Mastering Snowflake Platform

Generate, fetch, and automate Snowflake data as a skilled data practitioner

Pooja Kelgaonkar

www.bpbonline.com

First published: 2024

Published by BPB Online
WeWork
119 Marylebone Road
London NW1 5PU

UK | UAE | INDIA | SINGAPORE

ISBN 978-93-55519-764

www.bpbonline.com

Dedicated to

My beloved Parents,
Neelkrishna
&
My Daughter ***Anvi***

About the Author

Pooja Kelgaonkar is a distinguished figure in the realm of data, standing as one of the Snowflake data superheroes, boasting nearly two decades of experience in the field. She has honed her expertise through diverse roles on various data platforms and distributed systems, emerging as a specialist in data modernization. She is currently serving as a senior data architect at Rackspace Technology in Toronto. As a key data resource, she leads the data implementations, ensuring seamless integration and optimal performance, ultimately contributing to the success and satisfaction of the clients.

Her commitment to knowledge extends beyond her professional role, as she is also the founder of an edutech platform. She is a strong advocate for knowledge sharing and community growth. Her commitment is evident in her contributions to the tech community through blogs on Snowflake and Google Cloud Platform (GCP). These insightful articles are regularly featured on Google and Snowflake community pages, solidifying her position as a thought leader.

Pooja's passion for community growth extends beyond the digital arena; she thrives as a public speaker, frequently participating in events, conferences, and tech webinars. Her involvement in these platforms allows her to share her wealth of knowledge and insights, contributing to the collective learning of the community. Pooja's multifaceted engagement with data, education, and community development showcases a profound commitment to advancing the field and empowering those within it.

About the Reviewer

Dipika Aher is a passionate data engineer with extensive professional experience in the domain of healthcare and manufacturing. With a skill set that includes Snowflake, DBT, Python, Unix, SQL, AWS, Informatica, Oracle, she is a holder of multiple industry certifications such as the AWS Solutions, Snowflake Certifications and more.

She is currently working as part of a development team and has been using cloud technologies to enhance her Snowflake related knowledge.

She believes that education is the key to empowering people and creating positive change in the world. She uses her expertise and enthusiasm to inspire and motivate others to pursue their passions and goals in the field of data engineering.

Acknowledgement

I extend my deepest gratitude to my parents and family, especially my daughter Anvi, for their unwavering support and encouragement throughout the writing of this book. Their constant belief in me has been my greatest motivation.

I am grateful to BPB Publications for presenting me the opportunity to author a book on mastering Snowflake. My heartfelt thanks go to the entire BPB team, whose collaboration in both editorial and technical aspects has made this journey an enriching experience.

Special acknowledgment goes to my mentor as well as the Snowflake community. They have consistently encouraged me to pursue my aspirations. It is an honor to be part of the Snowflake's elite group that is committed to continuous knowledge and learning.

To all the readers who have shown interest in my book and contributed to its realization, I express my sincere thanks. Your support and encouragement have been truly invaluable in bringing this project to life.

Preface

Data is continuously evolving and hence the world around it. Many have witnessed the transition from Mainframes to the cloud implementations. It is the need of the moment to have a robust, scalable, performance and cost efficient data platform that cater to the ever growing need of the data applications. In the rapidly evolving landscape of data management, Snowflake stands as a powerful platform, and offers unified experience to the customers. This book is your first step to unlocking its full potential.

In this book, you embark on a transformative exploration, blending theoretical insights with practical applications to empower you on your data mastery voyage. Whether you are a newcomer to delve into the world of Snowflake or an experienced practitioner seeking to refine your skills, this guide is crafted to meet you at your current level and propel you forward.

As the book unfolds, you will traverse the intricacies of Snowflake architecture, discover its key features, and navigate the nuances of working with different data types. You will continue the journey with a deep dive into advanced concepts such as data security, governance, and collaboration.

What sets this guide apart is its relation with real time use cases and references from traditional enterprise data architectures along with hands-on approach. Throughout the chapters, you will find practical exercises, real-time use cases, and reference architectures that bridge the gap between theory and application.

