



Education

IS THE CLOUD READY FOR DEDUPE?

Jered Floyd, Permabit Technology Corporation

Author: Jered Floyd, Permabit Technology Corporation

- The material contained in this tutorial is copyrighted by the SNIA unless otherwise noted.
- 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.

➤ Is the Cloud Ready for Deduplication?

- ◆ Most organizations are trying to reduce total storage costs while facing 50% annual data growth rates and an increased demand to store data long term. As cloud storage increasingly becomes a part of IT storage strategies, customers expect a highly cost-effective, secure, and scalable solution that provides seamless access to their data. This session will explore why deduplication is a necessary and effective strategy for data in the cloud.

➤ Why Cloud Storage?

- ◆ Agile IT environments across coordinated layers
- ◆ Reduce capital costs, optimize operational expenses
- ◆ Leverage expertise from all service dimensions
- ◆ Pay for the IT you use, not just what you need for peak times

➤ Where does it live?

- ◆ Public – secure, multi-tenant, externally hosted
- ◆ Private – secure single or multi-tenant usually hosted inside a firewall or by a 3rd party
- ◆ Hybrid – combination of public and private

➤ How is it stored?

- ◆ Block-oriented – provide block storage services for cloud computing
- ◆ Object-oriented – access objects to/from the cloud via an Object API

Data Deduplication [Storage System]

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.

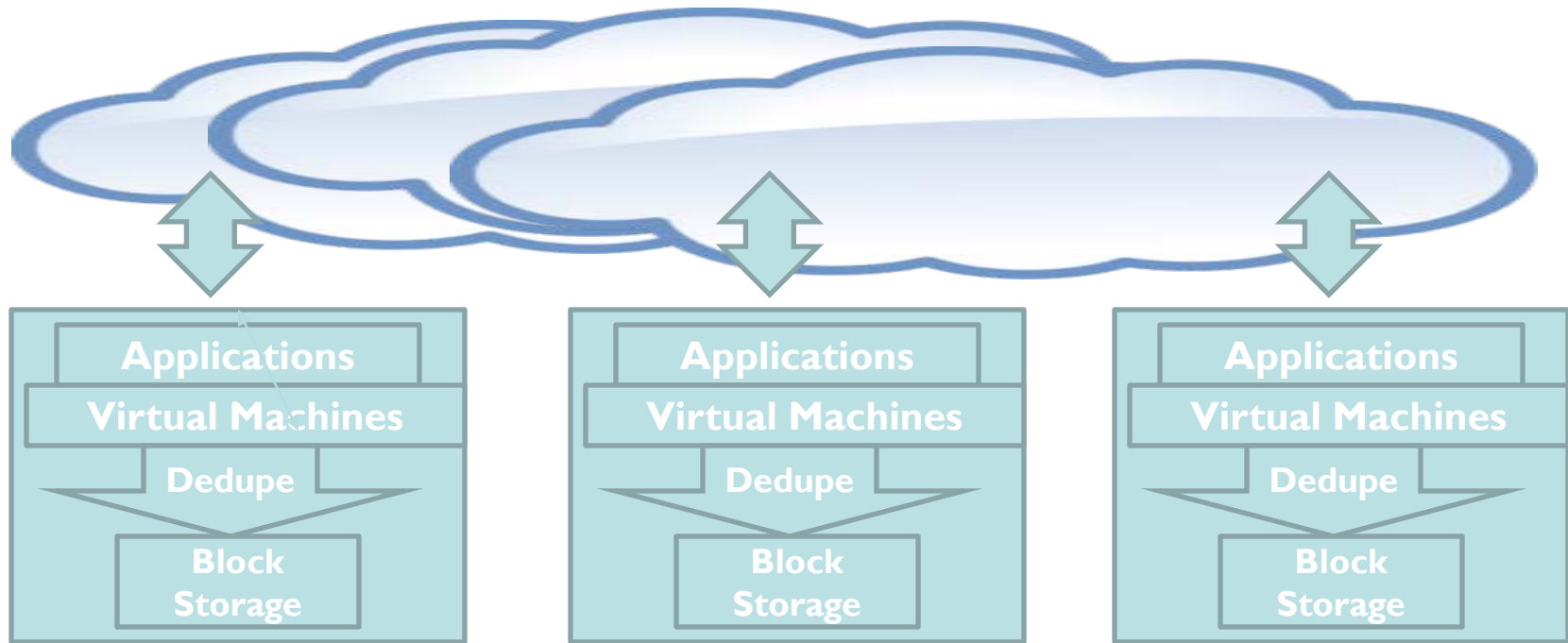


Check out these **SNIA Tutorials**:

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

- Deduplication provides competitive differentiation on performance and price to cloud providers:
 - ◆ Less bandwidth required to store data
 - ◆ Reduced cloud storage costs
 - ◆ Less storage to manage
- Integrates with block and object based cloud storage

