

# DEMO



# PostgreSQL 安定運用のレシピ

マニュアルには書かれていないPostgreSQLの真実

篠田典良/ 日本ヒューレット・パッカード株式会社/ 2014年12月5日

PostgreSQL Conference 2014 Hands-On 2

# 1. アーキテクチャとOS設定

# DEMO

## プロセスの確認

```
rel65-9:22 - postgres@rel65-9:~ VT
ファイル(F) 編集(E) 設定(S) コントロール(O) ウィンドウ(W) ヘルプ(H)
[postgres@rel65-9 ~]$ ps -ef|grep postgres
postgres 2192      1      0 09:53 ?        00:00:00 /usr/local/pgsql/bin/postgres -D data
postgres 2193    2192      0 09:53 ?        00:00:00 postgres: logger process
postgres 2195    2192      0 09:53 ?        00:00:00 postgres: checkpointer process
postgres 2196    2192      0 09:53 ?        00:00:00 postgres: writer process
postgres 2197    2192      0 09:53 ?        00:00:00 postgres: wal writer process
postgres 2198    2192      0 09:53 ?        00:00:00 postgres: autovacuum launcher process

postgres 2199    2192      0 09:53 ?        00:00:00 postgres: archiver process    last was 0
000000100000000000000002
postgres 2200    2192      0 09:53 ?        00:00:00 postgres: stats collector process
root      2234    1763      0 09:54 ?        00:00:00 sshd: postgres [priv]
postgres  2239    2234      0 09:54 ?        00:00:00 sshd: postgres@pts/1
postgres  2240    2239      0 09:54 pts/1    00:00:00 -bash
postgres  2278    2240      0 09:55 pts/1    00:00:00 ps -ef
postgres  2279    2240      0 09:55 pts/1    00:00:00 grep postgres
[postgres@rel65-9 ~]$
```



# DEMO

## 共有メモリの確認

```
rel65-9:22 - postgres@rel65-9:~ VT
ファイル(F) 編集(E) 設定(S) コントロール(O) ウィンドウ(W) ヘルプ(H)
[postgres@rel65-9 ~]$ ipcs

----- 共有メモリセグメント -----
キー          shmid      所有者  権限   バイト  nattch  状態
0x0052e2c1  98307      postgres 600    56      5

----- セマフォ配列 -----
キー          semid      所有者  権限   nsems
0x0052e2c1  98306      postgres 600    17
0x0052e2c2  131075     postgres 600    17
0x0052e2c3  163844     postgres 600    17
0x0052e2c4  196613     postgres 600    17
0x0052e2c5  229382     postgres 600    17
0x0052e2c6  262151     postgres 600    17
0x0052e2c7  294920     postgres 600    17
0x0052e2c8  327689     postgres 600    17

----- メッセージキュー -----
キー          msqid      所有者  権限   使用バイト数  メッセージ

[postgres@rel65-9 ~]$
```

System V Shared  
Memoryのサイズは56  
バイト

Key = port \* 1,000 + 1  
= 5,432,001  
= 0x52e2c1



# DEMO

## データベース・クラスタと WAL の確認

```
rel65-9:22 - postgres@rel65-9:~ VT
ファイル(F) 編集(E) 設定(S) コントロール(O) ウィンドウ(W) ヘルプ(H)
[postgres@rel65-9 ~]$ ls data
PG_VERSION      pg_hba.conf      pg_notify         pg_stat_tmp      postgresql.auto.conf
base            pg_ident.conf    pg_replslot       pg_subtrans      postgresql.conf
global          pg_log           pg_serial         pg_tblspc        postmaster.opts
pg_clog         pg_logical       pg_snapshots     pg_twophase      postmaster.pid
pg_dynshmem     pg_multixact     pg_stat          pg_xlog
[postgres@rel65-9 ~]$ ls data/pg_xlog
00000001000000000000000002  000000010000000000000003  archive_status
[postgres@rel65-9 ~]$ █
```



