Education

Deduplication's Role in Disaster Recovery

Gene Nagle, EXAR

Thomas Rivera, SEPATON

SNIA Legal Notice

- The material contained in this tutorial is copyrighted by the SNIA.
- Member companies and individual members may use this material in presentations and literature under the following conditions:
  - Any slide or slides used must be reproduced in their entirety without modification
  - The SNIA must be acknowledged as the 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.
- Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.
- The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

  NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.

About the SNIA DPCO

- This tutorial has been developed, reviewed and approved by members of the Data Protection and Capacity Optimization (DPCO) Committee, a group of more than 70 people representing 40 SNIA members
- The mission of the DPCO is to foster the growth and success of the market for data protection and capacity optimization technologies
- 2011 goals include educating the vendor and user communities, market outreach, and advocacy and support of any technical work associated with data protection and capacity optimization

Check out these SNIA Tutorials:

- **Understanding Data Deduplication**
- **Advanced Deduplication Concepts**

Abstract

Data deduplication can enhance Disaster Recovery (DR) because deduplication significantly reduces the amount of bandwidth required to replicate data.

This technical session will address the question of how deduplication fits into DR strategies, and what the various architectural choices available today are - for implementation.

This technical session will:

- Review data deduplication & Disaster Recovery
- Address the architectural choices for implementation
- Cover the impact of deduplication on WAN replication
- Discuss deduplication effects on meeting SLAs for DR

Terminology from SNIA Dictionary

Data Deduplication is the replacement of multiple copies of data - at variable levels of granularity - with references to a shared copy in order to save storage space and/or bandwidth

Data Replication is continuously maintaining a secondary copy of data – possibly at a remote site – from a primary volume for the purposes of providing high availability and redundancy.

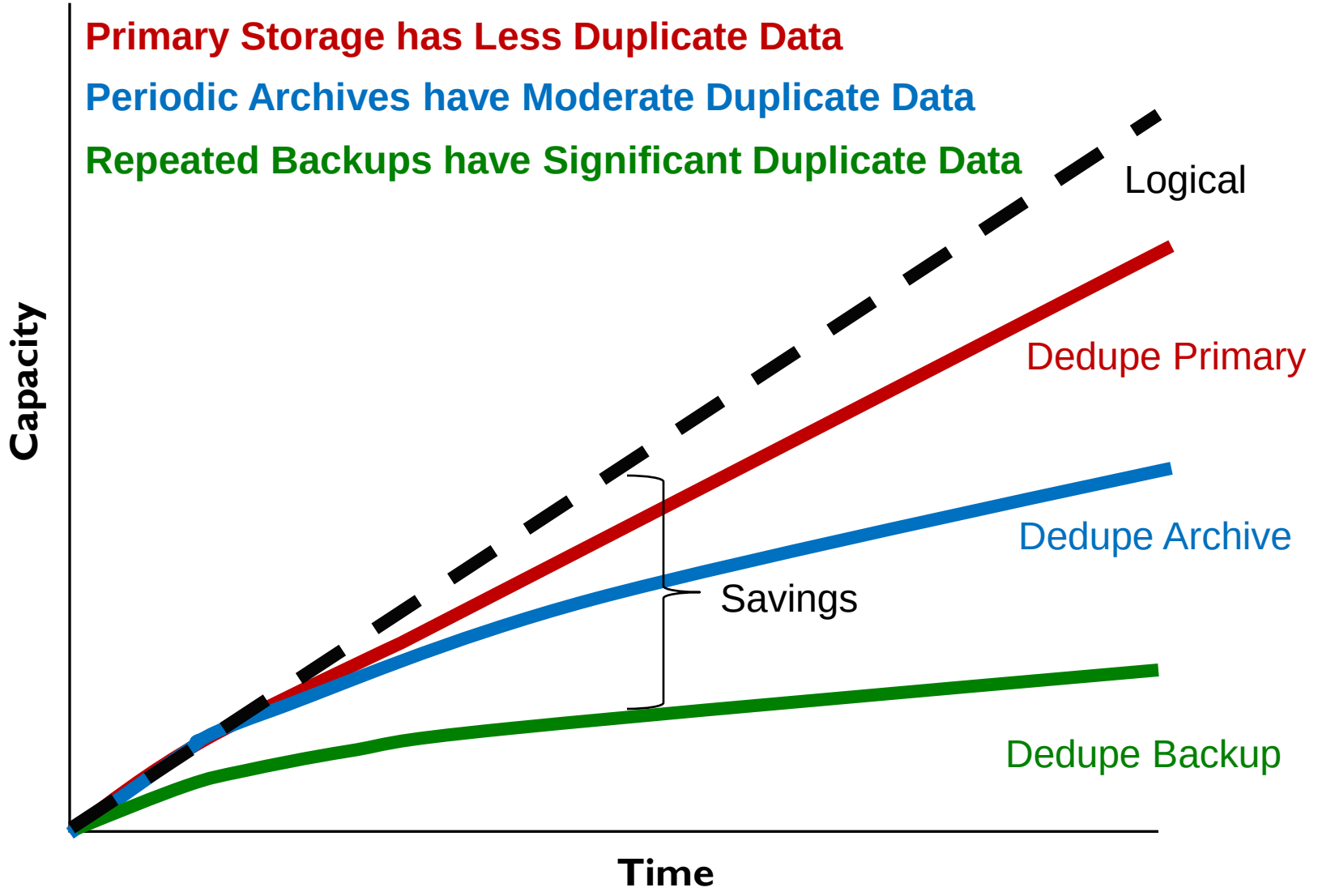
Disaster Recovery is the recovery of data, access to data and associated processing through a comprehensive process of setting up a redundant site (equipment and work space) with recovery of operational data to continue business operations after a loss of use of all or part of a data center.

