



# Chapter 1: An Introduction to Data Warehousing

## Preface

---

Tackling the subject of Data Warehousing is a daunting task. Tackling the writing of a book about the subject is equally intimidating. Most textbooks about Data Warehousing take an easy course and specialize in a tiny portion of the topic. This leaves the reader more educated about one specific part but without the perspective necessary to put all the pieces of the puzzle together. In other words, this approach teaches us how to build each room of a house, with little regard as to whether the rooms will fit together as a cohesive building when finally constructed. This book will take an opposite approach. Although different topics will be broached in different chapters, they will all be tied together by continually referring to 'the big picture:' the entire Data Warehousing process.

Of course, what is referred to as 'the big picture' is not consistent across different organizations, or even within them. This is one of the problems in formulating generic Data Warehousing strategies: what are the common denominators that will make practical sense to all those who read a book

## 2 Strategic Data Warehousing Principles

about it? In the case of this book, it is SAS software, which can optionally be used in every step of the Data Warehousing process. For those of you who have only a passing acquaintance or no knowledge of SAS software, this book will still be of use to you in learning about the Data Warehousing process. It will also give you a bonus: an insight into one of the most robust, powerful, and flexible software products available.

Data Warehousing is not an academic exercise, but is a practical one that should have direct and measurable benefits to any organization that invests in it. This means that any writing about the topic should be grounded in the practical, as opposed to the academic. Sometimes what is logically the best approach is not always the most practical. In real data warehouse processes, there are often resource limitations or political realities that prevent the most logical approach from being used. This should not, however, ever prevent each data warehousing decision that has to be made from being put into a framework (or process). If the decision that has the best chance of leading to a successful implementation is not practical, then this should be documented.

Only through the careful consideration of each step within the data warehousing process can it lead to a responsive enterprise tool that is capable of responding to changes in the enterprise itself or the circumstance in which the enterprise finds itself. This book is intended to describe this framework so that a successful process can be formulated. Although this process will differ greatly in content from enterprise to enterprise, in all situations it will be the foundation upon which a successful Data Warehouse can be built.

### 1.1 Who Will Benefit from Reading this Book?

---

This book is primarily aimed at three different types of readers:

1. Those who are, or will be, involved in the administration and design of the Data Warehousing process (*Data Warehouse facilitators*).
2. Those who will be involved in the implementation of the Data Warehouse (*Data Warehouse developers*).
3. Those who will have a management responsibility or a stake in the outcome of the Data Warehousing process (*Data Warehouse Managers*).

#### Data Warehouse Facilitators

These people are absolutely essential to the successful implementation of a Data Warehousing project. They are the facilitators that turn just another Information Systems project into an enterprise that is responsive and success-

ful. Although it is not necessary for these facilitators to be technical, they very often are, in a practical sense. They perform the role of a conduit between the enterprise goals and the physical implementation of the Warehouse itself.

Data Warehouse Facilitators have to mediate between the often outlandish desires of the business community and the inherent caution of the Information Systems folks. A practical knowledge of both the business and the technology available is therefore essential to successfully perform this role. Not only must the Warehouse Facilitators have the negotiating skills to keep the Warehousing project moving forward but they must be able to see the big picture at all times. This book will give the Warehouse Facilitators the framework within which the enterprise needs can be realized through the Data Warehouse process.

These people usually possess practical knowledge as well as an appreciation for technology. They understand how technology can be used to support and sometimes change business processes. Usually, this group of people has a clear vision of what is needed from the Data Warehouse process.

There is a good argument that suggests that this role can best be divided into two. First, the facilitators make sure that the enterprise needs are fully understood, so that the Data Warehouse can fully meet these needs. Second, part of the role translates these needs into a technological lexicon, so that the developers can understand fully what they are trying to achieve.

These Facilitators will benefit from reading this book by understanding the concepts and how different parts of the SAS System may be used to meet the very real technological challenges that Data Warehousing produces.

#### Data Warehouse Implementers

This book will give developers a firm grounding in how SAS software can be used to meet the goals that are set for them by enterprise needs that can be addressed through the Data Warehousing process. There is no inherent link between technical knowledge and conceptual understanding, so this book will help bridge that potential gap. Developers can use this book not only to gather the 'how to' but also the 'why', which will put them in a position where they can become more involved in the design as opposed to just the building of the implemented physical Data Warehouse.

In this day and age, building a database or a software application based upon a pre-defined specification is an anachronism, due to the phenomenal improvements made in both software and hardware. This is one of the principal reasons that Data Warehousing has gained in popularity: the nature of Data Warehousing, being both iterative and evolutionary, means

## 4 Strategic Data Warehousing Principles

that traditional software design techniques by definition could not work. It takes a new breed of developer, one that is willing to become involved in the entire process, not just the programming, to successfully implement a Data Warehouse. This means that the tools needed to implement a Data Warehouse go far beyond just the coding or database design. This book will address these needs. Alongside the many examples of SAS code and techniques, there will be a constant referral back to the big picture, explaining why and where the code and techniques fit into the entire Data Warehousing process.

Throughout the book, there will be references to other sources that the developer can use to gather more information about SAS. Both the topic of Data Warehousing and the SAS software are immense. This means that the code used within the book is there as an outline, an illustration of how it can be done. This does not mean that there are not alternative or better techniques to achieve the same results.

