



申请创建Elasticsearch代码仓进行自维护演进



openEuler公众号





1 / 背景

2 / 自维护策略

3 / 后续演进



► 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			DBMS	Database Model	Score		
Jan 2024	Dec 2023	Jan 2023			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

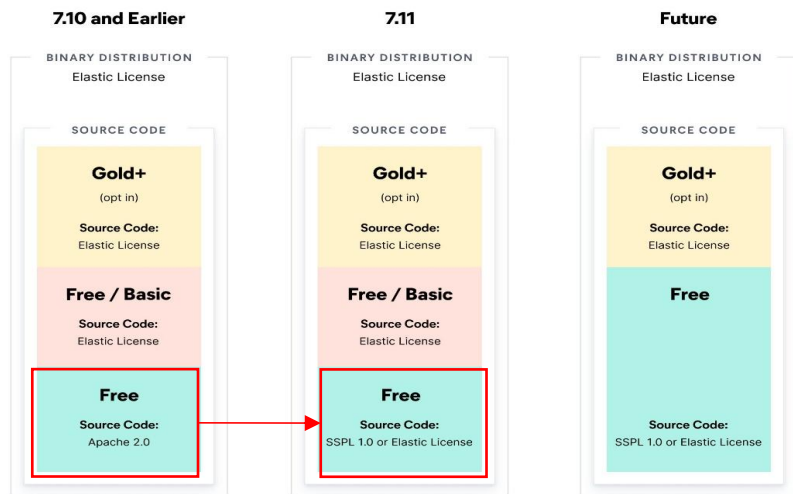
搜索引擎类DB排名第1

Rank			DBMS	Database Model	Score		
Jan 2024	Dec 2023	Jan 2023			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

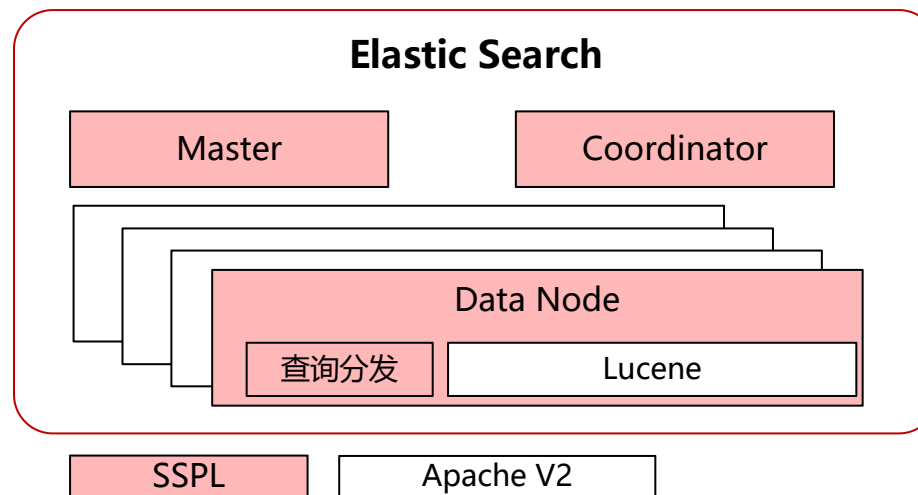
► 背景：Elasticsearch宣布自ES 7.10.2版本后修改License 到 SSPL

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 基于7.10.2做内核增强

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、增强性能。

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、优化了读取向量数据的内存消耗
1.5.1	bug修复 1、修复了hns索引带filter检索场景，在有nested字段的情况下影响召回的问题

阿里ES云服务 基于7.10.2提供内核增强

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](#)

