

Accelerate the Exploration of the Value of Genomic Data with Intel® QuickAssist Technology



“According to the test, PCIe*-based Intel® QuickAssist Adapter DH8950 showed surprisingly powerful data compression ability, reduced compression time by up to 2.7 times and server processor occupancy rate from 78 to 19% compared with traditional software compression practices, when 32 compression processes were running simultaneously with the same compression ratio. What’s more, by using the plug-and-play technology, there is no need to modify the application code before using the adapter. With these significant advantages, BGI could not only enhance the storage density of genetic data, reduce storage costs, but also reduce the bandwidth occupancy of data transmission, and at the same time make full use of processor computing resources for genomic researchers on life science.”

Fang Lin

**Vice President/ Director of IT Center
BGI Shenzhen**

Contents

Challenges:	1
Solutions:	1
Impacts:.....	1
Reduce data compression time	2
Reduce processor occupancy in data compression	2

The storage and transmission of massive sequencing data are becoming a formidable challenge facing the IT infrastructure of genome analysis system where data compression is a crucial technology to solve this problem. However, as a typical computing-intensive workload, data compression requires massive computing resources from the processors. BGI*, as one of the world’s largest genomics research institutes, has worked closely with Intel in testing hardware-based data compression solutions with Intel® QuickAssist Technology to increase storage density, reduce storage costs, optimize the overall performance of genome analysis system and accelerate the innovation from sequencing data.

Challenges:

- **Increase data compression efficiency:** Compress data with acceleration technology; reduce data compression time and occupancy of server resources.
- **Reduce storage system cost:** In the face of a 10 Tb genetic data growth on a daily basis, BGI needs to increase data compression ratio, improve storage density and reduce data storage costs.
- **Optimize overall system performance:** BGI needs to optimize different data storage solutions (e.g., high-performance processing of genetic data, archiving and the database) and save disk input/output and network bandwidth to optimize the system’s overall performance.

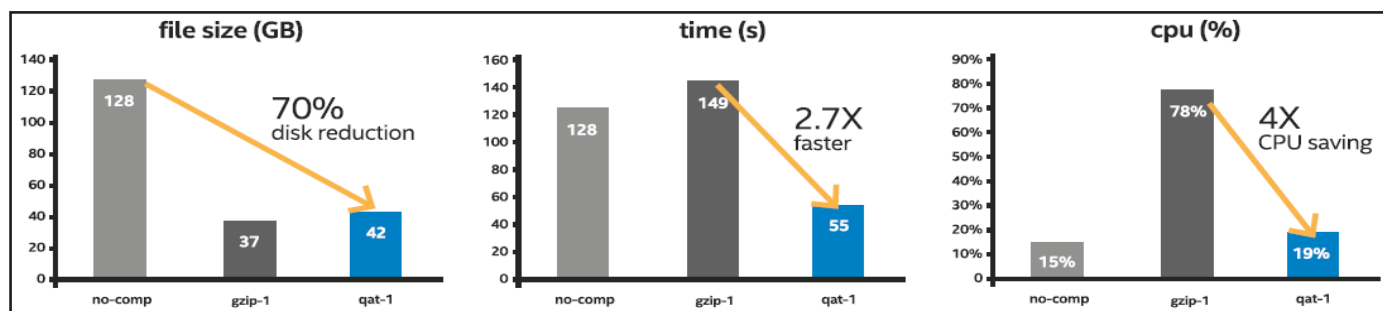
Solutions:

- **Intel® QuickAssist Adapter DH8950 Series:** BGI and Intel jointly tested the hardware-based data compression solution. Equipped with two Intel® Xeon® Processor E5-2620 v3, three 800 GB Intel® SSD DC S3500 and 32 GB memory, and one PCIe*-based Intel® QuickAssist Adapter DH8950 plug-in card, the benchmark program writes 128 GB raw data with 16, 24, and 32 processes respectively to the SSDs. The companies monitored the data compression ratio, compression time, and processor utilization when running the benchmark program.
- **ZFS Mainstream:** Z File System (ZFS) on Linux *(ZOL*) has supported Intel® QuickAssist since v0.7.0. Once the latest ZFS filesystem is deployed underneath, data written to the disk can be compressed by Intel® QuickAssist hardware transparently, and decompressed by Intel® QuickAssist when the data is read from disk.
- **Overall system performance is enhanced:** When read or written into the disk by the application system, genetic data is compressed or decompressed in a transparent and automated manner so that more data can be provided for computing to meet more complicated needs of genetic information research and improve research efficiency and quality.

