IMPORT语句将表格数据（例如CSV）导入到单个表中。

有关import SQL dumps的详细信息，请参阅[Import Data](https://www.cockroachlabs.com/docs/stable/import-data.html)。

要求

在使用IMPORT之前，你应该准备：

* 导入的表的schema。
* 导入的表数据（例如CSV），最好托管在云存储上。 使用相同地址的所有节点都必须可以访问此位置。 这意味着你无法使用节点本地存储文件。

为便于使用，我们建议使用云存储。如果没有的话，我们还提供了[轻松创建自己的文件服务器的指南](https://www.cockroachlabs.com/docs/stable/(/maintain/create-a-file-server/)

细节

导入目标

导入的表事先不能存在，必须在IMPORT语句中创建。 如果要导入的表已存在，则必须使用DROP TABLE删除它。

一次只能导入一个表。

你可以在IMPORT语句中的表名中指定目标数据库。 如果未在此处指定，则使用SQL会话中的active数据库。

创建表

你的IMPORT语句必须包含一个CREATE TABLE语句（表示你要导入的数据的schema），使用以下方法之一：

* 对包含CREATE TABLE语句的文件的引用
* 内联CREATE TABLE语句

我们还建议你在CREATE TABLE语句中声明所有二级索引，虽然后续也可以添加二级索引，但在导入期间指定它们要快得多。

CSV数据

要导入的表格数据必须是有效的CSV文件，逗号分隔符可以设置为另一个单个字符。 以下几点尤其需要注意是：

* 文件必须采用UTF-8编码。
* 如果字段中出现分隔符（默认,），双引号（"），换行符（\n）或回车符（\r），则该字段必须用双引号括起来。
* 如果使用双引号括起字段，则必须通过在其前面加上另一个双引号来转义出现在字段内的双引号。 例如： "aaa","b""bb","ccc"

CockroachDB特殊要求：

* 如果列的类型为BYTES，则它可以是有效的UTF-8字符串，也可以是以\x开头的十六进制编码的字节文字。 例如，其值应为字节1, 2的字段将写为\x0102。

对象依赖关系

导入表时，必须注意以下规则，因为IMPORT只创建一个不存在的表：

* 导入的表所依赖的对象必须已存在
* 只能在导入完成后创建依赖导入表的对象

可用存储要求

群集中的每个节点都分配了转换后的CSV数据的相等部分，因此必须有足够的临时空间来存储它。 此外，数据作为普通表保留，因此还必须有足够的空间来容纳最终的复制数据。 节点的第一个列出的/默认的store目录必须有足够的可用存储空间来保存其部分数据。

在cockroach start中，如果设置--max-disk-temp-storage，它也必须大于节点将在临时空间中存储的数据部分。

导入文件位置

你可以使用远程云存储（Amazon S3，Google Cloud Platform等）存储要导入的表格数据。 或者，你可以使用从所有节点都可访问的HTTP服务器。

