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.
SAP HANA smart data integration and smart data quality are the key data integration and data quality capabilities of the SAP HANA platform. These capabilities are delivered with one unified interface that radically simplifies IT landscapes, while harnessing the power of SAP HANA for in-memory-speed processing.

Quickly integrate, transform, and improve your enterprise data to unlock meaning from all of your organization’s data. Transform your data into a trusted, actionable asset for business insight, streamline processes, and maximize efficiency.
What is SAP HANA smart data integration and smart data quality?

SAP HANA smart data integration and smart data quality deliver real-time data integration, quality and transformation capabilities, as part of the SAP HANA platform, to support both operational and analytical use cases.

- Smart Data Quality
- Smart Data Integration
- Breakthrough real-time performance
- Data of any style, shape, size
- Simplified IT landscape
SAP HANA smart data integration | smart data quality
Functionality for the whole team

Source Experts
- Understands source system/s – navigation, data contents

Analyst
- Translates business requirements to functional requirements
- Virtualized read
- Validate results

Developer
- Translates functional requirements to Flowgraph requirements
- Acquires data for Flowgraph
- Develops Flowgraph
- Test Flowgraph

Manager
- Works with IT Administrator to defines user security
- Work with development teams to deploy Flowgraphs

SAP HANA smart data integration / smart data quality team
SAP HANA smart data integration and smart data quality
Radical simplification of IT landscape, plus in-memory processing power

**Any Source**
- RDBMSs (SAP & Non-SAP)
- Packaged Apps (SAP & Non-SAP)
- SaaS-based Applications
- Cloud & Social Media
- Big Data
- Un-/Semi-Structured Data

**Unified Integration, Access, Transform, & Quality Services**

**SAP HANA PLATFORM**
In-Memory Data Management Platform

**SAP HANA smart data quality**
Data Quality & Governance
Transformation | Cleanse | Profile | Match | Geocode | Text Data Processing

**SAP HANA smart data integration**
All Methods of Integration
ETL/ELT | Replication | Virtualization

**Any Target**
- SAP Databases
- 3rd Party apps (OData interface)
- 3rd Party Databases & DWs
- Hadoop
- Files
- Cloud (OData interface)
SAP HANA smart data integration
SAP HANA smart data integration
Any data source, any integration style – one unified framework

Native integration capabilities
- Supports all styles of data delivery, one framework
  - Bulk/Batch – ETL/ELT
  - Real-time Replication
  - Virtualization
- On-premise, cloud, or hybrid deployments

Advanced design tools
- SAP HANA Web-based Development Workbench
- SAP WebIDE

Open framework
- Out-of-the-box adapters for common sources
- Open SDK for customers and partners to build custom adapters
SAP HANA SDI
Data Provisioning Agent

Lightweight installation inside the customer’s firewall
- Small download from SAP Service Marketplace
- A Deployment Unit (DU) for HANA
- An onPremise agent
- Simple installation and configuration
- Runs on Windows or Linux

Securely transfers data
- Agent is configured to communicate with DP server using TCP/IP or HTTPS

Operates without firewall exceptions
- Communication is always from the agent to the cloud
- No need for VPN, reverse proxy, or other firewall exceptions
- Agent uses long polling: places request to server and waits for response when a task is ready to execute
SAP HANA smart data integration
Architecture – TCP/IP (SSL)

The Data Provisioning Server and Data Provisioning Agent are the two main components of SAP HANA smart data integration. By default, the Data Provisioning Agent is configured to communicate with the Data Provisioning Server using TCP/IP.
The Data Provisioning Agent can be configured to communicate with the data Provisioning Server, via the Proxy application, using HTTP(S). This allows data transfers from on-premise source systems to an SAP HANA instance in the cloud without having to reconfigure firewalls, implement a reverse proxy or VPN solution.
SAP HANA smart data integration
Architecture – Transformations

Data Integration transformations are executed in the Calc Engine, which is a component of the Index Server.
Built-in batch adapters for common sources

- e.g. Success Factors
- ODP-enabled / BW extractors
- Hadoop
- Facebook
- Delimited / Fixed format
- MS-Excel
- Apache Cassandra

