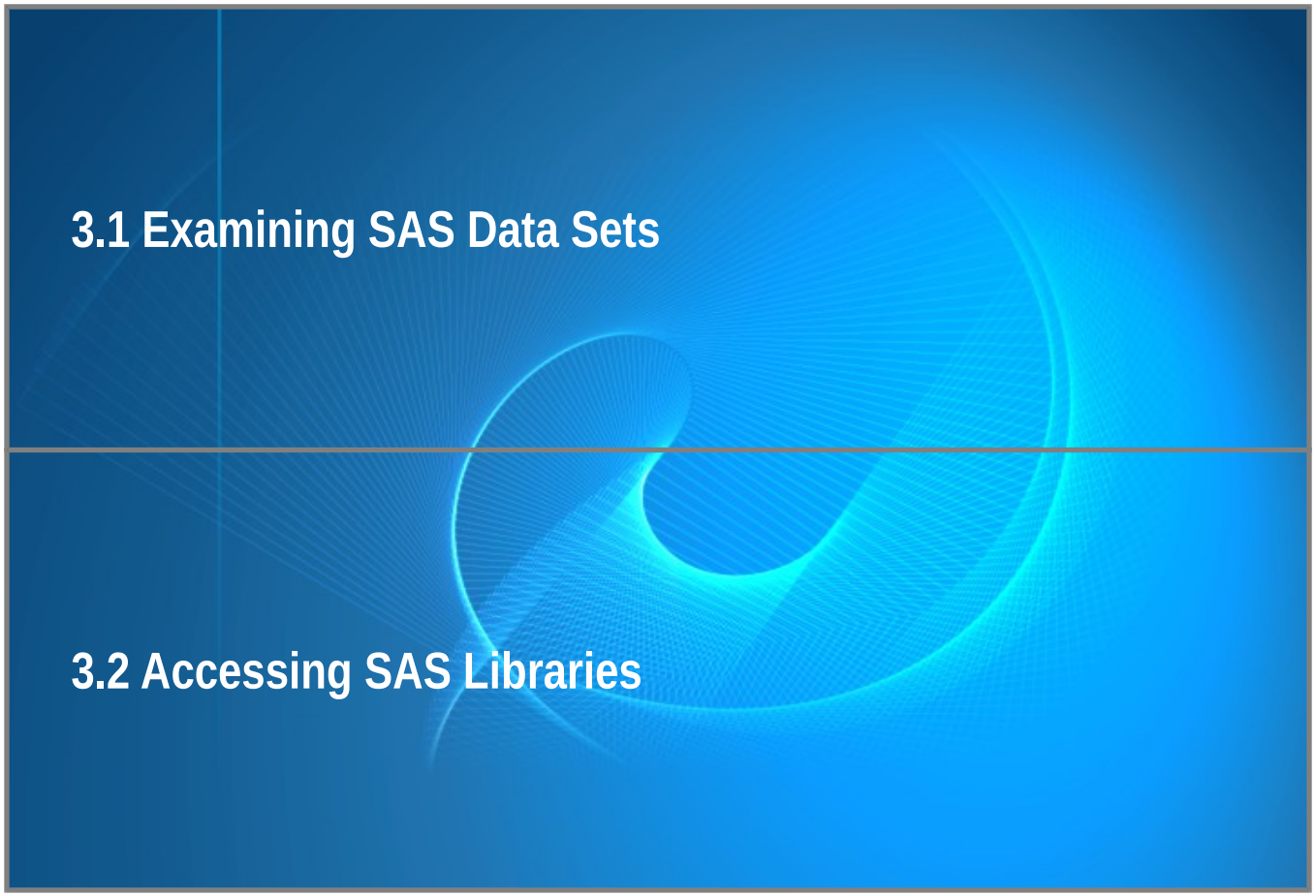
Chapter 3: Accessing Data



Chapter 3: Accessing Data

3.1 Examining SAS Data Sets

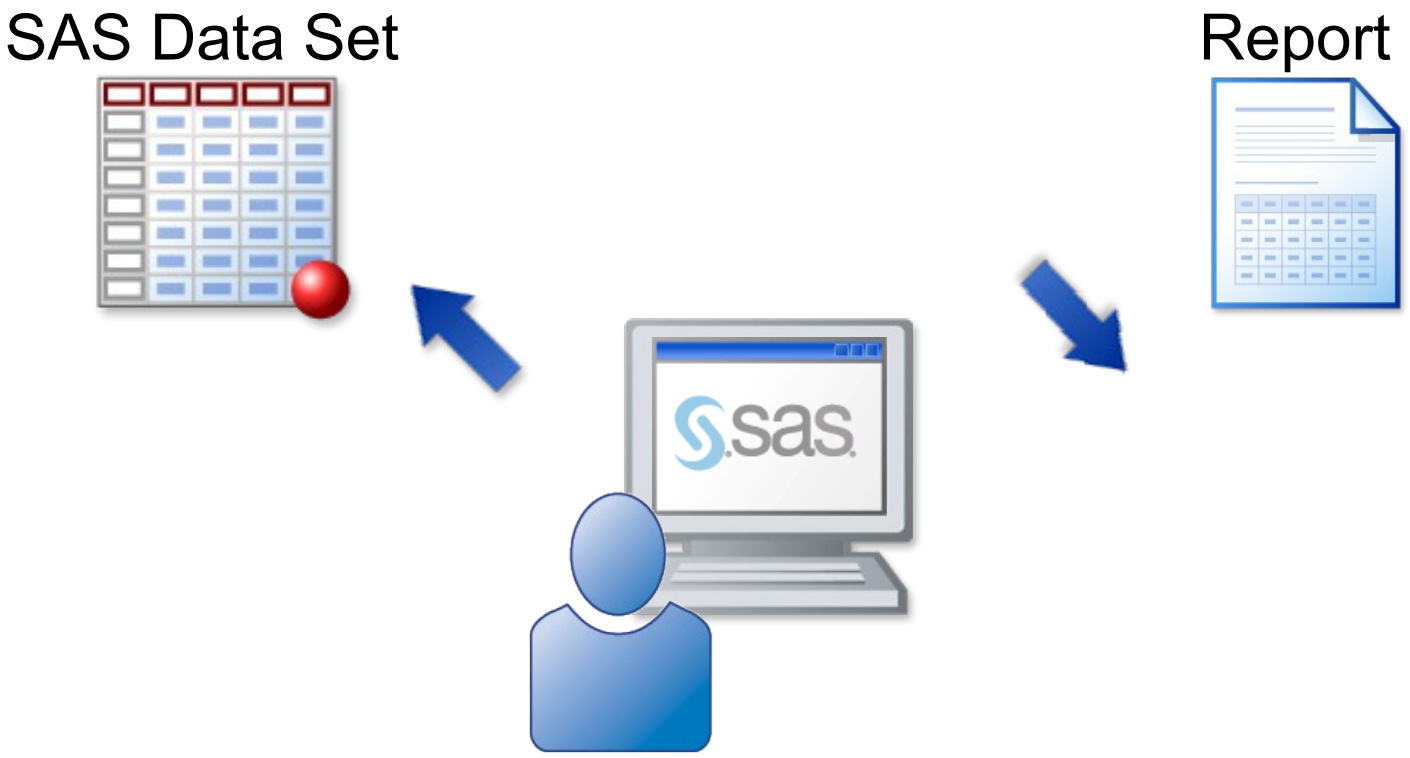
3.2 Accessing SAS Libraries

Objectives

- Define the components of a SAS data set.
- Browse the descriptor portion of a SAS data set using the CONTENTS procedure.
- Browse the data portion of a SAS data set using the PRINT procedure.
- Define a SAS variable.
- Define a missing value.
- Define a SAS date value.

