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

Creative Commons Attribution License  http://momjian.us/presentations

Last updated: September, 2017
1. Installation
2. Configuration
3. Maintenance
4. Monitoring
5. Recovery
Installation

- Click-through Installers
  - MS Windows
  - Linux
  - OS X

- Ports
  - RPM
  - DEB
  - PKG
  - other packages

- Source
  - obtaining
  - build options
  - installing
$ initdb
All of PostgreSQL successfully made. Ready to install.
PostgreSQL installation complete.
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 "en_US.UTF-8".
The default database encoding has accordingly been set to "UTF8".
The default text search configuration will be set to "english".

Data page checksums are disabled.

fixing permissions on existing directory /u/pgsql/data ... ok
creating subdirectories ... ok
selecting default max_connections ... 100
selecting default shared_buffers ... 128MB
selecting dynamic shared memory implementation ... posix
creating configuration files ... ok
running bootstrap script ... ok
performing post-bootstrap initialization ... ok
syncing data to disk ... ok

WARNING: enabling "trust" authentication for local connections
You can change this by editing pg_hba.conf or using the option --A, or
--auth-local and --auth-host, the next time you run initdb.
$ pg_controldata
pg_control version number: 960
Catalog version number: 201608131
Database system identifier: 6442063657695921205
Database cluster state: in production
pg_control last modified: Wed 12 Jul 2017 09:42:12 PM EDT
Latest checkpoint location: 0/14EEA58
Prior checkpoint location: 0/14EE9C0
Latest checkpoint's REDO location: 0/14EEA58
Latest checkpoint's REDO WAL file: 000000010000000000000000
Latest checkpoint's TimeLineID: 1
Latest checkpoint's PrevTimeLineID: 1
Latest checkpoint's full_page_writes: on
Latest checkpoint's NextXID: 0:551
Latest checkpoint's NextOID: 24576
Latest checkpoint's NextMultiXactId: 1
Latest checkpoint's NextMultiOffset: 0
Latest checkpoint's oldestXID: 543
Latest checkpoint's oldestXID's DB: 1
Latest checkpoint's oldestActiveXID: 0
Latest checkpoint's oldestMultiXid: 1
Latest checkpoint's oldestMulti's DB: 1
Latest checkpoint's oldestCommitTsXid: 0
Latest checkpoint's newestCommitTsXid: 0
Time of latest checkpoint: Wed 12 Jul 2017 09:42:12 PM EDT
Fake LSN counter for unlogged rels: 0/1
Minimum recovery ending location: 0/0
Min recovery ending loc's timeline: 0
Backup start location: 0/0
Backup end location: 0/0
End-of-backup record required: no
wal_level setting: minimal
wal_log_hints setting: off
max_connections setting: 100
max_worker_processes setting: 8
System Architecture

- **Main**
  - **Postmaster**
    - **Postgres**
      - **Parse Statement**
      - **Traffic Cop**
        - **Rewrite Query**
          - **Generate Paths**
            - **Optimal Path**
              - **Generate Plan**
                - **Execute Plan**

- **Utility**
  - **Libpq**

- **Storage Managers**
  - **Catalog**
  - **Utilities**
  - **Access Methods**
  - **Nodes / Lists**

- **Optimal Path**
  - **Query**
    - **SELECT, INSERT, UPDATE, DELETE**

- **Query**
  - **Utility Command**
    - e.g. **CREATE TABLE, COPY**

- **Execute Plan**
  - **Execute Plan**
  - **Execute Plan**
Starting Postmaster

LOG: database system was shut down at 2012-01-24 09:33:29 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
  - ’peer’ socket user name passing
  - 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
# TYPE   DATABASE   USER   ADDRESS   METHOD

# "local" is for Unix domain socket connections only
local  all    all    all         trust

# IPv4 local connections:
host  all    all    127.0.0.1/32 trust

# IPv6 local connections:
host  all    all    ::1/128 trust

# Allow replication connections from localhost, by a user with the
# replication privilege.
#local  replication  postgres    trust
#host  replication  postgres    127.0.0.1/32 trust
#host  replication  postgres    ::1/128 trust
# TYPE  DATABASE  USER  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

# disable connections from the gateway machine
host  all  all  192.168.1.254/32  reject

# enable local network
host  all  all  192.168.1.0/24  md5

# require SSL for external connections, but do not allow the superuser
hostssl all  postgres  0.0.0.0/0  reject
hostssl all  all  0.0.0.0/0  md5
Permissions

