



申请创建Elasticsearch代码仓进行自维护演进

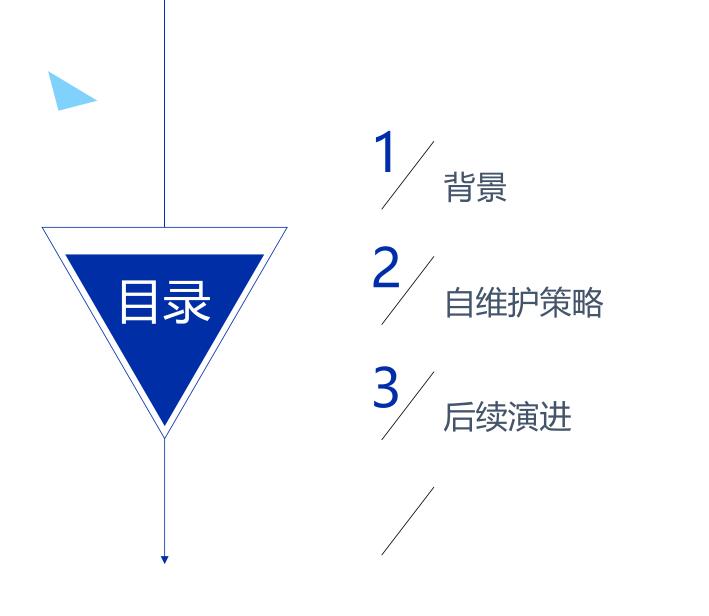














▶ 背景: Elasticsearch 简介 & 软件价值

ES简介:

- Elasticsearch是一个基于Apache Lucene实现的分布式近实时检索和分析引擎
- 支持多种输入数据:结构化、非结构化、文本、数值、地理、向量
- 支持多种检索方式: 倒排检索、向量检索
- 支持丰富的数据分析、聚合能力

业界广泛使用

目前DB-Engines排名中,总体排名第7

	Rank				Score		lan
Jan 2024	Dec 2023	Jan 2023	DBMS	Database Model	Jan 2024	Dec 2023	Jan 2023
1.	1.	1.	Oracle 🚻	Relational, Multi-model 📵	1247.49	-9.92	+2.33
2.	2.	2.	MySQL 🚹	Relational, Multi-model 🛐	1123.46	-3.18	-88.50
3.	3.	3.	Microsoft SQL Server 🖽	Relational, Multi-model 👔	876.60	-27.23	-42.79
4.	4.	4.	PostgreSQL 🚹	Relational, Multi-model 🛐	648.96	-1.94	+34.11
5.	5.	5.	MongoDB 🔠	Document, Multi-model 🛐	417.48	-1.67	-37.70
6.	6.	6.	Redis 🖽	Key-value, Multi-model 🔞	159.38	+1.03	-18.17
7.	7.	↑ 8.	Elasticsearch	Search engine, Multi-model 🛐	136.07	-1.68	-5.09

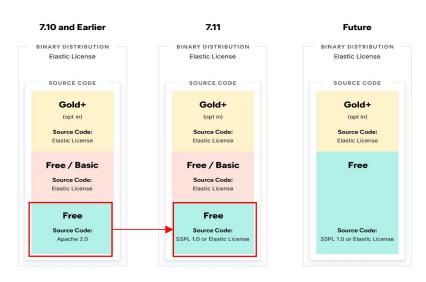
搜索引擎类DB排名第1

	Rank				Score			
Jan 2024	Dec 2023	Jan 2023	DBMS	Database Model	Jan 2024	Dec 2023	Jan 2023	
1.	1.	1.	Elasticsearch	Search engine, Multi-model 🛐	136.07	-1.68	-5.09	
2.	2.	2.	Splunk	Search engine	92.72	-3.57	+4.32	
3.	3.	3.	Solr	Search engine, Multi-model 🔞	45.13	+0.93	-1.26	
4.	4.	4.	OpenSearch 🔠	Search engine, Multi-model 🛐	13.95	+0.36	+3.51	

