

CASE STUDY

USE INTEL® ENTERPRISE EDITION FOR LUSTRE SOFTWARE-BASED COST-EFFECTIVE STORAGE SOLUTION TO SPEED UP EXPLORATION OF THE VALUE OF GENETIC INFORMATION

To discover the inherent law of life phenomena using genetic information can not only help mankind prevent diseases, but also lay the cornerstone for humans to continuously inquire into the secret of life. BGI, as the world's largest genomics R&D institution, has deployed the Supermicro open software-defined storage solution based on Intel® Enterprise Edition for Lustre software, to address the challenge to the storage system resulting from the average daily genetic data growth at 10TB and the demand for continuous improvement of data r/w performance. This has significant improvement in data throughput performance at lower costs, and accelerated the process of mining the value of gene sequencing data, and will meet the growing demand for data storage in the future.



Case Study
October 2016

Super Micro Computer, Inc.
980 Rock Avenue
San Jose, CA 95131 USA
www.supermicro.com



Use Intel Enterprise Edition for Lustre software-based cost-effective storage solution to speed up exploration of the value of genetic information

Gene sequencing data archiving and storage clusters based on Intel® Enterprise Edition for Lustre software have improved data throughput performance and accelerated value mining and insights into genetic information.

“With outstanding performance and enhanced scalability delivered by Lustre storage system, BGI has upgraded data throughput performance of its genetic information archiving up to 8GiB/s. This highly cost-effective storage solution has satisfied the expected requirements of genetic information research on gene sequencing data r/w performance, and the open storage architecture has allowed us to cope with the growing demand for storage expansion and performance extension at lower costs.

Moreover, Intel® Enterprise Edition for Lustre software has strengthened stability and reliability of the Lustre storage system, simplified system installation, configuration and monitoring, and reduced management and maintenance costs of the storage system.”

Lin Fang
Duty Director of BGI Research
Director of IT Center

CHALLENGE

- **Improvement of data r/w performance:** The existing storage system with a data throughput capacity of 400 MiB/s can no longer meet the requirement of BGI's high performance computing clusters for scientific research of genetic information on data throughput performance.
- **Reduction of TCO of the storage system:** In face of average daily growth of gene sequencing data at 10TB, BGI needs a storage solution featuring lower costs and higher storage density.
- **Enhancement of system scalability:** BGI has to strengthen flexibility of its storage system to proactively respond to future challenges to capacity and performance introduced by increasingly complex genetic research.

SOLUTION

- **Intel® Enterprise Edition for Lustre software-based storage solution:** BGI has deployed for its data center in Wuhan the solution using Supermicro servers based on Intel® Xeon® Processor E5-2650 v3, with 2 Supermicro 6028R-E1CR12N servers for Lustre metadata storage (MDS), 1 Supermicro CSE-216BE26-R920LPB server equipped with 15*600GB hard drive and 2*200GB solid state disks for Lustre metadata targets (MDT), 4 Supermicro 6028R-E1CR12N servers for object storage servers (OSS), 4 Supermicro CSE-946ED-R2KJBOD servers equipped with 74x 6TB disks and OpenZFS file system for object storage targets (OST). This is combined with configuration parameter tuning to achieve 8GiB/s of data throughput performance and reach the expected target of BGI.
- **Intel® Enterprise Edition for Lustre software:** Deployment of Intel® Enterprise Edition for Lustre software has improved stability and reliability of the storage system, and simplified system installation, configuration, monitoring and management.



Use Intel Enterprise Edition for Lustre software-based cost-effective storage solution to speed up exploration of the value of genetic information

LESSONS LEARNED

- The Lustre storage system has been developed and optimized for processing scientific research data, and will add to performance of mass storage devices, which makes it the ideal choice for gene sequencing data research. Migration to the Lustre-based storage solution has realized 20X improvement in storage performance.
- Intel® Xeon® Processor E5-2650 v3 provides full support for computing performance of the Lustre storage system. When the object storage server (OSS) is involved in 96 data r/w operations, the occupancy rate of the processor is about 40%, and that of memory about 30%.¹
- The Lustre storage system with flexible scalability was adopted to meet the growing business demand for data capacity and performance, and Intel® EE for Lustre software was deployed to improve stability and reliability of the storage system.
- The storage system, based on an open architecture platform and the OpenZFS file system independent of hardware RAID, were used to deal with the challenge from increasing data volume and continuous improvement in data throughput performance with lower costs.

