

# Biegnij SaSie,<sup>®</sup> biegnij

czyli kilka trików  
na przyspieszenie pracy w SAS

Bartosz Jabłoński

Wydział Matematyki i Nauk Informatycznych (PW) / Citi Europe PLC

✉ yabwon@gmail.com

19-20 listopada 2018

SAS dla Administratorów i Praktyków, Warszawa

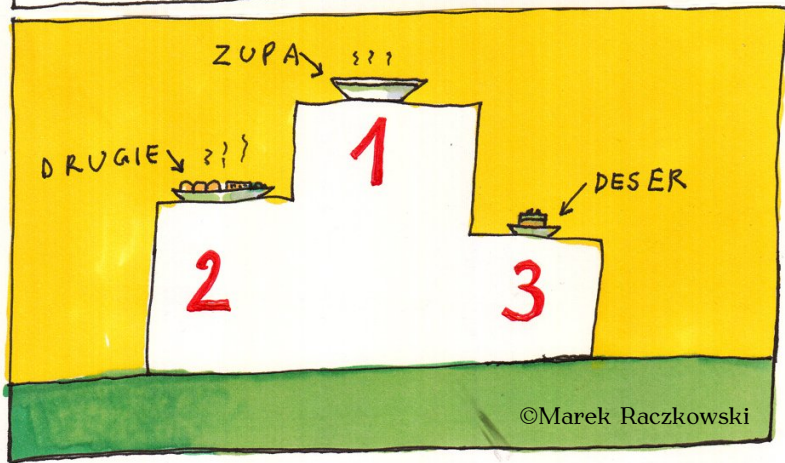


<https://www.linkedin.com/groups/8675451>

- Transmisje video z SAS Global Forum (np. <https://video.sas.com/sasgf17>),
- "Proceedingsy SASowe" na <https://www.lexjansen.com/>,
- Lista dyskusyjna SAS-L na serwerze <https://listserv.uga.edu>,
- Interaktywny quiz SASowy: [www.sasensei.com](http://www.sasensei.com),
- ...oraz moi studenci (<3).

Eleganckie podsumowanie on-line:  
<http://slides.com/allanbowe/find-your-sas-sensei#/>

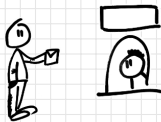
# WYNIKI KONKURSU KULINARNEGO



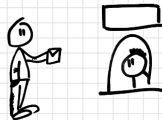
- Q1. Jak zrównoleglać/przyspieszać przetwarzania z pomocą polecenia SYSCALL oraz PROC CATALOG?
- Q2. Jak wykorzystać PROC FCMP i funkcje użytkownika oraz %MACRO'a do przekazywania danych w HASH tablicach pomiędzy DATA STEPami?
- Q3. Jak mając do wykonania klauzulę WHERE "typu" OR (na różnych zmiennych) nadal móc wykorzystywać zalety INDEXów?

## Zagadnienie pierwsze

Jak zrównoleglać/przyspieszać przetwarzania z pomocą polecenia SYSCALL oraz PROC CATALOG?



```
1
2 data tmp1;
3   set _BoB_have_;
4   where "some conditions for BoB";
5   ...
6 run;
7
8 proc sql;
9   create table tmp2 as
10  select ...
11    from tmp1
12   join "with some data for BoB"
13    on ...
14  ;
15 quit;
16
17 proc univariate data = tmp2 ...
18   ...
19   output out = _BoB_want_ / "some stats for BoB" ...;
20 quit;
21
```



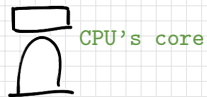
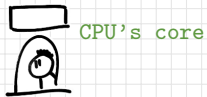
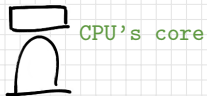


```

1 %do Iterator = 1 %to 1000000;
2   data tmp1;
3     set _&Iterator._have_;
4     where "some conditions for &Iterator.";
5     ...
6   run;
7
8   proc sql;
9     create table tmp2 as
10    select * from tmp1
11    where tmp1.some_condition = 1;
12    on ...
13    ;
14  quit;
15
16
17  proc univariate data = tmp2 ...
18    ...
19    output out = _&Iterator._want_ / "some stats for &Iterator." ...;
20  quit;
21 %end;

