The Checkpoint Blues

Baron Schwartz
MySQL UC 2011
Percona Server

- Replaces MySQL
- Faster Queries
- More Consistent
- More Measurable
- More Features
Percona XtraBackup

- Backs Up InnoDB
- Non-Blocking
InnoDB in 2007

Figure 2-1. Results from a 30-minute dbt2 benchmark run
InnoDB in 2011

From http://www.mysqlperformanceblog.com/2010/12/20/mysql-5-5-8-and-percona-server-being-adaptive/
How InnoDB Does a Checkpoint

• Continuously moving “fuzzy” checkpoint
  • c.f. Gray & Reuter, or Weikum & Vossen
  • Or http://www.xaprb.com/blog/2011/01/29/how-innodb-performs-a-checkpoint/
InnoDB Flushing

• There are two lists for flushing in InnoDB
  • LRU list, when a page read forces out a dirty page
  • Flush list, when checkpointing is needed
• See
  http://www.mysqlperformanceblog.com/2011/01/13/different-flavors-of-innodb-flushing/
Why is Flushing Important?

- Don't flush dirty data immediately
- This is an optimization
- I call it *write combining*
InnoDB's Constraints

- Don't run out of space in the logs
- Don't let the buffer pool get more than X% dirty

See
InnoDB Adaptive Flushing

• Flush routinely as background work
• How much to flush at a time?
  • How many dirty blocks there are
  • How fast the log sequence number increases
  • How large the logs are
  • But respect innodb_io_capacity configuration option
XtraDB Adaptive Flushing

• There are several algorithms, improving over time
• The current stable algorithm in our GA release:
  • Defer flushing, then estimate, then flush storm
• How much to flush?
  • How many dirty blocks there are
  • How fast the LSN increases
  • Average of modified age of dirty blocks
What Causes Flushing Storms?

- Running into one of the limits
  - Too many dirty blocks
  - Running out of space in the logs (oldest dirty block is too old)
What Makes Life Hard?

- Dirty block age is not uniformly distributed
- LRU flushing can be counter productive
- Neighbor block flushing doesn't really work
- Disks and SSDs need different optimizations
- Some blocks are very hot
Ongoing Research

• XtraDB flushing is very stable, in most cases
• But even XtraDB can get into trouble when...
  • Slow disks, so they can't keep up
  • Workload fits in memory, so even fast disks can't keep up (or XtraDB can't use all available capacity)
Ongoing Research

• Vary the flushing speed, based on how fast pages are being dirtied
• We hope to release this algorithm very soon
We're Hiring!  www.percona.com/about-us/careers/

baron@percona.com