

PREDICTION BASED WORKLOAD PERFORMANCE EVALUATION FOR DISASTER MANAGEMENT SPATIAL DATABASE

N. Suryana^{1*}, M. S. Rohman², F. S. Utomo^{1,3}

¹ Fakultas Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia - nsuryana@utem.edu.my

² Department of Informatics Engineering, Universitas Dian Nuswantoro, Semarang, Indonesia - syaifurplus@gmail.com

³ Department of Information Systems, STMIK AMIKOM Purwokerto, Purwokerto, Indonesia - fandy_setyo_utomo@amikompurwokerto.ac.id

Commission VI, WG VI/4

KEY WORDS: Disaster Management, Large Spatial Database, Shortest Path, Dijkstra Algorithm, Case-Based Reasoning

ABSTRACT:

This paper discusses a prediction based workload performance evaluation implementation during Disaster Management, especially at the response phase, to handle large spatial data in the event of an eruption of the Merapi volcano in Indonesia. Complexity associated with a large spatial database are not the same with the conventional database. This implies that in coming complex workloads are difficult to be handled by human from which needs longer processing time and may lead to failure and undernourishment. Based on incoming workload, this study is intended to predict the associated workload into OLTP and DSS workload performance types. From the SQL statements, it is clear that the DBMS can obtain and record the process, measure the analysed performances and the workload classifier in the form of DBMS snapshots. The Case-Based Reasoning (CBR) optimised with Hash Search Technique has been adopted in this study to evaluate and predict the workload performance of PostgreSQL. It has been proven that the proposed CBR using Hash Search technique has resulted in acceptable prediction of the accuracy measurement than other machine learning algorithm like Neural Network and Support Vector Machine. Besides, the results of the evaluation using confusion matrix has resulted in very good accuracy as well as improvement in execution time. Additionally, the results of the study indicated that the prediction model for workload performance evaluation using CBR which is optimised by Hash Search technique for determining workload data on shortest path analysis via the employment of Dijkstra algorithm. It could be useful for the prediction of the incoming workload based on the status of the predetermined DBMS parameters. In this way, information is delivered to DBMS hence ensuring incoming workload information that is very crucial to determine the smooth works of PostgreSQL.

1. INTRODUCTION

Research by Firdaus et al. (2013) has described Merapi volcano is recognized to have resulted in many disadvantages which include the biggest ever eruption in the world. Merapi volcano can be categorised to be very dangerous as it has been erupting (peak activity) every two to five years. It is surrounded by a very dense settlement. Since 1548, the volcano has erupted 68 times. The risks of Merapi's eruption can be reduced by managing the disaster. The use of technology may help in the management of disaster such as in Merapi volcano. GIS technology offers many advantages (Qaddah and Abdelwahed, 2015). It is a powerful tool that integrates graphical interfaces and a variety of information to solve complicated problems, such as modelling the exact sites that have complex criteria and a type of data. GIS had been used in decision making at different approaches to disaster management regarding preparedness, mitigation, prevention, recovery, response and rehabilitation.

The Spatial data is necessary to conclude and fully function in GIS map to display all the data layers that were included in its design specially to manage a disaster. The number and size of spatial databases now are growing rapidly due to the significant amount of data needed to be obtained from X-ray crystallography, satellite images, and other scientific equipment. The need to specifically manage and analyse large spatial data as well as to support the operation of such data which is

determined by specific systems, algorithms and techniques becomes crucial. As being part of the GIS, the advantage and disadvantage of large spatial database structure have a close relation as to whether it can succeed or not in the entire study.

Complexity and fussiness of large spatial database are different with the conventional database. It makes relational database management systems fail to full fill request and need of query mentioned from (Cao et al., 2015). Every DBMS experiences complex workloads that are difficult to be managed by humans due to the current situation. Human experts take more time to handle database workload fast and accurately. Even in some cases, it may result in failures and thus leading towards undernourishment. System performance management includes identifying the causes of performance problems, measuring performance, and applying the tools and techniques available to handle issues of large spatial databases. According to Flores-Contreras et al. (2015), performance prediction on the database is useful for different purposes such as for capacity planning, load balancing, and resource usage optimisation among others since it allows the estimations of the response time of a system under a particular workload. Besides, performance prediction methods also provide insights for resource provisioning, workload management, and scheduling. In some cases, some scheduling algorithms, or resource managers use them as an auxiliary technique, for example, to improve the resource usage of the system. However, analysing the performance of these

systems under varying workloads and hardware configurations can be costly and time-consuming (Molka and Casale, 2015).

