ALIBABA CLOUD

阿里云

阿里云Elasticsearch
Logstash实例

文档版本：20201224

阿里云

法律声明

阿里云提醒您在阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。

2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。

3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。

4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的“现状”、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。

5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含“阿里云”、“Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。

6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

格式	说明	样例
危险	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	危险 重置操作将丢失用户配置数据。
警告	该类警示信息可能会导致系统重大变更甚至故障，或者导致人身伤害等结果。	警告 重启操作将导致业务中断，恢复业务时间约十分钟。
注意	用于警示信息、补充说明等，是用户必须了解的内容。	注意 权重设置为0，该服务器不会再接受新请求。
说明	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	说明 您也可以通过按Ctrl+A选中全部文件。
>	多级菜单递进。	单击**设置**> **网络**> **设置网络类型**。
粗体	表示按键、菜单、页面名称等UI元素。	在**结果确认**页面，单击**确定**。
Courier字体	命令或代码。	执行 `cd /d C:/window` 命令，进入Windows系统文件夹。
斜体	表示参数、变量。	`bae log list --instanceid` *Instance_ID*
[] 或者 [a|b]	表示可选项，至多选择一个。	`ipconfig [-all|-t]`
{} 或者 {a|b}	表示必选项，至多选择一个。	`switch {active|stand}`

目录

1.什么是阿里云Logstash

阿里云Logstash（简称Logstash）作为服务器端的数据处理管道，提供了100%兼容开源Logstash的能力。Logstash能够动态地从多个来源采集数据、转换数据，并且将数据存储到所选择的位置。通过输入、过滤和输出插件，Logstash可以对任何类型的事件加工和转换。

为什么选择阿里云Logstash

阿里云Logstash除了支持所有官方预置插件外，还致力于打造包含logstash-input-sls、logstash-input-oss、logstash-output-oss等适用各类场景的插件中心，为您提供更为强大的数据处理和搬迁能力，实现云上数据生态打通。

在阿里云ELK（Elasicsearch、Logstash、Kibana）生态下，Elasticsearch作为实时分布式搜索和分析引擎，Kibana为Elasticsearch提供了强大的可视化界面，Logstash提供了数据采集、转换、优化和输出的能力，可以被广泛应用于实时日志处理、全文搜索和数据分析等领域。

Logstash数据传输原理

- 数据采集与输入：Logstash支持各种输入选择，能够以连续的流式传输方式，轻松地从日志、指标、Web应用以及数据存储中采集数据。
- 实时解析和数据转换：通过Logstash过滤器解析各个事件，识别已命名的字段来构建结构，并将它们转换成通用格式，最终将数据从源端传输到存储库中。
- 存储与数据导出：Logstash提供多种输出选择，可以将数据发送到指定的地方。

特点与优势

- 快速部署、轻松管理、简化复杂的运维操作。
- 集成官方全部Input、Output、Filter插件。
- 支持Log Service、OSS等阿里云产品输入或输出插件。
- 开放灵活的插件中心。
- 关联Elasticsearch实例进行集中式管道管理。

2.产品定价

2.1. 包年包月

阿里云LogstashService以实例规格和单节点存储空间为单位，对实例进行计费。 同时提供包年包月和按量付费两种计费模式供您选择。本文介绍包年包月的计费详情。

中国内地区域包括：华东 1（杭州）、华东 2（上海）、 华北 1（青岛）、华北 2（北京）、华北 3（张家口）和华南 1（深圳）。全国统一价，各资源使用单价详情如下。

实例规格和价格一览表

华东 1（杭州）、华东 2（上海）、华北 1（青岛）、华北 2（北京）、华南 1（深圳）价格如下。

实例规格族	实例规格	CPU	内存（GB）	价格（元/月）
1:1规格族	logstash.ic5.xlarge	4	4	374
	logstash.ic5.2xlarge	8	8	748
	logstash.ic5.3xlarge	12	12	1122
	logstash.ic5.4xlarge	16	16	1496
1:2规格族	logstash.sn1ne.large	2	4	216.7
	logstash.sn1ne.4xlarge	16	32	1733.6
	logstash.sn1ne.8xlarge	32	64	3467.2
1:4规格族	logstash.sn2ne.large	2	8	298.87
	logstash.sn2ne.xlarge	4	16	597.74
	logstash.sn2ne.2xlarge	8	32	1195.48
	logstash.sn2ne.4xlarge	16	64	2390.96
1:8规格族	logstash.r5.large	2	16	340.67
	logstash.r5.xlarge	4	32	681.34

实例规格族	实例规格	CPU	内存（GB）	价格（元/月）
	logstash.r5.2xlarge	8	64	1362.68

华北 3（张家口）

实例规格族	实例规格	CPU	内存（GB）	价格（元/月）
1:1规格族	logstash.ic5.xlarge	4	4	280.5
	logstash.ic5.2xlarge	8	8	561
	logstash.ic5.3xlarge	12	12	841.5
	logstash.ic5.4xlarge	16	16	1122
1:2规格族	logstash.sn1ne.large	2	4	162.8
	logstash.sn1ne.4xlarge	16	32	1300.2
	logstash.sn1ne.8xlarge	32	64	2600.4
1:4规格族	logstash.sn2ne.large	2	8	224.675
	logstash.sn2ne.xlarge	4	16	448.305
	logstash.sn2ne.2xlarge	8	32	896.61
	logstash.sn2ne.4xlarge	16	64	1793.22
1:8规格族	logstash.r5.large	2	16	256.025
	logstash.r5.xlarge	4	32	511.005
	logstash.r5.2xlarge	8	64	1022.01

存储计费（元/GB/月）

存储类型	华东 1（杭州）	华东 2（上海）	华南 1（深圳）	华北 2（北京）	华北 1（青岛）	华北 3（张家口）
高效云盘	0.35	0.35	0.35	0.35	0.35	0.28
SSD云盘	1	1	1	1	1	0.8

2.2. 按量付费

阿里云LogstashService以实例规格和单节点存储空间为单位，对实例进行计费。 同时提供包年包月和按量付费两种计费模式供您选择。本文介绍按量付费的计费详情。

中国内地区域包括：华东 1（杭州）、华东 2（上海）、 华北 1（青岛）、华北 2（北京）、华北 3（张家口）和华南 1（深圳）。全国统一价，各资源使用单价详情如下。

实例规格和价格一览表

华东 1（杭州）、华东 2（上海）、华北 1（青岛）、华北 2（北京）、华南 1（深圳）价格如下。

实例规格族	实例规格	CPU	内存（GB）	价格（元/小时）
1:1规格族	logstash.ic5.xlarge	4	4	1.298
	logstash.ic5.2xlarge	8	8	2.596
	logstash.ic5.3xlarge	12	12	3.894
	logstash.ic5.4xlarge	16	16	5.192
1:2规格族	logstash.sn1ne.large	2	4	0.748
	logstash.sn1ne.4xlarge	16	32	6.017
	logstash.sn1ne.8xlarge	32	64	12.034
1:4规格族	logstash.sn2ne.large	2	8	1.089
	logstash.sn2ne.xlarge	4	16	2.189
	logstash.sn2ne.2xlarge	8	32	4.367
	logstash.sn2ne.4xlarge	16	64	8.734

实例规格族	实例规格	CPU	内存（GB）	价格（元/小时）
1:8规格族	logstash.r5.large	2	16	1.243
	logstash.r5.xlarge	4	32	2.486
	logstash.r5.2xlarge	8	64	4.983

华北 3（张家口）

实例规格族	实例规格	CPU	内存（GB）	价格（元/小时）
1:1规格族	logstash.ic5.xlarge	4	4	0.979
	logstash.ic5.2xlarge	8	8	1.947
	logstash.ic5.3xlarge	12	12	2.926
	logstash.ic5.4xlarge	16	16	3.894
1:2规格族	logstash.sn1ne.large	2	4	0.561
	logstash.sn1ne.4xlarge	16	32	4.51
	logstash.sn1ne.8xlarge	32	64	9.031
1:4规格族	logstash.sn2ne.large	2	8	0.825
	logstash.sn2ne.xlarge	4	16	1.639
	logstash.sn2ne.2xlarge	8	32	3.278
	logstash.sn2ne.4xlarge	16	64	6.556
1:8规格族	logstash.r5.large	2	16	0.935
	logstash.r5.xlarge	4	32	1.87
	logstash.r5.2xlarge	8	64	3.74

存储计费（元/GB/小时）

存储类型	华东 1（杭州）	华东 2（上海）	华南 1（深圳）	华北 2（北京）	华北 1（青岛）	华北 3（张家口）
高效云盘	0.00049	0.00049	0.00049	0.00049	0.00049	0.00038
SSD云盘	0.0014	0.0014	0.0014	0.0014	0.0014	0.00112

2.3. 按量付费转包年包月

logstash按量付费转包年包月

创建一个按量付费阿里云Logstash实例后，您可以将实例的计费模式转为包年包月，提前预留资源，同时享受更大的价格优惠。本文介绍如何将按量付费实例转换为包年包月。

前提条件

待转换的实例需要满足以下条件：

- 归属于您的账号下。
- 不能有未支付的转换订单。

  如果待转换的Logstash实例有未支付的转换订单，您必须先作废未支付的订单，然后再执行新的转换操作。
- 处于正常状态。

操作步骤

1. 登录阿里云Elasticsearch控制台。
2. 在左侧导航栏，单击**Logstash实例**。
3. 在顶部菜单栏处，选择地域，然后在**实例列表**中单击目标实例ID。
4. 在**实例列表**中，找到待转换的Logstash实例。
5. 在**操作**列中，单击**更多 > 转包年包月**。
6. 在**确认订单**页面，选择购买时长。购买以月为单位，至少购买一个月。
7. 勾选**阿里云LogstashService（包月）服务协议**，单击**去开通**。
8. 按照页面提示完成支付。
   支付成功后，即可完成按量付费转包年包月。

2.4. 欠费与释放说明

当实例欠费后，系统会提醒或通知您。请及时续费，避免对您的服务造成影响。

按量付费

- 欠费即时提醒。
- 欠费24小时后停止服务。
- 停止服务7天后释放阿里云Logstash实例，释放后数据将被永久删除无法恢复。

包年包月

- 服务到期之前7天、3天、1天均会通知。

  > 注意 包年包月类型的实例支持手动续费和自动续费，请及时续费，避免对您的服务造成影响，详情请参见手动续费实例和开通自动续费。

- 服务到期即时停止服务。
- 停止服务7天后释放阿里云Logstash实例，释放后实时提醒，数据将被永久删除无法恢复。

2.5. 续费

2.5.1. 手动续费实例

logstash手动续费

在实例自动释放前，您随时可以手动续费包年包月的阿里云Logstash实例，延长对实例的使用时间。本文介绍如何手动续费实例。

前提条件

已创建包年包月类型的实例。具体操作，请参见创建阿里云Logstash实例。

操作步骤

1. 登录阿里云Elasticsearch控制台。
2. 在左侧导航栏，单击**Logstash实例**。
3. 在顶部菜单栏，选择地域。
4. 在**实例列表**中，找到待续费的Logstash实例，选择一种方式进行续费。
   - 在**操作**列中，单击**更多 > 续费**。
   - 单击实例ID，在**基本信息**页面，单击**续费**。
5. 在**续费**页面，选择**续费时长**。续费以月为单位，至少续费一个月。
6. 勾选**阿里云LogstashService（包月）服务协议**，单击**去支付**。
7. 按照页面提示完成支付。
   支付成功后，即可完成续费。

2.5.2. 开通自动续费

logstash自动续费

自动续费可以减少手动续费的管理成本，避免因忘记手动续费而导致Logstash服务中断，仅支持包年包月类型的实例。本文介绍如何开通自动续费。

在创建实例页面开通自动续费

您可以在创建阿里云Logstash实例页面开通自动续费，如下图所示。更多创建阿里云Logstash实例的信息，请参见创建阿里云Logstash实例。

在实例列表页面开通自动续费

1. 登录阿里云Elasticsearch控制台。
2. 在顶部菜单栏，单击**费用 > 续费管理**。
3. 选择到期时间范围、产品和地域，过滤待操作的阿里云Logstash实例。
4. 在**手动续费**页签，选择一种方式开通自动续费。
   - 为一个Logstash实例开通：找到实例，在**操作**列中，单击**开通自动续费**。
   - 为多个Logstash实例开通：勾选实例，在实例列表底部单击**开通自动续费**。
5. 选择自动续费时长，然后单击**开通自动续费**。
   单击**自动续费**页签，实例出现在列表中即表示已成功开通自动续费。

2.5.3. 取消自动续费

取消logstash自动续费

如果当前计费周期结束后不再需要自动续费实例，您可以提前取消自动续费。开启自动续费后，系统会在实例到期前第9天开始自动扣款。如果需要取消自动续费，请在自动扣款前操作。

取消logstash自动续费

操作步骤

1. 登录阿里云Elasticsearch控制台。
2. 在顶部菜单栏，单击**费用 > 续费管理**。
3. 选择到期时间范围、产品和地域，过滤待操作的阿里云Logstash实例。
4. 单击**自动续费**页签。
5. 选择一种方式取消续费。

取消续费的方式	操作	说明
恢复手动续费	○ 为一个Logstash实例恢复：找到实例，在**操作**列中，单击**恢复手动续费**。 ○ 为多个Logstash实例恢复：勾选实例，在实例列表底部单击**恢复为手动续费**。	恢复后，需要在实例到期前手动续费，详情请参见手动续费实例。
设置到期不续费	○ 为一个Logstash实例设置：找到实例，在**操作**列中，单击**不续费**。 ○ 为多个Logstash实例设置：勾选实例，在实例列表底部单击**设置为不续费**。	选择该方式后，您可以在到期前手动为实例续费，但是实例到期即停止服务并且只会提醒一次，您可以在停止服务前变更该设置。

6. 单击**确认**。
   单击**手动续费**或**到期不续费**页签，实例出现在列表中即表示已成功取消自动续费。

2.6. Logstash产品定价FAQ

本文档为您介绍阿里云Logstash实例定价方面的常见问题。

包年包月的阿里云Logstash实例是否可以退款？

包年包月的阿里云Logstash支持5天内退余款。超过5天后，将不再支持退款。

购买阿里云Logstash实例时，是否有优惠条件？

目前购买Logstash实例时，在相同时长下选择付费方式为**包年包月**会比**按量计费**的价格优惠。并且对于**包年包月**的Logstash实例，当选择购买时长大于等于一年时会有更多的优惠，1年享8.5折，2年享7折，3年享5折。

如何设置阿里云Logstash实例到期自动续费？

自动续费只适用于包年包月实例，具体操作步骤请参见开通自动续费。

3.快速入门

3.1. 入门概述

本教程指引您快速创建一个阿里云Logstash实例，并通过Logstash的管道配置，在阿里云Elasticsearch间同步数据。

背景信息

在开始本教程前，请先了解以下背景信息：

- 什么是阿里云Elasticsearch
- 什么是阿里云Logstash

操作流程

1. 准备工作。

   具体说明请参见准备工作。准备工作包括创建专有网络和虚拟交换机、创建源和目标Elasticsearch实例、开启目标Elasticsearch实例的自动创建索引功能、准备数据。

2. 步骤一：创建阿里云Logstash实例
3. 步骤二：创建并运行管道任务

   创建并配置阿里云Logstash管道任务，运行任务完成数据同步。

4. 步骤三：查看数据同步结果

   通过目标Elasticsearch实例的Kibana控制台，查看数据同步结果。

3.2. 准备工作

本文介绍在开始本教程前，需要完成的准备工作。

创建专有网络和虚拟交换机

创建专有网络和虚拟交换机的具体步骤，请参见搭建IPv4专有网络。

创建阿里云Elasticsearch实例

创建2个阿里云Elasticsearch实例，分别作为Logstash的input和output，创建方式请参见创建阿里云Elasticsearch实例，实例配置如下。

本教程选择实例版本为通用商业版6.7.0，使用的数据迁移方案为阿里云Elasticsearch 6.7.0 > 阿里云Logstash 6.7.0 > 阿里云Elasticsearch 6.7.0，提供的脚本仅适用于该数据迁移方案，其他方案不保证兼容。

> 注意
>
> - 源阿里云Elasticsearch实例需要与Logstash实例在同一VPC下，否则需要配置网络与安全，使用外网访问Logstash，详情请参见配置NAT公网数据传输。
> - 目标阿里云Elasticsearch实例需要与Logstash实例在同一区域、同一可用区、同一VPC下，且版本满足兼容性要求。
> - 访问阿里云Elasticsearch实例的账号，默认为elastic（本文以此为例）。如果需要使用自建用户，要给予自建用户相应的角色和权限，详情请参见创建角色和创建用户。

