

云搜索服务

# 快速入门

文档版本 01  
发布日期 2024-07-30



版权所有 © 华为云计算技术有限公司 2024。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

## 商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

## 注意

您购买的产品、服务或特性等应受华为云计算技术有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为云计算技术有限公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

# 华为云计算技术有限公司

地址：贵州省贵安新区黔中大道交兴功路华为云数据中心 邮编：550029

网址：<https://www.huaweicloud.com/>

---

## 目录

---

1 使用 Elasticsearch 搜索数据.....	1
2 使用 OpenSearch 搜索数据.....	10
3 使用 Logstash 迁移数据.....	18
4 入门实践.....	27

# 1 使用 Elasticsearch 搜索数据

本章节为您提供了一个通过Elasticsearch集群完成商品搜索的示例，您可以参考此示例来了解云搜索服务，包括创建索引、导入数据、搜索数据等操作。

## 场景描述

某女装品牌在网上经营电商业务，以前使用传统数据库来为用户提供商品搜索功能，但随着用户数量和业务的增长，使用传统数据库的弊端愈来愈明显。主要问题表现为响应速度慢、准确性低。为了改善用户体验从而避免用户流失，该电商网站开始使用云搜索服务为用户提供商品搜索功能，不仅解决了之前使用传统数据库产生的问题，而且实现了用户数量的增长。

假设该电商网站经营商品的数据如下所示：

```
{
  "products":[
    {"productName":"2017秋装新款文艺衬衫女装","size":"L"}
    {"productName":"2017秋装新款文艺衬衫女装","size":"M"}
    {"productName":"2017秋装新款文艺衬衫女装","size":"S"}
    {"productName":"2018春装新款牛仔裤女装","size":"M"}
    {"productName":"2018春装新款牛仔裤女装","size":"S"}
    {"productName":"2017春装新款休闲裤女装","size":"L"}
    {"productName":"2017春装新款休闲裤女装","size":"S"}
  ]
}
```

## 操作流程

以下是使用Elasticsearch集群完成数据搜索的操作步骤。

开始迁移数据前，请务必按[准备工作](#)指导完成必要操作。

- 步骤1：创建集群：** 创建一个Elasticsearch非安全集群用于搜索数据。
- 步骤2：导入数据：** 在Kibana上使用开源Elasticsearch API导入数据。
- 步骤3：搜索数据：** 在Elasticsearch集群中对数据进行全文检索和聚合结果显示。
- 步骤4：删除索引：** 当不再使用对应的索引数据时，可以删除对应索引节约资源。
- 步骤5：删除集群：** 无需继续使用集群时，可删除集群释放资源。

## 准备工作

已注册华为账号并开通华为云，进行了实名认证，且在使用云搜索服务前检查账号状态，账号不能处于欠费或冻结状态。

如果您还没有华为账号，请参考以下步骤创建。

1. 打开[华为云网站](#)。
2. 在页面右上角单击“注册”，根据提示信息完成注册。
3. 勾选服务条款，单击“开通”。
4. 实名认证，请参考：
  - 个人账号：[实名认证](#)
  - 企业账号：[实名认证](#)

## 步骤 1：创建集群

创建一个Elasticsearch非安全集群用于搜索数据。

1. 登录云搜索服务管理控制台。
2. 在左侧菜单栏选择“集群管理 > Elasticsearch”。
3. 单击右上角的“创建集群”，进入“创建集群”页面。
4. 配置集群的计费模式和可用区。

表 1-1 计费模型和可用区参数说明

参数	说明	取值样例
计费模式	集群支持“包年/包月”和“按需计费”两种模式。 <ul style="list-style-type: none"><li>● 包年/包月：根据集群购买时长，一次性支付集群费用。最短时长为1个月，最长时长为3年。如果购买时长超过9个月，建议包年购买，价格更优惠。一年计费为购买10个月得12个月。</li><li>● 按需计费：按实际使用时长计费，计费周期为一小时，不足一小时按一小时计费。</li></ul>	按需计费
当前区域	选择集群的所在区域。 不同区域的云服务产品之间内网互不相通。请就近选择靠近您业务的区域，可减少网络时延，提高访问速度。	华北-北京四
可用区	选择集群工作区域下关联的可用区。最多支持配置3个“可用区”。	可用区1

5. 配置集群基本信息。

图 1-1 配置集群信息

The screenshot shows a configuration interface with three main fields:
 

- 集群类型 (Cluster Type):** A button-based selector with 'Elasticsearch' highlighted in blue, and 'Logstash' and 'OpenSearch' as alternative options.
- 集群版本 (Cluster Version):** A dropdown menu currently showing '7.10.2'.
- 集群名称 (Cluster Name):** A text input field containing 'Sample-ESCluster'.

表 1-2 基本参数说明

参数	说明	取值样例
集群类型	选择“Elasticsearch”。	Elasticsearch
集群版本	选择所需的集群版本，支持的版本以界面可选项为准。	7.10.2
集群名称	自定义集群名称，可输入的字符范围为4~32个字符，只能包含数字、字母、中划线和下划线，且必须以字母开头。	Sample-ESCluster

6. 配置集群的规格信息。

图 1-2 配置集群规格

The screenshot displays a configuration interface for cluster specifications:
 

- 节点数量 (Node Count):** A numeric input field set to '1'.
- CPU架构 (CPU Architecture):** A button selector with 'x86计算' (X86 Computing) selected and '鲲鹏计算' (Kunpeng Computing) as an alternative.
- 节点规格 (Node Specification):** A list of instance types with a table below:
 

规格名称	vCPUs   内存	建议存储范围
ess.spec-4u8g (已售罄)	4 vCPUs   8 GB	40 GB - 1,800 GB
ess.spec-8u16g (已售罄)	8 vCPUs   16 GB	80 GB - 1,600 GB
ess.spec-16u32g (已售罄)	16 vCPUs   32 GB	100 GB - 3,200 GB
ess.spec-32u64g (已售罄)	32 vCPUs   64 GB	320 GB - 10,240 GB
- 节点存储 (Node Storage):** A button selector with '高I/O' (High I/O) selected.
- 节点存储容量 (Node Storage Capacity):** A numeric input field set to '40' GB.
- 节点类型 (Node Type):** Three checkboxes for '启用Master节点', '启用Client节点', and '启用冷数据节点', all of which are currently unchecked.

