

Intel data center CPUs set new benchmark for advanced bioinformatics

A successful case study in genomic analysis demonstrates tremendous potential to make bioinformatics more insightful, efficient and easy to scale.



Executive Summary

Intel forged a collaboration between innovators in bioinformatics to demonstrate that Intel® Xeon® Scalable processors can help make genomic analysis more scalable and accessible. The resulting proof-of-concept analytic pipeline, running on Dell PowerEdge Servers with Intel data center CPUs, showed stunning performance for processing genomic data.¹ Leveraging optimizations for Intel x86 architecture, the pipeline delivered a gene expression matrix from single-cell long-read sequencing data 14x faster than a standard workflow.¹



The proof-of-concept genomic analysis pipeline running on 4th Generation Intel® Xeon® Scalable processors delivered a result **14x faster** than an industry standard pipeline.¹



Solving the problems of scalability and cost

Scaling genomic analysis applications has been limited by demanding computational workloads that generally require specialized HPC solutions. This has led to a high cost-per-sample as well as a significant carbon footprint. However, it also presents bioinformatics applications with a new path to scalability and efficiency.

Proof of concept pipeline delivers exceptional speed

To find a new way to make genomic insights more accessible, Intel teamed up with a premier research institute, the Berlin Institute of Health at Charité (Charité BIH), and precision medicine focused bioinformatics software developer, Sentieon. This collaboration developed an optimized analytic pipeline that performed analysis on long-read next-generation sequencing (NGS) data using standard data center CPUs from Intel. The proof-of-concept pipeline built by Sentieon cut the standard time-to-result by more than 14x.¹

In further testing, 4th Generation Intel® Xeon® Scalable processors showed a 21% advantage in speed over 3rd Generation Intel® Xeon® Scalable processors. This offers tremendous potential to bring economies of scale to bioinformatics as each successive generation of Intel technology helps make genomic analysis software faster and more efficient.¹

New 4th Gen Intel® Xeon® Scalable processors increased the speed of genomic analysis by 21% compared to 3rd Gen Intel® Xeon® Scalable processors on Sentieon's optimized workflow.¹

Achieving far more than speed

This test revealed more than raw power and performance. The highest degree of accuracy was maintained, and thanks to built-in features on Intel Xeon processors, this new approach to genomic analysis offers advancements in trust, scalability and cost effectiveness.

Spotlight on The Berlin Institute of Health at Charité

The Berlin Institute of Health at Charité (Charité BIH) has a special mission to push the boundaries of medical translation to improve health for everyone. Rather than developing specialized, purpose-built solutions for analytics and bioinformatics, Charité BIH sees the future of healthcare in the cloud. At the heart of this effort is the Charité Clinical Cloud. It promises to offer health data platforms that bring hospitals, doctors, researchers and other specialists together while serving the entire healthcare value chain.

This focus on making advanced healthcare technology more available is what has driven the company's long-standing collaboration with Intel. Charité BIH and Intel have worked together to optimize workflows, pipelines and data center infrastructure for healthcare research. Together, they have achieved breakthroughs such as [identifying a method COVID-19 uses](#) to target the body's cells for infection. The Charité Clinical Cloud enables bioinformaticians to scale dynamically, move workloads around and customize applications. It also centralizes complex security and compliance measures. With offerings like this, Intel and Charité BIH want to maximize speed, performance and patient experience quality, while bringing down cost and complexity.

Challenge: Deep genomic insights have been hard to scale

Our genome has the key to so many of the secrets about how our bodies work and why we get sick. Employing genomic analysis in diagnostics, treatment and research offers opportunities to tailor treatments and improve results for patients. Studying genomic data can also improve population health and enable early detection and prevention measures that reduce healthcare costs.

Genomic analysis, however, is costly and resource-intensive to perform. Closed and specialized hardware systems have hampered scalability. Together with its global ecosystem of collaborators, Intel is advancing the performance, efficiency and openness of bioinformatics applications to help make them more powerful, cost effective and accessible.

