Industry Solution Brief

Financial Services

Delivering Optimized HPC+AI Technologies for Financial Services

Intel® HPC+AI technologies and FSI expertise deliver optimized solutions for risk management, security, and pricing, speeding answers and reducing computing costs

Table of Contents

Executive Summary

The Financial Services Industry needs High Performance Computing and Artificial Intelligence (HPC+AI) resources for enormous calculations, data analysis, and customer services. Myriad financial products exist around the world. New offerings appear as the industry and technology evolve. With new regulations for banking and securities, companies have had to perform finer-grained analyses of its positions and portfolios, requiring greater computing capacity than ever before.

In addition to traditional computational methods, such as Monte Carlo and Black-Scholes, AI and machine learning/deep learning (ML/DL) are being applied in risk management, pricing, security, and retail banking. Adding to the growing complexity of use cases, FSI institutions often implement unique codes for their operations.

Intel has been a leader in providing optimized solutions to FSI for decades. The company continues to address FSI needs with advanced silicon, software, and teams of experts knowledgeable in architecture, software, and data science.

With new regulations for banking and securities, companies have had to perform finer-grained analyses of its positions and portfolios, requiring greater computing capacity than ever before.

FSI workloads benefit from the advancements of the 3rd Gen Intel® Xeon® Scalable processor family. Key FSI workloads run faster compared to the previous generation processors.[1]

Challenges

Understanding risk to manage operations and meet regulatory requirements takes a large part of HPC+AI resources in FSI. Fraud and financial crime plus pricing and hedging also impact IT's HPC+AI computing capacity.

Risk Management: Traders and Portfolio Managers want to know their exact portfolio risk at any time. That means being able to evaluate their risk in real time and run complex risk scenarios without having to worry about the cost of computation. The limits and cost of available computing capabilities require many financial institutions to rely on the previous day's risk metrics. With the limited time frame in which answers are needed and the volume of computing to be done in a single period, exhaustive analysis is not always feasible.

Security and Financial Crime: Fraud detection and some cybersecurity workloads need to run in real-time to be truly effective. Others, running in overnight batches, must process massive amounts of data with complex calculations. Achieving the real-time or near real-time performance where needed, plus accessing deeper analyses, require HPC+AI solutions optimized for the workloads being run on the most suitable architecture with acceptable TCO.

Regulatory Compliance: Frequent stress testing (Comprehensive Capital and Review (CCAR)) and review of financial position (Fundamental Review of the Trading Book (FRTB)) require compute-intensive operations that run complex calculations on an institution's data.

Landscape of FSI Workloads and Disciplines

Both institutional and retail traders, investors, strategists, and operational managers use a variety of workloads to drive their decision-making day to day and minute to minute. IT decision-making takes into account performance and efficiency of the infrastructure on these key workloads. They include many highly parallelized codes and innovative AI methods, such as the following:

- Monte Carlo simulations
- Finite difference method
- Tree-Based (e.g., Binomial, Trinomial)
- Closed Form (e.g., Black-Scholes)
- Artificial Neural Networks (ANN), computer vision, and recommender engines
- STAC benchmarks (A2, A3, M3)

While the math of the methods is typically well understood, companies will deploy them differently in infrastructures to meet their needs and budgets. This practice results in a variety of interesting challenges.

- Unique codes require unique optimization approaches, with algorithm complexities presenting their own performance challenges.
- Many applications are highly parallelized and run best on many cores while others are more serial.
- Multi-purpose systems run a variety of algorithms and are built to provide a best overall performance.
- Generalized systems implement different hardware (high core counts, memory capacity, memory bandwidth, etc.) to support a range of requirements (high throughput, low latency, low power, etc.), resulting in complexity.
- Heterogeneity in the computing infrastructure adds complexity for hardware maintenance and software implementation and optimizations.
- Thermal Design Power (TDP) priorities vary from customer to customer to meet power constraints.

The varying combinations of these challenges with each customer require approaching an FSI IT solution individually with appropriate expertise across software, hardware, and architecture. No single recipe exists for optimal performance. The same algorithm type (e.g., a Monte Carlo simulation) can behave differently based on the problem being solved. Data inputs, code complexity and design, system architecture, development approach, and other constraints all impact performance.

Generally, workloads are applied for different operational needs: risk management and compliance, security and financial crime, algorithmic trading, pricing and hedging, and customer experience. Characteristics of each can vary. And similarities can be optimized for better performance.

