PIQUE SOLUTIONS

Cloud Data Warehousing

Customer Experiences with Oracle Autonomous Data Warehouse versus Amazon Redshift

PIQUE SOLUTIONS

October 2018

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

Nearly all companies recognize the importance of harvesting value from their enterprise data. Data warehousing is a foundational element to that end, but one that has been a challenge for companies with first-generation cloud data warehouse solutions. Cloud data warehouse services provide all customers with the opportunity to realize the value of data warehousing, including cost savings, agility, and enablement of new business opportunities.

Pique Solutions researched two cloud data warehouse services—Oracle Autonomous Data Warehouse (ADW) and Amazon Redshift—and interviewed a dozen companies on their experience implementing, operating, and using these services. The key findings of the study were as follows:

- The performance of Oracle ADW is often more than an order of magnitude (10x+) faster than Amazon Redshift. Oracle ADW customers could execute complex queries in a matter of seconds, whereas Redshift customers experienced queries most often measured in minutes.
- Customers found Oracle ADW very cost effective compared to Amazon Redshift, particularly when factoring in performance and other operational costs. Redshift customers cited examples of hundreds of thousands to in excess of a million dollars per year in operational costs.
- Oracle ADW was found easier to scale up or down in place and without downtime, whereas Redshift required "hours to days" to reconfigure the service and cannot scale compute and storage independently. Redshift customers also found it difficult to reclaim deleted space, meaning costs did not go down unless they redeployed and started from scratch.
- Redshift customers found issues with query performance that led to sub-par user experience, which required proactive usage guidance as to when and how to make queries to avoid performance issues. As a result, our study found that Redshift customers spend about 25% of their management effort on tuning.
- Customers interviewed also cited the ease of management of the Oracle ADW service, driven by built-in machine-learning automation, vis-à-vis Redshift, finding that ADW reduces effort by 100% for simple tasks and 50% to 60% for complex tasks.
- Oracle ADW customers also cited the value in being able to move data back and forth between on-premises systems and the cloud. Despite the AWS Database Migration service, this was not noted as a strength related to the Redshift deployments in customers we interviewed.
- Nearly all study participants were satisfied with their cloud data warehousing deployments relative to their existing on-premises environments in terms of cost, preferring the "pay-as-you-go model" that cloud delivers.

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. Cloud-based data warehouse services should promise and deliver to customers:

- Agility by being easy to set up and load data quickly; data volume is increasing and the window to ingest and process is continually shrinking.
- Flexibility with the ability to easily scale up or down based on business requirements.
- A lower cost and sustainable cost model for analytics.
- A reduction in operational costs through built-in automation and machine-learning capabilities.

Given these value drivers, many companies are migrating their data warehouses to the cloud or developing new capabilities using cloud services. As an early entrant in the cloud data warehouse market, Amazon Redshift has garnered adoption particularly by companies using other AWS infrastructure. A 2018 survey by Panoply.io confirmed the usage of Redshift but also concluded that those same customers experienced challenges with complexity and performance in their Redshift data warehouse.[1] Customers also cited cost as a reason for being unsatisfied with Redshift.

The study specifically said, "The bottom line is whether a company is big or small and even if they are already in the cloud—managing a data warehouse is often far too complicated." The study also asked which areas customers would want to see automated and they listed four key areas:

- Ingesting different data sources
- Transforming data
- Managing data
- Query optimization

Clearly, the promise of cloud-based data warehousing is falling short with respect to Redshift customer experiences. As the cloud services market matures, customers are finding a broader choice of vendors and services for data warehousing including Oracle ADW. This paper will discuss Amazon Redshift and Oracle's ADW and how they address the promises of cloud, particularly related to cost/performance, complexity, operational costs, and the automation of the existing pain points of cloud data warehousing.

[1] https://learn.panoply.io/data-warehousing-trends-report-2018

A Look at Two Popular Data Warehouse Cloud Services

Before diving into the primary research and customer experiences, it is important to provide some context on the solutions included in the study and each provider's approach and key solution capabilities.

