International Journal of Advanced Technology in Engineering and Science
Vol. No.3, Special Issue No. 01, September 2015
www.ijates.com
ijates
ISSN 2348 - 7550

IMPLEMENTING HIPPOCRATIC DATABASE IN LEGACY APPLICATIONS

Sonali Ganguly[1], Dr. S. P. Singh[2]

[1,2]*Department of Computer Science, BIT Mesra, Noida Campus, (India)*

ABSTRACT

In many organizations, legacy applications are bread earners. Even though the technology, language, technique or platform may be obsolete, such legacy applications either earn money for the business or helps in smooth functioning of the business. Legacy applications handle sensitive information thereby making security of such crucial legacy applications equally critical for an organization.

The concept of Hippocratic Databases evolved in 2002 which focuses on the responsibility of database to keep the data that rests in the database private through ten defined principles. Hippocratic Database (HDB) has been implemented in various domain like e-Learning, virtual community, web services etc. where privacy of information is essential. The paper has proposed a legacy modernization strategy called Divide and Rule Approach and also presents the method to implement principles of Hippocratic Databases on legacy applications and highlights the benefits and challenges of implementing HDB in legacy applications.

Keyword: Data Privacy, Hippocratic Database, Legacy Modernization, Legacy System

I. INTRODUCTION

The paper is divided into 5 sections. Section 1 provides a general introduction about Hippocratic Databases (HDB) and its principles and privacy of legacy application. Section 2 describes the legacy modernization strategies where a new strategy has been proposed called 'Divide and Rule Approach'. Section 3 highlights the benefits of the HDB and its implementation in legacy applications. Section 4 identifies the challenges in implementing the HDB in Legacy Applications and Section 5 concludes the paper. This paper is an extension of the paper mentioned in references [1] and [2].

1.1 Introduction to Hippocratic Database

The concept of Hippocratic Databases evolved from the Hippocratic Oath of medical or law profession. A part of the oath is given below.

'What I may see or hear in the course of the treatment or even outside of the treatment in regard to the life of men, which on no account one must spread abroad, I will keep to myself, holding such things to be unutterable.'[3]

Based on the Hippocratic Oath, Hippocratic Databases takes the responsibility for the privacy of data that rests in the database. The ten principles of HDB that supports the Hippocratic Oath as lay down by Agrawal et al. 2002 [3][1] are:

1. Purpose Specification: For every information stored in the database, the purpose for which the information was collected should be associated with the information. This will also ensure that the relevance of the table is known and any modification or updates are implemented keeping the purpose in mind.
2. Consent: The purpose associated with personal information shall have consent of the donor of the personal information. Consent is needed to determine what information should be disclosed to which group.
3. Limited Collection: The information collected should be limited to minimum necessary for accomplishing the specified purposes. There should be difference between collection and use of data.
4. Limited Use: The database should run only those queries that are consistent with the purposes for which the information has been collected.
5. Limited Disclosure: The personal information stored in the database should not be communicated outside the database for purposes other than those for which there is consent from the donor of the information.
6. Limited Retention: Personal information should be retained only as long as necessary for the fulfilment of the purposes for which it has been collected. The period of retention should be reasonable. It is possible that the information is retained for an extended period of time due to statistical computation. In such cases, information for personal identification can be removed.
7. Accuracy: The personal information stored in the database should be accurate and up-to-date. Accuracy is needed to eliminate data inconsistency.
8. Safety: Personal information should be protected by security safeguards against theft and other misappropriations.
9. Openness: A donor should be able to access all information about the donor stored in the database.
10. Compliance: A donor should be able to verify compliance with the above principles. Similarly, the database should be able to address a challenge concerning compliance

    Selected principles of HDB have been implemented in various domains based on the requirements and environment of the domain area.

1.2 Introduction on Privacy in Legacy Applications

Many organizations have legacy applications that are running the core business of the organization. Legacy Applications (LAs) were built when security was not a major design issue and the number of threats and risks were comparatively fewer. Assuming LAs are free from risk in the present environment is a myth since LAs are growing old with each passing day while the world around them is evolving. Thus, these applications are prone to new security risks and privacy threats making them vulnerable and open [4].

