Indian Journal of Science and Technology, Vol 10(30), DOI: 10.17485/ijst/2017/v10i30/115510, August 2017
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Application of ETLR in Telecom Domain

Sachin Sharma[1*], Sandip Kumar Goyal[1] and Kamal Kumar[2]

[1]Department of Computer Engineering, M. M. Engineering College, M. M. University, Ambala – 133207, Haryana, India; er.sachinsharma@gmail.com

[2]School of Computer Science and Engineering (SoCSE), UPES, Dehradun – 248007, Uttarakhand, India; kkumar@ddn.upes.ac.in

Abstract

Objectives: To apply ETLR (Extraction, Transformation, Loading and Retrieval) paradigm to build an efficient, effective and cost effective data warehouse for telecom industry. The focus point is to optimize every layer of telecom DWH. **Methods:** The data techniques used are making use of telecom infrastructure, i.e. MSC files and applying segregation logic at the source layer i.e. mediation layer. Files are pushed towards predefined separate destinations and applying multiple technology mix mainly of database inbuilt utilities and custom scripts to avoid use of commercial ETL tools; and at the same time achieving enhanced performance at every front. Technology mix includes source optimization, external table implementation and switching, DB copy utility and retrieval level optimization. We have used data loading statistics to compare the results. **Findings:** The ultimate result is a telecom data warehouse and the result that we have achieved using ETLR paradigm improved the data processing of the data many folds. The motive is to optimize every layer that comes in between the data warehouse building process. Source level optimization leads at the data cleaning at the source level itself, thus shifting the load at the source system and reduced the load on the DWH servers. We have supplied bunch of files to the external tables and thus utilizing the OS storage for tabular data. Transforming data using views and push them into partitioned tables using DB copy utility improved the overall performance. Using query optimization techniques and DB level tuning ensures the data availability in minimum time. The data availability of a standard DWH is sysdate-1; but in our case, we have reduced it to approx. 4 hours with indexes intact. The scalability is also a very strong point of our ETLR paradigm. Now telecom operators have a better system available for building their data warehouse without taking care of heavy license fee for commercial tools. **Application/Improvements:** The application of the paradigm is in mostly every sector where data processing is a big challenge and cost is a major factor. We have given its application in telecom sector in this paper. The same can be implemented in Banking Sector, Insurance Sector, social media etc. and we can put it on cloud also in case hardware is a constraint.

Keywords: Big-Data, ETL, Mobile Data, Retrieval, Scripts, Telecom Sector

1. Introduction

Big sized data is reality today. Most reliable businesses like banking, telecom, e-commerce and aviation sectors which in-fact are service sectors results in generation of huge data at ever increasing rate. The data may be structured, unstructured. Huge servers and clouds often come to rescue in such scenario. Storage is only one aspect of ever increasing data. Extraction, Transformation and Load is always a matter of concern for data scientists. Various ETL tools are available, offering multi-functionalities and customization with advanced technology; still lacking in overall view of the system as far as system building is concerned. ETL can be achieved by the tools or by custom made scripts which depends upon the organization structure and technical capabilities of the team concerned. There are various schema topologies adapted while constructing a Data warehouse e.g. Star Schema, Snowflake schema etc. When we talk about ETL, generally focus has been given on ETL jobs, its scheduling and the ETL streams. Somehow the reporting for analytics which is the ultimate aim of constructing a data

*Author for correspondence

warehouse is ignored, the researchers could not able to focus on the Retrieval part which is an integral part of the ETL process; in fact ETL process should be called ETLR (Extraction, Transformation, Loading and Retrieval)[1]. The framework was proposed in[1] and through this paper, application of ETLR paradigm is explored in telecom domain. Telecom sector is fastest growing service sector and has seen revolutionary break troughs in quick succession spanning over less than last five years. From 2G to now 4G, a massive user base and phenomenal increase in service coverage, mobile data traffic has seen huge growth. Mobile data traffic growth is shown in Figure 1.