Impacts:

- **1Data compression capacity is significantly improved:** With a 70% data compression ratio, compression time is reduced by 2.7 times, from 149s to 55s. CPU utilization decreased from 78% to 19%.
- **1Data storage cost is reduced:** By using Intel® QuickAssist Adapter DH8950 for hardware-based data compression, BGI can store more data with the same disk space and reduce storage costs. Meanwhile, higher storage density could reduce the cost of physical space footprint and power consumption.

Using Intel® QuickAssist Adapter DH8950 series for hardware-based data compression, free computing resources, improve data compression efficiency and overall system performance to accelerate scientific research on genetics.



Comparison of data compression performance based on Intel® Quick Assist Technology

BGI, a Chinese Company established in 1999, and after 17 years' development can now be found in over 60 countries and regions around the world. Up till now, its total data storage capacity has reached 35.09 PB. According to Fang Lin, "At a time when biological information explodes, and genetic research technology advances continuously, the demand for storage and computing grows by ten times every 12 to 18 months. BGI's computing systems are facing formidable challenges from data storage."

Reduce data compression time

As a result, BGI decided to test data compression technology. Data were compressed before they were written into the disk so that more data could be stored in the same disk space. BGI used a server equipped with two Intel® Xeon® Processor E5-2620 v3 series and tested 128 GB data with different numbers of concurrent processes and different software compression algorithms.

The test showed that, on average, it took 149 seconds to compress 128GB data when GZIP software was used, and a 70% compression ratio was achieved. "Although 70% storage space was saved, it took as long as 2.5 minutes. We hope the data could be compressed within 1 minute to achieve the best combination of processors, network bandwidth, and disk capacity and make the best use of computing resources," said Fang."

Reduce processor occupancy in data compression

Moreover, since data compression is a typically computing-intensive workload, the server CPU's utilization reaches 78% in BGI's testing environment. It isn't worthwhile to exchange space and cost with such data compression performance which is not suitable to BGI's production environment. Lin said,

"It's hard to strike a balance between storage cost savings and the returns on genetic research if the precious computing resources that should be used for scientific research were applied to data compression."

Based on investigation and research, BGI shifted its attention to hardware-based data compression technology, and finally, with the help of Intel, completed the data compression test based on the Intel® QuickAssist Adapter DH8950 series and achieved satisfactory results.

Intel provided Intel® QuickAssist Adapter DH8950 series and drivers for the test, which can be used on Intel® architecture-based servers with PCIe* slots. Intel also offered the ZFS file system patches based on this hardware compression scheme, where BGI can achieve seamless integration by taking full advantage of hardware acceleration without modifying the applications.

Data compression performance is improved significantly

BGI is exceptionally satisfied with the test results. Results are showing a 70% compression ratio and 32 processes running at the same time, the average compression time of 128 GB of raw data was reduced by 2.7 times, from 149 seconds to 55 seconds, meeting BGI's expectations. Besides, according to the monitoring, CPU utilization decreases from 78 to 19%.

