Researching PostgreSQL Performance

Fernando Ike de Oliveira

PGCon 2008

Researching PostgreSQL Performance

http://www.midstorm.org/~fike/researching_postgresql.pdf

Questions...

- "What version do we use? 8.2 or 8.3? (2007-12)"
- PostgreSQL 8.3 was a beta version
- PostgreSQL 8.3 was promising to be faster than PostgreSQL 8.2
- PostgreSQL 8.2 doesn't scale well in a high number of transactions

Hardware

- 2 Server DELL PowerEdge 6850, 16GB RAM, 4 dual-core processors
- Storage DELL Clarion of 1TB
- 3 RAID 5 (data, index and wal)
- Operation System: Debian Etch 4.0 AMD64.

Tests

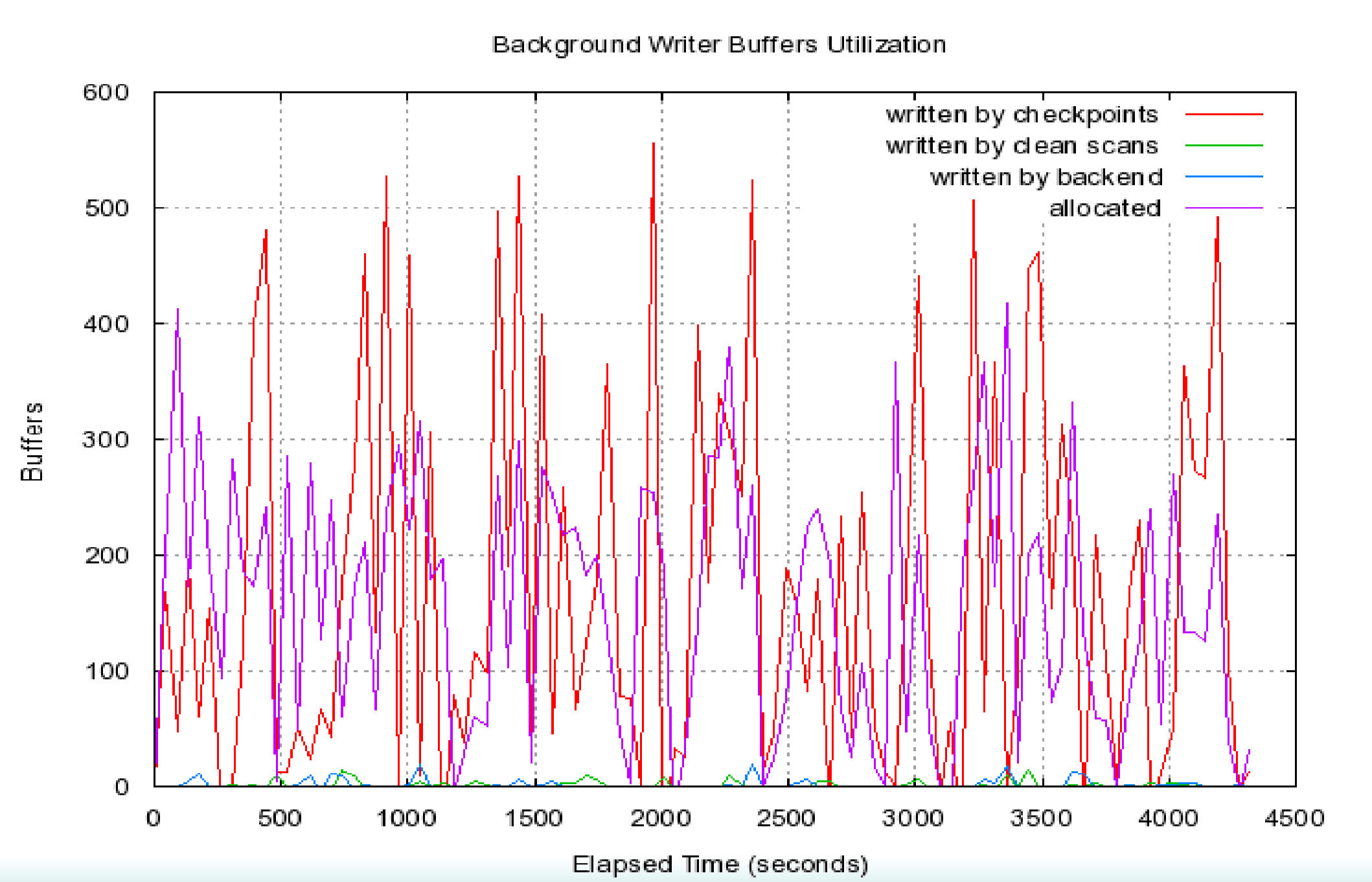
- fixed connections = 100
- scaling factor = 100
- transactions = 100, 100, 1000000

