Enterprise Data Warehouse Roundtable

Manish Umarwadia and Pravin Gupta, TekLink International
Las Vegas
September 18, 2016

Reducing the HANA Footprint through an Agile Development Approach

Agenda

Reducing HANA footprint

Agile development approach

Dynamic Tiering

SDA and HANA Connectivity using HANA VORA for Big Data

BW/4HANA – what does this bring you

Discussion

Reducing HANA footprint

Reducing HANA Footprint

Archive to NLS

Seamless integration between NLS archive and HANA database in reporting.

Data can be easily loaded back and forth between BW and NLS.

NLS data is highly compressed (~1:10 ratio)

- It is possible to store data in SAP IQ, Hadoop, ORACLE, DB2, DB2BLU, file storage and many more.
- Hadoop software is available online for download free of cost.
- SAP IQ NLS solution provided by SAP
- NLS solution also provided by various SAP partners (eg. PBS Software, Con4Pas, Informatica, DataVard)

Non-Active Data

House keeping Activities

Run HANA sizing report and pick

- Top 10 contributors in master data node
- Top 10 contributors in Slave node.

Our experience shows that with the combination of NLS archiving and housekeeping in SAP BW can result in shrinking the database size by 43%.

Housekeeping Activities

House Keeping Transaction Data:

Maintain PSA

Maintain Change log

Manage DTP Temporary storage

House Keeping Administration Data:

Delete or Archive large basis table created by application logs, change documents, BW Request Administration document etc.

Delete/Archive Application log tables:

BAL* Tables contains logs from different application

Not automatically deleted

Archive/Delete IDOCs.

Delete using transaction WE11

Delete unused Bookmarks Using Report
RSWR_BOOKMARK_DELETE

Delete table entries generated by LISTCUBE Tables
DYNPSOURCE and DYNPLOAD Using report RSDQ_DYNP_GP_CLEAN

Housekeeping Activities

Delete BW Batch Runtime Data using transaction RSBATCH or report RSBATCH_DEL_MSG_PARM_DTPTEMP

Delete unused workbooks

Delete Authorization log **Table RSECLOG Using transaction RSECADMIN**

Maintain change history tables CDCLS and CDHDR **Delete using Program RSCDOK99 Archive using Archive object CHANGEDOCU**

Delete BW Statistics table BW statistics table RSDDSTAT* and the BPS statistics tables UPC_STATISTIC*

Remove unused dimension table entries of an Infocube: Run FM RSDRD_DIM_REMOVE_UNUSED

Remove Temporary Tables Using program SAP_DROP_TMPTABLES

RSAN_RTT_CLEAR_TEMP_TABLES (to drop temporary tables created by analysis process that are no longer used)

FM RSDDS_CHANGERUN_TMPTABLS_DEL

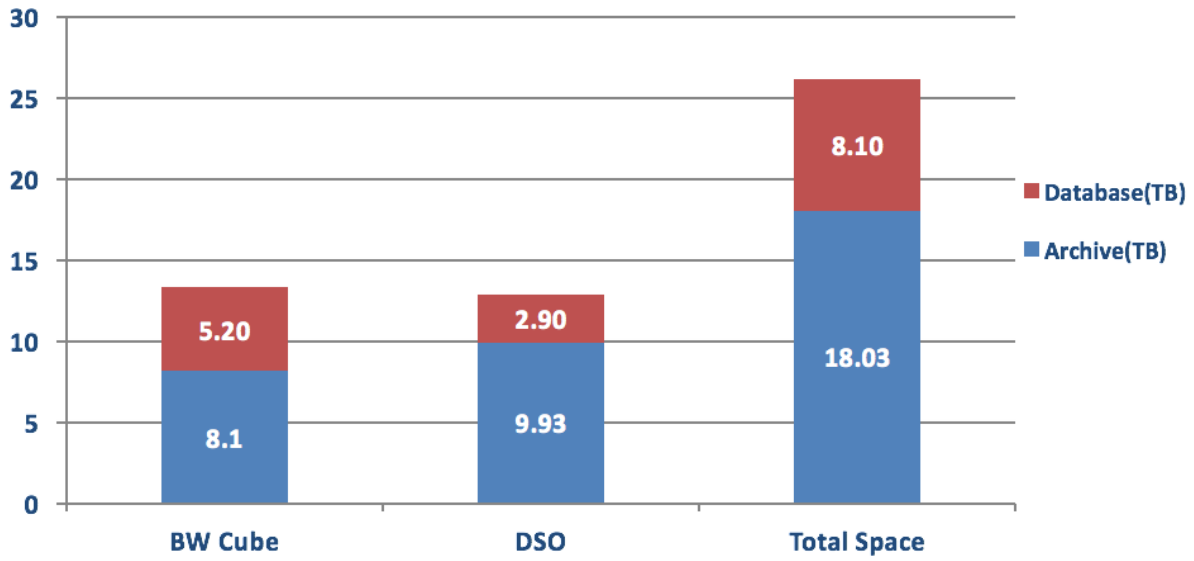
Reducing HANA Footprint

Total BWP Database Size (Nov. 2014) = ~14 Terabytes

Savings with NLS and Archiving = ~18 Terabytes

Benefits:

Savings on storage costs: 1TB = $70,000 annually



ILM Strategy

	Online Database	Near-Line Storage
Frequently read / changed data (actual)	✓	
Infrequently read data (mature)	✓	✓
Very rarely read data (aged)	✓	✓

Agile Development

Agile Development

Project types:

Proof-of-Concept (POC)

Pilot

Data Discovery

Commercialized Solution

Current Options:

Y Queries

BW Workspaces

Analytical Process Designer

Frequent Test/SBX system copies from Production

Challenges with POCs and Pilots:

Migration to Production

Production-like Data

Iterative approach/transports since I don’t know what I don’t know

Testing reveals gaps that were not envisioned

Functional Spec is incomplete since prototyping was done with mock-data

Agile Development – New Options

Virtual Objects with Object Modifiability (changeable original)

Composite Providers

Open ODS views

HANA Analytical Process

Open Hub Destinations

Separate Prototyping Namespace (e.g. /PROT/)

Set to be modifiable in production

Allows creation of all BW object types (cubes/DSO/datasource/transformations)

Naming conventions still enforced

Periodic clean-up of unused objects in this namespace

Normal development process to be followed if decision to commercialize.

No option to limit HANA memory consumption

Proper governance essential

HANA Views (separate package)

Accelerated Transports to Production

Other Options:

Webi – Merge Dimensions

Universe creation (IDT)

Lumira-based Discovery

Dynamic Tiering

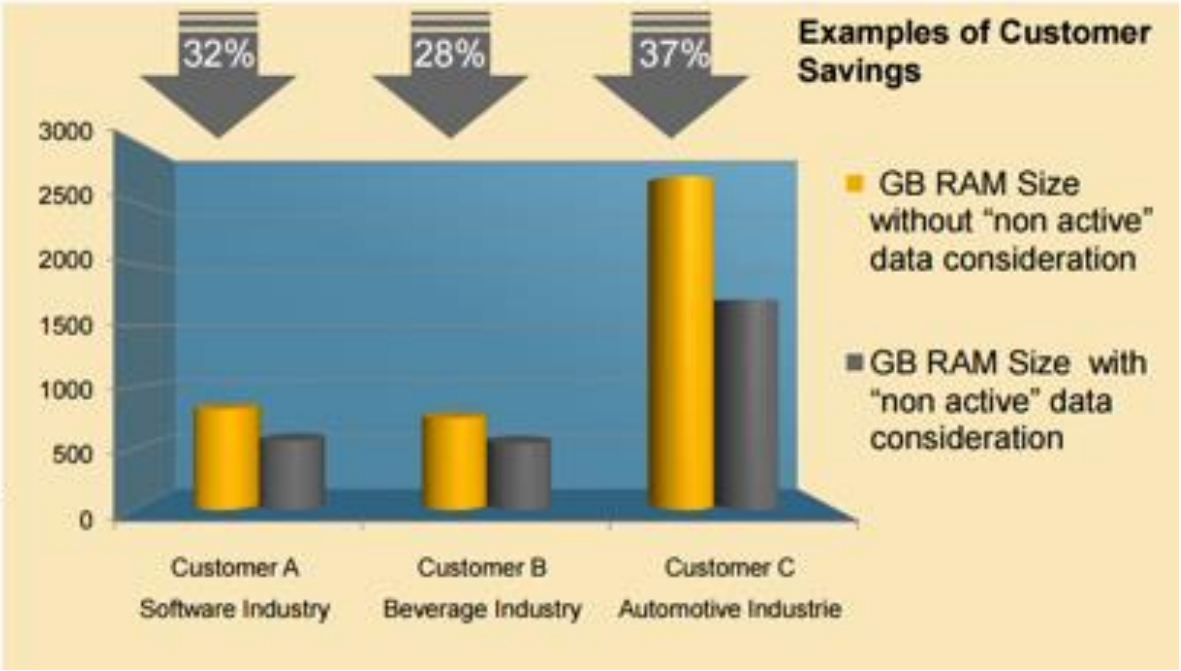
Non-Active Data

"Non Active" Data Concept – SAP NetWeaver BW

Reduces RAM sizing, thereby reducing TCO

Optimized RAM sizing

- "Non active" data concept has substantial impact on SAP NetWeaver BW on SAP HANA RAM sizing
- Assumption: only ~ 20% of "not active" tables are required in memory, rest resides on disk only.
- ABAP Sizing Report (note 1736976) for SAP NW BW on HANA will take "not active" data into account



