One-to-One, One-to-Many, and Many-to-Many Joins Using PROC SQL

Kirk Paul Lafler, Software Intelligence Corporation
Charu Shankar, SAS Institute Inc.

Abstract

A powerful and essential PROC SQL programming technique that all SAS® users should understand, and be comfortable performing, is the process of joining (or combining) two or more tables of data. This paper and presentation describes and illustrates the join process, including what a join is, exploration of one-to-one, one-to-many, and many-to-many data relationships, identifying a primary key (or unique identifier), special preparation requirements for each table being specified in a join, and popular join techniques available to SAS® users. To demonstrate the power of the join process, examples of conventional (symmetrical matching) and unconventional (asymmetrical left, right and full matching) using PROC SQL join programming techniques are illustrated.

Introduction

The SQL procedure is a wonderful tool for querying and subsetting data; restructuring data by constructing case expressions; constructing and using virtual tables known as a view; access information from Dictionary tables; and joining two or more tables to explore data relationships. Occasionally, an application problem comes along where the SQL procedure is either better suited or easier to use than other more conventional DATA and/or PROC step methods. As each situation presents itself, PROC SQL should be examined to see if its use is warranted for the task at hand.

Why Join Anyway?

Relational database systems continue to grow in popularity, the need to access normalized data stored in separate tables becomes increasingly important. A join of two or more tables provides a means of gathering and manipulating data in a single SELECT statement. Joins are specified on a minimum of two tables at a time, where a column from each table is used to connect the two tables. Connecting columns should have ***"like"*** values and the same datatype attributes since the join's success is dependent on these values. By relating matching values in key columns in one table with key columns in two or more tables, information can be retrieved as if the data were stored in one huge file. Consequently, the process of joining data from two or more tables can provide new and exciting insights into data relationships.

What Happens During a Join?

Joins are processed in two distinct phases. The first phase determines the names of the tables referenced in the FROM clause. An internal *virtual* table, known as a Cartesian product, is created resulting in each row in the first table being combined with each row in the second table, and so forth. Due to its size, the Cartesian product is managed by the SAS software. The second phase of every join processes the WHERE clause, when present.

Data Relationships and Cardinality

Tables of data and the relationships of data in a Relational Database Management System (RDBMS) environment represent an important topic among data base and SAS users. When it comes to understanding tables, data and the relationship of data, the term, **Cardinality**, is often used to associate the uniqueness of data values contained in a table's columns. Many working in an RDBMS environment associate a table that contains columns possessing a high concentration of unique values as having high cardinality. Conversely, a table with columns possessing a low concentration of unique values (or many "repeats" of values) is referred to as having low cardinality.

Cardinality is also associated with the relationship between two, or more, tables of data in a RDBMS. Cardinality between tables can be represented as one-to-one, one-to-many, and many-to-many data relationships. Typically, table relationships are identified and processed using a "key" column in each table.

- ✓ **One-to-one** – Each row in one table is linked (or related) to a single row in another table using a "key" column.
- ✓ **One-to-many** – Each row in one table is linked (or related) to one, or more, rows in another table using a "key" column.
- ✓ **Many-to-many** – One, or more, rows in one table is linked (or related) to one, or more, rows in another table using a "key" column.

The SQL Optimizer and Join Algorithms

The RDBMS environment and SQL processing is constantly being monitored to maximize efficient processing. The monitoring process is performed by the SQL Optimizer. Classified as a cost-based optimizer, the SQL Optimizer analyzes the available query execution plans available to it, determines the costs associated with each execution plan, and selects the execution plan with the lowest cost. To help achieve optimal processing efficiencies , the SQL optimizer has four algorithms to choose from when performing a join:

- ✓ **Nested Loop** – A nested loop join algorithm may be selected by the SQL optimizer when processing small tables of data where one table is considerably smaller than the other table, the join condition does not contain an equality condition, first row matching is optimized, or using a sort-merge or hash join has been eliminated from consideration.
- ✓ **Sort-Merge** – A sort-merge join algorithm may be selected by the SQL optimizer when the tables to be joined are small to medium size and a nested loop, index and hash join algorithm have been eliminated from consideration.
- ✓ **Index** – An index join algorithm may be selected by the SQL optimizer when indexes created on the columns participating in the join relationship are inclined to improve performance.
- ✓ **Hash** – A hash join algorithm may be selected by the SQL optimizer when sufficient memory is available to the system, and the BUFFERSIZE option is large enough to store at least the smaller of the tables into memory.

Join and the Cartesian Product

When a WHERE clause is omitted from a join query, all possible combinations of rows in each table is produced. This form of join is known as a ***Cartesian Product***. Say, for example, you join two tables where one table has 1,000 rows and the other table has 20 rows without a WHERE-clause – the resulting Cartesian product would have 20,000 rows. The real importance of a Cartesian product join is to serve as a base (or internal representation) for all joins, as illustrated below.

Authors
AuthorID
AuthorName

Books
BookID
BookTitle
AuthorID

The next example illustrates a Cartesian product join using a SELECT query without a WHERE clause. Note: The AuthorID column is displayed twice – once from the Authors table and the second time from the Books table.

PROC SQL Join Query

```
PROC SQL ;
 SELECT *
  FROM mydata.Authors,
       mydata.Books(keep=BookID BookTitle AuthorID) ;
QUIT ;
```

PROC SQL Join Query Results

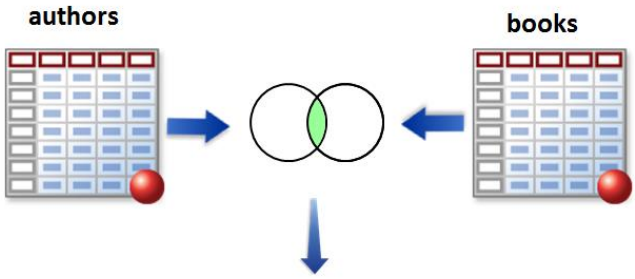
AuthorID	AuthorName	AuthorBio	BookID	BookTitle	AuthorID
A001	Tricia Aanderud		A00101	Building Business Intelligence Using SAS: Content Development Examples	A001
A002	Robert Allison		A00101	Building Business Intelligence Using SAS: Content Development Examples	A001
B001	William Benjamin		A00101	Building Business Intelligence Using SAS: Content Development Examples	A001
B002	Jonas V. Bilenas		A00101	Building Business Intelligence Using SAS: Content Development Examples	A001
B003	Michele M. Burlew		A00101	Building Business Intelligence Using SAS: Content Development Examples	A001
C001	Art Carpenter		A00101	Building Business Intelligence Using SAS: Content Development Examples	A001
C002	Goutam Chakraborty		A00101	Building Business Intelligence Using SAS: Content Development Examples	A001
C003	Ron Cody		A00101	Building Business Intelligence Using SAS: Content Development Examples	A001
D001	Lora D. Delwiche		A00101	Building Business Intelligence Using SAS: Content Development Examples	A001
D002	Barry de Ville		A00101	Building Business Intelligence Using SAS: Content Development Examples	A001
D003	Craig Dickstein		A00101	Building Business Intelligence Using SAS: Content Development Examples	A001
D004	Paul Dorfman		A00101	Building Business Intelligence Using SAS: Content Development Examples	A001
E001	Peter Eberhardt		A00101	Building Business Intelligence Using SAS: Content Development Examples	A001
E002	Jane Eslinger		A00101	Building Business Intelligence Using SAS: Content Development Examples	A001
F001	Lisa Fine		A00101	Building Business Intelligence Using SAS: Content Development Examples	A001
G001	Sunil K. Gupta		A00101	Building Business Intelligence Using SAS: Content Development Examples	A001
H001	Angela Hall		A00101	Building Business Intelligence Using SAS: Content Development Examples	A001
H002	Lauren Haworth		A00101	Building Business Intelligence Using SAS: Content Development Examples	A001
H003	Dan Heath		A00101	Building Business Intelligence Using SAS: Content Development Examples	A001
H004	Chris Hemedinger		A00101	Building Business Intelligence Using SAS: Content Development Examples	A001
H005	Don Henderson		A00101	Building Business Intelligence Using SAS: Content Development Examples	A001
H006	Philip Holland		A00101	Building Business Intelligence Using SAS: Content Development Examples	A001
J001	Mark Jordan		A00101	Building Business Intelligence Using SAS: Content Development Examples	A001
K001	Warren F. Kuhfeld		A00101	Building Business Intelligence Using SAS: Content Development Examples	A001
L001	Kirk Paul Lafler		A00101	Building Business Intelligence Using SAS: Content Development Examples	A001

. . . . . .