IMPACT

- **Acceleration of genetic information research to benefit mankind:** With improvement in data throughput performance, BGI can quicken its data supply for genetic information research, meet the demand of more complex scientific research, improve the research level, and make better use of gene sequencing data.
- **Decrease of the cost of genetic data archiving:** Open infrastructure architecture has helped BGI reduce the costs of the storage system and its capacity expansion, and a higher storage density has contributed to lower costs of physical space occupation and power consumption.
- **Reduction of storage management and maintenance costs:** Intel® Enterprise Edition for Lustre software has enhanced storage stability and reliability, and meanwhile provided simple but powerful management tools, significantly reducing management and maintenance costs of the storage system.
- **Enhancement of scalability:** Based on scalability of the Lustre system, BGI can flexibly perform data storage capacity expansion and performance extension caused by genetic information research.

BGI, founded in 1999, has set up large bioinformatics supercomputing centers in Shenzhen, Hong Kong, Beijing, Wuhan and Hangzhou, with overall peak computing capacity up to 288.5T ops, memory capacity up to 67.22 TB and storage capacity up to 35.09 PB. Gene sequencing identifies through large-scale computing analytics genes organisms carry and sequences of these genes from vast amounts of data, and ultimately obtain the genetic information. Continuous research on gene sequencing data using latest technologies will bring revolutionary changes to prediction and prevention of diseases, accurate diagnosis and treatment and drug R&D.

DATA THROUGHPUT PERFORMANCE AS A HANDICAP TO GENETIC RESEARCH

For gene sequencing data archiving, BGI used a proprietary storage system with 400MiB/s data throughput performance. In this system, data were first stored and then made available to several projects or laboratories for life science research. BGI Lin Fang said, "Massive, complex and changeable data computing has been the major issue in the way of genetic research, and application and rapid development of high performance computing clusters have played an important role in development of gene sequencing.

However, limited by data throughput capacity, the existing data storage system with insufficient data r/w performance has been unable to meet the requirements of relevant research, and provide high performance clusters with intime data



Use Intel Enterprise Edition for Lustre software-based cost-effective storage solution to speed up exploration of the value of genetic information

- ‘Just a bunch of disks’ (JBOD) has helped increase the storage density and reduce the physical space, which together with tuning of configuration parameters (config), has optimized data throughput performance of the storage system.
- Intel® EE for Lustre software with built-in intuitive and powerful management tools has simplified installation, configuration, monitoring and management of the Lustre storage system, and enjoys strong and complete technical support from Intel.

supply. Consequently, performance of the storage system has not only affected the research efficiency and but also become a bottleneck for further development of genetic research and technology.”

CHALLENGE TO STORAGE COSTS CAUSED BY EXPLOSION OF DATA

In addition, along with business expansion, BGI has increased the quantity of its gene sequencing machines from the initial few to the current several hundred, producing around 10TB of new data every day. “Historical experience shows that as genetic research technology advances, biological data for genetic research to be stored increases tenfold every 12 to 18 months.” Lin Fang continued, “Restricted by the expensive special storage system and faced with extremely high data storage and capacity expansion costs, we have to evolve the current storage system and reduce storage costs using latest technology.”

Therefore, BGI is looking for a highly cost-effective storage system to meet the demand of current and future genetic research for performance and costs. After performance tests, BGI deployed cluster storage solution based on Intel® EE for Lustre software in its Wuhan Data Center for permanently archiving gene sequencing.