# DEMO

## テーブルとファイルの関係確認

```
rel65-9:22 - postgres@rel65-9:~ VT
ファイル(F) 編集(E) 設定(S) コントロール(O) ウィンドウ(W) ヘルプ(H)
[postgres@rel65-9 ~]$ psql demodb demo
psql (9.4rc1)
Type "help" for help.

demodb=> \d+

               List of relations
 Schema | Name   | Type  | Owner  | Size  | Description
-----+-----+-----+-----+-----+-----
 public | demo1  | table | demo   | 4368 kB
 public | large1 | table | demo   | 2341 MB
 public | main   | table | demo   | 8192 bytes
 public | part100 | table | demo   | 8192 bytes
 public | part200 | table | demo   | 8192 bytes
(5 rows)

demodb=> select pg_relation_filepath('large1');
pg_relation_filepath
-----
base/16385/24608
(1 row)

demodb=>
```


テーブルとファイル名  
の対応を確認



# DEMO

## テーブルとファイルの関係確認

```
rel65-9:22 - postgres@rel65-9:~ VT
ファイル(F) 編集(E) 設定(S) コントロール(O) ウィンドウ(W) ヘルプ(H)
[postgres@rel65-9 ~]$ ls -l data/base/16385/24608*
-rw-----. 1 postgres postgres 1073741824 11月 24 10:11 2014 data/base/16385/24608
-rw-----. 1 postgres postgres 1073741824 11月 24 10:13 2014 data/base/16385/24608.1
-rw-----. 1 postgres postgres 306790400 11月 24 10:13 2014 data/base/16385/24608.2
-rw-----. 1 postgres postgres 622592 11月 24 10:10 2014 data/base/16385/24608_fsm
-rw-----. 1 postgres postgres 40960 11月 24 10:12 2014 data/base/16385/24608_vm
[postgres@rel65-9 ~]$
```



- セグメント・ファイル
- Free Space Map
- Visibility Map

# DEMO

## ALTER SYSTEM 文

```
rel65-9:22 - postgres@rel65-9:~ VT
ファイル(E) 編集(E) 設定(S) コントロール(Q) ウィンドウ(W) ヘルプ(H)
[postgres@rel65-9 ~]$ psql
psql (9.4rc1)
Type "help" for help.

postgres=# show work_mem ;
work_mem
-----
4MB
(1 row)

postgres=# ALTER SYSTEM SET work_mem = '8MB' ;
ALTER SYSTEM
postgres=# show work_mem ;
work_mem
-----
4MB
(1 row)

postgres=# ¥q
[postgres@rel65-9 ~]$ cat data/postgresql.auto.conf
# Do not edit this file manually!
# It will be overwritten by ALTER SYSTEM command.
work_mem = '8MB'
[postgres@rel65-9 ~]$
```

ALTER SYSTEM 文の実行でインスタンスの  
パラメータ値が変更されるわけではない



## 2. 安定稼働のために必要な設定

# DEMO

## ローケールとエンコーディング

```
rel65-9:22 - postgres@rel65-9:~ VT
ファイル(E) 編集(E) 設定(S) コントロール(Q) ウィンドウ(W) ヘルプ(H)
[postgres@rel65-9 ~]$ psql demodb demo
psql (9.4rc1)
Type "help" for help.

demodb=> select datname,datcollate,datatype from pg_database where datname='demodb';
 datname | datcollate | datatype
-----+-----+-----
 demodb  | ja_JP.utf8 | ja_JP.utf8
(1 row)

demodb=> \d+ locale1
                                Table "public.locale1"
  Column |          Type          | Modifiers | Storage | Stats target | Description
-----+-----+-----+-----+-----+-----
   c1    | character varying(10)  |           | extended |               |
   c2    | character varying(10)  |           | extended |               |
Indexes:
    "idx1_locale1" btree (c1)
    "idx2_locale1" btree (c2 varchar_pattern_ops)

demodb=>
```

