IBM ProtecTIER Deduplication for z/OS

John Webster
March 04, 2010
Announcement Summary

The many data deduplication technologies available today have the ability to dramatically lower enterprise IT costs for both storing and moving data. “Dedupe” has become widely used in open systems environments to, in some cases, significantly reduce storage capacity that would otherwise be required for backup, archive, and even primary storage supporting critical applications. Until now, the only way mainframe storage administrators could take advantage of this increasingly popular technology was to insert an ESCON/FICON-to-TCP/IP channel emulation device into a data stream between a mainframe channel and an open systems virtual tape library (VTL) that supports data deduplication (Bus-Tech MDL 100V + FalconStor VTL for example).

With the announcement of the IBM System Storage TS7680 ProtecTIER Deduplication Gateway for System z (TS7680), IBM now offers its mainframe customers an advanced data deduplication solution that can be used for a number of application scenarios including backup and other data stream-intensive applications where data is first streamed to tape for subsequent processing. One of the results of implementing data deduplication on System z is that a variety of disk platforms, both current generation and legacy, can now be considered as a cost-effective storage platform for these types of applications as compared to tape.

IBM acquired Diligent Technologies Corporation in April, 2008. IBM subsequently introduced a number of IBM branded products including the IBM System Storage TS7650 Appliance and TS7650G Gateway based on ProtecTIER with HyperFactor® (discussed in more detail below1) and has installed under the IBM logo more than 600 ProtecTIER solutions for open systems. With the announcement of the TS7680, IBM extended its portfolio of enterprise-class data deduplication solutions by providing one for System z environments based on proven technology.

Dedupe addresses capacity, performance, and bandwidth issues

Data deduplication is able to dramatically decrease the amount of disk space required for backup data when disk is used as a backup target, while retaining the significant performance improvements that disk based backup devices have over tape. Thus, data deduplication should be considered for any IT environment looking to contain storage costs associated with backup, while preserving the delivery of required service levels for data protection. Some storage administrators have decided to replace tape with disk for applications requiring rapid access to data precisely because the cost per GB of deduplicated data on disk made it more affordable to maintain tape data on disk.

Business continuance and disaster recovery-related data replication processes within and outside of a system can also take significant amounts of time depending on the volume of data and the size of the interconnecting data “pipe.” Deduplicating the data objects within these replication streams to in many cases a small fraction of their original size will allow them to be moved in much less time. Reduced bandwidth requirement could also be translated into reducing communications costs between sites for replication-related data transfers.

---

1 See also Evaluator Group Announcement Summary of IBM’s TS7650 VTL Systems published February 9, 2009.
Post-process vs. Inline

The storage industry’s approach to data deduplication has evolved to the point where today there are essentially two different processes that yield deduplicated data objects. Real-time or streaming data deduplication is known as “in-line” while data deduplication that occurs later is commonly referred to as “post-process” deduplication. The in-line process deduplicates data “in flight” and in real time as it is being sent to a backup device for example. Post-processing refers to data deduplication performed at some point in time after the data has been sent to a storage device—a Virtual Tape Library (VTL) for example that runs deduplication after data has been stored.

As with most options, the optimal method to use depends upon the goals the storage administrator has in mind. Consider the backup process. Storage administrators looking to simply minimize the backup window often choose the post-process method. The potential advantage is that, because the deduplication process is not in the path of the data stream, there will be no performance impact during the write operation and therefore no elongation of the backup window. That is, backup data is sent to a temporary holding area within the disk array to negate potential performance impact. Once the backup job completes, the data is later examined for duplicates, with duplicate data removed at a later “post-process” time. The disadvantage of this method is that additional storage space is required when compared with the in-line process.

An alternative to deduplicating after a backup is to perform deduplication “in-line” as data is being sent to the backup device. The first advantage with this method is that no extra disk space is required. The data stored to disk is in deduplicated form right from the start. Second, no additional processing step to deduplicate the data is required. Another advantage of in-line processing is that once the data is de-duplicated and stored, deduplicated data may be replicated immediately to off-site storage. As a result, the time to complete the entire business continuance process—including backup—is reduced, and as mentioned earlier, the bandwidth and/or the time required to replicate is also reduced. As noted above, in some implementations in-line processing impacts performance and therefore backup time. IBM claims “negligible” performance impact due to using a light-weight index of no more than 4GB maximum that maps to the contents of the data repository supporting up to 1PB.”

