SSD Architectural Considerations for Data Center Workloads

Paul Truong, Principal Engineer

Data Center Solutions, SanDisk Corp

Forward-looking Statements

During our meeting today we will make forward-looking statements.

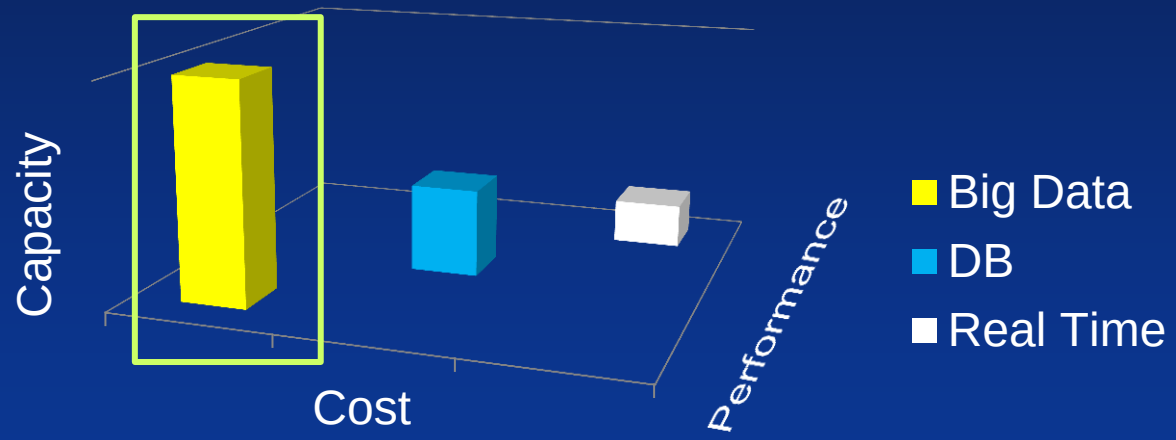
Any statement that refers to expectations, projections or other characterizations of future events or circumstances is a forward-looking statement, including those relating to products and their capabilities, performance and compatibility, cost savings and other benefits to customers.

Actual results may differ materially from those expressed in these forward-looking statements due to a number of risks and uncertainties, including the factors detailed under the caption “Risk Factors” and elsewhere in the documents we file from time to time with the SEC, including our annual and quarterly reports.

We undertake no obligation to update these forward-looking statements, which speak only as of the date hereof.