The global bioinformatics market is estimated to reach **\$45.2 billion** by 2030, growing at a **CAGR of 11.9%** from 2022-2030.²

NGS is changing genomics, but presents tradeoffs

NGS has introduced cutting-edge processes that dramatically speed up DNA or RNA sequencing. Typically, NGS scales throughput by running short-read sequencing (SRS) on bulk genomic data. An average gene transcript has thousands of base pairs (bp), but SRS reads around only 150 bp. This is enough data to identify specific genes and its expression that could be related to disease or an underlying health condition. Nonetheless, SRS leaves out much of what's occurring inside the cell or organ, which can make it hard to detect splicing, breaks, fusions, or structural variants in the cell that can be disease-causing.

Pursuing a more precise picture of health

Long-read single-cell RNA sequencing (scRNA-seq), however, 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.

The challenges of long-read scRNA-seq hold back scalability

Oxford Nanopore Technologies (ONT) has advanced the availability of long-read scRNA-seq with high sequencing throughput, generating large volumes of data. ONT sequencers also generate very large volumes of data. Analytics pipelines running ONT libraries often require dedicated hardware acceleration via a GPU or FPGA. This computational complexity has made long-read scRNA-seq prohibitively expensive and time-consuming. It also limits scalability in bioinformatics and precision medicine due to reliance on specialized, inflexible infrastructure and closed software.

Analyzing long-read scRNA-seq data also generates a significant carbon footprint as it requires computing resources that consume large amounts of energy. It can be hard for healthcare organizations to pursue carbon emission objectives, such as net zero compute, while also developing and using advanced bioinformatics.

Solution: Genomic analysis pipeline optimized for Intel data center CPUs

Bringing about the future of bioinformatics means overcoming the performance, efficiency and scalability issues presented by long-read sequencing data. Running genomic analysis workloads on data center CPUs—as opposed to specialized high-performance computing (HPC) workstations — can reduce cost-per-sample through economies of scale and efficiency advantages.

Bringing innovators together to improve healthcare

Intel has collaborated with Charité BIH for years to help advance its bioinformatics initiatives. Intel saw that another of its collaborators, Sentieon, could offer Charité BIH cutting-edge software that could advance the capabilities of the Charité Clinical Cloud.

Intel introduced the teams, and Sentieon developed a new genomic analysis pipeline and workflow for long-read sequencing data. This workflow is an improvement over the recently released open-source solution (Sockeye) implemented by ONT, and optimizations available for Intel architecture helped amplify performance and efficiency.

Evaluating the power of Intel Xeon Scalable processors for advanced bioinformatics

The test sequencing was performed on Intel Xeon Scalable processors. No specialized hardware acceleration was employed. The Intel CPUs employed for the proof of concept are those available in standard cloud instances across the globe.

Streamlined workflow, optimized pipeline

Sentieon designed the pipeline using the readily available Sockeye ONT pipeline as a starting point. The pipeline also combined the adapter scanning and barcode extraction stages into a single tool inside the Sentieon software. This removed a data processing bottleneck and eliminated the reliance on vsearch for adapter identification. ONT long-reads are aligned using Sentieon’s minimap2, which delivered greatly increased performance after being optimized for Intel architecture.

Unique Molecular Identifiers (UMI) and barcode tasks such as extraction, deduplication and correction present a major part of the workload challenge. The Sentieon workflow takes advantage of the power of Intel Xeon Scalable processors to run scanning, extraction and alignment