Dynamic Tiering (DT) with BW

BW Data is classified into 3 categories from a Data Management perspective:

Hot – Frequent access

Warm – less frequent access:

- Non-active (aka early unload)
- Dynamic Tiering (Using Extended storage server concept)

Cold – archive, data lakes

New options for DT only for BW on HANA (Beta in mid-2016)

Use standard HANA nodes instead of extended storage server

- This node uses a different sizing formula and CPU/RAM ratio
- Asymmetric scale-out with standard sizing for hot data and relaxed sizing for warm data
- Hot/Warm configuration to be done via BW application and standard HANA relocation techniques
- Relies on partitioning, pruning and access patterns
- Available for aDSO with 7.50 SP05 using range partitioning

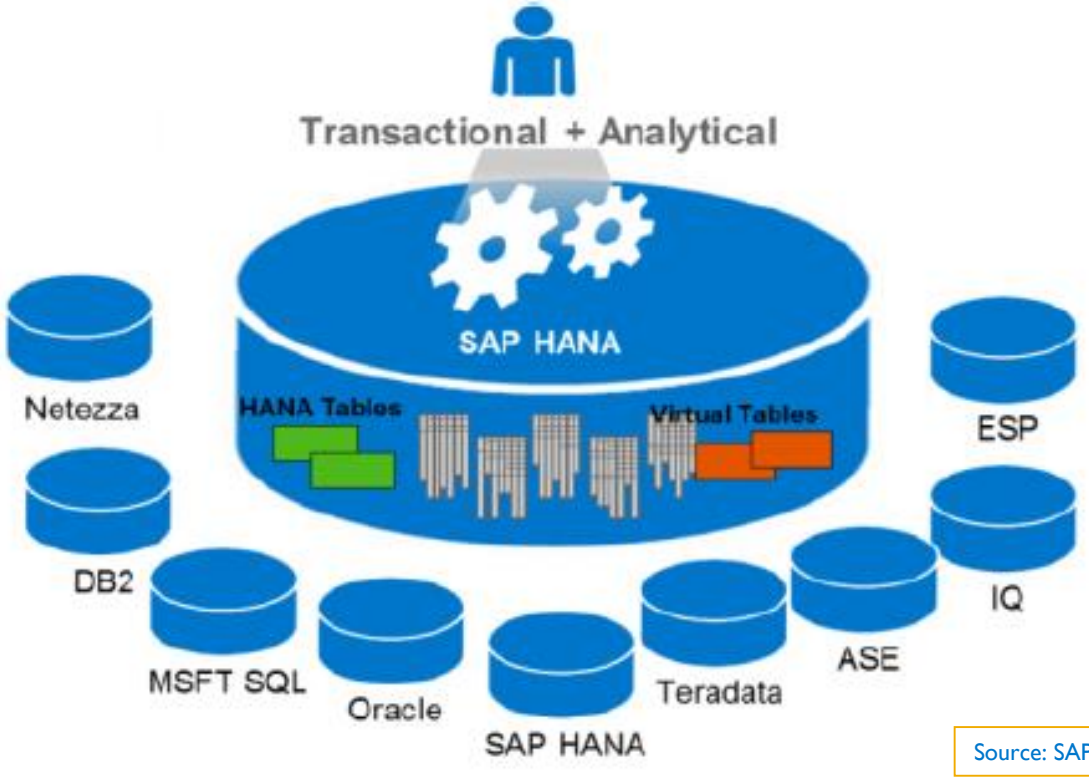
PSA tables, Change log tables and write-optimized DSOs should be able to use the early unload option since there's only one active partition "in-play" even now

- With 7.50 SP01 we can use DT with advanced DSOs (aDSO) provided we range partition them
- With 7.50 SP05 range partitioning of aDSO will be made available using functional tools
- Infocubes and classic DSOs can NOT use DT but could use early unload if they pertain to historical years

SDA and HANA Connectivity using HANA VORA

SAP HANA SDA : Definition

Smart Data Access (SDA) is a data virtualization feature in SAP HANA that allows to access data virtually from remote sources such as Hadoop, Oracle, Teradata, SQL Server and SAP databases and combine it with data that resides in an SAP HANA database.



Source: SAP

SAP HANA Vora – What does it do?

SAP HANA Vora is an in-memory query engine that leverages and extends the Apache Spark execution framework to provide distributed computing at scale and deliver enriched interactive analytics on Hadoop.

