Reducing Your Data Footprint

Richard Vining, Tivoli Storage Product Marketing

PulseANZ 2010

Meet the people who can help advance your infrastructure

© 2010 IBM Corporation

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.

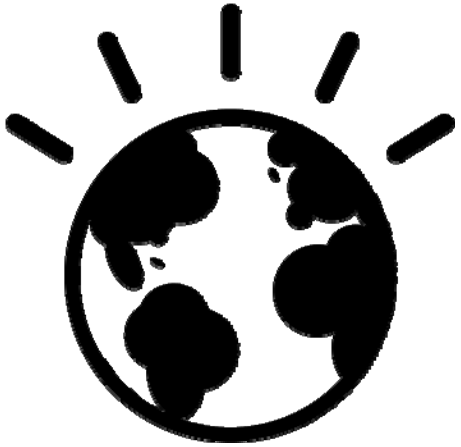
Speaker profile

Mr. Richard Vining is the Product Marketing Manager responsible for the IBM Tivoli Storage Manager portfolio of products.

Mr. Vining joined IBM in April 2008 as part of the acquisition of FilesX, where he served as Director of Marketing. Mr. Vining has more than 20 years of experience in the data storage industry, holding senior management roles in marketing, alliances, customer support and product management at a number of leading edge companies, including Signiant, OTG Software, Plasmon and Cygnet.

Effective Data Management for a Smarter Planet

- The planet is: smaller and flatter
- The world is connected: economically, socially and technically
- It is becoming more instrumented, interconnected and intelligent
- And it is creating an enormous amount of digital information (data)
- We need smarter ways of managing all this data; we need to do more with less



"On the Internet, things are communicating with things more than people are communicating with people" **– Al Gore, former U.S. Vice President, at PULSE 2010**

too much

got data?

And not enough (blank) to store it all?

Time Money People Floor Space Electricity Air Conditioning

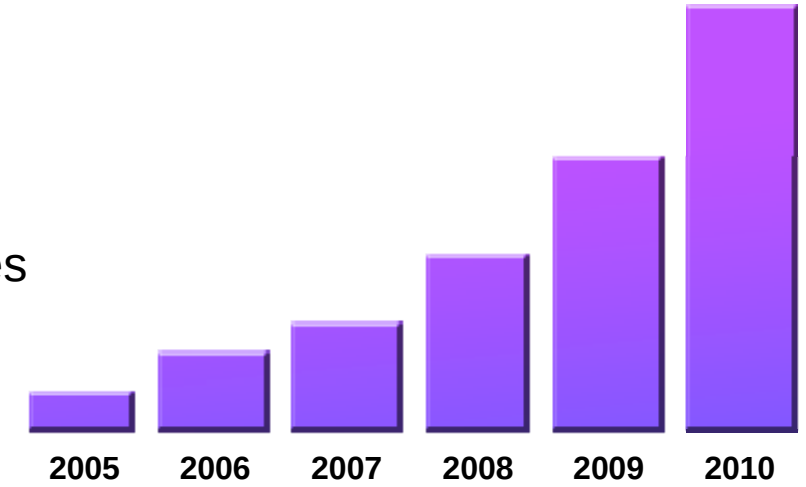
Agenda

- Are you drowning in a tidal wave of data?
- Approaches to reducing your data footprint
- Why IBM?
- 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
- But budgets are not increasing

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



Data created and copied is expected to grow at 48% CAGR through 2010

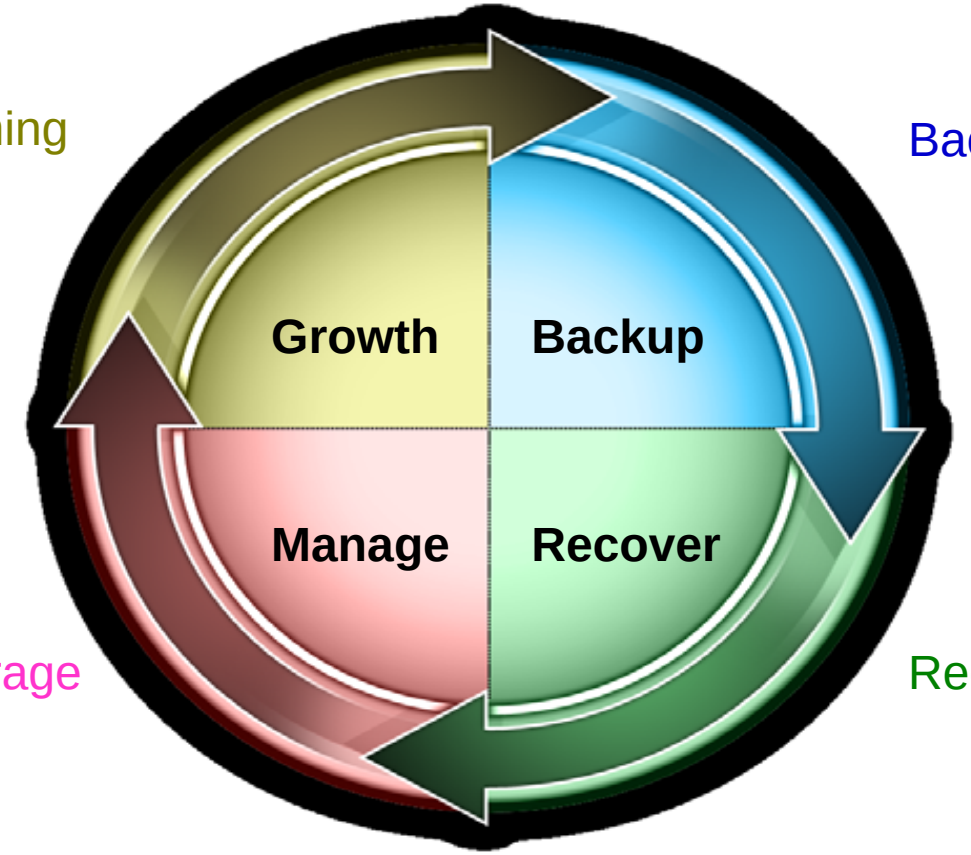
Source: Various external consultant reports

