

Mastering Snowflake Platform

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

Pooja Kelgaonkar



www.bpbonline.com

Copyright © 2024 BPB Online

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, without the prior written permission of the publisher, except in the case of brief quotations embedded in critical articles or reviews.

Every effort has been made in the preparation of this book to ensure the accuracy of the information presented. However, the information contained in this book is sold without warranty, either express or implied. Neither the author, nor BPB Online or its dealers and distributors, will be held liable for any damages caused or alleged to have been caused directly or indirectly by this book.

BPB Online has endeavored to provide trademark information about all of the companies and products mentioned in this book by the appropriate use of capitals. However, BPB Online cannot guarantee the accuracy of this information.

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.

Did you know that BPB offers eBook versions of every book published, with PDF and ePub files available? You can upgrade to the eBook version at www.bpbonline.com and as a print book customer, you are entitled to a discount on the eBook copy. Get in touch with us at :

business@bpbonline.com for more details.

At **www.bpbonline.com**, you can also read a collection of free technical articles, sign up for a range of free newsletters, and receive exclusive discounts and offers on BPB books and eBooks.

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

1. Getting Started with Snowflake	1
Introduction.....	1
Structure.....	1
Objectives.....	2
Why Snowflake?	2
History of Snowflake	3
Snowflake certifications.....	3
Snowflake community	4
Setting up a trial account with Snowflake	5
Connecting to Snowflake.....	7
<i>Snowflake Web User Interface</i>	<i>8</i>
<i>Classic console</i>	<i>8</i>
<i>Context setting</i>	<i>9</i>
<i>Snowsight</i>	<i>10</i>
<i>Command line interface</i>	<i>11</i>
<i>Download and setup SnowSQL.....</i>	<i>11</i>
<i>Connecting to Snowflake.....</i>	<i>12</i>
Conclusion.....	14
Points to remember	14
Practical.....	15
Multiple choice questions.....	15
Answers	16
2. Three Layered Architecture	17
Introduction.....	17
Structure.....	17
Objectives.....	18
Traditional data warehouse challenges.....	18
Legacy data warehouse challenges.....	19
Typical data lake challenges.....	20
<i>Hadoop ecosystem</i>	<i>21</i>
<i>Extended support to NoSQL</i>	<i>22</i>

Snowflake architecture.....	23
Cloud services layer.....	25
<i>Authentication and access control</i>	26
<i>Optimizer</i>	26
<i>Metadata storage and management</i>	26
<i>Result cache</i>	26
<i>Security</i>	26
<i>Availability of cloud services layer</i>	27
<i>Uses of cloud services layer</i>	27
Virtual warehouse layer.....	28
<i>Virtual warehouses: Single and multi-cluster</i>	29
<i>Single cluster warehouses</i>	30
<i>Multi-cluster warehouses</i>	30
<i>Virtual warehouse properties</i>	30
<i>Virtual warehouse sizes</i>	31
<i>Warehouse scaling</i>	31
<i>Scaling up</i>	31
<i>Scaling out</i>	32
<i>Scaling modes</i>	32
<i>Scaling policies</i>	33
<i>Virtual warehouses: Creation and modifications</i>	33
<i>Warehouse setup using console</i>	33
<i>Warehouse: Using SQL commands</i>	37
<i>Creating warehouse</i>	37
<i>Other warehouse commands</i>	38
<i>Virtual warehouses: Billing and usage</i>	38
Storage layer.....	39
<i>Time-travel</i>	40
<i>Failsafe</i>	40
<i>Data cloning</i>	40
<i>Storage usage and billing</i>	40
Cache in Snowflake.....	41
<i>Metadata cache</i>	41
<i>Query result cache</i>	42
<i>Warehouse cache</i>	43

Conclusion.....	43
Points to remember	44
Multiple choice questions.....	44
Answers	45
Questions.....	45
3. Data Types, Data Objects and SQL Commands	47
Introduction.....	47
Structure.....	47
Objectives.....	48
Understanding data types.....	48
<i>Numeric data types</i>	48
<i>String and binary data types</i>	49
<i>Binary data types</i>	50
<i>Logical data types</i>	50
<i>Date and time data types</i>	51
<i>Date and time intervals and constants</i>	54
<i>Supported date and time arithmetic operations</i>	55
<i>Semi-structured data types</i>	56
VARIANT.....	56
OBJECT.....	57
OBJECT constant.....	57
OBJECT elements.....	57
OBJECT considerations	58
ARRAY	58
ARRAY elements	59
ARRAY considerations	59
<i>Geospatial data types</i>	59
<i>Geography</i>	59
<i>Geometry</i>	59
Using SQL commands	60
Data definition language	60
Account and session DDL.....	60
User and security DDL.....	62
Warehouse and resource monitor DDL.....	63
Database, schema and share DDL.....	64

