The DATA Step

Tyler Cole

SECTIONS

Datasets

The DATA Step

SAS Techniques

Summary Points

Datasets

DATASETS

Describes and physically stores data

- Descriptor portion
  - Metadata about the contents of the dataset
  - Supplies dataset attributes and variable attributes

- Data portion
  - Physical values organized into observations and variables

DATASETS

descriptor component

descriptor information

data component

	PT	**WGT**
1	001	176
2	002	187

DATASETS

The CONTENTS Procedure

Data Set Name	WORK.ONE	Observations	2
Member Type	DATA	Variables	2
Engine	V9	Indexes	0
Created	03/27/2017 21:54:04	Observation Length	16
Last Modified	03/27/2017 21:54:04	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

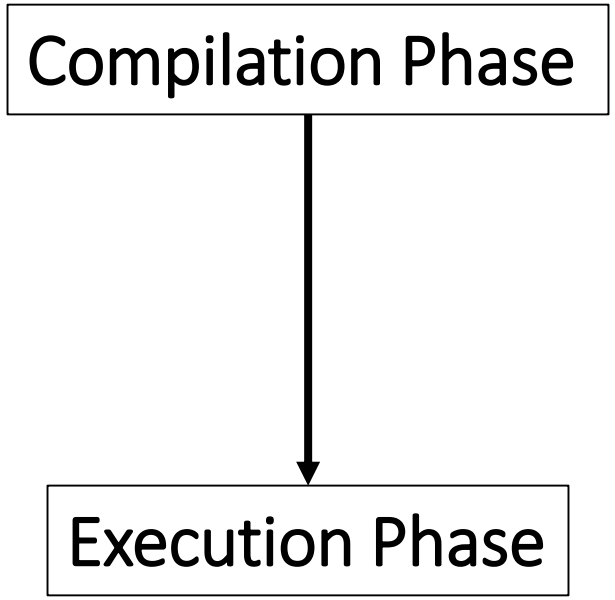
Engine/Host Dependent Information	
Data Set Page Size	65536
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	4061
Obs in First Data Page	2
Number of Data Set Repairs	0
Filename	/tmp/SAS_workDC7700000B99_localhost.localdomain/SAS_workC9D500000B99_localhost.localdomain/one.sas7bdat
Release Created	9.0401M4
Host Created	Linux
Inode Number	147846
Access Permission	rw-rw-r--
Owner Name	sasdemo
File Size	128KB
File Size (bytes)	131072

Variables in Creation Order

#	Variable	Type	Len	Label
1	pt	Char	5	Subject Identifier
2	wgt	Num	8	Weight in Lbs

The DATA Step

THE DATA STEP



THE DATA STEP

- Compilation
  - Build descriptor portion of the new dataset
  - Create the Program Data Vector

- Execution
  - Load values into the Program Data Vector
  - Output observations to the new dataset

