



Six Key Considerations

When Choosing a Cloud Data Integration Solution



Contents

Introduction	3
<hr/>	
FACTOR ONE: Freedom of Choice	4
<hr/>	
FACTOR TWO: A Modern Solution	6
<hr/>	
FACTOR THREE: Multi-Cloud	8
<hr/>	
FACTOR FOUR: Is All Functionality Included?	10
<hr/>	
FACTOR FIVE: Downtime	12
<hr/>	
FACTOR SIX: Security	14
<hr/>	
Conclusion	16
<hr/>	

Introduction

Adopting cloud-based solutions as part of your data strategy might have been forward-thinking even just a few years ago, but today, they're essential components. With the proliferation of cloud adoption, organizations have an increasing need for replication and integration solutions that can deliver data to the cloud and between clouds. And delivering the data alone is not enough—it must be on time, accurate, and fresh for analysis. While many organizations have adopted Change Data Capture (CDC) solutions, not all of these solutions address the complex needs of data-driven businesses, including expanding data volumes and types as well as the growing suite of data management platforms. In this eBook, we examine six crucial considerations to aid in the evaluation of data replication solutions.

1. Does the solution provide the freedom to choose the data sources and targets that make sense for your organization?
2. Does it support modern target platforms?
3. Will the solution enable a multi-cloud strategy that simplifies integration and data availability?
4. Does it include all the functionality necessary to perform a data integration job?
5. What downtime is required for maintenance?
6. How is the data routed from on-premise environments to the cloud?



This guide was written by HVR CTO, Mark Van de Wiel, and his team of data integration experts. Mark worked at Oracle for eight years and was part of the team at Oracle GoldenGate. He and the team at HVR have an extensive background in working with the major data integration and replication tools.

FACTOR ONE

Freedom of Choice

Does the solution provide the freedom to choose the data sources and targets that make sense for your organization?



Freedom of Choice

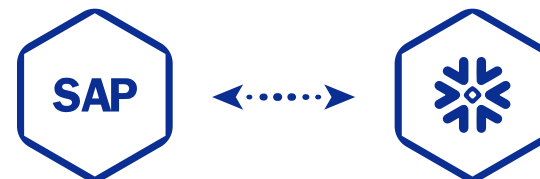
Does the solution provide the freedom to choose the data sources and targets that make sense for your organization?

For relational database transaction processing, it is rare for the same database technology to be used across departments within an organization. For legacy reasons, or as the result of a merger or acquisition, your organization might use multiple database technologies. If your goal is to standardize on a single technology for log-based Change Data Capture (CDC) with the goal being real-time integration across these technologies, you may want to consider your entire data management infrastructure and look for a solution that supports all of the database technologies within your organization.

Solutions that help you avoid “vendor lock-in” and instead support a variety of sources and targets within the Big Data ecosystem will allow you the freedom to appropriately scale your business with the technology that you determine is best. HVR is an independent solution built to be flexible in supporting support common and emerging platforms.

When evaluating your current replication solution, it is recommended to inquire about the depth of data integration capabilities that are provided for the platforms supported. An example is how HVR supports change data capture from SAP.

HVR supports data decompression and decoding for cluster and pool tables through the SAP transform capability. HVR also integrates with the SAP dictionaries to understand cluster and pool tables, including any custom Z columns in the table definition. If/when customers migrate their SAP ECC to S4/HANA, they can continue using log-based CDC.





FACTOR TWO

A Modern Solution

Does the solution support modern target platforms?



A Modern Solution

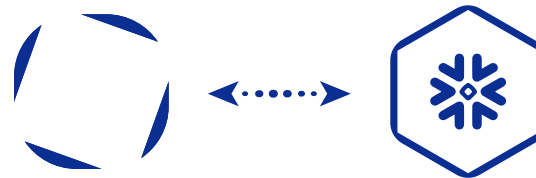
Does the solution support modern target platforms?

