Snowflake® for SAP Analytics

The agile, scalable and cost-effective approach to data management.

Snap ANALYTICS

Contents

Preface

In the ever-evolving landscape of data analytics, organizations are constantly seeking the most robust and flexible platforms to harness the power of their data. As enterprises embark on their journey towards data-driven decision-making, the choice of the right data platform becomes paramount. This book, **"Snowflake for SAP Analytics,"** delves into the rationale behind the increasing trend of organizations choosing Snowflake over SAP as their preferred data platform.

In recent years, Snowflake has emerged as a trailblazer in the world of cloud-based data warehousing, offering unparalleled scalability, performance, and ease of use. This book aims to explore the reasons why an increasing number of customers are opting for Snowflake, specifically for their SAP analytics needs.

The decision to choose Snowflake is not merely about a preference for one platform over another; it's about embracing a paradigm shift in how organizations manage and derive insights from their data. Snowflake's architecture, built for the cloud, provides a foundation that seamlessly integrates with SAP analytics solutions, allowing for a more agile, scalable, and cost-effective approach to data management.

To aid organizations in their journey, this book provides practical recommendations on how to get started with Snowflake for SAP analytics. From the initial setup to data migration strategies, readers will find valuable insights and guidance to ensure a smooth transition and optimal utilization of the Snowflake platform.

Additionally, we showcase accelerators developed by Snap Analytics: Tools, techniques, and strategies that expedite the implementation and adoption of Snowflake for SAP data. These accelerators are designed to empower organizations to derive value from their data faster, making the transition to Snowflake a strategic advantage.

As the authors of this book, we have drawn on our collective experience in the fields of data management, analytics, and cloud technologies to provide a comprehensive resource for organizations navigating the complexities of choosing and implementing a data platform for SAP analytics. We hope that this book serves as a valuable companion on your journey towards unlocking the full potential of your data with Snowflake.

Happy reading!

The Snap data & analytics team

Prologue:

Connecting data, technology and teams

Snap Analytics helps enterprises to solve complex data problems. They enable organisations to leverage the latest cloud-based solutions, to combine, store, enrich and visualise their data. They give decision-makers access to reliable cross-functional data enriched with AI and third-party sources to drive automation and enable more effective decision making.

Enterprise strategies over the past few years have shifted from on-premise to cloud, and from monolithic ERP (Enterprise Resource Planning) solutions to best of breed SaaS (Software as a Service) solutions. Whilst this enables functions to be more autonomous and less dependent on IT functions, it has resulted in enterprise data that is more siloed than ever. Snap Analytics specialises in helping organisations that run SAP to connect and democratise the power of their SAP data by leveraging the modern data stack.

Contributors

Jan van Ansem

(Head of SAP Practice)

"Jan van Ansem is a data warehouse expert with a focus on architecture, strategy and capability development. He specializes in data strategy for organisations with a large SAP footprint – either by using SAP technology or cloud data platforms like Snowflake. Jan has an authentic passion for technology. He enjoys helping organisations to understand trends and innovations in data warehousing, and ensure that enterprises get the best value out of their data and analytics spend.

Paul Johnson

(Head of Data Engineering)

"Paul is an accomplished Data Engineering and Analytics professional with over a decade of experience. Starting as a data engineer and architect, he later led Matillion's data integration portfolio as Director of Product, driving the company to unicorn status. Before becoming Head of Product, he launched the world's first data sharing platform at Bobsled. Paul is recognized as an expert in data integration and ETL for cloud data warehouses. He is a sought-after speaker at major events like AWS re:Invent and Big Data LDN."

A quote from the CEO of Snap - Dave Rice

"SAP is complex and due to limited access to skills and legacy technology, companies are struggling to unlock the power of this data. By fusing the rich data you have in SAP with data from external sources, AI and other SaaS applications you can unlock this tremendous potential. Utilising the modern cloud stack with security at it's core and specialist SAP, Snowflake and AI expertise, the opportunities for businesses today are truly endless! We love seeing the amazing results our clients have achieved by connecting data, technology and teams to maximise the potential of SAP and non-SAP data, together in the cloud. I hope this eBook provides some useful information to guide you in unlocking your SAP data using Snowflake"

Chapter One:
Introduction

In today's data-driven world, organizations are relentlessly pursuing ways to improve their data management and analytics capabilities. Amidst the overwhelming flow of information, these organizations stand at the crossroads of innovation and obsolescence. With every byte of data lies an opportunity: to gain a competitive edge, to unearth novel insights, or to redefine entire industries. But harnessing this power is no small feat.

