Voldemort on Solid State Drives

Vinoth Chandar, Lei Gao, Cuong Tran
LinkedIn Corp, Mountain View, CA
What is Voldemort?

- Distributed key-value storage system
  - Based on Amazon Dynamo
- Open source
  - [https://github.com/voldemort/voldemort](https://github.com/voldemort/voldemort)
- Widely used
  - 100K ops/sec, 80TB at Linkedin
  - Eharmony, Nokia, JIVE Software etc.
- We are on SSD!
SSD Migration

- Improve average latency
  - 20 ms to 2 ms
- Cluster expansion & data restoration
  - Improved 10x
- 95\textsuperscript{th} & 99\textsuperscript{th} latencies shot up
  - 30ms to 130ms and 240ms to 380ms
- GC Issues !!

Client operations

BDB Cache
App1  App2  .. AppN

BDB Cleaner

Retention Job

Server JVM Heap

Linux
Benchmarking for SSD

- Java based systems
- Need for End-End Correlation
  - Linux page stealing
  - Use AlwaysPreTouch
- Run Workload causing Fragmentation
  - Promotion failure more likely
  - Defragmentation costly
- Run Workload with High memory churn
  - cost of CMS Initial Mark
  - High CMS initiating fraction
- Factor in higher IO Parallelism
  - Sequential IO benchmarks no longer use single thread
Takeaways

• GC impact significant
  • Big heaps = Big pauses
• Longer runs to account for SSD GC
  • SSD is log structured
  • Worst case 2 ms
• Special attention to 95\textsuperscript{th} & 99\textsuperscript{th}
  • Fragmentation eats up gains