PIQUE SOLUTIONS

Cloud Data Warehousing Platforms

Customer Experiences with Oracle Autonomous Data Warehouse and Snowflake

PIQUE SOLUTIONS

July 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 management consulting and market analysis firm supporting primarily Fortune 500 companies in the information technology sector. Pique is based in San Francisco, California.

Visit www.piquesolutions.com to learn more about our consulting and market research services.

Executive Summary

Effective data management allows companies to efficiently run their operations, identify new business opportunities, and improve customer experiences. Modern cloud data platforms incorporate workloads for data engineering, data warehousing, data lakes, and increasingly data science applications. These platforms offer significant benefits over traditional on-premises deployments when leveraging a combination of proven database capabilities with the cloud's obvious benefits in terms of ease of provisioning, unlimited scale, agility, and reduction in effort associated with fully managed operations.

Pique Solutions researched two cloud data platforms—Oracle Autonomous Data Warehouse (ADW) and Snowflake—to understand the capabilities and the experiences of customers using these cloud data services in a variety of deployment scenarios. The study involved both secondary and primary research, and analyzed the common activities a typical company undergoes in a cloud data platform deployment. It also captured the voice of the customer who have used these services over a period of time.

The key findings of the study are as follows:

- Customers value a cloud data platform with comprehensive capabilities that enable them to gain insights quickly and effortlessly without the need to patch together a disparate set of tools and services. Customers we talked with found value in Oracle ADW providing a complete enterprise solution that combines the best of fully managed cloud automation and unlimited scale with the best of proven core data warehouse capabilities. In contrast, Snowflake customers lauded Snowflake for its cloud capabilities, but it lagged behind Oracle in core data warehouse features and capabilities. Moreover, Snowflake required licensing and implementation of a variety of third-party tools to comprise a complete solution. Many Snowflake customers cited licensing tools for ETL costing more than $10,000 per user per year, and for analytics and reporting costing upward of $1,000 per user per year. These figures add significantly to the cost of ownership of the cloud data platform.
- Migrating from traditional on-premises data warehouses was more challenging with Snowflake than with ADW due to the architectural differences between the two data warehouses. Most Snowflake customers we interviewed did a significant redesign or focused on net new deployments. One customer told us it took a 30-member team more than 6 months to redesign and replace the capabilities of a 600-TB on-premises MPP-style data warehouse deployment in Snowflake, which resulted in more than $1.6 million of transition costs. In comparison, Oracle customers found that migrating from existing data warehouses, particularly from Oracle databases, to ADW was much easier and less costly compared with customers' experiences with Snowflake.
- Customers using Oracle ADW found storage consumption optimized due to efficient data loading and advanced compression, whereas many Snowflake customers cited issues with storage bloating due to how data is loaded and distributed when features like Time Travel and Fail-safe are activated. One Snowflake customer saw their 10 GB base table expand in size, by a factor of 500 times, to 5 TB when they turned on Time Travel with a 90-day retention period. They learned later this was due to how Snowflake inserted and updated the data. Ultimately, they were led to shorten their retention

period to just 30 days, thus limiting their analysis horizon. Oracle ADW provides 60 days of backup included in the subscription cost, meaning no additional charges for storage.

- Both Oracle and Snowflake provide for robust data encryption; however, Snowflake currently lacks advanced data security features, such as ADW's built-in Data Safe, Risk Monitoring, Privilege Analysis, Database Vault, and Real Application Security. Even with Snowflake's new dynamic data-masking feature, several Snowflake customers found it to be a "work in progress" relative to broader security capabilities. For the most stringent security and compliance requirements, Oracle also offers customers the ability to run ADW in their own data centers with the Exadata Cloud@Customer service.
- Oracle ADW is well ahead of Snowflake in mature support of unstructured data, as well as a broader set of data models. Our research finds that, compared to Oracle ADW, Snowflake's geospatial only supports one coordinate system and is missing native support for many other unstructured data types such as 3D/LiDAR and raster, as well as missing tools for map visualization and self-service visual analytics. Therefore, Snowflake customers are limited in use cases involving unstructured data and the inherent value in those use cases in terms of identifying new insights and new growth opportunities.
- We found numerous examples of Oracle customers leveraging in-database machine learning and predictive analytics, capabilities that are natively built into ADW and provide immediate value. One customer identified more than $150 million in savings from fraud detection in just the first 3 months using Oracle. Snowflake customers rely on third-party tools for these capabilities, adding both cost and time to realize value.
- Query performance in Oracle ADW was described as "a rocket ship," whereas Snowflake query performance met expectations but required significant optimization by both customer and Snowflake resources. Oracle ADW was 2 to 3 times faster than Snowflake in large deployments with more concurrent users. While Snowflake does automatically scale to accommodate for query performance and higher concurrency, this comes at a cost, as the credit consumption doubles each time it scales to the next cluster size.

