



# Backup to Disk with DataDomain

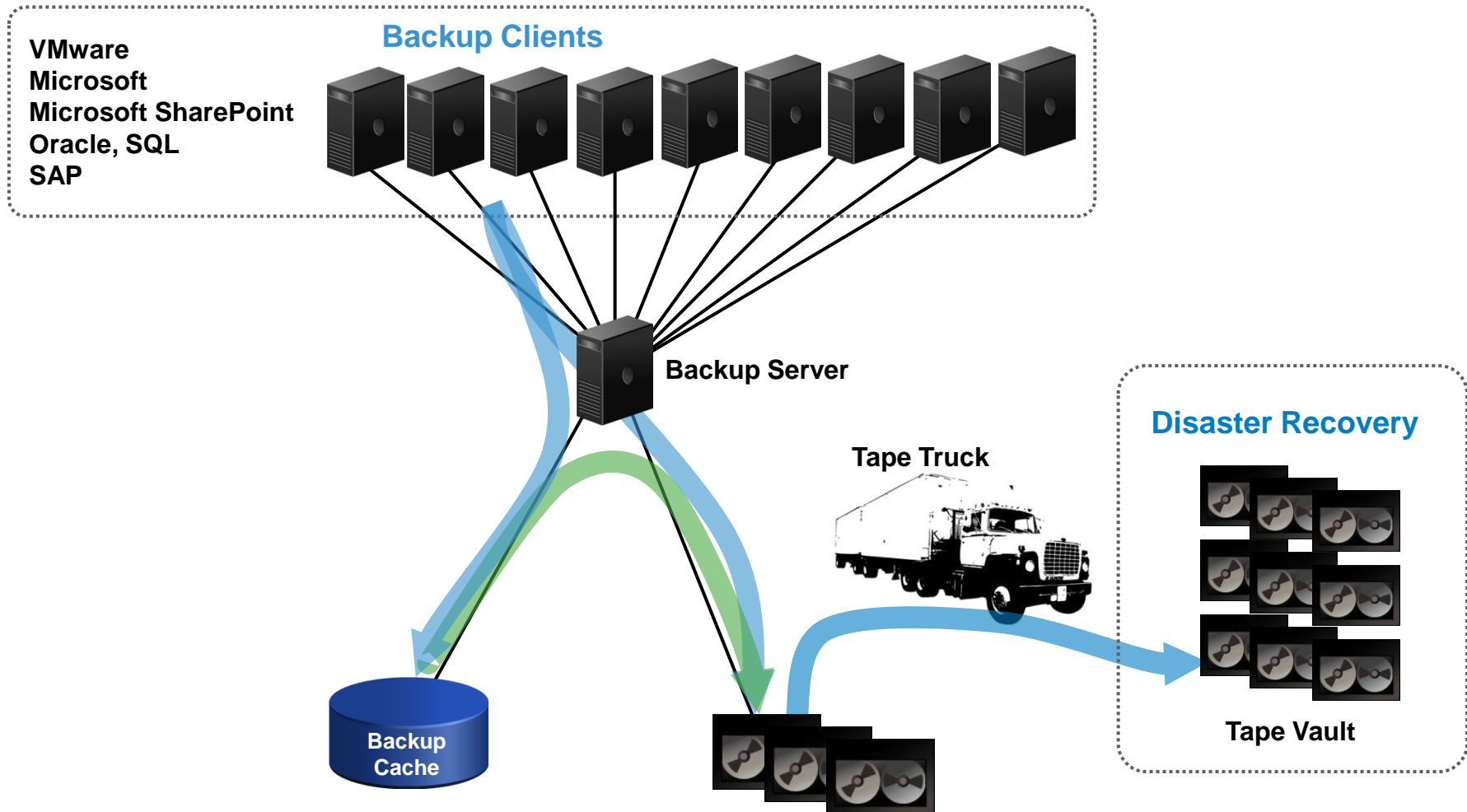
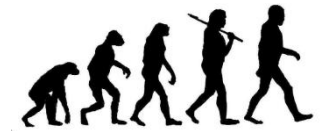
Why to use target based deduplication?

