A Modified Hybrid AES with SHA based Data De-duplication with Secure Encryption in Cloud

Deepak Kumar[#1], Dr. Umesh Lilhore[*2], Prof. Ankita Singh[*3] Prof. Nitesh Gupta[*3]
[#1]Research Scholar of CSE, [+2]Head of department, *[3]Assistant Professor CSE
NRI Institute of Information Science & Technology Bhopal, (M.P.) India

Abstract: Dynamic data is the big problem in big data and clouds. There are different resource available that create duplicate data. In this survey paper discuss the different data duplication avoiding methods. In the last decade there are different data duplication methods are proposed. In this present work discuss different method as well as compare those methods in the literature survey on the basis on survey problem proposed a modified hybrid AES-SHA 512 based Data De-duplication with Secure Encryption in Cloud. The main problem originate due to data duplication that is bandwidth consumption and space wastage due to same data. To avoid these problem researchers are focus on different data duplication techniques in last decade. In this proposed work use simple hash algorithm with advance encryption system for avidness of de-duplication in cloud. The proposed method output shows better result is compare to other previous method such as MD5, SHA – 1 as well as DSA in terms of security key generation time, encryption time and data duplicity find out time. Also the proposed method preserve same of the cloud server and bandwidth of communication channel.

Keywords – Dynamic Data, Deduplication, Social Media, SaaS, PaaS, IaaS, SHA -1, MD -5 , MD 2 , DSA , AES and Cloud Model.

I. INTRODUCTION

Cloud computing (on-demand computing), it is a branch of computing which provide services on demand. In this cloud computing, computing resources, data and information can be shared between one system to another system on demand. Cloud Computing may be a model obtainable anyplace, to access on-demand to a shared pool of configurable computing resources. Cloud computing provides solution for storage to clients and different organization with various capability to save and process their data in third-party data centre. It depends on sharing of resources to induce consistency and economies of scale, same because the utility into a network. [1]

To preserve privacy of data, encryption and decryption techniques are used. The plaintext is the original form of data and the cipher text is the encrypted form of data. Encryption strategies take plain content (unique type of information) as an info and change over it into figure content (scrambled type of information), in view of calculation utilizing a key. A key is a component on the basis of which data is encrypted. Unscrambling systems takes the figure content (scrambled type of information) and change over it into plain content (unique type of information) in light of calculation utilizing a key. [2]

Cryptography algorithms are divided into two different categories on the basis of the keys used-

1. Symmetric Key Cryptography and
2. Asymmetric Key Cryptography

In Symmetric Key Cryptography algorithms a single key is used for encrypting and decrypting the data. A key is transmitted to both sender and receiver before communication. It is also called private or secret key cryptography. Then again, Asymmetric Key Cryptography calculations utilize a key combine known as open key and private key. People normally key's utilized for coding the message although personal key's utilized for unscrambling the message. They are likewise called open key cryptography techniques. Symmetric key encryption is better than asymmetric key encryption. [3]

II. DE-DUPLICATION

Data de-duplication is sometimes called an intelligent compression or single-instance storage; is a process that removes redundant copies of information and reduces storage aloft. Information de-duplication strategies guarantee that just a single diverse occurrence of data is held on capacity, for example, plate, blaze or tape. Repetitive information squares are supplanted with a pointer to the one of a kind information duplicate. In that way, information de-duplication nearly lines up with incremental reinforcement, which duplicates just the information that has changed since the past reinforcement. Aligns with incremental backup, which copies only the data that has changed since the previous backup.[4]

Information de-duplication is one of the current advances away right now since it empowers organizations to spare a ton of cash on capacity expenses to store the information and on the transfer speed expenses to move the information while imitating it offsite for DR. This is awesome news for cloud suppliers, in light of the fact that in the event that you store less, you require less equipment. In the event that you can de-copy what you store, you can better use your current storage room, which can spare cash by utilizing what you have all the more proficiently. [5]

If you store little, you also back up little, which again means less hardware and backup of media. On the off chance that you store less, you likewise send less information over the system if there should be an occurrence of a calamity, which implies you spare cash in equipment and system costs after some time. The business benefits of data deduplication will be:

- Reduced hardware costs;
-
- Reduced backup values;
- Reduced values for business continuity / disaster recovery;
- Increased storage efficiency; and
- Increased network efficiency.