Risk Management & Compliance

Risk assessment and compliance workloads are often highly parallelized codes that are compute-intensive with double-precision floating-point calculations. The volume (or throughput) of calculations is often quite high. Some draw upon small enough data that can fit into a typical L2 cache. Others require a large memory footprint. Typical applications for these workloads include the following:

- Credit, market, liquidity, foreign, and interest rate risk modeling
- Variety of Valuation Adjustments (xVAs)
- Stress testing (CCAR and DFAST)
- Regulatory compliance (MiFID II, SOX, BASEL III, and FRTB/BASEL IV)

Security & Financial Crime

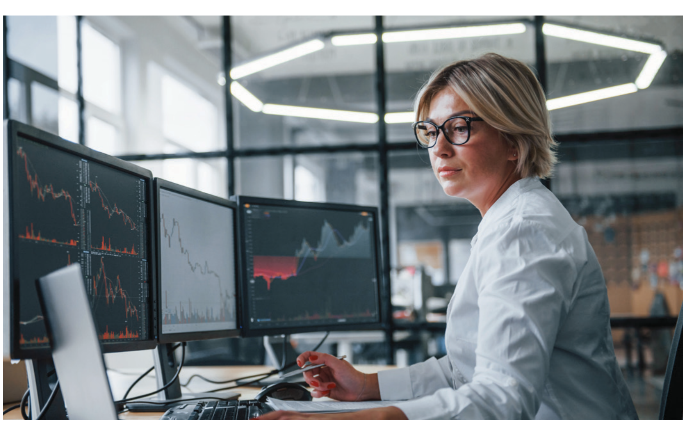
Some workloads for security and crime detection and prevention must run in real-time or near-real-time to be effective. Others run as batch jobs overnight. For example, Fraud Detection is a real-time operation that must process massive data with very low latency in order to mitigate fraud before it can take place. Anti-money laundering processes a lot of data in nightly batch runs using AI inferencing quickly to detect patterns indicative of the practice. Cybersecurity/malware detection might require low-latency, real-time processing, and large data to detect different patterns.

Algorithmic Trading

Algorithmic trading gets smarter with real-time analysis of current activities and historic patterns for decision-making. While the practice has been used for a while, applying AI methods to make intelligent trades is being explored and implemented in the industry. Both training and inference require specific technologies for optimal solutions.

Pricing & Hedging

Traders and Portfolio Managers can easily access current pricing of exchange-traded products to assess their risk and hedging. Some might want to run their own pricing scenarios at any time. Over the counter (OTC) derivatives, however, often do not have a market price. Thus, portfolio managers and traders using OTC might not have an accurate real-time view of the valuations and risk of their derivative positions. While traders might want real-time valuation and analysis, institutions might rely on previous day pricing.



Some workloads for security and crime detection and prevention must run in real-time or near-real-time to be effective. Fraud Detection must process massive data with very low latency in order to mitigate fraud before it can take place.

Customer Experience

Whether online or in-person, richer customer experiences in banking are mandatory for operational success and growth. New technologies, such as computer vision and AI-enabled behavior analysis, are helping bridge the gap between the physical space and digital channels. But, they require domain knowledge and expertise to implement.

Addressing FSI Challenges

With emerging technologies, the variety of commercial and custom codes used across FSI, and their unique applications by customers, solving challenges can be difficult. Strong tools and skills, plus knowledge of particular algorithms, software frameworks, and hardware architectures are essential to extracting the best possible performance from the codes. This becomes especially true as the financial environment changes due to new financial strategies and regulations.

Intel is a market leader of HPC+AI solutions in FSI. The company brings expertise across compute architectures, software development, and data science to solve the toughest problems in FSI, while helping to reduce TCO and improve operational efficiency. Intel serves a wide range of financial industry customers all over the globe—from some of the largest companies and most influential institutions to small trading firms and banking institutions. Each customer and challenge is unique, with the same level of expertise and commitment brought to them.

Solutions

The Intel Advantage

Intel is an early pioneer in HPC+AI and a leader in FSI computing. Dedicated Intel FSI teams are uniquely prepared to deliver consistent, long-term success for customers. The company offers an unmatched portfolio of extensive capabilities to solve today's FSI computational challenges, including the following:

- A dedicated data science and development team with deep FSI workload expertise to speed optimization of unique codes and their use cases through collaboration with the customer.
- Experience and innovation in multi-architecture, multi-framework, and multi-language programming to apply the right solution to the problem.
- Diverse compute architectures—CPUs, GPUs, IPUs, FPGAs, and accelerators—to address the unique demands of different workloads across traditional codes and emerging AI/ML.
- Built-in AI acceleration and instruction sets (e.g., Intel® Deep Learning Boost and Intel® Advanced Vector Extensions 512 (Intel® AVX-512) to reduce (and help eliminate) the reliance on expensive GPUs.

- Memory and storage innovations, such as Intel® Optane™ persistent memory and Distributed Asynchronous Object Storage (DAOS) have redefined the memory/storage hierarchy to accelerate data movement and to significantly expand data capacity in the compute complex.
- Proven networking technology, including intelligent 100 gigabit Ethernet controllers, to accelerate and balance cluster performance.
- Powerful monitoring and optimization tools that help identify code bottlenecks and increase performance and efficiency.
- Optimized software stacks, distributions and frameworks (e.g., Python, TensorFlow, openVINO), and libraries (e.g., Intel® Math Kernel Library, etc.). These assets leverage Intel® technologies in the silicon and achieve top performance and efficiency on Intel architecture for FSI workloads.

