CABOT PARTNERS
OPTIMIZING BUSINESS VALUE

# Total Value of Ownership (TVO) Assessment Highlighting Advantages of IBM Netezza Performance Server as a Service (NPSaaS) over Snowflake

Ravi Shankar, Ph.D., MBA

Srini Chari, Ph.D., MBA

Inesh Chari

**March 2022**

**Cabot Partners Group, Inc.**

100 Woodcrest Lane, Danbury, CT 06810.

www.cabotpartners.com

info@cabotpartners.com

Sponsored by IBM

# Table of Contents

## Analytics are increasingly moving to the cloud

Analytics and artificial intelligence (AI) solutions profoundly transform how businesses and governments engage with consumers and citizens. Across many industries, high-value transformative use cases are rapidly emerging. The economic impact of just AI is enormous: an estimated US $13 trillion is expected to be added to the global economy in the next decade.[1]

Impact of AI estimated to be $13 trillion

However, for analytics and AI to become an integral part of an organization, numerous deployment challenges with data and infrastructure must be overcome[2], especially as data volumes and variety continue to grow. Enterprise Data Warehouses (EDWs) – a critical component for Enterprise AI – have been used for over 20 years to integrate and analyze data from disparate sources for faster decision-making and improve business processes.

Until recently, EDWs were on-premises systems with fixed capacity for processing and storage that could not scale easily or quickly for spiky workloads. Today, cloud EDWs have replaced some legacy on-premises systems and are appropriate for organizations adopting new EDWs. Figure 1 depicts some key advantages of cloud EDWs:[3]

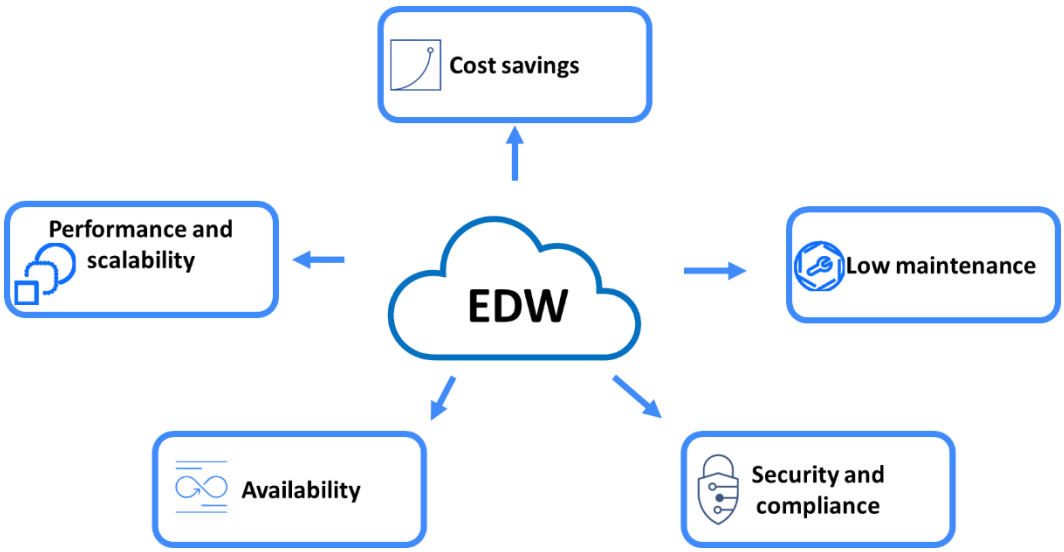



**Figure 1: Key benefits of Cloud Enterprise Data Warehouses (EDWs)**

Cloud EDWs are growing rapidly and deliver performance and scalability, cost savings, low maintenance, security and compliance, and availability

- **Performance and scalability:** Unlike legacy systems, the cloud platform provides the ability to quickly and easily scale up or down to meet nearly any processing demand efficiently with high performance.
- **Cost savings:** Infrastructure is available on a cost-effective subscription-based pay-as-you-go model, and software updates are automatic and included in the subscription.
- **Low maintenance:** The EDW-as-a-service model relieves stakeholders from the difficulties of purchasing, setting up, managing, and upgrading costly on-premises hardware and IT resources.

[1] Tim Fountaine, Brian McCarthy and Tamim Saleh, "Building the AI-Powered Organization", Harvard Business Review, July-August 2019.

[2] Ritu Jyoti, "Accelerate and Operationalize AI Deployments Using AI - Optimized Infrastructure", IDC Technology Spotlight, June 2018

[3] What is an enterprise data warehouse? | Stitch resource (stitchdata.com)

- **Security and compliance:** Handles data security with always-on, end-to-end data encryption and built-in protection against data loss (accidental or malicious). New security threats are mitigated by quickly deploying countermeasures. Can address a variety of compliance standards, such as SOC 1 and SOC 2, PCI DSS Level 1, and HIPAA.
- **Availability:** Built for high availability, across many availability zones or data centers. If a data center goes down, work shifts to another available data center with no user disruption.

Growth in Data Warehouse as a Service is 22.3% annually

Consequently, the global Data Warehouse as a Service (DWaaS) market size (public and private cloud) is expected to grow annually at 22.3%, to reach USD 12.9 billion by 2026 from USD 4.7 billion in 2021[4]. However, not all cloud EDWs are the same, and each has specific capabilities that improve the Total Value of Ownership (TVO) for clients.

## Key differentiators of Cloud EDWs

While there are several Cloud EDWs (Amazon Redshift, Google BigQuery, Microsoft Azure SQL Datawarehouse, Snowflake, and IBM Netezza SaaS) in the market, they differ (Figure 2) in several ways:[5]

Key differentiators of Cloud Data Warehouses (CDW) are performance and scale, pricing, elasticity, ease of use

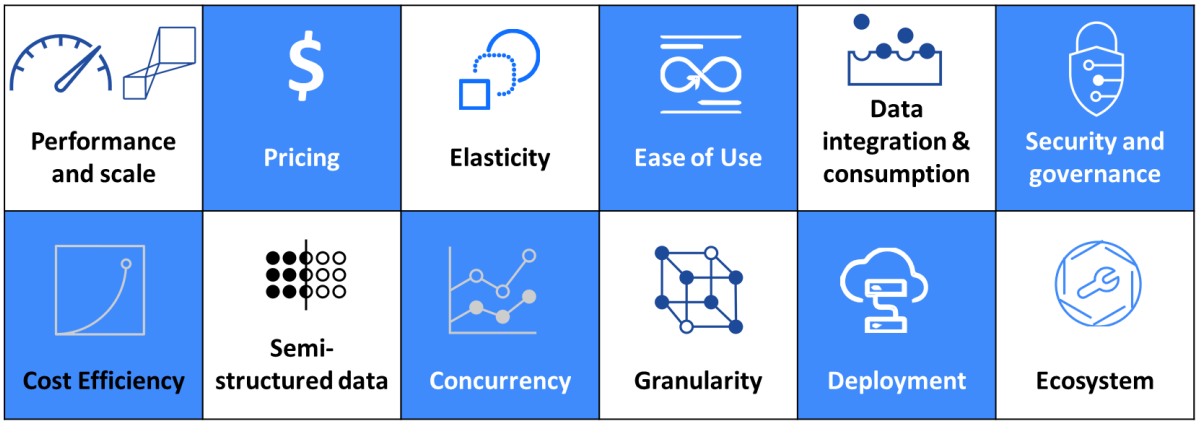



**Figure 2: Key differentiators of Cloud Data Warehouses (CDWs)**

**Performance and scale:** How fast can it run, on how much data, and with how many resources (affecting costs)?

**Pricing:** How is the solution priced? What capabilities does it include? How does the pricing vary with data size, workload volatility, and other resources consumed?