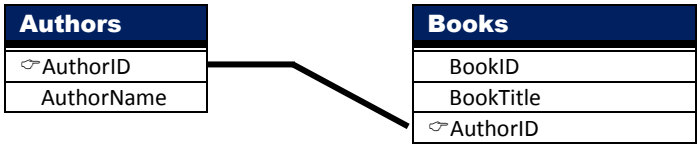
L001	Kirk Paul Lafler		L00201	Output Delivery System: The Basics and Beyond	Z002
L002	Lauren Haworth Lake		L00201	Output Delivery System: The Basics and Beyond	Z002
L003	Ann Lehman		L00201	Output Delivery System: The Basics and Beyond	Z002
M001	Phil Mason		L00201	Output Delivery System: The Basics and Beyond	Z002
M002	Sanjay Matange		L00201	Output Delivery System: The Basics and Beyond	Z002
M003	Thomas Miron		L00201	Output Delivery System: The Basics and Beyond	Z002
M004	Derek P. Morgan		L00201	Output Delivery System: The Basics and Beyond	Z002
O001	Rebecca A. Ottesen		L00201	Output Delivery System: The Basics and Beyond	Z002
P001	Olivia Parr-Rud		L00201	Output Delivery System: The Basics and Beyond	Z002
P002	Frederick Pratter		L00201	Output Delivery System: The Basics and Beyond	Z002
R001	Michael Raithel		L00201	Output Delivery System: The Basics and Beyond	Z002
S001	John Sall		L00201	Output Delivery System: The Basics and Beyond	Z002
S002	Armistead Sapp		L00201	Output Delivery System: The Basics and Beyond	Z002
S003	Marge Scerbo		L00201	Output Delivery System: The Basics and Beyond	Z002
S004	Sandra D. Schlotzhauer		L00201	Output Delivery System: The Basics and Beyond	Z002
S005	Brian C. Shilling		L00201	Output Delivery System: The Basics and Beyond	Z002
S006	Charles E. Shipp		L00201	Output Delivery System: The Basics and Beyond	Z002
S007	Jack Shostak		L00201	Output Delivery System: The Basics and Beyond	Z002
S008	Susan Slaughter		L00201	Output Delivery System: The Basics and Beyond	Z002
T001	Michael Tuchman		L00201	Output Delivery System: The Basics and Beyond	Z002
V001	Robert Virgile		L00201	Output Delivery System: The Basics and Beyond	Z002
W001	Perry Watts		L00201	Output Delivery System: The Basics and Beyond	Z002
W002	Rick Wicklin		L00201	Output Delivery System: The Basics and Beyond	Z002
Z001	Mike Zdeb		L00201	Output Delivery System: The Basics and Beyond	Z002
Z002	Cynthia Zender		L00201	Output Delivery System: The Basics and Beyond	Z002

Conventional Join with Two Tables and a Where Clause

A conventional join of two or more tables, uses a WHERE- or HAVING-clause to produce a result set of "matched" rows. Joining two tables together is a relatively easy process in SQL, as is illustrated in the following diagram. As can be seen, the Authors and Books table are joined together using the "key" AuthorID.



AuthorId	AuthorName	BookId	BookTitle	AuthorID
S008	Susan Slaughter	S00801	The Little SAS Book: A Primer, Fifth Edition	S008



The next example illustrates a join of the contents of the tables, Authors and Books, using the "key", AuthorID in a WHERE-clause. When the value of AuthorID is equal in both tables, the rows are combined together. Note: The AuthorID column is displayed twice in the results – once from the Authors table and a second time from the Books table.

PROC SQL Join Query

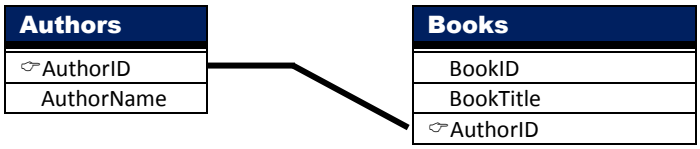
```
PROC SQL ;
 SELECT *
  FROM mydata.Authors(drop=AuthorBio),
       mydata.Books(keep=BookID BookTitle AuthorID)
   WHERE Authors.AuthorID = Books.AuthorID ;
QUIT ;
```

PROC SQL Join Query Results

AuthorID	AuthorName	BookID	BookTitle	AuthorID
A001	Tricia Aanderud	A00101	Building Business Intelligence Using SAS: Content Development Examples	A001
A001	Tricia Aanderud	A00102	An Introduction to SAS Visual Analytics: How to Explore Numbers, Design Reports, and Gain Insight Into Your Data	A001
B003	Michele M. Burlew	B00304	SAS® Macro Programming Made Easy, Third Edition	B003
B003	Michele M. Burlew	B00303	SAS® Hash Object Programming Made Easy	B003
B003	Michele M. Burlew	B00302	Combining and Modifying SAS® Data Sets: Examples, Second Edition	B003
B003	Michele M. Burlew	L00201	Output Delivery System: The Basics and Beyond	B003
C001	Art Carpenter	C00101	Carpenter's Complete Guide to the SAS® REPORT Procedure	C001
C001	Art Carpenter	C00102	Carpenter's Guide to Innovative SAS® Techniques	C001
C001	Art Carpenter	C00103	Carpenter's Complete Guide to the SAS® Macro Language, Third Edition	C001
C003	Ron Cody	C00309	Cody's Data Cleaning Techniques Using SAS®, Third Edition	C003
C003	Ron Cody	C00308	Biostatistics by Example Using SAS® Studio	C003
C003	Ron Cody	C00307	An Introduction to SAS® University Edition	C003
C003	Ron Cody	C00306	Test Scoring and Analysis Using SAS®	C003
C003	Ron Cody	C00305	Cody's Collection of Popular SAS® Programming Tasks and How to Tackle Them	C003
C003	Ron Cody	C00304	SAS® Statistics by Example	C003
C003	Ron Cody	C00303	SAS® Functions by Example, Second Edition	C003
C003	Ron Cody	C00302	Learning SAS® by Example: A Programmer's Guide	C003
C003	Ron Cody	C00301	Longitudinal Data and SAS®: A Programmer's Guide	C003
D001	Lora D. Delwiche	S00803	The Little SAS® Enterprise Guide® Book	D001
S008	Susan Slaughter	S00803	The Little SAS® Enterprise Guide® Book	S008
D001	Lora D. Delwiche	S00801	The Little SAS® Book: A Primer, Fifth Edition	D001
S008	Susan Slaughter	S00801	The Little SAS® Book: A Primer, Fifth Edition	S008
D001	Lora D. Delwiche	O00102	Exercises and Projects for The Little SAS® Book, Fifth Edition	D001
S008	Susan Slaughter	000102	Exercises and Projects for The Little SAS® Book, Fifth Edition	S008
O001	Rebecca A. Ottesen	O00102	Exercises and Projects for The Little SAS® Book, Fifth Edition	O001
D002	Barry de Ville	D00201	Decision Trees for Analytics Using SAS® Enterprise Miner™	D002
L001	Kirk Paul Lafler	L00101	PROC SQL: Beyond the Basics Using SAS®, Second Edition	L001
L002	Lauren Haworth Lake	L00201	Output Delivery System: The Basics and Beyond	L002
Z002	Cynthia Zender	L00201	Output Delivery System: The Basics and Beyond	Z002

Join and Table Aliases

Table aliases provide a "short-cut" way to reference two or more tables in a join operation. Table aliases are specified in the FROM-clause so columns can be referenced in a minimal number of keystrokes in the SELECT and WHERE-clause. To illustrate how table aliases in a join works, a two-table join is linked together as shown in the following diagram.



The following SQL code illustrates a join on two tables with TITLE specified as the connecting column. The table aliases are specified in the SELECT statement as qualified names, the FROM clause, and the WHERE clause.

