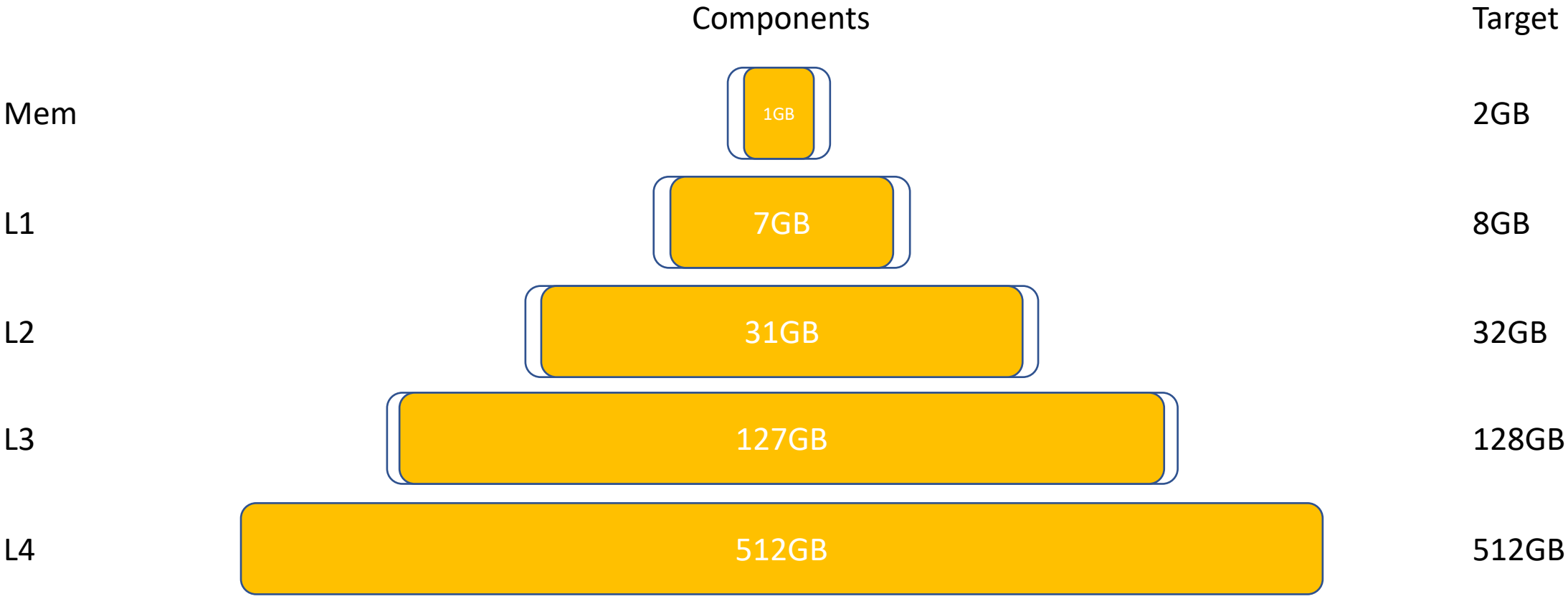
Compaction styles

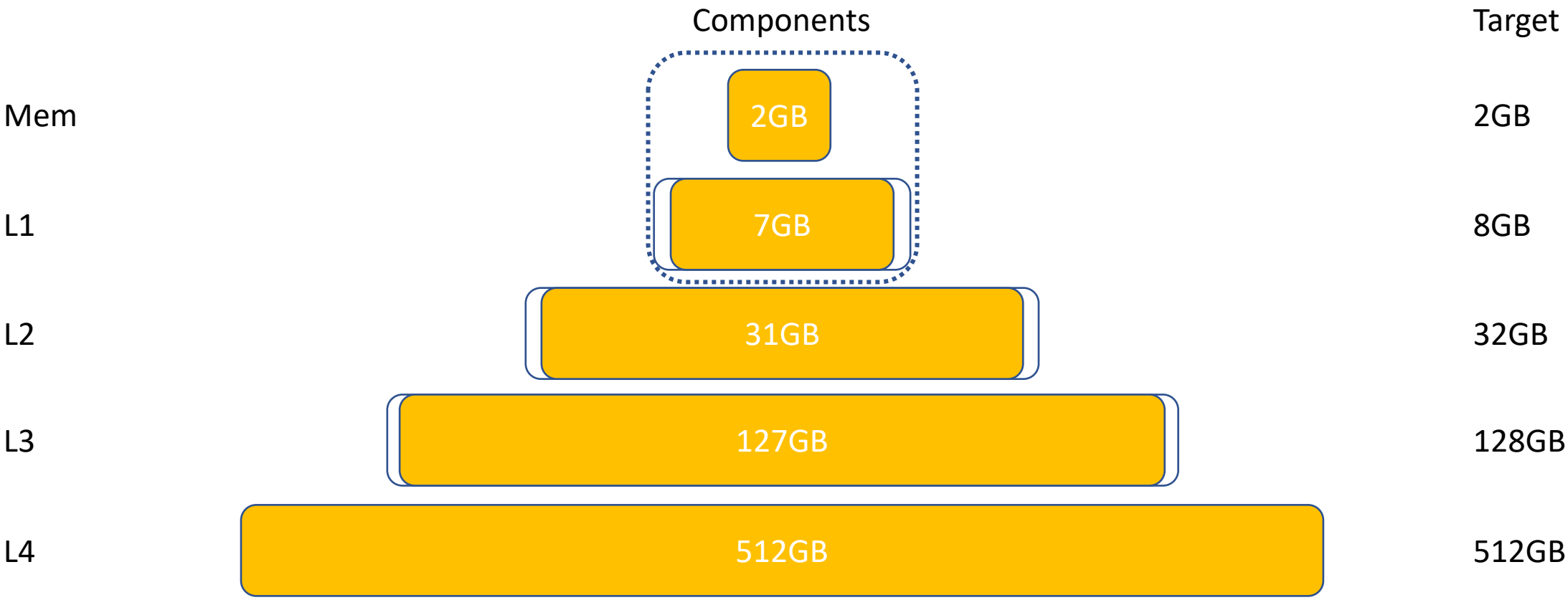
Andrew Kryczka

July 17, 2020

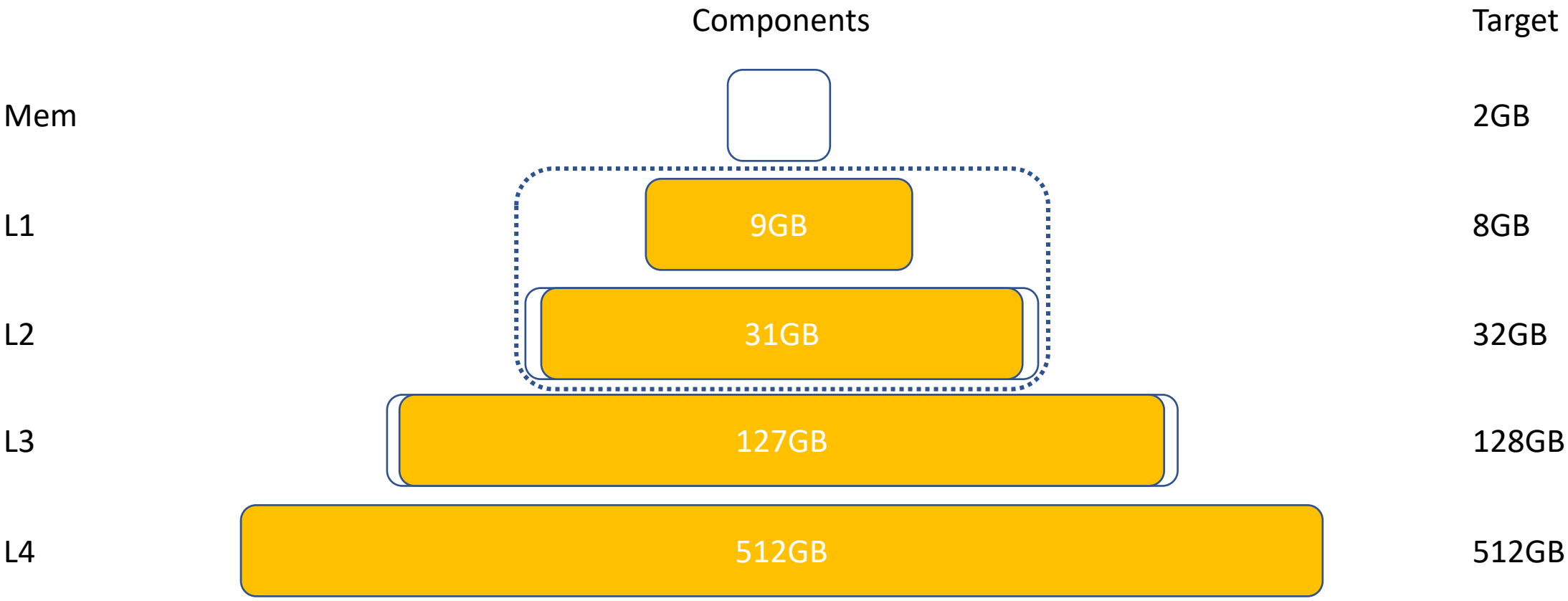
Pure leveled (T=4)