Oracle Autonomous Data Warehouse

- This cloud service is based on intellectual property and subject matter expertise from Oracle's lengthy history of enterprise-scale data management.
- It is built on key Oracle Database capabilities including parallelism, columnar processing, and compression.
- It includes advanced features such as SQL pattern matching and JSON capabilities for unstructured data. Approximate Queries is a powerful set of analytics that allows data discovery quickly and frees up compute resources.
- All aspects of management such as performance tuning are automatically managed via machine learning, so the service requires little to no human involvement, thus increasing productivity and lowering on-going operational costs.
- It integrates with a broad ecosystem of business analytics, data integration, and IoT services within Oracle's comprehensive range of integrated cloud solutions.
- ADW runs on engineered systems and offers extreme performance based on the proven Exadata.
- Pricing is based on compute resources per hour and storage based on capacity per month in either pay-as-you-go or monthly flex models.

Amazon Redshift

- Amazon launched Redshift in late 2012 based on technology licensed from ParAccel, with the underlying Amazon database capabilities built upon Postgres back in 2005.
- With the addition of Amazon Redshift Spectrum, it can also include analysis of unstructured data in S3.
- Service is uniquely dependent on AWS infrastructure services and relies on S3 for data ingestion and backup.
- Customers are required to perform tuning, query optimization, and other maintenance activities to obtain sustained performance.
- Redshift often requires the use of third-party tools for ETL and to round out analytic solutions, with most customers using Tableau for front-end query and analysis.
- Pricing is based on a regional model and on compute and storage according to predetermined shapes. All are priced on an hourly basis, with additional fees for clustering and S3 storage.

Study Approach and Methodology

The primary research phase consisted of an in-depth data collection and interview process. Pique identified and qualified 12 customers and partners involved in implementations inside medium and large organizations. These experts provided detailed primary research and data. The research focused on the cloud data warehousing options and the impact of choice of vendor platform on the IT environment and the broader business requirements.

The research process and methods were as follows:

- Reviewed publicly available information and secondary research on cloud application development trends, drivers of adoption, use cases, and key value drivers.
- Identified and qualified 12 customer interviewees who participated in multiphase in-depth interviews and data gathering for each of the different cloud data warehouse solutions.
- Synthesized data and research findings.

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

Table 1. Companies and Participant Roles Included in Primary Research

Company	Title
Financial Services	Director of IT
Regional Hospital	Director of IT
Multinational Conglomerate	Data Architect
Manufacturer	Data Warehouse Manager
Global Risk, Retirement, and Health Services	Executive Director, Technology
Global Auto Rental Brand	Director of BI and Analytics
Sports Information/Data Science Company	Chief Operation Officer
Laboratory Management Provider	CEO and Founder
State College/University	Enterprise Architect
Advanced Data Analytics Provider	CEO
Data Warehouse Consulting Company	CEO
Global Systems Integrator	Data Warehouse Practice Director

Customer Experiences with Oracle ADW and Amazon Redshift

In this section we discuss what our study participants and users of both Oracle and Amazon services shared relative to service cost, performance, manageability, and enterprise capabilities.

Service Cost and Performance Considerations

Cloud data warehouse services do provide publicly available pricing based on common scaling factors such as processing power and storage. That said, Amazon and Oracle take different approaches to service pricing that makes an "apples-to-apples" cost comparison challenging and not representative of actual or total cost given other factors such as performance and other operational costs. Customers we interviewed found that the performance and operational aspects added significant levels of expense to Amazon Redshift deployments.

> "I have 25 years of database experience and the performance of ADW was simply astonishing."
>
> **Data Warehouse Practice Director**
>
> Global Systems Integrator

Oracle ADW customers shared many examples of very fast query performance more than an order of magnitude (10x+) faster than Redshift and other more traditional database approaches. A data warehouse practice director for a global systems integrator had performed some benchmarking tests with a variety of queries and found that "even complex analytical queries were executed in under 2 seconds on average." He also did some benchmarking of ADW to the more general Oracle Database Cloud Service and concluded, "I have 25 years of database experience and the performance of ADW was simply astonishing."

