

云搜索服务

# 快速入门

文档版本 01  
发布日期 2023-10-19



版权所有 © 华为技术有限公司 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。

# 安全声明

## 漏洞处理流程

华为公司对产品漏洞管理的规定以“漏洞处理流程”为准，该流程的详细内容请参见如下网址：

<https://www.huawei.com/cn/psirt/vul-response-process>

如企业客户须获取漏洞信息，请参见如下网址：

<https://securitybulletin.huawei.com/enterprise/cn/security-advisory>

---

## 目录

---

1 快速开始使用云搜索服务.....	1
2 入门实践.....	11

# 1 快速开始使用云搜索服务

本章节为您提供了一个简单的商品搜索示例，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

- **步骤1：创建集群**
- **步骤2：导入数据**
- **步骤3：搜索数据**
- **（可选）步骤4：删除索引**
- **（可选）步骤5：删除集群**

## 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。

如果您已有一个华为账号，请跳到下一个任务。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：
  - 个人账户：[实名认证](#)
  - 企业账号：[实名认证](#)

## 步骤 1：创建集群

本示例您需要创建一个名称为“Sample-ESCluster”的Elasticsearch集群。此集群仅用于入门指导使用，建议选用“节点规格”为“ess.spec-4u8g”，“节点存储”为“高I/O”，“节点存储容量”为“40GB”。

1. 登录云搜索服务管理控制台。
2. 单击右上角的“创建集群”，进入“创建集群”页面。
3. 计费模式支持“按需计费”和“包年/包月”2种模式。本示例计费模式选择“按需计费”，按需计费按照实际使用时长计费，计费周期为一小时，不足一小时按一小时计费。
4. 选择“当前区域”和“可用区”。

表 1-1 区域和可用区参数说明

参数	说明
当前区域	在右侧下拉框中选择集群的所在区域。
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”。 本场景示例选择1个可用区。

5. 配置集群基本信息。

图 1-1 配置集群信息

The image shows a configuration form for creating a cluster. It has three rows of configuration options:

- 集群类型 (Cluster Type):** Three buttons are shown: 'Elasticsearch' (highlighted in blue), 'Logstash', and 'OpenSearch'.
- 集群版本 (Cluster Version):** A dropdown menu is set to '7.10.2'.
- 集群名称 (Cluster Name):** A text input field contains 'Sample-ESCluster'.

表 1-2 基本参数说明

参数	说明
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。 本场景示例选择7.10.2版本。

参数	说明
集群名称	自定义集群名称。 本场景示例您可以创建一个名称为“Sample-ESCluster”的集群。

6. 配置集群的规格信息。

图 1-2 配置集群规格



表 1-3 规格参数说明

参数	说明
节点数量	集群中的节点个数。可选节点数为1~32。 本场景示例选择1个节点。
CPU架构	目前支持“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。
节点规格	集群中的节点规格。 本场景示例节点规格建议选择“ess.spec-4u8g”。
节点存储	选择存储类型。 本示例节点存储建议选择高I/O。
节点存储容量	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。 本场景示例存储容量建议选择40GB。
启用Master节点	Master节点用于管理集群中的所有节点。当需要存储和分析的数据量大，所需节点数量大于20个节点时，建议启用Master节点，保证集群的稳定性。 本场景示例无需用Master节点。
启用Client节点	Client节点用于提供客户端接入集群和分析数据的服务。当需要存储和分析的数据量大，所需节点数量大于20个节点时，建议启用Client节点，保证集群的稳定性。 本场景示例无需启用Client节点。

参数	说明
启用冷数据节点	冷数据节点用于存放对于历史数据要求分钟级别的返回。当用户对历史数据返回时间要求不是很高的话，可以将这部分数据存储冷数据节点上，从而降低成本。 本场景示例无需启用冷数据节点。

#### 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。您可以在右侧下拉框中选择当前用户下已创建的企业项目，也可以通过单击“查看项目管理”按钮，前往“企业项目管理”管理控制台，新建企业项目和查看已有的企业项目。