Check out SNIA Tutorial:

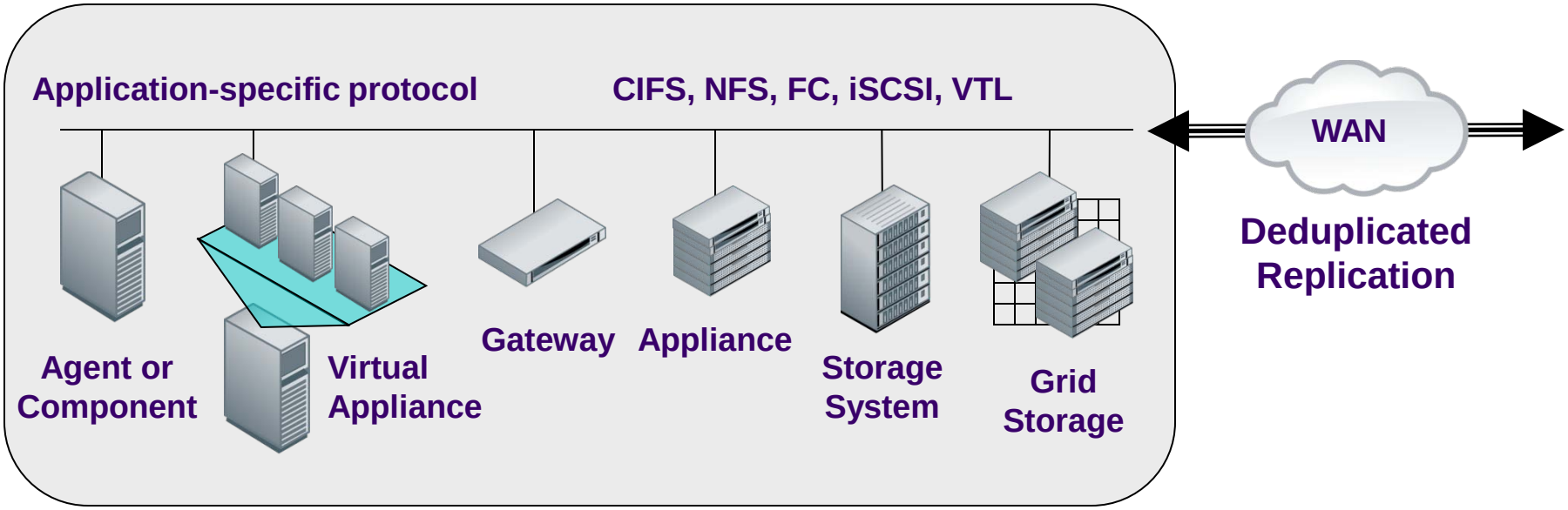
Understanding Data Deduplication

Deduplication Savings: Depends on Use Case and Time

Education SNIA



Deduplication Implementations



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

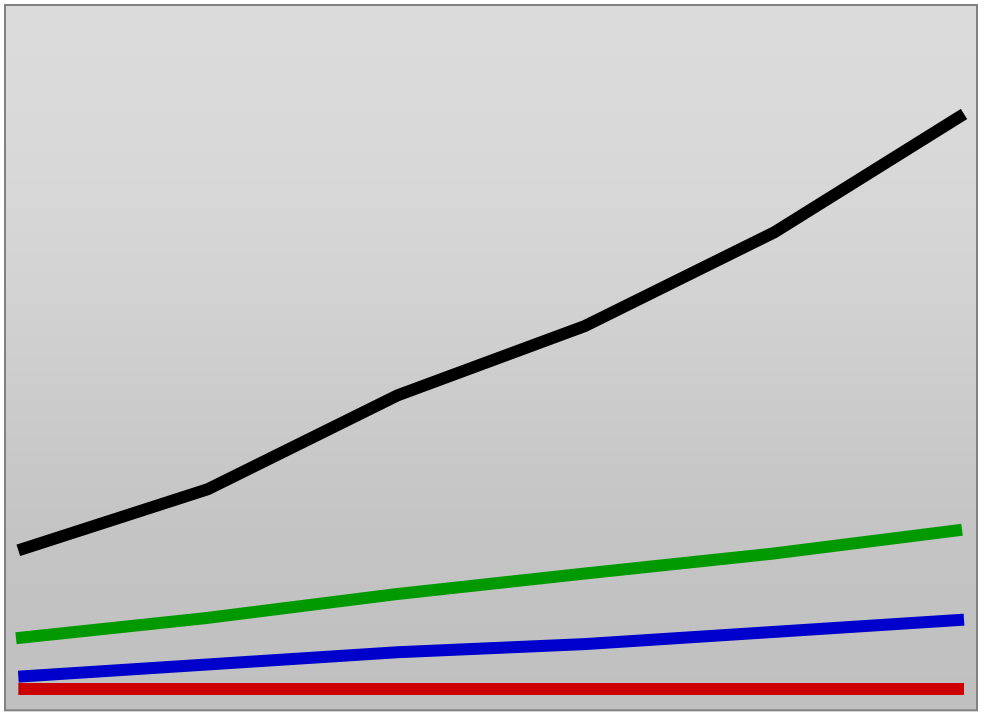
Data Deduplication Benefits

- Data Deduplication can help organizations:
  - Help satisfy ROI/TCO requirements
  - Manage data growth costs
  - Increase efficiency of storage and backup
  - Reduce overall expenditure on storage
  - Reduce network bandwidth
  - Reduce operational costs including:
    - Infrastructure costs for space, power and cooling
  - Reduce administrative costs
  - More easily implement and manage Disaster Recovery (DR)

Challenges of Implementing DR

- Data volumes too large for timely replication
- Bandwidth constraints / costs
- Added complexity and cost
  - Cost $$ for admin, HW/SW,
- Backup may delay or prevent DR
- Satisfying RPO/RTO metrics
- Efficient Replication
