



**Write a Foreign Data Wrapper in  
15 minutes**

## Table des matières

Write a Foreign Data Wrapper in 15 minutes.....	1
1 About me.....	4
2 Foreign Data Wrappers ?.....	5
3 Goals.....	5
4 Agenda.....	5
5 Part 1 - SQL/MED.....	6
6 MED ?.....	6
7 DBLink and DBI-Link.....	6
8 In other RDBMS.....	7
9 Overview.....	7
10 Part 2 - How To .....	8
11 Install.....	8
12 Create server.....	8
13 Create Foreign Table.....	8
14 Select.....	9
15 Errors.....	9
16 Beware.....	10
17 Uninstall.....	10
18 Part 3 : Current List (feb. 2012).....	10
19 RDBMS fdw.....	11
20 NoSQL fdw.....	11
21 Text fdw.....	11
22 Web fdw.....	12
22.1 ldap_fdw.....	12
22.2 s3_fdw.....	12
22.3 www_fdw.....	13
23 Others.....	14
24 What about pgsq1_fdw ?.....	14
25 Full list.....	14
26 Ideas.....	15
27 Part 4 : Write a fdw.....	15
28 Write a FDW in C : the basics.....	15
29 Problems with that.....	16
30 Multicorn.....	16
31 Multicorn is your friend !.....	16
32 Install multicorn.....	17
33 ical_fdw.py.....	17
34 make install.....	18
35 Plug - 1.....	18
36 Plug - 2.....	18
37 Play - 1.....	18
38 Play - 2.....	19
39 There's more : Optimize.....	19
40 There's more : Log.....	20

41 Conclusions.....	20
42 In a nutshell.....	20
43 Limitations.....	21
44 ETL No More.....	21
45 Release on pgxn.....	21
46 Links.....	22
47 Contact.....	22
48 KTHXBYE.....	23

# I CAN HAS



# DATA WRAPPERS

## 1 About me



- Damien Clochard
- COO of DALIBO, French PostgreSQL Company
- Admin of [www.postgresql.fr](http://www.postgresql.fr)

## 2 Foreign Data Wrappers ?



« One Database to Rule Them All »

## 3 Goals



- Know more about SQL/MED
- Use the existing data wrappers
- Write your own in 15 minutes

## 4 Agenda



1. SQL/MED ?
2. How to use a FDW
3. Current list
4. How to write a FDW

## 5 Part 1 - SQL/MED



- MED ?
- DBLink & DBI-LINK
- In other RDBMS
- Overview
- Why SQL/MED is cool

## 6 MED ?



- **Management of External Data**
- Appeared in SQL:2003
- implemented in PostgreSQL 9.1
- Also present in IBM DB2

<http://www.ibm.com/developerworks/data/library/techarticle/0203haas/0203haas.html>

## 7 DBLink and DBI-Link



- DBLink
  - Connect to other PostgreSQL databases
  - Contrib module since 8.3
  - Will be replaced by `pgsql_fdw`
- DBI-Link
  - Partial implementation written in PL/PerlU
  - Started in 2004
  - Now obsolete

More information :