> "When we started to test big data warehousing systems, with a lot of concurrent transactions, with a lot of data, we noticed that Oracle was performing approximately 2 to 3 times faster than Snowflake."
>
> Database Practice Lead
>
> Global SI

- Finally, Snowflake customers reported significant effort and trial and error were required for table design and to optimize both the sizing of the data warehouse and the performance of queries. Many also reported that they had to either build their own reporting and monitoring capability or rely on Snowflake support resources to help optimize performance, which led to higher unexpected operational cost of the data warehouse. Snowflake customers reported that the business impact of these costs, particularly in the early stages of deployments, was as high as thousands of dollars' worth of credits because of suboptimal table design and poor SQL query performance.

Introduction

Cloud data platforms promise a unified set of services that enables analysis and actionable insights while supporting a variety of structured, semi-structured, and unstructured data sources. The platforms include data engineering—the ingestion and transformation of data—and use cases for data marts, warehouses for analytics and reporting, data lakes for exploration, and machine learning to support data science operations.

An effective cloud data platform offers the best of traditional data warehouse capability combined with the power, agility, elasticity, and lower cost of the cloud. As a result, a cloud data platform should adhere to the following principles:

1. Offers a complete solution for a single platform to serve a variety of use cases.
2. Provides strong security to protect sensitive data and assess risks.
3. Provides high performance and scale for high concurrence and highly complex queries.
4. Minimizes the effort and cost involved in table design, performance tuning, and performance monitoring.
5. Provides analytics for not only historical data but also predictive insights.

In this white paper, Pique Solutions evaluates Oracle ADW and Snowflake cloud data platforms, and how they align with these principles. Specifically, we look at customer and implementation partner experiences based on the typical activities involved in the life cycle of a data warehouse implementation:

- ⊕ Provisioning an account
- ⊕ Loading data using native tools
- ⊕ Table design and performance tuning
- ⊕ Securing data
- ⊕ Running highly concurrent and complex queries
- ⊕ Sizing and scaling the data warehouse
- ⊕ Data visualization and self-service analytics

Study Approach and Methodology

The primary research phase consisted of an in-depth data collection and multiphase interview process. Pique Solutions identified and qualified 12 customers and partners involved in implementations inside mid-size and large organizations. These experts also provided detailed primary research and data.

The research and methods comprised the following:

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

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

Table 1. Companies and Vendor Solutions Included in Primary Research

Company	Title
Healthcare Provider	VP, Business Innovation
Retailer	Snowflake Solution Architect
Global Auto Rental Brand	Director of Business Intelligence (BI) and Analytics
Auto Manufacturer	Snowflake Architect & Developer
State College/University	Enterprise Architect
Global Systems Integrator	Practice Lead, Data Warehouse and Database
IT Services Firm	Snowflake Consultant
Retailer and Finance Services	Data Management Consultant
Fortune 500 Company	Data Engineer
Global IT Services Company	Data Warehouse Solution Architect
National Health System	Head of Information Services
Entertainment and Sports Information Data Provider	Chief Data Officer

Customer Experiences with Oracle ADW and Snowflake Services

This section summarizes what customers of both Oracle and Snowflake shared relative to the vendors' overall approaches. It also provides a comparison of the two offerings based on customers' experiences performing through the life-cycle activities of a data warehouse deployment.

A Complete and Comprehensive Cloud Data Platform Approach

To start, it is important to have some context related to the commercial and technical approach for both Oracle and Snowflake, when it comes to broad support for a variety of workloads, including data warehouse, data lake, and data mart deployment use cases. Functionally speaking, this progresses from data engineering, to storage, computing, and querying a variety of data types, and finally to the end-user analytics and insights.

Customers we interviewed found value in Oracle providing an optimized cloud infrastructure, powerful database capabilities, data-loading tools, and analytics all from a single vendor. Oracle ADW provides the "best of both worlds" in terms of the benefits of its cloud architecture while also retaining the power and engineering of decades of advanced database capabilities. Furthermore, customers also reported they still retained the flexibility to use popular third-party tools.

> "Snowflake is a product not a solution. It just serves as a repository for data and running queries. You need to use third-party ETL tools, and third-party BI tools to draw insights. Furthermore, you also rely on a separate relationship with cloud infrastructure providers. The cost of ownership doesn't sink in until well into the implementation when customers realize they need all of these third-party products."
>
> Data Engineer
>
> Fortune 500 Company

By contrast, Snowflake is not a complete solution and requires a multi-vendor approach. Snowflake's approach was to build a specialty cloud data platform, whose core data warehouse features are still maturing. Some customers commented that enterprise data warehouse features typically take years to perfect. Moreover, with a relatively narrow product focus, Snowflake provides the data warehouse but leaves it up to the customer to purchase and implement the other solution components they use. This means that customers must manage multiple vendor relationships while also managing an inherent lack of a complete integration of features, coordinated testing, and performance optimization. Snowflake customers did find benefits in their Snowflake deployments, particularly for net new use cases and the ability to support a mix of cloud infrastructure providers, but when it came to migrating existing, larger on-premises data warehouses, they found the reengineering effort to be significant.

