```
OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
                      *create a sample brands data table with brand names of between 1 & 5 words;
 52
53
54
55
56
57
58
                     data _null_;
x=datetime();;
put x datetime18.;
    05AUG16:18:07:08
Timestamp 08/05/2
Step Count
Page Faults
Page Reclaims
Page Swaps
Voluntary Context Switches
                                                                                 0
331
            Involuntary Context Switches
Block Input Operations
Block Output Operations
                     data brands (drop=i j);
length brand $ 60;
brand="Apple";output;
brand="Apples";output;
brand="Applese";output;
brand="Oco=Cola";output;
brand="Head & Shoulders";output;
 61
62
63
64
65
66
67
70
71
72
73
74
75
76
                     do i=0 to 5e6-1 by 1; *number of brands in data table will be 5 times this number; *change to 5E6 to get 25million individual brandnames;
                      do j=1 to 5;
brand=strip(brand)||' b'||strip(put(i+j, 10.));
brand=strip(brand);
output;
                       end;
                      run;
OS Memory 26536.00k
Timestamp 08/05/2016
Step Count
Page Faults
Page Reclaims
Page Swaps
Voluntary Context Switches
Involuntary Context Switches
Block Input Operations
Block Output Operations
                                                                                 167 Switch Count 94
                                                                                  359
                                                                                 2873
                                                                                 2937360
 78
79
80
81
82
83
84
85
86
87
88
89
90
91
                      *create the sample phrases data table with rows defined by the obs statement;
*1 in 3 phrases wil contain a brand name;
*phrases have 25 words + a brand name;
                     data phrases (drop=i brand);
set brands (keep=brand obs=400000);
length phrase $ 2000;
phrase="';
do i=1 to 25; *defines number of words in a phrase;
phrase=catx('', phrase, put(floor((1E8)*Rand("Uniform")), 12.));
if i=3 and (_N_<=5 or mod(_N_,3)=0) then
phrase=catx('', phrase, brand);
end;
                       end;
                       output;
 93
                       run;
Timestamp 08/05/201
Step Count
Page Faults
Page Reclaims
Page Swaps
Voluntary Context Switches
Involuntary Context Switches
Block Input Operations
Block Output Operations
                                                                                   724
                                                                                 0
393
5
48
94
95
96
97
98
99
                       *create a column to hold a copy of the original phrase, which will be editted;
                      data phrases_2;
                      set phrases;
length phrasecopy $ 2000;
phrasecopy=phrase;
phraseid=_N_;
 101
 102
 104
                       run;
NOTE: There were 400000 observations read from the data set WORK.PHRASES.

NOTE: The data set WORK.PHRASES 2 has 400000 observations and 3 variables.

NOTE: DATA statement used (Total process time):
real time 41.23 seconds
user cpu time 0.21 seconds
system cpu time 1.43 seconds
memory 3404.00k
```

```
OS Memory 28588.00k
Timestamp 08/05/201
Step Count
Page Faults
Page Reclaims
Page Swaps
Voluntary Context Switches
Involuntary Context Switches
Block Input Operations
Block Output Operations
                                                                                    28588.00k
08/05/2016 05:08:59 PM
169 Switch Count 48
                                                                                                                                     0
                                                                                                                                     3166
                                                                                                                                    1
88
3200016
 105
106
107
108
109
                                    *add a brand word count to the brand data set; data brands_2; set brands; brands; ' ');
 110
111
                                     run;
NOTE: There were 25000005 observations read from the data set WORK.BRANDS.
NOTE: The data set WORK.BRANDS_2 has 25000005 observations and 2 variables.