Beyond the technical aspects of query performance, several Oracle customers cited the linkage to business value resulting from faster analytics. The CEO and founder of a laboratory management services provider discussed the performance of ADW, sharing that "a lot of the competitors can take anywhere from one to two weeks to get results, and time is everything in healthcare. Patients do not want to wait one to two weeks for blood results. So, with the Oracle tools and the ADW platform, we can find efficiencies within our laboratory and provide results as quickly as one day, and in some instances even as fast as half an hour."

An enterprise architect with a state college/university system commented on the performance and scale of ADW, stating, "The query performance of Autonomous Data Warehouse is amazing. On-premises you are constrained by physical size, and when you run out of space it's hard to scale, so with ADW to be able to scale very elastically is phenomenal."

By contrast, several Redshift customers we spoke to shared experiences of query performance challenges. Many shared their use of on-premises Tableau to access Redshift data, which often results in "poor query performance and data egress challenges commonly attributed to Redshift."

> "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

One such customer, the IT director of a regional hospital, shared that although they are pleased with the cost of their Redshift

deployment relative to on-premise, their query performance is not good even with a low number of users. Their admins 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."

He found that even with considerable optimization by his skilled DBAs, 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."

Other Redshift customers shared the effort and workaround solutions to bolster the performance of Redshift to meet their needs related to query performance. For data scientists running more complex queries, they were forced to offload to a separate, Hadoop-based platform (EMR) due to performance issues. A data architect at a multinational conglomerate required a caching solution on top of Redshift to do column encoding. "Our goal was that any user interface–driven query or end-user–driven query should return results within 3 seconds. Redshift fell far short of this goal. So we had to build a lot of additional data structures and add aggregation strategies to improve the end-user experience."

Beyond the service subscription costs, other Redshift customers shared other costs associated with their data warehouse deployments. For example, many Redshift customers cited the need for VPC connections to compartmentalize deployments for security and chargeback reasons. A global risk, retirement, and health services company cited the need to build their own DR/HA solution, which ended up being about a third of the cost of their overall spend and amounted to hundreds of thousands of dollars per year in their environment.

One interesting point about the overall service cost and the impact of cost/performance is how it translates back to the business. Several Oracle customers cited the ability to lower the cost of services to their end customers. The chief operating officer of a sports information company explained that they seek not only to lower the cost of their services to their customers but also to improve the fan experience. "We are seeing savings from sheer performance of ADW and through the capabilities such as machine learning and AI. And we expect to share those savings with our customers. So those cost savings will be reflected in richer services, lower costs, and a better experience for our fans and customers."

"We are seeing savings from sheer performance of ADW and through the capabilities such as machine learning and AI. And we expect to share those savings with our customers."

Director of IT

Sports Information Company

Another Oracle customer, a laboratory management provider, discussed the impact of rapid analysis on patient outcomes. "The benefit of using Oracle's tool and Autonomous Data Warehouse is saving patient lives faster than our competitors. When we can predict and analyze how our machines and how our business operates and provide the results that they need faster than the competition, time is everything in healthcare, and time is saving human lives." He went on to say that by using Oracle ADW they can save a huge amount in human labor resources on the order of 15 to 20 financial analysts per lab. In other words, he shared that the dramatic increase in productivity related to the analytic capability enabled by Oracle Autonomous Data Warehouse actually reduced the number of analysts required, equating to over a million dollars per year per lab.

Manageability and Operational Cost Considerations

Equally important, if not more so, than the cost and performance aspects of cloud data warehousing are the costs associated with implementation, management, and ongoing operations.

Implementation

From an implementation standpoint, Oracle ADW customers we interviewed said it was very fast and easy to use. A director of BI and analytics for a global auto rental brand shared that "[with Oracle ADW] we were able to create the data warehouse very quickly; 3 minutes to create the instance, 5 minutes to load the data, and then you can use the database. It's so fast and easy to use."

