JDT 京东科技

SPDK Engineering Practices of Software and Hardware Integrated Virtualization for JD.com Cloud

Qiming zhang

JDT 京东科技

Agenda

01 Background

02 Integrated Virtualization platform

03 Engineering optimization practices

04 Summary

JDT 京东科技

01

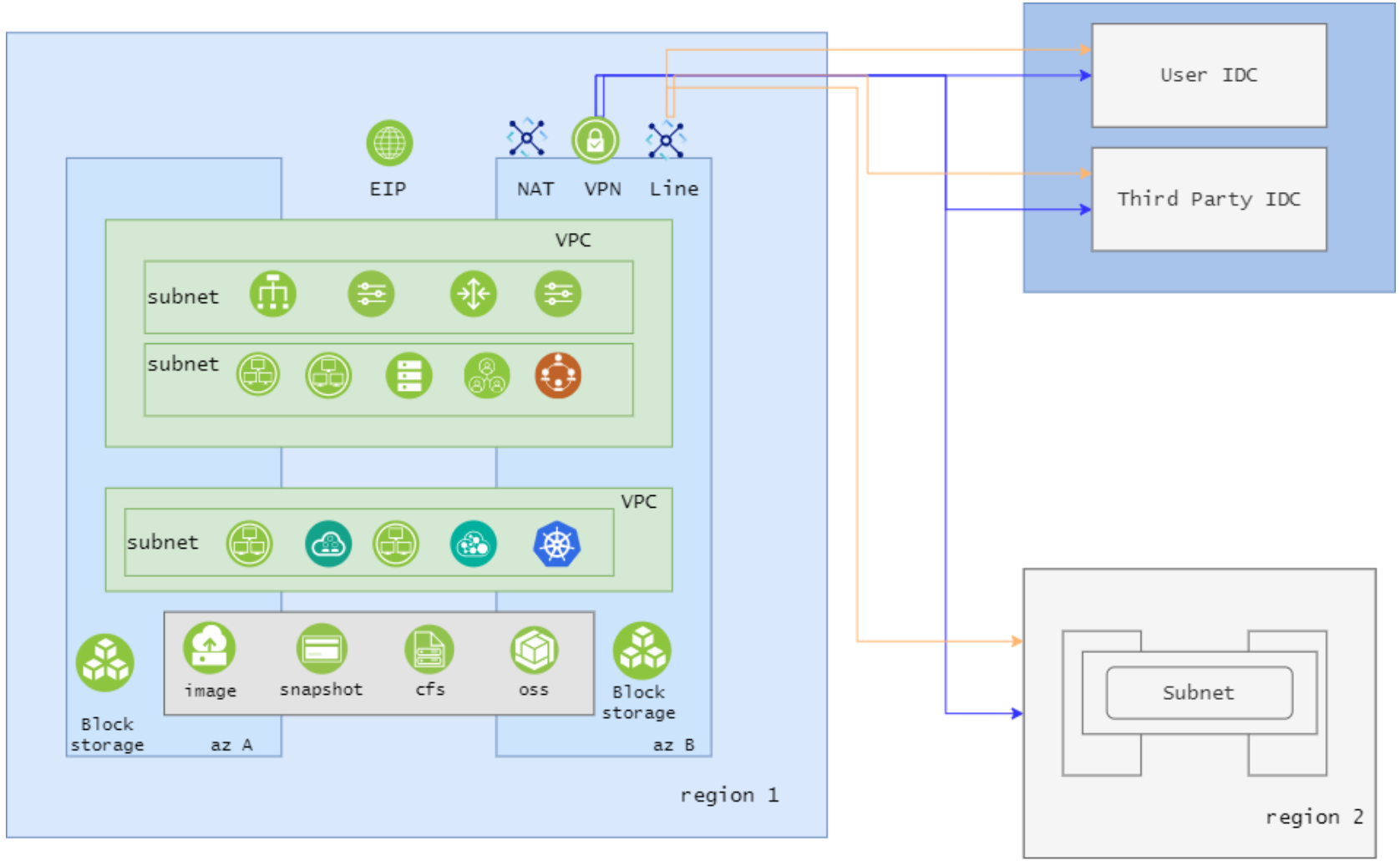
Background

The challenge of public cloud storage

JDT 京东科技

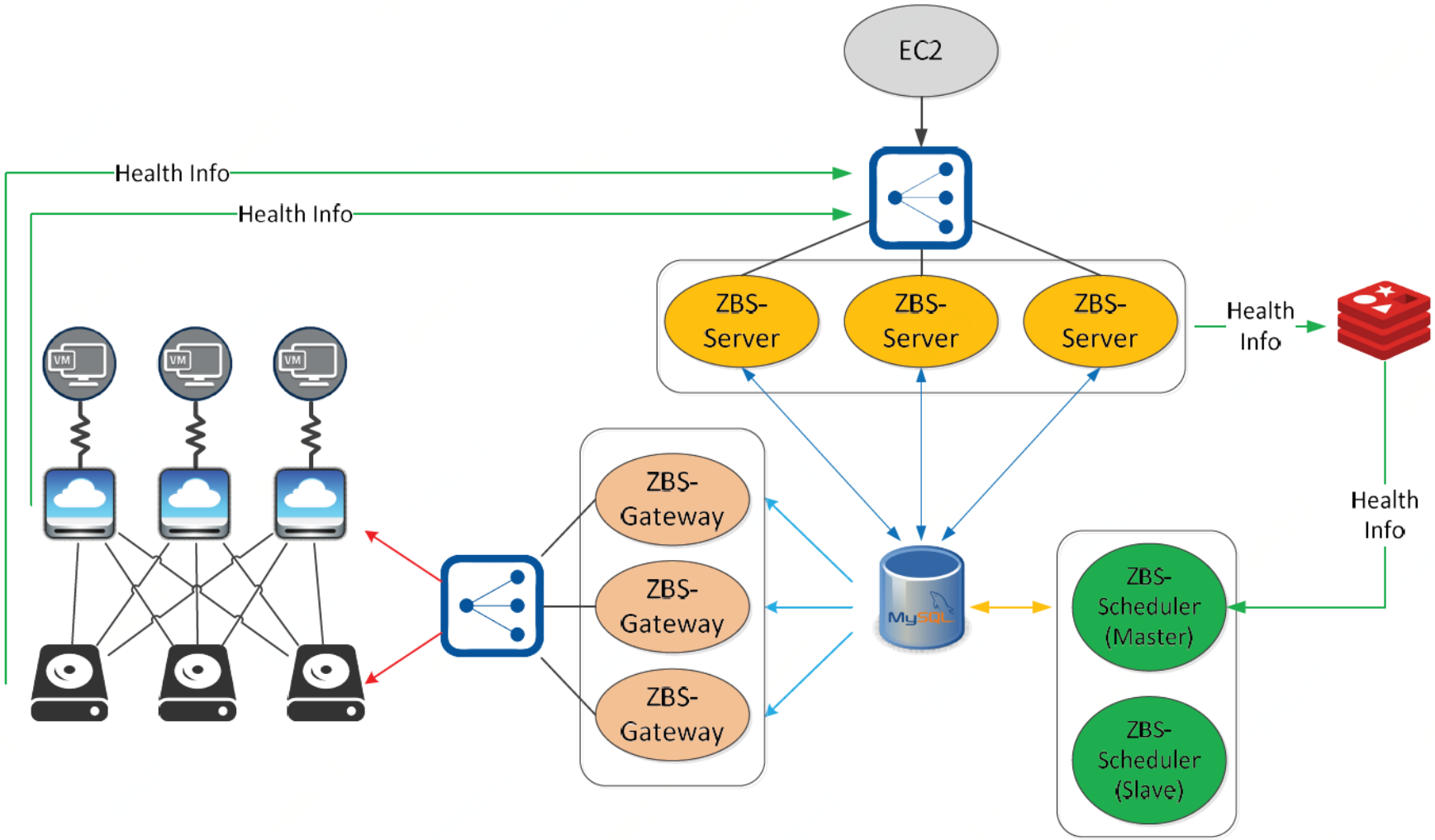
Faced with the growing storage needs of internal and external users on the cloud, storage is in urgent need of

