

Microsoft Power BI Data Analyst Exam Guide

A practical guide to becoming a Power BI professional

Peter ter Braake



www.bpbonline.com

First Edition 2026

Copyright © BPB Publications, India

ISBN: 978-93-65894-325

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 cannot 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 and correct to the best of author's and publisher's knowledge. The author has made every effort to ensure the accuracy of these publications, but the 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:



Dedicated to

My family

About the Author

Peter ter Braake is currently working as an independent Microsoft data platform professional. He studied physics at the University of Utrecht, the Netherlands, in the late 1980s and early 1990s. Quickly after graduating, he went into IT and found his passion to be data. He became a **Microsoft Certified Teacher (MCT)** in 2002. This was also the time that he became Microsoft Certified. His first certification was Microsoft Certified Database Administrator: Microsoft SQL Server 2000. He has kept his Microsoft certifications up to date ever since. He was a Microsoft **Most Valued Professional (MVP)** for a couple of years.

Currently, Peter works a lot with Microsoft Fabric as a data engineer, Power BI as a data analyst, and Azure Machine Learning as a data scientist.

As an MCT, he teaches a lot of courses at various training providers and on-site for his customers. These courses can be official Microsoft courses or customized courses that target specific goals for their customers.

As a consultant, Peter helps his customers with advice on various data-related issues. These are often design or performance issues.

About the Reviewer

Riccardo Perico began his career in IT back in 2010. After a brief stint as an ERP consultant, he transitioned into the data realm.

Over the years, he explored the world of data from multiple angles, working as both a database administrator and a BI architect. His expertise spans across databases, BI, and cloud technologies, and he is a Microsoft Certified Professional in several areas, including databases, BI, and Azure.

A four-time Microsoft MVP in the data platform category and a recognized Microsoft Fabric Super User, Riccardo is a well-known figure in the data community. He is one of the leaders of the Power BI User Group Italy and has spoken at major events such as the Power Platform World Tour and Global Power Platform Bootcamp in Italy.

His speaking engagements extend across Europe, with appearances at Azure Saturday, SQL Saturday, Data Saturday, and various other community-driven events.

In addition to speaking, Riccardo has contributed to the data community as a technical reviewer for several notable publications.

Last but not least, he is a hard rock and heavy metal enthusiast.

Acknowledgement

I would like to acknowledge the valuable contributions of my colleagues and co-workers during the many years working in the tech industry who have taught me so much and inspired me to look beyond my daily work routine. Moreover, I would like to specially thank Karthikeyan Sabesan for taking an active interest in contributing to the book's contents.

I am thankful to my family for their patience, motivation and encouragement at every juncture and for constantly pushing me to pursue my ambitions and goals. We appreciate each other as a team, to have kept up the resolve to take this book to completion and inspiring each other.

I am grateful to BPB Publications for their guidance and expertise in bringing this book to fruition. It was a long journey of revising this book, with valuable participation and collaboration of technical reviewers, and editors.

Most of all, I would like to thank all the readers who have taken an interest in this book and for their support in making it a reality. Your encouragement has been invaluable.

Preface

Working with data, or better phrased, understanding your world by the use of data, is core to almost every successful business nowadays. Power BI is a powerful tool to help you with that. This book aims to help you get the most out of Power BI and, with that, out of the opportunities that lie ahead of you. It also aims to guide you in passing Microsoft's official Power BI exam: Microsoft Certified: Power BI Data Analyst Associate.

Comprising 16 insightful chapters, this book covers all topics essential for working with Power BI and passing the exam successfully. The book starts with some necessary background on why data is important and why data modeling is even more important.

Chapters 3 and 5 focus on preparing data using Power Query. Chapter 3 starts with applying basic data transformations. Chapter 5 focuses on automating solutions or being prepared to get new data in the future.

Chapter 4 and 6 focus on visualizing the data, where Chapter 4 starts with basic theory on what an effective report is. Not all visuals are insightful, so careful consideration on how to setup your report is essential in creating effective reports. Chapter 6 focuses on more advanced concepts and especially on creating compelling interactive reports.

Chapters 7 through 10 focus on creating a semantic model. Chapter 7 explains the basics of semantic models and especially the relationships between tables, which are an important aspect of the model. Chapters 8 and 9 focus on **Data Analysis Expression (DAX)**. DAX is the language behind the semantic model that allows us to enrich our data analysis capabilities. Chapter 10 introduces concepts and best practices to make Power BI scalable and allow for greater datasets.

Chapters 11 through 13 describe the Power BI portal and collaborating with the reports that have been created. Chapter 11 focuses on security, a topic with ever-increasing significance. Chapter 12 provides a tour of the portal, and Chapter 13 focuses on the actual end result: creating a Power BI app.

To finish the book, Chapter 14 discusses the monitoring of the entire Power BI system. Chapter 15 introduces **artificial intelligence (AI)** capabilities in Power BI.

This book is designed to cater to all professionals who want to start working with Power BI and/or want to pass the exam PL-300 Microsoft Certified: Power BI Data Analyst Associate. There are no prerequisites for reading the book. However, affinity with data is a must, and some affinity with the use of code is a prerequisite.

Through practical examples, comprehensive explanations, and a structured approach, this book aims to equip readers with a solid understanding of Power BI. Whether you are a novice or an experienced learner, I hope this book will serve as a valuable resource in your journey of exploring Power BI.

Chapter 1: Introduction to Data and Power BI- **Business intelligence (BI)** is a process to prepare and use data to gain a competitive advantage. Lessons learned over the last decades have made people understand that data needs to be combined, prepared, and modeled before it can be used. Those lessons learned can be found in the setup of Power BI. Understanding the evolution of business intelligence is understanding the components of Power BI. The first chapter explains the components found in the tool Power BI from a historical perspective.

Chapter 2: Dimensional Modeling- Power BI is a tool that allows you to create compelling visualizations of your data with just drag-and-drop. However, to make that work as easily as it sounds, the data must be modeled correctly. This chapter teaches you the basics of dimensional modeling. It is a commonly accepted best practice to model your data into a dimensional model before starting to use the data.

