INDIUM

Modern Data Warehousing with Snowflake

A Whitepaper

Introduction

The need for data, and even access to data, is on the rise. But, setting up a data warehouse at scale is at times expensive, and it can deter businesses from leveraging the data to gather insights.

Snowflake, a cloud-native platform, is a disruptive solution that provides businesses with a cost-effective solution to create data marts, data lakes, and data warehouses and share data securely across the organization.

Snowflake can be built on top of AWS, Azure, and GCP infrastructure. It does not require hardware or software to be selected, installed, configured, or managed.

As a result, organizations can save on resources for setting up, maintaining, and providing support for in-house servers. Organizations can also use an ETL solution like Stitch to move data into Snowflake easily.

Snowflake's architecture allows storage and compute capabilities to be decoupled in a way that the two can scale independently.

This lowers costs for customers who pay-per-need. Snowflake also allows organizations to share secure, governed data quickly, in real-time.

Autodesk - A Case Study

Autodesk, a US-based software giant, found its existing data lake architecture to be difficult to operate and expensive to scale and maintain. It was using homegrown solutions for data ingestion and found the data to be unreliable and the code needing constant troubleshooting. Security and governance were also becoming a challenge. The company needed scalability, performance, and speed, and it found Snowflake to be the ideal fit.

Once Autodesk transitioned to Snowflake, its employees gained access to reliable data which improved BI performance and enhanced customer experience.

Snowflake's extensive network of connectors, drivers and programming languages helped the organization build less complex end-to-end data pipelines. Snowflake also reduced the administrative work due to zero maintenance requirements, thus freeing up IT staff to focus more on analytics.

Snowflake Architecture

The Snowflake architecture is made up of three layers - storage, compute, and services - and each can be scaled independently.

Storage: All structured, semistructured, and unstructured data is stored in the storage layer and and all aspects such as organization, file size, compression, structure, metadata, and statistics are automatically managed by Snowflake.

Compute: Data processing tasks to respond to queries are executed in the compute layer, which is made up of virtual warehouses. Each individual or cluster of virtual warehouse(s) access data in the storage layer and use it to execute queries independently, thus avoiding sharing or competing for compute resources.

As a result, nondisruptive, automatic scaling becomes possible even when the queries are running, without redistributing or rebalancing the data in the storage layer.

Services: Automating the coordination of the entire system using ANSI SQL, elimination of manual management and tuning of the data warehouse takes place in the cloud services layer. The service includes:

- Infrastructure management
- Metadata management
- Authentication
- Query parsing and optimization
- Access control
- Delivering data to Snowflake from more than 90 sources

Benefits of Snowflake

Snowflake is a cloud-native platform which helps organizations overcome problems found in older legacy-based data warehouses, including limited scalability, issues with data transformation, and large volumes of queries which cause delays or failures.

The five benefits of Snowflake for businesses include;

Performance and Speed: Snowflake allows businesses to take advantage of cloud's elasticity by enabling faster data loading or running larger volumes of queries by scaling up virtual warehouses and leveraging extra compute resources. When the need is fulfilled, organizations can scale it down and pay only for the resources used.

Supporting Structured and Semi-structured Data: It allows organizations to combine structured and semi-structured data to analyze and load into the cloud database without converting or transforming to a fixed relational schema. Snowflake automatically optimizes data storage and querying.

Accessibility and Concurrency: Concurrency can be an issue in a traditional data warehouse when a large number of users compete for resources or use cases try to access it parallelly. Snowflake's multicluster architecture enables concurrency, allowing multiple queries to be processed simultaneously. It allows each virtual warehouse to scale up or down based on need. It enables data analysts and data scientists to access the data they need anytime without waiting for the completion of other loading and processing tasks.

Seamless Data Sharing: Snowflake enables users to access data seamlessly with other data consumers, even those outside the Snowflake ecosystem, by creating reader accounts directly from the user interface.

Availability and Security: By being distributed across the platform's availability zones, Snowflake ensures sustained operations even during component and network failures, thereby improving service uptime. This ensures higher level of security as it is SOC 2 Type II certified, supports PHI data for HIPAA customers, and enables encryption across all network communications.

Analytics and Insights: Databases such as PostgreSQL and MySQL are row-based and can record transactions quickly. They are especially useful in applications where customer sales or other similar transactional data needs to be recorded at high speeds. Snowflake is a columnar database and is optimized to analyze large volumes of data at high speed. This makes Snowflake ideal for Business Intelligence.

Cloning: Snowflake cloud data warehouse enables the creation of a copy of any Snowflake objects such as tables, database and schema instantly, in near real-time. It is a metadata operation where the data is not duplicated and is used for generating the test environment from the production database.

Indium for Building Modern Data Warehouse Using Snowflake

Indium is a data engineering company with capabilities and experience in various data management solutions and platforms. The company's team of experts work closely with the customers to understand their business objectives and design and deploy solutions that break barriers to innovation.

We can help businesses leverage Snowflake architecture to build a modern data warehouse to accelerate growth and improve digital experience. To know more, visit: https://www.indium.tech/?s=snowflake

FAQs

What kind of data warehouse modeling does Snowflake support?

Snowflake's platform is versatile and supports different types of data modeling approaches with equal ease. Online analytical processing (OLAP) is a foundational part of Snowflake's database schema.

INDIUM

About Indium

Indium is an AI-driven digital engineering company that helps enterprises build, scale, and innovate with cutting-edge technology. We specialize in custom solutions, ensuring every engagement is tailored to business needs with a relentless customer-first approach. Our expertise spans Generative AI, Product Engineering, Intelligent Automation, Data & AI, Quality Engineering, and Gaming, delivering high-impact solutions that drive real business impact.

With 5,000+ associates globally, we partner with Fortune 500, Global 2000, and leading technology firms across Financial Services, Healthcare, Manufacturing, Retail, and Technology–driving impact in North America, India, the UK, Singapore, Australia, and Japan to keep businesses ahead in an AI-first world.

USA	INDIA	UK	SINGAPORE
Cupertino | Princeton Toll-free: +1-888-207-5969	Chennai | Bengaluru | Mumbai Hyderabad | Pune Toll-free: 1800-123-1191	London Ph: +44 1420 300014	Singapore Ph: +65 6812 7888

www.indium.tech

For Sales Inquiries
sales@indium.tech

For General Inquiries
info@indium.tech