阿里云 Elasticsearch

用户指南

文档版本: 20190816

为了无法计算的价值 | [-] 阿里云

<u>法律声明</u>

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。

- 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。
- 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
•	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。	禁止: 重置操作将丢失用户配置数据。
A	该类警示信息可能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,恢复业务所需 时间约10分钟。
	用于补充说明、最佳实践、窍门等,不 是用户必须了解的内容。	道 说明: 您也可以通过按Ctrl + A选中全部文件。
>	多级菜单递进。	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。	单击 确定。
courier 字体	命令。	执行 cd /d C:/windows 命令,进 入Windows系统文件夹。
##	表示参数、变量。	bae log listinstanceid 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。	<pre>swich {stand slave}</pre>

目录

法律声明	I
通用约定	I
1	1
11 	1
1.2 基本信息	
1.3 集群升配	7
1.4 ES集群配置	11
1.5 yml配置	14
1.6 集群监控	21
1.7 日志查询	23
1.8 安全配置	25
1.9 同义词配置	
1.10 数据备份	39
1.10.1 数据备份	
1.10.2 宣看备份状态	
1,10.3 目列 笛 衍指閈 1 11 长 件码署	
1.11 1	
1.11.1 阀灶	47
1.11.3 自定义插件列表	
1.11.4 analysis-aliws插件使用说明	
1.12 集群数据节点缩容	63
2 可视化控制	73
2.1 Kibana	
2.1.2 基本配置(6.7.0)	75
2.1.3 访问配置	77
2.1.4 插件配置	79
2.1.5 BSearch-QueryBuilder插件使用介绍	81
2.1.6 BSearch-Label插件使用介绍	91
3 Elasticsearch自带功能	95
4 快照备份及恢复	96
5 访问控制	
○ % 1 47 - 17 - 17 - 17 - 17 - 17 - 17 - 17	105
5.2 访问鉴权规则	
5.3 临时访问令牌	
6 ElasticFlow	115
6.1 数据源概述	
6.2 加工函数及参数说明	

6.2.1 加工函数概述	
6.2.2 CONCAT	119
6.2.3 CURRENT_TIMESTAMP	
6.2.4 PARSE_URL	
6.2.5 REGEXP_EXTRACT	
6.2.6 REGEXP_REPLACE	123
6.2.7 REPLACE	
6.2.8 SPLIT_INDEX	
6.2.9 SUBSTRING	127
6.2.10 TO_TIMESTAMP	
6.3 算子	
6.3.1 算子类型与逻辑表	
6.3.2 数据导入算子	
6.3.3 数据过滤	131
6.3.4 数据加工算子	132
6.3.5 数据目标算子	133
6.4 快速开始	
6.4.1 创建数据源(跨云服务授权)	
6.4.2 创建项目和任务	
6.4.3 创建数据导入算子	
6.4.4 创建数据加工算子	140
6.4.5 创建数据目标算子	146
6.4.6 启动任务	147
6.4.7 删除任务	147
6.5 日志查询	
6.6 任务	
6.6.1 任务概述	148
6.6.2 任务类型	149
6.6.3 任务资源及分配策略	

1 实例管理

1.1 实例管理

本文档为您介绍阿里云Elasticsearch的实例管理功能,包括集群监控、重启实例、刷新、任务列 表。

管理功能

阿里云Elasticsearch为您提供了集群监控、重启实例、刷新、任务列表等实例管理功能。

es-cn--v k 重启实例 問新 三

集群监控

阿里云Elasticsearch实例已支持对实例进行监控,并允许通过短信的方式报警。您可以根据需求 自定义报警阈值,详情请参见Elasticsearch云监控报警。

重启实例

阿里云Elasticsearch支持重启和强制重启两种重启实例的方式。请参考以下说明,根据您的场景 选择合适的方式。

- 前提条件:在重启前,您需要确保Elasticsearch实例的状态为正常(显示为绿色)、索引 至少包含1个副本、资源使用率不是很高(可在#unique_6页面查看,例如节点CPU使用 率为80%左右或以下,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)。
 - 重启:阿里云Elasticsearch实例在重启过程中可持续提供服务(需要首先满足以上<mark>前提条</mark> 件),但耗时较长。

!! 注意:

- 在进行重启前,请确保实例的状态为正常(显示为绿色),否则需要进行<mark>强制重启</mark>。
- 节点在重启期间,对应的CPU和内存使用率会存在临时突增的情况,您的服务可能会出现抖动,正常情况下过一段时间后会恢复正常。
- 目前阿里云Elasticsearch实例重启耗时与实例的数据总量、节点数量、总索引数量和分片 数量等因素有关,暂无法预估具体的耗时长短。您可以在任务列表中查看实例的重启进度。

· 强制重启:强制重启操作可能会导致阿里云Elasticsearch实例在重启阶段的服务不稳定,但耗时较短。

!! 注意:

当磁盘的使用率超过85%时,阿里云Elasticsearch实例的状态可能显示为非健康状态(黄 色/红色),此时将不支持重启操作,需要进行强制重启。

- 如果创建的阿里云Elasticsearch实例处于非健康状态(黄色/红色),此时不建议对实例进行节点扩容、磁盘扩容、重启、修改密码或其他变更配置类的操作,请务必保证集群状态变为绿色后再进行这些操作。
- 当您的阿里云Elasticsearch实例包括了两个及两个以上节点,在上述情况下进行变更配置
 时,如果集群一直处于生效中的状态,您可以提交工单,联系阿里云Elasticsearch技术工
 程师处理。

刷新

当控制台中的信息没有及时更新时,例如刚创建完成的阿里云Elasticsearch实例状态显示失败,可以通过该刷新功能,手动刷新界面中的状态信息。

任务列表

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。

· 当前实例没有进行中的任务。

制台	集群监控	重启实例	刷新	≔
任务	列表			×
	当前实例》	没有进行中的任务		

・正在进行中的任务信息。

冶	集群监控	实例生效中 刷新	1
【任务	务列表		×
实例	变更中	56.88% 展开注情	
		33.00 % /////	

・展开正在进行中的任务详情。

日	[务列表	×
实例	列变更中	
	/0% 展开评情	
- T		
	□元成⊤京数:4/4, 进度:100%	
•	准备容器服务	
	进度:100%	
•	变更节点(重启,扩节点,扩规格)	
	已完成节点数:3/4, 进度:80%	
	准备配置	
	进度:0%	

1.2 基本信息

预付费(包年包月)实例基本信息

阿里云Elasticsearch包年包月类型实例的基本信息如下图所示,相关参数说明请参见下文内容和#unique_10。

es-cn-v	100.000.0	t		集群监控	重启实例	刷新	Ξ
基本信息						续费	
		实例D: es-cn-v	创建对词: 2019年6月19日 16:12:14				
		名称: es-cn-v 编辑	状态: ● 正常				
	Elas	icsearch 版本: 5.5.3 with Commercial Feature	付据类型: 包年包月 2019年6月19日 00:00:00 到期				
		区域: 华东1 (杭州)	可用区: cn-hangzhou-b				
		专有网络: vpc-b t	VSwitch信息: vsw-b 11				
		内网地址: es-cn-v t.elasticsearch.aliyuncs.com	内网姗曰: 9200				
		公网地址: 请开启公网访问地址后使用					
		协议: HTTP 修改					
配置列表	节点可视化			集體	¥数据节点缩容	集群升音	3
		数编节点规措:elasticsearch.n4.small(1核 2G)	数编节点数量:2				
		存储规格: SSD云曲	存储容量: 20 GiB				
		协调节点规格: elasticsearch.sn2ne.large(2核 8G)	协调节点数: 2				
	协调	节点存储类型:高效云鱼	协调节点存储空间: 20 GIB				联系
	Ki	bana节点现格: elasticsearch.n4.small(1核 2G)	Kibana节点数 1				我们

- · 名称: 阿里云Elasticsearch实例的名称默认与实例ID相同,支持自定义实例名称,也支持按照 名称搜索实例。
- · 内网地址: 支持在专有网络(VPC)下, 通过指定内网地址访问阿里云Elasticsearch实例服务, 依赖阿里云ECS实例。

(!) 注意:

通过公网访问阿里云Elasticsearch实例的安全性较差。如果您对访问环境的安全性有要求,可 以购买一个与阿里云Elasticsearch实例在同一VPC下的ECS实例,通过专有网络(VPC)指 定内网地址访问来阿里云Elasticsearch实例。

- · 内网端口: 主要支持的端口如下。
 - 支持9200端口,基于HTTP/HTTPS。
 - 支持9300端口,基于TCP。主要支持阿里云Elasticsearch 5.5.3 with Commercial Feature版本。

蕢 说明:

阿里云Elasticsearch 6.3.2 with Commercial Feature和阿里云Elasticsearch 6.7.0 with Commercial Feature版本不支持通过Transport Client访问9300端口。

- · 公网地址: 支持在公网环境下, 通过指定公网地址来访问阿里云Elasticsearch实例。
- · 公网端口: 主要支持的端口如下。
 - 支持9200端口,基于HTTP/HTTPS。
 - 支持9300端口,基于TCP。主要支持阿里云Elasticsearch 5.5.3 with Commercial Feature版本。



- 阿里云Elasticsearch 6.3.2 with Commercial Feature和阿里云Elasticsearch 6.7.0 with Commercial Feature版本不支持通过Transport Client访问9300端口。
- 需要添加#unique_11/unique_11_Connect_42_section_ux5_yct_zgb, 默认禁止 所有地址访问。
- ・协议:默认使用HTTP协议,您可以单击修改进行切换。目前支持HTTP和HTTPS协议之间的 切换,详情请参见使用HTTPS协议。
- ・续费:您可以单击基本信息右侧的续费,对实例进行续费。续费以月为单位,至少续费1个月。



后付费(按量付费)实例基本信息

阿里云Elasticsearch按量付费类型实例的基本信息如下图所示,相关参数说明请参见上文的预付 费(包年包月)实例基本信息和#unique_10。

es-cn-v	t		集群监控重启实例	刷新 〓
基本信息				转包年包月
	实例D: es-cn-vit t	创理时间: 2019年6月19日 16:12:14		
	名称: es-cn-v t 编辑	状态:● 正常		
	Elasticsearch 版本: 5.5.3 with Commercial Feature	付擲类型: 后付 跳		
	区域:华东1 (杭州)	可用区: cn-hangzhou-b		
	专有网络: vpc-b t	VSwitch信息: vsw-b		
	内网地址: es-cn-v Lelasticsearch.aliyuncs.com	内网满口: 9200		
	公网地址: 请开启公网访问地址后使用			
	协议:HTTP 修改			
配置列表 节点可	JRK		集群数据节点缩容	集群升配
	散振节点规格: elasticsearch.n4.small(1核 2G)	数据节点数量 2		
	存储规格: SSD云盘	存储容量: 20 GiB		
	协调节点规格: elasticsearch.sn2ne.large(2核 8G)	协调节点数: 2		
	协调节点存储关型:高效云盘	协调节点存储空间: 20 GiB		10 10
	Kibana节点规略: elasticsearch.n4.small(1載 2G)	Kübana节后题: 1		我们

转包年包月

阿里云Elasticsearch按量付费实例支持转为包年包月实例类型,您可以单击基本信息右侧 的转包年包月,按照界面提示进行开通。通过转包年包月功能,将后付费(按量付费)类型 的Elasticsearch实例转为预付费(包年包月)类型的Elasticsearch实例时,不支持折扣优惠。

配置列表

己 置列表 节点可视化		集群数据节点缩容	集群升配
数据节点规格: elasticsearch.n4.small(1核 2G)	数据节点数量: 3		
存储规格: SSD云盘	存储容量: 20 GiB		
冷数据节点规格: elasticsearch.sn2ne.large(2核 8G)	冷数据节点数: 2		
冷数据节点存储类型: 高效云盘	冷数据节点存储空间: 500 GiB		
专有主节点规格: elasticsearch.sn2ne.large(2核 8G)	专有主节点数: 3		
专有主节点存储类型:SSD云盘	专有主节点存储空间: 20 GiB		
协调节点规格: elasticsearch.sn2ne.large(2核 8G)	协调节点数: 2		
协调节点存储类型: 高效云盘	协调节点存储空间: 20 GiB		
Kibana节点规格: elasticsearch.n4.small(1核 2G)	Kibana节点数: 1		

相关参数说明请参见#unique_13。

节点可视化

配置列表 节点可视化	集群数据节点缩容	集群升配
生詳 <u>cn-hangzhou-b</u> 数据节点 ● ● ● 冷数据节点 ● ● 章有主节点		■ 联系我们
防调节点		
Kibana 节点		

集群数据节点缩容

集群数据节点缩容功能目前只支持后付费和单可用区部署的阿里云Elasticsearch实例,暂不支持 预付费和跨可用区部署的阿里云Elasticsearch实例。并且只支持数据节点个数的变更,不支持 专有主节点、协调节点、Kibana等类型节点的节点规格和磁盘缩容。详情请参见集群数据节点缩 容。

集群升配

主要支持按实例规格、节点数量、专有主节点规格、单节点存储空间等进行扩容。详情请参见#unique_15。

1.3 集群升配

本文档为您介绍阿里云Elasticsearch集群升配的相关说明、注意事项和操作方法。

阿里云Elasticsearch主要支持按照实例规格、节点数量、专有主节点规格、协调节点数、协调节 点规格、冷数据节点数量、冷数据节点规格、冷数据节点存储空间及单节点存储空间等进行升级配 置。

兰 说明:

以上个别属性可能会受到关联条件限制而无法扩容,详情请参见下文的配置变更。

登录阿里云Elasticsearch控制台,单击实例ID > 基本信息,进入实例的变配页面。



变配页面包括了当前配置和配置变更模块,相关说明请参见下文的当前配置和配置变更。

当前配置

当前配置模块展示了当前阿里云Elasticsearch实例的相关配置信息,便于您在进行升配操作时做 参考。

当前配置			
实例名称: es-cn-(
专有主节点:是	协调节点:是	协调节点数:2	协调节点存储空间:20G
协调节点存储类型:高效云盘	协调节点规格:2核8G	存储类型:SSD云盘	单节点存储空间:20
版本 : 6.3 with Commercial Feature	强制变更:false	专有主节点数:3(默认)	专有主节点存储空间:20G
专有主节点存储类型:SSD云盘	专有主节点规格:2核8G	网络类型:专有网络	数量:3
实例规格:2核4G	地域:华东1	规格族:云盘型	专有网络:vpc-
虚拟交换机:vsw-	可用区:杭州可用区B	冷数据节点:是	冷数据节点数量:3
冷数据节点存储空间:500	冷数据节点存储类型:高效云盘	冷数据节点规格:2核8G	

注意事项

在进行变更配置操作前,您需要认真阅读以下注意事项。

- ·如果您的业务有变更需求,请先进行评估后再进行变更。
- ・每次变更操作只支持更新集群的其中一个可变更的属性。
- · 一般情况下,升配操作会重启阿里云Elasticsearch实例。但是如果您已经购买过专有主节 点,那么在进行变更节点数量的操作时不会重启阿里云Elasticsearch实例。
- ·如果您的阿里云Elasticsearch实例的状态为非健康状态(黄色/红色),在进行变更操作时需要 勾选强制变更,忽略集群健康状态,此操作可能会影响服务。
- · 变更操作不支持更改节点的存储类型,但可以修改节点存储空间的大小。
- · 阿里云Elasticsearch支持Kibana节点规格的升配,在进行升配时需要收费。
- ·预付费类型的阿里云Elasticsearch实例暂不支持降配(例如减少节点个数、减少磁盘大小、减 少节点规格)。
- ·后付费类型的阿里云Elasticsearch实例目前降配功能只支持数据节点个数的变更(通过集群 数据节点缩容实现,会有限制),暂不支持其它类型的降配(例如减少磁盘大小、减少节点规 格)。
- · 在变更配置的过程中, 您可以在变配页面中实时观察更新后的订单消费金额。
- · 当您提交了变更配置的订单后,阿里云Elasticsearch实例将按照更新后的订单计费。

配置变更

(!) 注意:

在进行变更配置操作前,请首先认真阅读上文的注意事项。

请根据实际业务需求,参考页面中的提示修改实例的配置。参数详情请参见#unique_18。

ž	地域	经 东1 (杭州)		
	2009			
	可用区数量	单可用区 ▼ 普通邮晷模式,适用于非关键任务型工作负载		
	规格族	云截型		
	实例规格	11款29 ▼ 11款26规格只适合测试,不适用于生产环境,不在SLA售后保建范围内		
	数量	3 全两个节点集群有脑裂风险,谨慎选择		
	专有主节点	否 是		
	专有主节点数	3(数以)		
	专有主节点规格	288G 👻		
	专有主节点存储 类型	SSD元曲		
	专有主节点存储 空间	20G		
	协调节点	否 是		
	协调节点数	2		
-	协调节点规格	2核86 👻		
	协调节点存储类 型	而效元量		
	协调节点存储空 间	20G		
	冷数据节点	否 <u>是</u>		
	冷数据节点数量	2		
	冷数据节点规格	288G +		
	冷数据节点存储 类型	高效云盘		
		高效云曲最大存储空间支持5T,提供较为低廉的存储能力,适合大规模数层量的日本及分析场景 2.5TB以上的高效云曲通过磁曲种列及RAID 0的方式提供服务,不支持扩容		
	冷数据节点存储 空间	500 🗢		
	Kibana节点	高波云曲步 新时髦大支持扩展型21,新购能大支持5120Gib(51),超过2048Gibb],只能制: 2560、3072、3584、4096、4608、5120		
	Kibana规格	21243		
	强制变更			
		强制变更,忽略集群健康状态		
HALEL	单节点存储空间	20 全 单位为Gib, SSD 云曲最大存储空间支持2048Glb(2T) 高效云曲扩容时最大支持扩容到2T,新购时最大支持5120Glb(5T),超过2048Glb时,只能取: 2560, 3072, 3584, 4096, 4608, 5120		
	一次	升配变更资源过多,目前只能更新以下属性中的一个(各类节点的节点数量,存储空间、规格)。	成者没有某类节点的	时候购买该节

部分参数说明如下。

·规格簇及实例规格

规格簇不支持变更。如果规格簇为本地盘类型,则不支持实例规格的变更。

・专有主节点

在变配页面单击专有主节点右侧的是,可购买专有主节点。支持升配已购买的专有主节点的规格,默认为3个专有主节点,2核8G,云盘型存储类型,20GB存储空间。配置变更后,系统将按照新规格的定价进行重新计费。

📕 说明:

如果您已经购买了1核2G规格的专有主节点,可在变配页面重新购买更高规格的专有主节 点,系统将按照新规格的定价进行重新计费。如果之前的专有主节点是系统为您赠送的,变更 后该专有主节点将变为计费模式。

・协调节点

在变配页面单击协调节点右侧的是,可购买协调节点。支持升配已购买的协调节点的规格,默认为2个协调节点,2核8G,云盘型存储类型,20GB存储空间。配置变更后,系统将按照新规格的 定价进行重新计费。

・冷数据节点

在变配页面单击冷数据节点右侧的是,可购买冷数据节点。支持升配已购买的冷数据节点的规格,默认为2个冷数据节点,2核8G,云盘型存储类型,500GB存储空间。配置变更后,系统将按照新规格的定价进行重新计费。

Kibana节点

在变配页面单击Kibana节点右侧的是,可购买Kibana节点。支持升配已购买的Kibana节点的 规格,默认为2核4G。

🕛 注意:

当您购买了阿里云Elasticsearch实例后,系统默认会为您赠送一个1核2G的Kibana节点。您 可以通过集群升配操作变更Kibana节点的规格,规格变更后,系统将按照新规格的定价进行计 费。 ・强制変更

如果您的阿里云Elasticsearch实例状态为非健康状态(红色/黄色),且您的服务已受到严重影响,急需扩容解决。此时可以勾选强制变更选项,忽略阿里云Elasticsearch的实例健康状态进行强制变更,耗时相对较短。

<u>!</u>注意:

- 强制变更操作会重启阿里云Elasticsearch实例。
- 强制变更操作可能会导致阿里云Elasticsearch实例服务在重启阶段不稳定。
- 不勾选强制变更时,系统默认使用重启的方式进行集群升配,相关注意事项请参考重启实
 例。
- 如果您的阿里云Elasticsearch实例状态为非健康状态(红色/黄色),系统会自动勾选强制
 变更,将不支持通过重启方式进行集群升配。

・存储

单位为Gib, SSD云盘最大支持2048Gib(2T)的空间。

高效云盘扩容时最大支持扩容到2T,新购时最大支持5120Gib(5T)。超过2048Gib时,只能 取2560、3072、3584、4096、4608、5120。

1.4 ES集群配置

分词配置

作用于ES的同义词库,新的索引将会采用更新后的词库,配置操作详情请参见#unique_20。

基本信息	分词配置 同义词配置
ES集群配置	同义词配置: 智无
〕 说明:	
・上传同义は	司词典文件保存提交后,不会触发重启阿里云ES实例,但需要一些时间来生效。

· 在上传同义词词典文件保存生效之前创建的索引如果需要使用同义词功能,需要重建索引并进 行同义词相关配置。

每行一个同义词表达式,保存为utf-8编码的.txt文件。例如:

西红柿,番茄 =>西红柿,番茄 社保,公积金 =>社保,公积金

配置步骤:

- 1. 在阿里云ES控制台上传同义词词典文件,保存并生效成功。
- 在创建索引配置 setting时, 配置 "synonyms_path": "analysis/your_dict_name.
 txt", 再为该索引配置 mapping为指定字段设置同义词。
- 3. 校验同义词并上传测试数据进行搜索测试。

YML文件配置

控制台界面中的YML文件配置表示当前阿里云ES实例的使用配置。

集群监控	YML文件配置	传改配置
日志查询	自动创建索引:不允许自动创建索引 ?	制除素引指定名称 制除时间确素引名称 ⑦
安全配置	Auditlog素引:不开启auditlog素引 ⑦	开启Watcher: 关闭 ⑦
数据备份	其他configure配置: ⑦	
▼ 智能运维		
集群概况		
健康诊断		
历史报告		

修改配置

修改YML文件配置后,需要通过重启阿里云ES实例生效。

📕 说明:

修改YML文件配置完成后,再滑动到页面底部勾选该操作会重启实例,请确认后操作确认提交 后,会自动触发重启阿里云ES实例进行生效。

YML参数配置		×
自动创建索引:	 ● 不允许自动创建索引 ● 允许自动创建索引 ● 自定义 +.*,-* 	3
删除素引指定名称:	 删除时明确索引名称 删除时索引名称支持通配符 	0
Auditlog索引:	● 不开启auditlog索引 ○ 开启auditlog索引	0
开启Watcher:	 ● 关闭 ○ 开启 	⑦ <mark>咨</mark> 询

	其他configure配置:		?
1			
			の
			询
			建议
	该操作会重启实例,请确认后操作。		
		确认	取消

- · 自动创建索引:此设定允许您的阿里云ES实例在接收到新文档后,如果没有对应索引,是否允 许系统帮您自动新建索引。自动创建的索引有可能不符合您的预期,不建议开启。
- · 删除索引指定名称:此配置表示在删除索引时是否需要明确指定索引名称,如果配置为支持通配 符删除索引,则可以使用通配符进行批量删除索引。索引删除后不可恢复,请谨慎使用此配置。
- · Auditlog索引:此设定允许记录您的阿里云ES实例对应的增、删、改、查等操作产生的索引日志,该日志信息会占用您的磁盘空间,同时也会影响性能,不建议开启,请谨慎使用此配置。
- ・ 开启Watcher: 开启后可使用 X-Pack 的 Watcher 功能,请注意定时清理 .watcherhistory* 索引,避免占用大量磁盘空间。
- · 其他configure配置: 支持部分配置项如下, 详情请参考#unique_21 文档。



以下没有标识具体适用于某个ES版本的参数,默认兼容ES 5.5.3 和 6.3.2 版本。

- http.cors.enabled
- http.cors.allow-origin
- http.cors.max-age
- http.cors.allow-methods
- http.cors.allow-headers
- http.cors.allow-credentials
- reindex.remote.whitelist
- action.auto_create_index
- action.destructive_requires_name
- thread_pool.bulk.queue_size (只适用于 Elasticsearch 5.5.3 with X-Pack版本)
- thread_pool.write.queue_size (只适用于 Elasticsearch 6.3.2 with X-Pack版本)
- thread_pool.search.queue_size

1.5 yml配置

自定义CORS访问(跨域)

配置信息



- ·表格中的配置信息是阿里云Elasticsearch为支持HTTP协议开放的自定义配置。
- ・表格中的字段配置仅支持静态配置。如果您想使配置生效,请将配置信息写入elasticsea rch.yml文件中。
- ·以下配置依赖于集群网络设定(Network settings)。

配置项	描述
http.cors.enabled	跨域资源共享配置项,即配置Elasticsearch是 否允许其他域资源下的浏览器向其发送请求。可 设置为true或false。
	 设置为true,表示启用跨域资源访问。 即可使Elasticsearch处理OPTIONS CORS请求。如果发送请求中的域信息已在 http.cors.allow-origin中声明,那 么Elasticsearch会在头信息中附加 Access -Control-Allow-Origin以响应跨域请 求。 设置为false(默认),表示禁止跨域资 源访问。即可使Elasticsearch忽略请求 头中的域信息,Elasticsearch将不会以 Access-Control-Allow-Origin信息头 应答,以达到禁用CORS目的。如果客户端 不支持发送附加域信息头的 pre-flight请 求,或者不校验从服务端返回的报文的头信 息中的Access-Control-Allow-Origin 信息,那么跨域安全访问将受到影响。如 果Elasticsearch关闭CORS支持,则客户端 只能尝试通过发送OPTIONS请求,以了解此 响应信息是否存在。
http.cors.allow-origin	域资源配置项,可设置接受来自哪些域名的请 求。默认不允许且无配置。 如果在该配置值前后添加/,则此配置信息会被 识别为正则表达式。允许您使用正则方式兼容支 持HTTP和HTTPs的域请求信息。例如/https ?:\/\/localhost(:[0-9]+)?/,表示可响
	应符合此正则的请求信息。 说明: *被认定为合法配置,可被识别为使集 群支持来自任意域名的跨域请求,这将 给Elasticsearch集群带来安全风险,不建议 使用。

