vs-b 無法妥當配置

現況說明

- 已在 kafka 建立新的 consumer id 給 b 站使用： `logstash-4-vs-b`

```
## 在 prod-logstash-sys-01/02/03 查看

$ sh /usr/local/kafka/bin/kafka-consumer-groups.sh --list --bootstrap-server `hostname -i`:9092
consumer-weblog
graylog2-service
logstash-4-vs-b
backend-logstash-srv
```

- 確認 b 站 logstash 狀態：運行中，也**持續**有讀取到資料

```
## 在 prod-b-logstash-backend 查看

$ sudo systemctl status logstash -l
● logstash.service - logstash
   Loaded: loaded (/etc/systemd/system/logstash.service; enabled; vendor preset: disabled)
   Active: active (running) since Fri 2022-12-23 11:00:09 CST; 2h 44min ago
 Main PID: 23784 (java)
    Tasks: 53
   Memory: 876.3M
   CGroup: /system.slice/logstash.service
           └─23784 /bin/java -Xms1g -Xmx1g -XX:+UseParNewGC -XX:+UseConcMarkSweepGC -XX:CMSInitiatingOccupancyFraction=75 -XX:+UseCMSInitia

Dec 23 13:44:25 prod-b-logstash-backend logstash[23784]: "createTime" => 1671774263087,
Dec 23 13:44:25 prod-b-logstash-backend logstash[23784]: "device" => "2448692999",
Dec 23 13:44:25 prod-b-logstash-backend logstash[23784]: "fields" => {
Dec 23 13:44:25 prod-b-logstash-backend logstash[23784]: "kafka_topic" => "backend-operator-log"
Dec 23 13:44:25 prod-b-logstash-backend logstash[23784]: },
Dec 23 13:44:25 prod-b-logstash-backend logstash[23784]: "username" => "",
Dec 23 13:44:25 prod-b-logstash-backend logstash[23784]: "content" => "查询游戏注单列表",
Dec 23 13:44:25 prod-b-logstash-backend logstash[23784]: "businessType" => "OTHER",
Dec 23 13:44:25 prod-b-logstash-backend logstash[23784]: "userId" => ""
Dec 23 13:44:25 prod-b-logstash-backend logstash[23784]: }
```

- 在 kafka 端查看 `logstash-4-vs-b` 這 group ID 對於資料的消費情形

```
## 在 prod-logstash-sys-01/02/03 查看
sh /usr/local/kafka/bin/kafka-consumer-groups.sh --bootstrap-server `hostname -i`:9092 --group logstash-4-vs-b --describe

GROUP           TOPIC                PARTITION  CURRENT-OFFSET  LOG-END-OFFSET  LAG             CONSUMER-ID
logstash-4-vs-b backend-operator-log  1          9875926         9875926         0               logstash-1-27b7b9c3-f588-4aa4-84ed-c938b29
logstash-4-vs-b backend-operator-log  0          9876525         9876525         0               logstash-0-7a54b0d4-76a7-4308-976e-34e8029
```