NOTE: DATA statement used (Total process time):
real time 48.20 seconds
user cpu time 3.16 seconds
system cpu time 1.65 seconds
memory 3404.81k
OS Memory 28588.00k
Timestamp 08/05/2016 05:09:47 PM
Step Count 170 Switch Count 36
                    Timestamp 08/05/2016 05
Step Count
Page Faults
Page Reclaims
Page Swaps
Voluntary Context Switches
Involuntary Context Switches
Block Input Operations
Block Output Operations
                                                                                                                                    170 Switch Count 36
                                                                                                                                     695
                                                                                                                                     0
4926
                                                                                                                                     3524368
                 *Need to know the maximum word count in brand name; proc sql noprint; select max(brandwc) into :brandmaxwc from brands_2; quit;

PROCEDURE SQL used (Total process time): real time 1.51 seconds user cpu time 1.51 seconds user cpu time 0.34 seconds memory 6775.46k

OS Memory 31916.00k
Timestamp 08/05/2016 05:09:49 PM
Step Count 171 Switch Count 24 Page Faults 040 Page Reclaims 240 Page Reclaims 240 Page Swaps 0 Voluntary Context Switches 40 Involuntary Context Switches 3 Block Input Operations 0 Block Output Operations 0
 117
                                   %put &brandmaxwc;
 118
                                   /*delete any existing 'matches' datatable*/
proc datasets lib=work nolist;
delete matches;
quit;
 119
120
 121
NOTE: Deleting WORK.MATCHES (memtype=DATA).

NOTE: PROCEDURE DATASETS used (Total process time):
real time 0.00 seconds
user cpu time 0.00 seconds
system cpu time 0.00 seconds
memory 344.96k
OS Memory 26024.00k
Timestamp 08/05/2016 05:09:49 PM
Step Count 172 Switch Count 30
Page Reclaims 12
Page Reclaims 12
Page Swaps 0
                    Page Faults
Page Reclaims
Page Swaps
Voluntary Context Switches
                    Involuntary Context Switches
Block Input Operations
Block Output Operations
 124
                                     %macro repeat;
                                     *do phrasepartwords=&brandmaxwc %to 1 %by -1;
*Then split phrases into word portions with &phrasepartwords words;
  125
                                data phraseslit (drop=i j wrds phrasewc);

set phrases_2 (drop=phrase);

length partphrase $ 200;

phrasewc=countw(phrasecopy, '');

wrds=sphrasepartwords;

do i=0 to phrasewc-wrds;

partphrase='';

do j=1 to wrds;

partphrase=catx('', partphrase, scan(phrasecopy, i+j, ''));

end;

output;

end;

run;
 126
 127
 129
 131
132
133
134
135
136
137
138
139
140
141
142
143
144
                                      run;
                                   proc sql ;
create table temp as
select p.phrasecopy, p.phraseid,b.brand
from phrasesplit p inner join brands_2 (where =(brandwc=&phrasepartwords)) b
on p.partphrase=b.brand
order by phraseid;
run:
 146
147
148
149
                                     run;
quit;
 150
 151
                                     data temp;
 152
                                     set temp;
phrasecopy=transtrn(phrasecopy, strip(brand), '');
  153
 154
  155
  156
```

157

```
proc append base=matches data=temp (drop=phrasecopy); *Update phrase_2 data with the matches.
This removes the found brand name from phrasecopy, ready for the next iteration; data phrases_2; merge phrases_2 temp (drop=brand); by phraseid; run;
 158
  159
160
 161
162
163
164
165
166
167
168
                                    data _null_;
x=datetime() ;;
put "Loop &phrasepartwords finished at " x datetime18.;
  170
171
                                     run;
%end;
 172
173
174
                                      %mend;
%repeat;
 3464.09k
28588.00k
08/05/2016 05:18:06 PM
                     memory
OS Memory
                    OS Memory 28588.00k
Timestamp 08/05/2016 05
Step Count
Page Faults
Page Reclaims
Page Swaps
Voluntary Context Switches
Involuntary Context Switches
Block Input Operations
Block Output Operations
                                                                                                                                     173 Switch Count 96
0
702
                                                                                                                                     69137
                                                                                                                                     38183448
 NOTE: Table WORK.TEMP created, with 26666 rows and 3 columns.
NOTE: PROC SQL statements are executed immediately; The RUN statement has no effect.

NOTE: PROCEDURE SQL used (Total process time):
    real time 7:13.64
    user cpu time 3:13.73
    system cpu time 2:30.10
    memory 130052.42k
    OS Memory 156492.00k
    Timestamp 08/05/2016 05:25:19 PM
    Step Count 174 Switch Count 177
    Page Faults 0
    Page Reclaims 6392
    Page Swaps 0
    Voluntary Context Switches 10911
    Involuntary Context Switches 207
    Block Input Operations 0
    Block Output Operations 38790216
NOTE: There were 26666 observations read from the data set WORK.TEMP.