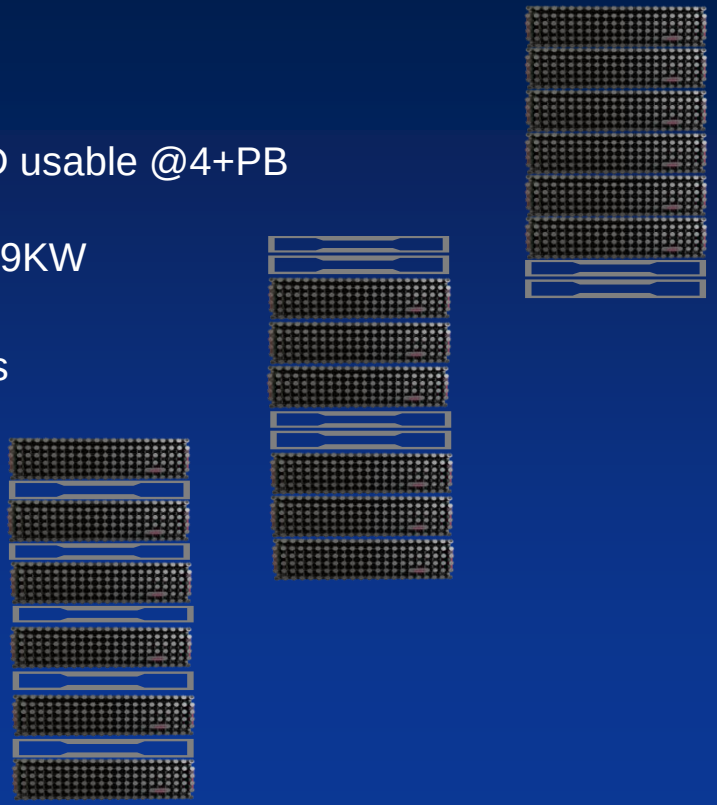
Data Center Workloads

- Application Filter is a modifier to Value



Big Data Use Cases Where Flash is Displacing Traditional Media… Perspective

- Active Archive (Rack)
  - 15 InfiniFlash ™ System IF100s, 2 Servers
  - Capacity 7.5PB (6.8PB Usable) vs. 8TB HDD usable @4+PB
  - Bandwidth 225 GB/s vs. HDD @ 30GB/s
  - Power 3KW Idle, 8KW All Active vs. HDD @ 9KW
- Data Lake Model (Rack)
  - 8-12 InfiniFlash System IF100s, 8-12 Servers
  - Capacity 4-6PB
  - Bandwidth up to 180GB/s
  - Power 6KW Active
- Active Analytics Model (Rack)
  - 4-8 InfiniFlash System IF100s, 8-24 Servers
  - Capacity 2-4PB
  - Bandwidth up to 120GB/s
  - Power 10KW Very Active

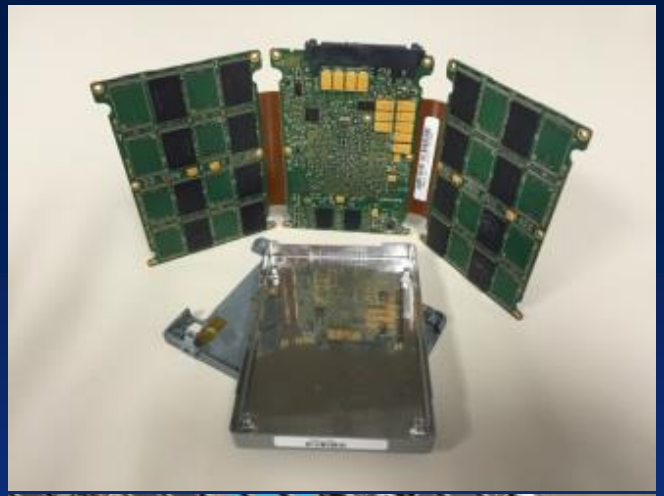


Solutions using SSDs

- Meet performance target but has cost overhead
  - In some cases results in unutilized excess performance
- HDD Form Factor
  - Mechanical (Case, Tri-Fold)
  - Thermal
  - Higher Cost Components (PFAIL Capacitors)
- Lower Capacity Point Availability (up to 4TB)
  - Lower Density Chassis, Higher Chassis Count

Card Form Factor

- Throw away the legacy HDD form factor
  - No case
  - No flex connectors
  - Low cost high profile capacitors
  - 8TB Capacity
  - Improved thermal profile
  - 64 Cards/Chassis (512TB)



1 GB = 1,000,000,000 bytes. Actual user capacity less. The performance results discussed herein are based on internal testing. Results and performance may vary according to configurations and systems, including drive capacity, system architecture and applications.

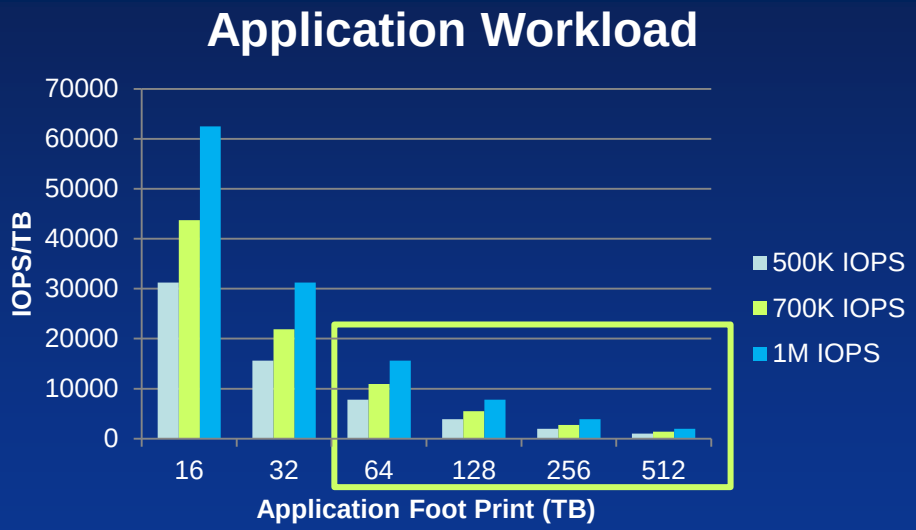
Optimizing Costs for Data Center Workloads