  - Without it SLAs for Disaster Recovery may not be met

Challenge$ of Implementing DR



IT Challenges

- Reworking inflexible storage architectures
- Increased management burden, labor costs
- Potential of doubling storage capex & opex (power & cooling)
- Outsourcing can save $$, but - -
- Must find the right cloud storage provider

Gartner (2009)

Meeting Service Level Agreements

Education SNIA

Business Issues & SLAs

Rapid Data Growth

- ✓50% CAGR
- ✓Increased backup costs

SLAs for BC/DR

- ✓Downtime costs
- ✓RTO / RPO

Space/Power Limitations

- ✓Data center footprint
- ✓Power costs

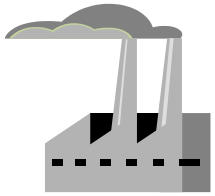
Regulatory Requirements

- ✓Online retention
- ✓Added complexity

Achieving Lower TCO & Better ROI

Education SNIA

Measurable TCO & ROI



Reduce Capital Expense

- ✓Lower acquisition cost
- ✓Scalability
- ✓Compare in-house vs. out-source

Reduce Operating Expense

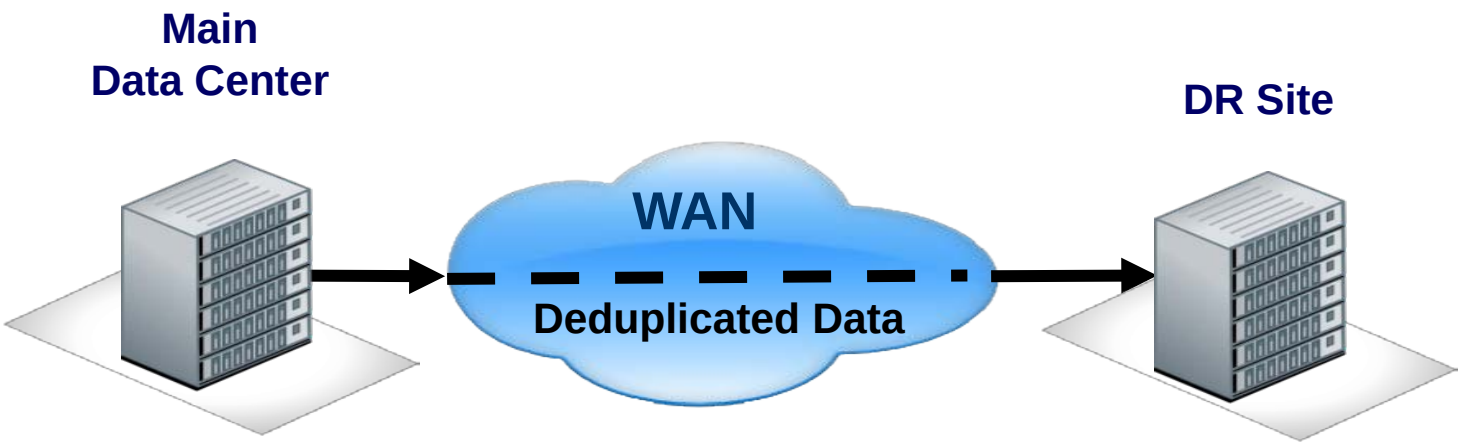
- ✓Less labor: automation
- ✓Less power & space
- ✓Non-Disruptive changes

