

# Unstructured Data Accelerator (UDA)



## Unstructured Data Accelerator

With the current data explosion and focus on data analytics across a range of sectors including technology, web2.0, cloud, banking, government and entertainment, companies in these sectors are adopting Hadoop at an exponential rate to achieve their business goals.

Hadoop is a popular open-source implementation of the MapReduce programming model for data analytics. It is an easy-to-use programming interface for numerous organizations to process explosive amounts of data, perform massive computation, and extract critical knowledge for business intelligence. Apache Foundation currently maintains Hadoop and it also has support from leading technology companies such as Google, Yahoo!, Facebook and LinkedIn.

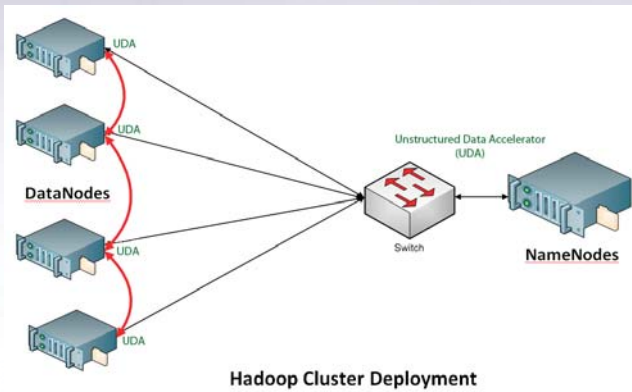
Traditional Ethernet networks are no longer capable of delivering the performance required for Hadoop clusters. Typical hierarchical Hadoop cluster utilizing TCP/IP over one or more Gigabit Ethernet (GbE) network interface cards connected to a GbE fabric can achieve only 125 MB/s of bandwidth per port. Multi-socket, multi-core server have outgrown this capacity, and with advances in processor technology there will soon be compute servers with hundreds of cores in the mass market. To achieve highest efficiency with these servers there must be enough bandwidth available for each with CPU offloads that prevent data movement from overwhelming the server's CPU. High bandwidth technologies InfiniBand and Ethernet deliver up to 40 Gb/s of bandwidth, and each have RDMA (Remote Direct Memory Access) capabilities to offload data movement. However, to utilize RDMA Hadoop needs a special interface to the network card driver.

## Mellanox UDA Solution

Mellanox UDA, a software plugin, accelerates Hadoop network and improves the scaling of Hadoop clusters executing data analytics intensive applications. A novel data moving protocol which uses RDMA in combination with an efficient merge-sort algorithm enables Hadoop clusters based on Mellanox InfiniBand and 10GbE RoCE (RDMA over Converged Ethernet) adapter cards to efficiently move data between servers accelerating the Hadoop framework. It is the first networking solution to tackle Hadoop node bandwidth and compute issues from a cluster perspective by leveraging InfiniBand and RoCE based superior lossless fabric. UDA is transparent to Hadoop users and current applications will run as is. Cluster administrators need only be aware of UDA in order to configure the cluster to leverage UDA advantages, while users can focus on Hadoop benefits.

## UDA Performance

Unstructured Data Accelerator more than doubles the data processing throughput and reduces total job execution time by half per Hadoop node. It is designed to scale with larger datasets to provide similar or better performance benefits. With increased CPU efficiency per node due to lower total job execution time, clusters will increase their corresponding power efficiency for the defined configurations which directly adds to datacenter power savings and aligns with the datacenter green initiative. Higher bandwidth with scale out architecture based on RDMA over InfiniBand and Ethernet provides consolidated single networking pipe to transfer larger datasets across the wire.



### UDA Availability

UDA acceleration kit is available for interested customers to deploy and evaluate tremendous benefits of UDA in their datacenters. Kit consists of Mellanox UDA acceleration software and networking hardware. For further information about how to order the kit, contact your Mellanox sales representative or contact us via [http://www.mellanox.com/content/pages.php?pg=buy\\_overview](http://www.mellanox.com/content/pages.php?pg=buy_overview)

UDA 2.0 is jointly developed by the Parallel Architecture and System Laboratory headed by Dr. Weikuan Yu from Auburn University and Mellanox.

Figure1. Hadoop Cluster Deployment

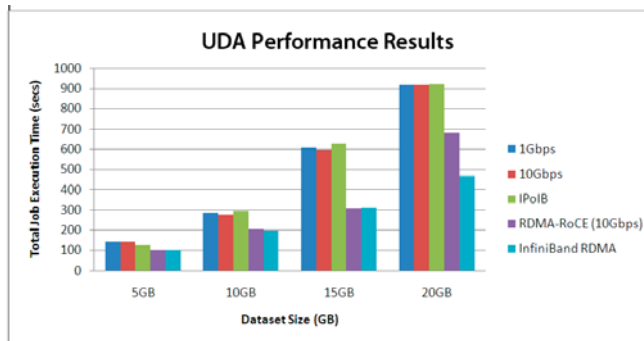


Figure2. UDA Performance Results

### UDA Key Advantages

- Leverages world's fastest interconnect that supports 40Gb/s InfiniBand or Ethernet fabric
- Increases Hadoop node efficiency by processing data with RDMA technology and efficient merge-sort algorithm
- Lowers total job execution time per node
- Lossless scalable fabric solution



350 Oakmead Parkway, Suite 100, Sunnyvale, CA 94085  
 Tel: 408-970-3400 • Fax: 408-970-3403  
[www.mellanox.com](http://www.mellanox.com)