PostgreSQL - Import

Freitag, 11. Juli 2025 10:29

Introduction

+ you need to create an empty table first
+ you need to create all columns before importing
+ and then you can import the .csv file
+ best practice - put it into the sql_files folder

Start PostgreSQL in Terminal

psql -U postgres

+ After using *"CleanMyMac"*
+ Use **"psql postgres"**

Jump straight into the right Database

psql -U postgres -d deine_datenbank

Jump separately into the Database

\c deine_datenbank

Quit PSQL Session

\q

Show Tables

\dt

Show Database Information:

\l

Show Columns Of a Table

\d table_name

Use Queries

SELECT * FROM deine_tabelle;

Delete Table

DROP TABLE table_name;

Make a Table Empty

TRUNCATE TABLE table_name;

Create a new Dataset

CREATE DATABASE dataset_name;

Change Column Name

ALTER TABLE tabellenname
RENAME COLUMN alter_spaltenname TO neuer_spaltenname;

Pretty Print Modus

\x
+take a free row
+ enter: =""""" & A1 & """" TEXT,"

+
copy all header

Spreadsheet Column Copy

+ paste them right under to have the right count
+ copy/paste our formula until the exact length of all header rows
+ take another free row
+ enter: =TEXTJOIN("," & CHAR(10), TRUE, WX:YZ)
remember: in excel we have to activate "wrap text" -> also remember : Copy & Paste as Value in a new cell !!!
+ continue with *"How to Create a Table"*

How to Create a Table

```
CREATE TABLE your_table_name (
   "column1"     TEXT,
   "column2"     TEXT,
   "column3" TEXT,
    ....
);
```

How to Import a Table

```
COPY your_table_name
FROM '/Users/patrick/sql_files/customer_data.csv'
DELIMITER ',' CSV HEADER;
```

Use \copy in terminal !!!

Data Types Comparison!

Bedeutung	BigQuery Typ	PostgreSQL Typ	Kurzbeschreibung
Ganzzahl	INT64	BIGINT	64-Bit ganze Zahl
Ganzzahl klein	INT32 (INTEGER)	INTEGER	32-Bit ganze Zahl
Gleitkommazahl	FLOAT64	DOUBLE PRECISION	64-Bit Fließkommazahl
Gleitkommazahl klein	FLOAT32	REAL	32-Bit Fließkommazahl
Zeichenkette (lang)	STRING	TEXT	Variable Länge Text

Datei a

Zeichenkette (begr.)	STRING / VARCHAR	VARCHAR(n)	Begrenzte Länge Text
Wahrheitswert	BOOL	BOOLEAN	Wahr/Falsch
Datum & Zeit (ohne TZ)	DATETIME	TIMESTAMP WITHOUT TIME ZONE	Datum und Zeit ohne Zeitzone
Datum & Zeit (mit TZ)	TIMESTAMP	TIMESTAMPTZ (TIMESTAMP WITH TIME ZONE)	Datum und Zeit mit Zeitzone
Datum	DATE	DATE	Nur Datum (Jahr, Monat, Tag)
Zeit	TIME	TIME	Nur Uhrzeit (ohne Datum)
Bytes / Binärdaten	BYTES	BYTEA	Binäre Daten
JSON	JSON	JSON / JSONB	JSON-Daten (JSONB ist binary-optimiert)
Numerisch (präzise)	NUMERIC	NUMERIC / DECIMAL	Dezimal mit definierter Genauigkeit

Cases - When Syntax Form Matters!

🔍 1️⃣ Einfache Anführungszeichen '...' → Für Strings / Textwerte in SQL

Wenn du einen **Text-Wert (string literal)** schreibst, musst du ihn in **einfache Anführungszeichen** ' setzen.

✅ **Beispiele:**

```sql
SELECT * FROM kunden WHERE name = 'Peter';
SELECT 'Hallo Welt';
```

Merke: ' ist **immer für Daten / Inhalte**, nicht für Tabellennamen oder Spalten!

🔍 2️⃣ Doppelte Anführungszeichen "..." → Für Bezeichner (Identifier) wie Tabellen-, Spaltennamen, wenn...

- ... der Name **Großbuchstaben** enthält.
- ... der Name **Leerzeichen oder Sonderzeichen** enthält.
- ... der Name **gleich einem SQL-Schlüsselwort** ist (z.B. `user`).

Ohne Anführungszeichen behandelt PostgreSQL alles als **kleingeschrieben**.

✅ **Beispiele:**

```sql
SELECT "KundenName" FROM "Bestell-Tabelle";
SELECT "user" FROM accounts;
```

Ohne ":

```sql
SELECT kundenname FROM bestell_tabelle;
```

PostgreSQL macht automatisch kundenname klein.

