

Improve your IT Analytics Capabilities through Mainframe Consolidation and Simplification

Ros Schulman – Hitachi Data Systems
John Harker – Hitachi Data Systems



#SHAREorg



SHARE is an independent volunteer-run information technology association that provides education, professional networking and industry influence.



Improve your IT Analytics capabilities through Mainframe consolidation and simplification

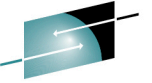


- When large amounts of data need to be analyzed quickly, consolidated and simplified mainframe environments deliver improved results at less cost. Whether you are running Big Data analytics packages or simply cloning databases for quality testing purposes consolidating delivers big operational cost savings while at the same time making it easier to flexibly deliver the power and capacity where it is needed and when it is needed.
- Adding flash into an automatic tiering environment can save even more while delivering significant performance improvements to existing and new applications. Come to this session to learn about how consolidation can improve your mainframe analytics environment.

Complete your session evaluations online at www.SHARE.org/Seattle-Eval



Access More Data For Insight



SHARE
Educate • Network • Influence

Don't just collect data. Put it to work.

**Consolidate
and Integrate**



**Expand data
access for
analysis**

Complete your session evaluation online at www.SHARE.org/Seattle-Era

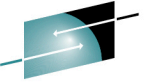
Gain competitive advantage





1. Redefine Your Storage

What is Software Defined Storage?



SHARE
Educate • Network • Influence

- Separates logical from physical storage
- Breaks the link between applications and physical storage
- Consolidates heterogeneous storage into a single managed set
- Software defined infrastructure provides automation, operational flexibility and utilization efficiency



Complete your session evaluations online at www.SHARE.org/Seattle-Eval



Software Defined Storage Abstract for Agility

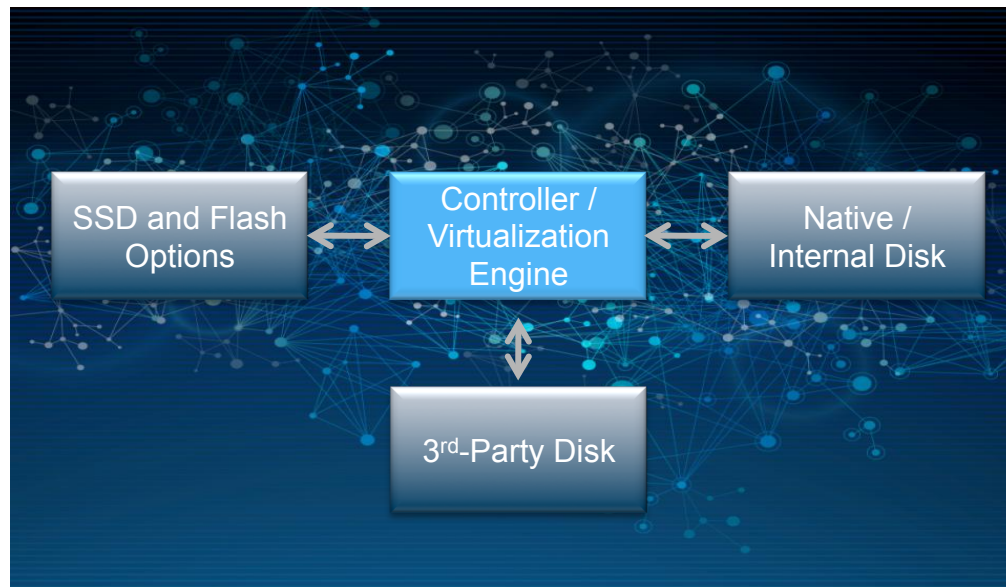
- Transform the fixed into the flexible
- Create tailored IT services



...on a common software-defined infrastructure

Virtualize and Pool to Consolidate and Simplify

- Virtualization with scalable high performance storage
- Extend capabilities to your existing storage
- Automate data placement
- Automate performance and capacity optimization
- Choose the storage to match your application needs



