Visualizing Postgres

michael.glaesemann@myyearbook.com

PostgreSQL Conference 2009 Japan
2009-11-20

http://seespotcode.net/postgres/jpug-2009-visualizing-postgres.pdf

myYearbook.com

casual social network
founded in 2006
Google Analytics

by gender	
55%	female
45%	male

by age	
43%	13–17
23%	18–24
21%	25–34
13%	35+

comScore Teens Category (April 2009)

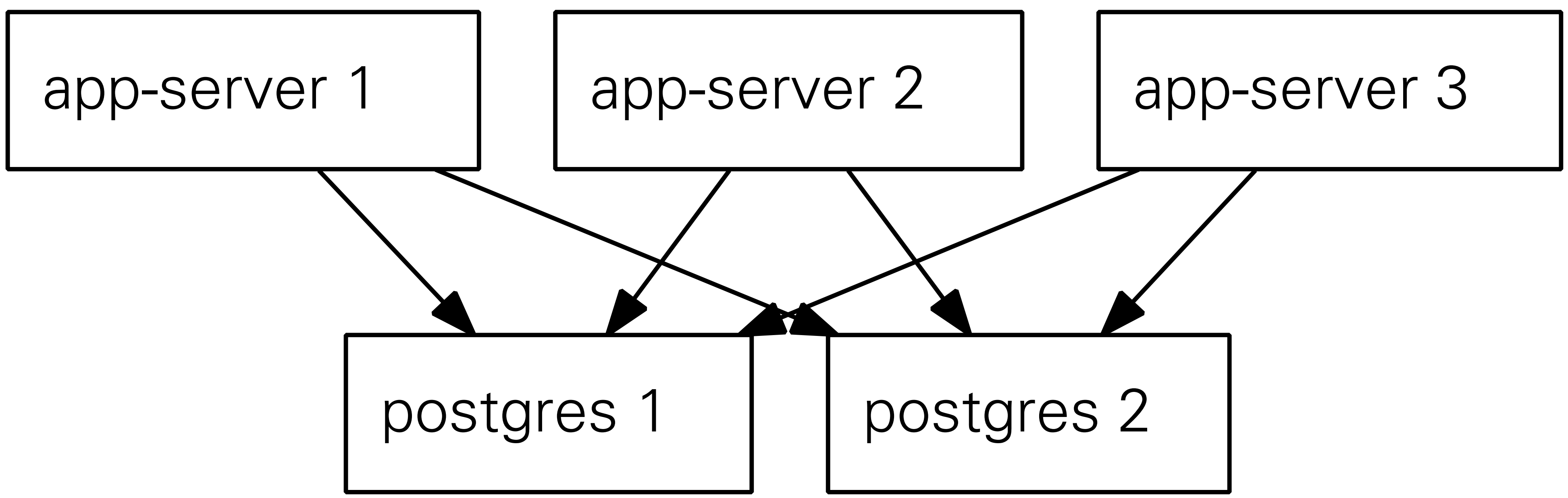
rank	site	visits (K)	uniques (K)	minutes (M)	page views (M)
1	**myYearbook**	**55,808**	**4,604**	**851**	**1,630**
2	MEEZ	8,629	1,407	226	469
3	Zwinky	11,558	3,691	153	108
4	Hearst Teen Network	6,940	2,314	53	117
5	Quizilla	7,924	2,058	75	75

comScore page views
July 2009

rank	site	views (M)
20	GaiaOnline	1,105
21	Chase	1,056
22	ESPN	984
23	**myYearbook**	**953**
24	Wikipedia	903
25	Onemanga	840
26	Mapquest	833
27	Foxsports	815

comScore time spent
July 2009

rank	site	minutes (M)
20	Amazon	806
21	CNN	744
22	GaiaOnline	709
23	Bing	707
24	MSNBC	691
25	**myYearbook**	**678**
26	Iwin	670
27	NickJr	665


app-server 1
app-server 2
app-server 3
postgres 1
postgres 2

more activity?

