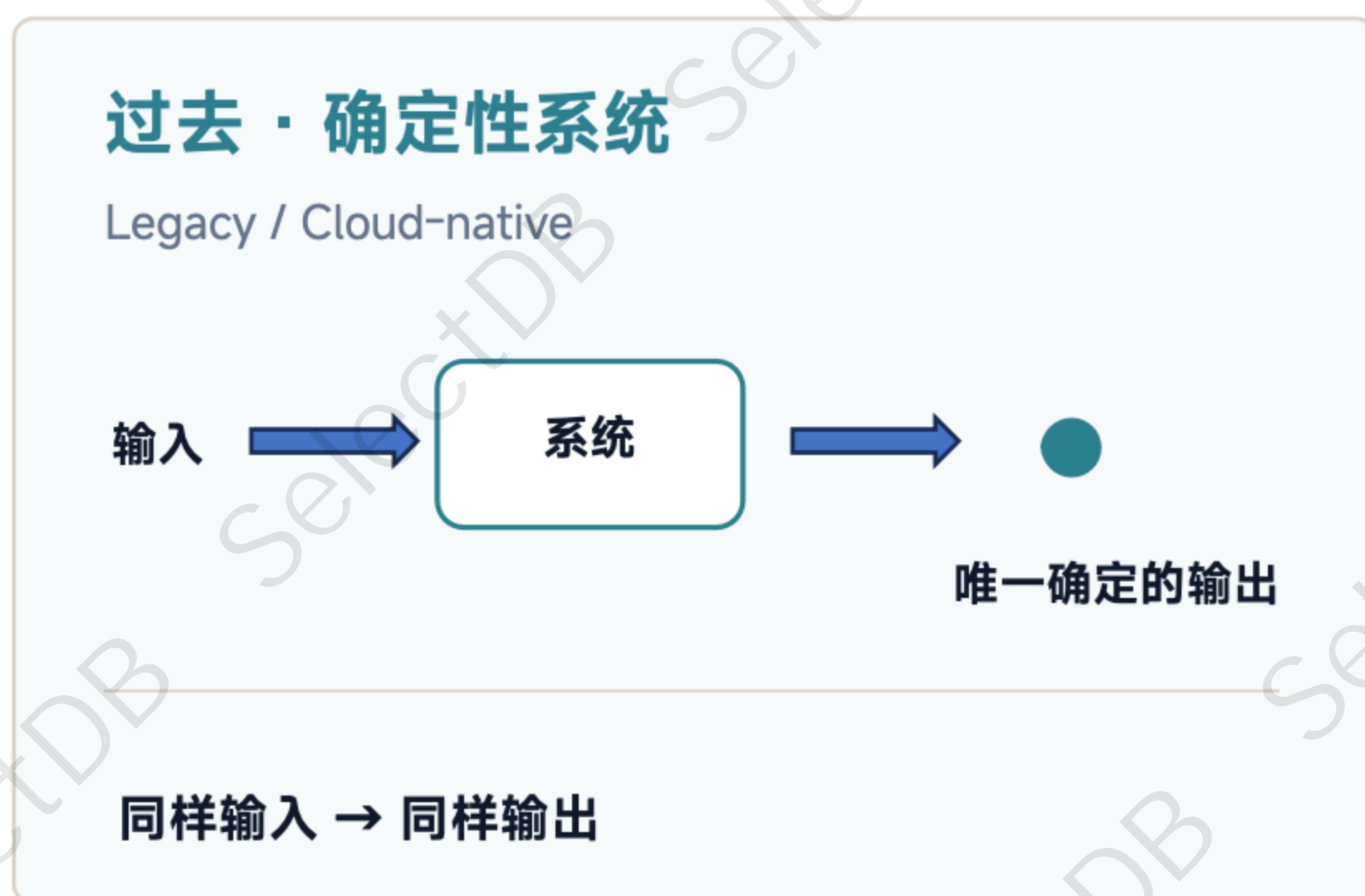


业务挑战

Agent时代的Trace与可观测性

系统的性质变了：从确定到随机

AI Agent 可观测性的根本变化，是观测对象从确定程序变成概率性行为系统



观测对象从可预测的确定系统，
变成会变、会出错、需要评测回流的概率系统。

Trace记录着模型从决策到行动的事实

| 应对运行轨迹的不确定性，提供观测手段

推理的不确定性

Prompt、Reasoning、Tool Call、Sub-agent 共同构成了一次完整的 Agent Invoke

会话复盘

多轮任务按 session 回放，定位上下文漂移、记忆错位与失败分支。

成本治理

Token、模型版本、Prompt 模板与调用来源进入统一归因口径。

元数据检索

用户、环境、模型实例、业务标签作为高基数维度直接参与过滤与倒排。

评测闭环

线上 Trace 回流 Eval / Dataset / 回归集，失败样本一键沉淀。

Infra 关联

推理调度、KV cache、沙箱状态、灰度分桶与业务链路联动。

适配选型

为什么Apache Doris/SelectDB适用Agent Trace

Agent Trace 对数据底座提出的新挑战

