SOLUTION BRIEF

Intel® Select Solutions
Distributed Flash Database
2nd Generation Intel® Xeon® Scalable Processors
June 2020

Intel® Select Solutions for Transwarp ArgoDB

With the continuous implementation of digital innovation, data is increasingly valued by enterprises as one of their most important assets. As big data technology is increasingly being used in business, user demands have also become more complex, which is mainly reflected in scenarios such as the coexistence of offline and online services, analytical and retrieval services, as well as structured and non-structured data. It is increasingly difficult for traditional single big data architecture to satisfy these ever-increasing demands, and many types of architectures and products are needed. This directly leads to issues that urgently need to be resolved by business, such as a rapid increase in platform complexity, slower system response time, increased operation and management, and implementation costs.

All-flash storage architectures have played an increasingly important role in the design of database systems. When compared to traditional storage layer design, all-flash storage architectures are often able to provide higher data throughput and greatly reduce the latency of read and write operations. Furthermore, in database applications, all-flash storage enables processors to avoid wasting time waiting for IO operations, which allows full use of the capabilities of the processor and significantly improves service quality.

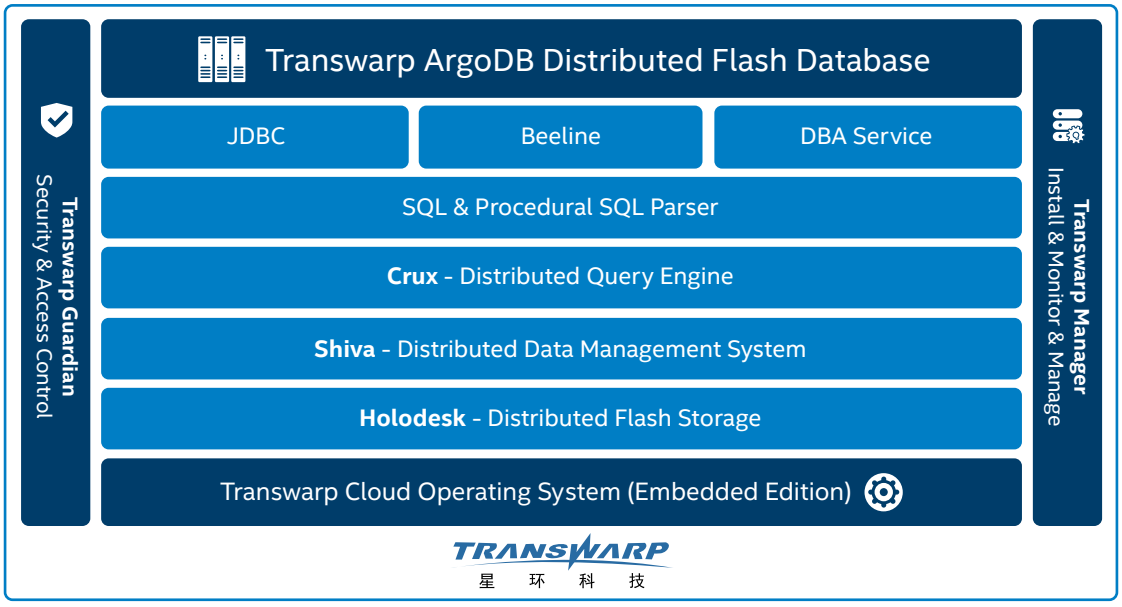
Transwarp and Intel have jointly released Intel Select Solutions for Transwarp ArgoDB based on the 2nd Generation Intel® Xeon® Scalable Processors, Intel® SSD DC P4510, Intel® SSD D3 S4510, Intel Network Adapter, and other products. These solutions are performance optimized and benchmarked for Transwarp ArgoDB workloads and provide a useful reference for the deployment of distributed all-flash architecture-based high-performance databases for users in the finance, government, logistics, transportation, and other industries.

Transwarp ArgoDB Helps Unleash the Value of Big Data

Transwarp ArgoDB, a distributed flash database released by Transwarp, is designed for all-flash servers and can replace multiple data analysis and processing hybrid architectures. It is a one-stop solution that can meet the various demands of enterprises for big data platforms and allows enterprises to use big data platforms more efficiently and to better explore and use the commercial value of big data.

Transwarp ArgoDB is designed with a refined storage structure and efficient algorithms for all-flash servers. Benefits from ArgoDB's outstanding features and performance, users only need a single product suite to meet various demands including offline batch data warehouses, real-time data analysis platforms, OLAP databases, full-text retrieval databases and others.

