OpenFabrics Enterprise Distribution:
# Driving Performance and Efficiency in the TOP500

Introduction

The OpenFabrics Enterprise Distribution (OFED™) software developed and licensed by the OpenFabrics Alliance (OFA) has been a catalyst in some of the key growth trends enjoyed by the fastest supercomputers of the world as compiled and listed by the TOP500 organization (www.TOP500.org). Most notable are the growth of commodity server-based clusters using industry-standard high performance interconnect technologies that are enabled by OFED, and the resultant increase in the performance and efficiency of the TOP500 systems.

The TOP500 Project

The TOP500 project ranks and details the 500 most powerful known computer systems in the world. The project aims to provide a reliable basis for tracking and detecting trends in high performance computing and bases rankings on an implementation of the High Performance LINPACK benchmark. It publishes an updated list of the supercomputers twice a year. The TOP500 list has garnered popularity as an indicator of trends for how systems will be built and deployed in mainstream commercial applications of the future.

Growth of OFED-based Interconnects in the TOP500

The growth of the use of OFED-based interconnects, in the TOP500 has been stellar. As indicated in Figure 1 below, its share on the TOP500 has grown from about 2% in 2004 to over 42% in 2010.

% of OFED-based Interconnects in Top500
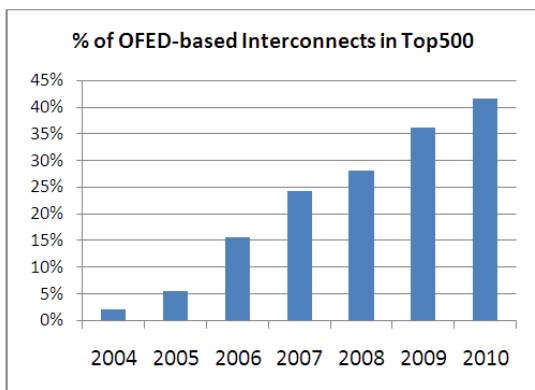
Key OFED Events in time

| Year	Event
2004	Alliance is formed. Initial support on InfiniBand.
2005	OFA software in Linux kernel. Major foundation for common Linux stack, elimination of proprietary stacks
2006	OFED develop starts. 10GbE iWARP support added.
2007	First OFED version released. OFED available in major Linux OS distributions
2008	OFA and UNH-IOL interoperability test events held. Expansion to commercial and datacenter apps.
2009	Fourth OFED version released. Major enhancements to Linux and Windows OFED.
2010	Fifth OFED version released. Support for 10GbE RoCE added.

Figure 1: OFED in the TOP500 and Key OFED Events in time

Key factors driving this growth are (a) the maturity and commoditization of server and RDMA-capable interconnect hardware, (b) standardization of the high performance networking software stack used in these applications – the latter instrumented by the contributions of OFED to the industry, and (c) the compelling performance and efficiency benefits delivered by (a) and (b) together. OFED, by eliminating proprietary RDMA (Remote DMA) stacks and through its inclusion in the kernel and popular operating system distributions, has made use of high performance interconnect solutions like InfiniBand and 10GigE – for iWARP and RDMA over Converged Ethernet (RoCE) – easy and cost effective, making these interconnect solutions leading contenders to the ubiquitous non-RDMA Gigabit Ethernet, which is the other most popular interconnect in the TOP500. It is also noteworthy that the share of proprietary interconnects in the TOP500 has shrunk from almost 40% in 2005 to about 10% in 2010.
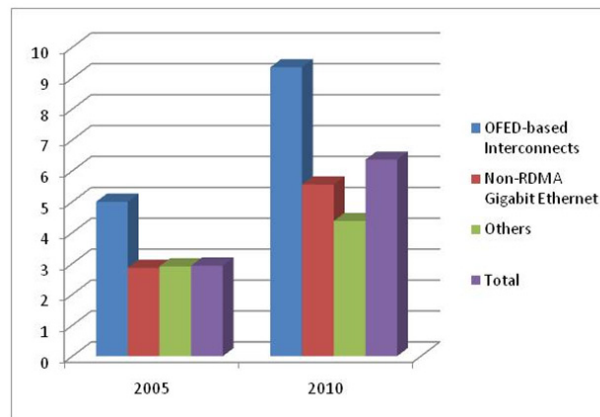
OFED Performance and Efficiency

While non-RDMA Gigabit Ethernet is still used in more than 48% of the TOP500 sites, it is primarily used in bottom half of the TOP500. As shown in Figure 2 below, the aggregate performance of TOP500 systems using OFED-based interconnects in 2010 was 15.89 Peta Flops – which is more than two times the aggregate performance of TOP500 systems using non-RDMA Gigabit Ethernet and 1.7 times that of other interconnects. Comparatively, in 2005 and before the advent of widespread OFED use in the TOP500, the aggregate performance using high performance RDMA-capable interconnects in 2010 was 0.05 Peta Flops, with non-RDMA Gigabit Ethernet accounting for 0.42 Peta Flops. Overall, the chart shows the significant growth in performance of the TOP500 systems in just a span of five years overlapping with the growth of OFED adoption with RDMA-capable interconnects, and how most of that growth is driven by the performance boost delivered by OFED-based interconnects.
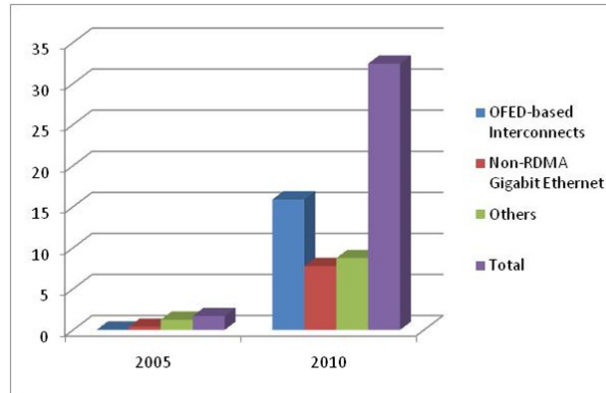
Figure 2: Performance of TOP500 systems in PetaFlops

It is also important to assess the efficiency trends in the TOP500 systems. Higher efficiency implies that more performance is delivered using fewer servers or CPU cores, resulting in more Peta Flops delivered in lesser space, and consuming less power. In the analysis below, Efficiency Index (EI) is defined as Performance in Peta Flops / Processor Sum in Millions and data from the TOP500 is used. As shown in Figure 3 below, in 2010, systems utilizing high performance OFED-based interconnect delivered 68% higher EI than systems connected using non-RDMA Gigabit Ethernet (EI of 9.4 versus 5.6) and more than twice that of other proprietary interconnects (EI of 9.4 versus 4.4). Overall efficiency of TOP500 systems has increased 117% from 2005 to 2010 (EI of 2.9 versus 6.4), and the chart below shows that most of that efficiency gain has been driven by the stellar efficiencies delivered by the systems connected using OFED-based interconnects.

Figure 3: Efficiency Index (PetaFlops / # of Processors) of TOP500 systems

The OFED Solution

As evidenced above, OFED provides high-performance and efficiency networking that enables the majority share of FLOPs amongst the fastest TOP500 supercomputers in the world. OFED is a proven solution in supercomputing and is freely downloadable open source available at www.openfabrics.org.