

ESTG_EDP: External sorting as a Tool Used to Process Gas Holder Details Using Early Data Preprocessing

Hrushikesava Raju S.^{1*} and Nagabhushana Rao M.²

^{1*}PP.CSE.0158, Rayalaseema University, Kurnool (A.P.), India

²Department of CSE, K L University, Vijayawada (A.P.), India

*Corresponding Author: hkesavaraju@gmail.com,

Available online at: www.ijcseonline.org

Received: 16/Jan/2017

Revised: 23/Jan/2017

Accepted: 10/Feb/2017

Published: 28/Feb/2017

Abstract— This paper takes external sorting as a tool in which processing made simpler. External sorting usually processes the huge data stored on the external storage devices such as a tape, disks or flash or combination etc. When there is huge gas holders details stored in a database, Some process is required in order to process the operations like filtering (sorting), searching over the gas holders details. The kind of sorting (process) required is external sorting which is used to process huge data (databases or big data) available on secondary storage devices. The external sorting is a useful and efficient technique to perform the operations on the gas holders database. To process the information and producing reports or results, an automated tool is required which asks required user details for processing the searching operation, filtering operation is to be performed if any duplication is found in the data set. The results are specified or illustrated in the Results chapter. Data pre-processing is required when there is any inconsistency present in the given dataset which set-right into quality or neat data set.

Keywords- External sorting, Database, searching, filtering, tool, ESTG_EDP

I. Introduction

The traditional gas offices are the places where customers approach for new gas connections. Getting a gas cylinder also needs the manual approach (by going to the branch), in which booking of filled cylinder is done. Many works to be done requires manual procedure. The drawback of existing methodology is any problem to resolve, that takes more time in the sense problem (complaint) is represented at the concerned gas center, which in turn communicates the same to the head office. The acknowledgement or response to be obtained from head office to the concerned gas center. It is time consuming process. Another is, a person can take multiple gas connections which cause misuse of the resource rights. Second approach is a little advanced approach in which booking for a cylinder can be done through messaging to the standard mobile number. Getting a new gas connection can be done through mee-seva(e-seva) where required documents can be submitted along with filled application form. This also have few drawbacks which are (i) A person can have multiple gas accounts can't be identified when accounts are in different gas agencies like hp, bharath, indane etc. This requires all different gas agencies data bringing into one platform which avoid duplication or identification of such scenarios occurred. (ii) Any complaint to post, can require manual approach in which customer can put a complaint in the concerned agency center. It also takes time

to resolve the issue. (iii) To do operations such as filtering, deleting on a huge data store in processing is also cumbersome. There were also other problems associated with first and second approaches like maintaining records of individual is cumbersome because the details are on manual paper, the ensure of safe maintenance of records is lagging here. This requires a platform where all these solutions are discussed and resolved. Hence, some tool is required to store all the gas information. The advanced data structure called External sorting which stores all the gas holders details in the fields such as name, father name, mother name, spouse if any, sibling names, date of birth, occupation, income, caste etc. as information. Why here the concept external sorting to be taken is huge amount of data to be stored on databases that can be placed or stored in external devices. This can be processed in terms of operations such as filtering, searching etc. requires a tool that acts like external sorting. Hence, the focus is on the automated tool which behaves like a tool that processes a particular gas account in the huge gas holders. The operations that are going to be performed, the process is all going to be explained in the proposed study chapter.

II. Proposed Study

In this, a simulation environment to be created. This environment is considered as a tool where operations should be included such as registration in which user details are

asked who wish to get new gas connection, searching operation where a particular gas applicant id can be searched, if any misleading cases are there in such situations also, searching is possible along with few other additional fields, filtering operation where customer location is searched and nearest location gas connection is preserved and other duplicate connections are disclosed (deleted) in case of domestic customers. Whereas, in case of business customer, based on limit, the booking facility is provided. In this case, number of cylinder eligible for a month takes normal fare but that number crosses the limit, double amount fare is imposed on each additional cylinder. All this gas holders' information to be stored, external devices are required and they have this information in the form of data bases, and other forms. To process all this in the automated tool called external sorting as a tool used to process gas holders details using optimal early data preprocessing" should be required. In processing this, an advanced data structure called external sorting is needed. In external sorting, information from external device database is taken, the defined operations are imposed based on priority. This is the place all vendor gas agencies customer details are available. So far, there is no tool to analyze different gas agencies customers. This is the first tool that retrieve customers belonging to different gas agencies. The power of this tool is process any customer when the person has gas connection id (identifier).

The tool has the following operations:

- A) Login
- B) Registration with the help of data preprocessing(Also called validation)
- C) Searching
- D) Filtering

Where login, and registration modules are user side modules whereas searching and filtering are admin modules. Whatever the work the admin done over user, can be communicated to customer mobile number.