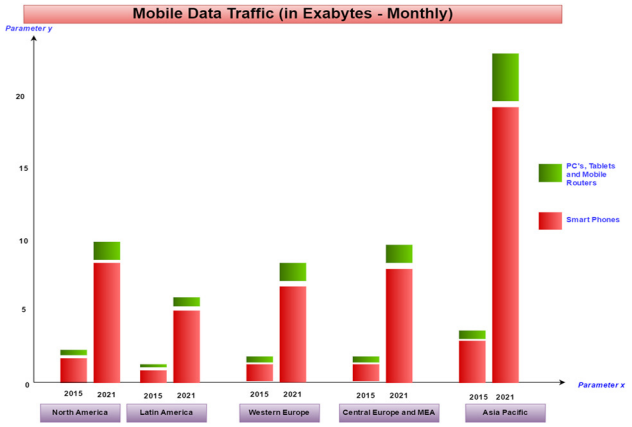

Figure 1. Data growth rate in Telecom[18].

Telecom domain is one of the domains handling huge data volume and handling data in telecom domain is still a big challenge. Also the growth in data volume rate for telecom is among highest in the industry.

Researchers have put in their efforts for ETL optimization, job efficiency and scheduling. In telecom data warehouse, research has been put into the network bandwidth and network availability[2]. Telecom data warehouse contains heterogeneous data sources and extremely complex and different network elements. Every network element support different programming languages and produced different kind of data which needs to be processed. Only network bandwidth optimization and network availability will not be able to achieve the ultimate goal. Performance improvement in the ETL part while migrating data of telecom domain has been lightened and a new technology named DMT has been designed[3]. Data migration is basically one time activity and DMT focused on the data migration of the telecom domain. While telecom data warehouse building process is a continuous task in which ongoing data is important and the data arrival volume is the major concern of the data scientists. DMT can work in the migration of the data but needs a system which can support end to end solution for the telecom data warehouse. In telecom billing migration, we have mostly postpaid customers so when we talk about telecom billing data migration then it refers to the data migration of the postpaid customers only for which solution is given in[3]. In current scenario; maximum data volume is of prepaid customers for which billing is not required as real time charging mechanism is already in place in the form of Telecom IN (Intelligent Network). The data migration technology DMT[3] is only for limited data volume and particular for migration purpose only. A universal ETL system architecture has been put in place for decision support system which facilitates ETL and monitoring of the overall system with abnormal scenarios handling capabilities[4]. The system proposed in[4] addressed specific organizational needs for Decision Support System and is designed for Hebei Province Netcom Corporation. The view area is very specific viz. source data. Moreover, retrieval part has not touched in the solution provided in[4]. Modeling and optimization of ETL process in the data warehouse is been detailed[5]. The author in[5] review the ETL process execution plan for optimization. ETL tools activities have been analyzed for the optimization whereas discrete solution without the use of any commercial tool and that can fit anywhere has not been come out. Tuning of ETL processes using distributed processing is also explored by the researchers by dividing the complex SQL queries into small subsets and executing the same in parallel distributed environment[6]. Real time data ETL framework has been proposed which process historical data and real time data separately[7]. Author in[8] have worked on the security aspect during extraction of the data in ETL process. We use data segmentation and workflow optimizations in large and complex data warehouse as detailed in[9].

Efforts have been made by the author in[10] for logical ETL optimization and to reduce the overall processing cost of ETL execution cycle. Data processing transition from traditional ETL process to real time ETL process is proposed by the authors in[11] and high level architecture for the partitioning of real time ETL flows is detailed. With the emerging technologies like map reduce, authors have explored the parallel ETL using map reduce technology[12]. Quality oriented approach during ETL process[13] has been bring into the consideration. ETL process without tools is a major concern as all type of use cases needs

Table 1. Fact Table Structure

Column Name	Description	Column Name	Description
TYPE_OF_RECORD	CDR Record Type i.e. Mobile Originating or Terminating etc.	CELL_ID_LAST	First Cell Id of calling Number
A_PARTY	Calling Party	CIRCLE_ORIGINATING	Call Originating Circle
B_PARTY	Called Party	TYPE_OF_SERVICE	Prepaid / Post-paid
DATE_AND_TIME	Time at which call has been matured	CALL_REFERENCE_ NETWORK	Call Reference of Network
CALL_DURATION	Total duration of call	IN_TRUNK	In Trunk
LRN	LRN Number	OUT_TRUNK	Out Trunk
MNPINFO_CALLED_ PARTY	MNP Information of called party	MSC_ID	MSC Identification Number
FILE_NAME	Mediation File Name from which CDR Generated	SMSC_CENTRE	SMSC Centre Identification Number
IMSI	IMSI of Calling Number	SEARCH_NO	Search Number (Calculated from A_Party and B_Party)
IMEI	IMEI of Calling Number	MOBILE_CIRCLE	Circle populated from IMSI
CELL_ID_FIRST	First Cell Id of calling number		

