SOFTWARE REQUIRMENT SPECIFICATION

Mudassar Shahbaz
mudassarshahbaz@gmail.com

CONTENTS

Technical Guide Concepts and Procedures Release Table

DOCUMENT VERSION	
v1.0	12-Feb-2023

SCENARIO

The company has a SQL backend to store machine statistic values, but they are concerned about its scalability limits. They want to replace it with a more suitable large-scale timeseries data backend. To evaluate different products, they need a simulator that can generate statistical data and feed it into different backends.

The simulator will be used to assess various performance metrics such as maximum write and query speeds, write and query performance with increasing amount of data stored, and the behavior of different backends at different scaling levels. The simulator will generate data for around 30 machine families, with each machine generating between 5 and 30 different types of statistics. Each statistic will have a machine identifier, timestamp, decimal/floating point value, statistic type, and 0 to 10 additional string-based dimensions.

The simulator needs to generate at least 1 trillion data points to simulate a year's worth of data from 5000 machines. The company has chosen 5 backend technologies for evaluation, which vary in terms of interface (ODBC, REST API, or dedicated SDKs). Queries will always follow a similar structure, aggregating values by time bucket and group by 0, 1, or 2 dimensional properties. The simulator will generate random queries across all machines and statistic types and output metrics to assess query speed.

Purpose

The current SQL backend used to store machine statistic values has shown scalability limitations and as a result, the company has decided to evaluate and replace it with a more suitable large-scale timeseries data backend. A simulator is needed to generate statistical data and feed it into different backends for evaluation.

The purpose of the simulator is to assess the performance metrics of the different backends being evaluated including maximum write and query speeds, write and query performance with increasing amount of data stored, and the behavior of different backends at different scaling levels.

Data Generation Requirements

- 30 different machine families with 1 machine belonging to 1 specific family.
- An individual machine generates between 5 and 30 different types of statistics, with a specific set of statistic types for each machine family.
- Each statistic datapoint consists of a machine identifier, timestamp, decimal/floating point value, statistic type, and 0 to 10 additional string-based dimensions.
- The goal is to generate 1 trillion data points, simulating a year's worth of data from 5000 machines.
- The simulator must generate data faster than real-world time and output metrics to assess write speed.

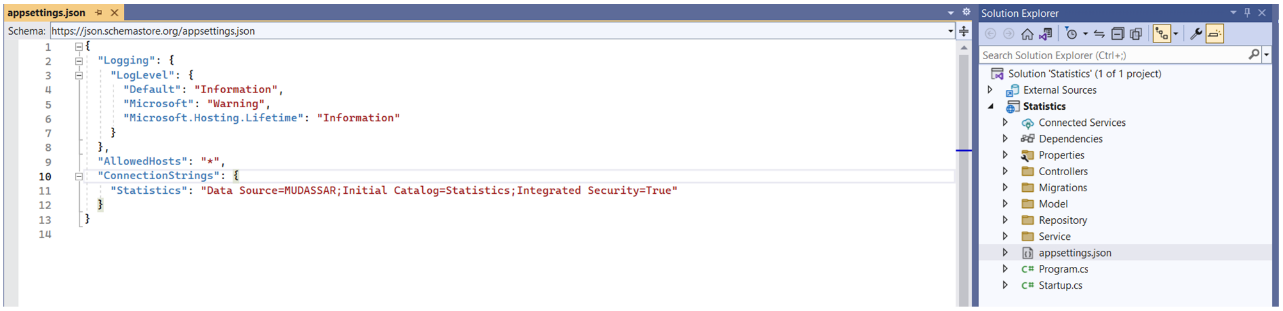
Backend Support Requirements

- The simulator will be evaluating 5 specific backend technologies, with the possibility of discovering additional candidates during the evaluation period.