- Host connection permissions
- Role permissions
  - create roles
  - create databases
  - table permissions
- Database management
  - template1 customization
  - system tables
  - disk space computations
$ ls -CF
base/  pg_hba.conf  pg_replslot/  pg_subtrans/  postgresql.auto.conf
global/ pg_ident.conf  pg_serial/  pg_tblspc/  postgresql.conf
pg_clog/ pg_logical/  pg_snapshots/  pg_twophase/  postmaster.opts
pg_commit_ts/ pg_multixact/ pg_stat/  PG_VERSION  postmaster.pid
pg_dynshmem/ pg_notify/  pg_stat_tmp/  pg_xlog/
$ ls -CF global/
1136   1214_fsm  1261_vm  2671  2846  2967  6000_vm
1136_fsm  1214_vm  1262  2672  2846_vm  3592  6001
1136_vm  1232  1262_fsm  2676  2847  3592_vm  6002
1137   1233  1262_vm  2677  2964  3593  pg_control
1213   1260  2396  2694  2964_vm  4060  pg_filenode.map
1213_fsm  1260_fsm  2396_fsm  2695  2965  4060_vm  pg_internal.init
1213_vm  1260_vm  2396_vm  2697  2966  4061
1214   1261  2397  2698  2966_vm  6000

$ ls -CF base/
1/  12406/  12407/  16384/

$ ls -CF base/16384
112   1249_fsm  2606_vm  2652  2699  3081  3598_vm
113   1249_vm  2607  2653  2701  3085  3599
12242  1255  2607_fsm  2654  2702  3118  3600
12242_fsm  1255_fsm  2607_vm  2655  2703  3118_vm  3600_fsm
12242_vm  1255_vm  2608  2656  2704  3119  3600_vm
12244  1259  2608_fsm  2657  2753  3164  3601
12246  1259_fsm  2608_vm  2658  2753_fsm  3256  3601_fsm

...
$ ls -CF pg_xlog/
00000001000000000000000000000001  archive_status/

$ ls -CF pg_clog/
0000
$ ls -CF share/
conversion_create.sql postgres.bki snowball_create.sql
extension postgres.description sql_features.txt
information_schema.sql postgresql.conf.sample system_views.sql
pg_hba.conf.sample postgres.shdescription timezone
gid.shp.sample psqlrc.sample timezonesets
pg_service.conf.sample recovery.conf.sample tsearch_data
Configuration of postgresql.conf

https://www.flickr.com/photos/mwichary/
# -----------------------------
# 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.
# 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
#  TB = terabytes  h  = hours
#  d  = days
# The default values of these variables are driven from the -D command-line # option or PGDATA environment variable, represented here as ConfigDir.

# 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 = ''  # write an extra PID file
# (change requires restart)
Connections and Authentication

#listen_addresses = 'localhost'

#port = 5432
max_connections = 100
#superuser_reserved_connections = 3
#unix_socket_directories = '/tmp'

#unix_socket_group = ''
#unix_socket_permissions = 0777

#bonjour = off

#bonjour_name = ''

# what IP address(es) to listen on;
# comma-separated list of addresses;
# defaults to 'localhost'; use '*' for all
# (change requires restart)
# (change requires restart)
# (change requires restart)
# comma-separated list of directories
# (change requires restart)
# (change requires restart)
# begin with 0 to use octal notation
# (change requires restart)
# advertise server via Bonjour
# (change requires restart)
# defaults to the computer name
# (change requires restart)
Security and Authentication

#authentication_timeout = 1min # 1s-600s
#ssl = off # (change requires restart)
#ssl_ciphers = 'HIGH: MEDIUM:+3DES:!aNULL' # allowed SSL ciphers
#ssl_prefer_server_ciphers = on # (change requires restart)
#ssl_ecdh_curve = 'prime256v1' # (change requires restart)
#ssl_cert_file = 'server.crt' # (change requires restart)
#ssl_key_file = 'server.key' # (change requires restart)
#ssl_ca_file = '' # (change requires restart)
#ssl_crl_file = '' # (change requires restart)
#password_encryption = on # (change requires restart)
#db_user_namespace = off # (change requires restart)
#row_security = on # (change requires restart)

# GSSAPI using Kerberos
#krb_server_keyfile = ''
#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
shared_buffers = 128MB
# huge_pages = try
# temp_buffers = 8MB
# max_prepared_transactions = 0

