International Journal of Electrical Electronics & Computer Science Engineering
Volume 4, Issue 3 (June, 2017) | E-ISSN : 2348-2273 | P-ISSN : 2454-1222
Available Online at www.ijeecse.com

Analysis of Disks, Relational and Non-Relational Databases for forensics Investigations

Shagufta Praveen[1], Dr. Umesh Chandra[2]
[1]Research Scholar, [2]Assistant Professor, Computer Science Department, Glocal University. Saharanpur, Uttar Pradesh, India
[1]shaguftasaeed125@gmail.com,[2]uck.iitr@gmail.com

Abstract: ***The main objective of this paper is to represent analysis of various databases for forensics Investigation. It mainly discuss about file system, relational and Non-relational database. For File system, data analysis and recovery of data is discussed where FAT and NTFS both are mentioned. For relational database, MySQL database is used to analyze data from different schemas. For Non-relational Database, HBase a NoSQL product is used that describes analysis of data with the help of external schema and database structure. Different areas of databases are mentioned and analyzed from where investigation of data can be done. This paper concludes different aspects that should be touched by investigator during database investigation and those aspects are discussed here with results.***

Keywords: Database Forensics, Disks, Non- Relational Database, Relational Database.

I. INTRODUCTION

We are living in digital world where every task is done by computer, mobiles, e-mails and laptops. It is easier to believe that many of the criminal cases are done with the help of same digital gadgets. Not only in cyber crimes but also in general crimes investigators try to scan all the digital objects present at the time of the event. Digital forensics includes Computer examination, Data analysis, database study, Mobile devices analysis, Network analysis, Photography analysis, video analysis and audio analysis[1]. Among all, database study is least touched topic which has a few research papers and a single book published [2]. The main aim of this paper is to focus on how data can be investigated from different databases and to highlight that most of the confidential data is not even seen by investigators while investigation due to lack of internal structure knowledge of various databases. This lack of knowledge becomes a useful technique for criminals to hide their data.

Steps involved while doing investigation are:

Investigation is performed to collect evidences. These evidences are only proof that are presented before court in favor or in against of a claim. Evidence can be any logical data or content but for addressing a file, record, content to be a proof, it's really important for an investigator to properly classified, collect, extract and documents the data. Figure 1 discusses steps involved while doing investigation.

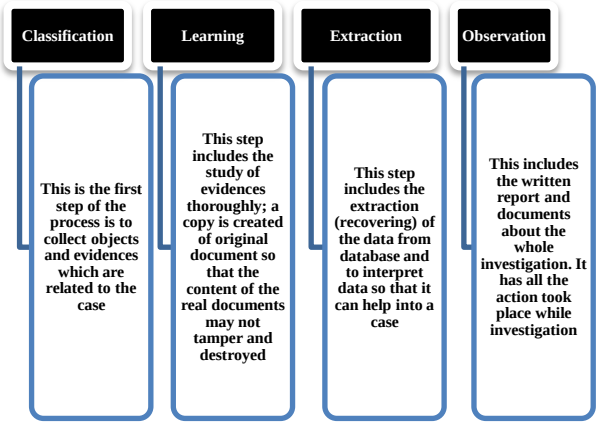


Fig 1: Steps Involved During Investigation

II. DISK FOR FORENSICS ANALYSIS

Disk is one of the most common devices to store data. The method that is used to store data on a disk is called file system. File system mainly control all tracking of files from different location of disk. Different aspects of file system are [3]: space management, files name, directories, Metadata but useful aspects in forensics during investigation of disk are [4]: Content [5], Metadata [6, 7], recovery of information[7] and reconstruction[8]. Table1 represents all the useful aspects in forensics.

