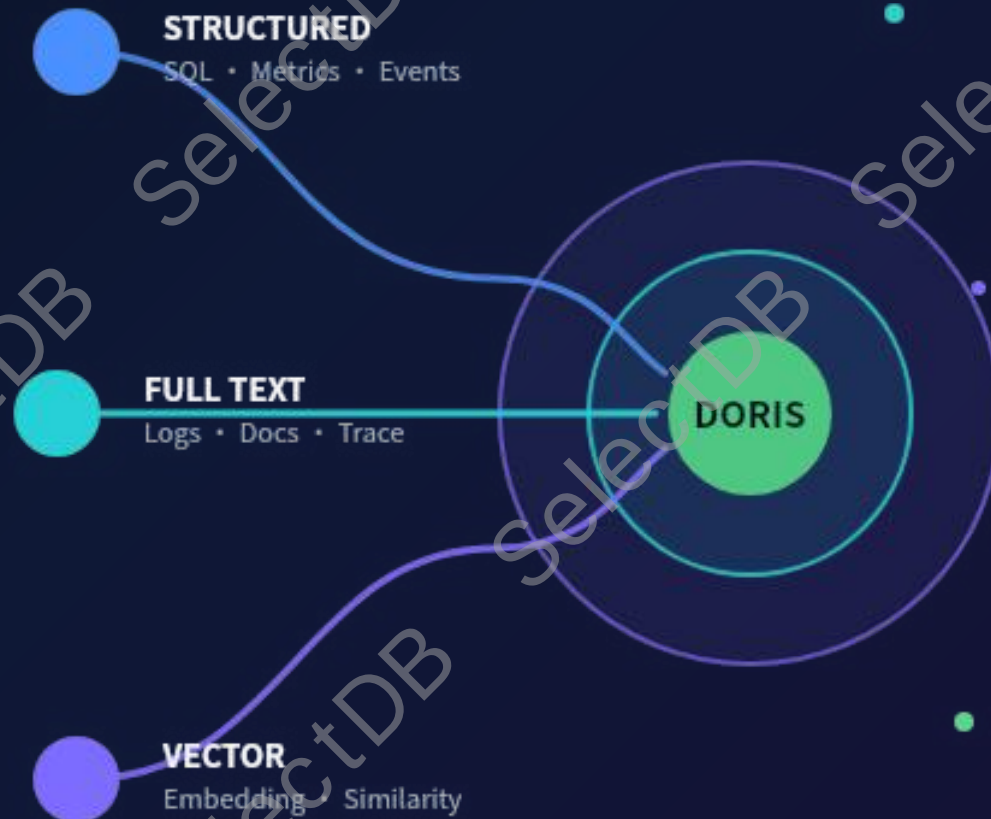




Apache Doris 的 下一次智能跃迁

从人类分析引擎
到 Agent 的实时数据底座

李昊鹏 | Apache Doris PMC





AI 时代，数据库增加了新的消费者

Human-facing Analytics

低频、交互式查询
分钟 / 秒级新鲜度
结构化指标为主
人为判断、系统展示



Agent-facing Decisions

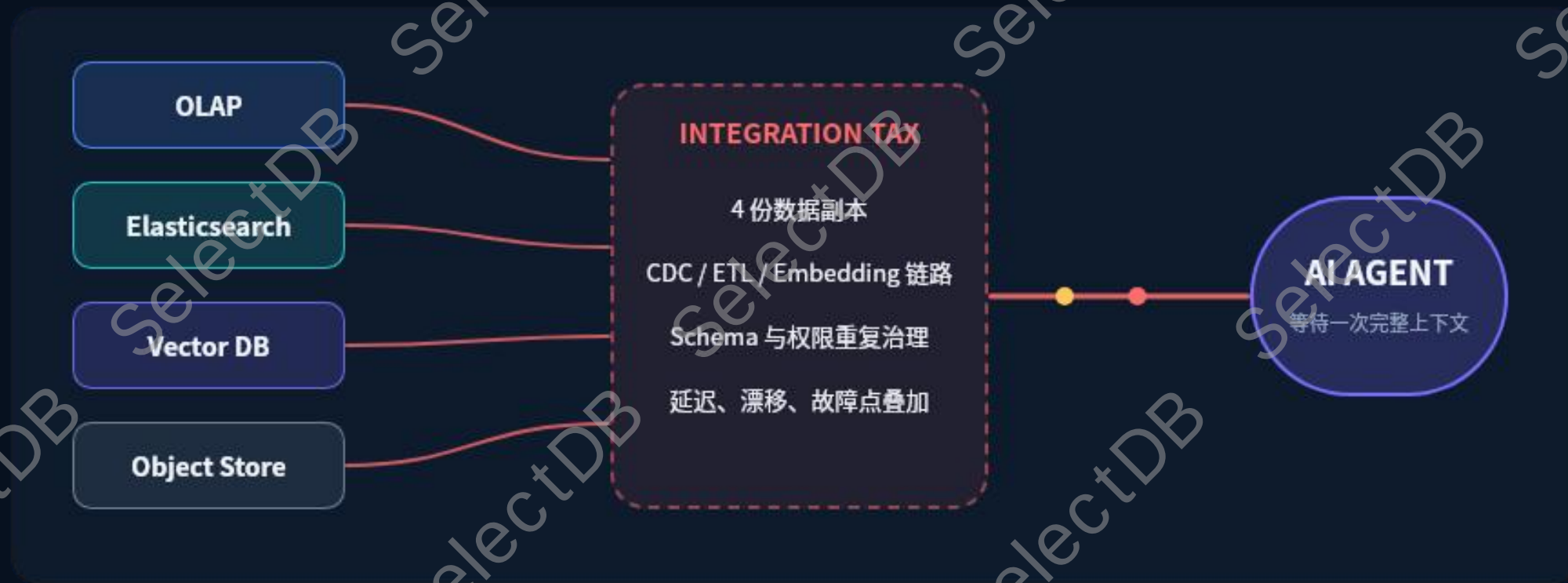
高频、程序化调用
毫秒级反馈与持续写入
文本 / 向量 / Trace / JSON
系统检索、模型决策、结果回流

