Performance degrades quickly beyond 1 process when using 2 vCore

Last updated by | Dandan Zhang | Apr 6, 2021 at 6:25 PM PDT

Contents

Issue

When customer run single query it takes between 16 - 18 seconds. but when he runs same query in 3 connections the query takes 25s

There are no high resource usage, and he needs to know why query time takes less when he ran it in concurrent connections even no resource consumption was done

Investigation/Analysis

Detailed investigation can be found in
https://portal.microsofticm.com/imp/v3/incidents/details/218801040/home

We further simplified the query

cmd="EXPLAIN (ANALYZE, VERBOSE, BUFFERS, COSTS, timing) select * from mike_test;"

```
                                                              QUERY PLAN

Seq Scan on public.mike_test  (cost=0.00..281819.45 rows=10000045 width=353) (actual time=0.066..1314.994 rows
 Output: ask, axioma_data_id, bid, mid, quote_convention, quote_date, source_asset_id, source_code, transactio
Buffers: shared hit=8098 read=173721
Planning Time: 0.222 ms
Execution Time: 1746.352 ms
(5 rows)

                                                              QUERY PLAN

 Seq Scan on public.mike_test  (cost=0.00..281819.45 rows=10000045 width=353) (actual time=0.026..2509.892 row
 Output: ask, axioma_data_id, bid, mid, quote_convention, quote_date, source_asset_id, source_code, transactio
Buffers: shared hit=14569 read=167250
Planning Time: 0.281 ms
Execution Time: 3331.903 ms
(5 rows)
```

I guess the expectation is one connection take N ms, two connections also take N ms i.e. scale linearly. But in this case, we are doing a Sequential scans which don't cache their data pages in the shared buffer and end up reading all the pages, likely they might be reading from the OS/Storage cache. This is a 2 V-core machine, I'm

not sure if that meant 2 Hyper threads or two real cores? Either way, let's assume two compute units. Any engine will have it's own house-keeping tasks, plus Azure service house-keeping tasks (containers) will usually take up one 1 core, and now we are left with 1 core. Running one query vs running two queries on the same core is what causing this behavior, as you can see, two sessions are taking approximately N * 2 ms against one session N ms. Please scale the server to 8 cores and see what happens to the test.

Mitigation

Scale up to have more vCores

How good have you found this content?