Table, view and sequence DDL.....	65
Data loading / unloading DDL.....	66
User-defined functions, external functions, and stored procedures DDL.....	67
Data pipeline DDL.....	68
Data manipulation language.....	69
Standard DML.....	69
Data loading/unloading DML.....	69
File staging commands.....	70
Creating database objects.....	70
Creating users and roles.....	71
Creating roles using SQL commands.....	71
Creating users using SQL commands.....	72
Creating roles using Web Console.....	72
Creating users using Web Console.....	73
Creating warehouses.....	73
Creating warehouse using SQL command.....	74
Creating warehouse using Web Console.....	74
Creating databases and schemas.....	75
Creating database using SQL command.....	75
Creating database using web console.....	75
Creating schemas using SQL command.....	75
Setting up context.....	75
Setup context using SQL command.....	76
Creating tables and views.....	76
Creating table using SQL command.....	76
Implementing extended SQL objects.....	77
UDF example (JavaScript).....	78
UDF example (SQL).....	78
Tabular functions.....	79
Sample function.....	79
Creating function.....	79
Using function.....	79
SQL UDFT function.....	80
Creating function.....	80
Using function.....	81

<i>External functions</i>	81
Conclusion.....	82
Points to remember	82
Practical: Create Snowflake objects.....	82
Multiple choice questions.....	83
Answers	83
Questions	83
4. Data Loading and Unloading	85
Introduction.....	85
Structure.....	85
Objectives.....	86
Understanding data load needs	86
Creating Snowflake load objects	87
<i>Source files</i>	87
<i>Loading from file locations</i>	87
<i>Loading files from local system</i>	94
<i>Loading data from Web UI</i>	94
Using COPY INTO to load data	97
<i>Bulk load using COPY INTO</i>	97
COPY syntax	97
<i>Transformations supported in COPY INTO</i>	98
COPY with transformation syntax.....	98
VALIDATE command.....	99
COPY optional parameters.....	99
COPY INTO options.....	100
Options.....	101
COPY INTO using for bulk loads.....	102
COPY for local system.....	103
PUT command	103
Understanding streaming loads.....	104
Implementing Snowpipe	104
Automate the load using cloud messaging	104
Using Snowpipe REST endpoints	105
Snowpipe versus bulk load.....	105
Snowpipe features and recommendations.....	106

Snowpipe commands.....	106
<i>SQL commands</i>	106
<i>CREATE PIPE</i>	106
<i>ALTER PIPE</i>	107
<i>DROP PIPE</i>	108
<i>DESCRIBE PIPE</i>	108
<i>SHOW PIPE</i>	108
Loading real time data.....	109
<i>Snowpipe</i>	109
<i>Snowflake connector for Kafka</i>	109
<i>Snowpipe streaming</i>	110
<i>Snowpipe streaming versus Snowpipe</i>	111
<i>Streaming cost</i>	111
Loading near real time data.....	111
Implementing Snowpipe.....	112
Understanding data unloading.....	113
Database feed.....	114
File feed.....	114
Using COPY to unload data.....	115
<i>COPY INTO</i>	115
<i>Exported feeds, formats and locations</i>	115
<i>Compression supported</i>	116
<i>Encryption supported</i>	116
Bulk unloading from a table or query.....	116
Bulk unloading to one or multiple files.....	117
<i>File naming in case of multiple file exports</i>	117
<i>Bulk unloading using PARTITIONS</i>	117
<i>COPY INTO options</i>	118
<i>Exporting data to named internal stages</i>	118
<i>GET command</i>	120
<i>Exporting data to named external stages</i>	120
<i>Steps to unload data to named external stages</i>	120
Practical: Data loading and unloading.....	121
<i>Creating stages</i>	121
<i>Creating file formats</i>	121

