

Utilization of Cloud Database on Criminal Background Check

Michael N. Mutisya, Andrew M. Kahonge

^{1,2}School of Computing and Informatics, University of Nairobi, Kenya

Abstract

The rate of Criminal background check has recently gone high in Kenya. Career candidates have been subjected to pre-screening before they are engaged by employers or before they vie for any elective post. The Directorate of Criminal Investigation Kenya is a unit of National Police Service that is tasked with the responsibility of carrying out criminal background screening of the citizens and issuing of Police Clearance Certificate. The current system is centralized and rocked by some challenges. It takes about two weeks to process and release one Police Clearance certificate. The increased demand of these clearance certificates has strained the current system due to handling of increased workload.

Cloud computing has been used by law enforcement agencies in crime reporting and storage.

This study used Mixed methods research. Its central premise is that the use of quantitative and qualitative approaches in combination provides a better understanding of research problems. Under this approach interviews and questionnaires were used to evaluate current processes as well as enquire the applicability of intended solution. Then Extreme software development methodology was employed in development of the prototype.

In this paper, we propose application of Cloud Database model and its collaboration with nodes in enhancing progression in criminal background screening. The findings of testing and evaluation of the prototype revealed that integration and utilization of Cloud database has positive implications in criminal background check where service delivery in processing and issuing of Police Clearance Certificate from every Criminal Investigation Office is improved.

Keywords

Cloud Database, Cloud Computing, Database as a Service (DBaaS), Node Splitting, Service-Level Agreements (SLA)

I. Introduction

There has been great progression in screening mostly the use of criminal background check in Kenya. Career candidates have been undergoing pre-screening before they are offered employment. Employers have been slack in hiring or retaining employees who have been having defective history and who subsequently engage in violence. Pre-screening has also been extended to tenants, vendors and volunteers. At several circumstances citizens might be asked to produce a criminal background check. Nowadays, companies, government sectors and NGO's demand one during pre-employment vetting or when applying for a visa. Companies who wish to place tenders for the supply of goods are also requested to produce the certificate from the director of companies.

Criminal background records provide substantial information for an individual who is under investigation by the Criminal Investigation Department (CID).

Criminal history record is information collected by criminal justice agencies on individual consisting of identifiable descriptions and notations, detention and indictments or other formal criminal charges and any disposition arising therefrom, including acquittal, sentencing, correctional supervision or release [1]. In Kenya, the certificate is essentially a search into an individual's records held by the Directorate of Investigations Department (CID) and is commonly known as Police Clearance Certificate. The certificate confirms that the holder has or has not engaged in criminal, civil or administrative criminal offenses in the preceding 6 months and the verdict of the case, a record which is entered into criminal records database after the court proceedings. In current situation, the process that is involved in the processing and issuance of these certificates is centralized and therefore accompanied by risks of having a centralized system.

Emergence of new technologies has spearheaded adoption of better systems that are able to solve real world problems. This paper proposes the use of cloud database as a solution to the most challenges facing the criminal background check process focusing handling of distributed queries and node splitting and reporting

on criminal background information

Cloud database will provide substantial potential benefits to law enforcement and government agencies. Cost savings, offsite storage and disaster recovery, and dynamic provisioning of new and additional resources when needed are among the tangible benefits that cloud computing can offer in criminal record review.

Recent growth of data gathering by law enforcement officers through, for example, the use of CCTV cameras, only oblige to reemphasize the requirement for clearly articulated policies concerning cloud-based data storage. The volume of data has been increasing in the database. It is therefore foreseeable that some of that information will be stored and processed using cloud services.

To meet the dynamic operational needs, while maintaining the security of systems and data, the law enforcement agency can contemplate on the use of Database as a Service to administer criminal background check.

