从阿里云日志服务 sls 导出到 oss

作者: fish2018

- 原文链接: https://ld246.com/article/1603967672649
- 来源网站:链滴
- 许可协议:署名-相同方式共享 4.0 国际 (CC BY-SA 4.0)



背景

sls收费存储和索引各占一部分,而且索引的收费不便宜,毕竟查看很久之前的历史数据是比较低频的 作。

为节约费用,可以将日志直接导出到归档规格的oss,sls本身只保留较短时间即可(比如一周)。需要 看历史日志时,再创建个临时project将日志从oss导入sls,看完删除即可。

一、新写入日志导出到oss

sls可以直接开启投递到oss,但是只支持新写入的日志

1. 点开logstore,在导出中选择OSS(对象存储)

Ξ	C-D 阿里云		Q 搜索文档、控	制台、API、解决方	案和资源	费用	工单 备案
<	k8s-log-custom-logtail 切换	â					
0	日志库 我的关注	概览					
Eb	搜索logstore Q 十	┃访问域名					
₩	 ✓ ■ k8s-tes Ø Ø ⑦	私网域名	cn-sh	nanghai-intranet.log	g.aliyuncs.com		公网域名
8	 ✓ 參 数据接入 > ◎ logtail配置 	基础信息					
3	> ② 数据导入	地域	华东2(上海)				注释
C	∨ ∋ 数据处理	全球加速	未开启				创建时间
জ	> 血 加工 > 🕞 快速查询	自定义域名	暂无配置				
≣	> <u></u> 倍警	■ 服务日志 收起	1				
ш	◇ 參 导出						
L. 点元	• MaxCompute (原OD • OSS (对象存储) • EMR > 參 AnalyticDB 告开启投递	OSS (对象存储) 0.1 和	-: якажак эк -: к 37.62 %	8≏8↑8 4⊼ 3-1 4 94 ↑	еляра ок. (п 2 ÷	■皮根出版制印度 0 末	Ę
<u>م</u>	◎ OSS投递(k8s ×						
055	奴 地 官 理 (属 f kös-tes 📄						
1/]\8	对 3小时 6小时 12小时 1天 25	F					开启投递
任务	开始时间 任务结束时间	接受日志数据时间	数据行数	任务类型	状态	1	注释
		没有	有数据		/		
重词	《全部失败任务				总数: 0 <	上一页	1 下一页 >

3. 填写用于保存日志的OSS信息

Q 搜索文档、控制台、AI	OSS 投递功能		×
⑥ ⑧ OSS投递(k8s ×	* Logstore名称:	k8s-te.	
OSS投速管理(属于k8s-tet ■)	OSS投递属性 帮助		
1小时 3小时 6小时 12小时 1天	*OSS投递名称:		
任务开始时间任务结束时间	* OSS Bucket:		
	OSS Prefix:		
重试全部失败任务	分区格式:	从日志服务同步到OSS的数据将存放到Bucket的该目录下 %Y/%m/%d/%H/%M	
0		按照时间动态生成目录,默认值为%Y/%m/%d/%H/%M,对应生成目录例如 2017/01/23/12/00,注意分区格式不能以/开头结尾,如何结合E- MapReduce(Hive/Impala等计算引擎)进行查询分析请查看 帮助	
	*RAM角色:		
		用于访问权限控制, OSS Bucket拥有者创建角色的标示, 如 "acs:ram::13234:role/logrole"	
	• 投递大小:	256	
		自动控制投递任务创建间隔并设置OSS的一个Object大小(以未压缩计算)上限,单位:MB	
		确定	取消

二、历史日志导出到oss

准备信息:

- 源sls的project
- 目标oss的bucket
- AccessKey ID 和 AccessKey Secret

历史日志导出到oss目前只能借助dataworks来实现

1. 开通dataworks服务,在空间列表创建新工作空间

DataWorks	DataWorks / 工作空间列表	创建工作空间						
极变	当前使用的是 輸進版 、版本》	1 基本配置		2 选择	引撃			3 引擎详情
资源组列表	他跟工作空间 错输入工	基本信息 • 工作空间名称	sis os	s sync				
根爾德維	工作空间名称/显示名							
计算引擎列表 ^		显示名	sis_os	s_sync				
MaxCompute		• 模式 😡	简单模	式(单环境)			~	
Graph Compute	<	描述						
Hologres	×							
		高级设置						
		• 能下载Select结果						
点击数据集成 三 (-) 啊聖云 ***2	(上海) •	下→步 Q 撥素文性、控制	取消 时,API,解决方案和竞赛	费用 工单 备案	企业 支持	管网 匹	ଦ ି କ୍ର	iii (\$
DataWorks	DataWorks / 工作空间列表							* #
概定								
工作空间列表	当前使用的是 标准版 , 版本到期	日为 2020年11月23日	۱.			版本升级 版2	本延期 查看版本详	i情 购买独享资源:
资源组列表	创建工作空间 请输入工作	空间/显示名	Q					
报警资源	工作空间名称/显示名	模式	创建时间	管理员		状态	操作	
计算引擎列表 へ MaxCompute	sis_oss_sync sis_oss_sync	简单模式 单环境	2020年10月28日 15:27:40			✓ 正常	进入首页 进入 进入运维中心 工作空间配置	入数据集成进入数据 进入数据地图进入 修改服务配置更重
Graph Compute Hologres								共1页

2. 进入数据源管理

≡	们 DataWorks	数据集成	sls_oss_sync sls_oss_sync	~		
<u>₀</u>	三 蓄页 批量上云	欢迎	使用数	据集成		
ŵ ■ 小	自定义资源组 整库迁移 数据源		□ 新建离线同步		【 ↓ 新建实时任务	↑ 数据源管理
μ.	同步解决方案 🏎		使 。 任务运维管理			
		帮助文档 支持的离射	國政 國家 支持的实时算	收据源 高线任务配置	网络连通和解决方案	

3. 点击新增数据源



4. 添加sls数据源,选择loghub,网络选择阿里云vpc,其他根据具体信息填写即可

新增数据源	
Elasticsearch	
消息队列	
LogHub	

. White yes do the	Take when A.L. who when	
* 数据源名称:	自定义名称	
数据源描述:		
网络连接类型:	请选择	
* LogHub Endpoint :	如: http://cn-shanghai.log.aliyuncs.com	?
* Project :	请输入Project	
* AccessKey ID :		?
* AccessKey Secret :		
资源组连通性:	数据集成 任务调度 ?	
如果数据同步时使 决方案。	用了此数据源,那么就需要保证对应的资源组和数据源之间是可以联通的。请参考资源组的详细概念和网	络解
查看当前最佳网络方案推	「花」	
	连诵状态	



半结构化存储 OSS HDFS

新增OSS数据源					×
*数据源名称:	自定义名称				
数据源描述:					
网络连接类型:	阿里云VPC		~]	
地域:	✔ 与当前DataWorks同region				
* Endpoint :				?	
* Bucket :				?	
* AccessKey ID :				?	
* AccessKey Secret :					
资源组连通性:	数据集成任务调度 ?				
可如果数据同步时使用 决方案。	用了此数据源,那么就需要保证对应	的资源组和数据源之间是可以联通	通的。请参考资源组的详 细概 念和M	网络解	
. 进入数据开发页面	☰ DataWorks 数据集组	sis_oss_sync sis_oss_sync			
	田 全部产品 >>	(1) 使用数据集(1)	1 0		
	(作) DataWorks(首页)		~		
	☆ DataStudio(数据开发) ×	Ū.	E		
	益维中心(工作流)	新建离线同步	新建实时任务		数提
	∽ 数据地图				
		任务运维管理			



7. 点击带"+"的图标, 创建业务流程



	新建业务流程			×	
	业务名称:	sis2oss		1	:/
	描述:	请输入业务描述			u7228
660 xit 020-	10-28 19:10:29 [NFD] Sandbox context cleanup temp file success.	新建 取消	1	

8. 创建离线同步,目标文件夹中就是刚才创建的业务流程



新建节点		×
节点类型:	离线同步	~
节点名称:	write_result	
目标文件夹:	sls2oss/业务流程	~
0 t		提交取消

9. 配置同步信息

来源选择loghub(sls)

