Buffer Pool Extension

How it works?

Murilo Miranda
@murilocmiranda | Pythian
murilo.miranda@gmail.com

About me

Murilo Miranda

Lead Database Consultant @ Pythian

- http://www.sqlshack.com/author/murilo-miranda/
- http://www.pythian.com/blog/author/murilo/
- @murilocmiranda
- http://pt.linkedin.com/in/murilomiranda/

MVP Microsoft® Most Valuable Professional

Microsoft CERTIFIED Solutions Expert

Data Platform

Pythian love your data

Agenda

- What’s buffer pool?
- Storage vs. Memory
- Buffer Pool Extension
  - Benefits.
  - How it works?
  - Considerations/Recommendations.
- In-Memory OLTP Challenge
- Troubleshooting

Starting from the basics…

ABOUT BUFFER POOL

Buffer Pool

VAS: Range of virtual addresses available for a process in memory.

- SQL Server's Virtual Address Space (VAS) is divided into two regions:
  - Buffer Pool.
  - Other components (a.k.a The Rest :).
- Most of the SQL Server VAS is occupied by Buffer Pool.

Buffer Pool

- Why Buffer Pool Exists?
  - Disk reads and writes are **resource-intensive** operations.

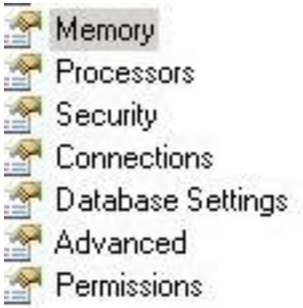
  - The goal of **Buffer Pool** is minimize disk I/O.
    - Pages from a database are held in memory.

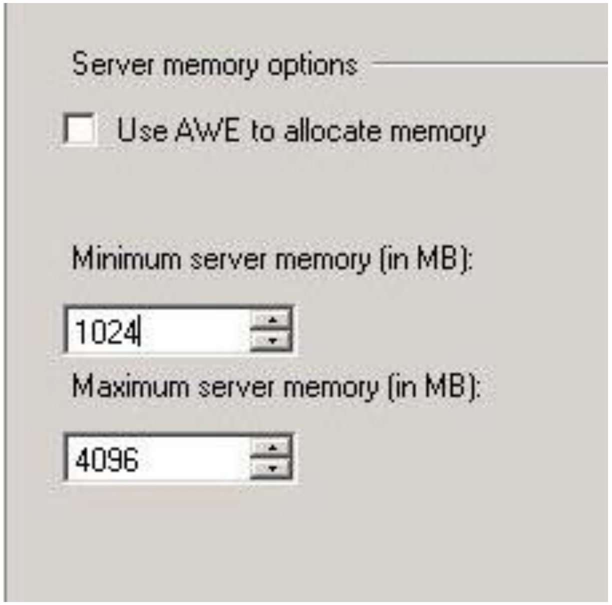
↓ Physical reads

↑ Buffer Pool

Buffer Pool

- Before SQL Server 2012 **MAX** and **MIN memory setting** were used to define the size of buffer pool…





Server memory options

□ Use AWE to allocate memory

Minimum server memory (in MB):

1024

Maximum server memory (in MB):

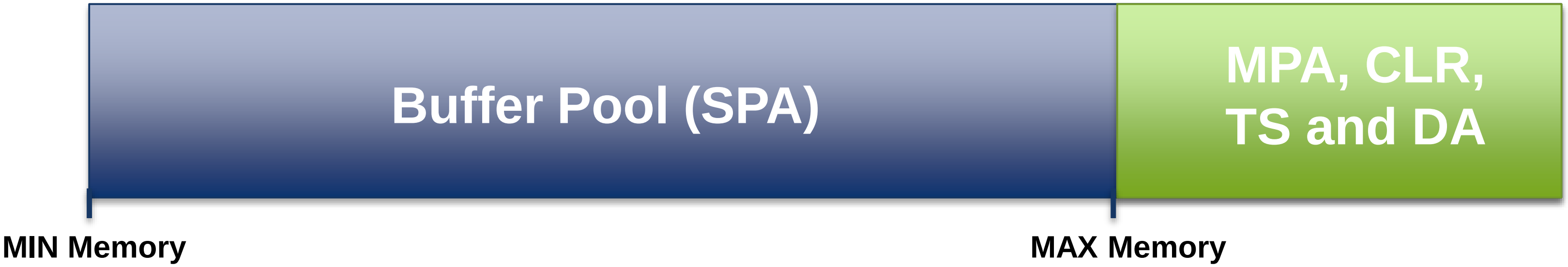
4096

Buffer Pool

- … this changed from SQL Server 2012.
- Now, MAX/MIN Memory settings affects more than buffer pool size, including:
  - Multi-page Allocations (MPA)
  - CLR Allocations.

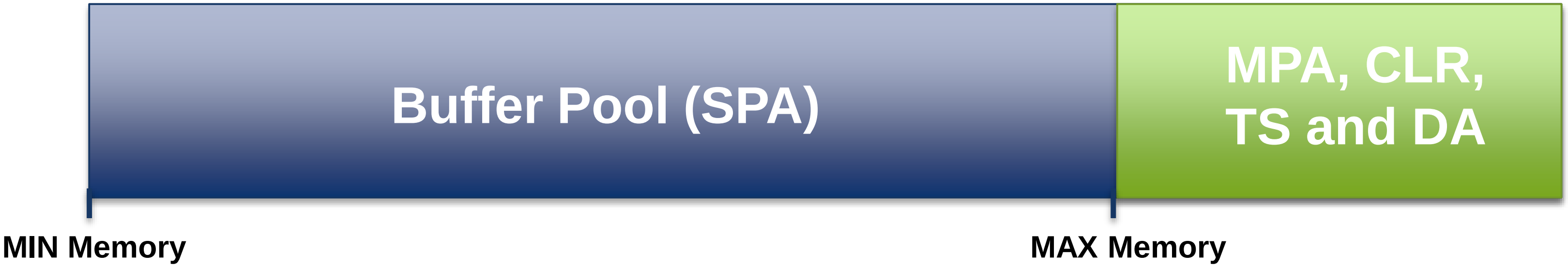
Buffer Pool

- Before SQL Server 2012:

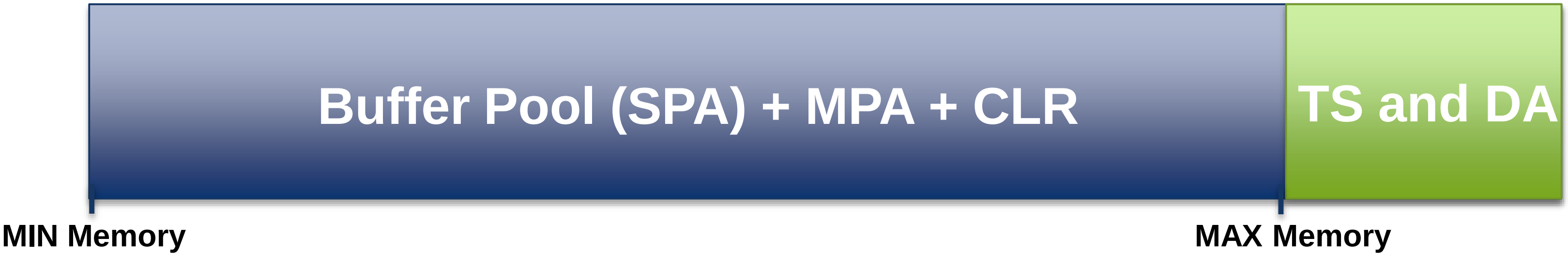


Buffer Pool

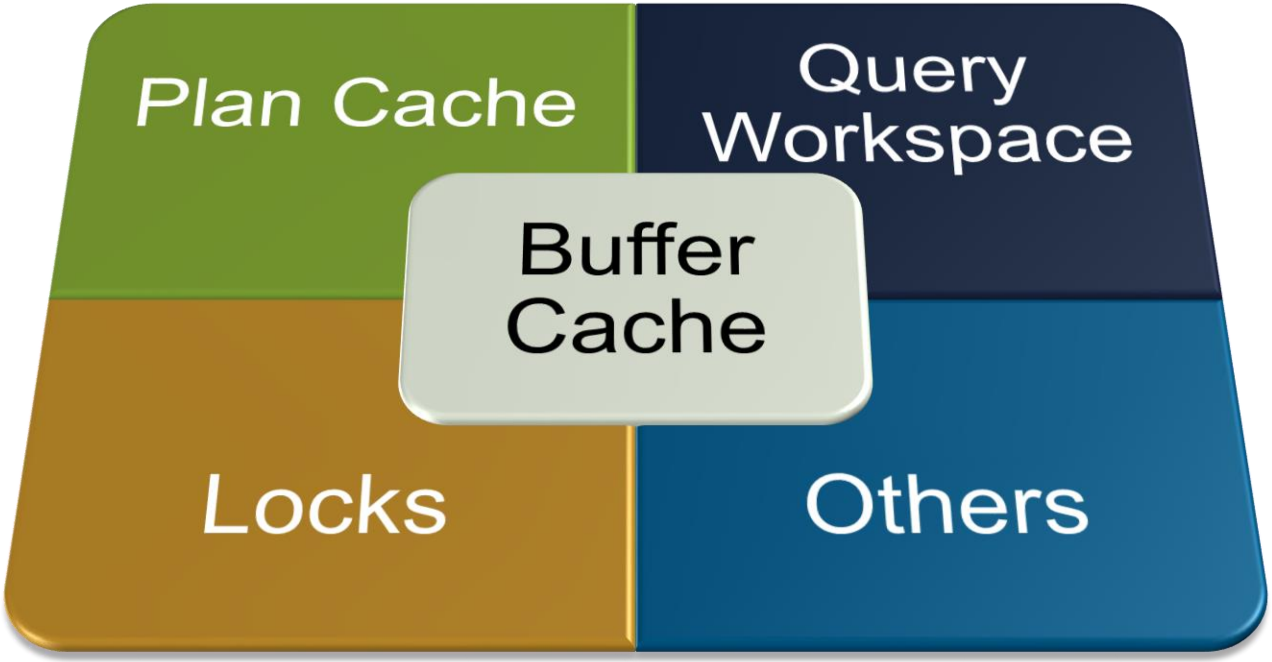
- Before SQL Server 2012:



- From SQL Server 2012:



Buffer Pool



Buffer Pool

- Dirty vs. Clean
  - A page read from disk into memory is a **Clean Page** - while it’s not modified.
  - Once modified, it id marked as dirty – **Dirty Page**.

Buffer Pool

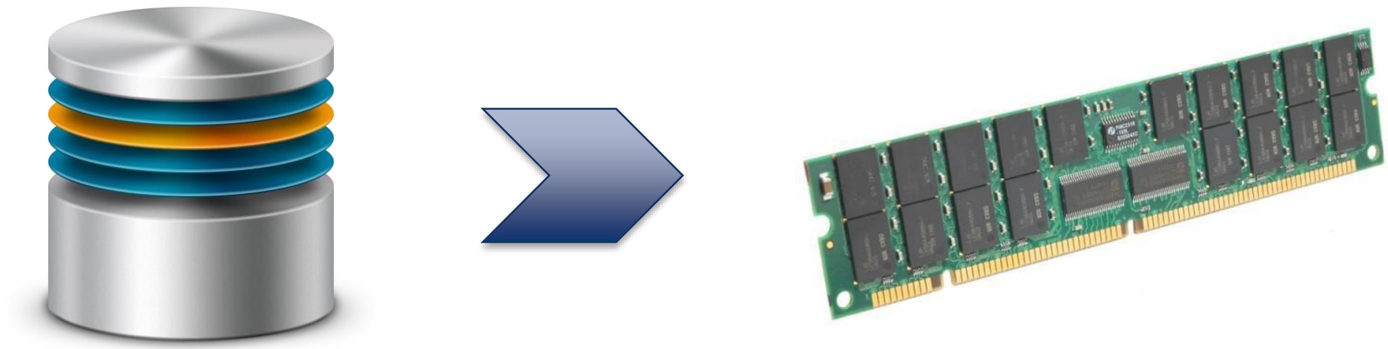
- Dirty vs. Clean
  - A page read from disk into memory is a **Clean Page** - while it's not modified.
  - Once modified, it id marked as dirty – **Dirty Page**.

  - We can flush clean pages by using:

DBCC DROPCLEANBUFFERS

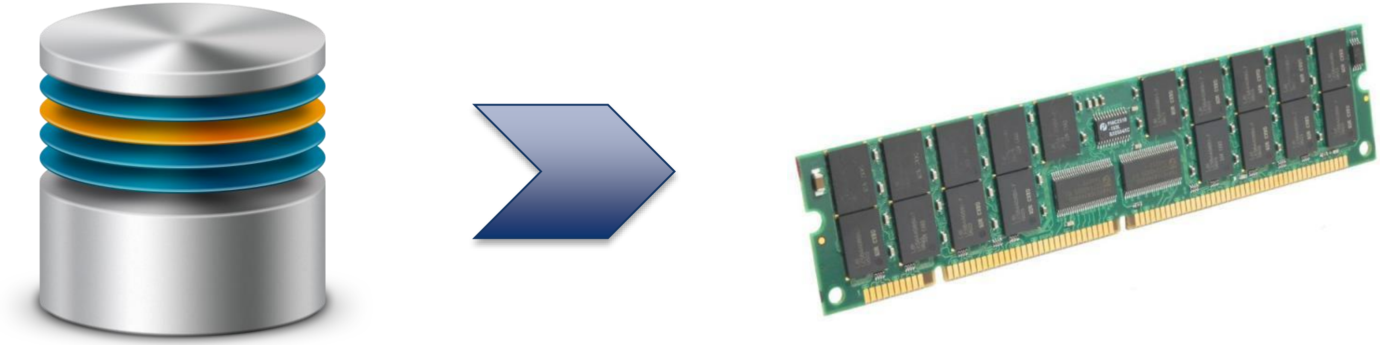
Buffer Pool

- Buffer Pool stores “buffers”.
  - A buffer is an 8-KB page in memory.