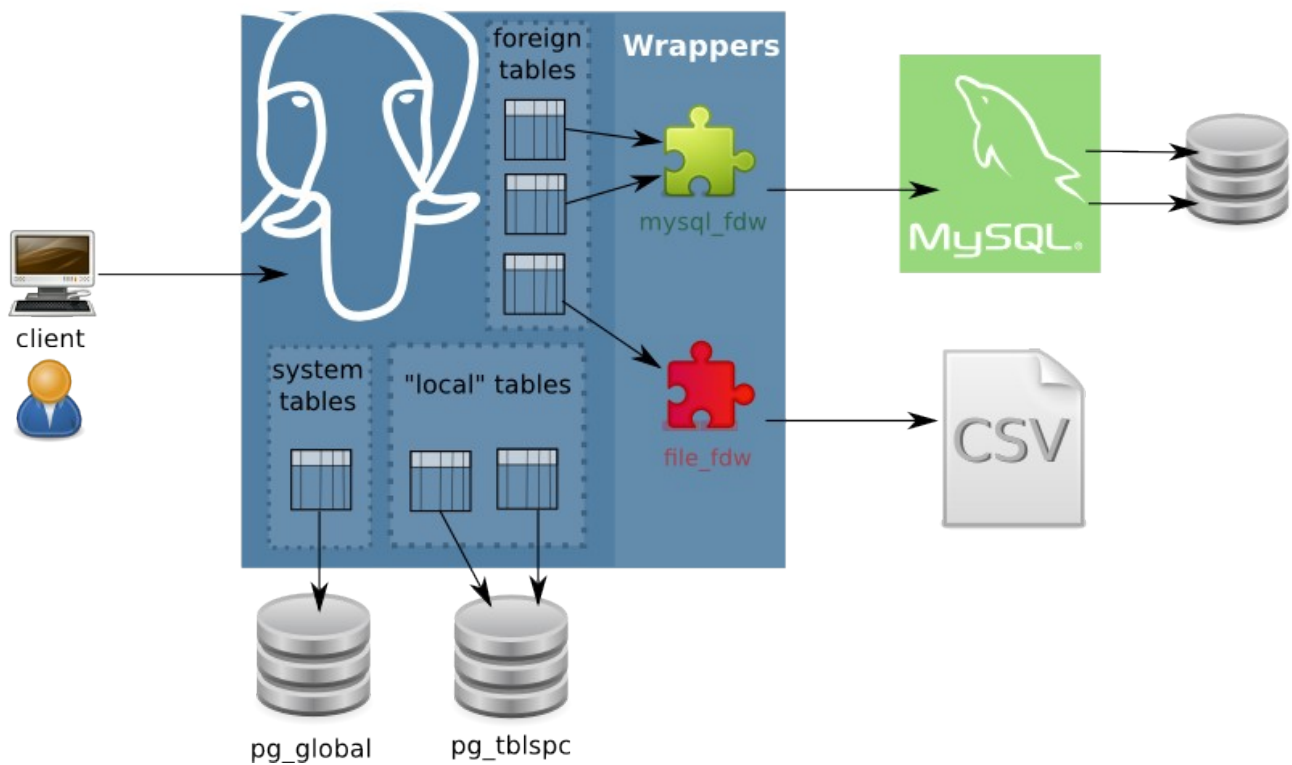
- <http://www.postgresql.org/docs/9.1/static/dblink.html>
- <https://github.com/davidfetter/DBI-Link>

## 8 In other RDBMS



- Informix: External Tables (flat files)
- Oracle: External Tables + DATABASE LINK (odbc)
- MySQL: FEDERATED Storage Engine (mysql only)
- MSSQL: Text File Driver (cvs only)
- Firebird: External Table (cvs only)
- DB2: Complete SQL/MED implementation

## 9 Overview



## 10 Part 2 - How To ...



- Install
- Create server
- Create foreign table
- Select
- Uninstall

## 11 Install

```
CREATE EXTENSION file_fdw;
```

## 12 Create server

```
CREATE SERVER fileserver  
  FOREIGN DATA WRAPPER file_fdw;
```

## 13 Create Foreign Table

```
CREATE FOREIGN TABLE meminfo (  
  stat text,  
  value text  
) SERVER fileserver  
OPTIONS ( filename '/proc/meminfo',  
          format 'csv',  
          delimiter ':'  
);
```



## 14 Select

```
SELECT value
FROM meminfo
WHERE stat = 'MemFree';
```

```
      value
-----
      55024 kB
(1 line)
```

## 15 Errors



- CREATE FOREIGN TABLE
  - won't check if the data storage exists/
  - won't check if the data can be read
  - won't check if the data format is ok
- But SELECT will return an error

## 16 Beware

- There's no cache
- SELECT will scan the whole file every time
- To avoid this :



```
CREATE TABLE localtable  
AS SELECT * FROM foreigntable
```

- Which is more or less equal to :

```
COPY localtable FROM '/your/file';
```

## 17 Uninstall



- DROP FOREIGN TABLE meminfo;
- DROP SERVER fileserver CASCADE;
- DROP EXTENSION file\_fdw CASCADE;

## 18 Part 3 : Current List (feb. 2012)



- SQL Databases Wrappers
- NoSQL Databases Wrappers
- File Wrappers
- Web Wrappers
- Others

In a nutshell, you can now use various Foreign Data Wrappers (FDW) to connect a PostgreSQL Server to remote data stores. This page is an incomplete list of the Wrappers available right now. Another fdw list can be found at the PGXN website. Please keep in mind that most of these wrappers are not officially supported by the PostgreSQL Global Development Group (PGDG) and that some of these projects are still in Beta version. Use carefully!

