Mastering PostgreSQL Administration

BRUCE MOMJIAN,
ENTERPRISEDB

March, 2010

Abstract

POSTGRESQL is an open-source, full-featured relational database. This presentation covers advanced administration topics.

 http://momjian.us/presentations

Introduction

- Installation
- Configuration
- Maintenance
- Monitoring
- Recovery

Installation

- Click-Through installers
  - MS Windows
  - Linux
  - OS/X
  - Solaris
- Ports
  - RPM
  - DEB
  - PKG
  - other packages
- Source
  - obtaining
  - build options
  - installing

Initialization (initdb)

```
$ initdb
The files belonging to this database system will be owned by user "postgres".
This user must also own the server process.
The database cluster will be initialized with locale C.
The default database encoding has accordingly been set to SQL_ASCII.
The default text search configuration will be set to "english".
fixing permissions on existing directory /u/pgsql/data ... ok
creating subdirectories ... ok
selecting default max_connections ... 100
selecting default shared_buffers ... 32MB
creating configuration files ... ok
creating template1 database in /u/pgsql/data/base/1 ... ok
initializing pg_authid ... ok
initializing dependencies ... ok
creating system views ... ok
loading system objects' descriptions ... ok
creating conversions ... ok
creating dictionaries ... ok
setting privileges on built-in objects ... ok
creating information schema ... ok
loading PL/pgSQL server-side language ... ok
vacuuming database template1 ... ok
copying template1 to template0 ... ok
copying template1 to postgres ... ok
```

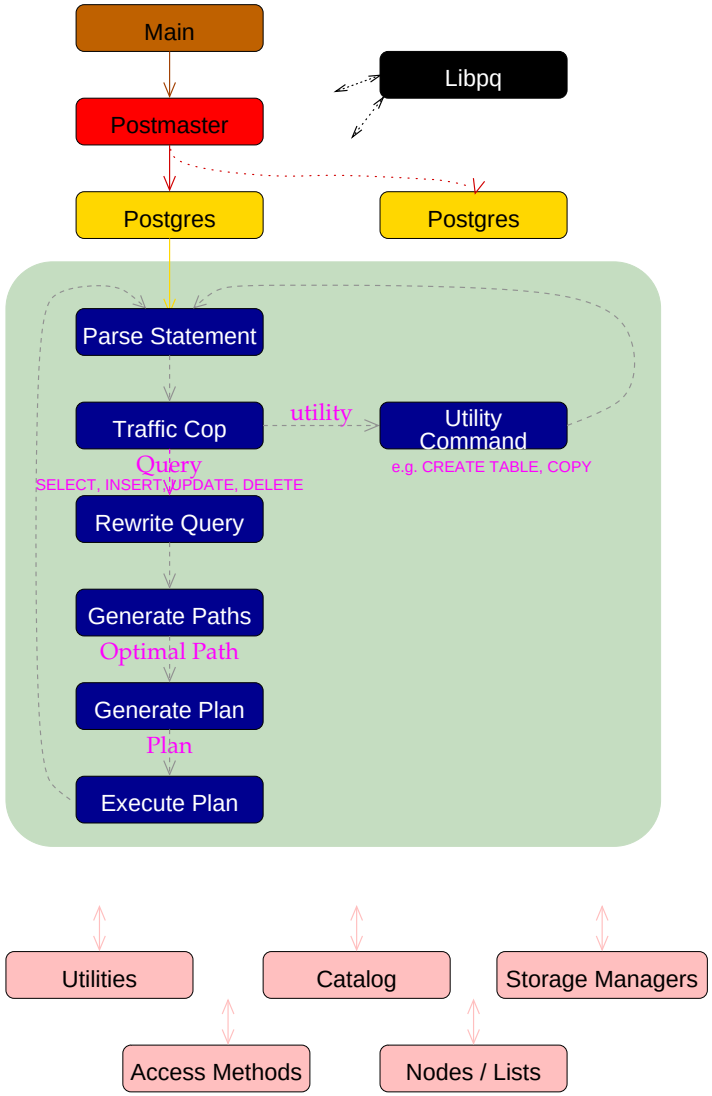
Initialization (continued)

```
WARNING: enabling "trust" authentication for local connections
You can change this by editing pg_hba.conf or using the -
A option the
next time you run initdb.
Success. You can now start the database server using:

    /u/pgsql/bin/postgres -D /u/pgsql/data

or

    /u/pgsql/bin/pg_ctl -D /u/pgsql/data -l logfile start
```

pg_controldata

```
pg_control version number:            901
Catalog version number:               201002161
Database system identifier:           5444498809957564411
Database cluster state:               in production
pg_control last modified:             03/03/10 14:56:16
Latest checkpoint location:           0/5C51C8
Prior checkpoint location:            0/5C513C
Latest checkpoint's REDO location:    0/5C51C8
Latest checkpoint's TimeLineID:       1
Latest checkpoint's NextXID:          0/663
Latest checkpoint's NextOID:          24576
Latest checkpoint's NextMultiXactId:  1
Latest checkpoint's NextMultiOffset:  0
Latest checkpoint's oldestXID:        654
Latest checkpoint's oldestXID's DB:   1
Latest checkpoint's oldestActiveXID:  0
Time of latest checkpoint:            03/03/10 14:56:16
Minimum recovery ending location:     0/0
Backup start location:                0/0
Maximum data alignment:               4
Database block size:                  8192
Blocks per segment of large relation: 131072
WAL block size:                       8192
Bytes per WAL segment:                16777216
Maximum length of identifiers:        64
Maximum columns in an index:          32
Maximum size of a TOAST chunk:        2000
```

System Architecture



Starting Postmaster

```
LOG:  database system was shut down at 2010-03-03 14:56:12 EST
LOG:  database system is ready to accept connections
LOG:  autovacuum launcher started
```

- manually
- pg_ctl start
- on boot

Stopping Postmaster

```
LOG:  received smart shutdown request
LOG:  autovacuum launcher shutting down
LOG:  shutting down
LOG:  database system is shut down
```

- manually
- pg_ctl stop
- on shutdown

Connections

- local — unix domain socket
- host — TCP/IP, both SSL or non-SSL
- hostssl — only SSL
- hostnossl — never SSL

Authentication

- trust
- reject
- passwords
  - md5
  - password (cleartext)
- local authentication
  - socket permissions
  - local socket ident
  - host ident using local identd

Authentication (continued)

- remote authentication
  - host ident using pg_ident.conf
  - kerberos
    - gss
    - sspi
  - pam
  - ldap
  - radius
  - cert

Access

- hostname and network mask
- database name
- role name (user or group)
- filename or list of databases, role
- IPv6

Pg_hba.conf

```
# TYPE  DATABASE        USER            CIDR-ADDRESS            METHOD

# "local" is for Unix domain socket connections only
local   all             all                                     trust
# IPv4 local connections:
host    all             all             127.0.0.1/32            trust
# IPv6 local connections:
host    all             all             ::1/128                 trust
```

Permissions

- host connection permissions
- role permissions
  - create roles
  - create databases
  - table permissions
- Database creation
  - template1 customization
  - system tables
  - disk space computations

Data Directory

```
$ ls -CF
PG_VERSION          pg_ident.conf       pg_tblspc/
base/               pg_multixact/       pg_twophase/
global/             pg_notify/          pg_xlog/
pg_clog/            pg_stat_tmp/        postgresql.conf
pg_hba.conf         pg_subtrans/        postmaster.opts
```

Database Directories

```
$ ls -CF global/
11604                   11785                   11797_vm
11604_fsm               11786                   11799
11604_vm                11786_fsm               11801
11606                   11786_vm                11802

$ ls -CF base/
1/      11510/  11511/  16384/

$ ls -CF base/16384
11584                   11650_vm                11742_fsm
11584_fsm               11652                   11742_vm
11584_vm                11653                   11744
11586                   11654                   11746
11586_fsm               11654_fsm               11747
```

