Data protection on Business Intelligence Analytics Platform with Masking Level Adjustment

Mykhailo V. Kolomytsev[1], Nataliia M. Kussul[1] , Svitlana O.Nosok[1]

[1] *National Technical University of Ukraine "Igor Sikorsky Kyiv Politechnic Institute", 37 Peremohy Avenue, Kyiv, 03056, Ukraine*

Abstract

Security of data, that is stored on platforms' storage, is an important factor in BI analytical platform. Gathering of data on BI platform and its storing in integrated storage is not end in itself, whilst it serves to obtain new knowledge and improve business processes with using of statistics tools and data mining. Confidential data (both numerical and qualitative) is important attribute of analysis process. In the meantime, it represents the greatest threat as well, since it allows to draw out important conclusions about business performance and trends within considered subject area. Intellectual analysis of confidential data on BI platform supports development of such information transformation methods that on one hand protect data and on the other hand must preserve usefulness of data for further analysis (Privacy-Preserving Data Mining -PPDM).

This article considers BI platform data masking technique that preserves their statistical characteristics. Implementation of the technique is presented in the form of a data protection framework with masking level adjusting.

Keywords

Privacy-Preserving Data Mining, Big Data, Data Security, Data Masking

1. Introduction

Nowadays analytical computing goes though popularization of business intelligence (BI) platform. Key components to such platform are data warehouse (DW), online analytical processing (OLAP) and means of extraction, transformation and loading of data received from different sources (ETL). Gigantic volumes of processed and stored information is an important feature of such platform. This feature leads to the use of big data processing technologies. Confidential information extracted from operational databases - personal information, protected medical information, payment cards details, intellectual property, information about business processes –gets to the storages.

Business intelligence (BI) platform can be integrated with big data technologies. Today, big data is easily accessible to any organization through a public cloud infrastructure. Such integration significantly accelerates the implementation of big data mining methodologies. Security management of DW complicates itself by diversity and multidimensionality of data available to multiple users with different level of authorization, variety of data access processes. This leads to occurrence of multiple threats of data confidentiality breaching. The main purpose of collecting data in the BI platform and storing it in an integrated data warehouse is to extract information/knowledge from data to improve business processes using statistics and data mining tools. From a security standpoint, the main concern is maintaining data confidentiality. Misuse of data analysis may lead to the disclosure of personal data and other data classified as confidential. In particular, user preferences can be analyzed using various big data analytics tools, resulting in privacy invasion. While analyzing data of an organization, various

XXI International Scientific and Practical Conference "Information Technologies and Security" (ITS-2021), December 9, 2021, Kyiv, Ukraine
EMAIL: box144a@ukr.net (M.V.Kolomytsev); nataliia.kussul@gmail.com (N.M.Kussul); nos.sv.ol@gmail.com (S.O.Nosok)
ORCID: 0000-0001-8460-3041 (M.V.Kolomytsev); 0000-0002-9704-9702 (N.M.Kussul); 0000-0002-0016-9346 (S.O. Nosok)

business trends can be identified. Particular attention should be paid to numerical data confidentiality protection. Analysis of such data allows to define business performance efficiency and its trends. Moreover, data of a non-numeric type, such as diagnoses, geographic location codes, and others, also require protection.

Traditional security mechanisms designed to protect small, static data across firewalls and networks are no longer sufficient. Data privacy threats emerge at every stage of the big data lifecycle. However, the greatest risk occurs when the data is located in a non-production environment. Such environment includes databases used for staff training, software development, as well as databases that integrate information from operational databases (DB) and are used to solve business intelligence problems.

With the growth company, the necessity for new applications arises, and these applications require development, debugging, and testing. Realistic data is required for application debugging and testing activities. Existing practice of real data bases passing to development teams carries risks of confidential information falling into wrong hands. This also applies to analysts who need such data for their work. This approach increases the likelihood of data theft. As a rule, a larger project involves a few development teams each needing a separate copy of data.

This means that there will be multiple copies of the production database. According to statistics, there can be 6-8 of such databases. Security infrastructure of production and non-production spheres differ significantly. Such security mechanisms as reliable authorization, logical and physical distinction of access, firewalls are necessary in a production environment.

However, such measures are generally not available in non-production environments. This can lead to misuse of master data copies. In such situation, for hackers that have gotten inside companies' infrastructure, it's not hard work to steal the data used for application development.