How to run application:

1. Connection String Name Changed

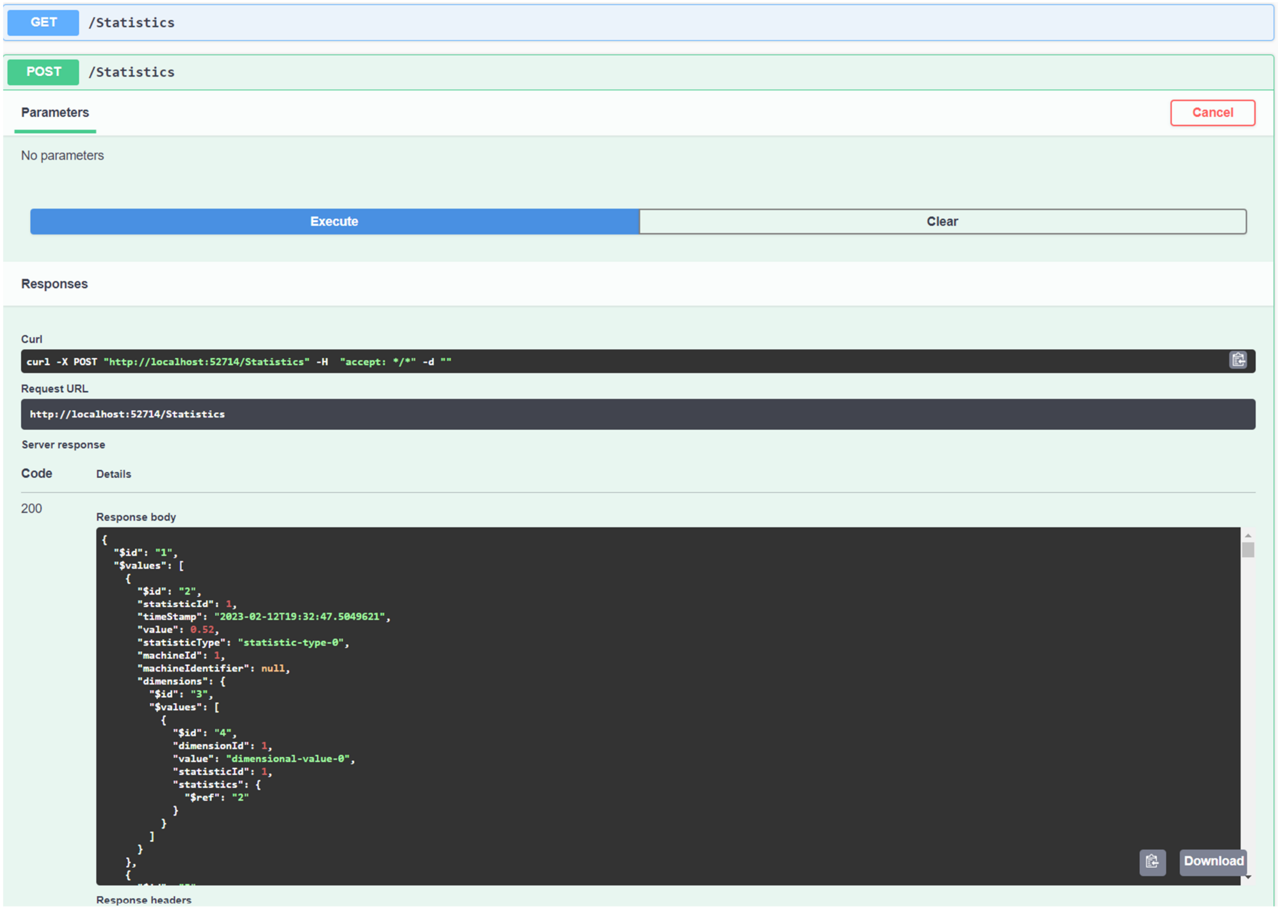
    - Change Connection String name to connect to the database in the file of **appsettings.json**



```json
{
  "Logging": {
    "LogLevel": {
      "Default": "Information",
      "Microsoft": "Warning",
      "Microsoft.Hosting.Lifetime": "Information"
    }
  },
  "AllowedHosts": "*",
  "ConnectionStrings": {
    "Statistics": "Data Source=MUDASSAR;Initial Catalog=Statistics;Integrated Security=True"
  }
}
```