Virtualized Pooled Unified Storage Services

Efficiently Meet Service Levels

DATA DEMANDS



Virtualized Storage Pool

Internal High-Performance Flash



External High-Performance SSD

Internal High-Performance SAS



External High-Performance Fibre Channel

Internal High-Capacity SAS



External High-Capacity SAS

Complete your session evaluations online at www.SHARE.org/Seattle

Virtualized Pooled Resources

Common Management and Monitoring

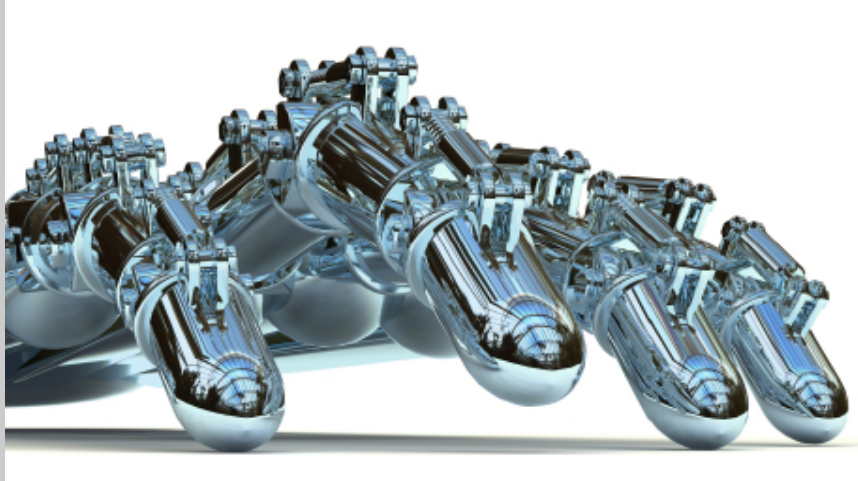
Automated Performance, Capacity Optimization

Application Service-Level Control

Common Data Protection, Disaster Recovery

Automate to Simplify

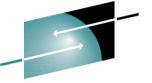
Automate
optimal
service-levels



Simplify
provisioning,
performance
and protection

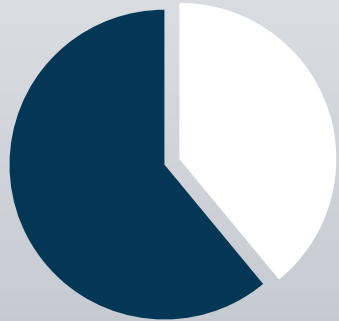
Let your software-defined infrastructure sweat the details

The Need For Automation Is Clear



29%

of time spent on **lower-value** admin and provisioning



39%

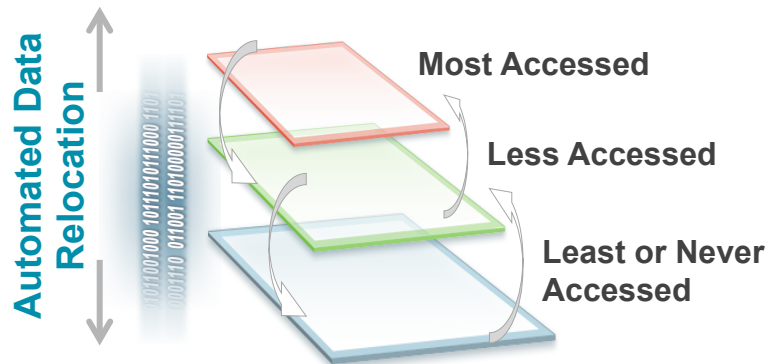
of storage outages caused by **human error**

Automated Optimized Data Placement



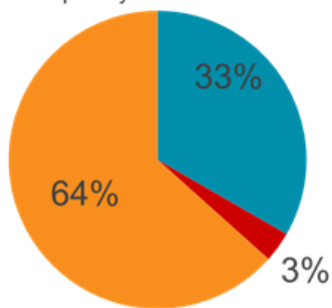
AUTOMATED

- Automated page-based data movement for performance and cost efficiency
- Frees users from hands-on tier management and data layout

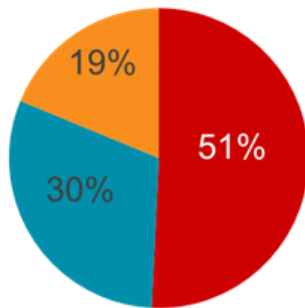


Large Financial Services Customer

Capacity Distribution



Workload Distribution



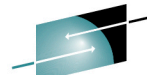
■ SAS ■ Flash ■ NL-SAS

Advantages

- Most efficient use of **flash** ensures that investments are fully utilized
- Reduce costs with self-managed and self-optimized storage tiers