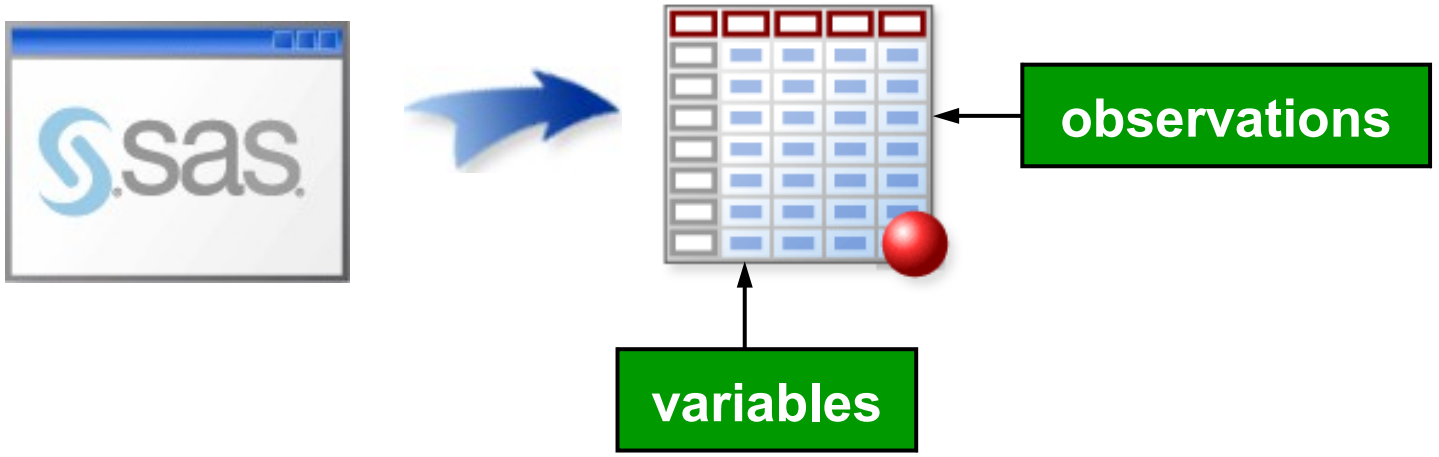
Business Scenario

Many SAS data sets related to the Orion Star project already exist. The programmers need to know how to display the structure and contents of the data sets.



What Is a SAS Data Set?

A *SAS data set* is a specially structured data file that SAS creates and that only SAS can read. A SAS data set is a table that contains observations and variables.

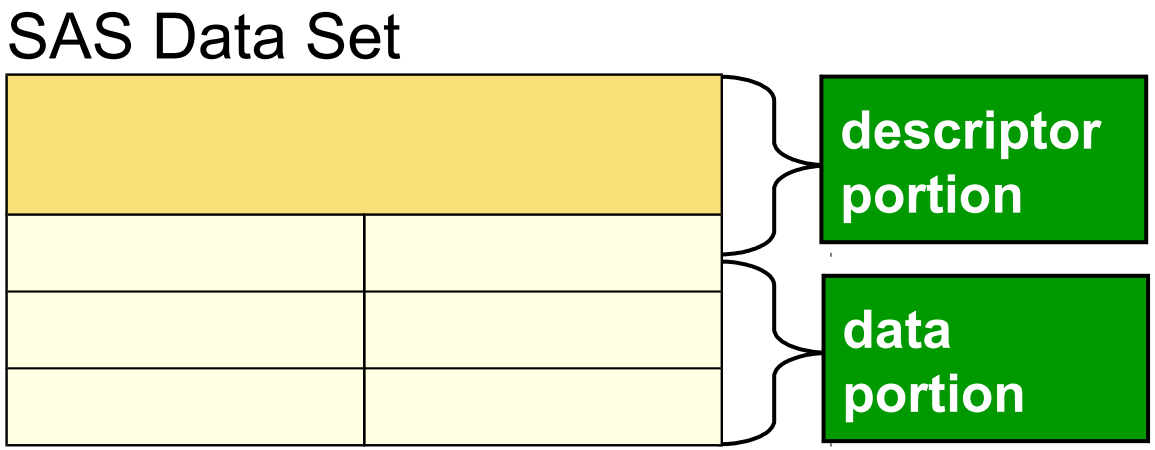


SAS Data Set Terminology

SAS Terminology	Database Terminology
SAS Data Set	~~Table~~
Observation	~~Row~~
Variable	~~Column~~

SAS Data Set Terminology

A SAS data set contains a descriptor portion and a data portion.



Descriptor Portion

The *descriptor portion* contains the following metadata:

- general properties (such as data set name and number of observations)
- variable properties (such as name, type, and length)

Partial **work.newsalesemps**

Data Set Name WORK.NEWSALESEMPS **Engine** V9 **Created** Mon, Feb 27, 2012 01:28 PM **Observations** 71 **Variables** 4 ...			
First_Name **$ 12**	**Last_Name** **$ 18**	**Job_Title** **$ 25**	**Salary** **N 8**

general properties

variable properties

Browsing the Descriptor Portion

Use *PROC CONTENTS* to display the descriptor portion of a SAS data set.

```
proc contents data=work.newsalesemps;
run;
```

PROC CONTENTS DATA=*SAS-data-set*;
RUN;

Viewing the Output

Partial PROC CONTENTS Output

The CONTENTS Procedure

Data Set Name	WORK.NEWSALESEMPS	Observations	71
Member Type	DATA	Variables	4
Engine	V9	Indexes	0
Created	Mon, Feb 27, 2012 01:28:51 PM	Observation Length	64
Last Modified	Mon, Feb 27, 2012 01:28:51 PM	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO

Engine/Host Dependent Information

...

Alphabetic List of Variables and Attributes

#	Variable	Type	Len
1	First_Name	Char	12
3	Job_Title	Char	25
2	Last_Name	Char	18
4	Salary	Num	8

3.01 Quiz

How many observations are in the data set **work.donations**?

- Retrieve program **p103a01**.
- After the DATA step, add a PROC CONTENTS step to view the descriptor portion of **work.donations**.
- Submit the program and review the results.

3.01 Quiz – Correct Answer

How many observations are in the data set **work.donations**? **124 observations**

```
data work.donations;
   infile "&path\donation.dat";
   input Employee_ID Qtr1 Qtr2 Qtr3 Qtr4;
   Total=sum(Qtr1,Qtr2,Qtr3,Qtr4);
run;

proc contents data=work.donations;
run;
```

Data Portion

The *data portion* of a SAS data set contains the data values, which are either character or numeric.

Partial **work.newsalesemps**

First_Name	Last_Name	Job_Title	Salary
Satyakam	Denny	Sales Rep. II	26780
Monica	Kletschkus	Sales Rep. IV	30890
Kevin	Lyon	Sales Rep. I	26955
Petrea	Soltau	Sales Rep. II	27440

variable names

data values

character values

numeric values

Browsing the Data Portion

Use *PROC PRINT* to display the data portion of a SAS data set.

```
proc print data=work.newsalesemps;
run;
```

PROC PRINT DATA=*SAS-data-set*;
RUN;

p103d02

Viewing the Output

Partial PROC PRINT Output

Obs	First_Name	Last_Name	Job_Title	Salary
1	Satyakam	Denny	Sales Rep. II	26780
2	Monica	Kletschkus	Sales Rep. IV	30890
3	Kevin	Lyon	Sales Rep. I	26955
4	Petrea	Soltau	Sales Rep. II	27440
5	Marina	Iyengar	Sales Rep. III	29715

SAS Variable Names

SAS variable names

- can be 1 to 32 characters long.
- must start with a letter or underscore. Subsequent characters can be letters, underscores, or numerals.
- can be uppercase, lowercase, or mixed case.
- are not case sensitive.

Salary

score2

cust_ID

month1

FirstName

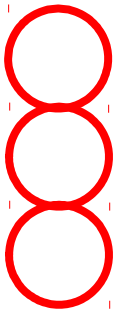
3.02 Multiple Answer Poll

Which variable names are invalid?

- data5mon
- 5monthsdata
- data#5
- five months data
- five_months_data
- FiveMonthsData
- fivemonthsdata

3.02 Multiple Answer Poll – Correct Answer

Which variable names are invalid?



- data5mon
- 5monthsdata
- data#5
- five months data
- five_months_data
- FiveMonthsData
- fivemonthsdata

Data Types

A SAS data set supports two types of variables.

Character variables

- can contain any value: letters, numerals, special characters, and blanks
- range from 1 to 32,767 characters in length
- have 1 byte per character.

Numeric variables

- store numeric values using floating point or binary representation
- have 8 bytes of storage by default
- can store 16 or 17 significant digits.

Missing Data Values

Missing values are valid values in a SAS data set.

