

目录

0.引言	1
1.将数据文件上传至 ECS 服务器	1
2.使用 copy 命令将数据文件中的数据导入到数据库表中的操作失败	3
3.导入数据到 employees 表时出现编码错误	8
4.导入数据到 employees 表时出现时间戳不匹配	10
5.使用带逗号作为定界符的 copy 命令导入数据到 customers 表时出现异常	13
6.数据文件表头问题	15
7.openGauss 数据库服务器的集群状态参数 Cluster_state 出现 unavailable	16

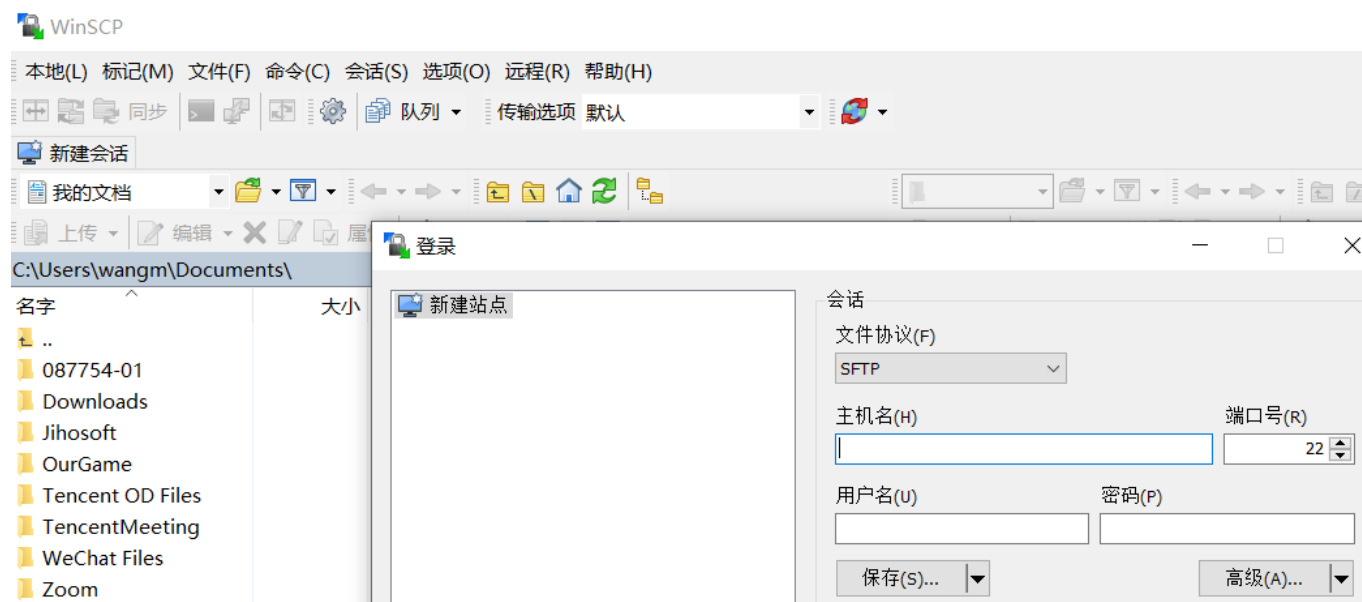
0.引言

利用 openGauss 的 COPY 命令可将数据文件的数据批量导入数据库表的操作分两步：1.将数据文件（通常为 CSV 格式）上传到 ECS 服务器；2.使用 COPY 命令将数据文件中的数据导入表中。

