

Data Protector on OpenVMS improving backup times

OpenVMS Technical Journal V18

Table of contents

Case scenario	2
The investigation.....	2
Data Protector settings.....	3
HP tests locally	4
The company tests	7
Summary	8



Data protector is not a native OpenVMS backup solution; however it can be used for multiple operating systems' data life-cycle goals. As there is very little benchmark data available for OpenVMS systems with Data Protector, it was decided to try and identify possible improvements using various techniques available within the product and from the OpenVMS operating system side. The research provides some useful results for consideration when time considerations threaten to become a problem. The document uses an actual case scenario to identify the options.

Case scenario

The company has two data centers and uses multiple hardware as well as multiple operating systems. The IT strategy is to eliminate physical tapes for data life cycle management and to use a single tool for data life cycle management. The single tool chosen is Data Protector (DP) and a SAN solution, such as a virtual tape library, to eliminate physical tapes and local drives.

DP was introduced more than two years ago and works fine for Microsoft® Windows® and UNIX® platforms. The throughput is not an issue in the data life cycle management for these platforms. However, the OpenVMS cluster platform with DP takes too long to complete. For redundancy purposes, not all the cluster nodes are in the same data center. OpenVMS platforms are used primarily with a large database application. One of the disk volumes has approximately 700 GB of data and the total disk volumes data is more than 1.5 Terabyte. The disks are SAN disks, mostly shadowed between the two data centers.

Using DP, they achieve a throughput of approximately 11 MB/sec but need to achieve 20 MB/sec to meet their time frame window when they can backup data.

At present they use physical tapes and local tape units for OpenVMS systems with the OpenVMS Archive Backup layered product and want to move to an integrated SAN solution and stop using physical tapes for the OpenVMS platform.

To illustrate how poor throughput affects time frames: The DP poor throughput was a volume size—143.24 GB 1 Disk; Agent running runtime was—3:29:56; and the throughput was—11 MB/sec.

What is causing the poor throughput and can DP be used for OpenVMS platforms to achieve their IT strategy?

How can this be investigated and tested?

The investigation

To try and isolate a cause or identify room for improvement, the data end-to-end flow chain needs to be specified and examined for any potential bottlenecks. Data Protector communicates via a disk agent to a media agent using TCP over the network to transfer data. The disk agent runs on the OpenVMS node and the media agent runs on a Windows server with a virtual tape library system. The disk data is on a SAN that uses EVA storage.

Some possible bottlenecks could be:

1. The disk agent read I/O's on the OpenVMS system from the EVA storage
2. Resource problems on the OpenVMS cluster
3. The network data transfer between the disk agent and the media agent
4. The media agent write I/O's to the virtual tape library

OpenVMS has a native backup facility built into the operating system called Backup. DP does not use this native backup. Using the Backup utility to make physical and image backups of the disks to the OpenVMS null device can help to identify the maximum throughput of the data and the OpenVMS systems. The results obtained showed that 20 MB/sec was achievable. The Monitoring tool—T4 was installed on the OpenVMS cluster to gather statistics to see if there were resource problems with the Fiber Channel adapters, Disk IO, and network interfaces.

A test cluster was made available consisting of four nodes with SAN data. Initially, a DP backup was carried out to get baseline data. The backups consisted of all data, the disks being spread across a single disk agent per node. T4 was used to monitor the OpenVMS systems, and statistics were gathered from the EVA and VLS hardware. The EVA statistics and the VLS showed no bottlenecks and had capacity over to handle more data. The OpenVMS T4 data showed that the network interfaces, the disk IO rates, and fiber channel adapters were not a bottleneck. It handled the volume of data without problems. The DP backup however was only delivering data at 11 MB/sec.

To see if TCP was a possible problem, TTCP tests were done on the OpenVMS system and also between OpenVMS and a windows media agent system. This also showed that 20 MB/sec was attainable.

What features are available within DP to get it to pipe more data and increase the throughput?

- Use multiple disk agents and do directory backups of disks
- Have the disk and media agent running on OpenVMS
- Change the settings in the DP software for the OpenVMS nodes
- Possibly have multiple agents on a node doing specific disk backups

When tests were done using multiple disk agents the DP throughput increased significantly. However the company found that this was not an option (for instance, directory backups) as any change on the OpenVMS disk structure would mean manually changing the DP backup templates. If template changes were not done, the risk was too high that data would not be backed up.

Using the disk and media agent both on OpenVMS showed an improved throughput from 11 MB/sec to approximately 16 MB/sec. However for security and management purposes, the HP Virtual Library System is presented via the SAN to a single windows server. This establishes a central point of control for use of the device. Thus, having both agents on OpenVMS nodes is not a valid option for the company.

Data Protector settings

Do not preserve access time attributes

This feature modifies the file headers on the OpenVMS system to the time it was last accessed. It is part of the backup template defined to run the DP backups. The disk agent does this action and switching it off may speed up the disk agent.

