Solution brief

Health and life sciences
Health analytics

Intel CPUs deliver breakthrough speed for genomic analysis

A collaboration between Intel, the Berlin Institute of Health at Charité, Dell and Sentieon led to an accelerated software pipeline for analyzing long-read single-cell RNA sequencing data that's 14x faster than the industry standard.[1]

Executive Summary

Major advances in healthcare — from population health policies to personalized drugs — are being driven by genomic data. Next-generation sequencing (NGS) has made gene sequencing more available than ever, but a large portion of the value of genomics data is only revealed in secondary analysis. It's through data analytics and bioinformatics that healthcare practitioners and researchers get the insights they need to deliver precise care or to achieve breakthroughs.

Bringing the power of genomics to more people

Intel believes that we can improve healthcare across the world by making genome informatics faster, more accurate, more efficient and more available. In pursuit of this goal, Intel collaborated with leading innovators in scalable bioinformatics to break down critical barriers to genomic analysis.

Testing the power of Intel CPUs for genomic analysis

As the Berlin Institute of Health at Charité (Charité BIH) sought to advance its cloud-based bioinformatics initiatives, Intel brought them together with Sentieon, a solution developer in precision medicine. Sentieon developed an optimized pipeline for analyzing long-read single-cell RNA sequencing (scRNA-seq) that delivered stunning speed on samples sequenced with Oxford Nanopore Technologies (ONT) devices. This incredible performance was achieved on standard Intel data center CPUs in Charité BIH's private cloud and without dedicated hardware acceleration.

> The proof-of-concept workflow running on 4th Generation Intel® Xeon® Scalable processors completed complex genomic analysis **14x faster** than the industry standard workflow.[1]

Setting a new standard for NGS data analysis

By employing readily available optimizations for Intel CPUs, Sentieon's pipeline lowered the time to perform the analysis of long-read scRNA-seq data by more than 14x, compared to the recently released solution implemented by ONT.[1] The same analysis benchmarking was also performed on 3rd Gen Intel® Xeon® Scalable processors. Testing showed that 4th Gen Intel® Xeon® Scalable processors delivered a 21% gen-over-gen boost in speed over the previous generation.[1]

4th Gen Intel® Xeon® Scalable processors performed genomic analysis **21% faster** than 3rd Gen Intel® Xeon® Scalable processors on Sentieon's optimized workflow.[1]

Bringing genome sequencing to the world

The Sentieon proof-of-concept solution shows that it's possible to move genomic analysis solutions from specialized laboratory workstations onto centralized, standardized data centers that can dramatically improve scalability and access for many use cases. It has established a new pathway for Charité BIH to bring ultra-fast RNA sequencing (RNA-seq) analyses to a wider audience via its Charité Clinical Cloud.

Spotlight on Sentieon

Sentieon works with some of the biggest clinical diagnostic organizations and hospitals in the world. The company's mission is to improve healthcare through cutting-edge bioinformatics software. Sentieon's world-class engineering ensures that genomic data gets treated quickly and accurately. They've been routinely recognized for exceptional accuracy by leading regulatory bodies.

What really drives Sentieon's work is a vision to improve the scalability and flexibility of genomic analytics in order to make precision medicine more effective and bring it to more people in more places. Sentieon is overcoming infrastructure limitations by capitalizing on the power of Intel® Xeon® Scalable processors. Solutions that necessitate GPUs or FPGAs for acceleration can't go as far as those that run on standard Intel CPUs. Leveraging continual advancements in performance and capabilities from Intel, Sentieon is working to make genomic analysis fast, cost effective, and easy to access.

Challenge: Overcoming barriers to cloud bioinformatics

The high sequencing throughput that NGS makes possible is opening new avenues for understanding the complexities of the human genome. However, while NGS is transforming biological research and healthcare practices it also presents major computing challenges.

