Education

GROWING THE GREEN STORAGE ENVIRONMENT

Miklos Sandorfi, Chief Technology Officer, SEPATON, Inc.

SNIA Legal Notice

- The material contained in this tutorial is copyrighted by the SNIA.
- Member companies and individuals may use this material in presentations and literature under the following conditions:
  - Any slide or slides used must be reproduced without modification
  - The SNIA must be acknowledged as source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of the SNIA Education Committee.

Abstract

Growing the Green Data Center

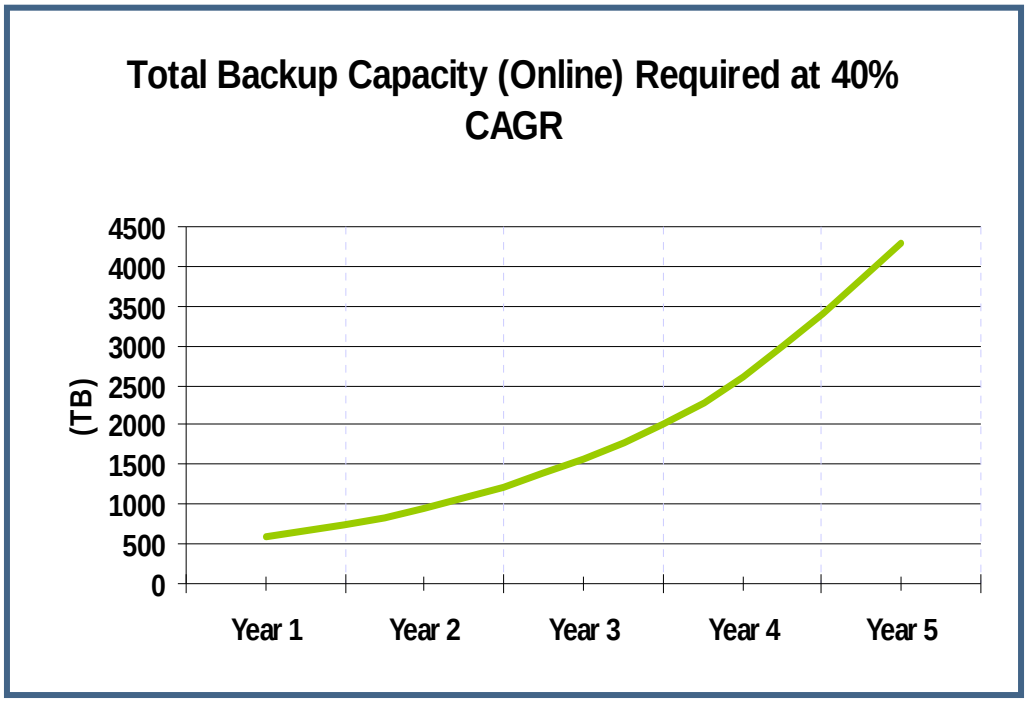
- This tutorial will describe the state of the green data center: why the challenge is very real, what technologies exist today to counter that challenge, and why innovative software that shrinks and streamlines storage is key. It will explain why space-saving technologies like compression, deduplication and thin provisioning are crucial to achieving the green data center. These capabilities not only serve as the foundation for the energy-efficient data center, but also increase storage-related ROI across the board.

The Need for Green

- Data Volumes are Growing Exponentially
- Data Protection Requirements are More Complex
- Risks Associated with Data Loss are Increasing
- Higher Energy Consumption, Higher Energy Cost

Data Volumes are Growing

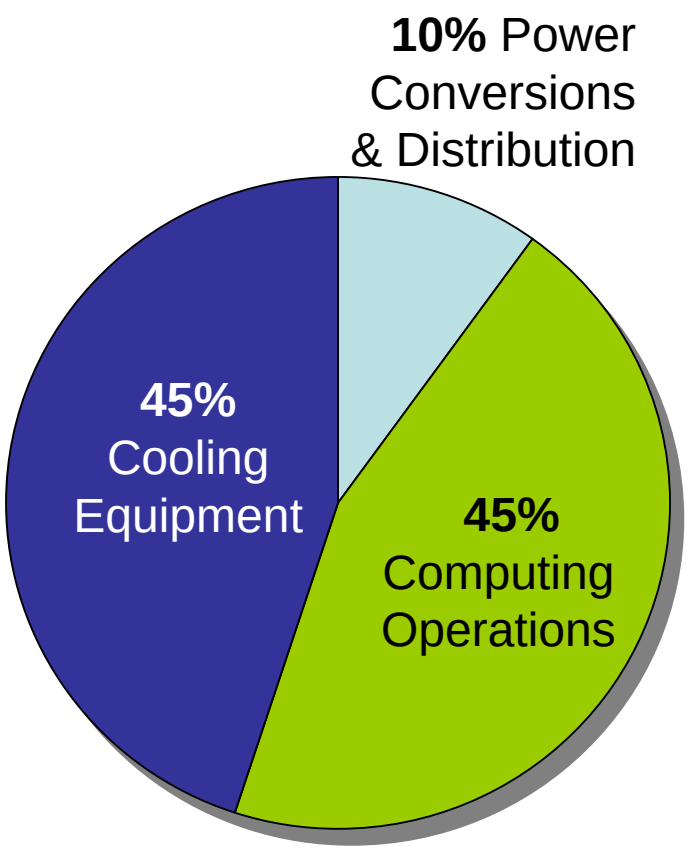
- 40-65% CAGR in Stored Data
- Increased Use of Email and Business Applications
- Higher Risks Associated with Loss
  - Increased dependency on data for productivity
  - Media coverage of data loss



