SIMON KUZNETS KHARKIV NATIONAL UNIVERSITY OF ECONOMICS

Serhii Minukhin, Yurii Savin

INVESTIGATION OF THE DTU PURCHASING MODEL FOR BIG DATA ON AZURE PLATFORM

Kharkiv, 2023

Scope of the study

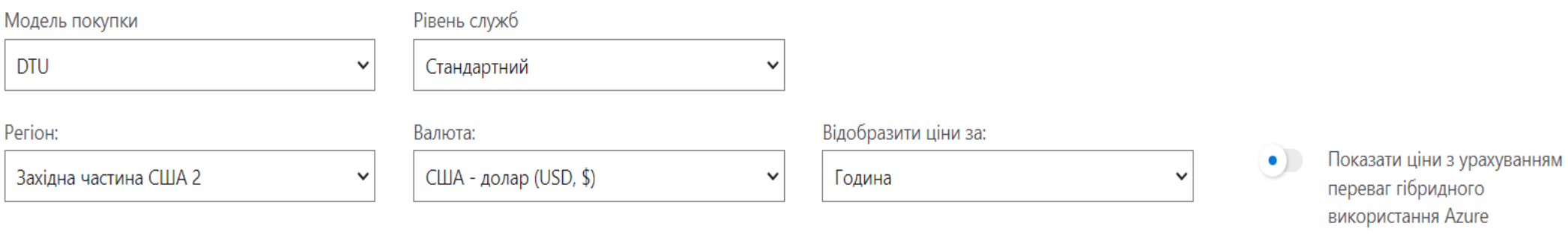
The purpose of this work is to investigate the possibilities of using the SQL Database service of the Azure platform for big data processing.

The object - relational database development processes on cloud platforms.

The subject is a study of the DTU purchase model for implementing different types of queries in SQL database service for big data processing on the Azure platform.

PAYMENT CHOICES

In DTU-based SQL purchasing models, a fixed set of resources is provided to a database or elastic pool at the following performance levels: "Basic, Standard, and Premium.
This model is best suited for users who prefer simple set monthly payments for simple predefined options.



Database transaction units (DTUs)

A database transaction unit (DTU) represents a blended measure of ***CPU, memory, reads, and writes***. Service tiers in the DTU-based purchasing model are differentiated by a range of compute sizes with a fixed amount of included storage, fixed retention period for backups, and fixed price. All service tiers in the DTU-based purchasing model provide flexibility of changing compute sizes with minimal downtime.

For a single database at a specific compute size within a service tier, Azure SQL Database guarantees a certain level of resources for that database (independent of any other database). This guarantee provides a predictable level of performance. The amount of resources allocated for a database is calculated as a number of DTUs and is a bundled measure of compute, storage, and I/O resources.

Azure SQL Database - DTU purchasing model

Basic service tier - *CPU-intensive workloads, a service tier of S3 or greater*

Compute size	Basic
Max DTUs	5
Included storage, GB	2
Max storage, GB	2
Max in-memory OLTP storage, GB	N/A
Max concurrent workers	30
Max concurrent external connections [1]	3
Max concurrent sessions	300

Standard service tier - *CPU-intensive workloads, service tier of S3 or greater*

Compute size	S0	S1	S2	S3
Max DTUs	10	20	50	100
Included storage, GB	250	250	250	250
Max storage,GB	250	250	250	1024
Max in-memory OLTP storage, GB	N/A	N/A	N/A	N/A
Max concurrent workers	60	90	120	200
Max concurrent external connections	6	9	12	20
Max concurrent sessions	600	900	1200	2400

Standard service tier - *CPU-intensive workloads, service tier of S3 or greater*

Compute size	S4	S6	S7	S9	S12
Max DTUs	200	400	800	1600	3000
Included storage,GB	250	250	250	250	250
Max storage, GB	1024	1024	1024	1024	1024
Max in-memory OLTP storage, GB	N/A	N/A	N/A	N/A	N/A
Max concurrent workers	400	800	1600	3200	6000
Max concurrent external connections [2]	40	80	150	150	150
Max concurrent sessions	4800	9600	19200	30000	30000