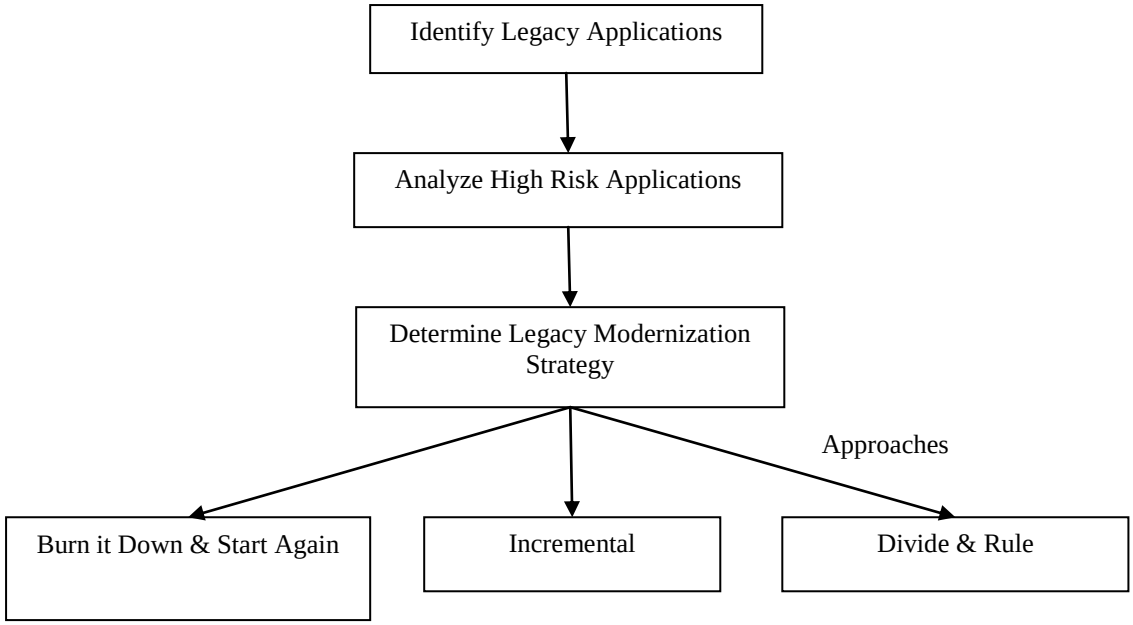

Figure 1: Steps to Ensure Data Privacy in Organization

As presented in Fig. 1, following are the steps to ensure data privacy in organizations:

1. Identify Legacy Applications

   An organization has many systems and applications which either performs business functions or internal processing to run business. Every such system or application cannot be categorized as a legacy system or application. An old application/system running over a decade, applications hidden behind other applications and forgotten or ignored applications performing background tasks can be regarded as legacy applications [4]. It becomes a difficult task to identify such legacy applications for IT based organizations. Many times organizations are not fully aware of all the legacy applications due to change in manpower, management structure etc.

2. Analyze High Risk Applications

   Once legacy applications have been identified, applications with high risks needs attention. Major factors to determine high risk applications are:

a. Data Sensitivity: Every application handles different type of data. Sensitivity of data depends upon the organizational objective.

b. Critical Functionality: Functions that are critical in running the business determines the importance of the legacy application.

   Applications that run critical functions of the business and handles sensitive data requires data protection strategies. Based on the above analysis, selection of the legacy application is important which would incur minimum cost and maximum benefit.

3. Determine the Legacy Modernization Strategy

   Legacy Modernization is the transformation of a legacy system/application to a new platform by rewriting code, porting to new environment, upgrading to new protocols, updating libraries or migration to new technology. Legacy Modernization is usually is large, complex and time consuming task. Involvement of

senior management, practitioners, clients and support staff is required before making the decision of converting the legacy application to new platform.

