RESEARCH NOTE
PROGRAM: INFRASTRUCTURE
DOCUMENT NUMBER: U138 SEPTEMBER 2020

NUCLEUS RESEARCH

REDEFINING DATA WAREHOUSING WITH SNOWFLAKE

ANALYSTS
Nick Grizzell, Daniel Elman

THE BOTTOM LINE

Nucleus evaluated several Snowflake customers to understand the value delivered from Snowflake's advanced cloud-based data warehouse solution. Snowflake provides data warehousing for data analytics, and it differentiates itself as operating entirely on a cloud infrastructure with no hardware or software for users to install, configure, or manage. Snowflakes architecture takes a hybrid approach utilizing shared-disk database architectures and shared-nothing database architectures to achieve data management simplicity without sacrificing scalability or performance.

OVERVIEW

Snowflake is a cloud data platform connecting data sources with data consumers through a range of workloads, including data engineering, data lakes, data warehouses, data science, data applications, and data exchange. A data source can represent IoT, enterprise applications, databases, or third-party applications. In contrast, data consumers are the products that deliver value, such as operation reporting, ad hoc analysis, data monetization, and real-time analytics. Snowflake enables users to build a modern data architecture to reduce the complexity and cost of adopting a cloud data platform. With Snowflake, there is no hardware, virtual or physical, for a user to select, install, configure, or manage. Furthermore, there is no software for the user to install, configure, or manage, and Snowflake handles all ongoing maintenance, management, and tuning.

Snowflake is cloud-agnostic, and the platform available as-a-service on multiple cloud providers, including Google Cloud, AWS, and Azure allowing users to transition to the solution seamlessly. A key differentiator for Snowflake is its architecture as the platform is built entirely for the cloud allowing for centralized storage giving users access to unlimited amounts of structured and semi-structured data. Additionally, the platform is built for multi-cluster compute for running multiple workloads and enables cloud services to automate administrative and database tasks surrounding transactions, security and governance, and metadata.

Snowflake's architecture is a hybrid of shared-disk database architectures and shared-nothing database architectures. The architecture uses a central data repository for persisted data that is accessible from all compute nodes in the data warehouse and gives users the data management simplicity of a shared-disk architecture with the performance of a shared-nothing architecture. Additionally, Snowflake processes queries using massively parallel processing (MPP) compute clusters where each node in the cluster stores a portion of the entire data set locally. Snowflake's architecture consists of three key layers involving database storage, query processing, and cloud services. The underlying file system in Snowflake is built on Amazon S3, where data is encrypted, compressed, and distributed to optimize performance. Snowflake manages all aspects of how the data is stored from the organization, file size, compression, metadata, and statistics. For query processing, Snowflake enables virtual warehouses to act as compute clusters in EC2, which can be used to load data or run queries concurrently. The virtual warehouses can be scaled up or down on-demand as well as paused to reduce compute spend.

KEY BENEFIT AREAS

Nucleus identified the following specific benefits from the experiences of Snowflake customers. Companies realized benefits across multiple segments, including reduced project development time, reduced report lifecycle time, reduced maintenance and monitoring costs, improvements to customer insights, reduced manual processes, scalability, and enterprise-scale performance.

Reduce time to offload ETL jobs by 83 percent with Snowflake's optimized architecture

- Reduced Report Development Lifecycle. Developing detailed reports on customer insights are often lengthy tasks requiring multiple data analysts to extract benefits and insights from customer behavior. With Snowflake's capabilities, users can have truly automated databases without the need to create indices or maintain them. Analysts and managers can spend more time gaining insights into customer behavior and retrieve actual results as opposed to spending time managing and maintain the solution to generate a report. Companies can have all different data sources within one data warehouse, meaning analysts and data engineers can begin looking for links between data sources and disparate business units to tie data together into a measurable and understandable report. One company reduced the time to deliver a report by 95 percent from five days to a matter of hours, allowing teams to redirect attention to more value-driven processes.
- Scalability. For many organizations, growth is a positive factor but can lead to an increase in costs surrounding employees, infrastructures, and maintenance. As the company expands to new landscapes, users will adopt more applications and data management tools, leading to an increase in workflow processes. With one company, Snowflake unlocked untapped potential within sales teams as they could not seek large customers without worrying about the system's ability to scale with the company's growth. The company can now scale up 100 times its current capacity with just a few clicks allowing managers and data engineers to redirect efforts to more value-driven processes within the company. If the company had moved forward with the legacy system, it would have needed to hire ten new operations employees and purchased new equipment. Therefore, we can also attribute hundreds of thousands of dollars in savings associated with avoided hires and avoided hardware costs to the Snowflake deployment.
- Increased Employee Efficiency. The Snowflake cloud data platform eliminates hardware and software installation needs while also handling ongoing maintenance, management, and tuning. Additionally, the platform supports automated processes,

unlimited scalability, zero-maintenance, and zero-monitoring to allow users to redirect efforts within a company to drive projects and revenues. The projects associated with extracting customer insights are often too time-consuming and requiring frequent reworks and updates to stay on top of continuing trends and customer behaviors. Efficiency is key for these projects, and developers need every tool on hand to not only help maintain current projects but also move on to new ones. The platform enabled one company to implement accelerated querying to the business structure to quickly gain insights into the current data set. The data engineers could now focus more on development rather than query optimization due to Snowflake's out-of-the-box performance. The time to offload ETL jobs was reduced by 83 percent from three hours to just under 30 minutes.