“As a leader in server and storage innovation, this integrated SDS solution for Lustre, based on Intel® Enterprise Edition for Lustre software has maximized price/performance, density, capacity and reliability of the storage system, and significantly reduced the costs of capacity expansion and operation. By eliminating the need for legacy proprietary storage products with open industry standard Supermicro Systems, the solution accelerates innovation by enabling the use of leading edge systems such as the 4U 90 Bay JBOD. Our total solution provides the complete hardware, software and services to rapidly and efficiently deploy and maintain a complete Lustre system.”

Dr. Tau Leng
Vice President of HPC and Marketing
Supermicro

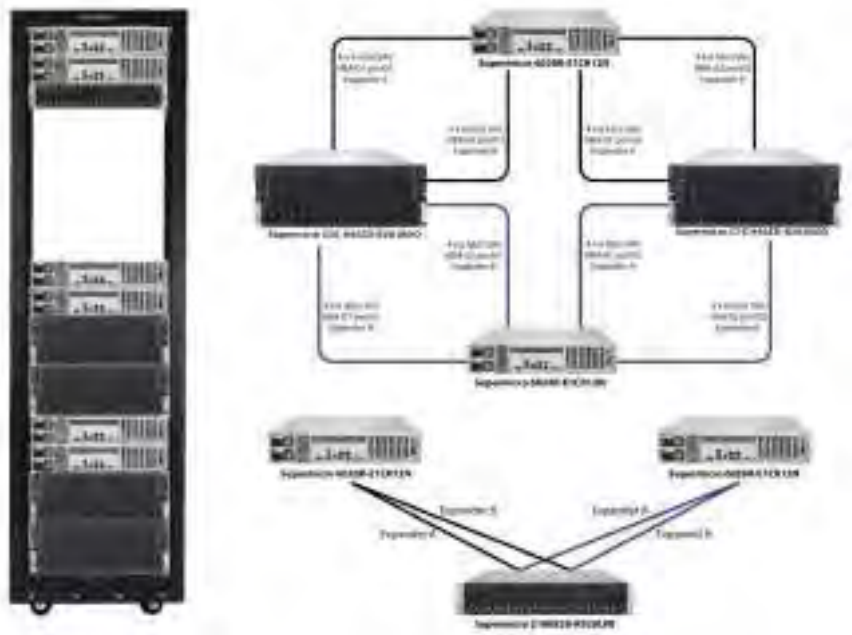


Figure 1. Storage Architecture of BGI’s Archived Gene Sequencing Data

HIGHLY COST-EFFECTIVE SDS SOLUTION

Intel® Enterprise Edition for Lustre software, Supermicro servers and storage



Use Intel Enterprise Edition for Lustre software-based cost-effective storage solution to speed up exploration of the value of genetic information

FOR MORE INFORMATION

Supermicro® SuperServer®
SYS-6028R-E1CR12N Datasheet

Supermicro® SuperServer®
CSE-946ED-R2KJBOD Datasheet

Supermicro® SuperServer®
216BE26-R920LPB Datasheet

Intel® Xeon® Processor E5-2600 v4 Product
Family

www.intel.com/content/www/us/en/processors/xeon/xeon-e5-solutions.html

Intel® Lustre
www.intel.com/lustre

clusters based on Intel® Xeon® Processor E5-2650 v3 product family were integrated to deliver for BGI a highly reliable, cost-effective software-defined storage (SDS) solution.

