

Elasticsearch Service

产品简介



腾讯云

【 版权声明 】

©2013–2026 腾讯云版权所有

本文档（含所有文字、数据、图片等内容）完整的著作权归腾讯云计算（北京）有限责任公司单独所有，未经腾讯云事先明确书面许可，任何主体不得以任何形式复制、修改、使用、抄袭、传播本文档全部或部分内容。前述行为构成对腾讯云著作权的侵犯，腾讯云将依法采取措施追究法律责任。

【 商标声明 】



及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。未经腾讯云及有关权利人书面许可，任何主体不得以任何方式对前述商标进行使用、复制、修改、传播、抄录等行为，否则将构成对腾讯云及有关权利人商标权的侵犯，腾讯云将依法采取措施追究法律责任。

【 服务声明 】

本文档意在向您介绍腾讯云全部或部分产品、服务的当时的相关概况，部分产品、服务的内容可能不时有所调整。您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或默示的承诺或保证。

【 联系我们 】

我们致力于为您提供个性化的售前购买咨询服务，及相应的技术售后服务，任何问题请联系 4009100100或 95716。

文档目录

产品简介

产品概述

产品版本

Elasticsearch 版本支持说明

日志增强版介绍

智能搜索开发介绍

产品功能

高级特性(X-Pack)

产品优势

应用场景

能力与限制说明

相关概念

产品简介

产品概述

最近更新时间：2024-10-30 17:12:01

腾讯云 Elasticsearch Service (ES) 是云端全托管海量数据检索分析服务，拥有高性能自研内核，集成 X-Pack。ES 支持通过自治索引、存算分离、集群巡检等特性轻松管理集群，也支持免运维、自动弹性、按需使用的 Serverless 模式。使用 ES 您可以高效构建信息检索、日志分析、运维监控等服务，它独特的向量检索还可助您构建基于语义、图像的 AI 深度应用。

腾讯云 ES 集成了腾讯云计算在计算、存储、安全等领域的领先技术优势，又保持了 Elasticsearch 本身的兼容与开放，拥有丰富的集群管理功能以及安全、弹性、高可用等特性，同时也集成了官方的高级商业特性 (X-Pack)，在开源的基础上，增加了权限管理、SQL、机器学习、告警等功能，可以帮助您简化集群部署、运营管理等基础运维工作，更加聚焦于业务本身。

通过腾讯云 ES，您可以快速构建海量数据存储搜索、实时日志分析等应用，例如网站搜索导航、企业级搜索、服务日志异常监控、点击流分析等。

主要组件

- **Elasticsearch**

分布式搜索引擎，可以对海量数据进行存储、全文检索、统计分析等，提供了 RESTful API 以及各类语言客户端，可以灵活地按照业务需求进行开发。

- **Kibana**

数据可视化工具，可以方便地对存入 Elasticsearch 集群的数据进行查询和分析。

- **高级特性 (X-Pack)**

Elasticsearch 官方高级商业插件，提供了多项高级功能，包括数据权限管控，可以配置字段级别的权限控制；SQL、JDBC 的连接方式，可以很方便地同原有的业务系统进行集成；机器学习和告警，可以针对存入集群的数据，进行分析和波动趋势的学习，预测数据变化趋势，并在数据出现大的波动时进行告警。

产品版本

最近更新时间：2026-03-06 11:30:09

我们深知每个业务都有其独特的需求和挑战，因此腾讯云 ES 提供了多个产品版本，来满足不同的使用场景。本文将各版本的功能特性进行对比说明，旨在帮助您根据自身应用场景、业务规模、性能需求以及技术栈做出合适的选择。

标准版、日志增强版与AI 搜索增强版的差异

项目	标准版	日志增强版	AI 搜索增强版
主要特点	适用各场景，社区版本跟进更快	更高的写入性能，更低的成本	更多 AI 增强支持、更高的稳定性保障
适用场景	日志、搜索、分析、AI 增强等各类场景	<ul style="list-style-type: none"> 日志分析：Web 日志、App 日志、数据库日志等 指标分析：服务器指标、网络指标、存储指标等 应用性能追踪：Python、NodeJS 等 	<ul style="list-style-type: none"> RAG：与大模型配合构建基于检索增强的智能问答应用等 向量搜索：作为向量数据库用以构建各种 AI 应用 经典搜索：电商搜索、全文检索等
核心优势	<ul style="list-style-type: none"> 对社区版本覆盖更多，跟进更快 同时包含部分腾讯云 ES 自研技术 	<ul style="list-style-type: none"> 成本：基于对象存储的存算分离，成本更低 写入：特有索引写入加速技术，性能提升3 - 5倍 	<ul style="list-style-type: none"> AI 增强：支持调用腾讯自研向量化、语义重排模型、 可用性：自研内核的基础上对查询延时、并发、排序性能全面优化，在多副本+多可用区条件下提供 99.95%可用性保障 向量裁剪：支持去除向量行存以及原始高位向量，可大幅节省磁盘空间
架构模式	通用于日志/搜索/分析等各类场景的经典 ES 架构、热温架构	专为日志场景打造的存算分离架构、热温架构	适用于搜索场景的经典 ES 架构
支持版本	9.13、8.16.1、7.14.2等	7.14.2	9.1.3、8.16.1、7.14.2

高级特性	白金版、基础版、开源版	白金版	白金版
计费	按集群节点规格、存储空间、节点数计费	<ul style="list-style-type: none"> 按集群节点规格、存储空间、节点数计费 存算分离架构后会产生分布式共享存储费用（日志场景下，费用更低） 	按集群节点规格、存储空间、节点数计费

独有自研技术对比

- 日志增强版相比标准版在内核层面提供更多的高级功能，提升集群写入和查询性能，降低海量数据的存储成本。
- 搜索增强版相比标准版提供更多AI 增强功能和更高稳定性保障。

分类	特性	描述	标准版	日志增强版	AI 搜索增强版	
支持的 Elasticsearch 版本			9.1、8.16、7.14	7.14	7.14	9.1、8.16
稳定性	SLA	服务级别协议，承诺服务质量、可用性和责任	最高 99.9%	最高 99.9%	最高 99.95%	最高 99.95%
	全链路熔断限流	结合弹性漏斗和全链路内存熔断，综合考虑 CPU、IO、内存矛盾，支持高并发和大查询的混合场景	✓	✓	✓	-
	分散置放群组	将集群节点置放在不同底层硬件上，提升容灾性和高可用性	-	-	✓	✓
AI 增强	文本向量化模型	支持调用腾讯自研文本向量化模型	-	-	-	✓
	语义重排序模型	支持调用腾讯自研语义重排序模型	-	-	-	✓

	向量存储裁剪	支持去除向量行存以及原始高位向量，可大幅节省磁盘空间	-	-	-	√
存储成本	存算分离	自研混合存储架构，缓存分层、智能卸载、实现冷热一体搜索，存储成本节约50% - 80%	-	√	-	-
	FST Off-Heap	自研 FST 堆外存储，降低堆内内存使用量，大幅提升单节点磁盘管理能力	√	√	√	-
	数据深度压缩	针对行存、列存、倒排索引字典文件做了系统性编码优化，并支持细粒度开关控制，存储成本下降20%+	√	√	√	-
写入性能	写入加速	自研无共享、自闭环读写分离架构，隔离资源的同时提升写入性能5 - 20倍	-	√	-	-
	物理复制	主从副本物理复制，消除副本冗余计算，写入性能提升50%	-	√	-	-
	定向路由	分片定向路由优化，优化写入高扇出、长尾问题，写入性能提升30%+	√	√	-	-
查询性能	查询/IO 并行化	自研多级并行查询框架，支持全部查询场景，查询性能提升3 - 5倍	-	√	-	-