Buffer Pool

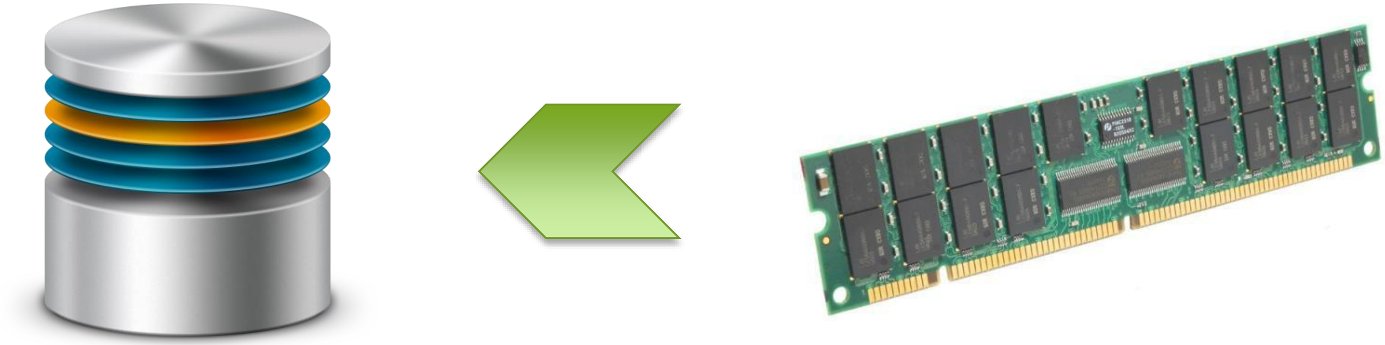
- Buffer Pool stores “buffers”.
  - A buffer is an 8-KB page in memory.



- A buffer is written back to disk only if it is **modified**.

Buffer Pool

- Crash Recovery
  - Dirty Pages are written to the disk periodically.
    - The “**Lazy Writting**”, “**Eager Writing**” and “**Checkpoint**” processes are responsible for this.

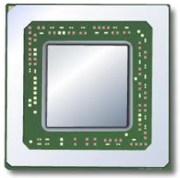
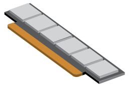
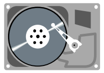


DEMO Time!

Comparison

STORAGE VS. MEMORY

Storage Vs. Memory

STORAGE			RELATIVE ACCESS TIME	
	L1 cache	▶	Fractions of a second	Volatile
	L2 cache	▶	1 second	
	L3 cache	▶	Few seconds	
	DRAM	▶	Few minutes	
	SSD	▶	Few hours -1 day	Non-Volatile
	HDD	▶	3-5 years	

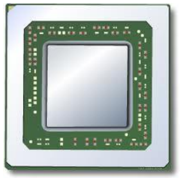
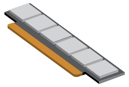
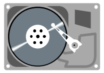
HDD are a storage architecture, not performance

Storage Vs. Memory

STORAGE		RELATIVE ACCESS TIME	
L1 cache	▶	Fractions of a second	Volatile
L2 cache	▶	1 second	
L3 cache	▶	Few seconds	
DRAM	▶	Few minutes	
SSD	▶	Few hours -1 day	Non-Volatile
HDD	▶	3-5 years	

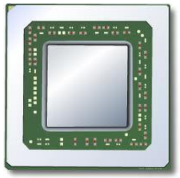
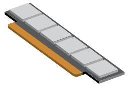
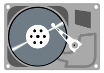
HDD are a storage architecture, not performance

Storage Vs. Memory

STORAGE			RELATIVE ACCESS TIME	
	L1 cache	▶	Fractions of a second	Volatile
	L2 cache	▶	1 second	
	L3 cache	▶	Few seconds	
	DRAM	▶	Few minutes	
	SSD	▶	Few hours -1 day	Non-Volatile
	HDD	▶	3-5 years	

HDD are a storage architecture, not performance

Storage Vs. Memory

STORAGE			RELATIVE ACCESS TIME	
	L1 cache	▶	Fractions of a second	Volatile
	L2 cache	▶	1 second	
	L3 cache	▶	Few seconds	
	DRAM	▶	Few minutes	
	SSD	▶	Few hours -1 day	Non-Volatile
	HDD	▶	3-5 years	

HDD are a storage architecture, not performance

Storage Vs. Memory

	STORAGE	RELATIVE ACCESS TIME	
	L1 cache	Fractions of a second	Volatile
	L2 cache	1 second	
	L3 cache	Few seconds	
Buffer Pool	**DRAM**	Few minutes	
Buffer Pool Extension	**SSD**	Few hours -1 day	Non-Volatile
Storage	**HDD**	3-5 years	

HDD are a storage architecture, not performance

Introducing…

BUFFER POOL EXTENSION (BPE)

Buffer Pool Extension

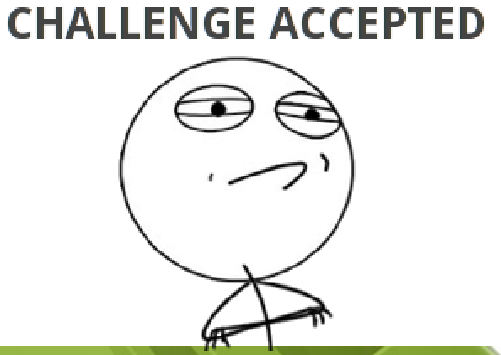
- **Introduction**: SQL Server 2014.

Buffer Pool Extension

- **Introduction**: SQL Server 2014.
- **Mission**: Create a “hot-area” based on evicted pages from Buffer Pool.

Buffer Pool Extension

- **Introduction**: SQL Server 2014.
- **Mission**: Create a “hot-area” based on evicted pages from Buffer Pool.
- **How**: Using fast non-volatile drives (SSD) to extend buffer pool.



Buffer Pool Extension

- **System requirements**
  - SQL Server 2014.
    - Standard and Enterprise.
    - Only supported for 64-bit SQL Server.
  - A fast disk (SSD).

Microsoft

SQL Server 2014

Buffer Pool Extension

- **Benefits**:
  - **Improves OLTP** query performance.
  - **Transparent.**
    - No application changes are required.
  - **Easy** to setup.
    - Enable online.
  - No **data loss.**
    - Deals with **clean pages** only.

Buffer Pool Extension

- **General Syntax**

```
ALTER SERVER CONFIGURATION SET BUFFER
POOL EXTENSION
 {  ON
            ( FILENAME = 'os_file_path_and_name' , SIZE =
[ KB | MB | GB ] ) | OFF
}
```

Buffer Pool Extension

- **Creation Syntax**

```
ALTER SERVER CONFIGURATION SET BUFFER POOL EXTENSION
 {  ON
          ( FILENAME = 'os_file_path_and_name' , SIZE = [ KB | MB | GB ] )
}
```

Buffer Pool Extension

- **Disable Syntax**

```
ALTER SERVER CONFIGURATION SET BUFFER
POOL EXTENSION OFF
GO
```

Buffer Pool Extension

- **Changing BPE Size**

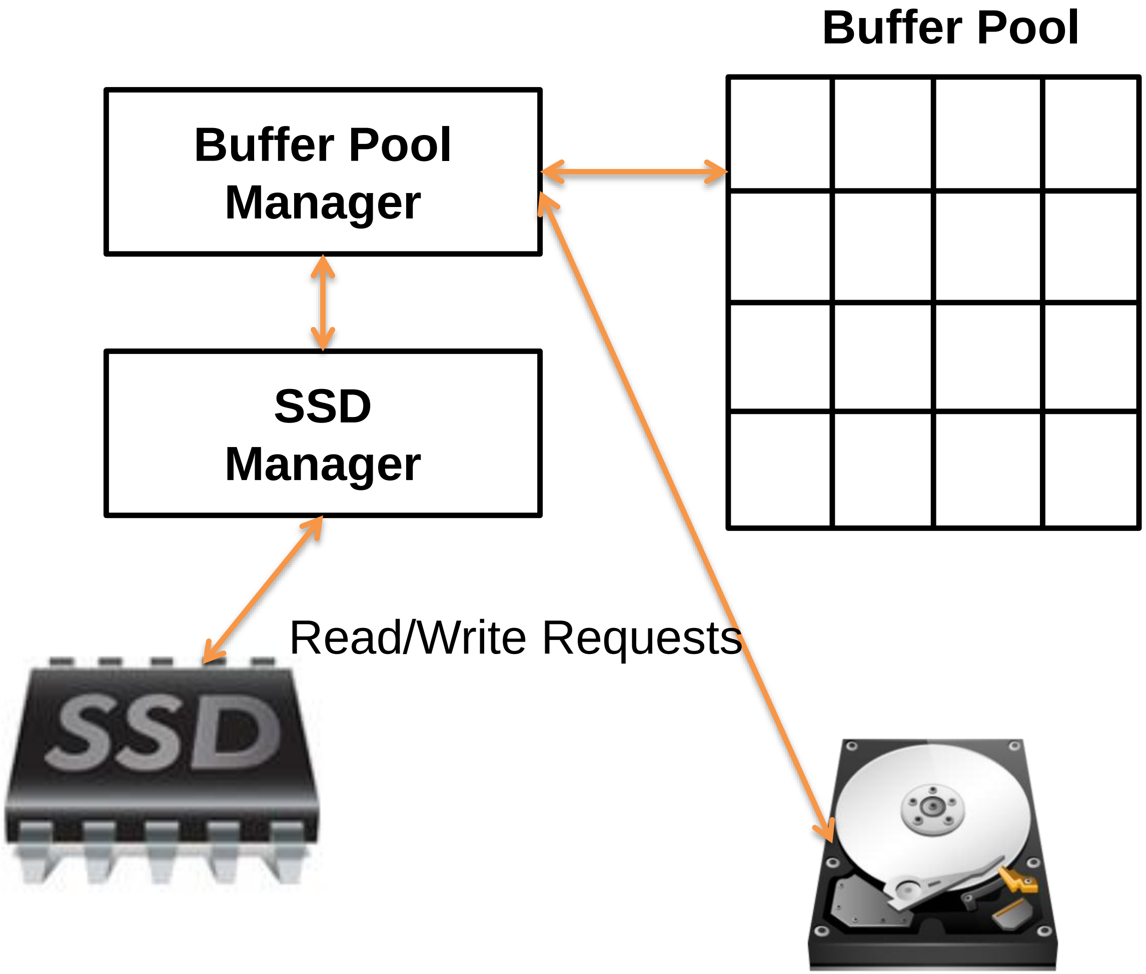
```
ALTER SERVER CONFIGURATION SET BUFFER POOL
EXTENSION OFF
GO

ALTER SERVER CONFIGURATION SET BUFFER POOL
EXTENSION
 {  ON
           ( FILENAME = 'os_file_path_and_name' , SIZE = [ KB
| MB | GB ] ) | OFF
}
```

Buffer Pool Extension

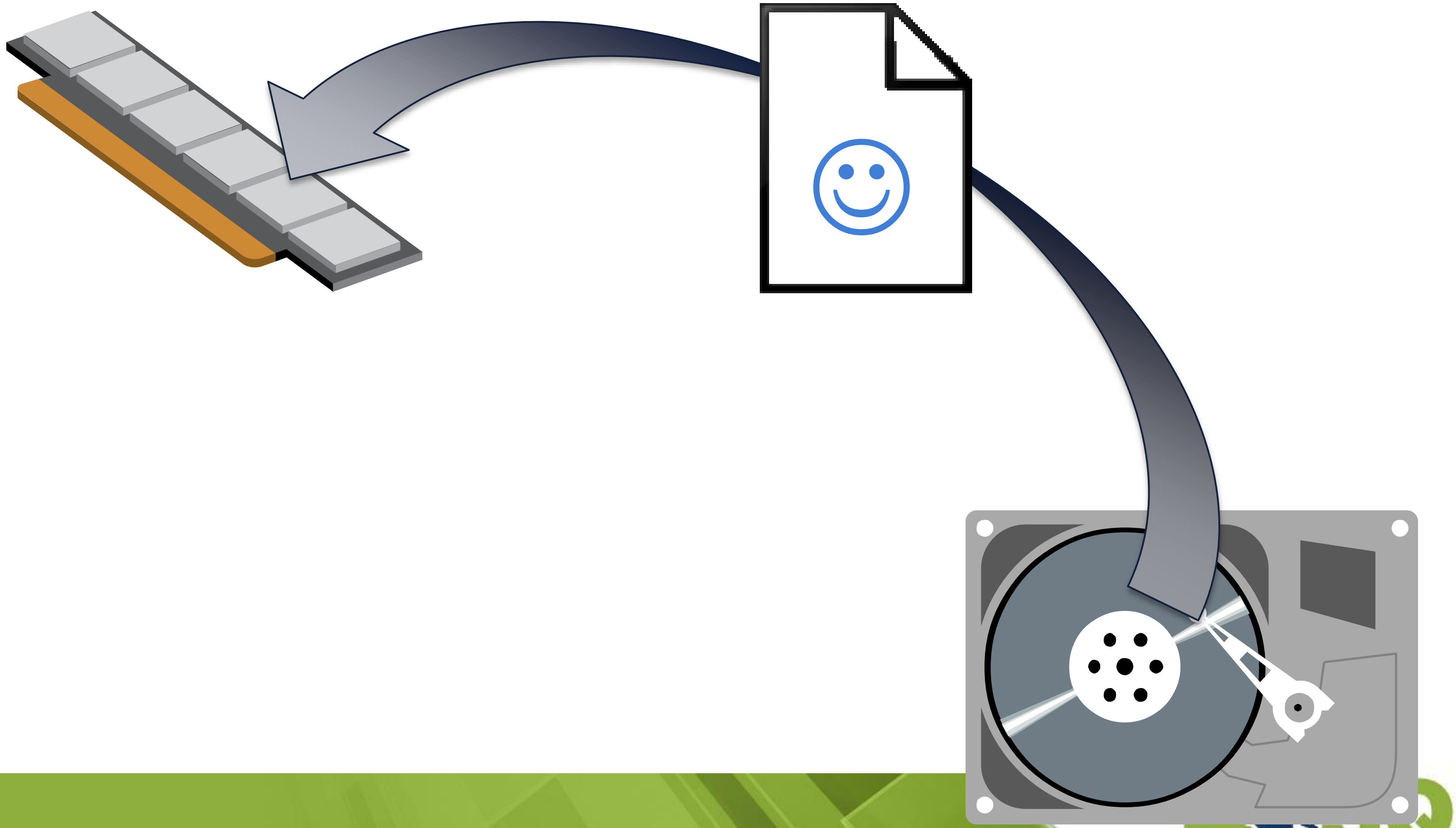
HOW

IT WORKS ?

