

MASTERING POSTGRESQL ADMINISTRATION

BRUCE MOMJIAN,
SOFTWARE RESEARCH ASSOCIATES

MARCH, 2005



ABSTRACT

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

INSTALLATION

- SOURCE
 - OBTAINING
 - INSTALLING

- BUILD OPTIONS

- RPM
 - OBTAINING
 - INSTALLING

- MS WINDOWS
 - OBTAINING
 - 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.

```
creating directory /u/pg/data ... ok
```

```
creating directory /u/pg/data/global ... ok
```

```
creating directory /u/pg/data/pg_xlog ... ok
```

```
creating directory /u/pg/data/pg_xlog/archive_status ... ok
```

```
creating directory /u/pg/data/pg_clog ... ok
```

```
creating directory /u/pg/data/pg_subtrans ... ok
```

```
creating directory /u/pg/data/base ... ok
```

```
creating directory /u/pg/data/base/1 ... ok
```

```
creating directory /u/pg/data/pg_tblspc ... ok
```

```
selecting default max_connections ... 100
```

INITIALIZATION (CONTINUED)

```
selecting default shared_buffers ... 1000
creating configuration files ... ok
creating template1 database in /u/pg/data/base/1 ... ok
initializing pg_shadow ... ok
enabling unlimited row size for system tables ... ok
initializing pg_depend ... ok
creating system views ... ok
loading pg_description ... ok
creating conversions ... ok
setting privileges on built-in objects ... ok
creating information schema ... ok
vacuuming database template1 ... ok
copying template1 to template0 ... 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:

```
postmaster -D /u/pg/data
```

or

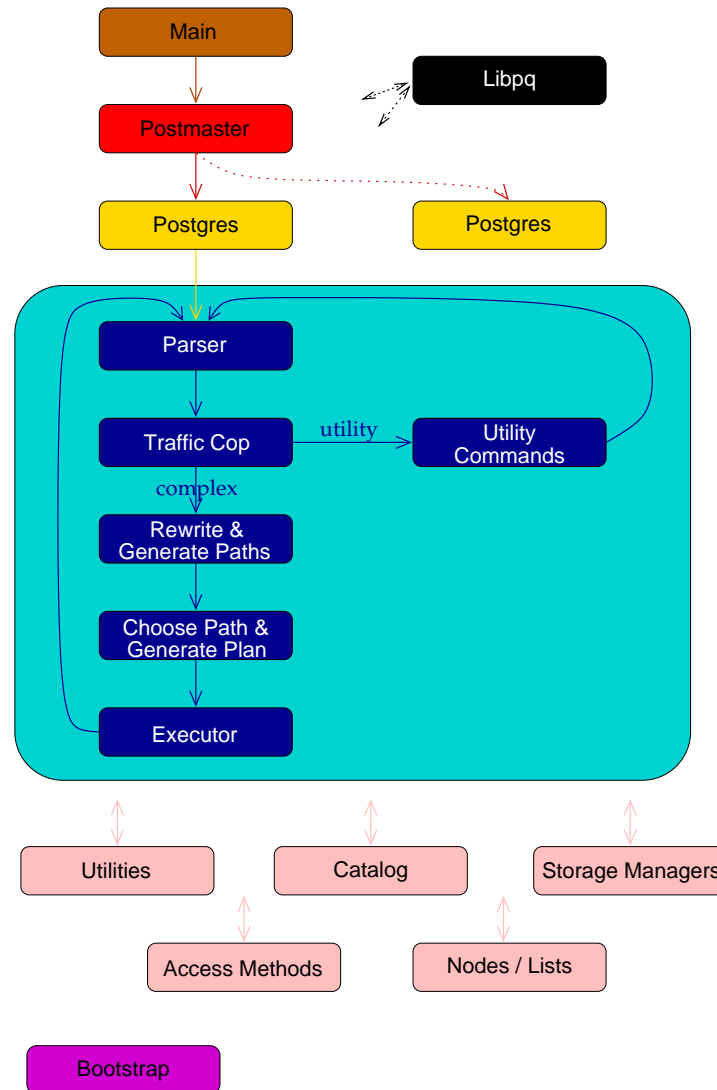
```
pg_ctl -D /u/pg/data -l logfile start
```

PG_CONTROLDATA

\$ pg_controldata

```
pg_control version number:      74
Catalog version number:        200502281
Database system identifier:     4766833642862247929
Database cluster state:        shut down
pg_control last modified:       03/03/05 10:49:18
Current log file ID:           0
Next log file segment:         1
Latest checkpoint location:     0/A34010
Prior checkpoint location:      0/A2D5C0
Latest checkpoint's REDO location: 0/A34010
Latest checkpoint's UNDO location: 0/0
Latest checkpoint's TimeLineID: 1
Latest checkpoint's NextXID:    545
Latest checkpoint's NextOID:    17233
Time of latest checkpoint:      03/03/05 10:49:18
Database block size:           8192
Blocks per segment of large relation: 131072
Bytes per WAL segment:         16777216
Maximum length of identifiers:  64
Maximum number of function arguments: 32
Date/time type storage:        floating-point numbers
Maximum length of locale name:  128
LC_COLLATE:                     C
LC_CTYPE:                       C
```

SYSTEM ARCHITECTURE



STARTING POSTMASTER

```
LOG: database system was shut down at 2005-03-03 10:49:18 EST
LOG: checkpoint record is at 0/A34010
LOG: redo record is at 0/A34010; undo record is at 0/0; shut-
down TRUE
LOG: next transaction ID: 545; next OID: 17233
LOG: database system is ready
```

- MANUALLY
- PG_CTL
- ON BOOT

STOPPING POSTMASTER

LOG: received smart shutdown request

LOG: shutting down

LOG: database system is shut down

- MANUALLY
- PG_CTL
- ON SHUTDOWN

CONNECTIONS

- LOCAL — UNIX DOMAIN SOCKET
- HOST — TCP/IP
- HOSTSSL

AUTHENTICATION (PG_HBA.CONF)

- TRUST

- PASSWORDS
 - MD5
 - CRYPT
 - PASSWORD

- REMOTE AUTHENTICATION
 - HOST IDENT USING PG_IDENT.CONF
 - KERBEROS

- LOCAL IDENT

- HOST IDENT USING LOCAL IDENTD

- SOCKET PERMISSIONS

- PAM

- REJECT

ACCESS

- HOSTNAME AND NETWORK MASK
- DBNAME
- USERNAME
- GROUPNAME
- FILENAME OR LIST OF DATABASES, USERS, GROUPS
- IPV6

PERMISSIONS

- HOST CONNECTION PERMISSIONS

- USER/GROUP PERMISSIONS
 - CREATE USERS
 - CREATE DATABASES
 - TABLE PERMISSIONS

- DATABASE CREATION
 - TEMPLATE1 CUSTOMIZATION
 - SYSTEM TABLES
 - DISK SPACE COMPUTATIONS

DATA DIRECTORY

```
$ ls -CF
```

```
PG_VERSION
```

```
base/
```

```
global/
```

```
pg_clog/
```

```
pg_hba.conf
```

```
pg_ident.conf
```

```
pg_xlog/
```

```
postgresql.conf
```

```
postmaster.opts
```

```
postmaster.pid
```

DATABASE DIRECTORIES

```
$ ls -CF global/
```

```
1260          16432          16454          16475          pg_group
1261          16434          16467          16485          pg_pwd
1262          16435          16469          16487          pgstat.stat
16431         16453          16473          pg_control
```