These reasons necessitate the development of methods for such information transformation, which, on one hand, protect the data, and, on the other hand, must preserve their usefulness for analysis (PPDM). There are various PPDM methods [1], the most popular being data de-identification, i.e. removing confidential information from data or replacing it with anonymized information. In a broad sense, de-identification means the impossibility of subject area processes details recovering. The most actively developing direction of de-identification is called masking.

Data masking inherently protects sensitive data from unauthorized access by changing its values while maintaining original restrictions of data values. Masking methods must preserve original characteristics of information and maintain integrity of data and links. Quality of development, testing and learning directly depends on quality of testing data and its' realism. Such data can be used to solve analytical problems. At the same time, data privacy solutions must be easily customizable to meet needs of any organization.

Well known methods of masking can demonstrate (though not always) good results at application functionality and users interface testing

However, they are completely unacceptable in terms of data analysis and building business models. Statistical characteristics of masked data differ significantly from original ones, that has a significant negative impact on the accuracy of analytical constructions.

2. Data masking methods

There are various masking methods. For example, in [2] the following methods are considered:

Substitution. In this case, replacement of one value with another is used. For example, the subject's last name is replaced with a randomly selected last name from a lookup table generated from a telephone directory. Complexity occurs if it is necessary to generate a very large lookup table. For example, if it's needed to generate several million realistic customer email addresses.

Redaction/Nulling. This method is a special case of the substitution method, when all masked characters are replaced by the same character, for example, "X". In this case, the masked phone number will look as "(XXX) XXX-XXXX". This is the easiest and fastest masking method. It is widely used as a built-in masking method in the DBMS, but it's not greatly valuable.

Shuffling. Shuffling is a method of randomizing existing values vertically in the dataset, i.e., permutation in a table column. However, if only permutation is applied, masking is unreliable. A person with any knowledge of real values can consistently reconstruct original data. Permutation method is

effective in case of large massive of initial data. Since there is no need to generate new values, the method is simple and fast enough. During implementation of the method, special attention should be paid to randomization of permutation process.

Blurring. Original value is replaced with a random but close (within a certain range) value. For example, replacing real sales data with a random value that differs in the range of 5% from the original. The method can be useful if it's needed to hide correlation between the original numerical values.

Averaging. In this method, original numbers are replaced with random ones in such a way that average value of the entire set of masked values remains the same as in the original set.

De-identification. General name for the methods that allow original information identifying a person to be transformed in such a way that connection with this person disappears. De-identification is used to mask complex data sets spanning multiple columns of a database table.

Tokenization. In this method, data elements are replaced with random placeholders (tokens). Representing of data as tokens is irreversible because token is not logically related to the original value.

Format-preserving encryption. In this masking method, data is converted into an encrypted form in such a way that general appearance of the original value is preserved.

Static and dynamic masking are distinguished according to the method of masking process organization. Static masking creates a copy of production database, replacing protected data with masked values. Static masking allows to create realistic test databases and reduce risks of information disclosure in a non-production environment as test database does not contain unmasked data. Dynamic masking is performed on production database. Conversion process is carried out in an intermediate software layer, between production database and application. Dynamic masking is triggered at the time of accessing the database and modifies its responses in such a way that anonymized data is submitted. Dynamic masking solutions are designed to protect data stored in production databases. This approach reduces risks of data leakage from insider actions.

3. Related works

A significant number of works are devoted to masking, see, for example, [3] - [5]. Loss of quality of analyzed data is classic masking methods serious problem, although they handle de-identification task well.

Works [5] - [8] are devoted to development of masking methods that protect data from disclosure, and at the same time, preserve their useful properties for analysis. These works consider masking methods that preserve statistical properties of the original data. The focus is on preserving of set properties, but level of protection of sensitive data becomes low.

In [9], it is proposed to protect data confidentiality by decomposing the original structure of tables and separating confidential data into a separate table. Such transformation is significantly complicated by the presence of relationships between tables and cannot be performed in real time proximity.

Security of big data at different stages of its' life cycle is considered in [10]. Authors identify 5 stages of database life cycle, and for each stage, threats and security methods are considered.

In [11, 17], to ensure confidentiality the use of so-called K-anonymity is considered. This approach requires that each entry in the table is not different at least from K-1of other entries. From view point of confidentiality maintaining, this approach has disadvantages, noted for example in [12]. To address these shortcomings, K-Anonymity methodology has been extended. One of examples is L-Diversity, which requires that each group of individuals that are indistinguishable by quasi-identifiers (such as age, gender, zip code, etc.) should not have the same sensitive attribute value, but should have L of separate well-represented (approximately in the same proportion) values [13].