2. Run Post API

- They automatically create database and required tables.
- EnsureCreatedAsync (If Database already EXISTED then it returns false)



```
curl -X POST "http://localhost:52714/Statistics" -H  "accept: */*" -d ""
```

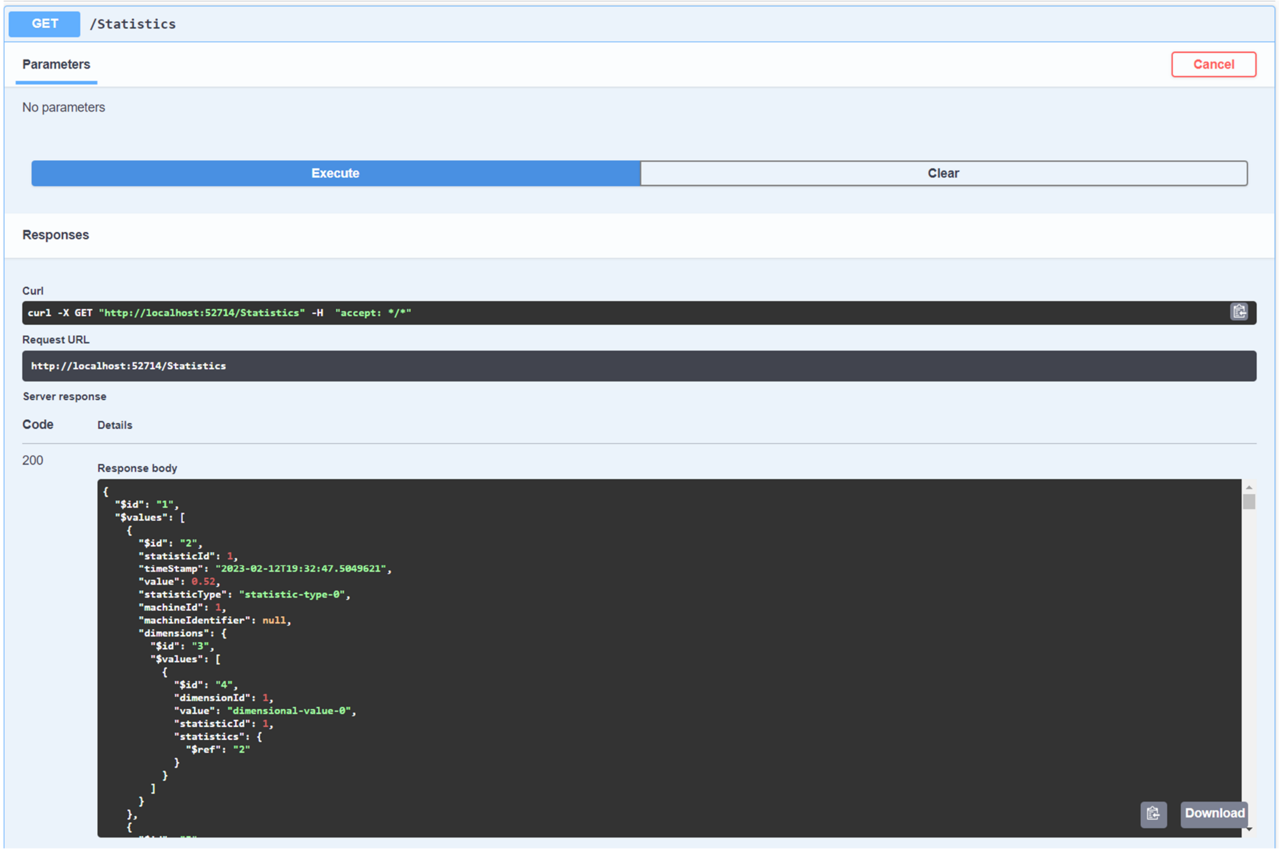
```
http://localhost:52714/Statistics
```

200

```
{
  "$id": "1",
  "$values": [
    {
      "$id": "2",
      "statisticId": 1,
      "timeStamp": "2023-02-12T19:32:47.5049621",
      "value": 0.52,
      "statisticType": "statistic-type-0",
      "machineId": 1,
      "machineIdentifier": null,
      "dimensions": {
        "$id": "3",
        "$values": [
          {
            "$id": "4",
            "dimensionId": 1,
            "value": "dimensional-value-0",
            "statisticId": 1,
            "statistics": {
              "$ref": "2"
            }
          }
        ]
      }
    },
    {
```

3. Run Get API

- They displayed the records in the Json Format.