# Caution: it is not advisable to set max_prepared_transactions nonzero unless
# you actively intend to use prepared transactions.
# work_mem = 4MB
# maintenance_work_mem = 64MB
# replacement_sort_tuples = 150000
# autovacuum_work_mem = -1
# max_stack_depth = 2MB
dynamic_shared_memory_type = posix

# min 128kB
# (change requires restart)
# on, off, or try
# (change requires restart)
# min 800kB
# zero disables the feature
# (change requires restart)
# min 64kB
# min 1MB
# limits use of replacement selection sort
# min 1MB, or -1 to use maintenance_work_mem
# min 100kB
# the default is the first option ...
Memory Usage (Continued)

Query and Checkpoint Operations

- PostgreSQL Shared Buffer Cache
- Kernel Disk Buffer Cache

Transaction Durability

- Write-Ahead Log
- fsync
- fsync

Recovery

Postgres Backend

Disk Blocks
Sizing Shared Memory

- Kernel Disk Buffer Cache
- Shared Buffer Cache (shared_buffers)
- Postgres Session (work_mem)
# Disk and Kernel Resources

# - Disk -

#temp_file_limit = -1  # limits per-process temp file space
# in kB, or -1 for no limit

# - Kernel Resource Usage -

#max_files_per_process = 1000  # min 25
# (change requires restart)
#shared_preload_libraries = ''  # (change requires restart)
Vacuum and Background Writer

```plaintext
#vacuum_cost_delay = 0  # 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 multiplier on buffers scanned/round
#bgwriter_flush_after = 512kB  # measured in pages, 0 disables

# - Asynchronous Behavior -
#effective_io_concurrency = 1  # 1-1000; 0 disables prefetching
#max_worker_processes = 8  # (change requires restart)
#max_parallel_workers_per_gather = 0  # taken from max_worker_processes
#old_snapshot_threshold = -1  # 1min-60d; -1 disables; 0 is immediate
#backend_flush_after = 0  # (change requires restart)
# measured in pages, 0 disables
```
#wal_level = minimal

#fsync = on

#synchronous_commit = on

#wal_sync_method = fsync

#full_page_writes = on
#wal_compression = off
#wal_log_hints = off

#wal_buffers = -1

#wal_writer_delay = 200ms
#wal_writerFlush_after = 1MB

#commit_delay = 0
#commit_siblings = 5

# minimal, replica, or logical
# (change requires restart)
# flush data to disk for crash safety
# (turning this off can cause
# unrecoverable data corruption)
# synchronization level;
# off, local, remote_write, remote_apply, or
# the default is the first option
# supported by the operating system:
# open_datasync
# fdatasync (default on Linux)
# fsync
# fsync_writethrough
# open_sync
# recover from partial page writes
# enable compression of full-page writes
# also do full page writes of non-critical
# (change requires restart)
# min 32kB, -1 sets based on shared_buffers
# (change requires restart)
# 1-10000 milliseconds
# measured in pages, 0 disables

# range 0-100000, in microseconds
# 1-1000
Write-Ahead Logging (Continued)

Query and Checkpoint Operations

PostgreSQL Shared Buffer Cache

Kernel Disk Buffer Cache

Disk Blocks

Transaction Durability

Write-Ahead Log

fsync

fsync

Postgres Backend

Postgres Backend

Postgres Backend

Recovery
# Checkpoints and Archiving

```plaintext
#checkpoint_timeout = 5min  # range 30s-1d
#max_wal_size = 1GB
#min_wal_size = 80MB
#checkpoint_completion_target = 0.5  # checkpoint target duration, 0.0 - 1.0
#checkpoint_flush_after = 256kB  # measured in pages, 0 disables
#checkpoint_warning = 30s  # 0 disables

# - Archiving -

#archive_mode = off  # enables archiving; off, on, or always
# (change requires restart)
#archive_command = ''  # command to use to archive a logfile segment
# placeholders: %p = path of file to archive
#               %f = file name only
# e.g. 'test ! -f /mnt/server/archivedir/%f && cp %p
#archive_timeout = 0  # force a logfile segment switch after this
# number of seconds; 0 disables
```
Write-Ahead Logging (Continued)