So-called differential privacy work [14] is dedicated to eliminating of shortcomings attributed to K-Anonymity and L-Diversity. This approach largely eliminates privacy protection issues of K-anonymity, L-variety and their extensions. A masking mechanism is said to guarantee ε-differential privacy if, for each pair of inputs D1 and D2, probability of masked values matching must be very high (within the coefficient **exp(ε)**).

As a rule, masking process is implemented at ETL (extract, transform and load) stage - i.e. in the process of data transferring from operational databases to storage.

Each of discussed masking methods is efficient in protecting privacy. However, relational database applications introduce additional complexity. In particular, one of qualities of relational model is referential integrity maintaining. If masked data item is a primary or a foreign key in database table relationship, then this newly masked data value must be propagated to all related tables in the database. Propagation of key ensures referential integrity of transformed data. Without propagation of key, relationship between parent and subsidiary tables will be broken, resulting into loss of integrity of masked data.

Analysis of the works allows us to formulate general requirements for masking process:

1. Masking should not be reversible. Possibility of unmasking procedures creates threats of data disclosure and lowers level of data security. During implementation of masking process, it must be taken into account that the result of the transformation should not be reversible, which means that persons who do not have access to key information should not be able to restore the original text using the masked one.
2. Only sensitive data needs to be masked. This requirement points to the need of metadata expert analysis in order to identify exact attributes to be protected.
3. The results should represent original data. Data masking is to provide information that still resembles real data and that is useful for analytical processing.
4. It is necessary to maintain referential integrity. Masking should not violate integrity of foreign keys of tables. This means that if masked attribute is a foreign key, it is also necessary to mask the primary key in the parent table and cascade it in related tables. Therefore, it is also necessary to ensure entities integrity. If masked attribute is a primary key, masked values must be unique.
5. Masking should be an iterative process, meaning that for different instances of certain data structures, masking process does not require changes in settings. At the same time, masking the same data instance again should produce values that are different from the previous ones.
6. Applicability of the masking algorithm to the entire data set of the domain. If, for example, technological process data is being masked, then masking algorithm must adequately handle the entire range of input data.

In addition to the basic requirements, additional requirements can occur, due to the characteristics of the subject area, for example:

- Saving generalized values. Total and average values for masked column of the table must match the original ones (precisely or with a certain tolerance).
- Statistical distribution of values. In some cases, it is important to retain information about such statistical characteristics as nature of the distribution. For example, if the database contains information about geographical distribution of cancer patients by postal codes, then an arbitrary replacement of postal codes can distort results of the analysis.

In general, known implementations of masking methods, as a rule, do not ensure fulfillment of points 3, 4 of the requirements. Masked data format, belonging to the same domain as the input data, referential integrity are generally not supported.

4. Methodology

Given the peculiarity of BI platform, the following requirements added to listed above:

1. Preservation of data belonging to a specific domain. This restriction can be represented as a set of particular requirements, for example:
   - Saving data type. Masking results must belong to the same base data types as the original data.
   - Save format. This means that if table attribute is a symbolic type and line limitation is from 10 to 15 symbols, then masked data should also respond to this term. A typical example is date masking, which must occur in the correct ranges for day, month and year. This means that the masking algorithm must determine data format and besides generate a suitable in the same format.

2. Speed of operations performing at ETL stage or at query level on masked data should be comparable to the speed of performing data operations without masking. This requirement is especially important in cases of big data.
3. Slight change in size of the masked table compared to the original. Ideally, tables should not be resized when masked.
4. Masking method should provide masking of both quantitative and qualitative data.
5. Individual masked value cannot be used to find out true value, but actual average behavior of masked data should be close to that of the original data.
6. Masking process must be customizable. This means possibility to change process settings in such a way that a compromise is reached between level of data confidentiality and its usefulness (reliability).

In this work masking technique that meets the above requirements is considered. During creation of the technique results of authors given in [15,16] were used. Based on the methodology data masking framework has been created that functions at ETL stage.