Another aspect of ADW that customers touted was the ability to access and query external data sources and not necessarily have to execute ETL processes for data they want to analyze. A director at a data warehouse consulting company stated, "With ADW you remove the pressure to pull every single data source into the warehouse as you have to with other data warehouse solutions. This often derails projects as you never stop writing ETL." He went on to share, "You can run analytics in ADW on data residing in other databases and cloud services. ADW is extremely flexible in that respect which is a very powerful capability. We see it as kind of a federation engine, if you will, and because Oracle has a particularly powerful engine, we can do a lot with leaving the data where it is, but organizing it, running analytics on it from within the data warehouse cloud service."

That said, some Oracle customers were still able to ingest large volumes of data into ADW, often within increasingly shorter time windows. According to the chief operations officer of the sports information company, "One of the challenges with data ingestion is that every team, every customer has a different footprint. So getting the data inside the data warehouse is increasingly more challenging as we add and enrich more data to the warehouse. The window of opportunity throughout the night to get into the data warehouse requires more efficiency. Oracle Autonomous Data Warehouse Cloud is one answer to getting the data into the warehouse within the window of opportunity that we have, which typically ends by 6:00 AM."

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 700 million rows of data in just 107 seconds. "This is exceptionally fast. It's the architecture and Exadata that help you do that."

> "ADW gives us the ability to insert almost 700 million rows of data in just 107 seconds. This is exceptionally fast. It's the architecture and Exadata that help you do that."
>
> **Data Warehouse Practice Director**
>
> Global Systems Integrator

By contrast, Amazon customers cited challenges in getting data into and out of Redshift with respect to other elements of Amazon services. For example, Redshift requires a two-step process for getting data into Redshift—first into S3 buckets and then into Redshift. "Redshift doesn't allow moving the data directly from sources outside Amazon. It must be in S3 bucket or something. There were times where we had to move the data to S3 bucket and then load it, obviously, to Redshift. We had to jump through hoops."

Another customer, due to some security considerations, had to engage Amazon to write custom JSON scripts and use Virtual Network Gates to be able to ingest data in Redshift. Initially standing up and getting data into a cloud data warehouse is certainly important, but study participants also relayed the importance of ongoing efforts required by their teams.

Tuning, Optimization, and Automation

Customers interviewed consistently cited the 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 machine learning and AI to automate many database manageability tasks.

While several Amazon Redshift customers did tout the ease of initial implementation, others found it more complex and requiring the help of third-party implementation partners, and even Amazon directly, to get data into Redshift and to tune and optimize the performance. Roughly 25% of ongoing management effort is dedicated to performance tuning. These activities are performed by highly skilled DBAs that typically command salaries in the range of $90 to $130K per year. According to a financial services company, standing up Redshift "required specialized Amazon consultants to optimize and fine-tune the environment." Paying external consultants adds a 30% to 50% cost premium over using in-house resources.

Autotuning was a significant capability for Oracle ADW customers. According to the director of BI and analytics for the global auto rental brand, "I will say the one key attribute [of ADW] is autotuning, which means that you don't have to spend time with your DBA, logical DBA, physical DBA to tune the database, to tune the queries; Oracle is doing that for you. So, for us, it's a big revolution." Another customer, a multinational conglomerate echoed, "And that's where autonomous data warehouse comes in, where they are essentially reducing the need for customer-tuning of a lot of routine activities."

Finally, the data warehouse practice director for the global systems integration company discussed in further detail the artificial intelligence and machine-learning (AI/ML) capabilities, explaining that "ADW's AI/ML capability and its implementation automate tasks that are manual in Redshift such as performance tuning and node balancing. As part of the service, Oracle on the backend uses SQL tuning advisor that can analyze workload that is causing performance issues, and it can tune itself. He went on to say that one of the most important things in data warehouse when you don't have to worry about indexes and data models as this is all automated in ADW. The system is so well engineered and architected that automation works extremely well.