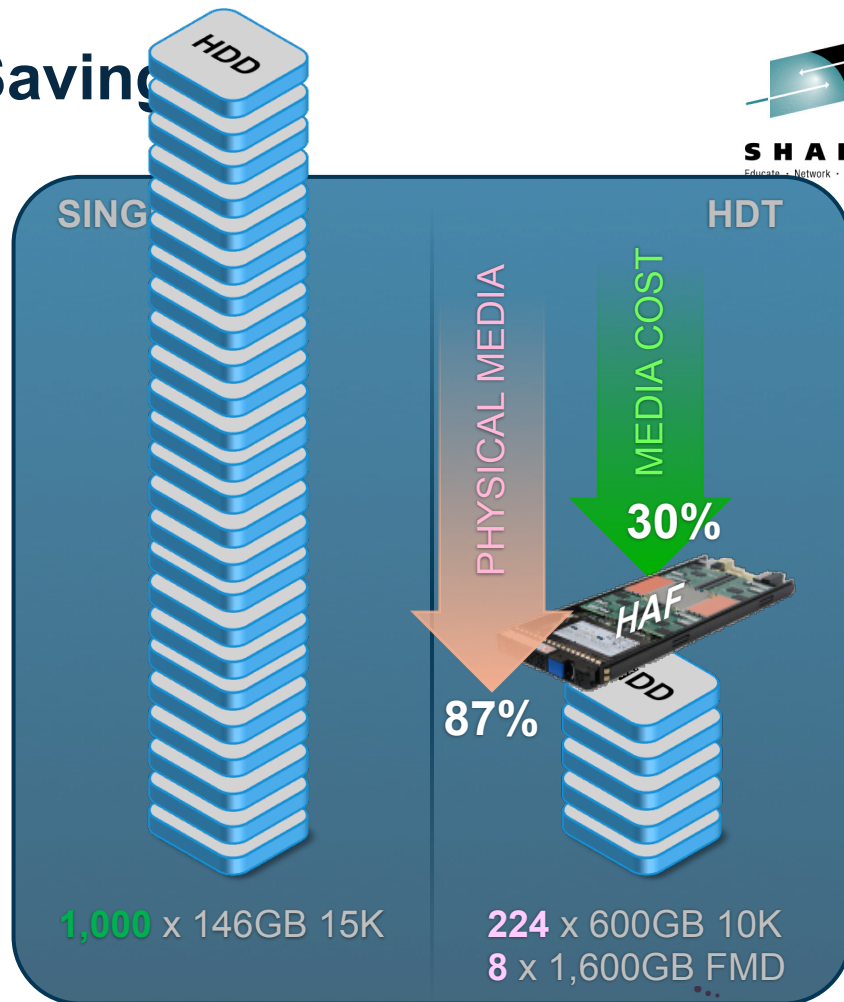


Hybrid Storage Offers Cost Savings



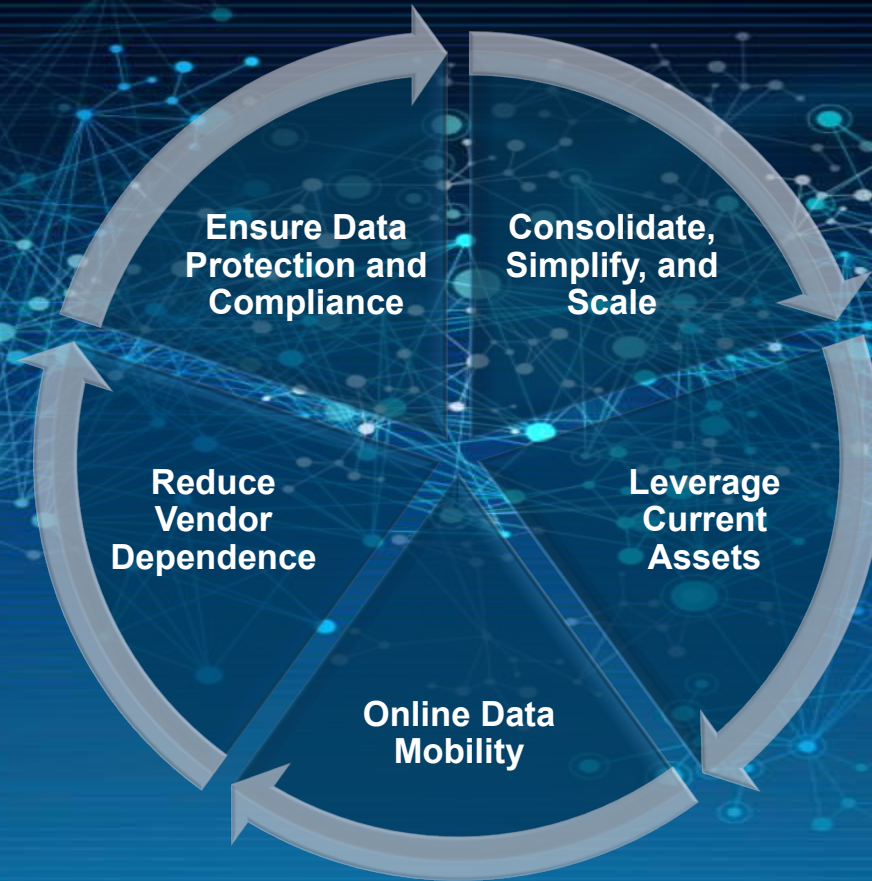
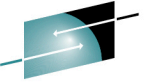
SHARE
Educate · Network · Influence

