Reduce Your Data Footprint

Manage More Data with Less Infrastructure

Pulse2010
The Premier Service Management Event

Vishal Maheshwari

Sales Leader – Tivoli Storage ISA

Session abstract

One of the primary IT-related solutions that companies are investing in today is any technology that will help them survive the tidal wave of data growth, especially those solutions that help reduce the overall data storage footprint. IBM is uniquely positioned to help our customers meet this challenge with a holistic approach to data reduction that addresses the major cause of data proliferation, as well as providing meaningful solutions that optimize storage. These solutions help to reduce capital and administrative costs, while improving service levels. This session will review an effective, four-step process to reduce your data storage footprint, and will cover techniques such as incremental backup, data categorization, space management, compression and data deduplication.

got data?

too much

**And not enough (blank) to manage it all?**

Time   Money   People   Floor Space   Electricity   Air Conditioning

Agenda

- Are you drowning in a tidal wave of data?
- Five approaches reducing your data footprint
- Why IBM?
- Next steps
- More information

The tidal wave continues …

- The amount of digital information continues to grow exponentially …
- And we need to keep more of it, longer …
- And the costs of losing data are increasingly unacceptable
  - Lost revenues
  - Lost customer confidence
  - Embarrassment in the market
  - Fines from contracts, government agencies
  - CEO and CFO could go to jail

We Need to do More with Less, and we need to do it smarter

2005   2006   2007   2008   2009   2010
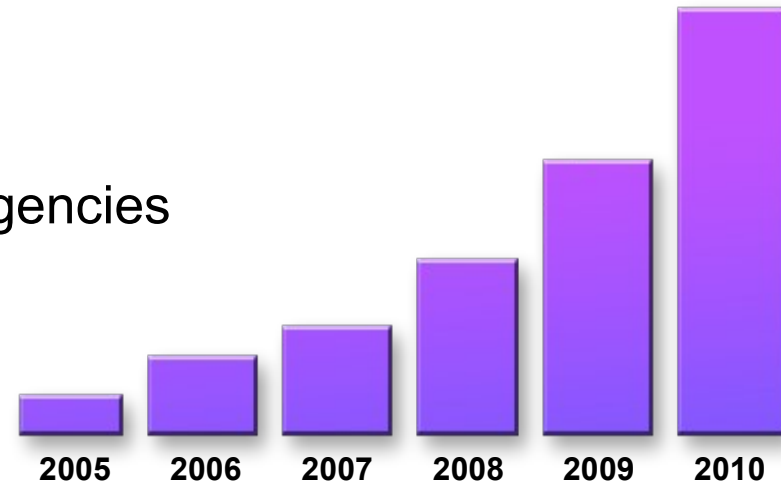
Data created and copied is expected to grow at 48% CAGR through 2010
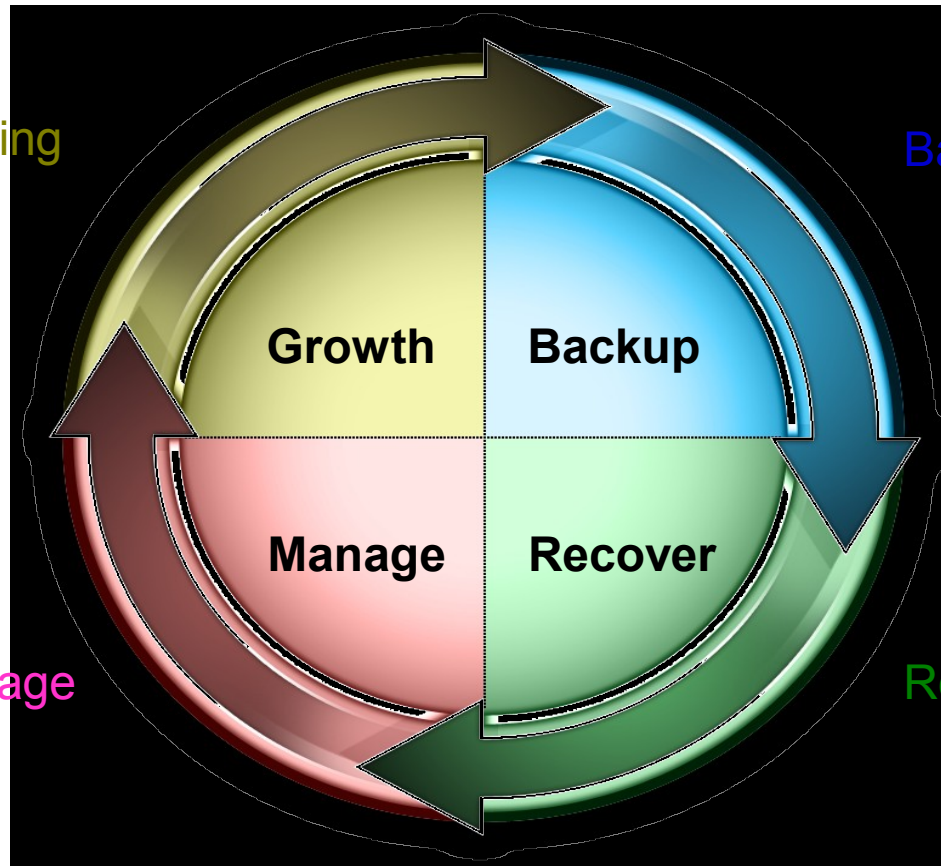
Source: Various external consultant reports

The pressures on administrators are growing

The consequences of data growth:

More new data coming

Backup takes longer

Can't buy more storage

Recovery takes longer

Growth   **Backup**

Manage   **Recover**

Surviving the tidal wave

Reducing your data storage footprint will:

- Reduce your costs
  - Less storage = less capital expenditures
  - Less data = simplified management and administration

- Improve service levels
  - Less downtime = higher application availability
  - Meet SLAs and customer expectations

- Mitigate risks
  - Eliminate consequences of data loss
 - Respond faster to events and legal/government inquiries

