HVR

THE CLOUD IS NOW

OVERCOMING CLOUD INTEGRATION CHALLENGES

www.hvr-software.com

info@hvr-software.com | +1 415.489.3427 | @hvr_software

Contents

HVR

TRENDS IN CLOUD ADOPTION

Trends in cloud adoption

In 2018, over 300 attendees were surveyed at data industry events such as AWS, Kafka, and PASS Summit. When asked about their organization's current primary IT focus, over a quarter chose "Cloud Migration" as their response.

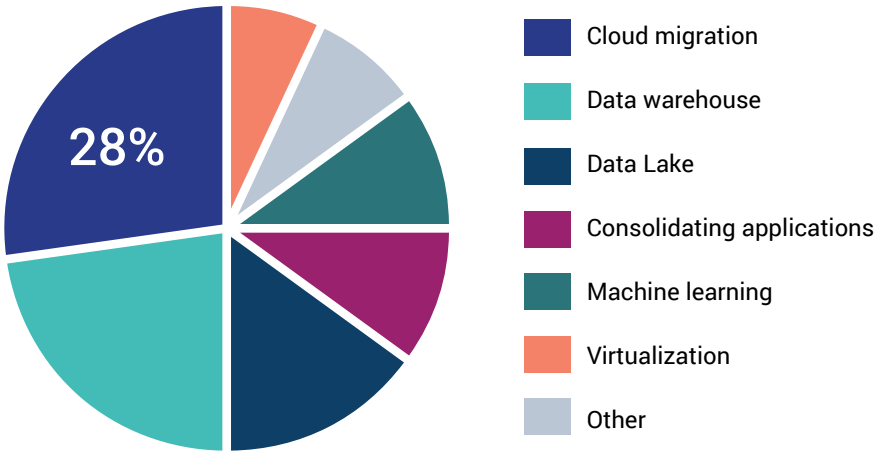
This is not surprising. Organizations large and small are embracing the cloud. Whether we're talking about infrastructure, middleware, applications or business process services, **Gartner reports that by 2021 approximately 28% of all IT spending will be for cloud-based services.**

In this e-book, we address common data integration challenges faced with cloud adoption, specifically, hybrid cloud adoption. With the right tools, common challenges can be mitigated.

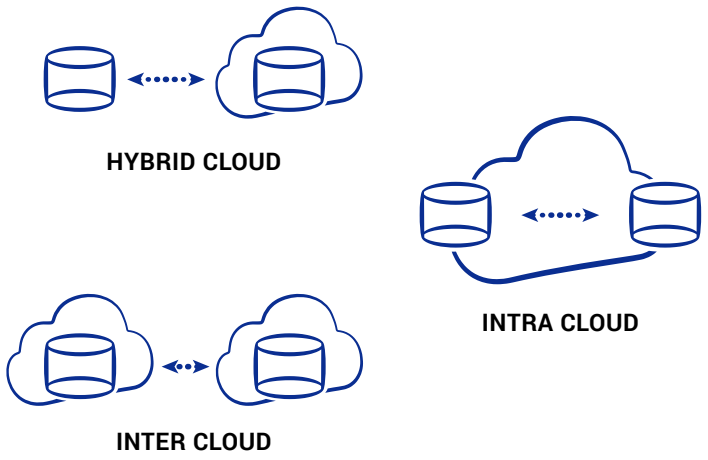
The phrase *Hybrid Cloud Computing* is used to describe the co-existence of multiple environments, at least one of which is a cloud environment.

Read more about how and why organizations are adopting the cloud.

What is your primary IT focus?



HVR cloud integration examples:



Hybrid cloud indicates the co-existence of on-premises and cloud servers and services as well as the use of multiple cloud providers, sometimes also referred to as multi-cloud. Because compute resources are not available at all levels, at least not cost-effectively, some of the largest environments will retain on-premises resources, at least for the time being. Most organizations will adopt a hybrid strategy, even if only during the initial migration when on-premises and cloud-based applications will run side-by-side to minimize disruptions.

Beyond initial setup, very few, if any, organizations use only a single cloud provider. For example, an organization may run some applications in AWS cloud and others in Microsoft's Azure cloud. Or, more likely, multiple applications are consumed as Software as a Service (SaaS), e.g., Salesforce for CRM, ServiceNOW for customer support, and SuccessFactors for HR, while home-grown applications are run in a Platform or Infrastructure as a Service (PaaS or IaaS) environment like AWS, Azure, or Google Cloud.