- 查看 logstash 自己的 log，疑似有兩個問題：

```
## 在 prod-b-logstash-backend 查看
## 從最近一次 restart 開始查看

$ grep -A200 '2022-12-23T11:0' /var/log/logstash/logstash-plain.log
[2022-12-23T11:00:03,802][WARN ][logstash.runner          ] SIGTERM received. Shutting down.
[2022-12-23T11:00:08,916][WARN ][logstash.shutdownwatcher ] {"inflight_count"=>0, "stalling_thread_info"=>{}}
[2022-12-23T11:00:08,923][ERROR][logstash.shutdownwatcher ] The shutdown process appears to be stalled due to busy or blocked plugins. Chec
[2022-12-23T11:00:08,960][INFO ][logstash.pipeline        ] Pipeline has terminated {:pipeline_id=>"main", :thread=>"#<Thread:0x2de7c656 ru
[2022-12-23T11:00:21,434][INFO ][logstash.runner          ] Starting Logstash {"logstash.version"=>"6.3.0"}
```

```
[2022-12-23T11:00:24,644][WARN ][logstash.outputs.elasticsearch] You are using a deprecated config setting "document_type" set in elasticse
[2022-12-23T11:00:24,749][INFO ][logstash.pipeline        ] Starting pipeline {:pipeline_id=>"main", "pipeline.workers"=>2, "pipeline.batch
[2022-12-23T11:00:25,153][INFO ][logstash.outputs.elasticsearch] Elasticsearch pool URLs updated {:changes=>{:removed=>[], :added=>[http://
[2022-12-23T11:00:25,162][INFO ][logstash.outputs.elasticsearch] Running health check to see if an Elasticsearch connection is working {:he
[2022-12-23T11:00:25,325][WARN ][logstash.outputs.elasticsearch] Restored connection to ES instance {:url=>"http://10.26.1.170:9200/"}
[2022-12-23T11:00:25,374][INFO ][logstash.outputs.elasticsearch] ES Output version determined {:es_version=>6}
[2022-12-23T11:00:25,378][WARN ][logstash.outputs.elasticsearch] Detected a 6.x and above cluster: the `type` event field won't be used to
[2022-12-23T11:00:25,401][INFO ][logstash.outputs.elasticsearch] New Elasticsearch output {:class=>"LogStash::Outputs::ElasticSearch", :hos
[2022-12-23T11:00:25,416][INFO ][logstash.outputs.elasticsearch] Elasticsearch pool URLs updated {:changes=>{:removed=>[], :added=>[http://
[2022-12-23T11:00:25,417][INFO ][logstash.outputs.elasticsearch] Running health check to see if an Elasticsearch connection is working {:he
[2022-12-23T11:00:25,421][WARN ][logstash.outputs.elasticsearch] Restored connection to ES instance {:url=>"http://10.26.1.170:9200/"}
[2022-12-23T11:00:25,425][INFO ][logstash.outputs.elasticsearch] ES Output version determined {:es_version=>6}
[2022-12-23T11:00:25,425][WARN ][logstash.outputs.elasticsearch] Detected a 6.x and above cluster: the `type` event field won't be used to
[2022-12-23T11:00:25,427][INFO ][logstash.outputs.elasticsearch] New Elasticsearch output {:class=>"LogStash::Outputs::ElasticSearch", :hos
[2022-12-23T11:00:25,485][INFO ][logstash.inputs.tcp      ] Starting tcp input listener {:address=>"0.0.0.0:4560", :ssl_enable=>"false"}
[2022-12-23T11:00:25,645][INFO ][logstash.pipeline        ] Pipeline started successfully {:pipeline_id=>"main", :thread=>"#<Thread:0x4eab9
[2022-12-23T11:00:25,701][INFO ][org.apache.kafka.clients.consumer.ConsumerConfig] ConsumerConfig values:
  auto.commit.interval.ms = 5000
  auto.offset.reset = latest
  bootstrap.servers = [10.23.1.160:9092, 10.23.1.162:9092, 10.23.1.163:9092]
  check.crcs = true
  client.id = logstash-0
  connections.max.idle.ms = 540000
  enable.auto.commit = true
  exclude.internal.topics = true
  fetch.max.bytes = 52428800
  fetch.max.wait.ms = 500
  fetch.min.bytes = 1
  group.id = logstash-4-vs-b
  heartbeat.interval.ms = 3000
  interceptor.classes = null
  internal.leave.group.on.close = true
  isolation.level = read_uncommitted
  key.deserializer = class org.apache.kafka.common.serialization.StringDeserializer
  max.partition.fetch.bytes = 1048576
  max.poll.interval.ms = 300000
  max.poll.records = 500
  metadata.max.age.ms = 300000
  metric.reporters = []
  metrics.num.samples = 2
  metrics.recording.level = INFO
  metrics.sample.window.ms = 30000
  partition.assignment.strategy = [class org.apache.kafka.clients.consumer.RangeAssignor]
  receive.buffer.bytes = 65536
  reconnect.backoff.max.ms = 1000
  reconnect.backoff.ms = 50
  request.timeout.ms = 305000
  retry.backoff.ms = 100
  sasl.jaas.config = null
  sasl.kerberos.kinit.cmd = /usr/bin/kinit
  sasl.kerberos.min.time.before.relogin = 60000
  sasl.kerberos.service.name = null
  sasl.kerberos.ticket.renew.jitter = 0.05
  sasl.kerberos.ticket.renew.window.factor = 0.8
  sasl.mechanism = GSSAPI
  security.protocol = PLAINTEXT
  send.buffer.bytes = 131072
  session.timeout.ms = 10000
  ssl.cipher.suites = null
  ssl.enabled.protocols = [TLSv1.2, TLSv1.1, TLSv1]
  ssl.endpoint.identification.algorithm = null
  ssl.key.password = null
  ssl.keymanager.algorithm = SunX509
  ssl.keystore.location = null