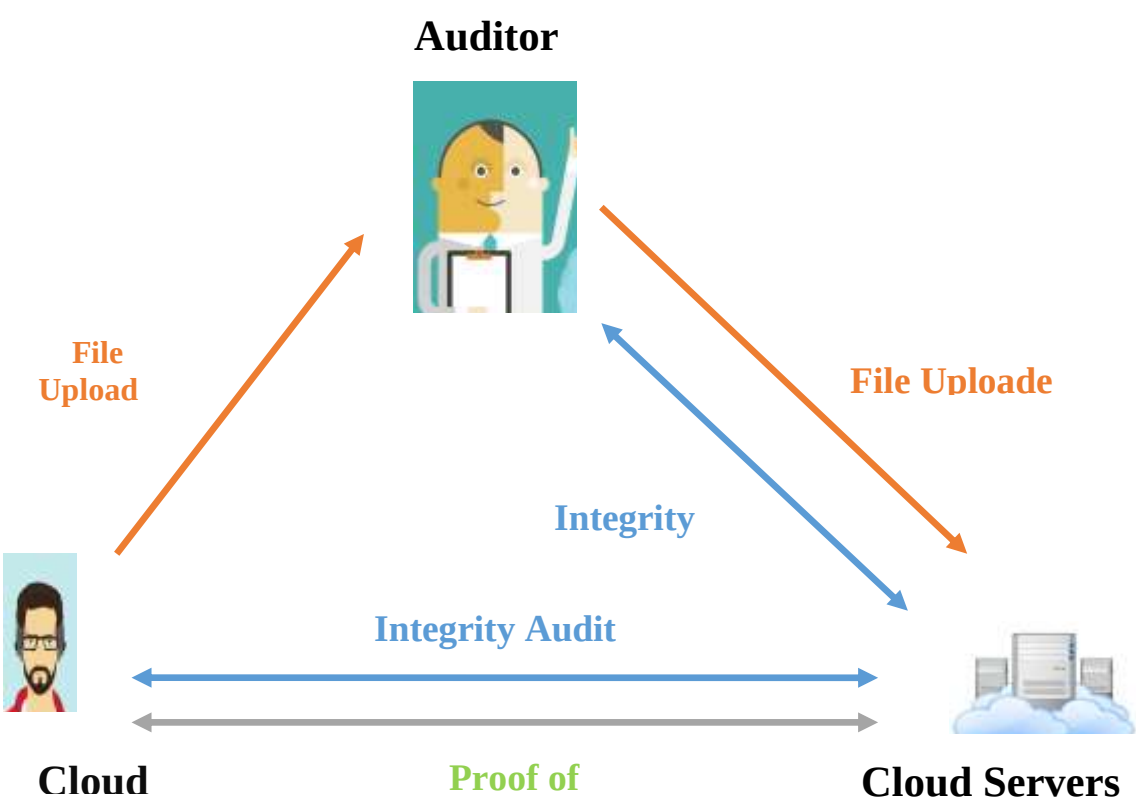

Fig. 1. Architecture of Auditor based Data Proving

III. PROPOSED METHOD

The overall proposed method is divided into two sections. In the first section, we upload the file on cloud and also check the duplicate file present or not in the cloud. To avoid the duplication problem in the cloud. That why when upload any file cloud check same file available or not on cloud if same is already exists in cloud. In this case file is not upload on the cloud. After the verification of file, in the second stage upload file on cloud data server with advance encryption system using AES.

