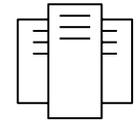


IBM LinuxONE Emperor 4 – Exceptional by Design

Reduce your carbon footprint by 75%¹ and achieve massive scalability by migrating your Linux® workloads from hundreds of x86 servers to a single IBM LinuxONE Emperor 4™. Reduce your TCO and carbon footprint further by leveraging integrated data compression to reduce your storage needs on a 5:1 ratio. Deliver exceptional experiences through excellent performance and engineered 99.99999% resiliency. Leverage innovative on-chip AI-inferencing and quantum-safe technologies to accelerate insights in every customers interaction and protect against “harvest data now to decrypt later” security threats.



Capability	Value	Proof points and use cases	LinuxONE Emperor 4	x86	Public Cloud
Sustainability	Sustainability features to reduce Green House Gas (GHG) emissions and costs around energy consumption, floor space utilization and cooling. A better option vs comparable workloads on x86 or public cloud with proven TCO studies and consolidation.	An IBM LinuxONE Emperor 4 can consolidate hundreds x86 Linux cores with 75% better energy efficiency & 50% less floor space. ¹ IBM internal tests show that when running WebSphere® and Db2® workloads, IBM LinuxONE Emperor 4 requires 16X fewer cores than compared x86 servers. If you scale this up to a complete IT solution this means when running this workload, the IBM LinuxONE Emperor 4 Max 125 would be doing the work of about 2000 cores of the compared x86 servers. ² Run the Yahoo Cloud Serving Benchmark (YCSB) on MongoDB without sharding on IBM LinuxONE Emperor 4 with 6 cores in total and achieve the same throughput as on MongoDB with 4 shards on compared x86 systems with 144 cores in total, which provides a 24:1 core consolidation ratio in favor of IBM LinuxONE Emperor 4. ³	✓		
Massive Scalability	Grow your workloads without disrupting your service. Utilize capacity on demand while delivering high QoS.	A single IBM LinuxONE™ core often does the work of 16 x86 cores. The IBM LinuxONE Emperor 4 can execute up to 20 billion HTTPS transactions per day with OLTP microservice applications running on the Red Hat® OpenShift® container platform. The LinuxONE Emperor 4 boasts 200 cores and up to 40TB RAM ⁴ so you can add more processor capacity or memory on-demand without physically adding more racks or measurably increasing your carbon footprint. Scale up your I/O intensive Linux applications and protect your data at rest with up to 12 million read-only I/O operations per second and 10 million R/W operations per second to an encrypted filesystem with FCP attached storage. ⁵	✓		
Performance	Implement containerized cloud applications and deliver exceptional customer experiences with blazing fast and predictable performance	Deliver exceptional customer experiences with blazing fast and predictable performance. New cache architecture with 1.5X more cache reduces latency and improves performance. New 7nm processor chip technology running at 5.2 GHZ with encrypted memory. New FICON Express32S has 2X the bandwidth as 16 Gbps adapters for faster data transfer systems.	✓		
Integrated Data Compression	Compress your data 5:1 using integrated hardware data compression	Reduce your disk and tape storage space by 80% to reduce your carbon footprint, data center space and lower your Total Cost of Ownership (TCO)	✓		
Quantum-Safe Security	Leverage an industry-first quantum-safe Linux server to protect against “collect now and decrypt later” quantum computing threats	Mitigate the risk of a cyber attack now and in the future, by using advanced Quantum-safe cryptography that makes your data more resilient to decrypting. Secure boot technology helps protect IBM LinuxONE Emperor 4 firmware from quantum attacks through a built-in dual signature scheme with no changes required. IBM LinuxONE Emperor 4 quantum-safe APIs enable clients to use quantum-safe cryptography along with classical cryptography on existing and new applications. Quantum-safe APIs will enable clients to build hybrid quantum-safe key exchange systems with Crypto Express 8S protection of the keys.	✓		
Continuous Compliance	IBM LinuxONE Security and Compliance Center and IBM LinuxONE Emperor 4 make regulatory compliance easier and reduce human errors and risk	IBM LinuxONE Security and Compliance Center on the IBM LinuxONE Emperor 4 mitigates regulatory risk and reduces audit preparation time. Sponsored user client reporting of projected savings after implementing the solution was a reduction of a month to one week and reduced the number of skilled resources needed for audit preparation functions by 40%. ⁶	✓		
AI Acceleration	Embed AI into every customer interaction at scale; low latency accelerates insights and reduces cost	Leverage IBM LinuxONE Emperor 4 on-chip AI acceleration to integrate inferencing easily into transactions with low latency and high performance to improve business results, reduce risks and deliver customer value with every interaction. IBM LinuxONE Emperor 4 is the only commercial enterprise class Linux server in the world with real-time latency optimized acceleration integrated into the microprocessor chip, designed to execute up to 300 Billion inference requests per day with only 1 ms latency. ⁷ Using one Integrated Accelerator for AI on an OLTP workload on IBM LinuxONE Emperor 4 matches the throughput of a compared remote x86 server running inferencing on 18 cores. ⁸	✓		

