Tuning PostgreSQL

Shoaib Mir
Fujitsu Australia Software Technology
LCA 2011

Agenda

- Procedures for diagnosing performance problems
- Benchmarking the database server
- Monitoring

Identifying Areas

- Application analysis
- SQL
- Memory
- Storage
- File System
- PostgreSQL configuration file (postgresql.conf)

Application Analysis

- What type of IO patterns (reads or writes)
  - Reading large amounts of data?
  - Writing large amounts of data?
- Bulk loading jobs
- Analytical queries

SQL

- Use query analysis tools on database server log files to find long running queries (EPQA, pgfouine)
- Use EXPLAIN ANALYZE to debug
- Avoid EXPLAIN ANALYZE with DML style queries in production environment

Memory

- Size of the database
  - Small to fit in memory
  - DW like DB (get faster disks)
- Buffer cache usage (pg_buffercache)
- Query plans (using disk?)

Storage

- What type of storage setup? Direct attached storage or SAN
- RAID setup
  - RAID 10 for write-heavy activity
  - RAID 5 problems with parity information
- iostat output and looking for waits and queue sizes
- RAID controller settings
  - Write cache (write-back)
  - Battery backed
  - Monitor battery health

Storage

- Tablespaces can help with distributing data
  - Start with indexes and tables on separate
- Transaction logs should be on a separate storage area

File System

- What type of filesystem?
  - XFS recommended
  - Better journaling then ext3 (only meta-data)
  - Use nobarrier with xfs when using battery backed write controller

PostgreSQL configuration file (postgresql.conf)

- shared_buffers
- effective_cache_size
- work_mem
- maintenance_work_mem
- Autovacuum
- default_statistics_target
- checkpoint_segments

shared_buffers

- Value for database buffer cache
- Allocated on database server start
- Start with 25% of the available RAM
- Use pg_buffercache to help you find the optimal value

effective_cache_size

- Will not allocate memory on database start
- Used during picking an optimal query plan (better index usage)
- ~75 percent of available memory (look for free and cached numbers)

work_mem

- Used for sorting operations
- Avoid disk sort (look for EXPLAIN ANALYZE output)
- Not easy to find an optimal work_mem setting at server level, for queries looking at large amounts of data use custom values per session

maintenance_work_mem

- Used by DDL operations
- Value depends on the size of tables
- Can be done for a specific session

Autovacuum

- Automatically vacuums and analyzes tables when analyze/vacuum thresholds are met
- Thresholds can be tuned by looking at pg_stat_user_tables
- Worker threads introduced in 8.3

default_statistic_target

- Value for gathering sampling statistics for a table when ANALYZE is done
- Default of 10 is 8.3 is low which should be set to 100, 8.4 already defaults to 100
- Custom statistics settings for each table can be set using ALTER TABLE

checkpoint_segments

- Number of files in pg_xlog (each of 16MB in size
  - 2 * checkpoint_segments + 1
  - In 8.4 and above using checkpoint_completion_target
- Tune this for large number of writes

Benchmarking

- Becnhmark IO performance using 'dd' and Bonnie++
- pgBench for database server
- Do the benchmarks when setting up a new database server and keep on doing them oftenly to find bottlenecks

Monitoring

- iostat, dstat, top
- check_postgres.pl plugin for database health monitoring
  - Backends
  - Bloat
  - Vacuum/Analyze activity
  - Checkpoints
  - Free space memory
  - Disk space
- A combination of Nagios and tools like Ganglia for trend analysis and alert monitoring

Questions?