配置项	描述
http.cors.max-age	浏览器可发送OPTIONS请求以获取CORS配置信息,此配置项可设置获取的信息在浏览器中的缓存时间,默认为1728000秒 (20 天)。
http.cors.allow-methods	请求方法配置项,默认为OPTIONS,HEAD, GET,POST,PUT,DELETE。
http.cors.allow-headers	请求头信息配置项, 默认为X-Requested- With, Content-Type, Content-Length 。
http.cors.allow-credentials	凭证信息配置项目,即是否允许响应头中返回 Access-Control-Allow-Credentials信 息。默认为false,表示不允许返回此信息。设 置为true表示允许返回此信息。

自定义跨域访问配置示例如下。

```
http.cors.enabled: true
http.cors.allow-origin: "*"
http.cors.allow-headers: "X-Requested-With, Content-Type, Content-
Length, Authorization"
```

如果需要了解更多设置,请参见Elasticsearch官方网站,查看HTTP相关信息。

自定义远程索引重建(白名单)

索引重建组件支持用户在远程的Elasticsearch集群中重建数据索引,适用于您所能找到的任意版本的远程Elasticsearch服务。

您可以使用自定义远程索引重建功能,将旧版本的Elasticsearch服务中的数据索引到当前发布版 本中。使用示例如下,详细说明请参见Reindex API。

```
POST _reindex
{
    "source": {
        "remote": {
            "host": "http://otherhost:9200",
            "username": "user",
            "password": "pass"
        },
        "index": "source",
        "query": {
            "match": {
               "test": "data"
        }
      },
      "dest": {
            "index": "dest"
      }
    }
}
```

}

・ host: 远程主机的地址,必须包含支持协议、域名和端口等信息,例如https://otherhost :9200。

<u>!</u>注意:

远程主机地址需要在elasticsearch.yaml中使用reindex.remote.whitelist属性进行 声明,才可以在远程使用此API功能。允许以host和port组合,并使用逗号分隔多个主机配 置(例如otherhost:9200, another:9200, 127.0.10.**:9200,localhost:**)。 白名单不识别协议信息,只使用主机和端口信息用于实现安全策略设定。

· username和password为可选参数,如果您所请求的远程Elasticsearch服务需要使用Basic
 Authorization,请在请求中一并提供此参数信息。通过Basic Authorization鉴权需要
 使用https协议,否则密码信息将以文本形式进行传输。

📕 说明:

- ·如果机器地址信息已在白名单中设定,将不会验证和修改query,而是直接发送请求至远端服务。
- ·从远端集群索引数据,不支持手动切片或自动切片。详情请参见手动切片或自动切片。

•

批量设定

远端服务使用堆缓存索引数据,默认最大设定值为100(MB)。如果远端索引中包含大文档,请您 将批量设定值设置为较小值。

以下示例中,设置批量数值为10。

}

超时时间

- · 使用socket_timeout设置socket读取超时时间,默认超时为30s。
- · 使用connect_timeout设置连接超时时间,默认超时为1s。

以下示例中,设置socket读取超时为1分钟,连接超时为10秒。

```
POST _reindex
{
    "source": {
        "nemote": {
            "host": "http://otherhost:9200",
            "socket_timeout": "1m",
            "connect_timeout": "10s"
        },
        "index": "source",
        "query": {
            "match": {
               "test": "data"
        }
      },
     "dest": {
            "index": "dest"
      }
}
```

自定义Accesslog

开启Accesslog索引

阿里云Elasticsearch实例不支持查看存盘请求的相关log文件,因此如果您想了解阿里 云Elasticsearch实例请求的相关信息(例如access_log),那么需要在控制台中开启阿里 云Elasticsearch实例对应的Accesslog索引功能。



修改生效后,Accesslog将输出到阿里云Elasticsearch实例中,并且使用.security_a udit_log-*开头的索引名称。

配置Accesslog索引



- · 您无法实现对Accesslog索引过程中的信息进行过滤。当您的Accesslog事件中包含request body信息时,有可能在文本中暴露敏感信息。
- · 当您将Accesslog日志计入索引中时,该日志将占用您阿里云Elasticsearch实例的存储空间。
 因为阿里云Elasticsearch不支持自动过期清除策略,需要您手动触发清除陈旧的Accesslog索引。

Accesslog索引配置项如下。

```
xpack.security.audit.index.bulk_size: 5000
xpack.security.audit.index.events.emit_request_body: false
xpack.security.audit.index.events.exclude: run_as_denied,anonymous_
access_denied,realm_authentication_failed,access_denied,connection
_denied
xpack.security.audit.index.events.include: authentication_failed,
access_granted,tampered_request,connection_granted,run_as_granted
xpack.security.audit.index.flush_interval: 180s
xpack.security.audit.index.rollover: hourly
xpack.security.audit.index.settings.index.number_of_replicas: 1
xpack.security.audit.index.settings.index.number_of_shards: 10
```

特性	默认设置	描述
<pre>xpack.security.audit. index.bulk_size</pre>	1000	控制Accesslog事件完成一次 写入动作,并归档到文件的文 档数量。
<pre>xpack.security.audit. index.flush_interval</pre>	1s	控制缓冲事件刷新到索引的频 率。
<pre>xpack.security.audit. index.rollover</pre>	daily	控制滚动构建到新索引的频 率,可以设置为hourly、 daily、weekly或者monthly。
xpack.security.audit. index.events.include	<pre>access_denied, access_granted, anonymous_access_denied , authentication_faile d, connection_denied , tampered_request , run_as_denied, run_as_granted</pre>	控制何种Accesslog事件可以 被计入到索引中。完整列单请 参见Accesslog事件类型
<pre>xpack.security.audit. index.events.exclude</pre>		构建索引过程中排除的 Accesslog事件。

特性	默认设置	描述
<pre>xpack.security.audit. index.events.emit_reque st_body</pre>	false	当触发明确的事件类型(比如 authentication_failed),是否忽略或包含以REST发 送的请求体。

您也可以对存储Accesslog的索引进行配置,将以xpack.security.audit.index.settings

为命名空间, 配置在elasticsearch.yml文件中。

以下设置构建Accesslog索引的分片和副本均为1。

```
xpack.security.audit.index.settings:
    index:
        number_of_shards: 1
        number_of_replicas: 1
```



如果您希望通过传入参数配置生成Accesslog索引,请在开启Accesslog索引的同时传入 此配置。当阿里云Elasticsearch实例完成变更后,Accesslog索引将会出现在您的阿里 云Elasticsearch实例中。否则阿里云Elasticsearch实例的Accesslog索引将以默认的 number_of_shards: 5number_of_replicas: 1进行设置。

更多详细信息请参见Auditing Security Settings。

自定义queue大小

可以通过自定义thread_pool.bulk.queue_size、thread_pool.write.queue_size和 thread_pool.search.queue_size, 分别调整文档写入和搜索queue大小。

以下设置文档写入和搜索queue大小为500,需要您根据实际业务情况自行调整合适参数值。



说明:

以下示例默认兼容阿里云Elasticsearch 5.5.3和6.3.2版本。

```
thread_pool.bulk.queue_size: 500 (只适用于 Elasticsearch 5.5.3 with X-
Pack版本)
thread_pool.write.queue_size: 500 (只适用于 Elasticsearch 6.3.2 with X-
Pack版本)
```

thread_pool.search.queue_size: 500

参数调优

配置项	描述
index.codec	Elasticsearch的数据压缩算法默认为LZ4。当 您使用高效云盘的warm或cold集群时,通常 将该参数设置为best_compression,就可使 用更高压缩率的DEFLATE算法。在更改压缩算 法后,segment merge时会使用新的压缩算
	法。 ① 注意: 使用best_compression会导致写性能降低。

REST API设置

```
可以通过REST API设置index.codec参数,示例如下。
```

```
PUT $index_name/_settings
{
    "index": {
        "codec": "best_compression"
    }
}
```

📕 说明:

- ・执行命令前需要先关闭对应索引,否则会报错。
- · \$index_name: 替换为需要设置的索引名。

1.6 集群监控

集群告警

1. 支持一键报警功能,默认是关闭状态,可在集群列表页单击一键报警功能按钮进行开通或关闭。

阿里云Elasticsearch	集群列表			产品动态	
实例列表	创建	一键报警	刷新	请输入实例名称	Q

2. 如果为关闭状态,单击一键报警功能按钮提示如下,通过单击前往开启进行手动开启。



自定义报警

可自定义报警规则,通过单击前往云监控配置功能按钮进行创建报警规则。更多信息请参见#unique_5进行创建报警规则。

集群监控

可查看阿里云ES实例重要参数状态或负载情况。

预置时间

可通过单击预置时间查看对应时间段重要参数监控信息。

集群监控

 小时
 3小时
 6小时
 12小时
 1天
 3天
 7天
 14天

自定义

可通过单击自定义功能按钮配置开始和结束时间,查看对应时间段重要参数监控信息。

■ 自定义		
开始时间:	2018-07-24	12 🔨 : 46 🔨
结束时间:	2018-07-24	13 🔨 : 46 🔨
31天内的 最多连续到	查询提供分钟粒度 查询7天数据	的数据,确定

〕 说明:

31天内的查询提供分钟粒度的数据,最多连续查询7天数据。

重要参数



1.7 日志查询

阿里云Elasticsearch提供主日志、searching慢日志、indexing慢日志、GC日志的查询与展示。

通过输入关键字和设置时间范围,可以快速锁定需要查询的日志内容。阿里云Elasticsearch日志 最多支持7天内查询,默认按时间倒序展示。

阿里云Elasticsearch搜索日志语法,支持Lucene查询语法,详情请参见Query String Query。



由于Elasticsearch限制查询条件,最大支持返回10000条日志。如果在返回的10000条日志中,未覆盖到您所需要的日志内容,可以通过缩短查询时间范围来获取您需要的日志。

查询样例

以查询主日志中, content含有关键字为health, level为info, host 是192.168.1.123为例。

- 1. 登录阿里云Elasticsearch控制台,单击实例管理选择左侧菜单栏中的日志查询,再选择主日 志选项卡。
- 2. 输入框中输入: host:192.168.1.123 AND content:health AND level:info。
- 3. 选择需要查询的时间范围, 点击查询按钮。

基本信息	主日志 searchingt慢日志 indexingt慢日志 GC日志	
高级配置		
日志查询	host:192.168.1.123 AND content:health AND level:info	2018-05-15 08:00:04 😵 🖆 2018-05-15 11:46:48 🐼 😥
	时间节点IP内容	
	2018-05-15 08:00:10 192:168.1	7] Cluster health status changed from [YELLOW] to [GREEN] (reason: [shards started [[monitor
		每页显示 20 🗸 1 💙

▋ 说明:

- ·结束时间为空,默认为当前时间。
- ・开始时间为空,默认为结束时间减去1小时。
- · 上面提到的 AND 必须大写。

日志展示详解

在日志查询页面,根据您的查询条件会返回日志查询结果。主要有时间、节点IP、内容三部分。

时间

日志产生时间。

节点IP

ES节点所在Node的IP。

内容

主要由 level、host、time、content 组成。

- · level: 日志级别 trace、debug、info、warn、error 等(GC日志没有 level)。
- · host: ES节点所在的Node的IP(可通过Kibana展示页面中Node列表中获取IP列表)。
- time: 日志产生时间。

· content: 日志主要内容。

1.8 安全配置

本文档为您介绍阿里云Elasticsearch的安全配置功能,包括访问密码、公网/私网访问白名单以及HTTPS协议等。

集群网络设置

集群网络设置		
ES实例访问密码: 已设置 重置	VPC14网访问白玄单: 0.0.0.0/0	修改
公网地址:		
使用HTTPS协议: ?		

支持重置ES实例访问密码、配置VPC私网访问白名单、开启公网地址、配置公网地址访问白名单以 及使用HTTPS协议。

ES实例访问密码

单击重置,输入elastic管理员账号的新密码,确认后即可重置您的密码。密码重置后,新密码会 在5分钟左右之后生效。

重置			×
	! 您以后每次登录	ElasticSearch时都需要输入该信息,请妥善保存	
	用户名:	elastic	
	密码:	0/30	
		8-30个字符,且同时包含三项 (大写字母,小写字母,数字,特殊 字符)	
	确认密码:	两次输入密码保持一致 0/30	

密码重置后会影响您使用elastic管理员账号访问阿里云Elasticsearch实例,生效后的密码将作为 新的管理密码,用来登录Kibana控制台以及访问阿里云Elasticsearch实例。



- · 重置Elasticsearch实例访问密码不影响其它非elastic管理账号的账号访问该实例,不建议在 程序中通过elastic管理账号来访问您的阿里云Elasticsearch实例。
- · 单击重置, 密码修改完成并确认提交后, 不会触发重启阿里云Elasticsearch实例。

VPC私网访问白名单

当您需要通过内网来访问阿里云Elasticsearch实例时,可将待访问设备的IP加入到阿里云 Elasticsearch实例的VPC私网访问白名单中。

单击修改,在VPC白名单输入框中输入您需要添加的IP地址,单击确认即可完成添加。

白名单支持配置为单个IP或IP网段的形式,格式为192.168.0.1或192.168.0.0/24,多个IP之 间用英文逗号隔开。127.0.0.1代表禁止所有IPV4地址访问,0.0.0/0代表允许所有IPV4地 址访问。

📕 说明:

- ・默认允许所有内网IPV4地址访问。
- ・主要针对内网地址进行访问控制。

公网地址

单击公网地址开关,可开启公网地址。开启后开关显示为绿色,默认显示为灰色,即关闭状态。公 网地址开启后,才可使用公网地址访问您的阿里云Elasticsearch实例。

公网地址访问白名单

在配置公网访问白名单前,请首先打开公网地址开关。打开公网地址后,默认禁止所有公网地址访问。

当您需要通过外网来访问阿里云Elasticsearch实例时,可将待访问设备的IP加入到阿里云 Elasticsearch实例的公网地址访问白名单中。

白名单支持配置为单个IP或IP网段的形式,格式为192.168.0.1或192.168.0.0/24,多个IP之 间用英文逗号隔开。127.0.0.1代表禁止所有IPV4地址访问,0.0.0.0/0代表允许所有IPV4地 址访问。

目前杭州区域支持公网IPV6地址访问,并可以配置IPV6白名单,格式为2401:b180:1000:24:: 5或2401:b180:1000::/48。::1代表禁止所有IPV6地址访问,::/0代表允许所有IPV6地址访问。 问。

使用HTTPS协议

HTTPS全称是Hyper Text Transfer Protocol over Secure Socket Layer,是一种能够保障数 据安全的HTTP通道。它是 HTTP 协议的安全版,其保障数据传输安全的方式就是在HTTP协议的 基础上增加了SSL(Secure Sockets Layer,安全套接层)。也就是说HTTPS 还是通过HTTP进 行通信,只是传输的内容经过了SSL加密。

操作步骤

!〕注意:

- · 阿里云Elasticsearch支持HTTPS协议的开启和关闭能力,为了保障您数据的安全性,建议开 启HTTPS协议。
- · 在开启HTTPS协议前,您需要首先购买协调节点。

数量	2		
	两个节点集群有脑裂风	险,谨慎选择	
专有王节点	合	是	
协调节点	否	是	
协调节点数	2		
协调节点规格	2核8G	•	
协调节点存储类 型	高效云盘		
协调节点存储空 间	20G		
冷数据节点	否	是	
Kibana节点	否	是	

1. 登录阿里云Elasticsearch控制台,单击实例ID > 安全配置,然后打开使用HTTPS协议开关。

= (-)阿里云	Q 搜索
<	Encoder State (Contraction Contraction Con
基本信息 ES集群配置	集群网络设置
插件配置	ES实例访问密码: 已设置 重置
集群监控	公网地址
日志查询	使用HTTPS协议: 2
安全配置	
数据备份	
可视化控制	



- · 启用HTTPS协议前,您需要首先对访问Elasticsearch的代码进行变更,否则会导致无法使 用程序访问您的Elasticsearch实例。变更方式请参见开启和关闭HTTPS协议代码变更示 例。
- ・ 启用和关闭HTTPS服务会中断服务,并且会触发集群重启,为保证您的业务不受影响,请 确认后操作。
- 2. 在弹出的操作提示对话框中,勾选确认已经修改访问Elasticsearch实例的代码,单击确认。

ES实例访问密码: 已设置重置		VPC私网访问白名单
公网地址:		
使用HTTPS协议:		
	操作提示	
	 ● 请先按照说明文档对访问Elasticsearch的代码做变更后,再启用HTTPS协议,否则将导致当前Elasticsearch实例无法访问! 查看说明 ☑ 确认已经修改访问Elasticsearch实例的代码 	
	当前操作会触发集群重启,请确认操作。如需取消操作,请将访问Elasticsearch的代 码做协议相关的配置变更。	
	論认取消	
道 说明:		

如果您还未购买协调节点,在打开使用HTTPS协议开关后,系统会提示您购买协调节点。您可 以按照提示购买协调节点后再进行操作。

确认后集群会进行重启,可单击右上角的图标查看实例变更进度。重启完成后,就可以使用HTTPS协议访问您的实例了。

	集群监控 实例生效中	刷新 ⊒
	任务列表	×
VPC私网访问白名单: 0.0.0.0/0	实例变更55%	展开详情
	 准备ECS资源 已完成节点数: 5/5, 进度: 100% 准备容器服务 进度: 100% 变更节点 (创建、重启, 扩节点, 扩规格) 已完成节点数: 0/5, 进度: 22.5% 准备配置 进度: 0% 	

开启和关闭HTTPS协议代码变更示例

下面以Elasticsearch官方Rest Client访问方式为例,为您介绍在开启HTTPS前后,客户端代码 需要做的变更点。

· 开启HTTPS前的示例代码如下。

```
final CredentialsProvider credentialsProvider = new BasicCrede
ntialsProvider();
       credentialsProvider.setCredentials(AuthScope.ANY,
           new UsernamePasswordCredentials("elastic", "Your
password"));
RestClientBuilder restClientBuilder = RestClient.builder(
           new HttpHost("es-cn-xxxxx.elasticsearch.aliyuncs.com",
9200));
       RestClient restClient = restClientBuilder.setHttpCli
entConfigCallback(
           new RestClientBuilder.HttpClientConfigCallback() {
              @Override
              public HttpAsyncClientBuilder customizeHttpClient(
sProvider(credentialsProvider);
              }
```

```
}).build();
```

・开启HTTPS后的示例代码如下。

```
final CredentialsProvider credentialsProvider = new BasicCrede
ntialsProvider();
        credentialsProvider.setCredentials(AuthScope.ANY,
            new UsernamePasswordCredentials("elastic", "Your
password"));
RestClientBuilder restClientBuilder = RestClient.builder(
            new HttpHost("es-cn-xxxxx.elasticsearch.aliyuncs.com",
9200, "https"));
        RestClient restClient = restClientBuilder.setHttpCli
entConfigCallback(
            new RestClientBuilder.HttpClientConfigCallback() {
                @Override
                public HttpAsyncClientBuilder customizeHttpClient(
HttpAsyncClientBuilder httpClientBuilder) {
                    return httpClientBuilder.setDefaultCredential
sProvider(credentialsProvider);
            }).build();
```

由以上示例可知, 开启HTTPS后, HttpHost中需要加上https参数, 即new HttpHost("escn-xxxxx.elasticsearch.aliyuncs.com", 9200, "https"));。

1.9 同义词配置

配置说明

!) 注意:

- · 阿里云Elasticsearch上传同义词词典操作不会重启节点,后台会进行同义词词典的下发,生效时间与节点数量相关。
- · 假设现存索引index-aliyun使用了aliyun.txt同义词文件,当aliyun.txt文件内容变更并重新 上传后,现存索引不会动态加载更新后的同义词词典。建议您在词典文件内容发生变化后进行 索引重建操作,否则可能会造成只有新增数据使用新词典的情况。

您可以使用filter过滤器配置同义词,示例代码如下。
- filter: 配置一个synonym(同义词)过滤器,其中包含一个路径analysis/synonym.txt
 (路径是相对于config的位置)。
- · tokenizer: 用于控制标记同义词的分词器, 默认为whitespace分词器, 其他设置有:
 - ignore_case: 默认值为false。
 - expand: 默认值为true。

目前同义词分词器支持Solr和WordNet两种同义词格式。

・Solr同义词

文件的示例格式如下。

```
# Blank lines and lines starting with pound are comments.
# Explicit mappings match any token sequence on the LHS of "=>"
# and replace with all alternatives on the RHS. These types of
mappings
# ignore the expand parameter in the schema.
# Examples:
i-pod, i pod => ipod,
sea biscuit, sea biscit => seabiscuit
# Equivalent synonyms may be separated with commas and give
# no explicit mapping. In this case the mapping behavior will
# be taken from the expand parameter in the schema. This allows
# the same synonym file to be used in different synonym handling
strategies.
# Examples:
ipod, i-pod, i pod
foozball , foosball
universe , cosmos
lol, laughing out loud
# If expand==true, "ipod, i-pod, i pod" is equivalent
# to the explicit mapping:
ipod, i-pod, i pod => ipod, i-pod, i pod
# If expand==false, "ipod, i-pod, i pod" is equivalent
# to the explicit mapping:
ipod, i-pod, i pod => ipod
# Multiple synonym mapping entries are merged.
foo => foo bar
foo => baz
# is equivalent to
```

foo => foo bar, baz

您也可以在配置文件中直接给过滤器定义同义词(请注意使用synonyms而不是synonyms_p ath),示例如下。

▋ 说明:

建议您使用synonyms_path在文件中定义大型同义词集,因为使用synonyms定义会增加群集 大小。

・WordNet同义词

示例格式声明如下。

以上示例使用synonyms定义WordNet同义词,您也可以使用synonyms_path在文本中定义WordNet同义词。

使用示例一

以下示例使用filter过滤器配置同义词,操作步骤如下。

- 1. 上传同义词词典文件。
 - a. 登录阿里云Elasticsearch控制台。
 - b. 单击页面左上角的创建,新建一个阿里云Elasticsearch实例。
 - c. 单击实例链接, 进入实例配置页面。
 - d. 选择左侧菜单栏的ES集群配置,单击同义词配置。

<	es-cn-v 5	同义词配置	×
基本信息	分词配置	▲ 当前实例规格最大支持5M词典文件	
ES集群配置 插件配置	用火间配造 響光	同义词配置 ②	
集群监控	YML文件配置		
日志查询 安全配置	日均衡建康引:不允许日前物理集制 ③		咨询、建
数据备份	Audunggest. Arthaedunggest (9) 其他configure記言. (9)		议
▼ 智能运進			
柴田和元		上传文件	4
历史报告		保伊	取消

e. 单击上传文件,选择您要上传的同义词词典(按照上述规则生成的txt文件),单击保存。

等待阿里云Elasticsearch实例生效并提示状态正常后即可使用。本示例中使用aliyun_syn onyms.txt作为测试文件,内容为begin, start。

- 2. 配置并测试同义词。
 - a. 单击页面右上角的Kibana控制台。
 - b. 在Elasticsearch实例的kibana控制台中,单击左侧菜单栏的Dev Tool。
 - c. 在Console中执行如下命令, 创建索引。

```
PUT aliyun-index-test
{
    "index": {
        "analysis": {
            "by_smart": {
                "type": "custom",
                "tokenizer": "ik_smart",
                "filter": ["by_tfr","by_sfr"],
                "char_filter": ["by_cfr"]
                },
                "by_max_word": {
                "type": "custom",
                "tokenizer": "ik_max_word",
                "filter": ["by_tfr","by_sfr"],
                "filter": ["by_tfr","by_sfr"],
                "char_filter": ["by_cfr"]
                },
                "filter": ["by_cfr"]
                      "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                      "filter": ["by_cfr"]
                      "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                      "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                     "filter": ["by_cfr"]
                    "filter": ["by_c
```

```
"by_tfr": {
    "type": "stop",
    "stopwords": [" "]
    },
    "by_sfr": {
        "type": "synonym",
        "synonyms_path": "analysis/aliyun_synonyms.txt"
        }
    },
    "char_filter": {
        "by_cfr": {
            "type": "mapping",
            "mappings": ["| => |"]
        }
    }
}
```

d. 执行以下命令, 配置同义词字段title。

```
PUT aliyun-index-test/_mapping/doc
{
    "properties": {
        "title": {
            "type": "text",
            "index": "analyzed",
            "analyzer": "by_max_word",
            "search_analyzer": "by_smart"
    }
}
```

e. 执行以下命令,校验同义词。

```
GET aliyun-index-test/_analyze
{
"analyzer": "by_smart",
"text":"begin"
}
```

命令执行成功后,返回数据如下。

```
{
    "tokens": [
        {
            "token": "begin",
            "start_offset": 0,
            "end_offset": 5,
            "type": "ENGLISH",
            "position": 0
        },
        {
            "token": "start",
            "start_offset": 0,
            "end_offset": 5,
            "type": "SYNONYM",
            "position": 0
        }
]
```