DD Workshop – ADD 9.9.2010

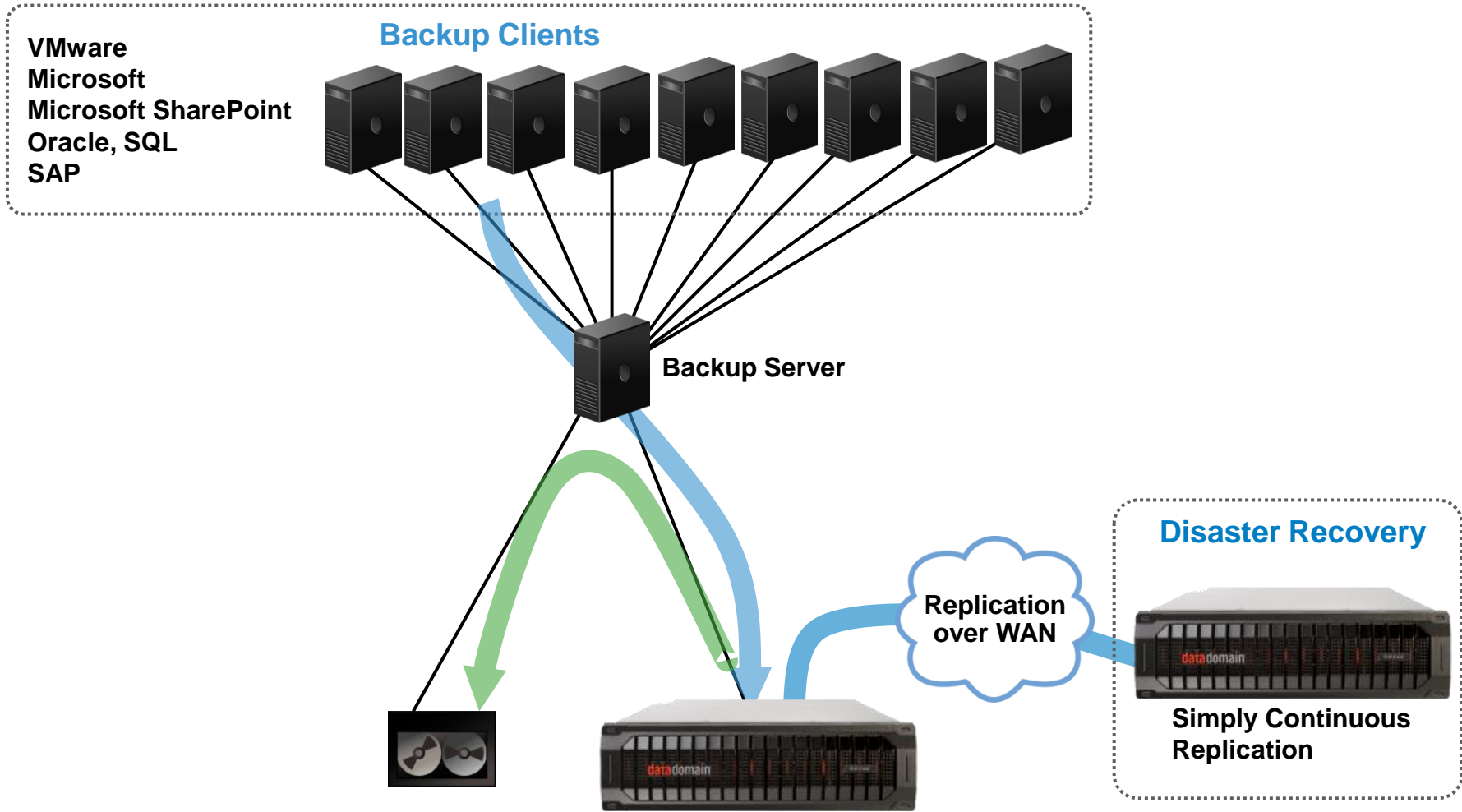
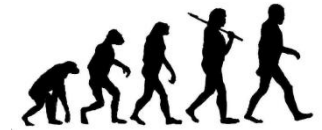
Richard Schmidt

