

# Dell™ Compellent™ Storage Center 5.3 Data Progression with Exchange 2010

---

Dell Compellent Technical Solutions



THIS WHITE PAPER IS FOR INFORMATIONAL PURPOSES ONLY, AND MAY CONTAIN TYPOGRAPHICAL ERRORS AND TECHNICAL INACCURACIES. THE CONTENT IS PROVIDED AS IS, WITHOUT EXPRESS OR IMPLIED WARRANTIES OF ANY KIND.

© 2011 Dell Inc. All rights reserved. Reproduction of this material in any manner whatsoever without the express written permission of Dell Inc. is strictly forbidden. For more information, contact Dell.

*Dell*, the *DELL* logo, and the *DELL* badge are trademarks of Dell Inc. *Microsoft*, *Windows*, *Windows Server*, and *Active Directory* are either trademarks or registered trademarks of Microsoft Corporation in the United States and/or other countries. Other trademarks and trade names may be used in this document to refer to either the entities claiming the marks and names or their products. Dell Inc. disclaims any proprietary interest in trademarks and trade names other than its own.

November 2011

## Contents

- Introduction ..... 2
- Background/Overview ..... 3
- Virtualization Advantages..... 3
- Exchange Efficiencies..... 3
- Simulation Findings..... 4
- Summary ..... 5
- Appendix ..... 6

## Figures

- Figure 1. Configuration environment ..... 2
- Figure 2. Effect of Data Progression ..... 4
- Figure 3. Amount of space used ..... 5
- Figure 4. Data Residing in Tier 3 ..... 6
- Figure 5. Data Progression charts over time ..... 7
- Figure 6. Data Progression charts over time ..... 8
- Figure 7. Data Progression charts over time ..... 9
- Figure 8. Data Progression charts over time .....10
- Figure 9. Data Progression charts over time .....11
- Figure 10. Data Progression charts over time .....12
- Figure 11. Data Progression charts over time .....13
- Figure 12. Data Progression charts over time .....14
- Figure 13. Data Progression charts over time .....15

## Introduction

The purpose of this document is to show results from a simulation of the data patterns of Exchange 2010 mailboxes hosted on a hosted on a Dell™ Compellent™ Storage Center clustered array enabled with Data Progression. The system included 3Gb SAS drives, two trays of 450GB 15k drives, and two trays of 1TB 7k drives. The simulation environment included 5,000 mailboxes using two 24 core servers. The storage profile used was the recommended, with daily replays. The front-end connectivity was 4Gb Fiber-Channel.

Though the mailboxes were configured to generate 2GB each, because of daily replays, the charts show a larger volume of data. Replays were used to simulate a real-world production environment, with recovery points available. The servers were configured with five databases each, which would support up to 10TB of email per server. The database and log files were moved to separate volumes.

Automated tiered storage from Dell Compellent, called Data Progression, is designed to optimize the use of storage in Exchange environments. Data Progression reduces the need for large numbers of high-performance, high-cost disks by moving frequently used data to higher performance tiers of storage while moving infrequently used data to lower cost, higher-density disks. The migration of data is done on the block level, so data within a volume can be moved based on performance characteristics.

The activity level for the mailboxes was set at 150 messages/day/user. In most cases this will be heavier than the standard customer load. The goal was to provide sufficient load to give a good picture of what Data Progression will accomplish, and to show how much data movement will occur.

Figure 1 below shows the logical configuration:

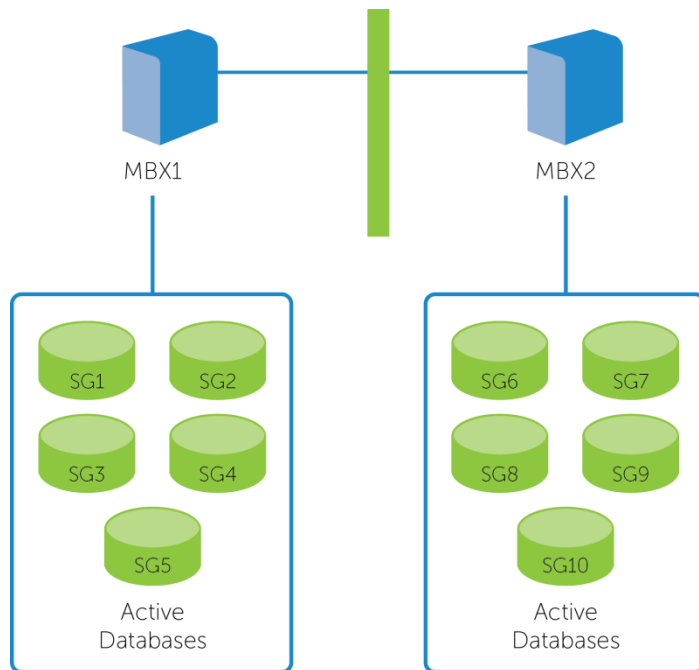


Figure 1

## Background/overview

Industry studies show that as much of 80 percent of Exchange data is inactive. This means that a lot of fast, higher-cost storage is being unnecessarily utilized. This is where automated tiering technology like Data Progression can deliver immense value. The ability to automatically classify and migrate inactive portions of an Exchange Server database at the block-level onto lower-cost drives with D is causing many organizations to rethink their Exchange Server storage strategy.

Storage Center's patented Data Progression technology uses real-time intelligence to identify inactive blocks in an Exchange Server configuration and automatically migrates those blocks off an expensive storage tier to a lower-cost tier of storage based on frequency of access. This solution does not require time consuming data classification and the repetitive manual transfer of data between tiers. Each volume is configured by default with a recommended storage profile that manages the RAID configuration and provides optimal operation and performance for Exchange on the Dell Compellent Storage Center. With this configuration, all data written to each volume is written at RAID10 providing the best possible I/O performance for Exchange database and log operations.

## Virtualization advantages

In Exchange Server environments using traditional SAN platforms, up to 60 percent of disk space ends up allocated but unused. A key feature in the performance of a Dell Compellent-backed Exchange Server solution is its advanced virtualization capabilities. Dell Compellent virtualizes at the disk level within the SAN, accelerating data access for Exchange Server by spreading read/write operations across all the disk drives so multiple requests can be processed in parallel. Full disk parallelization increases I/O performance and reduces hot spots in Exchange Server environments without requiring performance tuning. Each volume created spans multiple tiers as well as drives within the pool, with the option to utilize all shared drives or a specific subset of drives for each volume.

## Exchange efficiencies

Based on the changes in behavior in Exchange 2010, best practices have also changed. A big part of the change involves the architecture and data structure of databases and the shift away from Single Instance Storage. More specifically, the database has been restructured and the schema flattened to produce a more efficient, sequential-like write pattern. This allows greater control over disk IO, as a mailbox is more self-contained within the database. Now when a mailbox is loaded, all indexes or views required to render a mailbox are contained within that mailbox, requiring fewer random reads, and allowing sequential reads to open a mailbox.

