Logical Data Warehousing with SAP HANA

Glen Leslie
October 2016
SAP Data Warehousing Strategy

Key capabilities of SAP HANA
Does the Big Data mean the end of the EDW?

**Myth**: Hadoop and In Memory make an EDW obsolete

**Fact**: The EDW will leverage Big Data capabilities

No … but the role is changing
Why is Data Warehousing still necessary?

Characteristics
- Consolidates data across the enterprise
- Standardized data model
- Supports decision making

Main Tasks
- Define common semantics
- Harmonize data values
- Establish a ‘single version of truth’
- Provide a single, comprehensive source of current and historical information
In a world of “Cloud Everything”…

**Gartner**¹ “Emerging data sources, trends and technologies challenge the effectiveness of data warehouses in supporting analysis and decision making.”

**IDC**²: “The data warehousing market based on relational databases will continue to be disrupted by several nonrelational and/or nonschematic information management software categories. Data warehouses will not disappear as they have a key place in an organization's data architecture.”

---

¹ “2016 Strategic Roadmap for Modernizing Your Data Warehouse Initiatives” Mark Beyer and Lakshmi Randall, Gartner, October 2016
SAP HANA DW – Strategy

Planning and definition
2015

Execution and delivery
2016 - 2018

Vision

Analytics
(SAP BI Suite, Predictive, Planning)

SAP HANA DW

SAP Power
Designer

SAP DW
Foundation

SAP BW

SAP HANA
EIM

SAP HANA
Plattform

SAP HANA Vora

Market presence in data warehousing with a clear roadmap

Strong and simplified offering with tight integration

Convergence into one technology stack addressing BW and SQL-based DW needs

© 2016 SAP SE or an SAP affiliate company. All rights reserved.
Serve standard **SQL-based** and **BW-style** data warehousing in order to …

meet future demands

- **Logical Data Warehouse** for dynamically changing system landscapes
- **Cloud and hybrid deployment**
- Integration of any data types and **Big Data technologies**
- Scale out to high volumes and **data lakes**

go beyond other DW offerings

- Top **out-of-the-box integration** to SAP solutions – on-premise and in cloud environments
- **Real-time processing power** of SAP HANA
- Hadoop integration with **SAP HANA Vora**
- HANA-based **analytic business services**
- HANA-optimized re-usable **business content**
Planned Delivery Focus

**Modeling & Metadata**
- Integrated top-down modeling of DW artefacts with SAP Power Designer
- Consistent release management and impact analysis across DW models
- Consolidation of BW modeling objects optimized for SAP HANA

**Data Management & Processing**
- Unified data processing across databases and data lakes
- High performance business services, e.g. for inventory handling, planning and resource allocation

**Data Access Services**
- SAP HANA EIM becomes the central data integration component of the SAP HANA DW
- Flexible adapters for logical data warehousing covering SAP, third party and Big Data sources

**End-to-End Operations**
- Uniform scheduling and monitoring services for SAP HANA DW data flows
- Advanced data distribution services for scale out and dynamic tiering
- Comprehensive life cycle services for SAP HANA DW components
SAP HANA DW – Component View

SAP HANA data warehouse

- SAP Agile Data Preparation
- SAP Power Designer
- SAP Web IDE
- SAP BW
- XS advanced
- Git Hub
- HDI*
- SAP HANA EIM Service
- SAP HANA Application Lifecycle Management
- SAP HANA Data Warehousing Foundation

* SAP HANA Deployment Infrastructure
Extend the Warehouse: SAP HANA as a Platform

Run and scale applications on premise and in the cloud

- Polyglot runtime containers for SAP HANA application development like Node.js or Java
- New SAP HANA Deployment Infrastructure (HDI)
  - Multiple times deployment of isolated native SAP HANA content
- Established standard development tools and processes
  - E.g. Git for version control
- Run SAP XS classic applications (XSJS) as first class citizen of SAP XS Advanced
- Unified web based development environment for end-to-end native SAP HANA applications
Reinventing the Data Warehouse as a Logical Data Warehouse

