# PostgreSQL As Data Integration Tool

PostgreSQL Berlin User Group 2025-06-19 Stefanie Janine Stölting





#### **About Me**



Owner of Tallinn/Estonia based ProOpenSource OÜ

Website: www.ProOpenSource.eu

Blog: www.ProOpenSource.it

Mastodon: https://digitalcourage.social/@sjstoelting

Telegram: https://t.me/postgreschat



#### **About Me**



#### Member of the PGEU Diversity Committee





#### Data Integration Tools



- Azure Data Factory
- Dell Boomi
- Informatica
- Keboola
- Pentaho Data Integration
- Qlik
- Snaplogic
- Talend Open Studio



#### SQL/MED



First defined 2008 in ISO/IEC 9075-9:2008
Revised 2016 in ISO/IEC 9075-9:2016
Supported by:

- DB2
- MariaDB
   With the CONNECT Storage Engine,
   the implementation differs from the standard
- PostgreSQL



# PostgreSQL SQL/MED Implementation







# Implementation in PostgreSQL



#### Foreign Data Wrapper

- Read Only
- Read and Write

The installation is realized with Extensions.



#### Available FDW



#### **Examples:**

Oracle (pgxn.org)

MS SQL Server / Sybase ASE read-only (pgxn.org)

MongoDB (GitHub)

MariaDB / MySQL (pgxn.org)

Please use the GitHub Repository, the extension on PGXN is old.

SQLite (GitHub)

The version on PGXN hasn't been updated for a while.

There is another one (read-only) on GitHub, which is not further developed.

Hadoop (HDFS) read-only (GitHub)

ODBC (GitHub)

There is another one on PGXN, which is not further developed.

Apache Kafka read-only (GitHub)



#### Special FDW



```
file_fdw
postgres fdw
```

foreign\_table\_exposer

Some BI sysems have problems with foreign tables, this extentsion solves that problem.



# Write Your Own FDW



#### Multicorn 2

With Python and Multicorn one can create foreign data wrapper and access several sources, for example:

- IMAP
- HTML



#### Chinook



The example data used in this talk are available by the Chinook Database:

- PostgreSQL
- CSV
- SQLite



#### **Chinook Tables**



	<b>T</b> tablename					
1	Artist					
2	Invoice					
3	Employee					
4	Customer					
5	Playlist					
6	InvoiceLine					
7	Album					
8	Genre					
9	PlaylistTrack					
10	MediaType					
11	Track					

	I table_name ↔	■ column_name	■ data_type		
1	Artist	ArtistId	integer		
2	Artist	Name	character varying (120)		

	I table_name ↔	▼ column_name	T data_type ♣	
1	Album	AlbumId	integer	
2	Album	Title	character varying (160)	
3	Album	ArtistId	integer	

	📅 GenreId 🍫	🗷 Name 🖖
1	1	Rock
2	2	Jazz
3	3	Metal
4	4	Alternative & Punk
5	5	Rock And Roll
6	6	Blues
7	7	Latin
8	8	Reggae
9	9	Pop
10	10	Soundtrack
11	11	Bossa Nova
12	12	Easy Listening
13	13	Heavy Metal
14	14	R&B/Soul
15	15	Electronica/Dance

	T table_name ↔	🔳 column_name 🍫	፲ data_type	
1	Track	TrackId	integer	
2	Track	Name	character varying (200)	
3	Track	AlbumId	integer	
4	Track	MediaTypeId	integer	
5	Track	GenreId	integer	
6	Track	Composer	character varying (220)	
7	Track	Milliseconds	integer	
8	Track	Bytes	integer	
9	Track	UnitPrice	numeric	



#### **Show Time**



#### Slides are boring.







-- Create the SQLite foreign data wrapper extension in the current database CREATE EXTENSION sqlite\_fdw;









-- Create a schema for the SQLite tables
CREATE SCHEMA IF NOT EXISTS sqlite;





```
-- Create a foreing table pointing to the SQLite database
-- The column definition must be correct to the one in SQLite
CREATE FOREIGN TABLE sqlite.artist(
     "ArtistId" integer,
     "Name" text
)
SERVER sqlite_server
OPTIONS(
    table 'Artist'
);
```





```
-- Query data
SELECT * FROM sqlite.artist;
```

