

Deduping & Eliminating with Inaccurate Read & Write Metrics

1.3 p1

Agenda

The two SQL Server index usage views:

- By plan: sys.dm_db_index_usage_stats
- By index: sys.dm_db_index_operational_stats

Why they're not as accurate as you might suspect

How sp_BlitzIndex shows the results

How I interpret the results to do the D.E. parts



SQL Server has a lot of metrics.

Old-school, operating system level: Perfmon counters

System & database level: "system tables & views"

- Dynamic Management Views (DMVs)
- Dynamic Management Functions (DMFs)

Tracing: Profiler, Extended Events

1.3 p3



Dynamic Management Views

The good:

- Well-documented by both Microsoft and blogs
- · It's easy to find scripts and tools that use 'em

The bad:

- A lot of the user-written documentation is wrong
- Many of the DMVs don't mean what you think
- Contents can reset at surprising times
- Hit-or-miss coverage in Azure, keeps changing



sys.dm_db_index_usage_stats

Shows # of executions where a plan included an operator

 Does NOT show if the operator was used (or how often it was accessed)

Number and last date of reads (seeks, scans, lookups)

Number and last date of last write

• Insert/update/deletes all called "updates"

Data is since startup or when the index was modified

1.3 p5



sys.dm_db_index_operational_stats

Lower level, more transitory

Lock waits (page and row)

Access counts

 Doesn't distinguish between full scans/range scans, or even range scans and seeks

Data only persisted while object's metadata is in memory

No good way when to tell it was last cleared



A lot of tools use this data.

T-SQL scripts: sp_BlitzIndex, Glenn Berry's scripts

Apps: SentryOne Plan Explorer, lots of monitoring tools like Quest Spotlight, Red Gate SQL Monitor

You may have written your own scripts too (let's be honest, you copied it from someone else's online, and you don't really know what it's doing)

1.3 p7



I'm going to be using my favorite.

sp_BlitzIndex in the First Responder Kit:

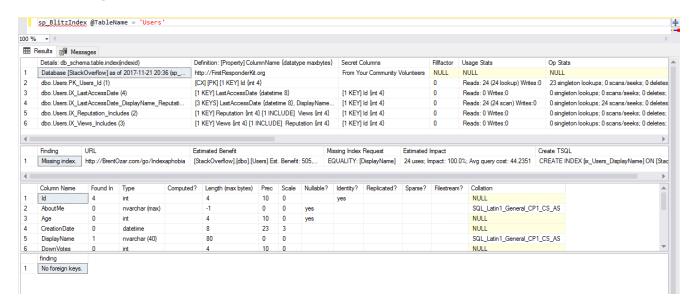
- Github repo: FirstResponderKit.org
- · Zip download: BrentOzar.com/first-aid
- · Slack chat: SQLslack.com, #FirstResponderKit

Lots of code contributors, used all over the world

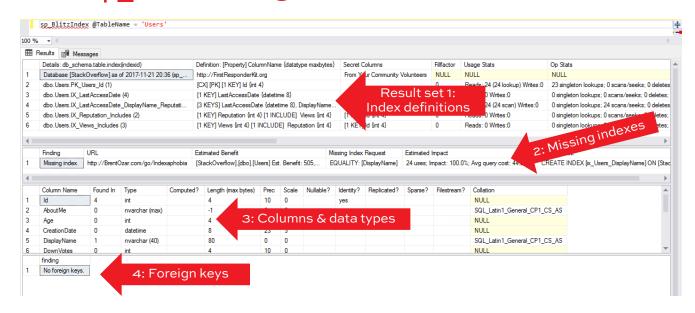
Open source, free as in speech, MIT License



Running it for one table: sp_BlitzIndex @TableName = ' Users'



Running it for one table: sp_BlitzIndex @TableName = 'Users'