}

f. 执行以下命令,添加数据,进行下一步测试。

```
PUT aliyun-index-test/doc/1
{
"title": "Shall I begin?"
}
PUT aliyun-index-test/doc/2
```

```
{
"title": "I start work at nine."
}
```

g. 执行以下命令,测试查询。

```
GET aliyun-index-test/_search
{
    "query" : { "match" : { "title" : "begin" }},
    "highlight" : {
        "pre_tags" : ["<red>", "<bule>"],
        "post_tags" : ["</red>", "</bule>"],
        "fields" : {
            "title" : {}
        }
    }
}
```

命令执行成功后,返回数据如下。

```
"took": 11,
"timed_out": false,
"_shards": {
"total": 5,
 "successful": 5,
 "failed": 0
},
"hits": {
 "total": 2,
 "max_score": 0.41048482,
 "hits": [
   {
     "_index": "aliyun-index-test",
"_type": "doc",
     "_id": "2",
"_score": 0.41048482,
     "_source": {
       "title": "I start work at nine."
     },
     "highlight": {
       "title": [
          "I <red>start</red> work at nine."
        ]
     }
   },
   {
     "_index": "aliyun-index-test",
     "_type": "doc",
"_id": "1",
     "_score": 0.39556286,
```

```
"_source": {
    "title": "Shall I begin?"
    },
    "highlight": {
        "title": [
            "Shall I <red>begin</red>?"
        ]
      }
    }
}
```

使用示例二

- 以下示例直接引用同义词并使用IK过滤,操作步骤如下。
- 1. 设置一个同义词过滤器my_synonym_filter,并配置同义词词库。
- 2. 设置一个my_synonyms解释器,使用ik_smart分词。

经过ik_smart分词,把所有字母小写并作同义语查找。

```
PUT /my_index
"analysis": {
           "analyzer": {
               "my_synonyms": {
    "filter": [
                        "lowercase",
                        "my_synonym_filter"
                    ],
"tokenizer": "ik_smart"
               }
          },
"filter": {
    cyne

               "my_synonym_filter": {
                    "synonyms": [
                        "begin, start"
                    ],
"type": "synonym"
               }
          }
      }
}
}
```

3. 执行以下命令,设置同义词字段title。

```
PUT /my_index/_mapping/doc
{
"properties": {
  "title": {
    "type": "text",
    "index": "analyzed",
    "analyzer": "my_synonyms"
}
}
```

}

4. 执行以下命令,校验同义词。

```
GET /my_index/_analyze
{
    "analyzer":"my_synonyms",
    "text":"Shall I begin?"
}
```

命令执行成功后,返回数据如下。

```
"tokens": [
 {
   "token": "shall"
   "start_offset": 0,
   "end_offset": 5,
   "type": "ENGLISH",
   "position": 0
 },
 {
   "token": "i"
   "start_offset": 6,
   "end_offset": 7
   "type": "ENGLISH",
   "position": 1
 },
 {
   "token": "begin",
"start_offset": 8,
"end_offset": 13,
   "type": "ENGLISH",
   "position": 2
 },
{
   "token": "start"
   "start_offset": 8,
"end_offset": 13,
   "type": "SYNONYM",
   "position": 2
 }
]
}
```

5. 执行以下命令,添加数据,进行下一步测试。

```
PUT /my_index/doc/1
{
"title": "Shall I begin?"
}
PUT /my_index/doc/2
{
"title": "I start work at nine."
}
```

6. 执行以下命令,测试查询。

GET /my_index/_search

```
"query" : { "match" : { "title" : "begin" }},
"highlight" : {
    "pre_tags" : ["<red>", "<bule>"],
    "post_tags" : ["</red>", "</bule>"],
    "fields" : {
        "title" : {}
    }
}
```

命令执行成功后,返回数据如下。

```
"took": 11,
"timed_out": false,
"_shards": {
 "total": 5,
"successful": 5,
 "failed": 0
"max_score": 0.41913947,
 "hits": [
   {
     "_index": "my_index",
"_type": "doc",
      "_id": "2",
      "_score": 0.41913947,
      "_source": {
       "title": "I start work at nine."
     "I <red>start</red> work at nine."
        ]
      }
   },
    {
     "_index": "my_index",
"_type": "doc",
     "_id": "1",
      "_score": 0.39556286,
      "_source": {
"title": "Shall I begin?"
     "title": [
          "Shall I <red>begin</red>?"
        ]
     }
   }
 ]
}
}
```

本文档部分内容参考了官方Elasticsearch文档,详细信息请参见Synonym Token Filter和 Using Synonyms。

1.10 数据备份

1.10.1 数据备份

本文档为您介绍阿里云Elasticsearch的数据备份功能。

进入阿里云Elasticsearch控制台,单击实例名称 > 数据备份,进入数据备份(免费试用)页面。

<	es-cn-v		東野盆地 重前尖州	刷新	Ξ
基本信息 ES集群配置	数据备份 (免费试用)			停改	武王
插件配置	开启自动备份:		自动保密开始时间。每天 12.00 🕢		
集群监控	醫份恢履:	点由查看校程	發台状态 <i>向</i> 由查 要 教理		
日志宣闻					
数据备份					
可视化控制					

参数	说明
开启自动备份	开启自动备份开关为绿色时表示启用,默认为关闭状态。
自动备份开始时间	如果没有开启自动备份功能,会提示请开启自动备份后配置。
	 注意: 如果已经开启自动备份功能,则自动备份启动时间是 当前Region的时间,请避免在自动数据备份时做集 群Snapshot操作。

参数	说明	
修改配置	如果已经开启自动备份功能,可单击右上角的修改配置,在 备份周期设置页面修改自动备份时间点。	E自动
	自动备份周期设置	\times
	备份周期: 每天 省份时间点: 02:00 へ 00:00 へ 01:00 02:00 ✓ 03:00 04:00	
	05:00 06:00	
	07.00	
	 注意: 备份周期是每天。 备份时间点单位是小时,可选时间范围为[0~23]。 自动备份只保存最近3天的快照数据。 	
备份恢复	可单击点击查看教程查看对应的文档。	
备份状态	可单击点击查看教程查看对应的文档。	

1.10.2 查看备份状态

自动备份状态

开启自动备份功能后,可以在阿里云ES集成的Kibana控制台中的Dev Tools中通过Elasticsearch snapshotAPI查看自动备份的快照状态。

查看所有快照

通过以下命令查看aliyun_auto_snapshot仓库中所有快照信息。

GET _snapshot/aliyun_auto_snapshot/_all

返回内容如下:

```
{
    "snapshots": [
        {
            "snapshot": "es-cn-abcdefghijklmn_20180628092236",
            "snapshot": "es-cn-abcdefghijklmn_20180628092236",
```

```
"uuid": "n7YIayyZTm2hwg8BeWbydA",
    "version_id": 5050399,
    "version": "5.5.3",
    "indices": [
       ".kibana"
    ],
"state": "SUCCESS",
    "start_time": "2018-06-28T01:22:39.609Z",
    "start_time_in_millis": 1530148959609,
"end_time": "2018-06-28T01:22:39.923Z",
     "end_time_in_millis": 1530148959923,
    "duration_in_millis": 314,
"failures": [],
    "shards": {
"total": 1,
       "failed": 0
       "successful": 1
    }
  },
    "snapshot": "es-cn-abcdefghijklmn_20180628092500",
     "uuid": "frdl1YFzQ5Cn5xN9ZWuKLA",
    "version_id": 5050399,
    "version": "5.5.3",
    "indices": [
       ".kibana"
    ],
"state": "SUCCESS",
2018
    "start_time": "2018-06-28T01:25:00.764Z",
    "start_time_in_millis": 1530149100764,
    "end_time": "2018-06-28T01:25:01.482Z",
    "end_time_in_millis": 1530149101482,
    "duration_in_millis": 718,
    "failures": [],
    "shards": {
       "total": 1,
       "failed": 0,
       "successful": 1
    }
  }
]
```

· state: 显示快照状态, Elasticsearch快照共有5种状态:

- IN_PROGRESS: 快照正在执行。
- SUCCESS: 快照执行结束, 且所有shards都存储成功。
- FAILED: 快照执行结束, 但有数据存储不成功。
- PARTIAL: cluster级别的数据存储成功,但至少有1个shard的数据没有存储成功。
- INCOMPATIBLE: 快照与阿里云ES实例版本不兼容。

查看指定快照

}

通过以下命令查看aliyun_auto_snapshot仓库中指定快照的详细信息。

GET _snapshot/aliyun_auto_snapshot/<snapshot>/_status

· <snapshot>: 替换为自动备份快照名称,例如es-cn-abcdefghijklmn_20180628092236

0

返回内容如下:

```
{
  "snapshots": [
    {
       "snapshot": "es-cn-abcdefghijklmn_20180628092236",
       "repository": "aliyun_auto_snapshot",
       "uuid": "n7YIayyZTm2hwg8BeWbydA",
       "state": "SUCCESS",
       "shards_stats": {
         "initializing": 0,
         "started": 0,
         "finalizing": 0,
         "done": 1,
         "failed": 0,
         "total": 1
      },
"stats": {
         "number_of_files": 4,
         "processed_files": 4,
         "total_size_in_bytes": 3296,
         "processed_size_in_bytes": 3296,
         "start_time_in_millis": 1530148959688,
         "time_in_millis": 77
      },
"indices": {
         ".kibana": {
            "shards_stats": {
              "initializing": 0,
              "started": 0,
              "finalizing": 0,
              "done": 1,
"failed": 0,
              "total": 1
           },
"stats": {
    "mer_l
              "number_of_files": 4,
"processed_files": 4,
              "total_size_in_bytes": 3296,
              "processed_size_in_bytes": 3296,
"start_time_in_millis": 1530148959688,
"time_in_millis": 77
           },
"shards": {
"?"· {
              "0": {
                "stage": "DONE",
                 "stats": {
                   "number_of_files": 4,
"processed_files": 4,
                   "total_size_in_bytes": 3296,
                   "processed_size_in_bytes": 3296,
                   "start_time_in_millis": 1530148959688,
                   "time_in_millis": 77
                }
              }
          }
         }
      }
    }
```

] }

1.10.3 自动备份指南

开启自动备份功能

- 1. 登录阿里云Elasticsearch控制台。
- 2. 单击Elasticsearch实例ID链接,跳转到Elasticsearch实例信息页面。
- 3. 单击左侧导航栏的数据备份。
- 4. 在数据备份页面,打开开启自动备份开关,开启自动备份功能。

数据备份 (免费试用)				修改配置
开启自动备份:)	自动备份开始时间:	每天 04:00 ⑦	
备份恢复: 点	点击查看教程	备份状态:	点击查看教程	

5. 单击界面右侧的修改配置,选择每日自动备份触发的时间。

自动备份周期设置				
备份周期:	每天			
备份时间点:	04:00	^		
	00:00	^		
	01:00			
	02:00			
	03:00			
	✓ 04:00			
	05:00			
	06:00			
	07:00			
	08:00	-		

从自动备份恢复数据

如果您对指定阿里云Elasticsearch实例开启了自动备份功能,每天会自动为该Elasticsearch实例备份数据,您可以通过Elasticsearch的snapshot API恢复数据到原阿里云Elasticsearch实例中。



- ・您的第一个快照是阿里云Elasticsearch实例数据的完整拷贝,但所有后续的快照保留的是已存 快照和新数据之间的增量,这意味着首次快照备份会耗时较长,后续备份会比较快。
- · 快照不保存阿里云Elasticsearch实例自身的监控数据(例如以.monitoring和. security_audit为前缀的索引)。
- · 快照只能用于恢复到原阿里云Elasticsearch实例。
- · 自动快照仓库会在首次做快照时生成。

查看所有仓库

使用GET _snapshot命令查看所有仓库信息。

返回内容如下:

```
{
    "aliyun_auto_snapshot": {
        "type": "oss",
        "settings": {
            "compress": "true",
            "base_path": "xxxx",
            "endpoint": "xxxx",
            }
      }
}
```

- ・aliyun_auto_snapshot: 仓库名称。
- · type: 快照存储介质, 阿里云OSS。
- · compress:true:采用压缩模式,进行快照时会对索引的metadata信息进行压缩。
- · base_path: 快照在OSS中的存储位置。
- · endpoint: OSS所处区域的信息。

查看全部快照

通过GET _snapshot/aliyun_auto_snapshot/_all命令查看aliyun_auto_snapshot仓库 内所有快照信息。

返回内容如下:

```
"snapshots": [
     {
       "snapshot": "es-cn-abcdefghijklmn_20180627091600",
       "uuid": "MMRniVLPRAiawSCm8D8Dug",
       "version_id": 5050399,
       "version": "5.5.3",
       "indices": [
          "index_1",
          ".security",
          ".kibana"
       ],
"state": "SUCCESS"
: "2018
       "start_time": "2018-06-27T01:16:01.009Z",
       "start_time_in_millis": 1530062161009,
"end_time": "2018-06-27T01:16:05.632Z",
       "end_time_in_millis": 1530062165632,
       "duration_in_millis": 4623,
"failures": [],
       "shards": {
"total": 12,
          "failed": 0,
          "successful": 12
       }
     }
  ]
}
```

默认参数

自动快照还有以下未显示的默认参数:

- ・ max_snapshot_bytes_per_sec:40mb: 单节点最大备份数据速度为40mb/秒。
- · max_restore_bytes_per_sec:40mb: 单节点最大恢复数据速度为40mb/秒。
- · chunk_size: Max 1Gb: 做快照过程中大文件会被拆分成若干个小文件,拆分出的文件最大为1GB。

从快照恢复索引数据

您可以通过Elasticsearch的_restore命令从快照中恢复索引数据。

·恢复aliyun_auto_snapshot仓库中指定快照的所有索引(后台执行):

POST _snapshot/aliyun_auto_snapshot/<snapshot>/_restore

<snapshot>: 替换为自动备份快照的名称,例如es-cn-abcdefghijklmn_20180 627091600。 ·恢复aliyun_auto_snapshot仓库中指定快照的所有索引,并等待任务处理完成。

_restore为异步命令, Elasticsearch实例在确认可执行恢复操作后会立即返回, 该恢复任 务会在后台执行, 可以通过追加wait_for_completion参数阻塞命令直到恢复完成再返回信 息。

```
POST _snapshot/aliyun_auto_snapshot/<snapshot>/_restore?wait_for_c
ompletion=true
```

<snapshot>: 替换为自动备份快照名称, 例如es-cn-abcdefghijklmn_20180627091600

0

・恢复aliyun_auto_snapshot仓库中指定快照的指定索引,并为恢复的索引重命名(后台执行)。

```
POST _snapshot/aliyun_auto_snapshot/<snapshot>/_restore
{
    "indices": "index_1",
    "rename_pattern": "index_(.+)",
    "rename_replacement": "restored_index_$1"
}
```

- <snapshot>: 替换为自动备份快照名称,例如es-cn-abcdefghijklmn_20180 627091600。
- indices:选择需要恢复的索引名称。
- rename_pattern: (可选)正则匹配需要恢复索引的名称。
- rename_replacement: (可选)为匹配上的索引按规则重命名。

1.11 插件配置

1.11.1 概述

阿里云Elasticsearch在开源社区插件的基础上,提供了丰富的插件,在此基础上又扩展并丰富了 部分插件提供给您使用。本章节为您介绍阿里云Elasticsearch的插件配置功能,帮助您快速使用 阿里云Elasticsearch提供的插件完成特定的业务需求。

使用插件配置功能

登录阿里云Elasticsearch控制台,单击实例ID > 插件配置。

<	es-cn-v	-			集群监控重用实例	周新 ☰
基本信息 ES集群配置	插件配置					
揚件配置	系统默认插件列表 自我	宝义插件列表				
集群监控	刷新				请输入插件名称	Q
日志童海	插件名称	英型	状态	描述	操作	
安全配度	analysis-icu	系统默认	 Egit 	Elasticsearch ICU分析攝件,集成了ICU機块,调加了ICU机关的分析组件。	卸載	
可视化控制	analysis-ik	系统数认	 E安装 	Elasticsearch IK分析简件。	冷距新	
 ■ 智能运進	analysis-kuromoji	医结果认	 B安装 	Elasticsearch日语(kuromoji)分析插件, 集成了Lucene kuromoji分析模块。	80 M	
健康诊断	analysis-phonetic	系统默认	 已安装 	Elasticsearch 音标分析插件,集成了词元过漆器。	卸載	
历史报告	analysis-pinyin	系统默认	 巴安陵 	Elasticsearch 拼音分析插件。	卸載	
	analysis-smartcn	系统默认	 已安装 	Elasticsearch 智能中文分析描件,集成了Lucene智能中文分析器。	50 \$t)	
	analysis-stconvert	系统赢动	• 未安談	STConvert is a analysis plugin that convert Chinese characters between traditis simplified.	onal and 安装	联系
	elasticsearch-repositor	ry-oss 至皖默认	 已安装 	支持使用阿里云OSS存储Elasticsearch Snapshot		

在插件配置页面,您可以查看系统默认插件列表和自定义插件列表。

· 系统默认插件列表

系统默认插件列表中的analysis-ik和elasticsearch-repository-oss插件默认不能卸

载。analysis-ik插件支持IK词典的冷更新和热更新操作,可以实现自定义扩展词典的更新功能。详情请参见系统默认插件列表。

・自定义插件列表

支持自定义标准类型的插件的上传、安装及卸载,以满足您特定场景的需求。详情请参见自定义 插件列表。

1.11.2 系统默认插件列表

本文档为您介绍阿里云Elasticsearch支持的默认插件,以及IK分词插件的冷更新和热更新的操作 方法。

阿里云Elasticsearch的默认插件列表如下图所示。

	插件翻	記置					
55来6H和2直 插件配置	3	系统默	认插件列表				
集群监控	ţ	刷新					请输入插件名称 Q
日志查询			括任冬款	2071	壮态	描述	爆作
安全配置 数据备份			analysis-icu	系统默认	• 已安業	Elasticsearch ICU分析插件,集成了ICU模块,添加了ICU相 关的分析组件。	5082
▼ 智能运维			analysis-ik	系统默认	 已安装 	Elasticsearch IK分析插件。	冷更新
集群概况 健康诊断			analysis-kuromoji	系统默认	 ● 已安装 	Elasticsearch日语(kuromoji)分析插件 ,集成了Lucene kuromoji分析模块。	卸戰
历史报告			analysis-phonetic	系统默认	 已安装 	Elasticsearch 音标分析插件,集成了词元过谗器。	卸戰
			analysis-pinyin	系统默认	 ● 已安装 	Elasticsearch 拼音分析插件。	卸载
			analysis-smartcn	系统默认	● 已安装	Elasticsearch 智能中文分析插件,集成了Lucene智能中文分析器。	卸载
			elasticsearch-repository-oss	系统默认	 ■ 已安装 	支持使用阿里云OSS存储Elasticsearch Snapshot	
			ingest-attachment	系统默认	 已安装 	预处理器,使用Apache Tika抽取内容。	卸载
			ingest-geoip	系统默认	 已安装 	预处理器,根据IP地址在MaxMind数据库中查找对应的地理 数据。	卸载
			ingest-user-agent	系统默认	 已安装 	预处理器,从用户Agent抽取信息。	卸载
		批量安排	走 批量印载				毎页显示 10 ~ 〈 1 2 〉
基本信息							
ES集群配置	插件面	置					
播件配置	3	系统默认	人插件列表				
集群监控	用	创新					请 協入 插件名称 Q
日志查询			插件名称	类型	状态	描述	操作
安全配置数据备份]	mapper-attachments	系统默认	 已安装 	The mapper attachments plugin adds the attachment type to Elasticsearch using Apache Tika.	卸戰
* 智能运进]	mapper-murmur3	系统默认	 已安装 	Mapper Murmur3允许用户在创建索引时计算字段值的哈希 值,并将获得的哈希值存储到索引中。	印載
乘6+140元 健康诊断 历史报告]	mapper-size	系统默认	 已安装 	Mapper Size播件允许在素引时中记录文档压缩前的大小。	印載
]	repository-hdfs	系统默认	 □安装 	HDFS存储库插件提供了对Hadoop分布式文件系统 (HDFS)存储库的支持。	卸载
]	sql	系统默认	● 未安装	Query elasticsearch using SQL	安装
			x-pack	系统默认	 已会装 	Elasticsearch Expanded Pack Plugin	
	10	比量安装	批量卸载				毎页显示 10 ~ 〈 1 2 〉

当您购买了阿里云Elasticsearch实例后,系统会默认安装系统默认插件列表中的插件,您可以 根据需求进行插件的<mark>卸载和安装</mark>。其中analysis-ik和elasticsearch-repository-oss插件为阿里 云Elasticsearch的扩展插件,默认不能卸载。

- · analysis-ik: IK分词插件。在开源插件的基础上,我们扩展支持了OSS词典文件的动态加载。 支持IK词典的冷更新和热更新操作。
- · elasticsearch-repository-oss: 在开源插件的基础上,支持通过阿里云OSS文件系统实现索引快照的创建及恢复。

卸载/安装系统默认插件

! 注意:

安装或卸载插件都会触发集群重启,并且卸载插件时会删除当前选中的插件,请确认后操作。

以analysis-kuromoji插件为例。

- 1. 单击analysis-kuromoji插件右侧操作栏下的卸载。
- 2. 认真阅读操作提示对话框中的注意事项,确认无误后再单击确认。

<	es-cn-v				黄群监控 重音实例 刷新 三
基本信息 ES集群配置	插件配置				
插件配置	系统默认插件列表 自定义插件列表				
集群监控	Ritr				请编入插件名称 Q
日志查询			100		
安全配置	2014年4月2	<u>22</u>	0X	20 8 1	SMTE
数据备份	analysis-icu	系统默认	 已安装 	Elasticsearch ICU分析操件,集成了ICU模块,添加了ICU模块的分析组件。	50 M
可视化控制	analysis-ik	系统默认	操作提示	× 、 ≤> 考慮す。	冷更新
▼ 智能逻维	analysis-kuromoji	系统默认	当前操作会触发集群重启,并删除当前选中	的操件,请确认操作。 日语(kuromoji)分析插件,集成了Lucene kuromoji分析媒块。	5088
集群概况				201A IN 1	
健康诊断	analysis-phonetic	系统默认		音标分析描件,集成了词元过滤器。	50號
历史报告	analysis-pinyin	系统默认	 已安装 	Elasticsearch 拼音分析循件。	10HE

确认后,集群会进行重启。重启成功后,如果看到analysis-kuromoji插件的状态变为未安装,表示插件卸载成功。

ES集群配置	插件配置				
插件配置	系统默认插件列表自定义指	种列表			
集群监控	刷新			清编入	插件名称 Q
日志查询	插件名称	类型	状态	描述	操作
安全配置		11/2012/02/		Flasticsearch ICII 分析活性 集成了ICII 描述 3	表記でには
数据备份	analysis-icu	系统默认	 ● 已安装 	关的分析组件。	卸载
可视化控制	analysis-ik	系统默认	• 已安装	Elasticsearch IK分析插件。	冷更新 热更新
▼ 智能运维 集群概況	analysis-kuromoji	系统默认	• 未安装	Elasticsearch日语(kuromoji)分析插件, 集成 kuromoji分析模块。	版 文了Lucene 安装 我 们
健康诊断	analysis-phonetic	系统默认	● 已安装	Elasticsearch 音标分析播件,集成了词元过读器	者。 卸載
NJS-JKH	analysis-pinyin	系统默认	 已安装 	Elasticsearch 拼音分析插件。	卸载

如果您需要再次使用此插件,可继续执行以下步骤进行安装。

3. 单击插件右侧的安装,使用同样的方式完成插件的安装。

插件安装会触发集群重启。重启成功后,如果看到插件的状态显示为已安装,表示插件安装成功。

IK词典更新

阿里云Elasticsearch的IK分词插件支持以下两种IK词典的更新方式。

- · IK词典冷更新。
- ・IK词典热更新。

对于已经配置了IK分词器的索引,在IK词典冷/热更新操作完成后将只对新数据生效。如果您希望 对全部数据生效,需要重建索引。

IK词典冷更新

IK词典冷更新需要对整个集群的词典进行更新操作。在使用IK冷更新时,会将您上传的词典文件传送到Elasticsearch节点上,并修改IKAnalyzer.cfg.xml文件,然后重启Elasticsearch节点加载词典文件。

IK词典冷更新支持修改IK自带的主词库及停用词主词库。在IK词典冷更新页面可以看到系统自带的主词典即SYSTEM_MAIN.dic,系统自带的停用词主词典即SYSTEM_STOPWORD.dic。

蕢 说明:

- ·如果您需要修改系统自带主词典文件,可以上传以SYSTEM_MAIN.dic为名的主词典。
- ·如果您需要修改系统自带的停用词主词典文件,可以上传以SYSTEM_STOPWORD.dic为名的 停用词主词典。

使用示例(IK词典冷更新)

- 1. 登录阿里云Elasticsearch控制台,单击您需要进行IK更新的Elasticsearch实例ID。
- 2. 单击左侧导航栏的插件配置, 找到需要更新的IK插件, 单击其右侧操作栏下的冷更新。

<	es-cn-v u				集群监控 重用实例	周新 Ξ
基本信息 15世期和第	插件配置					
53朱中山直 掛件配置	系统默认插件列表					
集群监控	周新				请输入插件名称	Q
安全配置	□ 播件名称	英型	状态	描述	操作	
数据备份	analysis-icu	系统默认	 已安装 	Elasticsearch ICU分析操件,集成了ICU機块,添加了ICU機关的分析组件。	卸载	
可视化控制	analysis-ik	系统默认	 已安装 	Elasticsearch IK分析描件。	冷更新 热更新	
▼ 智範运標 集群概況	analysis-kuromoji	系统默认	 已安装 	Elasticsearch日语(kuromoji)分析操件, 集成了Lucene kuromoji分析模块。	10%	
健康诊断	analysis-phonetic	系统默认	 已安装 	Elasticsearch 音标分析插件,集成了词元过漆器。	10%	
历史报告	analysis-pinyin	系统默认	 已安装 	Elasticsearch 拼音分析插件。	10転	
	analysis-smartcn	系统默认	 已安装 	Elasticsearch 智能中文分析描件,集成了Lucene智能中文分析器。	卸数	