PROC SQL Join Query

```
PROC SQL ;
 SELECT A.AuthorID, AuthorName, BookID, BookTitle
  FROM mydata.Authors(drop=AuthorBio) AS A,
       mydata.Books(keep=BookID BookTitle AuthorID) AS B
   WHERE A.AuthorID = B.AuthorID ;
QUIT ;
```

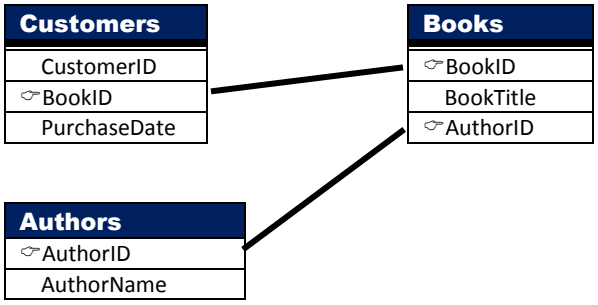
PROC SQL Join Query Results

AuthorID	AuthorName	BookID	BookTitle
A001	Tricia Aanderud	A00101	Building Business Intelligence Using SAS: Content Development Examples
A001	Tricia Aanderud	A00102	An Introduction to SAS Visual Analytics: How to Explore Numbers, Design Reports, and Gain Insight into Your Data
B003	Michele M. Burlew	B00304	SAS® Macro Programming Made Easy, Third Edition
B003	Michele M. Burlew	B00303	SAS® Hash Object Programming Made Easy
B003	Michele M. Burlew	B00302	Combining and Modifying SAS® Data Sets: Examples, Second Edition
B003	Michele M. Burlew	L00201	Output Delivery System: The Basics and Beyond
C001	Art Carpenter	C00101	Carpenter's Complete Guide to the SAS® REPORT Procedure
C001	Art Carpenter	C00102	Carpenter's Guide to Innovative SAS® Techniques
C001	Art Carpenter	C00103	Carpenter's Complete Guide to the SAS® Macro Language, Third Edition
C003	Ron Cody	C00309	Cody's Data Cleaning Techniques Using SAS®, Third Edition
C003	Ron Cody	C00308	Biostatistics by Example Using SAS® Studio
C003	Ron Cody	C00307	An Introduction to SAS® University Edition
C003	Ron Cody	C00306	Test Scoring and Analysis Using SAS®
C003	Ron Cody	C00305	Cody's Collection of Popular SAS® Programming Tasks and How to Tackle Them
C003	Ron Cody	C00304	SAS® Statistics by Example
C003	Ron Cody	C00303	SAS® Functions by Example, Second Edition
C003	Ron Cody	C00302	Learning SAS® by Example: A Programmer's Guide
C003	Ron Cody	C00301	Longitudinal Data and SAS®: A Programmer's Guide
D001	Lora D. Delwiche	S00803	The Little SAS® Enterprise Guide® Book
S008	Susan Slaughter	S00803	The Little SAS® Enterprise Guide® Book
D001	Lora D. Delwiche	S00801	The Little SAS® Book: A Primer, Fifth Edition
S008	Susan Slaughter	S00801	The Little SAS® Book: A Primer, Fifth Edition
D001	Lora D. Delwiche	O00102	Exercises and Projects for The Little SAS® Book, Fifth Edition
S008	Susan Slaughter	000102	Exercises and Projects for The Little SAS® Book, Fifth Edition
O001	Rebecca A. Ottesen	O00102	Exercises and Projects for The Little SAS® Book, Fifth Edition
D002	Barry de Ville	D00201	Decision Trees for Analytics Using SAS® Enterprise Miner™
L001	Kirk Paul Lafler	L00101	PROC SQL: Beyond the Basics Using SAS®, Second Edition
L002	Lauren Haworth Lake	L00201	Output Delivery System: The Basics and Beyond
Z002	Cynthia Zender	L00201	Output Delivery System: The Basics and Beyond

Join with Three Tables

In an earlier example, a two-table join was shown using the Authors and the Books tables. To illustrate what book was purchased as well as the date a book was purchased, a three table join will be constructed. To accomplish this, three different tables will be accessed: Authors, Books, and Customers.

A join with three tables adheres to the same rules as a two-table join. Each table needs to be listed in the FROM clause with the appropriate subsetting (or matching) restrictions specified in a WHERE clause. The diagram, below, illustrates the table relationships along with the keys used in a three table join.



The next example references a three table join with BookID acting as the connecting column for the Customers and Books tables, and AuthorID acting as the connecting column for the Authors and Books tables.

PROC SQL Join Query

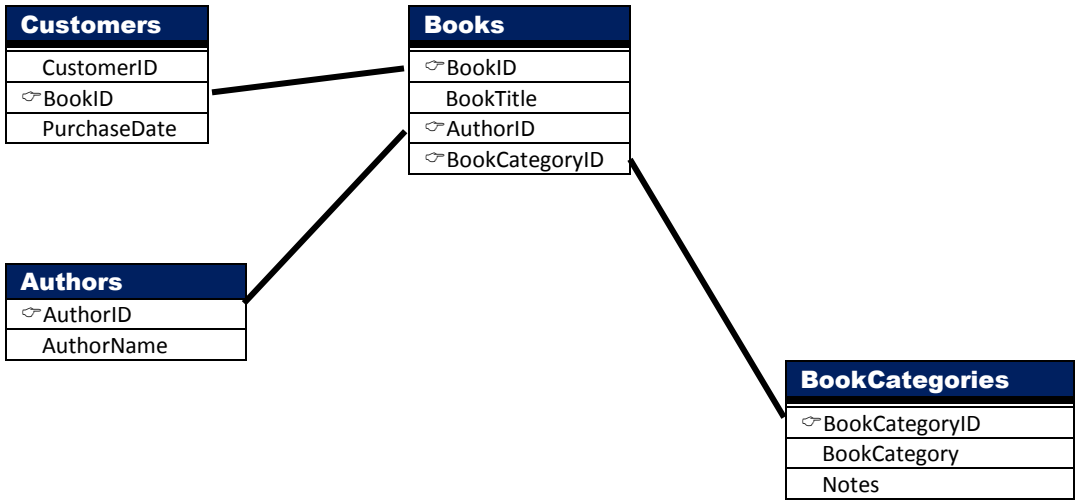
```
PROC SQL ;
 SELECT A.AuthorID, AuthorName, B.BookID, BookTitle, PurchaseDate
  FROM mydata.Authors(drop=AuthorBio) AS A,
       mydata.Books(keep=BookID BookTitle AuthorID) AS B,
       mydata.Customers AS C
    WHERE A.AuthorID = B.AuthorID
      AND B.BookID = C.BookID ;
QUIT ;
```

PROC SQL Join Query Results

AuthorID	AuthorName	BookID	BookTitle	PurchaseDate
A001	Tricia Aanderud	A00102	An Introduction to SAS Visual Analytics: How to Explore Numbers, Design Reports, and Gain Insight into Your Data	05/11/2017
B003	Michele M. Burlew	B00304	SAS® Macro Programming Made Easy, Third Edition	02/28/2015
B003	Michele M. Burlew	L00201	Output Delivery System: The Basics and Beyond	11/03/2015
C001	Art Carpenter	C00103	Carpenter's Complete Guide to the SAS® Macro Language, Third Edition	02/28/2015
C003	Ron Cody	C00309	Cody's Data Cleaning Techniques Using SAS®, Third Edition	04/18/2017
C003	Ron Cody	C00309	Cody's Data Cleaning Techniques Using SAS®, Third Edition	03/27/2016
C003	Ron Cody	C00309	Cody's Data Cleaning Techniques Using SAS®, Third Edition	11/03/2015
C003	Ron Cody	C00308	Biostatistics by Example Using SAS® Studio	12/07/2016
C003	Ron Cody	C00307	An Introduction to SAS® University Edition	10/21/2015
D001	Lora D. Delwiche	S00601	The Little SAS® Book: A Primer, Fifth Edition	04/18/2017
D001	Lora D. Delwiche	S00601	The Little SAS® Book: A Primer, Fifth Edition	10/03/2016
S008	Susan Slaughter	S00601	The Little SAS® Book: A Primer, Fifth Edition	04/18/2017
S008	Susan Slaughter	S00601	The Little SAS® Book: A Primer, Fifth Edition	10/03/2016
L001	Kirk Paul Lafler	L00101	PROC SQL: Beyond the Basics Using SAS®, Second Edition	07/07/2017
L001	Kirk Paul Lafler	L00101	PROC SQL: Beyond the Basics Using SAS®, Second Edition	11/03/2015
L002	Lauren Haworth Lake	L00201	Output Delivery System: The Basics and Beyond	11/03/2015
Z002	Cynthia Zender	L00201	Output Delivery System: The Basics and Beyond	11/03/2015

Join with Four Tables

A join with four tables adheres to the same rules as a two- and three-table join. Each table needs to be listed in the FROM clause with the appropriate subsetting (or matching) restrictions specified in a WHERE clause. The diagram, below, illustrates the table relationships along with the keys used in a four table join.



In the next example, a four-table join is constructed for identifying the book category along with the author name, book title, hardcopy price, and customer purchase date from the Authors, Books, Customers, and BookCategories tables. To accomplish this specific join, the AuthorID from the Authors and Books tables are matched, the BookID from the Books and Customers tables are matched, and the CategoryID from the Books and BookCategories tables are matched in a WHERE-clause. Finally, the results are arranged and displayed in ascending order.