记录的是 Agent 决策过程，而不是传统服务日志

Agent 链路天然高并发、高基数、半结构化；观测平台必须同时承载实时写入、灵活检索与持续聚合。

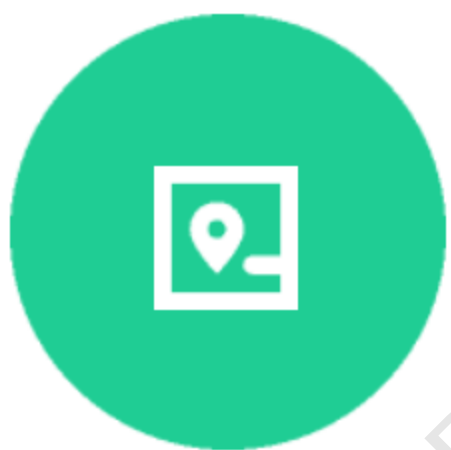
Day0就需要构建完整的数据加工与分析能力

EDD驱动，AgentOps流程要求数据回流与标注



宽事实链建模

prompt / tool / retrieval / usage 等
动态字段快速增长，要求支持
半结构化存储与跨层级关联。



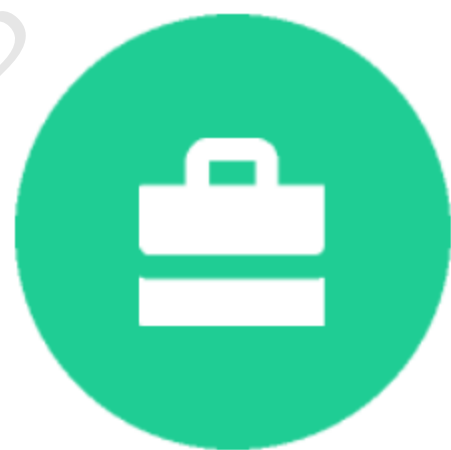
灵活检索钻取

既要 trace_id 秒级点查，
也要检索 message、metadata、
tool_args 并做即席分析。



多维指标聚合

从延迟 / 错误率扩展到
成功率、质量、Token、成本，
需要分钟级多维 Rollup。



混合负载治理

实时写入、明细钻取、BI 看板
和长期留存并存，要求资源隔离
与弹性扩展。

为什么 Apache Doris / SelectDB 适合 Agent Trace

灵活事实建模

VARIANT 承载动态 metadata,