Further, many companies are using more external data sets for completely new forms of analytics. These large, external, streaming data sets are a source of insights on their own, but companies are quickly realizing the tremendous value in integrating these with their existing operational systems and databases in hybrid cloud environments. As these hybrid strategies become increasingly common, incorporating them into your current business practices takes careful consideration.

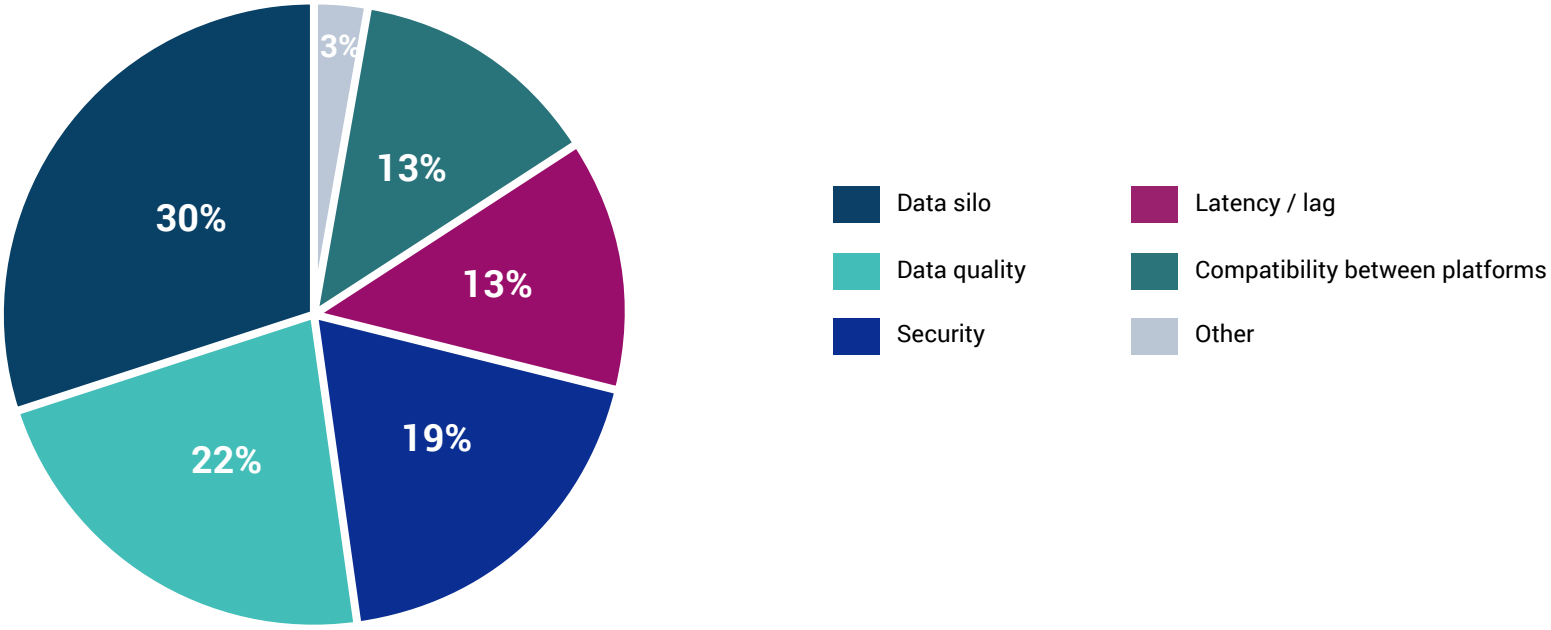
In the next section, you'll find some best practices to consider during migration or when adopting a hybrid strategy.

But, first a quiz:

Which of the following is your organization's biggest data integration challenge?

- Compatibility between data platforms
- Data quality
- Security
- Data silos
- Latency / lag
- Other

What is your biggest data challenge?



Over 30% of respondents indicated that ***data silos*** were their biggest challenge, with ***data quality*** and ***security*** as the close second and third responses, respectively.

How did your answer compare?

Read on to see how these challenges can be addressed.