PROC SQL Join Query

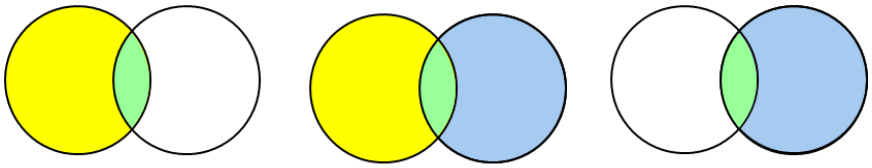
```
proc sql ;
 select AuthorName, BookTitle, HardcoverPrice, PurchaseDate, BookCategory
  from mydata.Authors,
       mydata.Books,
       mydata.Customers,
       mydata.BookCategories
   where Authors.AuthorID = Books.AuthorID
     AND Books.BookID = Customers.BookID
     AND Books.BookCategoryID = BookCategories.BookCategoryID
    order by PurchaseDate ;
quit ;
```

PROC SQL Join Query Results

AuthorName	BookTitle	HardcoverPrice	PurchaseDate	BookCategory
Michele M. Burlew	SAS® Macro Programming Made Easy, Third Edition	59.95	02/28/2015	Self-help: SAS Macro Language - Usage
Art Carpenter	Carpenter's Complete Guide to the SAS® Macro Language, Third Edition	74.95	02/28/2015	Self-help: SAS Macro Language - Usage
Ron Cody	An Introduction to SAS® University Edition	39.95	10/21/2015	Self-help: SAS Essentials
Ron Cody	Cody's Data Cleaning Techniques Using SAS®, Third Edition	47.95	11/03/2015	Self-help: SAS Essentials
Cynthia Zender	Output Delivery System: The Basics and Beyond	39.98	11/03/2015	Self-help: SAS ODS - Usage
Michele M. Burlew	Output Delivery System: The Basics and Beyond	39.98	11/03/2015	Self-help: SAS ODS - Usage
Kirk Paul Lafler	PROC SQL: Beyond the Basics Using SAS®, Second Edition	59.95	11/03/2015	Self-help: SAS PROC SQL - Usage
Lauren Haworth Lake	Output Delivery System: The Basics and Beyond	39.98	11/03/2015	Self-help: SAS ODS - Usage
Ron Cody	Cody's Data Cleaning Techniques Using SAS®, Third Edition	47.95	03/27/2016	Self-help: SAS Essentials
Susan Slaughter	The Little SAS® Book: A Primer, Fifth Edition	54.95	10/03/2016	Self-help: SAS Essentials
Lora D. Delwiche	The Little SAS® Book: A Primer, Fifth Edition	54.95	10/03/2016	Self-help: SAS Essentials
Ron Cody	Biostatistics by Example Using SAS® Studio	44.95	12/07/2016	Self-help: SAS Biostatistics - Usage
Susan Slaughter	The Little SAS® Book: A Primer, Fifth Edition	54.95	04/18/2017	Self-help: SAS Essentials
Ron Cody	Cody's Data Cleaning Techniques Using SAS®, Third Edition	47.95	04/18/2017	Self-help: SAS Essentials
Lora D. Delwiche	The Little SAS® Book: A Primer, Fifth Edition	54.95	04/18/2017	Self-help: SAS Essentials
Tricia Aanderud	An Introduction to SAS Visual Analytics: How to Explore Numbers, Design Reports, and Gain Insight into Your Data	49.95	05/11/2017	Self-help: SAS Visual Analytics - Usage
Kirk Paul Lafler	PROC SQL: Beyond the Basics Using SAS®, Second Edition	59.95	07/07/2017	Self-help: SAS PROC SQL - Usage

Outer Joins

A conventional join is a process of relating rows in one table with rows in another symmetrically. But occasionally, you may want to select the matching rows along with the preservation of unmatched rows from one, both or all tables. This approach is sometimes referred to as an asymmetric (or unconventional) join because its basic purpose is to select the matching rows from all tables, and to capture the rows without a match row from all tables. This type of join construct serves a significant purpose when working with tables of data and is referred to as an outer join construct.



Left Full Right

There are syntax and operational differences between inner (natural) and outer joins. The obvious difference between an outer and inner join is the way the syntax is constructed. Outer joins use keywords such as LEFT JOIN, RIGHT JOIN, and FULL JOIN, and has the WHERE clause replaced with an ON clause. These distinctions help identify outer joins from inner joins. But, there are important operational differences as well.

Unlike an inner join, the maximum number of tables that can be specified in an outer join construct is two. Similar to an inner join, an outer join relates rows in both tables. But this is where the similarities end since the resulting set of data also includes rows with no related rows from one or both of the tables. This special handling of "matched" and "unmatched" rows of data is what differentiates an outer join from an inner join. Essentially the resulting set of data from an outer join process contains rows that "match" the ON-clause plus any "unmatched" rows from the left, right, or both tables.

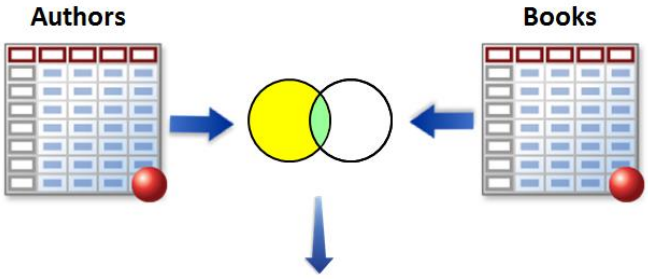
An outer join can accomplish a variety of tasks that would require a great deal of effort using other methods. This is not to say that a process similar to an outer join cannot be programmed – it would probably just require more work. Let's take a look at a few hypothetical tasks that are possible using outer joins:

- List all customer accounts with rentals during the month, including customer accounts with no purchase activity.
- Compute the number of rentals placed by each customer, including customers who have not rented.
- Identify movie renters who rented a movie last month, and those who did not.
- Identify a list of movie titles that an actor/actress appeared in, and movies they did not appear in.

Finally, specifying a left or right outer join is a matter of choice. Simply put, the only difference between a left and right join is the order of the tables used to relate rows of data. As such, the two types of outer joins can be specified based on convenience.

Left Outer Joins

Outer joins process data relationships from two tables differently than inner joins. In the next example a left outer join is constructed to select the "matched" AuthorIDs from both the Authors and Books tables, plus all the "unmatched" rows from the Authors table. The result contains all rows for which the SQL expression, referenced in the ON clause, matches the rows from the left table (Authors) that did not match any row in the right (Books) table. Essentially any "unmatched" rows from the left table are preserved and displayed as they appear in the table itself.



AuthorID	BookTitle	HardcoverPrice
S008	The Little SAS® Book, Fifth Edition	$54.95

PROC SQL Join Query

```
PROC SQL ;
 SELECT Authors.AuthorID, BookTitle, HardcoverPrice format=Dollar8.2
  FROM mydata.Authors
    LEFT JOIN
       mydata.Books
   ON Authors.AuthorID = Books.AuthorID ;
QUIT ;
```

PROC SQL Join Query Results

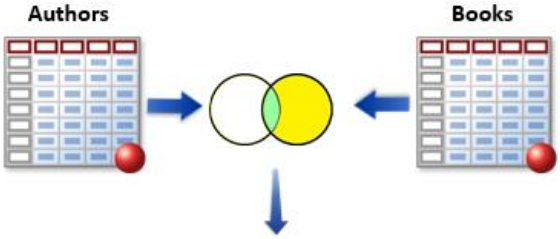
AuthorID	BookTitle	HardcoverPrice
A001	Building Business Intelligence Using SAS: Content Development Examples	$63.95
A001	An Introduction to SAS Visual Analytics: How to Explore Numbers, Design Reports, and Gain Insight into Your Data	$49.95
A002		.
B001		.
B002		.
B003	Output Delivery System: The Basics and Beyond	$39.98
B003	SAS® Hash Object Programming Made Easy	$29.95
B003	SAS® Macro Programming Made Easy, Third Edition	$59.95
B003	Combining and Modifying SAS® Data Sets: Examples, Second Edition	$48.95
C001	Carpenter's Complete Guide to the SAS® Macro Language, Third Edition	$74.95
C001	Carpenter's Complete Guide to the SAS® REPORT Procedure	$74.95
C001	Carpenter's Guide to Innovative SAS® Techniques	$84.95
C002		.
C003	Cody's Collection of Popular SAS® Programming Tasks and How to Tackle Them	$29.95
C003	An Introduction to SAS® University Edition	$39.95
C003	Test Scoring and Analysis Using SAS®	$39.95
C003	SAS® Statistics by Example	$49.95
C003	Biostatistics by Example Using SAS® Studio	$44.95
C003	Cody's Data Cleaning Techniques Using SAS®, Third Edition	$47.95
C003	Longitudinal Data and SAS®: A Programmer's Guide	$32.95
C003	Learning SAS® by Example: A Programmer's Guide	$89.95
C003	SAS® Functions by Example, Second Edition	$69.95

. . . . . .

S007		.
S008	Exercises and Projects for The Little SAS® Book, Fifth Edition	$24.95
S008	The Little SAS® Book: A Primer, Fifth Edition	$54.95
S008	The Little SAS® Enterprise Guide® Book	$54.95
T001		.
V001		.
W001		.
W002		.
Z001		.
Z002	Output Delivery System: The Basics and Beyond	$39.98

Right Outer Joins

The next example illustrates the result of using a right outer join to identify and match AuthorIDs from the Authors and Books tables. The result contains all rows for which the SQL expression, referenced in the ON clause, matches the rows from the right table (Books) that did not match any row in the left (Authors) table.