Chapter 3: The Basics of Power Query- This chapter introduces the basic working of Power Query. Power Query is used to transform and prepare the data. There are a lot of data transformations available in order to deal with miscellaneous data or data modeling issues. This chapter focuses on the often-used basic transformations. The chapter uses the theory of the previous chapter and transforms our sample data into facts and dimensions.

Chapter 4: The Basics of Visualizations- This chapter provides an introduction to visualizing data. There are a lot of different types of visualizations to choose from. There are a lot of questions to be asked of the data. Choosing which visualization answers the question in the most effective way is an important part of developing reports and dashboards. This chapter is about choosing the right visuals and how to actually create the visual. The chapter also describes how to change the formatting of the visual to your personal liking.

Chapter 5: Advanced Techniques of Power Query- This chapter focuses on automating Power Query solutions. Depending on the data you have and the questions you need answered using that data, Power Query can be straightforward. It can also become rather complex. You need to be prepared for complex situations, where, for instance, new data will be stored in new files. This chapter focuses on the language M of Power Query and on how to use it to automate your data loading and preparation solutions.

Chapter 6: Create Interactive Reports- This chapter explores the interactivity capabilities of Power BI reports. The chapter builds on the skills learned in *Chapter 4, The Basics of Visualizations*

to make the reports more effective by allowing users to be actively involved with the data in the report. It introduces drill down and drill through, amongst other things, to allow users to really get insights into their business processes.

Chapter 7: The Basics of Semantic Models- You cannot create reports in Power BI without a semantic model. The semantic model is basically the part that makes drag-and-drop reporting possible. The reports created so far in the book are based on a default semantic model that is automatically created when data is imported into Power BI. This chapter explores the semantic model. It explains the various types of relationships between tables of data and when to use which relationship. It also explains a lot of other settings that make the result of drag-and-drop reporting better.

Chapter 8: DAX- This chapter introduces DAX. DAX is the programming language behind the semantic model. It can be used to enrich the data model. With DAX, we can create calculations, calculated columns, and calculated measures. Sooner or later, you will have to use DAX to answer the questions that live within an organization. This chapter uses a lot of practical examples but focuses mainly on understanding the concept behind DAX. It explains the concept of context and shows how to play with context to make reports even more informative.

Chapter 9: Advanced DAX Concepts- This chapter goes one step further than the previous one. 80% of what you need to do using DAX can be simply done by using all the DAX functions that are readily available. Some of the real added benefits will take a little more effort. You need to understand iterators and context switches. This chapter teaches the slightly more difficult techniques of DAX.

Chapter 10: Scalable Power BI Solutions- This chapter focuses on larger datasets. Larger can be defined here as the scenario when you have issues within Power BI with everything taught so far in the book, because of the amount of data you have. This chapter teaches you concepts and tricks to use when you have large datasets. You will learn about different storage modes, aggregations, incremental refresh, and more.

Chapter 11: Security- This chapter focuses on securing Power BI. Security can be found at different levels throughout Power BI. The chapter teaches how to configure **row level security (RLS)** within a report. It also describes object level security and how to secure semantic models, reports, and dashboards once they are published to the Power BI service.

Chapter 12: Working with the Power BI Service- The Power BI service is the cloud portal where semantic models and reports are published to, once we need to collaborate on them with multiple people. It is also the place where we can create Power BI dashboards. This chapter focuses on functionality and settings found in the service. It describes workspaces,

domains, folders, and a whole lot of settings that you can adjust to customize your Power BI environment.

Chapter 13: Create App- After the previous chapter, you will realize that the Power BI portal can be a daunting environment for users who just need to use the reports. For these users, we can create Power BI apps. This chapter describes what Power BI apps are for, how you can create and customize them for different audiences, and how to share them with the intended users.

Chapter 14: Monitor Power BI and Fabric- This chapter focuses on keeping an eye on the system. Our reports are made available to users, and you need to keep track of their usage. Some reports might not be used at all or not anymore over time. Some reports might grow, either in how often they are used, by how many different users they are used by, or in the amount of data behind the report. Knowing how the system is used helps in keeping the system healthy. It also helps in knowing whether you are using the right license. This chapter describes the tools available for administrators to monitor Power BI and its usage.

Chapter 15: Copilot in Power BI- This chapter introduces AI in Power BI. Copilot, Microsoft's AI assistant, is integrated into Power BI in multiple ways. This chapter describes where and how you can use Copilot throughout Power BI.

Chapter 16: Practice Exam- Since this book helps you prepare for the official Microsoft exam PL-300 Microsoft Certified: Power BI Data Analyst Associate, it helps to know what sort of questions you can expect on the exam. This chapter provides some example questions that help you prepare for the exam even better.

Code Bundle and Coloured Images

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

<https://rebrand.ly/d952da>

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

<https://github.com/bpbpublications/Microsoft-Power-BI-Data-Analyst-Exam-Guide>.

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 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 feedback are highly appreciated by the BPB Publications' Family.

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. You can check our social media handles below:



Instagram



Facebook



Linkedin



YouTube

Get in touch with us at: **business@bpbonline.com** for more details.