IMPROVE SERVICE
REDUCE COST
MANAGE RISK

Dynamic Infrastructure
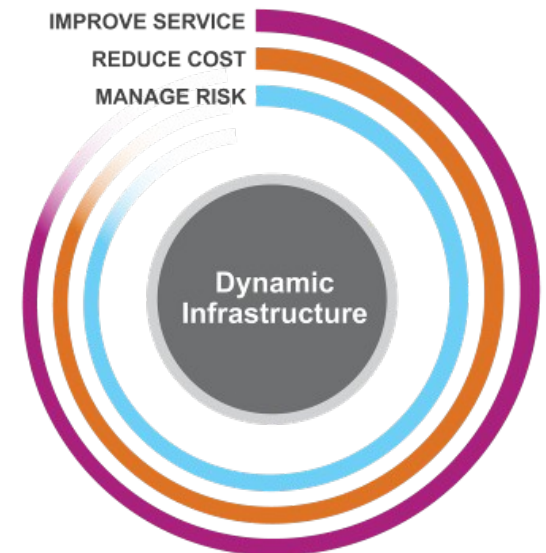
IBM can help you build a dynamic storage infrastructure that will intelligently improve service levels, reduce costs and manage risks

Surviving the tidal wave

Choices for reducing your data storage footprint:

- Discover & categorize your data
- Automate data lifecycle management
- Avoid data duplication
- Compress and deduplicate
- Maximize storage utilization
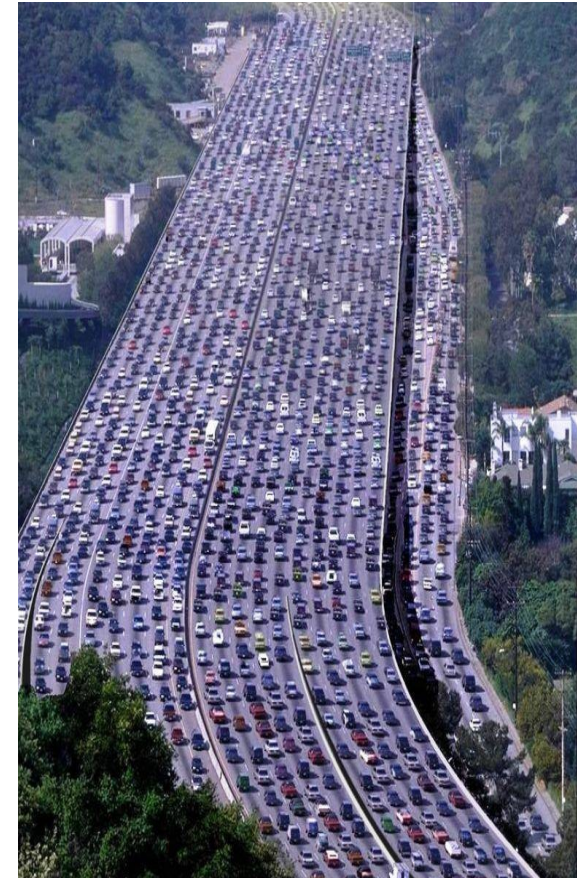
Discover and categorize

Determine what you have before you try to fix it

- Systems are bursting w/ data that is old and rarely used

- Is some of your data a liability?
  - Think e-discovery: do you know what was saved 5 years ago?

- Categorize & then migrate or archive this data from production systems to:
  - Reduce capacity requirements and lower CAPEX and OPEX
  - Improve backup and restore performance
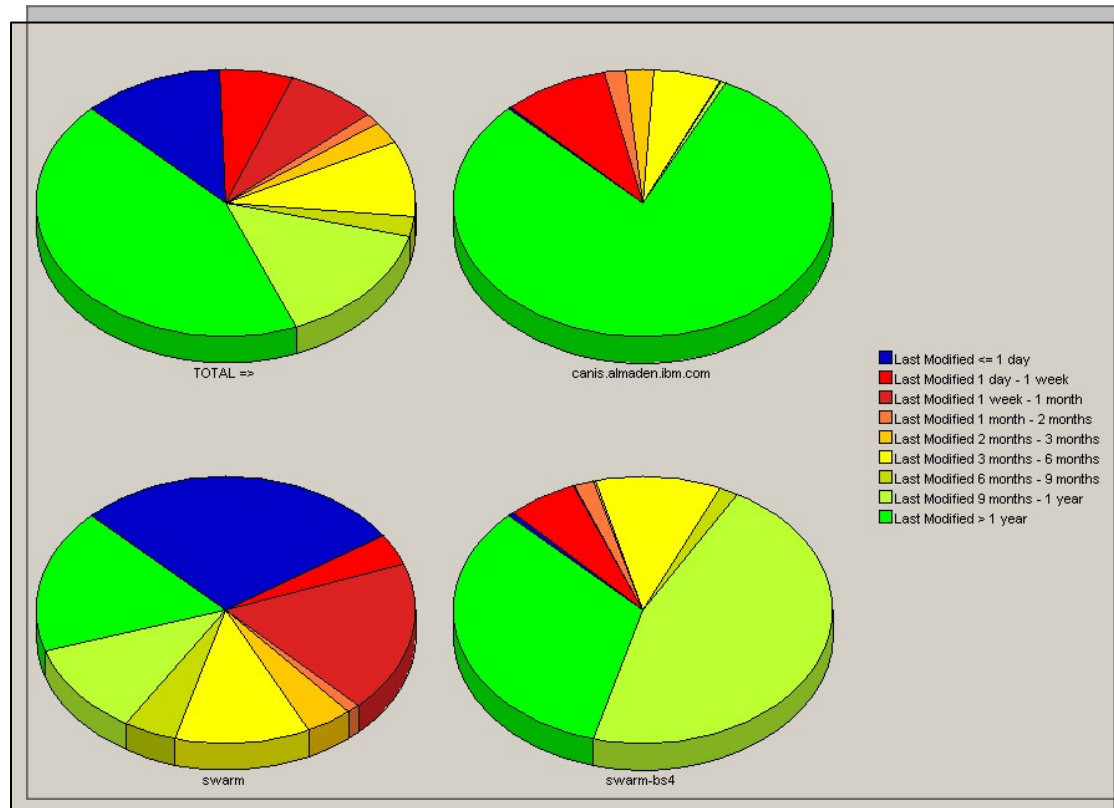 - Help meet data retention and expiration mandates