Scale operations 100 times faster and save hundreds of thousands in avoided hires

CUSTOMER PROFILES

CLOUD COMPUTING COMPANY

The company is an enterprise-level cloud computing company focused on a content delivery network (CDN), cybersecurity, and cloud services. The company delivers a network of servers around the world to help businesses set up websites by distributing content from locations near the user. After a new acquisition of a company deploying a legacy, on-premise data warehouse, the CDN company looked to improve storage capacity as it rose to unmanageable levels. The company began with 16 database servers and began adding one or two servers sporadically, leading to cost and maintenance issues for the management team.

The company looked towards an affordable and fully elastic cloud data warehouse to allow the company to scale up or down automatically. The company decided to move forward with Snowflake as the solution provided a no-maintenance alternative allowing the company to execute queries at varying performance levels without adjusting the pricing model. With Snowflake, the sales teams within the company found the scalability to be almost freeing in a sense. Sales teams could not seek large customers without worrying about the system's ability to scale with the company's growth. The company can now scale up 100 times its current capacity with just a few clicks allowing managers and data engineers to redirect

efforts to more value-driven processes within the company. If the company had moved forward with the legacy system, it would have needed to hire ten new operations employees and purchased new equipment as well. Therefore, we can also attribute hundreds of thousands of dollars in savings associated with avoided hires and avoided hardware costs to the Snowflake deployment.

DIGITAL ADVERTISING COMPANY

Reduce the report development lifecycle by 95 percent

The company is a small digital advertising company that aims to deliver a real-time bidding platform that surveys website visitors and provide insights to help customers bid on targeted advertising. With the growth of the company, data analysts lost business performance metrics as customer demand was outpacing the capacity of the existing infrastructure. The company sought a solution to provide insights on how its platform performed with existing customers in terms of decision-making processes. The company considered implementing Snowflake as it needed a micro-service architecture to stream and query platform events and near real-time data modeling to understand customer behavior.

After implementing Snowflake, the company soon realized the benefits of an elastic and fully scalable infrastructure that could react to its business needs. Additionally, the company deployed Snowpipe, Snowflake's serverless data ingestion service that handles raw data flows into data warehouses. Within two weeks, the team built the first data warehouse allowing business-level users to extract insights on customer behaviors. Furthermore, the company's engineering team reduced the time to deliver a report by 95 percent from five days to a matter of hours. Every 15 minutes, management can retrieve updated system reports giving account managers a head start on advertising campaigns. Data engineers also used Snowflake to build an A/B testing platform to replace Google Analytics allowing for savings of $13,000 per month.

MARKETING AND ADVERTISING COMPANY

The company is a medium-sized marketing and advertising company centered around improving the consumer lifecycle. The company aims to automate customer engagement through technology and software to drive customer relationships and deliver significant ROI. The platform uses AI to create a personalized customer experience to increase sales and services profits for brands, agencies, and dealers.

The data engineering team within the company noticed increasing stress within systems as the company continued to grow and expand operations. ETL processes and querying

became lengthy tasks as ETL jobs could take multiple hours per day, and this was a nightly process. The company originally deployed Azure SQL, and while searching for alternate solutions, the management team found itself impressed with Snowflake's zero-maintenance infrastructure that offered unlimited scalability. After deploying Snowflake, the data engineering team began using the platform, offload a substantial amount of ETL processes to Snowflake. The platform also enabled the team to implement accelerated querying to the business structure to quickly gain insights into the current data set. Altogether, the data engineers were able to focus more on development rather than query optimization due to Snowflake's out-of-the-box performance. The time to offload ETL jobs was reduced by 83 percent from three hours to just under 30 minutes.

LOOKING AHEAD

Data warehousing, and the larger challenge of data management, are becoming increasingly important and legacy technologies and architectures are struggling to keep up with new demands. Buying into data-driven decision making at the organizational level means that business data from all areas needs to be stored, indexed, and managed in such a way that makes it easily accessible for both regular reporting and ad hoc analytical queries. As a fully cloud-native data warehousing technology, Snowflake is well-positioned to capitalize on the wave of massive enterprise investment in cloud technologies. Further, the addressable market for its technologies spans virtually all industries and sizes of business – any company can benefit from more effective management, and Snowflake offers the scalability to meet any data volume or reporting schedule.

Recently, we conducted an ROI case study with Ewals Cargo Care who modernized their data management and reporting architecture from a legacy Oracle database to be instead built on Snowflake (Nucleus Research *u129 – Qlik ROI case study – Ewals Cargo Care –* September 2020). It dramatically increased the reporting and analytical agility, and was much faster to add new data sources, a must in a modern business ecosystem where companies often have hundreds or thousands of applications producing data and working together. We anticipate that this will become a trend as more businesses look to modernize their architecture to be more agile and cloud-native, with Snowflake currently in the pole position to address the demand.