Power Consumption Analysis for Proxy-based Space Saving

Nurul A. Emran*, Noraswaliza Abdullah, Nor Hafeizah Hassan

Fakulti Teknologi Maklumat dan Komunikasi,
Advanced Computing Technologies Centre (C-ACT),
Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia

*Corresponding e-mail: nurulakmar@utem.edu.my

ABSTRACT – With the rising concern for eco-friendly computing devices, many organizations seek greener options in managing their computing environment, especially their data centers. Thus, the green data center is a green computing concept that advocates efficient data storage with less energy consumption. This paper presents the results of implementing proxy-based space-saving to reduce energy consumption by optimizing space storage in data centers. A scatter plot analysis is conducted to examine the distribution of proxies discovered within a sample microbial data set. The proxies are evaluated in terms of the amount of space-saving and power-saving offered by them. The results yielded show the discovery of proxies within the data set under study, which can be used as the basis for proxy adoption decisions.

Keywords: *space-saving; functional dependency; scatter-plot analysis; proxy; green data center*

1. INTRODUCTION

In many organizations, data center consists of networked computers and storage, which is established for organizing, processing, storing, and disseminating large amounts of data. As an infrastructure that deals with large amounts of data, data centers consume a lot of electrical energy to ensure all the processes in data centers running smoothly. Consequently, it is undeniable that data center contributes to carbon footprint and global warming [1]. One characteristic of establishing a green data center is the efficient use of its storage because lesser storage requires less energy consumption. Nevertheless, studies on how to reduce energy consumption by optimizing space storage in data centers is limited.

The proxy-based space saving is a data management approach that offers space-saving via database schema modification [2]. Using proxies, redundant data can be removed, and free spaces are gained by deleting selected attributes from tables. The proxies substitute the deleted attributes in answering database queries.

The proxies are discovered by using the Functional Dependency Mining algorithm, where the proxy candidate and the droppable attribute for the table in the database must be discovered before the space-saving can be calculated. Power saving can be calculated after the amount of space saving is known.

The server is one of the essential components in the data centers. These servers handle hundreds to thousands of database users in large networks. As servers take a large amount of storage, they consume high power to run the data centers. The data centers are ongoing, which means they work every day, 24/7 per week. Based on this figure, we can estimate the power consumed by one server.

This paper addresses the question of how proxy-based space saving correlates with power saving to fulfill the gap in the current study.

2. METHODOLOGY

An experiment is conducted to evaluate the amount of space-saving and power-saving offered by proxies. The flow of the experiment is as shown in Figure 1. The experiment adopted the Comprehensive Microbial Resource (CMR) data sets that cover microbial bacterial genome types. The raw data set consists of genome collections with annotations, which are downloadable from the website link: ftp://ftp.tigr.org/pub/data. The raw data set needs to be converted into a format that is compatible with the proxy discovery step.

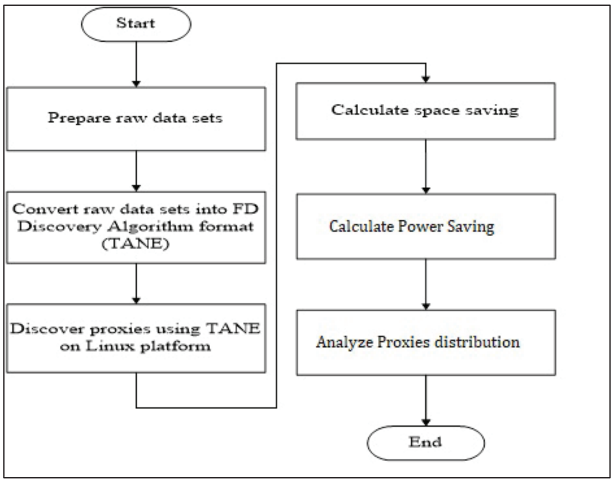

Figure 1: Experiment Flow for Proxy Evaluation

CMR datasets consist of 24 tables of data. In the experiment, four data sets with varying sizes were taken as samples which are Taxon (723 rows), Bug_attribute (10,165 rows), and ORF_Attribute (1 158 696 rows), and Feat_Link (854 379). FDTool (see [3]) is used to mine functional dependency (FD) scores among the columns. Based on the FD scores, the list of proxy candidates is generated. The proxy candidates need to be related to the dropped attribute to ensure that the replacement works. All the proxy candidates will be stored in a proxy map table, which must be considered in the calculation of space-saving (in kilobytes unit).

