

Bartosz Jabłoński, PhD

- A mathematician working and "playing" with data.
- Seasoned **SAS** user with experience in: education, telecommunication, clinical trials, and banking.
- As "active-learner" and "hobbyist-teacher" leads **SAS** classes and courses at the Faculty of Mathematics and Information Science at Warsaw University of Technology.
- During free time coordinates meetups of Polish **SAS** Users Group (#PoLSUG), and plays www.sasensei.com.

3722-2019

**Use the advantage of INDEXes
even if a WHERE clause
contains an OR condition**

Bartosz Jabłoński

Warsaw University of Technology / Citibank Europe PLC Poland

✉ yabwon@gmail.com

April 28th - May 1st

SAS Global Forum 2019, Dallas, Texas



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

the question

How to handle a situation in which one have a WHERE clause with the OR condition on two different INDEXed variables and want to use advantage of INDEXing to subset the data faster?



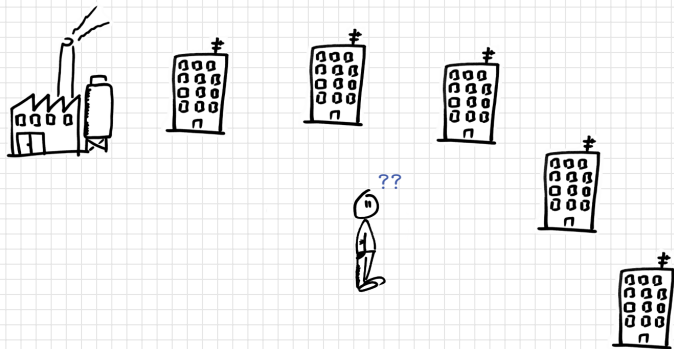


```
WHERE prod_from_Factory = "ABC";
```



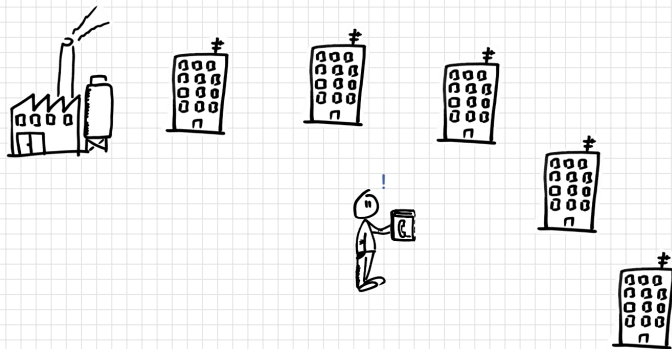
```
WHERE prod_from_Factory = "ABC";
```

```
small dataset? - read sequentially
```



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

big dataset?



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

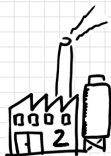
big dataset? - use INDEX



```
WHERE prod_from_FactOne = "ABC" OR prod_from_FactTwo = "XYZ";  
/* you can't use INDEX here... */
```



small dataset?
read sequentially



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



small dataset?
read sequentially

index on small set
is like a car-alarm
for Fiat Multipla...

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



big dataset?

no INDEX...



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



big dataset?



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

```
output;
```

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

```
output;
```



big dataset?



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



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

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

Live Demo?

Go to the following web address
and read and run the code to see the "bare metal".

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

codes starting with:

`OR-condition-in-WHERE-clause-with-INDEX-*.sas`
(and `Countries.sas`)



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

thank you



remember to complete the session survey

- Art Carpenter, "Carpenter's Guide to Innovative SAS Techniques", SAS Press
- Paul M. Dorfman, Koen Vyverman, "Data Step Hash Objects as Programming Tools", SUGI 30 Proceedings, 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 Management Solutions Using SAS Hash Table Operations: A Business Intelligence Case Study", SAS Press
- Paul M. Dorfman, "Fundamentals of the The SAS Hash Object", SESUG 2016 Proceedings, https://analytics.ncsu.edu/sesug/2016/H0W-195_Final.PDF.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
- Mark Keintz, "A Faster Index for Sorted SAS Datasets", SAS GF 2009 Proceedings, <https://support.sas.com/resources/papers/proceedings09/024-2009.pdf>