In recent years, more organizations are building data lakes to consolidate data in its raw, original state, continuously, and with minimal latency. Data replication technologies are suited to deliver data from a variety of sources into the data lake. While some organizations build their data lake using a traditional relational database such as Oracle, it is more common for this use case to be implemented using a distributed file system such as Hadoop Distributed File System (HDFS), AWS S3, Azure Data Lake Store, or Google Cloud Storage. When evaluating a solution for data lake consolidation, ask if your replication and integration solution requires a separate license. HVR does not require an additional license.

Snowflake Support

For a database with relational properties, a cloud-native technology such as Snowflake is often used. HVR natively supports Snowflake as a database destination and is able to keep a copy of the source application tables in sync, use soft deletes on the tables, or create an audit trail of changes to the data.

If a data lake is being used by your organization, make sure the data replication technology you choose supports your organization's technologies and budget considerations for the foreseeable future.

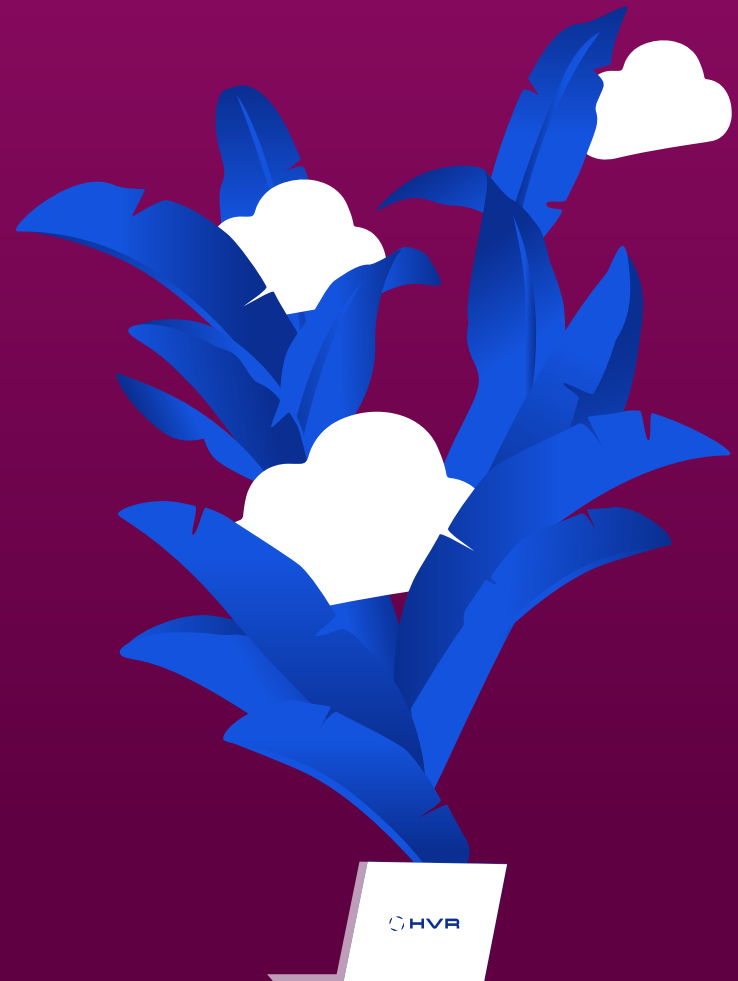




FACTOR THREE

Multi-Cloud

Will the solution enable a multi-cloud strategy that simplifies integration and data availability?

