



Was wir vorhersagen, soll auch eintreffen!



Natives Data Warehousing in SAP HANA mit der Data Warehousing Foundation – Teil 1: NDSO //

Philipp Schlechter



Einleitung

Die SAP HANA Data Warehousing Foundation (DWF) ist eine Zusammenstellung von Tools, die es Entwicklern und Adminstratoren von SAP HANA ermöglicht, native Data Warehouses zu modellieren und zu managen. Sie umfasst folgende Tools: Native DataStore Object (NDSO), Data Distribution Optimizer (DDO), Data Lifecycle Manager (DLM), Data Warehouse Scheduler (DWS) und Data Warehouse Monitor (DWM). Analog zum Advanced DataStore Object (ADSO) in SAP BW/4HANA ist das zentrale Objekt der DWF das NDSO. Dies ist der erste Teil unserer White Paper-Serie „Natives Data Warehousing in SAP HANA mit der Data Warehousing Foundation“, in welchem das NDSO mit seinen Funktionalitäten genauer betrachtet wird.

Um den Zweck der DWF zu erkennen, muss man sich zunächst mit dem HANA SQL Data Warehouse beschäftigen. SAP bedient bereits seit längerer Zeit zwei verschiedene Ansätze des Data Warehousing. Der erste Ansatz ist die weit verbreitete Anwendung SAP Business Warehouse (BW), welche sich ebenfalls mit SAP BW/4HANA zu einer HANA gestützten Applikation entwickelt hat. Beim zweiten Ansatz handelt es sich um das SQL Data Warehouse, das in der Vergangenheit auf jeder beliebigen Datenbank betrieben werden konnte und von SAP Produkten, wie den Data Services unterstützt wurde. Mit dem Release von SAP HANA 2 wurde das Toolset des SQL Data Warehouse aktualisiert und unter dem Namen HANA SQL Data Warehouse zur Verfügung gestellt. Dabei wurden die Tools des SQL DW übernommen und um einige Neuerungen erweitert. Eine dieser Neuerungen ist das Release neuer SAP HANA DWF Funktionen. Die zentrale Funktion, um die es sich hierbei handelt ist das NDSO.

Was ist das NDSO?

Ziel des NDSO ist die Bereitstellung der Funktionalitäten des ADSO in SAP BW4/HANA für die Entwicklung von native Data Warehouses. Im HANA SQL Data Warehouse stellt das NDSO das zentrale Persistenzobjekt dar. Betrachtet man das NDSO aus technischer Sicht, so stellt man fest, dass es sich dabei um eine standardisierte Objektdefinition handelt, welche die Technologie der Core Data Services (CDS) nutzt. Dementsprechend wird die Richtlinie umgesetzt, CDS als zentrales Datenmodellierungskonzept in HANA zu verwenden.

Der nähere Vergleich mit dem ADSO zeigt ein grundsätzlich gleiches Konzept in Bezug auf die Tabellen, welche der Datenverarbeitung innerhalb des NDSOs bzw. ADSOs dienen. Im NDSO, Inbound Table im ADSO, existiert jeweils eine Tabelle, in die Daten initial geladen werden (Inbound Queue (IQ)). Von dieser Tabelle aus können die Daten aktiviert und anschließend in die Active Data Table transferiert werden. Auf die Dokumentation der Delta Loads im Change Log wird im folgenden Abschnitt näher eingegangen. Durch die Architektur des NDSO und ADSO in den drei genannten Tabellen, wird die technische Grundlage zur Datenkonsolidierung und -speicherung gelegt. Die Anwendung beider DSOs unterscheidet sich lediglich dadurch, dass das ADSO die Möglichkeit zur Integration weiterer BW Objekte, wie zum Beispiel Info Objects, bietet, welche im nativen HANA SQL DW nicht existieren.