开启目标Elasticsearch实例的自动创建索引功能

阿里云Elasticsearch为了保证用户操作数据的安全性，默认将**自动创建索引**设置为**不允许**。Logstash在传输数据时，使用的是提交数据的方式创建索引，而不是使用Create index API方式。因此在使用Logstash上传数据前，需要把目标Elasticsearch实例的**自动创建索引**设置为**允许**，详情请参见开启自动创建索引。

准备数据

进入源阿里云Elasticsearch实例的Kibana控制台，在**Dev Tools**页面的**Console**中，执行如下命令创建索引和文档。

> 注意
>
> - 进入Kibana控制台的具体步骤，请参见登录Kibana控制台。
> - 以下示例以Elasticsearch 6.7.0版本为例，仅供测试，不一定适用于其他版本。

1. 创建名称为my_index的索引。

```
PUT /my_index
{
  "settings" : {
   "index" : {
    "number_of_shards" : "5",
    "number_of_replicas" : "1"
   }
  },
  "mappings" : {
    "my_type" : {
      "properties" : {
       "post_date": {
          "type": "date"
       },
       "tags": {
          "type": "keyword"
       },
        "title" : {
          "type" : "text"
        }
      }
    }
  }
}
```

2. 创建名称为1的文档。

```
PUT /my_index/my_type/1?pretty
{
 "title": "One",
 "tags": ["ruby"],
 "post_date":"2009-11-15T13:00:00"
}
```

3. 创建名称为2的文档。

```
PUT /my_index/my_type/2?pretty
{
 "title": "Two",
 "tags": ["ruby"],
 "post_date":"2009-11-15T14:00:00"
}
```

3.3. 步骤一：创建实例

3.3.1. 创建阿里云Logstash实例

在创建并配置阿里云Logstash管道任务前，需要先创建一个实例。本文介绍如何创建阿里云Logstash实例。

前提条件

您已经完成以下操作：

- 注册阿里云账号。

  具体操作，请参见账号注册。

- 开通专有网络和虚拟交换机服务。

  具体操作，请参见创建专有网络和虚拟交换机。

> 注意 目前阿里云Logstash数据推送只支持同一专有网络，相同版本的Elasticsearch。

操作步骤

1. 登录阿里云Elasticsearch控制台。
2. 在左侧导航栏，单击**Logstash实例**。
3. 在**实例列表**页面，单击**创建**。
4. 在购买页面的前三个配置页面，完成实例启动配置。本教程选择实例的付费模式为按量付费，版本为6.7，其余配置均保持默认。更多配置信息，请参见购买页面参数。

   > 说明
   > - 在前期程序研发或功能测试期间，建议购买**按量付费**类型的实例进行测试。
   > - 购买**包年包月**类型实例，可以享受优惠条件。

5. 单击**下一步：确认订单**，预览实例配置。

   配置不符合预期时，可单击图标进行修改。

   本教程的实例配置预览如下图。

6. 勾选**阿里云LogstashService（按量付费）服务协议**，单击**立即购买**。
7. 提示开通成功后，单击**管理控制台**，进入Logstash的**实例列表**页面。

后续步骤

等待实例**状态**变为**正常**，即可开始步骤二：创建并运行管道任务。

3.3.2. 购买页面参数

本文介绍阿里云Logstash购买页面的参数说明。购买实例时，您可以参考本文的说明进行配置。

基础配置

参数	说明
付费模式	支持包年包月和按量付费两种购买方式，请根据需求选择合适的方式： • 按量付费：在前期程序研发或功能测试期间，建议购买按量付费类型的实例进行测试。 支持在控制台手动单击**更多 > 释放实例**，释放实例。 • 目前在购买包年包月类型的实例时，可以享受优惠条件。购买后，支持5天内退余款。超过5天后，将不再支持退款。 支持手动续费和自动续费，详情请参见续费章节。不支持在控制台手动释放实例。
Logstash版本	支持7.4和6.7版本。

集群配置

- **地域和可用区**

  阿里云Logstash支持的地域和可用区如下。

国家	地域	可用区
中国	华北 2（北京）	可用区C、可用区D、可用区E、可用区F、可用区G、可用区H、可用区J
	华东1（杭州）	可用区E、可用区F、可用区G、可用区H、可用区I、可用区J
	华北1（青岛）	可用区B、可用区C
	华东 2（上海）	可用区B、可用区D、可用区E、可用区F、可用区G
	华南 1（深圳）	可用区A、可用区B、可用区C、可用区D、可用区E、可用区F
	华北3（张家口）	可用区A、可用区B、可用区C
	中国（香港）	可用区B、可用区C、可用区D
亚太	新加坡	可用区A、可用区B、可用区C
	澳大利亚（悉尼）	可用区A、可用区B
	马来西亚（吉隆坡）	可用区A、可用区B

国家	地域	可用区
	印度尼西亚（雅加达）	可用区A、可用区B
	日本（东京）	可用区A、可用区B
欧洲与美洲	美国（弗吉尼亚）	可用区A、可用区B
	美国（硅谷）	可用区A、可用区B
	德国（法兰克福）	可用区A、可用区B
	英国（伦敦）	可用区A、可用区B
中东与印度	印度（孟买）	可用区A、可用区B

- **实例规格**

  单击**修改**，可展开Logstash节点配置，并根据需求修改。

参数	说明
规格族	根据节点的CPU和内存配比，阿里云Logstash提供了1:1、1:2、1:4和1:8四种比例的规格族，各规格族支持的规格如下（实际以界面为准）： ○ 1:1规格族：4核4GB、8核8GB、12核12GB、16核16GB ○ 1:2规格族：2核4GB、16核32GB、32核64GB ○ 1:4规格族：2核8GB、4核16GB、8核32GB、16核64GB ○ 1:8规格族：2核16GB、4核32GB、8核64GB
存储类型	支持**SSD云盘**和**高效云盘**： ○ **SSD云盘**（默认）：支持最大2T的存储空间，适合拥有高IOPS，数据响应度较高的在线分析和搜索场景。 ○ **高效云盘**：支持最大5T的存储空间，提供较为低廉的存储能力，适合大规模数据量的日志及分析场景。
单节点存储空间	单节点存储空间与节点的存储类型有关： ○ **SSD云盘**：最大支持2048GB（2T），最小支持20GB。 ○ **高效云盘**：最大支持5120GB（5T）。 **注意** 高效云盘扩容时最大支持扩容到2T。2.5T以上的高效云盘通过磁盘阵列及RAID 0的方式提供服务，不支持扩容。
数量	表示需要购买几个数据节点，可选范围为1~20个。

网络及系统配置

参数	说明
网络类型	目前仅支持**专有网络**。
专有网络	选择对应区域下的专有网络。 注意 如果您需要通过ECS访问Logstash实例，且该ECS位于专有网络下，则Logstash实例与ECS实例必须在同一个专有网络下。
虚拟交换机	只能显示所选的专有网络下，与Logstash实例在相同可用区下的虚拟交换机。
计费周期	仅当**付费模式**为**包年包月**时显示。默认购买时长为一个月。可以自定义选择购买时长（单位：1~9月，1~3年）。
到期自动续费	仅当**付费模式**为**包年包月**时显示。勾选后，可开启自动续费功能。 • 按月购买：自动续费周期为1个月。 • 按年购买：自动续费周期为1年。

订单配置

订单配置中展示了实例的所有配置，可单击图标，修改对应配置。对于包年包月实例，您还可以配置购买时长和到期自动续费。

参数	说明
计费周期	仅当**付费模式**为**包年包月**时显示。默认购买时长为一个月。可以自定义选择购买时长（单位：1~9月，1~3年）。
到期自动续费	仅当**付费模式**为**包年包月**时显示。勾选后，可开启自动续费功能。 • 按月购买：自动续费周期为1个月。 • 按年购买：自动续费周期为1年。

3.4. 步骤二：创建并运行管道任务

实例创建完成后，您可以创建并运行管道任务进行数据同步。本文介绍具体的操作方法。

前提条件

- 完成准备工作。

  具体操作，请参见准备工作。
- 创建阿里云Logstash实例。

  具体操作，请参见创建阿里云Logstash实例。

操作步骤

1. 登录阿里云Elasticsearch控制台。

2. 在左侧导航栏，单击**Logstash实例**。
3. 在顶部菜单栏处，选择地域，然后在**实例列表**中单击目标实例ID。
4. 在左侧导航栏，单击**管道管理**。
5. 在**管道列表**区域，单击**创建管道**。

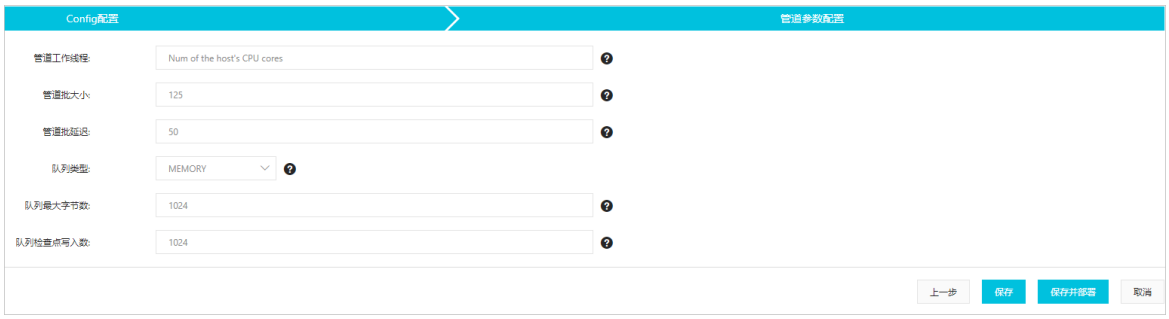
6. 在**Config配置**中，输入**管道ID**并配置管道。本案例使用的配置如下。

```
input {
  elasticsearch {
  hosts => ["http://es-cn-0pp1f1y5g000h****.elasticsearch.aliyuncs.com:9200"]
  user => "elastic"
  password => "your_password"
  index => "*"
  docinfo => true
 }
}
filter {
}
output {
 elasticsearch {
  hosts => ["http://es-cn-mp91cbxsm000c****.elasticsearch.aliyuncs.com:9200"]
  user => "elastic"
  password => "your_password"
  index => "%{[@metadata][_index]}"
  document_type => "%{[@metadata][_type]}"
  document_id => "%{[@metadata][_id]}"
 }
 file_extend {
   path => "/ssd/1/ls-cn-v0h1kzca****/logstash/logs/debug/test"
  }
}
```

参数	说明
hosts	阿里云Elasticsearch服务的访问地址。input中为http://<源实例ID>.elasticsearch.aliyuncs.com:9200；output中为http://<目标实例ID>.elasticsearch.aliyuncs.com:9200。
user	访问阿里云Elasticsearch服务的用户名，默认为elastic。
password	对应用户的密码。elastic用户的密码在创建实例时设定，如果忘记可进行重置，重置密码的注意事项和操作步骤请参见重置实例访问密码。
index	指定同步索引名。设置为%{[@metadata][_index]}，表示匹配元数据中的index，即同步后索引的名称和源索引名称相同。
docinfo	设置为true，将会提取Elasticsearch文档的元信息，例如index、type和id。
document_type	指定同步后索引的类型。设置为%{[@metadata][_type]}，表示匹配元数据中的type，即同步后索引的类型和源索引类型相同。
document_id	指定同步后文档的ID。设置为%{[@metadata][_id]}，表示匹配元数据中的id，即同步后文档的ID和源文档ID相同。
file_extend	开启调试日志功能，并通过path参数配置调试日志的输出路径（使用前需要先安装logstash-output-file_extend插件），详情请参见使用Logstash管道配置调试功能。 **注意** path参数默认为系统指定路径，请勿修改。您也可以单击**开启配置调试**获取path路径。

更多Config配置信息，请参见Logstash配置文件说明。

7. 单击**下一步**，配置管道参数。



管道配置参数说明

参数	说明
管道工作线程	并行执行管道的Filter和Output的工作线程数量。当事件出现积压或CPU未饱和时，请考虑增大线程数，更好地使用CPU处理能力。默认值：实例的CPU核数。

参数	说明
管道批大小	单个工作线程在尝试执行Filter和Output前，可以从Input收集的最大事件数目。较大的管道批大小可能会带来较大的内存开销。您可以设置LS_HEAP_SIZE变量，来增大JVM堆大小，从而有效使用该值。默认值：125。
管道批延迟	创建管道事件批时，将过小的批分派给管道工作线程之前，要等候每个事件的时长，单位为毫秒。默认值：50ms。
队列类型	用于事件缓冲的内部排队模型。可选值： ○ **MEMORY**：默认值。基于内存的传统队列。 ○ **PERSISTED**：基于磁盘的ACKed队列（持久队列）。
队列最大字节数	请确保该值小于您的磁盘总容量。默认值：1024MB。
队列检查点写入数	启用持久性队列时，在强制执行检查点之前已写入事件的最大数目。设置为0，表示无限制。默认值：1024。

警告 配置完成后，需要**保存并部署**才能生效。**保存并部署**操作会触发实例重启，请在不影响业务的前提下，继续执行以下步骤。

8. 单击**保存**或者**保存并部署**。
   - **保存**：将管道信息保存在Logstash里并触发实例变更，配置不会生效。保存后，系统会返回**管道管理**页面。可在**管道列表**区域，单击**操作**列下的**立即部署**，触发实例重启，使配置生效。
   - **保存并部署**：保存并且部署后，会触发实例重启，使配置生效。
9. 在创建成功提示框中，单击**确认**。
   确认后，可在**管道列表**中查看创建成功的管道。等待实例变更完成后，并且管道的**状态**显示为**运行中**时，表示阿里云Logstash开始执行同步任务。

后续步骤

执行步骤三：查看数据同步结果，查看数据同步结果。

3.5. 步骤三：查看数据同步结果

数据同步任务配置完成并开始运行后，您可以通过目标阿里云Elasticsearch的Kibana控制台，查看数据同步结果。

前提条件

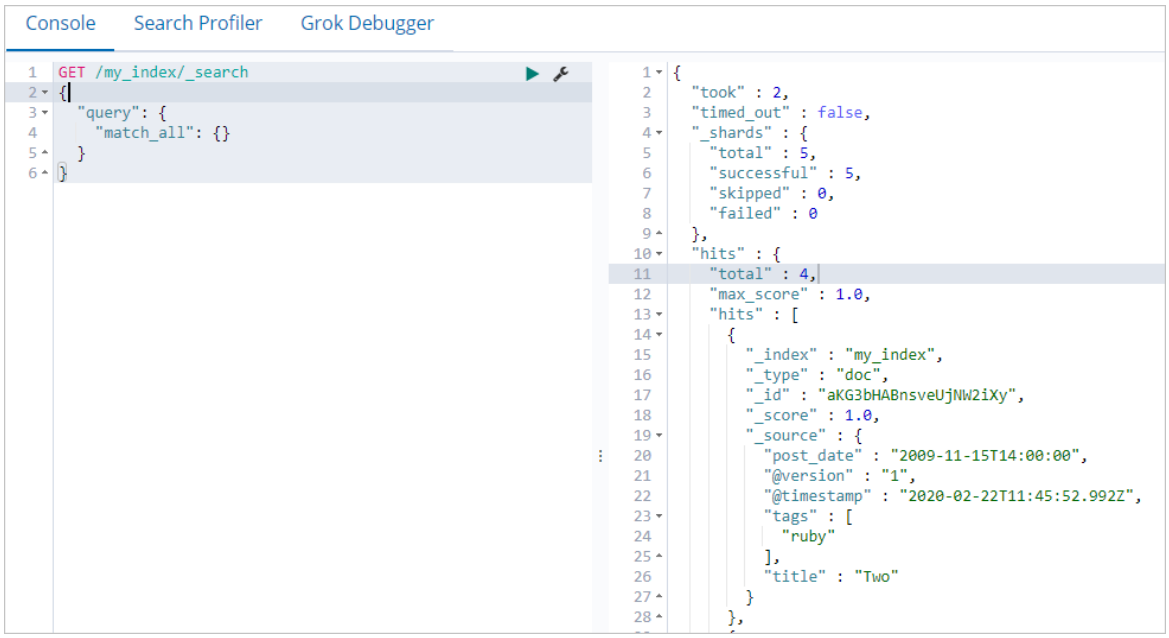
确保阿里云Logstash管道配置中的output为Elasticsearch，详情请参见步骤二：创建并运行管道任务。

操作步骤

1. 登录目标阿里云Elasticsearch实例的Kibana控制台。具体操作，请参见登录Kibana控制台。
2. 在左侧导航栏，单击**Dev Tools**（开发工具）。
3. 在**Console**中执行如下命令，查看数据同步结果。

```
GET /my_index/_search
{
 "query":{
  "match_all":{}
 }
}
```