```
$ ls -CF base/
```

```
1/          16569/  16640/  16652/
```

```
$ ls -CF base/16569
```

```
1247          16422          16450
1249          16423          16451
...
```

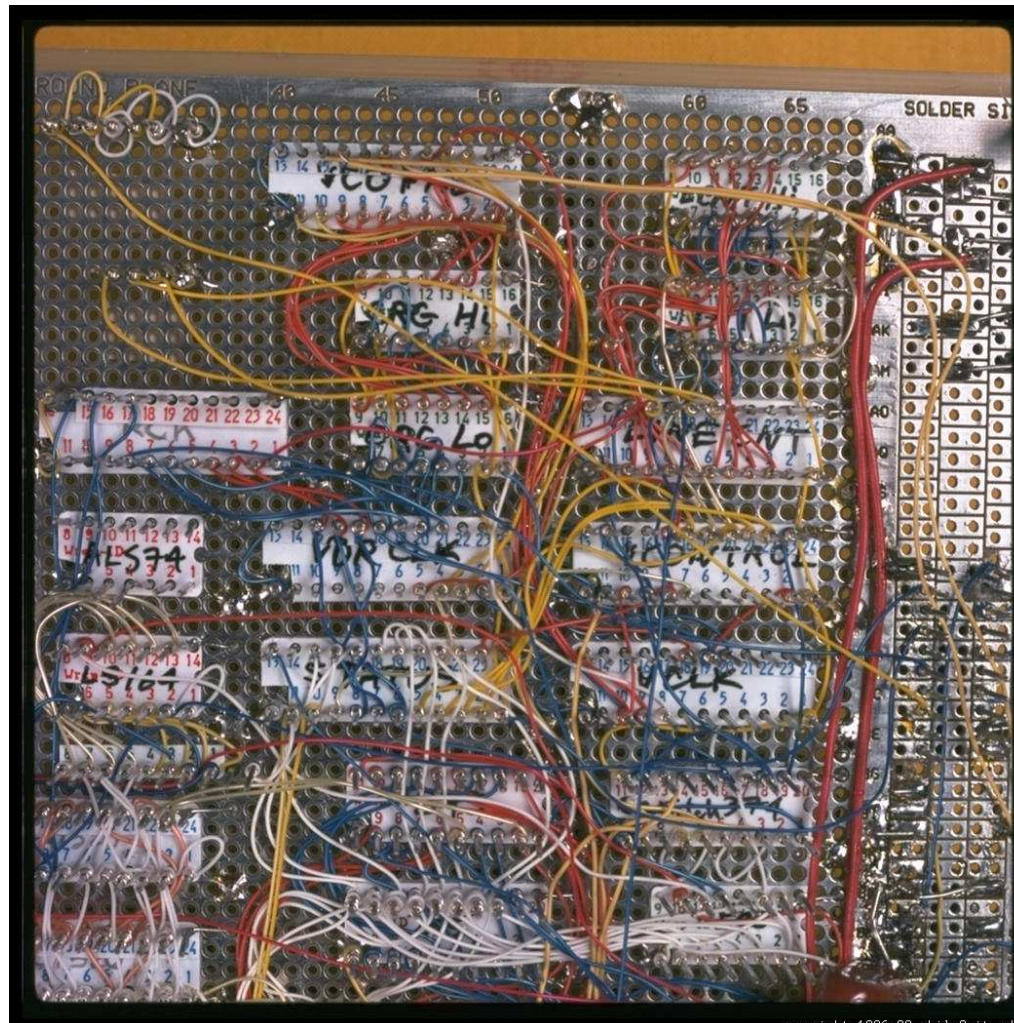
TRANSACTION/WAL DIRECTORIES

```
$ ls -CF pg_xlog/  
00000001000000000000000000000000 archive_status/  
$ ls -CF pg_clog/  
0000
```

CONFIGURATION DIRECTORIES

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

CONFIGURATION POSTGRESQL.CONF



copyright 1986-98 philg@mit.edu

POSTGRESQL.CONF

```
# -----  
# PostgreSQL configuration file  
# -----  
#  
# This file consists of lines of the form:  
#  
#   name = value  
#  
# (The '=' is optional.) White space may be used. Comments are introduced  
# with '#' anywhere on a line. The complete list of option names and  
# allowed values can be found in the PostgreSQL documentation. The  
# commented-out settings shown in this file represent the default values.
```

POSTGRESQL.CONF (CONTINUED)

```
# Please note that re-commenting a setting is NOT sufficient to revert it
# to the default value, unless you restart the postmaster.
#
# Any option can also be given as a command line switch to the
# postmaster, e.g. 'postmaster -c log_connections=on'. Some options
# can be changed at run-time with the 'SET' SQL command.
#
# This file is read on postmaster startup and when the postmaster
# receives a SIGHUP. If you edit the file on a running system, you have
# to SIGHUP the postmaster for the changes to take effect, or use
# "pg_ctl reload". Some settings, such as listen_address, require
# a postmaster shutdown and restart to take effect.
```

CONFIGURATION FILE LOCATION

```
# The default values of these variables are driven from the -D command line
# switch or PGDATA environment variable, represented here as ConfigDir.
# data_directory = 'ConfigDir'           # use data in another directory
# hba_file = 'ConfigDir/pg_hba.conf'     # the host-based authentication file
# ident_file = 'ConfigDir/pg_ident.conf' # the IDENT configuration file
# If external_pid_file is not explicitly set, no extra pid file is written.
# external_pid_file = '(none)'          # write an extra pid file
```

CONNECTIONS AND AUTHENTICATION

```
#listen_addresses = 'localhost' # what IP interface(s) to listen on;  
                                # defaults to localhost, '*' = any  
  
#port = 5432  
max_connections = 100  
    # note: increasing max_connections costs about 500 bytes of shared  
    # memory per connection slot, in addition to costs from shared_buffers  
    # and max_locks_per_transaction.  
#superuser_reserved_connections = 2  
#unix_socket_directory = ''  
#unix_socket_group = ''  
#unix_socket_permissions = 0777 # octal  
#rendezvous_name = ''          # defaults to the computer name
```

SECURITY AND AUTHENTICATION

```
#authentication_timeout = 60      # 1-600, in seconds  
#ssl = false  
#password_encryption = true  
#krb_server_keyfile = ''  
#db_user_namespace = false
```

RESOURCE USAGE

- Memory -

```
shared_buffers = 1000          # min 16, at least max_connections*2, 8KB each
#work_mem = 1024              # min 64, size in KB
#maintenance_work_mem = 16384 # min 1024, size in KB
#max_stack_depth = 2048      # min 100, size in KB
```

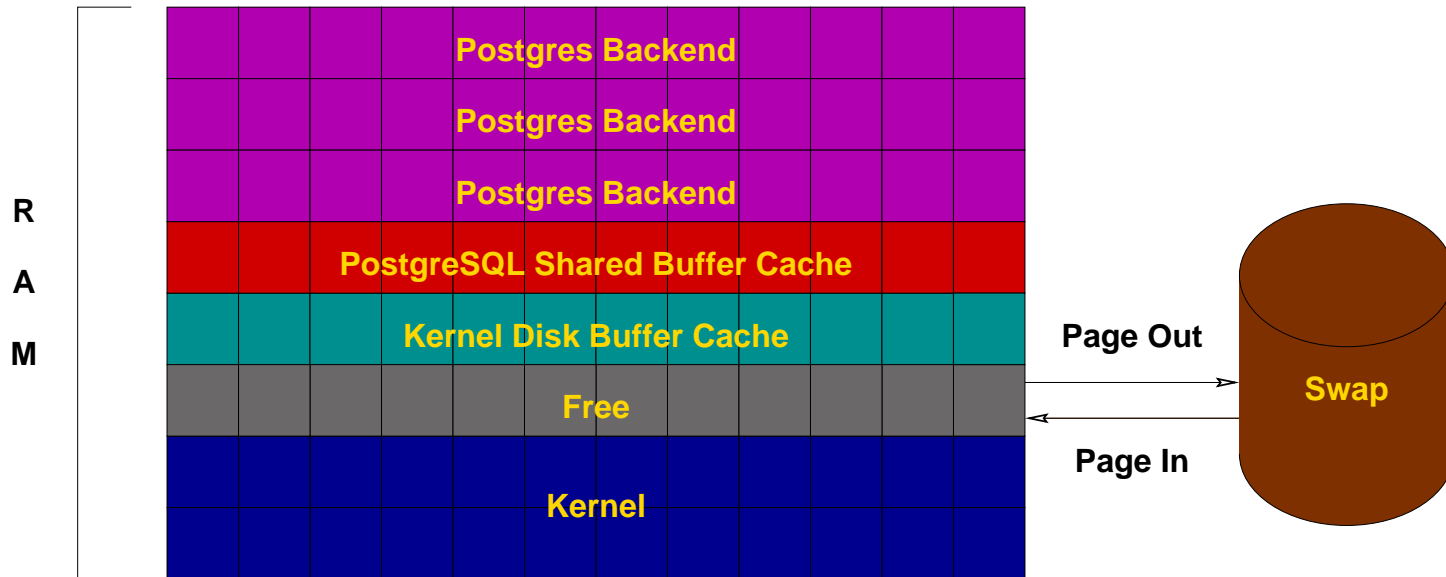
- Free Space Map -