0	123 ArtistId	A-z Name
1	1	AC/DC
2	2	Accept
3	3	Aerosmith
4	4	Alanis Morissette
5	5	Alice In Chains
6	6	Antônio Carlos Jobim
7	7	Apocalyptica
8	8	Audioslave
9	9	BackBeat





```
-- Join SQLite with PostgreSQL 17
SELECT artist."Name" AS album_name
    , album.title
FROM sqlite.artist AS artist
INNER JOIN public.album
    ON artist."ArtistId" = album.artist_id;
```

0	A-z album_name	A-Z title ▼
1	AC/DC	Let There Be Rock
2	AC/DC	For Those About To Rock We Salute You
3	Accept	Restless and Wild
4	Accept	Balls to the Wall
5	Aerosmith	Big Ones
6	Alanis Morissette	Jagged Little Pill
7	Alice In Chains	Facelift
8	Antônio Carlos Jobim	Chill: Brazil (Disc 2)





-- Create the PostgreSQL foreign data wrapper
CREATE EXTENSION postgres\_fdw;













-- Create a schema for the foreign PostgreSQL tables
CREATE SCHEMA IF NOT EXISTS pg14;





```
-- Create the foreign table
IMPORT FOREIGN SCHEMA public LIMIT TO("Track")
FROM SERVER pg_localhost_chinook
INTO pg14
;
```





-- Show data from PostgreSQL 14
SELECT \* FROM pg14."Track";

0	123 TrackId	*	A-z Name ▼	123 AlbumId		123 MediaTypeId	-	123 GenreId	-	A-Z Composer
1		1	For Those About To Rock (We Salute You)		1		1		1	Angus Young, Malcolm Young, Brian Johnson
2		2	Balls to the Wall		2		2		1	[NULL]
3		3	Fast As a Shark		3		2		1	F. Baltes, S. Kaufman, U. Dirkscneider & W. Hoffman
4		4	Restless and Wild		3		2		1	F. Baltes, R.A. Smith-Diesel, S. Kaufman, U. Dirkscneider & W. Hoffman
5		5	Princess of the Dawn		3		2		1	Deaffy & R.A. Smith-Diesel
6		7	Let's Get It Up		1		1		1	Angus Young, Malcolm Young, Brian Johnson
7		8	Inject The Venom		1		1		1	Angus Young, Malcolm Young, Brian Johnson
8		9	Snowballed		1		1		1	Angus Young, Malcolm Young, Brian Johnson
9	,	10	Evil Walks		1		1		1	Angus Young, Malcolm Young, Brian Johnson





```
-- Join SQLite and PostgreSQL tables
SELECT artist."Name"
    , album.title
    , track."Name"
FROM sqlite.artist AS artist
INNER JOIN public.album
    ON artist."ArtistId" = album.artist_id
INNER JOIN pg14."Track" AS track
    ON album.album_id = track."AlbumId"
;
```

0	A-z Name	Az title 🔻	A-Z Name
1	AC/DC	Let There Be Rock	Whole Lotta Rosie
2	AC/DC	Let There Be Rock	Hell Ain't A Bad Place To Be
3	AC/DC	Let There Be Rock	Overdose
4	AC/DC	Let There Be Rock	Problem Child
5	AC/DC	Let There Be Rock	Bad Boy Boogie
6	AC/DC	Let There Be Rock	Let There Be Rock
7	AC/DC	Let There Be Rock	Dog Eat Dog
8	AC/DC	Let There Be Rock	Go Down





```
-- Update data in a foreign table
UPDATE pg14."Track" SET
    "Name" = 'Der Track Name ist geändert!'
WHERE "TrackId" = 6;
```





```
-- Show the changed data
SELECT *
FROM pg14."Track"
WHERE "TrackId" = 6;
```

0	123 TrackId	A-Z Name ▼	123 AlbumId	123 MediaTypeId	•	123 GenreId	•	A-z Composer 🔻
1	6	Der Track Name ist geändert!	1		1	500	1	Angus Young, Malcolm Young, Brian Johnson





```
-- Update data in a foreign table
UPDATE pg14."Track" SET
"Name" = 'Der Track Name ist geändert!'
WHERE "TrackId" = 6;
```





```
-- Show the changed data
SELECT *
FROM pg14."Track"
WHERE "TrackId" = 6;
```





```
-- Update data in a foreign table
UPDATE pg14."Track" SET
"Name" = 'Put The Finger On You'
WHERE "TrackId" = 6;
```





```
-- Show the changed data
SELECT *
FROM pg14."Track"
WHERE "TrackId" = 6;
```