NOTE: The data set WORK.TEMP has 26666 observations and 3 variables.
NOTE: DATA statement used (Total process time):
real time 0.94 seconds
user cpu time 0.04 seconds
system cpu time 0.05 seconds
memory 3447.46k
OS Memory 29356.00k
Timestamp 08/05/2016 05:25:20 PM
Step Count 175 Switch Count 20
Page Faults 0
Page Reclaims 738
Page Swaps 0
Voluntary Context Switches 36
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 108552
                   Appending WORK.TEMP to WORK.MATCHES.

BASE data set does not exist. DATA file is being copied to BASE file.
There were 26666 observations read from the data set WORK.TEMP.
The data set WORK.MATCHES has 26666 observations and 2 variables.
PROCEDURE APPEND used (Total process time):
real time 0.01 seconds
user cpu time 0.00 seconds
system cpu time 0.01 seconds
memory 3575.21k
OS Memory 29616.00k
Timestamp 08/05/2016 05:25:20 PM
Step Court 176 Switch Count 26
                    Timestamp 08/05/2016 05
Step Count
Page Faults
Page Reclaims
Page Swaps
Voluntary Context Switches
Involuntary Context Switches
Block Input Operations
Block Output Operations
                                                                                                                                    176 Switch Count 26
0
483
                                                                                                                                     0
50
                                                                                                                                     3856
```

```
user cpu time 0.00 secon system cpu time 0.00 secon memory 776.28k 0S Memory 26532.00k Timestamp 08/05/2010 Step Count Page Faults Page Reclaims Page Reclaims Page Swaps Voluntary Context Switches Involuntary Context Switches Block Input Operations Block Output Operations
                                                                                                                  0.00 seconds
0.00 seconds
776.28k
                                                                                                              0.00 SC.
776.28k
26532.00k
08/05/2016 05:25:59 PM
178 Switch Count 18
0
224
 NOTE: There were 400000 observations read from the data set WORK.PHRASES 2.

NOTE: The data set WORK.PHRASESPLIT has 9066672 observations and 3 variables.

NOTE: DATA statement used (Total process time):
real time 7:52.46

user cpu time 11.37 seconds
system cpu time 17.40 seconds
memory 3464.09k
OS Memory 29356.00k
Timestamp 08/05/2016 05:33:52 PM
Step Count 179 Switch Count 95
Page Faults 0
                          Timestamp 08/05/2016
Step Count
Page Faults
Page Reclaims
Page Swaps
Voluntary Context Switches
Involuntary Context Switches
Block Input Operations
Block Output Operations
                                                                                                                                                                                0
702
                                                                                                                                                                                0
55715
                                                                                                                                                                                39340568
  NOTE: Table WORK.TEMP created, with 26667 rows and 3 columns.
NOTE: PROC SQL statements are executed immediately; The RUN statement has no effect.

NOTE: PROCEDURE SQL used (Total process time):
    real time 7:42.30
    user cpu time 3:17.57
    system cpu time 2:40.16
    memory 130050.18k
    OS Memory 156492.00k
    Timestamp 08/05/2016 05:41:34 PM
    Step Count 180 Switch Count 202
    Page Faults 0
    Page Reclaims 6233
    Page Swaps 0
    Voluntary Context Switches 11933
    Involuntary Context Switches 224
    Block Input Operations 0
    Block Output Operations 39934536
NOTE: There were 26667 observations read from the data set WORK.TEMP.
NOTE: The data set WORK.TEMP has 26667 observations and 3 variables.
NOTE: DATA statement used (Total process time):
real time 1.07 seconds
user cpu time 0.04 seconds
system cpu time 0.06 seconds
memory 3447.46K
OS Memory 29356.00k
Timestamp 08/05/2016 05:41:35 PM
Step Count 181 Switch Count 20
Page Paults 0
Page Realaims 723
Page Swaps 0
Voluntary Context Switches 36
Involuntary Context Switches 0
Block Input Operations 8
Block Output Operations 108552
 NOTE: Appending WORK.TEMP to WORK.MATCHES.