II. Existing system

The current system allows applicants of Police Clearance Certificate to open e-Citizen account through which they can apply for the certificate. This system also allows the applicants to pay processing fee online via credit and debit cards or through mobile payment platforms like Mpesa, Airtel money etc. The applicant downloads and print two copies of their invoice, and one copy of C24 on both sides On an A4 paper and then present the C24 and invoices with an original National ID for adults and Original Birth Certificate for minors, to the Directorate of Criminal Investigations for finger print processing. Foreign citizens are required to submit original copies of their passport and a documentary proof of being in the country for six months or more. The Directorate of criminal Investigation uses APFIS to process fingerprint scans. This is an integrated database that has two other databases i.e. Manual database that stores backup of fingerprints and Integrated/Automated database that store soft copy of all fingerprints.

This system stores soft copy of fingerprints taken from Police

station once a person has been arrested for committing any offense. When an applicant delivers the downloaded from the e-Citizen, fingerprints are taken on the hard copy and the application is send to APFIS. In APFIS, before search start, they compare fingerprints on the hard copy taken from the applicant with the fingerprints of the thump prints taken when the applicant applied for the National ID. Mostly the Right and the Left thump fingerprints are used for comparison

This process of comparing fingerprints is done manually using magnifying lenses. After the process, they confirm that the document has been certified before taken to the next phase. Here, they fill the bio data from the application form into the system After feeding the data they run APFIS scanner that scans and send the fingerprint to the database for comparison with the criminal database that stores all criminal records. This compares the two to see if they match for every individual. If the fingerprints match that's a hit. Meaning the applicant has criminal records in the database. This record is then printed on the certificate and then uploaded on the e-Citizen for downloading by the applicant. This takes an average of two weeks.

III. Proposed System

The use of database nodes will help improve service delivery in access of criminal information. The system will enable CID detectives and staff in different CID offices located countrywide to search for data in police databases concerning a particular individual under investigation or applicants of police certificate. Due to increased rate of certificate forgery, employees and other interested parties will be able to verify the authenticity and validity of a certificate by sending the ID number of the certificate holder and get a confirmation SMS on whether the certificate is authentic.

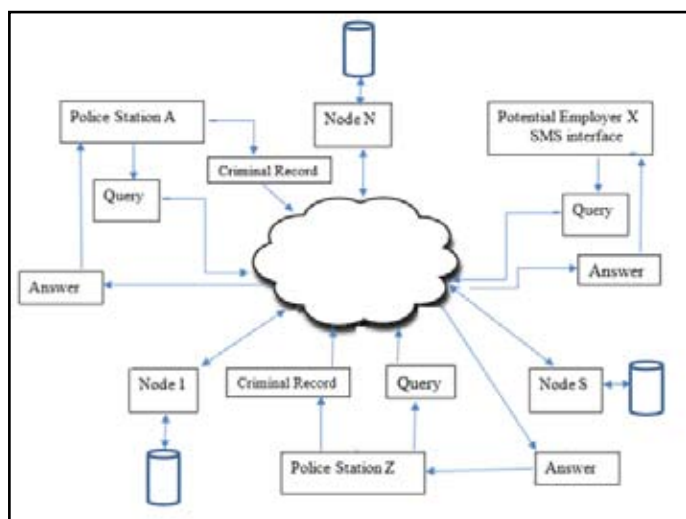


Fig. 1: Proposed System

A. Multiple Nodes

Due to increasing workload or multiple access of the cloud database, it becomes difficult for the cloud to service the increasing number of queries. This aspect becomes a complex mechanism for the cloud database to handle increasing number of queries, however multiple nodes in a data centre shares the workload hence ensuring timely response. When a criminal record is entered into the system, the central node replicates the record in all other nodes hence all nodes will be containing similar data. Multiple nodes serve an advantage in that failure of one node does not

affect other nodes thus affecting the retrieval of criminal record. During retrieval of the criminal record, the request is made to the central node. It therefore searches for any node that is idle and dispatches the task to that particular node which processes the Police Clearance Certificate. If the there is no record tied to the entered fingerprints and the National ID, this triggers for manual search. If no record is found, new record for that applicant is keyed into table that contains clean criminal records.

B. SMS Interface

For the verification of the Police Clearance Certificate, the employer sends prospective employees ID number to a certain dedicated telephone number via SMS.