Partial **work.newsalesemps**

First_Name	Last_Name	Job_Title	Salary
Monica	Kletschkus	Sales Rep. IV	.
Kevin	Lyon	Sales Rep. I	26955
Petrea	Soltau		27440

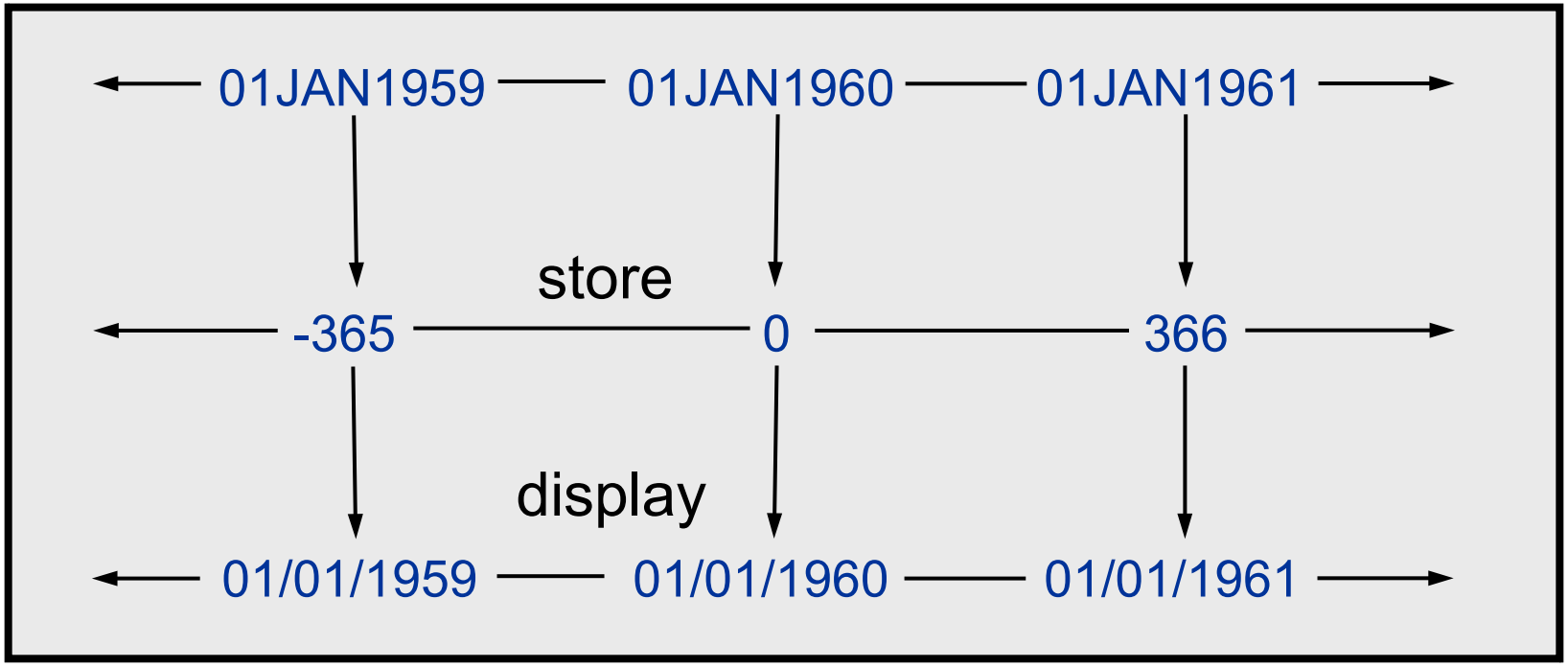
A blank represents a missing character value.

A period represents a missing numeric value.

- A value must exist for every variable in every observation.

SAS Date Values

SAS stores calendar dates as numeric values.



A *SAS date value* is stored as the number of days between January 1, 1960, and a specific date.

3.03 Quiz

What is the numeric value for today's date?

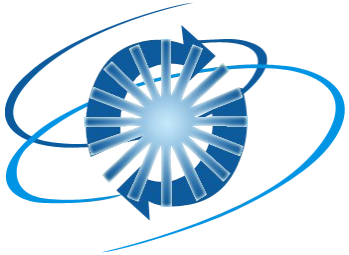
- Submit program **p103a02**.
- View the output to retrieve the current date as a SAS date value (that is, a numeric value referencing January 1, 1960).

3.03 Quiz – Correct Answer

What is the numeric value for today's date?

The answer depends on the current date.

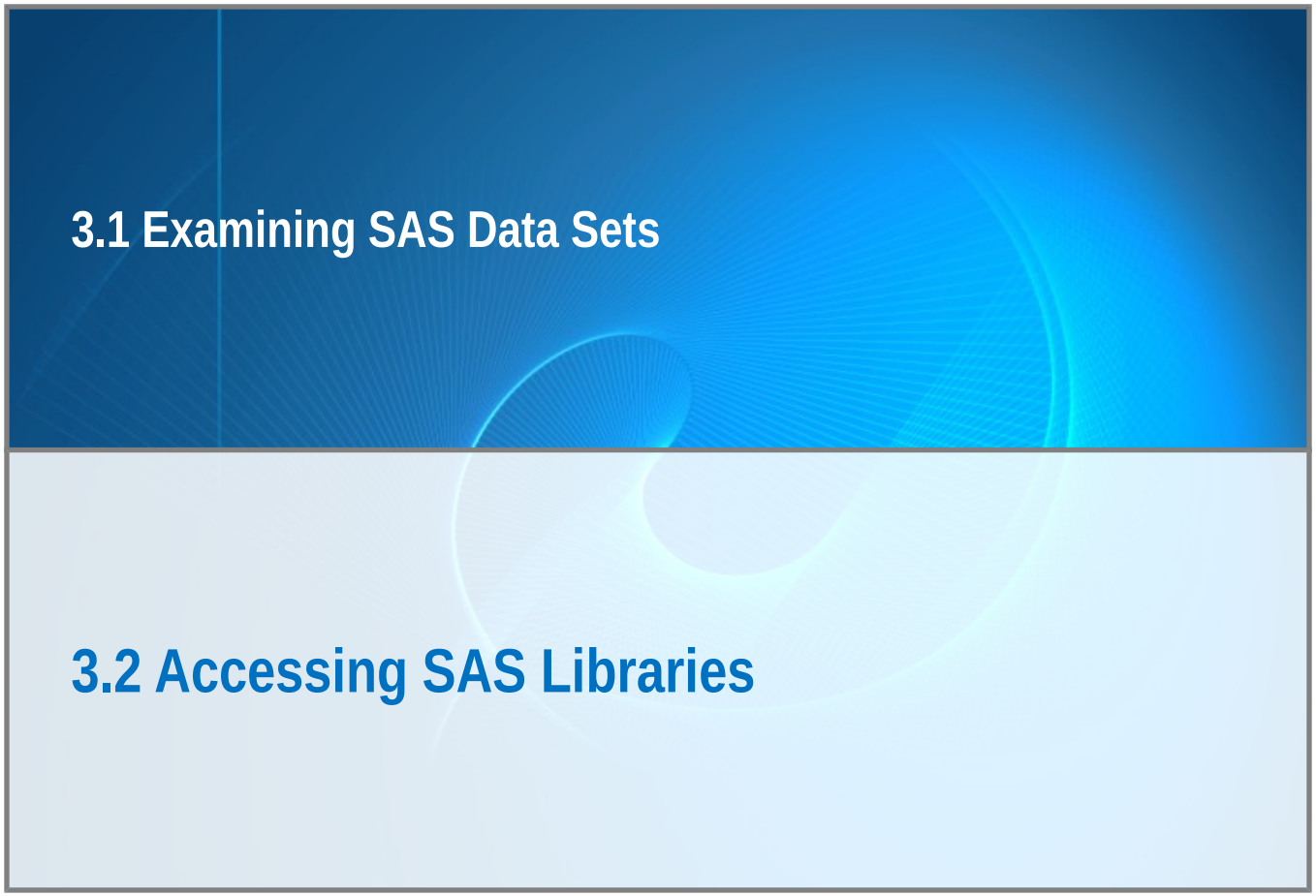
Example: If the current date is February 27, 2012, the numeric value is 19050.



Exercise

This exercise reinforces the concepts discussed previously.