```
#max_fsm_pages = 20000        # min max_fsm_relations*16, 6 bytes each
#max_fsm_relations = 1000     # min 100, ~50 bytes each
```

- Kernel Resource Usage -

```
#max_files_per_process = 1000 # min 25
#preload_libraries = ''
```

SIZING SHARED MEMORY



VACUUM AND BACKGROUND WRITER

- Cost-Based Vacuum Delay -

```
#vacuum_cost_delay = 0           # 0-1000 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        # 0-10000 credits
```

- Background writer -

```
#bgwriter_delay = 200           # 10-10000 milliseconds between rounds
#bgwriter_percent = 1           # 0-100% of dirty buffers in each round
#bgwriter_maxpages = 100        # 0-1000 buffers max per round
```

WRITE-AHEAD LOG (WAL)

- Settings -

```
#fsync = true                # turns forced synchronization on or off
#wal_sync_method = fsync    # the default varies across platforms:
                             # fsync, fdatasync, open_sync, or open_datasync
#wal_buffers = 8            # min 4, 8KB each
#commit_delay = 0          # range 0-100000, in microseconds
#commit_siblings = 5      # range 1-1000
```

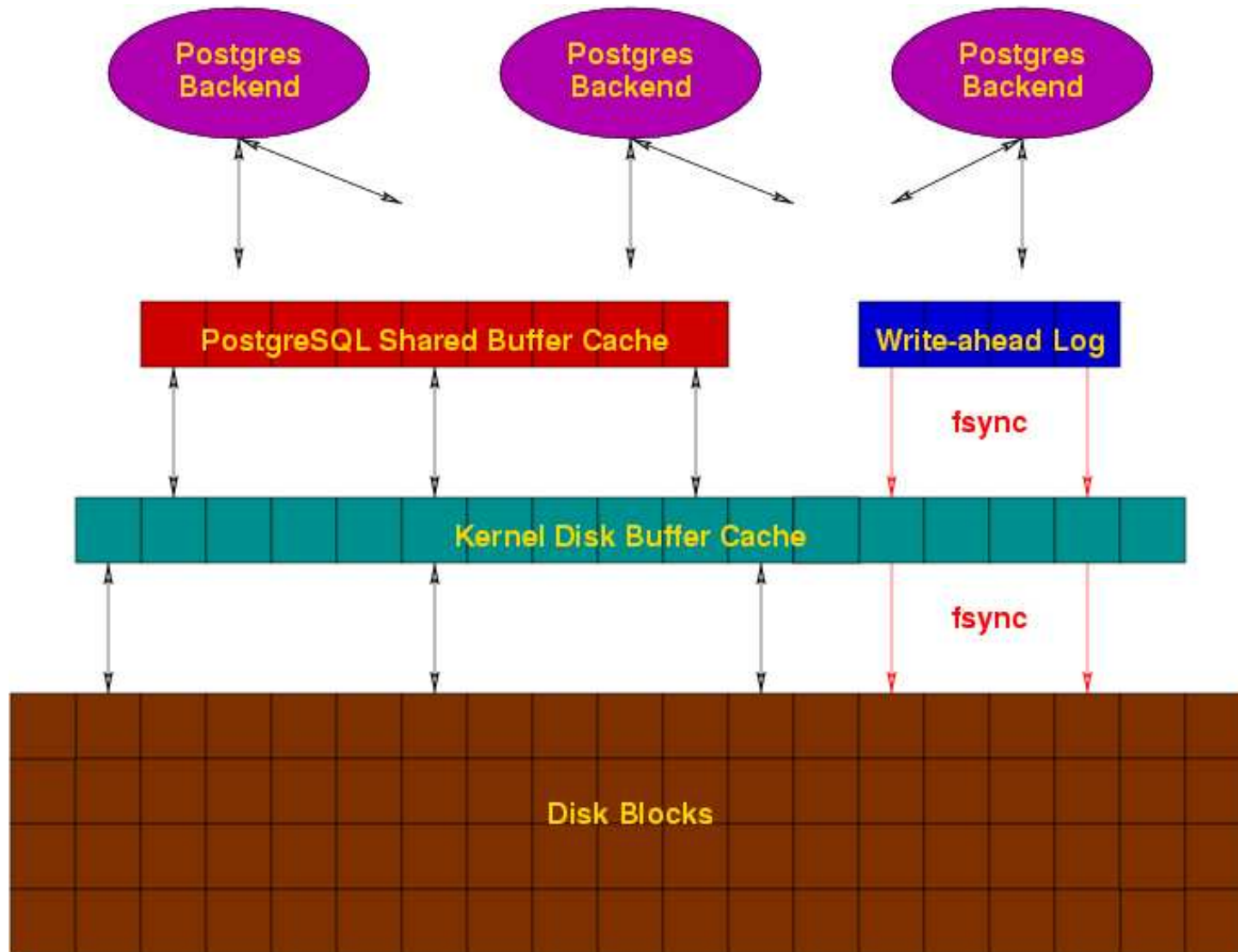
- Checkpoints -

```
#checkpoint_segments = 3    # in logfile segments, min 1, 16MB each
#checkpoint_timeout = 300   # range 30-3600, in seconds
#checkpoint_warning = 30    # 0 is off, in seconds
```

- Archiving -

```
#archive_command = ''      # command to use to archive a logfile segment
```

WRITE-AHEAD LOGGING (CONTINUED)



QUERY TUNING

- Planner Method Configuration -

```
#enable_hashagg = true
#enable_hashjoin = true
#enable_indexscan = true
#enable_mergejoin = true
#enable_nestloop = true
#enable_seqscan = true
#enable_sort = true
#enable_tidscan = true
```

- Planner Cost Constants -

```
#effective_cache_size = 1000    # typically 8KB each
#random_page_cost = 4          # units are one sequential page fetch cost
#cpu_tuple_cost = 0.01         # (same)
#cpu_index_tuple_cost = 0.001  # (same)
#cpu_operator_cost = 0.0025    # (same)
```

QUERY TUNING (CONTINUED)

- Genetic Query Optimizer -

#geqo = true

#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

- Other Planner Options -

#default_statistics_target = 10 # range 1-1000

#from_collapse_limit = 8

#join_collapse_limit = 8 # 1 disables collapsing of explicit JOINS

ERROR REPORTING AND LOGGING

- Where to Log -

```
#log_destination = 'stderr'    # Valid values are combinations of stderr,  
                                # syslog and eventlog, depending on  
                                # platform.
```

This is relevant when logging to stderr:

```
#redirect_stderr = false      # Enable capturing of stderr into log files.
```

These are only relevant if redirect_stderr is true:

```
#log_directory = 'pg_log'    # Directory where log files are written.
```

```
                                # May be specified absolute or relative to PGDATA
```

```
#log_filename = 'postgresql-%Y-%m-%d_%H%M%S.log' # Log file name pattern.
```

```
                                # May include strftime() escapes
```

ERROR REPORTING AND LOGGING (CONTINUED)

```
#log_truncate_on_rotation = false # If true, any 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 false, meaning append to existing
# files in all cases.

#log_rotation_age = 1440 # Automatic rotation of logfiles will happen after
# so many minutes. 0 to disable.

#log_rotation_size = 10240 # Automatic rotation of logfiles will happen after
# so many kilobytes of log output. 0 to disable.