To summarize, he shared that "the automation provided by ADW eliminates manual intervention by 100% for simple tasks and by 50% to 60% for complex tasks." Simple tasks were defined as patching, configuration, tuning, performance optimization, replication, capacity planning, database/code/schema/object maintenance, tablespace and data file management, and testing and monitoring clusters and scalability, to name a few.

> "The automation provided by ADW eliminates manual intervention by 100% for simple tasks and by 50% to 60% for complex tasks."
>
> **Data Warehouse Practice Director**
>
> Global Systems Integrator

Redshift does not provide commensurate capabilities for autotuning, optimization, and other common database tasks. One large Redshift customer shared that they had eight dedicated resources managing and administering Redshift, costing the company well in excess of a million dollars per year in administration labor cost. He told us, "So initially there were a lot of performance challenges. Over time they spent more time on optimizing the data structures but usually, in the first six-month period, the goal was to get it functional and not worry as much about the response time."

Scaling Compute and Storage Independently

Another important aspect of cloud data warehousing is the ability to scale based on business needs. Generally, this means scaling the data warehouse up as usage and the workloads expand, but increasingly companies want the ability to scale down and control their costs during idle periods or as certain data sets age. Again, this was noted as a key area of differentiation in our research comparing Oracle ADW to Amazon Redshift. The primary challenge for Redshift customers is that it cannot scale "in place" and that there are dependencies on the compute and storage aspects of the service based on the way the service is designed.

Customers cited both issues in scaling up and scaling down Redshift. According to an IT director at a multinational conglomerate, "You cannot just scale in place. Those are the limitations, particularly compared to let's say Oracle's autonomous data warehouse. If you must build a larger environment in Redshift, the existing one is kind of a throwaway environment."

Another IT director at a regional hospital shared his experience in terms of the expectation of Redshift being easy to scale not matching with his actual experience: "We scaled once in the last two years and it is a large issue to scale up or down because you have to go in and rebuild and configure everything again."

An advanced data analytics provider shared how easy it is to scale ADW service up and add additional space and processing capability. "We have scaled up on the number of processors and memory we are using, and we saw immediate results across the board in terms of dashboard responsiveness to mobile devices. And then being able to add additional space into the data warehouse as well, on the fly—It has been fantastic and such an easy process."

A data warehouse consulting firm echoed the ease of scaling up and down: "Scalability is another thing that ADW does very well." They shared the difficulty to determine the right number of CPUs and the amount of storage needed. ADW strikes the right balance in the scalability equation. This eliminates the need to redesign with a greater number of users, "You just add CPUs."

A final point relative to cost is the ease at which Oracle ADW customers shared they were able to scale CPU to zero while maintaining storage during periods when not actively using the data warehouse. According to the CEO of a data warehousing consulting company, "ADW enables you to do an unlimited number of pilot projects and then put them on ice without the need to keep paying the compute charges."

By contrast, Redshift customers we interviewed were not able to scale processing power and storage separately and thusly continued to pay for processing along with storage even during periods of idle activity.

> "ADW enables you to do an unlimited number of pilot projects and then put them on ice without the need to keep paying the compute charges."
>
> **CEO**
>
> Data Warehousing Consulting Company

Reclaiming Space

Finally, Redshift customers told us that it does not reclaim deleted space and hence does not reduce the cost of the service, which was the expectation going in. A director of IT for a risk, retirement, and health services company said, "Every time we deleted something in Redshift it didn't reclaim the space. You should be able to scale up and scale down, and when you scale down you should be able to see the cost savings and that's not the case. The only way to do this is that you wipe everything, and you start it from scratch." It requires a significant amount of time on the part of the DBAs to perform such work.

"Every time we deleted something in Redshift it didn't reclaim the space. You should be able to scale up and scale down, and when you scale down you should be able to see the cost savings and that's not the case. The only way to do this is that you wipe everything, and you start it from scratch."