The system compares the ID number of the applicant to see if there is any match to the records contained in the database (both the criminal records table and the clean records table). If a criminal record is found, the record is sent to the user via text. If the user is found in the clean records table, a message is sent to the user informing the user that he/she has no criminal record.

If there is no match in either of the tables, a message is sent to the user informing him/her that currently there are no records available in the database.

IV. Related Studies

A. Cloud computing Concept

Cloud Computing is a computation paradigm in which the resources of an IT system are offered as services, available to the users through net connections, frequently the Internet [2]. One of the paradigm of cloud computing is Database as a Service (DBaaS)

B. Database as a Service (DBaaS)

Cloud database is mostly used as a service although is less talked about compared to Software as a service (SaaS), Platform as a service (PaaS), Infrastructure as a service (IaaS). This technology is presented as a new transactional “database-as-a-service” (DBaaS) called Relational Cloud. DBaaS promises to migrate most of the operational problem of provisioning, configuration, scaling, performance tuning, backup, privacy, and access control from the database users to the service operator, offering lower overall costs to users [3] based on Service Level Agreement (SLA).

Many corporations and e-commerce companies have already adopted Database as a Service and they are currently reaping full benefits from this service.

Clients access the database on demand through the internet from the database service provider hence cloud database is designed for virtualized computer environment. Cloud computing enables implementation of cloud database where the service consumers utilize the software and hardware resources of the cloud computing service provider. The database should also be managed with ease and should cut down the cost as well. In Cloud computing information recovery is very efficient after a disaster in the database. In event of client's computer crash there is no data loss as everything is stored in the cloud. [4]. Backup is not a necessity as users can create a repository of information where is can be accessed and shared by all authorized parties.

C. Criminal Background Check using cloud database

With the changing trends in technology, emergence of cloud and its utilization law enforcement agencies can pursue smarter

policing having one huge potential benefit on the horizon. Cloud computing can minimize up-front investment and ongoing costs for IT systems and applications, makes sense in this era of fiscal austerity. With some law enforcement agencies utilizing cloud database to securely store criminal data, it is possible for criminal Investigation Department to catch up with the FBI and CIA as the cloud allows them to have access to the same IT infrastructure, data centre and applications. This will enable the department to efficiently serve all citizens from every corner of the country. Like other government agencies, police hope they can save some money and get rid of legacy hardware and software by using the cloud. With cloud database storage of criminal records on the cloud will help in disaster recovery and backup, crime analysis and records management. The cloud opens up sophisticated technology tools and services to smaller agencies that don't have the funds to purchase an entire application on their own. Cloud database will give a leg up to the CID department as it would follow that the small police department also could move quickly as far as using the latest digital productivity tools without big up-front costs. What is holding back many police departments whether large or small are security concern. The cloud providers therefore should meet stringent security standards for data sharing on the network.

D. Structure of Cloud Database

The cloud database houses the data on data centers distributed at various locations. This aspect draws a disparity between the cloud database structure and the normal database management system hence leading to a complex structure of the cloud database. Cloud database has multiple nodes that services queries for data centers that are dispersed in different geographical locations and the enterprise data centers as well. For comprehensive access of the cloud database services this joining is obligatory. [5] demonstrates the structure of cloud database through Business Intelligence (BI) application used for storing huge customers data by the enterprise. The customers usually access the cloud database via the internet which acts as bridge among the data centres, cloud databases and the client querying the database.

Cloud database is not composed of a single node but there are different nodes are used [3]. For this purpose, peer-to-peer communications are favoured. The purpose to implement peer-to-peer communication is that, a single node can service any kind of query generated by the user. The goal of having different node is to minimize the number of cross-node distributed transactions, which incur overhead both because of the extra work done on each node and because of the increase in the time spent holding locks at the back-ends.

When a user executes a query through their computer, the central node determines the most suitable node to handle the query and then forwards the query to that particular node. The assigned node services the query and returns an answer to the client. For example, when the query is received by central node, it carries out search to see which free node is holding the requested criminal record and then forwards the query to that node for servicing. The figure below shows the basic architecture of the cloud database [5].

