



Technical Whitepaper

# SIMPLEMENT DATA LIBERATOR™



Deployment Considerations for Enterprises  
running SAP on SQL Server

Simplement, Inc.  
Spring, 2014

## CONTENTS

Overview.....	3
What is the Data Liberator™?.....	3
What this White Paper Covers.....	3
Solution Architecture & Typical Deployment Approach.....	4
Solution Architecture.....	4
RTP Features.....	6
Deployment Patterns.....	7
Scaling.....	7
Installation/Configuration.....	8
Target Liberator - RTP Database Server and Replication.....	8
Source - SAP ECC and other SAP ComponentS on SQL SERVER Databases.....	8
Non-SAP Oracle Source Database.....	9
Non-SAP DB2 Source Database.....	9
Non-SAP SQL Server Source Database.....	9
Non-SAP Sources in General.....	10
Solution Operation.....	10
Complete Data Refresh / Snapshot.....	10
Add a Table to Replication.....	10
Remove a Table from Replication.....	11
Process to handle Changes To Tables in SAP.....	12
Solution Monitoring.....	13
High Availability / Disaster Recovery.....	13
High Availability.....	13
Storage Level Disaster Recovery.....	14
Application Level Disaster Recovery.....	14
Conclusion.....	15

This white paper is for informational purposes only. SIMPLEMENT, INC. MAKES NO WARRANTIES, EXPRESS, IMPLIED, OR STATUTORY, AS TO THE INFORMATION IN THIS DOCUMENT. This white paper is an overview only, and not a technical manual. Only trained users should control replication.

Simplement, Inc. may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. The furnishing of this document does not give you any license to these patents, trademarks, copyrights, or other intellectual property.

Data Liberator is a trademark of Simplement, Inc. All other trademarks are property of their respective owners.

Copyright 2013 – 2014 Simplement, Inc. All rights reserved.

## OVERVIEW

### WHAT IS THE DATA LIBERATOR™ ?

The Simplement Data Liberator™ is the catalyst to help you achieve better decisions, faster, with your SAP and multi-source data.

The Data Liberator eliminates the usual process of ETL (extraction/transformation/loading) that has been the traditional method for sourcing SAP data for reporting environments outside of SAP. The Data Liberator also eliminates the need for ABAP development in all phases of the reporting architecture.

As the only solution to provide these benefits, the Data Liberator achieves these results by starting from the foundation of database replication from a SAP source database, and providing processing and transforms on-the-fly to automatically deliver easily used views in the target SQL Server 2012 Data Liberator relational database.

The primary output of the Data Liberator is an ODS (operational data store) that is:

- always available – even if the SAP system is down
- always current, as any changes in the underlying SAP database are streamed in real-time into the Data Liberator
- always free of the need for ABAP, since the Data Liberator transforms the data and makes it useable as a modern, SQL ANSI data source

The outputs of the Data Liberator are used in many ways: to enable real-time operational reporting and dashboards using SSRS and other tools, to automatically refresh standardized reports, to support ad-hoc BI, to load tabular databases and cubes, to populate lists in SharePoint and kick-off SharePoint workflows, to provide real-time data via web services and to feed real-time data to additional systems (e.g., shop floor manufacturing systems).

Because the Data Liberator leverages log-reading on the source database, Simplement customers have reported no discernible impact by the Data Liberator on their underlying SAP source system.

### WHAT THIS WHITE PAPER COVERS

This white paper provides a technical introduction to the following topics related to the Simplement Data Liberator for enterprises running SAP on SQL Server:

1. Solution Architecture & Typical Deployment Pattern
2. Installation and Configuration of the Simplement Data Liberator for SAP ECC, other SAP components, other non-SAP Oracle or DB2 sources and non-SAP SQL Server sources
3. Overview of the Data Liberator and Replication operating procedures that include:
  - a. Complete Data Refresh/Snapshot
  - b. Adding New Tables
  - c. Removing Existing Tables

- d. Re-Snapshotting Existing Tables that have changed
4. Overview of High Availability (HA) configuration options and procedures.
5. Overview of Disaster Recovery (DR) configuration options and procedures