Resource (DTU) consumption of workload

To gain deeper insight into the resource (DTU) consumption of your workload, use query-performance insights to:
Identify the top queries by CPU/duration/execution count that can potentially be tuned for ***improved performance***.
For example, an I/O-intensive query might benefit from in-memory optimization techniques (https://learn.microsoft.com/en-us/azure/azure-sql/in-memory-oltp-overview?view=azuresql) to make better use of the available memory at a certain service tier and compute size.
Drill down into the details of a query to view its text and its history of resource usage.
Access performance-tuning recommendations that show actions taken by SQL Database Advisor.

https://learn.microsoft.com/en-us/azure/azure-sql/database/service-tiers-dtu?view=azuresql

Determine DTU utilization

Limit of a database *or an elastic pool, use the following* ***formula:***
avg_dtu_percent = MAX(avg_cpu_percent, avg_data_io_percent, avg_log_write_percent).

Limit of a database or an elastic pool, pick the largest percentage value *from the following:*
avg_cpu_percent, avg_data_io_percent, and avg_log_write_percent at a given point in time.
sys.dm_db_resource_stats, sys.resource_stats, and sys.elastic_pool_resource_stats DMVs.
https://learn.microsoft.com/en-us/azure/azure-sql/database/service-tiers-dtu?view=azuresql

