Flash Memory Summit

In-Storage Computing SSD specifications and applications

Feng Zhu, Longxiao Li, Hua Sun, Xiaohui Yan, Kun Wang,
Kai Tao, Ming Zeng, Wenqiang Wang, Wei Li, Zhushi Cheng,
Zhenjun Liu, Sheng Qiu, Xulin Yu, Shu Li

Alibaba Group

08/08/2019

Alibaba Group 阿里巴巴集团

AliFlash product family

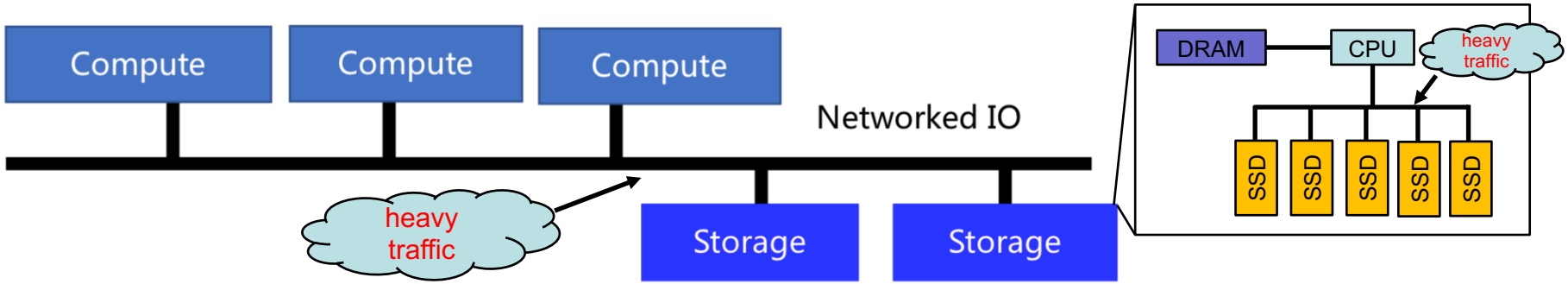
- Host-Based PCIe SSD
- Deployed since 2016
- > 50k pcs

- NVMe SSD
- Deployed since 2017

- Open Channel SSD
- Volume ramping up
- Targeting DB/RDS/ Search/EBS etc.

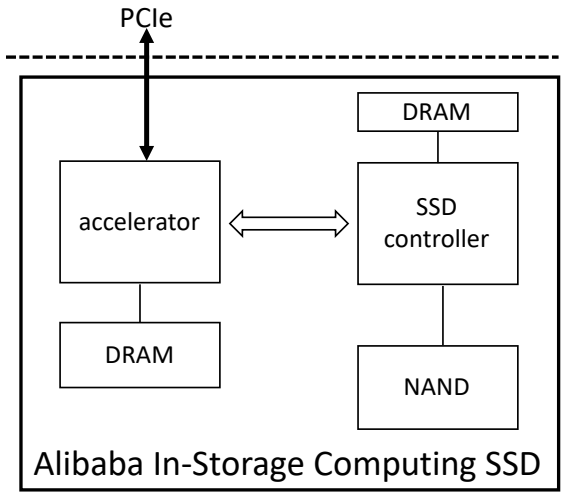
- In-storage computation SSD
- FTL offloading
- Dev in progress
- Targeting DB/RDS etc.

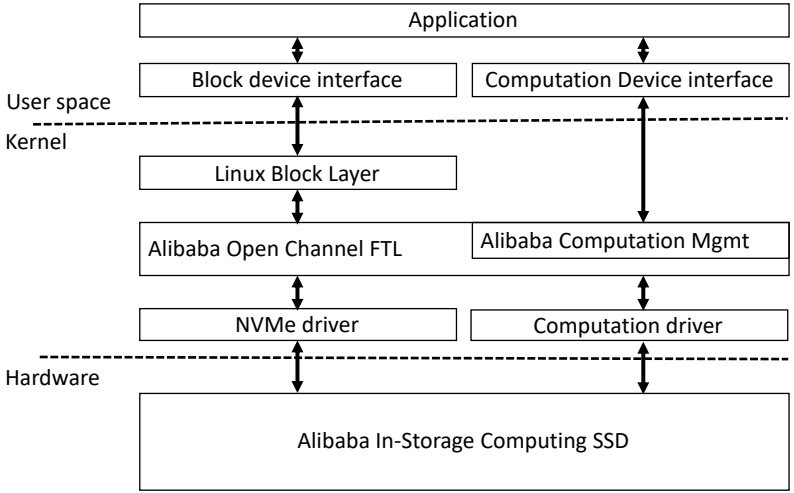
Why In-Storage Computing



- Fast network enabled the Disaggregated Storage and Compute Architecture for cost efficiency and scalability
- However the intense domain specific compute not only stresses the network and PCIe bus due to large data movement, but also causes the IO BW pressure on CPU and DRAM
- In-storage computing enables the data analytics and computing in SSD to mitigate above issues

