



How Flash Changes the Hardware Refresh Cycle in the Data Center

Matt Key
IBM Corporation

Historically



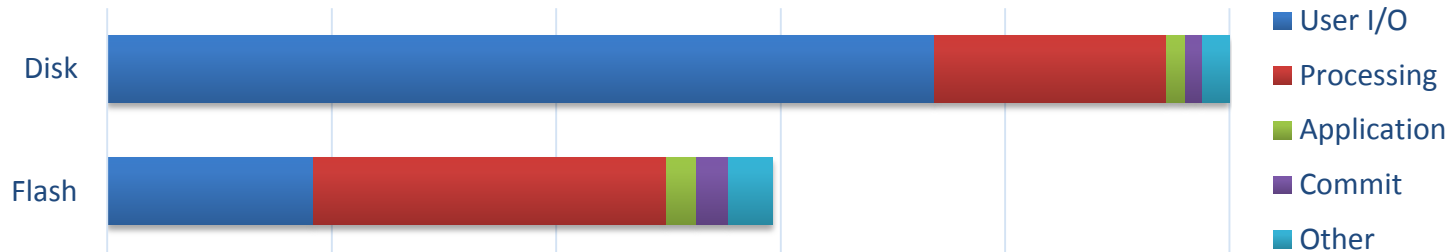
Performance from
Processors

Capacity from
Storage

Reality

IT as Investment for Time

Before & After Flash
(Distribution of Wait Percentage)

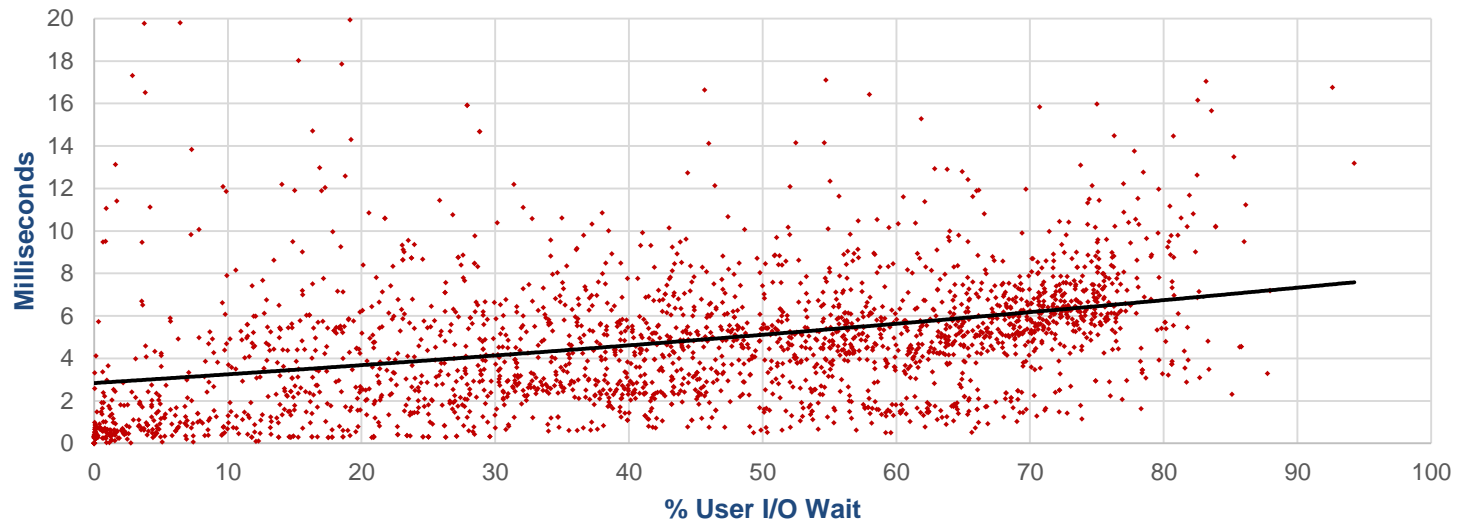


- Storage is a contribution of time
- Evolution of Storage Performance
Persistence → Capacity → Bandwidth → IOPS → Latency
- Focus on \$/Used Capacity & \$/Used Performance

Most Apps are 'Latency Sensitive'

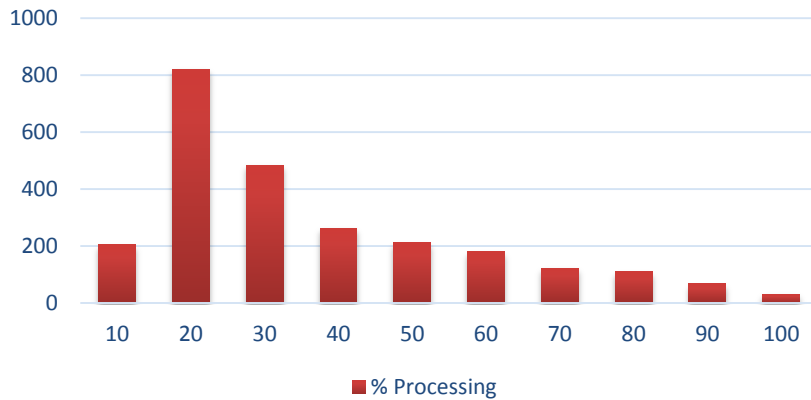
- Every I/O is wasted time; storage time is overhead
- Brief analysis of 183 databases
 - 5,984 database reports
 - Remove idle workloads
 - Mixed workloads, storage, and CPU

% I/O Wait vs Response Time

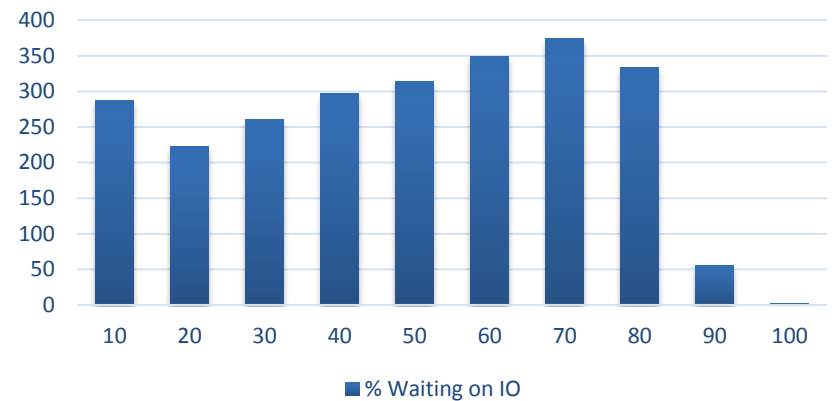


Focus Refresh on Storage

Histogram of CPU Contribution



Histogram of User I/O Contribution



- Majority of applications will benefit greater from lower latency
 - More so than increasing CPU or lowering CPU latency
- Easier to add CPU than add storage
- Greenfield investment should be in storage and high clock speed CPU
- License additional cores for scaling performance

NAND Flash Performance Adoption

- Great things about NAND Flash Evolution
 - Denser & cheaper
 - Higher throughput per chip
- Not getting quicker/slower for getting first 4KB
 - Access latency marginally increasing for increased density
 - Transfer time of data getting better
 - Staying ~100 μ s to get first 4KB data
- Large block apps reaping most benefit from each newer NAND
 - Focusing on getting Small-block / Indexed apps to NAND
- Ride the wave