-- Create the file extension
CREATE EXTENSION file\_fdw;





```
-- It does need a server object, that can be used forall CSV files
-- The user postgres must have read and write access
-- not only on the file but also on the folder
CREATE SERVER chinook_csv
FOREIGN DATA WRAPPER file_fdw
;
```





-- Create a schema for the CSV files
CREATE SCHEMA IF NOT EXISTS CSV;





```
-- Create a foreign table based on a CSV file
-- The options are the same as of the COPY command
CREATE FOREIGN TABLE csv.genre (
    "GenreId" integer,
    "Name" text
) SERVER chinook_csv
OPTIONS (
    filename '/var/sqlite/Genre.csv',
    format 'csv',
    HEADER 'true'
);
```





-- Show data
SELECT \* FROM csv.genre;

0	123 GenreId 💌	A-Z Name 🔻
1	1	Rock
2	2	Jazz
3	3	Metal
4	4	Alternative & Punk
5	5	Rock And Roll
6	6	Blues
7	7	Latin
8	8	Reggae
9	9	Pop





```
-- Join data from SQLite, zwei PostgreSQL Servern, and
-- a CSV file
SELECT artist."Name"
    , album.title
    , track."Name"
    , genre."Name"
FROM sqlite.artist AS artist
INNER JOIN public.album
    ON artist."ArtistId" = album.artist_id
INNER JOIN pg14."Track" AS track
    ON album.album_id = track."AlbumId"
INNER JOIN csv.genre AS genre
    ON track."GenreId" = genre."GenreId"
;
```

0	A-z Name	A-Z title	A-Z Name	A-z Name	
1	AC/DC	Let There Be Rock	Whole Lotta Rosie	Rock	
2	AC/DC	Let There Be Rock	Hell Ain't A Bad Place To Be	Rock	
3	AC/DC	Let There Be Rock	Overdose	Rock	
4 AC/DC		Let There Be Rock	Problem Child	Rock	
5 AC/DC		Let There Be Rock	Bad Boy Boogie	Rock	
6	AC/DC	Let There Be Rock	Let There Be Rock	Rock	









```
-- Show data
SELECT *
FROM mv_album_artist
WHERE upper(artist) LIKE 'A%'
ORDER BY artist;
```

0	A-z artist	album_titles	123 sum 🔻
> 1	Aaron Copland & London Symphony Orchestra	> A Copland Celebration, Vol. I	1
> 2	Aaron Goldberg	> Worlds	1
> 3	Academy of St. Martin in the Fields Chamber Ensemble & Sir Neville Marriner	> Sir Neville Marriner: A Celebration	1
> 4	Academy of St. Martin in the Fields, John Birch, Sir Neville Marriner & Sylvia McNair	> Fauré: Requiem, Ravel: Pavane & Others	1
> 5	Academy of St. Martin in the Fields & Sir Neville Marriner	> The World of Classical Favourites	1
> 6	Academy of St. Martin in the Fields, Sir Neville Marriner & Thurston Dart	> Bach: Orchestral Suites Nos. 1 - 4	1
> 7	Academy of St. Martin in the Fields, Sir Neville Marriner & William Bennett	NULL	[NULL]
> 8	Accept	> Balls to the Wall [+1]	2





-- Create the Multicorn extension
CREATE EXTENSION multicorn;





```
-- Create a server, it is only a place holder
CREATE SERVER rss_srv foreign data wrapper multicorn options (
    wrapper 'multicorn.rssfdw.RssFdw'
)
;
```





-- Create a schema for the multicorn tables CREATE SCHEMA IF NOT EXISTS multicorn;





```
-- Create a foreign table based on a RSS feed
CREATE FOREIGN TABLE multicorn.rss_mi2nbandnews (
        title text,
        link text,
        description text,
        "pubDate" TIMESTAMPTZ,
        guid text
) server rss_srv OPTIONS (
        url 'https://www.visions.de/feeds/news.rss'
)
;
```