AuthorID	BookTitle	HardcoverPrice
S008	The Little SAS® Book, Fifth Edition	$54.95

PROC SQL Join Query

```
PROC SQL ;
 SELECT Authors.AuthorID, BookTitle, HardcoverPrice format=Dollar8.2
  FROM mydata.Authors
    RIGHT JOIN
       mydata.Books
   ON Authors.AuthorID = Books.AuthorID ;
QUIT ;
```

PROC SQL Join Query Results

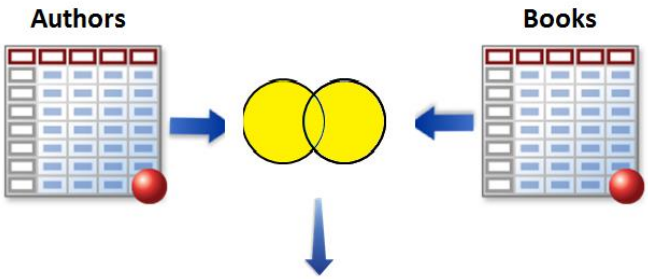
AuthorID	BookTitle	HardcoverPrice
A001	Building Business Intelligence Using SAS: Content Development Examples	$63.95
A001	An Introduction to SAS Visual Analytics: How to Explore Numbers, Design Reports, and Gain Insight Into Your Data	$49.95
B003	Output Delivery System: The Basics and Beyond	$39.98
B003	SAS® Hash Object Programming Made Easy	$29.95
B003	SAS® Macro Programming Made Easy, Third Edition	$59.95
B003	Combining and Modifying SAS® Data Sets: Examples, Second Edition	$48.95
C001	Carpenter's Complete Guide to the SAS® Macro Language, Third Edition	$74.95
C001	Carpenter's Complete Guide to the SAS® REPORT Procedure	$74.95
C001	Carpenter's Guide to Innovative SAS® Techniques	$84.95
C003	Cody's Collection of Popular SAS® Programming Tasks and How to Tackle Them	$29.95
C003	An Introduction to SAS® University Edition	$39.95
C003	Test Scoring and Analysis Using SAS®	$39.95
C003	SAS® Statistics by Example	$49.95
C003	Biostatistics by Example Using SAS® Studio	$44.95
C003	Cody's Data Cleaning Techniques Using SAS®, Third Edition	$47.95
C003	Longitudinal Data and SAS®: A Programmer's Guide	$32.95
C003	Learning SAS® by Example: A Programmer's Guide	$89.95
C003	SAS® Functions by Example, Second Edition	$69.95
D001	Exercises and Projects for The Little SAS® Book, Fifth Edition	$24.95
D001	The Little SAS® Book: A Primer, Fifth Edition	$54.95
D001	The Little SAS® Enterprise Guide® Book	$54.95
D002	Decision Trees for Analytics Using SAS® Enterprise Miner™	$49.95
L001	PROC SQL: Beyond the Basics Using SAS®, Second Edition	$59.95
L002	Output Delivery System: The Basics and Beyond	$39.98
O001	Exercises and Projects for The Little SAS® Book, Fifth Edition	$24.95
S008	Exercises and Projects for The Little SAS® Book, Fifth Edition	$24.95
S008	The Little SAS® Book: A Primer, Fifth Edition	$54.95
S008	The Little SAS® Enterprise Guide® Book	$54.95
Z002	Output Delivery System: The Basics and Beyond	$39.98

Full Outer Joins

A full outer join essentially represents the result of a left outer join and a right outer join. The result of a full outer join can be sizeable because it contains all "matched" and "unmatched" rows from both the left and right table. In the next example a full outer join query is constructed that selects the AuthorID, BookTitle, and HardcoverPrice from the Authors and Books tables. The result contains all rows that satisfy the SQL expression, referenced in the ON clause, by matching the rows from the right table (Books) that did not match any row in the left (Authors) table.

PROC SQL Join Query

```
PROC SQL ;
 SELECT Authors.AuthorID, BookTitle, HardcoverPrice format=Dollar8.2
  FROM mydata.Authors
    FULL JOIN
       mydata.Books
   ON Authors.AuthorID = Books.AuthorID ;
QUIT ;
```



AuthorID	BookTitle	HardcoverPrice
C001	Carpenter's Guide to Innovative SAS® Techniques	$84.95
C002		.

PROC SQL Join Query Results

AuthorID	BookTitle	HardcoverPrice
A001	Building Business Intelligence Using SAS: Content Development Examples	$63.95
A001	An Introduction to SAS Visual Analytics: How to Explore Numbers, Design Reports, and Gain Insight into Your Data	$49.95
A002		.
B001		.
B002		.
B003	Output Delivery System: The Basics and Beyond	$39.98
B003	SAS® Hash Object Programming Made Easy	$29.95
B003	SAS® Macro Programming Made Easy, Third Edition	$59.95
B003	Combining and Modifying SAS® Data Sets: Examples, Second Edition	$48.95
C001	Carpenter's Complete Guide to the SAS® Macro Language, Third Edition	$74.95
C001	Carpenter's Complete Guide to the SAS® REPORT Procedure	$74.95
C001	Carpenter's Guide to Innovative SAS® Techniques	$84.95
C002		.
C003	Cody's Collection of Popular SAS® Programming Tasks and How to Tackle Them	$29.95
C003	An Introduction to SAS® University Edition	$39.95
C003	Test Scoring and Analysis Using SAS®	$39.95
C003	SAS® Statistics by Example	$49.95
C003	Biostatistics by Example Using SAS® Studio	$44.95
C003	Cody's Data Cleaning Techniques Using SAS®, Third Edition	$47.95
C003	Longitudinal Data and SAS®: A Programmer's Guide	$32.95
C003	Learning SAS® by Example: A Programmer's Guide	$89.95
C003	SAS® Functions by Example, Second Edition	$69.95

. . . . . .

S007		.
S008	Exercises and Projects for The Little SAS® Book, Fifth Edition	$24.95
S008	The Little SAS® Book: A Primer, Fifth Edition	$54.95
S008	The Little SAS® Enterprise Guide® Book	$54.95
T001		.
V001		.
W001		.
W002		.
Z001		.
Z002	Output Delivery System: The Basics and Beyond	$39.98

Displaying a Query's Execution Plan

The SQL procedure offers users with a very important option to help understand a query's execution plan. Using a **_METHOD** option in a PROC SQL statement, users can display on the SAS Log the hierarchy of processing incurred by the SQL Optimizer. A list of the _METHOD codes, along with their descriptions, is displayed in the table, below.

Code	Description
SQXCRTA	Create table as Select
SQXSLCT	Select
SQXJSL	Step loop join (Cartesian)
SQXJM	Merge join
SQXJNDX	Index join
SQXJHSH	Hash join
SQXSORT	Sort
SQXSRC	Source rows from table
SQXFIL	Filter rows
SQXSUMG	Summary stats with GROUP BY
SQXSUMN	Summary stats with no GROUP BY

In the next example a _METHOD option is specified to display the three-table join query's execution plan.

PROC SQL Join Query

```
PROC SQL _METHOD ;
 SELECT A.AuthorID, AuthorName, B.BookID, BookTitle, PurchaseDate
  FROM mydata.Authors(drop=AuthorBio) AS A,
       mydata.Books(keep=BookID BookTitle AuthorID) AS B,
       mydata.Customers AS C
   WHERE A.AuthorID = B.AuthorID
     AND B.BookID = C.BookID ;
QUIT ;
```

PROC SQL Join Query Results

```
NOTE: SQL execution methods chosen are:

        sqxslct
            sqxjm
                sqxjm
                    sqxsrc( MYDATA.BOOKS(alias = B) )
                    sqxsrc( MYDATA.CUSTOMERS(alias = C) )
                sqxsrc( MYDATA.AUTHORS(alias = A) )
```

Specifying the BUFFERSIZE= Option

If surplus virtual (or real) memory is available, in-memory processing may be utilized by the SQL Optimizer to improve the performance associated with a query's matching, aggregation, and intersection processing. One technique uses a **BUFFERSIZE=** option to specify the internal buffer page size for paged memory (up to the size of the MEMSIZE= value) processing including performing in-memory, and faster, operations. Since memory speeds are generally faster (nanoseconds) in comparison to secondary storage (milliseconds), a query's performance can be improved. In the next example, a BUFFERSIZE=1024000 option (other values can be specified based on the amount of available memory) is specified to tell the SQL Optimizer the amount of memory that is available for in-memory processing.