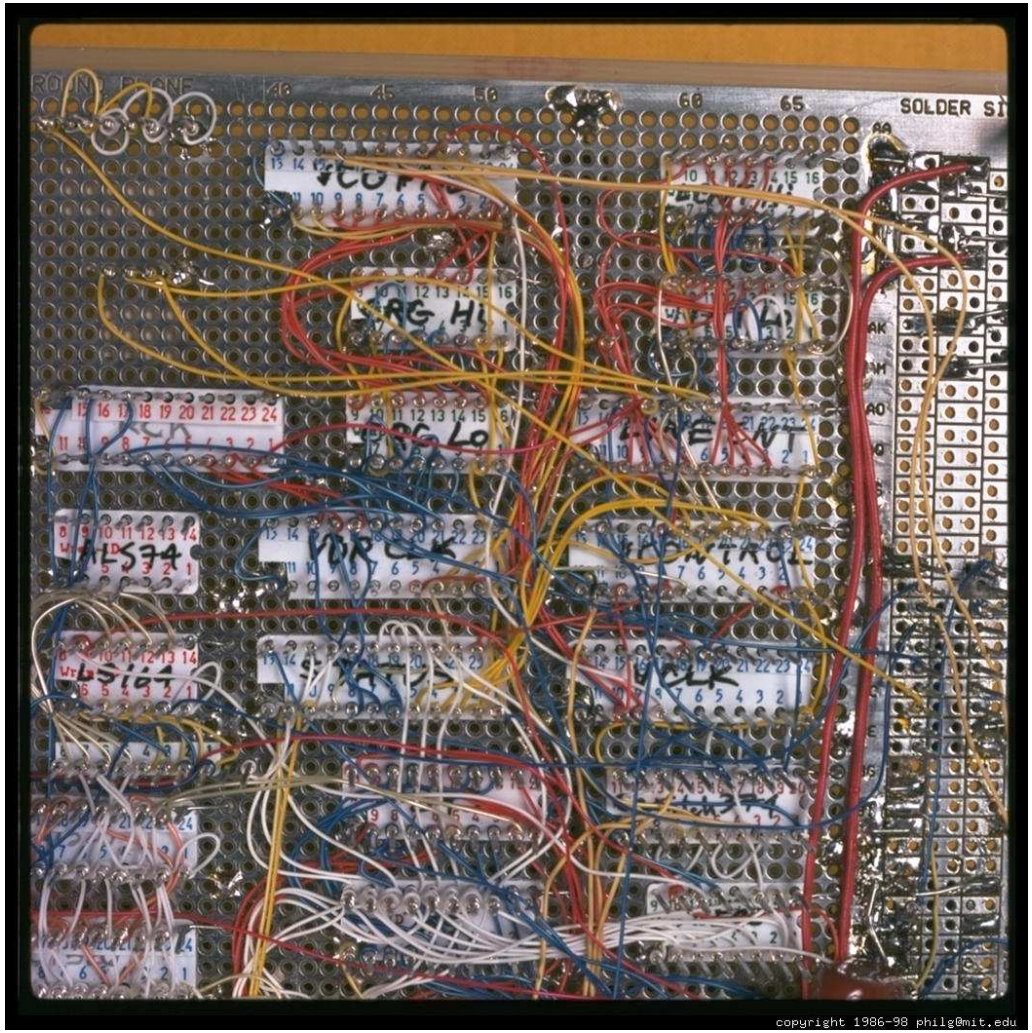
Transaction/WAL Directories

```
$ ls -CF pg_xlog/
000000010000000000000000        archive_status/
$ ls -CF pg_clog/
0000
```

Configuration Directories

```
$ ls -CF share/
conversion_create.sql   postgres.description    sql_features.txt
information_schema.sql  postgres.shdescription  system_views.sql
pg_hba.conf.sample      postgresql.conf.sample  timezone/
pg_ident.conf.sample    psqlrc.sample           timezonesets/
pg_service.conf.sample  recovery.conf.sample    tsearch_data/
postgres.bki            snowball_create.sql     unknown.pltcl
```

Configuration of postgresql.conf



postgresql.conf

```
# -----------------------------
# PostgreSQL configuration file
# -----------------------------
#
# This file consists of lines of the form:
#
#   name = value
#
# (The "=" is optional.)  Whitespace may be used.  Comments are introduced with
# "#" anywhere on a line.  The complete list of parameter names and allowed
# values can be found in the PostgreSQL documentation.
#
# The commented-out settings shown in this file represent the default values.
# Re-commenting a setting is NOT sufficient to revert it to the default value;
# you need to reload the server.
```

postgresql.conf (Continued)

```
# This file is read on server startup and when the server receives a SIGHUP
# signal.  If you edit the file on a running system, you have to SIGHUP the
# server for the changes to take effect, or use "pg_ctl reload".  Some
# parameters, which are marked below, require a server shutdown and restart to
# take effect.
#
# Any parameter can also be given as a command-line option to the server, e.g.,
# "postgres -c log_connections=on".  Some parameters can be changed at run time
# with the "SET" SQL command.
#
# Memory units:  kB = kilobytes        Time units:  ms  = milliseconds
#                MB = megabytes                     s   = seconds
#                GB = gigabytes                     min = minutes
#                                                   h   = hours
#                                                   d   = days
```

Configuration File Location

```
#data_directory = 'ConfigDir'           # use data in another directory
                                        # (change requires restart)
#hba_file = 'ConfigDir/pg_hba.conf'     # host-based authentication file
                                        # (change requires restart)
#ident_file = 'ConfigDir/pg_ident.conf' # ident configuration file
                                        # (change requires restart)
# If external_pid_file is not explicitly set, no extra PID file is written.
#external_pid_file = '(none)'           # write an extra PID file
                                        # (change requires restart)
```

Connections and Authentication

```
#listen_addresses = 'localhost'         # what IP address(es) to listen on;
                                        # comma-separated list of addresses;
                                        # defaults to 'localhost', '*' = all
                                        # (change requires restart)
#port = 5432                            # (change requires restart)
max_connections = 100                   # (change requires restart)
# Note:  Increasing max_connections costs ~400 bytes of shared memory per
# connection slot, plus lock space (see max_locks_per_transaction).
#superuser_reserved_connections = 3     # (change requires restart)
#unix_socket_directory = ''             # (change requires restart)
#unix_socket_group = ''                 # (change requires restart)
#unix_socket_permissions = 0777         # begin with 0 to use octal notation
                                        # (change requires restart)
#bonjour = off                          # advertise server via Bonjour
                                        # (change requires restart)
#bonjour_name = ''                      # defaults to the computer name
                                        # (change requires restart)
```

Security and Authentication

```
#authentication_timeout = 1min          # 1s-600s
#ssl = off                              # (change requires restart)
#ssl_ciphers = 'ALL:!ADH:!LOW:!EXP:!MD5:@STRENGTH'      # allowed SSL ciphers
                                        # (change requires restart)
#ssl_renegotiation_limit = 512MB        # amount of data between renegotiations
#password_encryption = on
#db_user_namespace = off
# Kerberos and GSSAPI
#krb_server_keyfile = ''
#krb_srvname = 'postgres'               # (Kerberos only)
#krb_caseins_users = off
```

TCP/IP Control

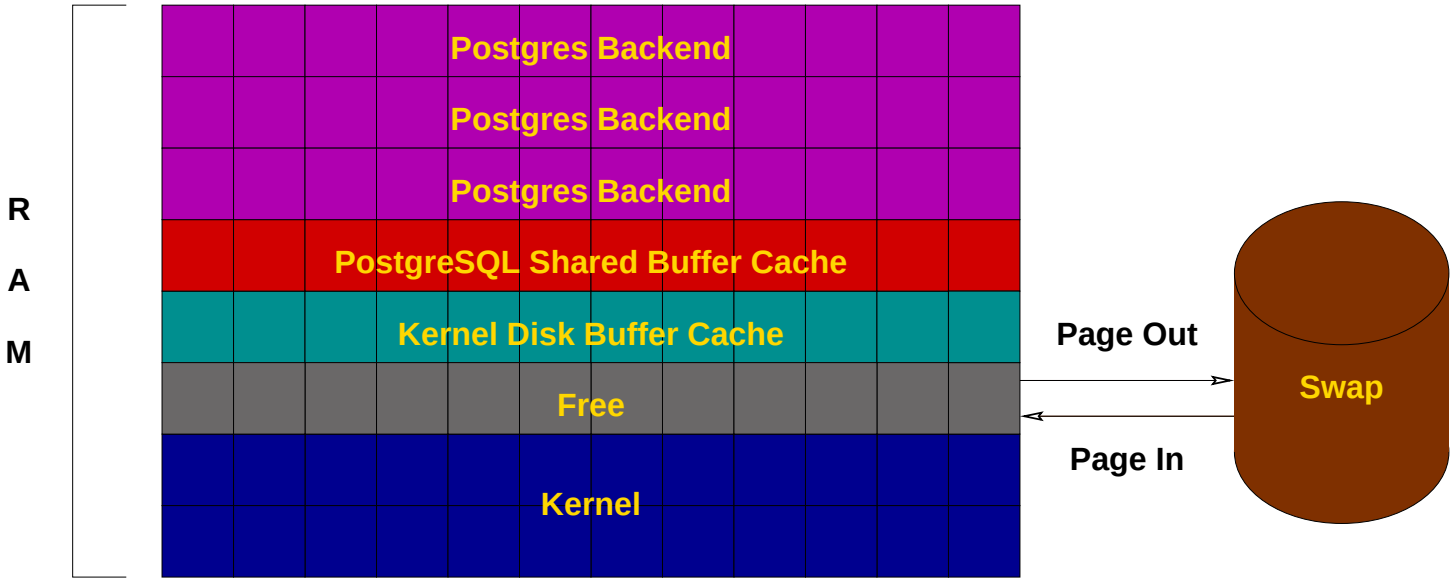
```
#tcp_keepalives_idle = 0                # TCP_KEEPIDLE, in seconds;
                                        # 0 selects the system default
#tcp_keepalives_interval = 0            # TCP_KEEPINTVL, in seconds;
                                        # 0 selects the system default
#tcp_keepalives_count = 0               # TCP_KEEPCNT;
                                        # 0 selects the system default
```

Memory Usage

```
shared_buffers = 32MB                   # min 128kB
                                        # (change requires restart)
#temp_buffers = 8MB                     # min 800kB
#max_prepared_transactions = 0          # zero disables the feature
                                        # (change requires restart)
# Note:  Increasing max_prepared_transactions costs ~600 bytes of shared memory
# per transaction slot, plus lock space (see max_locks_per_transaction).
# It is not advisable to set max_prepared_transactions nonzero unless you
# actively intend to use prepared transactions.
#work_mem = 1MB                         # min 64kB
#maintenance_work_mem = 16MB            # min 1MB
#max_stack_depth = 2MB                  # min 100kB
```

Kernel changes often required.

