Using DPUs
for Database Management
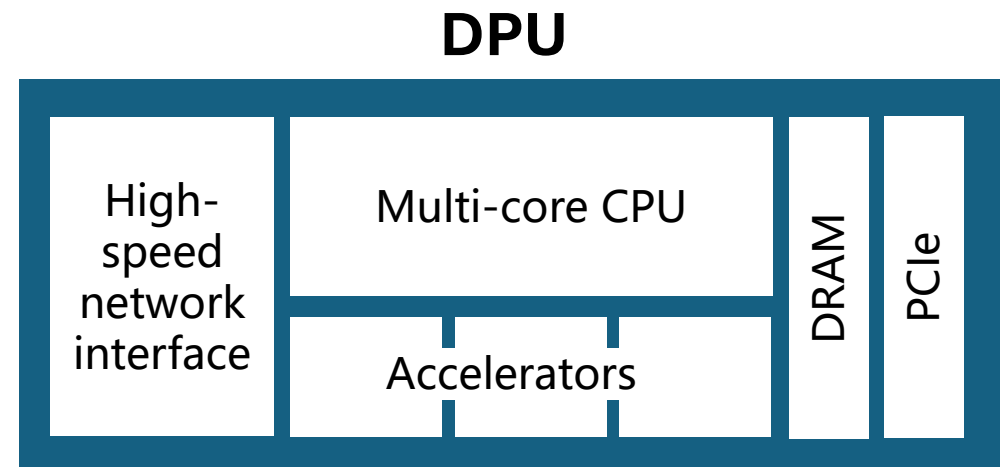
Phil Bernstein
Microsoft Research

HPTS Gong Show

September 16, 2024

What's a DPU

- A computer

- A SmartNIC implemented by a system-on-a-chip

DPU Anatomy

- Multi-core CPUs - Arm or MIPS
  - Cheap compute cycles
  - Low power consumption

- ASICs for compression, encryption, regex, dedup, crypto hash, …

- High-speed network connection

- Onboard memory, e.g., 8 GB - 32 GB

- PCIe support to access SSDs, etc.

DPU

High-speed network interface	Multi-core CPU	DRAM	PCIe
	Accelerators		

Everyone is making them

- **Achronix Speedster7t**
- **AMD Pensando**
- **Asterfusion: Helium**
- **AWS: Nitro**
- **Broadcom: Stingray**
- **Intel: IPU**
- **Kalray: K200**
- **Marvell: OCTEON**
- **Microsoft: Catapult, Azure Boost, Fungible**
- **Netronome: Agilio**
- **Nvidia: BlueField**
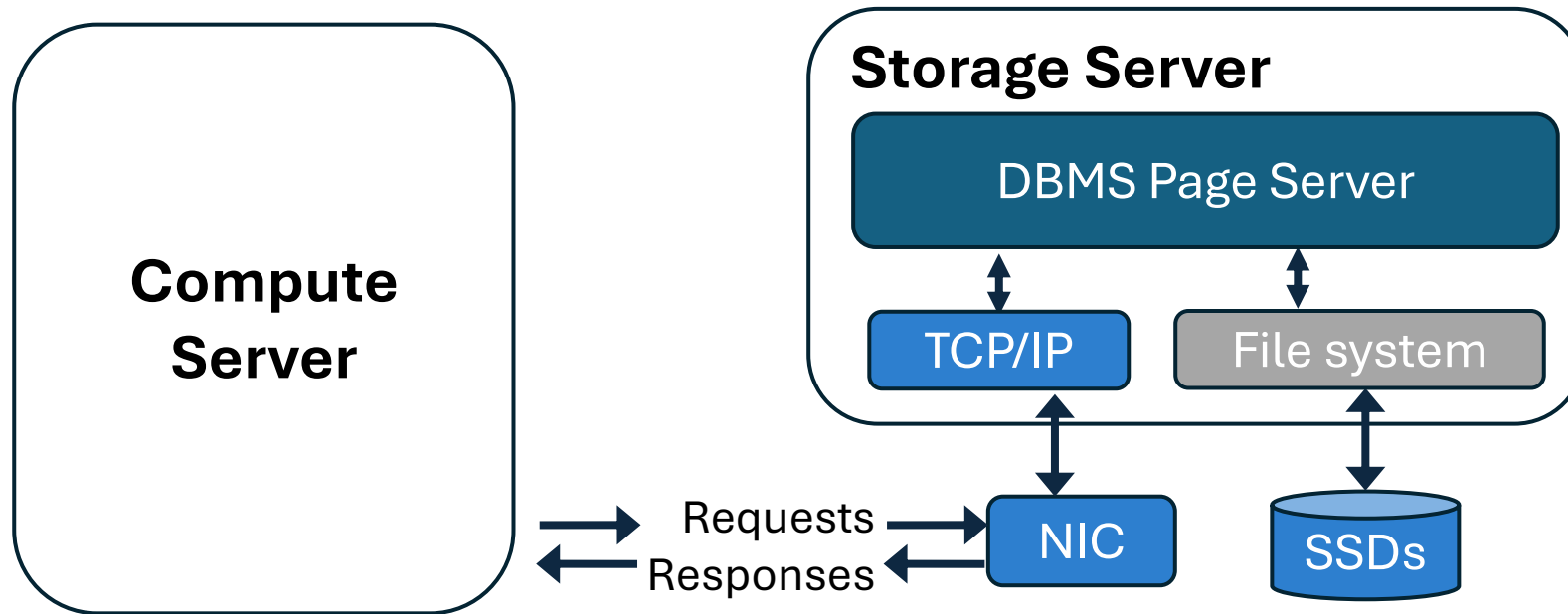- **Silicom**