PROC SQL Join Query

```
PROC SQL _METHOD BUFFERSIZE=1024000 ;
 SELECT A.AuthorID, AuthorName, B.BookID, BookTitle, PurchaseDate
  FROM mydata.Authors(drop=AuthorBio) AS A,
       mydata.Books(keep=BookID BookTitle AuthorID) AS B,
       mydata.Customers AS C
   WHERE A.AuthorID = B.AuthorID
     AND B.BookID = C.BookID ;
QUIT ;
```

PROC SQL Join Query Results

```
NOTE: SQL execution methods chosen are:

       sqxslct
           sqxjhsh
               sqxjhsh
                   sqxsrc( MYDATA.BOOKS(alias = B) )
                   sqxsrc( MYDATA.CUSTOMERS(alias = C) )
               sqxsrc( MYDATA.AUTHORS(alias = A) )
```

Conclusion

The SQL procedure is a tool for SAS users to explore and use in a variety of application situations, including powerful and essential join constructs using PROC SQL to combine two or more tables of data together. This paper and presentation described and illustrated the join process, including what a join is, exploration of one-to-one, one-to-many, and many-to-many data relationships, identifying a primary key (or unique identifier), and popular join techniques available to SAS users. The application of conventional (or symmetrical matching) and unconventional (or asymmetrical left, right and full matching) join queries were also illustrated.

References

Choate, Paul A. (2007). "*Quick Record Lookup without an Index*", Proceedings of the 2007 SAS Global Forum (SGF) Conference.

Hermansen, Sigurd W. (2003). "*Sublime Secrets of the SAS® SQLheads*", Proceedings of the 2003 SouthEast SAS Users Group (SESUG) Conference.

Lafler, Kirk Paul and Charu Shankar (2017), *"One-to-one, One-to-many, and Many-to-many Joins Using PROC SQL,"* Proceedings of the *2017* Western Users of SAS Software (WUSS) Conference.

Lafler, Kirk Paul (2015 and 2014), *"Five Little Known, But Highly Valuable and Widely Usable, PROC SQL Programming Techniques," Proceedings of the 2015 SouthEast SAS Users Group (SESUG) Conference; 2015* Ohio SAS Users Group (OhioSUG) Conference; and 2014 Wisconsin-Illinois SAS Users Group (WIILSUG) Conference.

Lafler, Kirk Paul (2013). *PROC SQL: Beyond the Basics Using SAS, Second Edition*, SAS Institute Inc., Cary, NC, USA.

Lafler, Kirk Paul (2013, 2011), *"Quick Results with PROC SQL,"* Proceedings of the 2013 North East SAS Users Group (NESUG) Conference; *Proceedings of the 2011* Western Users of SAS Software (WUSS) Conference; and Proceedings of the 2011 MidWest SAS Users Group (MWSUG) Conference.

Lafler, Kirk Paul (2011), *"Powerful and Sometimes Hard-to-find PROC SQL Features,"* Proceedings of the 2011 Pharmaceutical SAS Users Group (PharmaSUG) Conference.

Lafler, Kirk Paul (2010), *"Exploring DATA Step Merges and PROC SQL Joins,"* South Central SAS Users Group (SCSUG) 2010 Conference.

Lafler, Kirk Paul (2009), *"Exploring DICTIONARY Tables and SASHELP Views,"* South Central SAS Users Group (SCSUG) Conference; Proceedings of the 2009 Western Users of SAS Software (WUSS) Conference; and Proceedings of the 2009 Pharmaceutical SAS Users Group (PharmaSUG) Conference.

Lafler, Kirk Paul (2008), *"Kirk's Top Ten Best PROC SQL Tips and Techniques,"* Wisconsin Illinois SAS Users Conference (June 26th, 2008 One-day Conference).

Lafler, Kirk Paul (2008 and 2007), *"Undocumented and Hard-to-find PROC SQL Features,"* Greater Atlanta SAS Users Group (GASUG) Meeting (June 11th, 2008); Proceedings of the 2008 Pharmaceutical SAS Users Group (PharmaSUG) Conference; 2008 Michigan SAS Users Group (MSUG) Conference (May 29th, 2008); 2008 Vancouver SAS Users Group Meeting (April 23rd, 2008); 2008 Philadelphia SAS Users Group (PhilaSUG) Meeting (March 13th, 2008); and Proceedings of the 2007 Pharmaceutical SAS Users Group (PharmaSUG) Conference.

Lafler, Kirk Paul and Ben Cochran (2007), *"A Hands-on Tour Inside the World of PROC SQL Features,"* Proceedings of the SAS Global Forum (SGF) 2007 Conference, Software Intelligence Corporation, Spring Valley, CA, and The Bedford Group, USA.

Lafler, Kirk Paul (2006), *"A Hands-on Tour Inside the World of PROC SQL,"* Proceedings of the Thirty-first Annual SAS Users Group International (SUGI) Conference.

Lafler, Kirk Paul (2005), *"A Hands-on Tour of the 5 Most Exciting Features Found in PROC SQL,"* Proceedings of the 2005 Western Users of SAS Software (WUSS) Conference.

Lafler, Kirk Paul (2005), "*Manipulating Data with PROC SQL*," Proceedings of the Thirtieth Annual SAS Users Group International (SUGI) Conference.

Lafler, Kirk Paul (2004). *PROC SQL: Beyond the Basics Using SAS*, SAS Institute Inc., Cary, NC, USA.

Lafler, Kirk Paul (2003), "*Undocumented and Hard-to-find PROC SQL Features*," Proceedings of the Eleventh Annual Western Users of SAS Software Conference.

Lavery, Russ (2005). "*The SQL Optimizer Project: _Method and _Tree in SAS® 9.1*", Proceedings of the 2005 SAS Users Group International (SUGI) Conference.

Phelps, Kent; Ronda Phelps and Kirk Paul Lafler (2014). "*The Joinless Join; Expand the Power of SAS® Enterprise Guide® in a New Way*", Proceedings of the 2014 MidWest SAS Users Group (MWSUG) Conference.

Ross, Barbara and Mary Jessica Bennett (2016). "*PROC SQL for SQL Die-hards*".

Williams, Christianna S. (2012). "*Queries, Joins, and WHERE Clauses, Oh My!! Demystifying PROC SQL*", Proceedings of the 2012 SAS Global Forum (SGF) Conference.

Acknowledgments

The authors thank Clarence Jackson and Greg Gengo, SouthCentral SAS Users Group (SCSUG) Conference Co-Chairs for accepting our abstract and paper; the SouthCentral SAS Users Group (SCSUG) Executive Board; and SAS Institute for organizing and supporting a great conference!

Trademark Citations

Tables Used in Examples

The examples used throughout this paper utilize a database of five tables (A relational database is a collection of tables.) The tables and data correspond to a selection of SAS authors, book categories, book formats, books and customers. The Authors table consists of two columns: AuthorID and AuthorName, both of which are defined as character columns, illustrated below.

Authors Table

AuthorID	AuthorName
A001	Tricia Aanderud
A002	Robert Allison
B001	William Benjamin
B002	Jonas V. Bilenas
B003	Michele M. Burlew
C001	Art Carpenter
C002	Goutam Chakraborty
C003	Ron Cody
D001	Lora D. Delwiche
D002	Barry de Ville
D003	Craig Dickstein
D004	Paul Dorfman
E001	Peter Eberhardt
E002	Jane Eslinger
F001	Lisa Fine
G001	Sunil K. Gupta
H001	Angela Hall
H002	Lauren Haworth
H003	Dan Heath
H004	Chris Hemedinger
H005	Don Henderson
H006	Philip Holland
J001	Mark Jordan
K001	Warren F. Kuhfeld
L001	Kirk Paul Lafler
L002	Lauren Haworth Lake
L003	Ann Lehman
M001	Phil Mason
M002	Sanjay Matange
M003	Thomas Miron
M004	Derek P. Morgan
O001	Rebecca A. Ottesen
P001	Olivia Parr-Rud
P002	Frederick Pratter
R001	Michael Raithel
S001	John Sall
S002	Armistead Sapp
S003	Marge Scerbo
S004	Sandra D. Schlotzhauer
S005	Brian C. Shilling
S006	Charles E. Shipp
S007	Jack Shostak
S008	Susan Slaughter
T001	Michael Tuchman
V001	Robert Virgile
W001	Perry Watts
W002	Rick Wicklin
Z001	Mike Zdeb
Z002	Cynthia Zender

The BookCategories table consists of three columns: BookCategoryID, BookCategory, and Notes, all of which are defined as character columns. The physical data stored in the BookCategories table is illustrated below.