views/month 2007 100M ➔ 2009 1.5G

more TPS

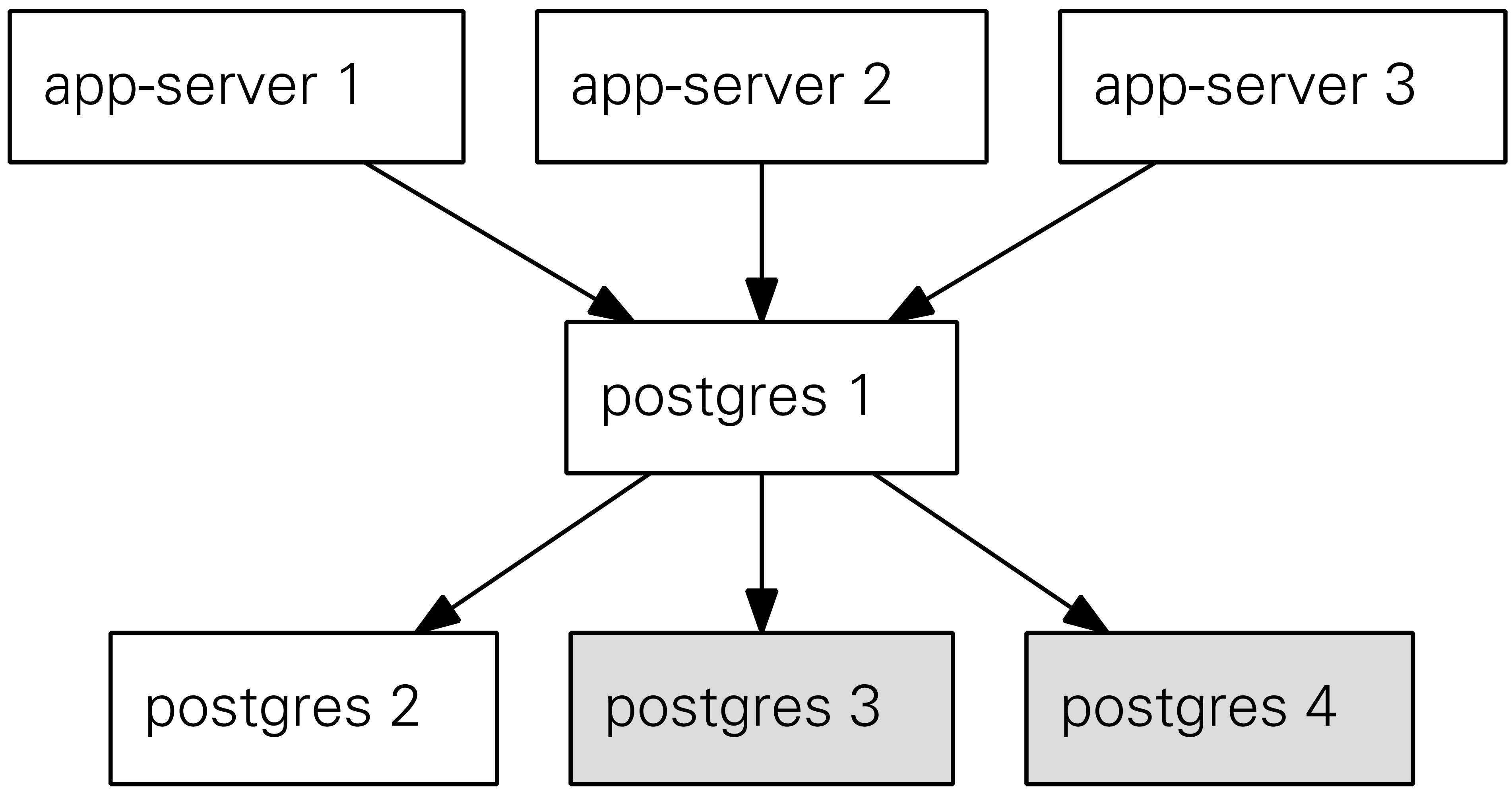
more servers

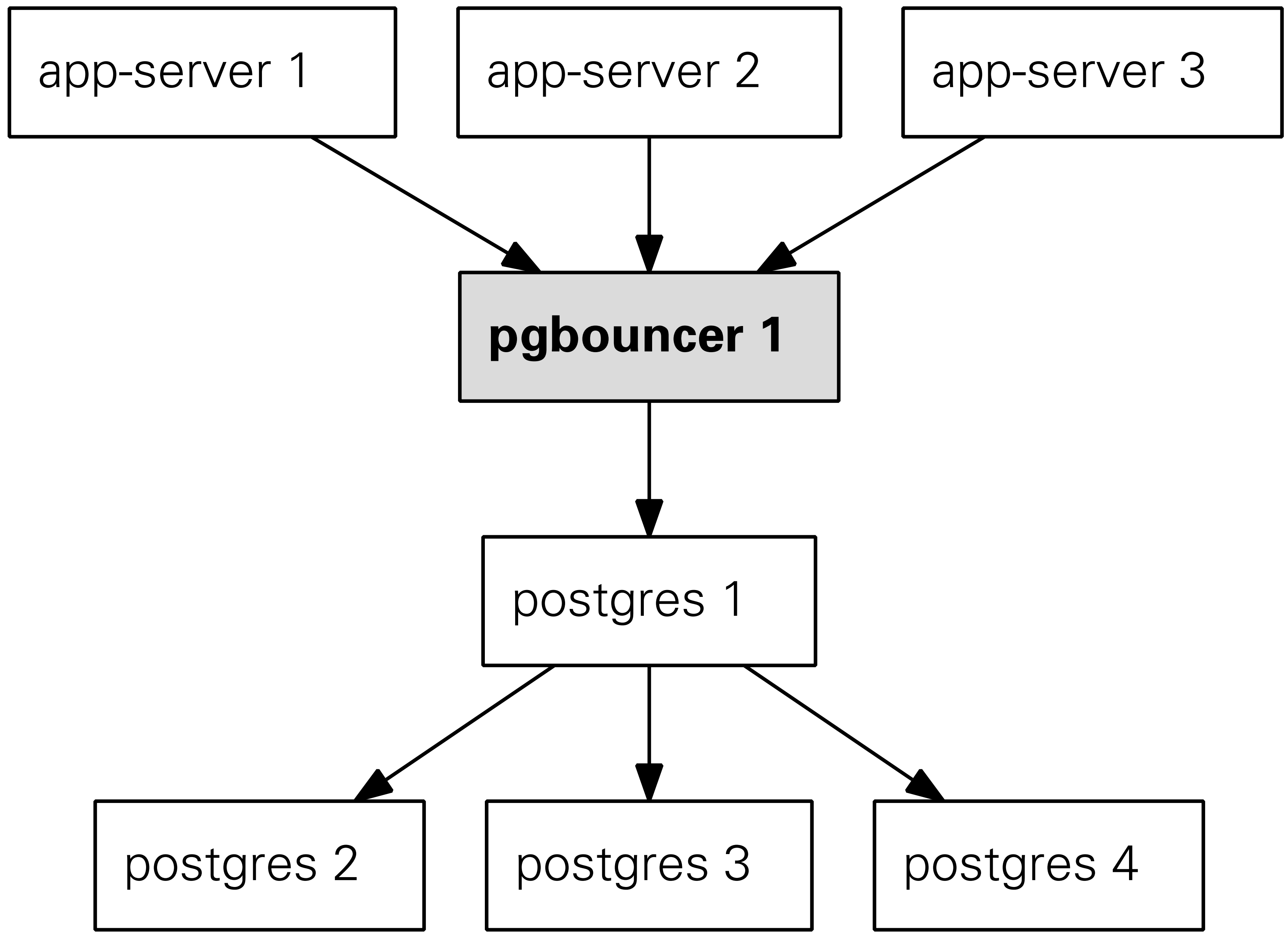
more connections

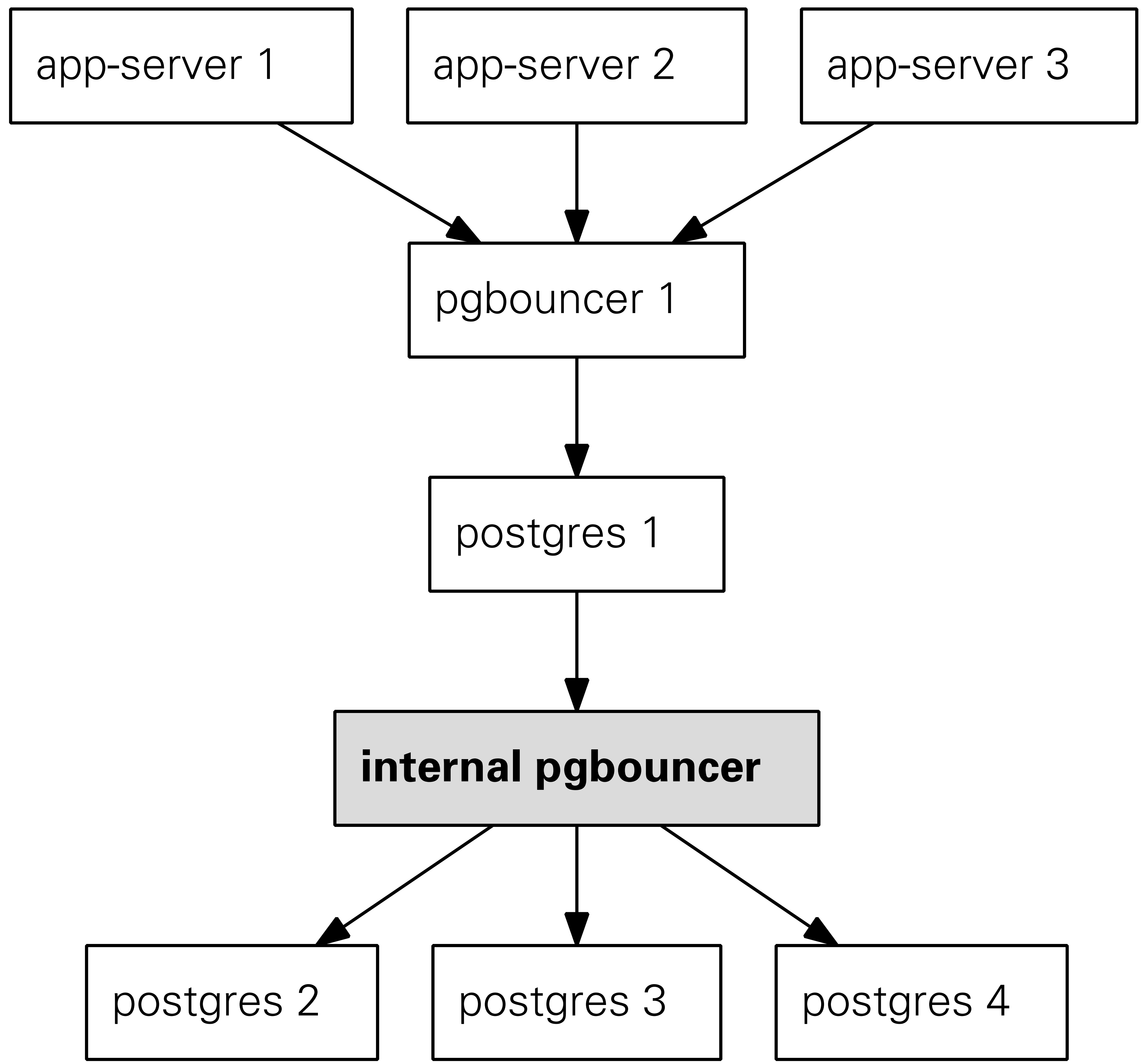
more configuration

more pain!