Suppose that the droppable column size is dv, original table size is tv, size of table with proxy is tpv and proxy map table size is pmv, the amount of space saving is calculated as follows:

$$\text{space saving } (\%) = (tv - tpv) / tv * 100\% \tag{1}$$

where,

$$tpv = (tv - dv) + pmv \tag{2}$$

A single server consumes energy between 500 to 1200 watts per hour. If the average power consumed by one server is 850 watts per hour, multiplied by 24, that equals 20 400 watts per day (20.4 kW) [3]. If we want to know the power consumed per week, we need to multiply 20 400 watts by seven days, which equals 142800 (142.8 kW). One server consumed about 7 446 000 (7446 kW) of power to run in a year. Given that one server has 2 TB disks attached and needs 850 watts per hour to run it, we can estimate the power consumed by one kilobyte of a disk, which is about 0.0000000004 kW of energy per day. The power consumption only can be estimated after we measure the space-saving offered by the proxies.

The power-saving is calculated in the kilowatts before converted into a percentage. Suppose that power consumption by droppable column is pd and the amount of estimated power usage per 1kB in a day is 0.0000000004 kW, the following formula is used:

$$\text{power saving } (\%) = (\text{space saving} * 0.0000000004 \text{ kW}) / pd * 100 \tag{3}$$

where,

$$pd = dv * 0.0000000004 \text{ kW} \tag{4}$$

Scatter plot analysis is used to examine the distribution of proxies based on their space-saving and power-saving scores. Scatter plots are among the widely used techniques for visual data exploration [4].

3. RESULTS AND DISCUSSION

Figure 2 shows a Scatter plot graph where ten proxies have been discovered in the experiment. The graph is plotted based on the increasing scores of space-saving and power-saving. If we divide the graph equally into four quadrants, we can see that there are four types of proxy quality. The proxies in this study occupy the top right and the bottom-left quadrants. The top right quadrant is the best-case scenario where the scores for both space-saving and power-saving are high. In contrast, the bottom left quadrant is the least desirable proxies as both space-saving and power-saving scores are low. None of the proxies can be found in the top left and bottom right quadrants.

The graph also depicts a positive correlation between the space-saving and power-saving scores as every increase in space-saving will also cause an increase in power-saving. Based on the behavior, the coarse selection of good proxies in terms of power-saving can be made based on space-saving percentage alone. Nevertheless, one's acceptability against the amount of reduction proxies may offer in terms of the total estimated kilowatts per day should also be considered.

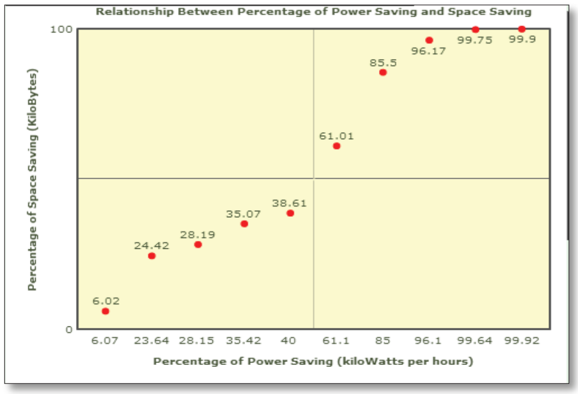

Figure 2: Distribution of proxies based on space-saving and power-saving scores

4. CONCLUSION

In conclusion, the results presented in this paper show the distribution of proxies by implementing the proxy-based space-saving approach against the CMR data set. The proxies were evaluated in terms of space-saving and power-saving criteria where, within the scope of the data set under study, useful proxies are present. The quality of proxies can be examined based on the Scatter plot quadrants. The adoption of proxies will depend on the space-saving and power-saving threshold that practitioners set. With the discovered correlation between the two variables, one can complement proxies with other green computing strategies to achieve greater energy saving. The findings contribute to the selection of proxies under the power consumption constraint.

ACKNOWLEDGEMENT

The authors would like to acknowledge the Universiti Teknikal Malaysia Melaka (UTeM) for financial support received through research grant PJP/2020/FTMK/PP/S01771.

REFERENCES

[1] A. Sabban, "*Introductory Chapter: Green Computing Technologies and Industry in 2021*," in Green Computing Technologies and Computing Industry in 2021, 2021

[2] N. A. Emran, N. Abdullah, and M. N. M. Isa, "Storage space optimisation for green data center," in *Procedia Engineering*, 2013, vol. 53, pp. 483–490.

[3] M. Buranosky, E. Stellnberger, E. Pfaff, D. Diaz-Sanchez, and C. Ward-Caviness, "FDTool: a Python application to mine for functional dependencies and candidate keys in tabular data," *F1000Research*, vol. 7, p. 1667, Jun. 2019

[4] D. A. Keim, M. C. Hao, U. Dayal, H. Janetzko, and P. Bak, "Generalized scatter plots," *Inf. Vis.*, 2010.