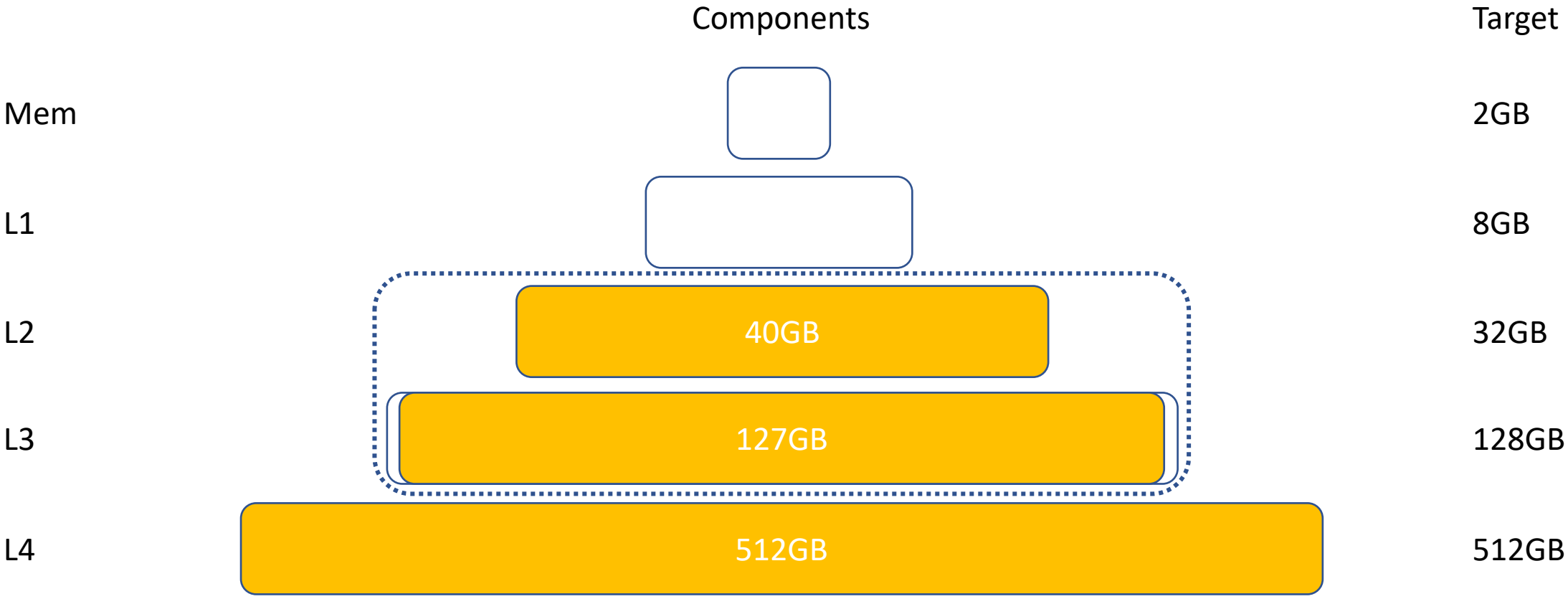
Pure leveled flush



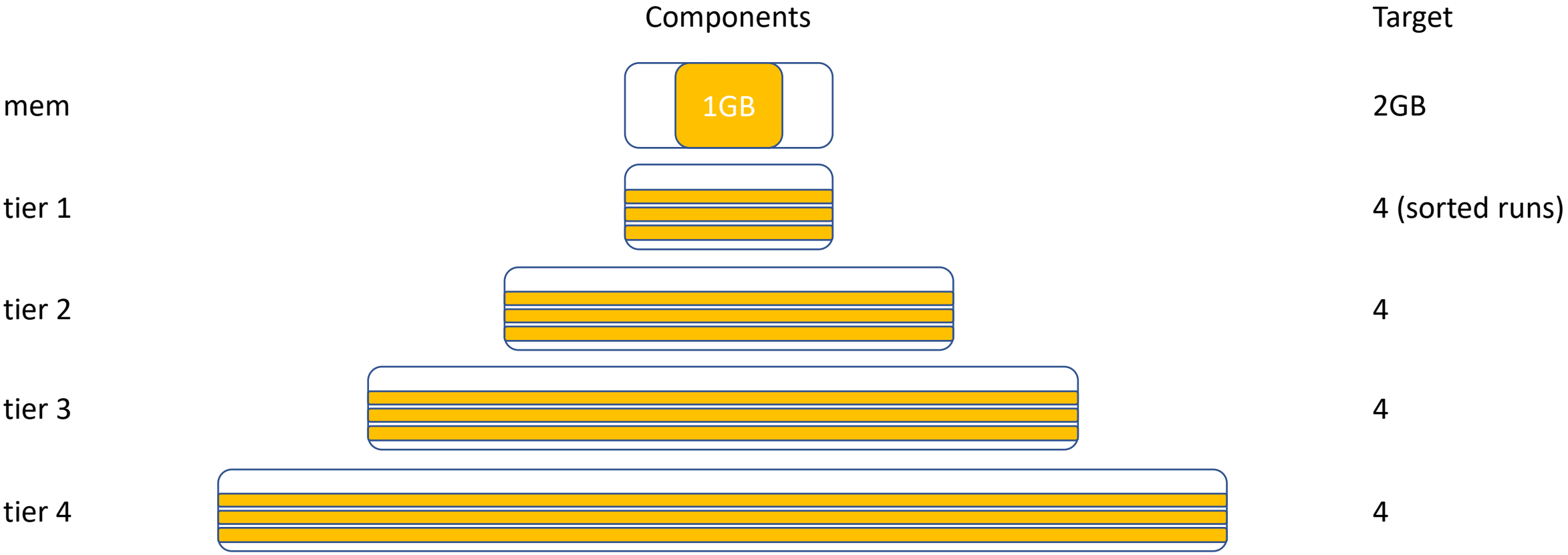
Pure leveled compaction



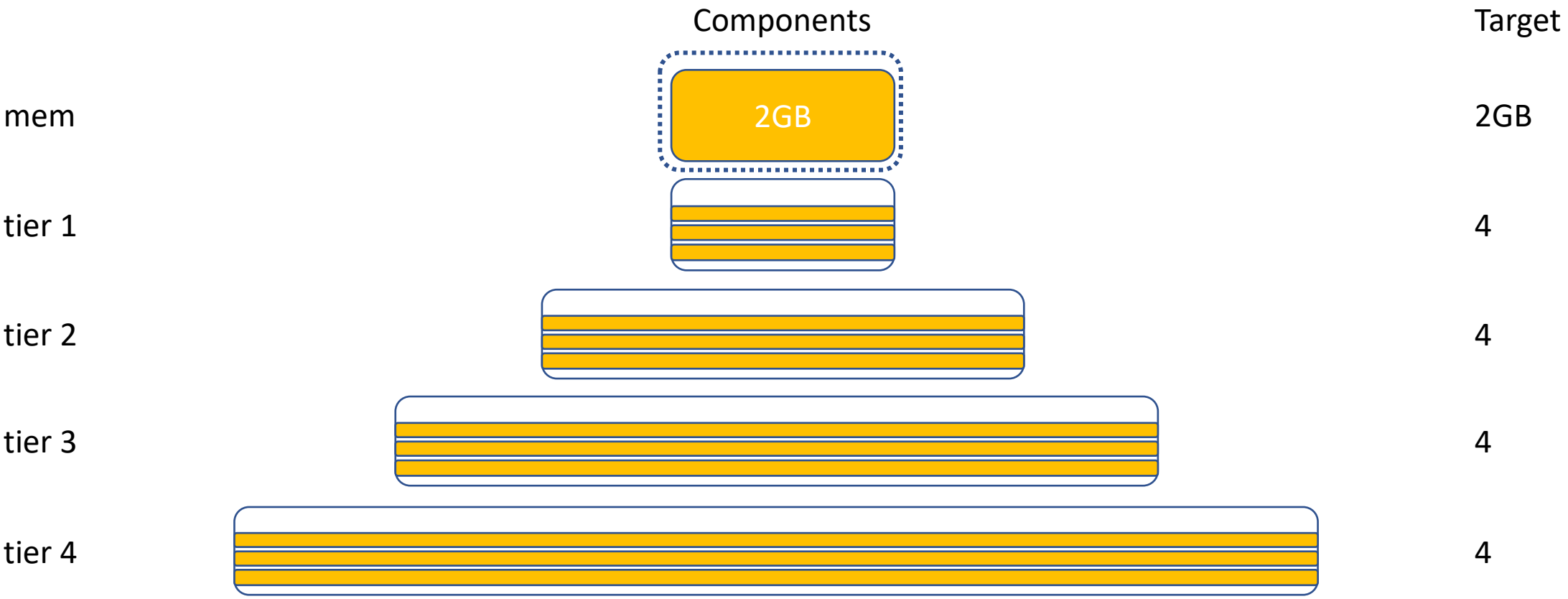
Pure leveled compaction



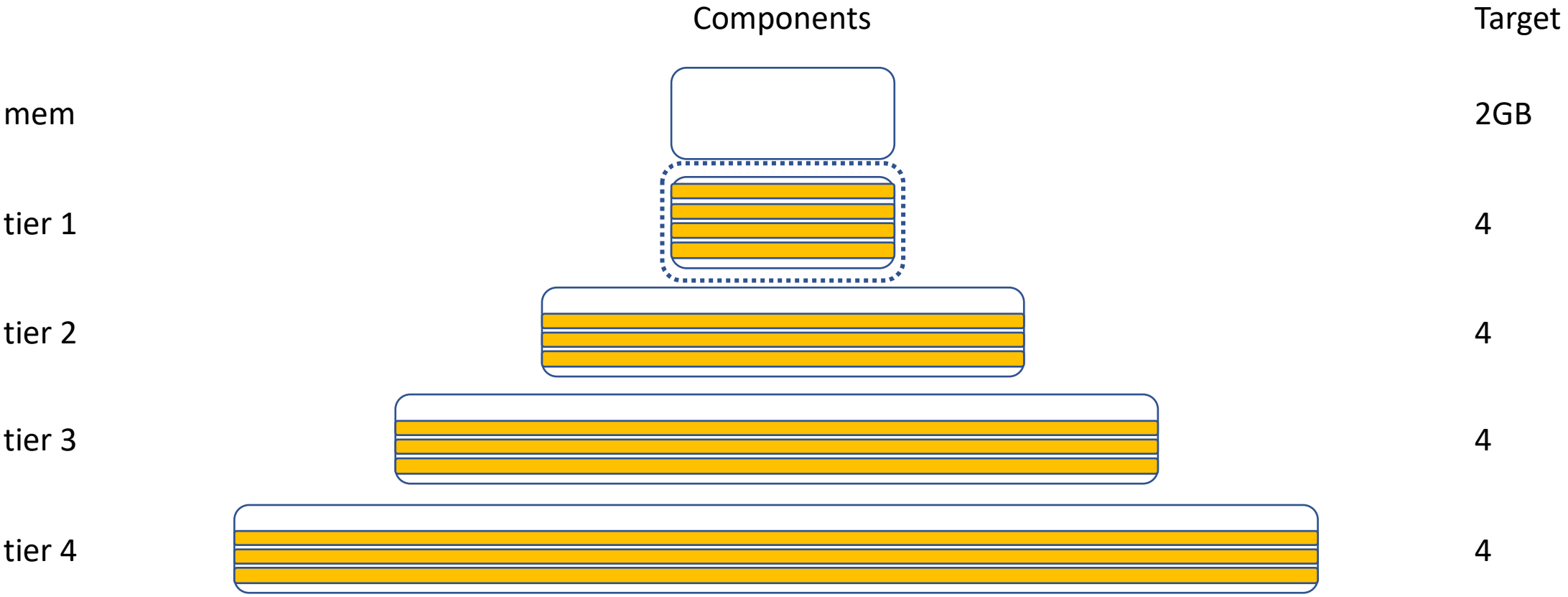
Pure tiered (T=4)