The journey from raw data to actionable intelligence is strewn with challenges, from the technical intricacies of data storage to the complexities of analysis. And at the heart of this journey is the complex nature of SAP, its underlying data structures and the challenges to extract, load and transform in Snowflake.

A Brief History of Data Warehousing on SAP

In the latter half of the 20th century, as businesses expanded and their operations grew complex, the need for effective data management systems became paramount. Data warehousing, a concept introduced in the late 1970s, emerged as a solution to address this need. By the 1990s, these warehouses became an integral part of enterprise IT infrastructure, paving the way for companies like SAP to profoundly influence the landscape.

The initial data warehousing concepts, put forth by pioneers like Bill Inmon and Ralph Kimball, set the stage for what would become a multi-billion-dollar industry. SAP, initially a player in the enterprise resource planning (ERP) space, recognized the potential of this trend early on. With its Business Information Warehouse offering introduced in 1998, SAP began its journey into the world of data warehousing.

In 2004, with the release of SAP NetWeaver, Business Information Warehouse became SAP Business Warehouse (SAP BW). The product received another significant update in 2013 when SAP customers could configure SAP BW to run on SAP HANA, an in-memory database. This offered several benefits, such as a simpler infrastructure and increased performance. In 2016, SAP BW/4HANA was released as a new product. SAP BW/4HANA was a simplified and optimized evolution of BW tailored specifically for SAP HANA – but its core was still built on the same ABAP technology stack.

With SAP being an early mover in offering a single platform for all data warehouse capabilities, and by delivering rich 'business content' for analytics based on the SAP ERP system, SAP managed to gain a large market share in data warehouse solutions in the enterprise world. For years, the customer base for SAP BW remained stable: The solution no longer appealed to new customers due to its complexity and aging technology, but existing companies often choose to stay with SAP BW because they had built so much content which would be difficult to migrate to a different technology stack.

SAP finally broke with tradition and announced a cloud-based data warehouse in 2019. The SAP Data Warehouse Cloud (since renamed to SAP DataSphere) offers data warehouse capabilities, powered by SAP HANA, as a service.

The Rise of Cloud Data Warehousing

In recent years, there has been a growing trend towards cloud data warehousing. This is due to several factors, including the increasing availability of cloud computing resources, the falling cost of cloud storage, and the growing

demand for agility and scalability.

Cloud-based data warehouses, in contrast to their traditional on-premises counterparts, offer superior scalability, allowing organizations to seamlessly adjust capacity without the burden of procuring new hardware or software. This flexibility typically translates to cost-effectiveness, as expenses are tied only to the resources consumed. Furthermore, these modern warehouses often surpass traditional ones in security, a top priority for any enterprise.

At the forefront of this rising trend has been Snowflake, the leader in cloud data warehousing and now supporting non-traditional use-cases with the Snowflake Data Cloud.

Snowflake is a popular alternative for organizations that want to use SAP data to drive business decisions and perform advanced analytics. Let's examine why...

Chapter Two:

Why SAP Customers are Choosing Snowflake

While SAP products, especially its data warehousing solutions, have been dominant forces in the enterprise software space for decades, there's a growing trend of businesses exploring alternatives. One of the top contenders that have emerged in recent years is Snowflake. This chapter will dive into the pain points experienced by SAP users and explore why Snowflake is increasingly being chosen as the remedy.

Challenges Faced by SAP Customers

While SAP data warehouse solutions many advantages, there remain notable challenges inherent to its architecture and the broader SAP ecosystem:

Rigid Design: One of the often-cited criticisms of SAP products is its rigidity. Unlike more flexible systems, making changes to the data model, or adjusting to evolving business needs can be cumbersome in SAP. This rigidity often results in extended timelines for implementing changes or incorporating new data sources.

Integration with Third-Party Software: SAP's ecosystem is, in many ways, a walled garden. Integrating with third-party applications, be it data sources, ETL tools, or other analytics platforms, is not always straightforward. While there are connectors and integration methodologies available, they are often not as seamless or efficient as users would prefer.