Data Protection Requirements

- Regulatory Compliance Pressures
  - Backup more data
  - Backup more frequently
  - Retain data online longer (eternal retention)
- Higher End-User Expectations, SLAs for Recovery
- Server Virtualization Increasing Storage Requirements
  - Virtual snapshots stored on secondary storage

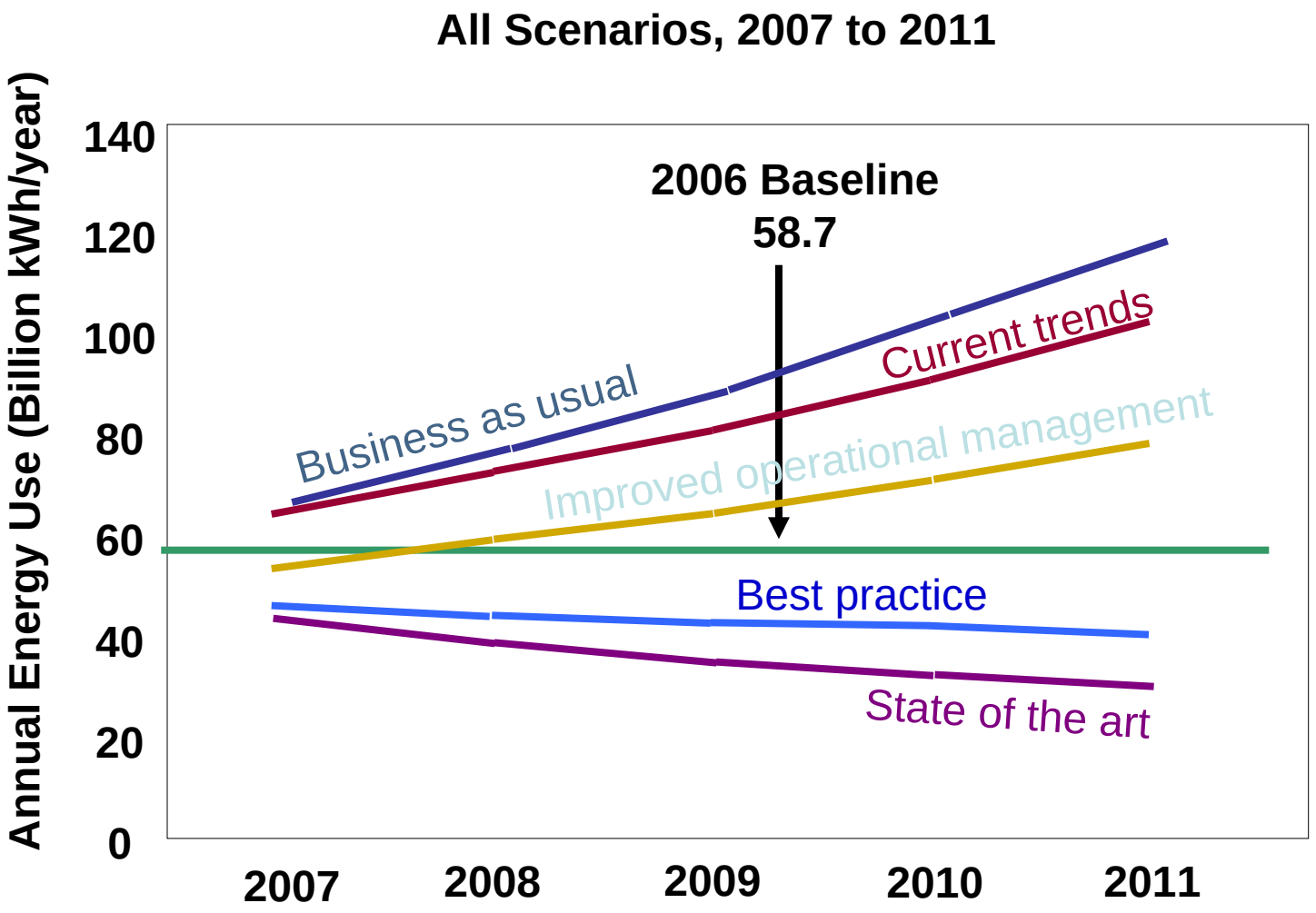
Higher Consumption/Cost of Energy



- Enterprise data centers consume 10 to 100 times more energy per square foot than a typical office building
  - Servers and data centers consumed 1.5 percent of the total US energy consumption (61 billion kWh) in 2006
  - Energy consumption more than doubled from 2000-2006. It will double again by 2011 unless current trends change
- Enterprise data centers are fastest growing segment
  - Among different types of data centers, 30 percent of power consumption is by enterprise-class data centers

Source: EPA Report to Congress on Server and Data Center Energy Efficiency, April 2007 and Creating Energy-Efficient Data Centers, Paul Scheihing, U.S. Department of Energy, Data Center Facilities and Engineering Conference, Washington, DC, May 18, 2007

Comparison of Projected Power Use



Source: Creating Energy-Efficient Data Centers, Paul Scheihing, U.S. Department of Energy, Data Center Facilities and Engineering Conference, Washington, DC, May 18, 2007

Green is Essential to Cost Control

"In 2005, data centers averaged *25 cents for every dollar the entire corporation spent on power and cooling*. In the same year, American corporations alone spent about $6 billion on energy. Global businesses are reporting similar costs."

"An average 25,000 foot data center will be able to save close to 45% in energy costs over today's budget. The converse is also true: in situations where the money is not as much an issue as is data center size, the 45% conservation rate can allow the corporation to double the size of its data center without increasing energy costs."

- Taneja Group, The Greening of the Data Center, August 2007

What is a Green Data Center?

- Not an Environmental Initiative
- Cost Saving Initiative
- Most Value/Utilization From Every Square Foot
  - Maximum Computing Power/Storage/Backup In Smallest Space
- Uses Least Power And Cooling Per Square Foot
- Enables Growth And Flexibility At Minimal Cost
- Uses Systems And Software To Enable Efficiency

Requirements for a Green Data Center

- *People:* Management Alignment for Energy Reduction
- *Process:* Power Consumption Measured, Reduced, Managed
- *Systems:* Optimized Data through Hardware and Software
  - Make optimal use of archiving capabilities in backup/restore software
  - Enterprise data centers are where energy use and footprint are growing fastest

