**Streamlining Data Management with Snowflake: DoorDash**

INDUSTRY CASE STUDY

horizontal line

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

# DATA 225 DATABASE SYSTEM FOR ANALYTICS

Chelsea Jaculina, Manjot Singh, Sai Prasad Thalluri

Group 4

April 24, 2024

# Table of Contents

[**Abstract 3**](#_tt7veohh8ndp)

[**Introduction 3**](#_arolcxe0i15c)

[**DoorDash 4**](#_cjh0k4nv7w4x)

[**Challenges Faced by DoorDash 4**](#_b1uwrxjgtdb2)

[**DoorDash’s Legacy Data Infrastructure 4**](#_bf65kbl1m8wm)

[**Data Warehouse 5**](#_8ee16hslgxse)

[Key Concepts of Data Warehouse 5](#_fu5arh9ahb0t)

[Benefits of Data Warehouse 6](#_cxdxiigljkm3)

[**Snowflake 6**](#_m2sjhy5vi701)

[Benefits Achieved from Snowflake 6](#_7jn9isowykk)

[Use Cases of Snowflake Particular to DoorDash 7](#_jfcqg2bccjjy)

[Multi-Cluster Data Architecture 9](#_9svtrjl7s8d4)

[Benefits of using Snowflake vs Legacy Data Warehouse **10**](#_lu76292u49ft)

[**Key Learnings 11**](#_qo3qj4tynqqw)

[**Conclusion 11**](#_p9el4v2c24q8)

[**References 12**](#_twf0hs4fhqta)

# Abstract

Doordash, a leading company in the food delivery market, uses data analytics to improve its operations and customer satisfaction. This study explores how Doordash uses data warehouses to make strategic decisions and improve efficiency. The main aim of this study is to analyze how Doordash harnesses the power of Snowflake to gain valuable insights, optimize processes, and drive business growth in the competitive food delivery market while also comparing their use of Snowflake as opposed to their legacy data warehousing method.

*Keywords:* DoorDash, Food Delivery, Data Warehouse, Snowflake

# Introduction

In the fast changing environment of the food delivery market, companies like Doordash have emerged as significant players, revolutionizing the way consumers access and enjoy their favorite meals. At the heart of Doordash's operational efficiency and data-driven decision-making is the use of innovative technologies, such as Snowflake.

Snowflake, known for its scalable and cloud-based design, has become a key component of modern data management systems, allowing enterprises to handle massive amounts of data with efficiency. For a dynamic and rapidly growing company like Doordash, leveraging the power of Snowflake has proven critical in streamlining processes, optimizing efficiency, and gaining useful insights from their data sources.

The study investigates the intricate connection between Doordash and Snowflake, examining how it enables Doordash to drive innovation, improve consumer experiences, and make significant business choices. By analyzing the complexities of Doordash's use of Snowflake, we aim to shed light on the synergies between technology and food delivery services, providing useful insights.

# DoorDash

DoorDash is an online food ordering platform that allows fresh groceries, meals, home goods, and alcohol to be delivered from local restaurants and convenience stores. It also offers takeaway and on-demand food delivery services. In 2013, Andy Fang, Stanley Tang, and Tony Xu launched the firm in San Francisco, California.

With partnerships with over 450,000 merchants, DoorDash offers a vast array of food and drink options to clients in over 4,000 locations in the US, Canada, and Australia.

# Challenges Faced by DoorDash

* **Large Data Volume:** Orders, delivery, customer interactions, and restaurant operations all contribute to DoorDash's massive data volume.
* **Data Silos:** The dispersion of data across several systems impeded the thorough examination of the data.
* **Scalability Limitations:** DoorDash's ever-increasing volume of data was too much for traditional data warehouses to manage.
* **Performance bottlenecks:** Timely insights and decision-making were impeded by slow data processing and querying.

# DoorDash’s Legacy Data Infrastructure

* DoorDash used to store their historical data on a local server until 2020.
* In 2020, DoorDash experienced accelerated growth in their marketplace with more customers ordering on their app, more merchants marketing their products on the app and more delivery drivers, or Dashers as a term coined by DoorDash.
* In order to increase the efficiency of the load and data handling and meet the same Service Level Agreements (SLAs) with their business partners, DoorDash turned to Snowflake instead of increasing their existing compute resources.
* This approach not only increased their ETL process efficiency, but also avoided increasing costs of compute resources.