This change alone has helped to reduce the overall I/O of the database and reduced mailbox IO requirements by 70 percent. This reflects an IOPS per mailbox decrease of 90 percent since Exchange 2003. With all of the changes in the mailbox database structure, Exchange now has the capability to offer mailboxes that are 5GB-10GB each. The database page size has been increased to 32K (up from 4K in Exchange 2003 and 8K in Exchange 2007). This means that more data can be written to a page which in turn requires less overall reads.

To offset the storage increase, Exchange 2010 does fast page compression. This essentially offsets the increase in space required, but requires a lot of processing power to compress and decompress real-time. This is the reason for an increased number of servers. The user count per processor now matters as well.

## Simulation findings

As data grows and usage patterns change, Data Progression can automatically move inactive blocks of data to a lower tier of storage (both disk class and RAID level) on-the-fly. With the recommended storage profile, active data is always written at RAID10, while any replays are initially stored at Tier 1 on RAID-5. This data eventually makes its way down the RAID levels and tiers. Figure 2 below shows the effect of Data Progression on Exchange 2010 volumes. Data will progress as it ages.

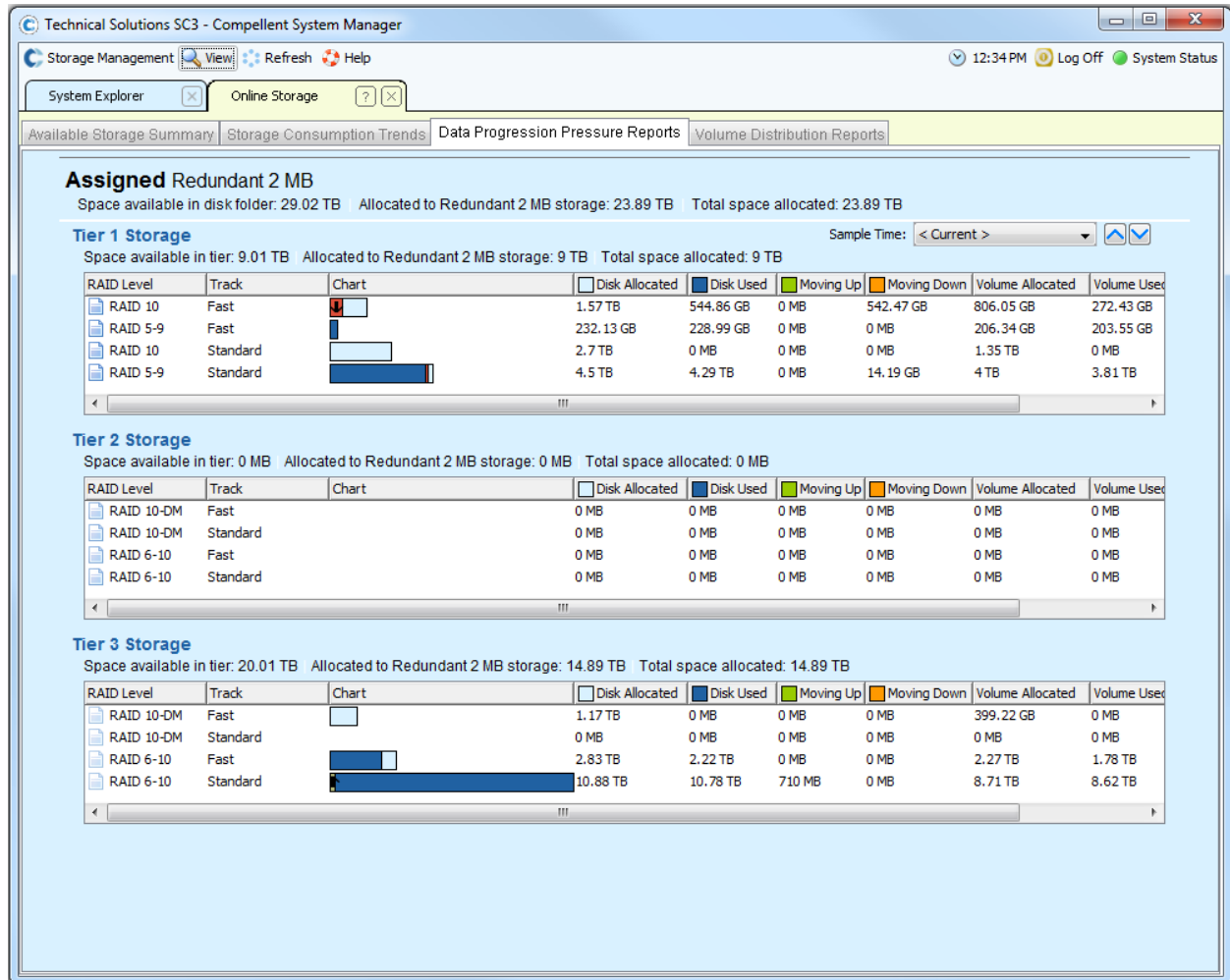


Figure 2

Leveraging a second or even third tier of storage for inactive data allows administrators to increase the number of recommended users without adding additional high-performance spindles. With Data Progression, increased user loads are a scalable benefit. As an Exchange Server configuration gets larger and user mailbox sizes increase, the amount of email that ages and is then automatically migrated to a lower tier increases. More users with more email can be managed with slower, less-expensive spindles on a second tier instead of constantly expanding by purchasing faster, more-expensive spindles.