to be handled by custom made scripts but it reduces the overall cost of the project to an extremely large extent and the scripting framework has been proposed in[14] for automating the ETL process. Dynamic ETL processes which will lead to the metadata based process and enhance the concept of reusability in the ETL have been proposed by researchers in[15]. Scalable ETL (SETL) system which aimed at the scalability aspect of the data warehouse has been proposed in[16]. Significant work has been done in the area of land and resources data warehouse and detailed in[17]. Data growth rate in telecom in recent past is as shown in Figure 1.

2. Problem Definition

Telecom data is growing without limitation and there is regulatory and statutory requirement to provide the data to various intelligence agencies, judiciary and various other agencies as prescribed by the government time to time and there is a huge penalty on the telecom operator in case of not providing or providing incomplete data. Generally data required from telecom operators is of the form MOC (Mobile Originating Call), MTC (Mobile Terminating Call), SMS Originating, SMS Terminating, Call forwarding record and SGSN (Serving GPRS Support Node) data. A telecom operator may use same network equipment like Mobile Switching Centers; SMSC's or may use different network equipment. Network equipment's are called network nodes in telecom domain. The data is originating from GSM, CDMA, Wimax, Landline, SMSC, Over the Air (OTA) and many other nodes. Some telecom operators work in centralized manner and some in distributed manner i.e. some operators have central data centers for processing of the information and some have zone wise/area wise data centers which are taking care of the organizational data processing requirements and other technical requirements. Now there is continuous flow of the data day and night and to process that data in data warehouse in a time bound manner while taking care of all constraints and maintaining indexing mechanism for speedy reporting is a big challenge. Different nodes are providing data in different formats and in heterogeneous form. The data of the customer in telecom is called as CDR (Call Data Record) and the data generated from MSC's is called raw CDR which can be processed by using tools/utilities or specialized software and configure them for file conversion and segregation logic. As data processing is dependent on the arrival of the files and there is lag between the originating time and the file generation time. The major pain area is the time taken between the actual Call/SMS/Data generation and the organization's technical capabilities to make the data available to the various

agencies. Organizations spend tremendous money to get it done and highly rely on commercial ETL tools but the data processing time needs to be reduced and retrieval part has to be taken care of. If there can inbuilt retrieval oriented approach while managing data bases centrally or in distributed manner by horizontal, vertical or hybrid fragmentation.

3. Preliminary: ETLR

The concept of Extraction, Transformation, Loading and Retrieval as suggested is a paradigm used for building a data warehouse solution to handle huge volume of data with improved performance at every stage while maintaining the data constraints all the way. ETLR facilitates us the use of various technologies and integrate them in such a way that optimization at every phase can be achieved. The various technologies that have been utilized are a) Source data cleansing b) Concept of external table c) database copy mechanism and query optimization during reporting with the help of partition table with indexes. The various techniques have been integrated in such a way that parallel data loading with parallel reporting will provide a wonderful experience to the technical and end users at very minimum maintenance efforts. The ETLR has provided us the alternate to the costly ETL and reporting tools for building organizations data warehouse and giving the flexibility to customize the solution to any extent as per organizational needs. It gives us the speed much beyond the traditional data loading and ETL tools. The beauty of the ETLR over ETL tools is that it cut off one extra layer that ETL tools used for interfacing with different data sources. ETLR uses the custom made scripts and database utilities itself and reduces the interfacing overhead during the data processing. A simple comparison of conventional loading and external table loading statistics are shown in Figure 2 and 3.