Deduplication for Block Storage

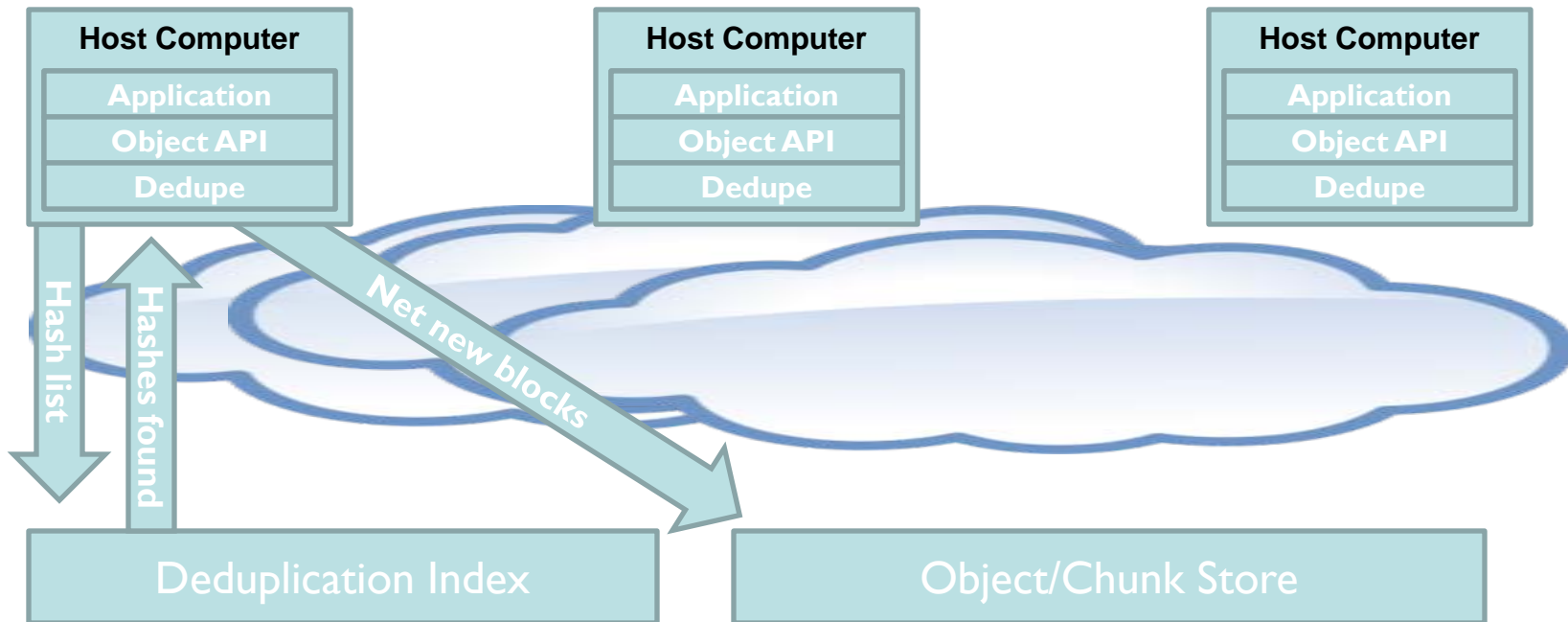


Benefits for Block Storage

- Improved cache performance
- Space efficiency – can be used to keep a single copy of application and OS blocks per physical compute server
- Makes low latency flash storage cost effective
- Speeds block-level replication/migration tasks

- Performance overhead - disk latency is a limiting factor in achieving high density of VMs in cloud compute environments
- Resource overhead can't cost more than savings

Deduplication for Object Storage



◆ Key Benefits

- ◆ Reduces the cost of network writes because chunks that have already been stored don't need to get sent
- ◆ Reduces the cost of storage for the provider

◆ Potential Issues

- ◆ Benefits depend on data type
- ◆ Must scale to cloud capacities
- ◆ Objects/chunks need to be encrypted before dedupe
- ◆ Variable chunk dedupe yields substantially greater savings here
- ◆ Cost of resource overhead can't be higher than savings

Benefits for Object Storage

- Reduces the cost of network writes because chunks that have already been stored don't need to get sent
- Reduces the cost of storage for the cloud provider (and potentially the end-customer)

- Benefits depend on type of data being stored and the workflow being used to generate and store the data
- Must scale to cloud capacities to really benefit
- Objects/chunks should be compressed/encrypted after any chunking is done, and before dedupe to maximize space savings and ensure security
- Variable chunk dedupe yields substantially greater savings for some object types (ZIP archives, office files, etc...)
- Cost of resource overhead can't be higher than savings

- **Dedupe saves costs:**
 - ◆ Reduces physical storage requirements
 - ◆ Improves network efficiency
- **Cloud saves costs:**
 - ◆ Reduces outlay for capital equipment
 - ◆ Lowers operational expense
- **Dedupe and Cloud together can reduce costs associated with both cloud computing and cloud storage**

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

- SNIA Education Committee

Louis Imershein

Mike Ivanov

Jered Floyd

Send any questions or comments on this presentation to SNIA: tracktutorials@snia.org