3. 在插件配置页面,单击右下方的配置。

插件配置	×
	-
· 当前实例规格最大支持5M词典文件	
IK主分词词库 😮	
SYSTEM_MAIN.dic	
IK停用词库 ?	
SYSTEM_STOPWORD.dic	
	● 联
	系我们
雨翌日日	出
「「「」「「」「」「」「」「」「」「」「」「」「」「」「」「」「」「」「」「	肖

4. 单击IK主分词词库右下方的上传dic文件,上传一个扩展主词典文件。

插件配置	
· 当前实例规格最大支持5M词典文件	
IK主分词词库 2	
SYSTEM_MAIN.dic \times	
上传dic文件 🗸	上传dic文件
▲ 当前实例规格最大支持5M词典文件	
IK主分词词库 ⑦	
SYSTEM_MAIN.dic \times a_10words.dic success \times	
上传dic文件 V	上传dic文件

	 说明: 系统支持上传dic文件和添加OSS文件两种词典更新方式。在远程词典文件内容变更后, 式都需要重新上传词典文件才能生效。 	两种方
5.	注意:该操作会重启实例,为保证您的业务不受影响,请确认后操作。	
	滑动到页面底端,勾选该操作会重启实例按钮,请确认后操作,单击保存。	
	K停用词库 @ SYSTEM_STOPWORD.dic ×	●联系我们
	上传dic文件 上传dic文件 ● ▼ 该操作会重启实例,请确认后操作。	7消
6.	重启成功后,登录kibana控制台,执行以下命令测试词典是否生效。	

GET _analyze { "analyzer": "ik_smart", "text": ["您词典中包含的词"] }

! 注意:

- ·系统不支持删除自带的主词典和停用词主词典。
- ·无论是词典文件变化,还是词典内容发生变化,冷更新操作都会触发集群重启。
- · 只能在集群健康的状态下进行更新。

IK词典热更新

IK词典热更新是当词典文件内容发生变化的时候,上传词典文件后Elasticsearch节点能自动加载词典文件,实现词典的更新操作。

在使用IK词典热更新功能时,如果词典文件列表发生变化,例如上传新词典文件或删除词典文件,那么整个集群都需要重新加载词典的配置(因为会涉及到修改IKAnalyzer.cfg.xml文件)。

IK词典热更新与上文的IK词典冷更新操作类似,如果您是第一次上传词典,因为需要修改 IKAnalyzer.cfg.xml文件,所以是需要进行重启集群操作。

使用示例(IK词典热更新)

- 1. 登录阿里云Elasticsearch控制台,单击您需要进行IK更新的Elasticsearch实例ID。
- 2. 单击左侧导航栏的插件配置, 找到需要更新的IK插件, 单击其右侧操作栏下的热更新。

<	es-cn-v u				集群监控 重启实例 刷新
基本信息 ES果群配置	插件配置				
損件配置	系统默认插件列表				
樂群监控	用的新				请输入插件名称 (
日志查询	播件名称	类型	状态	描述	操作
安全配置 数据备份	analysis-icu	服銃駅以	● 已安装	Elasticsearch ICU分析插件,集成了ICU模块,添加了ICU相关的分析组件。	50 9 0
可视化控制	analysis-ik	系统默认	 E安装 	Elasticsearch IK分析描件。	冷更新
▼ 智能运進 集群概況	analysis-kuromoji	系统默认	 已安装 	Elasticsearch日语(kuromoji)分析循件, 親成了Lucene kuromoji分析環境。	2080
健康诊断	analyzis-phonetic	系统默认	 已安装 	Elasticsearch 普标分析描件,集成了词元过漆器。	10 10
历史报告	analysis-pinyin	系统默认	 已安装 	Elasticsearch 拼音分析插件。	20 2 0

3. 在插件配置页面,单击右下方的配置。

插件配置	×
9 当前实例规格最大支持5M词典文件	
IK主分词词库 😮	
SYSTEM_MAIN.dic	
STSTEM_STOPWORD.uic	
	系我
	- ſŊ
配置取	肖

4. 单击IK主分词词库右下方的上传dic文件,上传一个扩展主词典文件。

插件配置	
· 当前实例规格最大支持5M词典文件	
IK主分词词库 😮	
SYSTEM_MAIN.dic \times	
上传dic文件 ~	上传dic文件
插件配置	>
IK主分词词库 ⑦	
a_10words.dic success \times	
上传dic文件 V	上传dic文件
上传dic文件 V	上传dic文件

系统支持上传dic文件和添加OSS文件两种词典更新方式。在远程词典文件内容变更后,两种方 式都需要重新上传词典文件才能生效。

5. <u>()</u> 注意:

该操作会重启实例,为保证您的业务不受影响,请确认后操作。

滑动到页面底端,勾选该操作会重启实例按钮,请确认后操作(第一次上传词典文件,需要重 启),单击保存。

IK停用词库 😮	
SYSTEM_STOPWORD.dic \times	
	L L L L L L L L L L L L L L L L L L L
	"" 我
上传dic文件 ~	上传dic文件
Ⅰ ✓ 该操作会重启实例,请确认后操作。	
	保存取消

单击保存后,集群会触发Rolling操作,等待集群Rolling结束后,词典会自动生效。 词典使用一段时间后,如果您需要扩充或者减少词典中的内容,可继续执行以下步骤对上述 a_10words.dic文件进行修改。 6. 进入词典热更新页面,先删除之前上传的同名词典文件,重新上传修改过的a_10words.dic同名词典文件。

因为修改的是已存在的同名词典文件的内容,所以本次上传修改过的同名词典文件不需 要Rolling整个集群,如下图所示。

插件配置	×
a_10words.dic success \times	
IK停用词库 ②	
	ざ 道 ・
	建议
上传dic文件 V 上传dic文件	
· · · · · · · · · · · · · · · · · · ·	取消

7. 单击保存。

因为使用的是Elasticsearch节点上插件自动加载的词典文件,所以每个节点获取词典文件的可 能时间不同,请耐心等待词典生效。大概两分钟后再使用更新之后的词典,为了保证准确性,请 登录Kibana控制台,多次执行以下命令进行验证。

```
GET _analyze
{
"analyzer": "ik_smart",
```

```
"text": ["您词典中包含的词"]
}
```

▋ 说明:

IK热更新不支持修改系统自带的主词典,如果您需要修改系统主词典请使用IK冷更新的方式。

参考文档: elasticsearch-analysis-ik。

1.11.3 自定义插件列表

本文档为您介绍阿里云Elasticsearch自定义插件的上传、安装及卸载功能。

上传并安装自定义插件



- · 上传自定义插件会触发Elasticsearch集群重启安装,并且插件本身可能影响Elasticsearch集群的稳定性,请务必保证自定义插件的可用性和安全性。
- ·Elasticsearch升级时插件不会自动升级更新,需要重新上传。
- ・如果您的插件不涉及保密协议,我们也希望您能开源,一起共建和丰富开源社区的插件。
- 1. 在插件配置页面,单击自定义插件列表 > 上传。

<	es-cn-v	
基本信息 ES集群配置	插件配置	
插件配置	系统默认插件列表 自定义插件列表	
集群监控	上传刷新	
日志查询 安全配置	▲ 插件名称 對	趐
数据备份		

 在弹出的插件上传对话框中,单击添加多个文件或将多个文件拖放到此区域,添加您需要上传的 自定义插件。

插件上传		×
选择文件:	添加多个文件或将多个文件拖放到此区域	
	elasticsearch-sql-6.7.0.0.zip (4.44M) ×	
	自定义插件上传会触发集群重启安装,并且插件本身可能影响集群稳定,请 务必保证自定义插件的可用性和安全性。ES升级时插件不会自动升级更新, 需要重新上传。	
	上传安装取	肖

您也可以将自定义的插件文件直接拖放到指定区域,完成添加。如上图成功添加了文件名为elasticsearch-sql-6.7.0.0的自定义插件。



使用同样的方式可添加多个自定义插件。

3. 仔细阅读并勾选对话框中的注意事项,单击上传安装。

插件上位	ŧ		×
选择	文件:	添加多个文件或将多个文件拖放到此区域	
		elasticsearch-sql-6.7.0.0.zip (4.44M)	
	~	自定义插件上传会触发集群重启安装,并且插件本身可能影响集群稳定,请 务必保证自定义插件的可用性和安全性。ES升级时插件不会自动升级更新, 需要重新上传。	
		上传安装	消

插件的上传安装会触发集群重启。集群重启成功后,如果在自定义插件列表中看到您上传的插 件,且状态显示为已安装,表示插件上传并安装成功。

系统默认插件列表 自定义插件列表					
上世剧新				请输入插件名称	Q
插件名称	英型	状态	捕送	操作	
sql	自定义播件	 ● 已安装 	Query elasticsearch using SQL	卸戳	

如果您不需要再使用此插件,可单击插件右侧的卸载,卸载此插件。详细操作方法及注意事项请 参见卸载/安装系统默认插件。

1.11.4 analysis-aliws插件使用说明

analysis-aliws是阿里云Elasticsearch自带的一个系统默认插件。通过analysis-aliws插件,您可以在Elasticsearch中集成对应的分析器和分词器,完成文档的分析和检索。

安装analysis-aliws插件

登录阿里云Elasticsearch控制台,单击实例ID > 插件配置 > 系统默认插件列表。在系统默认插件 列表列表中安装analysis-aliws插件,详情请参见卸载/安装系统默认插件。

基本信息 ES集群配置	插件配置					
插件配置	系统默认插件列表	自定义插件列表				
集群监控	周纬				请输入插件名称	Q
日志查询	插件名称	类型	状态	描述	操作	
安全配置数据备份	analysis-aliws	系统默认	● 未安装	Elasticsearch Aliws分析插件	安装 词车配置	

(!) 注意:

analysis-aliws插件默认为未安装状态。

使用analysis-aliws插件

analysis-aliws插件安装成功后,阿里云Elasticsearch默认会集成如下的分析器和分词器。

- ・分析器: aliws
- · 分词器: aliws_tokenizer

您可以使用上述的分析器和分词器完成文档的查询,具体步骤如下。

1. 创建索引。

以上代码创建了名称为index的索引,类型为fulltext,属性为content,并添加了aliws分析器。

2. 添加文档。

```
POST /index/fulltext/1
{
    "content": "中华人民共和国国歌。"
}
```

以上代码创建了名称为1的文档,并设置了文档中的content字段的内容。

3. 查询。

```
POST /index/fulltext/_search
{
    "query": {
        "match": {
            "content": "共和国"
        }
    }
}
```

以上代码在所有fulltext类型的文档中,使用aliws分析器,搜索content字段中包含共和国的文档。

📕 说明:

如果您在使用analysis-aliws插件时,得到的结果不符合预期,可通过下文的分析器测试和分词器 测试进行排查调试。

分析器测试

```
GET _analyze
{
"text": "中华人民共和国国歌。",
"analyzer": "aliws"
}
```

分词器测试

```
GET _analyze
{
    "text": "中华人民共和国国歌。",
    "tokenizer": "aliws_tokenizer"
}
```

1.12 集群数据节点缩容

集群数据节点缩容功能目前适用于后付费和单可用区部署的阿里云ES实例,暂不支持预付费和 跨可用区部署的阿里云ES实例。并且只支持数据节点个数的变更,不支持专有主节点、协调节 点、Kibana等类型节点的节点规格和磁盘缩容。

数据节点缩容

- 1. 在阿里云Elasticsearch控制台首页选择要进行数据节点缩容的ES实例,单击实例ID。
- 2. 在基本信息标签页,单击集群数据节点缩容。



3. 在缩容配置页面,选择对应节点类型,并在节点列表中单击选择需要缩容的数据节点。

缩容配置		
请选择节点类型: 数据节点	\sim	
当前节点个数: 13		
缩节点个数: 4		
192.168.	192.168.	192.168
192.168.	192.168.	1 192.168.
192.168	192.168.	



为保证数据的安全,进行缩容的数据节点中不应该存在数据,如果所选数据节点中有数据,可 以通过单击数据迁移辅助工具进行数据迁移。迁移后所选节点上不再有任何索引数据,新的索 引数据也不会被写入该节点。

容配置
选择节点类型: 数据节点 🗸
前节点个数: 4
节点个数: 1
192.168.
保证集群健康状况和数据安全,当前集群无法缩容1个节点。请您迁移或清空完部分节点中数据后缩容。点击前往 <mark>数据迁移辅助工具</mark> 。

您可以通过系统建议和自定义两种方式进行集群数据节点迁移:

・系统建议

通过系统建议自动选择需要缩容的数据节点,勾选同意数据迁移协议并确认。

提示 安全, 192	系统根据数据分布和集群健康状况智能推荐迁移节点IP,无法盘点业务数据重要性,请二次确认数据 2.168.3.152	

・自定义

您也可以通过自定义标签项,手动选择需要缩容的数据节点,勾选同意数据迁移协议并确 认。

集群数据节点迁移				
迁移节点类型:冷数据节点				
新玩玩建议 日花文				
提示:慎重自定义方式选择	廷移数据节点,可能会影 ······	响集群健康状况。		
192.168.	192.168.	192.168	192.168.	
✓ 我同意《阿里云Elasticsearch数据记	E移服务协议》,授权阿里云El	asticsearch服务协助迁移目标	数据节点中的数据。	
			确认	取消

进度查看

您可以在右上角的任务中心查看缩容或者数据迁移进度。




迁移回滚

在数据迁移过程中,可以根据需要选择停止迁移任务。

	Kibana控	制台	集群监控	实例生效中	刷新
任务列	表				
集群数据	重新分布			_	
		50%	展开详情	停止	
 待迁 	移节点选择中				
进度	: 100%				
• 迁移	数据				

Kibana控制台	集群监控	重启实例	刷新	=
			>	×
新分布(已取消)		屈开详情		
50%		ЖЛ म н		
00%				
居				
%,迁移服务器IP列表	長(0台):			
	Kibana控制台 新分布(已取消) 50% 方点选择中 00% 屠 9%,迁移服务器IP列表	Kibana控制台 集群监控 新分布(已取消) 50% 方点选择中 00% 8 9%, 迁移服务器IP列表(0台):	Kibana控制台 集群监控 重启实例 新分布(已取消) 新分布(已取消) 展开详情 方の、 展开 00% 8 0%, 迁移服务器IP列表(0台):	Kibana控制台 集群监控 重启实例 刷新 新分布(已取消)

迁移失败处理

数据迁移是一个周期很长的过程,在此期间集群状态和数据的变更可能会导致迁移失败,具体可 在任务中心中查看。当数据迁移失败或者迁移完成后需进行回滚,可以进行如下操作:

1. 获取迁移的IP列表

可以在任务中心中查找,或者直接通过使用以下ES API命令查找:

```
// 获取集群配置信息ES API命令
GET _cluster/settings
// 返回类似以下结果
{
  "transient": {
    "cluster": {
      "routing": {
        "allocation": {
         "exclude": {
           "_ip": "192.168.***.***,192.168.***.***,192.168.***.***"
          }
       }
     }
   }
 }
}
```

2. 回滚迁移节点

使用类似以下命令进行迁移节点回滚:

```
// 回滚部分节点ES API命令,配置中去掉需要回滚的节点,但要保留不回滚的节点。
PUT _cluster/settings
{
  "transient": {
    "cluster": {
      "routing": {
        "allocation": {
          "exclude": {
            "_ip": "192.168.***.***,192.168.***.***"
          }
        }
      }
    }
  }
}
// 回滚全部节点
PUT _cluster/settings
{
  "transient": {
    "cluster": {
      "routing": {
"allocation": {
          "exclude": {
            "_ip": null
          }
       }
     }
 }
}
```

3. 校验

可以通过GET _cluster/settings进行IP地址的二次确认,同时也可以通过观察相应节点是 否被重新分配shard来判断回滚完成情况。

数据迁移或回滚时均可以通过ES API GET _cat/shards?v进行任务的观察。

异常说明

常见异常及解决方案

数据迁移或者缩容时,常见异常提示如下:

· 该操作会导致当前集群资源(disk/cpu/mem)不足或shard分配异常。

原因与解决方案:集群在迁移或者缩容后,磁盘、内存、CPU等资源不足,不足以承担当前系统数据或者负载。使用GET _cat/indices?v查看集群中索引的副本数是否超出缩容后的节点数,检查磁盘等资源是否可以在安全阈值内承担当前的数据量和请求量。

・集群当前状态异常或有未完成任务,请稍后再试。

原因与解决方案:使用GET _cluster/health查看集群健康状况,或者在智能运维中查看集群异常原因。

- ・ 集群节点有数据, 请先执行迁移。
- ・保留节点数需大于2且大于当前节点数一半。

原因与解决方案:为保证集群的可靠性,保留的节点数必须大于2;为保证集群的稳定性,每次 迁移或者缩容节点数不得超过当前数据节点数的一半。

・当前ES集群配置不支持该操作,请检查ES配置。

原因与解决方案:使用GET _cluster/settings查看集群配置,查看是否存在不允许数据分配的配置。

auto_expand_replicas

部分用户使用了xpack提供的权限管理功能,在早期版本中,该功能对应的.security或 者.security-6索引默认会使用 "index.auto_expand_replicas" : "0-all",该配置会使得 数据迁移或者节点缩容失败。当您使用 auto_expand_replicas索引时,建议参照类似以下方式修 改:

```
// 查看索引配置
GET .security/_settings
// 返回结果
{
    ".security-6" : {
        "settings" : {
        "settings"
```

```
"index" : {
         "number_of_shards" : "1",
"auto_expand_replicas" : "0-all",
         "provided_name" : ".security-6",
         "format" : "6",
         "creation_date" : "1555142250367",
         "priority" : "1000",
"number_of_replicas" : "9",
         "uuid" : "9t2hotc7S50pPuKEÍJ****",
         "version" : {
          "created" : "6070099"
        }
      }
    }
  }
}
// 修改配置 (二选一)
PUT .security/_settings
{
  "index" : {
    "auto_expand_replicas" : "0-1"
  }
}
PUT .security/_settings
{
  "index" : {
    "auto_expand_replicas" : "false",
"number_of_replicas" : "1"
  }
}
// 副本个数可以根据实际需要选择,但要保证至少有1个,且不大于可用数据节点数。
```

2 可视化控制

2.1 Kibana

2.1.1 登录Kibana控制台

本文档为您介绍登录阿里云Kibana控制台的方法。当您购买了阿里云Elasticsearch实例后,我们 会为您赠送一个1核2G的Kibana节点。通过Kibana,您可以完成数据查询,数据可视化等操作。

前提条件

在登录Kibana控制台之前,您需要首先购买一个Elasticsearch实例。并且确保Kibana公网访问为开启状态。

背景信息

阿里云Elasticsearch实例提供Kibana控制台,为您的业务提供扩展的可能性。Kibana控制台 作为Elastic生态系统的组成部分,支持无缝衔接Elasticsearch服务,可以让您实时了解阿里 云Elasticsearch实例的运行状态并进行管理。

操作步骤

- 1. 登录阿里云Elasticsearch控制台,单击实例ID > 可视化控制。
- 2. 在可视化控制页面,单击Kibana模块中的进入控制台。



3. 在登录页面, 输入用户名和密码, 单击登录。

Welcome to Kibana Your window into the Elastic Stack	
Username	
Log in	

- · 用户名:默认为elastic。
- · 密码: 您在购买Elasticsearch实例时设置的密码。

使用阿里云Elasticsearch 6.7版本(其他版本略有差异)登录成功后的页面如下图所示。

	12haaa	Dev Tools Help			
	kibana	Welcome to Console			
Ê	Visualize	Quick intro to the UI			
50	Dashboard	The Console UI is split into two panes: an editor pane (left) and a response pane (right). Use the editor to type requests and submit them to Elasticsearch. The results will be displayed in the response pane on the right side.			
₽	Timelion	Console understands requests in a compact format, similar to cURL:			
寙	Canvas	While typing a request. Console will make suggestions which you can then accept by hitting Enter/Tab. These suggestions are made based on the request structure as well as your indices and types.			
\$	Maps	A few quick tips, while I have your attention			
ø	Machine Learning	Submit requests to ES using the green triangle button.			
G	Infrastructure	Use the wrench menu for other useful timgs. You can paste requests in cURL format and they will be translated to the Console syntax.			
I	Logs	You can resize the editor and output panes by dragging the separator between them. Study the keyboard shortcuts under the Help button. Good stuff in there!			
G	АРМ	Get to work			
্ত	Uptime	Canada - Canada Bardina - Carl Daharana			
÷	Graph	Console Search Profiler Grok Debugger			
ę	Dev Tools	1 Eff search F 1 1			
æ	Monitoring	4 ""match_all": () 5^ }			
۵	Management	0* }			
2	elastic	1			
В	Logout				
	Default				
÷	Collapse				

后续步骤

登录成功后,您就可以在Kibana控制台上完成数据查询、制作数据展示仪表板等操作。详情请参 见官方文档。

2.1.2 基本配置(6.7.0)

本文档为您介绍Kibana节点的基本配置功能。通过基本配置功能,您可以切换Kibana控制台的语 言,包括中文和英文。



Kibana节点的基本配置功能仅适用于阿里云Elasticsearch 6.7.0 with Commercial Feature版本。

Kibana语言配置

1. 登录阿里云Elasticsearch控制台,单击实例ID > 可视化控制。

2. 在Kibana模块中,单击修改配置,进入Kibana配置页面。

<	es-cn-vt 9	
基本信息		
ES集群配置	Kibana	V DataV
插件配置		
集群监控		
日志查询		
安全配置	QueryBuilder DSL	
数据备份	• 新增Querybuilder插件,可在修改配置中查看。	• 天猫双11、阿里云城市大脑同款数据大屏。
可视化控制	• 强大而便速的开发工具。	• 高性能三维渲染引擎。
▼ 智能运進	• 时序数据分析、数据图谱分析。	 多种图表组件,支撑多种数据类型的分析展示。
集群概况	相关文档参看Kibena言方介绍 修改配置 进入控制台	相关文档参看DataV官方介绍进入控制台
健康诊断		

在Kibana配置页面即可看到基本配置模块。在基本配置模块中,您可以继续执行以下步骤配 置Kibana控制台的语言,默认为English。

<	l es-cn-v k	集群监控	重启实例	刷新	≡
Kibana 🏹	基本設置			修改配到	2
	Kharutānītāli Ergish				

3. 单击基本配置右侧的修改配置。

(!) 注意:		
此配置需要重启Kibana节点才会生效。	为避免影响您的Kibana操作,	请确认后再执行以下步
骤。		

4. 在修改Kibana基本配置页面,选择修改语言,单击确定。

修改Ki	修改Kibana基本配置			
	修改语言: 💿 中文			
	🔘 English			
~				

Kibana控制台支持English和中文两种语言,默认为English。

确认完成后,Kibana节点会自动进行重启。重启后,#unique_44,即可看到系统使用您设置的语言进行显示。

	kibana	◎ 若養在 Kibana 中可規化和政策政策, 完美要创建委引成式, 以从 Elasticsearch 检索效率。			
Ø	Discover	Dev Tools	历史记录	设置	帮助
旈	可视化	欢迎使用 Console			0
5	仪表板	U linh?			
₩	Timelion	Console UI 分成两个警惕:鏡編器藝幣(左),他用編編器使入请求并符它们建交到 Elasticeard,结果符號不在右側的地位警察中。			
命	Canvas	Console 理解素质物成的演变,类似于 cURL:			
	Maps	键入遗疗时,Console 将提供建议,您可以通过技 Enter/Tab 键未接受建议,这些建议超手请求结构以双密的按书 IK 类型组织			
	Machina Learning	有几个景要想注意的有用退示			
	HILLOW	• 使用导色三角绘图片唐求理文列 55.			
	an and the second s	· 使用形式探索性方规的周期起始计。 · 您可以结点 ULL_能诊的意义,这些描字对转物感《 Console 语法指式、			
8	日本	・ 名可以通过語時機構構成は営業をご用の分類集構構造と10分入、 ・学习電行話中で経緯性性がついた問題が生まった。			
,4	APM	开始工作			
ି	运行时间				
÷	Graph	Console Search Profiler Grok Debugger			
Ÿ	开发工具	1 laff search ▶ ≠ 1 2 · (
æ	Monitoring] = "ouery": {			
۲	管理	5 · } 6 · }			
2	elastic	1			
Β	注销				
	Default				
÷	折叠				

2.1.3 访问配置

本文档为您介绍Kibana集群的访问配置功能,包括开启公网访问和访问白名单配置。

进入Kibana访问配置页面

1. 登录阿里云Elasticsearch控制台,单击实例ID > 可视化控制。

2. 在Kibana模块中,单击修改配置,进入Kibana配置页面。

<	es-cn-vi 9	
基本信息		
ES集群配置	Kibana	🕎 DataV
插件配置		
集群监控		
日志查询		
安全配置	QueryBuilder DSL	
数据备份	• 新增Querybuilder插件,可在修改配置中查看。	• 天猫双11、阿里云城市大脑间款数据大屏。
可视化控制	• 强大而便捷的开发工具。	 高性能三维渲染引擎。
智能运维	• 时序数据分析、数据图谱分析。	 多种图表组件,支撑多种数据类型的分析展示。
集群概况	相关文档参看Kibana官方介绍 修改配置 进入控制台	相关文档条看DataV官方介绍进入控制台
健康诊断		