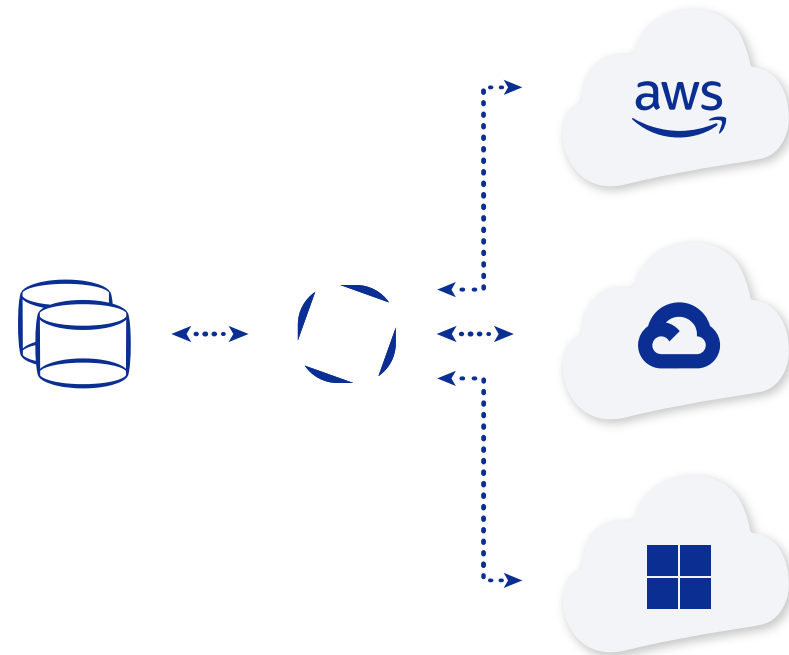


Multi-Cloud

Will the solution enable a multi-cloud strategy that simplifies integration and data availability?

An increasing number of both large and small organizations are moving to the cloud. Chances are your organization has plans to adopt the cloud or is currently adopting the cloud. The top three cloud choices include (in order) Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Services (GCS). Oracle Cloud Infrastructure is also one of the many other options. Large organizations often use multiple cloud vendors, but even small organizations can end up with a multi-cloud environment, and you may need to capture from or integrate into databases or data-centric technologies across clouds. Very few, if any, organizations will use only a single cloud provider.

Further, many organizations are adopting a hybrid cloud strategy, as well as a multi-cloud strategy, which indicates the use of multiple cloud providers. Companies are increasingly using large external streaming data sets on their own as sources of significant insight, but they are also quickly realizing the tremendous value in integrating these with their existing operational systems and databases in hybrid cloud environments. As these hybrid and multi-cloud strategies become increasingly common, your replication solution should enable data to move freely between multiple cloud-based providers and on-premises servers.





FACTOR FOUR

Is All Functionality Included?

Does the solution include all of the functionality necessary to perform a data integration job, or are multiple licenses and installations needed?



Is All Functionality Included?

Does the solution include all of the functionality necessary to perform a data integration job?

Typically, having the ability to set up data replication between one source and target is easy to implement, as is monitoring of the setup. This is true even if metrics need to be independently gathered from both the source and destination. However, you might want access to tools and capabilities that simplify management of the setup as you introduce additional data replication and/or complexity, such as scenarios that involve active/active replication, replication from many sources into one target, data distribution from one source to many destinations, or capture once and deliver multiple times. A graphical user interface (GUI), graphical monitoring, or automated alerts help to simplify management of the setup.

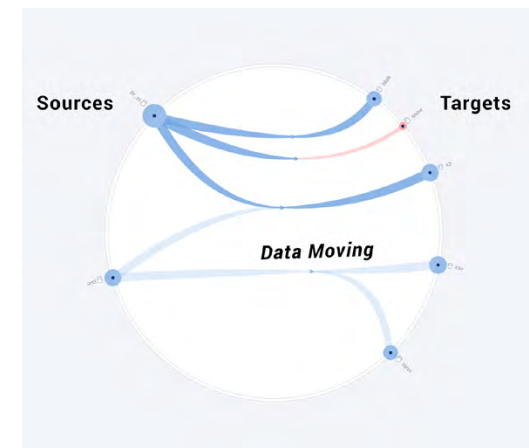
HVR includes all of these capabilities in a single offering because, if you need a transformation, you need to consider that transformation for all the aspects of data integration. Consider the following example of a soft delete transformation—when rows are physically deleted in the source, they are marked as deleted:

- Tables must be created with an extra column to mark the row as deleted.
- During initial load, every row must unset the “is deleted” indicator column.
- For CDC and continuous integration, rows that are deleted get updated with the “is deleted” indicator set.
- When comparing data, the rows with the “is deleted” indicator set must be excluded.