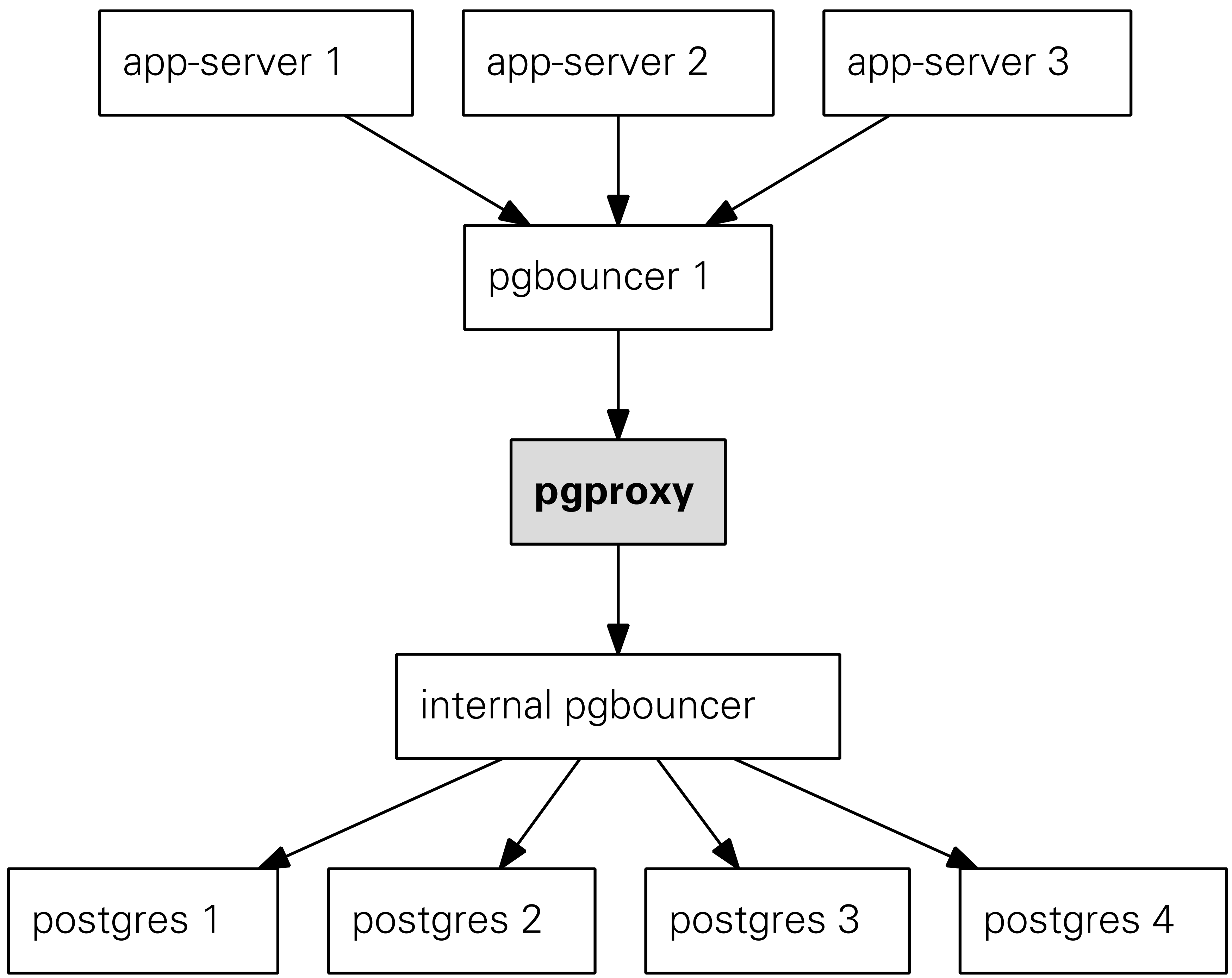
fewer connections!



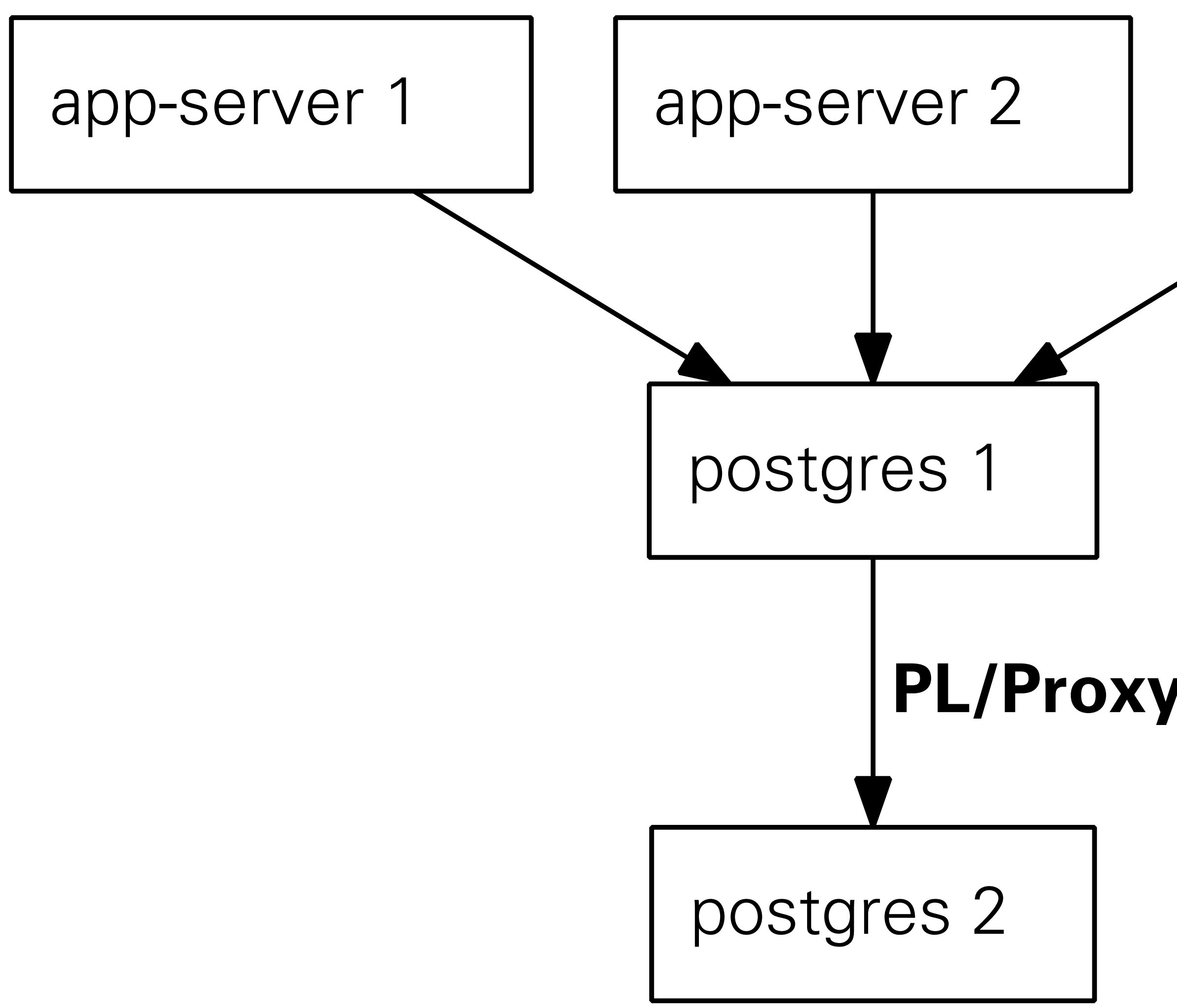

app-server 1
app-server 2
app-server 3
pgbouncer 1
postgres 1
postgres 2
postgres 3
postgres 4