- OData Adapter
- SAP ABAP Adapter
- HIVE Adapter
- Facebook Adapter
- File Adapter
- Excel Adapter
- Cassandra Adapter
SAP HANA SDI – Broad Access
Connectivity to Databases, Applications, Files, Social Media and Mainframe

**Databases**
- Access
- ASE
- DB2
- Informix
- Impala
- HANA
- Hadoop Hive
- MS SQL Server
- Oracle
- PostgreSQL
- Teradata
- GreenPlum
- Netezza
- Vertica
- MySQL
- Cassandra

**Applications**
- SAP Bus Suite
  - Extractors
  - BAPI
  - Table

**Files**
- Text (delimited /fixed width)
- Sharepoint
- Hadoop - HDFS
- Outlook PST files
- SOAP Web Services
- ODATA

**Social Media**
- Facebook
- Google+
- Twitter

**Mainframe**
- DB2 z/OS
- DB2 iSeries (Formerly AS400)
Friendly user experience
Highly productive drag-and-drop graphical environment

Fully Graphical user interface

- Drag-and-Drop design paradigm
- Same approach for simple to complex flows, single or multiple flows
SAP HANA smart data integration
Flowgraph – Batch

Overview
Design simple or complex data flows

Available Transformations
Basic SQL
• Aggregation, Filter, Join, Sort and Union
Advanced SQL
• Case, Lookup, Pivot and Unpivot
Data Lifecycle
• Date Generation, History Preserving Map Operation, Row Generation, Table Compare and Data Mask
Data Quality
• Cleanse, Match and Geocode
Code Execution
• AFL, Procedure and R Script
SAP HANA smart data integration
Flowgraph – Real Time ETL

Overview
Design simple or complex data flows
Enable for real time by selecting a checkbox

Available Transformations
Basic SQL
• Aggregation, Filter, Join, Sort and Union
Advanced SQL
• Case, Lookup
Data Lifecycle
• Date Generation, History Preserving, Map Operation, Row Generation, Table Comparison and Data Mask
Data Quality
• Cleanse and Geocode
**SAP HANA smart data integration**  
**Replication Task – Batch & Real Time**

**Overview**
- Single table or mass replication
- Multiple replication behaviors

**Configuration Options**
- Add, edit or remove target columns
- Filter for the records you need
- Change load behavior in order to…
  - Identify changed records for consuming applications: SAP Business Warehouse, SAP Data Services, and others
  - Create a history table
- Create logical partitions to decrease initial load times
Deep SAP Application Integration
Native access to SAP Business Suite

Over 5,000 Content Extractors certified
Business content instead of native tables
No ABAP coding
Initial & delta loads

De-facto SAP data integration

SAP Business Suite (via Tables)

SAP extractors (via Full + Delta)

SAP HANA SDI

3rd Party Databases
Delimited Files

Fully understands SAP application model and metadata

Extractors

SAP

0AA_ACCDET_TEXT - Account Determination
0AA_RECDATA_TEXT - Posted Deprciation/Profit
0AA_RECTYPE_TEXT - Record Type in Report
0AA_SHEETCL_TEXT - Classification of Transaction

Fully able to leverage pre-built SAP Extractors
Treat data as high value asset
Protect and secure data with encryption and masking of sensitive data

Data Encryption via SSL

Data Masking

<table>
<thead>
<tr>
<th>Cust ID</th>
<th>Name</th>
<th>Salary</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>17231</td>
<td>John Carter</td>
<td>140,000</td>
<td>54601-WI</td>
</tr>
<tr>
<td>21113</td>
<td>Eddie Smith</td>
<td>90,000</td>
<td>54601-WI</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cust ID</th>
<th>Credit Card</th>
<th>Exp. Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>21113</td>
<td>1234-2345-3288-1223</td>
<td>Jan 13, 2000</td>
</tr>
<tr>
<td>21113</td>
<td>4567-1111-2222-3458</td>
<td>Apr 2, 2002</td>
</tr>
<tr>
<td>21113</td>
<td>8979-2222-3488-0002</td>
<td>Dec 10, 2003</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cust ID</th>
<th>Name</th>
<th>Salary</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>10000</td>
<td>John Carter</td>
<td>74,456</td>
<td>98321-KZ</td>
</tr>
<tr>
<td>10001</td>
<td>Eddie Smith</td>
<td>134,976</td>
<td>76540-BA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cust ID</th>
<th>Credit Card</th>
<th>Exp. Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>10001</td>
<td>####-####-####-####-1223</td>
<td>Dec 1, 1933</td>
</tr>
<tr>
<td>10001</td>
<td>####-####-####-####-3458</td>
<td>Jan 30, 1951</td>
</tr>
<tr>
<td>10001</td>
<td>####-####-####-####-0002</td>
<td>Sep 15, 1925</td>
</tr>
</tbody>
</table>
## Monitoring & Scheduling