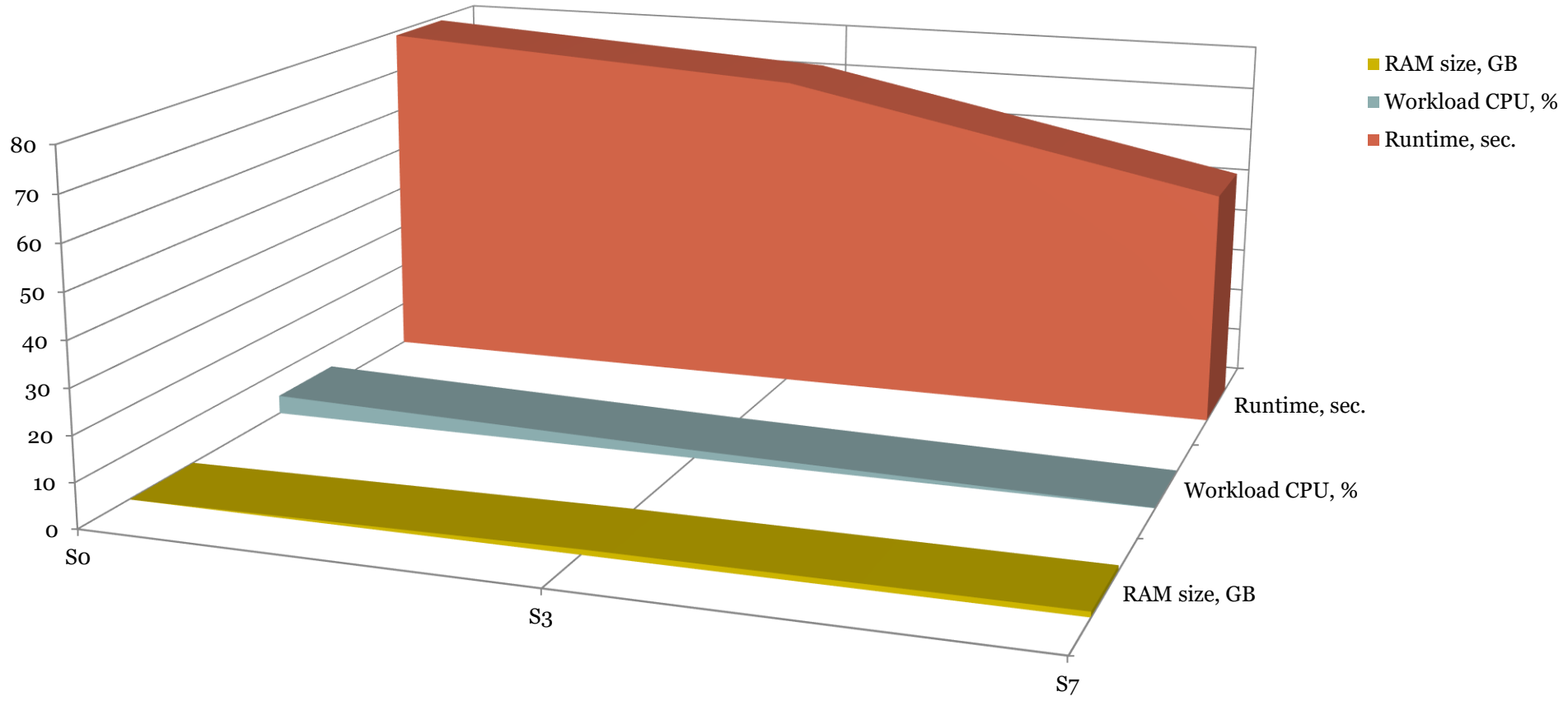
MODELING RESULTS ON A LOCAL RESOURCE

Type of query	Runtime, sec.	RAM size, GB	Workload CPU, %
SQ 1	15	1,17	23
SQ 2	1	0,9	3.8
SQ 3	3	1,05	20,1
CQ 1	1	1,43	25
CQ 2	1	1,22	4,5
CQ 3	5	1,14	23

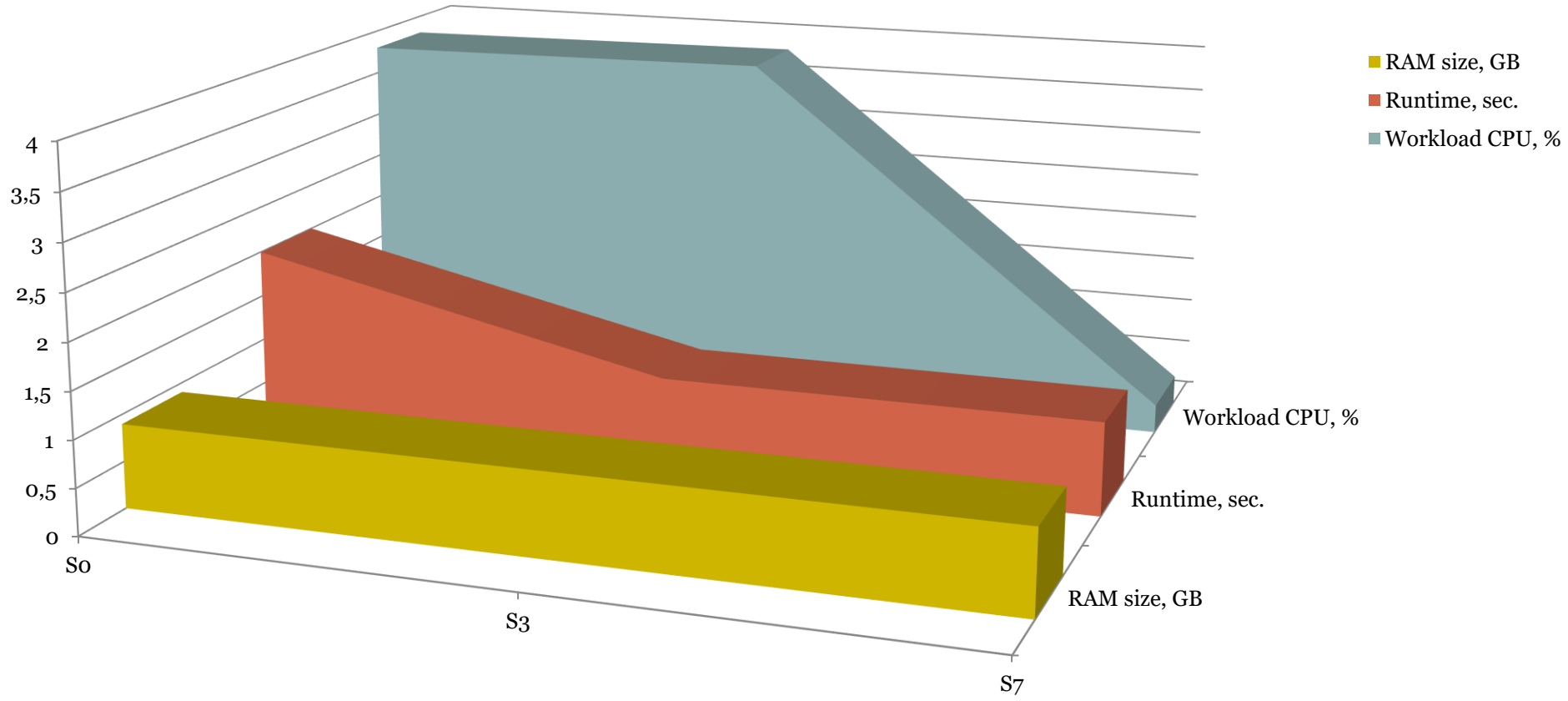
MODELING RESULTS ON AZURE (SQs)

