

Об ИТ банка, обеспечивающей развитие

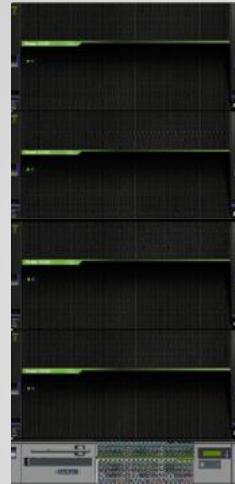
Алексей Перевозчиков
82189117@ru.ibm.com

IBM Power10 processor-based systems roadmap

2021

Power10

Enterprise
16 Socket System



Power E1080

- Best core performance
- Best memory bandwidth / core
- Best RAS capabilities

P9 HMC



1H

2H

2022

Power10

Enterprise
4-socket System



- Significant core density increase
- Significant memory bandwidth increase

Scale Out
1-2 Socket Systems



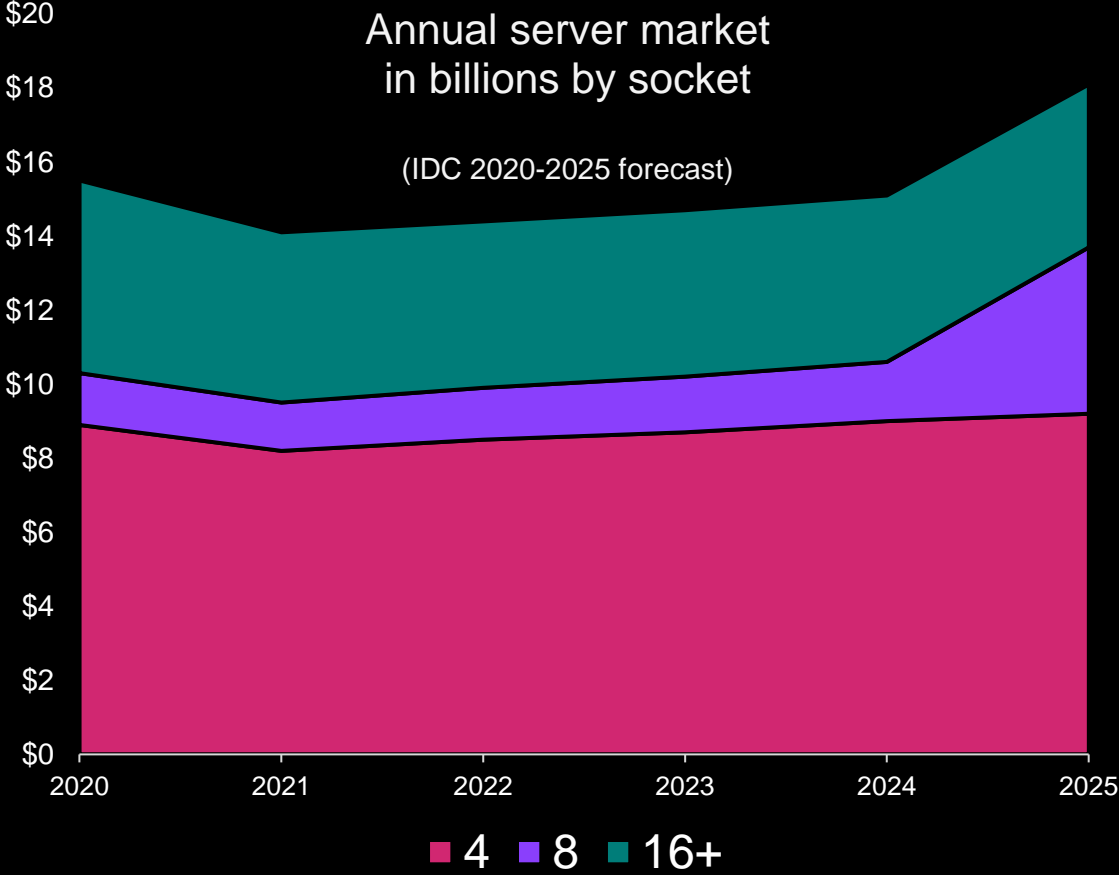
- Significant core density increase
- 2x increase in system memory bandwidth

“S10xx”

1H

2H

Marketplace opportunity



IBM Power addresses all areas of the market. Providing extreme performance per core, the Power E1080 is specifically ***architected and priced to compete with 4-socket and larger servers***

Что на кону? Стоит ли с этим связываться?