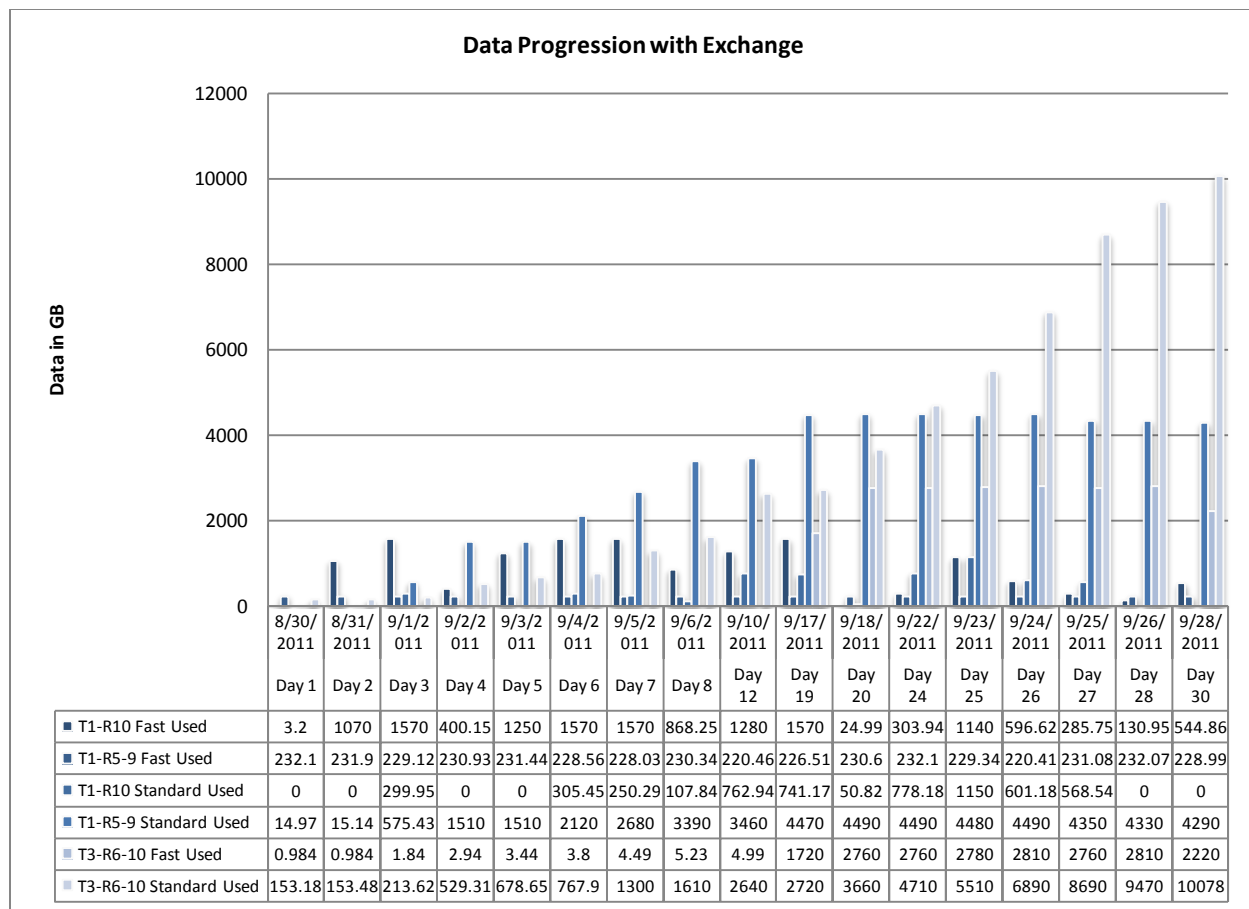


Figure 3

**Summary**

For most companies, an Exchange Server database represents perhaps the most significant amount of stored and inactive data. Within a week after receiving an email, most users no longer access attachments and calendar items. In fact, for many, Exchange Server becomes a kind of online repository. Users don't have or take the time to delete inactive items.

For some companies, just migrating calendar data can result in significant savings. This unique approach to intelligently managing data is changing traditional views of storage management and Exchange Server storage platform implementation. Using Data Progression results in as much as a 74 percent drop in hardware costs alone. And by eliminating manual data classification and movement, Dell Compellent Data Progression cuts storage administration time in half, allowing companies to achieve the benefits of a tiered storage environment without the traditional costs or complexities.

Figure 3 shows the amount of space used and the progression of data within the tiers and RAID levels over time. Historical data moves down the tiers very similar to other types of data. The most active data stays on Tier 1, the historical data moves to Tier 3. The data will continue to progress over time.

At the end of the test runs, the majority of the data has moved down to Tier 3, (Figure 4) showing how Data Progression can save space on Tier 1 disk. By moving the historical data to Tier 3, the Tier 1 disk

will be available for other applications to use. Exchange can also continue to benefit from the Tier 1 disk available for new email.

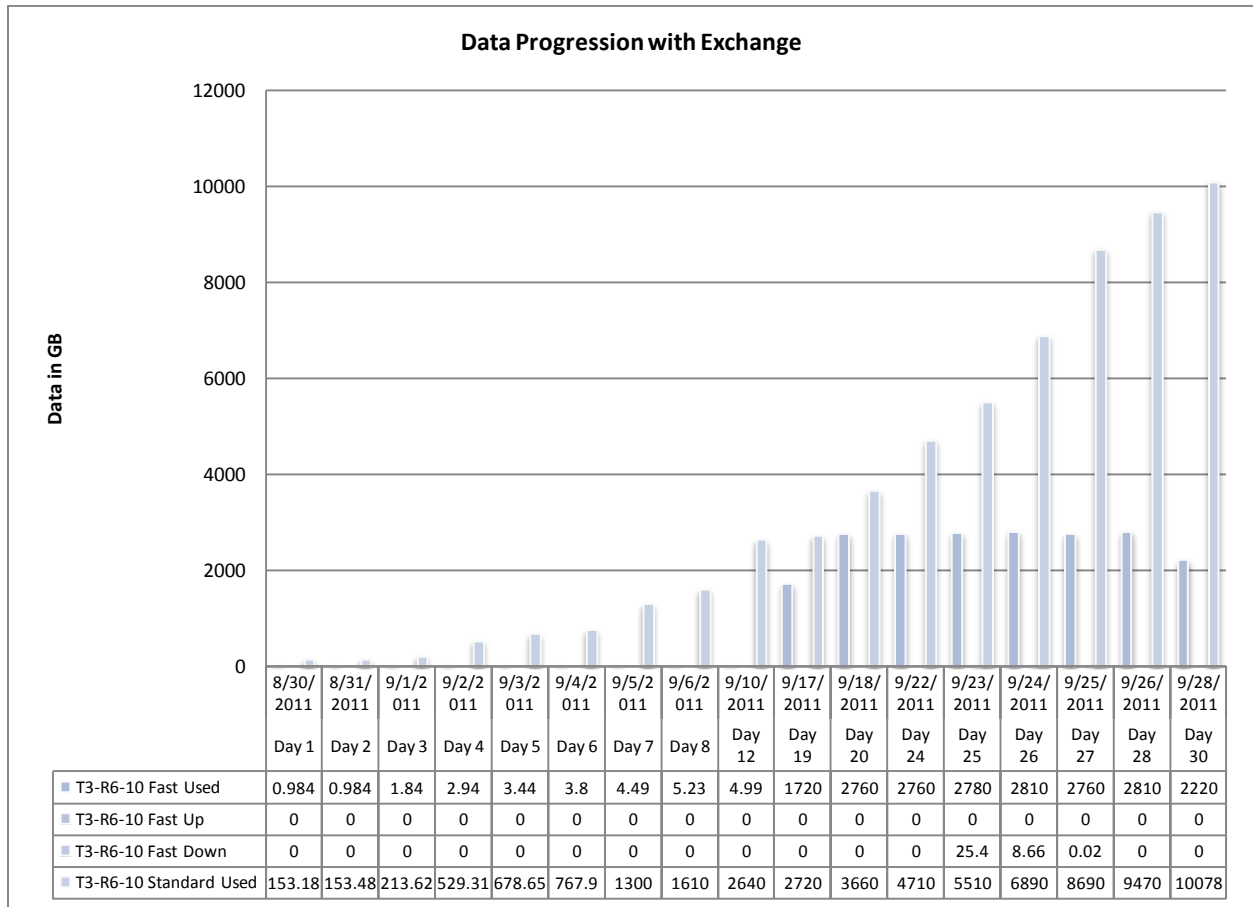


Figure 4

**Appendix**

Figures 5-13 in the attached appendix show the progression of the data over time as the mailboxes grow and data is generated in them. The time for progression was four weeks. These charts show Exchange in an isolated environment. Other applications running on the same Storage Center will affect data flow as well.