Curl

```
curl -X GET "http://localhost:52714/Statistics" -H  "accept: */*"
```

Request URL

```
http://localhost:52714/Statistics
```

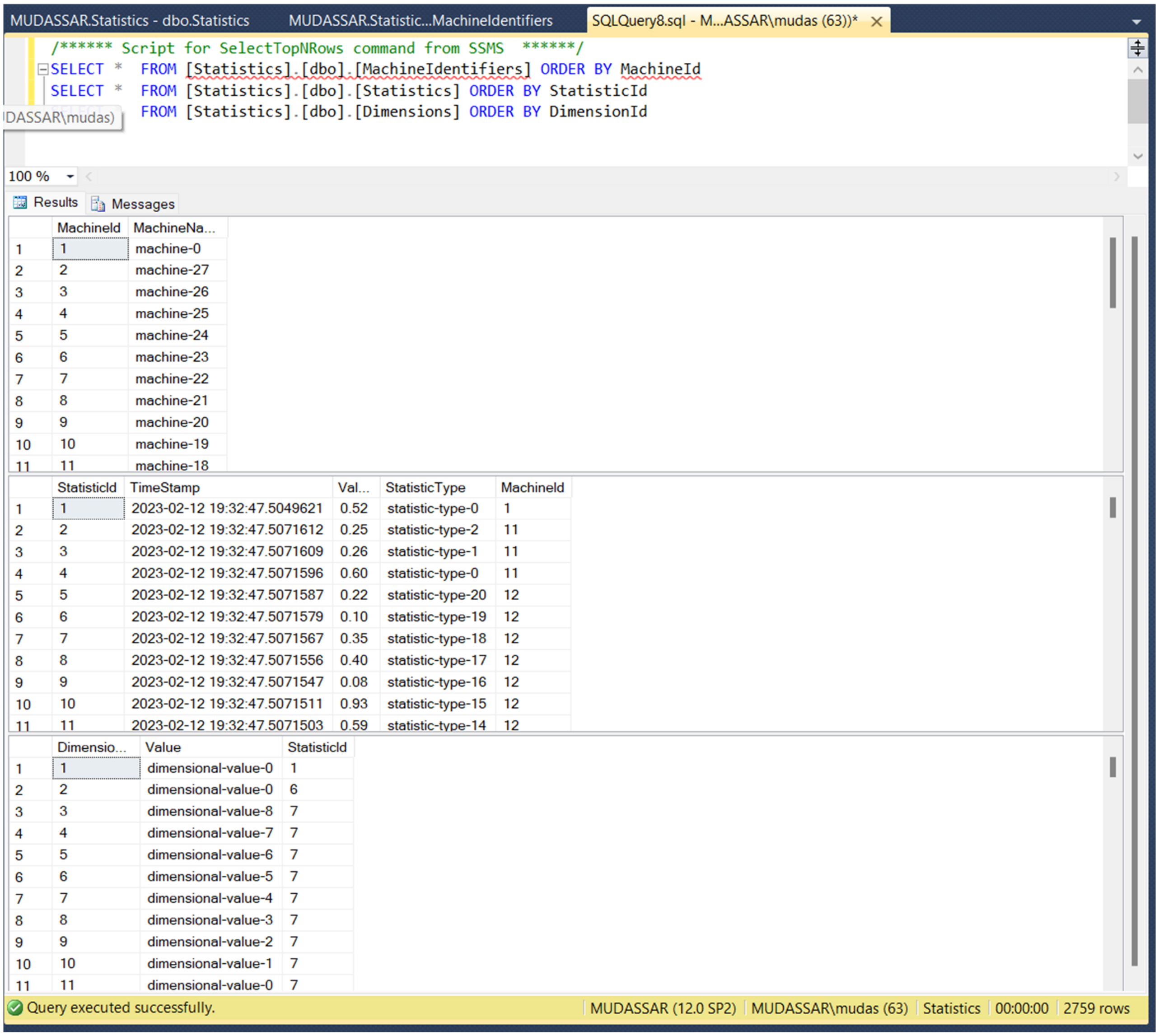
Server response

Code	Details
200	Response body

```
{
  "$id": "1",
  "$values": [
    {
      "$id": "2",
      "statisticId": 1,
      "timeStamp": "2023-02-12T19:32:47.5049621",
      "value": 0.52,
      "statisticType": "statistic-type-0",
      "machineId": 1,
      "machineIdentifier": null,
      "dimensions": {
        "$id": "3",
        "$values": [
          {
            "$id": "4",
            "dimensionId": 1,
            "value": "dimensional-value-0",
            "statisticId": 1,
            "statistics": {
              "$ref": "2"
            }
          }
        ]
      }
    },
    {
```

Database Table Structure:

Table name:	Information	DB Fields
MachineIdentifiers	Save Machine names	MachineId
		MachineName
Statistics	Save Statistics Information	StatisticId
		TimeStamp
		Value
		StatisticType
		MachineId
Dimensions	Save Dimensions	DimensionId
		Value
		StatisticId



MUDASSAR.Statistics - dbo.Statistics
MUDASSAR.Statistic...MachineIdentifiers
SQLQuery8.sql - M...ASSAR\mudas (63))*
DASSAR\mudas)
100 %
Results
Messages

```sql
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT *  FROM [Statistics].[dbo].[MachineIdentifiers] ORDER BY MachineId
SELECT *  FROM [Statistics].[dbo].[Statistics] ORDER BY StatisticId