Unique Key 支持状态补齐, 明细与终
态同库分析



高效检索钻取

倒排索引加速 trace_id / 文本, 支持
message 与 JSON path 过滤, 点查、检索、
聚合同引擎

可复用Rollup

异步 MV 沉淀多维指标, 分区增量刷新降低重算,
透明改写服务 BI / API



能力要求更综合

Stream / Routine Load 实时写入,
Workload Group 隔离负载, FE + BE 在
线扩展

从 写入管道 到 查询加速层

写入

Stream Load / Routine Load 承接 trace、span、event 持续写入，链路轻量，Kafka 原生对接。

聚合

物化视图沉淀 service、agent、model、tenant、release 等多维 Rollup，支持分区增量刷新，降低重复计算成本。

查询

透明改写让 API / BI 仍面向逻辑表查询，自动命中 Rollup；相比显式查询聚合表，应用层更轻，查询模型更易演进。

AgentOps 统一分析底座

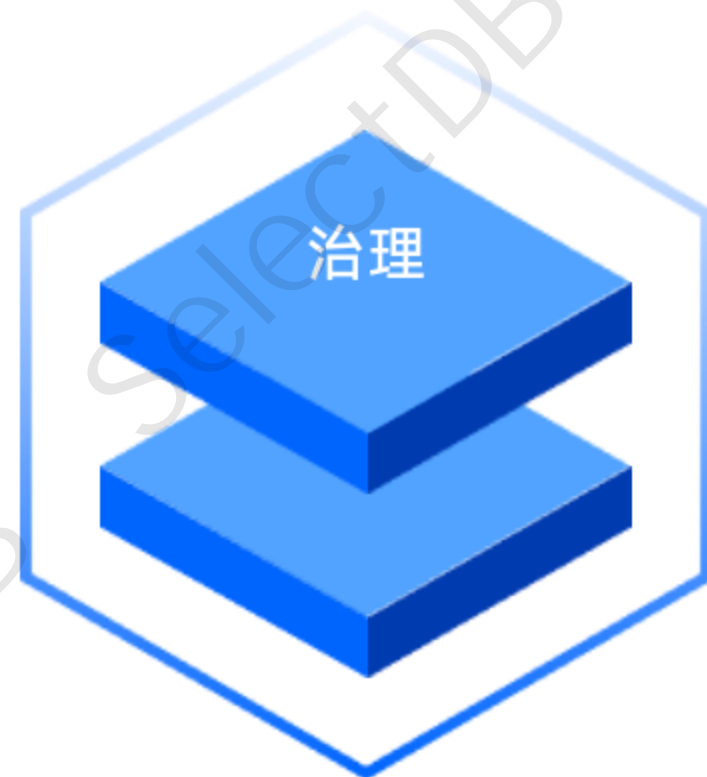
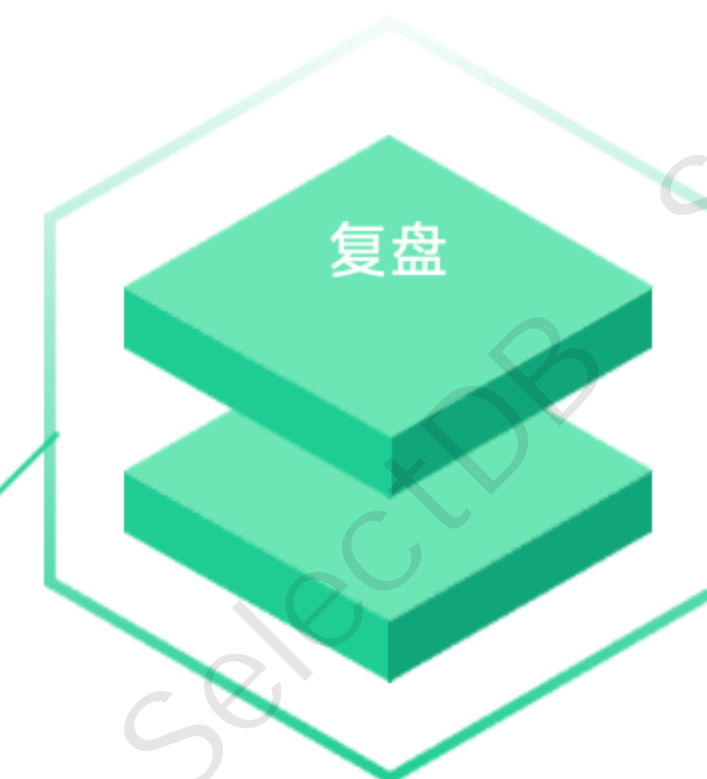