They're everywhere

Applications

- Network protocols

- Security barriers

- Distributed computing

What about Database Folks?

Phil Bernstein

Cloud DBMS Architecture

Compute Server

Storage Server

DBMS Page Server

TCP/IP   File system
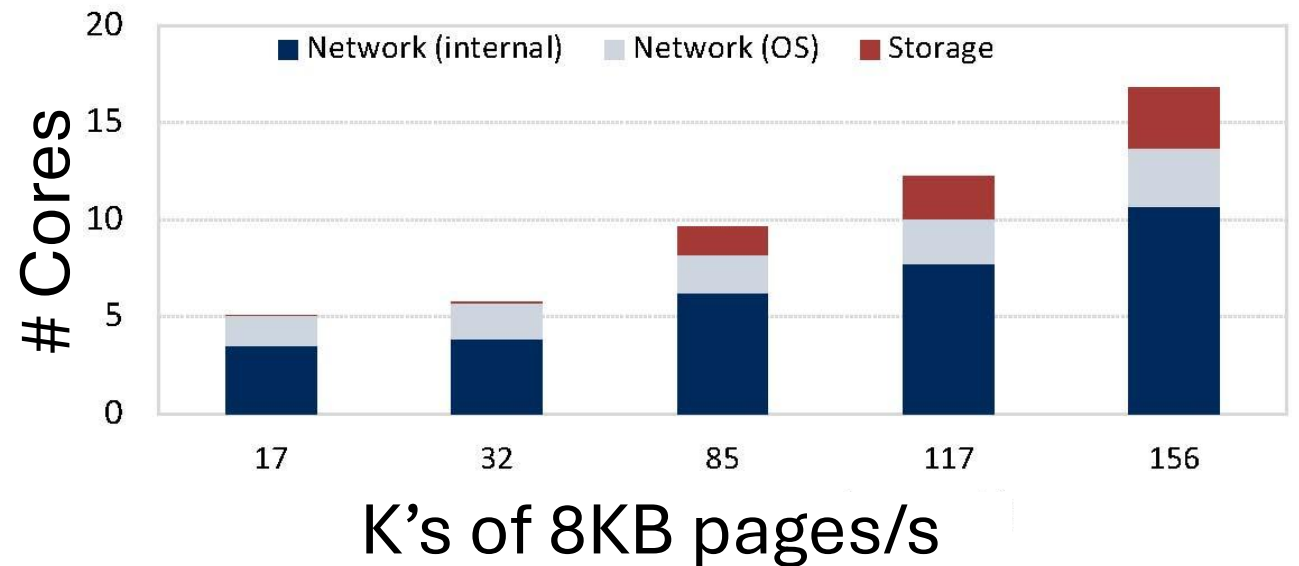
Requests →   NIC
Responses ←