Flash + Dynamic Tiering
= *EFFICIENCY*



CONFIGURATION = 144TB and 400,000 IOPS

Software Defined Storage Economic Benefits



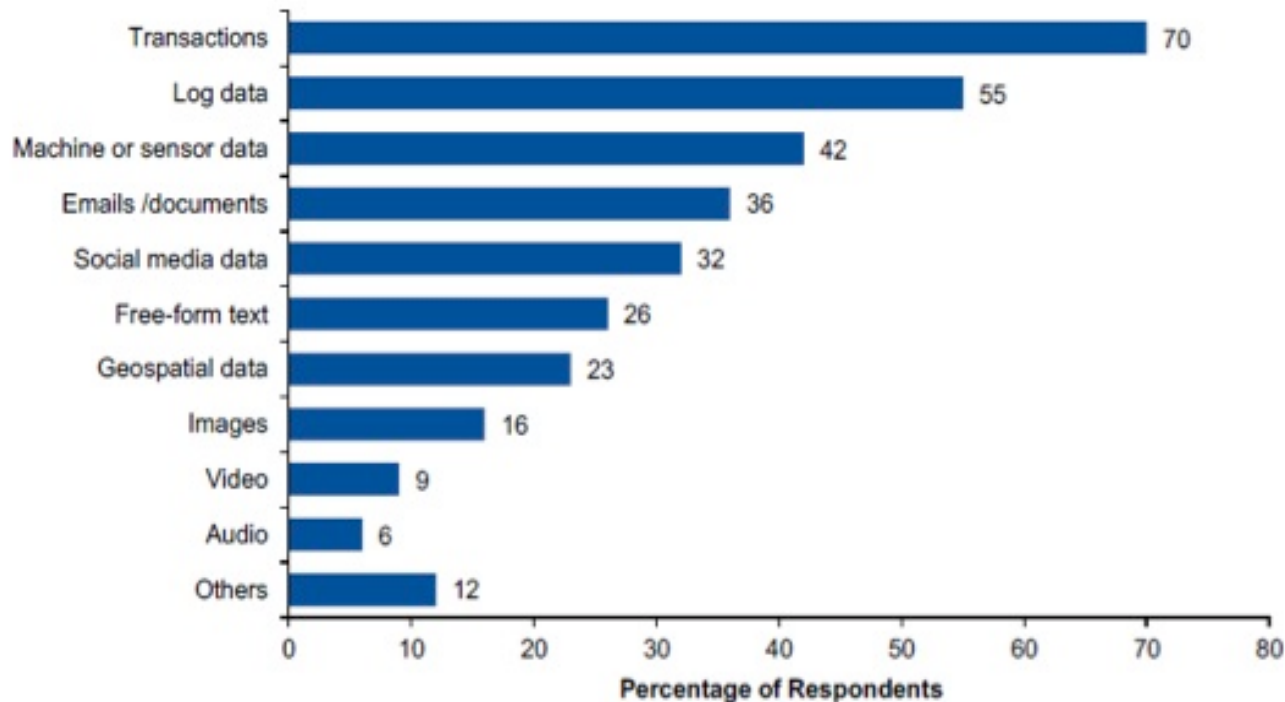
2. Improve Working With Data



What Do You Want To Analyze with What?

