Estimating Hardware Storage Costs

Jenny Woolley
William Black

ICEAA 2014 – Denver, CO

Scitor Corporation

* The views expressed in this presentation are those of the presenters and do not imply endorsement by the Office of the Director of National Intelligence or any other US Government agency

Overview

Scitor Corporation

- **RAID storage deflation research became a priority when data was provided for estimates in terms of storage volume, rather than a bill of materials (BOM)**
  - Needed way to translate volume into cost to support future estimates
  - Past research across multiple agencies has provided conflicting results on cost and how it changes over time
- **Agenda**
  - Storage Background
  - RAID Details
  - Storage Deflation
  - Data Analysis
  - Conclusions
  - Next Steps

Types of Storage

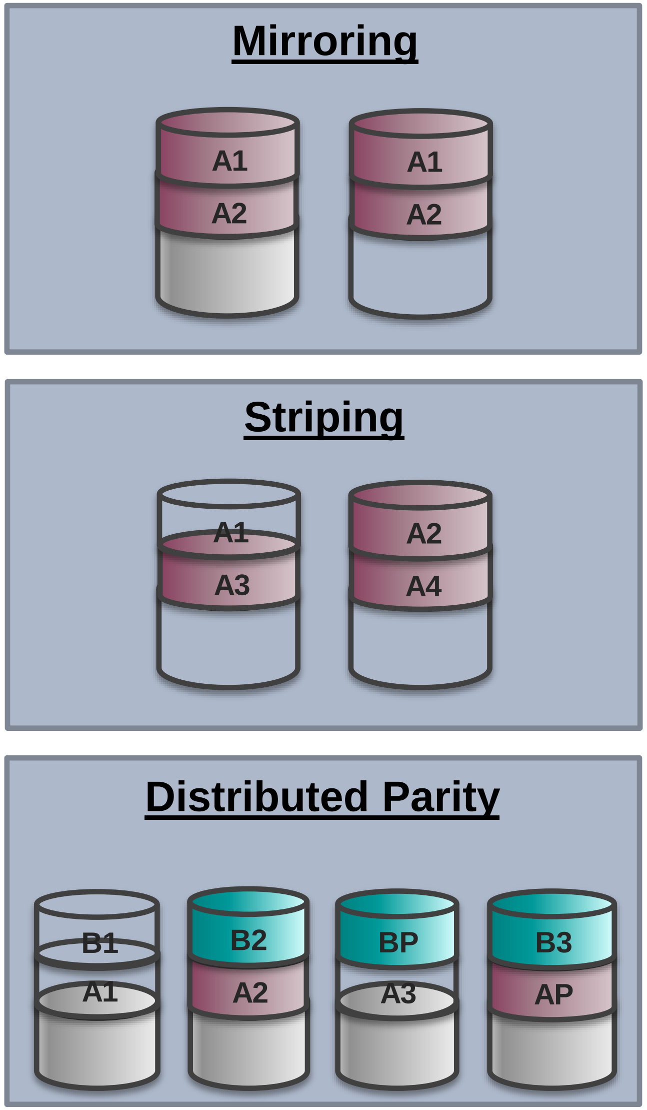
Scitor Corporation

- **Types of storage typically seen in government programs to be estimated:**
  - Tape storage
    - Utilizes tape drives and tape libraries
    - Still used for long-term storage, but has become less prevalent as costs for disk storage have decreased
  - Disk storage
    - Current standard for short to mid-term storage and often used for long-term archival purposes
    - Most commonly seen type of storage in recent cost estimates
  - Solid-state storage
    - High-performance plug-and-play storage device that contains no moving mechanical components
    - Likely to see expanded use in the next several years, but pricing is currently prohibitive for many government programs
- **RAID = Redundant Array of Independent Disks**
  - Combines two or more physical drives into a logical unit presented as a single hard drive to the operating system
  - Different configurations (called “levels”) of RAID utilize multiple techniques to provide varying degrees of reliability (ability to withstand drive failure) and availability (speed of Input/Output)