-- Show data from an RSS feed directly queried from the web  ${\bf SELECT}~\ast$ 

FROM multicorn.rss\_mi2nbandnews;

0	A-Z title 🔻	A-Z link	A-Z description   T		A-Z guid
1	Sólstafir: 2x2 Konzerttickets zu ge	☑ https://www.visions.de/verlo	>Die isländische Post-Metal	2025-05-06 15:50:43.000 +0200	☑ https://www.visions.de/?
2	Killswitch Engage kündigen erste E	☑ https://www.visions.de/news	Nachdem Killswitch Engag	2025-05-06 15:12:43.000 +0200	☑ https://www.visions.de/?
3	Singer-Songwriterin Jill Sobule bei	☑ https://www.visions.de/news	Folk-Pop-Singer/Songwrite	2025-05-06 13:50:28.000 +0200	☑ https://www.visions.de/?
4	John Coffey gehen im November a	☑ https://www.visions.de/news	Ende des Jahres sind John	2025-05-06 13:29:20.000 +0200	☑ https://www.visions.de/?
5	Baxter Dury kündigt neues Album	https://www.visions.de/news	Baxter Dury veröffentlicht	2025-05-06 12:38:35.000 +0200	☑ https://www.visions.de/?
6	André 3000 veröffentlicht überras	☑ https://www.visions.de/news	Ex-Outkast-Rapper André :	2025-05-06 12:15:31.000 +0200	☑ https://www.visions.de/?
7	The Mars Volta und ihr neues Albu	☑ https://www.visions.de/featu	Ende Januar wird das neue	2025-05-06 11:44:32.000 +0200	☑ https://www.visions.de/?
8	Jacknife Lee über Britpop: »Das Ma	☑ https://www.visions.de/news	Starproduzent Jacknife Lee	2025-05-06 11:40:37.000 +0200	☑ https://www.visions.de/?
9	Skegss spielen kostenloses Akustil	☑ https://www.visions.de/news	Vor ihrer morgigen Show i	2025-05-06 11:02:21.000 +0200	☑ https://www.visions.de/?
10	Alle Alben der Donots im Ranking	https://www.visions.de/feature	Die Donots sind echt. Eine	2025-05-05 16:40:18.000 +0200	☑ https://www.visions.de/?





```
-- Join the RSS data with an existing table
SELECT a."Name" AS artist
    , r.*
FROM multicorn.rss_mi2nbandnews AS r
INNER JOIN sqlite.artist AS a
        ON r.description ilike '%' || a."Name" || '%';
```

0	A-Z artist	A-Z title	A-Z link	A-Z description		A-z guid
1	Lost	Sólstafir: 2x2 Konzerttickets zu gewinnen!	https://www.visions.de/verlosungen/s	>Die isländische Post-Metal-Bar	2025-05-06 15:50:43.000 +0200	☑ https://www.visions.de/?p=496444
2	JET	The Mars Volta und ihr neues Album "Luci	https://www.visions.de/features/story	Ende Januar wird das neue Alb	2025-05-06 11:44:32.000 +0200	☑ https://www.visions.de/?p=496388





```
-- Create a materialized view
CREATE MATERIALIZED VIEW multicorn.rss_mi2nbandnews_mv AS
SELECT ROW_NUMBER() OVER() AS rn
    , r.title
    , r.link
    , r.description
    , r."pubDate" AS data_publication
    , r.guid
FROM multicorn.rss_mi2nbandnews AS r;
```









```
-- Show data
SELECT *
FROM multicorn.rss_mi2nbandnews_mv
;
```

