



Data model overview

OnCommand Insight

NetApp
March 08, 2024

Table of Contents

- Data model overview 1
 - OnCommand Insight data models 1
 - Data model fact and dimension tables 6
 - Colors used in data model elements 6
 - Using multiple data models in one report 6

Data model overview

OnCommand Insight provides several data models for use in report development. Each data model is an aggregation that summarizes data so that it can be queried and searched. For example, reports about capacity planning use the Capacity data model.

The OnCommand Insight enterprise reporting data models provide data elements and interactive relationships among data elements that yield business views of the data. Using the data elements and relationships, you can create reports using the IBM Cognos Analytics report generation tools recommended by NetApp.

OnCommand Insight also provides data marts that can be used to develop your own SQL queries. There is a distinction between these SQL query data marts and the data models used in reporting. The individual OnCommand Insight reporting data models use the underlying OnCommand Insight database schema provided in the data marts; however, the data models use additional tables and sometimes new elements in the tables. For instance, the data model includes a Monthly Capacity Fact table in the Storage Capacity data model that is based on the Capacity Fact table from the database schema and data mart. The data model filters out the values from the database schema table to show only month information.

Another example of a difference between the database schema used in data marts and the data model is in the Violation table and the Violation Type column. The data model translates programmatic-named values in the database to match the text displayed in the OnCommand Insight web UI.

OnCommand Insight data models

OnCommand Insight includes several data models from which you can either select predefined reports or create your own custom report.

Each data model contains a simple data mart and an advanced data mart:

- Simple data mart provides quick access to the most commonly used data elements and includes only the last snapshot of Data Warehouse data; it does not include historical data.
- Advanced data mart provides all values and details available from the simple data mart and includes access to historical data values.
- **Capacity data model**

Enables you to answer questions about storage capacity, file system utilization, internal volume capacity, port capacity, qtree capacity, and virtual machine (VM) capacity. The Capacity data model is a container for several capacity data models. You can create reports answering various types of questions using this data model:

- **Storage and Storage Pool Capacity data model**

Enables you to answer questions about storage capacity resource planning, including storage and storage pools, and includes both physical and virtual storage pool data. This simple data model can help you answer questions related to capacity on the floor and the capacity usage of storage pools by tier and data center over time.

If you are new to capacity reporting, you should start with this data model because it is a simpler, targeted data model. You can answer questions similar to the following using this data model:

- What is the projected date for reaching the capacity threshold of 80% of my physical storage?

- What is the physical storage capacity on an array for a given tier?
- What is my storage capacity by manufacturer and family as well as by data center?
- What is the storage utilization trend on an array for all of the tiers?
- What are my top 10 storage systems with the highest utilization?
- What is the storage utilization trend of the storage pools?
- How much capacity is already allocated?
- What capacity is available for allocation?

◦ **File System Utilization data model**

Enables you to answer questions about file system utilization. This data model provides visibility about capacity utilization by hosts at the file system level. Administrators can determine allocated and used capacity per file system, determine the type of file system, and identify trending statistics by file system type. You can answer the following questions using this data model:

- What is the size of the file system?
- Where is the data kept and how is it accessed, for example, local or SAN?
- What are the historical trends for the file system capacity? Then, based on this, what can we anticipate for future needs?

◦ **Internal Volume Capacity data model**

Enables you to answer questions about internal volume used capacity, allocated capacity, and capacity usage over time:

- Which internal volumes have a utilization higher than a predefined threshold?
- Which internal volumes are in danger of running out of capacity based on a trend?
- What is the used capacity versus the allocated capacity on our internal volumes?

◦ **Port Capacity data model**

Enables you to answer questions about switch port connectivity, port status, and port speed over time. You can answer questions similar the following to help you plan for purchases of new switches:

- How can I create a port consumption forecast that predicts resource (port) availability (according to data center, switch vendor and port speed)?
- Which ports are likely to run out of capacity, providing data speed, data center, vendor and number of Host and storage ports?
- What are the switch port capacity trends over time?
- What are the port speeds?
- What type of port capacity is needed and which organization is about to run out of a certain port type or vendor?
- What is the optimal time to purchase that capacity and make it available?

◦ **Qtree Capacity data model**

Enables you to trend qtree utilization (with data such as used versus allocated capacity) over time. You can view the information by different dimensions—for example, by business entity, application, tier, and service level. You can answer the following questions using this data model:

- What is the used capacity for qtrees versus the limits set per application or business entity?
- What are the trends of our used and free capacity so that we can do capacity planning?
- Which business entities are using the most capacity?
- Which applications consume the most capacity?

- **VM Capacity data model**

Enables you to report your virtual environment and its capacity usage. This data model lets you report on changes in capacity usage over time for VMs and data stores. The data model also provides thin provisioning and virtual machine chargeback data.