Section -1 File Uploading on cloud and duplicate checking
Section -2 Data verification by Data Auditor

Algorithm Codes –

1. User file Uploading Algorithm on Cloud
2. User – log in
3. User enter in the system
4. Select file for uploading
5. Check hash value if similar hash already exist in the system file not upload)
6. Fn(file id) = SHA-2 (Hash value == data log)
7 Send message to user ---("This file is already exist in the cloud ")
8. Stop operation
9. Else
10. Fn(1st file) = SHA-2(Hash value ≠ data) // hash value not matched
11. Apply encryption on selected file first fread // read test file
12. Fn (file id) = encryption_AES ('file', 'key') ;
13. F (ency.) // Store in the admin Log files
14. end

IV. ADVANCED ENCRYPTION STANDARD (AES)

Sub Bytes () Transformation

The Sub Bytes () transformation is a non-linear byte substitution that operates independently on each byte of the State using a substitution table (S-box). This S-box (Fig 1), which is invertible, is constructed by composing two transformations [1]:

1. Take the multiplicative inverse in the finite field GF(2^8), described in Sec the element {00} is mapped to itself.
2. Apply the following affine transformation

$$b_i' = b_i \oplus b_{(i+4 mod 8)} \oplus b_{(i+6 mod 8)} \oplus b_{(i+7 mod 8) \oplus c_i} \quad (1)$$

$for\ 0 \le i < 8, where\ b_i is\ the\ i^{-th}$ Bit of a byte c with the value {63} or {01100011}. Here and elsewhere, a prime on a variable (e.g., b¢) indicates that the variable is to be updated with the value on the right [3].

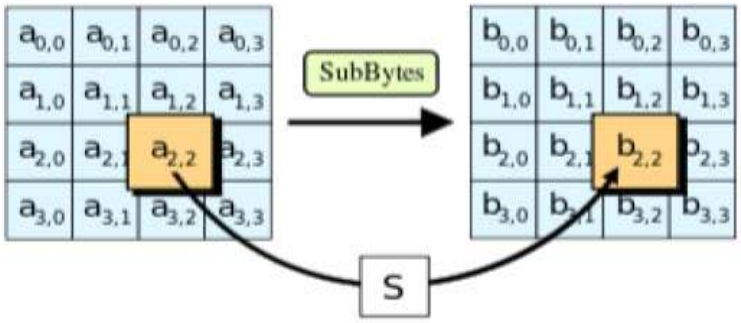

Fig. 2 Block diagram of SubBytes Transformation.

The various transformations (e.g., Sub Bytes (), Shift Rows (), etc.) act upon the State array that is addressed by the „state" pointer. Add Round Key () uses an additional pointer to address the Round Key. In matrix form, the affine transformation element of the S-box can be expressed as:

$$\begin{bmatrix} b'_0 \\ b'_1 \\ b'_2 \\ b'_3 \\ b'_4 \\ b'_5 \\ b'_6 \\ b'_7 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \quad (2)$$

Shift Rows () Transformation

In the Shift Rows () transformation, the bytes in the last three rows of the State are cyclically shifted over different numbers of bytes (offsets). The first row, r = 0, is not shifted. Specifically, the Shift Rows () transformation proceeds as follows:

$$S_{r,c} = s_{r,(c+shift(r,Nb))modNb} \ for\ 0 < r < 4\ and\ 0 \le c < Nb \quad (3)$$

Where the shift value shift(r, Nb) depends on the row number, r, as follows (recall that Nb= 4): Shift(1,4)=1 ; shift(2,4) =2; shift (3,4)=3.

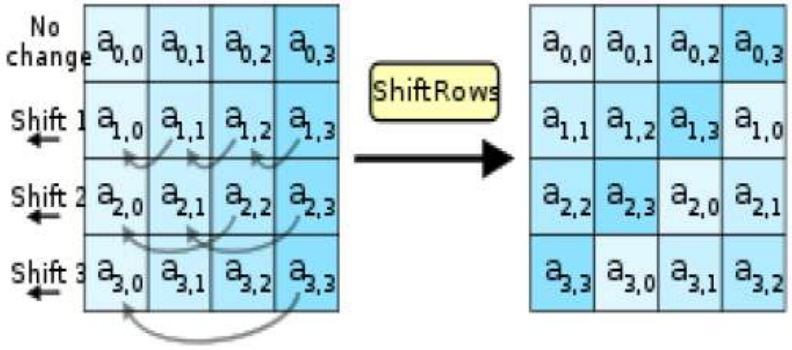

Fig 3 Shift Rows, cyclically shift the last three rows in the state.