Decision to implement legacy modernization depends upon the following factors:

a. Size of the Application: Task of transforming a legacy application partly or wholly into a new platform is easier for a simple and small scale application. Since size of the application determines its scope, the same task becomes difficult if the application is large and complex.

b. Investment: Capital and Manpower investment is a cost that the organization has to bear with during transformation phase. Since size of the application plays a major role, the size of the organization to invest in this transformation is equally considerable. Resources like manpower, hardware/software, time, space etc may be involved depending upon the scale of legacy modernization.

c. Change Management: A change management structure should govern the process of legacy modernization. Without an effective change management structure in place, tracking and monitoring of the application will become cumbersome.

d. Maintenance: An organization has to incur maintenance cost of the upgraded application. Though, maintenance costs are generally huge, the amount of maintenance required defines the cost.

II. LEGACY MODERNIZATION STRATEGIES

There are various strategies to implement legacy modernization.

a) Burn it Down and Start Again Approach[5]

The simple 'burn it down and start again' approach is one of the direct method to eliminate the legacy application completely and replacing it by building a new system/application with higher configurations and technology that performs the tasks of LAs with improved functionalities. But this approach is not a feasible option for all where LAs perform daily critical operations. It also faces operational risks wherein there is a likelihood of unknown errors or bugs after deployment. This approach may become a multiyear project where the organization has to bear resources like time, manpower, hardware/software, space and money but it may be effective where the legacy applications are smaller in size and perform simple tasks consuming minimum resources.

b) Incremental Approach

Incremental approach is a safer method of upgrading or improving the existing legacy application by focussing on the application part by part to improve the legacy application. This approach is suitable for large sized applications/systems wherein with each increment, the legacy code decreases. Eventually the application is completely modernized.

c) Divide and Rule Approach

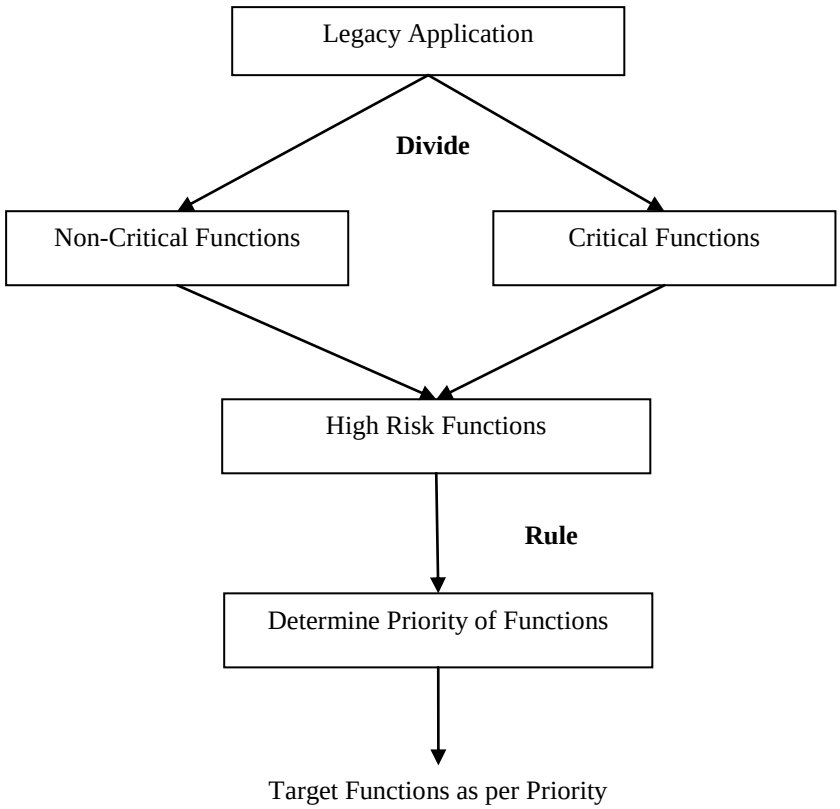

Figure 2: Divide and Rule approach

In the divide and rule approach, once the legacy application has been identified, divide all the functions of the application into two categories namely critical functions and non critical function. Among the critical and non critical functions, identify the functions with high security risks. There is a probability that non critical functions are more vulnerable to threats which may compromise the security of the whole legacy application. If critical functions too have security risks then define priority to all such functions and cater them one by one. If feasible integrate functions that can be handled together to implement legacy modernization.