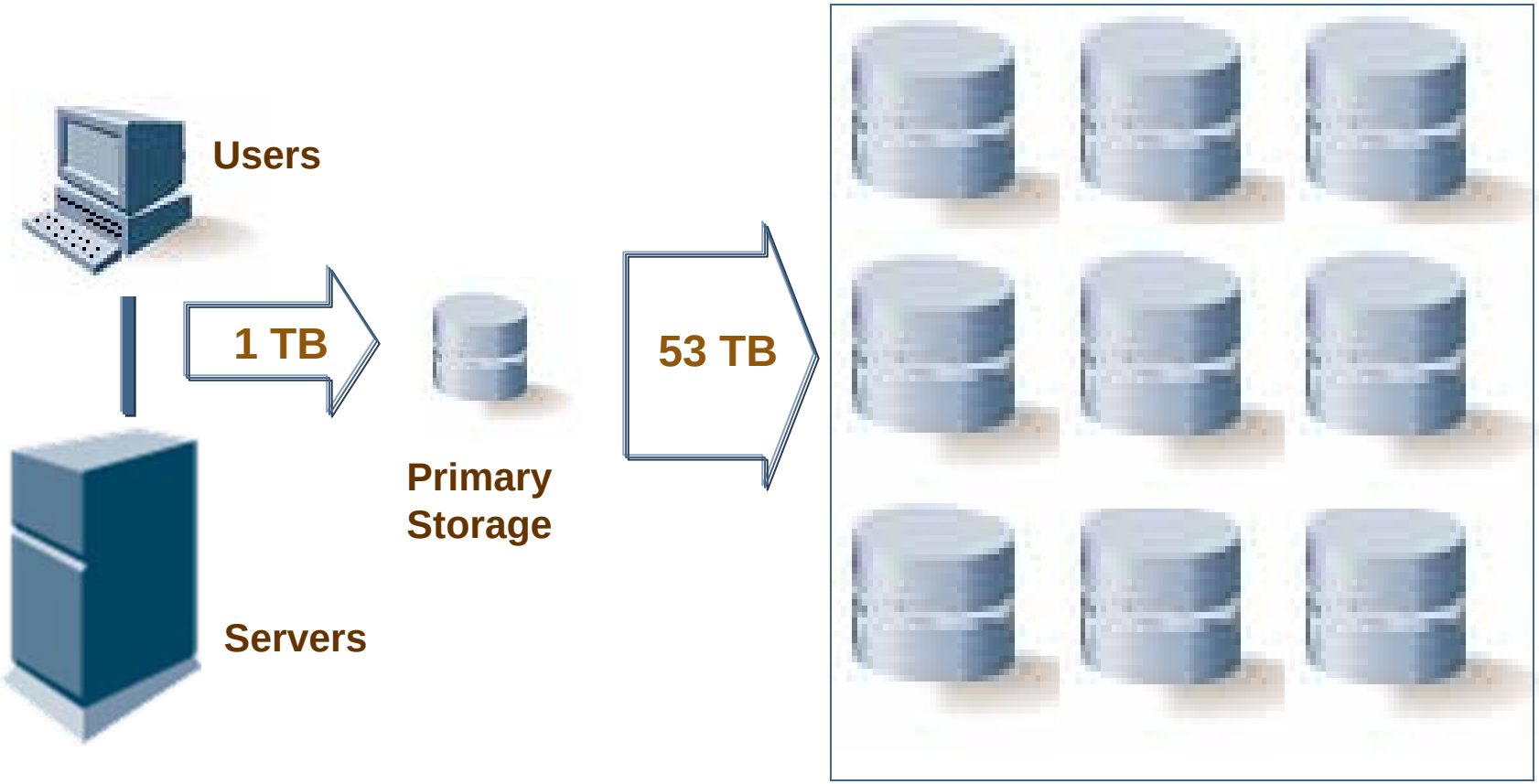
People: Managing for Efficiency

- Companywide Management Alignment
- Clear Metrics
  - Measure Current Usage by Department
  - Set Achievable Goals and Milestones Toward Goals
- Communicate Progress Toward Milestones
- Reward Achievement

Process: Driving Reduction

- Reduce Capacity Through ILM, Tiered Storage
  - Move data to secondary storage efficiently
  - Take advantage of backup software archive capabilities
- Focus on Backup and Storage
  - Greatest impact on capacity/consumption reduction
  - Policies and procedures
  - Technologies

Focus on Backup and Storage



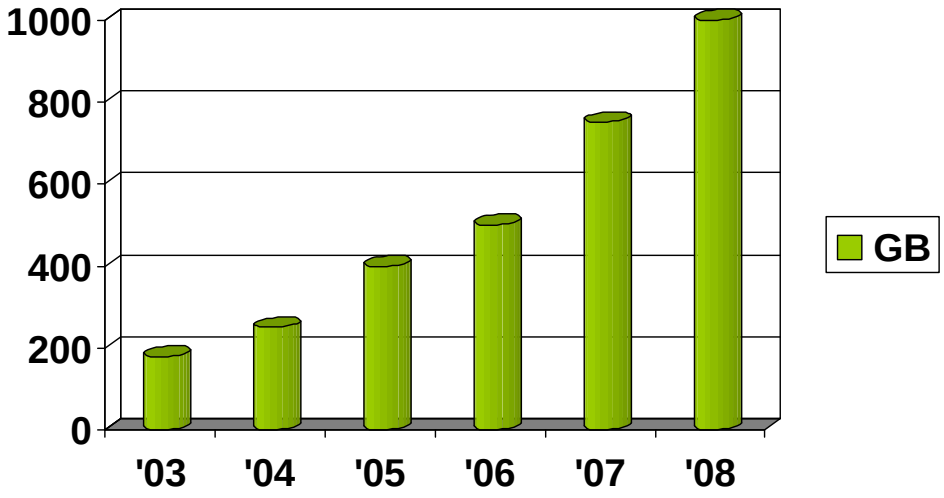
1 TB X 52 Weeks (1 year retention)
+ .1 (10% incremental change) X 10 days (retention)
53 TB

Systems: Hardware and Software

- Efficient Hardware
  - Lower draw, high efficiency systems
  - SATA disks
  - Power-managed arrays
- Optimize Data
  - Thin Provisioning
  - Data compression
  - De-duplication

