

PostgreSQL

PGDay.IT 2011

Monash University Prato Centre  
Venerdì 25 Novembre 2011

# Foreign Table e fonti di dati esterne

**Giulio Calacoci**

Italian PostgreSQL Users Group

[www.itpug.org](http://www.itpug.org)

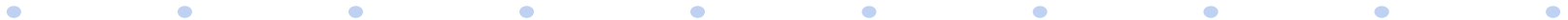
[www.postgresql.org](http://www.postgresql.org)

Giulio Calacoci - [Giulio.Calacoci@2ndQuadrant.it](mailto:Giulio.Calacoci@2ndQuadrant.it) - [ITPUG.org](http://ITPUG.org)



## Chi sono?

- 2ndQuadrant Italia:
  - Business Intelligence
  - Software Development:
    - Java
    - PHP



## Sommario

- SQL/MED
- Perché usare una Foreign Table
- Concetti di base e Sintassi
- Casi d'uso

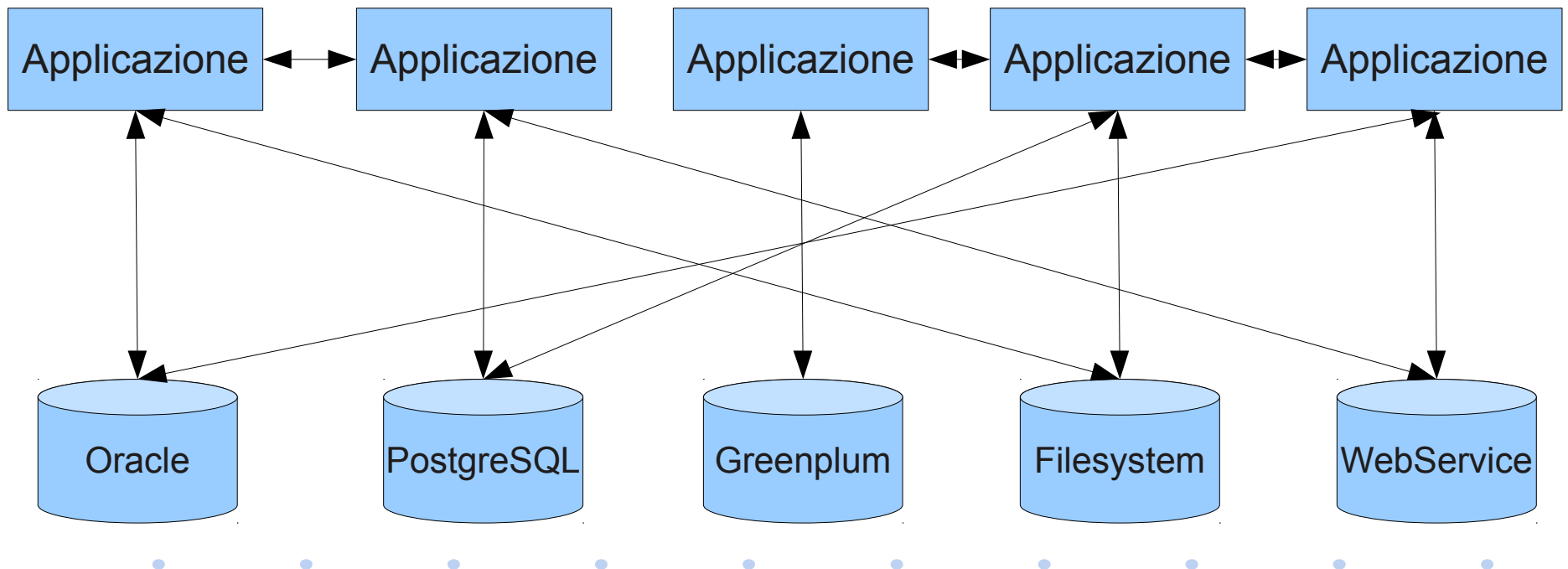


## SQL/MED

- **Management of External Data**
- *“Provides extensions to SQL that define foreign-data wrappers and datalink types to allow SQL to manage external data. External data is data that is accessible to, but not managed by, an SQL-based DBMS.”*  
(fonte: Wikipedia)

