

iFLYTEK Accelerates Machine Learning Data Center with Mellanox Ethernet Solutions

iFLYTEK, a world leader in intelligent speech and artificial intelligence technologies, has created a portfolio of applications that place high demand on their underlying network. Based in China, iFLYTEK’s speech synthesis, speech recognition, spoken language assessment and natural language processing applications involve the use of Hadoop, MPI, Lustre, and GlusterFS, which require 25Gb/s per host and uplinks reaching 100Gb/s.

To achieve this, iFLYTEK sought to employ a converged network architecture to meet the computing and data storage communications needs. Furthermore, the network required simplified management and the potential to scale in the future as their business grows.

Mellanox Ethernet Solutions

Mellanox produces the world’s fastest network adapters, switches, cables, and management software, which can be integrated to create an end-to-end non-blocking high-performance Ethernet solution featuring excellent flexibility and the 25 and 100Gb/s speeds that iFLYTEK required.

Mellanox’s Spectrum™ family of switches offers a wide range of flexible options toward construction of a data center network. With switches that provide 12 to 64 ports, and bit rates ranging from 10 to 100Gb/s, network administrators can design and deploy cost-effective solutions that fit the specific needs of their data center. Moreover, by

enabling Open Ethernet, Mellanox separates the choice of hardware from the software that runs on top, allowing admins to perfectly match the operating system to the application requirements of their cluster.

Furthermore, Mellanox integrates Software-Defined Networking (SDN) into its Ethernet solutions, enabling easy, efficient management by empowering the infrastructure with automated scheduling and allows simplified expansion from small networks to massive data centers.

Tailored for Machine Learning and High-Performance Computing applications, Spectrum offers the highest available bandwidth and ultra-low latency with zero packet loss in all kinds of traffic models, regardless of the complexity and size of the network. Spectrum’s unique ASIC design reduces its power consumption intelligently, even when it functions at 100% performance. The result is that Mellanox provides the most efficient solution for all applications, both in performance and in operating expense.

Ideal Solution for Machine Learning Network

When iFLYTEK sought a partner to enable its Machine learning data center, Mellanox was the logical choice. Mellanox’s Open Ethernet 25/100GbE Spectrum switches enabled iFLYTEK to build a high-performance architecture that is

“Mellanox’s solution has enabled iFLYTEK to build a next-generation Machine Learning center that will accelerate our application performance and provide for our future needs.”

*- Dr. Zhiguo Wang
Executive Vice President,
iFLYTEK Research Institute*

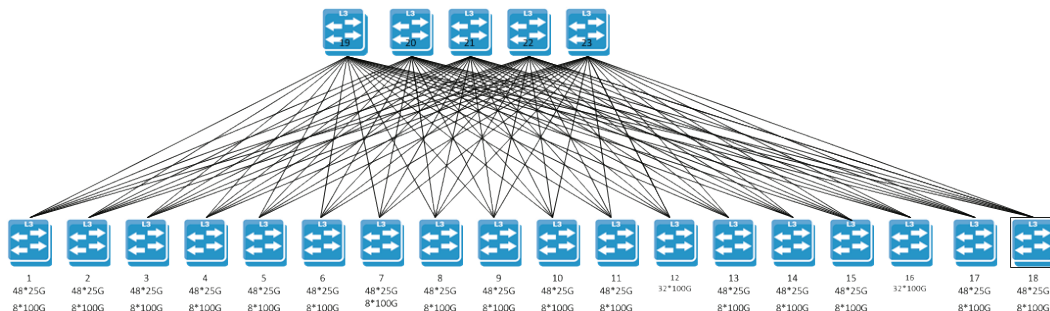


Figure 1. Example of Mellanox Spectrum Ethernet Switch-Based Leaf-Spine Topology Architecture