Intel® architecture has driven FSI datacenters for decades. The Intel® Xeon® Scalable processor family has delivered groundbreaking performance to financial services computing. The continuing deployment of Intel architecture has led to an ecosystem of FSI software and partnerships with software and solution providers, enabling

- Codes that run efficiently to reach insights faster
- Easier development efforts to improve accuracy and reduce risk
- Painless growth to help future-proof the business
- Reduced carbon footprint and technologies that assist institutions in their Sustainability/Social Responsibility/Government initiatives

Highlights of FSI Workload Optimizations on Intel® Architecture

Each financial institution has unique strategies and codes for their risk assessments, security, valuation, and pricing. This means that no single optimization strategy and recipe will address all customer needs. Each solution takes unique and individual approaches. That's why Intel has built a team of experts dedicated to solving customer problems.

Through work by Intel experts and collaboration with the ecosystem and customers, Intel optimizations for FSI workloads on 3rd Generation Intel Xeon Scalable processors are achieving new levels of performance. Some of these workloads are highlighted below.

Monte Carlo

Monte Carlo method is used in pricing and risk assessments. Highly parallelizable, large Monte Carlo simulations can be run on clusters or compute grids and will benefit from larger core counts. Monte Carlo simulation datasets are usually small enough to fit into the CPU's L2 cache, thus it is not memory bandwidth or capacity bound. Performance benchmarking has shown Monte Carlo runs up to 2.3X faster on the latest Intel Xeon Scalable processors compared to a top-bin AMD CPU (Figure 1).[1]

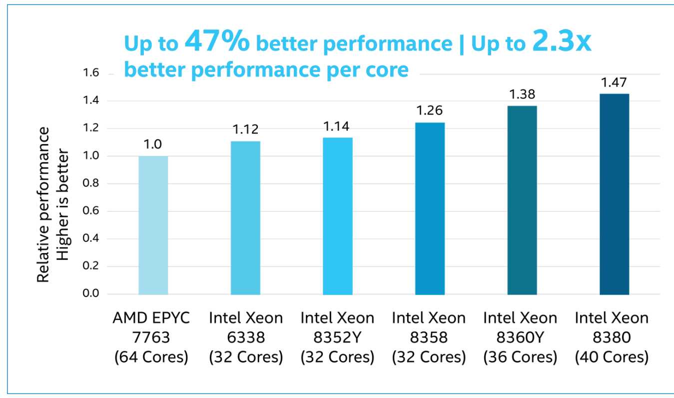


Figure 1. Monte Carlo benchmarks for 3rd Gen Intel Xeon Scalable processors versus AMD[1]

Matlogica offers algorithm differentiation using innovative solutions for running xVAs on Intel architecture. Matlogica's solution has achieved over 1000X speedup for one pricing application by leveraging Intel AVX-512 in 3rd Gen Intel Xeon Scalable processors.[2] See the Matlogica Case Study summary below or read the white paper.

Black-Scholes

Black-Scholes runs partial differential equation calculations. Black-Scholes is typically used to perform quick calculations of options pricing by estimating the variation of financial instruments over time. Easily parallelized, it benefits from access to more cores and efficiently scales up and down. Black-Scholes jobs are complex with large datasets, and thus can become memory bandwidth bound. Performance benchmarking has shown Black-Scholes runs up to 2.8X faster on the latest Intel Xeon Scalable processors compared to a top-bin AMD CPU (Figure 2).[3]

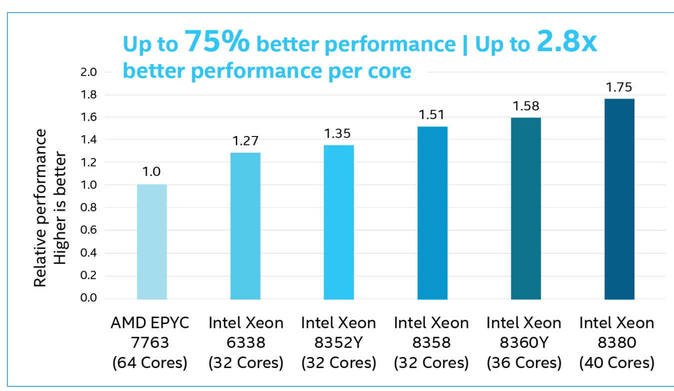


Figure 2. Black-Scholes benchmarks for 3rd Gen Intel Xeon Scalable processors versus AMD[3]

Binomial

Binomial calculations are also used for options pricing by breaking down time to expiration into intervals. This method offers greater flexibility than Black-Scholes, because traders can alter inputs at each step in the process. But models are complex to construct. Parallelized approaches can take advantage of more cores in a cluster.

