

Cognitive SSD: A Deep Learning Engine for In-Storage Data Retrieval

Shengwen Liang^{1,2}, Ying Wang^{1,2}, Youyou Lu³, Zhe Yang³
Huawei Li^{1,2}, Xiaowei Li^{1,2}

¹State Key Laboratory of Computer Architecture,
Institute of Computing Technology, Chinese Academy of Sciences, Beijing,

²University of Chinese Academy of Sciences,

³Tsinghua University



中国科学院大学

University of Chinese Academy of Sciences



清華大學

Tsinghua University

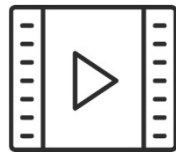
Unstructured Data



Text file and document



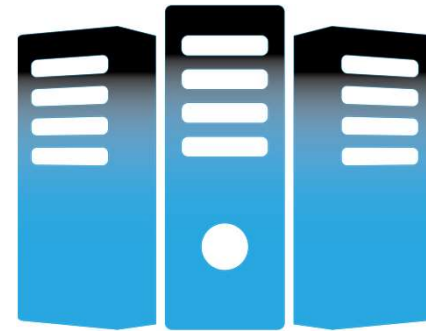
Audio



Video

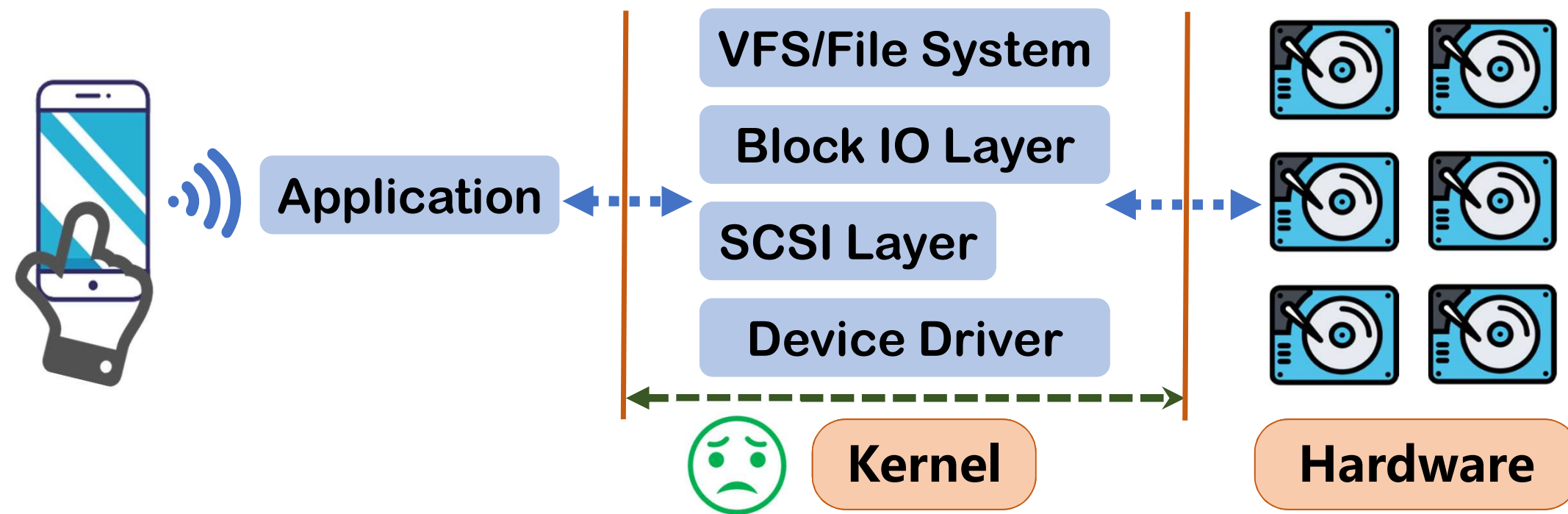


Image



80%





• Performance bottleneck migrates from hardware (SSD(75-50us) to software (IO Stack (60.8us))