---

2 Depending on implementation, a second backup may not be able to start until the post processing de-duplication completes.
3 EGI has not yet been able to validate this claim with ProtectTIER users
The TS7680 for IBM’s System z

The TS7680 is implemented as a gateway to disk arrays within a System z ESCON® or FICON® channel.

Figure 1: Data Deduplication for System z (Source: IBM and Evaluator Group)

Shown above in Figure 1 is a typical deployment of a TS7680 system to provide data deduplication and offline tape storage in a System z environment. As illustrated below in Figure 2, is a depiction of how the ProtecTIER TS7680 system operates between the System z host and the disk cache.

Figure 2: IBM TS7680 ProtecTIER Host Connectivity (Source: IBM and Evaluator Group)

Key points to bear in mind when evaluating the TS7680 include:

- Deduplication is performed in-line as described above.
- Components within the TS7680 solution include a single frame containing two clustered ProtecTIER servers for failover redundancy, FICON interfaces, and the ProtecTIER software. No System z host-resident software is required.
- Maximum capacity of the back-end disk array storage is 1PB meaning that the TS7680 supports up to 1PB of disk for storage of deduplicated data. If a deduplication ratio of 10:1 is assumed, one could expect to store 10PB of normally formatted data within this 1 PB space after
deduplication. Deduplication ratios can vary widely however depending on the amount of data redundancy encountered by the system. It is misleading to translate deduplication ratios seen in open systems environments to System z. It is also the case that data deduplication ratios can increase over time as the system processes an increasing amount of data, and consequently encounters more redundancy.

- Backend disk is Fibre Channel-attached and can be IBM System Storage DS8000®, IBM XIV® Storage Systems(SATA disk), IBM System Storage DS5000®/4000® for mid-range System z environments, and any combination of third-party disk arrays already supported for attachment to IBM’s TS7650G.
- The TS7680 emulates an automated tape library with IBM System Storage 3592 Model J1A tape drives and supporting MEDIA5 (3592 JA) cartridges.
- From the perspective of the storage administrator, the TS7680 is managed transparently using system-managed tape (SMStape) facilities. No host application, tape management, or JCL changes are required. Virtual tapes are returned to scratch processing after deletion. Alerts are sent to the administrator if available capacity is running low.
- Backend tape attachment is not supported. Data objects that need to be migrated to tape must first be “rehydrated” i.e. returned to normal format and then sent via the System z host to a tape device.

The HyperFactor Process

Storage vendors now offer a variety of ways to deduplicate data. As mentioned, the process can occur in-line or run sometime after data is stored. In addition, there are differing deduplication processes that can be applied. File level deduplication has been available for a number of years. Deduplication using hashing algorithms to generate a code that represents stored data objects is more recent, and now more common.

ProtecTIER’s HyperFactor uses a series of algorithms to identify elements within a data stream that have been previously stored by ProtecTIER. Once similar elements have been found, HyperFactor compares the new data to the similar data already stored and writes only the byte-level changes to disk. HyperFactor uses a memory resident index of no more than 4GB to identify similar data. A copy of the index is maintained on TS7680-attached disk. IBM reports a maximum measured throughput of 500 MB/s using HyperFactor’s data deduplication in-line processing.

Comparing the TS7680 to Other IBM System z Virtual Tape Solutions

IBM Virtualization Engine™ TS7700 Family

Although the TS7680 leverages disk storage capabilities, it does nevertheless emulate IBM’s 3592 tape and should be compared first to other IBM virtualized tape subsystems. While both the TS7720 and 7740 offer compression, they do not support or deliver the reduction in storage capacity that data deduplication is capable of. The TS7700 offerings do provide “Grid” replication functionality, which supports the replication of tape data between up to four sites. In addition, Grid supports capabilities
such as access to state-consistent tape volumes from any site. However, the TS7680 is planning a less sophisticated two site replication capability in a future release expected early next year.

