

On the Fly Sensitive Data Redaction in Database: Implementing Over Decentralized Cross-Agency and Inter-Agency Information Sharing Data Transactions

RavishankarBelkunde
Sr. Database Administrator & Researcher
Boston, Massachusetts, 02108

Abstract: In 2009, Department of Homeland Security(DHS) Data Privacy and Integrity Advisory Committee published White Paper on DHS Information Sharing and Access Agreements the paper reflects consensus recommendations of Fair Information Practice Principles Policy Framework(FIPPs)I. The FIPPs formed foundational principles for Information Sharing and Access Agreements (ISAAs) privacy policy implementation at DHS. In data sharing major challenge is to identify and deploy Information System procedures, business flow to avoid the inevitable 'scope creep' of the specified purpose. In information sharing data accessibility and sensitivity could be categorized at agency level and at individual level. Different government agencies require subset of data from other agencies. Within these government agencies as per the role of an individual and designation data accessibility is decided. For example, DHS have memorandum of understanding between Department of Health and Human Services (HHS), Department Of Labor (DOL), Department of Commerce for ISAAs to only access and use information that DHS is permitted to, protecting privacy of citizens and non-citizens. It becomes difficult to share and manage information from decentralized databases; it's challenging to securely share data to individuals with different security clearance. This work shows how data redaction at database level could address above challenges, what different data redaction techniques we could use to share the sensitive data without making any change to original data.

I. INTRODUCTION

The United States Federal agencies have requested multi billion for cyber security funding for. Our need is to protect data from international cyber threats and share controlled information data within agencies and individuals, to prevent any terrorist attacks. This could be achieved in most effective way by implementing data redaction at the database level. By implementing data redaction, we mask original data on the fly without changing original data value. The data could be redacted at front end or at back end. In front end data redaction process data is masked at the graphical user interface level using front end programming, in this approach there is possibility of eavesdropping sensitive data. Other issue with front end data redaction is it degrade application performance as data is processed twice. Backend data redaction occur in the database using structured query language (SQL) and database programming, the data is redacted at the source without changing original data value. This process is more efficient as data is redacted while data processing requests from application, this is more secure approach as redacted data gets transferred through the network.

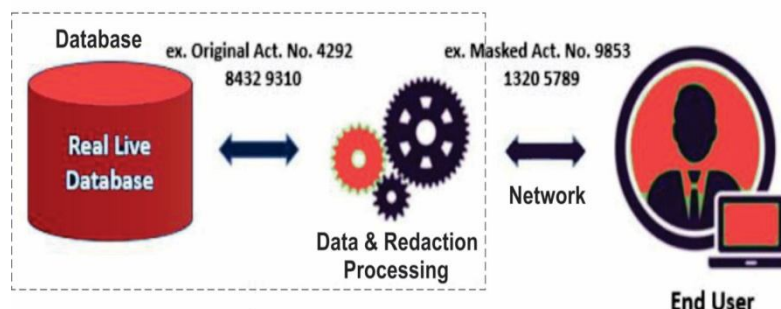


Fig. 1.1 On the Fly Data Redaction

Fig. 1.1 shows block diagram of on the fly data redaction, when application send request to database through network database processes the request if the requested data fall in confidential data category the credentials of the application/users are checked and it's corresponding data redaction policy and method is applied on the data, this redacted data is sent back to the end user, in the fig 1. Act number information was requested as the end user was not authorized to access this information data was redacted and sent back to the user.

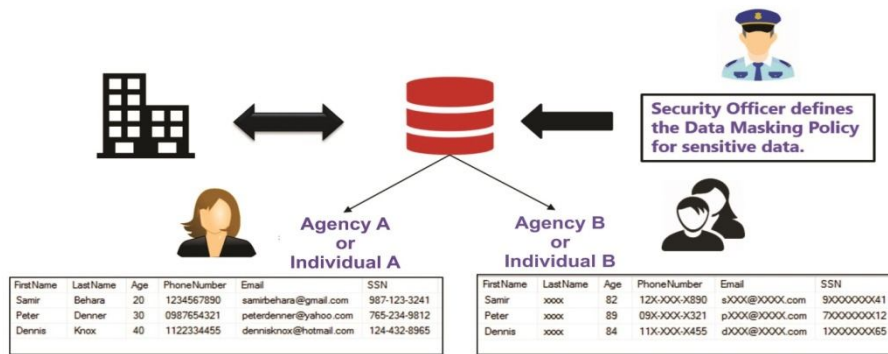


Fig. 1.2 How Data Redaction Works?

In Fig. 1.2 the security officer defines the data redaction policies as per the security clearance and data authorization agency and individuals level, these policies are applied in the database, whenever data request is sent to database these policies are applied, in the example Agency/Individual A have more privileges and access compared to Agency/Individual B. Fig.2 depicts appropriate redaction policies applied and sensitive data redacted at individual level within the agency.

II. COMPONENTS OF DATA REDACTION

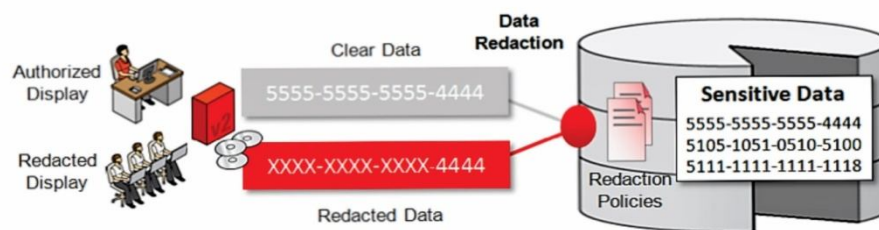


Fig. 2 Data Redaction Example

Following components are used for data redaction.