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 Discord space

Join our Discord workspace for latest updates, offers, tech happenings around the world, new releases, and sessions with the authors:

<https://discord.bpbonline.com>



Table of Contents

1. Introduction to Data and Power BI.....	1
Introduction.....	1
Structure.....	1
Objectives	2
Introduction to business intelligence.....	2
<i>Data in organizations.....</i>	<i>4</i>
Historical overview	5
<i>Data complexity</i>	<i>6</i>
<i>Impact on source systems</i>	<i>7</i>
<i>Data quality.....</i>	<i>7</i>
<i>Lack of a consistent 360-degree view.....</i>	<i>9</i>
<i>Lack of historical data.....</i>	<i>9</i>
<i>Data warehouse</i>	<i>10</i>
Introducing Power BI.....	12
<i>Power BI ideal sources.....</i>	<i>14</i>
Power BI use case	15
<i>Corporate reporting</i>	<i>15</i>
<i>Ad-hoc analysis</i>	<i>16</i>
<i>Other use cases</i>	<i>16</i>
Power BI Desktop and the Power BI service	16
<i>Power BI Desktop.....</i>	<i>17</i>
<i>Power BI service</i>	<i>17</i>
Power BI licensing.....	18
<i>Free</i>	<i>18</i>
<i>Power BI Pro</i>	<i>18</i>
<i>Power BI Premium Per User.....</i>	<i>19</i>
<i>Power BI Embedded</i>	<i>19</i>
<i>Fabric capacity reservation.....</i>	<i>19</i>
<i>Fabric capacity pay-as-you-go.....</i>	<i>19</i>

Conclusion.....	20
Multiple choice questions	20
<i>Answer</i>	21
2. Dimensional Modeling.....	23
Introduction.....	23
Structure.....	24
Objectives	24
Dimensional modeling	24
<i>Normalizing data</i>	25
<i>Data Vault</i>	25
<i>Dimensional modeling</i>	26
Steps to design a star schema	27
<i>Choose process</i>	27
<i>Determine grain</i>	28
<i>Northwind database example</i>	29
<i>Defining facts</i>	30
<i>Defining dimensions</i>	32
Designing a fact table.....	34
<i>Additive facts</i>	35
<i>Non-additive facts</i>	35
<i>Semi-additive facts</i>	36
<i>Designing facts</i>	36
<i>Northwind database example</i>	37
Design dimension tables	39
<i>Attribute design</i>	39
<i>Snowflaking</i>	40
<i>Slowly changing dimensions</i>	42
<i>Date dimension</i>	43
<i>Northwind example</i>	45
Conclusion.....	48
Multiple choice questions	48

<i>Answers</i>	49
3. The Basics of Power Query	51
Introduction.....	51
Structure.....	51
Objectives	52
Get data.....	52
<i>Text or comma separated values</i>	56
<i>Excel</i>	58
<i>SQL Server</i>	58
Select columns.....	60
De-duplicate rows and generate keys	62
Merge tables	63
Custom columns.....	68
Data types.....	71
Filtering rows	77
Other transformations	79
Challenge.....	79
Close & Apply.....	80
Conclusion.....	81
Multiple choice questions	82
<i>Answers</i>	82
4. The Basics of Visualizations	83
Introduction.....	83
Structure.....	83
Objectives	84
A tour of the interface	84
Creating a bar chart.....	88
Setting visualization properties.....	90
Adding a slicer to the report.....	93
Table visuals	96
Adding filters to your report	98

Pie chart	100
Visual interactions	101
Configuring report page	102
Paginated reports	103
Conclusion	104
Multiple choice questions	104
<i>Answers</i>	104
5. Advanced Techniques of Power Query	105
Introduction	105
Structure	105
Objectives	106
Date dimension	106
Analyzing the raw data	109
Unpivot data	110
Creating a custom function	112
Reading multiple files from a folder	114
Understanding M-code	118
Understanding the date table code	121
Using parameters	123
Organizing queries	125
<i>Optimizing queries and query folding</i>	125
Exam tip	126
Conclusion	127
Multiple choice questions	127
<i>Answers</i>	128
6. Create Interactive Reports	129
Introduction	129
Structure	130
Objectives	130
Pitfalls and considerations	130
<i>Pitfalls</i>	131

<i>Important considerations</i>	132
<i>Determine the message</i>	132
<i>Determine the user</i>	133
Preferred visualization types	134
<i>Using tooltips</i>	135
<i>Using hierarchies</i>	139
<i>Using drill through</i>	142
<i>Using bookmarks</i>	143
<i>Sync slicers</i>	145
Hiding items on a report	146
Using buttons	147
<i>Back, bookmark and page navigation</i>	148
<i>Web URL</i>	148
<i>Drill through</i>	149
Questions and answers	149
<i>Apply all slicers and clear all slicers</i>	150
Page navigator and bookmark navigator	150
Shapes and images	151
Using visual calculations	151
Conditional formatting	154
Other visuals	156
<i>Gauge, KPI</i>	156
<i>Map and filled map</i>	156
<i>Scatter chart</i>	156
<i>Get more visuals</i>	156
<i>Key influencers</i>	157
<i>Decomposition tree</i>	157
<i>Narrative</i>	157
Themes and templates	158
Conclusion	158
<i>Answers</i>	159

7. The Basics of Semantic Models.....	161
Introduction.....	161
Structure.....	162
Objectives	162
Introduction.....	162
Active and inactive relationships.....	163
Types of relationships	166
Cross-filter direction	169
Attribute properties	172
<i>Data type and default aggregation</i>	<i>173</i>
<i>Formatting.....</i>	<i>174</i>
<i>Hierarchies.....</i>	<i>174</i>
<i>Hide columns.....</i>	<i>175</i>
<i>Data categories</i>	<i>176</i>
<i>Grouping and binning.....</i>	<i>176</i>
<i>Display folder</i>	<i>177</i>
<i>Sort by column</i>	<i>177</i>
Table properties	178
<i>Mark as date table.....</i>	<i>178</i>
Other table properties.....	179
Conclusion.....	179
Multiple choice questions	180
<i>Answers.....</i>	<i>180</i>
8. DAX.....	181
Introduction.....	181
Structure.....	181
Objectives	182
Introduction to DAX	182
Calculated columns.....	183
Parent-child hierarchy	187
Calculated measures	190

<i>Implicit vs. explicit measures</i>	190
<i>% of parent</i>	192
Time intelligence functions	197
<i>Calculating growth</i>	200
<i>Quick measures</i>	201
Conclusion.....	203
Multiple choice questions	203
<i>Answers</i>	204
9. Advanced DAX Concepts	205
Introduction.....	205
Structure.....	205
Objectives	206
Variables.....	206
<i>Creating a date dimension</i>	209
Calculation groups	212
Calculate and context switches	216
<i>Moving average</i>	218
DAX queries and the <i>valuate</i> function.....	220
Conclusion.....	222
Multiple choice questions	222
<i>Answers</i>	223
10. Scalable Power BI Solutions	225
Introduction.....	225
Structure.....	226
Objectives	226
Connectivity mode.....	226
<i>Import</i>	226
<i>DirectQuery</i>	228
<i>Live connection</i>	231
<i>Direct Lake</i>	231
<i>Composite models</i>	232