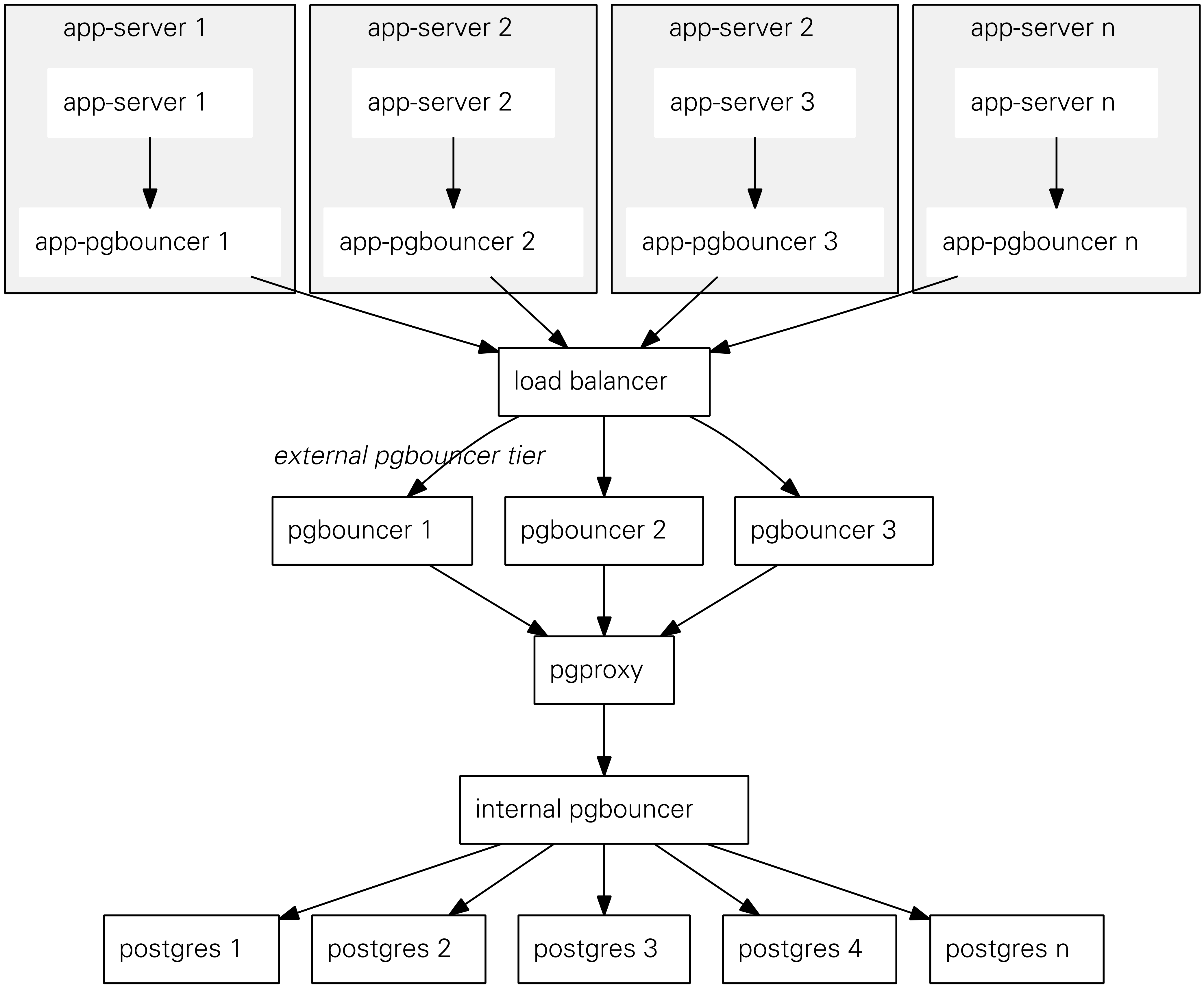

app-server 1
app-server 2
app-server 3
pgbouncer 1
postgres 1
internal pgbouncer
postgres 2
postgres 3
postgres 4

reduce TPS
less configuration



less configuration
fewer connections




app-server 1
app-server 1
app-pgbouncer 1
app-server 2
app-server 2
app-pgbouncer 2
app-server 2
app-server 3
app-pgbouncer 3
app-server n
app-server n
app-pgbouncer n
load balancer
external pgbouncer tier
pgbouncer 1
pgbouncer 2
pgbouncer 3
pgproxy
internal pgbouncer
postgres 1
postgres 2
postgres 3
postgres 4
postgres n

reduce TPS!

- ➛ memcached
  - 1 TB
  - get 140K/s, set 15K/s
  - pgfouine

simplify interface

➜ PL/Proxy function API

asynchronous

➤ message queues

28 servers avg 90% idle

464 cores

3.3 TB memory

3.8 TB on disk

35 TB total disk

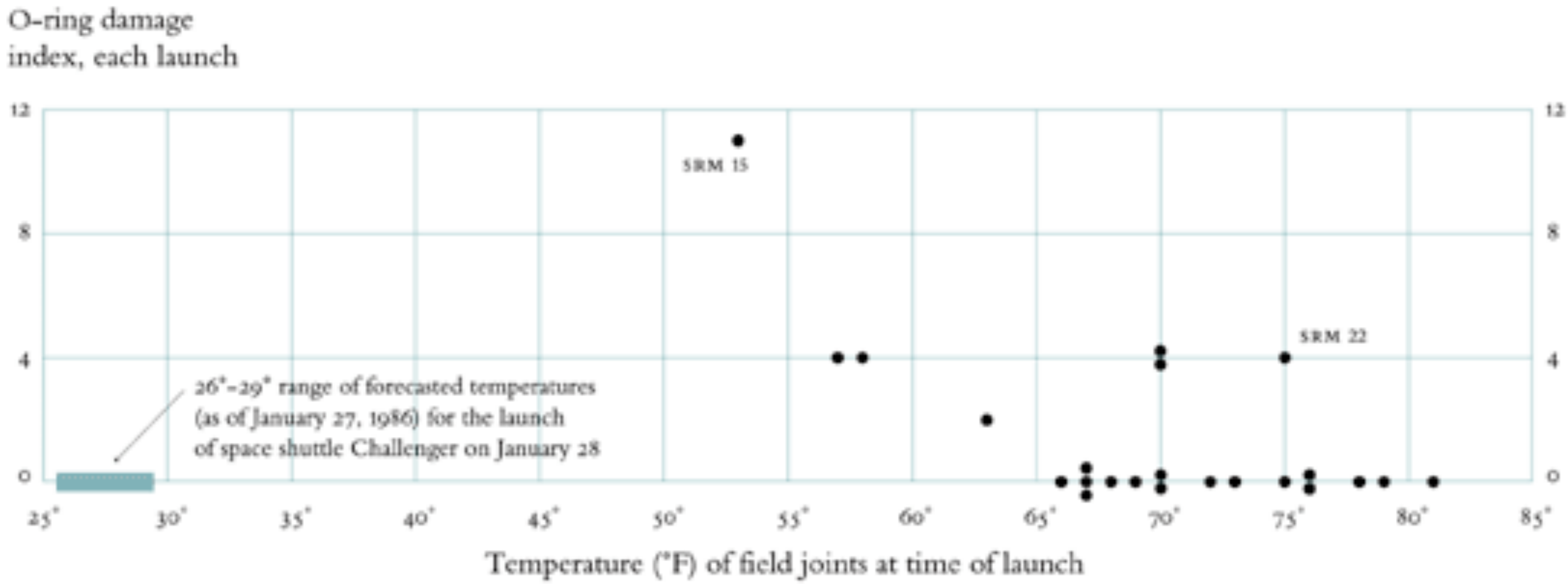
15K avg TPS (> 27K)

