

Mastering Power BI

*Build business intelligence applications powered
with DAX calculations, insightful
visualizations, advanced BI
techniques, and loads of data sources*

2nd Edition

Chandraish Sinha



www.bpbonline.com

Second Revised and Updated Edition 2024

First Edition 2022

Copyright © BPB Publications, India

ISBN: 978-93-55517-166

All Rights Reserved. No part of this publication may be reproduced, distributed or transmitted in any form or by any means or stored in a database or retrieval system, without the prior written permission of the publisher with the exception to the program listings which may be entered, stored and executed in a computer system, but they can not be reproduced by the means of publication, photocopy, recording, or by any electronic and mechanical means.

LIMITS OF LIABILITY AND DISCLAIMER OF WARRANTY

The information contained in this book is true to correct and the best of author's and publisher's knowledge. The author has made every effort to ensure the accuracy of these publications, but publisher cannot be held responsible for any loss or damage arising from any information in this book.

All trademarks referred to in the book are acknowledged as properties of their respective owners but BPB Publications cannot guarantee the accuracy of this information.

To View Complete
BPB Publications Catalogue
Scan the QR Code:



www.bpbonline.com

Kup ksi k

Dedicated to

Ishie

My daughter, who taught me that patience is a virtue

About the Author

Chandraish Sinha is the Founder and President of Ohio Computer Academy, a company dedicated to providing IT education.

An enthusiastic IT trainer, Chandraish embodies his company's motto: Inspire, Educate and Evolve.

He is a **Business Intelligence (BI)** enthusiast and explorer, having successfully implemented multiple large and medium-scale BI solutions. Over his 24-year career, Chandraish has worked extensively with various BI and dashboarding applications, including Power BI, Tableau, QlikView, Qlik Sense, IBM Cognos, Business Objects, and Actuate. His passion lies in data, and he actively explores applications that offer enhanced data insights.

About the Reviewers

- ❖ **Preeti Sahu** is an expert in Power BI and has an extensive background working with SharePoint and the Power Platform. She is currently employed by TSinfo Technologies as a Power Platform specialist for Power BI, Power Apps, Power Automate, and Power Virtual Agents. She is the author of the Microsoft Power Platform A Deep Dive book and the co-author of the SharePoint Online Modern Experience Practical Guide book (BPP Publications). Also, she has contributed a significant amount of time on technical articles on the Power Platform to SPGuides.com and EnjoySharePoint.com. In addition to traveling, she enjoys spending time with her family during her free time.
- ❖ **Sriram Chaitanya Puvvada** is a seasoned Business Intelligence and Data Engineering Consultant with a rich professional history spanning over 9 years. His expertise extends across various industry verticals, including SAP Retail, Pharma, Manufacturing, and Transportation, with a remarkable track record of delivering projects for prestigious clients globally.

Having served as a Solutions Architect, Project Manager, Scrum Master, and Product Owner, Sriram is adept at leading diverse teams, both onshore and offshore, to successful project outcomes. His proficiency encompasses a wide array of tools and technologies such as Data Bricks, ADF, SAP BI, Power BI, Tableau, and more, enabling him to implement and maintain BI analytical solutions effectively.

Sriram's dedication to excellence has been recognized through accolades like the Star Employee award and commendations from prominent clients like Arriva Plc and SEG Automotive. He is also known for his contribution as a Technical Content Writer for Google Cloud Publications and his pivotal role in obtaining ISO certifications for organizations. Sriram's commitment to delivering high-quality solutions within schedule and budget constraints underscores his reputation as a valuable asset in the field of Business Intelligence and Data Engineering.

- ❖ **Vivek Menon M** is an AI and Data Science professional with six years of experience in the industry. In his independent career, he has worked in advanced areas of NLP, Computer Vision, Generative AI, Data Analytics, Python, Machine Learning, etc., to gain comprehensive knowledge and skills in the field. As an eloquent speaker, Vivek frequently presents seminars, and workshops to share his insights and experience in the field, thereby adding value to the broader AI community. He enjoys the natural

ability to break down even the most complex topics to make them understandable and accommodative among listeners, strengths that make him a compelling and authoritative speaker. Beyond a successful job, Vivek is a reader, a characteristic that allows him to learn a wide range of subjects, from computational theories to classic books. His diverse learning style allows him to mold an integrative approach to AI implementations that foster success. He is also involved in mentoring activities for young data scientists, reflecting that he is committed to educating and mentoring young professionals to mold their future as suitable data science professionals.