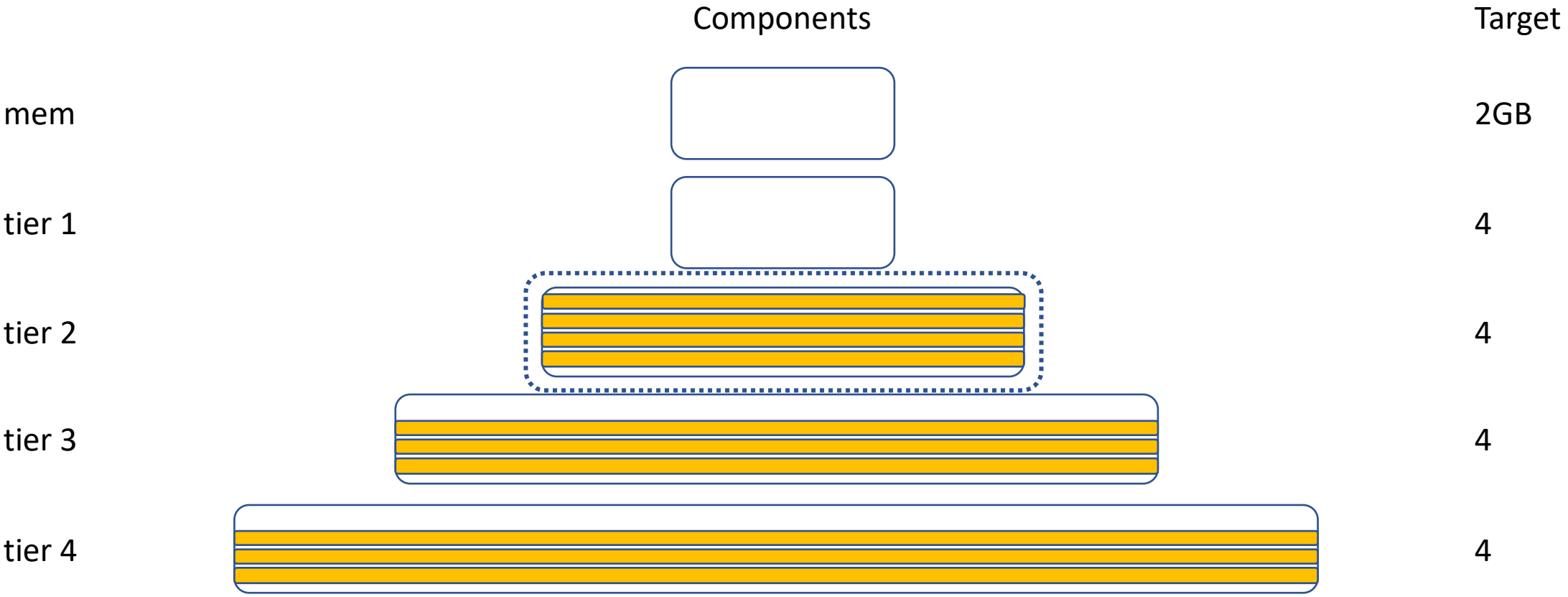
Pure tiered flush



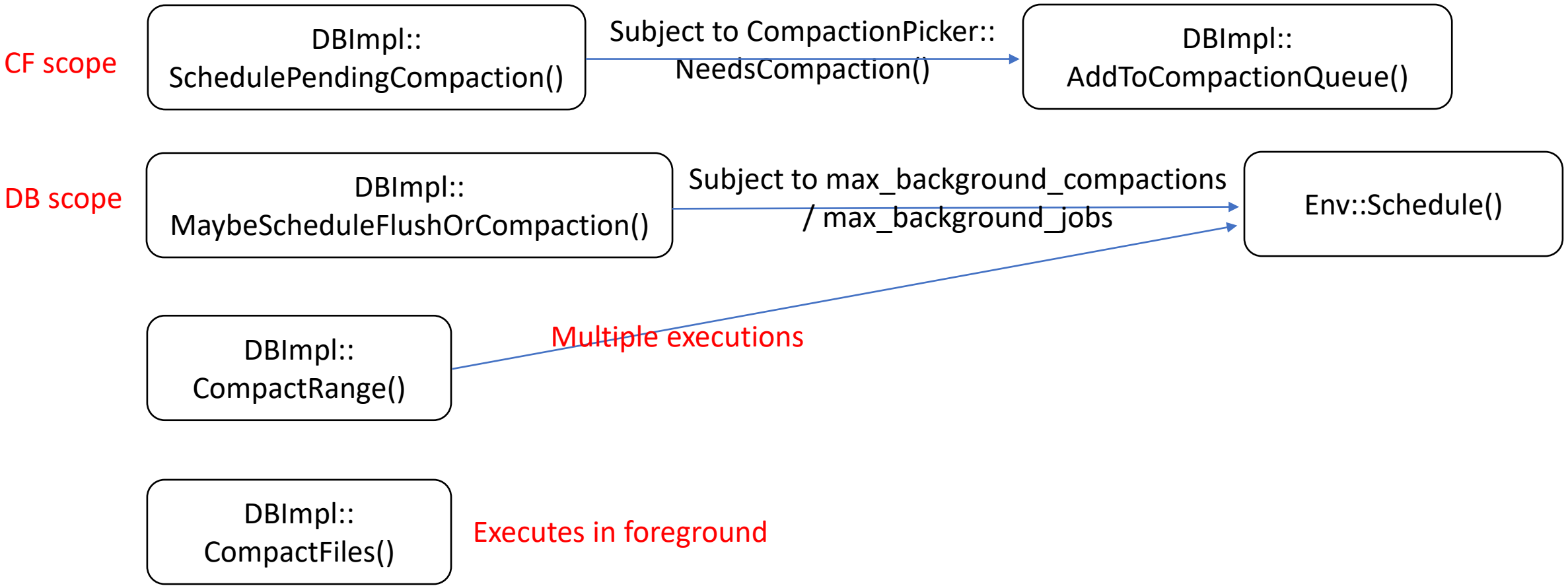
Pure tiered compaction



Pure tiered compaction

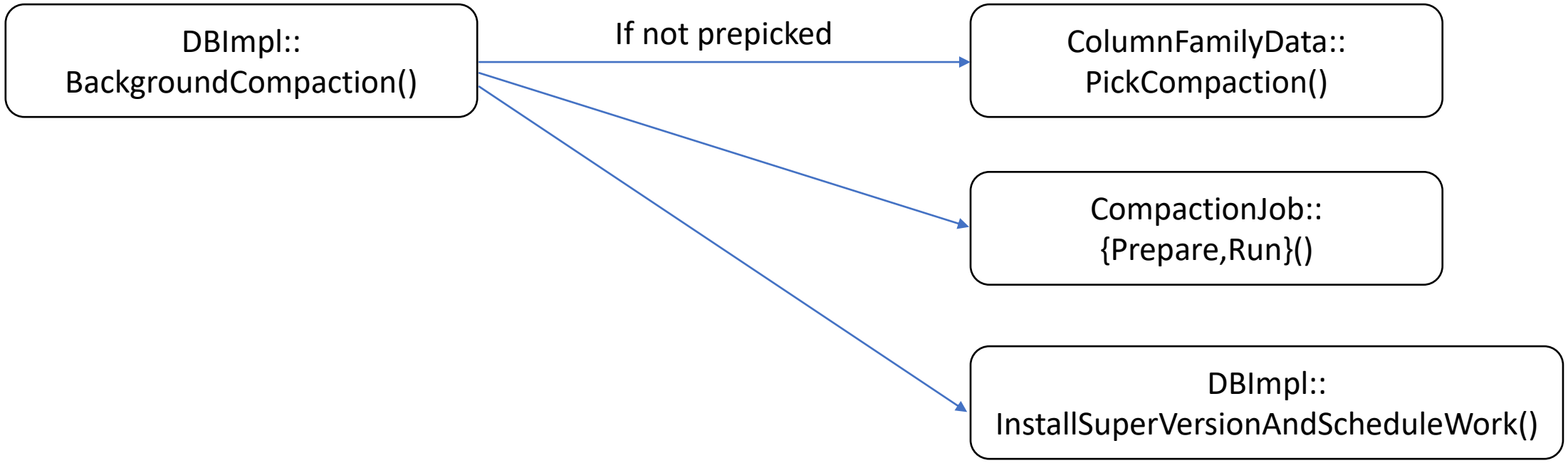


Code: compaction scheduling



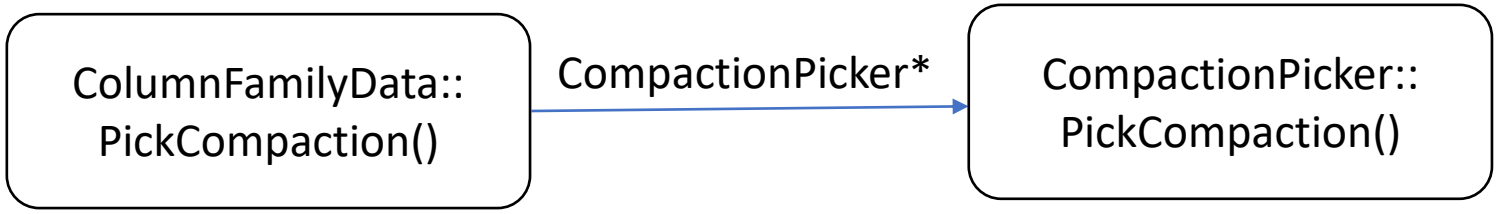
Code: compaction execution

In either Env::Priority::LOW or Env::Priority::BOTTOM ThreadPool

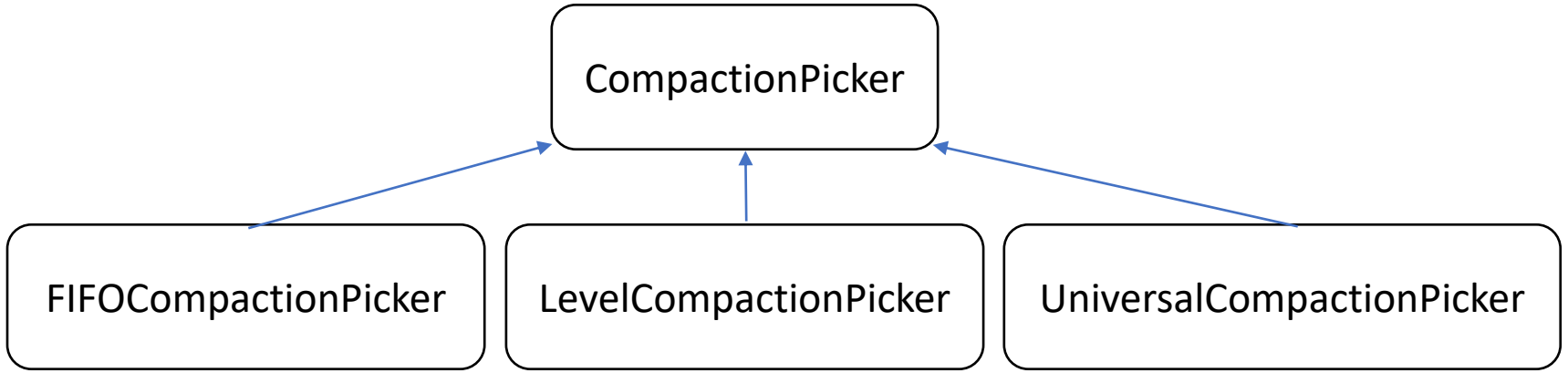


Schedule based on new state – chaining!

Code: compaction picking

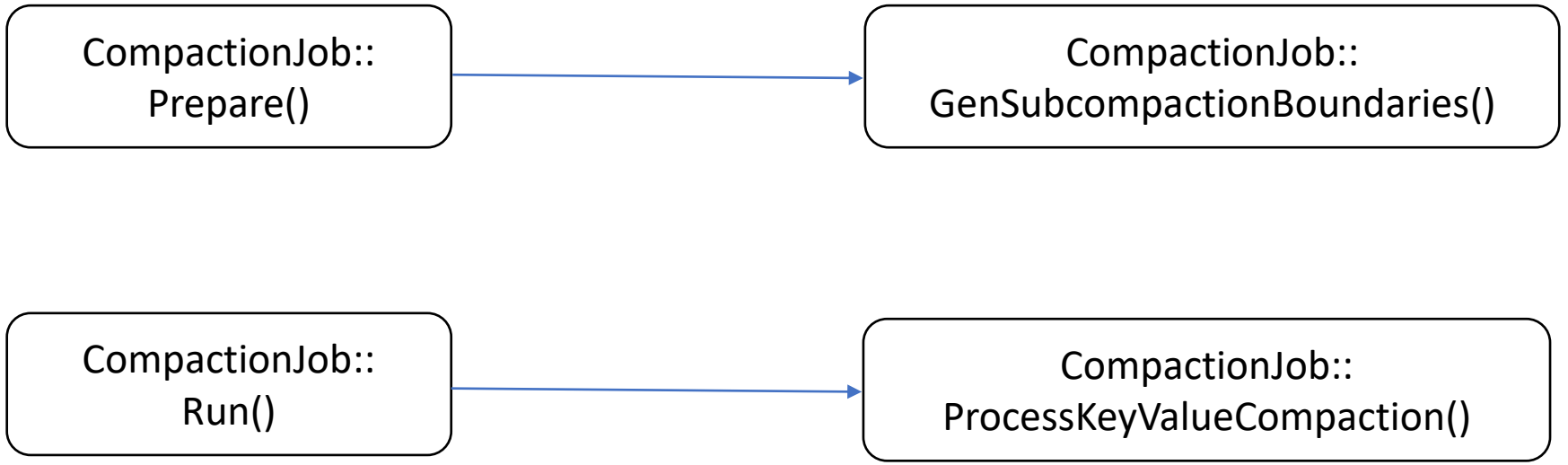


ColumnFamilyOptions::compaction_style determines CompactionPicker subclass



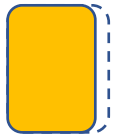
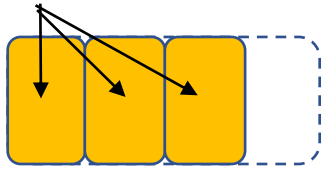
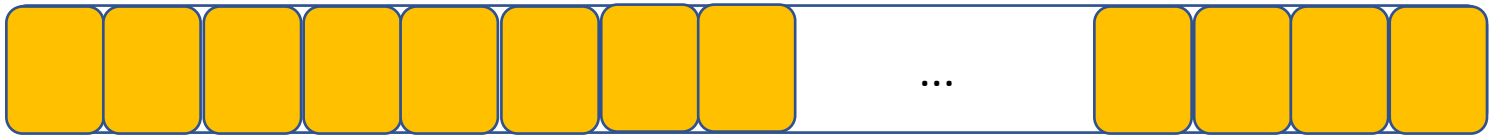
ColumnFamilyOptions::compaction_pri
determines order of
VersionStorageInfo::FilesByCompactionPri()
(used by LevelCompactionPicker only)

Code: CompactionJob

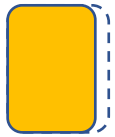
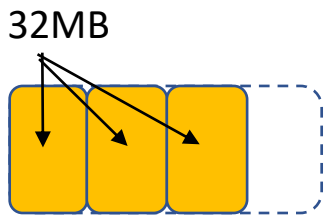
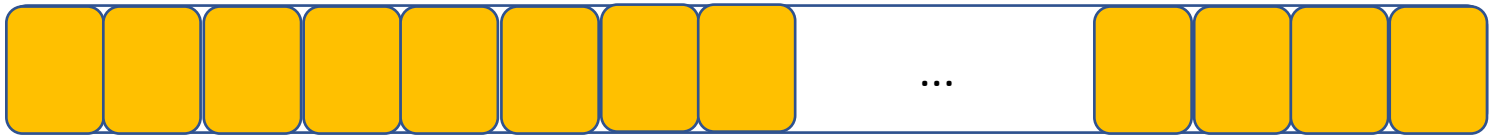


- Separate thread per subcompaction
- Uses CompactionIterator

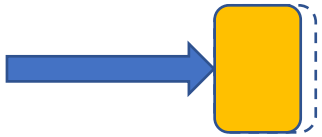
RocksDB leveled

	Components	Target
Mutable memtable		write_buffer_size
Immutable memtables		max_write_buffer_number
L0		level0_file_num_compaction_trigger
...		
Base level	target_file_size_base 	max_bytes_for_level_base
...		
num_levels-1		DB size * (1 – 1/max_bytes_for_level_multiplier)

RocksDB leveled, MyRocks options

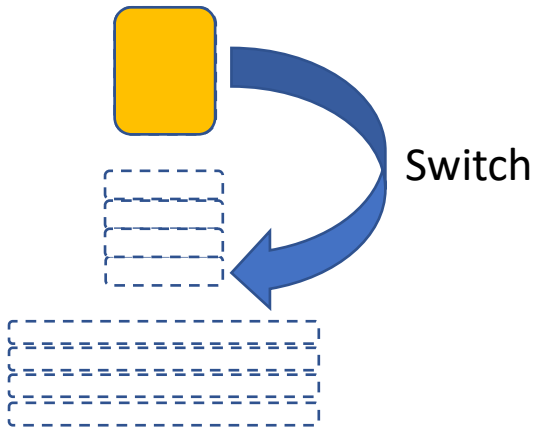
	Components	Target
Mutable memtable		128MB
Immutable memtables		4
L0		4
...		
Base level		512MB
...		
L6		90% DB size

RocksDB leveled flush

	Components	Target
Mutable memtable	Writes 	128MB
Immutable memtables		4
L0		4

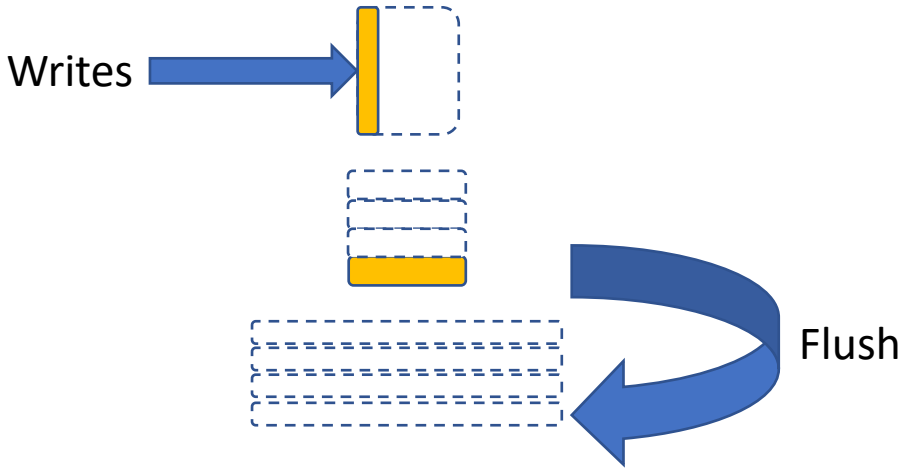
RocksDB leveled flush

Components	Target
Mutable memtable	128MB
Immutable memtables	4
L0	4

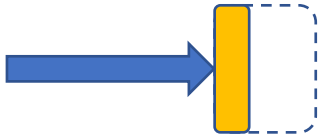


RocksDB leveled flush

Components	Target
Mutable memtable	128MB
Immutable memtables	4
L0	4

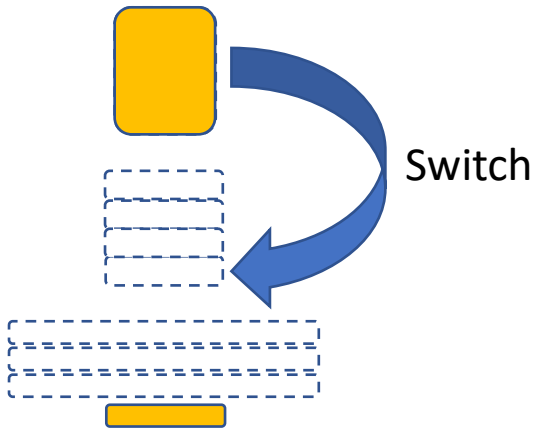


RocksDB leveled flush

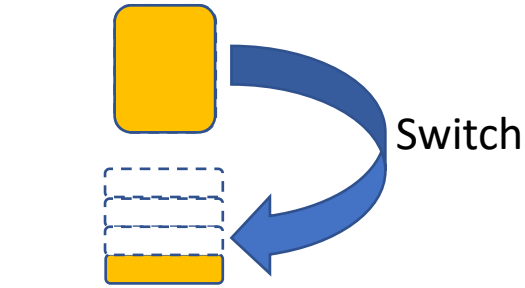
	Components	Target
Mutable memtable	Writes 	128MB
Immutable memtables		4
L0		4