HVR

DATA SILOS

Data silos

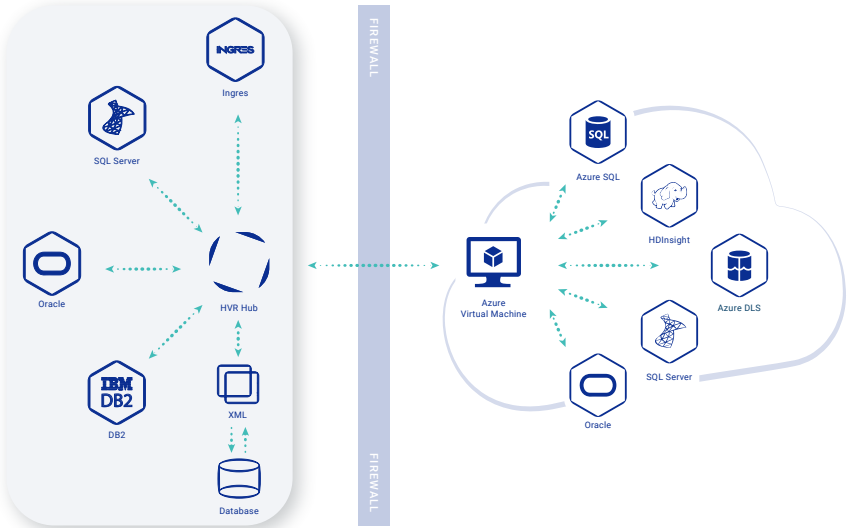
Why is this a concern?

The push toward data-driven decisions requires us to integrate new data faster and know our data better. This compels us to tear down the walls of siloed data. Think of the many possible applications that different departments in your organization use, from CRM to expense management to HR. Some of these may be cloud-based, while others are managed in-house. Or, your company may have data sets from different customers that need to be unlocked/integrated. Data from manufacturing plants, sales organizations, financial systems, and customer support structures are all typically managed independently. Despite this, your company's goals are supposed to transcend them all, bringing data from different business units and multiple databases into one location so you can access and use the data to make informed and effective analyses to support your organization's success.

When sitting in silos, the full power of the data cannot be realized.

What should you consider?

As we've already discussed, companies are often working with large amounts of data from multiple sources. Continuous, real-time data integration breaks down silos by consolidating large volumes of data from many sources into one location. In this way, adopting a hybrid cloud solution allows for more informed business intelligence with the freshest, most accurate data. For example a large, multi-divisional, global manufacturer uses such a strategy to consolidate financial data from all ERP systems across all divisions into a single cloud-based data lake, to facilitate cross-divisional analyses and quick, consolidated financial reporting.



HVR

DATA QUALITY

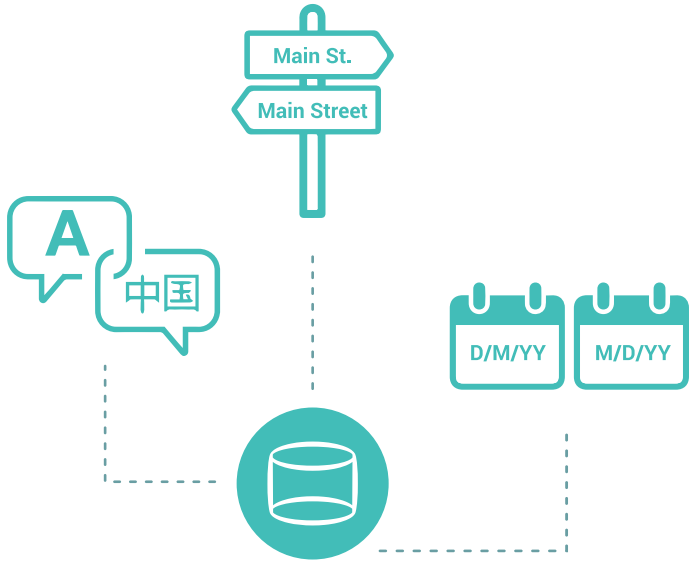
Data quality

Why is this a concern?

In a recent poll conducted at multiple major industry events, nearly a quarter of respondents whose primary IT focus was cloud migration listed data quality as their top concern. And, for good reason – data quality is an essential component of reliable, effective business intelligence. As you move between cloud and on-premises environments or between different cloud providers, inconsistencies in different data types, structures, and formatting of the data can affect quality. Quality can also be compromised as the data becomes stale. Further, the scalable nature of cloud-based environments means that it can support a large volume of data, but as that volume increases, the chance of inconsistencies in the data also increases.