“Экономический эффект систем ИИ на наше общество к 2030 году составит 15 триллионов долларов. Ни один сектор экономики не останется в стороне.”

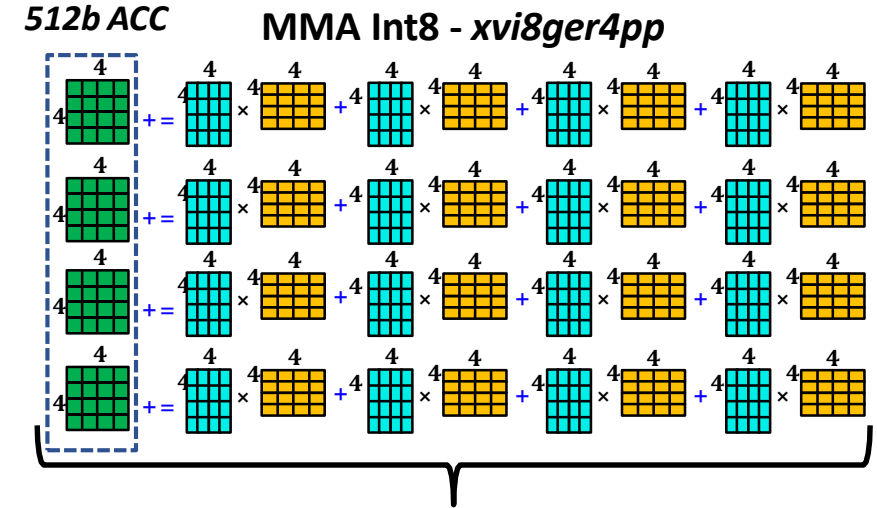
PwC report “Sizing the prize”

PwC, report “Sizing the prize”, Source: <https://preview.thenewsmarket.com/Previews/PWC/DocumentAssets/476830.pdf>



AI Infused Core: Inference Acceleration

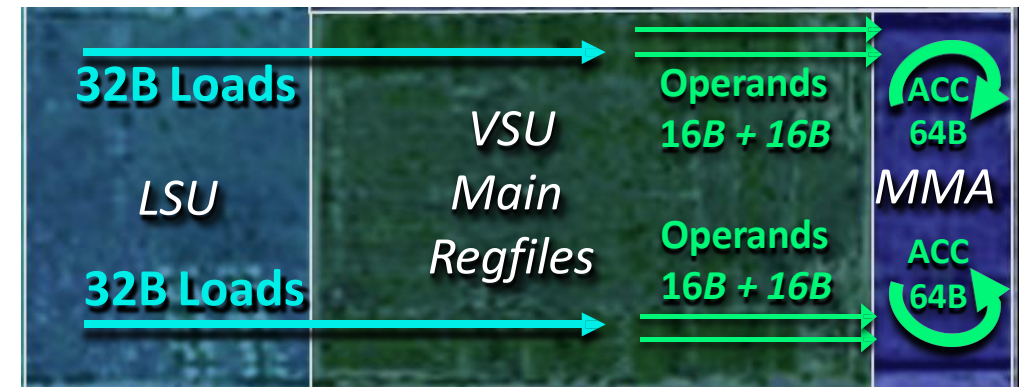
- **4x+ per core throughput**
- **3x -> 6x thread latency reduction (SP, int8)***
- **POWER10 Matrix Math Assist (MMA) instructions**
 - 8 512b architected Accumulator (ACC) Registers
 - 4 parallel units per SMT8 core
- **Consistent VSR 128b register architecture**
 - Minimal SW ecosystem disruption – no new register state
 - Application performance via updated library (OpenBLAS, etc.)
 - POWER10 aliases 512b ACC to 4 128b VSR's
 - Architecture allows redefinition of ACC
- **Dense-Math-Engine microarchitecture**
 - Built for data re-use algorithms
 - Includes separate physical register file (ACC)
 - 2x efficiency vs. traditional SIMD for MMA