measure

http://explain-analyze.info

statistics

rocket science



Tufte

7

analyze
host
bloat
DTrace/SystemTap
logs
contrib
statistics collector

ANALYZE

`pg_stats`

`pg_class`

cpu

memory

io

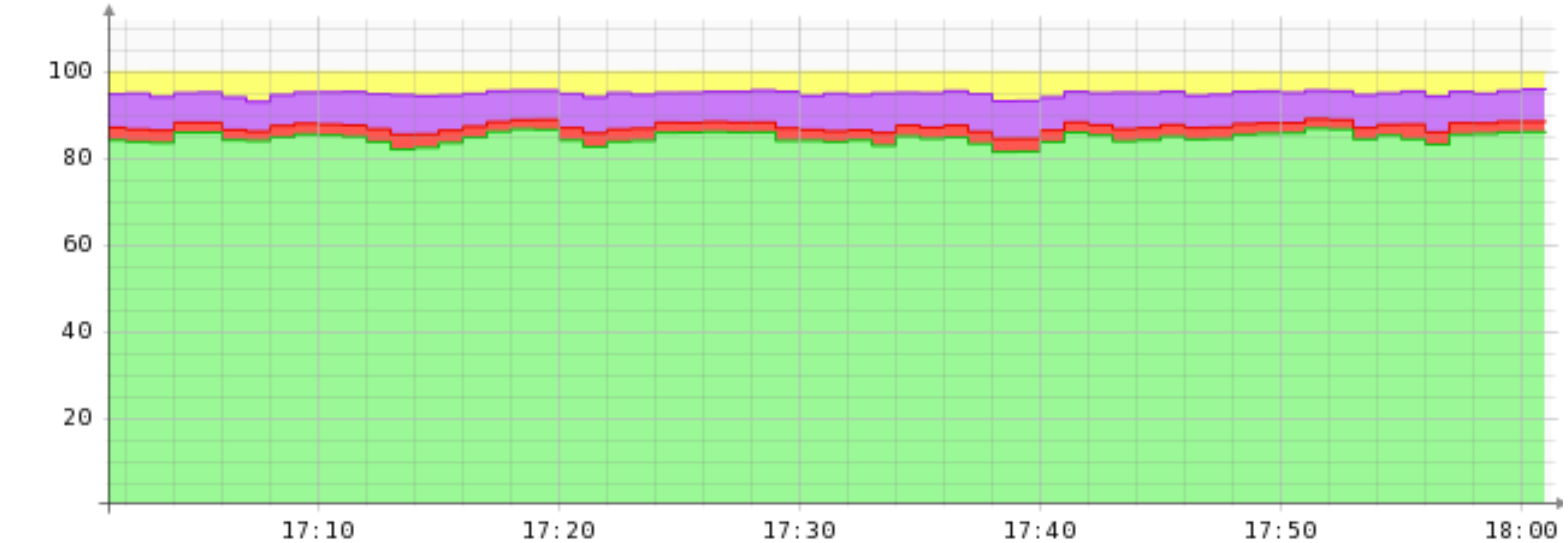
snmp

rrd

Staplr

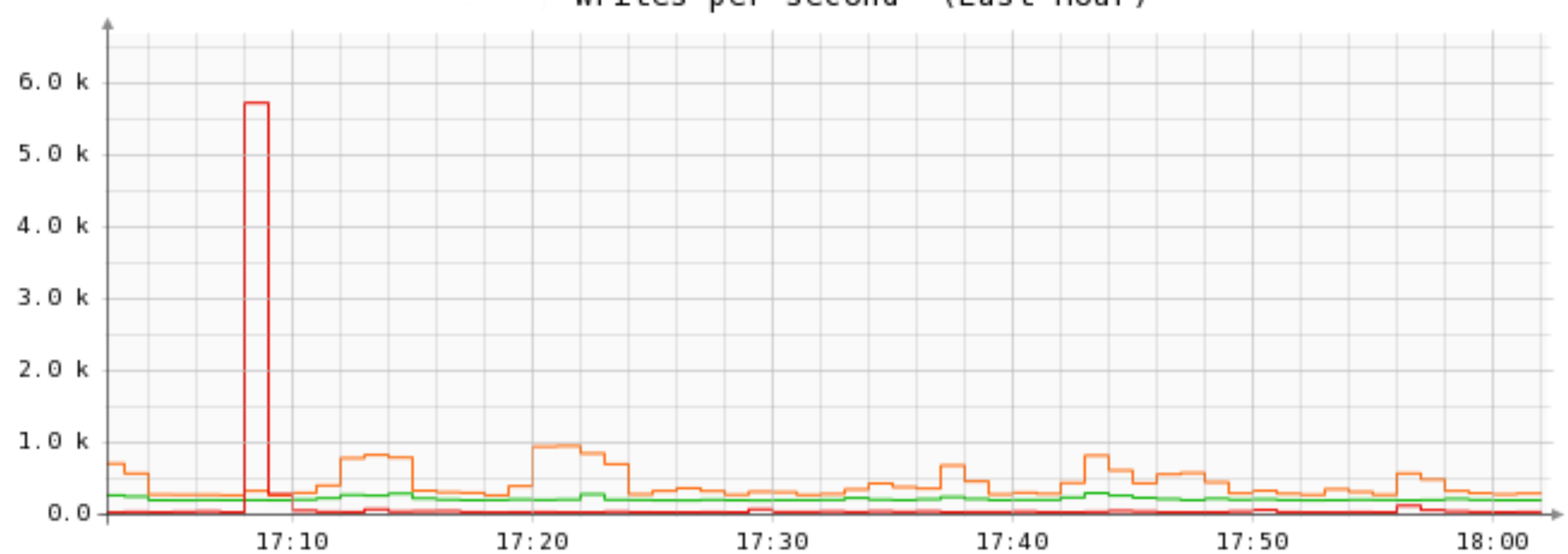
http://area51.myyearbook.com

CPU Utilization per second (Last Hour)