SSDs

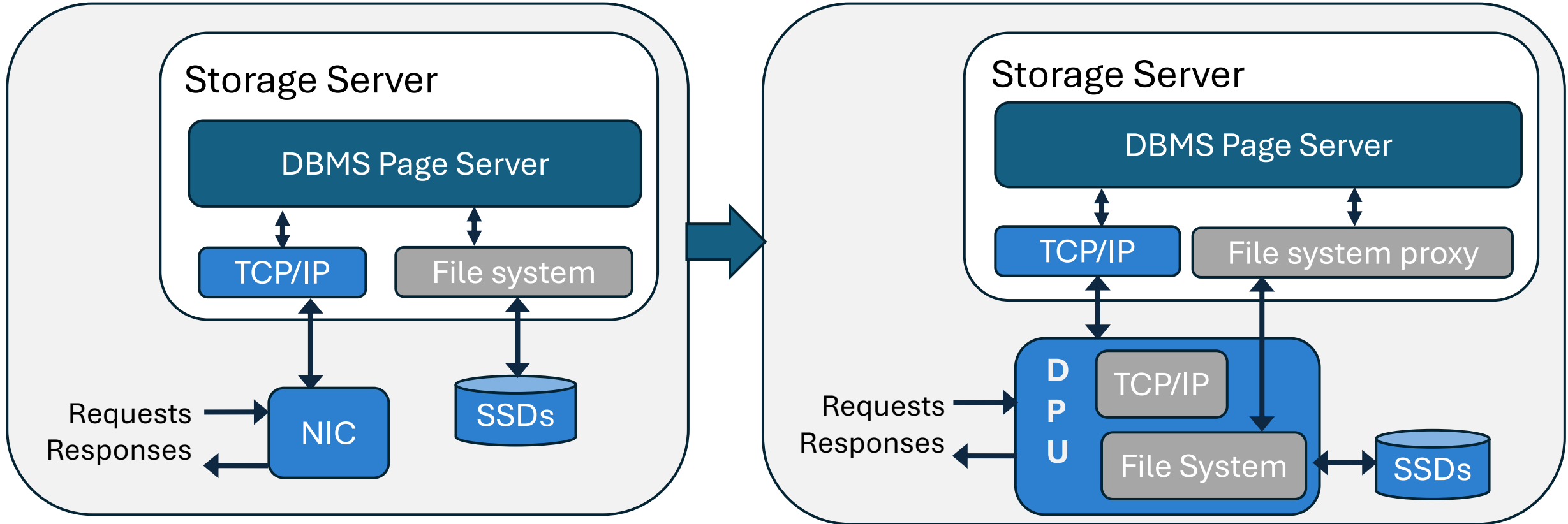Storage server's network and storage stacks are costly and inefficient

Storage Server I/O is Costly

- Uses 17 cores to get 1.25 GB/s (156K pages/s)

- TCP/IP is the biggest slice

- SSD accesses are costly too

Legend: ■ Network (internal) ■ Network (OS) ■ Storage

Y-axis: # Cores (0, 5, 10, 15, 20)

X-axis: 17, 32, 85, 117, 156 — K's of 8KB pages/s

Solution – Offload Storage Server I/O to DPU

17 Intel cores on Storage Server → 3 Arm cores on its SmartNIC

Further information

Qizhen Zhang, Philip A. Bernstein, Badrish Chandramouli,

Jiasheng Hu, Yiming Zheng:

DDS: DPU-optimized Disaggregated Storage, *VLDB 2024*

[Extended report] https://arxiv.org/abs/2407.13618