# Data Warehouse

A data warehouse serves as a specialized database that stores structured, integrated, and historical data from various sources within an organization. It acts as a central repository where data from disparate systems such as transactional databases, Customer relationship management (CRM) systems, and other operational sources are extracted, transformed, and loaded for analytical purposes. By consolidating data into a single location, a data warehouse provides a comprehensive and unified view of the organization's information landscape, enabling users to perform in-depth analysis and gain valuable insights for strategic decision-making.

## Key Concepts of Data Warehouse

* **Integration:** Data warehouses consolidate data from different sources across an organization to create a unified view.
* **Structure:** Data in a warehouse is organized in a structured format optimized for querying and analysis.
* **Historical Data:** Data warehouses store historical information, allowing users to analyze trends and patterns over time.
* **Query and Analysis:** Data warehouses support complex queries and analytics for deriving insights and making informed decisions.
* **Business Intelligence:** Data warehouses are often used with business intelligence tools to visualize data and conduct advanced analytics.
* **Data Quality:** Maintaining data quality is crucial in data warehouses to ensure accuracy and consistency.
* **Scalability and Performance:** Data warehouses are designed to handle large volumes of data efficiently and are optimized for fast query processing.

## Benefits of Data Warehouse

* **Improved Decision-Making:** Data warehouses provide a consolidated view of data, enabling decision-makers to access accurate and timely information for strategic planning and decision-making.
* **Enhanced Data Quality:** By integrating and cleaning data from multiple sources, data warehouses help maintain data consistency, accuracy, and integrity, improving overall data quality.
* **Advanced Analytics:** Data warehouses support complex queries, data mining, and advanced analytics, allowing organizations to derive valuable insights and uncover trends that drive business growth.
* **Historical Analysis:** Storing historical data in data warehouses enables trend analysis, forecasting, and comparison of performance over time, aiding in identifying patterns and making informed decisions.
* **Cost Savings:** By streamlining data management processes, eliminating data silos, and improving data accessibility, data warehouses can lead to cost savings through increased efficiency and productivity.

# Snowflake

Snowflake is a cloud-based data warehousing platform provided as SaaS (software-as-a-Service). The platform provides multiple services including storing, managing, and analyzing data. It offers a range of solutions that provides high performance, scalability, ease of use, and the ability to work under multiple loads.

## Benefits Achieved from Snowflake

* **Centralized Data Platform:** Snowflake created a single data lake for thorough analysis by combining data from several sources.
* **Improved Scalability:** Elastic scalability is made possible by Snowflake's cloud-based architecture, which can handle DoorDash's increasing data volume.
* **Improved Performance:** DoorDash was able to produce real-time insights because of quicker data processing and querying.
* **Streamlined Operations:** Snowflake made data management easier, saving time on data integration and maintenance.
* **Data-Driven Decision Making:** DoorDash gained a competitive edge, optimized delivery routes, and enhanced customer satisfaction because of easier access to information.

## Use Cases of Snowflake Particular to DoorDash

* **Client Segmentation and Targeting:** DoorDash may segment its client base for tailored marketing efforts and suggestions by examining customer behavior and preferences inside Snowflake.
* **Delivery Optimization:** To improve delivery routes and guarantee quicker delivery times, Snowflake makes it easier to analyze real-time traffic data, delivery routes, and driver positions.
* **Fraud Detection:** By helping to spot questionable payment trends and fraudulent activity, Snowflake's analytics capabilities improve security for DoorDash's clients as well.
* **Restaurant Performance Analysis:** To find opportunities for cooperation and development with partner restaurants, DoorDash may use Snowflake to evaluate restaurant performance indicators including wait times, customer satisfaction scores, and order volumes.
* **Machine Learning Integration:** According to the case study, DoorDash makes use of Snowflake's machine learning tool integration. As a result, they are able to develop predictive models for demand forecasting, resource allocation optimization, and seamless operations.