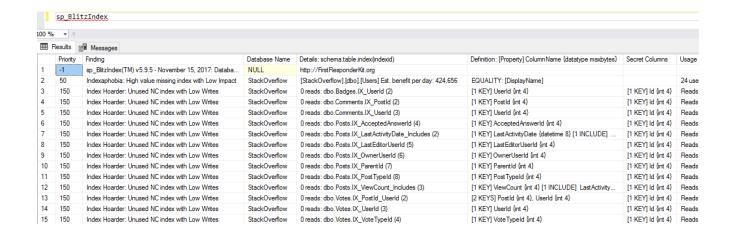


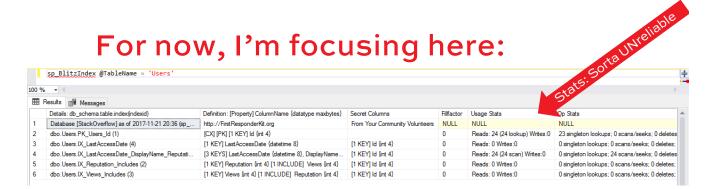
For now, I'm going to focus on usage stats & operational stats to explain the DMV gotchas.



Later, I'll use it across a database.

Defaults to the current database, or you can pick one with @DatabaseName parameter





Usage stats: sys.dm_db_index_usage_stats

Operational: sys.dm_db_index_operational_stats



Build the execution plan for this.

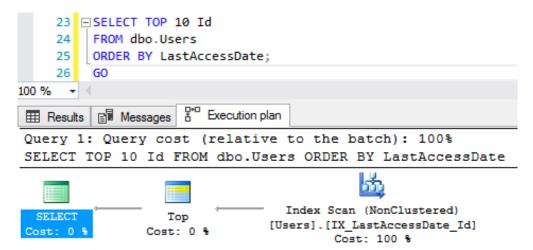
```
SELECT TOP 10 Id
FROM dbo.Users
ORDER BY LastAccessDate;
GO
```

Flash back to How to Think Like the Engine:

- Clustered index on Id (white pages)
- Nonclustered on LastAccessDate, Id (black pages)



It's a "scan" - but what does that mean?



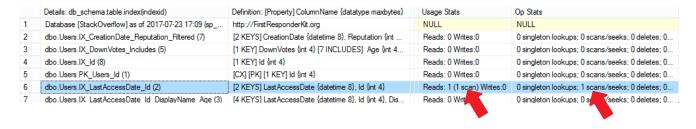
1.3 p15

It's efficient – it doesn't scan the whole index.

It's just labeled a scan

The DMVs don't distinguish between types of scans

sp_BlitzIndex @TableName = 'Users'

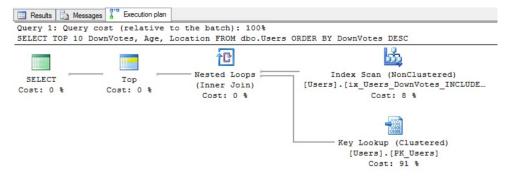


1.3 p17



This plan has a key lookup

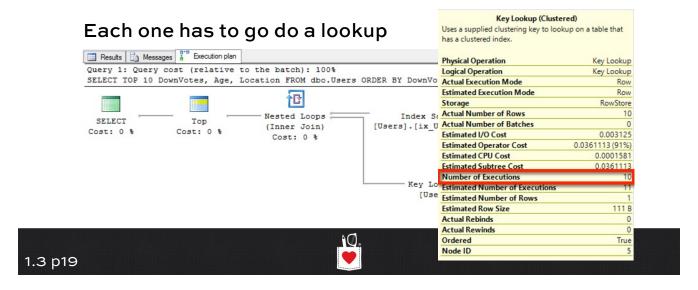
For every row from the nonclustered index scan, it looks up related values in the clustered index