Other several problems influence database availability, such as the database can be inaccessible due to network problems or a virus. In these cases the database can be too slow and therefore does not satisfy the user's requests (Molka and Casale, 2015). When similar issues occur in the Disaster Management phase, the process of decision-making can be slow. It may lead into more losses and even increase the number of deaths. Predicting workload performance is a way to avoid contention rate that affects database performance by modelling how a workload reacts to changes in resource availability, users can make informed purchasing decisions, and providers can better meet their users' expectations.

A study by Sarwat (2015) upon discussing spatial data elucidates that delays are not tolerated in fundamental spatial data management system as queries need to be executed accurately and timely. Instead, the user is required to observe the specific information quickly and interactively change any query if it is necessary especially during the response phase of the disaster management. Then, the primary spatial database system is required to figure out efficient and effective ways to process user request as workload.

The implementation of workload management can be further utilised in handling this spatial data of the Indonesian Merapi volcano (Marhaento, 2016). Spatial data for this study is located in two different provinces, namely the Central Java and Yogyakarta of Indonesia, with an altitude of 2,914 metres elevation above sea level as depicted in the satellite images in Fig. 1. This satellite image shows typical vector and raster typical large spatial data.

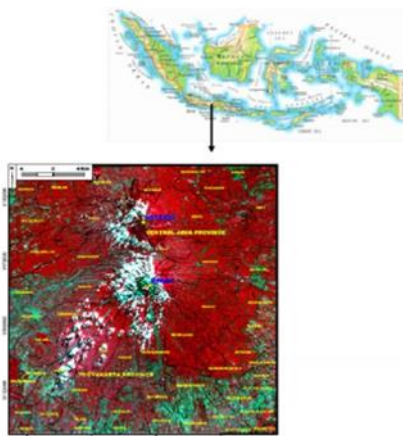


Figure 1. Merapi Volcano Spatial Data

This study requires large volume of spatial data in which implies we need the advanced management of workload to handle several requests from DBMS such as to locate needy people as well as to find the shortest route from and to the nearest places. Database administrators (DBAs) tune a DBMS based on their knowledge of the system and its workload. The type of the workload, specifically whether it is Online Transactional Processing (OLTP) or Decision Support System (DSS), is an essential criterion for workload tuning (Chiba and Onodera, 2016). Memory resources, for example, are allocated very differently for OLTP and DSS workloads.

Dijkstra Algorithm can be employed for the identification of the shortest route to an isolated area that is affected by the disaster and also identifies the nearest meeting point to the point of evacuation. Chen et al. (2014) posit that Dijkstra approach is the most effective and efficient to make shortest path analysis since improvements have been implemented by some researchers. The processes of workload encompassed in Dijkstra include sorting, scanning and joining. Elnaffar et al. (2007) state that is mainly the sort join and scan are primarily the kind of workload type associated with Decision Support System (DSS).

The job of DSS is to access the whole databases to perform Decision support and to join one Table with another Table to fulfil the needs of Dijkstra to be achieved. Another kind of workload besides DSS is OLTP. In the first place, OLTP performs workload that include daily activities like updating, inserting and deleting contain numbers of small transactions and that involve retrievals of individual records based on key values and updates. The OLTP workload does not include processing of whole databases as it only deals with specific Tables and an entity of fields. There are relatively few sorts and joins for this workload. DBAs therefore typically allocate memory to areas such as the buffer pools and log buffers while minimising the sort heap. Besides memory resources that are involved in the workload processes, there are also CPU Utilization and I/O activity trend.

However, there is no clear availability of the spatial database workload performance evaluation framework as well as to the parameters that may be suitable and applicable for the evaluation of spatial database workload performance. MySQL and PostgreSQL have their parameters. It has also been identified that different parameters of workload performance from several databases are generated based on the capability of each DBMS to monitor. In this regard, even in using the same database different researchers have used various parameters to evaluate workload performance. In other words, selection of suitable parameters as well as the availability of a framework is considered to be one of the crucial steps that may affect spatial database workload performance evaluation.