# These are relevant when logging to syslog:
#syslog_facility = 'LOCAL0'
#syslog_ident = 'postgres'
```

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 = notice # Values, in order of decreasing detail:
# debug5, debug4, debug3, debug2, debug1,
# info, notice, warning, error, log, fatal,
# panic

#log_error_verbosity = default # terse, default, or verbose messages

#log_min_error_statement = panic # Values in order of increasing severity:
# debug5, debug4, debug3, debug2, debug1,
# info, notice, warning, error, panic(off)

#log_min_duration_statement = -1 # -1 is disabled, in milliseconds.

#silent_mode = false # DO NOT USE without syslog or redirect_stderr
```

WHAT TO LOG

```
#debug_print_parse = false
#debug_print_rewritten = false
#debug_print_plan = false
#debug_pretty_print = false
#log_connections = false
#log_disconnections = false
#log_duration = false
#log_line_prefix = ''

# e.g. '<%u%%d> '
# %u=user name %d=database name
# %r=remote host and port
# %p=PID %t=timestamp %i=command tag
# %c=session id %l=session line number
# %s=session start timestamp %x=transaction id
# %q=stop here in non-session processes
# %%='%'

#log_statement = 'none'
#log_hostname = false

# none, mod, ddl, all
```

RUNTIME STATISTICS

```
# - Statistics Monitoring -
```

```
#log_parser_stats = false  
#log_planner_stats = false  
#log_executor_stats = false  
#log_statement_stats = false
```

```
# - Query/Index Statistics Collector -
```

```
#stats_start_collector = true  
#stats_command_string = false  
#stats_block_level = false  
#stats_row_level = false  
#stats_reset_on_server_start = true
```

CLIENT CONNECTION DEFAULTS

- Statement Behavior -

```
#search_path = '$user,public' # schema names
#default_tablespace = '' # a tablespace name, or '' for default
#check_function_bodies = true
#default_transaction_isolation = 'read committed'
#default_transaction_read_only = false
#statement_timeout = 0 # 0 is disabled, in milliseconds
```

- Locale and Formatting -

```
#datestyle = 'iso, mdy'
#timezone = unknown # actually, defaults to TZ environment setting
#australian_timezones = false
#extra_float_digits = 0 # min -15, max 2
#client_encoding = sql_ascii # actually, defaults to database encoding
```

LOCALIZATION

```
# These settings are initialized by initdb -- they might 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
```

OTHER DEFAULTS

```
#explain_pretty_print = true  
#dynamic_library_path = '$libdir'
```

LOCK MANAGEMENT

```
#deadlock_timeout = 1000          # in milliseconds  
#max_locks_per_transaction = 64 # min 10, ~200*max_connections bytes each
```

VERSION/PLATFORM COMPATIBILITY

- Previous Postgres Versions -

#add_missing_from = true

#regex_flavor = advanced # advanced, extended, or basic

#sql_inheritance = true

#default_with_oids = true

INTERFACES

- Installing
 - Compiled Languages
 - Scripting Language
 - SPI
- Connection Pooling

INCLUDE FILES

```
$ ls -CF include/
```

```
ecpg_informix.h      internal/          pg_config_os.h     pgtypes_timestamp.h
ecpgerrno.h          libpq/            pgtypes_date.h     postgres_ext.h
ecpglib.h            libpq-fe.h        pgtypes_error.h    server/
ecpgtype.h           pg_config.h        pgtypes_interval.h sql3types.h
informix/            pg_config_manual.h pgtypes_numeric.h  sqlca.h
```

LIBRARY FILES

```
$ ls -CF lib/
```

```
ascii_and_mic.so*      libecpg_compat.a      libpq.so.3.2*        utf8_and_gb18030.so*
cyrillic_and_mic.so*  libecpg_compat.so@    pgxs/                utf8_and_gbk.so*
euc_cn_and_mic.so*    libecpg_compat.so.1@  plperl.so*           utf8_and_iso8859.so*
euc_jp_and_sjis.so*   libecpg_compat.so.1.1* plpgsql.so*          utf8_and_iso8859_1.so*
euc_kr_and_mic.so*    libpgport.a           pltcl.so*            utf8_and_johab.so*
euc_tw_and_big5.so*   libpgtypes.a          utf8_and_ascii.so*   utf8_and_sjis.so*
latin2_and_win1250.so* libpgtypes.so@        utf8_and_big5.so*    utf8_and_tcvn.so*
latin_and_mic.so*     libpgtypes.so.1@     utf8_and_cyrillic.so* utf8_and_uhc.so*
libecpg.a             libpgtypes.so.1.2*    utf8_and_euc_cn.so*  utf8_and_win1250.so*
libecpg.so@          libpq.a               utf8_and_euc_jp.so*  utf8_and_win1256.so*
libecpg.so.4@        libpq.so@             utf8_and_euc_kr.so*  utf8_and_win874.so*
libecpg.so.4.2*      libpq.so.3@          utf8_and_euc_tw.so*
```

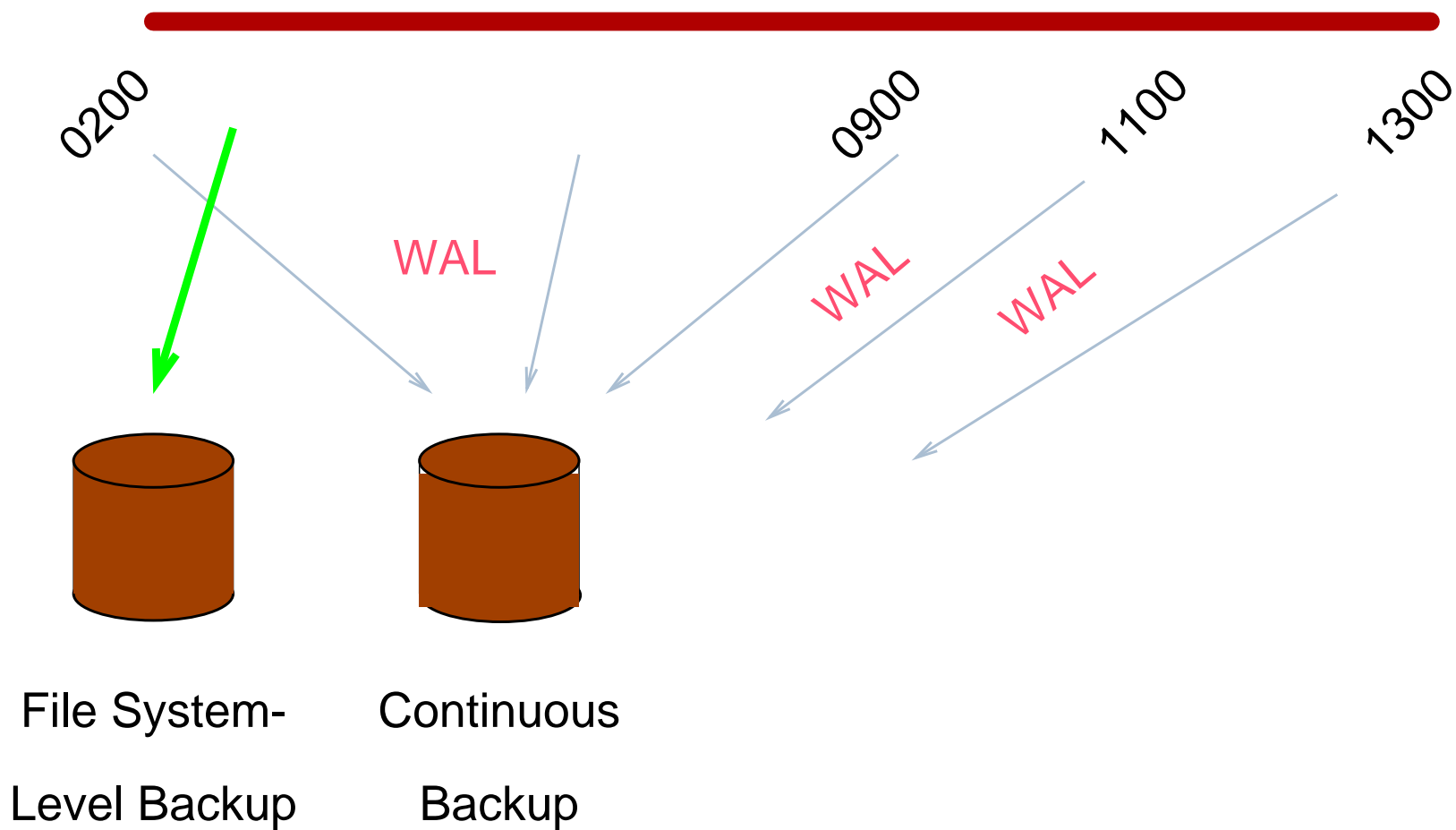
DAILY CHORES



BACKUP

- File system-level
 - TAR, CPIO WHILE SHUTDOWN
 - FILE SYSTEM SNAPSHOT
 - RSYNC, SHUTDOWN, RSYNC, RESTART
- pg_dump/pg_dumpall
- restore/pg_restore with custom format