**Elasticity:** How simple is it to allocate different compute/storage resources for various tasks dynamically? Does the pricing follow demand elasticity?

**Ease of Use:** How much time and resources will the user need to spend on scaling and maintenance instead of discovering insights?

[4] https://www.researchandmarkets.com/reports/5403276/global-data-warehouse-as-a-service-market-with?utm_source=GNOM&utm_medium=PressRelease&utm_code=9gj8gf&utm_campaign=1581078+-+Global+Data+Warehouse+as+a+Service+Market+(2021+to+2026)+-+Rapid+Deployment+of+Large-Scale+Cloud+Data+Warehouses+Presents+Opportunities&utm_exec=jamu273prd

[5] https://www.firebolt.io/blog/data-warehouse-requirements-template-for-evaluation

CDWs also differ in data integration and consumption, security and governance, cost efficiency, semi-structured data, concurrency, granularity, deployment, and ecosystem

**Data integration and consumption:** How to handle highly distributed and siloed data? Is continuous ingestion of data supported to enable data freshness? Does it allow self-service data consumption and collaboration?

**Security and governance:** How are these handled? Is it automated?

**Cost efficiency:** What are the customer's data set size and performance requirements today, and what will it be in 1-2 years? Is the data warehouse efficient enough to support future growth?

**Semi-structured data:** Will the user get the same experience as working with structured and semi-structured data? What are the tradeoffs of working with semi-structured data?

**Concurrency:** How many users can query the data? Can queries run simultaneously? What are the future needs?

**Granularity:** Will the user have to sacrifice granularity to achieve performance?

**Deployment:** Is the data warehouse available on the cloud platform used today? How vital is multi-cloud support for the user?

**Ecosystem:** Does the current data stack (ETL tools, BI tools, etc.) integrate easily?

This paper compares Snowflake with IBM Netezza Performance Server as a Service (NPSaaS).

Two CDWs, Snowflake and IBM Netezza Performance Server as a Service (NPSaaS), are compared

## Snowflake

As one of the Cloud EDW providers, Snowflake provides companies the performance, flexibility, and scalability to easily load, integrate, analyze, and securely share, consolidate, and govern data/analytics across data warehouses, data marts, and data lakes. Its architecture separates storage resources from computing resources so that all workloads can simultaneously scale without contention.

Snowflake also enables very efficient collaboration across the enterprise without having to copy/move data often and automating menial data maintenance tasks such as updating metadata and vacuuming. It also provides pricing flexibility by charging for computing and storage separately and offering several sizes of data warehouse resources. It can run on multiple cloud providers — Amazon Web Services, Microsoft Azure, and Google Cloud Platform.

## Netezza Performance Server as-a-Service (NPSaaS)

IBM recently announced the general availability of IBM Netezza Performance Server for IBM Cloud Pak for Data-as-a-Service (NPSaaS) as a fully managed service on Microsoft Azure. NPSaaS is IBM's public cloud EDW. It leverages the same software building blocks that underpin IBM Netezza – the industry's first data warehouse appliance (DWA) designed to run complex analytics workloads on very large data volumes.

Built on an open and extensible cloud-native AI platform (Figure 3), IBM NPSaaS modernizes how businesses store and analyze their enterprise data. It can scale from small to multi-petabyte capacity to accelerate a client's analytics/AI journey.

IBM Cloud Pak for Data and NPSaaS offer a data fabric to automate data discovery, data governance, and data consumption in hybrid multi-cloud environments

NPSaaS delivers unprecedented simplicity, speed, sophistication, and scale

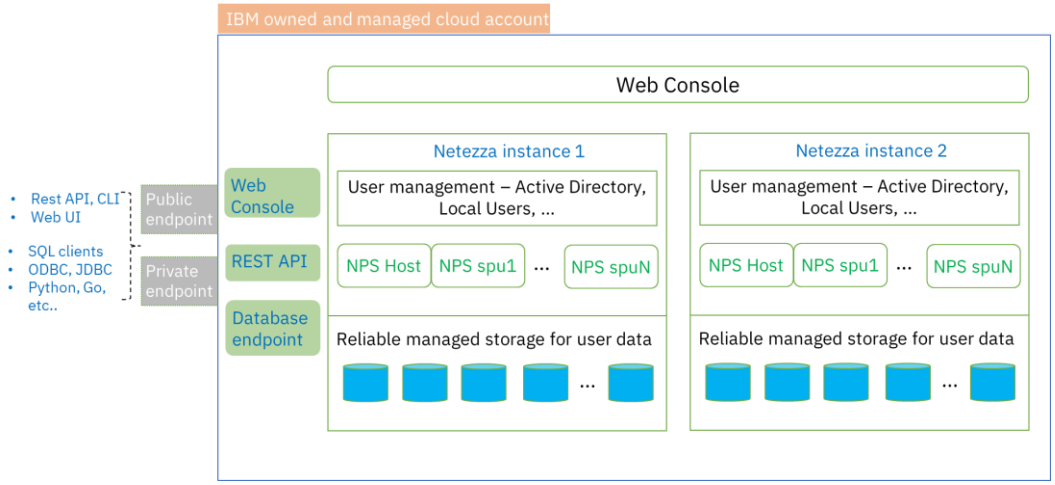



**Figure 3: IBM Netezza Performance Server for IBM Cloud Pak for Data-as-a-Service Architecture**

Together, IBM Cloud Pak for Data and NPSaaS offer a data fabric to automate data discovery, data governance, and data consumption in hybrid multi-cloud environments. It also makes it easier to contextualize data into data models and enables reusability for new use cases:

- Provide multiple business units with direct data access via a self-service insight platform
- Use fit-for-purpose compute capacity to run models on billions of rows of data efficiently
- Collaboratively develop models and easily deploy those models to infuse insight throughout the organization.