Results at the Speed of Memory

Data Fabric Layer

- Real-time Events/Machine-generated Data
- Column Storage
- Operational RDBMS
- Petabytes of Structured Data
- Other Sources

In-memory Platform

- SDA
- MapReduce/Hive
- ETL & Rep for RT sync

Business Applications

Tightly integrated orchestration for management, monitoring, and control

Stream

SQL
Logical Data Warehouse

Key capabilities of SAP HANA
SAP HANA Data Warehousing Foundation 1.0 delivers specialized SAP HANA XS 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 Dynamic Tiering, SAP IQ or Hadoop
- **Data Warehouse Monitor** to provide a comprehensive overview about current and past activities in the data warehouse
- **Data Warehouse Scheduler** to maintain dependencies between single processes with the focus to provision data warehouse models

**Outlook**
Define a data temperature management strategy with DLM – available with DWF

- Persistent storage option for Dynamic Tiering (“Warm Store”)
- Federation to other data storage via DWF (Hadoop, Sybase IQ)
- Displace aged data with varying degree of persistence intelligence
SAP Data Warehousing Foundation
Data Lifecycle Manager (DLM)

Tool-based management of data lifecycle
- Leverage SAP HANA Dynamic Tiering, Hadoop or SAP Sybase IQ with a tool-based approach to setup an aging strategy
- Optimize memory footprint in SAP HANA native use cases
- Define data slices on SAP HANA tables to be displaced from SAP HANA memory
- Optimize union access between the tables

Automated Data Movement between stores
- Generated Stored Procedures to do data mass movement – in and out
- Schedule data movement using HANA tasks

Hardware layout view
SAP HANA - SAP HANA dynamic tiering (specific integration of Extended Storage)

(*) The hosts for SAP HANA dynamic tiering do not need to be based on hardware certified for SAP HANA (you may of course choose HANA-certified hardware)
SAP HANA EIM: Next generation data integration and data quality
Design once, execute anywhere

SAP HANA Platform
Secure, Performance, Monitor, Manage, Accessible

Flowgraph: DI & DQ

HANA Repository
XML

Prebuilt: Common Data Model
EPIC Consortium contributions

IT User
SAP HANA
Web IDE

Business User UI
SAP Agile Data
Preparation

Design

SAP HANA Integration & Quality Services

ON-PREMISE | CLOUD | HYBRID

Native Data Processing
Distributed Data Processing

Any DB

Smart Data Integration
Smart Data Quality
Smart Data Streaming

Secure, Performance, Monitor, Manage, Accessible

ON-PREMISE | CLOUD | HYBRID

Native Data Processing
Distributed Data Processing

Any DB

Smart Data Integration
Smart Data Quality
Smart Data Streaming

Secure, Performance, Monitor, Manage, Accessible

ON-PREMISE | CLOUD | HYBRID

Native Data Processing
Distributed Data Processing

Any DB

Smart Data Integration
Smart Data Quality
Smart Data Streaming
SAP HANA Smart Data Access: Core Logical DW Capability

Key Capabilities:

- Enables access to remote data just like “local” table
- Smart federated query processing
- Supports data location agnostic development
- No special syntax to access heterogeneous data sources
- Cost savings
- Benefit from HANA functionality without moving all data to HANA
- Rapid deployment of high-performance, data intensive transactional and analytical applications
Core Concept: Virtual Tables

SAP HANA smart data access is based on local virtual tables that map to an existing object at the remote data source site.
Modelling HANA Smart Data Access

1. Install Source Driver
   - ODBC / JDBC / Other Driver

2. Define Data Sources
   - Data Source
   - User Credentials

3. Create Virtual Tables
   - Virtual Tables
   - Includes
   - Refers to

4. Create Federated Model
   - HANA View
   - Includes
   - Query

5. Consume
   - HANA Tables
   - Refers to
   - Includes

- HANA Tables
- HANA Client
- ODBC / JDBC / Other Driver
- User Credentials
Technical Architecture

Main features

- Integration with Hana Studio and DDL support (create remote sources and virtual tables)
- Optimized for reducing data transfer
- Support of virtual tables in the calculation views
Supported Data Sources