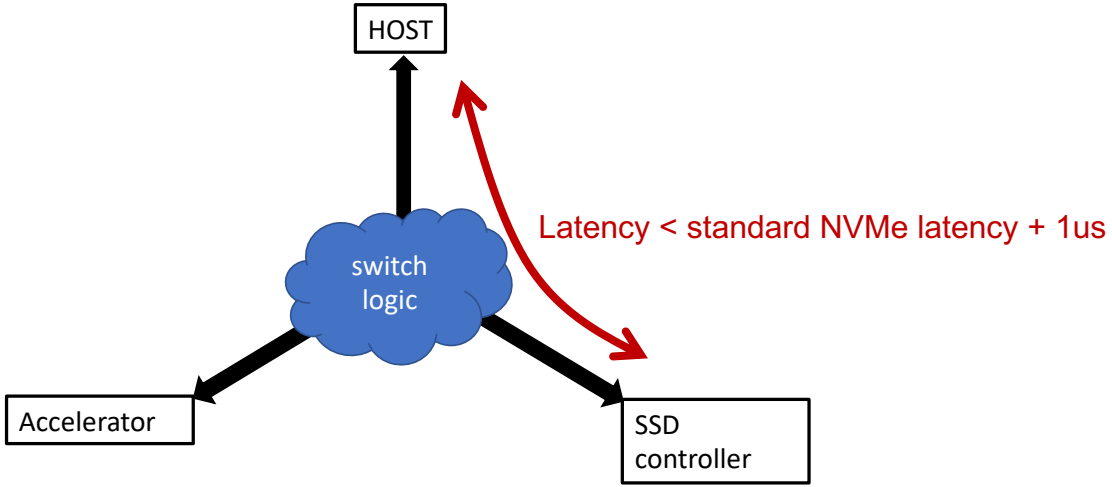
HW/SW Architecture





- Hybrid architecture with accelerator for computing and SSD controller for storage
- Separate storage and computation devices exposed to host (with separate drivers)
- Inherited from Alibaba open channel SSD

Flexible data paths



- 3-way switches provide the flexible data paths among host, accelerator and SSD controller
- Implementation shall be carefully optimized to ensure the Host-SSD latency adder < 1us compared to that of normal NVMe SSD

Alibaba In-Storage Computing Ecosystem

SAMSUNG

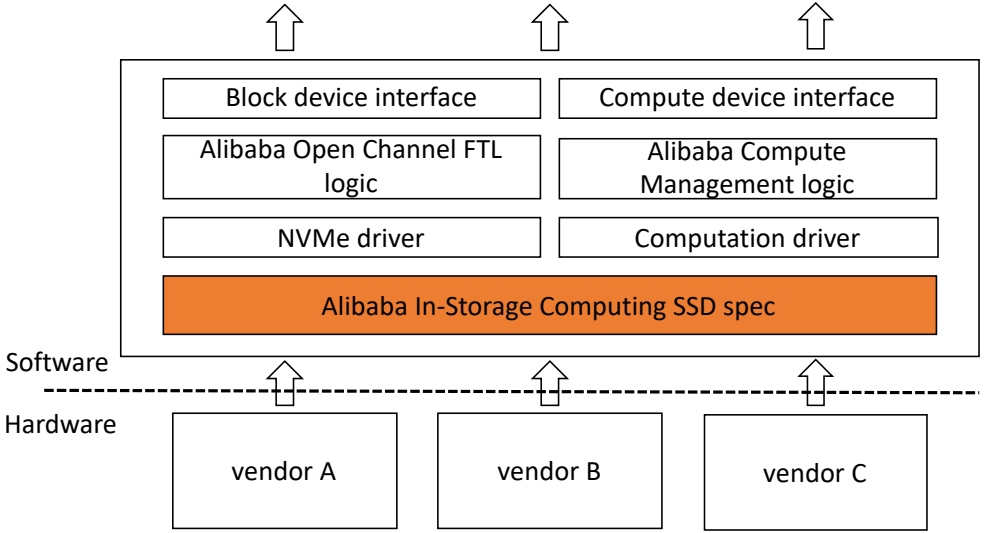
ScaleFlux™

Shannon Systems
宝 存 科 技

XILINX®

- Collaborating with major vendors in in-storage computing
- More vendors are joining