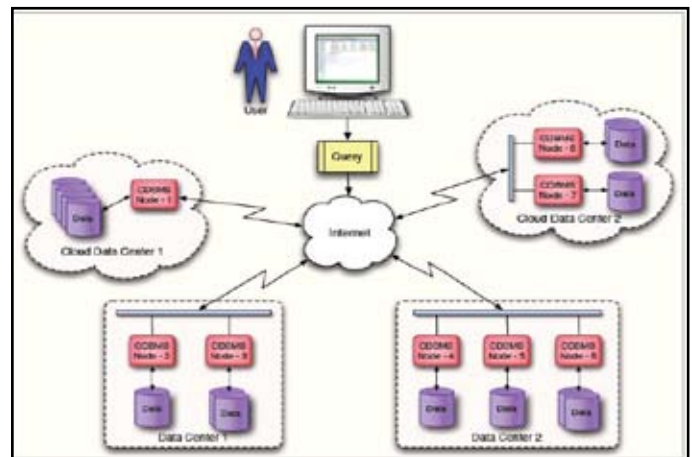


Fig. 2: Structure of Cloud database

V. Methodology

This study sought to establish the extend at which database nodes can be developed and run in a distributed environment for adoption in criminal background check and issuance of Police Clearance Certificate in every Criminal Investigation Office in Kenya.

A. Research design

This study used descriptive research approach. This approach is used when the researcher wants to provide answers to the questions of who, what, when, where, and how associated with a particular research problem [6]. We used Mixed methods research, a research design with philosophical assumptions as well as methods of inquiry that guides the direction of the collection and analysis of data and contains the mixture of qualitative and quantitative data in a single study or series of studies. Its central premise is that the use of quantitative and qualitative approaches in combination provides a better understanding of research problems that either approach alone. [7]. Questionnaires were used to collect quantitative data from prospective job seekers, applicants of Police Clearance Certificate, and selected employees while interviews were used to collect qualitative data from CID officers. After collecting the relevant data, the researcher developed a prototype to demonstrate how distributed database nodes can make criminal background check process effective

B. Population of the Study

This study categorized target population into two: Officers at the Directorate of criminal investigation, a unit of Kenya Police responsible for investigating complex cases, applicants of Police Clearance Certificate together with selected employers. Currently there are 200 officers who are involved in criminal investigations and issuance of these certificates at the CID headquarters. The employers were examined to answer whether they have been verifying the authenticity of Police Clearance Certificate presented to them by applicants. The total population of job seekers and employers is too large thus treated it as 'unknown' for the purpose of research and used 200 as a representative population for the unknown number of employers and job seekers [8] or applicants of Police Clearance Certificate

C. Population Sample

Qualitative research methods use a smaller sample size as compared to quantitative research methods because they are aimed at gathering an in-depth understanding of phenomenon which

focus is on how many and why a particular process or situation. Anywhere between 5-50 participants is viable sample size for qualitative research [9]. The main factor of consideration is not generalization by saturation. This implies that the sample size is just good enough just before the redundancy begins [10] [11] [12]. Therefore, this study focused on 5 respondents for interview [9] since from the fifth respondent the data collected had already reached *saturation*. This formed a pilot study to inform the research. To acquire the sample size of applicants and employers who formed part of respondents, this study will use several approaches and formulas.

$$n = \frac{Z^2 \times P(1-P)}{e^2}$$

where n=desired sample size z-score is the confidence level, for this study 95% thus z-score is 1.96 P is the expected true proportion which will be 0.5 for this study margin of error (e²) will be 5%. Thus, necessary sample size

$$n = \frac{1.96^2 \times 0.5(1-0.5)}{0.05^2}$$

=385 persons

The following formula was used to calculate sample size [14] where n_f= desired sample size when the population is less than 10,000, n = desired sample when the population is more than 10,000, n = estimate of the population size (this study will set it at 200).