在Kibana配置页面即可看到访问配置模块。在访问配置模块中,您可以开启或关闭Kibana公 网访问权限(默认开启),并配置Kibana访问白名单。

<	es-cn-\u	員群進拉 重用实例 刷新 三
Kibana配置	基本配置	修改動量
	Kibana衛言配置: English	
	Kitana公网访问:	Kiberal访问日名单. 费定

Kibana公网访问

Kibana公网访问开关默认显示为绿色,表示为开启状态。单击Kibana公网访问开关,可关闭Kibana公网访问权限,关闭后开关显示为灰色。Kibana公网访问关闭后,将无法通过外网进入Kibana控制台。

Kibana访问白名单

单击修改,在Kibana访问白名单输入框中输入您需要添加的IP地址,单击确认即可完成添加。



Kibana访问白名单支持配置为单个IP或IP网段的形式,格式为192.168.0.1或192.168.0.0/ 24,多个IP之间用英文逗号隔开。127.0.0.1代表禁止所有IPV4地址访问,0.0.0.0/0代表允许 所有IPV4地址访问。

目前杭州区域支持公网IPV6地址访问,并可以配置IPV6白名单,格式为2401:b180:1000:24:: 5或2401:b180:1000::/48。::1代表禁止所有IPV6地址访问, ::/0代表允许所有IPV6地址访问。 问。

2.1.4 插件配置

阿里云Kibana在开源社区插件的基础上,提供了丰富的插件。本文档为您介绍阿里云Kibana自定 义插件的种类、安装和卸载方法及注意事项。

自定义插件列表

BSearch-QueryBuilder插件。

BSearch-QueryBuilder又称高级查询,是阿里内部纯自研的一个纯前端的工具插件,具有如下特性。

- 简单易用: BSearch-QueryBuilder插件提供了可视化的界面点选操作来构造Elasticsearch的 DSL查询请求,无编码即可完成自定义条件的数据查询,减少了复杂的DSL的学习成本。也可辅助开发人员编写或验证DSL语句的正确性。
- · 方便快捷:已经定义的复杂查询条件会保存在Kibana中,避免重复构造查询条件。
- · 小巧轻盈:约占用14MB的磁盘空间,不会常驻内存运行,不影响Kibana和Elasticsearch的 正常运行。
- · 安全可靠: BSearch-QueryBuilder插件不会对用户的数据进行改写、存储和转发,源代码已 经通过了阿里云安全审计。

] 说明:

BSearch-QueryBuilder插件仅支持6.3和6.7版本的阿里云Elasticsearch实例(不支持5.5.3)。

安装插件

!) 注意:

当您购买了阿里云Elasticsearch实例后,我们会为您赠送一个1核2G的Kibana节点。由于插件 需要耗费较多的资源,所以在安装插件前,您需要将该Kibana节点升级为2核4G或以上规格,详 细请参见#unique_47。



- 1. 登录阿里云Elasticsearch控制台,并购买一个Elasticsearch实例。
- 2. 单击实例ID > 可视化控制。
- 3. 在Kibana模块中,单击修改配置。

<	es-cn-vi 9	
基本信息	_	
ES集群配置	Kibana	🕎 DataV
插件配置		44320172727
集群监控		
日志查询		
安全配置	QueryBuilder DSL	
数据备份	• 新增Querybuilder插件,可在修改配置中查看。	・ 天猫双11、阿里云城市大脑同款数据大解。
可视化控制	• 强大而便捷的开发工具。	 ・ 高性能三维道染引擎。 ・ (本) (本) (本) (本) (本) (本) (本) (本) (本) (本)
▼ 智能运進	 ロリキ級とは第シュアリア、家区は第221日フェフリア。 	• 多种型类组件,又增多种数据类型的方所展示。
集群概况	相关文档参看Kibana言方介绍	相关文档参署DataV官方介绍进入控制台
健康诊断		

4. 在Kibana配置页面的插件配置区域,单击插件列表操作栏下的安装。

! 注意:

- ·确认安装后会触发Kibana节点重启,所以在安装过程中Kibana不能正常提供服务,为避免 影响您的Kibana操作,请确认后操作。
- ·如果您的Kibana规格低于2核4G,系统会提示您进行集群升配,请按照提示将您的Kibana节点升级到2核4G或以上规格。
- 5. 确认安装并重启Kibana节点。

重启后即可完成插件的安装,安装成功后,插件的状态显示为已安装。

<	es-cn-v 9			
Kibana配置	访问配置			
	Kibans公网访问: 【】			Kibana访问白名单: 0.0.0.0/0 修改
	1四1十月20日			
	□ 插件名称	状态	描述	操作
	In the second	● 已安装	Res rational	卸載
	network_vis	 未安装 	通过家引的两个字段,可以生成这两个字段的网络关系图	安装
门 说明	月:			

此过程可能耗时较长,请耐心等待。

卸载插件

1. 参考安装插件的操作步骤,在Kibana配置页面的插件配置区域,单击插件列表操作栏下的卸载。

!) 注意:

确认卸载后会触发Kibana节点重启,所以在卸载过程中Kibana不能正常提供服务,为避免影响您的Kibana操作,请确认后操作。

2. 确认卸载并重启Kibana节点。

重启后即可完成插件的卸载,卸载成功后,插件的状态显示为未安装。

2.1.5 BSearch-QueryBuilder插件使用介绍

BSearch-QueryBuilder又称高级查询,是一个纯前端的工具插件。通过BSearch-QueryBuilder插件,您可以无需编写复杂的DSL语句,而是以可视化的方式完成复杂的查询 请求。本文档为您介绍BSearch-QueryBuilder插件的使用方法,帮助您快速使用BSearch-QueryBuilder插件完成查询业务。

BSearch-QueryBuilder的特性

BSearch-QueryBuilder具有如下特性。

- ·简单易用: BSearch-QueryBuilder插件提供了可视化的界面点选操作来构造Elasticsearch的 DSL查询请求,无编码即可完成自定义条件的数据查询,减少了复杂的DSL的学习成本。也可辅助开发人员编写或验证DSL语句的正确性。
- · 方便快捷:已经定义的复杂查询条件会保存在Kibana中,避免重复构造查询条件。
- · 小巧轻盈:约占用14MB的磁盘空间,不会常驻内存运行,不影响Kibana和Elasticsearch的 正常运行。
- · 安全可靠: BSearch-QueryBuilder插件不会对用户的数据进行改写、存储和转发,源代码已 经通过了阿里云安全审计。

背景信息

Query DSL是一个Java开源框架,用于构建安全类型的SQL查询语句,能够使用API代替传统的拼 接字符串来构造查询语句。目前Query DSL支持的平台包括JPA、JDO、SQL、Java Collections 、RDF、Lucene以及Hibernate Search。

Elasticsearch提供了一整套基于JSON的DSL查询语言来定义查询。Query DSL是由一系列 抽象的查询表达式组成,特定查询能够包含其它的查询(如bool),部分查询能够包含过滤 器(如constant_score),还有的可以同时包含查询和过滤器(如 filtered)。您可以从ES支 持的查询集合里面选择任意一个查询表达式,或者从过滤器集合里面选择任意一个过滤器进行组 合,构造出复杂的查询。但编写DSL容易出错,仅有少数专业程序人员精通,QueryBuilder能够 帮助对Elasticsearch DSL不甚了解或者想提升编写效率的用户快速生成DSL。



准备工作

在使用BSearch-QueryBuilder插件前,请先购买一个Elasticsearch实例,实例版本为6.3或6.7(不支持5.5.3)。

	预付费	后付费					
	地域	华东1 (杭州)	华北2 (北京)	华东2 (上海)	华南1 (深圳)	印度 (孟买)	新加坡
1.2		香港	美国 (硅谷)	马来西亚 (吉隆坡)	德国 (法兰克福)	日本 (东京)	澳大利亚 (悉尼)
		印度尼西亚 (雅加达)	华北1 (青岛)	华北3 (张家口)			
	可用区	上海可用区B	*				
资源组	资源组	全部	▼ \$	认资源组	*		
	版本	6.7 with Commercial Feature	6.3 with Commercial Feature	5.5.3 with Commercial Feature			

畄 说明:

您也可以使用已经创建的实例,如果实例版本不符合要求,可进行版本升级。

安装BSearch-QueryBuilder插件

! 注意:

当您购买了阿里云Elasticsearch实例后,我们会为您赠送一个1核2G的Kibana节点。由于插件 需要耗费较多的资源,所以在安装插件前,您需要将该Kibana节点升级为2核4G或以上规格,详 细请参见#unique_47。



- 1. 登录阿里云Elasticsearch控制台,并购买一个Elasticsearch实例。
- 2. 单击实例ID > 可视化控制。

3. 在Kibana模块中,单击修改配置。

<	es-cn-v(9	
基本信息		
ES集群配置	Kibana	V DataV
插件配置		44120121212
集群监控		
日志查询		
安全配置	QueryBuilder DSL	
数据备份	• 新増Querybuilder插件,可在修改配置中查看。	• 天猫双11、阿里云城市大脑间款数据大屏。
可视化控制	• 强大而便捷的开发工具。	• 高性能三维渲染引擎。
智能运维	• 时序数据分析、数据图谱分析。	• 多种图表组件,支撑多种数据类型的分析展示。
集群概况	相关文档象看Kibena首方介绍	音相关文档参看DataV官方介绍进入控制台
健康诊断		

4. 在Kibana配置页面的插件配置区域,单击插件列表操作栏下的安装。



- ·确认安装后会触发Kibana节点重启,所以在安装过程中Kibana不能正常提供服务,为避免 影响您的Kibana操作,请确认后操作。
- ·如果您的Kibana规格低于2核4G,系统会提示您进行集群升配,请按照提示将您的Kibana节点升级到2核4G或以上规格。
- 5. 确认安装并重启Kibana节点。

重启后即可完成插件的安装,安装成功后,插件的状态显示为已安装。

<	es-cn-v 9											
Kibana配置	访问和著											
	Kibana公司访问: Kibana访问台名单: 0.0.0/0 例文: 病化亞常											
	插件配置											
	播件名称	状态	描述	操作								
	 Instruction 	 已安装 	Resultance and	卸载								
	network_vis	• 未安装	通过家引的两个字段,可以生成这两个字段的网络关系图	安装								
道 说明	月:											

此过程可能耗时较长,请耐心等待。

使用BSearch-QueryBuilder插件

- 1. 返回Elasticsearch实例的可视化控制页面,单击Kibana模块中的进入控制台。
- 2. 输入Kibana控制台的用户名和密码,单击登录。

默认的用户名为elastic,密码为您购买实例时设置的密码。

3. 在Kibana控制台中, 单击Discover > Query。



4. 在查询区域选择查询和过滤条件,单击提交。

提交成功后,系统会显示查询结果。

	·	2 hits	New	Save	Open	Share	Query Reportin	g C Auto-refresh
	kibana	email v 通配符(Wildcard) v "iga" X					2	۵
Ø	Discover							
1	Visualize							
	Darbhoard							
	Territor	Search (e.g. status:200 AND extension:PHP)						Options Q
	Martin Landa	querytouikder0 Add a filter +						Actions •
-	Machine Learning	All filters: Enable Disable Pin Unpin Invert Toggle Remove						
=	APM	megacorp G _source						
*	Graph	mail: p2ipa0123.con first_mass: p2 last_mass: Smith age: 27 about: I love to go rock climbing interests: sports, music_id:	5 _typ	c enplo	yee _ind	ma: megaco	mp _score: 1	
2	Dev loois	Selected Fields						4
*	Monitoring	7 _source / email: John_ighthhg.com first_news: John last_news: Smith age: 25 about: I love to go rock climbing interests: sports, music	_id: 1	_type:	employee	_index: n	egacorp _seare:	1
•	Management	Available Helds O						
		t_index						
		5 500 B						
单	击查询	区域的	动]—∙	个	子达	滤条	件;单
击	×	可删除一个查询或过滤条件。						

具体的查询方式请参见下文的BSearch-QueryBuilder插件使用示例。

BSearch-QueryBuilder插件使用示例

BSearch-QueryBuilder支持模糊查询、多条件组合查询和自定义时间范围查询等多种查询方式。

・模糊查询

下图中表示对email这个条件进行模糊查询,并要求email中模糊匹配iga。

email	•	通配符(Wildcard)	▼*iga*		•
+ - =					
submit					

最终得到的匹配结果如下。

Add a filter +													Actions	
kibana_sample_data_ecommerce	• 0					2	018-04-24 14:11:12 - 2	1019-04-24 14:11:12 - Auto	•					
Belected fields 1 customer_frat, name 1 customer_pender 1 email	3	- 02 - 02 - 02 - 02 - 02 - 02 - 02 - 02	2016-06-01	2018-07-01	2018-08-01	2018-08-01	2018-10-01	2018-11-01	2016-12-01	2018-01-01	2019-02-01	2019-03-01	2019-04-01	
# taxtul_stal_price Available fields	•	Terr			and and an and			order_date per week						
t_id		2010-03-00 11:33:07	EDWALE		Abiasil			abiasil deslayar-famil				60.00		
t _index			TUPEL		Pergert			Border Concerns - Inite	,			07120		
# _score		2019-04-04 19:00:58	FDMLE		Abigail			abigail@banks-family.	222			53.98		
t_type		2819-03-29 04:21:07	FEMALE		Abigail			abigail@potter-family	. 222			47.99		
t category		2819-83-22 84:38:24	FEMALE		Abigail			abigail@farmer-family	. 222			66.98		
t currency		2819-83-14 18:58:34	FEMALE		Abigeil			abigail goowell-family	. 222			95.98		
t customer_full_name	- 5	2019-04-06 17:00:38	FEMALE		Abigail			abigail@willis-family	. 222			117.98		
# customer_id	5	2019-03-17 05:48:00	FEMALE		Abigail			abigail@goodwin-famil	y. 222			116.98		
t customer_last_name	1	2019-04-01 00:00:29	FEMALE		Abigail			abigail@webb-family.z	22			147.98		
t day_of_week	10	2019-03-22 09:17:46	FEMALE		Abigail			abigail@webb-family.z	22			53.98		
# day_ot_week_i		2019-03-20 00:55:38	FEMALE		Abigail			abigall@perry-family.	222			43.98		

・ 多条件组合查询

下图的查询条件表示index必须为tryme_book,同时要对type 进行过滤,要求type等于大学教辅、数学、对外汉语教学或大学教材。

	type	•	匹配(Match)	\$ 大学教辅	×
_	type	\$	匹配(Match)	\$ 数学	×
-	type	\$	匹配(Match)	\$ 对外汉语教学	×
-	type	\$	匹配(Match)	\$ 大学教材	×
	+				

最终得到的匹配结果如下。

_querybuilder Add a filter +		Actions
tryme_book		erve, e
Selected fields		· Date: Control Name: IN FIGURE Matter Dat. of Patient & Mather Dow patient, Statistic Dow patient, Statistic Date: Date
7 _source Available fields	•	* index: (TypE_Dock type: A TR H ender: ASH prior: 4 minut: jon paina: ASHRIDE amo: BFARSTONGAS HEDDREGHES index: jon bala: jon index: GEDREGHESDREGHESDREGHESDREFF, DFDREFF. DFDREGHESDREGHESDREGHESDREGHESDREFF. DFDREFF. DFDREFF
Popular 1 author		b Linker, Type, Dok Type, X T Mit Heller, S.A. TREE, C.A. REE, A.T. BOR prior, F Ministi Jon patient REETING AND TREETING - DECKY JUNCE DIAL CADE (SUB) Linket of belief on Linke on Linket of Linket of Linket on Li
t helo t index		b inter (species uppe) 2.9 R.W maker, 88 price, 0 minui (so paline) ANDALDE ANDALDE AND EXTERNAL OF COMPANY OF COMPANY.
t into t label		* Index (1992)49 April 27 B.C. Barry (2017) 1 Andrew (1992)40 ADDREW (2017)102 ADDREW (2017)102 (1992)40 (1990
t name # price		J Lines: Type_Lines Type (A TRANS. (D) 80 Derrol5.) priors: F Annalis (on patient AFRETARS Lones: DYRETAG BEEN Lanes: (on halies (on line) = 542858078. BORDEREL-FEDERELON-FEDEREDERE, REDEREDERE, REDEREDEREDERE, REDEREDERE, REDEREDEREDERE, REDEREDEREDERE, REDEREDERE, REDEREDEREDERE, REDEREDEREDERE, REDEREDEREDERE, REDEREDERE, REDEREDERE, REDEREDERE, REDEREDERE, REDEREDERE, REDEREDERE, REDEREDEREDERE, REDEREDERE, REDEREDEREDERE, REDEREDEREDERE, REDEREDEREDERE, REDEREDEREDERE, REDEREDEREDEREDEREDERE, REDEREDEREDEREDEREDERE, REDEREDEREDEREDEREDEREDEREDEREDEREDERED
t publish t type		* Johns Tyne, John XVEN Hannes NEE PER print # Annals jon palant ERETARYTREASUMMERPES CLEPPERDOCERE DID Jahns jon Jahns (CLEPPERDOCERETING Jahns Tyne, Johns) Jahns Tyne, Johns
t Animal t jid t jindex		b intent (type:loss type: X R M content RD/NET# April: 200 paine: LARGERT water EARSET +-C'REGUINT ATTEXAGE (T) intent (LARGERT +-C'REGUINT +-C'R
# _score 1 _type		* Jahon 1990,000 Sppc ATEX andar: UPH prior & Andari pro patient ATEX.DE same DESIGNER USE Jahon pro balanci pro b

・ 自定义时间范围查询

当您需要对时间字段进行筛选时,可使用时间类型的筛选功能。下图中对utc_time进行时间范围筛选,查询[当前时间-240天,当前时间]范围内的数据。

utc_time	\$ 最近几天(In the last) ◆	240	days 🗶
+ - =			
submit			

最终得到的匹配结果如下。

_querybuilderAdd a filter +													Actions +
kibana_sample_data_logs	• ()					2018-04-24 14:19:53 -	2019-04-24 14:19:53 - Auto	0				
Selected fields		1,500											
7 _source		1,000											
Available fields	•	500											
t _index		2018-05-01	2018-06-01	2018-07-01	2018-08-01	2018-09-01	2018-10-01	2018-11-01	2018-12-01	2019-01-01	2018-02-01	2019-03-01	2019-04-01
# _score													
t_type		Time -	_source										
t agent		> 2019-04-24 14:00:49	referer: http://faceboo	k.com/success/lloyd-ham	and elientip: 230.1	4.21.43 responses 2	tagai success,	info message: 230.114.21.43	[2018-09-12706:00	8:49.5992] "GET /styles/	app.css HTTP/1.1" 200	8700 "-" "Mozilla/5.0	8 (X11; Linux 1686) AppleMebKit/
# bytes			s wels https://cdn.elas	tic-elastic-elastic.org	styles/app.css memor	yı - hostı cdn.ela	stic-elastic-elastic	.org utc_time: 2018-09-12	14:00:49 machine.rem:	7,516,192,768 machine	.es: ios agent: Mozil	lla/5.0 DK11; Linux S	186) AppleWebKit/534.24 (KHTML,
t clertip			like Gecko) Chrome/11.0.	696.50 Safari/534.24 👪	: 230.114.21.43 byte	s: 8,700 timestamp:	2019-04-24 14:00:4	_id: J6EffWk8Qm8RQo-IrSpw	_type: _60C _index:	kibana_sample_data_log	score: -		
t extension		· 2019-04-24 13:53:53	referer: http://ww.ela	stic-elastic-elastic.com	/success/steven-lind	ey elientip: 158.64	134.248 response:	200 taga: success, info me	158.64.134.248	8 [2018-09-12705:53:	53.1192] "GET / HTTP/1	.1" 200 5139 "-" "Mo	rilla/4.0 (compatible; MSIE 6.0;
# geo.coordinates.lat			Windows NT 5.1; SV1; .NE	T CLR 1.1.4322)" reques	t: / geo.srcdest: R	I:RU geo.src: RU geo	.dest: RU geo.coos	dinates.lat: 42.557 geo.coc	ordinates.lon: -92.4	phpnenory: - extension	a: index: kibana_sar	ple_data_logs wris	https://www.elastic.co/downloads
# geo.coordinates.ion			senery: - host: www.e	lastic.co ute_time: 28	18-09-12 13:53:53 mad	hise.ran: 19,327,352	832 machine.os: 0	ix agent: Mozilla/4.0 (comp	atible; MSIE 6.0; Wind	dows NT 5.1; SV1; .NET C	UR 1.1.4322) ip: 158.	64.134.248 bytes: 5	,139 timestamp: 2019-04-24 13:5
t geo.dest				and share she s		and the particular							
t geo.src		> 2019-04-24 13:43:11	referer: http://faceboo	k.com/success/donald-ho	nquest clientip: 11	.214.211.68 response	a 200 tages succe	is, info message: 117.214.2	11.68 [2018-09-12]	105:43:11.2832] "GET /el	asticsearch/elasticsea	irch-6.3.2.deb HTTP/1	1" 200 8491 "-" "Mozilla/5.0 (X
t geo.srcdest			<pre>ll; Linux x86_64; rv:6.0 bana sample data loos w</pre>	<pre>al) Gecko/20110421 Fire fix https://artifacts.e</pre>	ox/6.dol" requests ;	elasticsearch/elastic asticsearch/elasticse	sarch-6.3.2.deb mee	eo.srodest: CN:ER geo.src: ory: - host: artifacts.eU	ON geo.dost: DR geo ostic.co uto time: 20	coordinates.lat: 40.52	2 geo.coordinates.los ine.ram: 6.442.450.944	nachine.os: win 7	seent: Mozilla/5.0 (X11: Linux
t host			x86_64; rv:6.0a1) Gecks/	20110421 Firefox/6.0a1	ip: 117.214.211.68 b	ytes: 8,491 timestas	p: 2019-04-24 13:4	1:11 _14: A6EffwkBQreRQo-Ir	Spw _type: _doc _ind	lex: kibana_sample_data_	logs _seares -		
t index		> 2810-84-24 13:42:14											
t ip		·	(KHTML, Like Gecko) Chro	.com/warning/patrick-ba me/11.0.695.50 Safari/5	4.24" request: /ent:	rprise con.arcdost:	IN: CN geo.sre: IN	pee.dest: Ol pee.coordinat	12010-09-12705:43	wordinates.los: -88.549	150 H119/1.1" 503 0 "-	elen: index: kiban	a sample data logs writ http
t machine.os			s://www.elastic.co/downl	oads/enterprise memory	- bost: www.elast	c.co ute_time: 2018	-09-12 13:42:14 mad	hime.ram: 15,832,385,536 m	achine.os: 105 agent:	Mozilla/5.0 (X11; Linu	ox 1686) AppleWebKit/5	14.24 (KHTML, like Ge	cko) Chrome/11.0.696.50 Safari/5
# machine.ram			34.24 ip: 122.62.233.59	bytes: 0 timestamp:	019-04-24 13:42:14	id: xaEffWk8Qm8RQo-I	pikc _type: _doc _	<pre>index: kibana_sample_data_b</pre>	ogs _score: -				

结合以上说明,构造一个复杂的查询条件,如下图所示。

0 hits				
	Match) 🔻 同寺			
AND \$				
type \$	匹配(Match) 💲	数学	×	
type 🜲	匹配(Match) 🗘	学习方法	×	
price 🛊	小于等于(<=) 🗘	20		×
OR ¢	× 通配符(Wildcard ND ↓ × info ↓ 匹配(Mato + i言) ◆*数学* ch) ◆*数学*	- ×	
+ - =				
submit				

而实际对应的DSL如下图所示。



可以看出通过使用BSearch-QueryBuilder插件,可以极大地降低Elasticsearch查询的难度。

2.1.6 BSearch-Label插件使用介绍

BSearch-Label是一个纯前端的数据打标插件。通过BSearch-Label插件,您可以无需编写复杂的DSL语句,而是以可视化的方式完成数据打标。本文档为您介绍BSearch-Label插件的使用方法,帮助您快速使用BSearch-Label插件完成查询业务。

背景信息

通常情况下,在分析数据的时候,您可能不仅是单纯的浏览,而是希望通过某些查询条件对数据进 行分析,并对某个字段(或者新增一个字段)赋予一个特殊的值(标签)来标注不同的数据,这一 过程被称为"打标"。对数据打标后,您可以根据这个标签进行聚合分类统计,也可以根据标签的 不同值进行快速过滤。标注的数据还可以直接为后续的流程所使用。

安装BSearch-Label插件

! 注意:

当您购买了阿里云Elasticsearch实例后,我们会为您赠送一个1核2G的Kibana节点。由于插件 需要耗费较多的资源,所以在安装插件前,您需要将该Kibana节点升级为2核4G或以上规格,详 细请参见#unique_47。

Vibana#c	
	是
Kibana规格 2核4G	•

- 1. 登录阿里云Elasticsearch控制台,并购买一个Elasticsearch实例。
- 2. 单击实例ID > 可视化控制。
- 3. 在Kibana模块中,单击修改配置。

<	es-cn-v(9	
基本信息		
ES集群配置	Kibana	V DataV
插件配置		
集群监控		
日志查询		
安全配置	QueryBuilder DSL	
数据备份	• 新道Quervbuilder插件。可在惨战配置中查看。	 天猫双11、阿里云城市大脑同款数据大屏。
可视化控制	• 强大而便捷的开发工具。	 高性能三维渲染引擎。
▼ 智能运维	 时序数据分析、数据图谱分析。 	• 多种图表组件,支撑多种数据类型的分析展示。
集群概况	相关文档参看Kibana窗方介绍 修改配置 进入控制台	相关文档参覆DataV官方介绍进入控制台
健康诊断		