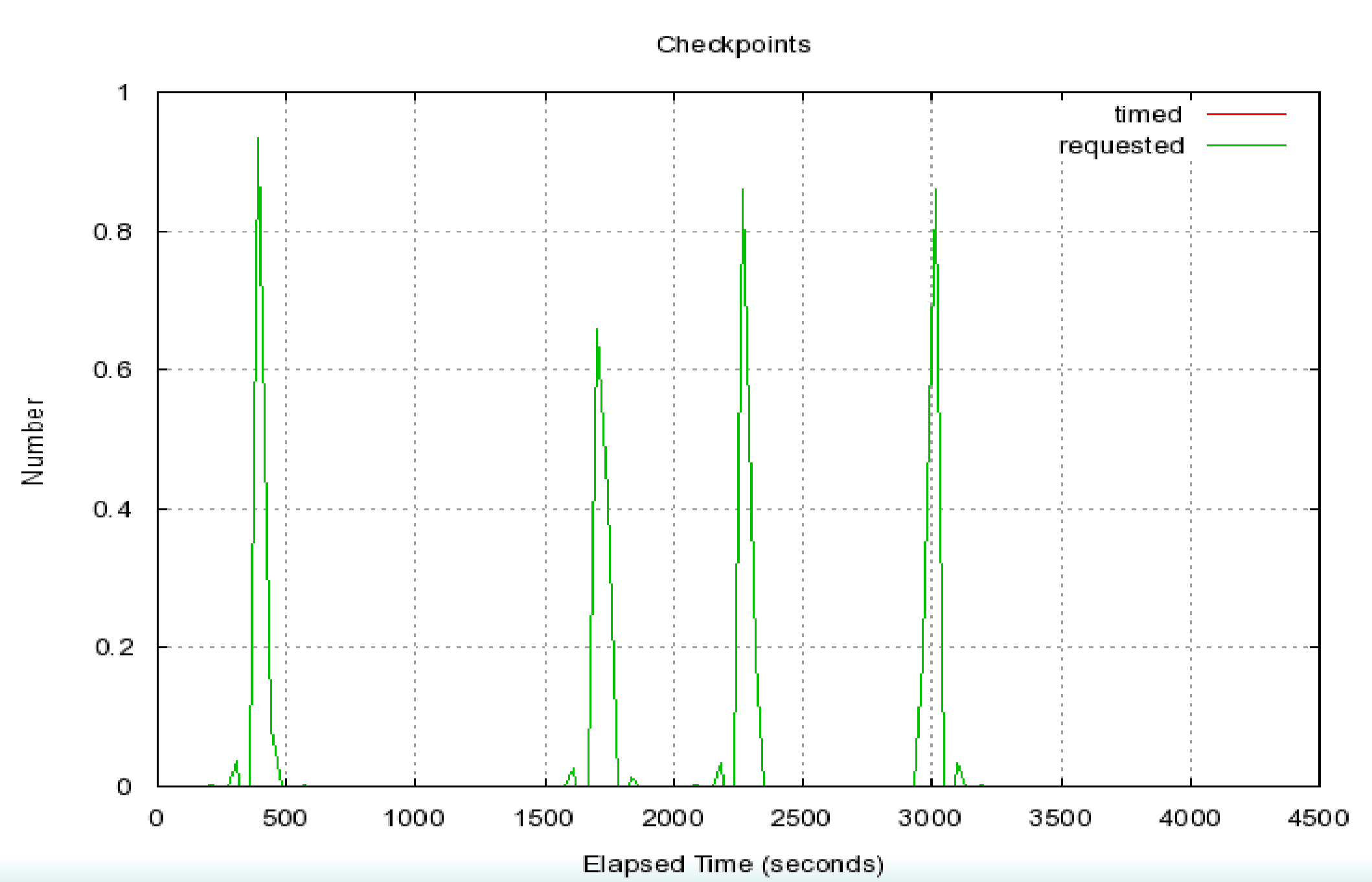
Tests

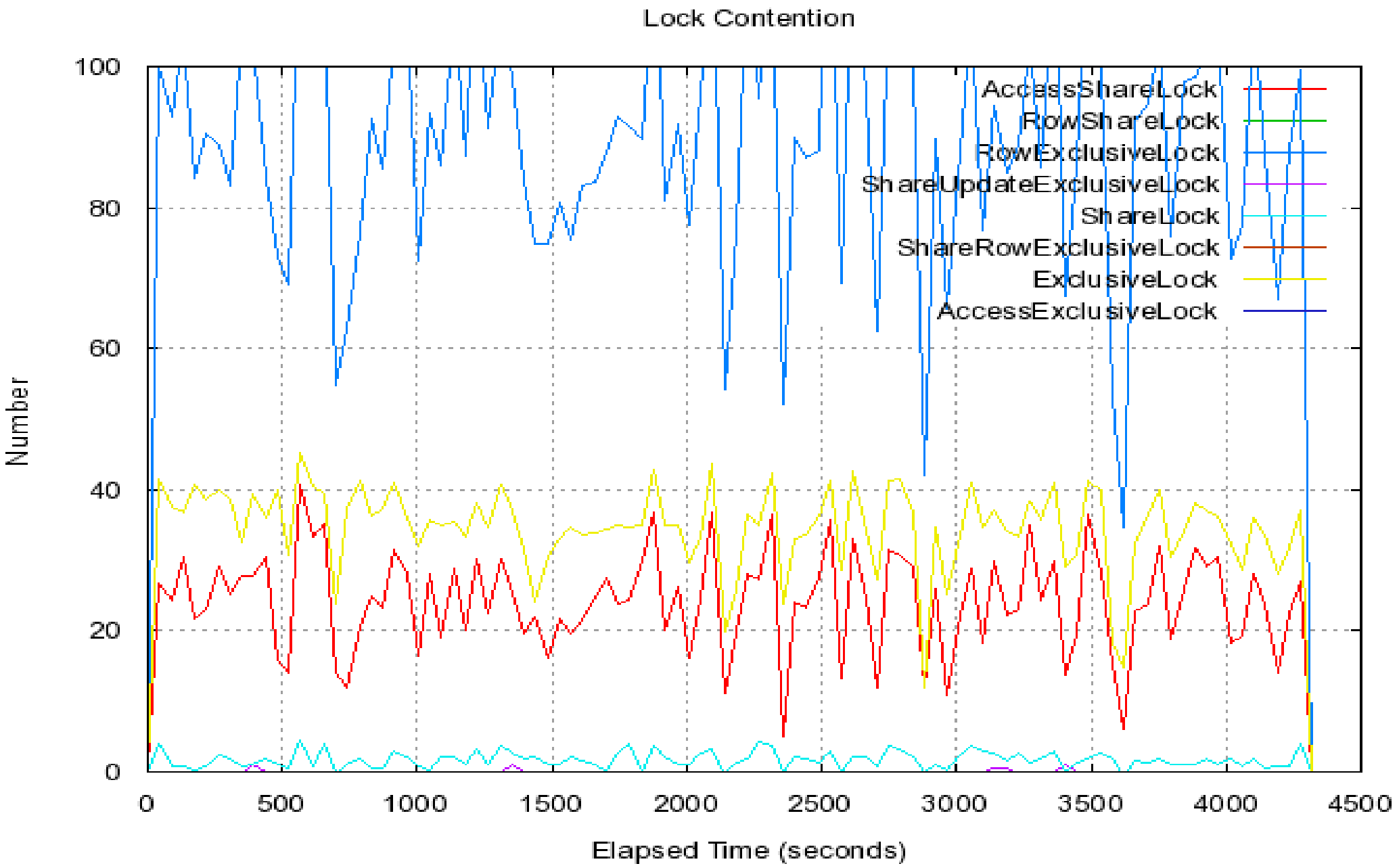
- ~ 340 tests
- initial test with pgbench
- DBT-2 and pgbench don't attended expected
- Euler developed/developing pgtesttool