## SOLUTION ARCHITECTURE & TYPICAL DEPLOYMENT APPROACH

### SOLUTION ARCHITECTURE

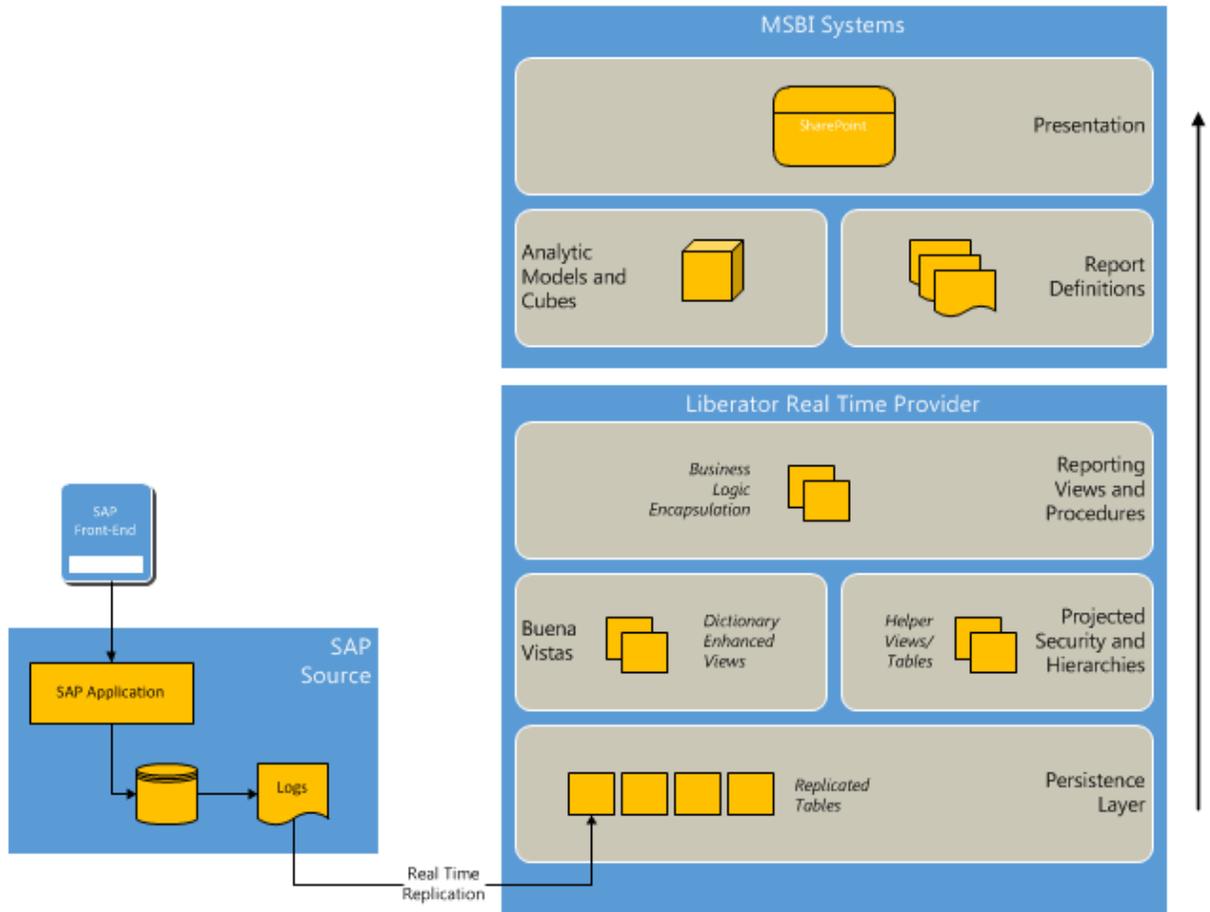
The Simplement Data Liberator works with one or more source SAP databases and uses log based replication to stream the data into a single target Operational Data Store (ODS) in SQL Server 2012 Enterprise. A Simplement Data Liberator installation typically refers to this target database as the Real Time Provider (RTP).

**Figure 1-1** below shows the flow of data from the SAP user in the lower left, through the SAP application layer, to an underlying database, the logs of which are read in real-time to populate a persistence layer in the target Real Time Provider (“RTP”) database at the lower right corner of the diagram. Because of this log-reading architecture, the Data Liberator never touches the SAP Application layer and has no discernable impact on the source SAP system. This architectural benefit contrasts with extractors that use the SAP application layer and can have a large negative impact on the SAP system.

Building on top of the RTP persistence layer, the Data Liberator automatically generates “Buena Vistas” which are easy to use views of the data collected in SAP. The Buena Vistas include English (or customer’s choice of language) column names, making it easy for developers to understand the data. In addition, the Buena Vistas identify primary keys to aid developers, make available true dates and true times to support date and time arithmetic in reporting outputs, and otherwise harmonize the peculiarities of the legacy SAP data types and structures to make the data readily usable as a SQL data source for a wide variety of business information and other systems that consume SQL data.

The Data Liberator also automatically declusters/depools data, instantiates SAP domains, creates easy to use meta data dictionaries, and automatically generates objects that allow developers to easily leverage existing SAP security and organizational hierarchy models in real time.