Data discovery and categorization

IBM Tivoli Storage Productivity Center for Data

Identifies data eligible for migration, archiving and deletion:

Provides:

- Date saved or last accessed

- Location and owner
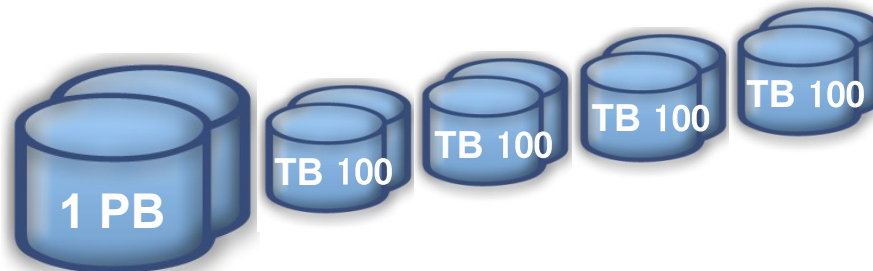
- File type and size

- Duplicate files

Automate Data Lifecycle Management

Data lifecycle management using tiered storage

Start with 1PB, add 100TB new data per quarter (40% CAGR)

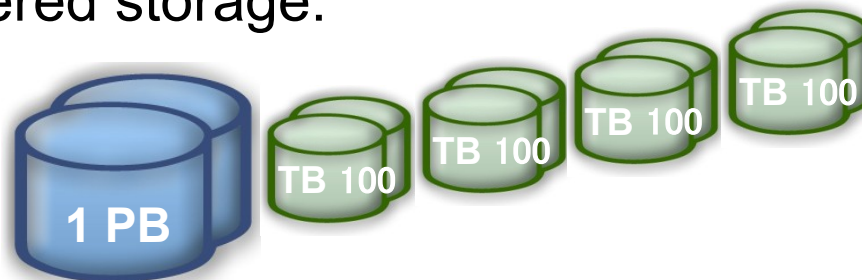- Storage growth w/o tiered storage:

Add new primary capacity each qtr.

NEW data grows into new capacity ⟶
At $50,000 per TB, new storage costs $20 million*

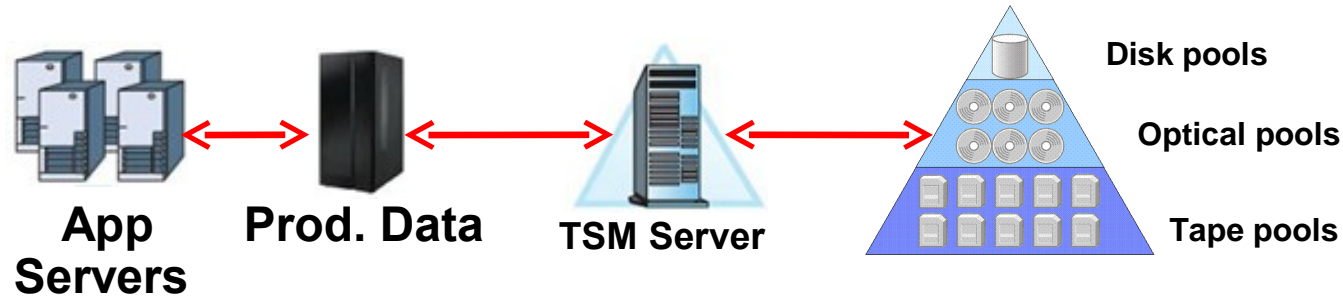## Storage growth <u>with</u> tiered storage:

Add new secondary capacity each qtr.

OLD data migrates into new capacity ⟶
At $15,000 per TB, new storage costs $6 million*

* fully-burdened cost

Data lifecycle management

App Servers → **Prod. Data** → **TSM Server** → Disk pools / Optical pools / Tape pools

- Regularly scrub production systems of old/stale data
  - Automated processes based on business-driven policies
- Reduce the growth of primary storage – reduce CAPEX
  - Faster backup / restore
- Migration / Hierarchical Storage Management (HSM):
  - Leaves a pointer; enables transparent access to migrated data
- Archive
  - Completely removes files from production systems
  - Supports data retention and version control policies

Automating data migration

IBM Tivoli Storage Manager for Space Management
IBM Tivoli Storage Manager HSM for Windows

- Get control of and efficiently manage data growth and its associated storage costs

- Storage pool "virtualization"

- Optimized restore management
  - Based on location of data in hierarchy

- Transparent to the users and applications
  - Simple pointer (stub file) replaces data in original location
  - Fast, direct restore from disk to client

- Migrations
  - Scheduled, automated, outside the backup window

Archive

Features

- Long-term storage on cost-effective media
- Point in time copy; revision history and auditability
- Retention period and 'retention hold' enforcement
- Fast expiration processing
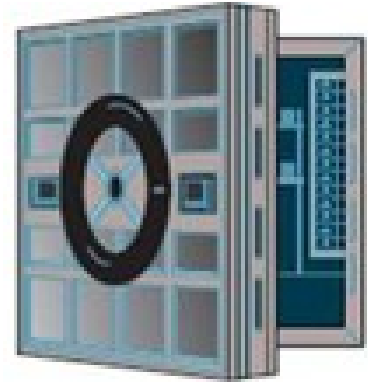
Benefits