- Capacity & Performance
- Over Provisioning
- Endurance

Capacity

- Increase NAND / non-NAND ratio
  - 8TB/Controller
  - 512TB/Chassis
- Constraints
  - Controller Performance (IOPS/TB)
  - DRAM Size

Right Sizing Performance



1 GB = 1,000,000,000 bytes. Actual user capacity less. The performance results discussed herein are based on internal testing. Results and performance may vary according to configurations and systems, including drive capacity, system architecture and applications.

Right Sizing Capacity

- Typical SSDs @ 80-100K RR IOPs

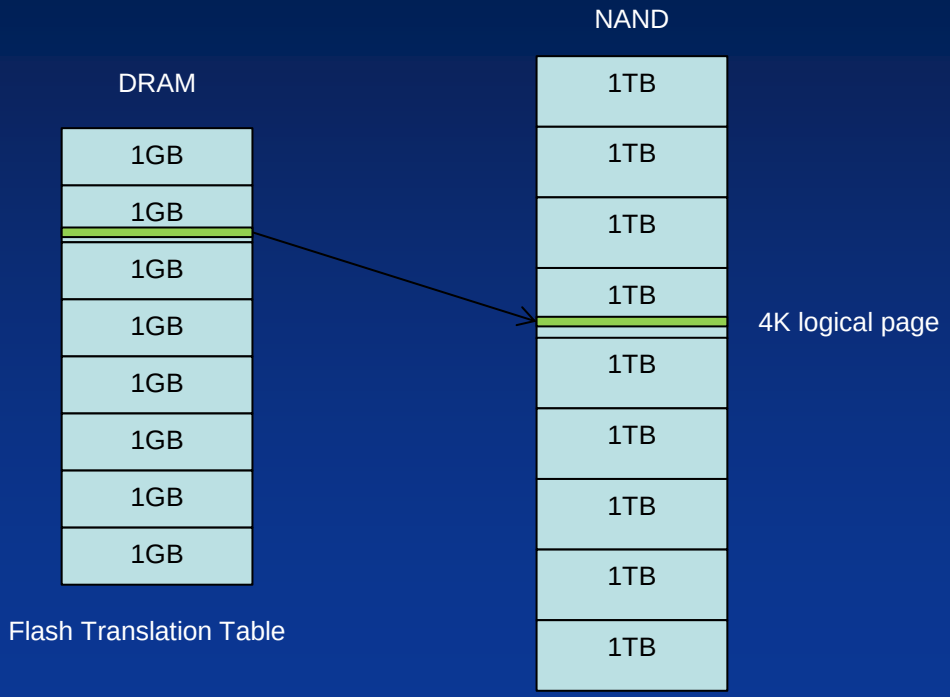
Capacity	IOPS/TB
512GB	160K
1TB	80K
2TB	40K
4TB	20K
8TB	10K