Acknowledgement

I extend my sincere gratitude to all my readers who supported the first edition of Mastering Power BI. Your feedback and responses guided me in crafting the second edition of this book.

I am deeply appreciative of all my readers who have supported my various books and motivated me to continue writing. I carefully consider each review and feedback provided, and your encouragement fuels my passion for sharing knowledge.

A special thank you to Preeti, Vivek, and Sriram, the technical reviewers of this book, for meticulously reviewing the entire manuscript and offering invaluable suggestions.

Lastly, I express my gratitude to BPB Publications for granting me this opportunity. Writing a book involves the collaboration of many individuals, and I am thankful to everyone involved for their support and assistance throughout this endeavor.

Preface

Welcome to the second edition of Mastering Power BI. Since the release of the first edition, the world of data analytics has continued to evolve at a rapid pace. Power BI desktop has introduced new tools, techniques, and features, reshaping the way we analyze and visualize data. In this dynamic landscape, staying ahead requires not just keeping up with the latest advancements, but also mastering the core principles that underpin them.

This book provides a comprehensive guide in unleashing the full potential of Power BI. Whether you are a seasoned data professional or just beginning your journey into the world of business intelligence, this book is designed to equip you with the knowledge and skills needed to succeed.

In this second edition, we have updated and expanded upon the content of the first edition to ensure it remains relevant and valuable in today's data-driven world. It incorporates new features and functionalities of Power BI, explores advanced data transformation techniques, covers **Data Analysis Expressions (DAX)**, introduces the much-needed **Artificial Intelligence (AI)** concepts, and offers practical step-by-step exercises to illustrate all the concepts.

Throughout the book, readers will find step-by-step instructions and hands-on exercises designed to reinforce learning and empower them to tackle real-world challenges with confidence. Whether you are building robust data models, creating visually stunning reports and dashboards, or uncovering hidden data insights using AI, Mastering Power BI will be your trusted companion on your journey to becoming a Power BI expert.

The primary challenge I faced in writing this book was the changes introduced by Power BI. A few times, the look-and-feel of the application has changed. I have tried my best to keep up with the changes but in case you find some discrepancies, remember the changes are only in the interface not in the concepts.

With this book, you will gain the knowledge and skills to become a proficient Power BI developer.

I invite you to explore the possibilities, and embark on a transformative journey with Power BI. Let us unlock the full potential of your data together. I hope you will find this book informative and helpful.

Chapter 1: Understanding the Basics – This chapter covers the basics of Business Intelligence and explains all the important terms and definitions. It also explains the different components of Power BI and what they do. It will assist you in installing the Power BI desktop and provide an overview of the data tables used in the book.

Chapter 2: Connect and Shape – This chapter explains how Power BI connects to data from disparate sources, such as database tables, XLS files, relational database and many more. It also introduces you to Query Editor and describes how to use it to shape your data.

Chapter 3: Advanced Data Transformations – In this chapter, we will cover advanced data transformations in Power BI. It explains the data profiling techniques using Query Editor and how to perform data transformations on text, numeric data, and dates. It also covers the use data source parameters to manage data connections.

Chapter 4: Optimize Your Data Model – This chapter deals with creating and optimizing a data model in Power BI. It covers relationships and how to create them in Power BI. It also explains how to create joins by using merge and append functionalities.

Chapter 5: Data Analysis Expressions – This chapter explains DAX, which is essential for any Power BI implementation. The chapter covers the concepts of calculated columns and measures. Additionally, it explains all the main functions and shows how to implement them.

Chapter 6: Visualizations in Power BI – Visualizations are essential as they assist in understanding the data. This chapter covers all the visualizations available in Power BI and explains when to use them.

Chapter 7: Drill Through and Drill Down Reports – This chapter covers how to get more insights into the data by using the Drill through and drill down reports in Power BI. It also elucidates how to create custom data hierarchies.

Chapter 8: Artificial Intelligence in Power BI – This chapter shows the importance of AI and how it is used in Power BI. It explains the use of AI visuals to get better data insights.

