PIQUE SOLUTIONS

# Modern Cloud Data Platforms

## Customer Experiences with Oracle Autonomous Data Warehouse versus Amazon Redshift

PIQUE SOLUTIONS

October 2020

THE DEVELOPMENT OF THIS WHITE PAPER WAS SPONSORED BY ORACLE. THE UNDERLYING RESEARCH AND ANALYSIS WERE EXECUTED INDEPENDENTLY BY PIQUE SOLUTIONS.

795 Folsom Street, 1st Floor | San Francisco, CA 94107 | Tel.: 415.685.3392 | www.piquesolutions.com

# Contents





Pique Solutions is a competitive research and market analysis firm supporting Fortune-500 companies in the information technology sector. Pique is based in San Francisco, California.

## Executive Summary

A modern cloud data platform should be designed from the ground up for high performance, unlimited and flexible scale, advanced security, and built-in support for a wide range of analytic workloads. It should also combine the best of proven database capabilities with a cost-effective cloud model. Customers in a variety of industries want to leverage these attributes to lower costs while driving superior business outcomes.

Pique Solutions researched two cloud data warehouse services—Oracle Autonomous Data Warehouse (ADW) and Amazon Redshift—and interviewed 10 companies about their experience implementing, operating, and using these services. We wanted to understand the impact of these services on customer experience, operational cost, and business enablement. The key findings of the study were as follows:

- While both Oracle ADW and Redshift are easy to initially provision and set up, customers found Oracle ADW to have more built-in capabilities as compared to Redshift. Starting with loading data, ADW provided advantages in using Oracle native tools and also the greater data compression to lower ongoing storage costs relative to Redshift.
- Oracle ADW was also found to better support multiple data types and a broad array of workloads, including departmental and enterprise data warehouse, graph analytics, spatial analytics, and machine learning out of the box, whereas Redshift has limited support and requires separate databases that add complexity and cost.
- Oracle ADW requires virtually no effort from DBAs or other IT specialists to operate as contrasted with Redshift, which requires extensive initial and ongoing query and performance tuning. An online directory service for small businesses in Germany found ADW 4 times faster than Redshift and, after moving to ADW, they were able to reduce their operational labor cost by 81%. The decision to move to ADW instead of Redshift came down to better performance and automation, which enabled dramatically lower operational resources and cost to run the cloud data warehouse.
- The performance of Oracle ADW is several times up to an order of magnitude (10+ times) faster than Amazon Redshift depending on the use case; Oracle ADW customers could execute complex queries in a matter of seconds, whereas Redshift customers experienced queries most often measured in minutes. Despite recent Redshift enhancements that claim to dramatically improve performance, customers we talked to found these new features targeted at niche use cases that may be 1% of all use cases and generally not worth the time and cost of investing in implementing them for the other 99%. An existing customer with a large deployment estimated it would cost them nearly $1.3 million in change management costs to take advantage of new Redshift features and, with uncertainty related to performance improvements, they opted not to deploy them.

> "It's very challenging for Amazon to retrofit, or 'bolt-on', high performance to Redshift with new features; it is not built from the ground up for consistently high performance like Oracle ADW is."
>
> **Senior Database Architect**
>
> Systems Integration Firm

- Oracle ADW was found easier to scale up or down with more granularity in terms of scaling compute and storage separately and without downtime. Redshift's new RA3 nodes with managed storage allow customers to scale compute and storage independently, but customers found they cannot easily transition "in place" and they are limited to only two fixed t-shirt sizes (12 and 48 vCPUs), which render the cost to benefit not compelling for many customers we interviewed. Specifically, scaling up in Redshift increases cost by 400%.
- Customers find Oracle far more advanced when it comes to database and data warehouse features supported, whereas Redshift is lagging. According to a chief innovation officer from a cloud services consulting firm, "When it comes to database features and functionality, on a scale of 1 to 100, Oracle is about 98 whereas Redshift is in the 40 to 50 range. The Oracle features have taken decades to build, so we wouldn't expect Redshift to match them anytime soon."
- In evaluating database security, Oracle was found to provide a high level of security out of the box. All data is encrypted by default, security updates are applied automatically with no downtime, and Database Vault hides user data from Oracle Cloud administrators. In particular, study participants called out Oracle Data Safe, saying that users can register ADW with one click: "It does all the analysis for you creating non-prod environments from prod environments and enable data masking in half a day. Redshift has nothing like this."
- Redshift lacks built-in advanced analytics (e.g., advanced statistical models, predictive analytics) and AI/ML features that come built into Oracle ADW. Redshift customers must license separate tools, such as Amazon SageMaker or third-party machine learning tools, to get advanced analytics and predictive insights. This comes at a cost in terms of integration costs as well as data movement and coordination between systems. Oracle ADW customers also noted how quickly they were able to drive insights from their data, with one customer telling us, "I never would have dreamed we would be using AI to get such fast benefits. With the pre-built models, we didn't need much training, so now we can make a huge impact without needing a lot of resource."
- Finally, many enterprise customers with strict security, regulatory compliance, or data sovereignty requirements need deployment choices that include running their data warehouse in their own datacenter. With Oracle Exadata Cloud@Customer and Oracle Dedicated Region Cloud@Customer, Oracle is the only vendor that offers a full public cloud data warehousing experience as if it were a private cloud. Customers value this choice of location, single tenancy, and the fact that the service is delivered with the same SLAs, pricing, security, and operating models as in the public cloud. Because Redshift is limited to the public cloud, Amazon does not offer its customers this deployment option.