After taking the differences between the variables used for workload prediction into account, research by Zewdu et al. (2009) took 4 of 10 status variable before and after the experiment of another study. Abdul et al. (2014) have come up with three variables that give more information about the type of workload. The paper states that three variables can be gained from the process of key write, key read and Table lock. PostgreSQL does not include the parameters for evaluating workload performance as much as MySQL. In benchmarking PostgreSQL to evaluate like MySQL workload performance, three variables could also be monitored with the variable in PostgreSQL that include buffers activity, block read and the amount of specific lock.

Before predicting workload performance, variables are required to enable the decision of Workload predictor to produce the result. Previous researchers have developed some techniques prior to the problem of workload to handle several variables. One of the researchers has used machine learning approach to make workload prediction. The CBR (Cased Based Reasoning) is currently the most popular machine learning technique. CBR has been involved in earlier Workload prediction work (Abdul et al., 2014). It is believed that CBR can provide a suitable paradigm for microarray analysis of prediction, where the rules

that define the domain knowledge are difficult to obtain because usually, only a small number of training samples are available. Moreover, to select the most informative genes, another research is also implemented with Workload prediction for CBR using hash search technique hence enabling the problem to be solved. To give a maximum result, this study applied another machine learning such as Neural Network (NN) and Support Vector Machine (SVM) for comparison purpose with CBR using hash search technique. Therefore, this study proposes prediction based performance evaluation that could handle several types of workload besides processing user workload effectively and efficiently. The workload that is predicted may help the shortest route technology thus providing fast and accurate solutions.

The rest of the paper is organised as follows: Section 2 related work. Section 3 illustrates data acquisition workload prediction purpose. Section 4 draws Results and Discussion, Section 5 conclusion and the direction of future work.

2. RELATED WORK

Complexity and fussiness of large spatial database are different with the conventional database. It makes relational database management systems fail to fulfill request and need of query mentioned from (Cao et al., 2015). Sarwat (2015) elaborates on spatial data investigation session in which the user in the research would not tolerate delays introduced by the fundamental spatial data management system in the execution of effective and efficient queries. Abdul et al. (2014) utilised Artificial Intelligent (AI) technique in the prediction of workload. This specific AI technique used Fuzzy Logic. The Fuzzy Based Scheduler (FBS) which is based on OLTP and DSS percentage places the separated workload through the Fuzzy rules and membership functions. In another research by Holze et al. (2010), Discrete Fourier Transform (DFT) was used to predict the types of incoming workload. The workload periodicity detection was used in the research to simplify the perceiving periodicities in the initiation of timestamps of a single a model. Although the implementation of Decision and Classification trees as in Elnaffar et al. (2007) was very promising especially after pruning the tree to handle such large data, unfortunately, it led to inaccurate results as the value was cut leaf (knowledge) that was rarely used to improve the speed of analysis. Besides Decision and Classification trees, there are various approaches to machine learning but the most promising algorithm to be implemented is Case-Based Reasoning (CBR) as mentioned by Abdul et al. (2014).

3. DATA ACQUISITION

Data Acquisition involves the completed analysis to retrieve data for prediction purposes. Before data is captured, an analysis needs to be performed to get data based on the selected predictor. Workload data analysis discusses the workload parameters based on the PostgreSQL provided and query analysis that matches with the requirements of the study.

3.1 Workload Parameter

Workload prediction should provide predictions of incoming workload so that Database Management System can prepare memory resources, CPU utilisation and I/O activity. To predict incoming workload, it is necessary to identify the characteristic that identify and classify the workload into be a DSS or an OLTP workload types. When database receives query, there are

other parts of the system that are affected. This means that the affected part of the system should prepare some processing like CPU utilization and I/O activity that have some key block as to whether it comes from or enters disk and as for Table locks, there were number of requests to the disk that could be run immediately. So the three variables that affect memory resources in CPU utilization and I/O activity include key write, key read and Table lock activity (Abdul et al., 2014).

To deal with such Large Spatial Data, MySQL lacks the speed to extract overlapping regions in comparison to PostgreSQL (Khushi, 2015). Based on paper experience from Matuszka and Kiss (2014) for benchmarking large spatial data, PostgreSQL over performs MySQL even in other most widespread databases while PostgreSQL has proven to be the best in terms of query response time. It has also proven to be good for loading time. In benchmarking PostgreSQL for evaluating MySQL workload performance, three variables could also be monitored with the variables in PostgreSQL with buffers clean activity same with key write, block read same with key read and the amount of specific lock same with Table lock activity of parameters MySQL.