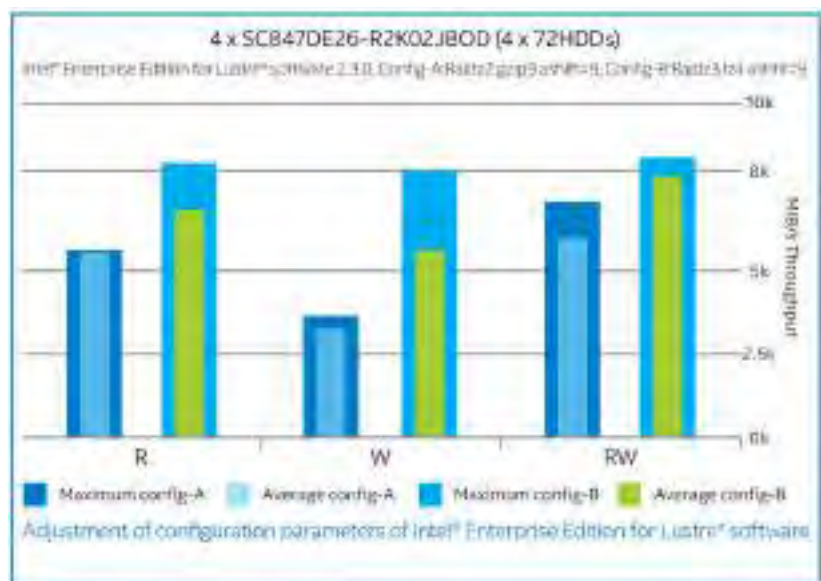
20X IMPROVEMENT IN DATA THROUGHPUT PERFORMANCE ACCELERATING GENETIC INFORMATION RESEARCH

With the help of Intel, BGI has optimized the storage system by adjusting parameters of Intel® Enterprise Edition for Lustre software and eventually upgraded the data throughput performance to 8GiB/s.

BGI Lin Fang said, "The new storage system offers 20X improvement in data r/w capacity with lower costs and supplies data to high-performance computing clusters for scientific research at a faster speed, making full use of computing resources and obviously raising efficiency of genetic research. More importantly,

Intel® EE for Lustre software has enhanced the system stability and manageability, and facilitated system installation, deployment, monitoring and maintenance. All this has given us full confidence in managing to meet the demand for continuous performance and capacity upgrades caused by increasingly complex genetic information research."

Looking into the future, BGI will continue to work with Intel to further explore their cooperation on worms and directory-based disk quota under the Lustre storage system, to continuously promote advance in the study on the mysteries of life and enable genetic information to benefit mankind.



About Super Micro Computer, Inc.

Supermicro® (NASDAQ: SMCI), the leading innovator in high-performance, high-efficiency server technology is a premier provider of advanced server Building Block Solutions® for Data Center, Cloud Computing, Enterprise IT, Hadoop/Big Data, HPC and Embedded Systems worldwide. Supermicro is committed to protecting the environment through its “We Keep IT Green®” initiative and provides customers with the most energy-efficient, environmentally-friendly solutions available on the market.

www.supermicro.com

¹ Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.

This document contains information on products, services and/or processes in development. All information provided here is subject to change without notice.

The products and services described may contain defects or errors known as errata which may cause deviations from published specifications. Current characterized errata are available on request.

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer or learn more at www.intel.com.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.

All performance tests were performed and are being reported by BGI. Please contact BGI for more information on any performance test reported here.

Cost reduction scenarios described are intended as examples of how a given Intel-based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Supermicro does not guarantee any costs or cost reduction.

No part of this document covered by copyright may be reproduced in any form or by any means — graphic, electronic, or mechanical, including photocopying, recording, taping, or storage in an electronic retrieval system — without prior written permission of the copyright owner.

Supermicro, the Supermicro logo, Building Block Solutions, We Keep IT Green, SuperServer, TwinPro™, TwinPro²™, SuperDoctor are trademarks and/or registered trademarks of Super Micro Computer, Inc.

Ultrabook, Celeron, Celeron Inside, Core Inside, Intel, Intel Logo, Intel Atom, Intel Atom Inside, Intel Core, Intel Inside, Intel Inside Logo, Intel vPro, Itanium, Itanium Inside, Pentium, Pentium Inside, vPro Inside, Xeon, Xeon Phi, and Xeon Inside are trademarks of Intel Corporation in the U.S. and/or other countries.

© Copyright 2016 Super Micro Computer, Inc. All rights reserved.

Printed in USA

MKT-0006-10/2016-12-R1

 Please Recycle

