Improve Hadoop Economics, Performance, and Security with Compression and Encryption

Ravi Lambi
Director of Software Engineering
Data Compression and Security Business Unit
Exar Corporation
The Storage IO Bottleneck Performance Gap

Santa Clara, CA USA November 2014
The Storage IO Bottleneck

Current Server Solution

More Disks and more rack space

This will increase management cost, and also require more expensive storage controller. Additionally, there is a limit to scale the width – each server has a hard physical limit.
The Storage IO Bottleneck
Summarized Challenge

It Is Difficult to Balance Performance, Capacity Scaling and Cost Associated with the Storage IO
Compression Technology
Where In Hadoop To Apply Compression?
Exar’s Hadoop Acceleration - AltraHD
Offload Compression and Accelerate

File System Filter Driver

Compression Codec

Ingest
Map
Compress/Distribute
Decompress/Compute
Reduce
Output

Disk

File System Filter Driver

5 GB/sec HW Compression & Encryption Accelerator

Network
AltraHD Overview
Core Technology

- **File System Filter Driver**
  - Kernel plug-in at the file system layer
  - Compresses/decompresses ALL files independent of application

- **Transparent to the Application**
  - No modification to Applications or Workflow

- **Seamlessly Layers over File System**
  - Supports EXT3, EXT4, or XFS

- **Fast, Easy Deployment**
  - No APIs – Software installs in minutes

- **Hardware Acceleration Offloads Host CPU**
Value Proposition – Performance
MapReduce 1 Terasort Benchmark

Santa Clara, CA  USA
November 2014
Value Proposition - Performance
MarReduce2 Job Execution Time

Santa Clara, CA  USA
November 2014
Value Proposition – Storage
Increased Storage Capacity

MR2 Effective Storage Capacity
- Native: 192 terabytes
- AltraHD: 672 terabytes

MR1 Effective Storage Capacity
- Native: 192 terabytes
- AltraHD: 1344 terabytes
Value Proposition – Storage

Increased Storage Capacity

AltraHD – Effective Storage

Native - Storage
Value Proposition – Security

- Exar’s Compression Acceleration Card Supports Compression, Encryption, and Hashing in a Single Pass
- Aligned with Hadoop Security Roadmap
Value Proposition – Indirect Values

Other Savings

- Reduce Indirect Costs:
  - Power
  - Rack Space
  - Cooling
  - Disk Life
  - etc.
Thank You!

Ravi Lambi
ravi.lambi@exar.com