Powerful data analysis is required to derive observations and insights from NGS data. The data must be read, corrected, sorted and prepared for consumption by a bioinformatician, medical practitioner or researcher. The key to unlocking the value of NGS is through advancing the field of bioinformatics.

What is long-read genomic sequencing?

This next-generation sequencing (NGS) technique sequences much longer DNA fragments than traditional short-read sequencing methods. This allows for the detection of complex structural variants that may be difficult to detect otherwise.

A deeper read that unlocks hidden insights

Many bioinformatic solutions for NGS employ short-read sequencing on bulk DNA data. This offers high analytics throughput but delivers only an average of the genomic data in the tissue or organism. To look deeper into gene expression — and the health of the individual — long-read scRNA-seq reads the majority of full transcripts in individual cells. This enables comparative analysis between cells within the same organ, as well as the identification of distinct cell populations, cell states, and rare cell types. The detailed information provided by long-read scRNA-seq provides deeper insights into gene expression dynamics at a single-cell level and can reveal more information about protein function and variances that might otherwise be missed.

When read and analyzed, long-read scRNA-seq data enables clinicians and researchers to make in-depth comparisons between individual cells and gene expressions. Understanding the subtleties of cellular diversity reveals variations in our genetic code, which could point to either signs of concern or a more effective way to treat a specific condition.

Maximizing both speed and efficiency

While highly valuable in precision medicine and research, analyzing long-read scRNA-seq requires processing massive data libraries. It also employs computationally demanding workloads that require powerful compute resources. This is why solutions for analyzing genomic data are often engineered around an HPC workstation with a dedicated hardware accelerator, such as a GPU or FPGA. This increases the cost, complexity and energy consumption of achieving an in-depth genomic analysis. This tradeoff is especially apparent when considering the flexibility and economies of scale found in the cloud.

Solution value: Speeding sequencing while making it more scalable

Charité BIH and Intel envision a future version of the Charité Clinical Cloud that offers readily available bioinformatics solutions capable of advanced analysis on large data sets. Demanding processes, such as genomic analysis on long-read scRNA-seq data, offer a deeper understanding of disease and can help genomic researchers and clinicians achieve a more complete picture of the health of a person or a population.

A platform for the future of healthcare: 4th Generation Intel® Xeon® Scalable processors

The latest generation of Intel Xeon Scalable processors offers a cost-effective server architecture.[2*] Both increased memory bandwidth and I/O bandwidth with DDR5 memory and PCIe 5.0 allow you to work with larger datasets with less latency. Higher speeds and up to 60 cores means you need fewer CPUs to complete tasks. 4th Gen Intel Xeon Scalable processors also offer up to 53% greater general purpose compute performance over 3rd Gen Intel Xeon Scalable processors.[3*]

By bringing down the costs of genomic analysis, both in terms of infrastructure and energy consumption, different types of innovators can build solutions that will gain new insights from genomic data analysis. The more we can remove special requirements by bringing bioinformatics workloads onto computing resources that are virtually ubiquitous — such as x86 Intel cloud and server CPUs — the sooner we will see those breakthrough solutions.

Solution architecture: Optimizations greatly enhance performance

Sentieon has optimized processes for Intel CPUs that accelerate and improve multiple approaches to NGS data analysis. The proof-of-concept workflow designed by Sentieon combines ONT long-reads with a single-cell RNA sequencing method, which was developed by 10x Genomics. The workflow outputs a standard gene expression matrix and uniform manifold approximation and projection (UMAP) plot of the analyzed cell population. This allows data analysts to easily find associations between genes and variations between cells.

Sentieon set out to improve ONT analysis performance with software optimizations available on Intel data center CPUs, as opposed to employing a specialized workstation with a dedicated hardware accelerator.

Custom-built sequencing pipeline from Sentieon

Sentieon wrote a custom pipeline based on the readily available Sockeye ONT open-source workflow. The Sentieon software pipeline applies novel strategies to eliminate steps and accelerate processes. For example, the Sentieon pipeline combines the adapter scanning and barcode extraction stages into a single tool implemented inside the Sentieon software.