PostgreSQL Shared Buffer Cache

Begin 1

<table>
<thead>
<tr>
<th>1</th>
<th>1</th>
</tr>
</thead>
</table>

Write-Ahead Log

Write-Ahead Log

End 1

| 1 | 2 | 2 | 1 | 2 |

Write-Ahead Log

Rotate

| 2 | 2 | 2 | 2 |

Write-Ahead Log
#max_wal_senders = 0            # max number of walsender processes
wal_keep_segments = 0           # (change requires restart)
wal_sender_timeout = 60s       # in logfile segments, 16MB each; 0 disables
#max_replication_slots = 0      # in milliseconds; 0 disables
#track_commit_timestamp = off   # max number of replication slots
# (change requires restart)
# collect timestamp of transaction commit
# (change requires restart)
#synchronous_standby_names = ''  # standby servers that provide sync rep
# number of sync standbys and comma-separated list
# from standby(s); '*' = all
#vacuum_defer_cleanup_age = 0  # number of xacts by which cleanup is delayed
Standby Replication Server

#hot_standby = off

#max_standby_archive_delay = 30s

#max_standby_streaming_delay = 30s

#wal_receiver_status_interval = 10s

#hot_standby_feedback = off

#wal_receiver_timeout = 60s

#wal_retrieve_retry_interval = 5s

# "on" allows queries during recovery
# (change requires restart)
# max delay before canceling queries
# when reading WAL from archive;
# -1 allows indefinite delay
# max delay before canceling queries
# when reading streaming WAL;
# -1 allows indefinite delay
# send replies at least this often
# 0 disables
# send info from standby to prevent
# query conflicts
# time that receiver waits for
# communication from master
# in milliseconds; 0 disables
# time to wait before retrying to
# retrieve WAL after a failed attempt
#enable_bitmapscan = on
#enable_hashagg = on
#enable_hashjoin = on
#enable_indexscan = on
#enable_indexonlyscan = on
#enable_material = on
#enable_mergejoin = on
#enable_nestloop = on
#enable_seqscan = on
#enable_sort = on
#enable_tidscan = on
Planner Constants

- \#seq_page_cost = 1.0  
- \#random_page_cost = 4.0  
- \#cpu_tuple_cost = 0.01  
- \#cpu_index_tuple_cost = 0.005  
- \#cpu_operator_cost = 0.0025  
- \#parallel_tuple_cost = 0.1  
- \#parallel_setup_cost = 1000.0  
- \#min_parallel_relation_size = 8MB  
- \#effective_cache_size = 4GB  

# measured on an arbitrary scale  
# same scale as above  
# same scale as above  
# same scale as above  
# same scale as above  
# same scale as above  
# same scale as above
Planner GEQO

#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
#miscellaneous_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

#force_parallel_mode = off

# 1 disables collapsing of explicit
#  JOIN clauses
#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

#log_file_mode = 0600

# creation mode for log files, begin with 0 to use octal notation
Where To Log (rotation)

#log_truncate_on_rotation = off

# If on, an existing log file with 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 (syslog)

#syslog_facility = 'LOCAL0'
#syslog_ident = 'postgres'
#syslog_sequence_numbers = on
#syslog_split_messages = on

# This is only relevant when logging to eventlog (win32):
#event_source = 'PostgreSQL'
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
```
#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
# 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
#  %n = timestamp with milliseconds (as a
#  %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_statement = 'none'
#log_replication_commands = off
#log_temp_files = -1

log_timezone = 'US/Eastern'

# - Process Title -
#cluster_name = ''
#update_process_title = on

# log lock waits >= deadlock_timeout
# none, ddl, mod, all

# log temporary files equal or larger than the specified size in kilobytes;
# -1 disables, 0 logs all temp files

# added to process titles if nonempty
# (change requires restart)
# - Query/Index Statistics Collector -

#track_activities = on
#track_counts = on
#track_io_timing = off
#track_functions = none  # none, pl, all
#track_activity_query_size = 1024  # (change requires restart)
#stats_temp_directory = 'pg_stat_tmp'

# - Statistics Monitoring -

#log_parser_stats = off
#log_planner_stats = off
#log_executor_stats = off
#log_statement_stats = off
# Enable autovacuum subprocess? 'on'
# requires track_counts to also be on.
# -1 disables, 0 logs all actions and
# their durations, > 0 logs only
# actions running at least this number
# of milliseconds.
# max number of autovacuum subprocesses
# (change requires restart)
# time between autovacuum runs
# min number of row updates before
# vacuum
# min number of row updates before
# analyze
# fraction of table size before vacuum
# fraction of table size before analyze
# maximum XID age before forced vacuum
# (change requires restart)
# maximum multixact age
# before forced vacuum
# (change requires restart)
# default vacuum cost delay for
# autovacuum, in milliseconds;
# -1 means use vacuum_cost_delay
# default vacuum cost limit for
Statement Behavior