# Dell Compellent Data Progression with Exchange 2010

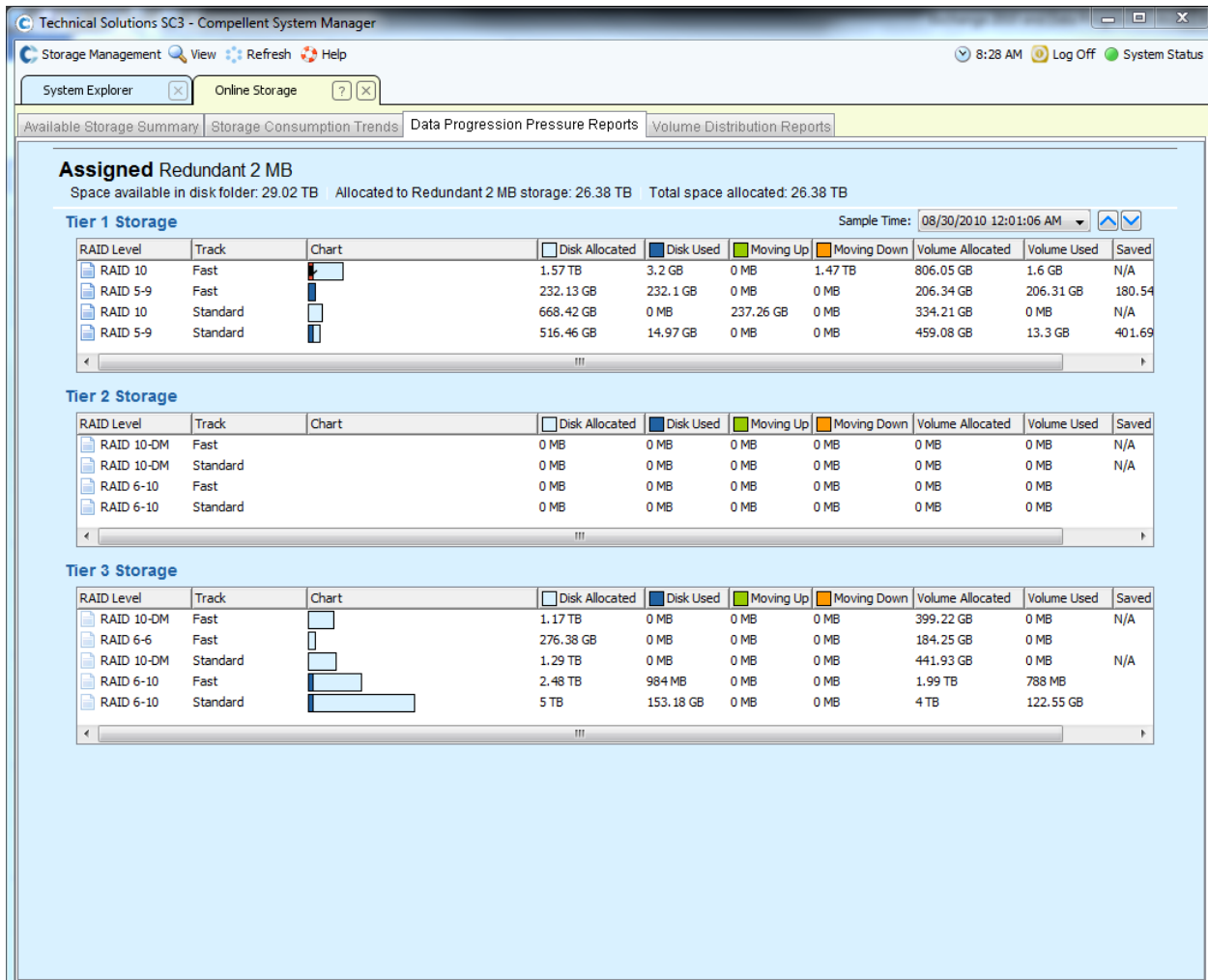


Figure 5

# Dell Compellent Data Progression with Exchange 2010

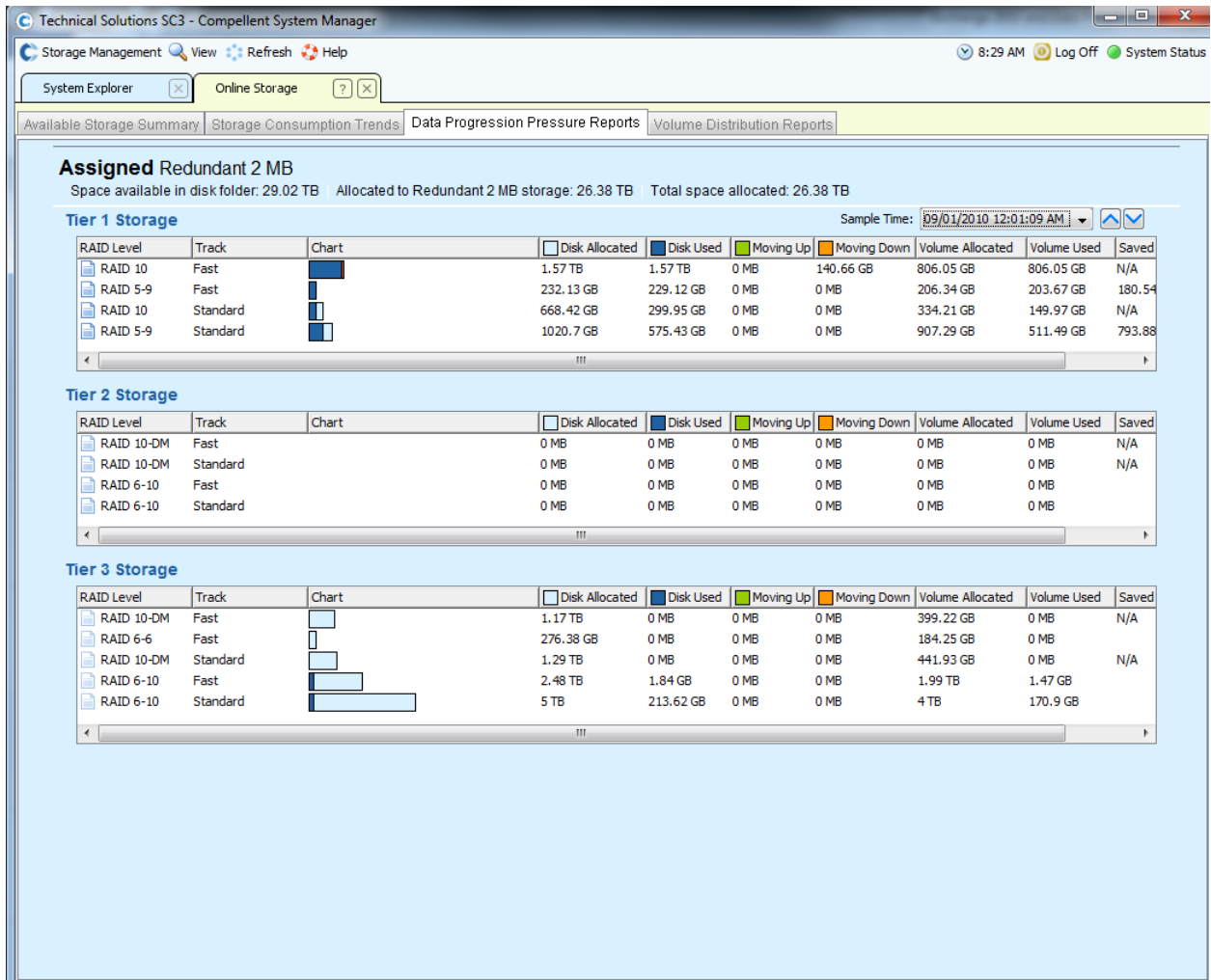


