

What the heck is a checkpoint, and why should I care?

Taiob Ali
He/Him/His



SQL Saturday Bangladesh 2022 (#1039)



Taioab Ali

Database Solutions Manager, GMO LLC



<http://sqlworldwide.com/>



/sqlworldwide



@sqlworldwide



taioab@sqlworldwide.com

Data Professional

Microsoft Data Platform MVP. Seventeen years working with Microsoft Data Platform. Microsoft and MongoDB certified. Worked in eCommerce, healthcare, and finance industries.

Giving Back

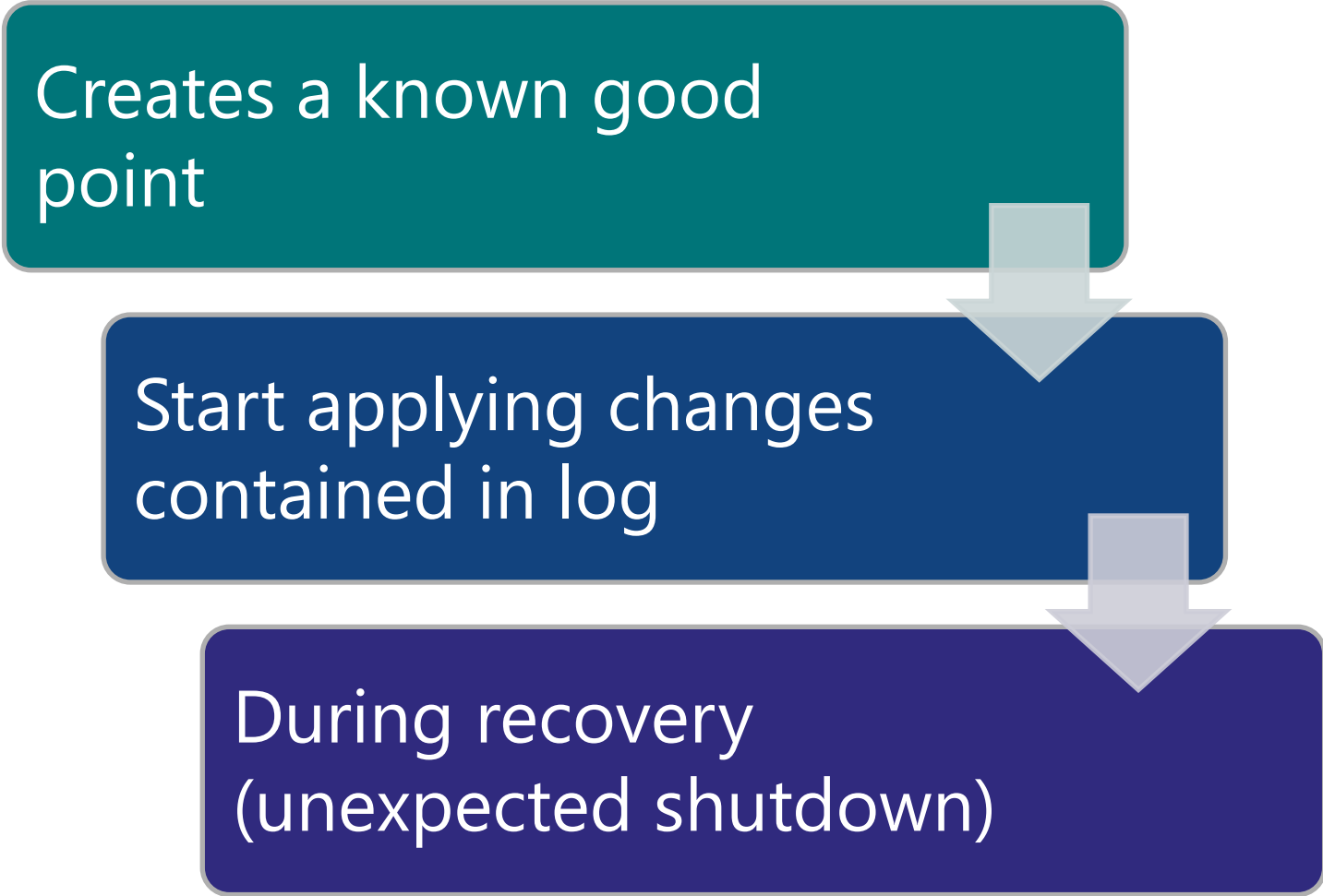
Board member NESQL user group and Founder of DBA virtual group. The organizer of Boston SQL Saturday. A frequent speaker at local and virtual user groups, SQL Saturdays, and Azure conferences.