Avoid Costs

- ✓Free IT staff time
- ✓More data per FTE
- ✓Reduced human error

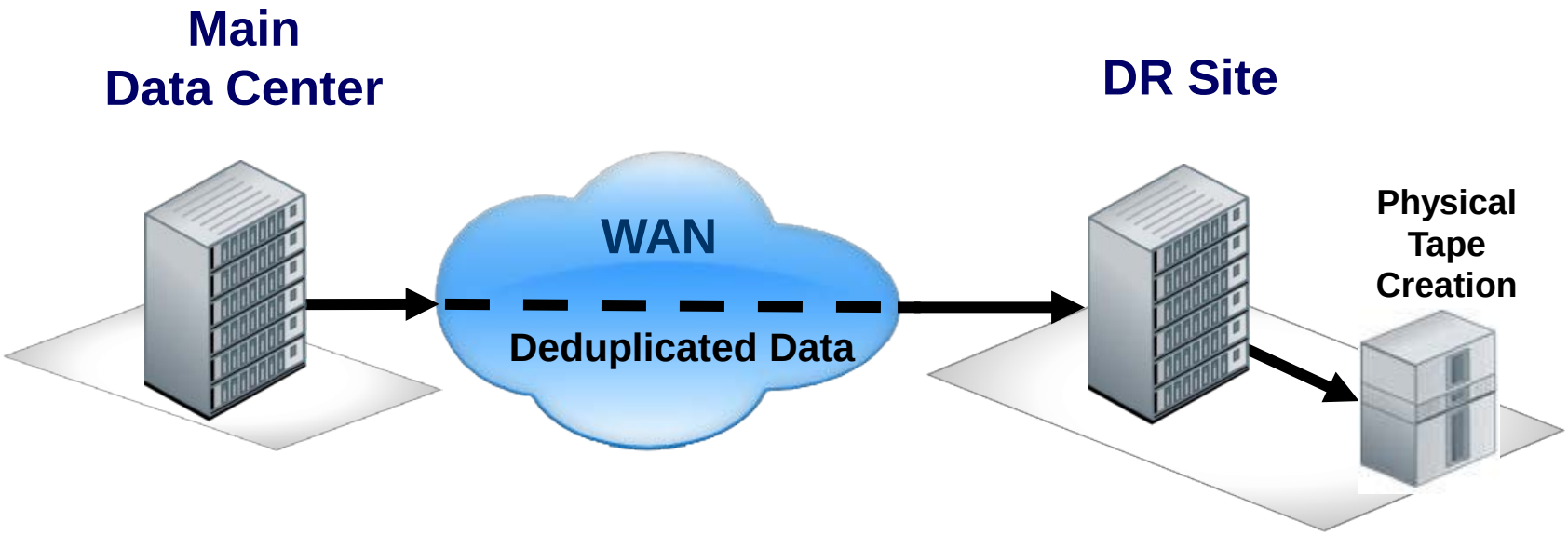
Dedupe in DR: Network Efficiency

- **Bandwidth Optimization for Increased WAN Efficiency**
  - Transfer more information per pipe
  - Supplement deduplication with compression
- **Support Remote Office Protection**
  - Enable Backup Centralization
- **Leverage existing bandwidth**