4 per cycle per SMT8 core

Matrix Optimized / High Efficiency

Result data remains local to compute



Inference Accelerator dataflow (2 per SMT8 core)

* versus POWER9

AI Deployment Simplification w/ ONNX

Client Requirement –
Uniform Model Behavior Across Platforms

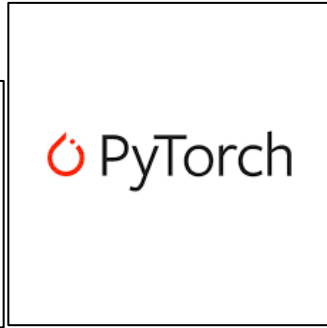
AC922 or x86

Common Model
Format, Common Runtime
(IBM, Community)



Business Processes
Hosted on IBM
Enterprise Servers

Train with Any Framework, Run with ONNX-Runtime on Any Platform



IBM Power E1080 – Memory Subsystem Highlights

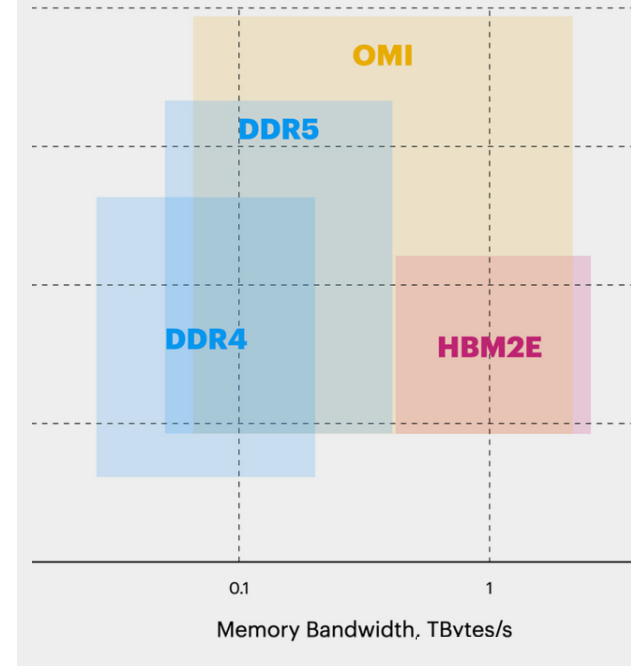
- ✓ New OMI (Open Memory Interface)
- ✓ New Differential DIMMs - DDIMMs
- ✓ 16 DDIMM slots per socket
- ✓ 64 DDIMM slots per drawer
- ✓ 4U DDIMMs - Enterprise buffer, N+1 power management design, and spare DRAM support
- ✓ DDIMM options
 - 32GB DDIMM (DDR4 @ 3200 Mbps)
 - 64GB DDIMM (DDR4 @ 3200 Mbps)
 - 128GB DDIMM (DDR4 @ 2933 Mbps)
 - 256GB DDIMM (DDR4 @ 2933 Mbps)
 - DDIMMs planned to carry forward to P11 systems