NPSaaS delivers unprecedented capabilities (Figure 4):

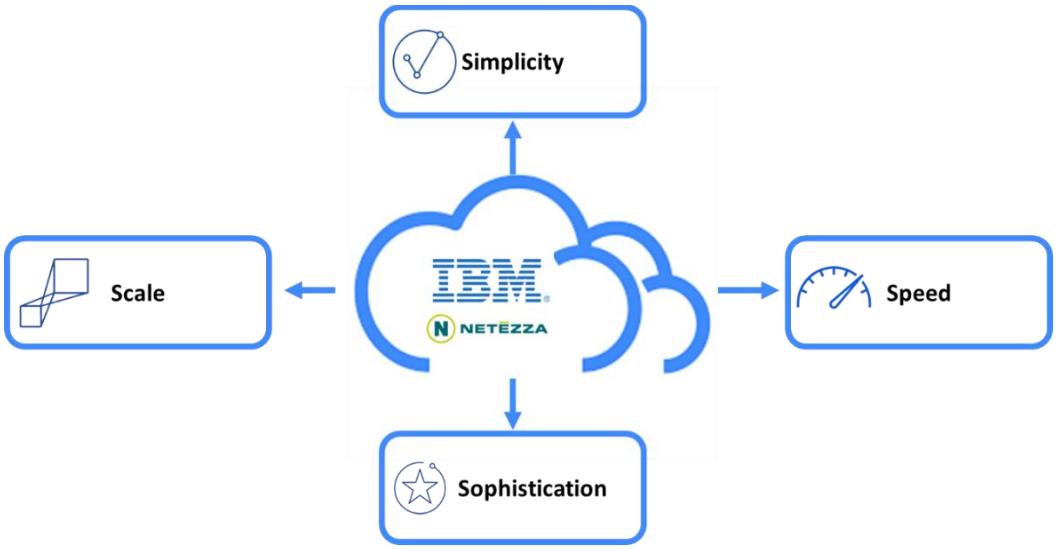



**Figure 4: Unprecedented capabilities of IBM Netezza Performance Server as a Service (NPSaaS)**

Existing Netezza customers will enjoy a risk-free and frictionless upgrade to NPSaaS with a single command

One cloud platform unifies data and AI at scale

Granular scaling lowers need to over-provision and waste resources

- **Scale:** Highly granular elastic scaling can be fine-tuned for the specific workload. It is possible to scale and manage compute and storage independently when needed. Pay only for what is used and pause/resume as needed.

  NPSaaS is spun up only when needed (at customer's schedule) to consume only when and what's required. There are no awkward pauses, and consumption is managed with automated or scheduled pause and resume.

- **Simplicity:** It is easy to load and go, requires minimum administration, no tuning, no indexing, and existing Netezza customers will enjoy a risk-free and frictionless upgrade to NPSaaS with a single command. To further simplify selection, NPSaaS has performance profiles and storage density, independent of each other. Users can select the performance characteristics to meet their workload requirements and increase or decrease performance as the workload demand ebbs and flows.

  Cloud-managed compute and highly available cloud storage are managed by a hybrid multi-cloud foundation of Red Hat OpenShift. The storage density can be increased independent of performance demands as capacity needs require.

- **Speed:** Faster insight and faster time to value is possible with vastly superior and predictable price-performance, hourly credit consumption model, with no overcharges or over-provisioning.

- **Sophistication:** One cloud platform unifies data and AI at scale with IBM Cloud Pak for Data's analytical ecosystem services, Netezza's in-database analytics, and mature support. With its rich built-in analytics capabilities, Netezza brings analytics to the data, eliminating the need to deal with the additional overhead associated with data preprocessing and transformations.

Granular scaling with cost predictability capabilities of NPSaaS can uniquely improve the costs and benefits for customers compared with Snowflake and/or other cloud EDWs.

## Granular scaling tailored to workloads with NPSaaS lowers costs

Like Snowflake, IBM has attractively priced NPSaaS to provide customers considerable choice and flexibility. However, one of the most differentiating features of NPSaaS is how customers can order and pay only for the resources they need. The granular design of its resource elements (i.e., Performance Profile and Storage Density – Figure 5) enables customers to access resources dynamically and easily track demand. This granularity ensures that customers do not over-provision and waste resources in anticipation of higher usage in the future.

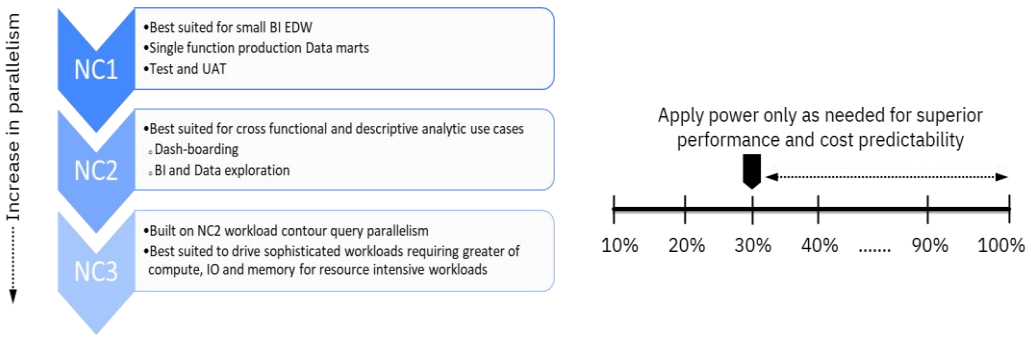



**Figure 5: NPSaaS workload contours, performance profiles, and storage**

NPSaaS Performance Profiles enable flexible scaling with predictable costs

Multiple performance profiles for varying time periods to optimize usage and to fit workload patterns

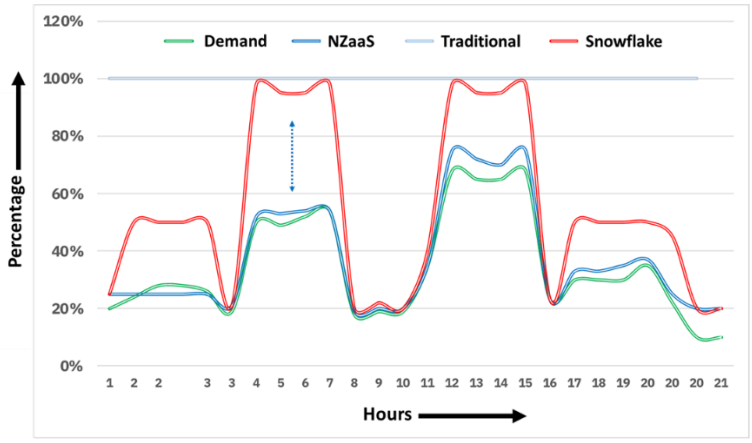
Real-life Retail workload example

IBM announced the general availability of three performance profiles, NC1, NC2 and NC3, with corresponding storage optimized for costs and performance for the workload contours shown in Figure 5. Later this year, other Performance Profiles are planned to be announced. Customers can easily scale within a contour by adjusting the profile (higher or lower) online and in real-time without any assistance from external or internal resources. This flexibility allows them to innovate with speed in response to customer demand, market opportunity, regulatory changes, or a competitor's move while maintaining predictable costs.

**NPSaaS pricing details:** Each NPSaaS performance profile comes with a range of storage (e.g., NC1 to NC100 comes with arrange of 2 TB to 61TB) and is associated with Resource Unit (RUs/hour). For example, the NC-25 profile costs 40/hour, and 18 hours of workload in a day will result in a credit consumption of 720 RUs for that day. The RU rates vary with geography. Assuming that NPSaaS is hosted in an eastern United States cloud center at $2/Hour, the above example will cost $1,440.