Binomial jobs are quite complex, with large data sets. Thus, they can become memory bandwidth bound. Performance

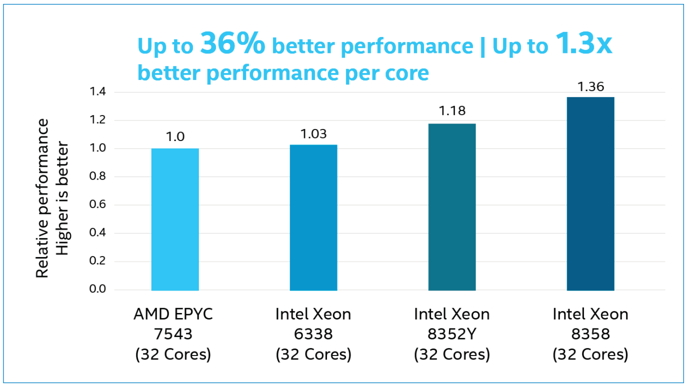


Figure 3. Binomial benchmarks for 3rd Gen Intel Xeon Scalable processors versus AMD4[4]

benchmarking has shown that optimized Binomial calculations run up to 1.3X faster on the latest Intel Xeon Scalable processors compared to a top-bin AMD CPU (Figure 3).[4]

STAC Benchmarks Set New Records

STAC benchmarks on Intel® architecture have achieved breakthrough performance for STAC-A2, STAC-A3, and STAC-M3.

- STAC-A2—New record for space efficiency with over 2.4X better than latest GPU solution.[5]
- STAC-A3—New record for storage efficiency with over 3X better than latest GPU solution.[6]
- STAC-M3—New standard for memory optimization with over 20X speedup for 2 Kanaga market snap benchmarks.[7]

AI for Risk Calculations/Pricing

Emerging AI methods and innovative services are being applied to FSI challenges to accelerate time to solutions and reveal insights from an institution's data. Quantifi Solutions delivers risk management and analytics services to companies and individuals around the world. The company offers real-time views of valuations and risk of trading positions using Artificial Neural Network (ANN) models. Quantifi's ANN model delivers 0.01 percent deviation compared to theoretical fair value for tested derivatives. Running on 3rd Gen Intel Xeon Scalable processors, the solution is up to 700X faster than conventional methods. Read the full white paper here.

Intel® Technologies Make a Difference in FSI

Leveraging the capabilities of different Intel technologies across silicon, optimization libraries and distributions, and software enables FSI codes to run fast and efficiently. This helps speed time to solution and reduce cost of computing.

3rd Generation Intel® Xeon® Scalable Processors

FSI workloads benefit from the advancements of the 3rd Gen Intel Xeon Scalable processor family. Key FSI workloads run faster compared to the previous generation processors (Figure 4).[8]

Intel® Advanced Vector Instructions 512 (Intel® AVX-512)—These instructions along with very wide (512-bit) registers, allow floating-point calculations to execute faster in hardware than software. Many optimizations in Intel libraries take advantage of Intel AVX-512.

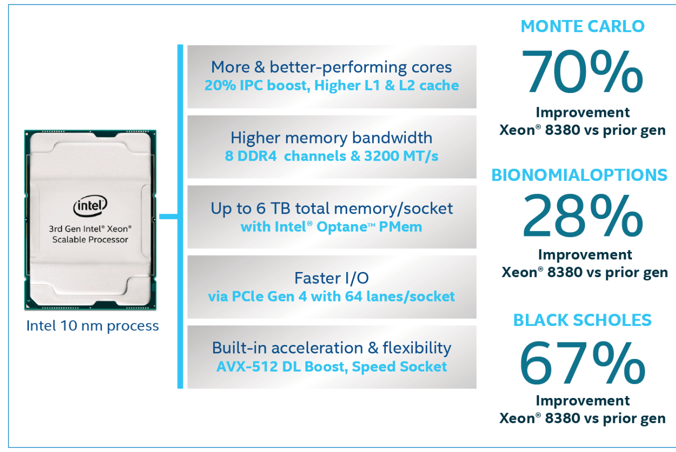


Figure 4. 3rd Gen Intel® Xeon® Scalable processors deliver improved performance for FSI workloads over previous generation[8]

Intel® Deep Learning Boost (Intel® DL Boost)—An evolving suite of integrated technologies, Intel DL Boost accelerates certain operations used in AI, DL, and ML. As FSI codes and approaches evolve, AI is becoming a more important tool for operations, such as fraud detection, money laundering, and others. Intel continues to integrate new technologies, such as Intel DL Boost, into its silicon to accelerate AI.

Intel® Optane™ Persistent memory (Intel® Optane™ PMem)—Intel redefined the memory/storage hierarchy with Intel Optane PMem. PMem can accelerate large database operations by storing up to 6 TB of data per socket closer to the CPU. PMem also enables persistence of data, eliminating the need to reload data from storage after a shutdown. See the MemVerge Case Study summary.