- See Figure 1-1 on next page -



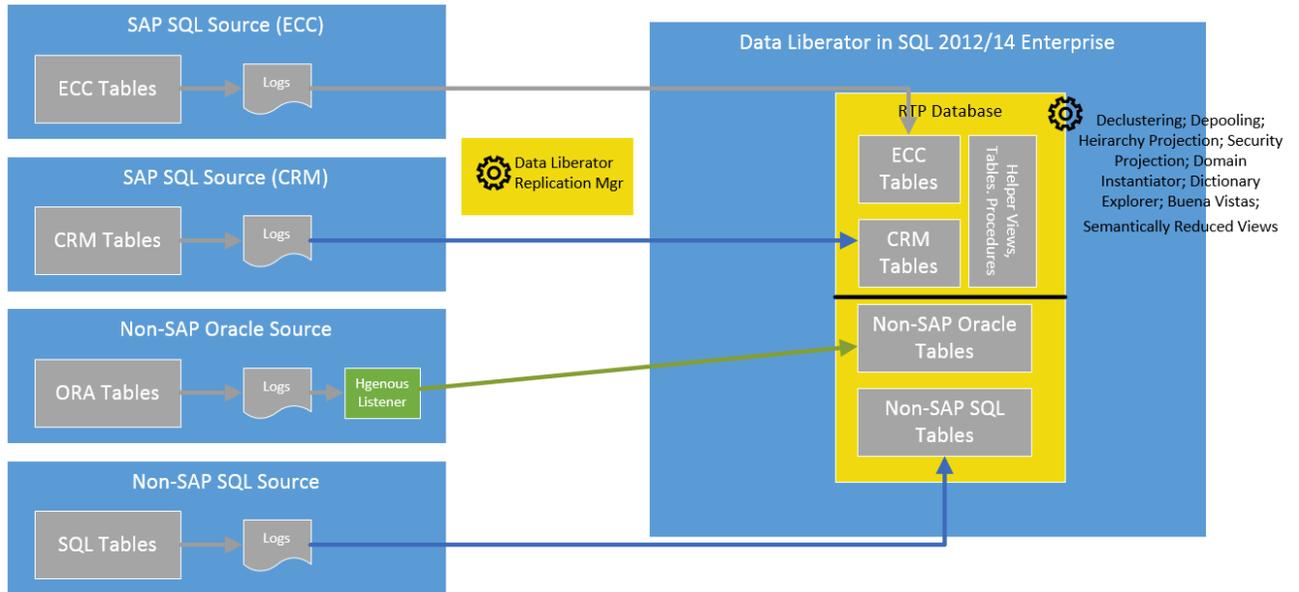
**FIGURE 1-1**

As a best practice, SAP and non-SAP data is best integrated **outside of SAP**. Rather than bring multi-source data into a SAP environment, the solution entails bringing data collected in SAP into a SQL Server environment where it can more easily be mashed up with non-SAP data for reporting and analytics.

**Figure 1-2** below shows a sample architecture consisting of two SAP component sources (ECC and CRM), both running on SQL Server, together with two non-SAP sources, one running on Oracle or DB2 and the other running on SQL Server. This shows the versatility of the solution in integrating a variety of data sources.

1. SAP ECC – Windows/SQL Server – Native SQL Server replication services will be used to replicate data into the Data Liberator SQL Server target.
2. SAP CRM – Windows/SQL Server – Native SQL Server replication services will be used to replicate data into Data Liberator SQL Server target.
3. Non-SAP Oracle/DB2 source – Oracle on Linux or Windows, or DB2 on select OS – an embedded third party replication tool is used to perform heterogeneous replication from Oracle/DB2 into the Data Liberator SQL Server target.
4. Non-SAP SQL source – Windows/SQL Server – Native SQL Server replication services will be used to replicate data into Data Liberator SQL Server target.

Log reading replication is done at the table level, and only the tables relevant for reporting are selected for replication. A typical number of SAP ECC tables for replication would be in the range of 400 – 500 tables. Replication from all systems is based on log reading in order to minimize impact on the source systems. Any additional processing of source data (such as creation of Buena Vistas, processing of cluster tables, etc. ) will be performed by Data Liberators processes on the target server.



**FIGURE 1-2**

## RTP FEATURES

1. Real-time data. The Data Liberators database is continuously updated, concurrently with the creation of database logs on the source system.
2. Customer chooses tables to replicate – including Z Tables and tables with Z Columns
3. Automatic generation of Buena Vistas that include business-friendly nomenclature.
4. Automatic generation of computed Date and Time columns to allow date calculations, etc.
5. Automatic Projection of SAP Hierarchies into developer-friendly data formats.
6. Automatic Projection of SAP security into developer-friendly data formats.
7. Automatic resolution of database collations between SAP source and target system.
8. Automatic de-clustering and de-pooling of cluster and pool tables.
9. Automatic instantiation of SAP domain values for reporting.
10. Automatic generation of meta data dictionaries for all SAP tables, including custom Z tables and Z columns.
11. No logon or processing on the source transactional systems. That means there is no impact on the source transactional system.