Locale=ja\_JP

列 c1 と列 c2 は  
インデックスのオプション  
以外同一

# DEMO

## ロケールとエンコーディング

```
rel65-9:22 - postgres@rel65-9:~ VT
ファイル(F) 編集(E) 設定(S) コントロール(Q) ウィンドウ(W) ヘルプ(H)

demodb=> explain analyze select * from locale1 where c1 like '12345%';
                                QUERY PLAN
-----
Seq Scan on locale1  (cost=0.00..179.00 rows=1 width=8) (actual time=2.885..2.885 rows=0 loops=1)
  Filter: ((c1)::text ~~ '12345%'::text)
  Rows Removed by Filter: 10000
  Planning time: 0.999 ms
  Execution time: 2.946 ms
(5 rows)

demodb=> explain analyze select * from locale1 where c2 like '12345%';
                                QUERY PLAN
-----
Index Scan using idx2_locale1 on locale1  (cost=0.29..8.31 rows=1 width=8) (actual time=0.023..0.023
rows=0 loops=1)
  Index Cond: (((c2)::text >= '12345'::text) AND ((c2)::text < '12346'::text))
  Filter: ((c2)::text ~~ '12345%'::text)
  Planning time: 0.789 ms
  Execution time: 0.084 ms
(5 rows)
```

インデックスが使われない

インデックスが使われる

# DEMO

## ロケールとエンコーディング

```
rel65-9:22 - postgres@rel65-9:~ VT
ファイル(E) 編集(E) 設定(S) コントロール(Q) ウィンドウ(W) ヘルプ(H)
[postgres@rel65-9 ~]$ psql demodb demo
psql (9.4rc1)
Type "help" for help.

demodb=> explain select * from locale1 where c1 >= 'A' ;
               QUERY PLAN
-----
Index Scan using idx1_locale1 on locale1  (cost=0.42..5.77 rows=1 width=12)
  Index Cond: ((c1)::text >= 'A'::text)
(2 rows)

demodb=> explain select * from locale1 where c2 >= 'A' ;
               QUERY PLAN
-----
Seq Scan on locale1  (cost=0.00..17905.01 rows=100 width=12)
  Filter: ((c2)::text >= 'A'::text)
(2 rows)

demodb=>
```

