Abstract
The number of use cases for Dell EMC Data Domain protection storage has grown over time and overall Data Domain maximum capacity per system has increased. The use of shared Data Domain system services by large Enterprises and Service Providers has also become more commonplace. For these reasons, Data Domain customers have been looking for a mechanism to more easily and effectively report on and manage consumption of Data Domain physical capacity. This paper provides an overview of Data Domain physical capacity measurement capabilities which can be used to facilitate chargeback/billing, capacity planning, and data migration planning.

August, 2017
# TABLE OF CONTENTS

**EXECUTIVE SUMMARY** .................................................................................................................. 3  
  The Challenge ................................................................................................................................. 3  
  Solution Overview ......................................................................................................................... 3  

**INTRODUCTION** ............................................................................................................................ 3  
  Audience ......................................................................................................................................... 3  

**PHYSICAL CAPACITY MEASUREMENTS** ...................................................................................... 3  
  Background ...................................................................................................................................... 4  

**CUSTOMER USE CASES** .............................................................................................................. 5  
  Chargeback/Billing Metrics ............................................................................................................ 5  
  Capacity Planning ........................................................................................................................... 5  
  Data Migration and Replication Planning ...................................................................................... 5  
  Identification of Datasets Achieving Poor Deduplication ............................................................. 6  

**BEST PRACTICES** .......................................................................................................................... 6  
  Reporting on Subsets of an MTree ................................................................................................. 6  
  Secure Multi-tenancy Chargeback ............................................................................................... 7  
  Additional Best Practices Notes .................................................................................................... 9  

**PRODUCT INTEGRATION AND USER INTERFACE MATRIX** ....................................................... 9  

**CUSTOMER BENEFITS** ................................................................................................................. 10  

**CONCLUSION** ................................................................................................................................ 10  

**ADDITIONAL RESOURCES** .......................................................................................................... 10
Executive Summary

The Challenge
The number of use cases for Data Domain protection storage has grown over time and overall Data Domain maximum capacity per system has increased. The use of shared Data Domain system services by large Enterprises and Service Providers has also become more commonplace. For these reasons, Data Domain customers have been looking for a mechanism to more easily and effectively report on and manage consumption of Data Domain physical capacity. This paper provides an overview of Data Domain physical capacity measurement which can be used to facilitate chargeback/billing, capacity planning, migration planning, and a way to identify datasets that are not deduplication efficient.

Solution Overview
Dell EMC® Data Domain physical capacity measurement provides Data Domain Enterprise customers and Service Providers with an effective mechanism for managing shared Data Domain protection storage capacity between individual departments or tenants. Physical capacity measurement enables efficient chargeback/billing, capacity planning, migration planning, and can help identify individual datasets that are not achieving a high degree of deduplication efficiency.

Introduction
The purpose of this white paper is to describe how Enterprise customers and Service Providers can leverage Dell EMC Data Domain physical capacity measurement and its reporting capabilities along with secure multi-tenancy features to more efficiently manage shared Data Domain protection storage capacity for the purpose of improved chargeback/billing, capacity planning, migration planning, and to help identify individual datasets that are not achieving a high degree of deduplication efficiency. This paper includes also key technical considerations and best practices.

Audience
This white paper is intended for Data Domain customers’ (particularly Enterprises and Service Providers implementing data protection as a service) technical IT staff, Dell EMC and Partner SEs, and anyone else looking to better understand how to more efficiently manage shared Data Domain system capacity using physical capacity measurement.

Physical Capacity Measurement
Data Domain physical capacity measurement measures the physical capacity consumed by a subset of files within the file system, based on how the files in the subset deduplicate with other files in the subset. Said differently, it measures the physical capacity that would be consumed on a Data Domain system by a set of files, if that set of files were the only files on the Data Domain system. This is a point in time measurement, based on when the measurement is requested.

You can specify the file system subset to measure in several ways: as a pathset (a set of files and directories), an MTree, a tenant unit (all files within a tenant unit), or a tenant (all files within a tenant).

The Data Domain system maintains a historical record of physical capacity measurements, which is available in tabular, graphical, and Excel spreadsheet formats, depending on the UI being used.

For example, Data Domain Management Center provides a tenant view of the capacity utilization of its managed Data Domain systems. This view shows the logical and physical capacity consumed by each tenant, its tenant units, and its MTrees (see Figure 1),
Background
With Dell EMC Data Domain secure multi-tenancy, a Data Domain system can isolate and securely store the backup and/or replication for multiple tenants. Each tenant has logically secure and isolated data and control paths on the Data Domain system.

MTree(s) and DD Boost storage unit(s) are allocated to each tenant to store their data.

