SOLUTION BRIEF

上海储迅信息技术有限公司
Shanghai StorSwift Information Technology Co., Ltd.

Storswift and Mellanox HPC Joint Storage Solution

BACKGROUND

Achievements in High Performance Computing (HPC), a cutting-edge computer application technology, have significantly impacted application performance in several important areas, including scientific research, engineering construction and military technology, and more.

Increasingly scalable parallel computing is required to address the growing needs of these high performance computing applications. This together with the performance improvements in the CPU, memory and interconnection networks introduces new challenges to I/O performance, extensibility and high availability of storage systems.

According to G. M. Amdahl's rule of thumb, a balanced computer system needs 1 Mbit/sec of I/O bandwidth for each 1 MIPS of calculation performance improvement. As of today, single core floating-point computing power of the mainstream Haswell CPU has reached as high as hundreds of GFLOPS, while the corresponding demand for I/O bandwidth has also reached a level of 100Gb/s. In addition, hardware processing performance is improving at an exponential rate.

As per Moore's law, new instruction sets and heterogeneous accelerator cards, such as general-purpose computing on graphics processing units (GPGPUs) have arrived on the market, accelerating the traditional HPC application model compute performance by several orders of magnitude. However, in contrast, storage has not improved at the same rate, and its bottleneck effect is becoming more and more prominent. Thus, storage systems are a major factor affecting cluster performance.

The I/O model associated with HPC applications is characterized by high concurrency and high throughput. The core requirements for storage include:

1. Allowing computing tasks to be allocated to any cluster node for parallel processing; the storage platform should be able to provide storage systems with a unified view through the same namespace, so that all nodes in the cluster can access, read and write the application data in the computing process, in parallel.
2. Enabling the parallel access to shared data sets by thousands of high performance computing nodes in clusters.
3. In computing, even the average throughput of a single node is only a few hundreds of MB/s; however, due to the high concurrency of the cluster I/O, the I/O throughput can reach hundreds or even thousands of GB/s, requiring extremely high throughput of the storage system. Moreover, the performance requirement index of the aggregate bandwidth is proportional to the cluster scale.
4. With the increase of computing tasks, the cluster scale expands continuously, while capacity, throughput and I/O bandwidth of the storage should be able to expand dynamically. This is an important factor in ensuring the extensibility of the cluster system.

KEY BENEFITS

- Petabyte (PB)-level high performance storage with easy online expansion
- Supports InfiniBand network and RDMA mode perfectly
- Aggregate bandwidth of 5 nodes is over 8GB/s
- All servers can run complex sequencing tasks concurrently
- Provides a disaster tolerance solution for massive genetic data

"Storswift is the main contributor to the mainstream distributed storage RDMA code. It has joined with Mellanox to build the most stable InfiniBand RDMA storage network. In the process of deployment and application, the solution can quickly solve the problems encountered, while its technical support has also won the high appreciation of the industry users."

Liu Zheng, GM of Storswift

In the process of building an HPC computing cluster, the distributed file system meets the demands of high performance computing, with low cost input. With its good extensibility and high cost performance, the system becomes the preferred structure of HPC storage, and is the core technology for alleviating the HPC cluster I/O bottleneck.

STORSWIFT & MELLANOX JOINT SOLUTION

The CX-CLOUD series distributed storage software introduced by Shanghai Storswift Information Technology Ltd. (Storswift) has made an omni-directional optimization for high performance computing running environments. This optimized solution is characterized by:

- Full support for high-speed Ethernet and RDMA networks
- Enterprise-level characteristics, such as ultra-high performance, online scaling-out, Petabyte (PB) level data-volume, high availability, multiple data redundancy mechanisms, complex authority management, ease-of-use and maintenance
- Wide adoption in several of high performance computing areas, such as gene sequencing, geophysical prospecting for petroleum, meteorological computation, material innovation, environmental engineering, weapon research, geological prospecting and engineering calculation; also provides a robust IT infrastructure and supporting environment for scientific research, government decision-making and enterprise innovation

Storswift's network-based extensible storage solution provides sustained and reliable performance and simplifies network management. Mellanox EDR 100Gb/s InfiniBand switch and ConnectX-5 VPI EDR adapter provide the network transmission performance of a single 100Gb/s full wire-speed port with no packet loss, and do not require complex configuration of the operation mode nor link speed of each port. The InfiniBand network is easy to deploy and manage, which can greatly reduce network complexity, and has a significant advantage in the HPC system solutions on which extensibility and high reliability are required.

RDMA TECHNICAL ADVANTAGE

By applying the remote direct memory access (RDMA) technology of EDR InfiniBand networks, the CPU is not required for the I/O task, freeing this computing resource speeds up application performance and improves transmission efficiency of the data. RDMA enables network adapters to access application data directly, thereby bypassing the kernel, CPU and protocol stacks, and enables the CPU to perform more valuable core application tasks during I/O transmission. This serves to greatly improve the performance of servers and enables the workload processing capability of applications to extend efficiently in high bandwidth networks.