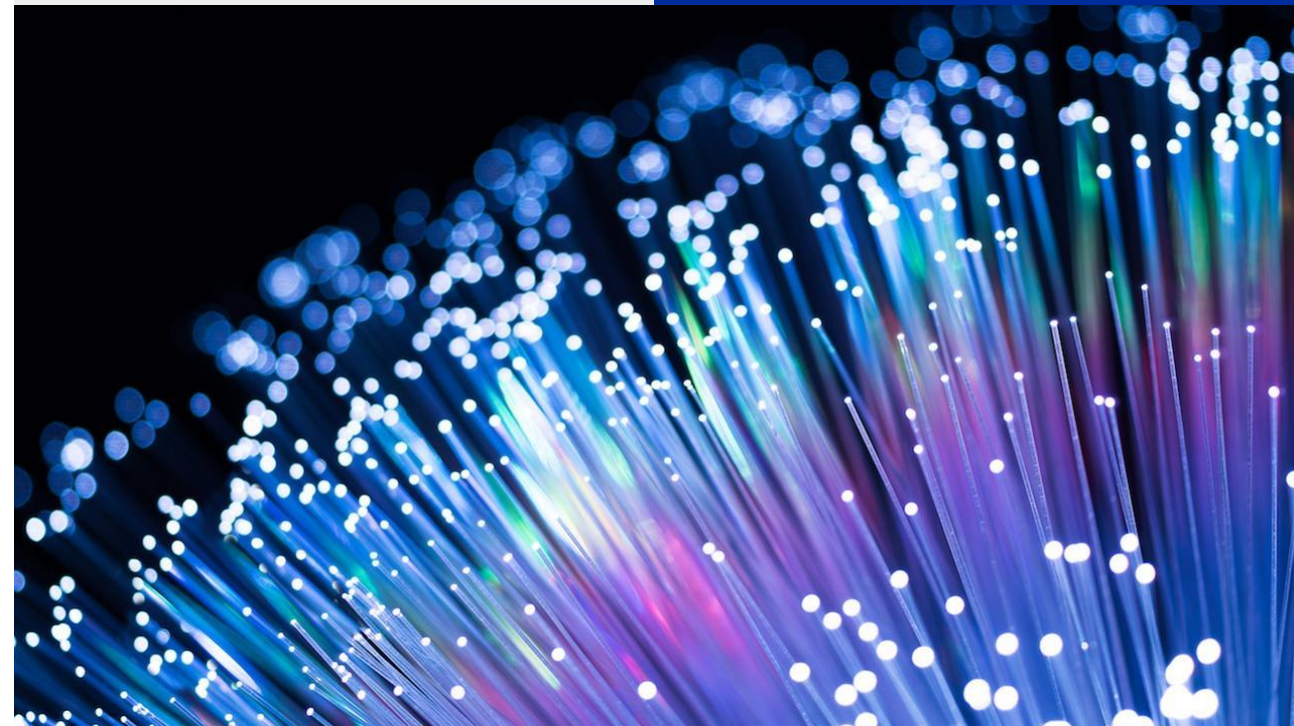
Industry leading memory bandwidth > 400 GB/s

In memory databases and enterprise application workloads place intense demands on systems hardware infrastructure, specifically taxing the memory bandwidth architecture and design.

IBM's Power E1080 server delivers **1.8x more peak theoretical main memory bandwidth per socket** than POWER9 and more than the leading competitor supporting extreme data center efficiency and sustainability.



Power10's new Open Memory Interface (OMI) offers industry leading memory bandwidth per socket to handle even the largest in-memory database workloads