4. 在Kibana配置页面的插件配置区域,单击插件列表操作栏下的安装。



重启后即可完成插件的安装,安装成功后,插件的状态显示为已安装。

<	es-cn-v 9			
Kibana配置	访问配署			
	WJP中国上重 Kibana公開访问:	Kibana访问白名单: 0.0.0.0/0 修改		
	☐ 插件名称	状态	描述	操作
	 Networks 	 ● 已安装 	MALLON-MAL	如戰
	network_vis	● 未安装	通过家引的两个字段,可以生成这两个字段的网络关系图	安装



此过程可能耗时较长,请耐心等待。

使用BSearch-Label插件

- 1. 返回Elasticsearch实例的可视化控制页面,单击Kibana模块中的进入控制台。
- 2. 输入Kibana控制台的用户名和密码,单击登录。

默认的用户名为elastic, 密码为您购买实例时设置的密码。

3. 在Kibana控制台中, 单击Discover > 打标。

!) 注意:

```
在查询前,请确保您已经创建了一个索引模式。否则需要在Kibana控制台中,单
击Management,再单击Kibana模块中的Index Patterns > Create index pattern,按照
提示创建一个索引模式。
```

- 4. 根据您的需求,选择以下任意一种方式完成数据打标。
 - ・对已有字段进行打标。

如下示例,先查询到名字是张三的数据,然后选择age字段,将其标记为18,单击确认打标。



打开历史打标开关,可查看历史打标任务详情。

	kibana	1 次命中					新建	保存	打开 共	享 检查	打标	C自动刷新
	KIDalla	□ 自定义打标字段										0
Ø	Discover	确认打标 请选择打标字印	9 -	请输入打标内容								
얍	可视化									Entri		
50	仪表板	「11 小时细 任冬 ID		表引之	任冬描述	开始时间		毛Rt(新り)))))))))))))))))))))))))))))))))))))))	在(面东教)	(首数)
₩	Timelion	DpTSx75xSiOvvI7rT6HdYw:		tryme_book	update-by-query [tryme_bo	2019-08-08 15:23:02	(0.2860		ALC	1/1	ICARA/
	Maps											
A	藏础设施	>_ 搜索(例如, status:2	200 AND	extension:PHP)						选项	e	Refresh
			Class de									Actions
		tryme book	()	FOURTE								Actions
9	运行时间	选定字段			male and 10 hobby thit 00 Fit	id: 020652.0EEACDC178884	tupor dos	indexe	trumo be	ok score	1 575	
Å	开发工具	? _source		name: SR gender:	mate age: 10 mobby: 26.28, 39-6-38	_10: 0211102WDE5A5D511NWN44	_cypedoc	_mdex:	cryme_oc	lok _score	. 1.575	
٩	管理	可用字段										
		t_id										
		# score										
		t_type										
		# age										
		t gender										
	默认值	t hobby										
4	50	t name										

・新増字段进行打标。

如下示例,先查询到名字是张三的数据,然后勾选自定义打标字段,新增一个字段tag,将其 标记为teenager,单击确认打标。

	kibana	1 次命中						新建	保存	打开	共享	检查	打标	C自动刷新
	Kibana	■ 自定义打标字段												0
Ø	Discover	确认打标 tag		teenager										
斺	可视化												_	
罰	仪表板	▶ 打标明细									历	史打标(> 1/1
57	Timelion	任务 ID DoTSy75ySiOuxI7rT6HdVwr		索引名 tryme_book	任: undate-by-gi	务描述	开始时间 2019-08-08 15:23:02		耗时(秒)			进度(更新数//	总数)
Ŵ	Inneion	oproxisiowintendriw.		dynie_book	upuate-by-qu	ery (d yme_bo	2013-00-00 13:23:02		0.2000				17.1	
\$	Maps													
G	基础设施	>_ 搜索(例如, status:	200 AND	extension:PHP)								选项	C	Refresh
I	日志	name: "张三" Add	a filter 🕇											Actions
ۍ	运行时间	tryme_book	0	_source										
<u>ب</u>	五金丁目	选定字段		name: 张三 gender:	male age: 18	nobby: 篮球、羽毛球	_id: 02nhb2wBE5ASBS1TNWN4 _t	ype: _doc	_index	tryme	_book	_score:	1.575	
		? _source												
భ	管理	可用字段 •												
		t index												
		# score												
		t_type												
		# age												
		t gender												
	ma217#	t hobby												
	N/VIII	t name												
	折叠													

查看打标结果。

	kibana	1 次命中				新建 保存 打开	共享 检查 打标	C自动刷新
		自定义打标字段						9
ø		确认打标 请选择打标字段	 					
尬	可視化	. +T+2004m						() 10
50	仪表板	F 1119/F11曲 任色 ID	委司名	任岳描述	开始时间	\$35.8±1/3001	の実行林 ●	(总数)
₩	Timelion	DpTSx75xSiOvvl7rT6HdYw:	tryme_book	update-by-query [tryme_bo	2019-08-08 16:34:21	0.0849	1/1	
2	Maps	DpTSx75xSiOwI7rT6HdYw:	tryme_book	update-by-query [tryme_bo	2019-08-08 15:23:02	0.2860	1/1	
G	基础设施							
E	日志	>_ 搜索 (例如, status:20	0 AND extension:PHP)				选项	C Refresh
্য	运行时间	name: "姚三" Add a fi	lter +					Actions •
ę	开发工具	tryme_book 🔹	G _source					
@	管理	选定字段 2 source	• name: 💥 = gender:	male tag teenager age: 18 hob	by: 篮球、羽毛球 _id: 02nhb2wB	ESASBS1TNWN4 _type: _d	loc _index: tryme_bool	k
		可用学段	_30010. 1.373					
		t_id						
		t _index						
		# _score						
		t_type						
		# age						
	默认信	t gender						
		t hobby						
÷	折叠	t name						

3 Elasticsearch自带功能

Elasticsearch官方文档

阿里云Elasticsearch是基于Elasticsearch官方开源的对应版本进行构建,详情请参见Elasticsea rch权威指南中文版。

SDK客户端

客户端仅支持使用HTTP的方式进行访问,您可以使用Elasticsearch官方提供的Java REST Client。

Elasticsearch Clients

- Java REST Client [6.4] other versions
- Java API [6.4] other versions
- JavaScript API
- Groovy API [2.4] other versions
- .NET API [6.x] other versions
- PHP API [6.0] other versions
- Perl API
- Python API
- Ruby API
- Community Contributed Clients

4 快照备份及恢复

您可以使用snapshot API来备份您的阿里云Elasticsearch集群。snapshot API会获取集群当前的状态和数据,然后保存到一个共享仓库里。这个备份过程是智能的。

第一次进行快照时,系统会备份您所有的数据,但后续所有的快照仅备份已存快照和新数据之间的 差异。随着数据快照的不断进行,备份也在增量的添加和删除。这意味着后续备份会相当快速,因 为它们只传输很小的数据量。

🗾 说明:

您可以参考OSS快照迁移Elasticsearch,使用快照方式完成自建Elasticsearch迁移至阿里 云Elasticsearch。

!! 注意:

本文代码中的<1>、<2>、<3>这三个标记用于标识位置,方便对指定位置代码描述。实际执行对应代码时,需去掉有包含这三个类型的标记。

前提条件

在进行快照备份前,您需要首先#unique_52并新建一个OSS Bucket(建议购买标准存储类型的OSS Bucket,不支持归档存储类型的OSS Bucket),且OSS Bucket的区域必须与您需要快照备份的阿里云Elasticsearch集群所在区域保持一致。

Bucket 名称	0/63
区域	坐东1 (杭州) ✓
	相同区域内的产品内网可以互通;订购后不支持更换区域,请谨慎选择
	您在该区域下没有可用的 <mark>存储包、流量包。</mark> 建议您购买资源包享受更多优惠,点 击 <mark>购买。</mark>
Endpoint	oss-cn-hangzhou.aliyuncs.com
存储类型	标准存储 低频访问 归档存储
	标准: 高可靠、高可用、高性能, 数据会经常被访问到。
	如何选择适合您的存储类型?
读写权限	私有 公共读 公共读写
	私有:对文件的所有访问操作需要进行身份验证。
实时日志查询	
	USS 与口志服务深度结合,光器提供取近7天内的 USS 买时口志宣闻。并通该切能后,用户可对 Bucket 的访问记录进行实时查询分析, <mark>了解详情</mark>

创建仓库

- · <1>: endpoint为您的OSS Bucket对应的内网地址。详情请参见#unique_54中的ECS访问的内网Endpoint一栏。
- · <2>: OSS Bucket的名称, 需要一个已经存在的OSS Bucket。
- · <3>: base_path设置仓库的起始位置,默认为根目录。

限制分块大小

当您上传的数据非常大时,可以限制snapshot过程中分块的大小,超过这个大小,数据将会被分块上传到OSS中。

```
POST _snapshot/my_backup/ <1>
{
    "type": "oss",
    "settings": {
        "endpoint": "http://oss-cn-hangzhou-internal.aliyuncs.com",
        "access_key_id": "xxxx",
        "secret_access_key": "xxxxxx",
        "bucket": "xxxxxx",
        "bucket": "500mb",
        "base_path": "snapshot/" <2>
    }
}
```

・ <1>: 请求方式,注意使用POST而不是PUT,这会更新已有仓库的设置。

· <2>: base_path设置仓库的起始位置,默认为根目录。

列出仓库信息

GET _snapshot

您也可以使用GET _snapshot/my_backup获取指定仓库的信息。

备份快照迁移

按照以下步骤将当前快照迁移到另外一个集群。

- 1. 将当前快照备份到OSS。
- 2. 在新的集群上创建一个快照仓库(相同的OSS)。
- 3. 设置base_path为备份文件的路径。
- 4. 执行恢复备份的命令。

快照所有打开的索引

一个仓库可以包含多个快照,每个快照跟一系列索引相关(比如所有索引,一部分索引,或者单个 索引)。当创建快照的时候,您可以指定感兴趣的索引,然后给快照取一个唯一的名字。

快照命令

最基础的快照命令如下。

PUT _snapshot/my_backup/snapshot_1

以上命令会备份所有打开的索引到my_backup仓库下,并保存在名称为snapshot_1的快照中。这个调用会立刻返回,然后快照会在后台运行。

如果您希望在脚本中一直等待到完成,可通过添加wait_for_completion实现。

PUT _snapshot/my_backup/snapshot_1?wait_for_completion=true

以上命令会阻塞调用直到快照完成。如果是大型快照,需要很长时间才能返回。

快照指定索引

系统默认会备份所有打开的索引。如果您在使用Kibana,并且考虑到磁盘空间大小因素,不想把 所有诊断相关的.kibana索引都备份起来,那么可以在快照您的集群时,指定需要备份的索引。

```
PUT _snapshot/my_backup/snapshot_2
{
    "indices": "index_1,index_2"
}
```

以上快照命令只会备份名称为index1和index2的索引。

列出快照信息

有时您可能会忘记仓库里的快照细节,特别是快照按时间划分命名的时候(比如backup_201 4_10_28)。

当您需要查看仓库中的某个快照时,可对仓库和快照名发起一个GET请求,获取单个快照信息。

GET _snapshot/my_backup/snapshot_2

返回结果中包括了快照相关的各种信息,如下所示。

```
{
"snapshots": [
   {
      "snapshot": "snapshot_2",
      "indices": [
         ".marvel_2014_28_10",
         "index1",
"index2"
      ],
"state": "SUCCESS",
      "start time": "2014-09-02T13:01:43.115Z",
      "start_time_in_millis": 1409662903115,
      "end_time": "2014-09-02T13:01:43.439Z",
      "end_time_in_millis": 1409662903439,
      "duration in millis": 324,
      "failures": [],
      "shards": {
         "total": 10,
         "failed": 0,
         "successful": 10
      }
   }
٦
```

}

您也可以使用_all替换掉具体的快照名称,获取一个仓库中所有快照的完整列表。

```
GET _snapshot/my_backup/_all
```

删除快照

您可以对仓库和快照名发起一个DELETE请求,删除所有不再使用的快照。

```
DELETE _snapshot/my_backup/snapshot_2
```

!) 注意:

- · 请使用delete API来删除快照,而不能使用其他机制(比如手动删除)。因为快照是增量的,很多快照可能依赖于之前的备份数据。delete API能够过滤出还在被其他快照使用的数据,会只删除不再被使用的备份数据。
- ・如果您进行了一次人工文件删除,您将会面临备份严重损坏的风险,因为您删除的文件可能是 还在使用中的数据。

监控快照进度

您可以使用wait_for_completion,对快照进行监控,但其仅提供了基础的监控形式。如果您需要对中等规模的集群进行快照监控,可能会不够用。您可以通过以下两种方式获取详细的快照进度 信息。

· 给快照ID发送一个GET请求。

```
GET _snapshot/my_backup/snapshot_3
```

如果您在执行这个命令时,快照还在进行中,那么您会看到它什么时候开始,运行了多久等信 息。

! 注意:

这个API用的是与快照机制相同的线程池,当您在快照非常大的分片时,状态更新的间隔会很大,因为API在竞争相同的线程池资源。

・使用_status API获取快照的状态信息。

```
{
    "snapshots": [
    {
        "snapshot": "snapshot_3",
        "repository": "my_backup",
        "state": "IN_PROGRESS", <1>
        "shards_stats": {
             "initializing": 0,
        "
```

```
"started": 1, <2>
                 "finalizing": 0,
                "done": 4,
"failed": 0,
                 "total": 5
 },
"stats": {
    bor
                "number_of_files": 5,
"processed_files": 5,
"total_size_in_bytes": 1792,
                 "processed_size_in_bytes": 1792,
                "start_time_in_millis": 1409663054859,
"time_in_millis": 64
},
"indices": {
    "index_3": {
        "shards_stats": {
        "initializing"
        "initialing"
        "initializing"

                                             "initializing": 0,
                                             "started": 0,
"finalizing": 0,
                                             "done": 5,
"failed": 0,
"total": 5
                             },
"stats": {
    "sumber"
}
                                             "number_of_files": 5,
"processed_files": 5,
                                             "total_size_in_bytes": 1792,
                                             "processed_size_in_bytes": 1792,
                                             "start_time_in_millis": 1409663054859,
                                             "time_in_millis": 64
                             },
"shards": {
                                             "0": {
                                                          "stage": "DONE",
                                                           "stats": {
                                                                         "number_of_files": 1,
                                                                         "processed_files": 1,
                                                                         "total_size_in_bytes": 514,
                                                                         "processed_size_in_bytes": 514,
                                                                         "start_time_in_millis": 1409663054862,
                                                                         "time_in_millis": 22
                                                           }
                                             },
```

• • •

- <1>: 快照的状态。一个正在运行的快照,会显示为IN_PROGRESS。
- <2>: 正在快照传输的分片数量。为1时,表示这个特定快照有一个分片还在传输(另外四个 已经完成)。

shards_stats响应不仅包括快照的总体状况,也包括下钻到每个索引和每个分片的统计 值。此参数为您展示了有关快照进度的非常详细的视图。分片可以在不同的完成状态:

- INITIALIZING: 分片在检查集群状态, 查看是否可以被快照。此过程一般是非常快的。
- STARTED: 数据正在被传输到仓库。
- FINALIZING: 数据传输完成,分片正在发送快照元数据。
- DONE: 快照完成。
- FAILED: 快照过程中碰到了错误,这个分片/索引/快照不可能完成。查看您的日志获取更 多信息。

取消快照

如果您想取消一个快照,可以在任务进行中的时候,执行以下命令删除快照。

DELETE _snapshot/my_backup/snapshot_3

以上命令会中断快照进程并删除仓库中进行到一半的快照。

从快照恢复

在需要恢复索引的目标阿里云Elasticsearch实例上,再执行和之前相同的创建仓库命令。您可以 根据实际情况,通过以下两种方式进行快照恢复。

・如果您已经备份过数据,可直接在需要恢复的快照ID后面加上_restore。

```
POST _snapshot/my_backup/snapshot_1/_restore
```

系统默认会恢复这个快照里的所有索引。例如snapshot_1包括五个索引,那么这五个索引都会 被恢复到您的集群中。您也可以指定需要恢复的索引,具体请参考上文的快照指定索引。

·恢复指定的索引,并对恢复的索引进行重命名(如果您需要在不替换现有数据的前提下,恢复旧版本的数据来验证内容,或者做其他处理,可以使用此方式进行快照恢复)。

```
POST /_snapshot/my_backup/snapshot_1/_restore
{
    "indices": "index_1", <1>
    "rename_pattern": "index_(.+)", <2>
    "rename_replacement": "restored_index_$1" <3>
```
}

以上命令会恢复index_1到您集群里,并且重命名成了restored_index_1。

- <1>: 只恢复index_1索引,忽略快照中的其他索引。
- <2>: 查找正在恢复的索引,该索引名称需要与提供的模板匹配。
- <3>: 重命名查找到的索引。
- ・使用wait_for_completion, 查看恢复信息, 直到恢复完成。

```
POST _snapshot/my_backup/snapshot_1/_restore?wait_for_completion=
true
```

和快照类似,_restore API会立刻返回,恢复进程会在后台进行。如果您更希望您的HTTP调用阻塞直到恢复完成,可以参考以上命令添加wait_for_completion。

监控恢复操作

■ 说明:

从仓库恢复数据借鉴了Elasticsearch里已有的现行恢复机制。在内部实现上,从仓库恢复分片和 从另一个节点恢复是等价的。

您可以通过recovery API来监控恢复的进度。

 ・
 监控指定索引的恢复状态。

```
GET restored_index_3/_recovery
```

recovery API是一个通用的API,可以用来展示您集群中移动着的分片状态。

・ 查看您集群中的所有索引的恢复信息(可能包含跟您的恢复进程无关的其他分片的恢复信息)。

GET /_recovery/

示例输出结果如下(根据您集群的活跃度,输出可能会非常多)。

```
{
"restored_index_3" : {
    "shards" : [ {
        "id" : 0,
        "type" : "snapshot", <1>
        "stage" : "index",
        "primary" : true,
        "start_time" : "2014-02-24T12:15:59.716",
        "stop_time" : 0,
        "total_time_in_millis" : 175576,
        "source" : { <2>
            "repository" : "my_backup",
            "snapshot" : "snapshot_3",
            "index" : "restored_index_3"
        },
        "target" : {
    }
}
```

```
"id" : "ryqJ5l05S4-lSFbGntkEkg",
                                         "hostname" : "my.fqdn",
"ip" : "10.0.1.7",
"name" : "my_es_node"
                      },
"index" : {
    "files" : {
        tol" :
        tol" :

                                                             "total" : 73,
                                                            "reused" : 0,
"recovered" : 69,
                                                            "percent" : "94.5%" <3>
                                    },
"bytes" : {
    "total" : 79063092,
    "reused" : 0,
    "recovered" : 68891939,
    "sercent" : "87.1%"
                                          },
                                           "total_time_in_millis" : 0
                         },
                          "translog" : {
"recovered" : 0,
                                           "total_time_in_millis" : 0
                        },
"start" : {
    tack in

                                           "check_index_time" : 0,
                                           "total_time_in_millis" : 0
                           }
       } ]
}
}
```

- <1>: type字段定义了您恢复的类型。snapshot表示这个分片是在从一个快照恢复的。
- <2>: source字段定义了作为恢复来源的特定快照和仓库。
- <3>: percent字段定义了恢复的状态。94.5%表示这个特定分片目前已经恢复了94.5%的 文件。

输出结果会展示所有正在恢复中的索引,并列出这些索引里的所有分片。同时每个分片中会显示 启动/停止时间、持续时间、恢复百分比、传输字节数等统计值。

取消恢复

您可以通过删除正在恢复的索引,取消一个恢复(因为恢复进程其实就是分片恢复,发送一个DELETE API修改集群状态,就可以停止恢复进程)。

```
DELETE /restored_index_3
```

如果restored_index_3正在恢复中,以上删除命令会停止恢复,同时删除所有已经恢复到集群中的数据。

参考文档: Snapshot And Restore。

5 访问控制

5.1 授权资源类型

资源类型及描述

阿里云Elasticsearch支持资源类型及完整资源描述如下:

资源类型	完整资源描述
instances	acs:elasticsearch:\$regionId:\$accountId: instances/*
instances	acs:elasticsearch:\$regionId:\$accountId: instances/\$instanceId
vpc	acs:elasticsearch:\$regionId:\$accountId: vpc/*
vswitch	acs:elasticsearch:\$regionId:\$accountId: vswitch/*

· \$regionId: 应为某个region的id, 或使用*代替。

· \$accountId: 是您云账号主账号的数字id, 或使用*代替。

· \$instanceId: 应为某个具体 es 实例id, 或使用*代替。

instances授权列表

🗾 说明:

以下资源描述为简写形式,已省略相同部分内容,完整资源描述参见文档顶部。

・ 实例基本操作

Action	Action 描述	资源描述
elasticsearch:CreateInst ance	创建实例	instances/*
elasticsearch:ListInstance	查看实例列表	instances/*
elasticsearch:DescribeIn stance	查看实例描述	instances/* 或 instances/\$instanceId
elasticsearch:DeleteInst ance	删除实例	instances/* 或 instances/\$instanceId

Action	Action 描述	资源描述
elasticsearch:RestartIns tance	重启实例	instances/* 或 instances/\$instanceId
elasticsearch:UpdateInst ance	更新实例	instances/* 或 instances/\$instanceId

・插件相关

Action	Action 描述	资源描述
elasticsearch:ListPlugin	获取插件列表	instances/\$instanceId
elasticsearch:InstallSys temPlugin	安装系统插件	instances/\$instanceId
elasticsearch:UninstallP lugin	卸载插件	instances/\$instanceId

・网络相关

Action	Action 描述	资源描述
elasticsearch:UpdatePubl icNetwork	是否开启公网访问	instances/\$instanceId
elasticsearch:UpdatePubl icIps	修改公网白名单	instances/\$instanceId
elasticsearch:UpdateWhit eIps	修改VPC私网白名单	instances/\$instanceId
elasticsearch:UpdateKiba naIps	修改kibana白名单	instances/\$instanceId

・词典

Action	Action 描述	资源描述
elasticsearch:UpdateDict	修改词典(IK、同义词)	instances/\$instanceId

底层云监控授权列表



说明:

以下资源描述为简写的`*`通配符形式。

Action	Action 描述	资源描述
cms:ListProductOfActiveA lert	获取用户云监控开通的产品	*

Action	Action 描述	资源描述
cms:ListAlarm	查询指定或全部报警规则设置	*
cms:QueryMetricList	查询一段时间内指定产品实例 的监控数据	*

购买页VPC和Vswitch授权列表

▋ 说明:

以下资源描述为简写形式,已省略相同部分内容,完整资源描述参见文档顶部。

Action	Action 描述	资源描述
DescribeVpcs	获取vpc列表	vpc/*
DescribeVswitches	获取vswitch列表	vswitch/*

智能运维授权列表

送明:

以下资源描述为简写形式,已省略相同部分内容,完整资源描述参见文档顶部。

Action	Action 描述	资源描述
elasticsearch:OpenDiagno sis	开启智能诊断	instances/* 或 instances /\$instanceId
elasticsearch:CloseDiagn osis	关闭智能诊断	instances/* 或 instances /\$instanceId
elasticsearch:UpdateDiag nosisSettings	更新诊断配置	instances/* 或 instances /\$instanceId
elasticsearch:DescribeDi agnosisSettings	获取诊断配置	instances/* 或 instances /\$instanceId
elasticsearch:ListInstan ceIndices	获取实例索引	instances/* 或 instances /\$instanceId
elasticsearch:DiagnoseIn stance	开始智能诊断	instances/* 或 instances /\$instanceId
elasticsearch:ListDiagno seReportIds	获取诊断报告标号列表	instances/* 或 instances /\$instanceId
elasticsearch:DescribeDi agnoseReport	获取诊断报告详情	instances/* 或 instances /\$instanceId

Action	Action 描述	资源描述
elasticsearch:ListDiagno	获取诊断报告详情列表	instances/*或instances
seReport		/\$instanceId

支持区域

Elasticsearch 区域	RegionId
华东1(杭州)	cn-hangzhou
华北2(北京)	cn-beijing
华东2(上海)	cn-shanghai
华南1(深圳)	cn-shenzhen
印度(孟买)	ap-south-1
新加坡	ap-southeast-1
香港	cn-hongkong
美国(硅谷)	us-west-1
马来西亚(吉隆坡)	ap-southeast-3
德国(法兰克福)	eu-central-1
日本(东京)	ap-northeast-1
澳大利亚(悉尼)	ap-southeast-2
印度尼西亚(雅加达)	ap-southeast-5
华北1(青岛)	cn-qingdao
华北3(张家口)	cn-zhangjiakou

5.2 访问鉴权规则

通用权限策略

我们有为常见访问需求的用户提供以下2种通用权限策略,可选择适合自己的权限。可直接从"可 选授权策略名称"中搜以下括号中策略名进行选择。

- ·只读访问Elasticsearch的权限,可用于只读用户(AliyunElasticsearchReadOnlyAccess)
- · 管理Elasticsearch的权限,可用于管理员(AliyunElasticsearchFullAccess)

🗐 说明:

如果以上2种通用权限策略无法满足您的需求,可参考下面的描述自定义符合您需求的权限策略。

购买实例权限(后付费&预付费)

访问主用户VPC权限

• ["vpc:DescribeVSwitch*" , "vpc:DescribeVpc*"]