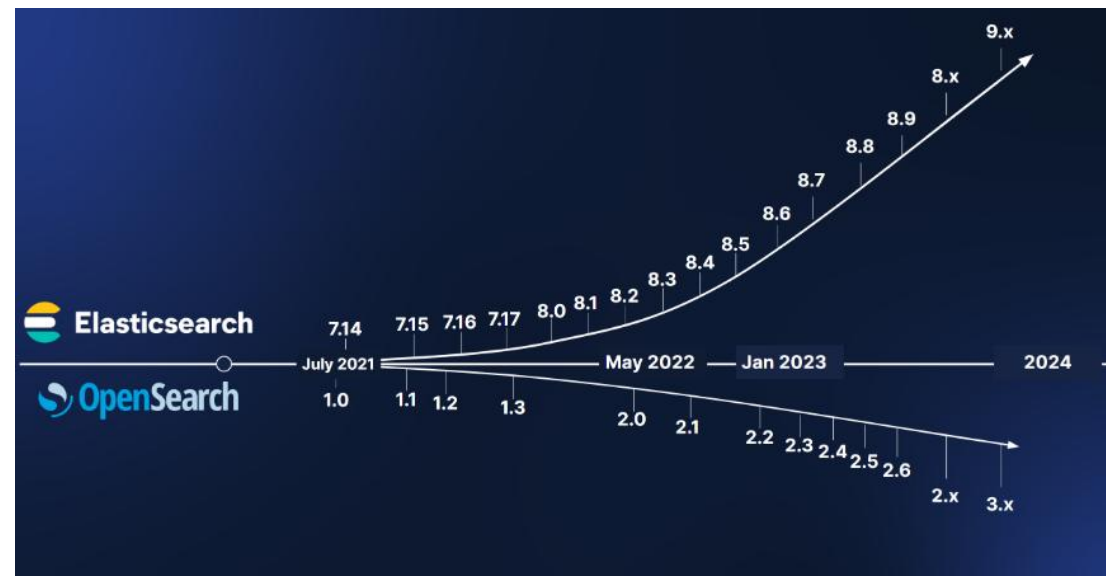
内核增强版

荐

- 100%兼容开源及X-pack高级特性
- 云上深度定制独立内核引擎，10+增强特性
- 助力业务实现性能增强、成本优化、稳定性提升

[了解更多](#)

AWS OpenSearch 基于7.10.2分叉社区演进



► 当前具备对ES内核开发和维护能力

针对一些业务场景
性能、效果、功能需求做定制开发
累积代码量超过10w+

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

索引编码：doc级别定制属性 & 查询接入属性打分

读写分离
数据离线spark集群构建
外部分词服务接入
离线数据在线加载
多集群在线写一致性

请求鉴权
Hmac请求鉴权
STS请求鉴权& 加密

安全漏洞修复：
CVE-2021-22144
CVE-2021-22145
CVE-2021-44228
CVE-2023-31419
CVE-2023-46673