The Mix Columns () transformation operates on the State column-by-column, treating each column as a four-term polynomial. The columns are considered as polynomials over GF (2^8) and multiplied modulo x4+ 1 with a fixed polynomial a(x), given by

$$a(x) = \{03\}x^3 + \{01\}x^2 + \{01\}x + \{02\} \quad (4)$$

This can be written as a matrix multiplication. Let

$$S'(x) = a(x) \otimes s(x)$$

$$\begin{bmatrix} S'_{0c} \\ S'_{1c} \\ S'_{2c} \\ S'_{3c} \end{bmatrix} = \begin{bmatrix} 02 & 03 & 01 & 01 \\ 01 & 02 & 03 & 01 \\ 01 & 01 & 02 & 03 \\ 03 & 01 & 01 & 02 \end{bmatrix} \begin{bmatrix} S_{0'c} \\ S_{1'c} \\ S_{2'c} \\ S_{3'c} \end{bmatrix} for\ 0 \le c < Nb \quad (5)$$

As a result of this multiplication, the four bytes in a column are replaced by the following:

$$\left.\begin{aligned} S'_{0c} &= (\{02\} \bullet S_{0'c}) \oplus (\{03\} \bullet S_{1'c}) \oplus S_{2'c} \oplus S_{3'c} \\ S'_{1c} &= S_{0'c}(\{02\} \bullet S_{1'c}) \oplus (\{03\} \bullet S_{2'c}) \oplus S_{3'c} \\ S'_{2c} &= S_{0'c} \oplus S_{1'c}(\{02\} \bullet S_{2'c}) \oplus (\{03\} \bullet S_{3'c}) \\ S'_{3c} &= S_{0'c}(\{02\} \bullet S_{1'c}) \oplus (\{03\} \bullet S_{2'c}) \oplus S_{3'c} \end{aligned}\right\} \quad (6)$$

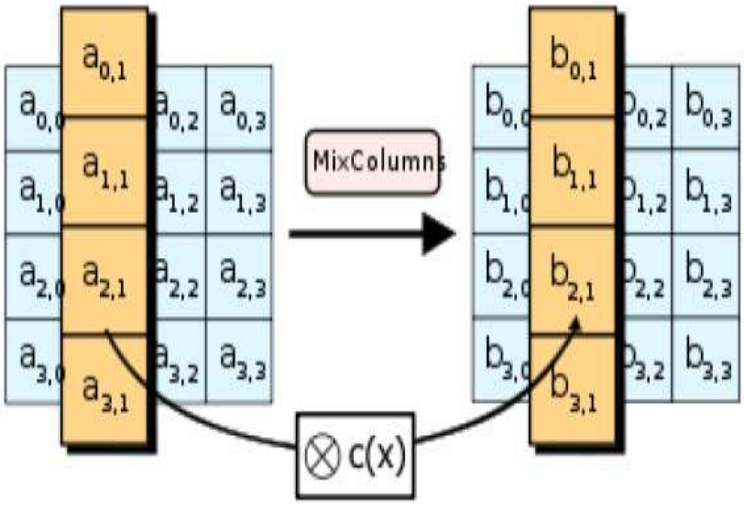

Fig. 4 Mix Columns operates on the state column by column.

Add Round Key () Transformation

In the Add Round Key () transformation, a Round Key is added to the State by a simple bitwise XOR operation. Each Round Key consists of Nb words. Those Nb words are each added into the columns of the State, such that

$$[S'_{0f}, S'_{1f}, S'_{2f}, S'_{3f},] = \left[S_{0f}, S_{1f}, S_{2f}, S_{3f} \oplus [W_{round \cdot Nb+c}]\right] for\ 0 \leq c < Nb, \quad (7)$$

Where [wi] are the key schedule words, and round is a value in the range 0 £ round £ Nr. In the Cipher, the initial Round Key addition occurs when round = 0, prior to the first application of the round function.

The application of the Add Round Key () transformation to the Nr rounds of the Cipher occurs when 1 £ round £ Nr.