Systems integrator SVA led the process of building the cloud server to run Sentieon's software pipeline and integrating into the Charité BIH's private cloud.

Optimizations for Intel dramatically boost speed

Sentieon and Intel worked closely to optimize the performance of the genomics informatics workload for Intel Xeon Scalable processors.

ONT wf-single-cell vs Sentieon's modified wf-single-cell on Clinical cloud

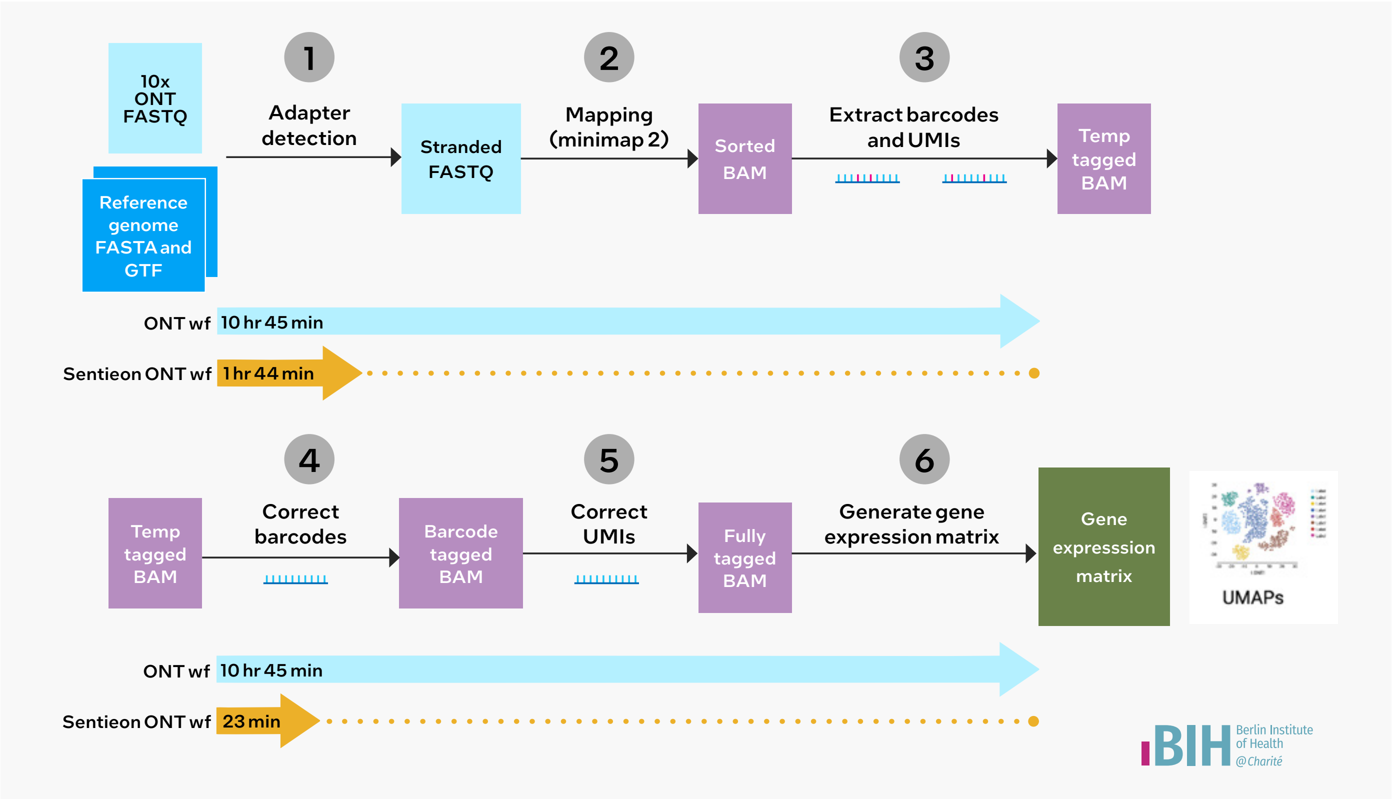

Figure 1. The custom sequencing software pipeline developed by Sentieon can dramatically accelerate long-read scRNA-seq on standard cloud infrastructure.