表 1-3 规格参数说明

参数	说明	取值样例
节点数量	设置集群中的数据节点个数。可选节点数为1~32。	1
CPU架构	支持“X86计算”和“鲲鹏计算”两种类型。具体支持的类型由实际区域环境决定。	X86计算
节点规格	选择集群节点的规格。	ess.spec-4u8g
节点存储	选择集群节点的存储类型。	高I/O

参数	说明	取值样例
节点存储容量	设置集群节点的存储空间大小，其取值范围与节点规格关联，不同的规格允许的取值范围不同。节点存储容量只支持配置为20的倍数。	40GB
启用Master节点	Master节点负责管理集群中所有节点任务。	不勾选
启用Client节点	Client节点负责接收并协调外部请求，如search和write请求	不勾选
启用冷数据节点	冷数据节点用于存储对查询时延要求不高，但数据量较大的历史数据。	不勾选

7. 设置集群的企业项目。

如果您开通了“企业项目”，在创建集群时，可以给集群绑定一个企业项目。本示例选择默认企业项目“default”。

8. 单击“下一步：网络配置”，设置集群的网络配置。

图 1-3 配置网络



表 1-4 网络配置参数说明

参数	说明	取值样例
虚拟私有云	指定集群节点使用的虚拟专用网络，实现不同业务的网络隔离。 <b>说明</b> 此处您选择的VPC必须包含网段（CIDR），否则集群将无法创建成功。新建的VPC默认包含网段（CIDR）。	vpc-default
子网	集群使用子网实现与其他网络的隔离，并独享所有网络资源，以提高网络安全。	subnet-default
安全组	安全组起着虚拟防火墙的作用，为集群提供安全的网络访问控制策略。 <b>说明</b> 为了确保您能够正常访问集群，需要放通安全组9200规则。	default

参数	说明	取值样例
安全模式	集群支持选择是否开启安全模式，开启之后将对集群进行通讯加密和安全认证。	关闭

- 单击“下一步：高级配置”，可以设置集群自动快照和其他高级功能。  
本集群仅做入门指导使用，无需开启集群快照和高级功能。
- 单击“下一步：确认配置”，确认完成后单击“立即创建”开始创建集群。
- 单击“返回集群列表”，系统将跳转到“集群管理”页面。您创建的集群将展现在集群列表中，且集群状态为“创建中”，创建成功后集群状态会变为“可用”。

图 1-4 创建集群

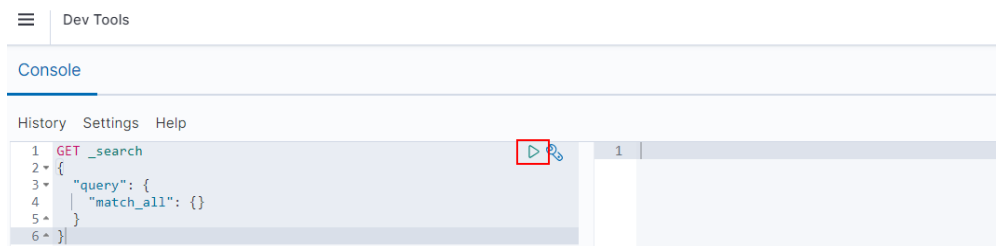
名称ID	集群状态	任务状态	版本	创建时间	企业项目	内网访问地址	计费模式	操作
css-055b	可用	-	7.10.2 elasticsearch	2024/03/15 15:24...	default	192.168.0.153.92...	按需计费	Kibana 监控信息 更多

## 步骤 2：导入数据

Elasticsearch 集群支持通过多种方式导入数据，本示例选择在 Kibana 使用开源 Elasticsearch API 导入数据。

- 在 Elasticsearch 集群管理页面，选择已创建的“Sample-ESCluster”集群，单击操作列中的“Kibana”进入 Kibana 控制台。
- 在 Kibana 的左侧导航中选择“Dev Tools”，进入 Console 界面。  
Console 左侧区域为输入框，输入框右侧的三角形为执行命令按钮，Console 右侧为结果输出区域。

图 1-5 Console 界面



- 在 Console 界面，执行如下命令创建索引“my\_store”。

```
PUT /my_store
{
  "settings": {
    "number_of_shards": 1
  },
  "mappings": {
    "properties": {
      "productName": {
        "type": "text",
        "analyzer": "ik_smart"
      },
      "size": {
        "type": "keyword"
      }
    }
  }
}
```



```
}  
}
```

返回结果如下所示。

```
{  
  "acknowledged" : true,  
  "shards_acknowledged" : true,  
  "index" : "my_store"  
}
```

4. 在Console界面，执行如下命令，将数据导入到“my\_store”索引中。

```
POST /my_store/_doc/_bulk  
{ "index": {} }  
{ "productName": "2017秋装新款文艺衬衫女装", "size": "L" }  
{ "index": {} }  
{ "productName": "2017秋装新款文艺衬衫女装", "size": "M" }  
{ "index": {} }  
{ "productName": "2017秋装新款文艺衬衫女装", "size": "S" }  
{ "index": {} }  
{ "productName": "2018春装新款牛仔裤女装", "size": "M" }  
{ "index": {} }  
{ "productName": "2018春装新款牛仔裤女装", "size": "S" }  
{ "index": {} }  
{ "productName": "2017春装新款休闲裤女装", "size": "L" }  
{ "index": {} }  
{ "productName": "2017春装新款休闲裤女装", "size": "S" }
```

当返回结果信息中“errors”字段的值为“false”时，表示导入数据成功。

### 步骤 3：搜索数据

在Elasticsearch集群中对数据进行全文检索和聚合结果显示。

- **全文检索**

假设用户进入该电商网站，她想要查找名称包含“春装牛仔裤”的商品信息，可以搜索“春装牛仔裤”。

在Kibana执行如下搜索命令。

```
GET /my_store/_search  
{  
  "query": { "match": {  
    "productName": "春装牛仔裤"  
  } }  
}
```