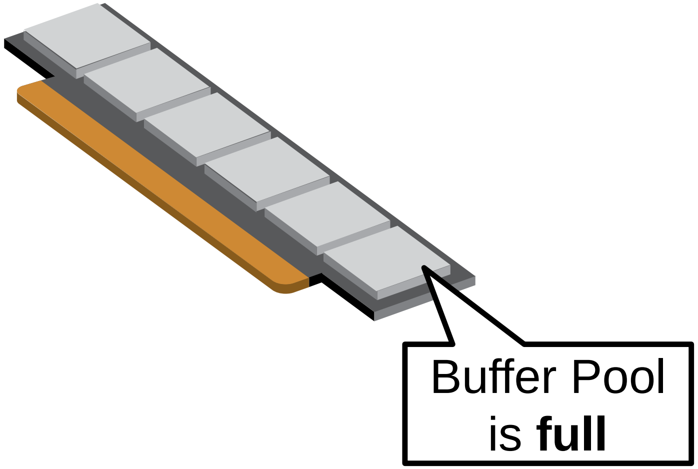
Buffer Pool Extension



Buffer Pool Extension

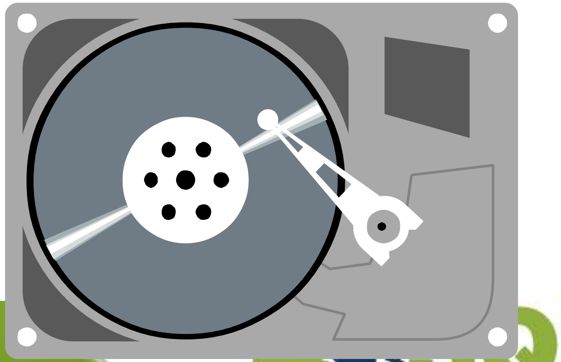


Buffer Pool Extension

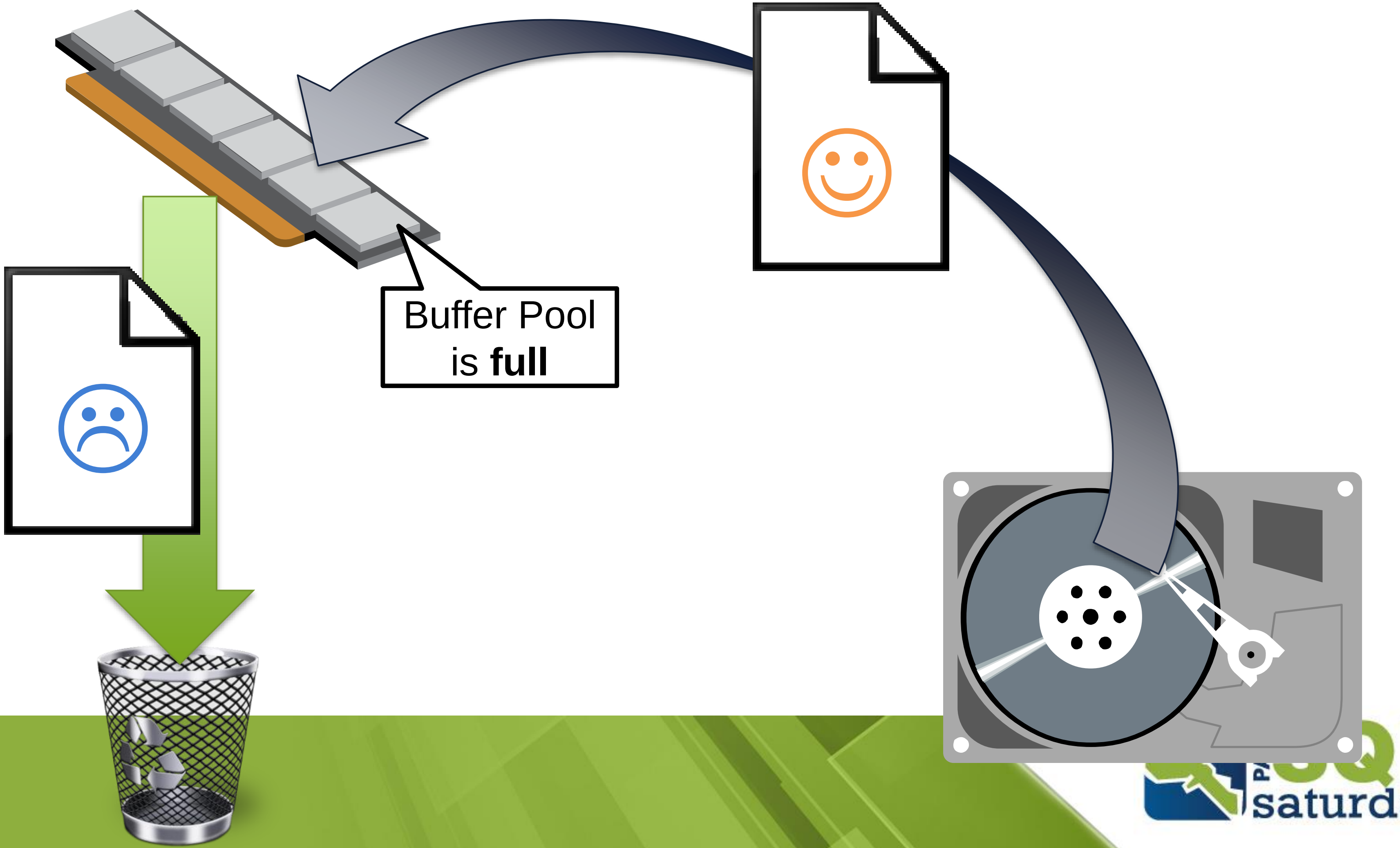


Buffer Pool is **full**

After a while….