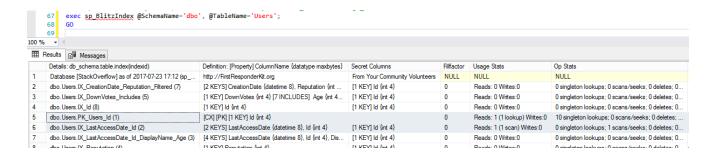
Key lookup is executed 10 times

Index scan returns 10 rows



The stats are different

- sys.dm_db_index_usage_stats: Number of times the operator appeared in an execution plan since last reset
- sys.dm_db_index_operational_stats: Number of times the operator was executed (recently)



Similar query, but no rows match

1.3 p21

It doesn't execute the lookup

```
/* This plan is slightly different, it has a key lookup - but it doesn't get executed. */
   □ SELECT TOP 10 Id, Location
     FROM dbo.Users
     WHERE LastAccessDate > GETDATE()
     ORDER BY LastAccessDate;
     GO
150 %
Query 1: Query cost (relative to the batch): 100%
SELECT TOP 10 Id, Location FROM dbo.Users WHERE LastAccessDate > GETDATE() ORDER BY LastAccessDate
                                                 ų.
              Index Seek (NonClustered)
                         Nested Loops
  Top
                                        [Users].[IX_LastAccessDate]
                         (Inner Join)
             Cost: 0 %
 SELECT
                          Cost: 0 %
                                              Cost: 50 %
              0.001s
                                                0.001s
              0 of
                            0 of
                                                0 of
              1 (0%)
```

What the DMVs say

Usage stats says the plan included the seek

Operational stats says it wasn't executed

```
/* How does that show up in the DMVs? */
        exec sp_BlitzIndex @SchemaName='dbo', @TableName='Users';
150 %
 Results Messages
                                                  Definition: [Property] ColumnName {datatype maxbytes}
                                                                                                                               Op Stats
      Details: db_schema.table.index(indexid)
                                                                                                Usage Stats
     Database [StackOverflow2013] as of 2019-06-11 10... http://FirstResponderKit.org
                                                                                                 Server: SQL2019 Days Uptime: 0.04 NULL
                                                 [CX] [PK] [1 KEY] ld {int 4}
 2 dbo.Users.PK_Users_ld (1)
                                                                                                Reads: 1 (1 lookup) Writes:0
                                                                                                                               0 singleton lookups; 0 scans/seeks; 0
     dbo.Users.IX_LastAccessDate (2)
                                                 [1 KEY] LastAccessDate {datetime 8}
                                                                                                Reads: 1 (1 seek) Writes:0
                                                                                                                               0 singleton lookups; 1 scans/seeks; 0
```



So, the contents don't match.

When I use them, I'm really just asking:

- Is this index helping? (reads)
- Is this index hurting? (writes)
- Roughly how much? (quantity: millions, billions)

But your next question is, "When did these numbers reset?"



Uh, well...I can't tell.

When SQL Server restarts (which we can measure)

When an Availability Group failed over (harder to tell)

When Azure SQL DB fails over, restarts (can't see)

SQL 2012, 2014: resets on ALTER INDEX REBUILD

- SQL 2012: fixed in SP2 CU12, or SP3 CU3
- SQL 2014: fixed in RTM CU14, or SP1 CU8, or SP2
- SQL Server 2016 & newer: unaffected

1.3 p25



Index DMVs: Your takeaways

- "Scan" may not be the whole table
- · "Seek" might actually be the whole table

sys.dm db index usage stats - "usage stats"

- Show # of times an operator appeared in a query plan that was run
- The operator may have been accessed many times, or not at all
- · Reset by system restart, or by index rebuild if on buggy versions

sys.dm db index operational stats - "op stats"

- Show number of times an operator was accessed
- Very volatile, can be reset by memory pressure

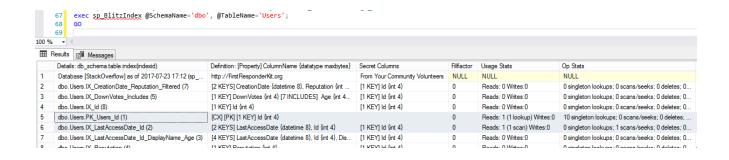
Only check when you have enough uptime to reflect business processes.