Security: *Pandemic Evolution into a “Cyber Pandemic”*

SECURITY

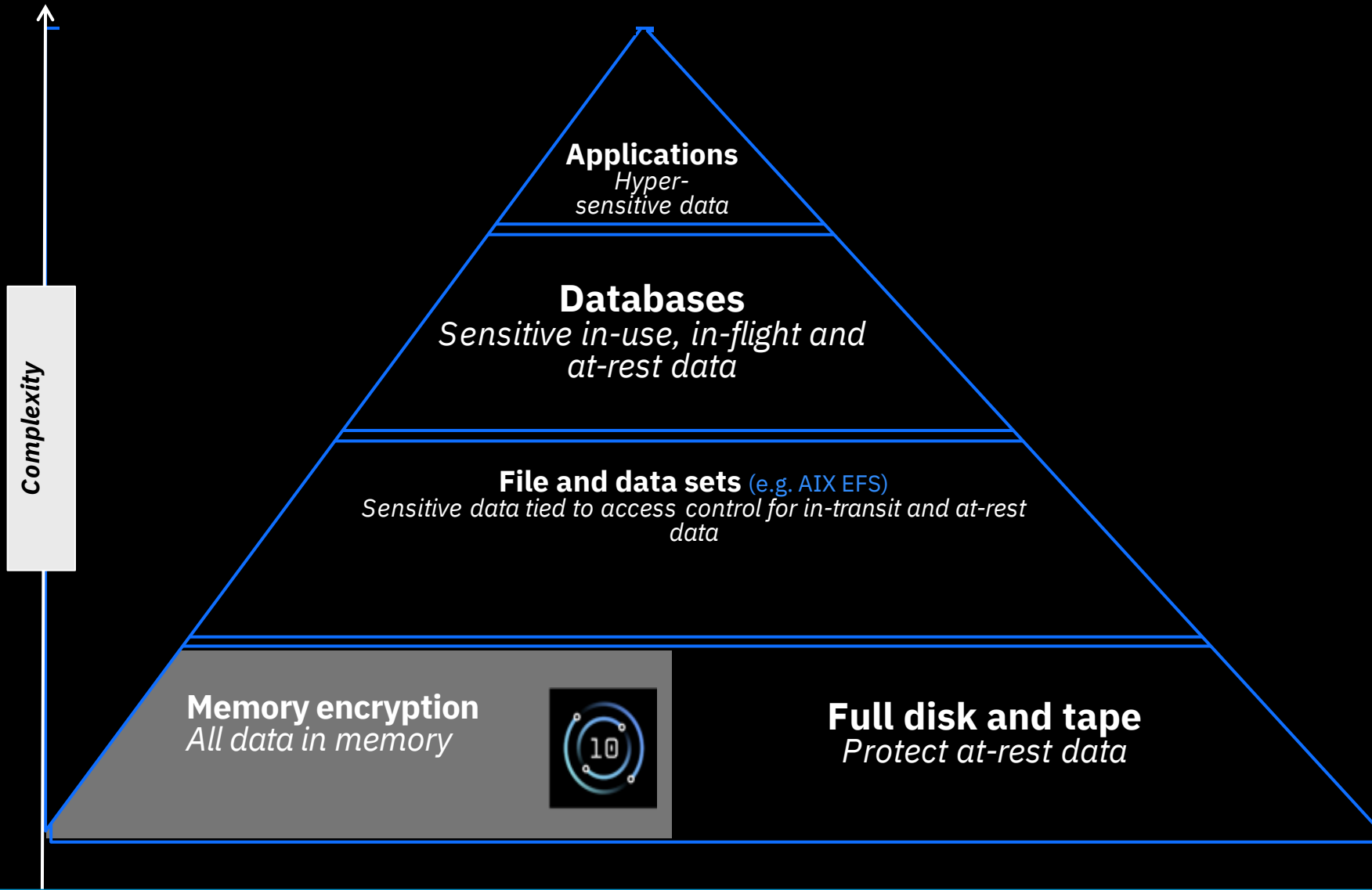
Florida water system cyberattack – the dangers of remote access

The attack plays into the rising use of remote access options among critical organizations amid the pandemic.

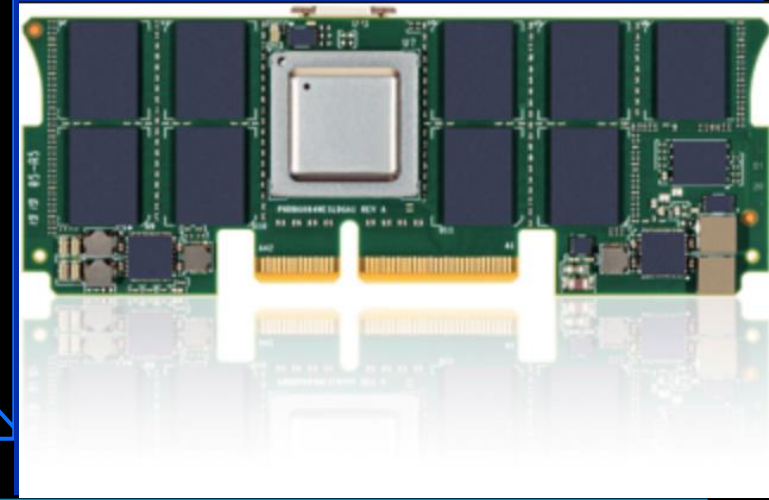
9 February 2021 | [9](#) Shares

- **15,000 Florida residents were at risk of consuming poisoned water after a cyberattacker gained control of systems at a water treatment plant**
- **The attacker began increasing the amount of sodium hydroxide in the water by a factor of 100**
- **Compromising remote access software, the incident highlights the current vulnerabilities of ‘connected’ critical infrastructure and operational technology**

