Maximizing Hadoop Performance and Storage Capacity with AltraHD™

Executive Summary

The explosion of internet data, driven in large part by the growth of more and more powerful mobile devices, has created not only a large volume of data, but a variety of data types, with new data being generated at an increasingly rapid rate. Data characterized by the "three Vs" – Volume, Variety, and Velocity – is commonly referred to as big data, and has put an enormous strain on organizations to store and analyze their data. Organizations are increasingly turning to Apache Hadoop to tackle this challenge. Hadoop is a set of open source applications and utilities that can be used to reliably store and process big data.

What makes Hadoop so attractive?

- Hadoop runs on commodity off-the-shelf (COTS) hardware, making it relatively inexpensive to construct a large cluster.
- Hadoop supports unstructured data, which includes a wide range of data types and can perform analytics on the unstructured data without requiring a specific schema to describe the data.
- Hadoop is highly scalable, allowing companies to easily expand their existing cluster by adding more nodes without requiring extensive software modifications.
- Apache Hadoop is an open source platform.
- Hadoop runs on a wide range of Linux distributions.

The Hadoop cluster is composed of a set of nodes or servers that consist of the following key components.

Hadoop Distributed File System (HDFS)

HDFS is a Java based file system which layers over the existing file systems in the cluster. The implementation allows the file system to span over all of the distributed data nodes in the Hadoop cluster to provide a scalable and reliable storage system.

MapReduce

MapReduce is a two stage analytic function used to search and sort large data sets, leveraging the inherent parallel processing of Hadoop. Mapping involves reading the dataset from HDFS and organizing it into an appropriate format for analysis. Reducing then analyzes the mapped data to yield the desired information.

In the evolution of Hadoop, there have been two versions of MapReduce.
MapReduce 1 (MR1) incorporates MapReduce and the resource manager into a single module.

MapReduce 2 (MR2) was introduced with hadoop-0.23. In MapReduce 2, MapReduce and the resource manager are separated into discrete modules. The resource manager for MR2 is commonly called YARN (Yet Another Resource Negotiator). MR2 allows additional applications to be run on Hadoop because they plug directly into YARN. For example, YARN enables different versions of Hadoop to be run in the same cluster, or run applications that do not follow the MapReduce model.

When data is written to HDFS, each block of data written into the file system is replicated, resulting in three copies of each block of data by default. If there are enough datanodes in the cluster, each of the data blocks will be stored on different nodes. In this way, the cluster can sustain the loss of a disk or even a server because the data is distributed. The cost of this redundancy is that it requires three times the storage and the distributed data can create I/O bottlenecks.

MapReduce also adds additional I/O overhead when splitting the data set into smaller segments that can be sorted and filtered. The MapReduce operation generates disk I/O traffic when data is moved to storage devices, and networking I/O traffic, or "shuffle" traffic, when data is moved between nodes. Hadoop attempts to localize the data to minimize the amount of data movement, but bottlenecks still occur in the processing of a job, often due to the transfer speed and rotational latency of spinning disks.

This paper describes a solution that may be seamlessly added to a Hadoop cluster to transparently compress HDFS data and intermediate data. Reducing the amount of data being transferred greatly improves I/O performance for both storage and networking which translates into enhanced system performance. Compressing the file system and intermediate data results in increased storage capacity.