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

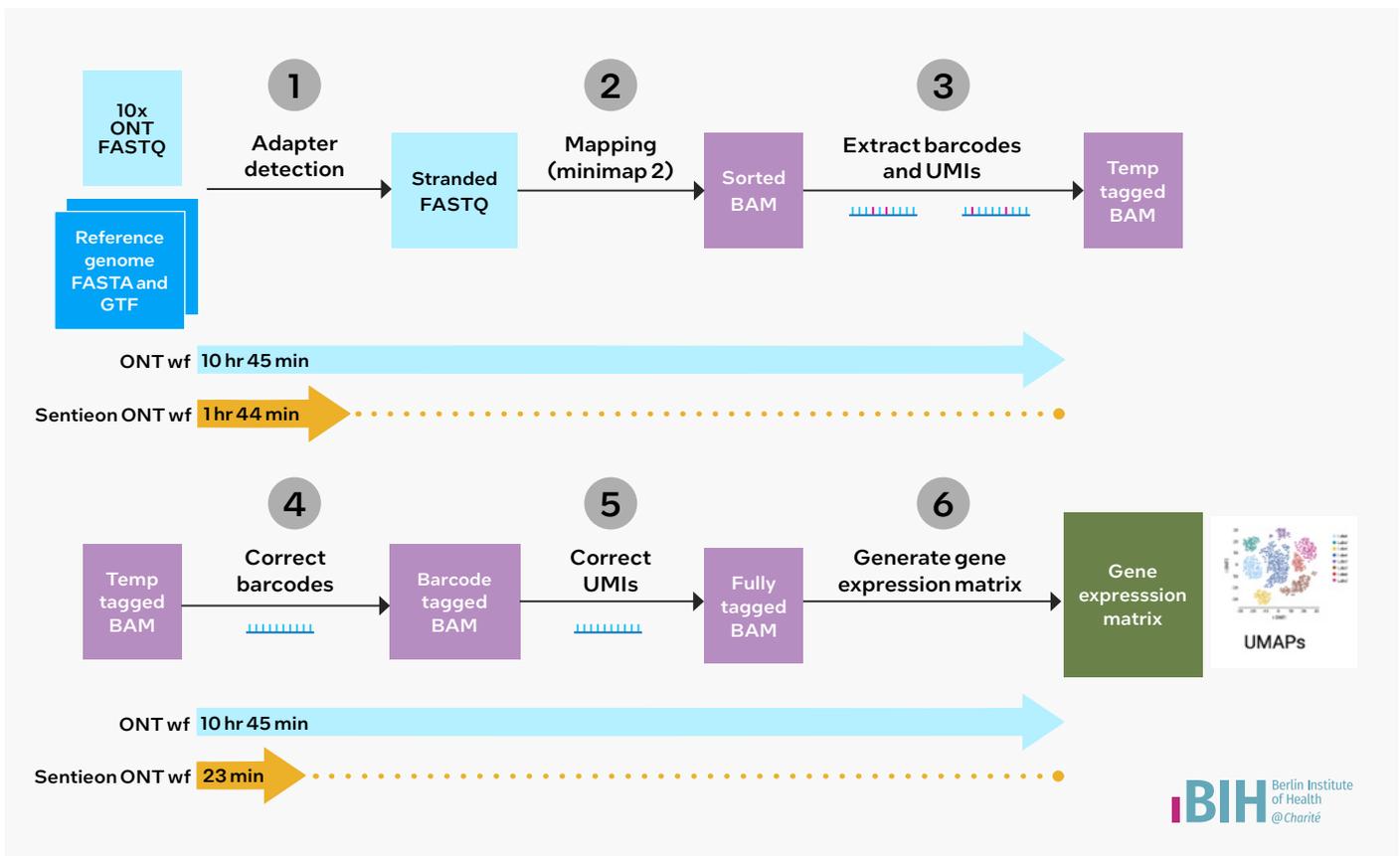


Figure 1. The custom sequencing software pipeline developed by Sentieon can dramatically accelerate long-read scRNA-seq on standard cloud infrastructure.