Chapter 3: Accessing Data

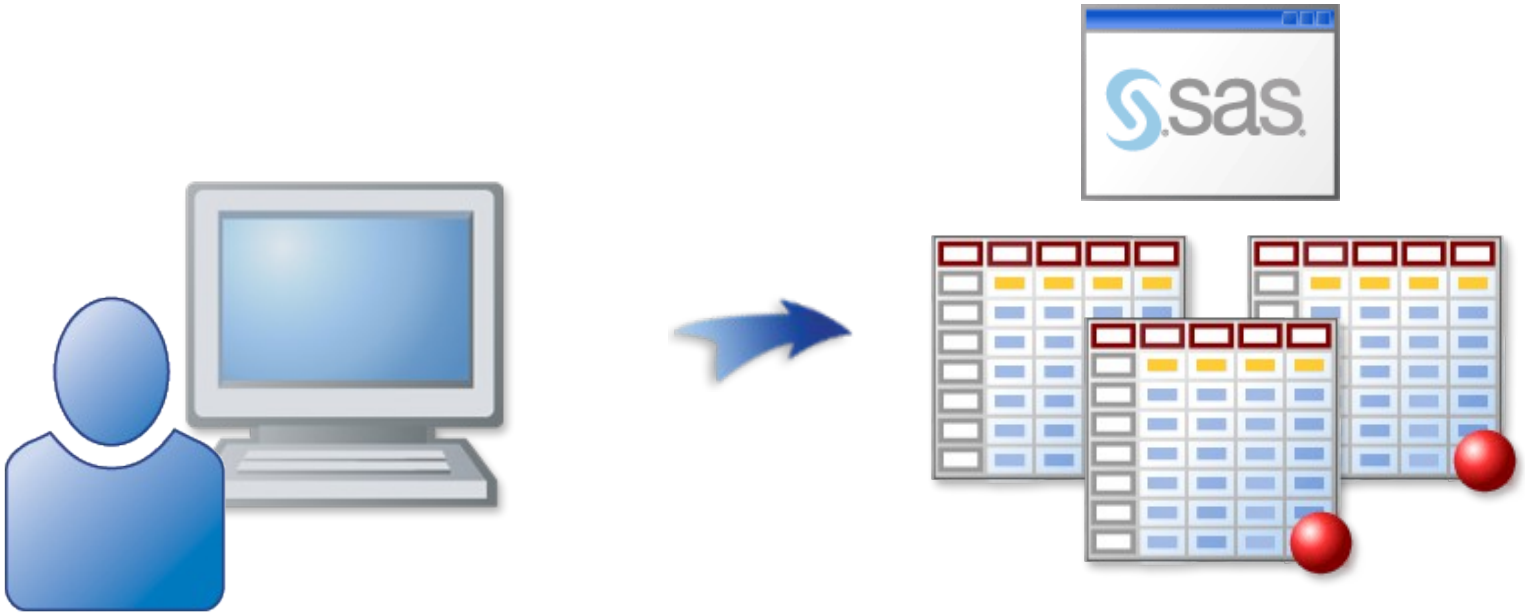


Objectives

- Explain the concept of a SAS library.
- State the difference between a temporary library and a permanent library.
- Assign a library reference name to a SAS library using a LIBNAME statement.
- Investigate a SAS library programmatically and interactively.
- Access a data set in a user-created permanent library.

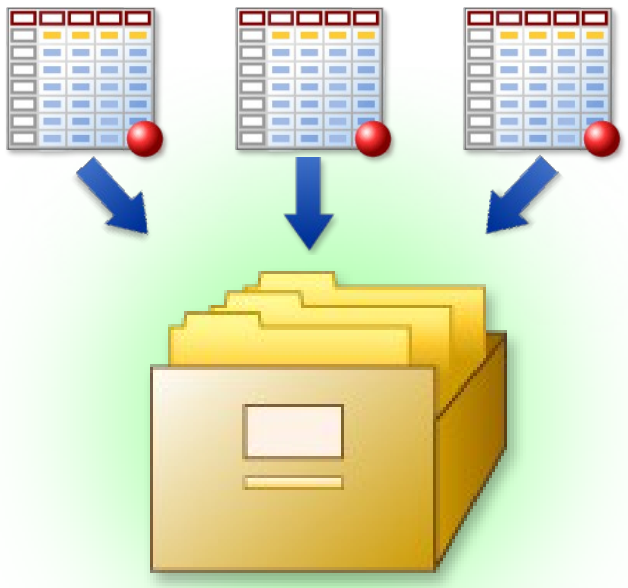
Business Scenario

Orion Star programmers need to access existing SAS data sets, so they need to understand how the data sets are stored in SAS.



SAS Libraries

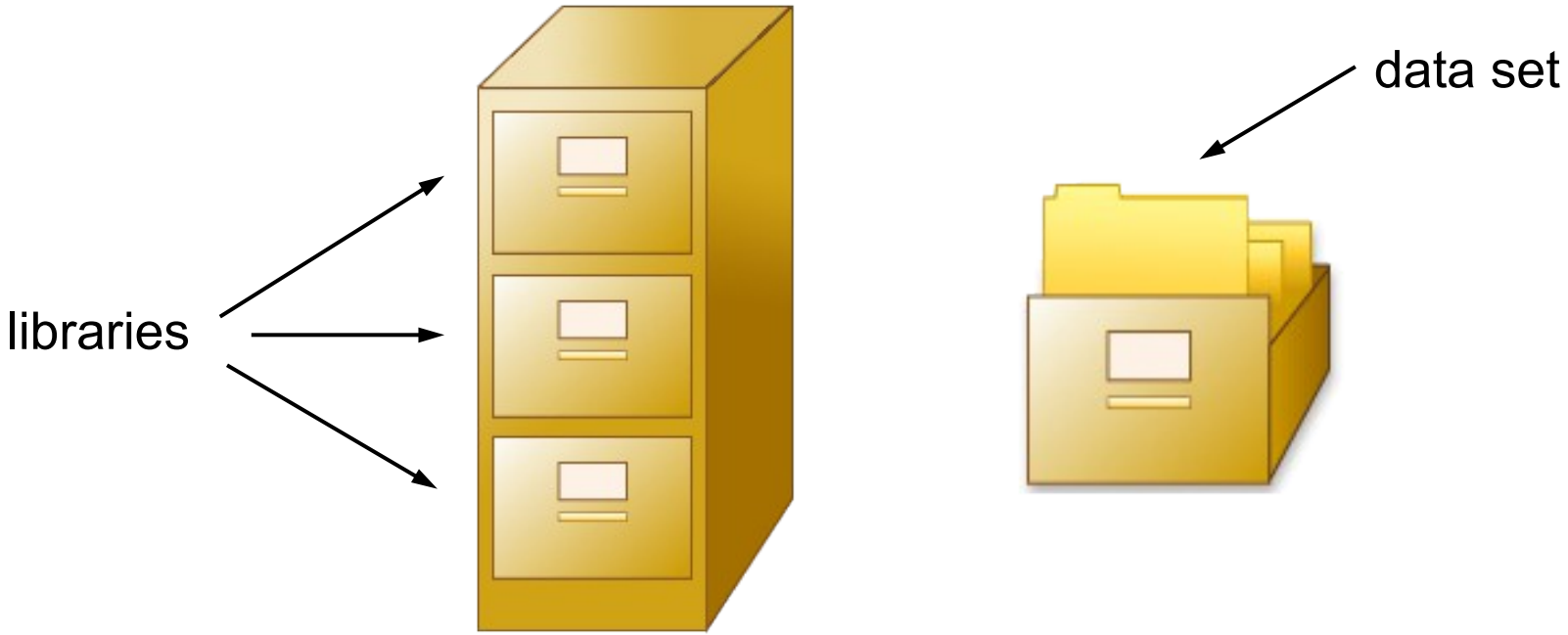
SAS data sets are stored in *SAS libraries*. A SAS library is a collection of SAS files that are referenced and stored as a unit.



Each file is a member of the library.

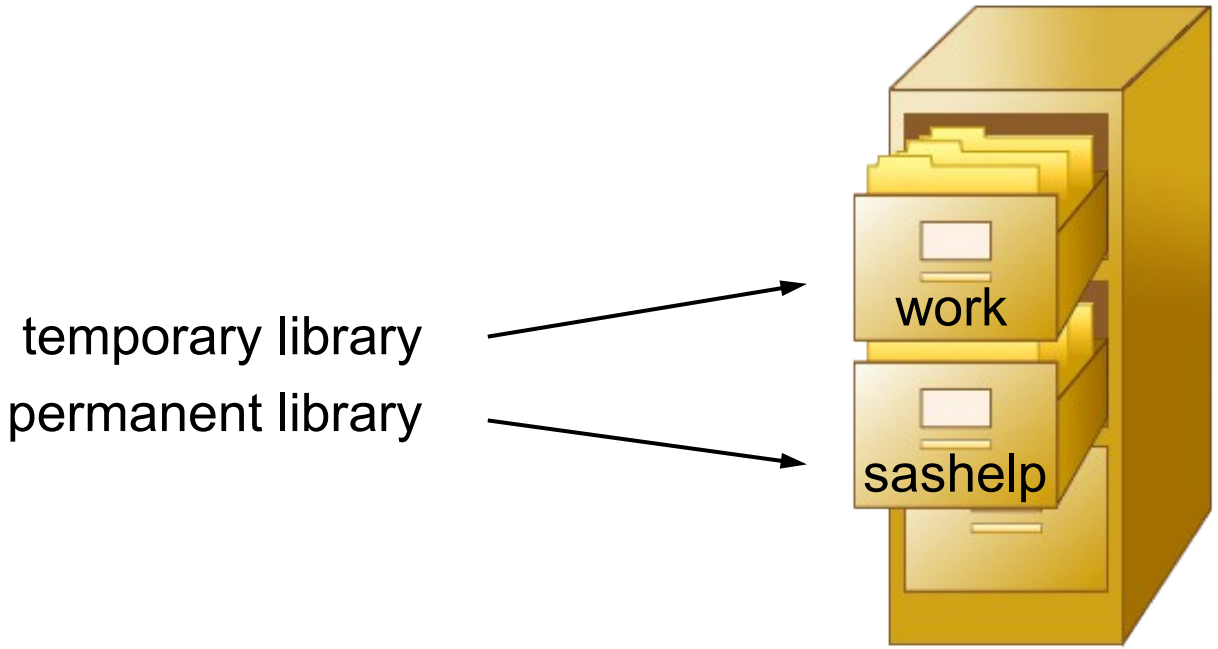
SAS Libraries

You can think of a SAS library as a drawer in a filing cabinet and a SAS data set as one of the files in the drawer.