Sizing Shared Memory



Kernel Resources

```
#max_files_per_process = 1000          # min 25
                                       # (change requires restart)
#shared_preload_libraries = ''         # (change requires restart)
```

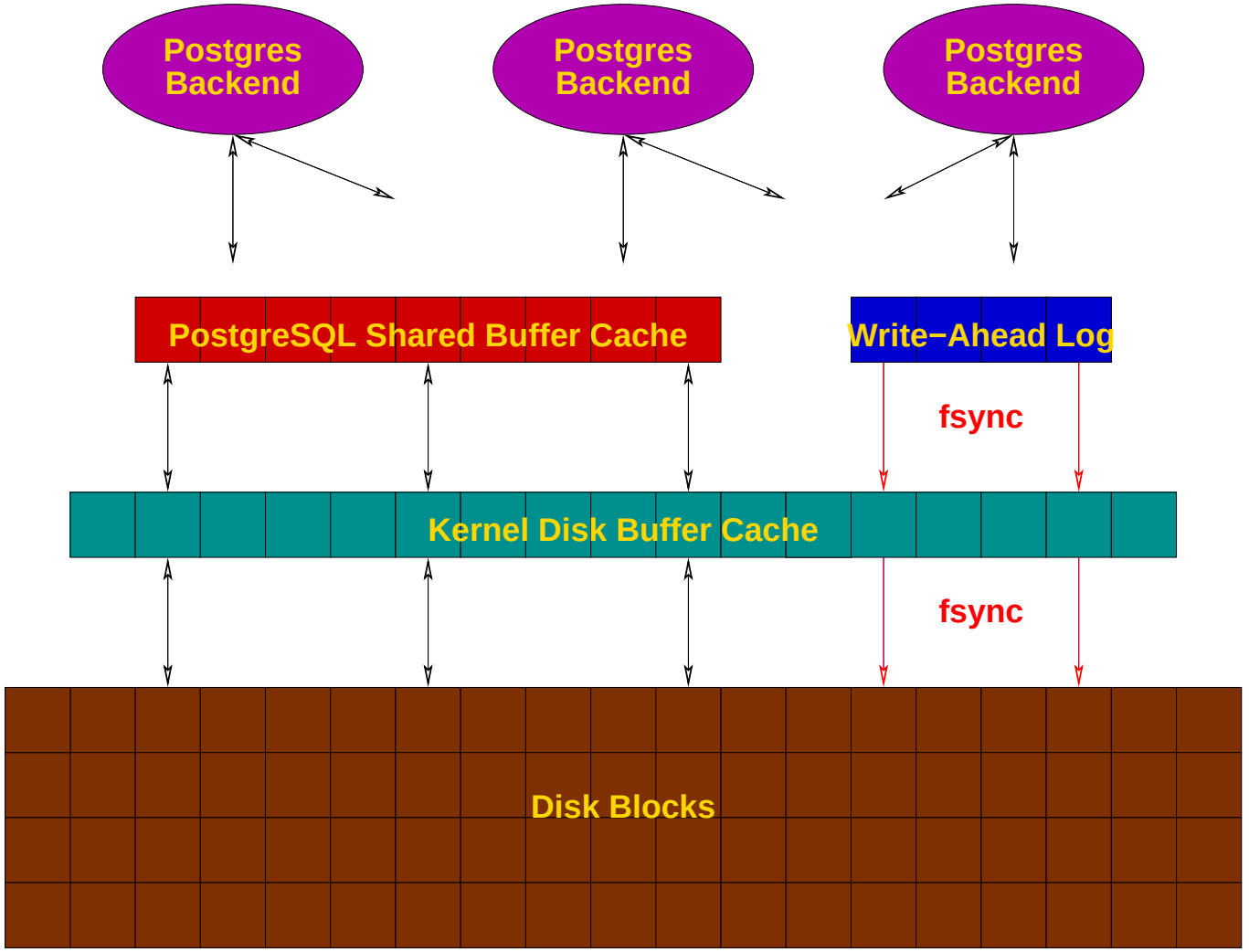
Vacuum and Background Writer

```
# - Cost-Based Vacuum Delay -

#vacuum_cost_delay = 0ms                # 0-100 milliseconds
#vacuum_cost_page_hit = 1               # 0-10000 credits
#vacuum_cost_page_miss = 10             # 0-10000 credits
#vacuum_cost_page_dirty = 20            # 0-10000 credits
#vacuum_cost_limit = 200                # 1-10000 credits

# - Background Writer -

#bgwriter_delay = 200ms                 # 10-10000ms between rounds
#bgwriter_lru_maxpages = 100            # 0-1000 max buffers written/round
#bgwriter_lru_multiplier = 2.0          # 0-
10.0 multipler on buffers scanned/round

# - Asynchronous Behavior -

#effective_io_concurrency = 1           # 1-1000. 0 disables prefetching
```

Write-Ahead Log (WAL)

```
#fsync = on                              # turns forced synchronization on or off
#synchronous_commit = on                # immediate fsync at commit
#wal_sync_method = fsync                # the default is the first option
                                        # supported by the operating system:
                                        #   open_datasync
                                        #   fdatasync
                                        #   fsync
                                        #   fsync_writethrough
                                        #   open_sync
#full_page_writes = on                  # recover from partial page writes
#wal_buffers = 64kB                     # min 32kB
                                        # (change requires restart)
#wal_writer_delay = 200ms               # 1-10000 milliseconds
#commit_delay = 0                       # range 0-100000, in microseconds
#commit_siblings = 5                    # range 1-1000
```

Write-Ahead Logging (Continued)



Checkpoints and Archiving

```
# - Checkpoints -

#checkpoint_segments = 3                # in logfile segments, min 1, 16MB each
#checkpoint_timeout = 5min              # range 30s-1h
#checkpoint_completion_target = 0.5     # checkpoint target duration, 0.0 - 1.0
#checkpoint_warning = 30s               # 0 is off

# - Archiving -

#archive_mode = off             # allows archiving to be done
                                # (change requires restart)
#archive_command = ''           # command to use to archive a logfile segment
#archive_timeout = 0            # force a logfile segment switch after this
                                # time; 0 is off
```

Hot Standby and Replication

```
# - Hot Standby -

#recovery_connections = on      # allows connections during recovery
#max_standby_delay = 30         # max acceptable standby lag (s) to allow queries
                                # to complete without conflict; -1 disables
# - Replication -

#max_wal_senders = 0            # max number of walsender processes
#wal_sender_delay = 200ms       # 1-10000 milliseconds
```

Query Tuning (1)

```
# - Planner Method Configuration -

#enable_bitmapscan = on
#enable_hashagg = on
#enable_hashjoin = on
#enable_indexscan = on
#enable_mergejoin = on
#enable_nestloop = on
#enable_seqscan = on
#enable_sort = on
#enable_tidscan = on
```

Query Tuning (2)

```
# - Planner Cost Constants -

#seq_page_cost = 1.0                    # measured on an arbitrary scale
#random_page_cost = 4.0                 # same scale as above
#cpu_tuple_cost = 0.01                  # same scale as above
#cpu_index_tuple_cost = 0.005           # same scale as above
#cpu_operator_cost = 0.0025             # same scale as above
#effective_cache_size = 128MB
```

Query Tuning (3)

```
# - Genetic Query Optimizer -

#geqo = on
#geqo_threshold = 12
#geqo_effort = 5                        # range 1-10
#geqo_pool_size = 0                     # selects default based on effort
#geqo_generations = 0                   # selects default based on effort
#geqo_selection_bias = 2.0              # range 1.5-2.0
#geqo_seed = 0.0                        # range 0.0-1.0
```

