

Litefuse - Agent

可观测与效果评估平台

肖康

Apache Doris PMC Member & 飞轮科技副总裁



目录

01

Agent 可靠性挑战

02

EDD 评估驱动开发

03

Litefuse 解决方案

Agent 可靠性挑战

软件与 Agent 的双重不确定性

为什么 Agent 没有报错、响应速度快， 却没有给出满意的结果？

软件可靠性通用挑战

逻辑正确性

业务高峰压力

运行健壮性

基础设施稳定性

Agent 不确定性挑战

大模型幻觉

路径规划错误

工具调用失败

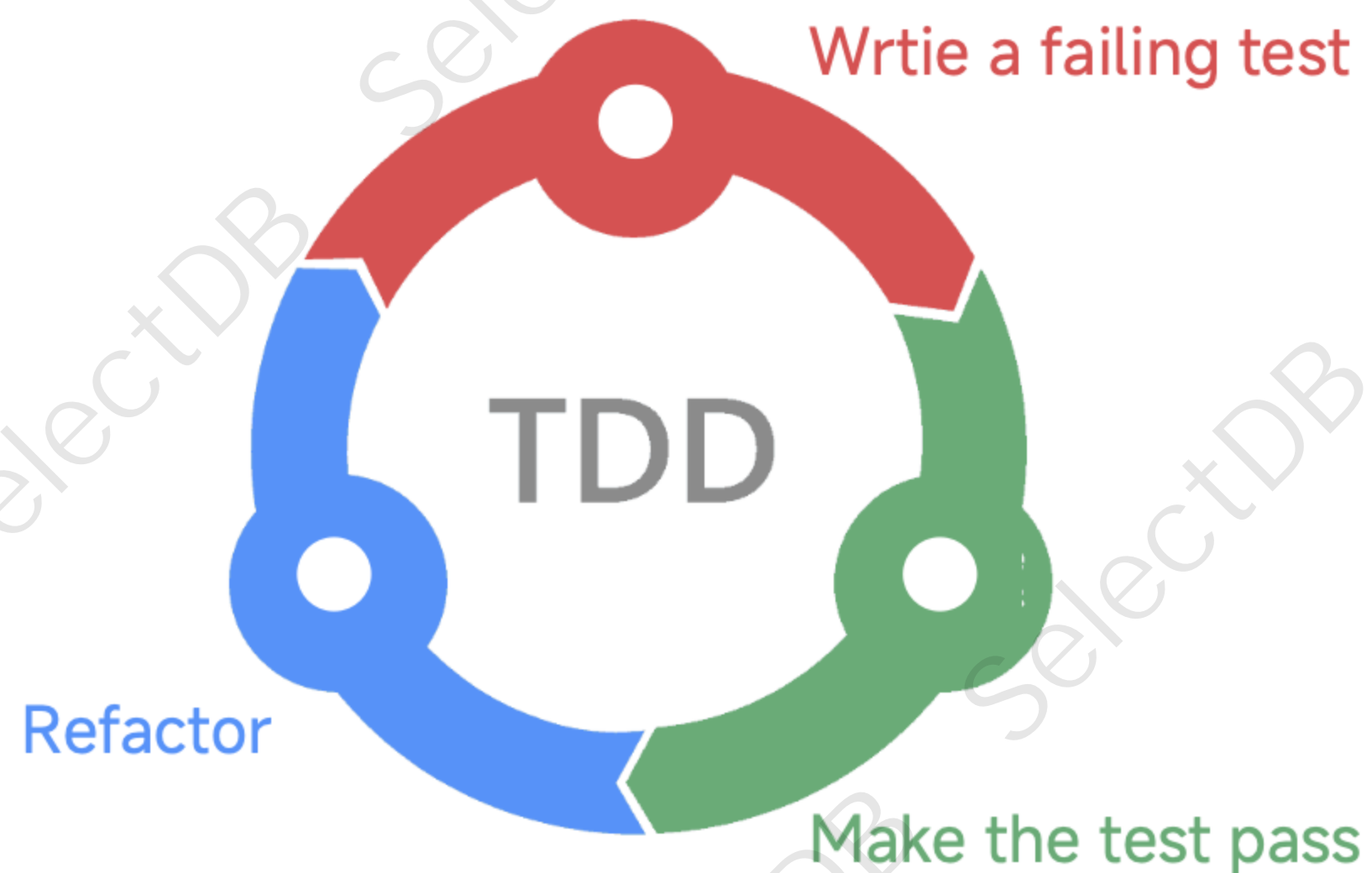
记忆腐化

Litefuse 解决方案

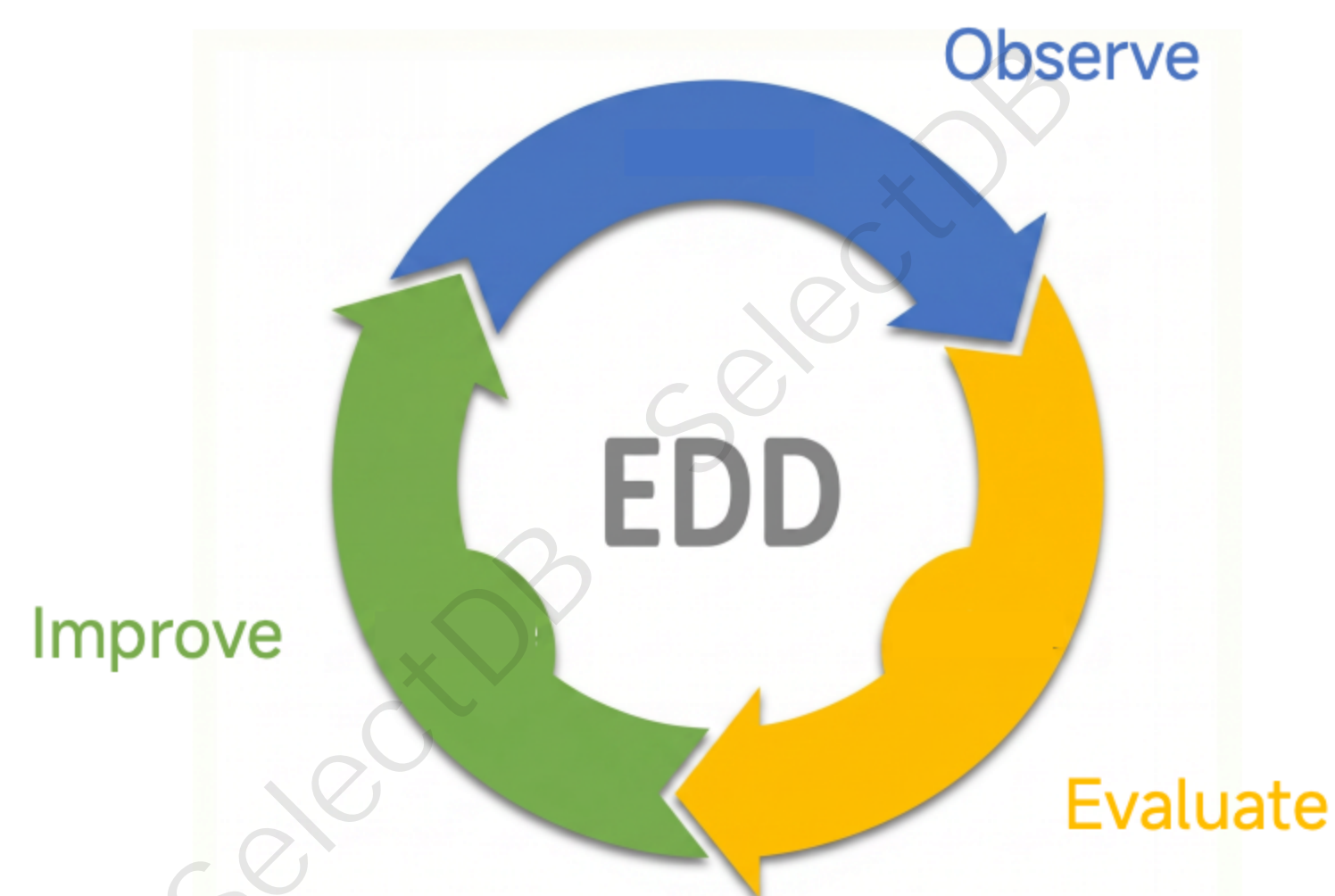
| Agent 可观测与效果评估平台

TDD 解决通用挑战， Agent 特有挑战需要新的方法论 - EDD

Test-Driven Development



Evaluation-Driven Development



Litefuse - 将 EDD 产品化的 Agent 可观测与效果评估平台



Litefuse - Observability 解开 Agent 执行黑盒

Litefuse

Go to... %K

Home

Dashboards

Observability

Tracing

Sessions

Users

Prompt Management

Prompts

Playground

Evaluation

Scores

LLM-as-a-Judge

Human Annotation

Datasets

Settings

Kang

My Agents

Tracing

Traces Observations

Hide filters Views 0 Search...