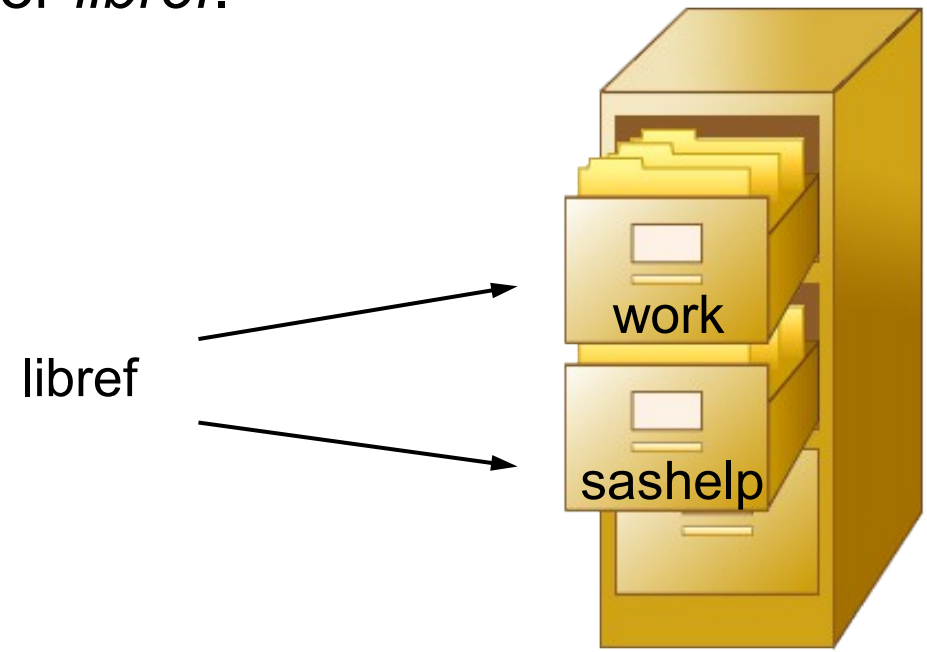
How SAS Libraries Are Defined

When a SAS session starts, SAS automatically creates one temporary and at least one permanent SAS library that you can access. These libraries are open and ready to be used.



Assigning a Libref

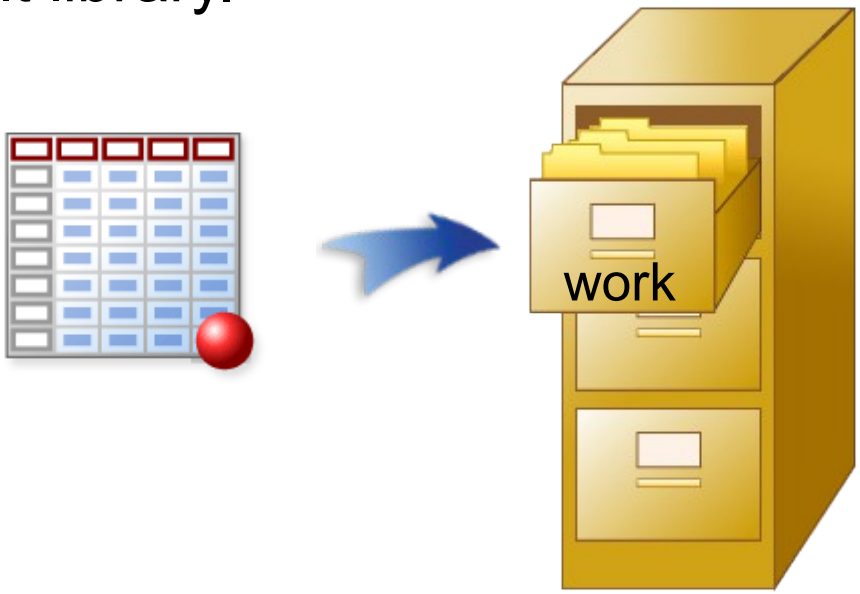
Regardless of the operating system that you use, you refer to a SAS library by a logical name called a library reference name, or *libref*.



A libref is a shortcut to the library.

Temporary Library

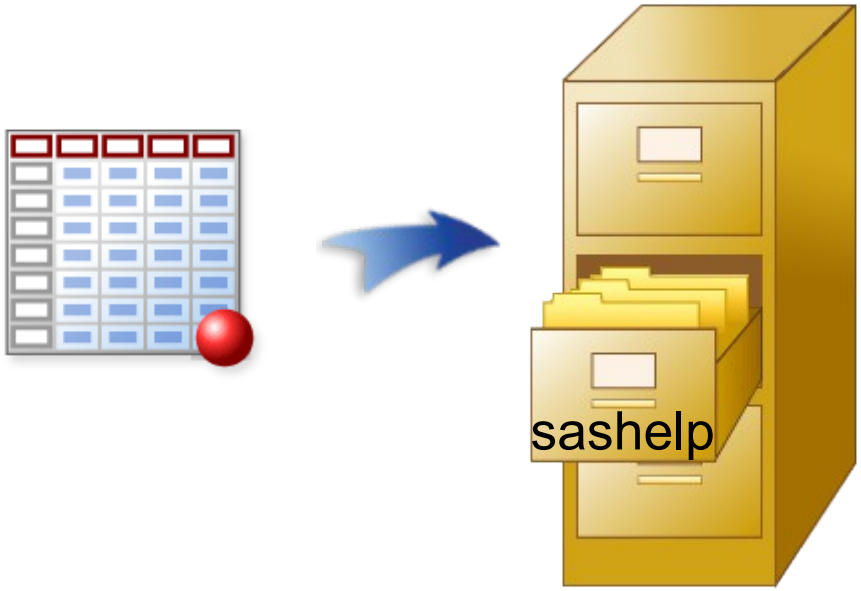
Work is a temporary library where you can store and access SAS data sets for the duration of the SAS session. It is the default library.



SAS deletes the **work** library and its contents when the session terminates.

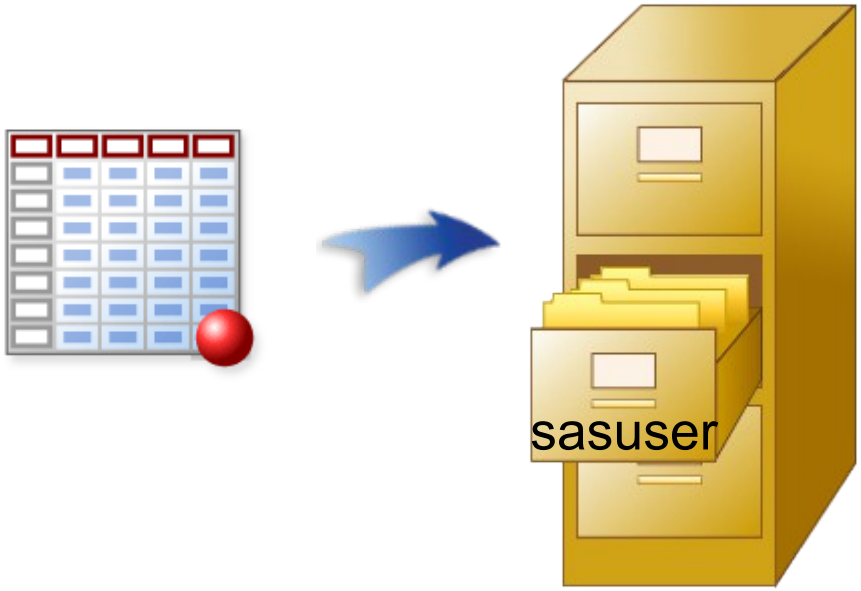
Permanent Libraries

Sashelp is a permanent library that contains sample SAS data sets you can access during your SAS session.