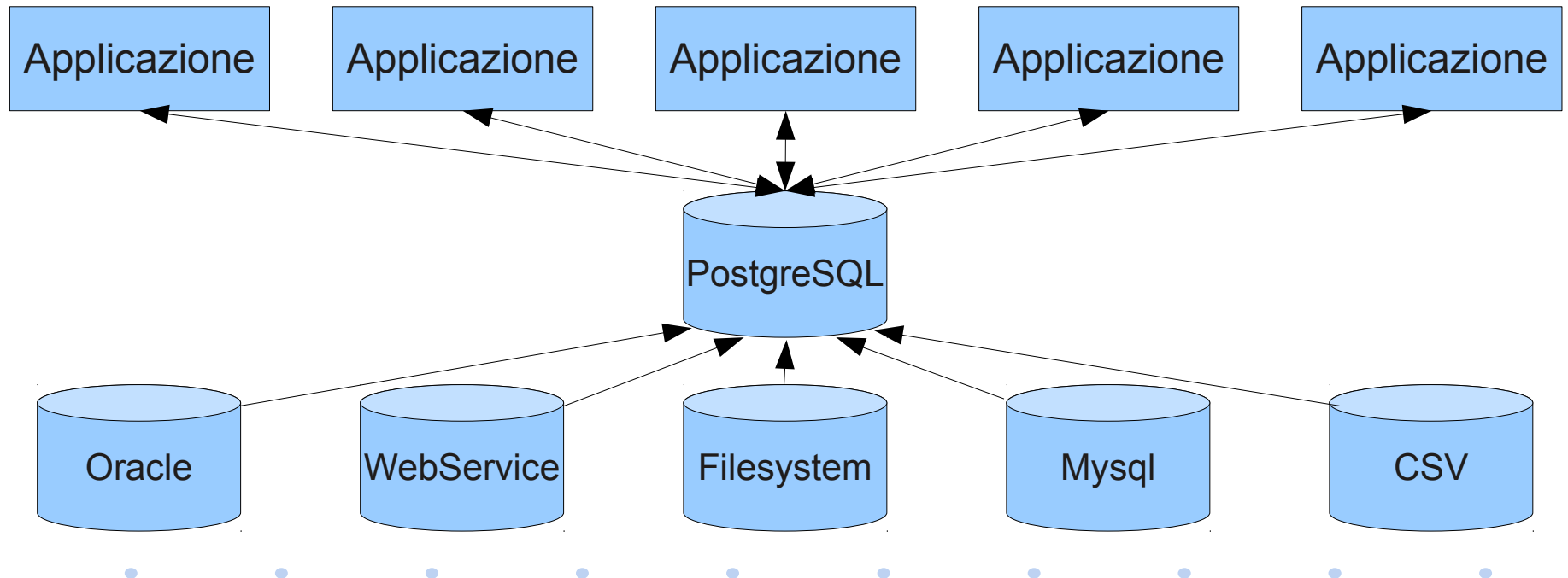
## Perché?

- Permette di trasformare un sistema strutturato così:



## Perché?

- In uno strutturato così :



## Perché? - I vantaggi

- Semplificazione e centralizzazione
- Rende PostgreSQL punto nevralgico per l'accesso ai dati
- Dati esterni collegati come tabelle
- Possibilità di gestire i permessi di accesso ai dati
- ETL con SQL



## Perché no? Gli Svantaggi

- Al momento le Foreign Table sono in sola lettura
- Potrebbe non essere disponibile un Foreign data wrapper adatto alle nostre esigenze
- Accesso più lento, il degrado delle prestazioni dipende dal tipo della fonte
  - No ANALYZE ( 9.2 ? )
  - No caching