执行成功后，返回如下结果。



4.实例管理

4.1. 创建实例

本文介绍创建阿里云Logstash实例的方法。

前提条件

您已完成以下操作：

- 注册阿里云账号。

  具体操作，请参见账号注册。

- 开通专有网络和虚拟交换机。

  具体操作，请参见搭建IPv4专有网络。

操作步骤

1. 前往实例创建页面。
2. 在购买页面的前三个配置页面，完成实例启动配置。详细配置信息，请参见购买页面参数。

   > 说明
   > - 在前期程序研发或功能测试期间，建议购买**按量付费**类型的实例进行测试。
   > - 购买**包年包月**类型的阿里云Logstash可以享受优惠条件。

3. 单击**下一步：确认订单**，预览实例配置。

   配置不符合预期时，可单击图标进行修改。

4. 勾选服务协议，单击**立即购买**。
5. 提示开通成功后，单击**管理控制台**，进入阿里云Logstash的**实例列表**页面，查看创建成功的实例。

4.2. 实例列表

logstash实例列表

阿里云Logstash的实例列表展示了实例的基本信息，并提供了创建实例、刷新实例状态、管理管道、管理实例等功能的入口。

logstash实例列表

登录阿里云Logstash控制台，系统直接进入Logstash的**实例列表**页面。**实例列表**页面展示了您账号下当前区域的所有阿里云Logstash实例，并提供了以下操作功能。

在Logstash实例列表页面，您可以完成以下操作。

功能	说明
查看实例的列表信息	包括**实例ID/名称**、**状态**、**版本**、**数据节点数**、**规格**、**可用区**、**付费类型**、**网络类型**和**创建时间**等。
查看实例的基本信息	单击**实例ID/名称**链接，在**基本信息**页面查看实例的基本信息。

功能	说明
创建实例	单击**创建**，可在购买页面购买实例，详情请参见创建实例。
刷新实例	单击**刷新**，可获取实例的实时状态。实例创建后，默认为**待生效**状态，可单击**刷新**查看实例的最新状态，当**状态**变为**正常**时，即可正常使用实例。
管理管道	单击右侧**操作**列下的**管道管理**，可在**管道管理**页面创建并配置管道，详情请参见管道任务管理。
管理实例	单击右侧**操作**列下的**实例管理**，可在实例管理页面查看实例基本信息、进行集群配置、插件配置、集群监控、日志查询、管道管理等操作。
转包年包月	此功能仅适用于按量付费类型的实例。单击右侧**操作**列下的**更多 > 转包年包月**，可在**确认订单**页面变更付费类型，详情请参见按量付费转包年包月。
变更配置	单击右侧**操作**列下的**更多 > 变更配置**，可在**变配**页面，修改集群的配置。
释放实例	单击右侧**操作**列下的**更多 > 释放实例**，在**释放实例**页面确认后，即可释放实例。 警告 实例释放后，数据将不可恢复，请谨慎操作。

4.3. 重启实例或节点

重启logstash

当您修改了实例或节点的配置或进行其他操作时，可能需要重启阿里云Logstash实例或节点才能生效。本文介绍如何通过控制台，重启实例或节点。

前提条件

实例的**状态**为**正常**（绿色），且资源使用率不是很高。

> 说明 资源使用率可在**集群监控**页面查看，例如**节点CPU使用率**为80%左右，**节点HeapMemory使用率**为50%左右，**节点load_1m**低于当前数据节点的CPU核数。详细信息，请参见配置云监控报警。

操作说明

重启分为实例级别重启和节点级别重启，实例级别重启是指重启实例中所包含的所有节点，节点级别重启是指重启所选的单个节点。阿里云Logstash的重启方式和相关注意事项与阿里云Elasticsearch类似。详细信息，请参见重启实例或节点。

4.4. 查看实例任务进度详情

您可以通过任务列表查看正在进行中的任务信息，例如实例的创建进度和重启进度。

前提条件

确保实例处于**生效中**状态。

操作步骤

1. 登录阿里云Elasticsearch控制台。
2. 在左侧导航栏，单击**Logstash实例**。
3. 在顶部菜单栏处，选择地域，然后在**实例列表**中单击目标实例ID。
4. 单击右上角的 ☰ 图标。
5. 在**任务列表**页面，查看**实例变更**进度。
6. 单击**展开详情**，查看各任务的进度详情。

变更过程中，您还可以单击**查看日志**，跳转到**日志查询**页面查看实例的操作日志，详情请参见查询日志。

如果需要暂停变更任务，可单击**中断变更**。变更中断后，可单击**恢复变更**，继续完成之前的实例变更任务。

> 注意
>
> - 实例处于**变更中断**状态时，可能会导致集群服务受到影响，此时可通过二次变更或手动操作恢复变更 。二次变更支持集群升配和插件管理。
> - 触发**恢复变更**操作后，整个重启流程会重新执行一遍，集群中的节点会再进行一次重启，请耐心等待。

5.集群配置

5.1. 配置扩展文件

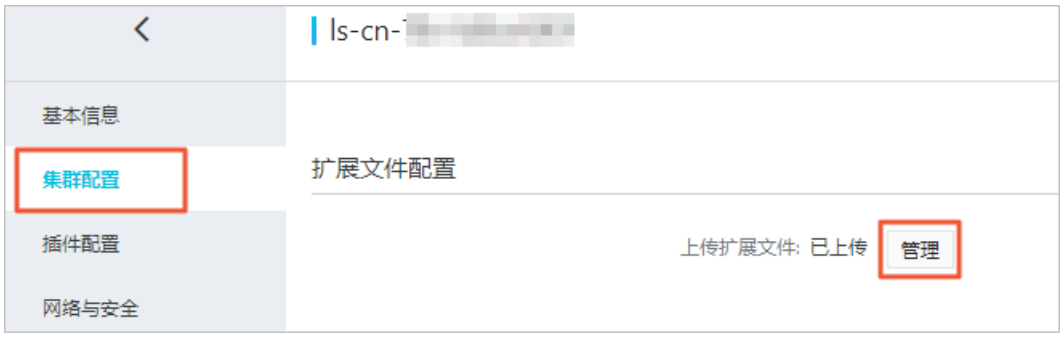
logstash驱动文件

当您需要在阿里云Logstash的配置文件中定义驱动文件时，可通过扩展文件配置功能，上传所需的驱动文件。同时扩展文件配置功能也提供了对所有扩展文件进行管理的能力。

logstash扩展文件 logstash mysql驱动文件

操作步骤

1. 登录阿里云Elasticsearch控制台。
2. 在左侧导航栏，单击**Logstash实例**。
3. 在顶部菜单栏处，选择地域，然后在**实例列表**中单击目标实例ID。
4. 在左侧导航栏，单击**集群配置**。
5. 在**扩展文件配置**区域，单击**上传扩展文件**右侧的**管理**。



6. 在**修改配置**页面，单击下方的**配置**。
7. 单击**上传文件**，在弹出框中选择本地文件进行上传。 阿里云Logstash支持批量上传，且上传前会对文件进行文件名及md5值校验（文件后缀必须是.jar，文件名不支持中文，且长度不超过100个字符），校验失败会进行提示，无法上传。

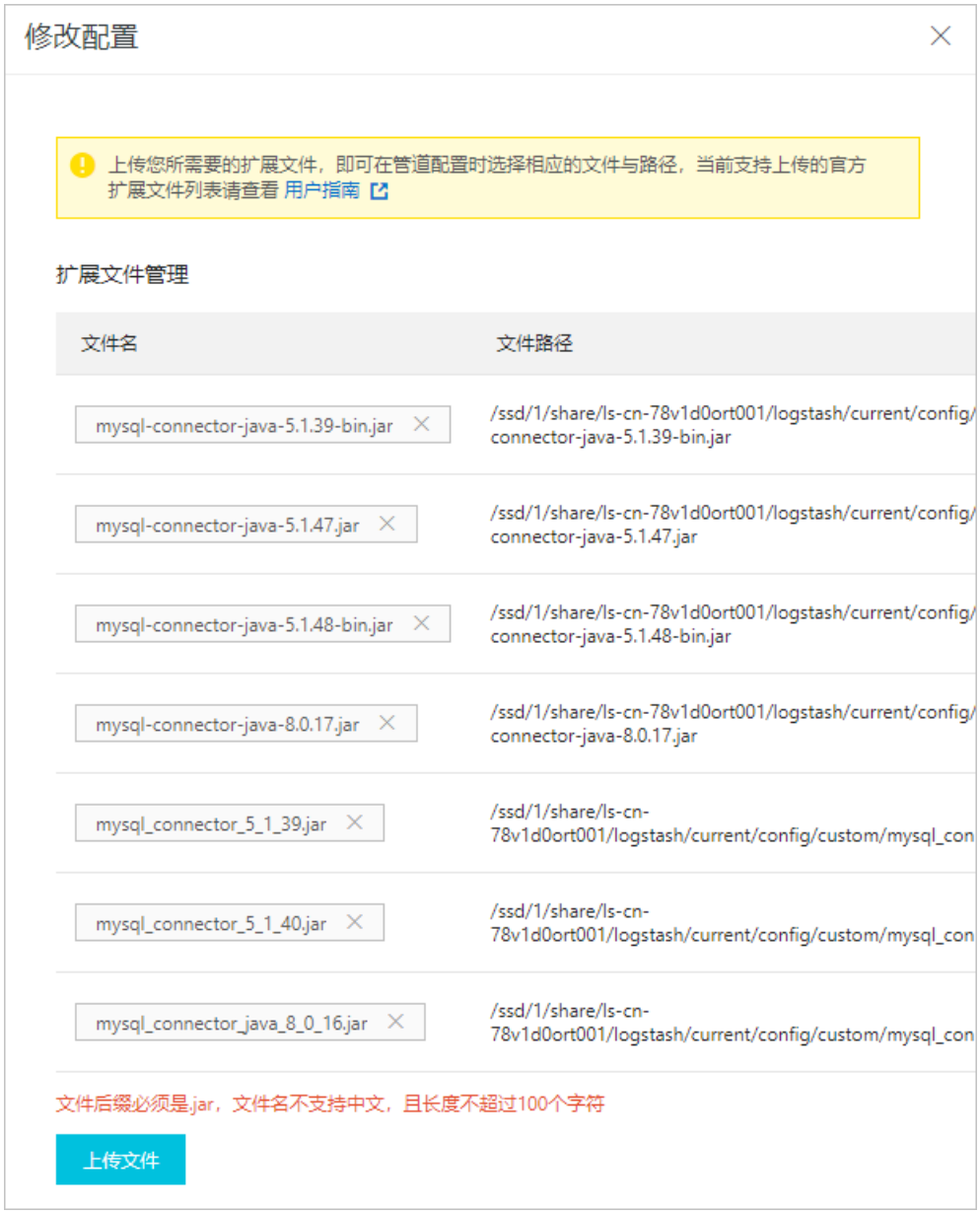
   目前，阿里云Logstash支持MySQL JDBC、PostgreSQL JDBC、PolarDB JDBC驱动文件。

驱动文件类型	驱动文件

驱动文件类型	驱动文件
MySQL JDBC driver	○ mysql-connector-java-5.1.27.jar ○ mysql-connector-java-5.1.35.jar ○ mysql-connector-java-5.1.39-bin.jar ○ mysql-connector-java-5.1.39.jar ○ mysql-connector-java-5.1.43.jar ○ mysql-connector-java-5.1.47.jar ○ mysql-connector-java-5.1.48.jar ○ mysql-connector-java-5.1.9.jar ○ mysql-connector-java-6.0.2.jar ○ mysql-connector-java-6.0.6.jar ○ mysql-connector-java-8.0.11.jar ○ mysql-connector-java-8.0.17.jar ○ mysql-connector-java-8.0.18.jar
PolarDB JDBC driver	○ polardb-jdbc16.jar ○ polardb-jdbc17.jar ○ polardb-jdbc18.jar
PostgreSQL JDBC driver	○ postgresql-42.0.0.jar ○ postgresql-42.1.4.jar ○ postgresql-42.2.0.jar ○ postgresql-42.2.1.jar ○ postgresql-42.2.8.jar ○ postgresql-42.2.10.jar ○ postgresql-42.2.13.jar

警告 修改扩展文件会触发实例重启，请在不影响业务的情况下继续执行以下步骤。

8. 单击**保存**。
   保存后，系统返回**扩展文件配置**页面，并触发集群重启。重启完成后，即可完成扩展文件的添加。
9. （可选）再次单击**上传扩展文件**右侧的**管理**，在**修改配置**页面查看已上传的扩展文件信息。 扩展文件信息包括**文件名**和**文件路径**。单击文件右侧的**X**，可移除对应文件。



注意

- 为了提升安全性，如果在配置管道时使用了JDBC驱动，需要在 `jdbc_connection_string` 参数后面添加 `allowLoadLocalInfile=false&autoDeserialize=false` ，否则在添加Logstash配置文件时，调度系统会抛出校验失败的提示，例如 `jdbc_connection_string => "jdbc:mysql://xxx.drds.aliyuncs.com:3306/test-database?allowLoadLocalInfile=false&autoDeserialize=false"` 。
- 如果不再使用扩展文件，可在**修改配置**页面，单击下方的**配置**，再单击扩展文件右侧的X图标，移除对应的扩展文件。

5.2. 配置YML文件

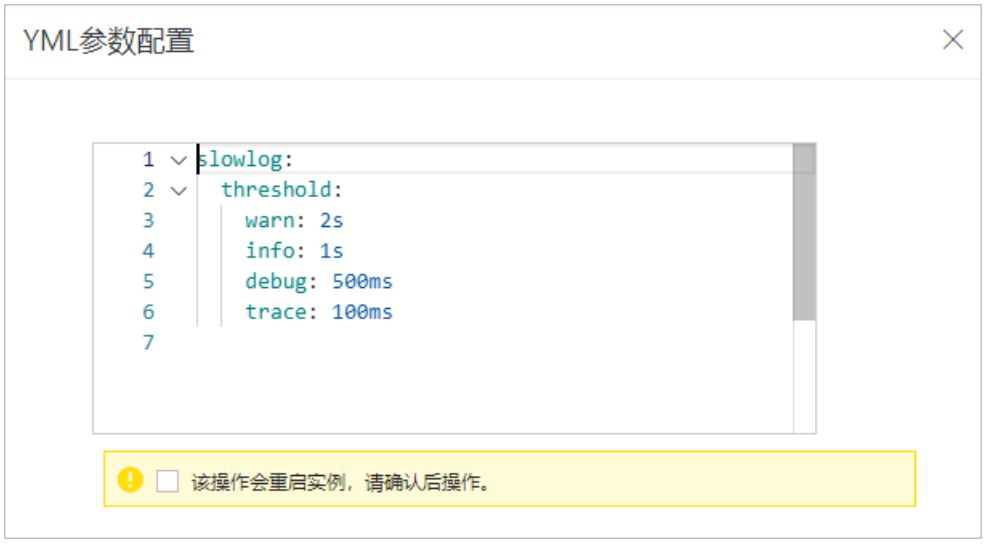
配置logstash yml文件

当您需要通过参数设置来控制阿里云Logstash执行的任务时，可通过Logstash的配置YML文件功能，修改YML文件的参数。

配置logstash yml文件

操作步骤

1. 登录阿里云Elasticsearch控制台。
2. 在左侧导航栏，单击**Logstash实例**。
3. 在顶部菜单栏处，选择地域，然后在**实例列表**中单击目标实例ID。
4. 在左侧导航栏，单击**集群配置**。
5. 在**集群配置**页面，单击**YML文件配置**右侧的**修改配置**。
6. 在**YML参数配置**页面，根据实际业务场景需求修改YML参数配置。



配置参数详情请参见官方Logstash 6.7.0参考文档。

在修改YML参数配置时，请注意：

- 为了方便后续排查与定位阿里云Logstash问题，YML文件配置中默认开启了慢日志，请不要移除该慢日志配置。
- 为了保证服务运行的稳定性，阿里云Logstash不支持修改以下参数值。

```
node.name
path.data
path.config
http.host
http.port
log.level
path.logs
path.plugins
log.format
path.settings
pipeline.workers
xpack.management.enabled
xpack.management.pipeline.id
xpack.management.elasticsearch.username
xpack.management.elasticsearch.password
xpack.management.elasticsearch.hosts
xpack.monitoring.enabled
xpack.monitoring.elasticsearch.username
xpack.monitoring.elasticsearch.password
xpack.monitoring.elasticsearch.hosts
```

警告 修改YML文件需要重启阿里云Logstash实例才能生效，为保证您的业务不受影响，请确认后再执行以下步骤。

7. 勾选**该操作会重启实例，请确认后操作**，单击**确认**。
   确认后，阿里云Logstash实例会进行重启，重启过程中可查看实例任务进度详情。重启成功后，即可完成YML文件的配置。

6.插件配置

6.1. Logstash默认插件列表

本文介绍阿里云Logstash实例默认集成的插件。

> **注意**
>
> - 当input插件需要监听Logstash所在机器的端口时，需要使用8000～9000之间的端口。
> - 阿里云Logstash不支持input的file插件，如有需求，可以使用Filebeat作为本地文件的采集器，以及Logstash的输入源。

