



欢迎使用**OnCommand Insight** 数据仓库

OnCommand Insight

NetApp
October 24, 2024

目录

- 欢迎使用OnCommand Insight 数据仓库 1
 - 数据仓库功能 1
 - 数据仓库组件 1
 - 数据仓库流程 2
 - 数据仓库如何提取数据 3
 - ETL频率和日期数据 3
 - 历史数据在数据仓库中的保留方式 4

欢迎使用OnCommand Insight 数据仓库

OnCommand Insight 数据仓库是一个集中式存储库、用于存储来自多个OnCommand Insight 服务器的数据、并将数据转换为一个通用的多维数据模型以供查询和分析。

通过OnCommand Insight 数据仓库、您可以访问由多个数据集市组成的开放式数据库、这些数据集市可用于生成自定义容量和性能报告、例如成本分摊报告、包含历史数据的趋势报告、使用情况分析和预测报告。

数据仓库功能

OnCommand Insight 数据仓库是一个由多个数据集市组成的独立数据库。

数据仓库包括以下功能：

- 当前和历史配置以及清单数据、可用于创建对预测和规划有用的趋势报告
- 多个多维历史数据集市和一个额外的纯当前清单数据集市
- 用于预定义查询或用户定义查询的优化数据库
- 与第三方报告和业务智能引擎集成的平台、包括：
 - 配置管理数据库
 - 财务会计系统
 - 资产管理系统

数据仓库组件

数据仓库包含多个组件。

- 数据仓库门户
- OnCommand Insight 报告门户
- 报告创作工具

使用数据仓库门户可以执行的操作

数据仓库门户是一个基于Web的用户界面、可用于配置选项并设置固定计划以检索数据。您还可以从数据仓库门户访问OnCommand Insight 报告门户。

使用数据仓库门户、您可以执行以下操作：

- 访问OnCommand Insight 报告门户可查看预先设计的报告或使用报告创作工具创建自定义报告。
- 整合多个OnCommand Insight 数据库。
- 管理与OnCommand Insight 服务器的连接。
- 检查当前正在运行的作业或查询的状态。
- 计划数据仓库构建。

- 编辑站点名称。
- 查看数据仓库版本和升级历史记录、包括模块版本、站点和许可证等特定信息。
- 导入标注。
- 配置从历史记录构建。
- 查看数据仓库文档和数据库架构。
- 重置数据仓库数据库。
- 备份和还原数据仓库数据库。
- 对数据仓库问题进行故障排除。
- 管理用户帐户。

数据仓库软件组件

OnCommand Insight 数据仓库包含多个软件组件。

- MySQL 数据库
数据集市表的后端存储库
- IBM Cognos
OnCommand Insight 的报告引擎
- Apache Derby数据库
用于存储Cognos配置和内容
- WildFly
托管OnCommand Insight 组件的Java Enterprise应用程序服务器

数据仓库流程

数据仓库执行多种类型的流程。

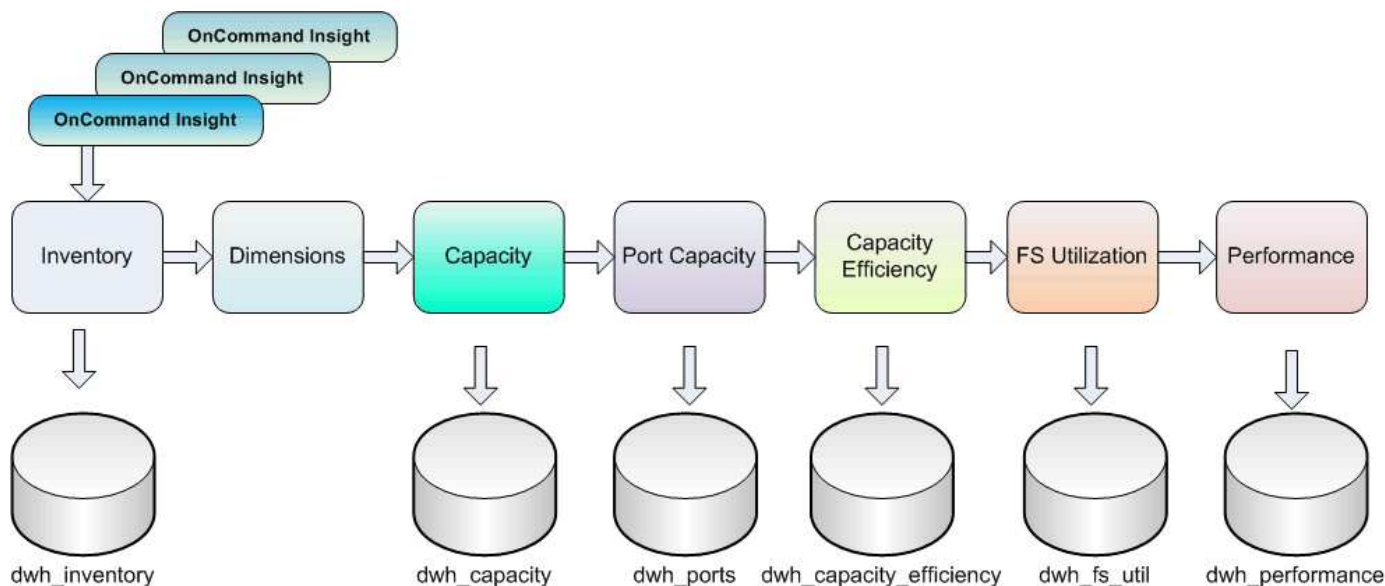
- * ETL进程*
提取转换和加载(ETL)过程从多个OnCommand Insight 数据库中检索数据、转换数据并将其保存到数据集中。数据仓库构建过程是一个ETL过程。
- * 作业 *
数据仓库执行和报告以下作业：清单、维度、容量、端口容量、VM容量、文件系统利用率、性能、容量效率、许可证、历史记录构建、动态标注、连接器删除、跳过的构建、ASUP选项和维护作业。
- 整合流程
数据仓库支持将多个OnCommand Insight 服务器整合到同一个数据仓库数据库中。在许多配置中、可能会

