Manage Data Access, Integration, and Quality Across Your Digital Enterprise

Table of Contents

2 Simplify Information Management to Improve Analytics and Operations

4 Use Cases for SAP HANA Smart Data Integration and SAP HANA Smart Data Quality

6 Using SAP HANA Smart Data Integration for Data Provisioning

8 Using SAP HANA Smart Data Quality to Deliver Data You Can Trust

10 Gaining Powerful New Data Quality and Data Integration Functionality

11 Empowering Users with Self-Service Data Preparation

13 Tallying the Benefits of EIM with SAP HANA
Simplify Information Management to Improve Analytics and Operations

You swim today in exploding amounts of data from internal and external sources. Keep afloat by taking advantage of the integrated data quality, data integration, and data virtualization capabilities of SAP HANA®, a next-generation platform for in-memory computing. Help simplify your IT landscape and enable successful information management for all your analytical and operational implementations.

DELIVERING TRUSTED, COMPLETE, AND RELEVANT INFORMATION

Job one in the 21st century is digital transformation. But how exactly do you become a digital enterprise? At SAP, we designed SAP HANA to help you harness the latest digital technologies in the service of these overarching goals:

• Faster time to insight and action with real-time analytics
• Improved operational efficiency and productivity
• Agility to respond to rapidly changing market conditions
• Simplification of the IT landscape with integrated capabilities

Managing your valuable information successfully is the key to thriving today. With the data access, integration, and quality capabilities of SAP HANA, you can deliver immediate and targeted insight, so all stakeholders can act in the moment and make quick decisions based on complete, timely, and accurate data. You can also streamline processes to make your operations more efficient and effective – from your inner core outward to the edge of your global business network, including cloud platforms, mobile devices, and the Internet of Things. And you can seize fleeting moments of opportunity, reimage long-standing business models, and create radical new sources of value that propel you to a leadership position in your industry.

LEARNING TO RUN SIMPLE

If you’re like many other companies today, you’re challenged merely to survive with an inherited web of solutions that maintain your data in numerous silos. That kind of complexity slows progress to better business results – stunting growth, reducing profits, and increasing risk. And in the enterprise information management realm, if you bring your current data problems with you, you lose the chance to Run Simple and realize the long-term benefits of digital transformation.

Deliver immediate and targeted insight with the powerful capabilities in SAP HANA for managing data access, integration, and quality.
STRIVING FOR INFORMATION EXCELLENCE

With SAP HANA, you no longer need to rely on a bundle of stand-alone data management solutions. You can radically simplify your IT landscape by using a consolidated set of data management capabilities offered on a single unified platform, as shown in Figure 1.

SAP HANA offers native real-time data replication; data virtualization; bulk and batch data movement through extract, transform, load (ETL) and extract, load, transform (ELT); data transformation; and data quality management. It provides all this functionality on a single in-memory computing platform that consolidates all data sources and accelerates data-related tasks exponentially. It slashes the number of moving parts in your information management landscape and lets you build processes once and deploy them multiple times – on premise and in the cloud.

As part of this revolutionary simplification, you can shorten the journey to information excellence and achieve the following goals:

• Acquire information from any source using any style of data integration, including bulk and batch data movement (ETL and ELT), virtualization, and real-time replication
• Deliver the information that’s fit for use in your business processes and analytics by improving data quality
• Keep information accessible and trusted for decision making by improving data integration, making data available when and where it’s needed

Figure 1: SAP HANA Platform with Native Data Access, Integration, and Quality

Any source

Unified integration, access, transformation, and quality services

Any target

RDBMSs (SAP and non-SAP)
Packaged apps (SAP and non-SAP)
SaaS-based applications
Cloud and social media
Big Data
Unstructured and semistructured data

SAP HANA* platform
In-memory data management platform

SAP HANA smart data quality
Data quality and governance
Transformation | Cleanse | Profile | Match | Geocode | Text data processing