We find that customers with heavy investments in AWS cloud infrastructure chose Redshift as a seemingly default choice for cloud data warehousing, but based on our research many are considering Oracle ADW as a modern cloud data platform that provides better price and performance, more advanced database capabilities, better security, more deployment options, and significantly lower operational cost.

# Introduction

Data warehousing in the cloud is a compelling proposition for many customers given the infrastructure and operational cost associated with building and maintaining an on-premises data warehouse deployment. A modern data cloud platform should promise and deliver to customers key benefits through the typical deployment life-cycle stages:

- Agility by being easy to set up and analyze all forms of data, from highly structured to an increasing array of semistructured and unstructured forms.
- Maximum and consistent performance across use cases while limiting database tuning, query optimization, and other costly labor.
- Securing data automatically and providing platform-level security features to proactively monitor risk and protect sensitive data.
- Flexibility with the ability to easily scale compute and storage up or down independently without downtime based on business requirements.
- A complete and integrated solution with built-in capabilities for an array of data types, data integration, machine learning, and predictive analytics that can drive rapid results.
- Choice of deployment options to account for a range of regulatory compliance and data sovereignty requirements.

Given these value drivers, many companies are migrating their data warehouses to the cloud or developing new capabilities using cloud services. As the market matures, customers are finding a broad choice of vendors and services for cloud data warehousing. Deciding which service provides proven, advanced database capabilities with the best of cloud infrastructure services can be daunting.

This paper will evaluate Oracle ADW and Amazon Redshift and discuss how they address the promises of a modern cloud data platform, particularly related to cost and performance, complexity, elasticity, security, and the operations of the services.

Our analysis, based on both secondary research and interviews with numerous customers, is organized along the lines of typical activities through the life cycle of a deployment:

- Provisioning an account and loading data
- Table design and performance tuning
- Securing data
- Querying in the data warehouse
- Sizing, scalability, and availability
- Machine learning and predictive analytics
- Deployment options for compliance and data sovereignty

## Study Approach

The research phase consisted of both secondary research combined with an in-depth data collection and interview process. Pique carefully reviewed publicly available vendor information, documentation, and relevant technical forums. In parallel, Pique identified and qualified 10 customers and partners involved in implementations inside medium-size and large organizations. These experts provided detailed experiential research and data on their data warehouse deployment. The research focused on the cloud data warehousing options and the impact of these services on IT and the broader business.

Table 1 lists the companies analyzed in the data-gathering phase of the research project.

**Table 1. Companies and Participant Roles Included in Primary Research**

| Company | Title |
|---|---|
| **Financial Services Firm** | Director of IT |
| **Commercial Real Estate Company** | Director of IT |
| **Online Directory Services for Small Business** | Chief Executive Officer |
| **Global Risk, Retirement, and Health Services** | Executive Director, Technology |
| **Global Auto Rental Brand** | Director of Business Intelligence and Analytics |
| **Systems Integration Firm** | Senior Database Architect |
| **Database and Cloud Consulting Firm** | Chief Innovation Officer |
| **Global Systems Integrator** | Data Warehouse Practice Director |
| **Regional Hospital System** | Director of IT |
| **Private University** | Director of IT |

# Customer Experiences with Oracle ADW and Amazon Redshift

In this section we discuss what our study participants—users of both Oracle ADW and Amazon Redshift—shared relative to the common activities through the life-cycle stages of a data warehouse deployment. A summary of our research and customer experiences across the stages of ADW and Redshift is listed in Table 2, which is followed by an in-depth analysis of each stage.

