



# The Business Value of Data Deduplication

## DDSR SIG



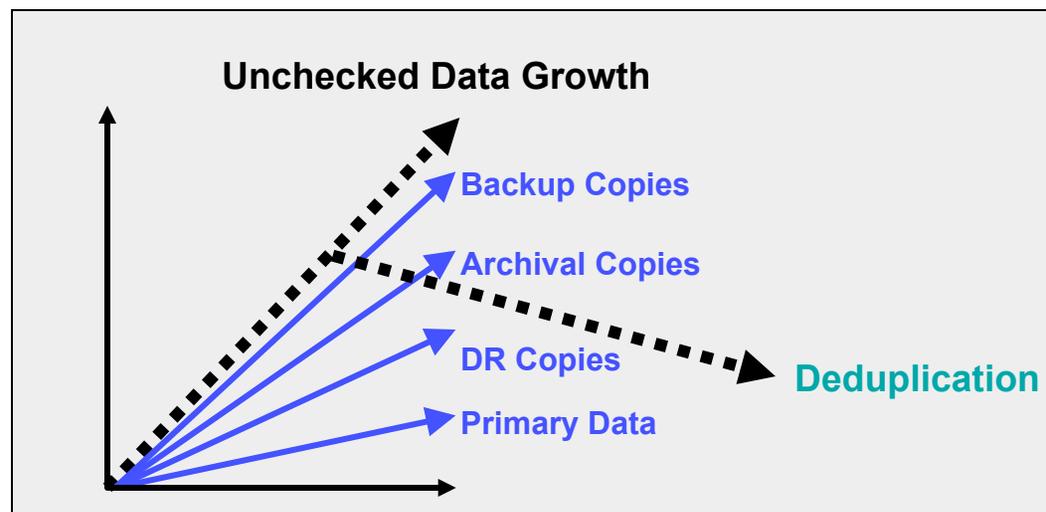
- The purpose of this presentation is to provide a base level of understanding with regards to data deduplication and its business benefits. Consideration will be given to current storage pain points and business goals. This session will focus on the benefits derived from data deduplication including cost savings, storage efficiency, and reduced storage administration. Example use cases will explore considerations related to backup and recovery, replication, archival, primary storage applications, and network bandwidth reduction.

# Agenda

- Storage Pain Points
- Storage Business Goals
- Data Deduplication Explained
- Data Deduplication Benefits
- Data Deduplication Implementations
- Data Deduplication Savings Ratios
- Summary

# Storage Pain Points

- Rising storage costs due to unchecked data growth
- Exponential data growth and shrinking backup windows
  - ◆ While dealing with same staffing levels
- Longer data retention on disk required for SLA's and regulatory requirements
- Limited space, power, cooling, budget
  
- Data deduplication technology can dramatically reduce physical disk capacity requirements while still meeting data storage requirements