SAP HANA SDA can access heterogeneous data sources through leveraging both native SDA adapters as well as SAP HANA Smart Data Integration (SDI) adapters

- Supported data sources include: Oracle, MS SQL, Teradata, Hadoop, SAP ASE, SAP IQ, IBM DB2, SAP MII, and many more…
- SDK Support for building custom adapters is also provided
- Key Suppliers (Simba Technologies etc) provide (highly optimized) drivers…

All supported SAP HANA SDA adapters are listed in a central note – 1868209
All supported SAP HANA SDI adapters can be found in the SDI PAM

*Please note: SDI adapters are available for SDA use cases at no additional cost, however unless SAP HANA SDI is included in your license agreement, this does not entitle you to use SDI adapters for any other purpose than SDA*
From SPS 11+ SDA leverages SDI infrastructure
SAP HANA SDA provides access to remote data via both ODBC and JDBC adapters depending on if native SDA adapters or SDI adapters are leveraged for the remote access.

Remote server data types are transformed to HANA data types.

Support for converting SDA adapters to SDI adapters is also provided for certain adapter types.

- Advantages of converting the adapter include that if the ODBC drivers of remote sources are unreliable it does not affect the core database, as well as access to many more remote sources and versions.

### Supported Conversions

<table>
<thead>
<tr>
<th>SDA Adapter</th>
<th>SDI Adapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSSQL (GENERIC ODBC)</td>
<td>MssqlLogReaderAdapter</td>
</tr>
<tr>
<td>ORACLE (GENERIC ODBC)</td>
<td>OracleLogReaderAdapter</td>
</tr>
<tr>
<td>DB2 (GENERIC ODBC)</td>
<td>DB2LogReaderAdapter</td>
</tr>
<tr>
<td>TERADATA (ODBC)</td>
<td>TeradataAdapter</td>
</tr>
</tbody>
</table>
HANA SDA SPARK adapter

SAP HANA SDA leveraging SPARK SQL adapters

• SAP HANA Spark SQL Adapter is a plugin for HANA Smart Data Access framework
• It provides access to the Spark Controller and moderates query execution and data transfer
• SAP HANA Spark Controller enables virtual tables on underlying Hive tables (SPS10)
• SAP HANA Spark Controller enables virtual tables on HANA Vora tables
• Download and Install HANA Spark Controller (see also SAP note: 2177933)
DDL Support

- Select, Insert, Update, Upsert, Delete
- Create/Drop Source (allows for the creation of data source targets)
- Create/Drop Virtual Tables (allows virtual tables, which point to remote tables to be created and modified).
Query Processing and Execution

SAP HANA SDA provides high performance by reducing data transfer between HANA and the remote data source – via query push down (push down filters, aggregates, semi-joins, etc.)

HANA functionality is extended to the remote data source via functional compensation

- Optimizer is aware when HANA specific features cannot be used and compensates for them
  - Built-in functions
  - SQL features not supported in the remote server

Functional translation is supported to include syntactical differences between HANA and the remote server

Parallel execution is also supported
Examples of Functional Compensation

Few examples of functional compensation happening in HANA

• Full outer join – A FULL OUTER JOIN B ON (A.a = B.b).
  • Even if both the tables are on the same ASE server, we would still do the outer join in HANA as ASE doesn’t support it

• STDDEV/VAR for Teradata –
  • We will compensate this in HANA if the STDDEV and VAR are on integer columns as Teradata doesn’t support that

• Also, there might other built-ins that remote server doesn’t support which we will compensate in HANA
Security

SAP HANA SDA allows for accessing remote sources via secondary credentials/technical user, or SSO via Kerberos Constrained Delegation in HANA to HANA scenarios

Kerberos (SSO) for HANA to HANA SDA scenarios

1. SAP HANA user logs on to SAP HANA 1 using any authentication method
2. SAP HANA 1 requests delegation ticket on behalf of SAP HANA user
3. KDC issues Kerberos constrained delegation ticket for user
4. SAP HANA 2 uses the constrained delegation ticket to authenticate the user
SAP HANA SDA is primarily administered through SAP HANA Studio, although monitoring can also be done in SAP HANA Cockpit