Table 1: Useful Aspects in Forensics

S.No.	Aspects	Reasons
1	Content	To search evidences at physical level
2	Metadata	Complete information about data (Name of File, Time and date of created file , Owner of File, position on storage, etc)
3	Reconstruction	Reconstruct data with the help of reverse query and relational algebra manipulated while doing a criminal act.
4	Recovery of information	To get back the lost information which actually exist but not visible physically

International Journal of Electrical Electronics & Computer Science Engineering
Volume 4, Issue 3 (June, 2017) | E-ISSN : 2348-2273 | P-ISSN : 2454-1222
Available Online at www.ijeecse.com

A. Various Stages While Disk Investigation:

1) Slack Space:

For learning phase during investigation disk goes through a process called Imaging. Imaging is about making an exact of copy of the disk (including setup, programs, data, file format). This copy is used while investigation so that the real content doesn't effect at all. To avoid write operation on evidence disc Physical write blockers are used [4].This approach of making replica of evidence is called integrity.

For extraction phase, data inside the disk is invisible at times. Some of the data is extract through tools as well as some of the is hide by the culprit so that investigator wont able to reach the proof. The space where data is made to be hidden is called slack space. Figure 2 represents sectors where data are stored in file systems

Sector 1	Sector 2	Sector 3	Sector 4	Sector 5	Sector 6

Fig 2. Sectors of The File

An Operating System allocates sector to a file when it writes a file in disk. Sectors allocation is determined by two things first as per the limitation of operating system and another decision by system administrator [9].

So for an example a File G allocated to sector 1 to sector 6 in Fig.3.

File G	File G	File G	File G	File G	File G

Fig.3. Sectors Allocated with File G

After deletion of File G, another File H allocated to sectors in Fig.4. Deletion unallocated the sectors and made them empty.

File G deleted	File G deleted	File G deleted	File G deleted	File G deleted	File G deleted

Fig.4. File G Unallocated from Sectors

File H	File H	File H	File H	empty	Empty

Fig.5. Sectors Allocated with File H

The last two sectors seems to be empty in fig.5. but deleted G file is not deleted actually it remain in the leftover space and this space is called slack space and criminal thinks the file has got deleted and new file allocated but that's not true. File G remains there but in slack space which appeared to be empty but has file content.

2) Swap Space:

There is another space called swap space which also contain sensitive data. Some of the operating systems like windows and Linux swap their data from RAM to disk as represented in Fig.6.Virtual memory concept moves data so that free memory can be achieved in RAM, due to which investigator doesn't able to find sensitive and confidential data in RAM.-

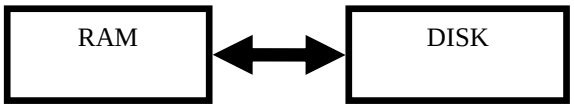


Fig.6. Transfer of Data from RAM to Disk and Vice versa

3) During Hibernation Mode:

Another way to hide important data is, when a system goes into hibernation state then content of the RAM moves to hibernated files shown in Fig.7. During investigation RAM doesn't contain data.

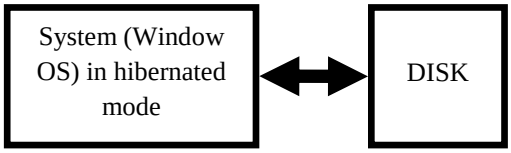


Fig.7. Transfer of Data from RAM to Disk while Hibernation

B. File Carving and Disk:

File carving is a technique to recover deleted data from disk with the help of file structure like file tables. Meta-data is not useful in file carving whereas file structure plays main role for this technique [10]. Different file systems have different file structures and these file systems organize data on the drive. Every file system has specific technique to store data on a particular area of the drive this area is called cluster and assigning of data to this particular area is called allocation [11]. A Cluster is a collection or group of different sectors (it is a smallest unit of space on hard disk which consist of 512 bytes).Most common file systems used today are FAT-32 and NTFS. Windows 95, 98, Linux uses FAT-32 whereas Window XP, 2K and NT uses NTFS.Table.2. represents the comparison between NTFS and FAT files.