- Higher performance
- Lower costs
- More stable service



Storage Architecture

JDT 京东科技

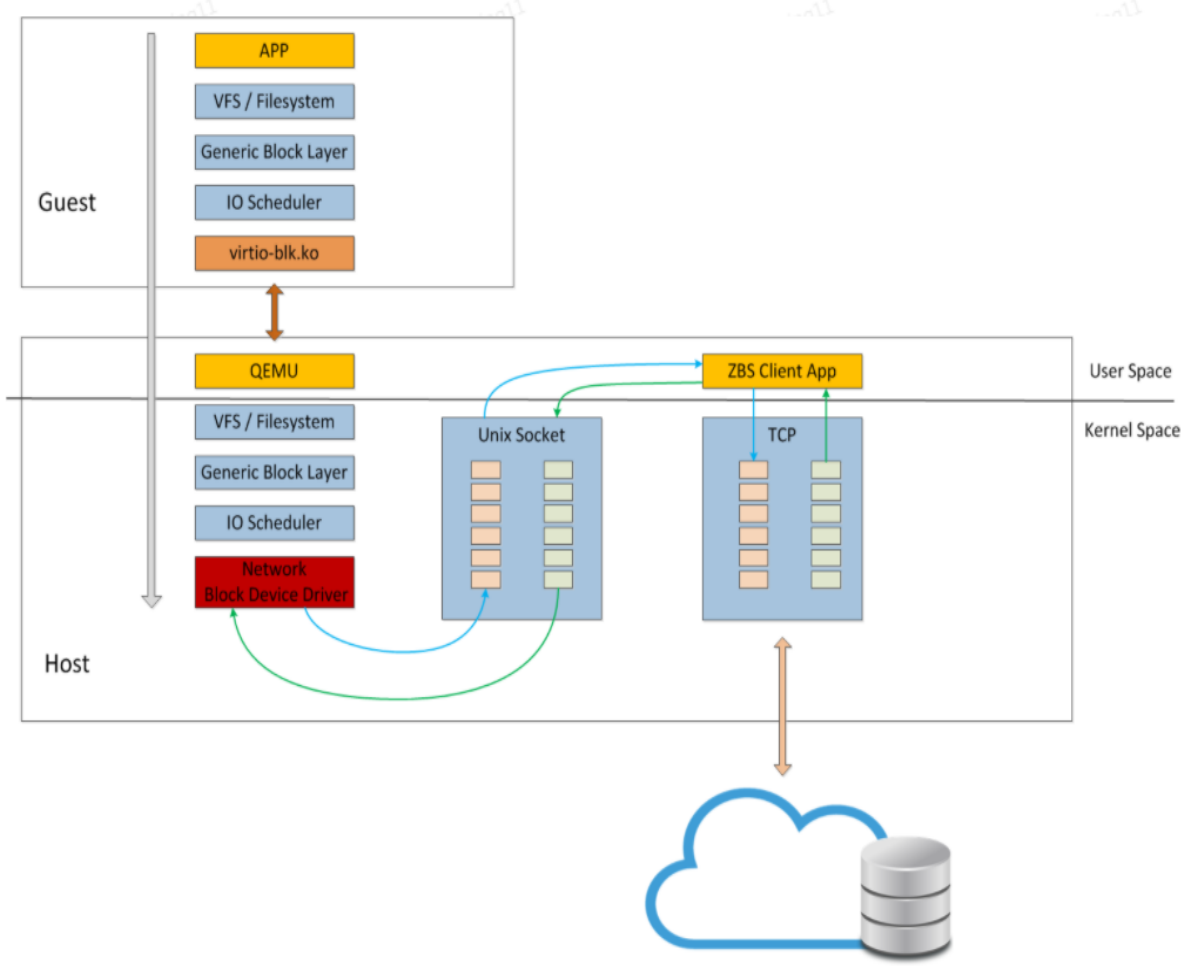


Architectural defects

JDT 京东科技

qemu && nbd

- Longer io path
- Expensive virtualization overhead
- Service stability impact



JDT 京东科技

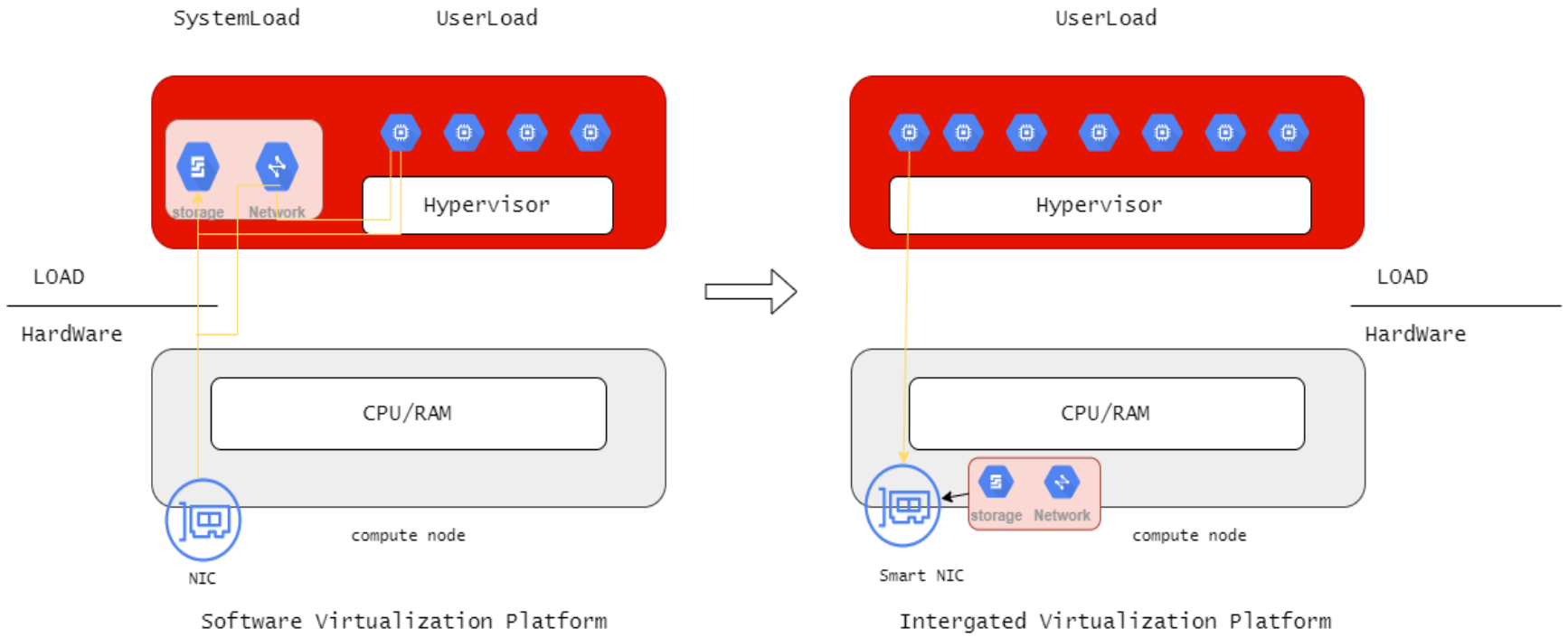