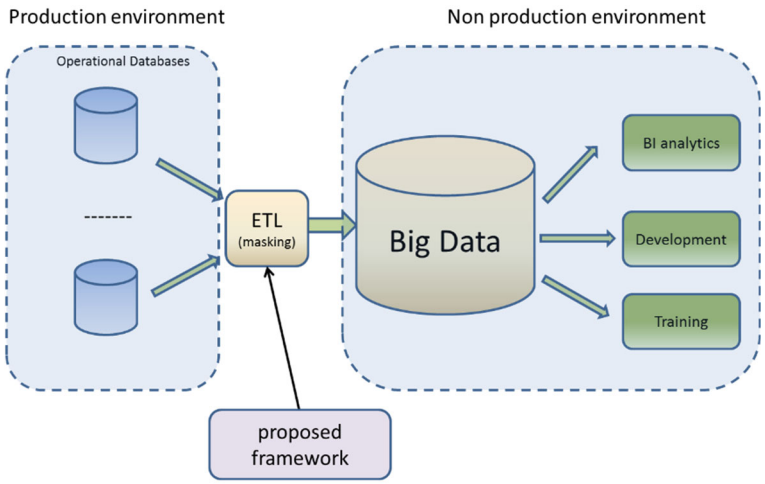

Figure 1: General masking structure in BI platform

Figure 2 shows model for protecting data privacy using masking and preserving usefulness of the data.

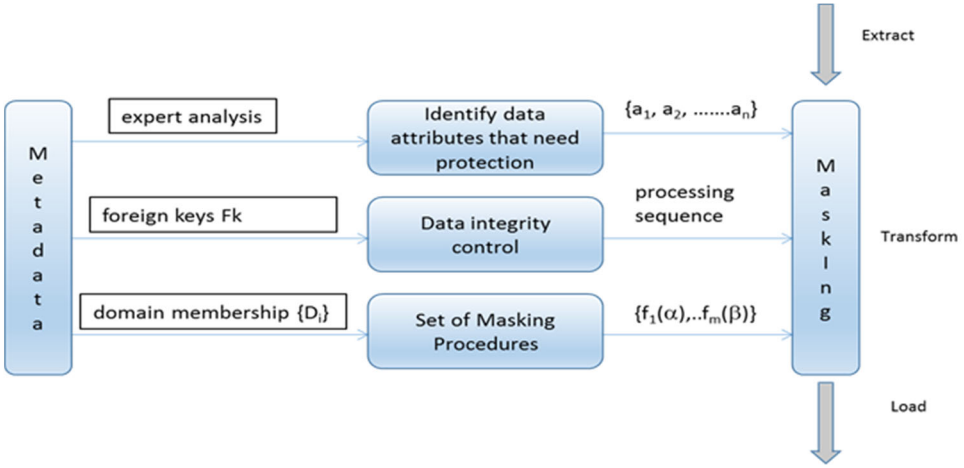

Figure 2: Data utility-preserving privacy model

Purpose of main components of the model:

- Identification of sensitive data. Based on results of metadata analysis, user (expert) selects those that need to be masked. User (expert) analyzes all attributes of sample data by studying subject

area, analyzing reports and documents. Data format, privacy policy, reliability requirements (usefulness) of data, level of security are determined

- Integrity control. By analyzing metadata, relationships between tables are determined. Masked foreign key values must match masked primary key of the parent table. Subsequently masked data is loaded into DW without errors.
- Masking procedures. Depending on type and format of data, one of masking procedures is selected. Masking procedures allow to keep data realistic. For string variables, code page, presence of upper- and lower-case characters, and length are saved. For numbers, bit depth and sign are preserved. Masking date and time data produce values in correct ranges for day, month, year, and time. This property is one of factors that ensure realism of data. Some data items can be set as non-maskable. Moreover, size of that part of the initial data that does not change can be adjusted (parameter α of the model). This parameter allows to maintain such data property as frequency distribution. In cases where it is necessary to maintain a logical grouping of values, technique allows to save elements of data template, forbidding masking of a certain part of the data. Obviously, such attribute as postal index should not be randomly masked since geographical component might be destroyed. In this case, it is possible to customize masking procedure in such a way that part of the index that determines, for example, belonging to a region or city, does not change. This option also allows to make masked data more similar to the original.
- Masking. First of all, foreign keys are masked along with corresponding primary keys. They do not participate in further transformation. Depending on data type of table column, masking procedure is automatically selected and applied. The α parameter specifies the part of data that is not masked. Value of the parameter should be able to provide a compromise between requirements of confidentiality and usefulness of data. Masked values are unique, and masking the same value again produces a different result.

