



SOLUTION BRIEF

**A PRACTICAL APPROACH TO REAL-TIME REPLICATION
FOR ANALYTICS**

GROWING DEMAND FOR REAL-TIME INSIGHT

In today's world of real-time ride sharing, traffic monitoring, weather forecasting, inventory control, and dynamic pricing, it's not surprising that expectations are rising for more current business information. The acceptability of batch-oriented daily data warehouse reports is rapidly diminishing. Businesses and government entities are increasingly relying on real time analytics to optimize customer service, marketing campaigns, logistics, inventory, cash flow, human capital and public safety. In financial services and other industries, analytics are being embedded into business processes to reduce fraud and operational risk.

The following examples are illustrative of the wide range of applications that cross industries and geographies.

"The impact of integrating real-time analytics with business operations is immediately apparent to business people because it changes the way they do their jobs..."

*Jim Sinur,
Research Vice President, Gartner*

Lufthansa delivers real-time flight planning services to a great number of national and international airlines around the world. The system populates hundreds of databases distributed across diverse geographical locations with massive volumes of continuously changing data pertaining to routes, flight paths, air traffic, flight restrictions, weather conditions and more. Airline operations teams and pilots receive updated, in-flight information needed to adjust schedules, routes, and flight patterns.

The U.S. Coast Guard has deployed a real-time data warehouse that continuously replicates data from its Aviation Logistics and Maintenance Information System. Detailed information relating to parts, material, and maintenance schedules for 211 aircraft is used by repair and logistics personnel to support search/rescue, law enforcement, environmental response, ice operations, and air interdiction operations.

Aspen Marketing Services, a division of Epsilon, is a leading provider of response and engagement marketing solutions. The company leverages real-time analytics to target the right message to the right customer at the right time for measurable impact.

NEW CHALLENGES

Although use cases for real-time analytics have expanded in recent years, early applications of "active data warehousing" date back to the 1990s. However, several emerging trends are making it more challenging than ever to implement real-time analytic solutions.

More sources and targets – Most of the early real-time analytic architectures featured a single data warehouse that was fed by one or two transactional source systems. Modern architectures often feature hundreds of sources and targets, on-premise and in the cloud, distributed across multiple geographic locations.

More diverse data types and schemas – Traditional analytic systems are built using relational databases with star or snowflake schemas, and primarily leverage data from relational databases and flat files. Although implementing these solutions can be very challenging, the emergence of file systems such as HDFS (part of Hadoop), columnar databases, NoSQL databases, and the "Internet of Things" has added yet another layer of complexity.

Greater data volumes – It's no secret that the era of big data has arrived. Organizations of every size, in every industry, continuously struggle to deliver acceptable system performance while controlling costs, maintaining security, and managing risk.

ALTERNATIVE APPROACHES

There are many different ways that real-time analytics can be achieved, but most approaches can be grouped within three basic architectures:

Federated data warehouses federate queries across multiple source systems utilizing a logical data model. The concept is appealing because it theoretically eliminates the need to physically move and integrate data. Unfortunately, the promise of the federated data warehouse, which has been in existence since the 1990s, has not been realized on a widespread basis. Organizations that attempt to implement them usually discover that query performance is unacceptably slow. Queries can be frozen if one of the sources is unavailable, or if data joins are too complex. At the same time, queries can impede performance on transactional systems as well. Furthermore, because source systems are often modified or replaced, ongoing maintenance can be extremely difficult.

Event and trigger-based replication entails continuously updating a data warehouse or operational data store in real time. Event triggers that initiate a replication action, such as updates, deletions or creation of new records, are stored in the source system databases. This approach comes with a major drawback. All trigger-based systems, including native database trigger services or third party tools, impose significant overhead that slows down transactional database processing. Third party tools, needed for heterogeneous trigger-based replication, add an additional layer of complexity and performance degradation to transactional systems.

Log-based replication provides a way of delivering real-time analytics without the performance problems and complexity issues associated with other approaches. Rather than relying on intrusive event-based triggering services that must continuously run on top of each transactional source system, log-based replication tools leverage standard log files that are natively generated by all common databases.