Netezza instance and its associated RUs are purchased using Cloud Pak for Data credits, which are the common currency for all of IBM's cloud service offerings. Credits are available in two ways: On-Demand or in credit deposit blocks appealing to customers looking to optimize credit acquisition cost. Clients can add multiple performance profiles for varying time periods (scheduled or on-demand) to optimize the usage and fit the workload patterns. Credit consumption includes provisioned IBM Cloud Pak for Data, Software entitlement, Managed services, and support. Data transfer is priced separately (IBM does not add any uplift to the price charged by Azure for Data transfer).

**A real-life Retail workload** provided by IBM (Figure 6) illustrates how customers can dynamically change Performance Profiles throughout the day to align with workload demands (green line) closely.




| Netezza Performance Profile | Hours Per Day | Storage Density (TBs) | Credit Velocity (RU/hour) | Credits Projection per day (RU/day) |
|---|---|---|---|---|
| NC2-25 (baseline) | 18 | 31 | 40 | 720 |
| NC2-30 (surge) | 4 | 31 | 48 | 192 |
| NC2-60 (surge) | 2 | 31 | 76 | 152 |
| **Total** | **24** | | | **1064** |

**Figure 6: Retail workload depicting NC2 Performance Profiles**

Granular scaling with NPSaaS provides significant cost advantages over Snowflake

Demand is low earlier and later in the 24-hour cycle, moderate in the morning, and higher late afternoon. If the customer were to meet this demand with an on-premises solution, they would need a system shown by the "traditional" line. It will result in significant over-provisioning and wasted excess capacity. The granularity of NPSaaS allows for provisioning to track demand (dotted green line) tightly. For example, for 18 hours a day, a lower performance profile is used (NC-25), and surges are satisfied with higher performance profiles (NC-30 and NC-60) with a total consumption of 1,064 resource credits. However, Snowflake lacks this level of granular elastic scaling and has to provision XL and 2XL sizes to meet this demand. It will result in over-provisioning, wasted resources, and higher prices.

In addition to granular scaling, other factors drive performance and cost advantages for NPSaaS over Snowflake, especially as the analytics problems grow in size.

## Cost and performance advantages of NPSaaS over Snowflake

Figure 7 depicts results of IBM's NPSaaS and Snowflake benchmark/sizing studies for Use Case 1, Use Case 2 (Figure 7), and several larger NC1 and NC2 profiles.

| Use Case 1 | | |
|---|---|---|
| NPSaaS | Snowflake | Hours/day |
| NC1-25 | L | 18 |
| NC1-50 | XL | 4 |
| NC1-100 | 2XL | 2 |

| Use Case 2 | | |
|---|---|---|
| NPSaaS | Snowflake | Hours/day |
| NC1-25 | L | 10 |
| NC1-50 | XL | 9 |
| NC1-100 | 2XL | 5 |

**Figure 7: Details of Use Case 1 and Use Case 2**

Benchmarks and sizing studies for NPSaaS vs. Snowflake

Figure 8 shows the annual license cost for these workloads. NPSaaS license cost is between 8% and 47% lower than Snowflake. Cost savings from the larger profiles (NC2) with more storage are greater. It implies that as the data size gets larger, the cost savings from NPSaaS would be a greater percentage of the Snowflake license costs.

NPSaaS license cost is between 8% and 47% lower than Snowflake

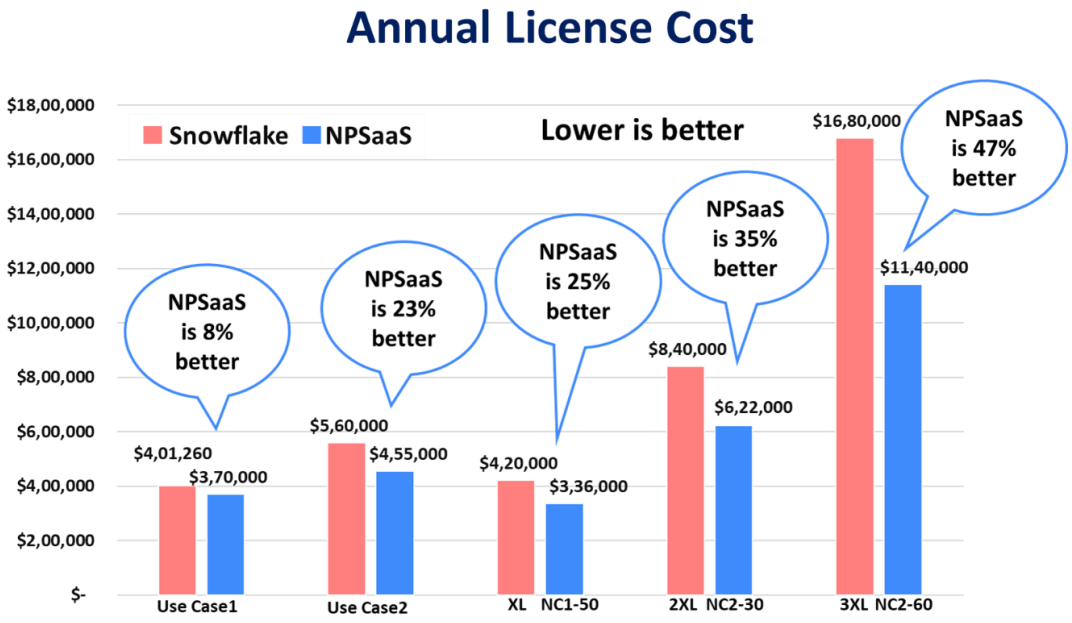



**Figure 8: Cost comparisons NPSaaS versus Snowflake for various workload contours**

In addition to lower license costs, NPSaaS also performs better. Figure 9 shows the elapsed time for NC2 profiles for NPSaaS and Snowflake. Again, the performance differential of NPSaaS over Snowflake gets larger as the data size grows.