THE DATA STEP

SET Statement

one.sas7bdat

PT	WGT
001	176
002	187

THE DATA STEP

```
data final;
set one;
study='Z999';
run;
```

Program Data Vector

_N_	PT[R]	WGT[R]	STUDY

R = retained

THE DATA STEP

```
data final;
set one;
study='Z999';
run;
```

Program Data Vector

_N_	PT[R]	WGT[R]	STUDY
1			

R = retained

THE DATA STEP

```
data final;
set one;
study='Z999';
run;
```

Program Data Vector

_N_	PT[R]	WGT[R]	STUDY
1	001	176	

R = retained

THE DATA STEP

```
data final;
set one;
study='Z999';
run;
```

Program Data Vector

_N_	PT[R]	WGT[R]	STUDY
1	001	176	Z999

R = retained

THE DATA STEP

```
data final;
set one;
study='Z999';
run;
```

Program Data Vector

_N_	PT[R]	WGT[R]	STUDY
1	001	176	Z999

PT	WGT	STUDY
001	176	Z999

R = retained

THE DATA STEP

```
data final;
set one;
study='Z999';
run;
```

Program Data Vector

_N_	PT[R]	WGT[R]	STUDY
2	001	176	

PT	WGT	STUDY
001	176	Z999

R = retained

THE DATA STEP

```
data final;
set one;
study='Z999';
run;
```

Program Data Vector

_N_	PT[R]	WGT[R]	STUDY
2	002	187	

PT	WGT	STUDY
001	176	Z999

R = retained

THE DATA STEP

```
data final;
set one;
study='Z999';
run;
```

Program Data Vector

_N_	PT[R]	WGT[R]	STUDY
2	002	187	Z999

PT	WGT	STUDY
001	176	Z999

R = retained

THE DATA STEP

```
data final;
set one;
study='Z999';
run;
```

Program Data Vector

_N_	PT[R]	WGT[R]	STUDY
2	002	187	Z999

PT	WGT	STUDY
001	176	Z999
002	187	Z999

R = retained

THE DATA STEP

```
data final;
set one;
study='Z999';
run;
```

Program Data Vector

_N_	PT[R]	WGT[R]	STUDY
3	002	187	

PT	WGT	STUDY
001	176	Z999
002	187	Z999

R = retained

THE DATA STEP

```
data final;
set one;
study='Z999';
run;
```

Program Data Vector

_N_	PT[R]	WGT[R]	STUDY
3	002	187	

PT	WGT	STUDY
001	176	Z999
002	187	Z999

R = retained

THE DATA STEP

MERGE Statement

one.sas7bdat

PT	WGT
001	176
002	187

many.sas7bdat

PT	VAL
001	11
002	21
002	22

THE DATA STEP

```
data final;
merge one many;
by pt;
wgt=wgt/2.2;
run;
```

Program Data Vector

_N_	PT^R	WGT^R	VAL^R

THE DATA STEP

```
data final;
merge one many;
by pt;
wgt=wgt/2.2;
run;
```

Program Data Vector

_N_	PT^R	WGT^R	VAL^R
1			

THE DATA STEP

```
data final;
```

- `merge one many;`
- `by pt;`

```
wgt=wgt/2.2;
run;
```

Program Data Vector

_N_	PT[R]	WGT[R]	VAL[R]
1	001	176	11

THE DATA STEP

```
data final;
merge one many;
by pt;
wgt=wgt/2.2;
run;
```

Program Data Vector

_N_	PT^{R}	WGT^{R}	VAL^{R}
1	001	80	11

THE DATA STEP

```
data final;
merge one many;
by pt;
wgt=wgt/2.2;
```

- `run;`

Program Data Vector

_N_	PT^R	WGT^R	VAL^R
1	001	80	11

PT	WGT	VAL
001	80	11

THE DATA STEP

```
data final;
merge one many;
by pt;
wgt=wgt/2.2;
run;
```

Program Data Vector

_N_	PT^R	WGT^R	VAL^R
2	001	80	11

PT	WGT	VAL
001	80	11

THE DATA STEP

```
data final;
merge one many;
by pt;
wgt=wgt/2.2;
run;
```

Program Data Vector

_N_	PT^{R}	WGT^{R}	VAL^{R}
2	002	187	21

PT	WGT	VAL
001	80	11

THE DATA STEP

```
data final;
merge one many;
by pt;
```

- `wgt=wgt/2.2;`
  `run;`

Program Data Vector

_N_	PT^R	WGT^R	VAL^R
2	002	85	21

PT	WGT	VAL
001	80	11

THE DATA STEP

```
data final;
merge one many;
by pt;
wgt=wgt/2.2;
```

- `run;`

Program Data Vector

_N_	PT^R	WGT^R	VAL^R
2	002	85	21

PT	WGT	VAL
001	80	11
002	85	21

THE DATA STEP

```
data final;
merge one many;
by pt;
wgt=wgt/2.2;
run;
```

Program Data Vector

_N_	PT^R	WGT^R	VAL^R
3	002	85	21

PT	WGT	VAL
001	80	11
002	85	21

THE DATA STEP

```
data final;
```

- `merge one many;`
- `by pt;`

```
wgt=wgt/2.2;
run;
```

Program Data Vector

_N_	PT^{R}	WGT^{R}	VAL^{R}
3	002	85	22

PT	WGT	VAL
001	80	11
002	85	21

THE DATA STEP

```
data final;
merge one many;
by pt;
wgt=wgt/2.2;
run;
```

Program Data Vector

_N_	PT^R	WGT^R	VAL^R
3	002	38.6	22

PT	WGT	VAL
001	80	11
002	85	21

THE DATA STEP

```
data final;
merge one many;
by pt;
wgt=wgt/2.2;
```

- `run;`

Program Data Vector

_N_	PT^R	WGT^R	VAL^R
3	002	38.6	22

PT	WGT	VAL
001	80	11
002	85	21
002	38.6	22

THE DATA STEP

```
data final;
merge one many;
by pt;
wgt=wgt/2.2;
run;
```

Program Data Vector

_N_	PT^R	WGT^R	VAL^R
4	002	38.6	22

PT	WGT	VAL
001	80	11
002	85	21
002	38.6	22

THE DATA STEP

```
data final;
merge one many;
by pt;
wgt=wgt/2.2;
run;
```

Program Data Vector

_N_	PT^R	WGT^R	VAL^R
4	002	38.6	22

PT	WGT	VAL
001	80	11
002	85	21
002	38.6	22

THE DATA STEP

Alternative approach

```
data final;
merge one many;
by pt;
wgt=wgt/2.2;
run;
```

THE DATA STEP

Alternative approach

```
data final;
merge one many;
by pt;
wgt_kg=wgt/2.2;
run;
```

THE DATA STEP

Alternative approach


```
data final;
merge one many;
by pt;
run;

data final;
   set final;
   wgt=wgt/2.2;
run;
```

SAS Techniques

SAS TECHNIQUES

Zero records

```
data final;
if nobs=0 then put 'No observations';
else set one nobs=nobs;
run;
```

SAS TECHNIQUES

Count of observations

```
data final;
if 0 then do;
   set one  nobs=nobs1;
   set many nobs=nobs2;
end;
run;
```

SAS TECHNIQUES

Copy descriptor information

```
data base;
if 0 then set final;
[…]
run;
proc append base=base data=final;
run;
```

SAS TECHNIQUES

Compute previous value

```
data final;
if last.pt=0 then prev_val=val;
set many;
by pt;
run;
```

Summary Points

SUMMARY POINTS

- SET *[…]* NOBS= value can be used immediately after compilation
- Variables entering via SET or MERGE are always retained
- Values may or may not be updated on variables entering via one-many MERGE