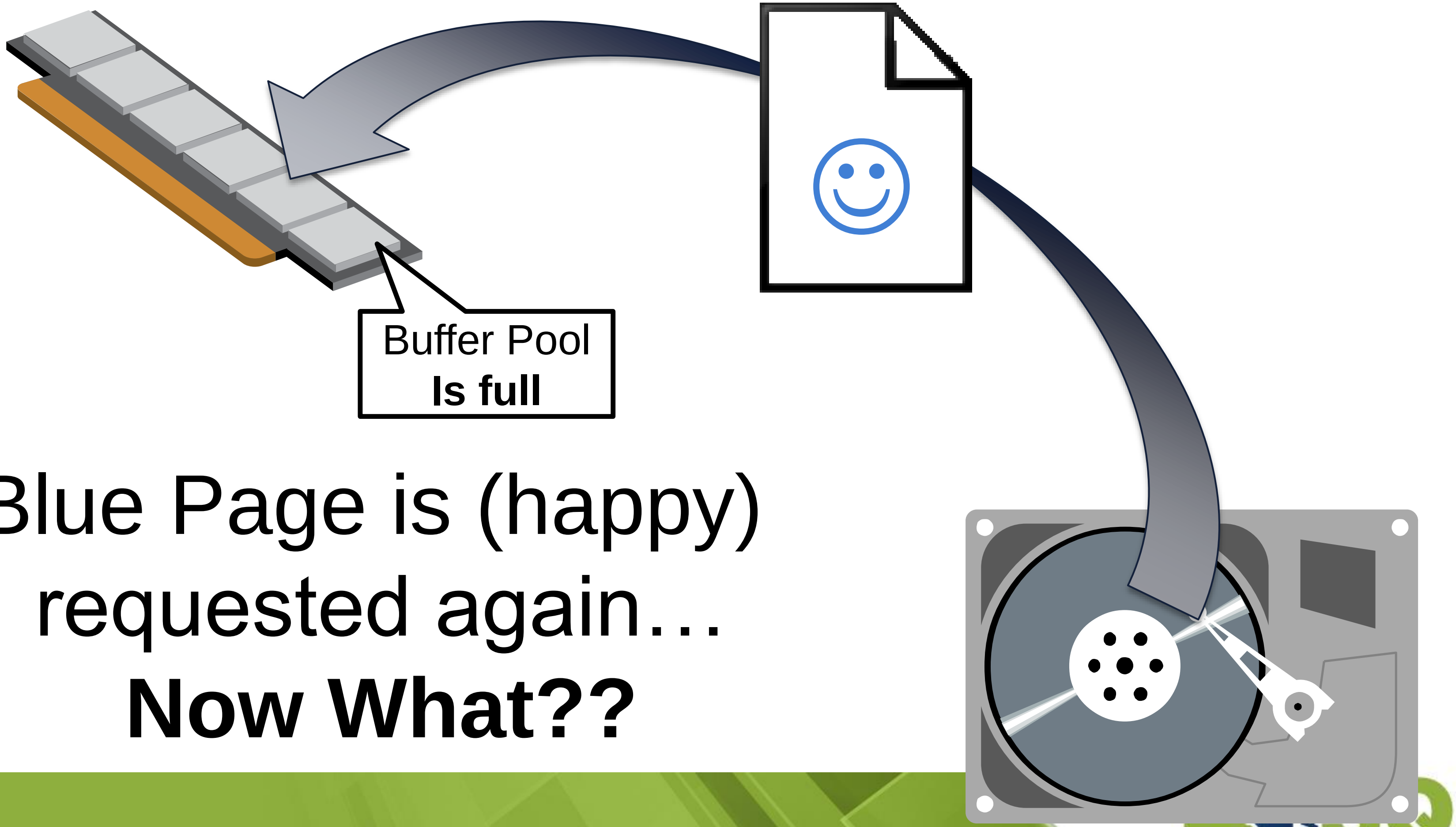


Buffer Pool Extension



Buffer Pool is **full**

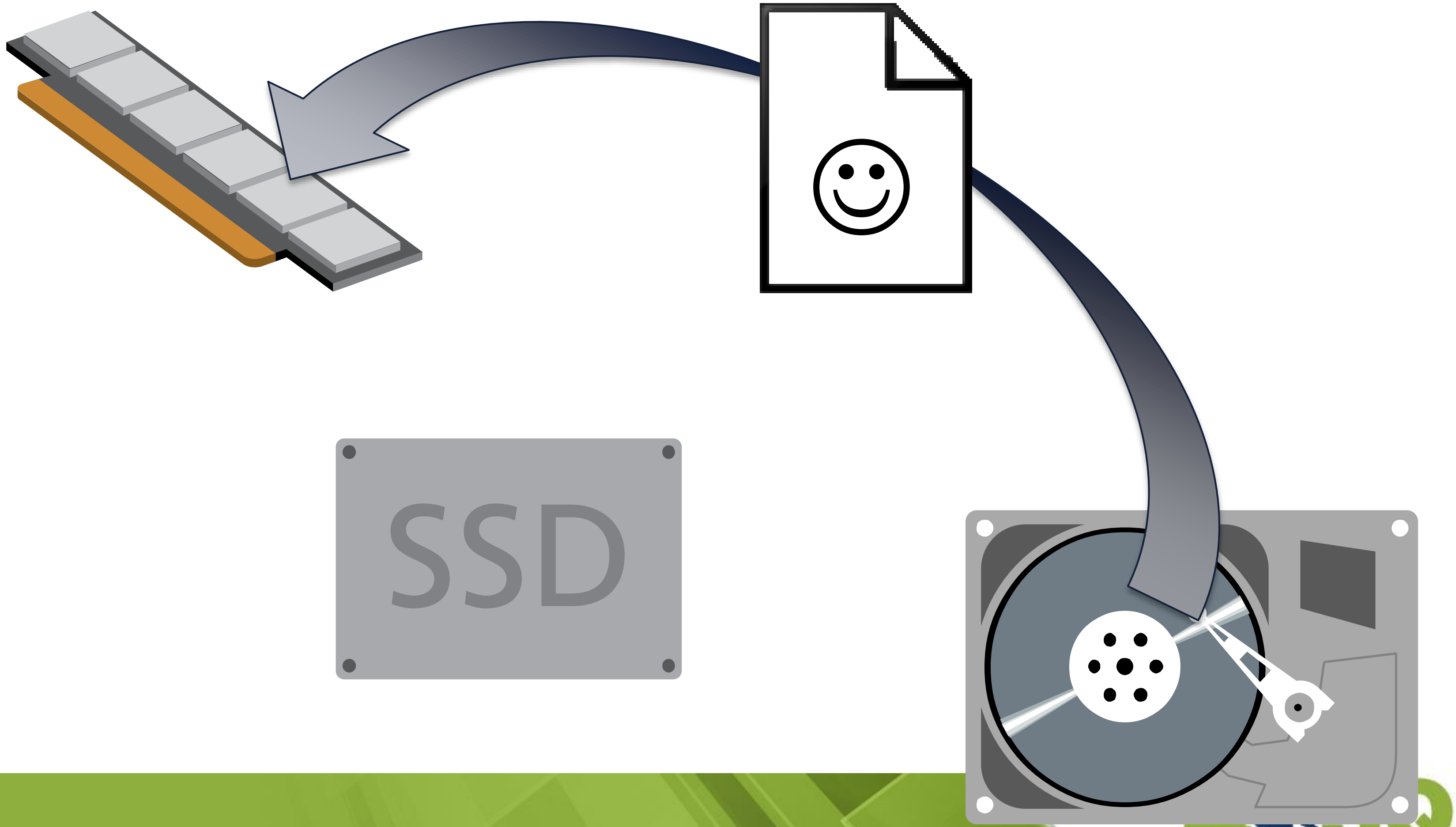
Buffer Pool Extension



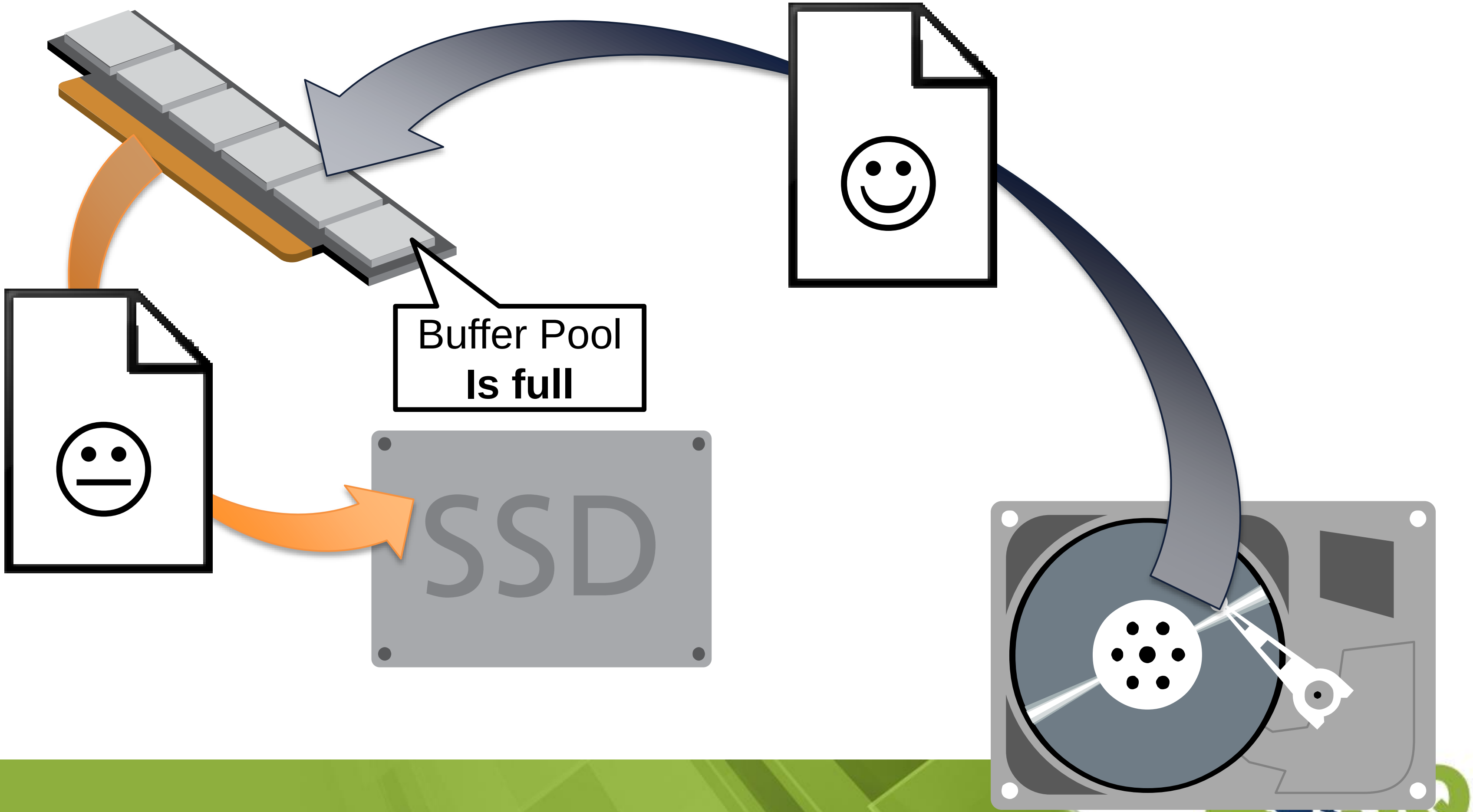
Buffer Pool
Is full

Blue Page is (happy) requested again…
Now What??