**Table 2. Summary of Research Findings**

| Stage | Oracle ADW | Amazon Redshift |
|---|---|---|
| **Provisioning an Account** | • Fast and easy initial provisioning | • Fast and easy initial provisioning, particularly for Amazon cloud storage customers |
| **Loading Data** | • Efficient data loading, using GoldenGate, Datapump, and other Oracle tools free of charge<br>• Supports a wide array of data types, including structured, semi-structured, and unstructured data<br>• >10× compression leads to lower storage costs | • Efficient data loading from S3; loading from other non-AWS sources requires additional tools and effort<br>• Supports limited data types, thus requiring separate databases to support many semi- and unstructured data use cases<br>• AZ64 encoding provides roughly only 3× compression compared to RAW encoding |
| **Securing Data** | • Autoencrypts data at rest and in motion by default<br>• Data Safe cited as key platform-level capability for assessing and monitoring risks and masking and protecting sensitive data | • Data encryption must be set manually, resulting in potential misconfiguration and exposing risks<br>• No built-in features to conduct ongoing security assessments, user and privilege analysis, sensitive data discovery, sensitive data protection, and activity auditing; all must be done manually through disparate tools |
| **Table Design** | • Minimal to no effort required by customers<br>• No need to consider any details about parallelism, partitioning, compression, and SQL plans; the service automatically configures the database for high-performance queries | • Extensive effort required for table design, testing, and planning to achieve good performance<br>• Customers need to define primary and foreign keys, data distribution styles as well as other tasks, which require special database expertise |
| **Query Tuning** | • Auto-tuning and query optimization | • Manual tuning and query optimization<br>• Requires ongoing effort as patterns and usage change |
| **Performance** | • Benchmarks and customer experience show that query performance is several to as much as 10×+ higher; no additional investment required | • Query performance is far lower<br>• Higher service costs and change management involved in new unproven features (e.g., AQUA) |
| **Sizing, Scaling, and Availability** | • Easy and fast to scale up or down without service disruption<br>• Fine granularity in scaling compute and storage independently and to the exact number of OCPUs and terabytes | • Limited options for independent compute and storage scaling with RA3<br>• Scaling may put the service in read-only mode, which could take hours or days to take effect depending on the size of the data |
| **Built-in Predictive Analytics** | • Built-in machine learning enables users to get predictive insights out of the box with no additional investment in service, integration, or specialized skillsets | • Machine learning not available out of the box; Redshift customers must implement additional service such as SageMaker or other third-party tools, which add integration costs, time and effort for moving data, and operational complexity |
| **Deployment Options: Compliance & Data Sovereignty** | • Can be deployed and fully managed in the public cloud, Cloud@Customer, or on premises to support regulatory compliance or data sovereignty requirements | • Only available in the public cloud, limiting the use cases for strict regulatory compliance or data sovereignty requirements |

## Provisioning an Account and Loading Data

From an account provisioning standpoint, Oracle ADW customers we interviewed said it was very fast to provision and easy to use. A director of business intelligence (BI) and analytics for a global auto rental brand shared, "[With Oracle ADW] we were able to create the data warehouse very quickly. It required only a few minutes to create the instance, 5 minutes to load the data, and then the database is ready to use. It's really fast and easy to use."

Similarly, Redshift customers found provisioning a data warehouse relatively easy, and those existing AWS infrastructure customers appreciated the ability to access data already stored in S3 buckets.

That said, some Oracle customers were still able to ingest large volumes of data into ADW, often within increasingly shorter time windows. In terms of benchmarking data ingestion, a data warehouse practice director for a global systems integration company shared his experience getting data into ADW and the ability to insert almost 500 million rows of data in less than 3 minutes. In the context of benchmarking, he stated, "This is exceptionally fast due to Oracle's architecture and it would simply not be possible in other cloud data warehouse services where it would be many times slower."

> "We were able to load into ADW 500 million rows in less than 3 minutes on average. This is exceptionally fast due to Oracle's architecture and it would simply not be possible in other cloud data warehouse services where it would be many times slower."
>
> **Data Warehouse Practice Director**
>
> Global Systems Integrator

Oracle was also cited by customers as providing higher compression when loading data into the data warehouse, which created efficiencies in terms of overhead and required less storage. Oracle was found to provide 10 times the storage compression whereas, even with Amazon's AZ64 encoding, it only reduced storage by roughly 3 times versus RAW encoding. According to a chief innovation officer for a database and cloud consulting firm that benchmarked Oracle ADW and Redshift, "The compression ratio is much better for us with ADW, so we save money there for storage as well."