Director of IT

Risk, Retirement, and Health Services Firm

Enterprise Capabilities

The final aspect of our research related to making enterprise-class capabilities available to customers, particularly for less sophisticated, smaller and mid-size business users that would not be able to build, maintain, or leverage advanced data warehouse features—specifically, capabilities such as advanced analytics, data security, and ease-of-use to support various customer scenarios.

Based on our research, Oracle customers touted the ability to bring enterprise-class analytics capabilities to even SMB companies that don't have the technical resources or expertise to maintain a data warehouse. Several of the study participants said that "it levels the playing field."

The data warehouse consulting company shared that ADW is an easy entry for customers trying to build a new data warehouse. They found it suited for customers who want to quickly set up a warehouse and solve specific business problems. "With the autonomous data warehouse cloud offering, the design of the system becomes, frankly, a lot simpler because you don't think as much about how you have to index your tables, partition the tables, that type of thing, and instead you can focus on the value coming out of your data warehouse," he said.

Customers also touted Oracle security as a foundational capability. According to the chief operations officer for the sports information firm, "The Oracle Data Warehouse Cloud has the security that we always envisioned from Oracle. Oracle is certainly the database management systems par excellence and the security has no comparison."

Customers found Redshift lacking in some key data warehouse features. One example is the lack of enforced uniqueness in the data structures. "One of the biggest problems with Redshift is that there's no uniqueness, which is enforced. So especially if we have got data that is across multiple market applications, things are extreme. We have 77 different applications that are writing into different things. So, what we have to do is we have to enforce these rules ourselves."

Another aspect related to security and a limitation for customers who maintain multiple datasets for their unique clients is that Redshift requires configurations of multiple VPCs to contain customer data. According to our study participants, this introduces complexity and adds cost for Redshift deployments.

Conclusion

With the promise of cloud services being cost-effective services based on need, having low operational costs, and featuring the ability to scale up or even down, Pique Solutions' research found that Oracle ADW fulfills this promise far better than Amazon Redshift. While Redshift has experienced adoption, according to customers this first-generation data warehouse service has not evolved significantly in terms of its key performance, manageability, and automation features.

Oracle has surpassed Redshift with the release of ADW and provides a compelling value for customers of all sizes due to its lower cost, higher performance, and the reduction of management and operational costs. This comes from a combination of better service architecture, automation, and machine-learning capabilities.

> "Oracle ADW is so well engineered and architected, and the automation works extremely well."
>
> **Data Warehouse Practice Director**
>
> Global Systems Integrator

The key differences our research identified based on customer experiences with the two services are the following:

- Oracle ADW is very cost effective compared to Amazon Redshift, particularly when factoring in performance and other operational costs.
- Oracle ADW was found to be very easy to set up, configure, ingest data, and even analyze data in external systems, whereas Amazon Redshift customers cited issues related to ingesting data and implementing ETL processes.
- The performance of Oracle ADW is often more than an order of magnitude (10x+) faster than Amazon Redshift. Many of the Amazon Redshift customers cited challenges relative to query performance.
- Machine-learning driven automation provided by ADW eliminates manual intervention by 100% for simple tasks and by 50% to 60% for complex tasks. No tuning or query optimization was required for the Oracle ADW deployments; in contrast, Amazon Redshift customers spent a significant portion of their administration time on this activity, both initially and on an ongoing basis.
- Scaling the data warehouse – both compute and storage, up or down, was found far easier in Oracle ADW than in Amazon Redshift, which impacted agility, staffing efforts, and service cost.
- Lastly, study participants shared that ADW offers enterprise-grade analytics, security, and flexibility to smaller and mid-size companies, masking the underlying complexities of data warehouse design and management, so organizations can focus on analyzing and deriving insights from data more quickly.

The data warehouse practice director of the global systems integrator summed up succinctly what our research consistently revealed, telling us that "Oracle ADW is so well engineered and architected, and the automation works extremely well."