When Not Working

Running – 1x26.2 and 30+x 13.1, Learning US history. Shuttling three kids.

Why Do We Need Checkpoint?

Creates a known good
point



Start applying changes
contained in log

During recovery
(unexpected shutdown)

ACID

Atomicity

Consistency

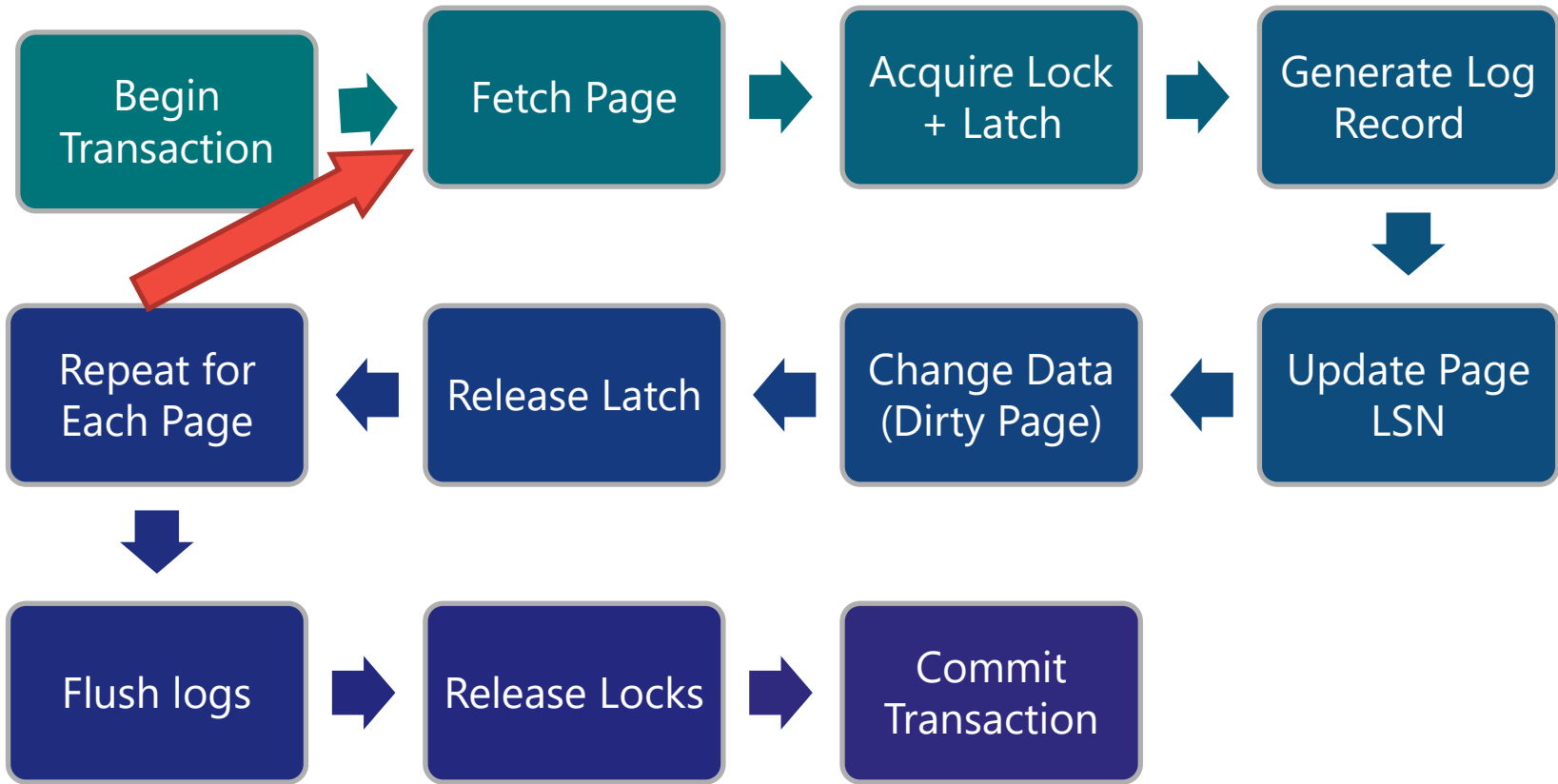
Isolation

Durability

Clean vs Dirty Page



Write Ahead Logging



Allocation Status

GAM (1:2) = ALLOCATED

SGAM (1:3) = NOT ALLOCATED

PFS (1:1) = 0x60 MIXED_EXT ALLOCATED 0_PCT_FULL

DIFF (1:6) = NOT CHANGED

ML (1:7) = NOT MIN_LOGGED

Slot 0 Offset 0x60 Length 73

Record Type = PRIMARY_RECORD

Record Size = 73

Memory Dump @0x000000B22

0000000000000000:

.....

0000000000000014:

0.....

0000000000000028:

.....

000000000000003C:

.....0

Did You Forget
to Flush Dirty
Pages!!