As described by a data engineer for a Fortune 500 company, "Snowflake is a product not a solution. It just serves as a repository for data and running queries. You need to use third-party ETL tools and third-party BI tools to draw insights. Furthermore, you also rely on a separate

relationship with cloud infrastructure providers. The cost of ownership doesn't sink in until well into the implementation, when customers realize they need all of these third-party products."

In the following sections, we summarize our research on the deployment life-cycle journey—the steps a typical customer takes in a data platform deployment—to better understand the differences between Oracle ADW and Snowflake. Our research demonstrates that certain activities are easy in a cloud data warehousing environment, such as standing up an account and loading modest datasets. However, Oracle ADW also proves itself as a modern data platform that provides more advanced capabilities across the solution stack, such as security and support for a broader range of enterprise use cases.

Provisioning an Account

For both Oracle ADW and Snowflake, standing up a net new cloud data warehouse and loading data to enable queries were considered easy and straightforward activities by the customers we interviewed, but many of them had an existing data warehouse or other on-premises data they wanted to integrate with.

Standing up a data warehouse using Oracle ADW was described by the director of BI and Analytics for a global auto rental brand as a simple process: "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."

"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."

Director of BI and Analytics

Auto Rental Company

Similarly, provisioning a new Snowflake account, loading initial data via cloud storage, and running queries also took just a matter of minutes. Snowflake customers also reported that they liked the ability to share data easily across teams. A Snowflake architect and developer for an auto manufacturer shared that "Snowflake gives you the ability to clone data to provide for different groups' analytics without impacting storage."

That said, for many enterprises, account provisioning is a very small step in the ultimate journey of data warehousing. Most customers we interviewed experienced some level of data migration, from either on-premises databases or traditional data warehouses.

This is where we found a significant difference between Oracle ADW and Snowflake, with customers in our study citing the ease of migrations with Oracle ADW as compared to Snowflake, which often required a significant reengineering effort. Oracle provides the best path by taking advantage of a combination of existing database technology and advanced cloud infrastructure designed for high performance.

As an example, a Snowflake solution architect for a U.K. retailer said that moving their on-premises MPP-style data warehouse to Snowflake was not a simple lift-and-shift but, rather, required a complete reengineering because of how Snowflake structures and stores the data. He shared, "Because the architecture of Snowflake is so fundamentally different, it really doesn't make sense to lift and shift from one platform to another." The result was more than 6 months

of effort to transition to Snowflake and, based on the 30-person transition team dedicated to the deployment, more than $1.6 million of investment in labor cost alone.

Another customer we interviewed was a healthcare provider wanting to implement a modern cloud data platform to replace the one provided by their key electronic medical records partner. After an evaluation of Oracle ADW versus Snowflake, the vice president of business innovation realized, "We would have had to make a massive change moving from Oracle database to Snowflake, whereas with Oracle ADW we pretty much kept the exact same data model. We tweaked it slightly, which caused a little bit of change. But theoretically you would not have had to change anything. You could have dropped it in ADW from an 11g database and all your reporting would have worked."

> "We would have had to make a massive change moving from Oracle database to Snowflake, whereas with Oracle ADW we pretty much kept the exact same data model."
>
> VP, Business Innovation
>
> Healthcare Provider

Loading Data Using Native Tools

Irrespective of on-premises migrations, the common step after provisioning an account is loading data. For almost all deployments, this means structured data, but increasingly it includes unstructured and multimodal data.

With Snowflake, most customers we interviewed acknowledged that, while Snowflake does offer some capabilities for loading data using copy commands or Snowpipe, they use third-party ETL tools for loading data into Snowflake. They also shared that Snowflake is better suited to address batch data than real-time connectivity. With respect to third-party tools, the Snowflake architect and developer for the auto manufacturer stated, "We bought Talend after we implemented Snowflake and then we found we needed AutoSys for scheduling jobs. These are all add-ons to Snowflake." With Talend costing roughly $12,000 per user per year, this adds a significant amount to the data platform cost of ownership.