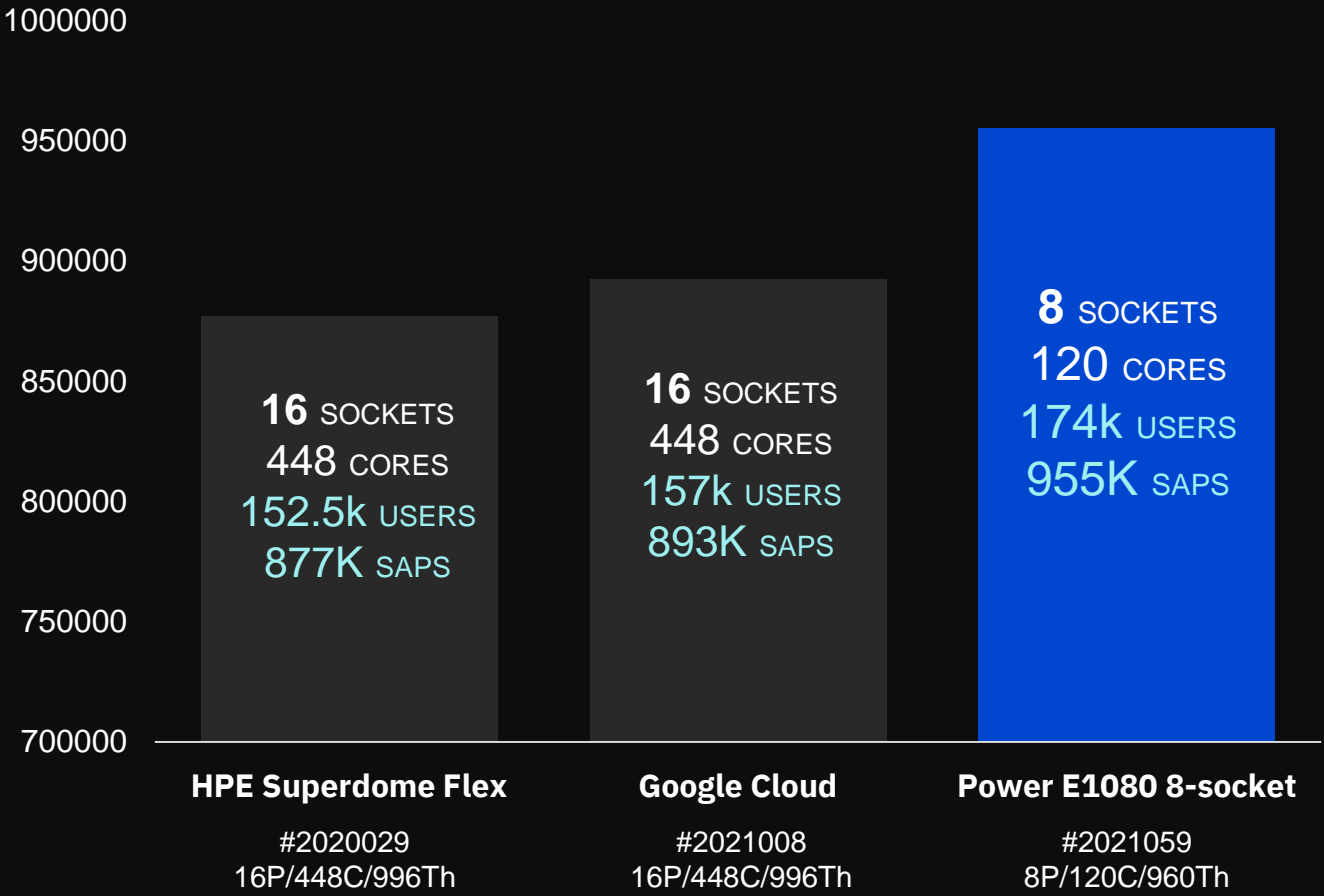
Power10 - Protect Data with Simplified Encryption!