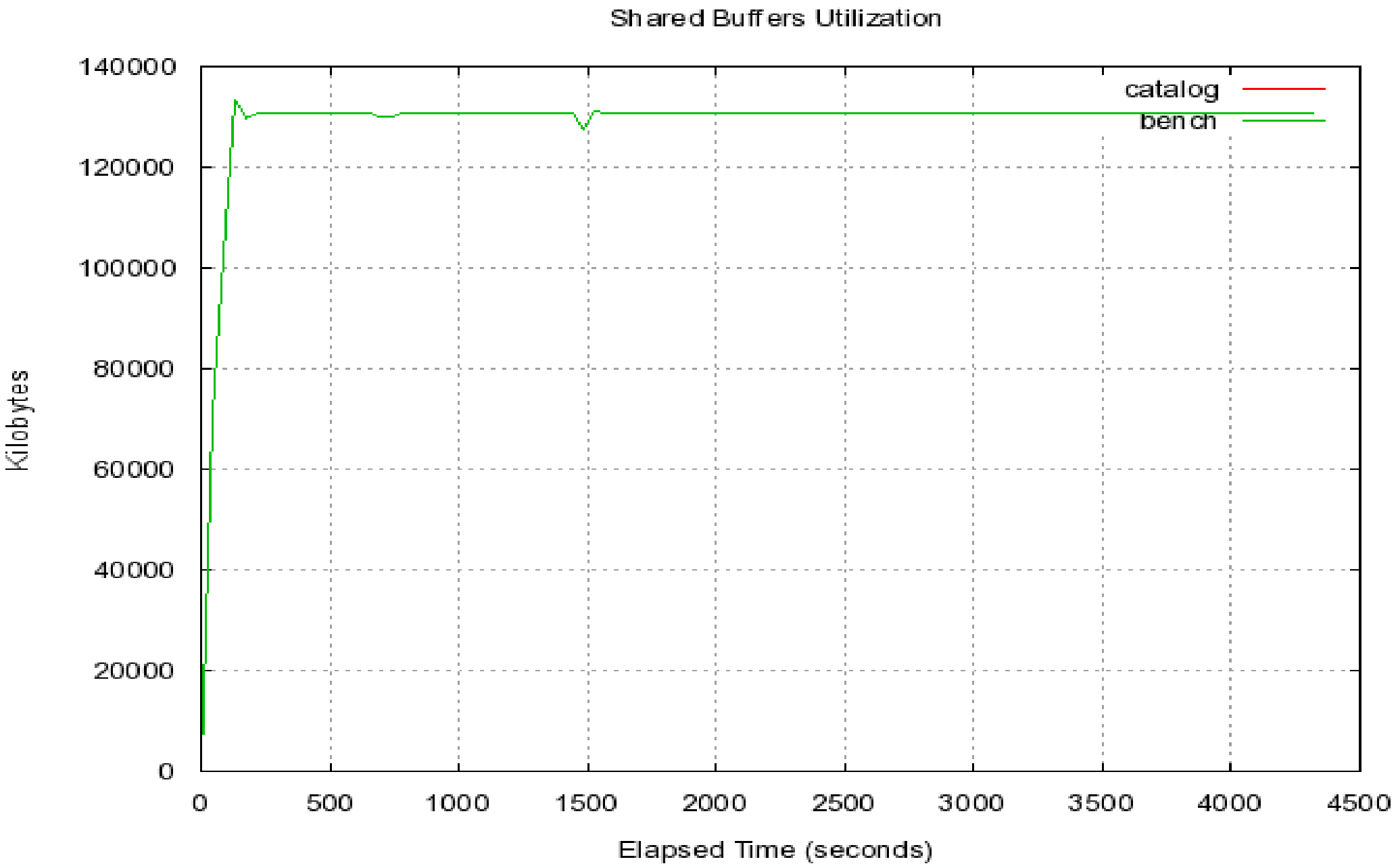
pgtesttool

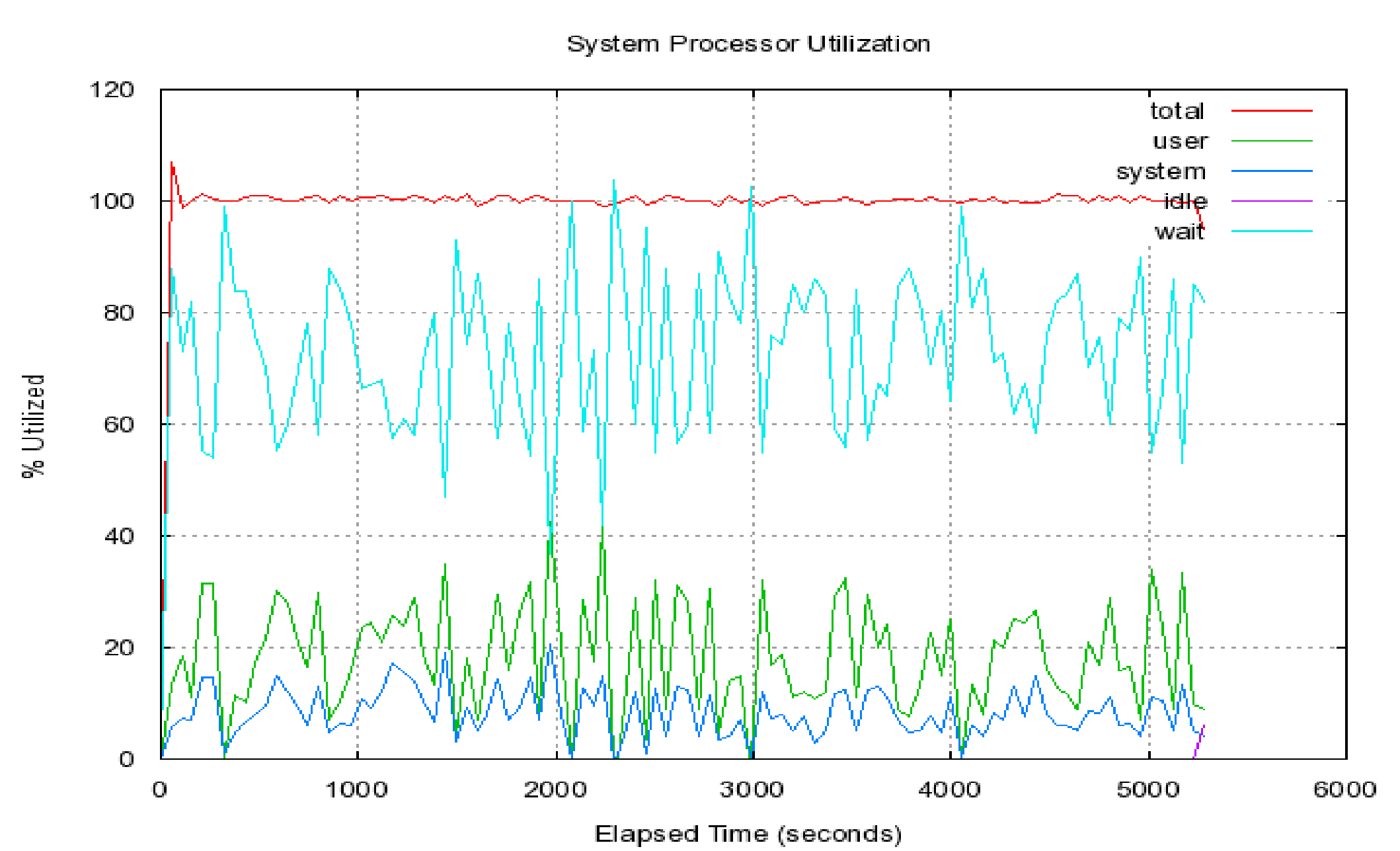
- automatic install postgresql and pgbench
- automatic data collection for server during test
- local and remote pgbench execution
- many postgresql.conf tests in one round
- generation of graphic results (OS statistics, bwgriter, checkpoints, database blocks, lock contention, pgbench results, shared buffers)











OS Tuning

```
echo "8589934592" > /proc/sys/kernel/shmmax
echo "deadline" > /sys/block/sdX/queue/scheduler
echo "250 128000 32 256" > /proc/sys/kernel/sem
echo "2" > /proc/sys/vm/overcommit_memory
echo "16777216" > /proc/sys/net/core/rmem_default
echo "16777216" > /proc/sys/net/core/wmem_default
echo "16777216" > /proc/sys/net/core/wmem_max
echo "16777216" > /proc/sys/net/core/rmem_max
```

/etc/security/limits.conf

```
postgres        soft    nofile          63536
postgres        soft    nproc            2047
postgres        hard     nofile         63536
postgres        hard     nproc          16384
```

basic postgresql.conf

listen_addresses = '*'

max_connections = 110

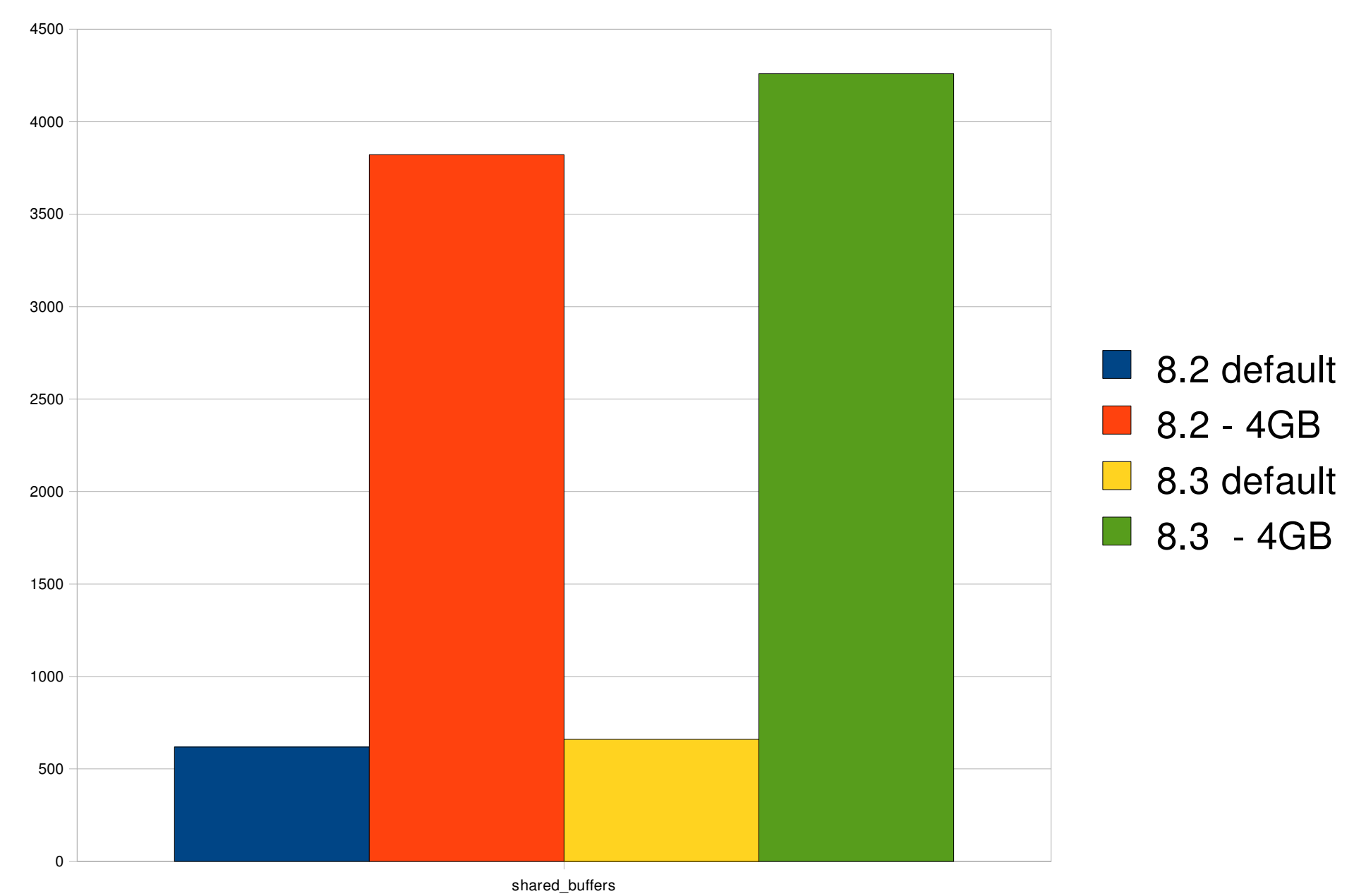
max_fsm_pages = 204800

effective_cache_size = 10GB

shared_buffers = 4GB (25%)

1 - **618.4320** tps, 8.2 w/ default conf
2 - **3,822.1502** tps, 8.2 w/ shared_buffers = 4GB
3 - **660.3667** tps, 8.3 w/ default conf
4 - **4,259.6078** tps, 8.3 w/ shared_buffers = 4GB, (full_page_write = off, wal_sync_method = open_sync).

