Assessing Your Journey From Netezza to Snowflake Data Cloud

NTT DaTa

For organizations large and small, data migration can be overwhelming. Many tools and solutions are available to help migrate, transform, organize and secure data. But not every solution is right for every organization. It's important to know that any migration may include different use-case scenarios, types of data, existing investments and ongoing costs, and a host of other factors.

So, how can you achieve the highest value when migrating from IBM Netezza to Snowflake Data Cloud?

It's always best to consult experts to understand which path is best for your business based on your specific needs. That's why an assessment is so key to getting the most out of your data when migrating from Netezza to Snowflake.

Why you need an assessment

An assessment is an in-depth review of your organization's systems, processes, technology, people and design considerations in the context of a specific business problem you want to solve. It gives you time to understand the specific project requirements in-depth — while also helping you better understand the project's activities, timelines and milestones. Assessments lower or remove risk so you can get a more accurate estimate and potentially reduce costs for a successful engagement and outcome.

But it also presents some unique concerns that you need to watch out for to avoid problems later. It's a rule, not an exception, that data teams use a wide range of technologies for data persistence, reporting and analysis beyond traditional databases for data manipulation and processing.

Frequently, data processing systems involve hundreds or thousands of sources of information, each with its own process (and target database) for managing the movement and manipulation of data. The orchestration system your data teams use to manage data processing may be complicated, magnifying the cost of making changes to processing jobs and leading to issues that your most vocal stakeholders will quickly bring to your attention.

Where to start

The consulting sales and executive team at your organization will have already set the groundwork for understanding your project's goals and requirements, so you can use the assessment period to refine that understanding. Specifics to look for include:

- **Hard requirements:** Consider things like budget, security, timeliness and technology.
- **Methodology:** How will the solution be delivered — through parallel systems, moving processes one at a time or hard cutover?
- **Updates versus historical data:** Often, only a portion of the data is updated each processing period. Previously existing data is often processed differently than newly arriving data.
- **Networking concerns:** Netezza systems are almost always onsite or within some sort of private corporate network. Be mindful of what shifting to a cloud solution will involve.
- **Existing technology and Snowflake compatibility:** Snowflake is still relatively new, and not everything works with it.
- **Timelines and milestones:** When are you going to finish? If you've already set the timeline before the assessment, does it still make sense?

Let NTT DATA be your partner for this journey. We have the expertise to help you successfully modernize your organization's approach to data, reducing the time and resources you spend maintaining data and delivering immense value.

Understanding how data is consumed, processed and moved

The real value in data processing systems like Netezza or Snowflake is their ability to drive insights into data. Data migration impacts everyone who uses the system, running reports and queries, and they're all likely to care that the system continues to work. These stakeholders are often excellent resources for quality control and testing during development. They also have some of the highest domain knowledge of the data you'll find and can be valuable resources for diagnosing data consistency issues or understanding undocumented business logic.

Assessing Your Journey From Netezza to Snowflake Data Cloud

By far, most of the work in a typical data migration involves updates to the ETL (extract, transform, load) process that loads the target system with information. The key things to consider when understanding and analyzing data processing are:

- Number of Netezza systems to migrate
- Data volumes involved
- Sources of data: Format, access mechanisms, frequency of updates and schedule, and processing
- Technology used to process the data
- How the data is processed: Cleaned, business rules applied, merged with existing data or replaced
- Management of historical data as it ages
- Any Netezza-specific SQL, stored procedures or functions
- Data access roles: Updates, processing, queries and consumption
- Software development lifecycle for managing changes to the data processing
- Table distribution

It's also essential to understand the flow of data. For ETL processing, you should know how data from the database is utilized, as well as how and where new source data is obtained. You should fully understand any Netezza-specific considerations related to the current system for migration planning. It's necessary to know if these same patterns will work when migrating to Snowflake using ELT (extract, load, transform) processing.

In Netezza to Snowflake migrations, NTT DATA recommends that assessments include several key activities:

Establish a relationship with key stakeholders within your organization

Understand the high-level goals and requirements of your project

Understand how your data is ultimately consumed and used

Understand the technology involved in collecting, processing and updating your data

Ensure you observe all security or data governance guidelines

Establish a baseline for automating database conversion

Build in time for a representative pilot conversion with a subset of data to confirm your findings

Form a plan for how you're going execute the migration while balancing all of the above

Another critical area is the software development lifecycle. Throughout the project, you'll need to make changes to the data processing software and understand how it manages data processing changes. How will the migration project coexist with existing processing and follow your planned methodology for migration?