The pressures on administrators are growing

The consequences of data growth:

More new data coming

Backup takes longer

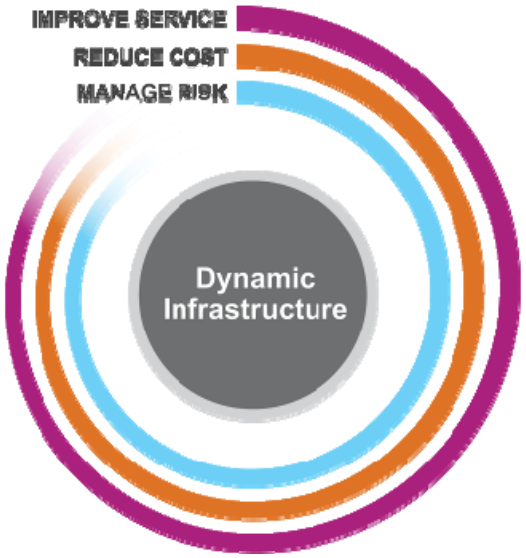


Can’t buy more storage

Recovery takes longer

Reducing your data storage footprint will:

- Improve service levels
  - Less downtime = higher application availability
  - Improved competitiveness and customer satisfaction
- Reduce your costs
  - Less storage = less capital expenditures
  - Less data = simplified management and administration
- Mitigate risks
  - Eliminate consequences of data loss
  - Respond faster to events and legal/government inquiries



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

Choices for reducing your data storage footprint

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

1. Discover and Categorize

PulseANZ 2010

Meet the people who can help advance your infrastructure

© 2010 IBM Corporation

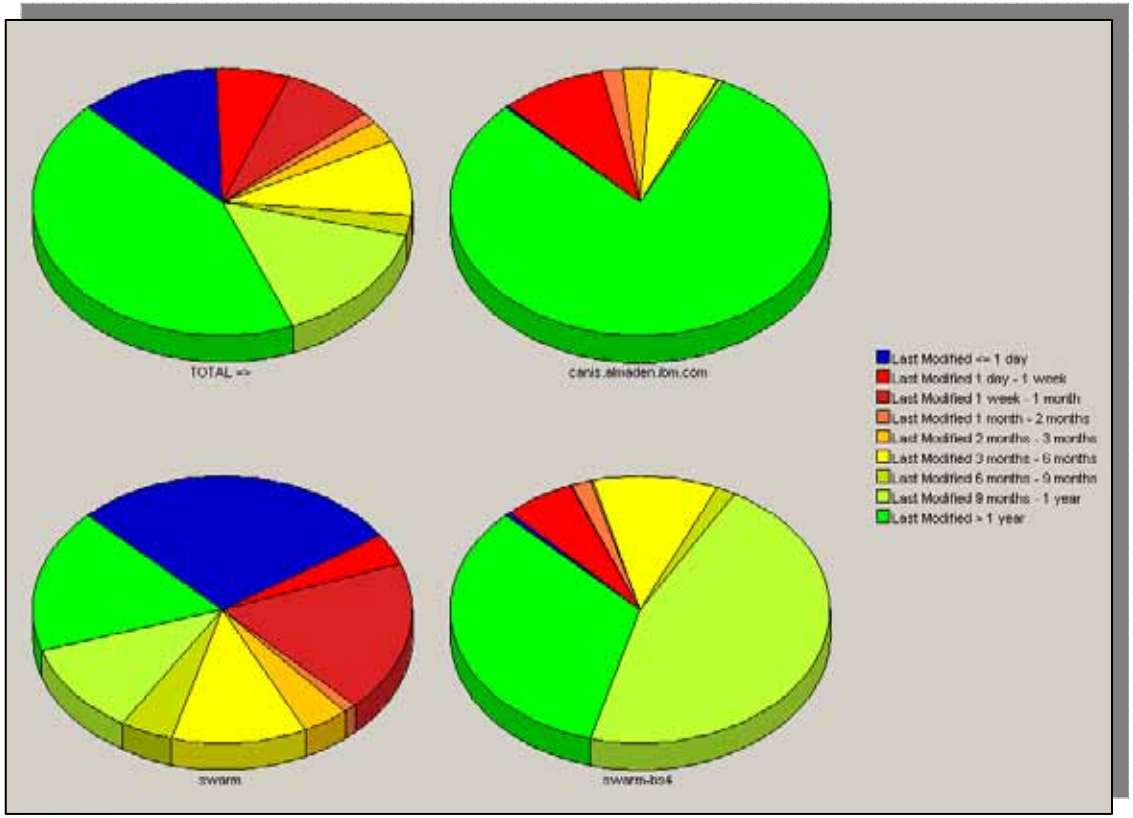
Determine what you have before you try to fix it

- 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?
  - Do you have policies for data deletion?
- Categorize & then migrate, archive or delete 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 reports on:
  - Date saved or last accessed
  - Location and owner
  - File type and size
  - Duplicate files