Filters

Environment

Mode:

SELECT TEXT

sdks-experiment

default

langfuse-llm-as-a...

Trace Name

Mode:

SELECT TEXT

Filter values

experiment-it... 778

Execute evalu... 737

Execute evalu... 737

Execute evalu... 737

Hermes Agen... 58

Claude Code ... 41

Hermes Agen... 34

Hermes Agen... 32

Claude Code ... 30

Trace 786eeef887ac88e5d7ff9ca1912677e

Search

Timeline

Claude Code - Turn 1 2m 28s

Claude Code - Turn 1 2m 28s

user message

Thinking (#1) 6.56s

Text response (#2) 0.76s

Decision to call tool: ToolSearch (#3) 0.35s 6 → 566 (Σ 29,717)

Tool call: ToolSearch (#3) 0.00s

Thinking (#4) 6.16s

Decision to call tool: TodoWrite (#5) 2.17s 6 → 444 (Σ 35,025)

Tool call: TodoWrite (#5) 0.00s

Decision to call tool: WebSearch (#6) 2.50s

Tool call: WebSearch (#6) 12.77s

Claude Code - Turn 1 ID

+ Add to datasets Annotate Add comment

2026-05-13 14:12:15.669

Latency: 2m 28s

Session: ccf50a66-7c37-4b5c-be2f-4b8ae2493505

User ID: xiaokang Env: default Release: 2.1.128

5,626 prompt → 9,853 completion (Σ 343,202)

Preview Log View Scores Formatted JSON

Tags

claude-code lifefuse model:claude-opus-...

Input

Path Value

content "research and write a report about agent observability and evaluation"

role "user"

Output

Path Value

content "Report written to agent-observability-report.md. Eight sections covering:

1. **Executive summary** - the tracing + eval + feedback-loop stack

2. **Why agents break traditional APM** - non-determinism, silent failures, causal amnesia, cost as a signal

3. **Observability primitives** - traces/spans, sessions, metrics, evals-

Litefuse - Evaluation 评估 Agent 效果



Litefuse 基于 Langfuse 与 SelectDB 构建

Langfuse

功能丰富 · 优势

- ✓ 100+ AI 生态集成 (OpenAI / LangChain / Dify...)
- ✓ AI Native 数据模型 (LLM · Tool · Token)
- ✓ Prompt 管理 + Evaluation

规模化痛点

存储成本高 (TB 级, 占 Agent 成本大头)

架构复杂 (6 组件, 部署运维重)

文本检索慢 (LIKE 全量扫描)

SelectDB · Apache Doris

SelectDB 存储优势

- ◆ VARIANT · 列存高压压缩比, 上万动态字段
- ◆ 存算分离 · 数据只写一次、存对象存储
- ◆ 倒排索引 MiniMax、阶跃、字节等 PB 级生产验证

正好解决左侧痛点

存储成本 ↓ 88%

架构 6 进程 → 1

检索快 5-10x

Litefuse 产品优势

01

单机版极简轻量

368MB · 无
Docker · 20 秒启动

02

成本大幅降低

存储成本实测
最高降 88%

03

Agent 生态丰富

新增
Hemes · OpenClaw 等

Litefuse 产品优势 - 单机极简轻量

安装包 368MB

无需外部依赖

无需 Docker 环境

单进程运行



Litefuse · 20 秒装完

(此处放置安装录屏)

