

(June 10th, 2013)

If you know someone who you think would benefit from being an Insider, feel free to forward this PDF to them so they can sign up [here](#).



Quick Tips for our Insider friends!

Hey Insiders!

This newsletter is coming to you from London where we've just started our IE2 on Performance Tuning. Last weekend we went to the famous [Kew Botanical Gardens](#), and Joe and Jon presented a pre-con and sessions at SQL Saturday in Edinburgh – I wish I could have been there with them, as Edinburgh is where I lived and studied for 8 years.

Our latest Pluralsight course has been published: [Developing and Deploying SQL Server ISV Applications](#) by Erin Stellato – check it out!

And don't forget to check out our three brand-new classes that we're debuting in September:

- IE0: Immersion Event for the Accidental DBA (3 days, 9/30 – 10/2, Erin and Jon)
 - [Dates, logistics, and registration](#)
 - [Detailed curriculum](#)
- IEHW: Immersion Event on SQL Server Hardware (2 days, 9/18 – 19, Glenn)
 - [Dates, logistics, and registration](#)
 - [Detailed curriculum](#)
- IETS: Immersion Event on Advanced T-SQL (4 days, 9/30 – 10/3, Bob)
 - [Dates, logistics, and registration](#)
 - [Detailed curriculum](#)

We're really excited about these new classes and hope you are too!

Our Fall SQLintersection conference is also open for registrations, with a stellar line-up of speakers and sessions, coming to Las Vegas at the end of October. Read all about it [here](#).

The most recent books I've read (back-to-back over the last week) are Colin Thubron's [Among the Russians](#) and [In Siberia](#).

The first is an account of traveling through the western Soviet Union in 1980. The people he meets mostly have a horribly blinkered view of the western world and deny the Soviet Union's culpability in the various invasions (e.g. Afghanistan, Hungary, and Czechoslovakia). At that time the Soviet economy was terrible, with shortages of most goods, people supplementing meager incomes with black-market activity, and the whole thing causing massive depression, hopelessness, and rampant alcoholism in the population.

The second book was written as the Soviet Union was breaking up and reflects quite a different experience from the *Among the Russians*. The people he speaks to are looking forward to reclaiming their ethnic roots and religions, after shaking off the constrictive yolk of central Soviet control. Siberia was becoming very run down at that time, with withdrawn central economic support and rampant inflation, and I noticed this too when we were there in 2010. In the Far East of Siberia, Thubron encounters the ruins of some very depressing old Stalinist labor camps, for instance at Butugychag, where exiled convicts mined for uranium with no protection and life expectancy was months (killing about 380 thousand in ten years).

Both books were excellent and highly recommended.

Please [let us know](#) if you liked what you read/saw here and/or have any suggestions for future Quick Tips.

Note: you can get all the prior Insider newsletters [here](#).

Paul's Ponderings

In the [previous newsletter](#) I discussed data compression and encouraged those of you with Enterprise Edition to check it out as a way to reduce storage costs and increase buffer pool efficiency.

The benefits of data compression include:

1. The database and backups are smaller
2. Data density increases for large range scans, reducing the number of read I/Os
3. More data fits in memory for general workloads, reducing the number of read I/Os

However, just because data compression is available, that does not mean it's the right solution for reducing the amount of disk space and memory used by your data. Also, if you *don't* have access to data compression (most likely because you're not on Enterprise Edition, or maybe because data compression results in too much of a performance degradation) there are other solutions for you.

Even if you *are* using data compression, there are extra things you should do to further reduce wasted space in the database.

First, check to make sure that you don't have duplicate or unused nonclustered indexes. Any extra indexes take up disk space, take up memory in the buffer pool, and their maintenance (during inserts/updates/deletes and periodic fragmentation removal) uses CPU and generates extra transaction log. They're pure overhead and by removing them you can realize benefits #1-3 above.

Check out the following posts for details of what to do:

- [Removing duplicate indexes](#)
- [Indexes From Every Angle: How can you tell if an index is being used?](#)

Second, make sure you're managing index fragmentation correctly. This means setting appropriate FILLFACTORs to help avoid costly page splits that produce low page densities, logical fragmentation, and lots of extra transaction log when they occur.

