

INFINI Easysearch × 达梦数据库

达梦数据库 搜索与分析加速解决方案

基于 CDC 实时同步的查询卸载与智能检索架构

国产 · 国密 · 等保

已验证 CDC 链路

信创全栈

CONTENTS

目录

01 公司简介：极限科技

实时搜索与数据分析

02 达梦数据库检索挑战

搜索体验 · 核心库承压 · 扩展成本高

03 解决思路：达梦 + Easysearch

事务库与检索分析库的互补定位

04 核心场景

全文搜索 · 海量检索分析 · 数据库加速

05 方案收益

搜索体验 · 性能 · 成本

06 客户案例

ES国产化 · 非结构化数据检索

PART 01

公司简介：极限科技

实时搜索与数据分析

01

极限数据（北京）科技有限公司

极限数据（北京）科技有限公司（品牌：INFINI Labs 极限实验室）是一家专注于实时搜索与数据分析的软件公司，位居国内搜索引擎产品厂商第一梯队，致力于成为企业大数据实时搜索分析产品的首选。

核心产品与品牌

INFINI Easysearch 分布式搜索型数据库

- 可替代 Elasticsearch
- 国产自主可控，适配信创 OS/CPU
- 集成国密算法，满足等保要求

工程师团队

核心成员来自 Elastic、Red Hat、AWS等
拥有 ECE 认证工程师 10+

精通 Elasticsearch / Easysearch /
OpenSearch 原理、架构与 Lucene 内核，
熟悉 Logstash / Kibana / Beats 生态

开源与生态

开源 Coco AI、Console、Gateway、Agent、
Loadgen 等搜索领域产品（GitHub：
infinilabs）

积极参与全球开源社区建设，推动中国技术品
牌输出

我们致力于

01 搜索引擎研发

Easysearch / Pizza 自研搜索内核，追
求极致性能与易用性

02 国产化解决方案

Elasticsearch 平滑替代，信创环境深度
适配

03 专业技术服务

搜索技术支持、迁移服务、定制方案与
培训认证

04 全球开源建设

开源产品与社区贡献，提升中国开源全
球影响力

我们的用户

金融	制造业	运营商	互联网/高科技
         	    	 	        

PART 02

达梦数据库检索挑战

DM 8

02

数据库检索的三大挑战

01

搜索体验不够，客户与坐席难满意

- 交易明细、合同、制度文件检索依赖模糊匹配，
 - 不支持拼音、纠错、同义词与相关性排序，
 - 搜索结果"能出但不够好用"。

02

分析查询拖垮核心交易库

- 报表导出、多维聚合、复杂检索直接压在核心库上，
 - CPU / IO 高位运行，与交易争抢资源，
 - 影响核心账务系统稳定性与响应。

03

海量数据"存得起、查不起"

- 日志、交易明细、客户行为数据量持续增长，
 - 堆数据库算力扩展成本高，
 - 历史数据长期保存与检索的经济性差。

根本原因：事务型数据库并非为"海量检索 + 多维分析"而设计——需要一个专职的检索与分析加速层。

PART 03

解决思路：达梦 + Easysearch

事务库与检索分析库的互补定位

03

达梦数据库：强项与能力边界

达梦的强项 · 做得好

- ✓ OLTP 事务处理、强一致性（ACID）
- ✓ Oracle / MySQL 高度兼容，迁移成本低
- ✓ PL/SQL、存储过程、触发器、物化视图
- ✓ 三权分立、强制访问控制、审计、加密引擎
- ✓ 国产 + 国密，满足等保 / 密评
- ✓ DPC、DMDSC 共享存储集群、数据守护

特定搜索场景 · 可强化

- 全文检索仅 5 种固定分词
 - 不支持自定义词典
 - 热更新词库
- 无相关性评分，无法实现“越匹配越靠前”排序
- 高级检索
 - 同义词扩展
 - 拼写纠错
 - 搜索联想
 - 高亮片段
 - ...