HANA Studio
- Create/Drop data sources and virtual tables
- Execute and monitor queries
- Analyze query plans
- Calculation view support for virtual tables

HANA Cockpit
- No differences to SDA monitoring in HANA Studio
- Additionally remote user and transaction-id of the remote query are displayed in the SDA HANA Cockpit
**Overview**
- SAP HANA
- SAP PowerDesigner
- 3rd party monitoring and backup solutions
- “Data vault” modeling as best fit, as it provides the full data traceability and historization a financial institution needs

**Benefits & Goals:**
- Consolidation of several RDBMS data warehouses into one
- Reduction of interfaces → a single source of truth
- Traceability of all data and processing
- High performance reports and simplified slicing and dicing
- Free choice of tools and data schema provided optimal fit for the customer’s scenario
Logical Data Warehousing

Other Key Features of SAP HANA to consider
SAP Agile Data Preparation – Key Capabilities
Discover, Profile, Combine, and Share Data Sets

Ingest data from variety of sources
2. Profile data
3. Combine, shape, enrich, or cleanse data
4. Output data for downstream uses
5. IT Governance team - analyze and optimize user processes
SAP HANA Vora
What’s Inside and What Does It Do?

SAP HANA Vora is an in-memory query engine which leverages and extends the Apache Spark execution framework to provide enriched interactive analytics on Hadoop.

- Make Precision Decisions
- Democratize Data Access
- Simplify Big Data Ownership

Drill Downs on HDFS
Mashup API Enhancements
Compiled Queries
HANA-Spark Adapter
Unified Landscape
Open Programming

Any Hadoop Clusters
Enable Precision Decisions
With Contextual Insights In Enterprise Systems

Gain business coherence with business data and big data

<table>
<thead>
<tr>
<th>In-Memory Store</th>
<th>Application Services</th>
<th>Database Services</th>
<th>Integration Services</th>
<th>Processing Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAP HANA Platform</td>
<td>HANA-Spark Adaptor</td>
<td>Vora Spark</td>
<td>Vora Spark</td>
<td>Vora Spark</td>
</tr>
<tr>
<td>Compile queries enable applications &amp; data analysis to work more efficiently across nodes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Familiar OLAP experience on Hadoop to derive business insights from big data such as drill-down into HDFS data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

HANA-Spark Adapter for improved performance between distributed systems

Compiled Queries
Spark Adapter
Drill Downs

Customer

© 2016 SAP SE or an SAP affiliate company. All rights reserved.
Democratize Data Access for Data Science Discovery

Pursue new inquiries without compromise on data and easily integrate these insights with all data

Enable data scientists and developers who prefer Spark R, Spark ML to mash up corporate data with Hadoop/Spark data easily

Optionally, leverage HANA’s multiple data processing engines for developing new insights from business and contextual data.

Extensive programming support for Scala, python, C, C++, R, and Java allow data scientists to use their tool of choice,
Key Features
Vora SQL Engine

```
int main()
{
    return 0;
}
```

Native Integration

In-Memory Storage

Fast Column Scans

Hadoop Components

Compressed Columns

Cache Efficient Algorithms

Multi Platform

Parallel Query Processing

Code Generation
Bringing Enterprise Grade Capabilities to Hadoop
Release 1.3*

SAP HANA Vora provides sophisticated in-memory distributed computing engines to enable complex analysis of Hadoop data

- OLAP engine - enhancing Spark SQL to include data hierarchies for OLAP and drilldown analysis
- Graph engine - graph specific algorithms to facilitate simple and rapid analysis of complex network structures
- Time series engine - enabling efficient compression and analysis of data collected at regular time intervals
- Document store - analyzing schema-less JSON format files
- Disk store - intelligently moving less frequently accessed data from in-memory to conventional disk storage

Additional enterprise grade features

- Hierarchy processing
- Currency conversion
- Bi-directional data virtualization between SAP HANA and Hadoop
- Data modeler - web interface with SQL editor, data browser, and graphical drag-and-drop modeling to communicate with various Vora engines

SAP HANA Vora provides an integrated framework to combine relational, time series, JSON documents, and graph processing without the need to stitch together multiple systems.