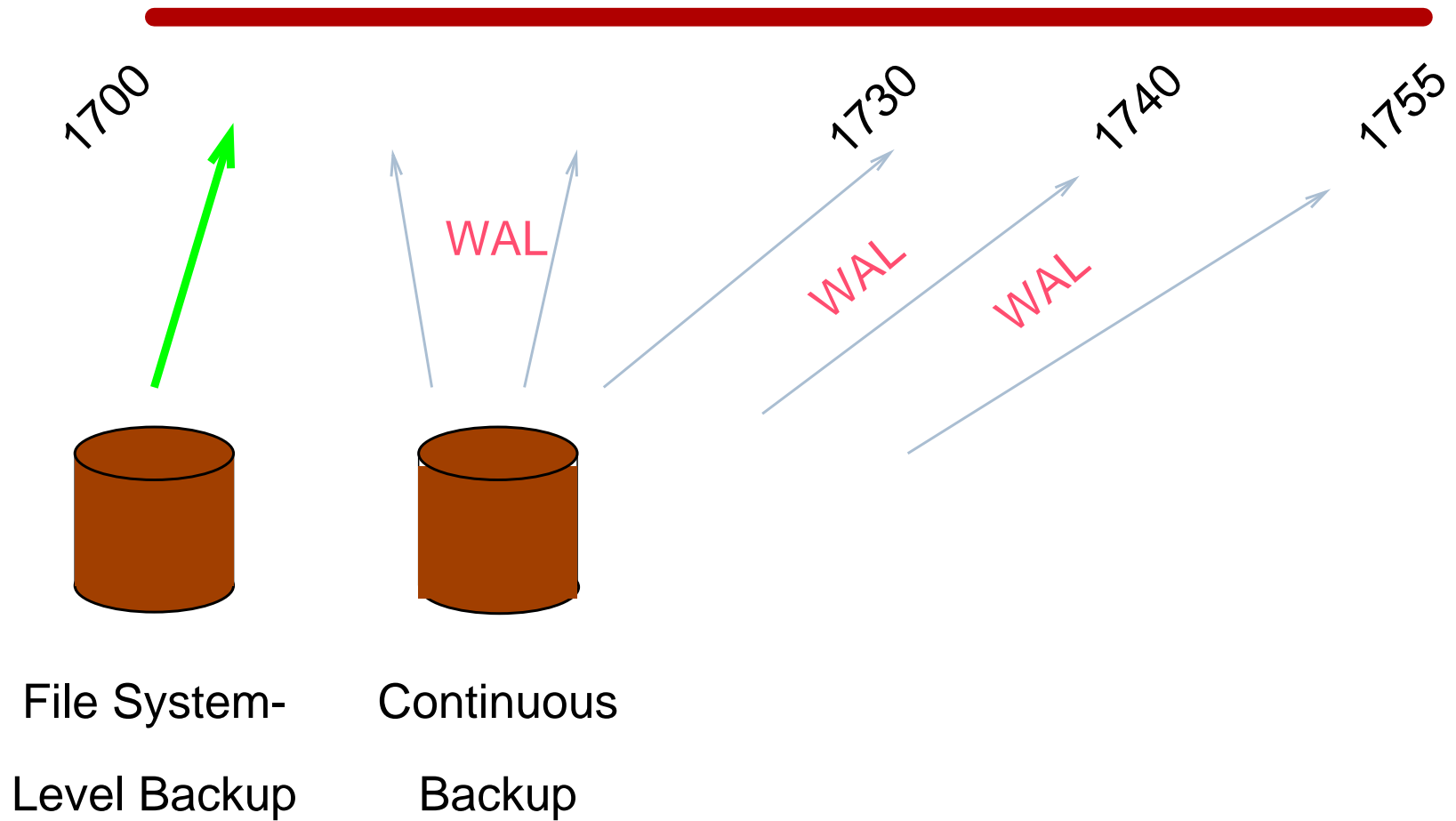
CONTINUOUS LOGGING POINT-IN-TIME RECOVERY (PITR)



PITR BACKUP PROCEDURES

1. `ARCHIVE_COMMAND = 'CP %P /MNT/SERVER/PGSQL/%F'`
2. `SELECT PG_START_BACKUP('LABEL');`
3. PERFORM FILE SYSTEM-LEVEL BACKUP (CAN BE INCONSISTENT)
4. `SELECT PG_STOP_BACKUP();`

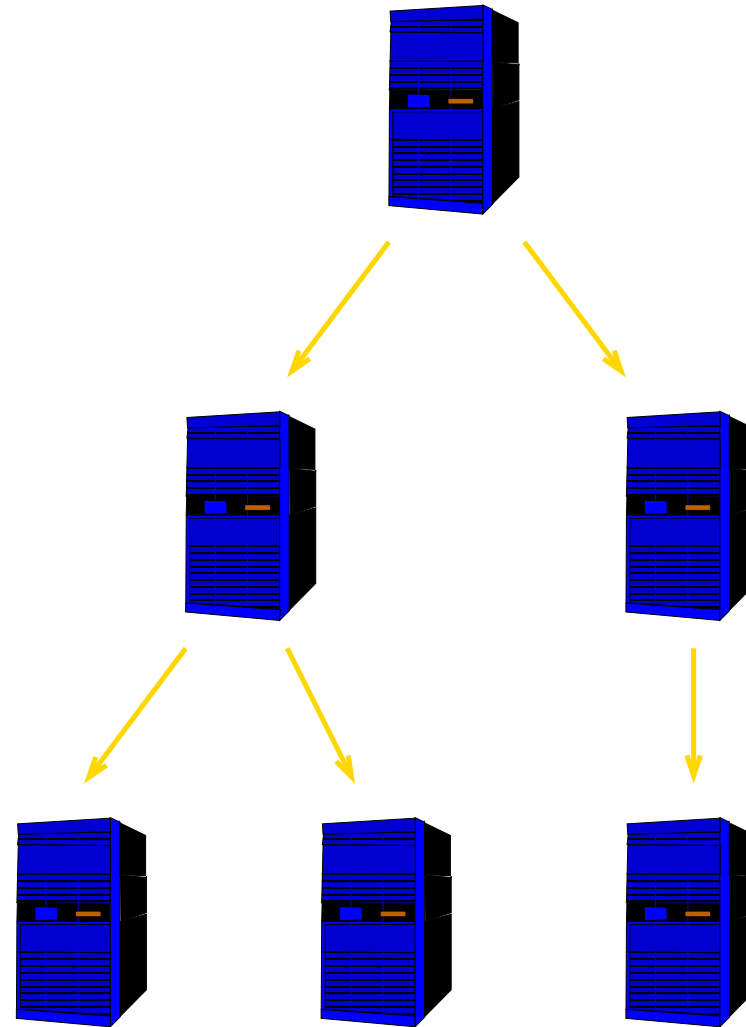
PITR RECOVERY



PITR RECOVERY PROCEDURES

1. STOP POSTMASTER
2. RESTORE FILE SYSTEM-LEVEL BACKUP
3. MAKE ADJUSTMENTS AS OUTLINED IN THE DOCUMENTATION
4. CREATE RECOVERY.CONF
5. ADD `RESTORE_COMMAND = 'CP /MNT/SERVER/PGSQL/%F %P'`
6. START THE POSTMASTER