Intel® Speed Select Technology (Intel® SST)—A family of features built into the silicon, Intel SST is designed to provide more active and granular control over CPU performance. Intel SST enables IT better to manage power consumption, match workloads to the CPU, and dynamically reconfigure the processors on the fly. Intel SST enables more efficient system designs and reduces complexities and cost of computing.

Tuning Infrastructure

Besides Intel Xeon Scalable processors, Intel offers a range of compute technologies to address the unique needs of critical workloads. Intel FSI experts can help IT departments make informed choices about which technologies and optimizations are best suited for their operations.

FSI infrastructures can also benefit from other Intel and ecosystem technologies and appliances. These include 100 Gbps Intel® Ethernet Network Adapter E810 family to help balance and improve performance across clusters.

Optimizing FSI Solutions In-House, Grid, and Cloud

Intel technologies and products benefit FSI workloads wherever they are run. Leading Cloud Service Providers (CSPs) offer high-performance instances built on 3rd Gen Intel Xeon Scalable processors, Intel Optane PMem, and

Traders and Portfolio Managers can easily access current pricing of exchange-traded products to assess their risk and hedging. Some might want to run their own pricing scenarios at any time.

other Intel technologies. Intel FSI system architects can help customers define the most optimal configurations for deployments that use any—or all—of these approaches.

Examples

Across FSI, Intel technologies have helped accelerate financial calculations and improve operations. Following are some examples.

Quantifi—Accelerating Derivative Valuations by 700X Using AI

Over the Counter (OTC) derivative traders often don't have a real-time view of valuations and risk of their positions. Obtaining real-time valuations of OTC derivatives has been challenging. The commonly used techniques for these products require significant machine time, making them computationally expensive.

Quantifi Solutions' ANN models deliver real-time pricing with an accuracy considered equivalent to conventional approaches, such as numerical integration and Monte Carlo methods.

Quantifi's solution, running on 3rd Gen Intel Xeon Scalable processors, delivers 0.01 percent deviation when compared to theoretical fair value for tested derivatives. Their ANN model is also orders of magnitude faster—up to 700X faster—than conventional methods, enabling the approach to deliver real-time valuations. Read the full white paper here.

Matlogica—Over 1000X Speedup for xVA Pricing

Financial institutions must report sensitivities to multiple scenarios to regulators, requiring banks to recalculate their trade portfolios thousands of times with fine granularity. Multiple valuation adjustments (xVAs) represent price correction at a portfolio level, but they require thousands of valuations for each trade accurately to determine the Credit Value Adjustment (CVA) and Debit Value Adjustment (DVA) needed for analysis and reporting. With this volume of computation, traditional bump-and-revalue methods are infeasible or demand large—and expensive—cloud deployments. Instead, Automatic Adjoint Differentiation (AAD) has become a highly desirable method.

Matlogica has demonstrated its innovative technology and approach to financial derivative computations can accelerate pricing as much as 1770X on one application. Building upon Intel AVX-512 on 3rd Gen Intel Xeon Scalable processors and its Matlogica Vector Accelerator Library, workloads that include xVA pricing and Greeks together achieve up to 832X performance gains. Read the entire white paper here.

Pico

Pico provides mission critical technology, data, and analytic services for the financial markets community. The company's resilient proprietary network, PicoNet, interconnects all major financial data centers around the world, including all major public cloud providers. Customers include the world's largest banks, exchanges, quantitative hedge funds, electronic market makers, and asset managers.

Pico Corvil Analytics bare-metal appliances run in Pico's co-located data centers and in private data centers to ingest, timestamp, process, and store trade information used by financial strategists and traders. As data volumes have continued to grow, and with significant adoption of 100 Gbps Ethernet in the trading network, Pico introduced the Corvil 10000 appliance with 3rd Gen Intel Xeon Scalable processors.

3rd Gen Intel Xeon Scalable processors deliver the capabilities Corvil appliances need to continue to deliver its industry-leading services. With the new processors, the Corvil 10000 is targeting 100 million packets per second from ingest to storage, rather than supporting bursts. This level of performance was unachievable without the new technology. Read the case study here.

Additional Examples

As a leader in providing HPC+AI solutions for FSI, the above are only a few examples. Others include the following:

Aerospike with Intel Optane PMem achieves real-time fraud detection for PayPal. Read more about Intel and Aerospike here. Read the PayPal case study here.

In-memory database performance benefits from large memory footprints. MemVerge, using Intel Optane PMem, delivers significant performance improvements for FSI applications, such as kdb+, without using more expensive DRAM populations. Read the case study here.