Chapter 9: Power BI Service – This chapter describes the components of Power BI Service. You will deploy the application created in the Power BI desktop to Power BI Service. This chapter also teaches how to connect to data and create visualizations in Power BI Service.

Chapter 10: Securing Your Application – We have introduced the concept of **Row-level security (RLS)**. It is essential to secure the data to prevent unauthorized access and users see only the data they are authorized to view.

Code Bundle and Coloured Images

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

<https://rebrand.ly/9h0nxf0>

The code bundle for the book is also hosted on GitHub at

<https://github.com/bpbpublications/Mastering-Power-BI-2nd-Edition>.

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. Understanding the Basics	1
Introduction.....	1
Structure.....	1
Objectives	2
Understanding business intelligence	2
<i>Advantages of BI system</i>	<i>2</i>
<i>Components of BI</i>	<i>3</i>
<i>Data sets</i>	<i>3</i>
<i>Extract, Transform, and Load</i>	<i>3</i>
<i>Data warehouse</i>	<i>3</i>
<i>Data mart</i>	<i>4</i>
<i>Data model</i>	<i>5</i>
Concepts of the star and snowflake schema.....	7
<i>Star schema</i>	<i>7</i>
<i>Snowflake schema</i>	<i>8</i>
Key performance indicator	9
<i>Visualization</i>	<i>9</i>
<i>Dashboard.....</i>	<i>10</i>
Power BI as a business intelligence application.....	10
<i>Functions of Power BI</i>	<i>11</i>
<i>Power BI components</i>	<i>11</i>
<i>Power BI environment</i>	<i>13</i>
Different users of Power BI.....	14
<i>Power BI desktop developer</i>	<i>14</i>
<i>Power BI analyst</i>	<i>15</i>
<i>Power User</i>	<i>15</i>
<i>Executive user</i>	<i>16</i>
Power BI licensing.....	16
Installation of Power BI desktop	17
<i>Installation</i>	<i>17</i>
<i>Configuring Power BI desktop options and settings.....</i>	<i>18</i>
Power BI desktop interface	21

Development life cycle in Power BI desktop	23
<i>Initiate a Power BI implementation as a developer</i>	24
<i>Overview of the data used in this book</i>	25
<i>Power BI development setup for learning</i>	26
Conclusion	27
Questions	27
<i>Answers</i>	28
2. Connect and Shape	29
Introduction	29
Structure	29
Objectives	30
Data connections in Power BI	30
Connecting to data	38
<i>Connecting to database tables</i>	38
<i>Loading the table</i>	38
Understanding Power Query Editor	40
<i>Verifying the loaded query</i>	44
<i>Changing the default summarization</i>	44
<i>Changing the display format of the numbers</i>	45
<i>Review your model</i>	46
Loading customers and products tables	47
<i>Verifying the loaded query</i>	49
<i>Changing the default summarization for the Products table</i>	49
<i>Review your model</i>	50
Loading data from an Excel file	50
<i>Loading Category data from an Excel file</i>	50
<i>Loading multiple CSV files from a folder</i>	52
Loading data from MS SQL server	54
<i>Import and DirectQuery</i>	54
<i>Creating static table in Power BI</i>	56
<i>The need for a static table in the model</i>	57
<i>Review the data model</i>	59
Bonus section	60
<i>Connecting to web data source</i>	60

Loading data from SharePoint.....	62
Connecting to SharePoint using ODataFeed	63
Connecting to the SharePoint Lists using the SharePoint connector.....	65
Useful transformations using Query Editor.....	70
Applying Group By to CustomerSales_Report.....	75
Group By to calculate Total by Customer	76
Advanced group by to calculate totals	77
Creating a Column in Query Editor.....	78
Conclusion.....	79
Questions.....	80
Answers.....	80
3. Advanced Data Transformations.....	81
Introduction.....	81
Structure.....	81
Objectives	82
Data profiling using Query Editor.....	82
Data profiling	82
Data profiling options in Query Editor	83
Performing data profiling on a table	84
Validating Column quality	86
Understanding the errors reported by Column quality.....	87
Validating column distribution	89
Caution before taking Query Editor suggestion	90
Validating column profile.....	90
Data transformation on text data	91
Trim leading and trailing spaces	92
Delimiter option to create a new column.....	93
Using the Merge Columns option.....	94
Data transformation on Numeric Data	95
The statistics option	95
The standard option	96
Data transformation on Dates	97
The gear symbol in applied steps.....	98
Utilizing Data Source parameters.....	100
Using a parameter to change the data source	100