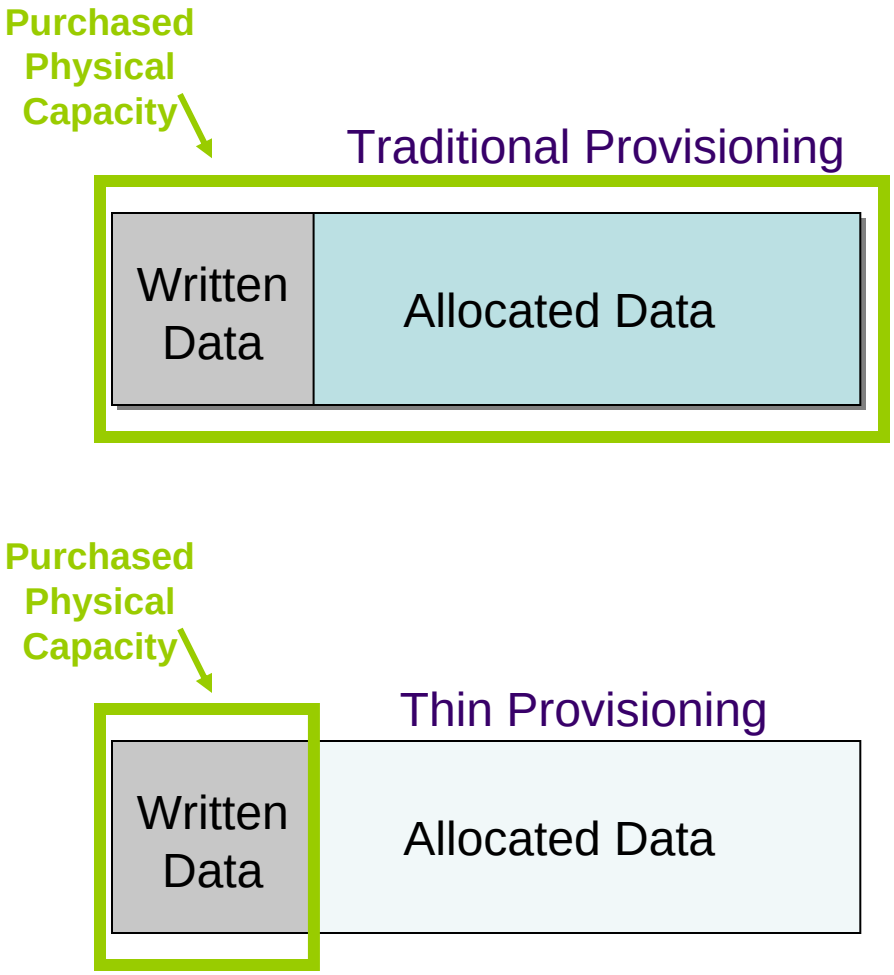
Efficient Hardware

- Load-optimized power supplies
  - Higher power factor conversion rates
- Higher capacity per watt
  - SATA density growth, 2.5” disk drives
- Power Managed Storage
- “On The Spot” Cooling