RAID Terms

Scitor Corporation

- **<u>Mirroring</u>: Duplicating data to more than one disk**
  - Can speed read times because the system can read data from more than one disk
  - Can slow write times if the system must confirm that data is correctly written to each disk
- **<u>Striping</u>: Writing data across a number of disks in parallel**
  - Speeds read/write performance
- **<u>Parity</u>: Redundancy information is calculated for each piece of data stored**
  - If a drive fails, the missing data can be reconstructed from the remaining data and parity data



RAID Levels

Scitor Corporation

Level	Striping	Mirroring	Parity	Notes	Storage for ~1 TB
RAID 0	X			Provides no fault tolerance	1 TB
RAID 1		X		Provides fault tolerance, but can cause a slight drag on performance	2 TB
RAID 2	X			Striping at bit (rather than block) level; not currently used	Not Used
RAID 3	X		X	Byte level striping with a dedicated parity disk; rare in practice	Not Used
RAID 4	X		X	Block level striping with a dedicated parity disk; rare in practice	Not Used
RAID 5	X		X	Block level striping with parity data distributed across all member disks; fault tolerance against one drive failure	1.5 TB
RAID 6	X		X	Block level striping with two parity blocks distributed across all member disks; fault tolerance against two drive failures	2 TB
RAID 10	X	X		Stripe set composed of two or more mirrored sets; can operate as long as drives on both mirror sets do not fail	2 TB
RAID 0+1	X	X		Mirror set composed of two or more stripe sets; low level of scalability	2 TB
RAID 50	X		X	Striping data across multiple RAID 5 sets; can sustain up to 4 drive failures	1.5 TB

Calculating Storage Volume

Scitor Corporation

- **Typically estimate storage costs in one of two ways:**
  - Include cost of BOM
    - Other individuals determine hardware needs and obtain vendor quotes
  - Estimate storage based on the amount of data to be received, which requires consideration of
    - Downlink limitations
    - Amount of data compression that will occur
    - Removal of data that is not usable
    - Products, reports, and metadata that must be stored (in addition to the original data)
    - Storage policies (i.e., requirements for duration of storage)
    - Chosen RAID level
    - Standard 15-20% additional open storage recommended to ensure the system does not slow down

Notional Example

1 TB/day received from source → **.5 TB/day** after removing unusable data → **.25 TB/day** after compression → **.75 TB/day** with products and metadata → **1.5 TB/day** For RAID level 6 → **1.8 TB/day** with 20% open storage

Scitor Corporation

Storage Deflation

Storage Deflation

Scitor Corporation

- **As technology has advanced over time, the cost of storage has decreased**
  - Change largely driven by decreasing cost of disk drives
  - While other types of HW have also evolved, they have not demonstrated the same consistent decrease in cost
    - Ex: Capability of servers increase while price stays about the same
- **Groups estimate the changing cost of storage differently**
  - Some estimate a consistent annual decrease in storage costs (X% each year), leading to a lower total cost of storage over time
  - Others have indicated that while storage deflation does occur, a group may just purchase additional storage to compensate for the cost of deflation (results in no cost change over time, but expanded storage capability)

Deflation Estimating Challenges

Scitor Corporation

- **Estimating using projected volume and a $/TB that includes the cost of all peripheral HW/SW may be misleading**
  - While storage costs deflate, other associated COTS HW/SW costs may not
    - Need a full breakout of COTS purchases to apply deflation to only storage
  - Must ensure that the $/TB used is applicable
    - Avoid double counting or underestimating other COTS HW/SW products
  - Using a $/TB that includes multiple types of HW makes capturing unique recap costs difficult (e.g., recap the physical rack every 15 years, but replace COTS SW every 3 years)
- **Deflation has occurred historically, but previous research does not indicate when/if deflation might slow or cease entirely (i.e., does a floor exist?)***
  - Storage may have already deflated so much that it is only a small portion of the total $/TB
  - Cost of materials and production may limit how low the cost of storage can become
- **Storage deflation does not take into account the possibility of additional technology advances**
  - Deflation estimates may not cover a program's transition to a new, more expensive type of storage
  - If a type of storage becomes obsolete, the cost of that storage may actually begin to increase
- **Existing burdens (e.g., SEITPM, maintenance) do not take deflation into account**
  - Unlikely that it will become less expensive to manage and maintain more complex HW
  - Burdens may be correlated more with volume of storage rather than cost