发生多个连接器报告同一对象的情况(即、同一交换机存在于两个OnCommand Insight 实例中)。在这种情况下、数据仓库会将多个对象整合到一个对象中(选择一个主连接器、并且仅从该连接器中获取对象的数据)。

数据仓库如何提取数据

提取、转换和加载(ETL)过程从多个OnCommand Insight 数据库中检索数据、转换数据并将其保存到数据集中。

OnCommand Insight 连接器会调用一系列批处理作业、以便从多个OnCommand Insight MySQL数据库中提取数据并将数据发布到各种数据集中、如下图所示。



ETL 流程包括以下各个流程：

- 提取

此过程会从多个OnCommand Insight 数据库中获取数据、转换数据并将其保存到数据集中。系统会同时对每个OnCommand Insight 实例执行此过程。为了确保执行数据清理和重复数据删除、不能将ETL进程拆分为多个计划的ETL操作。

- 转换

此过程将应用业务逻辑规则或函数从OnCommand Insight 数据库中提取数据。

- 加载

此过程会将转换后的数据加载到公共数据集中。

ETL频率和日期数据

您应至少每天运行一次提取、转换和加载(ETL)进程；但是、如果需要、您可以选择多次运行ETL。

默认情况下、Cognos报告引擎会将所有容量和性能数据视为附加数据。因此、如果在没有适当时间筛选器的情况下每天运行ETL流程多次、则存在容量数据重复计数的风险。

"Date"维度中的两个日期数据元素与每日ETL流程相关。在多个数据模型中使用的"日期"维度包括以下受ETL影响的数据元素：

- 是日代表

在任意给定日期运行的第一个ETL进程期间、"Is Day Representative"数据元素的值将设置为1 (true)。如果第一个ETL流程在凌晨1：00运行、则对于在凌晨1：00期间加载的所有数据、IS Day代表性将设置为1ETL流程。如果稍后计划第二个ETL (例如、下午1：00)、则对于在该ETL过程中加载的数据、IS Day代表性将设置为0 (false)。

- 为最新

每个ETL过程完成后、"Is Latest"成员的值将设置为1 (true)。如果第一个ETL进程在凌晨1：00运行、则对于在凌晨1：00期间加载的所有数据、"Is Latest"将设置为1ETL流程。如果稍后计划另一个ETL进程(例如下午1：00)、则对于下午1：00期间加载的数据、Is Latest将设置为1ETL流程。ETL过程还会设置凌晨1：00ETL负载的最新条目为0 (false)。

历史数据在数据仓库中的保留方式

数据会按计划和数据仓库中进行维护。随着数据变旧、数据记录保留率也会降低。

数据仓库会根据数据集市和数据粒度保留历史数据、如下表所示。

数据集市	测量对象	粒度	保留期限
性能集市	卷和内部卷	每小时	14 天
性能集市	卷和内部卷	每天	13 个月
性能集市	应用程序	每小时	13 个月
性能集市	主机	每小时	13 个月
性能集市	端口的交换机性能	每小时	5周
性能集市	主机，存储和磁带的交换机性能	每小时	13 个月
性能集市	存储节点	每小时	14 天
性能集市	存储节点	每天	13 个月
性能集市	虚拟机性能	每小时	14 天

性能集市	虚拟机性能	每天	13 个月
性能集市	虚拟机管理程序性能	每小时	14 天
性能集市	虚拟机管理程序性能	每天	13 个月
性能集市	VMDK 性能	每小时	14 天
性能集市	VMDK 性能	每天	13 个月
性能集市	磁盘性能	每小时	14 天
性能集市	磁盘性能	每天	13 个月
容量集市	全部（单个卷除外）	每天	13 个月
容量集市	全部（单个卷除外）	每月代表	14 个月及以上
清单数据集市	单个卷	当前状态	1 天（或直到下一 ETL）