返回结果如下所示。

```
{  
  "took" : 3,  
  "timed_out" : false,  
  "_shards" : {  
    "total" : 1,  
    "successful" : 1,  
    "skipped" : 0,  
    "failed" : 0  
  },  
  "hits" : {  
    "total" : {  
      "value" : 4,  
      "relation" : "eq"  
    },  
    "max_score" : 1.7965372,  
    "hits" : [  
      {  
        "_index" : "my_store",  
        "_type" : "_doc",  
        "_id" : "9xf6VHIBfClT6SDjw7H5",  
        "_score" : 1.7965372,  
        "_source" : {
```

```
    "productName": "2018春装新款牛仔裤女装",
    "size": "M"
  },
  {
    "_index": "my_store",
    "_type": "_doc",
    "_id": "-Bf6VHIBfClT6SDjw7H5",
    "_score": 1.7965372,
    "_source": {
      "productName": "2018春装新款牛仔裤女装",
      "size": "S"
    }
  },
  {
    "_index": "my_store",
    "_type": "_doc",
    "_id": "-Rf6VHIBfClT6SDjw7H5",
    "_score": 0.5945667,
    "_source": {
      "productName": "2017春装新款休闲裤女装",
      "size": "L"
    }
  },
  {
    "_index": "my_store",
    "_type": "_doc",
    "_id": "-hf6VHIBfClT6SDjw7H5",
    "_score": 0.5945667,
    "_source": {
      "productName": "2017春装新款休闲裤女装",
      "size": "S"
    }
  }
]
}
```

- Elasticsearch支持IK分词，上面执行命令会将“春装牛仔裤”分词为“春装”和“牛仔裤”。
- Elasticsearch支持全文检索，上面执行命令会在所有商品信息中搜索包含“春装”或“牛仔裤”的商品信息。
- Elasticsearch与传统数据库不同，它能借助倒排索引在毫秒级返回结果。
- Elasticsearch支持评分排序，在上面返回结果中，前两条商品信息中同时出现了“春装”和“牛仔裤”，后两条商品信息中只出现了“春装”，所以前两条比后两条与检索关键词的匹配度更高，分数更高，排序也更靠前。

- **聚合结果显示**

该电商网站可以提供聚合结果显示功能，例如：对“春装”对应的产品按照尺码分类，统计不同尺码的数量。

在Kibana执行如下聚合结果命令。

```
GET /my_store/_search
{
  "query": {
    "match": { "productName": "春装" }
  },
  "size": 0,
  "aggs": {
    "sizes": {
      "terms": { "field": "size" }
    }
  }
}
```

返回结果如下所示。

```
{
  "took" : 3,
  "timed_out" : false,
  "_shards" : {
    "total" : 1,
    "successful" : 1,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 4,
      "relation" : "eq"
    },
    "max_score" : null,
    "hits" : [ ]
  },
  "aggregations" : {
    "sizes" : {
      "doc_count_error_upper_bound" : 0,
      "sum_other_doc_count" : 0,
      "buckets" : [
        {
          "key" : "S",
          "doc_count" : 2
        },
        {
          "key" : "L",
          "doc_count" : 1
        },
        {
          "key" : "M",
          "doc_count" : 1
        }
      ]
    }
  }
}
```

## 步骤 4：删除索引

当不再使用对应的索引数据时，可以在 Kibana 执行如下命令删除对应索引，避免造成资源浪费。

```
DELETE /my_store
```

返回结果如下所示。

```
{
  "acknowledged" : true
}
```

## 步骤 5：删除集群

已完成数据搜索业务，无需继续使用集群时，可删除集群释放资源。

### 📖 说明

由于集群删除后，数据无法恢复，请谨慎操作。

1. 登录云搜索服务管理控制台。
2. 在左侧菜单栏选择“集群管理 > Elasticsearch”。
3. 在集群列表，选择“Sample-ESCluster”集群，在操作列单击“更多” > “删除”。

4. 在弹出的确认对话框中，确认要删除的集群名称，单击“确定”完成操作。

# 2 使用 OpenSearch 搜索数据

本章节为您提供了一个通过OpenSearch集群完成商品搜索的示例，您可以参考此示例来了解云搜索服务，包括创建索引、导入数据、搜索数据等操作。

## 场景描述

某女装品牌在网上经营电商业务，以前使用传统数据库来为用户提供商品搜索功能，但随着用户数量和业务的增长，使用传统数据库的弊端愈来愈明显。主要问题表现为响应速度慢、准确性低。为了改善用户体验从而避免用户流失，该电商网站开始使用云搜索服务为用户提供商品搜索功能，不仅解决了之前使用传统数据库产生的问题，而且实现了用户数量的增长。

假设该电商网站经营商品的数据如下所示：

```
{
  "products":[
    {"productName":"2017秋装新款文艺衬衫女装","size":"L"}
    {"productName":"2017秋装新款文艺衬衫女装","size":"M"}
    {"productName":"2017秋装新款文艺衬衫女装","size":"S"}
    {"productName":"2018春装新款牛仔裤女装","size":"M"}
    {"productName":"2018春装新款牛仔裤女装","size":"S"}
    {"productName":"2017春装新款休闲裤女装","size":"L"}
    {"productName":"2017春装新款休闲裤女装","size":"S"}
  ]
}
```

## 操作流程

以下是使用OpenSearch集群完成数据搜索的操作步骤。

开始迁移数据前，请务必按[准备工作](#)指导完成必要操作。

1. **步骤1：创建集群**：创建一个OpenSearch非安全集群用于搜索数据。
2. **步骤2：导入数据**：在OpenSearch Dashboards上使用开源Elasticsearch API导入数据。
3. **步骤3：搜索数据**：在OpenSearch集群中对数据进行全文检索和聚合结果显示。
4. **步骤4：删除索引**：当不再使用对应的索引数据时，可以删除对应索引节约资源。
5. **步骤5：删除集群**：无需继续使用集群时，可删除集群释放资源。

## 准备工作

已注册华为账号并开通华为云，进行了实名认证，且在使用云搜索服务前检查账号状态，账号不能处于欠费或冻结状态。

如果您还没有华为账号，请参考以下步骤创建。

1. 打开[华为云网站](#)。
2. 在页面右上角单击“注册”，根据提示信息完成注册。
3. 勾选服务条款，单击“开通”。
4. 实名认证，请参考：
  - 个人账号：[实名认证](#)
  - 企业账号：[实名认证](#)

## 步骤 1：创建集群

创建一个OpenSearch非安全集群用于搜索数据。

1. 登录云搜索服务管理控制台。
2. 在左侧菜单栏选择“集群管理 > OpenSearch”。
3. 单击右上角的“创建集群”，进入“创建集群”页面。
4. 配置集群的计费模式和可用区。