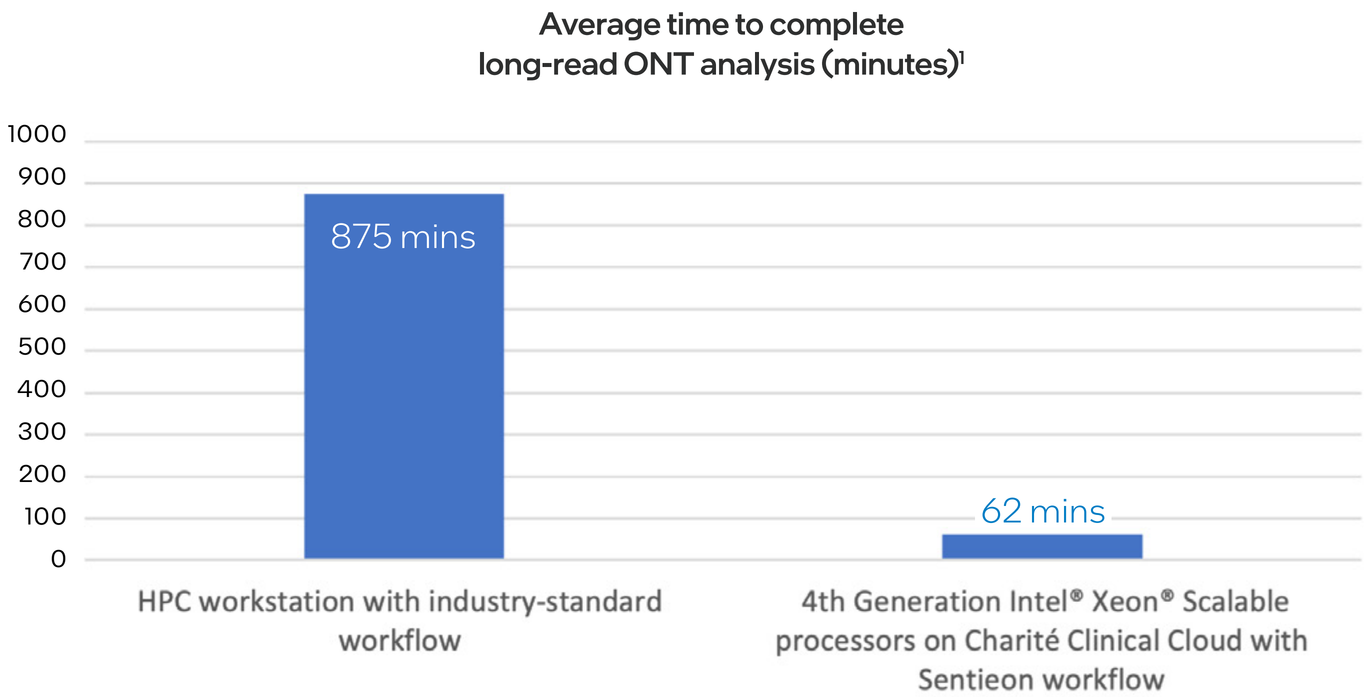

Figure 2. Compared to an industry-standard genomic analysis pipeline, 4th Generation Intel® Xeon® Scalable processors offer a leap forward in performance.[1]

> "SVA was available throughout the project as the operator of the infrastructure, supporting communication between the various parties. In addition, SVA designed, built and operated the clinical cloud and supports the Charité team. All new workloads were integrated to the cluster with the help of SVA."
>
> – Daniel Vois, head of sales,
> SVA Healthcare/Germany

Results: Dramatically faster time for long-reads

After successive testing on different ONT FASTQ samples, the proof-of-concept solution delivered a result an average of 14x faster than a standard ONT workflow running on an standalone workstation.[1] This massive improvement promises breakthrough throughput and efficiency. 4th Gen Intel Xeon Scalable processors are designed to maximize workload efficiency for analytics to help applications deliver better price/performance in the cloud.