V. RESULT

The presented de-duplication system based on SHA -512, it's consume less time and give higher hash output as compare to other previous methods. In the below table 1, shows the comparison of proposed de- duplication finding time as compare to de-duplication finding time MD -5, SHA- 1, SHA- 256 and SHA – 256 and SHA 512. In the below figure and table clearly shows that proposed method shows better result as compare to other previous methods. Also shows in the figure 2, compare in bar graph.

Table 1: Compare with different previous methods

File Size / Methods	Time (mile second)			
	MD 5	SHA1[1] [3]	SHA 256	SHA 512
5 kb	2.086	1.854	1.55	0.781
10 kb	2.085	1.901	1.606	0.925
20 kb	2.226	1.976	1.734	0.955
30 kb	2.141	2.269	2.045	1.083
40 kb	2.139	2.098	2.975	1.404
50kb	2.951	3.348	2.988	1.731

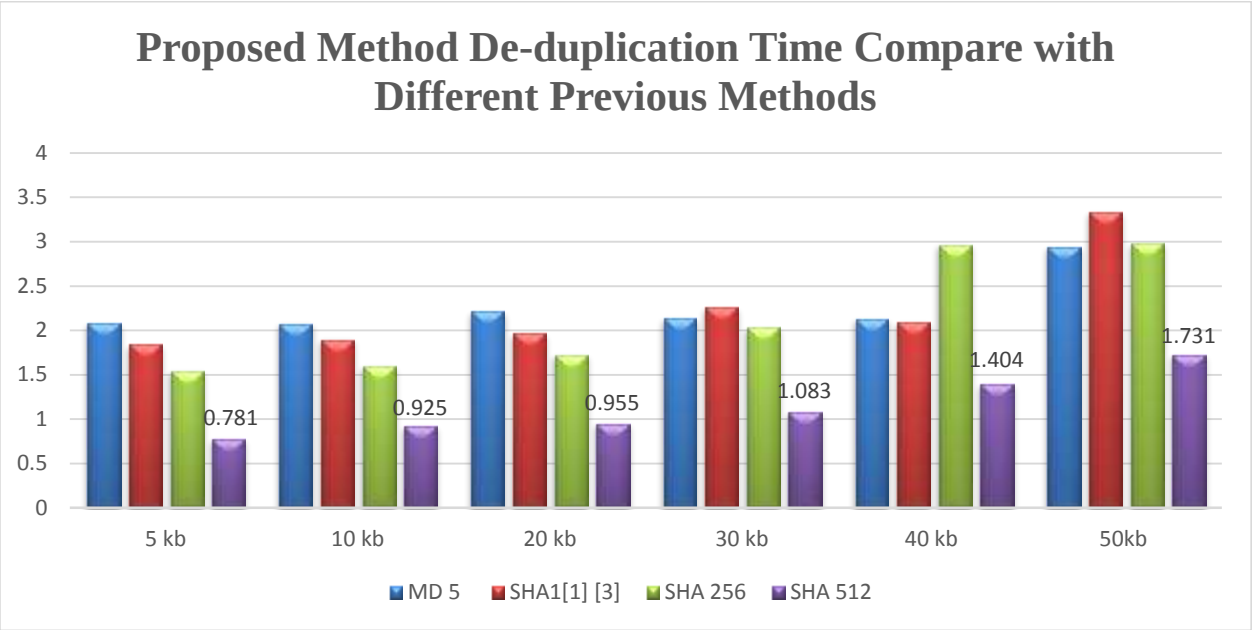

Figure 2: Graphical Comparison of proposed work in different previous work

IV. CONCLUSION

Cloud computing is an emerging area of research, where most of the IT infrastructure is moving to make their service and delivery more efficient. Cloud computing make it more scalable, more reliable, secure and accessible with plenty of option to perform its best. In this Dissertation our work approach lead behind the multiple copy and duplicate data get upload over the cloud and its different data centre due to different ownership. The concept behind the research is taken a secure and reliable algorithm, approach which can find the solution for security as well as de-duplication redundancy optimization over the data store.

