SAP HANA SPS 09 - What’s New?
HANA IM Services: SDI and SDQ

(Delta from SPS 08 to SPS 09)

SAP HANA Product Management

November, 2014
Agenda

HANA IM Services – overview
Smart Data Integration
Smart Data Quality
Summary of value proposition
HANA IM Services : Overview
HANA EIM Services
Simplifying the landscape, lowering data latency, open and extensible framework

- **Simplified**: Simplified landscape; one common modelling environment to provision and consume
- **Real time**: Lower latency due to real time replication and in memory performance
- **Open and Extensible**: Supports data of any shape and size; open framework for new data sources
HANA EIM Services – What’s New in SPS 09?

SDI: smart data integration
Extends HANA by providing a architecture supporting
• All styles of data delivery: real-time, batch and federation (SDA)
• All styles of deployment: on premise and cloud
HANA’s SDI
• Provides both data replication and data transformation services
• Is open and extensible, works on SAP and non-SAP data of any style, shape and size
• Its modeling environment is part of HANA Studio and HANA Web-based Development Workbench
• Extends HANA’s transformation capability by integrating ETL-type transformations natively in HANA
• Accelerates performance through a native HANA implementation

SDQ: smart data quality
Provides advanced transformations to support DQ functionality
Business value of HANA EIM services

Today, we have several EIM products available for specific use cases:
Replication products (Sybase Replication Server, SLT) to build on premise, operational BI platforms
Integration and Data Quality products to build analytic platforms or solutions
  • On premise platforms, with Data Services
  • In the cloud solutions, with HCI/DSoD

With HANA IM services we now have a single product that
Is able to address all the use cases above
Allows developers to address new use cases, where we can freely combine
  • Data sources: SAP and non-SAP (on DBMS, Hadoop, Hive, OData)
  • Data source location and target deployment: on premise, cloud
  • Data delivery mode from source to target: real time, batch, federation

E.g., integrate, in a real time analytic platform deployed in the cloud: (i) on premise, ECC SAP data, (ii) on premise DBMS data and (iii) federated data in the cloud exposed by an OData service
SDI: Smart Data Integration
Smart Data Integration
Extending HANA by integrating real time delivery mode

SDI provides *real time push mode* to replicate sources
• On selected sources with change data capture (CDC) capability
• Leverages proven Sybase Replication Server technology

Provides *batch pull mode* for all types of sources

Interacts well with HANA’s federation technology (SDA)
• Leverages SDA anytime remote data is requested (initial data load, queries, etc).
• Extends SDA connectivity by providing an Adapter SDK

Provides Transactional Integrity for real time push
By listening to changes in the DBMS transaction logs and only replicating committed changes

Provides Guaranteed Delivery for real time push
• SDI can resume processing if replication stream is halted or disrupted
• SDI can continue to operate during a temporary absence of the HANA target
Smart Data Integration
Transformations

Extends HANA’s capability by integrating ETL-type transformations natively in HANA
Transformation inputs include
• HANA tables
• Virtual HANA tables
• Views

Provides advanced UI to define transformation flows
In HANA studio’s Application Function Modeler (AFM)

Execution modes
Batch
Real time
Smart Data Integration
Available transformations

Basic SQL – oriented
Filter, Join, Union, Sort, Map

Advanced SQL – oriented
Aggregation, lookup, sort, case, and pivot/unpivot

Addressing the data movement lifecycle
Row generation, date generation, table comparison and history preserving

Executing code
Script, function, procedure
Smart Data Integration
Real time replication

Real time replication
Replication UI (in HANA Web-based Development Workbench)
Simple (only filter and project) transformations
Real time push
Smart Data Integration

Batch data integration

Data flow UI (in HANA Studio)
Complex transformation flows
Smart data access pull
Smart Data Integration
New use case: real time data integration

Real time data integration
Data flow UI (in HANA Studio)
Complex transformation flows
Real time push

Benefits
Transformed data now flows in real time
Improved productivity: no need to write “delta load” flows
Smart Data Integration
Native HANA Implementation of Transformations as Tasks

A task is a long running program executing transforms

Task has a plan that compiles to a HANA calculation scenario
Operations in plan may be implemented by existing calculation nodes or by IM-specific native nodes