3.2 Workload Query Requirement

To support workload query requirements, there is analysis for determining the query that could be called OLTP and DSS workload simulation data that is related to shortest path analysis by Dijkstra algorithm and the existence of sort, join and scan characteristics. The existence of sort, join and scan characteristics for each query could be represented by OLTP workload. There must also be the insert, delete and update query to the support shortest path analysis by Dijkstra algorithm as follows:

3.2.1 Insert Query

In shortest path analysis, Dijkstra algorithm in this study is intended to generate the shortest and safest route. This includes how this algorithm avoiding the blocked route, and obstacle from the meeting point to the evacuation point. This computing process should be stored in specific fields. But it can be retrieved at anytime when it is needed.

3.2.2 Update Query

Update in this simulation of shortest path analysis exists because the cost was not defined and Dijkstra algorithm needs it to make route analysis. It also includes of update change value in the existing field together with the calculation of length between point to point.

3.2.3 Delete Query

Delete in specific field/Table was necessary to recycle memory so that there is no redundant data. After the result of Dijkstra analysis is stored in the Table of results, it must be deleted to make it available for another analysis with a different meeting point.

3.2.4 Select Query

Selecting route Table by accessing the fields to generate the result of Dijkstra algorithm result is one of another query of select. Another select query work is by requesting for the result of Dijkstra data after it has been stored in the taste of results.

3.3 Workload Data Snapshotting

Snapshotting was a necessary step to take the parameters report data due to incoming workload. Snapshotting could be done after the incoming workload by Dijkstra algorithm had successfully generated shortest path avoiding blocked route and thus providing a route to evacuate people from isolated areas to the evacuation place. After successfully generating some of the routes, workload data based on the proposed parameter was monitored as shown in Table 1. The following data in Table 1 was used to evaluate CBR with Hash Technique, NN, and SVM for Workload prediction. Buffers clean activity same with key write, block read same with key read and the amount of specific lock. All recorded data were monitored and generated with 520 workload data.

Queryid	Block read	Buffer clean	lock	W. Type
463466295	0	73	1	OLTP
484360484	178	0	2	DSS
493296969	0	40	4	OLTP
499873258	61628	15407	0	DSS
500158311	0	106	1	OLTP
509421166	279	93	2	DSS
513369138	174	0	0	DSS
514290828	0	50	3	OLTP
531690097	174	0	4	DSS
537658207	174	0	2	DSS

Table 1. Floating-point operations to classify a sample

4. RESULTS AND DISCUSSION

Some of the main topics in this section include construct prediction model, explanation about CBR prediction model. Moreover, the building of prediction model using CBR is optimized with Hash Search technique. The evaluation performed is used the optimization of Hash Search technique with Support Vector Machine and Neural Network so as to compare the performance of the proposed optimization. The technique used to validate and evaluate the model were cross-validation and confusion matrix. Cross-validation was a validation technique which divides the sample data randomly into k partition or subset. In this study, the value of k is set to 10. So mainly in 10 folds cross-validation, the sample data was divided into ten different subsets with the same percentage of information data in each subset. Since there were ten subsets of the sample, the system will run in 10 periods for testing procedure for each subset. Thus, the proportion of accuracy in this experiment was calculated by summing the single accuracy level in each run of testing.

4.1 Support Vector Machine

In 1992 at the Annual Workshop on Computational Learning Theory, SVM was first presented by Boser, Guyon, and Vapnik. SVM was introduced to address pattern recognition problems (Boser et al., 1992). Support Vector Machine is a relatively new technique to make predictions, both for the method of classification and regression. In SVM, we try to find an optimal separator function (hyperplane) that can separate two sets of data sets from two different classes. Where the function we are looking for is a linear function which can be defined as follows:

$$g(x) = \text{sgn}(f(x)) \quad (1)$$

with $f(x) = w^T x + b$, where $x, w \in R^n$ and $b \in R$. What we are looking for is the set of parameters (w, b), so that f