By defining transformations once in a channel, all HVR functions take advantage of the transformations. With Oracle, given that individual tools are used for the different integration steps, the same logic must be defined multiple times.

Data validation is another important consideration for continuous data integration, especially in a heterogeneous environment. Data accuracy not only provides data integration experts with peace of mind but also builds trust with data consumers. Data analysts or anyone consuming the data can trust that the data in the data lake or data warehouse destination correctly represents the data sources.

Additionally, HVR provides a topology chart that gives insights into the data replication flows, all data replication via a single overview, and the status, with access to detailed statistics on the data flows. The information is stored in a relational database table for easy integration with other monitoring solutions, and alerts can be integrated through SNMP, even into Oracle Enterprise Manager.

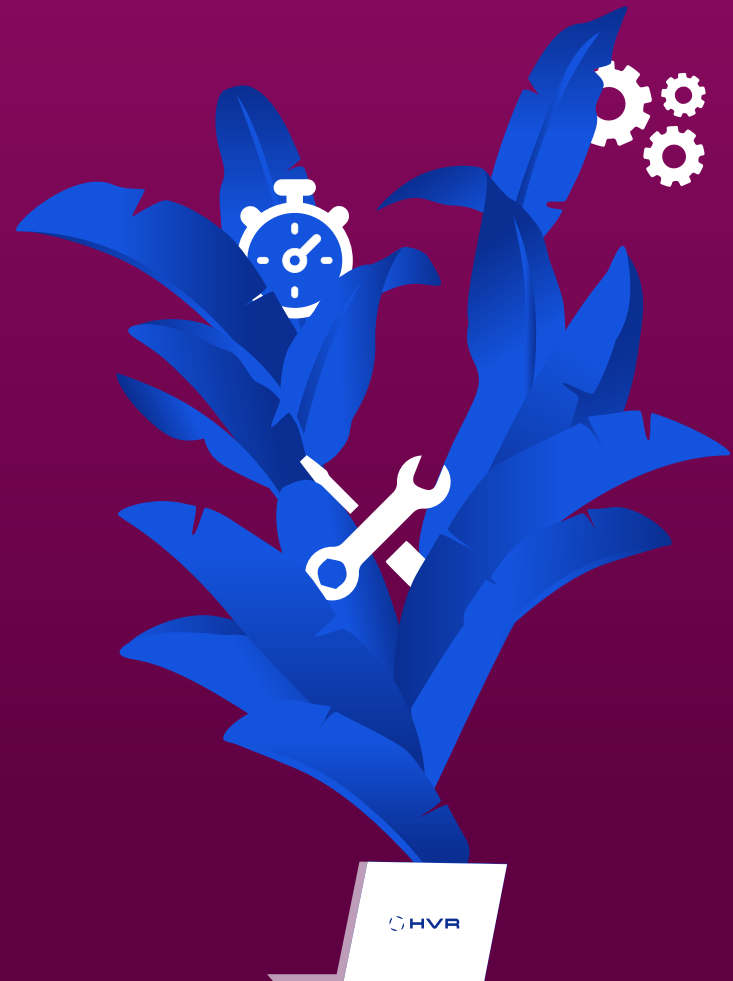


HVR topology chart UI

FACTOR FIVE

Downtime

What downtime is required for maintenance?