```sql
#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
#default_transaction_deferrable = off
#session_replication_role = 'origin'
#statement_timeout = 0  # in milliseconds, 0 is disabled
#lock_timeout = 0  # in milliseconds, 0 is disabled
#idle_in_transaction_session_timeout = 0  # in milliseconds, 0 is disabled
#vacuum_freeze_min_age = 50000000  # in milliseconds, 0 is disabled
#vacuum_freeze_table_age = 150000000
#vacuum_multixact_freeze_min_age = 5000000
#vacuum_multixact_freeze_table_age = 150000000
#bytea_output = 'hex'  # hex, escape
#xmlbinary = 'base64'
#xmloption = 'content'
#gin_fuzzy_search_limit = 0
#gin_pending_list_limit = 4MB
```
locale, formatting, and full text search

datestyle = 'iso, mdy'
#intervalstyle = 'postgres'
timezone = 'US/Eastern'
#timezone_abbreviations = 'Default'

#extra_float_digits = 0
#client_encoding = sql_ascii

# These settings are initialized by initdb, but they can be changed.
1c_messages = 'en_US.UTF-8'
# locale for system error message
# strings
1c_monetary = 'en_US.UTF-8'
# locale for monetary formatting
1c_numeric = 'en_US.UTF-8'
# locale for number formatting
1c_time = 'en_US.UTF-8'
# locale for time formatting

# default configuration for text search
default_text_search_config = 'pg_catalog.english'
#dynamic_library_path = '$libdir'
#local_preload_libraries = ''
#session_preload_libraries = ''
Lock Management

```plaintext
#deadlock_timeout = 1s
#max_locks_per_transaction = 64 # min 10
#max_pred_locks_per_transaction = 64 # (change requires restart)
```
# - 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
#operator_precedence_warning = off
#quote_all_identifiers = off
#sql_inheritance = on
#standard_conforming_strings = on
#synchronize_seqscans = on

# - Other Platforms and Clients -

#transform_null_equals = off
Error Handling

`#exit_on_error = off`  # terminate session on any error?
`#restart_after_crash = on`  # reinitialize after backend crash?
## Config File Includes