数据库正在成为 Agent 的实时记忆

不是只在数据库里多一个 AI API，而是把一次智能决策的闭环压到更短。

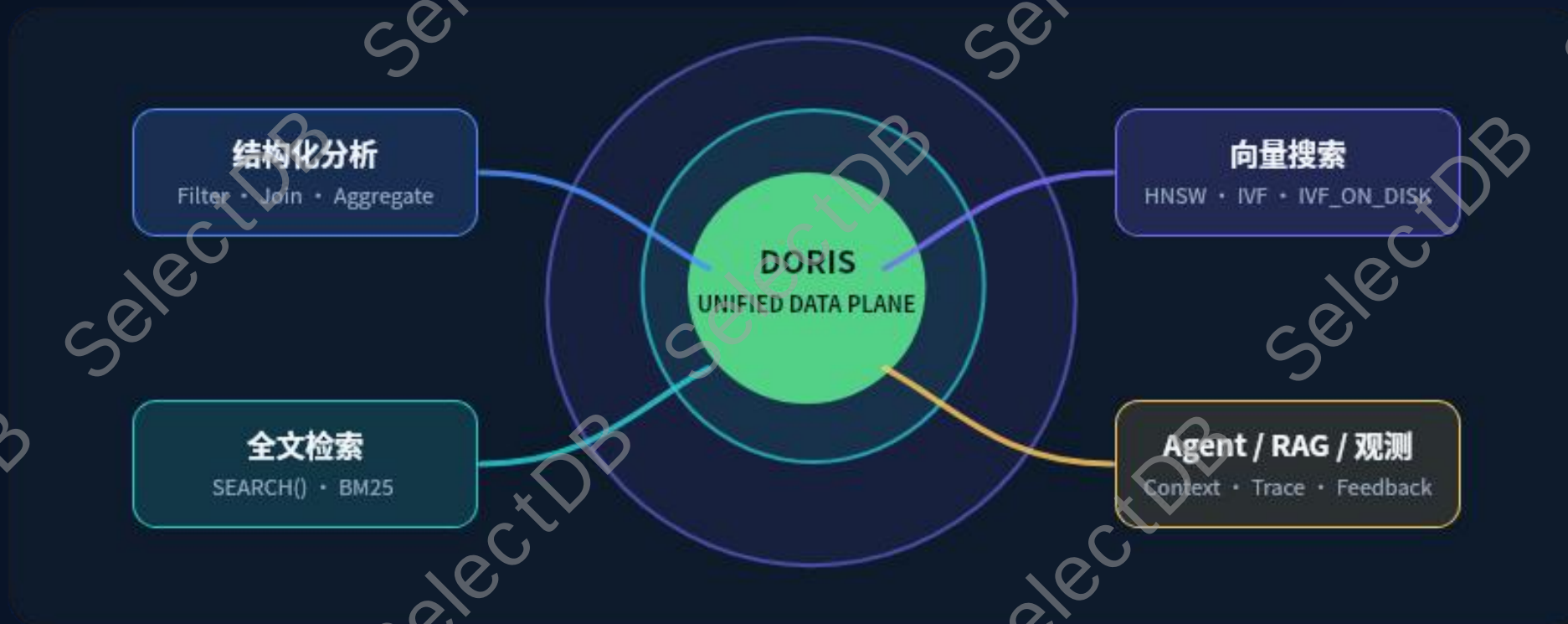


多引擎拼装, 正在成为 Agent 的系统性瓶颈





一份数据，三种检索，一个 SQL





Doris AI 能力全景：存、搜、懂、观测

存

原生承载 AI 数据

100MB JSON · VARIANT
长上下文、超宽动态 Schema

搜

混合检索与分析

BM25 · SEARCH() · ANN
结构化、关键词、语义召回同一查询

懂

模型能力进入 SQL

Embed · Classify · Extract · Generate
Filter · Summarize · Translate · Similarity