(	123 rn	*	A-Z title	A-Z link	A-Z description   T		A-z guid ▼
1		1	Kneecap-Eklat: Billy Bragg	https://www.visions.de/news/billy-bragg-aeussert-sich-	Singer/Songwriter Billy Bragg hat ein Sta	2025-05-06 15:56:09.000 +0200	☑ https://www.visions.de/?p=496435
2		2	Sólstafir: 2x2 Konzertticke	https://www.visions.de/verlosungen/solstafir-konzerttic	Die isländische Post-Metal-Band Sólstafi	2025-05-06 15:50:43.000 +0200	☑ https://www.visions.de/?p=496444
3		3	Killswitch Engage kündige	https://www.visions.de/news/killswitch-engine-this-con	Nachdem Killswitch Engage nach der Ve	2025-05-06 15:12:43.000 +0200	☑ https://www.visions.de/?p=496423
4		4	Singer-Songwriterin Jill So	https://www.visions.de/news/jill-sobule-ist-tot/	Folk-Pop-Singer/Songwriterin Jill Sobule v	2025-05-06 13:50:28.000 +0200	☑ https://www.visions.de/?p=496413
5		5	John Coffey gehen im Nov	https://www.visions.de/news/john-coffey-im-november	Ende des Jahres sind John Coffey wieder	2025-05-06 13:29:20.000 +0200	☑ https://www.visions.de/?p=496420
6		6	Baxter Dury kündigt neue	https://www.visions.de/news/baxter-dury-neues-album	Baxter Dury veröffentlicht mit "A	2025-05-06 12:38:35.000 +0200	☑ https://www.visions.de/?p=496380
7		7	André 3000 veröffentlicht	https://www.visions.de/news/andre-3000-veroeffentlick	Ex-Outkast-Rapper André 3000 bringt of	2025-05-06 12:15:31.000 +0200	☑ https://www.visions.de/?p=496379
8		8	The Mars Volta und ihr ne	https://www.visions.de/features/storys-und-interviews/	Ende Januar wird das neue Album von Th	2025-05-06 11:44:32.000 +0200	☑ https://www.visions.de/?p=496388
9		9	Jacknife Lee über Britpop:	https://www.visions.de/news/jacknife-lee-britpop-war-c	Starproduzent Jacknife Lee lässt seinen l	2025-05-06 11:40:37.000 +0200	☑ https://www.visions.de/?p=496390
10		10	Skegss spielen kostenlose	https://www.visions.de/news/skegss-kostenlose-akustil	Vor ihrer morgigen Show in Berlin spiele	2025-05-06 11:02:21.000 +0200	☑ https://www.visions.de/?p=496378





```
-- Link the RSS data in the materialized view with a table
SELECT a."Name" AS artist
    , r.*
FROM multicorn.rss_mi2nbandnews_mv AS r
INNER JOIN sqlite.artist AS a
        ON r.description ilike '%' || a."Name" || '%'
;
```

•	A-z artisl 🔻	123 rn 🔻	A-Z title 🔻	AZ link	A-Z description 🔻	Ø data_publication ▼	A-z guid 🔻
1	JET	8	The Mars Volta und ihr neues Alb	https://www.visions.de/features/story	Ende Januar wird das neue Album von T	2025-05-06 11:44:32.000 +0200	☑ https://www.visions.de/?p=496388
2	Lost	2	Sólstafir: 2x2 Konzerttickets zu ge	https://www.visions.de/verlosungen/s	Die isländische Post-Metal-Band Sólsta	2025-05-06 15:50:43.000 +0200	☑ https://www.visions.de/?p=496444





-- Create a cron extension to schedule refreshs
CREATE EXTENSION pg\_cron;





```
-- This table is used to log cron calls
CREATE TABLE cron.log (
    log_id bigserial NOT NULL,
    time_stamp_begin timestamp WITH time ZONE NOT NULL,
    time_stamp_end timestamp WITH time ZONE NOT NULL,
    executed text NOT NULL,
    CONSTRAINT log_pk PRIMARY KEY (log_id)
);
```





```
-- Create a procedure to refresh and log the call
CREATE OR REPLACE PROCEDURE refresh every minute() AS $$
DECLARE
    ts start timestamp WITH time ZONE DEFAULT current timestamp;
BEGIN
     -- Refresh the materialized view
    REFRESH MATERIALIZED VIEW CONCURRENTLY multicorn.rss mi2nbandnews mv;
     -- Add a log entry
    INSERT INTO cron.log (executed, time stamp begin, time stamp end)
    VALUES (
          'REFRESH MATERIALIZED VIEW CONCURRENTLY multicorn.rss mi2nbandnews mv',
         ts start,
         current timestamp
    );
END;
$$
LANGUAGE plpgsql
```





```
-- Create a cron job inside PostgreSQL that will refresh the
-- Materialized view with RSS feed data every one minute
INSERT INTO cron.job (schedule, command, nodename, nodeport, database, username)
VALUES (
    '* * * * *',
    'CALL refresh_every_minute()',
    '',
    5435,
    'chinook',
    'postgres'
);
```