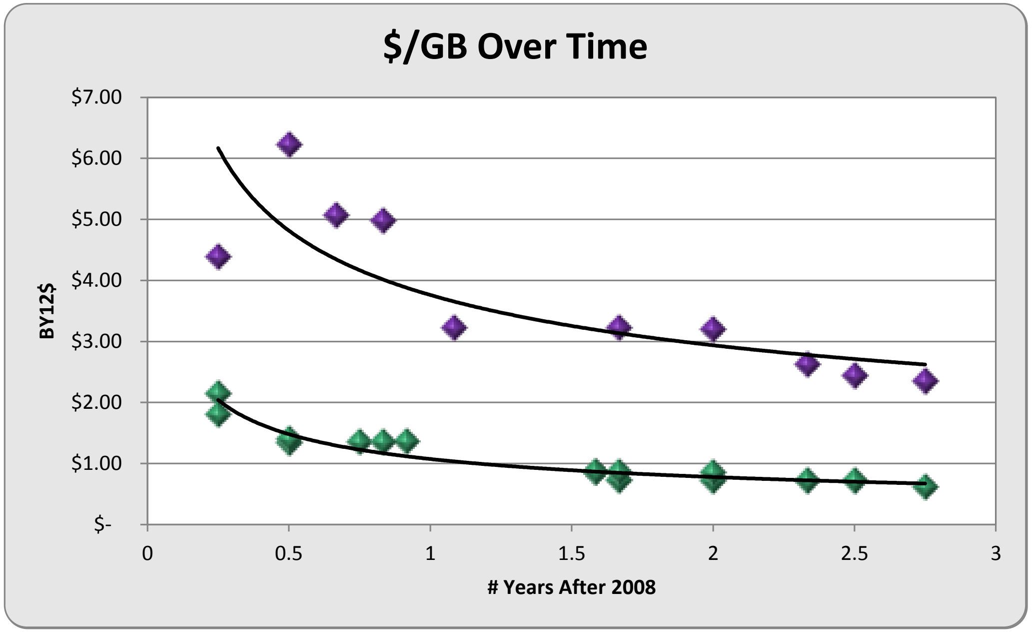
* Earlier Research: "Cost Deflation vs. Technology Inflation of RAID Storage Systems" Converse, Watkins, SCEA, 2006

Data Collection

Scitor Corporation

- **General research**
  - Confirms trend of decreasing cost per TB, but unclear on future impacts
    - How long will cost continue to decrease?
    - How will external factors impact cost?
      - State of economy
      - Transition to Infrastructure as a Service approach
      - Recovery from 2011 Asian tsunami
- **Data collected from available BOMs**
  - Searched for commonality within a single BOM and between different BOMs
  - Looked for procurement of the same piece of equipment in multiple years
    - Avoided HW with vague descriptions because a piece of HW with the same general name can have multiple configurations
    - Ensured apples-to-apples comparison
      - Did not compare prices that include maintenance with ones that did not
      - Evaluated individual pieces of HW, rather than aggregate $/TB (based on level of information available)
  - Unable to determine whether costs were influenced by purchasing agreements or enterprise licenses