INFINI Easysearch: 国产搜索与分析型数据库

全文检索

IK / 拼音 / HanLP 等 20+ 中文分词
相关性评分 · 高亮 · 纠错 · 同义词 · 搜索建议

向量与语义

原生 KNN 向量字段 · embedding 服务
混合检索 · LangChain / RAG 知识库

聚合分析

50+ 桶 / 指标 / 管道聚合
列式 doc values · PB 级亚秒级检索

空间检索

Geo Point / Shape · 距离 / 多边形
Geohash 网格聚合 · LBS / 地理

分布式弹性

分片 / 副本 · 跨集群搜索
可搜索快照 · 索引生命周期管理

信创合规

国产自主可控 + 国密
支持信创 CPU / OS

核心理念：事务库 + 检索分析库

让金库专心管账，让检索分析室专心找资料——两者各司其职，体验、性能与成本才能同时最优。

达梦 = 银行金库

- 安全存放、精确记账每一笔交易
- 写入、更新、事务一致性、主数据
- 强一致、高可靠、合规审计

CDC
实时同步

Easysearch = 检索分析室

- 从海量内容里快速"找出来、关联起来"
- 搜索、聚合、向量、空间、AI 加速
- 横向扩展承接查询与分析负载

达梦（写入 / 事务） + CDC 实时同步 + Easysearch（检索 / 分析 / AI） = 最优架构

CDC 实时同步方案：

同步链路工作方式

1

全量同步

首次将达梦存量数据一次性同步到 Easysearch，建立基线

2

增量同步（CDC）

实时捕获达梦 INSERT / UPDATE / DELETE 变更，同步到目标端

3

持续运行

同步任务持续运行，保持 Easysearch 与达梦近实时一致

CDC 工具：

- 已通过达梦 × Easysearch 兼容认证测试
- 支持高并发、万级变更稳定近实时同步
- 读写一致、无积压、可观测

其他可选接入方式

- Logstash / JDBC 批量同步（适合低频场景）

关键设计原则

- 职责分离：达梦管事务与主数据，Easysearch 承接查询与分析
- 近实时一致：CDC 增量同步，写入仍走达梦保证事务完整
- 一份数据多种服务：同步一次，同时支撑搜索、分析、向量、RAG

产品兼容认证

结论：通过

达梦 DM8 × INFINI Easysearch

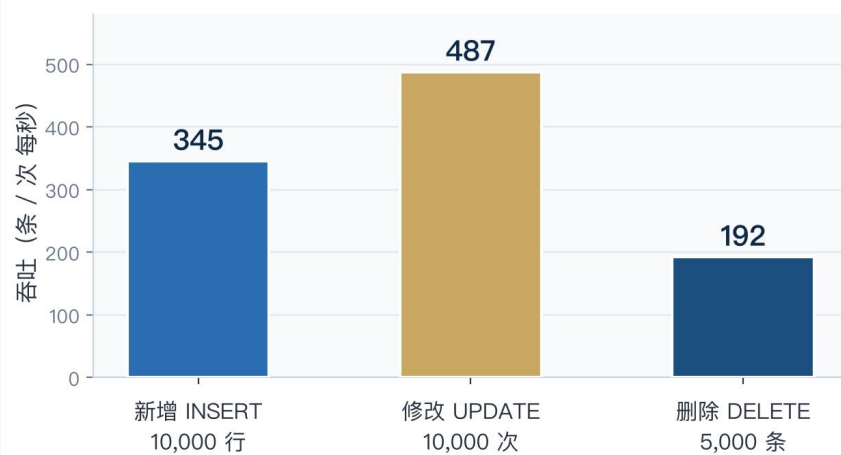
全量同步 + 增量同步均验证通过

CDC 高压同步测试

万级变更 · 30 秒内完成 · 读写一致 · 无积压

三类变更全部实时同步，目标端数量与样例一致

三类 CDC 变更实测吞吐对比



测试环境

单节点 Easysearch 环境(单分片)

万级 CDC 变更可稳定近实时同步

具备客户压测场景下的可用性与数据一致性

新增 ≈ 345 · 修改 ≈ 487 · 删除 ≈ 192 条/秒

Easysearch支持动态扩容，提升处理能力

PART 04

核心场景

全文搜索 · 海量检索分析 · 数据库加速

04

企业级全文搜索 / 站内搜索

金融场景

- 交易明细 / 流水检索（按金额、对手、摘要模糊搜索）
- 合同 / 制度 / 合规文件全文检索
- 客服知识库搜索（坐席快速检索 FAQ 与话术）
- 客户 360 视图（客户、账户、交易关联检索）
- APP 检索