<table>
<thead>
<tr>
<th>Feature</th>
<th>TS7680</th>
<th>TS7740</th>
<th>TS7720</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max. disk capacity (raw)</td>
<td>1PB</td>
<td>14TB or 56TB w/ 4-way grid</td>
<td>70TB or 280 TB w/ 4-way grid</td>
</tr>
<tr>
<td>Max. number of virtual drives supported</td>
<td>256</td>
<td>256 or 1024 (4-way grid)</td>
<td>256 or 1024 (4-way grid)</td>
</tr>
<tr>
<td>Max. number of virtual volumes supported</td>
<td>1M</td>
<td>1M</td>
<td>1M</td>
</tr>
<tr>
<td>Direct tape attachment</td>
<td>No</td>
<td>Yes (grid)</td>
<td>Yes when configured in TS7740 grid</td>
</tr>
<tr>
<td>Deduplication</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Device-to-device Replication</td>
<td>Future (see below)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table 1: Comparison of IBM ProtecTIER® TS780 and IBM Virtualization Engine™ Family

IBM VTF™ Mainframe

VTF Mainframe is based on software acquired in the Diligent Technologies acquisition. VTF Mainframe is z/OS® host-resident software that provides emulation of IBM and IBM-compatible cartridge devices and tape volumes and redirects tape-targeted data streams to ESCON/FICON channel-attached disk. It does not support HyperFactor deduplication, but it does support remote mirroring between storage devices and could be considered along with the TS7680 when there is a need to reduce the time required to run batch jobs that are heavy users of tape. Also unlike the TS7680, VTF Mainframe supports multiple concurrent access to a single tape data set (Parallel Access Tape).
<table>
<thead>
<tr>
<th>Feature</th>
<th>TS7680</th>
<th>VTF Mainframe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supported disk</td>
<td>IBM DS Series, IBM XIV, and/or any disk supported for attachment to ProtecTIER TS7650G</td>
<td>Any ESCON/FICON 3380 or 3390-compatible</td>
</tr>
<tr>
<td>Deduplication</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Max. number of virtual drives supported</td>
<td>256</td>
<td>256 Per LPAR</td>
</tr>
<tr>
<td>Max. Theoretical Factoring Ratio</td>
<td>25:1 (^4) (HyperFactor deduplication)</td>
<td>2:1 (standard compression)</td>
</tr>
<tr>
<td>Native tape attachment</td>
<td>No</td>
<td>N/A (Runs as z/OS-resident software which directs tape data stream to disk)</td>
</tr>
<tr>
<td>DFSMS Support</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Replication</td>
<td>Future – TS7680 to TS7680 (see below)</td>
<td>Yes – between ESCON/FICON-attached 3380/3390 compatible disk subsystems</td>
</tr>
<tr>
<td>Maximum Physical Disk Capacity/System</td>
<td>1PB</td>
<td>No limit other than that imposed by z/OS</td>
</tr>
<tr>
<td>Parallel Access Tape</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Tape stacking support</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table 2: Comparison of IBM ProtecTIER® TS780 and IBM VTF™ Mainframe

Replication as a Future Deliverable

As part of this announcement, IBM also announced planned support for TS7680 device to device replication. This will be a significant enhancement to the TS7680 product set in that it will deliver the benefits of deduplication to business continuance and disaster recovery planners. During replication, only the deduplicated data will be sent from a primary site to a secondary site over the communications link between the two, be it LAN, MAN, or WAN. This capability could reduce the overall cost of a robust business continuance plan—one that also includes disaster recovery capabilities. Indeed, the ability to send deduplicated data between sites could put a more robust DR plan within reach of organizations that cannot now afford one.

Replication will be configured at the tape volume level i.e. the smallest data unit that will be sent between primary and secondary sites will be a tape volume. Replication can proceed before the volume is unloaded. Volumes will be visible to one active site at a time.

The trade-off here will be in determining whether or not to use the significant reduction in data transmitted between sites to reduce the cost of a DR-related communications link by reducing the bandwidth required, or to improve on recovery time objectives by maintaining the communications link.

---

\(^4\) Ratio highly dependent on the amount of time data resides within the target storage device and the degree of variability in the data stream. Some data streams dedupe better than others.
already in place. Under the right circumstances, a storage administrator could also consider eliminating the need to send physical tapes off-site.

**Conclusion**

IBM’s TS7680 delivers a form of data deduplication that is consistent with mainframe production environments. The inline deduplication process implemented here should have minimal impact on performance when data written to TS7680-attached disk. The fact that deduplicated data is immediately available for replication (once this capability is delivered) means that there is no impact to disaster other processes needing to use the replicated copies.

The TS7680 gives mainframe administrators another tool to improve service levels with disk based tape processing while repurposing tape for other longer-term storage requirements. The fact that the TS7680 supports some legacy disk arrays means that previous generation disk can now be used in place of tape to accelerate application performance.

The open system environment has enjoyed the benefits of deduplication for some time now. Mainframe customers looking to leverage those same benefits for an IBM solution now have an IBM option to evaluate.