Technology Consultant  
Backup and Recovery Systems

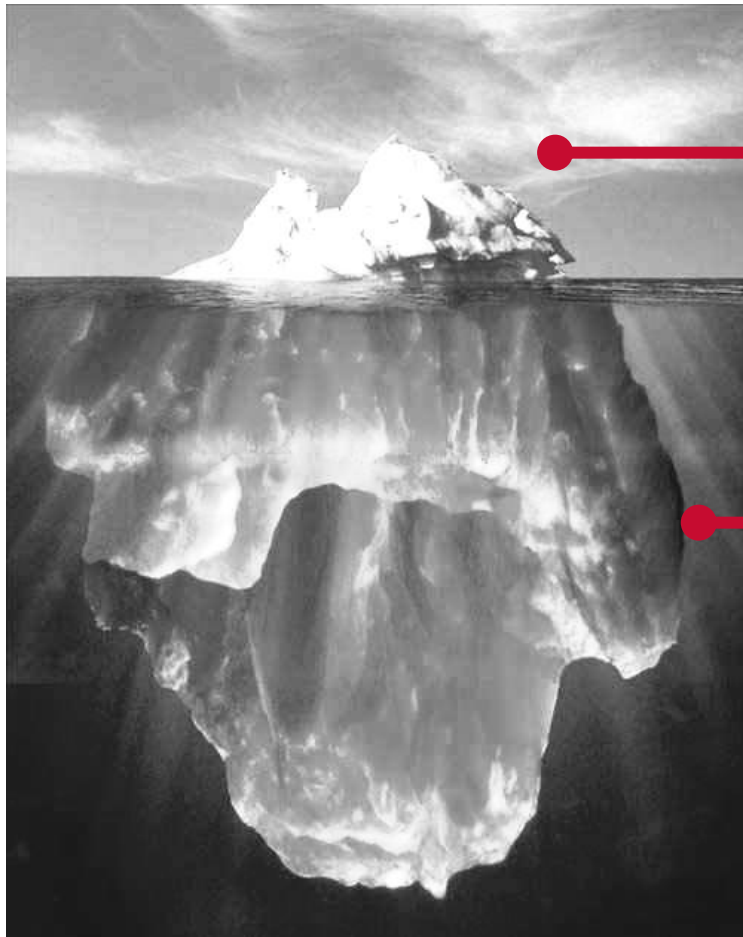
# Evolution of Backup Infrastructure



# Evolution of Backup Infrastructure



# Back-up to Disk: Capacity = €€€€€€€€



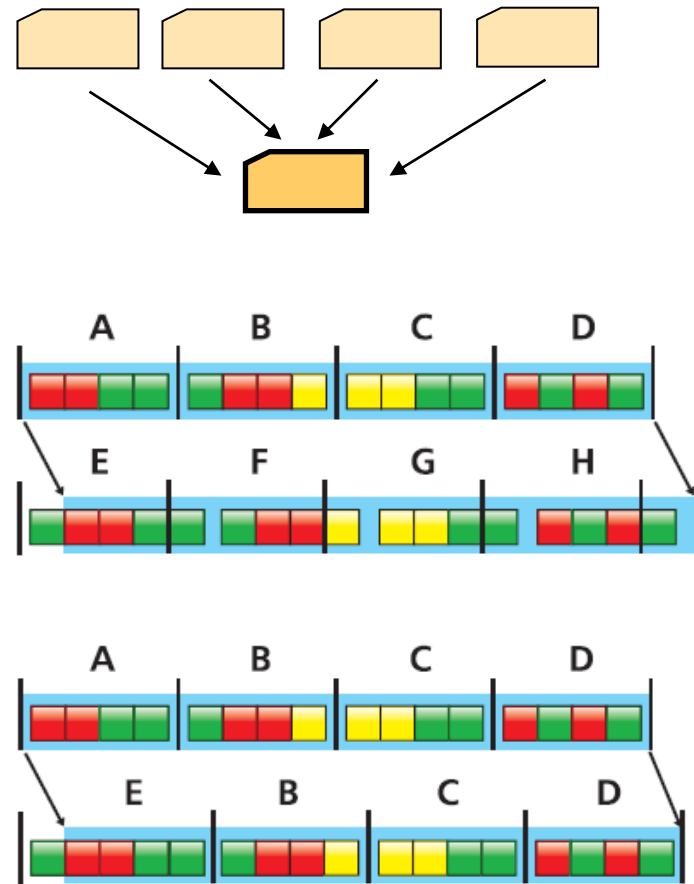
**Primary  
Data  
1x**

**Back-up  
Data  
15x-20x**

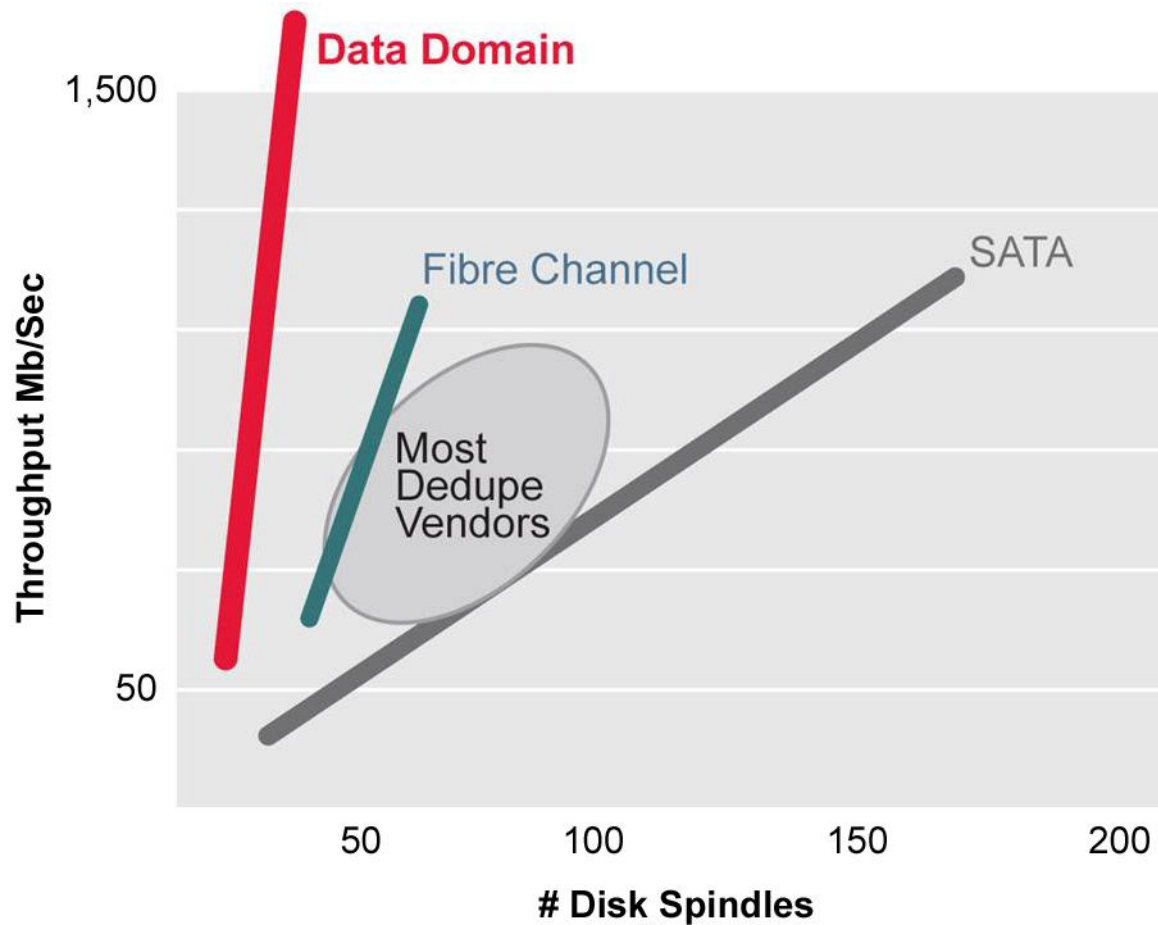
