



**FORSYTHE<sup>®</sup>**

# Efficient Backup with Data Deduplication:

Which Strategy is Right for You?  
Part 1 - Brian Peterson, Forsythe  
Part 2 - Mike Salins, EMC

# Forsythe solutions address...



## Cost

View economic models to analyze, & manage capital & operating expenses.

## Growth

Architect an infrastructure focused on adaptability, scalability, & flexibility.

## Risk

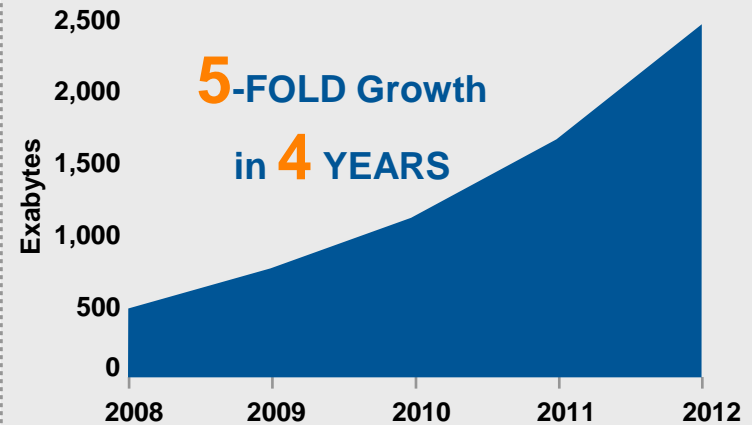
Address and mitigate compliance issues & design or maximum availability, recoverability & business continuance.

...to provide optimized infrastructure through people, process & tools

# Why is Data Deduplication Interesting?

- Data Growth
  - Backups consume 4 to 30x primary storage
- New Regulations
- Remote Office Consolidation
- Server Virtualization
  - Dynamic Environments
  - Virtual Server Sprawl
  - Heavily utilized servers

Digital Information Created and Replicated Worldwide



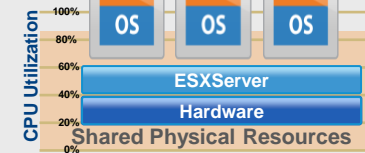
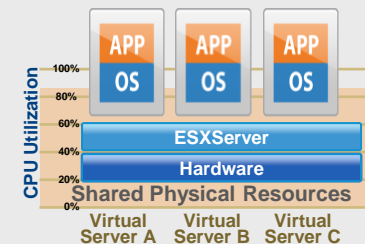
Source: IDC Digital Universe white paper, sponsored by EMC, May 2009