I'm not saying not to use these.

I'm just saying don't put too much faith in the details.

Your goal is just to know, "Are these indexes kinda getting used, or totally ignored?"



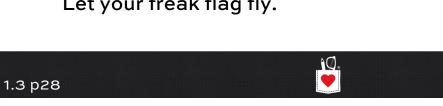
Our scripts try to prioritize stuff.

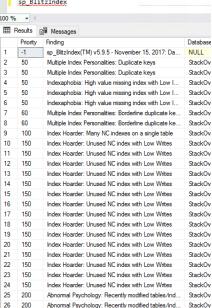
We want you to focus on the biggest bang-for-the-buck first.

Findings here are a little more fun - we use psychiatry terms.

There's nothing wrong with some psychiatric disorders, either.

Let your freak flag fly.

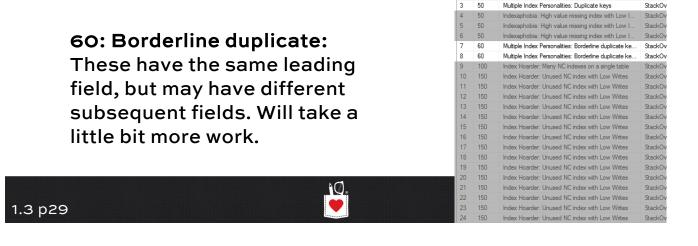




Our scripts try to prioritize stuff.

50: Duplicate keys:

These are no-brainers to dedupe.



.00 % -

Results Messages

sp_BlitzIndex(TM) v5.9.5 - November 15, 2017: Multiple Index Personalities: Duplicate keys

How I use sp_BlitzIndex to D/E

- sp_BlitzIndex @GetAllDatabases = 1
 (and figure out what database to tune)
- 2. Run sp_BlitzIndex in the database I want to tune (and figure out what table I want to focus on)
- Scroll across to the More Info column and run it for the particular table I want to tune



Advanced sp_BlitzIndex tips

@ThresholdMB: default 250MB, only alerts you for problems with indexes at least this large

@Mode:

- O = default, most urgent problems
- 4 = more analysis, includes more warnings
- 2 = inventory of all your indexes & metrics
- 1 = summary of space usage

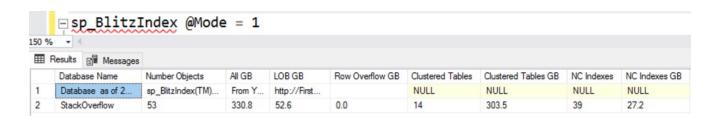


Example writeup

"On the Users table, we have:

- 3 duplicate indexes of 15GB total
- 2 unused indexes of 8GB total

By removing these, I saved 23GB of drive space, made deletes/updates/inserts go faster, and now I can add more appropriate indexes."



Azure SQL DB? Log it regularly.

Since your database can (and will) restart without warning, log this data weekly so it's there when you want to do index analysis, and use the most recent:

```
sp_BlitzIndex @Mode = 2,
@OutputDatabaseName = 'MyDB',
@OutputSchemaName = 'dbo',
@OutputTableName = 'BlitzIndex_Mode2'
2 = inventory, 3 = missing indexes
```

1.3 p33



What we covered

The two SQL Server index usage views:

- Usage by plan: sys.dm_db_index_usage_stats
- Usage by index: sys.dm_db_index_operational_stats

Why they're not as accurate as you might suspect:

- Seek doesn't mean one row
- Scan doesn't mean the whole table
- Reads doesn't mean the index was actually read
- 1 write doesn't mean 1 row was updated
- They even reset at unusual times
- Analyze with enough uptime to reflect business processes