Transwarp ArgoDB has two core components: a distributed computing engine and a distributed storage engine. The ArgoDB distributed computing engine is a high-performance engine designed for data marts and real-time data warehousing. Moreover, a “purely vectorized” computing



engine is customized for column-oriented storage; it can read batch storage files quickly and respond to simple or complex queries of a small amount of data at high speed. The ArgoDB distributed storage engine has been deconstructed into two modules - a general distributed database service layer and an underlying storage engine. The underlying storage engine can be abstracted as a set of interfaces, where any storage engine that supports these interfaces can access ArgoDB via plug-ins.

Furthermore, Transwarp ArgoDB also includes several major management components such as the Transwarp Container Operating System (a custom-made container operating system for big data applications), the Transwarp Manager (graphical tool for the configuration, management, operation and maintenance of ArgoDB clusters), and the Transwarp Guardian (provides ArgoDB with integrated security and resource management services), which can be used to implement the agile management of databases.

In general, the new generation Transwarp ArgoDB database has the following outstanding features:

- **Comprehensive SQL support:** ArgoDB provides comprehensive support for SQL 2003 and also supports Oracle PL/SQL and DB2 SQL PL. To adapt to various database languages, ArgoDB also allows users to set database dialects. Currently, Oracle, DB2 and Teradata are well supported.
- **Support for distributed transactions:** A serializable distributed transaction algorithm is designed for data warehouse and data mart applications in ArgoDB, which provides transaction management guarantees without compromising analytical performance. Serializable transaction isolation is implemented in ArgoDB, while the Raft protocol ensures strong data consistency.
- **Real-time and batch data access:** ArgoDB supports real-time and batch data importing. Online services use real-time data access to ensure data timeliness; offline services such as data warehouses use batch data imports to achieve higher throughput.
- **Massive data OLAP and offline analysis:** ArgoDB supports offline analysis and high concurrency real-time/quasi-real-time data mart services simultaneously with PB-level data. When high performance flash disks are used as the storage medium, ArgoDB's storage and computing performance can be brought into full play, which results in even better performance advantages.
- **Support for RAM/flash/magnetic drive three-level hybrid storage:** ArgoDB supports three-level hybrid storage of RAM/flash/magnetic drive. The replica storage strategy for a table can be defined by the DBA; for example, a replica can be stored in flash memory, and two can be stored in magnetic drives. Multi-level storage allows users to find a better balance between performance and hardware costs.

Intel Select Solutions for Transwarp ArgoDB

Intel Select Solutions for Transwarp ArgoDB are solutions that include server hardware, database software, and related verification results. As a customized reference design for distributed flash-based database structures, it can help the customer's IT departments optimize the costs for deploying and verifying Transwarp ArgoDB. While meeting the various demands of Transwarp ArgoDB in performance, availability, scalability, etc., it also greatly reduces the evaluation time, and provides valuable reference for users in finance, government, telecommunications, logistics, and other industries.

This solution is based on the Transwarp ArgoDB flash database and uses 2nd Generation Intel® Xeon® Scalable Processors, Intel® SSD DC P4510, and Intel Ethernet Network Adapters. The solution has the following features:

- With a storage engine based on distributed consistency protocol and a pre-verified infrastructure platform, the compatibility and stability are fully verified. Faster online deployment is also achieved.
- Using the selected Intel Xeon Scalable Processors along with the exclusive Transwarp "purely vectorized" computing engine, the solution delivers powerful computing capabilities and can fully satisfy the computing requirements of offline batch data warehouses, real-time data analysis platforms, OLAP databases, full-text retrieval databases, and other workloads.
- The solution recommends the use of the Intel SSD DC P4510, which comes with the NVMe interface, as the high-performance storage layer, which can take the data throughput performance of all-flash servers to new heights. Coupled with Transwarp ArgoDB's special design for all-flash architecture, not only outstanding performance is achieved, hybrid loads are also effectively supported.

Hardware Selections

Intel Select Solutions for Transwarp ArgoDB come with the Intel high performance computing technical framework. The 2nd Generation Intel Xeon Scalable Processors, Intel SSD DC P4510 with NVMe interface, Intel SSD D3 S4510, and Intel Ethernet Converged Network Adapters are recommended hardware. Industry users may use the above as a basis to quickly deploy Transwarp ArgoDB distributed flash databases.

- **Intel SSD DC P4510:** The Intel SSD DC P4510 Series is based on Intel 3D NAND Technology and uses the NVMe interface to provide higher storage density and support a wider range of workload applications. The P4510 comes with enhanced smart firmware algorithm, error correction, power failure protection, and is designed to be durable. It provides better QoS and is able to guarantee the optimal balance in read and write protection for data storage.
- **Intel SSD D3 S4510:** This SSD can satisfy demanding service level requirements while improving server efficiency. Thanks to the innovative SATA firmware and the latest generation Intel 3D NAND, the Intel SSD D3-S4510 is compatible with existing SATA equipment, and allows storage upgrades to be implemented easily. This improves the efficiency of read-intensive workloads while maintaining infrastructure compatibility.
- **2nd Generation Intel Xeon Scalable Processor:** Fully optimized for data center workloads, the processor has industry-leading performance, implements platform-wide innovations across compute, networking, and storage, and provides enhanced hardware

What Are Intel Select Solutions?