SDA and SAP HANA Connectivity

SAP HANA & Hadoop Integration

SQL on Hadoop via SDA (virtual tables) – Hive (SPS06)

Remote caching with Hive (SPS07)

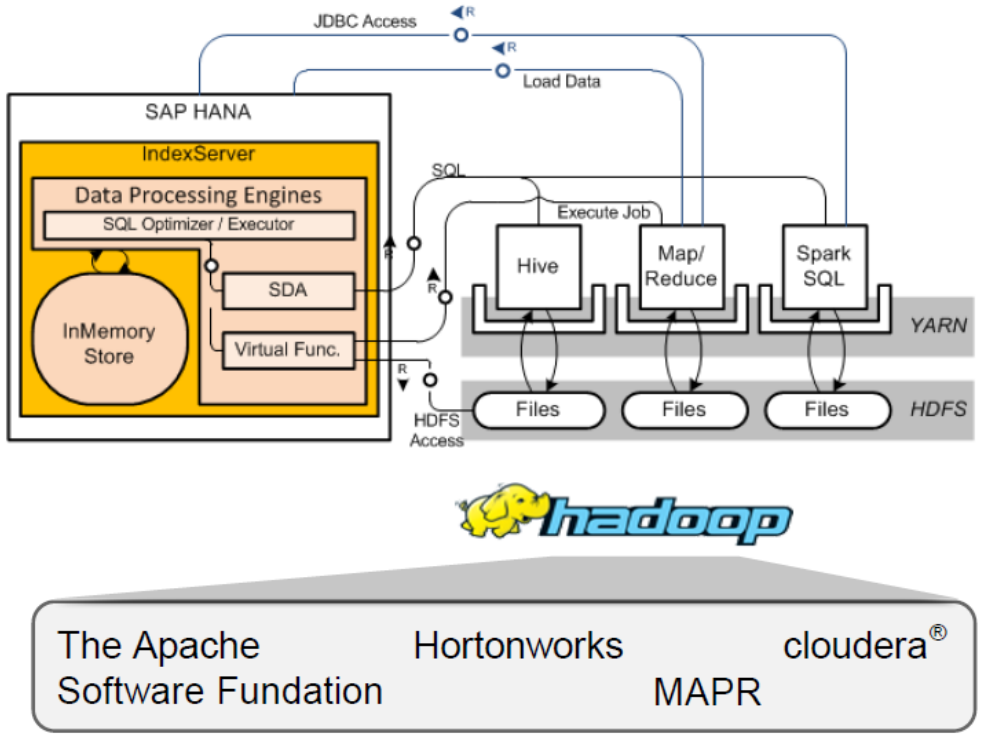
Connectivity to Apache Spark using ODBC

Execution of MR-Jobs via SAP HANA (Virtual Functions) and direct access

SAP HANA Spark controller via SDA (SPS10)