Figure 6

# Dell Compellent Data Progression with Exchange 2010

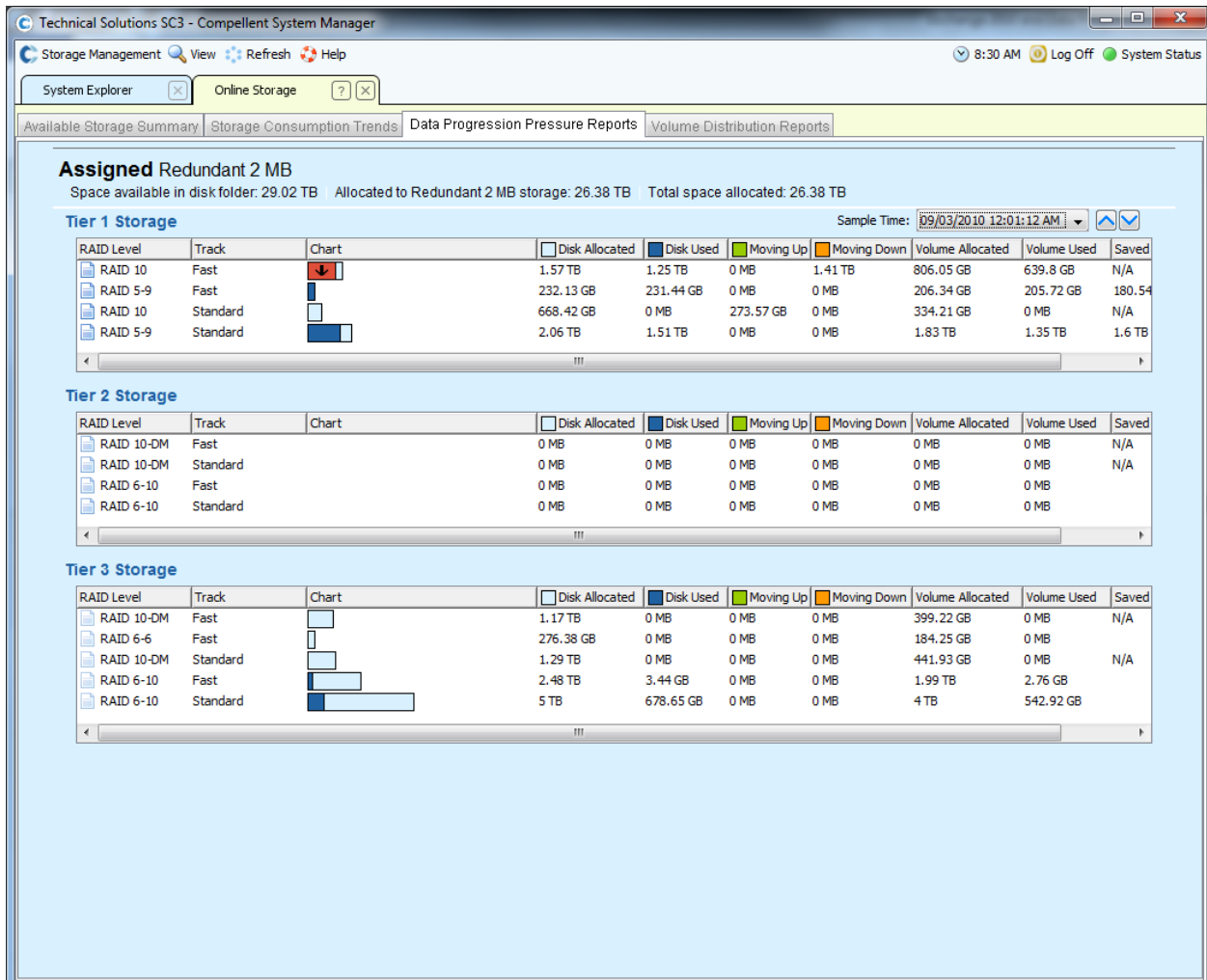


Figure 7

# Dell Compellent Data Progression with Exchange 2010