Combined Data View

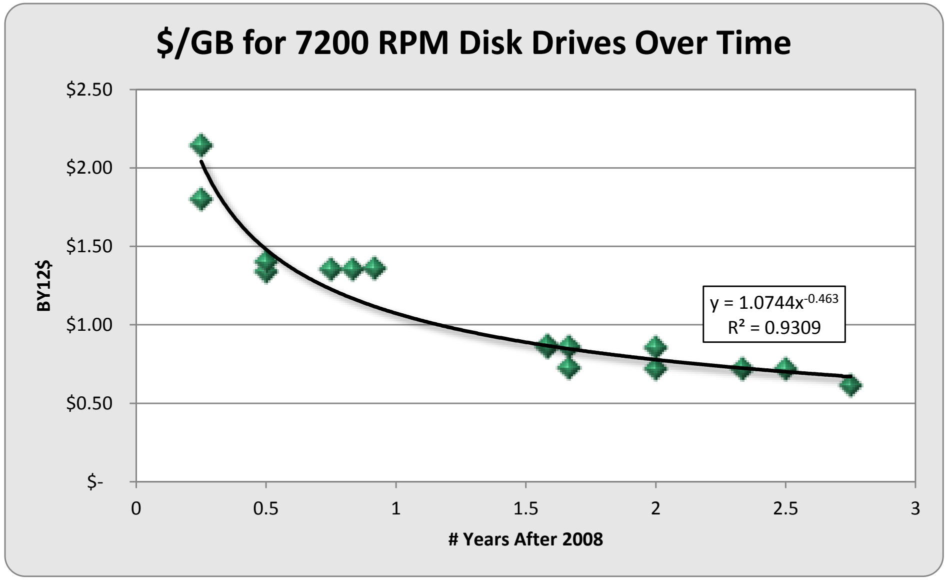


- **All available data points shown**
  - Data available from 2008 to 2011
  - Color indicates different disk drive Revolutions Per Minute (RPMs: **green** = 7200 RPM, **purple** = 15K RPM)
- **Graph indicates that disk drive RPM is a determining factor for cost and rate of deflation**

7200 RPM Disk Drive Data

Scitor Corporation

- **RPM: Revolutions per minute**
  - The faster the disk spins, the faster the drive operates
- **Data used**
  - Equipment: Disk Drive, 1 TB or 500 GB, 7200 RPM, SATA
  - 16 data points available
- **Strong R^2 for data set**
  - Line of best fit indicates that annual change in cost is not as consistent as other research has indicated



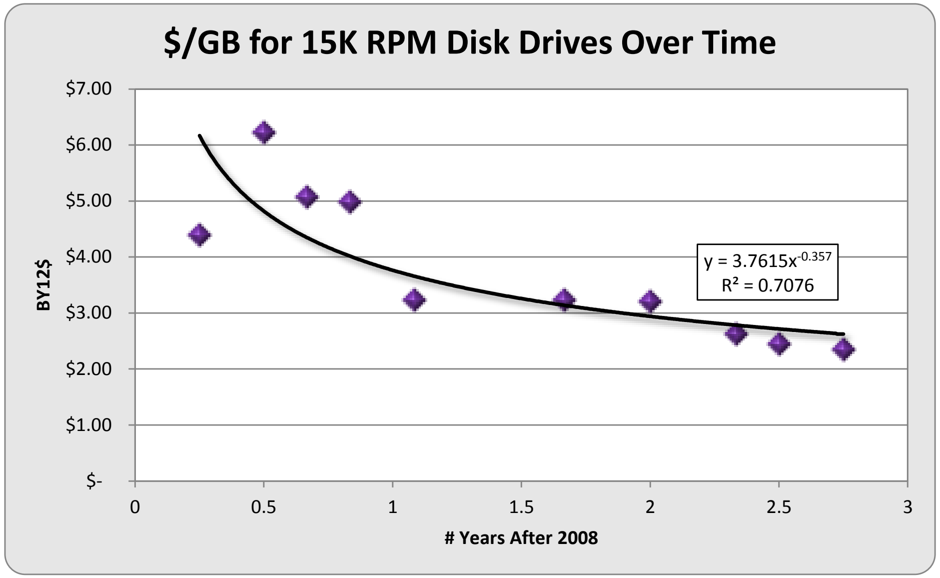
Trendline Predictions - 7200 RPM

FY	BY12$/GB	% Change
2009	$1.07	
2010	$0.78	27%
2011	$0.65	17%
2012	$0.57	12%
2013	$0.51	10%
2014	$0.47	8%
2015	$0.44	7%
2016	$0.41	6%
2017	$0.39	5%
2018	$0.37	5%
2019	$0.35	4%
2020	$0.34	4%
2021	$0.33	4%
2022	$0.32	3%
2023	$0.31	3%
2024	$0.30	3%
2025	$0.29	3%
2026	$0.28	3%
2027	$0.27	2%
2028	$0.27	2%
2029	$0.26	2%
2030	$0.26	2%

15K RPM Disk Drive Data

Scitor Corporation

- **Higher RPM = higher performance disks**
- **Data used**
  - Equipment: Disk Drive, 300 or 144 GB, 15K RPM, 4 GB, FC
  - 10 data points available
- **R^2 not as strong as for 7200 RPM disk drives**
  - Limited number of available data points
- **Cost per TB higher than 7200 RPM, but deflation occurs more slowly**