收益

搜索体验从“能用”升级到“好用”——输错字、输拼音也能搜到，结果按相关性智能排序。

达梦数据库 × Easysearch 方案

- IK / 拼音 / HanLP 等 20+ 中文分词器
- BM25 相关性评分 + Function Score 自定义打分
- 结果高亮 + 拼音检索 + 同义词图
- Completion / Autocomplete 搜索建议
- More Like This 相似内容检索

海量数据的实时检索与分析

金融场景

- 交易日志 / 系统日志检索与审计追溯
- 交易明细多维分析（按网点、产品、时段聚合）
- 客户行为数据画像与趋势分析
- 风控实时关联分析与异常监测

收益

卸载分析查询压力、核心库受保护，检索与分析提速降本，历史数据查询更友好。检索体验升级，毫秒级响应

达梦数据库 × Easysearch 方案

- 倒排索引 + 列式 doc values，PB 级亚秒级检索
- 50+ 桶 / 指标 / 管道聚合，实时多维分析
- 索引生命周期管理，冷热分层成本可控
- 可搜索快照，冷数据“查得起”
- 同时支持 SQL 访问，迁移成本低

Easysearch 通用使用场景

场景		解决的核心问题
01	统一日志平台与监报告警	日志分散在大量服务器，无法集中查询、分析与告警
02	业务搜索与高并发查询	海量业务数据的低延迟、高并发、灵活组合查询
03	业务实时监控与可视化	交易与业务指标的实时监控、异常发现与汇报
04	用户画像与标签系统	基于标签的精准营销、个性化推荐与客户分析
05	非结构化数据检索	海量保单影像、文件的内容索引与场景化查询
06	OLAP 报表与聚合分析	多维度统计聚合、报表生成与决策支持

AI 场景：Easysearch 为 AI 应用提供检索底座

RAG 检索增强生成

向量 + 全文混合检索，回答有据可依、减少幻觉；作为智能体的“长期记忆库”，节约 token，提升使用体验；
集成 LangChain / Spring AI；

语义搜索

Embedding 向量化 + k-NN 检索，理解用户意图而非关键词；与拼音、同义词、纠错能力融合

智能客服与 AI 助手

企业知识库智能问答；Coco AI 开源助手统一连接内外部异构数据源，融合大模型能力辅助决策

多模态搜索

图像、视频指纹、音频等特征向量相似检索，实现跨模态内容理解与快速定位

推荐与个性化

基于用户画像 / 商品特征向量相似召回，实现资讯、产品与服务的智能推荐

反欺诈与异常检测

业务特征向量化，与已知欺诈模式实时相似度比对，金融风控场景实时识别与预警

AI 就绪底座：原生 KNN 向量字段 · Ingest / Search Pipeline 自动 Embedding · 向量与全文同索引混合检索 · 亿级向量毫秒级响应

PART 05

方案收益

达梦数据库 × Easysearch 方案

05

方案收益

体验

提供毫秒级相关性搜索 + 拼音 / 纠错 / 高亮 / 搜索建议"

性能

降低核心交易库查询负载，避免分析查询拖垮 OLTP；

成本

Easysearch 横向扩展承接检索分析，数据存储高压缩比，存算分离（S3），比堆数据库算力更经济；

合规

达梦 + Easysearch = 双国产 + 双国密 + 信创全栈，满足等保 / 密评，无协议风险

AI 就绪

原生 KNN 向量字段支持余弦 / 内积 / L2 近邻检索，Embedding Pipeline；结合 LangChain / Mem0 构建 RAG 知识库，支撑智能客服、合规问答、反洗钱语义关联等金融 AI 场景