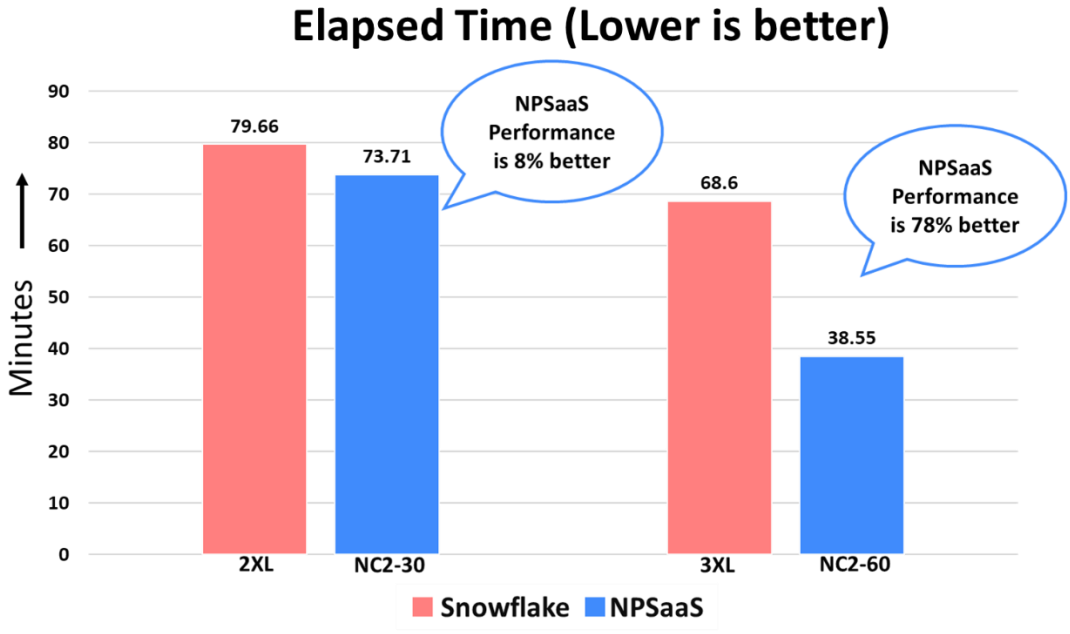



**Figure 9: Performance advantage of NPSaaS over Snowflake**

NPSaaS is up to 78% faster than Snowflake

Some of the additional factors that consistently drive better price-performance for NPSaaS over Snowflake include:

- Persistent block storage with hybrid columnar assist
- No indexing, no tuning, mature and better workload management, auto compression
- Granular elastic scaling on storage independent of computing with predictable cost profile.

NPSaaS cost and performance advantages provide a better TVO

These cost and performance advantages and other unique features of NPSaaS can help customers maximize their TVO and get a better return on investment (ROI) than Snowflake.

## High-level framework for the TVO of Cloud EDWs

The Total Value of Ownership (TVO) framework (Figure 9) can illustrate the value of NPSaaS. It goes beyond just the Total Cost of Ownership (TCO) and categorizes interrelated *cost/benefit drivers (circles)*[6] for cloud EDWs by each quadrant: Costs, Productivity, Revenue, and Risks. The drivers along the horizontal axis are primarily Technology or Business drivers. Similarly, drivers along the vertical axis depict ease of measurability: Direct or Derived.

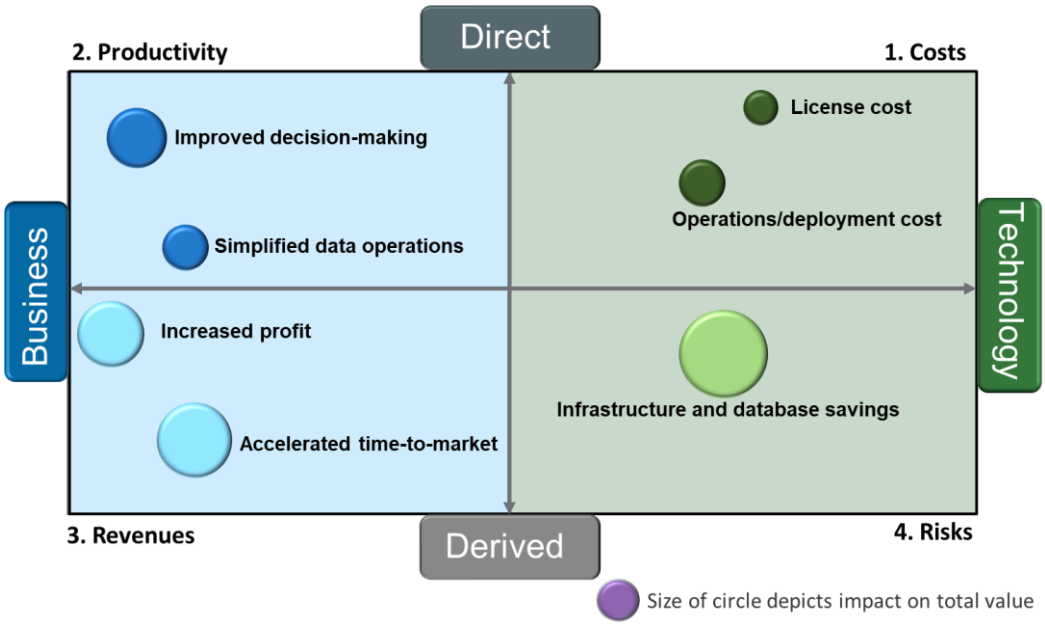



**Figure 10: Total Value of Ownership (TVO) framework for Cloud EDWs**

Total Value of Ownership (TVO) framework for Cloud EDWs

The circles depict cost/value drivers for cloud EDWs, and their size is proportional to the potential impact on a client's Total Value (Benefits – Cost) of Ownership or TVO as follows:

[6] Forrester, "The Total Economic Impact™ Of Snowflake's Cloud Data Platform", July 2020.

Value includes lower costs, enhanced productivity, higher revenues/ profits, and lower risk

1. **Costs:** Typical costs include one-time acquisition costs for migration and *deployment and annual costs for software, maintenance, and operations.*
2. **Productivity:** The TVO model quantifies the value of productivity gains of data scientists, data engineers, applications developers, and the organization through *improved decision-making and simplified data operations.*
3. **Revenues:** *Accelerated time-to-market* with better performance. Greater innovation and better decision-making capabilities spur growth, revenues, and *increased profit.*
4. **Risks:** Lower risk of project failure (even well-planned Analytics projects have up to 60% failure rate[7]) with a streamlined workflow with reusable components, better ecosystem, business/IT collaboration, and enhanced security/ privacy. Improved governance with better data cleansing/quality and process consistency. All these translate to *infrastructure and database savings.*

A recent study quantified the economic impact of using Snowflake[6] for a "composite" organization (a proxy for several Snowflake customers interviewed) with $5 billion in revenues and a maximum data size of 563 terabytes. This study compared Snowflake with an alternative of using multiple ad-hoc on-premises or cloud data warehouse solutions with siloed data sets scattered across the enterprises. We used this study as a basis for comparing the TVO of NPSaaS with Snowflake.

Cost-Benefit analysis over 3-year horizon

## TVO of NPSaaS vs. Snowflake – Assumptions/Results

The Cost-Benefit Analysis presented here for the same 'composite" organization quantifies the Total Value (Total Benefits – Total Costs) for three years of operations and compares NPSaaS with Snowflake.

**Key assumptions:** The cost and benefit data for Snowflake used were identical to the prior study. We conservatively assumed that the deployment and operations costs are the same for Snowflake and NPSaaS.

TVO analysis for three cases

To study the impact on TVO of lower annual license costs of NPSaaS over Snowflake, we examined three cases with 8%, 25%, and 47% lower license costs (Figure 8). Annual license cost savings with NPSaaS typically grow with data size. So, these lower license costs assumed for this TVO analysis are probably conservative assumptions because the data size of the composite organization (563 terabytes) is much larger than the previously mentioned NPSaaS benchmark cases.