The other point brought up by customers was related to the performance and cost of loading data into a single Snowflake data warehouse used for both ingestion and query activity. The way Snowflake works with clustering and autoscaling via virtual warehouses impacts the choice of sizing the data warehouse. The customers who needed rapid ingestion of data based on a larger-sized compute service found it best to split their Snowflake deployment into two separate warehouses, one with a large size for data ingestion and another smaller-sized, more dynamic warehouse for query and reporting. The result was an increase in Snowflake compute costs by having two separate instances.

In contrast, loading data into Oracle ADW and ingesting large volumes of data in small windows was a significant benefit cited by the Oracle customers. They appreciated the ability to use native Oracle tools to achieve this in both batch and real-time use cases. Oracle provides mature data integration offerings, including Data Pump, Data Integrator, and Golden Gate.

> "So we were able to load into Oracle ADW approximately 500 million rows in less than 3 minutes on average. You can't really do that in Snowflake or in Redshift—It would be simply impossible."
>
> Database Practice Lead
>
> Global Systems Integrator

Furthermore, due to the optimization of Oracle ADW and the underlying cloud infrastructure, loading data into Oracle ADW was extremely fast. For example, a database practice lead for a global systems integrator reported that there is a significant performance difference between Oracle ADW and Snowflake as deployments get larger: "So we were able to load into Oracle ADW approximately 500 million rows in less than 3 minutes on average. You can't really do that in Snowflake or in Redshift—It would be simply impossible."

Another key use case involves the addition of unstructured and multimodal data to the data warehouse. Companies are doing this to augment more traditional data models to understand more connected and complex relationships for areas such as geographical data analysis, fraud detection, recommendation engines, and customer insights.

Oracle ADW supports unstructured data, including graph and spatial, to augment structured and semi-structured data. Specifically, Oracle Database Spatial supports comprehensive data types, services, and tools for spatial data management, analytics, and visualization—including points, lines, polygons, geofencing, vector, 3D/LiDAR, raster, topologies, networks, map visualization, geocoding, routing, and Spatial Studio (a self-service visual analytics tool).

A European energy production company shared their experience using Oracle to streamline the collection and processing of data and documentation for water and electrical pipe networks in their mining region. This required migrating large volumes of spatial data and nonspatial functional data from the legacy application within a short time frame to avoid hindering daily mine surveying and basic spatial data processing. The manager of mine surveying summarized, "We have relied on Oracle Database for many years for storing our functional data. When it came to optimizing spatial data management, we found that the OGC-compliant data format offered by Oracle Spatial and Graph, combined with its powerful map engine for web visualization, was the way to go. We don't know any other solution in the market that would allow us to do that."

Snowflake supports geospatial data with a unique data type that uses a round earth coordinate system. The efficacy of this capability is not yet clear, as it depends on third-party tools to support the data type for data processing, analytics, and visualization, and customers would likely need to find a way to rearrange or migrate their existing data to take advantage of the new geo data type. Snowflake's geospatial also lags behind Oracle in support of other unstructured data types such as 3D/LiDAR and raster, as well as missing tools for map visualization and self-service visual analytics. As a result, Snowflake is limited in its ability to support applications that require unstructured data and the inherent value in those use cases in terms of identifying new insights and new growth opportunities.

Table Design and Performance Tuning

Another key area of contrast in customer experience between Oracle ADW and Snowflake is table design and performance tuning operations.

Oracle ADW customers shared that little to no effort was required for table design and tuning. Specifically, auto-tuning was found to be a significant benefit for Oracle ADW customers. The director of BI and Analytics for the global auto rental brand explained, "I will say the one key attribute is auto-tuning, which means our DBAs don't have to spend time tuning the database or tuning queries; Oracle is doing that for us. So, for us, it's a big revolution."

The global systems integrator shared that the machine learning–enabled automation with Oracle ADW makes a big difference in both performance and enabling lesser-skilled resources to implement data warehouses. According to their database practice lead, "Nearly everything has been now embedded into the [ADW] database and it really makes an impressive impact on productivity and efficiency. It is fair to say someone who has only a few weeks or few months of database education can very quickly become productive."

"Nearly everything has been now embedded into the [ADW] database and it really makes an impressive impact on productivity and efficiency. It is fair to say someone who has only a few weeks or few months of database education can very quickly become productive."

Database Practice Lead

Global Systems Integrator

Customers who used or evaluated Snowflake told us that Snowflake requires manual intervention and significant trial and error to optimize the data warehouse and query performance. Optimizing query performance requires prescriptive table design, user mapping, and query patterns that are often unknowable. Snowflake has historically lacked tooling to show the performance of a query before running it in production.

"With Snowflake you have to ensure you have good queries and good table structures. That is where we are still struggling and where I spend most of my time."

Snowflake Architect and Developer

Auto Manufacturer

The Snowflake architect and developer for the auto manufacturer told us, "With Snowflake you have to ensure you have good queries and good table structures. That is where we are still struggling and where I spend most of my time. We have weekly sessions where we go through code and discuss queries and table structures."