Table 2: Comparisons Between FAT and NTFS

Attribute	FAT	NTFS
Created by	Bill gates and Marc McDonald	Microsoft and IBM
Year	1977	Mid of 1980
Directory Content	File Table	B+ Tree
File Allocation	Link List	Bit Map

International Journal of Electrical Electronics & Computer Science Engineering
Volume 4, Issue 3 (June, 2017) | E-ISSN : 2348-2273 | P-ISSN : 2454-1222
Available Online at www.ijeecse.com

Performance	N/2 (N=number of files) N are items and N=1000 , time taken to search is 1000/2 =500 seconds	Log N N are items and N=1000 , time taken to search is Log 1000, time taken to search = 3
Encryption	No	Yes
Compression	No	Yes
Size	4GB	16TB
Conversion	Yes	No
Fault Tolerance	No	Auto Repair

File Allocation Table	
Cluster Number	**Next Number**
0	
100	EOF
200	600
300	500
400	EOF
500	EOF
600	300
700	0

Directory Table		
File name	**Starting Cluster**	**Meta Data**
File.txt	200	

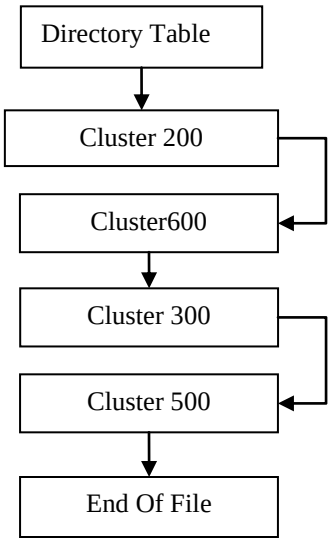


Fig 8. File Allocation Table

C. FAT During Allocation:

Each Directory table has name, starting cluster number and metadata about the file as shown in Fig.8. All the sequence of clusters having content of the file 'file.txt' is connected by link list (as per File allocation table) from starting block to end of file. After Allocation, Data disk area has content of file.txt in the clusters with their respective cluster number such as, Cluster 200=File.txt, Cluster 600=File.txt, Cluster 300= File.txt, Cluster 500= File.txt [10]

After Deletion of file File.txt

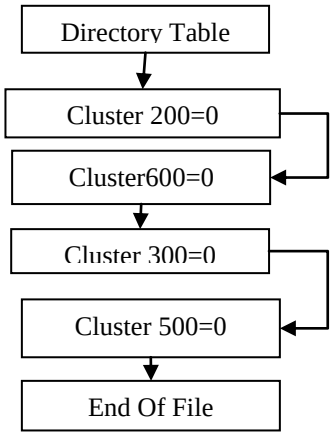


Fig .9. Table on Deletion

After deletion allocated cluster will become unallocated and in directory table connected cluster sequence will be marked 0 and space will be marked unallocated but data disk area will remain there with the content Cluster 200= File.txt, Cluster 600=File.txt, Cluster 300= File.txt, Cluster 500= File.txt as shown in fig.9. This data will help investigator to extract deleted file. Data will be deleted apparently but it actually remains there in data disk area. In NTFS, Allocation of file to clusters is represents through Bitmap. In this cluster assigned a value '1' if file exists otherwise it assigned a value '0' on deletion. On deletion of file, file cluster link doesn't delete. It remain there hence, recovery of data in NTFS is easier [10].