送 说明:

可参考系统模板 AliyunVPCReadOnlyAccess

子用户订单权限

· ["bss:PayOrder"]



可参考系统模板 AliyunBSSOrderAccess

API 对应权限

Method	URI	Resource	Action
GET	/instances	instances/*	ListInstance
POST	/instances	instances/*	CreateInstance
GET	/instances/\$ instanceId	instances/\$ instanceId	DescribeInstance
DELETE	/instances/\$ instanceId	instances/\$ instanceId	DeleteInstance
POST	/instances/\$ instanceId/actions/ restart	instances/\$ instanceId	RestartInstance
PUT	/instances/\$ instanceId	instances/\$ instanceId	UpdateInstance

授权样例参考

- · #unique_59 (例如: \$regionid, \$accountid, \$instanceId 等)。
- · Resource 中的es实例也可以使用通配符 * 来表示。

授权样例(1)

给accountId为1234的主账号下的某个子账号,赋予华东1区所有实例在控制台中拥有(除了创建 实例权限)之外的所有操作权限,同时限制只能指定的IP才能访问。

该策略在主账号控制台中创建后,需再通过主账号在 RAM 控制台中对子账号授权,或通过 RAM SDK对子账号授权。

1. 创建一个策略

```
{
  "Statement": [
    {
      "Action": [
        "elasticsearch:ListInstance",
        "elasticsearch:DescribeInstance",
        "elasticsearch:DeleteInstance",
        "elasticsearch:RestartInstance",
        "elasticsearch:UpdateInstance"
      ],
"Condition": {
        "IpAddress": {
          "acs:SourceIp": "xxx.xx.xx.x/xx"
        }
      "Éffect": "Allow",
      "Resource": "acs:elasticsearch:cn-hangzhou:1234:instances/*"
    }
  "Version": "1"
}
```

2. 把当前策略授权给您指定的子账号。

授权样例(2)

给accountId为1234的主账号下的某个子账号,赋予华东1区指定实例在控制台中拥有(除了创建 实例权限)之外的所有操作权限,同时限制只能指定的IP才能访问。

该策略在主账号控制台中创建后,需再通过主账号在 RAM 控制台中对子账号授权,或通过 RAM SDK对子账号授权。

1. 创建一个策略

```
{
  "Statement": [
    {
      "Action": [
        "elasticsearch:ListInstance"
      ],
"Condition": {
        "IpAddress": {
          "acs:SourceIp": "xxx.xx.xx.x/xx"
        }
      },
      "Éffect": "Allow",
      "Resource": "acs:elasticsearch:cn-hangzhou:1234:instances/*"
    },
    {
      "Action": [
        "elasticsearch:DescribeInstance",
        "elasticsearch:DeleteInstance"
        "elasticsearch:RestartInstance",
        "elasticsearch:UpdateInstance"
      "Condition": {
        "IpAddress": {
```

```
"acs:SourceIp": "xxx.xx.x/xx"
        }
        },
        "Effect": "Allow",
        "Resource": "acs:elasticsearch:cn-hangzhou:1234:instances/$
instanceId"
        }
    ],
    "Version": "1"
}
```

2. 把当前策略授权给您指定的子账号。

授权样例(3)

给accountId为1234的主账号下的某个子账号,赋予所有阿里云Elasticsearch支持区域,所有实例在控制台中拥有所有操作权限。

该策略在主账号控制台中创建后,需再通过主账号在 RAM 控制台中对子账号授权,或通过 RAM SDK对子账号授权。

1. 创建一个策略

```
{
   "Statement": [
        {
            "Action": [
               "elasticsearch:*"
              ],
            "Effect": "Allow",
            "Resource": "acs:elasticsearch:*:1234:instances/*"
        }
   ],
   "Version": "1"
}
```

2. 把当前策略授权给您指定的子账号。

授权样例(4)

给accountId为1234的主账号下的某个子账号,赋予所有阿里云Elasticsearch支持区域,指定实例在控制台中拥有(除了创建实例权限,查看实例列表权限)之外的所有操作权限。

该策略在主账号控制台中创建后,需再通过主账号在 RAM 控制台中对子账号授权,或通过 RAM SDK对子账号授权。

1. 创建一个策略

```
{
    "Statement": [
        {
          "Action": [
            "elasticsearch:DescribeInstance",
            "elasticsearch:DeleteInstance",
            "elasticsearch:UpdateInstance",
            "elasticsearch:RestartInstance"
        ],
```

```
"Effect": "Allow",
    "Resource": "acs:elasticsearch:*:1234:instances/$instanceId"
    }
],
"Version": "1"
}
```

2. 把当前策略授权给您指定的子账号。

5.3 临时访问令牌

本文档为您介绍使用临时访问令牌STS(Security Token Service,它是RAM的一个扩展授权服务)给临时用户授权的方法。

前提条件

进行操作前,请确保您已经注册了阿里云账号。如还未注册,请先完成账号注册。

账号注册成功后,请使用该账号登录RAM控制台。

背景信息

有时存在一些用户(人或应用程序),并不经常访问您的云资源,只是偶尔需要访问一次,称这些 用户为临时用户。您可以通过STS来为这些用户(子账号)颁发访问令牌。颁发令牌时,您可以根 据需要来定义令牌的权限和自动过期时间。

使用STS访问令牌,给临时用户授权的好处是让授权更加可控。您不必为临时用户创建一个RAM用 户账号及密钥,因为RAM用户密钥都是长期有效的,但临时用户并不需要长期的资源访问。具体使 用场景可参见#unique_61和#unique_62。

创建可信实体为阿里云账号的RAM角色

- 1. 在左侧导航栏,单击RAM角色管理。
- 2. 单击新建RAM角色,选择可信实体类型为阿里云账号,单击下一步。
- 3. 输入角色名称和备注。
- 4. 选择云账号为当前云账号或其他云账号。

📃 说明:

若选择其他云账号,需要填写其他云账号的ID。

5. 单击完成。

临时访问授权

在使用STS来授权访问前,需要先对上文创建RAM角色第三步中的受信云账号的子账号进行扮演角 色的授权。若任意子账号都可以扮演这些角色,会带来不可预估的风险,因此有扮演对应角色需求 的子账号需要被赋予相应的配置权限,具体操作方式如下。

- 1. 创建自定义策略。
- 2. 为RAM用户授权。
- 3. 使用RAM角色。

创建自定义策略

- 1. 在左侧导航栏的权限管理菜单下,单击权限策略管理。
- 2. 单击新建权限策略。
- 3. 填写策略名称和备注。
- 4. 配置模式选择脚本配置。
- 5. 在策略内容脚本编辑区域输入如下脚本内容。

```
{
"Version": "1",
"Statement": [
{
    "Effect": "Allow",
    "Action": "sts:AssumeRole",
    "Resource": "acs:ram::${aliyunID}:role/${roleName}"
}
]
}
```

📕 说明:

Resource的具体内容,可以从RAM角色管理 > 基本信息页面的ARN中获取。

RAM访问控制 / RAM角色管理 / Aliyur					
← Aliyun					
基本信息					
RAM角色名称	Aliyur	创建时间	2019年6月21日 15:04:19		
备注	云盾DDoS高防默认使用此角色来访问您在其他云产品中的资源	ARN	acs:ram:: :role/aliyun		

- · \${aliyunID}:表示创建角色的用户ID。
- · \${roleName}:表示角色名称的小写。
- 6. 单击确认。



自定义策略创建成功后,您还需要在用户管理页面,对子账号授权上文中已经创建的扮演角色的权限,详情请参见下文的为RAM用户授权。

为RAM用户授权

- 1. 在左侧导航栏的人员管理菜单下,单击用户。
- 2. 在用户登录名称/显示名称列表下,找到目标RAM用户。
- 3. 单击添加权限, 被授权主体会自动填入。
- 4. 在左侧权限策略名称列表下,单击需要授予RAM用户的权限策略。



在右侧区域框,选择某条策略并单击×,可撤销该策略。

- 5. 单击确定。
- 6. 单击完成。

使用RAM角色

为了安全起见,阿里云不允许受信云账号以自己的身份扮演角色,如果一个实体用户想扮演某个 RAM角色,该实体用户必须先以自己身份登录,然后将自己从实体身份切换到RAM角色身份。

子账号登录控制台后,可以切换到刚刚被授予扮演的角色,行使角色的权限,步骤如下。

- 1. 将鼠标悬停在右上角头像的位置,单击切换身份。
- 2. 在角色切换页面,输入相应企业别名/默认域名以及角色名。
- 3. 单击切换。

📃 说明:

切换成功后,用户将以RAM角色身份登录控制台,控制台右上角头像位置将显示角色身份(即 当前身份)和登录身份,此时用户只能执行该角色身份被授权的所有操作。

6 ElasticFlow

6.1 数据源概述

ElasticFlow通过读取数据源中的信息来拉取对应数据,再进行数据处理发送至下游。数据源与数 据处理任务的配置是分开的,即可以先配置好数据源,再在配置数据处理任务时指定之前配置好的 数据源。

数据源类型

ElasticFlow支持以下四种类型的阿里云数据源,部分类型的数据源需要做一次访问授权,在创建数据源时可以根据控制台提示引导完成:

• RDS for MySQL

可以将阿里云上的RDS实例ID、数据库名以及具备读权限的账号密码配置到ElasticFlow 中,作为一个数据源。再在创建数据同步任务时指定该数据源中需要拉取数据的表名,完成以上 步骤并启动任务后将从配置的数据源表中拉取数据。

新建MySQL数据源		×
* 数据源名称:	请输入自定义名称	
数据源描述:	请输入数据源描述信息	
* RDS 实例ID:	请输入实例ID	
* 数据库名称:	请输入数据库名称	
* 用户名:	请输入数据库用户名	
* 密码:	请输入数据库密码	日本
授权及连通性测试.	授权并测试连通性	
	确保数据库可以被网络访问 确保数据库没有被防火墙禁止 确保数据库过经启动 暂不支持MYSQL 5.7版本 确保实例的网络类型为专有网络	
	确认 上一步	取消

• MaxCompute

可以将阿里云上的MaxCompute项目名以及具备读权限的access key/secret配置到 ElasticFlow中,作为一个数据源。再在创建数据同步任务时指定该数据源中需要拉取数据的表 名,完成以上步骤并启动任务后将从配置的数据源表中拉取数据。

新建MaxCompute数据源		×
* 数据源名称:	清輸入自定义名称	
<u> </u>	項制/XX储源:(田达信忌.	
* 项目名称:	请输入MaxCompute项目名称	
* Access ID:	请输入数据库用户名	咨询
* 密码:	请输入数据库密码	建议
授权及连通性测试:	测试连通性	
	确认 上一步	取消

• LogService

可以将阿里云上的LogService项目名配置到ElasticFlow中,作为一个数据源。再在创建数据 同步任务时指定该数据源中需要拉取数据的表名,完成以上步骤并启动任务后将从配置的数据源 表中拉取数据。



不支持LogService的Shard扩容和缩容,如果分裂或者合并了某个数据同步任务正在使用的Shard,会导致数据丢失或者同步任务出错。有关Shard的具体操作,请参见操作Shard。

新建LogService数据源		×
* 数据源名称:	请输入自定义数据源名称	
数据源描述:	请输入自定义数据源描述	
* Project:	请输入Project	
测试连通性:	授权及连通性测试	
		2 谷
		间 建
		1X

确认

取消

• ElasticSearch

可以将阿里云上的ElasticSearch实例ID以及具备读权限的access key/secret配置到 ElasticFlow中,作为一个数据源。再在创建数据同步任务时指定该数据源中需要拉取数据的表 名,完成以上步骤并启动任务后将从配置的数据源表中拉取数据。

新建Elasticsearch数据源		×
* 数据源名称:	请输入自定义名称	
数据源描述:	请输入数据源描述信息	
* 实例ID:	请输入elasticsearch实例ID	
* 用户名:	请输入数据库用户名	となって
* 密码:	请输入数据库密码	吗 · 建
授权及连通性测试:	测试连通性	
\triangle	目标elasticsearch集群elasicsearch.yml中要事先配置 reindex.remote.whitelist,并保证生效(配置后重启)	
	确认 上一步	取消

数据源管理

在数据源管理页面,您可以创建、搜索、编辑、删除数据源。有关创建数据源,详情请参见创建数 据源(跨云服务授权)。

阿里云Elasticse	数据源列表							
实例列表	创建刷新				数据源ID	∨ 请输入搜索内容		Q
项目列表	数据源ID	数据源名称	数据源类型	链接信息			数据源描述	操作
数据源列表	HeCHe4w-PNtm_27-T0	elasticflow_rds_jcj	MySQL	数据库名称: elasticflow_jcj RDS 实例ID: rm-uf664g9p19(m);gg 用户名: jmg			elasticflow_rds_jcj	编辑 删除

6.2 加工函数及参数说明

6.2.1 加工函数概述

相关加工函数及对应功能概述列表如下:

函数名	功能概述	参考文档
CONCAT	连接两个或多个字符串值组成 一个新的字符串。	CONCAT
TO_TIMESTAMP	将VARCHAR类型的日期转换 成TIMESTAMP类型。	TO_TIMESTAMP
REPLACE	字符串替换函数。	REPLACE
CURRENT_TIMESTAMP	返回当前UTC时间(GMT+0)的 时间戳,小于北京时间8小时。	CURRENT_TIMESTAMP
SUBSTRING	获取字符串子串。例如从位置 start开始截取长度为len的子 串。	SUBSTRING
SPLIT_INDEX	以sep作为分隔符,将字符串 str分隔成若干段,取其中的第 index段。	SPLIT_INDEX
PARSE_URL	解析url,获取partToExtract 的值。例如partToExtract=' QUERY',获取url参数key的 值。	PARSE_URL
REGEXP_REPLACE	正则匹配替换。用字符串 replacement替换字符串str中 正则模式为pattern的子串,返 回新的字符串。	REGEXP_REPLACE
REGEXP_EXTRACT	使用正则模式pattern匹配抽 取字符串str中的第index个子 串。	REGEXP_EXTRACT

6.2.2 CONCAT

本文为您介绍如何使用实时计算字符串函数CONCAT。

功能描述

连接两个或多个字符串值从而组成一个新的字符串。当参数为NULL时,则跳过该参数。

语法

VARCHAR CONCAT(VARCHAR var1, VARCHAR var2, ...)

输入参数

参数	数据类型	说明
----	------	----

var1	VARCHAR	普通字符串值
var2	VARCHAR	普通字符串值

示例

测试数据

var1(VARCHAR)	var2(VARCHAR)	var3(VARCHAR)
Hello	Му	World
Hello	null	World
null	null	World
null	null	null

测试语句

CONCAT(var1, var2, var3)

测试结果

var(VARCHAR)	
HelloMyWorld	
HelloWorld	
World	
N/A	

6.2.3 CURRENT_TIMESTAMP

本文为您介绍如何使用日期函数CURRENT_TIMESTAMP。

功能描述

返回当前UTC时间(GMT+0)的时间戳,小于北京时间8小时。

语法

TIMESTAMP CURRENT_TIMESTAMP

示例

测试语句

CURRENT_TIMESTAMP

测试结果

result(TIMESTAMP)

10:02.0

6.2.4 PARSE_URL

本文为您介绍如何使用字符串函数PARSE_URL。

功能描述

解析url,获取partToExtract的值。如partToExtract='QUERY',获取url参数key的

值。partToExtract可

取HOST、PATH、QUERY、REF、PROTOCOL、FILE、AUTHORITY、USERINFO。

📋 说明:

参数为null,则返回null。

语法

```
VARCHAR PARSE_URL(VARCHAR urlStr, VARCHAR partToExtract [, VARCHAR key
])
```

输入参数

参数	数据类型	说明
urlStr	VARCHAR	url的字符串。
partToExtract	VARCHAR	解析后获取的值。
key	VARCHAR	参数名。

示例

测试数据

url1(VARCHAR)	nullstr(VARCHAR)
http://facebook.com/path/p1.php?query =1	null

测试语句

```
PARSE_URL(url1, 'QUERY', 'query') // var1,
PARSE_URL(url1, 'QUERY') // var2,
```

```
PARSE_URL(url1, 'HOST') // var3,
PARSE_URL(url1, 'PATH') // var4,
PARSE_URL(url1, 'REF') // var5,
PARSE_URL(url1, 'PROTOCOL') // var6,
PARSE_URL(url1, 'FILE') // var7,
PARSE_URL(url1, 'AUTHORITY') // var8,
PARSE_URL(nullstr, 'QUERY') // var9,
PARSE_URL(url1, 'USERINFO') // var10,
PARSE_URL(nullstr, 'QUERY', 'query') // var11
```

测试结果

var1(var2(var3(var4(var5(var6(var7(var8(var9(var10(var11(
VARCH	AFARCH	AFARCH	AFARCH	AFARCH	AFARCH	AFARCH	AFARCH	AFARCH	AFARCH	AFARCH	AR
)))))))))))	
1	query =1	faceboo .com	/path /p1. php	null	http	/path /p1. php? query =1	faceboo .com	otrull	null	null	

6.2.5 REGEXP_EXTRACT

本文为您介绍如何使用字符串函数REGEXP_EXTRACT。

功能描述

使用正则模式pattern匹配抽取字符串str中的第index个子串,index从1开始,正则匹配提取。参 数为null或者正则不合法返回null。

语法

```
VARCHAR REGEXP_EXTRACT(VARCHAR str, VARCHAR pattern, INT index)
```

输入参数

参数	数据类型	说明
str	VARCHAR	指定的字符串。
pattern	VARCHAR	匹配的字符串。
index	INT	第几个被匹配的字符串。



正则常量请按照Java代码来写。codegen会将常量字符串自动转化成Java代码。如果要描述一个 数字(\d),需要写成 '\d',也就是像在Java中写正则一样。

示例

测试数据

str1 (VARCHAR)	pattern1(VARCHAR)	index1 (INT)
foothebar	foo(.*?)(bar)	2
100-200	(\d+)-(\d+)	1
null	foo(.*?)(bar)	2
foothebar	null	2
foothebar	空	2
foothebar	(2

测试语句

REGEXP_EXTRACT(str1, pattern1, index1)

测试结果

result(VARCHAR)
bar
100
null
null
null
null

6.2.6 REGEXP_REPLACE

本文为您介绍如何使用字符串函数REGEXP_REPLACE。

功能描述

用字符串replacement替换字符串str中正则模式为pattern的子串,返回新的字符串。正则匹配 替换。



参数为null或者正则不合法返回null。

语法

VARCHAR REGEXP_REPLACE(VARCHAR str, VARCHAR pattern, VARCHAR replacement)

输入参数

参数	数据类型	说明
str	VARCHAR	指定的字符串。
pattern	VARCHAR	被替换的字符串。
replacement	VARCHAR	用于替换的字符串。

请您按照Java代码来写正则常量。codegen会自动将SQL常量字符串转化为Java代码。如果要描述一个字符串(\d),需要写成 '\d',同在Java中书写正则一样。

示例

测试数据

str1(VARCHAR)	pattern1(VARCHAR)	replace1(VARCHAR)
2014/3/13	-	空
null	-	空
2014/3/13	-	null
2014/3/13	空	s
2014/3/13	(s
100-200	(\ d +)	num

测试语句

REGEXP_REPLACE(str1, pattern1, replace1)

result(VARCHAR)	
20140313	
null	
null	

2014/3/13	
null	
num-num	

6.2.7 REPLACE

本文为您介绍如何使用字符串函数REPLACE。

功能描述

字符串替换函数。

语法

VARCHAR REPLACE(str1, str2, str3)

输入参数

参数	数据类型	说明
str1	VARCHAR	原字符
str2	VARCHAR	目标字符
str3	VARCHAR	替换字符

示例

测试数据

str1(INT)	str2(INT)	str3(INT)
alibaba es	es	eflow

测试语句

REPLACE(str1, str2, str3)

result(VARCHAR)	
alibaba eflow	

6.2.8 SPLIT_INDEX

本文为您介绍如何使用字符串函数SPLIT_INDEX。

功能描述

以sep作为分隔符,将字符串str分隔成若干段,取其中的第index段。index从0开始,取不到字段,则返回null。

任一参数为NULL,则返回null。

语法

VARCHAR SPLIT_INDEX(VARCHAR str, VARCHAR sep, INT index)

输入参数

参数	数据类型	说明
str	VARCHAR	被分隔的字符串。
sep	VARCHAR	分隔符的字符串。
index	INT	截取的字段位置。

示例

测试数据

str(VARCHAR)	sep(VARCHAR)	index(INT)
Jack,John,Mary	,	2
Jack,John,Mary	,	3
Jack,John,Mary	null	0
null	,	0

测试语句

SPLIT_INDEX(str, sep, index)

Value(VARCHAR)	
Mary	
null	

null	
null	

6.2.9 SUBSTRING

本文为您介绍如何使用字符串函数SUBSTRING。

功能描述

获取字符串子串。截取从位置start开始,长度为len的子串。若未指定len则截取到字符串结尾。 start从1开始,start为0当1看待,为负数时表示从字符串末尾倒序计算位置。

语法

```
VARCHAR SUBSTRING(VARCHAR a, INT start)
VARCHAR SUBSTRING(VARCHAR a, INT start, INT len)
```

输入参数

参数	数据类型	说明
a	VARCHAR	指定的字符串。
start	INT	在字符串a中开始截取的位置。
len	INT	类截取的长度。

示例

测试数据

str(VARCHAR)	nullstr(VARCHAR)
k1=v1;k2=v2	null

测试语句

sql SUBSTRING('', 22222222) // var1, SUBSTRING(str, 2) // var2, SUBSTRING(str, -2) // var3, SUBSTRING(str, -2, 1) // var4, SUBSTRING (str, 2, 1) // var5, SUBSTRING(str, 22) // var6, SUBSTRING(str, -22) // var7, SUBSTRING(str, 1) // var8, SUBSTRING(str, 0) // var9, SUBSTRING(nullstr, 0) // var10

var1(var2(var3(var4(var5(var6(var7(var8(var9(var10(
VARCHA	RARCHA	A R								
))))))))))	

空	1=v1;	v2	v	1	空	空	k1=v1;	k1=v1;	null
	k2=v2						k2=v2	k2=v2	

6.2.10 TO_TIMESTAMP

本文为您介绍如何使用日期函数TO_TIMESTAMP。

功能描述

将VARCHAR类型的日期转换成TIMESTAMP类型。

语法

TIMESTAMP TO_TIMESTAMP(VARCHAR date)
TIMESTAMP TO_TIMESTAMP(VARCHAR date, VARCHAR format)

输入参数

参数	数据类型	说明
date	VARCHAR	
format	VARCHAR	默认格式为yyyy-MM-dd HH: mm:ss[.SSS]。

示例

测试数据

timestamp1(VARCHAR)	timestamp2(VARCHAR)	
2017/9/15 0:00	20170915000000	

测试用例

<pre>TO_TIMESTAMP(timestamp1)</pre>	// result1
<pre>TO_TIMESTAMP(timestamp2,</pre>	<pre>'yyyyMMddHHmmss') // result2</pre>

测试结果

result1(TIMESTAMP)	result2(TIMESTAMP)
2017-09-15 00:00:00.0	2017-09-15 00:00:00.0

6.3 算子

6.3.1 算子类型与逻辑表

本章节为您介绍与算子相关的基本概念。

算子

任务由多个算子构成,各算子通过逻辑表串联,算子类型包含数据导入算子、数据加工算子、数据 目标算子三种。

| ■ 说明:

使用算子与逻辑表时请注意以下几点:

· 数据导入算子、目标算子只能有一个。

· 流计算任务加工算子可以是多个,索引迁移任务不支持数据加工处理。

· 产出的所有逻辑表都必须使用到。

数据导入算子

配置拉取数据的数据源,产出是逻辑表,详情请参见数据导入算子。

数据加工算子(非必须)

对逻辑表进行加工处理,例如对逻辑表进行数据过滤、对逻辑表中的某些字段通过加工函数进行转换,产出是逻辑表,详情请参见数据加工算子。

数据目标算子

设置导入到目标Elasticsearch实例中的逻辑表,详情请参见数据目标算子。

逻辑表

逻辑表由字段构成,是一个虚拟的概念,类似数据库视图。

6.3.2 数据导入算子

数据导入算子用于配置拉取数据的数据源,产出为逻辑表。

任务配置				创建任务配置	rg lawd
 完整的任务配置流程需要经历 分别对这三步进行添加和编辑。 	1. 导入数据 -> 2. 数据加工(添加数据加工算子) -> 3. 详细说明	数据目标(添加目标算子)三个步骤,选择创建任	务配置,您可以按步骤整体创建	任务。创建之后,您也可以	
导入数据 数据加工	数据目标				2 咨询
添加数据导入算子 刷新			请输入算子名称		· 建 Q议
数据导入算子名称	全量数据源类型	增量数据源类型	产出逻辑表	操作	
data_import_test	RDS (elasticflow_rds_jcj)		data_import_table	编辑删除	

支持数据源

数据导入算子支持以下数据源:

- RDS for MySQL
- · MaxCompute
- · LogService
- · Elasticsearch

尻明	Ë)	说明
----	----	----

- ・需要为产出的逻辑表设置主键。
- · RDS for MySQL需要在VPC网络下,数据库版本需要为5.6,暂时不支持5.7。
- · MaxCompute必须有分区,设置数据导入算子时需填写全部分区信息,多个分区间以英文逗号 分割。
- · LogService需要用户自己手动添加逻辑表字段,并为字段设置类型。
- ·Elasticsearch只需要设置产出逻辑表名称,不需要设置逻辑表结构。