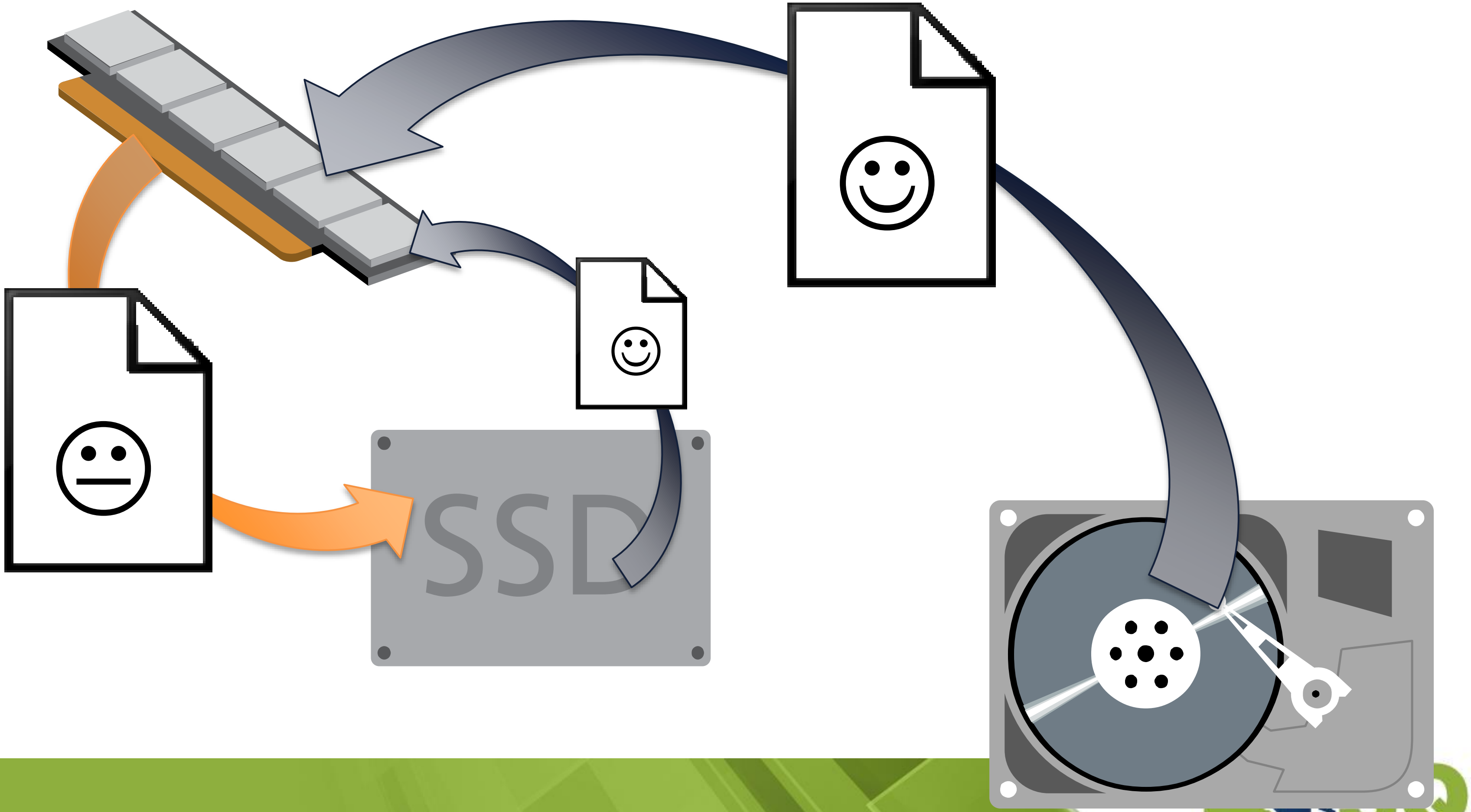
Buffer Pool Extension



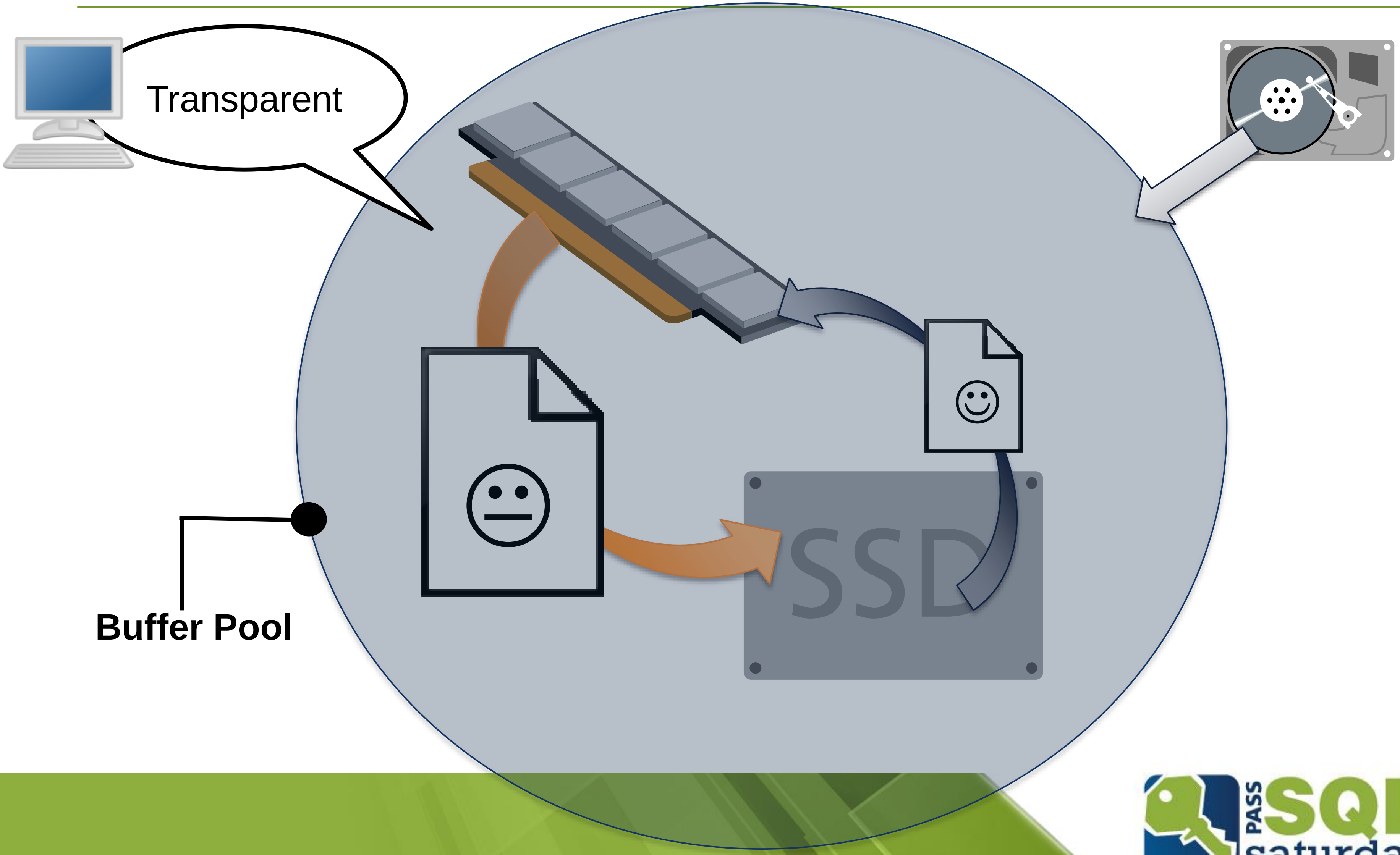
Buffer Pool Extension



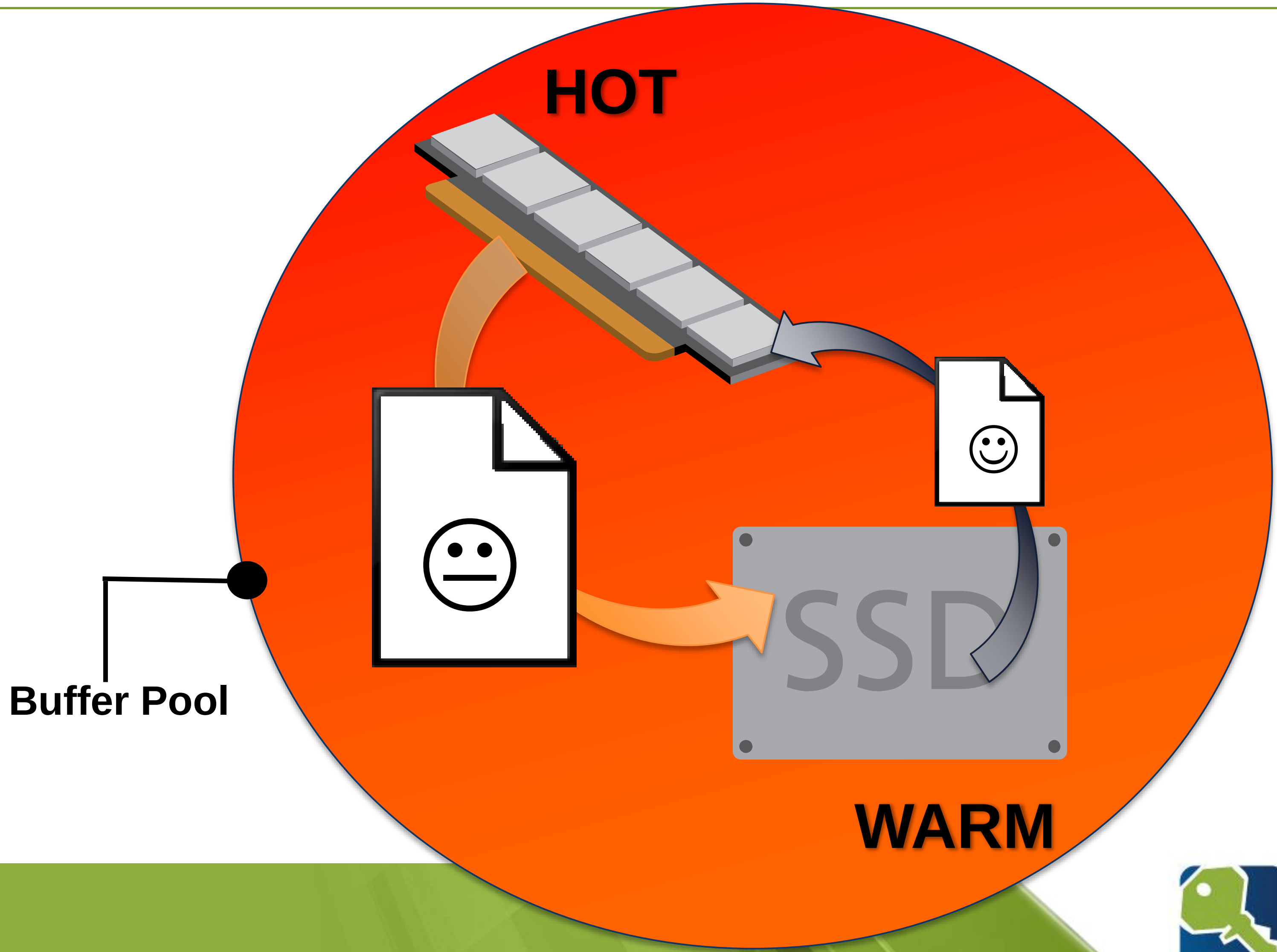
Buffer Pool Extension



Buffer Pool Extension



Buffer Pool Extension

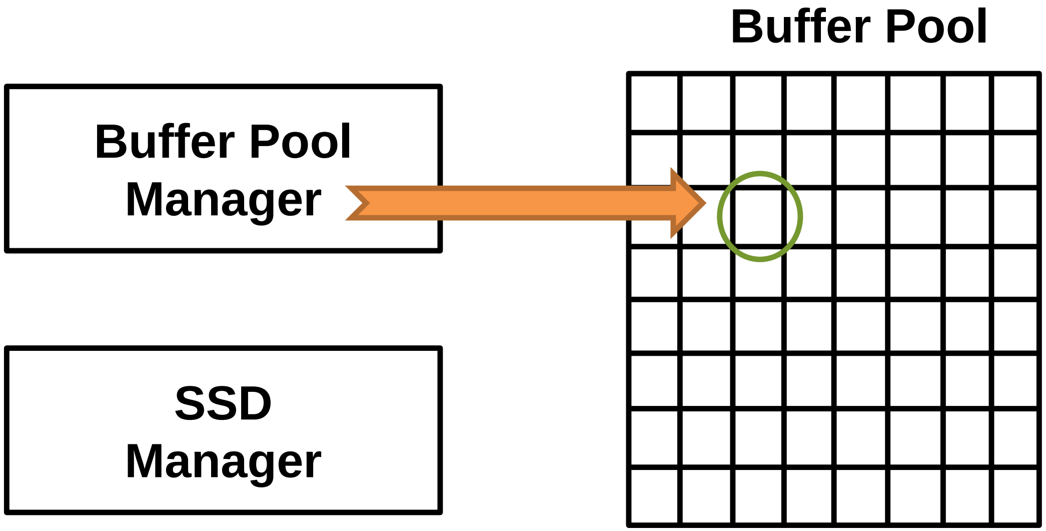


Buffer Pool Extension

How BPE

is **FILLED**?

Buffer Pool Extension



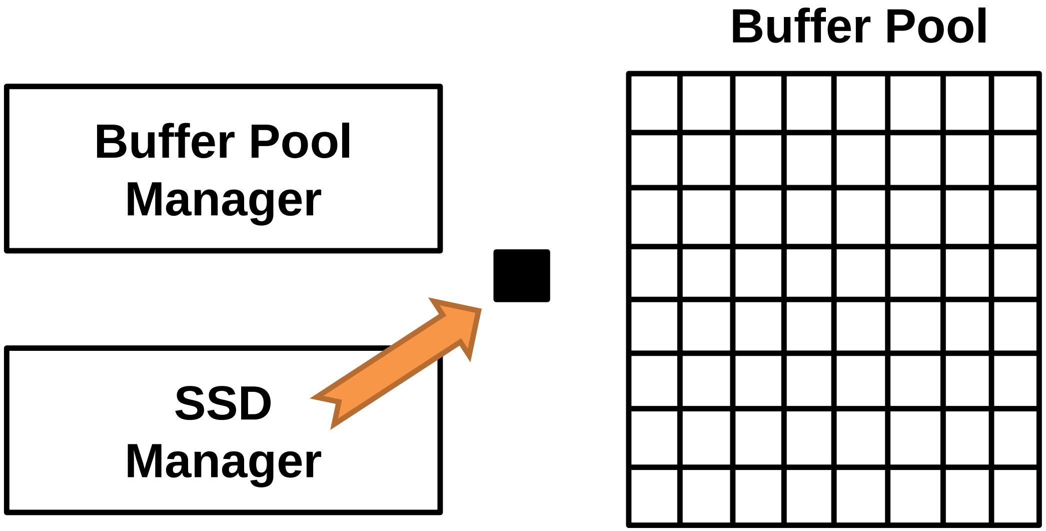
Buffer Pool

Buffer Pool Manager

SSD Manager

BM decides to **evict a page** from the BP.

Buffer Pool Extension



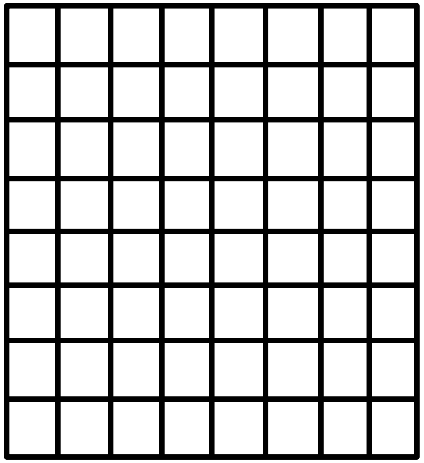
SSD Manager decides whether or not to **cache the page** on BPE.
(**SSD Admission Policy**)

Buffer Pool Extension

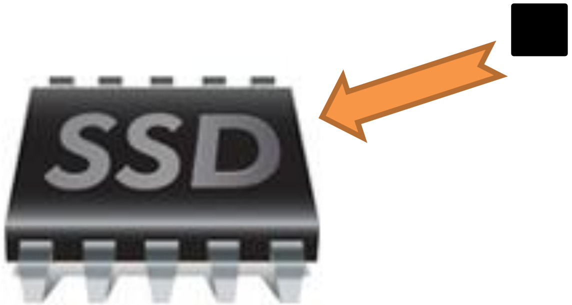
Buffer Pool Manager

SSD Manager

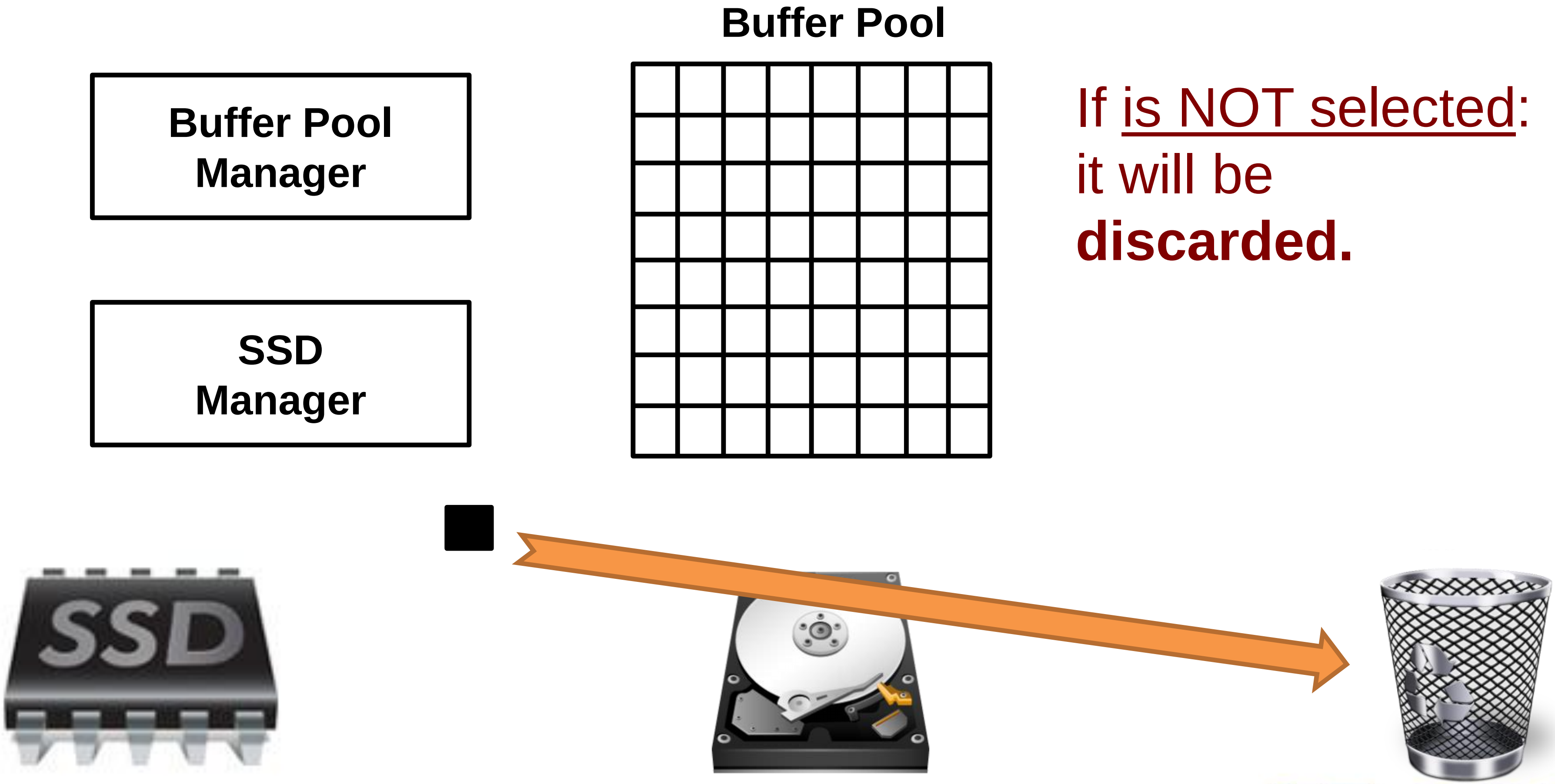
Buffer Pool



If <u>page is selected</u>: it is **written into the BPE** asynchronously.

