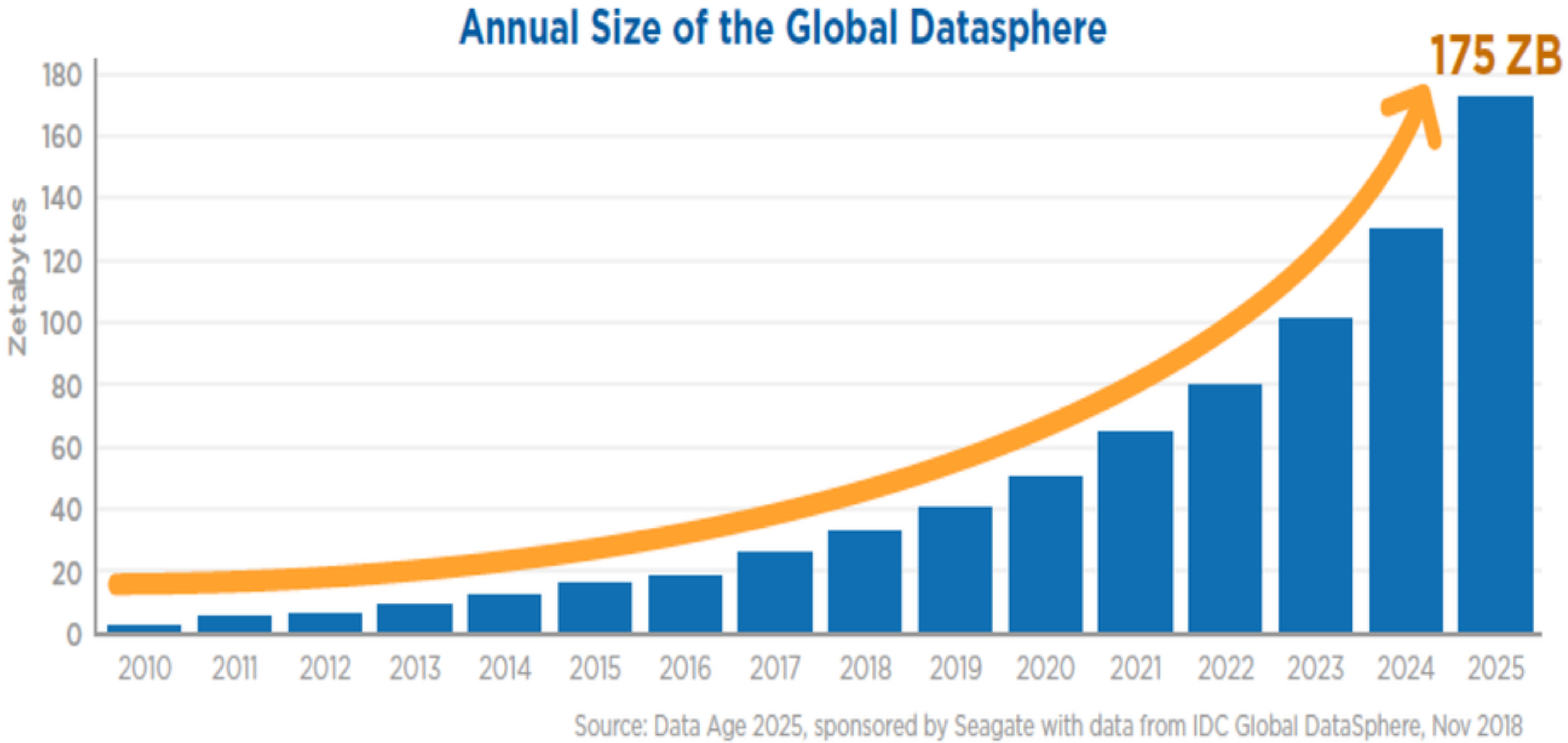
STAC Live 2021

Accelerating Analytics with Intel® Optane™ Technology

Steve Scargall - Persistent Memory Architect, Intel®

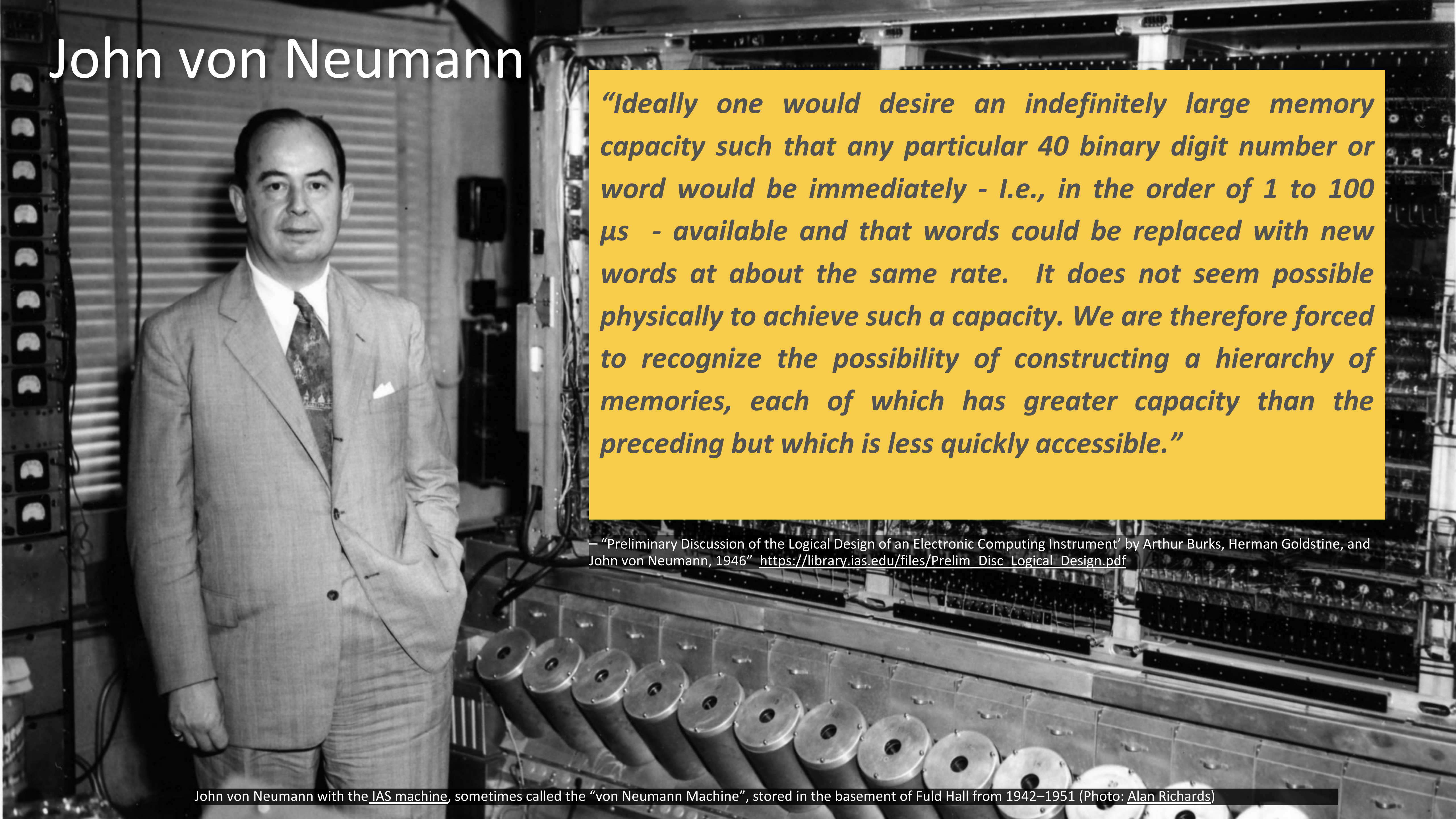
intel®

Data Growth



Annual Size of the Global Datasphere
175 ZB
Zetabytes
180
160
140
120
100
80
60
40
20
0
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
Source: Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018