Weil das NDSO als DWF Tool nur mit dem Release von HANA 2 gelauncht wurde, ist es ausschließlich über XS Advanced verfügbar. Nach Installation der DWF Tools über das HANA XS Advanced Lifecycle Management (HALM) können die Tools in die Web IDE integriert werden. Es gibt diverse Möglichkeiten, ein NDSO in der Web IDE anzulegen. Für erste Übungszwecke ist es empfehlenswert, zunächst ein DWF Projekt anzulegen, in dem alle Konfigurationen für die Verwendung der DWF Tools bereits vorhanden sind. Wie Abbildung 1 zeigt, sollte bei der Erstellung eines neuen Projektes die Option „Data Warehouse Project based on SAP Data Warehousing Foundation“ ausgewählt werden. Das angelegte Projekt beinhaltet somit alle notwendigen Konfigurationen und Objekte zur Modellierung von NDSOs.

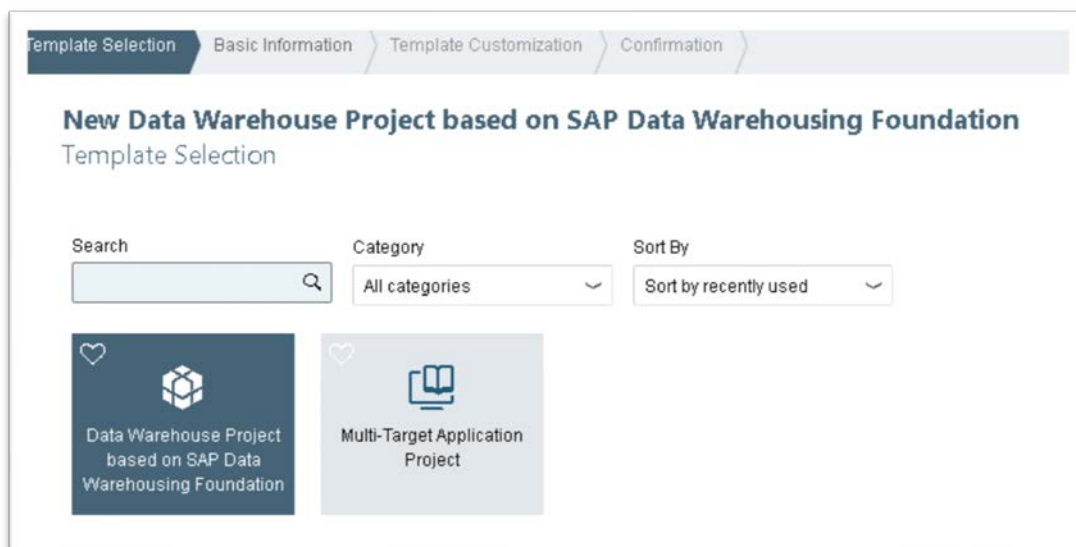


Abbildung 1: Erstellung eines DWF Projektes

Alternativ kann ein NDSO in einem bestehenden Projekt angelegt werden. Dabei wählt man per Rechtsklick auf ein HDB Modul, welches bereits im Projekt angelegt wurde, die Option „New -> HDB CDS Artifact“. Dies ist notwendig, da das NDSO, wie bereits erwähnt, technisch gesehen auf CDS basiert. Über den grafischen Editor kann das erstellte Objekt, das die Dateiendung „.hdbcds“ trägt, bearbeitet werden.

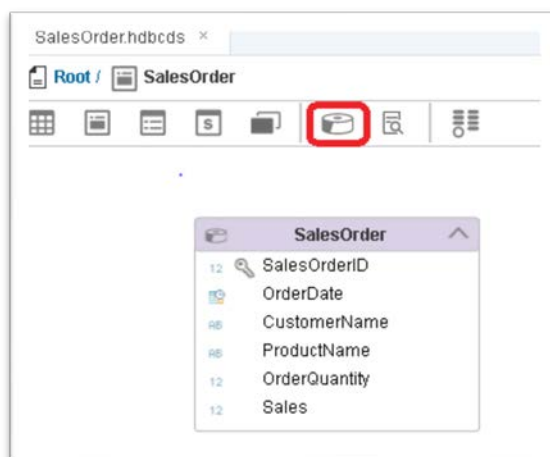


Abbildung 2: DSO über den grafischen Editor anlegen

Das Rot markierte Symbol in der Toolbar des grafischen Editors in Abbildung 2 dient dazu, ein NDSO anzulegen. Die Abbildung zeigt ein bereits angelegtes NDSO „SalesOrder“. Durch Doppelklick auf das NDSO kann dieses bearbeitet werden. In der Bearbeitungsansicht stehen mehrere Tabs zur Verfügung, unter denen das NDSO angepasst werden kann. Es können Elemente, welche die Datenstruktur des NDSO bestimmen, hinzugeführt und entfernt sowie erweiterte Konfigurationen definiert werden.