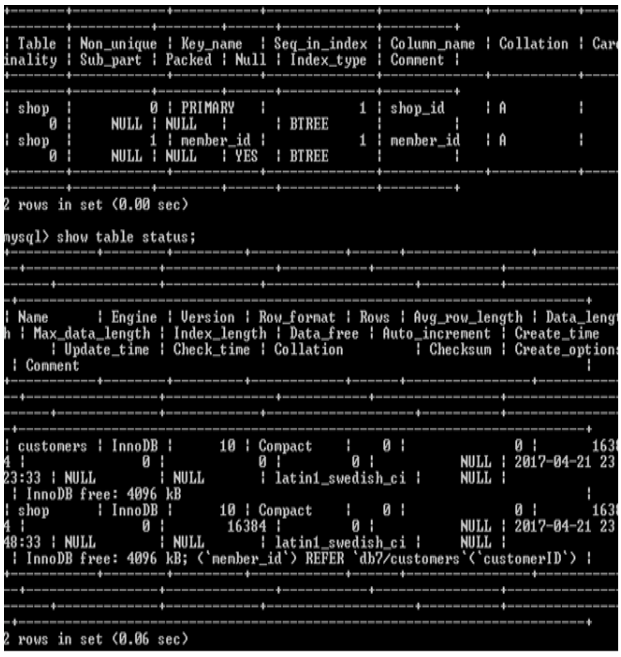
III. RELATIONAL DATABASE AND FORENSICS

Relational database model is on the renowned model used for transactions. Organization mainly banks and transactions related to money are made consistent and secure with relational database concept. SQL is a standard language for Relational Database management system and most of the relational database management system like MySQL, Oracle, MS Access, Sybase, Informix, SQL Server use SQL as their standard language [12].

A. Analysis on Basis of Schema in RDBMS:

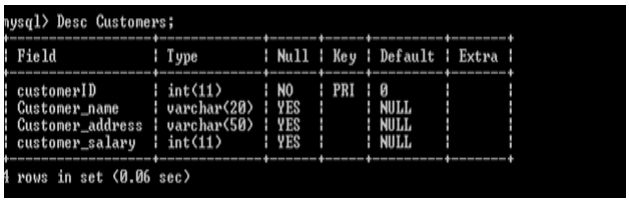
RDBMS has three schema architecture: Internal schema, Conceptual schema, External Schema.

International Journal of Electrical Electronics & Computer Science Engineering
Volume 4, Issue 3 (June, 2017) | E-ISSN : 2348-2273 | P-ISSN : 2454-1222
Available Online at www.ijeecse.com



```
| Table | Non_unique | Key_name  | Seq_in_index | Column_name | Collation | Car
inality | Sub_part | Packed | Null | Index_type | Comment |

| shop  |          0 | PRIMARY   |            1 | shop_id     | A         |
     0 |     NULL | NULL   |      | BTREE      |         |
| shop  |          1 | nember_id |            1 | nember_id   | A         |
     0 |     NULL | NULL   | YES  | BTREE      |         |

2 rows in set (0.00 sec)

mysql> show table status;

| Name      | Engine | Version | Row_format | Rows | Avg_row_length | Data_leng
h | Max_data_length | Index_length | Data_free | Auto_increment | Create_time
      | Update_time | Check_time | Collation         | Checksum | Create_option
 | Comment                                                                    |

| customers | InnoDB |      10 | Compact    |    0 |              0 |      163
4 |               0 |            0 |         0 |           NULL | 2017-04-21 23
23:33 | NULL        | NULL       | latin1_swedish_ci |     NULL |
 | InnoDB free: 4096 kB                                                       |
| shop      | InnoDB |      10 | Compact    |    0 |              0 |      163
4 |               0 |        16384 |         0 |           NULL | 2017-04-21 23
48:33 | NULL        | NULL       | latin1_swedish_ci |     NULL |
 | InnoDB free: 4096 kB; (`nember_id`) REFER `db7/customers`(`customerID`) |

2 rows in set (0.06 sec)
```