Intel Select Solutions are pre-defined, workload-optimized solutions designed to minimize the challenges of infrastructure evaluation and deployment. Solutions are validated by OEMs/ODMs, certified by ISVs, and verified by Intel. Intel develops these solutions in extensive collaboration with hardware, software and operating system vendor partners and with the world's leading data center and service providers. Every Intel® Select Solution is a tailored combination of Intel data center compute, memory, storage, and network technologies that delivers predictable, trusted, and compelling performance.

To qualify as an Intel® Select Solution, solution providers must:

1. Meet the software and hardware stack requirements outlined by the solution's reference-design specifications.
2. Replicate or exceed established reference-benchmark test results.
3. Publish solution content to facilitate customer deployment.

Solution providers can develop their own optimizations in order to give end customers a simpler, more consistent deployment experience.

virtualization features. With 48 PCIe 3.0 channels, it is applicable for demanding I/O intensive workloads and helps accelerate the transformative impact of data. Furthermore, the resource allocation technology that comes with the processor provides guarantee for the flexible scheduling of various resources. For example, the latest resource allocation technology enables resource control and orchestration, so in cache management and memory bandwidth management, the orchestration feature can now be utilized to help provide the best service for specific applications.

- **Intel Ethernet Converged Network Adapter:** The Intel Ethernet Converged Adapter features specialized network functions and supports advanced features such as Virtual Machine Device Queue (VMDq), and Single-Root Input/Output Virtualization (SR-IOV). It also provides comprehensive compatibility, a wide range of product choices, performance and acceleration, as well as simple installation and great reliability.

Technology Selections

Besides the required hardware, the below technology can further improve the performance and reliability of the Select Solutions:

- **Intel 3D NAND Technology:** SSDs based on Intel 3D NAND Technology provide outstanding performance, QoS, larger capacity, and optimized storage efficiency. These SSDs can help data centers improve the efficiency of each flash-based server, minimize server disruptions while achieving effective large-scale management. In addition to accelerating the transitioning and scalability using proven manufacturing processes, Intel 3D NAND Technology also uses an architecture designed for higher capacity and optimal performance which further elevates the advantages of flash memory.
- **NVMe SSD:** Non-volatile memory express (NVM Express or NVMe) is a logical device interface. NMVe SSDs have features such as low latency and high efficiency, and can achieve up to 6 times the performance of SATA hard disks[1].
- **Integrated Intel QuickAssist Technology (Intel QAT):** Hardware acceleration based on chipsets can continuously increase compression and encryption workloads to achieve higher efficiency. At the same time, enhanced data transmission and protection are provided for the server, storage and network infrastructures.
- **Intel Ultra Path Interconnect (Intel UPI):** 4 Intel UPI (9200 series) and up to 3 Intel UPI (8200 series) channels extend the scalability of the platform to 2 channels (9200 series) and up to 8 channels (8200 series), to achieve a balance between high throughput and energy efficiency.

2nd Generation Intel Xeon Scalable Processors

2nd Generation Intel Xeon Scalable processors:

- Provide high scalability to corporate data centers
- Provide higher performance for virtual infrastructures compared to previous generation processors
- Implement outstanding resource utilization and agility
- Improve data and workload integrity for data center solutions and ensures regulatory compliance

Intel Select Solutions for Transwarp ArgoDB recommends using 2nd Generation Intel Xeon Scalable processors.

- **Intel Infrastructure Management Technology (Intel IMT):** A type of resource management framework which combines multiple Intel functionalities together and supports platform-level detection, reporting, and configuration. It can perform hardware-enhanced monitoring, management, and control of resources and helps in improving data center resource efficiency and utilization.

Verified Performance Through Benchmark Testing

All Intel Select Solutions are verified and performance benchmarked to meet the minimum level of workload-optimized performance. Intel Select Solutions for Transwarp ArgoDB are tested using TPC-DS as the benchmarking tool.

TPC-DS is a decision support benchmark for decision support systems and includes several generally applicable aspects of a decision support system, including queries and data maintenance. These decision support system performance benchmarks are useful for the purpose of comparing the Base and Plus configurations.

The Base and Plus configurations are shown in the table below. Test data shows that the total completion time for Intel Select Solutions for Transwarp ArgoDB Base configuration in the TPC-DS (10TB) test was 93000s, while the total completion time for the Plus configuration

was 17000s. The performance for the Plus configuration is as high as 5.4 times of the Base configuration.[2] To improve computing efficiency, users only need to expand the number of server nodes according to actual needs to achieve approximately linear performance expansion.