To consider the positive impact on TVO of better performance of NPSaaS over Snowflake and other factors (better workload management, more granularity, integration with Cloud Pak for Data, etc.), the benefits of NPSaaS over Snowflake were scaled up by a few percent. We used a greater percentage for those benefits (*accelerated time-to-market, improved profits*) that improve with better EDW performance. Likewise, we scaled up the benefits for larger cases (more NPSaaS license cost savings).

**Results with 8% lower license cost:** Figure 11 depicts the costs and benefits mapped by each quadrant and value driver. NPSaaS license costs are lower, and benefits gained from improved productivity, higher revenues, and lower risks are higher.

[7]Why big data projects fail and how to make 2017 different, Expansion of Gartner's prediction that 60% of big data projects fail; By Sameet Agarwal, Network World.

For the 8% case, NPSaaS provides more productivity, faster time to value, better governance, and higher revenue/profits

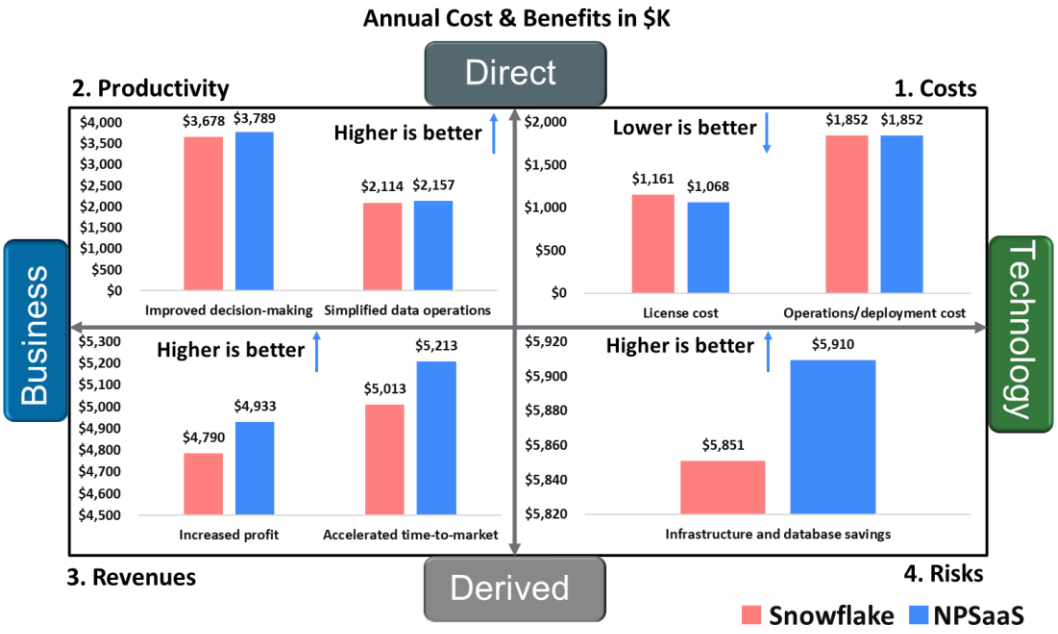



**Figure 11: Annual Costs and Benefits for NPSaaS versus Snowflake with 8% lower license costs**

**Results with 25% lower license cost:** Figure 12 depicts the costs and benefits mapped by each quadrant and value driver. NPSaaS license costs are much lower, and benefits gained from improved productivity, higher revenues, and lower risks are much higher.

For the 25% case, NPSaaS provides even more productivity, faster time to value, better governance, and higher revenue/profits

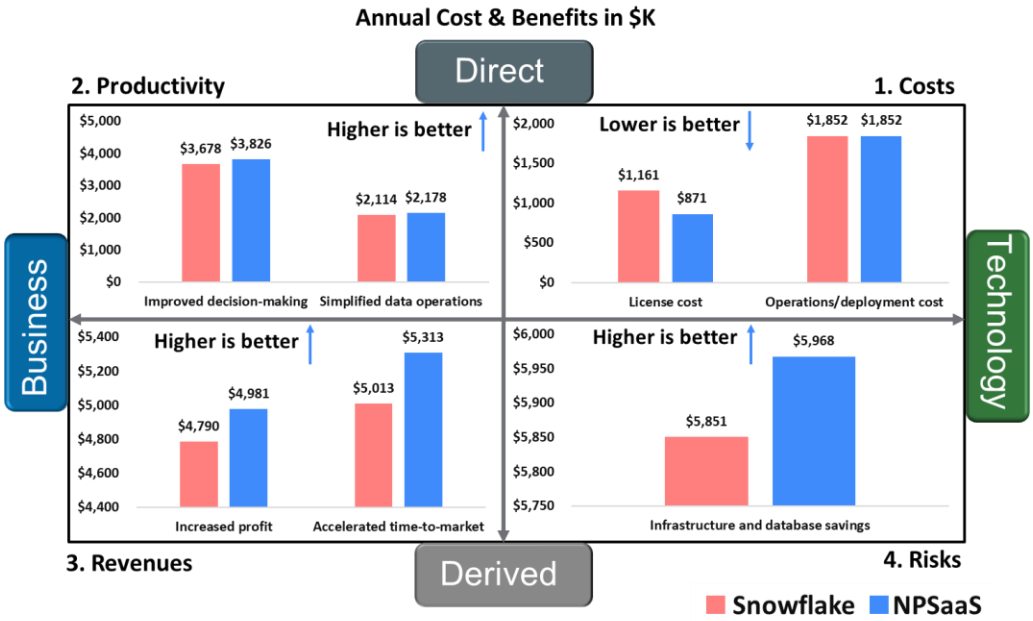



**Figure 12: Annual Costs and Benefits for NPSaaS versus Snowflake with 25% lower license costs**

**Results with 47% lower license cost:** Figure 13 depicts the costs and benefits mapped by each quadrant and value driver. NPSaaS license costs are far lower, and benefits gained from improved productivity, higher revenues, and lower risks are far higher.

