

初识云搜索服务入门操作指导

1 任务介绍

本次任务，介绍华为[云搜索服务](#)集群创建、使用、删除。通过本次任务，您将学会云搜索服务的基本操作流程，理解 Elasticsearch 的基本概念和关键技术的应用。

2 任务执行

使用云搜索服务的基本流程如下所示：



下面，我们用一个简单场景，来快速了解云搜索服务的使用流程。

2.1 场景准备

某女装品牌在网上[经营电商业务](#)，提供不同尺寸的 3 个产品，该电商网站使用云搜索服务来为用户提供商品搜索功能。

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"
}
]
```

2.2 创建[虚拟私有云](#)

步骤 1 从华为云官网进入虚拟私有云产品页



步骤 2 点击“立即购买”登录华为云控制台。



步骤 3 选择左上角区域为“华北-北京一”。



步骤 4 在“网络控制台”页面右上角单击“创建虚拟私有云”。



步骤 5 根据界面提示配置虚拟私有云参数后，单击“立即创建”。

- “区域”保持默认为“华北-北京一”。
- “名称”配置为“vpc-css-demo”，也可以按规范命名。
- “网段”保持默认值。
- “标签”保持默认值，不填写。
- “可用分区”保持默认值“可用区 2”。
- “子网名称”配置为“subnet-css-demo”，也可以按规范命名。
- “子网网段”保持默认值。
- “高级配置”选择默认配置。

参数配置完成后，如图所示。

创建虚拟私有云 ?

[返回虚拟私有云列表](#)

基本信息

区域

华北-北京一

* 名称

vpc-css-demo

* 网段

192.168.0.0 / 16

建议使用网段：10.0.0.0/8~24，172.16.0.0/12~24，192.168.0.0/16~24

标签 ?

经过最佳实践的总结，建议在给资源关联标签之前先在TMS上创建预定义标签。[查看预定义标签](#)

请输入标签键

请输入标签值

您还可以创建10个标签。

子网配置

可用区 ?

可用区2

可用区1

* 子网名称

subnet-css-demo

* 子网网段

192.168.0.0 / 24 ?

可用IP数:250

子网创建完成后无法修改

高级配置

默认配置

自定义配置

步骤 6 确认安全组 Sys-default 的出入方向都允许所有协议和所有端口访问。



步骤 7 若没有上述规则，请添加。否则，略过此步骤。



2.3 创建云搜索服务集群

步骤 1 从华为云官网进入云搜索服务产品页



步骤 2 点击“立即使用”按钮进入控制台页面。



步骤 3 在“云搜索服务”页面右上角单击“创建集群”



步骤 4 根据界面提示配置参数后，单击“立即申请”。

- “当前区域” 从下拉框选择 “华北-北京一”
- “集群版本” 从下拉框选择，是 Elasticsearch 的软件版本，选择 5.5.1。
- “集群名称” 按规范命名配置，可以输入 Es-test。
- “节点数量” 自由选择配置，生产集群一般建议 3 个以上。这里选择 1。

- “节点规格” 自由选择，生产集群一般建议 `ess.spec-2u16g` 以上。这里我们选择 `ess.spec-1u8g`。
- “节点存储” 自由选择，这里选择 “高 I/O(推荐)”。
- “节点存储容量” 可以在允许范围内自由选择。这里输入 40GB。
- “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。选择刚才创建的 “vpc-css-demo”。
- “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 “subnet-css-demo”。
- “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。选择 “Sys-default”。
- “集群快照”：为了增强数据可靠性保障，系统默认开启自动快照功能。这里，我们将其关闭。

创建集群 ?

[返回集群列表](#)

计费模式

按需计费

当前区域

华北-北京一

不同区域的资源之间内网不互通。请选择靠近您客户的区域，可以降低网络时延、提高访问速度。

可用区 ?

cn-north-1a

集群版本

5.5.1

集群名称

Es-test ?

