

Backup e Restore di un database PostgreSQL

Sandro Fioravanti
INFN-LNF

Abstract

Vengono descritte le procedure di backup e restore dei dati memorizzati in un database di tipo PostgreSQL 8.1 e superiori, prendendo come esempio il database di SPARC.

1. Introduzione

Ci sono 2 modi per fare il backup ed il restore dei dati di una database postgresQL: si possono usare i semplici tool presenti sul pannello di controllo pgAdmin III oppure scrivere comandi manuali. Prima di inviare un qualsiasi comando, bisogna essere assolutamente sicuri di avere ben chiaro cosa si sta facendo, perché i comandi sono irreversibili, e possono durare delle ore, se non giorni. Chiudere il programma pgAdmin non aiuta a fermare il processo. L'unico modo per fermare un processo avviato è quello di fermare il server con i comandi di stop/start o di restart.

2. Eseguire un backup (pg_dump)

Il comando da eseguire per creare un file contenenti i dati del database è pg_dump
pg_dump [connection-option...] [option...] [dbname]

Questo comando estrae i dati dal database e li scrive su un file unico.

Per come è pensato il database di SPARC, il backup è stato organizzato basandosi sull'importanza dei dati e sulla dimensione delle tabelle. Creare un file unico è sconsigliato, perché il ripristino di un file molto grande può richiedere anche dei giorni. Quindi è bene suddividere in più file i sottogruppi delle tabelle. Nel caso del db di SPARC, abbiamo 79 tabelle di vario genere. Ho suddiviso il backup in 3 blocchi principali (system, daq, extra system) e vari file singoli per ragioni di spazio, oltre a questo si è pensato di dividere per anno le tabelle di grandi dimensioni. Per i 3 blocchi principali, è indispensabile usare righe di comando, perché solo tramite riga di comando possiamo abilitare o escludere delle tabelle dal backup dello schema.

Il primo file da creare, "sparc_system_[DATA].backup", riguarda tutte le tabelle inerenti alla parte statica del sistema, e sono tutte quelle tabelle il cui nome termina con "system".

La riga di comando da inviare è:

```
pg_dump.exe -h 192.168.197.253 -p 5432 -U postgres -F c -D -v -f
"C:\Users\sparc\Desktop\BackUp db\sparc_system_2012-01-25.backup" -t
"\sparc\." "AMPSsystem" -t "\sparc\." "APSSystem" -t "\sparc\." "BCMsystem" -t
"\sparc\." "BPMsystem" -t "\sparc\." "CAMsystem" -t "\sparc\." "CARsystem" -t
"\sparc\." "CHLsystem" -t "\sparc\." "CHNsystem" -t "\sparc\." "CORsystem" -t
"\sparc\." "DDGsystem" -t "\sparc\." "FACsystem" -t "\sparc\." "MGIsystem" -t
"\sparc\." "MODsystem" -t "\sparc\." "MOTsystem" -t "\sparc\." "PMCsystem" -t
"\sparc\." "RDSsystem" -t "\sparc\." "RFSsystem" -t "\sparc\." "SDGsystem" -t
"\sparc\." "SERsystem" -t "\sparc\." "SHTsystem" -t "\sparc\." "SIPsystem" -t
"\sparc\." "SSAsystem" -t "\sparc\." "SWTsystem" -t "\sparc\." "THZsystem" -t
"\sparc\." "TMPsystem" -t "\sparc\." "TRJsystem" -t "\sparc\." "UNDsystem" -t
"\sparc\." "VALsystem" -t "\sparc\." "VGAsystem" -t sparc
```

Questa riga di comando va scritta su terminale (dos o unix) da qualsiasi pc collegato in rete che può accedere al server. L'accesso al server viene stabilito con gli appositi comandi di connessione.

Spiegazione riga di comando:

pg_dump.exe

Richiama l'eseguibile

[connection-option...]