The standard process to read and analyze long-read ONT sequencing data typically takes more than 14 hours to complete.[1] The custom software pipeline, developed by Sentieon and running on 4th Gen Intel Xeon Scalable processors in the Charité's cloud, completed the same analysis in just over an hour.[1]

Improving the speed of this workload so dramatically, using only software optimizations, offers the chance to deliver advanced bioinformatics capabilities with fewer resources. What's more, by employing standard data center hardware without a dedicated accelerator, accessible genomic analytics in the cloud seems closer than ever.

More speed and efficiency with every generation

The same benchmarking with Sentieon's optimized workflow was performed on 3rd Gen Intel Xeon Scalable processors in Charite's cloud, showing an average speed that is 11x faster than the the ONT standard Sockeye workflow.[1]

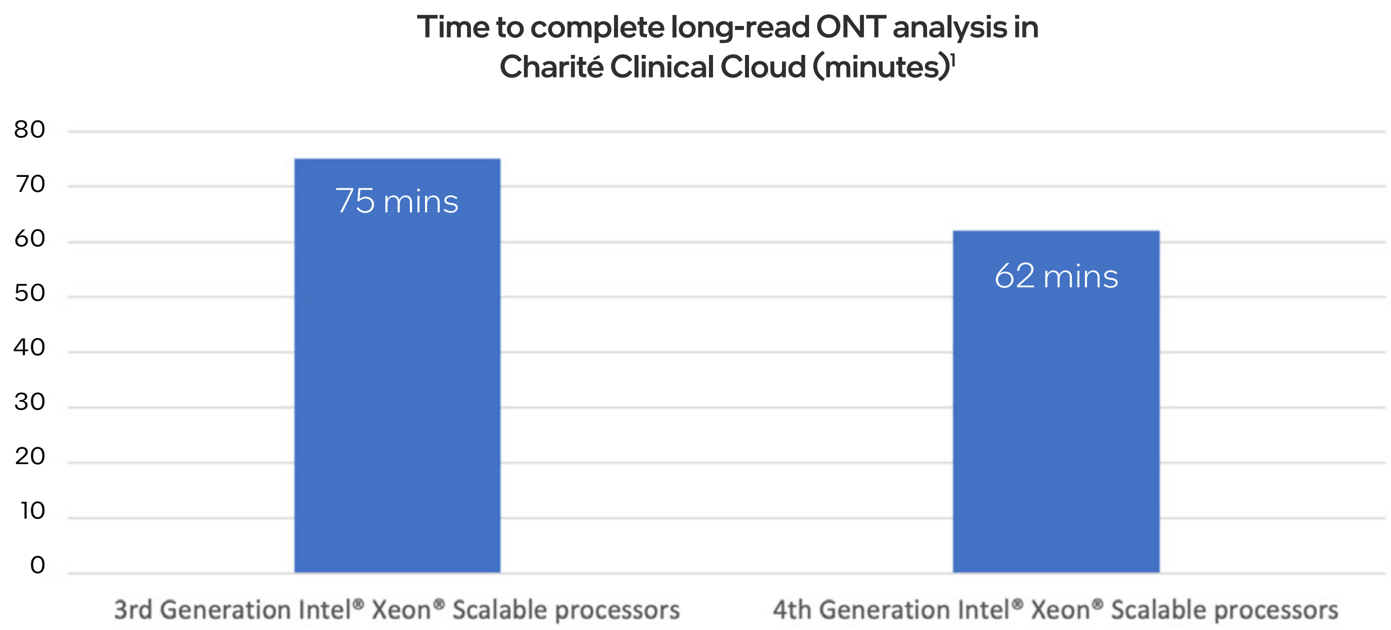

Figure 3. Comparing the genomic analysis performance of Intel® Xeon® Scalable processors gen-over-gen reveals a significant advantage for the latest generation.[1]