Table aggregations.....	233
Incremental refresh and real-time data	238
Fabric.....	242
Best practices.....	243
Conclusion.....	245
Multiple choice questions	246
<i>Answers</i>	246
11. Security.....	247
Introduction.....	247
Structure.....	248
Objectives	248
Row level security	248
Publish a report to a workspace.....	252
Object level security	254
Sharing a report	255
Manage report permissions	258
Configure workspace access.....	259
Sensitivity labels	261
Conclusion.....	263
Multiple choice questions	263
<i>Answers</i>	264
12. Working with the Power BI Service	265
Introduction.....	265
Structure.....	265
Objectives	266
Configure semantic model connectivity and data refresh.....	266
Creating a dashboard.....	271
Working with reports in the Power BI service	274
Configure workspace settings	278
Power BI settings.....	279
Conclusion.....	281

Multiple choice questions	281
<i>Answers</i>	282
13. Create App	283
Introduction.....	283
Structure.....	283
Objectives	284
Deployment pipeline	284
Power BI projects.....	286
Creating a Power BI app.....	288
Working with apps.....	291
Template apps.....	292
Conclusion.....	293
<i>Answers</i>	294
14. Monitor Power BI and Fabric	295
Introduction.....	295
Structure.....	296
Objectives	296
Monitoring report usage metrics	296
View item or workspace lineage	299
Admin monitoring workspace	300
Microsoft Fabric Capacity Metrics app	301
Power BI Desktop tools	303
Power BI activity log.....	305
Application programming interface.....	306
Conclusion.....	307
Multiple choice questions	307
<i>Answers</i>	308
15. Copilot in Power BI	309
Introduction.....	309
Structure.....	309
Objectives	310

Generative AI prompts	310
Copilot for business users	310
Copilot for developers	312
<i>Understanding, enhancing and using the data model</i>	312
<i>Write DAX expressions and queries</i>	314
<i>Creating narrative visuals on your report</i>	316
Prep data for AI	317
Conclusion	317
Multiple choice questions	318
<i>Answers</i>	318
16. Practice Exam	319
Introduction	319
Question 1	320
Question 2	320
Question 3	320
Question 4	321
Question 5	321
Question 6	322
Question 7	322
Question 8	322
Question 9	323
Question 10	323
Question 11	323
Question 12	323
Question 13	324
Question 14	324
Question 15	325
<i>Answers</i>	325
Index	327-333

CHAPTER 1

Introduction to Data and Power BI

Introduction

This chapter introduces you to the concept of data analysis and analytics. It describes the context in which a data analyst performs his or her role. It also provides a historical overview of data warehousing. Data analysis is not new. Organizations have been using data for years already. By now, we have a good idea of what works and what pitfalls to watch out for. Lessons learned in the past in **business intelligence (BI)** in general, and data warehousing specifically, are used today in Power BI development. So, it is worth your time to learn about those experiences from the past.

Power BI uses what we learned about successful data warehousing, which is why Power BI is a lot more than a visualization tool. It comprises a lot of components. This chapter introduces you to all Power BI components. It also outlines Power BI licensing.

Structure

This chapter covers the following topics:

- Introduction to business intelligence
- Historical overview
- Introducing Power BI

- Power BI use case
- Power BI Desktop and the Power BI service
- Power BI licensing

Objectives

By the end of this chapter, readers will know what BI is. You will also know and understand typical data warehousing terms like **online analytical processing (OLAP)**, **extract, transform, and load (ETL)**, and cubes. You will understand data quality issues and the need to prepare data before use.

Besides more theoretical terms, you will learn about Power BI Desktop and the Power BI service and understand the difference between them. You will also understand Power BI licensing.

Introduction to business intelligence

Data has become increasingly important to businesses and is generated with everything you do. Using all the generated data to gain insight is paramount to every business. The Japanese scientist, *Shinya Yamanaka*, once stated that the main business model for companies in the future would revolve around data; by that, he meant that a bakery would not earn money by selling bread but by collecting, using, and potentially selling data. More recently, it is people like *Mark Zuckerberg* and *Jeff Bezos* who declare similar things. *Facebook (Meta)* did not become rich from the license money people pay to use Facebook. It is free. That is to say, Facebook is gathering data, your data, and making money with that data.

Of course, most companies are not Facebook or Amazon, but small local companies also gather data. Each company and individual works with data. This may be customer-related data, patient data when in health care, product or service-related data, sales data, measurements, or event data from **Internet of Things (IoT)** sensors, and so on. The list is endless. Gathering this data and then not using it would be plain silly. We can analyze the data, learn from the data, and turn that knowledge into good use. Turning data into information, insight, and action is what we call BI, or to put it in other words:

BI is providing the right people with the right information in the right format at the right time in order to make the right decision.

This definition is straightforward. The challenges come from the word right that is used five times in the definition. In real life, you need to pay close attention to each occurrence of the word right. Let us analyze this definition.

The most important word here is decision. It is all about decision-making. Learning something and doing nothing with it can be ok in fundamental science (and it sometimes seems to be ok in politics as well), but generally speaking, we want to do something with the knowledge we gained. Gaining a competitive edge is what is said oftentimes about BI, but that means that

you change things because you have new or more insights into processes, customer behavior, the market in general, etc. and by applying changes, you hope to, for instance, increase sales, reduces costs, increase productivity, decrease sick leave, etc.

The definition speaks about the right decision. However, the data analyst may not be responsible for decision-making; however, a manager is, but if the data analyst provides the manager with data to support the decision-making, the data analyst is responsible for the correct interpretation of the data. Proper visualizations are a part of making sure the audience understands the data. Good visualizations can turn data into information. So, we make the right visualization, so decision-makers have the right information to make the right decisions.