```plaintext
#include_dir = 'conf.d'
#include_if_exists = 'exists.conf'
#include = 'special.conf'
```
Interfaces

➤ Installing
  ➤ Compiled Languages (C, ecpg)
  ➤ Scripting Language (Perl, Python, PHP)
  ➤ SPI

➤ Connection Pooling
Include Files

$ ls -CF include/
ecpg_config.h    libpq/          pgtypes_date.h    sql3types.h
ecpgerrno.h      libpq-events.h  pgtypes_error.h   sqlca.h
ecpg_informix.h  libpq-fe.h      pgtypes_interval.h sqllda-compat.h
ecpglib.h        pg_config_ext.h  pgtypes_numeric.h sqlda.h
ecpgtype.h       pg_config.h      pgtypes_timestamp.h sqllda-native.h
informix/        pg_config_manual.h postgres_ext.h
internal/        pg_config_os.h   server/
$ ls -CF lib/
ascii_and_mic.so*
cyrillic_and_mic.so*
dict_snowball.so*
euc2004_sjis2004.so*
euc_cn_and_mic.so*
euc_jp_and_sjis.so*
euc_kr_and_mic.so*
euc_tw_and_big5.so*
latin2_and_win1250.so*
latin_and_mic.so*
libecpg.a
libecpg_compat.a
libecpg_compat.so@
libecpg_compat.so.3@
libecpg_compat.so.3.8*
libecpg.so@
libecpg.so.6@
libecpg.so.6.8*
libpgcommon.a
libpgfeutils.a
libpgport.a
libpgtypes.a
libpgtypes.so@
libpgtypes.so.3@
libpgtypes.so.3.7*
libpq.a
libpq.so@
libpq.so.5@
libpq.so.5.9*
libpqwalreceiver.so*
pgxs/
pkgconfig/
plperl.so*
plpgsql.so*
plpython2.so*
utf8_and_ascii.so*
libpgtypes.so.3.7*
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)

File System−Level Backup

Continuous Archive (WAL)

02:00

WAL

09:00 11:00 13:00
PITR Backup Procedures

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

`pg_basebackup` does this automatically and can be run on version 9.2+ standbys.
PITR Recovery

File System-Level Backup

Continuous Archive (WAL)
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
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
<table>
<thead>
<tr>
<th>Process</th>
<th>PID</th>
<th>PPID</th>
<th>Username</th>
<th>Time</th>
<th>Command</th>
</tr>
</thead>
<tbody>
<tr>
<td>postgres</td>
<td>825</td>
<td>1</td>
<td></td>
<td>0:06:57</td>
<td>/u/pgsql/bin/postmaster -i</td>
</tr>
<tr>
<td>postgres</td>
<td>829</td>
<td>825</td>
<td></td>
<td>0:35:03</td>
<td>writer process (postmaster)</td>
</tr>
<tr>
<td>postgres</td>
<td>830</td>
<td>825</td>
<td></td>
<td>0:16:07</td>
<td>wal writer process (postmaster)</td>
</tr>
<tr>
<td>postgres</td>
<td>831</td>
<td>825</td>
<td></td>
<td>0:11:34</td>
<td>autovacuum launcher process (postmaster)</td>
</tr>
<tr>
<td>postgres</td>
<td>832</td>
<td>825</td>
<td></td>
<td>0:07:63</td>
<td>stats collector process (postmaster)</td>
</tr>
<tr>
<td>postgres</td>
<td>13003</td>
<td>825</td>
<td></td>
<td>0:00:01</td>
<td>postgres test [local] idle (postmaster)</td>
</tr>
<tr>
<td>postgres</td>
<td>13002</td>
<td>12997</td>
<td>tty1</td>
<td>0:00:03</td>
<td>/u/pgsql/bin/psql test</td>
</tr>
</tbody>
</table>
$ 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
```sql
SELECT * FROM pg_stat_activity;
```

<table>
<thead>
<tr>
<th>datid</th>
<th>16384</th>
</tr>
</thead>
<tbody>
<tr>
<td>datname</td>
<td>test</td>
</tr>
<tr>
<td>procpid</td>
<td>29964</td>
</tr>
<tr>
<td>usesysid</td>
<td>10</td>
</tr>
<tr>
<td>usename</td>
<td>postgres</td>
</tr>
<tr>
<td>application_name</td>
<td>psql</td>
</tr>
<tr>
<td>client_addr</td>
<td></td>
</tr>
<tr>
<td>client_port</td>
<td>-1</td>
</tr>
<tr>
<td>backend_start</td>
<td>2011-04-04 08:27:33.089199-04</td>
</tr>
<tr>
<td>xact_start</td>
<td>2011-04-04 08:27:47.901121-04</td>
</tr>
<tr>
<td>query_start</td>
<td>2011-04-04 08:27:47.901121-04</td>
</tr>
<tr>
<td>waiting</td>
<td>f</td>
</tr>
<tr>
<td>current_query</td>
<td>SELECT * FROM pg_stat_activity;</td>
</tr>
<tr>
<td>Access Statistics</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td></td>
</tr>
<tr>
<td>pg_stat_all_indexes</td>
<td>view</td>
</tr>
<tr>
<td>pg_stat_all_tables</td>
<td>view</td>
</tr>
<tr>
<td>pg_stat_database</td>
<td>view</td>
</tr>
<tr>
<td>pg_stat_sys_indexes</td>
<td>view</td>
</tr>
<tr>
<td>pg_stat_sys_tables</td>
<td>view</td>
</tr>
<tr>
<td>pg_stat_user_indexes</td>
<td>view</td>
</tr>
<tr>
<td>pg_stat_user_tables</td>
<td>view</td>
</tr>
<tr>
<td>pg_statio_all_indexes</td>
<td>view</td>
</tr>
<tr>
<td>pg_statio_all_sequences</td>
<td>view</td>
</tr>
<tr>
<td>pg_statio_all_tables</td>
<td>view</td>
</tr>
<tr>
<td>pg_statio_sys_indexes</td>
<td>view</td>
</tr>
<tr>
<td>pg_statio_sys_sequences</td>
<td>view</td>
</tr>
<tr>
<td>pg_statio_sys_tables</td>
<td>view</td>
</tr>
<tr>
<td>pg_statio_user_indexes</td>
<td>view</td>
</tr>
<tr>
<td>pg_statio_user_sequences</td>
<td>view</td>
</tr>
<tr>
<td>pg_statio_user_tables</td>
<td>view</td>
</tr>
</tbody>
</table>
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