Old Paradigm



20 percent resource utilization

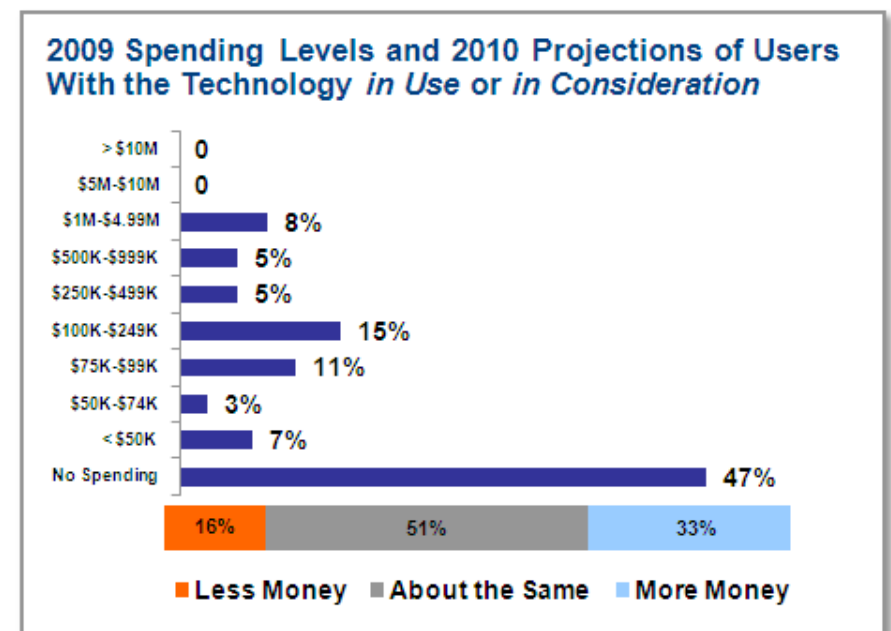
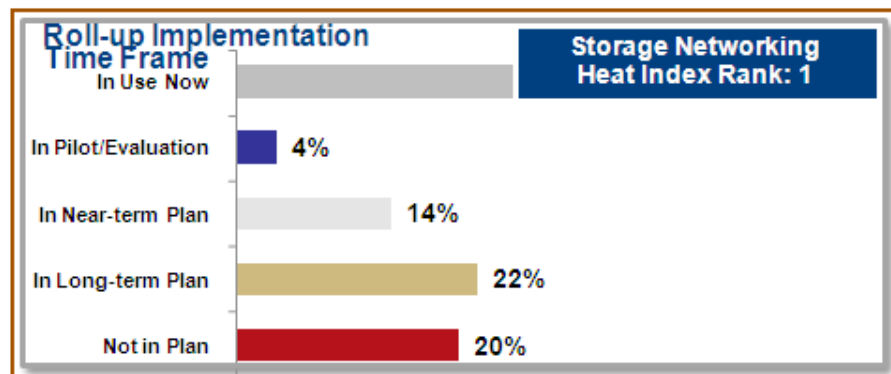
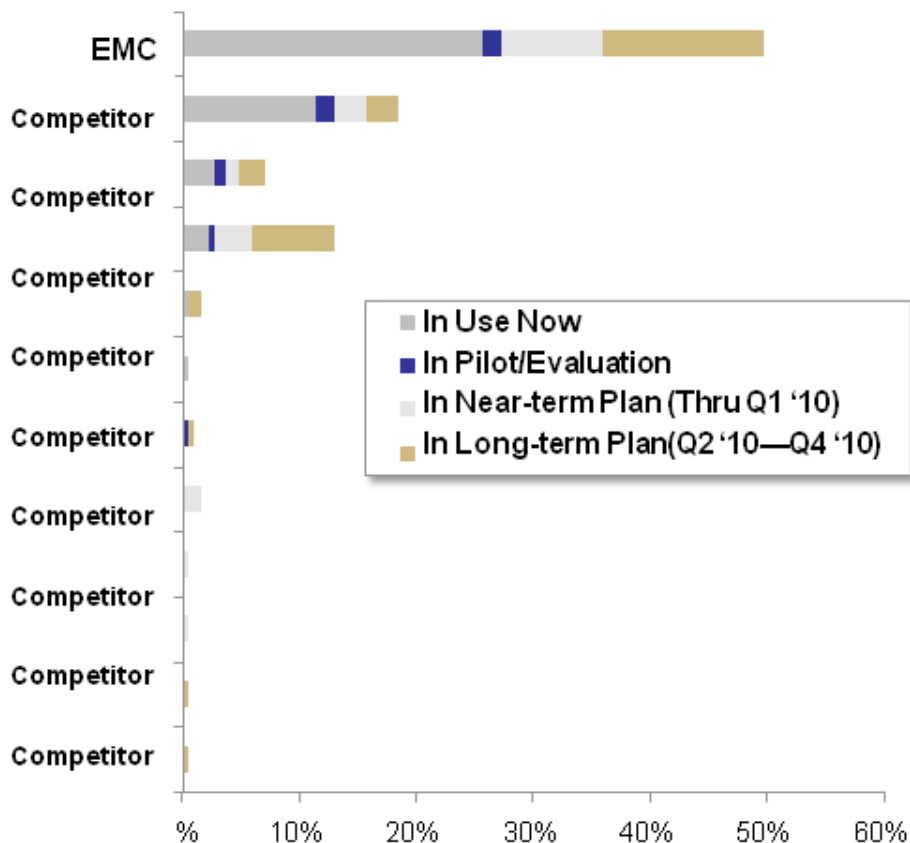
New Paradigm