类别	名称	说明	介绍
input	azure_event_hubs	使用Azure事件中心中的事件。	Azure Event Hubs plugin
	beats	从Elastic Beats框架中接收事件。	Beats input plugin
	dead_letter_queue	从Logstash的死信队列中读取事件。	Dead_letter_queue input plugin
	elasticsearch	从Elasticsearch集群中读取数据。	Elasticsearch input plugin
	exec	定期运行shell命令，将shell命令的全部输出作为事件捕获。	Exec input plugin
	ganglia	通过用户数据协议UDP（User Datagram Protocol）从网络读取Ganglia包。	Ganglia input plugin
	gelf	在网络上将GELF格式信息作为事件读取。	Gelf input plugin
	generator	生成随机日志事件。	Generator input plugin
	graphite	从Graphite工具读取指标。	Graphite input plugin
	heartbeat	生成心跳消息。	Heartbeat input plugin
	http	通过HTTP或HTTPS接收单行或多行事件。	Http input plugin
	http_poller	调用HTTP API，将输出解码为事件，并发送事件。	Http_poller input plugin
	imap	从IMAP服务器读取邮件。	Imap input plugin
	jdbc	通过JDBC程序界面，将任一数据库数据读取到Logstash中。	Jdbc input plugin

类别	名称	说明	介绍
	kafka	从Kafka主题读取事件。	Kafka input plugin
	pipe	从长时间运行的管道命令中流式读取事件。	Pipe input plugin
	rabbitmq	从RabbitMQ队列中读取事件。	Rabbitmq input plugin
	redis	从Redis实例中读取事件。	Redis input plugin
	s3	从S3 Bucket中的文件流式读取事件。	S3 input plugin
	snmp	使用简单网络管理协议（SNMP）轮询网络设备，获取当前设备操作状态信息。	SNMP input plugin
	snmptrap	将SNMP trap消息作为事件读取。	Snmptrap input plugin
	sqs	从Amazon Web Services简单队列服务（Simple Queue Service, SQS）队列中读取事件。	Sqs input plugin
	stdin	从标准输入读取事件。	Stdin input plugin
	syslog	在网络上将syslog消息作为事件读取。	Syslog input plugin
	tcp	通过TCP套接字读取事件。	Tcp input plugin
	twitter	从Twitter Streaming API接收事件。	Twitter input plugin
	udp	在网络上通过UDP，将消息作为事件读取。	Udp input plugin
	unix	通过UNIX套接字读取事件。	Unix input plugin
	kafka	向Kafka主题写入事件。	Kafka output plugin
	lumberjack	使用lumberjack协议发送事件。	Lumberjack output plugin
	nagios	通过Nagios命令文件，向Nagios发送被动检查结果。	Nagios output plugin
	pagerduty	根据预先配置的服务和升级政策发送通知。	Pagerduty output plugin
	pipe	将事件输送到另一个程序的标准输入。	Pipe output plugin
	rabbitmq	将事件推送到RabbitMQ exchange。	Rabbitmq output plugin

类别	名称	说明	介绍
output	redis	使用RPUSH命令将事件发送到Redis队列。	Redis output plugin
	s3	向亚马逊简单存储服务（Amazon Simple Storage Service, Amazon S3）批量上传Logstash事件。	S3 output plugin
	sns	向采用托管pub/sub框架的亚马逊简单通知服务（Amazon Simple Notification Service）发送事件。	Sns output plugin
	sqs	将事件推送到Amazon Web Services（AWS）SQS队列。	Sqs output plugin
	stdout	将事件打印到运行shell命令的Logstash标准输出。	Stdout output plugin
	tcp	通过TCP套接字写入事件。	Tcp output plugin
	udp	通过UDP发送事件。	Udp output plugin
	webhdfs	通过webhdfs REST API向HDFS中的文件发送Logstash事件。	Webhdfs output plugin
	cloudwatch	聚合并发送指标数据到AWS CloudWatch。	Cloudwatch output plugin
	csv	以逗号分隔（CSV）或其他分隔的格式，将事件写入磁盘。基于文件输出共享配置值。内部使用Ruby CSV库。	Csv output plugin
	elastic_app_search	向Elastic App Search解决方案发送事件。	App Search output plugin
	email	收到输出后发送电子邮件。 您也可以使用条件来包含，或者排除电子邮件输出执行。	Email output plugin
	file	向磁盘文件写入事件。您可以将事件中的字段作为文件名和/或路径的一部分。	File output plugin
	graphite	从日志中读取指标数据，并将它们发送到Graphite工具。Graphite是一个用于存储和绘制指标的开源工具。	Graphite output plugin
	http	向通用HTTP或HTTPS端点发送事件。	Http output plugin
	aggregate	聚合单个任务下多个事件（通常为日志记录）的信息，并将聚合信息推送到最后的任务事件。	Aggregate filter plugin
	anonymize	将字段值替换为一致性哈希值，以实现字段匿名化。	Anonymize filter plugin

类别	名称	说明	介绍
filter			
	cidr	根据网络块列表检查事件中的IP地址。	Cidr filter plugin
	clone	检查重复事件。将为克隆列表中的每个类型创建克隆。	Clone filter plugin
	csv	解析包含CSV数据的事件字段，并将其作为单独字段存储（名字也可指定）。 该过滤器还可以解析带有任何分隔符（不只是逗号）的数据。	Csv filter plugin
	date	解析字段中的日期，然后使用该日期或时间戳作为事件的logstash时间戳。	Date filter plugin
	de_dot	将“.”字符替换为其他分隔符，以重命名字段。 在实际应用中，该过滤器的代价较大。它必须将源字段内容拷贝到新目的字段（该新字段的名字中不再包含点，“.”），然后移除相应源字段。	De_dot filter plugin
	dissect	Dissect过滤器是一种拆分操作。	Dissect filter plugin
	dns	对“reverse”阵列下的各个或指定记录执行DNS查找（A记录或CNAME记录查找，或PTR记录的反向查找）。	Dns filter plugin
	drop	删除满足此过滤器的所有事件。	Drop filter plugin
	elasticsearch	搜索Elasticsearch中的过往日志事件，并将其部分字段复制到当前事件。	Elasticsearch filter plugin
	fingerprint	创建一个或多个字段的一致性哈希值（指纹），并将结果存储在新字段。	Fingerprint filter plugin
	geoip	根据Maxmind GeoLite2数据库的数据添加关于IP地址的地理位置信息。	Geoip filter plugin
	grok	解析任意非结构化文本并将其结构化。	Grok filter plugin
	http	整合外部网络服务或多个REST API。	HTTP filter plugin
	jdbc_static	使用预先从远程数据库加载的数据丰富事件。	Jdbc_static filter plugin
	jdbc_streaming	执行SQL查询，并将结果集存储到“目标”字段。 将结果缓存到具有有效期的本地最近最少使用（LRU）缓存。	Jdbc_streaming filter plugin
	json	JSON解析过滤器，将包含JSON的已有字段扩展为Logstash事件中的实际数据结构。	JSON filter plugin

类别	名称	说明	介绍
	kv	自动解析各种“foo=bar”消息（或特定事件字段）。	Kv filter plugin
	memcached	将外部数据整合到Memcached。	Memcached filter plugin
	metrics	聚合指标。	Metrics filter plugin
	mutate	在字段上执行转变。您可以重命名、删除、更换并修改您事件中的字段。	Mutate filter plugin
	ruby	执行Ruby代码。 该过滤器接受内联Ruby代码或文件。 这两个方案互斥，在工作方式方面略有不同。	Ruby filter plugin
	sleep	按照指定的休眠时间长度休眠。 在休眠时间内，Logstash将停止。这有助于限流。	Sleep filter plugin
	split	通过分解事件的一个字段，并将产生的每个值嵌入原始事件的克隆版，从而克隆事件。被分解的字段可以是字符串或字符串数组。	Split filter plugin
	syslog_pri	解析Syslog（RFC3164）消息前部的“PRI”字段。 如果没有设置优先级，它会默认为13（每个请求注解）。	Syslog_pri filter plugin
	throttle	限制事件的数量。	Throttle filter plugin
	translate	一款普通搜索和替换工具，基于配置的哈希和/或文件决定替换值。	Translate filter plugin
	truncate	截断超过一定长度的字段。	Truncate filter plugin
	urldecode	解码URL编码字段。	Urldecode filter plugin
	useragent	基于BrowserScope数据，将用户代理字符串解析为结构化数据。	Useragent filter plugin
	xml	XML过滤器。将包含XML的字段，扩展为实际数据结构。	Xml filter plugin
	cef	根据《实施 ArcSight 通用事件格式》第二十次修订版（2013 年 6 月 5 日），使用Logstash编解码器处理ArcSight通用事件格式（CEF）。	Cef codec plugin

类别	名称	说明	介绍
codec	collectd	在网络上通过UDP从collectd二进制协议中读取事件。	Collectd codec plugin
	dots	该编解码器生成一个点（.），代表它处理的一个事件。	Dots codec plugin
	edn	读取并产生EDN格式数据。	Edn codec plugin
	edn_lines	读取并产生以新行分隔的EDN格式数据。	Edn_lines codec plugin
	es_bulk	将Elasticsearch bulk格式解码到单独的事件中，并将元数据解码到[@metadata](/metadata)字段。	Es_bulk codec plugin
	fluent	处理fluentd msgpack模式。	Fluent codec plugin
	graphite	编码并解码Graphite格式的行。	Graphite codec plugin
	json	解码（通过输入）和编码（通过输出）完整的JSON消息。	Json codec plugin
	json_lines	解码以新行分隔的JSON流。	Json_lines codec plugin
	line	读取面向行的文本数据。	Line codec plugin
	msgpack	读取并产生MessagePack编码内容。	Msgpack codec plugin
	multiline	将多行消息合并到单个事件中。	Multiline codec plugin
	netflow	解码Netflow v5、v9和v10（IPFIX）数据。	Netflow codec plugin
	plain	处理事件之间没有分隔的明文。	Plain codec plugin
	rubydebug	使用Ruby Awesome Print库输出Logstash事件数据。	Rubydebug codec plugin

6.2. 安装Logstash插件

安装Logstash插件

在使用插件前，您必须先安装插件。本文介绍安装阿里云Logstash插件的方法。

前提条件

已创建阿里云Logstash实例。具体操作，请参见创建阿里云Logstash实例。

操作步骤

1. 登录阿里云Elasticsearch控制台。
2. 在左侧导航栏，单击**Logstash实例**。
3. 在顶部菜单栏处，选择地域，然后在**实例列表**中单击目标实例ID。
4. 在左侧导航栏，单击**插件配置**。
5. 在**系统默认插件列表**中，单击目标插件右侧的**安装**。

> **警告** 安装插件会触发实例重启，请确认后再执行以下步骤。

6. 在**安装提示**对话框中，阅读系统提示，单击**确认**。
   确认后，实例会重启。重启时，可在任务列表中查看任务进度。重启成功后，即可完成插件的安装。

> **说明** 插件安装成功后，如果不再使用，可单击对应插件右侧的**卸载**，使用同样的方式进行卸载。卸载插件也会触发集群重启，请确认后操作。

6.3. logstash-input-sls插件使用说明

logstash-input-sls插件

logstash-input-sls插件是阿里云Logstash自带的默认插件。作为Logstash的input插件，logstash-input-sls插件提供了从日志服务获取日志的功能。

logstash-input-sls插件 logstash从日志服务获取日志

> **说明** logstash-input-sls是阿里云维护的开源插件，详情请参见logstash-input-logservice。

功能特性

- 支持分布式协同消费：可配置多台服务器同时消费一个Logstore服务。
- 高性能：基于Java ConsumerGroup实现，单核消费速度可达20MB/s。
- 高可靠：消费进度保存到服务端，宕机恢复时，会从上一次checkpoint处自动恢复。
- 自动负载均衡：根据消费者数量自动分配shard，消费者增加或退出后会自动进行负载均衡。

前提条件

您已完成以下操作：

- 安装logstash-input-sls插件。

  具体操作步骤请参见安装Logstash插件。
- 创建日志服务项目和Logstore，并采集数据。

  具体操作步骤请参见日志服务快速入门教程。

使用logstash-input-sls插件

满足以上前提条件后，您可以通过配置文件管理管道的方式创建管道任务。在创建管道任务时，按照以下说明配置管道参数。配置完成后进行保存与部署，即可触发Logstash从日志服务获取日志。

以使用阿里云Logstash消费某一个Logstore，并将日志输出到阿里云Elasticsearch为例，配置示例如下。

```
input {
 logservice{
  endpoint => "your project endpoint"
  access_id => "your access id"
  access_key => "your access key"
  project => "your project name"
  logstore => "your logstore name"
  consumer_group => "consumer group name"
  consumer_name => "consumer name"
  position => "end"
  checkpoint_second => 30
  include_meta => true
  consumer_name_with_ip => true
 }
}
output {
 elasticsearch {
   hosts => ["http://es-cn-***.elasticsearch.aliyuncs.com:9200"]
   index => "<your_index>"
   user => "elastic"
   password => "changeme"
 }
}
```

假设某Logstore有10个shard，每个shard的数据流量1M/s；每台阿里云Logstash机器处理的能力为3M/s，可分配5台阿里云Logstash服务器；每个服务器设置相同的 `consumer_group` 和 `consumer_name` ，并且将 `consumer_name_with_ip` 字段设置为 `true` 。

这种情况每台服务器会分配到2个shard，分别处理2M/s的数据。

参数说明

logstash-input-sls支持的参数如下。

参数名	参数类型	是否必填	说明
endpoint	string	是	VPC网络下的日志服务项目的Endpoint，详情请参见经典网络及VPC网络服务入口。
access_id	string	是	阿里云Access Key ID，需要具备ConsumerGroup相关权限，详情请参见使用消费组消费。

参数名	参数类型	是否必填	说明
access_key	string	是	阿里云Access Key Secret，需要具备ConsumerGroup相关权限，详情请参见使用消费组消费。
project	string	是	日志服务项目名。
logstore	string	是	日志服务日志库名。
consumer_group	string	是	自定义消费组名。
consumer_name	string	是	自定义消费者名。同一个消费组内消费者名不能重复，否则会出现未定义行为。
position	string	是	消费位置，可选： • begin：从日志库写入的第一条数据开始消费。 • end：从当前时间点开始消费。 • yyyy-MM-dd HH:mm:ss：从指定时间点开始消费。
checkpoint_second	number	否	每隔几秒checkpoint一次，建议10~60秒，不能低于10秒，默认30秒。
include_meta	boolean	否	传入日志是否包含Meta，Meta包括日志source、time、tag以及topic，默认为 true 。
consumer_name_with_ip	boolean	否	消费者名是否包含IP地址，默认为 true 。分布式协同消费下必须设置为 true 。

性能基准测试信息

- 测试环境
  - 处理器：Intel(R) Xeon(R) Platinum 8163 CPU @ 2.50GHz，4 Core
  - 内存：8GB
  - 环境：Linux
- 阿里云Logstash配置

```
input {
 logservice{
 endpoint => "cn-hangzhou-intranet.log.aliyuncs.com"
 access_id => "***"
 access_key => "***"
 project => "test-project"
 logstore => "logstore1"
 consumer_group => "consumer_group1"
 consumer => "consumer1"
 position => "end"
 checkpoint_second => 30
 include_meta => true
 consumer_name_with_ip => true
 }
}
output {
 elasticsearch {
  hosts => ["http://es-cn-***.elasticsearch.aliyuncs.com:9200"]
  index => "myindex"
  user => "elastic"
  password => "changeme"
 }
}
```

- 测试过程
  i. 使用Java Producer向Logstore发送数据，每秒分别发送2MB、4MB、8MB、16MB、32MB数据。
     每条日志约500字节，包括10个Key和Value对。
  ii. 启动阿里云Logstash消费Logstore中的数据，并确保消费延迟没有上涨（消费速度能够跟上生产的速度）。
- 测试结果

流量（MB/S）	CPU使用率（%）	内存占用量（GB）
32	170.3	1.3
16	83.3	1.3
8	41.5	1.3
4	21.0	1.3
2	11.3	1.3

6.4. logstash-input-oss插件使用说明

logstash-input-oss插件

logstash-input-oss插件基于阿里云消息服务MNS（Message Notification Service），实现了当关联的对象存储服务OSS（Object Storage Service）文件变化时，触发MNS通知阿里云LogstashService（简称Logstash）从OSS文件系统中获取最新的数据。您可以在OSS的事件通知区域，配置当文件发生变化时，自动发送消息给MNS。

logstash-input-oss Logstash OSS事件通知 Logstash MNS

> **说明** logstash-input-oss是阿里云维护的开源插件，详情请参见logstash-input-oss。

注意事项

- 当**logstash-input-oss**插件接收到MNS通知消息后，阿里云Logstash会全量同步关联的文件。
- 如果OSS存储的是.gz或.gzip结尾的文本文件，阿里云Logstash会以.gzip的文件格式对其进行处理，其他格式的文件以文本文件进行处理。
- 文件是以文本文件的方式读取的，如果您的文件是不可解析的格式（例如.jar、.bin等格式），有可能读取出来是乱码。

