

Preface

SAS is used for various functions in the development and maintenance of data warehouses because of its reputation of being able to handle so-called big data. SAS software has been in existence a long time, and has been implemented in many large, data-intensive environments, including data warehouses.

This book provides end-to-end coverage of the practical programming considerations to make when involving SAS in a data warehouse environment. Complete with step-by-step explanations of essential concepts, practical examples, and self-assessment questions, the book helps you begin to make decisions about the roles SAS should play in your data warehouse. It will teach you how to design arrays and macros to standardize extract-transform-load protocols, as well as how to develop strategies to optimally serve data warehouse customers.

You will learn the pros and cons of storing data in SAS, how to document and design ETL protocols for SAS processes, and how the use of SAS arrays and macros can help improve input/output (I/O) efficiency. You will also examine approaches to serving up data using SAS, and how to connect SAS to other systems to enhance the data warehouse user's experience. By the end of this book, you will have a foundational understanding of the roles SAS can play in a warehouse environment, and be able to choose wisely when designing your data warehousing processes involving SAS.

Who this book is for

This book is aimed at programmers using SAS products who are working on a data lake, data mart, or data warehouse. It is also aimed at managers heading up projects involving using SAS to maintain a data lake, data mart, or data warehouse. To benefit from this book, it is helpful to have a background in working on data projects that require serving data or reports to customers. Also, some experience of working with big datasets will be helpful in understanding this book.

What this book covers

Chapter 1, Using SAS in a Data Mart, Data Lake, or Data Warehouse, explains the origins of SAS, and how data **input/output (I/O)** are managed in SAS. It also provides context for how SAS products are used today, in modern data warehouses.

Chapter 2, Reading Big Data into SAS, covers how to read data in different formats into SAS. It also talks about **SAS data formats**, and packaging data for **import** and **export** in SAS.

Chapter 3, Helpful PROCs for Managing Data, provides an introduction to PROC CONTENTS, PROC SQL, and PROC PRINT, and describes how to deal with SAS **formats** and **labels**. It also provides different strategies for **viewing data** in SAS.

Chapter 4, Managing ETL in SAS, explains how to prepare an analytic environment, including developing **naming conventions**, and SAS format and label policies. It also describes the designation of data storage and **user groups**.

Chapter 5, Managing Data Reporting in SAS, introduces you to the **output delivery system (ODS)**, and explains how the ODS is used for outputting graphics files from SAS. This chapter also covers how to use PROCs that were developed specifically for the ODS, including PROC TABULATE and PROC SGPLOT.

Chapter 6, Standardizing Coding Using SAS Arrays, explains how to do array processing in a SAS data warehouse, how to add **conditions** to arrays, and how to deal with naming conventions in arrays. In SAS, because of I/O limitations, the use of **arrays** is usually necessary in ETL code.

Chapter 7, Designing and Developing ETL Code in SAS, goes over how to plan ETL code, using PROC UNIVARIATE and PROC FREQ to study our data and help us plan how to serve up variables. The second part of the chapter focuses on how to develop optimal ETL code based on our plans.

Chapter 8, Using Macros to Automate ETL in SAS, describes how to convert data step code used in ETL to **SAS macro language** in order to automate the process. It also covers how to store and call macros, and how to use them to load transformed data.

Chapter 9, Debugging and Troubleshooting in SAS, covers **debugging** approaches in SAS. Advice for forming and formatting code is given, and special attention is given to debugging **do loop code** and macros.

Chapter 10, Considering the User Needs of SAS Data Warehouses, describes a method by which to classify users, and then apply **data stewardship policies** that help ensure their needs are met. For analyst users, providing data access, **foreign keys**, and **crosswalk variables** is described. For developer users, providing **data curation** and other support is delineated.

Chapter 11, Connecting the SAS Data Warehouse to Other Systems, talks about serving SAS to other data systems, which is typically done asynchronously. Next, it describes connecting SAS to other data systems, which is typically done synchronously through an **open database connectivity (ODBC)** protocol using **SAS/ACCESS**.

Chapter 12, Using the ODS for Visualization in SAS, describes the differences with using the ODS and visualization in SAS when done in print compared to on the web. Next, ways to serve SAS data to the web using the **SAS Enterprise Guide** aided by **SAS Viya** are explained, and how to visualize SAS data in other programs, such as **R** and **Tableau**, is described.

To get the most out of this book

You will need access to a version of SAS. If you do not have access to a SAS server environment or PC SAS, you can use the free version of SAS, called SAS University Edition (available here: https://www.sas.com/en_us/software/university-edition/download-software.html). SAS University Edition is available for Windows, macOS, and Linux. All code examples have been tested using SAS University Edition in Windows, but they should work on any version of SAS.

Example data curation files in this book were developed using Microsoft Word, Excel, and PowerPoint. These files can be developed in the same or comparable software.

If you are using the digital version of this book, we advise you to type the code yourself or access the code via the GitHub repository (link available in the next section). Doing so will help you avoid any potential errors related to the copying/pasting of code.

You may benefit from following the author on YouTube (<https://www.youtube.com/channel/UCCHcm7rOjf7Ruf2GA2Qnxow>) and LinkedIn (<https://www.linkedin.com/in/dethwench/>), where she posts video tutorials and information about data science.

Download the example code files

You can download the example code files for this book from GitHub at <https://github.com/PacktPublishing/Mastering-SAS-Programming-for-Data-Warehousing>. In case there's an update to the code, it will be updated on the existing GitHub repository.

We also have other code bundles from our rich catalog of books and videos available at <https://github.com/PacktPublishing/>. Check them out!

Download the color images

We also provide a PDF file that has color images of the screenshots/diagrams used in this book. You can download it here: https://static.packt-cdn.com/downloads/9781789532371_ColorImages.pdf.

Conventions used

There are a number of text conventions used throughout this book.

`Code in text`: Indicates code words in text, database table names, folder names, filenames, file extensions, pathnames, dummy URLs, user input, and Twitter handles. Here is an example: "We will map LIBNAME to X, with X being the folder where we put the dataset."

A block of code is set as follows:

```
LIBNAME X "/folders/myfolders/X";  
PROC CONTENTS data=X.Chap5_1;  
run;
```

When we wish to draw your attention to a particular part of a code block, the relevant lines or items are set in bold:

```
ODS TRACE ON / label;  
PROC UNIVARIATE data=X.chap5_1;  
    var _AGE80;  
run;  
ODS TRACE OFF;
```

Bold: Indicates a new term, an important word, or words that you see on screen. For example, words in menus or dialog boxes appear in the text like this. Here is an example: "If you are using SAS University Edition, the **RESULTS** tab will display the graphic."

<p>Tips or important notes Appear like this.</p>

Get in touch

Feedback from our readers is always welcome.

General feedback: If you have questions about any aspect of this book, mention the book title in the subject of your message and email us at customercare@packtpub.com.

Errata: Although we have taken every care to ensure the accuracy of our content, mistakes do happen. If you have found a mistake in this book, we would be grateful if you would report this to us. Please visit www.packtpub.com/support/errata, selecting your book, clicking on the Errata Submission Form link, and entering the 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 copyright@packt.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 authors.packtpub.com.

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 Packt can understand what you think about our products, and our authors can see your feedback on their book. Thank you!

For more information about Packt, please visit packt.com.