Beyond loading data into ADW, Oracle also provides secure access to data from ADW and Object Storage using CloudSQL. ADW provides direct, transparent, in-place access to data stored in Oracle's Object Storage, Azure's Blob storage, and Amazon's S3. Users can even join and analyze datasets from all three object stores within a single SQL query. The key differentiator for Oracle is that the security policies in ADW are leveraged when accessing data from Object Storage. In contrast, Redshift requires customers to purchase additional ETL tools and/or connectors to load data from non-AWS object stores to Amazon S3.

Lastly, our study found that customers want to have a unified view of their data by incorporating structured, semi-structured, and unstructured data in their analytics and data warehouses. Oracle ADW satisfies their requirements by supporting a broad array of structured, semi-structured (XML, JSON, Apache Avro and Parquet), and unstructured use cases, including graphs, geospatial, and character large object (CLOB) in one single database. In contrast, Redshift does not support semi-structured datatypes such as XML, Apache Avro and Parquet, and JSON. Relative to unstructured data, Redshift is missing graphs. While it does support spatial data, it supports only a subset of spatial data types and is missing raster, 3D/Lidar, and map visualizations. Furthermore, Redshift Spectrum does not support spatial data.

As a result, Redshift customers are limited in data warehouse workloads and need separate databases to analyze semi-structured and unstructured data, which fragments their data and makes it more complex to

secure. For example, customers are required to license Amazon DynamoDB for semi-structured data and separately Amazon Neptune database for graph data, which entails additional expenses and learning multiple sets of databases and tools.

According to a senior database architect for a systems integration firm, “Redshift’s support for semi-structured and unstructured data is limited as compared to Oracle ADW. This means that Redshift customers will need multiple database deployments, which adds complexity, operational cost and additional tools for customers to learn.”

## Table Design and Performance Tuning

Study participants consistently cited the operational and manageability benefits of Oracle ADW as compared to Redshift, which required significantly more resources for a variety of operational activities. In contrast, Oracle ADW took far less effort to manage and operate; in large part due to built-in automation that handles many database manageability tasks autonomously.

Customers shared that ADW does not require tuning or consideration of details about table design and query optimization. The database is optimized automatically to ensure the highest level of performance across use cases. In contrast, Redshift was found to require a lot of manual tuning, including trial and error in analysis, design, data loading, and organization. DBAs also need to consider data distribution styles, identify the most costly queries and tables used by these queries, and consider how the tables are joined and aggregated. There is also a long-term cost aspect of this activity as query patterns change over time.

Often, the choice of data warehouse service is driven by the resource constraints of a particular customer. One company in our study, an online directory service serving small businesses in Germany, was looking to modernize its data warehouse capabilities and evaluated Oracle ADW and Redshift as they shifted to cloud data warehousing. As a medium-sized business themselves, they lacked IT budget and resources and preferred to concentrate on business and new developments rather than IT and maintenance work. They conducted a query performance benchmark of ADW and Redshift and found that ADW was 4 times faster. In the words of the chief executive officer, “ADW is fully-tuned and provides good performance out of the box, so no tuning or special database expertise is required.” As they realized this would not be the case with Redshift, they selected and implemented ADW and, in that transition, they were able to reduce their operational labor cost by 81% and achieve return on investment in just 12 months.

> “ADW is fully-tuned and provides good performance out of the box, so no tuning or special database expertise is required.”
>
> **Chief Executive Officer**
>
> Online Directory Services for Small Businesses

Auto-tuning in ADW was also found to be a key value driver for another cloud data warehouse customer, a global auto rental brand. Their director of BI and analytics shared, “The one key attribute of ADW is auto-tuning, which means that we don’t need DBAs to tune the database or to tune the queries; Oracle is doing that for you. So, for us, it’s a big revolution.” He added that the selection of Oracle ADW was driven in large part by the fact they did not have the in-house resources to support a cloud data warehouse that would need extensive tuning and database optimization.

Our research found that Redshift does not provide commensurate capabilities for auto-tuning, performance optimization, and other common database administrative tasks. One large financial services

customer shared that they had eight dedicated resources managing and administering Redshift, costing the company more than a million dollars per year in administration labor cost alone. The director of IT told us, “Initially there were a lot of performance challenges, so for the first six-month period the goal was to get it functional and not worry as much about the response time. Once it was stabilized, the DBAs spent more time on optimizing the data structures and queries as the usage patterns changed.”