2. Automate Data Lifecycle Management

PulseANZ 2010

Meet the people who can help advance your infrastructure

© 2010 IBM Corporation

Data lifecycle management using tiered storage

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

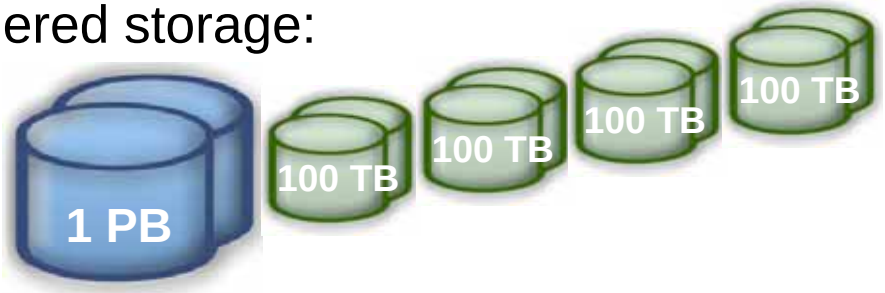
- Storage growth w/o tiered storage:

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:

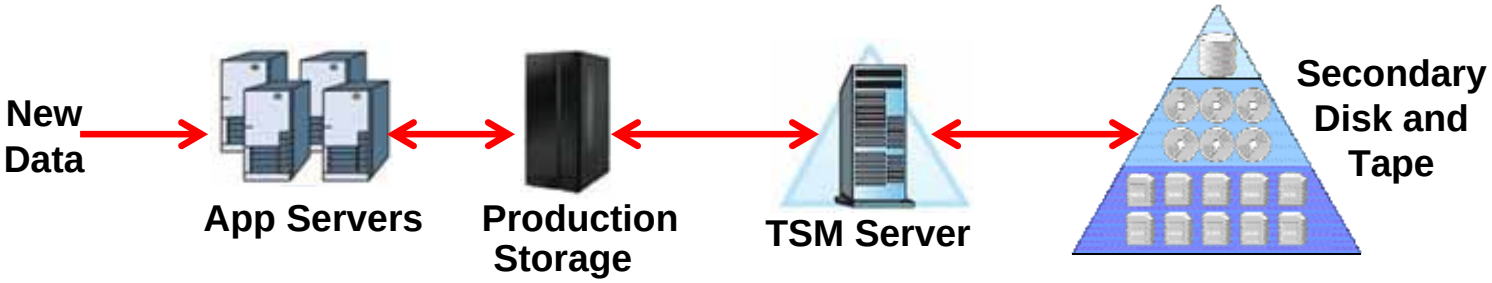
Add new secondary capacity each qtr.



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

* fully-burdened cost

Data lifecycle management



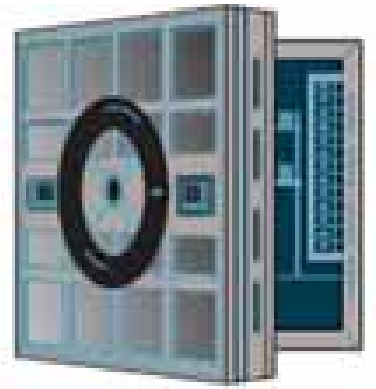
- Regularly scrub production systems of old/stale data
  - Automated processes based on business-driven policies
- Reduce the growth of primary storage – reduce CAPEX
  - Faster backup / restore
- Migration / Hierarchical Storage Management (HSM):
  - Leaves a pointer; enables transparent access to migrated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