HVR REAL-TIME REPLICATION FOR ANALYTICS – PRAGMATIC, EFFECTIVE AND AFFORDABLE

HVR is one of the few technology companies focused on problems relating to high-speed data replication. Our products have been built ground up exclusively to solve this problem around the following values.

Flexible architecture - Real-time replication technology is applied to a wide variety of use cases that extend beyond analytics and business intelligence. Examples include hybrid cloud computing, geographic data distribution, data and application migrations, and high availability. HVR's real-time replication solution is designed to serve any combination of use cases while adhering to core design principles.

Easy to use, fast to deploy – In contrast to most log-based replication solutions that rely on multiple tools and technologies to perform functions such as capturing log files, generating target DDL, and initiating data load, HVR delivers all functionality required for real-time analytics in a single, unified product that is based on a hub and spoke architecture. A centralized hub that communicates with agents running on sources and targets acts as the command center for building, managing and maintaining data integration flows.

Installation and set-up can be completed in a matter of minutes. Unlike point-to-point solutions, HVR places no limits on data flow complexity or number of sources, targets, and mappings.

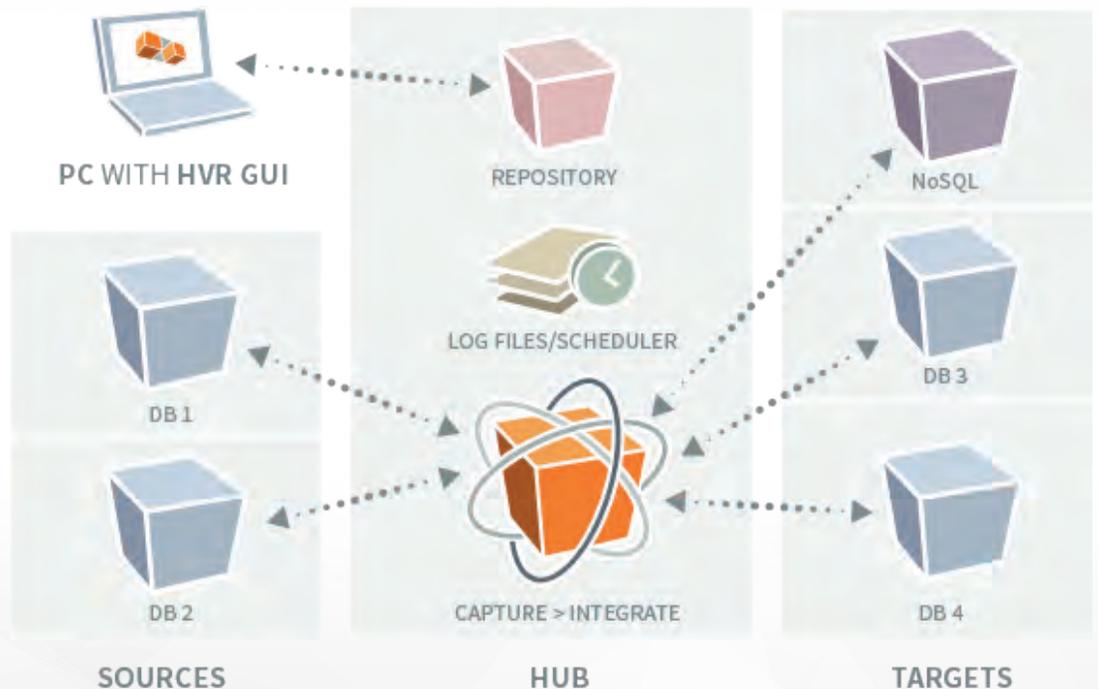
It is easy to implement and maintain, delivers comprehensive functionality with unlimited scalability, and provides exceptionally low total cost of ownership.

HVR provides access to all capabilities through an easy-to-use graphical user interface. Designers and administrators can get started quickly without the steep learning curve associated with non-intuitive command line interfaces found in most replication tools.