Query Tuning (4)

```
# - Other Planner Options -

#default_statistics_target = 100        # range 1-10000
#constraint_exclusion = partition       # on, off, or partition
#cursor_tuple_fraction = 0.1            # range 0.0-1.0
#from_collapse_limit = 8
#join_collapse_limit = 8                # 1 disables collapsing of explicit
                                        # JOIN clauses
```

Where To Log (1)

```
#log_destination = 'stderr'             # Valid values are combinations of
                                        # stderr, csvlog, syslog and eventlog,
                                        # depending on platform.  csvlog
                                        # requires logging_collector to be on.
# This is used when logging to stderr:
#logging_collector = off                # Enable capturing of stderr and csvlog
                                        # into log files. Required to be on for
                                        # csvlogs.
                                        # (change requires restart)
# These are only used if logging_collector is on:
#log_directory = 'pg_log'               # directory where log files are written,
                                        # can be absolute or relative to PGDATA
#log_filename = 'postgresql-%Y-%m-%d_%H%M%S.log'        # log file name pattern,
                                        # can include strftime() escapes
```

Where To Log (2)

```
#log_truncate_on_rotation = off          # If on, an existing log file of the
                                         # same name as the new log file will be
                                         # truncated rather than appended to.
                                         # But such truncation only occurs on
                                         # time-driven rotation, not on restarts
                                         # or size-driven rotation.  Default is
                                         # off, meaning append to existing files
                                         # in all cases.
#log_rotation_age = 1d                   # Automatic rotation of logfiles will
                                         # happen after that time.  0 disables.
#log_rotation_size = 10MB                # Automatic rotation of logfiles will
                                         # happen after that much log output.
                                         # 0 disables.
```

Where to Log (3)

```
# These are relevant when logging to syslog:
#syslog_facility = 'LOCALO'
#syslog_ident = 'postgres'
#silent_mode = off                          # Run server silently.
                                            # DO NOT USE without syslog or
                                            # logging_collector
                                            # (change requires restart)
```

When to Log

```
#client_min_messages = notice          # values in order of decreasing detail:
                                       #   debug5
                                       #   debug4
                                       #   debug3
                                       #   debug2
                                       #   debug1
                                       #   log
                                       #   notice
                                       #   warning
                                       #   error
#log_min_messages = warning            # values in order of decreasing detail:
                                       #   debug5
                                       #   debug4
                                       #   debug3
                                       #   debug2
                                       #   debug1
                                       #   info
                                       #   notice
                                       #   warning
                                       #   error
                                       #   log
                                       #   fatal
                                       #   panic
```

When to Log (Continued)

```
#log_min_error_statement = error        # values in order of decreasing detail:
                                        #   debug5
                                        #   debug4
                                        #   debug3
                                        #   debug2
                                        #   debug1
                                        #   info
                                        #   notice
                                        #   warning
                                        #   error
                                        #   log
                                        #   fatal
                                        #   panic (effectively off)
#log_min_duration_statement = -1        # -1 is disabled, 0 logs all statements
                                        # and their durations, > 0 logs only
                                        # statements running at least this number
                                        # of milliseconds
```

What to Log

```
#debug_print_parse = off
#debug_print_rewritten = off
#debug_print_plan = off
#debug_pretty_print = on
#log_checkpoints = off
#log_connections = off
#log_disconnections = off
#log_duration = off
#log_error_verbosity = default          # terse, default, or verbose messages
#log_hostname = off
```

What To Log: Log_line_prefix

```
#log_line_prefix = ''
```

```
# special values:
#   %a = application name
#   %u = user name
#   %d = database name
#   %r = remote host and port
#   %h = remote host
#   %p = process ID
#   %t = timestamp without milliseconds
#   %m = timestamp with milliseconds
#   %i = command tag
#   %e = SQL state
#   %c = session ID
#   %l = session line number
#   %s = session start timestamp
#   %v = virtual transaction ID
#   %x = transaction ID (0 if none)
#   %q = stop here in non-session
#        processes
#   %% = '%'
# e.g. '<%u%%%d> '
```

What to Log (Continued)

```
#log_lock_waits = off                   # log lock waits >= deadlock_timeout
#log_statement = 'none'                 # none, ddl, mod, all
#log_temp_files = -1                    # log temporary files equal or larger
                                        # than the specified size in kilobytes;
                                        # -1 disables, 0 logs all temp files
#log_timezone = unknown                 # actually, defaults to TZ environment
                                        # setting
```

Runtime Statistics

```
# - Query/Index Statistics Collector -

#track_activities = on
#track_counts = on
#track_functions = none                 # none, pl, all
#track_activity_query_size = 1024
#update_process_title = on
#stats_temp_directory = 'pg_stat_tmp'

# - Statistics Monitoring -

#log_parser_stats = off
#log_planner_stats = off
#log_executor_stats = off
#log_statement_stats = off
```

Autovacuum

```
#autovacuum = on                        # Enable autovacuum subprocess?  'on'
                                        # requires track_counts to also be on.
#log_autovacuum_min_duration = -1       # -1 disables, 0 logs all actions and
                                        # their durations, > 0 logs only
                                        # actions running at least this number
                                        # of milliseconds.
#autovacuum_max_workers = 3             # max number of autovacuum subprocesses
#autovacuum_naptime = 1min              # time between autovacuum runs
#autovacuum_vacuum_threshold = 50       # min number of row updates before
                                        # vacuum
#autovacuum_analyze_threshold = 50      # min number of row updates before
                                        # analyze
#autovacuum_vacuum_scale_factor = 0.2   # fraction of table size before vacuum
#autovacuum_analyze_scale_factor = 0.1  # fraction of table size before analyze
#autovacuum_freeze_max_age = 200000000  # maximum XID age before forced vacuum
                                        # (change requires restart)
#autovacuum_vacuum_cost_delay = 20ms    # default vacuum cost delay for
                                        # autovacuum, in milliseconds;
                                        # -1 means use vacuum_cost_delay
#autovacuum_vacuum_cost_limit = -1      # default vacuum cost limit for
                                        # autovacuum, -1 means use
                                        # vacuum_cost_limit
```

Statement Behavior

```
#search_path = '"$user",public'          # schema names
#default_tablespace = ''                 # a tablespace name, '' uses the default
#temp_tablespaces = ''                   # a list of tablespace names, '' uses
                                         # only default tablespace
#check_function_bodies = on
#default_transaction_isolation = 'read committed'
#default_transaction_read_only = off
#session_replication_role = 'origin'
#statement_timeout = 0                   # in milliseconds, 0 is disabled
#vacuum_freeze_min_age = 50000000
#vacuum_freeze_table_age = 150000000
#bytea_output = 'hex'                    # hex, escape
#xmlbinary = 'base64'
#xmloption = 'content'
```

Locale and Formatting

```
datestyle = 'iso, mdy'
#intervalstyle = 'postgres'
#timezone = unknown                    # actually, defaults to TZ environment
                                       # setting
#timezone_abbreviations = 'Default'    # Select the set of available time zone
                                       # abbreviations.  Currently, there are
                                       #   Default
                                       #   Australia
                                       #   India
                                       # You can create your own file in
                                       # share/timezonesets/.
#extra_float_digits = 0                # min -15, max 3
#client_encoding = sql_ascii           # actually, defaults to database
                                       # encoding
# These settings are initialized by initdb, but they can be changed.
lc_messages = 'C'                      # locale for system error message
                                       # strings
lc_monetary = 'C'                      # locale for monetary formatting
lc_numeric = 'C'                       # locale for number formatting
lc_time = 'C'                          # locale for time formatting
```

Full Text Search

```
# default configuration for text search
default_text_search_config = 'pg_catalog.english'
```

Other Defaults

```
#dynamic_library_path = '$libdir'
#local_preload_libraries = ''
```

Lock Management

```
#deadlock_timeout = 1s
#max_locks_per_transaction = 64         # min 10
                                        # (change requires restart)
# Note:  Each lock table slot uses ~270 bytes of shared memory, and there are
# max_locks_per_transaction * (max_connections + max_prepared_transactions)
# lock table slots.
```

Version/Platform Compatibility

```
# - Previous PostgreSQL Versions -

#array_nulls = on
#backslash_quote = safe_encoding        # on, off, or safe_encoding
#default_with_oids = off
#escape_string_warning = on
#lo_compat_privileges = off
#sql_inheritance = on
#standard_conforming_strings = off
#synchronize_seqscans = on

# - Other Platforms and Clients -

#transform_null_equals = off
```

Customization

```
#custom_variable_classes = ''           # list of custom variable class names
```

Interfaces

- Installing
  - Compiled Languages (C, ecpg)
  - Scripting Language (Perl, Python, PHP)
  - SPI
- Connection Pooling