Fig.10. Represent Index, Engine and Status of a Table

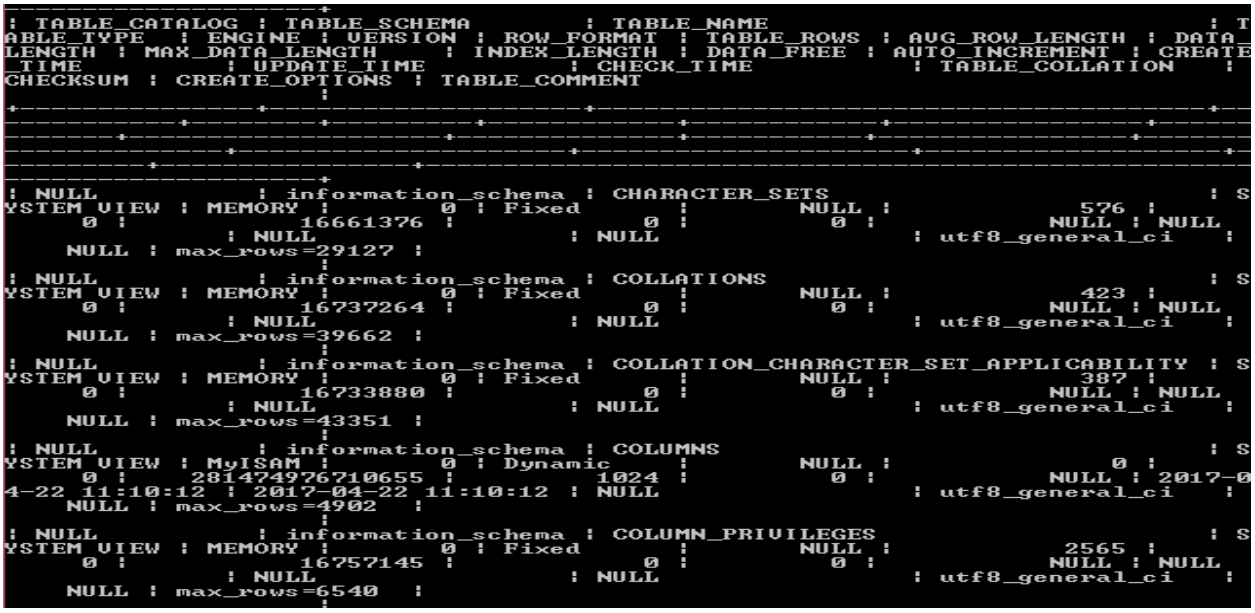


```
mysql> Desc Customers;
| Field            | Type        | Null | Key | Default | Extra |
| customerID       | int(11)     | NO   | PRI | 0       |       |
| Customer_name    | varchar(20) | YES  |     | NULL    |       |
| Customer_address | varchar(50) | YES  |     | NULL    |       |
| customer_salary  | int(11)     | YES  |     | NULL    |       |
4 rows in set (0.06 sec)
```

Fig.11. Represent External Schema

The above snapshot is of conceptual schema of RDBMS 'MySQL' which represents various fields (Name, engine, version, format, update time, create time, cardinality, index type, index length, Data free, Fields, their types) that makes investigator comes to know about database. This helps investigator to know about the database and related data presented in a table of the organization.

In figure 10,' index_type is BTree' that represents that data stored in internal schema is in B Tree form. Every file system has different structures, FAT use Link List whereas NTFS use BTree. So, it's easy to guess the file system for this table is NTFS. Whereas, fig.11. gives us a figure of external schema.



```
| TABLE_CATALOG | TABLE_SCHEMA       | TABLE_NAME                                 | T
ABLE_TYPE  | ENGINE | VERSION | ROW_FORMAT | TABLE_ROWS | AVG_ROW_LENGTH | DATA_
LENGTH | MAX_DATA_LENGTH    | INDEX_LENGTH | DATA_FREE | AUTO_INCREMENT | CREATE
_TIME      | UPDATE_TIME | CHECK_TIME | TABLE_COLLATION   |
CHECKSUM | CREATE_OPTIONS | TABLE_COMMENT

| NULL          | information_schema | CHARACTER_SETS                             | S
YSTEM VIEW | MEMORY |      0 | Fixed      |          NULL |            576 |
     0 |          16661376 |            0 |         0 |           NULL | NULL
               | NULL                | utf8_general_ci   |
    NULL | max_rows=29127 |

| NULL          | information_schema | COLLATIONS                                 | S
YSTEM VIEW | MEMORY |      0 | Fixed      |          NULL |            423 |
     0 |          16737264 |            0 |         0 |           NULL | NULL
               | NULL                | utf8_general_ci   |
    NULL | max_rows=39662 |

| NULL          | information_schema | COLLATION_CHARACTER_SET_APPLICABILITY      | S
YSTEM VIEW | MEMORY |      0 | Fixed      |          NULL |            387 |
     0 |          16733880 |            0 |         0 |           NULL | NULL
               | NULL                | utf8_general_ci   |
    NULL | max_rows=43351 |

| NULL          | information_schema | COLUMNS                                    |
YSTEM VIEW | MyISAM |      0 | Dynamic    |          NULL |              0 |
     0 |  281474976710655 |         1024 |         0 |           NULL | 2017-0
4-22 11:10:12 | 2017-04-22 11:10:12 | NULL                | utf8_general_ci   |
    NULL | max_rows=4902  |

| NULL          | information_schema | COLUMN_PRIVILEGES                          | S
YSTEM VIEW | MEMORY |      0 | Fixed      |          NULL |           2565 |
     0 |          16757145 |            0 |         0 |           NULL | NULL
               | NULL                | utf8_general_ci   |
    NULL | max_rows=6540  |
```