Data ownership and data access may also be a concern. Often, one or a few sets of roles process all data. However, if fine-grained data access is a design requirement, pay attention to how it will impact your project. This added complexity may mean the project will take longer to complete.

Also consider how you'll confirm the data has been migrated successfully to Snowflake. Complex data processing can often involve undocumented business logic within the data processing code. When changing the way data is processed, you mustn't alter the intent of these business rules. Some strategies to confirm successfully migrated data include:

- Picking a small subset of data to check
- Comparing row counts or other aggregates grouped by various criteria
- Comparing hash-based checksums of rows
- Leveraging subject-matter experts at your organization to help validate data

It may be useful to migrate sets of related tables, jobs and reports and to work through validation in parallel with the migration effort.

Addressing operational concerns, security and data governance

This is where you'll need to consult your organization's networking and system administrators. Most Netezza to Snowflake migrations involve changes in where data will be processed and moving data off site to a cloud environment. Administrators can often facilitate systems access and network policy changes, and they're essential gatekeepers for actions that can block large portions of the project if not complete.

Many data migration projects deal with sensitive data containing personally identifiable information (PII) or other information that needs to be handled with care. It's important to understand what sensitive data exists and how this data is currently managed/governed. For example, PCI compliance requires encrypted data at rest within a table, along with access selectively granted through a stored procedure or function. Understanding the roles, technology and processes in place is essential for defining this within Snowflake. It's also key to finding a solution that's the best fit for the organization.

Data governance may be a concern within your organization, too. Many organizations manage who, what, when and how data is accessed. It's essential to ensure this knowledge is modeled in Snowflake for a successful engagement. Sometimes data governance also covers areas within master data management (MDM) and may have policies and rules in place governing the naming of various entities and processes. Be sure to follow these guidelines and express their intent within Snowflake.

Choosing the right data migration strategy Organizations generally take one of three data migration approaches: "lift and shift," organize or modernize. Each approach has shortcomings and some key advantages.

Using what you've learned: Planning and deliverables

The work you've performed up to this point should support the migration's design and a plan to execute it. Think through the various large components of the project, the types of work needed for each and the work patterns. Then decide how to divide the work among team members. Whiteboarding a solution and discussing it with key stakeholders can help you understand issues or concerns you may have missed.

Break the major components into a set of milestones, and for each milestone, think about how you'll track progress toward it. What constitutes success for you and your team? Consider potential blockers, such as systems access, network policy change authorization, data movement timing or other things that can cause delays. Discuss the plan and work with your team to balance their timing, budget and technology constraints. Do your best to set yourself and your team up for success. If anyone sees significant concerns, now is the time to raise them.

You should come out of the assessment project with several key deliverables:
- A strong understanding of the business needs and how your organization uses the data you want to migrate
- A clear plan for how to proceed with the migration
- A set of milestones against which to track progress
- A design for the migration
- Documentation, diagrams and other materials that support your decisions
- Relationships with the key people within your organization who will facilitate the project's completion

No matter what strategy you choose to take in a Netezza to Snowflake Cloud Data migration, you need to carefully consider every angle, from use cases (application requirements and desired outcomes) and existing technologies to your team's skill sets and organizational experience. Each plays a part in the decision.

NTT DATA recommends a blended, iterative approach to ensure our clients fully own all parts of the process. You can achieve this mix in three waves:

1. Assess and discover, which includes static analysis, an automation estimate and effort sizing.
2. Migrate, which includes iterative replication of data ingestion, transformation and consumption workflows, validation, new workload monitoring, legacy pipeline retirement and documentation.
3. Modernize, which includes developing net-new capabilities as well as modernizing replicated workflows in the new architecture.

Key considerations:

- Is there an opportunity to purge or archive unwanted historical data before migration?
- Do any data elements need to be tokenized while migrating?
- During data validation, is the row count enough, or does the row correctness also need to be verified by row checksum?
- Are there any Netezza proprietary SQL functions in BI reports that need to be replaced with appropriate Snowflake ANSI SQL functions?
- Is there any sensitive data in BI reports that needs to be restricted (masked by policies/views/secure views or external de-tokenization)?

Let NTT DATA be your partner for this journey. We have the expertise to help you successfully modernize your organization's approach to data, reducing the time and resources you spend maintaining data and delivering immense value.