Hadoop Benchmarking

Owen O’Malley
owen@hortonworks.com
@owen_omalley

May 2012
Who Am I?

- Was working in Yahoo Search in Jan 2005, when we decided to work on Hadoop.
- My first patch went in before it was Hadoop.
- Was the first committer added to the project.
- Was the first Apache VP of Hadoop.
- Was the tech lead for:
  - MapReduce
  - Security
- Now I’m working on Hive
First Hadoop Benchmarks

• Created random writer and sort benchmark.
• Created 10 GB/machine and sorted it.
• Very good for measuring scalability
• Tracked performance of Hadoop weekly
  • We were releasing Hadoop monthly!
  • Generated trending graphs to spot problems
• Challenged by
  • Random variances
  • One particular job profile
Sort Benchmark

[Graph showing the decrease in hours to sort (10 GB/Node) for different node counts over time from May 2006 to July 2007. The graph includes lines for 20, 100, 500, and 900 nodes, demonstrating a significant reduction in hours required as the number of nodes increases.]
Terasort Wins

- Nice simple micro-benchmark
  - Straight disk to disk sort
- Applies to wide range of solutions
  - Provide a consistent set of requirements
  - Allow flexibility where possible
- Separating out general software from benchmark specific software
- Deterministic data generator is consistent
- The test and the data scale
- Publicity
Terasort Misses

• Rules changes
  • Gave a few months of notice
    • A missed deadline means no benchmark
    • Need > 1 year notice of changes
  • Flattened distribution
    • Generated data should *not* be flat
  • Solution-based requirements
    • No requirements like, if MapReduce…
Terasort Questions

- How do you test scalability of cluster software?
- Without testing the scalability of their wallet?
  - Have resource limited variants:
    - Money – Too small (1₵)
    - Energy – hard to determine
- How do you do verification?
  - Have them document what they did including hardware and software.
  - Limit to open source?
GridMix

- How do you measure Hadoop performance?
  - Often between versions of Hadoop
  - Security work had to be < 3% slower
- Need to simulate real workloads
- Running on clusters that may be smaller
- Focused on measuring the size of the data in the MapReduce pipeline.
  - Input bytes, Shuffle bytes, Output bytes
  - Input records, Shuffle records, Output records
- Three versions of GridMix in Hadoop
GridMix Versions

• Version 1
  • Parameters for the number of machines
  • Ran specific mix of MapReduce jobs
• Version 2
  • Parameters controlling types of jobs
  • Includes MapReduce and Pig jobs
• Version 3
  • Allows replay of workloads
  • Scaled to size of cluster
  • Time delay or saturation
PigMix

- Set of queries that model realistic ones
  - Based on TPC-H
  - Isn’t balanced in terms of frequency
- Set of generated data
- Lack of scalability
  - Assumes a particular cluster size
- Both Hive and Pig have used the Brown MapReduce benchmarks
Hints for Hadoop Benchmarking

- Always ensure that you have enough parallelism.
  - Customer once benchmarked Hadoop scalability
    - 4 nodes versus 8 nodes
    - Didn’t see much improvement
    - Had a 2 map and 1 reduce job
- Use much much larger tasks than you expect.
- Flush the buffer cache
- Turn off logging
- Pay attention to bad hardware.
- Watch for bottlenecks in the system including NFS
Final Thoughts

• Needs to be vendor independent
• Have both micro & high-level benchmarks
• Publicize the rules changes early
• Ensure the benchmark is scalable
• Include resource limited variants
• Include the winners’ papers on the website
Thank You!

Questions & Answers