表 2-1 计费模型和可用区参数说明

参数	说明	取值样例
计费模式	集群支持“包年/包月”和“按需计费”两种模式。 <ul style="list-style-type: none"><li>● 包年/包月：根据集群购买时长，一次性支付集群费用。最短时长为1个月，最长时长为3年。如果购买时长超过9个月，建议包年购买，价格更优惠。一年计费为购买10个月得12个月。</li><li>● 按需计费：按实际使用时长计费，计费周期为一小时，不足一小时按一小时计费。</li></ul>	按需计费
当前区域	选择集群的所在区域。 不同区域的云服务产品之间内网互不相通。请就近选择靠近您业务的区域，可减少网络时延，提高访问速度。	华北-北京四
可用区	选择集群工作区域下关联的可用区。最多支持配置3个“可用区”。	可用区1

5. 配置集群基本信息。

图 2-1 配置集群信息

表 2-2 基本参数说明

参数	说明	取值样例
集群类型	选择“OpenSearch”。	OpenSearch
集群版本	选择所需的集群版本，支持的版本以界面可选项为准。	1.3.6
集群名称	自定义集群名称，可输入的字符范围为4~32个字符，只能包含数字、字母、中划线和下划线，且必须以字母开头。	Sample-OSCluster

6. 配置集群的规格信息。

图 2-2 配置集群规格

规格名称	vCPUs   内存	建议存储范围
ess.spec-4u8g (已售罄)	4 vCPUs   8 GB	40 GB - 1,800 GB
ess.spec-8u16g (已售罄)	8 vCPUs   16 GB	80 GB - 1,600 GB
ess.spec-16u32g (已售罄)	16 vCPUs   32 GB	100 GB - 3,200 GB
ess.spec-32u64g (已售罄)	32 vCPUs   64 GB	320 GB - 10,240 GB

表 2-3 规格参数说明

参数	说明	取值样例
节点数量	集群中的节点个数。可选节点数为1~32。	1
CPU架构	支持“X86计算”和“鲲鹏计算”两种类型。具体支持的类型由实际区域环境决定。	X86计算
节点规格	选择集群节点的规格。	ess.spec-4u8g
节点存储	选择集群节点的存储类型。	普通I/O

参数	说明	取值样例
节点存储容量	设置集群节点的存储空间大小，其取值范围与节点规格关联，不同的规格允许的取值范围不同。节点存储容量只支持配置为20的倍数。	40GB
启用Master节点	Master节点负责管理集群中所有节点任务。	不勾选
启用Client节点	Client节点负责接收并协调外部请求，如search和write请求	不勾选
启用冷数据节点	冷数据节点用于存储对查询时延要求不高，但数据量较大的历史数据。	不勾选

7. 设置集群的企业项目。

如果您开通了“企业项目”，在创建集群时，可以给集群绑定一个企业项目。本示例选择默认企业项目“default”。

8. 单击“下一步：网络配置”，设置集群的网络配置。

图 2-3 配置网络



表 2-4 网络配置参数说明

参数	说明	取值样例
虚拟私有云	指定集群节点使用的虚拟专用网络，实现不同业务的网络隔离。 <b>说明</b> 此处您选择的VPC必须包含网段（CIDR），否则集群将无法创建成功。新建的VPC默认包含网段（CIDR）。	vpc-default
子网	集群使用子网实现与其他网络的隔离，并独享所有网络资源，以提高网络安全。	subnet-default
安全组	安全组起着虚拟防火墙的作用，为集群提供安全的网络访问控制策略。 <b>说明</b> 为了确保您能够正常访问集群，需要放通安全组9200规则。	default



参数	说明	取值样例
安全模式	集群支持选择是否开启安全模式，开启之后将对集群进行通讯加密和安全认证。	关闭

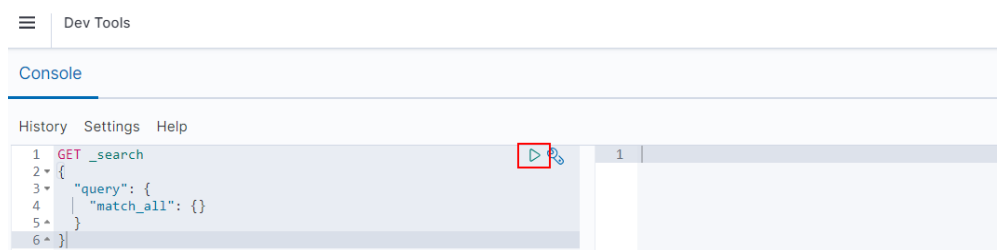
- 单击“下一步：高级配置”，可以设置集群自动快照和其他高级功能。  
本集群仅做入门指导使用，无需开启集群快照和高级功能。
- 单击“下一步：确认配置”，确认完成后单击“立即创建”开始创建集群。
- 单击“返回集群列表”，系统将跳转到“集群管理”页面。您创建的集群将展现在集群列表中，且集群状态为“创建中”，创建成功后集群状态会变为“可用”。

## 步骤 2：导入数据

OpenSearch 集群支持通过多种方式导入数据，本示例选择在 OpenSearch Dashboards 上使用开源 Elasticsearch API 导入数据。

- 在 OpenSearch 集群管理页面，选择已创建的“Sample-OSCluster”集群，单击操作列中的“Kibana”进入 OpenSearch Dashboards 控制台。
- 在 OpenSearch Dashboards 的左侧导航中选择“Dev Tools”，进入 Console 界面。  
Console 左侧区域为输入框，输入框右侧的三角形为执行命令按钮，Console 右侧为结果输出区域。

图 2-4 Console 界面



- 在 Console 界面，执行如下命令创建索引“my\_store”。

```
PUT /my_store
{
  "settings": {
    "number_of_shards": 1
  },
  "mappings": {
    "properties": {
      "productName": {
        "type": "text",
        "analyzer": "ik_smart"
      },
      "size": {
        "type": "keyword"
      }
    }
  }
}
```

返回结果如下所示。

```
{
  "acknowledged": true,
```

```
"shards_acknowledged" : true,  
"index" : "my_store"  
}
```

