Data Lifecycle Management and Data Aging Concepts for SAP HANA
Disclaimer

The information in this presentation is confidential and proprietary to SAP and may not be disclosed without the permission of SAP. Except for your obligation to protect confidential information, this presentation is not subject to your license agreement or any other service or subscription agreement with SAP. SAP has no obligation to pursue any course of business outlined in this presentation or any related document, or to develop or release any functionality mentioned therein.

This presentation, or any related document and SAP's strategy and possible future developments, products and or platforms directions and functionality are all subject to change and may be changed by SAP at any time for any reason without notice. The information in this presentation is not a commitment, promise or legal obligation to deliver any material, code or functionality. This presentation is provided without a warranty of any kind, either express or implied, including but not limited to, the implied warranties of merchantability, fitness for a particular purpose, or non-infringement. This presentation is for informational purposes and may not be incorporated into a contract. SAP assumes no responsibility for errors or omissions in this presentation, except if such damages were caused by SAP’s intentional or gross negligence.

All forward-looking statements are subject to various risks and uncertainties that could cause actual results to differ materially from expectations. Readers are cautioned not to place undue reliance on these forward-looking statements, which speak only as of their dates, and they should not be relied upon in making purchasing decisions.
Introduction

The **SAP HANA Data Warehousing Foundation** product provides specific data management tools, to support large scale SAP HANA use cases like:

- **HANA Table-, Partition- & Application Distribution Management in SAP HANA Scale-Out Configurations**
  - SAP BW/4HANA and SAP BW powered by SAP HANA
  - SAP HANA SQL Data Warehouse (native) & „Mixed scenarios“
  - Multiple applications running on a single SAP HANA Scale-Out database
- **Data Temperature Management** (Hot- to Warm- and Cold-Store including Bi-Directional Data Movement) based on Data Movement Rules and Data Access across the different data stores
- **Modeling** of HANA native Content Objects including Data Management Capabilities and TaskChains
- **Execute and Monitor** the Sequence of HANA Tasks

- **SAP HANA DWH Foundation** complements the data warehouse offerings of SAP BW/4HANA, SAP BW powered by SAP HANA and SAP HANA SQL Data Warehouse
SAP HANA Data Warehousing Foundation delivers specialized SAP HANA based applications

- **Data Distribution Optimizer** to plan, adjust and analyze landscape reorganizations for SAP HANA scale out systems
- **Data Lifecycle Manager** to deliver a possibility to archive / displace data from a SAP HANA persistency to Multi-Store Table, HANA Extension Node, Dynamic Tiering, SAP IQ, Hadoop or SAP Vora **
- **Native DataStore Object** to provide a central persistence object with the same capabilities of the BW standard advanced DataStore Object (ADSO models
- **Data Warehouse Scheduler** to maintain dependencies between single processes with the focus to provision data warehouse models
- **Data Warehouse Monitoring** to provide a comprehensive overview about current and past activities in the data warehouse

- DWF 1.0 SP05 >= HANA 1.0 SPS12
- DWF 2.0 SP01 >= HANA 2.0 SPS01

* available with DWF 2.0 SP01
** restricted shipment with DWF 2.0 SP02
Data Lifecycle Manager (DLM)
Define a data temperature (tiering) management strategy with DLM

Leverage SAP HANA tables (Hot-Store), HANA Extension Node (Warm-Store), SAP HANA Dynamic Tiering (Warm-Store), SAP Vora*, Hadoop or SAP Sybase IQ (Cold-Store) in SAP HANA native use cases with a tool based approach to model aging rules on tables to displace ‘aged’ data to optimize the memory footprint of data in SAP HANA.

* restricted shipment with DWF 2.0 SP02
SAP Data Warehousing Foundation
Data Lifecycle Manager (DLM) – Data Movement

Orchestrate and optimize the HANA memory footprint of data in SAP HANA tables

- Data Modification on primary Application table (e.g. Hot-Store) - only on records in specific “current” / open periods
- ‘Aged’ Data within “closed” periods to be archived / displaced to another Storage Destination
- Define Data Movement rule (in and out) to displace data between HANA-, Extended-, Hadoop-, SAP Vora* or SAP Sybase IQ-tables
- Data Movement rule supports regular SQL WHERE Clause operators and a set of DATE-functions, like:
  - "DATEFROM" < ADD_DAYS (CURRENT_DATE, -1095) AND “STATUS” = ‘C’
  (displace all “closed/completed” records older than ~3 years (1095 days) to another Storage Location)
- Data Movement rules generated into HANA Stored Procedures to perform mass data movement
- Execution of HANA Stored Procedures using HANA tasks (Manual and Scheduled execution)
- Selective data deletion for proper housekeeping with DLM

* restricted shipment with DWF 2.0 SP02
Optimized reading access to DLM managed data