NOTE: There were 26667 observations read from the data set WORK.TEMP.

NOTE: 126667 observations added.

NOTE: 16 data set WORK.MATCHES has 53333 observations and 2 variables.

NOTE: PROCEDURE APPEND used (Total process time):

real time 0.02 seconds

user cpu time 0.01 seconds

system cpu time 0.01 seconds

memory 3654.28k

OS Memory 29104.00k

Timestamp 08/05/2016 05:41:35 PM

Step Count 182 Switch Count 24

Page Faults 0
                          OS Memory 29104.00k
Timestamp 08/05/2016 05
Step Count
Page Faults
Page Reclaims
Page Swaps
Voluntary Context Switches
Involuntary Context Switches
Block Input Operations
Block Output Operations
                                                                                                                                                                                0
467
                                                                                                                                                                                0
41
                                                                                                                                                                                4368
Loop 4 finished at 05AUG16:18:42:18

NOTE: DATA statement used (Total process time):
real time 0.00 seconds
user cpu time 0.00 seconds
system cpu time 0.00 seconds
```

```
memory 776.28k
OS Memory 26532.00k
Timestamp 08/05/201
Step Count
Page Faults
Page Reclaims
Page Swaps
Voluntary Context Switches
Involuntary Context Switches
Block Input Operations
Block Output Operations
                                                                                26532.00k
08/05/2016 05:42:17 PM
                                                                                                                          184 Switch Count 18
0
222
NOTE: There were 400000 observations read from the data set WORK.PHRASES_2.

NOTE: The data set WORK.PHRASESPLIT has 9360004 observations and 3 variables.

NOTE: DATA statement used (Total process time):
real time 7:59.58

user cpu time 9.70 seconds
system cpu time 17.82 seconds
memory 3464.09k
OS Memory 29356.00k
Timestamp 08/05/2016 05:50:17 PM
Step Count 185 Switch Count 102
Page Faults 0
                  Timestamp 08/05/2016 05
Step Count
Page Faults
Page Reclaims
Page Reclaims
Voluntary Context Switches
Involuntary Context Switches
Block Input Operations
Block Output Operations
                                                                                                                          0
700
                                                                                                                          0
61086
                                                                                                                          24
632
                                                                                                                          40613400
NOTE: Table WORK.TEMP created, with 26667 rows and 3 columns.
NOTE: There were 26667 observations read from the data set WORK.TEMP.
NOTE: There were 26667 observations read from the data set WORK.TEMP NOTE: The data set WORK.TEMP has 26667 observations and 3 variables.

NOTE: DATA statement used (Total process time):
real time 0.88 seconds
user cpu time 0.04 seconds
system cpu time 0.07 seconds
memory 3447.46k
OS Memory 29356.00k
Timestamp 08/05/2016 05:58:18 EM
                  system opu time 0.07 seconds memory 3447.46k OS Memory 29356.00k Timestamp 08/05/2016 OS Step Count Page Faults Page Reclaims Page Reclaims Page Swaps Voluntary Context Switches Involuntary Context Switches Block Input Operations Block Output Operations
                                                                                                                         187 Switch Count 20
0
746
                                                                                                                         0
35
0
                                                                                                                          108552
                 Appending WORK.TEMP to WORK.MATCHES.
There were 26667 observations read from the data set WORK.TEMP.
26667 observations added.
The data set WORK.MATCHES has 80000 observations and 2 variables.