4. 在Console界面，执行如下命令，将数据导入到“my\_store”索引中。

```
POST /my_store/_doc/_bulk  
{  
  "index":{}  
  {"productName":"2017秋装新款文艺衬衫女装","size":"L"}  
  {"index":{}  
  {"productName":"2017秋装新款文艺衬衫女装","size":"M"}  
  {"index":{}  
  {"productName":"2017秋装新款文艺衬衫女装","size":"S"}  
  {"index":{}  
  {"productName":"2018春装新款牛仔裤女装","size":"M"}  
  {"index":{}  
  {"productName":"2018春装新款牛仔裤女装","size":"S"}  
  {"index":{}  
  {"productName":"2017春装新款休闲裤女装","size":"L"}  
  {"index":{}  
  {"productName":"2017春装新款休闲裤女装","size":"S"}  
}
```

当返回结果信息中“errors”字段的值为“false”时，表示导入数据成功。

### 步骤 3：搜索数据

在OpenSearch集群中对数据进行全文检索和聚合结果显示。

- **全文检索**

假设用户进入该电商网站，想要查找名称包含“春装牛仔裤”的商品信息，可以搜索“春装牛仔裤”。

在OpenSearch Dashboards执行如下搜索命令。

```
GET /my_store/_search  
{  
  "query": {"match": {  
    "productName": "春装牛仔裤"  
  }}  
}
```

返回结果如下所示。

```
{  
  "took" : 3,  
  "timed_out" : false,  
  "_shards" : {  
    "total" : 1,  
    "successful" : 1,  
    "skipped" : 0,  
    "failed" : 0  
  },  
  "hits" : {  
    "total" : {  
      "value" : 4,  
      "relation" : "eq"  
    },  
    "max_score" : 1.7965372,  
    "hits" : [  
      {  
        "_index" : "my_store",  
        "_type" : "_doc",  
        "_id" : "9xf6VHIBfClT6SDjw7H5",  
        "_score" : 1.7965372,  
        "_source" : {  
          "productName" : "2018春装新款牛仔裤女装",  
          "size" : "M"  
        }  
      },  
      {  
        "_index" : "my_store",
```

```
{
  "_type": "_doc",
  "_id": "-Bf6VHIBfClT6SDJw7H5",
  "_score": 1.7965372,
  "_source": {
    "productName": "2018春装新款牛仔裤女装",
    "size": "S"
  }
},
{
  "_index": "my_store",
  "_type": "_doc",
  "_id": "-Rf6VHIBfClT6SDJw7H5",
  "_score": 0.5945667,
  "_source": {
    "productName": "2017春装新款休闲裤女装",
    "size": "L"
  }
},
{
  "_index": "my_store",
  "_type": "_doc",
  "_id": "-hf6VHIBfClT6SDJw7H5",
  "_score": 0.5945667,
  "_source": {
    "productName": "2017春装新款休闲裤女装",
    "size": "S"
  }
}
]
```

- OpenSearch支持IK分词，上面执行命令会将“春装牛仔裤”分词为“春装”和“牛仔裤”。
  - OpenSearch支持全文检索，上面执行命令会在所有商品信息中搜索包含“春装”或“牛仔裤”的商品信息。
  - OpenSearch与传统数据库不同，它能借助倒排索引在毫秒级返回结果。
  - OpenSearch支持评分排序，在上面返回结果中，前两条商品信息中同时出现了“春装”和“牛仔裤”，后两条商品信息中只出现了“春装”，所以前两条比后两条与检索关键词的匹配度更高，分数更高，排序也更靠前。
- **聚合结果显示**

该电商网站可以提供聚合结果显示功能，例如：对“春装”对应的产品按照尺码分类，统计不同尺码的数量。

在OpenSearch Dashboards执行如下聚合结果命令。

```
GET /my_store/_search
{
  "query": {
    "match": { "productName": "春装" }
  },
  "size": 0,
  "aggs": {
    "sizes": {
      "terms": { "field": "size" }
    }
  }
}
```

返回结果如下所示。

```
{
  "took": 3,
  "timed_out": false,
  "_shards": {
    "total": 1,
    "successful": 1,
```

```
"skipped": 0,
"failed": 0
},
"hits": {
  "total": {
    "value": 4,
    "relation": "eq"
  },
  "max_score": null,
  "hits": [ ]
},
"aggregations": {
  "sizes": {
    "doc_count_error_upper_bound": 0,
    "sum_other_doc_count": 0,
    "buckets": [
      {
        "key": "S",
        "doc_count": 2
      },
      {
        "key": "L",
        "doc_count": 1
      },
      {
        "key": "M",
        "doc_count": 1
      }
    ]
  }
}
}
```

## 步骤 4：删除索引

当不再使用对应的索引数据时，可以在 OpenSearch Dashboards 执行如下命令删除对应索引，避免造成资源浪费。

```
DELETE /my_store
```

返回结果如下所示。

```
{
  "acknowledged": true
}
```

## 步骤 5：删除集群

已完成数据搜索业务，无需继续使用集群时，可删除集群释放资源。

### 📖 说明

由于集群删除后，数据无法恢复，请谨慎操作。

1. 登录云搜索服务管理控制台。
2. 在左侧菜单栏选择“集群管理 > OpenSearch”。
3. 在集群列表，选择“Sample-OSCluster”集群，在操作列单击“更多” > “删除”。
4. 在弹出的确认对话框中，确认要删除的集群名称，单击“确定”完成操作。

# 3 使用 Logstash 迁移数据

Logstash是一款收集、转换、清洗、解析数据的工具，本章节为您提供了一个使用Logstash集群完成Elasticsearch集群间数据迁移的示例，您可以参考此示例来了解Logstash服务，包括创建集群、导入导出数据、任务配置等操作。

## 操作流程

以下是使用Logstash集群将源Elasticsearch集群迁移到目标Elasticsearch集群的操作步骤。

开始迁移数据前，请务必按[准备工作](#)指导完成必要操作。

1. **步骤1：获取Elasticsearch集群信息**：获取源Elasticsearch集群和目标Elasticsearch集群的访问地址。
2. **步骤2：创建Logstash集群**：创建一个Logstash集群用于迁移Elasticsearch集群的数据。
3. **步骤3：配置迁移任务**：在Logstash集群配置Elasticsearch集群的迁移任务。
4. **步骤4：启动迁移任务**：在Logstash集群启动配置好的迁移任务。
5. **步骤5：停止任务**：当数据迁移完成，不再需要迁移任务时，您可以停止任务。
6. **步骤6：删除集群**：无需继续使用集群时，可删除集群释放资源。

