

Secure Data Entry in Operational Systems for Data Warehouse

Er. Muheet Ahmed Butt

Abstract

A 'data warehouse' is a repository of an organization's electronically stored data. Emergence of Huge Data Warehouses has also put a big challenge in these Data Entry Processing Systems. Data Warehouses contain a wide variety of data that present a coherent picture of business conditions at a single point in time [1]. Development of data warehouse includes development of systems to extract data from operational systems plus installation of a warehouse database system that provides managers flexible access to data. Data entry is processing of data using application programs that are created with the objective of the output. It is a process of getting information into a database, usually done by people working from home by typing it in by way of data entry which is pre designed to simplify the process. In this paper we have proposed a security measure for a distributed for the manual data entry module pertaining to Examination's Operational System of University of Kashmir Data Warehouse. The proposed System tries to identify invalid entries done by faulty data entry operators who have gained privilege over the system and generates a report for further necessary action.

Introduction

One of the most important assets of any organization is its information. This asset is almost always kept by and organization in two forms: the operational systems of record and data warehouse. Crudely speaking the operational systems are where the data is put in, and the data warehouse is where we get data out. As our focus is on operational systems of records that capture the transaction of business[2][3]. The source systems should be thought of as outside the data warehouse because presumably we have little to no control over the content and format of data in these operational legacy systems. The main priorities of the source systems are processing performance and availability. Queries against source systems are narrow, one record at a time queries that are part of normal transaction flow and severely restricted in their demands on the operational system. We make the assumption that source systems are not queried in the broad and unexpected ways that data warehouse typically are queried in the broad and unexpected ways that data warehouse typically are queried. The school systems maintain little historical data, and if you have a good data warehouse, the source systems can be relieved of much of the responsibility for representing the past. Each source system is often a natural stovepipe application, where little investment has been made to sharing common data such as product, customer [4], geography or calendar with other operational systems in the organization. It would be great if your source systems were being reengineered with a consistent view. Such enterprise application integration (EAI) effort will make the data warehouse design task far easier.

Another very important challenge of Information Security within an Operational System of a data warehouse have gone two major changes in the last one decade. Before the wide spread use of data processing equipment, the security of information felt to be valuable to an organization was primarily provided by administrative means. An example of the former is the use of rugged filing cabinets with a combination lock for storing sensitive documents. An example of the latter is personnel screening procedures used during hiring process. The second major challenge that affected security is the introduction of distributed systems and the use of networks and communication facilities for carrying data between terminal user-computer and computer-computer. In this paper we have tried to use the above concepts and procedures for strengthening the security mechanisms pertaining to data entry module for University of Kashmir Data Warehouse.

The rest of the paper is organized as proceeding sections describe the existing system and its working discuss table structures for various data groups pertaining to enrolment and marks, discusses the disadvantages of the existing system in terms of security concerns, and the Trailing sections discuss various measures to overcome these disadvantages and provide algorithms for the implementation of the same. The references used in this paper are listed at the end.

Existing System

The Examination System forms one of the most important units of the University. It forms the linkage between the general public and the University. The Examination System requires delivering its services efficiently for developing a better environment of teaching and learning. Our Examination System is completely computerized and latest trends of Information Technology are used to accomplish various tasks pertaining to Examination Automation. The major two operations carried by our examination system are conduction of examination and declaration of result. Our University offers around 300 courses and declare around 600 results each year having student strength of around 2 lakhs which is increasing every year.

Handling such a massive system involves a huge data entry support which enables us to move ahead with our academic calendar. There are around 40 data entry operators working round the clock to make it a success. Therefore security threat to such data entry systems increases and is more vital to any sort of attack.

Database Design

The database structure is given below in the table.

Considering the Enrollment Dimension

Rollno	Regno	Name	Fname	Sub1	Sub2	Sub3	Sub4.....
101	2345-ANG-2008	Masood Ahmed	Mohd Altaf Malik	GE	ED	PS	UR
102	3456-ANG-2008	Manzoor Ahmed	Saleem Ahmad Khan	GE	GG	PS	ED

The table above shows rollno, registration number, name, parentage and subjects offered like (GE for General English, ED for Education, UR for Urdu). Each student is allowed to take a combination of four subjects each year called as group where General English is compulsory and rest three pertain either to science, arts and commerce stream. Each subject can have a sub category of A, B, or C paper with a Practical/Internal Marks. The roll no/code no and marks entry is done using the following database schema. The assignment of jobs for various data entry operators is random. Any Operator can perform the data entry for any of the subjects on first come first enter basis. The awards pertaining to the marks of the candidates who have appeared for a particular examination are distributed evenly among all the data entry operators.

Consider

Marks entry in General English A (GE A)

File Name: MFGEA.dbf having following structure.

Rollno	Code	Marks
101	231	23
102	435	45

Marks entry in General English A (GE B)

File Name: MFGEA.dbf having following structure.

Rollno	Code	Marks
101	221	22
102	455	21

Hazards/Drawbacks

The above existing system leads to the following major following drawbacks/hazards.