Market Dynamics

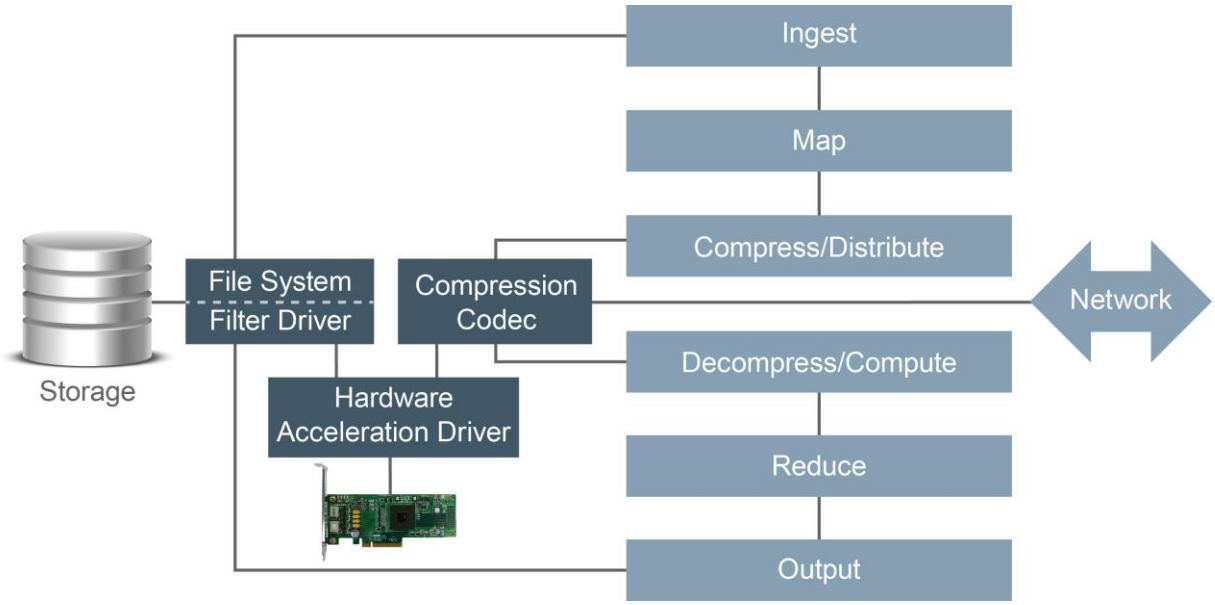
IDC estimates the 2.8 ZetaBytes of data generated in 2012 is expected to grow to 40 ZetaBytes by 2020[1]. Eighty-five percent of this growth is expected to come from new data sources with a 15x increase in machine-generated data. As the quantity of data continues to grow, the amount of processing power needed to analyze the data and the amount of storage needed will increase. Typically the solution is to add more datanodes to the cluster, but doing so increases the CAPEX, along with the OPEX from powering, cooling, and cabling the additional datanodes, and results in escalating TCO.

Exar provides a solution that transparently compresses data in the Hadoop cluster, provides more storage capacity to the cluster, and enables more efficient execution of Hadoop jobs. Exar's state of the art AltraHD solution combines Exar's software technology with its leading edge hardware compression accelerators to remove costly I/O bottlenecks and optimize the storage capacity for Hadoop applications.

The Exar AltraHD™ Solution

Exar's AltraHD's solution for Hadoop includes the following components:

1. An application transparent file system filter driver that sits below HDFS and automatically compresses/decompresses all files using HDFS. This enables transparent compression for all modules that interface to HDFS, including MapReduce, HBase, and Hive.
2. A suite of compression codecs that automatically compress/decompress intermediate data during the MapReduce phase of Hadoop processing.
3. A high performance PCIe-based hardware compression card that automatically accelerates all compression and decompression operations, maximizing throughput while offloading the host CPU. Each DX2040 card provides up to 5 GB/sec of compression/decompression throughput, optimizing workloads and delivering maximum system performance. The DX2040 card also supports encryption and hashing to support future Hadoop roadmaps.



[1] THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East- http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf

AltraHD integrates seamlessly into the Hadoop stack, and *transparently* compresses all files, including files that are stored using HDFS, as well as files stored locally as intermediate data outside of HDFS.

AltraHD is a plug and play solution that installs easily and quickly on each Hadoop datanode without requiring kernel recompilation or modification of user applications. Once installed, all file accesses are automatically accelerated and optimized.

The performance increase and storage optimization gained by using the Exar AltraHD solution can reduce the number of nodes required as well as the associated power, cooling, and space requirements. This minimizes both CAPEX and OPEX, reducing the overall TCO for the solution. The storage savings provided by AltraHD is proportional to the source data compression ratio, and will therefore vary based on the data sets.

Benchmark Results

This section presents two benchmarks. The Hadoop cluster's performance was measured using Terasort, the industry standard benchmarking tool. The following table describes the test cluster configuration.

Hardware Configuration		
	Name Node	**Data Node**
Number of Nodes	1	8
CPU Model	Intel E5520	Intel E5-2630
Speed	2.27 GHz	2.3 GHz
# of Cores/Threads	8/16	12/24
RAM	16 GB	96 GB
Network	10Gbit	10Gbit
Software Configuration		
AltraHD Release	AltraHD 1.1.0L	
Hadoop Version (applies to both MR1 and MR2)	Hadoop 2.2.0-cdh5.0.0-beta-1	
Operating System Version	CentOS 6.2 (2.6.32-220)	

The first set of benchmark data was collected using Hadoop MapReduce 1. The second set of benchmark data was collected using MapReduce 2. For comparison, data for each benchmark was collected in both a native environment and in an AltraHD environment. The Hadoop configuration settings were optimized for both the native and AltraHD environments.