## Securing Data

Data security and compliance are of utmost concern for companies seeking to leverage cloud data warehouse services, particularly for those regulated organizations and those with strict compliance requirements.

Data encryption is a critical feature for cloud data warehouses, and for Oracle it was found to have automated encryption whether data is at rest or in motion. ADW has always-on encryption to ensure that sensitive data is always protected, meets compliance requirements, and minimizes human errors during the encryption operations. In contrast, with Amazon Redshift encryption is an optional setting and therefore requires customers to enable encryption manually. DBAs have to choose a hierarchy of encryption keys to encrypt the database at the cluster level. The process that Amazon Redshift uses for encryption differs depending on how customers manage keys. If they are using AWS Key Management Service, there is an integrated process, but if they are using a hardware security module (HSM) they must use client and server certificates to configure a trusted connection between Amazon Redshift and the HSM. Therefore, it is more complex and may introduce human errors during the encryption process.

Our study also found that Oracle ADW includes built-in security via Data Safe, which helps customers understand the sensitivity of their data, evaluate risks to data, mask sensitive data, implement and monitor security controls, assess user security, monitor user activity, and address data security compliance requirements. Redshift currently does not have comparable features and requires users to implement and integrate additional security components to achieve similar benefits.

The database and cloud consulting firm shared their experience with security relative to a recent ADW deployment. The chief innovation officer told us, “We are in the middle of doing a project right now and Data Safe has been unbelievable for us. It takes just one click to register ADW and it does all the analysis for you. In less than half a day, it creates non-prod environments from prod environments and enables data masking. Redshift has nothing like this.”

“Data Safe has been unbelievable for us. It takes just one click to register ADW and it does all the analysis for you. In less than half a day, it creates non-prod environments from prod environments and enables data masking. Redshift has nothing like this.”

**Chief Innovation Officer**

Database and Cloud Consulting Firm

## Querying the Data Warehouse

Querying the data warehouse, as well as the length of time required to return the results of a query, is a key performance indicator for customers of data warehouse services. It not only has a significant bearing on the user experience but also on the cost of the data warehouse service.

In our research, Oracle ADW customers shared many examples of very fast query performance on the order of several times to as much as an order of magnitude (10+ times) faster than Redshift and other more traditional database approaches. A data warehouse practice director for a global systems integrator had performed benchmarking tests with a variety of queries and data warehouse sizes and found that Oracle ADW significantly outperformed Redshift. He reported, "ADW has shown to have better performance compared to Redshift. But as the size increases, the difference in performance becomes even more drastic. After conducting a series of Star Schema benchmarks on 96 million rows of data, the average performance advantage of Oracle ADW is at least 10:1 or greater versus Redshift." He went on to say, "No database brand can beat the parallelism capabilities of Oracle, which contributes significantly to the extreme performance of ADW."

> "The average performance advantage of Oracle ADW is at least 10:1 or greater versus Redshift. No database brand can beat the parallelism capabilities of Oracle, which contributes significantly to the extreme performance of ADW."
>
> **Data Warehouse Practice Director**
>
> Global Systems Integrator

In contrast, many Redshift customers cited performance issues and concerns related to query performance and the corresponding toll it takes from a tuning and optimization perspective. One such customer, an IT director of a regional hospital, shared that their query performance is not good even with a low number of users. Their administrators spend a lot of time tuning and optimizing, but executives are at a point where "they don't want to invest in further capacity as they aren't certain it will improve performance."

> "With Redshift, queries that took 12 minutes can be optimized down to 6 minutes but that still doesn't solve the root cause of poor performance."
>
> **Director of IT**
>
> Regional Hospital

The IT director found that even with considerable optimization by his skilled DBAs over time, query performance suffers from root-cause performance issues, saying, "Queries that took 12 minutes can be optimized down to 6 minutes but that still doesn't solve the root cause of poor performance. Ironically, the promise of Redshift was to open up user access to a variety of user groups/use cases, but the performance is so poor we can't support or enable them technically."

Redshift has recently released new capabilities aimed at improving performance. These features include Advanced Query Accelerator (AQUA) and materialized views. The AQUA capability does require separate investment and is also dependent on the new RA3 nodes, with a minimum investment of 12 vCPUs.

We spoke with several existing Redshift customers who evaluated these features and found that the performance results did not meet their expectations. This was due to their use cases, the uncertainty regarding the investment and potential benefits in the new services, and the costly change management required to implement the new features.