```







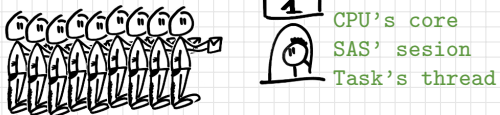
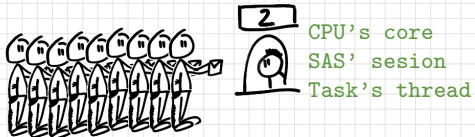
CPU's core  
SAS' session

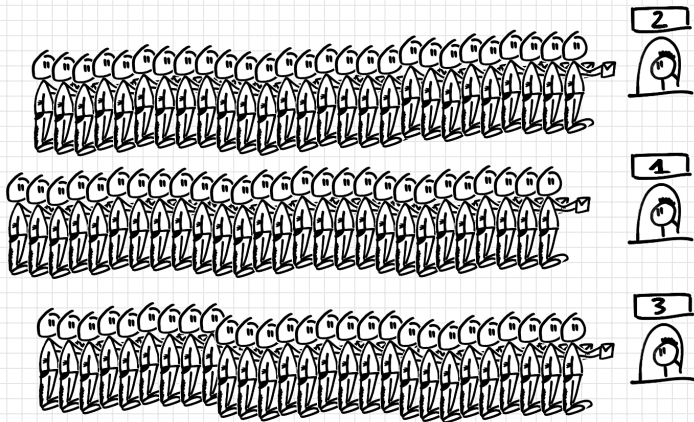


CPU's core  
SAS' session



CPU's core  
SAS' session





Demo 1.

www: `http://www.mini.pw.edu.pl/~bjablons/SASpublic/`

kod: `Parallel-processing-in-BASE-SAS.sas`

Lista pytań

## Zagadnienie drugie

Jak wykorzystać PROC FCMP i funkcje użytkownika oraz %MACRO'a do przekazywania danych w HASH tablicach pomiędzy DATA STEPami?

## OpenKoldLand



## OpenKoldLand



OpenKöldLand



Data step 1



Data step 2

OłpenKołdLand



Data step 1

Proc FCMP i funkcja  
z Hash Tablicą

Data step 2

OłpenKołdLand



Data step 1

Proc FCMP i funkcja  
z Hash Tablicą

Data step 2

Proc FCMP i funkcja  
z Hash Tablicą

OłpenKołdLand



Data step 1

Proc FCMP i funkcja  
z Hash Tablicą

FCMP+Resolve(%Macro)



Data step 2





Data step 1

Proc FCMP i funkcja  
z Hash Tablicą

OłpenKołdLand



Data step 2



FCMP+Resolve(%Macro) to "most" pozwalający przekazywać dane pomiędzy granicami data step'ów w hash tablicy zaszytej w funkcji bez ponownego wczytywania danych (autorem pomysłu użycia FCMP oraz %MACRO jest pan Art Carpenter, rozszerzenie na datastep'y z użyciem funkcji RESOLVE() zasugerowaliśmy mu niezależnie z Thomasem Billingssem)

Demo 2.

www: `http://www.mini.pw.edu.pl/~bjablons/SASpublic/`

kod: `Function-Hash-Macro-sandwich.sas`

(suplement

`Function-Hash-Macro-sandwich-approach-2.sas`)

Lista pytań

## Zagadnienie trzecie

Jak mając do wykonania klauzulę WHERE "typu" OR (na różnych zmiennych) nadal móc wykorzystywać zalety INDEXÓW?