          FROM [Statistics].[dbo].[Dimensions] ORDER BY DimensionId
```

	MachineId	MachineNa...
1	1	machine-0
2	2	machine-27
3	3	machine-26
4	4	machine-25
5	5	machine-24
6	6	machine-23
7	7	machine-22
8	8	machine-21
9	9	machine-20
10	10	machine-19
11	11	machine-18

	StatisticId	TimeStamp	Val...	StatisticType	MachineId
1	1	2023-02-12 19:32:47.5049621	0.52	statistic-type-0	1
2	2	2023-02-12 19:32:47.5071612	0.25	statistic-type-2	11
3	3	2023-02-12 19:32:47.5071609	0.26	statistic-type-1	11
4	4	2023-02-12 19:32:47.5071596	0.60	statistic-type-0	11
5	5	2023-02-12 19:32:47.5071587	0.22	statistic-type-20	12
6	6	2023-02-12 19:32:47.5071579	0.10	statistic-type-19	12
7	7	2023-02-12 19:32:47.5071567	0.35	statistic-type-18	12
8	8	2023-02-12 19:32:47.5071556	0.40	statistic-type-17	12
9	9	2023-02-12 19:32:47.5071547	0.08	statistic-type-16	12
10	10	2023-02-12 19:32:47.5071511	0.93	statistic-type-15	12
11	11	2023-02-12 19:32:47.5071503	0.59	statistic-type-14	12

	Dimensio...	Value	StatisticId
1	1	dimensional-value-0	1
2	2	dimensional-value-0	6
3	3	dimensional-value-8	7
4	4	dimensional-value-7	7
5	5	dimensional-value-6	7
6	6	dimensional-value-5	7
7	7	dimensional-value-4	7
8	8	dimensional-value-3	7
9	9	dimensional-value-2	7
10	10	dimensional-value-1	7
11	11	dimensional-value-0	7

Query executed successfully. | MUDASSAR (12.0 SP2) | MUDASSAR\mudas (63) | Statistics | 00:00:00 | 2759 rows

Tools to run for application.

- Visual Studio 2022
- SQL Server

Thanks & Best Regards,
Mudassar Shahbaz