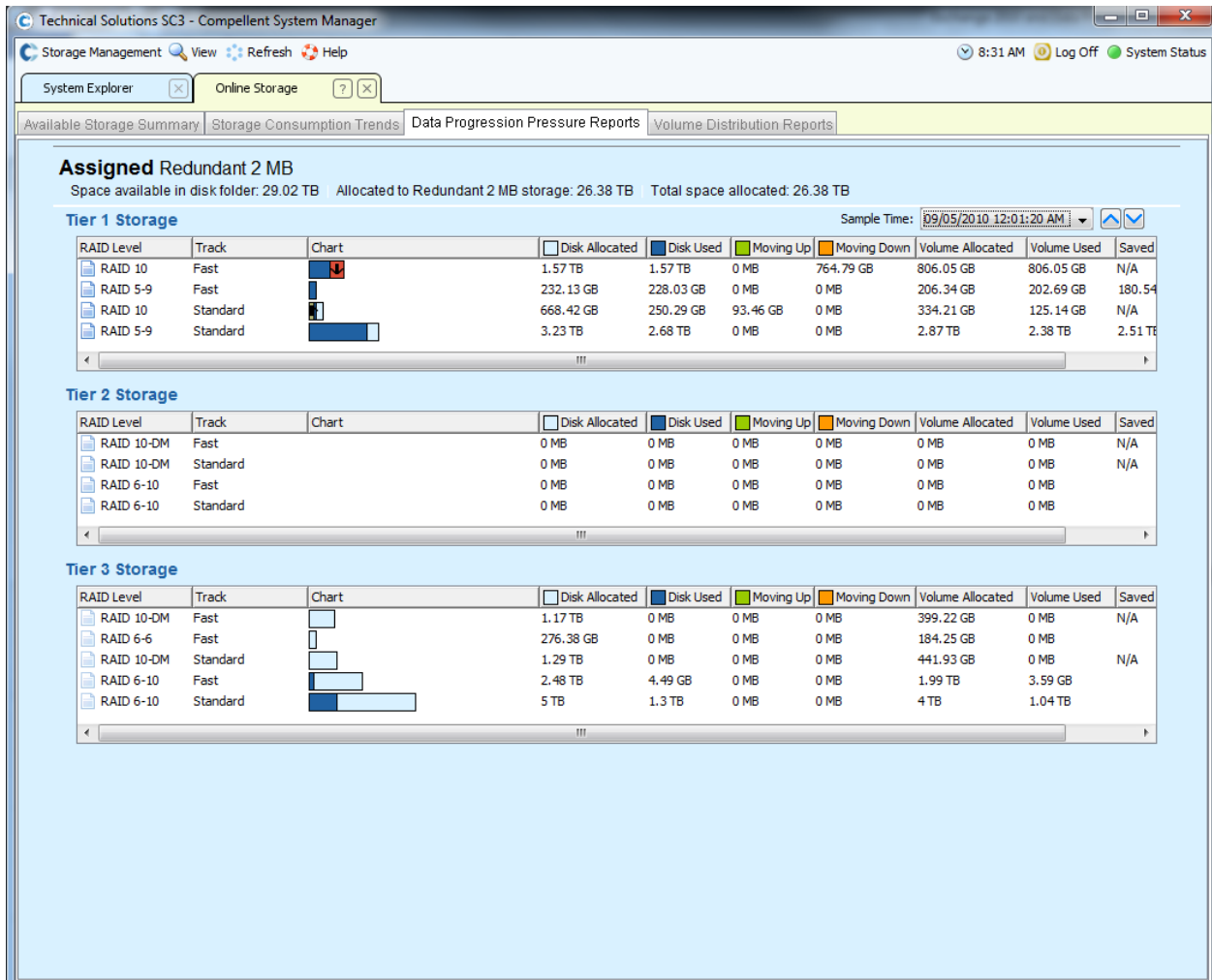


Figure 8

# Dell Compellent Data Progression with Exchange 2010

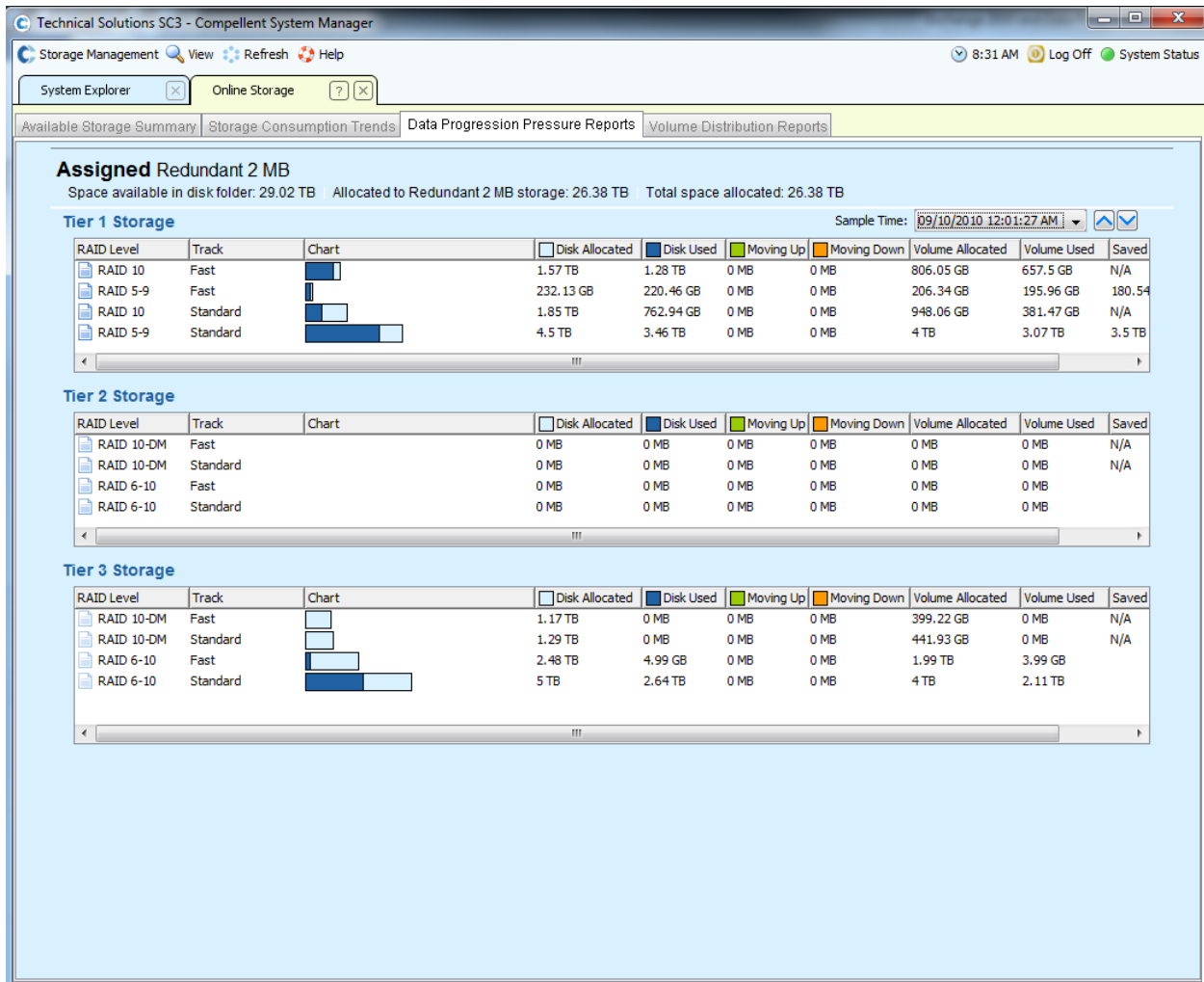


Figure 9

# Dell Compellent Data Progression with