What should you consider?

It is important to address data anomalies prior to database migrations, and a computer-driven analysis is the best way to do that. A data validation feature that tells you whether your data is in sync, alerts you if there's a problem, and can manage differences automatically can ensure quality data. By adopting a real-time integration solution, you can guarantee access to the freshest possible data.



HVR

DATA SECURITY

Data security

Why is this a concern?

According to a report by the data protection company, Bitglass, 90% of IT and IT security practitioners are moderately to **very concerned about data security in the cloud**. In the spectrum of cloud integration problems, failure to ensure adequate data security is potentially the most damaging and costly to an organization. When integrating data between systems within the cloud or, particularly in hybrid cloud environments where data is transmitted between cloud-based and on-premises systems, it is critical to address potential vulnerabilities that fall outside of the scope of the cloud or on-premises security provisions.

Vendor and software choices might be the most important considerations when adopting a hybrid cloud strategy.

What should you consider?

Although this may seem obvious, it cannot be overstated that all data in transit must be encrypted, and SSL encryption using AES256 is recommended. If using remote database connections, use the native encryption provided by most database connection drivers. If your data integration solution includes the deployment of local agents acting as proxy to communicate with the database, ensure that communications between the database and the agent are encrypted.

Another necessity is to limit exposure to production systems through the firewall when moving data in hybrid cloud environments. Although the firewall will need to be opened, you should consider whether it must be opened in both directions. Look for data integration solutions that allow you to open only a single machine and port pair in a single direction. Also, lock down the firewall wherever possible to a single server, ideally in a DMZ or perimeter network. Finally, deploy an agent on the server to act as proxy, handling communications with on-premises systems. By following these measures, you can ensure that your data is secure.

HVR

LATENCY / LAG

Latency / lag

Why is this a concern?

Reporting on near-real-time data is a must in today's business landscape, as companies need to make day-to-day tactical decisions to stay competitive in their industries. Cloud-based environments are built to be available 24x7, and users (as well as customers) have become accustomed to instant access to information. These combined factors are driving organizations to look for solutions that limit latency and offer continuous data integration solutions. This requires capturing data changes and propagating them from source to target, but this process can increase the load on the source system, causing latency in operational systems.

An ideal way to ensure maximized performance is to use a Log-based Change Data Capture (CDC) approach.

What should you consider?

Optimizing data transfer between the source and the cloud is essential to maximizing performance. An ideal way to do this is to employ a Change Data Capture (CDC) approach, which only transfers the changes, rather than the full data set, between environments.

Log-based CDC is generally considered the superior of CDC approaches. It provides a way to deliver the data needed to power real-time analytics without the performance problems associated with traditional approaches. The log-based method reads asynchronously from the database transaction log, requiring no additional table updates or query processing. Therefore, it has less impact on the database by placing no additional load on the source system transactions. Further, this approach can be applied to all possible scenarios, including systems with extremely high transaction volumes.

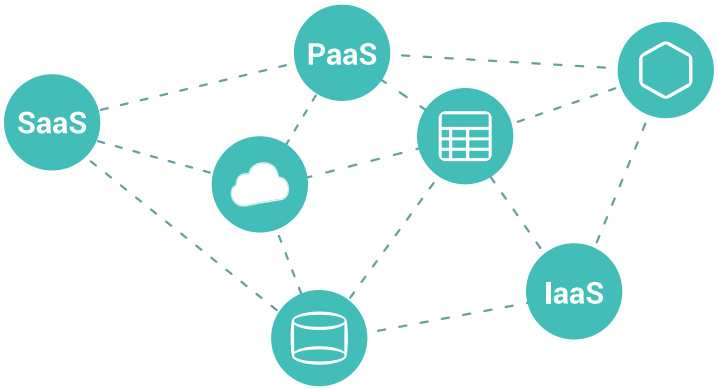
HVR

COMPATIBILITY BETWEEN DATA PLATFORMS

Compatibility between data platforms

Why is this a concern?

Most hybrid cloud deployments are mixed environments with a range of heterogeneous on-premises systems and diverse cloud services, requiring the platforms to be compatible. Any data integration solution must work in the initial environment as well as support changes to the environment over time. For example, an application running on a relational database today, deployed in the cloud or using a database as a service, may one day be replaced by a SaaS solution. Therefore, a hybrid data integration or migration solution needs to be flexible.



