DEVELOPING SYSTEM PERFORMANCE METRICS FOR CLOUD COMPUTING BASED ON HADOOP

Rema Hariharan, Gabriele Jost, Sanjiv Lakhanpal, Dave Raddatz, AMD (Advanced Micro Devices)

Project goals:

- Identify optimization opportunities for Hadoop workloads
- Establish a set of tools and techniques for Hadoop workload analysis
- Provide feedback to the user community
  - Eg: Best Practices Guide
- Provide feedback to the open source developer community:
  - Eg: Optimizations for the Hadoop Eco system
- **Performance Metrics** and relevant **Benchmark Use Cases** are crucial!
BENCHMARKS AND METRICS

Benchmarks

- How measurements are done?
  - Blackbox
  - Whitebox

- What type of service is tested?
  - IaaS
  - PaaS
  - SaaS

- What workload/use-case is run?
  - Data Analytics
  - Data Serving (reading and writing data)
  - Web Serving
  - Search
  - Streaming
  - Software testing

Metrics that make it a cloud benchmark

- Throughput
- Response time

- Elasticity
  - Provisioning Interval
  - Agility
  - Scale up
  - Elastic Speedup

- Variability
Approach

- SUT measurement tool set to measure, quantify performance metrics
- Hadoop for storage and analysis of Big Data workloads
  - Best known components HDFS MapReduce
  - Consider all components of Hadoop Eco System
- Code Optimization:
  - HWC and timings mapped to processes, threads, subroutines, source code, assembly
  - gprof, perf, CodeAnalyst, others?
- CPU Performance:
  - System wide HWC, system time, idle time etc mapped to cores
- Node and Cluster performance
  - CPU-to-memory, CPU-to-IO performance (dstat, netstat, etc.)
  - Network monitoring
  - Quality of service performance like Response time, latency
Trademark Attribution

AMD, the AMD Arrow logo and combinations thereof are trademarks of Advanced Micro Devices, Inc. in the United States and/or other jurisdictions. Other names used in this presentation are for identification purposes only and may be trademarks of their respective owners.

©2012 Advanced Micro Devices, Inc. All rights reserved.