前提条件

您已完成以下操作：

- 安装logstash-input-oss插件。

  详情请参见安装Logstash插件。
- 开通阿里云OSS服务和阿里云MNS服务，且两者在相同区域。

  详情请参见开通阿里云OSS服务和MNS服务。
- 在OSS中配置事件通知。

  详情请参见配置事件通知。

使用logstash-input-oss插件

满足以上前提条件后，您可以通过配置文件管理管道的方式创建管道任务。在创建管道任务时，请按照以下说明配置管道参数。配置完成后进行保存与部署，即可触发阿里云Logstash从OSS中获取数据。

以从OSS中获取数据，然后写入到阿里云Elasticsearch（简称ES）为例，配置示例如下。

```
input {
 oss {
  endpoint => "oss-cn-hangzhou-internal.aliyuncs.com"
  bucket => "zl-ossou****"
  access_key_id => "******"
  access_key_secret => "**********"
  prefix => "file-sample-prefix"
  mns_settings => {
   endpoint => "******.mns.cn-hangzhou-internal.aliyuncs.com"
   queue => "aliyun-es-sample-mns"
  }
  codec => json {
   charset => "UTF-8"
  }
 }
}
output {
 elasticsearch {
  hosts => ["http://es-cn-***.elasticsearch.aliyuncs.com:9200"]
  index => "aliyun-es-sample"
  user => "elastic"
  password => "changeme"
 }
}
```

注意 MNS Endpoint不能以http为前缀，并且需要internal域名，否则会报错。

参数说明

logstash-input-oss插件支持的参数如下。

参数	类型	是否必选	说明
endpoint	string	是	OSS对外服务的访问域名，详情请参见访问域名和数据中心。
bucket	string	是	OSS的Bucket名称。
access_key_id	string	是	阿里云账号的AccessKey ID。
access_key_secret	string	是	阿里云账号的Access Key Secret。

参数	类型	是否必选	说明
prefix	string	否	如果指定了该参数，则Bucket中文件名的前缀必须与之匹配（不是正则表达式）。
additional_oss_settings	hash	否	附加的OSS客户端配置。可选值：secure_connection_enabled 和 max_connections_to_oss 。
delete	boolean	否	是否从原始Bucket中删除已处理的文件。默认为 false 。
backup_to_bucket	string	否	用来备份已处理过的文件的OSS Bucket名称。
backup_to_dir	string	否	用来备份已经处理过的文件的本地目录路径。
backup_add_prefix	string	否	文件处理后，为key（OSS中包含文件名的完整路径）附加一个前缀。当您将数据备份到另一个（或同一个）Bucket时，这个参数将有效地让您选择一个新的文件夹来放置文件。
include_object_properties	boolean	否	是否在 [@metadata][oss] 中包含OSS对象的属性（last_modified，content_type，metadata）。如果不设置此参数， [@metadata][oss][key] 将始终存在。
exclude_pattern	string	否	要从Bucket中排除的key的ruby正则表达式。

参数	类型	是否必选	说明
mns_settings	hash	是	消息服务（MNS）配置。 可选值及说明如下： • endpoint：MNS端口链接。不能以http为前缀，并且需要internal域名，否则会报错。 • queue：队列名。 • poll_interval_seconds：当队列中没有消息时，针对该队列的ReceiveMessage请求最长的等待时间，默认为10秒。 • wait_seconds：本次ReceiveMessage请求最长的Polling等待时间，单位为秒。 ReceiveMessage的详细信息请参见ReceiveMessage。

常见问题

- Q：为什么基于MNS设计logstash-input-oss插件？

  A：因为OSS文件的变更需要有一种机制通知客户端，而目前OSS文件事件变更可以无缝的写入到MNS中。

- Q：为什么不使用OSS的ListObjects API获取变更的文件？

  A：OSS在记录未处理的文件及已经处理的文件时会增加本地存储，当本地存储较大时，ListObjects API性能会降低。目前其他文件存储系统，如S3开源社区，也将ListObjects API改为了消息通知机制。

6.5. logstash-output-oss插件使用说明

logstash-output-oss插件

通过logstash-output-oss插件，您可以将数据批量传送到阿里云对象存储服务OSS（Object Storage Service）中。本文介绍如何使用logstash-output-oss插件。

> **说明** logstash-output-oss是阿里云维护的开源插件，详细信息，请参见logstash-output-oss。

前提条件

您已完成以下操作：

- 安装logstash-output-oss插件。

  具体操作，请参见安装logstash-output-oss插件。

- 开通阿里云OSS服务。

  具体操作，请参见开通阿里云OSS服务。

- 创建可读写的OSS Bucket，并且获取拥有该Bucket写权限的Accesskey ID和Accesskey Secret。

  具体操作，请参见创建可读写的OSS Bucket。

- 准备输入数据源。

  输入数据源可以为input支持的所有输入源插件中的数据，详细信息，请参见input插件。

使用logstash-output-oss插件

满足以上前提条件后，您可以通过配置文件管理管道的方式创建管道任务。在创建管道任务时，按照以下说明配置Pipeline参数，保存并部署后，即可触发阿里云Logstash向OSS传送数据。

以将Beats采集文件中的数据传送到OSS为例。

```
input {
 beats {
  port => "8044"
  codec => json {
   charset => "UTF-8"
 }
}
output {
 oss {
  endpoint => "http://oss-cn-hangzhou-internal.aliyuncs.com"
  bucket => "zl-log-output-test"
  access_key_id => "LTAIaxxxxx******"
  access_key_secret => "zuxxxx8hBpXs3e6i******"
  prefix => "oss/database"
  recover => true
  rotation_strategy => "size_and_time"
  time_rotate => 1
  size_rotate => 1000
  temporary_directory => "/ssd/1/<Logstash实例ID>/logstash/data/22"
  encoding => "gzip"
  additional_oss_settings => {
   max_connections_to_oss => 1024
   secure_connection_enabled => false
  }
 }
}
```

> 说明
>
> - 阿里云Logstash目前只支持在同一专有网络下进行数据传输，如果源端数据在公网下，请参见配置NAT公网数据传输，在公网环境下进行数据传输。
> - logstash-output-oss插件的具体应用，请参见基于MNS事件通知迁移OSS数据。

参数说明

logstash-output-oss插件支持的参数如下。

参数	类型	是否必选	说明
endpoint	string	是	OSS对外服务的访问域名。详细信息，请参见访问域名和数据中心。
bucket	string	是	OSS的Bucket名称。
access_key_id	string	是	拥有对应Bucket写权限的Accesskey ID。
access_key_secret	string	是	拥有对应Bucket写权限的Accesskey Secret。
prefix	string	否	指定文件名前缀，不指定默认为空。 警告 此选项支持字符串，因此可能会创建很多临时本地文件。
recover	Boolean	否	程序出现异常退出时，保存在本地的数据是否可以继续上传。默认为true。
additional_oss_settings	hash	否	附加的OSS客户端配置。可选值： • server_side_encryption_algorithm：服务端加密方式，只支持AES256。 • secure_connection_enabled：是否开启https，默认false。 • max_connections_to_oss：最大连接数，默认1024。
temporary_directory	string	是	数据上传到OSS之前的临时目录路径定义，必须设置为/ssd/1/<Logstash实例ID>/logstash/data/。任务结束后，一般会在秒级被删除。
rotation_strategy	string	否	文件滚动更新策略。可选值：size、time、size_and_time（默认）。
size_rotate	number	否	如果文件大小大于等于size_rotate，OSS将滚动更新文件（依赖rotation_strategy）。默认为31457280 Bytes。
time_rotate	number	否	如果文件的生存时长大于等于time_rotate，OSS将滚动更新文件（依赖rotation_strategy）。默认为15分钟。
upload_workers_count	number	否	上传线程并发数。

参数	类型	是否必选	说明
upload_queue_size	number	否	上载队列大小。
encoding	string	否	消息在上传文件到OSS之前，支持纯压缩和gzip压缩。可选值：gzip、none（默认）。

临时文件说明

logstash-output-oss在传送数据到OSS时，会在Logstash本地创建一个临时文件。数据临时存储在该文件下，logstash-output-oss插件定期推送数据到OSS。可通过设置temporary_directory参数，设置该临时文件的地址。如果您对输出数据保存的路径有要求，可以设置该临时文件路径。

临时文件路径示例如下。

```
/ssd/1/<Logstash实例ID>/logstash/data/eaced620-e972-0136-2a14-02b7449b****/logstash/1/ls.oss.eaced620-e972-0136-2a14-02b7449b****.2018-12-24T14.27.part-0.data
```

路径	说明
/ssd/1/<Logstash实例ID>/logstash/data/	由temporary_directory指定的临时目录。
eaced620-e972-0136-2a14-02b7449b****	随机UUID。
logstash/1	OSS对象前缀。
ls.oss	临时文件，表示由logstash-output-oss插件生成。
2018-12-24T14.27	临时文件创建的时间。
part-0	临时文件的前缀。
.data	临时文件的后缀。如果设置 `encoding` 为 `gzip` ，将会以 `.gz` 结尾，其他以 `.data` 结尾。

6.6. logstash-input-maxcompute插件使用说明

logstash-input-maxcompute插件

通过logstash-input-maxcompute插件，您可以读取MaxCompute离线表的数据到其他数据源中。

logstash-input-maxcompute插件 读取MaxCompute离线表的数据

前提条件

您已完成以下操作：

- 安装logstash-input-maxcompute插件。

  详情请参见安装Logstash插件。
- 开通阿里云MaxCompute产品，并创建项目、创建表和导入数据。

  详情请参见MaxCompute官方文档的准备工作和快速入门章节。

使用logstash-input-maxcompute插件

满足以上前提条件后，您可以通过配置文件管理管道的方式创建管道任务。在创建管道任务时，按照以下说明配置Pipeline参数，保存并部署后，即可触发阿里云Logstash读取MaxCompute的数据到目标数据源中。

配置脚本如下，相关参数说明请参见参数说明。

```
input {
  maxcompute {
    access_id => "Your accessId"
    access_key => "Your accessKey"
    endpoint => "maxcompute service endpoint"
    project_name => "Your project"
    table_name => "Your table name"
    partition => "pt='p1',dt='d1'"
    thread_num => 1
    dirty_data_file => "/ssd/1/tmp/xxxxx"
  }
}
output {
  stdout {
    codec => rubydebug
  }
}
```

> **注意**
>
> - 目前阿里云Logstash只支持同一专有网络VPC（Virtual Private Cloud）下的数据传输，如果源端数据在公网环境下，请参见配置NAT公网数据传输，通过公网访问Logstash。
> - logstash-input-maxcompute插件会全量同步数据到目标数据源中。

参数说明

logstash-input-maxcompute插件支持的参数如下。

参数	类型	是否必选	说明
endpoint	string	是	MaxCompute对外服务的访问域名，详情请参见开通MaxCompute服务的Region和服务连接对照表。

参数	类型	是否必选	说明
access_id	string	是	阿里云账号的AccessKey ID。
access_key	string	是	阿里云账号的Access Key Secret。
project_name	string	是	MaxCompute的项目名称。
table_name	string	是	MaxCompute的表名称。
partition	string	是	分区字段。分区表按照字段来定义，例如：sale_date='201911'，region='hangzhou'。
thread_num	number	是	线程数，默认为1。
retry_interval	number	否	重试的间隔，单位为秒。
dirty_data_file	string	是	指定文件，用于记录处理失败的日志。

6.7. logstash-input-datahub插件使用说明

logstash-input-datahub插件

通过logstash-input-datahub插件，您可以读取DataHub中的数据到其他数据源中。

前提条件

您已完成以下操作：

- 安装logstash-input-datahub插件。

  详情请参见安装Logstash插件。
- 开通DataHub产品，并完成创建项目、创建Topic和导入数据。

  详情请参见DataHub官方文档的用户指南章节。

使用logstash-input-datahub插件

满足以上前提条件后，您可以通过配置文件管理管道的方式创建管道任务。在创建管道任务时，按照以下说明配置Pipeline参数，保存并部署后，即可触发阿里云Logstash读取DataHub的数据到目标数据源中。

配置脚本如下，相关参数说明请参见参数说明。

```
input {
  datahub {
    access_id => "Your accessId"
    access_key => "Your accessKey"
    endpoint => "Endpoint"
    project_name => "test_project"
    topic_name => "test_topic"
    interval => 5
    #cursor => {
    #   "0"=>"20000000000000000000000003110091"
    #   "2"=>"20000000000000000000000003110091"
    #   "1"=>"20000000000000000000000003110091"
    #   "4"=>"20000000000000000000000003110091"
    #   "3"=>"20000000000000000000000003110000"
    #}
    shard_ids => []
    pos_file => "/ssd/1/ls-cn-v0h1kzca****/logstash/data/文件名"
  }
}
output {
  elasticsearch {
    hosts => ["http://es-cn-mp91cbxsm000c****.elasticsearch.aliyuncs.com:9200"]
    user => "elastic"
    password => "your_password"
    index => "datahubtest"
    document_type => "_doc"
 }
}
```

注意 目前阿里云Logstash只支持同一专有网络VPC（Virtual Private Cloud）下的数据传输，如果源端数据在公网环境下，请参见配置NAT公网数据传输，通过公网访问Logstash。

参数说明

logstash-input-datahub插件支持的参数如下。

参数	类型	是否必选	说明
endpoint	string	是	DataHub对外服务的访问域名，详情请参见DataHub域名列表。
access_id	string	是	阿里云账号的AccessKey ID。

参数	类型	是否必选	说明
access_key	string	是	阿里云账号的Access Key Secret。
project_name	string	是	DataHub的项目名称。
topic_name	string	是	DataHub的Topic名称。
retry_times	number	否	重试次数。-1表示无限重试（默认）、0表示不重试、大于0表示按照设置的次数重试。
retry_interval	number	否	重试的间隔，单位为秒。
shard_ids	array	否	需要消费的shard列表。默认为空，表示全部消费。
cursor	string	否	消费起点。默认为空，表示从头开始消费。
pos_file	string	是	Checkpoint记录文件，必须配置，优先使用checkpoint恢复消费。
enable_pb	boolean	否	是否使用pb传输，默认为true。如果不支持pb传输，请将该参数设置为false。
compress_method	string	否	网络传输的压缩算法，默认不压缩。可选项：lz4、deflate。
print_debug_info	boolean	否	是否打印DataHub的Debug信息，默认为false。设置为true时，会打印大量信息，这些信息仅用来进行脚本调试。

6.8. logstash-output-datahub插件使用说明

logstash-output-datahub插件

通过logstash-output-datahub插件，您可以将数据传输到DataHub中。

前提条件

您已完成以下操作：

- 安装logstash-output-datahub插件。

  详情请参见安装Logstash插件。

- 开通DataHub产品，并完成创建项目和创建Topic。

  详情请参见DataHub官方文档的用户指南章节。

- 准备输入数据源。

  输入数据源可以为input支持的所有输入源插件中的数据，本文以阿里云Elasticsearch为例，详情请参见input插件。

使用logstash-output-datahub插件

满足以上前提条件后，您可以通过配置文件管理管道的方式创建管道任务。在创建管道任务时，按照以下说明配置Pipeline参数，保存并部署后，即可触发阿里云Logstash向DataHub传送数据。

Logstash的Pipeline配置如下，相关参数说明请参见参数说明。

```
input {
  elasticsearch {
    hosts => ["http://es-cn-mp91cbxsm000c****.elasticsearch.aliyuncs.com:9200"]
    user => "elastic"
    index => "test"
    password => "your_password"
    docinfo => true
 }
}
filter{
}
output {
  datahub {
     access_id => "Your accessId"
     access_key => "Your accessKey"
     endpoint => "Endpoint"
     project_name => "project"
     topic_name => "topic"
     #shard_id => "0"
     #shard_keys => ["thread_id"]
     dirty_data_continue => true
     dirty_data_file => "/ssd/1/ls-cn-st21txlz****/logstash/data/文件名"
     dirty_data_file_max_size => 1000
  }
}
```

> **说明** 阿里云Logstash目前只支持在同一专有网络VPC（Virtual Private Cloud）下进行数据传输，如果源端数据在公网下，请参见配置NAT公网数据传输，在公网环境下进行数据传输。