For the 47% case, NPSaaS provides a lot more productivity, faster time to value, better governance, and higher revenue/profits

NPSaaS delivers more license cost savings and these grow with data size

Enhanced productivity with improved decision-making

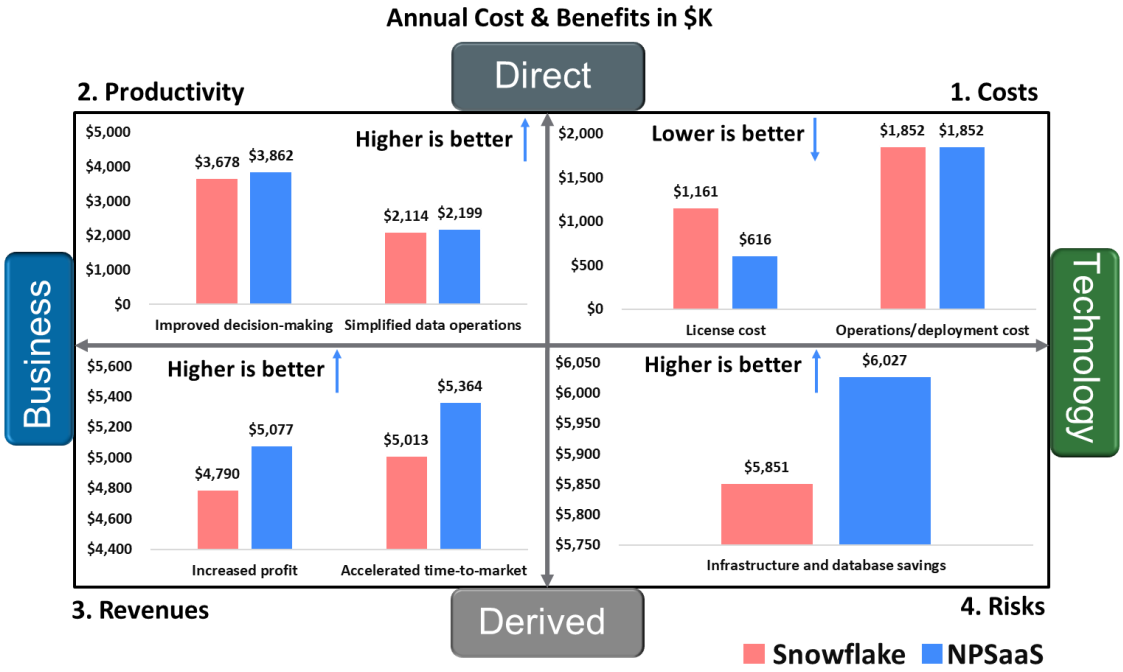



**Figure 13: Annual Costs and Benefits for NPSaaS versus Snowflake with 47% lower license costs**

## Discussion of Results from the TVO model

Most Total Cost of Ownership (TCO) models only quantify the costs in the top right quadrant (Figure 10). The TVO model outlined here considers these costs and the benefits from the value drivers in the remaining three quadrants. The TVO for NPSaaS is better than Snowflake in all cases with:

1. **Lower costs:** As discussed earlier, NPSaaS significantly reduces license costs compared with Snowflake. These license cost savings grow with data size. While the model conservatively assumes that *deployment/operations costs* are the same for both NPSaaS and Snowflake, the simplicity inherent in NPSaaS (described earlier) could drive additional reductions in *deployment/operations cost*.

2. **Enhanced productivity:** Data scientists, data engineers, applications developers, and the organization can become more productive with NPSaaS with *improved decision-making with faster access to data and simplified data operations.*

   *Improved decision-making:* As the volume and variety of data continues to explode, companies must handle large datasets (e.g., 2 Petabytes). NPSaaS handles large datasets well. On the other hand, Snowflake is designed to work optimally on smaller datasets. Hence, it suffers from performance and scale problems with larger datasets.

   As the data size gets larger, many cloud EDWs suffer cost, performance, SLA, and predictability issues because of resource (I/O and CPU) contention and collision among competing workload elements. So, they must separate workloads. Snowflake achieves workload separation by assigning separate compute and disk clusters of various sizes (Virtual Data Warehouses – VDWs) to specific workload elements, i.e., BI Reporting, ETL, Data Science, etc. However, this can become inefficient and expensive as these workload elements grow in variety and size.

Enhanced productivity with simplified data operations

Higher revenues and profits with greater cost savings and accelerated time to market

Increased profit from faster time-to-market

NPSaaS has performance profiles and storage density, independent of each other. Users can select the performance characteristics to meet their workload requirements and increase or decrease performance as the workload demand ebbs and flows. The storage density can be increased or decreased independent of performance demands and capacity needs.

*Simplified data operations:* NPSaaS has a single, modernized, unified GUI that helps administrators monitor and manage system resources, administer database objects, configure workload management, view active sessions, and monitor system resource utilization for capacity planning.

In addition, it is easy to load and go, requires minimum administration, no tuning, no indexing, and comes with risk-free frictionless upgrades. A consistent set of services helps streamline operations and manage costs as workloads scale.

3. **Higher revenues and profits:** Improved performance drives *cost savings from accelerated time-to-market*, increasing *profit.*

*Cost savings and increased profit from accelerated time-to-market:* NPSaaS runs faster, especially as the data size grows and its highly granular elastic scaling can be fine-tuned for the specific workload. It is also possible to independently scale and manage compute and storage when needed, pay only for consumed resources, and pause/resume as needed.

On the other hand, Snowflake processes queries using MPP compute clusters where each node stores a portion of the entire data set locally. This approach offers the data management simplicity of a shared-disk architecture. VDW and zero movement data sharing are some of Snowflake's notable and unique features. However, the lack of a robust workload management solution, particularly for very large datasets, presents scalability and governance challenges for Snowflake users. There are multiple functions that can't be cached. Consequently, this limits performance as well as scalability/elasticity. Also, Snowflake uses local NVMe for cache, but its persistent data resides on 100x slower S3/cloud object storage.

Snowflake depends on getting data to warm cache for optimal performance. Since caching is not available for all the Snowflake EDWs, it will slow the data access and analysis process. Also, as load increases, additional warehouses must be instantiated, and not every VDW will access the cache.

*Increased profit from faster time-to-market:* Sporadic and unplanned demand spikes challenge EDW operations and adversely impact end-user satisfaction. Many cloud EDW solutions such as Snowflake support concurrent spikes by spinning up additional processing clusters. This additional compute cost is billed to the customer and can be very large. As an alternative, Snowflake offers customers the ability to pause compute when there is no activity. However, pausing deletes the cache and significantly deteriorates performance when restarted. NPSaaS with workload management can address this without pausing and provides very granular elastic scaling to optimize even very spiky workloads.

For data scientists, NPSaaS offers a built-in analytical infrastructure and an extensive library of statistical and mathematical functions, supporting a breadth of analytic tools and programming languages, including Open-Source R. It is delivered with a

library of more than 200 prebuilt, scalable, in-database analytic functions that execute analytics in parallel while abstracting away the complexity of parallel programming from the developers, users, and DBAs.

Reduced risks and more infrastructure and database management savings

4. **Reduced risks:** Higher availability, better governance, and a robust ecosystem drive *infrastructure and database management savings.*

   *Infrastructure and database management savings*: NPSaaS, being part of the IBM Cloud Pak for Data ecosystem, has enterprise-grade containerization capabilities and can be integrated with enterprise management and lifecycle operations. It has utilities to load and unload large volumes of data and comes with extract-transform-load (ETL), master data management (MDM), and data governance tools under a single umbrella.

   In addition, built-in automation and auto-recovery in NPSaaS provide a highly available and fault-tolerant deployment with a minimal human touch, ensuring the EDW is available 24x7x365. Failure of a processing node or any element in the system causes no significant performance degradation providing a robust, production-ready environment throughout the lifecycle. The hot standby node offers increased uptime and availability.