Beyond table design, he also shared challenges with Snowflake clustering and configuration. He stated, "Initial clustering can take a while. It can take 12 to 15 hours of credits to auto-cluster a 1-TB table." As a result, he recommended auto-clustering at the individual

table level. Furthermore, when it comes to optimization, he emphasized, "You really have to understand how your data is being used by the business and how your data is laid out. It is a case-by-case, table-to-table consideration." In the end, he said that ultimately he ended up partitioning users/user groups to their own tables.

Customers raised several other points related to table design and optimization. Snowflake does not provide monitoring reports that enable companies to optimize their data warehouse. Several customers cited the need to create their own reports using third-party reporting tools (e.g., Tableau) to facilitate this. Other customers, particularly those with larger deployments, cited the need for Snowflake engineers to assist with identifying costly queries and other performance-related matters.

Finally, there is a significant learning curve for optimization of Snowflake beyond a single data warehouse deployment. A database consultant who worked on numerous Snowflake deployments across retail and financial services companies told us that he learned many lessons via trial and error over multiple deployments.

The business impact of this for Snowflake customers, particularly in the early stages of deployments, is numerous examples of burning through thousands of dollars' worth of credits per month as a result of suboptimal table design and SQL queries.

Securing Data

Securing data in a cloud data warehouse is critically important for many enterprise use cases to ensure regulatory compliance and protect sensitive corporate and customer data.

Both Oracle ADW and Snowflake provide for auto-encryption of data, but our study participants cited the strength and maturity of Oracle database security features, as well as the unified and integrated security across the entire stack of services from applications, to database, to infrastructure. Specifically, companies we talked with touted the Oracle ADW's Data Safe capability along with mature database security features such as Database Vault, Privileged Analysis, and Real Application Security.

The vice president of BI for the healthcare provider expressed concerns with security in a multivendor environment such as that when using Snowflake, third-party tools, underlying cloud infrastructure, and so forth. Regarding Oracle ADW, he concluded, "They've got literally the whole entire stack from top to bottom, only Oracle and not anybody else. So, from a security perspective they have all these extra wrapper tools like Oracle Data Safe for evaluating risk, monitoring, and intrusion detection and everything else that goes with that."

"They've got literally the whole entire stack from top to bottom, only Oracle and not anybody else. So, from a security perspective they have all these extra wrapper tools like Oracle Data Safe for evaluating risk, monitoring, and intrusion detection and everything else that goes with that."

VP, Business Innovation

Healthcare Provider

In contrast, while most Snowflake customers were satisfied with the basic security features, several shared the need for costly VPC connections and intimated that Snowflake was a "work in progress" as it related to advanced security features such as subsetting, data masking, and redaction.

Snowflake does offer a dynamic data-masking capability that provides the ability to define policies that allow for full or partial field redaction based on role access. Snowflake also offers integration to external tokenization services to augment the security controls. While this feature is important for protecting customers' sensitive data, Snowflake still lacks many enterprise-class governance capabilities such as row-level security, label security, sensitive data discovery, user risk scoring, fine-grained access controls and auditing, and database firewalls.

Snowflake also lacks in support of the most stringent security and compliance scenarios where, due to data sovereignty or data residency requirements, customers must retain data on-premises. Oracle supports this by giving customers the ability to run ADW in their own data centers via the Exadata Cloud@Customer service using the same database, analytics services, and tools as in the cloud.

Querying the Data Warehouse

After loading and securing data, the primary function of a data warehouse is performing and executing queries, which vary from those that are common and run frequently by the user community to those that require more sophisticated logic to draw unique insights. Our research investigated both types, as well as the customer experience related to concurrent and complex queries. This is another area where Oracle ADW customers find advantages over their counterparts using Snowflake.

Running Concurrent Queries

Study participants related that Oracle ADW has higher concurrency, as there is no limit on the concurrency levels of ADW because of its highly parallelized and scale-out architecture. ADW improves concurrency and parallelism by several orders of magnitude with scale-out compute, networking, storage, automated Database Resource Management, I/O Resource Management, and Network Resource Management.

In contrast, although Snowflake provides several warehouse sizes (e.g., X-Small, Small, Medium, Large), it advises the maximum per cluster is 8 concurrent user sessions or queries (soft limit), and its performance will start to degrade as the number of concurrent user sessions or queries increases beyond 8. To handle more concurrent user sessions or queries, Snowflake recommends that customers spin up additional clusters. While this could be considered an autoscaling advantage, the additional clusters will cost more in credits. Due to the doubling in credits consumed for each larger Snowflake warehouse size, this impacts the cost of scaling, as each time Snowflake scales it doubles the cost even if just an incremental amount of compute is needed.