1. Data entry operator can easily enter into the login of other data entry operator with stolen password and will try to mingle with the data.
2. Sometimes password is lost by the concerned Data Entry Operator.
3. This all leads to improper management of a Data Entry System.
4. It is very difficult to ascertain the Authenticity of a Data Entry Operator.
5. Passwords only known to data entry operators.
6. No time Management solution.
7. Poor track of data entry Operator/Operation.
8. No efficient reporting mechanism on operator-to-Entry basis.

Proposed System Architecture

The proposed three level Architecture comprises of a Data Entry Operator Level, Password Manager Level and Super Admin Level. These levels have been designed to provide efficient security mechanism for a huge and complex manual data entry for Operational Systems pertaining to a Data warehouse.

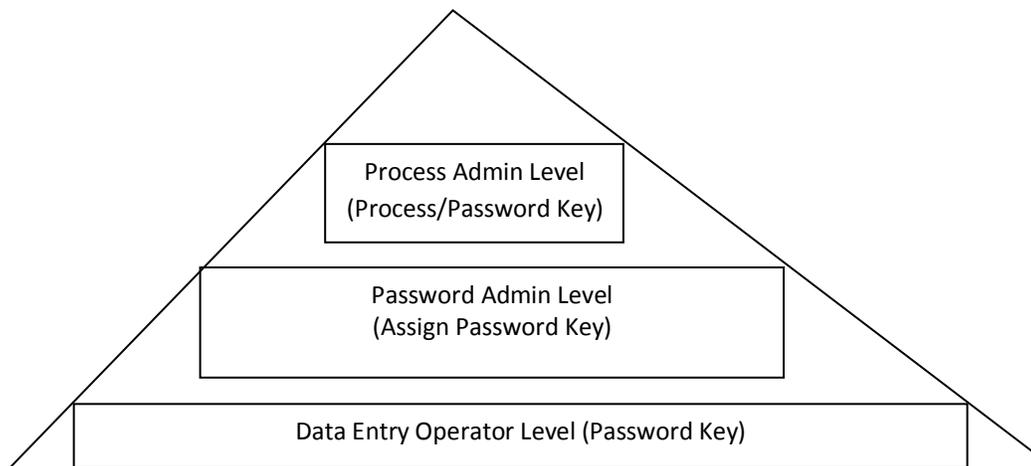


Fig: Shows a Three Level Proposed Architecture of a Data Entry Module for an Operational System pertaining to Data Warehouse.

Data Entry Operator Level

This level corresponds to Data Entry and Data Entry Work. In this level manual award corresponding either rollno - marks or code-marks of a candidate is entered. The operating system used for this purpose is SCO Unix 5.0. The operator using the numeric key pad has to type these roll numbers and marks pertaining to candidates appearing in various examinations conducted by the University of Kashmir. This data entry work is carried out by data entry operators working in a phase manner.

Password Manager Level

In this a password keeper is kept who allocates random username and passwords to operators. Every time the Operator initiates data entry work, he has to receive new password from a Password Manager who gets the new password just by entering the operator code. This password is only valid for a single login which is to be done with three minutes else the scope of the password is lost and it becomes invalid. After three minutes a new password is gets generated and become the active one.

Process Admin Level

This is the highest level of data security operation which is carried out by super user or a System Administrator. When the User logs in a random process key is allocated to the user while performing the data entry process. All the data entered by the data entry operator is identified with this process key, which remains hidden from the operators and can be only accessed by super user or main system Administrator.

Proposed Solution

A password bank database (PBD) is created at Super User Level for keeping track of passwords which are being used by data entry operators[13][14]. Each time the data entry operation is initiated by a data entry operator a new password is provided to him by a password manager. The password is a 5 to 8 letter alpha numeric value which is randomly generated by the password manager using a standard random password generation algorithm. When the password is inputted to another algorithm which generates a process id key value. Both the values the password and the process id key value are stored in the PBD. The access to PBD is limited to the super administrator only. Even the password manager has cannot access PBD. This password is valid for a single login and if not used between particular interval of time expires. Password once used by the data entry operator cannot be used again. Figure below shows the structure of PBG. When any password is randomly selected the operator code date and time are appended with the records and user environment variables.

3.1 PBD Structure

Pa_ID	Passwd	Op_cod	Proc_key	Date_Time_Login	DATE Time Logout	TOT_ENT	Status	MODE
10000	JiK23Q	OPA	IW?/V	July 10, 08 10:45:34:89			A	I
10001	256	unused	unused	Unused			N	
10002	UTIVEI	OPB	IVR&*@	July 10, 08 14:45:34:89			N	
10003	U83	unused	unused	Unused			N	
...	unused	unused	Unused			N	
99999	UE34A	unused	unused	Unused			N	

Algorithm for Password Generation

```

Begin
    for i = 10000 to 99999
        for j = 1 to rand_3to5()
            //rand_3to5 Will Generate number between 3 to 5 for length

            strpass = strpass & chr(alphaNumGenVal())
            //alphaNumGen will generate any numeric value perating to the ascii of alpha or numeric
            value
        end for j
        //Store password in data base
        open_connection
            open_recordset
            rsCon.AddNew
            rsCon.fields("pa_id") = I
            rsCon.fields("passwd") = strpass
            rsCon.Update //Save Record
            close_recordset
        close_connection
    end for I
    
```