The major difference in performance is due to the database library calling and file read and CPU time incorporated in the conventional loading. In ETLR loading mechanism, multiple files can be shown to external table and thus copy mechanism can be applied at once on the whole bunch which reduces the CPU time as well as the individual file read time. When we are dealing with millions of files at once, this time reduction mechanism matters a lot. In next section, we will detail the ETL paradigm for telecom domain.

```
Table     DATA1:
  26 Rows successfully loaded.
  0 Rows not loaded due to data errors.
  0 Rows not loaded because all WHEN clauses were failed.
  0 Rows not loaded because all fields were null.


Space allocated for bind array:                 255420 bytes(45 rows)
Read   buffer bytes: 1000000

Total logical records skipped:          0
Total logical records read:            26
Total logical records rejected:         0
Total logical records discarded:        0

Run began on Fri Aug 31 12:39:25
Run ended on Fri Aug 31 12:39:25
```

Figure 2. Conventional Data Load Speed.

```
Data Loading with External Table started @ 160816145212
ORACLE_SID=
CONNECTION_STRING=/
 files has not to move

Array fetch/bind size is 5000. (arraysize is 5000)
Will commit after every 2 array binds. (copycommit is 2)
Maximum long size is 80. (long is 80)
   2504074 rows selected from @
   2504074 rows inserted into
   2504074 rows committed into          at @

Data Loading with External Table Finished @ 160816151946
```

Figure 3. ETLR Loading Speed.

4. ETLR in Telecom Domain

Extraction, Transformation, Loading and Retrieval paradigm is best suited in the telecom data warehousing which supports lightning speed of the data processing including transformation and at the same time takes care of the scalability and optimized reporting from the production system. The beauty of the ETLR in telecom domain is that it does make use of the database inbuilt utilities and custom shell programming. Commercial ETL tools need not to be used and thus saving the cost to a large extent. In telecom sector, BSC (Base Station Controller) communicates with MSC (Mobile Switching Centre) and HLR (Home Location Register); thereafter raw CDR files are pushed towards middleware/mediation system to be processed for the various sub domains like billing, rating etc. The MSC captured all the parameters which are then filtered at mediation system for different systems. The data warehousing team of the telecom operator decides the parameters to be taken from mediation for data processing. The provision of the parameters (as suggested by data warehousing team) has to be done

at mediation layer and subsequently the file generation activity completes which in turn pushed towards the predefined destination at the destination system. ETLR mechanism will be incorporated at the staging system where the files have been put in by mediation. Call Data Record file generation process is shown in Figure 4.

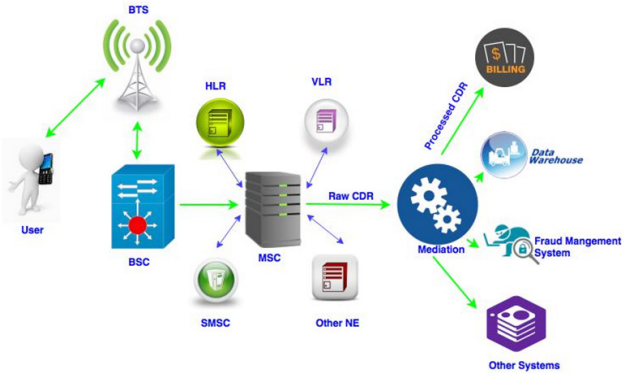

Figure 4. CDR File Generation Process.

As detailed in the Figure 2 when user makes a call or SMS, it interacts with BTS (Base Transceiver Station) which further communicates with BSC (Base Station Controller) and then with MSC (Mobile Switching Centre). MSC is a central high end server which is connected with various NE (Network Elements) like HLR, VLR etc., MSC generates the raw file containing CDR (Call Data Record) of end users which will be supplied to mediation system. Mediation system is used to convert the files and apply logic for various sub systems. Mediation system process the raw CDR files and give the processed CDR files to billing, FMS (Fraud Management System), Data Warehouse and other systems. Billing systems are getting postpaid data only and takes only the fields required for billing purpose only. Similarly other systems are getting the required data from mediation as per the requirements. When it comes at Data warehouse, prepaid as well as postpaid data needs to be processed and also GPRS, 3G data needs to be processed for the reporting purpose. Data volume for data warehouse is very high and CDR files are generating continuously from MSC day and night which needs to be processed in the telecom data warehouse keeping in view the fast retrieval process as well. ETLR (Extraction, Transformation, Loading and Retrieval) paradigm can play an important role while developing a data warehouse system for telecom domain as it covers all the key aspects of the data warehouse beginning from the optimized retrieval mechanism, followed by the parallel transformation, loading and at the same time takes care of the timely retrieval of the data from the production database system. We will discuss the implementation of the ETLR in the next section.

5. Implementation

Where to fit the ETLR paradigm in telecom? Data warehouse system for telecom depends on the files generated by mediation system. The generated files can be comma delimited (CSV), space delimited or other delimited files. A sample processed CDR file record is shown in Figure 5.