Masking algorithm:

1. Determine tables with data subjected to protection T1…Tm and attributes of tables Ti{a1, a2, …….an} to be protected. Steps 1 and 2 are performed by an expert. These steps are performed once.
2. Conditions for integrity of selected attributes are determined whether it is a foreign key. If yes, attributes of related tables are determined. Masking of related attributes occurs according to the method described in [16].
3. Data type for each attribute is determined and appropriate masking procedure is assigned.
4. Based on the confidentiality / usefulness requirements, parameters of masking procedures are determined (parameter α)
5. All rows of the table are processed sequentially, performing the transformation

$$T_{in}\{a_1, a_2, \ldots\ldots a_n\} \Rightarrow T_{in}\{s_1, s_2, \ldots\ldots s_n\}$$

where $\mathbf{a_k}$ is unmasked attribute value, $\mathbf{s_k}$ is masked

6. Procedure of masking value for each attribute is the following:
   - unmasked attribute value is converted into a sequence of characters
     $a_k = \{c_1, c_2, c_3, \ldots\ldots c_q,\}$,
   - masked value view is:
     $s_k = \{c_1, ..c_p, f_{p+1}, \ldots\ldots f_q,\}$,

   where part [0 – p] of characters remains unchanged. Remaining characters are replaced with random values from the domain to which original characters belong. Number of unmasked characters is determined by the parameter α
7. Since masked characters belong to the same domain as unmasked characters, masked values are converted into original value format.
8. Repeat steps 2 - 7 for each table defined in step 1.

Data masked using the proposed technique has the following properties:

- Preservation of the format and type of data. Initial data structure saving. For string type data this means saving of line length. Upper- and lower-case characters, vowels and consonant are in the same positions as in initial line. The masked number value has the same bit depth and sign as the unmasked one. Masking date and time data produces values in the correct ranges for day, month, year, and time. This property is one of the factors that ensure realism of the data.
- Ability to leave some part of the original value unchanged.
- Referential integrity. The masking technique maintains referential integrity between tables. Therefore, loading the masked data into the DW proceeds without errors.
- Frequency distribution. In some cases, it is necessary to support a logical grouping of values. The technique allows you to save the elements of the data template, prohibiting the masking of a certain part of the data. If source data contains zip code, you can customize the masking procedure in such a way that the part of the index that determines, for example, belonging to a region or city, does not change.
- Uniqueness. Masked values are unique, and masking the same value again produces a different result.

Empirical research has shown that as the degree of masking increases (the parameter α decreases), difference between statistical properties of actual and masked values increases. As degree of masking decreases, statistical properties of masked data become close to the original. User can perform masking with α setting for both privacy and realism.

5. Experimental study

To test masking results, dataset containing 700 rows of sales and profit information, sorted by market segments and countries [17] was used.

2 data masking variants were studied:

- all sales data is masked ($\alpha = 0$)
- first digit of sales data is not masked ($\alpha = 1$)

Statistical characteristics were analyzed:

- original and masked data
- sales data grouped by months in both original and masked forms.

5.1. Statistical analysis of masking results

Samples were compared by testing hypotheses using Student's t-test. Null hypotheses of equality of means and variances of two sets of data were tested.

Test results:

Table 1

t-Test: Paired two sample for means results

	No mask	$\alpha = 0$
Mean	169609,0718	222172,2758
Variance	56039363321	1,00298E+11
Pooled Variance	0,112643644	
t-Stat	-3,724140007	
P(T<=t) two-tail	0,000211775	
t Critical two-tail	1,963363576	

	No mask	$\alpha = 1$
Mean	169609,0718	175099,4411
Variance	56039363321	63814921540
Pooled Variance	0,317907741	
t-Stat	-0,50779584	
P(T<=t) two-tail	0,611756717	
t Critical two-tail	1,963363576	

Table 2

t-Test: Two-sample assuming equal variances results

	No mask	$\alpha = 0$
Mean	169609,0718	222172,2758
Variance	56039363321	1,00298E+11
t-Stat	-3,517221888	
P(T<=t) two-tail	0,000450047	
t Critical two-tail	1,961662333	

	No mask	$\alpha = 1$
Mean	169609,0718	175099,4411
Variance	56039363321	63814921540
t-Stat	-0,419588714	
P(T<=t) two-tail	0,674850403	
t Critical two-tail	1,961662333	