Disclaimers

1. Consolidating Linux workloads on 5 IBM LinuxONE Emperor 4 systems instead of running them on compared x86 servers under similar conditions can reduce energy consumption by 75%, space by 50%, and the CO2e footprint by over 850 metric tons annually. This is the equivalent of taking approximately 186 gas powered passenger vehicles off the road annually. **DISCLAIMER:** Compared 5 IBM LinuxONE Emperor 4 Max 125 model consists of three CPC drawers containing 125 configurable cores and two I/O drawers to support both network and external storage versus 192 x86 systems with a total of 10364 cores. IBM LinuxONE Emperor 4 power consumption was based on inputs to the IBM LinuxONE IBM Power Estimation Tool for a memo configuration. x86 power consumption was based on March 2022 IDC QPI power values for 7 Cascade Lake and 5 Ice Lake server models, with 32 to 112 cores per server. All compared x86 servers were 2 or 4 socket servers. IBM LinuxONE and x86 are running 24x7x365 with production and non-production workloads. Savings assumes a Power Usage Effectiveness (PUE) ratio of 1.57 to calculate additional power for data center cooling. PUE is based on Uptime Institute 2021 Global Data Center Survey (<https://uptimeinstitute.com/about-ui/press-releases/uptime-institute-11th-annual-global-data-center-survey>). Estimated annual CO2e reduction and gas- powered passenger vehicles driven for a year are based on the EPA GHG calculator (<https://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator>).
2. **DISCLAIMER:** This is an IBM internal study designed to replicate typical IBM customer workload usage in the marketplace. Results may vary. The core consolidation study targeted comparison of the following IBM LinuxONE and x86 servers: IBM LinuxONE Emperor 4 Max 125 system consists of three CPC drawers containing 125 configurable processor units (IFLs or zIIPs) and two I/O drawers to support both network and external storage. Lenovo ThinkSystem SR650 (2U) with two 2nd Gen Intel® Xeon® Platinum processors 2.1 GHz, 16 cores per CPU. Both the x86-based and LinuxONE solutions had access to the same storage array. The workloads consisted of a transactional application running on WebSphere Application Server and IBM DB2 simulating core online banking functions. The actual test results were extrapolated to the stated above x86 servers using IDC QPI metrics and IBM sizing methodology using the following assumptions on a typical IT environment of a banking client using x86 servers. The production IT environment has 16 x86 servers running at 50% average utilization. There are 48 x86 servers in the non-production IT environments: development (4 environments with 2 servers each, 8 servers total), development test environment (4 servers), system integration test environment (8 servers), performance test environment (16 servers), user acceptance test environment (4 servers), production fix test environment (8 servers). A typical average CPU utilization is 7% across all non-production environments. An equivalent LinuxONE Emperor 4 solution requires a single Max 125 server running at 85% average utilization across all IT environments separated using LPAR technology.
3. Run the Yahoo Cloud Serving Benchmark (YCSB) on MongoDB without sharding on IBM LinuxONE Emperor 4 with 6 cores in total and achieve the same throughput as on MongoDB with 4 shards on compared x86 systems with 144 cores in total, which provides a 24:1 core consolidation ratio in favor of IBM LinuxONE Emperor 4. **DISCLAIMER:** Performance results based on IBM internal tests running YCSB 0.10.0 benchmark (read-mostly) on MongoDB Enterprise Release 5.0.6 with 3-node replication. On IBM LinuxONE Emperor 4 MongoDB was setup without sharding but with two replicas. IBM LinuxONE configuration: LPAR with 4 dedicated cores and 2 LPARs with each 1 core, each with SMT and 128 GB memory, 1 TB FlashSystem 900. x86 config: 9 Intel® Xeon® Gold 5218 CPU @ 2.30GHz with Hyperthreading turned on, 192 GB memory, 1 TB local RAID5 SSD storage, RHEL 8.4 running MongoDB, driven remotely by YCSB using 2 x86 server with total 128 threads. Results may vary.