# Data deduplication: HOW

## Granularity:

- File level = Single Instance Store
  - EMC Centera
  - EMC Celerra
  - *Low resource utilization*
- Sub-File level = Block level
  - Fixed block length
  - *Medium resource utilization*
- Variable block length (dynamic)
  - EMC Avamar
  - EMC Data Domain
  - *High resource utilization*

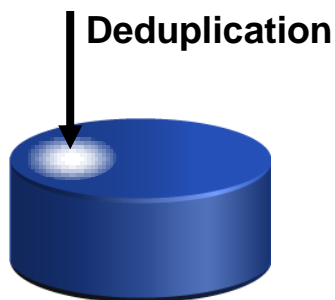


# Performance: CPU-centric vs. Spindle-bound



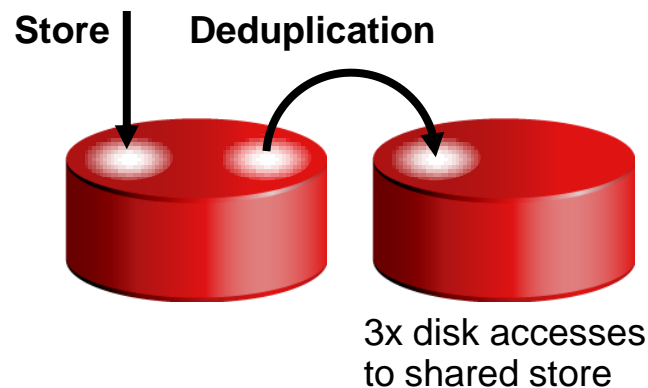
# Methodology: Inline vs. Post-Process Deduplication

## Inline: Deduplication Before Storing



- Other activities unimpeded
  - Predictable
  - Simpler

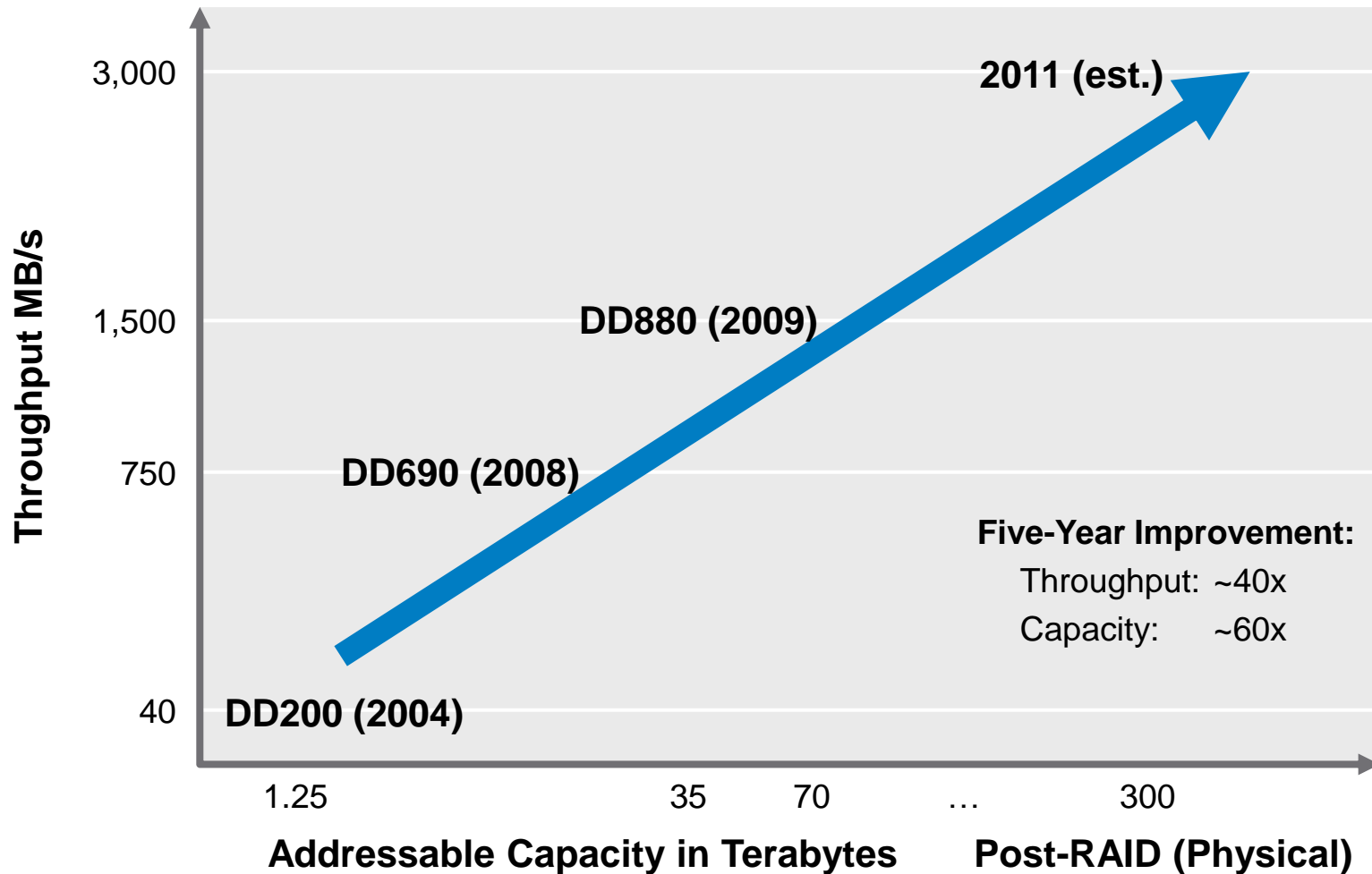
## Post-Process: Deduplication After Storing