Trendline Predictions - 15K RPM

FY	BY12$/GB	% Change
2009	$3.76	
2010	$2.94	22%
2011	$2.54	13%
2012	$2.29	10%
2013	$2.12	8%
2014	$1.98	6%
2015	$1.88	5%
2016	$1.79	5%
2017	$1.72	4%
2018	$1.65	4%
2019	$1.60	3%
2020	$1.55	3%
2021	$1.51	3%
2022	$1.47	3%
2023	$1.43	2%
2024	$1.40	2%
2025	$1.37	2%
2026	$1.34	2%
2027	$1.31	2%
2028	$1.29	2%
2029	$1.27	2%
2030	$1.25	2%

Disk Drive Summary

Scitor Corporation

Trendline Predictions - 7200 RPM

FY	BY12$/GB	% Change
2009	$1.07	
2010	$0.78	27%
2011	$0.65	17%
2012	$0.57	12%
2013	$0.51	10%
2014	$0.47	8%
2015	$0.44	7%
2016	$0.41	6%
2017	$0.39	5%
2018	$0.37	5%
2019	$0.35	4%
2020	$0.34	4%
2021	$0.33	4%
2022	$0.32	3%
2023	$0.31	3%
2024	$0.30	3%
2025	$0.29	3%
2026	$0.28	3%
2027	$0.27	2%
2028	$0.27	2%
2029	$0.26	2%
2030	$0.26	2%

Cost per TB significantly different

Deflation for 15K RPM consistently lower

7200 RPM disk drives deflate more quickly

Trendline Predictions - 15K RPM

FY	BY12$/GB	% Change
2009	$3.76	
2010	$2.94	22%
2011	$2.54	13%
2012	$2.29	10%
2013	$2.12	8%
2014	$1.98	6%
2015	$1.88	5%
2016	$1.79	5%
2017	$1.72	4%
2018	$1.65	4%
2019	$1.60	3%
2020	$1.55	3%
2021	$1.51	3%
2022	$1.47	3%
2023	$1.43	2%
2024	$1.40	2%
2025	$1.37	2%
2026	$1.34	2%
2027	$1.31	2%
2028	$1.29	2%
2029	$1.27	2%
2030	$1.25	2%

Estimating Applications

Scitor Corporation

- **Method described on previous slides is most accurate when storage is provided as a service or shared between multiple programs**
  - If storage is a service, a program can be charged for a portion of a rack
  - If storage is purchased by programs individually, a program might have to buy an entire rack when only half of a rack is needed
- **To fully estimate storage costs, must capture additional HW needed to support disk drives**
  - Includes chassis, servers, networking equipment, etc.
  - May need to include cost of COTS SW licenses if not captured separately
  - Equipment needs appear to change with relative frequency
    - Ex: Rarely buy the same server to support storage when it is time for a recap

Conclusions

Scitor Corporation

General storage conclusions

- Must consider the chosen storage strategy when estimating storage costs
  - RAID level contributes to the volume of storage required
  - RPMs and other attributes of storage impact total cost and rate of cost deflation
- Receipt of detailed BOMs improves costing insight, especially in conjunction with volume data
  - Provides details on storage costs vs. other HW/SW costs
  - Allows for accurate application of RAID deflation and estimation of recaps

Disk drive deflation conclusions

- Cost of disk drives decreases over time, approaching a floor
  - Suggests use of a single factor to estimate deflation may not be sufficient
  - Could be used as a cross check for programs with more detailed storage costs available
- Results of research conflicts with some previous storage deflation studies
  - Research shows deflation rate decreasing, rather than remaining constant
  - Need more detailed data to reconcile varying results

Next Steps

Scitor Corporation

- **Evaluate impacts of pricing agreements typically offered to the government**
- **Investigate life expectancy of disks at different RPMs**
  - Also consider different usage (e.g., long or short-term storage)
- **Research impacts to factors and CERs**
  - Assess changes over time due to variation in storage costs
- **Consider impacts of potential shifts in storage needs (e.g., solid state drives)**

Back-ups

Contact Information

Scitor Corporation

- **William Black: wblack@scitor.com**
- **Jenny Woolley: jwoolley@scitor.com**

Recognition

Scitor Corporation

- **Acknowledgments**
  - Carrie Gamble
  - Brian Wells