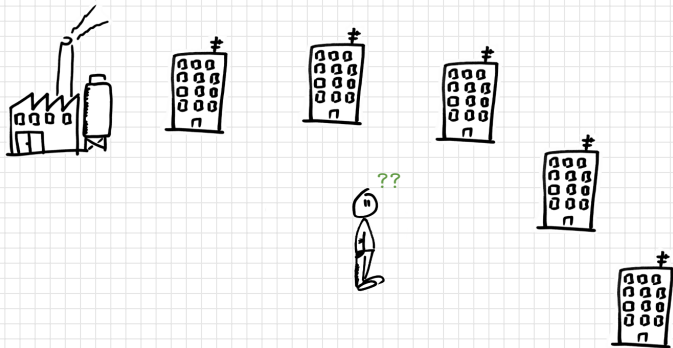


```
WHERE prod_from_Factory = "ABC";
```



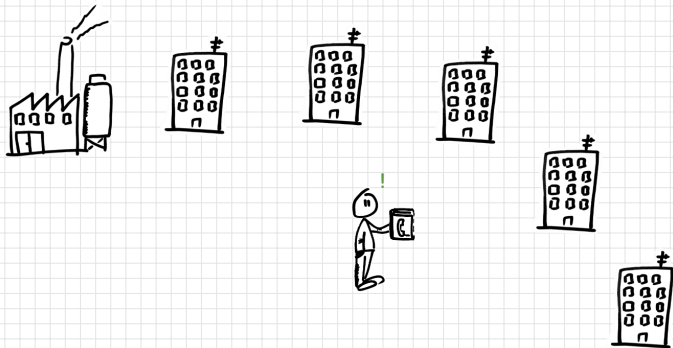
```
WHERE prod_from_Factory = "ABC";
```

mały zbiór? - przeczytaj sekwencyjnie



```
WHERE prod_from_Factory = "ABC";
```

duży zbiór?



```
WHERE prod_from_Factory = "ABC";
```

duży zbiór? - użyj INDEX'u



```
WHERE prod_from_FactOne = "ABC" OR prod_from_FactTwo = "XYZ";  
/* użycie INDEX'u jest niemożliwe... */
```



mały zbiór?  
przeczytaj sekwencyjnie



```
WHERE prod_from_FactOne = "ABC" OR prod_from_FactTwo = "XYZ";
```

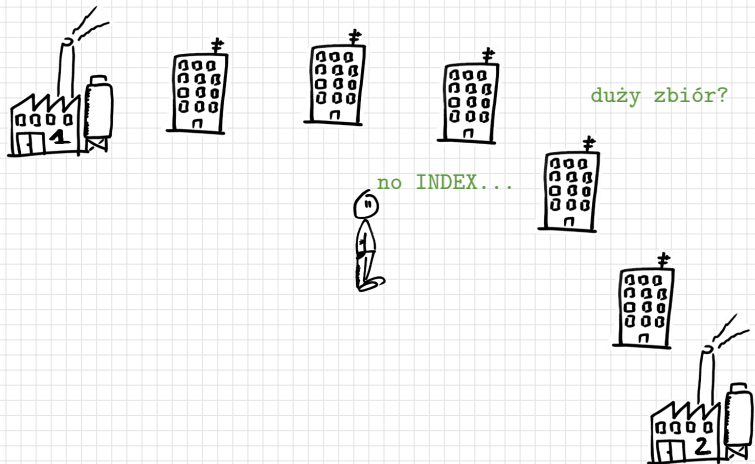


mały zbiór?  
przeczytaj sekwencyjnie

index na małym zbiorze  
to tak jak auto-alarm  
we Fiacie Multipla...

```
WHERE prod_from_FactOne = "ABC" OR prod_from_FactTwo = "XYZ";
```





```
WHERE prod_from_FactOne = "ABC" OR prod_from_FactTwo = "XYZ";
```



duży zbiór?



```
set ... ;  
WHERE prod_from_FactOne = "ABC";
```

```
output;
```

```
set ... ;  
WHERE prod_from_FactTwo = "XYZ";
```

```
output;
```



duży zbiór?



```
Declare HASH Lista();
Lista.DefineKey("CurObs");
Lista.DefineDone();
```



```
set ... curobs = CurObs;
WHERE prod_from_FactOne = "ABC";
Lista.ADD();
output;
```

```
set ... curobs = CurObs;
WHERE prod_from_FactTwo = "XYZ";
if Lista.FIND() then output;
```