- The more processes, the more resource contention
  - Copy to tape: Too slow to stream tape
  - Recovery: Service level agreement predictability
  - Replication: Poor time-to-disaster recovery
  - Deduplication : If interleaved with backup or restore
- More administration to fight these issues

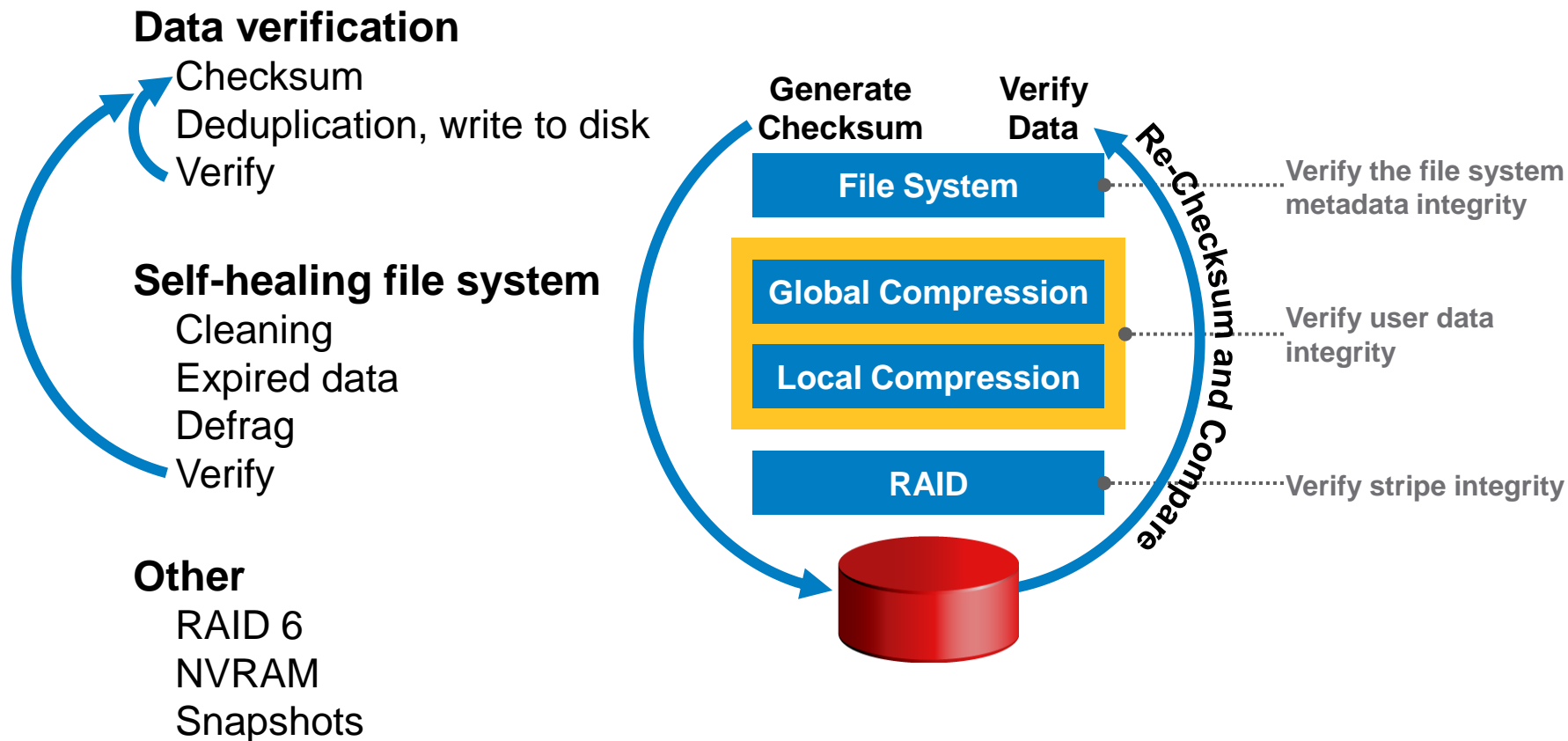
# Scalability: Data Domain Systems

Data Domain scaling architecture



# Data Integrity: Data Invulnerability Architecture

Trust but verify—"hope" is not a strategy



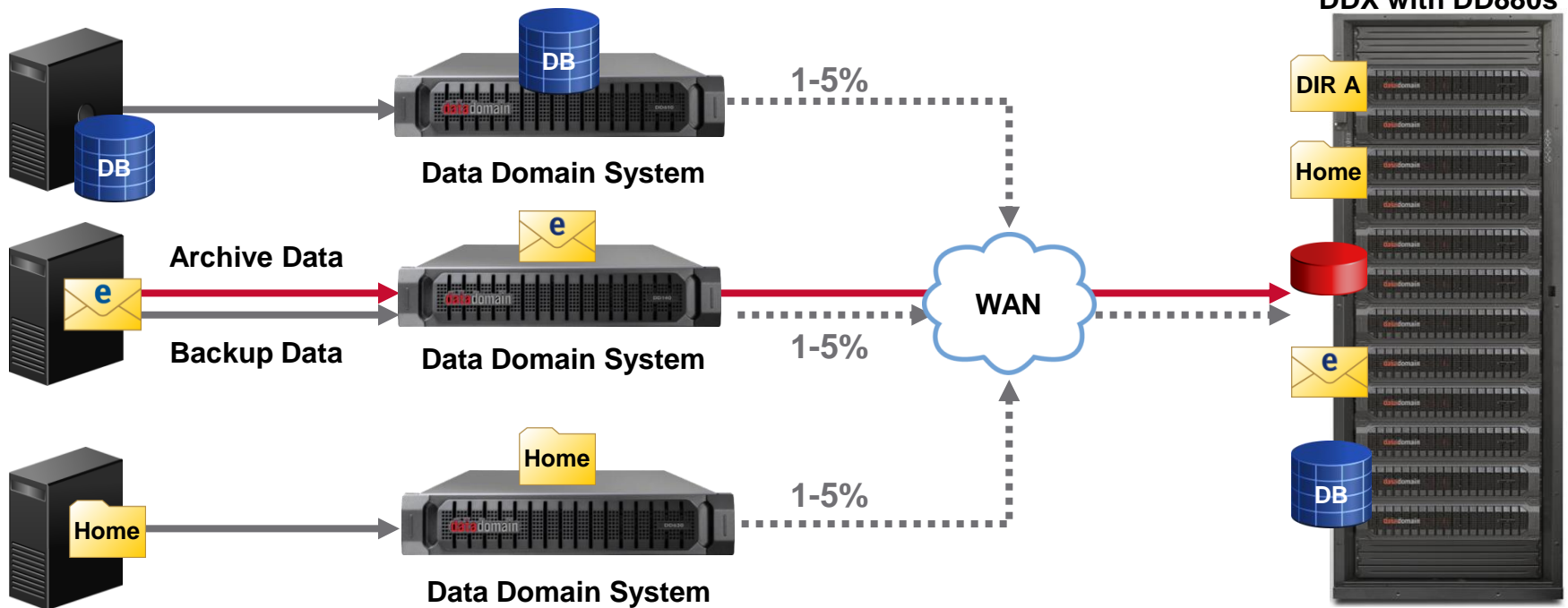
# Multi-Site Protection for Remote Office