80 percent resource utilization

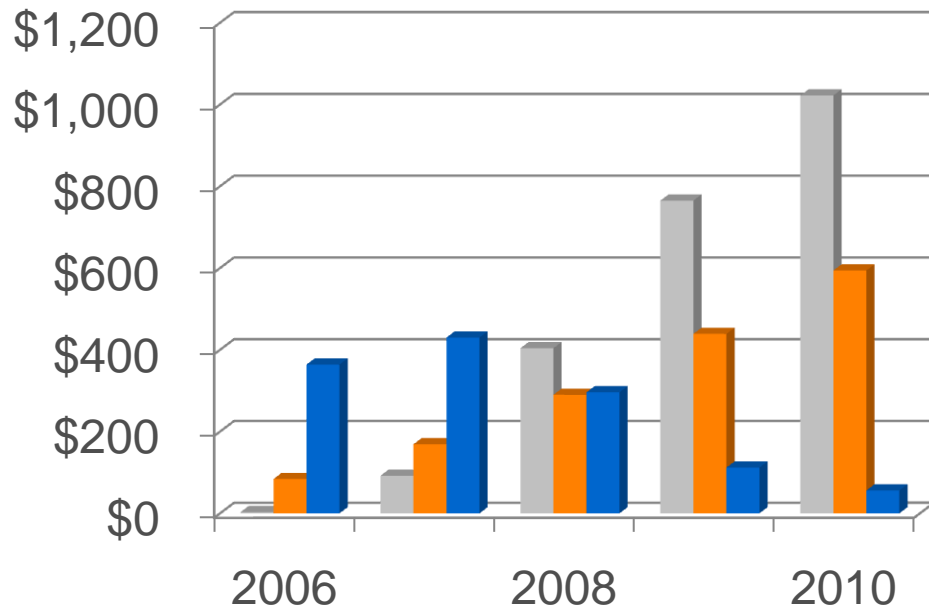
# Deduplication is “Top of Mind”

## F1000: Backup Data Reduction/Deduplication



# Data Protection Market Trends

## Data Protection Storage Forecast (\$M) by Type



- Capacity Optimized VTL
- Capacity Optimized Other Storage
- Non-Capacity Optimized Other Storage

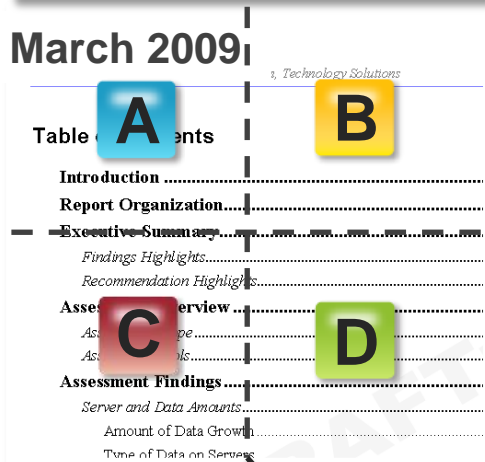
- Optimized (Deduplicated) disk capacity is growing rapidly
- Non-optimized disk capacity is less valued in the market
- Serious focus on reducing physical disk capacity & cost is expected to continue

# Deduplication Defined

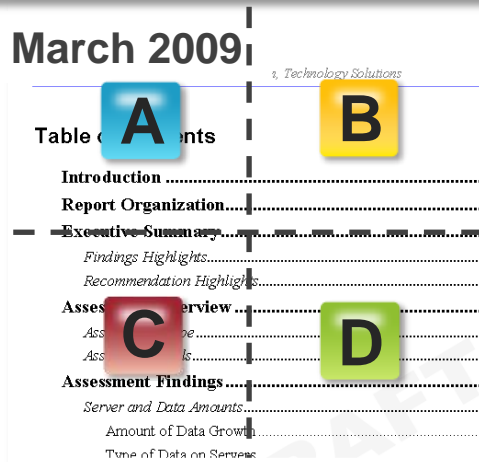
Technologies that are intended to **reduce** the amount of data stored on disk by only storing one **unique** copy of the data.