  ssl.keystore.password = null
  ssl.keystore.type = JKS
  ssl.protocol = TLS
  ssl.provider = null
  ssl.secure.random.implementation = null
  ssl.trustmanager.algorithm = PKIX
  ssl.truststore.location = null
  ssl.truststore.password = null
  ssl.truststore.type = JKS
  value.deserializer = class org.apache.kafka.common.serialization.StringDeserializer

[2022-12-23T11:00:25,717][INFO ][logstash.agent           ] Pipelines running {:count=>1, :running_pipelines=>[:main], :non_running_pipelin
[2022-12-23T11:00:25,766][INFO ][org.apache.kafka.common.utils.AppInfoParser] Kafka version : 1.0.0
[2022-12-23T11:00:25,766][INFO ][org.apache.kafka.common.utils.AppInfoParser] Kafka commitId : aaa7af6d4a11b29d
[2022-12-23T11:00:25,769][INFO ][org.apache.kafka.clients.consumer.ConsumerConfig] ConsumerConfig values:
  auto.commit.interval.ms = 5000
  auto.offset.reset = latest
```

```
  bootstrap.servers = [10.23.1.160:9092, 10.23.1.162:9092, 10.23.1.163:9092]
  check.crcs = true
  client.id = logstash-1
  connections.max.idle.ms = 540000
  enable.auto.commit = true
  exclude.internal.topics = true
  fetch.max.bytes = 52428800
  fetch.max.wait.ms = 500
  fetch.min.bytes = 1
  group.id = logstash-4-vs-b
  heartbeat.interval.ms = 3000
  interceptor.classes = null
  internal.leave.group.on.close = true
  isolation.level = read_uncommitted
  key.deserializer = class org.apache.kafka.common.serialization.StringDeserializer
  max.partition.fetch.bytes = 1048576
  max.poll.interval.ms = 300000
  max.poll.records = 500
  metadata.max.age.ms = 300000
  metric.reporters = []
  metrics.num.samples = 2
  metrics.recording.level = INFO
  metrics.sample.window.ms = 30000
  partition.assignment.strategy = [class org.apache.kafka.clients.consumer.RangeAssignor]
  receive.buffer.bytes = 65536
  reconnect.backoff.max.ms = 1000
  reconnect.backoff.ms = 50
  request.timeout.ms = 305000
  retry.backoff.ms = 100
  sasl.jaas.config = null
  sasl.kerberos.kinit.cmd = /usr/bin/kinit
  sasl.kerberos.min.time.before.relogin = 60000
  sasl.kerberos.service.name = null
  sasl.kerberos.ticket.renew.jitter = 0.05
  sasl.kerberos.ticket.renew.window.factor = 0.8
  sasl.mechanism = GSSAPI
  security.protocol = PLAINTEXT
  send.buffer.bytes = 131072
  session.timeout.ms = 10000
  ssl.cipher.suites = null
  ssl.enabled.protocols = [TLSv1.2, TLSv1.1, TLSv1]
  ssl.endpoint.identification.algorithm = null
  ssl.key.password = null
  ssl.keymanager.algorithm = SunX509
  ssl.keystore.location = null
  ssl.keystore.password = null
  ssl.keystore.type = JKS
  ssl.protocol = TLS
  ssl.provider = null
  ssl.secure.random.implementation = null
  ssl.trustmanager.algorithm = PKIX
  ssl.truststore.location = null
  ssl.truststore.password = null
  ssl.truststore.type = JKS
  value.deserializer = class org.apache.kafka.common.serialization.StringDeserializer

[2022-12-23T11:00:25,776][INFO ][org.apache.kafka.common.utils.AppInfoParser] Kafka version : 1.0.0
[2022-12-23T11:00:25,776][INFO ][org.apache.kafka.common.utils.AppInfoParser] Kafka commitId : aaa7af6d4a11b29d
[2022-12-23T11:00:25,924][INFO ][logstash.agent           ] Successfully started Logstash API endpoint {:port=>9600}
[2022-12-23T11:00:25,981][INFO ][org.apache.kafka.clients.consumer.internals.AbstractCoordinator] [Consumer clientId=logstash-1, groupId=lo
[2022-12-23T11:00:25,981][INFO ][org.apache.kafka.clients.consumer.internals.AbstractCoordinator] [Consumer clientId=logstash-0, groupId=lo
[2022-12-23T11:00:25,984][INFO ][org.apache.kafka.clients.consumer.internals.ConsumerCoordinator] [Consumer clientId=logstash-0, groupId=lo
[2022-12-23T11:00:25,984][INFO ][org.apache.kafka.clients.consumer.internals.ConsumerCoordinator] [Consumer clientId=logstash-1, groupId=lo
[2022-12-23T11:00:25,985][INFO ][org.apache.kafka.clients.consumer.internals.AbstractCoordinator] [Consumer clientId=logstash-1, groupId=lo
[2022-12-23T11:00:25,985][INFO ][org.apache.kafka.clients.consumer.internals.AbstractCoordinator] [Consumer clientId=logstash-0, groupId=lo
[2022-12-23T11:00:26,124][INFO ][org.apache.kafka.clients.consumer.internals.AbstractCoordinator] [Consumer clientId=logstash-0, groupId=lo
[2022-12-23T11:00:26,194][INFO ][org.apache.kafka.clients.consumer.internals.AbstractCoordinator] [Consumer clientId=logstash-0, groupId=lo
[2022-12-23T11:00:26,194][INFO ][org.apache.kafka.clients.consumer.internals.AbstractCoordinator] [Consumer clientId=logstash-1, groupId=lo
[2022-12-23T11:00:26,196][INFO ][org.apache.kafka.clients.consumer.internals.ConsumerCoordinator] [Consumer clientId=logstash-1, groupId=lo
[2022-12-23T11:00:26,196][INFO ][org.apache.kafka.clients.consumer.internals.ConsumerCoordinator] [Consumer clientId=logstash-0, groupId=lo
```