This paper has proposed the above approach 'Divide & Rule' to improve the legacy application.

III. HIPPOCRATIC DATABASE IN LEGACY APPLICATIONS

With the emergence of HDB principles in 2002 to shield the privacy of data that rests in database, the principles of HDB have been realized in various different platforms, data sets and domains like e-Learning [6] [7], web services [8], virtual community [9], social networks [10] [11] [12], log file architecture [13], relational databases [14] [15] [16] etc. Data being a very significant asset for any organization, these principles can lock sensitive information at database level.

3.1 Benefits of Hippocratic Database

Application of Hippocratic Database comes with the following benefits:

1. Simplicity: Simplicity in the concept of Hippocratic Database with respect to the ten defined principles focussed on privacy of data that rests in the database. These ten principles define the scope for the database modification and maintenance of legacy application.
2. Privacy: The focus of HDB is privacy of data that is stored in the database. Since data is a critical asset to an organization, its privacy is also essential. The best part is that the organization is able to minimize security threats to data at database level. Application of the principles ensures that only relevant information is passed to authorized user.
3. Flexibility: Flexibility in the implementation of the defined principles. Based on the organizational culture, technology and platform, complexity and nature of the system, selected principles of HDB can be implemented. The level of privacy to be maintained in the database can be categorized based on the selected principles and other affecting parameters. It is not necessary to implement all the ten principles of HDB but maximum principles should be implemented for maximum benefit.
4. Availability of Resources: The concept of Hippocratic databases emerged a decade ago. Multiple articles and research papers are available on the Internet for reference that explains the implementation of the concept on various platforms and environments for a new project. Examples of HDB implementation are available online for understanding of its model on various domains.

3.2 Application of HDB principles on Legacy Applications

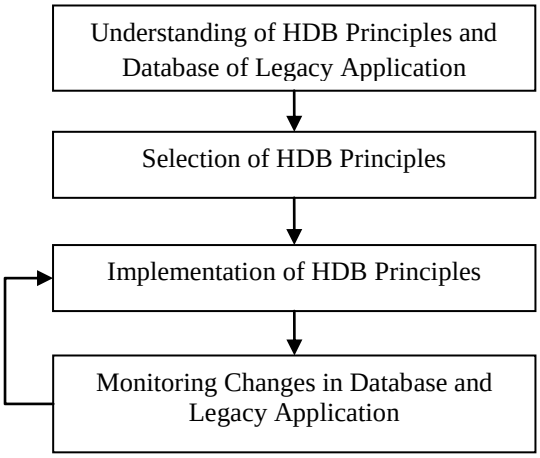

Figure 3 Steps to Implement HDB Principles in Legacy Application

To implement HDB principles in a legacy application, steps have been proposed as described in Fig. 3.

1. Understanding HDB Principles and Database of Legacy Application

   Awareness of the relation database structure of the legacy application with the knowledge of schemas and table structure is required. Study the database structure and analyze how the system interacts with the database to ensure appropriate measures are implemented to improve its security. Also thorough understanding of the HDB principles[] is necessary so that it is easier to perform the following steps.
2. Selection of principles

Once there is understanding of the application and the HDB principles, it is easier to correlate which principle is suitable for implementation in the legacy application. A description of the following principles can help in implementing them in legacy applications.

a. 'Purpose' is the first principle of Hippocratic Database and the most significant principle. This principle ensures that the purpose of creating a table or keeping a table in the database in case of legacy applications be maintained and stored in the database. A 'purpose' table may be created in the database that lists all the tables and their attributes with the purpose of creating them.