观

Agent 决策可复盘

Prompt · Reasoning · Tool Call · RAG
SEARCH + 聚合重建异常与成本链路

向量检索：从“能用”到“可生产”

900

QPS

97%

Recall

官方测试：100 万条 768 维向量, 16C / 64GB; Index Only Scan 避免读取原始向量列

HNSW

IVF

IVF_ON_DISK

SQ8 / SQ4 / PQ 量化

SEARCH(): 让全文检索成为 SQL 的一等能力

```
SELECT
  model_name,
  COUNT(*) AS errors,
  PERCENTILE_APPROX(latency_ms, 0.99) AS p99
FROM
  inference_logs
WHERE
  SEARCH('level:ERROR AND error_msg:"CUDA out of memory"')
  AND log_time > NOW() - INTERVAL 1 HOUR
GROUP BY
  model_name;
```

Query Tree, 一次解析与执行

SearchDslParser: TERM / PHRASE /
NESTED

Analyzer: 分词与多字段展开

Inverted Index: 共享 IndexReader

BM25 score() / Filter



把模型能力带到数据分析中

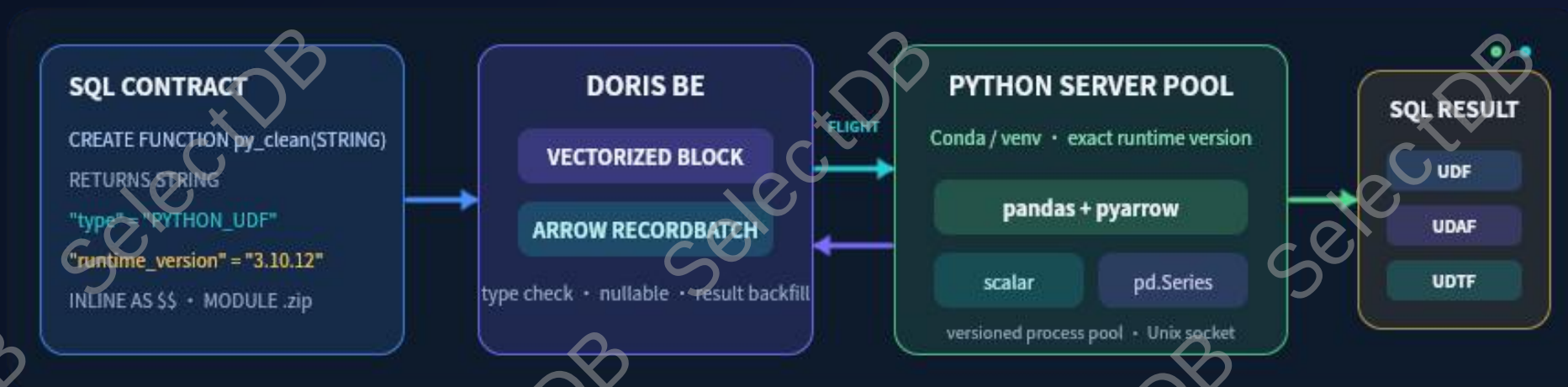
AI Resource 管理模型连接；SQL 函数把分类、抽取、生成和 Embedding 接进数据流水线。