	查询拦截与改写	自研关键词查询 DSL 拦截与改写优化，帮助拒绝不合理的大查询，保障集群稳定性	-	-	√	-
	自适应旁路 Merge 策略	通过对 Segment 进行合并收敛，减少查询随机 IO，总体查询性能提升 2 倍以上	-	√	√	-
	查询缓存优化	大幅提高了查询的并发能力，QPS 提升 50% 以上	√	√	√	√
	查询裁剪	节点时序索引查询裁剪，时间范围裁剪优化为边界定点裁剪，高基维时序检索性能提升 10 倍 +	√	√	√	-
	排序加速	基于 bkd 存储架构在遍历倒排表时跳过不必要的数据比对，排序性能提升 5 - 10 倍	√	√	√	-
	鉴权加速	X-Pack 鉴权性能优化，消除鉴权热点开销，提升查询性能 30% +	√	√	√	-
运维效率	自治索引	业界独有，支持分片自动调优、智能滚动、查询裁剪、故障自愈，实现索引全托管	√	√	-	-
	在线 Reindex	支持秒级完成分片数目的倍数调整	-	-	√	-

Elasticsearch 版本支持说明

最近更新时间：2026-05-08 11:54:21

腾讯云 ES 目前提供多个 Elasticsearch 版本供用户选择，并将不断跟进社区新版本。所有云上内核版本均保持与社区版本 100% 兼容，且在部分版本中提供腾讯云独有增强特性，为用户提供更强的性能、更丰富的能力与更高的稳定性。有关增强特性及其支持的版本，请参考 [腾讯云 ES 内核版本发布记录](#)。

腾讯云 ES 提供的 Elasticsearch 版本是由腾讯云独立维护的代码分支，统一由腾讯云 ES 团队独立维护和管理。为保障业务稳定性，请根据本文提供的版本生命周期规划您的升级节奏，确保及时升级至受支持的版本。

版本生命周期说明

腾讯云 ES 版本的生命周期包含以下阶段：

- **GA**: Generally Available，即正常商业化售卖中。
- **EOM**: End of Marketing，即停止销售，在 EOM 日期之后停止该版本的销售，仅提供必要的修复和更新。
- **EOS**: End of Service & Support，即停止服务，在 EOS 日期之后停止该版本的服务，不再提供该版本的任何技术服务支持。

腾讯云 ES 目前提供的版本生命周期如下，未在下表中列出的版本，将统一被视为 EOS 版本：

ES 版本	生命周期阶段	EOM 时间	EOS 时间
5.6.4	停止服务 (EOS)	2022.06.30	2024.06.30
6.4.3	停止服务 (EOS)	2023.09.30	2025.09.30
6.8.2	停止售卖 (EOM)	2025.12.31	2027.12.31
7.5.1	停止服务 (EOS)	2023.12.31	2025.12.31
7.10.1	停止售卖 (EOM)	2025.12.31	2027.12.31
7.14.2	商业化 (GA)	暂无计划	暂无计划
8.8.1	停止服务 (EOS)	2023.12.31	2025.12.31
8.11.3	停止售卖 (EOM)	2024.06.30	2026.06.30
8.13.3	停止售卖 (EOM)	2024.12.31	2026.12.31
8.16.1	商业化 (GA)	暂无计划	暂无计划
9.1.3	商业化 (GA)	暂无计划	暂无计划

升级与迁移指引

如您使用的 ES 版本已经或即将处于 EOM 或 EOS 阶段，建议您及时升级或迁移至正常商业化版本，详情请参考 [升级 ES 集群](#) 和 [迁移 ES 集群](#)。

日志增强版介绍

最近更新时间：2025-04-11 15:22:32

应用场景

日志分析，例如 Web 日志、App 日志、数据库日志、服务器指标、网络指标、存储指标等，这些场景的特点一般是写多读少、成本敏感。

版本特色

- 关键词：更高的写入性能，更低的成本。
- 采用存算分离、写入加速、查询/IO 并行化等亮点思想和技术，降低存储成本，提升写入性能。（推荐存算分离架构，也可以使用传统的热温架构，客户可根据实际业务按需选择）
- 在运维效率上，自治索引能实现索引创建、滚动、降冷、删除、故障自动修复的全自动化，并实现分片的智能调优，大幅减少了运维故障和资源投入。

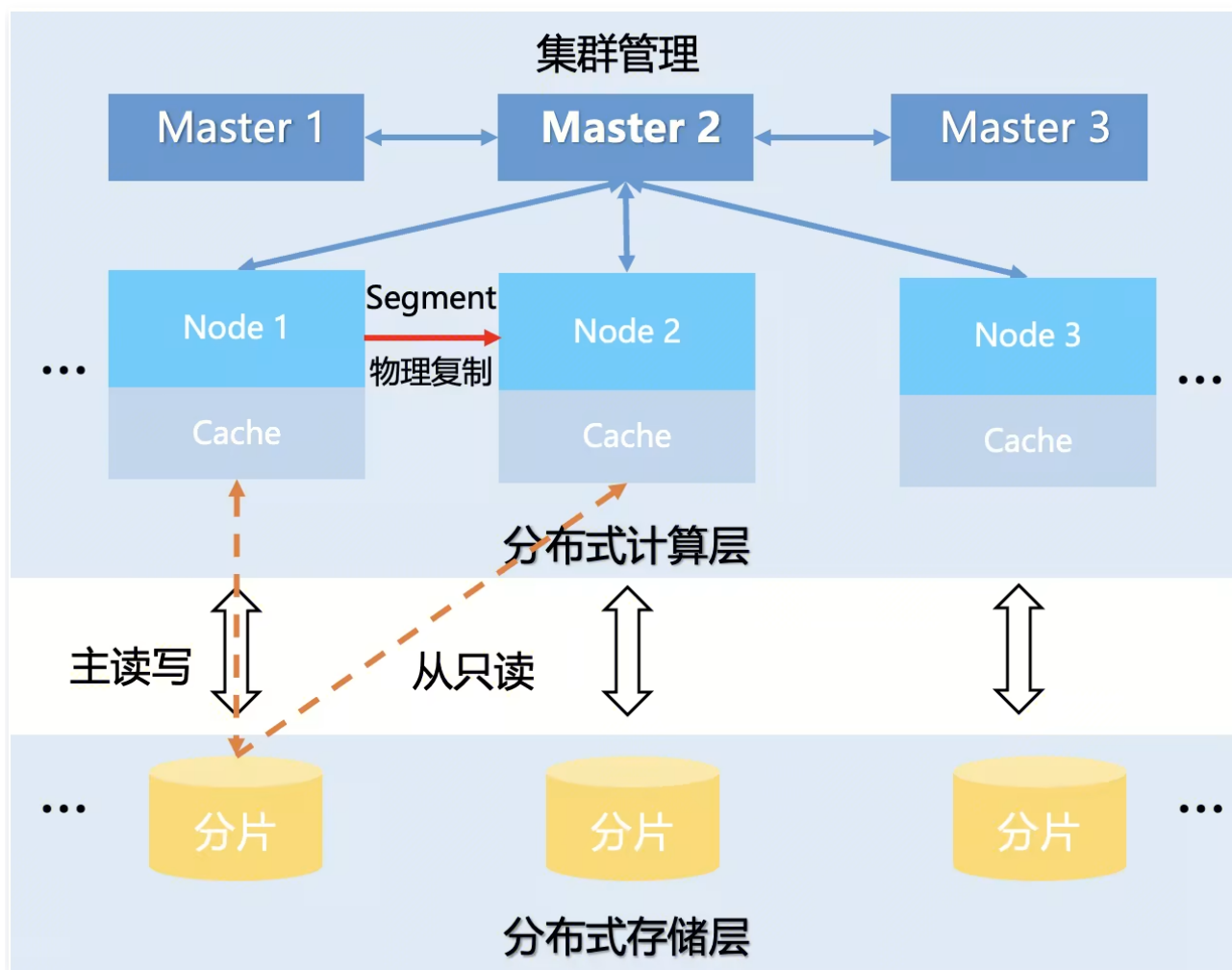
ⓘ 说明：解决的主要问题

- 在存储方面，用户为了保证稳定性一般会设置2 - 3个副本，如果使用云盘，云盘底层也会存储3个副本，在存储上有较大冗余。
- 在计算方面，由于 ES 的主副分片都会解析文档、分词、索引等，所以主副本之间也存在重复计算和重复写入的问题。
- 另外，原生架构存储与计算耦合，数据和计算在同一台节点，资源无法独立弹性扩缩容，在集群扩缩的时候会涉及大量的数据搬迁，不仅耗时比较长，也非常浪费资源。