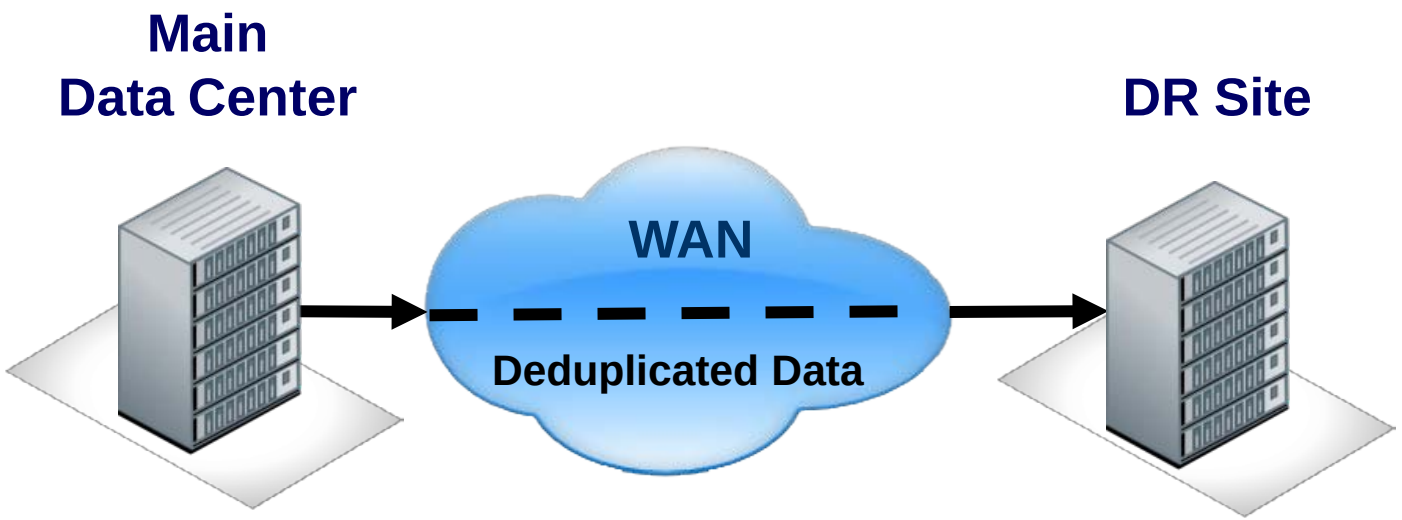
Dedupe in DR: Automation Benefits

- Simplify the offsite process
- Minimize risk of data loss/data theft
- Consolidate physical tape creation



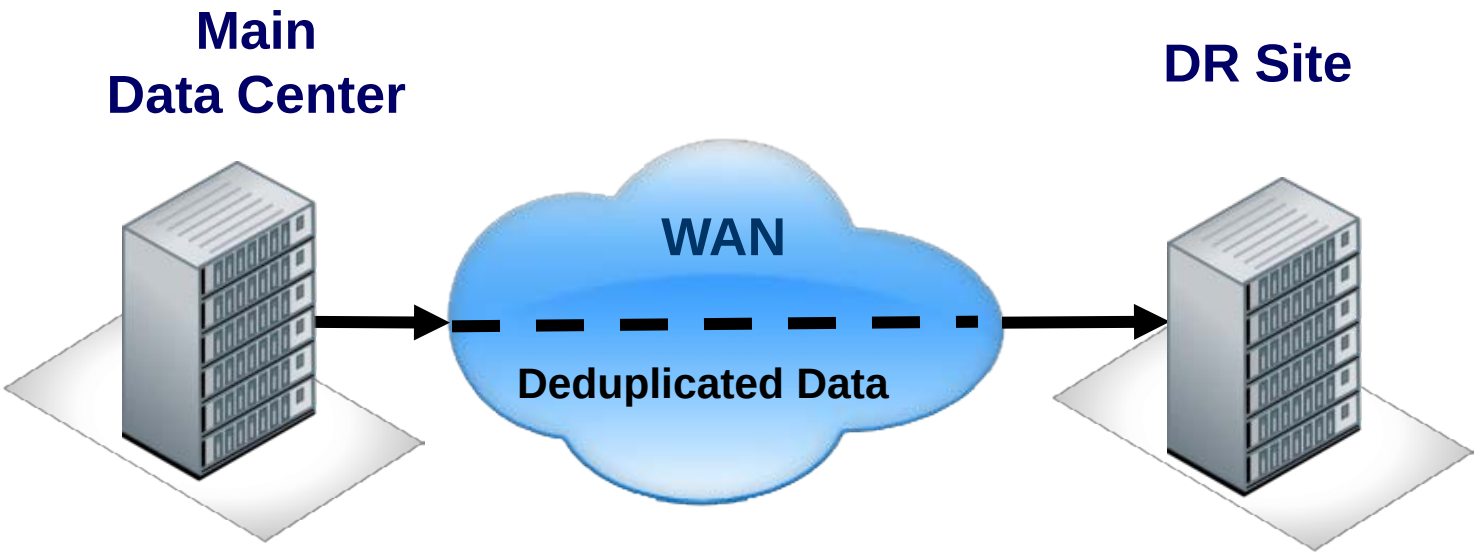
Dedupe in DR: Risk Reduction

- Human error reduced with automation
- Regulatory compliance more easily achieved
- Improve data access reliability