Our Continuing Commitment to Technology Evolution

Intel brings expertise across compute architectures, software development, and data science to solve the toughest problems in FSI, while helping to reduce TCO and improve operational efficiency. Every customer and challenge is unique; Intel delivers the same level of expertise and commitment to each of them. As FSI needs and applications evolve, Intel will continue to provide innovative solutions and expert services to help optimize FSI workloads on Intel architecture.

New Intel engineering and manufacturing sites being built around the world will allow the company to continue to develop and create these technologies. Intel's goal is to address future FSI challenges with next-generation solutions that accelerate calculations and reduce the cost of computing.

[1] **Monte Carlo—**AMD EPYC-7763; Config Date: 1/21/2021; # CPU Sockets: 2; # CPU Cores: 64; CPU Base Frequency: 2.45 GHz; CPU Max Frequency: 3.5 GHz; CPU Base TDP: 280 W; RAM: 512GB, 16x32GB 3200MHz DDR4, Samsung M393A4K40DB3-CWE; Hard Drive: 3.7TB SATA SAMSUNG MZ7LH3T8HMLT; Cluster File System: N/A; BIOS: Ver M02 Rev 5.21; BIOS Settings: SMT=Enabled, NUMA nodes per socket=NPS4, Power Policy Quick Setting=Best Performance: IOMMU=Enabled, Determinsim Control=Manual, Determism Slider=Power, cTDP Control=Manual, cTDP=280, Package Power Limit Control=Manual, Pack Power Limit=280; Microcode: 0xa001114; Intel Management: N/A; BMC: 12.49.06; Operating System: Red Hat Enterprise Linux 8.3; Kernel: 4.18.0-240.10.1.el8_3.x86_64

Intel ICX-6338; Config Date: 5/10/2021; # CPU Sockets: 2; # CPU Cores: 32; CPU Base Frequency: 2.0 GHz; CPU Max Frequency: 3.2 GHz; CPU Base TDP: 205 W; RAM: 256GB 16*16GB 3200MT/s DDR4, Hynix HMA82GR7CJR8N-XN; Hard Drive: SSDSC2KG96 960GB; Cluster File System: HDR Based Lustre; BIOS: SE5C6200.86B.0020.P23.2103261309; BIOS Settings: HT=on TURBO=ON; Microcode: 0xd000270; Intel Management: 04.04.04.053; BMC: 2.78; Operating System: CentOS Linux release 8.3.2011; Kernel: 4.18.0-240.22.1.el8_3.crt1.x86_64

Intel ICX-8352Y; Config Date: 5/10/2021; # CPU Sockets: 2; # CPU Cores: 32; CPU Base Frequency: 2.2 GHz; CPU Max Frequency: 3.4 GHz; CPU Base TDP: 205 W; RAM: 256GB 16*16GB 3200MT/s DDR4, Hynix HMA82GR7CJR8N-XN; Hard Drive: SSDSC2KG96 960GB; Cluster File System: HDR based Lustre; BIOS: SE5C6200.86B.0020.P23.2103261309; BIOS Settings: HT=on TURBO=ON; Microcode: 0xd000270; Intel Management: 04.04.04.053; BMC: 2.78; Operating System: CentOS Linux release 8.3.2011; Kernel: 4.18.0-240.22.1.el8_3.crt1.x86_64

Intel ICX-8358; Config Date: 5/10/2021; # CPU Sockets: 2; # CPU Cores: 32; CPU Base Frequency: 2.6 GHz; CPU Max Frequency: 3.4 GHz; CPU Base TDP: 250 W; RAM: 256GB 16*16GB 3200MT/s DDR4, Hynix HMA82GR7CJR8N-XN; Hard Drive: SSDSC2KG96 960GB; Cluster File System: HDR based Lustre; BIOS: SE5C6200.86B.0020.P23.2103261309; BIOS Settings: HT=on TURBO=ON; Microcode: 0xd000270; Intel Management: 04.04.04.053; BMC: 2.78; Operating System: CentOS Linux release 8.3.2011; Kernel: 4.18.0-240.22.1.el8_3.crt1.x86_64

Intel ICX-8360Y; Config Date: 5/10/2021; # CPU Sockets: 2; # CPU Cores: 36; CPU Base Frequency: 2.4 GHz; CPU Max Frequency: 3.5 GHz; CPU Base TDP: 250 W; RAM: 256GB 16*16GB 3200MT/s DDR4, Hynix HMA82GR7CJR8N-XN; Hard Drive: SSDSC2KG96 960GB; Cluster File System: HDR based Lustre; BIOS: SE5C6200.86B.0020.P23.2103261309; BIOS Settings: HT=on TURBO=ON; Microcode: 0xd000270; Intel Management: 04.04.04.053; BMC: 2.78; Operating System: CentOS Linux release 8.3.2011; Kernel: 4.18.0-240.22.1.el8_3.crt1.x86_64