First tree walk disabled

If millions of small files are backed up, the first tree walk could take considerable time and increase the overall runtime. The tree walk can be disabled by setting "NoTreeWalk=1" in the configuration file.

```
"<Data_Protector_home>\omnirc."
```

```
# OB2NOTREEWALK=0|1
```

```
# Default: 0
```

```
# Disables the first treewalk (Windows only). This is only applicable if
```

```
# hard links are not enabled. Setting this variable may improve
```

```
# performance on filesystems with many small files. If set, the progress
```

```
# is no longer displayed in percentage.
```

Block size KB

This is part of media device settings available under advanced options.

The Disk Agent (DA) record size limit is always set to 56 KB.

Matching the media agent device to, say 56 KB, has been seen to speed up the throughput.

Can the network transfer be influenced by Network Interface Controllers (NIC) settings?

The NICs on the OpenVMS nodes are capable of running up to 1 Gigabit running up to 1 Gigabit, although it was observed that they do not use not using jumbo frames.

Another test system was configured locally at HP to see what the effect of modifying the DP settings and OpenVMS NICs had on the throughput.

HP tests locally

Local tests were conducted using an **HP AlphaServer E47 system** under **HP OpenVMS Alpha Version 7.3-2-OpenVMS Systems**, with multiple NICs capable of 100 megabit/sec and 1000 megabit/sec.

This meant jumbo frames could also be tested as DP uses TCP to transport the data between the OpenVMS Disk Agent and the Media Agent system.

Two virtual SAN disks (EVA5000) were created seen in OpenVMS as DGA930 and DGA931 each with 100 GB data.

The test methods used were to identify potential benefits of different configurations of DP so that the company could achieve approximately 20 MB/sec backup throughput. The windows null device was used as media device.