- Generated SAP HANA Views (CalcScenarios) for optimized data access between Column-, Extended- and virtual-Table (HANA Application Integration Objects) – NOT relevant for HANA Extension Node and Multi-Store DLM-Profiles
- “Dynamic Pruning” based on SQL query WHERE-clause analysis to optimize the access to only relevant data store (HANA Pruning View)
- Flexible adjustment of DLM Lifecycle Profile (Relocation Direction, Package Size, Data Movement rule setting … ) without interfering with DLM generated SAP HANA Views
- Integration of DLM generated SAP HANA Views into HANA virtual data models (VDM) and native SAP HANA Applications
Details on the generated SAP HANA Views

- **DLM_UNION**: “UNION ALL” DB-View
  1. Table Primary Key or DLM Nominal Key specifies the data granularity
  2. Data-Set from cold AND hot storage location are accessed, by applying the SQL query WHERE-clause filter criteria (not relevant for HANA Extension Node or Multi-Store-Table only Scenario)

**Scenario:**
Unique record available on the cold storage location (Initial record)
More recent / current record (Update record) in the hot store, both following the same Primary Key

- 2 Records will be returned when querying the DLM_UNION View and filtering on the given Primary Key / Nominal Key
- UNION ALL will append relevant records from both stores to the query result set

3. Displace Records from HANA Column-Store table only if
   - NO further Update Records are expected (enhance DLM-Rule by “Status” Column value)
   - Execute the DLM-Profile more frequently to move/update all records with identical Primary Key value(s) to a single Storage Location

4. Adjust consuming OLTP-Application to “select records for update” using the DLM_UNION (not relevant for HANA Extension Node or Multi-Store-Table only Scenario)
Details on the generated SAP HANA Views

**DLM_PVIEW**: “HANA Calculation View“ (generated from HANA Calculation Scenario)

1. Table Primary Key or DLM Nominal Key specifies the data granularity
2. Data-Set from cold AND / OR hot storage location are accessed, by applying the SQL query WHERE-clause filter criteria (not relevant for HANA Extension Node or Multi-Store-Table only Scenario).

**Scenario:**

Unique record available on the cold storage location (Initial record)
More recent / current record (Update record) in the hot store, both following the same Primary Key

- 1 Record will be returned when querying the DLM_PVIEW and filtering on the given Primary Key / Nominal Key
- The distributed data set will be interpreted as a single table, where only the most recent / current record will be added to query result set

3. “Dynamic Pruning” based on the SQL query WHERE-clause analysis to optimize the access to only relevant data store

- access only hot storage location data if SQL query WHERE-clause doesn't match already displaced data set
- access cold AND hot storage location data SQL query WHERE-clause matches already displaced data set

Select the “most-current” record from the set of two tables (might cause performance issues on large datasets)

4. Relax the frequency to execute the DLM-Profile to move/update all records with identical Primary Key value(s) to a single Storage Location
5. Adjust consuming OLTP-Application to “select records for update” using the DLM_PVIEW (not relevant for HANA Extension Node or Multi-Store-Table only Scenario)
SAP Data Warehousing Foundation
Data Lifecycle Manager (DLM) – HANA Extension Node

Orchestrate and optimize the HANA memory footprint of data in SAP HANA partitioned tables

- Partitioned Column-Store table with Partitions located in Hot-Store and HANA Extension Node
- Low to No effort to integrate with existing Applications
  - existing tables remain unchanged / stable
  - existing In-Memory / Column-Store table to altered to partitioned table
- DLM specified aging rules to move complete Table-partitions between Hot-Store and HANA Extension Node
- No DLM generated SAP HANA View (Pruning / UNION) required, due to single partitioned table
- Data access managed by HANA incl. Partition pruning
- No impact to data update / delta handling, as records are moved to unique table-partition, based on partitioning criteria

For HANA native scenarios, Extension Node is under controlled availability.
Please check SAP Note 2415279 - How-To: Configuring SAP HANA for the SAP HANA Extension Node
Orchestrate and optimize the HANA memory footprint of data in SAP HANA Multi-Store tables

- Multi-Store table with Partitions located in Hot-Store and HANA Extension Node
- Low to No effort to integrate with existing Applications
  - existing tables remain unchanged / stable
  - existing In-Memory / Column-Store table to altered to partitioned table
- DLM specified aging rules to move complete Table-partitions between Hot-Store and Dynamic Tiering Node
- No DLM generated SAP HANA View (Pruning / UNION) required, due to single partitioned table
- Data access managed by HANA incl. Partition pruning
- No impact to data update / delta handling, as records are moved to unique table-partition, based on partitioning criteria
Orchestrate and optimize the HANA memory footprint of data in SAP HANA using DLM Modeled Persistence Object (MPO)

- Pursue a consistent data movement for multiple data-connected tables (e.g. Header – Item)
- MPO based on either a
  - Table Hierarchy (HANA Calculation View)
  - Table Group (set of connected tables) with identical column-names defining the connection between the tables
- Once defined, a MPO can be used in a DLM lifecycle profile.
- Data Movement rules to be defined on the common Calculation View columns (Table Hierarchy) or table columns (Table Group) of the MPO.
- Data movement from set of source tables (Table A – C) to the set of target table (IQ: Table A – C)
Orchestrate and optimize the HANA memory footprint of data in SAP HANA using DLM Modeled Persistence Object (MPO)