Runs in the background

Needs monitoring
## Smart Data Integration
### Adapter functions

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Open / Close</strong></td>
<td>Open or close a connection to a remote source</td>
</tr>
<tr>
<td><strong>Browse metadata</strong></td>
<td>Show the remote source objects (tables / views)</td>
</tr>
<tr>
<td><strong>Import metadata</strong></td>
<td>Add remote source objects as HANA virtual table</td>
</tr>
<tr>
<td><strong>Get remote source description</strong></td>
<td>Allow user inputs for connection</td>
</tr>
<tr>
<td><strong>Get capabilities</strong></td>
<td>Describe real time / batch (<em>push down</em>) and other capabilities</td>
</tr>
<tr>
<td><strong>Read data</strong></td>
<td>Read data from a remote source table or view</td>
</tr>
<tr>
<td><strong>Manage real time push</strong></td>
<td>Specify subscription details, start/stop replication, etc.</td>
</tr>
</tbody>
</table>

*(CDC capable adapters)*
Smart Data Integration
Built-in real time adapters for common sources

Log Reader adapters
For Oracle, MSSQL and DB2

SAP ECC adapters
Implements log reader adapter functionality on top the same DBMS sources as above; In addition, it takes into account
• SAP ABAP Transparent and Clustered tables
• SAP catalog hierarchy metadata

Twitter adapter
• Real time access to Twitter content on hashtags, users, etc.
Smart Data Integration
Built-in batch adapters for common sources

**OData adapter**
Adapter for oData services
Generally accessible from the cloud (e.g., Success Factors)

**Hive adapter**
Adapter for Hadoop
Smart Data Integration
Adapter SDK

Open framework allowing HANA to extend its connectivity to external sources
• Includes a collection of built-in adapters

Framework consist of Java API to build, test and debug adapters
• Adapters run in DP agent process, close to the data source
• DP agent can run in any windows/linux server

Architected for both on premise and cloud deployments
• HANA – DP agent communication occurs without needing to open a port in the on premise firewall
  – Cloud: HTTP
  – On premise: TCP/IP
HANA EIM services
EIM-specific monitoring

Remote subscriptions
Source and target, status, last processed transaction time…
General statistics

DP Agent monitor
Status/statistics on agents and on adapters within Agent

Task monitor
Task complete (yes/no), time it took
Statistics within a task
• Percentage completed
• number of records transformed,
• Transforms executed
SDQ: Smart Data Quality
Transformations enriching data

Cleanse
Parse, standardize and enrich person, title, phone, firm, email and address information within a specified input source.

Geocode
Enrich address data with associated latitude and longitude information
Smart Data Quality
Cleansing Transform Improvements

Simplify cleanse transforms
Single transform deals with
• Person names and titles, phone, email,
• Firm, and
• Address information
(in Data Services, it is in 2 transforms)

Consolidate available configuration options
Improved productivity at functional parity
Summary of Value Proposition
## HANA SPS09 EIM Services - Value Proposition

### Lower TCO
- Simplified Landscape, Integrated modeling environment
- Single product covering multiple use cases
- Build once; deploy on premise or on cloud

### Open & Extensible
- Open framework
  - Data – any style, shape and size
  - SAP and non-SAP

### Real-time
- Ability to replicate and transform data in real time
- Transactional consistency and guaranteed delivery
- Breakthrough Performance (natively built in HANA)
Appendix
Smart Data Integration
A new data provisioning architecture

A new Data Provisioning (DP) Server within HANA
Manages DP **Agents** running outside of HANA

**DP Agent**
Container running on any computer, hosting source **adapters**
Provides connectivity to sources not installed in DP server

**Adapter**
A Java or C++ program implementing the Adapter SDK, specific to a **remote source**

**Remote Source**
The actual connection; contains remote objects which can be imported; represented by **virtual tables** once user decides to import them
Replication UI (in HANA Web-based Development Workbench) allows user to specify simple transformations:

- Remote source and target schema
- Objects (e.g., tables or views) to replicate from the source
- For each object, define columns to delete/add and rows to filter out

**At activation time, SDI generates the following run-time objects:**

- A virtual table and a real, empty table for each object to replicate
- A stored procedure which, for real-time adapters:
  - defines a remote subscription, per object, to capture changes to it
  - A task to perform the initial real table load from the virtual table
  - A call to switch to real time push mode, distributing (pushing) changed queued during initial table load and future changes