Permanent Libraries

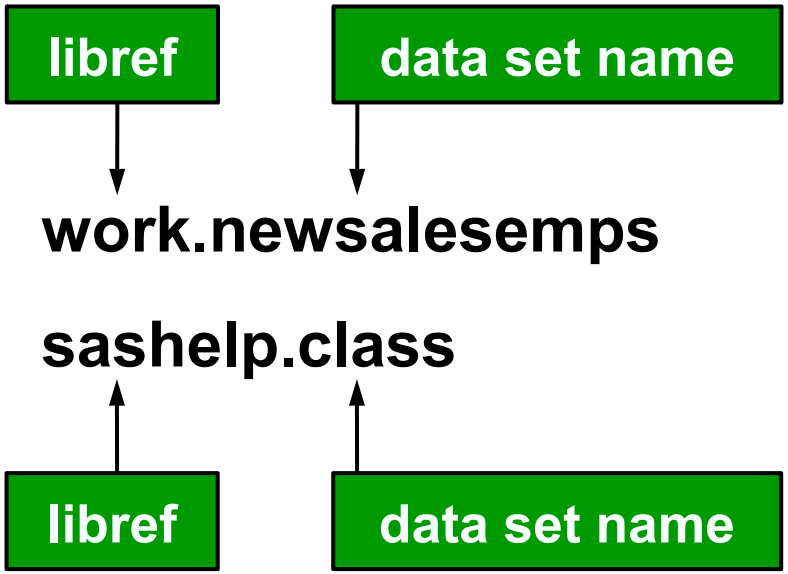
Sasuser is a permanent library that you can use to store and access SAS data sets in any SAS session.



SAS data sets in permanent libraries are saved after your SAS session terminates.

Accessing SAS Data Sets

All SAS data sets have a two-level name that consists of the libref and the data set name, separated by a period.

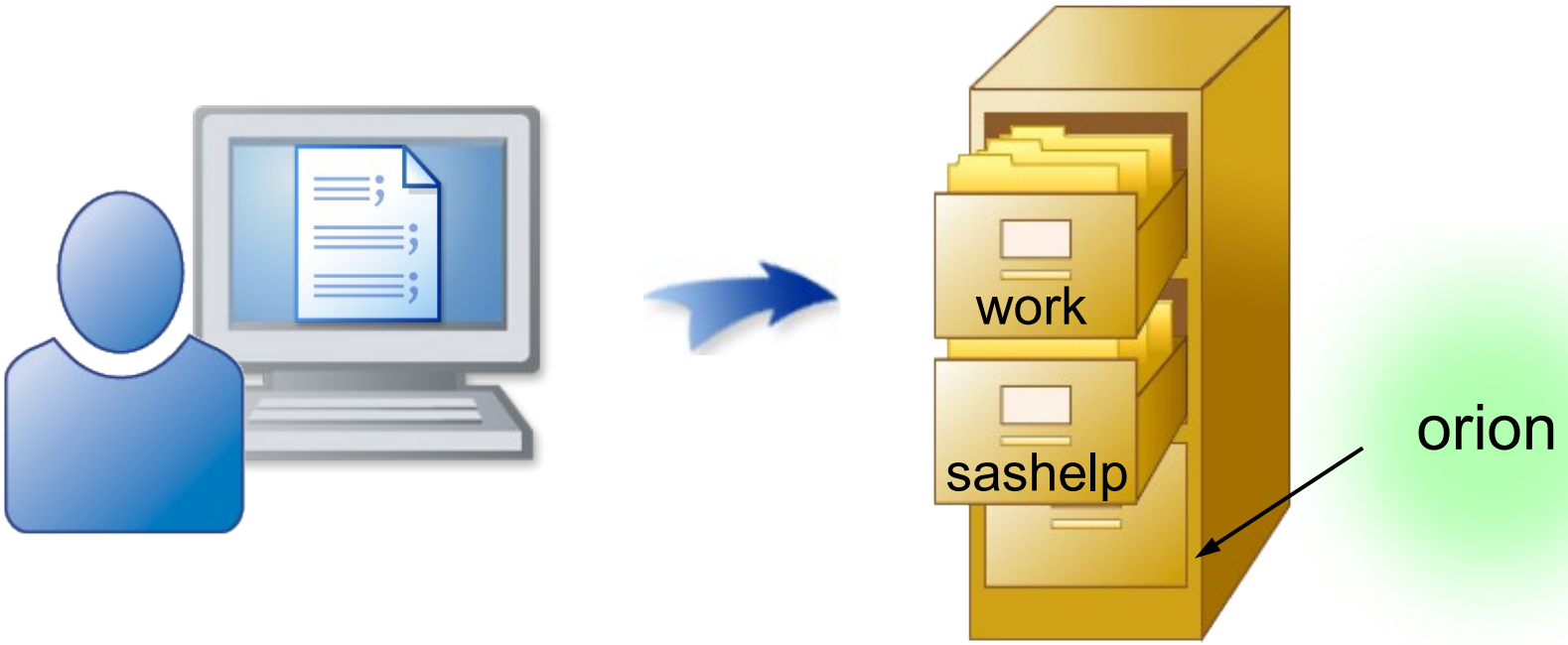


libref.data-set-name

When a data set is in the temporary **work** library, you can use a one-level name (for example, **newsalesemps)**.

Business Scenario

Orion Star programmers need to access and view SAS data sets that are stored in a permanent user-defined library.



User-Defined Libraries

Users can create their own SAS libraries. A user-defined library

- is permanent. Data sets are stored until the user deletes them.
- is implemented within the operating environment's file system.
- is not automatically available in a SAS session.

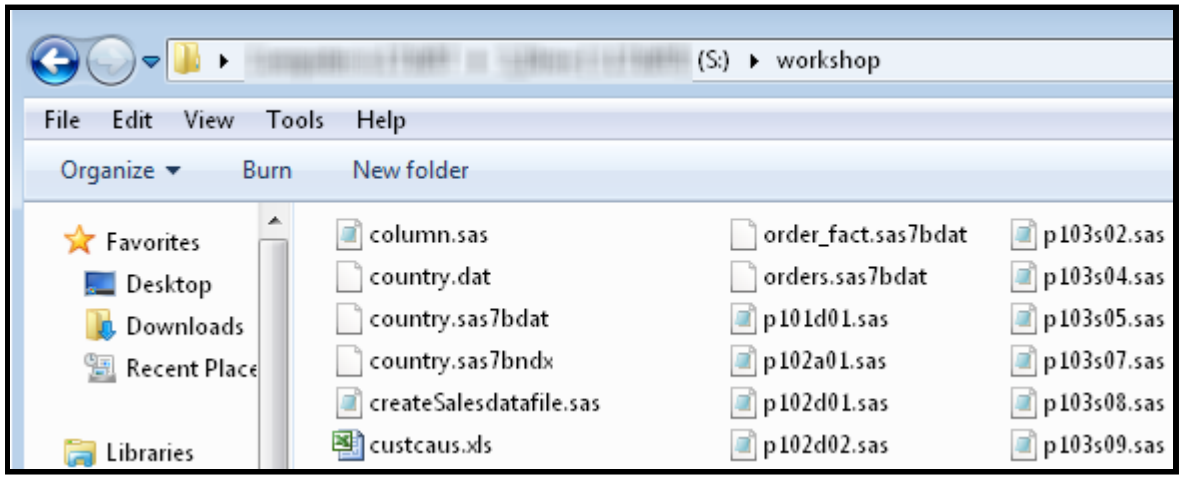
User-Defined Libraries

Operating Environment	A SAS library is…	Example
Microsoft Windows	A folder	s:\workshop
UNIX	A directory	~/workshop
z/OS (OS/390)	A sequential file	userid.workshop.sasdata

The user must assign a libref to the user-defined library to make it available in a SAS session.

Accessing a Permanent Library

Step 1 Identify the location of the library.