- How can I determine capacity chargeback based on capacity provisioned to VMs and data stores?
- What capacity is not used by VMs and which portion of unused is free, orphaned, or other?
- What do we need to purchase based on consumption trends?
- What are my storage efficiency savings achieved by using storage thin provisioning and deduplication technologies? Capacities in the VM Capacity data model are taken from virtual disks (VMDKs). This means that the provisioned size of a VM using the VM Capacity data model is the size of its virtual disks. This is different from the provisioned capacity in the Virtual Machines view in OnCommand Insight, which shows the provisioned size for the VM itself.

- **Volume Capacity data model**

Enables you to analyze all aspects of the volumes in your environment and organize data by vendor, model, tier, service level, and data center. You can view the capacity related to orphaned volumes, unused volumes, and protection volumes (used for replication). You can also see different volume technologies (iSCSI or FC), and compare virtual volumes to non-virtual volumes for array virtualization issues. You can answer questions similar to the following with this data model:

- Which volumes have a utilization higher than a predefined threshold?
- What is the trend in my data center for orphan volume capacity?
- How much of my data center capacity is virtualized or thin provisioned?
- How much of my data center capacity must be reserved for replication?

- **Chargeback data model**

Enables you to answer questions about used capacity and allocated capacity on storage resources (volumes, internal volumes, and qtrees). This data model provides storage capacity chargeback and accountability information by hosts, application, and business entities, and includes both current and historical data. Report data can be categorized by service level and storage tier.

You can use this data model to generate chargeback reports by finding the amount of capacity that is used by a business entity. This data model enables you to create unified reporting of multiple protocols (including NAS, SAN, FC, and iSCSI).

- For storage without internal volumes, chargeback reports show chargeback by volumes.
- For storage with internal volumes:
 - If business entities are assigned to volumes, chargeback reports show chargeback by volumes.
 - If business entities are not assigned to volumes but assigned to qtrees, chargeback reports show chargeback by qtrees.
 - If business entities are not assigned to volumes and not assigned to qtrees, chargeback reports

show the internal volume.

- The decision whether to show chargeback by volume, qtree or internal volume is made per each internal volume, so it is possible for different internal volumes in the same storage pool to show chargeback at different levels. Capacity facts are purged after a default time interval. For details, see Data Warehouse processes.

Reports using the Chargeback data model might display different values than those reports using the Storage Capacity data model.

- For storage arrays that are not NetApp storage systems, the data from both data models is the same.
- For NetApp and Celerra storage systems, the Chargeback data model uses a single layer (of volumes, internal volumes, or qtrees) to base its charges, while the Storage Capacity data model uses multiple layers (of volumes and internal volumes) to base its charges.

• **Inventory data model**

Enables you to answer questions about inventory resources including hosts, storage systems, switches, disks, tapes, qtrees, quotas, virtual machines and servers, and generic devices. The Inventory data model includes several submarts that enable you to view information about replications, FC paths, iSCSI paths, NFS paths, and violations. The Inventory data model does not include historical data. Questions you can answer with this data mart could include the following:

- What assets do I have and where are they?
- Who is using the assets?
- What types of devices do I have and what are components of those devices?
- How many hosts per OS do I have and how many ports exist on those hosts?
- What storage arrays per vendor exist in each data center?
- How many switches per vendor do I have in each data center?
- How many ports are not licensed?
- What vendor tapes are we using and how many ports exist on each tape?
- Are all the generic devices identified before we begin working on reports?
- What are the paths between hosts and storage volumes or tapes?
- What are the paths between generic devices and storage volumes or tapes?
- How many violations of each type do I have per data center?
- For each replicated volume, what are the source and target volumes?
- Do I have any firmware incompatibilities or port speed mismatches between Fibre Channel host HBAs and switches?

• **Performance data model**

Enables you to answer questions about performance for volumes, application volumes, internal volumes, switches, applications, VMs, VMDKs, ESX versus VM, hosts, and application nodes. Using this data model, you can create reports that answer several types of performance management questions:

- What volumes or internal volumes have not been used or accessed during a specific period?
- Can we pinpoint any potential misconfiguration for storage for an application (unused)?
- What was the overall access behavior pattern for an application?
- Are tiered volumes assigned appropriately for a given application?