**NAND FLASH
SSD**

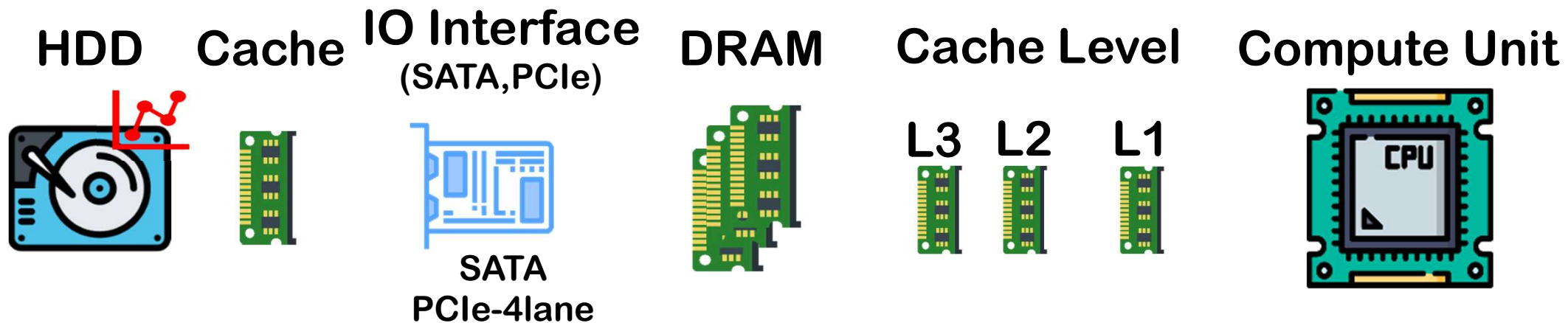


75-50us



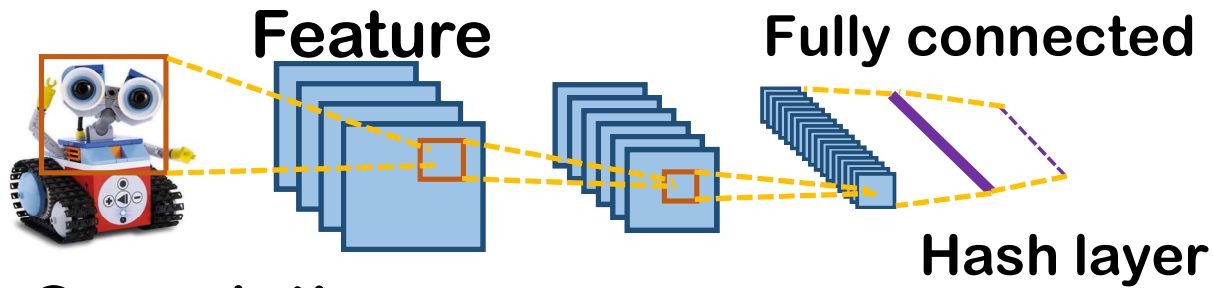
10us

**OPTANE
SSD**

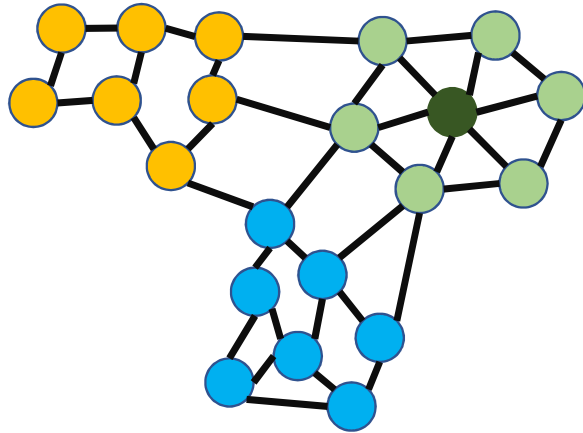


- Massive data movement incurs energy and latency overhead in the conventional memory hierarchy.

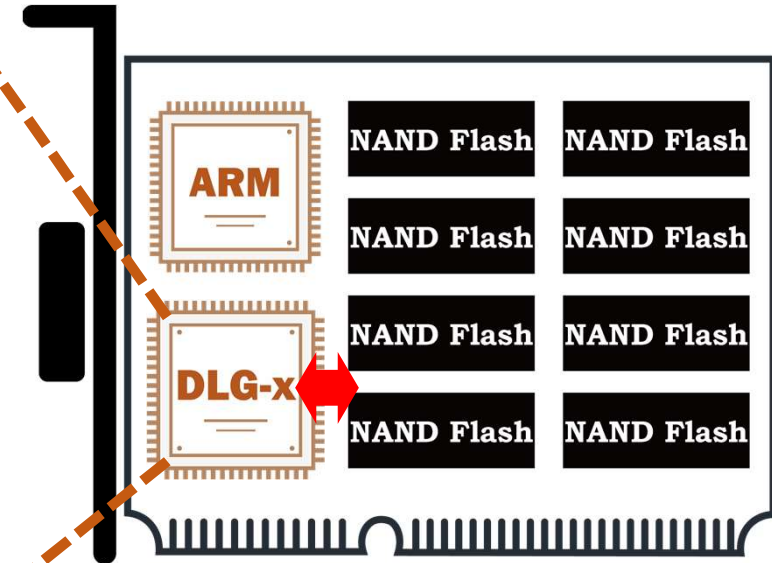
Deep learning for better feature representation