インデックスが使われる

インデックスが使われない

# 4. 障害発生時の動作と対処

# DEMO

## クラッシュとファイル削除(変更されていないテーブル)

```
rel65-9:22 - postgres@rel65-9:~ VT
ファイル(F) 編集(E) 設定(S) コントロール(O) ウィンドウ(W) ヘルプ(H)
[postgres@rel65-9 ~]$ psql demodb demo
psql (9.4rc1)
Type "help" for help.

demodb=> select pg_relation_filepath('crash1');
 pg_relation_filepath
-----
base/16385/24712
(1 row)

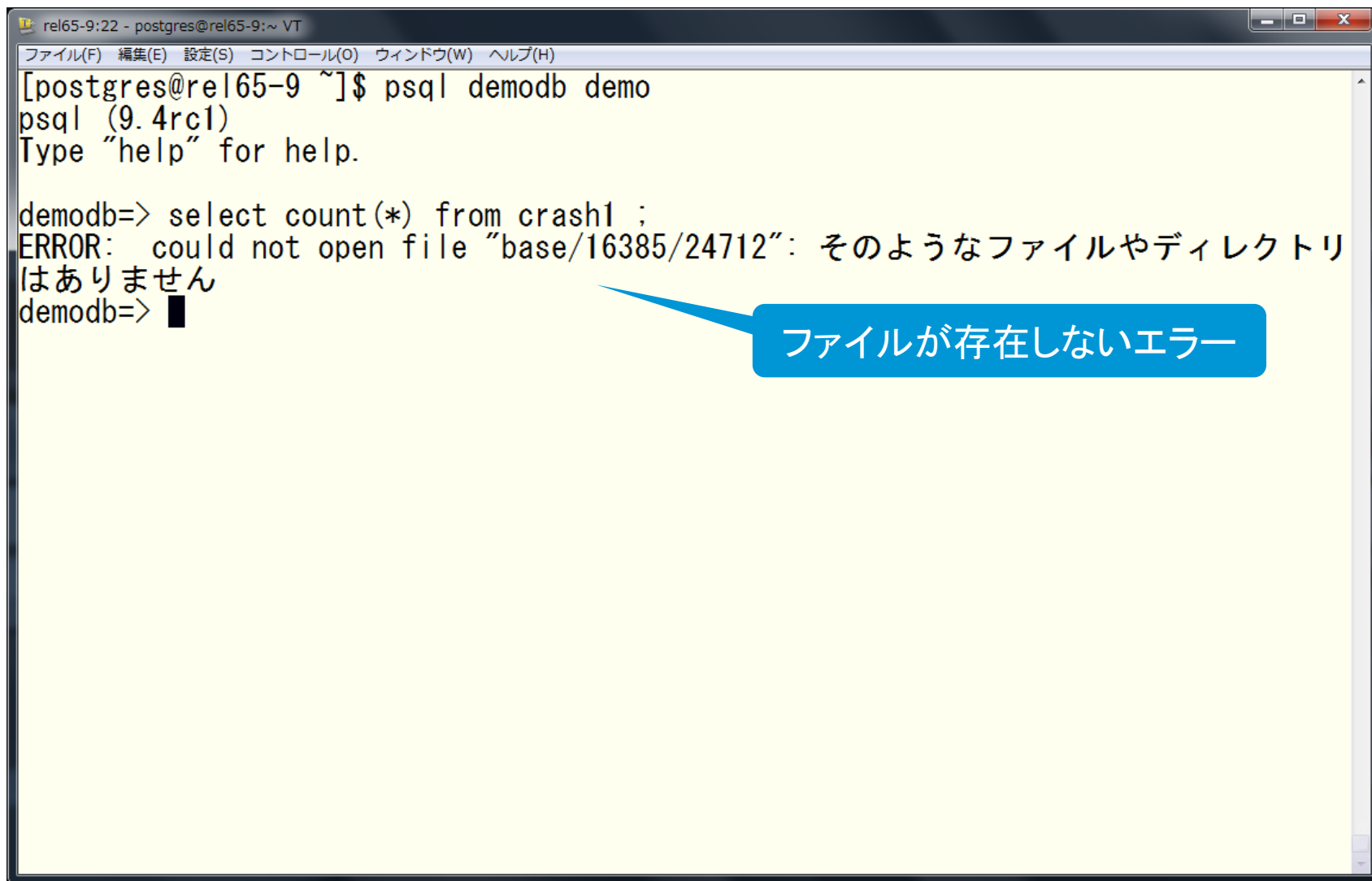
demodb=> \q
[postgres@rel65-9 ~]$ ps -ef|grep '/bin/postgres' | grep -v grep
postgres 11250      1    0 14:06 pts/1    00:00:00 /usr/local/pgsql/bin/postgres
[postgres@rel65-9 ~]$ kill -9 11250
[postgres@rel65-9 ~]$ mv data/base/16385/24712 data/base/16385/24712.org
[postgres@rel65-9 ~]$ pg_ctl start
pg_ctl: another server might be running; trying to start server anyway
server starting
[postgres@rel65-9 ~]$ LOG:  redirecting log output to logging collector process
HINT:  Future log output will appear in directory "pg_log".

[postgres@rel65-9 ~]$
```

疑似クラッシュとファイル削除

# DEMO

## クラッシュとファイル削除(変更されていないテーブル)



```
rel65-9:22 - postgres@rel65-9:~ VT
ファイル(F) 編集(E) 設定(S) コントロール(O) ウィンドウ(W) ヘルプ(H)
[postgres@rel65-9 ~]$ psql demodb demo
psql (9.4rc1)
Type "help" for help.

demodb=> select count(*) from crash1 ;
ERROR:  could not open file "/base/16385/24712": そのようなファイルやディレクトリ
はありません
demodb=> █
```