- 12. Leverages log-reading replication from the source database, meaning no discernible impact on source database performance.

## DEPLOYMENT PATTERNS

The typical Data Liberator deployment pattern involves two tiers. Production Data Liberator is installed in conjunction with the production SAP database(s). The Dev Data Liberator is installed in conjunction with the QA SAP database(s).

SAP ENVIRONMENT	DATA LIBERATOR ENVIRONMENT
Production	Production
QA	Dev

Note that the Dev Data Liberator in this typical pattern is NOT installed with Dev SAP database(s). The QA database(s) have more data than the Dev database(s), so QA makes a better source for a Dev reporting environment.

Some enterprises have deployed Data Liberator in three or four tiers. The three tier setup is:

SAP ENVIRONMENT	DATA LIBERATOR ENVIRONMENT
Production	Production
QA	QA
Dev	Dev

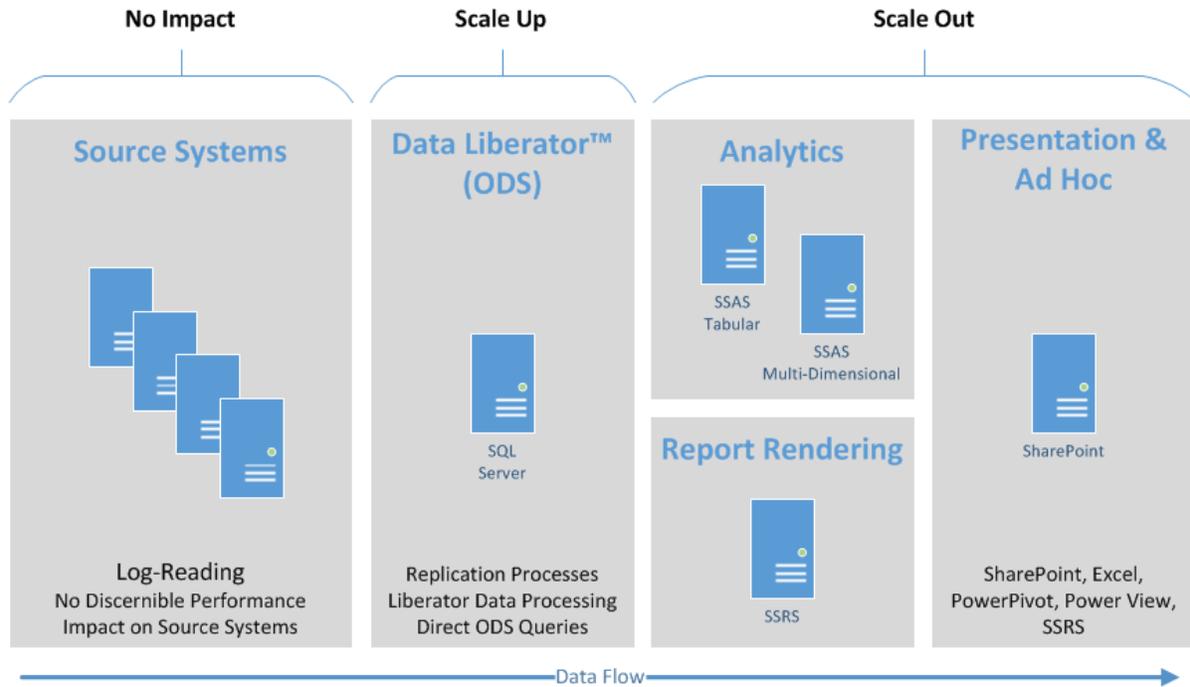
Some of these enterprises that begin with three tiers (the first installation being Dev) end up reducing their deployments to the typical two tiers described above after they have gained familiarity with the solution.

Optionally, some large enterprises install the Data Liberator in a fourth tier, Sandbox, however that is very unusual.

## SCALING

Recommended scaling of the overall solution is shown in **Figure 1-3** below. The Data Liberator ODS is scaled up (more CPU's and more memory in one server) while the analytic and reporting aspects of the solution are scaled out (more servers).

- See Figure 1-3 on next page -



**FIGURE 1-3**

## INSTALLATION/CONFIGURATION

### TARGET LIBERATOR - RTP DATABASE SERVER AND REPLICATION

The primary output of the Data Liberator is contained in a SQL Server instance that holds the Real Time Provider (RTP) database. As part of the native SQL Server replication services a small database named “distribution” will be installed on the Data Liberator server.

