**Data Science Final Project – Optimization of Servers**

**Introduction:**

The aim of this project is to find the impact of reducing servers and services on the entire Analytics application environment. When the initialization of this application, architect team proposed to have the maximum hardware resource, which includes RAM, Processors, Hard Disk to execute the application. The down size of having high configuration is ‘application license’ cost for higher Processors and its multiple instances on difference servers for load balancing and failover. The application was released to the user, and the performances of servers have been monitored to make sure not to cross 50% threshold. The result of servers’ performance showed that it did not cross 15% threshold in ordinary hours and 25% in peak hours, having these data the management would like to test the environment by reducing servers and services to get the optimum performance.

**Hypothesis:**

***Null Hypothesis***: By reducing servers and services will not impact the entire application environment

***Alternate Hypothesis:*** By reducing servers and services will impact the entire application environment

**Method/Model:**

As such the experiment has multiple variables to test; multiple regression method will be used

Y = β0 + β1X1 + β2X2 + β3X3 + β4X4 + β5X5 + ε

Where H0: β1X1 = β2X2 = β3X3 = β4X4 = β5X5 = 0 (There is no regression relationship)

**Data:**

For this experiment we have cross sectional data collected from six months periods which includes ‘load test’ data to compare before and after release of the environment.

***Independent Variables:***

X1 = Combination of ‘Number of Processor’, ‘RAM’,’ Services available on server 01 (can be labeled)

X2 = Combination of ‘Number of Processor’, ‘RAM’,’ Services available on server 02 (can be labeled)

X3 = Combination of ‘Number of Processor’, ‘RAM’,’ Services available on server 03 (can be labeled)

X4 = Combination of ‘Number of Processor’, ‘RAM’,’ Services available on server 04 (can be labeled)

X5 = Combination of ‘Number of Processor’, ‘RAM’,’ Services available on server 05 (can be labeled)

***Dependent Variable:***

Y = Total number of query processed by specified time with no error.

**Outstanding Questions, Assumptions, Risks, Caveats:**

1. ‘Load Test’ haven been conducted for different scenarios to get a real-time data, and this might lead into outlier.

Exp:

Single large query executed by 50 concurrent users

Single smaller query executed by 50 concurrent users

Single large and single smaller query executed by 25 users

2. Data collected for dependent variable might have different query executed on different servers. Data needs to be cleaned by selecting ‘Single Large’ or ‘Single Small’ query executed by all the servers.

3. All Servers can be a categorical variable

4. Check number of any other large query executed at the time of sample ‘Single Large’ queries.

5. Large or Small query can be compared by number of rows returned by database.

**Domain Knowledge:**

I architected this application, and tested with HP load test software to tune and get 50% threshold. Testing with ‘load test ‘software is time consuming and resource intensive (it took a moth), by using regression methods we believe we can reduce some of the steps for ‘load test’ and get the result quick with less cost.

**Goals and Criteria:**

The purpose of this experiment is to find the impact of reducing server, services will reduce the application performance by not changing any configuration in the existing environment. If the test is significant, then recommend the results and implement the changes on the environment. Making decision by this regression method is cost effective and not time and resource intensive compare to actual test. This method will also helpful when design and architect for newer version of this application.