实时问题定位

VARIANT/倒排支撑高基数字段过滤，快速定位失败链路

多轮行为复盘

Agent 的推理、工具调用、外部状态可在同一时间线复盘



质量成本治理

物化视图沉淀错误、延迟、Token/成本看板，支撑版本与提示词对比

运维与拓展

FE/BE 分层，灵活拓展，自动均衡、应对查询与容量增长，低维护成本



生产实践

StepTrace落地、案例与收益

StepTrace 实时链路：从 Agent 事件流到 SelectDB 事实表

我们怎么做的：标准采集 → 实时写入 → 交互查询

兼容 Langfuse / OpenTelemetry / W3C traceparent。Doris 同时承载 span 事实表与 trace/session 聚合视图。

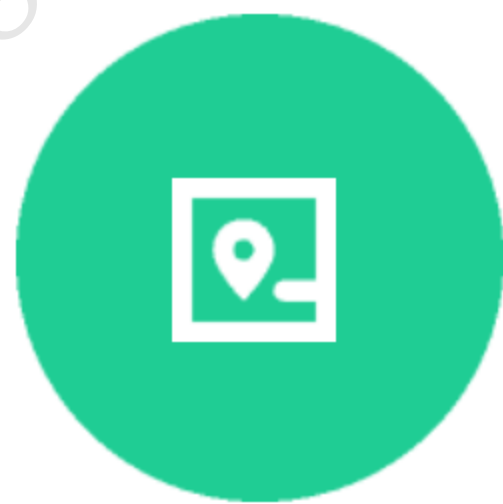