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



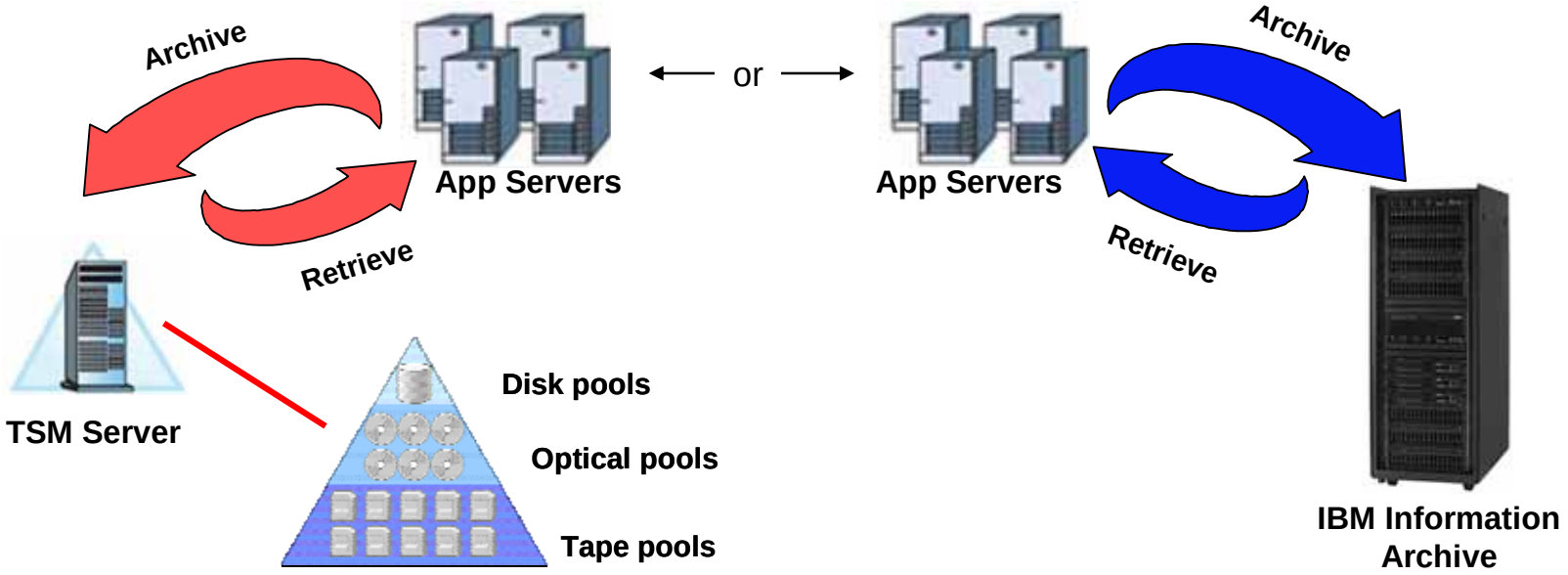
Features

- 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

3. Avoid Data Duplication

PulseANZ2010

Meet the people who can help advance your infrastructure

© 2010 IBM Corporation

Treat the cause, not the symptom

- 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
  - ***When 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** – continuous incremental
- **Tivoli Storage FlashCopy Manager** – incremental snapshot backups

Data reduction: progressive-incremental backup

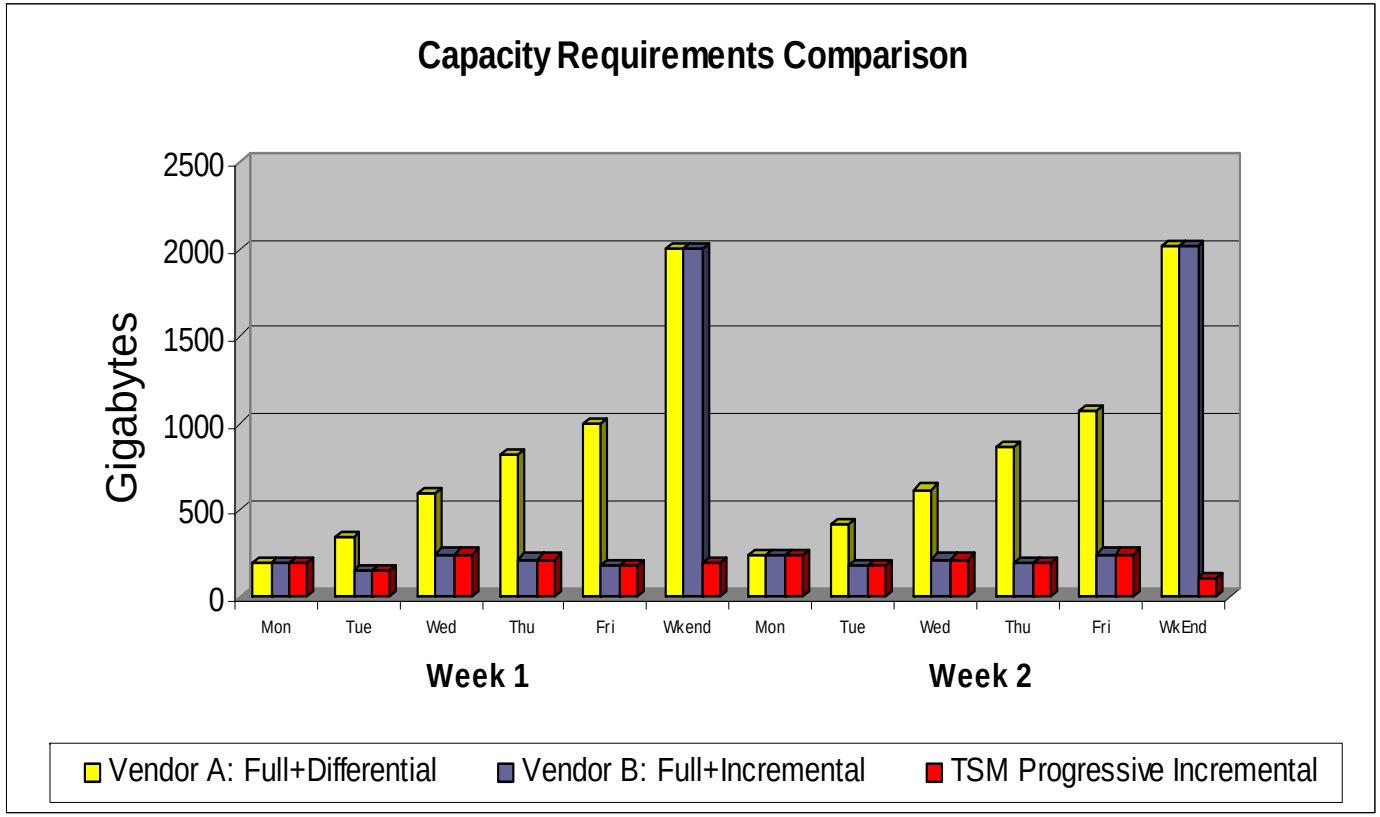
Features

- ONLY new or changed files are backed up
- Restores don't require the same file to be restored multiple times
- Data is tracked at a file level
- Supports multiple versions of files
- Accurate restores

Benefits

- Requires less storage space, less network bandwidth and less time
- Shorter backup windows
- Fast accurate restores

Benefits of progressive-incremental backup



Backup Capacity Needed for 1 Month:

Vendor A: 26TB

Vendor B: 14TB

IBM TSM: 7TB

Assumes: Full backup completed, 2TB data to start, 26% annual growth rate, 10% new/changed data per day
At $15,729/TB (burdened cost), TSM saves this customer **$409,954** per month in storage costs over Vendor A