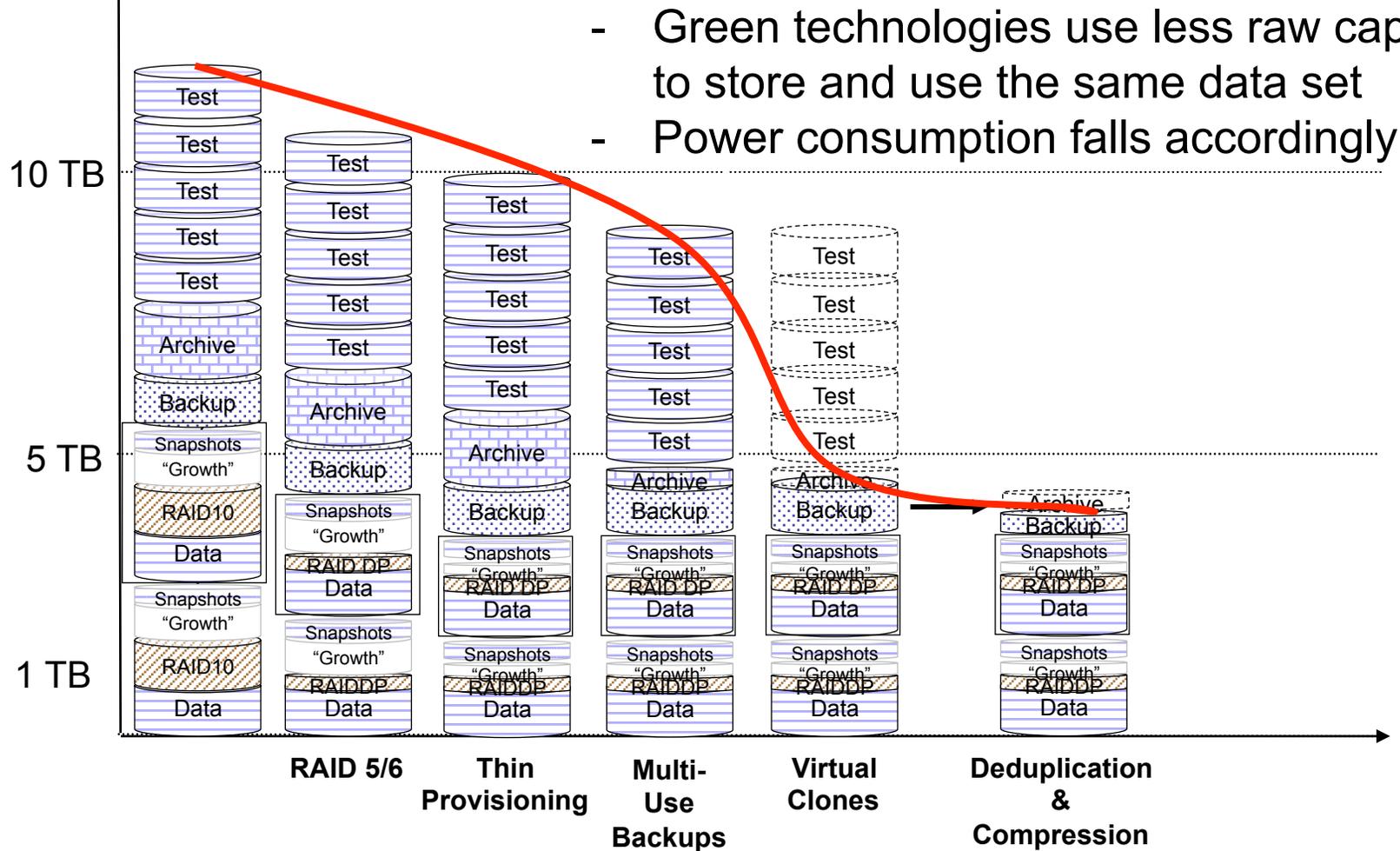
## ➤ Data Deduplication can help organizations:

- ◆ Satisfy ROI/TCO requirements
- ◆ Manage data growth
- ◆ Increase efficiency of storage and backup
- ◆ Reduce overall cost of storage
- ◆ Reduce network bandwidth
- ◆ Reduce operational costs including:
  - › Infrastructure costs required space, power and cooling
  - › Movement toward a greener data center – carbon footprint friendly
- ◆ Reduce administrative costs

# Data Deduplication Effect on Green Technologies



Advancing storage & information technology



Source: SNIA Green Storage Initiative

# Data Deduplication Explained

- Find duplicates at some level, substitute pointers to a single shared copy
- Block or **sub-file** based
- **Not to be confused with SIS**, which is content or name based
- Streaming (in-band) and offline techniques
- Savings increase with number of copies found
  - ◆ Up to 95% savings, depending on dataset