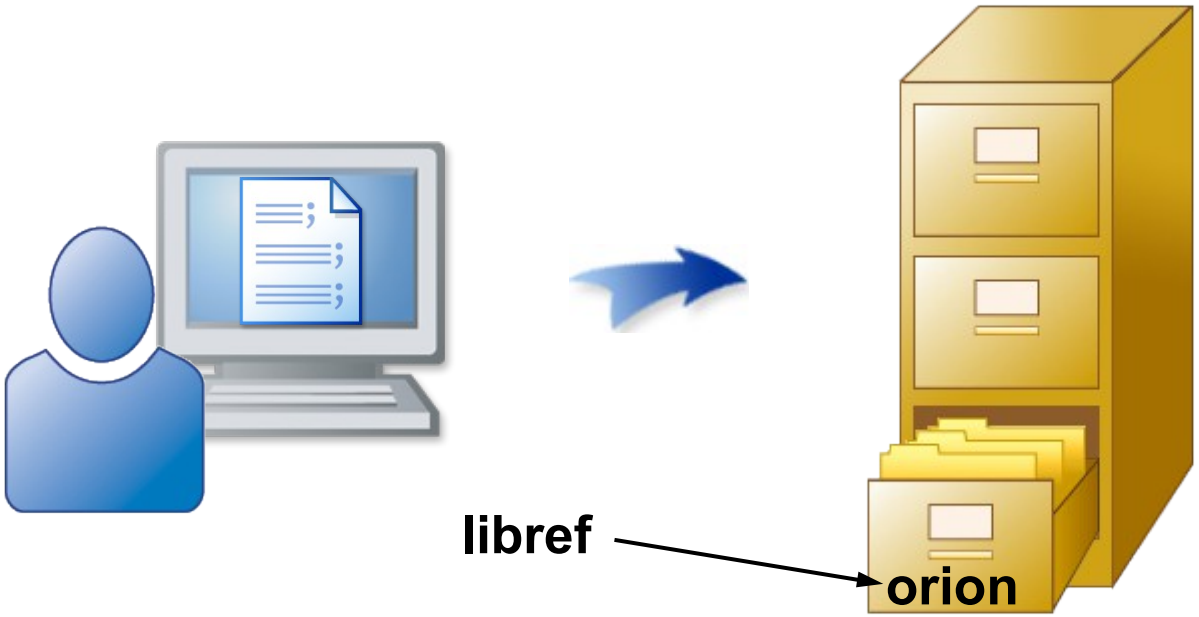


In this example, **s:\workshop**, a Microsoft Windows folder, is used as the SAS library.

Identify the location of ***your*** course data.

Accessing a Permanent Library

Step 2 Use a SAS LIBNAME statement to associate the libref with the physical location of the library.



Associate the libref **orion** with the Windows folder so that it is available to your SAS session.

LIBNAME Statement

The SAS LIBNAME statement is a *global* SAS statement.

```
libname orion "s:\workshop";
```

LIBNAME *libref* "*SAS-library*" <*options*>;

- It is not required to be in a DATA step or PROC step.
- It does not require a RUN statement.
- It executes immediately.
- It remains in effect until changed or canceled, or until the session ends.

Use the location of ***your*** course data in your LIBNAME statement.

Viewing the Log

Partial SAS Log

```
47   libname orion "s:\workshop";
NOTE: Libref ORION was successfully assigned as follows:
      Engine:        V9
      Physical Name: s:\workshop
```

3.04 Multiple Choice Poll

Which of the following correctly assigns the libref **myfiles** to a SAS library in the **c:\mysasfiles** folder?

- libname orion myfiles "c:\mysasfiles";
- libname myfiles "c:\mysasfiles";
- libref orion myfiles "c:\mysasfiles";
- libref myfiles "c:\mysasfiles";

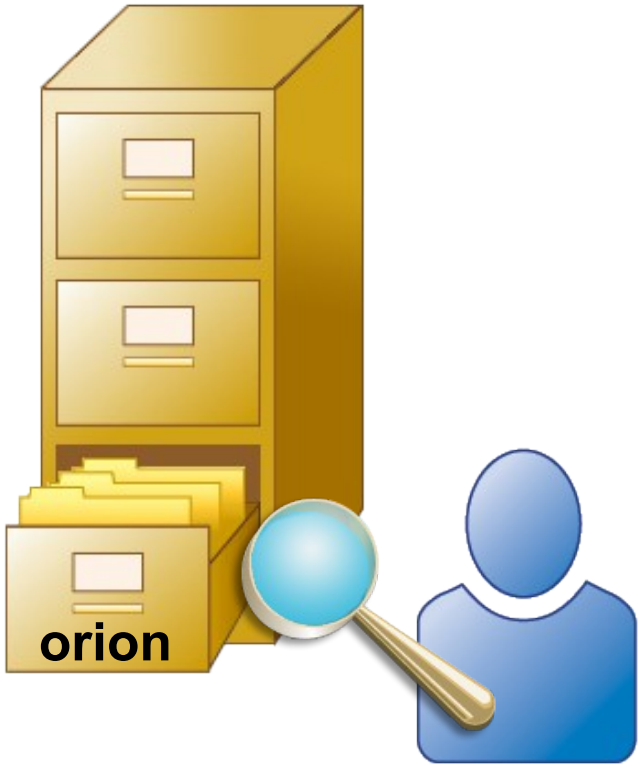
3.04 Multiple Choice Poll – Correct Answer

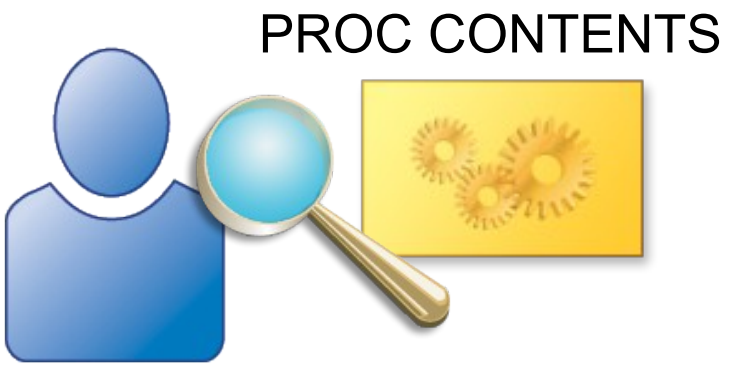
Which of the following correctly assigns the libref **myfiles** to a SAS library in the **c:\mysasfiles** folder?

- libname orion myfiles "c:\mysasfiles";
- libname myfiles "c:\mysasfiles";
- libref orion myfiles "c:\mysasfiles";
- libref myfiles "c:\mysasfiles";

Browsing a Library

Step 3 You can browse a library interactively in a SAS or SAS Enterprise Guide session, or programmatically using the CONTENTS procedure.





Browsing a Library Programmatically

Use PROC CONTENTS with the _ALL_ keyword to generate a list of all SAS files in a library.

```
proc contents data=orion._all_ nods;
run;
```

PROC CONTENTS DATA=*libref*._ALL_ NODS;
RUN;

- _ALL_ requests all files in the library.
- The NODS option suppresses the individual data set descriptor information.
- NODS can be used only with the keyword _ALL_.

Viewing the Output

Partial PROC CONTENTS Output

```
                The CONTENTS Procedure

                     Directory

                Libref        ORION
                Engine        V9
                Physical Name  S:\workshop
                Filename      S:\workshop


                  Member     File
#  Name              Type       Size  Last Modified

1  CHARITIES          DATA       9216  23Aug12:15:58:39
2  CONSULTANTS        DATA       5120  23Aug12:15:58:39
3  COUNTRY            DATA      17408  13Oct10:19:04:39
   COUNTRY            INDEX     17408  13Oct10:19:04:39
4  CUSTOMER           DATA      33792  04Nov11:09:52:27
5  CUSTOMER_DIM       DATA      33792  04Nov11:09:52:27
```

Accessing a Permanent Data Set

Step 4 After the libref is assigned, you can access SAS files in the library.

```
proc print data=orion.country;
run;
```

PROC PRINT Output

Obs	Country	Country_Name	Population	Country_ID	Continent_ID	Country_Former Name
1	AU	Australia	20,000,000	160	96	
2	CA	Canada	.	260	91	
3	DE	Germany	80,000,000	394	93	East/West Germany
4	IL	Israel	5,000,000	475	95	
5	TR	Turkey	70,000,000	905	95	
6	US	United States	280,000,000	926	91	
7	ZA	South Africa	43,000,000	801	94	

p103d04

Viewing the Log

Partial SAS Log

```
25   proc print data=orion.country;
26   run;

NOTE: There were 7 observations read from the data set
ORION.COUNTRY.
```

The libref **orion** remains in effect until you change or cancel it, or until you end your SAS session.

Changing or Canceling a Libref

To change a libref, submit a LIBNAME statement with the same libref but a different path.

```
libname orion "c:\myfiles";
```

To cancel a libref, submit a LIBNAME statement with the CLEAR option.

```
libname orion clear;
```

Browsing SAS Libraries: SAS Enterprise Guide

This demonstration illustrates defining and accessing a SAS library using SAS Enterprise Guide.

libname
p103d03

3.05 Poll

The library display in the Server List updates immediately when a libref is assigned or cleared using SAS Enterprise Guide.

- True
- False

3.05 Poll – Correct Answer

The library display in the Server List updates immediately when a libref is assigned or cleared using SAS Enterprise Guide.

- True
- False

The library display might or might not update immediately. Click Libraries → Refresh to update the list.