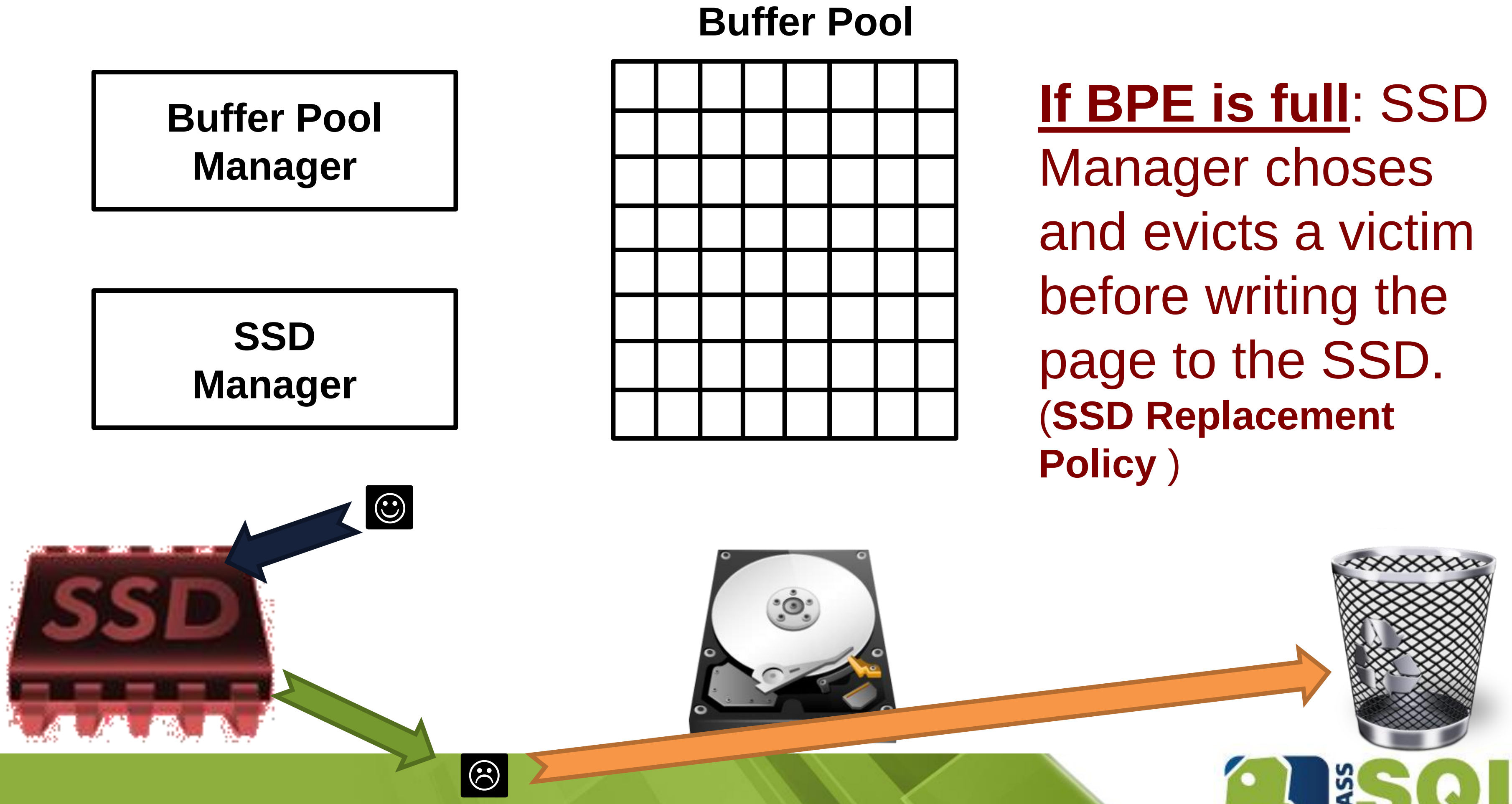


Buffer Pool Extension



If is NOT selected: it will be **discarded.**

Buffer Pool Extension



Buffer Pool

Buffer Pool Manager

SSD Manager

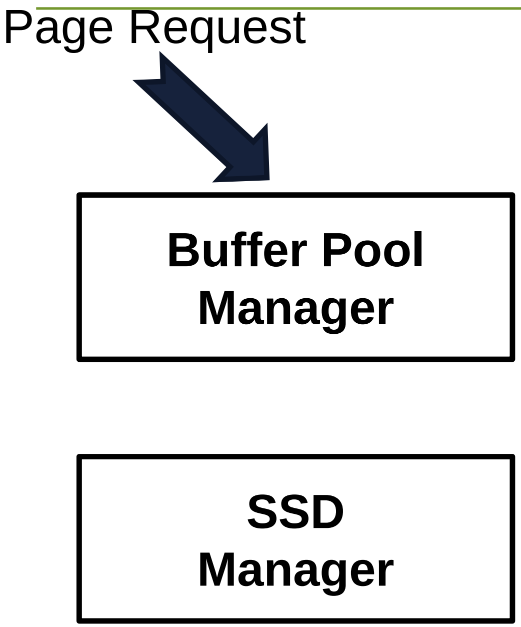
If BPE is full: SSD Manager choses and evicts a victim before writing the page to the SSD. (**SSD Replacement Policy**)

Buffer Pool Extension

How a **page request** works?

(when BPE is enabled)

Buffer Pool Extension



Page Request

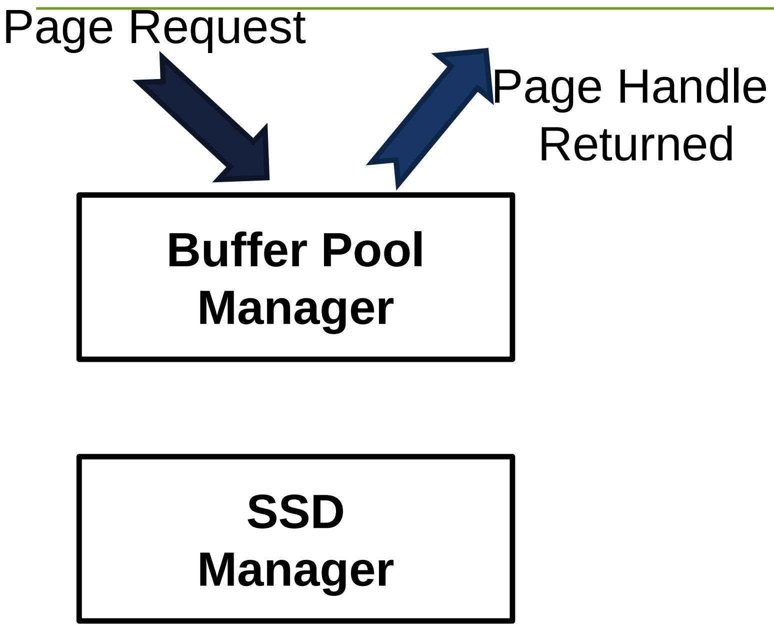
Buffer Pool Manager

SSD Manager

<u>Buffer Manager</u> receives a page request.

☹

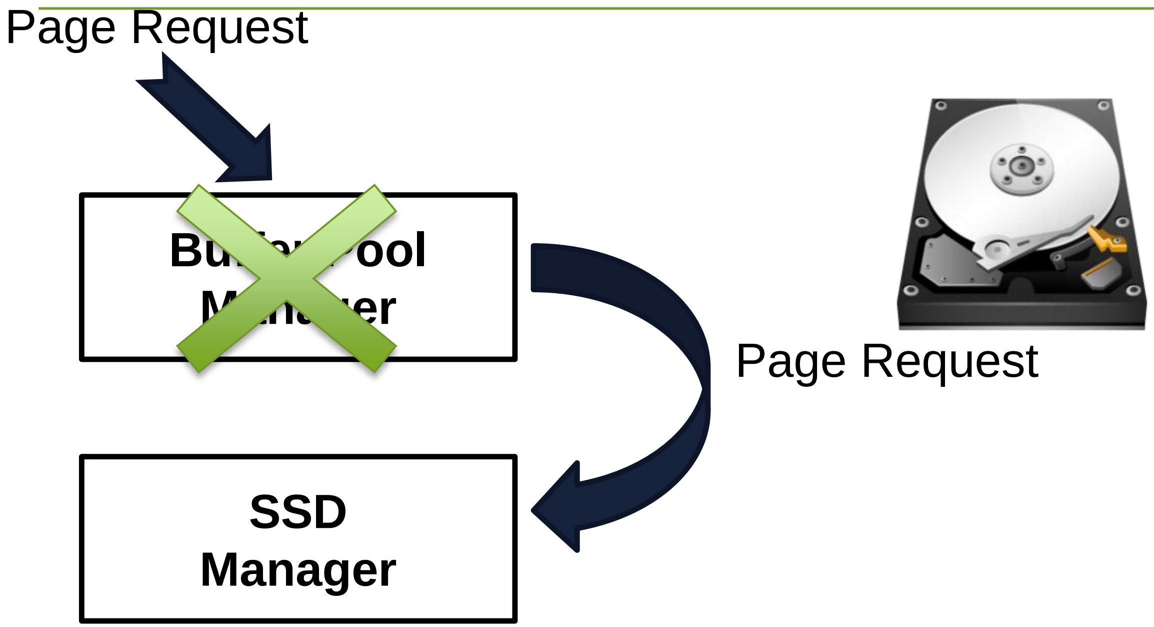
Buffer Pool Extension



If the page **is in the main memory BP**, a page handle is returned.

☹

Buffer Pool Extension

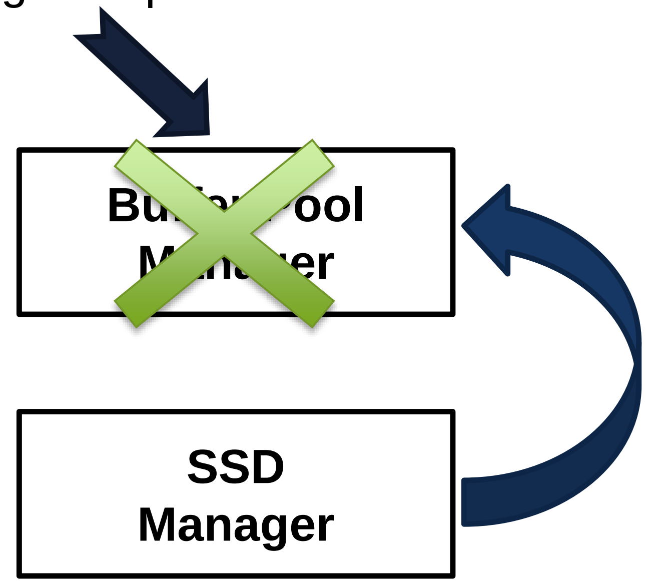


If the page **<u>is NOT in the main memory BP</u>**, a request is sent to the SSD Manager.

☹

Buffer Pool Extension

Page Request



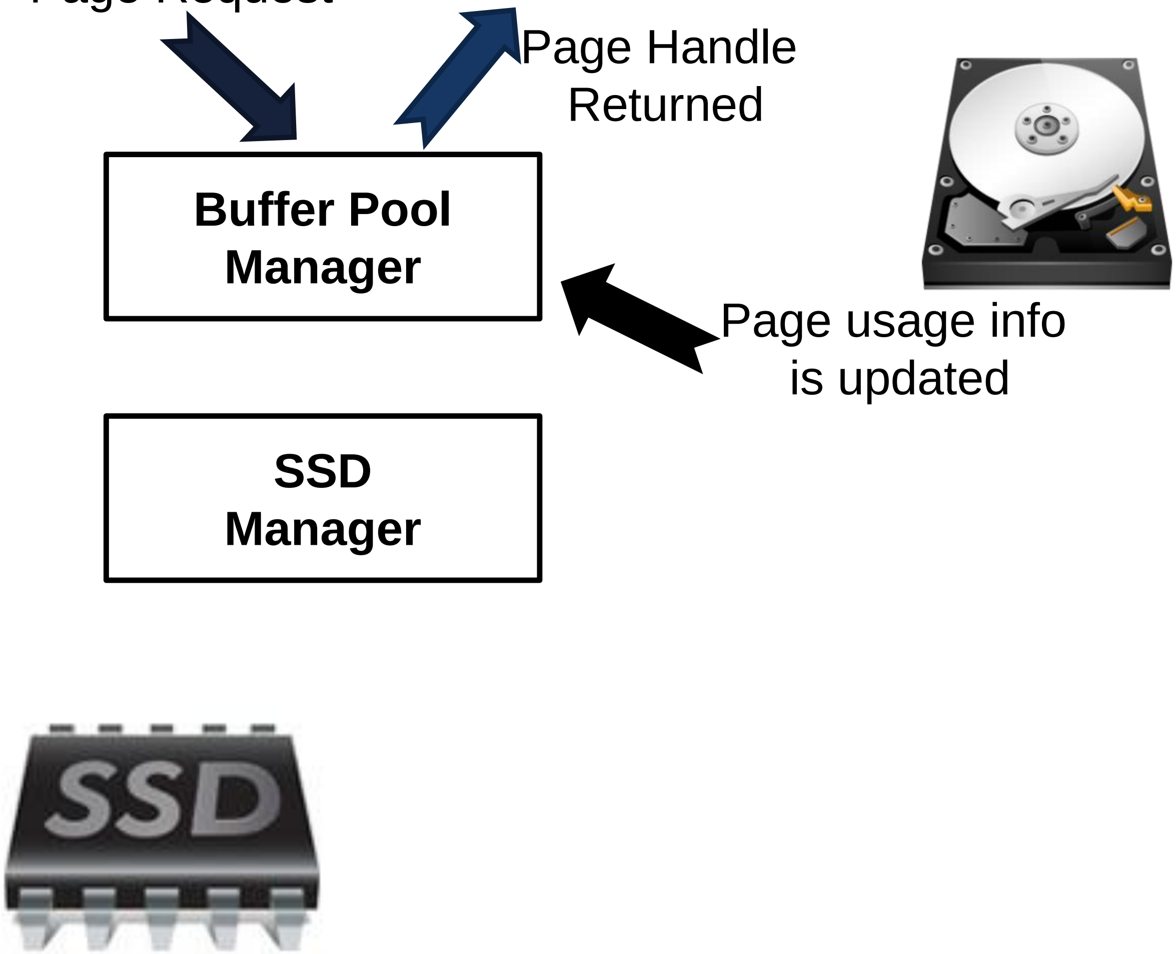
Buffer Pool Manager

SSD Manager

Page is copied to the BP

If the page **is in the main memory BPE**, a request is sent to the SSD Manager.