If the queries are written poorly, which our research discovered was the case in many instances, this compounds the concurrency issue. The Snowflake solutions architect for a retailer told us, "The performance is only as good as someone who can understand how the product works. For example, if you run a very bad query, Snowflake is not going to fix your query. Snowflake will still try to run that query and burn credits by auto-scaling. So, to maximize performance the key is still to understand your data."

> "If you run a very bad query, Snowflake is not going to fix your query. Snowflake will still try to run that query and burn credits by auto-scaling."
>
> Snowflake Architect
>
> Retailer

Oracle ADW customers shared their experiences of a high-performance cloud data platform, with the advantage of columnar data in the cloud and leveraging the power and performance of Oracle infrastructure optimized for large datasets. The database practice lead of the global systems integrator shared, "When we started to test big data warehousing systems, with a lot of concurrent transactions, with a lot of data, we noticed that Oracle was performing approximately 2 to 3 times faster than Snowflake. We were measuring several reports in the real data warehousing environment."

> "When we started to test big data warehousing systems, with a lot of concurrent transactions, with a lot of data, we noticed that Oracle was performing approximately two to three times faster than Snowflake."
>
> Database Practice Lead
>
> Global Systems Integrator

The vice president for the healthcare provider also was impressed, stating, "With ADW, it's been like a rocket ship. I mean, we are talking drastically faster. We are talking 2 to 3 times faster on average. Some of our processes that took an hour now take 5 to 10 minutes."

Finally, 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-prem you are constrained by the physical size and when you run out of space it's hard to scale, so with the Autonomous Data Warehouse to be able to scale so elastically is phenomenal."

Running Complex Queries

When developing and executing more complex queries, Oracle provides more advanced capabilities compared to Snowflake, which currently lacks support for proven database features such as joins, stored procedures, and triggers that are found in a traditional data warehouse. Snowflake acknowledged this shortcoming in their recent platform announcement and will be providing support for SQL stored procedures in the near future, but it remains only one feature, as compared to the many database capabilities that Oracle already provides.

The result of these limitations, according to Snowflake customers we talked to, was more complex query coding that ultimately led to additional computing resources to accommodate

for the shortcomings. As indicated in the earlier section on table design, Snowflake customers often depend on Snowflake technical support consultants to help them optimize complex query performance, as it is challenging for customers to understand what is happening under the covers of Snowflake.

The Snowflake data architect for the retailer explained this query complexity in terms of the difference between their on-premises MPP-style data warehouse and what they need to do in Snowflake. He said, "We have stored procedures in our on-premise data warehouse to move data from one layer to another. In Snowflake, we have to write everything in SQL to move layers and queries. The result is that you have to break up these complex transformation steps or else the SQL logic becomes 200 to 300 lines."

The database practice lead for the global systems integrator shared his view of what he called the "cloud pricing trap" with Snowflake. "It's when you have a tool with a normalized data model, then, of course the application will be doing a lot of joins. So, what happens with Snowflake is, it redistributes between the nodes and, as a consequence, you pay a lot for compute. Further, the second normal form data model leading to so many joins will probably come to not using the database features, which of course is exactly the opposite of Oracle, where Oracle is using a lot of data-based features."

Materialized views were another aspect that Snowflake customers pointed out relative to complex query performance. Those who had used the capability found it created performance issues and, as a result, they avoided using the feature altogether. The Snowflake solution architect and developer for the auto manufacturer told us, "My personal recommendation is not to use [Snowflake's] Materialized Views: Due to auto-clustering we found we ended up paying twice the cost."

"My personal recommendation is not to use [Snowflake's] Materialized Views: Due to auto-clustering we found we ended up paying twice the cost."

Snowflake Solution Architect and Developer

Auto Manufacturer

Sizing and Scalability

Efficient query performance ultimately relies on sizing and scaling of the data warehouse. Both Oracle ADW and Snowflake customers lauded the ability to scale compute and storage independently, both up and down. Horizontal scaling, or scaling out, can be done in an automated fashion using autoscaling.

Oracle customers we interviewed touted the ability to scale ADW up and down very easily and with the flexibility to scale storage and compute granularly. Study participants told us they were able to start with a small configuration of just a few OCPUs and scale, as needed, to as many OCPUs as they needed. The inherent performance of ADW resulted in fewer scaling activities for data ingestion and queries.

Snowflake does let customers scale storage and compute separately, but customers cited some inherent challenges based on their experience. Snowflake data warehouse sizing is based on

fixed “T-shirt sizes” that double for each larger size. Due to the performance considerations, many customers we spoke to divide the data warehouse into two, one larger one for data loading and another smaller one for analytics. With the smaller size warehouse, they would start small and let the autoscaling adjust to attempt to optimize query performance. The business impact of this was an increase in compute costs by having two instances of Snowflake potentially running at the same time.

