本小节将说明如何使用[IMPORT](http://doc.cockroachchina.baidu.com/" \l "develop/sql-statements/import)功能读取[pg\_dump](https://www.postgresql.org/docs/current/static/app-pgdump.html)生成文件，从Postgres迁移数据到CockroachDB。

以下示例将从[Amazon S3](https://aws.amazon.com/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。该数据将通过[pgloader](https://pgloader.io/)导入Postgres中，然后按照下述操作进行调整。

Step 1: 导出Postgres数据库

从Postgres数据库中导出数据（用于导入CockroachDB）有几种方式：

* 导出整个数据库
* 一次只导出单张表的数据

如果导出文件包含函数或是类型定义，则导入CockroachDB会失败。除了使用[pg\_dump](https://www.postgresql.org/docs/current/static/app-pgdump.html)，你可能还需要编辑导出文件，删除函数和数据类型相关的内容。

此外，CockroachDB的[IMPORT](http://doc.cockroachchina.baidu.com/" \l "develop/sql-statements/import)不支持从Postgres非公开的[Schemas](https://www.postgresql.org/docs/current/static/ddl-schemas.html)自动导入数据。解决方式是编辑导出文件，修改[CREATE TABLE](http://doc.cockroachchina.baidu.com/" \l "develop/sql-statements/create-table)语句中相关的表和Schema名称。

导出整个数据库

绝大多数用户需要一次性导出整个Postgres数据库，可以使用[pg\_dump](https://www.postgresql.org/docs/current/static/app-pgdump.html)命令导出整个数据库，如下所示：

pg\_dump employees > /tmp/employees-full.sql

具体到该数据集，需要对Postgres导出文件进行以下修改：

* [CREATE TABLE](http://doc.cockroachchina.baidu.com/#develop/sql-statements/create-table)语句中employees.gender列的数据类型需要从employees.employees\_gender变成[STRING](http://doc.cockroachchina.baidu.com/" \l "develop/data-types/STRING)。因为在Postgres中使用[CREATE TYPE](https://www.postgresql.org/docs/10/static/sql-createtype.html)语句创建的employees.employees\_gender类型，在CockroachDB中是不支持的。
* CREATE TYPE employee ...语句需要移除。

一次只导出单张表的数据

从Postgres数据库中导出employees表的数据，可以使用[pg\_dump](https://www.postgresql.org/docs/current/static/app-pgdump.html)工具执行以下命令：

pg\_dump -t employees employees > /tmp/employees.sql

具体到该数据集，需要对Postgres导出的文件进行以下修改：

* [CREATE TABLE](http://doc.cockroachchina.baidu.com/#develop/sql-statements/create-table)语句中employees.gender列的数据类型需要从employees.employees\_gender变成[STRING](http://doc.cockroachchina.baidu.com/" \l "develop/data-types/STRING)。因为在Postgres中使用[CREATE TYPE](https://www.postgresql.org/docs/10/static/sql-createtype.html)语句创建的employees.employees\_gender类型，在CockroachDB中是不支持的。

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Postgres导出文件

用户可以根据情况（导入整个数据库或是导入单张表）使用并调整[IMPROT](http://doc.cockroachchina.baidu.com/" \l "develop/sql-statements/import)语句来导入数据。

* 导入整个数据库的导出文件
* 导入整个数据库导出文件中的单张表
* 导入单张表的导出文件

本小节使用的[IMPORT](http://doc.cockroachchina.baidu.com/" \l "develop/sql-statements/import)语句将从[Amazon S3](https://aws.amazon.com/s3/)中拉取真实数据，开启后台导入任务。用户可以通过[SHOW JOBS](http://doc.cockroachchina.baidu.com/" \l "develop/sql-statements/show-jobs)查看该任务。