Table 1 Schema of Purpose Table

ID	TABLENAME	ATTRIBUTE	PURPOSE

Key features of purpose table are [2]:

i. The purpose table stores the purpose or reason of creating each table in the database. This ensures that the core cause of creating and maintaining a table in the database is captured. This feature would also aid in understanding the legacy application and the reason for the existence of a particular table in the database since databases generally contain several tables.

ii. If purpose is associated with each table then obsolete or unused tables can be revealed. Such tables can be discarded after scrutiny.

iii. This relation will capture the various reasons for accessing a particular attribute of its corresponding table. Thus access of an attribute shall communicate the purpose of the access.

iv. Purpose table can also highlight the unused attributes of a relation. Information like obsolete or useless attributes can be extracted from purpose table. It may also be revealed that new attributes have been introduced in the system for improvement .

In the purpose table as shown in Table 1, the rows with no value in 'attribute' column shall define the core cause of creating the table while the rows with attribute value shall capture the reason for accessing the particular attribute of a relation. This way purpose definition can be distinguished for a table creation and access of an attribute of a relation.

b. The 'Limited Collection' principle ensures that minimum necessary information is captured and stored in the database. Only those attributes that are essential and needed for some purpose are maintained [2]. If the purpose principle is implemented meticulously, then it will help implement limited collection principle. For example, in an organization, personal information like marital status of an employee is not required to be maintained during employment and corresponding change in name need not be updated.

Since data is a critical to an organization, its maintenance and consistency is also crucial. Hence, if minimum and relevant information is only stored in the database then cost of maintenance and data integrity also reduces.

c. Accuracy ensures regular up-to-date information in a database. To ensure accuracy of data during data insert, data modify and data delete, a process must be in place. For example, to enrol for Aadhaar Card, a person has to visit an authorized Aadhaar enrolment centre with multiple and defined documentary evidences and photographs. At the centre, fingerprints and iris scan will be taken and the person has to fill

out the form. Clearly, there is a strict process followed before the individual information is stored in the database. There is a similar process for change in personal details in the Aadhaar Card.

Identify the process applicable in the legacy application and improve the process if required. If a process is not in place then establish a simple and consistent process to maintain accuracy of data in database.

d. Retention principle focuses on retaining information in the database till its purpose of collection is fulfilled. Legacy Applications accumulate huge amount of information over the years. Organizations follow their own retention policy to manage the data. Review the retention policy for the legacy application and streamline it accordingly. Keeping sensitive information in the database may expose past data of the organization during threats. Past information can be archived in another medium or moved to a secure location.

3. Implementation of the HDB Principles

   Based on the requirements of the legacy application and culture of the organization, the selected principles can be applied to the database one by one. Monitor each implementation separately to ensure that it does not affect the working of the system. Make small changes so that reverting back to stable stage is easy in case the application becomes dysfunctional.

4. Monitor Changes in Database and Legacy Application

   Monitoring and implementing a change is an iterative process. It is necessary to track changes made in the database and the application to ensure that the changes do not introduce uncertainty in the legacy application.

VI. CHALLENGES IN IMPLEMENTING HIPPOCRATIC DATABASE IN LEGACY APPLICATIONS

Every system has some pros and cons. Before following any system, analysis of pros and cons is necessary to ensure whether the organization can live with the limitations of the system or not. A system that works well in one environment may or may not be suitable in another.

1. Due to lack of documentation available for LAs, only a handful of people in the organization understand the complete system. LAs either don't have any documentation or limited records.
2. It is easier to implement HDB when the system starts from scratch. As per the requirements of the new application/system, the HDB principles can be applied accordingly. On the other hand, it is difficult to implement HDB principles in legacy system due to limited understanding of the system. To implement HDB principles in legacy applications, knowledge of schemas and table structure of LA is mandatory so that HDB principles can be selected accordingly.
3. HDB principles are applicable to RDBMS.
4. A change in the system faces resistance from employees. It is necessary to implement a change management system simultaneously and train employees to minimize conflicts at implementation stage.