<i>Creating storage integrations</i>	122
<i>Creating LOAD commands and load using COPY</i>	122
<i>Load from local using COPY</i>	122
<i>Unload from local using COPY</i>	123
Conclusion.....	123
Points to remember	123
Questions	124
Multiple choice questions.....	124
Answers	125
5. Understanding Streams and Tasks	127
Introduction.....	127
Structure.....	127
Objectives.....	128
Understanding Change Data Capture.....	128
Understanding Snowflake Streams	130
<i>Types of Streams</i>	131
<i>Standard Streams</i>	131
<i>Append-only Streams</i>	131
<i>Insert-only Streams</i>	131
<i>Validity of change logs</i>	132
<i>Billing of Streams</i>	132
Implementing data capture with Streams.....	132
CREATE STREAM.....	132
ALTER STREAM.....	133
DESCRIBE STREAM.....	133
DROP STREAM.....	133
SHOW STREAMS.....	134
<i>Stream examples</i>	134
Understanding tasks	135
<i>Serverless tasks</i>	136
<i>User-managed tasks</i>	136
<i>Scheduling tasks</i>	137
Implementing scheduling with Tasks	137
TASK commands	137
CREATE TASK.....	138

ALTER TASK.....	140
DROP TASK.....	141
EXECUTE TASK.....	141
SHOW TASKS.....	141
Practical: Create Snowflake Tasks and Streams	142
Exercise 1	142
Exercise 2	142
Exercise 3	142
Conclusion.....	143
Points to remember	143
Multiple choice questions.....	143
Answers	143
Questions	144
6. Understanding Snowpark.....	145
Introduction.....	145
Structure.....	146
Objectives.....	146
Why Snowpark?	146
Understanding Snowpark.....	148
Snowpark-optimized warehouse	149
Snowpark-optimized warehouse billing.....	149
Snowpark-ML.....	150
Snowpark-benefits over Spark Connector.....	150
Implementing Snowpark.....	150
Environment setup	151
Using Python and developing code	151
How does Python dataframe API work?	152
Using Snowpark for ML.....	153
Use cases to implement Snowpark	153
Data engineering use cases.....	154
Practical: Setting up Snowpark and use cases.....	155
Snowpark engineering use cases.....	155
Conclusion.....	156
Points to remember	156
Questions	157

Multiple choice questions.....	157
Answers	157
7. Access Control and Managing Users Roles	159
Introduction.....	159
Structure.....	160
Objectives.....	160
Snowflake access control overview	160
<i>Snowflake role and users.....</i>	<i>161</i>
<i>Types of roles</i>	<i>162</i>
<i>Snowflake securable objects</i>	<i>163</i>
Understanding RBAC	163
Understanding default roles in Snowflake	166
Implementing RBAC with Snowflake	167
<i>Step 1: Creating Snowflake resources</i>	<i>168</i>
<i>Create environments</i>	<i>169</i>
<i>Create DEV databases and schemas.....</i>	<i>169</i>
<i>Create QA databases and schemas</i>	<i>169</i>
<i>Create BI databases and schemas</i>	<i>169</i>
<i>Create ML databases and schemas</i>	<i>170</i>
<i>Create warehouses for users and use cases.....</i>	<i>170</i>
<i>Step 2: Create custom roles.....</i>	<i>170</i>
<i>Create DEV custom role.....</i>	<i>170</i>
<i>Step 3: Granting access to the custom roles</i>	<i>171</i>
<i>Grant access of DATABASES to custom role.....</i>	<i>171</i>
<i>Grant access to WAREHOUSES to the custom role</i>	<i>171</i>
<i>Grant access to QA custom role</i>	<i>171</i>
<i>Grant access to DEV custom role in QA environment.....</i>	<i>171</i>
<i>Step 4: Create users (user onboarding).....</i>	<i>172</i>
<i>Create users</i>	<i>172</i>
<i>Step 5: Assign custom roles to the users</i>	<i>172</i>
<i>Assign custom roles to the users</i>	<i>172</i>
<i>Step 6: Using the custom roles</i>	<i>173</i>
<i>Using custom roles created.....</i>	<i>173</i>
<i>Testing and validating access of the custom roles created.....</i>	<i>173</i>
<i>Step 7: Validating the access privileges</i>	<i>173</i>