## 准备工作

- 已注册华为账号并开通华为云，进行了实名认证，且在使用云搜索服务前检查账号状态，账号不能处于欠费或冻结状态。  
如果您还没有华为账号，请参考以下步骤创建。
  - a. 打开[华为云网站](#)。
  - b. 在页面右上角单击“注册”，根据提示信息完成注册。
  - c. 勾选服务条款，单击“开通”。
  - d. 实名认证，请参考：
    - 个人账号：[实名认证](#)
    - 企业账号：[实名认证](#)
- 已准备好源Elasticsearch集群“Source-ES”和目标Elasticsearch集群“Dest-ES”，且都是单节点的非安全集群。

## 步骤 1：获取 Elasticsearch 集群信息

获取源Elasticsearch集群和目标Elasticsearch集群的访问地址，如果是安全集群还需要联系服务管理员获取集群的用户名和密码。

1. 登录云搜索服务管理控制台。
2. 在左侧菜单栏选择“集群管理 > Elasticsearch”。
3. 在集群管理列表页面，选择需要访问的集群，在“内网访问地址”列获取Elasticsearch集群的IP地址，一般是“<host>:<port>”或“<host>:<port>,<host>:<port>”样式。

例如本示例里，源Elasticsearch集群“Source-ES”的访问地址是“10.62.179.32:9200”，目标Elasticsearch集群“Dest-ES”的访问地址是“10.62.179.33:9200”。

图 3-1 获取 IP 地址

名称ID	集群状态	任务状态	版本	创建时间	企业版	内网访问地址	计费模式	操作
LO-E	可用	-	6.2.3	2024-07-16 16:34:38 GMT+08	edtest	192.168.0.5:9200	按需计费	Kibana
US-E	可用	-	6.2.3_24.2.0_0615					
CS-E	可用	-	7.10.2	2024-07-16 16:30:41 GMT+08	default	10.62.179.32:9200	按需计费	Kibana
SH-E	可用	-	7.10.2_24.2.1_0715					
DS-E	可用	-	7.10.2	2024-07-16 15:19:18 GMT+08	default	192.168.0.5:9200	按需计费	Kibana
DE-E	可用	-	7.10.2_24.2.1_0715					

## 步骤 2：创建 Logstash 集群

创建一个Logstash集群用于迁移Elasticsearch集群的数据。

1. 登录云搜索服务管理控制台。
2. 在左侧菜单栏选择“集群管理 > Logstash”。
3. 单击右上角的“创建集群”，进入“创建集群”页面。
4. 配置集群的计费模式和可用区。

表 3-1 计费模型和可用区参数说明

参数	说明	取值样例
计费模式	<p>集群支持“包年/包月”和“按需计费”两种模式。</p> <ul style="list-style-type: none"> <li>包年/包月：根据集群购买时长，一次性支付集群费用。最短时长为1个月，最长时长为3年。如果购买时长超过9个月，建议包年购买，价格更优惠。一年计费为购买10个月得12个月。</li> <li>按需计费：按实际使用时长计费，计费周期为一小时，不足一小时按一小时计费。</li> </ul>	按需计费
当前区域	<p>选择集群的所在区域。</p> <p>不同区域的云服务产品之间内网互不相通。请就近选择靠近您业务的区域，可减少网络时延，提高访问速度。</p>	华北-北京四

参数	说明	取值样例
可用区	选择集群工作区域下关联的可用区。最多支持配置3个“可用区”。	可用区1

5. 配置集群基本信息。

图 3-2 配置集群信息

The screenshot shows a configuration interface with three main sections:

- 集群类型 (Cluster Type):** Three buttons are visible: 'Elasticsearch', 'Logstash' (highlighted in blue), and 'OpenSearch'.
- 集群版本 (Cluster Version):** A dropdown menu is set to '7.10.0'.
- 集群名称 (Cluster Name):** A text input field contains 'Sample-Logstash'.

表 3-2 基本参数说明

参数	说明	取值样例
集群类型	选择“Logstash”。	Logstash
集群版本	选择所需的集群版本，支持的版本以界面可选项为准。	7.10.0
集群名称	自定义集群名称，可输入的字符范围为4~32个字符，只能包含数字、字母、中划线和下划线，且必须以字母开头。	Sample-Logstash

6. 配置集群的规格信息。

图 3-3 配置集群规格

The screenshot shows a configuration interface for cluster specifications:

- 节点数量 (Node Count):** A numeric input field is set to '1'.
- CPU架构 (CPU Architecture):** A button labeled 'X86计算' is selected.
- 节点规格 (Node Specification):** Three tabs are visible: '计算规格型' (selected), '通用计算型', and '内存优化型'. Below the tabs is a table of instance types:
 

规格名称	vCPU核   内存	建议存储范围
<input checked="" type="radio"/> ess.spec-4u8g	4 vCPUs   8 GB	40 GB - 1,500 GB
<input type="radio"/> ess.spec-8u16g	8 vCPUs   16 GB	80 GB - 1,500 GB
<input type="radio"/> ess.spec-16u32g	16 vCPUs   32 GB	100 GB - 3,200 GB
<input type="radio"/> ess.spec-32u64g	32 vCPUs   64 GB	100 GB - 10,240 GB
- 节点存储 (Node Storage):** Two buttons are visible: '通用' (selected) and '超等I/O'.
- 节点存储容量 (Node Storage Capacity):** A numeric input field is set to '40' GB.

表 3-3 规格参数说明

参数	说明	取值样例
节点数量	集群中的节点个数。可选节点数为1~100。	1
CPU架构	仅支持“X86计算”类型。	X86计算

参数	说明	取值样例
节点规格	选择集群节点的规格。	ess.spec-4u8g
节点存储	选择集群节点的存储类型。	高I/O
节点存储容量	设置集群节点的存储空间大小，默认配置40GB。	40GB

7. 设置集群的企业项目。  
如果您开通了“企业项目”，在创建集群时，可以给集群绑定一个企业项目。本示例选择默认企业项目“default”。
8. 单击“下一步：网络配置”，设置集群的网络配置。