您还可以创建199个集群实例。您的可用资源包含799核vCPU，1,592GB内存，32,668GB磁盘。

节点数量

- 1 + ?

节点规格

ess.spec-1u8g

选择的规格为：1 vCPUs | 8 GB

节点存储

普通I/O

高I/O（推荐）

超高I/O

节点存储容量

- 40 + GB ?

虚拟私有云 ?

vpc-css-demo

[查看虚拟私有云](#)

子网 ?

subnet-css-demo(192.168.0.0/24)

安全组 ?

Sys-default

[查看安全组](#)

集群快照

启用自动快照 ☒

数据保护已关闭，存在风险。创建索引时建议设置多个副本。

步骤 5 在规格确认界面确认完成后，单击“提交申请”开始创建集群。

步骤 6 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”。

云搜索服务		您还可以创建19个集群实例。您的可用资源包含800核vCPU，1,600GB内存，32,768GB磁盘。		
总览				
集群管理				
集群名称	集群状态	任务状态	搜索引擎版本	
Es-test	创建中	-	5.5.1	-

步骤 7 耐心等待十几分钟，创建成功后集群状态会变为“可用”。

云搜索服务		您还可以创建19个集群实例。您的可用资源包含799核vCPU，1,592GB内存，32,728GB磁盘。		
总览				
集群管理				
集群名称	集群状态	任务状态	搜索引擎版本	创建时间
Es-test	可用	-	5.5.1	2018/07/30 15:05:05 GMT+...

2.4 导入数据

云搜索服务支持通过[云数据迁移](#)（简称 CDM）、[Logstash](#)、[Elasticsearch API](#) 或 [Kibana](#) 将数据导入。其中 [Kibana](#) 是 [Elasticsearch](#) 的图形化界面，其 dev tool 提供了简单的交互验证功能。本例我们就采取这种方式。

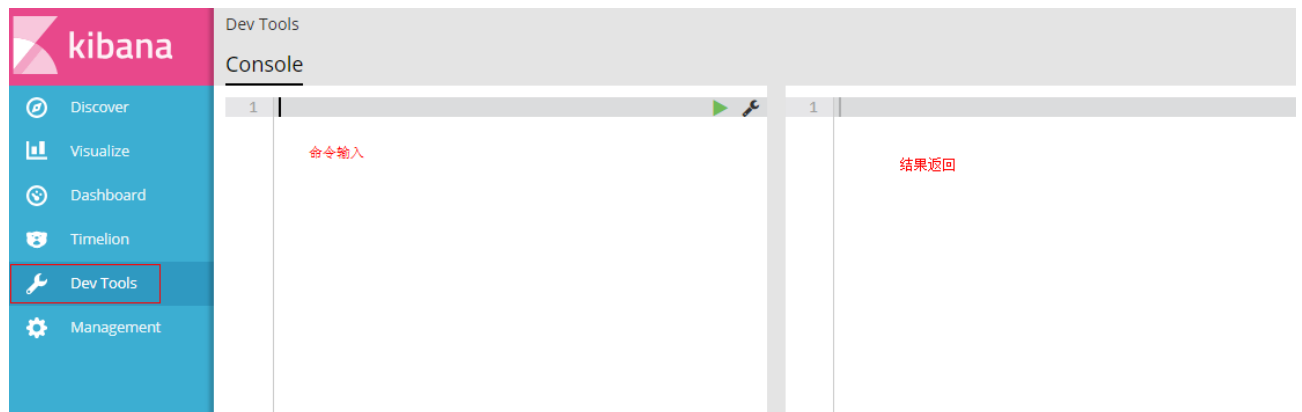
在导入数据前，我们可以选择自己创建索引定义，也可以让 Elasticsearch 引擎根据数据自动推测索引定义。本例中，我们采用自己创建的方式，来控制索引设置和字段类型定义。

步骤 1 在云搜索服务的“集群管理”页面上，单击集群“操作”列的“Kibana”访问集群。

云搜索服务		您还可以创建19个集群实例。您的可用资源包含799核vCPU，1,592GB内存，32,728GB磁盘。			请输入集群名称	
集群名称	集群状态	任务状态	搜索引擎版本	创建时间	内网访问地址	公网访问地址
Es-test	可用	-	5.5.1	2018/07/30 15:05:05 GMT+...	192.168.0.244:9200	-
				Kibana 更改规格 更多		

步骤 2 在 Kibana 的左侧导航中选择“Dev Tools”，单击“Get to work”，进入 Console 界面

说明：这里 dev tools 作为一个客户端，通过 REST 接口向 Elasticsearch 发起请求，并得到响应。



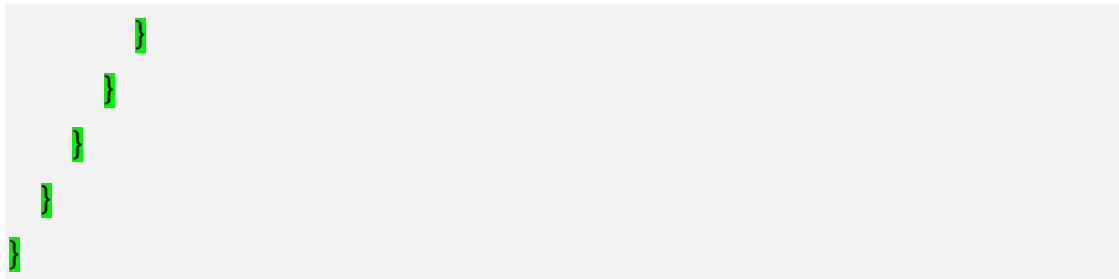
步骤 3 在 Console 界面，执行如下命令创建索引 “my_store”，并指定自定义映射来定义数据类型。

说明：创建索引使用 PUT 请求。

setting 中指定了索引的相关配置信息，因为数据量较少，我们的 shard 数设置为 1。

mapping 中指定 type 的名称为 “products”，field 依次为 “productName” 和 “size”。其中，“productName” 是文本类型，采用了 ik_smart 分词器进行分词。“size” 是 keyword 类型，表示不需要分词。

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



如下：



步骤 4 在 Console 界面，执行如下命令，将数据导入到 “my_store” 索引中。

说明：导入数据使用 **POST** 请求，通过 [Elasticsearch 的 bulk API](#)。

因为我们在 api 中已经指定了 my_store/product，所以 “index” 字段后面无需参数。