Fig.12. Represent Meta Data of the Table

In Figure 12, represents the metadata of the table which tells much more about the table create_time, update_time, check_time, data_free.The most interesting thing about metadata is it remains there even after deletion of the table and its respective database.

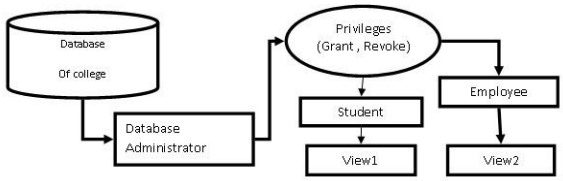


Fig .13. External Schema of Database

International Journal of Electrical Electronics & Computer Science Engineering
Volume 4, Issue 3 (June, 2017) | E-ISSN : 2348-2273 | P-ISSN : 2454-1222
Available Online at www.ijeecse.com

This figure 13 represents the privileges given by DBA to different members those who are accessing the database. That's help investigator to know what are the operations assigned to different members dealing with database. This can be done through some specific queries:

Grant Priv_name on Obj_name to {User_name|Public|Role_name};

Where Priv_name= Access rights, Obj_name=Database object,User_name= Name of User

B. Reconstruction of Data for Investigation

Relational algebra is a non procedural language that expresses queries of RDBMS in form of variables and formulas. Inverse relational algebra i.e. inverse of relational algebra can be used in reconstruction of database in forensics field. The inverse operators of the relational algebra can extract the value of attribute A of a tuple in relation R at time t.$\mathbf{Q^{-1}(Q(R_t)) = R_t^*}$ Inverse of a query Q of a current relation Rt until R_t^* until R_t^* relationship is achieved [13].

C. Software Vulnerability : A Door for Crime

At times vulnerability in software can also be the reason for the lost and manipulation of data. One of the famous RDBMS also had the same issue. Version of Oracle before 10g had various default accounts and passwords. As per "Internet security threat Report" published around 168 security threats in 2006 [14].

IV. NON-RELATIONAL DATABASE AND FORENSICS

Hbase, a non relational database belongs to column oriented family of NoSQL. The data base layer of Hbase is above HDFS so that HDFS can work as data store for Hbase. Hbase works with Hadoop (HDFS and Map reduce) to achieve data storage. Fig.14 shows the sequence of layers that has HBase and HDFS layers.

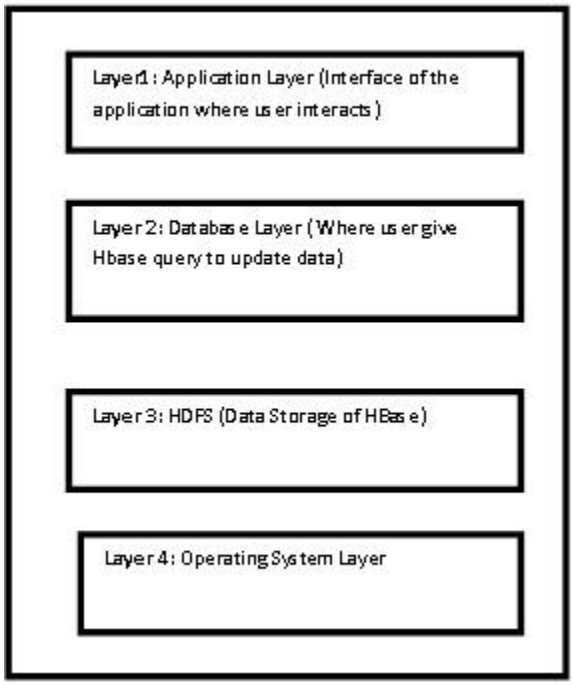


Fig:14. Sequence of Layers of an Application where Database is Hbase and HDFS is Data Storage