PROCEDURE APPEND used (Total process time):
real time 0.02 seconds
user cpu time 0.03 seconds
system cpu time 0.03 seconds
memory 3655.25k
OS Memory 29104.00k
Timestamp 08/05/2016 05:58:18 PM
Step Count 188 Switch Count 24
Page Faults 0
                   Step Count
Page Faults
Page Reclaims
Page Swaps
Voluntary Context Switches
Involuntary Context Switches
Block Input Operations
Block Output Operations
                                                                                                                          0
467
                                                                                                                          0
40
                                                                                                                          4368
Loop 3 finished at 05AUG16:18:59:08

NOTE: DATA statement used (Total process time):
    real time 0.00 seconds
    user cpu time 0.00 seconds
    system cpu time 0.00 seconds
    memory 776.28k
    OS Memory 26532.00k
```

```
Timestamp 08/05/2016
Step Count
Page Faults
Page Reclaims
Page Swaps
Voluntary Context Switches
Involuntary Context Switches
Block Input Operations
Block Output Operations
                                                                                 08/05/2016 05:59:08 PM
190 Switch Count 18
                                                                                                                                         224
NOTE: There were 400000 observations read from the data set WORK.PHRASES 2.

NOTE: The data set WORK.PHRASESPLIT has 9680003 observations and 3 variables.

NOTE: DATA statement used (Total process time):
real time 10:06.56

user cpu time 8.22 seconds
system cpu time 18.40 seconds
memory 3464.09k
OS Memory 29356.00k
Timestamp 08/05/2016 06:09:14 PM
Sten Count 109
                     Step Count
Page Faults
Page Reclaims
                                                                                                                                         191 Switch Count 109
                                                                                                                                         0
702
                    Page Reclaims
Page Swapon
Voluntary Context Switches
Involuntary Context Switches
Block Input Operations
Block Output Operations
                                                                                                                                         0
52227
                                                                                                                                         20
736
                                                                                                                                         42001688
 NOTE: Table WORK.TEMP created, with 26666 rows and 3 columns.
NOTE: There were 26666 observations read from the data set WORK.TEMP.
NOTE: The data set WORK.TEMP has 26666 observations and 3 variables.
NOTE: DATA statement used (Total process time):
real time 1.16 seconds
user cpu time 0.03 seconds
system cpu time 0.07 seconds
memory 3447.46k
OS Memory 29356.00k
Timestamp 08/05/2016 06:17:35 PM
                      Timestamp
Step Count
Page Faults
Page Reclaims
                                                                                                                                         193 Switch Count 20
0
721
                                                                                                                          72.
0
37
0
8
                    Page Reclaims
Page Swaps
Voluntary Context Switches
Involuntary Context Switches
Block Input Operations
Block Output Operations
                                                                                                                                         108552
NOTE: Appending WORK.TEMP to WORK.MATCHES.

NOTE: There were 26666 observations read from the data set WORK.TEMP.

NOTE: 26666 observations added.

NOTE: The data set WORK.MATCHES has 106666 observations and 2 variables.

NOTE: PROCEDURE APPEND used (Total process time):
real time 0.02 seconds
user cpu time 0.00 seconds
system cpu time 0.00 seconds
memory 3654.28K
OS Memory 29104.00K
Timestamp 08/05/2016 06:17:35 PM
Step Count 194 Switch Count 24
Page Faults 0
Page Reclaims 467
Page Swaps 0
Voluntary Context Switches 40
Involuntary Context Switches 0
                     Involuntary Context Switches
Involuntary Context Switches
Block Input Operations
Block Output Operations
                                                                                                                                         4112
NOTE: There were 400000 observations read from the data set WORK.PHRASES_2.

NOTE: The data set WORK.PHRASES_2 has 400000 observations and 3 variables.

NOTE: DATA statement used (Total process time):
real time 37.67 seconds
user cpu time 0.26 seconds
system cpu time 1.81 seconds
memory 5166.18k
OS Memory 31408.00k
Timestamp 08/05/2016 06:18:12 PM
Step Count 195 Switch Count 30
Page Faults 0
Page Reclaims 1018
Page Swaps 0
Voluntary Context Switches 2019
Involuntary Context Switches 3
Block Input Operations 72
Block Output Operations 3200016
Loop 2 finished at 05AUG16:19:18:13
NOTE: DATA statement used (Total process time):
real time 0.00 seconds
user cpu time 0.00 seconds
system cpu time 0.00 seconds
memory 776.28k
                     memory
OS Memory
Timestamp
                                                                                0.00 seconds
776.28k
26532.00k
08/05/2016 06:18:12 PM
                      Step Count
                                                                                                                                         196 Switch Count 18
```

```
Page Faults
Page Reclaims
Page Swaps
Voluntary Context Switches
Involuntary Context Switches
Block Input Operations
Block Output Operations
                                                                                                                                                                   222
   NOTE: There were 400000 observations read from the data set WORK.PHRASES_2.