#### 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
虚拟私有云	VPC即虚拟私有云，是通过逻辑方式进行网络隔离，提供安全、隔离的网络环境。 选择创建集群需要的VPC，单击“查看虚拟私有云”进入VPC服务查看已创建的VPC名称和ID。如果没有VPC，需要创建一个新的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）。
子网	通过子网提供与其他网络隔离的、可以独享的网络资源，以提高网络安全。 选择创建集群需要的子网，可进入VPC服务查看VPC下已创建的子网名称和ID。
安全组	安全组是一个逻辑上的分组，为同一个VPC内具有相同安全保护需求并相互信任的弹性云服务器提供访问策略。单击“查看安全组”可了解安全组详情。
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。 本场景示例选择关闭安全模式。



- 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

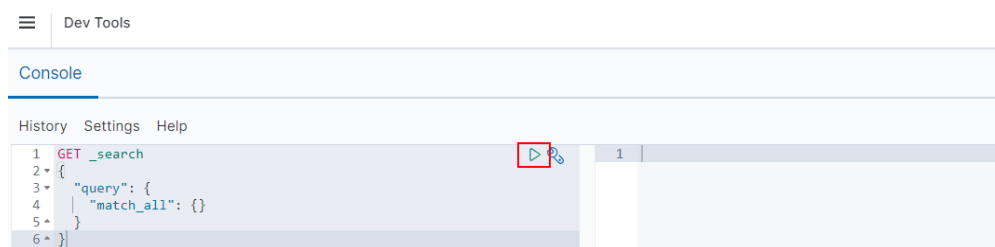
名称ID	集群状态	任务状态	版本	创建时间	企业项目	内网访问地址	计费模式	操作
css-055b	可用	-	7.10.2 elasticsearch	2024/03/15 15:24...	default	192.168.0.153.92...	按需计费	Kibana 监控信息 更多

## 步骤 2：导入数据

云搜索服务支持通过云数据迁移（简称CDM）、Logstash、Kibana或API将数据导入到Elasticsearch。Kibana是一个开源的数据分析与可视化平台，通过Kibana可以搜索、查看存放在Elasticsearch索引中的数据，也可以实现以图表、地图等方式展示数据。云搜索服务的Elasticsearch集群默认提供Kibana，本示例将以Kibana为例介绍将数据导入到Elasticsearch的操作流程。

- 在“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”。  
(低于7.x版本)

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

( 7.x版本及高于7.x版本 )

```
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。

( 低于7.x版本 )

```
POST /my_store/products/_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"}
```

( 7.x版本及高于7.x版本 )

```
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

- 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”。这里使用Kibana演示用户搜索数据在后台的执行命令和返回结果。

执行命令如下所示。

（低于7.x版本）

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

（7.x版本及高于7.x版本）

```
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" : "9xf6VHIBfCl6SDjw7H5",
        "_score" : 1.7965372,
        "_source" : {
          "productName" : "2018春装新款牛仔裤女装",
          "size" : "M"
        }
      },
      {
        "_index" : "my_store",
        "_type" : "_doc",
        "_id" : "-Bf6VHIBfCl6SDjw7H5",
        "_score" : 1.7965372,
        "_source" : {
          "productName" : "2018春装新款牛仔裤女装",
          "size" : "S"
        }
      },
      {
        "_index" : "my_store",
        "_type" : "_doc",
        "_id" : "-Rf6VHIBfCl6SDjw7H5",
        "_score" : 0.5945667,
        "_source" : {
          "productName" : "2017春装新款休闲裤女装",
          "size" : "L"
        }
      }
    ]
  }
}
```

```
{
  "_index": "my_store",
  "_type": "_doc",
  "_id": "-hf6VHIBfCl6SDJw7H5",
  "_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。这里使用Kibana演示聚合结果显示功能在后台的执行命令和返回结果。

执行命令如下所示。

（低于7.x版本）

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

（7.x版本及高于7.x版本）

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

返回结果如下所示。

（低于7.x版本）