Therefore

$$n_f = \frac{n}{1 + \frac{n}{N}}$$

$$n_f = \frac{385}{1 + \frac{385}{200}}$$

The sample size chosen (n_f)=132 persons

The researcher therefore chose to gather data from 107 applicants and 32 employers.

VI. Design and Implementation

1. Extreme Programming Development Methodology

This is a software development technique that is designed to improve the quality of software that is produced and also improve its ability to properly adapt to the changing requirements of the customer or client. Extreme programming consists of the following steps [13]: - Product Life Cycles, Releases, Iterations, Tasks, Development, Feedback

The diagram below summarizes the process of extreme programming.

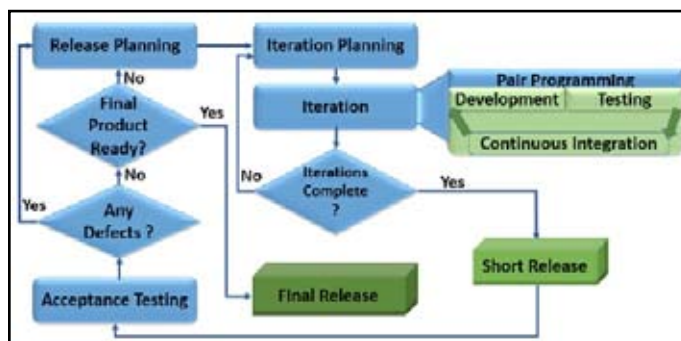


Fig. 3: Extreme Programming Process Cycle

2. Architectural System Design

The prototype will run under cloud architecture- Database as a Service (DBaaS). The system is web based and the client can access the application via laptops, desktops and mobile device to make SMS requests to the server side which used PHP Script and MYSQL database

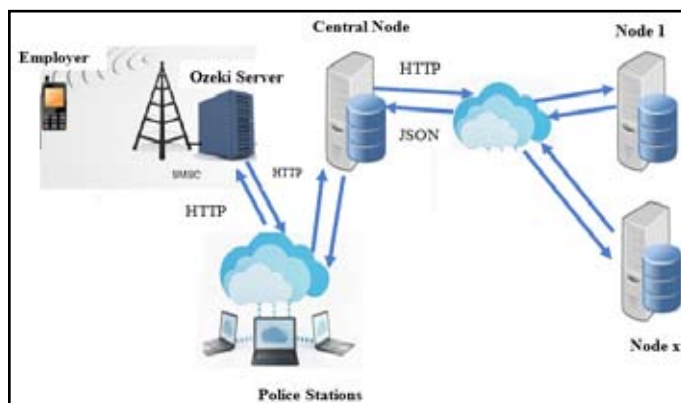


Fig. 4: Prototype architecture diagram

3. Use Case Model

This is a diagram that is used to display how different users of a system interact with it to solve a task. It is used to define the objectives of the users while using the system. It also shows the interactions between users and the system as well as how the system behaves to satisfy the set goals.

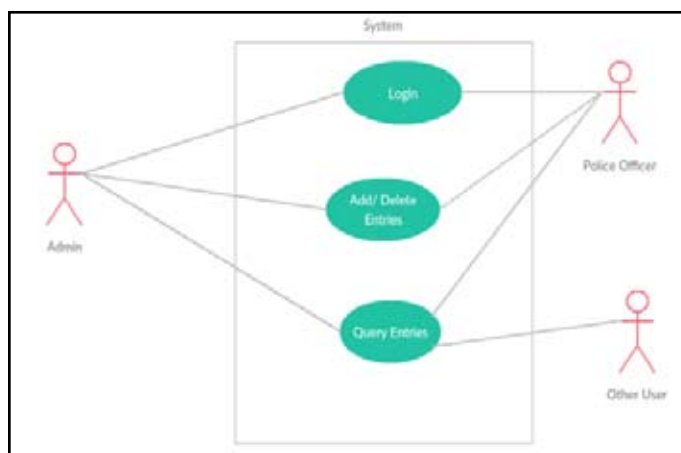


Fig. 4: System Use Case Model

4. Context Diagram

This is a diagram that is used to display the high-level view of a system in that it defines the boundaries between the system or a segment of the system with its environment. It also shows the

different entities that interact with the system.