-- Show the cron job
SELECT \*
FROM cron.job;

•	123 ™ jobid ▼	A-Z schedule 💌	A-Z command 🔻	A-Z nodename 🔻	123 nodepor ▼	A-Z databası 🔻	A-Z username 🔻	☑ active ▼	A-Z jobname 🔻
1	1	****	CALL refresh_every_minute()		5.435	chinook	postgres	[v]	[NULL]





-- Show the cron execution in the log table
SELECT \*
FROM cron.log;

	0	123 log_id 💌	② time_stamp_begin		A-Z executed
ĺ	1	1	2025-05-06 16:14:00.022 +0200	2025-05-06 16:14:00.022 +0200	REFRESH MATERIALIZED VIEW CONCURRENTLY multicorn.rss_mi2nbandnews_mv





-- Show the cron execution in the log table
SELECT \*
FROM cron.log;

0	123 log_id <b>~</b>		Ø time_stamp_end     ▼	A-Z executed
1	1	2025-05-06 16:14:00.022 +0200	2025-05-06 16:14:00.022 +0200	REFRESH MATERIALIZED VIEW CONCURRENTLY multicorn.rss_mi2nbandnews_mv
2	2	2025-05-06 16:15:00.022 +0200	2025-05-06 16:15:00.022 +0200	REFRESH MATERIALIZED VIEW CONCURRENTLY multicorn.rss_mi2nbandnews_mv
3	3	2025-05-06 16:16:00.020 +0200	2025-05-06 16:16:00.020 +0200	REFRESH MATERIALIZED VIEW CONCURRENTLY multicorn.rss_mi2nbandnews_mv
4	4	2025-05-06 16:17:00.025 +0200	2025-05-06 16:17:00.025 +0200	REFRESH MATERIALIZED VIEW CONCURRENTLY multicorn.rss_mi2nbandnews_mv





-- With the job id one can cancle a cron job
SELECT cron.unschedule(1);





```
-- Create another foreign table from a RSS feed
CREATE FOREIGN TABLE multicorn.rss_postgresql_events (
    title text,
    link text,
    description text,
    "pubDate" TIMESTAMPTZ,
    guid text
) server rss_srv OPTIONS (
    url 'https://www.postgresql.org/events.rss'
);
```





```
-- Show the upcoming PostgreSQL events
SELECT title
    , "pubDate"::DATE AS "Conference Start Date"
    , description
FROM multicorn.rss_postgresql_events
WHERE "pubDate"::DATE > NOW()::DATE
ORDER BY "pubDate" ASC
:
```

0	A-Z title ▼	Ø Conference Start Date ▼	A-Z description
1	PG Day France 2025¶	2025-06-03	¶The PG Day France is the annual conference of the Frence
2	POSETTE: An Event for Postgres 2025¶	2025-06-10	¶ <a href="https://posetteconf.com/2025/">POSETTE: An E</a>
3	Swiss PGDay 2025¶	2025-06-26	¶The Swiss PostgreSQL Conference. Take the opportunity t
4	FOSSY 2025¶	2025-07-31	¶ <a href="https://2025.fossy.us/">FOSSY</a> (Free and Op
5	pgDay México 2025¶	2025-08-01	¶pgDay Mexico is an international event that will take place
6	PGConf.Brazil 2025¶	2025-09-03	¶The Brazilian PostgreSQL conference, also known as PGC
7	PGDay Austria 2025¶	2025-09-04	¶ <a href="https://pgday.at">PGDay Austria</a> is back! Sa
8	PGDay Napoli 2025¶	2025-09-25	¶The first edition of PGDay Napoli is a one-day PostgreSQI
9	PGDay Israel 2025¶	2025-10-19	¶For the eighth time in Israel PostgreSQL enthusiast from
10	PostgreSQL Development Conference 2026 (pgconf.dev)	2026-05-19	¶PostgreSQL Development Conference 2026¶PGConf.dev



#### Link List



Slides and source codes are available on GitLab and GitHub:

https://gitlab.com/sjstoelting/talks/

https://github.com/sjstoelting/talks/



# PostgreSQL As Data Integration Tool



© <u>0</u>

This document by Stefanie Janine Stölting is covered by the Creative Commons Attribution 4.0 International