Pavel Málek

PASS SQLSATURDAY BRATISLAVA | 23 JUN 2018

Score a (row) goal and beat a query optimizer

Sponsors

Bio

- 10+ years of experience with SQL Server
- Czech PASS leader
- SQL Saturday Prague co-organizer

ČESKÁ SPOŘITELNA

- Lead SQL Server DBA in largest Czech bank for 6+ years

Joyful Craftsmen

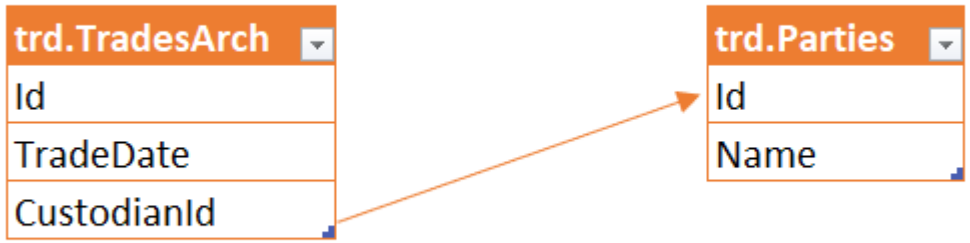
- Currently newbie at Joyful Craftsmen

@malekpav LinkedIn malekpav@gmail.com

Before we begin...

Challenge for you!

Try to write a query which will show you 10 last trades *(trd.TradesArch)* for each market trader *(trd.Parties)*?



It will be used on primary site for traders => it has to be faaaast!

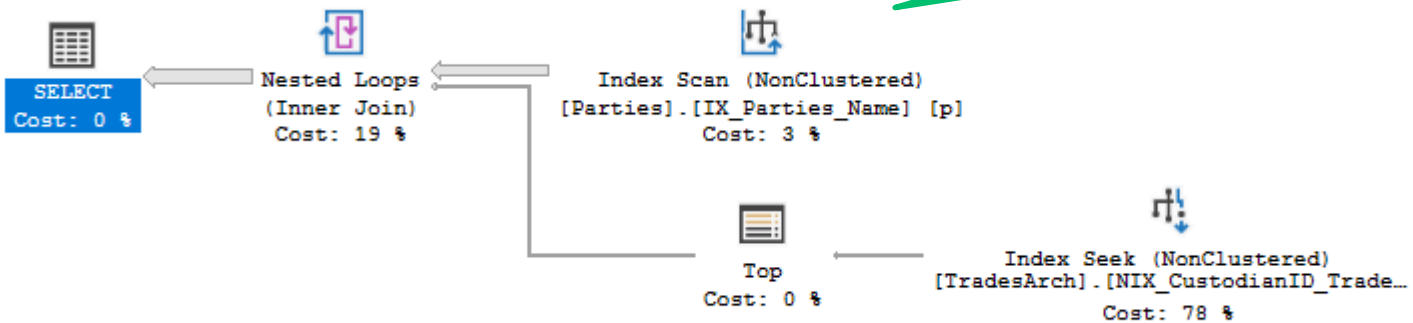
Solution

Nature of SQL

- SQL is a declarative language!

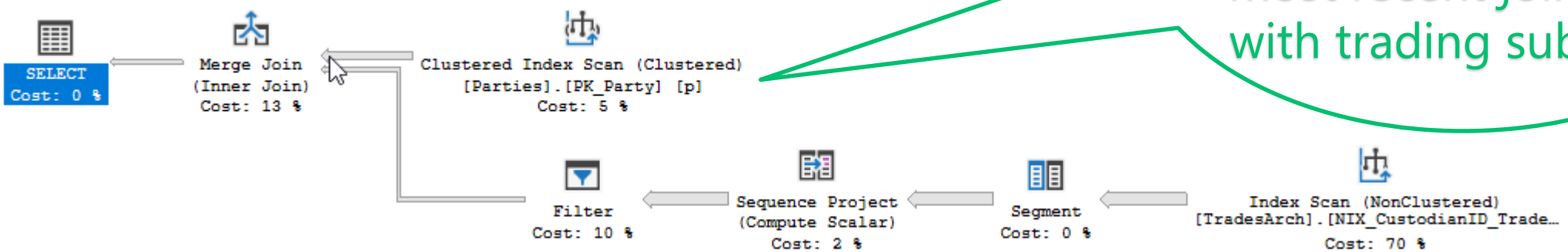
But how much “declarative”?