SAP HANA smart data integration
All methods of integration
ETL and ELT | Replication | Virtualization

Common user interface, modeling, and metadata management
High-performance in-memory processing platform

SAP* databases
Third-party apps (OData interface)
Third-party databases and data warehouses
Apache Hadoop
Files
Cloud (OData interface)

RDBMSs = Relational database management systems
SaaS = Software as a service
OData = Open Data Protocol
ETL = Extract, transform, load
ELT = Extract, load, transform
Use Cases for SAP HANA Smart Data Integration and SAP HANA Smart Data Quality

SAP HANA provides both smart data integration and smart data quality. This section describes three typical use cases for these capabilities.

DATA INTEGRATION
Few, if any, computer systems in major organizations stand in isolation. As businesses transition to the digital economy and rely on initiatives involving Big Data, the Internet of Things, and business value networks, the ability to move and integrate data efficiently, securely, and reliably becomes increasingly important.

There are three major styles of data integration: bulk and batch, real-time replication, and federation (now popularly known as virtualization). SAP HANA smart data integration software supports all three, making it the ideal choice for data integration tasks in SAP HANA–centric landscapes. By handling both initial loading and real-time data capture in a single flow graph, the smart data integration option simplifies the process of performing these data integration jobs.

ANALYTICS
Data quality is extremely important in analytic applications. If you are basing your analysis on faulty, incomplete, or incorrect data, you cannot get a clear view of your business.

Analytics have traditionally been used for historical reporting, that is, to tell you what happened last month or last week. With today’s real-time analytics, however, this historical view now sheds clear light on what happened just a few seconds ago. What’s more, companies increasingly depend on data for predictive analysis (to determine what will happen) or prescriptive analysis (to decide what you should do about it). As a result, data quality is even more important, as divining the future from an incorrect picture of the past and present leads you down the wrong path.

SAP S/4HANA® breaks new ground by allowing you to run analytics on the same copy of data used for transactional processing. In the past, if you had poor quality data in your SAP® ERP application, you could at least cleanse it en route to the data warehouse to improve your analytics. But, with just one copy of the data, all the information had better be correct from the start (or corrected as quickly as possible before negatively affecting business). SAP HANA smart data quality software is designed expressly to cleanse data for analytics within the SAP HANA platform.
CUSTOM DEVELOPMENT
Perhaps the most exciting thing about SAP HANA smart data integration and SAP HANA smart data quality is that they are created from the ground up as built-in platform services in SAP HANA. As a result, developers are now able to build data integration and data quality directly into their applications. They can both ingest and improve data using the same development tools that they use to build SAP HANA applications. This is an entirely new breakthrough in the development world.

Imagine developing a custom analytic application, but instead of hoping that users will deduplicate the data, actually taking care of that within your application. Imagine how much better and more accurate your application would be. Developers think of application development and integration development as separate domains, but with SAP HANA they are now merging into a single set of competencies. And developers can perform them both with a simple and unified tool set that allows them to focus on what they want to do without worrying about how to do it.

SAP HANA offers native real-time data replication, data virtualization, bulk and batch data movement through ETL and ELT, data transformation, and data quality management on a single in-memory computing platform.
Using SAP HANA Smart Data Integration for Data Provisioning

Data provisioning is the process of providing data to users (and to applications) in an orderly and secure way. In the context of this section, think of data provisioning as simply getting data into the SAP HANA database from external sources, as well as moving it out to other targets.

SAP HANA smart data integration consists of built-in, bidirectional data movement capabilities that allow the database to ingest data from a wide variety of sources and move data out to a large number of targets. You can move data using bulk and batch or real-time replication, or you can simply virtualize the data. Data virtualization is particularly useful if you want to create a quick proof of concept or construct test scenarios to prove the value of certain data before fleshing out permanent data movement processes. It is also valuable for accessing and exploring Big Data sources.