Intel ICX-8380; Config Date: 5/10/2021; # CPU Cores: 40; CPU Base Frequency: 2.3 GHz; CPU Max Frequency: 3.4 GHz; CPU Base TDP: 270 W; RAM: 256GB 16*16GB 3200MT/s DDR4, Hynix HMA82GR7CJR8N-XN; Hard Drive: SSDSC2KG96 960GB; Cluster File System: HDR based Lustre; BIOS: SE5C6200.86B.0020.P23.2103261309; BIOS Settings: HT=on TURBO=ON; Microcode: 0xd000270; Intel Management: 04.04.04.053; BMC: 2.78; Operating System: CentOS Linux release 8.3.2011; Kernel: 4.18.0-240.22.1.el8_3.crt1.x86_64

[2] System configuration: 2x Intel® Xeon® 8280 processors (28 cores/processor = 56 cores/112 threads), 192 GB DRAM, Matlogica AADC library (5/15 release), Intel® C++ Compiler 19.1.1. Results published by Matlogica at www.matlogica.com.

[3] **Black-Scholes—**AMD EPYC-7763; Config Date: 1/21/2021; # CPU Sockets: 2; # CPU Cores: 64; CPU Base Frequency: 2.45 GHz; CPU Max Frequency: 3.5 GHz; CPU Base TDP: 280 W; RAM: 512GB, 16x32GB 3200MHz DDR4, Samsung M393A4K40DB3-CWE; Hard Drive: 3.7TB SATA SAMSUNG MZ7LH3T8HMLT; Cluster File System: N/A; BIOS: Ver M02 Rev 5.21; BIOS Settings: SMT=Enabled, NUMA nodes per socket=NPS4, Power Policy Quick Setting=Best Performance: IOMMU=Enabled, Determinsim Control=Manual, Determism Slider=Power, cTDP Control=Manual, cTDP=280, Package Power Limit Control=Manual, Pack Power Limit=280; Microcode: 0xa001114; Intel Management: N/A; BMC: 12.49.06; Operating System: Red Hat Enterprise Linux 8.3; Kernel: 4.18.0-240.10.1.el8_3.x86_64

Intel ICX-6338; Config Date: 5/10/2021; # CPU Sockets: 2; # CPU Cores: 32; CPU Base Frequency: 2.0 GHz; CPU Max Frequency: 3.2 GHz; CPU Base TDP: 205 W; RAM: 256GB 16*16GB 3200MT/s DDR4, Hynix HMA82GR7CJR8N-XN; Hard Drive: SSDSC2KG96 960GB; Cluster File System: HDR Based Lustre; BIOS: SE5C6200.86B.0020.P23.2103261309; BIOS Settings: HT=on TURBO=ON; Microcode: 0xd000270; Intel Management: 04.04.04.053; BMC: 2.78; Operating System: CentOS Linux release 8.3.2011; Kernel: 4.18.0-240.22.1.el8_3.crt1.x86_64

Intel ICX-8352Y; Config Date: 5/10/2021; # CPU Sockets: 2; # CPU Cores: 32; CPU Base Frequency: 2.2 GHz; CPU Max Frequency: 3.4 GHz; CPU Base TDP: 205 W; RAM: 256GB 16*16GB 3200MT/s DDR4, Hynix HMA82GR7CJR8N-XN; Hard Drive: SSDSC2KG96 960GB; Cluster File System: HDR based Lustre; BIOS: SE5C6200.86B.0020.P23.2103261309; BIOS Settings: HT=on TURBO=ON; Microcode: 0xd000270; Intel Management: 04.04.04.053; BMC: 2.78; Operating System: CentOS Linux release 8.3.2011; Kernel: 4.18.0-240.22.1.el8_3.crt1.x86_64

Intel ICX-8358; Config Date: 5/10/2021; # CPU Sockets: 2; # CPU Cores: 32; CPU Base Frequency: 2.6 GHz; CPU Max Frequency: 3.4 GHz; CPU Base TDP: 250 W; RAM: 256GB 16*16GB 3200MT/s DDR4, Hynix HMA82GR7CJR8N-XN; Hard Drive: SSDSC2KG96 960GB; Cluster File System: HDR based Lustre; BIOS: SE5C6200.86B.0020.P23.2103261309; BIOS Settings: HT=on TURBO=ON; Microcode: 0xd000270; Intel Management: 04.04.04.053; BMC: 2.78; Operating System: CentOS Linux release 8.3.2011; Kernel: 4.18.0-240.22.1.el8_3.crt1.x86_64

Intel ICX-8360Y; Config Date: 5/10/2021; # CPU Sockets: 2; # CPU Cores: 36; CPU Base Frequency: 2.4 GHz; CPU Max Frequency: 3.5 GHz; CPU Base TDP: 250 W; RAM: 256GB 16*16GB 3200MT/s DDR4, Hynix HMA82GR7CJR8N-XN; Hard Drive: SSDSC2KG96 960GB; Cluster File System: HDR based Lustre; BIOS: SE5C6200.86B.0020.P23.2103261309; BIOS Settings: HT=on TURBO=ON; Microcode: 0xd000270; Intel Management: 04.04.04.053; BMC: 2.78; Operating System: CentOS Linux release 8.3.2011; Kernel: 4.18.0-240.22.1.el8_3.crt1.x86_64