The telecom operator may have multiple MSC and all the MSC generates raw CDR files continuously which are further processed by mediation system to generate processed CDR files. Now either data warehouse server needs to pull the files from mediation or mediation system can push the files to data warehouse server. In either of the case, we should apply segregation logic at mediation layer on processed files. Let us understand the concept; a MSC of state A contains maximum CDR's of the state "A" region and few CDR's of the other states of the country. The customer of every state is identified by the IMSI. Now segregation will facilitate the single file coming from MSC to be split into multiple files based on the IMSI series and will be pushed in the respective destination. CDR File segregation based on IMSI is shown in Figure 6.

Files received after segregation from mediation will be pushed to the data warehouse server at the predetermined destination which is separate for separate telecom circles. Now, transformation needs to be applied on the processed CDR files at mediation. We use concept of external table on the data warehouse server. The files placed at OS will be used as tables in the database by making use of database directories concept within the database. The advantage of this technique is that, as and when we get the files at the OS level, it is usable as tabular form in the database. Now, we can write the custom SQL scripts or

```
MTC,97602037XX   ,94667000XX   ,20160204083544 ,000006   ,   ,          ,FTP_MSS2_FF1566                ,4043412000366XX
,357652012949XXX ,04 443006XX52b  ,04 44300XX055b  ,40434,PP,9855d00001,CAMB1I  ,BAMBN1O ,9194662XXXX5                ,
,
```

Figure 5. Sample CDR File record.

views on the external table to transform the raw data to the desired format. If we use the views on the external table, no extra space will be used for database storage. The file system space will be only used and data is available in the view which will be mapped with the production fact table structure of the data warehouse. Fact table structure with column description is shown in Table 1.

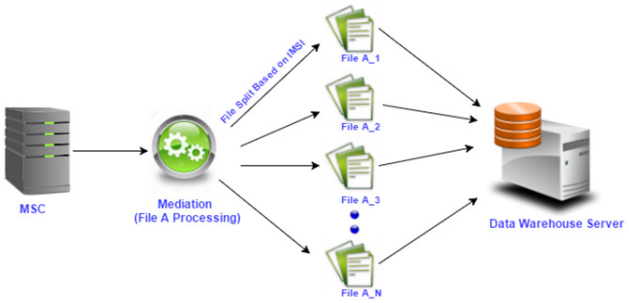

Figure 6. File Segregation based on IMSI at Mediation.

Loading of the data available at staging area in the form of the views will be done by a mix of various technologies and database utilities. The production tables have been designed in such a way that it has partitioned data (range partition) day wise and having global indexes on the three columns. The indexes will be maintained throughout the process to facilitate the faster retrieval process. The concept of loading will work in such a way that a file group has been created which contains number of files that can be defined as per the estimated data volume or file group can be made based upon the number of the CDR contained within the group (whichever is achieved). We call it as bunch of the CDR and that bunch are created real time and circle wise bunch will reflect the data within external table at a time and subsequently view on the external table transforms the external table data in the data suited for the production table.

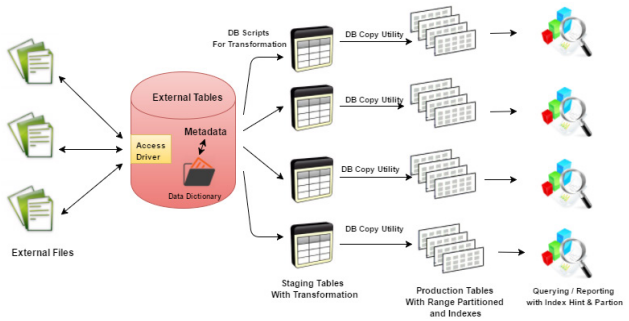

Figure 7. Detailed ETLR Loading Mechanism in Telecom Data warehouse.

The switching of bunches are frequent and is operated in parallel mode for various circles at a time. Now the "COPY" mechanism of the database will be used to push the data into the production database in the parallel mode. The data can be pushed into the fact table directly without dropping indexing mechanism on the tables. Secondly, data can be pushed in the multi tables identical to the production table and partition exchange mechanism can be used to sandwich the multi table data into the fact table within fraction of the minute. The overall mechanism is tuned in such a way that multiple streams of the above mentioned process is executing in distributed manner. When indexes are intact, the data availability is extremely fast and nearly real time. During retrieval, the optimized mechanism is used in such a way that first data will be identified based upon the range partition of the database so that the search of the required criteria can be narrowed. Further index hint will be used at the time of querying data so two ways optimization will result into faster retrieval of the information from the tables of terra bytes in size.