- 最後一步總是卡在：**... Setting newly assigned partitions...**
- 前幾步的訊息 **Revoking previously assigned partitions []** 代表：其實先前根本沒有分配 partition 成功過
- 網友討論：可能是回應等待時間過短而導致 timeout，而這不會顯示在這裡；所以有建議把回應時間調大，但是這似乎是治標不治本的方法。

現在的 config 檔案內容

```
## 在 prod-b-logstash-backend 查看
$ cat /etc/logstash/conf.d/kafka-service.conf
input {
  kafka {
    bootstrap_servers => "10.23.1.160:9092,10.23.1.162:9092,10.23.1.163:9092"
    topics => "backend-operator-log"
    group_id => "logstash-4-vs-b"
    consumer_threads => 2
    codec => "json"
    auto_offset_reset => "latest"
    decorate_events => "true"
  }
}

filter {
  if [@metadata][kafka][topic] == "backend-operator-log" {
    # message to json format
      json {
        source => "message"
        remove_field => ["message","@version"]
      }
      mutate {
        add_field => { "[topic_name]" => "%{[@metadata][kafka][topic]}"}
#        replace => { "type" => "logs" }
      }
  }
}


output {
        if [@metadata][kafka][topic] == "backend-operator-log" {
               stdout{codec =>rubydebug}
                elasticsearch {
                        hosts => ["10.26.1.170:9200"]
                        index => "%{indexname}"
                        manage_template => false
                        document_type => "logs"
                }
        }
}
```

- 說明：新增了 `consumer_threads => 2`，是因為知道 backend-operator-log 有兩個 partition，故增加這樣的配置內容。
- 參考來源：

> **曹伟雄**
>
> 所有日志由Rsyslog或者Filebeat收集，然后传输给Kafka，Logstash作为Consumer消费Kafka里边的数据，分别写入Elasticsearch和Hadoop，最后使用Kibana输出到web端供相关人员查看，或者是由Spark接手进入更深层次的分析。 在以上整个架构中，核心的几个组件Kafka、Elasticsearch、Hadoop天生支持高可用，唯独Logstash是不支持的，用单个Logstash去处理日志，不仅存在处理瓶颈更重要的是在整个系统中存在单点的问题， 如果Logstash宕机则将会导致整个集群的不可用，后果可想而知。 如何解
>
> https://www.cnblogs.com/caoweixiong/p/12691458.html