1-es接口和阿里云对接调研

Created by 杨超, last modified on 2018 Aug 13

> 这里记录一下和阿里云直接的对接，步骤比较繁琐

1 估算和购买es机器

相对来讲，es更类似于rds，成本较高，和nodes数和硬盘大小成正比，属于高投入高产出【可以建多个应用】

目前整个华北区只有华北2 e区 有资源，所以可能无法建立dev环境，目前的想法是直接在华北2区建立两个es，一个做test，一个做prod

```
test环境 -- 最低配置，一般来说2个nodes有脑裂风险不受SLA保护  只用来测试
1核2G
20G
2 nodes
```

tips:

创建速度较慢 因为要启动节点和集群，不是简单的分配资源
初始化的过程 ecs -> nodes -> SIA -> es集群
最后大约用了15-20分钟 比较慢

2 数据离线同步

2.1 参照文档

https://help.aliyun.com/document_detail/62584.html?spm=a2c4g.11186623.6.568.4jnQMx

https://help.aliyun.com/document_detail/52330.html?spm=a2c4g.11186623.2.7.iAOno5
https://help.aliyun.com/document_detail/62716.html?spm=a2c4g.11186623.6.608.z1yyVO

2.2 说明

1 es只支持使用数据集成进行离线同步，最多5min一次，所以可能产生数据集成费用

2 es需要一台配合的ecs来进行 数据同步，现在先用www-test顶着

3 es只能用新增资源组的方式进行调度，测试的时候无法直接测试，需要上线之后并更改调度组之后再进行测试，不过速度很快

最后的测试脚本如下

```
"job": {
"setting": {
...
},
"content": [
{
"reader": {
...
},
"writer": {
"name": "www-test-eswriter",
"parameter": {
"endpoint": "http://es-cn-v0h0q2d7l0001spov.elasticsearch.aliyuncs.com:9200",
"accessId": "",
"accessKey": "",
"index": "test-1",
"type": "default",
"cleanup": true,
"settings": {"index" :{"number_of_shards": 1, "number_of_replicas": 0}},
"discovery": false,
"batchSize": 1000,
"splitter": ",",
"column": [
{"name": "pk", "type": "id"},
{ "name": "col_ip","type": "ip" },
```

```
{ "name": "col_double","type": "double" },
{ "name": "col_long","type": "long" },
{ "name": "col_integer","type": "integer" },
{ "name": "col_keyword", "type": "keyword" },
{ "name": "col_text", "type": "text", "analyzer": "ik_max_word"},
{ "name": "col_geo_point", "type": "geo_point" },
{ "name": "col_date", "type": "date", "format": "yyyy-MM-dd HH:mm:ss"},
{ "name": "col_nested1", "type": "nested" },
{ "name": "col_nested2", "type": "nested" },
{ "name": "col_object1", "type": "object" },
{ "name": "col_object2", "type": "object" },
{ "name": "col_integer_array", "type":"integer", "array":true},
{ "name": "col_geo_shape", "type":"geo_shape", "tree": "quadtree", "precision": "10m"}
]
}
}
}
]
}
}
```

其他

1 kibana 确实很好用，不过功能有点复杂 需要摸索

2 数据查询正常

```
curl -u username:passwd -XGET "http://es-cn-v0h0q2d7l0001spov.elasticsearch.aliyuncs.com:9200/test-1/_search" -H 'Content-T
{
"query": {
"match_all": {}
}
}' | jq
```

```
{
  "took": 2,
  "timed_out": false,
  "_shards": {
    "total": 5,
    "successful": 5,
    "failed": 0
  },
  "hits": {
    "total": 33618,
    "max_score": 1,
    "hits": [
      {
        "_index": "test-1",
        "_type": "default",
        "_id": "AWUxcUGXN8kVhZVPLxfG",
        "_score": 1,
        "_source": {
          "score": 589,
          "investorid": 63,
          "name": "张野",
          "id": "63_3_1"
        }
      },
      {
        "_index": "test-1",
        "_type": "default",
        "_id": "AWUxcUGXN8kVhZVPLxfI",
        "_score": 1,
        "_source": {
          "score": 937,
          "investorid": 63,
          "name": "张野",
          "id": "63_3_5"
        }
      },
```