Many-to-one consolidation

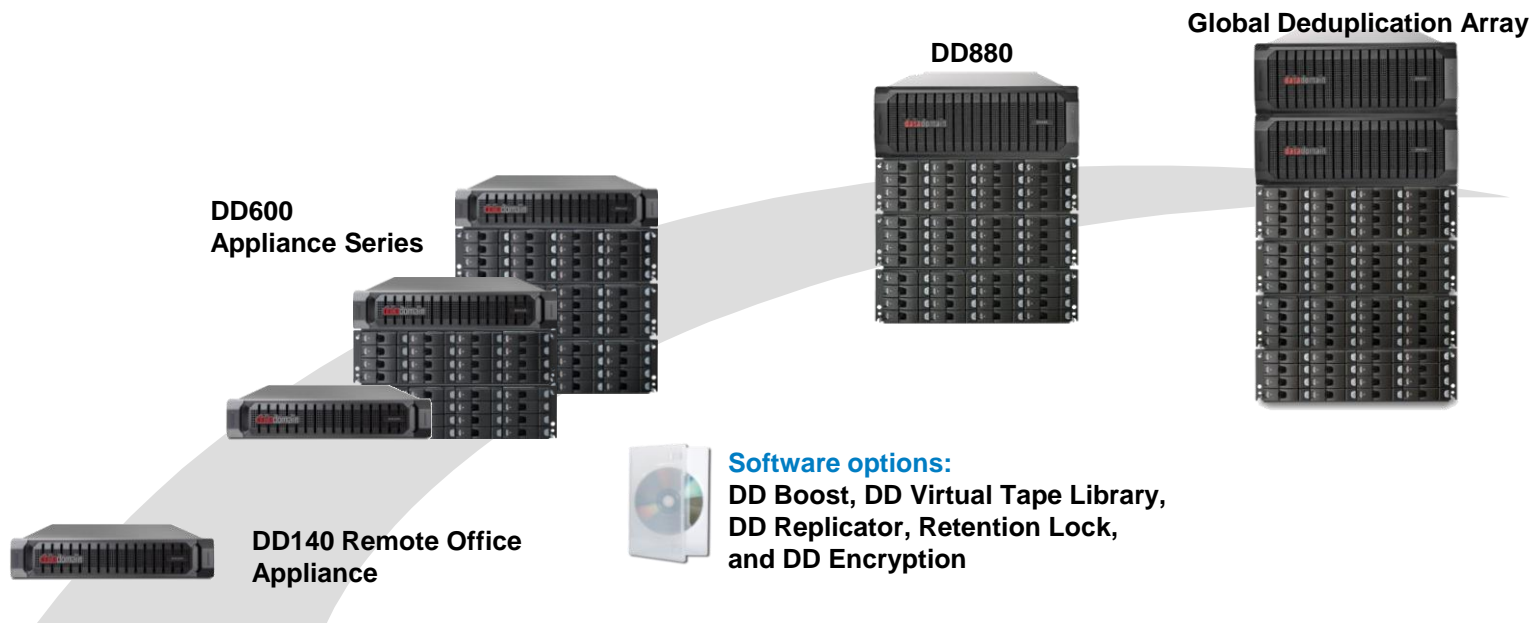
- **Fan-in scalability:**
  - Supports hundreds of remote sites
  - Partial Backup Replication (no full replication required)

## Data Center Hub

## Remote Sites



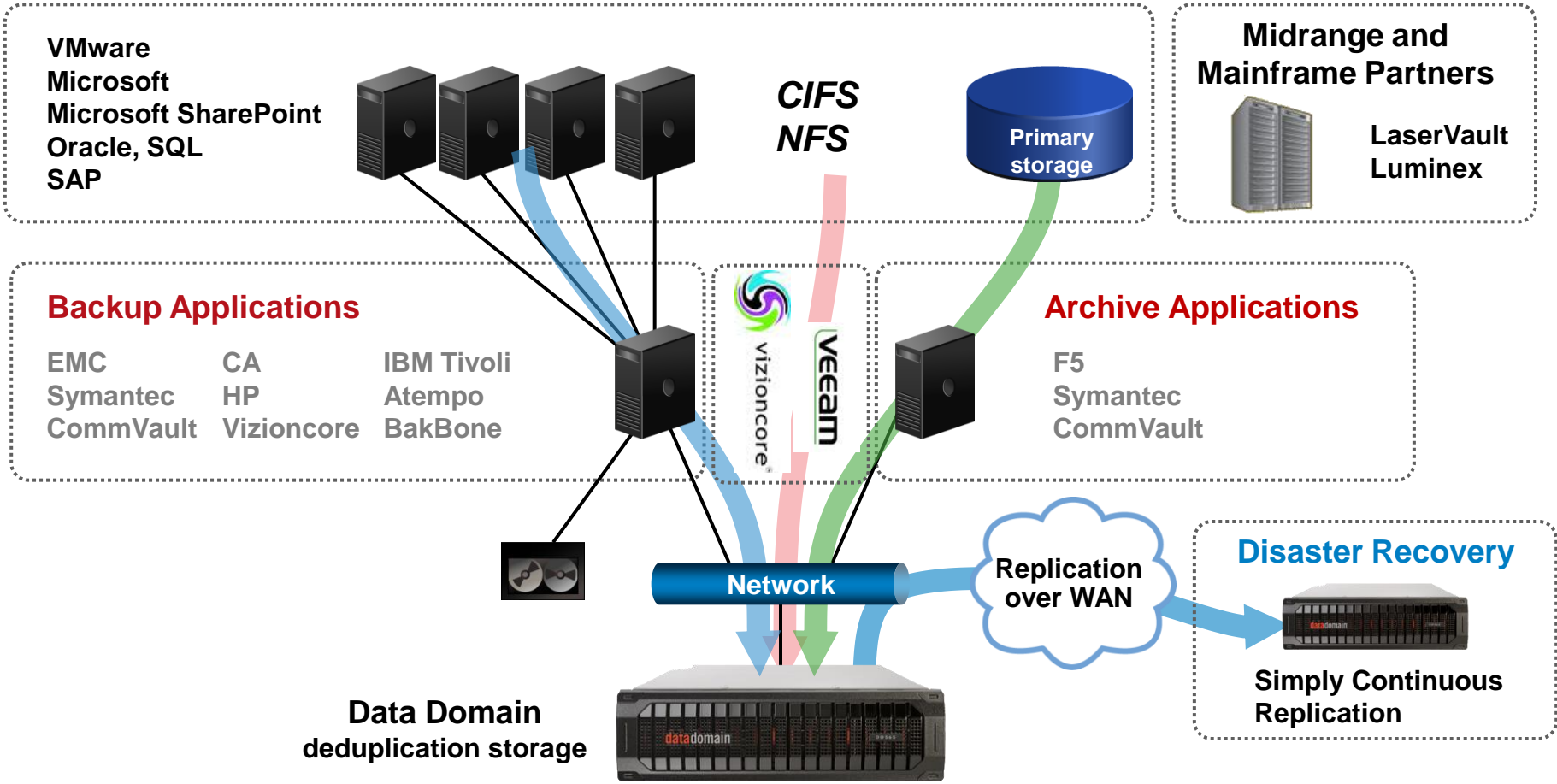
# Industry's Most Scalable Inline Deduplication Systems