Test results allow to draw the following conclusions:

- in the case of complete masking of sales data ($\alpha = 0$), the hypotheses about equality of means and variances are rejected. Pearson's correlation value = 0.112. Correlation is practically absent, which makes it impossible to calculate unmasked values from masked ones. However, masked data characteristics differ significantly from masked data, and they are not suitable for BI analytical processing.
- in case of partial masking of sales data ($\alpha = 1$), hypotheses about the equality of means and variances are not rejected. Pearson's correlation value = 0.317. The values of the series are weakly related, which makes it impossible to calculate the unmasked values from the masked ones. At the same time, statistical characteristics of masked data quite reliably repeat characteristics of unmasked data, and they are suitable for analytical processing of BI.

5.2. Statistical analysis of grouped data

Grouping data is an important element of BI analytics. Masking should not interfere with trending when summarizing large amounts of data.

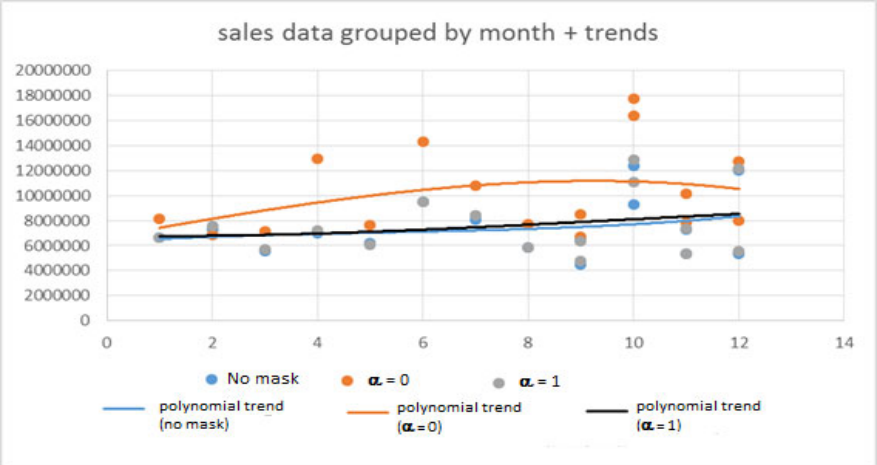

Figure 3: Comparison of data grouped by months with different masking parameters

Test results:

Table 3

t-Test: Paired two sample grouped data for means results

	No mask	$\alpha = 0$
Mean	7420396,891	10220037,07
Variance	5,33867E+12	1,2601E+13
Observations	16	16
t-Stat	1,753050356	
P(T<=t) two-tail	0,000165285	
t Critical two-tail	2,131449546	

	No mask	$\alpha = 1$
Mean	7420397	7660601
Variance	5,34E+12	6,25E+12
Observations	16	16
t-Stat	1,75305	
P(T<=t) two-tail	0,051593	
t Critical two-tail	2,13145	

Table 4
t-Test: Two-sample grouped data assuming equal variances results

	No mask	$\alpha = 0$
Mean	7420396,891	10220037,07
Variance	5,33867E+12	1,2601E+13
Observations	16	16
t-Stat	-2,64396268	
P(T<=t) two-tail	0,012906186	
t Critical two-tail	2,042272456	

	No mask	$\alpha = 1$
Mean	7420397	7660601
Variance	5,34E+12	6,25E+12
Observations	16	16
t-Stat	-0,28224	
P(T<=t) two-tail	0,7797	
t Critical two-tail	2,042272	

Here one can also see that if the data is masked entirely ($\alpha = 0$), then the hypothesis of equality of means and variance is rejected, and for a partially masked ($\alpha = 1$) sample, the hypothesis of equality of means is not rejected, and the hypothesis of equality of variances is confirmed with a high probability.

The graph clearly shows that by changing masking level parameter, it is possible to achieve a practical coincidence of trend lines (fig. 3).

Results of the study suggest that the proposed masking method combines properties of blurring methods group with possibility to adjust level of masking.

6. Conclusion

Compared to other data protection tools, masking provides a unique opportunity because it can provide privacy and preserve complex data relationships and characteristics. No other data protection method provides these benefits at the same time. Masking reduces security risks with minimal impact on operations of BI systems. Data masking in non-production environment becomes more and more common. Integration of data masking infrastructure into the BI platform is under study, and currently main results are focused on the use of traditional methods for data masking that do not take into account usefulness of data.