Sentieon Proof of Concept - Accelerated ONT wf-single-cell in the Clinical Cloud

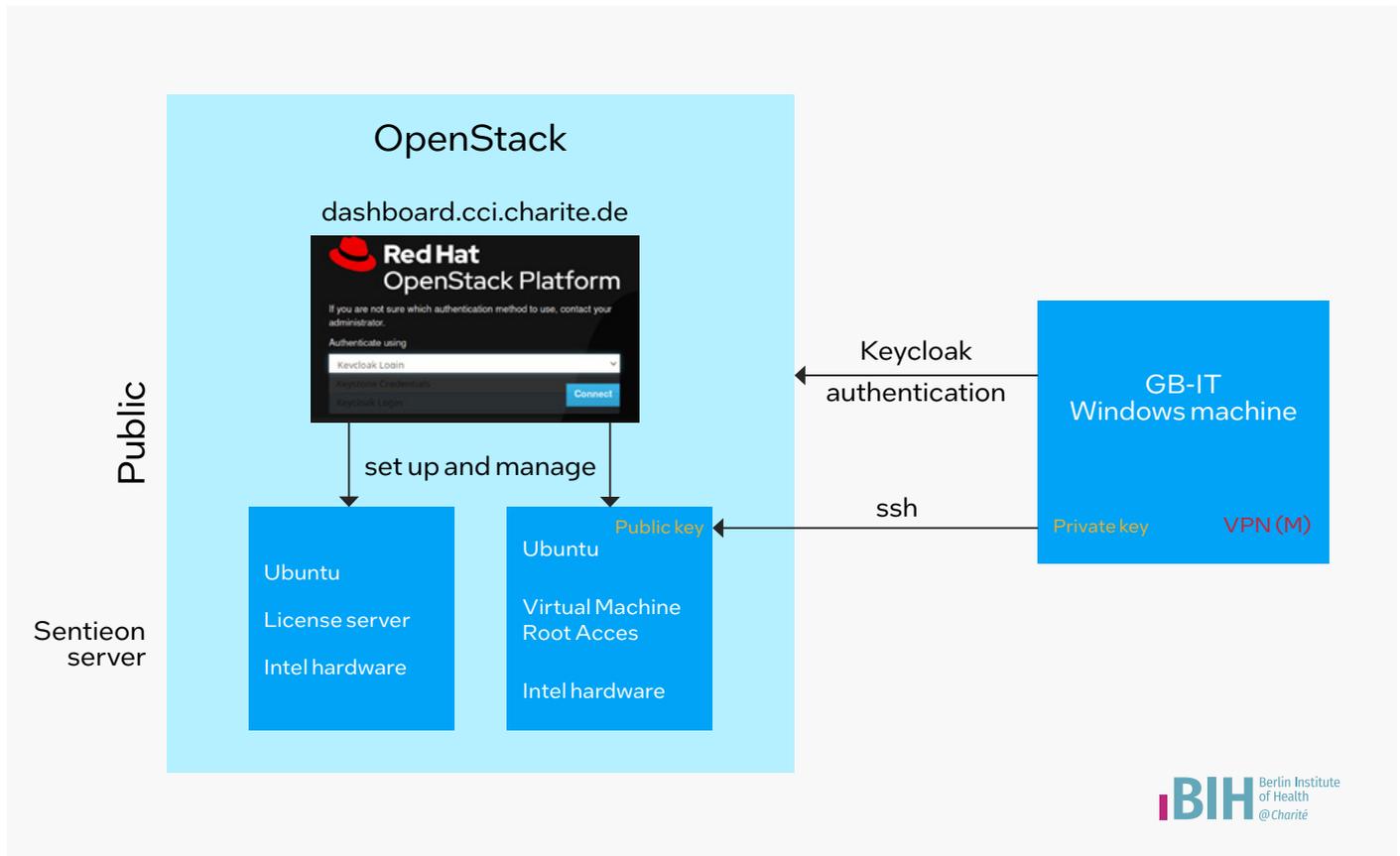


Figure 2. The proof-of-concept software stack running on Dell PowerEdge Servers with Intel® Xeon® Scalable processors in the Charité Clinical Cloud was designed by Sentieon and implemented by SVA.

concurrently. This reduced read/write time and eliminated the need for intermediary files. Sentieon’s experience in UMI handling also enabled the company to write code that improves deduplication performance by removing dependency on external tools for UMI handling.

The success of the initiative was enhanced by systems integrator SVA, which led the process of building the Dell PowerEdge Server to run Sentieon’s software pipeline and integrating it into Charité BIH’s private cloud.

Results: Dramatically faster times and the promise of efficiency and scalability

Benchmarking the proof-of-concept analytic pipeline from Sentieon against a standard ONT workflow demonstrated its capability to deliver results in much less time. In successive tests with different ONT long-read scRNA-seq data sets, the Sentieon workflow installed on 4th Gen Intel Xeon processors was 14x faster than running the standard ONT Sockeye workflow.¹ It was the fastest time to produce a gene expression matrix from long-read sequencing data that the researchers performing the testing had ever seen.