Include Files

```
$ ls -CF include/
ecpg_config.h          libpq-fe.h             postgres_ext.h
ecpg_informix.h        pg_config.h            server/
ecpgerrno.h            pg_config_manual.h     sql3types.h
ecpglib.h              pg_config_os.h         sqlca.h
ecpgtype.h             pgtypes_date.h         sqlda-compat.h
informix/              pgtypes_error.h        sqlda-native.h
internal/              pgtypes_interval.h     sqlda.h
libpq/                 pgtypes_numeric.h
libpq-events.h         pgtypes_timestamp.h
```

Library Files

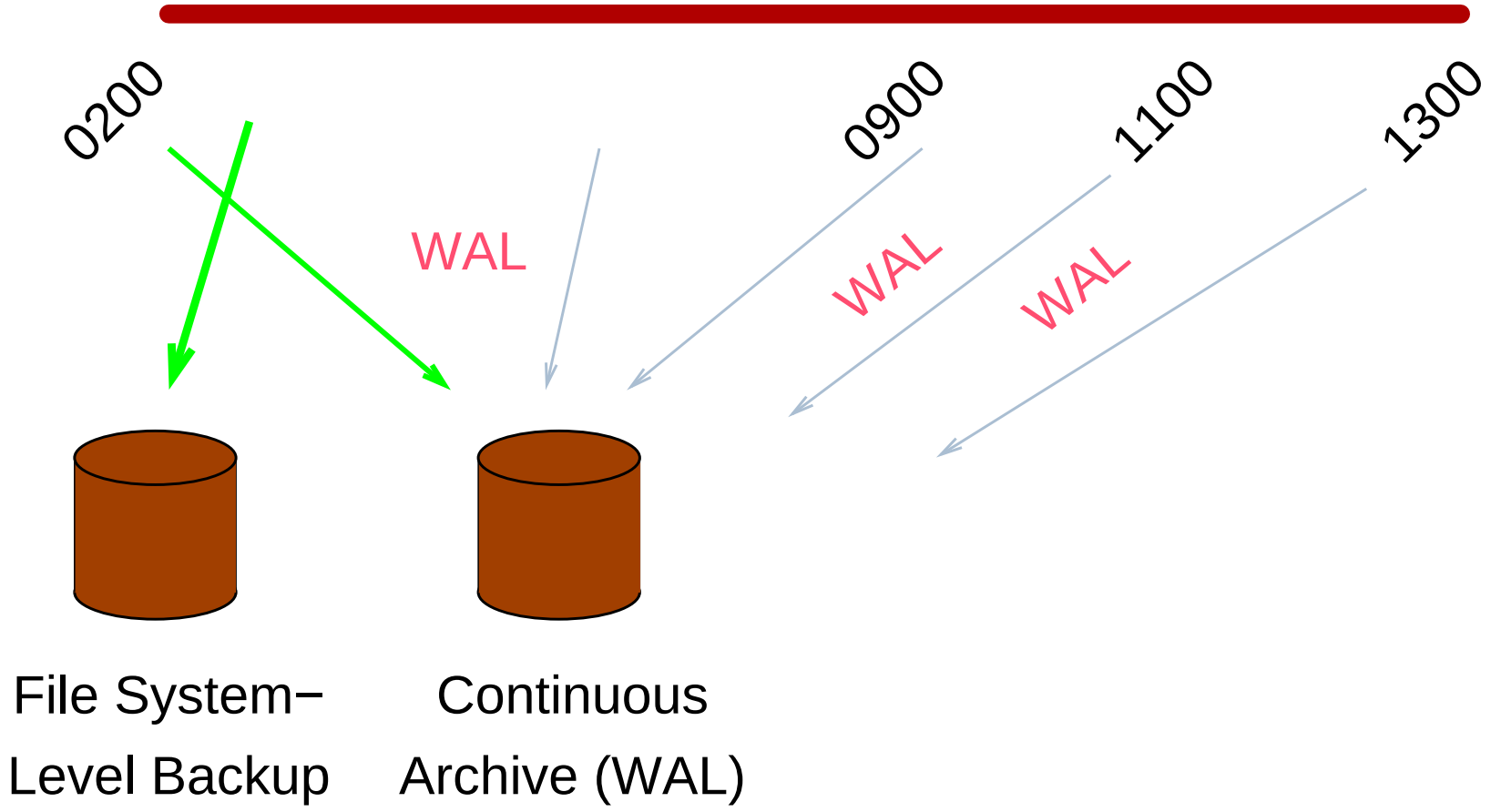
```
$ ls -CF lib/
ascii_and_mic.so*       libecpg_compat.so.3.2*  utf8_and_cyrillic.so*
cyrillic_and_mic.so*    libpgport.a             utf8_and_euc2004.so*
dict_snowball.so*       libpgtypes.a            utf8_and_euc_cn.so*
euc2004_sjis2004.so*    libpgtypes.so@          utf8_and_euc_jp.so*
euc_cn_and_mic.so*      libpgtypes.so.3@        utf8_and_euc_kr.so*
euc_jp_and_sjis.so*     libpgtypes.so.3.1*      utf8_and_euc_tw.so*
euc_kr_and_mic.so*      libpq.a                 utf8_and_gb18030.so*
euc_tw_and_big5.so*     libpq.so@               utf8_and_gbk.so*
latin2_and_win1250.so*  libpq.so.5@             utf8_and_iso8859.so*
latin_and_mic.so*       libpq.so.5.3*           utf8_and_iso8859_1.so*
libecpg.a               libpqwalreceiver.so*    utf8_and_johab.so*
libecpg.so@             pgxs/                   utf8_and_sjis.so*
libecpg.so.6@           plperl.so*              utf8_and_sjis2004.so*
libecpg.so.6.2*         plpgsql.so*             utf8_and_uhc.so*
libecpg_compat.a        pltcl.so*               utf8_and_win.so*
libecpg_compat.so@      utf8_and_ascii.so*
libecpg_compat.so.3@    utf8_and_big5.so*
```

Maintenance

Backup

- File system-level (physical)
  - tar, cpio while shutdown
  - file system snapshot
  - rsync, shutdown, rsync, restart
- pg_dump/pg_dumpall (logical)
- Restore/pg_restore with custom format

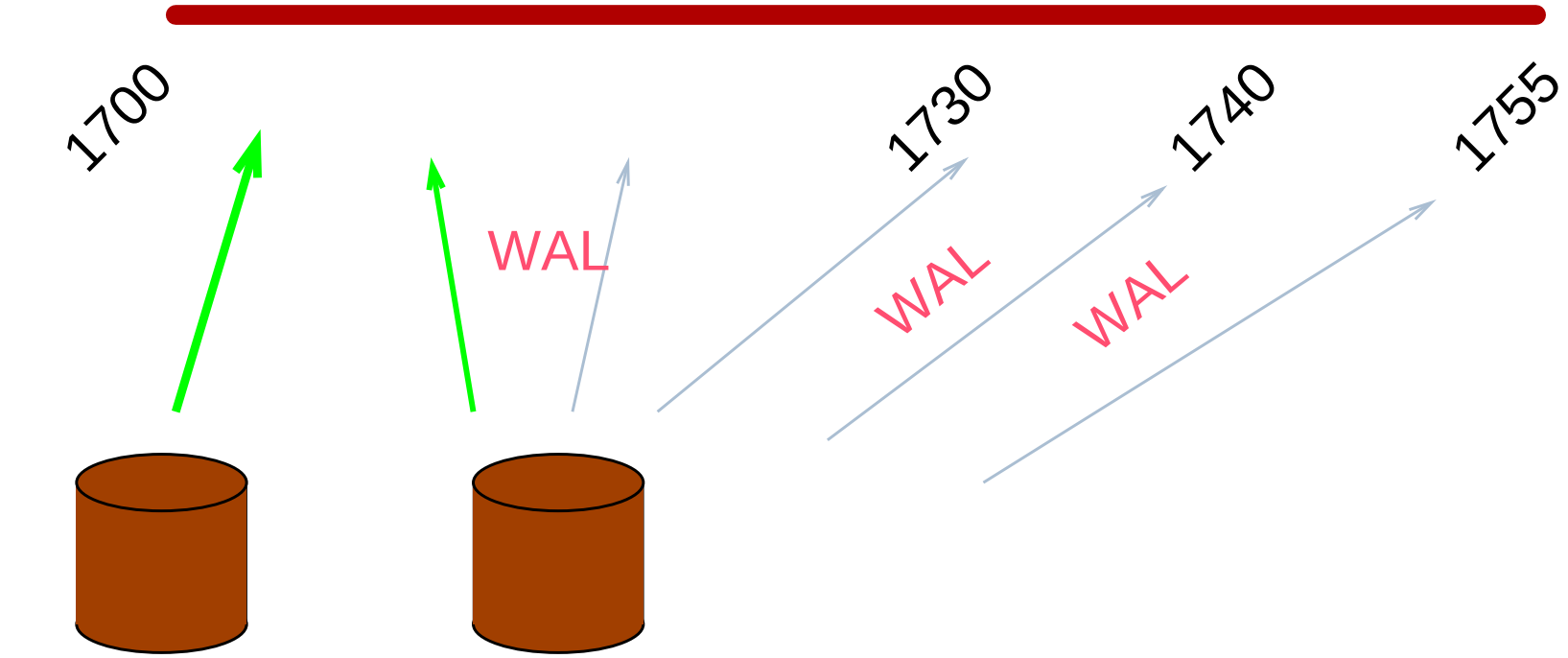
Continuous Archiving / Point-In-Time Recovery (PITR)