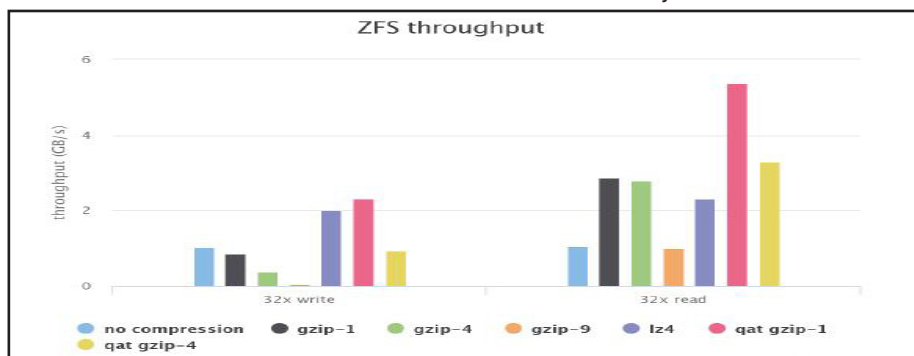
"Through the test, we could see the improvements in performance and other benefits brought by hardware-based adapter's automatic compression of data rather than modification of existing application codes. Obviously, this technology could be applied to various fields in genomic research to address data storage challenges." Fang Lin added.

Data throughput capacity promotes research on Life Science

Furthermore, system's data throughput capacity was significantly improved, both in data reading, and writing.

The 32-process synchronous writing speed increased from 0.9 GB/s to 2.3 GB/s, compared with a traditional software-based compression method. Data reading speed increased from 2.9 GB/s to 5.4 GB/s. Which means the data compression solution of the Intel® QuickAssist Adapter 8950 could not only boost storage capacity, reduce storage costs, but also help BGI further enhance the overall performance of the system, as writing and reading capacity are enhanced by 2.5 and 1.86 times respectively.

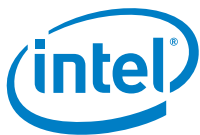
In the face of the future, BGI will continue to work with Intel for further explorations of cold storage systems, high-performance storage systems and progression of the research on the secret of life with genetic information to benefit humanity.



LESSONS LEARNED

- Intel® QuickAssist Technology focuses on data security and compression acceleration. It has different form factors such as chipset, PCIe* card, and System-on-a-Chip (SoC), supporting the overall performance upgrade of the data center.
- Hardware-based Intel® QuickAssist technology is implemented to accelerate the computing-intensive data compression, to reduce job run time, CPU utilization, storage space and costs, and optimize disk IO and network bandwidth utilization.
- Combine Intel® SSDs, Intel® QuickAssist Adapter with OpenZFS, Lustre*, and other file systems on Intel® architecture-based servers for high-performance storage solutions.
- Intel® processor with Intel® QuickAssist Adapter for low-cost, high-performance data archiving solutions.

Please contact your Intel representative or go to the Intel IT Center for solutions more suitable to your company.



¹For more complete information about performance and benchmark results, visit www.intel.com/benchmarks.

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. For more complete information visit www.intel.com/benchmarks.

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

Benchmark results were obtained prior to implementation of recent software patches and firmware updates intended to address exploits referred to as "Spectre" and "Meltdown". Implementation of these updates may make these results inapplicable to your device or system.

[Note the phrase "benchmark results" may be tweaked to say "Estimated results" when estimates are used]

[FTC disclaimer] 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. For more information go to www.intel.com/benchmarks.

Specific System Configuration and Workload Details

Any other disclaimer we would normally include, including a disclaimer if we are using estimated results.

Test configuration: CPU: Intel® Xeon® Processor E5 v3 Family; 2.40 GHz; Cores and DIMMS 4GB 2133 MHz x8; Intel® QuickAssist Adapter 8950; Storage: Intel® SSD DC S3510 Series (800GB, 2.5in SATA 6Gb/s, 16nm, MLC); SW cat*; Kernel 2.6.32-642.3.1.el6.x86_64; ZOL 0.6.5.3-1; Intel® QuickAssist Technology driver 1.6

For more information go to <http://www.intel.com/performance>. Optimization Notice: Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel.

Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel micro architecture are reserved for Intel microprocessor. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice. Results of performance tests are from test report of BGI. For more information about performance and standards, please contact BGI.

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. Intel does not guarantee any costs or cost reduction.

*Other names and brands may be claimed as the property of others.

Intel, the Intel logo, and Xeon are trademarks of Intel Corporation in the U.S. and other countries

© 2018, Intel Corporation. All rights reserved..

Printed in USA

Please Recycle

Document Number: 336873-001