SAP HANA smart data integration is available preconfigured with a host of built-in adapters for common data sources, including SAP and third-party databases, Big Data from applications and transactions, text documents, e-mail, and instant messaging. You can easily customize adapters with either on-premise or Web-based design tools and develop custom adapters with an open software development kit.

APPLYING MULTIPLE STYLES OF DATA PROVISIONING IN SAP HANA

You may need to use several different styles of data provisioning in one SAP HANA system. That’s why SAP HANA smart data integration supports all three of the major methods of data provisioning, as described below.

**Bulk and Batch Data Integration**

Bulk and batch data integration is used to move large quantities of data in a single job (or in multiple related jobs). Even for real-time replication projects, the first step is usually a bulk load.

Bulk data loading typically involves transformation, the process of changing the data from the source format to the target format. This can include structural transformation (for example, converting field names, joins, data type changes, and so on) and functional conversions (such as value mappings and lookups).

Turn raw operational data into enriched information that fuels real-time analytic and governance applications.
Real-Time Replication
SAP HANA smart data integration supports both single-table and mass-data replication. It lets you feed the process with an initial load or with the real-time capture of change data. You can set configuration options to add, edit, or remove target columns and filter for the records you need. You can manage load behavior to identify changed records for consuming applications or create a history table. And you can also create logical partitions to decrease initial load times.

Classically, replication moves data as is from one system to another. But SAP HANA smart data integration combines replication with transformation to change the data en route from the source system to the target system. You can also use the replication functionality for high availability and disaster recovery (HA/DR) use cases to provide ready back-up for load balancing or in case of system failure. In addition, you can use replication when data needs to be current in two separate systems.

SAP HANA smart data integration combines the real-time change data capture from SAP and non-SAP sources with complex data transformation capabilities. You can turn raw operational data into enriched information that fuels real-time analytic and governance applications.

Depending on the source, you can run replication as a change data capture (CDC) process, in which any change made to a table is automatically replicated to the target system.

Data Virtualization
In virtualization, data does not physically move from the source system(s) to the target system(s). Rather, virtualization presents a unified view of the data. Any updates to the data are effected on the source system without changing the target system. Data virtualization is provided by SAP HANA smart data access technology, which is incorporated into SAP HANA smart data integration.

Gain open and extensible support for any data volume, data type, and data source with SAP HANA smart data integration.
Using SAP HANA Smart Data Quality to Deliver Data You Can Trust

SAP HANA smart data quality lets you simplify data quality operations by parsing, standardizing, validating, correcting, and enhancing individual, company, and address data in a single transformation. These capabilities help you gain and maintain confidence in the data that your organization uses to run core processes and make strategic decisions.

PRESERVING AND PROTECTING DATA QUALITY
All organizations have data quality problems. Duplicate customer master records, incomplete and incorrect material data, incorrect vendor addresses, and many more quality issues affect profitability, customer satisfaction, marketing and production costs, returns, regulatory compliance, and operational efficiency.

Data cleansing is the process of removing duplicate records, creating a best-quality “golden record,” correcting addresses (with the help of regularly updated postal directories), and enriching data with geocoding or bureau data (such as Dun & Bradstreet). By so doing, you can ensure that all data is correct and complete. With SAP HANA smart data quality, we deliver unified cleansing, enrichment, deduplication, matching, geocoding, and other data management capabilities as callable services atop the native platform, through one user-friendly interface. You can:

- Cleanse individual, company, and address data
- Identify duplicates within a single source or across multiple sources of data
- Select a best record from a group of duplicates
- Enrich address data with geocode information
- Search for addresses by latitude and longitude

You can further support geodependent business operations by identifying nearby addresses and returning U.S. census data where applicable.
DESIGNING DATA FLOWS
With SAP HANA smart data integration and smart data quality, you can design simple or complex data flows. Available Structured Query Language transformations support basic aggregate, filter, join, sort, and union processes, along with advanced case, lookup, pivot, and unpivot operations. Data lifecycle management capabilities include data generation, history-preserving mapping, row generation, and table comparison. You can cleanse, match, and geocode data to preserve and enhance quality, and you can execute programs in the leading data management scripting languages. Best of all, enabling data flows for real time is as simple as checking a box on a screen.