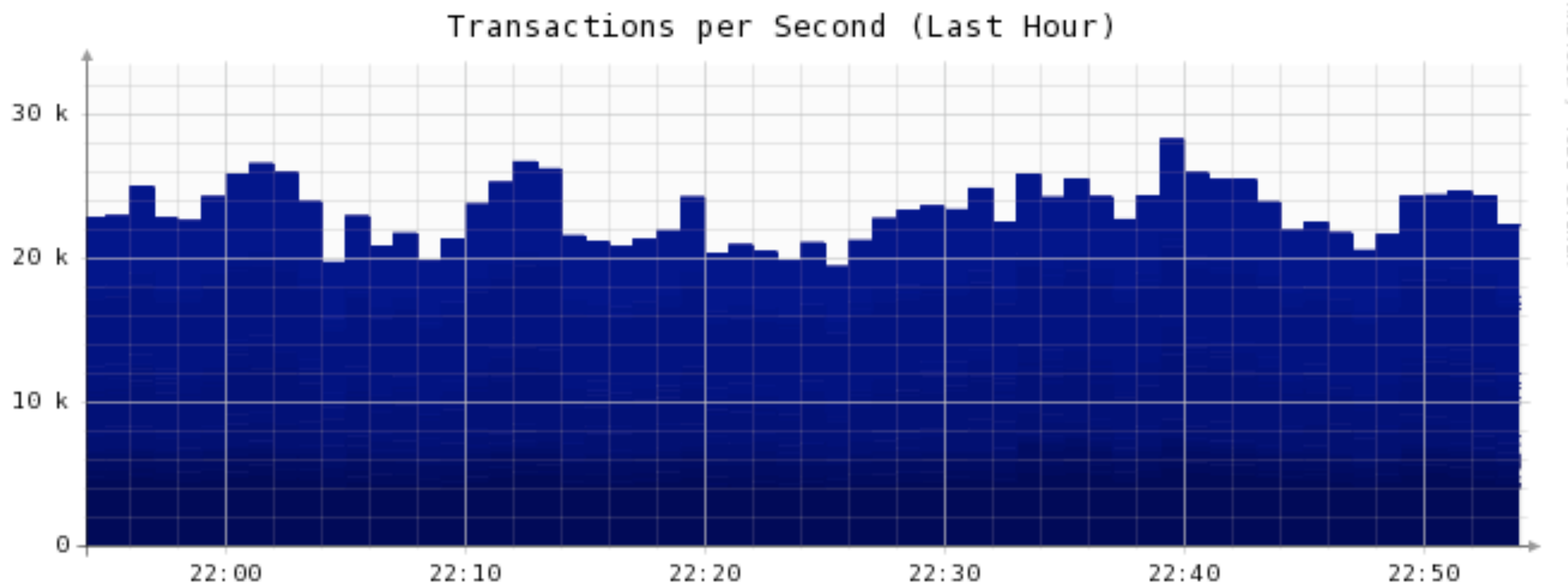


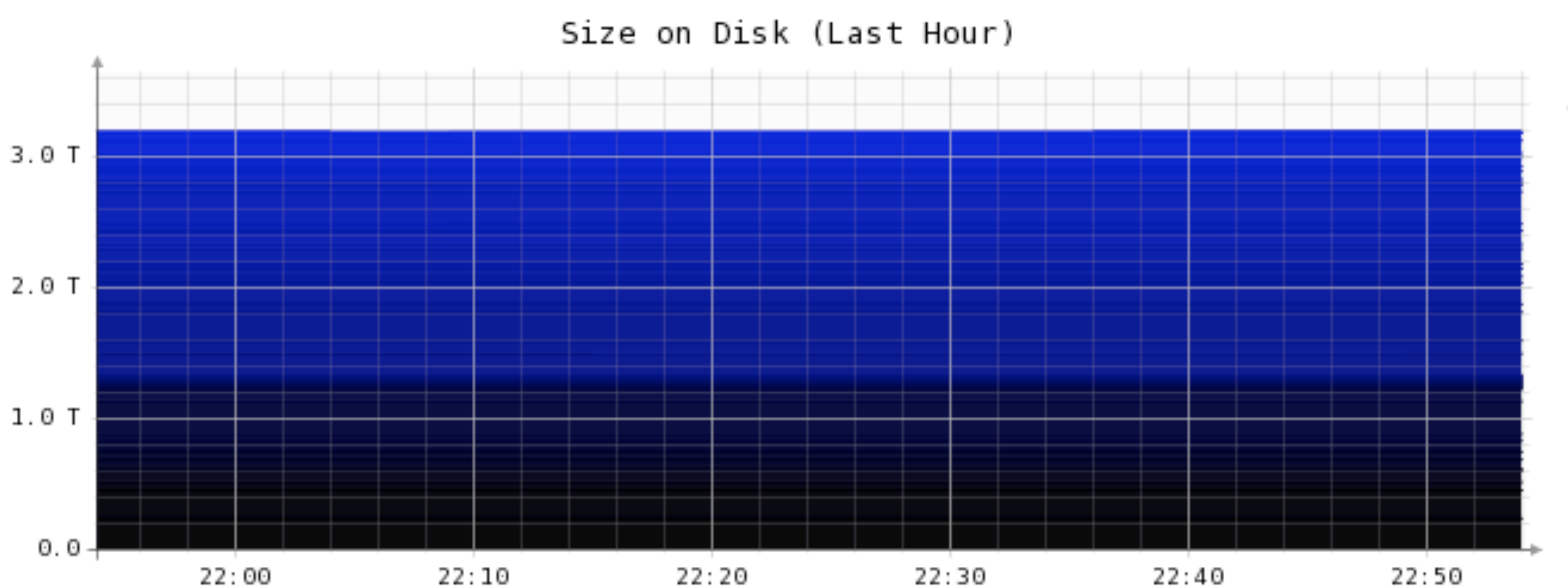
	Last	Maximum	Average	Minimum
Idle	86.24	87.02	84.84	81.62
Nice	0.00	0.00	0.00	0.00
System	2.37	3.44	2.60	2.13
Stolen	0.00	0.00	0.00	0.00
Application	7.46	9.14	7.64	6.56
Wait on IO	3.92	6.82	4.92	3.92

Writes per second (Last Hour)

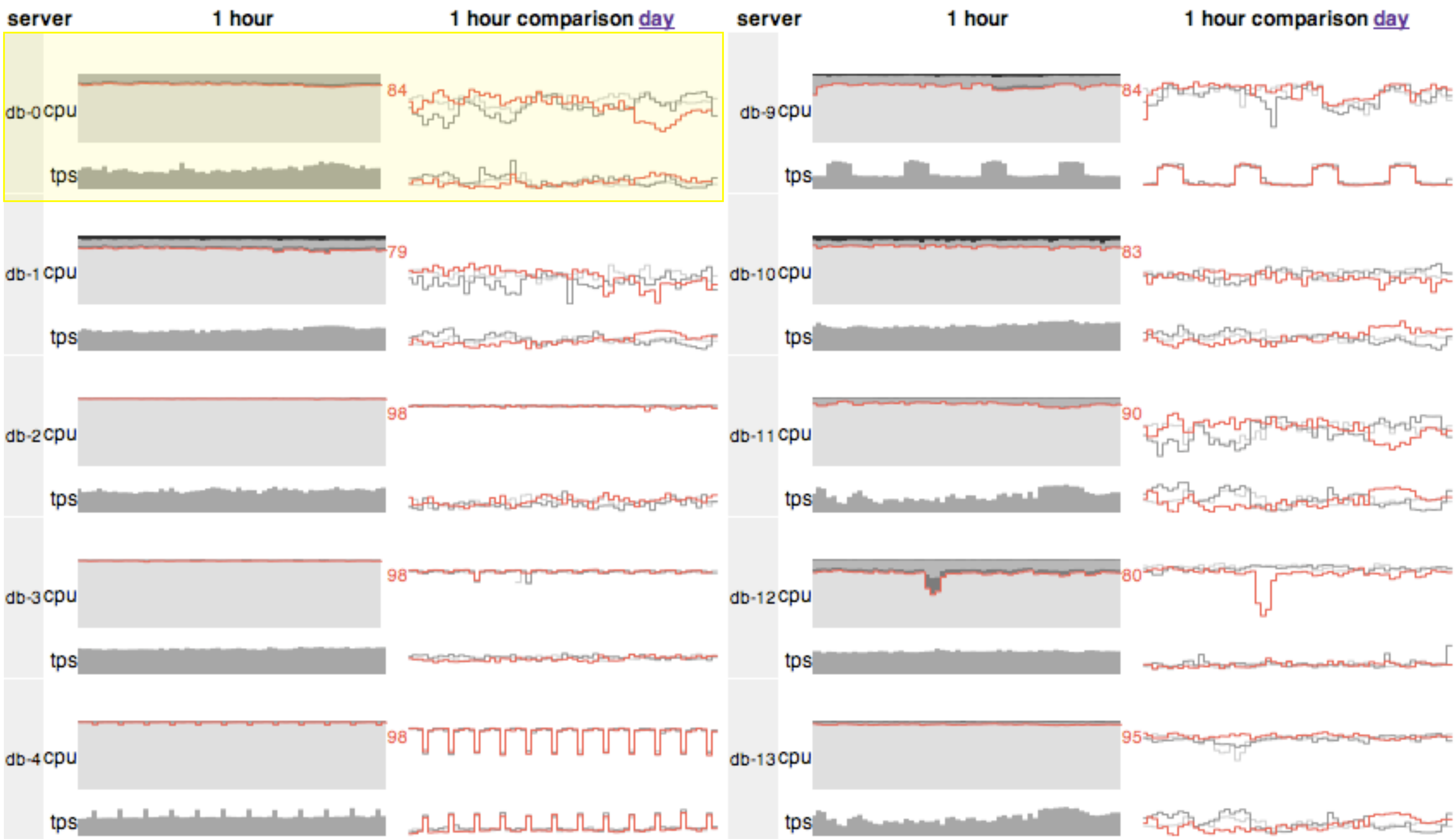


	Last	Maximum	Average	Minimum
Inserts	200	300	217	196
Updates	294	954	430	266
Deletes	37	5727	140	30