Dedupe in DR: Cost Savings

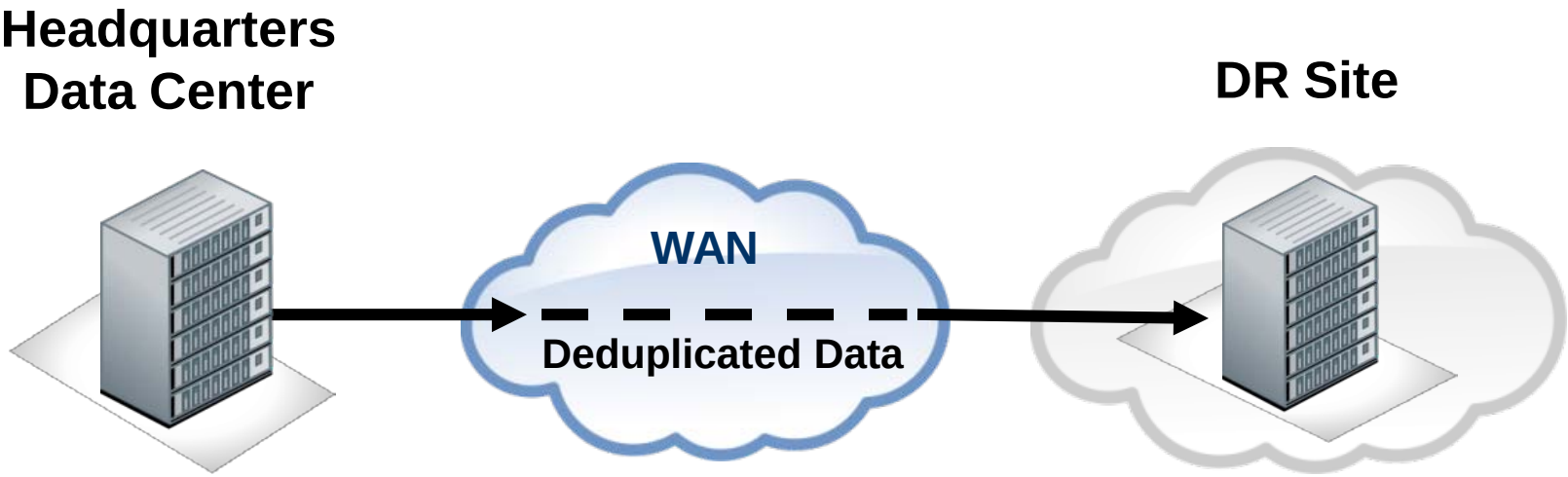
- Reduced network costs
- Reduced manual media handling
- Reduce tape archival services
- Minimize data loss



Dedupe in DR: Requirements

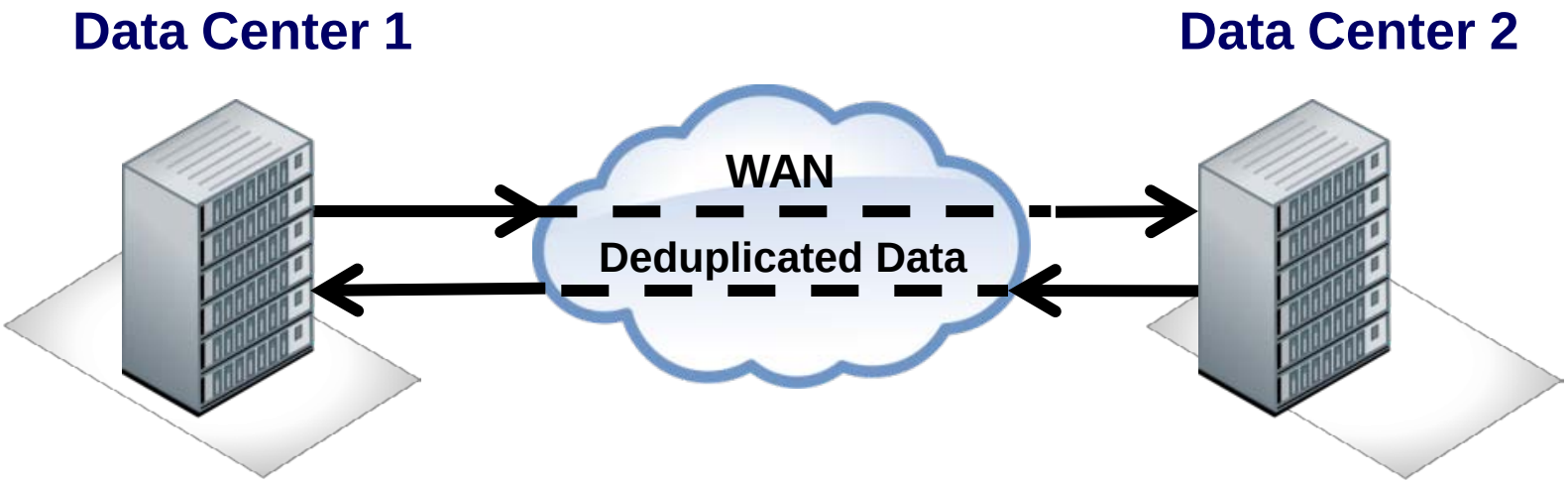
- Replicate large volumes of data over time
- Send only "Changed Data" over the Network
- Both sides must have same dedupe technique
- Perform fast sub-volume data restores from remote site
- Provide fast volume/system restores (failover) at remote site
- Provide resiliency/high availability and ease of management