## Concetti di base

- Definire una tabella “Esterna”
- Su un server “Esterno”
- Tramite un **wrapper** di dati “Esterni”
  - libreria per comunicazione con un tipo specifico di fonte esterna
  - Condiviso da più tabelle nello stesso database

## Alcuni esempi pratici

- Foreign table su file
- Foreign table su MySQL
- Foreign table su Oracle



## Esempio #1: file

- Caricamento del FDW :

```
CREATE EXTENSION file_fdw;
```

- Creazione di un oggetto “server” che definisce la sorgente dei dati:

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

- Creazione della tabella:

```
CREATE FOREIGN TABLE statistical_data  
  (field1 numeric, field2 numeric)  
  SERVER file  
  OPTIONS (filename '/tmp/statistical_data.csv',  
          format 'csv', delimiter ';') ;
```

• • • • • • • • • •

# /etc/passwd lettura tramite Foreign Table

```
CREATE EXTENSION file_fdw;  
CREATE EXTENSION  
  
CREATE SERVER passwd FOREIGN DATA WRAPPER file_fdw ;  
CREATE SERVER  
  
CREATE FOREIGN TABLE passwd (  
    username text, pass text, uid int4, gid int4,  
    gecos text, home text, shell text  
) SERVER passwd  
OPTIONS (format 'text', filename '/etc/passwd', delimiter ':', null '');
```

## /etc/passwd lettura tramite Foreign Table

```
> SELECT username, uid, home FROM passwd LIMIT 3;
```

username	uid	home
root	0	/root
daemon	1	/usr/sbin
bin	2	/bin

## Sicurezza

- > GRANT SELECT ON passwd TO test ;
- > GRANT
  
- GRANT e REVOKE come su ogni altra tabella
- Da non dimenticare:
  - Il controllo ai dati è definibile anche a livello di colonna!

## Connessione a MySQL

```
mysql> DESCRIBE test_ft;
```

Field	Type	Null	Key	Default	Extra
id	int(11)	NO	PRI	NULL	auto_increment
data	varchar(100)	YES		NULL	

```
mysql> SELECT * FROM test_ft WHERE id <= 2;
```

id	data
1	test
2	test1

## Connessione a MySql

```
CREATE EXTENSION mysql_fdw;  
CREATE EXTENSION  
  
CREATE SERVER p2m  
    FOREIGN DATA WRAPPER mysql_fdw  
    OPTIONS (address '127.0.0.1', port '3306');  
CREATE SERVER
```



## Connessione a MySql

```
CREATE FOREIGN TABLE testmysql (  
    id integer,  
    data text) SERVER p2m  
    OPTIONS( database 'test', table 'test_ft');  
CREATE FOREIGN TABLE  
  
CREATE USER MAPPING FOR PUBLIC  
    SERVER p2m  
    OPTIONS (username 'test', password 'test');
```

## Connessione a MySql

```
SELECT * FROM testmysql WHERE id <= 2;
```

id	data
1	test
2	test1

```
EXPLAIN SELECT * FROM testmysql WHERE id <=2;
```

QUERY PLAN

---

```
Foreign Scan on testmysql (cost=10.00..18.00 rows=8 width=36)
  Filter: (id <= 2)
  Local server startup cost: 10
  MySQL query: SELECT * FROM test_ft
```

## Connessione a Oracle

```
ORACLE> SELECT id, testdata FROM test WHERE id < 6;
```

ID	TESTDATA
2	test
4	test1

```
CREATE EXTENSION oracle_fdw;
```

```
CREATE SERVER p2o
```

```
    FOREIGN DATA WRAPPER oracle_fdw
```

```
    OPTIONS (dbserver '10.68.1.177:1521/XE');
```