### Data Provisioning Task Monitor

<table>
<thead>
<tr>
<th>Task Name</th>
<th>Task Id</th>
<th>Create Time</th>
<th>Is Realtime Task</th>
<th>Is valid</th>
</tr>
</thead>
<tbody>
<tr>
<td>sap.hana.im.ess.d...</td>
<td>155724</td>
<td>2015-11-25T16:...</td>
<td>FALSE</td>
<td>TRUE</td>
</tr>
<tr>
<td>sap.hana.eim.pm.r...</td>
<td>167263</td>
<td>2015-12-03T21:...</td>
<td>FALSE</td>
<td>TRUE</td>
</tr>
</tbody>
</table>

### Task Overview

- **Start**
- **Schedule**
- **Refresh**
- **Auto Refresh** 10 seconds

### Schedule Task VENDORS.sap.hana.eim.pm...
SAP HANA smart data quality
SAP HANA smart data quality
Powerful, natively integrated data quality services

In-memory data quality performance
- Data Quality processes performed in-memory break down performance barriers

Native data quality capabilities
- Cleanse person, firm, and address data
- Identify duplicates within a single or across multiple sources and select best record from group of duplicates
- Geocoding to enrich address data with latitude and longitude information
- Geocode is fully integrated with SAP HANA Spatial Geocoding Index
- Search for addresses using latitude and longitude coordinates

Simple user interface
- Parse, standardize, validate, correct, and enhance person, firm, address in ONE transformation

SAP HANA

Smart Data Quality
- DQ Assessment
- Cleansing
- Matching & Geocoding
- Best Record
- Metadata and Semantics
- Third-Party Enrichment

enterprise semantic services
- Dataset discovery
- Dataset relationships

Smart Data Integration
- ETL/ELT
- Replication
- Virtualization

On-premise and cloud sources of data
- Relational, semi-structured, and unstructured

Out-of-the-box adapters
- Custom adapters

RDBMS / Hadoop
- Text Data
- Structured Data
- Social Network

© 2019 SAP SE or an SAP affiliate company. All rights reserved. ǀ PUBLIC
Data Quality transformations are executed in the Script Server.
# SAP HANA smart data quality

Accurate data enables trusted operations and analytics

<table>
<thead>
<tr>
<th>Source Data</th>
<th>Parsed, Formatted, Validated and Corrected Data</th>
<th>Survival/Master Record</th>
</tr>
</thead>
</table>
| Sun & Ski
32 rue Gustave Eiffel,
Ris Orangis
Île-de-France
91001
FR | Sun & Ski
32 rue Gustave Eiffel
Ris Orangis, Île-de-France
91000
France | Sun & Ski
32 rue Gustave Eiffel
Ris Orangis Île-de-France
91000 France
+33-01-20-98-16-43 |
| Sun & Ski
32 rue Gustave Eiffel,
RIS-ORANGIS
Île-de-France
91000 France
01.20.98.16.43 | Sun & Ski
32 rue Gustave Eiffel
Ris Orangis, Île-de-France
91000 France
+33-01-20-98-16-43 | |
| Sunn & Ski
32 rue Gustave Eiffel
Île-de-France
91000 France
33 01 20 98 16 43 | Sun & Ski
32 rue Gustave Eiffel
Ris Orangis, Île-de-France
91000 France
+33-01-20-98-16-43 | |
SAP HANA smart data quality
Ensure accurate geolocation information

BI reports deliver more insight when accurate location data is available

• Rapidly process large volumes of data in both batch and real time modes

• Provides address cleansing to increase quality of geolocation results

• Enhance addresses with geocode information to support geo-dependent business operations

• Return nearby addresses and census data (U.S. only) for even more geo-dependent business scenarios