关键特性

存算分离

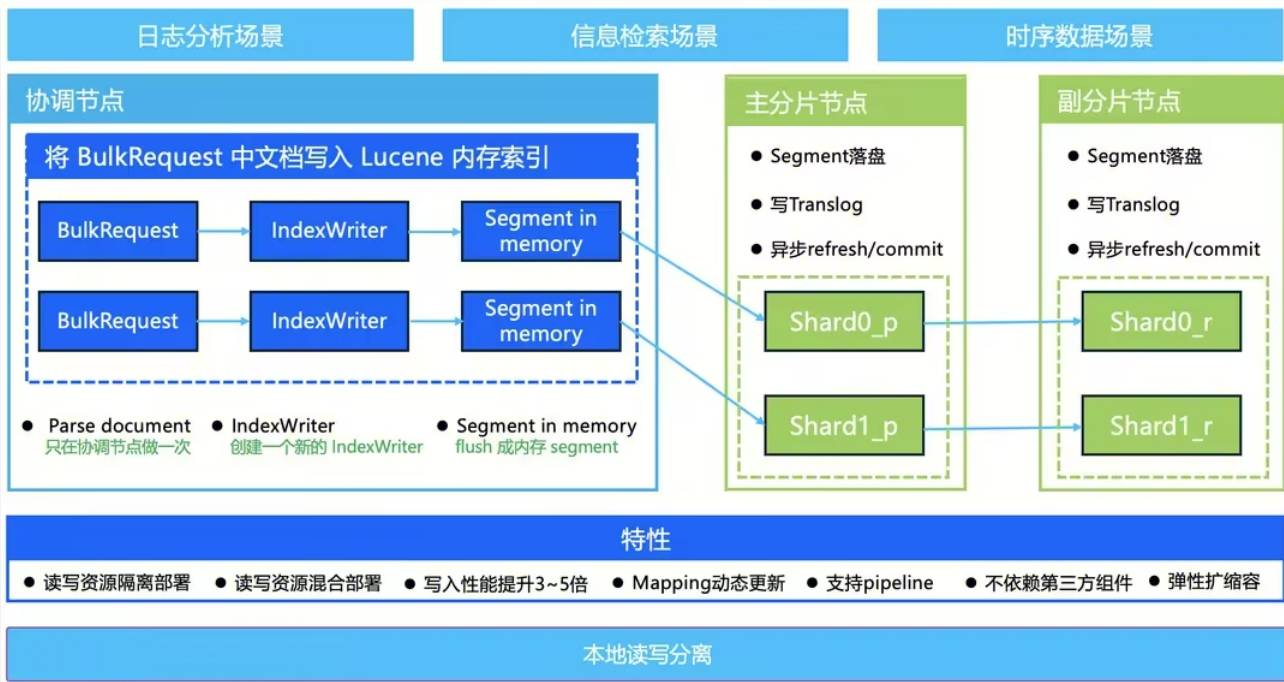
存算分离的核心思路是基于物理复制消除副本冗余计算，并确保主从副本分片完全一致，同时采用 delta + base 架构，本地 SSD 扛高并发写入并承载 Merge 计算开销，合并好较大的数据文件实时下沉至海量的对象存储，基于对象存储实现高可用，同时相比云盘大幅降低存储成本。缓存模块同时对高频访问数据进行缓存，降低对象存储的访问频次。针对对象存储和本地磁盘访问性能差异，采用 IO 并行化技术结合多级缓存实现冷热一体混合搜索能力。



具体特性说明请参见 [存算分离特性](#)。

写入加速

写入加速的核心思想是基于内存 segment 消除副本冗余计算、网络开销、锁的争抢等，ES 写入数据时最终是通过 Lucene 写入到内存中，一段时间后 refresh 成 segment。我们可以在协调节点提前通过 Lucene 的 API 构建好 segment 并保存在内存中，然后把内存 segment 转发给具体索引分片，索引分片收到内存 segment 后定时追加到 Lucene 中，通过 segment 内存生成和拷贝，内存 merge，自定义 merge policy、定向路由等亮点思想和技术，提升写入吞吐3 - 5倍。



这里的写入加速优化可以理解为本地读写分离，或单一集群读写分离架构，协调节点实现 segment 内存构建并物理拷贝至主从副本。集群可以按读写分区节点部署，也可以混合部署。

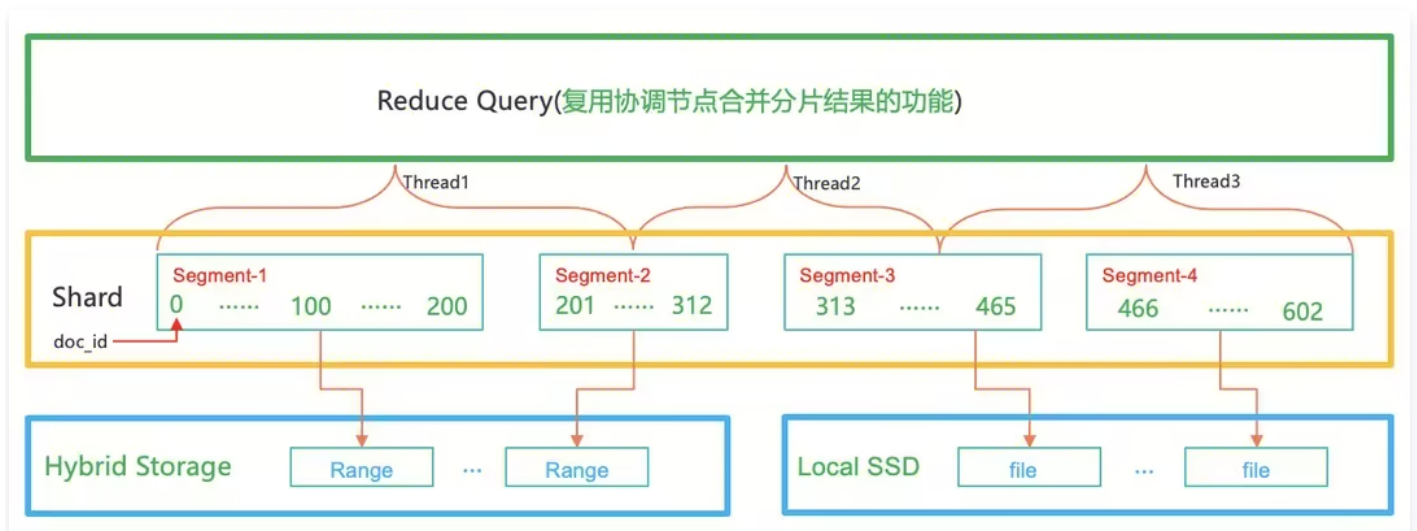
说明：

- **读写分区部署：**将集群中部分节点设置为专属协调节点，由专属协调节点提供写入计算能力，适合提升写入性能的同时更关心资源隔离的场景。
- **读写混合部署：**集群中每台节点都提供写入计算能力，对比分区部署的优点是有更多的节点提供写入计算能力，写入性能提升更多，适合只提升写入性能不关心资源隔离场景。