Cost Implications: One of the major hurdles with SAP is the total cost of ownership. SAP's pricing structure, particularly for its data warehousing solutions, can be a significant point of contention for many enterprises. The high licensing costs, coupled with the expenditure for infrastructure and maintenance, make it an expensive proposition for many companies, especially when compared to cloud-based solutions that adopt a more scalable pricing model. It's not just the licensing, but the esoteric nature of SAP systems means businesses often require specialist skill sets to manage, optimize, and troubleshoot the system. These SAP specialists are in high demand, which further increase operational costs.

Snowflake: Addressing the Challenges

Snowflake's Data Cloud has been able to address these challenges through a combination of architectural design and rich feature set.

Shared Storage and Shared-Nothing Benefits

This combination means that data is stored centrally (shared storage) but can be computed and processed by multiple, independent computational units (shared-nothing). This structure ensures that data is accessible and consistent across the platform while simultaneously allowing for high concurrency and performance.

Flexibility with Semi-Structured Data

Traditional systems often struggle or require complex processes to handle semi-structured data. Snowflake shines here. It natively supports semi-structured data types like JSON, Avro, and Parquet. This support means businesses can ingest, store, and directly query semi-structured data without cumbersome transformations, providing an agile response to varied data sources.

Automated and On-the-Fly Scaling

Snowflake's Virtual Warehouses can automatically scale, vertically or horizontally, based on workload. This feature ensures optimal performance even with varied workloads, without manual intervention.

The on-demand scaling aligns costs with actual usage, ensuring that businesses only pay for the computational power they consume.

Broad Third-Party Ecosystem

One of Snowflake's philosophical differences to SAP is its commitment to extensive integration. This gives SAP customers wider choice for best of breed tools in a wide variety of areas including ETL, BI & Analytics and AI & Machine Learning.

Data Sharing

All data in Snowflake can easily be shared in a secure way, directly from the Snowflake platform. Suppliers and customers can be given access to the specific data slices that you want to share with them and third parties can make data instantly available for you, through the data sharing function on Snowflake. In addition, you can effortlessly integrate data by commercial data providers (such as market research institutes or put your own data up for sale in Snowflake through the Data Exchange.

Usage-Based Pricing

Unlike the high fixed licensing costs associated with traditional data warehousing solutions like SAP BW and HANA, Snowflake has adopted a usage-based pricing model. This means enterprises pay for the actual compute and storage they use. That means instead of large upfront investments, costs scale with usage. This is particularly beneficial for businesses with fluctuating workloads, as they don't overpay during periods of lower activity.

For budgeting and financial planning, Snowflake's tagging capabilities give clear visibility into costs, allowing for more predictable budget forecasting.

Automation and Ease of Use

Snowflake has been designed with user-friendliness in mind. Many tasks that require specialist skills in systems like SAP are automated in Snowflake. This not only cuts down on operational costs but also speeds up processes and allows companies to eliminating the risk of a dependency on highly specialized expertise.

Snowflake can address the common complaints of a more rigid architecture and design through a combination of its cloud-native architecture, allowing separation of storage and compute and through compelling features that allow for greater flexibility such as support for non-structured data and automatically scaling based on workloads.

By focusing on a philosophy of openness and integration, Snowflake provides a compelling answer to many of the integration pain points experienced by SAP customers. The platform's data sharing and broad ecosystem support ensure that businesses can maximize the value from their SAP data while benefiting from the flexibility and innovation that comes with a wide array of third-party tools.

With a usage-based pricing model and features that support robust financial governance Snowflake customers have a transparent and auditable cost structure that provides greater financial agility. The cloud services layer and

ease of use, further mitigate the costs associate with niche expertise require to build and maintain solutions.

Comparative Overview

It is worth noting that SAP's own solution, SAP Datasphere, has emerged to address many of the challenges SAP customers face. While the Datasphere looks promising and looks to offer many of the features that are present in Snowflake today, it is currently less mature when compared to Snowflake. As the technology landscape evolves, it's essential for decision-makers to weigh the advantages and challenges of each platform to make an informed choice.

The following comparison table is based on our own experience and expertise.