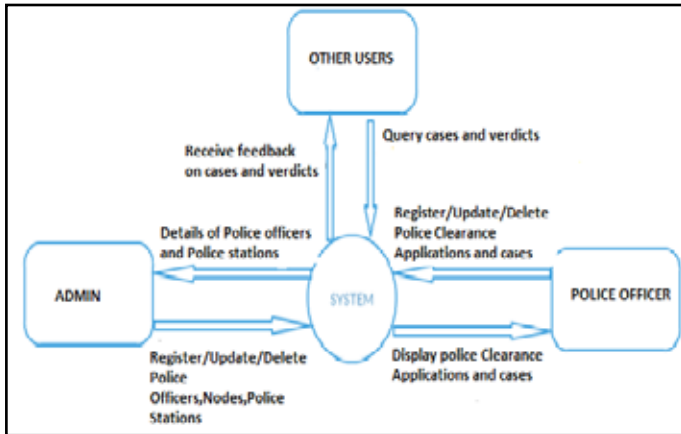


Fig.5: context diagram for this system

5. System flowchart

Most of the activities will be conducted by the Police officers. The application will allow them to perform the following functions:

- I. Input a new criminal record.
- II. Input a new record of Police Clearance Certificate applicant
- III. Input a new record of a person with clean record in case manual search has generated clean record.
- IV. View all the above data
- V. Search for a record
- VI. Generate a new certificate of good conduct

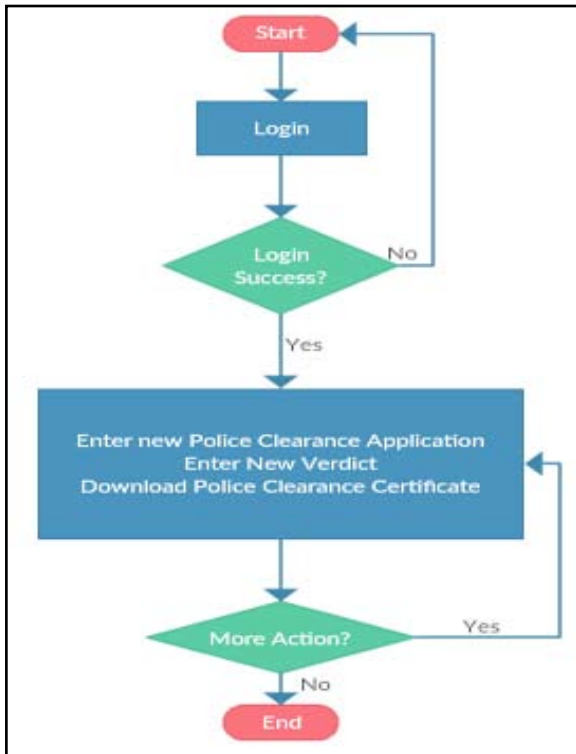


Fig. 6: Analytical algorithm flowchart

Code for Central node

```

<?php
//start the session
session_start();
ob_Start();
//include the database
require'db.inc.php';
    
```

```

//get the variables and urlencode them
$name=urlencode($_POST['fname']);
$mname=urlencode($_POST['mname']);
$lname=urlencode($_POST['lname']);
$email=urlencode($_POST['email']);
$telephone=urlencode($_POST['telephone']);
$idno=urlencode($_POST['idno']);
$pic_data=base64_encode(file_get_contents($_FILES['image']['tmp_name']));
//$pic_data = addslashes(file_get_contents($_FILES['image']['tmp_name']));
$pic_size=$_FILES['image']['size'];
$pic_data=urlencode($pic_data);
$pic_size=urlencode($pic_size);
    
```

```

//get all the nodes
$thenodes="SELECT * FROM nodes";
$thenodes2=mysql_query($thenodes);
//for each node, send the data
while ($thenodes3=mysql_fetch_array($thenodes2))
{
    $base_url=$thenodes3['url'];
    $page="newapplication.php?";
    $other="fname=$name&mname=$mname&lname=$lname&email=$email&telephone=$telephone&idno=$idno&pic_data=$pic_data&pic_size=$pic_size";
    $url="$base_url$page$other";
    echo $url;
    $response = file_get_contents($url);
    echo $response;
    //if the record is found in any node
    if ($response=="found")
    {
        $_SESSION[response]=$response;
    }
}
    
```