Data Load Prozess

Um die Funktion des NDSO als zentrales Persistenzobjekt in einem HANA SQL DW erfüllen zu können, sind sowohl der Data Load Prozess als auch die damit verbundenen Prozesse elementar. In diesem Abschnitt werden diese Aspekte kurz erläutert. Die Grundlage für die Prozesse besteht in der Konfiguration des NDSO Objektes. Diesbezügliche Einstellungen können im Tab „DSO Details“ in der Bearbeitungsansicht vorgenommen werden. Über die Checkbox mit der Beschriftung „Write Change Log“ wird festgelegt, ob die Change Log Tabelle bei Änderungen, die über die Inbound Queue bearbeitet werden, die entsprechenden Änderungen dokumentiert. Standardmäßig ist dies der Fall und sollte in den meisten Fällen auch beibehalten werden, da über den Change Log die Möglichkeit eines Roll-Back besteht. Über die tabellarische Darstellung auf der unteren rechten Seite kann für jedes einzelne Element des NDSO ein sogenannter Update Mode definiert werden. Ist dieser auf „Ignore“ gesetzt, so wird das Element beim Update von Daten in die Inbound Queue nicht befüllt. So ist es möglich, nicht-dynamische Werte von den Updates auszunehmen, um zu verhindern, dass sie überschrieben werden.

The screenshot shows the configuration interface for a SalesOrder NDSO. The 'DSO Details' tab is active, showing various settings and tables.

General

- Ensure Consistency (Digest)
- Support Snapshot
- Write Change Log `change_log`

Aggregation Behavior

Elements	Data Type	Aggregation
SalesOrderID	Integer	None
OrderDate	UTCDateTime	None
CustomerName	String	None
ProductName	String	None

Inbound Queue

Name	Type
inbound_queue	Entity

Details of inbound_queue

Elements	Data Type	Update Mode
SalesOrderID	Integer	Move
OrderDate	UTCDateTime	Move
CustomerName	String	Move
ProductName	String	Move

Abbildung 3: Bearbeitungsansicht des NDSO

Bevor das NDSO mit Daten beladen werden kann, müssen entsprechende Berechtigungen an den Anwender vergeben werden. Eine detaillierte Anleitung dazu gibt [1]. Anschließend können Daten mittels Flowgraph in das NDSO geladen werden. Zur Modellierung des Data Load Process bietet die Web IDE den Flowgraph Modeler, mit dem über eine grafische Benutzeroberfläche verschiedene Objekte eingefügt werden können. Für die Beladung eines NDSO werden nur eine Data Source und ein Data Target benötigt. Als Target wird das zuvor angelegte NDSO und als Source eine dazu passende aktivierte Tabelle definiert, welche in der Web IDE als CDS Datei vorliegt.