Get ready and fasten your seatbelt to embark on a transformative journey with Snowflake. Whether you are aiming to enhance your career, architect data solutions, or simply learn new skills, “Mastering Snowflake Platform” is your trusted companion to get you started.

Chapter 1: Getting Started with Snowflake – The first chapter is the foundation of the book, and it covers the history of the Snowflake, the need to implement Snowflake, and how readers can get started with Snowflake. This chapter guides to set up a trial or demo account to be used for various hands-on activities covered in subsequent chapters. This also covers the overview of Snowflake certifications and various community events.

Chapter 2: Three Layered Architecture – This chapter helps readers to get started with Snowflake with a detailed view of architecture. This covers the various data platform architecture challenges and how Snowflake helps to overcome most of these challenges. This chapter defines 3 layers of architecture, their distinguishing features and setup.

Chapter 3: Data Types, Data Objects and SQL Commands – This chapter focuses on various data types, data objects and SQL commands supported in Snowflake. Snowflake supports ANSI SQL standard. However this chapter guides on DDL, DCL and DML supported in this platform. This chapter also covers unique database objects used and created in Snowflake.

Chapter 4: Data Loading and Unloading – Data loading and unloading is one the most important parts of data platform implementation. This chapter guides you to learn native utilities, commands supported by Snowflake. This covers various commands used to load batch and streaming datasets. This also covers data extraction and sharing with consumer groups.

Chapter 5: Understanding Streams and Tasks – Change data capture is the most critical part of the data platform implementation. This chapter guides you to streams – Snowflake native objects used for change data capture. This also covers native scheduling and orchestrating objects – Tasks.

Chapter 6: Understanding Snowpark – Spark is one of the fastest ways to implement ETL or ELT while implementing data pipelines. This chapter covers Snowpark – which is a native programming language like Spark. It is also a guide on how to understand and implement Snowpark.

Chapter 7: Access Control and Managing Users Roles – Access control is one of the key pillars of data governance. This chapter helps to understand various access control implementations in Snowflake. This also covers role and user management, how users are created, roles are created and maintained. This guides to set up the most important pillar of governance.

Chapter 8: Data Protection and Recovery – Data protection and recovery is the next pillar of data governance. This chapter guides you to understand data protection mechanisms with Snowflake. This also covers various data recovery options in Snowflake. This chapter helps to define appropriate policies, data masking and using the right data recovery technique when needed.

Chapter 9: Snowflake Performance Optimization – This chapter guides to develop an understanding of performance measure, performance optimization techniques and need to implement optimization. This also covers understanding of Snowflake metadata objects that help to measure performance of the platform.

Chapter 10: Understanding Snowflake Costing and Utilizations – Cost is the most critical component of platform design, implementation, usage, and maintenance. This chapter

guides to develop an understanding of cost measure, cost optimization techniques and need to implement optimization.

Chapter 11: Implementing Cost Optimizations – This chapter guides how to implement cost optimizations. This helps to develop an understanding of real time use cases and best practices followed for cost optimizations.

Chapter 12: Data Sharing – Data collaborations, data integrations and integrating with consumer applications is essential for business growth and requirements. This chapter guides how to implement data sharing, secure data sharing with Snowflake as well as non-Snowflake consumer groups. This also helps to understand various sharing options available in Snowflake.

Chapter 13: Data Cloning – Managing environments like DEV, QA, UAT and PROD becomes challenging when a user wants to develop, test their features with greater data volumes or production datasets. This chapter guides how to setup environments using data cloning features and helps to understand how easy it is to clone and maintain copies, data across environments.

Chapter 14: Understanding Snowsight – Earlier Snowflake offered two versions of web based user interfaces – Classic console and Snowsight. Now with latest sign ups, Snowsight is the only UI made available to the users and all earlier accounts migrated to Snowsight. This chapter helps to understand Snowsight features and distinguishing factors of the interfaces. This also covers the intuitive dashboarding feature of Snowsight.

Chapter 15: Programming Connectors and Drivers – Snowflake offers various connectors, drivers to work with a set of programming languages. This chapter helps to understand the various connectors and drivers. This also guides to setup sample drivers and Snowflake native command line interface – Snowsql.