If we go back to our definition of BI, we get to the part where it says the right information. We need proper data for that, and very often (more often than not), we do not get proper data. Consider, for instance, a list of patients in a hospital with around 30% of duplicate rows in that table. The first step in being a data analyst is to analyze the raw data. This is where you first see that you might have a problem with duplicates (we will discuss more potential issues later in this chapter). This step is called **exploratory data analysis (EDA)**.

EDA is your first look at raw data. You spot potential issues like duplicate rows. You do a first check on what values are in your data, like calculating statistics like average, min, and max values of different columns. It is where you get to know your data.

The second step, after having done the initial EDA, is solving any issues you found, like in this case, de-duplicating the dataset. The third step is then to perform the actual data analysis to see what we can learn from the data. In our example, with patient data, we want to learn about our patients from this dataset. The main point here is that raw data is seldom good enough to use directly. It needs preparation before we can use it and make it into the right information.

We have now explained two occurrences of the word right in our definition. The other occurrences are slightly more straightforward. The person who has to make a decision is the right person to receive the information, and they need it at the time the decisions need to be made (or preferably before that). Delivering it in time, in some cases, proves to be a challenge, but the need for timely information is clear.

The last right of the definition is about the right format. Data can be visualized in a lot of different ways. Different formatting, like showing data in a 3D pie chart vs. showing the same data in a bar chart, might lead to completely different interpretations of the data, which can be something that you specifically were looking for, or it might lead to unintentional misinterpretation of the data. When visualizing data, you need to be aware of what message you want to tell what target audience. It will further be discussed in *Chapter 4, The Basics of Visualizations*.

To recap, we live in a world of data. Data plays an important role in our everyday lives. Being able to use all that data to your advantage is an important skill nowadays. Doing so will not always be easy, but at the end of this book, you will have a good head start to become a proficient data analyst using Power BI. First, let us have a closer look at the data within organizations.

Data in organizations

Data-driven working seems to be the holy grail for a lot of companies, which leads to situations where management at some point decides they want the company to work data-driven. More often than not, it starts a data warehouse project or, more recently, a data lake implementation within the IT department. On other occasions, BI software like Power BI is purchased so people can create reports and analyze data. However, data-driven work is a lot more than building a data warehouse or purchasing visualization software. People within the organization need to change how they perform their daily tasks, and the organization as a whole needs to be organized and managed differently.

Data-driven work is more a management style, a way to organize a company, than anything else. To be successful, we have to realize that. Maybe change management is more important than starting an IT project. Data-driven work is anything but an IT project or, even more generally, a project. It is a way of working.

Before you start any BI initiative, it can be useful to perform a maturity scan of the organization. Maturity models have been developed to get an idea on where a company stands in regard to using data.

Figure 1.1 shows a simplified maturity model. All the way to the left, you will see maturity Level 1: Unaware. Even in companies that should be classified as being in this stage, people work with data, and Power BI can be a very useful tool for those people. There are probably no formal requirements for these reports.

As can be seen in the following figure, there is no formal decision-making process. Maybe data plays a role in decision-making, maybe not. The data we use and how it is prepared is not formally described:

Level 1: Unaware	Level 2: Opportunistic	Level 3: Standardize	Level 4: Company wide	Level 5: Transformative
No formal decision making Process Ad-hoc reports	Limited number of users Spreadmarts Shadow systems Data inconsistencies	Executive sponsor Central data marts Central direction, both technical and in business definitions	Business Intelligence Competence Center Corporate datawarehouse	Agile Innovative

Figure 1.1: Maturity model

When a company states the ambition to become data-driven, it should first determine its current situation in regard to the maturity model. Then, they should decide on what their ambition is. With the current situation and the ambition stated clearly, a path towards that ambition should be determined. You cannot be on level 1 today and expect to be on level 5 tomorrow. That will take time, and more importantly, you will have to go through all the stages to successfully get to the desired level.

This book is not about how organizations can become data-driven and what they have to do to reach their ambitions, but it is important to have some understanding of the different maturity levels because Power BI can play an important role in all levels. Power BI may be used differently, though. Before we take a look at different Power BI use cases, let us first look at how BI has evolved over the years. Power BI takes advantage of lessons learned in the past, and we as Power BI developers can take advantage of those lessons learned in all the different scenarios Power BI supports.

Historical overview

In the 1980s, relational databases were introduced. If you take a look at the marketing from that time, you will read texts like information at your fingertips or be able to always make informed decisions. We came from the mainframe era, where there was limited data and information available, and it was hard to get, but the promise of the relational database was that everyone could send queries to the database and get answers immediately. Systems like **customer relationship management (CRM)** were built, and it was said we could use all the data generated by those applications in real-time. The ambitions back then were no different than our current ones.

In the real world, it turned out not to be as easy as marketing promised. A couple of problems were encountered when trying to do more complex reporting or analysis on these operational systems. Microsoft nowadays makes the same promise, that you can connect Power BI to basically any operational system that generates data, start analyzing the data and use it for reporting straight away. This approach did not work previously, and we now have to make sure we do not make the same mistakes today with Power BI. Let us see what we can learn from the past.

The most important problems that we encountered were as follows:

- Data complexity
- Impact on source systems
- Data quality
- Lack of a consistent 360-degree view
- Lack of historical data

Data complexity

Databases that support operational applications like **human resources (HR)** applications, CRM applications, **enterprise resource planning (ERP)**, **electronic patient dossiers (EPD)**, to name just a few, are most of the time databases that use a normalized data model. An example of how that may look can be seen in *Figure 1.2*.

The goal of *Figure 1.2* is not for you to understand this data structure at this point. It is just an illustration to show that normalized database structures can become really complex.