Join relocation to Hadoop through SparkRDD

Unified Admin through Ambari integration for Hortonworks



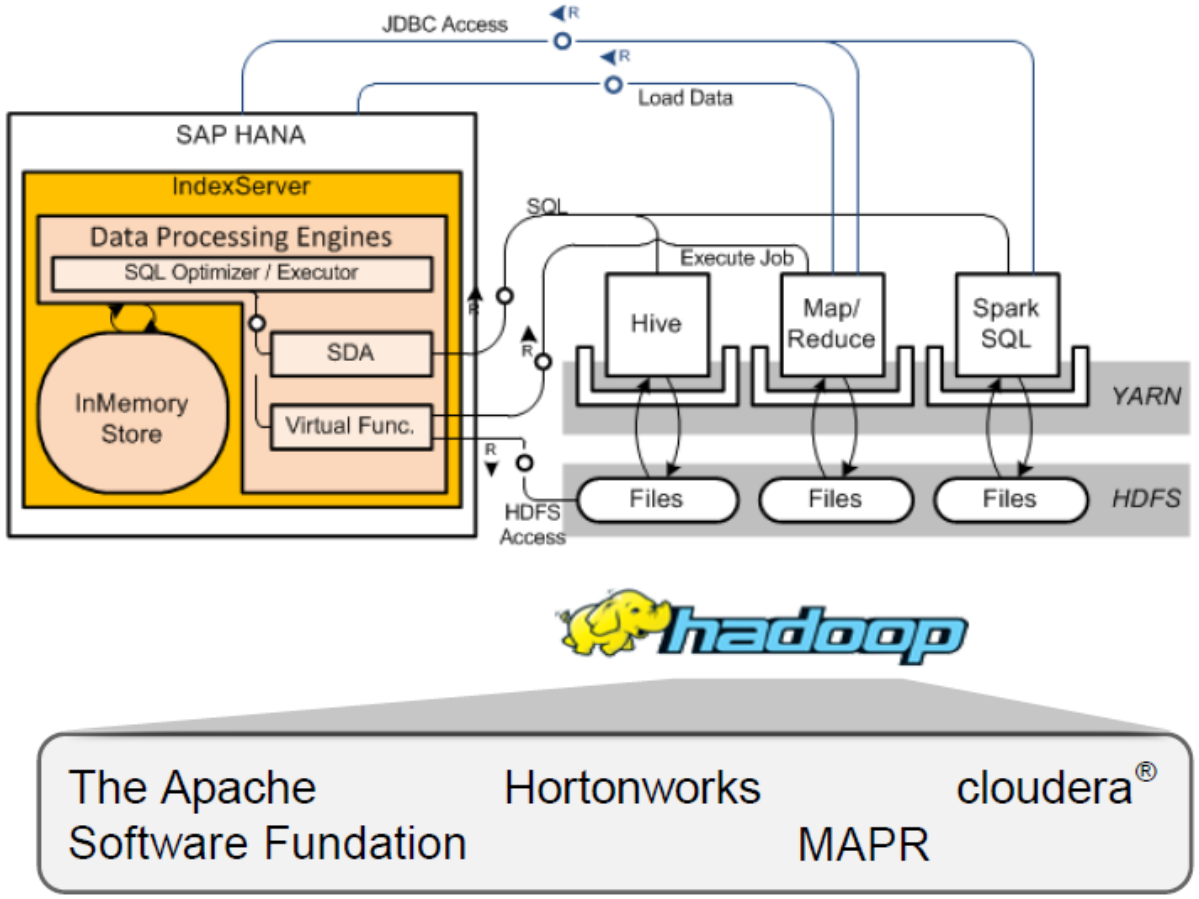
SDA and SAP HANA Connectivity

Key Benefits

Deep integration for storage & processing

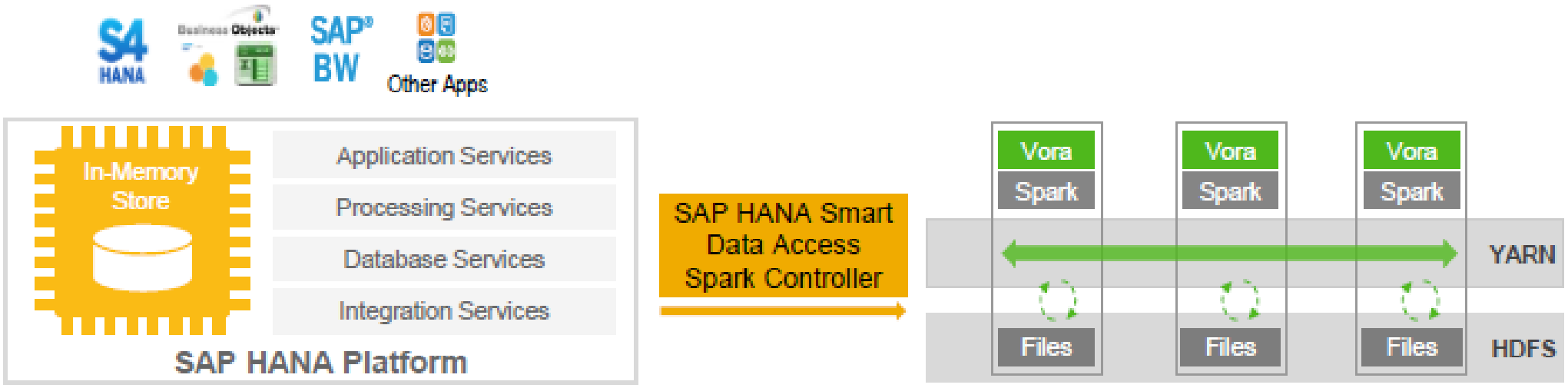
Optimized data access between SAP HANA & Hadoop

Data tiering to Hadoop for cold storage



Hadoop access in SAP HANA – “inside out” scenario

Leverage the SAP HANA Spark controller through SDA to virtually consume Hadoop data processed by SAP HANA Vora without any data duplication



Advantages of SAP HANA Vora connectivity over SDA

Replaces existing odbc/jdbc connectivity with direct access to Hadoop/Spark through SAP HANA Vora using the SAP HANA Spark controller

Deeper integration with Hadoop because the SAP HANA Vora engine runs natively on each of the Hadoop/Spark nodes, and hence has the data locality

Future integrations to Hadoop/Spark from SAP HANA will be driven through SAP HANA Vora. SAP HANA Vora delivers features for data consumption both from Hadoop/Spark side and from SAP HANA natively using calc. views. This extends the platform to data scientists to consume SAP data residing in SAP HANA through the SAP HANA Vora Spark Data Source API.