具体特性说明见 [写入加速特性](#) 说明文档。

查询/IO 并行化

ES 查询模型是将查询请求拆分成片级的子请求转发给各个分片执行，最后在协调节点合并各个分片的结果，在每个分片内部有多个 segment，默认情况下 ES 执行分片级查询时是单线程串行处理每个 segment 的，由于分片内的 segment 是独立的，可以尝试再拆分几个子请求，由多个线程并行处理，在数据节点合并多个线程的结果后再返回给协调节点，在数据节点合并每个线程的结果跟在协调节点合并每个分片的结果道理是相同的。该方案意在压榨空闲 CPU 资源，将 ES 的单个分片级请求拆分成3 - 5个子请求并行处理该分片下的 segment 或者 docs，根据 docs 或者 segment 切分，每个线程只处理一部分 docs 或者 segment，在数据节点合并每个线程的结果后再返回给协调节点，协调节点合并各个分片的结果返回给客户端，从而达到性能倍数级的提升。



效果测试

整体效果

腾讯云 ES 全新技术栈：采用存算分离、写入加速和查询/IO 并行化等先进技术，广泛应用于日志场景，实现冷热数据一体化搜索及弹性伸缩能力，成本可降低30% - 50%，最优情况下相比云盘可降低 80%。

- 存算分离：自研混合存储架构，实现冷热一体搜索，存储成本节约30% - 50%。
- 写入加速：无依赖、自闭环，读写资源隔离的同时提升 3+倍写入吞吐。
- 查询/IO 并行化：自研多级并行查询框架，支持全部查询场景，查询性能提升3 - 5倍。

性能压测

存算分离

压测环境：

- 集群：3台标准型 SA2 16核64G，1500GB SSD 云硬盘 x 1。
- 数据：http_logs。
- 工具：esrally。

压测结果：

从查询性能损耗可以看出，以本地盘为基准，Elastic 自研的 [可搜索快照](#) 查询性能损耗太大了，日志增强版自研的存算分离相较于本地盘查询性能损耗，仍在可接受的范围内，增加并行化压测后大部分场景比本地还要快2 - 3倍。

(查询延时单位：ms)

查询类型	本地 SSD 盘	可搜索快照	存算分离	存算分离 + 并行化
match_all (全部匹配)	3	59	3	4
term (单值精确匹配)	6	71	16	7

terms (多值精确匹配)	5	45	4	5
range (范围查询)	12	28	23	9
aggs (聚合查询)	1544	2575	2020	580
desc_sort_timestamp (按时间字段降序)	65	187	81	33
asc_sort_timestamp (按时间字段升序)	71	256	54	8
desc_sort_with_after_timestamp (降序排序中增加search_after)	1140	1863	1968	440
asc_sort_with_after_timestamp (升序排序中增加search_after)	936	1692	1389	395

写入加速

压测环境:

- 集群: 3台标准型 SA2 16核64G, 1500GB SSD 云硬盘 x 1。
- 数据: 随机生成。
- 工具: 编写代码消费 Kafka。

压测结果:

写入类型	副本	批次大小	写入性能	说明
ES 默认	1	5000	31w/s	默认 ES 的写入性能
写入加速	1	5000	119w/s	ES 的3.8倍

查询/IO 并行化

压测环境:

- 集群: 3台标准型 SA2 16核64G, 1500GB SSD 云硬盘 x 1。
- 数据: geonames。
- 工具: esrally。
- 并发度: IO 并行化并发度设置为3。

压测结果:

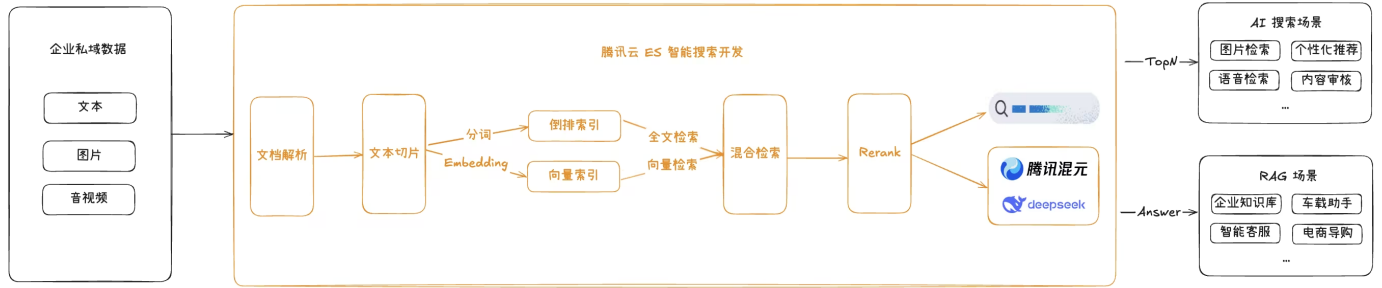
- IO 并行化并发度设置为3, 性能普遍提升3倍左右, 经过压测对比发现, P50, P90, P99耗时普遍减少5 - 10倍, 查询抖动更少更稳定。
- 如果 CPU 核数更多, 拆分的子请求可以更多, 性能会更好, 如果将并发度设置为5, 理论上性能会提升5倍左右。

类别	短语查询 (ms)	聚合查询 (ms)	脚本查询 (ms)	自定义打分 (ms)	排序(ms)
关闭并行化	43	43	374	408	143
打开并行化	5	4	132	125	44

智能搜索开发介绍

最近更新时间：2026-04-02 14:42:21

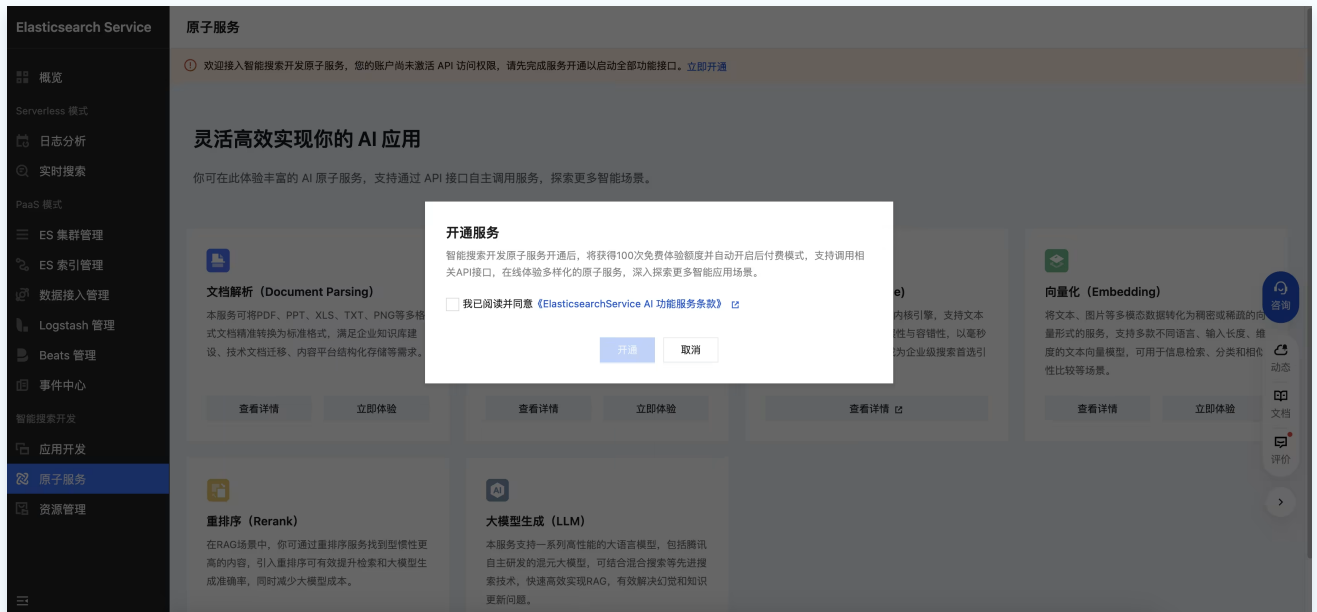
腾讯云 ES 发布的新特性智能搜索开发，进一步优化了对全文与向量混合搜索的能力支持，从原始文档解析、向量化等原子能力，到查询性能、混合排序效率、搜索结果精准度等提供了全方位的支持和优化，让搜索有了更多想象空间。在此基础上，还可以与混元、DeepSeek 等大语言模型无缝集成，从而帮助企业进一步高效、灵活地构建知识问答等 RAG 应用，在关键环节也可介入调优，整体流程如下：