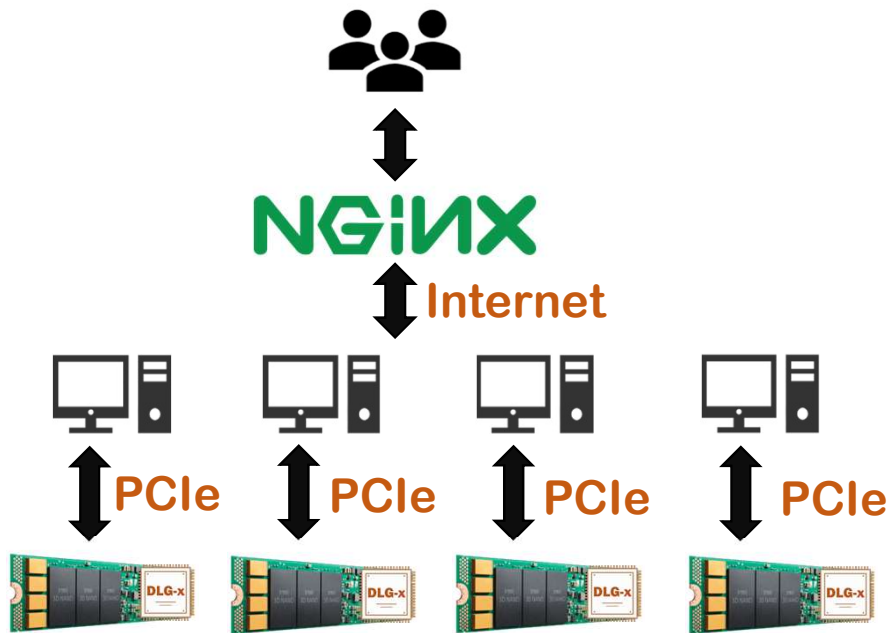
Convolution Pooling



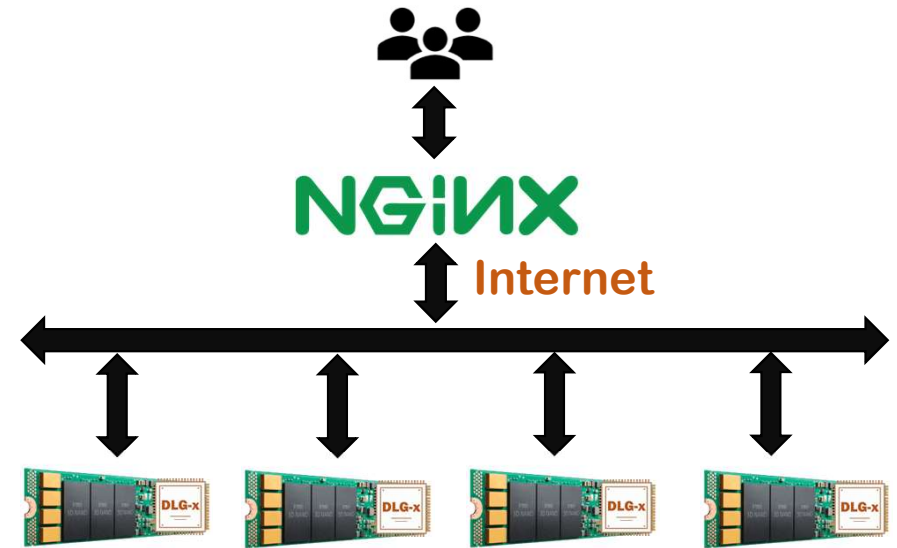
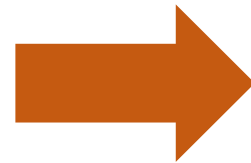
Graph search for database indexing



Cognitive SSD



Conventional cluster system



Host free cluster system

Cognitive SSD as a server

Cognitive SSD: A Deep Learning Engine for In-Storage Data Retrieval

Hope to see you at our presentation

Track I: Programmable I/O Devices

July 11, 2019, Thursday, 11:15am-12:35pm

Thank you



中国科学院大学
University of Chinese Academy of Sciences



清华大学
Tsinghua University