4. Compress and Deduplicate

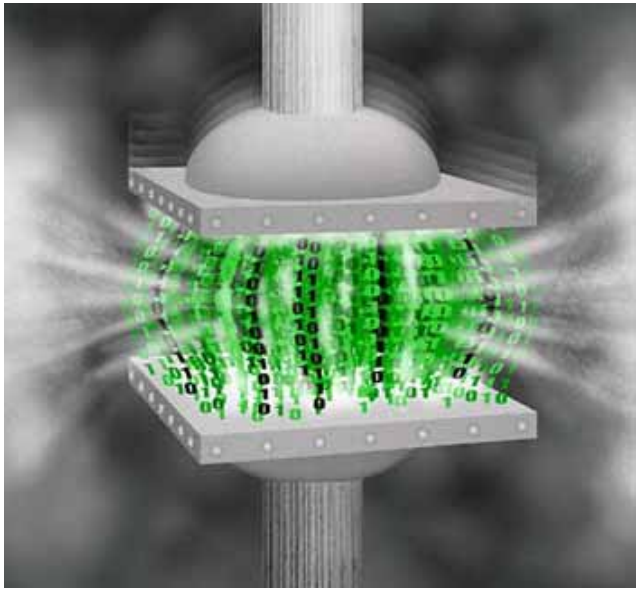
PulseANZ 2010

Meet the people who can help advance your infrastructure

© 2010 IBM Corporation

Data compression in Tivoli Storage Manager

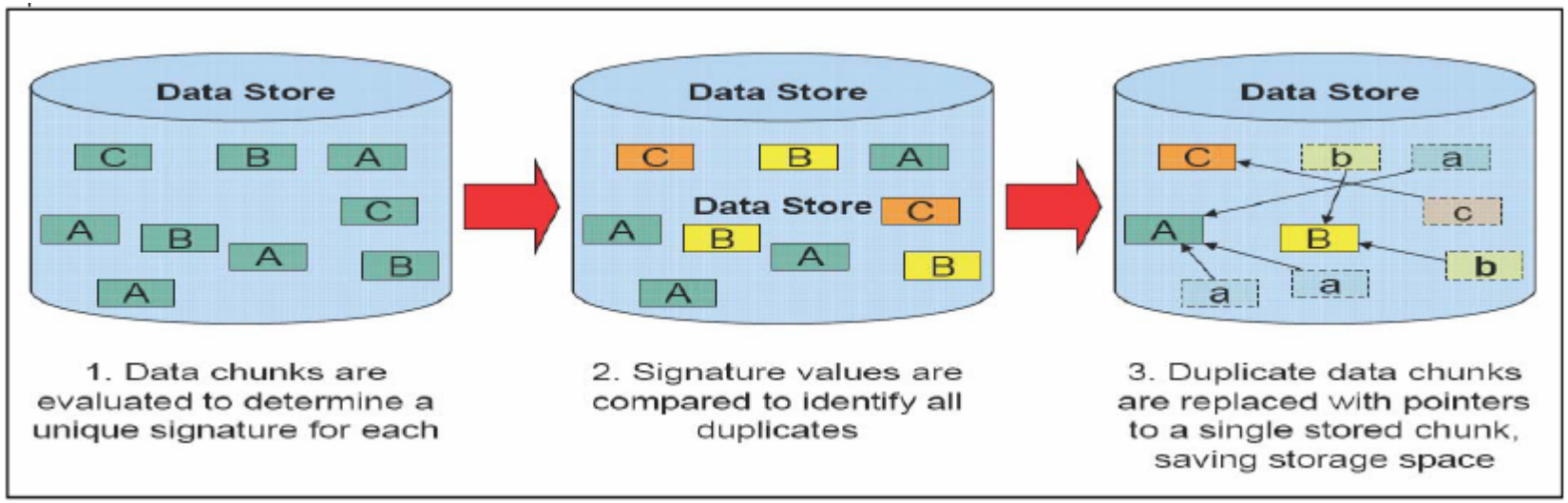
- Selectable option – on average, yields 2:1
- Sub-file backup when only parts of a file change
  - Byte-Level: for smaller files
  - Block-Level: for larger files
  - File-Level: if more than 60% of the file has changed, TSM backs up the whole file
- Tape Reclamation – increase tape utilization
- Result: lower storage, bandwidth and management costs