-h host

specifica il numero ip della macchina dove il server sta girando.

-p port

specifica la TCP port della macchina dove il server sta girando.

-U username

username per contattare il server

[option...]

-F format

seleziona il formato di uscita del backup, ne esistono di tre tipi, nel nostro caso usiamo:

c

custom

“Formato Custom” è un formato adatto per essere facilmente riutilizzato per il ripristino dati, inoltre è un formato altamente compresso.

t

Restituisce il backup in formato tar, utilizzabile dai sistemi Unix.

p

Restituisce il file in formato plain-text SQL script file (formato di default) è un formato leggibile da qualsiasi programma di testo, ma non essendo compresso occupa moltissimo spazio disco.

-D

with command

Con “-D” inseriamo anche i comandi per la creazione delle varie tabelle.

-v

verbose

restituisce sul terminale lo status del comando.

-f file

Percorso e nome del file che verrà creato, contenente il database.

"C:\Users\SandoR\Desktop\BACKUP POSTGRESS\2011-05-16\sparc_system_2011-05-16.backup"

-t table

con -t, possiamo selezionare le tabelle che vogliamo inserire nel backup.

```
-t "\"sparc\".\BPMsystem\""  
  "schema\".\"tabella\""
```

[dbname]

il nome del database, nel nostro caso sparc

Lo stesso procedimento va applicato per tutte le altre tabelle.

Il secondo file da creare, contiene tutti i dati che non si riferiscono alla parte statica delle “classi” SPARC ma che sono di vario uso nel sistema di controllo, che chiameremo “sparc_extra-system_[DATA].backup”

```
pg_dump.exe -h 192.168.197.253 -p 5432 -U postgres -F c -D -v -f  
"C:\Users\sparc\Desktop\BackUp db\sparc_extra-system_2012-01-25.backup" -t  
"\"sparc\".\"CPU_World\"" -t "\"sparc\".\"RFSatsDYN\"" -t "\"sparc\".\"RFSatsSTA\"" -t  
"\"sparc\".\"RFSdemDYN\"" -t "\"sparc\".\"RFSdemSTA\"" -t "\"sparc\".\"RFShciDYN\"" -t  
"\"sparc\".\"RFShciSTA\"" -t "\"sparc\".\"RFSklpDYN\"" -t "\"sparc\".\"RFSklpSTA\"" -t  
"\"sparc\".\"RFSshsDYN\"" -t "\"sparc\".\"RFSshsSTA\"" -t "\"sparc\".\"class\"" -t  
"\"sparc\".\"command\"" -t "\"sparc\".\"command_log\"" -t "\"sparc\".\"conversion\"" -t  
"\"sparc\".\"menu_table\"" -t "\"sparc\".\"switch-porte-info\"" -t "\"sparc\".\"switch-  
traffico\"" -t "\"sparc\".\"switches\"" -t "\"sparc\".\"sys_flags\"" -t  
"\"sparc\".\"uniMagDataSet\"" -t "\"sparc\".\"error_log\"" sparc
```

Il terzo file contiene tutte le tabelle di media dimensione che si riferiscono ai dati dinamici delle classi SPARC.

```
pg_dump.exe -h 192.168.197.253 -p 5432 -U postgres -F c -D -v -f  
"C:\Users\sparc\Desktop\BackUp db\sparc_daq_small_2012-01-25.backup" -t  
"\"sparc\".\"AMPdaq\"" -t "\"sparc\".\"BPMdaq\"" -t "\"sparc\".\"CAMdaq\"" -t  
"\"sparc\".\"CARdaq\"" -t "\"sparc\".\"CHLdaq\"" -t "\"sparc\".\"DDGdaq\"" -t  
"\"sparc\".\"FACdaq\"" -t "\"sparc\".\"MG1daq\"" -t "\"sparc\".\"MOTdaq\"" -t  
"\"sparc\".\"RDSdaq\"" -t "\"sparc\".\"SERdaq\"" -t "\"sparc\".\"SHTdaq\"" -t  
"\"sparc\".\"SSAdaq\"" -t "\"sparc\".\"SWTdaq\"" -t "\"sparc\".\"THZdaq\"" -t  
"\"sparc\".\"TRJdaq\"" -t "\"sparc\".\"VALdaq\"" sparc
```

