Performance differences due to Gen hardware changes

Last updated by | Sabrin Alsahsah | Mar 6, 2023 at 1:33 AM PST

Contents

Issue

Customer experiences a performance change following an internal reconfiguration (after load balancing or maintenance etc) or an SLO update.

Investigation/Analysis

The performance change can be caused by a change in the hardware generation following the reconfiguration. For certain types of workload if the generation hardware changes, for example, from Gen 8 to Gen 5 the customer may notice higher resource usage such as CPU or IO.

The **Performance / Config & Change History** section in the ASC report displays hardware changes. Perform regular performance troubleshooting steps first to evaluate the performance before and after the performance change, to confirm this not related to other factors, such as workload increase. Check if there was a Gen hardware change by correlating the time of the reconfiguration with the increase resource usage (CPU, IO etc.).

In Kusto, you can use the below query in order to check the database hardware generation:

```
let ClusterNames =
MonWiQdsExecStats
| where TIMESTAMP  > ago (1h)
| where LogicalServerName == '******'
| where database_name == '******'
| project ClusterName
| take 1
| union (
MonWiQdsExecStats
| where TIMESTAMP  > ago (1h)
| where LogicalServerName == '******'
| where database_name == '******'
| project ClusterName
| take 1
);
MonCapacityTenantSnapshot
| where TIMESTAMP  > ago (1h)
| where tenant_ring_name in (ClusterNames)
| summarize arg_max(TIMESTAMP, vm_size) by tenant_ring_name
```

Sample output

tenant_ring_name	TIMESTAMP ▲	vm_size
tr34685.westeurope1-a.worker.database.windows.net	2023-02-26 14:27:11.9390733	Internal_GPGen8HH_128iad

Mitigation

For internal use: As of Jun 2022: Unfortunately, this issue is currently by design. When new (faster) h/w was introduced we continued exposing it as "Gen5", so there is now a possibility that under the same SLO workloads can run faster, depending on what hardware they land.

The performance wouldn't be worse than original Gen5, but if the db initially landed on better h/w, and then db moved back to an older h/w, this would result in observed regression. Unfortunately, no control over that from the customer side today, and we cannot guaranty the hardware generation type that the customer app would land on after a internal reconfiguration or an SLO update. The SQL PG is aware of the issue and looking into how to address it in future. Given the current situation, however, it is not easy to resolve for current dbs. We are not able to retire all of the old hardware quickly. We cannot start charging more for those dbs running on better hardware either. So the effect is essentially some dbs may experience better performance for the same price. In order to actually guarantee better performance, the only mechanism currently is to explicitly request higher SLO.

As a side note, according to our experts CPU utilization increase on the older hardware (expect it to be as high as 20-40%; in this incident we observed avg. CPU utilization of 50% to go to 90%) does not always translate to proportionally lower throughput.

RCA

The behavior is publically documented in this article vCore purchasing model - Hardware configuration ☒. The relevant section:

Hardware configuration
Common hardware configurations in the vCore model include standard-series (Gen5), Fsv2-series, and DC-

series. Hardware configuration defines compute and memory limits and other characteristics that impact workload performance.

Certain hardware configurations such as Gen5 may use more than one type of processor (CPU), as described in Compute resources (CPU and memory). While a given database or elastic pool tends to stay on the hardware with the same CPU type for a long time (commonly for multiple months), there are certain events that can cause a database or pool to be moved to hardware that uses a different CPU type. For example, a database or pool can be moved if it is scaled up or down to a different service objective, or if the current infrastructure in a datacenter is approaching its capacity limits, or if the currently used hardware is being decommissioned due to its end of life.

For some workloads, a move to a different CPU type can change performance. SQL Database configures hardware with the goal to provide predictable workload performance even if CPU type changes, keeping performance changes within a narrow band. However, across the wide spectrum of customer workloads running in SQL Database, and as new types of CPUs become available, it is possible to occasionally see more noticeable changes in performance if a database or pool moves to a different CPU type.

Regardless of CPU type used, resource limits for a database or elastic pool remain the same as long as the database stays on the same service objective.

Internal Reference

IcM: [309162886: Performance variation after unplanned Outage](#) ⤢
IcM: [307549811: Elastic Pool stuck in update state when changing Maintenance time - Customer follow-up questions](#) ⤢

Root Cause Classification

Cases resolved by this TSG should be coded to the following root cause:
/Azure SQL v3/Performance/<related cause - e.g. CPU or IO>

How good have you found this content?

😊 😞