

数据库数据离线写入ES

1. 从数据库导出数据至文件中

业务方从数据库中导出数据

格式要求：csv格式，分隔符使用"|", 不要表头列名，空值不写null，如"a||c", "||"中间的值即为空值

2. 配置Logstash

业务方提供导出的数据各列对应elasticsearch中的字段名

如ngcct.csv中有7列数据，对应列名

为"serialNo","contactId","serviceTypeId","contactStartTime","mediaTypeId","fstCharAttrVal","secdCharAttrVal"

其中前5列为elasticsearch中对应字段，后两列无用，以contactStartTime作为业务时间字段

一条数据示例如下（fstCharAttrVal为空）

```
201806120929330001|201806120929333021|ynytcck|2018-06-12 09:29:33|1||aaa
```

对应的logstash配置如下:

```
input {
  file {
    path => [ "/home/zyzx/ngcct.csv" ]
    start_position => "beginning"
  }
}

filter {
  csv {
    columns => [
      "serialNo","contactId","serviceTypeId","contactStartTime","mediaTypeId"
    ]
    separator => "|"
  }

  ruby {
    code => 'event.set("logdate",
      (Time.parse(event.get("contactStartTime")) + 8 * 60 * 60).to_s)'
  }

  date {
    match => ["logdate", "yyyy-MM-dd HH:mm:ss Z"]
    target => "@timestamp"
  }

  ruby {
```

```

        code =>'event.set("timeDistance", (event.get("@timestamp")-
::Time.now).abs - 3 * 30 * 24 * 3600)'
    }

    mutate {
        remove_field => [ "logdate", "@version", "host", "path", "tags",
"message", "fstCharAttrVal","secdCharAttrVal" ]
    }

}

#logstash 输出插件
output {
    if[serviceTypeId] and [timeDistance] and [timeDistance] < 0 {
        elasticsearch {
            hosts => ["192.168.100.107:9200"]
            index => "ngcct_contact_{serviceTypeId}_{+YYYY_MM}"
            document_type => "contact"
            action => "update"
            doc_as_upsert => true
            codec => "json"
            document_id => "%{serialNo}"
        }
    }else {
        stdout {
            codec => "rubydebug"
        }
    }
}
}

```

注：

以上例子需要根据具体业务修改的地方

1. input中的file路径
2. filter中的csv列名
3. ruby中的时间字段名（ruby只能解析常用的时间格式，具体以业务对应的logstash配置中的解析方式为准）
4. mutate中的remove_field列表增加除"logdate", "@version", "host", "path", "tags", "message"外的业务无用字段，若无则不增加
5. output内容全部更换为业务对应logstash配置的output中的内容

3. 从CSV文件中读取数据至Elasticsearch中

将csv文件放置到对应位置，启动logstash，查看elasticsearch中是否有数据写入

若无数据写入，一般情况为：logstash启动失败、文件中部分数据格式与elasticsearch定义不符（检查elasticsearch集群日志）

