

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

Bartosz Jabłoński

Polish SAS Users Group

✉ yabwon@gmail.com

Tuesday, May 21st, 2019

WCSUG Lunch and Learn Webinar



<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?

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?

the answer

SAS Proceedings' paper 3722-2019



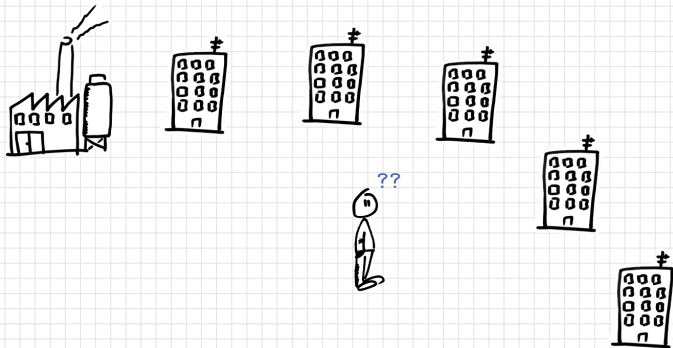


```
WHERE prod_from_Factory = "ABC";
```



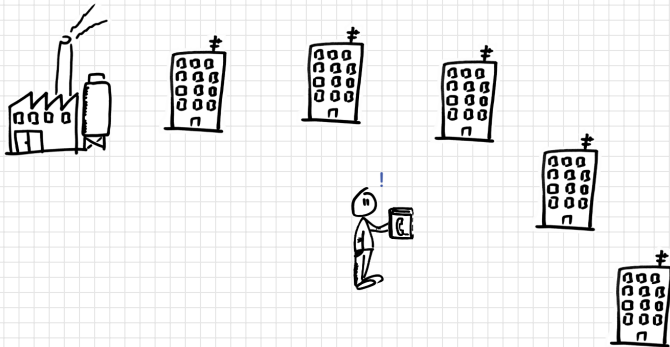
```
WHERE prod_from_Factory = "ABC";
```

```
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";
```

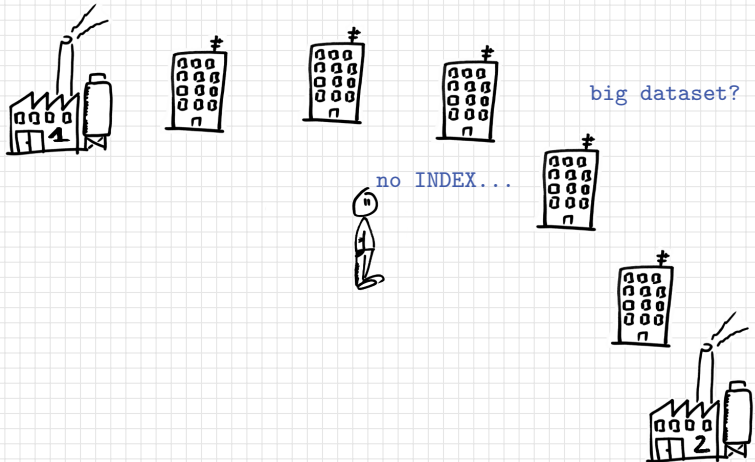


small dataset?
read sequentially

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

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





```
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;
```

Small Example

Let us look at a small simple example.

A dataset with two indexed variables

Obs.	VarL	VarN
1	A	10
2	A	10
3	A	20
4	B	20
5	B	20
6	B	30
7	C	30
8	C	30
9	C	40

WHERE VarL="B" or VarN=20



WHERE VarL="B"



WHERE VarN=20



Dataset with observations read by
WHERE VarL="B" and WHERE VarN=20

Current Obs.	VarL	VarN	
4	B	20	✓
5	B	20	✓
6	B	30	✓
3	A	20	✓
4	B	20	✗
5	B	20	✗

Observations read with WHERE VarL="B"

Current Obs.	VarL	VarN	
4	B	20	✓
5	B	20	✓
6	B	30	✓

A list of read observations

observation number	"4"	"5"	"6"
-----------------------	-----	-----	-----

Observations added by WHERE VarN=20

Current Obs.	VarL	VarN	
4	B	20	✓
5	B	20	✓
6	B	30	✓
3	A	20	✓
4	B	20	✗
5	B	20	✗

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>