Highly available and fault-tolerant deployment of a cloud native MPP architecture with built-in analytic capabilities reduce risks

   The Snowflake Connector for Spark ("Spark connector") brings Snowflake into the Apache Spark ecosystem, enabling Spark to read data from and write data to, Snowflake. From Spark's perspective, Snowflake looks like other Spark data sources (PostgreSQL, HDFS, S3, etc.). Snowflake does have a rich portfolio of analytical features and services, including from partners. However, this imposes additional infrastructure and network capacity requirements, and substantial integration is needed to integrate partner tools.

   Finally, the NPSaaS cloud native MPP architecture provides a significant boost in performance and scalability without the need to copy the data into a separate analytics server. With its 200+ built-in analytics capabilities and rich support for all major languages such as Python, C, C++, Java, Lua, R, Netezza operationalizes Data Science and Machine Learning at Scale.

NPSaaS delivers a better ROI than Snowflake and provides even better ROI gains at larger data volumes.

## Better ROI with IBM Netezza Performance Server as a Service (NPSaaS)

For the "composite" organization, Figure 14 depicts the ROI for a 3-year time horizon for Snowflake and all three NPSaaS cases considered in this study.

NPSaaS delivers a better ROI than Snowflake

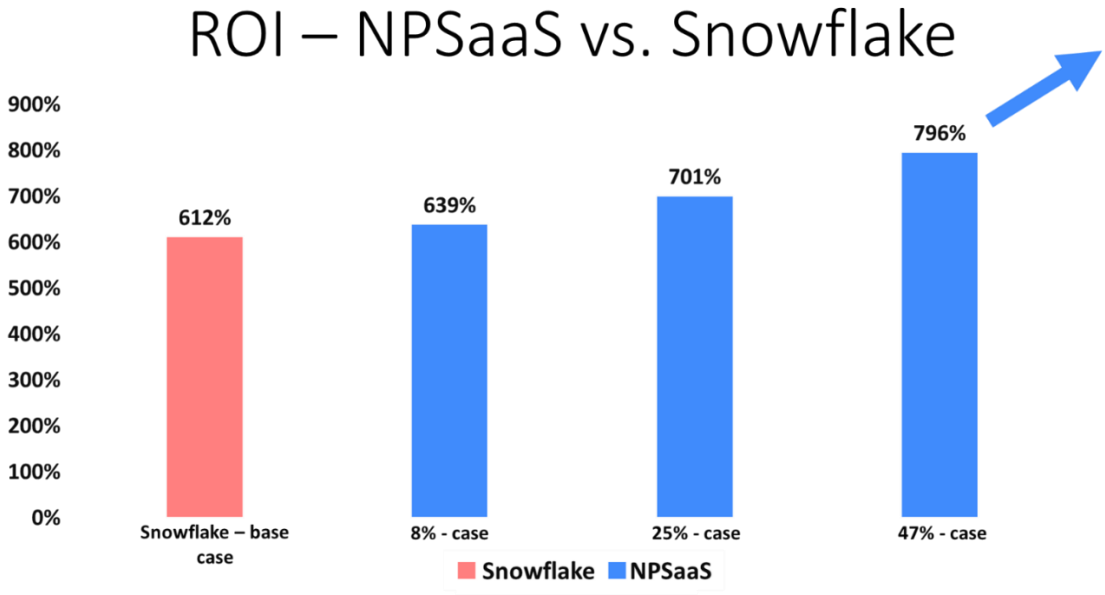



**Figure 14: ROI over three years of NPSaaS over Snowflake**

As previously mentioned, even a 47% license cost savings with NPSaaS is a very conservative estimate. If we run the benchmarks for the "composite" organization, the cost savings and ROI could be much more.

## Conclusions and Recommendations

Analytics is a game-changing business opportunity for companies to deliver exceptional customer experience, enhance marketing effectiveness, increase operational efficiencies, reduce financial risks, improve product quality and reliability, and more. To obtain faster actionable insights from a growing volume and variety of data, many organizations deploy analytics solutions across the enterprise.

Cloud EDWs growing in use and produce realistic, reliable, actionable, and time-critical analyses

However, as the volume and variety of data grow and organizations implement higher-value Analytics, they need enterprise-grade solutions, including cloud EDWs, to produce more realistic, reliable, actionable, and time-critical analyses. These are the reasons why Cloud EDW deployments are growing.

Snowflake is one of the earliest pioneers in this growing Cloud EDW market. It delivers the performance, flexibility, and scalability to load easily, integrate, analyze, securely share, consolidate, and govern data/analytics across data warehouses, data marts, and data lakes.

NPSaaS delivers highly granular elastic scaling with predictable costs from small to multi-petabyte capacity

IBM recently announced NPSaaS, a fully managed service on Microsoft Azure. It leverages the same software building blocks that underpin IBM Netezza, which has a large client base. Built on IBM Cloud Pak for Data, NPSaaS delivers highly granular elastic scaling with predictable costs from small to multi-petabyte capacity. In addition, it is simple to deploy and operate and provides the speed, stability, and sophistication to accelerate a client's Analytics/AI journey.

The ROI for Snowflake was 612%, whereas the ROI for NPSaaS ranges from 639% to 796%

Compared to Snowflake, NPSaaS delivers higher value and ROI especially for complex analytics/AI and for larger data sizes

Benchmarks and sizing analysis indicate that NPSaaS lowers license costs and elapsed time for analytics workloads compared to Snowflake. This cost and performance differential grows as the data size grows.

Using previously mentioned NPSaaS benchmarks, sizing information, and a Snowflake base case from a previous study, the TVO analysis presented here quantifies the costs and associated benefits for a "composite" organization for three NPSaaS cases. The ROI for Snowflake was 612%[6], whereas the ROI for NPSaaS ranges from 639% to 796%. These NPSaaS estimates are probably conservative as the data size of the composite organization is much larger than the benchmark cases used for this analysis.

Clients deploying Cloud EDWs should seriously consider NPSaaS for the following reasons:

1. The cost-benefit analysis is compelling for all cases
2. The business value and ROI differential improve as data sizes get larger
3. This investment is protected and can continue to deliver even greater marginal value for more complex analytics, including the rapidly growing use of Artificial Intelligence and Machine Learning (AI/ML) techniques coupled with the Internet of Things (IoT) – all areas where IBM continues to make substantial investments.
4. IBM is a reliable partner and offers one of the most comprehensive cloud solutions portfolios for Analytics and AI.

*Cabot Partners is a collaborative consultancy and an independent IT analyst firm. We specialize in advising technology companies and their clients on how to build and grow a customer base, how to achieve desired revenue and profitability results, and how to make effective use of emerging technologies including HPC, Cloud Computing, Analytics and Artificial Intelligence/Machine Learning. To find out more, please go to [www.cabotpartners.com](http://www.cabotpartners.com)*

CABOT PARTNERS

OPTIMIZING BUSINESS VALUE

**Cabot Partners Group, Inc.**

100 Woodcrest Lane, Danbury, CT 06810

www.cabotpartners.com| info@cabotpartners.com