02

Integrated Virtualization platform

Integrated Virtualization Platform

JDT 京东科技

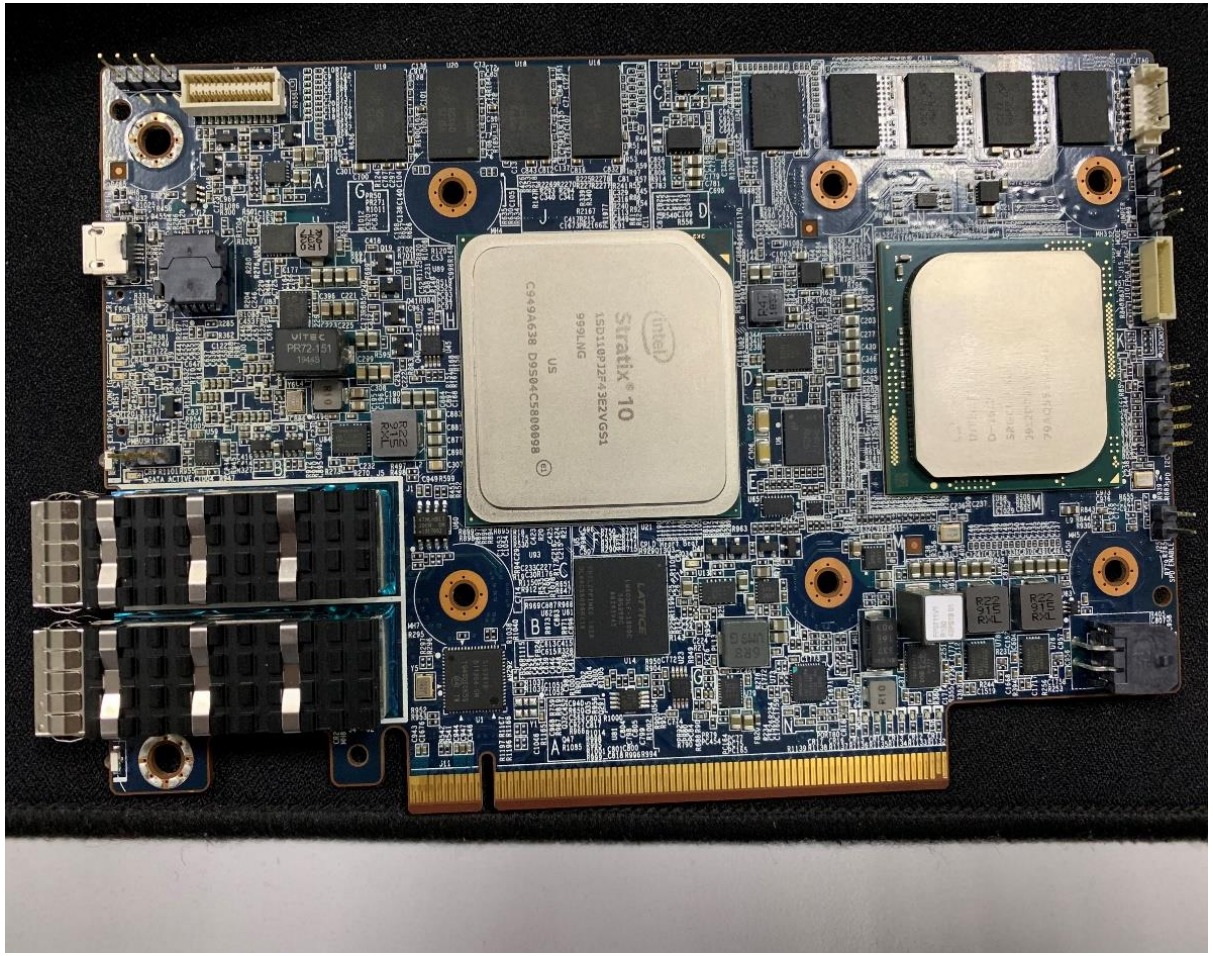
- Reduce virtualization overhead
- Reduce interference and improve stability
- Unified virtualization platform that supports heterogeneous
- Multi-vendor support