Browsing SAS Libraries: SAS Windowing Environment

This demonstration illustrates defining and accessing a SAS library using the SAS windowing environment.

libname
p103d03

3.06 Multiple Answer Poll

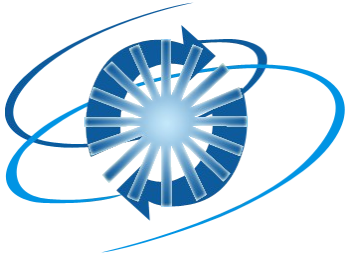
If you end your SAS or Enterprise Guide session and then need to access an **orion** data set again, which of the following must you do?

- start a new SAS or Enterprise Guide session
- use Windows Explorer to locate the files
- create a new folder to be used as a SAS library
- submit a LIBNAME statement to define a libref
- use a two-part name to access the data set

3.06 Multiple Answer Poll – Correct Answers

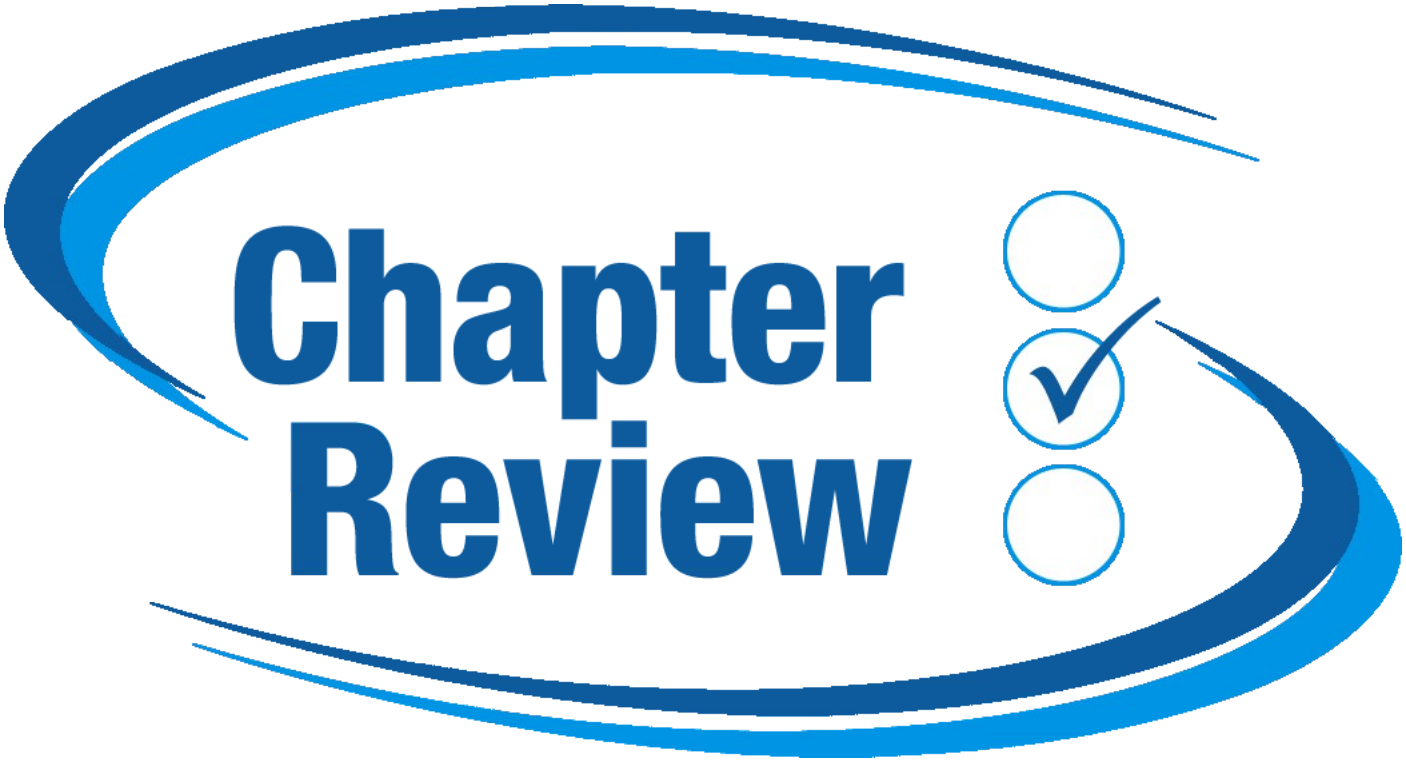
If you end your SAS or Enterprise Guide session and then need to access an **orion** data set again, which of the following must you do?

- start a new SAS or Enterprise Guide session
- use Windows Explorer to locate the files
- create a new folder to be used as a SAS library
- submit a LIBNAME statement to define a libref
- use a two-part name to access the data set



Exercise

This exercise reinforces the concepts discussed previously.



1. In which portion of a SAS data set are the following found?
   - the name of the data set
   - the type of the variable **Salary**
   - the creation date of the data set

   a. descriptor portion

   b. data portion

1. In which portion of a SAS data set are the following found?
   - the name of the data set
   - the type of the variable **Salary**
   - the creation date of the data set

   a. descriptor portion

   b. data portion

2. In this PROC CONTENTS output, what is the default length of the variable **Month**?

a. 2 bytes

b. 8 bytes

c. 16 or 17 bytes

d. 32,767 bytes

Alphabetic List of Variables and Attributes

#	Variable	Type	Len
1	Month	Num	?

2. In this PROC CONTENTS output, what is the default length of the variable **Month**?

a. 2 bytes

b. 8 bytes

c. 16 or 17 bytes

d. 32,767 bytes

Alphabetic List of Variables and Attributes

#	Variable	Type	Len
1	Month	Num	?

3. Which LIBNAME statement has the correct syntax?

   a. libname reports 's:\workshop';

   b. libname orion s:\workshop;

   c. libname 3456a 's:\workshop';

3. Which LIBNAME statement has the correct syntax?

a. libname reports 's:\workshop';

b. libname orion s:\workshop;

c. libname 3456a 's:\workshop';

4. Which PROC step successfully prints a list of all data sets in the **orion** library without printing descriptor portions for the individual data sets?

   - proc contents data=orion.nods _all_;
     run;
   - proc contents data=orion._all_ nods;
     run;
   - proc print data=orion._all_ noobs;
     run;
   - proc print data=orion._all_ nods;
     run;

4. Which PROC step successfully prints a list of all data sets in the **orion** library without printing descriptor portions for the individual data sets?

   - proc contents data=orion.nods _all_;
     run;
   - proc contents data=orion._all_ nods;
     run;
   - proc print data=orion._all_ noobs;
     run;
   - proc print data=orion._all_ nods;
     run;

5. In this data set, what type of variable is **Employee_ID**?

   - character
   - numeric
   - temporary
   - missing

Obs	Employee_ID	Last	Salary
1	.	Ralston	29250
2	120101	Lu	163040
3	120104	Billington	46230
4	120105	Povey	27110
5	120106	Hornsey	.

5. In this data set, what type of variable is **Employee_ID**?

   - character
   - numeric
   - temporary
   - missing

Obs	Employee_ID	Last	Salary
1	.	Ralston	29250
2	120101	Lu	163040
3	120104	Billington	46230
4	120105	Povey	27110
5	120106	Hornsey	.

6. What type of data set is the input data set in this PROC PRINT step?

```
proc print data=order_fact;
run;
```

- temporary
- permanent
- There is not enough information to determine the type.

6. What type of data set is the input data set in this PROC PRINT step?

```
proc print data=order_fact;
run;
```

- temporary
- permanent
- There is not enough information to determine the type.

7. A numeric variable can store numeric values with a maximum of eight digits.

   - True
   - False

7. A numeric variable can store numeric values with a maximum of eight digits.

- True
- False

8. Which of the following is not true of SAS date values?

- They are numeric.
- They can be positive or negative values.
- They represent the number of days between the day being stored and a base date.
- The base date is January 1, 1900.

8. Which of the following is not true of SAS date values?

- They are numeric.
- They can be positive or negative values.
- They represent the number of days between the day being stored and a base date.
- The base date is January 1, 1900.

9. Which statement about SAS libraries is true?

   - You refer to a SAS library by a logical name called a libname.
   - A SAS library is a collection of one or more SAS files that are referenced and stored as a unit.
   - A single SAS library can contain files that are stored in different physical locations.
   - At the end of each session, SAS deletes the contents of all SAS libraries.

9. Which statement about SAS libraries is true?

   - You refer to a SAS library by a logical name called a libname.
   - A SAS library is a collection of one or more SAS files that are referenced and stored as a unit.
   - A single SAS library can contain files that are stored in different physical locations.
   - At the end of each session, SAS deletes the contents of all SAS libraries.

10. Which of the following librefs is valid?

a. orionstar

b. orion/01

c. or_01

d. 1_or_a

10. Which of the following librefs is valid?

a. orionstar

b. orion/01

c. or_01

d. 1_or_a