	SAP BW/4HANA	SAP Datasphere	Snowflake
Deployment	Self-hosted	Cloud-native	Cloud-native
Pricing	Upfront license based on capacity	Upfront license based on capacity	Usage based
Ease of use	More complex, requires highly expert SAP skills	Easier to use, SAP skills are required	Easier to use, requires standard SQL and/or Python skills only
Resource usage – compute	Everything shared. Changes to memory requires lengthy planning process and is costly. You can only scale up/out, never down.	Some control of compute allocation by user group. Increasing overall available compute is costly/lengthy process. Scale down is not possible	Complete flexibility. Almost limitless scale up and out in seconds and can be scaled down and clusters reduced in seconds.
Scalability – Storage	Typically uses standard Memory /Storage ratios. Non-primary storage needs to be architected and is costly to implement	Comes with native datalake capability (up to 90TB / tenant). So far, integration with Databricks is the only non-SAP integration for storage.	Leverages the virtual limitless storage capacity of the Hyperscaler of choice.
Features	More mature platform	Less mature platform	More mature platform
ETL Tool Support	SAP Native tools + some third-party support	SAP Native tools + some third-party support	Rich partner ecosystem, providing a wide offering of different ETL options.

Chapter Three:

Discovering Use Cases for SAP data

Seamless integration between SAP and Snowflake requires strategic planning. With the right foundation, companies can harness the power of Snowflake to leverage their SAP data in more advanced and efficient ways. This chapter explores the initial steps that need to be taken to ensure that all analytics use cases are identified and a plan is created to deliver these in a Snowflake environment

Understanding Analytics Use Cases

Start by assessing the analytics use cases supported by the existing SAP solution. Additionally, recognize new analytics use cases vital to the business. This involves:

- Grasping the business processes underpinned by different analytics use cases.
- Identifying the current reports and dashboards, as well as the data contained within.

Understanding the decisions driven by this data.

Define the roadmap for future analytics use cases, using input from the business stakeholders about upcoming changes in their processes and operational systems.

Analyse the 'pain' felt by the business due to shortcomings of the current solutions. There might be low hanging fruit when providing solutions in Snowflake for gaps in the current offering.

By collaborating with key business stakeholders, you can gauge the importance and complexity of each use case. This knowledge will guide the prioritization of use cases in the Snowflake data platform.

Source System Inventory

It's essential to understand the data sources and their interactions:

- Start with the current systems landscape, understanding key components of source system integrations.
- Identify data load frequencies from each source system, the average data volume, and any relevant processes.

There is often important and relevant data existing outside of SAP, with other systems or even individual files. Catalog all data sources and types for integration into Snowflake.

Data Lineage & Data Model

Once the sources of data are known and understood it's important to understand how the data from those source systems ultimately ends up in the final reporting layer:

- Analyse the initial extractors or tables that are used in the initial landing layer of the existing data platform or analytics solution

- Map out how these initial tables travel through the different layers of the data warehouse. Ideally there should be documentation for this but more often than not this information is lacking. In that case, it makes sense to spend some time with the business users to capture their understanding of how data is transformed between the source system and the analytics application.

- Document and estimate the complexity of the data transformations and jobs to move data between each layer of the data warehouse

- Create the high-level data model mapping out the transactions and business processes (facts) and the core business entities involved in these transactions (dimensions).

Planning

Now armed with an understanding of how the analytics use cases are used by the business and how these are technically underpinned the overall implementation can be planned:

- Create a prioritisation matrix to ascertain complexity Vs value. This will help to plan out the order of delivery and estimations for the project

- Determine the resource mix that is needed to deliver the project based on key milestones and times required with dependencies between deliverables factored in

- Identify any key pieces of SAP analytics functionality which is currently used which may not be available out-of-the-box in Snowflake and any technical deliverables that might be required

- Create a decommissioning plan to determine when any existing technology can be decommissioned. This will help to determine the timeline for when benefits will be realised for the project.

Summary

In this chapter, we delved into the intricacies of preparing for a successful integration of SAP data into the Snowflake platform. Key to this process is a deep understanding of your current analytics use cases, the sources of your data, and the quality of that data. We highlighted best practices such as data profiling, cleansing, and cataloguing, and underscored the importance of mapping out a comprehensive implementation plan.

This plan encompasses aspects from scheduling and task allocation to data transformation, extraction, and validation. By adhering to these outlined best practices, organizations can ensure a smooth and efficient transition of their SAP data onto Snowflake. As we move forward, the next chapter will guide you through the practical steps of executing this integration, focusing on data extraction, transformation, and loading, complemented by strategies for thorough validation and testing.

Chapter Four:

Getting data out of SAP

SAP Data Structures and Formats

SAP is a complex system developed over decades, with a diverse range of data types and formats. Grasping these structures and formats lays the groundwork for a seamless integration between SAP and Snowflake.