## 19 RDBMS fdw



- oracle\_fdw
- mysql\_fdw
- sql\_lite
- odbc\_fdw

## 20 NoSQL fdw



- couchdb\_fdw
- redis\_fdw

## 21 Text fdw



- file\_fdw (delivered with PG 9.1)
- file\_text\_array\_fdw

## 22 Web fdw



- www\_fdw
- twitter\_fdw
- google\_fdw
- rss\_fdw
- imap\_fdw
- s3\_fdw

### 22.1 ldap\_fdw

Allows PostgreSQL to query an LDAP server and retrieve data from some pre-configured Organizational Unit.

```
CREATE EXTENSION ldap_fdw;

CREATE SERVER ldap_my_server_service
FOREIGN DATA WRAPPER ldap_fdw
OPTIONS ( address 'my_ldap_server', port '389' );

CREATE USER MAPPING FOR current_user
SERVER ldap_my_server_service
OPTIONS ( user_dn 'cn=the_ldap_user,dc=example,dc=com', password 'the_ldap_user_password' );

CREATE FOREIGN TABLE ldap_people (
    uid text,
    displayname text,
    mail text
)
SERVER ldap_my_server_service
OPTIONS ( base_dn 'OU=people,DC=example,DC=com' );

SELECT * FROM ldap_people WHERE mail = 'user@mail.com';
```

### 22.2 s3\_fdw

s3\_fdw reads files located in Amazon S3

```
db1=# CREATE EXTENSION s3_fdw;
CREATE EXTENSION

db1=# CREATE SERVER amazon_s3 FOREIGN DATA WRAPPER s3_fdw;
CREATE SERVER

db1=# CREATE USER MAPPING FOR CURRENT_USER SERVER amazon_s3
OPTIONS (
    accesskey 'your-access-key-id',
    secretkey 'your-secret-access-key'
```

```
);
CREATE USER MAPPING
db1=# CREATE FOREIGN TABLE log20110901(
    atime timestamp,
    method text, elapse int,
    session text
) SERVER amazon_s3
OPTIONS (
    hostname 's3-ap-northeast-1.amazonaws.com',
    bucketname 'umitanuki-dbtest',
    filename 'log20110901.txt',
    delimiter E'\t'
);
CREATE FOREIGN TABLE
```

## 22.3 www\_fdw

www\_fdw allows to query different web services: Here is example for google search usage:

```
DROP EXTENSION IF EXISTS www_fdw CASCADE;
CREATE EXTENSION www_fdw;
CREATE SERVER www_fdw_server_google_search FOREIGN DATA WRAPPER www_fdw
    OPTIONS (uri 'https://ajax.googleapis.com/ajax/services/search/web?v=1.0');
CREATE USER MAPPING FOR current_user SERVER www_fdw_server_google_search;
CREATE FOREIGN TABLE www_fdw_google_search (
    /* parameters used in request */
    q text,

    /* fields in response */
    GsearchResultClass text,
    unescapedUrl text,
    url text,
    visibleUrl text,
    cacheUrl text,
    title text,
    titleNoFormatting text,
    content text
) SERVER www_fdw_server_google_search;

postgres=# SELECT url, substring(title, 1, 25) || '...', substring(content, 1, 25) || '...'
FROM www_fdw_google_search WHERE q='postgres';
      url              |                ?COLUMN?                |                ?COLUMN?
-----+-----
http://www.postgresql.org/ | <b>PostgreSQL</b>: The wo... | Sophisticated open-
source...
http://www.postgresql.org/download/ | <b>PostgreSQL</b>: Downlo... | The core of the
<b>Postgr...
http://www.postgresql.org/docs/ | <b>PostgreSQL</b>: Docume... | There IS a wealth of
<b>P...
http://en.wikipedia.org/wiki/PostgreSQL | <b>PostgreSQL</b> - Wikip... | <b>PostgreSQL</b>, often
...
(4 rows)
```

## 23 Others



- git\_fdw
- ldap\_fdw
- process\_fdw
- pgStorm

## 24 What about pgsqld\_fdw ?



- Will be an “official” extension (like file\_fdw)
- Currently under development
- Probably in 9.2
- Until then keep using dblink

## 25 Full list



- [http://wiki.postgresql.org/wiki/Foreign\\_data\\_wrappers](http://wiki.postgresql.org/wiki/Foreign_data_wrappers)
- <http://pgxn.org/tag/fdw/>
- 1 new wrapper every month

## 26 Ideas



- XML
- Excel / OpenDocument Spreadsheet
- GIS file formats (KML,WKT,... )
- Images
- open data websites
- RRD

## 27 Part 4 : Write a fdw



- How to implement in C
  - Multicorn
- Let's write a wrapper for ICalendar files !