```sql
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 |
```
```sql
SELECT * FROM pg_statio_all_tables;
```

<table>
<thead>
<tr>
<th>relid</th>
<th>2602</th>
</tr>
</thead>
<tbody>
<tr>
<td>schemaname</td>
<td>pg_catalog</td>
</tr>
<tr>
<td>relname</td>
<td>pg_amop</td>
</tr>
<tr>
<td>heap_blks_read</td>
<td>3</td>
</tr>
<tr>
<td>heap_blks_hit</td>
<td>114</td>
</tr>
<tr>
<td>idx_blks_read</td>
<td>5</td>
</tr>
<tr>
<td>idx_blks_hit</td>
<td>303</td>
</tr>
<tr>
<td>toast_blks_read</td>
<td></td>
</tr>
<tr>
<td>toast_blks_hit</td>
<td></td>
</tr>
<tr>
<td>tidx_blks_read</td>
<td></td>
</tr>
<tr>
<td>tidx_blks_hit</td>
<td></td>
</tr>
</tbody>
</table>
Analyzing Activity

- Heavily used tables
- Unnecessary indexes
- Additional indexes
- Index usage
- TOAST usage
$ vmstat 5

<table>
<thead>
<tr>
<th>proc</th>
<th>memory</th>
<th>page</th>
<th>disks</th>
<th>faults</th>
<th>cpu</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>b</td>
<td>w</td>
<td>avm</td>
<td>fre</td>
<td>flt</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>501820</td>
<td>48520</td>
<td>1234</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>512796</td>
<td>46812</td>
<td>1422</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>542260</td>
<td>44356</td>
<td>788</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>539708</td>
<td>41868</td>
<td>576</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>547200</td>
<td>32964</td>
<td>454</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>556140</td>
<td>23884</td>
<td>461</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>535136</td>
<td>46280</td>
<td>1056</td>
</tr>
</tbody>
</table>

CPU
```
$ iostat 5

<table>
<thead>
<tr>
<th>tty</th>
<th>tin</th>
<th>tout</th>
<th>sps</th>
<th>tps</th>
<th>mps</th>
<th>sps</th>
<th>tps</th>
<th>mps</th>
<th>sps</th>
<th>tps</th>
<th>mps</th>
<th>usr</th>
<th>nic</th>
<th>sys</th>
<th>int</th>
<th>idl</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 119</td>
<td>244</td>
<td>11</td>
<td>6.1</td>
<td>0</td>
<td>0</td>
<td>27.3</td>
<td>0</td>
<td>0</td>
<td>18.1</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>86</td>
</tr>
<tr>
<td>0  86</td>
<td>20</td>
<td>1</td>
<td>1.4</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0  82</td>
<td>61</td>
<td>4</td>
<td>3.6</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0  65</td>
<td>6</td>
<td>0</td>
<td>0.0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12  90</td>
<td>31</td>
<td>2</td>
<td>5.4</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24  173</td>
<td>6</td>
<td>0</td>
<td>4.9</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>48</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0  91</td>
<td>3594</td>
<td>63</td>
<td>4.6</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>11</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>85</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```
test=> \df *size*

<table>
<thead>
<tr>
<th>Schema</th>
<th>Name</th>
<th>Result data type</th>
<th>Argument data types</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>pg_catalog</td>
<td>pg_column_size</td>
<td>integer</td>
<td>&quot;any&quot;</td>
<td>normal</td>
</tr>
<tr>
<td>pg_catalog</td>
<td>pg_database_size</td>
<td>bigint</td>
<td>name</td>
<td>normal</td>
</tr>
<tr>
<td>pg_catalog</td>
<td>pg_database_size</td>
<td>bigint</td>
<td>oid</td>
<td>normal</td>
</tr>
<tr>
<td>pg_catalog</td>
<td>pg_indexes_size</td>
<td>bigint</td>
<td>regclass</td>
<td>normal</td>
</tr>
<tr>
<td>pg_catalog</td>
<td>pg_relation_size</td>
<td>bigint</td>
<td>regclass</td>
<td>normal</td>
</tr>
<tr>
<td>pg_catalog</td>
<td>pg_relation_size</td>
<td>bigint</td>
<td>regclass, text</td>
<td>normal</td>
</tr>
<tr>
<td>pg_catalog</td>
<td>pg_size_pretty</td>
<td>text</td>
<td>bigint</td>
<td>normal</td>
</tr>
<tr>
<td>pg_catalog</td>
<td>pg_table_size</td>
<td>bigint</td>
<td>regclass</td>
<td>normal</td>
</tr>
<tr>
<td>pg_catalog</td>
<td>pg_tablespace_size</td>
<td>bigint</td>
<td>name</td>
<td>normal</td>
</tr>
<tr>
<td>pg_catalog</td>
<td>pg_tablespace_size</td>
<td>bigint</td>
<td>oid</td>
<td>normal</td>
</tr>
<tr>
<td>pg_catalog</td>
<td>pg_total_relation_size</td>
<td>bigint</td>
<td>regclass</td>
<td>normal</td>
</tr>
</tbody>
</table>