Smart NIC Choice

JDT 京东科技

Our choice is working on Intel BSC

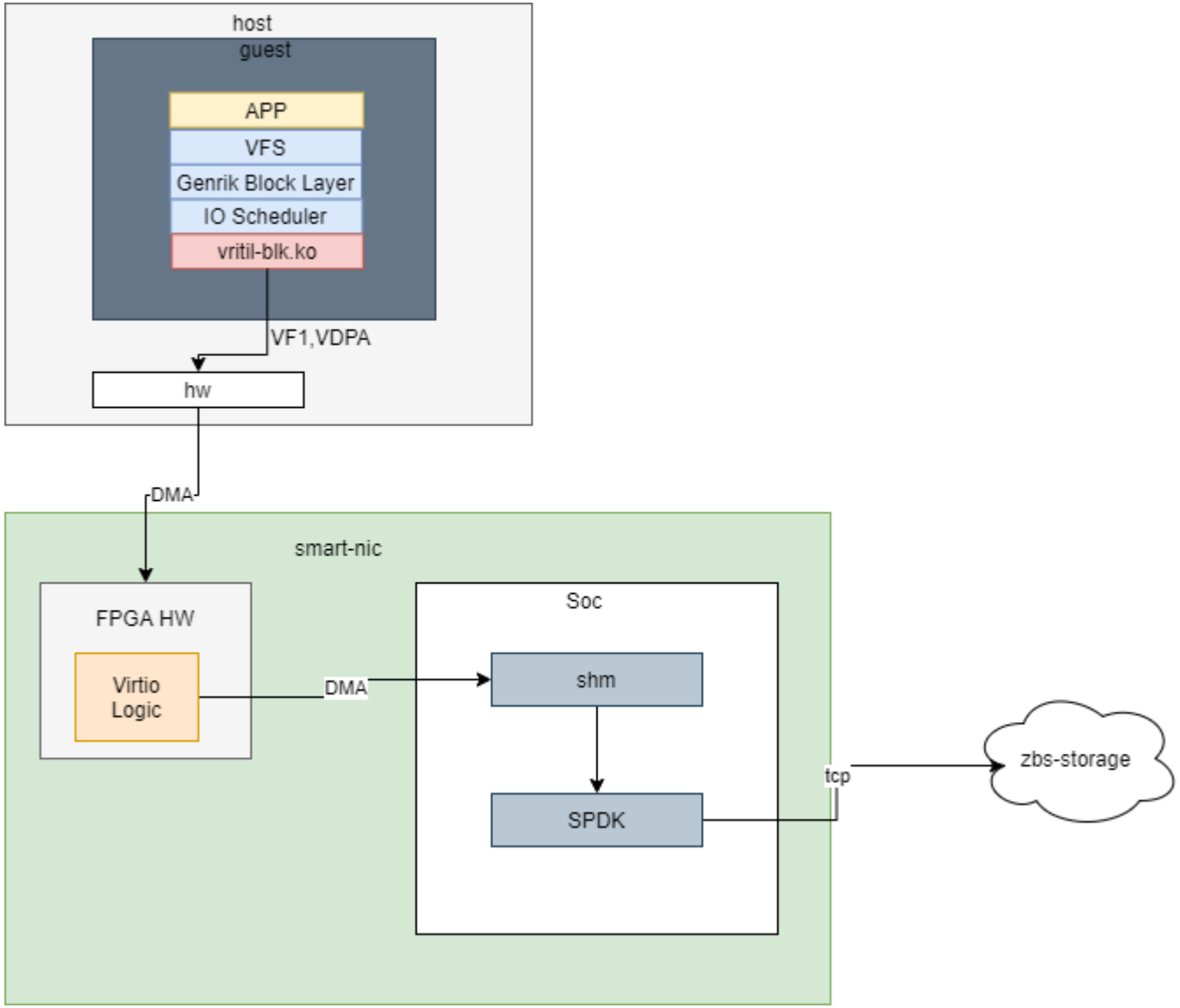


New Storage Architecture

JDT 京东科技

qemu + spdk

- Shorter io path
- Lower virtualization overhead
- Better storage performance
- Improved service stability