原子服务

⚠️ 原子服务开通说明：

原子服务开通后才支持接口调用，请及时前往 [控制台](#) 开通服务，具体操作步骤：[单击立即开通](#)，确认免费开通。



目前智能搜索开发提供以下类型的原子服务：

类型	描述	服务	说明	接口
----	----	----	----	----

文档解析	本服务可将 PDF、DOC、PPT、XLS、HTML、TXT、PNG 等多格式文档转换为标准格式，满足企业级知识库建设、技术文档迁移、内容平台结构化存储等需求。	doc-llm	腾讯优图实验室推出的精品OCR大模型，覆盖论文、书籍、说明书、试卷、PPT、海报等众多场景，可满足各行业的文档处理需求。	<ul style="list-style-type: none"> 实时文档解析 ParseDocument 文档解析（异步） ParseDocumentAsync 获取文档解析结果 GetDocumentParseResult
文本切片	文本切片是将长文本分割为短片段的技术，用于适配模型输入、提升处理效率或信息检索，平衡片段长度与语义连贯性，适用于 NLP、数据分析等场景，本服务支持语义切分。	doc-chunk	基于分隔符、文本长度进行切片，适用规则性较强的文本，仅支持 Markdown与TXT格式文件输入。	在线文本切片 ChunkDocument
		doc-tree-chunk	腾讯自研文档拆分模型，支持多种文件类型，能够解析并深入理解图表中的信息，语义切分成文本块。	<ul style="list-style-type: none"> 文档切片（异步） ChunkDocumentAsync 获取文档切片结果 GetDocumentChunkResult
Embedding	将文本、图片等多模态数据转化为稠密或稀疏的向量形式的服务，支持多款不同语言、输入长度、维度的文本向量模型，可用于信息检索、分类和相似性比较等场景。	bge-base-zh-v1.5	768 维，最大 512 token 经典轻量级中文 Embedding 模型，适合资源有限、轻量级 RAG。	获取特征向量 GetTextEmbedding
		bge-large-zh-v1.5	1024 维，最大 512 token 经典轻量级中文 Embedding 模型，适合资源有限、轻量级 RAG。	
		KaLM-embedding-multilingual-mini-v1	896 维，最大 131072 token 微信自研多语言模型，支持多语言。	

		<p>bge-m3</p> <p>1024 维，最大 8194 token 经典 BGE Embedding 模型，支持 100+ 语言，适合多语言检索场景。</p>	
		<p>Conan-embedding-v1</p> <p>1792 维，最大 512 token 腾讯内部自研，曾登顶中文CMTEB榜单，适合高精度中文检索。</p>	
		<p>WeCLI Pv2-Base</p> <p>768维，最大 72 token 微信自研图文 Embedding 模型，在中文场景下表现优秀，适合文搜图、图搜图等场景</p>	<p>获取多模态特征向量 GetMultiModalEmbedding</p>
		<p>WeCLI Pv2-Large</p>	
重排序	<p>在 RAG 场景中，可通过 RAG 排序服务找到相关性更高的内容，引入排序服务可有效提升检索和大模型生成准确率，同时减少大模型成本。</p>	<p>bge-reranker-large</p> <p>单个 doc 最大 514 token 经典 BGE 重排序模型，通过交叉编码提升 Top-K结果相关性，支持中文、英文，适合检索结果精排后进行问答。</p>	<p>重排序 RunRerank</p>
		<p>bge-reranker-v2-m3</p> <p>单个 doc 最大 8192 token 升级版 BGE 重排模型，有更大的文本长度，同时支持多语言。</p>	
大模型生成	<p>本服务支持一系列高性能的大语言模型，包括 DeepSeek 系列和自主研发的混元大模型，通过深度优化和精确训练，目前已具备强大的自然语言处理能力和高效的文本理解能力，结合混合搜索等先进搜索</p>	<p>deepseek-r1</p> <p>最大输入 128k，最大输出 8k 擅长复杂需求拆解、技术方案直译，提供精准结构化分析及可落地方案，实现了与GPT-4o和Claude Sonnet 3.5等模型相媲美的性能。</p>	<p>大模型生成 ChatCompletions</p>

技术，快速高效实现 RAG，有效解决幻觉和知识更新问题。

<p>deepseek-v3 (v3.2)</p>	<p>最大输入 128k，最大输出 8k 通用型AI模型，拥有庞大参数规模及强大多任务泛化能力，擅长开放域对话、知识问答、创意生成等多样化需求。</p>
<p>deepseek-r1-distill-qwen-32b</p>	<p>最大输入 128k，最大输出 8k deepseek-r1 32b 参数蒸馏模型，使用成本更低，性价比较高。</p>
<p>hunyuan-turbo</p>	<p>最大输入 28k，最大输出 4k 腾讯新一代旗舰大模型，混元Turbo模型，在语言理解、文本创作、数学、推理和代码等领域都有较大提升，具备强大的知识问答能力。</p>
<p>hunyuan-large</p>	<p>最大输入 28k，最大输出 4k 腾讯开发的开源业界参数规模最大、效果最好的 transformer 结构的 MoE 模型，适用对模型效果、复杂指令有较高的要求的场景。</p>
<p>hunyuan-large-longcontext</p>	<p>最大输入 128k，最大输出 6k 擅长处理长文任务如文档摘要和文档问答等，同时也具备处理通用文本生成任务的能力。在长文本的分析和生成上表现优异，能有效应对复杂和详尽的长文内容处理需求。</p>
<p>hunyuan-</p>	<p>最大输入 30k，最大输出 2k</p>

		standard	在通用效果提升的基础上，训练数据中融合了医疗、金融领域的长文数据、长文翻译数据和长文多文档问答等高质量精标数据。	
		hunyuan-standard-256K	最大输入 250k，最大输出 6k 256k极长窗口特化版，复用7B-MoE框架压缩显存占用，支持百页文献级处理，适用绝大部分场景，同时兼顾效果及推理性能。	
联网搜索	整合 Sogou、Bing 以及 Baidu 联网搜索，涵盖网页、新闻、学术论文等多源数据，支持指定域名搜索，旨在提供高质量的公共互联网信息，突破本地数据局限。	sogou	本服务基于原搜狗网页 3.0 版本提供稳定的联网搜索服务，曾为全球首个网页收录量达到 100 亿的中文搜索引擎。	-
		bing	微软旗下多语言搜索引擎，实时新闻索引（1 小时内更新），学术资源整合能力强。	本服务为第三方服务厂商提供，受限于第三方服务， 目前已暂时关停 ，原服务自动切到sogou。
		baidu	全球最大中文搜索引擎，在字词结合的中文处理技术上具有深厚积累。	本服务为第三方服务厂商提供，受限于第三方服务， 目前已暂时关停 ，原服务自动切到sogou。

应用场景

- **文本搜索**：依托倒排索引与分词优化，腾讯云 ES 支持亿级数据毫秒级检索，具备拼写纠错、同义词扩展及高亮显示功能，广泛应用于电商商品搜索、新闻推荐等场景，精准匹配用户查询意图。
- **多模态搜索**：提供多模态数据向量化服务，结合腾讯云ES的混合搜索能力，支持文本、图像等混合数据等联合检索，如图搜图、文搜图、文搜视频等场景，突破单一模态限制。
- **知识库问答**：通过自然语言处理与语义分析，快速匹配用户问题与知识库内容，将用户问题与检索的内容交给大模型进行总结生成，适用于企业FAQ、客户机器人等场景，实现高效知识检索与精准答案推送。
- **AI 助手**：为 AI 助手提供实时知识库检索和对话上下文管理，通过搜索增强与多轮对话支持，提升虚拟助手、智能客服等应答能力。