Originally launched in 1972, SAP's foundational data structures, present in modules such as Finance, Material Management, and Sales & Distribution, remain consistent. During its early development, space optimization was crucial due to database constraints.As a result, many table names in SAP, stemming from German abbreviations, are restricted to a mere 4 characters, with field names limited to 6. This gives rise to seemingly cryptic table names like LFA1, VBAP, and MARA. Considering a standard SAP system houses hundreds of thousands of such tables, it's no wonder constructing intuitive reports on top of them is a formidable task!

In their quest to simplify analytics, SAP provided 'business content extractors'. These extractors, known interchangeably as S-API extractors, SAP data sources, or BW extractors, bridge the gap between the raw SAP tables and data consumers. For instance, leveraging the 'Sales Orders' extractor saves you from wrestling with complex table relationships.

To further aid data analytics, SAP unveiled 'CDS Views' (Core Data Services). Unlike S-API extractors that are rooted in SAP's ABAP programming language, CDS Views utilize SQL with SAP-specific annotations. Not only do they simplify data access, but they also cater to an array of functions including SAP application development, embedded analytics, and data extraction processes.

Common SAP Data Management Obstacles

Navigating the complex data structures and unique formats of SAP presents distinct challenges. Here are some key hurdles faced when managing SAP data:

1. **Data Quality:**

   SAP's intricate nature can make data upkeep challenging, impacting the reliability of analytics and reports.

2. **Data Integration:**

   The proprietary nature of SAP's data formats, coupled with strict licensing terms, complicates the integration with external systems.

3. **Data Transformation:**

   SAP data often requires significant adaptation to suit the requirements of other systems.

4. **Data Security:**

   SAP databases frequently contain confidential information, necessitating rigorous security measures and adherence to data governance standards.

5. **Expertise:**

   The nuanced landscape of SAP makes finding skilled professionals a daunting and often expensive task.

SAP Data Extraction

Seamless integration of data from SAP systems into Snowflake calls for a well-planned data extraction process. Designed to reduce latency and preserve data integrity, effective extraction involves understanding and utilizing various methods based on an organization's specific needs.

Key data extraction approaches:

There's a variety extraction interfaces, each bearing unique strengths and challenges related to SAP system workload, latency, and incurred costs. Complex IT landscapes often necessitate the amalgamation of two or more techniques for optimal outcomes. These approaches can be largely categorized as:

SAP-Supported frameworks and technologies:

- Operational Data Provisioning (ODP)
- Access via the SAP Application layer
- OData interface
- ODBC/JDBC interface
- SAP Replication Server (either trigger or database log-based)

Third-party proprietary solutions:

- Database trigger/log replicators
- Application trigger/log replicators (The specifics of these solutions, while not extensively public, revolve around the assimilation of third-party proprietary software within the SAP ecosystem).

Delving into the interface approach

The act of pulling data from SAP is defined around the combination of the extraction pattern and the communication protocol. The interface approach is the combination of specific extraction patterns and communication protocols.

This combination is derived from the customer specific requirements, SAP system versions and the restrictions defined within the SAP license agreement.

The extraction pattern describes how datasets are prepared for interface to a data platform. The most common patterns are:

- Table extraction: Data is replicated using simple table queries. For transactional data, selection criteria based on creation/changed date can be used to create delta loads.
- Data collections: A semantic layer is used as a source for extraction processes. This semantic layer can be build in ABAP (programmatic) or in SQL
- Change Data Capture: All changes to data are captured in a 'queue' which can be 'pulled' by the data platform. Generic CDC mechanism use database triggers or the database log (both are available for SAP systems). Other options are available, where 3rd party vendors implement their proprietary processes on the SAP system for identifying changes in SAP data.
- The two main communication protocols for connecting a data platform to an SAP application are:
- Remote Function Call (RFC): SAP's proprietary protocol for communication from/to SAP systems.
- OData
- For a direct database connection, the ODBC / JDBC protocols can be used.

Selecting the right interface approach

Selecting the right pattern is key for successful data integration. We recommend the following approach when deciding on an extraction pattern.

1. **Data Source Insight:**

Gain a deep understanding of your organization's data sources and their utilization. Is your organization fluent in SAP tables? Or are semantic models (BW extractors, CDS views) already widely used for analytics?

2. **Performance & Scalability:**

Gauge the scalability and performance of potential patterns. Some might be better suited for granular datasets.