Transactions per Second (Last Hour)
30 k
20 k
10 k
0
22:00
22:10
22:20
22:30
22:40
22:50
RRDTOOL / TOBI OETIKER


Size on Disk (Last Hour)
3.0 T
2.0 T
1.0 T
0.0
22:00
22:10
22:20
22:30
22:40
22:50
RRDTOOL / TOBI OETIKER

overview



server
1 hour
1 hour comparison day
db-0 cpu
tps
84
db-1 cpu
tps
79
db-2 cpu
tps
98
db-3 cpu
tps
98
db-4 cpu
tps
98
server
1 hour
1 hour comparison day
db-9 cpu
tps
84
db-10 cpu
tps
83
db-11 cpu
tps
90
db-12 cpu
tps
80
db-13 cpu
tps
95

bloat

```
pg_database_size
pg_relation_size
pg_total_relation_size
pg_column_size
pg_size_pretty
```

bloat report

Treemap view

Table Sizes and Bloat (2009-05-19T08:55:01-0400)

Host	Database	Table	Total	Table Relation	Table Free	Table Bloat	Total Indexes Total	Total Indexes Ratio	Key Indexes Size	Key Indexes Free	Key Indexes Ratio	Nonkey Indexes Size	Nonkey Indexes Free	Nonkey Indexes Ratio	Toast Relation	Toast Index
		High bloat tables														
host_cc044	database_fd897	schema_4c918.table_26ec9	99 GB	50 GB	82 MB	0.002	49 GB	0.493	0 B	0 B	0.0	49 GB	16 GB	0.337		
host_cc044	database_fd897	schema_4c918.table_9c27a	32 GB	11 GB	26 kB	0.0	21 GB	0.659	3660 MB	392 MB	0.107	18 GB	3623 MB	0.202		
host_cc044	database_fd897	schema_4c918.table_edee8	22 GB	17 GB	440 MB	0.026	5956 MB	0.259	2316 MB	1591 MB	0.687	3639 MB	843 MB	0.232	0 B	40 kB
host_cc044	database_fd897	schema_4c918.table_b447a	20 GB	3637 MB	308 MB	0.085	17 GB	0.823	13 GB	12 GB	0.907	3326 MB	538 MB	0.162		
host_cc044	database_fd897	schema_4c918.table_5dbaa	16 GB	6214 MB	7760 B	0.0	10093 MB	0.619	0 B	0 B	0.0	10093 MB	1877 MB	0.186	0 B	24 kB
host_cc044	database_fd897	schema_4c918.table_64a1e	16 GB	3910 MB	609 MB	0.156	12 GB	0.755	588 MB	289 MB	0.492	11 GB	6621 MB	0.579		
host_cc044	database_fd897	schema_741f6.table_c839c	15 GB	8007 MB	3363 kB	0.0	7083 MB	0.467	0 B	0 B	0.0	7083 MB	1485 MB	0.21	382 MB	6480 kB
host_cc044	database_fd897	schema_4c918.table_c6aaa	12 GB	4879 MB	440 MB	0.09	7654 MB	0.611	639 MB	373 MB	0.584	7009 MB	3470 MB	0.495	0 B	120 kB
host_cc044	database_fd897	schema_4c918.table_c496c	12 GB	6129 MB	1257 MB	0.205	5896 MB	0.49	817 MB	306 MB	0.375	5080 MB	819 MB	0.161		
host_cc044	database_fd897	schema_9bc65.table_1a3e3	11 GB	3094 MB	637 MB	0.206	7917 MB	0.719	4399 MB	3292 MB	0.748	3518 MB	1119 MB	0.318		
host_cc044	database_fd897	schema_4c918.table_95fb5	10 GB	2765 MB	510 MB	0.184	7821 MB	0.739	693 MB	398 MB	0.575	7122 MB	4449 MB	0.625	0 B	72 kB
host_cc044	database_fd897	schema_4c918.table_7ebb0	7514 MB	1636 MB	127 MB	0.078	5878 MB	0.782	0 B	0 B	0.0	5869 MB	3631 MB	0.619	504 kB	336 kB
host_cc044	database_fd897	schema_4c918.table_26d77	6727 MB	3366 MB	13 MB	0.004	3361 MB	0.5	2576 MB	1161 MB	0.451	784 MB	107 MB	0.136		
host_cc044	database_fd897	schema_4c918.table_4c650	6099 MB	3824 MB	2354 MB	0.616	2274 MB	0.373	2274 MB	1937 MB	0.852	0 B	0 B	0.0		
host_cc044	database_fd897	schema_4c918.table_45d01	4237 MB	1190 MB	556 MB	0.467	3047 MB	0.719	430 MB	240 MB	0.558	2617 MB	1400 MB	0.535		
host_cc044	database_fd897	schema_4c918.table_0cecd	3084 MB	1270 MB	1026 MB	0.808	1814 MB	0.588	423 MB	381 MB	0.902	1392 MB	1059 MB	0.761		
host_cc044	database_fd897	schema_4c918.table_cf33d	2218 MB	609 MB	538 MB	0.884	1609 MB	0.725	133 MB	131 MB	0.99	1476 MB	1464 MB	0.992	0 B	88 kB
host_cc044	database_fd897	schema_4c918.table_e6247	1662 MB	437 MB	370 MB	0.848	1226 MB	0.737	57 MB	57 MB	0.999	1169 MB	1169 MB	1.0	0 B	48 kB

key indexes

key constraints

➤ unique indexes

cached plans (≤8.2)

reindex

host	table	reindexed on	total reclaimed	% saved
server-1	schema_a.table_a	2009-10-15	1.36 GB	29.94%
server-1	schema_a.table_b	2009-10-15	736.99 MB	5.02%
server-2	schema_b.table_c	2009-10-15	1.81 GB	24.97%
server-3	schema_c.table_d	2009-10-14	688.85 MB	16.89%
server-4	schema_d.table_e	2009-10-20	7.58 GB	22.71%
server-5	schema_e.table_f	2009-10-20	1.27 GB	51.20%
server-6	schema_f.table_g	2009-10-15	4.80 GB	25.02%
server-7	schema_g.table_h	2009-10-19	13.13 GB	89.18%

DTrace
SystemTap

logs

```
log_min_duration_statement
log_duration
log_lock_waits
deadlock_timeout
log_temp_files
log_connections
log_disconnections
```

```
log_statement_stats
log_parser_stats
log_planner_stats
log_executor_stats
```

```
LOG:  EXECUTOR STATISTICS
DETAIL:  ! system usage stats:
    !   0.017621 elapsed 0.004762 user 0.000816 system sec
    !   [6.012501 user 0.336354 sys total]
    !   0/0 [0/0] filesystem blocks in/out
    !   0/0 [0/0] page faults/reclaims, 0 [0] swaps
    !   0 [1] signals rcvd, 0/10 [4/14944] messages rcvd/sent
    !   2/0 [210/0] voluntary/involuntary context switches
    ! buffer usage stats:
    !   Shared blocks:          9 read,          0 written, buffer hit rate = 0.00%
    !   Local  blocks:          0 read,          0 written, buffer hit rate = 0.00%
    !   Direct blocks:          0 read,          0 written
STATEMENT:  select * from posuta.index_statistics limit 1000;
LOG:  duration: 42.422 ms
```