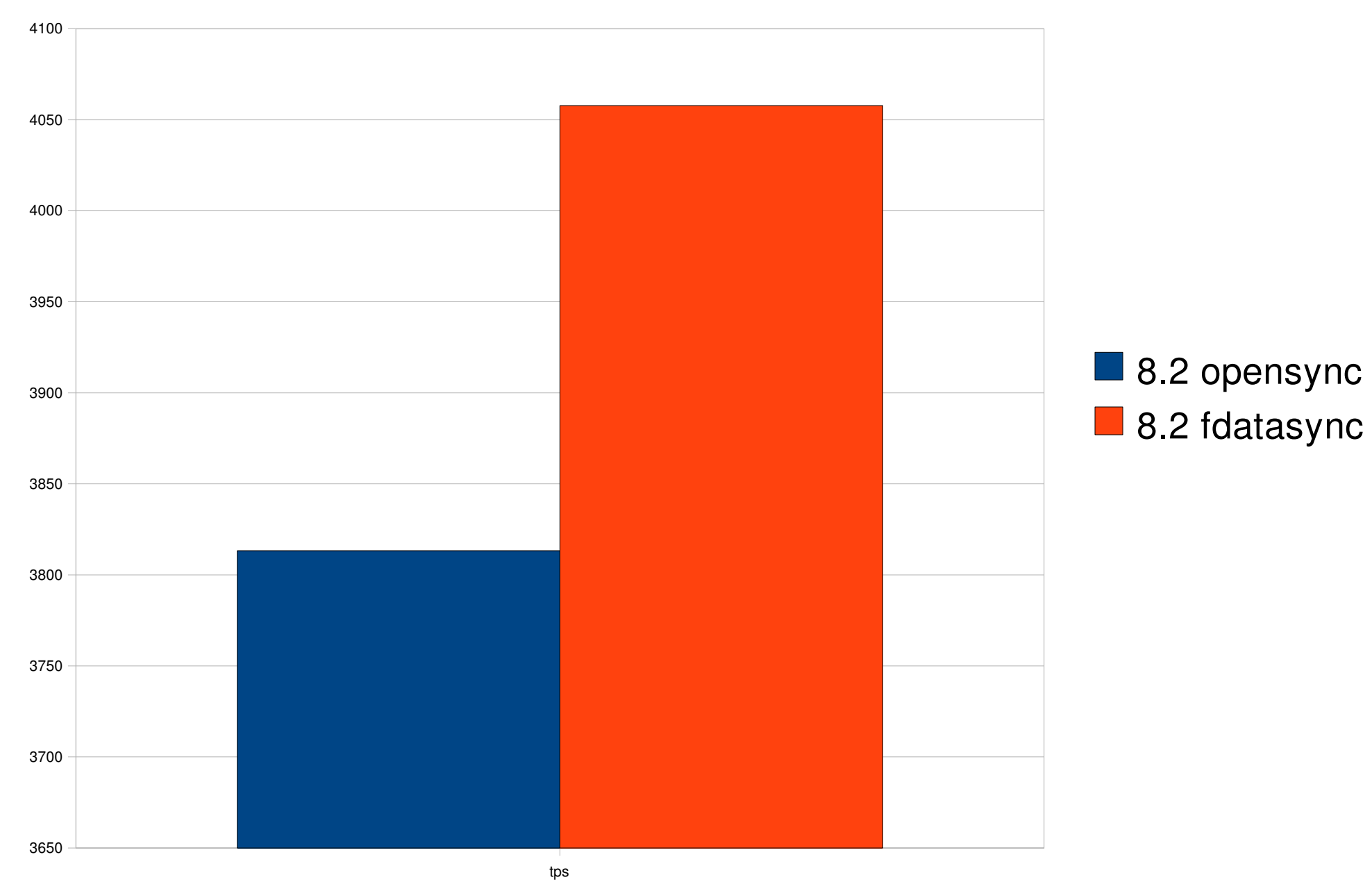
1 p/ 2 - **518,03%**
1 p/ 3 - **6.78%**
3 p/ 4 - **545.03%**


4500
4000
3500
3000
2500
2000
1500
1000
500
0
shared_buffers
8.2 default
8.2 - 4GB
8.3 default
8.3 - 4GB

fdatasync and opensync 100t

1 - **3813.2723** tps, 8.2 w/ opensync
2 - **4057.7641** tps, 8.3 w/ fdatasync

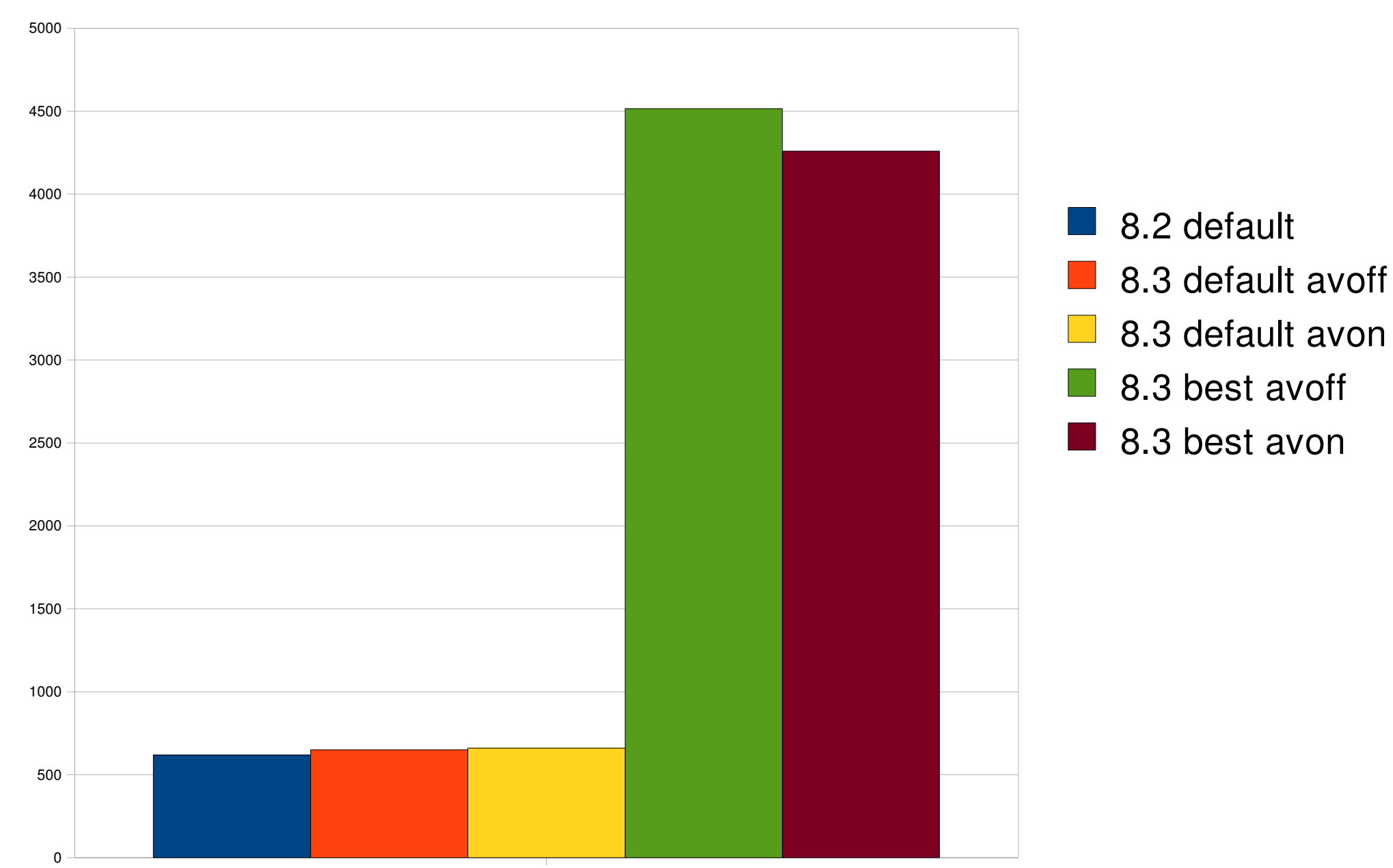
1 p/ 2 - **6,41**%



autovacuum

1 - **618.4320** tps, 8.2 default conf
2 - **650.4415** tps, 8.3 default conf, autovac off
3 - **660.3667** tps, 8.3 default conf, autovac on
4 - **4515.4117** tps, 8.3 better conf, (fullpgwrite off), autovac off
5 - **4259.6078** tps, 8.3 better conf, (fullpgwrite off), autovac on