NOTE: The data set WORK.PHRASESPLIT has 10026671 observations and 3 variables.

NOTE: DATA statement used (Total process time):
    real time 8:30,98
    user cpu time 6:71 seconds
    system cpu time 18:39 seconds
    memory 3464.09k

OS Memory 29356.00k
    Timestamp 08/05/2016 06:26:43 PM
    Step Count 197 Switch Count 108
    Page Faults 0
    Page Reclaims 700
    Page Swaps 0
                           Timestamp 08/05/2
Step Count
Page Faults
Page Reclaims
Page Swaps
Voluntary Context Switches
                                                                                                                                                                   0
64234
                          Involuntary Context Switches
Block Input Operations
Block Output Operations
                                                                                                                                                                   18
688
                                                                                                                                                                   43505944
   NOTE: Table WORK.TEMP created, with 26671 rows and 3 columns.
  NOTE: PROCE DURE SQL statements are executed immediately; The RUN statement has no effect.

NOTE: PROCEDURE SQL used (Total process time):
real time 9:12.74
user cpu time 3:41.45
system cpu time 2:52.14
memory 130052.67k
OS Memory 156492.00k
Timestamp 08/05/2016 06:35:56 PM
Step Count 198 Switch Count 270
Page Faults 0
Page Reclaims 6233
Page Swaps 0
Voluntary Context Switches 16966
Involuntary Context Switches 284
Block Input Operations 0
Block Output Operations 44089416
  NOTE: There were 26671 observations read from the data set WORK.TEMP.
NOTE: The data set WORK.TEMP has 26671 observations and 3 variables.
NOTE: DATA statement used (Total process time):
real time 1.08 seconds
user cpu time 0.04 seconds
system cpu time 0.06 seconds
memory 3447.46k
OS Memory 29356.00k
Timestamp 08/05/2016 06:35:57 PM
Step Count 199 Switch Count 20
Page Faults 0
                          OS Memory 29356.00k
Timestamp 08/05/2016 06
Step Count
Page Faults
Page Reclaims
Page Swaps
Voluntary Context Switches
Involuntary Context Switches
Block Input Operations
Block Output Operations
                                                                                                                                                                   0
745
                                                                                                                                                                  0
                                                                                                                                                                   108552
OS Memory 29104.00x
Timestamp 08/05/2016 06
Step Count
Page Faults
Page Reclaims
Page Swaps
Voluntary Context Switches
Involuntary Context Switches
Block Input Operations
Block Output Operations
                                                                                                                                                                   0
40
                                                                                                                                                                   4368
  NOTE: There were 400000 observations read from the data set WORK.PHRASES 2.

NOTE: There were 26671 observations read from the data set WORK.PHRASES 2.

NOTE: The data set WORK.PHRASES 2 has 400000 observations and 3 variables.

NOTE: DATA statement used (Total process time):
real time 41.20 seconds
user cpu time 0.24 seconds
system cpu time 1.79 seconds
memory 5166.15k
OS Memory 31408.00k
Timestamp 08/05/2016 06:36:39 PM
Step Count 201 Switch Count 30
Page Faults 0
Page Faults 0
Page Reclaims 1018
Page Swaps 0
Voluntary Context Switches 1229
Involuntary Context Switches 2
Block Input Operations 96
Block Output Operations 96
Block Output Operations 3200016
  Loop 1 finished at 05AUG16:19:36:39
NOTE: DATA statement used (Total process time):
    real time 0.00 seconds
    user cpu time 0.00 seconds
    system cpu time 0.00 seconds
    memory 776.28k
    OS Memory 26532.00k
                           memory
OS Memory
Timestamp
                                                                                                         08/05/2016 06:36:39 PM
                           Step Count
Page Faults
Page Reclaims
                                                                                                                                                          202 Switch Count 18
                                                                                                                                                                   224
```