RocksDB leveled flush

Components	Target
Mutable memtable	128MB
Immutable memtables	4
L0	4

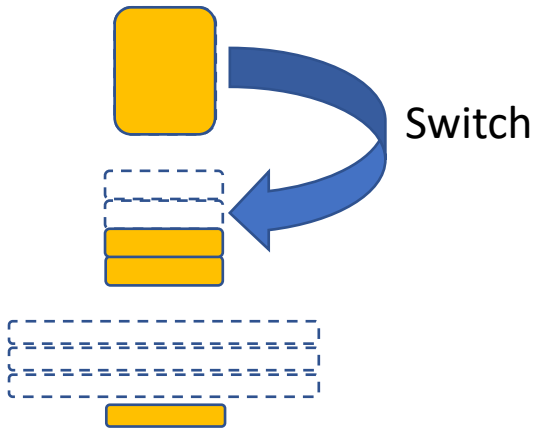


RocksDB leveled flush

	Components	Target
Mutable memtable		128MB
Immutable memtables		4
L0		4

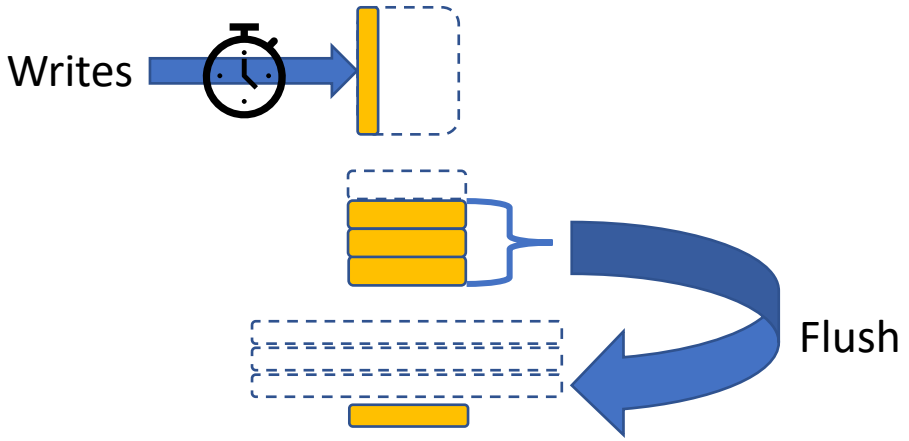
RocksDB leveled flush

Components	Target
Mutable memtable	128MB
Immutable memtables	4
L0	4



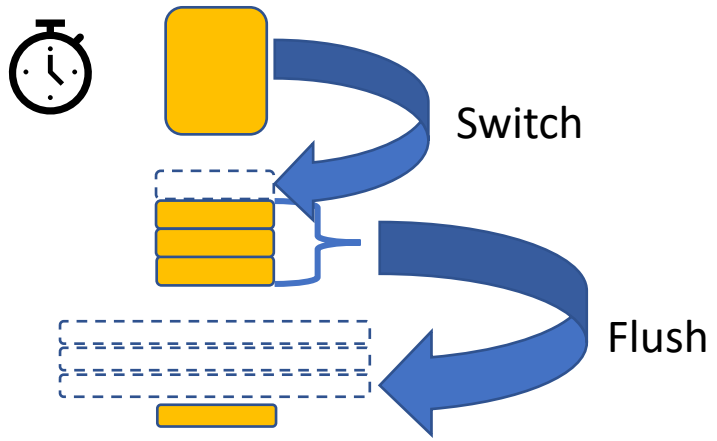
RocksDB leveled flush

Components	Target
Mutable memtable	128MB
Immutable memtables	4
L0	4



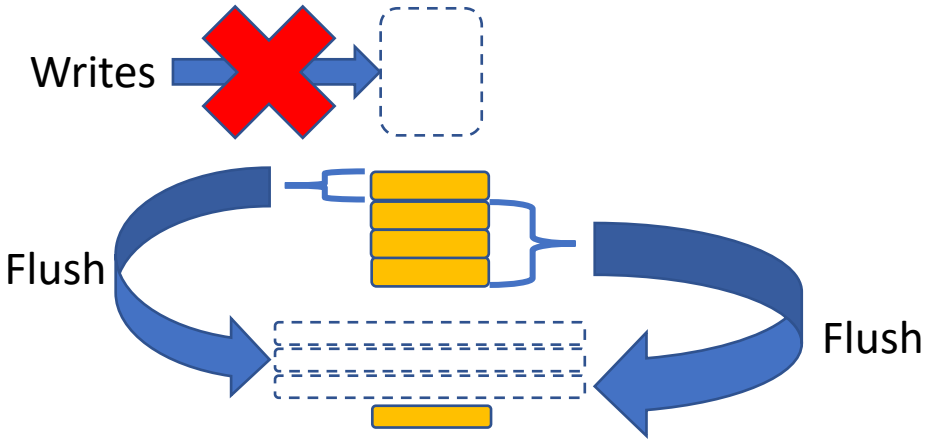
RocksDB leveled flush

Components	Target
Mutable memtable	128MB
Immutable memtables	4
L0	4

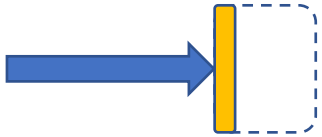


RocksDB leveled flush

Components	Target
Mutable memtable	128MB
Immutable memtables	4
L0	4



RocksDB leveled flush

	Components	Target
Mutable memtable	Writes 	128MB
Immutable memtables		4
L0		4

RocksDB leveled compaction: L0, base level

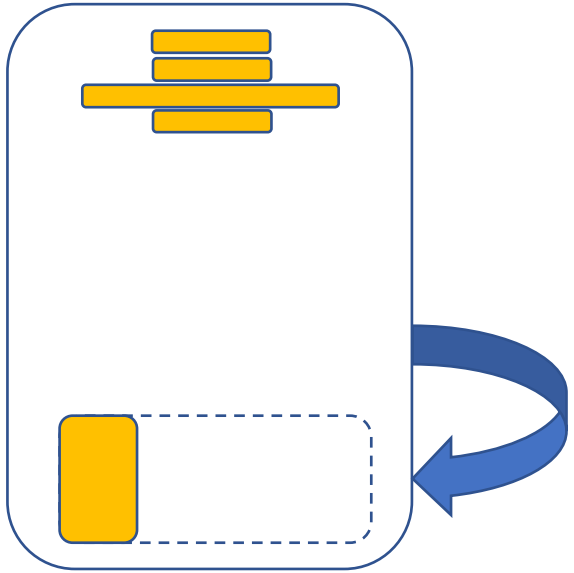
Components

Target

L0

4

...



Compaction

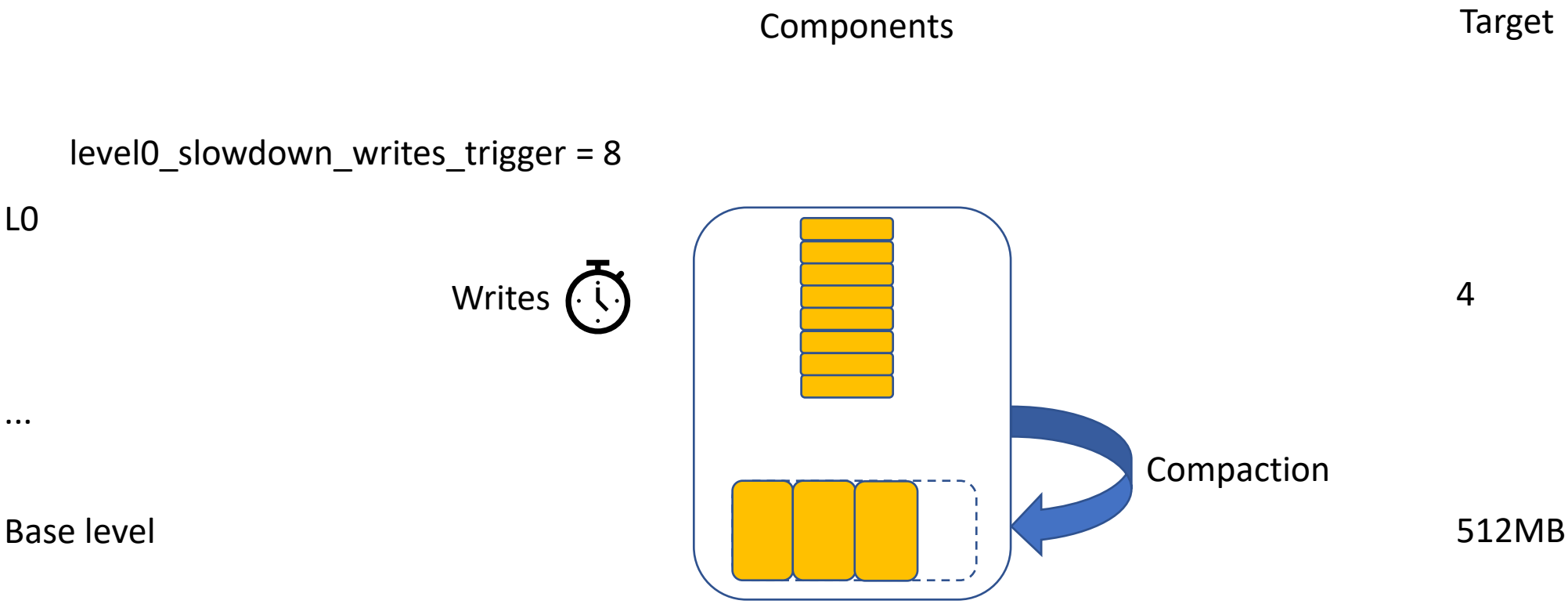
Base level

512MB

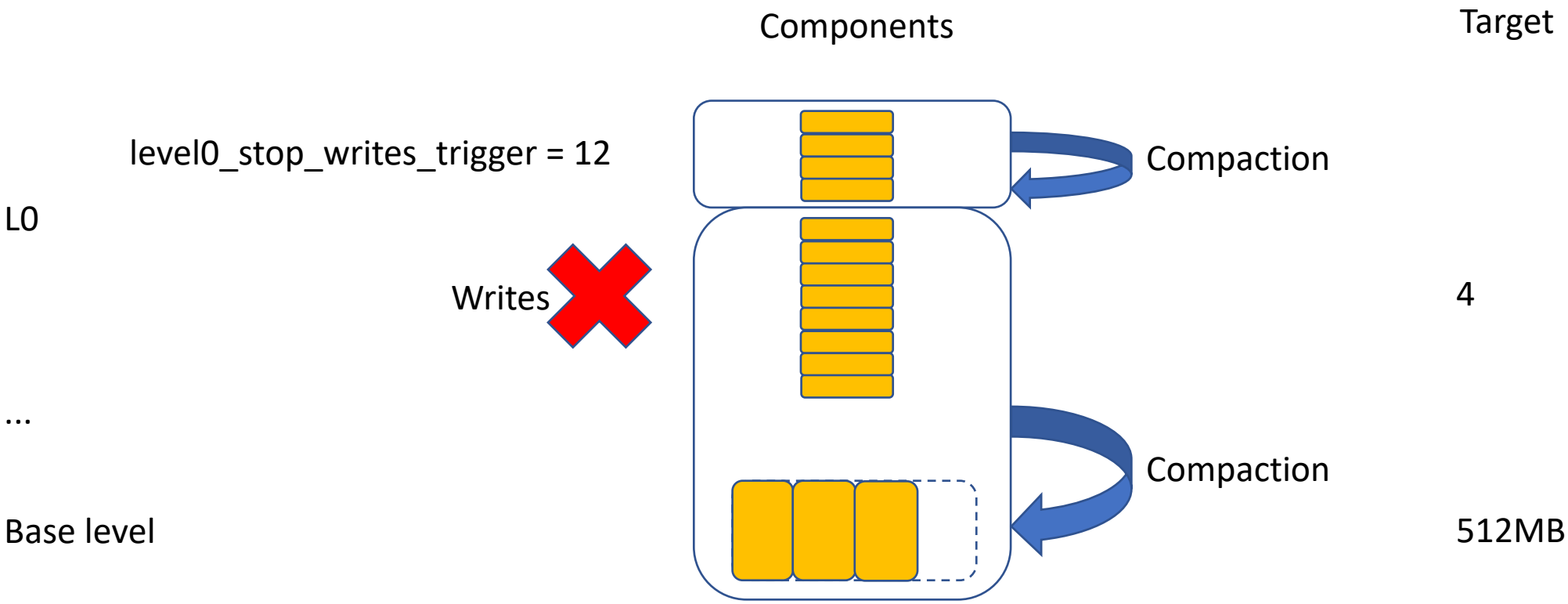
RocksDB leveled compaction: L0, base level