(xi) = <x, w> + b = yi for all i. in this SVM technique we look for the best hyperplane (separator/classifier function) that separates two kinds of objects/labels/classes. Finding the best hyperplane is equivalent to maximising margins or the distance between two sets of objects from different classes. If $wx_1 + b = +1$ is the supporting hyperplane of the +1 class and $wx_2 + b = -1$ is the supporting hyperplane of class -1, then the margin can be calculated by finding the distance of the two supporting hyperplanes. So that:

$$(wx_1 + b = +1) - (wx_2 + b = -1) \Rightarrow w(x_1 - x_2) = 2 \Rightarrow \left(\frac{w}{\|w\|} \cdot (x_1 - x_2)\right) = \frac{2}{\|w\|} \quad (2)$$

For linear classification in primal space, SVM optimisation formulation is as follows:

$$\min \frac{1}{2} \|w\|^2 \quad (3)$$

with $y_i (w x_i + b) \geq 1$, $i = 1, \dots, l$, where x_i is data input (parameter variable), y_i is data output (class variable) from x_i , w and b are parameters that value we looking for. If output data $y_i = +1$, then limit function are $(w x_i + b) \geq 1$, and if $y_i = -1$ then $(w x_i + b) \leq -1$. In case that is not infeasible, where the data cannot be grouped correctly then the formulation is $\min \frac{1}{2} \|w\|^2 + \sum_{i=1}^l t_i$, with $y_i (w x_i + b) + t_i \geq 1$, $t_i \geq 0$, $i = 1, \dots, l$, where t_i is variable slack. Since the data used is data with two classes then the technique used is linear SVM.

4.2 Neural Network

Neural Network (NN) is determined by three things: the relationship pattern between neurons, determining the linking weight and the activation function. There are three layers of process in the neural network method, called neural layers, the input layer (receiving the input data pattern from the outside illustrating the problem), the hidden layer / hidden layer and the output layer (a solution to a problem). In the case of this study, seeking the existing data, it is a simple matter with two classes, so the NN architecture is chosen as a single layers network. For NN formulation as follow:

If $\text{net} = \sum x_i w_i$ then activation function is $f(\text{net}) = f(\sum x_i w_i)$. Some activation function that used is:

1) Threshold limit

$$f(x) = \begin{cases} 1 & \dots & x \geq a \\ 0 & \dots & x \leq a \end{cases} \quad (4)$$

For bipolar number case, 0 change into a number -1 then equivalent change into:

$$f(x) = \begin{cases} 1 & \dots & x \geq a \\ -1 & \dots & x \leq a \end{cases} \quad (5)$$

2) Sigmoid function

$$f(x) = \frac{1}{1 + e^{-x}} \quad (6)$$

Usually used because of the value of function easy to be differentiated,

$$f(x) = f(x)(1 - f(x)) \quad (7)$$

3) Identity function

$$f(x) = x \quad (8)$$

Used if result output by NN is a random real number.

4.3 Case Based Reasoning

Case-Based Reasoning (CBR) is a reasoning model that unites problem solving, understanding and learning in memory processes. In workload prediction, those tasks could be done using the same case of the system, and the case is representative of the experience of essential learning to make workload predictor. In other words, the CBR method is used for problem-solving rather than other methods. CBR consists of several steps that can be taken to identify problem-solving through the prediction model. These steps were retrieved, reuse, revise and retain. These several steps should give a prediction of high accuracy and speed.

The searching of Euclidean distance on CBR could be optimised with Hash Search technique Fig. 2. Hash search distance searches the matching hash value between workload training data and new workload data. If it does not match, Hash search would find the nearest value and calculated with Euclidean distance formula. The model of CBR is as Fig. 2 (Abdul et al., 2014) since Hashing technique involves only finding the nearest value.