Model, tier	Runtime, sec.	Workload DTU, %	Workload, CPU, %	RAM size, GB	Number of operations per 1 sec.
query SQ1					
S0	79	44	4,2	1,17	12671
S3	73	19	1,82	1,17	13712
S7	53	1,7	0,08	1,17	18887
query SQ2					
S0	2	52	3,8	0,9	353
S3	1	23	3,85	0,9	706
S7	1	2	0,31	0.9	706
query SQ3					
S0	93	68	5,2	1,05	2161
S3	99	22	1,54	1,05	2030
S7	74	1,8	0,9	1,05	2716

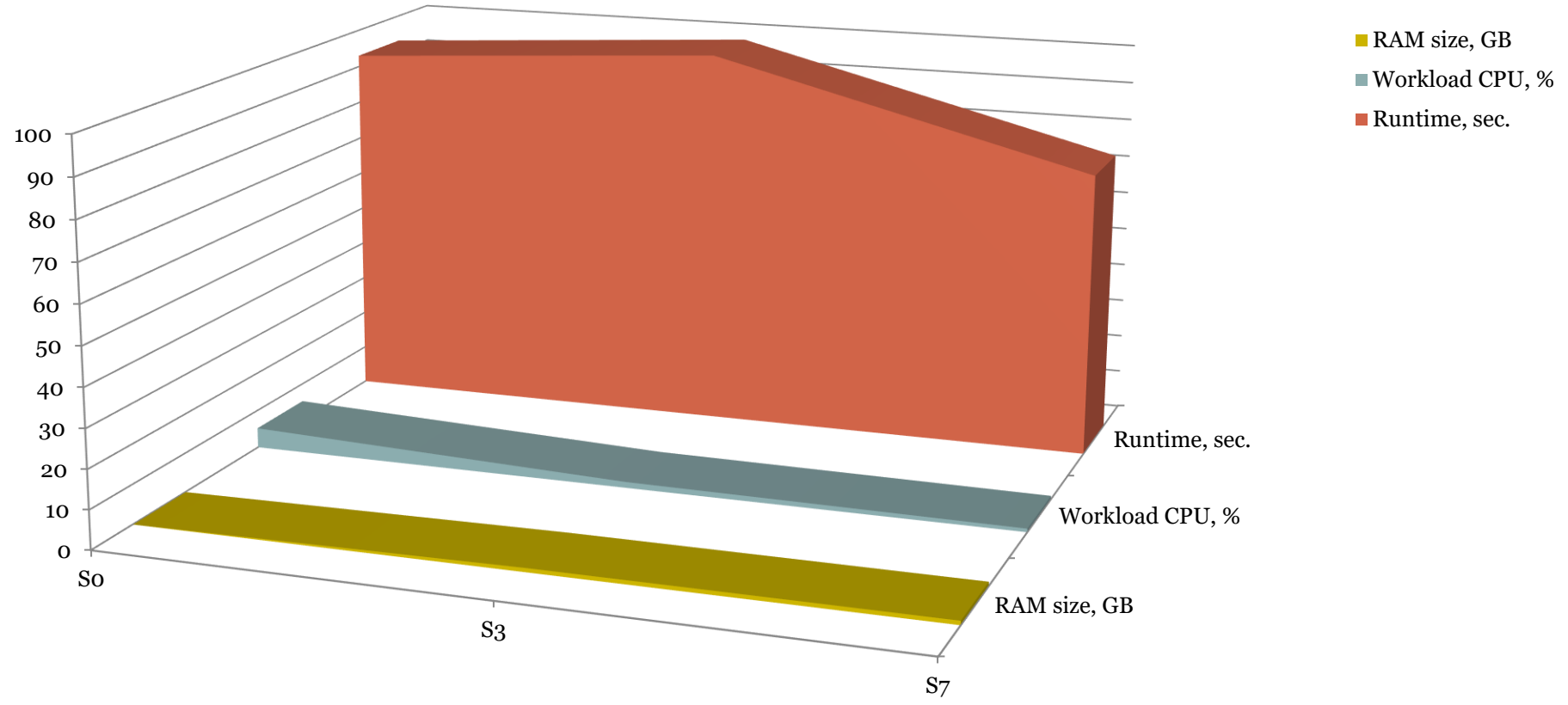
Comparative analysis of the simulation results for the SQ1 query