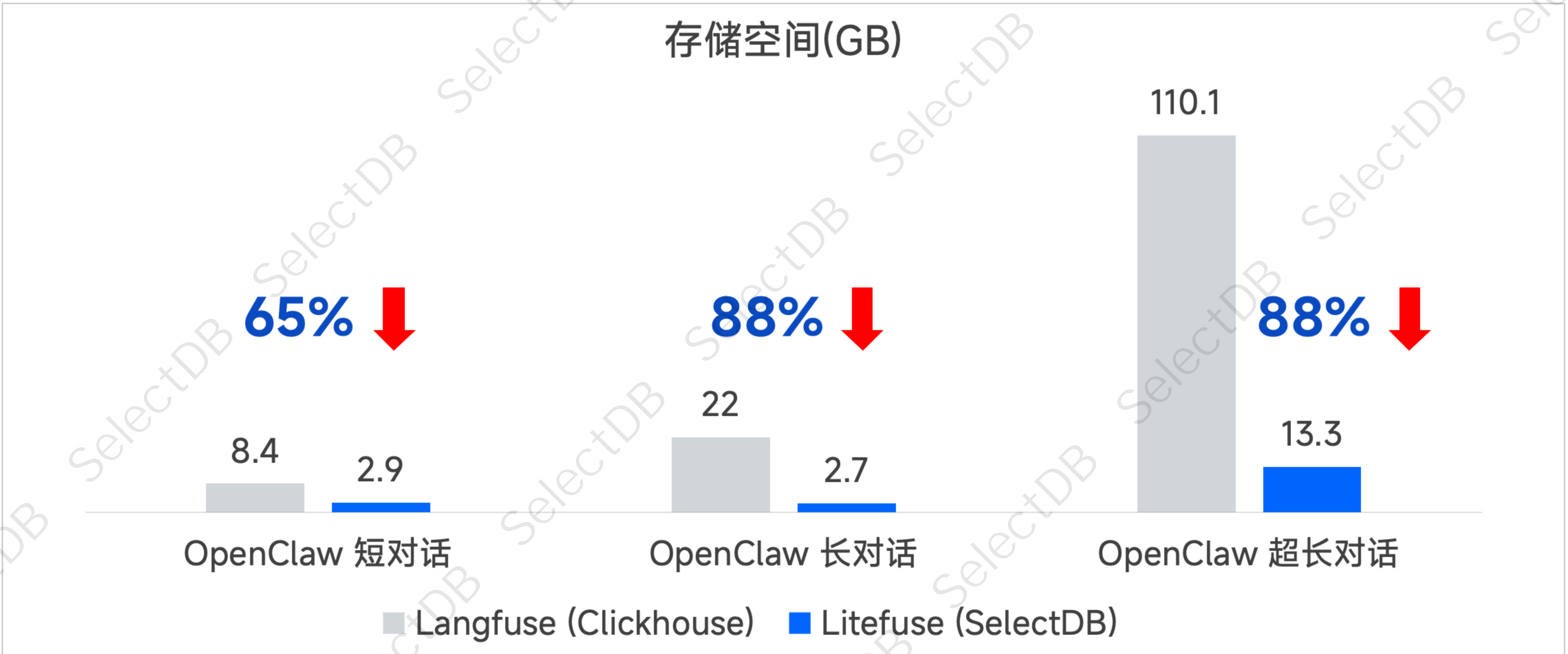
VS



Langfuse · 几分钟

(此处放置安装录屏)

Litefuse 产品优势 - 成本节省 实测最高 88%



Litefuse 产品优势 - Agent 生态丰富

继承自 Langfuse

Claude Code

Codex

Cursor

Litefuse 新增覆盖

OpenClaw

Hermes

Pi

MiniMax Coder

Kimi Code

StepClaw

Litefuse 产品形态

Open Source

- ✓ 单机版 · 极致简洁轻量
- ✓ 分布式版 · 成本节省 50%

开源免费

Enterprise

- + 含 OpenSource 全部能力
- ✓ 深度优化 · 成本节省 80%
- ✓ 企业级特性 Coming Soon

企业自建

Cloud

- + 含 Enterprise 全部能力
- ✓ 免维护 开箱即用
- ✓ litefuse.cloud
- ✓ 阿里云 SelectDB

云上托管

Demo

| 在 Litefuse 上实践 EDD - Agent 观测 评估 优化

从 litefuse.cloud 快速开始

10万条/月

开发者够用的免费额度

10万/30天/12小时/10 = 27 对话/小时

200+ 用户

上线不到一个月

真实开发者自发注册

基于 Litefuse 构建可靠的 Agent

<https://litedfuse.ai>