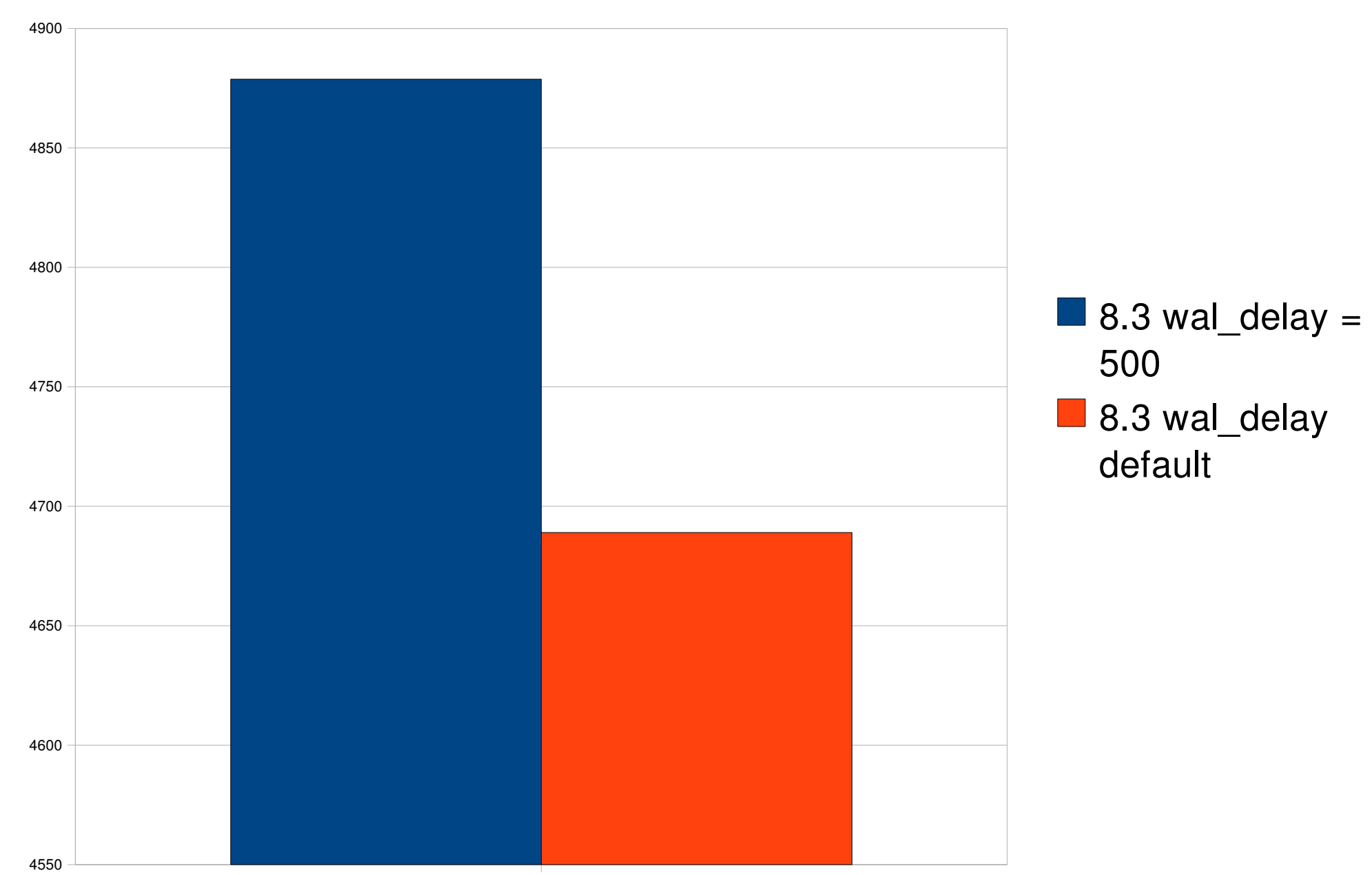
1 p/ 2 - **5,17**%
1 p/ 3 - **6,78**%
2 p/ 3 - **1,52**%
4 p/ 5 - **6,00**%


5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
8.2 default
8.3 default avoff
8.3 default avon
8.3 best avoff
8.3 best avon

wal_delay 1,000t

1 - **4878.7961** tps, 8.3 better conf, wal_delay = 500
2 - **4688.9773** tps, 8.3 better conf, default wal_delay

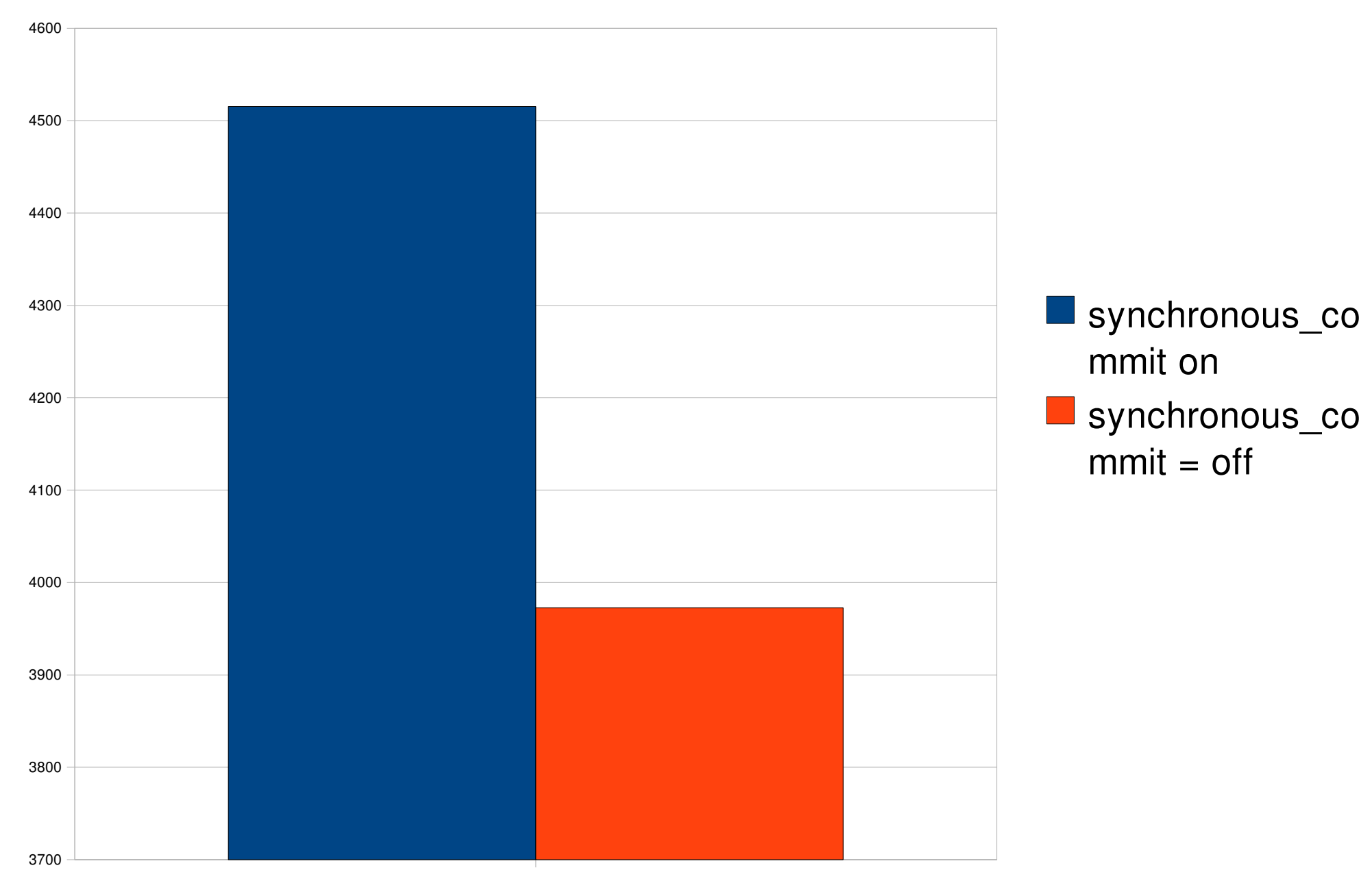
1 p/ 2 - **4.04**%


4900
4850
4800
4750
4700
4650
4600
4550
8.3 wal_delay = 500
8.3 wal_delay default

synchronous_commit 100t

1 - **4515.4117** tps, 8.3 synchronous_commit = on
2 - **3972.6742** tps, 8.3 synchronous_commit = off