Use Model: One Way



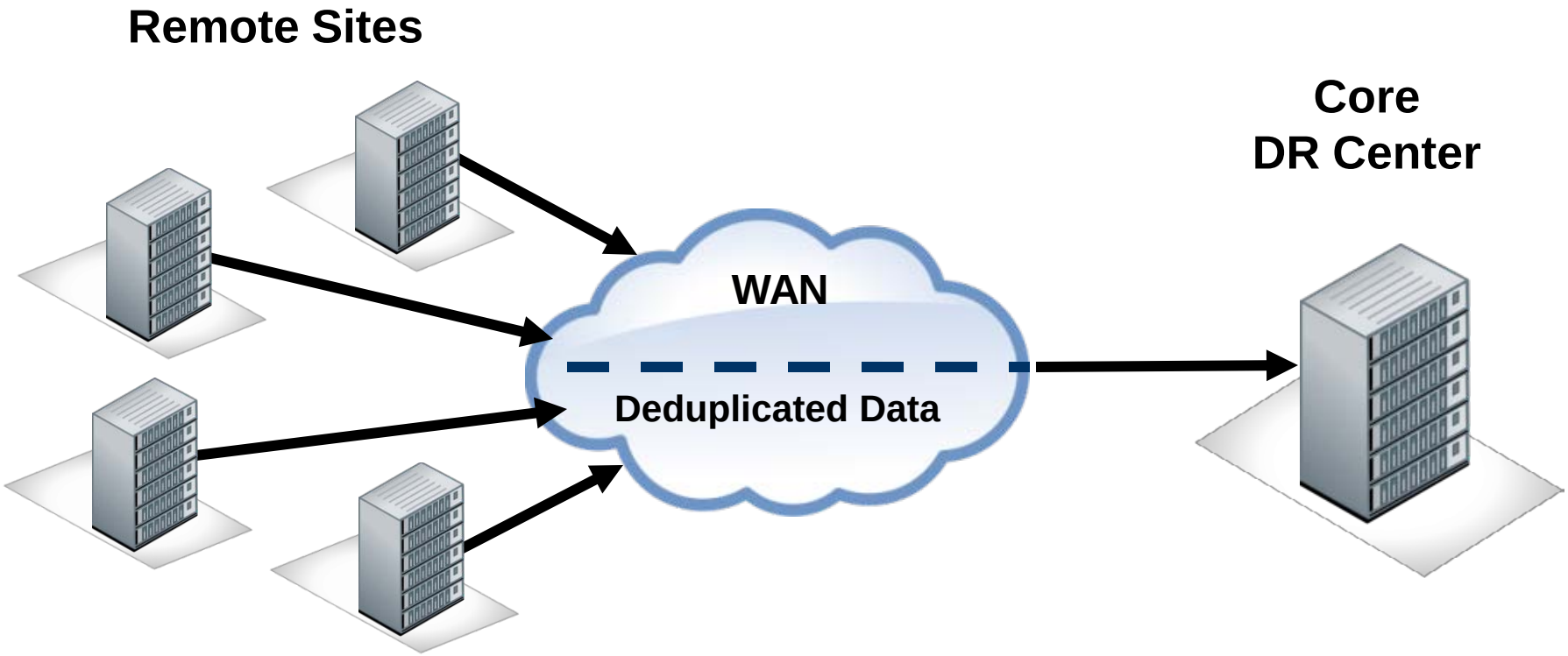
- Data is deduplicated and replicated to a DR site
  - May be private cloud or public cloud
- In event of data becoming unavailable at the headquarters data center:
  - Data can be restored to headquarters data center
  - Data can be used at the DR site (remote restore)