A director of IT at a large financial services firm shared that they evaluated the features and found that their unpredictable and changing query patterns did not match well with AQUA and materialized views. He told us, "Where they added AQUA, RA3 and some other features you can only draw the savings if your patterns are really, really specific. When there is a high degree of variance in query patterns, we're creating issues and unexpected billing." He went on to say that AWS's published performance claims were based on a very niche 1% use case that would not apply to many customers. He said, "Because the use cases they publish are very specific to the customer that got that type of savings, we don't think the savings would apply to 99% of the audience."

Another long-standing Redshift customer we interviewed, a global risk, retirement, and health services company, also evaluated AQUA and RA3, but they did not find them compelling enough to implement in production based on the change management cost they would incur to implement and the uncertainty of the benefits they would get after implementation. They have a very large deployment of Redshift. The executive director of technology estimated it would take a team of 10 people six months to redesign the system and implement and test the new features. Further, once implemented, he determined it would require another four months with 20 people to do the tuning and optimization required to get the value of the capabilities. Based on the salary cost of these technical resources, this would amount to nearly $1.3 million in change management costs alone to only start to realize the benefits of the new capabilities. He shared, "While we do feel AQUA and RA3 would add a performance improvement to our queries, it is hard to quantify what that would be, and given the large investment requirement in change management, and the fact we would pay more for the service, we decided not to proceed with these new services." They also evaluated materialized views and arrived at the same conclusion: They were not comfortable investing in a capability that would not provide a "proven" return on investment.

Our research found that Oracle ADW has extensive features that were built from the ground up to accelerate performance for all use cases, unlike what we heard from Redshift customers about their uncertainty of performance improvements and the challenge of retrofitting high performance as an add-on service element. It is considered an add-on service because it is paid for separately and it needs to be implemented. A senior database architect for a systems integration firm also evaluated Redshift AQUA relative to Oracle ADW. He found that Amazon's approach with offloading some data processing operations to AQUA mimics the approach Oracle has been using for years for data preprocessing using Exadata Storage servers, which are specifically tuned up for accelerated data preprocessing. Oracle introduced this concept roughly 10 years ago with the launch of the third generation of Exadata platform (X2), and it has matured over the years. Redshift with AQUA is far behind in their maturity of supported data-processing offloading operations. What Amazon offers to its customers with its AQUA solution represents only a small subset of what Oracle offers in Exadata under

> "It's very challenging for Amazon to retrofit, or 'bolt-on', high performance to Redshift with new features; it is not built from the ground up for consistently high performance like Oracle ADW is."
>
> **Senior Database Architect**
>
> Systems Integration Firm

> "We looked at RA3 and the economics of the service just didn't make sense for us, so we decided to stick with our DS2 nodes."
>
> **Executive Director of IT**
>
> Global Risk, Retirement and Health Services Provider

Oracle Smart Software and Hardware features. In his estimation, "It's very challenging for Amazon to retrofit, or 'bolt-on', high performance to Redshift with new features; it is not built from the ground up for consistently high performance like Oracle ADW."

The senior database architect also evaluated Redshift materialized views (MVs) and found that they are still new in Redshift, with general availability only since March 2020. There are several documented limitations for creating and using MVs, which makes it difficult to determine actual feasibility, possible performance benefits, and the management overhead while using them. Some of the challenging considerations are (a) whether or not a new MV can be created with an existing dataset, (b) whether or not it would be useful, and (c) what refresh method should be used. His overall synopsis is that Redshift is late to the game with this first-release feature that is difficult to determine whether or not, and if so where, it adds value relative to the cost of the feature.

## Sizing, Scalability, and Availability

Another important aspect of cloud data warehousing is the ability to scale up or down based on business needs. Generally, this means scaling the data warehouse up as usage and workloads expand, but organizations also want the ability to scale down and control their costs during idle periods or as certain datasets age. The primary challenge for Redshift customers is that, until recently with the announcement of the RA3 nodes, they are not able to independently scale compute and storage; therefore, when they scale storage, they have to scale compute at the same time, whether they need the additional compute or not, and vice versa. This leads to excess capacity and a more cumbersome process to upgrade Redshift instances. Further, there still are issues related to service availability when resizing a Redshift data warehouse.