```
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"}

```

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



2.5 搜索分析

2.5.1 全文检索

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。

说明：检索使用 Get 请求，通过 [Elasticsearch 的 search API](#)。

Elasticsearch 检索使用 [DSL 语法](#)，采用 json 格式。

```
GET /my_store/products/_search
```

```

{
  "query": {"match": {
    "productName": "春装牛仔裤"
  }}
}

```

```
}}  
}
```

返回结果如下：

```
Dev Tools  
Console  
1 GET /my_store/products/_search  
2 {  
3   "query": {"match": {  
4     "productName": "春装牛仔裤"  
5   }}  
6 }  
7  
8 },  
9 "hits": {  
10   "total": 4,  
11   "max_score": 1.7656698,  
12   "hits": [  
13     {  
14       "_index": "my_store",  
15       "_type": "products",  
16       "_id": "AWTqDMVbrsB_R_qRF19l",  
17       "_score": 1.7656698,  
18       "_source": {  
19         "productName": "2018春装新款牛仔裤女装",  
20         "size": "M"  
21       }  
22     },  
23     {  
24       "_index": "my_store",  
25       "_type": "products",  
26       "_id": "AWTqDMVbrsB_R_qRF19m",  
27       "_score": 1.7656698,  
28       "_source": {  
29         "productName": "2018春装新款牛仔裤女装",  
30         "size": "S"  
31       }  
32     },  
33     {  
34       "_index": "my_store",  
35       "_type": "products",  
36       "_id": "AWTqDMVbrsB_R_qRF19n",  
37       "_score": 0.5843511,  
38       "_source": {  
39         "productName": "2017春装新款休闲裤女装",  
40         "size": "L"  
41       }  
42     },  
43     {  
44       "_index": "my_store",  
45       "_type": "products",  
46       "_id": "AWTqDMVbrsB_R_qRF19o",  
47       "_score": 0.5843511,  
48       "_source": {  
49         "productName": "2017春装新款休闲裤女装",  
50         "size": "S"  
51       }  
52     }  
53   ]  
54 }
```

- 上面执行命令会将“春装牛仔裤”分词为“春装”和“牛仔裤”。这就是课程资料讲到的关键技术“分词”。
- 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，这里就用到了课程资料里讲到的关键技术“倒排索引”。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。
- 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。这里就是课程资料里讲到的关键技术“相关性评分”。

2.5.2 [聚合分析](#)

除了检索功能外，Elasticsearch 还可以支持数据分析。例如：可以对“春装”对应的产品按照尺码分类，统计不同尺码的数量。

执行如下命令：

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

返回结果如下：

```
1 GET /my_store/products/_search
2 {
3   "query": {
4     "match": { "productName": "春装" }
5   },
6   "size": 0,
7   "aggs": {
8     "sizes": {
9       "terms": { "field": "size" }
10    }
11  }
12 }
13
```

```
1 {
2   "took": 2,
3   "timed_out": false,
4   "_shards": {
5     "total": 1,
6     "successful": 1,
7     "failed": 0
8   },
9   "hits": {
10    "total": 4,
11    "max_score": 0,
12    "hits": []
13  },
14  "aggregations": {
15    "sizes": {
16      "doc_count_error_upper_bound": 0,
17      "sum_other_doc_count": 0,
18      "buckets": [
19        {
20          "key": "S",
21          "doc_count": 2
22        },
23        {
24          "key": "L",
25          "doc_count": 1
26        },
27        {
28          "key": "M",
29          "doc_count": 1
30        }
31      ]
32    }
33  }
34 }
```

2.6 删除集群

当集群资源使用完毕后，我们可以选择删除集群来释放资源。删除前记得先完成任务打卡。

步骤 1 在云搜索服务的“集群管理”页面上，单击集群“操作”列的“删除”。

集群名称	集群状态	任务状态	搜索引擎版本	创建时间	内网访问地址	公网访问地址	操作
Es-test	可用	-	5.5.1	2018/07/30 15:05:05 GMT+	192.168.0.244:9200	-	Kibana 更改规格 更多
							重启 删除 自定义词典 备份与恢复 集群迁移

步骤 2 点击“确定”，删除集群。



3 任务打卡

在云搜索服务的“集群管理”页面上，截图带有自己华为云用户名的集群列表，完成打卡。



4 扩展学习资料

https://support.huaweicloud.com/usermanual-css/css_01_0007.html