<table>
<thead>
<tr>
<th>Cust Name</th>
<th>Customer Address</th>
<th>Geo Code</th>
<th>Closest Store #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Margaret Roberts</td>
<td>1801 Dowling St., Fort Collins, CO 80522</td>
<td>(40.575874, -105.101652)</td>
<td>1544</td>
</tr>
<tr>
<td>Neil Nevue</td>
<td>942 California Ave, Salt Lake City, UT, 84115</td>
<td>(40.74043, -111.935701)</td>
<td>4403</td>
</tr>
</tbody>
</table>
Common Usecases
Common scenarios
Agile data warehouse / mart

Replicate or transform data – batch or real time – into a data warehouse or mart with / without BW

Most efficient way to populate SAP HANA
Common scenarios
SAP Master Data Governance

- SAP HANA smart data integration loads data into MDG hub
- MDG invokes SAP HANA smart data quality to perform cleansing or de-duplication process

**Most performant ETL / DQ support for SAP MDG**
Common scenarios
Data quality firewall for SAP S/4HANA

- Real-time data quality check on incorrect data entries (e.g. incomplete or wrongly structured address information)
- Real-time verification check against duplicate data entries

Provides data trust and confidence up-front
Common scenarios
Self-service data preparation

Underpin SAP Agile Data Preparation to support the ingestion of data, perform data shaping / enrichment / cleansing and distribution of data

Foundation for SAP Agile Data Preparation
SAP Data Hub – SDI Integration
SAP Data Hub
Simplify, understand and govern the flow of data at scale

**Metadata Catalog**
- Connection management
- Crawling, browsing & searching
- Data preview & profiling

**Data Pipelines**
- Batch & streams
- Structured & unstructured
- Machine Learning

**Data Orchestration**
- Multi system data workflows
- Enterprise application integration
- Scheduling and monitoring
SAP Data Hub
Outlook for Data Integration & Processing Portfolio

- Agile Data Preparation (ADP)
- Smart Data Integration (SDI)
- Smart Data Quality (SDQ)
- Data Services (DS)
- SAP SLT
- Data Quality Management (DQaas)
- Information Steward (IS)

SAP Data Hub
SAP VORA
SAP Data Hub
Current architecture view

Application
SAP HANA, XS Advanced Model

Platform Services
- UAA
- SDI
- Git
- ...
- Metadata Catalog
- Data Discovery & Profiling
- Scheduling & Monitoring
- Data Pipelines
- Access Policies
- Remote Orchestration

Distributed Runtime
Kubernetes Cluster
- SAP Vora
  - Containerized
  - Relational
  - Time-Series
  - Document
  - DB Engines
- SAP Data Hub Pipelines
  - Serverless infrastructure
  - Scripting (JS, Python)
  - Templates
  - Custom Operators
  - Flow-based applications
- SAP Data Hub Flowagent
  - Connectivity
  - Connections

Data Storages
Cloud / On Premise
- Cloud Stores
  - AWS S3, GCP GCS, Azure ADL & WASB
- Hadoop
  - HDFS (optional)
  - Adapter
  - Spark Extensions

Connected Systems
SAP Integration & Open Connectivity
- SAP LT Replication Server
  - Configurations
  - Replication Jobs
- SAP BW
  - Process Chains
  - Data Warehousing Processes
- SAP HANA
  - SDI Flowgraphs
  - Data Integration into SAP HANA
- SAP Data Services
  - Data Services Job
  - Heterogeneous Landscapes
- 3rd party, Open Source
  - Direct Connectivity
  - Messaging, APIs
Roadmap
## Recent innovations

### Connectivity
- Apache Cassandra, DB2 on AS/400
- Enhanced MSSQL adapter to support dynamic ports and windows authentication
- Enhanced Hive adapter to support advanced options
- Enhanced PostgreSQL adapter to support real-time CDC for LOB types
- Whitelists for adapters to limit access to source database
- Support for using credentials stored in data provisioning agent secure storage
- WS-Security support for SOAP adapter
- Support Kerberos authentication for file adapter when writing to HDFS target