	Components	Target
L0		4
...		
Base level		512MB

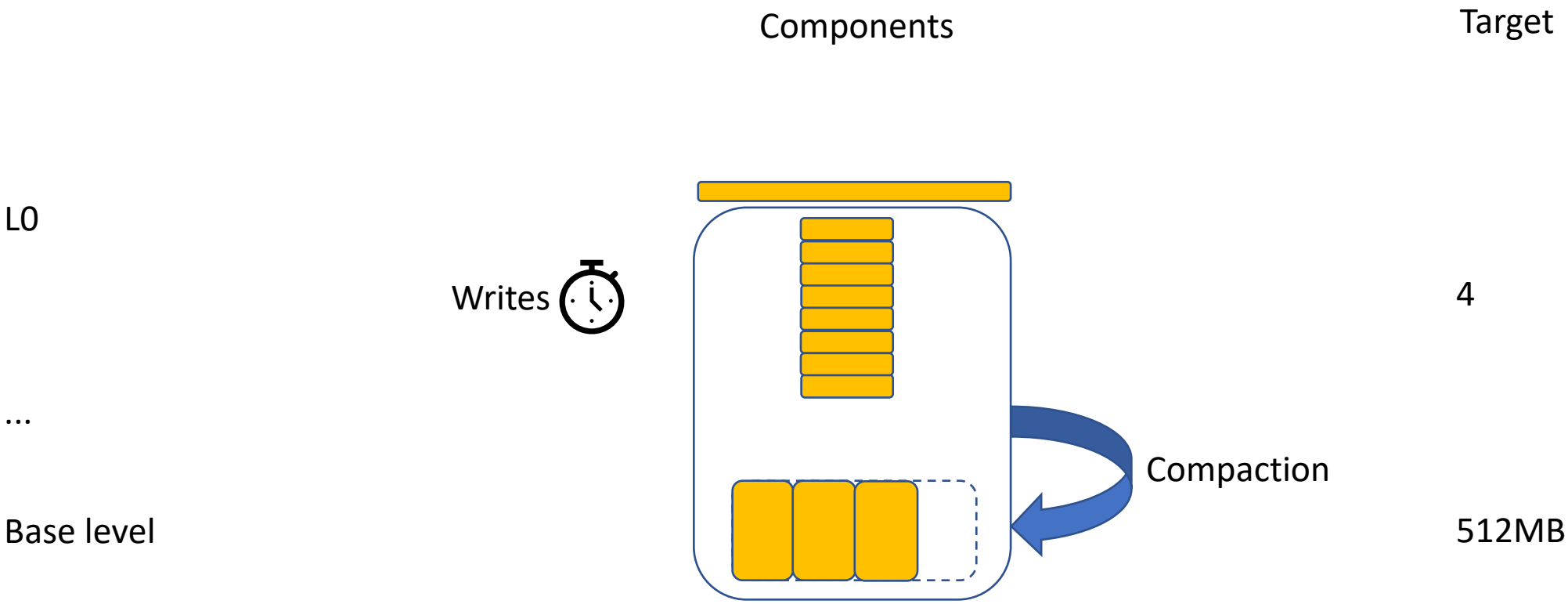
RocksDB leveled compaction: L0, base level



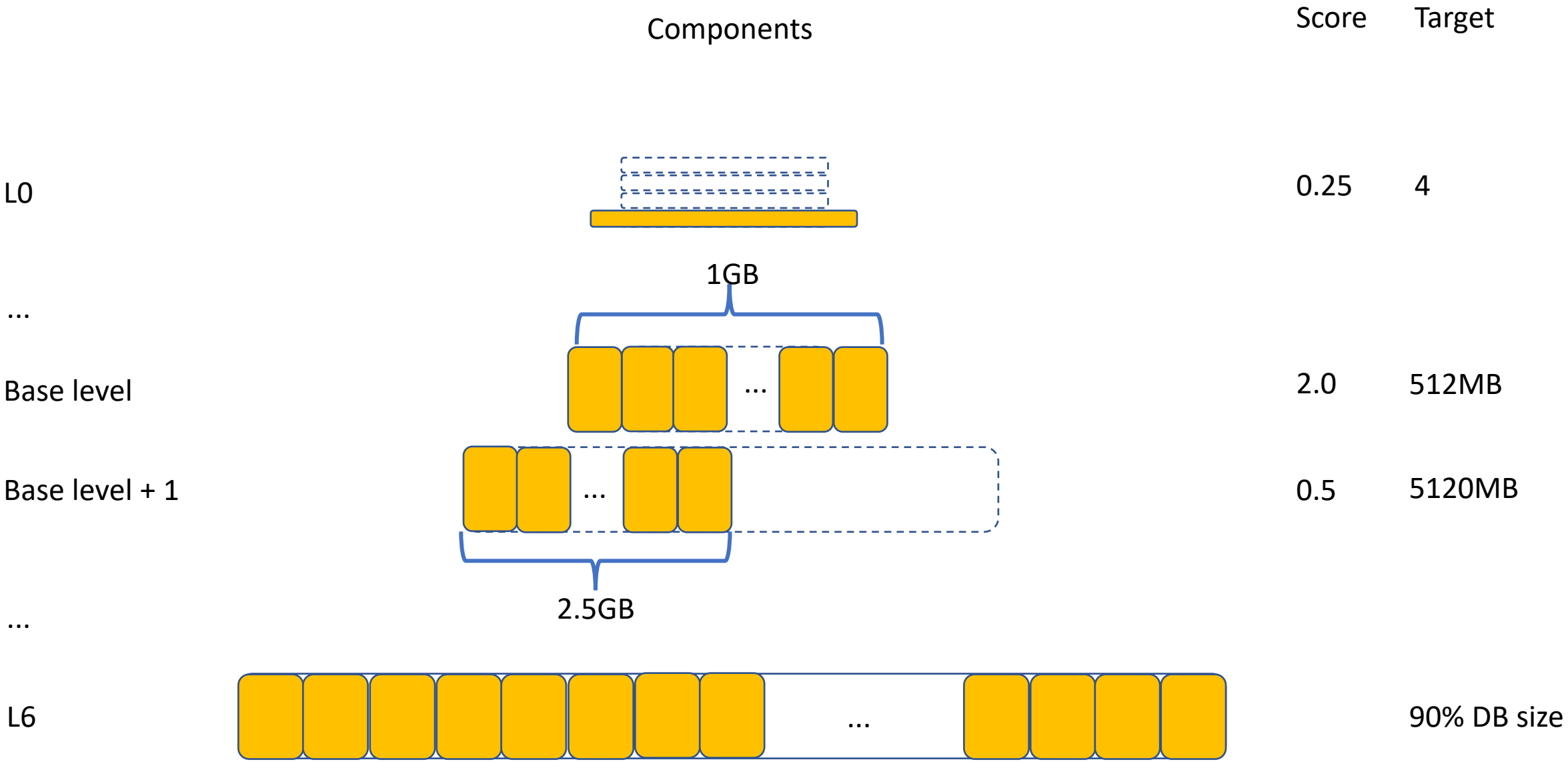
RocksDB leveled compaction: L0, base level



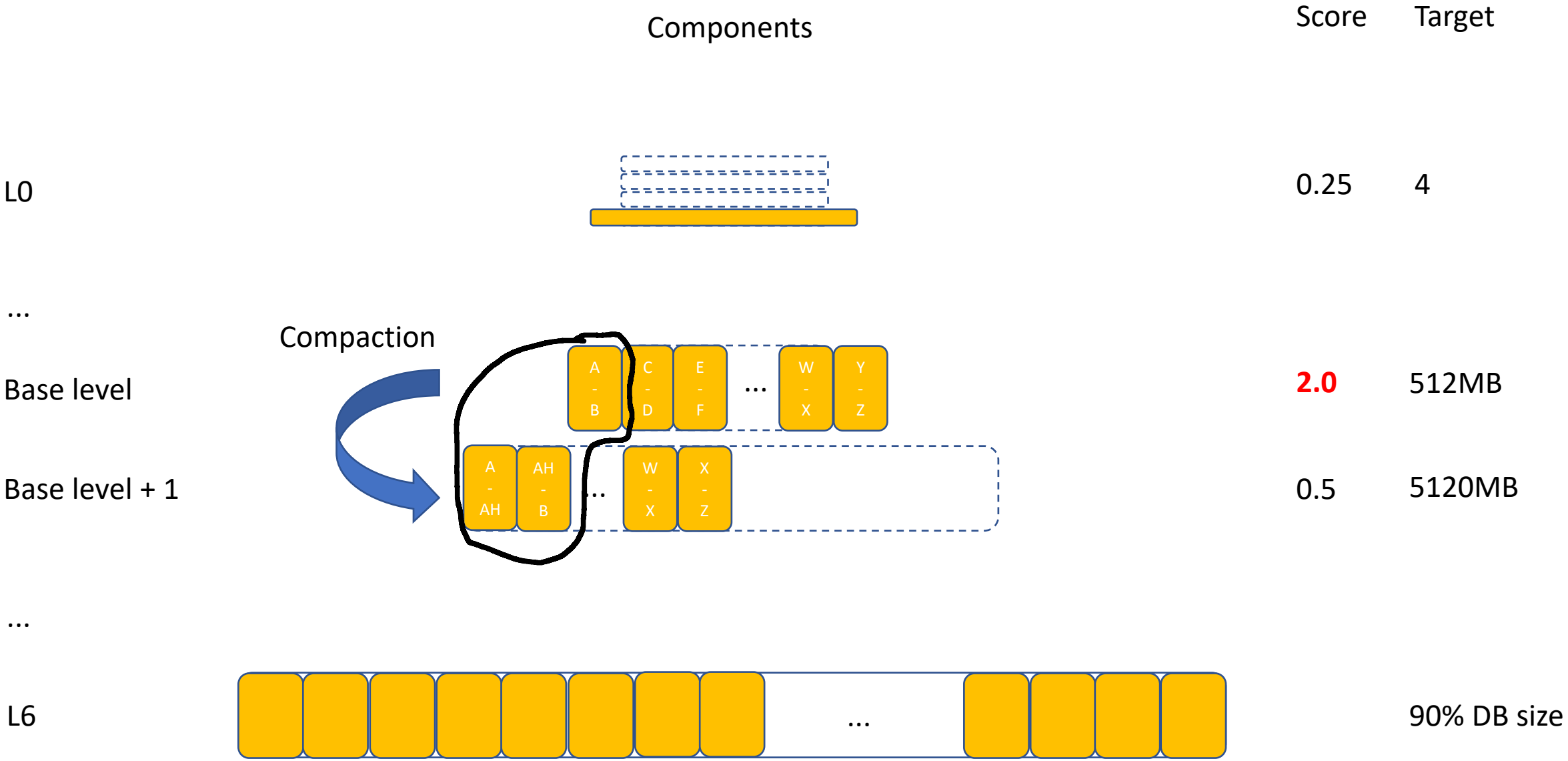
RocksDB leveled compaction: L0, base level



