DATASHEET

Apache Kafka

2020

HVR is a Confluent Certified Connector for Kafka

HVR provides native, high performance data delivery into Kafka and integration with schema registry.

Apache Kafka (“Kafka” for short) is a distributed streaming platform that is used to:

1. Distribute data between systems or applications
2. Build real-time applications that respond to data events

Kafka uses concepts like publish and subscribe similar to the Enterprise Service Bus (ESB), but is different in that it does not implement business logic or protocol translations. Unlike the ESB Kafka stores a history of messages for replay purposes. Many well-known organizations including Microsoft (LinkedIn), Netflix and Uber, but also non-technology companies like Target and CapitalOne use Kafka extensively.

Why HVR for Kafka

The HVR technology delivers end-to-end data integration capabilities to set up and manage real-time log-based change data capture and continuous data delivery with Kafka as a target. Starting with tables in an existing database or application, HVR supports:

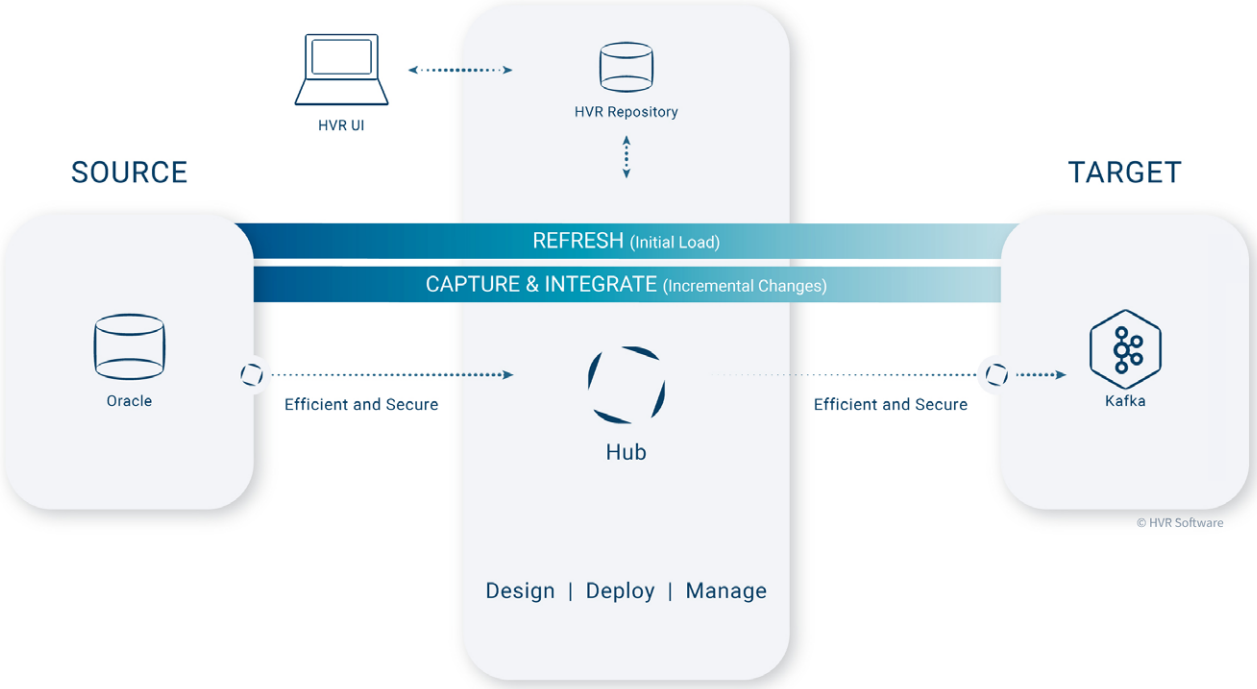
- Populating the schema registry.
- Flexible options to deliver data in JSON, Avro and other formats.
- One-time load into topics, referred to as refresh, integrated with continuous log-based Change Data Capture (CDC). This capability simplifies populating a topic with an initial data set.
- Non-intrusive, log-based, transactional CDC on many supported source technologies with optimized continuous delivery into Kafka.
- Transactionality can be propagated through manifests.
- A graphical management console to setup the data flows and any automated alerts, also giving access to rich data movement insights.

HVR Architecture

HVR provides a flexible, modular architecture that uses agents to gather change data most efficiently in a distributed setup. An agent, which is an additional installation of software on or near the source or target (and is optional), facilitates the movement of changes that occur between data end points. Communication between agents uses compression and other optimizations for maximum performance.

Agents are highly beneficial in setups that connect over a Wide Area Network (WAN) (e.g. between on-premises and the cloud, or between cloud-based systems running in different availability zones or hosted by different providers.) SSL/TLS encryption is available for agent communication, and authentication using custom certificates is optional.

A data integration setup always includes a central HVR installation, referred to as the hub, that controls CDC and routes data delivery. Use the Graphical User Interface to define data movement, set up automated alerts, and to gain data movement insights based on the rich statistics HVR gathers as data moves into Kafka.



Supported Source Technologies

With wide support for real-time, log-based CDC from relational database technologies, HVR can play in an important role in feeding Kafka. The following are relational databasesin which HVR can feed Kafka from using log-based CDC:

- **Oracle**, starting with version 9.2 up to current, supporting virtually all combinations of setups between storage on file systems, in ASM or even using raw devices, clustered or not, with support for advanced options like encryption and audit vault. HVR also supports capture from a standby or archived log only environment. Capture is supported directly or through SQL and includes cloud-based services like Amazon's Relational Database Service (RDS) and the Oracle Cloud.
- **SQL Server**, starting with version 2005 up to current. Log-based CDC can be set up locally or remotely through SQL calls, supporting capture from a standby node in an AlwaysOn configuration, or CDC by solely accessing transaction log backups.
- **SAP HANA** versions 1 and higher, supporting CDC through direct access to the transaction logs, or with additional latency by accessing just backups of the log, possibly on a separate server.
- **DB2 LUW** (Linux, Unix, Windows) starting with version 9.7.

- **DB2 iSeries** (AS400) starting with version 7.1.
- **PostgreSQL** 9.4 and higher, either by directly accessing the log files or through SQL statements.
- Log-based CDC is also available on **Amazon PostgreSQL RDS.**
- **Ingres** 9.1 and higher.
- **MariaDB** 10.0 and higher.
- **MySQL** 5.6 and higher, including support for MySQL RDS, and Amazon's MySQL compliant Aurora database.

With SAP ECC commonly in use by large organizations, HVR also features a decoder for cluster and pool tables to facilitate log-based CDC from all tables in older ECC deployments.

HVR also supports file capture from a variety of file systems including cloud-based data stores like AWS S3 and Azure Data Lake Store. Files can either be propagated as a separate object, or parsed into separate records when the file format is delimited e.g. CSV.

Beyond Kafka

Many organizations aim to deploy Kafka for all means of data sharing. Such a goal is ambitious and cannot always be achieved or achieved immediately. HVR supports a wide variety of destination technologies with a highly efficient and scalable capture once, deliver many times routing mechanism.

Refer to the HVR website for a current overview of supported technologies.

ABOUT HVR

We accelerate data movement so that you can revolutionize your business. HVR is designed to move large volumes of data FAST and efficiently in hybrid environments for real-time updates.