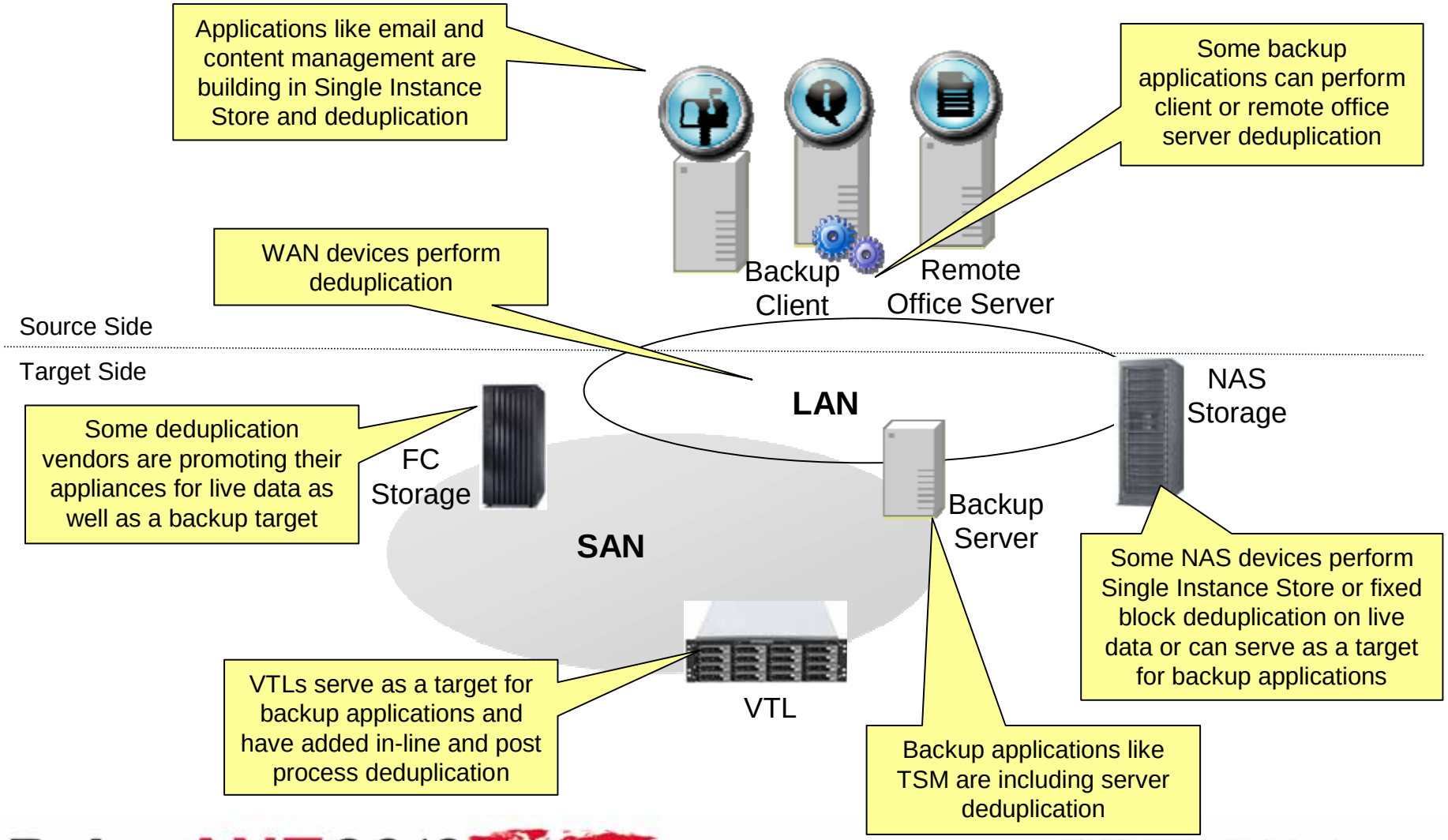
"Tivoli Storage Manager has been long recognized as having the best tape management capabilities on the market. These features are fully integrated into the product at no additional charge." **– Dave Russell, Gartner Inc.**

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



Where can data deduplication occur?



Deduplication in Tivoli Storage Manager

- **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

Deduplication with ProtecTIER Virtual Tape

- **Performance**
  - Fastest VTL in real customer environments
- **Scalable**
  - Easily scales in performance (1000MB/sec) AND capacity (1PB)
- **Data Integrity**
  - Doesn't rely on a hash algorithm; performs a byte level differential to ensure data is a duplicate for enterprise class data integrity
- **Reliable**
  - ProtecTIER features all IBM best of breed components
- **Production Proven**
  - There is more capacity (>25PBs) deployed behind ProtecTIER servers in production than any other vendor in the world

Effective data deduplication – you have a choice

- Use the deduplication capabilities built into **IBM Tivoli Storage Manager 6** and **TSM FastBack** when:
  - You have a single TSM server
  - You want to improve TSM recovery times by storing more backup data on disk
  - You want to reduce the amount of your disk-based backup capacity

- Use the **IBM TS7650 ProtecTIER** virtual tape solutions when:
  - You have multiple TSM servers
  - You have other sources of backup and archive data
  - You are using other (non-IBM) backup products that perform periodic full backups

Industry Canada

Expanding data protection to reduce risk

Business challenge:

Industry Canada's mission is to foster a growing, competitive, knowledge-based Canadian economy. In 1996, the organization's IT staff implemented IBM® Tivoli® Storage Manager to protect critical SAP-based financial data. Today, Tivoli Storage Manager runs more than 30,000 backup schedules on a monthly basis, backing up nearly 200 terabytes of changed data from 400 UNIX®, Windows® and Netware-based servers.

Solution:

Seeking the ability to restore user mailboxes, mailbox items, Microsoft® Active Directory individuals or groups that have been accidently deleted, Industry Canada participated in the beta test of Tivoli Storage Manager 6.1. The improved integration with the organization's Microsoft environment provides a unified recovery management platform that enhances Quality of Service (QoS) and reduces the risk associated with data loss.

Benefits:

- Reduction in backup servers, storage capacity and administration costs
- Reduces risk of data loss with expanded recovery management capabilities
- Provides seamless upgrade path to minimize business interruption
- Streamlines data recovery processes for improved productivity

"Our testing showed that the transition to Tivoli Storage Manager 6.1 will be transparent to our user community and our operational staff. This is a huge relief, enabling us to take advantage of valuable new capabilities with no business interruption."

— Ralph Zimmerling, Chief Informatics Office, Industry Canada

Solution components:

- IBM® Tivoli® Storage Manager 6
- IBM TotalStorage® 3494 Tape Library
- IBM TotalStorage 3592 Tape Drive
- IBM System Storage™ TS7650A ProtecTIER® Deduplication Appliance