ファイルが存在しないエラー

# DEMO

## クラッシュとファイル削除(更新されたテーブル)

```
rel65-9:22 - postgres@rel65-9:~ VT
ファイル(F) 編集(E) 設定(S) コントロール(O) ウィンドウ(W) ヘルプ(H)
[postgres@rel65-9 ~]$ psql demodb demo
psql (9.4rc1)
Type "help" for help.

demodb=> select pg_relation_filepath('crash2');
pg_relation_filepath
-----
base/16385/24718
(1 row)

demodb=> select count(*) from crash2;
count
-----
10000
(1 row)

demodb=> insert into crash2 values (generate_series(1, 1000), 'add');
INSERT 0 1000
demodb=> select count(*) from crash2;
count
-----
11000
(1 row)

demodb=> █
```

10,000レコードのテーブルに  
1,000レコード追加  
1,000レコード分の WAL 出力



# DEMO

## クラッシュとファイル削除(更新されたテーブル)

```
rel65-9:22 - postgres@rel65-9:~ VT
ファイル(F) 編集(E) 設定(S) コントロール(O) ウィンドウ(W) ヘルプ(H)
[postgres@rel65-9 ~]$ ps -ef|grep '/bin/postgres' | grep -v grep
postgres 11382      1    0 14:11 pts/0    00:00:00 /usr/local/pgsql/bin/postgres
[postgres@rel65-9 ~]$
[postgres@rel65-9 ~]$
[postgres@rel65-9 ~]$ kill -9 11382
[postgres@rel65-9 ~]$ mv data/base/16385/24718 data/base/16385/24718.org
[postgres@rel65-9 ~]$ pg_ctl start
pg_ctl: another server might be running; trying to start server anyway
server starting
[postgres@rel65-9 ~]$ LOG:  redirecting log output to logging collector process
HINT:  Future log output will appear in directory "pg_log".

[postgres@rel65-9 ~]$ psql demodb demo
psql (9.4rc1)
Type "help" for help.

demodb=> select count(*) from crash2;
 count
-----
   1010
(1 row)

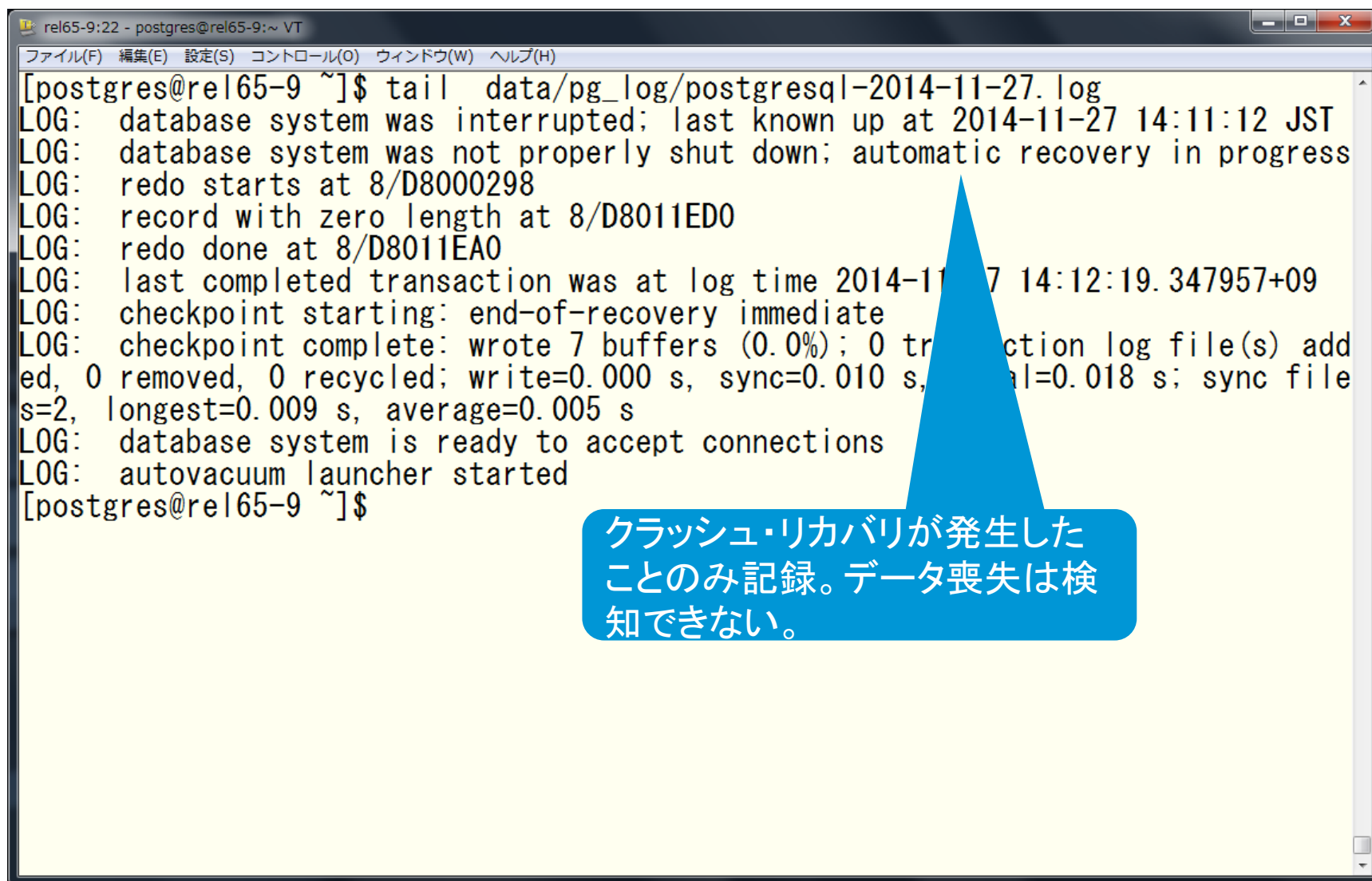
demodb=>
demodb=>
demodb=>
```