# How Does dedupe Work?

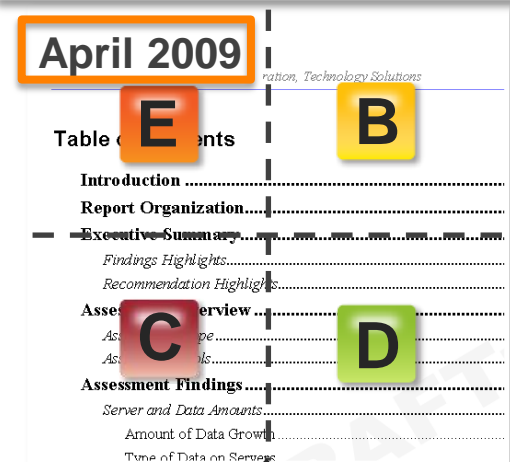
## 1 First Instance



## 2 Duplicate Instance



## 3 Modified Instance



Only unique data segments are backed up

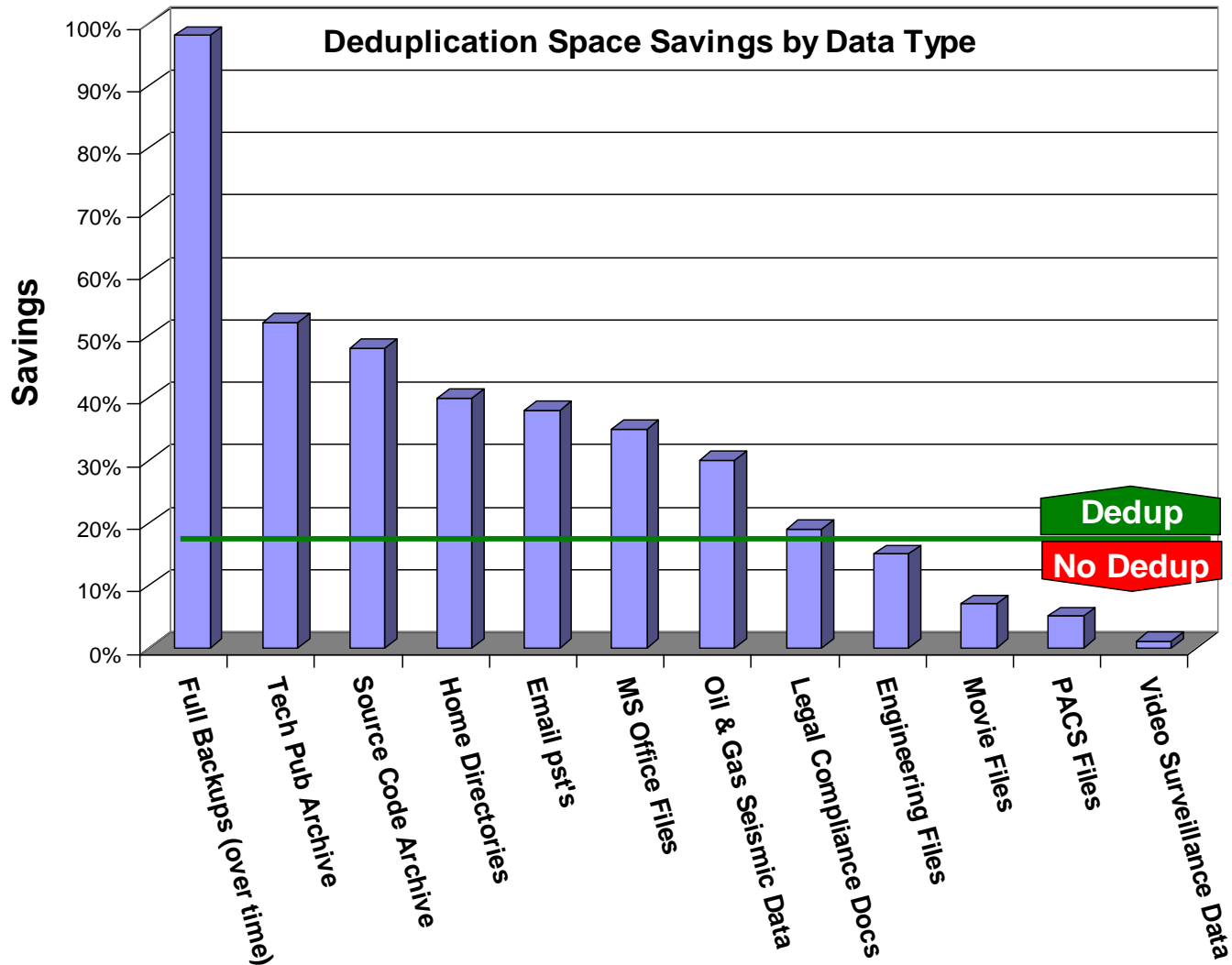
Data already backed up, so only a unique ID pointer is stored (20 bytes)