4. The IBM LinuxONE Emperor 4 can execute up to 20 billion HTTPS transactions per day with OLTP microservice applications running on the Red Hat® OpenShift® container platform. The Emperor 4 boasts 200 cores and up to 40TB RAM. **DISCLAIMER:** Performance result extrapolated from IBM internal tests running in an IBM LinuxONE Emperor 4 LPAR with 24 dedicated cores, 560 GB memory and DASD storage benchmark (<https://github.com/blueperf/acmeair-main-service-java>) on Red Hat OpenShift Container Platform (RHOC) 4.9 using RHEL 8.4 KVM on 4 RHOC compute nodes. The 4 Acme Air instances were running parallel, each driven remotely from JMeter 5.2.1 with 384 parallel users. The KVM guests with RHOC compute nodes were configured with 12 vCPUs and 64 GB memory each. The KVM guests with RHOC management nodes and RHOC infrastructure nodes were configured with 4vCPUs and 16 GB memory each. Results may vary.
5. With IBM LinuxONE Emperor 4, scale up your I/O intensive Linux applications and protect your data at rest with up to 12 million read-only I/O operations per second and 10 million R/W operations per second to an encrypted filesystem with FCP attached storage. **DISCLAIMER:** Performance result extrapolated from IBM internal tests running the fio 3.19 benchmark tool in an IBM LinuxONE Emperor 4 LPAR with 12 cores and 64 GB memory on RHEL 8.5 (SMT mode) using the XFS filesystem format with luks2 encryption and two FICON Express 32S features. The fio benchmarking tool was run with 128 parallel threads using 8 volumes on FS9200 equally distributed over the two nodes and a file size of 150GB on each volume. Results may vary.
6. **DISCLAIMER:** IBM does not ensure regulatory compliance. The intent is to provide a point in time statement of your current posture for a specific group of resources. The responsibility of ensuring systems are configured in accordance with regulatory controls is on the individual businesses who are using the IBM Z security and compliance Center and IBM does not take responsibility for any compliance oversights or penalties associated with data breaches. The survey consisted of 9 responses across 6 unique customers. Sourced from the IBM ZSCC Sponsor User Program and zDC.
7. **DISCLAIMER:** Performance result is extrapolated from IBM internal tests running local inference operations in an IBM z16 LPAR with 48 IFLs and 128 GB memory on Ubuntu 20.04 (SMT mode) using a synthetic credit card fraud detection model (<https://github.com/IBM/ai-on-X-fraud-detection>) exploiting the Integrated Accelerator for AI. The benchmark was running with 8 parallel threads each pinned to the first core of a different chip. The Iscpu command was used to identify the core-chip topology. A batch size of 128 inference operations was used. Results were also reproduced using a z/OS V2R4 LPAR with 24 CPs and 256GB memory on IBM z16. The same credit card fraud detection model was used. The benchmark was executed with a single thread performing inference operations. A batch size of 128 inference operations was used. Results may vary.
8. **DISCLAIMER:** Performance results is extrapolated from IBM internal tests running an OLTP workload with credit card transaction using the Credit Card Fraud Detection (<https://github.com/IBM/ai-on-z-fraud-detection>) model on IBM LinuxONE Emperor 4 vs running the OLTP workload (<https://github.com/IBM/megacard-standalone>) on IBM LinuxONE Emperor 4 and running inferencing on a remote x86 server running Tensorflow serving. IBM LinuxONE Emperor 4 configuration: Ubuntu 20.04 in an LPAR with 6 dedicated cores, 256 GB memory, and IBM FlashSystem 900 storage. x86 configuration: Ubuntu 20.04 on 18IceLake Intel® Xeon® Gold CPU @ 2.80GHz with Hyperthreading turned on, 1 TB memory, local SSDs. Results may vary.