Cloud Data Warehousing

Benefits and options for a modern cloud data warehousing strategy

Cloud Data Warehouses: What are the benefits?

Cloud based architecture and applications are in recent times becoming the preferred choice of organisations of varying sizes these days. With benefits such as those mentioned below, more and more organisations are choosing to base their data warehouse projects in a cloud-based architecture:

- Scalability and flexibility
- Built-in and manageable Disaster Recovery capabilities
- Cost Management
- Enhanced Security Management
- Performance

Users are finding that traditional on-premise based data warehouses cannot cope with the growing volume and demands. The global cloud data warehouse market size was worth approximately $4.6 billion in 2021 and is expected to grow to generate revenues of around $12.9 billion dollars by the end of 2026.

On-premise data warehouses have been found lacking in many areas including scalability, cost, deployment speed, and agility. The table below shows a contrast of On-Premise Data Warehousing vs Cloud Data Warehousing across various aspects.

	On-Premise Data Warehouse	Cloud Data Warehouse
Deployment	Deployed on physical servers on-site, or on a client managed Virtual Machine	Deployed on virtualized servers in the cloud
Scalability	Offers limited scalability, requires upfront hardware investment	Easily scalable with on-demand resources adjustment
Maintenance	Requires in-house IT management for updates and troubleshooting	Managed services, less maintenance burden
Cost Structure	Involves capital expenditure (CapEx) with upfront costs for hardware and infrastructure	Operational expenditure (OpEx), pay-as-you-go pricing model offers flexibility and efficiency
Flexibility	Fixed capacity, harder to adapt to changing needs	Flexible, can scale resources based on demand
Integration	Limited integration with cloud services	Seamless integration with various cloud services
Accessibility	Limited accessibility, tied to physical location	Accessible from anywhere with an internet connection

climber

Deployment Speed	Longer lead times for hardware procurement, setup, and configuration	Quick deployment with on-demand resources, reduced time-to-value
Updates and Upgrades	Manual updates and upgrades, potentially causing downtime	Automated updates, minimal downtime with managed services
Disaster Recovery	Relies on client managed or hosted backup and recovery solutions	Built-in disaster recovery options in the cloud

Cloud Data Warehouses: What are the challenges?

From a Cloud Data Warehousing perspective, there are indeed challenges to be made aware of, which are:

- **Data Integration**
  Data Integration challenges are all focused on the breadth and diversity of data sources. These days most modern organisations employ data sources which are both on-premise and cloud based. Being able to ingest and manage security around these data sources can prove to be a challenge.

- **Security**
  With security rules and restrictions varying within organisations and even within departments, it becomes a challenge to maintain security, ensuring visibility for datasets to the right recipients or audience. Ensuring consistency across various security models for users can be a challenge.

- **Vendor lock-in**
  By using features and functionality provided by cloud data warehouse vendors, users can easily become closely wedded with their chosen cloud data warehouse, with the risk of becoming locked-in to that provider. Whilst Cloud Data Warehouses can often increase efficiency and performance, it can prove to be a challenging task to transition to a different cloud provider when migrating the functionality and data warehouse deployment scripts.

- **Cost Management**
  While cloud data warehouses offer flexibility, the payment model can lead to unexpected costs if not carefully monitored. The challenge is around optimising resource utilisation to match variable workloads and data processing demands.

Which Cloud Data Warehouse Vendor is best for me?

Now that you have considered the benefits and the challenges of a Cloud Data Warehouse and likely chosen this over an on-premise option, it's at this point that a decision should be made as to which Cloud Data Warehouse vendor to choose. Here is a list of some of the vendors available, and some pros and cons to each:

Microsoft Azure Synapse

Microsoft Azure Synapse Analytics delivers capabilities around big data analytics and enterprise data warehousing. In particular, it uses SQL for data warehousing, Spark technologies to handle big data, and offers ETL and ELT based data integration. Azure Synapse Analytics also integrates seamlessly with BI tools like Power BI.

It can be a viable cloud data warehouse solution if your organization is involved in all, or most of, these data management areas. Additionally, if you already use multiple other Microsoft services, consider integrating Azure Synapse Analytics into your existing data stack since Microsoft's services integrate smoothly together.