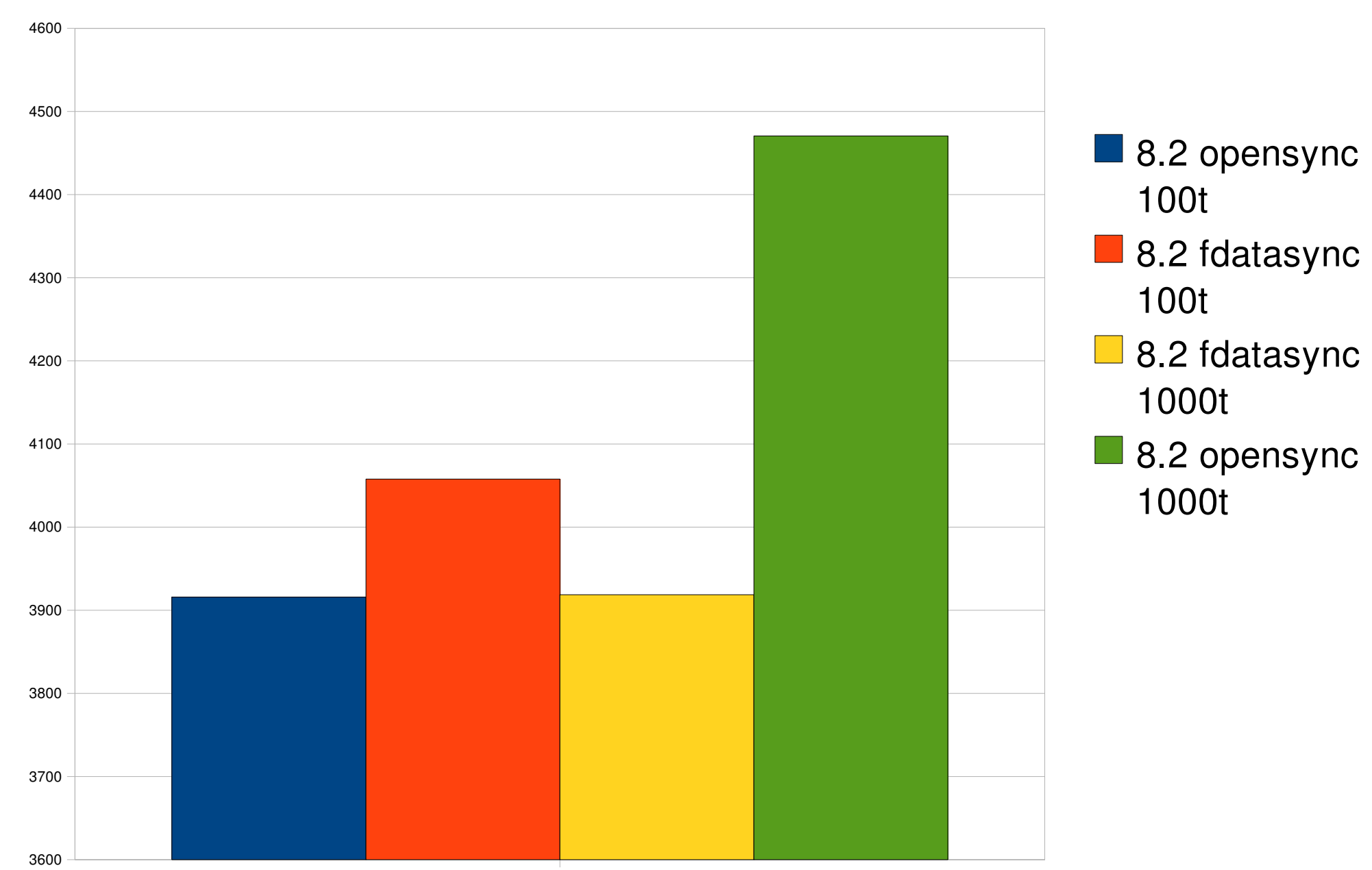
[1] p/ [2] - **13,66**%


4600
4500
4400
4300
4200
4100
4000
3900
3800
3700
synchronous_commit on
synchronous_commit = off

wal_method test

1 - **3915.7527** tps 8.2 w/ opensync - 100t
2 - **4057.7641** tps 8.2 w/ fdatasync - 100t
3 - **3918.6068** tps 8.2 w/ fdatasync - 1,000t
4 - **4470.5250** tps 8.2 w/ opensync - 1,000t

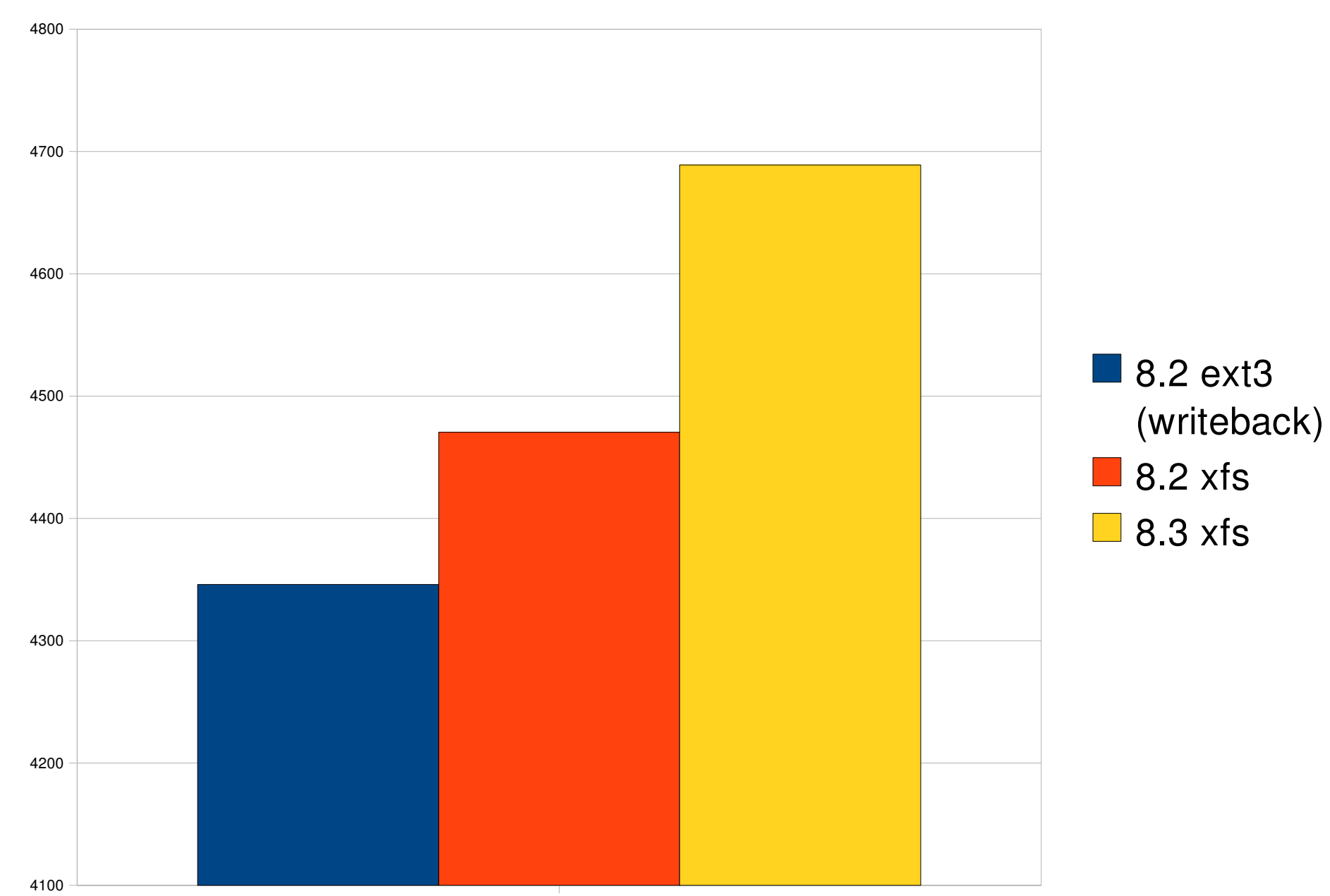
[1] p/ [2] - **3,62**%


4600
4500
4400
4300
4200
4100
4000
3900
3800
3700
3600
8.2 opensync 100t
8.2 fdatasync 100t
8.2 fdatasync 1000t
8.2 opensync 1000t

Comparison 8.2 X 8.3 and Filesystem

1 - **4345.9916** tps 8.2 w/ opensync, ext3 (writeback) - 1,000t
2 - **4470.5250** tps 8.2 w/ opensync, xfs - 1,000t
3 - **4688.9773** tps 8.3 w/ opensync, xfs - 1,000t