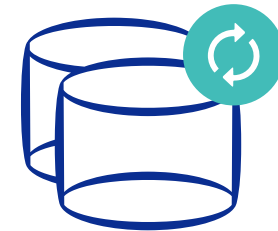


Downtime

What downtime is required for maintenance?

Complex software requires management. Software updates will be required at some point—to upgrade to a newer version for added features or to implement a fix.

HVR is different in that it runs outside of the database. While an upgrade requires data replication downtime, with HVR no downtime is required on the transaction processing database. This is because every HVR upgrade simply overwrites the previous installation. The complete software installation is less than 150 MB in size—including documentation and on-line help—so the upgrade can be performed in less than a minute. As a result, even though processes have to restart and maybe re-process a few minutes of work, the downtime for replication is low, and there is no downtime on the transaction processing database.

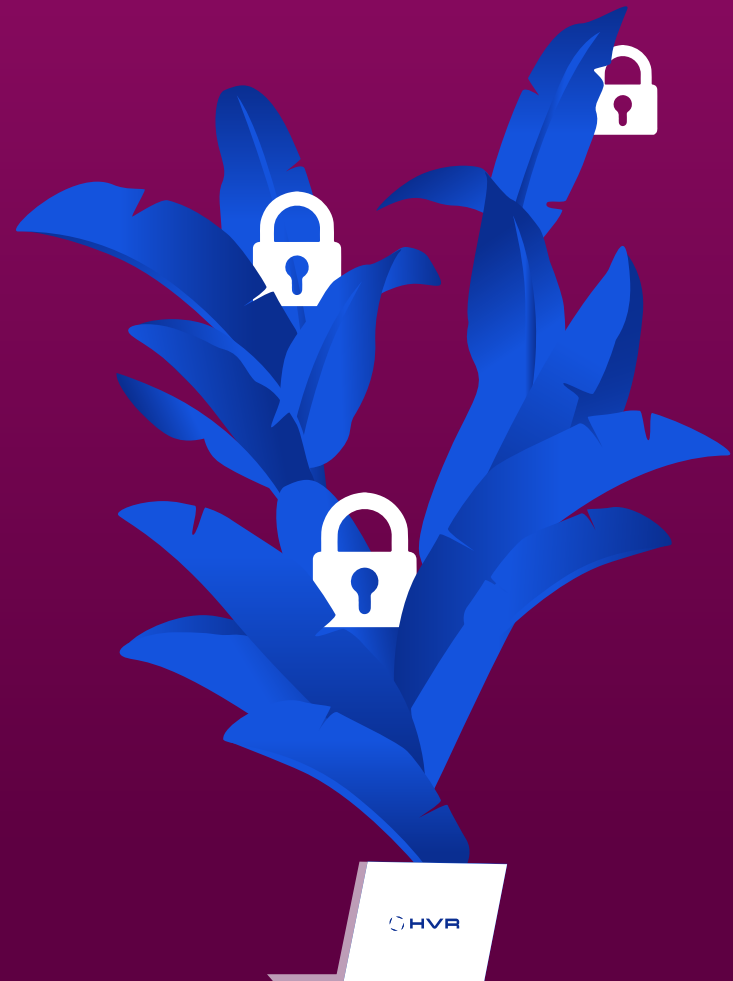




FACTOR SIX

Security

How is the data routed from on-prem environments to the cloud?



Security

How is the data routed from on-prem environments to the cloud?

Any time data is moving within an organization, and, especially when it crosses the corporate firewall, security is a top concern. In the spectrum of data integration problems, failure to ensure data security is damaging and costly to an organization. When integrating data between systems, it is critical to mitigate potential vulnerabilities.

HVR provides a rich set of security capabilities for enterprise deployments:

- SSL using AES256 encryption for data transfer over the wire.
- Authentication to agents through OS username/password, LDAP, or a plugin as well as the availability of certificate-based authentication for extra security and to avoid a “man-in-the-middle” attack.
- Requirement for only the hub to the agent to be opened because the hub always initiates the connection to the target and the hub can be installed anywhere in the data infrastructure.