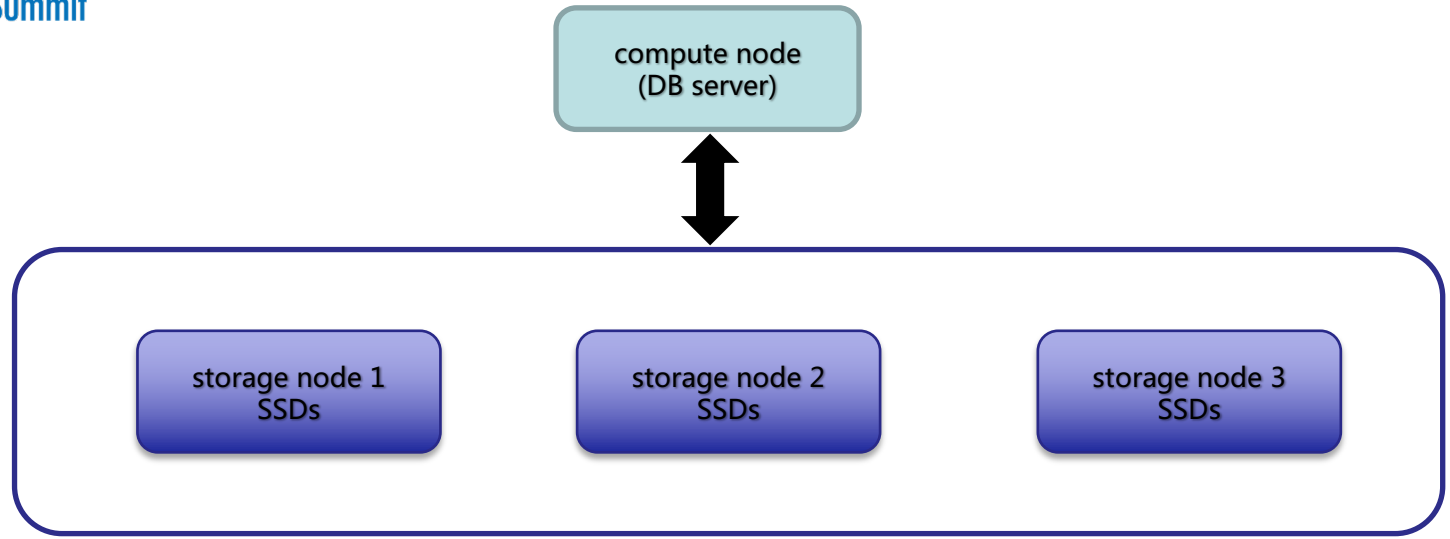
Alibaba In-Storage Computing specifications



- Alibaba owns the software/driver, the acceleration logic and FPGA shell
- Vendors provide the hardware platform

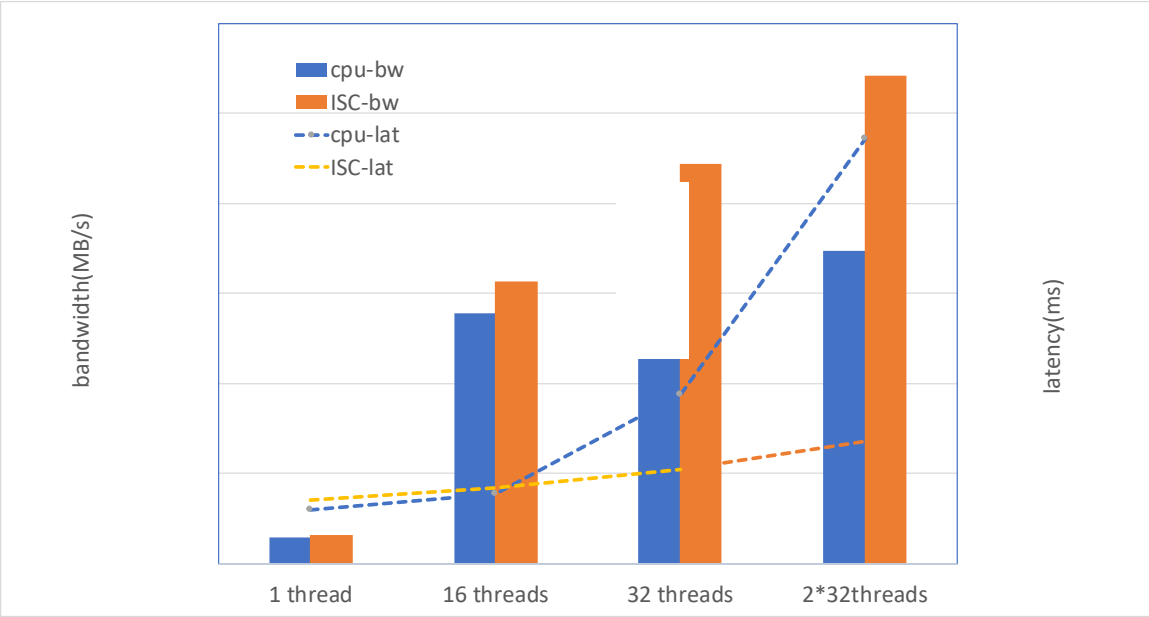
- Multiple vendors' hardware complied to Alibaba In-storage Computing spec, which defined the HW/SW architecture, performance/power requirement etc.
- Released to vendors with MOU and NDA signed

