Quantum®

TECHNOLOGY BRIEF

DXi Accent Technical Background

TABLE OF CONTENTS

INTRODUCTION

DXi Accent™ is software that allows the Quantum DXi® deduplication to be used in a collaborative or hybrid mode, moving a portion of the deduplication operations to a backup server so that only unique blocks are moved over the network to a target appliance. Use of DXi Accent can increase backup performance where network bandwidth is the limiting factor, and can extend duplication beyond the appliance in an integrated, easy-to-deploy, and cost-effective solutions set. DXi Accent is available as a standard feature on all current Quantum deduplication appliances for operation in environments using NetBackup or Backup Exec and the Symantec OpenStorage API (OST). This technical brief describes DXi Accent software, its mode of operation, general use cases, and considerations for deployment.

BACKGROUND ON DEDUPLICATION

Quantum's patented[1] deduplication system operates by segmenting a stream of data into variable-length blocks, identifying repeated blocks, then compressing and writing only the unique segments to a blockpool on disk. To identify repeated blocks in a transmitted stream, the data deduplication engine creates a digital signature—like a fingerprint—for each data segment, and it keeps an index of the signatures for each repository. The index provides the reference list used to determine whether blocks already exist in a repository, and it is used to determine which data segments need to be stored or copied during a replication operation. When a data deduplication system identifies a block it has processed before, instead of storing the block again, it inserts a pointer to the original block in the data set's metadata. If the same block shows up multiple times, multiple pointers to it are generated but only the new, unique blocks are written to the storage pool.

HYBRID DEDUPLICATION BRIDGES GAP BETWEEN TARGET AND SOURCE-BASED ARCHITECTURES

Most deduplication systems carry out the deduplication operation either on a dedicated target appliance or on a source server that is also executing the backup operation. For full target mode operation, all the data is transmitted to the device and the deduplication occurs in the dedicated system. Purpose-built appliances like Quantum's DXi-Series are designed with deduplication-specific features to optimize the process. In the case of the DXi appliances, these include tiers of different storage devices for data pools and indexes to allow very high-speed inline performance both for backup and for reads. In bandwidth-constrained environments, however, the requirement to move all data to the appliance can limit overall performance.

In source-based deduplication, the deduplication occurs in software as part of the backup process, taking place either on the backup client at the time of initial backup or on a backup server. Full source-side deduplication can speed up backups where network bandwidth is the limiting factor because only unique data needs to be transmitted to the target device. The trade-off in source-based operation is that the shared backup hardware (the backup server or client) experiences a greatly increased load when it assumes the entire deduplication task—this increased load can slow down the backup and any other functions for which the server is responsible. To minimize this performance impact, source-based systems typically use a fixed-block deduplication method. Fixed-block deduplication is less compute-intensive than the variable-block system used by most target based systems, so it slows down backups less, but this comes at a price. Variable-block is much more effective at reducing data, so in comparison fixed-block requires more storage to protect the same data, block pools are restricted in size, and unique-block replication is often not supported. The end result is greater cost and complexity vs. variable-block systems.

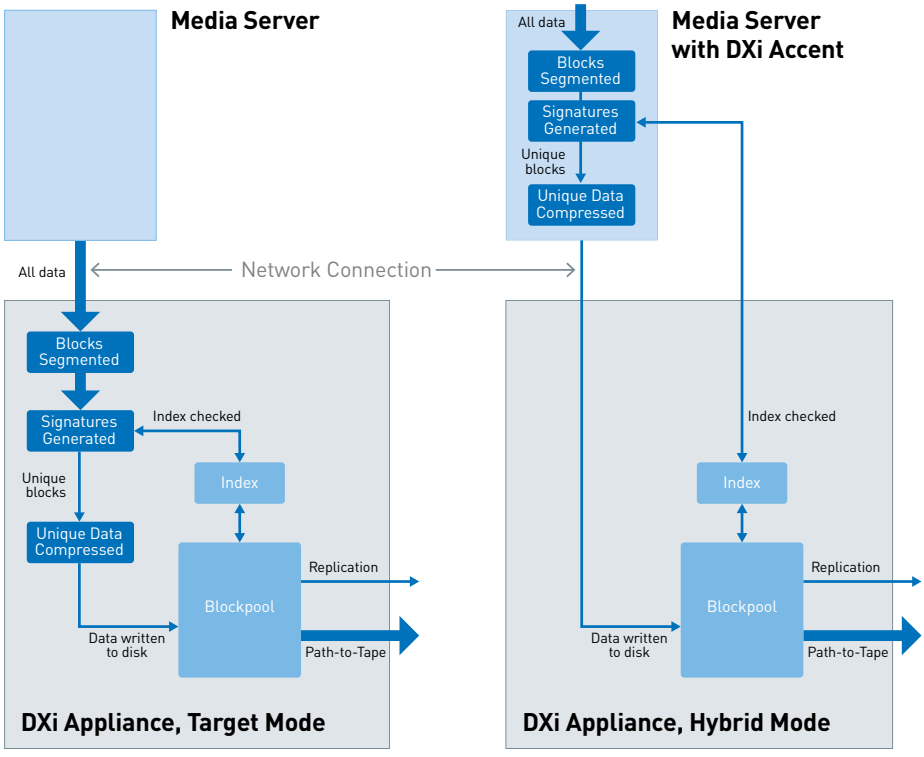
1. Quantum's deduplication technology is protected by Pending and Issued U.S. and Foreign Patent No. 5,990,810.