- (1) **Data:** Identifying sensitive data.
- (2) **Data Correlation:** Correlate sensitive data with other data sets, to maintain data integrity.
- (3) **Data Sub setting:** Create logical datasets to separate sensitive data and to differentiate between different agencies data.
- (4) **Redaction Policies:** Define data redaction policies as per the security clearance, needs of organization, data sharing agreements between agencies and apply these policies to data sets.
- (5) **Redaction Methods:** The data is redacted using different methods, complete redaction, partial redaction, random redaction, user defined function redaction etc.

DATA REDACTION METHODS

We will use test page of the E-Verify application of Department Of Homeland Security to review different methods of database level data redaction. Fig. 3 shows original data of the test sample.

E-Verify Company Information		
Company Name	Company ID Number	Doing Business AS (DBA) Name
New Company Test Account	7533	NCTA DBA
DUNS Number	Account Number	Total Number Of Employees
987654321	946498900983	10,000 and over
EIN	Email ID	Total Revenue
123456789	admin@ncta.com	30 Million
Physical Location	Mailing Address	Owner Name
123 ABC Street Los Angeles CA 90012	456 DEF Street New York NY 01234	Jefferson Hardy Owner SSN 981970952

Fig. 3 Original Data

E-Verify Company Information		
Company Name	Company ID Number	Doing Business AS (DBA) Name
New Company Test Account	7533	NCTA DBA
DUNS Number	Account Number	Total Number Of Employees
987654321	943479800874	10,000 and over
EIN	Email ID	Total Revenue
123456789	admin@ncta.com	30 Million
Physical Location	Mailing Address	Owner Name
123 ABC Street	456 DEF Street	Jefferson Hardy
Los Angeles	New York	Owner SSN
CA 90012	NY 01234	981970952

Fig. 4 Complete Data Redaction

E-Verify Company Information		
Company Name	Company ID Number	Doing Business AS (DBA) Name
New Company Test Account	7533	NCTA DBA
DUNS Number	Account Number	Total Number Of Employees
987654321	946498900983	10,000 and over
EIN	Email ID	Total Revenue
123456789	admin@ncta.com	30 Million
Physical Location	Mailing Address	Owner Name
123 ABC Street	456 DEF Street	Jefferson Hardy
Los Angeles	New York	Owner SSN
CA 90012	NY 01234	98197XXXX

Fig. 5 Partial Data Redaction

E-Verify Company Information		
Company Name	Company ID Number	Doing Business AS (DBA) Name
New Company Test Account	7533	NCTA DBA
DUNS Number	Account Number	Total Number Of Employees
987654321	900494900777	10,000 and over
EIN	Email ID	Total Revenue
123456789	admin@ncta.com	30 Million
Physical Location	Mailing Address	Owner Name
123 ABC Street	456 DEF Street	Jefferson Hardy
Los Angeles	New York	Owner SSN
CA 90012	NY 01234	981970952

Fig. 6 Random Data Redaction

E-Verify Company Information		
Company Name	Company ID Number	Doing Business AS (DBA) Name
New Company Test Account	7533	NCTA DBA
DUNS Number	Account Number	Total Number Of Employees
987654321	946498900983	10,000 and over
EIN	Email ID	Total Revenue
123456789	*****@ncta.com	30 Million
Physical Location	Mailing Address	Owner Name
123 ABC Street	456 DEF Street	Jefferson Hardy
Los Angeles	New York	Owner SSN
CA 90012	NY 01234	981970952

Fig. 7 User Defined Data Redaction



- (1) **Complete Redaction:** In this method original data is replaced by look-alike dummy data. The data type and data format is preserved data values are replaced with some vague values end user will get real feel of data but it will be a dummy data. Fig 4. is an example of complete data redaction, the account number value is redacted it is difficult to differentiate between original and redacted value.
- (2) **Partial Redaction:** In this method original data is partially replaced with some characters. In this method end user knows that data is redacted as visibility of the data is controlled. Fig. 5 is an example of partial redaction. The last four digits of SSN value is masked.
- (3) **Random Redaction:** In this method original data is replaced with some random values using random string or number generation function. Fig. 6 is an example of random redaction. The account number is replaced with some random values, in this method its easy to differentiate between original and redacted value.
- (4) **User Defined Redaction:** In this method original data is replaced with user defined values and criteria. This is powerful redaction method as it gives application owner complete control over redacted data values based on the original data values. Fig. 7 is an example of User Defined redaction. In the example all the values before @ symbol are replaced with asterisk.

CONCLUSION

Database level data redaction solves many technology challenges to meet expectations of data sharing law and policy without modifying data. Data redaction works best in read only environment, if limited accessibility end user gets write permission then there is possibility of permanently replacing original data with the redacted data. The data redaction at source further enhances the network security. We tested different methods of redaction and its usage and how this approach overcomes data sharing and accessibility control challenges overdecentralized Cross-Agency and Inter-Agency Information Sharing Data Transaction.

REFERENCES

1. FIPPS (Fair Information Practice Principles) PRIVACY POLICY GUIDANCE MEMORANDUM: 2008-01 Privacy Act of 1974, 5 U.S.C. § 552a, as amended. Homeland Security Act of 2002, as amended, 6 U.S.C. § 142. Research Service Reports Cybersecurity: (December 29, 2008).
2. Scott Gaetjen, David Knox, William Maroulis Oracle Database 12c Security V1.