See the following blog posts for details:

- [How expensive are page splits in terms of transaction log?](#)
- [Performance issues from wasted buffer pool memory](#)
- [How to choose a good index fillfactor](#)

Third, and this one's a lot harder to fix, you may be wasting lots of space in your nonclustered indexes from having large clustered index keys. Every nonclustered index record must include the cluster key from the clustered index record on which it is based. The longer the cluster key is, the more space is taken up in aggregate across the nonclustered indexes.

As an example, consider a cluster key made up of a 16-byte GUID and three 4-byte integers, for a total of 28 bytes. If the table has 1.5 billion rows and 12 nonclustered indexes, the space taken up in the nonclustered indexes for the cluster key is $1.5 \text{ billion} \times 12 \times 28 = \sim 469 \text{GB}$. If the cluster key was changed to an 8-byte bigint, the space taken up in the nonclustered indexes drops to only $1.5 \text{ billion} \times 12 \times 8 = \sim 134 \text{GB}$. That's a pretty big space saving!

See the following blog posts for details:

- [How cluster key size can lead to GBs of wasted space](#)
- [Code to list potential cluster key space savings per table](#)

Finally, don't discount backup compression. If you only need to reduce the size of your backups, you may find that data compression is overkill – just use backup compression. Even if you're using data compression for the benefits apart from backup size reduction, backup compression can also provide further compression of backups. Just make sure you test the backup compression ratio to make sure you're getting a decent amount of compression that makes it worthwhile to use the extra CPU used when backing up and restoring. See [SQL Server 2008: Backup Compression](#).

Call to action: Whether you're using data compression or not, make sure you're doing everything you can to reduce the size of your databases and make your buffer pool usage as efficient as possible. You'll likely be surprised at just how much space is being wasted in your databases.

I'm curious to hear your thoughts on alternatives and additions to data compression, including whether you used the various bits of code above to save space, so please feel free to [drop me a line](#), treated confidentially of course.

Video Demo

From Erin:

You probably try to avoid duplicate indexes in your environment, as you know they create additional overhead and do not provide any performance benefit. Do you do the same for statistics? It's not uncommon to find a separate column statistic for the leading column of an index – SQL Server automatically created the column statistic at some point *before* you created the nonclustered index that leads on that column. When this occurs, you have duplicate statistics for a column, but unlike duplicate indexes, duplicate statistics are not always *exactly the same*. Sample size, update date, density, the histogram – they can vary between the duplicate statistics.

In this Insider Video I'll explore the typical method by which duplicate statistics exist in a database, and step through code you can use to find and remove redundant column statistics. Want more information on statistics? Sign up for my remote session, *Updates to Statistics and the Effect on Query Performance*, which I'll present June 27 during the [Summer Performance Palooza](#) hosted by the PASS Performance Virtual Chapter. Hope to see you there!

The video is just over 12 minutes long and you can get it:

- In WMV format [here](#)
- In MOV format [here](#)

You can get the demo code [here](#).

Enjoy!

SQLskills Offerings

As I mentioned above, we have three brand-new classes that we've created, and they're available for registration for deliveries in Bellevue, WA in September/October 2013. They're sure to sell out fast, so don't delay!

Please know that these classes are final as the hotel contracts are signed, and the classes will not be cancelled or moved for any reason, nor will the dates change.

- June 17-21, 2013: High Availability & Disaster Recovery (**IE3**) in London – UK (registration closes on Wednesday!)
- September 16-20, 2013: Internals and Performance (**IE1**) in Bellevue, WA – USA
- September 18-19, 2013: Hardware (**IEHW**) in Bellevue, WA - USA

- September 23-27, 2013: Performance Tuning (**IE2**) in Bellevue, WA – USA
- September 30-October 2, 2013: Accidental DBA (**IE0**) in Bellevue, WA – USA
- September 30-October 3, 2013: Advanced T-SQL (**IETS**) in Bellevue, WA – USA

See [here](#) for the main Immersion Event Calendar page that allows you to drill through to each class for more details and registration links.

Summary

We hope you've enjoyed this issue - we really enjoy putting these together.

If there is anything else you're interested in, we'd love to hear from you - [drop us a line](#).

Thanks,
Paul and Kimberly

Paul@SQLskills.com and Kimberly@SQLskills.com