## 28 Write a FDW in C : the basics



- 2 functions :
  - handler
  - validator (optionnal)
- 6 callback routines (called by planner and executor)
  - PlanForeignScan / ExplainForeignScan
  - BeginForeignScan / EndForeignScan
  - IterateForeignScan / ReScanForeignScan

## 29 Problems with that...



- This is complex
- This is not well documented
- The code is generic
- What if i don't know C?



# MULTICORN

## 30 Multicorn

## 31 Multicorn is your friend !



- A simple python interface to write FDW
- “A wrapper over a wrapper”
- Handles all the dirty C code for you
- You can focus on the extraction
- ... and use python's librairies



## 32 Install multicorn



- You need postgresql-server-dev and python-dev
- `pgxn install multicorn --testing`
- Sources at [github.com/Kozea/Multicorn.git](https://github.com/Kozea/Multicorn.git)

```
apt-get install python-software-properties
add-apt-repository ppa:pitti/postgresql
apt-get update
apt-get install postgresql-9.1 postgresql-server-dev-9.1 libpq-dev python python-dev python-
setuptools
easy_install importlib
easy_install pgxnclient
```

then

```
pgxn install multicorn --testing
```

or

```
git clone git://github.com/Kozea/Multicorn.git
cd Multicorn
make && make install
```

## 33 ical\_fdw.py

```
from multicorn import ForeignDataWrapper
import urllib
from icalendar import Calendar, Event

class ICalFdw(ForeignDataWrapper):

    def __init__(self, options, columns):
        super(ICalFdw, self).__init__(options, columns)
        self.url = options.get('url', None)
        self.columns = columns

    def execute(self, quals, columns):
        ical_file = urllib.urlopen(self.url).read()
        cal = Calendar.from_string(ical_file)
        for v in cal.walk('vevent'):
            e = Event(v)
            line = {}
            for column_name in self.columns:
```

```
        line[column_name] = e.decoded(column_name)
    yield line
```

## 34 make install

```
cd Multicorn
cp icalfdw.py python/multicorn
make & make install
```

## 35 Plug - 1

```
CREATE EXTENSION multicorn;

CREATE SERVER multicorn_ical
  FOREIGN DATA WRAPPER multicorn
  options (wrapper 'multicorn.icalfdw.ICalFdw');
```

## 36 Plug - 2

```
CREATE FOREIGN TABLE fosdem_events (
  dtstart TIMESTAMP,
  dtend TIMESTAMP,
  summary TEXT
) server multicorn_ical options (
  url 'http://fosdem.org/schedule/ical'
);
```

## 37 Play - 1

```
SELECT count(*)
```

```
FROM fosdem_events;

count
-----
  437
(1 line)
```

## 38 Play - 2

```
SELECT to_char(e.dtend, 'HH24:MI') AS end_of_this_talk
FROM fosdem_events e
WHERE e.summary ILIKE '%foreign data wrapper%';

end_of_this_talk
-----
  12:50
(1 line)
```

## 39 There's more : Optimize

Reduce the amount of data fetched over the network :



- handle the “quals” argument
- a Qual object defines :
  - a field\_name
  - an operator
  - a value

## 40 There's more : Log



- multicorn.utils has a `log_to_postgres` function
- 3 parameters
  - message
  - level
  - hint (optional)

```
log_to_postgres('You MUST set the url', ERROR)
```

## 41 Conclusions



- In a Nutshell
- Limitations
- ETL No More
- Release on pgxn
- The future
- Links / Contact

## 42 In a nutshell



Useful for :

- Easier migrations
- Data Warehouse
- Reporting
- Avoid useless data replication

## 43 Limitations



- Read Only
- Most wrappers are beta or alpha
- Most are not supported by PGDG
- Forget about performance
- Forget about data integrity

## 44 ETL No More



- Instead of : EXTRACT - TRANSFORM - LOAD
- Just : CREATE FOREIGN TABLE - SELECT
- Use SQL or PL/SQL to manipulate data
- Transformation code stays in the DB

## 45 Release on pgxn



- If you write your own FDW:
- publish it on PGXN
  - update the wiki page

## 46 Links



- This talk : <http://dalibo.org/conferences>
- Multicorn : <http://multicorn.org>
- Ical FDW : <http://moourl.com/icalfdw>

## 47 Contact



- me : [damien@dalibo.info](mailto:damien@dalibo.info)
- Multicorn creator : Ronan Dunklau [rdunklau@gmail.com](mailto:rdunklau@gmail.com)



# QUESTIONS ???

48 KTHXBYE