### SOURCE - SAP ECC AND OTHER SAP COMPONENTS ON SQL SERVER DATABASES

During the initial setup with assistance of Simplement (typically one week), the enterprise will provide the information in the format of the sample below with regard to each SAP system running on a SQL Server database:

<b>SERVER NAME (SQL INSTANCE)</b>	
<b>DATABASE</b>	
<b>PUBLISHER</b>	
<b>PUBLICATION</b>	
<b>SUBSCRIBER</b>	
<b>DESTINATION DB</b>	

SQL Server replication requires that all source Tables or Views have primary keys.

Then a number of steps are performed to configure native SQL Server replication on the designated source database(s). These steps are fully documented, and Knowledge Transfer sessions will

supplement documentation in helping the enterprise attain self-sufficiency in connection with ownership and operation of the system.

#### NON-SAP ORACLE SOURCE DATABASE

The typical enterprise has non-SAP systems of importance for reporting purposes. To use these systems as a source of data for integration with SAP data, the enterprise will begin with identification of the system and its characteristics in the format of the sample below:

<b>HOST</b>	
<b>DATABASE</b>	
<b>SCHEMA</b>	
<b>ORACLE NLS_CHARACTERSET</b>	
<b>ORACLE HOME</b>	
<b>ORACLE SID</b>	
<b>ORACLE SERVICE PORT</b>	

When these systems are running on an Oracle database, a number of steps are performed to configure the heterogeneous listener. A light weight listener will be installed in the Oracle source to read log changes for replication. Additionally, a heterogeneous replication Hub instance is installed and configured for creating and directing heterogeneous replication channels from Oracle and source systems.

#### NON-SAP DB2 SOURCE DATABASE

The typical enterprise has non-SAP systems of importance for reporting purposes. To use these systems as a source of data for integration with SAP data, the enterprise will begin with identification of the system and its characteristics in the format of the sample below:

<b>HOST</b>	
<b>INSTHOME</b>	
<b>DB2INSTANCE</b>	
<b>DATABASE</b>	
<b>SCHEMA</b>	
<b>Database Code set</b>	

When these systems are running on a DB2 database, a number of steps are performed to configure the heterogeneous listener. These configuration steps are analogous to the steps taken with respect to the non SAP Oracle database. Any differences in setup are to be documented during the setup and configuration exercise.

#### NON-SAP SQL SERVER SOURCE DATABASE

If an enterprise running SAP on SQL Server also has non-SAP systems of importance for reporting that are running on SQL Server, the integration of this non-SAP data is straightforward. Because the target Data Liberator database runs in SQL Server, homogeneous (rather than heterogeneous) replication is

available. Specifically, the solution utilizes native SQL Server replication services. The sample below shows a setup and identification grid used during installation:

<b>SERVER NAME (SQL INSTANCE)</b>	
<b>DATABASE</b>	
<b>PUBLISHER</b>	
<b>PUBLICATION</b>	
<b>SUBSCRIBER</b>	
<b>DESTINATION DB</b>	

SQL Server replication requires that all source Tables or Views have primary keys defined.

**NON-SAP SOURCES IN GENERAL**

The Data Liberator core functionality applies to SAP ECC and certain other SAP components. While replication strategies can be used to integrate numerous non-SAP data sources, each such source will need to be individually analyzed. For example, replication of a non-SAP source may include all referential constraints that are defined in the source database. While this provides some advantages in the ODS, it comes at the expense of making replication harder to manage. Simplement has core expertise in multi-source data integration and assists customers in evaluation and documentation of optimal approaches to specific multi-source data scenarios.

**SOLUTION OPERATION**

**COMPLETE DATA REFRESH / SNAPSHOT**

Before the Data Liberator can start replicating transactional data (changes or Delta) from the source system in the form of Database Insert/Update/Delete/Truncate SQL statements, it needs to have an exact copy of the data from the source database. This process is called Data Refresh generally, but sometimes it is referred to as Snapshot Operation in a SQL Server environment.

This is usually a one-time occurrence done during the installation process. However, there are a few cases where it needs to be done after installation such as:

- a. Periodically the QA environment is refreshed from Production.
- b. If replication is down for any reason for a prolonged period of time (usually more than 3-5 days, depending on the log retention policy). In this case there would be no transaction log available on the source database, and it would require a refresh of all tables that are in replication.

**ADD A TABLE TO REPLICATION**

The Data Liberator Replication Manager, shown in **Figure 2-1** below, is a software tool that is part of the Data Liberator platform. It completely automates the process of adding SAP tables to replication. The

Replication Manager is fully SAP-aware and automatically works with standard tables, custom Z tables, tables with Z columns, transparent tables, cluster tables and pool tables.