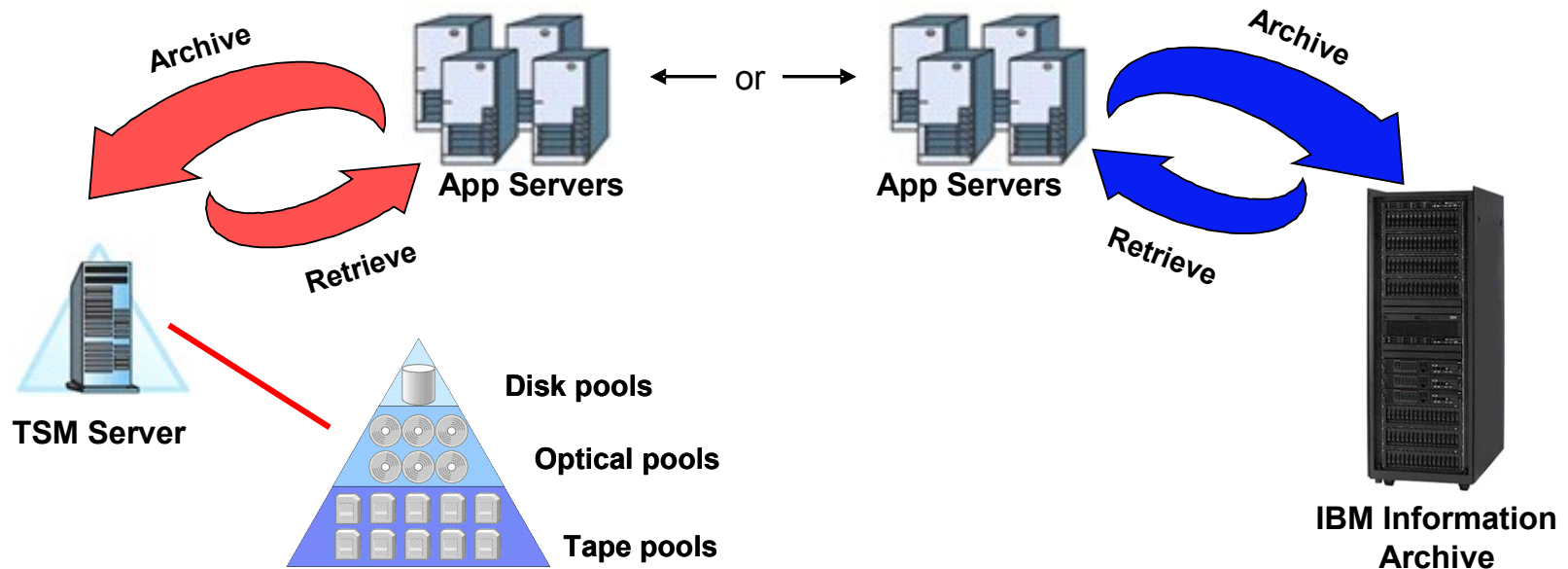Speed file-server recovery times – recover only active data

Reduce backup times and resource usage

Move archived files to a hierarchy of lower-cost storage

Archived files are indexed with descriptive metadata to aid in locating historical information

IBM Smart Archiving solutions

Tivoli Storage Manager 6

- Integrated Backup, HSM and Archive solution
- Leverage the same hierarchy of storage

IBM Information Archive

Dedicated, scalable archive appliance

Supported by more than 40 Apps

Avoid data duplication

Avoiding data duplication

- Performing periodic full backups is typically the largest contributor to data growth in a data center
- As much as 95% of your data doesn't change from week-to-week
- Are you making another copy of that data every weekend?
- Data deduplication solutions were created to address this problem
  - *They claim 95% reduction ratios, this is the data they're talking about*

Never perform a full backup again

- **Tivoli Storage Manager** – 'progressive-incremental' and sub-file backup
- **Tivoli Storage Manager FastBack** – block level incremental
- **Tivoli Storage Manager FastBack for Workstations**
- **Tivoli Continuous Data Protection for Files**

Data reduction: progressive-incremental backup

Features

- ONLY new or changed files are backed up
- Synthetic FULL created in the background
- Restores don't require the same file to be restored multiple times
- Supports multiple versions of files
- Accurate point-in-time restores

Benefits

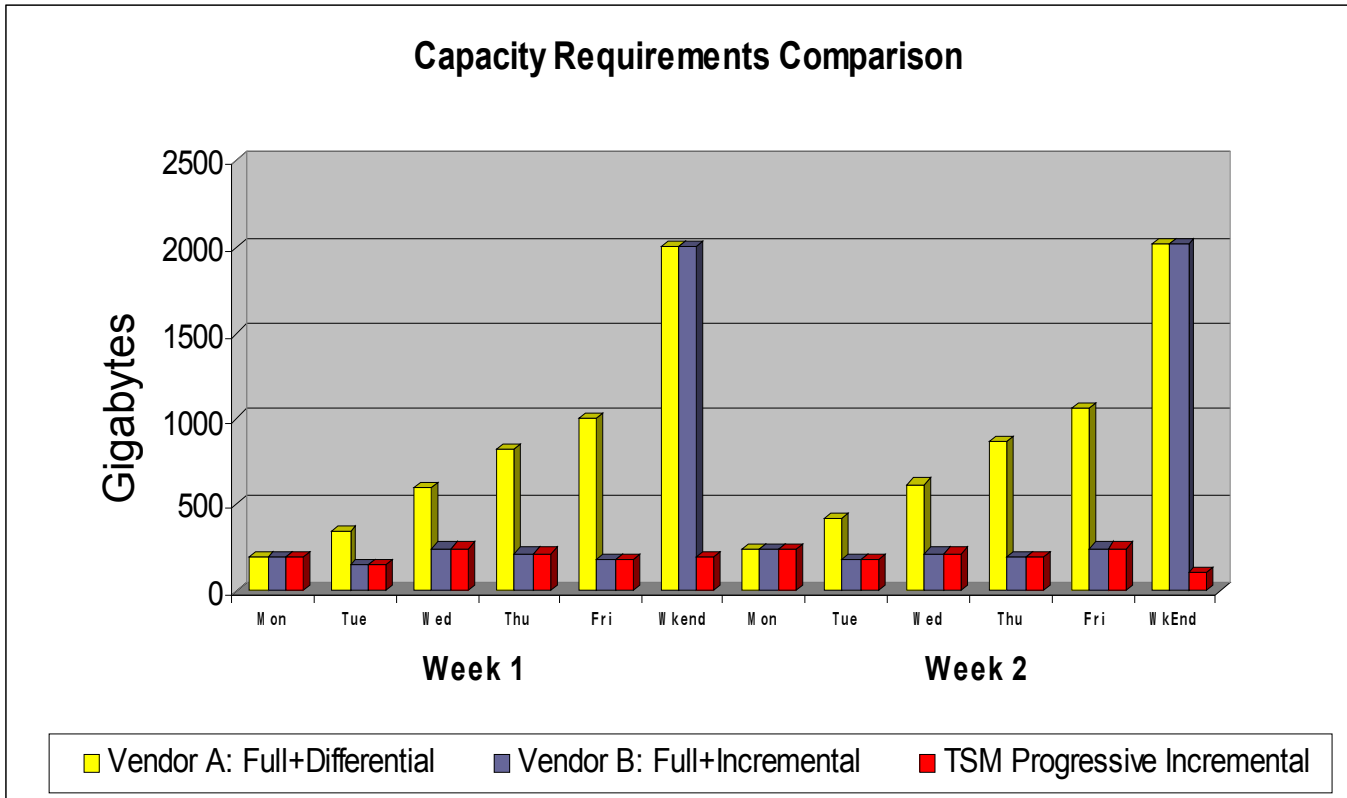Requires less storage space, less network bandwidth and less time