Canada

TIP14080-CAEN-00

5. Maximize storage utilization

PulseANZ 2010

Meet the people who can help advance your infrastructure

© 2010 IBM Corporation

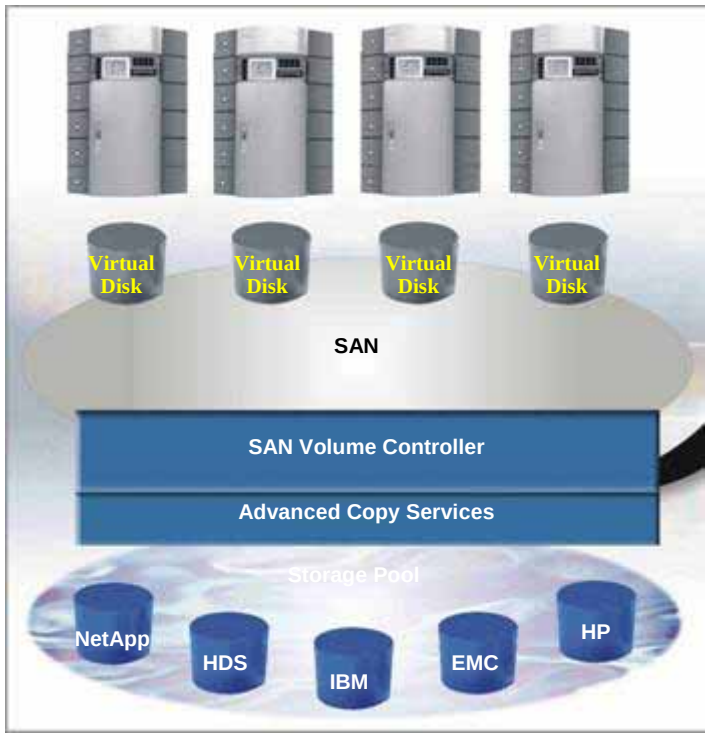
Gain visibility and control of storage resources

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

PulseANZ 2010

Meet the people who can help advance your infrastructure

© 2010 IBM Corporation

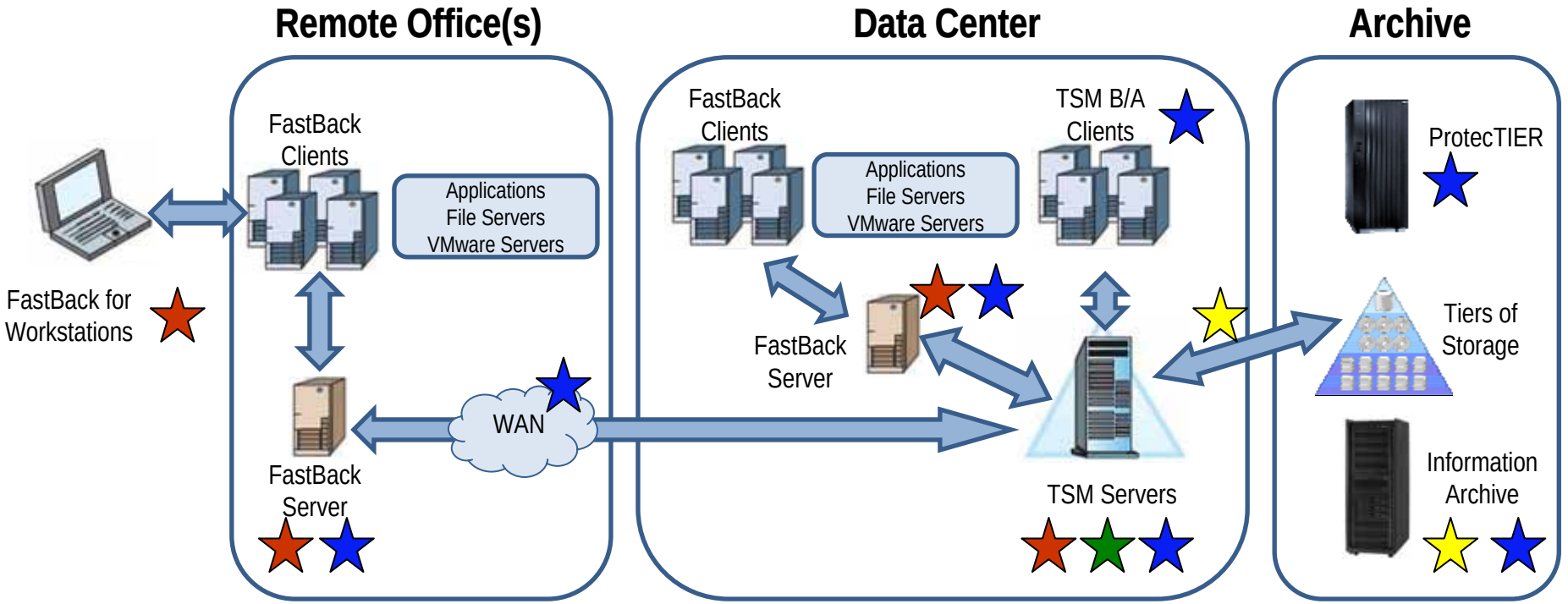
IBM’s unique position in the industry

- IBM is the only vendor with a comprehensive set of data reduction technologies
- IBM’s broad portfolio of data reduction solutions gives us the freedom to solve customer issues with the most effective technology
- IBM’s high quality global support services will ensure your investment in data reduction will meet your needs

Why IBM

- IBM has the global resources to help solve the IT and service management problems of even the largest organizations, almost anywhere they do business
- IBM is on solid financial footing and will be here to support our customers for the long term
- IBM continues to invest heavily in research and development, providing customers with product roadmaps that will continue to add value to their relationship with IBM
- IBM offers the broadest range of products, services and financing options to help customers realize significant return on their investment and faster time to value