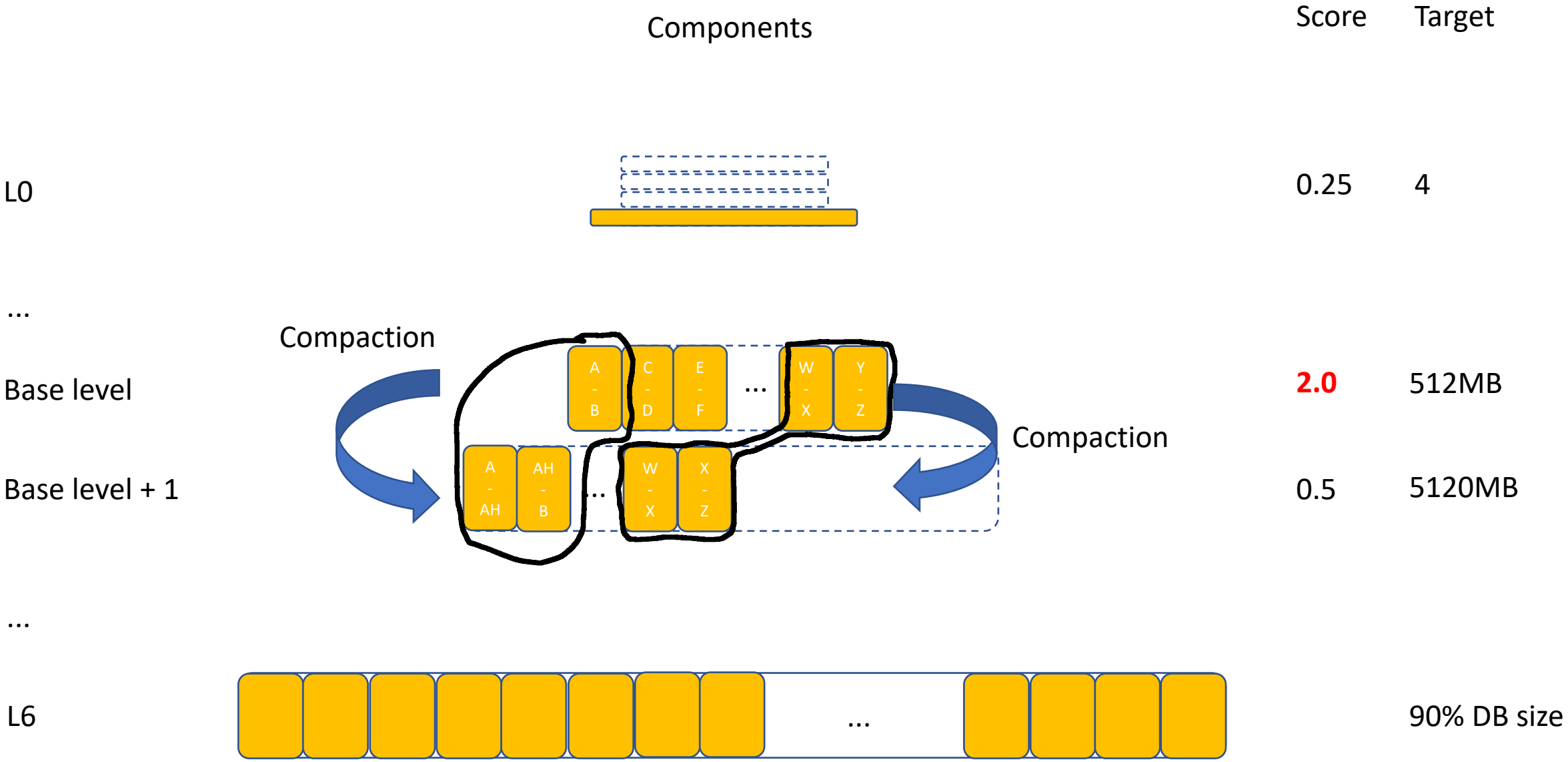
RocksDB leveled compaction



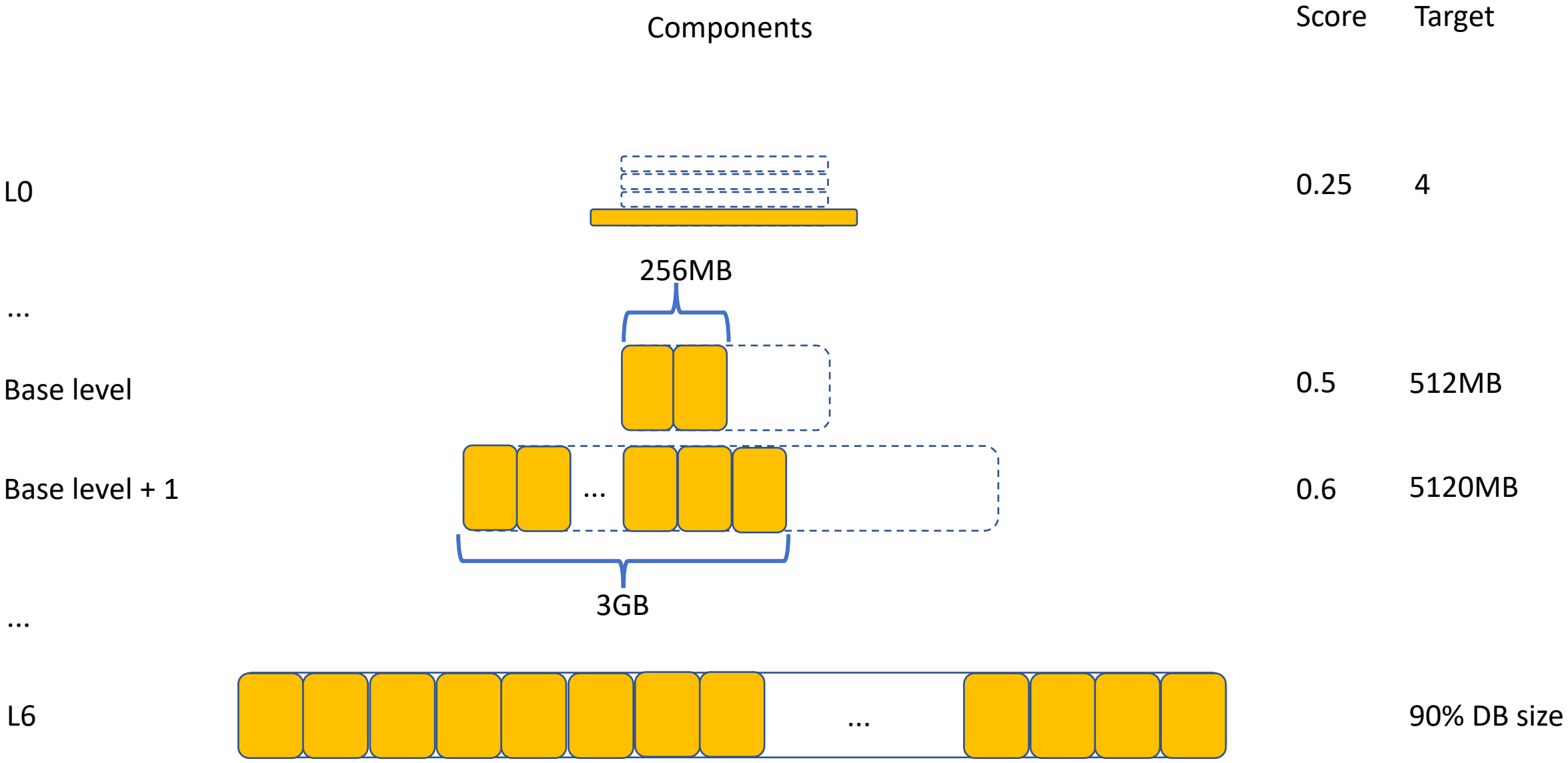
RocksDB leveled compaction



RocksDB leveled compaction



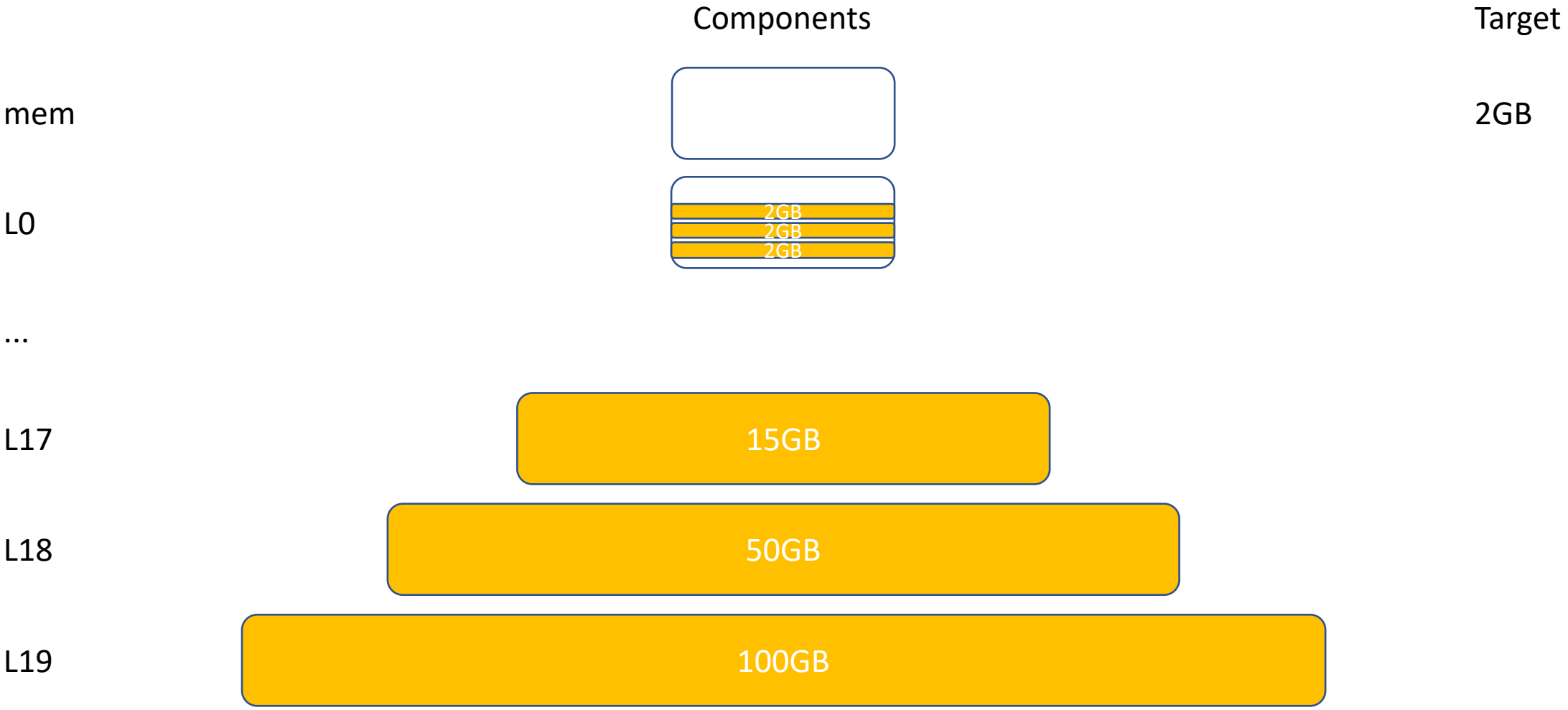
RocksDB leveled compaction



RocksDB universal heuristics

- Space amp
- Size ratio
- Read amp

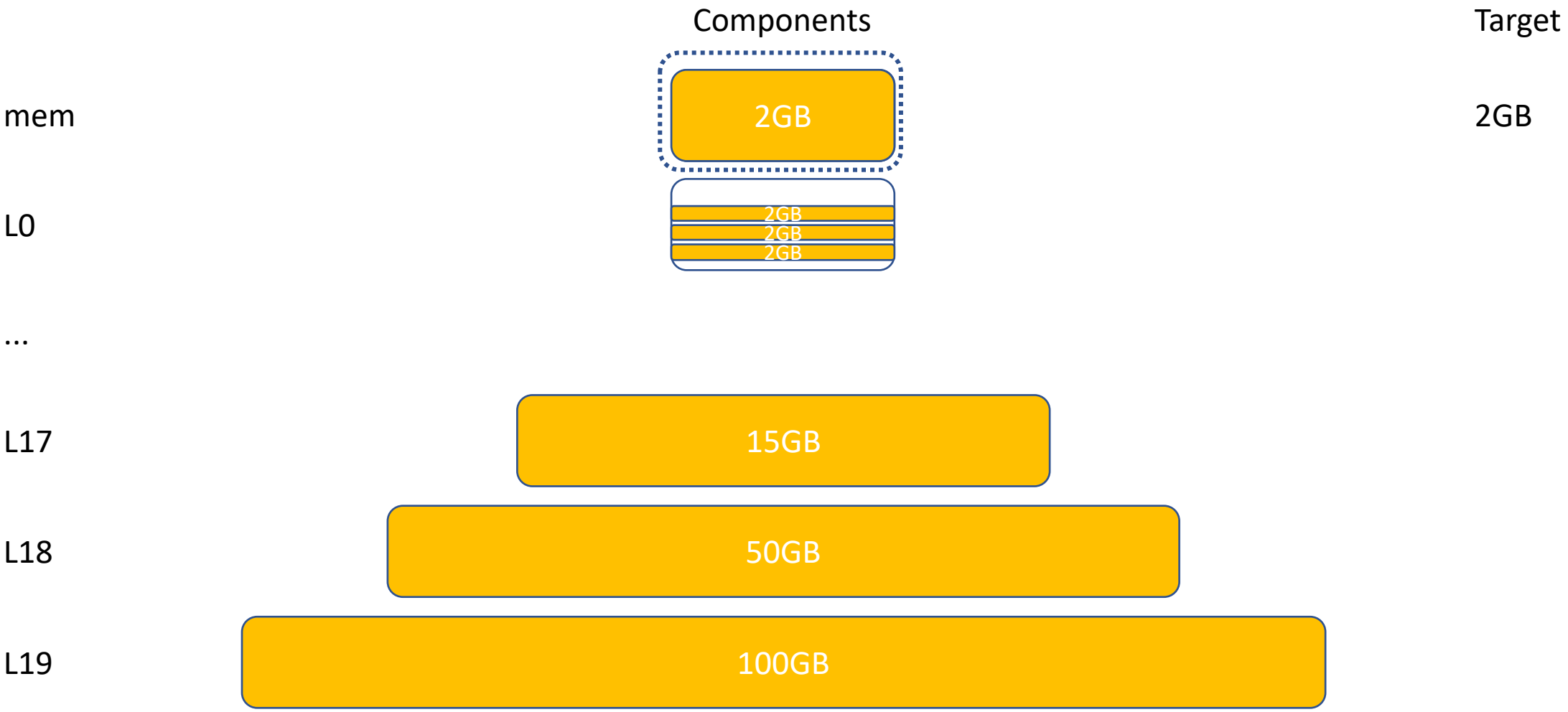
RocksDB universal