SUCCESSFUL CASE

A leading Chinese gene computation company employing Storswift distributed file storage and the Mellanox InfiniBand RDMA scheme, now uses only five nodes to obtain more than 8GB/s of continuous reading and writing performance, which fully satisfies the high I/O bandwidth requirements of lengthy concurrent runs in large-scale gene sequencing (see Figure 1).

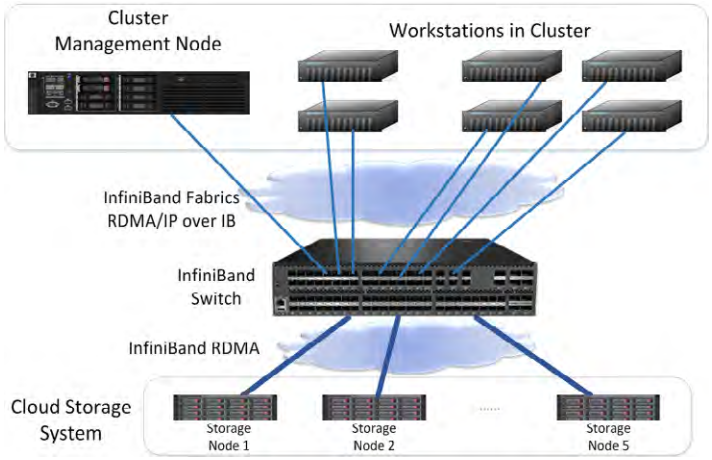

Figure 1: Storswift and Mellanox joint solution employed in a leading gene computation company

Weaknesses in the original storage platform:

- Storage file system: Lustre
- Capacity is exhausted
- Bottlenecks in the performance
- Fewer than 8 sequencing tasks running concurrently

Advantages of the CX-CLOUD-FS distributed storage system:

- 1PB capacity was deployed in the first term; supported online extension
- Supports InfiniBand network and RDMA mode perfectly
- Aggregate bandwidth of 5 nodes is over 8GB/s
- All servers can run complex sequencing tasks concurrently
- Provides a disaster tolerance solution for massive genetic data

User technical benefits:

- Special read-write optimization of unstructured data
- Ultra high performance, RDMA optimization, network read-write bandwidth greater than10GB/s
- Fully symmetric architecture with high stability and PB-level data volume
- Easy extension, online heterogeneous expansion, and horizontal extension, with nearly linear performance improvement
- Multiple data guarantees such as multiple copies, erasure codes, fault isolation
- Easy use and maintenance; centralized management of mass equipment
- With continuous updating and maintenance, it is a perfect substitute for Lustre.

CONCLUSION

Storswift and Mellanox provide a much-needed distributed storage solution for HPC business cluster, which greatly improves the parallel access performance of cluster storage, simplifies the management and maintenance, and enables high performance computing companies to focus on providing applications and services and helping their customers to develop business.

This solution has the following advantages: ultra high performance, convenient management, online horizontal linear extension and incomparable reliability, making it the perfect substitute for Lustre.

About Mellanox

Mellanox Technologies (NASDAQ:MLNX) is a leading provider of end to end Ethernet and InfiniBand intelligent interconnection solutions and services, which is server, storage, and hyper-converged infrastructure oriented. The Mellanox intelligent interconnection solution can provide the highest throughput and the lowest delay, faster transmission of data to the application and full system performance, thus to improve the efficiency of the data center. Mellanox provides a choice for high performance solutions: network and multi-core processor, network adapter, switch, cable, software and chip. They can speed up the application running speed and maximize the business results for a wide range of markets (including high performance computing, enterprise data center, Web 2.0, cloud, storage, network security, telecommunication and financial services).
More information is available at www.mellanox.com

About Storswift

Shanghai Storswift Information Technology Ltd is a high-tech company that provides high performance network storage products and related services. The founding team has more than ten years of experience in the storage industry, and has rich experience in large-scale storage research & development and marketing. The company has deep technology accumulation in the distributed storage field. With the leading technical architecture and product advantages, it has created high performance, high reliability and easy maintenance enterprise level storage products and are widely used in HPC, AI, Telecom operators, radio and television, medical, security and other industries. Especially in HPC industry, the distributed file storage products of Storswift provide full support for the RDMA model, have better cost performance while having ultra high performance, and meanwhile provide flexible data guarantee measures, product architecture which is easy to manage and easy to maintain, full technical support covering application scheme and underlying optimization, and provide the most comprehensive protection plan for the stable and efficient operation of HPC computing tasks.
More information is avaialble at www.storswift.com

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

60181SB
Rev1.0