# Data Deduplication Savings Calculation



Advancing storage & information technology

## ➤ Every TB of disks you don't buy saves you

- ◆ CAPEX for the equipment
- ◆ CAPEX for the footprint
- ◆ CAPEX and OPEX for power conditioning
- ◆ OPEX for the power to spin the drives
- ◆ OPEX for cooling
- ◆ OPEX for storage management

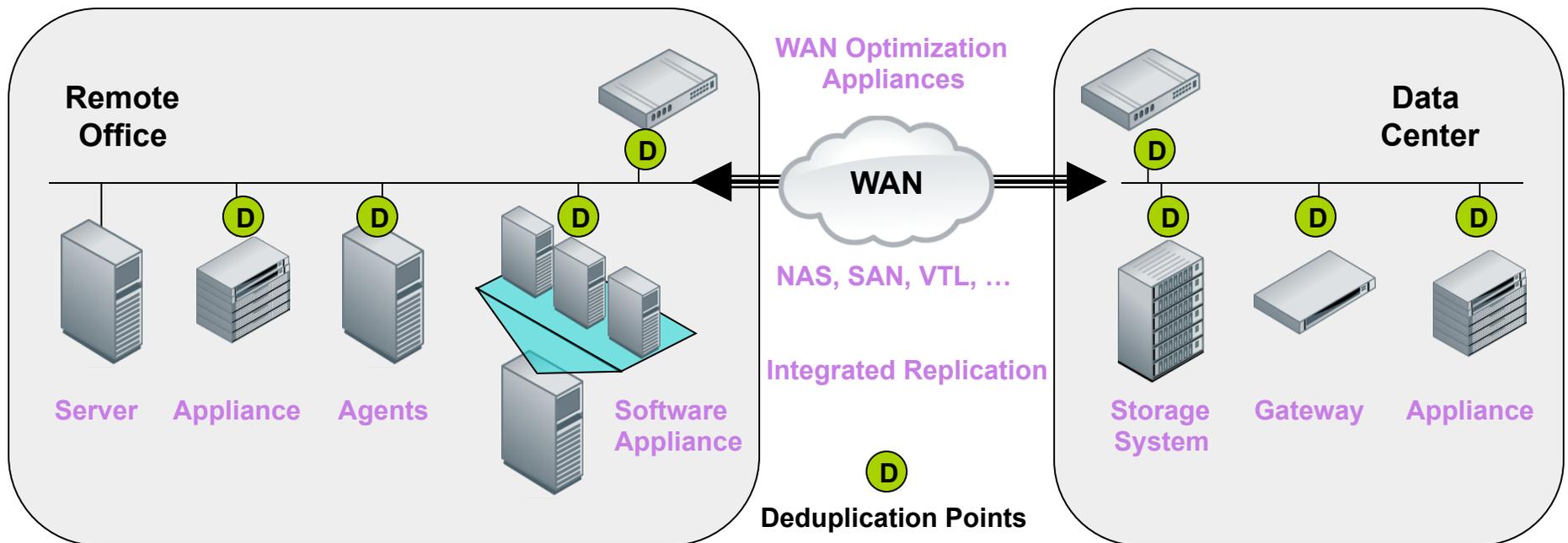
# Data Deduplication Benefits

- Reduced floor space requirements
- Reduced power & cooling costs
- Reduced capital costs
- Reduced operational costs
- Longer retention of data on disk

Space Reduction Ratio	Space Reduction Percentage = $1 - (1 / \text{Space Reduction Ratio})$
2:1	$1/2 = 50\%$
5:1	$4/5 = 80\%$
10:1	$9/10 = 90\%$
20:1	$19/20 = 95\%$
100:1	$99/100 = 99\%$
500:1	$499/500 = 99.8\%$

# Deduplication Implementations

- Vendors today provide deduplication solutions for nearly every point at which data is stored or transmitted
- The decision as to “where” to deduplicate is determined only by which problem you are trying to solve