BW/4HANA – what does this bring you

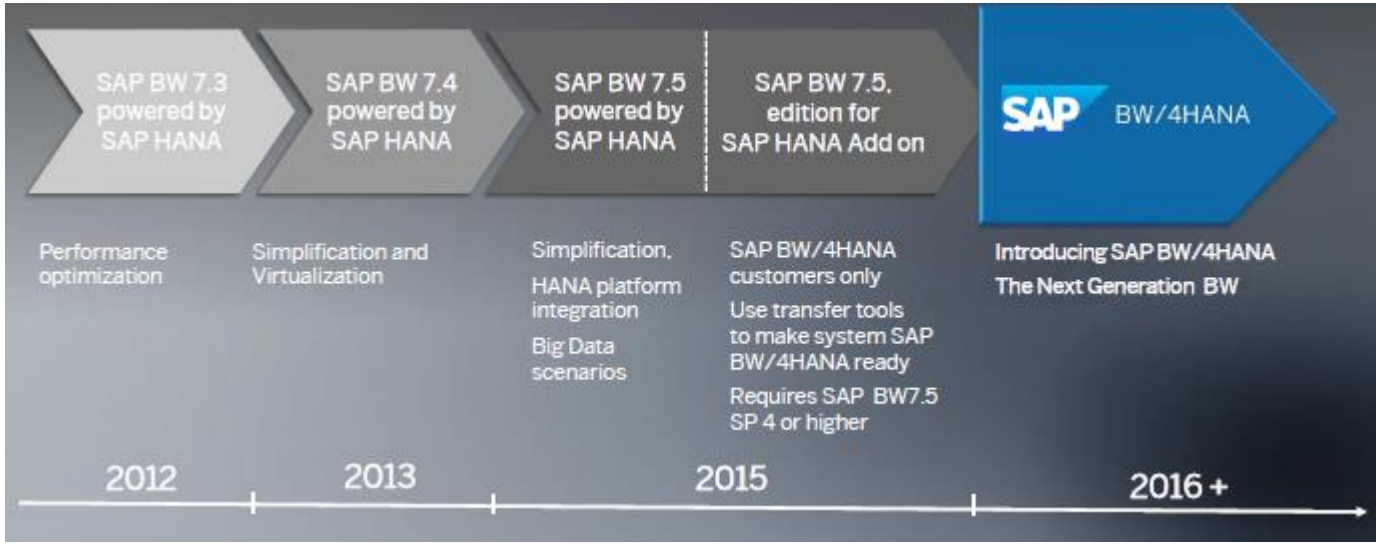
What is BW/4HANA

It is not BW 7.5 powered by HANA or w/ Edition Add-on.
It is a new EDW solution, highly optimized for HANA

- **Simplicity:** HANA optimized objects only
- **Modern UI:** Development and administration
- **Open SQL DW approach** Data Lake scenarios planned
- **Performance**: Push down to HANA

All new future innovations will take place for BW/4HANA

Not a pre-req for S/4HANA



Journey to Consolidation

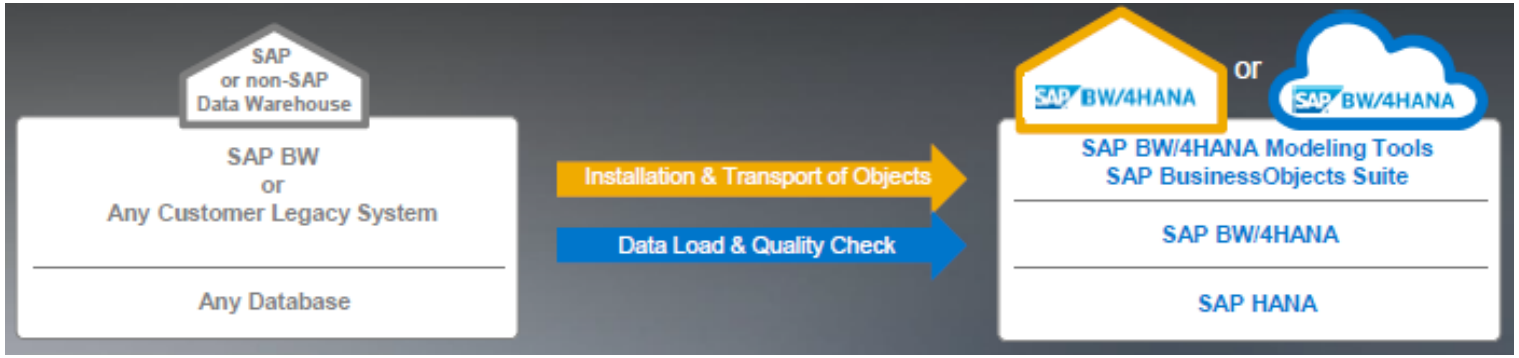
SAP BW/4HANA consolidates existing InfoProviders