Capacity – FTL Implication & Applications Considerations

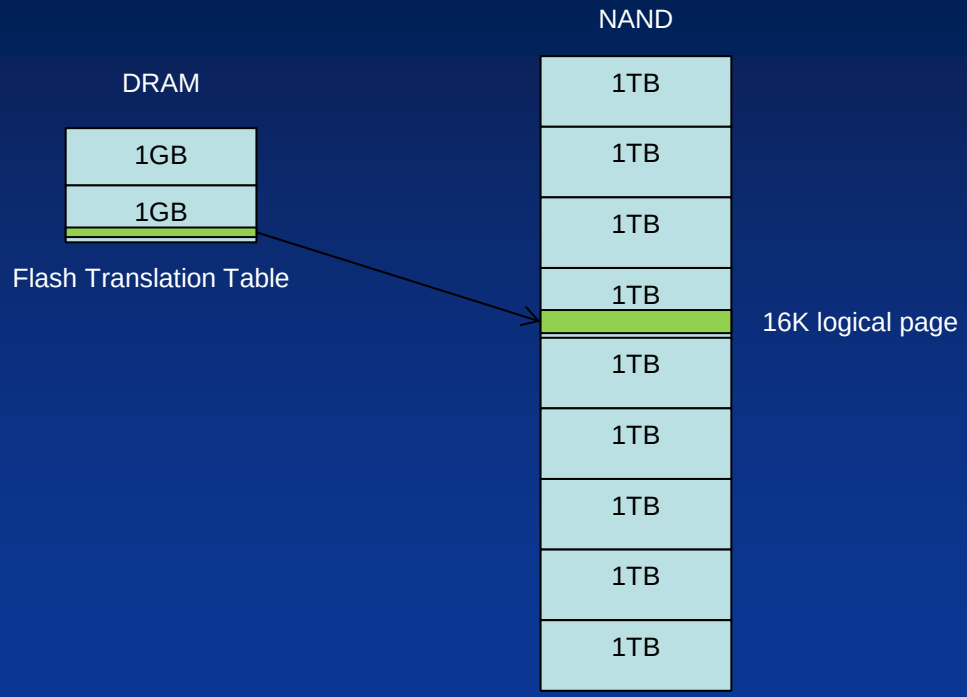
- Large logical space for translation table
- 4 bytes entry / 4K logical page
  - 1GB DRAM addresses 1TB of logical space

Application	Solution Option
Real-time Applications	Increase DRAM
Big Data	Increase granularity of logical page
DB	Caching of translation table

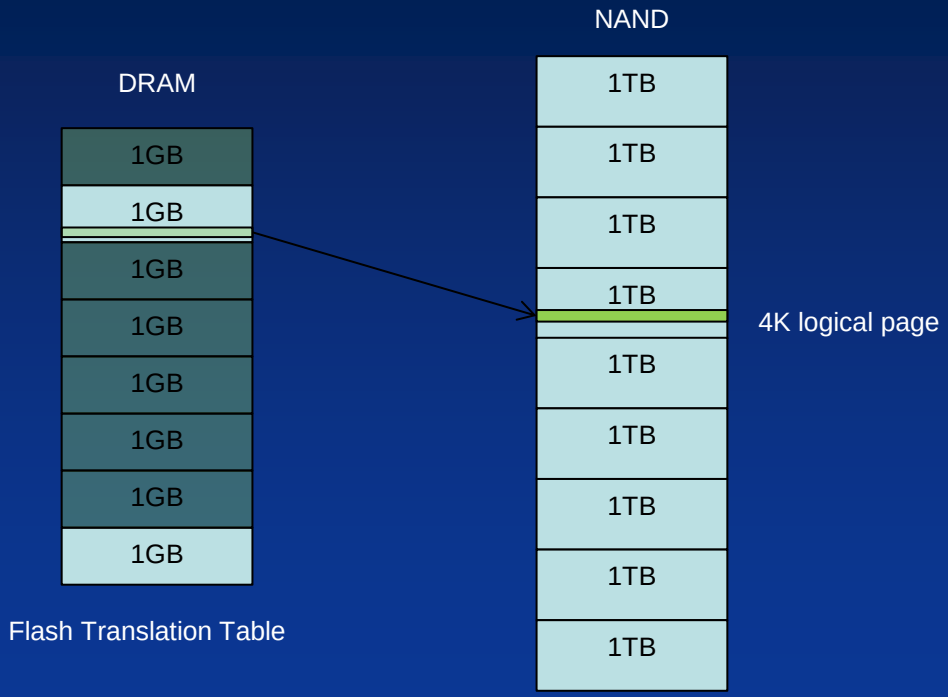
Increase DRAM



Increase Logical Page Size



Caching of Translation Table



Endurance at Scale

- Scale:
  - 512TB InfiniFlash box
  - Bandwidth maxed at 7GB/s
  - 90/10 Read/Write workload
  - 5 year run, 24/7/365
- Result:
  - 100 Petabytes written
  - 200 total box writes

1 GB = 1,000,000,000 bytes. Actual user capacity less. The performance results discussed herein are based on internal testing. Results and performance may vary according to configurations and systems, including drive capacity, system architecture and applications.