Chapter 16: Workload Patterns with Snowflake – Snowflake is data on cloud and supports various data implantations. This chapter helps to understand Data Warehouse, Data Lake, Lake house requirements and implementing with Snowflake. This also covers a variety of real time use cases, data architectures for reference.

Chapter 17: Introduction to Snowflake's Advance Features – Snowflake continuously works on their features, releases updates and introduces new features with every release. This chapter helps to understand new features introduced by Snowflake. This also covers a variety of real time use cases for reference.

Code Bundle and Coloured Images

Please follow the link to download the
Code Bundle and the ***Coloured Images*** of the book:

https://rebrand.ly/xar4q49

The code bundle for the book is also hosted on GitHub at
https://github.com/bpbpublications/Mastering-Snowflake-Platform.
In case there's an update to the code, it will be updated on the existing GitHub repository.

We have code bundles from our rich catalogue of books and videos available at **https://github.com/bpbpublications**. Check them out!

Errata

We take immense pride in our work at BPB Publications and follow best practices to ensure the accuracy of our content to provide with an indulging reading experience to our subscribers. Our readers are our mirrors, and we use their inputs to reflect and improve upon human errors, if any, that may have occurred during the publishing processes involved. To let us maintain the quality and help us reach out to any readers who might be having difficulties due to any unforeseen errors, please write to us at :

errata@bpbonline.com

Your support, suggestions and feedbacks are highly appreciated by the BPB Publications' Family.

Piracy

If you come across any illegal copies of our works in any form on the internet, we would be grateful if you would provide us with the location address or website name. Please contact us at **business@bpbonline.com** with a link to the material.

If you are interested in becoming an author

If there is a topic that you have expertise in, and you are interested in either writing or contributing to a book, please visit **www.bpbonline.com**. We have worked with thousands of developers and tech professionals, just like you, to help them share their insights with the global tech community. You can make a general application, apply for a specific hot topic that we are recruiting an author for, or submit your own idea.

Reviews

Please leave a review. Once you have read and used this book, why not leave a review on the site that you purchased it from? Potential readers can then see and use your unbiased opinion to make purchase decisions. We at BPB can understand what you think about our products, and our authors can see your feedback on their book. Thank you!

For more information about BPB, please visit **www.bpbonline.com**.

Join our book's Discord space

Join the book's Discord Workspace for Latest updates, Offers, Tech happenings around the world, New Release and Sessions with the Authors:

https://discord.bpbonline.com

Table of Contents

Chapter 1
Getting Started with Snowflake

Introduction

The first chapter is the foundation of the book and covers the history of the Snowflake, the need to implement Snowflake, and getting started with Snowflake. This chapter guides to setting up a trial or demo account to be used for various hands-on activities covered in subsequent chapters. This also covers the overview of Snowflake certifications and various community events.

You will learn more about Snowflake data platform capabilities in upcoming chapters of this book. Every chapter consists of real-time use cases and references to explain each concept. We have also provided lab questions with every chapter to help you learn with examples. There are a set of questions to check your knowledge at the end of each chapter.

This book is uniquely designed and well-written to help you understand Snowflake's features easily. You will learn Snowflake seamlessly as you read through this book. With every concept and practical lab in this, you will surely want to plan for certification.

Structure

This chapter consists of the following topics:

- Why Snowflake?
- History of Snowflake

- Snowflake certifications
- Snowflake community
- Setting up a trial account with Snowflake
- Connecting to Snowflake

Objectives

By the end of this chapter, you will be able to understand the need for a Snowflake data platform over traditional or enterprise platforms. You will also be able to set up a trial account for yourself. This trial account will be used to perform various exercises throughout this book.

Why Snowflake?

In the earlier era setting up a data platform needed tedious efforts to design, derive the capacity of the system, and purchase hardware-software or appliances to set up an ecosystem to support the data needs. These ecosystems had their own limitations in terms of scaling – horizontal vs vertical, cost-efficiency, performance efficiency, operational cost, and huge maintenance cost associated-support teams, upgrades, patches, EOLs activities, and so on.