Shorter backup windows

Fast accurate restores

Benefits of progressive-incremental backup

Never perform a full backup again

Capacity Requirements Comparison

Backup Capacity Needed for 1 Month:

Vendor A: 26TB

Vendor B: 14TB

IBM TSM: 7TB

Legend:
- Vendor A: Full+Differential
- Vendor B: Full+Incremental
- TSM Progressive Incremental

Assumes: Full backup completed, 2TB data to start, 26% annual growth rate, 10% new/changed data per day

Compress and deduplicate

Data compression in Tivoli Storage Manager

- Selectable option – on average, yields 2:1
- Sub-file backup when only parts of a file change

    – Byte-Level: for smaller files

    – Block-Level: for larger files

    – File-Level: if more than 60% of the file has changed, TSM backs up the whole file

- Tape Reclamation – increase tape utilization
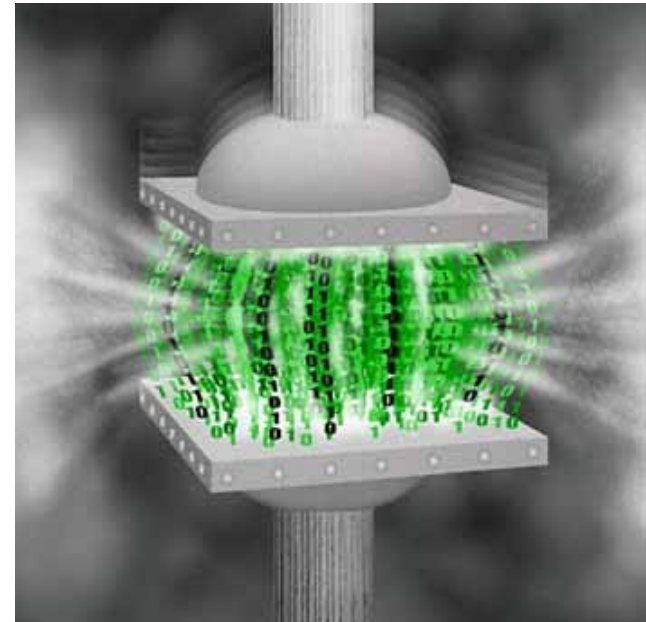
"Tivoli Storage Manager has been long recognized as having the best tape management capabilities on the market. These features are fully integrated into the product at no additional charge." **– Dave Russell, Gartner Inc.**

Tape reclamation

100% → 25 % full + 70 % full = 95 % full
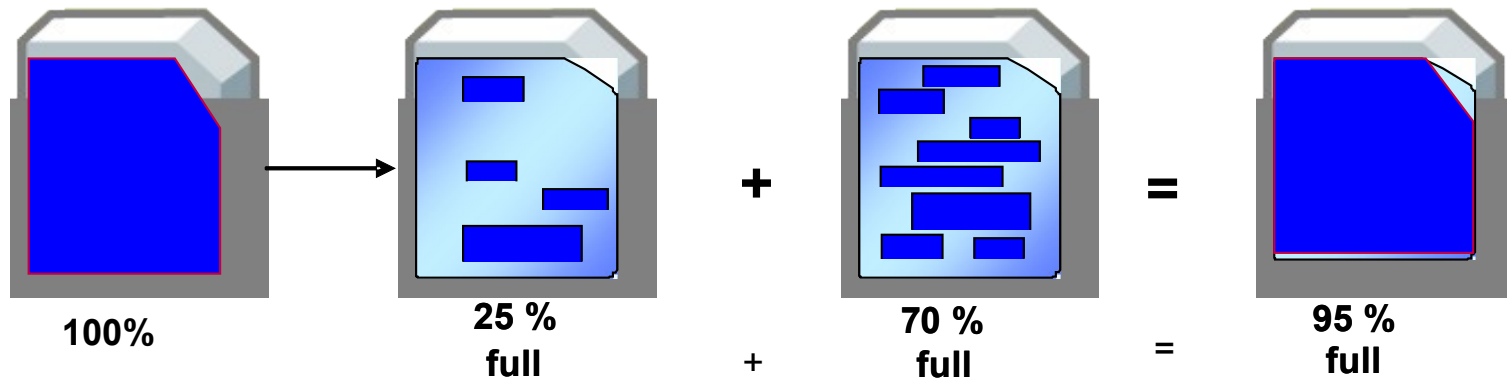
- Better utilizes tapes, thus, saving money
- Tape utilization constantly monitored
- User-defined reclamation threshold
- Can be scheduled to occur at specified times