New data segment identified and backed up



Unique data stored on disk, available for immediate recovery.  
dedupe ratios can range from 1:1 to 500:1 depending on data structures, dedupe techniques, and retention period

# What Data Types to dedupe?

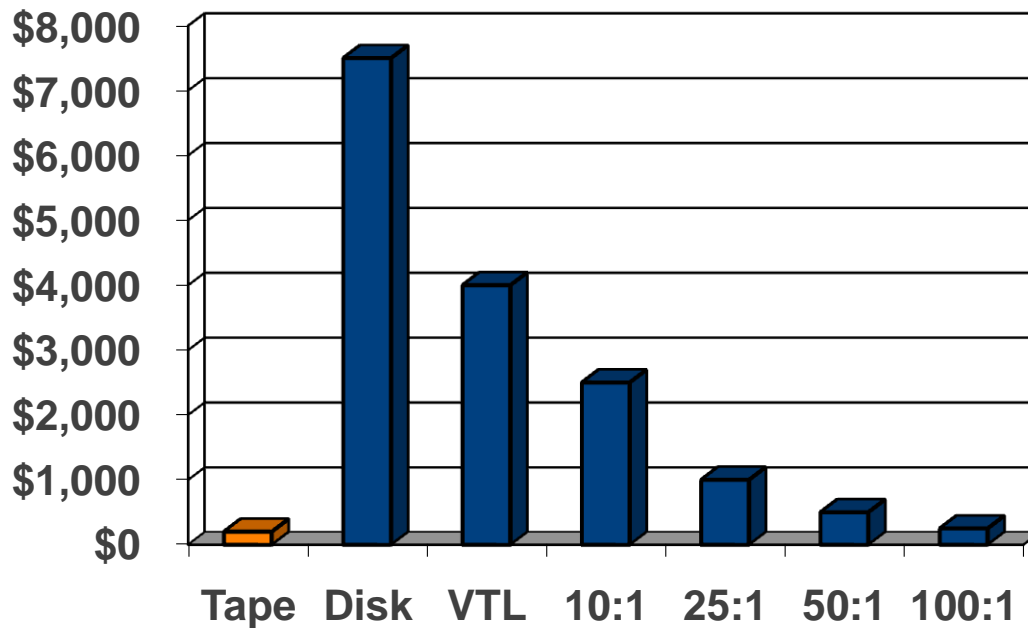


# Typical dedupe Ratios for Backup

- **First full backup: 1.5:1**
  - Less than typical tape compression of 2:1
- **Subsequent full backups: 30:1 to 50:1 (average 40:1)**
  - The NUMBER of full backups stored GREATLY effect the total dedupe ratio
- **Incremental backups: 5:1 to 7:1**
  - For backup systems like TSM, “incremental forever” scheme may be employed to achieve this effect
- **So “your mileage may vary”**
  - Specific analysis of your environment is required