在数据导入过程中易出现的错误可归纳为以下四种情形：权限不足、编码不统一、时间戳格式不一致和用逗号作为定界符异常。

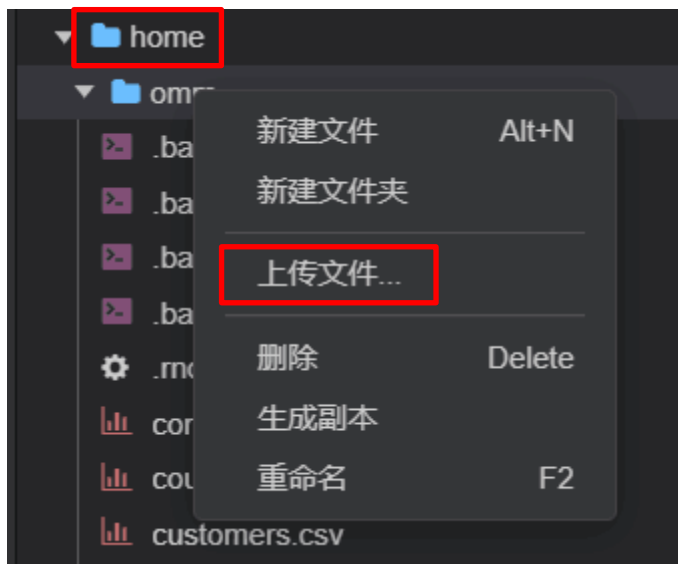
1.将数据文件上传至 ECS 服务器

方法一：使用 FTP 工具（此处示例工具为 WinSCP，其他 FTP 工具如 FileZilla 均可）（推荐）



主机名：ECS 公网 IP，用户名：root，密码为安装 openGauss 时设定的密码。

方法二：使用 ECS 提供的 Cloudshell 登录，进入 omm 的 home 目录，右键选中“上传文件...”完成相应的文件上传任务。



2.使用 copy 命令将数据文件中的数据导入到数据库表中的操作失败

问题描述：使用 copy 命令把数据文件中的数据导入数据库表失败，操作提示 “permission denied”。

```
db_jjh=> \copy orders FROM '/home/omm/orders.csv' WITH (delimiter',' ,IGNORE_EXTRA_DATA 'on');  
/home/omm/orders.csv: Permission denied  
db_jjh=>
```

原因：文件操作权限不足

解决方法：修改文件权限为可读 r，可写 w 和可执行 e，即 777 权限

步骤：

(1) 以 root 身份登录到 ECS，进入存放 CSV 格式数据文件的目录（使用 cd 命令进入，cd 命令用法见实验一之附录一 2.4）。（前提：CSV 数据文件已上传到 ECS 服务器）

```
[root@ecs-5555 omm]# ls
contacts.csv  customers.csv  employeesdata.sql  locations.csv  order_itmes.csv  product_categories.csv  regions.csv
countries.csv  employees.csv  inventories.csv  order_items.csv  orders.csv       products.csv            warehouses.csv
```

注：上图绿色文件为上传的 csv 文件

(2) 执行命令 chmod 777 文件名

```
[root@ecs-5555 omm]# chmod 777 orders.csv
[root@ecs-5555 omm]#
```