Slot 0 Column 1 Offset 0x4 Length 1

auid = 72057594037993472

Slot 0 Column 2 Offset 0xc Length 1 Length (physical) 1

type = 1

Slot 0 Column 3 Offset 0xd Length 8 Length (physical) 8



Lazy Writer

Eager Writer

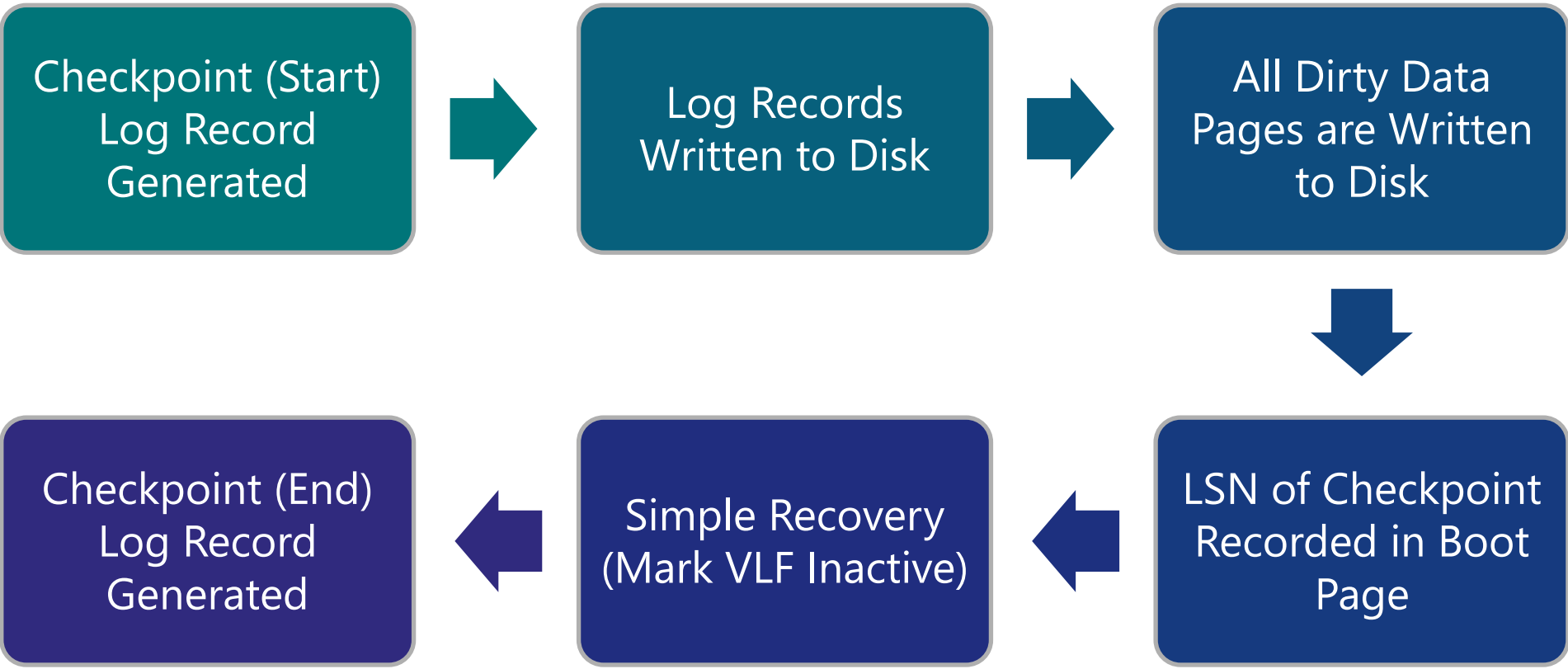
Checkpoint

<https://www.alphr.com/features/386074/out-of-disk-space-storage-options-for-your-media/>

Lazy Writer

Eager Writer

Back to Checkpoint



- IO Size
 - 1 MB
 - Pre 2016: 256 KB
- Throttle threshold
 - 50ms
 - Pre 2016: 20ms
- Startup parameter (-k chkpoint speed in MB)

DEMO

Checkpoint + Log Records





Automatic

Manual

Internal

Indirect

-Introduced SQL 2012

-Default in SQL 2016

<https://www.alphr.com/features/386074/out-of-disk-space-storage-options-for-your-media/>

Automatic

- “Recovery Interval” default zero
 - What does this value mean?
- How reliable is this setting?
- When do you change default?
- Changing to over 60 min require “OVERRIDE”

Manual

- Transact-SQL CHECKPOINT
- *'checkpoint_duration'* seconds to complete
 - Not guaranteed
 - Advanced option
- When do you use a Manual Checkpoint?
- When do you use *'checkpoint_duration'* ?



Aaron Bertrand

@AaronBertrand

Replying to @SqlWorldWide

Before a planned failover or other maintenance, because it can really put a dent in recovery time (especially if your system isn't using indirect yet).