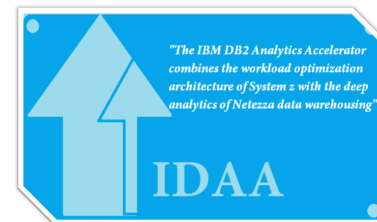


RE
Influence



N =465 (multiple responses allowed)

Source: Gartner (September 2013)



Achieving Business Insight



Where Do You Analyze?

Bring data to the
analytics
or
Bring analytics to
the data

Both are valid answers



Extract, Transform and Load (ETL) Challenges

- **Extraction** is the task of acquiring the data (in whatever format might be possible) from the source systems.
- **Transformation** is converting data formats
- **Load** is the phase where the captured and transformed data is deposited into the new data store (warehouse, mart, etc.).

Issues:

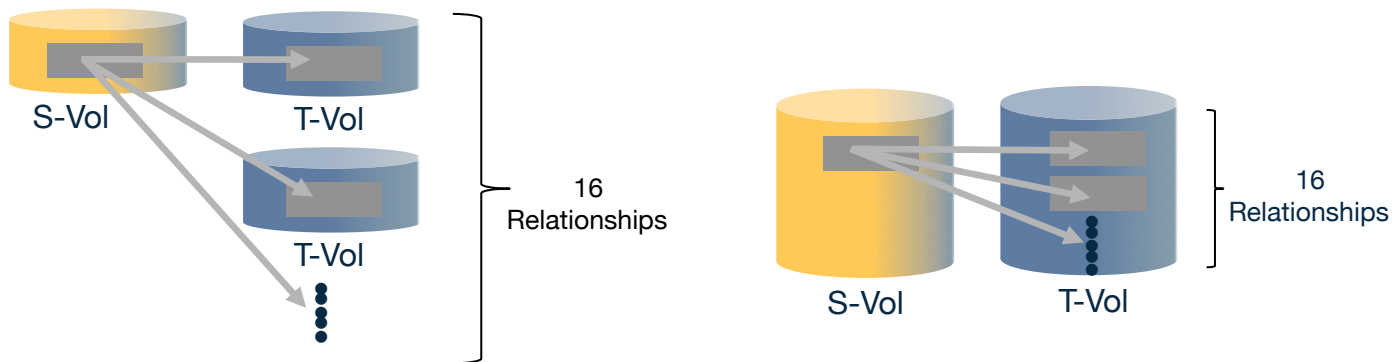
- Time Lag
- Processor Overhead
- Security



Use Analytics tools to help

- Tools that can mine data from many sources
 - RMF/SMF, our own SMF record etc
 - Mine the data
 - Report on the data
 - Export the data into common formats any tool can use
 - The data can be fed to an Hadoop instance or Websphere instance
- Look at proven tools to reduce MIPS consumption for things like DFHSM by avoiding migrations and recalls

Copies of Active Data – ex: FlashCopy Clones



- Provides fast data replication capability
- Allows a copy of source data to targeted volume virtually or physically
- A pair created by version 2 is called a “relationship”
- After a relationship is created, hosts can access data copied to targeted volume

Copies of Data



- Volume Clone and Rename (VCR) for DB2 or IMS
 - Clone large DB2/IMS systems quickly and easily
- Database Backup and Recovery for DB2 or IMS
 - Storage-aware backup and recovery solution for big data

Complete your session evaluations online at www.SHARE.org/Seattle-Eval



Bridging Two Worlds – Options

- **WebSphere MQ** – messaging application to exchange information across different platforms
- **S390-tools-zdsfs** – mount a z/OS DASD as Linux file system, reads files from z/VM from z/OS formatted volumes
- **Sqoop** - scriptable interface for pulling relational data from a SQL database and loading it into Hadoop
- **SyncSort** ingests, translate, process and distribute mainframe data with Hadoop
- **Hitachi Cross O/S file exchange** – simultaneous volume access from FICON/3390 and Fiber/Open-V connections
- **Infosphere Change Data Capture**
- **InfoSphere System z Connector** for Hadoop

Complete your session evaluations online at www.SHARE.org/Seattle-Eval

Bridging Two Worlds

- **IBM Infosphere Change Data Capture**
 - Real time replication of data from IBM DB2
 - Part of IBM InfoSphere Data Replication for DB2 for z/OS
 - Allows DB2 for z/OS data to be reconstituted into an Hadoop Distributed File System
 - Works directly from DB2 change logs
 - Supports replicated enterprise data volumes while maintaining transactional integrity and consistency.