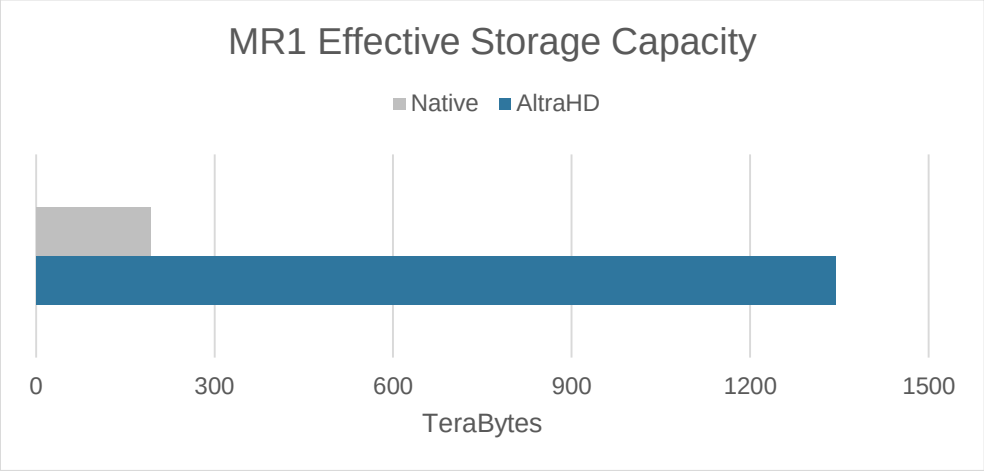
The native Hadoop cluster uses the industry standard software LZO codec to compress the intermediate data, but does not compress file system data. The AltraHD Hadoop cluster uses the AltraHD file System filter driver to compress file system data and the zlib AltraCodec to compress intermediate data, providing a complete compression solution for Hadoop.

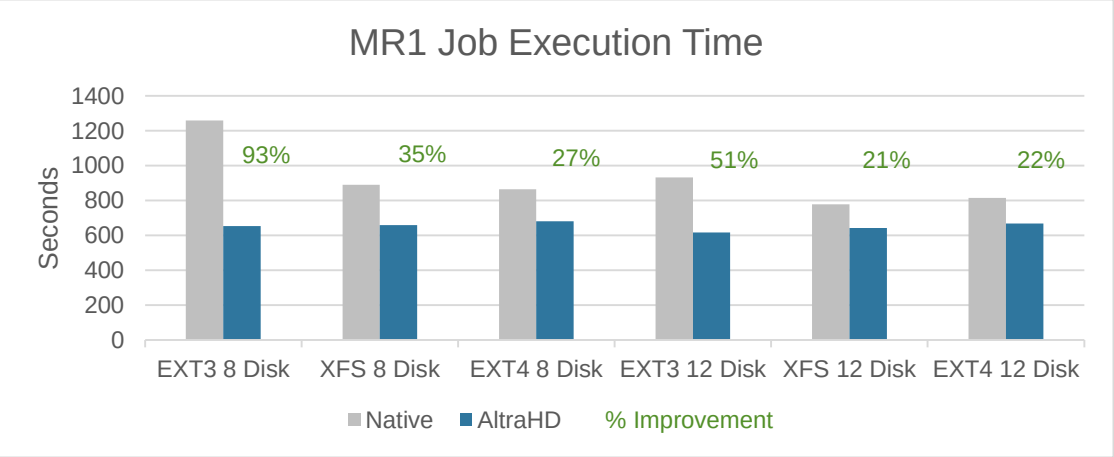
The performance metrics shown in the charts below include:

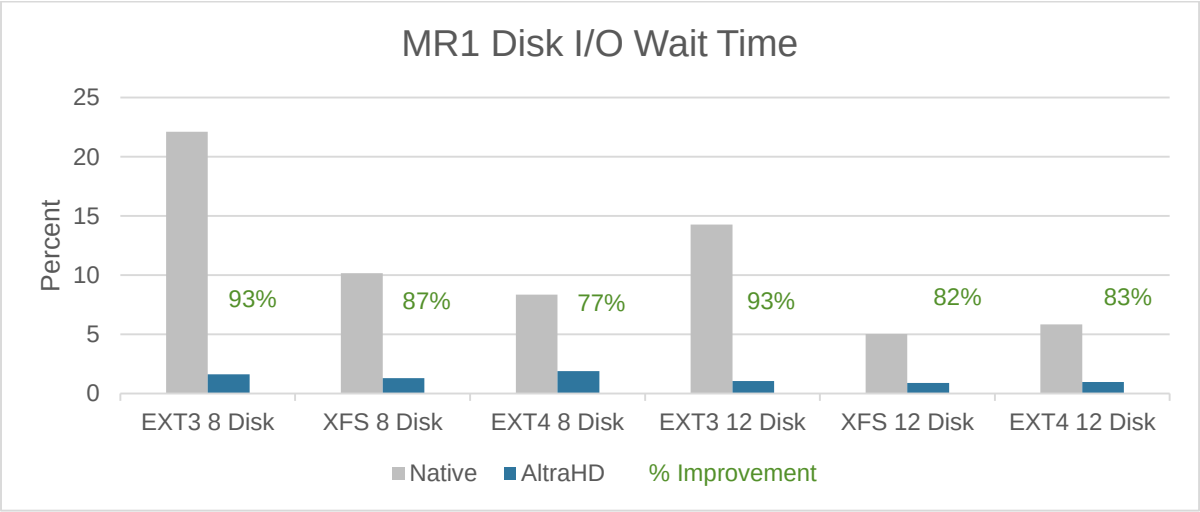
- Job Execution Time: the time required for the job to run. A smaller job time translates to more time available for the CPU to perform other tasks.
- Disk I/O Wait Time: the time the CPU spends waiting for disk I/O accesses to complete which indirectly measures the read latency. The lower the disk I/O wait time, the less time the CPU is waiting for an I/O access to complete.
- Disk I/O Utilization: The percentage of time that the disk is actually in use by Hadoop. The lower the disk I/O utilization, the more time other applications can utilize the disk.

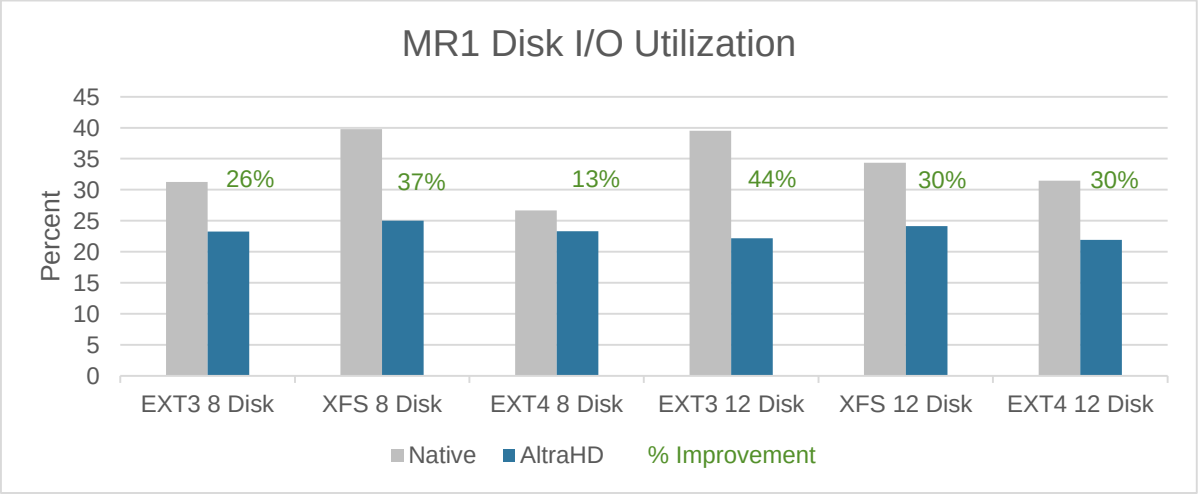
MapReduce 1 Benchmark Results

Using AltraHD in a Hadoop MR1 environment yields a disk compression ratio of 7:1 which corresponds to an increase of storage capacity of 7x.