End

SQL for Cleaning Data 1.0

Freitag, 11. Juli 2025 10:29

Spreadsheets VS SQL Databases

Features of Spreadsheets	Features of SQL Databases
Smaller data sets	Larger datasets
Enter data manually	Access tables across a database
Create graphs and visualizations in the same program	Prepare data for further analysis in another software
Built-in spell check and other useful functions	Fast and powerful functionality
Best when working solo on a project	Great for collaborative work and tracking queries run by all users

Putting 2 Databases Together

L> INSERT INTO `database-pathway`
L--> (column4, column3, column2, column1)
L> VALUES
L-->(2645, `Rachel DeSantos`, `33 SQL Road`, `Jackson`, `MI`, 49202, `US`)

Update The Databases

L> UPDATE `database-pathway`
L> SET address = `123 New Address`
L> WHERE column = 2645

Cleaning Duplicates after Moving

L> DROP TABLE IF EXISTS

Removing Duplicates

L> add a DISTINCT before the column
L> e.g. select DISTINCT customer_id

Defining a length of a row-result

L> using LENGTH or LEN
L--> SELECT LENGTH (column)
L--> it will only show you the lengths of the row
L> put LENGTH into "WHERE"
L--> WHERE LENGTH(country) > 2
L--> now e have what we want

Correct Error using "Substring"

L> can filter/specify our search
L> WHERE SUBSTR(column, which letter to start, how many letters) = X
L--> e.g. WHERE SUBSTR(country, 1, 2) = 'US'

Removing Extra Spaces

L> using the TRIM function
L--> e.g. WHERE LENGTH(state) > 2
L--> shows a character with 2 letters
L-->there must be a space extra
L> we need UPDATE - SET - WHERE
L> SET column = TRIM(column)
L> WHERE True; **(?)**

Finding "Nulls"

L> helping you to find empty fields
L> WHERE column IS NULL
L> e.g. WHERE car_id IS NULL
L> *and we can correct it ourselves:*

Editing Rows

L> with UPDATE *Path*, SET *column=result*, *WHERE column=parameter*
L> Example

```
UPDATE
  your project name.cars.car_info
SET
  num_of_doors = "four"
WHERE
  make = "dodge"
  AND fuel_type = "gas"
  AND body_style = "sedan";
```

Deleting Rows

```
DELETE your project name.cars.car_info
WHERE compression_ratio = 70;
```

Datei auswählen Keine ausgewählt

End

SQL For Cleaning Data 2.0

Freitag, 11. Juli 2025 10:29

BigQuery	PostgreSQL
FLOAT64	DOUBLE PRECISION
INT64	BIGINT
STRING	TEXT oder VARCHAR
DATETIME	TIMESTAMP

Float
L> a data type
L> a number that contains a decimal

Typecasting
L> converting data from one type to another

The Function ORDER BY
L> not putting it anywhere under
L> it's like SELECT, FROM, WHERE
L> ORDER BY column theorder
L> e.g. ORDER BY purchase_price DSC

ORDER BY Fails - Wrong Data Type
L> we have to change the Data Type
L> because they are sorted by letters

The Function CAST
L> convert a column into another data type
L> we're changing it for our illustration
L> we're not editing the data with it

```
SELECT CAST(purchase_price AS DOUBLE PRECISION)
FROM laurens_furniture_shop
ORDER BY CAST(purchase_price AS DOUBLE PRECISION) DESC;
```

Edit The Data Type
L> change it long lasting
L> after no need to use cast anymore
L> if you don't want to use ::
L> you can use CAST()

```
ALTER TABLE laurens_furniture_shop
ALTER COLUMN date TYPE DATE USING CAST(date AS DATE);
```

```
ALTER TABLE laurens_furniture_shop
ALTER COLUMN purchase_price TYPE DOUBLE PRECISION USING purchase_price::DOUBLE PRECISION;
```

Add Row-Number to PSQL
L> SELECT ROW_NUMBER() OVER() AS rows,

The BETWEEN Function
L> WHERE column BETWEEN '2020-12-01' AND '2020-12-31'

The CONCAT Function
L> add strings together
L> to create new text strings
L> can be used as unique keys
L> for example you have ID's and colours
L> but you don't have something unique for colours, right?
L> so you can not sort and find out the colour preference
L> CONDACT (column1, column2) AS choose_name
L> in example we filtered WHERE product = 'couch'

The COALESCE Function
L> used to return non-null values in a list
L> if we have optional fields, which can be a "Null"
L> we can tell SQL - if THIS column is a "Null", choose THAT column
L> SELECT COALESCE(column1, column2)

Datei auswählen Keine ausgewählt

Glossary

Freitag, 11. Juli 2025 10:29

Terms and definitions for Course 4, Module 3

CAST:

A SQL function that converts data from one datatype to another

COALESCE:

A SQL function that returns non-null values in a list

CONCAT:

A SQL function that adds strings together to create new text strings that can be used as unique keys

DISTINCT:

A keyword that is added to a SQL SELECT statement to retrieve only non-duplicate entries

Float:

A number that contains a decimal

Substring:

A subset of a text string

Typecasting:

Converting data from one type to another

Datei auswählen Keine ausgewählt