Average Time to Complete Test in Minutes	
Industry-standard workflow on standalone workstation	876
Sentieon workflow with 3rd Generation Intel® Xeon® Scalable processors on Charité cloud	75
Sentieon workflow with 4th Generation Intel® Xeon® Scalable processors on the Charité cloud	62

Figure 4. Comparing the average speeds to complete the genomic analysis shows the potential impact that Intel® Xeon® Scalable processors could have on bioinformatics use cases.

This testing also shows that 4th Gen Intel Xeon Scalable processors deliver a 21% advantage compared to the previous generation.[1] This means that as each generation of Intel Xeon Scalable processors become available in the cloud, bioinformatics applications do not require a major capital expenditure to get a boost in performance.

Conclusion

Analysis of data from DNA sequencing offers an unprecedented window into health and unlocks the secrets of biology. By making the analysis of NGS data more available and more cost-effective, this solution could revolutionize how the medical and research communities employ genomic data to improve patient outcomes.

When in the hands of skilled bioinformaticians, ultra-fast long-read scRNA-seq analysis has the potential to lead to new breakthroughs in genomic research. However, the most important achievement of this proof of concept are the levels of trust, scalability and cost-effectiveness that only Intel can offer.

Find the solution that's right for your organization.

Contact your Intel representative or visit **Intel.com/healthcare**.

About Intel

Intel strives to enrich the lives of everyone on the planet by improving the computing capabilities that make data-driven healthcare possible and speed the delivery of transformative healthcare technologies.

Explore 4th Generation Intel® Xeon® Scalable processors

About Sentieon

Sentieon develops highly optimized and accurate software and algorithms for bioinformatics applications, winning many precisionFDA awards and helping customers all over the world process their genomic data.

See how Sentieon is advancing bioinformatics

About Dell

With a team of professionals experienced in image processing, telecom, computational lithography, large-scale data mining, and bioinformatics, Sentieon develops highly optimized software and algorithms for bioinformatics applications.

Get more info on Dell PowerEdge Servers

About SVA

SVA System Vertrieb Alexander GmbH is one of the leading German system integrators. Founded in 1997 and based in Wiesbaden, Germany, the company has more than 2,700 employees at 27 branch offices all over Germany. The corporate objective of SVA is the combination of high-quality IT products from different vendors with the project know-how and flexibility of SVA to achieve optimal solutions for customers.

Explore the broad expertise of SVA

About Charité BIH

The Berlin Institute of Health (BIH) at Charité is dedicated to improving healthcare through medical translation. The company's mission is to transfer biomedical research findings into novel approaches to prevent, diagnose, and treat patients to greater effect.

Learn more about the work of Charité BIH

---

*4th Gen Intel® Xeon® Scalable processors versus 3rd Gen Intel® Xeon® Scalable processors

[1] Testing by Charité BIH on the Charité Clinical Cloud over one month (01/06/2023 – 01/07/2023) using 3 COVID-19 nasopharyngeal samples sequenced with ONT PROMETHION (FASTQ.GZ, 54GB, 69 million reads; FASTQ.GZ, 34GB, 44 million reads; FASTQ.GZ, 27GB, 40 million reads). ONT Sockeye software for comparisons was downloaded from https://github.com/nanoporetech/sockeye/releases (v0.2.1)

- Industry-standard workflow: 1-node workstation, 1x Intel® Xeon® CPU Gold 6132 , 2.6G, 14C/28T, GT/s, 19.25MB Cache, Turbo 3.7G, HT (140W), 8x 32GB/ 2666 total DDR4 memory