# Backup Storage Costs Compared

Average acquisition costs per TB stored

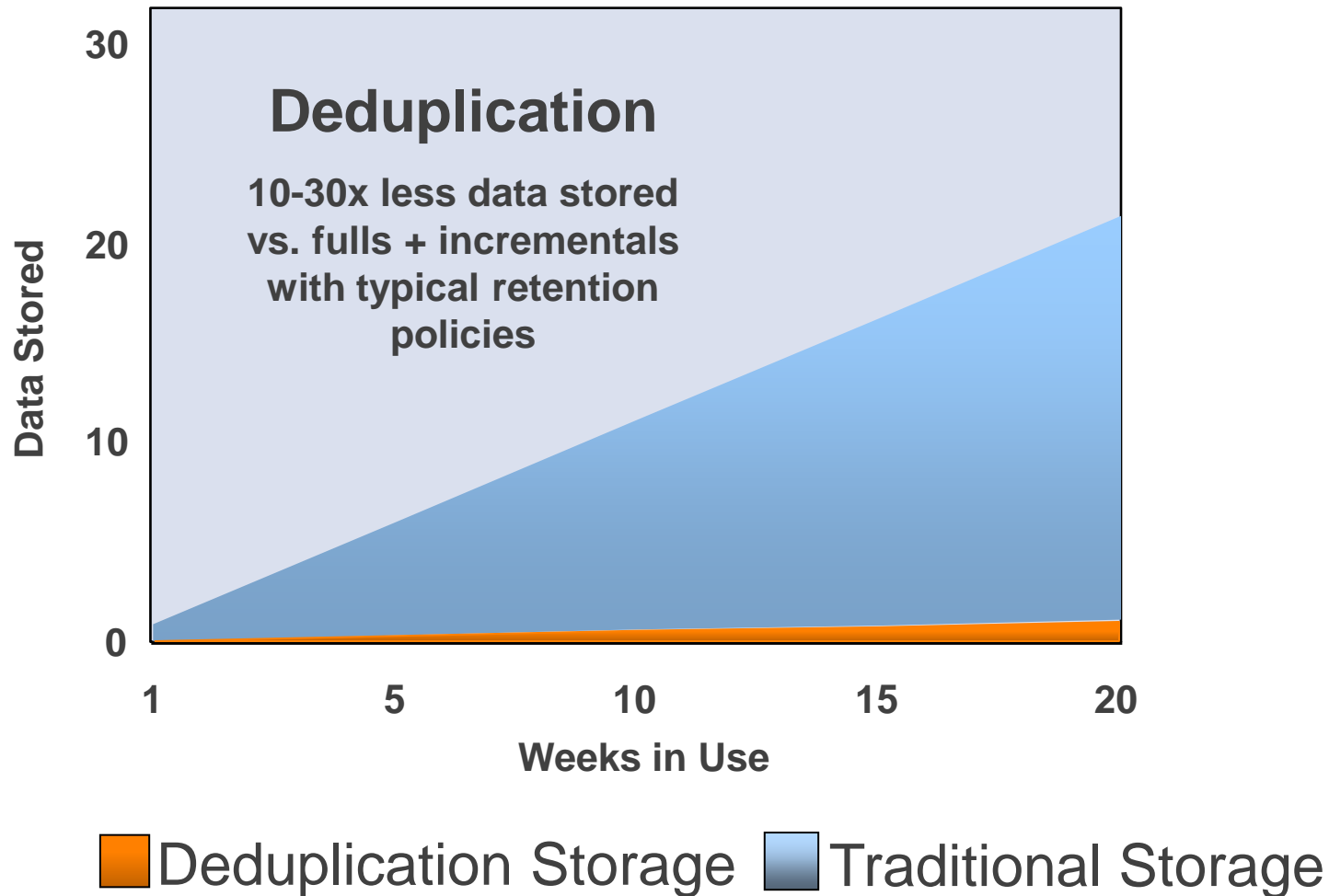


**Dedupe Ratio**

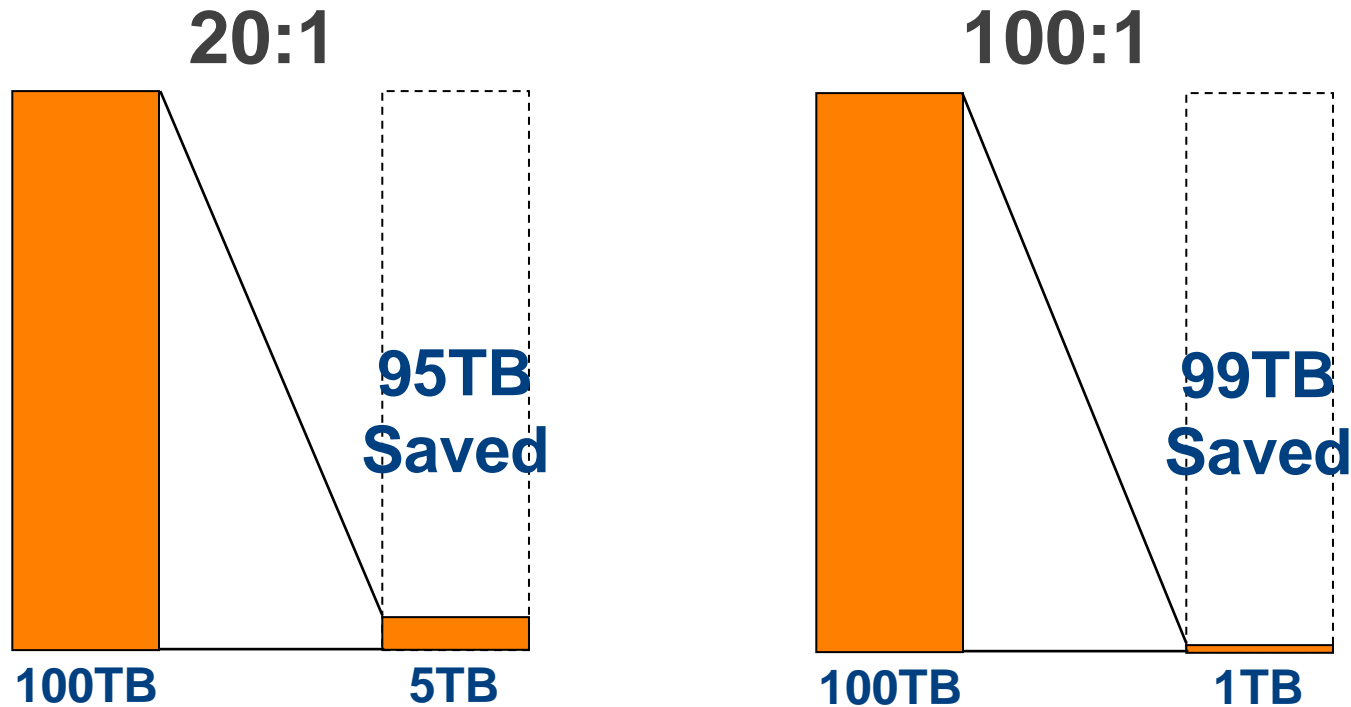
## Does not consider:

- Environmental savings
- Remote Replication Bandwidth savings
- Management savings
- SLA/recovery speeds vs. Tape/Vault