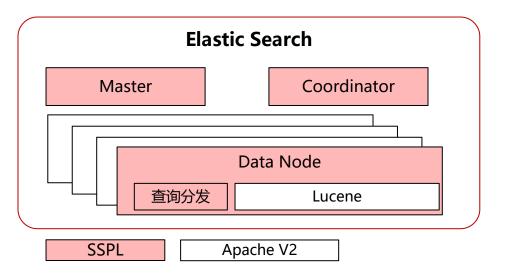


→ 背景: Elasticsearch宣布自ES 7.10.2版本后修改License 到 SSPL

- 2021.2 Elasticsearch软件 的 Apache 2.0-licensed 源码协议被修改成 SSPL, 对云服务提供商造成巨大影响。
- 2022.5 ES官方宣布最后一个Apache 2.0 版本 7.10.2 EOL,后续漏洞、缺陷将不会继续维护



自2021年2月 7.11版本发布时修改协议



License变更影响范围

License变化影响和风险:

- 1. 云服务和软件分发使用情景下,由于SSPL的强传染性**无法使用7.11及以后的版本**,影响使用ES产品和服务的技术先进性
- 2. 后续相关**漏洞、缺陷没人维护、修复**,影响使用ES产品和服务的安全性



▶ 背景: 业界依据7.10.2开源版本进行自演进案例

业界多家云服务均基于最后一个Apache 2.0版本 (7.10.2) 继续进行内核的自演进,针对性能、成本、功能进行深度增强,提升其竞争力

百度云 BES 基干7.10.2做内核增强

阿里ES云服务 强 基于7.10.2提供内核增强

AWS OpenSearch 基于7.10.2分叉社区演进

文档中心 > ElasticsearchBES > Elasticsearch > 用户手册 > 版本升级 > 内核版本说明

内核版本说明

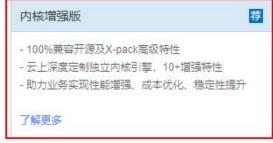
更新时间: 2023-12-18 产品详情 自助选购

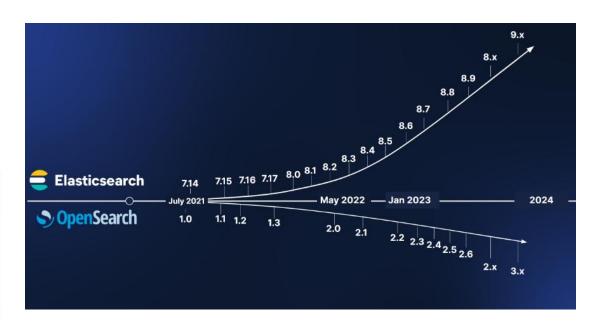
在完全兼容开源Elasticsearch内核的所有特性基础上,百度云Elasticsearch在冷热分离、向于百度云Elasticsearch团队在多种应用场景下所积累的丰富经验,提升集群稳定性、增强性性。

7.10.2

内核版本	发布说明	
1.6.1	bug修复 1、修复了大数据量下读取向量数据偶现的io错误	
1.6.0	新增、优化特性 1、支持了通过SIMD指令加速script距离计算 2、优化了读取向星数据的内存消耗	
1.5.1	bug修复 1、修复了hnsw索引带filter检索场景,在有nested字段的情况下影响召回的问题	