- 3rd Generation Intel® Xeon® processor: 2-node server, Intel® Xeon Platinum 8368 2.4G, 38C/76T, 11.2GT/s, 57MB Cache, Turbo 3.4G, HT (270W), 8x 32GB/8x 64GB 3200MT/s total DDR4 memory, 3x 1.92TB SSD SATA, 2x 100GbE SFP+PCIe, OS Ubuntu,

- 4th Generation Intel® Xeon® processor: pre-production 2-node server, Intel® Xeon® Platinum 8480+ 2G, 56C/112T, 16GT/s, 105MB Cache, Turbo 3.80G , HT (350W), 16x 64GB 4800MT/s total DDR5 memory, 1x 960MB SSD SATA, 2x 100GbE SFP+PCIe x8, OS Ubuntu

[2] Configuration details for a more cost-efficient server architecture: (see below)

[3] 4th Generation Intel® Xeon® Scalable processors vs. 3rd Generation Intel® Xeon® Scalable processors. See G1 at intel.com/processorclaims: 4th Gen Intel Xeon Scalable processors. Results may vary.

[4] 4th Generation Intel® Xeon® Scalable processors vs. 3rd Generation Intel® Xeon® Scalable processors. See E9 at intel.com/processorclaims: 4th Gen Intel Xeon Scalable processors. Results may vary.

ResNet50 Image Classification:

New Configuration: 1-node, 2x pre-production 4th Gen Intel® Xeon® Scalable 8490H processor (60 core) with Intel® Advanced Matrix Extensions (Intel AMX), on pre-production SuperMicro SYS-221H-TNR with 1024GB DDR5 memory (16x64 GB), microcode 0x2b0000c0, HT On, Turbo On, SNC Off, CentOS Stream 8, 5.19.16-301.fc37.x86_64, 1x3.84TB P5510 NVMe, 10GbE x540-AT2, Intel TF 2.10, AI Model=Resnet 50 v1_5, best scores achieved: BS1 AMX 1 core/instance (max. 15ms SLA), using physical cores, tested by Intel November 2022.

Baseline: 1-node, 2x production 3rd Gen Intel® Xeon® Scalable 8380 processor (40 cores) on SuperMicro SYS-220U-TNR , DDR4 memory total 1024GB (16x64 GB), microcode 0xd000375, HT On, Turbo On, SNC Off, CentOS Stream 8, 5.19.16-301.fc37.x86_64, 1x3.84TB P5510 NVMe, 10GbE x540-AT2, Intel TF 2.10, AI Model=Resnet 50 v1_5, best scores achieved: BS1 INT8 2 cores/instance (max. 15ms SLA), using physical cores, tested by Intel November 2022.

For a 50 server fleet of 3rd Gen Intel Xeon 8380 (RN50 w/DLBoost), estimated as of November 2022: CapEx costs: $1.64M OpEx costs (4 year, includes power and cooling utility costs, infrastructure and hardware maintenance costs): $739.9K Energy use in kWh (4 year, per server): 44627, PUE 1.6 Other assumptions: utility cost $0.1/kWh, kWh to kg CO2 factor 0.42394.

For a 17 server fleet of 4th Gen Intel Xeon 8490H (RN50 w/AMX), estimated as of November 2022: CapEx costs: $799.4K OpEx costs (4 year, includes power and cooling utility costs, infrastructure and hardware maintenance costs): $275.3K Energy use in kWh (4 year, per server): 58581, PUE 1.6 Other assumptions: utility cost $0.1/kWh, kWh to kg CO2 factor 0.42394.