Benchmark	TPC-DS (10TB) Total time (s)
Intel Select Base Configuration	93000
Intel Select Plus Configuration	17000

Accelerating Data Revolution with Intel Select Solutions for Transwarp ArgoDB

Through in-depth technical collaboration between Transwarp and Intel, Intel Select Solutions for Transwarp ArgoDB provide high performance, high availability, and high scalability Transwarp ArgoDB as well as pre-tested and pre-verified all-flash server configuration recommendations. This brings a flexible and highly effective business data storage solution to business users and facilitates the establishment of comprehensive integrated one-stop big data platforms. The partnership between both parties has brought an innovative database solution to the industry which helps businesses to better discover the value of big data and accelerate the "data revolution".

Table: The Base and Plus Configurations for Intel Select Solutions for Transwarp ArgoDB

ArgoDB: 4 nodes	Base	Plus
Processor	2x Intel Xeon Gold 5218 processor @ 2.30GHz 16C or higher	2x Intel Xeon Gold 6240 processor @ 2.60GHz 18C or higher
Memory	256 GB or higher (8x 32 GB DDR4-2666)	384 GB or higher (8x 32 GB DDR4-2666)
Boot Drive	480 GB or larger Intel SSD D3 S4510	480 GB or larger Intel SSD D3 S4510
Data Drive	10x 2 TB SATA HDD 7200RPM (RAID) or higher	8 x 2TB Intel SSD DC P4510 or higher
Cache disk	2x 2TB Intel SSD DC P4510	NA
Data Network	1x 10 GB dual-port Intel Ethernet CNA X710-DA2 SFP+ or higher version	1x 25 GB dual-port Intel Ethernet CNA XXV710-DA2 SFP28 or higher version
Management Network	Integrated 1 GbE or higher version	Integrated 1 GbE or higher version
ArgoDB Version	ArgoDB 1.3.2	ArgoDB 1.3.2
Operating system	CentOS 7.5, Kernel 5.5.2	CentOS 7.5, Kernel 5.5.2

About Transwarp

Transwarp focuses on the product R&D for enterprise-level container cloud computing, big data, and artificial intelligence core platforms, and on taking big data and artificial intelligence infrastructure software to a new era. After many years of R&D, Transwarp has established several product lines: a container-based intelligent big data cloud platform, Transwarp Data Cloud (TDC); a one-stop big data platform, Transwarp Data Hub (TDH); an artificial intelligence platform, Transwarp Sophon; the hyper-converged big data all-in-one TxData Appliance; as well as holding many patented technologies. In 2018, Transwarp successfully became the first database vendor in the world to complete the TPC-DS test and pass official auditing in the past 12 years.

About Intel

Intel makes possible the most amazing experiences of the future

Intel, a leader in the semiconductor industry, is shaping the data-centric future with computing and communications technology that is the foundation of the world's innovations. The company's engineering expertise is helping address the world's greatest challenges as well as helping secure, power and connect billions of devices and the infrastructure of the smart, connected world – from the cloud to the network to the edge and everything in between.

[1] https://www.intel.com/content/www/us/en/products/docs/memory-storage/solid-state-drives/intel-ssd-data-center-family-for-nvme-video.html?wapkw=nvme

[2] Transwarp Argo DB Base config: Tested by Intel as of 3/27/2020. 4-node, 2x Intel Xeon Gold 5218 Processor, 16 cores HT On Turbo ON Total Memory 256 GB (8 slots/ 32 GB/ 2666 MHz), 480 GB Intel SSD DC S4510, 2x 2TB Intel SSD DC P4510, 10 x 2TB SATA 7200RPM HDD, BIOS: SE5C620.86B.0D.01.0395.022720191340 (Microscode: 0x500001c), CentOS 7.5, Kernel 5.5.2, TPC-DS 2.11.0, Atranswarp ArgoDB 1.3.2 Plus config: Tested by Intel as of 4/24/2020. 4-node, 2x Intel Xeon Gold 6240Y Processor, 18 cores HT On Turbo ON Total Memory 384 GB (12 slots/ 32 GB/ 2666 MHz), 480 GB Intel® SSD DC S4510, 8 * 2TB Intel SSD DC P4510, BIOS: SE5C620.86B.0D.01.0395.022720191340 (Microscode: 0x500001c), Centos 7.5, Kernel 5.5.2, TPC-DS 2.11.0, Atranswarp ArgoDB 1.3.2

Performance tests of components are measured using specific computer systems. Any change of hardware, software or configuration may affect the actual performance. Intel technologies may require enabled hardware, software or service activation.

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

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. Costs and results will vary. Intel does not guarantee any costs or cost reduction.

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 product or component can be absolutely secure.

© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.