图 3-4 配置网络



表 3-4 网络配置参数说明

参数	说明	取值样例
虚拟私有云	指定集群节点使用的虚拟专用网络，实现不同业务的网络隔离。 和Elasticsearch集群选择同一个VPC。 <b>说明</b> 此处您选择的VPC必须包含网段（CIDR），否则集群将无法创建成功。新建的VPC默认包含网段（CIDR）。	vpc-default
子网	集群使用子网实现与其他网络的隔离，并独享所有网络资源，以提高网络安全。	subnet-default
安全组	安全组起着虚拟防火墙的作用，为集群提供安全的网络访问控制策略。 <b>说明</b> 为了确保您能够正常访问集群，需要放通安全组9200规则。	default

9. 单击“下一步：高级配置”，可以设置Logstash集群的高级功能。  
本集群仅做入门指导使用，无需开启集群高级功能。
10. 单击“下一步：确认配置”，确认完成后单击“立即创建”开始创建集群。



- 单击“返回集群列表”，系统将跳转到“集群管理”页面。您创建的集群将展现在集群列表中，且集群状态为“创建中”，创建成功后集群状态会变为“可用”。

图 3-5 创建集群

名称ID	集群状态	任务状态	版本	创建时间	企业项目	内部访问地址	计费模式	操作
Sample-Logstash cf600c39-4000-405a-8088...	U 创建中	-	7.10.0 logstash	-	default	-	按年包月	返回中心 查看详情 更多

### 步骤 3：配置迁移任务

在Logstash集群配置Elasticsearch集群的迁移任务。

- 在Logstash集群管理页面，选择已创建的“Sample-Logstash”集群，进入集群基本信息页面。
- 单击右侧“配置中心”，进入Logstash配置中心界面。

图 3-6 Logstash 配置中心界面



- 在配置中心界面，单击右上角“创建”，进入创建配置文件页面，编辑配置文件。

图 3-7 配置任务

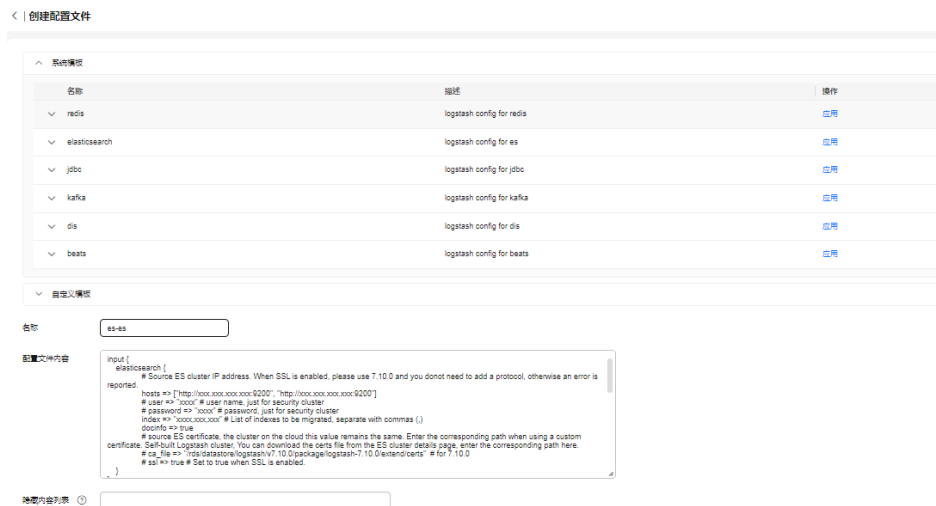


表 3-5 创建配置文件参数说明

参数	说明	取值样例
名称	自定义配置文件名称。 只能包含字母、数字、中划线或下划线，且必须以字母开头。必须大于等于4个字符。	es-es
配置文件内容	展开上方“系统模板”，选择“elasticsearch”，单击操作列的“应用”，在配置文件内容中，基于模板注释填写配置项。	关键配置项请参见表3-6，其他配置项保持默认值。
隐藏内容列表	配置隐藏字符串列表后，在返回的配置内容中，会将所有在列表中的字符串隐藏为“***”。 输入需要隐藏的敏感字符串列表，按“Enter”创建。 列表最大支持20条，单个字符串最大长度512字节。	不涉及

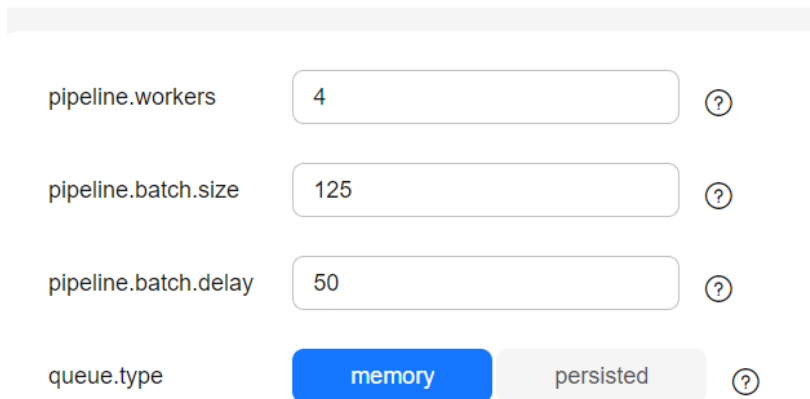
表 3-6 配置项说明

配置	说明	取值样例
hosts	分别在input和output中填写源Elasticsearch集群和目标Elasticsearch集群的访问地址。获取方式请参见步骤1：获取Elasticsearch集群信息。	input hosts: http:// 10.62.179.32:9200  output hosts: http:// 10.62.179.33:9200
user	访问Elasticsearch集群的用户名。仅安全集群涉及，如果是非安全集群则此项使用“#”注释掉。	使用“#”注释掉
password	访问Elasticsearch集群的密码。仅安全集群涉及，如果是非安全集群则此项使用“#”注释掉。	使用“#”注释掉
index	需要迁移的索引信息，可以使用通配符设置。	index*

4. 编辑完成后，单击“下一页”配置Logstash管道参数。

图 3-8 配置管道参数

&lt; | 创建配置文件



The screenshot shows a configuration interface for a Logstash pipeline. It includes four input fields and a dropdown menu:

- `pipeline.workers`: Input field with value 4.
- `pipeline.batch.size`: Input field with value 125.
- `pipeline.batch.delay`: Input field with value 50.
- `queue.type`: Dropdown menu with options `memory` (selected) and `persisted`.

