> 主讲老师：Fox
> 有道云笔记地址：https://note.youdao.com/s/626SjA9C

1. 什么是ElasticSearch

ElasticSearch（简称ES）是一个开源的分布式搜索和数据分析引擎，是用Java开发并且是当前最流行的开源的企业级搜索引擎，能够达到近实时搜索，它专门设计用于处理大规模的文本数据和实现高性能的全文检索。

2. ElasticSearch的优势

搜索引擎的排名：

☐ include secondary database models　　26 systems in ranking, July 2024

Rank Jul 2024	Rank Jun 2024	Rank Jul 2023	DBMS	Database Model	Score Jul 2024	Score Jun 2024	Score Jul 2023
1.	1.	1.	Elasticsearch	Search engine, Multi-model	130.82	-2.01	-8.77
2.	2.	2.	Splunk	Search engine	92.92	+3.82	+5.80
3.	3.	3.	Solr	Search engine, Multi-model	38.88	-2.15	-9.68
4.	4.	4.	OpenSearch	Search engine, Multi-model	16.64	+0.61	+3.77
5.	5.	↑ 8.	Sphinx	Search engine	5.96	+0.01	-0.12
6.	6.	↑ 7.	Microsoft Azure AI Search	Search engine, Multi-model	5.68	+0.15	-0.67
7.	↑ 8.	↓ 6.	Algolia	Search engine	5.57	+0.41	-2.36
8.	↓ 7.	↓ 5.	MarkLogic	Multi-model	4.42	-0.75	-4.17
9.	9.	9.	Virtuoso	Multi-model	3.98	-0.29	-0.91
10.	10.	10.	ArangoDB	Multi-model	3.42	+0.15	-1.15

> 参考网站：https://db-engines.com/en/ranking/search+engine

作为排名第一的搜索引擎，以下是一些 Elasticsearch 的优势：

- 分布式架构：Elasticsearch 采用分布式架构，可以轻松处理大规模数据，并支持水平扩展，提高系统的可扩展性和容错性。
- 全文检索功能：Elasticsearch 提供了强大的全文检索功能，可以对文本数据进行高效的搜索和分析，支持复杂的查询语法和自定义分析器。
- 多语言支持：Elasticsearch 支持多种语言的数据处理和检索，可以满足不同语言环境下的搜索需求。
- 高性能：Elasticsearch 采用了倒排索引等优化技术，能够实现高效的搜索和数据处理性能，满足大规模数据的实时查询需求。
- 实时性：Elasticsearch 提供近乎实时的搜索和分析功能，确保用户能够及时获取最新的数据和信息。
- 易用性：Elasticsearch 提供了丰富的 API 和插件，使得开发者可以轻松集成和使用，同时其查询语法简洁明了，易于上手。

> **官方网站:** https://www.elastic.co/
> **下载地址：**https://www.elastic.co/cn/downloads/past-releases#elasticsearch

3. Elastic Stack生态介绍

Elastic Stack由Logstash、Beats、Elasticsearch和Kibana四大核心产品组成，在数据采集、存储、分析及数据可视化方面有着无可比拟的优势。

- Elasticsearch

作为Elastic Stack的基石，Elasticsearch是一个高度可扩展的开源全文搜索与分析引擎。它利用分布式架构提供近乎实时的数据搜索、分析和可视化能力。Elasticsearch通过其强大的索引和查询功能，能够处理PB级的数据量，支持复杂的数据分析和挖掘需求，是构建现代数据驱动应用的理想选择。

- Logstash

Logstash是一个灵活的服务器端数据处理管道，能够同时从多个源采集数据，转换数据，然后将数据发送到您指定的目的地。它支持丰富的插件生态系统，使得数据收集、解析和转换过程变得高效且易于配置。Logstash在Elastic Stack中扮演着数据预处理和传输的关键角色，确保数据以正确的格式和结构进入Elasticsearch，为后续的分析和可视化提供坚实的基础。

- Beats

Beats是一个轻量级的数据采集器家族，专为发送数据到Logstash、Elasticsearch或Kafka等目的地而设计。每个Beat都是一个独立运行的守护进程，用于从系统或应用程序中收集数据，并将这些数据转发到指定的数据收集和处理系统中。Beats家族包括Filebeat（用于文件日志）、Metricbeat（用于系统和应用性能指标）、Heartbeat（用于监控服务可用性）等多个成员，它们共同构成了强大的边缘数据采集网络，覆盖了广泛的监控和日志收集需求。

- Kibana

Kibana是Elastic Stack的可视化和管理界面，为Elasticsearch数据提供了强大的可视化功能。通过Kibana，用户可以轻松创建仪表板、图表和地图，以直观的方式展示Elasticsearch中的数据。此外，Kibana还提供了交互式查询和过滤功能，使用户能够深入挖掘数据，发现隐藏的趋势和模式。作为Elastic Stack的用户界面，Kibana使得数据分析变得更加直观、便捷和高效。

Elastic Stack通过整合Elasticsearch、Logstash、Beats和Kibana这四大核心组件，帮助用户实现从数据收集、处理、存储到分析和可视化的一体化解决方案。这一方案不仅简化了数据处理的复杂性，还提高了数据处理的效率和准确性，是现代数据分析和监控领域不可或缺的强大工具。

4. ElasticSearch 应用场景

只要用到搜索的场景，Elasticsearch几乎都可以是最好的选择。结合Kibana、Logstash、Beats，ElaticSearch可以用于全文检索、日志分析、商业智能场景。

- **全文检索**

首先，Elasticsearch支持各类应用、网站等的全文搜索，包括淘宝、京东等电商平台的搜索，360手机助手、豌豆荚等应用市场平台的搜索，以及腾讯文档、石墨文档等平台的全文检索服务。

其次，Elasticsearch支持用户通过自定义打分、自定义排序、高亮等机制召回期望的结果数据，通过跨机房/跨机架感知、异地容灾等策略，为用户提供高可用、高并发、低延时、用户体验好的搜索服务。
许多知名企业，如阿里巴巴、腾讯、携程、滴滴出行、美团、字节跳动、贝壳找房等，都将Elasticsearch作为关键技术之一，以提升用户体验和满足业务需求。

- **日志分析**

Elasticsearch支持的日志包含但不限于如下类型：

a. 用户行为日志、应用日志等业务日志。
b. 慢查询、异常探测等状态日志。
c. Debug、Info、WARN、ERROR、FATAL等不同等级的系统日志。

基于倒排索引技术，Elasticsearch能够实现高效且灵活的搜索分析功能。从产生日志到生成相应的倒排索引并将其写入Elasticsearch，再到最终用户可以访问这些信息，整个过程所需时间仅为秒级。这确保了Elasticsearch能够快速处理和检索大量数据，满足实时搜索和分析的需求。
许多知名企业，如58集团、唯品会、日志易、国投瑞银等，都使用Elasticsearch来快速分析和处理大量的日志数据，从而对业务运行状况进行实时的监控和故障排查。

- **商业智能场景**

大型业务数据给电子商务、移动App开发、广告媒体等领域的企业的数据收集和数据分析带来了巨大的挑战。而Elasticsearch具有结构化查询功能，能实现全文数据检索和聚合分析，所以能有效帮助客户对上述大数据进行高效且个性化的分析，进而发现问题、辅助业务决策，并从数据中挖掘真正的商业价值。
许多知名企业的商业智能系统，如睿思BI、百度数据可视化Sugar BI、永洪BI等，都借助Elasticsearch的高效、实时的数据分析和可视化能力，帮助企业更好地理解市场趋势、优化决策过程。