* in beta
SAP HANA Vora Deployment Scenarios (1/3)

* as of rel. 1.3 (in beta)
SAP HANA Vora Deployment Scenarios (2/3)
The Lambda Architecture

Batch Layer
- High latency, high throughput
- Compute official result

Speed Layer
- Low latency
- Compute approximate update to last known result

Serving Layer
- Real-time
- Merge batch/speed results
Lambda Architecture
Customer Example

- **LTE**
  - Things
  - MQ
  - Kafka
  - Alerts
  - Validate & Aggregate Messages

**Low Latency – High throughput**

- **SAP HANA Smart Data Streaming**
- **Record/Replay**
- **Immutable Copy**
- **SAP HANA**
  - Predictive
  - Spatial
  - Libraries
- **Modern Developer Tools**
- **Reporting with standard BI Tools**
- **High Speed Analytics**
- **All Thing History +++**

**Observations**

- Intelligent distribution eliminates replication of data for analytics
- Each component provides fit-to-purpose analytics
- Each component scales independently for the use case at hand

- Reduced TCO
- Increased Analytical Agility
- Brings the code to the data
- Supports additional usage models
SAP HANA Data Warehouse – Customer Example

DESIGN TIME

Web IDE (DevX)
Power Designer
EIM Services
XS Advanced
HANA Lifecycle Management
DWF

RUNTIME

Agile Data Preparation (ADP)
UI5 / SAP BI
DB Cockpit

SAP HANA DW

DWH Layer

MTA

DLM

Data Mart Layer

SAP HANA EIM (SDA, SDI, SDQ, ESS)

Replicate
ETL
Streaming
ETL
ELT
ETL / SOA Load
Virtual Access

S4 HANA
Cloud / OnPremise
BW on HANA

Calculation Views (.hdbcalcview)
persisted Table (.hdbcds)
FlowGraph (.hdbflowgraph)
SDATables (.hdbvirtualtable)
HANA EIM Services
SDI
SDQ
ESS

HANA Lifecycle Management
Power Designer
EIM Services
XS Advanced
HANA Lifecycle Management
DWF
SAP HANA Data Warehouse
Future-Proof Data Management Platform for Analytics

Serve standard **SQL-based** and **BW-style** data warehousing in order to …

meet future demands

• **Logical Data Warehouse** for dynamically changing system landscapes
• **Cloud and hybrid deployment**
• Integration of **any data types** and **Big Data technologies**
• Scale out to **high volumes** and **data lakes**

go beyond other DW offerings

• Top **out-of-the-box integration** to SAP solutions – on-premise and in cloud environments
• **Real-time processing power** of SAP HANA
• Hadoop integration with **SAP HANA Vora**
• HANA-based **analytic business services**
• HANA-optimized re-usable **business content**
Further information

**SAP Public Web**
SAP In Memory Data Warehousing: [http://www.sap.com/imdf](http://www.sap.com/imdf)
SAP Data Warehousing Customer Stories: Click here
SAP HANA Vora Test Drive - [http://testdrive.saphanavora.com](http://testdrive.saphanavora.com)
SAP HANA Express download - [https://go.sap.com/developer/topics/sap-hana-express.html](https://go.sap.com/developer/topics/sap-hana-express.html)
SAP HANA Cloud Platform - [https://account.hanatrial.ondemand.com/register](https://account.hanatrial.ondemand.com/register)
SAP HANA Cloud Appliance Library (CAL) - [https://cal.sap.com/](https://cal.sap.com/)

**SAP Education Opportunities**
SAP HANA Academy: [https://www.youtube.com/user/saphanaacademy](https://www.youtube.com/user/saphanaacademy)
SAP Education: [http://training.sap.com](http://training.sap.com)

**Watch SAP TechEd Online**
[www.sapteched.com/online](http://www.sapteched.com/online)
Thank you

Contact information:
Glen Leslie
SAP HANA COE
glen.leslie@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.