Using traditional masking technique in the BI platform is inefficient. Therefore, in this work, a masking technique with support for the usefulness of the data was proposed. The proposed technique identifies sensitive data and stores it securely in DW while providing usefulness of masked data. Data transformed according to the above technique can be used in non-production databases for such purposes as application testing, staff training and analytical processing.

References

[1] J. Koo, G. Kang, Y.-G. Kim,Security and Privacy in Big Data Life Cycle: A Survey and Open Challenges, URL: https://www.mdpi.com/2071-1050/12/24/10571.

[2] W. Ahmed, Data Masking & Techniques, IJTRS, v. IV Issue VIII, 2019 URL: https://www.ijtrs.com/uploaded_paper/Data Masking & Techniques.pdf. DOI: https://doi.org/10.30780/IJTRS.V04.I08.003.

[3] Privacy Preserving for Sensitive Data using Data Masking Technique. URL: https://www.ijrte.org/wp-content/uploads/papers/v8i6/F9422038620.pdf.

[4] S. F. Shah , Z. Hussain, M. Riaz, S. A. Cheema, Shewhart-Type Charts for Masked Data: A Strategy for Handling the Privacy Issue, Mathematical Problems in Engineering Volume 2020. DOI: https://doi.org/10.1155/2020/5104753.

[5] A. Lane, Understanding and Selecting Data Masking Solutions: Creating Secure and Useful Data. 2012. URL: https://securosis.com/assets/library/reports/ UnderstandingMasking_FinalMaster_V3.pdf.

[6] O. Ali, Secured Data Masking Framework and Technique for Preserving Privacy in a Business Intelligence Analytics Platform. 2018. URL: https:// https://ir.lib.uwo.ca/etd/5995.

[7] Ravikumar GK, B. Justus Rabi , Manjunath T N, A Study on Dynamic Data Masking with its Trends and Implications, International Journal of Computer Applications, V. 38– No.6, 2012. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.735.9965.
[8] O. Ali, A. Ouda, A Classification Module in Data Masking Framework for Business Intelligence Platform in Healthcare, URL: https://https://ieeexplore.ieee.org/document/7746327.
[9] S. Ali, A. Rauf, J. Ahmad, Protecting unauthorized big data analysis using attribute (data) relationship, 2015. URL: http://www.sci-int.com/pdf/11068939081 a 5075-5077 Shaukat Ali Khan --IT--KPK--TR.pdf.
[10] J. Koo, G. Kang, Y. Kim, Security and privacy in big data life cycle: a survey and open challenges. 2020. URL: https://www.mdpi.com/2071-1050/12/24/10571.
[11] L. Sweeney, K-Anonymity: A Model For Protecting Privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10 (5), 2002. URL: https://epic.org/wp-content/uploads/privacy/reidentification/Sweeney_Article.pdf.
[12] R. Mendes, J. Vilela, Privacy-Preserving Data Mining: Methods, Metrics and Applications. 2017. URL: https://ieeexplore.ieee.org/document/7950921. DOI: 10.1109/ACCESS.2017.2706947.
[13] A. Machanavajjhala, J. Gehrke, D. Kifer, ℓ-Diversity: Privacy Beyond k-Anonymity. 2006. URL: https://ieeexplore.ieee.org/document/1617392. DOI: 10.1109/ICDE.2006.1
[14] C. Dwork, A. Roth, The Algorithmic Foundations of Differential Privacy. 2014. URL: https://www.cis.upenn.edu/~aaroth/Papers/privacybook.pdf. DOI: 10.1561/0400000042.
[15] M. Kolomytsev, S. Nosok, A. Mazurenko, Database tables masking using the SQL CLR technology. Ukrainian Information Security Research Journal, Vol. 19, N 1(2017). P.16-22. DOI: https://doi.org/10.18372/2410-7840.19.11440.
[16] M. Kolomytsev, S. Nosok, A. Mazurenko, Integrity control of masked database foreign key. Ukrainian Information Security Research Journal, Vol. 17, N 4(2015). P.306-311. DOI: https://doi.org/10.18372/2410-7840.17.9789.
[17] M. Sparkman, Financial Sample Excel workbook for Power BI. 2022. URL: https://github.com/MicrosoftDocs/powerbi-docs/blob/live/powerbi-docs/create-reports/sample-financial-download.