- Could we use cheaper storage for an application currently running without impact to application performance?
- What are the applications that are producing more accesses to currently configured storage? When you use the switch performance tables, you can obtain the following information:
- Is my host traffic through connected ports balanced?
- Which switches or ports are exhibiting a high number of errors?
- What are the most used switches based on port performance?
- What are the underutilized switches based on port performance?
- What is the host trending throughput based on port performance?
- What is the performance utilization for last X days for one specified host, storage system, tape, or switch?
- Which devices are producing traffic on a specific switch (for example, which devices are responsible for use of a highly utilized switch)?
- What is the throughput for a specific business unit in our environment? When you use the disk performance tables, you can obtain the following information:
- What is the throughput for a specified storage pool based on disk performance data?
- What is the highest used storage pool?
- What is the average disk utilization for a specific storage?
- What is the trend of usage for a storage system or storage pool based on disk performance data?
- What is the disk usage trending for a specific storage pool? When you use VM and VMDK performance tables, you can obtain the following information:
- Is my virtual environment performing optimally?
- Which VMDKs are reporting the highest workloads?
- How can I use the performance reported from VMDs mapped to different datastores to make decisions about re-tiering. The Performance data model includes information that helps you determine the appropriateness of tiers, storage misconfigurations for applications, and last access times of volumes and internal volumes. This data model provides data such as response times, IOPs, throughput, number of writes pending, and accessed status.

• **Storage Efficiency data model**

Enables you to track the storage efficiency score and potential over time. This data model stores measurements of not only the provisioned capacity, but also the amount that is used or consumed (the physical measurement). For example, when thin provisioning is enabled, OnCommand Insight indicates how much capacity is taken from the device. You can also use this model to determine efficiency when deduplication is enabled. You can answer various questions using the Storage Efficiency data mart:

- What is our storage efficiency savings as a result of implementing thin provisioning and deduplication technologies?
- What are the storage savings across data centers?
- Based on historical capacity trends, when do we need to purchase additional storage?
- What would be the capacity gain if we enabled technologies such as thin provisioning and deduplication?
- Regarding storage capacity, am I at risk now?

Data model fact and dimension tables

Each data model includes both fact and dimension tables.

- Fact tables: Contain data that is measured, for example, quantity, raw and usable capacity. Contain foreign keys to dimension tables.
- Dimension tables: Contain descriptive information about facts, for example, data center and business units. A dimension is a structure, often composed of hierarchies, that categorizes data. Dimensional attributes help describe the dimensional values.

Using different or multiple dimension attributes (seen as columns in the reports), you construct reports that access data for each dimension described in the data model.

For explanations of all data elements used in creating reports, see the Data Glossary.

Colors used in data model elements

Colors on data model elements have different indications.

- Yellow assets: Represent measurements.
- Non-yellow assets: Represent attributes. These values do not aggregate.

Using multiple data models in one report

Typically, you use one data model per report. However, you can write a report that combines data from multiple data models.

To write a report that combines data from multiple data models, choose one of the data models to use as the base, then write SQL queries to access the data from the additional data marts. You can use the SQL Join feature to combine the data from the different queries into a single query that you can use to write the report.

For example, say you want the current capacity for each storage array and you want to capture custom annotations on the arrays. You could create the report using the Storage Capacity data model. You could use the elements from the Current Capacity and dimension tables and add a separate SQL query to access the annotations information in the Inventory data model. Finally, you could combine the data by linking the Inventory storage data to the Storage Dimension table using the storage name and the join criteria.

Copyright information

Copyright © 2024 NetApp, Inc. All Rights Reserved. Printed in the U.S. No part of this document covered by copyright may be reproduced in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or storage in an electronic retrieval system—without prior written permission of the copyright owner.

Software derived from copyrighted NetApp material is subject to the following license and disclaimer:

THIS SOFTWARE IS PROVIDED BY NETAPP “AS IS” AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, WHICH ARE HEREBY DISCLAIMED. IN NO EVENT SHALL NETAPP BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

NetApp reserves the right to change any products described herein at any time, and without notice. NetApp assumes no responsibility or liability arising from the use of products described herein, except as expressly agreed to in writing by NetApp. The use or purchase of this product does not convey a license under any patent rights, trademark rights, or any other intellectual property rights of NetApp.

The product described in this manual may be protected by one or more U.S. patents, foreign patents, or pending applications.

LIMITED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (b)(3) of the Rights in Technical Data -Noncommercial Items at DFARS 252.227-7013 (FEB 2014) and FAR 52.227-19 (DEC 2007).

Data contained herein pertains to a commercial product and/or commercial service (as defined in FAR 2.101) and is proprietary to NetApp, Inc. All NetApp technical data and computer software provided under this Agreement is commercial in nature and developed solely at private expense. The U.S. Government has a non-exclusive, non-transferrable, nonsublicensable, worldwide, limited irrevocable license to use the Data only in connection with and in support of the U.S. Government contract under which the Data was delivered. Except as provided herein, the Data may not be used, disclosed, reproduced, modified, performed, or displayed without the prior written approval of NetApp, Inc. United States Government license rights for the Department of Defense are limited to those rights identified in DFARS clause 252.227-7015(b) (FEB 2014).

Trademark information

NETAPP, the NETAPP logo, and the marks listed at <http://www.netapp.com/TM> are trademarks of NetApp, Inc. Other company and product names may be trademarks of their respective owners.