注：文件权限的修改是在 root 用户而非 omm 用户下执行，即在执行 su - omm 命令之前完成。否则，将出现如下错误：

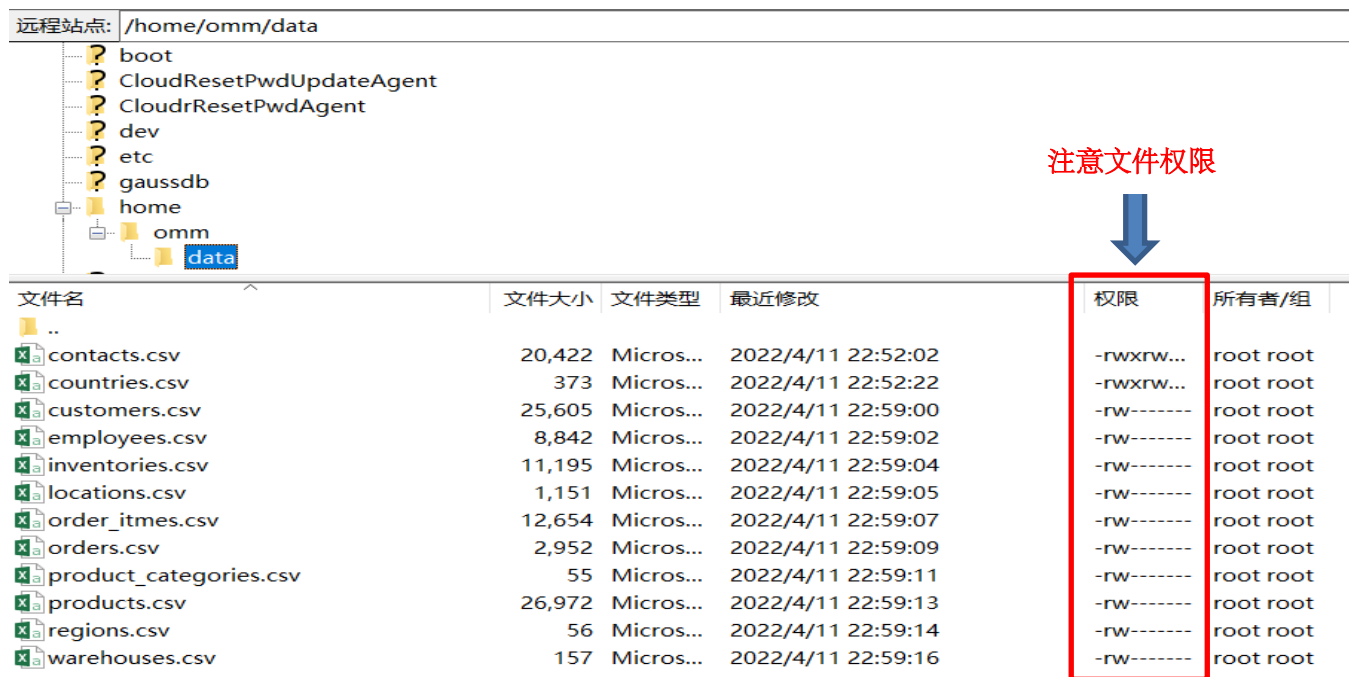
```

chmod: changing permissions of 'contacts.csv': operation not permitted
[omm@ecs-4324 ~]$ ls
contacts.csv  customers.csv  inventories.csv  order_itmes.csv  product_categories.csv  regions.csv
countries.csv  employees.csv  locations.csv  orders.csv  products.csv  warehouses.csv
[omm@ecs-4324 ~]$ chmod 777 orders.csv
chmod: changing permissions of 'orders.csv': Operation not permitted

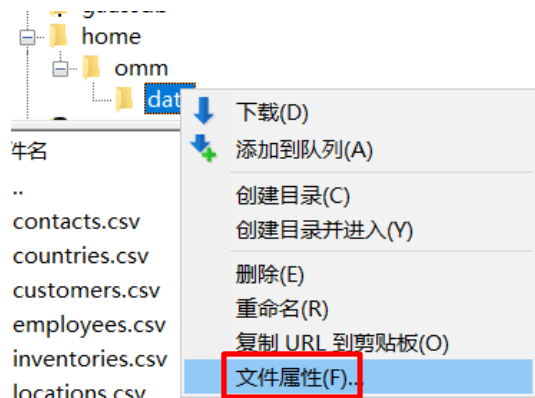
```

方法二、使用 FTP 工具在上传数据文件之前修改其权限为 777。该方法将数据文件上传和权限修改同时完成，效率高，推荐。

远程站点: /home/omm/data




文件名	文件大小	文件类型	最近修改	权限	所有者/组
..					
contacts.csv	20,422	Micros...	2022/4/11 22:52:02	-rwxrw...	root root
countries.csv	373	Micros...	2022/4/11 22:52:22	-rwxrw...	root root
customers.csv	25,605	Micros...	2022/4/11 22:59:00	-rw-----	root root
employees.csv	8,842	Micros...	2022/4/11 22:59:02	-rw-----	root root
inventories.csv	11,195	Micros...	2022/4/11 22:59:04	-rw-----	root root
locations.csv	1,151	Micros...	2022/4/11 22:59:05	-rw-----	root root
order_itmes.csv	12,654	Micros...	2022/4/11 22:59:07	-rw-----	root root
orders.csv	2,952	Micros...	2022/4/11 22:59:09	-rw-----	root root
product_categories.csv	55	Micros...	2022/4/11 22:59:11	-rw-----	root root
products.csv	26,972	Micros...	2022/4/11 22:59:13	-rw-----	root root
regions.csv	56	Micros...	2022/4/11 22:59:14	-rw-----	root root
warehouses.csv	157	Micros...	2022/4/11 22:59:16	-rw-----	root root



选中文件夹 data 并右键在弹出的菜单选择‘文件属性...’