John von Neumann



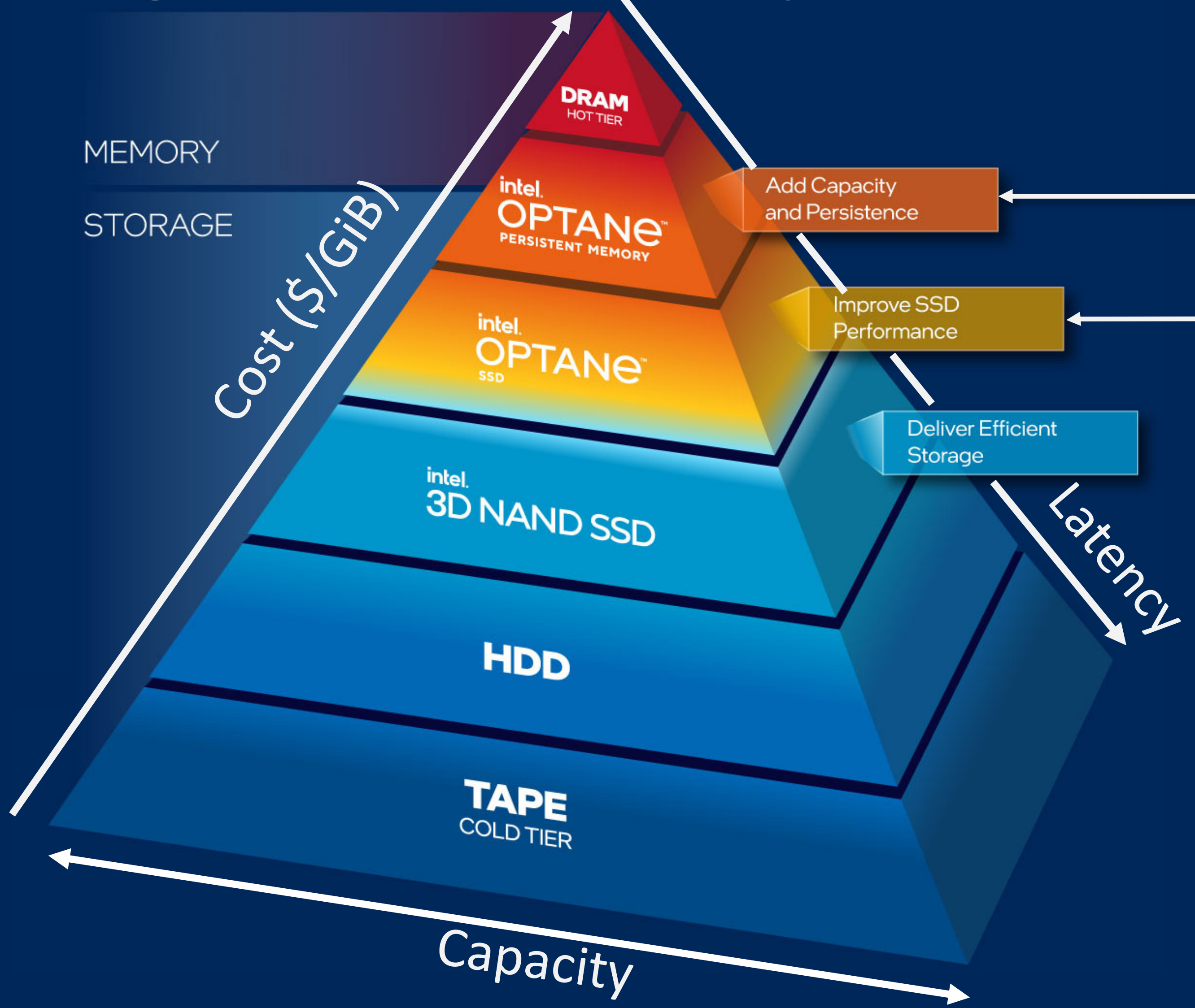
“Ideally one would desire an indefinitely large memory capacity such that any particular 40 binary digit number or word would be immediately - I.e., in the order of 1 to 100 µs - available and that words could be replaced with new words at about the same rate. It does not seem possible physically to achieve such a capacity. We are therefore forced to recognize the possibility of constructing a hierarchy of memories, each of which has greater capacity than the preceding but which is less quickly accessible.”