## ➤ Storage System

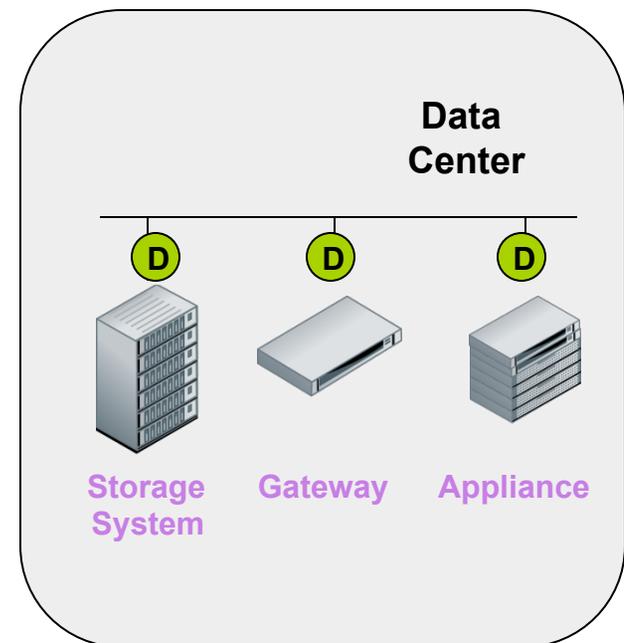
- ◆ General purpose storage system that is capable of deduplicating data

## ➤ Gateway

- ◆ Hardware capable of deduplicating data on external storage hardware

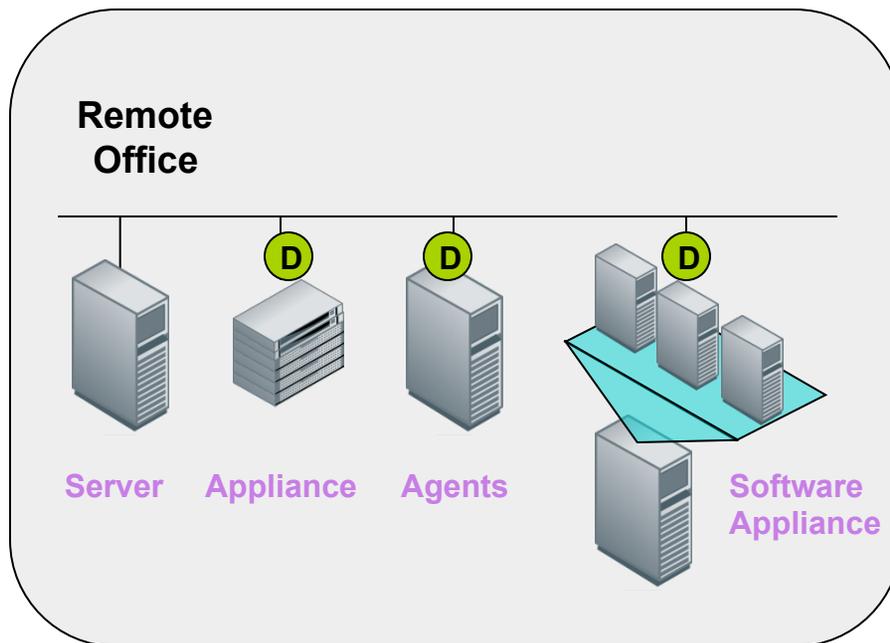
## ➤ Appliance

- ◆ Hardware capable of deduplicating and storing data in an appliance dedicated for this purpose



Refer to DPI Buyer's Guide For Vendor Descriptions

# Example Design Approach – Remote Office



## ➤ Appliance

- ◆ Hardware capable of deduplicating and storing data in an appliance dedicated for this purpose

## ➤ Agents

- ◆ Client-side software capable of deduplicating data for a specific application