Cross apply



For every subject trading on market find the most recent related trade

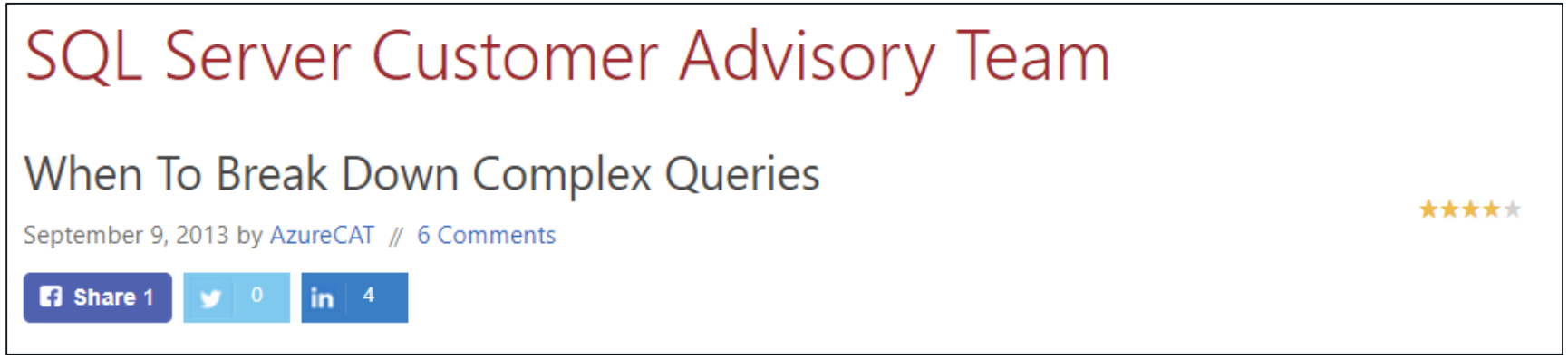
RowNb



Go through all the trades and show only those who are the most recent joined with trading subjects

Query Optimizer is in charge, but…

- Sometimes QO simply cannot have correct inputs



- OR logic in the WHERE clause
- Aggregations in intermediate results sets
- A large number of very complex joins
- A CASE clause in the WHERE or JOIN clause

Query optimizer having troubles

RE: Long responses in BO

Pavel Málek <pmalek@csas.cz>
komu: Jan, sITS, Martin

15.11.17 ☆ Odpovědět všem

There are currently only two long running queries:

```
SELECT TOP (80) * FROM trd.TradesView ORDER BY [Id] DESC, [Version] DESC

SELECT TOP (80) * FROM trd.TradesArchView ORDER BY [Id] DESC, [Version] DESC
```

Query plan is attached. Very first impressions - nested loop everywhere, estimated nr of rows is not accurate. Wrong stats? I am going to update them...

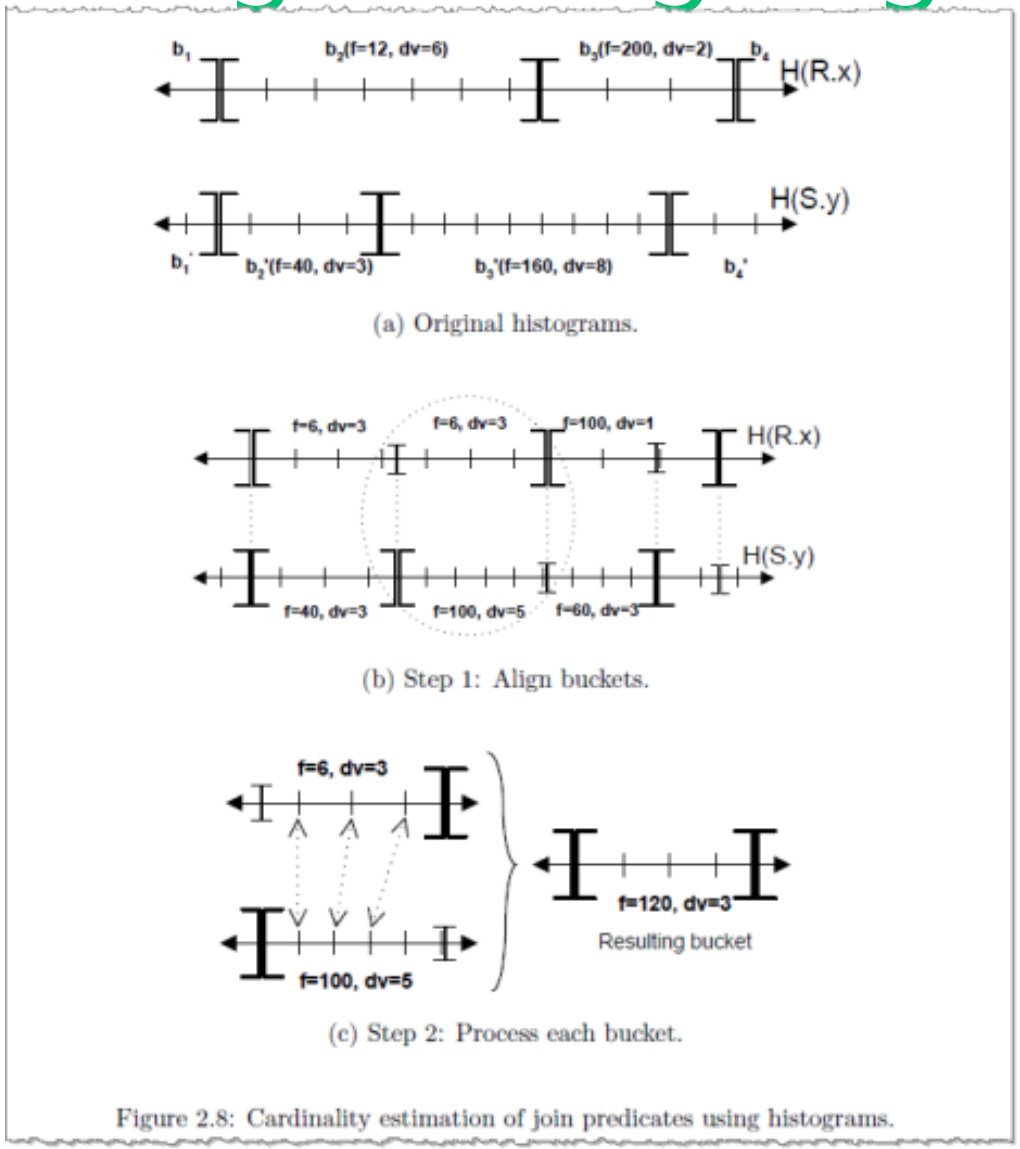
P.

```
SELECT TOP (80) * FROM trd.TradesView ORDER BY [Id] DESC, [Version] DESC

SELECT TOP (80) * FROM trd.TradesArchView ORDER BY [Id] DESC, [Version] DESC
```

