Big Data Flash

Roark Hilomen presented by Jim Fitzpatrick
Eng Fellow – Data Center Solutions

9/14/2017
Agenda

1. Introduction
2. Primer to WDC (HGST/SanDisk)
   A. HDD Technology
   B. Flash Technology
3. Competitive Media Comparison
4. A System Level View
5. The Future (3+ years)
6. Summary
About Western Digital Corporation

TRUST
- A data storage leader
- “Under the hood” of major OEMs
- Millions of devices shipped annually

VALUE
- Among the highest SSD/HDD MTBF rates
- Certified solutions (Msft, Oracle, VMware)
- Single source for storage devices & systems

INNOVATION
- 12,000+ active patents, 500 new/yr
- $2.5 billion invested annually in R&D
- Leader in Open Community initiatives
Broad Storage Portfolio

**Consumer**
- Imaging
- WD Software
- microSD™ Cards & Mobile Storage
- Wearable Audio
- USB Drives

**Mobile & Industrial**
- Embedded & Removable
- Surveillance
- Solid State Drives (internal and external)
- Wearable Audio
- USB Drives

**Client Computing**
- Solid State Storage
- Solid State Storage
- HDD Storage
- Solid State Storage
- Wearable Audio
- USB Drives

**Data Center/Enterprise**
- Solid State Storage
- Wearable Audio
- USB Drives
- Solid State Storage
- Wearable Audio
- USB Drives

©2016 Western Digital Corporation or affiliates. All rights reserved. Confidential.
Primer to WDC Media Offering (HDD)

Capacity targets next 3 years

- 12TB
- 14TB
- 20TB+
Primer to WDC Media Offering (Flash)

Rohit Gupta: SanDisk Blogger
https://itblog.sandisk.com/hdds-vs-ssds-evolution-storage/

Bits per Cell
- MLC: 2 Bits / Cell
- TLC: 3 Bits / Cell
- QLC: 4 Bits / Cell

Layers
- Planar (1 Layer)
- 32 Layer (Samsung)
- 48 Layer
- 64 Layer (Today)
- 96 Layer (Next Leap)
- 128 Layer
- ?

©2016 Western Digital Corporation or affiliates. All rights reserved. Confidential.
Primer to WDC Media Offering (NVMe 96 Layer, QLC)

Legal: The following are technology projections and not a specific product offerings

Key Characteristics
Capacity: 100TB per 2.5”/15mm
Interface: PCIE/NVMe
Read/Write IOPS: 100K/25K (Large Seq Block)
WORM (Write Once): 15+ years
Write Seldom (<60): 10+ years
Throughput (R/W): 4GB/s up to 1GB/s
Power (goal): 12w (Full write), 6-9w (mixed workload), less than 2w (Idle)
** Power Management can reduce avg power use significantly
## Competitive Comparison (HDD vs SSD)

<table>
<thead>
<tr>
<th>Spec</th>
<th>HDD (14TB)</th>
<th>LTO8</th>
<th>96L QLC</th>
<th>Ratio SSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacity</td>
<td>14TB</td>
<td>12.8TB</td>
<td>100TB</td>
<td>7x / 7.8x</td>
</tr>
<tr>
<td>Uncompressed W</td>
<td>180MB/s</td>
<td>470MB/s</td>
<td>1GB/s</td>
<td>5x, 2x</td>
</tr>
<tr>
<td>Uncompressed R</td>
<td>200MB/s</td>
<td>470MB/s</td>
<td>4GB/s</td>
<td>27x, 10x</td>
</tr>
<tr>
<td>Power Operating</td>
<td>7w</td>
<td>*</td>
<td>9w (mixed)</td>
<td></td>
</tr>
<tr>
<td>Power Idle</td>
<td>5w</td>
<td>0</td>
<td>2w**</td>
<td></td>
</tr>
</tbody>
</table>

* Based on tape drive, controller and robotics
** Based on maintenance and potential Power Management
## A System Level View

### Rack Level Comparison (40U Usable rack space)
- HDD Rack – 4U Server w/ 60HDD
- Tape – Quantum Scaler i6000 extrapolated using LTO8
- QLC Rack - 1 2U Server + 3x 2U PCIE Storage Chassis (96 SSD in 8U)