Comparative analysis of the simulation results for the SQ2 query



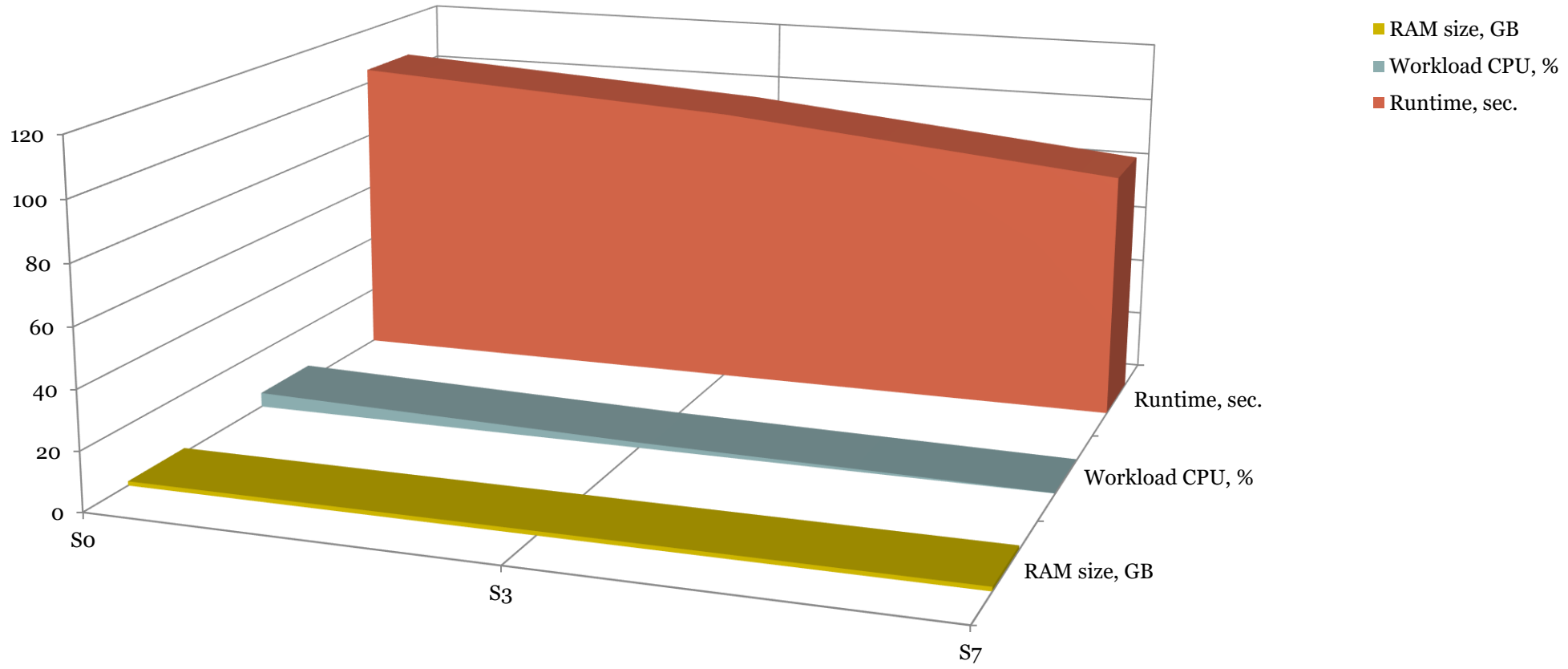
Comparative analysis of the simulation results for the SQ3 query