Use Model: Two Way



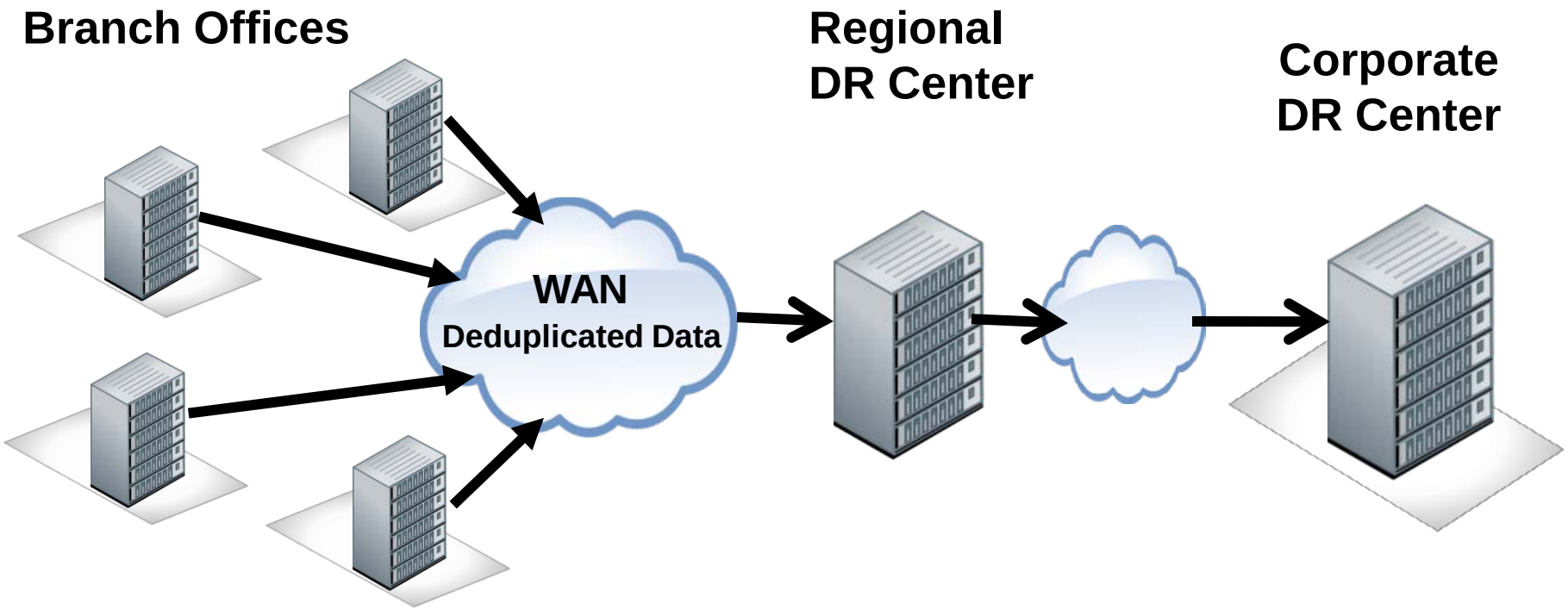
- Data is deduplicated & replicated bi-directionally between two production data centers
  - Each data center acting as a “DR Site” for the other

Use Model: Multi-Node



- Data is deduplicated and replicated from multiple regional data centers to a main DR center
  - Core DR center acting as a "DR Site" for all production data centers

Use Model: Multi-Hop



- Data is deduplicated and replicated from multiple branch offices to a regional data center and then to a main DR center

Deduplication: Potential Issues

- Be Aware of the Potential Pitfalls / Trade-Offs
  - May Decrease Data Ingestion Performance
  - Can Negatively Impact Restore Performance
  - May not Scale in Performance
  - Encrypted Data Limits Deduplication
- Use of Cloud Provider Introduces Special Considerations
  - Careful selection required
  - Service provider needs ability to reconstitute deduplicated data if remote restore required
- Easy to Under-Estimate the Bandwidth Required
  - Maximum Changed Data Size ÷ Replication Window = Data Rate Needed

Dedupe in DR: What Really Matters?

Education SNIA

- Focus on your Service Level Agreements (SLAs)
  - Needs to meet window for *Replication*
  - Needs to meet SLA for System *Recovery or Data Restore*
- Is it Necessary to Dedupe All Data?
  - May have regulatory issues for some data
  - Some data types not conducive to deduplication
- Can the Dedupe Solution Scale to Meet Your Needs?
  - Needs to scale in capacity & performance
  - Different dedupe approaches yield different reduction ratios
  - Capex & Opex savings should increase in proportion

Review

- Using Dedupe in DR Can Help Organizations:
  - Satisfy ROI/TCO requirements
  - Manage data growth
  - Increase efficiency of replication and DR
  - Reduce overall cost of storage
  - Reduce required network bandwidth
  - Reduce operational costs including:
    - Infrastructure costs - required space, power and cooling
  - Reduce administrative costs
  - Avoid risk of physical transfer of tapes

Summary of Considerations

- Multiple Elements to Consider when Evaluating Deduplication Technologies for DR Projects:

WAN Efficiency of Deduped Data

Public Vs. Private DR Site

Restore Performance of Deduped Data

Overall Power Consumption

Scalability of Deduplication Solution

Resiliency/HA of Deduplication Solution

Summary (continued)

- There is no “Right” Solution for Everyone!
  - The Appropriate Solution will Vary by Environment and Requirements
- As with All Backup and DR Systems - -
  - TEST TEST TEST

Q&A / Feedback

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

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

- SNIA Education Committee

Michael Alvarado
Mike Dutch
Larry Freeman
Bernd Henning
Gene Nagle
Ronald Pagani
Thomas Rivera
Tom Sas