通用商业版
- 100%兼容开源及X-pack高级特性
- 云原生管控与智能运维,提高运维管理效率
- 多版本覆盖支持,提供开源最新特性
了解更多







▶ 当前具备对ES内核开发和维护能力

查询能力:多样化召回、匹配特征抽取、查询跳变控制、检索结果proto序列化

查询扩展:近似短语匹配、dis_max_const、conditional boosts、扩展function score

读写分离 针对一些业务场景 索引编码: doc级别定制属性 & 查询接入属性打分 数据离线spark集群构建 性能、效果、功能需求做定制开发 外部分词服务接入 采集和扩充 数据存储 搜索和分析 累积代码量超过10w+ 离线数据在线加载 灵活性 全文本搜索 多集群在线写一致性 理和运行 客户端和 API 数据类型 倒排索引 采集节点 全文本搜索 (倒排索引 运行时字段 可扩展性和弹性 安全性 文档存储库(针对非结构化数据) Beats 查找运行时字段 语言客户端 集群和高可用性 Elasticsearch 安全设定 Elasticsearch-Hadoor Logstash 时序/分析(列式存储库) 跨集群搜索 白动节点恢复 加來通信 Flasticsearch DSI Anache Hive 社区贡献的采集器 地理空间 (BKD 树) 相关性评分 自动数据再平衡 数据静态加密支持 Elasticsearch SQI Apache Pig 矢量搜索 (ANN 水平可扩展性 基于角色的访问控制 (RBAC) 事件查询语言 (EQL) Apache Spark 查询 DSL 数据扩充 基于属性的访问控制 (ABAC) 非同步搜索 请求鉴权 机架意识 JDBC 客户端 Apache Storm 字段级和文档级 API 安全性 Elastic Common Schema Highlighter Hmac请求鉴权 跨集群复制 字段和文档级别安全性 ODBC 客户端 商业智能 (BI) 处理器 数据静态加密支持 自动补全 (auto-complete) STS请求鉴权& 加密 跨数据中心复制 审计日志 面向 Elasticsearch 的 Tableau 连接器 插件和集成 分析器 提示器 (did-you-mean) 分词器 IP 筛选 管理 更正 (拼写检查) 管理 部署 筛洗器 Security Realm 集群式索引 Percolator 从快照恢复 下载并安装 REST API 语言公析器 数据快照和还原 单点登录 (SSO) 查询剖析器/优化器 索引生命周期管理 文档 API Elastic Cloud 动态映射 汇总索引 第三方安全性集成 基于许可的搜索结果 数据层 搜索 API Elastic Cloud Enterprise 动态可更新同义词 Geo-match enrich 处理器 快昭生命周期管理 聚合 API Elastic Cloud on Kubernete ALERTING 结果同定 快照和还原 高可用性、可扩展警报 采集 API Helm Charts 安全漏洞修复: 可搜索快照 通过电子邮件、Webhook、IBM 管理 API Docker 容器化 CVE-2021-22144 分析 Resilient, Jira, Microsoft Teams CVE-2021-22145 数据汇总 PagerDuty、ServiceNow、Slack和 聚合 xMatters 传送的涌知 CVE-2021-44228 数据流 Graph 探索 检索性能: 交集超时, 检索超时 CVE-2023-31419 Transforms 集群可维可测 控制强化, 堆外内存管理索引I/O CVE-2023-46673 升级助手 API 性能监控日志, coor层慢日志, 请求时延分解, 多segment并发, 异步缓存, 定向 MACHINE LEARNING API 密钥管理 DSL日志打印、检索深度 时序型预测 id查询优化 时序型数据异常监测 集群自动化部署

能力总结

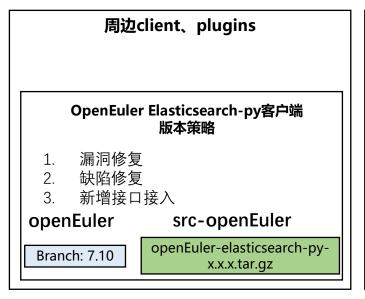
内部维护大量的特性定制能力,涵盖集群管理、检索能力、索引编码、可维可测、自研单机引擎,具备特性演进、缺陷发现及修复能力

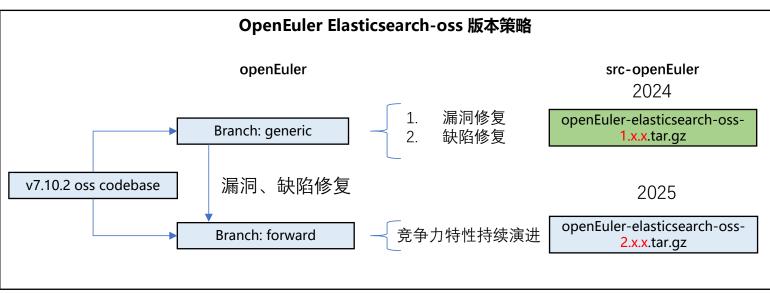


▶ 自维护策略:代码维护方式及SIG组织归属

基于Elasticsearch-oss 7.10.2版本,进行漏洞修复,竞争力特性研发工作,版本系列分为2个系列发布

- ◆ generic: 基于7.10.2 oss的稳定版本,进行漏洞、缺陷修复,发布1.x.x版本
- ◆ forward: 业务竞争力版本, 针对社区重点业务需求以及检索引擎发展方向, 持续提供竞争力特性









▶ 后续演进

2024 **7.10.2版本自维护**

1、社区建立Elasticsearch 衍生版本,进

行漏洞、bug修复

2、启动竞争力版本规划、开发

2025+

主力版本演进

- 1、社区需求规划及开发
- 2、竞争力特性持续演进
- 3、联合社区发展伙伴,合作共建

4、.....

欢迎大家共同参与构建社区Elasticsearch自维护、自演进版本,满足社区各行业搜索、分析诉求