Exchange 2010

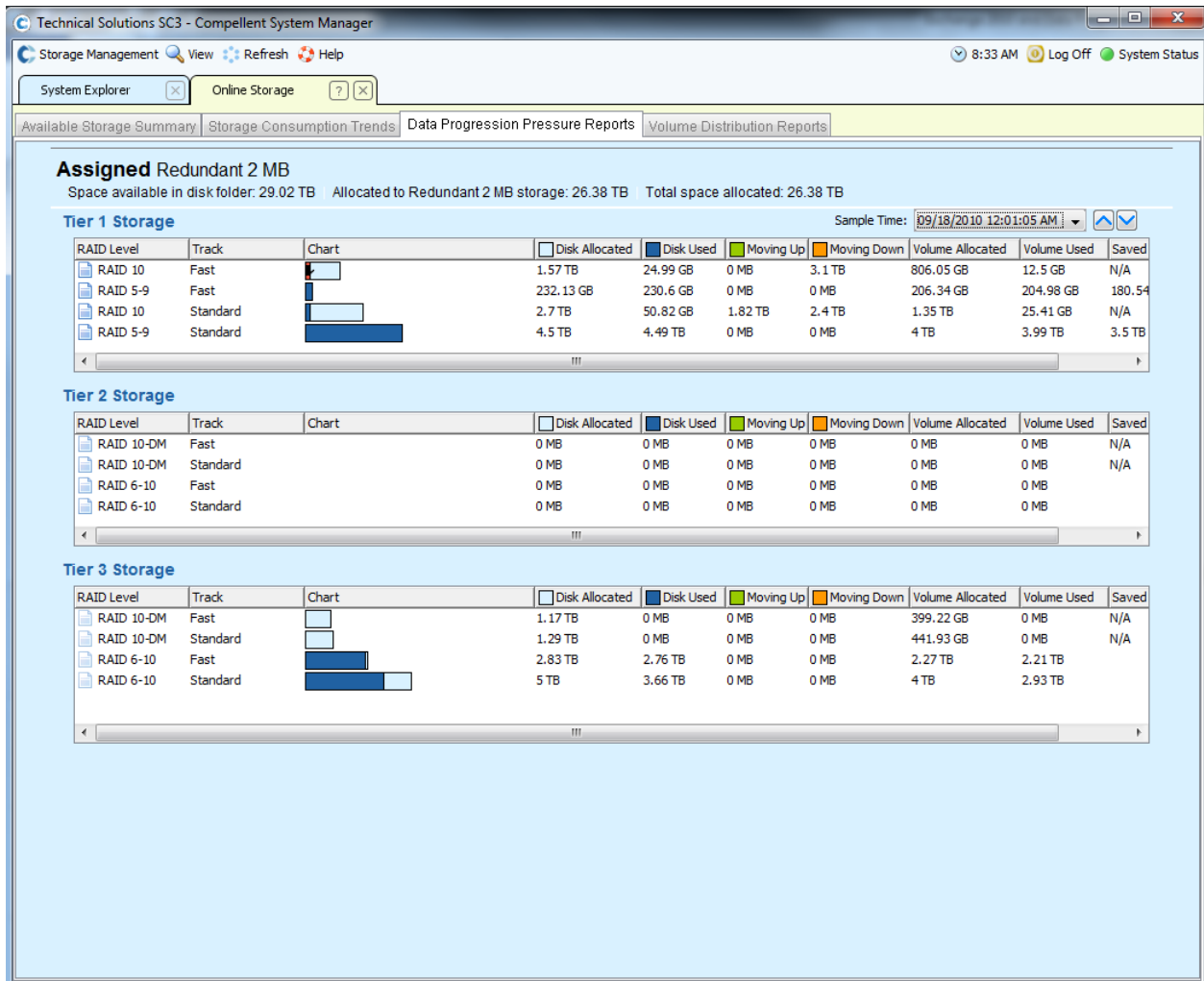


Figure 10

# Dell Compellent Data Progression with Exchange 2010

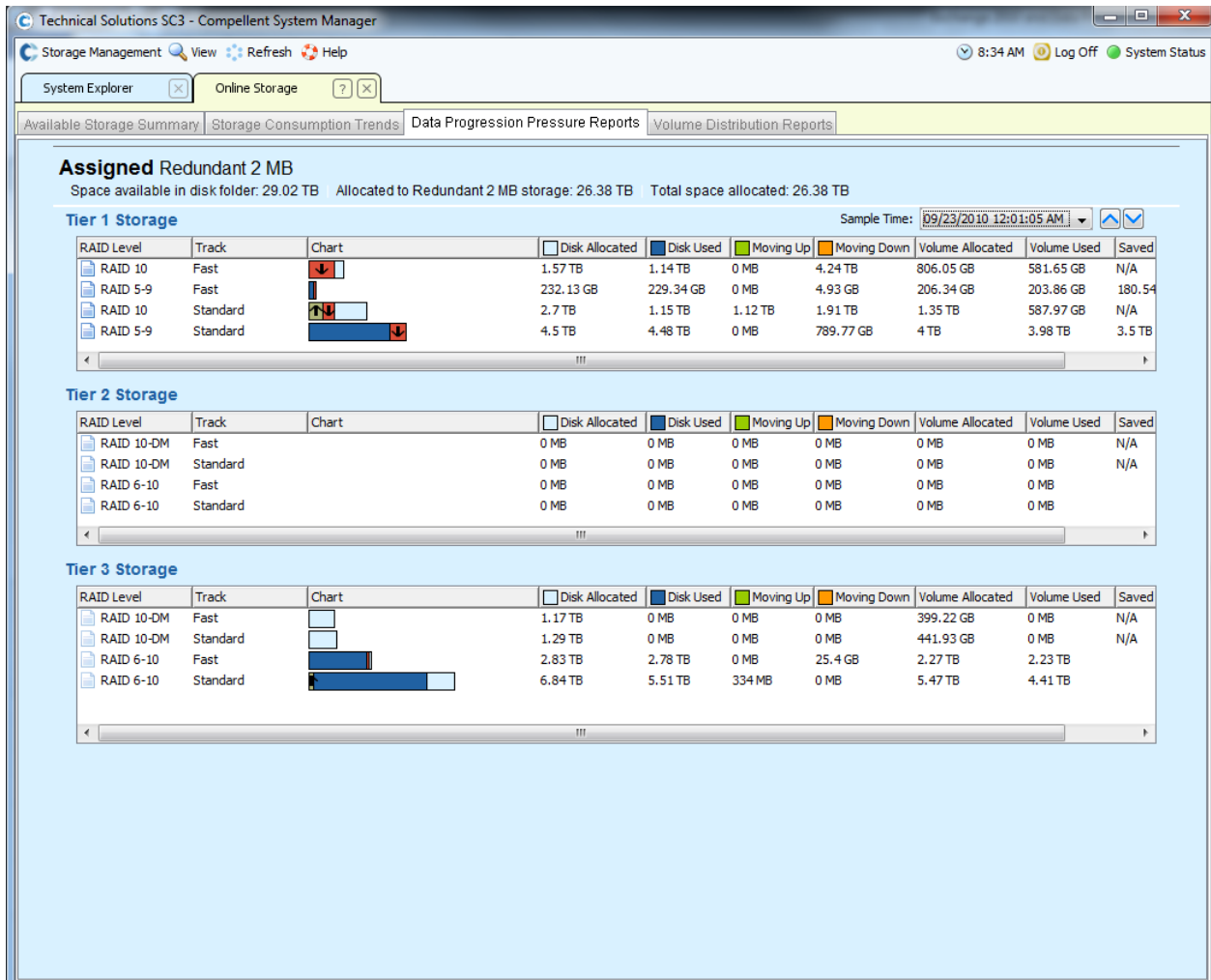


Figure 11