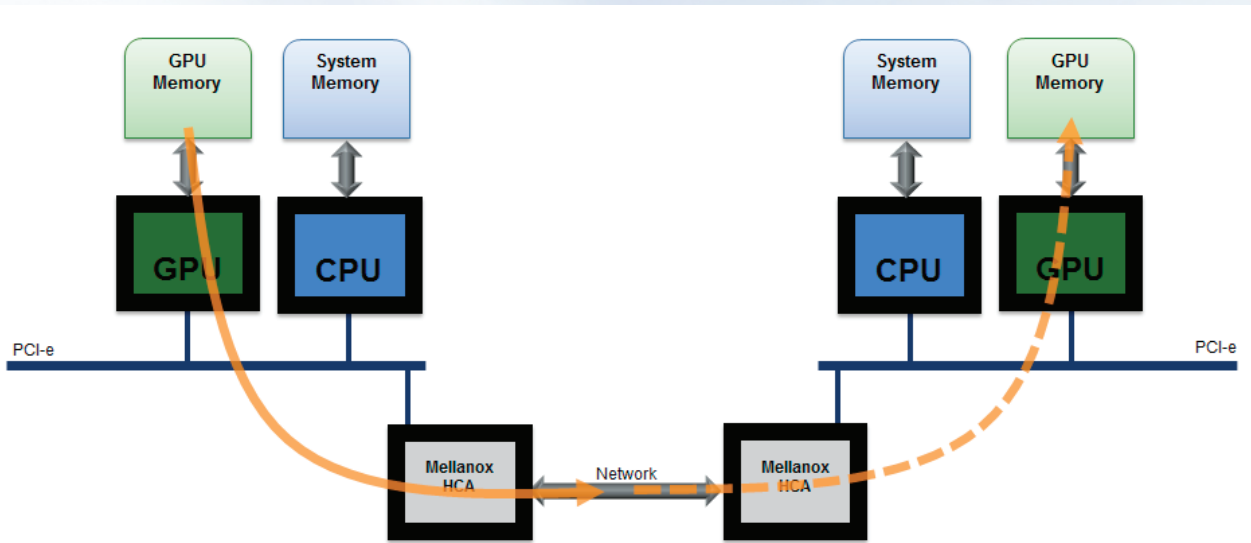


Figure 2. GPUDirect RDMA Eliminates the Bottlenecks of System Memory and CPU Capacity

flexible, scalable and easily managed. The iFLYTEK machine learning systems can handle massive concurrent traffic and are futureproofed to support unpredictable business growth. Mellanox’s intelligent networking solutions provide iFLYTEK with automatic network provisioning and management, as well as a convergence network infrastructure with Quality of Service and RDMA over Ethernet (RoCE).

Even better, Mellanox employs an offload architecture that reduces strain on the Central Processing Unit (CPU), improving the computing power and overall system efficiency as the CPU concentrates on its intended tasks and not on networking.

Mellanox could also provide GPUDirect® RDMA technology, which enables peer-to-peer communication among Graphic Processing Units (GPUs) within the network by creating a shared buffer and eliminating redundant copies to host memory. GPUs are frequently the processors of choice for Machine Learning networks, as the architecture of a GPU, as opposed to that of a CPU, is well-suited to the algorithms that are the building blocks of Machine Learning platforms. With GPUDirect RDMA, Mellanox was able to improve transmission efficiency between remote GPU nodes in the data center.

This mechanism worked especially well in iFLYTEK’s speech recognition projects, helping iFLYTEK achieve an amazing 97% speech recognition rate.

“Mellanox’s solution has enabled iFLYTEK to build a next-generation Machine Learning center that will accelerate our application performance and provide for our future needs,” said Dr. Zhiguo Wang, executive vice president of iFLYTEK Research Institute. “Moreover, we have leveraged the scalability of Mellanox Ethernet solutions to grow our compute

and storage needs in the most efficient manner.”

Summary

Mellanox has long been known as the world leader for interconnect technologies to enable the fastest and most powerful supercomputers, and now those same principles are being applied to today’s Machine Learning data centers. With high speed, low-latency Ethernet that that capitalized on Mellanox’s offload architecture and GPUDirect RDMA-enabled infrastructure, was compatible with the company’s current infrastructure, and is scalable for future computing and storage requirements, Mellanox accelerated iFLYTEK’s Machine Learning network and laid the foundation for iFLYTEK to grow its next generation data center.

About iFLYTEK

IFLYTEK CO., LTD (IFLYTEK) is a Chinese key software enterprise dedicated to the research of intelligent speech and language technologies, development of software and chip products, provision of speech information services, and integration of E-government systems. iFLYTEK’s intelligent speech technology, the core technology of the company, represents the top level of such products in the world. For more information, please visit <http://www.iflytek.com/en/>

About Mellanox

Mellanox Technologies is a leading supplier of end-to-end Ethernet interconnect solutions and services for enterprise and Machine Learning data centers, Web 2.0, cloud, storage and financial services. More information is available at www.mellanox.com.



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