“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

Average time to complete long-read ONT analysis (minutes)¹

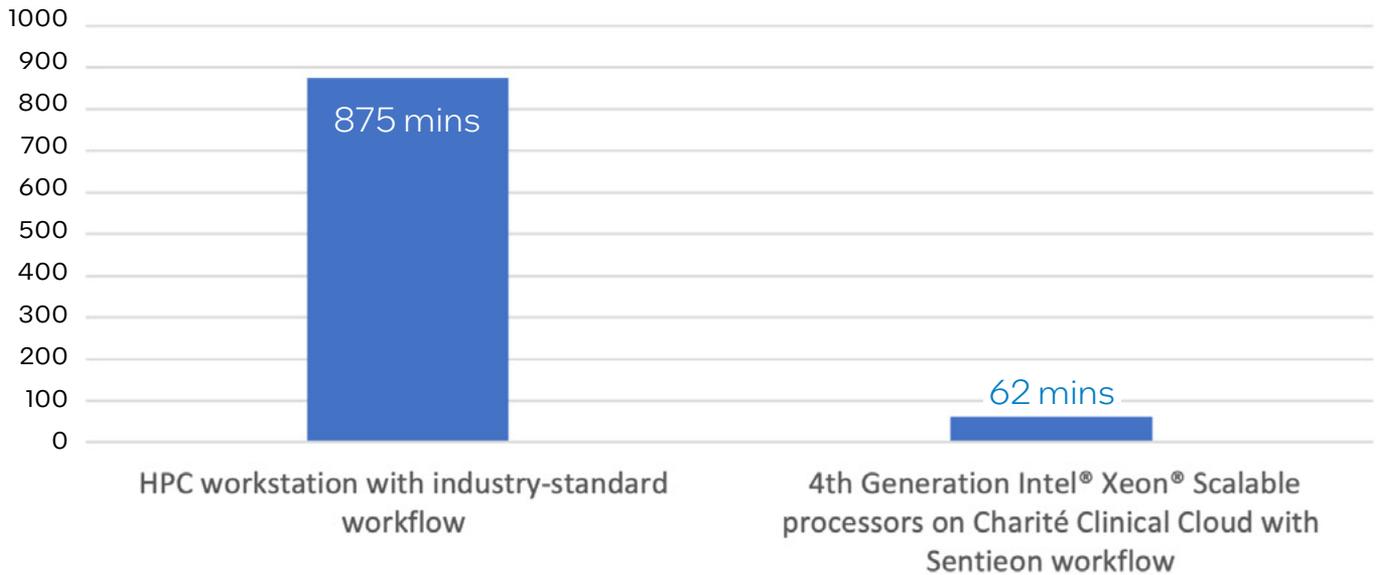


Figure 3. 4th Generation Intel® Xeon® Scalable processors delivered incredible speeds in genomic analysis on ONT data.¹

Generational gains show potential for long-term scaling

The case study also tested how 3rd Gen Intel Xeon Scalable processors, available throughout the world in the cloud, performed on the streamlined workflow. Thanks to Sentieon’s optimizations for Intel server architecture, 3rd Gen Intel Xeon Scalable processors completed the genomic analysis task nearly 12x faster than the benchmark HPC workstation.¹

Performance and efficiency 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 that allows you to work with larger datasets with less latency.^{3*} Highlights include:

- DDR5 memory and PCIe 5.0 for increased memory and I/O bandwidth
- Higher clock speeds and up to 60 cores to handle bigger tasks with fewer racks
- Built-in accelerators that can speed up key workloads like AI and analytics
- Up to 53% general purpose compute performance over 3rd Generation Intel® Xeon® Scalable processors.^{3*}

Average Time to Complete Test in Minutes	
Industry-standard workflow on HPC 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. Sentieon’s optimized workflow for Intel® Xeon® Scalable processors showed potential to perform genomic analysis with scalable infrastructure and fewer resources.