Also during performance testing or batch loading on simple (to minimize log impact).

Never used `checkpoint_duration`.

[#sqlhelp](#)



Travis Page

@pagerwho

Replying to @SqlWorldWide

I've used it during performance testing to ensure that when I issue `DBCC DROPCLEANBUFFERS` anything writes are fully cleared out.

I've also used it for t-log shrinking. Usually this is because of high VLF numbers or something blew up the logs, not a regular occurrence.



L_N__

@sql_handLe

Replying to @SqlWorldWide

[#sqlhelp](#)

if indirect checkpoint (and no [#sqlserver](#) 2019++ ADR), issue manual checkpoint when txlog is close to full and filling rapidly to prevent `log_reuse_wait_desc = 'oldest_page'` and a full txlog due to txlog hotspot in a small number of data pages. `no checkpoint_duration`.

Internal

- Guarantee disk image match what is in logs
- When?
 - Backup
 - Snapshot (DBCC Checkdb)
 - Stop SQL Server Engine
 - Add/Remove a database file

Indirect

- Introduced in SQL 2012
- Uses number of dirty pages
- Keep number of dirty page below threshold
- Set by "*target recovery time*"
 - Default is 60 seconds
- Set per database
 - Persist with backup/restore

Indirect

- SQL 2016
 - Default behavior for new databases
 - Including Model and Tempdb
- SQL 2019
 - Fixed '*non-yielding scheduler errors*'