Demo

Histogram aligning



Nicolas's Bruno (Software Developer, Microsoft) work *"Statistics on Query Expressions in Relational Database Management Systems" COLUMBIA UNIVERSITY, 2003*.

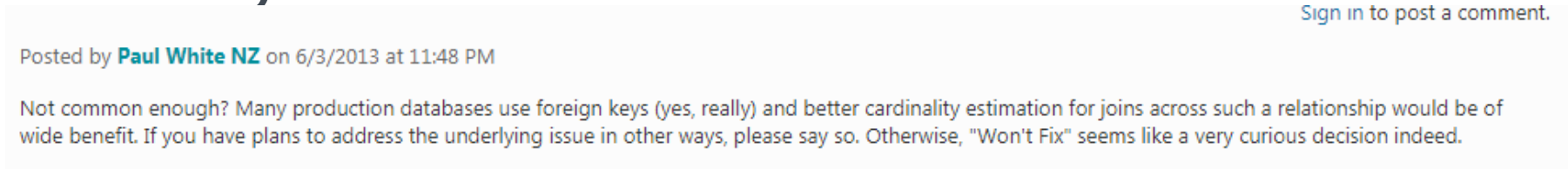
"Mere mortals" reading:
Hugo Kornelis – SQL Sat presentation
Dimo – queryprocessor.com

Other possibilities?

Simple Join: Simple join ignores histograms from the two relations, and employs a containment model that simply assumes all distinct values from one child are contained in the other (originally introduced in [8]). This simplistic strategy typically sacrifices estimation quality for faster optimization time, providing a balance that better suits some workloads.

Demo Takeaways

- FKs are good
- Indexes on FKs is almost always good
- But don’t expect them to be “always” used for row estimation!
- **Updates of stats via full scan matters!!!**
- TF 2301 sometimes helps
- Carefully with TF 2371



Demo case continues

What if I don’t want to use any traceflags

TOP operator introduced an interesting behavior

What the hell “a row goal” is?

Case Trading System#2: Key Takeaways

Via TOP + ORDER BY you are able to create your own “blocks”

Via TOP (@x) OPTIMIZE FOR you are able to decrease estimated nb of rows

Via MANY() you are able to enlarge estimated nb of rows

CREDITS: Adam Machanic

Another example of fixed QO behavior

Summary

Now you know:

- How much statistics matter
- How joins are calculated
- How to force QO to do what you want to
  - (TOP blocks, MANY(), TOP (@x) + OPTIMIZE FOR)
- CROSS APPLY vs ROWNB()
- What are the possibilities for MAX() and MIN()

Thank you!

Evaluations

Event Evaluation

http://www.sqlsaturday.com/764/EventEval.aspx

Session Evaluation

http://www.sqlsaturday.com/764/Sessions/SessionEvaluation.aspx