# Dell Compellent Data Progression with Exchange 2010

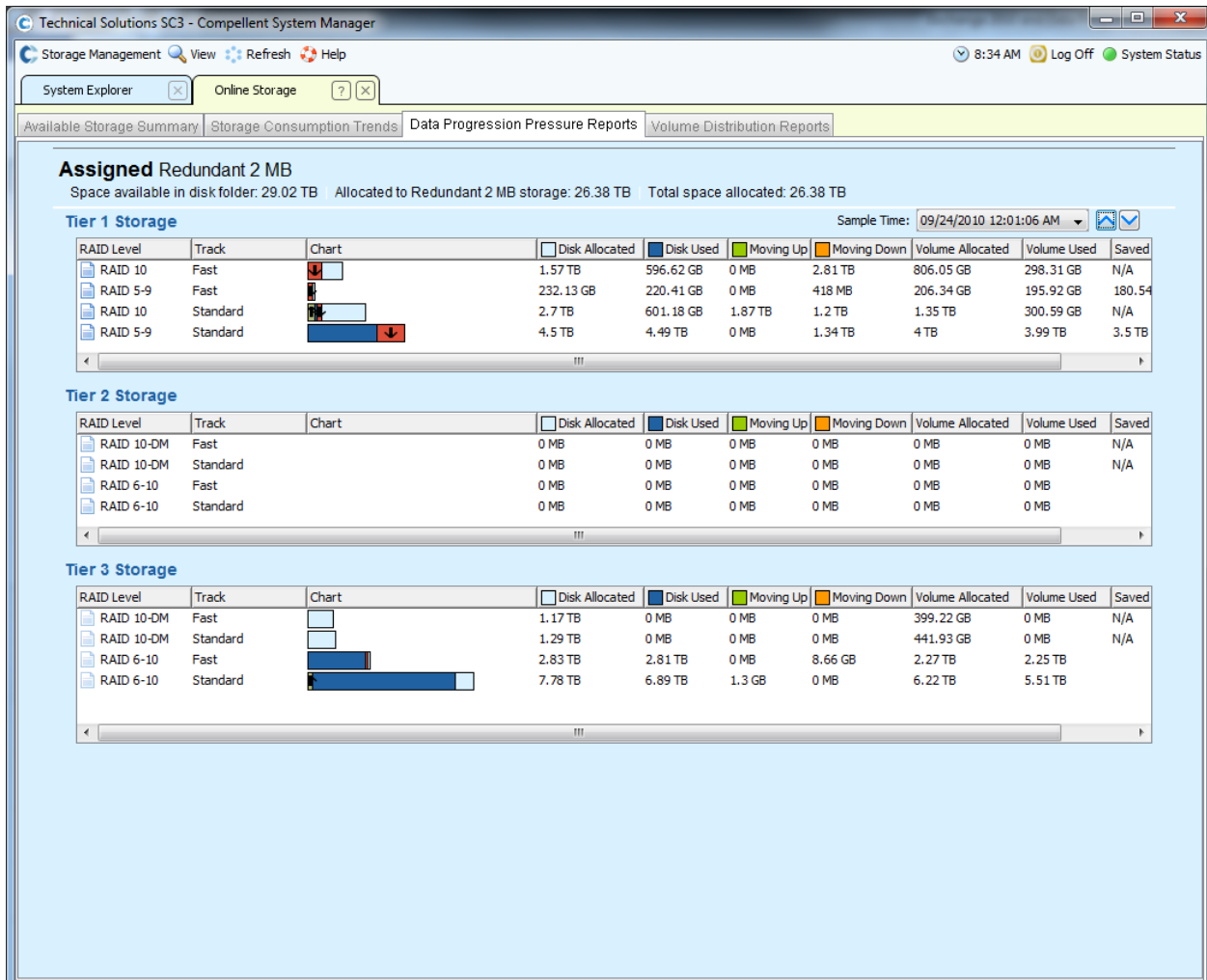


Figure 12

# Dell Compellent Data Progression with Exchange 2010

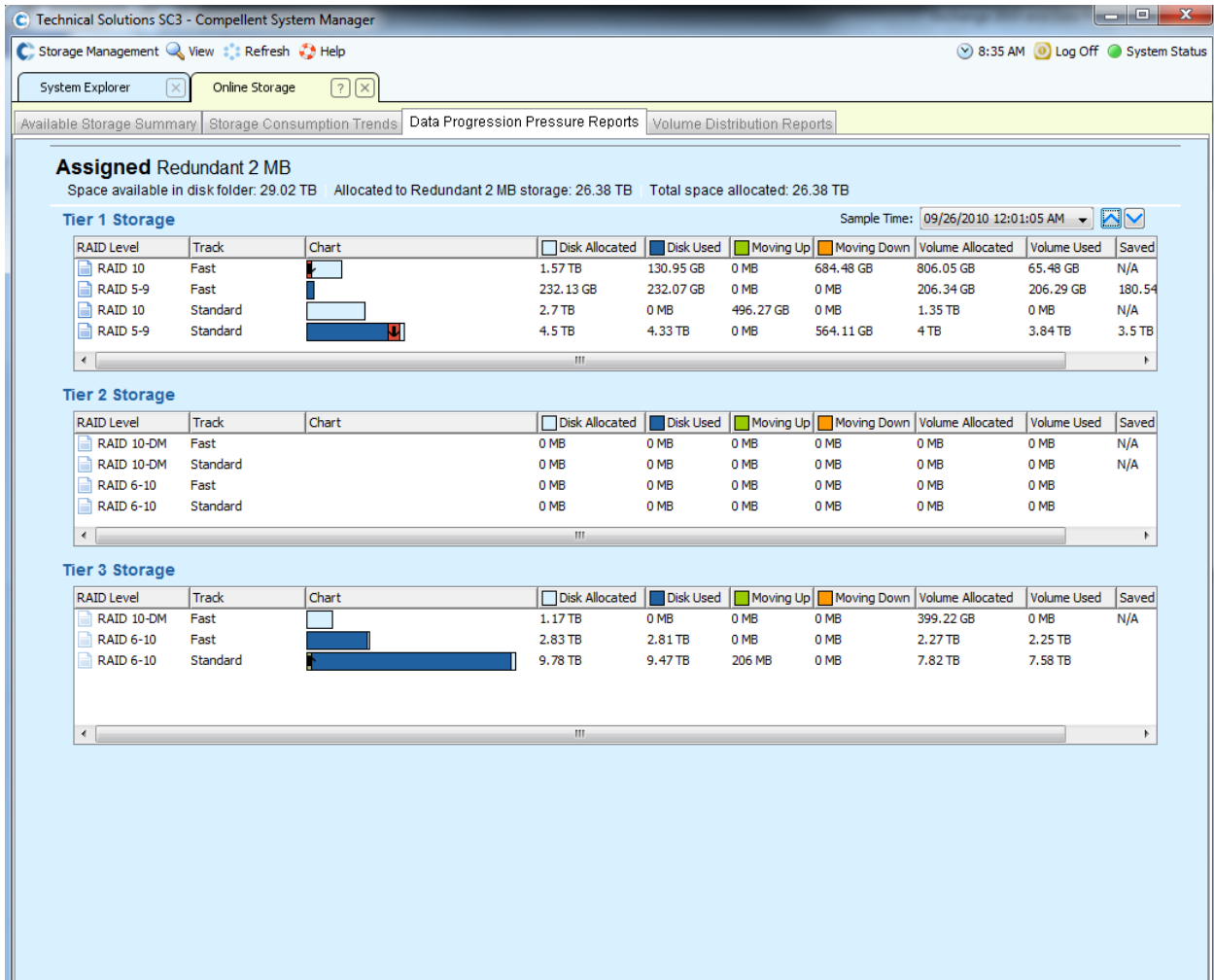


Figure 13