JDT 京东科技

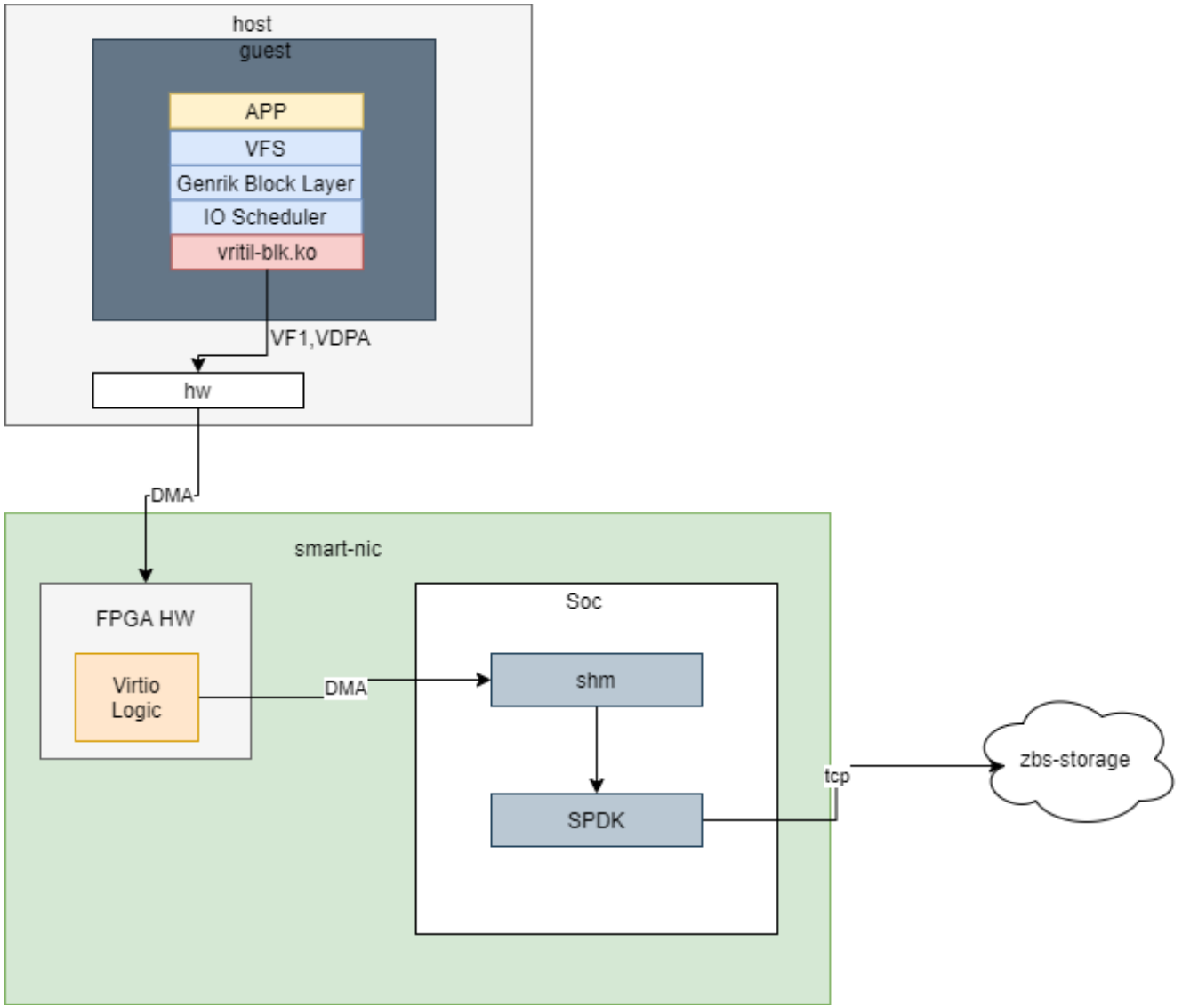
03

Engineering optimization practices

Problems on the new platform

JDT 京东科技

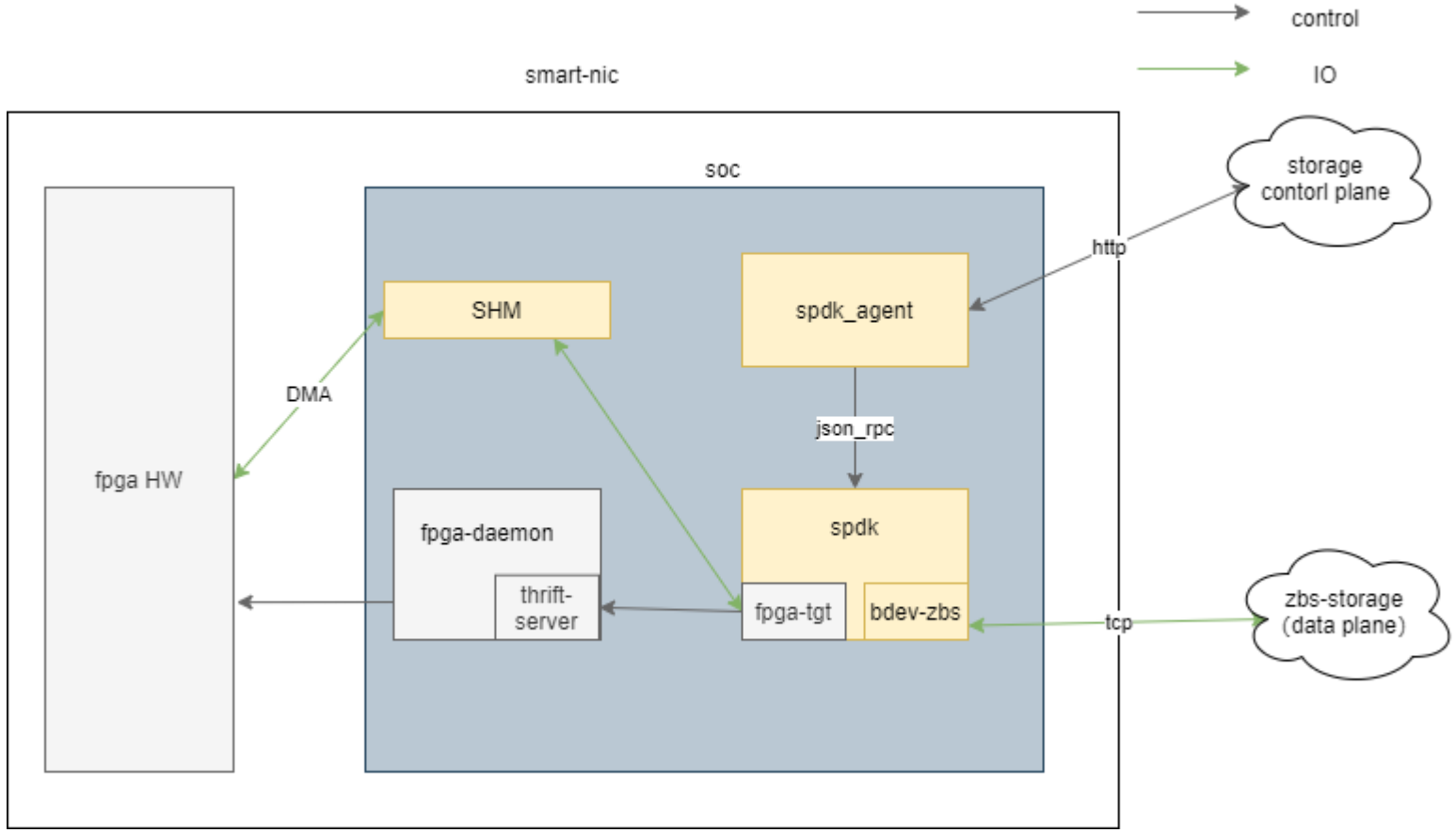
- R&D effectiveness trade-offs
  - Ability reuse
  - Technical selection
- Hardware capability limitations
  - Longer IO path
  - Host/SOC Status synchronization
  - Resource restrictions