可以分别勾选所有者权限，组权限和公共权限，也可以直接输入数字值 777

远程站点: /home/omm/data						
<div> <div> ? boot ? CloudResetPwdUpdateAgent ? CloudrResetPwdAgent ? dev ? etc ? gaussdb home omm data </div> </div>						
				所有文件的权限都变为 777		
						
文件名	文件大小	文件类型	最近修改	权限	所有者/组	
..						
contacts.csv	20,422	Micros...	2022/4/11 22:52:02	-rwxrw...	root root	
countries.csv	373	Micros...	2022/4/11 22:52:22	-rwxrw...	root root	
customers.csv	25,605	Micros...	2022/4/11 22:59:00	-rwxrw...	root root	
employees.csv	8,842	Micros...	2022/4/11 22:59:02	-rwxrw...	root root	
inventories.csv	11,195	Micros...	2022/4/11 22:59:04	-rwxrw...	root root	
locations.csv	1,151	Micros...	2022/4/11 22:59:05	-rwxrw...	root root	
order_itmes.csv	12,654	Micros...	2022/4/11 22:59:07	-rwxrw...	root root	
orders.csv	2,952	Micros...	2022/4/11 22:59:09	-rwxrw...	root root	
product_categories.csv	55	Micros...	2022/4/11 22:59:11	-rwxrw...	root root	
products.csv	26,972	Micros...	2022/4/11 22:59:13	-rwxrw...	root root	
regions.csv	56	Micros...	2022/4/11 22:59:14	-rwxrw...	root root	
warehouses.csv	157	Micros...	2022/4/11 22:59:16	-rwxrw...	root root	

3.导入数据到 employees 表时出现编码错误

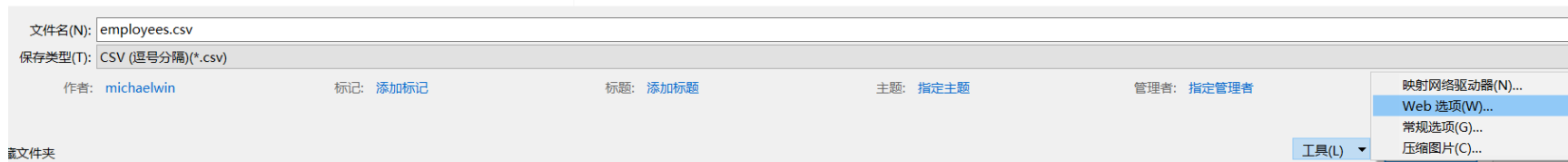
问题描述：使用 copy 命令在导入数据的过程中出现 “invalid byte sequence for encoding 'UTF8' ” 错误

```
demo=> \copy EMPLOYEES from '/home/omm/data/employees.csv' CSV;  
ERROR:  invalid byte sequence for encoding "UTF8": 0xd4 0xc2  
CONTEXT:  COPY employees, line 1
```