**Effect of Snowflake on DoorDash’s Data Infrastructure**

* **Optimized ETL Processes:** In order to optimize ETL processes, DoorDash decommissioned any redundant ETL processes by tracing the DAG (Directed Acyclic Graph) downwards to determine the impact of the removal.
* **Resource Availability:** By decommissioning unnecessary ETL jobs, compute resources can be freed up for other jobs, reducing overall compute costs and helping to meet SLAs.
* **Meeting Time Constraints:** In case any ETL process could not be decommissioned, DoorDash resorted to breaking down its dependencies by analyzing the DAG and identifying any unnecessary dependencies that can be removed without impacting the job. This allowed them to meet SLAs by allowing the job to complete before a specific time.
* **Incremental ETL Processes:** DoorDash optimized their ETL processes by converting them to incremental processes, which would only insert or update changed records instead of every single record. This technique avoided reprocessing of data that had not changed and improved performance.
* **Routine Checks:** DoorDash performed routine checks on the columns used in the ETL processes in order to remove any unnecessary columns and reduce the amount of data being processed.
* **Clustering:** DoorDash implemented clustering keys on large tables in order to reduce partition scans and improve query performance. They also analyzed frequent queries and joins to determine the appropriate dimension for clustering.
* **Examples of Snowflake functions used:** DoorDash provides an example of using HAVERSINE and QUALIFY functions in Snowflake. HAVERSINE function allowed them to calculate the distance between two points while QUALIFY allowed removal of duplicate values from queries

## Multi-Cluster Data Architecture

Snowflake not only differentiates itself from traditional data warehouse solutions through being a cloud-based data warehouse platform, but through its innovative multi-cluster architecture. Snowflake’s multi-cluster, shared data architecture (Figure 1) separates compute resources from data storage. This means that while queries are running, compute resources are able to scale without the need to redistribute the data. Therefore, Snowflake can automatically detect scaling when needed without any user or admin involvement (Snowflake. (2023, Jan 30)

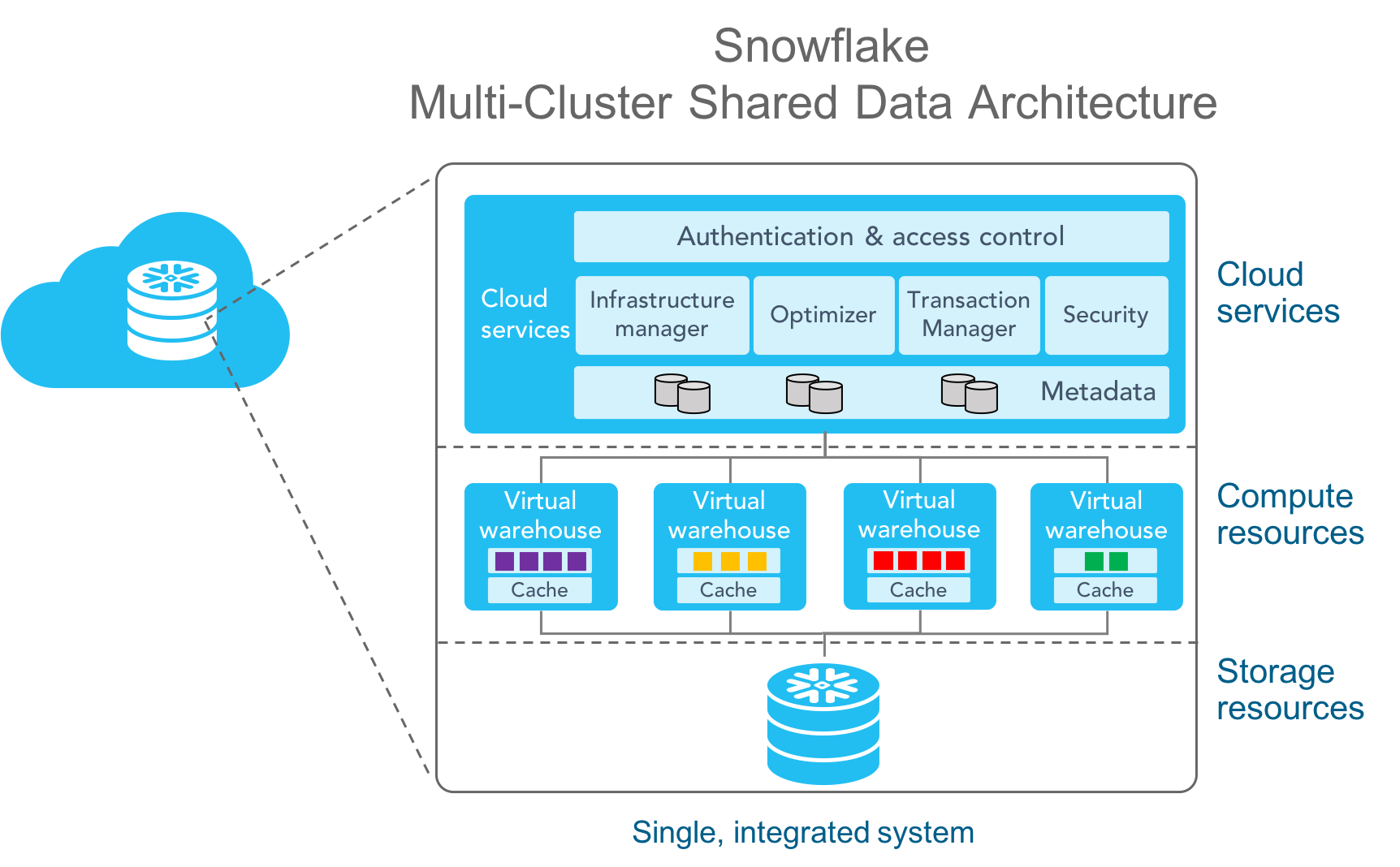


Figure 1.