The cloud has broken down almost all these limitations and taken over the majority of legacy ecosystems to the cloud. The cloud offers scalability, efficiency, low operations, and no or low-cost maintenance. Cloud-native and SaaS services replaced the legacy ecosystems however, had their own limitations or dependencies or locking with vendors. For example, if you are using Google Cloud Bigquery – you are tied with using it only for Google Cloud; though this can be integrated with any other platforms available. There are also limitations associated with most cloud-native or managed services in terms of the type of workload they support. Data Lake, Data Warehouse, data analytics, and data science are often treated as separate workloads and designed differently integrating with each other. There are various data platforms available in the market to support data workloads as per business requirements.

Snowflake eliminated the need to design and define data workloads separately. With Snowflake, you can use the same data platform to cater all types of workload needs - Data Lake, Data Warehouse, data analytics, and data science needs. This also caters to a workload where you can combine analytical and transactional workloads: **Unistore**.

You need to learn about Snowflake and its unique offerings, as they are one of the leading data platforms in the market. You can learn Snowflake and get started with your data career journey or change your career path as well. It will be beneficial to learn Snowflake, considering the ample opportunities available in the market and Snowflake's adaptability.

History of Snowflake

Snowflake is a one-stop solution for a variety of workloads. Snowflake's all-in-one platform enables organizations to quickly set up a centralized data platform. This platform can be used to generate values from the data stored within, extending implementations to various applications with data protection, security, and compliance.

This is the journey that started in 2012 when the founders met for the first time with a vision of building a data warehouse for the cloud from scratch to unlock the potential of unlimited analytics from heterogeneous data. They had aimed to build a solution that is not only secure and impactful but also cost-effective and easier to manage.

Within three years, Snowflake's data warehouse – built from scratch on the cloud – was available in 2015. Snowflake's unique, cloud-agnostic architecture disrupted the data warehousing market. With Snowflake, data engineering also changed from technical to business-oriented implementations. This has made data analytics simpler, that helps users generate data stats which in turn helped organizations make data-driven decisions.

In another three years, in 2018, Snowflake introduced – data sharing. This is the most critical feature that is used to share data with internal or external stakeholders with appropriate access controls and security. Interestingly, you can also share data with Snowflake as well as non-snowflake users. You will learn more about this in *Chapter 12: Data Sharing*.

In 2021, Snowflake announced its expansion to support the wider category of data engineering with Snowpark. This is a new development framework which is a unique combination that makes it simpler to design and develop data engineering workloads on Snowflake. This can also be used to extend support to data science capabilities. You will learn more about this in *Chapter 6: Understanding Snowpark*.

Also, in 2021, Snowflake announced Snowflake organizations. With this, it is very easy to manage multiple accounts for the same customer. You can tag various accounts that are active and required under an organization. You can also set up utilization and usage at the account as well as organization level for tracking.

In 2022, Snowflake added the Security data lake. This is a type of workload that enables the full visibility of security logs. Snowflake's newest workload: **Unistore,** is a very unique platform where you can combine the power of transactional and analytical operations.

Snowflake certifications

Snowflake offers basic and advanced-level certifications. Earning a certification badge definitely adds more weight to your profile.

Snowflake SnowPro Core is the first level of the foundation certification exam. Snowflake offers five advanced certifications based on your role. You can refer to *Figure 1.1* for basic and advanced certifications available:

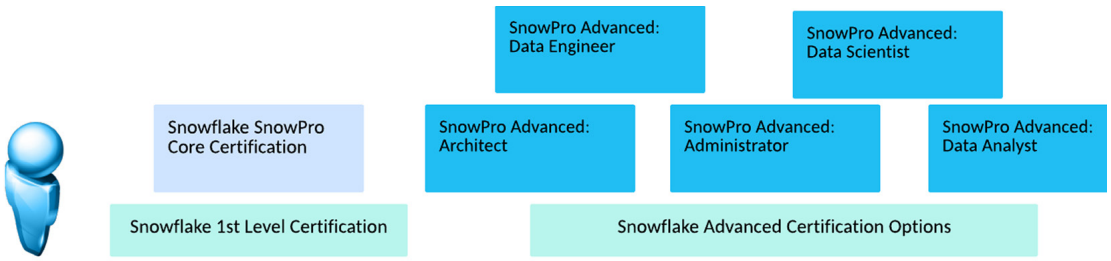

Figure 1.1: *Snowflake certifications*

You will need to complete the foundation - SnowPro Core Certification before you appear for the advanced certifications. Once you pass the certification, it is valid for two years, and you can appear for a re-certification exam to renew your certification for another two years. An advanced certification will automatically renew your SnowPro Core Certification for the next two years. Advanced certifications are available for various roles – Data Engineer, Architect, Administrator, Analyst, and Data Scientist. You can appear for the advanced one based on your role and experience. You can find more details about certifications here: **https://www.snowflake.com/certifications/**.

Snowflake community

Snowflake runs various community initiatives. This is one of the most active communities where many users can contribute and connect with other community members.

There are Snowflake user groups that you can join if you are interested in connecting and joining. User groups held events in person as well as virtually across various regions. You can find more details here: **https://usergroups.snowflake.com/**.

Snowflake runs the Data Superheroes community program every year. This program is for Snowflake experts who are highly active in the community. The active contributors are recognized as Data Superheroes, and Snowflake announces members of this elite group at the beginning of the year. You can learn more about this program here: **https://medium.com/snowflake/all-you-need-to-know-about-snowflake-data-superheroes-a36914e2e614**. Refer to the following figure:

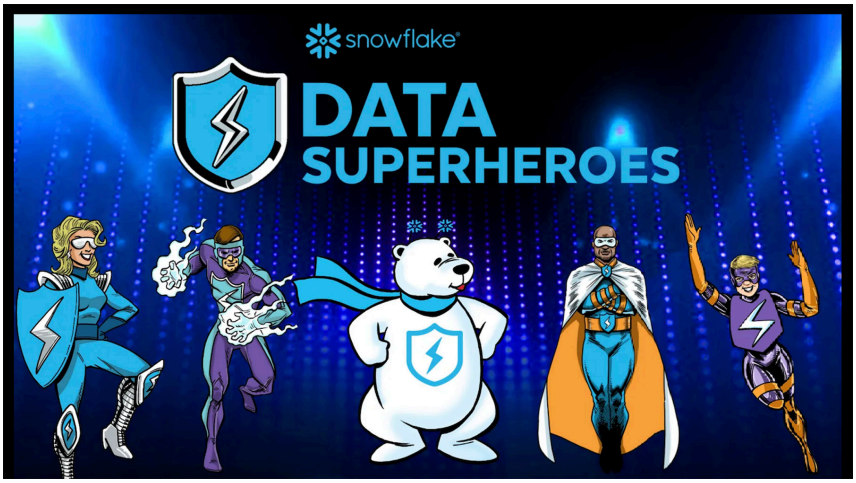

***Figure 1.2**: Snowflake data superheroes*

You can start contributing to the community as an individual user or through your partner account. There are various trainings available on the community portal as well as the partner network portal. This book is one guide to learning and getting started with Snowflake. Once you understand the concepts and complete practical labs from the book then you can refer to the quick starts to navigate through data engineering, data lakes, and other types of workload-specific use cases.

You can also use Snowflake's documentation to get started with Snowflake. There are also quick labs available that enable users to perform hands-on labs based on the workloads and learn Snowflake. You can find use case-specific hands-on labs here: **https://quickstarts.snowflake.com/**.

Setting up a trial account with Snowflake

Snowflake offers a 30-day trial account that is worth $400 to practice, perform hands-on labs, and learn easily. Snowflake offers three versions on all 3 public cloud platforms that users can choose while setting up a trial version.

You can follow these steps to set up a trial account:

1. Open the link: **https://signup.snowflake.com/**.
2. Fill in these details in the signup form specified as shown in *Figure 1.3*:
    a. **First name**: Your first name
    b. **Last name**: Your last name