### Flowgraph
- Autopartitioning
- Enhanced hierarchical transform to support virtual tables
- R-script node

### Replication task
- New editor for SAP HANA extended application services, advanced model

### Deployment
- SAP HANA smart data integration on Cloud Foundry, SAP HANA, enterprise edition

---

## 2018 – Planned innovations

### Connectivity
- AWS RDS for Oracle, SQL Server, PostgreSQL
- Enhancement to MSSQL Log reader adapter to support configuring multiple remote sources from single MSSQL server

### Flowgraph
- Enhancements to projection node in SAP HANA extended application services, advanced model, to support drag and drop
- Parameter passing to SAP HANA procedures

### Real-time replication
- IBM DB2 on z/OS

### Operations
- SAP HANA cockpit content in SAP HANA extended application services, advanced model
- REST API – support batch, real-time, transactional scenarios in XSA with EIM REST API

### Deployment
- Agent group load balancing
- Agent JDBC WebSockets configuration
- SAML security

---

## 2019 – Product direction

### Connectivity
- SOAP Web service as target
- REST adapter
- SharePoint Cloud Office 365
- Connectivity using SAP Cloud Platform Open Connectors service

### Flowgraph
- Autopartitioned data read
- Spatial data support
- Parsing, un-nesting, and nesting of NRDM data formats
- SAP HANA extended application services, classic model, to advanced model migration tool

### Real-time replication
- SAP S/4HANA ABAP core data services view
- SAP HANA to SAP HANA in the cloud

### Replication task
- ODP objects as source

### Operations
- DP agent-based access control

### Deployment
- Schema-level multitenancy
- Support active/active data systems (SAP HANA)

---

## 2020 – Product vision

### Connectivity
- Intelligent payload handling and routing
- Extended data type support and processing (such as geospatial)
- Self-healing data movement

### Operations
- Enhanced hybrid cloud capabilities
- Extended DevOps capabilities
- Advanced data governance
- Enhanced lifecycle management
- Data consistency verification and remediation

---

1. Potential data protection and privacy features include simplified deletion of personal data, reporting of personal data to an identified data subject, restricted access to personal data, masking of personal data, read access logging to special categories of personal data, change logging of personal data, and consent management mechanisms. 2. This is the current state of planning and may be changed by SAP at any time without notice.
SAP HANA smart data quality
Product road map overview – Key innovations

**Recent innovations**
- **Cleanse**
  - Supported country override option
  - Locality official output column
- **Geocoding**
  - Geocoding now supported for India addresses
- **Match**
  - Match transform enhancements for SAP HANA extended application services, advanced model
  - Candidate selection
  - New output fields for match_score and match_policy
- **Data mask**
  - Support seed value as substitution variable

**2018 – Planned innovations**
- **Cleanse**
  - Integration with SAP Data Quality Management, microservices for location data
  - New Chinese address directory for more accurate and current data
- **Geocode**
  - Integration with SAP Data Quality Management, microservices for location data

**2019 – Product direction**
- **Cleanse**
  - Cleanse transform enhancements for SAP HANA extended application services, advanced model
    - Country identification enhancements
    - Allow users to select side-effect tables
    - Set schema name for side-effect tables
- **Geocode**
  - Points of interest search
- **Best record**
  - New transform
- **Match**
  - Proximity matching
  - Cleanse transform enhancements for SAP HANA extended application services, advanced model
    - Match source output component when multiple input sources used in matching
    - Custom and tie-breaker strategies for the survival rule settings

**2020 – Product vision**
- **Cleanse**
  - Enhance cleansing to product and material master data domains
- **Data enrichment**
  - Enrich data from third-party sources for master data and transactional data

---

**SAP HANA 2.0 SPS3**
1. Potential data protection and privacy features include simplified deletion of personal data, reporting of personal data to an identified data subject, restricted access to personal data, masking of personal data, read access logging to special categories of personal data, change logging of personal data, and consent management mechanisms.
2. This is the current state of planning and may be changed by SAP at any time without notice.
SAP HANA smart data integration & smart data quality
Getting started

Find out more:
http://sap.com/eim
https://help.sap.com/viewer/p/HANA_SMART_DATA_INTEGRATION
https://www.sap.com/community.html

Subscribe to SAP HANA SDI & SDQ videos:
YouTube Channel

Contact your SAP Account Executive
Thank you.

Contact information:

Shibajee Dutta Gupta - shibajee.dutta.gupta@sap.com
Venkat Madireddi - venkat.madireddi@sap.com