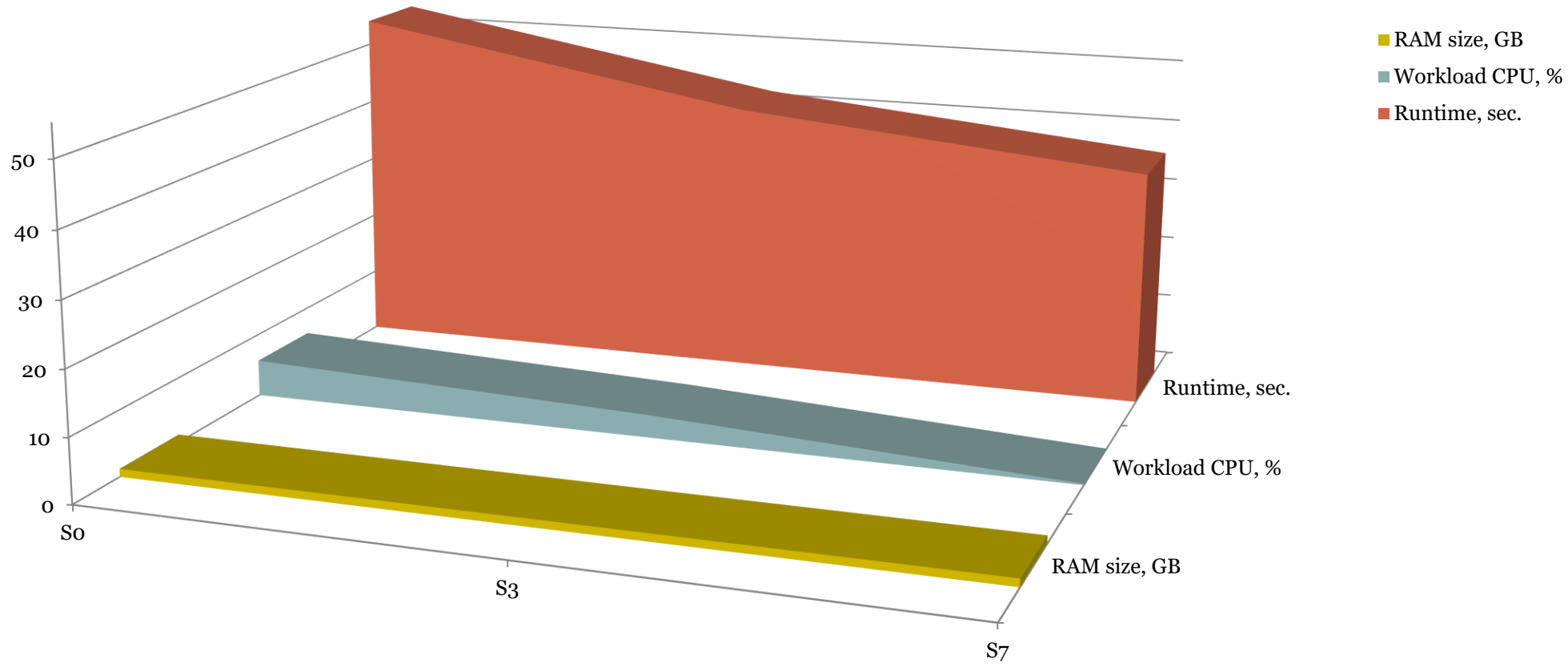
MODELING RESULTS ON AZURE (CQs)

Model, tier	Runtime, sec.	Workload DTU, %	Workload CPU, %	RAM size, GB	Number of operations per 1 sec.
query CQ1					
S0	105	94	4,95	1,43	9533
S3	97	37	1,82	1,43	10320
S7	84	3,7	0,08	1,43	11917
query CQ2					
S0	54	83	5,74	1,22	256
S3	43	32	3,57	1,22	322
S7	37	1,4	0,21	1,22	374
query CQ3					
S0	103	81	4,3	1,14	1951
S3	54	3,7	2,69	1,14	3722
S7	42	2,1	0,31	1,14	4786