Doris 负责数据存储、SQL查询的编排，模型的调用；模型推理由配置的外部服务完成。

Python UDF: 把 Python 生态带进 SQL 数据流

UDF / UDAF / UDTF 覆盖转换、聚合与展开; Arrow 批量交换, 多版本 Python 进程池执行。



UDF · 1 → 1

清洗 · 脱敏 · 校验 · 复杂转换

UDAF / OVER · N → 1

自定义指标 · 分布式合并 · 窗口计算

UDTF · 1 → N

CSV / JSON 拆解 · 序列生成 · 行展开



长上下文进入数据库: 100MB JSON 承载 Agent 记忆

100MB

单行 JSON 文档

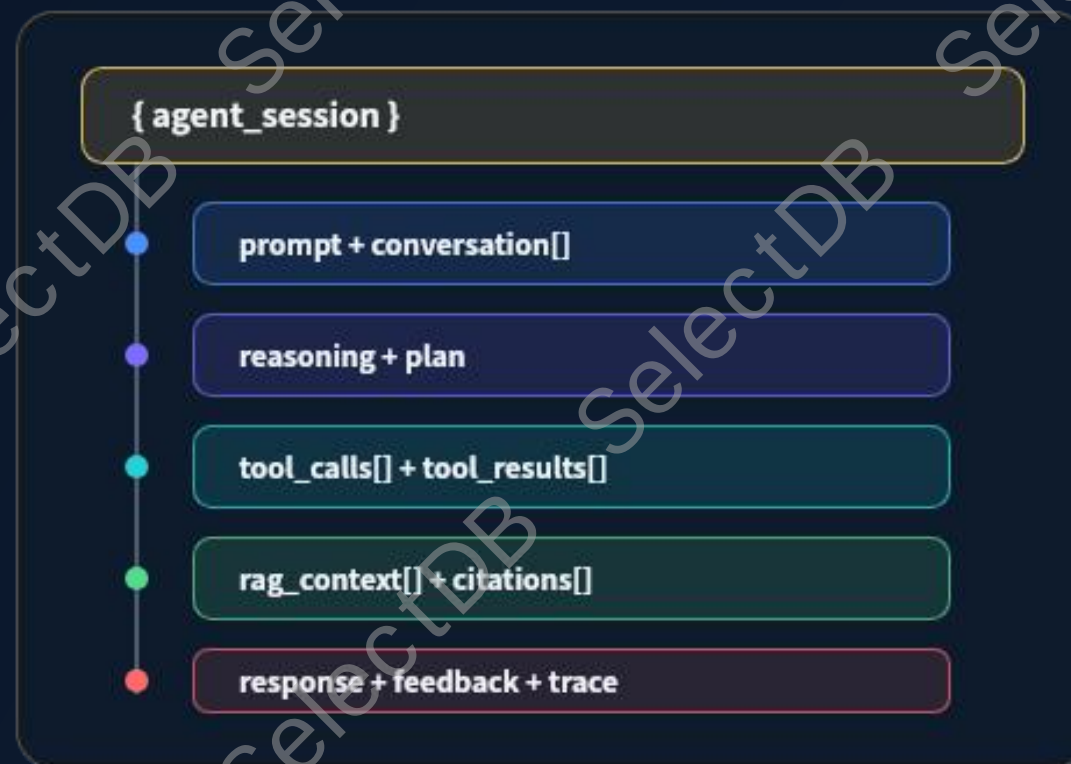
完整多轮会话, 不拆段

长文档 / 音视频转录, 不拼接

Agent Trace / Tool Call, 不丢上下文

RAG Context / Response, 可直接检索分析

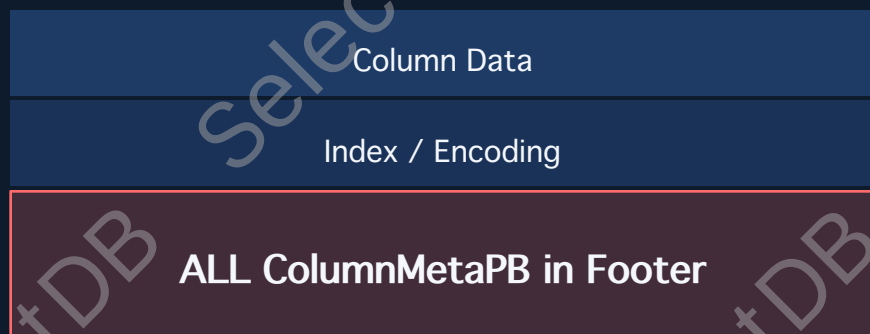
从“大字段支持”升级为可治理的长期记忆



Segment V3: 为万列动态 JSON 重写文件打开路径

Agent Trace 的字段多、变化快、冷热分化明显；瓶颈常常不是扫描，而是“打开一个超宽 Segment”。

Segment V2 - 大 Footer



Segment V3 - 元数据外置



一条生产链路：RAG + Agent + Observability 同底座闭环



下一次智能跃迁， 真正改变的是数据库的角色

01 · 统一数据：长上下文、Trace 与多模态，结构化数据同底座

02 · 统一检索：SQL + 全文 + 向量组合成一次决策上下文

03 · 统一闭环：日志、Trace 回流和复盘共享治理



感谢聆听

THANK YOU

李昊鹏 | Apache Doris PMC

让数据成为 Agent 的实时记忆。
让每一次决策更快、更完整、更可治理。

微信 | SelectDB
公众号



钉钉 | SelectDB
用户群