疑似クラッシュとファイル削除

エラーは発生しない  
10,000レコードは喪失

# DEMO

## クラッシュとファイル削除(更新されたテーブル)



```
rel65-9:22 - postgres@rel65-9:~ VT
ファイル(F) 編集(E) 設定(S) コントロール(O) ウィンドウ(W) ヘルプ(H)
[postgres@rel65-9 ~]$ tail data/pg_log/postgresql-2014-11-27.log
LOG: database system was interrupted; last known up at 2014-11-27 14:11:12 JST
LOG: database system was not properly shut down; automatic recovery in progress
LOG: redo starts at 8/D8000298
LOG: record with zero length at 8/D8011ED0
LOG: redo done at 8/D8011EA0
LOG: last completed transaction was at log time 2014-11-27 14:12:19.347957+09
LOG: checkpoint starting: end-of-recovery immediate
LOG: checkpoint complete: wrote 7 buffers (0.0%); 0 transaction log file(s) added, 0 removed, 0 recycled; write=0.000 s, sync=0.010 s, total=0.018 s; sync files=2, longest=0.009 s, average=0.005 s
LOG: database system is ready to accept connections
LOG: autovacuum launcher started
[postgres@rel65-9 ~]$
```

クラッシュ・リカバリが発生した  
ことのみ記録。データ喪失は検  
知できない。

# Thank you

篠田 典良  
テクノロジー事業統括  
サービス統括本部 オープンソース部  
シニアアーキテクト



[Noriyoshi.Shinoda@hp.com](mailto:Noriyoshi.Shinoda@hp.com)

日本ヒューレット・パカード株式会社  
本社  
〒136-8711  
東京都江東区大島2-2-1