HVR simplifies data replication by co-existing with sources, targets, and network topologies in a completely non-invasive manner. It captures log files generated by source system databases, can leverage native database loaders, works in conjunction with third party ETL tools and system management consoles, and runs on standard network protocols. Any installation of HVR can act as a hub, or as an agent, with minimal overhead.

“We have been using HVR now more than ten years for our flight planning system business...With the combination of flexibility, performance and robustness, HVR has proven to be a very good choice to embed in our flight planning system.”

*Rudi Koffer,
Senior Database Software Architect,
Lufthansa Systems Airlines Operations
Solutions Division*



HVR Software High Speed Data Replicator Architecture

Complete solution – HVR provides the flexibility needed to suit any replication use case, regardless of its complexity.

HVR is designed to operate in heterogeneous environments, with support for all major relational databases, analytic databases, log files, and HDFS. A sub-set of HVR capabilities commonly used for real-time analytics include:

- Automatic mapping between source and destination data types, including automatic mapping of character sets and code pages.
- Initial load capabilities on live systems to ensure systems are in sync once real-time replication catches up to current.
- Transformations including table and column name changes and derived values at column level.
- Built-in transformations to support soft deletes and last updated columns for slowly changing dimensions.
- Compare functionality including detailed reporting and SQL generation in case systems are out-of-sync.

- Bi-directional or multi-directional replication, including conflict detection and resolution.

A rich set of management and administration capabilities make operations and trouble-shooting easy and inexpensive. The system can even be configured for lights-out operation.

“HVR has proven to be very robust and reliable... We are all amazed by the performance of the HVR...It performed near flawlessly from the very beginning.”

*Steve Peck,
Technical Expert, US Coast Guard*

Scalable high-performance – Organizations that deploy real-time analytics environments need to be sure these systems can scale without compromising performance as data volumes grow exponentially and the number of sources and targets expand. The HVR architecture incorporates numerous features that collectively deliver unmatched performance with no limits on data volumes.

In contrast to other replication tools, HVR routes data directly to its destination(s) through a constant process flow that does not require staging or intermediate steps. Multiple jobs can be run in parallel and network transport efficiency is maximized by intelligent data packaging and advanced compression.

At the destination, HVR uses optimizations to ensure scalability for large data volumes irrespective of the type of target database:

- Combining transactions: multiple small transactions become fewer large transactions.
- Coalescing changes: multiple subsequent changes to the same row are combined into a single row change.
- Burst mode: changes are staged, coalesced and then frequently applied using bulk SQL statements to the actual target table(s) modifying multiple rows at the time.

HVR Software US

135 Main Street, Suite 850
San Francisco CA 94105
USA

+1 (415) 655-6361

HVR Software EMEA

Daalwijkdreef 47
1103 AD Amsterdam
The Netherlands

+31 20 820 1570

HVR Software APAC

23 Adams Street
French Forest NSW 2086
Australia

+61 4040 81962

HVR Software China

2/F ShanghaiMart
2299 Yan An Road West,
Changning District Shanghai
Shanghai, 20036
People's Republic of China

+ 86 150 0066 1981

GETTING STARTED

If you're an IT professional interested in learning more about how real-time replication technology can be applied in your analytic environment, HVR can help. You can request a live demo and consultation with an HVR representative to learn how the software works, how it's used in organizations like yours, and how it can potentially be used to deliver real time analytics to your organization.

ABOUT HVR

At HVR, we believe it should be easy to deliver large volumes of data efficiently, reliably and at the right time into your data store of choice. Our software, the HVR High Volume Replicator, does exactly this using real-time data capture between data sources including SQL databases, Hadoop, data warehousing and business intelligence data stores as well as the most commonly used file systems.

For those organizations where real-time data replication is a mission critical process, HVR has been proven to be a reliable, secure and scalable solution by some of the largest global companies and leading government and defense organizations. HVR Software is a privately held company with offices in North America, Europe and Asia Pacific.

For more information, please contact us at info@hvr-software.com