## Benefits of using Snowflake vs Legacy Data Warehouse

DoorDash outgrew its initial data warehouse solution and adopted Snowflake's cloud-based platform. This switch to Snowflake’s multi-cluster data architecture unlocked several key benefits as we can see in the comparison below:

| Comparison | Traditional  Data Warehouse | Snowflake |
| --- | --- | --- |
| Scalability | Limited scalability | Highly scalable storage and processing based on demand |
| Data Handling | Struggles with various data types | Handles diverse data types from multiple sources (transactions, customer data, operational data) |
| Querying | Complex queries | Uses familiar SQL queries for easier data analysis |
| Security | Limited security features | Advanced security features like encryption and access controls |
| Cost | High upfront costs and ongoing maintenance | Pay-as-you-go model,  reduces costs |
| Performance | Slower performance with large data volumes | Fast data processing and analysis |

# 

# Key Learnings

This case study of DoorDash and Snowflake showcases how a cloud-based data warehousing solution can empower businesses to effectively manage big data, gain actionable insights, and optimize operations for growth.

**Snowflake's Cloud-Based Advantage:** The case study illustrates how cloud-based data warehousing can manage enormous data volumes in a flexible and scalable manner.

**Advantages Beyond Scalability:** Although scalability is important, Snowflake also offers enhanced performance and streamlined data management, which promote quicker insights and improved decision-making.

**Better Analytics Capabilities:** Businesses like DoorDash can use Snowflake's centralized data to do a variety of analytical activities, which in turn spurs corporate growth.

The success of these technologies is driving industry adoption, creating a demand for employees with relevant expertise. Therefore, understanding the topics and concepts we are studying in this course becomes crucial as they are instrumental in shaping the future of many industries.

# Conclusion

In the fast-paced world of food delivery, DoorDash relies on real-time data insights to optimize operations and ensure customer satisfaction. Traditional data warehouses struggle to handle DoorDash's massive and ever-growing dataset. By adopting Snowflake's cloud-based data warehousing platform, DoorDash gained the scalability and agility needed to manage their data effectively. Snowflake's ability to store, process, and analyze data in real-time empowers DoorDash to make data-driven decisions that improve delivery routes, personalize customer experiences, and ultimately drive business growth.

# References

* *Key Concepts & Architecture — Snowflake Documentation*. Docs.snowflake.com. (2022). Retrieved 24 April 2024, from <https://docs.snowflake.com/en/user-guide/intro-key-concepts.html>.
* *Oracle.com*. (2022). What is a Data Warehouse?. Retrieved 24 April 2024, from <https://www.oracle.com/database/what-is-a-data-warehouse/>
* *RTS Labs*. (2023, April 4). Streamlining data management with Snowflake: A case study. Retrieved 24 April 2024, from <https://rtslabs.com/streamlining-data-management-with-snowflake-a-case-study/>
* *Snowflake*. (2018, July 18). 5 reasons to love Snowflake's architecture for your data warehouse. Retrieved 24 April 2024, from <https://www.snowflake.com/blog/5-reasons-to-love-snowflakes-architecture-for-your-data-warehouse/>
* *Snowflake*. (2023, Jan 30). How retailers optimize merchandising and assortment planning. Retrieved 24 April 2024, from <https://www.snowflake.com/blog/how-retailers-optimize-merchandising-and-assortment-planning/#:~:text=To%20eliminate%20data%20silos%20and,twice%20as%20fast%20as%20before>.
* *Softweb Solutions*. (2023, December 6). Snowflake data warehousing solutions. Retrieved 24 April 2024, from <https://www.softwebsolutions.com/resources/snowflake-data-warehousing-solutions.html>
* *DoorDash Engineering.* (2021, June 22).Overcoming Rapid Growth Challenges for Datasets in Snowflake. Retrieved 24 April 2024, from <https://doordash.engineering/2021/06/22/overcoming-rapid-growth-challenges-for-datasets-in-snowflake/>