3. **Check supported extraction patterns / communication protocols for your ETL tool:**

Not every ETL tool supports SAP specific objects/interfaces/protocols. Does your ETL tool support RFC? Can it handle CDS views? Call S-API extractors?

4. **Security & Compliance Considerations:**

In a world rife with data regulations, ensuring that the extraction aligns with standards like SOC 2, PCI DSS, HIPAA, and GDPR is non-negotiable.

A glimpse into real-world SAP-to-Snowflake successes:

- At Snap we have real-world experiences across industries that highlight the significance of making an informed choice and there is no one size fits all approach, some examples are:

- An energy provider migrated from SAP ECC to SAP S/4HANA and decided to break away from SAP BW as a data platform. Snap Analytics implemented Snowflake as the new data platform, which now brings together the world of SAP and non-SAP data, seamlessly integrated.

- One of the world's highest ranked professional sports teams partnered with Snowflake ad Snap Analytics to reduce the time-to-market for analytics solutions. With the Snap Analytics Accelerators for SAP data, the task of performing analytics on SAP data turned from 'nearly impossible' to 'easy' within a couple of months after starting the engagement.

- A food manufacturer wanted to break down the silos between the world of SAP data and 'everything else'. Using Snap Analytics consultants with their deep understanding of SAP and modern cloud data platforms, the legacy processes were unpicked, redesigned and optimized for today's state of the art data platforms. Having both SAP data and non-SAP data now integrated in one place has enabled the business to turn into a truly data driven enterprise.

The success of moving SAP data to Snowflake heavily hinges on the chosen interface and extraction pattern. By deeply understanding data sources, assessing performance metrics, and ensuring regulatory compliance, organizations can set themselves up for a successful implementation.

Chapter Five:

Loading & Transforming SAP Data in Snowflake

When building analytics solutions, the integration between systems is paramount to achieving reliable and actionable insights. This chapter delves into the intricacies of importing SAP data into Snowflake and preparing it for analytical queries.

Key Considerations for Data Loading

Choosing the Optimal Loading Method:

Snowflake is versatile and supports multiple data loading methods. Whether it's bulk loading for large datasets, continuous loading for ongoing data feeds, or real-time streaming for instantaneous updates, the key is to identify what aligns best with your organization's data requirements. Look into:

- Volume: How much data are you ingesting?
- Complexity: Does your data come with intricate hierarchies or relationships?
- Latency: How fresh does your data need to be?

Configuring the Loading Procedure:

From establishing the link between the SAP data source and Snowflake to defining the loading rules, this stage is all about optimization. Improve loading performance your data loading by:

- Utilizing parallel processing to divide and rule.
- Optimizing network bandwidth for faster transfers.
- Making informed decisions on specific loading options for your dataset.

Validating Your Data Load:

As the adage goes, garbage in, garbage out. A core component of this process is data validation. Integrate checks and balances during the loading stage. By nipping data quality issues in the bud even before the migration begins, you're ensuring a smoother analytical process down the line.

Data Transformation

SAP tables are highly normalized and interconnected. To extract value from it, it's essential to transform this data in a way that's ready for analytical workloads.

Navigating Complex SAP Data Structures:

Challenge:

Business objects, such as sales orders or stock levels, often made up of numerous tables. Deciphering and transforming these underlying tables, especially with their intricate naming patterns, into an analytics-ready data model can be a challenging.

Solution:

Use SAP provided metadata where available, for example by using CDS views or S-API extractors. You can also familiarize yourself with the SAP data model specific to the modules in use. SAP's data dictionary provides a wealth of information for those who know how to use it.

If you're looking to fast-track this process,

consider seeking assistance from specialists or leveraging accelerators from Snap Analytics.

Customizations & Z-tables:

- Challenge: Many SAP implementations have custom tables (often starting with the letter "Z") and fields added to standard tables. These customizations can vary between different SAP installations.

- Solution: Work closely with SAP developers and system administrators to understand the purpose and structure of custom objects. Maintain documentation of custom fields and tables for future reference. Check if the custom fields are already included in extractors or CDS views.

The act of loading and transforming data is more than just a technical procedure; it's an interplay of understanding the source system, the target platform, and the underlying business processes. This chapter emphasized the need for a deep knowledge of these elements to successfully build a transformed data model that is ready for analytics. As emphasized in the last chapter on choosing the right extraction method, making informed choices regarding data loading and transformation stands central to the success of your SAP to Snowflake implementation.