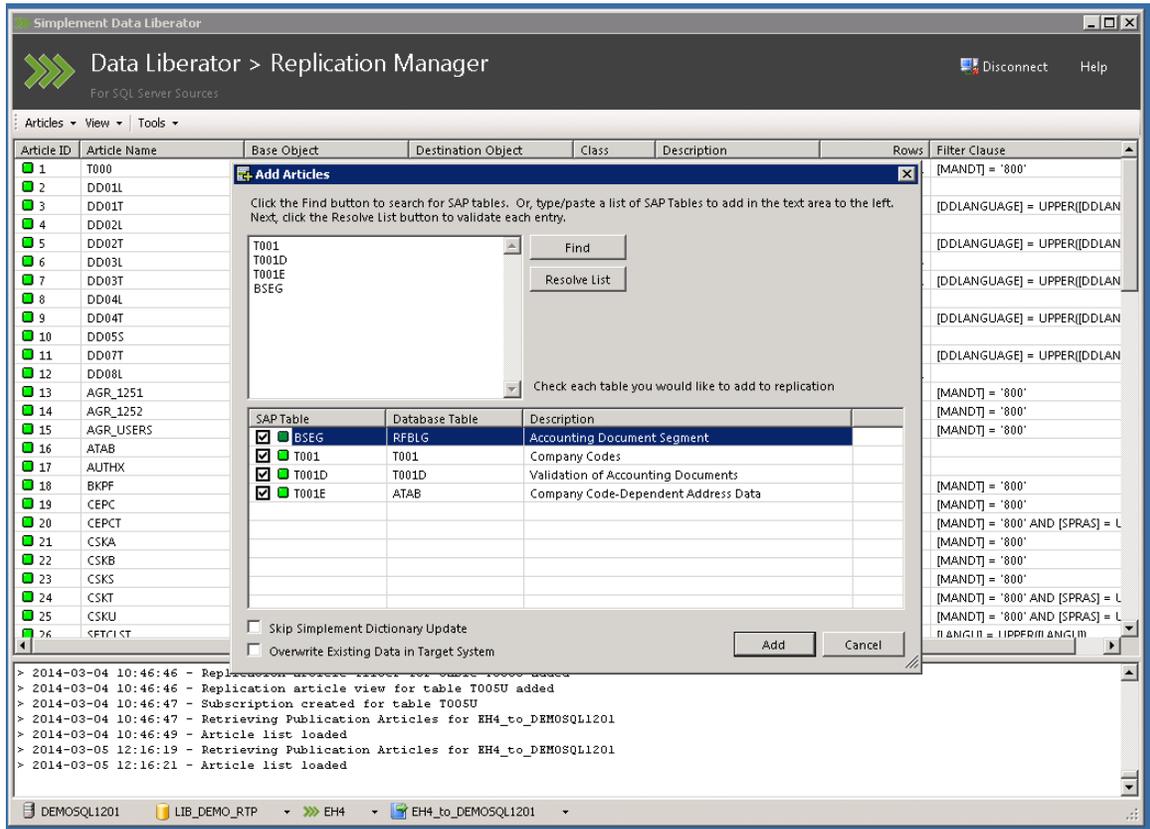


FIGURE 2-1

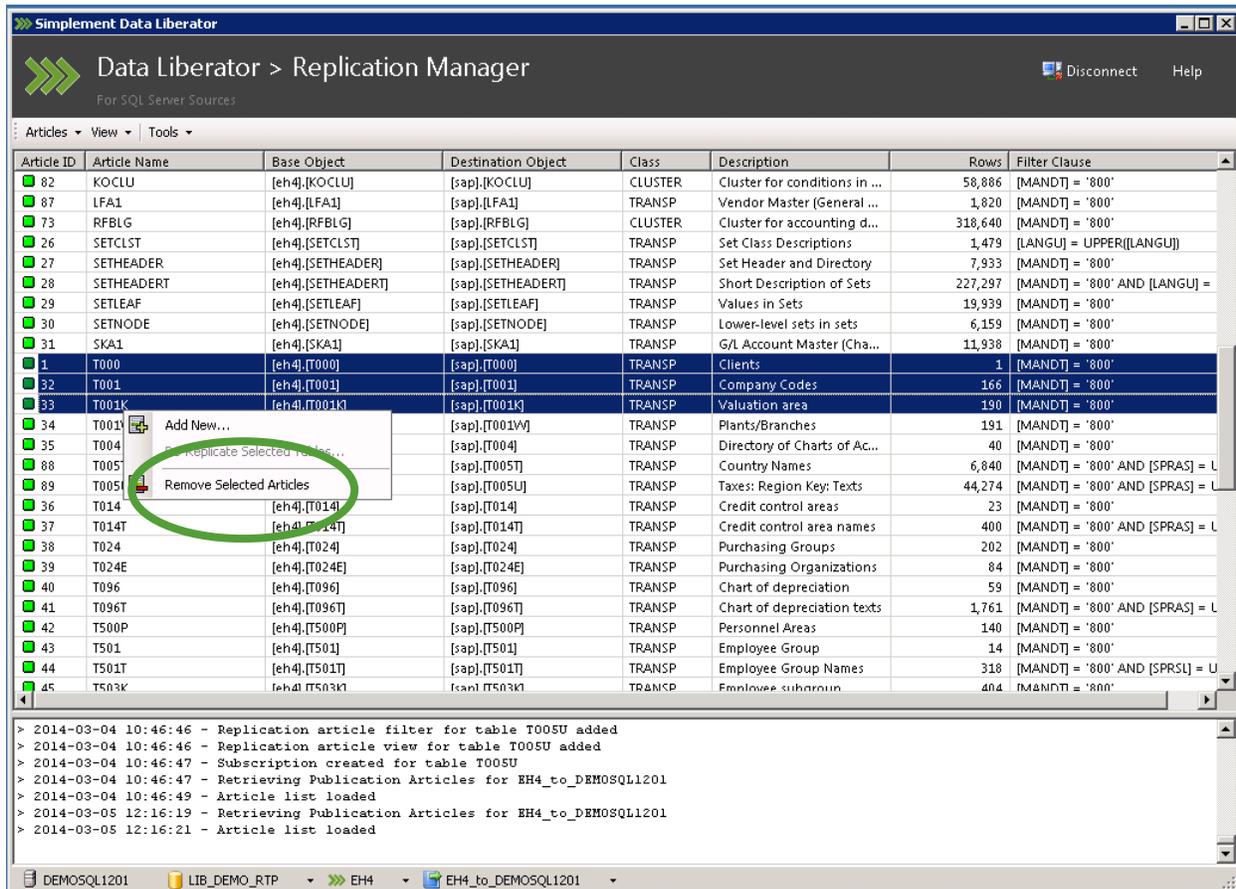
Most customers replicate a basic group of 200 to 400 tables during the initial week when the Data Liberator is first installed. Because the Replication Manager makes it very easy to add tables on demand, it's not necessary to spend a lot of time developing the list of tables for that first group. Additional tables can be easily added when needed. It is a best practice to build up the tables under replication in this fashion, rather than initially including hundreds or thousands of extra tables "just in case" they might be useful.

While the Replication Manager is not designed to work with non-SAP tables, the Data Liberator environment is an ideal location to integrate multi-source data with multiple replication feeds from a variety of SAP and non-SAP sources. Simplement provides knowledge transfer and consulting services specifically related to multi-source data integration.

Additional details and procedures related to use of the Replication Manager for adding SAP tables to replication are fully documented and provided by Simplement to the enterprise as part of the knowledge transfer exercise to allow the customer to achieve self-sufficiency.

**REMOVE A TABLE FROM REPLICATION**

The Replication Manager also completely automates the process of removing SAP tables from replication. For example, certain SAP tables may have been erroneously added into replication, or they may no longer be needed for reporting purposes. These unneeded tables are easily removed from the replication stream with a few mouse clicks using the “Remove Selected Articles” functionality as shown in **Figure 2-2** below depicting the Replication Manager.



**FIGURE 2-2**

Non-SAP tables don’t use the Replication Manager, and the specific steps for removing these non-SAP tables vary based on whether the source database is DB2, Oracle or SQL Server. If Simplement is engaged to provide knowledge transfer and consulting services for a customer scenario involving multi-source data integration, the specific steps will be identified and documented for customer self-sufficiency.

**PROCESS TO HANDLE CHANGES TO TABLES IN SAP**

Before your BASIS team releases transports that change the structure of SAP tables that are under replication, the affected tables need to be removed from replication and then added back using the graphical Simplement Replication Manager. This remove/add cycle only occurs for the affected tables, and does not require a re-snapshot of any other existing tables