Moreover, customers mentioned that scaling out Snowflake requires a lot of trial and error to get the T-shirt size right relative to the clustering and autoscale. Each virtual warehouse has a step function of scale. Clusters must be the same size, which, particularly in early stages of deployment, companies found burned through a lot of credits. The Snowflake consultant for an IT services firm with experience across multiple pharmaceutical companies explained, “It’s a long learning curve to determine how to best size and scale Snowflake, and it requires expertise learned over several deployments. You have to determine whether you utilize parallel processing on a small size cluster or increase the size of the data warehouse to optimize query performance. Snowflake doesn’t provide any tools to help with this.” He went on to say that this can lead to overprovisioning and added costs, citing examples of single queries “burning through thousands of credits.”

“It’s a long learning curve to determine how to best size and scale Snowflake, and it requires expertise learned over several deployments. You have to determine whether you utilize parallel processing on a small size cluster or increase the size of the data warehouse to optimize query performance. Snowflake doesn’t provide any tools to help with this.”

Snowflake Consultant

IT Services Firm

Relative to storage size and cost, Oracle customers cited the benefits of various choices for advanced compression to significantly reduce the amount of storage required. The practice lead for Data Warehouse and Database for the IT services firm found compression as a key difference between Oracle and Snowflake that results in larger storage requirements for Snowflake customers. Also relative to storage cost, Oracle ADW includes 60 days of backup at no additional charge.

Snowflake customers shared several examples of excess storage needs. One area that came up in several interviews was the impact on storage when using the features for Time Travel and Fail-safe. Customers reported that this can add significant size and cost to the Snowflake data warehouse deployment. According to the Snowflake solution architect for the retailer, their 600 TB of core data results in a 1-PB investment when factoring in Time Travel and Fail-safe. In his words, “You have to cater to Time Travel and Fail-safe for data backup.”

Another customer shared the same experience, finding that the way they loaded their data resulted in a huge impact on storage. The Snowflake data management consultant to retail and financial services companies told us he had set the Time Travel retention period to 90 days and found that with a base table size of 10 GB Time Travel expands to 5TB, or 200 times the footprint of the original table size. He found that it had to do with data loading, telling us, “It has

to do with how we load data into a table. If there's lots of inserts and updates, copies of data are created across partitions and clustering, which takes up humongous space. Nobody thinks about the load strategy; we never thought truncate and load would cause such a cost impact." The resulting business impact led them to shorten their retention period to 30 days, thus limiting their analysis horizon.

Insights and Analytics

The final area where there are substantive differences between Oracle ADW and Snowflake relates to how the data is used by the various groups within a company to visualize data and draw actionable and predictive insights. As discussed earlier, Snowflake relies on third parties for data analysis and advanced analytics, whereas Oracle is an end-to-end solution that includes native tools for self-service analytics and data analysis. While Oracle customers can leverage third-party tools for analysis of data resided in Oracle ADW, many we talked with view the single-vendor approach as an advantage unique to Oracle. This advantage was evident from many perspectives, including coordinated tests, integration, support, and ultimately lower cost of ownership.

Put simply, Snowflake's primary function is a query engine, and they leave the analytics, reporting, visualization, and augmented analytics up to customers to stitch together third-party tools. This results in numerous separate purchases and implementations, as well as coordination of support. These tools are also typically licensed on a per–named user basis and, with a cost on the order of $1,000 per year per user, can quickly become a significant portion of cost of ownership when users are in the hundreds. Some customers we interviewed acknowledged that this approach added significantly to the overall cost of ownership. Somewhat ironically, customers shared that they even needed to use third-party tools to build dashboards to monitor Snowflake usage data. With the release of Snowsight, Snowflake appears to be providing some level of their own insights and analytics; however, they acknowledge it publicly as a complement to the more robust third-party reporting and analytics tools. This will be a confusing message for Snowflake customers who are conditioned to using third party tools.

> "One of the limitations of Snowflake is it doesn't have an end-to-end solution with its own products and needs to rely on partners to deliver required capabilities for data analysis, reporting, and visualization."
>
> Data Warehouse Solution Architect
>
> Global IT Services Company

According to a data warehouse solution architect for a global IT services company, "One of the limitations of Snowflake is it doesn't have an end-to-end solution with its own products and needs to rely on partners to deliver required capabilities for data analysis, reporting, and visualization."

Many customers cited the benefits of Oracle ADW and Oracle Analytics Cloud (OAC) providing a complete, single-vendor solution for data warehousing, data marts, and data lakes, plus the tools for self-service analytics. OAC provides data analysts and data scientists with the

capabilities for self-service visualization, enterprise reporting, what-if modeling, and even mobile analytics.