<i>Using custom roles created</i>	174
Understanding access hierarchy	174
Managing users and roles	175
<i>User commands</i>	176
<i>CREATE command</i>	176
<i>DROP command</i>	176
<i>ALTER command</i>	176
<i>DESCRIBE command</i>	176
<i>SHOW command</i>	176
Practical.....	177
Conclusion.....	177
Points to remember	178
Questions	178
Multiple choice questions.....	178
Answers	179
8. Data Protection and Recovery	181
Introduction.....	181
Structure.....	182
Objectives.....	182
What is data protection?.....	182
Implementing dynamic masking	183
<i>Implementation commands</i>	183
<i>Create masking policy</i>	183
<i>Normal masking policy</i>	184
<i>Conditional masking policy</i>	185
<i>Alter masking policy</i>	186
<i>Drop masking policy</i>	187
<i>Show masking policy</i>	187
<i>Describe masking policy</i>	187
Using dynamic masking	187
<i>Step 1: Create custom role</i>	188
<i>Step 2: Grant custom role to a user</i>	188
<i>Step 3: Create masking policy</i>	188
<i>Step 4: Apply masking policy to the database objects</i>	189
<i>Step 5: Query data and validate masking policies</i>	189

Understanding data recovery options.....	190
<i>Understanding time travel</i>	190
<i>Time travel SQL extensions</i>	191
<i>Time travel data retention</i>	191
<i>Time travel setup</i>	192
<i>Time travel storage cost</i>	192
<i>Time travel storage for temporary and transient tables</i>	192
<i>Understanding Failsafe</i>	193
<i>Failsafe storage cost</i>	194
Implementing time travel.....	194
<i>Setup time travel at create table</i>	194
<i>Access historical data</i>	195
<i>Cloning objects</i>	195
<i>Commands to access dropped and restored objects</i>	196
<i>List dropped objects</i>	196
<i>Restore dropped objects</i>	196
<i>Reference use case</i>	197
Implementing data replication	198
<i>Database replication</i>	199
<i>Share replication</i>	199
<i>Replication group</i>	200
<i>Replication schedule</i>	200
<i>Business continuity</i>	201
Practical: Create Snowflake policies, tagging objects.....	202
Exercises.....	203
Conclusion.....	204
Points to remember	204
Questions	204
Multiple choice questions.....	205
Answers	206
9. Snowflake Performance Optimization	207
Introduction.....	207
Structure.....	207
Objectives.....	208
Understanding Snowflake performance	208

Query performance	208
Understanding Snowflake metadata objects	210
INFORMATION_SCHEMA.....	210
Using INFORMATION_SCHEMA.....	211
ACCOUNT_USAGE.....	211
DATA_SHARING_USAGE.....	212
ORGANIZATION_USAGE.....	212
Introduction to ACCOUNT_USAGE.....	212
Dropped object information	213
Latency of the details	213
Data retention.....	213
Account usage views.....	213
Account usage functions	214
Reader account usage views.....	214
Roles and permissions.....	215
Introduction to INFORMATION_SCHEMA	216
Information schema views	216
Information schema functions	217
Calculating and understanding performance measures	219
Understanding QUERY_HISTORY.....	221
Introduction to Query Acceleration Service	224
SYSTEM\$ESTIMATE_QUERY_ACCELERATION function.....	225
QUERY_ACCELERATION_ELIGIBLE view	225
Enable query acceleration service	226
Query acceleration usage.....	226
Query acceleration billing	227
Implementing performance optimization	227
Query performance optimization.....	228
View historical performance in Snowsight.....	228
View query profile Snowsight	230
Listing queries for performance optimization	231
Warehouse performance optimization.....	233
Reducing queue time.....	233
Resolving memory spillage.....	234
Changing warehouse size	235

<i>Optimizing cache</i>	235
<i>Limiting concurrent queries</i>	236
<i>Using query acceleration service</i>	236
Storage performance optimization	237
<i>Automatic clustering</i>	237
<i>Materialized views</i>	237
<i>Search optimization service</i>	238
<i>Choosing the right strategy</i>	238
<i>Performance considerations</i>	238
Conclusion.....	239
Points to remember	240
Questions	240
Multiple choice questions.....	240
Answers	241
10. Understanding Snowflake Costing and Utilizations.....	243
Introduction.....	243
Structure.....	243
Objectives.....	244
Understanding Snowflake costing.....	244
<i>Visibility</i>	245
<i>Control</i>	245
<i>Optimization</i>	246
Component costing in Snowflake	246
<i>Cloud services cost</i>	247
<i>Compute cost</i>	247
<i>Storage cost</i>	248
<i>Serverless cost</i>	249
<i>Data transfer cost</i>	249
Using Snowflake metadata objects	250
<i>View compute usage</i>	250
<i>View Cloud services usage</i>	251
<i>View serverless usage</i>	252
Monitoring Snowflake costs.....	253
Conclusion.....	255
Points to remember	255