Figure 1: Single disk DGA930, 1 DA

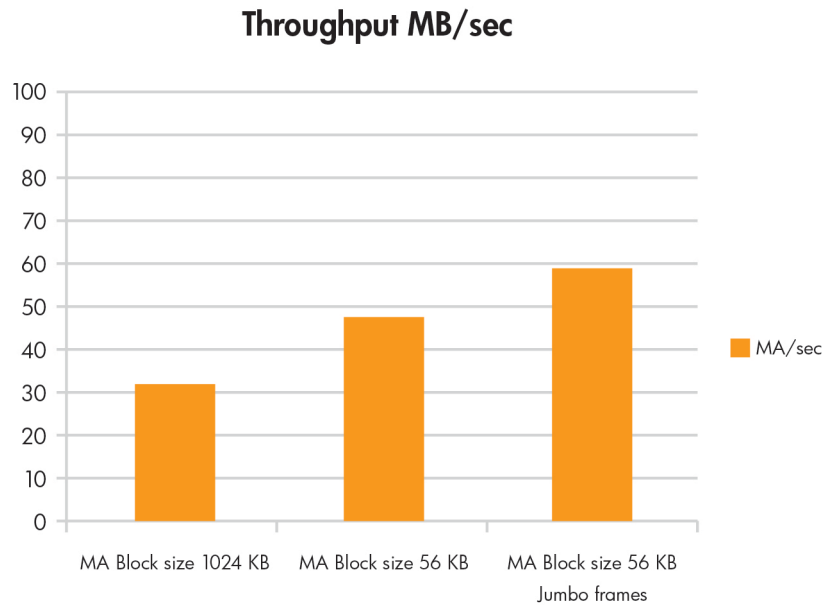


Figure 2: DGA930 single disk, 1 DA

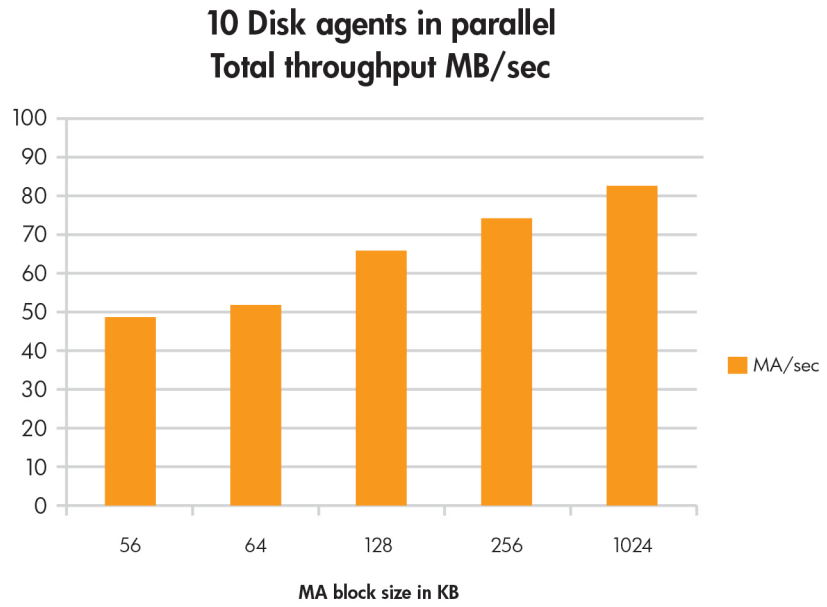


Figure 3: DGA930, 4 DA parallel, four times same disk

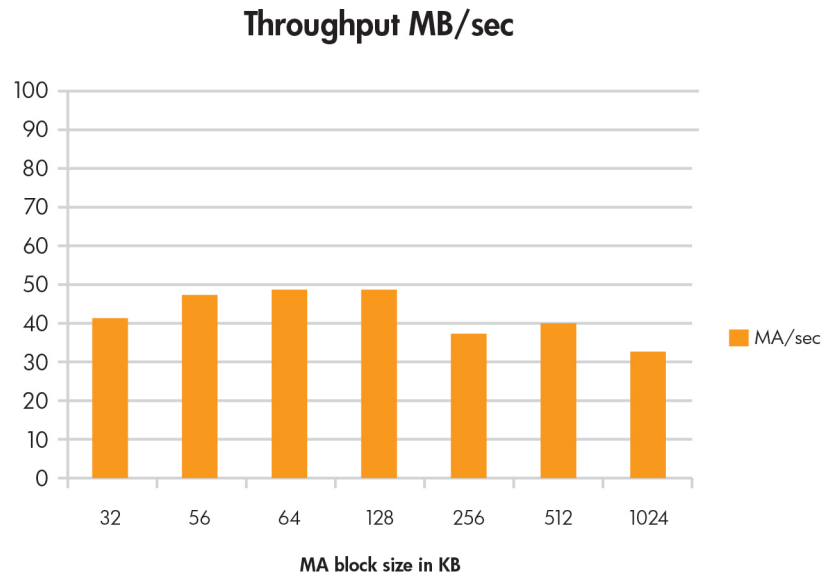


Figure 4: DGA930 and DGA931, 10 DA parallel, five times same disk

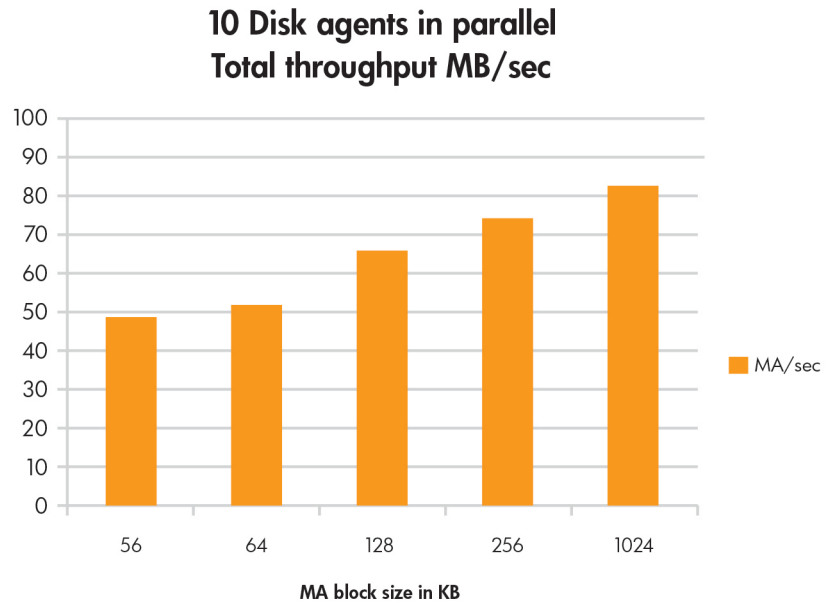


Figure 5: DGA930 and DGA931, 10 DA parallel, five times same disk

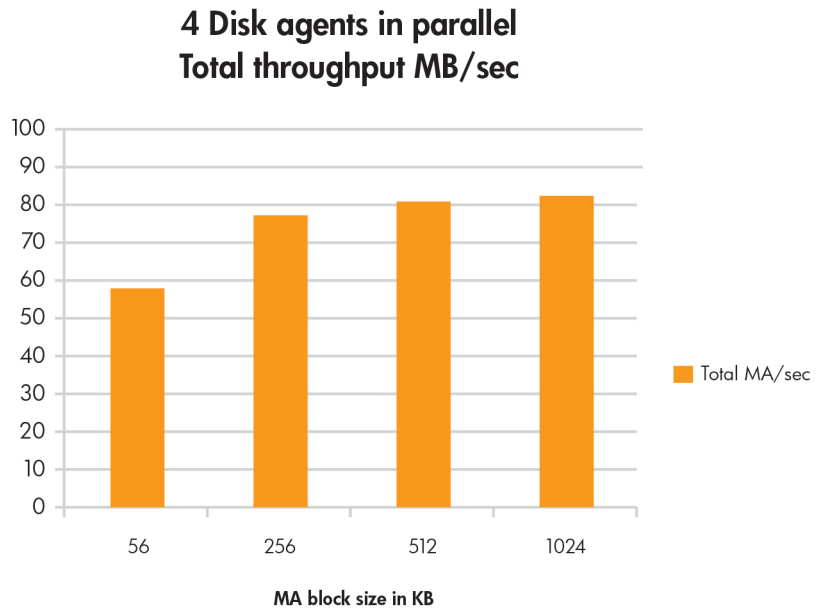
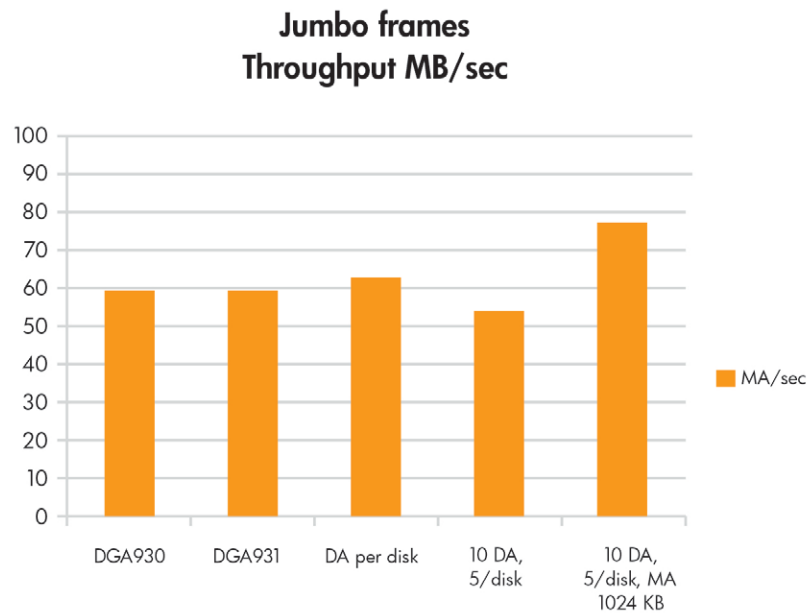


Figure 6: DGA930 and DGA931, MA block size 56 KB with jumbo frames



The evidence suggested that the following 3 factors can improve the Data Protector throughput:

- Modifying the media agent block size used
- Using multiple Disk Agents on a node
- Using Jumbo frames on the NIC

If the throughput can be increased, the backup window time can be reduced.

The company tests

The company tested having modified the Media Agent to 56 KB block size and formatted the media similarly. It was a test with four Disk Agents in parallel, one per node. Each agent had a subset of the total amount of disks and the total volume of data was approximately 1.5 TB.

Observations by the company convinced them of the following:

- The OpenVMS server has no problem serving the amount of data
- The network has no problem moving this amount of data (monitoring the transfer peaks of almost 60 MB/sec)
- The Media server had no problem receiving this amount of data

The conclusions from the test were:

- 30 MB/sec per agent is achievable
- The disk organization (number of files and size) had a large influence on the throughput
- On the production system, the largest disk has a large amount of data in a few files, which is an advantage for the throughput.