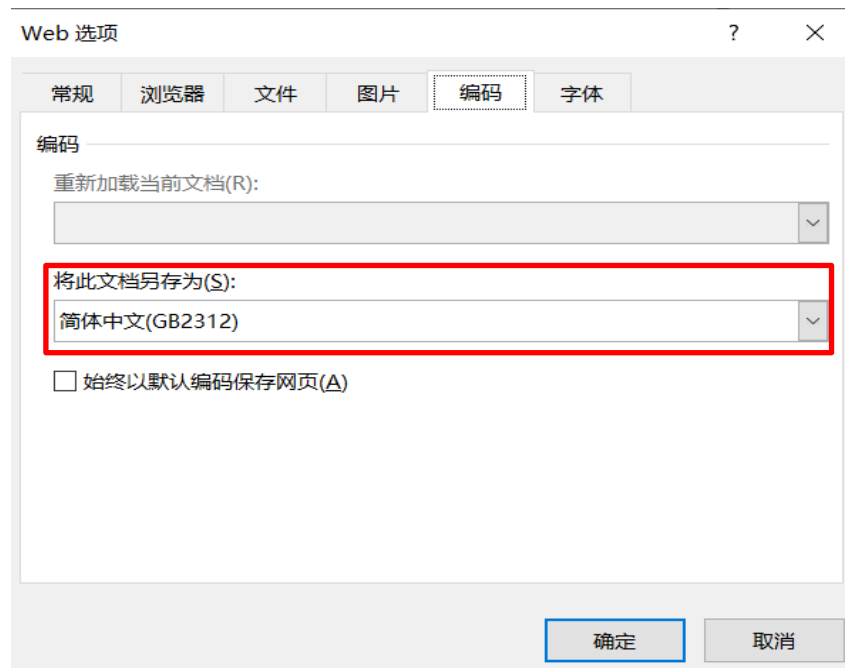
原因：CSV 文件的编码与 openGauss 数据库不一致

解决方法：修改 CSV 文件的编码为 UTF8

双击打开 CSV 文件>另存为>选择存储位置>在打开的页面，进行如下选择：

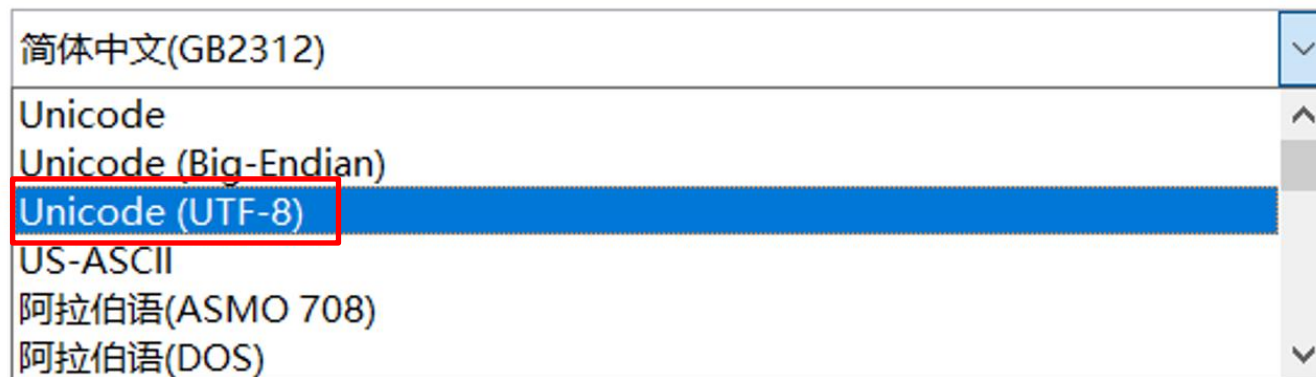


- 保存类型：CSV
- 点开“工具”的下拉三角形，在弹出的菜单选择 “web 选项” 并打开，点击选择 “编码” 标签



点击下拉选项

将此文档另存为(S):



选择 Unicode(UTF-8)，点击确定，保存文件即可，此时的 CSV 文件的编码为 UTF-8。

4.导入数据到 employees 表时出现时间戳不匹配

问题描述：导入数据到 employees 表中时出现 “invalid input syntax for type timestamp...” 错误

```
demo=> \copy employees from '/home/omm/data/employees.csv' CSV;
ERROR:  invalid input syntax for type timestamp: "07-6月 -16"
CONTEXT:  COPY employees, line 1, column hire_date: "07-6月 -16"
```

原因：employees.csv 文件中 hire_date 列的时间戳格式与 employees 表中对列的时间戳不匹配

解决方法：修改 CSV 数据文件中相关列的时间戳格式使之与导入表对应列的时间戳格式一致

以 employees.csv 数据文件为例。打开 employees.csv 文件，找到涉及时间戳的列，设置其与数据库表对应列有相同的时间戳格式。

(1) 在数据库环境下使用 **show** 命令查看日期风格 datestyle

```
demo=> show datestyle;  
DateStyle  
-----  
ISO, MDY  
(1 row)
```