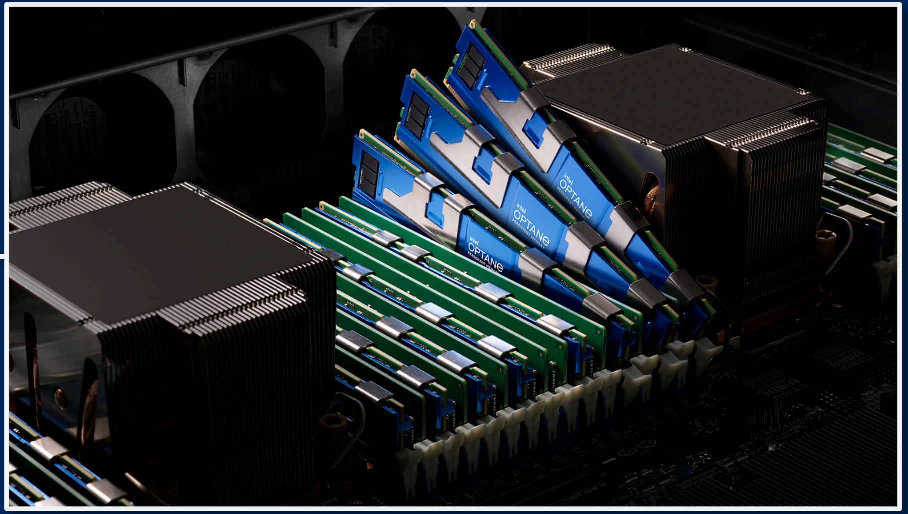
– “Preliminary Discussion of the Logical Design of an Electronic Computing Instrument’ by Arthur Burks, Herman Goldstine, and John von Neumann, 1946” https://library.ias.edu/files/Prelim_Disc_Logical_Design.pdf

John von Neumann with the IAS machine, sometimes called the “von Neumann Machine”, stored in the basement of Fuld Hall from 1942–1951 (Photo: Alan Richards)

Closing the Memory and Storage Divide

Data greater than Memory





Intel® Optane Persistent Memory Technology

AES 256-BIT encryption | Secure Erase | Up To **512 GB** modules

UP TO 32% higher average memory bandwidth over the previous generation+ | **AES 256-BIT** encryption | Secure Erase | Up To **512 GB** modules

Intel® Optane™ PMem 100 series

2nd Generation Intel® Xeon® Scalable processors on 2S/4S/8S platform

- **8-28** cores
- **6 channels** memory
- **3 TB** Intel Optane PMem per socket*
- **4.5 TB** Total system memory per socket*
- **2,666 MT/s** DDR4 + Intel Optane PMem
- **18 W Max** thermal design power

Intel Optane PMem 200 series

3rd Generation Intel Xeon Scalable processors on 4S platform

- **18-28** cores
- **6 channels** memory
- **3 TB** Intel Optane PMem per socket*
- **4.5 TB** Total system memory per socket*
- **2,666 MT/s** DDR4 + Intel Optane PMem
- **eADR**
- **15 W Max** thermal design power

Intel Optane PMem 200 series

3rd Generation Intel Xeon Scalable processors on 2S platform

- **16-40** cores
- **8 channels** memory
- **4 TB** Intel Optane PMem per socket**
- **6 TB** Total system memory per socket**
- **3,200 MT/s** DDR4 + Intel Optane PMem
- **eADR**
- **15 W Max** thermal design power

+ Based on testing by Intel as of April 27,2020 (Baseline) and March 31, 2020 (New).
* 3 TB Intel Optane PMem = 6 x 512 GB Intel Optane PMem per socket, 4.5 TB System Memory = 6 x 512 GB Intel Optane PMem per socket + 6 x 256 GB
** 4 TB Intel Optane PMem = 8 x 512 GB Intel Optane PMem per socket, 6 TB System Memory = 8 x 512 GB Intel Optane PMem per socket + 8 x 256 GB