- Pursue a consistent data movement for multiple data-connected tables (e.g. Header – Item)
- MPO based on either a
  - Table Hierarchy (HANA Calculation View)
  - Table Group (set of connected tables) with identical column-names defining the connection between the tables
- Once defined, a MPO can be used in a DLM lifecycle profile.
- Data Movement rules to be defined on the common Calculation View columns (Table Hierarchy) or table columns (Table Group) of the MPO.
- Data movement from set of source tables (Table A – C) to the set of target table (DT: Extended Table A – C)
Orchestrating and optimizing the HANA memory footprint of data in Multi-Tier scenarios (here: partitioned table + Cold Store)

- Multi-Tier Data Access via generated SAP HANA Views (CalcScenarios) for optimized data access between partitioned Column-Table and virtual-Table (Cold Store)

- DLM-Profile 1: DLM specified aging rules to move complete Table-partitions between Hot-Store and HANA Extension Node
  - Option to access ONLY partitioned Table (Hot- and Warm- data) via Table-Access
  - Data access managed by HANA incl. Partition pruning

- DLM-Profile 2: DLM specified aging rules to move dataset between partitioned Table and Cold Store Location
  - DLM generated SAP HANA View (Pruning / UNION) to access ALL data (Hot-, Warm- and Cold- data)
Orchestrate and optimize the HANA memory footprint of data in Multi-Tier scenarios (here: Multi-Store table + Cold Store)

- Multi-Tier Data Access via generated SAP HANA Views (CalcScenarios) for optimized data access between Multi-Store Table and virtual-Table (Cold Store)
- DLM-Profile 1: DLM specified aging rules to move complete Table-partitions between Hot-Store and Dynamic Tiering Node
  - Option to access ONLY partitioned Table (Hot- and Warm- data) via Table-Access
  - Data access managed by HANA incl. Partition pruning
- DLM-Profile 2: DLM specified aging rules to move dataset between partitioned Table and Cold Store Location
  - DLM generated SAP HANA View (Pruning / UNION) to access ALL data (Hot-, Warm- and Cold- data)
Modern SAP UI5 based administration

- Specify and manage Storage Destinations
- Create, adjust and manage Lifecycle Profiles
- Setup & Adjust Lifecycle Relocation Rule (via Rule Editor)
  - Define Data Movement rules to move data between HANA-, Extended-, Hadoop- or SAP Sybase IQ-tables - in and out
  - Define Modeled Persistence Objects (e.g. Header-Item Dep.)
  - Persistence criteria incl. Source Table w/o PK (Nominal Key)
  - Data Movement rules generated into HANA Stored Procedures to perform mass data movement
- Validate Data Relocation Rule Proposal vs. Current Distribution
- Schedule and Execute Data Relocation Run using HANA tasks
- Analyze logs of executed Data Relocation Run
SAP HANA Data Warehousing Foundation - Data Lifecycle Manager
Product road map overview - key themes and capabilities

Today
• Adapt Relocation Agent module for Table Groups & Table Hierarchies for supported persistent storage destinations
• Enhanced support for Table Hierarchies and Table Groups for certain storage destinations
• Feature completeness of Hadoop storage destinations (packet size, clash strategy)
• Support for multi-tiering

Planned Innovations
• XSA enablement
• Rule Editor enhancements (Semantic checks / Customer exit)
• SAP HANA Multi-Store table support
• Integration with SAP HANA Vora**

Future Direction
• Pre-delivered Application DLM templates
• Integrated lifecycle management
• Exploration module based on SQL Query Access Cache
• Parallelization of relocation packet processing
• Support for SAP HANA CDS Entities as source persistence object
• Automated substitution of data foundation in Calculation Views (table replaced by generated view)

SAP DW Foundation 2.0 SP01

*This is the current state of planning and may be changed by SAP at any time
** restricted shipment with DWF 2.0 SP02
Data Lifecycle Manager (DLM)

– Offer data warehouse developers functionality to model data movement strategies from HANA column store tables to Hadoop, Vora*, (SAP IQ, Dynamic Tiering or HANA Extension Node planned)

– Enable access to warm and cold data by generating pruning views (calculation views)

– Enable data movement by generating HANA database procedures

– Execution and Monitoring of DLM Profiles (Data Movement) embedded into HANA Data Warehousing Scheduler via generated “DLM task chains”

* restricted shipment with DWF 2.0 SP02
Data Warehousing Scheduler (DWS)
- provide a framework to define task chains as a sequences of single tasks
- Flexible start conditions
- Parallelization and Dependency Handling
- Provide capability to schedule flowgraphs, NDSO related tasks, project local database procedures (planned for DWF 2.0 SP02) and DLM related tasks (planned for DWF 2.0 SP02)
Thank you.

Contact information:

**Axel Meier**
HANA Data Warehouse Expert
ax.meier@sap.com