Conclusion.....	102
Questions.....	103
Answers.....	103
4. Optimize Your Data Model	105
Introduction.....	105
Structure.....	105
Objectives	106
Introduction to data modeling	106
<i>Best practices of data modeling</i>	<i>106</i>
Review and create relationships.....	107
<i>Establishing relationships</i>	<i>107</i>
Review the loaded tables.....	108
<i>Review the orders and products relationship</i>	<i>109</i>
<i>Manually linking orders and customers table</i>	<i>110</i>
<i>Manually linking products and category tables.....</i>	<i>112</i>
<i>Manually linking orders and sales tables.....</i>	<i>112</i>
<i>Review the data model.....</i>	<i>112</i>
<i>About model explorer in model view.....</i>	<i>113</i>
Combining queries in Power BI	114
<i>Combining queries using joins.....</i>	<i>114</i>
Joins	115
<i>Loading OrderDetails.xlsx</i>	<i>115</i>
<i>Left outer join.....</i>	<i>116</i>
<i>Right outer join.....</i>	<i>118</i>
<i>Full outer join</i>	<i>119</i>
<i>Left anti.....</i>	<i>119</i>
<i>Right anti.....</i>	<i>120</i>
<i>Handling Multiple Fact tables</i>	<i>120</i>
<i>Combining queries using append.....</i>	<i>120</i>
<i>Append records from NewCustomers table to Customers</i>	<i>121</i>
Conclusion.....	123
Questions.....	123
Answers	124

5. Data Analysis Expressions	125
Introduction.....	125
Structure.....	125
Objectives	126
Introduction to DAX	126
<i>DAX functions</i>	126
<i>Overview of DAX functions</i>	127
Calculated columns and measures	128
<i>Calculated columns</i>	128
Exercise 1	129
<i>Verify the Unit Price calculation</i>	130
Exercise 2	131
<i>Calculated measure</i>	132
Exercise 3	133
<i>Quick measures</i>	134
<i>Creating a quick measure</i>	135
Exercise 4	136
Mathematical functions	138
Sum	138
Exercise 5	138
SumX.....	139
SumX versus Sum.....	140
Using SumX	140
Exercise 6	140
Count functions	141
Count.....	141
Exercise 7	141
CountA.....	142
Exercise 8	142
Using Count and CountA.....	143
CountX.....	145
CountAX.....	145
CountBlank	145
CountRows.....	145
Information functions	146

ISERROR.....	146
Exercise 9	146
USERNAME	146
Exercise 10	146
LOOKUPVALUE.....	147
LOOKUPVALUE: SalesPerson_Location and Sales Person.....	147
Hiding the redundant tables.....	148
Logical functions	150
IF	150
Exercise 11.....	150
And.....	151
Exercise 12	151
Switch.....	152
Exercise 13	152
Filter functions.....	153
Filter	153
Calculate.....	153
Exercise 14	154
CalculateTable	154
Exercise 15	155
Optional example	155
ALL.....	156
Exercise 16	156
Related.....	157
Exercise 17	157
RelatedTable.....	159
Values	159
Exercise 18	159
Date and time functions	160
Calendar	160
Exercise 19	160
DateDiff.....	161
Exercise 20	161
Variables in DAX expressions.....	162
VAR statement.....	162

Exercise 21	163
Conclusion.....	164
Questions.....	164
Answers	164
6. Visualizations in Power BI.....	165
Introduction.....	165
Structure.....	165
Objectives	166
Review the data model.....	166
Understanding visualization	166
Introduction to Power BI reports	167
Creating a report	167
Standardizing report development.....	171
Report themes.....	171
Applying built-in theme.....	171
Creating and using templates	173
Power BI desktop projects	174
Creating multi-page reports using visualizations	174
Remove any pages or visualizations.....	175
Creating visualizations	175
Card.....	175
Exercise 1	176
Copy/paste a visual	177
Bar and column chart.....	179
Exercise 2	179
Filters and slicers	181
Line and area chart.....	185
Exercise 3	186
Line and stacked column chart.....	188
Exercise 4	188
Ribbon chart.....	189
Exercise 5	189
Waterfall, funnel, and scatter charts.....	191
Exercise 6	191
Scatter chart	192