Tenant units are a fundamental unit of multi-tenant organization on a Data Domain system. One or more tenant units are created for each tenant, and each tenant's MTree(s) and DD Boost storage unit(s) are then assigned to the tenant's tenant-unit(s).

Data access to each MTree is restricted to the owning tenant by configuring the protocol (DD Boost, CIFS, NFS, etc.) used to access each MTree.

Tenant administrative access to tenant units and the MTrees, which each tenant unit contains, is restricted by assigning management users or groups (AD or NIS) roles on the tenant's tenant units, and then providing the appropriate user credentials to each tenant (see Figure 2),

![Figure 1. DD Management Center Physical Capacity Measurement Report](image1.jpg)

![Figure 2. Secure Multi-tenancy Overview](image2.jpg)
In addition to tenant units, tenant objects can also be created on Data Domain systems. Tenant objects are a hierarchical object on top of tenant units which are used to group the tenant units belonging to tenant together. The same tenant object can be created on multiple Data Domain systems to track all of the resources (tenant units, MTrees, etc.) used by a tenant across multiple Data Domain systems. Tenant objects are also used to enforce that data can only replicated or fast copied from and to MTrees that belong to the same tenant.

For additional information on secure multi-tenancy for Data Domain systems please refer to the Why Secure Multi-tenancy with Data Domain Systems white paper.

Customer Use Cases
Data Domain physical capacity measurement provides tremendous customer value for Enterprise customers and Service Providers with 4 primary use cases which are described in more detail in the paragraphs below,

- Chargeback/billing metrics
- Capacity planning
- Data migration planning
- Identification of datasets achieving poor deduplication efficiency

Chargeback/ Billing Metrics
This use case refers to situations where backup admins can measure how much capacity is used per tenant or department and charge them accordingly. For instance, in a large Enterprise, a backup admin can implement chargeback as a policy in IT as a Service (ITaaS) segments in which each department or division is charged based on the utilization of a Data Domain system. With physical capacity measurement, chargeback can be implemented based on physical space utilization. Therefore, departments that are using more physical space can be charged more than departments that are using less physical space.

Similarly, physical capacity measurement can be used to provide a physical capacity billing metric to Service Providers. In this case, a group of tenants is sharing a Data Domain system that is owned by the service provider. Periodically, the service provider can use physical capacity measurement to obtain the amount of physical space being used by each tenant, so that a charge for their physical capacity consumed can be included in their bill.

Capacity Planning
If the capacity on a Data Domain system is being consumed at a fast pace, the Data Domain admin wants to understand which tenant is consuming most of the storage. In the capacity planning use case the Data Domain system may use physical capacity measurement to generate reports on physical space consumption rates by tenant. Based on these reports, the Data Domain admin can forecast how much physical space each tenant will consume at some time in the future. This can be used to plan expansions of physical capacity for Data Domain systems and to plan migrations of tenants to different Data Domain systems.

Data Migration and Replication Planning
Some tenants with similar workloads deduplicate very well among themselves, while other tenants may not have a lot of physical data in common. Ideally, Data Domain admins want tenants with a lot of common/shared data to use the same Data Domain system, since they will achieve better total compression. In the event that the Data Domain admin needs to move a tenant, or a group of tenants, from one Data Domain system to another (for example, because the available space on the Data Domain system is getting low), it is useful to know how much extra physical space that tenant, or group of tenants will consume on the destination Data Domain system.

Furthermore, in the case of replication, customers are interested in knowing how much physical space a tenant’s data will consume on the destination Data Domain system. This information will assist the decision on where to replicate data and how to best utilize the physical capacity of the destination Data Domain system.

Identification of Datasets Achieving Poor Deduplication
Physical capacity measurement can also help to identify datasets that are not achieving a high rate of deduplication efficiency. Imagine a scenario where in a highly-consolidated environment with a mixture of several different workloads there could be few clients that have a lot of multimedia audio or video files that may be causing a lower overall deduplication ratio on the Data Domain system. In such a case, a backup or storage admin would want to identify such poorly deduplicating datasets and possibly send them to a non-deduplicating file system.

**Best Practices**

The following section provides some best practice recommendations about using physical capacity measurement for specific scenarios and use cases.

**Reporting on Subsets of an MTree**

If you anticipate needing to measure the physical capacity of subsets of an MTree, for example the directories used by specific clients as specified in pathsets, it is best to periodically (probably on a weekly basis) measure the entire MTree, since the Data Domain system caches the physical capacity of the files that it samples. This can make the measurements of any part of the MTree faster to complete. Caveat: this performance benefit lessens as the number of files and the churn in the MTree increases. Periodic measurements can be scheduled using DD Management Center, DD System Manager or the CLI.

