1. 概述

备份过程通过二步走实现，首先通过logstash将生产环境的ES作为数据源，将日志数据按照索引逐天备份为本地文件（在开发环境的es-node1上进行）。接着，通过该主机的另一个logstash实例，以之前备份好的文件作为数据源，以webhdfs（开发环境的Hadoop集群master）作为输出源，将数据存入hdfs。

1. 实现

logstash的配置文件如下：

*input {*

*elasticsearch {*

*hosts => ["\*.\*.\*.\*:9200"]*

*index => "logstash-2018.01.\*"*

*codec => "json\_lines"*

*docinfo => true*

*}*

*}*

*filter {*

*mutate {*

*lowercase => [ "channelType", "log\_type" ]*

*update => ["message", "%{sourceId} %{callAdvTime} %{orderId} %{orderInputId} %{channelType} %{source} %{userid} %{uuid} %{from\_ip} %{delMode} %{log\_type} %{repeatTime} %{client\_ip} %{scid} %{responseMsg} %{orderSourceId} %{channelTime} %{advId} %{ideaId} %{port} %{appid} %{proId} %{status} %{cid} %{delPlat} %{landingPageId} %{clickTime} %{sc\_name} %{channelMoney} %{inputMoney} %{notifyChannelUrl} %{callAdvUrl} %{callChannelUrl}"]*

*gsub =>["message","%{[A-Za-z\_]\*}", 'null']*

*lowercase => [ "message" ]*

*remove\_field => ["level", "source","userid","tags","sc\_name","log\_type","port","thread\_name","level\_value","appid","@version","host","client\_ip","logger\_name","cid","status"]*

*}*

*}*

*output {*

*if "\_grokparsefailure" in [tags] {*

*stdout {*

*codec => rubydebug*

*}*

*}*

*if "\_grokparsefailure" not in [tags] {*

*file{ path => "/backup/%{+YYYY-MM-dd}/logstash-%{+YYYY-MM-dd}.log"}*

*webhdfs {*

*host => "192.168.1.83"*

*port => 50070*

*path => "/online\_backup/%{+YYYY-MM-dd}/logstash-%{+YYYY-MM-dd}.log"*

*user => "hadoop"*

*compression => "snappy"*

*snappy\_format => "stream"*

*}*

*}*

*}*