Intel ICX-8380; Config Date: 5/10/2021; # CPU Cores: 40; CPU Base Frequency: 2.3 GHz; CPU Max Frequency: 3.4 GHz; CPU Base TDP: 270 W; RAM: 256GB 16*16GB 3200MT/s DDR4, Hynix HMA82GR7CJR8N-XN; Hard Drive: SSDSC2KG96 960GB; Cluster File System: HDR based Lustre; BIOS: SE5C6200.86B.0020.P23.2103261309; BIOS Settings: HT=on TURBO=ON; Microcode: 0xd000270; Intel Management: 04.04.04.053; BMC: 2.78; Operating System: CentOS Linux release 8.3.2011; Kernel: 4.18.0-240.22.1.el8_3.crt1.x86_64

[4] **Binomial—**AMD EPYC-7543; Config Date: 3/29/2021; Sockets: 2; # CPU Cores: 32; CPU Base Frequency: 2.8 GHz; CPU Max Frequency: 3.7 GHz; CPU Base TDP: 225 W; RAM: 1024GB, 16x64GB 3200MHz DDR4, Samsung M393A4K40DB3-CWE; Hard Drive: 2x 480GB Micron 5300 Pro (MTFDDAV480TDS); Cluster File System: N/A; BIOS: Ver 2.0.3; BIOS Settings: Logical Processor=Enabled, NUMA nodes per socket=4, System Profile Custom, Determinism Slider: Power Determinism; Microcode: 0xa001119; Intel Management: N/A; BMC: 4.32; Operating System: Red Hat Enterprise Linux 8.3; Kernel: 4.18.0-240.15.1.el8_3.x86_64;

Intel ICX-6338; Config Date: 5/10/2021; # CPU Sockets: 2; # CPU Cores: 32; CPU Base Frequency: 2.0 GHz; CPU Max Frequency: 3.2 GHz; CPU Base TDP: 205 W; RAM: 256GB 16*16GB 3200MT/s DDR4, Hynix HMA82GR7CJR8N-XN; Hard Drive: SSDSC2KG96 960GB; Cluster File System: HDR Based Lustre; BIOS: SE5C6200.86B.0020.P23.2103261309; BIOS Settings: HT=on TURBO=ON; Microcode: 0xd000270; Intel Management: 04.04.04.053; BMC: 2.78; Operating System: CentOS Linux release 8.3.2011; Kernel: 4.18.0-240.22.1.el8_3.crt1.x86_64

Intel ICX-8352Y; Config Date: 5/10/2021; # CPU Sockets: 2; # CPU Cores: 32; CPU Base Frequency: 2.2 GHz; CPU Max Frequency: 3.4 GHz; CPU Base TDP: 205 W; RAM: 256GB 16*16GB 3200MT/s DDR4, Hynix HMA82GR7CJR8N-XN; Hard Drive: SSDSC2KG96 960GB; Cluster File System: HDR based Lustre; BIOS: SE5C6200.86B.0020.P23.2103261309; BIOS Settings: HT=on TURBO=ON; Microcode: 0xd000270; Intel Management: 04.04.04.053; BMC: 2.78; Operating System: CentOS Linux release 8.3.2011; Kernel: 4.18.0-240.22.1.el8_3.crt1.x86_64

Intel ICX-8358; Config Date: 5/10/2021; # CPU Sockets: 2; # CPU Cores: 32; CPU Base Frequency: 2.6 GHz; CPU Max Frequency: 3.4 GHz; CPU Base TDP: 250 W; RAM: 256GB 16*16GB 3200MT/s DDR4, Hynix HMA82GR7CJR8N-XN; Hard Drive: SSDSC2KG96 960GB; Cluster File System: HDR based Lustre; BIOS: SE5C6200.86B.0020.P23.2103261309; BIOS Settings: HT=on TURBO=ON; Microcode: 0xd000270; Intel Management: 04.04.04.053; BMC: 2.78; Operating System: CentOS Linux release 8.3.2011; Kernel: 4.18.0-240.22.1.el8_3.crt1.x86_64

[5] STAC-A2—Calculated based on data from 2 reports: https://stacresearch.com/INTC190903 and https://stacresearch.com/NVDA200909.

[6] STAC-A3—Source report: https://stacresearch.com/news/INTC200514.

[7] STAC-M3—Source report: https://stacresearch.com/news/KDB200603.

[8] See [37], [108] and [123] at www.intel.com/3gen-xeon-config. Results may vary.

Please Recycle 0522/RJMJ/CP/PDF 351172-001US