13个月后可配置)、数据仓库每月仅保留一条记录、而不是每天在以下数据表中保留一条容量、性能和资源数据记录：

- 成本分摊数据表(dwh_capacity.chargeback_fact)
- 文件系统利用率数据表(dwh_fs_util.fs_util_fact)
- 主机数据表(dwh_sa.sa_host_fact)
- 内部卷容量数据表(dwh_capacity.internal_volume_capacity_fact)
- 端口数据表(dwh_ports.ports_fact)
- qtree容量数据表(dwh_capacity.qtree_capacity_fact)
- 存储和存储池容量数据表(dwh_capacity.storage_and_storage_pool_capacity_fact)
- 卷容量数据表(dwh_capacity.vm_capacity_fact)
- 存储节点每小时性能(storage_node_hourly_performance_fact)和存储节点每日性能(storage_node_daily_performance_fact)数据表

数据保留、ETL和时间段

OnCommand Insight 数据仓库会根据数据的不同数据集市和时间粒度、将从提取、转换和加载(ETL)流程获取的数据保留不同时间段。

卷和内部卷的性能集市和每小时粒度

OnCommand Insight 数据仓库会记录14天内每天每小时的每小时平均值、每小时最大值和访问位(24个数据点

)。访问位是一个布尔值、如果访问了卷、则为true；如果在每小时间隔内未访问卷、则为false。前一天的所有24个数据点都是在当天的第一个ETL过程中获取的。

您无需每小时运行一个ETL流程来收集每小时数据点。在一天内运行其他ETL进程不会从OnCommand Insight服务器获取任何性能信息。

卷和内部卷的性能集市和每日粒度

每天处理ETL时、都会在数据仓库中计算和填充前一天的每日平均值。每日平均值是前一天24个数据点的汇总。性能数据集市会将卷和内部卷的每日摘要保留13个月。

容量集市和每日粒度

容量集市每天对13个月内的各种容量数据进行测量。数据仓库中的容量数据是在ETL之前最后一次为设备采集数据源时的最新数据。

容量集市和每月粒度

数据仓库会将每日容量数据保留13个月。达到13个月的阈值后、容量数据将按月汇总。每月数据基于月份代表性日期所反映的值。

下表显示了每月摘要中包含的每月数据：

Date	为月份代表值	已分配容量
1月1日	1 (正确)	50 TB
1月2日	0 (错误)	52 TB
...
1月31日	0 (错误)	65 TB
2月1日	1 (正确)	65 TB

根据该表、每月报告将显示1月份分配的容量为50 TB、2月份分配的容量为65 TB。1月份的所有其他容量值都不会包含在每月摘要中。

清单数据集市

清单数据集市不是历史数据集市。每次运行ETL进程时、都会擦除并重建清单数据集市。因此、从清单数据集市生成的任何报告都不会反映历史清单配置。

版权信息

版权所有 © 2024 NetApp, Inc.。保留所有权利。中国印刷。未经版权所有者事先书面许可，本文档中受版权保护的任何部分不得以任何形式或通过任何手段（图片、电子或机械方式，包括影印、录音、录像或存储在电子检索系统中）进行复制。

从受版权保护的 NetApp 资料派生的软件受以下许可和免责声明的约束：

本软件由 NetApp 按“原样”提供，不含任何明示或暗示担保，包括但不限于适销性以及针对特定用途的适用性的隐含担保，特此声明不承担任何责任。在任何情况下，对于因使用本软件而以任何方式造成的任何直接性、间接性、偶然性、特殊性、惩罚性或后果性损失（包括但不限于购买替代商品或服务；使用、数据或利润方面的损失；或者业务中断），无论原因如何以及基于何种责任理论，无论出于合同、严格责任或侵权行为（包括疏忽或其他行为），NetApp 均不承担责任，即使已被告知存在上述损失的可能性。

NetApp 保留在不另行通知的情况下随时对本文档所述的任何产品进行更改的权利。除非 NetApp 以书面形式明确同意，否则 NetApp 不承担因使用本文档所述产品而产生的任何责任或义务。使用或购买本产品不表示获得 NetApp 的任何专利权、商标权或任何其他知识产权许可。

本手册中描述的产品可能受一项或多项美国专利、外国专利或正在申请的专利的保护。

有限权利说明：政府使用、复制或公开本文档受 DFARS 252.227-7013（2014 年 2 月）和 FAR 52.227-19（2007 年 12 月）中“技术数据权利 — 非商用”条款第 (b)(3) 条规定的限制条件的约束。

本文档中所含数据与商业产品和/或商业服务（定义见 FAR 2.101）相关，属于 NetApp, Inc. 的专有信息。根据本协议提供的所有 NetApp 技术数据和计算机软件具有商业性质，并完全由私人出资开发。美国政府对这些数据的使用权具有非排他性、全球性、受限且不可撤销的许可，该许可既不可转让，也不可再许可，但仅限在与交付数据所依据的美国政府合同有关且受合同支持的情况下使用。除本文档规定的情形外，未经 NetApp, Inc. 事先书面批准，不得使用、披露、复制、修改、操作或显示这些数据。美国政府对国防部的授权仅限于 DFARS 的第 252.227-7015(b)（2014 年 2 月）条款中明确的权利。

商标信息

NetApp、NetApp 标识和 <http://www.netapp.com/TM> 上所列的商标是 NetApp, Inc. 的商标。其他公司和产品名称可能是其各自所有者的商标。