DXi Accent is software from Quantum that enables a hybrid or collaborative approach to deduplication, combining the best features of both target and source-based systems. DXi Accent uses variable-length deduplication for the most effective data reduction, and it takes advantage of purpose-built DXi appliances for scalability, performance, and ease of integration, but it moves a portion of the deduplication process to the backup server so that only unique blocks are transmitted to the target appliance. This system, which leverages much of the underlying functionality of the DXi replication, allows DXi Accent to accelerate backups where network bandwidth is the limiting factor while limiting the impact on the backup server and maintaining DXi features that integrate deduplication effectively into the larger data protection environment. The performance gains with DXi Accent in a particular customer environment vary with the amount of unique data in the data set and the nature of the bandwidth restrictions. For example, for a backup where 10% of the blocks are new, the potential effective transmission rate will be approximately 10 times more than when using a target-based approach alone. In general, DXi Accent does not increase the overall backup performance of DXi appliances for situations in which network bandwidth is not the constraining factor.

HIGHER PERFORMANCE, BANDWIDTH SAVINGS, MINIMIZING IMPACT ON THE BACKUP SERVER

With DXi Accent, the backup server collaborates in the deduplication process by carrying out the initial deduplication phases, specifically: 1) Dividing the stream of data into variable-length blocks and computing the signature for each one, 2) Collaborating with the DXi to identify the new unique blocks, and 3) Compressing the new unique blocks and transmitting them to the DXi appliance for storage in the blockpool. In order to determine the unique blocks, the signatures for all the blocks are sent by the server to the DXi appliance. The DXi compares the signatures to its central index and returns to the backup server a list of signatures for the unique blocks not already present in the blockpool. The backup server compresses these blocks and transmits them to the DXi to be stored. For blocks already present in the blockpool, the DXi simply stores a pointer to the existing block. This division of tasks between the backup server and DXi maximizes end-to-end performance while minimizing loading effects on the backup server because it leaves most of the processor-intensive tasks on the appliance, including index comparisons, data integrity verification, disk I/O, direct write to tape, remote replication, space reclamation, and blockpool maintenance. As a result, the backup server requirements for DXi Accent are significantly lower than for traditional source-based architectures, and much more data can be protected with the same resources. (Figure 1)

Figure 1.



USE CASES FOR DXi ACCENT AND HYBRID DEDUPLICATION

There are three primary use cases for DXi Accent: 1) Local servers on congested LANs, 2) Virtual server environments, and 3) Remote servers where WAN bandwidth and network latency are the limiting factors. Let's consider each of these scenarios in turn.

For local servers on congested LANs, DXi Accent's hybrid mode can accelerate backups by overcoming bandwidth limitations. By only sending new unique blocks to the target, much less data traverses the LAN, minimizing the effects of congestion. Backups that took too long may now be completed in their allotted windows without investing in network upgrades, and LAN capacity is freed for other applications.

Virtual server environments are nearly always network-constrained, often severely so. Every virtual machine (VM) has one or more virtual network interface cards (VNICs), all competing for host resources and sharing a small number of physical NICs. When using a Quantum DXi V-Series virtual appliance as a backup target, DXi Accent reduces congestion on the virtual networks just as it does with physical LANs.

For the third case, where servers are connected to the DXi over a WAN, the benefits are potentially even greater. With a traditional target-based approach, a target device (such as a deduplication appliance or a local tape library) must be deployed at the remote site to accept data from backup servers located there. That target then either creates tapes, which are then moved off-site for DR protection, or in the case of a deduplication appliance, replicates the unique data to another target device at a second site. DXi Accent makes it possible to entirely eliminate the need to deploy a hardware target at the remote site by allowing users to deploy DXi Accent on the backup server and send the backup data directly to a remote DXi over the WAN, whether that DXi is hosted at another customer site or in the cloud with a service such as Quantum's Q-Cloud™. Because DXi Accent is based on the same protocol used for DXi-to-DXi replication, it is exceedingly tolerant of the low bandwidth and high latency typical of WAN connections. This method has implications for the restore process (discussed below), but where it is a fit, it provides a low-cost alternative to traditional target deployment.