When free space reaches a set threshold:
- Tape is mounted
- Valid data is moved to another tape
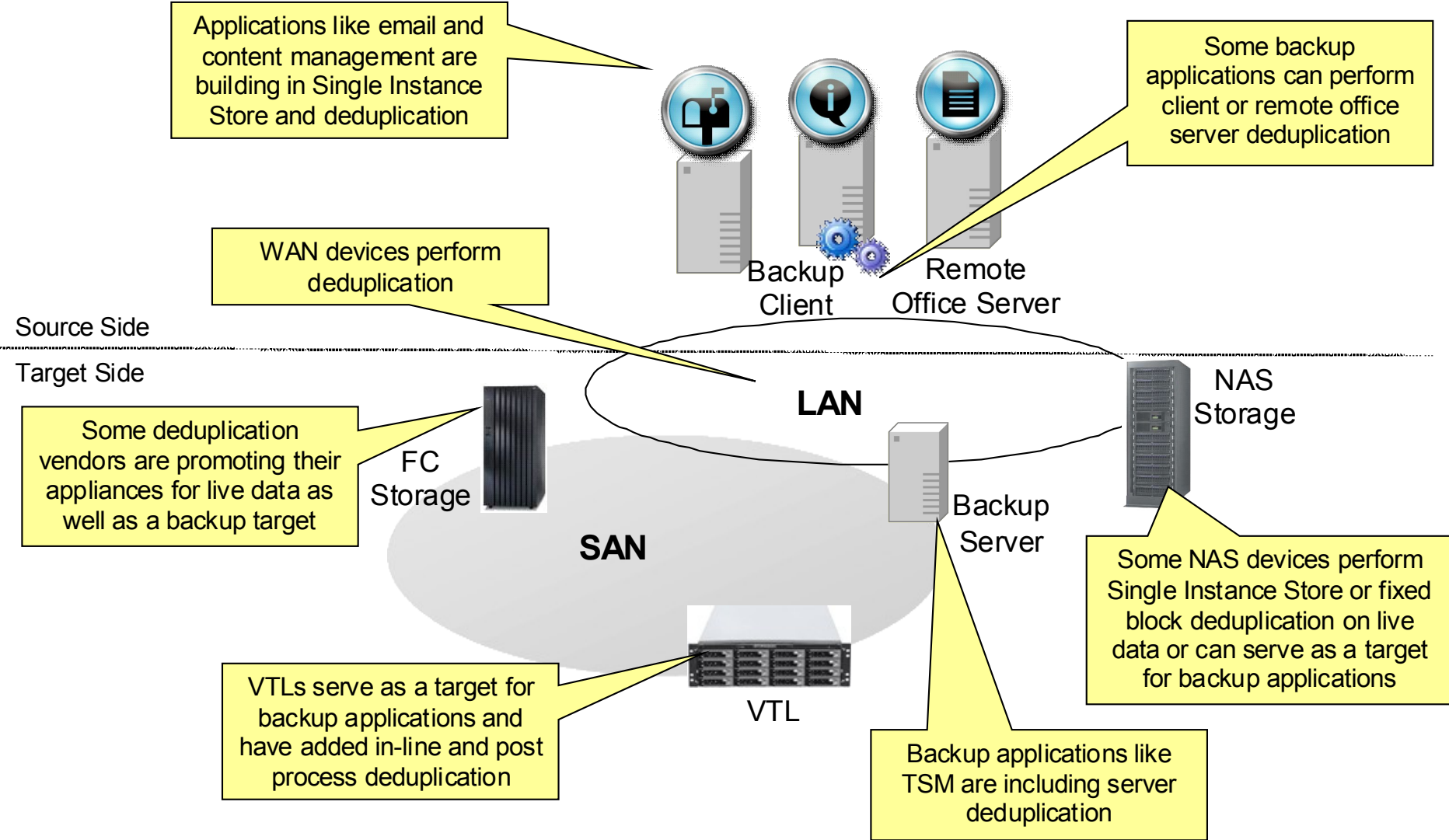- Original tape is returned to the scratch pool

Basics of data deduplication

- A 'hot' data reduction technology
- Eliminates redundant subfiles
  - Known as chunks, blocks, or extents
- Only one instance is stored for each common chunk
- Duplicate instances of the chunk point to the stored chunk

1. Data chunks are evaluated to determine a unique signature for each

2. Signature values are compared to identify all duplicates

3. Duplicate data chunks are replaced with pointers to a single stored chunk, saving storage space
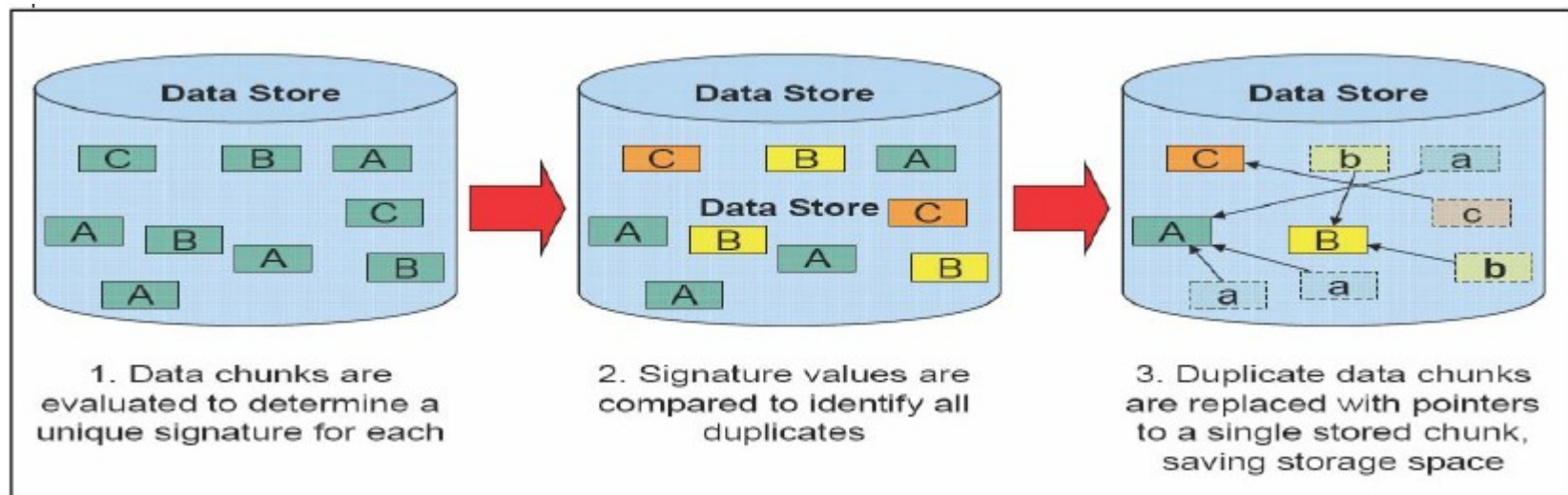
Where can data deduplication occur?