Lower OP

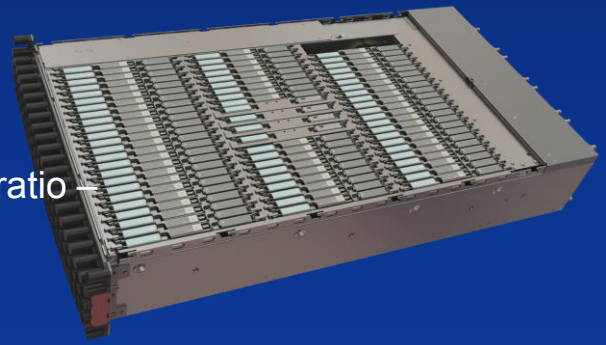
- Reduced endurance requirement
- Higher sequential write workload
- Enables the use of low 7% OP

Read Latencies

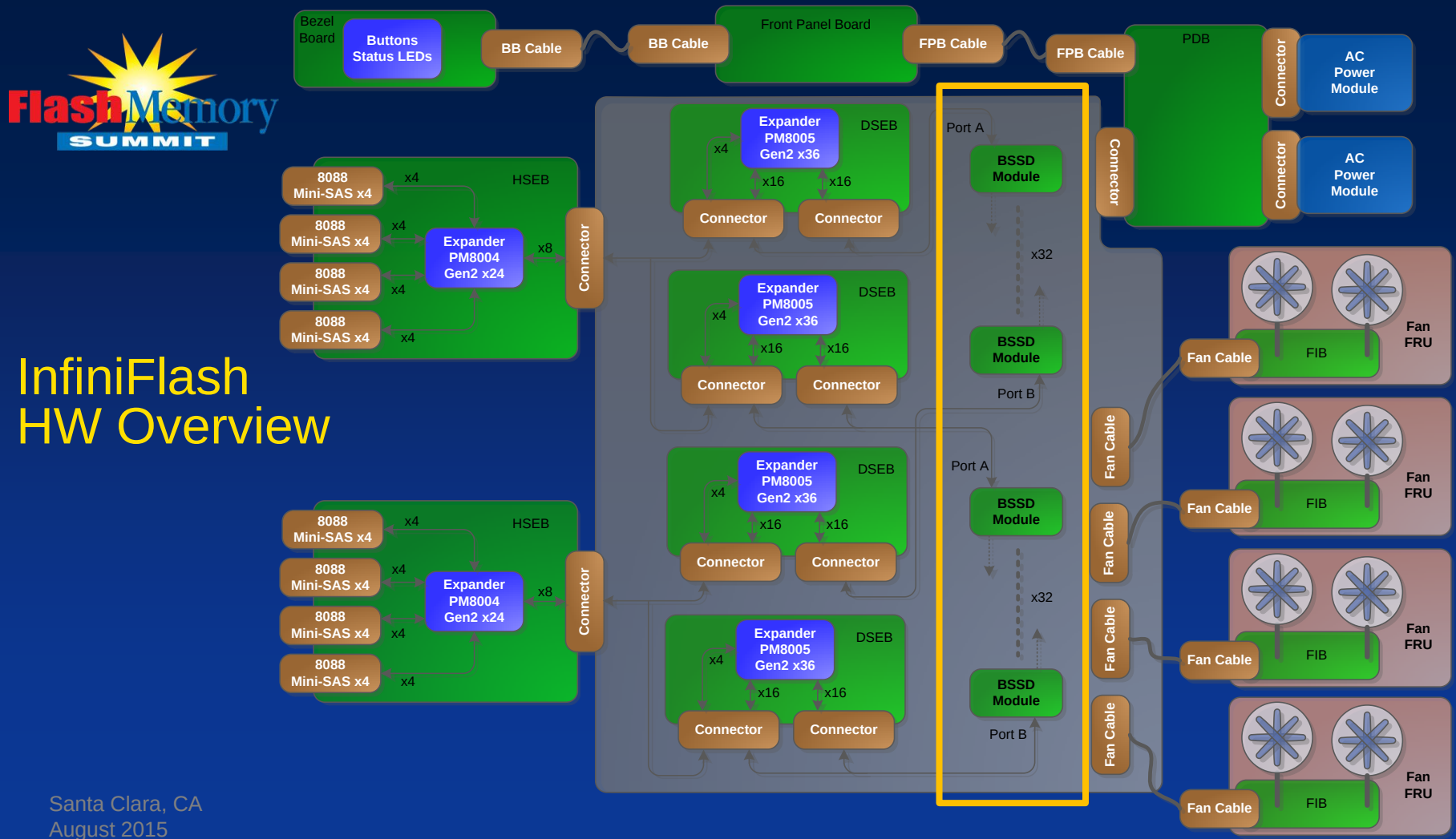
- Read latencies during ingest
- High read latencies incur when read access to die busy with write
- High capacity point reduces probability of event