```
{
  "took": 31,
  "timed_out": false,
  "_shards": {
    "total": 1,
```

```
"successful" : 1,
"skipped" : 0,
"failed" : 0
},
"hits" : {
  "total" : 4,
  "max_score" : 0.0,
  "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
      }
    ]
  }
}
}
```

( 7.x版本及高于7.x版本 )

```
{
  "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
        }
      ]
    }
  }
}
```

## （可选）步骤 4：删除索引

如果不再使用对应的索引数据，您可以执行如下命令，删除对应索引，避免造成浪费资源。

```
DELETE /my_store
```

返回结果如下所示。

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

## （可选）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。在左侧菜单栏选择“集群管理 > Elasticsearch”。
2. 进入集群管理页面，选中“Sample-ESCluster”集群所在行，在操作列单击“更多” > “删除”。
3. 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 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 2-1 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>源端为Kafka/MQ的数据迁移</b>	<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。 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。	
接入集群	<a href="#">使用Kibana和Cerebro组件快速访问集群</a>	云搜索服务创建的集群自带Kibana和Cerebro组件，支持一键打开Kibana和Cerebro，快速访问集群。	
	<a href="#">通过Curl命令行接入集群</a>	当云搜索服务和弹性云服务器ECS在同一VPC内时，在此ECS中可以通过Curl命令直接访问Elasticsearch 集群，此方法多用于前期调试Elasticsearch的连接性，排查访问集群的客户端是否和Elasticsearch节点的网络连通。	
	通过Java接入集群	<a href="#">通过Rest High Level Client接入集群</a>	Elasticsearch官方提供了SDK（Rest High level Client）方式连接集群，Rest Client客户端对Elasticsearch的API进行了封装，您只需要构造对应的结构即可对集群进行访问。
		<a href="#">通过Rest Low Level Client接入集群</a>	High Level Client是在Low Level Client基础上进行封装的，如果在High Level Client中的方法调用（例如.search，.bulk）不能满足使用需求，或存在兼容性问题，可以选择使用Low Level Client方式，甚至可以使用“HighLevelClient.getLowLevelClient()”方式直接获取Low Level Client。
		<a href="#">通过Transport Client接入集群</a>	本实践为您介绍如何使用Elasticsearch原生的Transport Client客户端连接非安全模式的集群集群。
		<a href="#">使用Spring Boot接入集群</a>	本实践为您介绍使用Spring Boot工具接入集群的详细流程。
	<a href="#">通过Python接入集群</a>	本实践为您介绍通过Python语言访问Elasticsearch集群。	
<a href="#">通过ES-Hadoop实现Hive读写Elasticsearch数据</a>	本实践通过MRS的ES-Hadoop与Elasticsearch集群连接作为示例，你可以配置其他任何需要使用Elasticsearch的应用。如有需要，也可以参考本文在其他服务中使用Elasticsearch，前提是要保证客户端与Elasticsearch集群网络连通。		
优化集群性能	<a href="#">写入性能优化</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。	



实践场景		描述
	<a href="#">查询性能优化</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。
管理索引生命周期	<a href="#">使用生命周期实现自动滚动索引</a>	对于时间序列数据，随着时间推移数据持续写入，索引会越来越大，通过生命周期管理来定期将数据滚动到新索引，并将历史老索引删除，实现自动滚动索引。
	<a href="#">使用生命周期实现自动存算分离</a>	CSS支持存算分离，即将索引冻结到OBS来降低冷数据的存储成本。本文介绍如何使用索引生命周期管理，在特定的时间自动冻结索引，实现存算分离。
加速数据查询分析	<a href="#">使用CSS加速数据库的查询分析</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。
建统一日志管理平台	<a href="#">使用CSS搭建统一日志管理平台</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。
自定义评分查询	<a href="#">使用Elasticsearch集群自定义评分查询</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。