- Quantum-safe cryptography
- Fully homomorphic encryption



IBM Power E1080 sets world record 8-socket two-tier SAP SD standard application benchmark result¹



- World record 8-socket performance
 - 955,050 vs. 670,830 SAPS
 - 174,000 vs. 122,300 users
 - More performance per core
 - 4x vs. 16-socket Intel²
 - 2.7x vs. 8-socket Intel³
- The most flexible and reliable SAP HANA platform⁴
- Power E1080 servers scale to 16 sockets

100 SAPS = 2,000 fully business processed order line items per hour

1. IBM Power E1080; two-tier SAP SD standard application benchmark running SAP ERP 6.0 EHP5; Power10 3.55-4.0 GHz processor, 4,096 GB memory, 8p/120c/960t, 174,000 SD benchmark users (955,050 SAPS), AIX 7.2, DB2 11.5. Certification # 2021059. All results can be found at sap.com/benchmark Valid as of 8/27/21

2. Google Cloud Platform; two-tier SAP SD standard application benchmark running SAP ERP 6.0 EHP5 (cloud); Intel Xeon Platinum 8280L 2.7 GHz, 16p/448c/896t, 157,000 SD benchmark users (892,270 SAPS), running Windows Server 2019 and Microsoft SQL Server 2017, Certification # 2021008.

3. HPE Superdome Flex; two-tier SAP SD standard application benchmark running SAP ERP 6.0 EHP5; Intel Xeon Platinum 8380H 2.9 GHz, 8p/224c/448t, 122,300 SD benchmark users (670,830 SAPS), Windows Server 2016 and Microsoft SQL Server 2012, Certification # 2021006.

4. Ranked most reliable server in its category for 12th year by ITIC. Flexible: Only platform that runs AIX, IBM i, Linux OS'es while supporting the ability to run 16 SAP HANA production environment in a single server

Respond faster to
business demands

Protect data from
core to cloud

Streamline insights
and automation

Maximize availability

Red Hat OpenShift & IBM Cloud Paks on Power

Efficient
Scaling

4.1X

more containerized throughput
per core than x86 running Red
Hat OpenShift*

Persistent
Security and
Reliability

Most secure workload isolation
Advanced data protection
Platform integrity

Optimize
Utilization

Automated core allocation across worker nodes
Gain performance and TCO advantages co-locating AIX,
IBM i and Red Hat OpenShift environments
Instant scaling, pay per use consumption

*Based on IBM internal testing of Red Hat OpenShift Container Platform 4.8.2 worker nodes running 80 pods each with 10 users using the Daytrader7 workload (<https://github.com/WASdev/sample.daytrader7/releases/tag/v1.4>) accessing AIX Db2 databases. Average cpu utilization for the OCP worker nodes is > 95%. Comparison: Power E1080 running OCP accessing AIX Db2 on an S922 versus OCP on Cascade Lake accessing AIX Db2 on the same S922. Valid as of 8/26/2021 and conducted under laboratory conditions. Individual result can vary based on workload size, use of storage subsystems & other conditions. IBM Power E1080 (40 cores/3.8 GHz/2 TB memory) in maximum performance mode, 25 Gb two-port SRIOV adapter, 1 x 16Gbps FCA, with PowerVM. Competitive system: Intel(R) Xeon(R) Gold 6248 CPU (Cascade Lake) in performance mode, 40 cores/3.9GHz/512GB memory), 25Gb two-port SRIOV adapter, 1 x 16Gbps FCA, RHEL 8.4 KVM.

Спасибо за внимание!

Вопросы?