Buffer Pool Extension



Page Request

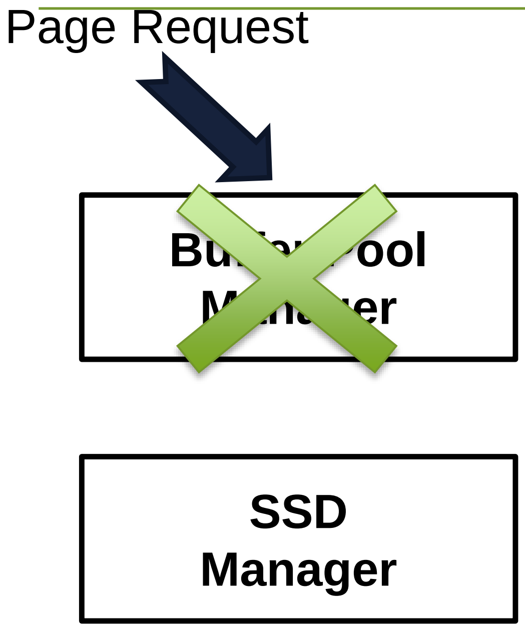
Page Handle Returned

Buffer Pool Manager

Page usage info is updated

SSD Manager

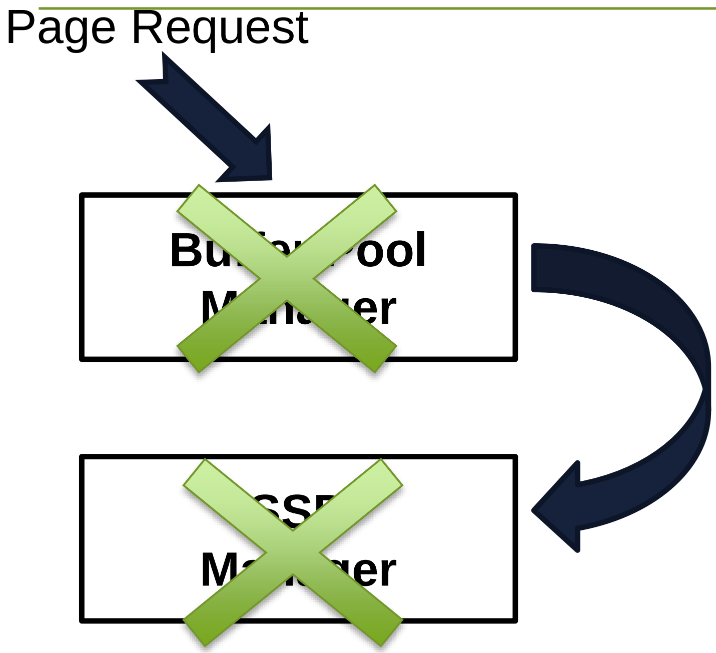
Buffer Pool Extension



If the page **<u>is NOT in the memory BP nor in the BPE</u>**, page is fetched from the disk.

☹

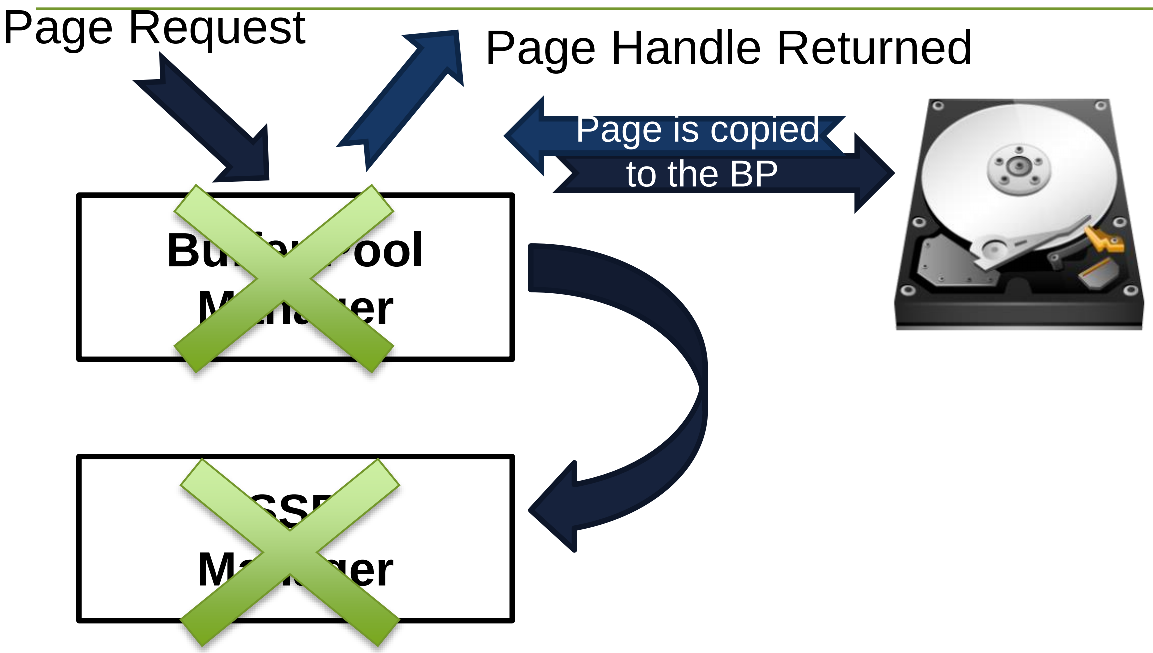
Buffer Pool Extension



If the page **<u>is NOT in the memory BP nor in the BPE</u>**, page is fetched from the disk.

☹

Buffer Pool Extension



Page Request

Page Handle Returned

Page is copied to the BP

Buffer Pool Manager

SSD Manager

If the page **<u>is NOT in the memory BP nor in the BPE</u>**, page is fetched from the disk.

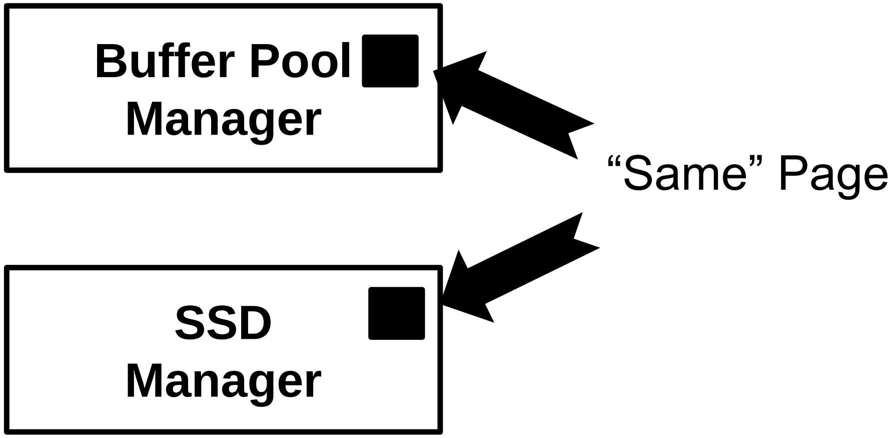
☹

Buffer Pool Extension

What if a **page**
is modified?

(and it is on both BP and BPE)

Buffer Pool Extension

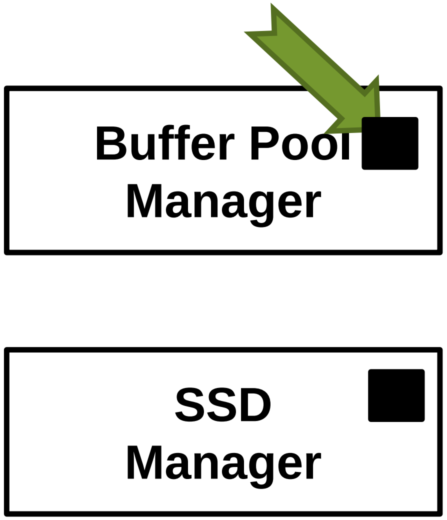


If the page **is NOT in the memory BP nor in the BPE**, page is fetched from the disk.

☹

Buffer Pool Extension

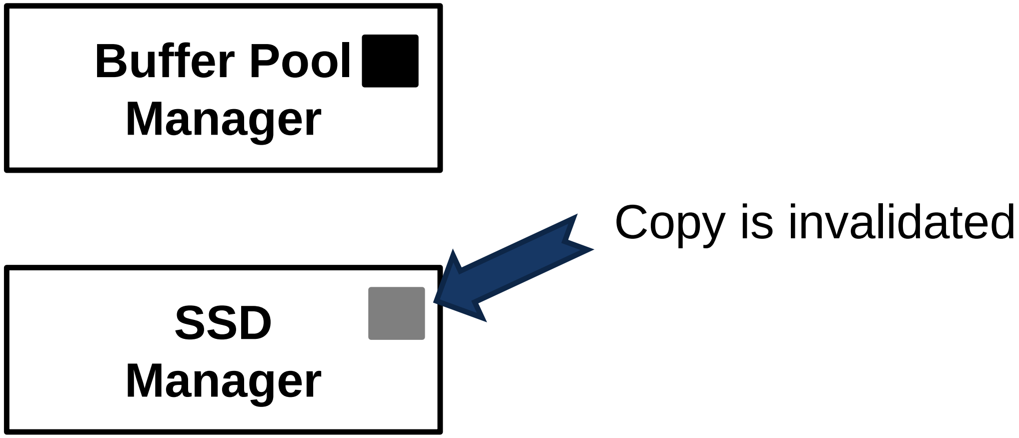
Page is modified (dirtied)



If the page **is NOT in the memory BP nor in the BPE**, page is fetched from the disk.

☹

Buffer Pool Extension

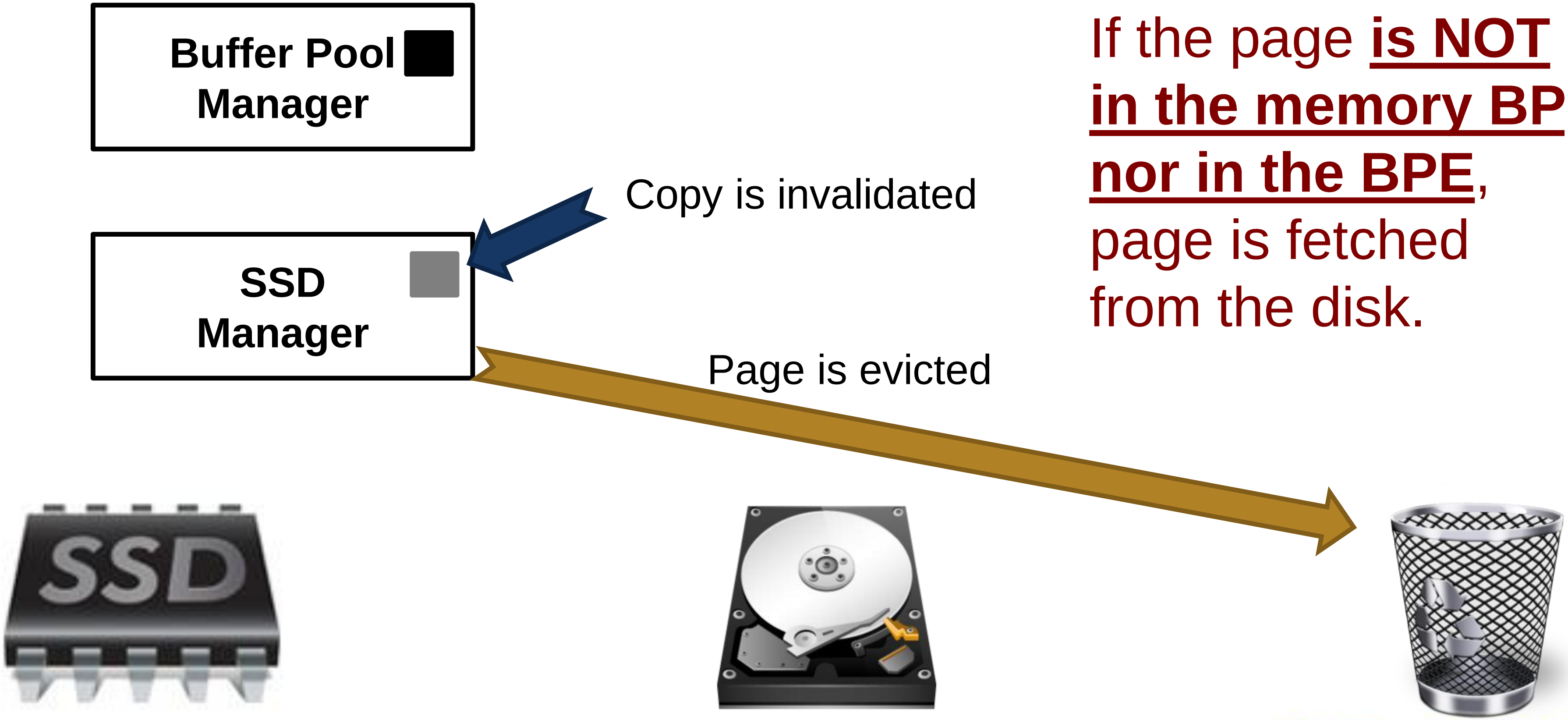


Copy is invalidated

If the page **<u>is NOT in the memory BP nor in the BPE</u>**, page is fetched from the disk.

☹

Buffer Pool Extension



If the page **is NOT in the memory BP nor in the BPE**, page is fetched from the disk.

☹

DEMO Time!

Buffer Pool Extension

MORE

About BPE

Buffer Pool Extension

- **Recommendations**
  - Use the fastest disk as possible.
  - Define the BPE file within 4 to 10 times the available memory size.