Chapter Six:

Optimizing Workloads on Snowflake

In this chapter, we'll provide guidance for optimizing workloads in Snowflake. The way we will approach this is to first consider how will set up our tables and views for efficient performance and then how do we optimally assign the compute resources to execute those queries.

Setting up tables for efficient query performance

- Leveraging Micro-Partitions in Snowflake: Micro-partitions are small, self-contained units of data that are stored together on the same physical storage node. This makes it possible for Snowflake to quickly access and process the data in micro-partitions. They play a pivotal role in ensuring efficient query performance by narrowing down the data segments to be scanned.

A cost-effective approach to maximizing the benefits of micro-partitioning, without incurring additional costs associated with clustering keys, is to naturally cluster your data upon ingestion. For instance, with sales order data, an inherent sequence might be the transaction date. Thus, when inserting this data into Snowflake tables, if we order by the transaction date we ensure that Snowflake partitions the data in alignment with its inherent order, optimizing query performance.

- Defining Clustering Keys in Snowflake: Snowflake can automatically re-organise data at its storage layer based on designated clustering keys. While this introduces an added cost, its benefits can be seen when querying very large tables. Additionally, clustering keys are not just confined to tables; they can also be applied to materialized views. Clustering materialized views becomes advantageous when faced with varied data access patterns necessitating distinct clustering keys. When employing clustering keys on materialized views, it's advisable to remove the clustering keys on the underlying base table.

- Materialized views: A materialized view is a pre-computed data. Because the data is pre-computed, querying a materialized view is faster than executing a query against the base table of the view. This performance difference can be significant when a query is run frequently or is sufficiently complex. As a result, materialized views can speed up expensive aggregation, projection, and selection operations, especially those that run frequently and that run on large data sets. It's recommended to use materialized views when the results are a small number of records relative to the base table and the base table does not change frequently.

- Search Optimization Service: The search optimization service can significantly improve the performance of certain types of analytical queries. It is particularly useful when performing lookups on VARIANT or GEO data types or point look-ups where the predicate is not a clustering key.

Optimally configuring Virtual Warehouses

- When to use Clustered Virtual Warehouses: Clustered Warehouses are designed for

high-concurrency workloads. They allow you to automatically scale out by adding more clusters to handle multiple simultaneous queries. If you expect many users or processes to query your data simultaneously, clustered warehouses distribute the load.

- When to select the right clustering mode: Clustered Virtual Warehouses can be set to operate in Standard or Economy Mode.
    - Standard Mode: Prioritizes performance. It resumes and suspends faster but incurs a bit more cost.
    - Economy Mode: Prioritizes cost savings. Warehouses might take a bit longer to resume, but they are more cost-effective for infrequent workloads.
- Query Acceleration Service: The query acceleration service can accelerate parts of the warehouse workload by offloading portions of the query processing work to shared compute resources that are provided by the service.

Examples of the types of workloads that might benefit from the query acceleration service include:

- Ad hoc analytics.
- Workloads with unpredictable data volume per query.
- Queries with large scans and selective filters.

The QUERY_ACCELERATION_ELIGIBLE view can help you identify queries and warehouses that might benefit from the query acceleration service.

Snowflake removes a lot of the traditional performance tuning that earlier generations of data warehouses required but there are still tweaks you can make that have a significant impact on performance and cost optimization.

Chapter Seven:

Snap Data Warehouse accelerators

Snap has helped companies with rapidly implementing a data warehouse on Snowflake using our Snap Data Warehouse and ETL accelerators. These accelerators are designed for implementing a standard layered data warehouse solution following industry best practices. The solution is fully configurable, meaning that you only build a data pipeline once, and use meta data to define the scope for regular data load processes. This makes it effortless to increase the scope of your data acquisition into the data warehouse, and reduces the cost of ongoing support and maintenance as all dataflows follow exactly the same process.

Accelerators for SAP data

Snap have developed accelerators for common design patterns used in SAP. Below is an overview of the accelerators which are currently available. Snap is continuously adding to the scope of accelerators for SAP data, so do let us know if you have suggestions for further SAP content on Snowflake.

Accelerator 1: Automatically create meaningful table/field name descriptions

As mentioned before, SAP uses very cryptic table and field names in their database and application. The Snap accelerator automatically creates a virtual layer over the tables acquired from SAP, and over all the tables as they are propagated through the different layers of the data warehouse. This virtual layer uses the SAP data dictionary to provide a description for every field name (in various languages). The solution also offers the flexibility to replace a name from the dictionary with your own description.