High Density Chassis

- Achieves higher density of Flash in the same chassis than building with standard SSDs
  - A 512TB InfiniFlash System IF100 has 8x higher flash density than today's densest all-flash array; and 40% higher density than the most aggressive HDD arrays
- IF100 components are designed to be field replaceable
  - Chassis is greater than 1.5 Million Hours MTBF (>99.97% reliability)
  - SAS expander boards, fans and drive cards are all hot swappable and built for HA
- Multiple design decisions made to optimize for cost while providing the required endurance
  - Drive card format chosen to maximize "NAND/non-NAND" cost ratio –
    - Also allows for 4TB, 8TB, 16TB capacities
  - Overprovisioning minimized to 7%



1 GB = 1,000,000,000 bytes. Actual user capacity less. The performance results discussed herein are based on internal testing. Results and performance may vary according to configurations and systems, including drive capacity, system architecture and applications.



InfiniFlash HW Overview

Thank You

Email: paul.truong@sandisk.com

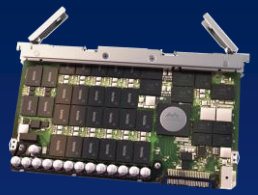
LinkedIn: https://www.linkedin.com/pub/paul-truong/4/27/81b

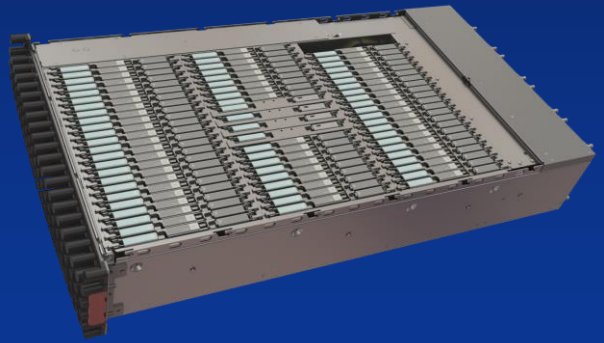
1 GB = 1,000,000,000 bytes. Actual user capacity less. The performance results and cost savings discussed herein are based on internal testing. Results and performance may vary according to configurations and systems, including drive capacity, system architecture and applications.

Back up

Performance at Scale

- InfiniFlash ™ System IF100 - Purpose built for Big Data
  - 64 x 8TB Flash Cards (500TB)
  - 780K IOPS 4K Read (aggregate)
  - 500+ KIOPS 4K Write (aggregate)
  - 7GB/s chassis level bandwidth
  - 1/10 AFR of HDD (20HDD failure for every InfiniFlash Card)
  - Chassis level idle @ less then 200W, Active @ 400-500W
    - (Comparative HDD @700W avg)
  - Endurance (multi-Exabyte at chassis level)
  - 3U





1 GB = 1,000,000,000 bytes. Actual user capacity less. The performance results discussed herein are based on internal testing. Results and performance may vary according to configurations and systems, including drive capacity, system architecture and applications.