集群可维可测：
性能监控日志，coord层慢日志，请求时延分解，DSL日志打印、检索深度
集群自动化部署

检索性能：交集超时，检索超时
控制强化，堆外内存管理索引I/O，多segment并发，异步缓存，定向id查询优化



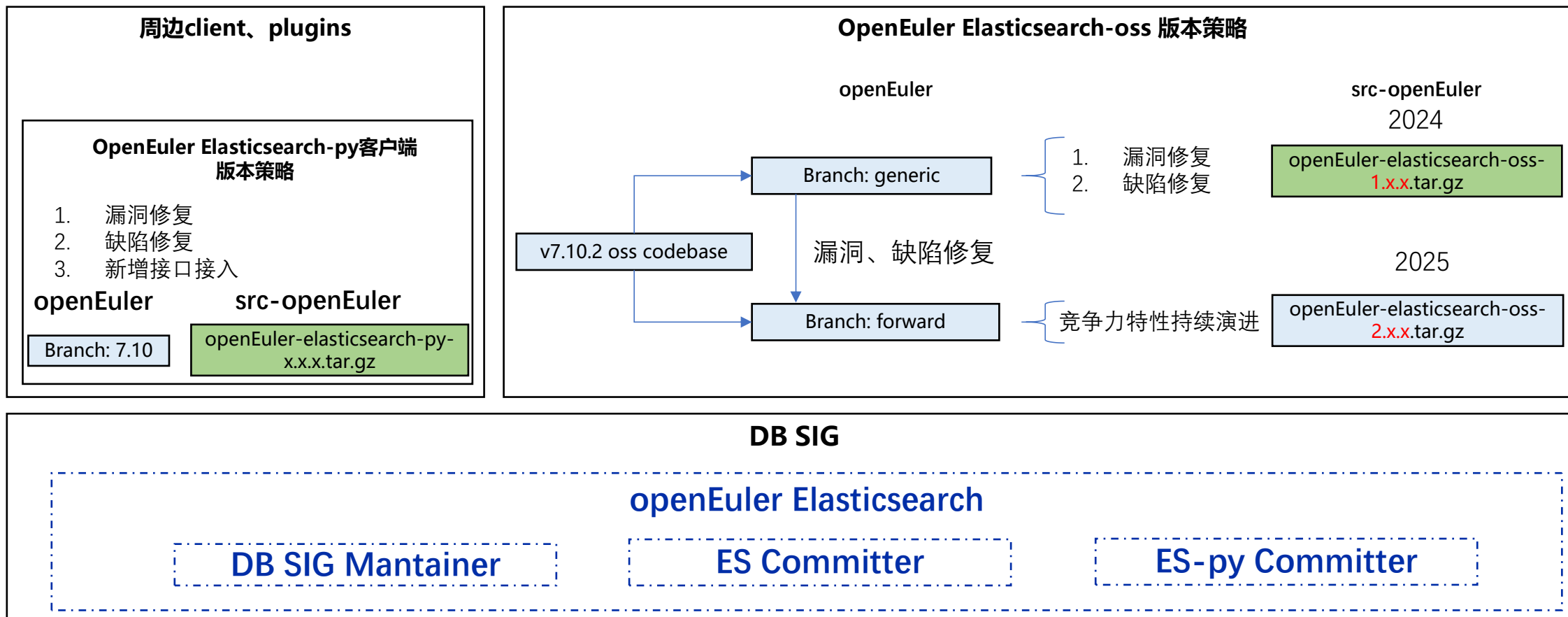
能力
总结

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

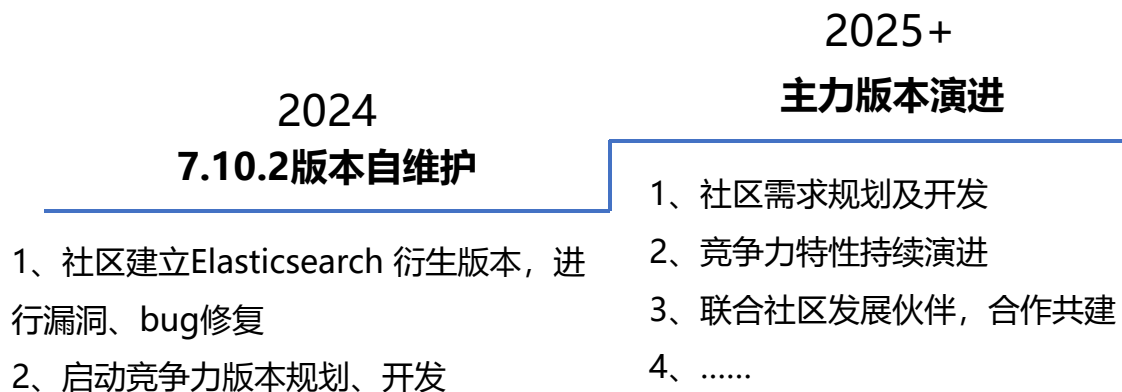
► 自维护策略：代码维护方式及SIG组织归属

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



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



THANKS