The IT director at the regional hospital shared his experience in terms of the expectation of Redshift being easy to scale not matching with his actual experience: "Whether we are scaling up or down or changing the number of nodes, we are finding issues in terms of availability of the warehouse and query execution which can stretch from hours to days." Specifically, he told us that scaling Redshift up or down requires taking the system to a read-only state while data is redistributed across all the nodes, resulting in a big impact to his business. When Redshift clusters are involved, the process is further complicated and involves more in-depth planning, analysis, and scheduling operations.

In contrast, Oracle ADW customers noted the ease of scaling up or down without downtime or disruption. They also cited the flexibility and granularity of adding compute or storage independently and at exactly the size they needed. A data warehouse consulting firm echoed the ease of scaling up and down: "Scalability is another thing that ADW does very well. We are able to add or reduce CPUs or storage with no impact to availability of the service or user queries."

Redshift customers we interviewed are limited in scaling compute and storage separately unless they opt for the new RA3 nodes, which only offer two large-capacity sizes starting at 12vCPUs (jumping to the next t-shirt size of 48 vCPUs). This requires a large, 4X step function in terms of computing cost. The executive director of IT for the global risk, retirement, and health services provider evaluated RA3 and, based on the service offering and pricing, was not convinced it would be cost effective to change from their existing DS2 nodes, in part because of the overall service discounts they were getting for the combined compute

and storage. He said, “We looked at RA3 and the economics of the service just didn’t make sense for us, so we decided to stick with our DS2 nodes.”

Further related to availability are several key points, the first of which regards maintenance. ADW can be updated without disruption while the service is running so there is no data warehouse downtime. Redshift does not support rolling updates and requires shutting down while undergoing regular maintenance. When Redshift is performing planned maintenance, it terminates any queries or other operations that are in progress.

The second point has to do with failover. ADW can also automatically fail over to a standby with no data loss and no further administration. All capabilities of ADW are transferred to the standby system, which is automatically kept up to date. In the case of Redshift, when a node fails, the whole cluster becomes unavailable until the failed node is restored.

The final point regarding availability, closely related to the first two, is the availability SLA. With ADW, Oracle’s SLA is 99.95% versus Redshift’s SLA of 99.9%, which results in 2 times more downtime hours compared to Oracle ADW customers. Furthermore, planned maintenance is factored into the ADW SLA, whereas that is not the case for Redshift, so the expectation of 99.9% availability is often not met during the course of production deployment when factoring in planned maintenance.

As a result of these sizing, scalability, and availability shortcomings, customers found Redshift limited in its ability to support enterprise, mission-critical requirements.

## Machine Learning

Yet another important aspect of the deployment life cycle and evaluation of Oracle ADW versus Amazon Redshift is in the area of machine learning and advanced analytics. Customers are increasingly interested in not just creating data warehouses for query use cases but also for drawing actionable insights without the need for specialized resources or third-party tools.

Our research finds that Oracle ADW customers can use the built-in machine learning and advanced analytics in ADW to automate discovery of new insights, generate predictions, and add “AI” to data. This requires no additional investment or integration on the part of the customers. Customers shared the key elements of Oracle’s data management and machine-learning capability, including in-database implementation of machine-learning algorithms, intelligent defaults, and fully parallelized machine-learning models that can be accessed using SQL. The benefits of this approach include the fact that data remains in the database, eliminating the need for data duplication or separate analytical servers. Moreover, Oracle Machine Learning supports a wide variety of users, including R and Python developers, data scientists, data analysts, and even citizen data scientists.

The result is ADW delivers rapid results across a variety of industries, including education, financial services, healthcare, marketing, consulting services, the public sector, and others. A private university in northern India provides a great example of using Oracle ADW predictive analytics to quickly conduct trend analyses and implement changes based on dynamic student data. The admissions department can now predict which specializations will fill fastest, so they can better allocate classrooms and infrastructure. The result was lab capacity planning that is 30% to 40% more accurate. In the area of student housing, they can predict when rooms are likely to become vacant as students move off site or transfer, resulting in a 20% improvement in filling vacancies. The director of IT explained, “Twelve months ago, I never would have dreamed we would be using AI to get such fast benefits. With the pre-built models, we didn’t need much training, so now we can make a huge impact without needing a lot of resource.”

Conversely, while Amazon positions Redshift as a data or data science platform, it does not have built-in machine learning and predictive analytics. Redshift's customers have no choice but to use other tools, provided by either AWS (such as SageMaker for AI/ML) or third-party vendor tools. Customers told us that Redshift involves implementing another service and learning another set of tools to access predictive insights. This comes at a cost in terms of integration costs as well as data movement and coordination between systems. It can also result in delays to data usage and the insights generated for customers.