VII. CONCLUSION

Relevance of data privacy in today's environment is expanding with the increasing threats and risks attached to private information of an individual or organization. It is the ability of the organization to ensure data protection.

Privacy of data can be accomplished to some extent through the implementation of ten principles laid down in Hippocratic Database. Steps to implement HDB principles in legacy applications have been proposed in the paper but these steps depend upon various decisional factors also mentioned in the paper.

Various legacy modernization strategies have been discussed in the paper where Divide and Rule Approach as a legacy modernization strategy has been proposed to handle the legacy application during improvements, enhancements or updates. It is true that diverse resources are involved in legacy modernization but there can be no compromise with security.

REFERENCES

[1] Sonali Ganguly, S. P. Singh, "A Literature Review on Database Privacy in Social Networks using Hippocratic Database", 5th International Conference on ACSEICT 2014

[2] Sonali Ganguly, S. P. Singh, "Proposed Methodology for Database Privacy in an Existing System Using Hippocratic Database", 7th International Conference on ACSEICT 2015

[3] Rakesh Agrawal, Jerry Kiernan, Ramakrishnan Srikant and Yirong Xu, "Hippocratic Databases", Proceedings of 28th VLDB conference, Hong Kong, China, 2002

[4] https://buildsecurityin.us-cert.gov/articles/best-practices/legacy-systems/assessing-security-risk-in-legacy-systems

[5] https://iapp.org/news/a/concerns-about-opm-notification-plan-legacy-systems-continue/

[6] Vanja Bevanda, Jasmin Azemović and Denis Mušić, "Privacy preserving in eLearning environment (Case of modeling Hippocratic database structure)", Fourth Balkan Conference in Informatics, 2009

[7] Jasmin Azemović, "Privacy Aware eLearning Environments Based on Hippocratic Database Principles", BCI, Nova Sad, Serbia, 2012

[8] Norjihan Abdul Ghani, Zailani Mohamed Sidek, "Privacy-Preserving in Web Services using Hippocratic Database" IEEE, 2008

[9] G Skinner, E Chang, M McMahon, J Aisbett and M Miller, "Shield Privacy Hippocratic Security Method for Virtual Community", The 30th Annual Conference of the IEEE Industrial Electronics Society, Busan,Korea, 2004

[10] Mohammad Reza Khayyambashi, Fatemeh Salehi Rizi, "An approach for detecting profile cloning in online social networks", 7th International Conference, Kish Island, Iran, IEEE, 2013

[11] Rajneesh Kaur Bedi, V.M. Wadhai and Nitinkumar Rajendra Gove, "Application of Hippocratic Principles for Privacy Preservation in Social Networks", World Congress on Information and Communication Technologies, IEEE, 2012

[12] Maryam Majedi, Kambiz Ghazinour, Amir H. Chinaei and Ken Barker, "SQL Privacy Model for Social Networks", Advances in Social Network Analysis and Mining, IEEE 2009

[13] Andrew Rutherford, Reinhardt Botha and Martin Olivier, "Towards a Hippocratic Log File Architecture", Proceedings of SAICSIT, 2004

[14] Norjihan Abdul Ghani and Zailani Mohamed Sidek, "Owner-Controlled Towards Personal Information Stored in Hippocratic Database", International Conference on Computer Technology and Development, 2009

[15] Oberholzer Hendrik JG, Ojo Sunday O and Olugbara Oludayo O, "A PET evaluation framework for relational databases" SocialCom, IEEE 2013

[16] Rakesh Agrawal, Paul Bird, Tyrone Grandison, Jerry Kiernan, Scott Logan, Walid Rjaibi, "Extending Relational Database Systems to Automatically Enforce Privacy Policies", Proceedings of the 21st International Conference on Data Engineering (ICDE 2005) IEEE