Advantages of Azure Synapse	Disadvantages of Azure Synapse
Easy integration with other Azure based services and BI and Machine Learning Platforms	High dependency on the Azure ecosystem
Support for a wide range of data types and unstructured data	Frequent technology updates mean users must constantly adapt and learn
Easily scale to handle large datasets	Potential cost escalation with increased usage
Robust security features for data protection	Steep learning curve for users unfamiliar with the Azure platform, users often need ongoing training plans to keep up with changes to the platform
On-Demand resource provisioning means the warehouse can scale at high volume peaks to offer flexibility	Fine tuning for optimisation and administration may be complex

You should consider Azure Synapse for use cases which as Big Data Analytics, Predictive Analytics and Forecasting, Integration of Advanced Analytics, and Machine Learning.

climber

Amazon Redshift

Amazon Redshift is a fully managed data warehouse service in the cloud, provided by Amazon Web Services (AWS). Redshift can manage petabytes of data in volume. It is designed to handle large datasets and deliver high-performance analytics for organisations seeking a scalable and cost-effective cloud data warehouse solution. Amazon Redshift is particularly well-suited for analytical workloads and business intelligence applications.

Advantages of Amazon Redshift	Disadvantages of Amazon Redshift
Easily scales to deal with both small and large data volumes	Optimised for analytics use cases, less suitable for transactional workloads
Offers fast query performance, particularly good for analytics use cases	Available features vary by each AWS region
Easy integration with other AWS based services	Users will likely need training and familiarisation with the AWS platform and ecosystem
Automated backups and maintenance	Can be cost effective; but large-scale usage can often lead to significant costs
Good level of security for protecting sensitive data	

Amazon Redshift can be considered for BI and Analytics use cases, Cloud Data Warehousing, Integration with other AWS services, and scalable data processing.

Google BigQuery

Google BigQuery is a fully managed, serverless cloud data warehouse solution provided by Google Cloud Platform (GCP). It is designed to handle large-scale analytics workloads and enables you to analyse and query large datasets in real-time. Its integration with other Google cloud services makes it a comprehensive platform for various data analytics needs.

Advantages of Google BigQuery	Disadvantages of Google BigQuery
Serverless operation allows the platform to scale automatically	Not designed for transactional data; better suited towards analytics
Optimised for fast query performance, ideal for real-time data analytics	Integration with GCP, can easily lead towards vendor lock-in
Efficiently handles large datasets, scaling automatically based upon workload demands	Cost effective for small to medium workloads, can be very expensive for large-scale usage

Seamless integration with other Google Cloud services	Users will likely need to get training and familiarise themselves with Google's platform and ecosystem
Familiar SQL syntax-based querying, for easy adoption by analysts and developers	
Support for real-time data streaming	

Google BigQuery can be considered for Real-time data provisioning for analytics and dashboards, Log Analytics, IoT Data Analytics, Predictive Analytics, and Cloud Data Warehousing.

Snowflake

Snowflake is a cloud-based data warehousing platform that provides a fully managed and scalable solution for storing and analysing data. It operates as a Software-as-a-Service (SaaS) platform and is designed to be simple, flexible, and efficient for organizations seeking a modern cloud data warehouse.

Advantages of Snowflake	Disadvantages of Snowflake
Snowflake can be deployed onto multi-cloud platforms to offer flexibility and avoid vendor lock-in	Users will need to familiarise themselves with the Snowflake platform and terminology
Automatic scaling to ensure performance during high data volume periods	Some complex workloads and administration tasks might need fine-tuning for optimal performance
Easy to share data between organisations and departments	While being cost-effective, large-scale usage might result in unexpected and significant costs
Efficient and easy cloning of databases or tables without an increase in storage	Transferring data between Snowflake instances on different cloud platforms could incur additional costs
Easy access to historical data and recovery/roll-back from unexpected changes	Without the correct monitoring in place, unexpected costs may occur when using Snowflake
Ability to scale storage and compute independently	

Snowflake should be considered for user cases like data-intensive application development, inter-company or inter-organisation data sharing, and cloud data warehousing.

climber

Summary

In summary, Cloud Data Warehousing certainly appears to be the direction going forward for users finding that traditional on-premise Data Warehouses are not fit for purpose and don't offer the functionality and capabilities required for a Modern Data Stack.

There are indeed many Cloud Data Warehouse vendors in the market these days and some offer different capabilities and strengths over others. It is advisable to really consider the market before making a final decision and investment in your chosen platform. As ever, the Climber Group are always available to provide consultation and guidance on your Data Warehouse strategy and would welcome the opportunity to be involved in any conversations going forward.

How Climber can help

Since the very start in 2007 we have helped many organisations to manage their Cloud migration. We have a talented team of business consultants, BI consultants, integration specialists, and web developers with different experience and backgrounds. Together, we will help you get the best out of your Cloud Data Warehousing strategy.

Get in touch

Mark James
Data Integration and BI Consultant
mark.james@climberbi.co.uk
+44 203 858 0668

Read our customer success stories

To find out more about how we have helped companies like yours, read our customer success stories at: www.climberbi.co.uk/how-can-we-help/customer-success-stories/