- **CompositeProvider as new object to define joins between InfoProviders**
- **Open ODS View for virtual access to external sources**
- **InfoObject to model semantically rich master data**
- **Advanced DataStore Object as the new object for persistence management**

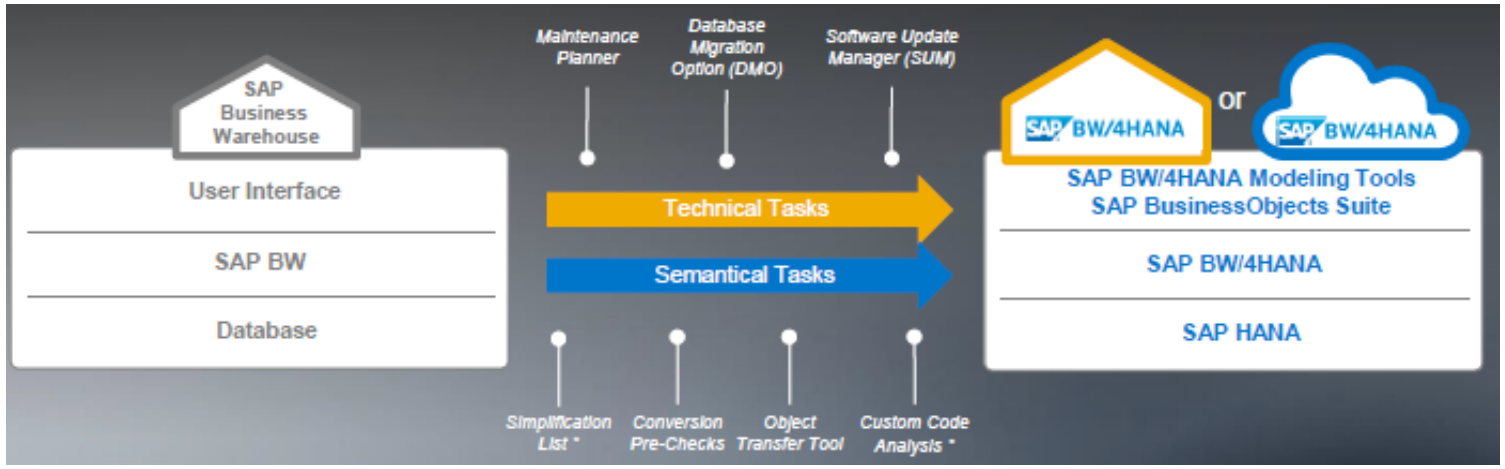
	10 Classic BW		BW/4HANA	4
Virtualization	1	MultiProvider	Composite Provider	1
	2	InfoSet		
	3	Composite Provider		
	4	Virtual Provider	Open ODS View	2
	5	Transient Provider		
Persistency	6	InfoCube	DataStore Object	3
	7	DataStore Object		
	8	Hybrid Provider		
	9	PSA Table		
	10	InfoObject	InfoObject	4

Path to BW/4HANA

Set up system as a Fresh Start. Transport option is available only for compliant objects.
Benefit from Re-engineering.



Convert (*Planned). Upgrade will be huge effort. Conversion Tools planned release Q2/2017



Path to BW/4HANA

Consolidate multiple BW systems (*Planned)

Benefits for Customers

- Stay with current data warehouse landscape and move gradually to SAP BW/4HANA innovations
- Harmonized data models and shared master data through consolidation
- Carve out of single entities of the company to SAP BW/4HANA and leverage process simplification

Implications

BEx is no longer supported

Conversion of data flows (tool available)