Periodic measurements of MTrees make it possible to see historical trends (graphed in DD System Manager and DD Management Center) that are useful to help plan for future migrations, and to identify deviations (e.g., sharply lower deduplication ratios) that need to be investigated and potentially corrected. The chart in Figure 3 shows logical capacity, physical capacity and associated compression.

![Figure 3. Periodic Measurement of MTrees](image)

For example, if two client hosts are backing up to the client1, and client2, respectively, subdirectories of MTree /data/col1/m0, then define 2 pathsets, 1 for each client, and use the CLI to submit a measurement for each pathset. In the example below, which shows the creation and measurement of the pathsets client1 is sending data that deduplicates...
well (a 16.43 compression ratio), but client2’s data does not deduplicate well (a 1.29 compression ratio). With this information you know to investigate what data client2 is sending to the Data Domain system. (see Figure 4),

Figure 4. Using Pathsets to measure client host deduplication ratios example

Secure Multi-tenancy Chargeback
For secure multi-tenancy chargeback first, use DD Management Center to schedule the tenant physical capacity measurements needed for your tenant bills. For instance, suppose you bill based on physical capacity consumed on the 1st, 15th and last day of the month for tenants T1 and T2 (see Figure 5).

Figure 5. Secure Multi-tenancy Chargeback Step 1
Next, use DD Management Center to schedule the tenant usage reports that cover your billing period (e.g., monthly) (see Figure 6).

![Add Report Template](image)

**Figure 6. Secure Multi-tenancy Chargeback Step 2**

Next at the beginning of each month, import the usage data from the excel spreadsheet usage report into your billing program.

Since tenant physical capacity measurements also include separate measurements of each of the tenant’s tenant units, and each of the tenant’s MTrees, there is no need to schedule separate measurements for the tenant’s tenant units and MTrees for chargeback purposes. For example, in Figure 7 below, DD Management Center’s Tenant Usage Report section shows the physical capacity used by each of a tenant’s tenant units over the period covered by the report.

![Physical Capacity Used (Post-Comp)](image)

**Figure 7. DD Management Center Tenant Usage Report**
Additional Best Practice Notes
The following are some additional best practice notes for using Data Domain physical capacity measurement,

- Only 3 measurement samples can be running at the same time on the same MTree, so plan your measurements accordingly.
- For sets of files that contain many (>5 million) small (<100K) files, it is recommended to run physical capacity measurement jobs weekly rather than monthly in order to ensure each measurement completes more quickly.
- If the total logical size of a set of files being measured is less than 100GB, the reported physical capacity can be off by more than 1%.
- It is more efficient to run many physical capacity measurements jobs at once, than to run those jobs serially.
- While file system cleaning is running, physical capacity measurement jobs are suspended and will complete after file system cleaning is done.
- When a physical capacity measurement job is started, a snapshot is created for each MTree on which the measurement will be performed. This ensures that the measurement will report the physical capacity at the desired point in time. If file system cleaning is running, all required snapshots are created, but the measurement will be done after file system cleaning is done. If a measurement is required as input to generate a bill, this delay should be taken into consideration since it might affect the timing of when bills can be sent to tenants.

Product Integration and User Interface Matrix
Data Domain physical capacity measurement is available from other Dell EMC products. The following matrix summarizes the PCM features available using each Data Domain UI CLI (DDSM), and using other Dell EMC products (DDMC, Avamar, DPA).

<table>
<thead>
<tr>
<th>Physical Capacity Measurements (PCM) Manageability Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PCM Functions</strong></td>
</tr>
<tr>
<td>Schedule PCM Jobs</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Customer Benefits
Data Domain physical capacity measurement can dramatically simplify Data Domain protection storage operations and improve efficiencies for large Enterprises and Service Providers. Data Domain physical capacity measurement will help Enterprises & Service Providers,

- Perform chargeback billing based on physical Data Domain system capacity used.
- Increase awareness of Data Domain system capacity usage trending by department, MTreer, or tenant.
- Simplify Data Domain systems capacity planning.
- Provide physical capacity usage information needed for data migration planning.
- Identify individual Data Domain datasets that are not achieving a high degree of deduplication efficiency

Conclusion
This paper has explained how Dell EMC Data Domain physical capacity measurement provides Data Domain Enterprise customers and Service Providers with an effective mechanism for managing shared Data Domain protection storage capacity between individual departments or tenants. Physical capacity measurement enables and greatly simplifies efficient chargeback/billing, capacity planning, data migration planning, and can help identify individual datasets that are not achieving a high degree of deduplication efficiency.

Additional Resources


**Why Data Domain** – The Business Value of Data Domain

**Data Domain Data Invulnerability Architecture** – A Detailed Technical Review White Paper

**Data Domain Replicator** – A Detailed Technical Review White Paper