表 3-7 管道参数说明

参数	说明	取值样例
<code>pipeline.workers</code>	并行执行管道的Filters和Outputs阶段的工作线程数。	4
<code>pipeline.batch.size</code>	单个工作线程在尝试执行其Filters和Outputs之前将从inputs收集的最大事件数，该值较大通常更有效，但会增加内存开销。	125
<code>pipeline.batch.delay</code>	创建管道事件批时，在将过小的批调度到管道工作线程之前，等待每个事件的时间。	50
<code>query.type</code>	用于事件缓冲的内部队列模型。 <ul style="list-style-type: none"><li>“memory”为基于内存的传统队列。</li><li>“persisted”为基于磁盘的ACKed持久化队列。</li></ul>	memory

- 配置完成后，单击“创建”，系统会自动进行配置文件校验，等待配置文件状态为“可用”，表示创建成功。

图 3-9 配置文件校验



## 步骤 4：启动迁移任务

在Logstash集群启动配置好的迁移任务。

1. 在Logstash集群管理页面，选择已创建的“Sample-Logstash”集群，进入集群基本信息页面。
2. 单击右侧“配置中心”，进入Logstash配置中心界面。
3. 选择状态为“可用”的配置文件，单击“启动”，启动配置任务。管道列表中“事件”列显示配置任务处理的任务数量。

图 3-10 启动配置任务



4. 数据迁移完毕，检查目标Elasticsearch集群和源集群数据的一致性。例如，分别在源集群和目标集群执行 `_cat/indices` 命令，对比两者的索引信息是否一致。
  - a. 在Elasticsearch集群管理页面，选择源Elasticsearch集群“Source-ES”或目标Elasticsearch集群“Dest-ES”，单击操作列中的“Kibana”进入Kibana控制台。
  - b. 在Kibana的左侧导航中选择“Dev Tools”，进入Console界面。
  - c. 在Console界面，执行如下命令查看索引信息。

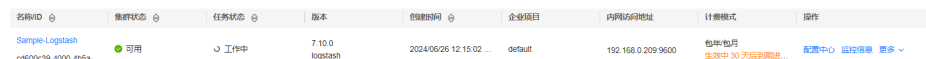
```
_cat/indices
```

## 步骤 5：停止任务

当数据迁移完成，不再需要迁移任务时，您可以停止任务。

1. 登录云搜索服务管理控制台。
2. 在左侧菜单栏选择“集群管理 > Logstash”。
3. 在集群列表，选择“Sample-Logstash”集群，单击操作列的“配置中心”，进入Logstash集群配置页面。

图 3-11 进入配置页面



4. 选择已经启动的管道名称，单击“全部停止”，停止正在运行的任务，等待所有管道任务停止完成。

图 3-12 停止任务



## 步骤 6：删除集群

已完成数据迁移业务，无需继续使用集群时，可删除集群释放资源，删除前请确保所有正在运行的管道任务全部停止。

### 说明

集群删除前请停止所有正在运行的任务，请谨慎操作，删除前请备份配置文件。

1. 登录云搜索服务管理控制台。
2. 在左侧菜单栏选择“集群管理 > Logstash”。
3. 在集群列表，选择“Sample-Logstash”集群，在操作列单击“更多” > “删除”。
4. 在弹出的确认对话框中，确认要删除的集群名称，单击“确定”完成操作。

# 4 入门实践

云搜索服务为您提供各种业务场景、方案架构的最佳实践，方便您根据自身的业务场景选择合适的实践方案。

表 4-1 常用最佳实践

实践场景		描述
数据迁移	源端为Elasticsearch的数据迁移	<ul style="list-style-type: none"><li>● <b>使用备份与恢复迁移集群数据（源端为CSS Elasticsearch）</b> 本实践为您介绍通过备份与恢复集群快照功能实现云搜索服务Elasticsearch集群之间的数据迁移。</li><li>● <b>使用备份与恢复迁移集群数据（端源为第三方Elasticsearch）</b> 本实践为您介绍通过备份与恢复集群快照功能实现自建ES到华为云ES之间的数据迁移和第三方友商ES到华为云ES之间的数据迁移。</li></ul>
		<b>使用CDM迁移数据</b> 华为云服务提供的CDM云迁移工具，可以实现不同云服务间的集群迁移能力，本实践介绍如何通过CDM将本地Elasticsearch整库迁移到云搜索服务中。
		<b>使用Logstash迁移集群数据</b> Logstash是Elasticsearch官方提供的数据库清洗工具，ELK生态中的一部分，功能强大，通过此实践可以完成不同数据源和Elasticsearch数据的迁移，还可以进行数据的清洗和加工。
	<b>源端为Kafka/MQ的数据迁移</b>	在泛IOT、新闻、舆情、社交等数据量较大的行业，通常会引入消息中间件（kafka、MQ等）对流量进行削峰填谷，然后借助Flink/Logstash等工具消费数据并进行数据预处理，最终将数据导入到搜索引擎。本实践为您介绍详细介绍源端为Kafka/MQ的集群迁移流程和操作流程。

实践场景		描述
	<a href="#">源端为数据库的数据迁移</a>	Elasticsearch支持全文检索和Ad Hoc查询，所以经常作为关系型数据库（例如MySQL、GaussDB for MySQL等）的补充，以此提升数据库的全文检索能力和高并发的Ad Hoc查询能力。
优化集群性能	<a href="#">优化Elasticsearch和OpenSearch集群写入性能</a>	集群在使用前，建议参考本实践进行集群的写入性能优化，便于提高集群的写入性能，提升使用效率。
	<a href="#">优化Elasticsearch和OpenSearch集群查询性能</a>	集群在使用前，建议参考本实践进行集群的查询性能优化，便于提高集群的查询性能，提升使用效率。
加速关系型数据库查询	<a href="#">使用Elasticsearch加速关系型数据库的查询分析</a>	关系型数据库（例如MySQL、GaussDB for MySQL等）受限于全文检索和Ad Hoc查询能力，因此会将Elasticsearch作为关系型数据库的补充，以此提升数据库的全文检索能力和高并发的Ad Hoc查询能力。
构建日志管理平台	<a href="#">使用Elasticsearch和Logstash构建日志管理平台</a>	使用云搜索服务搭建的统一日志管理平台可以实时地、统一地、方便地管理日志，让日志驱动运维、运营等，提升服务管理效率。
自定义规则排序搜索	<a href="#">使用Elasticsearch自定义规则排序搜索结果</a>	本实践为您介绍如何对搜索匹配的文档进行自定义评分。