TAKING ADVANTAGE OF SEMANTIC SERVICES
The enterprise semantic services feature of SAP HANA is an optional component that provides an application programming interface to simplify data discovery. These services help developers and business users search SAP HANA catalog objects and remote source objects to identify and suggest relevant business content. They use metadata, semantic information, and table contents to build knowledge graphs that answer users’ search requests. The business-focused search user interface delivers the following functionality:
• Keyword-based semantic search, including functions for stemming and synonym pickup
• User-defined acronyms and abbreviations
• Filters on content type within the search query
• Keyword autosuggest
• Delivery of data sets ranked by relevance

With SAP HANA smart data quality, we deliver unified cleansing, enrichment, deduplication, matching, geocoding, and other data management capabilities as callable services atop the native platform.

REVERSE GEOCODING IN SAP HANA®
A reverse-geocoding function in SAP HANA® smart data quality software can pinpoint a location with latitude and longitude and display the addresses within a radius of that point. This is particularly useful in scenarios such as targeting marketing campaigns in mobile apps or determining the households at risk in a natural disaster or public health crisis.
Gaining Powerful New Data Quality and Data Integration Functionality

At SAP, we understand that the data in your SAP Customer Relationship Management, SAP ERP, and SAP Master Data Governance applications—as well as in your SAP S/4HANA landscape—is vital to making sound business decisions. By enhancing the data integrity within your SAP solutions, data quality helps to establish and maintain trust in your data. To achieve this goal, it is very important to tackle on-entry validation, duplicate prevention, and periodic data cleansing to meet accuracy and completeness requirements.

**SAP Data Quality Management software, version for SAP solutions,** delivers prepackaged capabilities to enforce data accuracy for customer, vendor, and business partner data with a virtually seamless look and feel to your SAP software landscape. This integration brings the benefits of SAP HANA smart data quality directly into your SAP Business Suite applications and provides immediate value without the need for custom development. The add-on provides prebuilt data quality configurations that support real-time address validation, geocoding, batch data cleansing, and batch duplicate record detection.

**SAP Master Data Governance** provides built-in, domain-specific governance functionality to centrally create, change, distribute, or consolidate master data across enterprise system landscapes. It uses SAP HANA smart data quality and SAP HANA smart data integration to integrate and load master data from many different sources and to standardize, validate, and enrich address data. It also matches data to detect duplicates based on customer-specific matching rules and calculates best records.
Empowering Users with Self-Service Data Preparation

Business communities are increasingly moving toward self-service data preparation. The SAP Agile Data Preparation application, which runs on top of SAP HANA smart data integration and SAP HANA smart data quality, responds to the growing demands of business users to clean and shape data to suit their own needs – without involving IT resources. While empowering them to do so, the application lets IT retain the ability to monitor and govern user activity and optimize underlying processes.

Figure 2 shows how SAP Agile Data Preparation works with SAP HANA.

Figure 2: Key Capabilities of SAP Agile Data Preparation

- Ingest data from a variety of sources
- Profile data
- Combine, shape, enrich, or cleanse data
- Output data for downstream uses
- Analyze and optimize user processes
DEEPENING COLLABORATION
Business users, analysts, data stewards, and data scientists can all contribute to improving the quality of information assets across your enterprise. Intuitive, interactive workflows encourage the free exchange of data and ideas. Any stakeholder can enrich the data set with a single click.

GIVING IT A MORE STRATEGIC ROLE
With SAP Agile Data Preparation, you can free your IT staff from low-level data administration chores and give them more time to monitor and control user access rights and proactively deliver polished data sets to the business. You can also use the application to parse, correct, and standardize the quality and trustworthiness of your master data.