优势特性

- 文本/向量混合搜索：独有文本与向量的混合搜索能力，支持多模态数据联合检索，通过倒排索引优化与向量算法扩展，实现复杂场景下的精准匹配与召回扩大。
- 开放的原子服务：提供文档解析、文本切片、Embedding、Rerank、LLM 生成以及联网搜索原子服务，集成 ES Inference API，降低 RAG、AI 搜索应用等开发门槛。
- 灵活搭建AI应用：高度灵活的写入/检索工作流自定义能力，支持调用内置原子服务和自定义服务，快速构建智能问答、AI 搜索等场景。
- 高性能与稳定性：千亿级向量规模支持，毫秒级响应延迟，自研存储优化节省 70%–90% 资源，结合分布式架构与多可用区容灾，保障高并发场景稳定运行。
- 权威认证：作为国内首个端到端 RAG 技术标准制定者和测评通过者（信通院认证），腾讯云 ES 过去一年已助力微信读书 AI 问书、ima、腾讯会议智能助手等顶流应用落地。

产品功能

最近更新时间：2024-09-26 15:55:31

用户可以将 CVM、TencentDB、容器服务等其他云产品的实时日志、业务的存量及增量业务数据，汇聚传输到 ES 集群，进行数据的分布式存储、查询分析。

数据采集与同步

- 用户通过 Elasticsearch 中的 Beats 功能，可以把数据传输到 Elasticsearch 中进行存储，也可以传输到 Logstash 中进行自定义转换和解析后，再传输到 Elasticsearch 中。
- Elasticsearch 提供了易用的 RESTful API，用户可以自行开发客户端，调用数据存储 API，存储数据到 Elasticsearch 集群中。
- ES 构建在 VPC 内，用户可以非常方便地使用各种数据同步插件，将已有云产品的数据，同步到 ES 集群中。

数据存储

- 腾讯云 ES 提供了多种规格的节点类型和高性能 SSD 磁盘，有效保障数据的读写性能。
- 支持弹性扩展到上百个节点，能达到 PB 级数据的存储，满足用户不同类型的业务场景。
- 支持故障节点探测及替换，保障集群高可用性。
- 具备全文检索、向量检索以及两者混合搜索功能。

数据查询分析可视化

- Elasticsearch 拥有全文检索、结构化搜索、数据过滤和指标统计等搜索功能，可应用于信息搜索和数据分析等多种场景。
- Elasticsearch 提供了简单易用的 RESTful API 以及各种语言的客户端，用户可以很方便地构建自己的搜索服务。
- 使用 Kibana，用户可以方便地在浏览器里对集群的数据进行搜索和统计分析。

高级特性(X-Pack)

最近更新时间：2024-10-08 14:18:21

简介

高级特性，是指 Elasticsearch 官方商业特性（原 X-Pack 商业版插件包含的特性），包含了安全（Security）、SQL、机器学习（Machine Learning）、监控（Monitor）等高级功能，可以为 Elasticsearch 服务的应用开发和运维管理，提供更有力的帮助。腾讯云 ES 已提供了包含高级特性的版本，您可以在创建购买集群时选择，下文介绍各版本详细功能。

购买指引



如上图所示，在腾讯云 ES 创建购买页，有高级特性版本选项。腾讯云 ES 提供了3种可选的高级特性版本，版本说明如下：

对比项	基础版	白金版	开源版
是否包含 X-Pack	✓	✓	×
X-Pack 功能完整度	部分	全部	无

购买推荐为了能够使用腾讯云 Elasticsearch 更多高级功能，我们建议您在创建购买集群时，选择白金版，各版本具体功能介绍及区别见下文，产品的定价信息详见 [Elasticsearch Service 集群定价](#)。