PART 06

客户案例

06

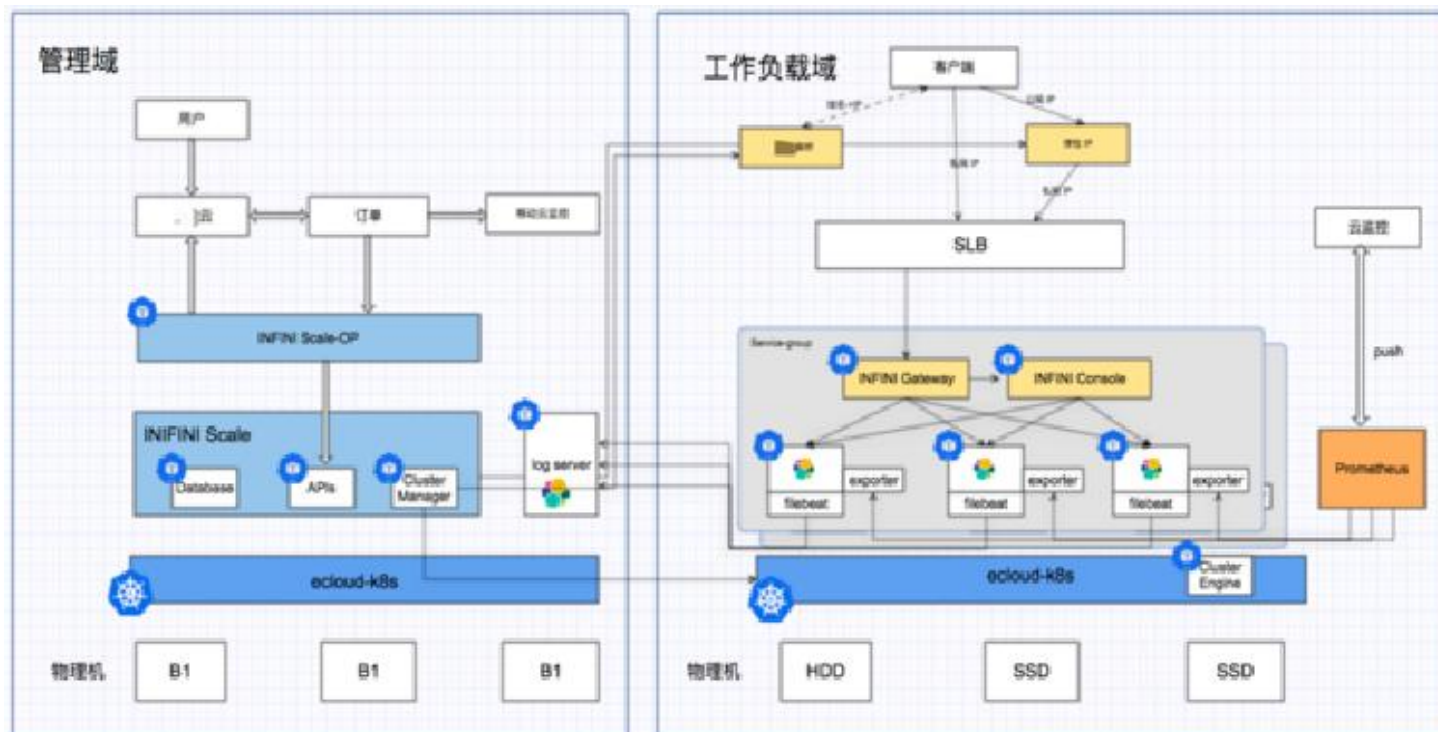
在公有云上对外提供托管 ES 服务，涉及多个使用区、机房。

早期使用 Elastic 产品支撑该业务，目前以完成国产化改造，使用 Easysearch 提供服务。该项目获得"2023年中国信通院大数据星河标杆案例"。

13 个
3000+
可用区
PB级
最大集群

节点
3年+
稳定运行

2023年度中国信通院大数据"星河"标杆案例



2022 年 9 月，国资委发布 79 号文，针对98 家央企提出信创要求，标志行业信创全面开展。

2023 年开始进行 ES 国产化替代工作。

几十个
业务场景

700+
节点

15 TB
日均写入

3年+
稳定运行

业务属性划分	
IT基础应用	24
大营销	109
大财务	7
大数据	12
大研发	5
分子公司	7
管理与支持	6

各领域超大系统		
IT基础应用	统一监控平台	API网关
大营销	智能网联云平台2.0	红旗营销管理系统
大财务	企业业务中台 (EIP)	财务共享中心
大数据	CDP平台	奔腾UBI
大研发	一汽云仿真平台	D-life6.0
分子公司	旗享智投平台	资产交易电商平台
管理与支持	协同办公系统	管理者工作台

总结

达梦保障事务与主数据，
Easysearch 补齐搜索、分析与 AI 能力。

一套 CDC 链路，三重收益——

搜索体验提升 · 核心库压力下降 · 扩展成本优化

CDC 链路

国产 · 国密

信创全栈适配



 400 139 9200