IBM's comprehensive data reduction portfolio

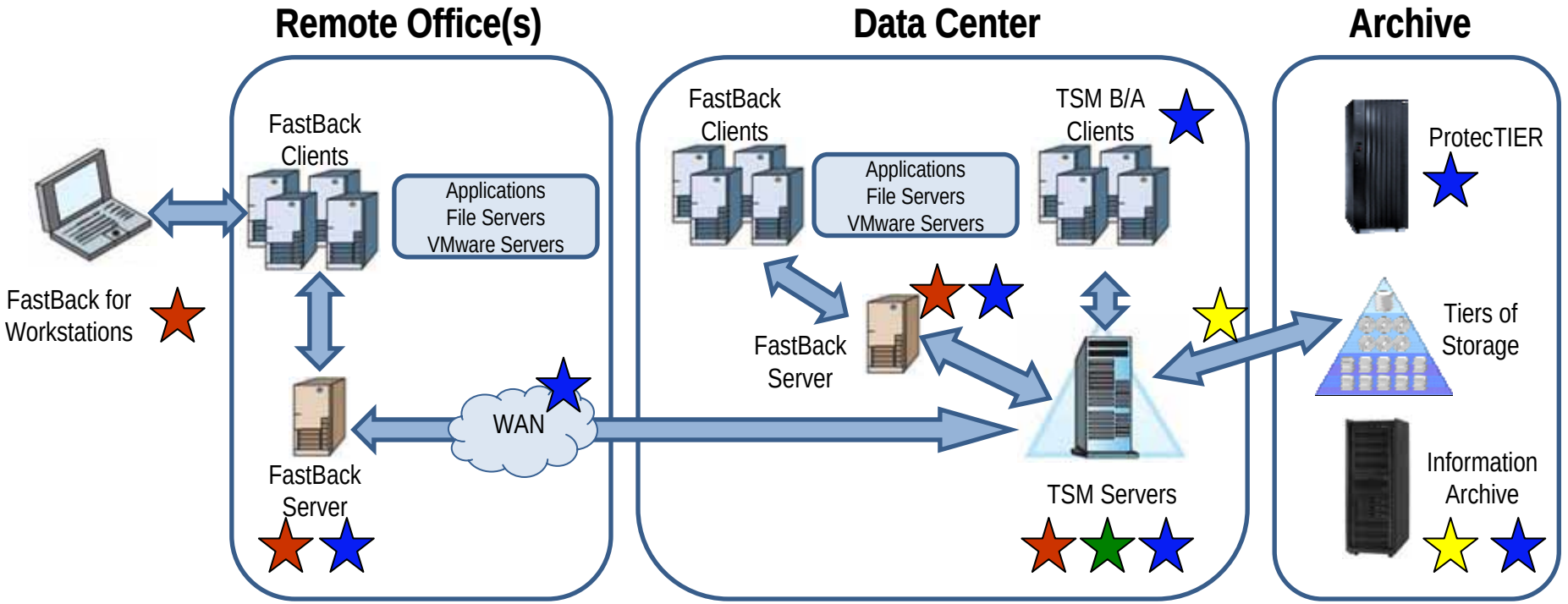


- Incremental-only backup
- Data compression and tape management
- Data deduplication
- Space management and archive

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



- 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 & Assoc.

For more information

- Data Reduction Solutions web page
- Whitepaper: “Using IBM data reduction solutions to manage more data with less infrastructure”
- Webcast: Nick Allen of the Wikibon Group reviews the options and trends for controlling data growth by reducing your data storage footprint
- Blog Series
- Products:
  - Tivoli Storage Manager 6
  - Tivoli Storage Manager FastBack, TSM FastBack for Workstations
  - ProtecTIER TS7650
  - Tivoli Storage Productivity Center, SAN Volume Controller
- Contact:
  - Richard Vining
  - Product Marketing Manager, IBM Tivoli Storage
  - rvining@us.ibm.com

Questions?

PulseANZ 2010

Meet the people who can help advance your infrastructure

© 2010 IBM Corporation

Khop Khun Mak

Cheers

Sukran

Tack

Merci

Dioich

Gracias

Arigato

Kamsa hamaida

Danke

Salamat

kiitos

Efharisto

Dankie

Grazie

Hvala

Xie xie

Thank You

Kam ouen

Shukrla

Tesekkurler

Köszönöm

Takk

Toda

Spasiba

Trademarks and disclaimers

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries./ Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both. IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency which is now part of the Office of Government Commerce. ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office. UNIX is a registered trademark of The Open Group in the United States and other countries. Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both. Other company, product, or service names may be trademarks or service marks of others. Information is provided "AS IS" without warranty of any kind.

The customer examples described are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by customer.

Information concerning non-IBM products was obtained from a supplier of these products, published announcement material, or other publicly available sources and does not constitute an endorsement of such products by IBM. Sources for non-IBM list prices and performance numbers are taken from publicly available information, including vendor announcements and vendor worldwide homepages. IBM has not tested these products and cannot confirm the accuracy of performance, capability, or any other claims related to non-IBM products. Questions on the capability of non-IBM products should be addressed to the supplier of those products.

All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Some information addresses anticipated future capabilities. Such information is not intended as a definitive statement of a commitment to specific levels of performance, function or delivery schedules with respect to any future products. Such commitments are only made in IBM product announcements. The information is presented here to communicate IBM's current investment and development activities as a good faith effort to help with our customers' future planning.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput or performance improvements equivalent to the ratios stated here.

Prices are suggested U.S. list prices and are subject to change without notice. Starting price may not include a hard drive, operating system or other features. Contact your IBM representative or Business Partner for the most current pricing in your geography.

Photographs shown may be engineering prototypes. Changes may be incorporated in production models.

© IBM Corporation 1994-2010. All rights reserved.

References in this document to IBM products or services do not imply that IBM intends to make them available in every country.

Trademarks of International Business Machines Corporation in the United States, other countries, or both can be found on the World Wide Web at http://www.ibm.com/legal/copytrade.shtml.