RocksDB New Configuration:

1-node, 2x pre-production 4th Gen Intel® Xeon® Scalable 8490H processor (60 cores) with integrated Intel In-Memory Analytics Accelerator (Intel IAA), on pre-production Intel platform and software, HT On, Turbo On, Total Memory 1024GB (16x64GB DDR5 4800), microcode 0x2b0000a1, 1x3.84TB P5510 NVMe, 10GbE x540-AT2, Ubuntu 22.04.1 LTS, 5.18.12-051812-generic, QPL v0.2.1,accel-config-v3.4.6.4, ZSTD v1.5.2, RocksDB v6.4.6 (db_bench), tested by Intel November 2022. Baseline: 1-node, 2x production 3rd Gen Intel Xeon Scalable 8380 Processor (40 cores) on SuperMicro SYS-220U-TNR , HT On, Turbo On, SNC Off, Total Memory 1024GB (16x64GB DDR4 3200), microcode 0xd000375, 1x3.84TB P5510 NVMe, 10GbE x540-AT2, Ubuntu 22.04.1 LTS, 5.18.12-051812-generic, ZSTD v1.5.2, RocksDB v6.4.6 (db_bench), tested by Intel November 2022.

For a 50 server fleet of 3rd Gen Intel® Xeon® 8380 (RocksDB), estimated as of November 2022: CapEx costs: $1.64M OpEx costs (4 year, includes power and cooling utility costs, infrastructure and hardware maintenance costs): $677.7K Energy use in kWh (4 year, per server): 32181, PUE 1.6 Other assumptions: utility cost $0.1/kWh, kWh to kg CO2 factor 0.42394.

For a 18 server fleet of 4th Gen Intel Xeon 8490H (RockDB w/IAA), estimated as of November 2022: CapEx costs: $846.4K OpEx costs (4 year, includes power and cooling utility costs, infrastructure and hardware maintenance costs): $260.6K Energy use in kWh (4 year, per server): 41444, PUE 1.6 Other assumptions: utility cost $0.1/kWh, kWh to kg CO2 factor 0.42394 OpenFOAM New Configuration: 1-node, 2x pre-production 4th Gen Intel Xeon CPU Max Series (56 cores) on pre-production Intel platform and software, HT On, Turbo On, SNC4 mode, Total Memory 128 GB (8x16GB HBM2 3200MT/s), microcode 0x2c000020, 1x3.5TB INTEL SSDPF2KX038TZ NVMe, CentOS Stream 8, 5.19.0-rc6.0712. intel_next.1.x86_64+server, OpenFOAM 8, Motorbike 20M @ 250 iterations, Motorbike 42M @ 250 iterations, Tools: ifort:2021.6.0, icc:2021.6.0, impi:2021.6.0, tested by Intel December 2022. Baseline: 1-node, 2x production 3rd Gen Intel Xeon Scalable 8380 Processor (40 cores) on SuperMicro SYS-220U-TNR, HT On, Turbo On, 512GB (16x32GB DDR4 3200 MT/s), microcode 0xd000375, 1x2.9TB INTEL SSDPE2KE032T8 NVMe, CentOS Stream 8, 4.18.0-408.el8.x86_64, OpenFOAM 8, Motorbike 20M @ 250 iterations, Motorbike 42M @ 250 iterations, Tools: ifort:2021.6.0, icc:2021.6.0, impi:2021.6.0, tested by Intel December 2022.

For a 50 server fleet of 3rd Gen Intel Xeon 8380 (OpenFOAM), estimated as of December 2022: CapEx costs: $1.50M OpEx costs (4 year, includes power and cooling utility costs, infrastructure and hardware maintenance costs): $780.3K Energy use in kWh (4 year, per server): 52700, PUE 1.6 Other assumptions: utility cost $0.1/kWh, kWh to kg CO2 factor 0.42394.

For a 16 server fleet of Intel Xeon CPU Max Series 56 core, estimated as of December 2022: CapEx costs: $507.2K OpEx costs (4 year, includes power and cooling utility costs, infrastructure and hardware maintenance costs): $274.9K Energy use in kWh (4 year, per server): 74621, PUE 1.6 Other assumptions: utility cost $0.1/kWh, kWh to kg CO2 factor 0.42394.