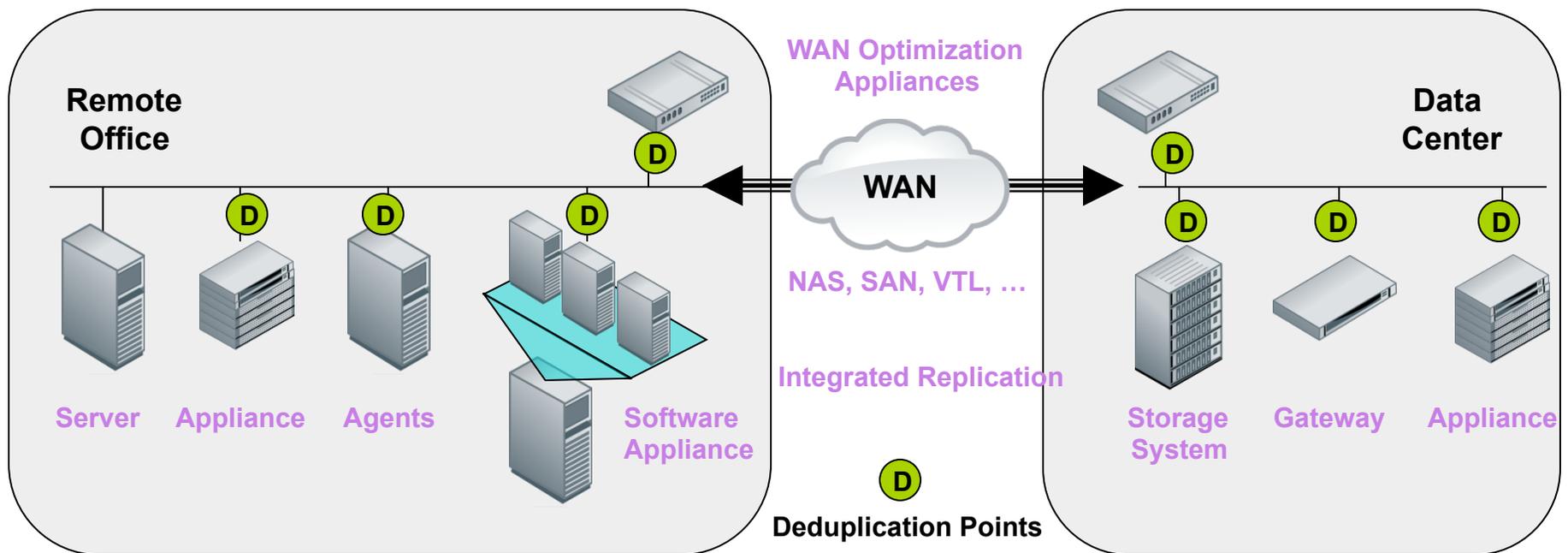
## ➤ Software Appliance

- ◆ Server-side software capable of storing deduplicated data from application agents

Refer to DPI Buyer's Guide For Vendor Descriptions

## ➤ WAN Optimization

- Many deduplication solutions are also capable of transmitting deduplicated data from point to point, resulting in reduced network traffic and “inherited” deduplication at the destination