- A backup of 714 GB with 30 MB/sec would have a window time of 6.6 hours, which falls well within the acceptable time frame

Summary

The company has decided to modify the media agent block size to 56 KB, to flag "Do not preserve access time attributes," and to proceed to test on the production systems.

Using these settings, Data Protector can possibly now be used as a single tool, as part of an integrated SAN solution and they can eliminate physical tapes and local tape drives.

WARNING

Using the above options showed the company under study, some possibilities on how to reduce their DP backup times. However, system resources and workloads for each system are unique, so there is no single fit to match all cases. It does, however, identify different options that can be tested in individual environments.

Other possible options:

- Use a standalone OpenVMS system for DP backups
- Multiple disk agents doing complete disk backups

Discover how Data Protector fits as a single backup tool for an integrated SAN solution for Data Life Cycle Management, if OpenVMS is one of your platforms.

<http://h71000.www7.hp.com/openvms/storage/enterprise.html>

For detailed information on HP Data Protector support matrices, please visit: <http://www.hp.com/go/dataprotector>



Get connected

www.hp.com/go/getconnected

Current HP driver, support, and security alerts
delivered directly to your desktop

© Copyright 2012 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

Microsoft and Windows are U.S. registered trademarks of Microsoft Corporation. UNIX is a registered trademark of The Open Group.

4AA4-0524ENW, Created April 2012