为简单起见，**强烈建议**将云/远程存储用于你要导入的数据。 虽然支持本地文件但是它们必须可以从群集中的所有节点以相同方式访问。

表用户和权限

导入的表被视为新表，因此你必须具有GRANT权限。

性能

在表格数据转换为kv数据期间使用所有节点，这意味着除了提供正常流量之外，IMPORT任务将消耗所有节点的部分CPU和RAM。

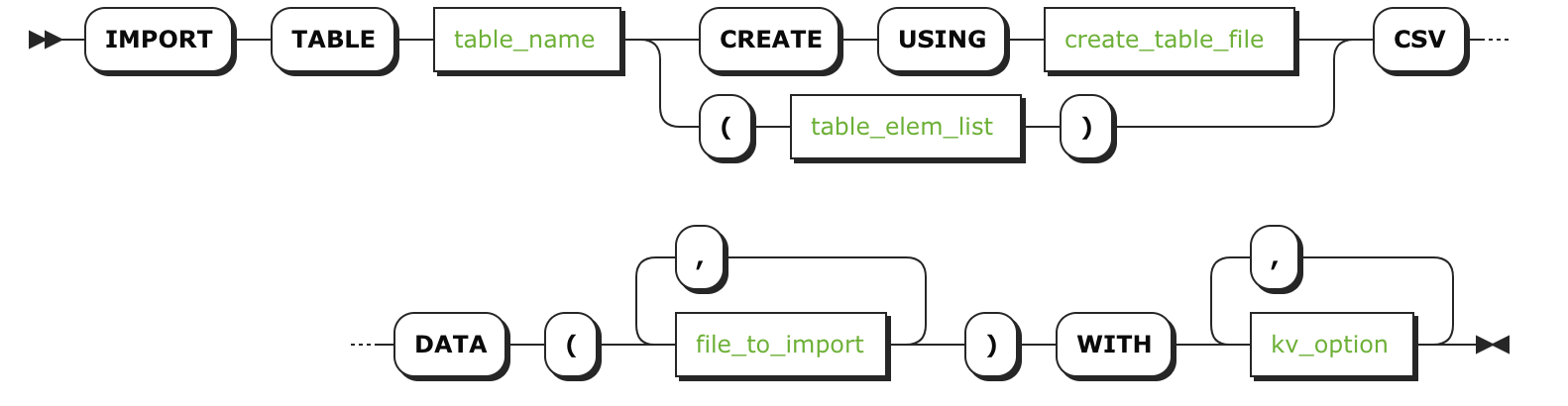
查看和控制导入job

在CockroachDB成功启动导入后，它将导入注册为job，你可以使用SHOW JOBS查看。

导入开始后，你可以使用PAUSE JOB，RESUME JOB和CANCEL JOB来控制它。

暂停然后恢复IMPORT作业将导致它从头开始重新启动

概要



IMPORT语句不能在事务中使用。

所需权限

只有root用户才能运行IMPORT

参数

| **Parameter** | **Description** |
| --- | --- |
| table\_name | 要导入/创建的表的名称。 |
| create\_table\_file | 包含要使用的CREATE TABLE语句的纯文本文件的URL。(有关语法，[请参阅此示例](http://doc.cockroachchina.baidu.com/#develop/sql-statements/IMPORT/#use-create-table-statement-from-a-file)). |
| table\_elem\_list | 要使用的表定义 |
| file\_to\_import | 要导入的文件的URL。 |
| WITH kv\_option | 使用这些选项控制导入行为。 |

导入文件URL

要导入的文件的URL必须使用以下格式：

[scheme]://[host]/[path]?[parameters]

| **Location** | **scheme** | **host** | **parameters** |
| --- | --- | --- | --- |
| Amazon S3 | s3 | Bucket name | AWS\_ACCESS\_KEY\_ID, AWS\_SECRET\_ACCESS\_KEY |
| Azure | azure | Container name | AZURE\_ACCOUNT\_KEY, AZURE\_ACCOUNT\_NAME |
| Google Cloud | gs | Bucket name | AUTH (optional): can be default or implicit |
| HTTP | http | Remote host | N/A |
| NFS/Local | nodelocal | File system location | N/A |
| S3-compatible services | s3 | Bucket name | AWS\_ACCESS\_KEY\_ID, AWS\_SECRET\_ACCESS\_KEY,AWS\_REGION, AWS\_ENDPOINT |

注意事项：

* 如果AUTH参数是implicit的，则所有GCS连接都使用Google的默认身份验证策略。 如果AUTH参数为default，则必须将群集配置cloudstorage.gs.default.key设置为将在身份验证期间使用的服务帐户文件的内容。 如果未指定AUTH参数，则将使用cloudstorage.gs.default.key设置（如果它为非空），否则将使用implicit行为。
* 你可以使用Caddy或nginx轻松创建自己的HTTP服务器。 可以通过设置群集配置cloudstorage.http.custom\_ca将自定义根CA附加到系统的默认CA，该设置将在验证来自HTTPS URL的证书时使用。
* NFS驱动器上的文件系统备份位置和启动节点时设置的--external-io-dir标志指定的路径相关。 如果该标志设置为disabled，则禁用从本地目录和NFS驱动器导入。
* 可以通过设置群集配置cloudstorage.http.custom\_ca将自定义根CA附加到系统的默认CA，该设置将在验证来自S3兼容服务的证书时使用。
* location参数通常包含需要进行URI编码的特殊字符。 使用Javascript的encodeURIComponent函数或Go语言的url.QueryEscape函数对参数进行URI编码。 其他语言也提供与URI编码特殊字符类似的功能。