R&D Efficiency Tradeoff

JDT 京东科技

- Re-use control plane capability
- Re-implement data plane：VDPA &&SPDK

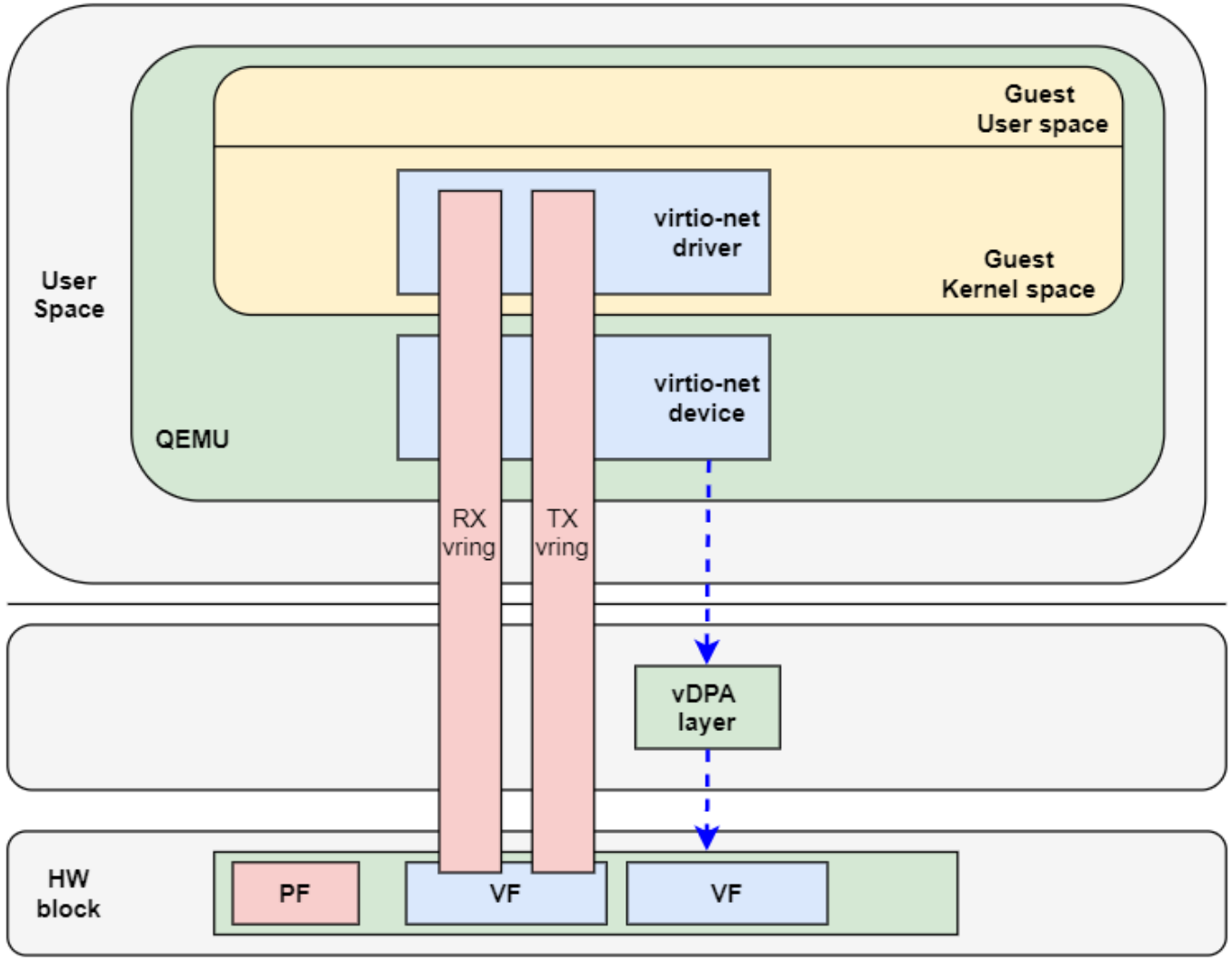


VDPA

JDT 京东科技

Virtio-net -> vdpa + sriov

- zero copy
- zero interrupt overhead
- Shorten io path

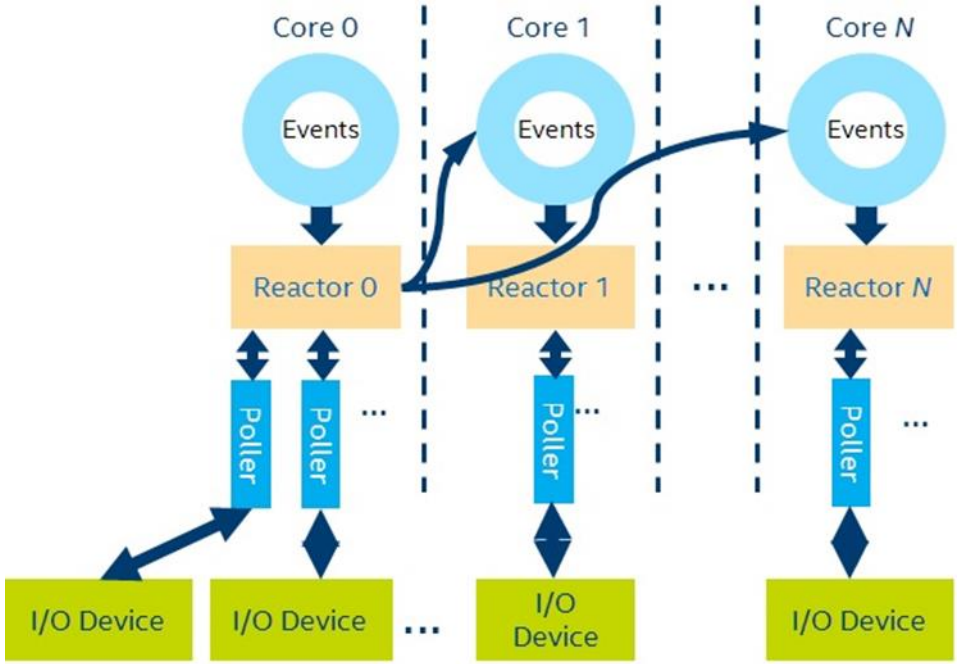


SPDK

JDT 京东科技

High Performance

- PMD
- User Space Driver
- Huge page mem