|                         | DD140     | DD610         | DD630        | DD670         | DD690          | DD880          |
|-------------------------|-----------|---------------|--------------|---------------|----------------|----------------|
| <b>Speed (Other)</b>    | 450 GB/hr | 675 GB/hr     | 1.1 TB/hr    | 3.6 TB/hr     | 2.7 TB/hr      | 5.4 TB/hr      |
| <b>Speed (DD Boost)</b> | 490 GB/hr | 1.3 TB/hr     | 2.1 TB/hr    | 5.4 TB/hr     | 3.9 TB/hr      | 8.8 TB/hr      |
| <b>Logical capacity</b> | 9 - 17 TB | 40 - 80 TB    | 80 - 160 TB  | 550 – 1.1 PB  | .350 - .700 PB | 1.4–2.8 PB     |
| <b>Raw capacity</b>     | 1.5 TB    | Up to 6 TB    | Up to 12 TB  | Up to 76 TB   | Up to 48 TB    | Up to 192 TB   |
| <b>Usable capacity</b>  | 0.86 TB   | Up to 3.98 TB | Up to 8.4 TB | Up to 55.9 TB | Up to 35.3 TB  | Up to 142.5 TB |

# Data Domain Infrastructure and Ecosystem

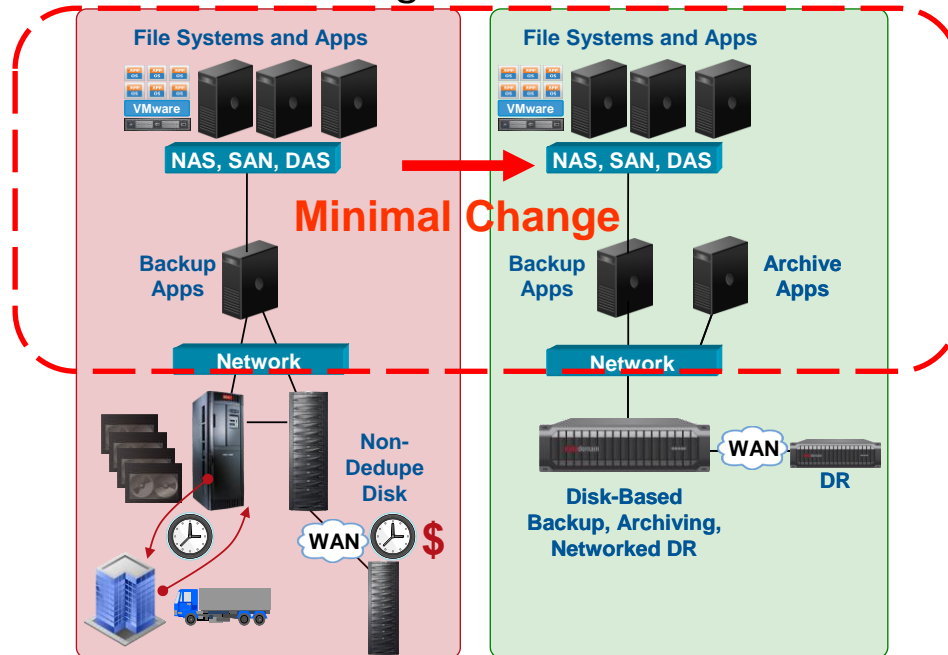
**Backup**      **NAS**      **Archive**



# EMC Data Domain

Resource Less. Manage Less

- Add deduplication to existing backup infrastructure through standard interfaces including NAS, virtual tape library, or OpenStorage API
- Integrated with popular archive applications (Symantec, EMC and CommVault) through standard NAS-based locking interfaces



How...

✓ Easy integration with existing backup software and archive applications

✓ Reduce back-end operational complexity

Thank you

**EMC<sup>2</sup>**<sup>®</sup>

**where information lives<sup>®</sup>**

*deduped*