参数说明

logstash-output-datahub插件支持的参数如下。

参数	类型	是否必选	说明
endpoint	string	是	DataHub对外服务的访问域名，详情请参见DataHub域名列表。

参数	类型	是否必选	说明
access_id	string	是	阿里云账号的AccessKey ID。
access_key	string	是	阿里云账号的Access Key Secret。
project_name	string	是	DataHub的项目名称。
topic_name	string	是	DataHub的Topic名称。
retry_times	number	否	重试次数。-1表示无限重试（默认）、0表示不重试、大于0表示按照设置的次数重试。
retry_interval	number	否	重试的间隔，单位为秒，默认为5。
skip_after_retry	boolean	否	当由DataHub异常导致的重试次数超过 retry_times 设置的值，是否跳过这一轮上传的数据。默认为false。
approximate_request_bytes	number	否	用于限制每次发送请求的字节数，是一个近似值，防止因Request body过大而被拒绝接收，默认为2048576（2MB）。
shard_keys	array	否	数据的字段名称，插件会根据这些字段的值计算Hash值，将每条数据写入到某个shard。 **注意** shard_keys 和 shard_ids 都未指定，默认轮询写入各shard。
shard_ids	array	否	所有数据写入指定的shard。 **注意** shard_keys 和 shard_ids 都未指定，默认轮询写入各shard。
dirty_data_continue	string	否	处理数据时遇到脏数据是否继续运行，默认为false。设置为true时，必须指定 dirty_data_file 文件，表示处理数据时忽略脏数据。

参数	类型	是否必选	说明
dirty_data_file	string	否	脏数据文件名称。 当 dirty_data_continue 为true时，必须指定该参数值。 **注意** 处理数据时，脏数据文件会被分割成两个部分part1和part2，part1为原脏数据，part2为替换后的脏数据。
dirty_data_file_max_size	number	否	脏数据文件大小的最大值。
enable_pb	boolean	否	是否使用pb传输，默认为true。如果不支持pb传输，请将该参数值设置为false。

7.网络与安全

7.1. 配置NAT公网数据传输

logstash与公网连通

本文档介绍通过配置NAT网关，实现专有网络VPC（Virtual Private Cloud）下的阿里云Logstash与公网连通的方法。

前提条件

您已完成以下操作：

- 创建专有网络VPC和虚拟交换机。

  具体操作，请参见搭建IPv4专有网络。
- 创建阿里云Logstash实例。

  具体操作，请参见创建阿里云Logstash实例。

操作步骤

1. 登录阿里云Elasticsearch控制台。
2. 在左侧导航栏，单击**Logstash实例**。
3. 在顶部菜单栏处，选择地域，然后在**实例列表**中单击目标实例ID。
4. 在左侧导航栏，单击**网络与安全**。
5. 在**网络配置**区域，单击**前往配置NAT网关**。关于NAT网关的详细说明和配置流程，请参见NAT网关。其中DNAT条目适用于公网服务向Logstash节点推送数据；SNAT条目适用于Logstash主动访问公网。
6. 在NAT网关配置页面，创建NAT网关。创建NAT网关时，所选的区域和VPC需要与阿里云Logstash保持一致。详细创建方法，请参见步骤二：创建NAT网关。
7. 绑定弹性公网IP。
   i. 单击NAT网关列表右侧**操作**列下的 > **绑定弹性公网IP**。
   ii. 在**绑定弹性公网IP**页面，选择**从已有弹性公网IP中选择**。如果还没有EIP，可选择**新购EIP并绑定NAT网关**，按照页面提示完成绑定。
   iii. 选择可用的EIP，单击**确定**。

   > 注意　一个NAT网关最多可绑定20个EIP（最多可绑定10个按流量计费的EIP，每个按流量计费的EIP的最大峰值不能超过200Mbps），您可以提交工单申请更多配额。
8. 创建DNAT条目。
   i. 在NAT网关列表中，单击对应网关右侧**操作**栏下的**设置DNAT**。
   ii. 在**DNAT条目列表**区域，单击**创建DNAT条目**。

iii. 在**创建DNAT条目**页面，填写相关参数。

参数	说明
选择公网IP地址	选择一个可用的公网IP。 **说明** 用于创建SNAT条目的公网IP不能再用来创建DNAT条目。
选择私网IP地址	选择**通过手动输入**，输入Logstash的IP地址。可在Logstash的基本信息页面获取。
端口设置	选择DNAT映射的方式： ■ **任意端口**：该方式属于IP映射，相当于为目标Logstash实例配置了一个弹性公网IP。任何访问该公网IP的请求，都将转发到目标Logstash实例上。 ■ **具体端口**：该方式属于端口映射，NAT网关会将指定协议和端口访问该公网IP的请求，转发到目标Logstash实例的指定端口上。 选择**具体端口**后，请根据业务需求输入**公网端口**（进行端口转发的外部端口）、**私网端口**（进行端口转发的内部端口）和**协议类型**（进行端口转发的协议类型）。
条目名称	输入DNAT条目的名称。 名称长度为2~128个字符，以大小写字母或中文开头， 可包含数字、下划线（_）和短横线（-）。

iv. 单击**确定创建**，完成创建。

9. 创建SNAT条目。

i. 返回NAT网关列表页面，单击对应网关右侧**操作**栏下的**设置SNAT**。

ii. 在**SNAT条目列表**区域，单击**创建SNAT条目**。

iii. 在**创建SNAT条目**页面，单击**交换机粒度**，并填写相关参数。

参数	说明
选择交换机	选择Logstash所属的专有网络VPC中的交换机。该交换机下所有ECS实例，都将通过SNAT功能进行公网访问。
选择公网IP地址	选择用来提供互联网访问的公网IP，支持选择多个公网IP，多个公网IP构建SNAT IP地址池。 当选择多个公网IP地址配置SNAT IP地址池时，请确保每个公网IP地址加入到一个共享带宽中。详情请参见加入共享带宽。

更多参数的详细信息，请参见创建SNAT实现访问公网服务。

iv. 单击**确定创建**，完成创建。

8.集群监控

8.1. 配置云监控报警

配置logstash监控报警

阿里云Logstash支持对实例进行监控，并支持自定义报警阈值以及通过短信接收报警。为避免出现集群状态不正常、节点磁盘使用率过高等问题而影响Logstash服务，强烈建议您进行监控报警配置，实时监控集群状态、节点磁盘使用率等信息，及时查收报警短信，提前做好防御措施。本文介绍如何为Logstash实例配置云监控报警。

背景信息

阿里云Logstash支持以下监控报警项。

监控项	说明
节点磁盘使用率（%）	必选。报警阈值控制在75%以下。
节点HeapMemory使用率（%）	必选。报警阈值控制在85%以下。
节点CPU使用率（%）	可选。报警阈值控制在95%以下。
节点load_1m	可选。以CPU核数的80%为参考值。

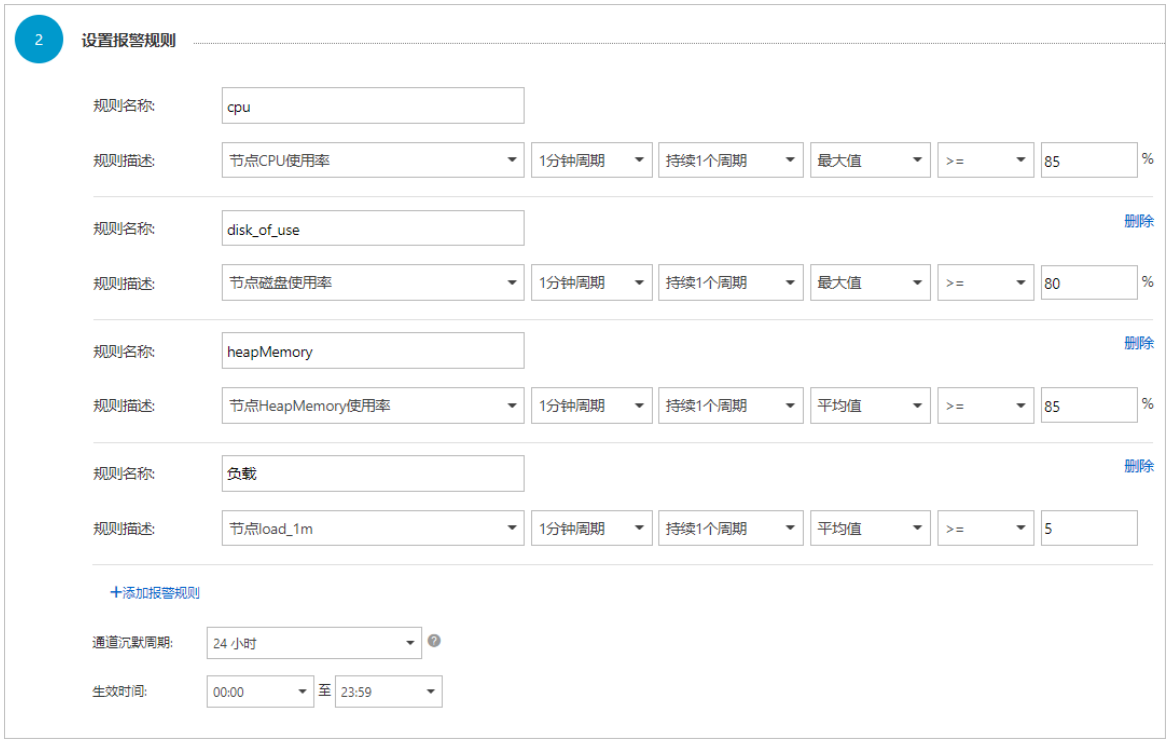
> **注意** 目前Logstash只支持在云监控中配置以上4个监控指标。如果您在配置项中观察到其他指标，请忽略。

操作步骤

1. 进入云监控报警控制台。
   i. 进入云监控控制台。
   ii. 在左侧导航栏，单击**报警服务 > 报警规则**。
   iii. 在**阈值报警**页签中，单击**创建报警规则**。
2. 配置关联资源。

参数	说明
产品	选择**阿里云LogstashService**。
资源范围	选择**实例**。
地域	选择实例所在地域。
实例	选择待监控的实例。

3. 设置报警规则。



通道沉默时间是指同一个指标在一定时间范围内，只会触发一次报警。

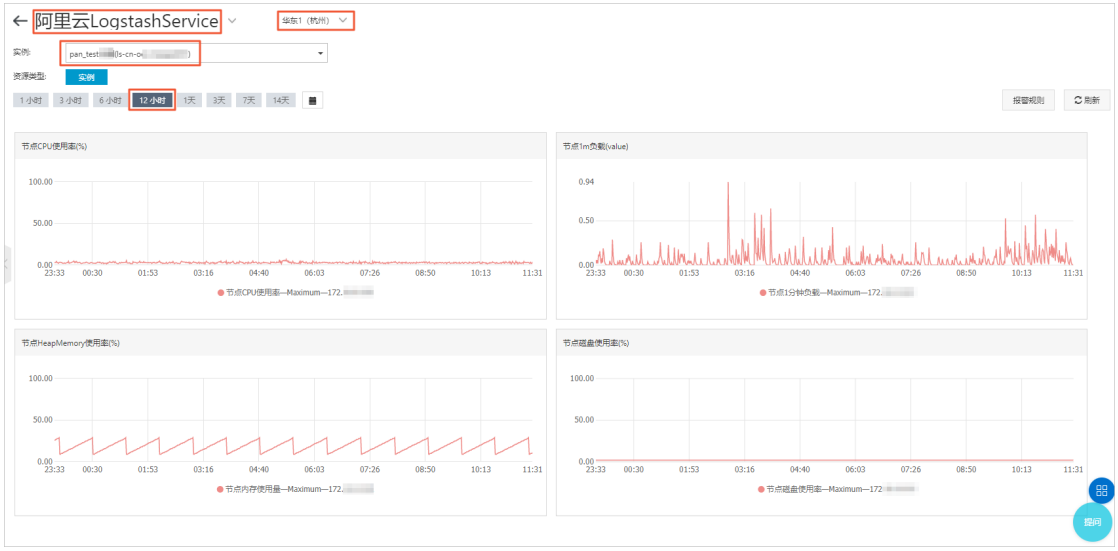
> **说明** 其他参数说明，请参见创建阈值报警规则。

4. 配置告警**通知方式**，选择**云账号报警联系人**。如果您还没有报警联系组，请单击**快速创建联系人组**，进行创建。

> **说明** 您可以在**报警回调**中填写可通过公网访问的URL，云监控会将报警信息通过POST请求推送到该地址，目前仅支持HTTP协议。

5. 单击**确认**。配置完成后，Logstash实例的监控信息将在实例正常运行后开始采集。当指标值超过您设置的报警阈值时，系统会为您发送报警通知。您可以通过以下方式查看Logstash监控大屏：
   i. 在云监控首页的左侧导航栏，单击**Dashboard > 云产品监控大盘**。
   ii. 选择**阿里云LogstashService**产品，并选择地域。

iii. 选择**实例**和监控时间段，查看该段时间内的监控大屏。



8.2. 配置X-Pack监控

本文介绍如何通过配置X-Pack来监控阿里云Logstash服务。开启X-Pack监控，并关联阿里云Elasticsearch实例后，即可在Kibana中监控Logstash服务。

前提条件

您已完成以下操作：

- 创建阿里云Logstash实例。

  具体操作，请参见创建阿里云Logstash实例。

- 创建阿里云Elasticsearch实例，要求与Logstash实例在同一专有网络下，且版本相同。

  具体操作，请参见创建阿里云Elasticsearch实例。

操作步骤

1. 登录阿里云Elasticsearch控制台。
2. 在左侧导航栏，单击**Logstash实例**。
3. 在顶部菜单栏处，选择地域，然后在**实例列表**中单击目标实例ID。
4. 在左侧导航栏，单击**集群监控**。
5. 在**监控报警配置**区域，单击**X-Pack监控**右侧的**修改**。

6. 在**修改配置**页面，开启**X-Pack监控**，并配置要关联的阿里云Elasticsearch实例。

参数	说明
Elasticsearch实例	选择需要关联的阿里云Elasticsearch实例，要求与Logstash实例在同一专有网络下，且版本相同。
用户名	访问阿里云Elasticsearch实例的用户名。
密码	访问阿里云Elasticsearch实例的密码。

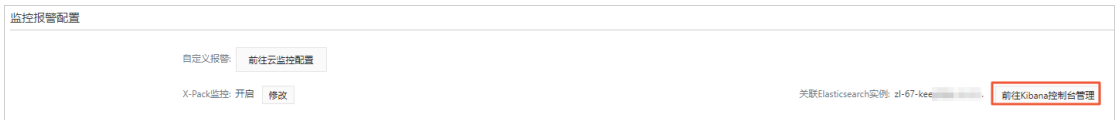
7. 单击**测试连通性**。无报错即连通成功。

> **警告** 修改X-Pack配置会触发实例重启，请在不影响业务的情况下，继续执行以下步骤。

8. 单击**确定**。
   确定后，系统返回**集群监控**页面，并触发实例重启。
9. 等待重启完成后，查看Logstash监控信息。重启完成后，**X-Pack监控**显示为**开启**，且在当前页面显示所关联的阿里云Elasticsearch实例。

> **注意** 重启完成后，才可在Kibana控制台上查看到Logstash监控信息。

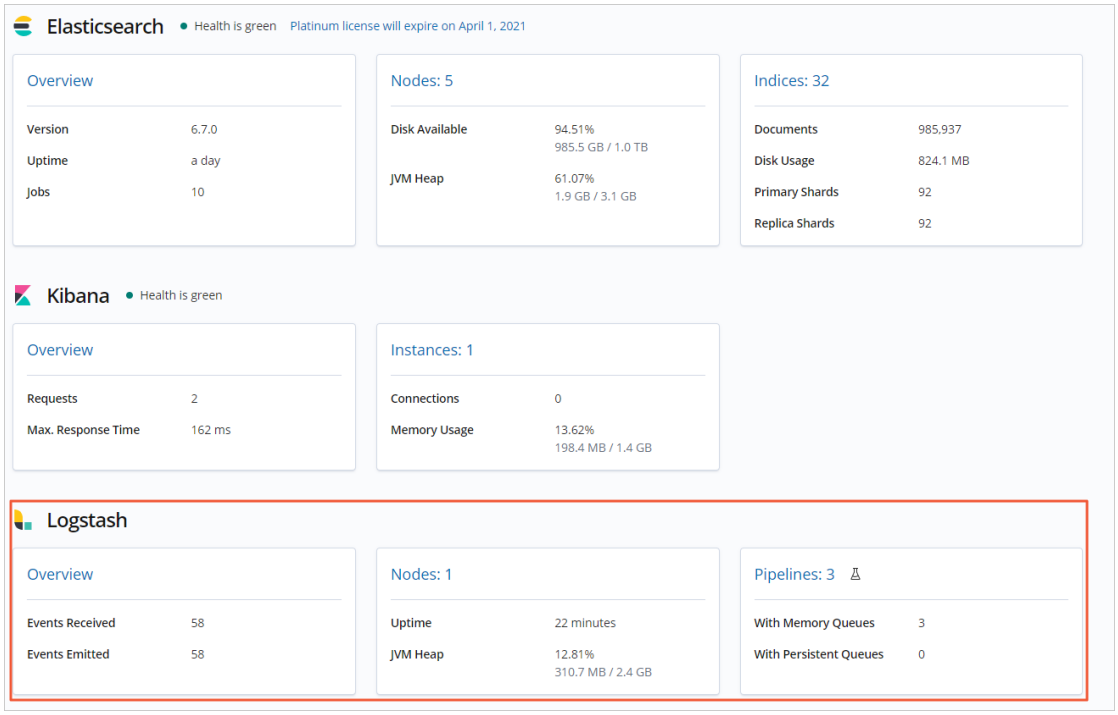
   i. 在**集群监控**页面，单击**前往Kibana控制台管理**。



   ii. 登录Kibana控制台。具体操作，请参见登录Kibana控制台。

   iii. 在左侧导航栏，单击**Monitoring**。

iv. 在Logstash区域，查看Logstash的监控信息。



9.查询日志

logstash日志

阿里云Logstash提供了查询与展示主日志、慢日志、GC日志以及调试日志的功能。通过输入关键字和设置时间范围，就可以快速锁定需要查询的日志内容。最多支持查询连续7天内的日志，日志默认按时间倒序展示。本文介绍如何查询Logstash实例的日志。