PITR Backup Procedures

1. `archive_command = 'cp -i %p /mnt/server/pgsql/%f < /dev/null'`
2. `SELECT pg_start_backup('label');`
3. Perform file system-level backup (can be inconsistent)
4. `SELECT pg_stop_backup();`

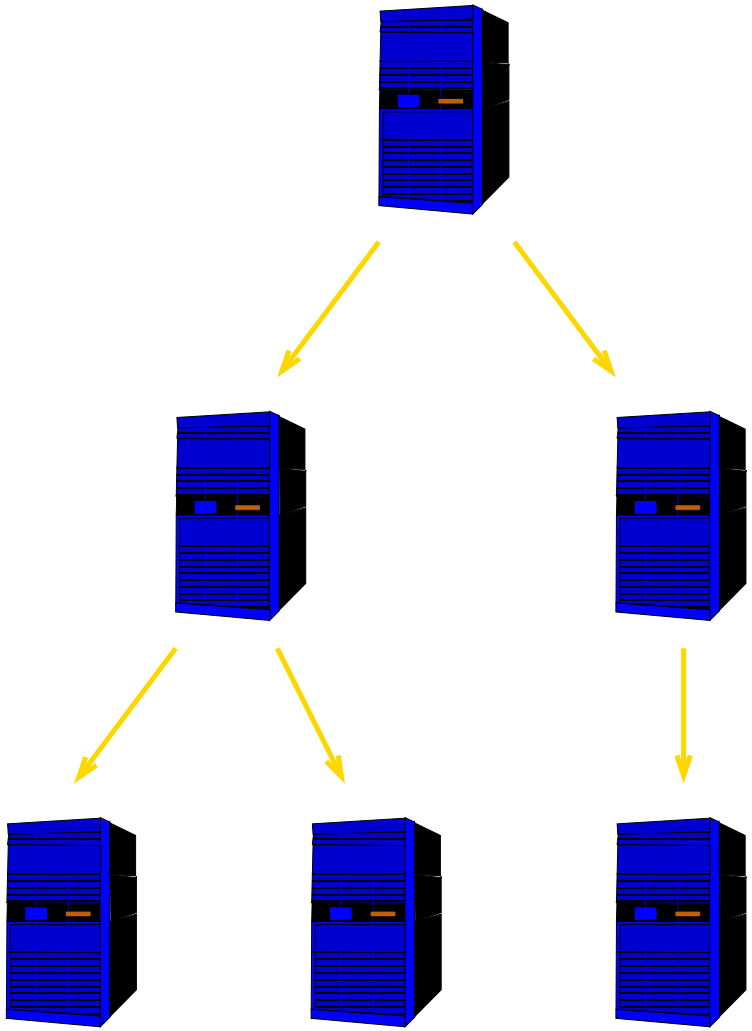
PITR Recovery



PITR Recovery Proceedures

1. Stop postmaster
2. Restore file system-level backup
3. Make adjustments as outlined in the documentation
4. Create recovery.conf
5. `restore_command = 'cp /mnt/server/pgsql/%f %p'`
6. Start the postmaster

Master-Slave Replication - Slony



Other Solutions

- Mutli-master replication: Bucardo, PgCluster
- Pooling: PgPool II, Londiste (Skytools)

Data Maintenance

- VACUUM (nonblocking) records free space into .fsm (free space map) files
- ANALYZE collects optimizer statistics
- VACUUM FULL (blocking) shrinks the size of database disk files

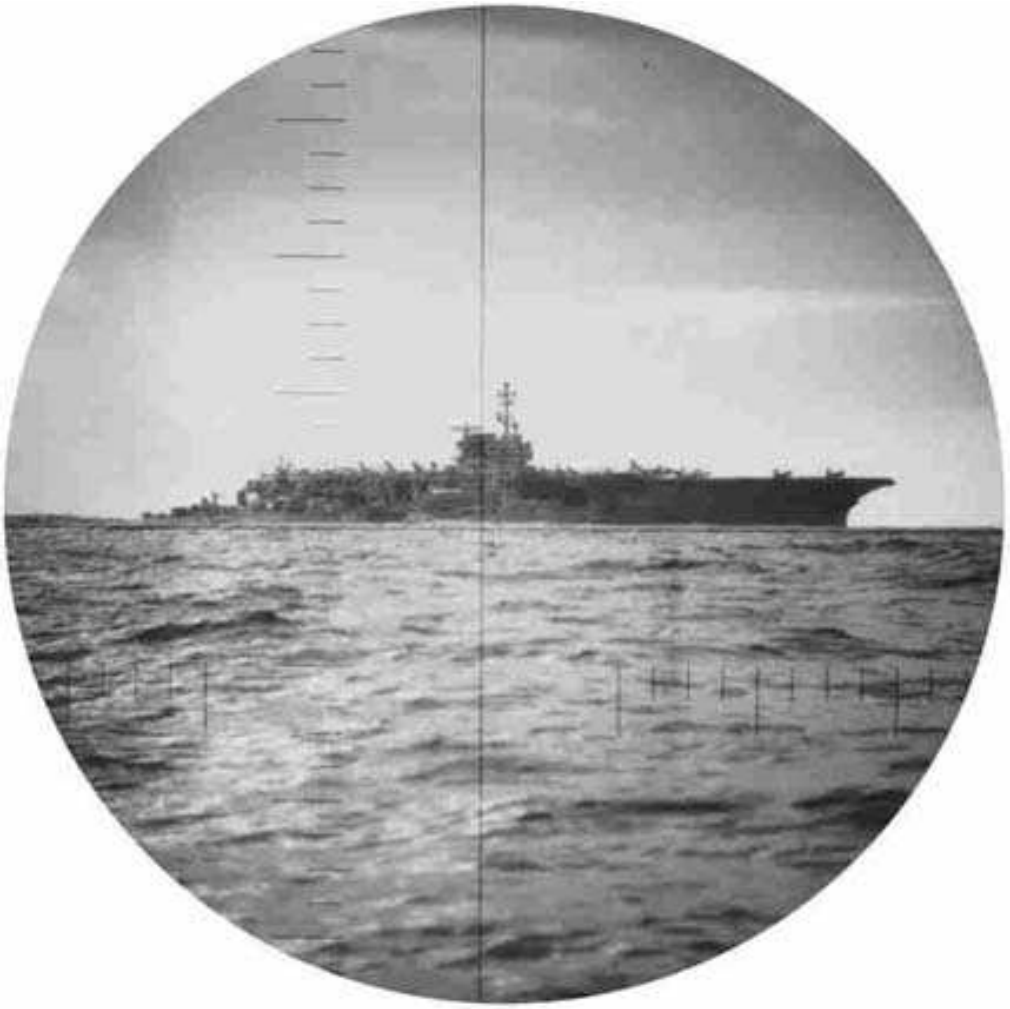
Automating Tasks

Autovacuum handles vacuum and analyze tasks automatically.

Checkpoints

- Write all dirty shared buffers
- Sync all dirty kernel buffers
- Recycle WAL files
- Check for server messages indicating too-frequent checkpoints
- If so, increase *checkpoint_segments*

Monitoring Active Sessions



ps

```
$ ps -Upostgres
postgres   825     1  0 Tue12AM  ??         0:06.57 /u/pgsql/bin/postmaster -i
postgres   829   825  0 Tue12AM  ??         0:35.03 writer process    (postmaster)
postgres   830   825  0 Tue12AM  ??         0:16.07 wal writer process    (postmaster)
postgres   831   825  0 Tue12AM  ??         0:11.34 autovacuum launcher process    (postmaster)
postgres   832   825  0 Tue12AM  ??         0:07.63 stats collector process    (postmaster)
postgres 13003   825  0  3:44PM  ??         0:00.01 postgres test [local] idle (postmaster)
postgres 13002 12997  0  3:44PM  ttyq1      0:00.03 /u/pgsql/bin/psql test
```

top

```
$ top
load averages:  0.56,  0.39,  0.36                                    18:25:58
138 processes: 5 running, 130 sleeping, 3 zombie
CPU states: 50.0% user,  0.0% nice,  0.0% system,  0.0% interrupt, 50.0% idle
Memory: Real: 96M/133M Virt: 535M/1267M Free: 76M

  PID USERNAME PRI NICE  SIZE   RES STATE   TIME   WCPU    CPU COMMAND
23785 postgres  57    0   11M 5336K run/0   0:07 30.75% 30.66% postmaster
23784 postgres   2    0   10M   11M sleep   0:00  2.25%  2.25% psql
```