BookCategories Table

BookCategoryID	BookCategory	Notes
SH001	Self-help: SAS Essentials	A Self-help book on SAS essentials.
SH002	Self-help: SAS Procedures - Usage	A Self-help book on SAS procedure usage with examples.
SH003	Self-help: SAS DATA Step - Usage	A Self-help book on SAS DATA Step usage with examples.
SH004	Self-help: SAS Enterprise Guide (EG) - Usage	A Self-help book on SAS Enterprise Guide (EG) usage with examples.
SH005	Self-help: SAS Functions - Usage	A Self-help book on SAS Functions usage with examples.
SH006	Self-help: SAS ODS - Usage	A Self-help book on SAS Output Delivery System (ODS) usage with examples.
SH007	Self-help: SAS PROC SQL - Usage	A Self-help book on SAS PROC SQL usage with examples.
SH008	Self-help: SAS PROC REPORT - Usage	A Self-help book on SAS PROC REPORT usage with examples.
SH009	Self-help: SAS PROC TABULATE - Usage	A Self-help book on SAS PROC TABULATE usage with examples.
SH010	Self-help: SAS PROC DS2 - Usage	A Self-help book on SAS PROC DS2 usage with examples.
SH011	Self-help: SAS Macro Language - Usage	A Self-help book on SAS Macro Language usage with examples.
SH012	Self-help: SAS Hash Object Programming - Usage	A Self-help book on SAS Hash Object Programming usage with examples.
SH013	Self-help: SAS Business Intelligence (BI) - Usage	A Self-help book on SAS Business Intelligence (BI) usage with examples.
SH014	Self-help: SAS Statistical Graphics - Usage	A Self-help book on SAS Statistical Graphics usage with examples.
SH015	Self-help: SAS Visual Analytics - Usage	A Self-help book on SAS Visual Analytics usage with examples.
SH016	Self-help: SAS Statistics - Usage	A Self-help book on SAS Statistics usage with examples.
SH017	Self-help: SAS Biostatistics - Usage	A Self-help book on SAS Biostatistics usage with examples.

The BookFormats table consists of three columns: BookCategoryID, BookCategory, and Notes, all of which are defined as character columns. The physical data stored in the BookCategories table is illustrated below.

BookFormats Table

FormatID	BookFormat	Style Sheets	Notes
B01	Hardcover		A Hardcover book is bound with rigid protective covers.
B02	Paperback		A paperback book has a thick paper or paperboard cover, and is held together with glue rather than stitches or staples.
B03	e-Book		An electronic book (or e-Book) is a book publication made available in digital form.
B04	AudioBook		An audiobook (or talking book) is a recording of text being read.
B05	Ring Connection		The ring connection, comb-bound or spiral connection, is a popular book connection, which binds pages using spiral wires.
B06	Bunkobon		In Japan, bunkobon are small-format paperback books, designed to be affordable and portable.
E01	eBook - PDF	N/A	Base Code: Portable Document Facility
E02	eBook - Mobipocket	CSS3	Base Code: Amazon-specific - HTML3ish
E03	eBook - KF8	CSS3	Base Code: HTML 4, XHTML1.1, (X)HTML 5
E04	eBook - ePub2	CSS2	Base Code: XHTML 1.1
E05	eBook - ePub3	CSS3	Base Code: (X)HTML5
E06	eBook - ePlb	N/A	Base Code: XHTML (similar to ePub2)
E07	eBook - IBooks Author	CSS3	Base Code: XHTML 1.1 with extensions (similar to ePub2)

The Books table consists of fifteen columns: BookID, BookTitle, AuthorID, BookCategoryID, Publisher, Edition, CopyrightDate, FormatID, HardcopyISBN, ePubISBN, MobiISBN, PDFISBN, PageCount, HardcoverPrice, and BookHyperlink, all of which are defined as character columns except PageCount and HardcoverPrice are defines as numeric columns. The physical data stored in the Books table is illustrated below.

Books Table

BookID	BookTitle	AuthorID	BookCategoryID	Publisher	Edition	CopyrightDate	FormatID	HardcopyISBN	ePubISBN	MobiISBN	PDFISBN	PageCount	HardcoverPrice	BookHyperlink
A00101	Building Business Intelligence Using SAS: Content Development Examples	A001	SH013	SAS Institute	1st Edition	January 2012	B01	978-1-60764-988-5	978-1-60764-994-6	978-1-61290-952-3	978-1-62959-020-2	378	63.95	https://www.sas.com/store/books/categories/examples/building-business-intelligence-using-sas-content-development-examples/prodBK_64393_en.html?storeCode=SAS_US
A00102	An Introduction to SAS Visual Analytics: How to Explore Numbers, Design Reports, and Gain Insight into Your Data	A001	SH015	SAS Institute	1st Edition	March 2017	B01	978-1-62960-291-2	978-1-63526-042-7	978-1-63526-043-4	978-1-63526-044-1	294	49.95	https://www.sas.com/store/books/categories/usage-and-reference/an-introduction-to-sas-visual-analytics-how-to-explore-numbers-design-reports-and-gain-insight-into-your-data/prodBK_69462_en.html?storeCode=SAS_US
B00304	SAS® Macro Programming Made Easy, Third Edition	B003	SH011	SAS Institute	3rd Edition	June 2014	B01	978-1-61290-693-5	978-1-62959-151-3	978-1-62959-152-0	978-1-61290-791-8	386	59.95	https://www.sas.com/store/books/categories/getting-started/sas-macro-programming-made-easy-third-edition/prodBK_66295_en.html?storeCode=SAS_US&storeCode=SAS_US
B00303	SAS® Hash Object Programming Made Easy	B003	SH012	SAS Institute	1st Edition	September 2012	B01	978-1-60764-801-7	978-1-61290-098-8	978-1-61290-951-5	978-1-62959-028-8	208	29.95	https://www.sas.com/store/prodBK_62230_en.html?storeCode=SAS_US&storeCode=SAS_US
B00302	Combining and Modifying SAS® Data Sets: Examples, Second Edition	B003	SH001	SAS Institute	2nd Edition	November 2009	B01	978-1-59047-920-9				332	48.95	https://www.sas.com/store/prodBK_60961_en.html?storeCode=SAS_US&storeCode=SAS_US#
L00201	Output Delivery System: The Basics and Beyond	B003	SH006	SAS Institute	1st Edition	October 2009	B01	978-1-59994-660-3				536	39.98	https://www.sas.com/store/books/categories/examples/output-delivery-system-the-basics-and-beyond/prodBK_61685_en.html?storeCode=SAS_US
C00101	Carpenter's Complete Guide to the SAS® REPORT Procedure	C001	SH008	SAS Institute	1st Edition	May 2007	B01	978-1-59994-195-0				490	74.95	https://www.sas.com/store/prodBK_60966_en.html?storeCode=SAS_US&storeCode=SAS_US
C00102	Carpenter's Guide to Innovative SAS® Techniques	C001	SH001	SAS Institute	1st Edition	March 2012	B01	978-1-60764-991-5	978-1-61290-202-9	978-1-61290-951-5	978-1-62959-019-6	570	84.95	https://www.sas.com/store/prodBK_62454_en.html?storeCode=SAS_US&storeCode=SAS_US
C00103	Carpenter's Complete Guide to the SAS® Macro Language, Third Edition	C001	SH011	SAS Institute	3rd Edition	August 2016	B01	978-1-62959-268-8	978-1-62960-237-0	978-1-62960-238-7	978-1-62960-239-4	540	74.95	https://www.sas.com/store/books/categories/usage-and-reference/carpenter-s-complete-guide-to-the-sas-macro-language-third-edition/prodBK_67815_en.html?storeCode=SAS_US
C00309	Cody's Data Cleaning Techniques Using SAS®, Third Edition	C003	SH001	SAS Institute	3rd Edition	March 2017	B01	978-1-62960-796-2	978-1-63526-057-0	978-1-63526-058-7	978-1-63526-059-4	234	47.95	https://www.sas.com/store/books/categories/usage-and-reference/cody-s-data-cleaning-techniques-using-sas-third-edition/prodBK_70074_en.html?storeCode=SAS_US
C00308	Biostatistics by Example Using SAS® Studio	C003	SH017	SAS Institute	1st Edition	September 2016	B01	978-1-62960-328-5	978-1-62960-493-0	978-1-62960-494-7	978-1-62960-495-4	262	44.95	https://www.sas.com/store/books/categories/getting-started/biostatistics-by-example-using-sas-studio/prodBK_69326_en.html?storeCode=SAS_US
C00307	An Introduction to SAS® University Edition	C003	SH001	SAS Institute	1st Edition	September 2015	B01	978-1-62959-770-6	978-1-62960-007-9	978-1-62960-008-6	978-1-62960-010-9	366	39.95	https://www.sas.com/store/prodBK_68380_en.html?storeCode=SAS_US&storeCode=SAS_US