Experiment setup



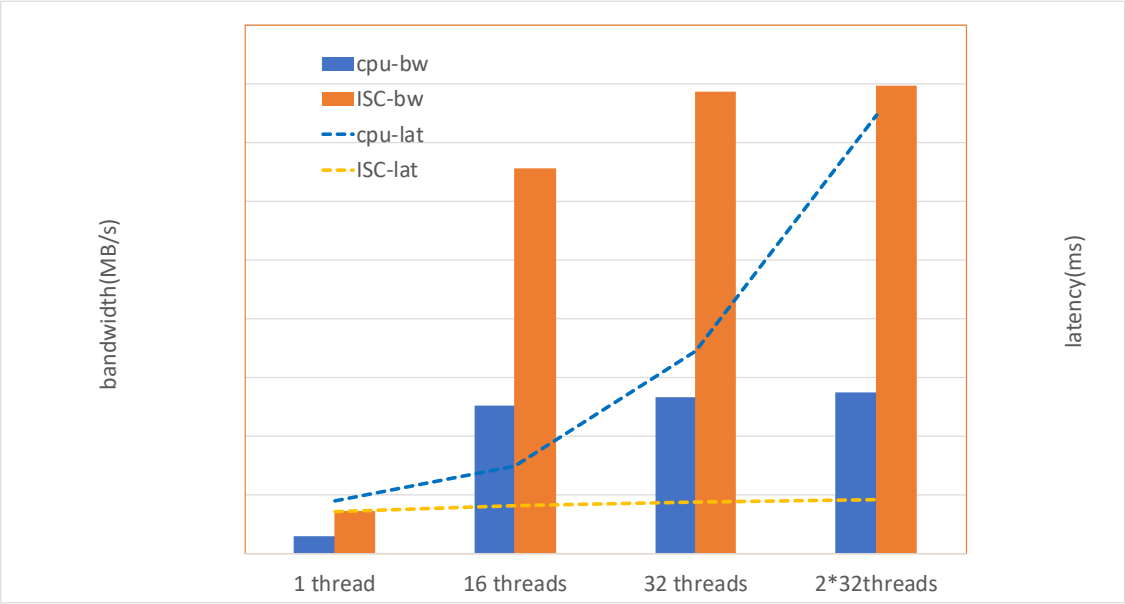
- Demonstrated the benefits of decompression and filtering for Database applications
  - Baseline: 1 compute node + 3 storage node. 12 standard NVMe SSDs per storage node. 2 core for data fetching and 9 cores for computation
  - In-storage computing solution: 1 compute node + 3 storage node, 12 Alibaba In-Storage Computing SSDs per storage node, 2 threads for query dispatch

Filtering only

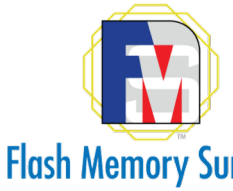


- Using CPU for computation, 9 cores are fully utilized and become the bottleneck at higher pressure
- Using In-Storage Computing SSD, 1.56X BW improvement and 71% latency reduction

Decompression + Filtering



- With decompression + filtering, even higher gain are observed: 2.88X BW improvement and 88% latency reduction.



Conclusion Remarks

- AliFlash In-Storage Computing architecture
- Multi-vendor ecosystem with the unified Alibaba In-Storage Computing spec
- Significant gain demonstrated with database applications
- Alibaba is open to industry collaboration

THANK YOU