导入整个数据库的导出文件

假使用户导出了整个数据库，可以通过以下的[IMPORT](http://doc.cockroachchina.baidu.com/" \l "develop/sql-statements/import)语句从导出文件中读取全部数据和[DDL](https://en.wikipedia.org/wiki/Data_definition_language)语句（包含存在的外键关系）。

IMPORT PGDUMP 'https://s3-us-west-1.amazonaws.com/cockroachdb-movr/datasets/employees-db/pg\_dump/employees-full.sql.gz';

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

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

381845110403104769 | succeeded | 1 | 300024 | 0 | 0 | 11534293

(1 row)

导入整个数据库导出文件中的单张表

假使用户导出了整个数据库，可以通过以下的[IMPORT](http://doc.cockroachchina.baidu.com/" \l "develop/sql-statements/import)语句从导出文件中读取某个表的数据和其[CREATE TABLE](http://doc.cockroachchina.baidu.com/" \l "develop/sql-statements/create-table)语句：

**CREATE** **DATABASE** **IF** **NOT** **EXISTS** employees;

**USE** employees;

IMPORT TABLE employees FROM PGDUMP 'https://s3-us-west-1.amazonaws.com/cockroachdb-movr/datasets/employees-db/pg\_dump/employees-full.sql.gz';

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

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

383839294913871873 | succeeded | 1 | 300024 | 0 | 0 | 11534293

(1 row)

导入单张表的导出文件

假使用户导出了单张表，最简单的导入方式是通过以下的[IMPORT TABLE](http://doc.cockroachchina.baidu.com/" \l "develop/sql-statements/import)语句从导出文件中读取该文件的数据和[CREATE TABLE](http://doc.cockroachchina.baidu.com/" \l "develop/sql-statements/create-table)语句：

**CREATE** **DATABASE** **IF** **NOT** **EXISTS** employees;

**USE** employees;

IMPORT PGDUMP 'https://s3-us-west-1.amazonaws.com/cockroachdb-movr/datasets/employees-db/pg\_dump/employees.sql.gz';

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

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

383855569817436161 | succeeded | 1 | 300024 | 0 | 0 | 11534293

(1 row)

如果因为一些原因需要指定为表的某些列，用户可以使用以下的[IMPORT TABLE](http://doc.cockroachchina.baidu.com/" \l "develop/sql-statements/import)语句，它将忽略文件中的[CREATE TABLE](http://doc.cockroachchina.baidu.com/" \l "develop/sql-statements/create-table)语句，选择使用用户指定的列创建表。

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

)

PGDUMP DATA ('https://s3-us-west-1.amazonaws.com/cockroachdb-movr/datasets/employees-db/pg\_dump/employees.sql.gz');

配置选项

IMPORT ... PGDUMP语句相关的配置选项有：

* 最大行大小
* 忽略外键

最大行大小

max\_row\_size选项可以用来调整行大小的限制，**默认是0.5MB**。如果导出的文件中包含特别长的行，则需要调整该选项：

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

)

PGDUMP DATA ('s3://your-external-storage/employees.sql?AWS\_ACCESS\_KEY\_ID=123&AWS\_SECRET\_ACCESS\_KEY=456') WITH max\_row\_size = '5MB';

忽略外键

默认情况下，IMPORT ... PGDUMP支持外键，即**该配置选项的默认值为false**。使用skip\_foreign\_keys选项能够忽略导出文件中DDL语句的外键约束，加快导入的速度。在导入某张表的时候，该选项也能够帮助用户避免可能的、因为引用其他表所导致的错误。

IMPORT TABLE employees (

emp\_no INTEGER NOT NULL,

birth\_date DATE NOT NULL,

first\_name STRING NOT NULL,

last\_name STRING NOT NULL,

gender STRING NOT NULL,

hire\_date DATE NOT NULL

) PGDUMP DATA ('s3://your-external-storage/employees.sql?AWS\_ACCESS\_KEY\_ID=123&AWS\_SECRET\_ACCESS\_KEY=456') WITH skip\_foreign\_keys;

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