Recovery Writer

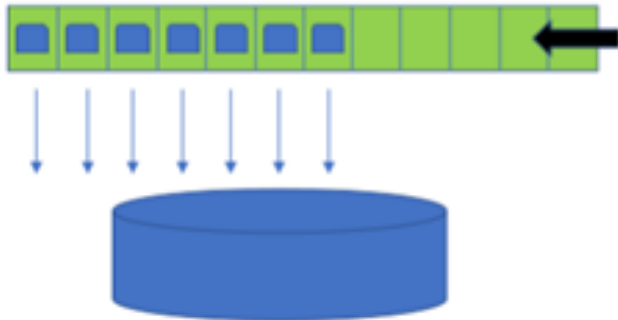
When a work for flush is enqueued

Do While

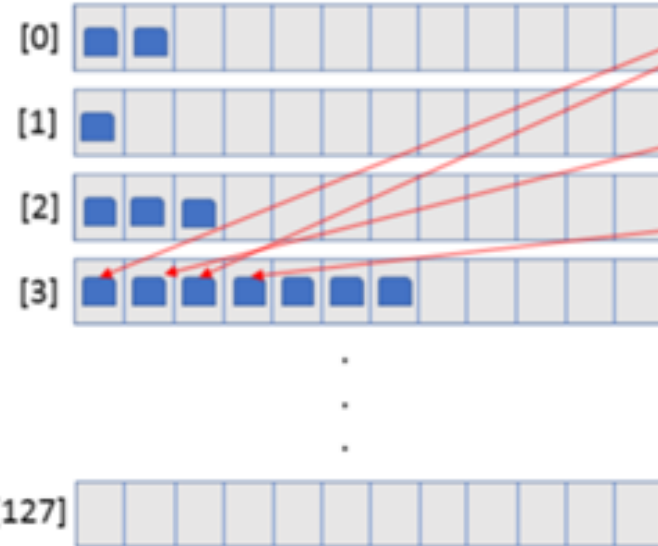
1. Find the longest DPList
2. **Collect Page IDs (under spinlock)**
3. Sort the pages
4. Write the pages

Until the target recovery time is met

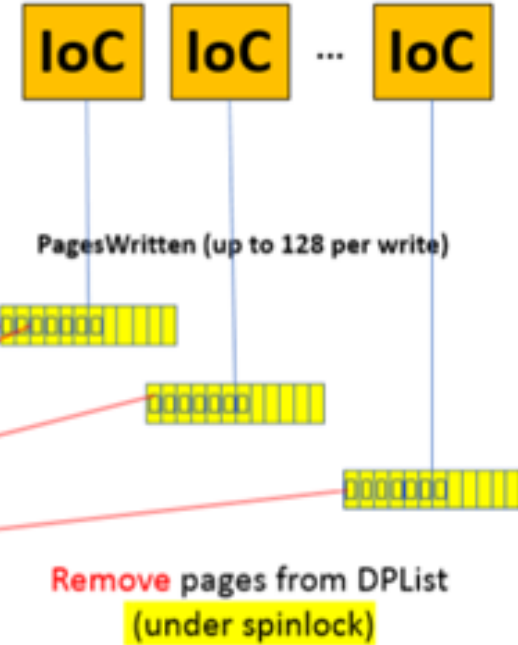
PagesToWrite[128]



DPLists[128]



IO Completion routines



@sqlworldwide

Monitor Checkpoint

- Performance Objects
 - Buffer Manager: Checkpoint pages/sec
- Extended Events
 - sqlserver.checkpoint_begin
 - sqlserver.checkpoint_end
- Trace Flag
 - 3502: writes to error log when a checkpoint starts and finishes
 - 3504: writes to error log information about what is written to disk
 - 3605: allows trace prints to go to the error log

Checkpoint tempdb

DEMO

- Monitor Checkpoint
- Simple Recovery Behavior
- Automatic vs Indirect



Reference

- [Database Checkpoints \(SQL Server\)](#)
- [SQL Server Transaction Log Architecture and Management Guide](#)
- [How do checkpoints work and what gets logged](#) by Paul Randal
- [What does checkpoint do for tempdb?](#) by Paul Randal
- [Changes in SQL Server 2016 Checkpoint Behavior](#) by Mike Ruthruff
- [Indirect Checkpoint and tempdb – the good, the bad and the non-yielding scheduler](#) by Parkshit Savjani
- ["0 to 60" : Switching to indirect checkpoints](#) by Aaron Bertrand
- [SQL Server Checkpoint Monitoring with Extended Events](#) by Aaron Bertrand
- [More Reasons to Enable SQL Server Indirect Checkpoints](#) by Aaron Bertrand
- [How do I interpret the log when I run DBCC TRACEON \(3502, 3504, 3605, -1\)](#)
- [How It Works: Bob Dorr's SQL Server I/O Presentation](#) by Bob Dorr^{@sqlworldwide}



@sqlworldwide



linkedin.com/in/sqlworldwide



sqlworldwide.com



taio@sqlworldwide.com