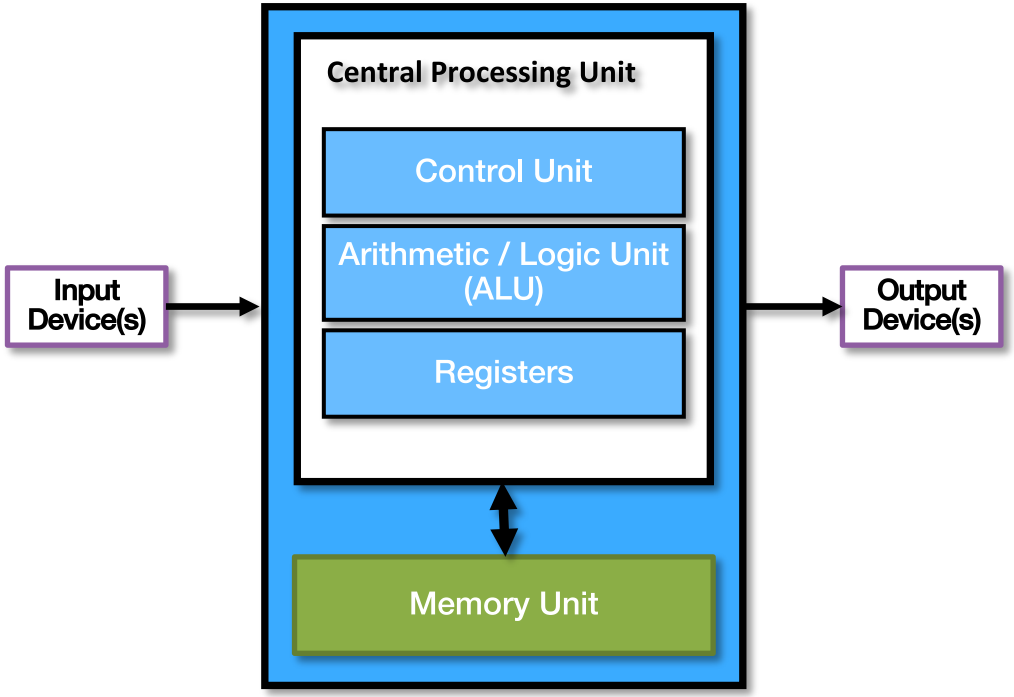
KX KDB+ Nano Benchmark

- **Analytics**
  - **Performed within 10% of DRAM** for queries involving table joins
  - **Performed 4× to 12× faster** than 24 NVMe storage in RAID configuration
  - DRAM performed 3× to 10× faster when performing single-threaded calculations and aggregations on data
- **Data processing and I/O operations**
  - **Processed 1.6× more data per second than NVMe-only** storage where data was read from PMem and written to NVMe storage
  - **2× to 10× faster** reading data from files in parallel
  - Seed of reading data similar to page cache (DRAM)
  - Single-threaded file-write performance within 10% in both configurations
  - Multithreaded file-write performance 42% slower
- **Infrastructure resources**
  - **Required 37% less RAM** to complete key I/O-intensive data processing
  - **Required no page cache** for querying or retrieving data stored in PMem
- **Business benefits**
  - PMem is cheaper than DRAM per GiB, and higher density
  - Collect and process more data with higher velocity sensors and assets
  - Accelerate analytics and queries on recent data by **4× to 12×**
  - Reduce cost of infrastructure running with less servers and DRAM to support data processing and analytic workloads
  - Align infrastructure more closely to the value of data by establishing a storage tier between DRAM and NVMe- or SSD-backed performance block storage