When asked about predictive analytics, the IT director for the commercial real estate company reported, "We evaluated AI/ML tools and talked to the vendors but just didn't want to integrate more stuff and create more dependencies with Redshift. We decided not to put those into a sprint and to create more interdependencies to keep using Redshift, particularly as we look to migrate to another cloud data warehouse."

"We evaluated AI/ML tools and talked to the vendors but just didn't want to integrate more stuff and create more dependencies with Redshift. We decided not to put those into a sprint and to create more interdependencies to keep using Redshift, particularly as we look to migrate to another cloud data warehouse."

**Director of IT**

Commercial Real Estate Company

## Deployment Options

A final difference we found between ADW and Redshift was related to the choice in deployment options. Redshift is limited to the public cloud and cannot be deployed in the customer's own data center to meet the stringent requirements for regulatory compliance and data sovereignty. Our research finds that ADW is available in more choices of deployment than Redshift, including the public cloud and Cloud@Customer. With Oracle Exadata Cloud@Customer and Oracle Dedicated Region Cloud@Customer, Oracle is the only vendor that offers a fully managed cloud data warehousing experience within the customers' own datacenters. Customers value this choice of location, single tenancy, and the fact that the service is delivered with the same SLAs, pricing, security, and operating model as in the public cloud. Organizations can therefore easily transition from on-premises to the cloud and vice versa to meet their business requirements while leveraging the same skillsets across deployments.

"Customers like everything the public cloud has to offer, but there are still plenty of customers who want it on-premises and dedicated to them."

Leading Industry Analyst Firm

A leading industry analyst firm shared the requirements as stated by many of their enterprise clients, reporting, "Customers like everything the public cloud has to offer, but there are still plenty of customers who want it on-premises and dedicated to them."

# Conclusion

Our research into customer experiences with Oracle ADW and Amazon Redshift shows clear and compelling advantages for ADW across the life-cycle stages of a data warehouse. Simply put, ADW delivers on the needs of a modern cloud data platform while Redshift was found lacking in many areas.

Study participants tell us that the architectural foundation and database maturity built into ADW, combined with a cost-effective cloud model, leads to substantially higher performance across a broad range of workloads and customer use cases. The higher performance, achieved and sustained without extensive customer involvement and resources, means a significantly lower cost of ownership.

Further, ADW includes numerous built-in capabilities that do not require additional service fees and integration as in the case of Redshift. Redshift recently released features that claim to address performance and service gaps, but customers told us the purported performance gains may not address their use cases or query patterns. Redshift customers we talked to are not willing to invest more in additional service fees and the expensive change management in testing and rolling out new features such as AQUA, RA3 nodes, and materialized views that have uncertain performance benefits. The experience of most Redshift customers is that sustained performance comes with a significant cost of people continuously tuning and optimizing the database and associated queries.

Consider the following key points from our customer research comparing experience with Oracle ADW versus Amazon Redshift:

- ADW loads data more efficiently and with greater compression than Redshift, as well as provides support for a wider array of data types, which lowers cost and complexity by having multiple data in one system.
- Despite Redshift claiming new features to improve performance, they are suited to a very small percentage of use cases. Existing Redshift customers found it too costly to implement features like AQUA, RA3, and materialized views, with one large deployment estimating it would cost $1.3 million to implement, test, and deploy the new features.
- There is a substantive difference in the operational labor cost required for ADW and Redshift. ADW customers found labor cost lowered by 81% when moving to ADW due to higher performance and more autonomous operations.
- Scaling the data warehouse up or down in ADW was easy with no disruption for customers, whereas Redshift customers experienced disruption on the order of “hours to days” due to the database switching to a read-only state. Redshift customers also complained about the complexities around selecting the correct cluster reconfiguration. Further, Redshift has limited scaling options. Even with the new RA3 nodes, scaling to the next larger t-shirt size will increase cost by 4 times.
- Machine learning is an important workload for data warehouse customers who want to drive predictive actions from data without additional investment in separate services, integration, and specialized skills. ADW customers find machine learning as a built-in capability is easy to use and requires minimal training, whereas their Redshift counterparts incur integration costs, time and effort for moving data, and operational complexity.

In summary, customers found Oracle ADW to be a highly performant, secure, and cost-effective alternative to Amazon Redshift. With Oracle ADW, companies can realize savings not only from lower cloud service fees but also from significant performance and automation advantages. They can lower operational costs via reduction in ongoing tuning, optimization, and integration costs associated with third-party tools and multiple databases.