

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。

* 当前区域 华东-上海二

不同区域的资源之间内网不互通。请选择靠近您客户的区域，可以降低网络时延、提高访问速度。

* 可用区 可用区1 可用区2 cn-east-2c

* 集群名称 cdm-48a2

* 版本 1.8.3

* 实例类型

规格名称	vCPUs/内存	基准/最大带宽	并发作业数	适用场景
<input checked="" type="radio"/> cdm.small	2核 4GB	0.2/0.8 Gbps	10	PoC验证，开发测试
<input type="radio"/> cdm.medium	4核 8GB	0.4/1.5 Gbps	20	单表规模<1000万条，适合大多数迁移场景
<input type="radio"/> cdm.large	8核 16GB	0.8/3 Gbps	30	单表规模≥1000万条

选择的实例规格为：cdm.small | 2vCPUs | 4GB

* 购买数量 1

* 虚拟私有云 vpc-container [查看虚拟私有云](#)

* 子网 subnet-bd95(192.168.0.0/24)

* 安全组 express-express1126-8000 [查看安全组](#)

集群名称：CDM的集群名称

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。

区域 华北-北京一

不同区域的资源之间内网互不相通，请选择靠近您业务的区域，可以降低网络时延，提高访问速度。桶创建成功后不支持变更区域，请谨慎选择。

桶名称 obs-5556

命名规则：
- 需全局唯一，不能与已有的任何桶名称重复。
- 长度范围为3到63个字符，支持小写字母、数字、中划线（-）、英文句号（.）。
- 禁止两个英文句号（.）或英文句号（.）和中划线（-）相邻，禁止以英文句号（.）和中划线（-）开头或结尾。
- 禁止使用IP地址。
- 如果名称中包含英文句号（.），访问桶或对象时可能会进行安全证书校验。

存储类别 标准存储 低频访问存储 归档存储

适用于有大量热点文件或小文件，且需要频繁访问（平均一个月多次）并快速获取数据的业务场景。
上传对象时，对象默认与桶的存储类别相同，也可以根据适用场景修改。 [了解更多](#)

桶策略 私有 公共读 公共读写

任何用户都可以对桶内对象进行读/写/删除操作。

多AZ ? 当前仅【华东-上海二、华北-北京四、西南-贵阳一】区域支持配置多AZ。

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：

对象存储服务

桶列表，hangtiantest

基础数据统计

存储统计

存储用量	对象数量
48.02 MB	7

GET: 445, PUT: 8

内网流量统计

流出 (KB)	流入 (Byte)
749.26	584

基本信息

桶名称	hangtiantest
存储类别	标准存储
桶版本号	3.0
区域	华东-上海二
所有者	hwstaff_pub_cides
账号ID	fc9f1080d94248a491a840f035749c3b
创建时间	2018/12/22 10:19:00 GMT+08:00
多版本控制	未启用
Endpoint	obs.cn-east-2.myhuaweicloud.com
访问域名	hangtiantest.obs.cn-east-2.myhuaweicloud.com
多AZ	未开启

基础配置

日志记录	未配置
静态网站托管	
生命周期规则	未配置
防盗链	
事件通知	未配置
CORS规则	
标签	未配置

点击左侧“对象”，如下图：

对象存储服务

桶列表，hangtiantest

对象 已删除对象 碎片

对象是数据存储的基本单位，在OBS中文本和文件夹都是对象。您可以上传任何类型（文本、图片、视频等）的文件，并在桶中对这些文件进行管理。 [了解更多](#)

上传对象 新建文件夹 恢复 删除 修改存储类别

名称	存储类别	大小	加密状态	恢复状态
bl_dm	-	-	-	-
dd-2019-03-19_19.39.00.135.csv	标准存储	26 byte	未加密	-
DWS_生成数据和导入数据样例.docx	标准存储	401.76 KB	未加密	-
javaudf-example.jar	标准存储	770 byte	未加密	-
TLE50000.csv	标准存储	14.78 MB	未加密	-
satcat.csv	标准存储	3.32 MB	未加密	-

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	HTTP	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
对象存储类型	华为云对象存储OBS
OBS终端节点 ?	obs.cn-east-2.myhwclouds.c
端口 ?	443
访问标识(AK) ?	*****
密钥(SK) ?	*****