导入选项

你可以使用以下任何kv对作为kv\_option来控制IMPORT的行为。

delimiter

如果不使用逗号作为列分隔符，则可以指定另一个Unicode字符作为分隔符。

|  |  |
| --- | --- |
| **是否必填？** | No |
| **Key** | delimiter |
| **Value** | 用于分隔行中列的unicode字符 |
| **Example** | 要使用制表符分隔的值： WITH delimiter = e'\ t' |

comment

不要导入以此字符开头的行。

|  |  |
| --- | --- |
| **是否必填？** | No |
| **Key** | comment |
| **Value** | 标识要跳过的行的unicode字符 |
| **Example** | WITH comment = '#' |

nullif

如果值与指定的字符串匹配，则将值转换为SQL *NULL*。

|  |  |
| --- | --- |
| **是否必填？** | No |
| **Key** | nullif |
| **Value** | 字符串应转换为*NULL* |
| **Example** | 将空列用作*NULL ： WITH nullif = ' '* |

示例

使用文件中的创建表语句

> IMPORT TABLE customers

**CREATE** **USING** 'azure://acme-co/customer-create-table.sql?AZURE\_ACCOUNT\_KEY=hash&AZURE\_ACCOUNT\_NAME=acme-co'

CSV **DATA** ('azure://acme-co/customer-import-data.csv?AZURE\_ACCOUNT\_KEY=hash&AZURE\_ACCOUNT\_NAME=acme-co')

;

使用Statement中的Create Table Statement

> IMPORT TABLE customers (

id UUID PRIMARY KEY DEFAULT gen\_random\_uuid(),

name TEXT,

INDEX name\_idx (name)

)

CSV DATA ('azure://acme-co/customer-import-data.csv?AZURE\_ACCOUNT\_KEY=hash&AZURE\_ACCOUNT\_NAME=acme-co')

;

导入制表符分隔文件

> IMPORT TABLE customers (

id UUID PRIMARY KEY DEFAULT gen\_random\_uuid(),

name TEXT,

INDEX name\_idx (name)

)

CSV DATA ('azure://acme-co/customer-import-data.tsc?AZURE\_ACCOUNT\_KEY=hash&AZURE\_ACCOUNT\_NAME=acme-co')

WITH

delimiter = e'\t'

;

跳过Commented行

> IMPORT TABLE customers (

id UUID PRIMARY KEY DEFAULT gen\_random\_uuid(),

name TEXT,

INDEX name\_idx (name)

)

CSV DATA ('azure://acme-co/customer-import-data.tsc?AZURE\_ACCOUNT\_KEY=hash&AZURE\_ACCOUNT\_NAME=acme-co')

WITH

comment = '#'

;

使用空白字符作为*NULL*

> IMPORT TABLE customers (

id UUID PRIMARY KEY DEFAULT gen\_random\_uuid(),

name TEXT,

INDEX name\_idx (name)

)

CSV DATA ('azure://acme-co/customer-import-data.tsc?AZURE\_ACCOUNT\_KEY=hash&AZURE\_ACCOUNT\_NAME=acme-co')

WITH

nullif = ''

;

已知限制

IMPORT有时会因“context canceled”错误而失败，或者多次重启但从不会完成。 如果发生这种情况，可能是由于磁盘冲突很大。 这可以通过将群集配置kv.bulk\_io\_write.max\_rate设置为低于最大磁盘写入速度的值来缓解。 例如，要将其设置为10MB/s，请执行：

> **SET** CLUSTER SETTING kv.bulk\_io\_write.max\_rate = '10MB';

See Also

* [Create a File Server](http://doc.cockroachchina.baidu.com/#maintain/create-a-file-server/)
* [Importing Data](http://doc.cockroachchina.baidu.com/#maintain/import-data/)