**Execution of the stored procedure starts replication**
Smart Data Integration
Batch data integration

Batch data integration
Data flow UI (in HANA Studio)
Complex transformation flows
Smart data access pull

Activation and execution
Batch mode execution as a HANA task

A built-in HANA scheduler is not in scope for SP09
Smart Data Integration
New use case: real time data integration

Real time data integration
Data flow UI (in HANA Studio)
Complex transformation flows
Real time push

Activation and execution
Real time mode execution of transformation flow pushes data directly to HANA task

Benefits
Transformed data now flows in real time
Improved productivity: no need to write “delta load” flows
This call allows your remote source table to be defined as a HANA virtual table

```java
public abstract Metadata importMetadataByName(String objectName, String schema, String owner) throws AdapterException;
```

Sample Implementation

```java
TableMetaData metas = new TableMetaData();
metas.setName(nodeId);
metas.setPhysicalName(nodeId);
try{
    ResultSet rs =
        stmt.executeQuery("use test;" + "SELECT TOP 1 * FROM " + objectName);

    ResultSetMetaData md = rs.getMetaData();
    List<Column> cols = new ArrayList<Column>();
    for(int i=1;i <= md.getColumnCount(); i++){
        Column col = new Column();
        col.setName(md.getColumnName(i));
        col.setType(getAdapterDataType(md.getColumnType(i)));
        cols.add(col);
    }
    metas.setColumns(cols);
}
```

Simple Java Object model to understand

Call underlying database to query the metadata

Return list of columns along with other information (precision, primary keys, etc.)
HANA EIM – SDI Adapter Strategy - 3 pronged

- OOTB adapters for common sources
- Partner adapters (e.g. for competitive packaged applications - SFDC) available on SAP store/HANA marketplace
- Open Source Adapters to spur community to develop additional (e.g. RSS feed adapter)
Disclaimer

This presentation outlines our general product direction and should not be relied on in making a purchase decision. This presentation is not subject to your license agreement or any other agreement with SAP.

SAP has no obligation to pursue any course of business outlined in this presentation or to develop or release any functionality mentioned in this presentation. This presentation and SAP’s strategy and possible future developments are subject to change and may be changed by SAP at any time for any reason without notice.

This document 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. SAP assumes no responsibility for errors or omissions in this document, except if such damages were caused by SAP intentionally or grossly negligent.
How to find SAP HANA documentation on this topic?

- In addition to this learning material, you can find SAP HANA platform documentation on SAP Help Portal knowledge center at http://help.sap.com/hana_platform.
- The knowledge centers are structured according to the product lifecycle: installation, security, administration, development:

  **SAP HANA Platform SPS**
  - What’s New – Release Notes
  - Installation
  - Administration
  - Development
  - References

- Documentation sets for SAP HANA options can be found at http://help.sap.com/hana_options:

  **SAP HANA Options**
  - SAP HANA Advanced Data Processing
  - SAP HANA Dynamic Tiering
  - SAP HANA Enterprise Information Management
  - SAP HANA Predictive
  - SAP HANA Real-Time Replication
  - SAP HANA Smart Data Streaming
  - SAP HANA Spatial
Thank you

Contact information

Subha Ramachandran & Fernando Velez
SAP HANA Product Management
AskSAPHANA@sap.com
No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP SE or an SAP affiliate company.

SAP and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP SE (or an SAP affiliate company) in Germany and other countries. Please see http://global12.sap.com/corporate-en/legal/copyright/index.epx for additional trademark information and notices.

Some software products marketed by SAP SE and its distributors contain proprietary software components of other software vendors.

National product specifications may vary.

These materials are provided by SAP SE or an SAP affiliate company for informational purposes only, without representation or warranty of any kind, and SAP SE or its affiliated companies shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP SE or SAP affiliate company products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

In particular, SAP SE or its affiliated companies have no obligation to pursue any course of business outlined in this document or any related presentation, or to develop or release any functionality mentioned therein. This document, or any related presentation, and SAP SE’s or its affiliated companies’ strategy and possible future developments, products, and/or platform directions and functionality are all subject to change and may be changed by SAP SE or its affiliated companies at any time for any reason without notice. The information in this document is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. 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.