MASTER-SLAVE REPLICATION - SLONY



OTHER SOLUTIONS

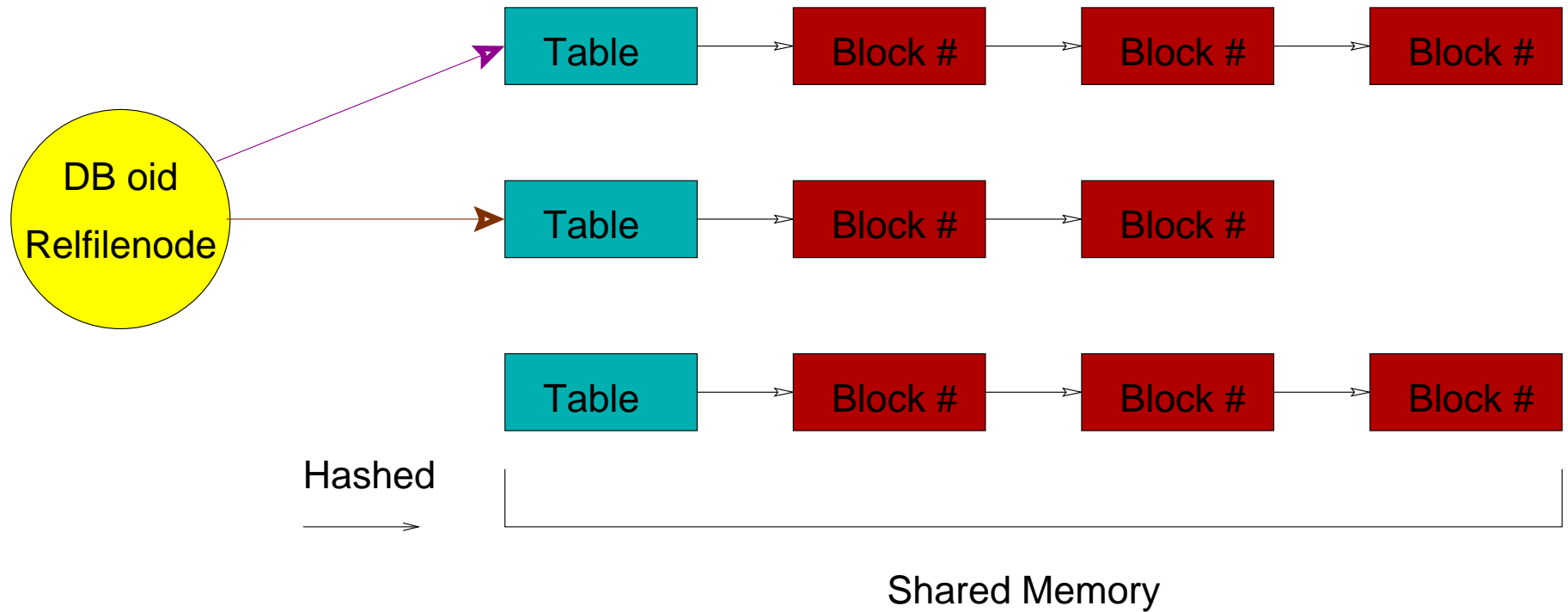
- MUTLI-MASTER REPLICATION: PGCLUSTER, SLONY II (UNDER DEVELOPMENT)
- POOLING: PGPOOL

DATA MAINTENANCE

- VACUUM (nonblocking), free space map
- VACUUM FULL
- ANALYZE

VACUUM

Free Space Map

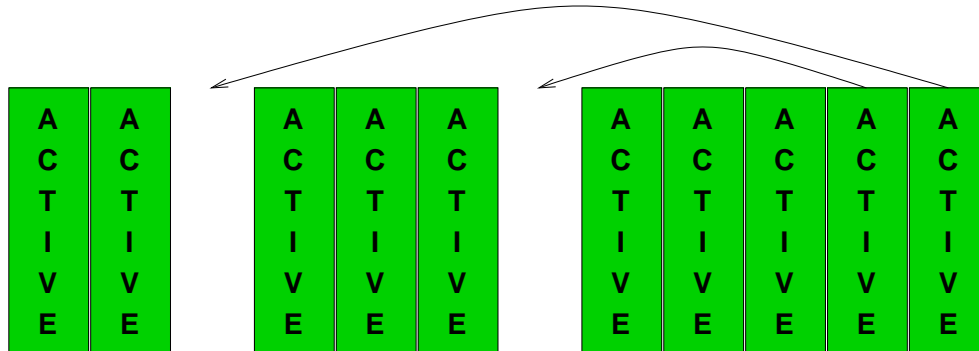


VACUUM FULL

Original Heap
With Expired
Rows Identified

A	A	E	A	A	A	E	A	A	A	A	A
C	C	X	C	C	C	X	C	C	C	C	C
T	T	P	T	T	T	P	T	T	T	T	T
I	I	I	I	I	I	I	I	I	I	I	I
V	V	R	V	V	V	R	V	V	V	V	V
E	E	E	E	E	E	E	E	E	E	E	E

Move Trailing
Rows Into Expired
Slots



Truncate File

A	A	A	A	A	A	A	A	A	A		
C	C	C	C	C	C	C	C	C	C		
T	T	T	T	T	T	T	T	T	T		
I	I	I	I	I	I	I	I	I	I		
V	V	V	V	V	V	V	V	V	V		
E	E	E	E	E	E	E	E	E	E		

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*

AUTOMATING TASKS

```
0 3 * * * root psql -c 'VACUUM FULL;' test
```

```
0 3 * * * root vacuumdb -a -f
```

MONITORING ACTIVE SESSIONS



PS

```
$ ps -Upostgres
```

PID	TT	STAT	TIME	COMMAND
2125	??	Ss	0:00.26	./bin/postmaster -i
2142	??	S	0:00.03	stats buffer process (postmaster)
2143	??	S	0:00.06	stats collector process (postmaster)
3341	??	I	0:00.07	postgres test [local] idle (postmaster)
3340	p6	I+	0:00.03	psql test

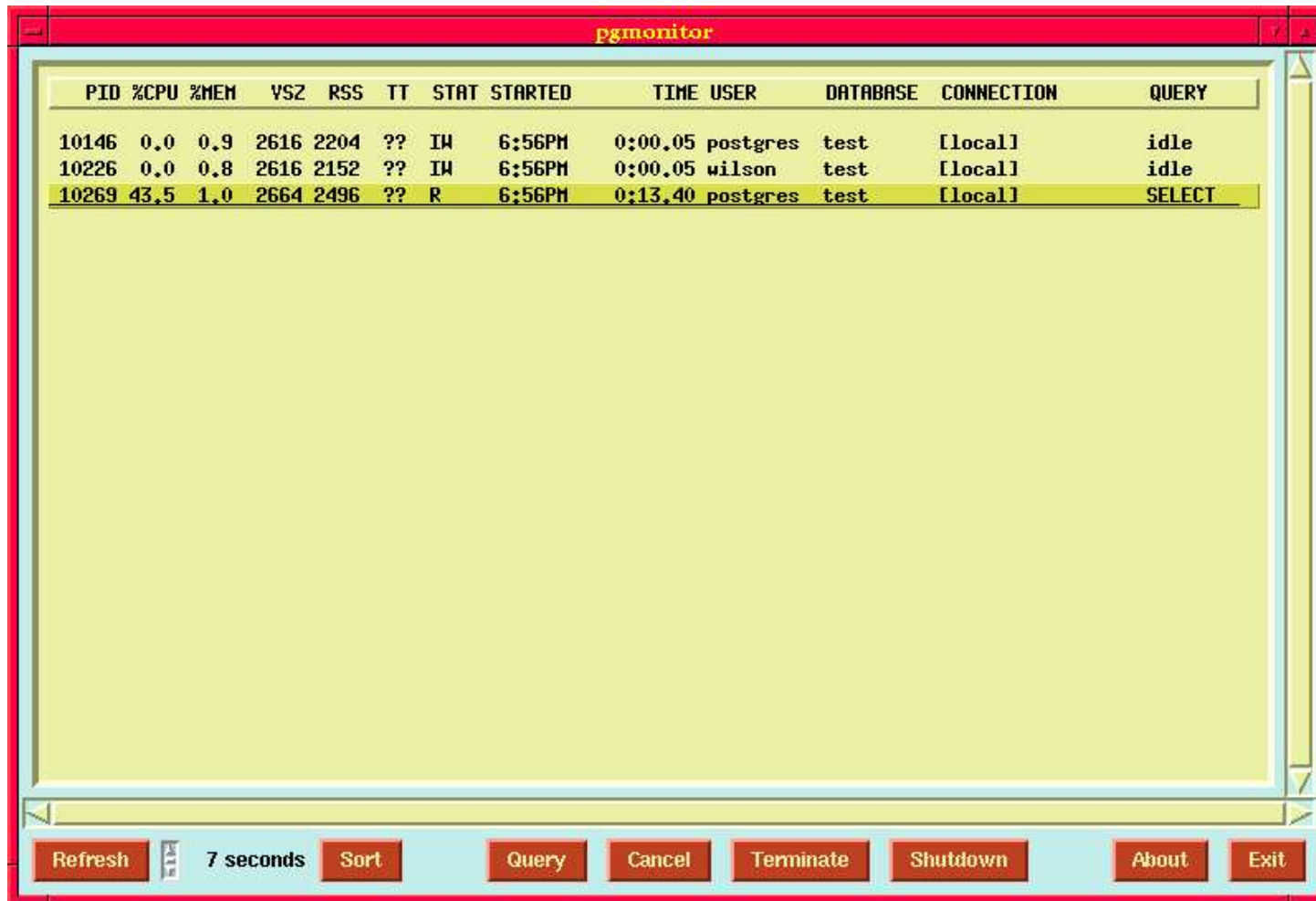
TOP

\$ 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

PGMONITOR



The screenshot shows a window titled "pgmonitor" with a table of process information and a control panel at the bottom. The table has the following columns: PID, %CPU, %MEM, VSZ, RSS, TT, STAT, STARTED, TIME, USER, DATABASE, CONNECTION, and QUERY. The data rows are:

PID	%CPU	%MEM	VSZ	RSS	TT	STAT	STARTED	TIME	USER	DATABASE	CONNECTION	QUERY
10146	0,0	0,9	2616	2204	??	IM	6:56PM	0:00,05	postgres	test	[local]	idle
10226	0,0	0,8	2616	2152	??	IM	6:56PM	0:00,05	wilson	test	[local]	idle
10269	43,5	1,0	2664	2496	??	R	6:56PM	0:13,40	postgres	test	[local]	SELECT

The control panel at the bottom includes a "Refresh" button, a refresh interval of "7 seconds", a "Sort" button, and several action buttons: "Query", "Cancel", "Terminate", "Shutdown", "About", and "Exit".

QUERY MONITORING

```
stats_command_string = true
```

```
$ pg_ctl reload
```

```
test=> SELECT * FROM pg_stat_activity;
```

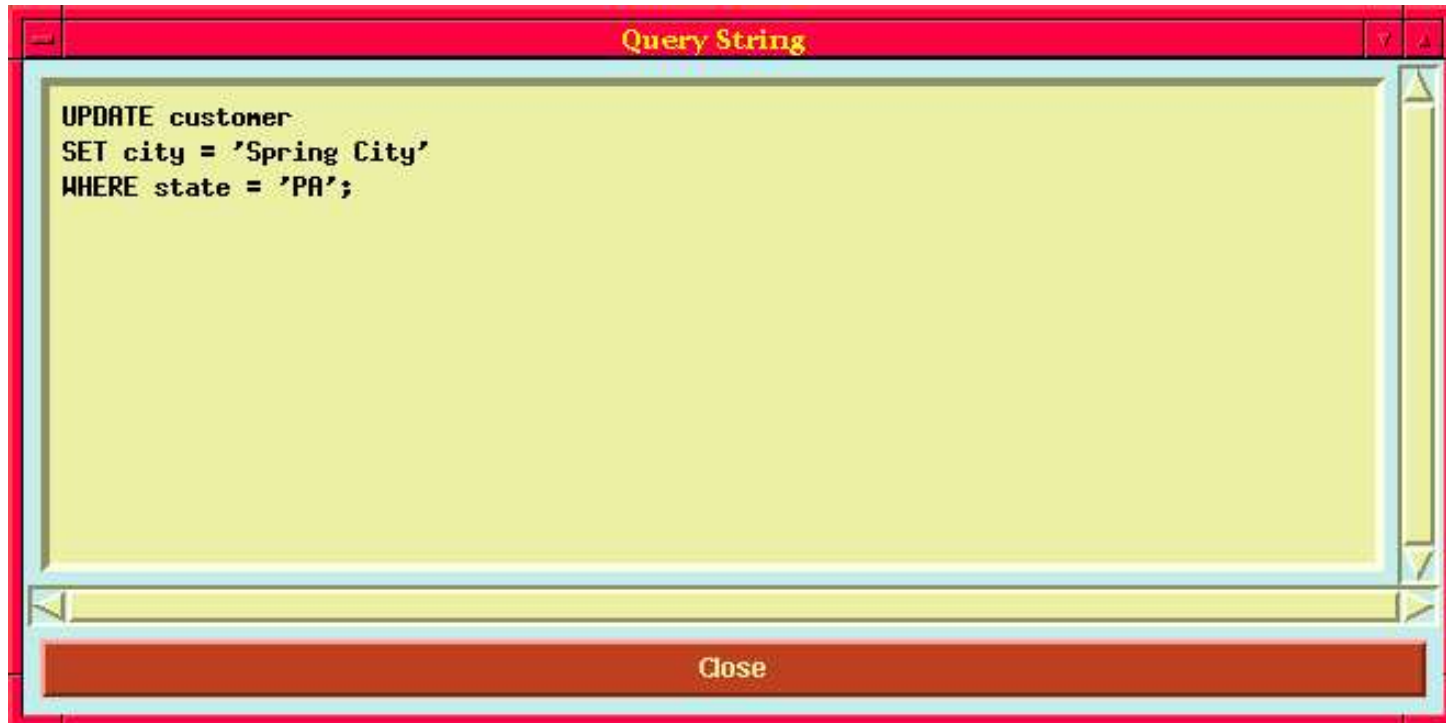
```
  datid | datname | procpid | usesysid | username | current_  
query   |         |         |          |          | query_start
```

```
-----+-----+-----+-----+-----+-----+-----+-----+-----
```

```
-----
```

```
 17230 | test   |    377 |         1 | post-  
gres   | <IDLE> |         | 2005-03-03 12:05:52.888076-05  
 17230 | test   |    417 |         1 | postgres | select * from pg_  
class | 2005-03-03 12:06:06.555737-05  
(2 rows)
```

PGMONITOR



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;
```

datid	datname	numbackends	xact_commit	xact_rollback	blks_read	blks_hit
16570	test	1	16	3	151	880
1	template1	0	0	0	0	0
16569	template0	0	0	0	0	0

(3 rows)

TABLE ACTIVITY

```
test=> SELECT * FROM pg_stat_all_tables;
```

relid	relname	seq_scan	seq_tup_read	idx_scan	idx_tup_fetch	n_tup_ins	n_tup_upd	n_tup_del
1247	pg_type	1	10	26	26	0	0	0
1249	pg_attribute	0	0	28	75	0	0	0
1255	pg_proc	1	1	60	55	0	0	0
1259	pg_class	194	21268	36	36	0	0	0
1260	pg_shadow	6	6	4	4	0	0	0

TABLE BLOCK ACTIVITY

```
test=> SELECT * FROM pg_statio_all_tables;
```

relid	relname	heap_blks_read	heap_blks_hit	idx_blks_read	idx_blks_hit	toast...
1247	pg_type	5	25	4	54	
1249	pg_attribute	13	88	9	93	
1255	pg_proc	9	47	33	149	
1259	pg_class	0	1147	13	93	
1260	pg_shadow	4	6	8	0	

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	mtps	sps	tps	mtps	sps	tps	mtps	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

```
play=# SELECT relfilenode, relpages * 8 AS kilobytes
play-# FROM pg_class
play-# WHERE relname = 'customer';
relfilenode | kilobytes
-----+-----
          16806 |          480
(1 row)
```

VACUUM REQUIRED. DBSIZE AVAILABLE.

TOAST USAGE

```
play=# SELECT relname, relpages * 8 AS kilobytes
play-# FROM pg_class
play-# WHERE relname = 'pg_toast_16806' OR
play-#         relname = 'pg_toast_16806_index'
play-# ORDER BY relname;
```

relname	kilobytes
pg_toast_16806	0
pg_toast_16806_index	1

INDEX USAGE

```
play=# SELECT c2.relname, c2.relpages * 8 AS kilobytes
play-# FROM pg_class c, pg_class c2, pg_index i
play-# WHERE c.relname = 'customer' AND
play-#         c.oid = i.indrelid AND
play-#         c2.oid = i.indexrelid
play-# ORDER BY c2.relname;
```

relname	kilobytes
customer_id_index	26

LARGEST TABLES

```
play=# SELECT relname, relpages * 8  
play=# FROM pg_class  
play=# ORDER BY relpages DESC;
```

relname		kilobytes
-----+-----		
bigtable		3290
customer		3144

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 space for every db object
```

```
$ du * | while read SIZE RELFILENODE
```

```
> do
```

```
>     echo "$SIZE      'oid2name -q -d test -o $RELFILNODE'"
```

```
> done
```

```
24      18737 = ips
```

```
36      18722 = cities
```

```
...
```

```

$ # same as above, but sort by largest first
$ du * | while read SIZE OID
> do
>     echo "$SIZE      'oid2name -q -d test -o $OID'"
> done |
> sort -rn
2048    19324 = bigtable
1950    23903 = customers
...

