

udwufile:在 ufile 上创建 udw 的外部表

一、udwufile 的使用简介

udwufile 打通了 udw 和 ufile 之间的联系，可以通过在 udw 上创建连接 ufile 的外部表，我们就可以通过 udw 直接访问 ufile 上的数据，也可以通过创建可写的外部表，把 udw 的数据写入 ufile。

1.1 创建 ufile 可读外部表

```
CREATE READABLE EXTERNAL TABLE tablename(columnname datatype [...] | LIKE
othertable)
LOCATION('udwufileprotocol')
FORMAT 'TEXT'
    [( [HEADER]
        [DELIMITER [AS] 'delimiter' | 'OFF']
        [NULL [AS] 'null string']
        [ESCAPE [AS] 'escape' | 'OFF']
        [NEWLINE [ AS ] 'LF' | 'CR' | 'CRLF']
        [FILL MISSING FIELDS] )]
    | 'CSV'
    [( [HEADER]
        [QUOTE [AS] 'quote']
        [DELIMITER [AS] 'delimiter']
        [NULL [AS] 'null string']
        [FORCE NOT NULL column [, ...]]
        [ESCAPE [AS] 'escape']
        [NEWLINE [ AS ] 'LF' | 'CR' | 'CRLF']
        [FILL MISSING FIELDS] )]
[ ENCODING 'encoding']
[ LOG ERRORS INTO error_table ]
SEGMENT REJECT LIMIT count [ROWS | PERCENT] ]
```

1.2 创建 ufile 可写外部表

```
CREATE WRITABLE EXTERNAL TABLE tablename(columnname datatype [...] | LIKE
othertable)
LOCATION('udwufileprotocol')
FORMAT 'TEXT'
    [( [HEADER]
        [DELIMITER [AS] 'delimiter' | 'OFF']
```

```

[NULL [AS] 'null string']
[ESCAPE [AS] 'escape' | 'OFF']
[NEWLINE [ AS ] 'LF' | 'CR' | 'CRLF']
[FILL MISSING FIELDS] ))
| 'CSV'
[( [HEADER]
[QUOTE [AS] 'quote']
[DELIMITER [AS] 'delimiter']
[NULL [AS] 'null string']
[FORCE NOT NULL column [, ...]]
[ESCAPE [AS] 'escape']
[NEWLINE [ AS ] 'LF' | 'CR' | 'CRLF']
[FILL MISSING FIELDS] ))
[ ENCODING 'write_encoding']

```

1.3 udwufile 协议说明

udwufileprotocol:

udwufile://**bucketname**/**filename**

public_key=your_public_key private_key=your_private_key

参数说明

udwufileprotocol 整个协议和相关参数描述如下

- 1) udwufile, 协议名。
- 2) bucketname, ufile 桶的名字
- 3) filename, ufile 上的文件名

对于只读外部表, filename 支持通配符 *。

例如 filename* 则可以被导入的文件是所有以 filename 为前缀的文件。

为了最大化导入性能, 建议拆分文件的个数是导入到的 udw primary segment 个数的倍数。

对于可写外部表, 指定 filename, 数据到导出为 filename0, filename1...;

文件的个数为 primary segment 的个数。

primary segment 的个数可通过下面的 sql 获得:

```
select count(*) from udw_segment_configuration where role ='p' and
content != -1;
```

4) public_key , ucloud 帐号的公钥

5) private_key, ucloud 帐号的私钥

其他参数说明:

FORMAT 支持的文件格式, 支持 text,csv

ENCODING 文件中数据的编码格式, 如 utf8

LOG ERRORS 容错模式数据导入, 指定该子句忽略掉导入中出错的数据

二、udwufile 使用举例

2.1 udwufile 可读外部表

1. 创建一个可读外部表

```
CREATE READABLE EXTERNAL TABLE ufile_read_example(id int, ,node_id varchar(20),
role varchar(20))
LOCATION('udwufile://testbucket/test_read* public_key=xxxxxxx private_key=xxxxxxx')
FORMAT 'csv' (DELIMITER ',')
LOG ERRORS error_table REJECT LIMIT 10;
```

2. 对于大文件, 将大文件拆分小文件, 小文件的个数最好是primary segment个数的整倍数, 然后将这些小文件上传到ufile上。上面创建的表会读取文件名以test_read为前缀的文件的数据。

3. 通过udw的sql就可以访问test_read为前缀的文件的数据。

```
SELECT * FROM ufile_read_example LIMIT 10;
```

4. 在 udw 上创建堆表, 数据就可以装载到这张表中

```
CREATE TABLE example(id int, ,node_id varchar(20), role varchar(20)) DISTRIBUTED BY
(id);
```

5. 数据并行的从 ufile_read_example 装载到 example 中。

```
INSERT INTO example SELECT * FROM ufile_read_example;
```

2.2 udwufile 可写外部表

1. 创建一个可写外部表

```
CREATE WRITABLE EXTERNAL TABLE ufile_write_example(id int, node_id varchar(20),  
role varchar(20))  
LOCATION('udwufile://testbucket/test_large_write.txt  public_key=xxxxxx  
private_key=xxxxxx') FORMAT 'csv';
```

2. 数据并行的从 example 导出到 ufile

```
INSERT INTO ufile_write_example SELECT * FROM example;
```

3. 在 ufile 上查看结果

<input type="checkbox"/>	test_large_write.txt0	8.24 GB	application/octet-stream 🔗	2016-08-23 13:37:44	📄 获取地址	📄 下载	🗑️ 删除
<input type="checkbox"/>	test_large_write.txt1	8.24 GB	application/octet-stream 🔗	2016-08-23 13:37:16	📄 获取地址	📄 下载	🗑️ 删除
<input type="checkbox"/>	test_large_write.txt2	8.24 GB	application/octet-stream 🔗	2016-08-23 13:35:44	📄 获取地址	📄 下载	🗑️ 删除
<input type="checkbox"/>	test_large_write.txt3	8.24 GB	application/octet-stream 🔗	2016-08-23 13:35:33	📄 获取地址	📄 下载	🗑️ 删除

2.3 使用注意

- 1、创建的 ufile 桶必须和 udw 在同一个地域,比如在北京创建的 udw 则必须创建北京地域的 ufile 桶。
- 2、数据导入的性能和 udw 实例的节点个数相关,也和 ufile 相关。