The vice president of BI for the healthcare provider, who uses Oracle ADW with a mix of Tableau for general users and OAC for financial planning and analysis, shared their vision of adopting OAC more widely: "OAC is much richer than Tableau. It is better integrated, with better dashboards and a more robust security model. It is really the difference between an enterprise application and a desktop application, to be perfectly honest with you. And then there's pricing. Tableau is really expensive because you have to do it on a per-seat basis, whereas with OAC, there are new models where you can base it upon CPU size. You can have unlimited users so you can scale it much better."

"OAC is much richer than Tableau. It is better integrated, with better dashboards and a more robust security model. And then there's pricing. Tableau is really expensive because you have to do it on a per-seat basis, whereas with OAC, there are new models where you can base it upon CPU size. You can have unlimited users so you can scale it much better."

VP of Business Innovation

Healthcare Provider

Several survey participants expressed that they leveraged the modern data platform to generate predictive insights to improve business performance. They found the embedded Oracle Machine Learning, whose large library of mathematical and statistical algorithms and machine learning SQL functions allow users to automate their discovery of new insights, generate predictions and add "AI" to data.

For example, a European national health system shared their experience using Oracle's machine learning and advanced analytics algorithms to improve patient care and drive cost savings via fraud identification. They found fraud prediction simplified from an automated in-database analytical methodology and, as a result, they were able to identify $156 million in savings in just the first 3 months of operation. The head of information services for the national health system shared the experience of building their data lab to find savings and cost reductions in their overall healthcare budget, citing, "With one vendor providing the whole solution, it is very easy for us."

Another example we found harvesting value from Oracle's machine learning was a consulting company that uses big data analytics to help entertainment companies and professional sports teams improve the experience for their fans by better understanding their behavior. According to the chief data officer, "Oracle gives us everything we need in terms of machine learning and predictive analytics that allow us to do in-database marketing. In-database marketing is very effective when you have all the algorithms you need, which leads to a lot of savings in terms of time and performance because your data is already in Autonomous Data Warehouse."

In the case of Snowflake, customers were required to purchase separate advanced analytics and machine learning service from a third party, which increased the cost of ownership of their data platform, required them to work with multivendor support, and resulted in a longer time to value for predictive insights.

Conclusion

Pique Solutions research found that an effective modern cloud data platform includes not only the expected benefits of the cloud but also the mature and proven capabilities of core database technologies. Standing up a new cloud data warehouse, loading basic structured data, and performing a few simple queries were found to be easy and fast activities for both Oracle ADW and Snowflake customers.

But when data warehouse deployments became more complex—in terms of advanced use cases, volume, and scale—Snowflake capabilities were found to be lagging behind those of Oracle ADW. Customers shared that Oracle provides a more comprehensive, enterprise-oriented solution with a broader set of advanced and proven capabilities. This resulted in optimized performance with less effort, more comprehensive security and governance, and support for more data models and advanced analytic workloads. Snowflake customers cited a roadmap of future capabilities, but they also said there is an inherent learning curve to compensate for feature maturity through manual effort, trial and error, and the incorporation of a variety of third-party tools.

Many Snowflake customers shared examples of unexpected costs incurred due to a lack of native built-in tools combined with limited visibility and understanding of how Snowflake works under the covers. These examples include the following:

- Third-party tools for ETL, self-service reporting and visualization, and predictive analytics added up to hundreds of thousands of dollars per year for deployments with hundreds of users.
- “Bad queries” individually burned through thousands of credits that, collectively, added thousands of dollars to the monthly bill for large deployments.
- Scaling up in Snowflake doubles the cost for each new warehouse size. Scaling out replicates the original size so, even if a customer slightly exceeds their concurrency limitations, it results in a doubling of their credit consumption.
- Customers chose to limit their historical data analysis horizon based on storage cost issues related to the way data is loaded into Snowflake and the impact of enabling Time Travel and Fail-safe. As a result, one Snowflake customer cut their analysis horizon by 67% due to high cost of storage. Oracle ADW includes 60 days of data retention and storage at no additional cost.
- Lastly, and a stark reality for many enterprise customers, Oracle was found to enable easier migration from, and integration with, on-premises data warehouses. One retailer spent more than $1.6 million in migration costs with Snowflake in trying to recreate the capabilities and data from a large on-premises MPP-style data warehouse deployment.

Payments to Snowflake will represent a fraction of the overall cost of ownership of their cloud data platform when factoring in costs for infrastructure, third-party tools, manual effort for performance optimization, unforeseen compute and storage fees, and migration or reengineering costs. As a result, the cloud data platform investment is much clearer and significantly lower for Oracle customers vis-à-vis their Snowflake counterparts.