操作步骤

本文以查询content包含关键字running，level为info，host为172.16.xx.xx的Logstash主日志为例。

1. 登录阿里云Elasticsearch控制台。
2. 在左侧导航栏，单击**Logstash实例**。
3. 在顶部菜单栏处，选择地域，然后在**实例列表**中单击目标实例ID。
4. 在左侧导航栏，单击**日志查询**。
   系统默认进入**主日志**页签。
5. 在搜索框中输入查询条件。

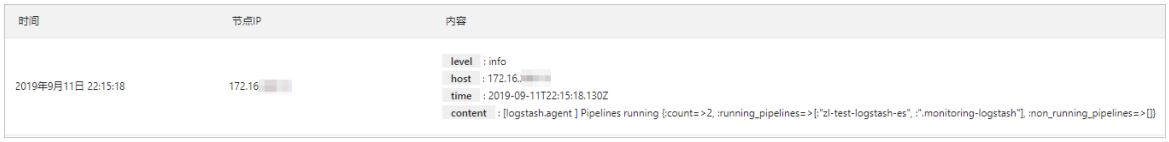
   本文的查询条件为：`host:172.16.xx.xx AND level:info AND content:running`。

   > 注意 查询条件中的`AND`必须为大写。

6. 选择开始时间和结束时间，单击**搜索**。

   > 注意
   > - 如果结束时间为空，那么结束时间默认为当前时间。
   > - 如果开始时间为空，那么开始时间默认为结束时间减去1小时。

   搜索成功后，Logstash会根据您的查询条件返回日志查询结果。日志查询结果主要包括**时间**、**节点IP**和**内容**三部分。

   - **时间**：日志产生时间。
   - **节点IP**：Logstash节点的IP地址。
   - **内容**：主要由level、host、time和content组成。

名称	描述

名称	描述
level	日志级别。包括trace、debug、info、warn、error等内容（GC日志没有 `level` ）。
host	Logstash节点的IP地址。可在实例的基本信息页面获取。
time	日志产生的时间。
content	日志的主要内容。

相关文档

ListLogstashLog

10.管道任务管理

10.1. 通过配置文件管理管道

logstash管道配置

当您需要使用Logstash采集数据时，可通过配置文件方式创建并配置管道，完成数据采集。阿里云Logstash支持多管道并行运行，目前最多支持20个。本文介绍如何通过配置文件管理管道，包括创建管道、修改管道、复制管道和删除管道。

前提条件

您已完成以下操作：

- 创建阿里云Elasticsearch实例，并开启自动创建索引功能。

  创建实例的具体操作，请参见创建阿里云Elasticsearch实例；开启自动创建索引功能的具体操作，请参见开启自动创建索引。

  > **说明** 阿里云Elasticsearch为了保证用户操作数据的安全性，默认将**自动创建索引**配置设置为**不允许**。阿里云Logstash在传输数据的时候，使用提交数据的方式创建索引，而不是Create index API的方式。所以在使用阿里云Logstash上传数据之前，需要先把集群的**自动创建索引**设置为**允许**。

- 创建阿里云Logstash实例。

  具体操作，请参见创建阿里云Logstash实例。

背景信息

Logstash支持通过以下几种方式配置管道：

- 通过配置文件配置Logstash管道（使用 `-f<path/to/file>` 指定配置文件）。
- 通过*pipelines.yml*配置文件，在同一进程中运行多个管道。
- 通过Kibana访问Logstash并配置单进程管道。

创建管道

1. 登录阿里云Elasticsearch控制台。
2. 在左侧导航栏，单击**Logstash实例**。
3. 在顶部菜单栏处，选择地域，然后在**实例列表**中单击目标实例ID。
4. 在左侧导航栏，单击**管道管理**。
5. 在**管道列表**区域，单击**创建管道**。

6. 在**Config配置**中，输入**管道ID**并配置管道。配置示例如下。

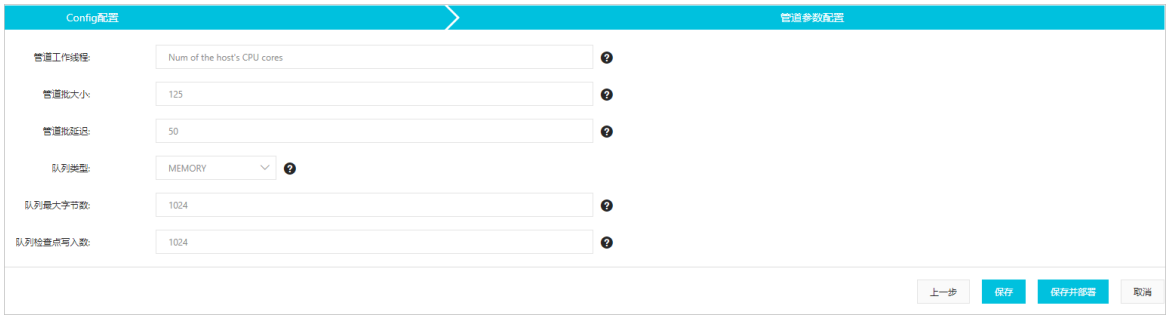
```
input {
  kafka {
  bootstrap_servers => ["192.168.xx.xx:9092,192.168.xx.xx:9092,192.168.xx.xx:9092"]
  group_id => "group_1"
  topics => ["logstash_test"]
  consumer_threads => 6
  decorate_events => true
  }
}
output {
elasticsearch {
hosts => ["http://es-cn-o40xxxxxxxxx****.elasticsearch.aliyuncs.com:9200"]
index => "logstash_test_1"
password => "es_password"
user => "elastic"
}
file_extend {
   path => "/ssd/1/ls-cn-v0h1kzca****/logstash/logs/debug/test"
 }
}
```

参数	说明
input	指定输入数据源。支持的数据源类型，请参见Input plugins。 **说明** ○ Input插件需要监听Logstash进程所在节点的端口，请使用8000~9000范围内的端口。 ○ 如果您需要在input中定义插件、驱动或其他文件，可单击**查看扩展文件路径**，在**本地文件管理**对话框中，单击**前往上传**，根据提示上传对应的文件。详细信息，请参见配置扩展文件。
filter	指定对输入数据进行过滤的插件。支持的插件类型，请参见Filter plugins。
output	指定目标数据源类型。支持的数据源类型，请参见Output plugins。 output中的file_extend参数用来开启调试日志功能，并通过path参数配置调试日志的输出路径。详细信息，请参见使用Logstash管道配置调试功能。 **注意** path参数默认为系统指定路径，请勿修改。您也可以单击**开启配置调试**获取path路径。

管道配置详情，请参见Logstash配置文件说明。

> **注意**
>
> - 为了提升安全性，在使用JDBC驱动并配置管道时，需要在jdbc_connection_string参数后面添加 allowLoadLocalInfile=false&autoDeserialize=false ，否则当您在添加Logstash配置文件的时候，调度系统会抛出校验失败的提示，例如 jdbc_connection_string => "jdbc:mysql://xxx.drds.aliyuncs.com:3306/test-database?allowLoadLocalInfile=false&autoDeserialize=false" 。
> - 如果Config配置中有类似last_run_metadata_path的参数，那么需要阿里云Logstash服务提供文件路径。目前后端开放了*/ssd/1/ls-cn-xxxxxxx/logstash/data/*路径供您测试使用，且该目录下的数据不会被删除。因此在使用时，请确保磁盘有充足的使用空间。
> - 由于阿里云Logstash创建在专有网络下，配置过程中涉及到阿里云系列产品时，建议使用同一专有网络下的实例。如果使用外网访问阿里云Logstash，需要配置网络与安全，详细信息，请参见配置NAT公网数据传输。

7. 单击**下一步**，配置管道参数。



管道配置参数说明

参数	说明
管道工作线程	并行执行管道的Filter和Output的工作线程数量。当事件出现积压或CPU未饱和时，请考虑增大线程数，更好地使用CPU处理能力。默认值：实例的CPU核数。
管道批大小	单个工作线程在尝试执行Filter和Output前，可以从Input收集的最大事件数目。较大的管道批大小可能会带来较大的内存开销。您可以设置LS_HEAP_SIZE变量，来增大JVM堆大小，从而有效使用该值。默认值：125。
管道批延迟	创建管道事件批时，将过小的批分派给管道工作线程之前，要等候每个事件的时长，单位为毫秒。默认值：50ms。
队列类型	用于事件缓冲的内部排队模型。可选值： - **MEMORY**：默认值。基于内存的传统队列。 - **PERSISTED**：基于磁盘的ACKed队列（持久队列）。
队列最大字节数	请确保该值小于您的磁盘总容量。默认值：1024MB。

参数	说明
队列检查点写入数	启用持久性队列时，在强制执行检查点之前已写入事件的最大数目。设置为0，表示无限制。默认值：1024。

> **警告** 配置完成后，需要**保存并部署**才能生效。**保存并部署**操作会触发实例重启，请在不影响业务的前提下，继续执行以下步骤。

8. 单击**保存**或者**保存并部署**。
   - **保存**：将管道信息保存在Logstash里并触发实例变更，配置不会生效。保存后，系统会返回**管道管理**页面。可在**管道列表**区域，单击**操作**列下的**立即部署**，触发实例重启，使配置生效。
   - **保存并部署**：保存并且部署后，会触发实例重启，使配置生效。
9. 在创建成功提示框中，单击**确认**，在**管道列表**中查看创建成功的管道。等待实例重启完成后，即可完成管道任务的创建。

修改管道

> **警告** 修改管道后，在**保存并部署**时会触发实例重启，请在不影响业务的情况下，执行操作。

1. 在**管道列表**区域，单击右侧**操作**列下的**修改管道**。
2. 在**修改管道任务**页面，修改管道的**Config配置**和**管道参数配置**（**管道ID**不可修改）。
3. 单击**保存**或**保存并部署**，等待实例重启完成后，即可完成管道修改。

复制管道

> **警告** 复制管道后，在**保存并部署**时会触发实例重启，请在不影响业务的情况下，执行操作。

1. 在**管道列表**区域，单击右侧**操作**列下的**复制管道**。
2. 在**复制管道任务**页面，输入**管道ID**，其他配置保持不变。
3. 单击**保存**或**保存并部署**，等待实例重启完成后，即可完成管道复制。

删除管道

> **警告**
> - 管道删除后无法恢复，正在运行的管道任务会被中断，请确认后操作。
> - 管道删除操作会触发实例变更，请在不影响业务的情况下，执行操作。

1. 在**管道列表**区域，单击右侧**操作**列下的**更多 > 删除管道**。
2. 在**删除管道**对话框中，查看风险提示。
3. 单击**确定**，等待实例变更完成后，即可删除管道。

10.2. 通过Kibana管理管道（旧实例）

目前，新购实例已不支持使用Kibana管理管道。如果您使用的是旧实例，建议将管道管理方式切换为配置文件管理。通过Kibana管理管道，您可以将阿里云Logstash与阿里云Elasticsearch（简称ES）实例进行关联，并通过该ES实例的Kibana控制台进行集中式管道管理配置，完成数据传输。

前提条件

您已完成以下操作：

- 创建阿里云ES实例。

  详情请参见创建阿里云Elasticsearch实例。

- 创建阿里云Logstash实例。

  详情请参见创建阿里云Logstash实例。

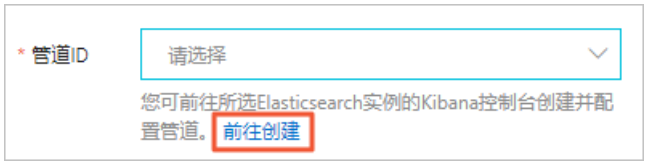
关联Elasticsearch实例

1. 登录阿里云Elasticsearch控制台。
2. 在左侧导航栏，单击**Logstash实例**。
3. 在顶部菜单栏处，选择地域，然后在**实例列表**中单击目标实例ID。
4. 单击左侧导航栏的**管道管理**。
5. 在**管道管理配置**区域，单击**配置文件管理**右侧的**修改**。旧实例的管道管理方式支持Kibana管道管理和配置文件管理（默认）两种方式。由于安全因素限制，新购实例已逐步关闭Kibana管道管理配置，建议优先选择配置文件管理方式，详情请参见通过配置文件管理管道。
6. 在**修改配置**页面，选择**管道管理方式**为**Kibana管道管理**。

   > 警告 更改管道管理方式，会导致原先配置的所有管道失效，正在执行的数据任务受到影响。您需要先删除原有管理方式下的所有管道任务，再进行切换。

7. 选择阿里云ES实例，并输入实例的用户名和密码。用户名为访问所选ES实例的用户名（一般为elastic），密码为创建实例时设置的密码。
8. 单击**测试连通性并获取管道列表**。
   连通成功后，系统显示**管道ID**下拉框。
9. 选择**管道ID**。

   如果您还没有**管道ID**，可单击**前往创建**链接，前往所选ES实例的Kibana控制台创建并配置管道，详情请参见通过Kibana控制台管理管道。



管道创建成功后，返回Logstash管道管理的**修改配置**页面，再次单击**测试连通性并获取管道列表**，获取新创建的管道ID。

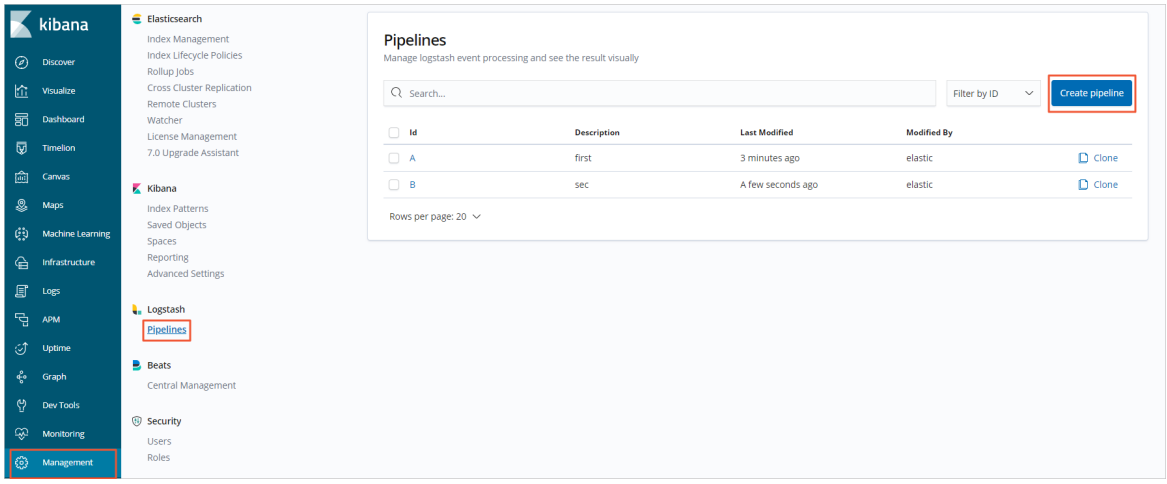
警告 管道配置变更需要重启Logstash进程，请在不影响业务的情况下，继续执行以下操作。

10. 勾选重启Logstash进程注意事项，单击**确定**。
确定后，Logstash会进行重启。重启过程中，可在任务列表中查看重启进度。重启成功后，即可完成Logstash实例与ES实例的关联，并启动相应的数据传输进程。