致 据导入奀型

扁辑数据导入算子		
* 导入算子名称:	data_import_test	8
* 选择数据源:	elasticflow_rds_jcj	→ 新増数据源
	✔ 全量导入数据 ✔ 増量导入数据	
* 选择表:	elasticflaow_jcj	选择数据表
* 产出逻辑表表名:	data_import_table	\odot

数据导入算子支持以下两种配置:

・全量

从数据源中将当前时刻已有的数据导入到目标Elasticsearch中,导入完成即任务完成。

・増量

从任务运行起,一直监控数据源中的数据信息变化,同步到目标Elasticsearch。任务不会停止,除非在控制台手动停止任务。

全量与增量的组合

在配置导入算子时,根据数据源类型可选组合如下:

全量	增量
MySQL	
	MySQL
MySQL	MySQL
MaxCompute	
MaxCompute	LogService
	LogService
Elasticsearch	

6.3.3 数据过滤

数据过滤是根据过滤条件从逻辑表中筛选出符合条件的数据。

过滤条件

过滤条件格式支持类SQL WHERE的booleanExpression表达式。连接符支持And和OR。操作 符支持类型如下:

操作符	描述
=	等于
<>	不等于
>	大于
>=	大于等于
<	小于
<=	小于等于

■ 说明:

- · 如果有数值型变量比较,需要显式的指定常数的类型。比如 x<= 5,其中如果x是浮点类型,需 要显式指定常数为浮点类型: x<=5.0
- ·如果有数值类型比较,操作符的左右不能都是变量。

举例

当筛选条件为City='Beijing'并且逻辑表数据如下所示时:

Address	City
Oxford Street	Beijing
Fifth Avenue	Beijing

Changan Street	shanghai
8	8

过滤出符合条件的数据如下图所示:

Address	City
Oxford Street	Beijing
Fifth Avenue	Beijing

6.3.4 数据加工算子

本章节为您介绍如何对数据导入算子产出的逻辑表进行加工处理,例如对逻辑表过滤数据、对逻辑 表中的某些字段用加工函数做转换,产出为逻辑表。此数据加工算子非必须,根据实际需要进行创 建

任务详情

任务详情主要展示任务ID、任务创建时间、任务名称(支持修改)、以及任务当前状态等参数。

pl-itp7663o1yc01w67c

任务详情			
任务ID	pl-itp7663o1yc01w67c	创建时间	2019-03-26 11:46:57
名称	total_quantity_test 编辑	状态	● 执行完成

任务配置

任务配置主要用于添加、编辑和删除数据加工算子操作,也支持根据算子名称进行搜索。

任务配置					创建任务	配置
(1) 完整的任务配置流程需要经历 1. 导入: 分别对这三步进行添加和编辑。详细说	数据 -> 2. 数据加工(添加数据加工算子) -> 词	3. 数据目标 (添加目标算子) 三个步骤 ,	选择创建任务配置,	您可以按步骤整体创建任	务。创建之后,您也可以	L
导入数据 数据加工 第	数据目标					
添加数据加工算子 刷新				请输入算子名称		Q
数据加工算子名称	数据加工类型	引用逻辑表	产出逻辑表	ł	操作	咨询
		没有数据				建议

添加数据加工算子

添加数据加工算子用于创建数据加工算子操作流程,详情请参见创建数据加工算子。

编辑数据加工算子

单击数据加工算子操作列的编辑可对数据加工算子进行编辑,该操作与创建数据加工算子流程类 似,详情请参见创建数据加工算子。

添加数据加工算子 刷新				请输入算子名称	Q	
数据加工算子名称	数据加工类型	引用逻辑表	产出逻辑表	操作		咨询
data_process_test	数据加工	data_import_table <	data_process_proc	luce 编辑 删除		· 建议

删除数据加工算子

- 1. 选择要删除的数据加工算子。
- 2. 单击删除。

添加数据加工算子 刷新				请输入算子名称	۹ 🕒
数据加工算子名称	数据加工类型	引用逻辑表	产出逻辑表	操作	咨询
data_process_test	数据加工	data_import_table ()	data_process_proc	duce 编辑 删除	建议

3. 在操作提示提示框中,单击确认。

?	操作提示	×
	是否要删除当前算子:pl-conf-9ct7q0n9ue3	50delr
	确认	取消

6.3.5 数据目标算子

数据目标算子用于设置导入到目标Elasticsearch实例中的逻辑表。目标数据源仅支持Elasticsearch。任务类型支持流计算资源和索引迁移两类。

添加流计算数据目标算子时,需要输入以下参数:

编辑数据目标算子			×
* 目标算子名称:	target_operator_test	8	
* 数据目标类型:	elasticsearch	V	
* 实例ID:	流计具资源任务类型不支持5.5.3版本的E	·S目标	
* 用户名:	elastic	8	
* 密码:		8	
* 引用逻辑表:	data_process_produce1	\vee	
* 目标索引:	 选择现有索引 自动生成索引 全量会导致现有索引的数据被覆盖,并且 	短时间不可见	」
* 索引名称:	my_index_new 如果索引来自5.X版本的ES并且有多个ty	↓ pe,则不支持6.X版本的ES作为目标	
* type名称:	default	\vee	

・目标算子名称

为数据目标算子命名。

· 数据目标类型

目标数据源仅支持Elasticsearch。

・实例ID

阿里云Elasticsearch产品对应目标实例 ID。

・用户名

访问阿里云Elasticsearch实例用户名(要有对应索引读写权限,默认用户名为elastic)。

・密码

访问阿里云Elasticsearch实例的用户对应密码。

・引用逻辑表

逻辑表为数据导入算子任务或数据加工算子任务产出的逻辑表。

・目标索引

支持现有索引和自动生成索引两种方式。

・索引名称

可选择目标阿里云Elasticsearch实例中现有索引或自动生成索引。

- 现有索引

选择当前索引已存在的type,继续使用已选择的索引的mapping和setting配置(不支持修改)。

- 自动生成索引

根据用户选择逻辑表生成新索引的mapping配置(支持修改),新索引的setting配置默认 使用系统默认配置(支持修改)。



6.4 快速开始

6.4.1 创建数据源(跨云服务授权)

在数据源管理页面您可以创建数据源,并且对创建的数据源进行管理。

创建数据源

- 1. 进入 ElasticFlow 控制台。
- 2. 单击左侧数据源列表标签页。
- 3. 在数据源列表管理页面,单击创建。
- 4. 选择对应数据源类型(例如MySQL),单击下一步。

5. 在新建数据源页面,输入相关参数并单击确认,验证配置信息,通过后完成创建。

新建MySQL数据源		\times
		1
*	· · · · · · · · · · · · · · · · · · ·	
数据源描述:	请输入数据源描述信息	
* RDS 实例ID:	请输入实例ID	
* 数据库名称:	请输入数据库名称	
* 用户名:	请输入数据库用户名	
* 密码:	请输入数据库密码	 咨 词
授权及连通性测试:	授权并测试连通性	建议
	确保数据库可以被网络访问 确保数据库没有被防火墙禁止 确保数据库域名能够被解析 确保数据库已经启动 暂不支持MYSQL 5.7版本 确保实例的网络类型为专有网络	
	确认 上 一步	取消

管理数据源

数据源创建完成后,您可以在数据源列表页面对数据源进行以下操作:

阿里云Elasticse	数据源列表							
实例列表	创建刷新				数据源ID ∖	/ 请输入搜索内容		Q
项目列表	数据源ID	数据源名称	数据源类型	链接信息			数据源描述	操作
数据源列表	HeCHe4w-PNtm_27-TD	elasticflow_rds_jcj	MySQL	数据库名称: elasticflow_jcj RDS 实例ID: m-uf664g9p19());g 用户名: mag	,		elasticflow_rds_jcj	编辑 删除

- · 在搜索框中输入数据源ID或数据源名称搜索对应数据源。
- ・创建数据源。
- ·编辑数据源(对应数据源不能被任务调用,否则需要先去掉调用再编辑)。
- ·删除数据源(对应数据源不能被任务调用,否则需要先去掉调用再删除)。

创建取消

6.4.2 创建项目和任务

数据源创建完成后,您可以创建项目和任务。

创建项目

- 1. 在ElasticFlow控制台,单击项目列表。
- 2. 在项目列表管理页面,单击创建。
- 在购买项目页面,选择对应区域和资源类型(例如流计算资源),并为对应项目后续将要创建的 增量或全量任务分配适当的处理资源,详情请参见#unique_87。

购买项目					返回项目列表
区域	区域	4/52		当前配置 区域: 华东2 资源类型: 流计算资源 项目配额: 10 CU ¥ 元小时:	
资源类型	资源类型	流计算资源 素引迁移			
项自配额	项目配额	10CU 	100CU		-
					· · · · · · · · · · · · · · · · · · ·

4. 单击创建,系统自动跳转至项目列表管理页面。

阿里云Elasticse	创建剧新				
实例列表	项目名称	资源类型	项目配额	创建时间	撮作
 ElasticFlow 	eh-cn-i29h6851085ppsixa jcj_test	流计算资源	100CU	2 天前	任务列表 项目配置 释放
项目列表	eh.cn_61vi0033v565				
数据源列表	暂无	流计算资源	10CU	3 天前	任务列表 项目配置 释放
	eh-cn-ie7ha0x74t031aatha RDS	流计算资源	43CU	3 天前	任务列表 项目配置 释放
	eh-cn-2896xrm8uq4	流计算资源	10CU	6 天前	任务列表 项目配置 释放 🗧 高) 词
					页 - 2 页 - 2 页 - 2 建 毎页显示 10 ∨ く上一页 1 下一页 > 建 议

创建任务

项目创建完成后,您可以在该项目中创建任务,任务详情请参见任务类型。

- 1. 在项目列表管理页面,选择要创建任务的项目,单击项目名称或任务列表。
- 2. 在任务列表管理页面,单击创建。
- 3. 在创建任务页面,输入任务名称单击确认


仕穷名称 长度为1-30个子付,以天小子母,数子或中又开头,可以包含_或-行号。								
<	eh-cn-kxb7ddm25fnp425aq						刷新	
基础信息	会15章 医1度症							
任务列表	(3)A2							
日志查询	任务名称	状态	进度	资源占用	创建时间	操作		
	pl-l2ae27te2bft25s7p test	● 待启动	0%	0 CU	2 小时前	启动 编辑 日志查询 删除		

6.4.3 创建数据导入算子

任务创建完成后,您可以为对应任务配置数据导入算子。

- 1. 在任务列表管理页面,单击任务名称或者编辑按钮。
- 2. 在任务基础信息页面,单击任务配置标签页下的添加数据导入算子。
- 3. 在添加数据导入算子页面, 输入以下参数:

<	pl-keaiqpt26zmpp85jn	添加数据导入算子		×
基础信息	Pr-KealqDtzb2mppbdjn 任务详情 任务に 任务に CS称 Test0329 編編 任务配置	 导入算子名称 选择数据度 选择表 产出逻辑表表名: 	 请输入导入算子名称 请选择 > 新提供的。 资料数据表 请输入高班产出的逻辑表表名 	
	① 完整的任务配置加度需要经历1.号人数据→2 数据加工(添加数 载,评控研明			● 咨询
	(水気)(株) (水気)(北) (×)			w.
				确认 取 消

- ・导入算子名称
- ・选择数据源

如果选择 RDS 或 MaxCompute 数据源,则需要配置数据导入算子,即需要通过全量还是 增量方式导入数据。



如果选择了RDS、MaxCompute数据源,请注意以下几点:

- MaxCompute: 除了设置对应表,还需要填写对应分区(必需)。
- LogService: 单击选择Logstore, 选择对应Logstore。

- Elasticsearch: 单击选择索引,选择对应索引。

・选择数据表

・产出逻辑表表名

如果选择了 RDS、MaxCompute、LogService数据源,可以单击侧边栏左下角增加字段,添加需要的字段到逻辑表。单击逻辑表中的主键 列设置逻辑表主键。

📋 说明:

设置逻辑表时请注意以下几点:

- Elasticsearch数据源不需设置逻辑表主键。
- 逻辑表表名称长度为1-30个字符,支持大小写字母、数字、_。
- 逻辑表表名不能以_开头。
- 不能与现有逻辑表表名重复。

4. 单击确认保存数据导入算子。

6.4.4 创建数据加工算子

数据导入算子创建完成后,您可以对数据导入算子产出的逻辑表进行加工处理,例如对逻辑表过滤 数据、对逻辑表中的某些字段用加工函数做转换,最终产出为逻辑表。此数据加工算子非必须,请 根据实际需要进行创建。

在数据加工标签中,可以添加多个数据加工算子,算子之间目前仅支持线性关系,如下图所示,一 个算子的应用逻辑表为其上游算子的产出逻辑表,多个算子被逻辑表连接,形成一个链式结构。因 此每个逻辑表都必须被引用,否则任务无法顺利执行。

任务配置					创建任务配	置
 完整的任务配置流程需要经历1. 分别对这三步进行添加和编辑。 	导入数据 -> 2. 数据加工(添加数据加工算子) -> 详细说明	3. 数据目标 (添加目标算子) 三个步骤 ,	选择创建任务配置,	您可以按步骤整体创建任务。仓	1建之后,您也可以	
导入数据 数据加工	数据目标					
添加数据加工算子 刷新				请输入算子名称		Q
数据加工算子名称	数据加工类型	引用逻辑表	产出逻辑表	操作		
data_process_test2	数据加工	data_process_produce1 ()	data_process_prod	luce2 编辑 册	删除	6
data_process_test3	数据加工	data_process_produce2 ()	data_process_prod	luce3 编辑 ff	删除	ì
data_process_test1	数据加工	data_import_table <	data_process_prod	uce1 编辑 f	删除	

1. 在任务列表管理页面,单击任务名称或者编辑按钮。

2. 在任务基础信息页面,单击任务配置下的数据加工标签页。

任务配置					创建任务配	置
(1) 完整的任务配置流程需要经历 1. 导入 分别对这三步进行添加和编辑。详细	数据 -> 2. 数据加工(添加数据加工算子) -> 兑明	> 3. 数据目标(添加目标算子)三个步骤	,选择创建任务配置,	您可以按步骤整体创建任务	5。创建之后,您也可以	
导入数据 数据加工	数据目标					
添加数据加工算子 刷新				请输入算子名称		Q
数据加工算子名称	数据加工类型	引用逻辑表	产出逻辑表	操	ſſĘ	咨询
		没有数据				建议

3. 单击添加数据加工算子。

4. 在添加数据加工算子页面, 输入以下参数:

添	加数据加工算子	2					×
	* カロユ	[算子名称:	data_process_test		8		
	* カロコ	[算子类型:	数据加工		\sim		
	* ਤੋ	用逻辑表:	data_import_table		~ (0)		
	* 7	产出逻辑表:	data_process_prod	uce	8		
		I	✔ 添加数据过滤条件				
			id>1		۲		
	产出表字段	主键	类型	引用表字段	加工函数	操作	咨询
	id	•	INT	id	_	数据加工配置 删除	建议
	name		VARCHAR	name		数据加工配置 删除	
	十						
						确认	取消

- ・加工算子名称
- ・加工算子类型

当前版本仅支持数据加工。

・引用逻辑表

选择一张逻辑表,可以是数据导入算子产出的逻辑表,或其它算子产出的逻辑表。

・产出逻辑表

为产出的逻辑表起一个名字,不可与当前已存在逻辑表重名。

・添加数据过滤条件

此处配置过滤出id>1的文档,相当于SQL语句中的where子句,是作用在引用逻辑表上的一个过滤条件,详情请参见数据过滤。

・増加字段

有两种方式向算子中添加字段,分别是"手动添加"和"从引用逻辑表批量添加":

 - 手动添加:一次只能添加一个字段到当前算子,适用于较为复杂的数据加工场景,可以修 改目标字段名称,也可以选择多个源字段。如果对应字段的数据加工配置函数中需要多个 输入参数,也必须进行手动添加。

添加产出逻辑表字	段				×
字段添加方式:	● 手动添加	○ 从引用;	逻辑表批量添加		^
* 字段名称:	TName		8		- 1
* 字段类型:	VARCHAR		\sim		- 1
添加引用表字段:					- 1
查找字段			name		
id			title		- 1
age		>			- 1
		<			- 1
2 项			2 项		
				确认	▼

单击确认后,在算子字段列表中显示如下:

产出表字段	主键	类型	引用表字段	加工函数	操作	l
id	•	INT	id		数据加工配置 删除	l
name		VARCHAR	name		数据加工配置 删除	
age		INT	age		数据加工配置 删除	の咨询
title		VARCHAR	title		数据加工配置 删除	□ · 建议
TName		VARCHAR	name title		数据加工配置 删除	1
十 増加字段						

- 从引用逻辑表批量添加:可以一次性从引用逻辑表中选择多个字段添加到当前算子,其字
 段名称不可修改。适用于较为简单的数据加工流程,例如不需要数据加工函数,或数据加工函数仅需要一个参数的场景。

添加	叩产出逻辑表字	段				\times
	字段添加方式:	○ 手动添加	Ο 从引用道	逻辑表批量添加		
	查找字段					
	🗸 name					
	🗸 age		>			
	✓ title		<			
	✔ 3/3 项			1项		
					确认 耴	消

在查找字段列勾选需要的字段,单击向右箭头,再单击确认,所选择的所 有字段被添加到算子中。每个字段自动生成一个同名的目标字段,如下图所

示:	编辑数据加工算子			
	*加工算子名称:	data_process_test	8	
	*加工算子类型:	数据加工	\sim	
	* 引用逻辑表:	data_import_table	\sim	0
	* 产出逻辑表:	data_process_produce	8	
		✔ 添加数据过滤条件		
		id>1	8	(?)

产出表字段	主键	类型	引用表字段	加工函数	操作
id	•	INT	id		数据加工配置 删除
name		VARCHAR	name		数据加工配置 删除
age		INT	age		数据加工配置 删除
title		VARCHAR	title	_	数据加工配置删除
十 増加字段					

数据加工配置

在添加字段完成后,单击数据加工配置开始配置。以TName为例,选择CONCAT函数,该 函数的功能是对两个字段值进行拼接。

在下拉列表中选择函数CONCAT,在下方输入框中,填写该函数所需要的输入参数(即需要进行连接的 name 和 title 这两个字段名称),各参数之间用英文逗号隔开。配置完成后单击确认,如下图所示:

数据加工配置		×
被引用字段:	name title	
* 选择数据加工函数:	CONCAT	~ ?
	name,title	
		确认 取消

确认提交后,在算子字段列表中显示如下:

产出表字段	主键	类型	引用表字段	加工函数	操作
id	•	INT	id	_	数据加工配置 删除
name		VARCHAR	name		数据加工配置 删除
age		INT	age	_	数据加工配置 删除 谷
title		VARCHAR	title		22 数据加工配置 删除 22 议
TName		VARCHAR	name title	CONCAT	数据加工配置 删除
十 増加字段					
					确认 取消

5. 当算子中的所有字段配置完成后,再点击右下方确认按钮保存配置。

6.4.5 创建数据目标算子

数据导入或数据加工算子创建完成后,您可以创建对应数据目标算子。

- 1. 在任务列表管理页面,单击任务名称或者编辑按钮。
- 2. 在任务基础信息页面,单击任务配置下的数据目标标签页。

任务配置					创建任务商	躍
① 完整的任务配置流程需要经历 1. 导入数 分别对这三步进行添加和编辑。详细说	牧据 -> 2. 数据加工(添加数据加工算子) -> 3. 對 明	数据目标 (添加目标算子) 三个步骤	,选择创建任务配置,	您可以按步骤整体创建任务。	创建之后 , 您也可以	
导入数据 数据加工	牧据目标					咨询
添加数据目标算子				请输入算子名称		· Q 议
数据目标算子名称	数据目标类型	引用逻辑表	操作			
		没有数据				

- 3. 单击添加数据目标算子,在添加数据目标算子页面,输入相关参数,详情见数据目标算子。
- 4. 单击确认。

6.4.6 启动任务

数据目标算子创建完成后,您可以启动对应任务运行处理。

1. 进入任务列表管理页面。

<	eh-cn-kxb7ddm25fnp425aq						刷新
基础信息	Allige Eligit						
任务列表	00Xe (#0221						
日志春询	任务名称	状态	进度	资源占用	创建时间	撮作	
	pI-I2ae27te2bft25s7p test	● 待启动	0%	0 CU	2 小时前	启动 编辑 日志查词 删除	

2. 选择要启动的任务,单击启动。



确认当前项目中的任务资源分配是合理的(项目资源分配可参考任务资源及分配策略),以及 各算子的配置信息正确,任务启动后后台会做完整性检查。

3. 单击日志查询查看已启动任务的处理进度或报错提示信息。

<	eh-cn-i29h6851085ppy3	x3				刷新
基础信息	pl-02j6j2ha0wij4yf97	~		2019年3月22日	- 2019年3月29日	的一般素
任务列表	时间	任务名称/ID	日志内容			
日志查询	2019-03-26 15:24:31	pl-02j6j2ha0wij4yf97	incbuild start successfully, taskId[0] nu	mTasks[1]		
	2019-03-26 15:24:21	pl-02j6j2ha0wij4yf97	Restarting EsInc success			
	2019-03-26 15:22:51	pI-02j6j2ha0wij4yf97	Restarting EsInc			
	2019-03-26 15:21:04	pl-02j6j2ha0wij4yf97	EsFull finished			

6.4.7 删除任务

当确认不再需要某个项目下的某个同步任务时,可以在对应任务列表界面中删除对应同步任务。

1. 进入任务列表管理页面。

<	eh-cn-kxb7ddm25fnp425aq						刷新
基础信息	台(活業 目)協任						
任务列表	Page 1						
日志音询	任务名称	状态	进度	资源占用	创建时间	操作	
	pl-l2ae27te2bft25s7p test	● 待启动	0%	0 CU	2 小时前	启动 编辑 日志查词 删除	

2. 选择要删除的任务, 单击删除。

道 说明:	
任务删除后,	相关算子配置也会一并被删除。

3. 在删除任务提示框中,单击确认。

?	删除任务		×
	确定要删除任务:	9000 finds	(dayo?
		确认	取消

6.5 日志查询

通过日志查询您可以查看任务运行时间、任务名称/ID、日志内容等信息。

- 1. 在项目列表管理页面,单击项目名称。
- 2. 在项目详情页面,单击日志查询。

<	eh-cn-		限制行
基础信息	pl-02j6j2ha0wij4yf97	\vee	2019年3月22日 - 2019年3月29日 📖 拨素
任务列表	时间	任务名称/ID	日志内容
日志查询	2019-03-26 15:24:31	pl-	incbuild start successfully, taskid[0] numTasks[1]
	2019-03-26 15:24:21	pl-	Restarting Esinc success
	2019-03-26 15:22:51	pl-	Restarting Esinc
	2019-03-26 15:21:04	pl-	EsFull finished

3. 选择任务ID和日期范围并单击搜索。

eh-cn-				刷新
pł]		2019年4月1日 - 2019年4月2日	的一般来
pl	任务名称/ID	日志内容		
✓ pl-				

6.6 任务

6.6.1 任务概述

任务列表

在任务列表管理页面,您可以查看任务名称、状态、进度、资源占用、创建时间等信息。

<	eh-cn-						刷新
基础信息	创建剧新						
任务列表							
日本春海	任务名称	状态	进度	资源占用	创建时间	操作	
	pl- test	● 待启动	0%	0 CU	2 小时前	启动 编辑 日志查询 删除	

创建任务

在任务列表管理页面,单击创建可进行任务创建,详情请参见创建项目和任务。

操作任务

- · 启动: 单击启动运行对应任务。
- ·编辑:单击编辑修改对应任务。
- · 日志查询: 单击日志查询查看对应任务运行进度和报错提示信息。
- ·删除:单击删除删除对应任务。

6.6.2 任务类型

任务类型

ElasticFlow支持任务类型如下:

- ・ 全量导入数据(全量): 针对RDS for MySQL或者MaxComputer数据, 全量拉取所有数据并
 建立索引, 有完整的开始和结束时间。资源分配评估详情请参见任务资源及分配策略。
- · 增量导入数据(增量):针对LogService或者DTS数据,持续的向Elasticsearch发送更新数 据,具有无界(Unbounded)的特点。资源分配评估详情请参见任务资源及分配策略。

类	型	对	ĿŁ
~	<u> </u>		~~

	全量	增量
数据源	MaxComputer、RDS for MySQL、Elasticsearch	DTS、LogService
计算类型	批处理	流处理
生命周期	有明确开始和结束时间。	有明确的开始时间。任务不会 停止,除非在控制台手动停止 任务。
所需CU资源	相对较多	相对较少

6.6.3 任务资源及分配策略

任务资源及分配策略

ElasticFlow 的一个任务拓扑结构是典型的Source->Processor->Sink模型,同时计费单位是CU,一个CU对应的是虚拟的1核4G内存。用户对增量和全量指定的计算资源会全部分配给相应的任务。为了让Source、Processor和Sink能够达到最大吞吐(没有瓶颈节点,不会反压),可以根据本文中的策略来平衡资源分配。

增量任务分配策略

- ·资源分配:各个节点和并发均匀分配(通常情况下5CU能够满足1000QPS)。
- · 在Source节点: DTS采用单并发策略, LogService 会根据partition来控制并发。
- · 在Sink节点: Sink节点单并发向Elasticsearch发送数据。

全量任务分配策略

- · 资源分配:根据数据源和Elasticsearch的index动态计算。通常资源量在20-200CU能够在2小时内建立100G数据的index。
- · 在Source节点: MaxCompute会水平切分成多个并发。RDS考虑到数据一致性采用单并发策略。
- · 在Sink节点:最大限度满足资源需求,同时考虑index的partition数量。