Query Monitoring

```
test=> SELECT * FROM pg_stat_activity;
-[ RECORD 1 ]-+-------------------------------
datid         | 16384
datname       | test
procpid       | 2373
usesysid      | 10
usename       | postgres
current_query | select * from pg_stat_activity;
waiting       | f
xact_start    | 2009-01-27 12:12:59.917695-05
query_start   | 2009-01-27 12:12:59.917695-05
backend_start | 2009-01-27 12:12:33.422001-05
client_addr   |
client_port   | -1
```

Access Statistics

```
pg_stat_all_indexes      | view    | postgres
pg_stat_all_tables       | view    | postgres
pg_stat_database         | view    | postgres
pg_stat_sys_indexes      | view    | postgres
pg_stat_sys_tables       | view    | postgres
pg_stat_user_indexes     | view    | postgres
pg_stat_user_tables      | view    | postgres
pg_statio_all_indexes    | view    | postgres
pg_statio_all_sequences  | view    | postgres
pg_statio_all_tables     | view    | postgres
pg_statio_sys_indexes    | view    | postgres
pg_statio_sys_sequences  | view    | postgres
pg_statio_sys_tables     | view    | postgres
pg_statio_user_indexes   | view    | postgres
pg_statio_user_sequences | view    | postgres
pg_statio_user_tables    | view    | postgres
```

Database Statistics

```
test=> SELECT * FROM pg_stat_database;
...
-[ RECORD 4 ]-+----------
datid         | 16384
datname       | test
numbackends   | 1
xact_commit   | 188
xact_rollback | 0
blks_read     | 95
blks_hit      | 11832
tup_returned  | 64389
tup_fetched   | 2938
tup_inserted  | 0
tup_updated   | 0
tup_deleted   | 0
```

Table Activity

```
test=> SELECT * FROM pg_stat_all_tables;
-[ RECORD 10 ]---+------------------------
relid            | 2616
schemaname       | pg_catalog
relname          | pg_opclass
seq_scan         | 2
seq_tup_read     | 2
idx_scan         | 99
idx_tup_fetch    | 99
n_tup_ins        | 0
n_tup_upd        | 0
n_tup_del        | 0
n_tup_hot_upd    | 0
n_live_tup       | 0
n_dead_tup       | 0
last_vacuum      |
last_autovacuum  |
last_analyze     |
last_autoanalyze |
```

Table Block Activity

```
test=> SELECT * FROM pg_statio_all_tables;
-[ RECORD 50 ]--+-----------------------
relid           | 2602
schemaname      | pg_catalog
relname         | pg_amop
heap_blks_read  | 3
heap_blks_hit   | 114
idx_blks_read   | 5
idx_blks_hit    | 303
toast_blks_read |
toast_blks_hit  |
tidx_blks_read  |
tidx_blks_hit   |
```

Analyzing Activity

- Heavily used tables
- Unnecessary indexes
- Additional indexes
- Index usage
- TOAST usage

CPU

```
$ vmstat 5
 procs      memory      page                    disks      faults      cpu
 r  b w    avm    fre  flt  re  pi  po  fr  sr  s0  s0   in   sy  cs us sy  id
 1  0 0 501820  48520 1234  86   2   0   0   3   5   0  263 2881 599 10  4  86
 3  0 0 512796  46812 1422 201  12   0   0   0   3   0  259 6483 827  4  7  88
 3  0 0 542260  44356  788 137   6   0   0   0   8   0  286 5698 741  2  5  94
 4  0 0 539708  41868  576  65  13   0   0   0   4   0  273 5721 819 16  4  80
 4  0 0 547200  32964  454   0   0   0   0   0   5   0  253 5736 948 50  4  46
 4  0 0 556140  23884  461   0   0   0   0   0   2   0  249 5917 959 52  3  44
 1  0 0 535136  46280 1056 141  25   0   0   0   2   0  261 6417 890 24  6  70
```

I/O

```
$ iostat 5
      tty             sd0             sd1             sd2                 % cpu
 tin tout   sps   tps msps   sps   tps msps   sps   tps msps  usr nic sys int idl
   7  119   244    11  6.1     0     0 27.3     0     0 18.1    9   1   4   0  86
   0   86    20     1  1.4     0     0  0.0     0     0  0.0    2   0   2   0  96
   0   82    61     4  3.6     0     0  0.0     0     0  0.0    2   0   2   0  97
   0   65     6     0  0.0     0     0  0.0     0     0  0.0    1   0   2   0  97
  12   90    31     2  5.4     0     0  0.0     0     0  0.0    4   0   3   0  93
  24  173     6     0  4.9     0     0  0.0     0     0  0.0   48   0   3   0  49
   0   91  3594    63  4.6     0     0  0.0     0     0  0.0   11   0   4   0  85
```

Disk Usage

```
test=> \df *size*
                                  List of functions
   Schema   |          Name          | Result data type | Argument data types |  Type
------------+------------------------+------------------+---------------------+--------
 pg_catalog | pg_column_size         | integer          | "any"               | normal
 pg_catalog | pg_database_size       | bigint           | name                | normal
 pg_catalog | pg_database_size       | bigint           | oid                 | normal
 pg_catalog | pg_indexes_size        | bigint           | regclass            | normal
 pg_catalog | pg_relation_size       | bigint           | regclass            | normal
 pg_catalog | pg_relation_size       | bigint           | regclass, text      | normal
 pg_catalog | pg_size_pretty         | text             | bigint              | normal
 pg_catalog | pg_table_size          | bigint           | regclass            | normal
 pg_catalog | pg_tablespace_size     | bigint           | name                | normal
 pg_catalog | pg_tablespace_size     | bigint           | oid                 | normal
 pg_catalog | pg_total_relation_size | bigint           | regclass            | normal
(11 rows)
```

Database File Mapping - oid2name

```
$ oid2name
All databases:
----------------------------------
18720  = test1
1      = template1
18719  = template0
18721  = test
18735  = postgres
18736  = cssi
```

Table File Mapping

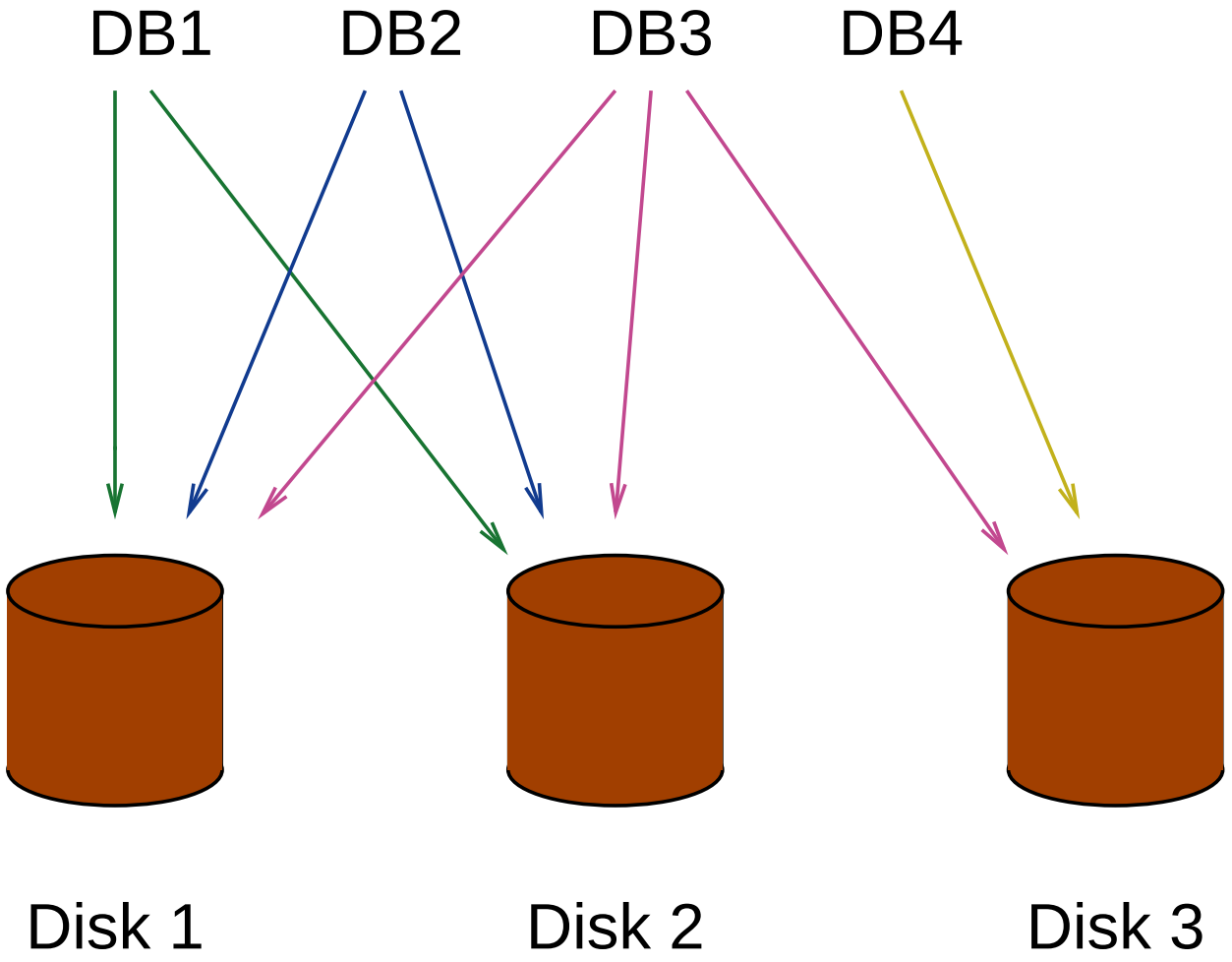
```
$ cd /usr/local/pgsql/data/base
$ oid2name
All databases:
----------------------------------
16817  = test2
16578  = x
16756  = test
1      = template1
16569  = template0
16818  = test3
16811  = floattest

$ cd 16756
$ ls 1873*
18730   18731   18732   18735   18736   18737   18738   18739
```

```
$ oid2name -d test -o 18737
Tablename of oid 18737 from database "test":
---------------------------------
18737  = ips

$ oid2name -d test -t ips
Oid of table ips from database "test":
---------------------------------
18737  = ips

$ # show disk usage per database
$ cd /usr/local/pgsql/data/base
$ du -s * |
> while read SIZE OID
> do
>       echo "$SIZE      'oid2name -q | grep ^$OID' ''"
> done |
> sort -rn
2256        18721  = test
2135        18735  = postgres
```