This way, you can have meaningful names for all data elements throughout your data warehouse, as well as full data lineage to the source data.

Accelerator 2: Capture time dependent master data or ‘slowly changing dimensions’

The SAP analytics solutions are well known for providing an elegant solution to what SAP calls ‘Time dependent master data’. In data warehousing, this is a very familiar concept, mostly referred to as ‘slowly changing dimensions’. The Snap data warehouse approach enables you to switch on the functionality for slowly changing dimensions for any type of master data – whether this originates from SAP or any other source system.

Accelerator 3: Convert SAP hierarchies to a generic BI hierarchy

SAP hierarchies are extremely flexible and support many different hierarchy varieties. Unfortunately, they are very difficult to use by any other tool than SAP tools. The Snap data warehouse comes with an accelerator for converting SAP hierarchies into a generic hierarchy which can be used in any data & analytics tool.

Accelerator 4: Data type conversion

SAP uses proprietary data types. These need to be converted to the more generic data types used by cloud data platforms, such as Snowflake. The Snap accelerator has a pre-defined mapping table to ensure all data types are correctly converted, without loss of information.

Accelerator 5: SAP data model and business content

Having done numerous SAP to Snowflake implementations, the team is now equipped with a standardised dimensional model for SAP data, including standard data transformation. The SAP model is currently set up around SAP Finance with the core dimensions such as Account hierarchy, Profit and Cost Center hierarchies and Planning/Forecast versions. The model also includes the main dimensions for subledgers including material, customer and vendor. Further content is on the roadmap, with standard content for purchasing and sales planned for 2024.

Making SAP data useful in Snowflake does not need to be challenging. Timelines can drastically be reduced by using the Snap accelerators for out-of-the-box models for complex SAP structures, and by using the people with years of experience with analytics on SAP data.

Chapter Eight:
Real world case study

The food manufacturer which saved 500+ days of manual work and managed a 6-figure cost saving.

A leading, UK based food manufacturer suffered from an aging SAP data platform. There was a lack of actionable insights and data was locked away in SAP and non-SAP business applications.

Snap Analytics helped to implement a connected data stack to bring together all data sources into one Cloud Data Warehouse – in a cost effective way and with the ability to easily extend the scope of the data warehouse without the need for source application experts.

In the initial phase, Snap helped identify the solutions required to achieve the vision of a modern data stack. According to the IT Director:

"The right technology partner has to have the right knowledge, but also be a good cultural fit and work with us as part of our team. We needed some thought leadership, and we needed to be challenged, and we got all of those things with Snap."

Three core objectives took centre stage:

1. The data platform will deliver new insights to a wide variety of areas in the business. The solution is generic for the entire business, and department specific adaptations will only be allowed in departmental data marts.

2. Improve efficiencies by automating repeatable tasks

3. Reduce risk associated with manual data practices

The biggest benefit of the new data platform is the simplicity of the end-to-end solution. There is no longer a need for expert users in proprietary technologies. Analysts can easily be trained on exploring the data themselves. The solution is agile and supports the team's "pilot, prove, scale" approach, allowing new solutions to quickly find their way to the business users.

The results?

- Delivered 15 different data solutions across the organization, including customer service, manufacturing, sales reporting, procurement and product development.

- Saved over 500 days of manual work and hundreds of thousands of pounds through decommissioning unnecessary data activities.

- Stakeholders across the entire business are now keen to join the in with the centralised data platform and get their use cases implemented.

Going forwards, there are plans to explore new ways of connecting data across departments and gaining strategic insights from its new data platform. Now the company's prospective data users have seen what's possible, the floodgates have opened in terms of demand, and everyone is eager to take advantage.

"This is only the beginning in terms of what's possible for us," says the IT director. "Realizing that we can do things that took months, in a matter of days with the new data platform was the spark that led to where we are now. We can see endless possibilities ahead."

Driving innovation together!

We believe in the power of collaboration to help our customers achieve their goals using their data. Our strong partnerships with leading technology providers help us bring the best ideas and solutions to our customers.

Contact us today to get started:
Email: hello@snapanalytics.co.uk
Phone: + 44 333 305 7772
Web: snapanalytics.co.uk


Snap
ANALYTICS