Le tabelle escluse dal blocco dei dati dinamici per causa di dimensione, sono:

```
-t "\"sparc\".\"APSdaq\""  
-t "\"sparc\".\"BCMdaq\""  
-t "\"sparc\".\"CHNdaq\""  
-t "\"sparc\".\"CORdaq\""  
-t "\"sparc\".\"MODdaq\""  
-t "\"sparc\".\"RFS1306daq\""  
-t "\"sparc\".\"RFS1307daq\""  
-t "\"sparc\".\"SIPdaq\""  
-t "\"sparc\".\"TMPdaq\""  
-t "\"sparc\".\"VGAdaq\""
```

il backup di queste tabelle è stato effettuato usando I tool di pgAdmin.

Date le enormi dimensioni di queste tabelle (come ad esempio VGAdaq che è composta da 70 milioni di righe con un file di backup di 1.5Gb di spazio) sono state suddivise per anno, cancellando gli anni passati tranne quello che precede l'anno corrente (esempio, VGAdaq_2007, VGAdaq2008 ecc...). In questo modo abbiamo accesso molto più velocemente ai dati dell'anno corrente e se si necessita di consultare i vecchi dati si può accedere allo schema "sparc_backup" dove si trovano tutte le tabelle divise per anno. Il comando per inserire i dati di un determinato periodo da una tabella ad un'altra è:

```
INSERT INTO sparc_backup."VGAdaq_2009" SELECT * FROM sparc."VGAdaq" WHERE tempo BETWEEN '2009-01-01 00:00' AND '2009-12-31 23:59'
```

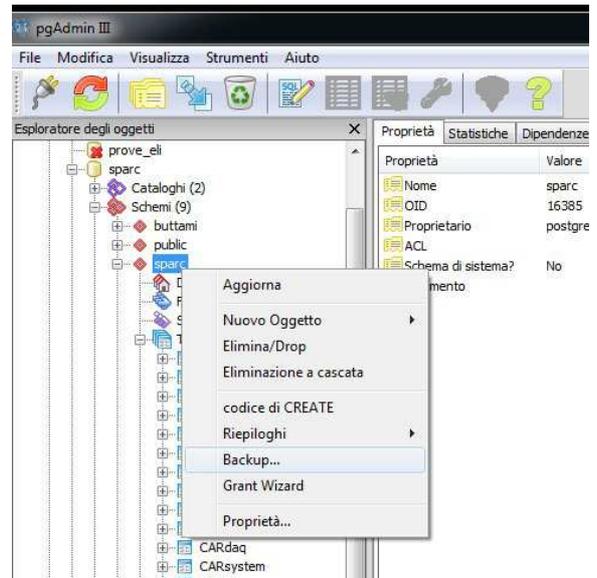
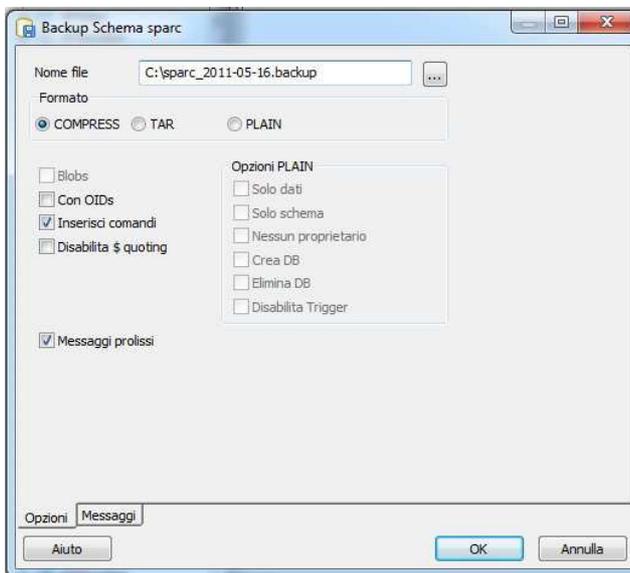
Alla fine del processo il terminale restituisce questo messaggio:

L'interrogazione è stata eseguita con successo: 1035915 righe , 337206 ms tempo di esecuzione.

3. Procedura di backup usando pgAdmin

Cliccare con il tasto destro sullo schema o sulla tabella che vogliamo salvare e cliccare su “Backup...”

Selezionare il percorso e il nome del file, spuntare il formato “COMPRESS” e la modalità “INSERISCI COMANDI” e “MESSAGGI PROLISSI”



Usando pg Admin, si può solo fare il backup di un intero schema o di una singola tabella.

4. Eseguire un Restore (pg_restore)

Il comando da eseguire per ripristinare un database è “pg_restore”

```
pg_restore [connection-option...] [option...] [filename]
```

La procedura di ripristino, è una procedura molto lenta, per file molto grandi può richiedere anche dei giorni.

Come per il backup, si può eseguire questa procedura sia tramite terminale con comando testuale, che con i tool di pgAdmin.

Esempio di restore:

```
pg_restore.exe -h localhost -p 5432 -U postgres -d sparc -v "C:\Users\sparc\Desktop\2011-05-18\sparc_turni_2011-05-18.backup"
pg_restore: connecting to database for restore
pg_restore: creating SCHEMA sparc_turni
pg_restore: creating TABLE prenotazioni_turni
pg_restore: creating TABLE sparc_turni
pg_restore: creating TABLE turni_not
pg_restore: restoring data for table "prenotazioni_turni"
pg_restore: restoring data for table "sparc_turni"
pg_restore: restoring data for table "turni_not"
pg_restore: creating CONSTRAINT sparc_turni_pkey
pg_restore: setting owner and privileges for SCHEMA sparc_turni
pg_restore: setting owner and privileges for TABLE prenotazioni_turni
pg_restore: setting owner and privileges for TABLE sparc_turni
pg_restore: setting owner and privileges for TABLE turni_not
pg_restore: setting owner and privileges for CONSTRAINT sparc_turni_pkey
```

Il processo ha ritornato il codice di uscita 0.

Spiegazione comando:

[connection-option...]

-h host

specifica il numero ip della macchina dove il server sta girando.

-p port

specifica la TCP port della macchina dove il server sta girando.

-U username

username per contattare il server

-d dbname

nome del database al quale ci si vuole connettere

[option...]

-v

verbose

restituisce sul terminale lo status del comando.

-c

--clean

Se la tabella già esiste, PostgreSQL non è in grado di riscriverla.

Quindi si deve aggiungere l'opzione -c che cancella la tabella prima di riscriverla.
[filename]