Figure 1.2 is a subset of a simple example database:

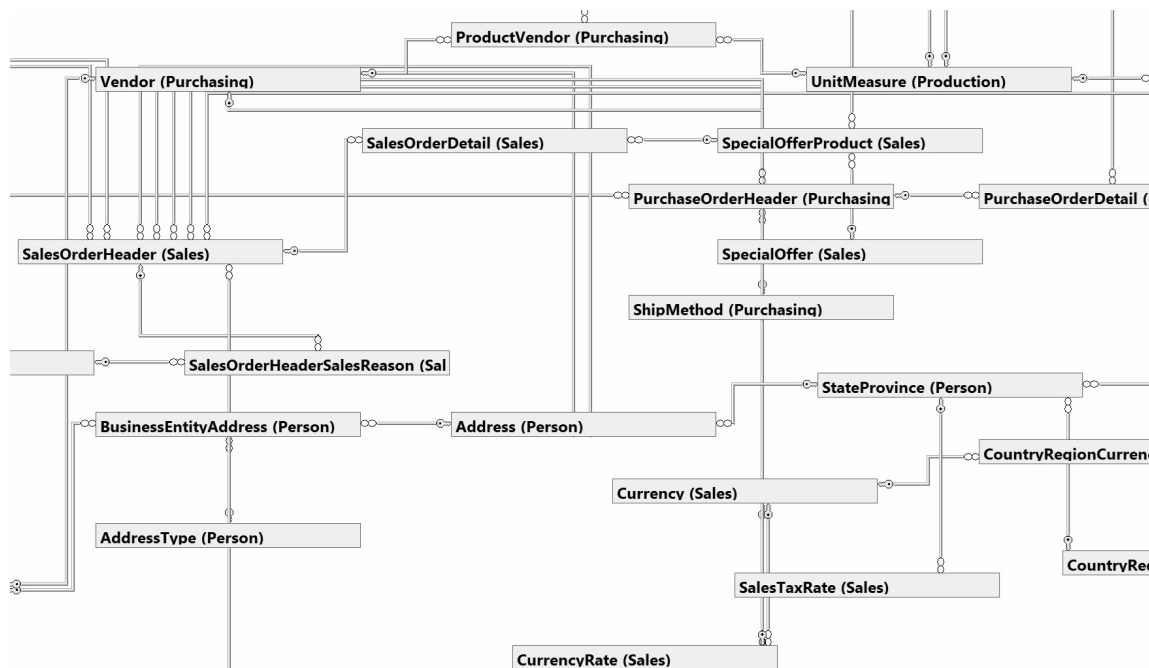


Figure 1.2: Normalized data model

The database, as shown in *Figure 1.2*, is an example of the adventure works database, which is a fairly simple example database that Microsoft uses a lot. Real world database can become a lot more complex, and some databases have over 100.000 tables. Complexity by itself is not necessarily bad. It however, leads to two negative effects as follows:

- Error-prone
- Productivity

The more the complexity increases, the bigger the chances of errors. Queries to get the data out of the database correctly become complex as well. People make more errors when writing complex code as opposed to simple code. You can import a table schema, as shown,

into Power BI and create reports based on these tables. Within Power BI, you will at some point need to write **Data Analysis Expressions (DAX)**. These DAX expressions can become incomprehensibly complex when using normalized databases, as shown in *Figure 1.2*. You want to stay away from that, especially when accountability becomes important. You will learn about dimensional modeling in *Chapter 2, Dimensional Modeling* where you create easy-to-use data models.

Another issue might be the productivity of report builders. When the complexity of the data, and especially the complexity of the data model, increases, so does the time it takes to develop good, effective reports. With HR being a big part of operational costs for a lot of companies, having people work more efficiently is important. Ideally, we want to enable data analysts and report builders alike to focus on content and not on complexity. By simplifying the data model, we simplify data analysis and report creation, making better use of the expensive resource that is human capital.

Impact on source systems

Data analysis or extensive use of reports puts an extra strain on the database being used. This may impact the database and, with that, the application that the database is supporting negatively. The application can very well be one of the primary processes of the company. Although effectively working with data is important, the primary processes are even more important.

With the compute power available today and the optimized database engines we have, this argument is not nearly as important as it was thirty years ago. We can develop Power BI reports in such a way that we minimize the impact the report has on the source it uses, but even so, a lot of database administrators do not want us to query their database directly.

Data quality

Companies nowadays have a lot of data, but more often than not, that data has issues that prevent companies from using the data as-is. The quality of the data is bad. There are always issues with data in operational sources. In our definition of BI at the start of this chapter, we used the word *right* a lot. The definition of BI states that we need the right information, and we need proper data to get the right information. More likely than not, you do not get proper data. Issues that you may encounter may be as follows:

- Duplicate rows
- Missing data
- Incorrect data
- Inconsistent data
- Homonyms and synonyms
- Other issues

Let us start at the top of the list. Consider a **Patient** table in a hospital. It would not be surprising that the table contains up to (or even more) than 30% of duplicate patients. *General Data Protection Regulation (GDPR)* may forbid you to use anything like a citizen service number, so you cannot be sure. However, looking at first name, last name, birth date, and address may give you the impression that many duplicates exist in your dataset. With all those duplicates, a lot of patient analysis that can potentially be done would be meaningless. You will need to deduplicate the data before starting any analysis. Doing so on the original table is probably out of the question. You need to extract the data first, prepare the copy, and do the analysis on the copy.

Missing data is a very common issue as well. Consider a web shop where you have to create an account before you can order anything. Some of the information you have to enter is obligatory, like your last name and email address, but the date of birth might, for instance, not be mandatory. A lot of people will leave it empty. That leaves you with a lot of missing values in the dataset.

Continuing on the date of birth example, suppose the date of birth was mandatory. Some people will enter their real birth date. Others will just type in a fake birth date. January 1st of 2000 is an often-seen birth date in customers' data of online shops. Other values, like January 1st of 1970, may show up more than statistically expected. It is the data represented by zero in Unix-based systems. Instead of missing data, we are now faced with incorrect data. There are a lot of examples of incorrect data. You have to figure out what this means for the data you are going to work with, especially when you pretend to present the right information.