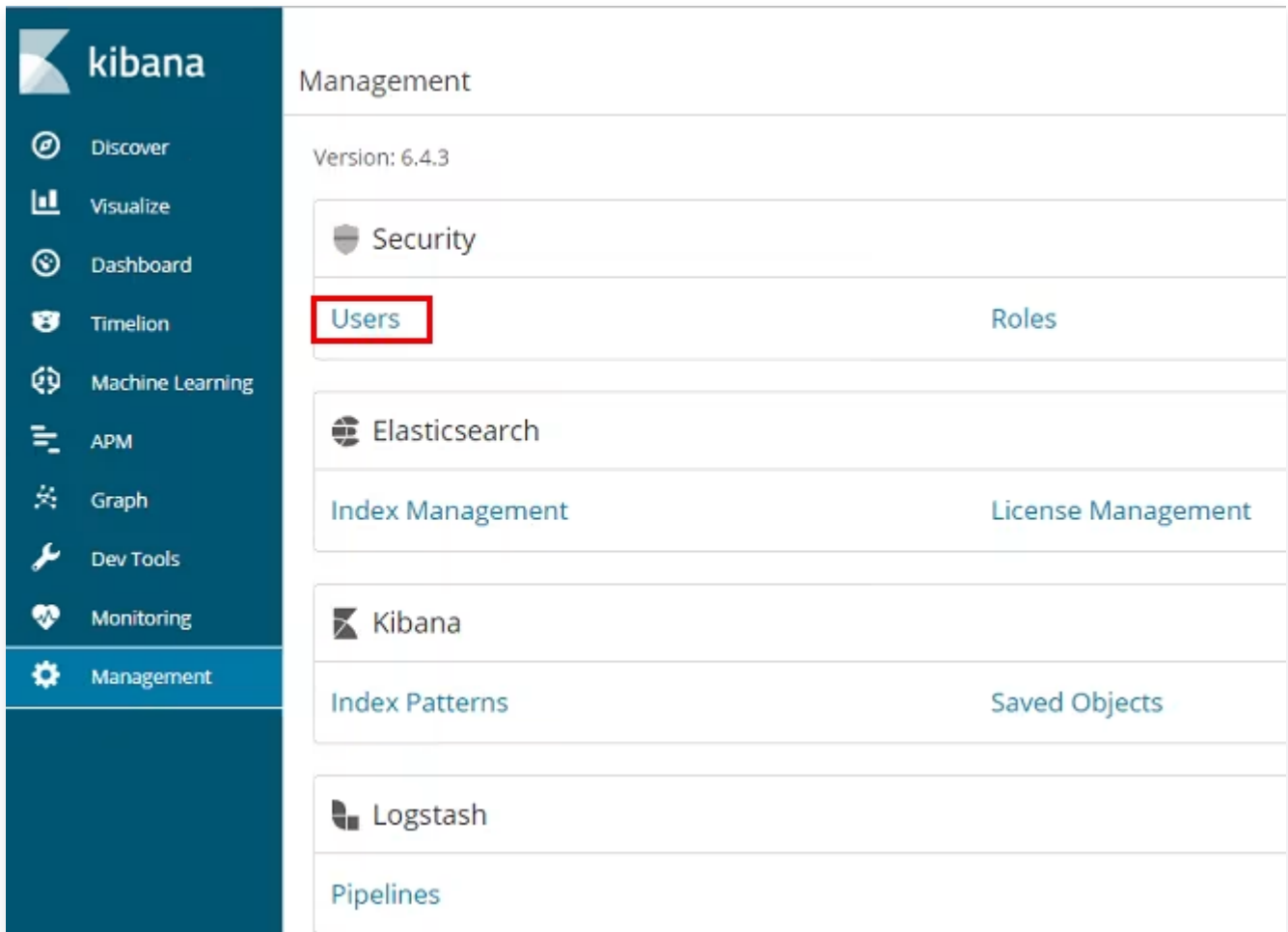
高级特性介绍

本文对部分常用高级特性进行了说明，完整的高级特性及说明，可查看官方说明 [Elastic Stack 订阅](#)、[API 文档](#)。

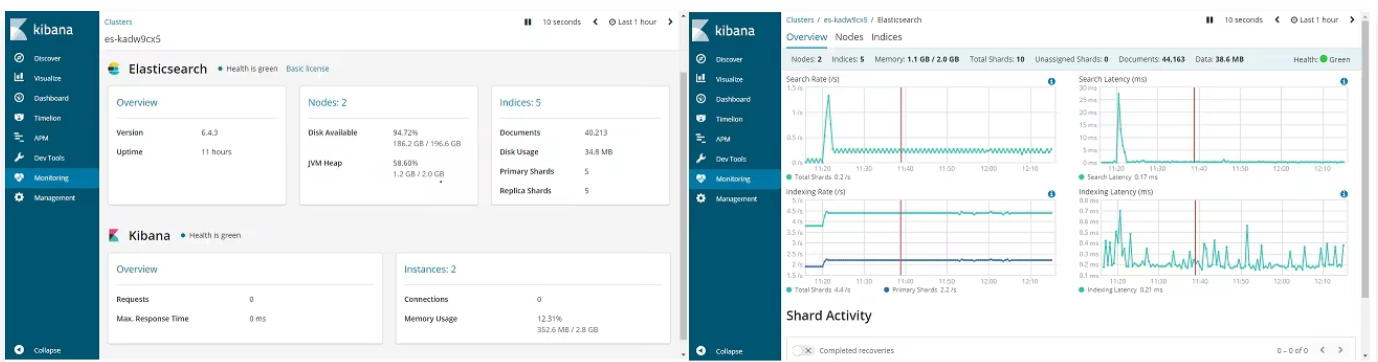
⚠ 注意

- 部分功能在不同的高级特性版本（基础、白金、开源）间有区别。
- 部分功能在较老的腾讯云 ES 版本不支持，若您遇到此问题，可通过 [售后支持](#) 咨询。

- 安全（Security）**支持索引和字段级别读写等细分权限的控制管理，实现数据安全防护、业务访问隔离，向正确的人员授予访问权限，阻止恶意破坏和数据泄露，有效的保障数据安全。



- **机器学习 (Machine Learning)** 在自定义数据告警的应用场景中，对于较难设置规则和阈值来定义的变化，可通过结合非监督型机器学习来预测数据的变化趋势和合理的波动范围，在数据偏离正常变化趋势时，发出告警通知。
- **监控集群、节点、索引多个维度**，全方位监控，实时了解集群运行情况，辅助应用开发及运维。



- **SQL** 提供了通过传统数据库 SQL 工具，实现对 Elasticsearch 数据进行全文检索、数据统计分析功能，支持 CLI、REST 等接入方式，**白金版还支持 JDBC 连接**。可以实现同原有业务系统的无缝对接，降低新技术

学习成本。

```

1 POST _xpack/sql?format=text
2 {
3   "query": "select * from website"
4 }

```

```

1      title
2      -----
3 My first blog post
4 My second blog post
5 My first blog post
6 My second blog post
7 My second blog post
8 My first blog post
9 My second blog post
10 My first blog post
11 My second blog post
12 My second blog post
13 My first blog post
14 My first blog post
15 My second blog post
16 My first blog post
17 My second blog post

```

说明

SQL 支持方面，开源版集成了其他的 SQL 插件，详细了解和使用可查看 [elasticsearch-sql](#)。

高级特性版本功能详细对比

本节主要对不同 ES 版本的部分重点功能做对比说明，方便用户了解不同版本中功能的区别。因 Elasticsearch 正处于快速发展阶段，不同版本对各功能的支持情况也在不断调整，所以我们不保证下述内容能及时跟进社区变化。了解最新准确的功能对比，可查看 Elasticsearch 官方介绍 [Elastic Stack 订阅](#)。

说明

下表中 、、 用于表示对应特性的功能完整度，：包含全部功能；：包含部分功能；：不包含。

模块	特性	开源版	基础版	白金版
Elasticsearch	可扩展性和弹性			
	查询和分析			
	数据扩充			
	管理和工具			
	Security			
	机器学习			

Kibana	探索和可视化	○	○	●
	堆栈管理和工具	○	○	●
	堆栈监测	—	○	●
	分享与合作	○	○	●
	Security	—	—	●
	机器学习	—	—	●
Beats	数据收集	○	○	●
	数据传输	○	○	●
	模块	○	○	●
	监测和管理	—	○	●
Logstash	数据收集	●	●	●
	数据扩充	●	●	●
	数据传输	●	●	●
	模块	○	●	●
	监测和管理	—	○	●
ELASTIC APM	APM 服务器	●	●	●
	APM 代理	●	●	●

	Kibana 中的 APM 仪表盘	●	●	●
	APM UI	—	●	●
	分布式跟踪	—	●	●
	Machine Learning 整合	—	—	●
ELASTIC 日志	日志采集器 (Filebeat)	●	●	●
	常用数据源的仪表盘	●	●	●
	Logs UI	—	●	●
ELASTIC 基础设施	指标采集器 (Metricbeat)	●	●	●
	常用数据源的仪表盘	●	●	●
	Infrastructure UI	—	●	●
ELASTIC 运行状态监控	运行状态监测 (Heartbeat)	●	●	●
	Kibana 里的运行状态仪表盘	●	●	●
	运行状态监测 UI	—	●	●

Elasticsearch 部分功能详细说明：

说明

下表中 ✓ 用于表示是否拥有对应特性的功能，✓：表示具备；-：表示不具备。

Elasticsearch 功能模块	细项	开源版	基础版	白金版
管理和工具	REST API	✓	✓	✓

	语言客户端	✓	✓	✓
	快照/恢复	✓	✓	✓
	_仅源快照	-	✓	✓
	SQL 解释器 CLI	-	✓	✓
	数据汇总	-	✓	✓
	索引生命周期管理	-	✓	✓
	冻结索引	-	✓	✓
	升级助手 API	-	✓	✓
	JDBC 客户端	-	-	✓
	ODBC 客户端	-	-	✓
Security	加密通信	-	✓	✓
	基于角色的访问控制	-	✓	✓
	文件和原生身份验证	-	✓	✓
	审核日志	-	-	-
	基于属性的权限控制	-	-	✓
	字段和文档级别安全性	-	-	✓
	LDAP 身份验证	-	-	✓
机器学习	时序型异常监测	-	-	✓
	输入/实体分析	-	-	✓
	日志消息分类	-	-	✓
	根本原因指示	-	-	✓
	异常情况警报	-	-	✓
	时序型预测	-	-	✓

产品优势

最近更新时间：2024-09-26 15:55:31

腾讯云 ES 提供了云端托管的服务，用户可容易地创建和管理 Elasticsearch 集群，并且能够保障生产环境中的高可用性。下文是对产品核心优势的介绍：

易于部署和管理

您只需通过简单的操作，即可在几分钟内创建一个 ES 集群，免去软硬件部署调试的复杂流程。此外，ES 还提供了便捷的集群运营管理工作、Kibana 页面以及完善的集群监控和告警系统，以满足客户日常的集群运营管理需求。

弹性扩缩容

ES 提供了多种类型的节点规格和存储介质，您可以根据业务需要进行选择。随着业务的发展，您可以动态调整集群的配置扩容或缩容，实时保障业务的发展需要并有效控制成本。

Elasticsearch 高级特性 (X-Pack)

集成 Security、SQL、Machine Learning 等 Elasticsearch 高级特性 (X-Pack)，提升了 Elasticsearch 集群的安全管控、操作使用和运维管理效率。

高可用

ES 提供了多可用区部署方案，可保证在单可用区网络、电力等不可抗力故障下不停服。COS 数据备份策略可定时备份数据，保障数据在意外情况下丢失时快速恢复。分散置放群组则是将集群的节点置放在不同底层硬件上，降低具有相同底层硬件的节点同时发生故障的风险。此外还有为保障集群稳定而进行的内核优化等策略，可以全方位地保障数据的安全和服务的稳定。

安全加固

通过部署在逻辑隔离的私有网络 VPC，客户可以完全掌控自己的环境配置，自定义网络访问控制列表 (Access Control List) 和安全组，提供了 Kibana 访问和 IP 访问黑白名单机制，高级特性 (X-Pack) 的 Security 能力提供了字段级的权限控制，切实保证您云上资源的安全性。