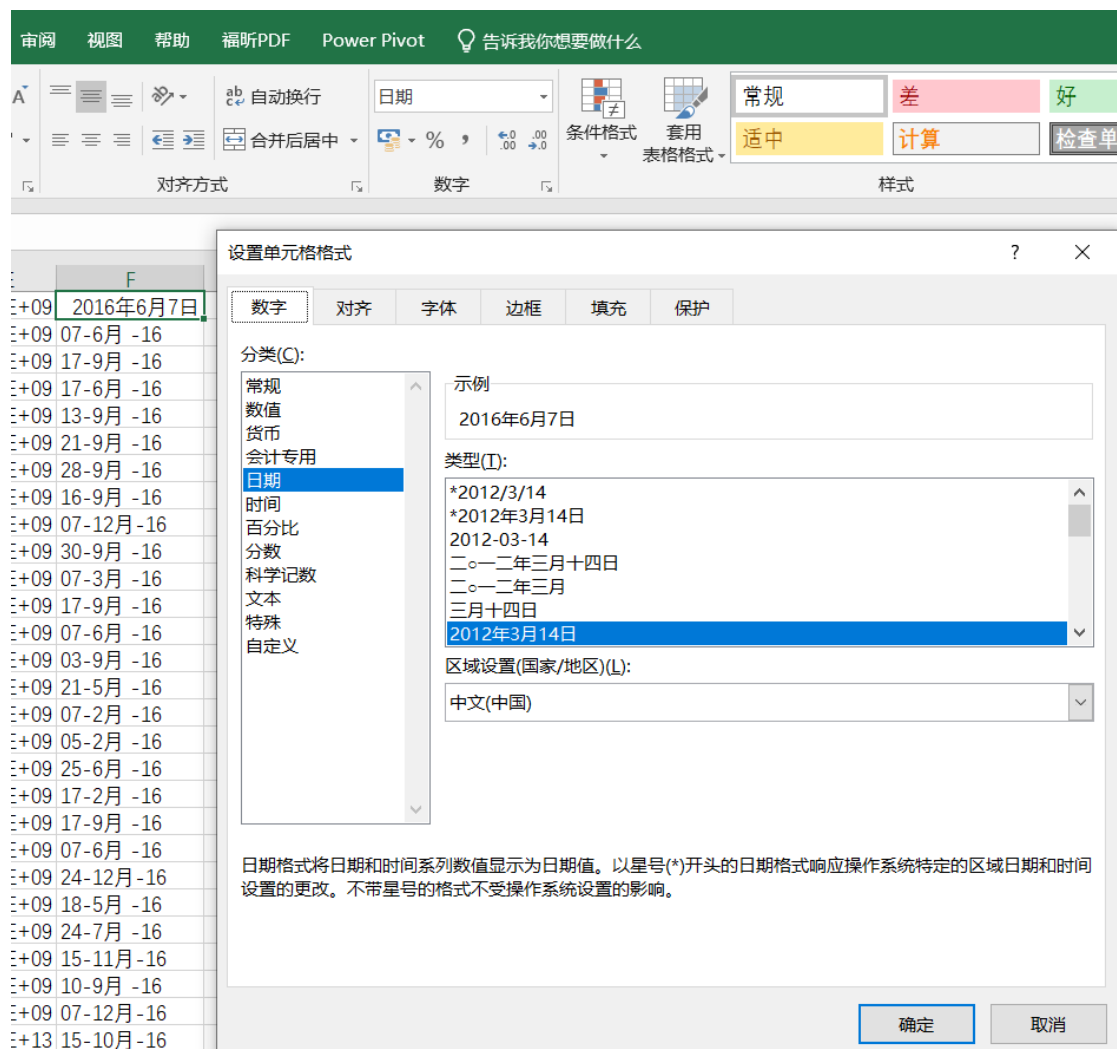
上图说明 openGauss 支持的日期风格有两种：ISO 和 MDY。

ISO 格式：yyyy-mm-dd

MDY 格式：mm/dd/yyyy

两种日期格式的详解和使用可参见：<https://www.modb.pro/db/30390>

(2) 设置 employees.csv 的 hire_date 列的时间戳格式为 ISO。



对 employees.csv 数据文件在修改日期格式之后再使用 COPY 命令导入数据到 employees 表中，可以查到数据量，说明导入成功。

```
demo=> \copy employees from '/home/omm/data/employees.csv' CSV;
demo=> select count(*) from employees;
      count
-----
       107
(1 row)
```

5.使用带逗号作为定界符的 copy 命令导入数据到 customers 表时出现异常

问题描述：copy 命令使用 WITH(delimiter ',', IGNORE_EXTRA_DATA 'on')参数将数据导入 customers 表时出现异常。

```
shiyanyi=> \copy customers FROM '/home/omm/customers.csv' WITH (delimiter ',',IGNORE_EXTRA_DATA 'on');
ERROR:  invalid input syntax for type numeric: " NY"
CONTEXT: COPY customers, line 1, column credit_limit: " NY"
```