Data can be correct but inconsistent at the same time. Suppose, for instance, you have a customer who lives in the city of *Seattle, Washington*. Most people working with Microsoft are familiar with this city. Now, suppose that the same customer lives in *Australia*. The value of Australia in itself can be correct, just as Seattle can be. However, the combination of both is not correct. They are inconsistent with each other. This does have an impact when analyzing where your customers are located.

As a last example, let us look at homonyms. Suppose we again analyze customer data, especially in which city our customers live. For this example, we focus on our *Belgium* customers solely. In Belgium, part of the people speak French, and the other part speaks Dutch. The first group calls their capital *Bruxelles*, the Dutch-speaking people call it *Brussel*, and English-speaking people call it *Brussels*. Using the data as-is, with different values for the same city, will not lead to correct results when used in drag-and-drop reporting like in Power BI. In terms of data, you have three entirely different cities. In the real world, it is of course one and the same city. You may also have other variations in the database where people made typos when entering the city name. Again, if you do not prepare your data before doing any sort of analysis on the data, you run the risk of getting incorrect results. Even a simple analysis of counts per city may be more complex than it seems at first glance.

Lack of a consistent 360-degree view

Companies never have just a single database or a single source of data. They have multiple sources. Those sources are partly complementary and partly overlapping. Companies can, for instance, connect to an order entry system and, by combining all orders of a single month, calculate their monthly sales, or they connect to their invoicing or bank system and derive monthly sales from there. More often than not, you will get two different answers.

Your first step when creating a Power BI report is to choose the data to use. You will have to decide which source to use for the report you are going to make. You may have to analyze first whether or not you get the same answers from both systems. When the numbers are different, the choice becomes more difficult but also a lot more important. These are difficult decisions to make. Sometimes there may be obvious reasons for the differences. The order entry system may have used the order date to calculate sales, where the other system used an invoice date. In this case, you simply choose the definition of sales you need for the report.

Another example might be when you need to combine information about sales people from the CRM system with identities coming from Microsoft Entra. There should be a match in Entra for every sales person, but in real life, it could be difficult to match the entries from both systems. Missing data, typos in email addresses, old email addresses from before Entra was introduced, or from before the merger with another company can all be reasons for this.

Lack of historical data

The last potential issue when working with data from operational systems is the potential lack of historical data. One issue may be that, for instance, six months ago, the company migrated to a new CRM system. Comparing today's business with last year is all of a sudden complex because you need to combine the data of the old system with the data of the new system, and even though it is the data about the same process, there is likely not a one to one match between how the data is stored.

More importantly, it is the lack of historical data of what we call the dimensions. Suppose you want to analyze what products are popular in what parts of a country. You do have all historical sales transactions available, so you know exactly who bought what, but also suppose that every time a customer moves to a new address, the old address is overwritten with the new address. This is what happens in a lot of operational applications.

In this scenario, you will match a historical sales transaction to a current address. So, products sold to a student living on a university campus may now very well be linked to an urban area where a lot of families with young children live. The sold products get linked to the wrong part of town.

As with issues described earlier, a fairly straightforward analysis becomes difficult or even impossible to do.

When marketing promised us all those years ago how easy it was to connect to the database and get information out of it to support decision-making, they overlooked these issues.

Today, with Power BI, the promise is the same as it was back then. It is easy to connect to almost any source. Once connected, you start building reports immediately using simple drag-and-drop, but all the issues described here are applicable to today's data sources. You have to be aware of the problems your data poses to you, and you have to find and implement fixes before you can do meaningful analysis on your data.

Instead of focusing on issues, let us focus on possible solutions. In the 90s, data warehousing became popular to overcome the issues mentioned in the section. So, let us have a look at data warehousing.

Data warehouse

In the 90s, companies started creating data warehouses to overcome the issues as discussed.

Figure 1.3 shows a simplified data warehouse architecture:

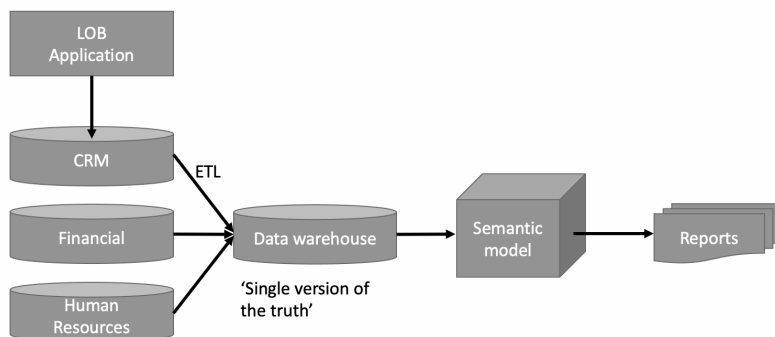


Figure 1.3: Simple data warehouse architecture

In Figure 1.3, there are three different databases supporting **line of business (LOB)** applications. This means there are three data sources available to use for reporting and analysis. With data warehousing, a BI team extracts the relevant data from the source systems to load the data into a new database, that is, the data warehouse. By doing so, it gives them the opportunity to transform the data while doing so. Probably, the most valuable part of data warehousing is the ETL process. Nowadays, people creating data warehouses and implementing ETL processes are called **data engineers**.

The transform part of the ETL process allows data engineers to first and foremost use a different table structure to store the data. The first issue described is the complexity of the normalized operational databases. Data engineers have the opportunity to make it simpler.

There are multiple architectures and data modeling techniques that can be used. This book is not about data warehouse architectures. Although the lessons learned in the past with data

warehousing, and especially the data modeling techniques applied in data warehousing, are relevant. The data modeling technique preferred for Power BI is **dimensional modeling**. The result of a dimensional model database is what is called a **star schema**.

Figure 1.4 is an example of a simple star schema: the star in Figure 1.4 emphasizes the design with a central table surrounded by other tables, together creating the form of a star:



Figure 1.4: Star schema

Refer to Figure 1.2 and compare it to Figure 1.4. A lot of people will agree it looks a lot simpler. It is a Power BI best practice to base all your work on data modeled into a star schema. It will improve your Power BI report dramatically. Easier report development leads to higher quality reports, less performance issues, and with current Fabric licensing, it will be cheaper too.