18.7. Error Reporting and Logging
18.8. Run-Time Statistics

CSV

```
2009-05-19 10:25:35.470 EDT,"grzm","posuta_production",99595,"[local]",
4a12c078.1850b,28,"SELECT",2009-05-19 10:21:44 EDT,2/30525,0,LOG,00000,"EXECUTOR
STATISTICS","! system usage stats:
!    1.786288 elapsed 0.065964 user 0.074493 system sec
!    [6.079580 user 0.412469 sys total]
!    2/0 [2/0] filesystem blocks in/out
!    0/0 [0/0] page faults/reclaims, 0 [0] swaps
!    0 [1] signals rcvd, 0/13 [5/14960] messages rcvd/sent
!    1008/0 [1230/0] voluntary/involuntary context switches
! buffer usage stats:
!    Shared blocks:       1073 read,          0 written, buffer hit rate = 0.00%
!    Local  blocks:          0 read,          0 written, buffer hit rate = 0.00%
!    Direct blocks:          0 read,          0 written",,,,,"select * from
posuta.index_statistics where index_id = 265 limit 1000;",,
```

contrib/

pg_freespacemap

pg_buffercache

pgrowlocks

pgstattuple

statistics collector

Chapter 26. Monitoring Database Activity track_activities

track_activity_query_size*

track_counts

track_functions*

stats_temp_directory*

pg_stat_activity

pg_locks

```
pg_stat_get_numscans

pg_stat_get_tuples_returned

pg_stat_get_tuples_fetched

pg_stat_get_tuples_inserted

pg_stat_get_tuples_updated

pg_stat_get_tuples_hot_updated

pg_stat_get_tuples_deleted
```

returned: table—rows fetched by bitmap scans
index—table rows fetched by simple index scans using the index
fetched: table—rows read by sequential scans
index—number of index entries returned

pg_stat_get_live_tuples

pg_stat_get_dead_tuples

pg_stat_get_blocks_fetched

pg_stat_get_blocks_hit

```
pg_stat_get_last_vacuum_time

pg_stat_get_last_autovacuum_time

pg_stat_get_last_analyze_time

pg_stat_get_last_autoanalyze_time
```

```
pg_stat_get_function_calls*

pg_stat_get_function_time*

pg_stat_get_function_self_time*
```

```
pg_stat_get_db_xact_commit

pg_stat_get_db_xact_rollback
```

pg_stat_get_bgwriter_timed_checkpoints

pg_stat_get_bgwriter_requested_checkpoints

pg_stat_get_bgwriter_buf_written_checkpoints

pg_stat_get_bgwriter_buf_written_clean

pg_stat_get_bgwriter_maxwritten_clean

snapshot

topHeapHitters

topIndexHitters

```
pgStatIO Heavy Index Hitters - 10/20/09 20:02:06 - Interval: 5s
------------------------------------------------------------------------------------
Table                                              Last Difference         Total Hits
------------------------------------------------------------------------------------
1.  schema_a.table_a                                        146502       710013840879
2.  schema_b.table_b                                         92171        55473259607
3.  schema_a.table_c                                         38684       106950242690
4.  schema_a.table_d                                         32797       110541228095
5.  schema_a.table_e                                         25096        43939803663
6.  schema_a.table_f                                         20940        75501250234
7.  schema_a.table_g                                         10982       126558207896
8.  schema_a.table_h                                          9337       147866373045
```

0300

Those who cannot remember the past are condemned to repeat it.

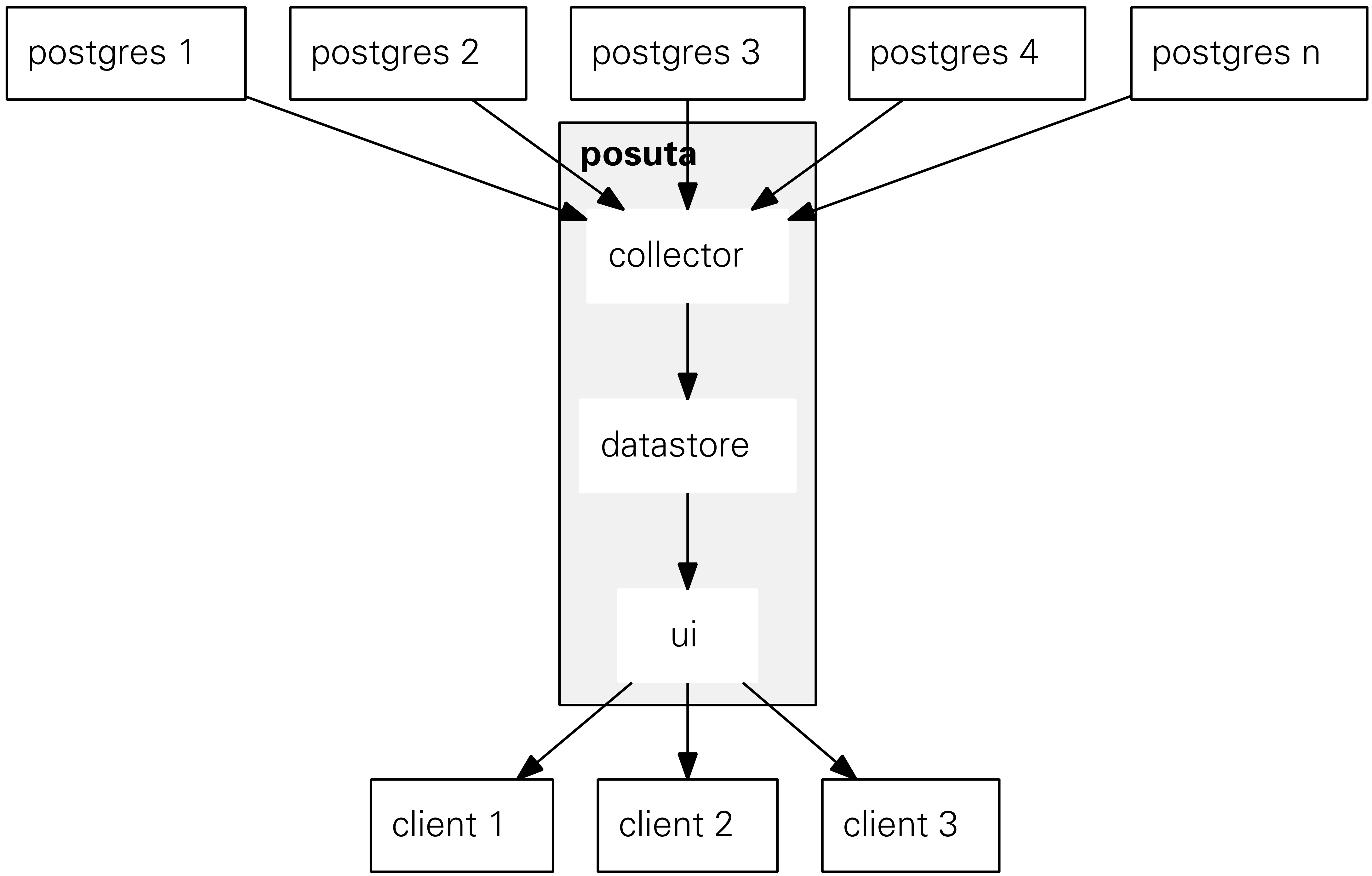
— George Santayana

posuta

Postgres **sta**tistics

ˈposu̥ta

ポスタ


postgres 1
postgres 2
postgres 3
postgres 4
postgres n
posuta
collector
datastore
ui
client 1
client 2
client 3

Future

partition datastore

dashboard/ui enhancements

bloat

multiple collectors

trending/anomaly detection

Posuta
http://code.google.com/p/posuta/

Postgres
http://postgresql.org

Clojure
http://clojure.org

Compojure
http://github.com/weavejester/compojure

jQuery
http://jquery.com

flot
http://code.google.com/p/flot