Adaptable Storage Acceleration Platforms for Exabyte-scale Data Centers

Manish Muthal
Vice President Datacenter Marketing
August 9, 2018
Three Big Trends

Computing After Moore’s Law

> Frequencies have hit a brick wall
> Cost per unit of yielded area is going up
> Compute is becoming a precious resource that must be conserved
Three Big Trends

Explosion of Data

- Video and IoT driving explosion of data
- Flash Storage driving lower latency and higher bandwidth
- Moving Data Back and Forth between CPU and Storage creates bottlenecks and Consumes Power

More Data from the Network

More Data per Server

Increasing SSD Densities

© Copyright 2018 Xilinx
The Dawn of AI

- Machine Learning, Video and Image Processing, Big Data Analytics…
- Increasingly being used in combination
- New Applications and Algorithms coming at a furious pace
The Need for Adaptable Storage Platforms

The Answer: Put Adaptable Acceleration Closer to Storage

- Move Intensive Workloads into Efficient Accelerators
- Put Acceleration Closer to Storage Endpoints
- Adaptable and Reconfigurable Acceleration

Different Accelerators for Different Workloads
- ML
- Analytics
- Video
- Storage

Adaptable Accelerators for Variable and Changing Workloads
The Need for Adaptable Storage Platforms

FPGAs Provide the Ideal Platform for Distributed Storage Acceleration

11x Machine Learning
34x Data Analytics
40x Video Processing
40x Storage Compression

11-40x Improvement in Efficiency (Perf/Watt) over CPU

Easy to Attach to a Variety of Storage Endpoints

Adaptable and Dynamically Reconfigurable
Moving Adaptable Acceleration Close to Storage

Solving The Needle In the Haystack Problem

- Search for Needle in CPU/Accelerator
- Sequential Scan of All Drive Data
- I/O and Processing Bottlenecks
- High Power Consumption to Move All Data

- Move the Needle down to Distributed Accelerators
- Individual Scan from Distributed Accelerators
- Eliminate I/O and Processing Bottlenecks
- Localized Data Inspection Reduces Power
## Improving Total Cost of Ownership

### Significant CAPEX/OPEX Reduction

- Accelerators free up general compute cycles
- Result is more work per server or less servers per work
- Accelerators add only incremental cost/power (10-20%)
- Results in Capex/Opex gains for DC Operators

<table>
<thead>
<tr>
<th>DC Workload Mix</th>
<th>Acceleration Ratio</th>
<th>Unaccelerated</th>
<th>Accelerated</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Purpose Compute</td>
<td>N/A</td>
<td>40%</td>
<td>78%</td>
</tr>
<tr>
<td>Storage Services</td>
<td>40</td>
<td>10%</td>
<td>0.25%</td>
</tr>
<tr>
<td>Database/Big Data</td>
<td>34</td>
<td>10%</td>
<td>0.29%</td>
</tr>
<tr>
<td>Video Transcoding</td>
<td>40</td>
<td>10%</td>
<td>0.25%</td>
</tr>
<tr>
<td>Machine Learning</td>
<td>11</td>
<td>10%</td>
<td>0.91%</td>
</tr>
<tr>
<td>Idle/Overhead</td>
<td>N/A</td>
<td>20%</td>
<td>20%</td>
</tr>
</tbody>
</table>

Server CPU Utilization

- Server Cost = X
- Server Cost Accelerated = X + 15%
- Server Power = Y
- Server Power Accelerated = Y + 20%

10,000 Servers

- 5,128 Servers

42% Capex Reduction
38% OPEX Reduction
Building a Comprehensive Ecosystem of Applications

The Application Ecosystem for the Adaptable Storage Platform is Strong and Growing Rapidly

DEVELOPERS

100% QoQ Growth of Published Applications in FY18

Hundreds of Developers Trained Every Quarter

1500+ Developers Trained by Year End

LIFE SCIENCES

FINANCIAL

VIDEO

MACHINE LEARNING

DATABASE

Framework, API, Python/Java/C++ Programmability

Adaptable Storage Platform

© Copyright 2018 Xilinx
Looking Towards the Future

7nm Rocket Fuel

> New Device Category for Adaptive Workload Acceleration
> Network-On-Chip
> HW/SW Programmable Engines
> Coming to Adaptable Storage in 2019!
Thank You

> Please Come Visit Xilinx in our Booth #313

> Live Demos!
  >> Storage Compression Offload
  >> Programmable Controller for Software-defined Flash
  >> Database Acceleration on FPGAs
  >> NVDIMM-N Solution
Adaptable. Intelligent.