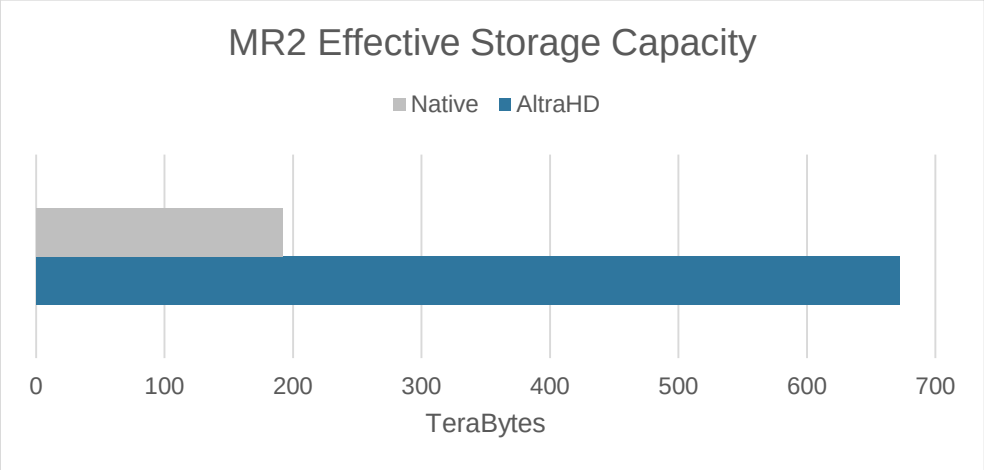


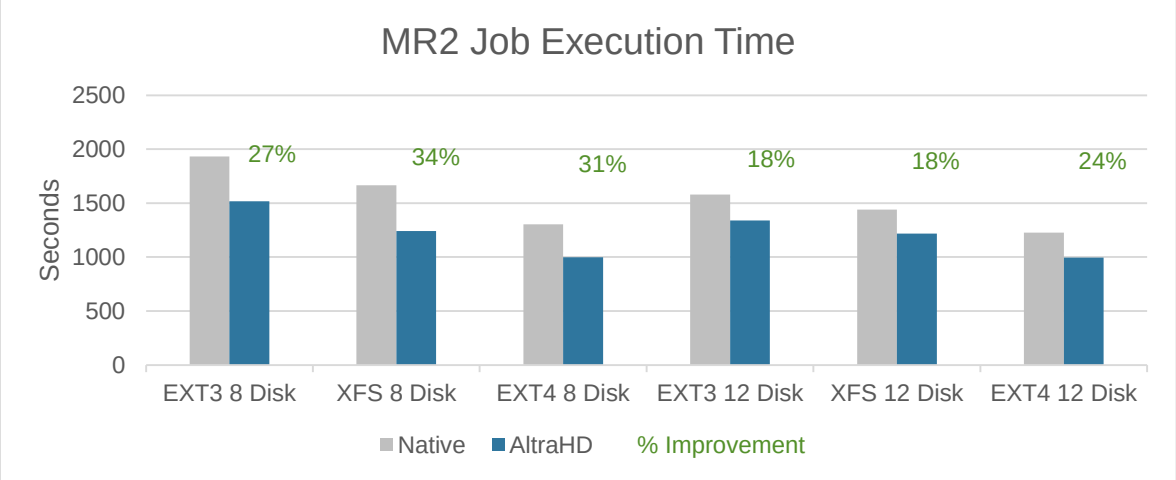

MR1 Disk I/O Wait Time
Percent
25
20
15
10
5
0
93%
87%
77%
93%
82%
83%
EXT3 8 Disk
XFS 8 Disk
EXT4 8 Disk
EXT3 12 Disk
XFS 12 Disk
EXT4 12 Disk
Native
AltraHD
% Improvement


MR1 Disk I/O Utilization
Percent
45
40
35
30
25
20
15
10
5
0
26%
37%
13%
44%
30%
30%
EXT3 8 Disk
XFS 8 Disk
EXT4 8 Disk
EXT3 12 Disk
XFS 12 Disk
EXT4 12 Disk
Native
AltraHD
% Improvement