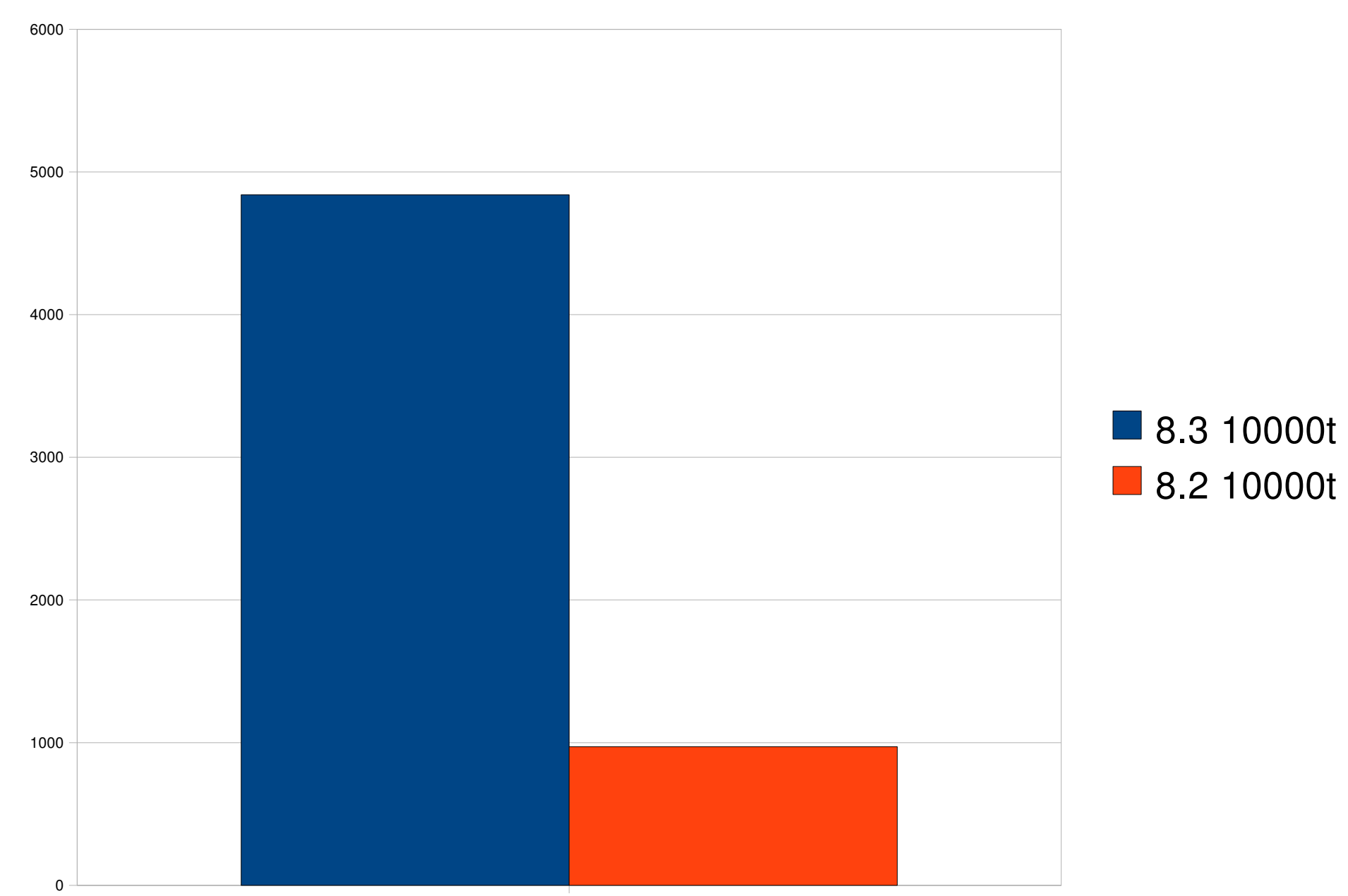
1 p/ 2 - **2,86**%
2 p/ 3 - **4,88**%


4800
4700
4600
4500
4400
4300
4200
4100
8.2 ext3 (writeback)
8.2 xfs
8.3 xfs

8.2 X 8.3 100000t

1 - **4839.7800** tps 8.3 100000t
2 - **971.7439** tps 8.2 100000t

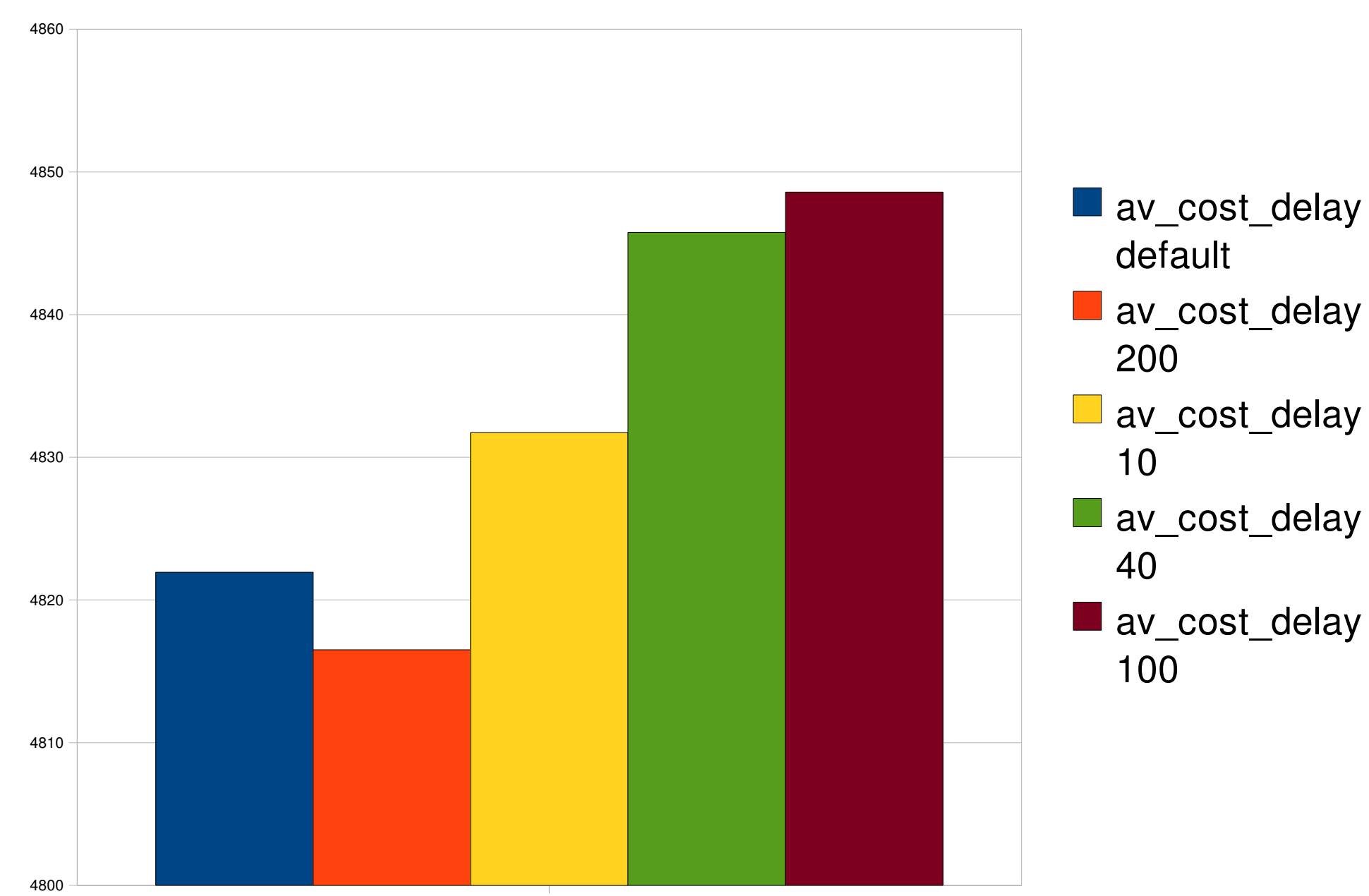
1 p/ 2 **-398**%


6000
5000
4000
3000
2000
1000
0
8.3 10000t
8.2 10000t

autovacuum delay

1 - **4821.9402** tps, 8.3 w/ default autovacuum_cost_delay
2 - **4816.5119** tps, 8.3 w/ autovacuum_cost_delay = 200
3 - **4831.7244** tps, 8.3 w/ autovacuum_cost_delay = 10
4 - **4845.7629** tps, 8.3 w/ autovacuum_cost_delay = 40
5 - **4848.5798** tps, 8.3 w/ autovacuum_cost_delay = 100

1 p/ 2 - **-0.11**%
1 p/ 3 - **0.20**%
1 p/ 4 - **0.49**%
1 p/ 5 - **0.55**%


4860
4850
4840
4830
4820
4810
4800
av_cost_delay default
av_cost_delay 200
av_cost_delay 10
av_cost_delay 40
av_cost_delay 100