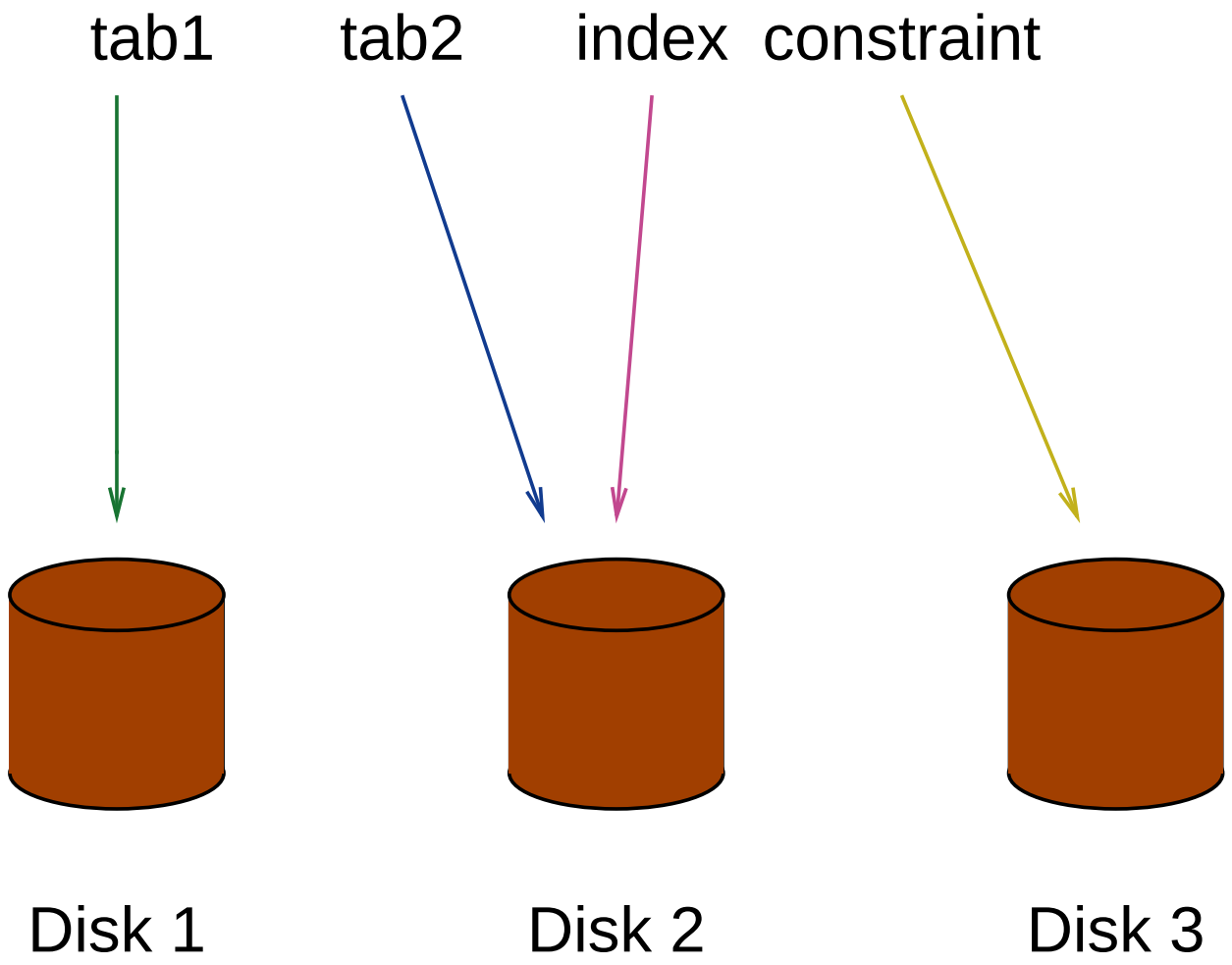
Disk Balancing

- Move pg_xlog to another drive using symlinks
- Tablespaces

Per-Database Tablespaces



Per-Object Tablespaces



Analyzing Locking

```
$ ps -Upostgres
  PID  TT  STAT      TIME COMMAND
 9874  ??  I      0:00.07 postgres test [local] idle in transaction (postmaster)
 9835  ??  S      0:00.05 postgres test [local] UPDATE waiting (postmaster)
10295  ??  S      0:00.05 postgres test [local] DELETE waiting (postmaster)

test=> SELECT * FROM pg_locks;
 relation | database | transaction | pid  |       mode       | granted
----------+----------+-------------+------+------------------+---------
    17143 |    17142 |             | 9173 | AccessShareLock  | t
    17143 |    17142 |             | 9173 | RowExclusiveLock | t
          |          |         472 | 9380 | ExclusiveLock    | t
          |          |         468 | 9338 | ShareLock        | f
          |          |         470 | 9338 | ExclusiveLock    | t
    16759 |    17142 |             | 9380 | AccessShareLock  | t
    17143 |    17142 |             | 9338 | AccessShareLock  | t
    17143 |    17142 |             | 9338 | RowExclusiveLock | t
          |          |         468 | 9173 | ExclusiveLock    | t
(9 rows)
```

Miscellaneous Tasks

- Log file rotation, syslog
- Upgrading
- Migration

Administration Tools

- pgadmin
- phppgadmin

Recovery

Client Application Crash

Nothing Required. Transactions in progress are rolled back.

Graceful Server Crash

Nothing Required. Transactions in progress are rolled back.

Abrupt Server Crash

Nothing Required. Transactions in progress are rolled back.

Operating System Crash

Nothing Required. Transactions in progress are rolled back. Partial page writes are repaired.

Disk Failure

Restore from previous backup or use PITR.

Accidental DELETE

Recover table from previous backup, perhaps using pg_restore. It is possible to modify the backend code to make deleted tuples visible, dump out the deleted table and restore the original code. All tuples in the table since the previous vacuum will be visible. It is possible to restrict that so only tuples deleted by a specific transaction are visible.

Write-Ahead Log (WAL) Corruption

See pg_resetxlog. Review recent transactions and identify any damage, including partially committed transactions.

File Deletion

It may be necessary to create an empty file with the deleted file name so the object can be deleted, and then the object restored from backup.

Accidental DROP TABLE

Restore from previous backup.

Accidental DROP INDEX

Recreate index.

Accidental DROP DATABASE

Restore from previous backup.

Non-Starting Installation

Restart problems are usually caused by write-ahead log problems. See pg_resetxlog. Review recent transactions and identify any damage, including partially committed transactions.

Index Corruption

Use REINDEX.

Table Corruption

Try reindexing the table. Try identifying the corrupt OID of the row and transfer the valid rows into another table using SELECT…INTO…WHERE oid != ###. Use http://sources.redhat.com/rhdb/tools.html to analyze the internal structure of the table.

Conclusion