● 注意时间戳格式

去向选择oss

- 文本类型选择text
- 列分隔符配置一个日志内容中不会包含的字符,可以用生僻汉字组合的unicode

● 并发数根据情况设置,公共资源组的机器网络能力都是共享的 高峰时段资源会有争抢的情况,任务 慢和资源组网络能力、源、目标负载都有关系



点击保存,运行,不要点击提交,否则会周期性运行。

三、从oss把日志导入sls

oss导入到sls有2个入口:

● project/接入数据

文档

1. 打开project,点击右侧的接入数据

文档

(-) 阿里云		Q 搜索文档,控制台, API	· 解决方案和资源 费用 工单	备案 企业 支持 官网 🖸	Q. A. Q 👷 🐧
k8s-log-custom-l	ogtail 👥 🙃				
概览					权限助手 接入数据
访问域名					
粘胸域名	cn-shanghai-intranel	log.allyuncs.com	公開城名	cn-shanghai.log.aliyuncs.com	
基础信息					
- 10 M	华东2(上海)		注释 kd	35收集日志	/
全球加速	未开启		创建时间 24	020-10-10 14:57:27	
自定义域名	智无配置				
■ 服务日志 👷	起へ				
					(E
选择OSS-ヌ	寸象存储				
接入数据			请选择分类	∨ 请输入关键词	Q X
友盟 - 数据开	放平台	SS - 对象存储	正则 - 文本日志	hot 单行 - 文本	x日志 hot
文档	× x	a 🕞 🥌	文档	.** 这档	
	hot		hot	new	new
Nginx - 文本		ubernetes - 标准.	主机监控	MySQL监持	₽ C
213				人口	
Redis监控	new	asticsearch监控	new	Tomcat版	new

文档

原文链接:从阿里云日志服务 sls 导出到 oss

文档

3. 如果之前OSS使用的规格是归档型,这里要打开解冻,填好信息点击预览,然后点击下个配置

数据 导入	_		Q 理重文档.	控制台、API、输决	方票和资源	費用 工单	备案	企业	支持 官同	∑ _m
	选择日志空间	导入	配置	1	询分析配置			结束		
	数据源设置		送格式配置	\rightarrow	调度间隔					
	配置名称:	oss2sis								
	OSS Region:	华东 2(上海)				~				
	Bucket:	詞region访问会产生公网混量 sis-k8s-log-custom-logtail				~				
		臺看文件列表								
	文件夹前缀:	logistics/ 只支持导入2g以内的文件								
	正则过速:									
		使用Java正则表达式语法								
	數据格式:	单行文本日志				~				
	压缩格式:	未压缩				~				
	编码格式:	UTF-8				V				
	解冻归档文件:	「白格存储需要解冻之后才可访问				预数				
	文件预览结果:	2020-10-27 19:20:10.297 [react 2020-10-27 19:20:10.474 [react 2020-10-27 19:20:10.474 [react 2020-10-27 19:20:10.565 [react 2020-10-27 19:20:10.645 [react	tor-http-epoil-3) [] INFC tor-http-epoil-4) [] INFC tor-http-epoil-1] [] INFC tor-http-epoil-2) [] INFC tor-http-epoil-3] [] INFC) c.i.lu.client.filter.Pr) c.i.lu.client.filter.Pr) c.i.lu.client.filter.Pr) c.i.lu.client.filter.Pr) c.i.lu.client.filter.Pr	ermissionFluxFilte ermissionFluxFilte ermissionFluxFilte ermissionFluxFilte ermissionFluxFilte	r - 放行url=/ai r - 放行url=/ai r - 放行url=/ai r - 放行url=/ai				
					上个能置	下个配置				

如果选择oss region时提示没有权限,需要点击提示连接授权后再继续操作

=	(-) 阿里云		Q 搜索文档、控制台、 ✔	API、解决方案和资源 云资源访问授权成功!	费用	工单	备案
		云资源访问授权 如需修改角色权限,请前往 RAM 控制台 角 (色管理 中设置,需要注意的是,错	i误的配置可能导致 Cloud	Monitor :	无法获取到	制必要的权限
•	 Log 请求获取订 下方是系统创建的 AliyunLogImpo 	防问您云资源的权限。 可供 Log 使用的角色,授权后,Log 拥有对: ortOSSRole 展开详情	您云资源相应的访问权限。				