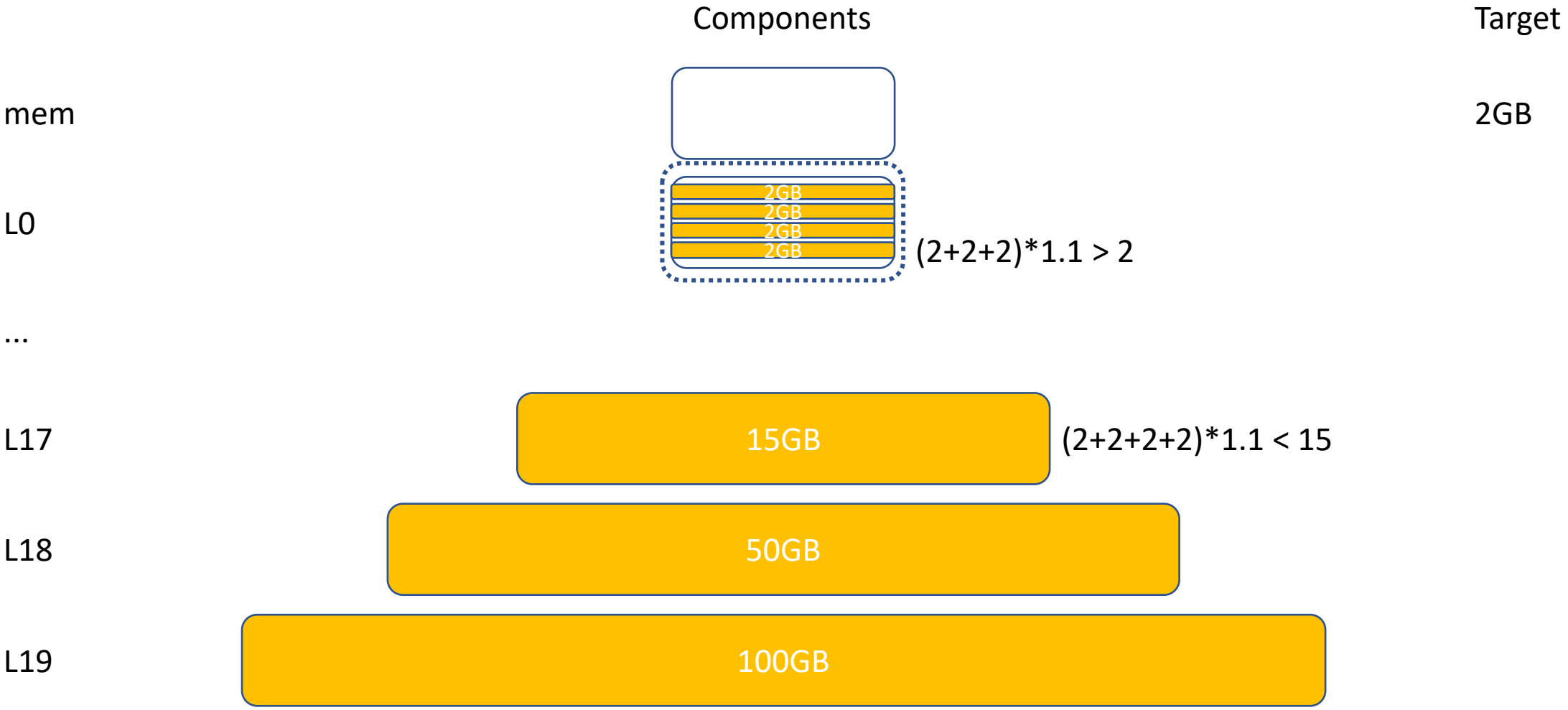
RocksDB universal options

- num_levels = 20
- level0_file_num_compaction_trigger = 6
- size_ratio = 10
- min_merge_width = 4
- max_merge_width = 8
- max_size_amplification_percent = 100

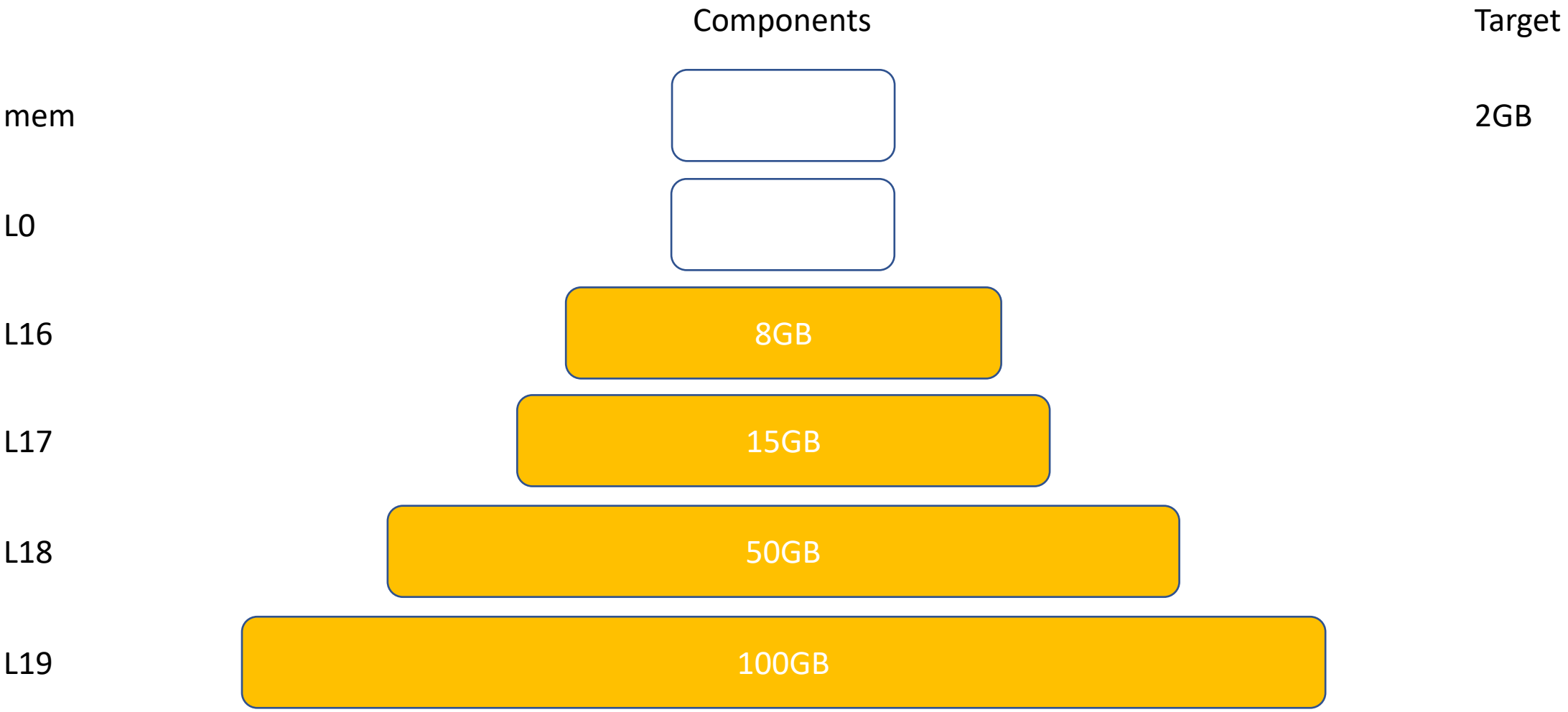
RocksDB universal flush



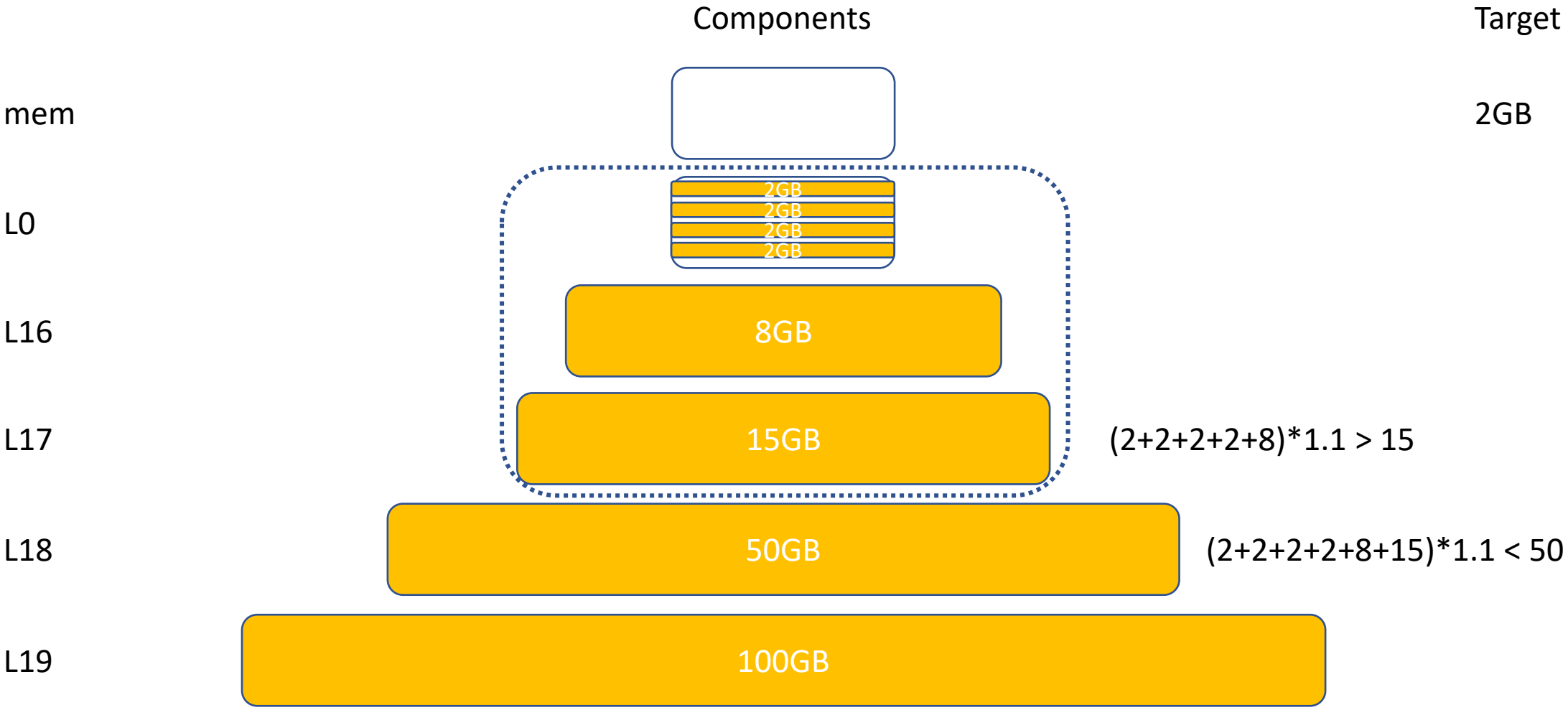
RocksDB universal size ratio compaction 1