CONSIDERATIONS FOR RESTORE, DR PROTECTION, AND LONG-TERM RETENTION

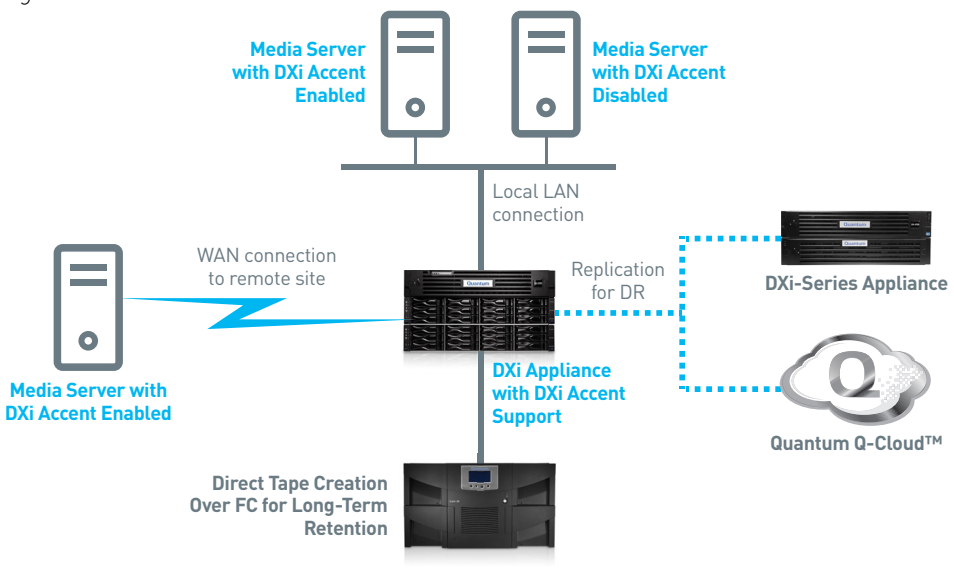
Users of hybrid backup architectures need to keep in mind the implications of restore profiles, DR protection needs, and long-term retention requirements when planning deployments. Although they let users send more backup data than the network's bandwidth would otherwise allow, any hybrid deduplication system—and many source-based systems—will have a different profile for restores. With these systems, all the data needing recovery is sent back along the same path as the backup, not just the unique blocks that were transferred during the backup. This means that more data must traverse the LAN or WAN during recovery and the process is likely to take more time. Where single-file or limited restores are the primary use case, the impact on restores will be minimal. In situations where large-scale restores are the norm or full-server recoveries must be accounted for, users need to understand the impact on restore times and plan accordingly.

One of the advantages of using DXi appliances is that DR protection and long-term retention can both be provided without having to move large amounts of data back over the network to the backup server. For automated DR protection, DXi systems provide remote replication to send duplicate backup sets to remote DXis. This process leverages the same variable-block deduplication technology to minimize network bandwidth and overcome latency, and it provides an automated process for creating off-site copies for DR purposes that has no impact on the backup server or the link between the server and the appliance. For long-term retention, and as an alternative for DR protection, DXi systems provide an integrated, direct path-to-tape, allowing them to send full data sets over a high-speed Fibre Channel connection, bypassing the media server and the network link from the server to the appliance to create removable media in a directly attached tape library. Data on tape can be used for DR protection, for large-scale restores at a remote site, or for low-cost storage for data that needs to be maintained for extended periods. Application-aware replication and direct tape creation are provided for DXi Accent through Symantec NetBackup's OpenStorage (OST) interface on supported platforms.

MAXIMUM FLEXIBILITY

Unlike typical source-based systems or other hybrid approaches, DXi Accent was designed to provide maximum flexibility to make distributed deduplication a practical part of extended backup environments. Its underlying transport protocol is designed for effective operation even in high-latency environments so it can support connections both to local servers over a LAN and also to remote servers over a WAN. The DXi appliance and DXi Accent software solution was designed to allow mixed operation—hybrid mode operation can be enabled or disabled on a server-by-server basis—allowing it to be deployed in a wide range of use cases and avoiding the downsides of an all-or-nothing solution. In addition, DXi Accent works with the DXi appliances to provide DR protection and low-cost, long-term retention through replication and the integration of DXi's direct path-to-tape. This combination of local and remote operation, server-by-server enablement, DR protection, and integrated long-term retention is an industry-unique combination. (Figure 2)

Figure 2.



DXi Accent is supported in both NetBackup and Backup Exec OST environments but is designed to be extensible to other applications. For details on current application and platform support, visit **www.quantum.com**

TB00017A-v04 July 2014