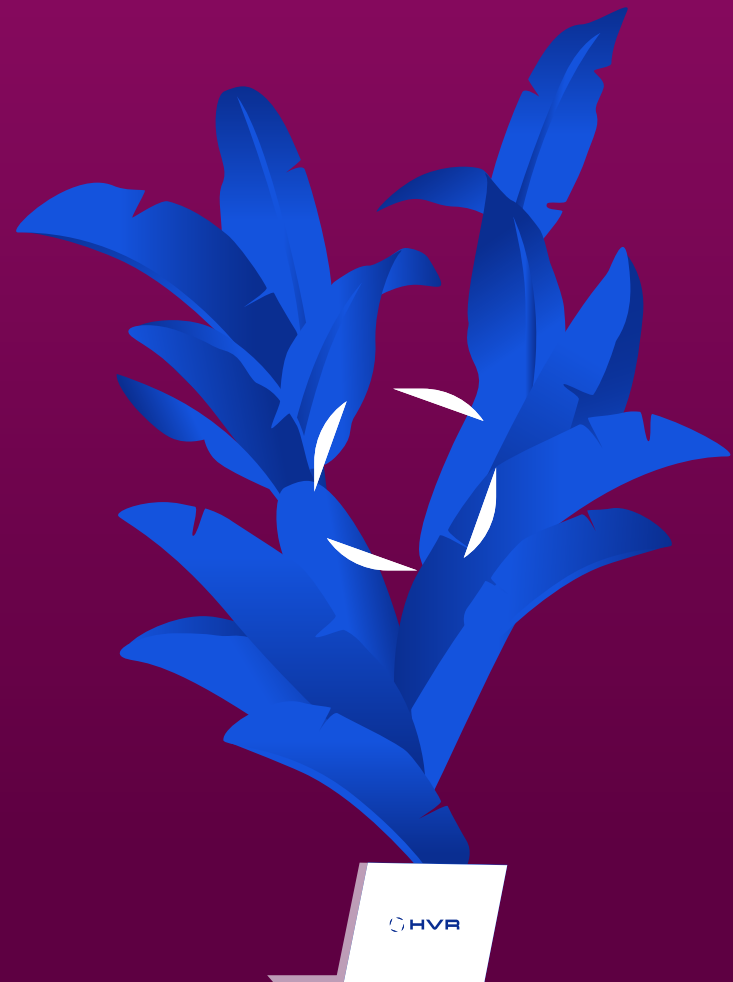
- This means customers will often choose to deploy the hub on-premises, rather than in the cloud, to avoid opening the firewall into the on-premises data center.
- Support for a proxy, which is another installation that acts as both the router and the gatekeeper for connections between the hub and agent end point.

A proxy is commonly used in a De-Militarized Zone (DMZ) of the data center to avoid opening the firewall to individual agents directly or to support multiple end points in the same data center through a single host.



Conclusion

Even if your replication solution checks all of these boxes, make sure it meets your day-to-day needs.



Even if your replication solutions checks all of these boxes, make sure it meets your day-to-day needs.

Choosing a data replication provider is an important decision—one that should consider both current and future organizational needs. At HVR, our single technology focus continues to be data replication, and we strive to ensure successful data replication implementations. From the CEO down, everyone at HVR is here to help. With HVR, you get direct access to a technical team of highly knowledgeable, expert engineers with a rich background in data management. Our ultra-responsive and accessible support staff means that support tickets are responded to in minutes.

Everyone at HVR is dedicated to understanding your challenges and helping you achieve your goals.

Our mission is to be your favorite replication solution. This means from the tool you use to the support you receive—from implementation to post-production—you feel that you made the right choice for you, your organization, for today, and for the future.

READY FOR THE NEXT STEP?

See the difference in our replication solution by taking it for a Test Drive.

[Test Drive HVR](#)



hvr-software.com