本小节介绍如何使用[IMPORT](http://doc.cockroachchina.baidu.com/" \l "develop/sql-statements/import/)语句从CSV文件迁移数据到CockroachDB当中。

以下示例从Amazon S3拉取真实的数据，其使用的[employees数据集](https://github.com/datacharmer/test_db)也被使用在[MySQL文档](https://dev.mysql.com/doc/employee/en/)当中，文件格式为CSV格式。

目前不支持导入数据到已经存在的表。

Step 1: 导出CSV格式的数据

根据导出CSV文件的源数据库，查看其相应的文档，导出CSV格式的数据文件。

用户为每张表导出的CSV格式的数据文件，需要满足以下要求：

* 文件必须是[有效的CSV格式](https://tools.ietf.org/html/rfc4180)，且分隔符必须是单个字符。如果使用逗号以外的其他字符（例如Tab），需要使用[delimiter选项](http://doc.cockroachchina.baidu.com/" \l "develop/sql-statements/import#delimiter)设置分隔符。
* 文件必须是UTF-8编码。
* 如果以下字符出现在数据当中，该数据域需要用双引号包裹。
  + 分隔符（默认,）
  + 双引号（"）
  + 换行符（\n）
  + 回车（\r）
* 如果使用双引号包裹数值域，则在域中出现双引号的时候需要在该双引号前使用另外的双引号进行转义，例如："aaa","b""bb","ccc"。
* 如果某一列的数据类型为[BYTES](http://doc.cockroachchina.baidu.com/" \l "develop/data-types/BYTES)，则该列的值可以为一个有效的UTF-8字符串，也可以是一个\x开头的[16进制编码字符](https://www.cockroachlabs.com/docs/stable/sql-constants.html" \l "hexadecimal-encoded-byte-array-literals)。例如：一个值为1，2的BYTES类型数值会输出为\x0102。

Step 2: 上传文件到集群能访问到的位置

CockroachDB集群中的每个节点都需要访问到导入使用的数据文件。有几种方式能够使得集群访问到数据文件。关于[IMPORT](http://doc.cockroachchina.baidu.com/" \l "develop/sql-statements/import)语句支持的存储位置的类型，可以查看[Import File URL](http://doc.cockroachchina.baidu.com/" \l "develop/sql-statements/import#import-file-urls)。

**TIPS:** 我们强烈推荐使用云存储（例如Amazon S3或是Google Cloud）托管需要导入的数据文件。

Step 3: 导入CSV格式文件

用户可以根据情况使用并调整IMPORT TABLE语句来导入数据。

例如，可以使用以下语句从employees.csv导入数据到employees表：

IMPORT TABLE employees (

emp\_no INT PRIMARY KEY,

birth\_date DATE NOT NULL,

first\_name STRING NOT NULL,

last\_name STRING NOT NULL,

gender STRING NOT NULL,

hire\_date DATE NOT NULL

) CSV DATA ('https://s3-us-west-1.amazonaws.com/cockroachdb-movr/datasets/employees-db/csv/employees.csv.gz');

job\_id | status | fraction\_completed | rows | index\_entries | system\_records | bytes

--------------------+-----------+--------------------+--------+---------------+----------------+----------

381866942129111041 | succeeded | 1 | 300024 | 0 | 0 | 13258389

(1 row)

对每个CSV文件通过以上方式导入数据。

在导入完成后可以使用[ALTER TABLE ... ADD CONSTRAINT](http://doc.cockroachchina.baidu.com/" \l "develop/sql-statements/add-constraint)命令添加[外键约束](http://doc.cockroachchina.baidu.com/" \l "develop/constraints/foreign-key)。

配置选项

IMPORT ... CSV语句相关的配置选项有：

* 列分隔符
* 注释语法
* 跳过首行
* 空字符串
* 文件压缩

列分隔符

delimiter选项是用来设置用于标记列结束位置的Unicode字符。默认是,。

示例：

IMPORT TABLE employees (

emp\_no INT PRIMARY KEY,

birth\_date DATE NOT NULL,

first\_name STRING NOT NULL,

last\_name STRING NOT NULL,

gender STRING NOT NULL,

hire\_date DATE NOT NULL

)

CSV DATA ('s3://acme-co/employees.csv?AWS\_ACCESS\_KEY\_ID=123&AWS\_SECRET\_ACCESS\_KEY=456')

WITH delimiter = e'\t';

注释语法

comment选项是用来设置跳过该行所有内容的Unicode字符。

示例：

IMPORT TABLE employees (

emp\_no INT PRIMARY KEY,

birth\_date DATE NOT NULL,

first\_name STRING NOT NULL,

last\_name STRING NOT NULL,

gender STRING NOT NULL,

hire\_date DATE NOT NULL

)

CSV DATA ('s3://acme-co/employees.csv?AWS\_ACCESS\_KEY\_ID=123&AWS\_SECRET\_ACCESS\_KEY=456')

WITH comment = '#';

跳过首行

skip选项是用来设置跳过首部的行数。

示例：

**IMPORT** TABLE employees (

emp\_no INT PRIMARY KEY,

birth\_date DATE NOT NULL,

first\_name STRING NOT NULL,

last\_name STRING NOT NULL,

gender STRING NOT NULL,

hire\_date DATE NOT NULL

)

CSV DATA ('s3://acme-co/employees.csv?AWS\_ACCESS\_KEY\_ID=123&AWS\_SECRET\_ACCESS\_KEY=456')

**WITH** **skip** = '2';

空字符串

nullif选项用来设置将转义为NULL值的字符串。

示例：

IMPORT TABLE employees (

emp\_no INT PRIMARY KEY,

birth\_date DATE NOT NULL,

first\_name STRING NOT NULL,

last\_name STRING NOT NULL,

gender STRING NOT NULL,

hire\_date DATE NOT NULL

)

CSV DATA ('s3://acme-co/employees.csv?AWS\_ACCESS\_KEY\_ID=123&AWS\_SECRET\_ACCESS\_KEY=456')

WITH nullif = '';

文件压缩

compress选项用来设置解压缩CSV文件用的编解码器，可选值包括：

* gzip：使用[gzip](https://en.wikipedia.org/wiki/Gzip)算法解压缩文件
* bzip：使用[bzip](https://en.wikipedia.org/wiki/Bzip2)算法解压缩文件
* none：关闭解压缩
* auto：（默认选项）根据文件后缀推测解压缩形式：.csv代表none，.gz代表gzip，.bz代表bzip。

示例：

**IMPORT** TABLE employees (

emp\_no INT PRIMARY KEY,

birth\_date DATE NOT NULL,

first\_name STRING NOT NULL,

last\_name STRING NOT NULL,

gender STRING NOT NULL,

hire\_date DATE NOT NULL

)

CSV DATA ('s3://acme-co/employees.csv.gz?AWS\_ACCESS\_KEY\_ID=123&AWS\_SECRET\_ACCESS\_KEY=456')

**WITH** compress = 'gzip';