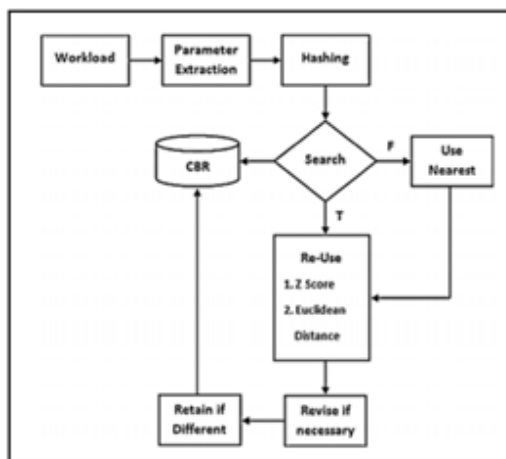


Figure 2. CBR structure

4.4 Validation and Evaluation

Validation is a process used to evaluate the prediction accuracy of a model. In k-fold cross-validation, the initial data are randomly partitioned into k mutually exclusive subsets or “folds”, D1, D2, ..., Dk, each of approximately equal size (Han et al., 2011). Training and testing are performed k times. In iteration i, partition Di is reserved as the test set, and the remaining partitions are collectively used to train the model. For the evaluation process, the confusion matrix is implemented to check the classification performance and accuracy. According to Han et al. (2011), a confusion Matrix is a tool used for model evaluation classification to predict an object which is true or not. A prediction matrix will be compared with the original input class. In other words, confusion matrix consists of actual information and prediction in classification.

Cross-Validation was being used in the experiment to validate the accuracy of the workload prediction model. For each experiment, a different result of confusion matrix was provided

depending on the performance. The cross-validation aims to make predictions of a new workload data that has never appeared in the dataset. Accuracy, precision, and recall were three key assessment and evaluation points used in this study. Since accuracy itself could be misleading especially for workload data problem. Thus, precision and recall were used to measure prediction exactness and completeness. Fig. 3 described confusion matrix table with two classes.

Confusion Matrix		Predicted Class		Class Recall
		Normal	Unclassified	
Actual Class	Normal	a	b	$\frac{a}{a+b}$
	Unclassified	c	d	$\frac{d}{c+d}$
		Class Precision		Accuracy
		$\frac{a}{a+c}$	$\frac{b}{b+d}$	$\frac{a+d}{a+b+c+d}$

Figure 3. Confusion Matrix Table with 2 Classes

As in the result of the performance of the proposed method had been shown with some percentages and value. The accuracy of the proposed Hash Search technique for CBR prediction model showed an excellent result of accuracy, precision and recall based on the determination of cross-validation in each experiment. The prediction for workload analysis using CBR with Hash search technique to determine the workload data speed improved well and had shown that execution time CBR with Hash Search technique smaller.

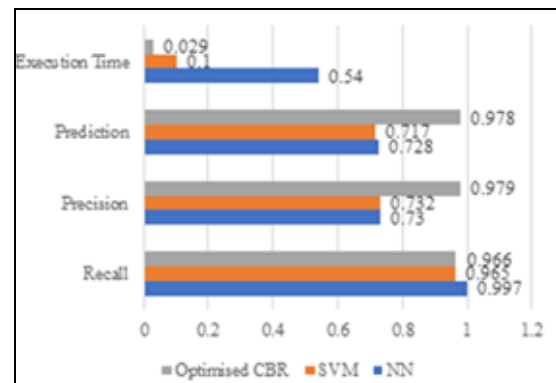


Figure 4. Comparison result of prediction for overall cross-validation

It was indicated that CBR with Hash Search technique provided a fast database management system to prepare the system resources than other machine learning algorithm Neural Network or Support Vector Machine. Comparison of CBR with Hash Search technique was presented in Fig. 4. CBR without and with Hash Search technique showed different of percentages accuracy, precision, and recall. CBR with Hash search showed a better percentage of accuracy, precision, recall and last CBR with Hash Search showed fastest execution time that needed for.

5. CONCLUSION

In this study, we have investigated the performance of CBR with Hash Search Technique algorithm in handling prediction of incoming workload. Since no standard parameters that may be suitable and applicable to evaluate spatial database workload performance. Therefore, this study is focused on PostgreSQL to analyse and find the parameters that matched workload performance status that was generated by MySQL. These parameters were a number of block read, buffer clean and lock

to make resources of the system applicable to receive a change of CPU allocation, I/O memory process and RAM. So, any possible change could be prepared by identifying the workload prediction so DBMS could be performed efficiently and effectively.

In handling large spatial database especially spatial data from Merapi Volcano satellite, there was a need to find an accurate and fast prediction of incoming workload so system recourses could be more efficient and effective especially in giving decision making of disaster management response. Case-Based Reasoning (CBR) prediction model as machine learning had been proposed to answer the research question. CBR had used old experiment data to predict new experimental data. When there is no past experiment data, the similarity of new experimental data performed until new experience data became new knowledge of CBR. So, the need for accurate prediction of workload could be the answer as shown in the results of the fourth section prediction evaluation to evaluate workload performance efficiently and efficiently. In handling fast prediction for incoming workload spatial data, CBR prediction model had been optimised with Hash Search technique to make matching similarity become fast. Optimization with Hash Search technique was involved in retrieving steps from CBR. In addition, with the help of Euclidean distance to minimise the finding to the nearest destination with smaller execution time than without using hash search optimisation.