```

```

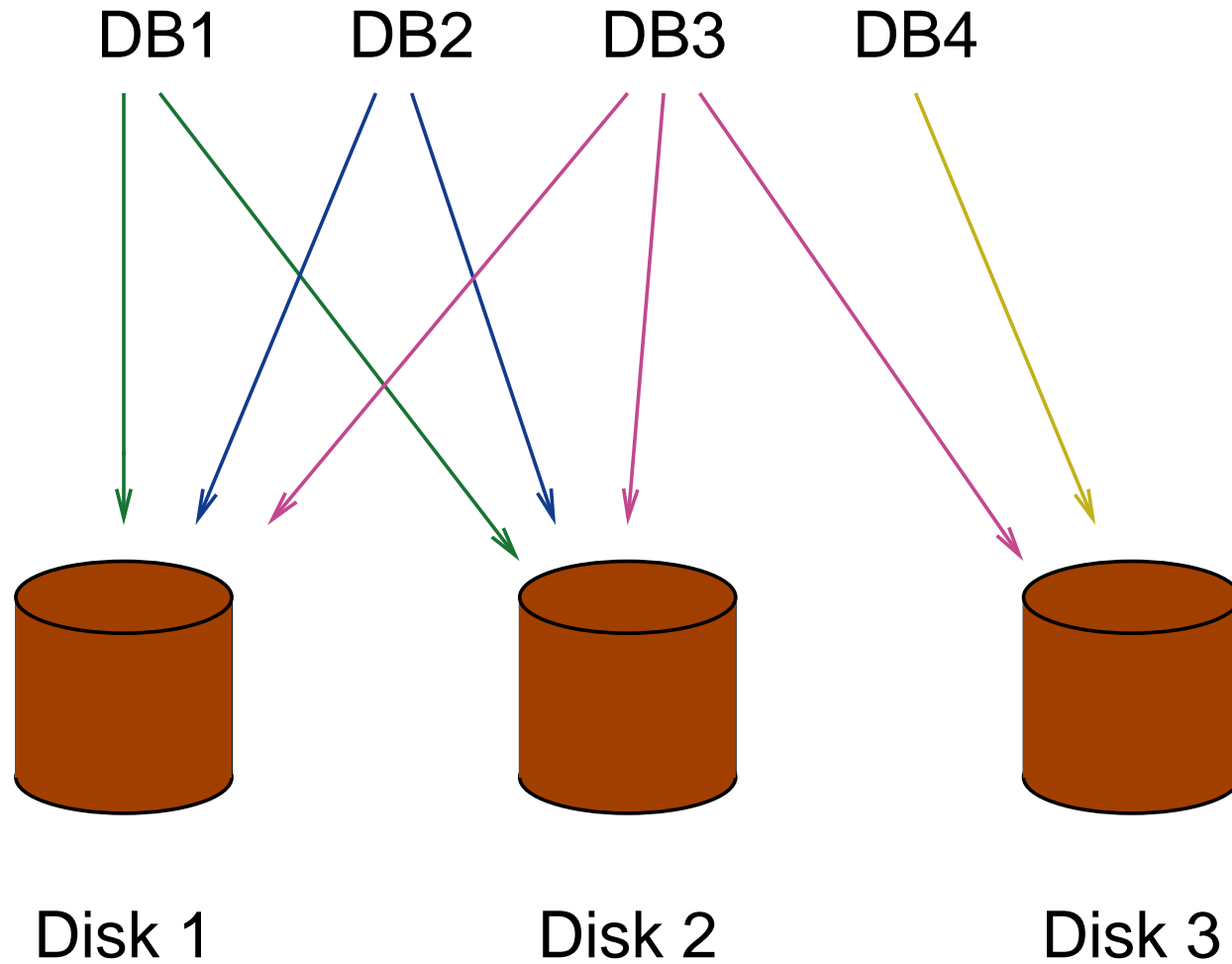
$ # show disk usage per database
$ cd /usr/local/pgsql/data/base
$ du -s * |
> while read SIZE OID
> do
>     echo "$SIZE      'oid2name -q | grep ^$OID' '"
> done |
> sort -rn
2256          18721 = test
2135          18735 = postgres

```

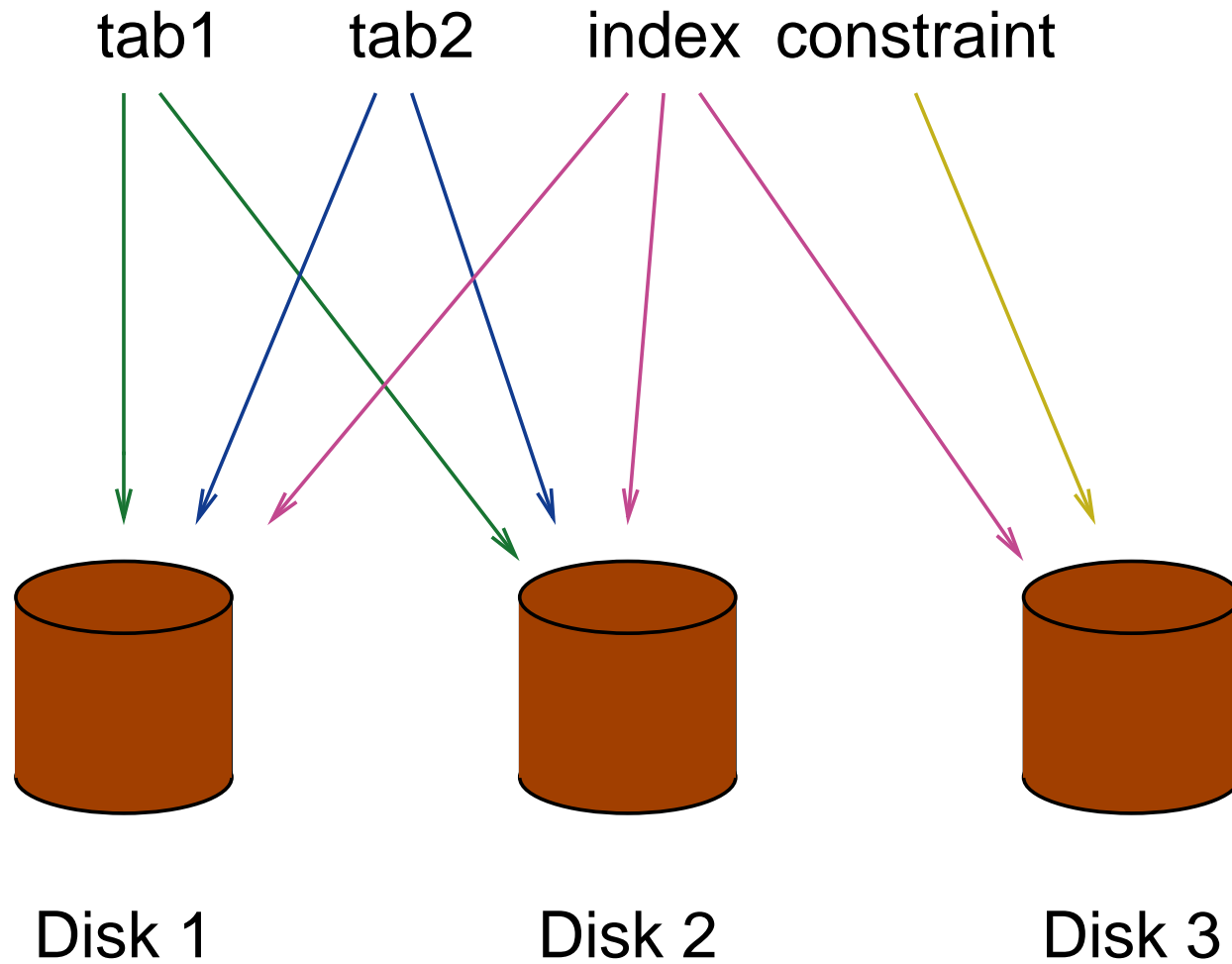
DISK BALANCING

- 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
- PGPHPADMIN
- PGACCESS

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