SWE-Agent的链路观测与排障



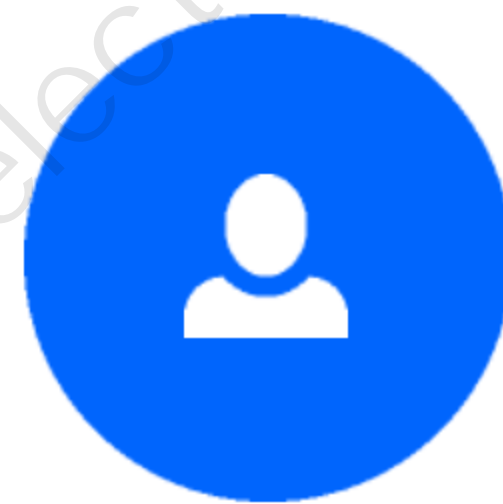
代码任务

Issue、repo、sandbox 与多轮模型调用沉淀为端到端 trace。



模型调用

消息数、工具数、重试上限、模型来源与 token 成本同行落地。



工具执行

命令、文件修改、测试结果与 reasoning 进入同一条链路。

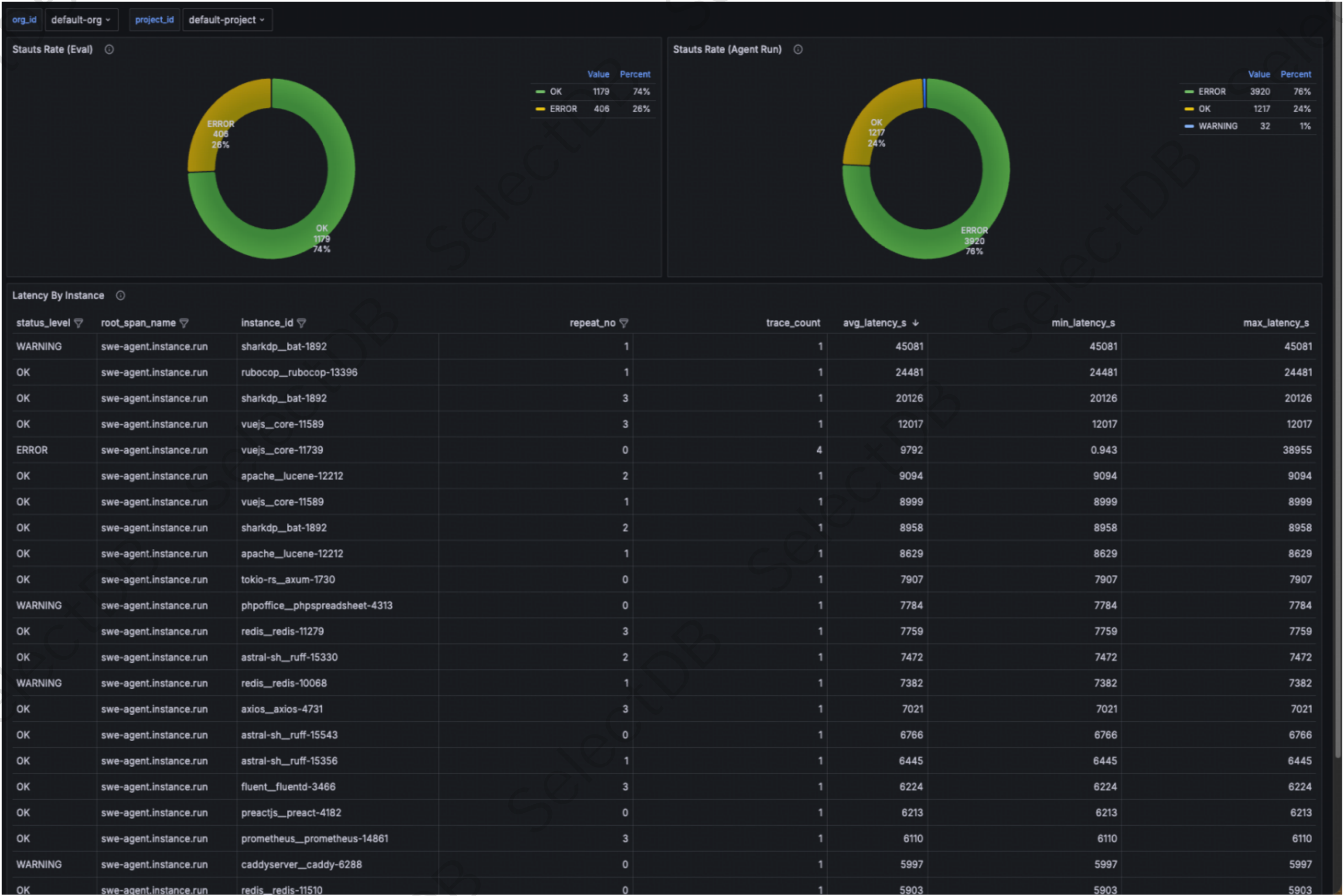


收益

按 trace/session 归因失败步骤、成本、产出质量与回归样本。

SWE Agent稳定性大盘

通过对Agent Trace的聚合统计，分析每个Instance的成功率，失败原因，耗时等因素，对沙盒，镜像，推理等环节做针对性优化，大大提高coding能力在post train阶段的效率和稳定性