- CPU affinity
- lockless

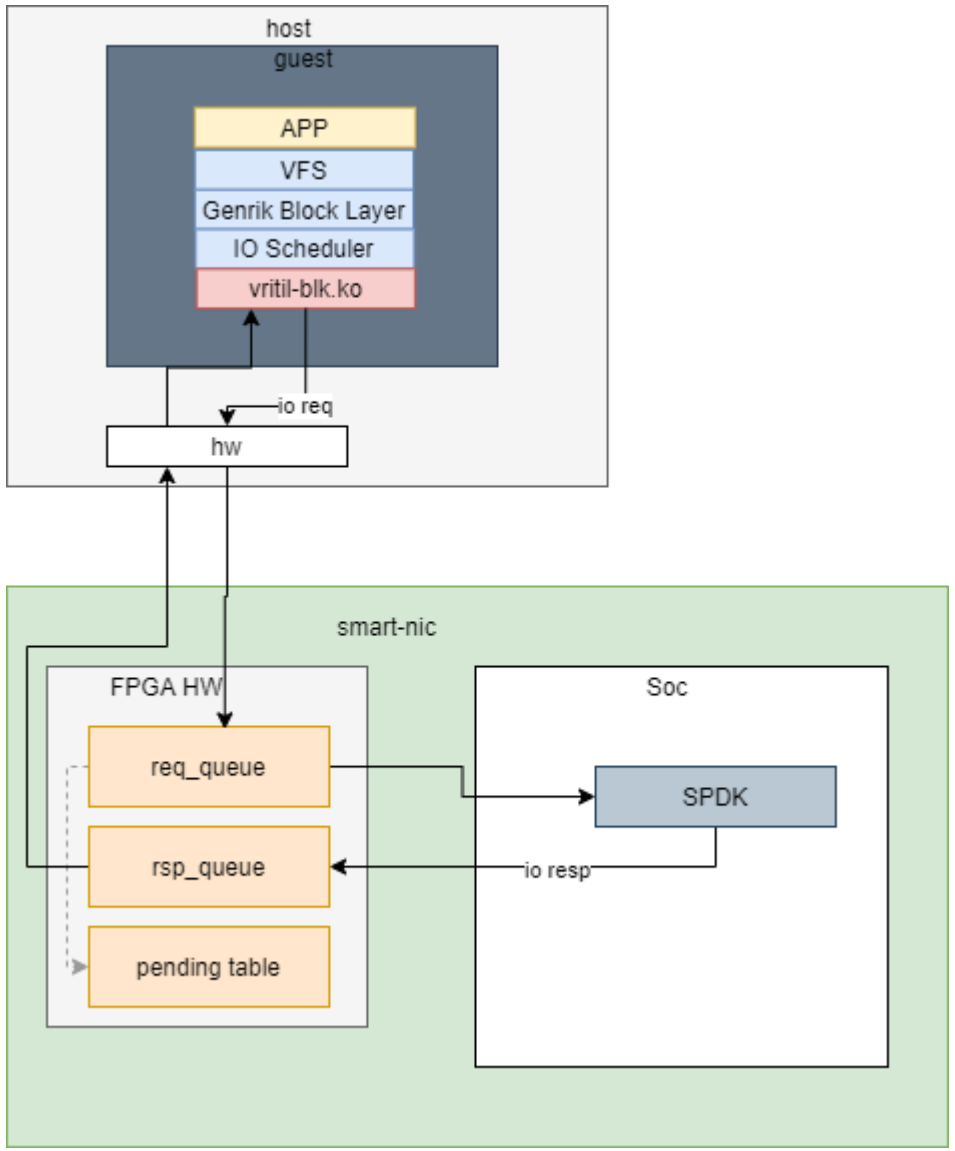


Hardware optimization

JDT 京东科技

Our FPGA team provides the ability to solve soc，host status synchronization issues

- Hold io
- Host status check

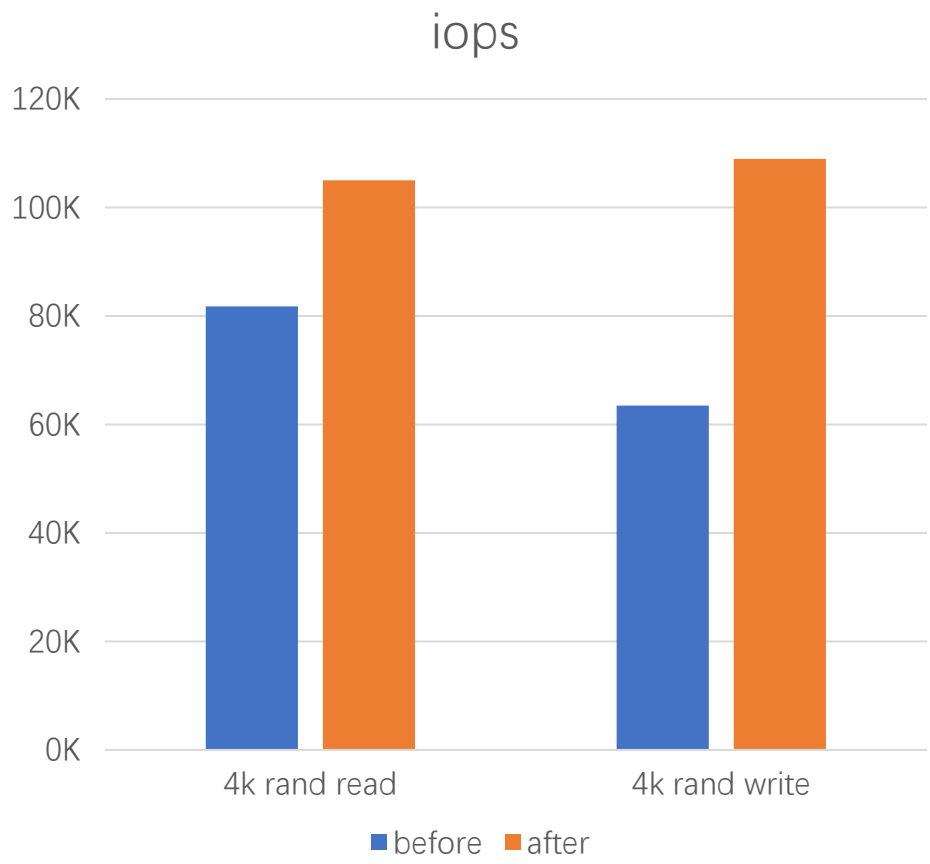


System optimization

JDT 京东科技

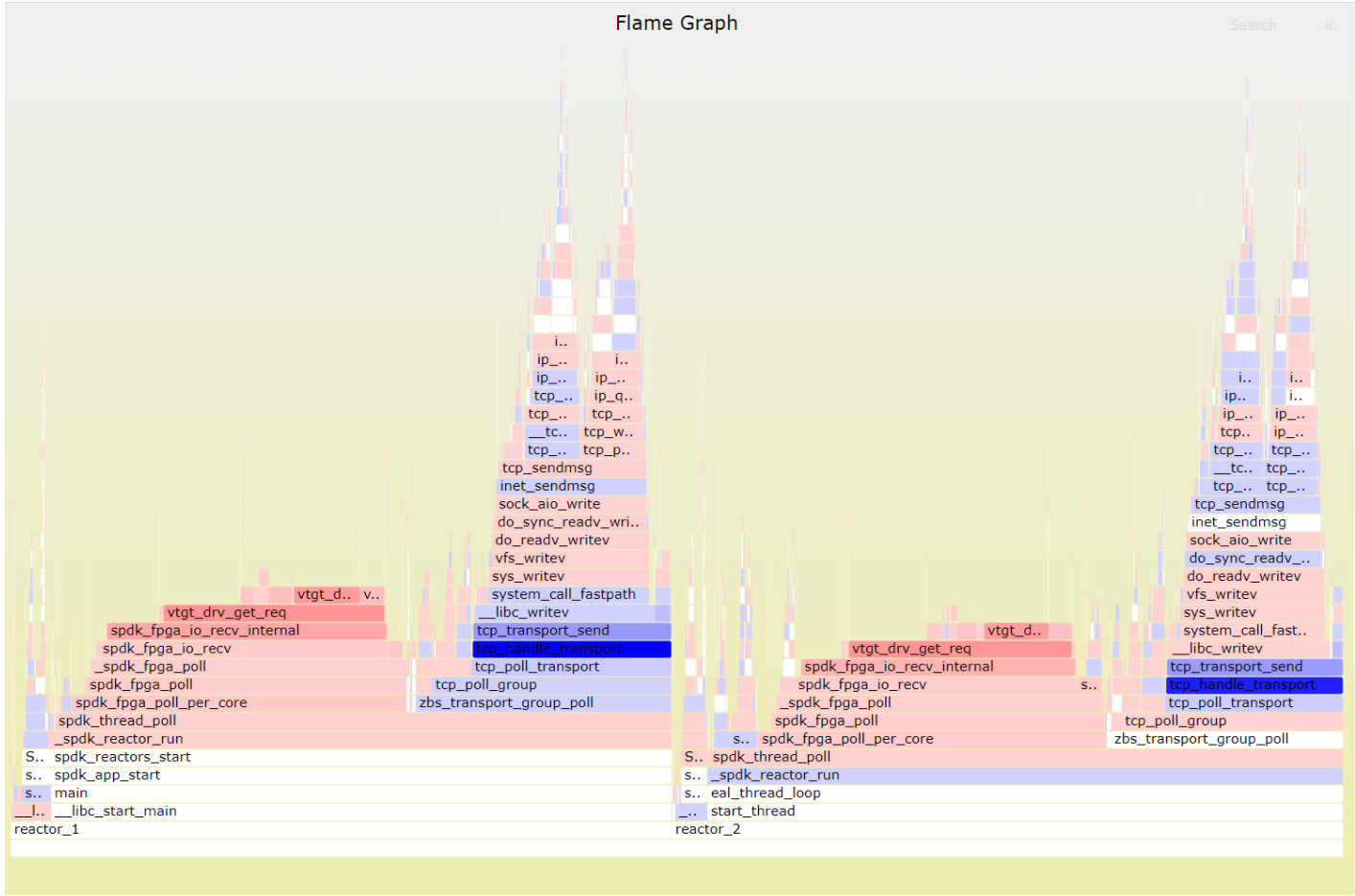
Reduce CPU losses and increase throughput

