AltraHD™

High Performance Application Transparent Compression for Hadoop

Product Overview

AltraHD combines Exar's state of the art software technology with its leading edge hardware compression accelerators to remove costly I/O bottlenecks and optimize the storage capacity for Hadoop applications. AltraHD integrates seamlessly into the Hadoop stack, and is able to transparently compress all files, including files that are stored using the HDFS file system as well as files stored locally as intermediate data outside of the HDFS file system. AltraHD is the only compression solution for Hadoop to offer all of the following key features:

- Exar's application transparent file system filter driver, which sits below the Hadoop Distributed File System (HDFS), automatically compresses/decompresses all files that are using HDFS. This enables transparent compression for all modules that interface to HDFS, including MapReduce, HBase, and Hive.
- Exar's family of compression codecs automatically compresses/decompresses intermediate data during the MapReduce phase of Hadoop processing.
- Exar's high performance PCIe-based compression acceleration card automatically accelerates all compression and decompression operations, maximizing throughput while offloading the host CPU. This optimizes workloads and delivers maximum system performance. A single card provides up to 5 GB/sec of compression/decompression throughput.

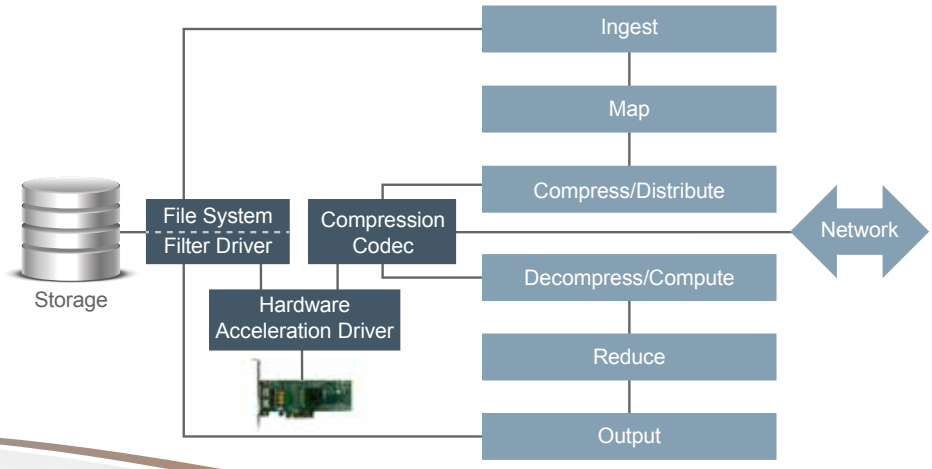
AltraHD is a plug and play solution that installs easily and quickly on each Hadoop datanode without requiring kernel recompilation or modification of user applications. Once installed, all file accesses are transparently accelerated and optimized.

Key Benefits

Exar's AltraHD addresses multiple issues with Hadoop clusters, delivering a multitude of benefits. The large amount of data processing that occurs with Hadoop can cause the system to become I/O bound, causing the CPU to wait for data to be retrieved from the storage or networking I/O subsystems and reducing system performance. In addition, the storage footprint can expand to a point where additional nodes are added to address the expanding storage requirements. AltraHD provides the following benefits to solve these problems:

- System performance is maximized by reducing or eliminating costly I/O bottlenecks, delivering up to 2x performance increase.
- The storage capacity is increased in proprtion to system data compressibility, resulting in an optimized storage subsystem.
- The performance increase and storage optimization 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 solution TCO.

AltraHD Functional Diagram within Hadoop Stack



AltraHD™

High Performance Application Transparent Compression for Hadoop

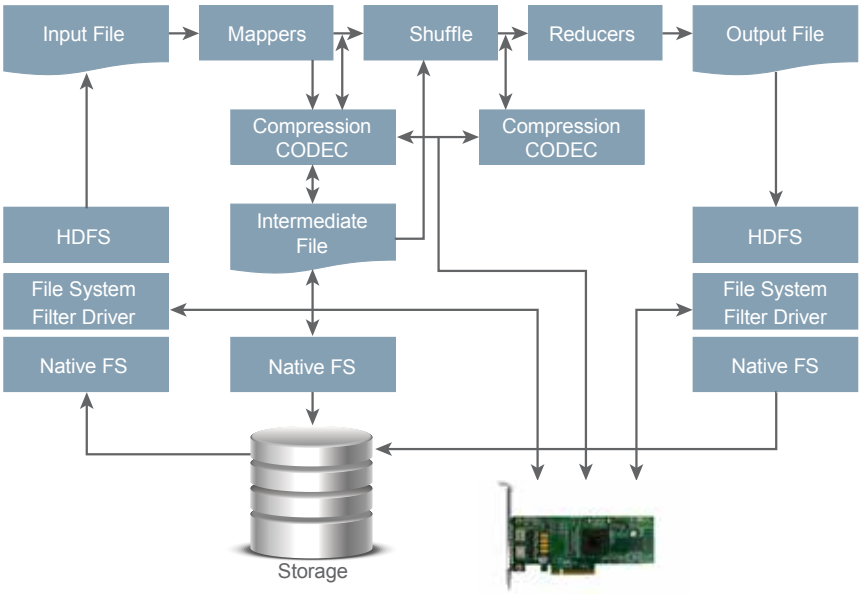
Data Flow

AltraHD's data flow during MapReduce is shown below, illustrating how the file system filter driver and compression codec work together to optimize system performance and efficiency:

1. The Map job begins by reading the input file from HDFS, and the compressed file is transparently decompressed by AltraHD's file system filter driver and given to the Mapper.
2. As the Mapper processes the data, any intermediate data that needs to be stored is compressed using AltraHD's compression codec. In addition, during the shuffle phase, compressed data is sent over the network to the Reducer.
3. The Reducer receives the data from the Mapper and decompresses it using AltraHD's compression codec and continues processing.
4. When the reducer finishes its processing, the output file is stored to HDFS and automatically compressed using AltraHD's file system filter driver.

All compression/decompression operations are accelerated using Exar's compression acceleration card for maximum performance.

AltraHD Data Flow During MapReduce



Flexible Deployment

AltraHD has a flexible, modular architecture that enables multiple deployment models. Typically both the file system filter driver and the compression codec are deployed together on each Hadoop datanode, delivering optimal results. In the case where AltraHD nodes are being added incrementally to an existing Hadoop Cluster, AltraHD has the ability to be deployed on the new nodes with the file system filter driver only, allowing the existing compression codec to continue to be used on each node within the cluster, with the new nodes benefitting from AltraHD's hardware accelerated file system compression. This enables AltraHD to be deployed in a variety of environments, and allows it to be added seamlessly on existing clusters.

System Support

AltraHD supports a wide range of Linux distributions, and is compatible with the EXT3, EXT4, and XFS file systems. AltraHD has been validated with multiple Hadoop distributions, including Cloudera, and Hortonworks.

EXAR CORPORATION
48720 Kato Road
Fremont, CA 94538
U.S.A.
T. +1.510.668.7000
F. +1.510.668.7001

FLY1014_ALTRAHD