Exercise 7	192
Pie, donut, and treemap charts	195
Exercise 8	195
Treemap	196
Exercise 9	196
Highlighting and Cross filtering.....	197
Exercise 10	197
Map	198
Exercise 11.....	199
Conclusion.....	201
Questions.....	201
Answers	202
7. Drill Through and Drill Down Reports	203
Introduction.....	203
Structure.....	203
Objectives	204
Drill through report in Power BI	204
<i>Creating a drill-through report</i>	204
<i>Exercise</i>	205
<i>Performing drill through action</i>	206
Drill down report in Power BI	207
<i>Exercise</i>	208
Creating custom hierarchy	209
<i>Creating a location hierarchy</i>	209
<i>Exercise</i>	209
<i>Changing the levels of hierarchy</i>	210
<i>Exercise</i>	211
Conclusion.....	212
Questions.....	212
Answers	212
8. Artificial Intelligence in Power BI.....	213
Introduction.....	213
Structure.....	213
Objectives	214

Artificial Intelligence in Power BI	214
Why AI visuals are used.....	214
AI visuals in Power BI desktop.....	215
<i>Preparing the application for AI</i>	215
<i>Insights using Analyze option</i>	216
<i>Exercise</i>	216
<i>Key influencers visual</i>	218
<i>Exercise</i>	218
<i>Explanation of key influencer visualization</i>	219
<i>Decomposition tree visual</i>	220
<i>Exercise</i>	220
<i>Performing manual analysis using decomposition tree</i>	222
<i>Performing AI suggested analysis using decomposition tree</i>	222
<i>Q&A Visual</i>	224
<i>Exercise</i>	225
Narrative visual.....	226
<i>Working of smart narrative</i>	226
<i>Exercise</i>	226
AI in Query Editor	228
<i>Text Analytics and Vision</i>	228
<i>Azure Machine Learning</i>	228
Conclusion.....	229
Questions.....	229
Answers	229
9. Power BI Service.....	231
Introduction.....	231
Structure	231
Objectives	232
Understanding the Power BI service.....	232
<i>Power BI service workflow</i>	232
<i>Power BI service interface</i>	232
Foundational elements of Power BI Service.....	234
Publishing reports from the Power BI desktop.....	235
Visualizations in Power BI Service	237

Creating a dashboard.....	238
Conclusion.....	241
Questions.....	241
Answers	241
10. Securing Your Application	243
Introduction.....	243
Structure.....	243
Objectives	244
Row-level security in Power BI.....	244
Implementing Row-Level Security	244
<i>Exercise</i>	244
Configure RLS in Power BI service.....	249
Conclusion.....	251
Questions.....	252
Answers	252
Index	253-258

CHAPTER 1

Understanding the Basics

Introduction

Welcome to the foundational chapter of Mastering Power BI — a comprehensive guide designed to empower you with the skills and insights needed to harness the full potential of Power BI. We will build a solid understanding of the basics, laying the groundwork for your mastery of this powerful business intelligence tool.

In this chapter, we will learn about the basics of Power BI. Power BI is growing in popularity due to the functionality it provides to the business users. The chapter will cover the basics of Power BI as a Business Intelligence application. It will start with the **Business Intelligence (BI)** fundamentals and explain the terms and technologies in the BI paradigm. You will learn about Star Schema, Snowflake schema, and gain more understanding on Dimensions and Facts the building blocks of any BI application.

We will learn about Power BI and understand how it works. This chapter is important as it will lay the foundation of all the subsequent chapters in this book.

Structure

In this chapter, we will discuss the following topics:

- Understanding business intelligence
- Concepts of the star and snowflake schema

- Key performance indicator
- Power BI as a business intelligence application
- Different users of Power BI
- Power BI licensing
- Installation of Power BI Desktop
- Power BI Desktop Interface
- Development life cycle in Power BI desktop

Objectives

Understanding the concepts of business intelligence is a key to success in Power BI. After completing this chapter, you will be able to explain BI and its terminology, like the **star schema**, and the **snowflake schema**. You will also be able to differentiate between the dimension and fact tables, which is the key to designing powerful data models and visualizations. You will also learn what is Power BI and how it works.