SATA Disk Capacity Growth, 2003-2008

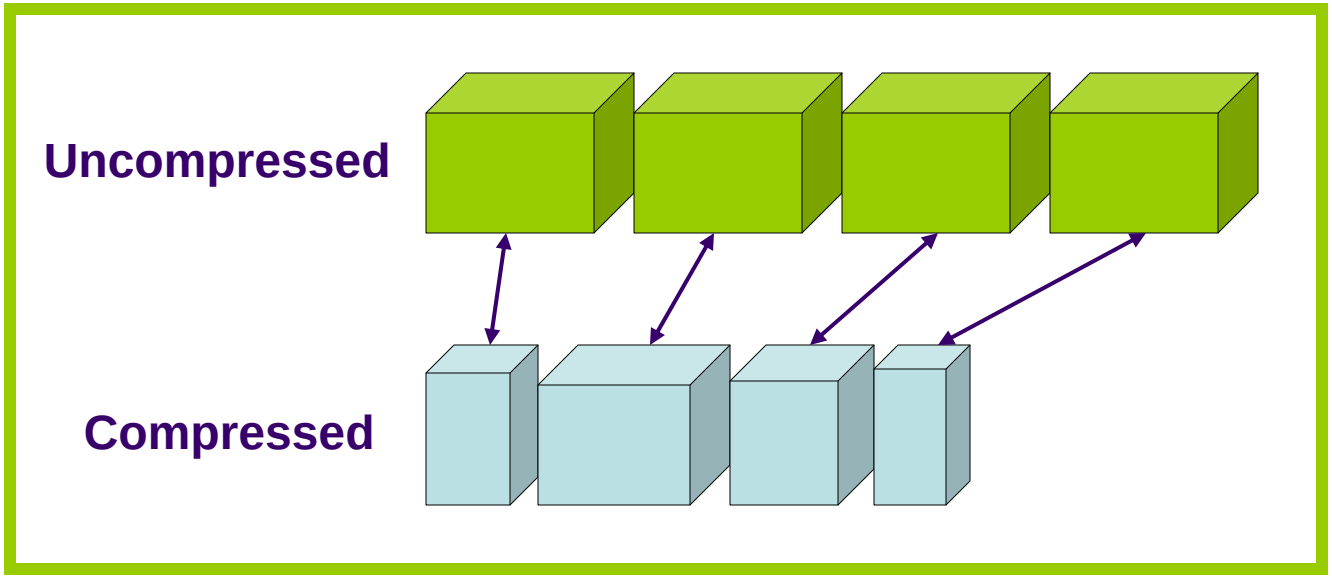
Software: Thin Provisioning



- Virtualized Backup Environments
- Provision Capacity as Required
  - Application requirements
  - Virtualization software requires 2 X capacity allocation
- Actual Capacity as Used
- Eliminates Over-Purchasing, Provisioning
- Enables Streamlined Data Center

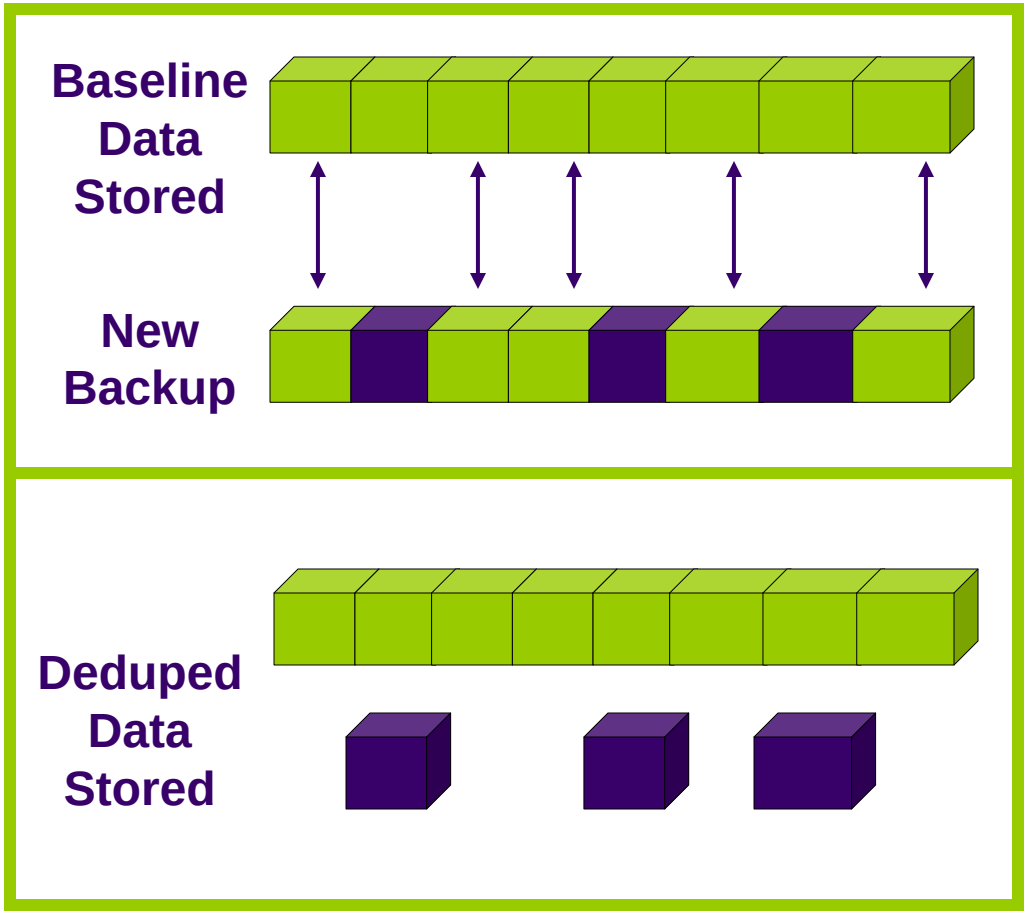
Software: Compression

- Hardware or Software
- Lempel-Ziv Algorithm
- Typical 2:1 Reduction



Software: Deduplication

- Compares blocks of data in backup set to data already stored on the backup device
- Stores a single set of data as a baseline
- Replaces duplicate data with a pointer to the baseline