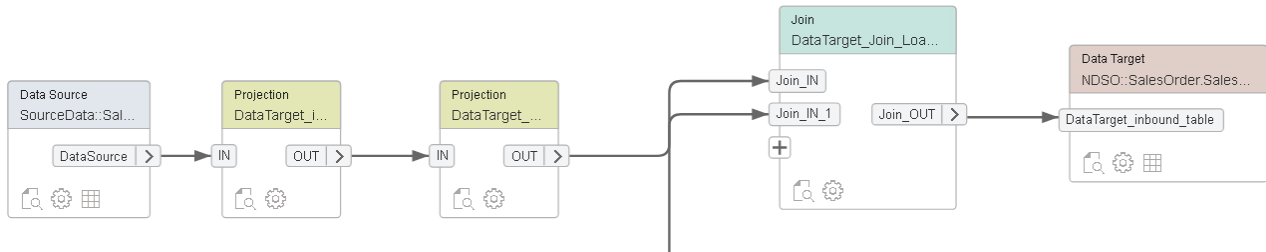


Abbildung 4: Flowgraph Modeler

Durch die Zuweisung des NDSO als Data Target wird automatisiert eine Reihe an Projections, Joins und Unions erstellt, die dazu dienen, das Record ID Handling und das Request ID Handling im Data Load Prozess zu strukturieren. Über den Flowgraph Modeler kann, wie aus SAP BW bekannt, ein Mapping von Spalten aus der Datenquelle auf Elemente des NDSO definiert werden. Durch die Ausführung des Flowgraph wird der Load in die Inbound Queue durchgeführt. Nachdem die Inbound Queue beladen wurde, können die neuen Daten zunächst validiert und dann aktiviert werden. In der Web IDE ist dies über den Database Explorer möglich. Per Rechtsklick auf das NDSO kann „Manage“ ausgewählt werden, wobei Load und Activation Requests angezeigt werden. Durch die Auswahl eines Load Requests kann dieser aktiviert werden. Hierbei werden die Daten von der Inbound Queue in die Active Data Tables geschrieben und im Change Log hinterlegt, wenn diese Option zuvor ausgewählt wurde. Falls es sich beim Data Load Prozess um die initialen Prozesse handelt, ist an dieser Stelle nichts Weiteres zu berücksichtigen, da es sich um einen Full Data Load handelt.

Bei einem Data Load in ein NDSO mit bereits aktivierte Daten, muss der jeweilige Record Mode berücksichtigt werden. Der Record Mode ist ein Konzept, das festlegt, wie jeder spezifische Datenpunkt beim Data Load verarbeitet wird. Es existieren folgende Record Modes für Delta Loads aus externen Datenquellen:

- **,N'** – gibt an, dass es sich um einen neuen Record handelt, der einfach in die Active Data Table eingetragen werden kann.
- **,A'** – gibt an, dass es sich um einen aggregierbaren Record handelt, sodass Daten bei der Übertragung in die Active Data Table additiv mit Daten der gleichen ID aggregiert werden.
- **,D'** – der Record mit der entsprechenden ID in der Active Data Table wird gelöscht. Dabei spielen alle Felder außer der ID keine Rolle.
- **,R'** – die Werte des Records werden durch die Aktivierung nicht verändert, aber der Record wird beim nächsten Update des NDSO gelöscht.
- **,X'** – gibt die Umkehr des Record Modes ‚A‘ an. Dadurch werden additive Data Loads um einen Schritt rückgängig gemacht.

Fazit

Für das native SQL Data Warehouse in SAP HANA stellen die DWF Tools eine sehr wichtige Erweiterung dar. Das NDSO als zentrales Persistenzobjekt liefert die Funktionalität des ADSO für die Verwendung in der Web IDE. Mit dem Data Flow Modeler wurde zusätzlich ein bekanntes Objekt aus der BW Welt in die neue Umgebung gebracht. Dies trägt dazu bei, dass die Modellierung von Data Load Prozessen einfach und intuitiv durchgeführt werden kann. Der Weg ins Web, den SAP bereits mit der Integration der DWF Tools in die Web IDE beschritten hat, soll mit den nächsten Releases noch weiter gegangen werden, indem die DWF über die HANA Cloud Platform als Service zur Verfügung gestellt wird. Der Grundstein für eine moderne, webbasierte Ergänzung zu SAP BW4/HANA ist dadurch gelegt. Wir sind gespannt, wie sich die DWF strategisch entwickelt.



Quellen und weiterführende Informationen

- [1] [SAP HANA Data Warehousing Foundation - Native DataStore Object Administration Guide \(XSA\)](#).
- [2] Heiko Schneider: Introducing the NDSO: Part Two – NDSO services (Delta Capabilities), <https://blogs.sap.com/2017/12/23/introducing-the-ndso-part-two-ndso-services-delta-capabilities/>
- [3] SAP HANA Data Warehousing Foundation 2.0, Road Map, <https://help.sap.com/viewer/disclaimer-for-links?q=https%3A%2F%2Fwww.sap.com%2Fdocuments%2F2017%2F06%2F4459c228-c37c-0010-82c7-eda71af511fa.html>



Kontaktieren Sie uns //

mayato GmbH
Friedrichstraße 121
10117 Berlin

info@mayato.com

+49 / 30 7001.4692-0