(11 rows)
$ oid2name

All databases:

---------------------------------
18720  = test1
1      = template1
18719  = template0
18721  = test
18735  = postgres
18736  = cssi
$ 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

DB1  DB2  DB3  DB4

Disk 1  Disk 2  Disk 3
Per-Object Tablespaces

Diagram:

- tab1
- tab2
- index
- constraint

Disk 1
Disk 2
Disk 3
$ ps -f -U postgres

<table>
<thead>
<tr>
<th>PID</th>
<th>TT</th>
<th>STAT</th>
<th>TIME</th>
<th>COMMAND</th>
</tr>
</thead>
<tbody>
<tr>
<td>9874</td>
<td>??</td>
<td>I</td>
<td>0:00.07</td>
<td>postgres test [local] idle in transaction (postmaster)</td>
</tr>
<tr>
<td>9835</td>
<td>??</td>
<td>S</td>
<td>0:00.05</td>
<td>postgres test [local] UPDATE waiting (postmaster)</td>
</tr>
<tr>
<td>10295</td>
<td>??</td>
<td>S</td>
<td>0:00.05</td>
<td>postgres test [local] DELETE waiting (postmaster)</td>
</tr>
</tbody>
</table>

test=> SELECT * FROM pg_locks;

<table>
<thead>
<tr>
<th>relation</th>
<th>database</th>
<th>transaction</th>
<th>pid</th>
<th>mode</th>
<th>granted</th>
</tr>
</thead>
<tbody>
<tr>
<td>17143</td>
<td>17142</td>
<td></td>
<td>9173</td>
<td>AccessShareLock</td>
<td>t</td>
</tr>
<tr>
<td>17143</td>
<td>17142</td>
<td></td>
<td>9173</td>
<td>RowExclusiveLock</td>
<td>t</td>
</tr>
<tr>
<td></td>
<td></td>
<td>472</td>
<td>9380</td>
<td>ExclusiveLock</td>
<td>t</td>
</tr>
<tr>
<td></td>
<td></td>
<td>468</td>
<td>9338</td>
<td>ShareLock</td>
<td>f</td>
</tr>
<tr>
<td></td>
<td></td>
<td>470</td>
<td>9338</td>
<td>ExclusiveLock</td>
<td>t</td>
</tr>
<tr>
<td>16759</td>
<td>17142</td>
<td></td>
<td>9380</td>
<td>AccessShareLock</td>
<td>t</td>
</tr>
<tr>
<td>17143</td>
<td>17142</td>
<td></td>
<td>9338</td>
<td>AccessShareLock</td>
<td>t</td>
</tr>
<tr>
<td>17143</td>
<td>17142</td>
<td></td>
<td>9338</td>
<td>RowExclusiveLock</td>
<td>t</td>
</tr>
<tr>
<td></td>
<td></td>
<td>468</td>
<td>9173</td>
<td>ExclusiveLock</td>
<td>t</td>
</tr>
</tbody>
</table>

(9 rows)
Miscellaneous Tasks

- Log file rotation, syslog
- Upgrading
  - `pg_dump`, `restore`
  - `pg_upgrade`
  - Slony
- Migration
Administration Tools

- pgadmin
- phppgadmin
- Alerting: check_postgres, tail_n_mail, Nagios
- Analysis: Munin, Cacti, Zabbix, Nagios, MRTG
- Queries: pgbadger, pgFouine
- Commercial: Circonus (or open-source Reconnoiter), Postgres Enterprise Manager (PEM), Hyperic
Recovery

https://www.flickr.com/photos/coastguardnews/
Nothing Required. Transactions in progress are rolled back.
Graceful Postgres Server Shutdown

Nothing Required. Transactions in progress are rolled back.
Abrupt Postgres 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.
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.
Restart problems are usually caused by write-ahead log problems. See `pg_resetxlog`. Review recent transactions and identify any damage, including partially committed transactions.
Use REINDEX.
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](http://sources.redhat.com/rhdb/tools.html) to analyze the internal structure of the table.
Conclusion

http://momjian.us/presentations