VII. Results

The main objective of this study was to develop a prototype that will provide solution to effective Criminal background check in the background screening industry by utilization of cloud database. We managed to analyze, design, develop and test the prototype. The system had three modules, the administrator module, the Police officer module and the other user's module. The administrator was able to login, and to view, register, update and delete nodes, Police officers, counties and Police stations.

The Police officer was able to register new clearance certificate applications, enter new cases and verdicts, query applications and cases and to generate Police Clearance Certificates in PDF format.

The other users were able to verify authenticity of Police Clearance Certificate by viewing cases and verdicts via an SMS based application that was able to query from a database.

1. Interface Design

These are mock up images that represent the flow of the different screens or interfaces through which the users can interact with the system. It exposes the functional components of the developed system.

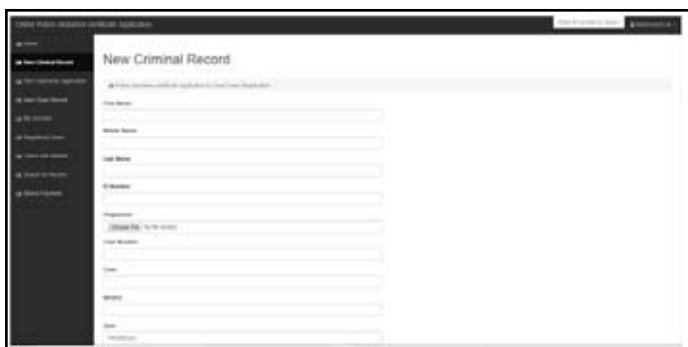


Fig.7: System Interface Design



Fig.9: Overall rating

2. System Evaluation

System evaluation was conducted to test the basic functionalities of the system and how user requirements of the system and user requirements were directly achieved. The main objective of this test was to examine the success of the prototype usage in the actual environment and identify issues that would arise for further refinement. The key parameters used on this test were: - User Interface, Response time, Nodes performance and System navigation. A sample of 20 respondents was used to evaluate the prototype as well as rating its performance. Among the respondents were 10 Police officers and ten professional developers. The table below shows the responses as captured from respondents.

Table 1: System Evaluation

	Excellent	Good	Poor	Very Poor
User Interface	13	6	1	0
Response Time	10	8	1	0
Nodes Performance	15	5	0	0
System Navigation	12	7	1	0
Overall rating %	62.5	32.5	3.75	0

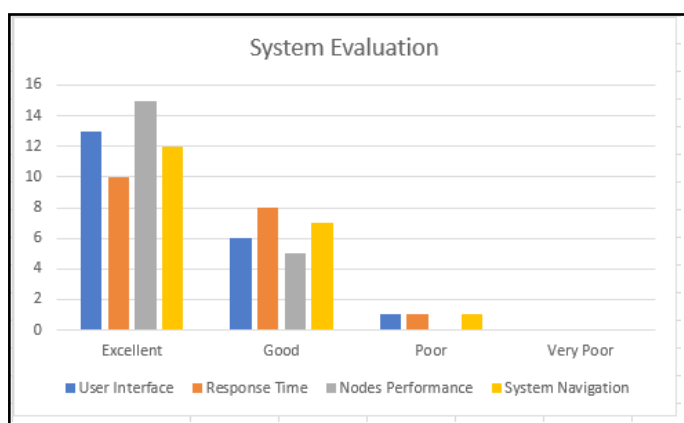


Fig 8: The summary System Evaluation

The greatest number of system users, represented by 62.5% supposed that the performance of the system was excellent, 32.5% believed that the system was good to some extent while 3.75% thought that the system was poor. The overall sample rating on the evaluation of the system showed that the system would be of great help when operationalized in processing and issuing of Police Clearance Certificate.