C00306	Test Scoring and Analysis Using SAS®	C003	SH016	SAS Institute	1st Edition	December 2014	B01	978-1-61290-924-0	978-1-62959-495-8	978-1-62959-496-5	978-1-62959-497-2	200	39.95	https://www.sas.com/store/prodBK_67044_en.html?storeCode=SAS_US
C00305	Cody's Collection of Popular SAS® Programming Tasks and How to Tackle Them	C003	SH001	SAS Institute	1st Edition	October 2012	B01	978-1-61290-333-0	978-1-62959-776-8	978-1-62959-777-5	978-1-61290-439-9	162	29.95	https://www.sas.com/store/prodBK_65193_en.html?storeCode=SAS_US&storeCode=SAS_US
C00304	SAS® Statistics by Example	C003	SH016	SAS Institute	1st Edition	August 2011	B01	978-1-60764-800-0	978-1-61290-012-4	978-1-61290-954-7	978-1-62959-023-6	274	49.95	https://www.sas.com/store/books/categories/examples/sas-statistics-by-example/prodBK_63671_en.html?storeCode=SAS_US
C00303	SAS® Functions by Example, Second Edition	C003	SH001	SAS Institute	2nd Edition	March 2010	B01	978-1-60764-340-1	978-1-60764-364-7	978-1-61290-956-1	978-1-62959-024-0	472	69.95	https://www.sas.com/store/prodBK_62857_en.html?storeCode=SAS_US&storeCode=SAS_US
C00302	Learning SAS® by Example: A Programmer's Guide	C003	SH001	SAS Institute	1st Edition	March 2007	B01	978-1-59994-165-3	978-1-59994-426-5	978-1-61290-946-2	978-1-62959-014-1	664	89.95	https://www.sas.com/store/prodBK_60864_en.html?storeCode=SAS_US&storeCode=SAS_US
C00301	Longitudinal Data and SAS®: A Programmer's Guide	C003	SH016	SAS Institute	1st Edition	October 2001	B01	978-1-58025-924-8	978-1-62959-249-7	978-1-62959-248-0	978-1-62959-247-3	208	32.95	https://www.sas.com/store/prodBK_58176_en.html?storeCode=SAS_US&storeCode=SAS_US
S00803	The Little SAS® Enterprise Guide® Book	D001	SH004	SAS Institute	1st Edition	February 2017	B01	978-1-62960-380-3	978-1-62960-888-4	978-1-62960-889-1	978-1-62960-890-7	330	54.95	https://www.sas.com/store/books/categories/getting-started/the-little-sas-enterprise-guide-book/prodBK_69524_en.html?storeCode=SAS_US
S00803	The Little SAS® Enterprise Guide® Book	S008	SH004	SAS Institute	1st Edition	February 2017	B01	978-1-62960-380-3	978-1-62960-888-4	978-1-62960-889-1	978-1-62960-890-7	330	54.95	https://www.sas.com/store/books/categories/getting-started/the-little-sas-enterprise-guide-book/prodBK_69524_en.html?storeCode=SAS_US
S00801	The Little SAS® Book: A Primer, Fifth Edition	D001	SH001	SAS Institute	5th Edition	October 2012	B01	978-1-61290-343-9	978-1-61290-400-9	978-1-61290-945-5	978-1-62959-013-4	378	54.95	https://www.sas.com/store/books/categories/getting-started/the-little-sas-book-a-primer-fifth-edition/prodBK_65423_en.html?storeCode=SAS_US
S00801	The Little SAS® Book: A Primer, Fifth Edition	S008	SH001	SAS Institute	5th Edition	October 2012	B01	978-1-61290-343-9	978-1-61290-400-9	978-1-61290-945-5	978-1-62959-013-4	378	54.95	https://www.sas.com/store/books/categories/getting-started/the-little-sas-book-a-primer-fifth-edition/prodBK_65423_en.html?storeCode=SAS_US
O00102	Exercises and Projects for The Little SAS® Book, Fifth Edition	D001	SH001	SAS Institute	5th Edition	July 2015	B01	978-1-62959-655-6	978-1-62959-805-5	978-1-62959-806-2	978-1-62959-807-9	148	24.95	https://www.sas.com/store/books/categories/getting-started/exercises-and-projects-for-the-little-sas-book-fifth-edition/prodBK_67330_en.html?storeCode=SAS_US
000102	Exercises and Projects for The Little SAS® Book, Fifth Edition	S008	SH001	SAS Institute	5th Edition	July 2015	B01	978-1-62959-655-6	978-1-62959-805-5	978-1-62959-806-2	978-1-62959-807-9	148	24.95	https://www.sas.com/store/books/categories/getting-started/exercises-and-projects-for-the-little-sas-book-fifth-edition/prodBK_67330_en.html?storeCode=SAS_US
C00102	Exercises and Projects for The Little SAS® Book, Fifth Edition	O001	SH001	SAS Institute	5th Edition	July 2015	B01	978-1-62959-655-6	978-1-62959-805-5	978-1-62959-806-2	978-1-62959-807-9	148	24.95	https://www.sas.com/store/books/categories/getting-started/exercises-and-projects-for-the-little-sas-book-fifth-edition/prodBK_67330_en.html?storeCode=SAS_US
D00201	Decision Trees for Analytics Using SAS® Enterprise Miner™	D002	SH016	SAS Institute	1st Edition	July 2013	B01	978-1-61290-315-6	978-1-62959-100-1	978-1-62959-101-8	978-1-62959-807-9	268	49.95	https://www.sas.com/store/books/categories/usage-and-reference/decision-trees-for-analytics-using-sas-enterprise-miner-/prodBK_63319_en.html?storeCode=SAS_US
L00101	PROC SQL: Beyond the Basics Using SAS®, Second Edition	L001	SH007	SAS Institute	2nd Edition	October 2013	B01	978-1-61290-027-8	978-1-61290-734-5	978-1-62959-089-9	978-1-62959-088-2	400	59.95	https://www.sas.com/store/prodBK_62432_en.html?storeCode=SAS_US
L00201	Output Delivery System: The Basics and Beyond	L002	SH006	SAS Institute	1st Edition	October 2009	B01	978-1-59994-660-3				636	39.98	https://www.sas.com/store/books/categories/examples/output-delivery-system-the-basics-and-beyond/prodBK_61685_en.html?storeCode=SAS_US
L00201	Output Delivery System: The Basics and Beyond	Z002	SH006	SAS Institute	1st Edition	October 2009	B01	978-1-59994-660-3				636	39.98	https://www.sas.com/store/books/categories/examples/output-delivery-system-the-basics-and-beyond/prodBK_61685_en.html?storeCode=SAS_US

The Customers table consists of three columns: CustomerID, BookID, and PurchaseDate, where CustomerID and BookID are defined as character columns and PurchaseDate is defined as a numeric column. The physical data stored in the Customers table is illustrated below.

Customers Table

CustomerID	BookID	PurchaseDate
9753111	C00309	04/18/2017
9753111	S00801	04/18/2017
1234567	S00801	10/03/2016
2244667	L00101	07/07/2017
8172639	B00304	02/26/2015
8172639	C00103	02/26/2015
3339876	C00307	10/21/2015
3339876	L00101	11/03/2015
3339876	L00201	11/03/2015
3339876	C00309	11/03/2015
4917563	C00309	03/27/2016
0053218	C00308	12/07/2016
1357964	A00102	05/11/2017
1357964	D00102	09/22/2016

Author Information

Kirk Paul Lafler is an entrepreneur, consultant and founder of Software Intelligence Corporation, and has been using SAS since 1979. Kirk is a SAS application developer, programmer, certified professional, provider of SAS consulting services, mentor, advisor and professor at UC San Diego Extension, educator to SAS users around the world, and emeritus sasCommunity.org Advisory Board member. As the author of six books including Google® Search Complete (Odyssey Press. 2014) and PROC SQL: Beyond the Basics Using SAS, Second Edition (SAS Press. 2013); Kirk has written hundreds of papers and articles; been an Invited speaker and trainer at hundreds of SAS International, regional, special-interest, local, and in-house user group conferences and meetings; and is the recipient of 25 "Best" contributed paper, hands-on workshop (HOW), and poster awards.

Charu Shankar is a technology trainer, wellness coach, writer and public speaker with proven ability in delivering top quality training resulting in a high degree of satisfaction among clients. As a SAS instructor Charu helps individuals and organizations leverage SAS to learn and use SAS creatively to solve practical business problems. Charu is a popular speaker at SAS user group conferences and has helped train thousands of SAS users. She also helps individuals land their dream SAS job to meet their technology goals and make a contribution to society with their skills. Charu's blog posts are found at blogs.sas.com/content/author/charushankar.

Comments and suggestions can be sent to:

Kirk Paul Lafler
SAS® Consultant, Application Developer, Programmer, Data Analyst, Educator and Author
Software Intelligence Corporation
E-mail: KirkLafler@cs.com
LinkedIn: http://www.linkedin.com/in/KirkPaulLafler
Twitter: @sasNerd

~ ~ ~ ~ ~

Charu Shankar
Technical Trainer, SAS Institute, Canada
E-mail: charu_shankar@hotmail.com
LinkedIn: https://www.linkedin.com/in/charushankar
Twitter: CharuSAS