6. Result and Discussion

The implementation of the ETLR paradigm in telecom domain provides us a better alternative for building telecom data warehouse with improved performance at every stage of the process. The cost factor is also reduced many folds because of the use of the inbuilt utilities and not using commercial ETL tools. When we closely examine the overall system, the system is scalable enough to handle organizational needs over the period of time without having much investment in maintenance. In addition to it, we have implemented a uniform mechanism which can be adapted by any organization and is highly reliable and efficient at every stage. Database techniques and utilities like external table, bunch mechanism for CDR, partitioned tables and copy command have been integrated in such a way that boost overall performance of the system. ETLR data warehouse mechanism for telecom is shown in Figure 7.

On adapting the ETLR in telecom domain, we have improved the speed of the overall ETL process and also focused on the Retrieval mechanism. The sample logs are of single run, there are multiple runs at the same time of similar type of the loading which are loading data with indexes intact on the fact tables so as to facilitate the

faster retrieval mechanism as already discussed. We can add multiple bunches/batches to load the CDR into the fact tables as per the requirement which make the system scalable. It has been observed that when we are handling multiple files per day (in millions), traditional data loading terminologies does not work as files read write time will lead to a huge time delay in the overall process. A new way of the loading with inbuilt utilities is a mix of technologies a) External Tables b) Range Partitioned Tables and c) copy utility to load data into fact tables. The resultant paradigm for telecom domain data warehouse is best suited. Multiple files are handled using external tables, parallel view executions on external tables to map the raw table data with fact table data, parallel threads execution in the form of bunches utilizing the hardware capabilities of the system for lightning loading. The sample log file for one bunch is shown in Figure 8 and a sample report output from the data warehouse is shown in Figure 9.

```
Data Loading With External Table Started @170616205222
ORACLE_SID=XXXXXXX
Array fetch/bind size is 5000. (arraysize is 5000)
Will commit after every 2 array binds. (copycommit is 2)
Maximum long size is 80. (long is 80)
2502784 rows selected from @XXXXXX
2502784 rows inserted into XXXXXX
2502784 rows committed into XXXXXXX at @XXXXXXX
Data Loading With External Table Finished @170616222252
Rejected Records - 0
Discarded Records – 0
```

Figure 8. Final Log File (One Bunch).

Calling (A) Party	Called (B) Party	Call Date	Call Time	Call Duration	First Cell ID	Call Type	IMEI	Connection Type	Circle
9868xxxxxx	9459xxxxxx	10-Jun-16	18:50:13	105	Mullana	MTC	35293102xxxxxxx	PREPAID	HP
9868xxxxxx	9459xxxxxx	11-Jun-16	13:17:29	186	Mullana	MTC	35293102xxxxxxx	PREPAID	HP
9868xxxxxx	9459xxxxxx	11-Jun-16	13:24:06	42	Mullana	MTC	35293102xxxxxxx	PREPAID	HP
9868xxxxxx	9459xxxxxx	11-Jun-16	7:44:54	91	Mullana	MTC	35293102xxxxxxx	PREPAID	HP
9868xxxxxx	9459xxxxxx	11-Jun-16	7:54:27	65	Mullana	MTC	35293102xxxxxxx	PREPAID	HP
9868xxxxxx	9459xxxxxx	10-Jun-16	13:49:37	291	Mullana	MTC	35293102xxxxxxx	PREPAID	HP
9868xxxxxx	9459xxxxxx	11-Jun-16	15:02:32	31	Mullana	MTC	35293102xxxxxxx	PREPAID	HP
9868xxxxxx	9459xxxxxx	11-Jun-16	9:58:15	100	Mullana	MTC	35293102xxxxxxx	PREPAID	HP
9868xxxxxx	9459xxxxxx	11-Jun-16	10:08:47	63	Mullana	MOC	35293102xxxxxxx	PREPAID	HP
9868xxxxxx	9459xxxxxx	11-Jun-16	17:53:36	13	Mullana	MTC	35293102xxxxxxx	POSTPAID	PB
*** END OF REPORT ***									

Figure 9. Sample Report from telecom data warehouse statistics.