### Data Warehouse Managers

Data Warehouse managers are people who might not actually take part in the design and building of the Data Warehouse process, but they have some form of responsibility toward its success. It is in their interests to understand the process of building the Warehouse so that they can be more responsive in managing the resources needed to make it successful. They take responsibility for educating the enterprise on the data warehouse process. It often falls into the hands of the Data Warehouse managers to 'sell' the concepts behind the data warehouse process that drive the delivery of the project.

This group of managers will come from both the business community and Information Systems. They all have a stake in the outcome of the Data Warehousing process. They all have some influence upon that outcome, but they might not be responsible for the actual implementation of the project. Understanding the Data Warehousing process will give these managers the information they need to make decisions that will result in a more responsive business tool.

### Summary

To benefit from reading this book, all that is needed is an interest in Data Warehousing. Because the book is written from a Warehousing perspective, each topic is addressed at a high level. Instead of being bombarded with details and expecting the reader to fit them into the Warehousing process, every part of the book is tied back to the 'big picture'. At no point should the

details not be directly linked to the reason that the book is being read. This is primarily a book about Data Warehousing, not about SAS software. It is written, however, from the perspective of a Data Warehouse that is being constructed by using SAS software.

This book is not written with the assumption that you are a highly skilled technician, but that you are interested in the Data Warehousing process and have a desire to know how the different parts of this process fit together from both a conceptual and a technical level.

## 1.2 How to Read this Book

---

One of the major differences between life today, and that of ten, twenty, or fifty years ago is the amount of information we are asked to use in our everyday life. The technology for creating information has increased incredibly, without a corresponding increase in our ability to assimilate that information. This means that we have had to find ways to deal with this situation. At school, when expected to read several books a week, we learn to skim, pulling what we hope will be the key pieces of information to, at best, pass the test with flying colors, or at worst, escape public humiliation. When was the last time any of us read a technical book from cover to cover?

The temptation to ignore those chapters that don't immediately arouse our interest in favor of those that do is irresistible. To find out that understanding the interesting chapters is dependent upon reading the boring ones is always a huge disappointment. Readers of this book need not despair. There is only one chapter after this one that needs to be read: Chapter 2: "Planning the Data Warehousing Process." After that, you can use your random-number generator (with or without replacement) to select the next chapter to read, happy in the knowledge that it will stand alone. This does not mean that there will be no references to other chapters, but that, on a conceptual level, it is not necessary to read those chapters in order to understand the one you are reading.

The examples used within this book are extremely simple. Another frustration of reading a technical book is often the complexity of the examples. This means that the reader spends so much time trying to decipher what the example is actually trying to do, that the points the book is trying to elucidate are lost. The examples used in this book are far simpler than those tasks that you will come across in deploying a Data Warehouse in real life, but the real-life tasks will be a natural extension of the examples used within the book.

## 6 Strategic Data Warehousing Principles

The book does not attempt to tell you how your Data Warehouse should be built (as many other books attempt) because no two Data Warehouses are alike. The concepts required to build a Warehouse are addressed, and then these concepts are available to you to help you design a Warehouse that works for your particular set of circumstances. From a technical, as opposed to a conceptual basis, the book will explain and demonstrate the use of SAS software in the Warehousing process.

### 1.3 Formatting Conventions

---

There are several formatting conventions in the layout that will make this book easier to read:

**Key Concepts/Special Topics:** Whenever special topics are addressed within the text, they will be reiterated within a box. This will help you to find topics that might be of particular interest to you. An example of a key concept follows:



*Key Concepts are those that are considered most important from a Data Warehousing standpoint, as opposed to a coding standpoint.*

**Syntax:** There will be extensive sets of examples of SAS code to show how the Warehousing concepts can be practically applied. This code will often be one step within many, so it might not be complete and executable in itself. In every situation, however, the code will be put into perspective, and it will be explained how it could be used.

To keep in line with other publications, all SAS keywords will appear in uppercase. For example, the following lines of code for printing a SAS data set using the PRINT procedure are written as follows:

```
PROC PRINT DATA=libname.dataset ;  
RUN ;
```

All the SAS keywords appear in uppercase letters. The code will be indented, whenever possible, to make it easier to read and to understand.

**Glossary of Terms:** There will be a glossary of terms included at the end of the book. *Any terms that are in the glossary will be distinguished in text because they will be in italics.*

**Cross-references:** This book is not intended to be an introduction to SAS programming or to replace current documentation. Many of the SAS

software modules will be described in some detail, but references will always be made. For instance, the chapter dealing with Client/Server processes will be addressing SAS/CONNECT in detail and will therefore cross-reference the existing applicable documentation alongside any that is included.

For example, an excellent introduction to using SAS is *The Little SAS Book*, by Lora D. Delwiche and Susan J. Slaughter. If there is a specific reference in this book to a particular item in that book, the reference will be detailed in a footnote at the bottom of the page. For example:

*It is possible that in this situation, your particular client might not handle the tasks you have asked of it. You might well receive a message that you have run out of memory or disk space,<sup>1</sup> in which case, given your current architectural limitations, another strategy will be required.*

---

<sup>1</sup>See: Delwiche, Lora D., and Slaughter, Susan J., *The Little SAS Book: A Primer*, Cary, NC: SAS Institute Inc., 1995. 228pp., pp 186-187.

