文档名称 文档密级

# 数据仓库服务使用—云上数据流动

本文档演示如何通过CDM将DWS的表数据归档到OBS上,同理,也可以将OBS上的数据文件通过CDM传输到DWS集群里。

### 一. 创建DWS集群

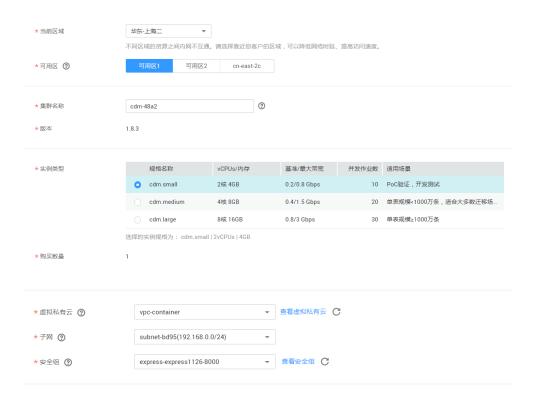
DWS创建流程已经在第一天的文档介绍,此处不再赘述。

# 二. 创建CDM集群

1. 登陆华为云控制台,进入"云数据迁移CDM"管理页面,点击进入控制台。



2. 点击右上角的"购买云数据迁移服务",进行创建CDM集群的配置。



集群名称: CDM的集群名称



文档名称 文档密级

实例类型: CDM集群规格, 选择cdm.small规格

虚拟私有云:需要和DWS集群的虚拟私有云保持一致

其余选项选择默认即可。

点击"点击立即购买",开始创建CDM集群,等待几分钟,集群创建完成。如下图 所示:



注: CDM的收费方式为运行作业时收费,此种规格运行作业时0.56元/小时, 关机时费用0.05元/小时,所以当没有作业需要运行迁移时,可以选择关机,这样 节省费用。当需要迁移数据时,在开机CDM集群。

# 三. 创建OBS桶并上传数据

1. 登陆华为云"对象存储服务OBS"的管理控制台。



2. 点击"管理控制台",点击右上角"创建桶"。然后进行配置。



文档名称 文档密级



区域:注意区域需要选择和DWS集群,CDM集群相同的区域,保证它们之间的数据是能够访问通的。点击"立即创建"。稍等几分钟后,桶会创建好。桶创建好后,点击桶名,进入详情界面,如下图:



## 点击左侧"对象",如下图:



文档名称 文档密级

可知,这里是存储数据文件的页面,DWS导出的数据之后也应该在这个页面可以 查看到。

### 四.配置CDM作业

- 1. 回到CDM管理页面,点击CDM集群上的"作业管理",开始配置连接管理和作业。
- 2. 首先配置源端OBS连接。点击"新建连接"。选择"对象存储"中的"对象存储服务(OBS)"

对象存储	对象存储服务(OBS)	阿里云对象存储(OSS)	七牛云对象存储(KODO)	亚马逊对象存储S3
文件系统	FTP	SFTP	НТТР	网络附加存储 (NAS)
关系型数据库	云数据库 MySQL	云数据库 PostgreSQL	云数据库 SQL Server	分布式数据库中间件 ( DDM )
	PostgreSQL	Microsoft SQL Server	Oracle	IBM Db2
	SAP HANA			

# 点击"下一步",

* 名称	OBS-DEMO		
* 连接器	OBS ▼		
ALLES AN			
对象存储类型	华为云对象存储OBS ▼		
OBS終端节点 ②	obs.cn-east-2.myhwclouds.c		
端口 ②	443		
访问标识(AK) ②	******		
密钥(SK) <b>②</b>			
× 取消 〈上一步	☎ 测试 🔒 保存		

名称: 给源端连接起个名字

连接器: OBS, 不可选

OBS终端节点:地址,不可选

端口: 默认443

访问标识/密钥:填写华为云账号的AK/SK

填写完成后,点击"测试",如果测试成功,点击保存。

3. 配置目的端DWS连接。点击"新建连接",选择数据仓库中的"数据仓库服务



文档名称 文档密级

(DWS) ".

数据仓库	数据仓库服务(DWS)	FusionInsight LibrA	数据湖探索(DLI)

点击"下一步",进行数据仓库的配置:

* 名称	jyz-test	
* 连接器	关系数据库   ▼	
数据库类型	数据仓库   ▼	
* 数据库服务器 ②	jyz-test.dws.myhuaweicloud	选择
* 端口 ②	8000	
* 数据库名称 ②	postgres	
* 用户名 ②	dbadmin	
* 密码 ②		
导入模式 ②	COPY	
显示高级属性		
▼ 取消	6 测试	

数据仓库的配置和创建DWS集群的配置相同,此处不再赘述。填写完成后,点击"测试",如果测试成功,点击保存。

4. 配置CDM作业,开始迁移。在"作业管理"页面,点击"新建作业"。



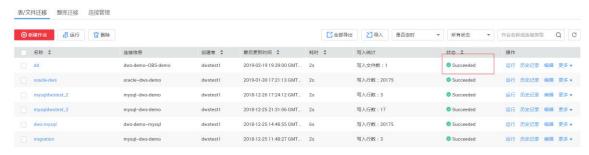
源端作业配置,我们选择刚刚创建的DWS连接,然后选择我们想要导出的特定模式下表;目的端连接我们选择创建OBS连接,选择我们需要将DWS表导出到OBS的哪个目录,配置好后点击"下一步",最后点击"保存并运行"。

5. 查看作业运行,并检查作业是否导出。回到"作业管理"页面,可以看到历史



文档名称 文档密级

运行的作业情况。如下图:



作业如果成功,会显示"Succeeded"。如果失败,可以更多里面"日志",来查看失败的原因。然后修改配置后,可以点击"运行",再次运行作业,知道作业成功。

当作业运行成功后,我们到OBS页面查看是否有数据导出。



可以看到,在OBS桶的根目录下,有文件导出,可以下载的本地进行查看。

至此,通过CDM导出DWS数据到OBS上的操作就全部完成了。

# 打卡任务:

通过CDM作业,完成将OBS的数据文件导入到DWS集群,查看CDM作业的状态,并且在DWS集群中查看数据是否导入成功。