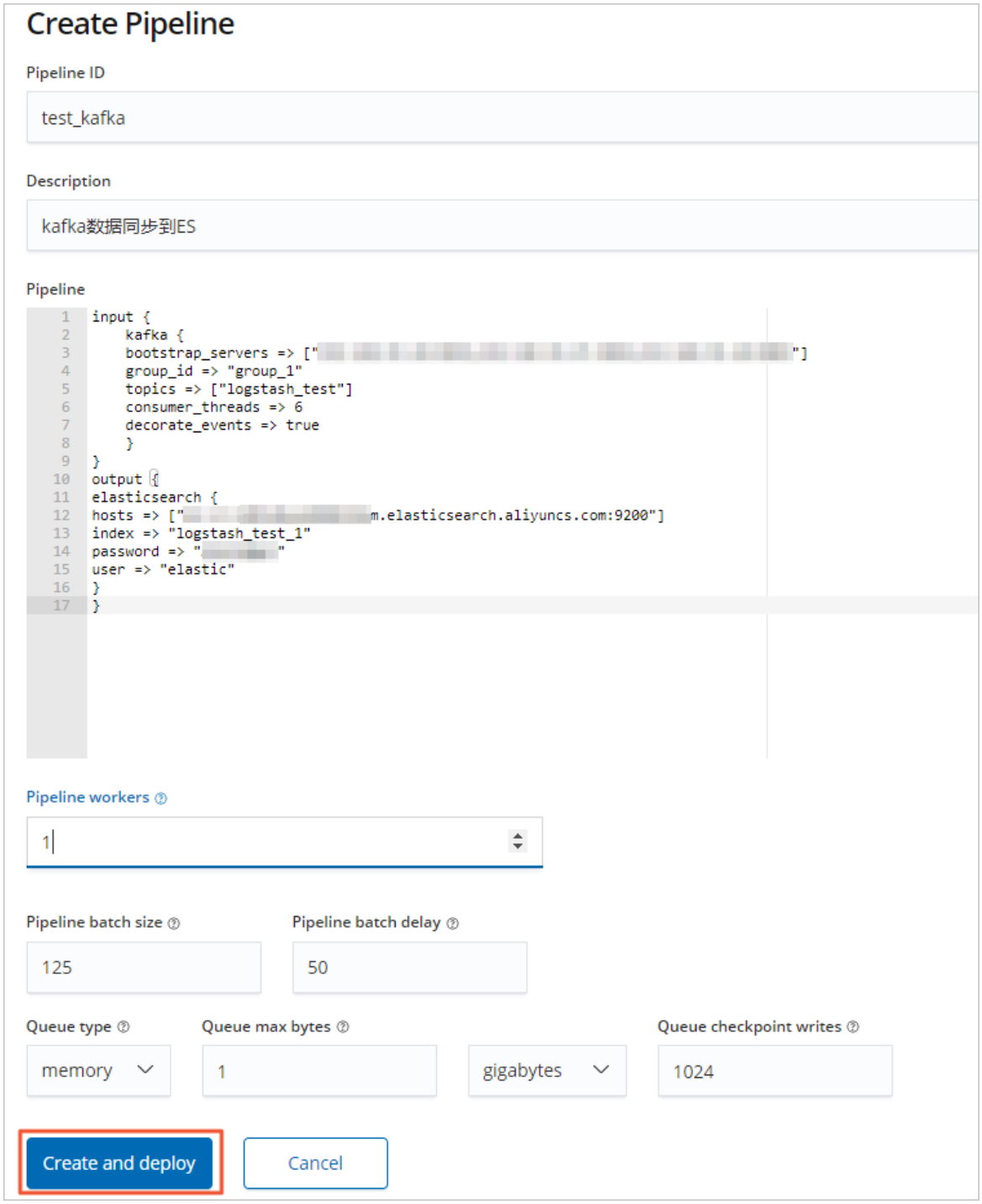
通过Kibana控制台管理管道

当关联Elasticsearch实例成功后，您就可以在所关联ES实例的Kibana控制台中，创建管道或修改管道配置。

1. 在**管道管理**页面，单击**关联Elasticsearch实例**右侧的**前往Kibana控制台管理**。
2. 输入用户名和密码，单击登录。
3. 在Kibana控制台中，单击左侧导航栏的**Management**（管理）。
4. 单击**Logstash**区域下的**Pipelines**（管道）。
5. 在**Pipelines**页面，单击**Create pipeline**（创建管道）。



6. 在Create Pipeline页面，输入Pipeline ID和Description，并根据需求配置其他参数。



配置时，可将鼠标移至参数上，查看相关说明。

参数	说明
Pipeline ID（管道 ID）	管道名称。
Description（描述）	管道配置的描述。

参数	说明

Pipeline（管道）	管道配置。需要配置正确的输入、输出源地址。例如：
	input {
	kafka {
	bootstrap_servers => ["192.168.xx.xx:9092,192.168.xx.xx:9092,192.168.xx.xx:9092"]
	group_id => "group_1"
	topics => ["logstash_test"]
	consumer_threads => 6
	decorate_events => true
	}
	}
	output {
	elasticsearch {
	hosts => ["http://es-cn-o40xxxxxxxxxxxwm.elasticsearch.aliyuncs.com:9200"]
	index => "logstash_test_1"
	password => "es_password"
	user => "elastic"
	}
	}

Pipeline workers（管道工作线程）	用于运行管道的过滤器和输出阶段的并行工作器数。
Pipeline batch size（管道批大小）	单个工作线程在执行过滤器和输出之前收集的最大事件数。
Pipeline batch delay（管道批延迟）	在将小型批处理发送给管道工作者之前等待每个事件的时间（以毫秒为单位）。
Queue type（队列类型）	事件缓冲的内部排队模型。选项是内存中队列的内存，或者是基于磁盘的确认队列的持久性。
Queue max bytes（队列最大字节数）	队列的总容量。默认值为1024MB（1GB）。您可以在右侧下拉列表中选择队列总容量的单位。
Queue checkpoint writes（队列检查点写入数）	启用持久队列时强制检查点之前写入的最大事件数。

7. 单击**Create and deploy**（创建并部署）完成创建。
   创建成功后，系统直接返回**Pipelines**页面，展示创建成功的管道。
8. 单击创建成功的**管道ID**，可在编辑管道页面修改管道配置。

10.3. Logstash配置文件说明

本文档提供阿里云Logstash管道配置文件的详细说明。

您可以通过配置文件管理方式，修改Logstash的配置文件，配置管道，完成数据传输。详情请参见通过配置文件管理管道。

配置文件结构

Logstash配置文件对每种类型的插件都提供了一个单独的配置部分，用于管道事件处理。

```
input {
  ...
}
filter {
  ...
}
output {
  ...
}
```

每个配置部分可以包含一个或多个插件。例如，指定多个filter插件，Logstash会按照它们在配置文件中出现的顺序，进行处理。

> 说明
>
> - 如果配置中有类似 last_run_metadata_path 的参数，那么需要Logstash服务提供文件路径。目前阿里云Logstash后端开放了*/ssd/1/ls-cn-xxxxxxx/logstash/data/*路径供您测试使用，且该目录下的数据不会删除，因此在使用时，请确保磁盘有充足的使用空间。
> - 为了提升安全性，如果在配置管道时使用了JDBC驱动，需要在 jdbc_connection_string 参数后面添加 allowLoadLocalInfile=false&autoDeserialize=false ，否则当您在添加Logstash配置文件的时候，调度系统会抛出校验失败的提示，例如 jdbc_connection_string => "jdbc:mysql://xxx.drds.aliyuncs.com:3306/test-database?allowLoadLocalInfile=false&autoDeserialize=false" 。

插件配置

插件的配置包括插件名称，以及名称中包含的一组插件配置属性。例如，以下input部分包含了两个beats插件，每个beats插件中都配置了 port 和 host 属性。

```
input {
 beats {
  port => 5044
  host => "118.11.xx.xx"
 }
 beats {
  port => 514
  host => "192.168.xx.xx"
 }
}
```

插件支持的配置因插件类型而异，各插件详情请参见输入插件、输出插件、过滤器插件、编解码器插件。

值类型

配置插件时，您可以设置插件的值类型，例如布尔值、列表、哈希等。插件支持以下值类型：

- 数组

  目前不推荐使用此类型，而建议使用标准类型（例如String），使用插件定义 `:list => true` 属性，以便更好地进行类型检查。同时仍然需要处理不需要类型检查的哈希表，或者混合类型列表，示例如下。

  ```
  users => [ {id => 1, name => bob}, {id => 2, name => jane} ]
  ```

- 列表

  列表本身不具备类型特征，但其所包含的属性具有类型特征。这样就可以键入检查多个值。这里可以通过列表的形式，在声明参数时指定启用检查 `:list => true` ，示例如下。

  ```
  path => [ "/var/log/messages", "/var/log/*.log" ]
  uris => [ "http://elastic.co", "http://example.net" ]
  ```

  以上示例，将 `path` 配置为一个列表，该列表中包含2个字符串。 `uris` 也为一个列表，如果所包含的URL无效，会导致事件处理失败。

- 布尔类型

  布尔类型的值必须为 `true` 或者 `false` ，且不需要引号标注，示例如下。

  ```
  ssl_enable => true
  ```

- 字节类型

  字节类型的字段，代表有效字节单位的字符串字段。这是在插件选项中，声明特定大小的便捷方法。字节类型支持SI（k M G T P E Z Y）和Binary（Ki Mi Gi Ti Pi Ei Zi Yi）。二进制单位为base-1024，SI单位为base-1000。该字段不区分大小写，并且接受值和单位之间的空格。如果未指定单位，则整数字符串表示字节数。示例如下。

```
my_bytes => "1113"  # 1113 bytes
my_bytes => "10MiB"  # 10485760 bytes
my_bytes => "100kib" # 102400 bytes
my_bytes => "180 mb" # 180000000 bytes
```

- 编解码器

  编解码器是用于对数据进行编码，或者解码后的目标数据类型，在输入和输出插件中都可以使用。

  输入编解码器，提供了在数据输入之前对其进行解码的功能。输出编解码器，提供了在数据输出之前对其进行编码的功能。使用输入或输出编解码器后，您不需要在Logstash管道中，单独使用过滤器。

  您可以参考官方提供的编译解释器插件文档，查找可用的编解码器。

  示例如下。

  ```
  codec => "json"
  ```

- 哈希

  哈希格式指定键值对的集合，例如 `"field1" => "value1"` 。

  > 注意 多个键值对使用空格进行分隔，而不是逗号。

  示例如下。

  ```
  match => {
   "field1" => "value1"
   "field2" => "value2"
   ...
  }
  # or as a single line. No commas between entries:
  match => { "field1" => "value1" "field2" => "value2" }
  ```

- 数值

  必须是有效的数值（浮点数或整数）。

  示例如下。

  ```
  port => 33
  ```

- 密码

  密码是一个没有记录或打印的单值字符串。

  示例如下。

  ```
  my_password => "password"
  ```

- URL

URL可以是任何内容，从完整的URL到简单的标识符（如foobar）。如果URL中包含类似 `http://user:paas@example.net` 的密码，那么密码部分不会被记录或打印。

```
my_uri => "http://foo:bar@example.net"
```

- 路径

  路径是代表有效操作系统路径的字符串。

  示例如下。

```
my_path => "/tmp/logstash"
```

- 字符串

  字符串必须是单个字符序列，且必须用双引号或单引号括起来。

- 转义序列

  默认情况下，Logstash不启用转义序列。如果您希望在带引号的字符串使用转义序列，需要在*logstash.yml*中设置 `config.support_escapes: true` 。当设置为 `true` 时，带引号的字符串（双精度和单精度）可以进行一下转换。

文本	结果
\r	回车（ASCII 13）
\n	换行（ASCII 10）
\t	跳格（ASCII 9）
\\	反斜杠（ASCII 92）
\"	双引号（ASCII 34）
\'	单引号（ASCII 39）

  示例如下。

```
name => 'It\'s a beautiful day'
```

10.4. 使用Logstash管道配置调试功能

Logstash管道配置调试

当Logstash管道配置出现错误时，可能导致输出数据结果不符合预期，需要反复去目标端确认数据格式，再返回控制台修改配置，这样会耗费大量的时间和人力。对于这种情况，您可以通过阿里云Logstash提供的管道配置调试功能，在管道配置完成后，直接在控制台上查看管道配置的输出结果，降低调试成本。本文介绍具体的实现方法。

前提条件

已安装logstash-output-file_extend插件。如果还未安装，请先安装该插件，安装方法请参见安装Logstash插件。

操作步骤

1. 登录阿里云Elasticsearch控制台。
2. 在左侧导航栏，单击**Logstash实例**。
3. 在顶部菜单栏处，选择地域，然后在**实例列表**中单击目标实例ID。
4. 在左侧导航栏，单击**管道管理**。
5. 单击**创建管道**。
6. 配置并启动管道。
   i. 在**Config配置**中，填写**管道ID**和**Config配置**。

参数	说明
管道ID	自定义输入。输入后，管道ID会自动映射到file_extend的path路径下。
Config配置	**Config配置**由三部分组成： ■ input：指定待读取的数据源，支持Logstash自带的input plugins（除过file插件）。 ■ filter：进一步加工处理数据源采集到的数据，支持丰富的Filter plugins。 ■ output：将过滤后的数据发送到特定的目的端。阿里云Logstash不仅支持开源的 Logstash output plugins，还可通过配置特有的file_extend插件，开启调试日志功能，即可在管道部署完成后直接查看输出结果，并进行验证与调试。

config配置示例如下。

```
input {
  elasticsearch {
   hosts => "http://es-cn-0pp1jxv000****.elasticsearch.aliyuncs.com:9200"
   user  => "elastic"
   index => "twitter"
   password => "<your_password>"
   docinfo => true
   schedule => "* * * * *"
  }
}
filter {
}
output {
  elasticsearch {
  hosts => ["http://es-cn-000000000i****.elasticsearch.aliyuncs.com:9200"]
  user => "elastic"
  password => "<your_password>"
  index => "%{[@metadata][_index]}"
  document_id => "%{[@metadata][_id]}"
 }
  file_extend {
   path => "/ssd/1/ls-cn-v0h1kzca****/logstash/logs/debug/test"
  }
}
```

> 注意
>
> - output中的file_extend配置默认为注释状态，如果需要使用调试功能，请先删除注释。
> - file_extend中的path参数默认为系统指定路径，请勿修改。您也可以单击**开启配置调试**获取path路径。
> - path中的{pipelineid}将自动映射为管道ID，请勿修改为其他名称，否则无法获取调试日志。

ii. 单击**下一步**，配置管道参数。管道配置参数的详细信息，请参见通过配置文件管理管道。

iii. 保存并部署管道。

- **保存**：将管道信息保存在Logstash里并触发实例变更，配置不会生效。保存后，系统会返回**管道管理**页面。可在**管道列表**区域，单击**操作**列下的**立即部署**，触发实例重启，使配置生效。
- **保存并部署**：保存并且部署后，会触发实例重启，使配置生效。

iv. 在创建成功提示框中，单击**确认**。

7. 查看调试日志。

i. 等待实例重启完成后，在**管道列表**中，单击目标管道右侧**操作**列下的**查看调试日志**。

ii. 在**日志查询**页面的**调试日志**页签中，获取管道处理后的输出数据。对于多个管道，您可在搜索框中输入pipelineId: <管道ID>过滤对应的日志。

11.访问控制

12.Logstash FAQ

本文介绍使用阿里云Logstash的常见问题。

Logstash是否支持将数据源配置为DRDS?

支持。可参考RDS MySQL数据迁移方案进行配置，具体操作步骤请参见通过Logstash将RDS MySQL数据同步至Elasticsearch。

如何将公网数据导入或导出到Logstash中?

Logstash实例部署在专有网络VPC（Virtual Private Cloud）下，可以通过配置NAT网关实现与公网的连通，详情请参见配置NAT公网数据传输。

使用自建Kafka作为Logstash的输入或者输出时，出现如下 `No entry found for connection` 错误日志，如何处理?

详细错误信息如下。

```
[2019-09-22T10:01:55,914][INFO ][org.apache.kafka.clients.consumer.internals.AbstractCoordinator] [Consumer clientId=logstash-3, groupId=group_1] Discovered group coordinator iZbp15qsax98n3ho*****:9092 (id: 2147483646 rack: null)
// 省略若干行日志
Error: No entry found for connection 2147483646
Exception: Java::JavaLang::IllegalStateException
```

原因：Logstash节点无法解析到Kafka服务的hostname对应的IP地址。

解决方法：请在 `server.properties` 中添加如下配置（假设Kafka服务运行在10.10.10.10的9092端口，需要替换为您自己的IP地址和端口号）。

```
listeners=PLAINTEXT://10.10.10.10:9092
advertised.listeners=PLAINTEXT://10.10.10.10:9092
```

> 注意 推荐您使用阿里云Kafka服务，并且需要保证Logstash所在节点的IP地址在Kafka的访问白名单内。

使用自建Kafka作为Logstash的输入或者输出时，出现 `could not be established. Broker may not be available` 错误日志，如何处理?

原因：Kafka服务不存在或者无法连接。

解决方法：请检查Kafka服务是否正常运行，或者Logstash管道配置中的 `bootstrap_servers` 配置是否正确。

阿里云Logstash的JDBC支持MySQL数据库吗?

支持。请参见配置扩展文件上传对应的mysql-connector-java包。