## Connessione a Oracle

```
CREATE USER MAPPING FOR PUBLIC SERVER p2o
  OPTIONS (user 'test', password 'test');

CREATE FOREIGN TABLE testora (
  id integer,
  testdata text)
  SERVER p2o OPTIONS (table 'test', schema 'test');
```

## Connessione a Oracle

```
SELECT id, testdata FROM testora WHERE id < 6;
```

```
Id | testdata  
---+-----  
2 | test  
4 | test1
```

```
EXPLAIN SELECT * FROM testora WHERE id <=6 ;
```

```
QUERY PLAN
```

```
-----  
Foreign Scan on testora (cost=10000.00..10000.00 rows=1000 width=36)
```

```
Filter: (id <= 6)
```

```
Oracle query: SELECT /*b5b0a38dbdd20f626501d6929a915a3e*/ "ID", "TESTDATA"  
FROM test2ndq.test WHERE ("ID" <= 6)
```

```
(3 rows)
```

## Nota sulle Query

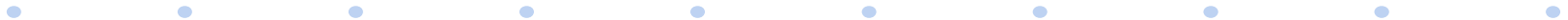
- Ogni FDW esegue le query in maniera differente
- Wrapper avanzati ottimizzano le query
  - Le prestazioni sono dipendenti dall'implementazione



## Connessione a:

Qualcosa di completamente differente:

Twitter



# Twitter Data Wrapper

```
create extension twitter_fdw;
```

```
CREATE EXTENSION
```

```
\d
```

List of relations

Schema	Name	Type	Owner
public	passwd	foreign table	postgres
public	twitter	foreign table	postgres



# Twitter Data Wrapper

Column	Type	Modifiers
id	bigint	
text	text	
from_user	text	
from_user_id	bigint	
to_user	text	
to_user_id	bigint	
iso_language_code	text	
source	text	
profile_image_url	text	
created_at	timestamp without time zone	
q	text	

# Twitter Data Wrapper

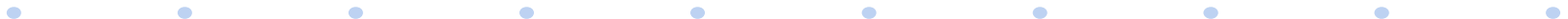
```
SELECT from_user, created_at, text FROM twitter WHERE q = '#PGDay'  
AND from_user = 'simonemartelli';
```

from_user	created_at	text
-----------	------------	------

simonemartelli	2011-11-24 07:21:32	beh decisamente per nerd sì. :-) però divertente. Puoi sempre passare da Ozne per la PgBeer delle 17.30 #pgday
simonemartelli	2011-11-24 07:16:50	#pgday è la giornata italiana dedicata a PostgreSQL, database open source. <a href="http://t.co/KLJAX33I">http://t.co/KLJAX33I</a>
simonemartelli	2011-11-24 07:12:00	Buongiorno. Ormai mancano 24 ore al #pgday a #Prato. Sono al banco della registrazione, arrivate presto e un po' per volta eh 2011.pgday.it

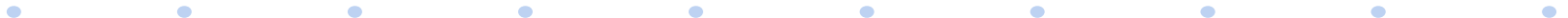
## Conclusioni

- Tabelle esterne:
  - Ulteriore strumento a disposizione per ETL / ELT
    - Interno al database (con SQL)
- SQL/MED in Postgres:
  - 8.4 = definizione connessioni
  - 9.1 = tabelle esterne in sola lettura
  - ?
    - supporto completo in lettura/scrittura
    - Proliferazione driver



## Domande?

- E-Mail: [giulio.calacoci@2ndquadrant.it](mailto:giulio.calacoci@2ndquadrant.it)
- URL: [www.2ndquadrant.it](http://www.2ndquadrant.it)
- Blog: [blog.2ndquadrant.com](http://blog.2ndquadrant.com)



## Licenza Creative Commons

Attribuzione

Non commerciale

Condividi allo stesso modo

2.5 Italia

<http://creativecommons.org/licenses/by-nc-sa/2.5/it/>

Copyright 2011 2ndQuadrant Italia- [www.2ndQuadrant.it](http://www.2ndQuadrant.it)