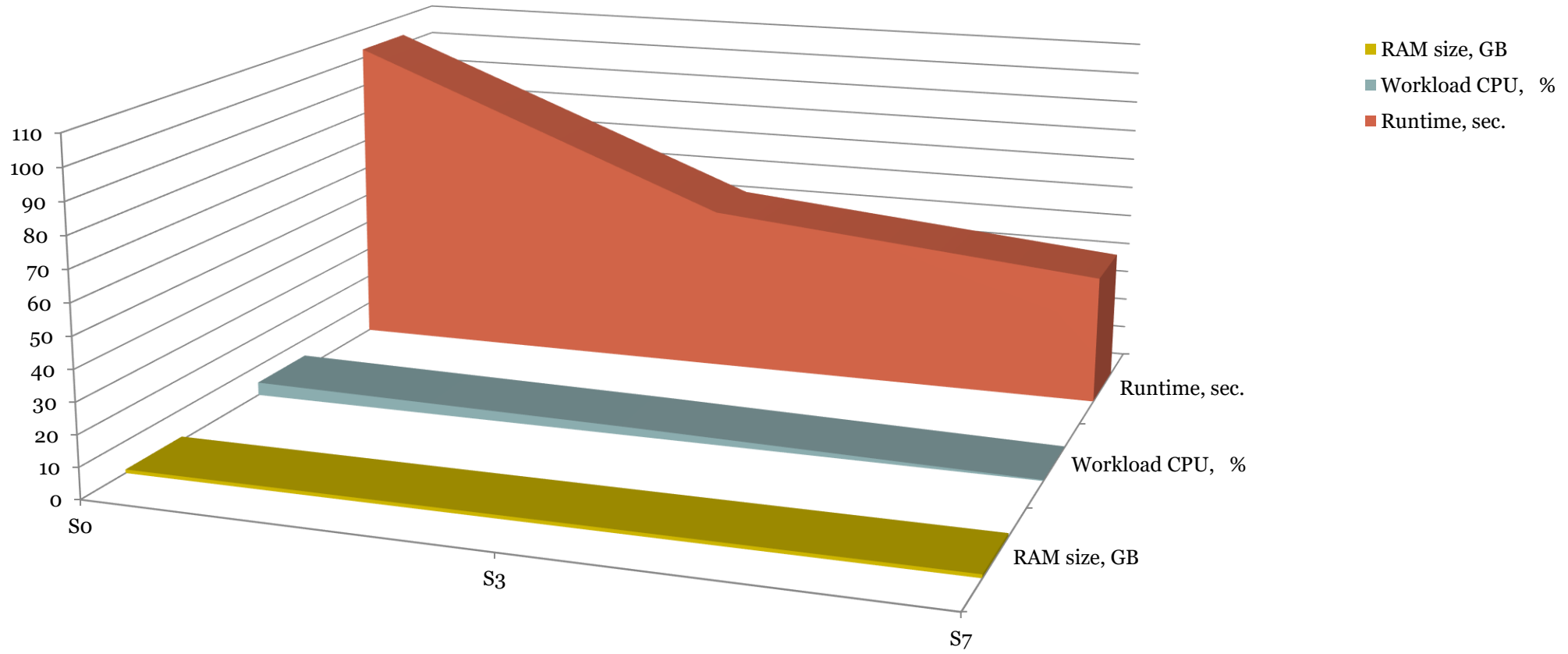
Comparative analysis of the simulation results for the CQ1 query



Comparative analysis of the simulation results for the CQ2 query

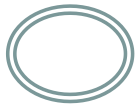


Comparative analysis of the simulation results for the CQ3 query



CONCLUSIONS

1. A model of working with relational databases to determine the performance of computing for the levels of the DTU acquisition model used for the Data Base SQL service on the Azure platform was developed.
2. To conduct a comparative analysis of the service performance, a 10 GB database was created on a local resource and, using the Data Base SQL service, on the Azure platform.
3. Queries to the database of varying complexity were created - 3 simple and 3 complex queries - to determine the effectiveness of using different levels of the DTU purchase model - S0, S3, S7.
4. For the comparative analysis, we used the database performance metrics - query execution time, processor load, and RAM capacity.
5. The results of the analysis showed that the following trends occur when using the Azure service: as the model level increases, the execution time and the level of CPU load decrease, while the amount of RAM remains almost unchanged. At the same time, there is no linear relationship between these indicators, which indicates the need for additional experiments to determine the possibility of other indicators affecting the work with the database.



Thanks for your attention !