SOLUTION BRIEF

High Performance Big Data Collection and Processing

Background

In the area of network data processing, organizations often collect millions of network events, and still have minimal-to-no visibility into the actual performance, health and security status of the network. The challenge is to be able to collect millions of network events per second for cyber forensics, network management, IoT, telco and mobile operations and receive real information that can lead to better operational and business decisions in these areas.

All-In-Image Big Data Platform

All-In-Image innovative software technology reduces the cost of data management and addresses the challenges of big data applications.

The All-In-Image product is a distributed file-based database for the holistic management of vast-scale structured and unstructured data. All-In-Image enables the aggregation, storage, and querying of terabytes and petabytes of heterogeneous data at significant speeds with reduced compute, network, storage and maintenance costs.

By enabling more data collection per server with linear scale, customers reduce energy, space and management costs in the data center. In addition, after the data collection, All-In-Image provides high performance data collection (e.g., 1.2 million events/second per core from NIC to DB), strong data compression (e.g., Netflow x15), and a parallel SQL query engine to process the data at Petascale. Target markets include cyber security, IoT, mobile/telco networks, media, and analytics and BI.

All-In-Image Network Data Collection and Analysis at Petascale

Each file shared within the database is a self-contained, encapsulated database, containing all metadata and compressed data, and can be easily copied and moved between processing elements and storage tiers across the network.

The All-In-Image distributed file-based database can run on top of any scalable file and storage system.

Enabling Enhanced Network-Based Services

With high performance data collection and processing, more advanced network security and network performance optimization can be implemented.

- Network Big Data Processing
- 1.2 million events/sec per core with linear scale from network (NIC) to Petascale DB with strong compression (Netflow x15)
- Fastest networks supported (10-100Gb/s)
- Querying (Parallel SQL, REST API) hundreds

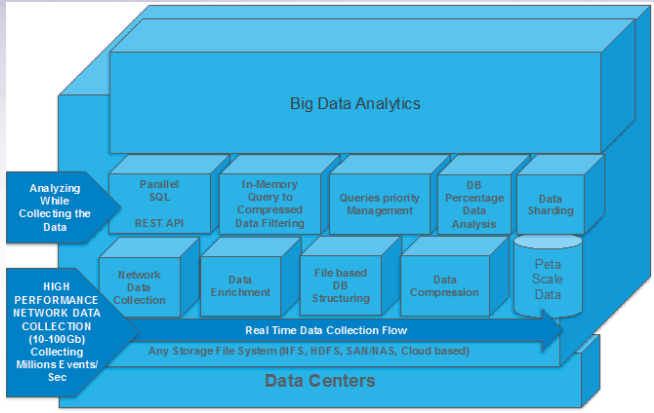


Figure 1. All-In-Image System Components

of billions of records, and collecting millions of new records

- Unified solution for both structured and unstructured (protocols, raw)
- Reduces data center costs (electricity, space, management)

Mellanox Ethernet Solution

Mellanox offers a complete product line of end-to-end 10/25/40/50/56/100Gb Ethernet solutions tailored for Big Data applications like Hadoop and NoSQL. Mellanox delivers industry-leading performance, scalability, reliability and power savings for advanced data center applications.

Big Data applications utilizing TCP or UDP over IP transport can achieve the highest throughput and application density using the market-leading hardware-based offloads and flow steering engines in ConnectX®-4 and ConnectX-4 Lx network adapters.

These advanced stateless offloads reduce CPU overhead in IP packet processing, thereby allowing completion of heavier analytic workloads in less time in the Big Data cluster. Socket acceleration software further increases performance for latency sensitive applications, with faster network speeds that range from 10GbE and 25GbE up to 100GbE.

Mellanox Ethernet switches feature consistently low latency and can support a variety of non-blocking, lossless fabric designs. Furthermore, with the Mellanox NEO™ network orchestration platform, network administrators can leverage existing data center fabric management solutions to deploy, orchestrate and monitor a large-scale cluster easily. To connect the clusters, Mellanox copper cables, active optical cables, and transceivers offer reliable connections at speeds from 10Gb/s to 100Gb/s with the highest quality, featuring error rates up to 100X lower than industry standards.

Mellanox-Based Platform Performance Results

We have simulated a real-world scenario of network data collection with Netflow protocol data, a typical scenario for enterprise Cyber security and network management. The more packets processed per second, the better granularity in network visibility and better decisions can be taken by security analysts and data center admins.

Setup

The setup consisted of 2 Linux Rhel 6.5 Servers, each with the following specs:

- Intel® Xeon® CPU E5-2680 v2 @ 2.80GHz (* 2 sockets * 10 cores per socket)
- 64GB RAM of physical memory
- 1.1TB SATA HDD
- Mellanox ConnectX-4 100Gb/s dual port Ethernet controller card
- Mellanox ConnectX-3 Ethernet controller card

The servers were connected to the management network via the Mellanox ConnectX-3 cards.