A. Architecture of HBase:

Architecture of HBase is shown in fig.15 that represents a row Id connected to several columns this how scalability improves in HBase than conventional databases. In Table.3. different aspects of HBase are also discussed with description.

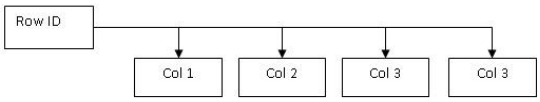


Fig.15. Relation of Row to Several Columns

Table.3. Different Elements of HBase

Aspects	**Description**
Row Key	Unique Number to identify a column
Column Qualifier	Name of a particular column
Column Family	Complete collection of columns

Row and Columns in Hbase

ROW ID -001				
Col 1-01	Col 2-02	Col3-03	Col4-04	Col 5-05
ROW ID-002				

International Journal of Electrical Electronics & Computer Science Engineering
Volume 4, Issue 3 (June, 2017) | E-ISSN : 2348-2273 | P-ISSN : 2454-1222
Available Online at www.ijeecse.com

COL1-06	Col 2-07	Col3-08	Col4-09	Col 5-10
ROW ID-003				
COL1-11	Col 2-12	Col3-13	Col4-14	Col 5-15
ROW ID-004				
COL1-16	Col 2-17	Col3-18	Col4-19	Col n-mn

Fig.16. Rows and columns in HBase

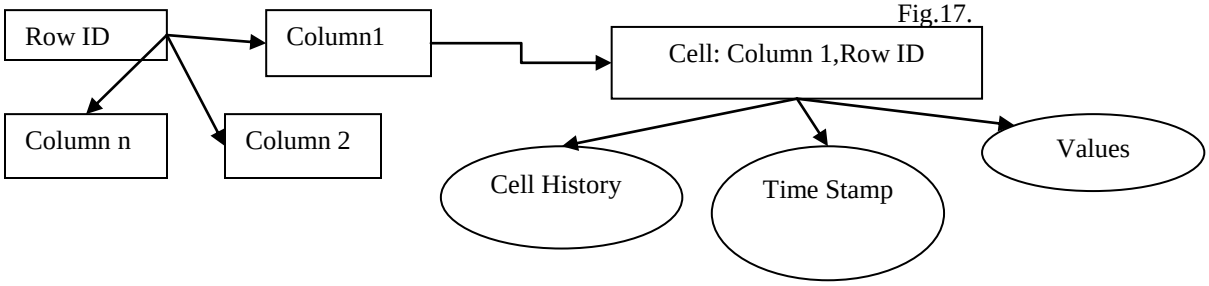


Fig.17.

Representing Cell of a Column

Structure of HBase is in row and columns form that can be seen by fig.16. A Row Id can have billions of columns. Each cell of the table has its own cell history, time stamps and values. Each cell can have multiple values with different time stamps. This way a row can have multiple values this can be seen in fig. 17.

The language that is used to load data from Hbase is Hbase Shell that can also express the status of Hbase that describes number of live servers, dead servers and average load. Schema of the Hbase can also be extracted with the help of Hbase Shell (Name, Data_Block_encoding, verson, Replication_Scope, In_memory,Compression,Encode_on_disk,Block_size). Metadata information stored Hbase in –Root-(list of Meta data) and .Meta Files (list of user region). HBase also keep preparing backup of all data having in a table and these backups can be collect by HBase command [15].