VIII. Conclusions

The findings from this study demonstrated that the current systems and processes involved in criminal background check have been taking too long to process and generate a Police Clearance Certificate. This study further demonstrated how integration and utilization of Cloud database can cure problem associated with a centralized system of criminal background check hence improving service delivery in processing and issuing of Police Clearance Certificate from every CID office countrywide. Every criminal record keyed into the system is received by the central node which in turn replicates the same in distributed database nodes. This will cut down on the cost involved even by the applicants who need not to travel long distances to the CID headquarters from every corner of the country for finger print processing as the developed system was able to accept fingerprints and carry out comparison with the criminal records already stored in the database. During the test, the system users retired several nodes to in order to see the performance of the system. This did not affect in anyway the processing and generation of Police Clearance Certificate. This was made possible through implementation of DBaaS where distributed nodes can share the workload and carry out criminal background check within a short period of time then generate Police Clearance Certificate. With the technology in place, applicants will be acquiring their Police Clearance Certificates at one stop without having to wait for one or two weeks.

Acknowledgement

I am greatly indebted to various people who have contributed immensely to the success of this project. First and foremost, I thank my supervisor, Andrew Kahonge, who has been on the forefront in stimulating new ideas and providing guidance in this research project. Secondly, I give special gratitude to my family for the moral support they have accorded me.

References

[1] Jacobs, J. and Crepet, T., 2007. *Expanding Scope, Use, and Availability of Criminal Records*, The. NYUJ Legis. & Pub. Pol'y, 11, p.177.
 [2] Buyya, R., Broberg, J., Goscinsky, A.: *Cloud Computing: Principles and Paradigms*. John Wiley and Sons (2011)

- [3] Curino, C., Jones, E., Popa, R., Malviya, N., Wu, E., Madden, S., Balakrishnan, H., And Zeldovich, N. 2011. *Relational Cloud: A database service for the cloud*. In *Proceedings of the 5th Biennial Conference on Innovative Data Systems Research*. 235–240
- [4] Pocatilu, P., Alecu, F., Vetrici, M.: *Measuring the efficiency of cloud computing for e-learning systems*. *W. Trans. on Comp.* 9, 42–51 (2010)
- [5] Al Shehri, W., 2013. *Cloud Database as a Service*. *International Journal of Database Management Systems*, 5(2), p.1.
- [6] Bickman, L., Rog, D.J. and Hedrick, T.E., 1998. *Applied research design: A practical approach*. *Handbook of applied social research methods*, 19, p.19.
- [7] Creswell, J.W. and Clark, V.L.P., 2007. *Designing and conducting mixed methods research*.
- [8] Muthoni, J., 2015. *E-verification—a case of academic testimonials*.
- [9] Dworkin, S.L., 2012. *Sample size policy for qualitative studies using in-depth interviews*.
- [10] Dr. Bonnie Nastasi, 2005. *Qualitative research: Sampling and sample size considerations*. *Director of School Psychology Program*.
- [11] Marshall, M. N. 2012. *Sampling for qualitative research*. *Family practice*, 13(6), 522-526.
- [12] Crouch, M., & McKenzie, H. 2006. *The logic of small samples in interview-based qualitative research*. *Social Science Information*, 45(4), 18. doi: 10.1177/0539018406069584.
- [13] Beck, K., 2000. *Extreme programming explained: embrace change*. *addison-wesley professional*.
- [14] Mugenda, O.M., 1999. *Research methods: Quantitative and qualitative approaches*. *African Centre for Technology Studies*.

Author's Profile



Mr. Michael Ndonye Mutisya, is currently pursuing Master of Science in Distributed Computing Technology, University of Nairobi, Kenya. Currently, he is a Lecturer at the School of Computing and Informatics of Mount Kenya University and Riara University, Kenya. His research interest is in the area of Information Security and database systems. E-mail: mutisyamike07@gmail.com