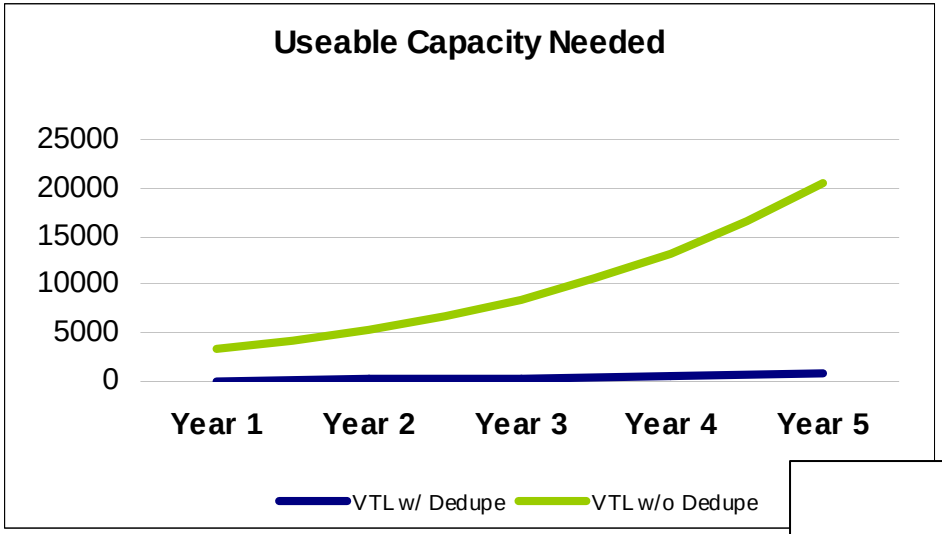
Focus on De-duplication

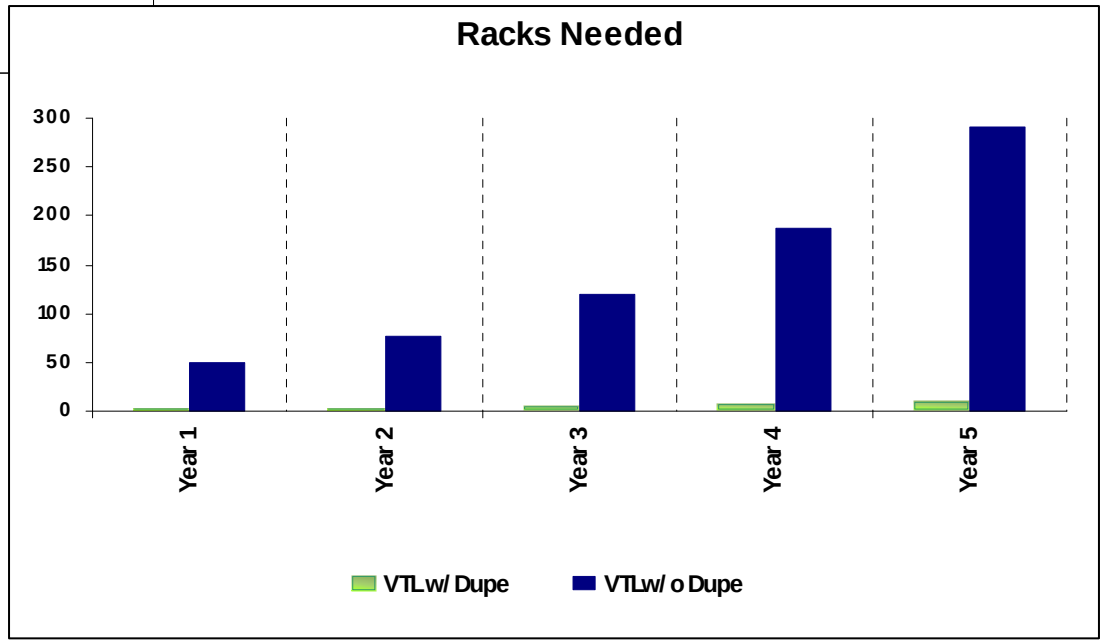
- Optimized For a VTL Environment
- Cut Capacity of Backup by as Much as 25x
  - 50x with Compression
- Different Approaches
  - Source ← → Target
  - Byte-level Comparison ← → Hash Digest
  - In-Line ← → Post-Process