From relational to non-relational, most of the development and changes have been found in databases[19] and due to this these various database has different ways of investigation. Table 4 concludes various aspects of investigation of different databases.

Table.4. Content About Different Database

S.N.	Aspects	Content		
		File System	**RDBMS (MySQL)**	**NoSQL (HBase)**
1	Metadata for Forensics	Yes Useful	Yes useful	No (Not Useful) **Reason:** Metadata has only information about data region and location of data blocks not much useful for investigation [15]
2	Extraction of Deleted Information	Slack Space	Information schema and log files	Hbase Configuration files and Hbase Log File [15]
3	Data stored in	Disk (Sectors or clusters)	Disk (FAT (Link List), NTFS(B-tree))	HDFS (Follow Map reduce due to large amount of data)
4	Software Vulnerability	Yes [18]	Yes[17] (Oracle 10g had weak configuration in release 1)	Yes [16]

5	Preservation of evidence	Imaging and MD5	Creating Back ups	Back up Mechanism
6	Layer of schema meant to be investigate most	File structure	External, internal and Conceptual	Database Layer and HDFS

V. CONCLUSION AND DISCUSSION

Database forensics is one of the least touched topics among various fields present in today's scenario. This paper demonstrates that deleted files in the databases are not actually found as deleted and this proves to be a big help to investigators in finding clues in a criminal case. But today we are dealing with different databases and every database has different selective area from where investigation becomes fruitful. This paper represents different areas and various approaches to analyze database for forensics purpose. Big Data is an emergence topic today; this topic can also be big help to database forensics. Data declared as a dead can be extracted through mining for forensics purposes like US government analyzes Data in TB everyday in order to seek knowledge for the better treat of their national security. Though there could be many challenges like scalability, data management, Analysis of large amount of complex and heterogeneous data but big data can also add some more to forensics field.

VI. REFERENCES

[1] https://en.wikipedia.org/wiki/Database_forensics

[2] Oracle Forensics ISBN 0-9776715-2-6 (May 2008)

[3] https://en.wikipedia.org/wiki/File_system

[4] M.S. Olivier, 2009, On Metadata Context in Database Forensics, ICSA Research Group, Computer Science, University of Pretoria, South Africa.

[5] N. L. Beebe, J. G. (Clark, 2007), Digital forensic text string searching: Improving information retrieval effectiveness by thematically clustering search results, Digital Investigation 4 (Supplement 1) 49–54.

[6] F. Buchholz, E. Spafford, (2004) On the role of file system metadata in digital forensics, Digital Investigation 1 (4) 298–309.

[7] K. Eckstein, (2004) Forensics for advanced UNIX file systems, in: Proceedings from the Fifth Annual IEEE SMC Information Assurance Workshop, IEEE, pp. 377–385.

[8] S. L. Garfinkel, (2007) Carving contiguous and fragmented files with fast object validation, Digital Investigation 4 (Supplement 1) 2–12.

[9] http://www.techrepublic.com/blog/it-security/computer-forensics-finding-hidden-data/By Tom Olzak | in IT Security, May 21, 2007, 1:16 AM PST

[10] P. Anandabrata , and M. Nasir, 2009 The Evolution of File carving, IEEE Signal processing Magazine.

[11] http://www.dewassoc.com/kbase/hard_drives/clusters.html.

[12] SQL by www.tutorial point.com , 2017, tutorial point pvt ltd.

[13] O. Mary and M. Olive, Reconstruction in database forensics, International federation for information processing, 2012.

[14] D. Litchfield , Oracle forensics: how Attacker break in, Chapter 3.

[15] J. Sremack, 2015,Packt Publishing ,Big Data Forensics- Learning Hadoop Investigation.

[16] https://secuniaresearch.flexerasoftware.com/community/advisories/64655

[17] http://www.infoworld.com/article/3119120/security/mysql-zero-day-exploit-puts-some-servers-at-risk-of-hacking.html

[18] https://www.dataretrieval.com/fat-structure-analysis.html

[19] S. Praveen, U. Chandra, 2017, A review literature on evolving database, International Journal of Computer Applications.