7. Conclusion and Future Works

ETLR is one of the best options available for the telecom data warehouse building. We have understood the telecom terminology of the data warehouse and proposed ETLR framework for telecom domain. Complete end to end solution from source data segregation and extraction, transformation layer at staging server using in built database utilities, optimized loading with external table and copy command followed by the retrieval of the reports from the fact tables has been discussed in a very detailed way with statistics and sample records.

The aim to represent this framework is to facilitate developers in the telecom domain to adapt ETLR and take benefit of the flexibility of the ETLR mechanism with saving of the overall project cost to a large extent. For future, ETLR should be explored in other area like retail, banking etc as it eliminates the use of the commercial ETL tools whereas at the same time providing flexibility, performance and maintainability of the data warehouse system in a much better and customizable way as per the organizations requirement. The data warehouse developers/programmers can customize the process at any stage during the development and can meet organization's needs in much efficient manner. In this paper, we have indicated many open research issues and provided future directions in the area of the telecom data warehouse building.

8. References

1. Sharma S, Kumar K. ETLR - Effective DWH Design Paradigm. Springer: Springer Science Business Media Singapore: Proceedings of the International Conference on Data Engineering and Communication Technology. 2016; p. 149–57.
2. Mohajir BE, Mohajir ME. Telecom data warehouse prototype for bandwidth and network throughput monitoring and analysis. 2011.
3. Agrawal H, Chafle G, Goyal S, Mittal S, Mukherjea S. An enhanced extract-transform-load system for migrating data in telecom billing. Proceedings - International Conference on Data Engineering. 2008. Crossref
4. Li H, Zang M, Liang C. Research and Implementation of a Universal ETL Management Platform based on Telecom Industry. Computational Intelligence and Software Engineering. 2009. Crossref
5. Anand N, Kumar M. Modeling and Optimization of Extraction-Transformation-Loading processes in Data Warehouse. 2013; p. 1–5.
6. Yang P, Liu Z, Ni J. Performance tuning in distributed processing of ETL. Proceedings 7th International Conference on Internet Computing for Engineering and Science. 2013 Sep. Crossref
7. Li X, Mao Y. Real-Time data ETL framework for big real-time data analysis. 2015.

8. Mrunalini M, Kumar TVS, Kanth KR. Simulating secure data extraction in Extraction Transformation Loading (ETL) processes. UKSim 3rd European Modelling Symposium on Computer Modelling and Simulation. 2009.
9. Tu Y, Guo C. An intelligent ETL workflow framework based on data partition. Intelligent Computing and Intelligent Systems. 2010; p. 358-63. PMid:19790034
10. Simitsis A, Vassiliadis P, Sellis T. Optimizing ETL processes in data warehouses. International Conference on Data Engineering. 2005; p. 564-75. Crossref
11. Simitsis A, Gupta C, Wang S, Dayal U. Partitioning real-time ETL workflows. International Conference on Data Engineering. 2010. Crossref
12. Bala M, Boussaid O, Alimazighi Z. P-ETL Parallel-ETL based on the MapReduce paradigm. Computer Systems and Applications. 2014. Crossref
13. Hamed I, Ghozzi F. A Knowledge-based approach for quality-aware ETL process. 2015.
14. Radhakrishna V, Sravankiran V, Ravikiran K. Automating ETL process with scripting technology. 3rd Nirma University International Conference on Engineering. 2012. Crossref
15. Wijaya R, Pudijoatmodjo. An Overview and Implementation of Process in Data Warehouse. 2015.
16. Sun K, Lan Y. SETL A scalable and high performance ETL system. 3rd International Conference on System Science Engineering Design and Manufacturing Informatization. 2012; p. 6–9. Crossref
17. Qin H, Jin X, Zhang X. Research on extract, transform and Load in land and resources star schema data warehouse. 5th International Symposium on Computational Intelligence and Design. 2012; p. 120–3.
18. Baidya A. Global Mobile Data Traffic 2016 - 2021 Asia Pacific Heads 10x Increase. Available from: http://dazeinfo.com/2016/01/07/global-mobile-data-traffic-2016-2021-4g-lte-5g-apac-us/. Accessed date 01/08/2016.