Bringing Two Worlds



- Integrate DB2 applications with wide range of data sources - on and off mainframe
- Access to mainframe data in place without ETL
- Breadth of data providers - Big Data, mainframe, distributed, machine data, syslogs
- Breadth of data consumers - Business Intelligence, Analytics, Cloud, Mobile, Web
- Extend the value of your data warehouse with real-time data virtualization

Complete your session evaluations online at www.SHARE.org/Seattle-Eval



Bridging Two Worlds



PROCESS MAINFRAME FILES IN HADOOP SO THAT MAINFRAME "DARK DATA" CAN BE PART OF BIG DATA ANALYTICS PROJECTS!

- LegStar z/OS File reader - community-developed plugin available through the Pentaho Data Integration (PDI) Marketplace. Uses mainframe JDK (Java) to compile a client library for parsing your specific COBOL Copybook file.
- Run on mainframe, LegStar reads mainframe records from file to transform them into PDI rows including complex transformations for use by things like MapReduce.
- FTP or other means is used to transfer the data for HADOOP processing.

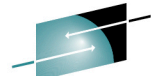


3. Achieve Performance

Consolidate Smartly

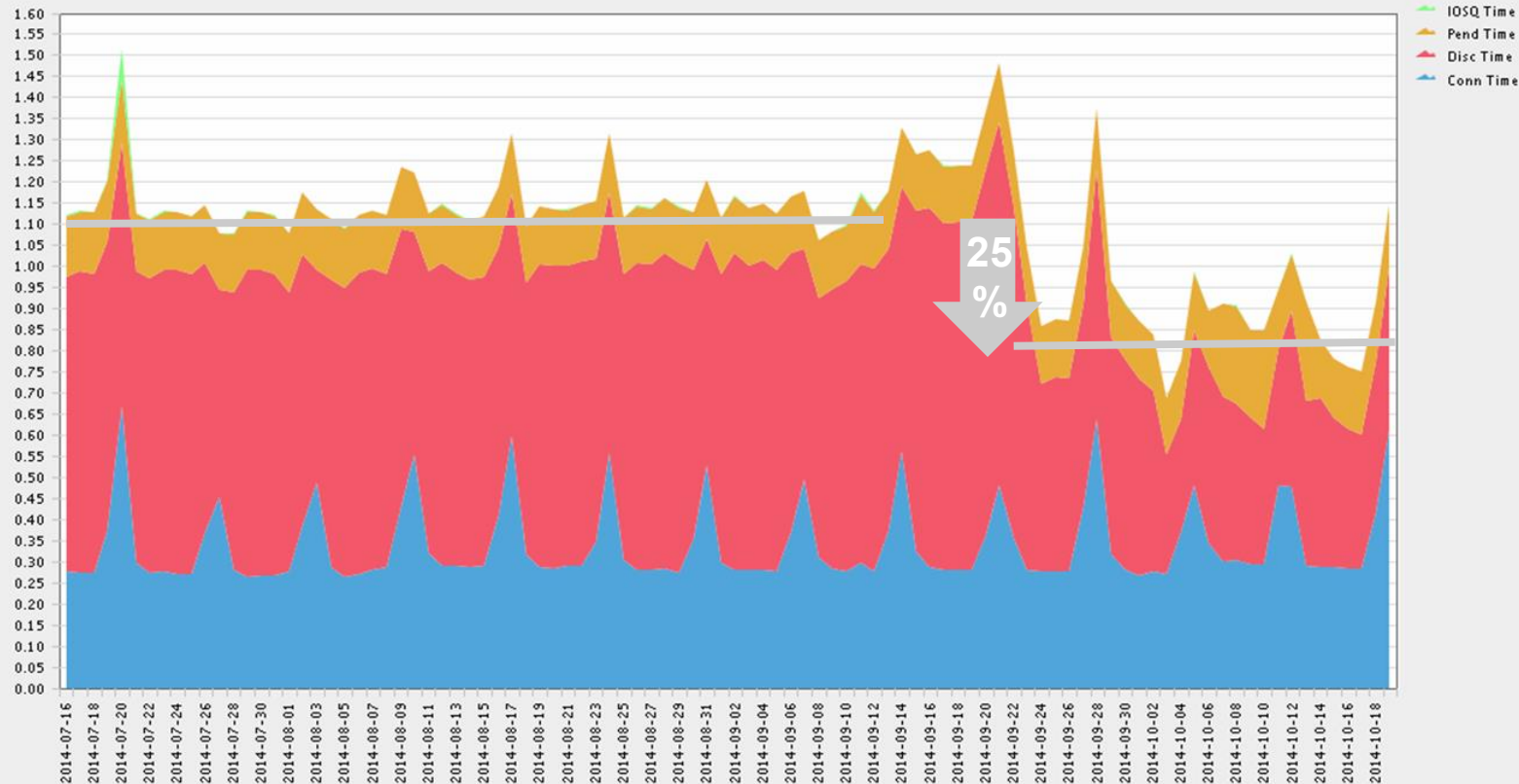
- Simplify by consolidating and virtualizing storage
- Take advantage of proven performance technologies
 - Dynamic Tiering
 - Flash
 - Use advanced storage to scale to 2 million IOPS +
- DO NOT have too many exceptions to the rules 😊
- Decide whether you want to analyze on the MF or offload the data
 - Plus and minus to both