Time to complete long-read ONT analysis in Charité Clinical Cloud (minutes)¹

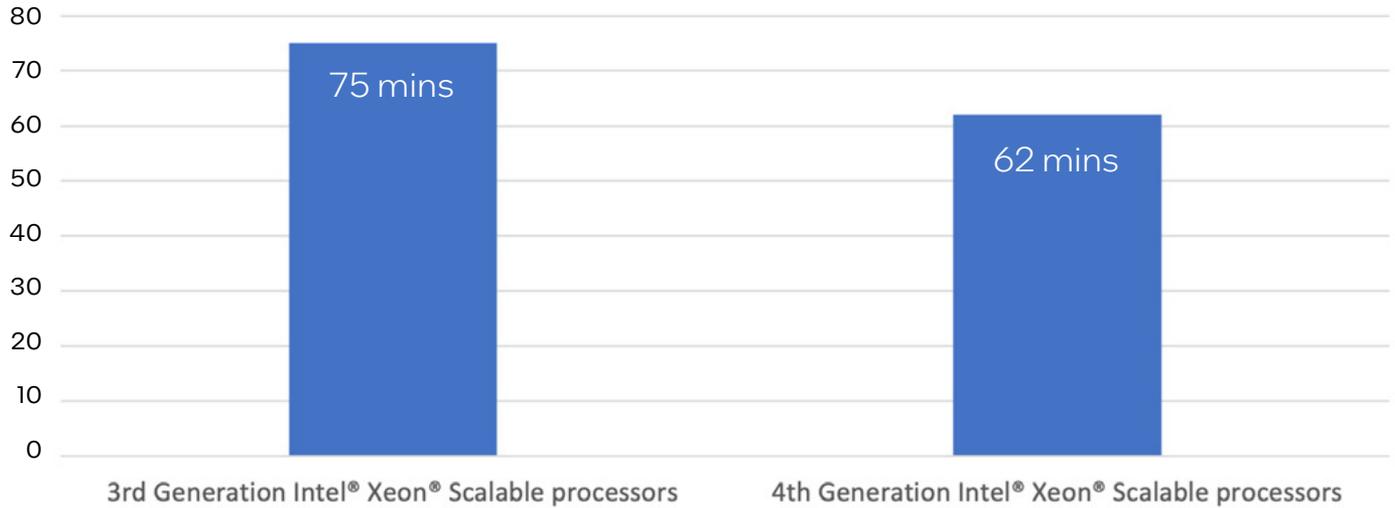


Figure 5. Comparing the genomic analysis performance of Intel® Xeon® Scalable processors gen-over-gen reveals a significant advantage for the latest generation.¹

Furthermore, 4th Gen Intel Xeon processors demonstrated a 21% advantage over the previous generation.¹ This shows the potential for bioinformatics applications to gain performance and efficiency every time workloads move to cloud instances with the latest generation of Intel data center CPUs. With Intel, such migrations are typically seamless and require little effort or expense.

The speed to launch the next generation of genomic analysis

Achieving this level of performance for long-read sequencing data analysis can enable hospitals and research institutions to get results at speeds that can change the way they work. What's more, delivering genomic analysis over the cloud can allow practitioners to make greater use of NGS data.

Conclusion: Maximize speed, efficiency, scalability and trust

"Omics" technologies are the future of healthcare. Enabling long-read sequencing in the Charité Clinical Cloud on standard hardware generates potential use cases that could have a huge impact. Genomic analysis can power precision medicine that offers personalized treatments. Screening populations of children can lead to early disease detection. Genomic data can allow virologists to see the impact a virus has on a molecular level.

Enabling fast and efficient genomics analysis on standard Intel data center CPUs offers a pathway to bring bioinformatics applications into the cloud. In a cloud environment powered by Intel technology, bioinformatics workloads can be easily migrated and accessed from virtually anywhere. What's more, data centers present an opportunity to centralize security, sustainability and regulatory compliance, which can eliminate major hurdles to launching an application that uses health data. Giving practitioners and researchers tools capable of inspiring their trust can empower everyone in the healthcare value chain to deliver better care.

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](#)

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](#)

Find the solution that's right for your organization.

Contact your Intel representative or visit [Intel.com/healthcare](https://www.intel.com/healthcare).



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

¹1. 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

²2 Global Bioinformatics Market to Reach \$45.6 Billion by 2030, Yahoo Finance, September 27, 2021, <https://finance.yahoo.com/news/global-bioinformatics-market-reach-45-155000066.html>

³ Configuration details for a more cost-efficient server architecture:

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.

Intel® technologies may require enabled hardware, software, or service activation. Your costs and results may vary.

No product or component can be absolutely secure.

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.

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

©2023 Intel Corporation 0923/GL/CAT/PDF PleaseRecycle 356708-001EN