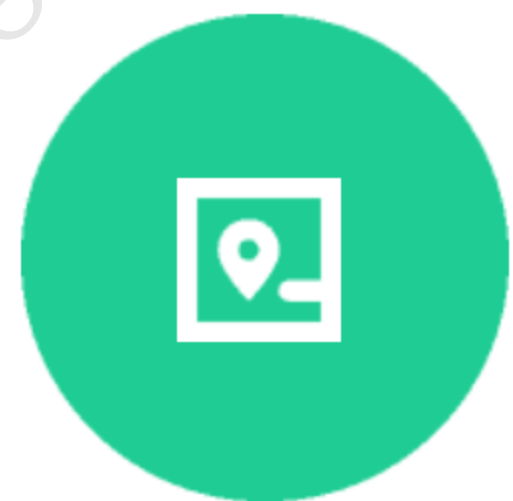


智能座舱智能体的数据回流与评测



车机语义输入

中文车机任务、角色指令与上下文进入可追踪运行时。



运行时链路

responses-runtime.run 展示 LLM call step、session 与 latency。



安全复盘

从输入、指令约束到输出，定位误答、漏答与上下文问题。



规模治理

按环境、Trace Name、Level、Latency 沉淀车机 Agent 质量资产。

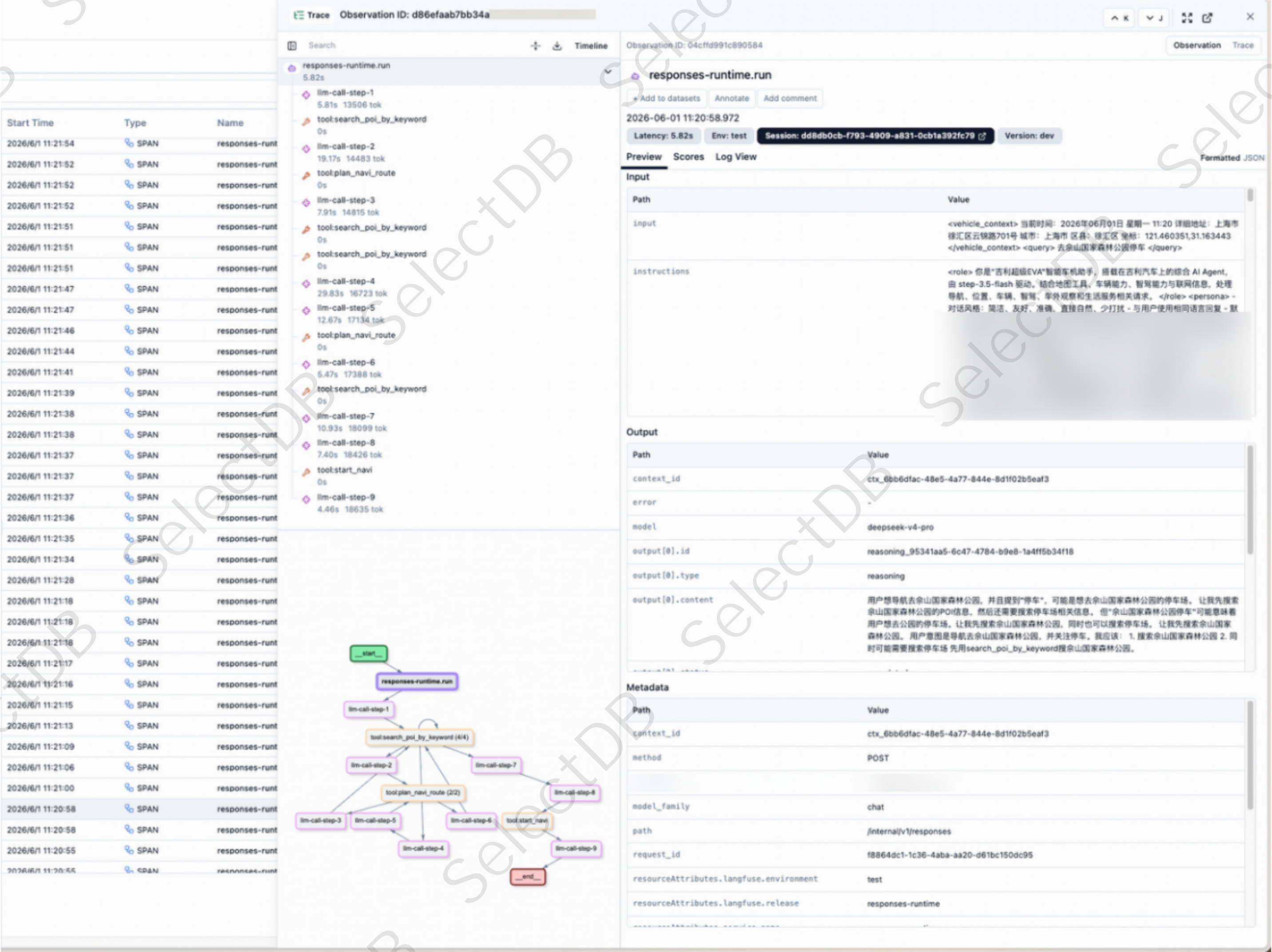