Refer to DPI Buyer's Guide For Vendor Descriptions

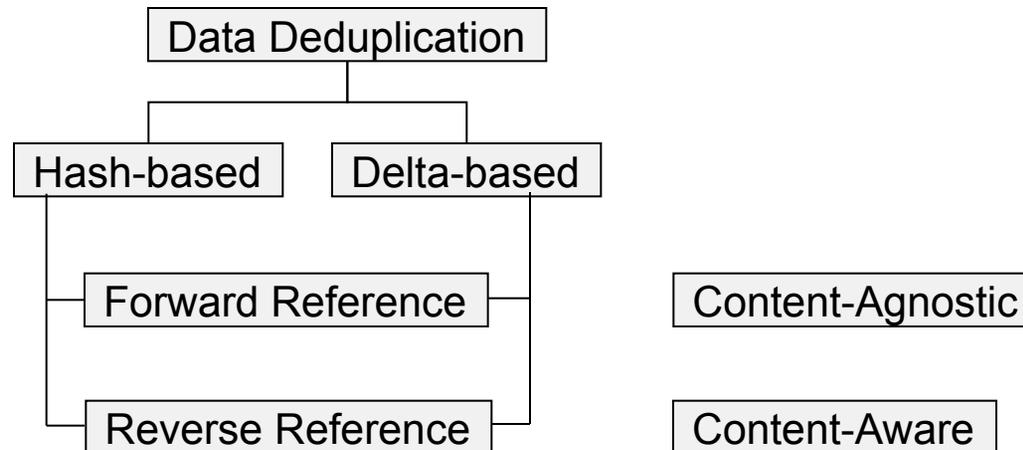
# Deduplication Savings Ratios

Savings ratios reflect design tradeoffs involving performance and space reduction

<b>Factors associated with higher data deduplication ratios</b>	<b>Factors associated with lower data deduplication ratios</b>
Data created by users	Data captured from mother nature
Low change rates	High change rates
Reference data and inactive data	Active data
Applications with lower data transfer rates	Applications with higher data transfer rates
Use of full backups	Use of incremental backups
Longer retention of deduplicated data	Shorter retention of deduplicated data
Wider scope of data deduplication	Narrower scope of data deduplication
Continuous business process improvement	Business as usual operational procedures
Smaller segment size	Larger segment size
Variable-length segment size	Fixed-length segment size
Content-aware	Content-agnostic
Temporal data deduplication	Spatial data deduplication

# Deduplication Considerations

## How



## Where

Source

Target

## When

Inline

Post processing

## What

Spatial

Temporal

Global

Local

## Resiliency

RAID

RAIN

## Benefits

TCO

Power

Cooling

Footprint

Labor

Retention

## ◆ Data Deduplication Technical Considerations

- ◆ *This technical session will address the nuances of data deduplication and the various design approaches available today. Space savings technologies including data deduplication are used to dramatically improve storage efficiency. The session will address the question of what data deduplication is, how it is performed, and the architectural choices available today. It will address various design approaches including fixed length and variable length segmentation, inline and post processing, source and target, availability and integrity of deduplicated data, and complementary use of removable media. It will also explore the factors affecting space reduction ratios relative to specific capacity optimization techniques.*



Check out SNIA Tutorial:

**“Data Deduplication Technical Considerations”**

- ◆ Data Deduplication reduces physical data storage requirements by removing duplicate data
- ◆ Data Deduplication also allows data to be retained on disk for longer periods, using less physical space
- ◆ Data Deduplication involves tradeoffs relating to degree of physical space reduction and performance considerations
- ◆ Thoughtful deployment of deduplication allows users to benefit in many ways, including overall data center “greening”

- Please send any questions or comments on this presentation to [ddsr-sig@snia.org](mailto:ddsr-sig@snia.org)

**Many thanks to the following individuals  
for their contributions to this presentation**

**Data Deduplication and Space Reduction SIG**

**Matt Brisse  
Mike Dutch  
Larry Freeman  
Devin Hamilton**

**Shane Jackson  
Gideon Senderov  
Tom Sas**

# Call To Action Join the SNIA DDSR SIG !!



## ➤ Join the DDSR SIG!

- ◆ Make an impact in helping solve today's data storage challenges
- ◆ Ensure your voice is heard as the industry reaches consensus on space reduction terminology and best practices
- ◆ Help define promote deduplication and space reduction benefits industry-wide
- ◆ For more information and membership, contact the DDSR chair: at [ddsr-sig-chair@snia.org](mailto:ddsr-sig-chair@snia.org)
- ◆ Visit the DDSR team at:
  - > If a SNIA member : <https://www.snia.org/apps/org/workgroup/ddsr-sig>
  - > If not a SNIA member: [http://www.snia.org/member\\_com/join/](http://www.snia.org/member_com/join/)