The overall conclusion of the prediction that was performed with Hash Search technique in CBR together with the selected parameters has proven to fulfil the requirements to handle large spatial database. Therefore, future works for evaluation can be continued to system resources after workload handle large spatial data. It is a must to have an enlarged system so that schedules could predict and thus, on the workload suggest the database management.

REFERENCES

- Abdul, M., Muhammad, A. M., Mustapha, N., Muhammad, S., Ahmad, N., 2014. Database workload management through CBR and fuzzy based characterization. *Applied Soft Computing*, 22, pp. 605–621.
- Boser, B. E., Guyon, I. M., Vapnik, V. N., 1992. A training algorithm for optimal margin classifiers. In: *Proceedings of the fifth annual workshop on Computational learning theory*, pp. 144–152.
- Cao, X., Cong, G., Guo, T., Jensen, C. S., Ooi, B. C., 2015. Efficient Processing of Spatial Group Keyword Queries. *ACM Transactions on Database Systems*, 40(2), pp. 1–48.
- Chen, Y. Z., Shen, S. F., Chen, T., Yang, R., 2014. Path optimization study for vehicles evacuation based on Dijkstra algorithm. *Procedia Engineering*, 71, pp. 159–165.
- Chiba, T., Onodera, T., 2016. Workload characterization and optimization of TPC-H queries on Apache Spark. In: *IEEE International Symposium on Performance Analysis of Systems and Software*, pp. 112–121.
- Elnaffar, S., Martin, P., Schiefer, B., Lightstone, S., 2008. Is it DSS or OLTP: Automatically identifying DBMS workloads. *Journal of Intelligent Information Systems*, 30(3), pp. 249–271.
- Firdaus, Hidayat, S., Sahroni, A., Setiawan, H., Akbar, R., 2013. Design technology in wireless mesh network system for eruption disaster mitigation of Merapi volcano. In: *3rd International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering*, pp. 255–259.
- Flores-Contreras, J., et al., 2015. A performance prediction model for database environments: A preliminary analysis. In: *IEEE 17th International Conference on High Performance Computing and Communications, IEEE 7th International Symposium on Cyberspace Safety and Security, IEEE 12th International Conf on Embedded Software and Systems*, pp. 1707–1712.
- Han, J., Kamber, M., Pei, J., 2011. *Data Mining Concept and Technique, 3rd ed.* San Francisco: Morgan Kaufmann.
- Holze, M., Haschimi, A., Ritter, N., 2010. Towards workload-aware self-management: Predicting significant workload shifts. In: *IEEE 26th International Conference on Data Engineering Workshops*, pp. 111–116.
- Khushi, M., 2015. Benchmarking database performance for genomic data. *Journal of Cellular Biochemistry*, 116(6), pp. 877–883.
- Marhaento, H., 2016. Gis-Based Analysis For Assessing Landslide And Drought Hazard In The Corridor Of Mt. Merapi And Mt. Merbabu National Park, Indonesia. *Journal of Geomatics and Planning*, 3(1), pp. 15–22.
- Matuszka, T., Kiss, A., 2014. Experimental evaluation of some geodata management systems. In: *9th International Conference on Computer Engineering and Systems*, pp. 92–97.
- Molka, K., Casale, G., 2015. Experiments or simulation? A characterization of evaluation methods for in-memory databases. In: *11th International Conference on Network and Service Management*, pp. 201–209.
- Qaddah, A. A., Abdelwahed, M. F., 2015. GIS-based suitability modeling for seismic Stations: Case study of the northern Rahat volcanic field, Saudi Arabia. *Computers & Geosciences*, 83, pp. 193–208.
- Sarwat, M., 2015. Interactive and Scalable Exploration of Big Spatial Data - A Data Management Perspective. In: *16th IEEE International Conference on Mobile Data Management*, pp. 263–270.
- Zewdu, Z., Denko, M. K., Libsie, M., 2009. Workload Characterization of Autonomic DBMSs Using Statistical and Data Mining Techniques. In: *International Conference on Advanced Information Networking and Applications Workshops*, pp. 244–249.