Algorithm for Generation of Process Key

```

Begin
    for i = 10000 to 99999
        for j = 1 to rand_3to5()
            //rand_3to5 Will Generate number between 5 to 8 for length

            strpass = strpass & chr(alphaNumUncGenVal())
            //alphaNumGen will generate any numeric value pertaining to the ascii of alpha or numeric value including
            Unicode values.
        end for j
        //Store password in data base
        open_connection
            open_recordset
    
```

```

rsCon.AddNew
rsCon.fields("pa_id") = I
rsCon.fields("passd") = strpass
rsCon.Update //Save Record
close_recordset
close_connection
end for I
    
```

Updating Active User Table

An active User database is created which keeps track of online users and their passwords. When the user logs in the corresponding password entry is updated in his record and when the user logs out the same entry is updated to NULL value. Using the Active User table the Administrator can easily know how many users are working at a particular instance of time. It also keeps track of number of entries done by each user at a particular session [10][11][12].

Modifications in Data Files

The successive modifications are done in the user data file shown in the table below. The time stamp [7][8][9] and the Process key fields are appended with the normal data file for ensuring that no invalid data entry operators can directly access the database without using an appropriate interface provided at the front end.

awdgea. (Award Table for General English A)

Rollno	Marks	Proc_key	date_time
201	39	IW?/V	Feb 14, 2008 12:34:30:23
202	43	><IV34	Feb 6, 2008 01:34:30:12
...

Cross Comparison for Faulty Entry

When the data file entry is complete the system administrator compares the PBD with the corresponding data file. Every process key entry is compared with the process key value with time stamp. All those process key values which are generated in the award files are and cannot be determined from PBD are identified and deleted and the integrity of data is maintained.

The Comparison algorithm

```

Begin
open password bank
store field records in variables (process key, time stamp values, operator codes etc.)
while not eof()

    open subject master
    while not eof()
        open subject file
        while not eof()
            traverse each marks file //mfgea, mfgeb, mfpsa, mfpsb etc

            //Compare stored field records with award file for a particular session.
            if no match then
                print record
            endif
            move next record
        end while
    end while
    move next record
end while
    
```

```
        move next record
    end while
end
```

Efficient Reporting Mechanism

Using the above technique of data entry for a data which has to be finally stored in a data warehouse an efficient reporting mechanism other than finding Invalid Entries in the data files has also been carried out. This include

- Getting Instant Status of Logged in and Idle Operations.
- Efficiency of Each Operator in terms of entry per minute/hour/day/month/year
- Entry wise payment for operators.

Conclusion

Data entry operations have always laid a big challenge to the programmers in designing various security mechanism schemes for such Centralized systems. These security measures not only ensure the data integrity but also provide a suitable environment for data acquisition services which can be then manipulated for transformation and reporting. SCADA (Supervisory Systems and Data Acquisition) is now become a very vital field in the field of Information Technology in which data is acquired from different sources in an automated manner. But still there are very big and bitter challenges ahead in the said field.

References

- [1] Denning, D. E., Denning, P. J., Schwartz, M.D. 1979. The tracker: a threat to statistical database security. ACM Trans. Database Syst. 4(1): 76-96
- [2] D.Ferraiolo, R.Chandramouli, R.Kuhn "Role Based Access Control", Artech House, 2003
- [3] S.Oliveira, O. Zaiane "Privacy Preserving Frequent Itemset Mining", Proc. IEEE ICDM Workshop, 2002
- [4] Denning D.E., Database-Security Annual Review Of Computer Science. 1988; 3: 1-22
- [5] B.Thuraisingham "Recursion Theoretic Properties of the Inference Problem," MITRE report MTP291, June 1990
- [6] W.Jonge "Compromising statistical databases responding to queries about means", ACM Transactions on Database Systems, Volume 8, Issue 1, March 1983
- [7] Multiprocessing Approach to Secure Database Management. ACM Trans. Database Syst. 8(3): 382-409
- [8] S.Jajodia, R.Sandhu, B.Blaustein "Solutions to the Polyinstantiation Problem", Information Security: An Integrated Collection of Essays" vol 1, IEEE CS Press, 1994
- [9] R.Sandhu, F.Chen "The Multilevel Relational Data Model", ACM Trans. Information and System Security, vol 1, 1998
- [10] E.Bertino, E.Ferrari "Secure and Selective Dissemination of XML Documents", ACM Trans", Information and System Security, vol 5, 2002
- [11] J.Crampton, G.Loizou "Administrative Scope: A Foundation for Role-Based Administration", Information and System Security, 2003
- [12] N.Li, M.Tripunita "Security Analysis in Role-Based Access Control", ACM SACMAT 2004
- [13] F. Rabitti, W.Kim, D.Woelk "A Model of Authorization for Next Generation Database System", Database Systems, 1991 CS 8701
- [14] R Agrawal, J Kiernan, R Srikant, Y Xu "Order preserving encryption for numeric data" Proceedings of the 2004 ACM SIGMOD international conference