The proposed base method discussed about the file level distribution and redundancy detection using file level chunking, where as to transmit and store the data AES (Asymmetric encryption system) algorithm is used to provide data security. For improvement of the hash calculation use SHA – 512, with the help of SHA -2 obtain duplicate file detection faster as compare to other methods. In future try to encrypted this image using the above methods also add impulse noise as an attack [11] [12] [13]. In future implement the proposed method on real time communication system such as LTE, TD LTE. [14]

REFERENCES

[1] Hur, Junbeom, Dongyoung Koo_, Youngjoo Shinz and Kyungtae Kangx, "Secure Data Deduplication with Dynamic Ownership Management in Cloud Storage (Extended Abstract)", IEEE 33rd International Conference on Data Engineering, 2017.

[2] Deshmukh, Ankush R., R. V. Mante, and P. N. Chatur. "Cloud Based Deduplication and Self Data Destruction." *2017 International Conference on Recent Trends in Electrical, Electronics and Computing Technologies (ICRTEECT)*. IEEE, 2017.

[3] Hur, Junbeom, et al. "Secure data deduplication with dynamic ownership management in cloud storage." *IEEE Transactions on knowledge and data engineering* 28.11 (2016): 3113-3125.

[4] Armknecht, Frederik, Jens-Matthias Bohli, Ghassan O. Karame, and Franck Youssef. "Transparent data deduplication in the cloud." In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pp. 886-900. ACM, 2015.

[5] Yang, Kan, and Xiaohua Jia. "TSAS: third-party storage auditing service." *Security for Cloud Storage Systems*. Springer, New York, NY, 2014. 7-37.

[6] Bellare, Mihir, Sriram Keelveedhi, and Thomas Ristenpart. "DupLESS: Server-Aided Encryption for Deduplicated Storage." *IACR Cryptology ePrint Archive* 2013 (2013): 429.

[7] Yinjin, Fu, Xiao Nong, and Liu Fang. "Research and development on key techniques of data deduplication [j]." *Journal of Computer Research and Development* 1 (2012): 002.

[8] Bose, Sumit Kumar, et al. "Cloud-Spider: Combining replication with scheduling for optimizing live migration of virtual machines across wide area networks." *Cluster, Cloud and Grid Computing (CCGrid), 2011 11th IEEE/ACM International Symposium on*. IEEE, 2011.

[9] Xiong, Jin, et al. "Metadata distribution and consistency techniques for large-scale cluster file systems." *IEEE Transactions on Parallel and Distributed Systems* 22.5 (2011): 803-816.

[10] Bhagwat, Deepavali, et al. "Extreme binning: Scalable, parallel deduplication for chunk-based file backup." *Modeling, Analysis & Simulation of Computer and Telecommunication Systems, 2009. MASCOTS'09. IEEE International Symposium on*. IEEE, 2009.

[11] Pranay Yadav, "Color Image Noise Removal by Modified Adaptive Threshold Median Filter for RVIN" , Electronic Design, Computer Networks & Automated Verification (EDCAV), *2015 International Conference on National Institution of Technology (NIT - Shilog) Conference, pp - 175 - 180*, 29-30, DOI, 10.1109/EDCAV.2015.7060562, Jan. 2015.

[12] P. Yadav and Parool Singh, "Color Impulse Noise Removal by Modified Alpha Trimmed Median Mean Filter for FVIN", *IEEE International Conference on Computational Intelligence and Computing (IEEE-ICCIC) in PARK College of Engineering and Tekhnology, Coimbatore-641659*, pp: 1 – 8, Dec – 2014.

[13] Sharma, S. and Yadav, P, "Removal of Fixed Valued Impulse Noise by Improved Trimmed Mean Median Filter" IEEE International Conference on Computational Intelligence and Computing (IEEE-ICCIC) in PARK College of *Engineering and Tekhnology, Coimbatore-641659, Tamilnadu, (IEEE-ICCIC)*, pp: 1 - 8, Dec 2014.

[14] **Yadav P.,** Sharma S., Tiwari P., Dey N., Ashour A.S., Nguyen G.N. "A Modified Hybrid Structure for Next Generation Super High Speed Communication using TDLTE and Wi-Max" accepted for publication in *Studies in Big Data, Springer Book Chapter* 2017.