Questions	256
Multiple choice questions.....	256
Answers	256
11. Implementing Cost Optimizations	257
Introduction.....	257
Structure.....	257
Objectives.....	258
Components costing in Snowflake.....	258
<i>Compute or virtual warehouse layer.....</i>	<i>258</i>
<i>Cloud services layer</i>	<i>259</i>
<i>Storage layer</i>	<i>259</i>
<i>Serverless services.....</i>	<i>259</i>
<i>Sample use case.....</i>	<i>259</i>
Implementing cost optimization	261
<i>Compute optimization</i>	<i>261</i>
<i>Access setup to warehouse.....</i>	<i>261</i>
<i>Setup the right size of the warehouse</i>	<i>261</i>
<i>Setup query time limit.....</i>	<i>261</i>
<i>Setup query queue time limit</i>	<i>262</i>
<i>Setup auto suspend and auto resume.....</i>	<i>262</i>
<i>Enforce spend limits</i>	<i>262</i>
<i>Storage optimization</i>	<i>262</i>
<i>Store only required data</i>	<i>262</i>
<i>Cloud services optimization.....</i>	<i>262</i>
Implementing cost dashboards	263
Conclusion.....	266
Points to remember	267
Questions	267
12. Data Sharing	269
Introduction.....	269
Structure.....	269
Objectives.....	270
Data sharing needs.....	270
Implementing data sharing.....	272

<i>How does data sharing work?</i>	273
Implementing sharing with Snowflake consumers.....	274
<i>Data sharing considerations</i>	274
<i>Data sharing commands</i>	274
<i>Data sharing pre-requisites</i>	275
<i>Creating a data share</i>	276
<i>Creating a data share with multiple databases</i>	277
<i>Using a data share</i>	279
Implementing sharing with non-Snowflake consumers.....	280
<i>Creating reader account</i>	281
Understanding data sharing options.....	283
<i>Listing</i>	283
<i>Marketplace listing</i>	284
<i>Private listing</i>	284
<i>Personalized listing</i>	284
<i>Data exchange</i>	284
<i>Providers tasks</i>	285
<i>Consumer tasks</i>	286
Practical: Create Snowflake shares.....	286
<i>Scenario</i>	286
<i>Technical requirement</i>	287
<i>Account setup</i>	287
<i>Data sharing</i>	287
Conclusion.....	288
Points to remember.....	288
Questions.....	289
13. Data Cloning	291
Introduction.....	291
Structure.....	292
Objectives.....	292
Data cloning needs.....	292
Zero copy cloning.....	293
Implementing data cloning.....	294
<i>Cloning considerations</i>	295
Cloning implementation.....	296

<i>Syntax of CREATE</i>	296
<i>Reference use cases and examples</i>	297
Understanding the usage of cloned objects.....	298
Practical: Create Snowflake clones.....	298
Conclusion.....	299
Points to remember	299
Questions	299
14. Understanding Snowsight	301
Introduction.....	301
Structure.....	301
Objectives.....	301
Understanding Snowsight	302
<i>Snowsight interface</i>	302
<i>Worksheet pages</i>	303
<i>Data pages</i>	305
<i>Dashboard page</i>	305
<i>Data page</i>	306
<i>Marketplace page</i>	306
<i>Activity page</i>	306
<i>Admin page</i>	307
Implementing dashboards with Snowsight	307
Practical: Create Snowflake dashboards	313
<i>Use case</i>	313
<i>Initial set up</i>	314
<i>Key asks</i>	314
Conclusion.....	314
Points to remember	315
Questions	315
15. Programming Connectors and Drivers	317
Introduction.....	317
Structure.....	317
Objectives.....	317
Understanding programming connectors	318
<i>Snowflake connector for Python</i>	318

Snowflake connector for Kafka.....	318
Snowflake connector for Spark.....	318
Understanding drivers	319
JDBC driver and ODBC driver	319
GO driver.....	319
Understanding Snowsql: CLI	320
Installing Snowsql.....	320
Connecting to Snowsql.....	320
Using password to connect via Snowsql	321
Conclusion.....	322
Annexure	322
Points to remember	323
Questions.....	323
16. Workload Patterns with Snowflake.....	325
Introduction.....	325
Structure.....	325
Objectives.....	326
Understanding platform needs	326
Extract, Transform, and Load.....	326
Extract, Load and Transform.....	327
Implementing data solutions with Snowflake	329
Snowflake's features	329
Data solutions.....	330
Reference use cases and architecture.....	331
Data warehouse reference use case	331
Use case	331
Technical requirements.....	331
Business requirements.....	331
Proposed solution design.....	332
Reference architecture.....	332
Pipeline design	333
Platform design	333
Data lake reference use case	333
Use case	333
Technical requirements.....	334