Enterprise Backup Scenario

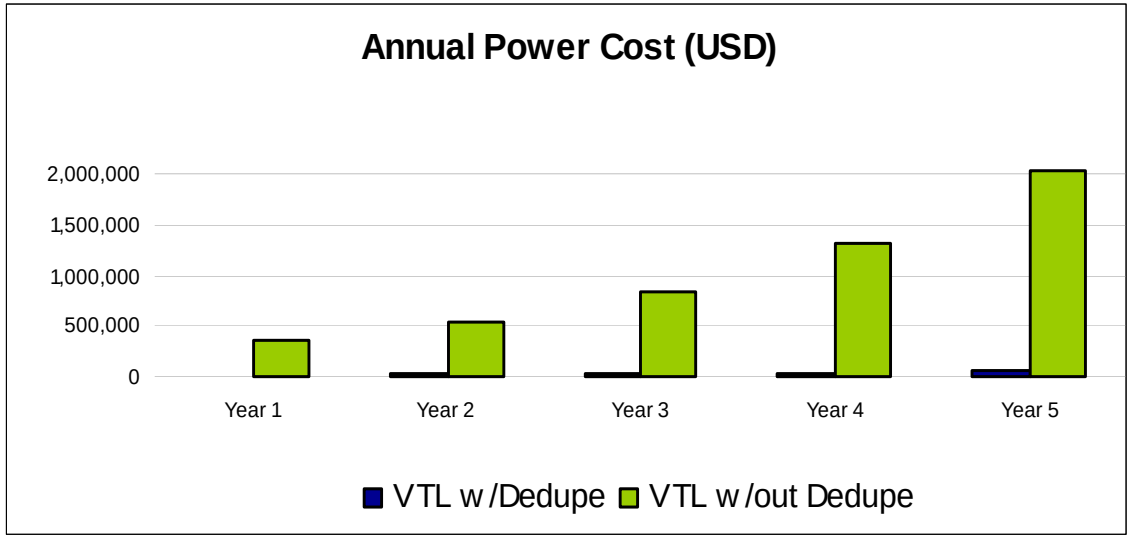
- Eight-hour maximum backup window
- Full weekly backups of 65 TB with one-year online retention period
- Daily incremental backups (10% daily change) with a 10-day online retention period
- Compounded annual growth (CAGR) of 56% in primary data
- Power consumption of 5,000 W per rack
- Average cost of power ($0.0867/kW) and floor space ($35/sft)
- Reduction ratio of 25:1

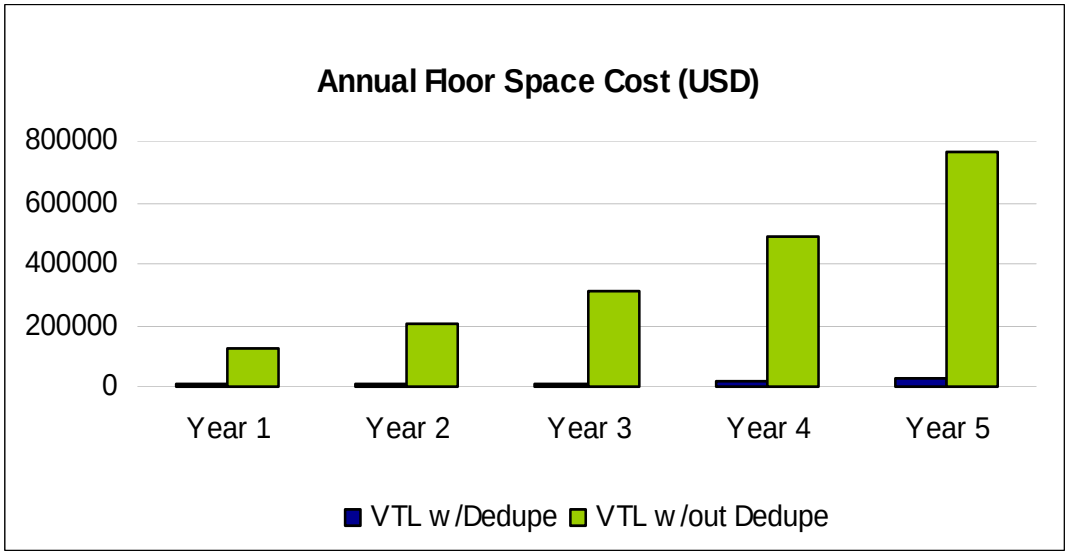
Reducing Floor Space, Power





Reducing Floor Space, Power





Conclusion

- Green Data Center is a Necessity
- Optimize Data in Areas of Greatest Impact
  - Primary and Secondary Storage
- Threefold Approach
  - People, Process, and Technology

Q&A / Feedback

- Please send any questions or comments on this presentation to SNIA: trackgreenstorage@snia.org

Many thanks to the following individuals for their contributions to this tutorial.

- SNIA Education Committee

Arun Taneja, Taneja Group