177	United Continental Holdings	2904 S Salina St, Syracuse, NY	http://www.unitedcontinentalholdings.com	5000
180	INTL FCStone	5344 Haverford Ave, Philadelphia, PA	http://www.intlfcstone.com	5000
184	Publix Super Markets	1795 Wu Meng, Muang Chonburi,	http://www.publix.com	1200
187	ConocoPhillips	Walpurgisstr 69, Munich,	http://www.conocophillips.com	2400
190	3M	Via Frenzy 6903, Roma,	http://www.3m.com	1200
192	Exelon	Via Luminosa 162, Firenze,	http://www.exeloncorp.com	500
208	Tesoro	Via Notoriosa 1942, Firenze,	http://www.tesoro.com	500
207	Northwestern Mutual	1831 No Wong, Peking,	http://www.northwesternmutual.com	3600
200	Enterprise Products Partners	Via Notoriosa 1949, Firenze,	http://www.enterpriseproducts.com	2400
204	Rite Aid	Piazza Cacchiatore 23, San Gimignano,	http://www.riteaid.com	3600
212	Qualcomm	Piazza Svizzera, Milano,	http://www.qualcomm.com	500
216	EMC	Via Delle Grazie 11, San Gimignano,	http://www.emc.com	700
220	Time Warner Cable	1597 Legend St, Mysore, Kar	http://www.twc.com	3700
223	Northrop Grumman	1606 Sangam Blvd, New Delhi,	http://www.northropgrumman.com	5000
39	Lear	2115 N Towne Ln Ne, Cedar Rapids, IA	http://www.lear.com	500
43	Facebook	5112 Sw 9Th St, Des Moines, IA	http://www.facebook.com	500
46	Supervalu	8989 N Port Washington Rd, Milwaukee, WI	http://www.supervalu.com	500

原因：customers.csv 文件中的 address 列的值包含了逗号，当 copy 命令使用逗号作为定界符时，默认将 NY 自动归到第四列，从而出现异常。

解决方法：把 WITH 和后面括号都去掉，改成大写的 CSV

```

shiyanyi=> \copy customers FROM '/home/omm/customers.csv' CSV
shiyanyi=> select*from customers,

```

customer_id	name	address	website
177	United Continental Holdings	2904 S Salina St, Syracuse, NY	http://wwwunitedcontinent
180	INTL FCStone	5344 Haverford Ave, Philadelphia, PA	http://wwwintlfcstonecom
184	Publix Super Markets	1795 Wu Meng, Muang Chonburi,	http://wwwpublixcom
187	ConocoPhillips	Walpurgisstr 69, Munich,	http://wwwconocophillipsc
190	3M	Via Frenzy 6903, Roma,	http://www3mcom
192	Exelon	Via Luminosa 162, Firenze,	http://wwwexeloncorpcom
208	Tesoro	Via Notoriosa 1942, Firenze,	http://wwwtsocorpcom
207	Northwestern Mutual	1831 No Wong, Peking,	http://wwwnorthwesternmut
200	Enterprise Products Partners	Via Notoriosa 1949, Firenze,	http://wwwenterpriseprodu
204	Rite Aid	Piazza Cacchiatore 23, San Giminiano,	http://wwwriteaidcom
212	Qualcomm	Piazza Svizzera, Milano,	http://wwwqualcommcom

6.数据文件表头问题

问题描述：如果 CSV 文件包含了表头，则在数据导入过程中可能因表头格式不符导致错误。

解决方法：

方法一：将 CSV 文件的表头去掉，但要保证 CSV 文件的各列与表的各列一致。

方法二：保留 CSV 文件的表头，但在导入数据时需要加上相应的参数，即表名第一行是表头，不是数据。

7.openGauss 数据库服务器的集群状态参数 Cluster_state 出现 unavailable

问题描述：cluster_state 是 unavailable

```
[omm@ecs-ed53 ~]$ gs_om -t status
-----
cluster_name      : dbCluster
cluster_state     : Unavailable
redistributing    : No
```

原因：openGauss 数据库未启动

解决方法：进入 omm 后发布启动命令：`gs_om -t start`

注：如果实验完成，一个好的习惯是先关闭数据库以释放相应的资源，如所占的内存空间。数据库关闭命令：`gs_om -t stop`