<table>
<thead>
<tr>
<th>Spec</th>
<th>HDD</th>
<th>Tape</th>
<th>QLC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacity No Compression</td>
<td>8.4PB</td>
<td>10.5PB</td>
<td>48PB</td>
</tr>
<tr>
<td>Throughput/W</td>
<td>108GB/s</td>
<td>*Based on Drives</td>
<td>480GB/s</td>
</tr>
<tr>
<td>Throughput/R</td>
<td>120GB/s</td>
<td>*Based on Drives</td>
<td>2.4TB/s</td>
</tr>
<tr>
<td>Throughput/M</td>
<td>12GB/s</td>
<td>*Based on Drives</td>
<td>*Based on Mix 400GB/s-2TB/s</td>
</tr>
<tr>
<td>Power Mixed</td>
<td>6KW</td>
<td>*</td>
<td>5KW</td>
</tr>
<tr>
<td>Power Idle</td>
<td>4.8KW</td>
<td>*</td>
<td>1.8KW</td>
</tr>
</tbody>
</table>

### Network Throughput (12 Uplinks)

<table>
<thead>
<tr>
<th>Network</th>
<th>Throughput</th>
</tr>
</thead>
<tbody>
<tr>
<td>10Gb</td>
<td>9.6GB/s</td>
</tr>
<tr>
<td>40Gb</td>
<td>36GB/s</td>
</tr>
<tr>
<td>100Gb</td>
<td>96GB/s</td>
</tr>
<tr>
<td>400Gb</td>
<td>384GB/s</td>
</tr>
</tbody>
</table>
A System Level View

• **Limiting Factors**
  – Network Bandwidth
    • All media can scale on the compute side to exceed Network Bandwidth @10Gb
    • Transition to 100Gb and faster, Only SSD justify /can fully utilize
  – Software (not Optimized for Flash)
    • Object Based Erasure Coded/Compressed data to fully utilize and keep data
    • Resilient over 15+ years

• **Deployment Model**
  – As the density increases, rack level distribution becomes key to realizing throughput.
    • New Models distributing ‘Archive’ across active racks (Top of Rack Archive, Top of Rack Storage)
  – At 7x capacity, almost comparable power use (minus real deep archive/vaults)
    • QLC may expand where flash can be utilized
    • 10-15yr TCO on 48PB raw / Rack cost

• **Risk**
  – Limited NAND Fabrication Capacity
    • Exabyte level as we enter Zetabyte scale world
    • Choice of Fast Data or Big Data
TCO Example: 2015 Infiniflash 8TB Flash Cards

Food for thought

Total Energy Cost

- InfiniFlash
- AFA
- External AFA
- DAS SSD
- DAS 15k HDD
- DAS 10k HDD
- DAS 7.2k HDD
- DAS 15k HDD

Data Center Racks

- InfiniFlash
- AFA
- External AFA
- DAS SSD
- DAS 15k HDD
- DAS 10k HDD
- DAS 7.2k HDD

© 2016 Western Digital All rights reserved.
3. The Future (3+ years)

• Capacity Growth
  – SSD Vendors now have 3 knobs to turn for capacity
    • Layers (3d Nand)
    • Bits per Cell (1,2,3,4.. Yes more is possible, but with distinct limits)
    • **Lithography (65nm , 3xnm... )
  – Packaging
    • Infiniflash 1st in custom packaging..
    • EDSFF Ruler (U.2 capacity in 1U package)
    • 3.5” Flash drives?
    • Various ‘Flash Stick’ standards popping up (M.2,others)

• As value of data increases and cost of storage decrease – more data will be stored
3. Summary (Thoughts)

• Shifts are not made overnight. Flash + Tape, Flash + HDD + Tape will continue to exist. Ratio and length of retention/data activity will be key... Even a $0.01/GB difference is equivalent to $1M at 100 PB, moving to a Zetabyte, new architecture will be key.

• As value of data increase and infrastructure cost decrease, trend is showing active data life is increasing from 1-6mo to 3-6 years and capacity of archive data growing.

• Flash has limits on its fabrication capacity. Decision must be made on Fast Data Flash, Big Data Flash or Archive/Compliance Flash (QLC).

• The dynamic is changing and it’s up to the people in the room to decide how it will change.