Applications like email and content management are building in Single Instance Store and deduplication

Some backup applications can perform client or remote office server deduplication

WAN devices perform deduplication

Backup Client

Remote Office Server

Source Side

Target Side

LAN

NAS Storage

Some deduplication vendors are promoting their appliances for live data as well as a backup target

FC Storage

SAN

Backup Server

Some NAS devices perform Single Instance Store or fixed block deduplication on live data or can serve as a target for backup applications

VTLs serve as a target for backup applications and have added in-line and post process deduplication

VTL

Backup applications like TSM are including server deduplication

Deduplication in Tivoli Storage Manager v6.2

- **Tivoli Storage Manager v6.1** and **TSM FastBack v6.1** include target-side data deduplication, *at no extra charge*
  - Improves recovery times and/or reduces capacity requirements
  - Uses data from any source including: API, backup, HSM, archive
  - Operates as a post-process; automatic space reclamation
  - Builds on automatic data compression & progressive-incremental

- **New in TSM v6.2**: client-side data deduplication
  - Reduces network traffic by determining if a chunk has already been backed up (maybe from a different client system)

"The combination of TSM progressive incremental backups and target-side data deduplication reduced disk capacity by a factor of 19:1 after just 10 backups"
  – *Tony Palmer, ESG Lab Validation Report, April 2009*

Tivoli.
Storage Manager 6

Maximize storage utilization

Gain visibility and control of storage resources

Tivoli Storage Productivity Center – Standard Edition

- Enable end-to-end storage management
  - A holistic topology view of entire storage infrastructure
  - Centralized management of heterogeneous storage

- Improve storage utilization, performance and service levels
  - Analytics, trending, change configuration, best practices guidance
  - Green features
  - High availability; transparency for Cloud deployments

- Reduce storage complexity
  - Built-in, customizable operational control and automation
  - Deep-dive reporting across host file systems, databases and virtualized storage environments

Maximize utilization and simplify management
IBM System Storage SAN Volume Controller (SVC)
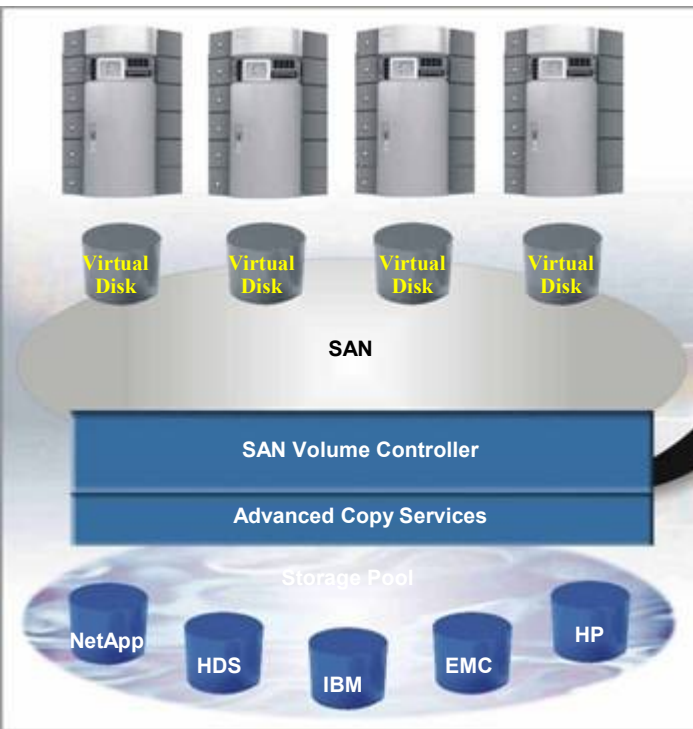
- Virtualization of total storage capacity across vendors
  - Eliminates storage silos
- Thin provisioning - auto-assigns capacity
- Non-disruptive data migration & data movement
- Greater efficiency and productivity
- Industry-leading virtualization software
  - Scales with your growth and minimizes costs

Why IBM

IBM's unique position in the industry

- IBM is the only vendor with a comprehensive set of data reduction technologies

- IBM does not force any particular approach
  - Our broad portfolio gives us the freedom to solve customer issues with the most effective and appropriate technology

- IBM's high quality global support services will ensure your investment in data reduction will meet your needs

- IBM is continuing to invest in research to deliver the advanced data reduction features our customers are requesting