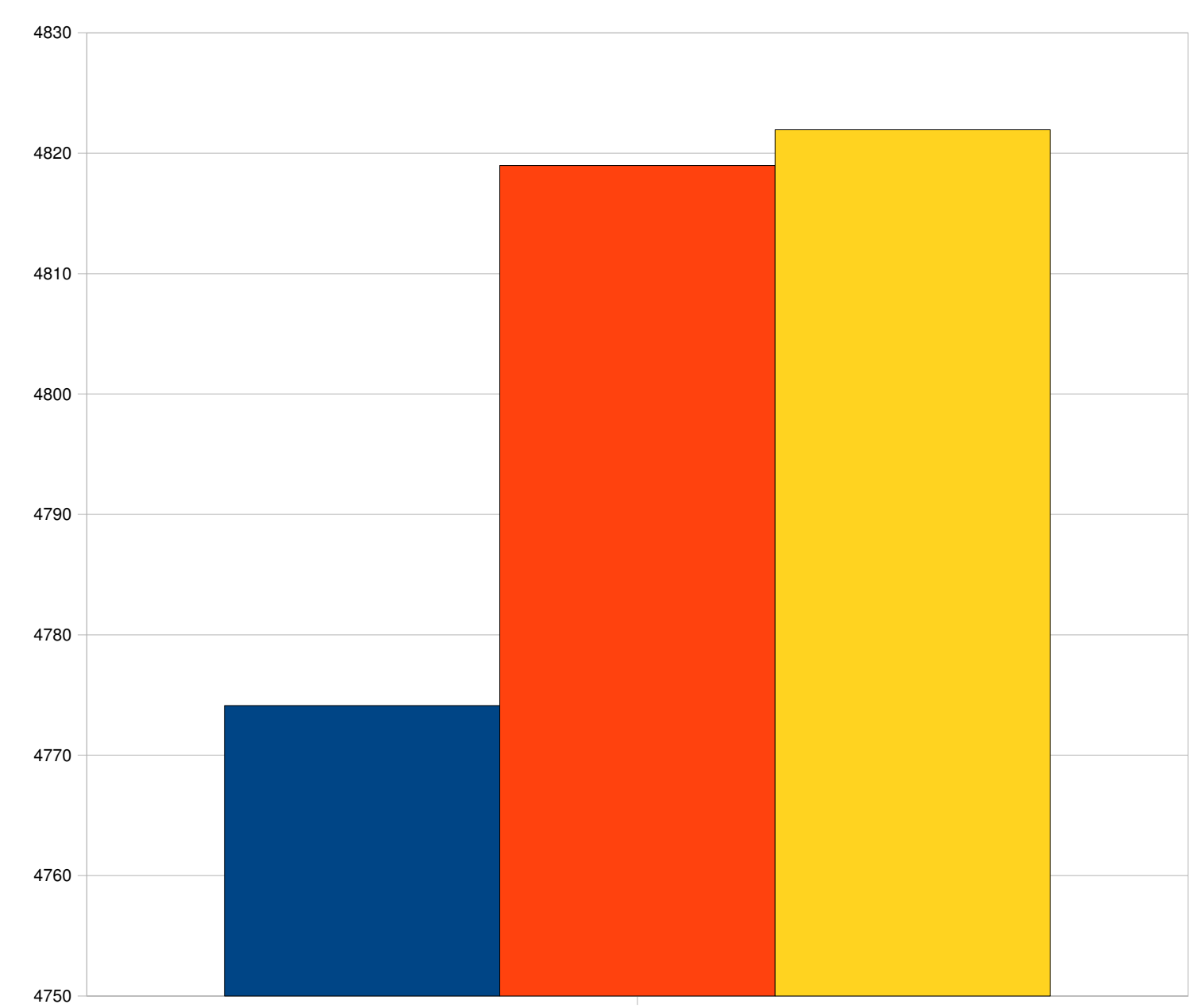
Checkpoints 1000t

1 - **4774.1278** tps, 8.3 w/ checkpoints_segments = 256, checkpoint_complementation_target = 05

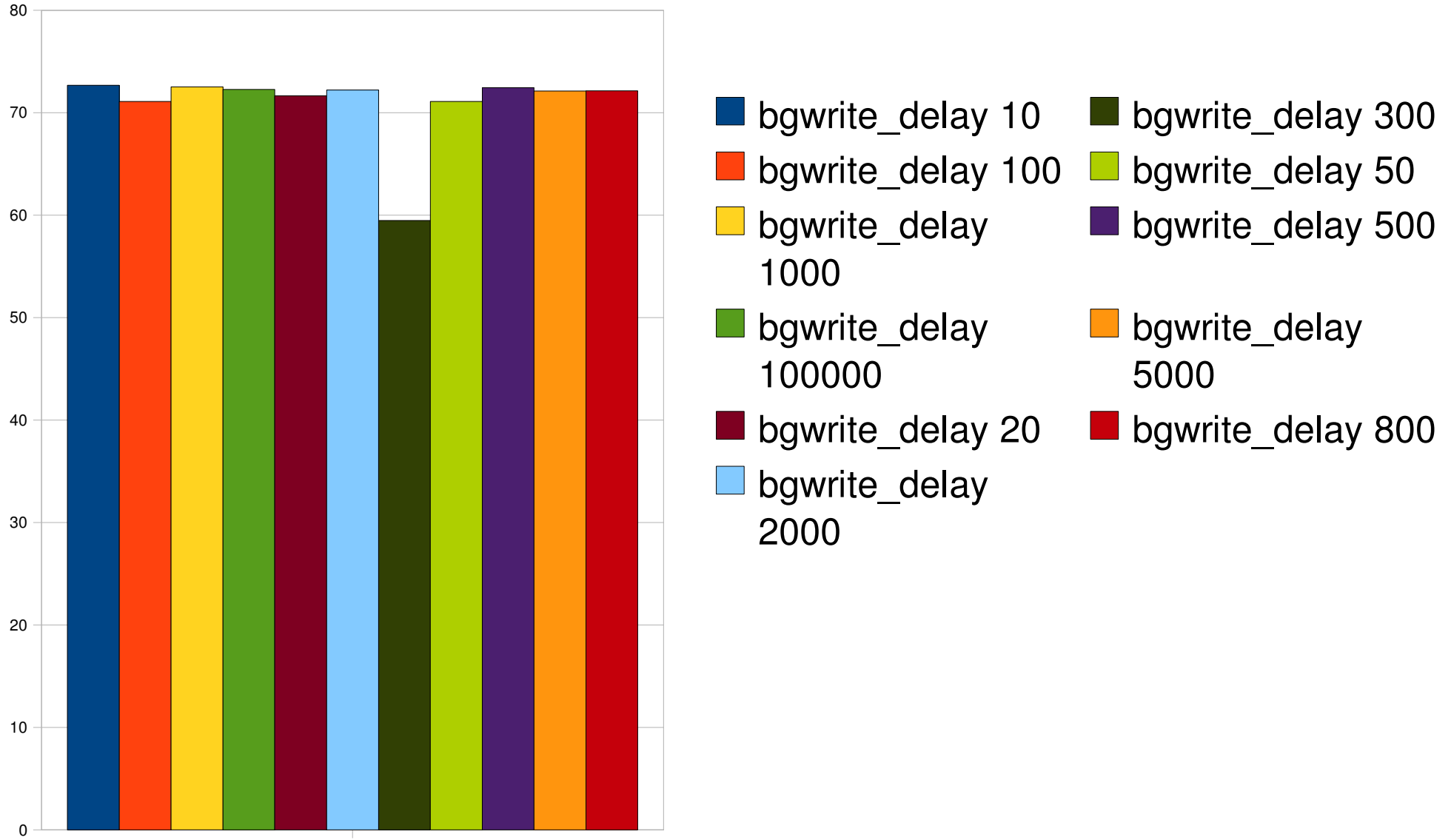
2 - **4818.9832** tps, 8.3 w/ checkpoints_segments = 128, checkpoint_complementation_target = 07

3 - **4821.9402** tps, 8.3 w/ checkpoints_segments = 256, checkpoint_complementation_target = 07

1 p/ 2 - **0.93**%
1 p/ 3 - **1.00**%


4830
4820
4810
4800
4790
4780
4770
4760
4750
Segments 256, target 5
Segments 128, target 7
Segments 256, target 7

Bgwriter, isolated test (s= 30; t/c = 2000; c = 20)



Data not registered but known by us

- magic numbers to max_connections
- 700 or more?
- 1500 connections decrease 3%
- 3000 connections decrease 5%

- opensync is better with pgbench transaction

- fdatasync is better with load tables of the pgbench

- full_page_write = off is more than faster

- deadline I/O scheduler > CFQ

Conclusions of the tests.

- Ext3 (writeback) - better in few transactions
- XFS - better in many transactions
- synchronous_commit disabled decrease performance
- to turn off autovacuum in the PostgreSQL 8.3 is a bad idea
- PostgreSQL 8.3 has better performance compared to PostgreSQL 8.2.

Still to be tested

- Filesystems: Ext4, JFS, Reiserfs
- Scheduling I/O: Anticipatory, Deadline, CFQ, Noop
- Linux Kernel 2.6.23 >= Completely fair scheduler
- OS: FreeBSD, OpenSolaris...
- Tests with TPC-C, TPC-H, TPC-E
- RAID 10 X RAID 5

Contacts and references

Euler Taveira de Oliveira - euler@timbira.com

Fernando Ike de Oliveira - fike@4linux.com.br
fike@midstorm.org

Result tests: - http://www.inf.ufrgs.br/~etoliveira/pg/resultados/

pgtesttool:

PGCon Brasil 2008:
http://pgcon.postgresql.org.br