<i>Business requirements</i>	334
<i>Proposed solution design</i>	334
<i>Reference architecture</i>	335
<i>Pipeline design</i>	335
<i>Platform design</i>	336
<i>Data mesh reference use case</i>	336
<i>Use case</i>	336
<i>Technical requirements</i>	336
<i>Business requirements</i>	337
<i>Proposed solution design</i>	337
<i>Reference architecture</i>	337
<i>Platform design</i>	338
Snowflake recommendations.....	339
<i>Recommendations</i>	339
<i>Warehouse recommendations</i>	339
<i>Storage recommendations</i>	339
<i>Load recommendations</i>	339
<i>Transformation recommendations</i>	340
<i>Usage recommendations</i>	340
Conclusion.....	340
Points to remember	341
Questions	341
17. Introduction to Snowflake’s Advance Features	343
Introduction.....	343
Structure.....	343
Objectives.....	343
Understanding new features and releases.....	344
<i>Snowflake releases</i>	344
<i>New features</i>	344
<i>SQL extensions</i>	344
<i>Schema features</i>	345
<i>SQL features</i>	345
Understanding new types of tables	346
<i>Dynamic tables</i>	347
<i>Event tables</i>	348

<i>Hybrid tables</i>	350
<i>Iceberg tables</i>	351
Introduction to new Snowsight features.....	352
<i>Worksheets</i>	353
<i>Data governance</i>	353
<i>Native apps</i>	354
<i>Budgets</i>	354
<i>Data loading</i>	355
<i>Stage options</i>	355
Introduction to new LLM capabilities.....	356
Conclusion.....	357
Points to remember	357
Index	359-369

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:

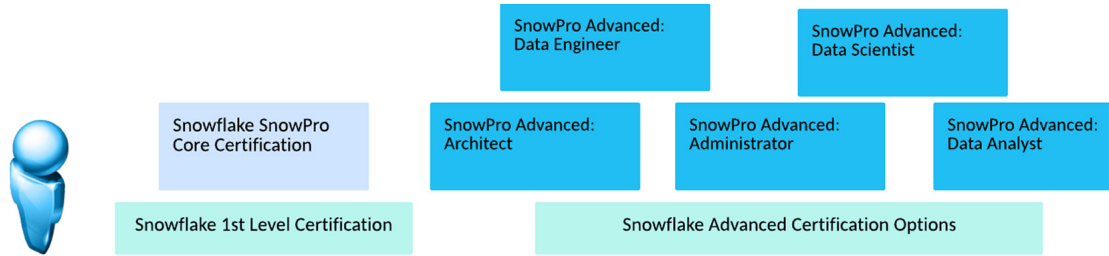


Figure 1.1: Snowflake certifications

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:

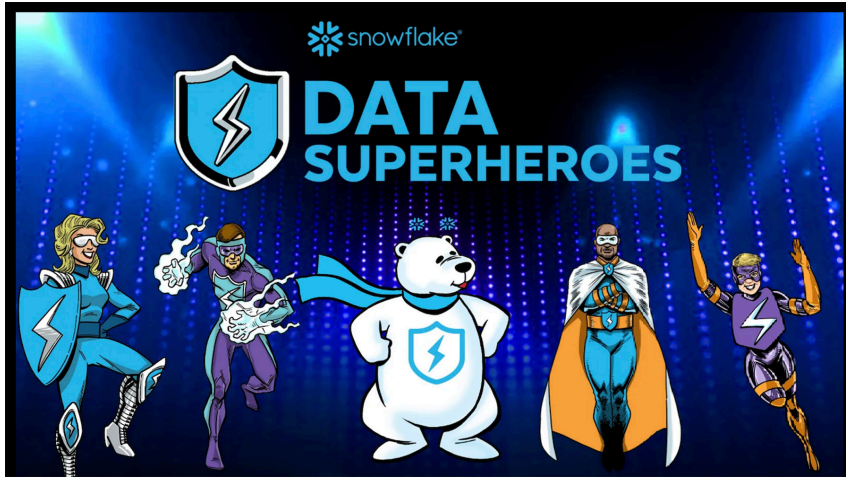


Figure 1.2: Snowflake data superheroes

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