IBM's comprehensive data reduction portfolio
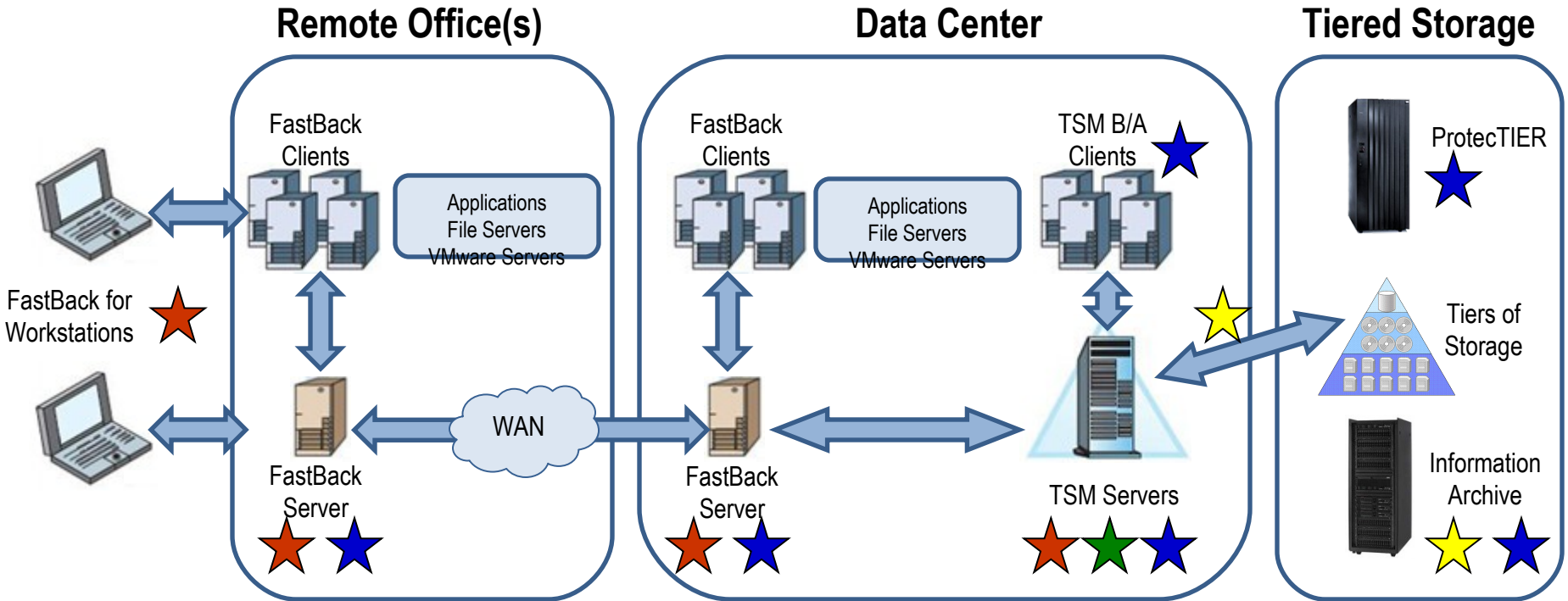
Remote Office(s)   **Data Center**   **Tiered Storage**

FastBack Clients

Applications
File Servers
VMware Servers

FastBack for Workstations ★

FastBack Server ★ ★

WAN

FastBack Clients

Applications
File Servers
VMware Servers

FastBack Server ★ ★

TSM B/A Clients ★

TSM Servers ★ ★ ★

ProtecTIER ★

Tiers of Storage

Information Archive ★ ★

★ Incremental-only backup

★ Data compression and tape management

★ Data deduplication

★ Space management and archive

Plus:
• TPC for Data categorization/deletion
• TPC-SE for storage resource management
• SVC for storage virtualization
• Optim for database archiving
• … and many others

IBM's comprehensive data reduction portfolio
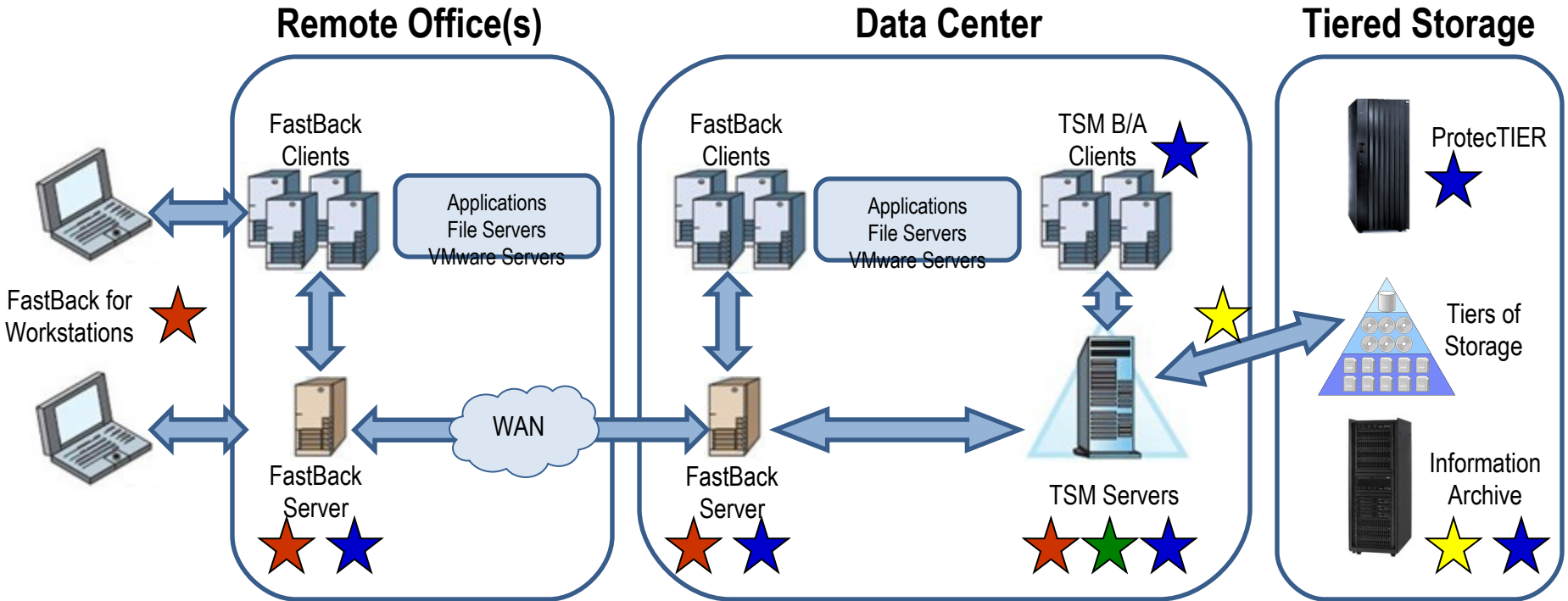
Remote Office(s) | Data Center | Tiered Storage

★ Incremental-only backup

★ Data compression and tape management

★ Data deduplication

★ Space management and archive

"IBM delivers data reduction in more places than any other vendor", analyst *Mike Karp with Ptak, Noel & Associates, 02/10*

Next steps

Next steps

- IBM Global Technology Services and IBM Business Partners stand ready to help you assess your current situation and recommend next steps.

- We can help you determine which data reduction techniques will have the most cost-effective impact on your operations.

- Ask for a comprehensive ROI analysis using IBM's comprehensive Business Value Analyst (BVA) tool.

THANK YOU

Trademarks and disclaimers