How historical data is retained in Data Warehouse

OnCommand Insight

NetApp
March 08, 2024

Table of Contents

How historical data is retained in Data Warehouse

Data is maintained in Data Warehouse according to a schedule. As data gets older, the data record retention is reduced.

Data Warehouse retains historical data based on the data marts and granularity of the data, as shown in the following table.

Data mart	Measured object	Granularity	Retention period
Performance marts	Volumes and internal volumes	Hourly	14 days
Performance marts	Volumes and internal volumes	Daily	13 months
Performance marts	Application	Hourly	13 months
Performance marts	Host	Hourly	13 months
Performance marts	Switch performance for port	Hourly	5 weeks
Performance marts	Switch performance for host, storage, and tape	Hourly	13 months
Performance marts	Storage node	Hourly	14 days
Performance marts	Storage node	Daily	13 months
Performance marts	VM performance	Hourly	14 days
Performance marts	VM performance	Daily	13 months
Performance marts	Hypervisor performance	Hourly	14 days
Performance marts	Hypervisor performance	Daily	13 months
Performance marts	VMDK performance	Hourly	14 days
Performance marts	VMDK performance	Daily	13 months
Performance marts	Disk performance	Hourly	14 days
Performance marts	Disk performance	Daily	13 months

Capacity marts	All (except individual volumes)	Daily	13 months
Capacity marts	All (except individual volumes)	Monthly representative	14 months and beyond
Inventory marts	Individual volumes	Current state	1 day (or until next ETL)

After 13 months (which is configurable), Data Warehouse retains only one record per month instead of one record per day for capacity, performance, and resource data in the following fact tables:

- Chargeback fact table (dwh_capacity.chargeback_fact)
- File System Utilization fact table (dwh_fs_util.fs_util_fact)
- Host fact table (dwh_sa.sa_host_fact)
- Internal Volume Capacity fact table (dwh_capacity.internal_volume_capacity_fact)
- Ports fact table (dwh_ports.ports_fact)
- Qtree Capacity fact table (dwh_capacity.qtree_capacity_fact)
- Storage and Storage Pool Capacity fact table (dwh_capacity.storage_and_storage_pool_capacity_fact)
- Volume Capacity fact table (dwh_capacity.vm_capacity_fact)
- Storage Node Hourly Performance (storage_node_hourly_performance_fact) and Storage Node Daily Performance (storage_node_daily_performance_fact) fact tables

Data retention, ETL, and time periods

OnCommand Insight Data Warehouse retains data obtained from the Extract, Transform, and Load (ETL) process for different time periods based on the different data marts and time granularity of the data.

Performance Marts and hourly granularity for volumes and internal volumes

The OnCommand Insight Data Warehouse records the hourly averages, hourly maximums, and access bit for each hour of the day (24 data points) for 14 days. The access bit is a Boolean value that is true if the volume is accessed or false if the volume is not accessed during the hourly interval. All 24 data points for the preceding day are obtained during the first ETL process of the day.

You do not need to run one ETL process per hour to gather the hourly data points. Running additional ETL processes during the day does not obtain any performance information from the OnCommand Insight Servers.

Performance Marts and daily granularity for volumes and internal volumes

Each day when the ETL is processed, the daily averages for the preceding day are calculated and populated within Data Warehouse. The daily average is a summary of the 24 data points for the previous day. The performance data marts retain daily summaries for volumes and internal volumes for 13 months.

Capacity marts and daily granularity

The Capacity marts provide daily measurements for various capacity facts on a daily basis for a period of 13 months. The capacity facts in Data Warehouse are current as of the last data source acquisition for the device prior to the ETL.

Capacity marts and monthly granularity

Data Warehouse retains daily capacity data for 13 months. After the 13-month threshold is reached, the capacity data is summarized on a monthly basis. The monthly data is based on the values reflected by the date that is the month representative date.

The following table shows which monthly data is included in the monthly summary:

Date	Is Month Representative value	Allocated capacity
Jan 1	1 (True)	50 TB
Jan 2	0 (False)	52 TB
…	…	…
Jan 31	0 (False)	65 TB
Feb 1	1 (True)	65 TB

Based on the table, a monthly report would show 50 TB allocated for January and 65 TB allocated for February. All of the other capacity values for January would not be included in the monthly summary.

Inventory mart

The Inventory data mart is not historical. Each time an ETL process is run, the Inventory mart is erased and rebuilt. Therefore, any reports generated out of the Inventory mart do not reflect historical inventory configuration.

Copyright information

Trademark information