## SOLUTION MONITORING

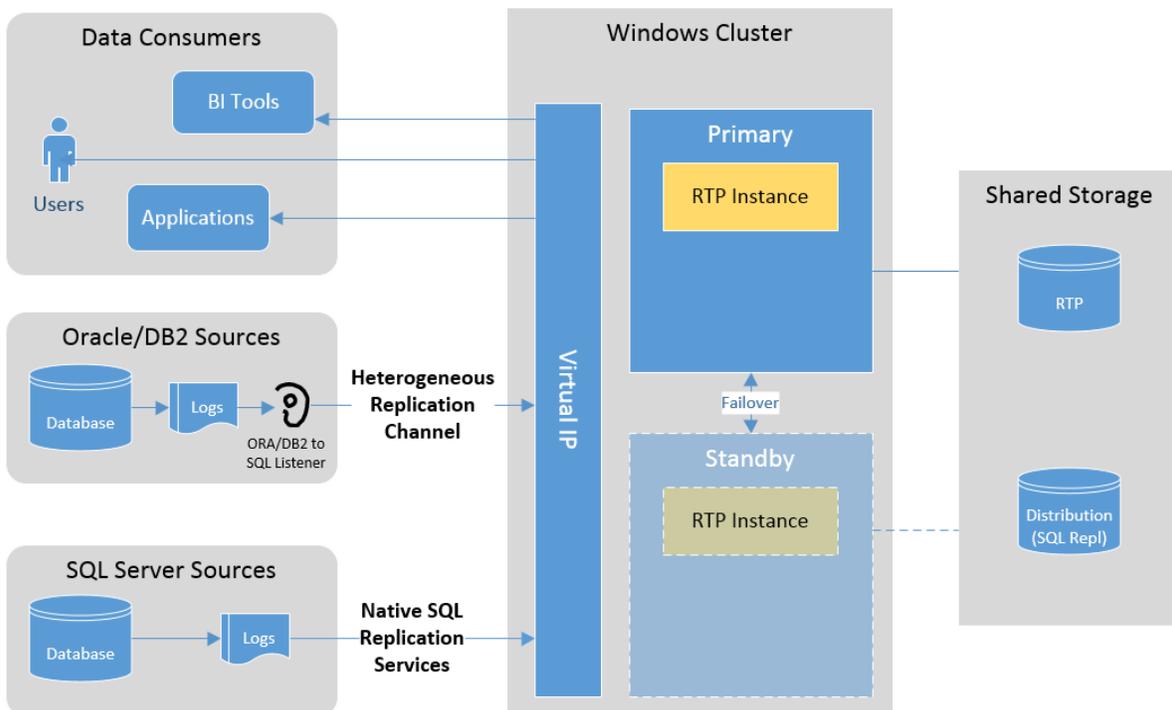
When SAP runs on SQL Server, the homogenous replication can be monitored with Microsoft tools. SQL Server replication is controlled through SQL Server Agent Jobs. Each of these jobs can be configured to provide notification through email.

## HIGH AVAILABILITY / DISASTER RECOVERY

### HIGH AVAILABILITY

There are many different potential strategies for making the Data Liberator replication highly-available. One recommended configuration uses windows clustering, as shown in **Figure 3-1** below. In a clustered environment, the SQL Server Instance that contains the RTP and homogenous replication processes can fail-over to standby hosts automatically. The cluster hosts use shared storage for the RTP (Real Time Provider) and Distribution (SQL Replication) databases as well as the homogenous router application and configuration.

Using this strategy, all replication sources and data consumers communicate with the virtual address of the cluster.



**FIGURE 3-1**

## STORAGE LEVEL DISASTER RECOVERY

SAN replication is one option for recovery that is used in rare circumstances. If you would like more information on SAN replication, please contact a Simplement representative.

## APPLICATION LEVEL DISASTER RECOVERY

This process involves rebuilding the server. Most customers prefer this option, as it is simple, cost effective, and saves time.

The Data Liberator system has three types of data from the point of system recovery, and each type is restored individually.

**a. Data Liberator Application**

Data Liberator Application components are installed from the standard stock scripts and installer programs. This also allows for change during the restoration, should any configurations (e.g. server name, client) need to be changed. The process is similar to a new installation of the Data Liberator.

**b. Transactional data**

Transactional data is a copy of the source transactional system, such as ECC, CRM, or a custom application. All data should be refreshed from the source system, so that the source and the Data Liberator are consistent.

**c. Reporting logic**

Reporting logic is additional logic built by the customer for developing reports in the form of Views, Configuration Tables, Stored Procedures, etc. The customer maintains a copy of the objects created in the database. This can be also be saved by putting the database in full recovery mode or by taking a script backup of the database objects after new changes are made.

*Intentionally Blank*

## CONCLUSION

For the enterprise running SAP on SQL Server, this white paper has provided an introductory overview of:

- Data Liberator Solution Architecture, technical features and deployment patterns
- Installation and configuration of the Simplement Data Liberator for SAP ECC and SAP CRM
- Replication operating procedures for adding tables, removing tables and re-snapshotting existing tables that have changed
- Configuration options and procedures for High Availability and Disaster Recovery.

Implementing the Data Liberator eliminates the cost, time and hassles of ETL. It gives the enterprise the ability for rapid and incremental BI development to meet changing business needs. By providing an enterprise grade method for self-service and Ad Hoc BI, it empowers business users to make better decisions, faster. At the same time, the IT department is freed of report development backlog, while maintaining IT control over governance and security. The Data Liberator allows the enterprise still running SAP Business Warehouse to develop a road map for the replacing SAP BW with a modern BI platform that is free of ABAP coding at all levels of the reporting solution.

*For questions on this document, please contact [info@simplement.us](mailto:info@simplement.us)*