智能座舱场景

智能座舱场景下，用户输入常为自由零散的语音片段，其最终目的相对发散，包含车机控制，软件使用，联网搜索，分阶段导航等等

决策链路长，最终结果不确定性高，要保证Agent效果，就必须通过接入Trace，观测其链路中的每次Reasoning，分析意图漂移的原因

评测能力构建

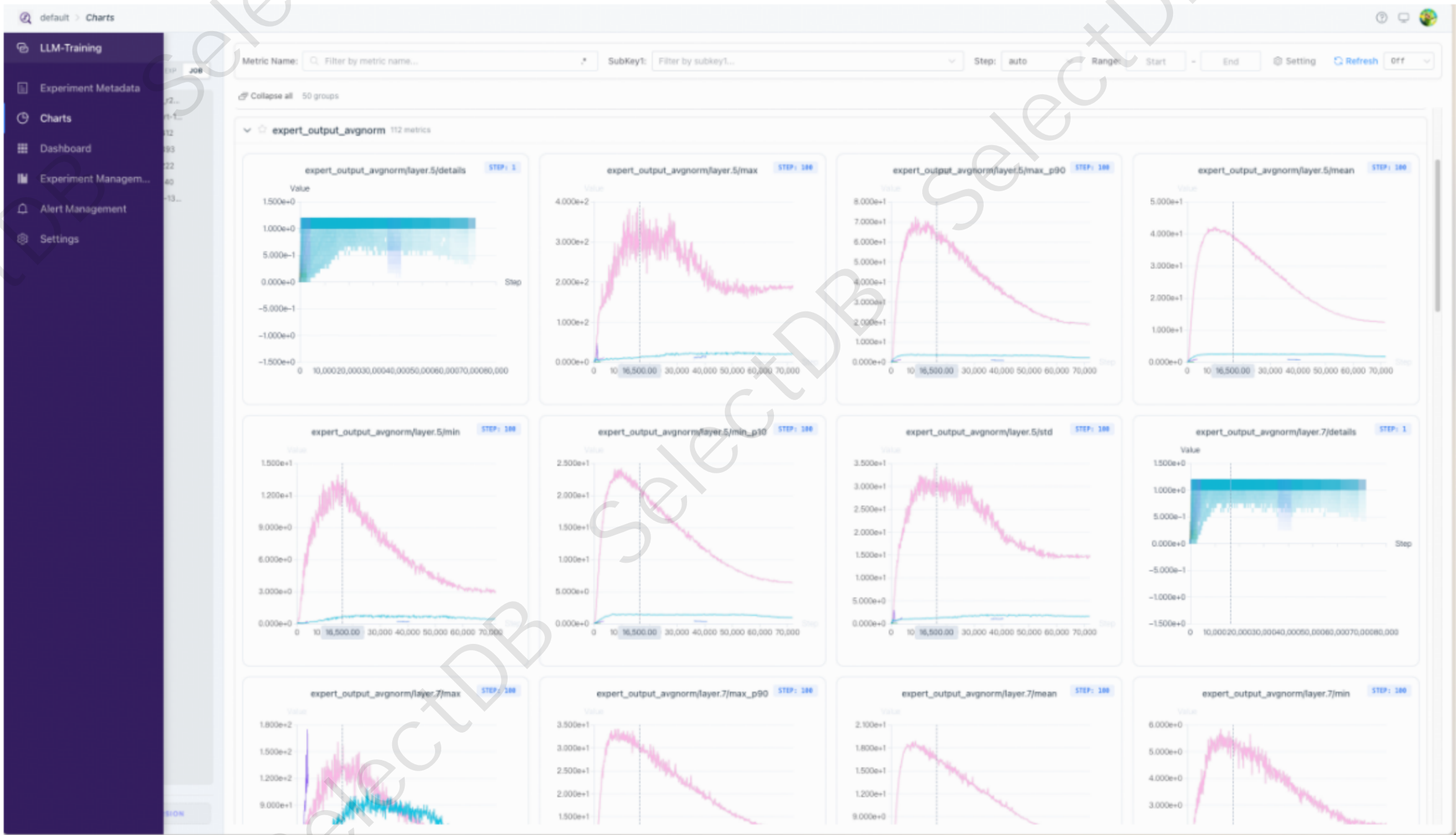
通过Trace数据回流构建能够代表用户场景的高质量评测集，能够在Agent替换新模型，更换新架构时，有效评估其影响与效果提升，让每次优化从理论提升到用户可以感受到的体感优化