Understanding business intelligence

Before defining the term BI, let us understand the terms data and information. In the world of **Information Technology (IT)**, data can be anything – text, numbers, or images in a digital format. The data is raw, unorganized, or arbitrary, but should be in a format that is understandable to a computer system. Once loaded, the data is transformed, processed, and interpreted by the system to produce meaningful and contextual information.

In the business world, data and information are closely related and thus used interchangeably.

BI relates to the set of technologies and techniques that collect and categorize an organization's data and present meaningful information in a format that helps make better decisions. The BI applications allow the developers to collect vast amount of data from diverse sources, transform the data according to the business requirements, and present it in a visual format like tables and charts. BI does not make decisions for an enterprise, but eases the analysis of data to arrive at actionable results.

Advantages of BI system

An enterprise can drive huge benefits by implementing a BI System, which is as follows:

- **Data management:** BI system facilitates the collection of data from diverse sources. This data is stored in an enterprise-wide data warehouse or a data mart. Since the data is centrally stored, it helps in producing a single version of the truth.

- **Information broadcasting:** The information is delivered in a visual format that is understandable to the users. It helps in the quick delivery of information in the form of interactive dashboards, tables, charts, and maps. The users can get to the data faster and collaborate with the information.
- **Secure delivery:** BI System also supports secure information delivery, that is, the data is contextual and is delivered on a need-to-know basis. The visualizations can be developed that display the different data based on the organization's role or organization structure.
- **Adhoc analysis:** The business users can use the self-service BI applications to perform their own data analysis. This will reduce the dependency on the IT technical team.

Components of BI

Before plunging into a BI application like Power BI, it is important to learn about some of the BI components.

Data sets

The core use of a BI application is to enhance the understanding of data. Data can come from disparate sources. It can be sourced from legacy systems, relational database, cloud, or from various file-based applications, such as Excel, CSV, or flat files. The data can be unstructured, such as emails or webpages.

Extract, Transform, and Load

Extract, Transform, and Load (ETL) is a process of data integration and is used to combine disparate data arriving from multiple sources such as Excel, CSV, RDBMS, ERP, and so on. In an ETL process, the data is extracted from the source, transformed to aggregate or to implement business rules, and then loaded into the target system. The data loaded in the target system is used for analysis. The ETL process is used to build a data warehouse.

Data warehouse

A data warehouse is a process of managing large amounts of data in an organization. It is designed to assist in the BI tasks, especially in analytics. Given a large amount of historical data, a data warehouse enables faster data query and analysis. A typical data warehouse contains historical data, which is derived from a variety of sources, such as operational or transactional databases.

A data warehouse works as a central repository of the aggregated data from multiple sources and provides an organization with a single version of truth. Since it contains historical records, it empowers the data scientists and data analysts in improved decision making and predictive analysis.

A typical data warehouse contains the following:

- A relational database to store and manage the data. This relational database is created in any of the applications, such as Oracle, SQL Server, or db2, and so on.
- An ETL process to extract data from the multiple sources, transform, and aggregate the data according to the organizational needs and load the data in the data warehouse.
- Data analysis and visualization applications, such as Power BI to assist in the analysis of the data.

The data warehouse is a core component of a BI implementation.

Data mart

Data mart is similar to the data warehouse but contains only the specific business data within an organization. A data warehouse is a central repository of an enterprise-wide data, while a data mart contains the subset of data pertaining to a specific business or user function. Data warehouse and the data marts are used for reporting and analysis. A data mart can be sourced from a data warehouse.

A typical data warehousing environment is shown in the following figure:

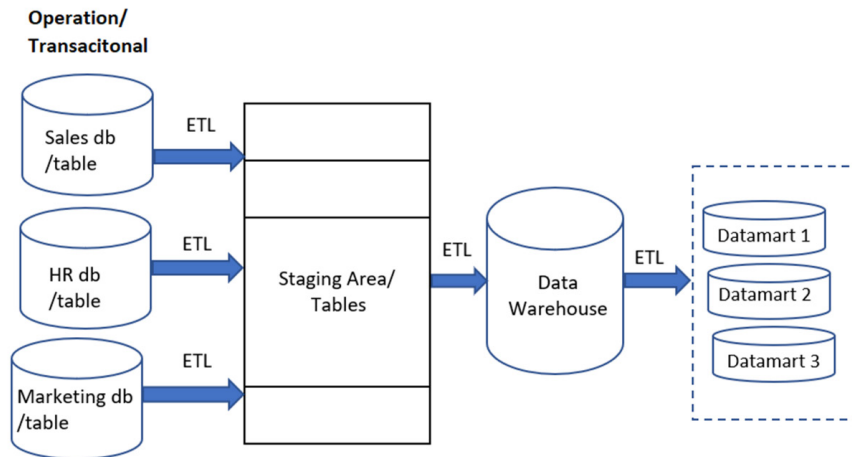


Figure 1.1: Data warehousing environment

The preceding *Figure 1.1* shows how the data warehouse and the data marts are created, and are explained in words as follows:

- An ETL process is the run process to extract the data from the various operational or transactional databases or tables.
- This data is stored in the staging tables.

- A different ETL process is created to extract and transform this data, which is loaded in the data warehouse database.
- From the data warehouse, separate business or user specific data marts are created.

A data warehouse is typically created for the reporting and analytical needs of the organization. It helps in the data analyses by reducing the number of tables and joining the ones that are atypical of an operational database.

Data model

A **data model** displays how the different data entities are related in a data warehouse environment. It presents a pictorial format, showing different tables and the relationships between them. In BI, a data model represents the organization's data and should be designed for faster data access. It should contain all the data categories, hierarchies, and filters. The most popular data model used in a data warehouse is the dimensional model, which is also called the star schema. The two types of tables in a star schema are the dimension table and the fact table. Let us have a look at them briefly:

- **Dimensions and facts:** Dimension and facts tables are the main ingredients of any BI implementation. These tables are used to form the star or the snowflake schemas, which are designed as part of building a data warehouse or a data mart.
- **Dimension table:** The dimension tables contain the descriptive or qualitative attributes of the data. For example, the customer dimension may contain information about the customer, such as the name, address, contact number, and so on. The dimension fields usually contain the characters or the textual type of data. The dimension tables are constructed from the operational or transactional relational database. The dimension tables contain the primary key with the respective foreign key in the fact table. A dimension table provides context to a fact table.

There are different types of dimension tables. Some of the commonly used ones are as follows:

- **Slowly Changing dimensions (SCD):** It is a dimension table where the row of the data in the table varies with time. It is used to track the current and historical data. SCD is implemented in the following three ways:
- **Type1:** In type1 SCD, the existing row of data is simply overwritten. No history is maintained, and the existing data is lost. For example, consider the following employee record:

EmpID	EmpName	Department
101	John	Administration

Table 1.1: Employee record

If employee John changes his department to HR, no history will be maintained. The record will simply be overwritten, as shown in the following table:

EmpID	EmpName	Department
101	John	HR

Table 1.2: Employee changes department

- **Type2:** The type2 SCD keeps the complete history of the data by creating a new record with the Start date and the End date. Only one record will be active at a time. This is the most popular way of storing the historical data. For example, consider the following employee record:

EmpID	EmpName	Department	Start_Date	End_Date	Active
101	John	Admin	01-15-2010	12-31-9999	Y

Table 1.3: Employee history with active indicator

If employee John moves to a new department, HR, a new record will be added to keep the history:

EmpID	EmpName	Department	Start_Date	End_Date	Active
101	John	Admin	01-15-2010	03-31-2019	N
101	John	HR	01-15-2010	12-31-9999	Y

Table 1.4: Employee history with only one department Active

- **Type3:** In type3 SCD, the history of the data is maintained by using the **Current_Value** and the **New_Value** columns. It is cumbersome to maintain the history, as it is limited by the number of columns needed to store the historical data. This technique is not frequently used. For example, consider the following employee record:

EmpID	EmpName	Department	Current_Value	Previous_Value
101	John	Admin	Admin	Admin

Table 1.5: Employee history using current and previous value columns

If employee John moves to a new department, HR, the history is maintained by putting the new department under the **Current_Value** column:

EmpID	EmpName	Department	Current_Value	Previous_Value
101	John	Admin	HR	Admin

Table 1.6: Employee record with changed current value