Take notice of the text 'Single version of the truth' under the data warehouse in Figure 1.3. One of the issues described previously is the lack of consistency between different source systems and the lack of a consistent 360-degree view of an organization because of it. With all the relevant data placed into a single data warehouse, which is now used as the source for reporting and analysis, the lack of consistency is solved. When you always go to the same place with questions, you always get the same answer. The underlying integration problem has been solved in the ETL process. A more modern line than 'Single version of the truth' is 'single version of the facts', referring to the fact that different people may still interpret data differently, even though they get the same data.

A more modern architecture involves the use of data lakes instead of data warehouses. It is just a modern version of old-fashioned data warehousing, more geared towards big data and the cloud. There is not much difference from a Power BI development perspective.

Figure 1.5 shows a data lake architecture:

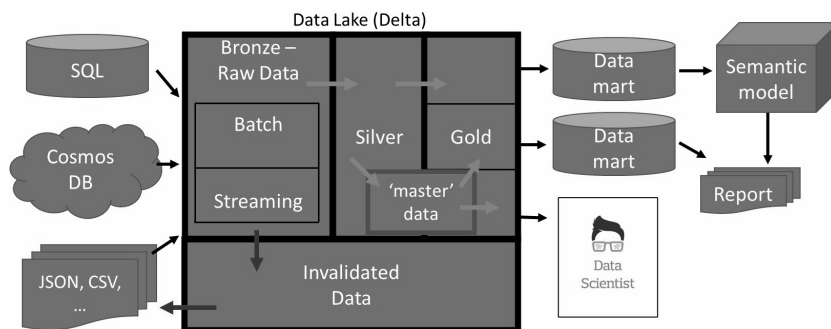


Figure 1.5: data lake

The data mart is in Figure 1.5 to the right of the data lake. The data mart is usually a database modeled into a star schema. It can be the perfect source for Power BI developers.

In creating a star schema, all preceding issues described are as good as possible solved. Data is deduplicated when applicable and necessary, synonyms are resolved, etc. The data has been cleansed.

Notice that in both Figure 1.3 and Figure 1.4, you can see a semantic model. Building data warehouses or data marts solved most issues discussed earlier. There is one thing left, though, we want to enable as many people as possible to create reports, and we want them to be as productive as possible as well. That means that it should be really straightforward and simple to create reports. We try to reach that ambition by making drag-and-drop reporting tools like Power BI. However, there is a gap between how well drag-and-drop works and the functionality provided by databases. In the 90s, we started creating OLAP cubes to bridge the gap between technical-oriented databases and drag-and-drop tools. Today, we call that extra layer a semantic model.

All the preceding is about lessons learned in the past. Applying these lessons learned increases the quality of any BI platform. It is, therefore, essential to take these lessons forward and take advantage of them when working with Power BI.

Introducing Power BI

Up to the zeroes of this century, you would use Microsoft SQL Server when implementing BI solutions. Data warehouse developers would create data warehouses in SQL Server using **SQL Server Integration Services (SSIS)** as an ETL tool. They would then implement an OLAP

cube (semantic model) using **SQL Server Analysis Services (SSAS)**. The reports and data analysis would then be done using either **SQL Server Reporting Services (SSRS)** or Excel. To enable more people to create compelling reports faster, Microsoft introduced some add-ins in Excel. Power Pivot, Power Query, and Power View. A bit later, these three add-ins were combined to create a new product, that is, Power BI.

Today, Power BI still has three main components. Power Query is still called **Power Query**. It is a visual low-code/no-code ETL tool. Power Pivot has evolved into what we now call a Power BI semantic model, and Power View does not exist anymore as a name, but it is what we know Power BI best for: the visualizations that make up our reports.

The three Power BI components, combined with the lessons learned from data warehousing, are shown in *Figure 1.6*:

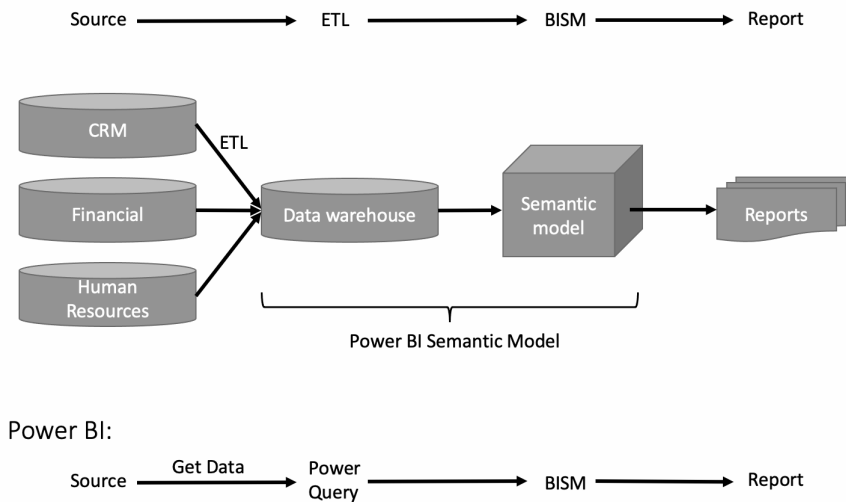


Figure 1.6: Power BI compared to data warehousing

When comparing *Figure 1.6* to *Figure 1.3*, we see that Power BI took the lessons learned from the past and basically allows the Power BI developer to create a data warehouse and semantic model. It uses an ETL tool to read data from various sources and allows transformations before loading the data into its own semantic model. Once the data is modeled, it can be visualized.

Power BI does not have a separate data warehouse and semantic model. However, you will learn that you load data into tables within your Power BI solution (you will learn about this later in the book). You can then set all sorts of settings that are typical for a semantic model, as shown in *Chapter 7, The Basics of Semantic Model*. On top of that, you will enrich the data by adding calculated measures using the semantic model language DAX, as can be read in *Chapter 8, DAX*.