What should you consider?

Due to both its technology and inherently scalable nature, a hybrid cloud solution can provide this flexibility, but make sure to choose a data integration solution that ensures compatibility across the broadest possible array of databases, file systems, applications, platforms, and cloud services, including IaaS, PaaS, and SaaS. Such a system will deliver a wide range of deployment choices and offer the flexibility to make changes in a heterogeneous environment. Since the cloud provides many options to store or process data, the ability to quickly add destinations to an existing data integration flow is an added bonus.

Now that we have addressed some of the common concerns, your vendor and software choices might be the most important considerations when adopting a hybrid cloud strategy.

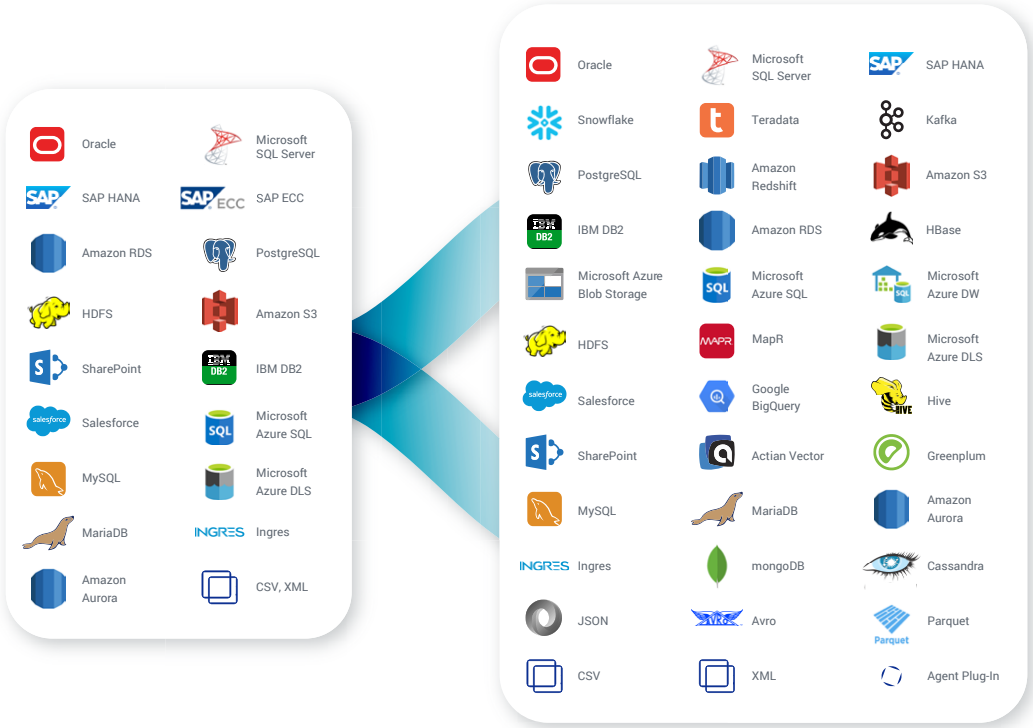
HVR

NINE CONSIDERATIONS FOR CHOOSING A CLOUD DATA INTEGRATION VENDOR

Nine considerations for choosing a cloud data integration vendor

1. Does the software support a wide array of data endpoints?

Organizations rarely run all their applications on a single data platform. Many run several different applications on different databases and distributed file systems in a hybrid cloud setup. If you want a single tool to support your needs, then that tool should replicate data in a hybrid cloud environment as well as between a wide range of databases, data lakes, and data streaming platform types.



Don't see your platform? Contact us to learn more about our API agent plug-in.

2. Does the software work with all the major cloud vendors?

As we've discussed, organizations using hybrid cloud environments often do so with multiple cloud vendors. Some organizations are reluctant to put all their applications and infrastructure eggs in one provider's basket. Others may strategically place specific workloads where they make the most sense.

For example, they may place general computing, relational databases and file storage in AWS; business applications and SQL Server in Microsoft Azure; and artificial intelligence (AI) in Google Cloud. The data integration solution should work consistently across all major cloud vendors for the sake of simplicity. You can learn one tool for all use cases and avoid vendor lock-in.