Major Financial Company - Effect of adding Flash and Dynamic Tiering – Average R/T



SHARE
Rate · Network · Influence

Avg Response Time per Day in ms



Compl

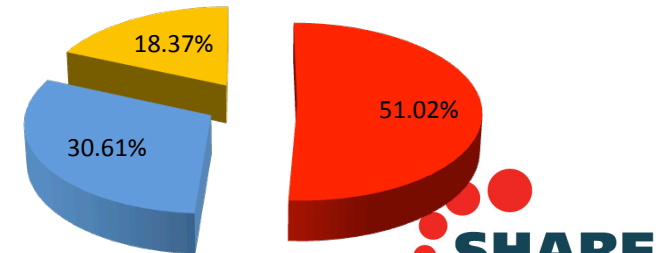
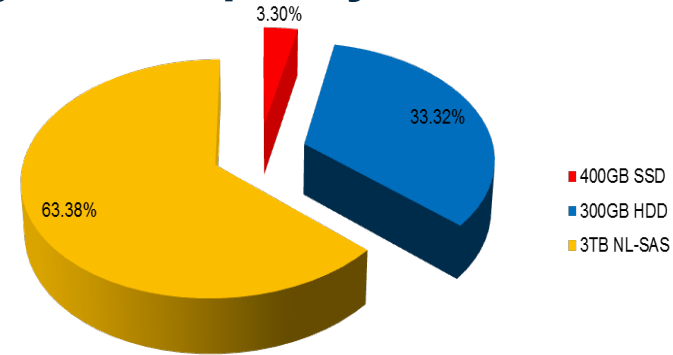
ARE
tle 2015

Major Global Financial Company (U.S. based) Back-End I/O Distribution vs. Physical Capacity



- Excellent performance benefit with a small amount of SSD
- Proven ability to take advantage of larger-capacity drives
- 2X improved response time
- Increased workload of up to 55% more I/O
- 50%+ environmental cost savings
- Significant software savings through consolidation
- Automated performance and capacity management

Physical Capacity Distribution



I/O Rate Distribution

Complete your session evaluations online at www.SHARE.org/Seattle-Eval



Summary – Efficiently Enabling New Uses

Virtualized Pooled Storage



Software Defined Agility

Extreme Performance



Automated

Bridge two worlds



Real-time Access

Room to Grow



>>> INNOVATE
WITH INFORMATION™



SHARE
Educate · Network · Influence

Thank You

© Hitachi Data Systems Corporation 2014. All rights reserved. HITACHI is a trademark or registered trademark of Hitachi, Ltd. Innovate With Information is a trademark or registered trademark of Hitachi Data Systems Corporation. Microsoft, Active Directory, Hyper-V, SQL Server and Windows Server are trademarks or registered trademarks of Microsoft Corporation. All other trademarks, service marks, and company names are properties of their respective owners.



#SHAREorg



SHARE is an independent volunteer-run information technology association that provides education, professional networking and industry influence.