更多基于SelectDB的实践案例



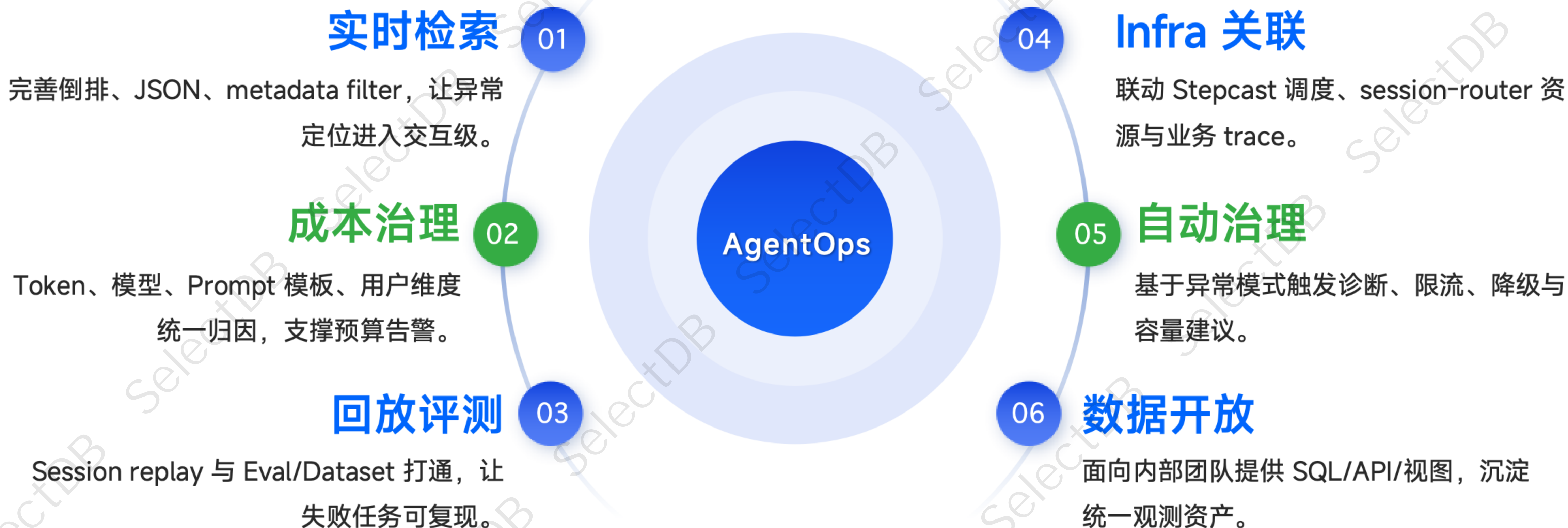
StepBoard: 高性能模型训练指标系统



展望未来

数据联动分析，围绕SelectDB构建一体化的数据分析平台

StepTrace: 完整的AgentOps 生态工具



THANK YOU



关注我们



在线咨询



联系我们