- Tso
- NIC interrupted
- NIC ring buffer
- IrqBalance
- IPtable
- CPU patch



Software optimization

- SIMD
- Message batch size tune
- Hotspot code optimization



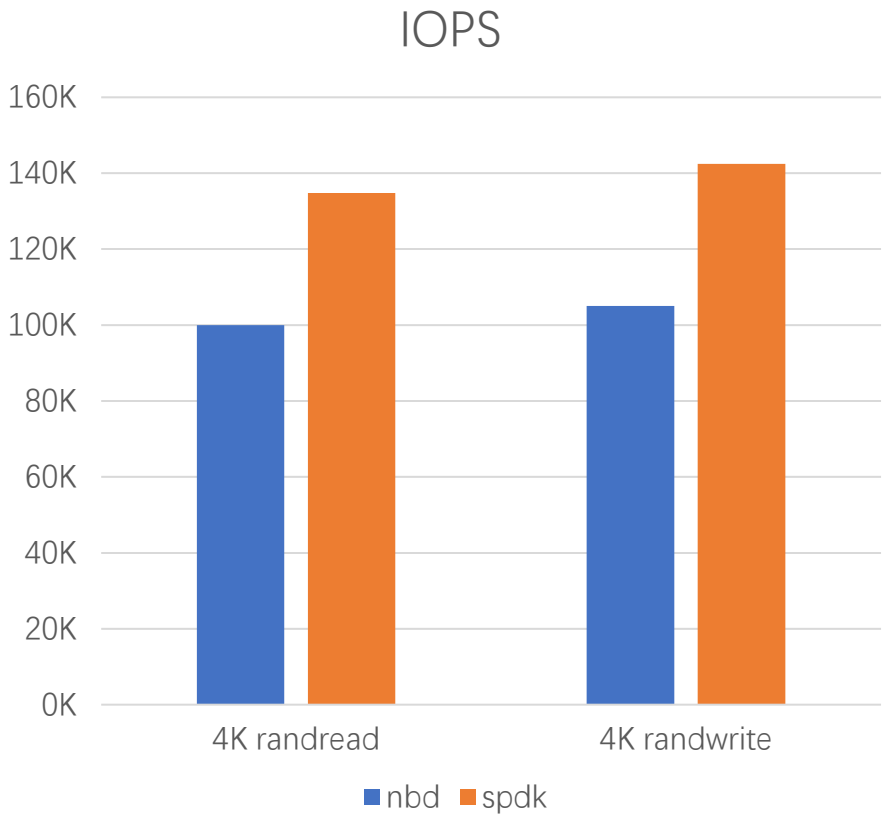
JDT 京东科技

04

Summary

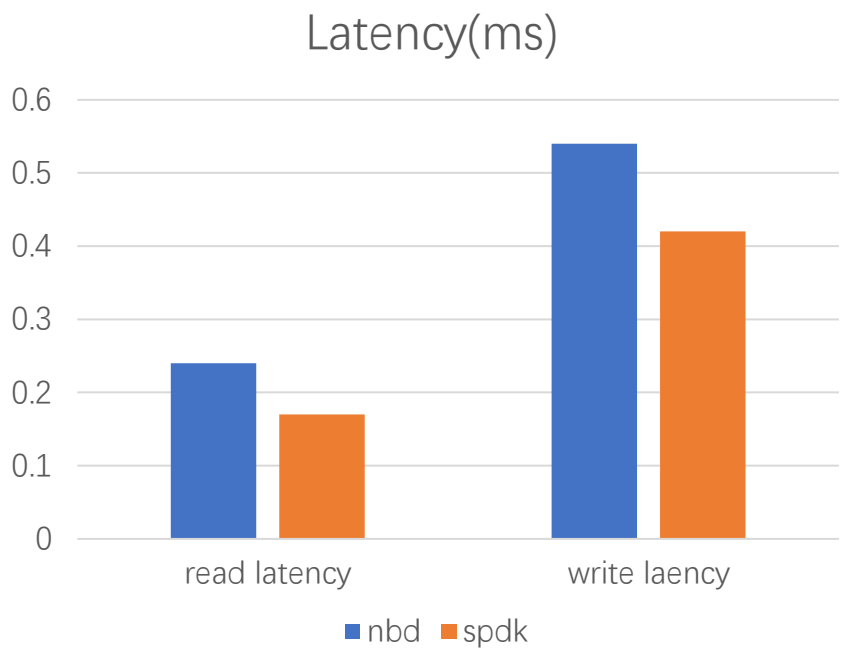
Summary

- 30% iops performance improvement on single volume
- Upper limit of IOPs of the whole SOC is lower than nbd-client， due to SOC performance limit



Summary

- 15% reduction in latency



Future work

JDT 京东科技

- User-level TCP
- RDMA
- Next-generation storage engine

JDT 京东科技

Q & A