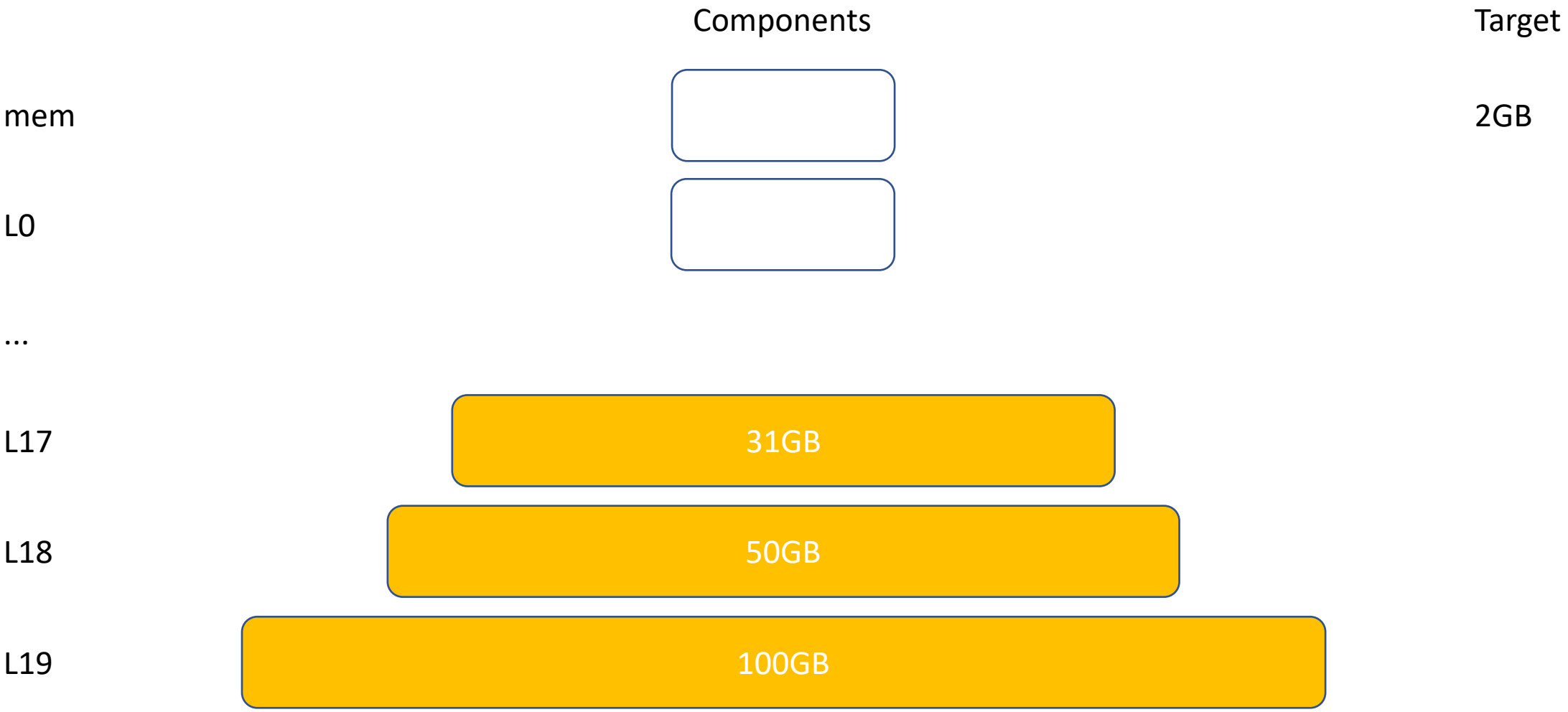
RocksDB universal



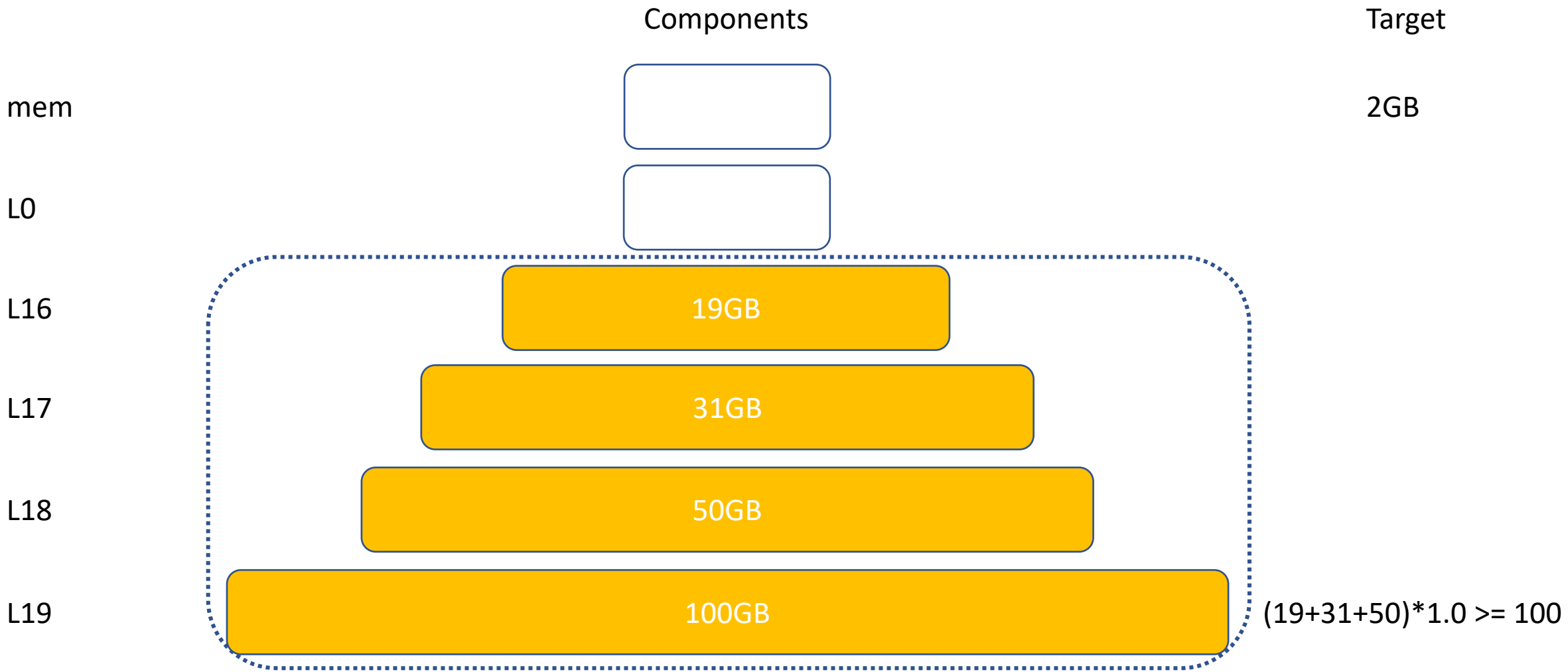
RocksDB universal size ratio compaction 2



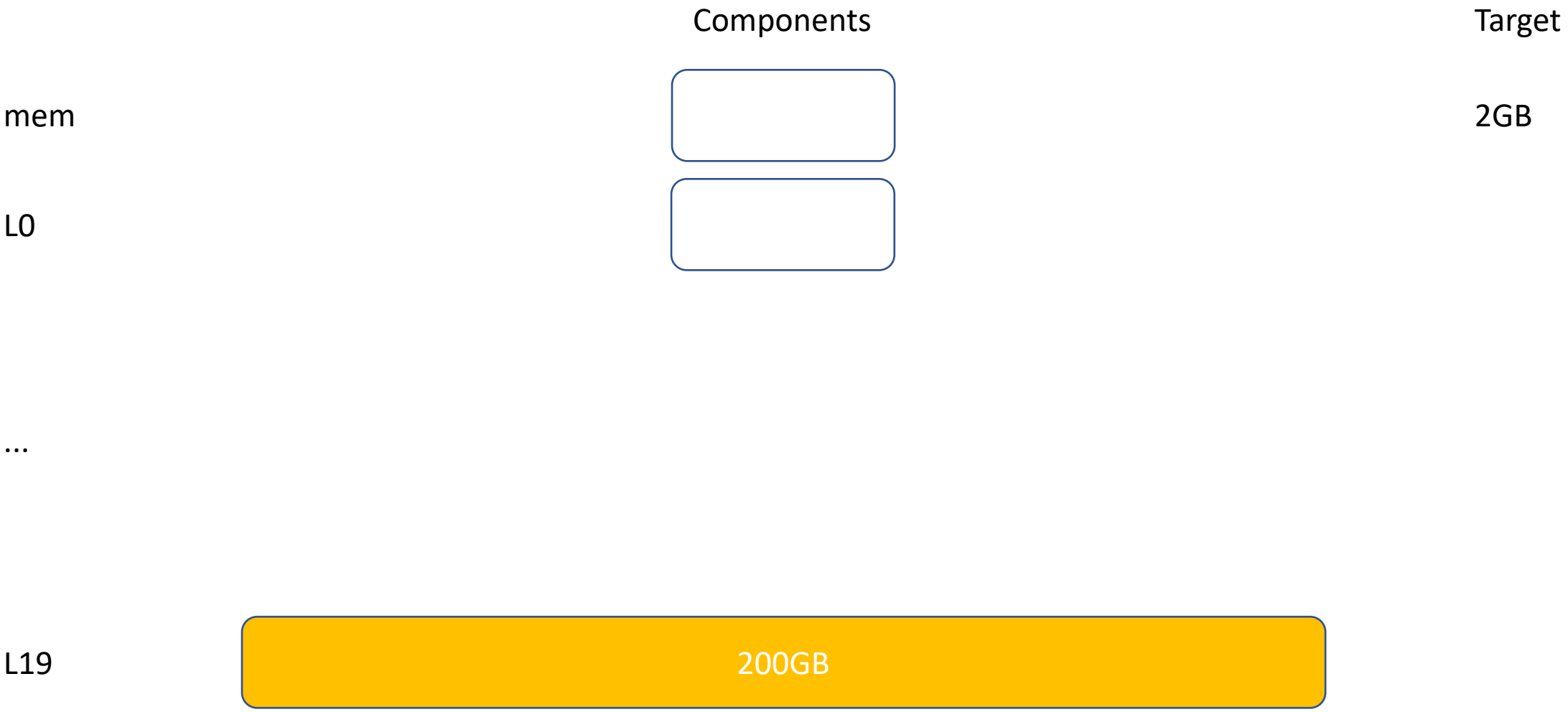
RocksDB universal



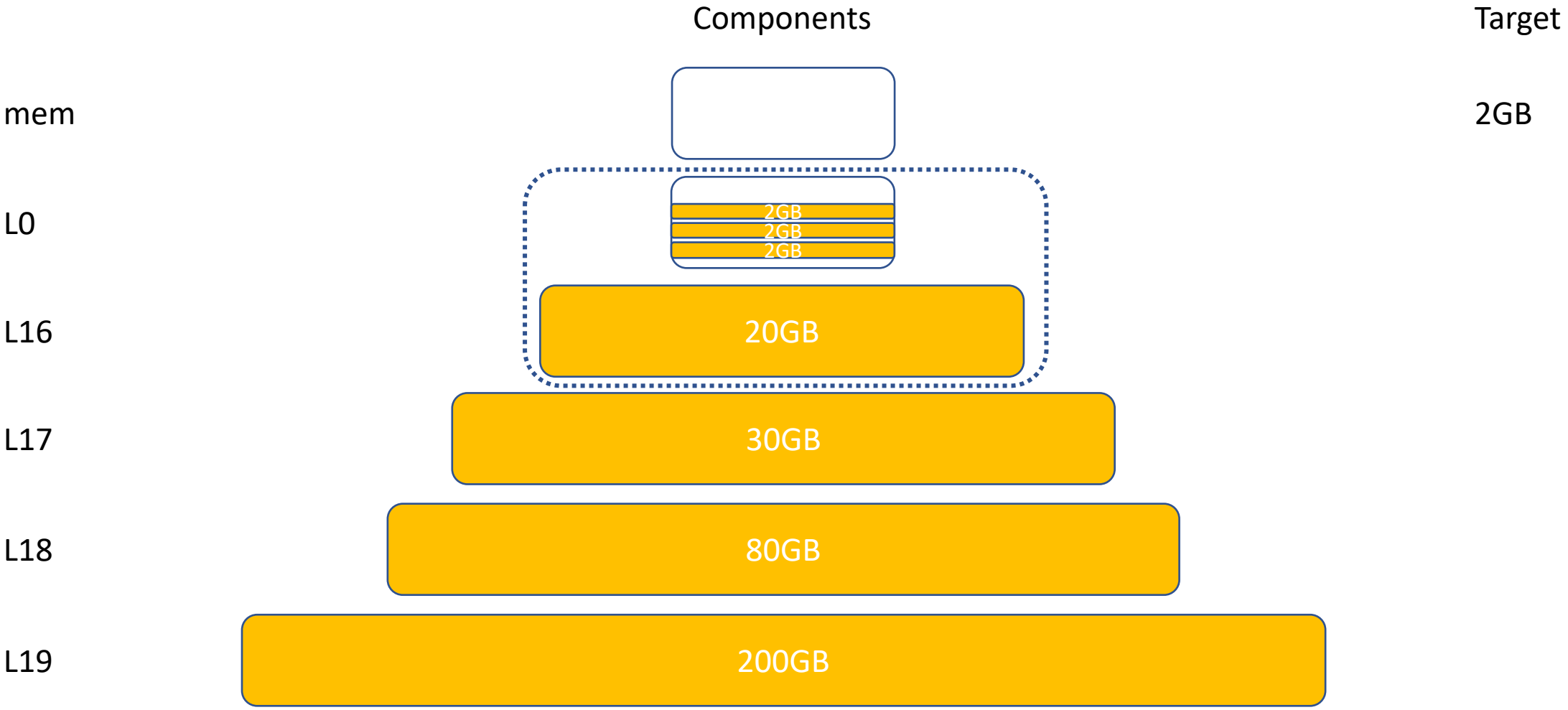
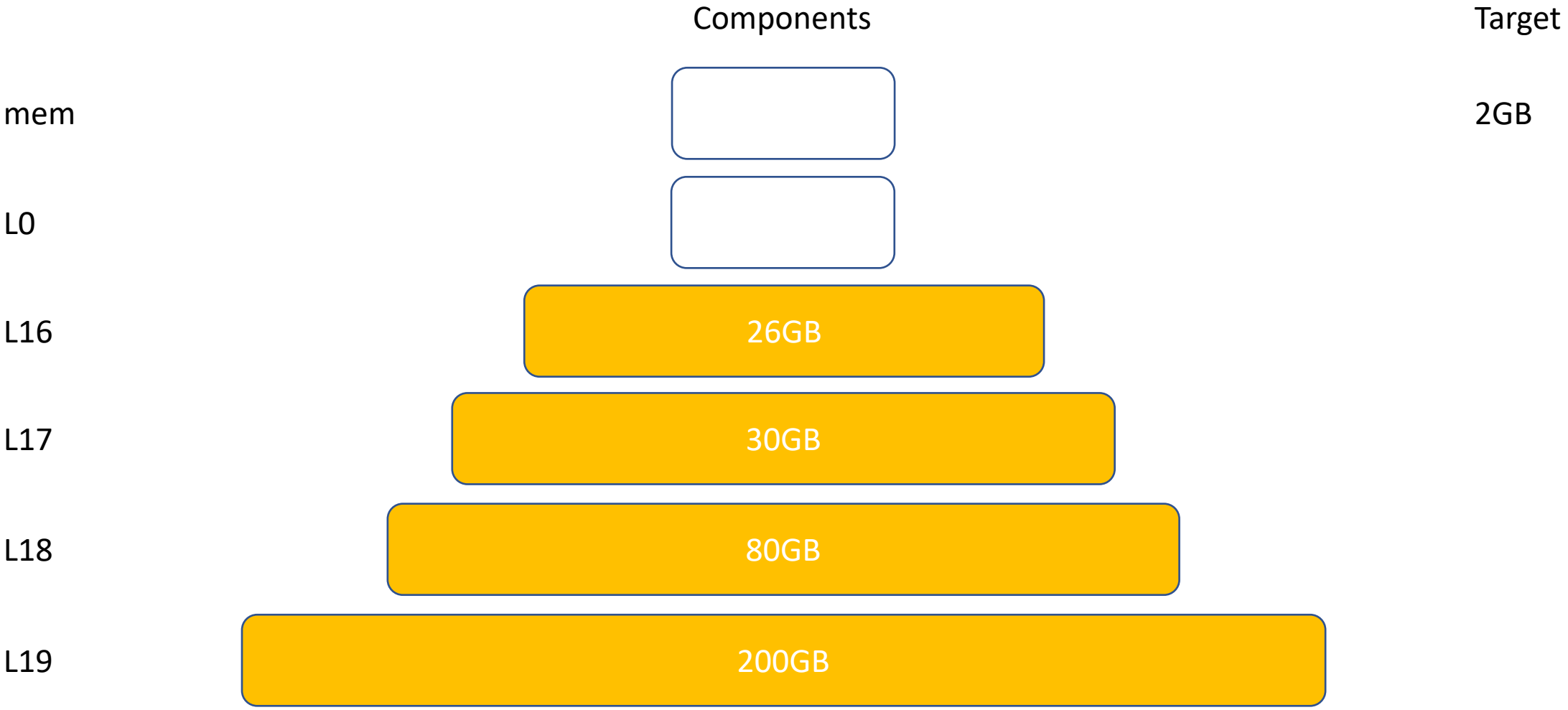
RocksDB universal size amp compaction



RocksDB universal



RocksDB universal read amp compaction



RocksDB universal read amp compaction