IMPROVING DECISION MAKING
SAP Agile Data Preparation helps users explore, integrate, and transform raw data directly into actionable information – with no interdepartmental delay. Users can spot hidden patterns, transform data with one click, export prepared data sets to any analytical tool, and share their insights readily with colleagues. As decisions get wiser, they also get faster through shorter reaction times to new data, business requirements, and market opportunities. Data aggregation options let business users select and acquire a subset of the data before importing it into their workspace. Filtering options let them group data by a specific column, preview a summary, and import the data as a new worksheet.

Empower business users to clean and shape data to suit their own needs with SAP Agile Data Preparation.
The enterprise information management (EIM) capabilities native to SAP HANA help you realize a 21st-century framework for managing your valuable information assets to help ensure information excellence for your digital enterprise. SAP HANA simplifies your IT landscape and speeds operations and analytics by combining transaction and application processing on a single in-memory database. It offers you revolutionary power for smart data integration, smart data quality, smart data access, and self-service data preparation.

SAP HANA smart data integration provides open and extensible support for any data volume, data type, and data source. It delivers structured, text, social, and spatial data from SAP and non-SAP applications on premise, on device, or in cloud or hybrid environments. SAP HANA smart data quality helps you unify, cleanse, enrich, match, and geocode data through callable services with a single user interface. SAP HANA smart data access lets you access remote data virtually, without copying and storing it in SAP HANA. Powered by SAP HANA smart data integration and smart data quality, SAP Agile Data Preparation helps users take decisive action without IT intervention. That means your IT staff has more time for process optimization and other activities that make a more strategic contribution to the business.

As Figure 3 shows, SAP HANA can be the single platform that supports the full functionality you need to run a successful modern business.

### Figure 3: A Comprehensive Platform to Support Your Digital Enterprise

<table>
<thead>
<tr>
<th>Application services</th>
<th>Processing services</th>
<th>Integration and quality services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web server</td>
<td>Spatial hadoop</td>
<td>Data virtualization</td>
</tr>
<tr>
<td>JavaScript</td>
<td>Graph*</td>
<td>ELT and replication</td>
</tr>
<tr>
<td>SAP Fiori* user experience</td>
<td>Predictive</td>
<td></td>
</tr>
<tr>
<td>Graphic modeler</td>
<td>Search</td>
<td></td>
</tr>
<tr>
<td>Application lifecycle management</td>
<td>Series data</td>
<td></td>
</tr>
<tr>
<td>Text analytics</td>
<td>Business functions</td>
<td></td>
</tr>
<tr>
<td>Streaming analytics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Series data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business functions</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Database services**

- Columnar OLTP and OLAP
- Multicore and parallelization
- Advanced compression
- Multitenancy
- Multitier storage
- Data modeling
- Openness
- Administration and security
- High availability and disaster recovery

OLTP = Online transaction processing  
OLAP = Online analytical processing  
ETL = Extract, transform, load  
*Graph is in controlled availability.
Because you can see where data comes from, what transformations it has undergone, and how well it conforms to your in-house standards, you increase trust in your data. By simplifying access across any source in the cloud or on premise and empowering users to find, combine, enrich, and share data on their own, you improve efficiency. And higher data quality and deeper data context help you improve the accuracy of modeling, planning, and analysis. Bottom line? You increase everybody’s productivity and streamline business processes, ultimately lowering operational costs and driving revenues.

LEARN MORE
For more on the features and benefits of the EIM capabilities in SAP HANA, talk to your SAP representative or visit us at [www.sap.com/product/technology-platform/hana/features/data-access.html](http://www.sap.com/product/technology-platform/hana/features/data-access.html).

Increase everybody’s productivity and streamline business processes, ultimately lowering operational costs and driving revenues.