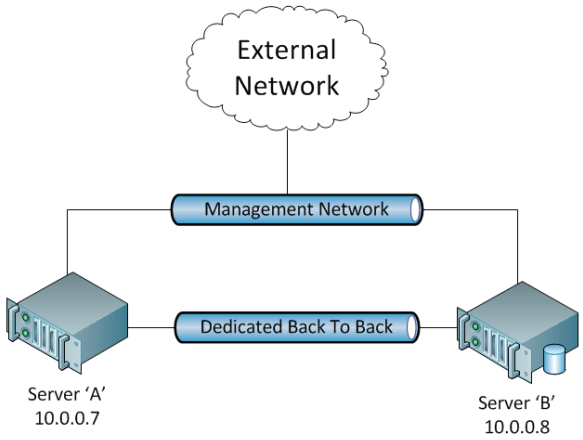


Figure 2. Test Setup

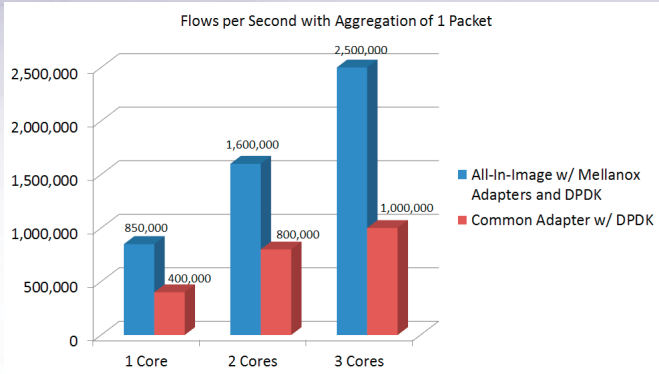
Packets and Injection

The two servers were assigned two different 'roles': Server 'A' was to generate and send data, while Server 'B' was to receive and process data.

Traffic generation: Server 'A' used previously generated captured traffic, where each network flow consisted of a ~100 bytes of information, saved in a pcap file.

The flows were injected into the back-to-back NIC using the 'tcpreplay' utility, using various settings for packet aggregation and packets-per-second.

Packet reception, processing and storage: Server 'B' used



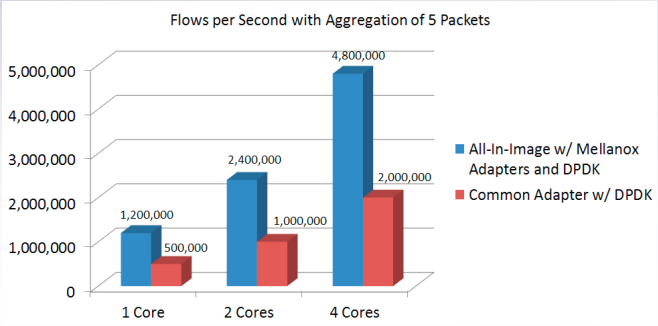


Figure 3. Performance Results (Aggregation = the number of flows in each Netflow packet)

All-In-Image software and Mellanox DPDK drivers for receiving the data, and the All-In-Image system for processing and storing the generated traffic

The test process included various optimizations for utilizing the Mellanox DPDK drivers and the 'tcpreplay' utility to generate data at as high a rate as possible, and the All-in-One technology to receive and process the generated traffic on the other end. Final optimizations allowed for sending and processing five network flows per packet, allowing for a total processing rate of ~1.2M packets per second for each worker CPU on the receiving server.

Results

Mellanox adapters and All-In-Image High Performance Big Data Collection and Processing, provide 2.2-2.5 times more packets per second and bandwidth.

Conclusion

When running big data processing frameworks, such as the All-In-Image Collection and Processing solution with Mellanox network equipment one can gain significantly better performance, security and ROI advantages.

This includes a 2.5X improvement in packet rate, translating into less lost data and more events detected, thereby enabling more overall data collection per server. Moreover, it allows for better network security and better analysis of the network's performance.

Furthermore, this system achieves linear scale, by which customers reduce their energy, space, and management costs by increasing predictability. Network data analysis and data center costs are reduced. This equates to the lowest available Total Cost of Ownership, maximizing system resources that support multiple workloads while building a network analysis solution economically.

Mellanox TECHNOLOGIES
350 Oakmead Parkway, Suite 100, Sunnyvale, CA 94085
Tel: 408-970-3400 • Fax: 408-970-3403
www.mellanox.com

15-51323SB
Rev1.0