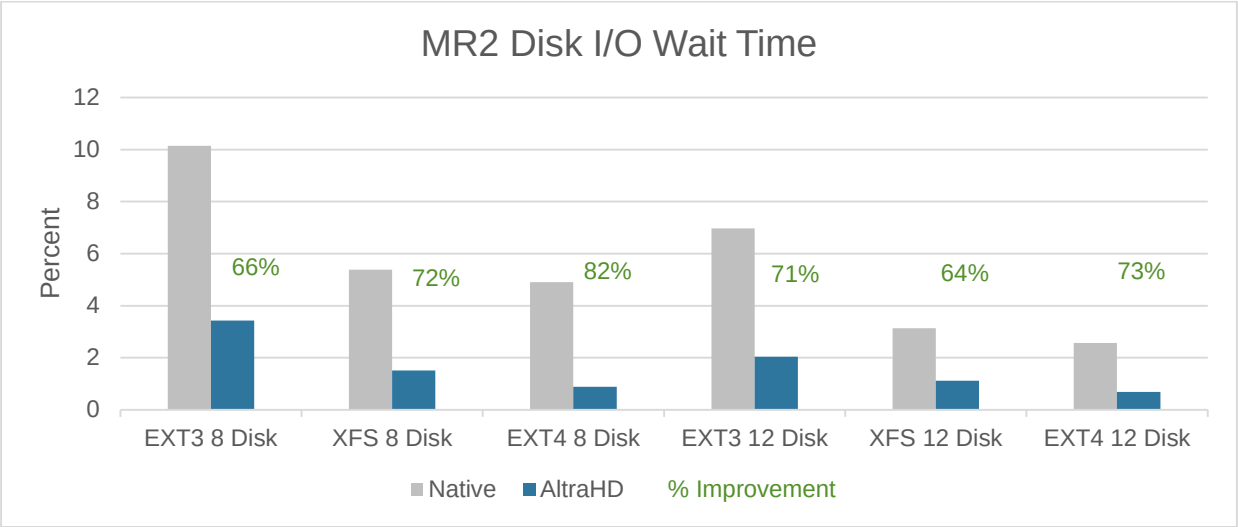
MapReduce 2 Benchmark Results

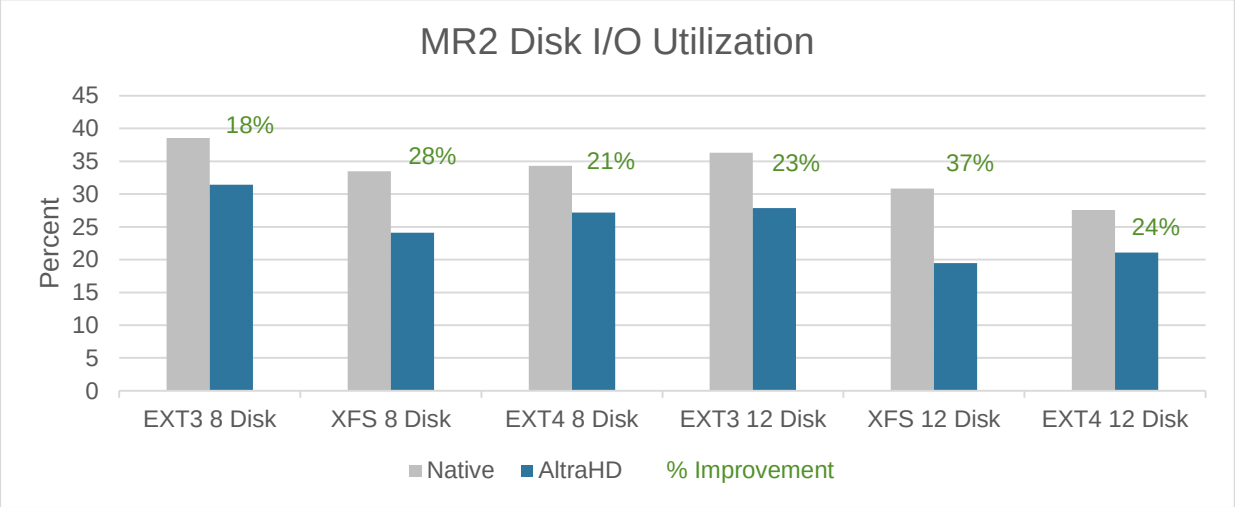
Using AltraHD in a Hadoop MR2 environment results in a disk compression ratio of 3.5:1 which corresponds to an increase of storage capacity of 3.5x. The difference in compression ratio compared to MR1 is due to changes to in the Terasort GraySort algorithm[2] that produces a less compressible data set.





[2] Getting MapReduce 2 Up to Speed - Cloudera Engineering Blog - http://blog.cloudera.com/blog/2014/02/getting-mapreduce-2-up-to-speed/





Conclusion

Hadoop is increasingly being used by leading edge companies to manage and analyze the explosion of big data on their servers. However, Hadoop replication adds a 3x multiplier to their already large storage footprint and creates I/O bottlenecks as the additional data is accessed on the network and disk. Exar's AltraHD solution addresses these issues by increasing the overall performance by up to 2x and by boosting the storage capacity by an amount proportional to the data compression ratio. The increase in performance and storage capacity translates to lower CAPEX and OPEX, resulting in a significant reduction in TCO.

For more information about the Exar AltraHD solution, visit Exar's website at www.exar.com or call 510-668-7000.