Buffer Pool Extension

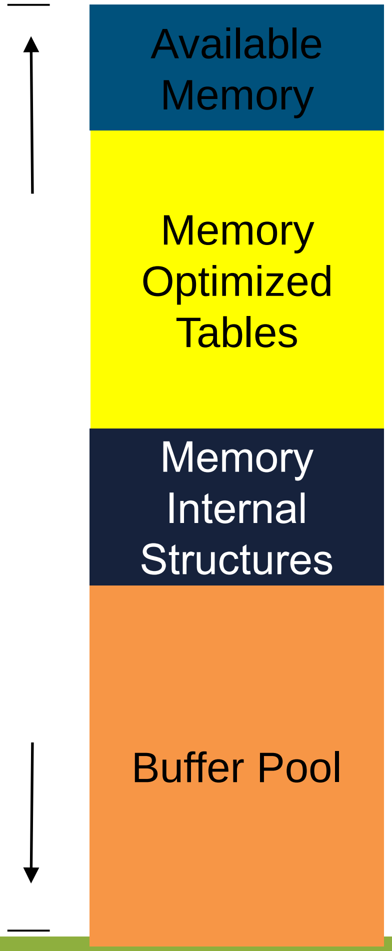
- **General Consideration**
  - The memory is always faster, use BPE only if there's no option to increase the RAM size.
  - Instances with a high amount of writes may not benefit from BPE.
  - BPE is not another point of failure, SQL Server behaves well if the BPE file have problems.
  - Servers with more than 64 GB of RAM may not take advantage.

DEMO Time!

Buffer Pool Extension

In-Memory OLTP

Memory Challenge



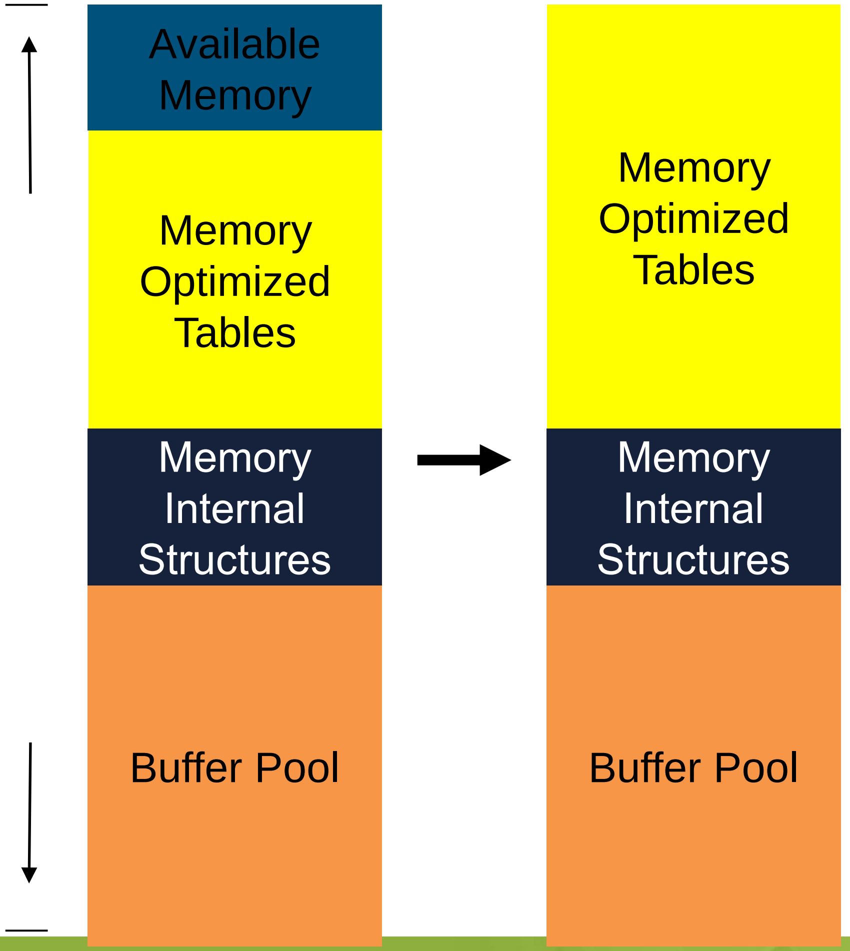
Available Memory

Memory Optimized Tables

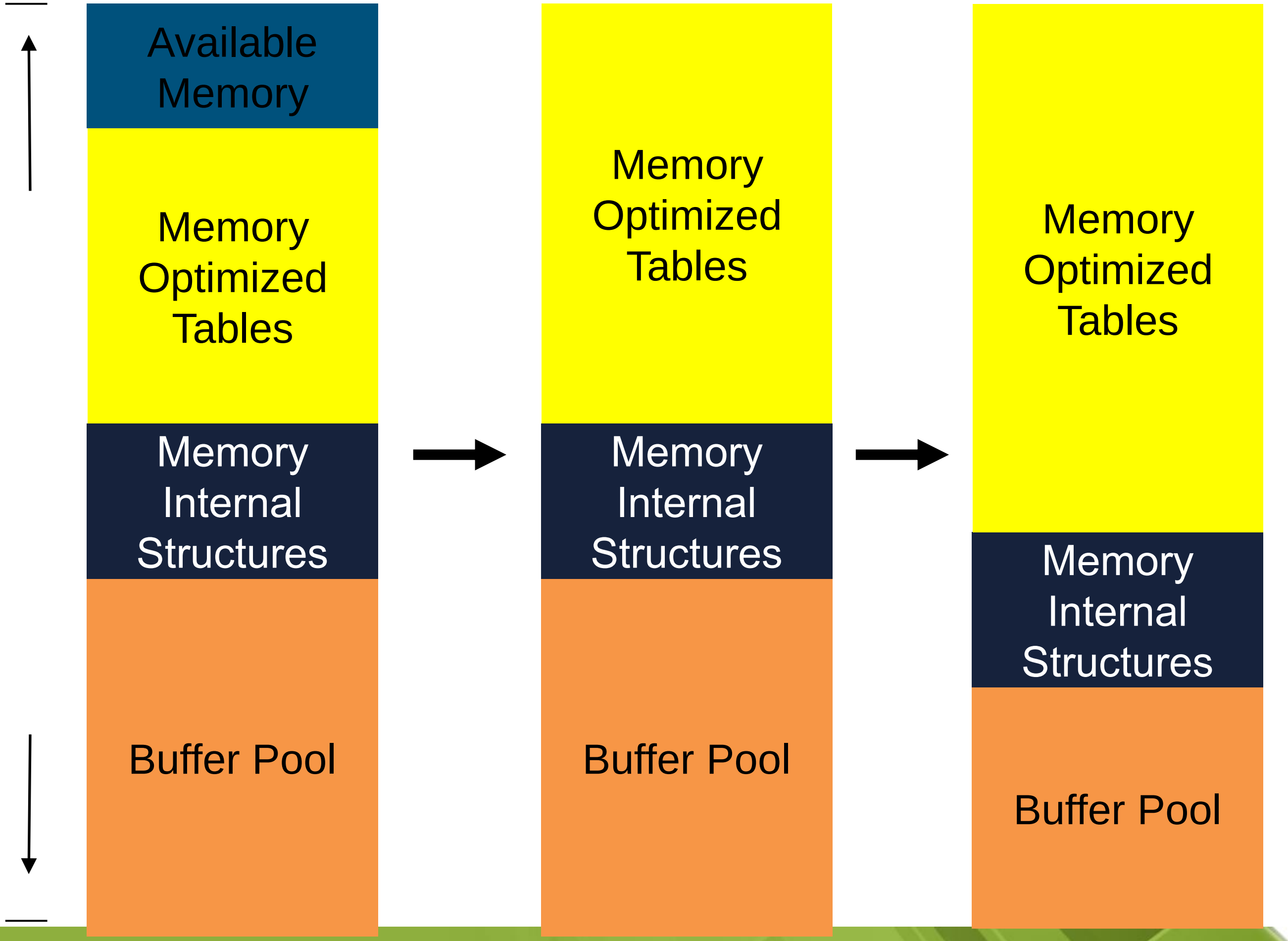
Memory Internal Structures

Buffer Pool

Memory Challenge



Memory Challenge



Available Memory

Memory Optimized Tables

Memory Internal Structures

Buffer Pool

Memory Optimized Tables

Memory Internal Structures

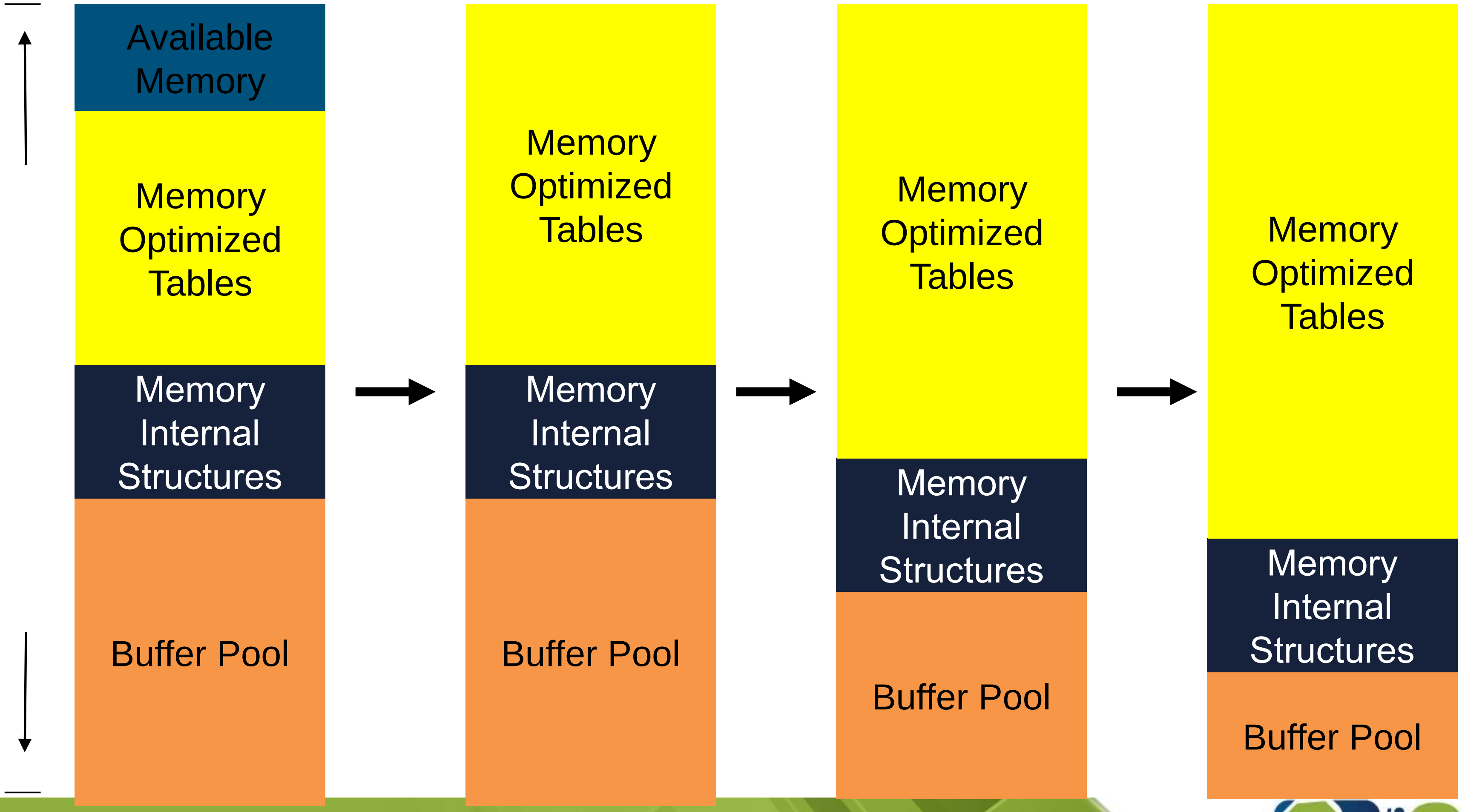
Buffer Pool

Memory Optimized Tables

Memory Internal Structures

Buffer Pool

Memory Challenge



Available Memory

Memory Optimized Tables

Memory Internal Structures

Buffer Pool

Memory Optimized Tables

Memory Internal Structures

Buffer Pool

Memory Optimized Tables

Memory Internal Structures

Buffer Pool

Memory Optimized Tables

Memory Internal Structures

Buffer Pool

Side Effects and Solutions

Side effects:

- Slow down of other workloads.
- Transactions on memory-optimized tables may fail due to out-of-memory.

Workaround:

- Adequate the memory size accordingly.
- Limit the memory Consumption using Resource Governor.
- Avoid the problem monitoring the system.
- **Enable the Buffer Pool Extension (BPE) feature.**

Ways to do…

TROUBLESHOOTING

Troubleshooting

- **Troubleshooting BPE**
  - DMVs
    - sys.dm_os_buffer_pool_extension_configuration
    - sys.dm_os_buffer_descriptors
  - XEvents
    - sqlserver.buffer_pool_extension_pages_written
    - sqlserver.buffer_pool_extension_pages_read
    - sqlserver.buffer_pool_extension_pages_evicted
    - sqlserver.buffer_pool_page_threshold_recalculated

Troubleshooting

- **Troubleshooting BPE**
  - Performance Counters
    - Extension page writes/sec
    - Extension page reads/sec
    - Extension outstanding IO counter
    - Extension page evictions/sec
    - Extension allocated pages
    - Extension free pages
    - Extension page unreferenced time
    - Extension in use as percentage on buffer pool level

DEMO Time!

QUESTIONS?

THANK YOU!