SK Telecom:  
All NVMe Flash Array  
Hyper-Efficiency in Rack Scale  
Eric H. Chang/Manager/SK Telecom  
(echang@sk.com)
This session will discuss...

- The need for a new level of efficiency within the infrastructure
  - Nice-To-Have has become Must-Have for real-time AI and Deep Learning applications

- New rack scale storage technology for real-time applications and for a more efficient infrastructure
  - The advantages of PCIe Storage
  - Performance improvements enabled by the new NV-Array

- The future of the NV-Array
New Requirements for Modern Telecom Infrastructures
New Efficiency Requirements

- Comparing Big Data batch processing vs. Real-time analytic applications

**Big Data Batch Processing Apps**
- Accumulated data analysis and prediction
- Batch processing
- Static data sets
- Petabytes of data
- Decisions in Minutes to Hours

**Real-Time Analytic Apps**
- Live data analysis and decision-making
- Low latency
- Live streaming data sets
- Terabytes of data
- Decisions in Seconds to Minutes

Accumulated data analysis and prediction
Batch processing
Static data sets
Petabytes of data
Decisions in Minutes to Hours

Live data analysis and decision-making
Low latency
Live streaming data sets
Terabytes of data
Decisions in Seconds to Minutes
Requirements for Real-time Apps (RTA)

**Time critical**
- Low Latency
- Dynamic response to live inputs

**Mission critical**
- No service interruptions or data drops
- High decision accuracy

**Diverse processing requirements**
- One or a few universal solutions can’t cover the local variations such as regions, languages and customs.
- Different implementations and optimizations for different environments
Required Storage Innovation for RTA

• Storage, as the one of major infrastructure elements, MUST see substantial innovation:
  – Needs to be re-implemented to support the RTA with low latency, high capacity and reliable design
Introducing the NV-Array D20
Cut-To-The-Chase: PCIe only, Flash only!

- SKT has focused on PCIe to directly connect CPUs to NVMe storage devices
  - Avoids the performance cost of bus translation

- NV-Array D20: PCIe JBOF with all NVMe SSDs
  - 52.8GB/s sequential access
  - 13.2M IOPS random access
NV-Array components

- Re-timer cards
- MiniSAS HD cables
- NV-Array (JBOF)
SKT has developed the more reliable re-timer card to connect NV-Array with Hosts based on the Microsemi switch.

[Block Diagram]

[PCB Layout]
NV-Array D20 Design

- Dual switch implementation doubles the performance to 52.8GB/s and 13.2M IOPS
Usage model of the NV-Array D20

- NV-Array used as a centralized DAS pool
  - Traditional servers can leverage the exceptional performance of the NV-Array
  - Servers dynamically assigns NVMe storage capacity as demand grows
  - Lightly loaded servers can release unused capacity
P2P Communication over PCIe fabric

- Communications over the PCIe fabric substantially improves latency (50-100% projected)
Evolution of the NV-Array

- Double the bandwidth by adopting dual switch chips
- Additional value-added features will be added as the design evolves

[Yesterday: C20]

[Today: D20]

[Tomorrow: TBD]

*Production version*
Performance Comparison: C20 vs. D20

- 2X Throughput/IOPS realized!
  - 6.7MIOPS vs. 13.2M IOPS

[Flash Memory Summit 2016: C20 ]

[OCP US Summit 2017: D20]
The Future of the NV-Array and Beyond...
Dynamic Pooled Storage Management

- Pooled NV-Arrays will be managed through RSD/Redfish compliant APIs

Composition #1

Composition #2

Composition #3
High-availability of NV-Array

- HA capable NV-Array targets Telco and Enterprise infrastructures
  - No single point of failure
  - Hot swappable switch boards
  - Hot swappable drives (Dual-port enabled)
  - Failover supported by the device driver
Features Under Evaluation

- **SR-IOV**
  - Adding SR-IOV to the NVMe Drives enables a number of new features

- **NVMe Over Fabrics**
  - NVMeOF enables system expansion across racks or PoDs
A Project Under Evaluation

- SKT is looking into the integration project beyond the NV-Array storage for Deep Learning and AI Infrastructure (Coming in late 2017)
Summary

• Real-time applications demand infrastructure innovation.

• Higher levels of operating efficiency and lower latency must be achieved.

• The SKT NV-Array is a essential building block with which to push the envelope of the infrastructure capabilities.

• The upcoming NV-Array will provide much higher levels of manageability and reliability.