 「意授权 取消 4. 点击测试,显示结果后点击下个配置 	数据导入	选择日志空间		2 导入配置	3
		数据源设置		数据格式配置	调度间隔
		使用系统时间:			
		测试结果:	日本計画	content	
			1603943382	2020-10-27 19:20:10.29	7 [reactor-http-epoll-3] []
			1603943382	2020-10-27 19:20:10.47	4 [reactor-http-epoll-4] []
			1603943382	2020-10-27 19:20:10.47	4 [reactor-http-epoll-1] []
			1603943382	2020-10-27 19:20:10.56	5 [reactor-http-epoll-2] []
			1603943382	2020-10-27 19:20:10.64	5 [reactor-http-epoll-3] []

5. 导入间隔改成5分钟,打开立即执行,点击下一步 (\cdot) 数据导入 4 选择日志空间 导入配置 数据源设置 数据格式配置 5分钟 ¥. 导入间期: 立即执行: 6. 打开索引 (\cdot) (\cdot) (4) 数据导入 选择日志空间 导入配置 查询分析配置 结束 预览数据 自动生成素引 * 全文索引 大小写敏感 时间/来源 内容 包含中文 分词符 . ":=000?@&->/:/n\//v 2020-10-29 15:03 content:2020-10-27 23:59:21.078 [reactor-http-epoil-4] [] INF O c.Liu.client.filter.PermissionFluxFilter - 放行url=/actuator/inf • 指定字段查询 sis-k8s-logcustom 开启查询 字段名称 包含中文 开启统计 删除 logtail 黄型 别名 大小写敏感 分词符 2020-10-29 **()** × text . ":=0007 content 15:03 content:2020-10-27 23:59:21.078 [reactor-http-epoil-1] [] INF O c.i.iu.client.filter.PermissionFluxFilter - 放行url=/actuator/inf sls-k8s-log customlogtail 统计字段 (🖙) 最大长度 😑 📃 + 2020-10-29 15:03 1. 全文索引属性和字段索引属性必须至少启用一种 content:2020-10-27 23:59:21.079 [reactor-http-epoil-2]]] INF O c.i.iu.client.filter.PermissionFluxFilter - 放行url=/actuator/inf sis-k8s-log 2. 索引类型为long/double时,大小写敏感和分词符属性无效 3. 如何设置索引请参考文档说明(查看帮助) custom logtail 4. 打开对应列的统计功能、才能使用SQL进行快速统计分析。(分析语法) 2020-10-29 15:03 content:2020-10-27 23:59:21.158 [reactor-http-epoil-3] [] INF O c.i.iu.client.filter.PermissionFluxFilter - 放行url=/actuator/inf sis-k8s-logcustom-7. 点击下一步后提示完成接入配置 ●返回概览 项目 (Project): k8s-log-custom-logtail 日志库 (Logstore): Import-log (\checkmark) (1) (\checkmark) 数据导入 选择日志空间 导入配置 查询分析配置 恭喜您完成数据接入配置,配置会在1分钟内生效,请耐心等待。 接下来您还可以: 查询日志 投递OSS 70 描述 立即尝试 帮助文档

- logstore/导入
- 1. 打开logstore,点击数据导入右边的"+"

参考资料

实时日志,可以在日志服务控制台开启oss数据投递

https://help.aliyun.com/document_detail/29002.html

dataworks创建同步任务

https://help.aliyun.com/document_detail/85555.html

dataworks来做数据同步:

LogHub Reader

https://help.aliyun.com/knowledge_detail/137732.html

OSS Writer

https://help.aliyun.com/knowledge_detail/137765.html

dataworks产生的费用: 开通DataWorks按量计费就可以, dataworks提供了免费额度 https://help. liyun.com/document_detail/118599.html

sls新版预付资源包折扣策略说明

https://help.aliyun.com/document_detail/173043.html

sls存储空间使用情况

账单详情里可以看到,登录https://usercenter2.aliyun.com/finance/expense-report/expense-det il-by-instance 选择"计费项"和"明细",切记,注意!产品选择日志服务,将csv导出,这里能 到每个logstore的日志size。

sls导入oss数据

https://help.aliyun.com/document_detail/147923.html

开启并配置索引

https://help.aliyun.com/document_detail/90732.html