# Impact of Deduplication on Backups



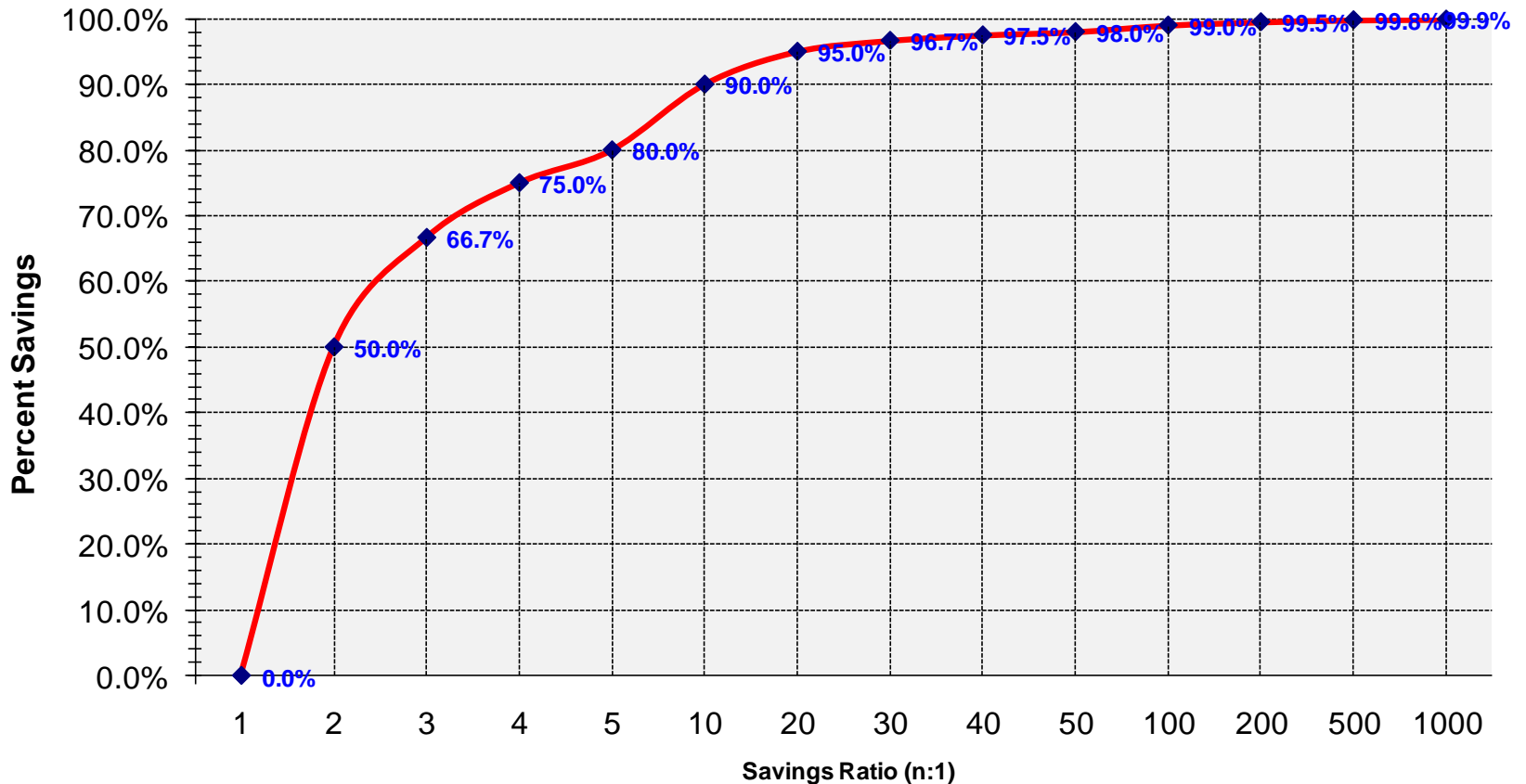
# Dedupe Ratios compared



**20:1 consumes only 5% more storage than 100:1**

# Dedupe Ratios vs. Storage Saved

Dedupe/Compression Ratios & Corresponding Percentage Saved



- 100:1 sounds like it is 5 times better than 20:1, but it isn't!
- Keep this perspective when comparing vendor ratio claims

# Dedupe technology choices

Where to dedupe

1. At the client
2. At the backup server
3. At the storage device (SAN or NAS connected)

When to process deduplication

1. In-line deduplication
2. Post process deduplication

What type of deduplication?

1. Object Based
2. Fixed Block
3. Variable Block

# Dedupe technology considerations

## Your Requirements

- What problems are you trying to solve?
- Service level needs (RPO, RTO)
- Is the solution justified for your environment?

## Architecture

- Data types
- Backup methods
- Server and network topology

## Technology

- Strategic roadmap
- Degree of HA, data integrity, recoverability
- Function, reporting, and “performance at scale”

## Integration

- Implementation
- Migration
- Backup system

# Forsythe can help reduce evaluation cycles

- Forsythe Expertise and Services:
  - Current requirements and environment analysis
  - Solution justification and design
  - Onsite POC coordination
  - Forsythe TEC Lab testing of function and payload limits
  - Backup software installation/ configuration/ testing/integration
  - Technology and manufacturer scorecards

# Forsythe can help you integrate the technology

- Complete solution integration
  - ✓ Supply, financing, and integration
  - ✓ Backup software
  - ✓ Disk target
  - ✓ Tape target
  - ✓ Replication
  - ✓ WAN
  - ✓ LAN
  - ✓ Training
  - ✓ Management and Monitoring



**FORSYTHE**

*Realize the business value of IT<sup>®</sup>*