Source: https://code.kx.com/q/architecture/optane-tests/#summary-results

Backup

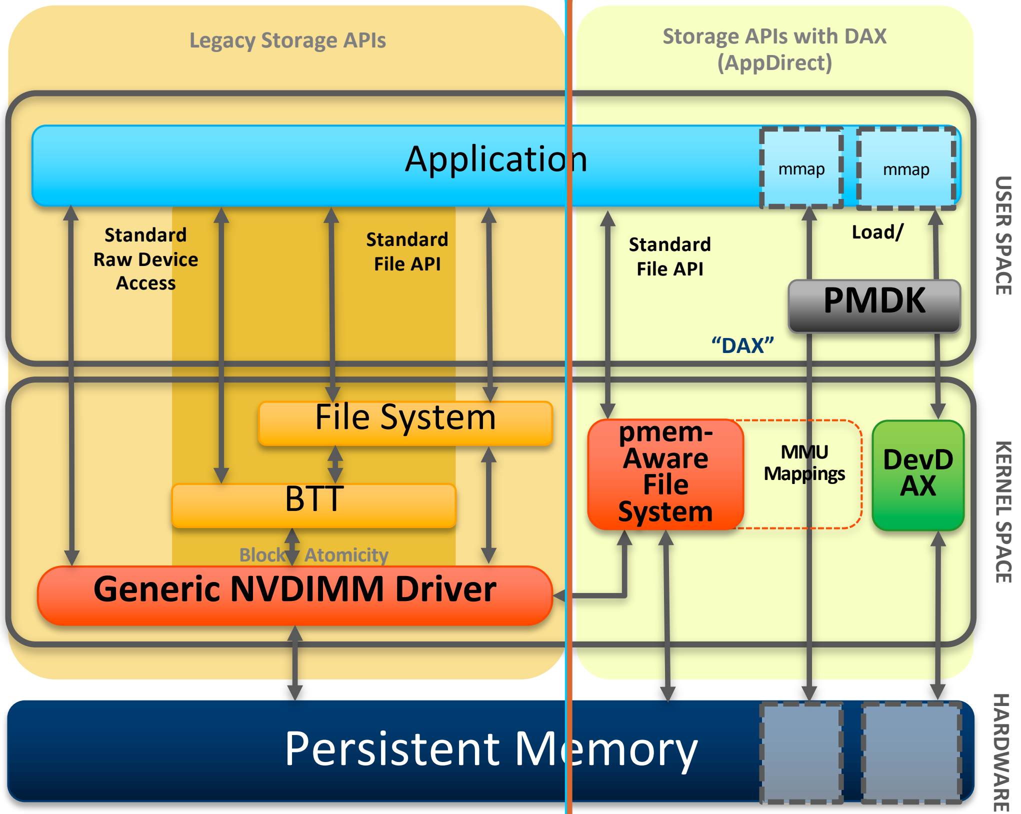
Von Neumann Computer Architecture



SNIA Programming Model

- No Code Changes Required
- Operates in Blocks like SSD/HDD
  - Traditional read/write
  - Works with Existing File Systems
  - Atomicity at block level
  - Block size configurable
    - 4K, 512B*
- NVDIMM Driver required
  - Support starting Kernel 4.2
- Configured as Boot Device
- High Performance Block Storage
  - Low Latency, higher BW, High IOPs

*Requires Linux



- Code changes may be required*
- Bypasses file system page cache
- Requires DAX enabled file system
  - XFS, EXT4, NTFS
- No Kernel Code or interrupts
- No interrupts
- Fastest IO path possible

* Code changes required for load/store direct access if the application does not already support this.

Current Problems

- DRAM Too Costly
- Scale Up Too Expensive
- Not Enough Capacity
- Operational Inefficiencies
- Poor Workload Performance
- Storage Too Slow

Use Intel® Optane™ Persistent Memory for...

Cost Savings

DRAM

Servers greater than 512 GB

Improve TCO

Workloads that need large or persistent memory

Productivity

Increase Memory Size

Reduce software license fees per core

Consolidate Workloads

Many VMs, with low CPU utilization

Performance

Break I/O Bottlenecks

High disk I/O traffic

Add High-Speed Storage

Byte-addressable storage tier

Database	Analytics	Virtualized Infrastructure	Networking
AEROSPIKE	KATANA GRAPH	vmware	**Content Delivery Network**
Accelerated Performance Through Expanded Capacity	Faster Analytics Insights	Lower Costs Per VM	Increase Live Video Resolution
Up To **2.5X** More Transactions Keep Data and Index in PMem	Up To **2X** Faster Graph Analytics Computations	Up To **25%** Lower Costs Per VM While Delivering the Same Performance	Up To **63%** Higher Throughput of Content Delivery Network