Demo 3.

www: `http://www.mini.pw.edu.pl/~bjablons/SASpublic/`

kod: `OR-condition-in-WHERE-clause-with-INDEX.sas`  
(dane `Countries.sas`)

Lista pytań



<https://www.linkedin.com/groups/8675451>

dziękuję



- Art Carpenter, "Using Hash Tables to Manage Your Macro Language Control Files", SAS GF 2018 Proceedings, <https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2018/2399-2018.pdf>
- Art Carpenter, "Using Hash Tables to Manage Your Macro Language Control Files" (final), WUSS 2018 Proceedings, [https://www.lexjansen.com/wuss/2018/41\\_Final\\_Paper\\_PDF.pdf](https://www.lexjansen.com/wuss/2018/41_Final_Paper_PDF.pdf)
- Art Carpenter, "Carpenter's Guide to Innovative SAS Techniques", SAS Press
- Art Carpenter, "Carpenter's Complete Guide to the SAS Macro Language", SAS Press
- Ian Whitlock, "The RESOLVE Function - What Is It Good For?", NESUG 1998 Proceedings, <https://www.lexjansen.com/nesug/neg98/code/p088.pdf>
- Paul M. Dorfman, Koen Vyverman, "Data Step Hash Objects as Programming Tools", SUGI 30 Proceedings, [www2.sas.com/proceedings/sugi30/236-30.pdf](http://www2.sas.com/proceedings/sugi30/236-30.pdf)
- Chris Schacherer, "Introduction to SAS Hash Objects", SAS GF 2015 Proceedings, <https://support.sas.com/resources/papers/proceedings15/3024-2015.pdf>
- Paul M. Dorfman, Don Henderson, "Data Aggregation Using the SAS Hash Object", SAS GF 2015 Proceedings, <https://support.sas.com/resources/papers/proceedings15/2000-2015.pdf>
- Paul M. Dorfman, "Fundamentals of the The SAS Hash Object", SESUG 2016 Proceedings, [https://analytics.ncsu.edu/sesug/2016/HOW-195\\_Final\\_PDF.pdf](https://analytics.ncsu.edu/sesug/2016/HOW-195_Final_PDF.pdf)
- Dylan Ellis, "RUN\_MACRO Run! With PROC FCMP and the RUN\_MACRO Function from SAS 9.2, Your SAS Programs Are All Grown Up", SAS GF 2013 Proceedings, <https://support.sas.com/resources/papers/proceedings13/033-2013.pdf>
- Andrew Henrick, Donald Erdman, Stacey Christian, "Hashing in PROC FCMP to Enhance Your Productivity", SAS GF 2013 Proceedings, <http://support.sas.com/resources/papers/proceedings13/129-2013.pdf>
- Mike Rhoads, "Use the Full Power of SAS in Your Function-Style Macros", SAS GF 2012 Proceedings, <https://support.sas.com/resources/papers/proceedings12/004-2012.pdf>
- Rick Langston, "Submitting SAS Code On The Side", SAS GF 2013 Proceedings, <https://support.sas.com/resources/papers/proceedings13/032-2013.pdf>
- Billy Clifford, "Frequently Asked Questions about SAS Indexes", SUGI 30 Proceedings, <http://www2.sas.com/proceedings/sugi30/008-30.pdf>
- Michael A. Raithel, "The Complete Guide to SAS Indexes", SAS Press

Chciałbym serdecznie podziękować moim kolegom z Citi Europe PLC za wiele cennych uwag, które znacząco poprawiły jakość niniejszej prezentacji i załączonych kodów.

Dziękuję w szczególności:

Krzysztofowi Sockiemu,

Filipowi Kulonowi,

i Michałowi Gostkowskiemu.