3. Is the software easy to set up and use?

Look for a solution that allows you to manage your entire set of data migrations between on-premises and multiple cloud vendors with a single tool on a single console with a single look and feel. This tool should provide comprehensive migration capabilities to manage data table creation, initial load, synchronization, monitoring, data validation, testing and deployment. It should also minimize the number of clicks necessary for each operation.

4. Can the solution handle the volume of changes in the hybrid cloud setup?

Many organizations need a solution that can handle multi-terabyte databases with heavy volumes of changes occurring continuously across the hybrid environment – changes that are happening in the on-premises and cloud databases. The solution should also be able to handle the increased spikes of activity your application requires.

5. How does it enable security?

As mentioned, look for a solution that minimizes your data security footprint. A solution with one server and one port acting as a proxy server on your on-premises system that talks over an encrypted connection to only one server and one port on your cloud solution will keep everything locked down behind the firewall.

6. Does it provide the flexibility to do phased migrations?

Usually, when you migrate to the cloud, you're moving an application and the database that supports it. Transferring the application can take considerable time because you must move the entire stack, including web servers, application servers, load balancers, and so on. Yet it's the database transfer that causes the most downtime for users. This downtime occurs because traditionally data needed to be offline when moved. The move also needed to be completed in full before the application could be made available to users and updated again.

By using a phased technique, however, you can minimize interruption by migrating the application stack first, followed by the database. This way, your application will only be offline for a few days or even hours during the database migration. The overall migration, which might take months, should be online throughout the migration process.

Another way to minimize downtime is to move your applications and users to the cloud gradually in a phased migration, rather than all at once. Look for a tool that enables multi-tenant migrations that provide this level of flexibility as well as bi-directional data movement so that any shared data in a multi-tenant stack can be kept in sync.

7. Can the solution instantiate the target database and perform automatic datatype mapping?

When transitioning to a hybrid cloud system, you may need to migrate some of your data into a new cloud platform. This requires you to instantiate (i.e., create) the target database structures, including user accounts, tables, indexes, constraints, stored procedure structures and database objects, and then load the data.

The right solution will automate the creation of the target database structures, automatically map the datatypes and load these data into the new tables. The solution should also accelerate these processes by efficiently performing operations in parallel while compressing the data on the fly to squeeze every bit of performance out of your existing bandwidth.

8. Does the solution transfer data reliably in a hybrid environment?

The solution should perform an initial bulk load of the data into tables on the target database. Simultaneously, it should capture all changes occurring on the on-premises source during the migration. When the initial load is complete, the system should automatically update the cloud target with the changes. When all these operations are complete, the solution should compare the on-premises source with the cloud target to validate that they are in sync.

Should something go wrong during the migration or the initial phases of use, the system should also be able to handle a failover of the source or target and perform data validation and repair if needed. In the worst-case scenario, the migration may need to failback to the original source without losing any data. To ensure that this failback occurs properly, the system needs to replicate data from the cloud to the on-premises system from the very first transaction. If it doesn't, you can never migrate back to the old system because the data won't be in sync.

9. Does it do real-time and historical monitoring and alerting?

How do you know if your replication systems are up and running and operating properly? A hybrid integration solution that can track real-time stats and send alerts will automatically tell you if stats reach pre-set thresholds. For example, if latency goes past 5 minutes, the solution might send an email, SNMP, Slack, or SNS message to inform you.

A solution that can integrate directly out of the box or be extended to integrate with third-party systems such as **ServiceNow**, **Atlassian**, and others will streamline service ticket automation, allowing you to more quickly resolve any issues.

You also want to track current and historical performance statistics regarding your hybrid cloud migration efforts, data throughput volumes, and latency, so that you can visualize and gain deeper insights into the nature of your changed data.

HVR

CONCLUSION

Conclusion

As your organization strategically integrates on-premises data centers and the cloud, you must continue to deliver the same high performance your users have come to expect. To do this, important consideration must be given to what cloud strategy you'll employ, how you'll get there, and the vendors and software you choose to complete it all. When planning your tactical execution, it pays to keep your long-term goals in mind.

HVR is a unique hybrid cloud data integration solution that provides you the flexibility, performance, and control you need in your cloud environments. Here at HVR, we handle complex cloud data integration challenges with one simple and efficient product. Our secure and reliable HVR distributed architecture supports a variety of topologies and gives you the flexibility to design your environment as well as the initial setup.

Contact us today for more information or to request a demo.

Contact us

Live demo

hvr-software.com