开放与服务集成

支持完整的 ELK 产品体系，兼容标准的开源 RESTful API 和生态组件，可以与对象存储 COS、网络流日志、消息队列、云数据库 TencentDB 等腾讯云产品集成，为用户提供数据传输和备份能力，以满足不同业务场景需要。

应用场景

最近更新时间：2024-09-26 15:55:31

日志分析

业务系统运行过程中，服务器、数据库和容器等会产生大量日志和监控数据，而且存储分散、种类繁多、规模庞大，很难进行检索和分析。ES 通过丰富的数据采集工具和分布式存储，方便日志统一管理和指标实时监控，通过一站式全观测优势，帮助用户快速定位问题，提升运维效率。

信息检索

ES 非常适合应用于网站搜索、移动应用搜索等场景，特别是针对大数据量、高并发以及对搜索灵活性和相关度要求比较高的情况，通过灵活的关键字、查询条件、模糊匹配等方式，可以从 PB 级的结构化和非结构化数据中毫秒级返回搜索结果。

向量检索

向量检索是一种基于向量空间模型的检索技术，它通过将文本、图像、视频等数据转换为数值向量，在向量空间进行相似性搜索，从而突破传统文本搜索只能基于关键字，不能基于语义搜索的限制。ES 提供从向量生成、到向量索引、存储、检索的一站式解决方案，帮助用户高效构建语义搜索、图片搜索、商品推荐等应用场景。

检索增强生成 RAG

大语言模型 LLM 在企业应用中遇到不少挑战，包括缺乏企业私域知识、幻觉和知识更新等，RAG 可以结合检索和生成技术，通过企业知识库的输入，提高 LLM 回答的准确性。ES 围绕 RAG 提供从数据切片、向量检索、文本和向量混合搜索、rerank、大模型集成等一站式服务，超越传统向量数据库的单点方案，助力企业轻松构建 AI 助手、知识问答等场景。

数据分析

在数据驱动运营的行业背景下，电子商务、移动应用、广告媒体等业务都需要借助数据分析和数据挖掘来辅助商业决策，而规模庞大的业务数据对数据的统计分析带来了很大的挑战。ES 拥有结构化查询的能力，支持复杂的过滤和聚合统计功能，帮助客户对海量数据进行高效地个性化统计分析、发现问题与机会、辅助商业决策，让数据产生真正的价值。

数据库查询加速

关系型数据库更偏向事务型查询，在海量数据规模的场景下，容易遇到查询性能不足、可扩展性差的挑战。ES 提供了弹性扩展及海量数据下的高并发低延时查询能力，通过数据同步工具保持与数据库同步，并支持 SQL 能力，满足客户数据库加速查询的需求，弥补了传统数据库的不足。

能力与限制说明

最近更新时间：2024-09-27 15:33:21

腾讯云 ES 是基于开源软件 Elasticsearch 开发的云上 PaaS 服务。通过腾讯云 ES 您可以快速搭建 Elasticsearch 集群服务，开发日志分析、数据搜索等应用。下面介绍产品能力和使用限制。

产品构成

腾讯云 ES 包含核心部件 Elasticsearch 集群和数据可视化分析工具 Kibana 两部分。数据采集和传输到 ES 集群的部分，您可以根据自己的业务需要，部署数据采集工具如 Beats、Logstash 或开发应用程序，将数据写入 ES 集群。

可选配置

SSD 云盘，单个节点，磁盘大小限制与机型规格有关。ES 版本为6.8及以上的，可购买的磁盘范围参考如下：

类型	机型规格	容量限制 (GB)
标准型实例族 (S1、SA5、SA2)	2核4G	20 - 2000
	2核8G	20 - 2000
	4核8G	20 - 4000
	4核16G	20 - 6000
	8核16G以上	20 - 30000
内存型实例族 (M1)	所有规格	20 - 30000

ⓘ 说明：

由于 ES 集群常以分布式多节点的形式构成，因此需要有主节点对集群进行统一的管理。为了防止可能的节点故障导致脑裂问题，建议您至少选择3个节点构建集群。配置选择可参见 [集群规格和容量配置评估](#)。

网络访问

VPC 内网访问

为了保证数据的安全，腾讯云 ES 构建在用户 VPC 内，也只能通过 VPC 访问集群进行数据的写入和查询。如果您需要通过公网访问 ES 集群，进行开发调试，可以通过 [VPN 连接](#) 的方式，打通本地 IDC 和云上 VPC，或通过 [外网代理的形式](#)。同时也提醒您做好数据安全防护。

Kibana 页面

支持通过公网访问 Kibana 页面，为了保证数据安全，Kibana 页面需要设置密码和访问 IP 黑白名单。

VPC 网络选择

腾讯云 ES 创建完成后，不支持切换 VPC 网络，请在创建集群时，提前规划业务部署。

相关概念

最近更新时间：2026-05-08 14:20:22

Elasticsearch 集群一般是由多个节点共同组成的分布式集群，节点之间互通，彼此配合，共同对外提供搜索和索引服务（节点之间能够将客户端请求转向到合适的节点）。不同的节点会负责不同的角色，有的负责一个，有的可能负责多个。Elasticsearch 中有多个节点角色，如数据节点、主节点、机器学习节点和协调节点。

数据节点（Data Node）

主要承担了数据存储和数据处理的工作，保存索引分片，处理与数据相关的操作，例如 CRUD、搜索、聚合等各种 I/O、内存和 CPU 密集型操作。集群使用过程中，需要注意监控数据节点的资源使用率，并在服务过载时通过添加更多的数据节点，进行集群扩容，以保障集群的稳定性。

主节点（Master Node）

负责轻量化整个集群范围内的操作，例如创建或删除索引、跟踪哪些节点是集群的一部分以及决定将哪些分片分配给哪些节点。对于集群健康来说，拥有一个稳定的主节点非常重要。

备选主节点（Master-eligible node）

指有资格被选为主节点的节点，任何符合主节点条件的节点（默认情况下为所有节点）都可以通过主选择过程被选为主节点。

默认情况下，所有节点都是数据节点，也都是备选主节点，对于小型集群来说非常方便。由于数据节点处理索引和搜索数据的请求都是 I/O、内存和 CPU 密集型工作，可能对节点的资源造成压力。随着集群的增长，为了确保主节点稳定且不受压力，保障集群的稳定，需将主节点和数据节点分离。

专用主节点（Dedicated Master Node）

指 Elasticsearch 集群中，设置了只能作为主节点的节点。

专用主节点配置建议

设置专用主节点主要是为了保障集群增大时的稳定性，建议专用主节点个数至少为3个。

- 专用主节点个数为1：只有1个备选主节点，`discovery.zen.minimum_master_nodes` 设置为1，网络发生故障时缺少备份。
- 专用主节点个数为2：有2个备选主节点，`minimum_master_nodes` 设置为1，有备份节点，但网络发生故障，在重新选主时，有发生脑裂的风险（每个备选主节点都将自己设置为主节点）；`minimum_master_nodes` 设置为2，发生故障时，可选主节点个数达不到要求，不能选出主节点。
- 专用主节点个数为3：有3个备选主节点，`discovery.zen.minimum_master_nodes` 设置为2，网络发生故障时，丢失一个可选主节点，在重新选主时，能有效选出主节点。

机器学习节点（Machine Learning Node）

机器学习节点用于创建机器学习任务，自动进行数据分析，识别异常数据，也可在此节点加载向量模型，提升向量生成和向量检索能力，同时可以与业务进行隔离，提升集群稳定性。

协调节点（Coordinating Node）

协调节点主要负责协调客户端的请求，如搜索请求或批量索引请求，将接收到的请求分发给合适的数据节点，每个数据节点在本地执行请求，并把结果汇集到协调节点。ES 每个节点都可以承担协调节点的角色，在高并发读写、多聚合查询等 CPU 密集型场景，设立独立的协调节点将有益于降低主节点和数据节点负担。

其他详细说明，可参考 [ES 节点说明](#)。