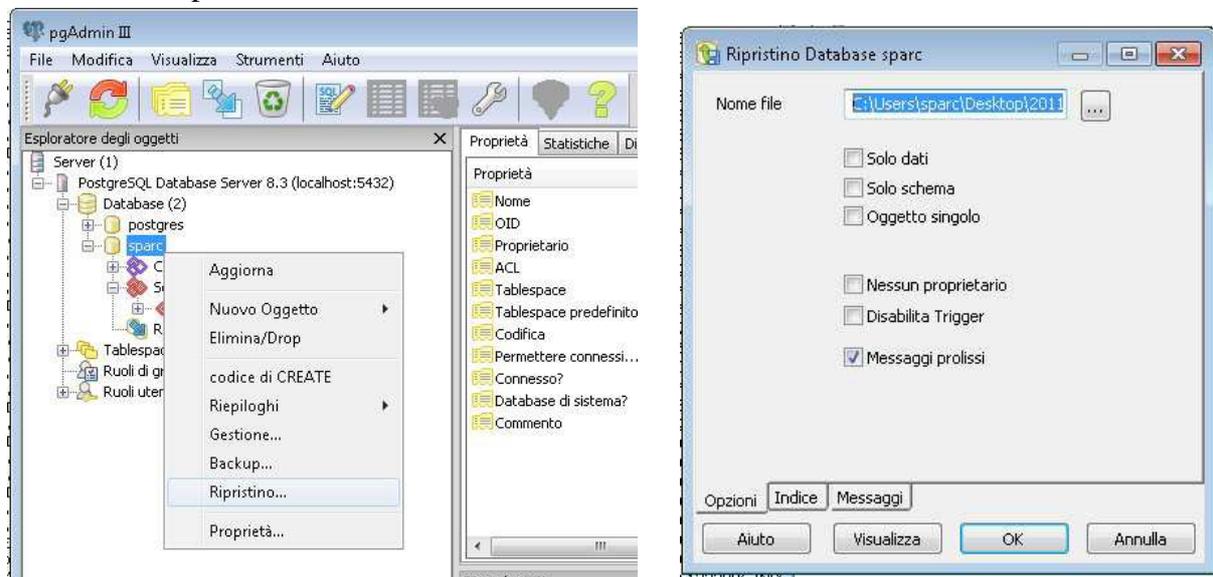
"C:\Users\sparc\Desktop\2011-05-18\sparc_turni_2011-05-18.backup"

Indirizzo e nome del file "backup" dove risiede la nostra tabella.

```
C:\Program Files\PostgreSQL\8.3\bin>pg_restore.exe -h localhost -p 5432 -U postgres -d sparc -v -c "C:\Users\sparc\Desktop\2011-05-18\sparc_turni_2011-05-18.backup"
pg_restore: connecting to database for restore
Password:
pg_restore: dropping CONSTRAINT sparc_turni_pkey
pg_restore: dropping TABLE turni_not
pg_restore: dropping TABLE sparc_turni
pg_restore: dropping TABLE prenotazioni_turni
pg_restore: dropping SCHEMA sparc_turni
pg_restore: creating SCHEMA sparc_turni
pg_restore: creating TABLE prenotazioni_turni
pg_restore: creating TABLE sparc_turni
pg_restore: creating TABLE turni_not
pg_restore: restoring data for table "prenotazioni_turni"
pg_restore: restoring data for table "sparc_turni"
pg_restore: restoring data for table "turni_not"
pg_restore: creating CONSTRAINT sparc_turni_pkey
pg_restore: setting owner and privileges for SCHEMA sparc_turni
pg_restore: setting owner and privileges for TABLE prenotazioni_turni
pg_restore: setting owner and privileges for TABLE sparc_turni
pg_restore: setting owner and privileges for TABLE turni_not
pg_restore: setting owner and privileges for CONSTRAINT sparc_turni_pkey
```

5. Procedura per fare un restore con pgAdmin

Cliccare con il tasto destro sullo schema o sulla tabella che vogliamo ripristinare e cliccare su "Ripristino..."



Selezionare il percorso e il nome del file, spuntare "MESSAGGI PROLISSI" per visualizzare I messaggi provenienti dal server e cliccare il tasto "OK"

Se la tabella è già esistente, il server comunicherà un messaggio di errore di non avvenuta scrittura. Quindi è necessario cancellare prima la tabella da ripristinare e ripetere la procedura.

Bibliografia

Manuali Online:

pg_dump:

<http://www.postgresql.org/docs/8.1/static/app-pgdump.html>

http://www.bigbiz.com/cgi-bin/manpage?1+pg_dump

pg_restore:

<http://www.postgresql.org/docs/current/static/app-pgrestore.html>