ODP extraction, only whitelisted extractors work

Custom objects need to be verified and adapted

New business content based on S/4HANA VDMs

BPC and SPM: Waiting for exact details…

Difference between BW 7.5 and BW/4HANA

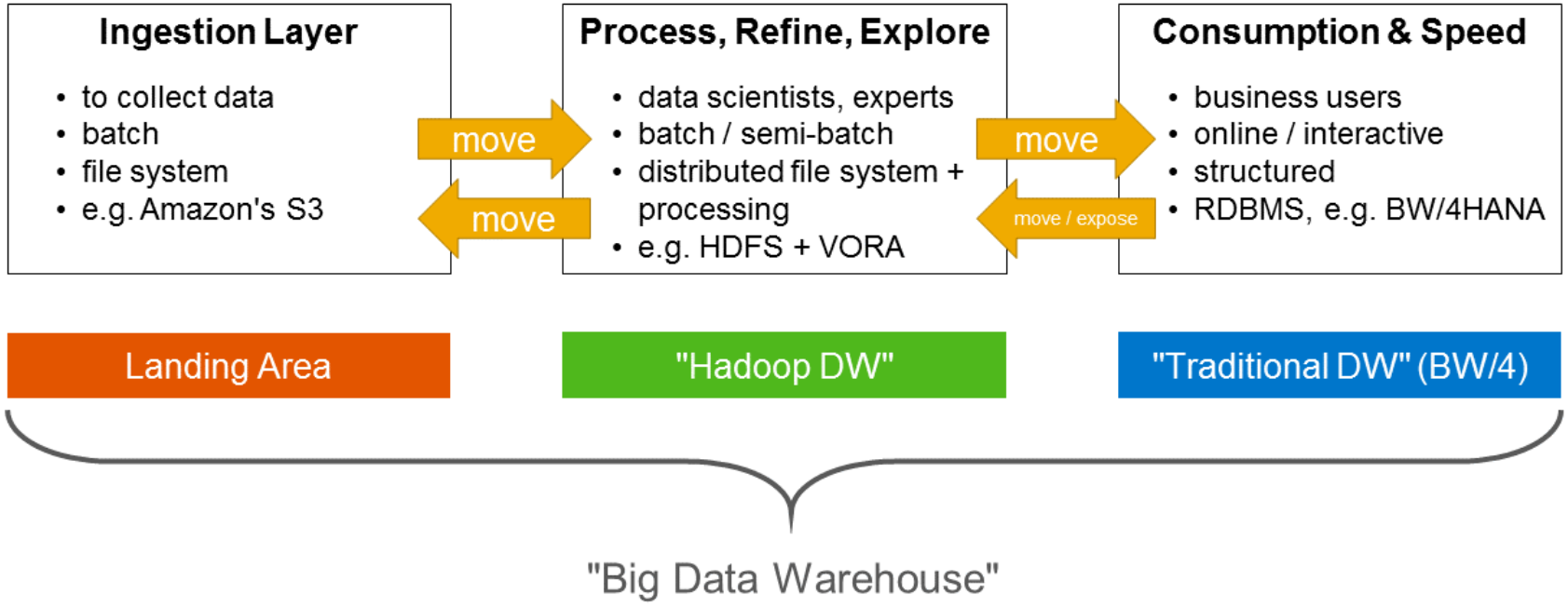
Data flow modeling which can be regarded as a successor to the DW workbench as it has existed for 2 decades. It introduces a new, modern UI and a different paradigm on how to handle, model, manage your DW.

The BW/4HANA roadmap for the next 12 months shows how the gap will open very quickly.

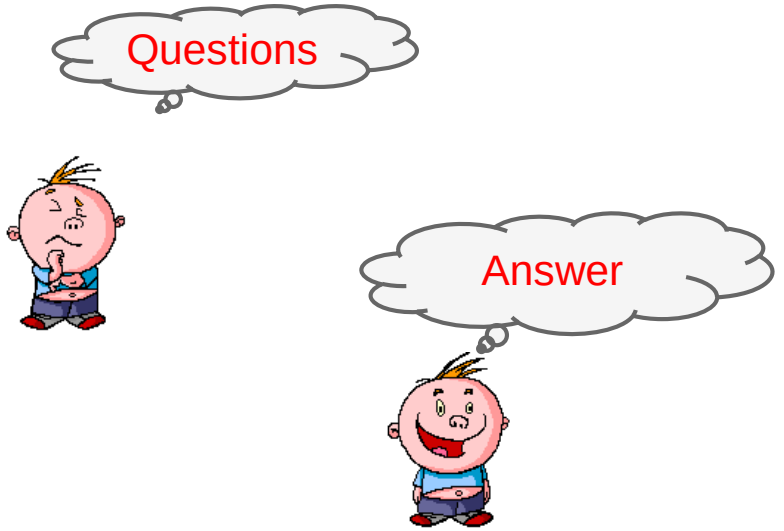
The combination between BW/4HANA, HANA, VORA and Hadoop will be extremely powerful. No backward compatibility.

Enabled for Big Data

BW/4HANA will be free of any burdens and will leverage any optimal access to SAP HANA which will be especially interesting in the context of big data scenarios as SAP HANA VORA offers a highly optimized "bridge" between the worlds of SAP HANA (RDBMS) and Hadoop/SPARK (distributed processing on a file system).



We know you have some questions…



Your Trusted Analytics and Planning Partner

BWoH & HANA Enterprise
Trade Promotions on HANA
Big Data Technologies
BusinessObjects

BPC Consolidation
BPC & IP Planning
S/4HANA Planning

Think TekLink High Performance Consulting

BI Strategy & Roadmap
Big Data Adoption Strategy
BI Center of Excellence
Global Delivery & Support

Innovation Labs & PoC
BPC Accelerators
Cloud & Technical Services

SAP Partner

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.