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

(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	<input type="text" value="jyz-test"/>
* 连接器	<input type="text" value="关系数据库"/>
数据库类型	<input type="text" value="数据仓库"/>
* 数据库服务器 ?	<input type="text" value="jyz-test.dws.myhuaweicloud"/> 选择
* 端口 ?	<input type="text" value="8000"/>
* 数据库名称 ?	<input type="text" value="postgres"/>
* 用户名 ?	<input type="text" value="dbadmin"/>
* 密码 ?	<input type="password" value="*****"/>
导入模式 ?	<input type="text" value="COPY"/>

[显示高级属性](#)

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”。

作业配置	
* 作业名称	<input type="text" value="DWS-CDM-OBS"/>

源端作业配置	目的端作业配置
* 源连接名称	<input type="text" value="dws-demo"/> +
使用SqI语句 ?	<input type="text" value="是"/> <input type="text" value="否"/>
* 模式或表空间 ?	<input type="text" value="dbadmin"/> ⊙
* 表名 ?	<input type="text" value="test"/> ⊙

[显示高级属性](#)

* 目的连接名称	<input type="text" value="OBS-demo"/> +
* 桶名 ?	<input type="text" value="hangtianetest"/> ⊙
* 写入目录 ?	<input type="text" value="/"/> ⊙
* 文件格式 ?	<input type="text" value="CSV格式"/>

[显示高级属性](#)

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

运行的作业情况。如下图：

表/文件迁移

整库迁移

连接管理

新建作业

运行

删除

全部导出

导入

是否定时

所有状态

作业名称或连接类型

名称

连接信息

创建者

最后更新时间

耗时

写入统计

状态

操作

dd

dws-demo-OBS-demo

dwestest1

2019-03-19 19:39:00 GMT...

2s

写入文件数：1

Succeeded

运行 历史记录 编辑 更多

oracle-dws

oracle-dws-demo

dwestest1

2019-01-30 17:21:13 GMT...

2s

写入行数：20175

Succeeded

运行 历史记录 编辑 更多

mysqldwestest_2

mysql-dws-demo

dwestest1

2018-12-26 17:24:12 GMT...

2s

写入行数：3

Succeeded

运行 历史记录 编辑 更多

mysqldwestest_3

mysql-dws-demo

dwestest1

2018-12-25 21:31:06 GMT...

2s

写入行数：17

Succeeded

运行 历史记录 编辑 更多

dws-mysql

dws-demo-mysql

dwestest1

2018-12-25 14:48:55 GMT...

6s

写入行数：20175

Succeeded

运行 历史记录 编辑 更多

migration

mysql-dws-demo

dwestest1

2018-12-25 11:48:27 GMT...

2s

写入行数：3

Succeeded

运行 历史记录 编辑 更多

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。

对象

已删除对象

碎片

对象是数据存储的基本单位，在 OBS 中文件和文件夹都是对象。您可以上传任何类型（文本、图片、视频等）的文件，并在桶中对这些文件进行管理。[了解更多](#)

上传对象

新建文件夹

恢复

删除

修改存储类别

输入对象名前缀搜索

名称	存储类别	大小	加密状态	恢复状态	最后修改时间	操作
<div><div></div><div>hl_dm</div></div>	标准存储	26 byte	未加密	--	2019/03/19 19:39:00 GMT+08:00	复制路径 删除 更多
<div><div></div><div>dd-2019-03-19_19:39:00:135.csv</div></div>	标准存储	401.76 KB	未加密	--	2019/01/12 15:43:55 GMT+08:00	下载 分享 更多
<div><div></div><div>DWS_生成数据和导入数据样例.docx</div></div>	标准存储	770 byte	未加密	--	2018/12/25 16:08:54 GMT+08:00	下载 分享 更多
<div><div></div><div>javaudf-example.jar</div></div>	标准存储	14.78 MB	未加密	--	2018/12/22 10:10:29 GMT+08:00	下载 分享 更多
<div><div></div><div>TLE50000.csv</div></div>	标准存储	3.32 MB	未加密	--	2018/12/22 10:10:13 GMT+08:00	下载 分享 更多
<div><div></div><div>satcat.csv</div></div>	标准存储					

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。