Unlocking SAP Data with Log-Based CDC

This datasheet is about how HVR can help you move data from your SAP applications to your data warehouse or data lake, today, as well as in the future for when you upgrade your SAP installation(s) and start using the SAP HANA database.

A common option for extracting data is to use ABAP (Advanced Business Application Programming) to bulk extract the data, or select the data out using BAPIs (Business Application Programming Interfaces).

Log-Based CDC reads all data changes asynchronously from the database transaction log. The approach avoids direct impact on transactions in contrast to trigger-based capture. Log-Based CDC ensures inserts, updates and deletes are always captured regardless of any create or last modified timestamps on the table, and transaction boundaries are maintained. HVR is a real-time data integration solution that features Log-based CDC from SAP ERP and SAP HANA.

The SAP database can be any one of HVR’s supported source databases. Data can be delivered into any of HVR’s supported destinations including cloud-native data platforms like Snowflake, data streaming technologies like Kafka, or cloud file systems such as S3, Azure Data Lake Store, or Google Cloud Storage. In line with HVR’s overall CDC support, transaction boundaries are maintained during the data integration.

Complete list of supported platforms
How It Works

HVR runs in a distributed architecture, which includes a hub and agents on the source and target side. HVR is modular and a single installation of the software can capture from a source, play the role of the hub, and/or deliver into a target.

<table>
<thead>
<tr>
<th>Capture from SAP</th>
<th>The Hub</th>
<th>Deliver from SAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOURCE AGENT</td>
<td>ROUTING DATA</td>
<td>TARGET AGENT</td>
</tr>
</tbody>
</table>
| Typically an agent will run on, or close to the source SAP database servers (or a standby of the source database) for capture from an SAP application. Note that in order to minimize the load on the SAP system and database, the source capture process leaves the data for cluster and pool tables in the original encoded state. For SAP HANA, the agent must run on the HANA database server for real-time capture, or only read backups of the log on another system. | The hub routes the data to one or more destinations. By default, no transformations take place on the hub and the data remains compressed as it is stored on the file system to manage recovery in case of any failures. The HVR administrator always connects to the hub to:  
  - Define and store connection details for the endpoints for the data integration.  
  - Create and manage so-called channels, the data flows between sources and targets.  
  - Initialize the data integration, and create the jobs to perform one-time data loads and compare runs.  
  - Monitor the application to get current state of the jobs, as well as to review the log files.  
  - Access rich data movement statistics to obtain insights in the data flows. | An HVR agent on the destination side, close to the destination, receives compressed data. It then:  
  - Decompresses (and if needed decrypts) the data.  
  - Decodes data for any cluster and pool tables through HVR’s SAP Transform. Note the SAP transform is available for a Windows or a Linux OS.  
  - Applies the data to the target in line with the rules the hub provided using optimized, platform specific data integration strategies. Metadata to support HVR is stored in one of HVR’s supported databases. |