The screen shots of all these modules are specified in the Results chapter. The connections among them are also described in the Results chapter only. Now, the functionality or behavior of this tool (considered ESTG_EDP) can be demonstrated in terms of operations as follows:

A) **Login:** It asks for authorized user name and password. If these details are matched with retrieving fields of gas holders database on the external device. The logic is written as follows:

Pseudo_procedure

ESTG_EDP_LOGIN(u_name,u_pwd):

Step1: Take user name where username must be gas_id and password in to separate variables
 Step2: Compare them with user name and password fields in the database.

Step2.1: Allow accessing the application tool if the details are matched

Step2.2: Deny accessing, suggested to register under any one gas agency

Step3: It have options as tabs like booking, history of booking, amount refunded, edit. User can select any one and perform the work.

Step 3.1: If the option is booking, enter on what date, the cylinder is required.

Step 3.2: The details of previous booked gas details are picked up and displayed.

Step 3.3: It display the amount that are refunded according subsidy percentage.

Step 3.4: It asks the user to provide changed address details, and also asked for unique number can be adhaar number or voter number along with proof of document to be uploaded.

This module expects the user either to enter into application further to do or asks for registration to click and registration is the next module.

B) Registration: It asks for many details about the customer along with the name, father, mother,date of birth, gender, address, identity proof using addhar or driving licence, occupation, sibling details such as daughter's, sons, number of dependents. The definition of this is written as a pseudo code procedure as follows:

Pseudo_procedure ESTG_EDP_REGISTRATION():

Step1: Provide data to the fields like name, father, mother,date of birth, gender, address, identity proof using addhar or driving licence, occupation, sibling details such as daughter's, sons, number of dependents.

Step2: Take those fields data into variables when submit is clicked. Use random long integer function to create one id for that user who entered details. The id created can be sent to registered user mobile along to the mail.

Step3: Connect to the database and insert details into the database using SQL command insert.

Step4: Respond with a positive message successfully inserted and your id(username) is some long number.

Step5: If any violation in the data entered against the fields rules, data preprocessing nothing but here validation checks the fields data and acknowledges the warning through dialog or alert windows.

Step6: This validation (data preprocessing step) process is repeated until all fields data entered are in correct notation.

This module expects entering of details about the user and sends an unique id as user name to the customer. It also processes all the entered field data base on rules such as type, size, constraints framed about concerned fields.

C) **Searching:** It asks for gas_id number first. Then, the tool displays all the information such as user details like address, date of birth, number of connections the customer has,etc about the customer. From the customer details, admin can

confirm about the status of the customer such as number of connections possessed. If the connections are multiple, check the type of customer. If it is again domestic, delete all far connections with respect to address and interest of the customer. If the type of customer is business or enterprise type like hotels, limit the number of connections and also limit the number of cylinders needed per month.

The following is the working of searching operation:

Pseudo_procedure ESTG_EDP_searching(gas_id):

Step1: verify this id against primary key of the database table.

Step 1.1: if match was found, display other column data of that row.

Step 1.2: If it is not matched, alert a warning message "invalid type user".

Step2: Once details of the customer are displayed, admin has to examine the number of connections based on type of user. Then, perform the filtering operation.

Step 2.1: In case of multiple connections, call filtering operation.

Step2.2 : In case of single connection, no need of performing filtering operation.

This module confirms whether a particular applicant have multiple connections or not. In case of former, filtering operation is called by the admin.

D) Filtering: It asks for far connections of the customer details like id as primary key by taking location and interest of the customer into account. It uses one of the SQL command delete for removing additional connections from this tool. The demonstration of filtering operation is as follows:

Pseudo_procedure

ESTG_EDP_FILTERING(extra_gas_ids[]):

Step1: Take SQL command delete. Use delete command over each index of extra_gas_ids[] array from first index to last index values. Each index represents additional gas connection.

Step2: Accept these changes on the database table using commit command. After accepted, display again customer gas details showing only one connection.

This module takes additional gas unique ids, and removes all those additional connections using delete command.

Finally, the tool can be demonstrated by having all these operations is as follows:

Pseudo_procedure ESTG_EDP()

Step1: Use Login module.

Step 1.1: In case of registered user, call login module.

Step 1.2: In case of unregistered user, call registration module.

Step2: The admin search for customers who have multiple gas connections using search operation. In case of multiple

connections, call filtering module which removes duplicate connections. In case of single connection, no action taken.

Step3: The background data structure is used here is external sorting in which customer details can be known by searching area-wise. In this, the external sorting going to be applied according to alphabetical order of mandal, and village under a district. This takes each village details on one tape each, internal sorting is applied over tapes based on varying run size. The mandal type data going to be obtained by sorting each village data separately, the bringing the details according to alphabetical order of the villages. The district type data going to be collected by sorting each mandal data separately, and brings the details according to alphabetical order of the mandals. This is going to be output depending on the area taken in the query.

The tapes or pages can be on either input or output. Initially, data taken on input tape or page and result on output tape or page. The sorting takes alternatively among input and output tapes. The main advantage is varying run size based on village data set, mandal dataset and district data set size. These sizes are based on data set having population.

The external sorting is depicted for gas holders application is as follows:

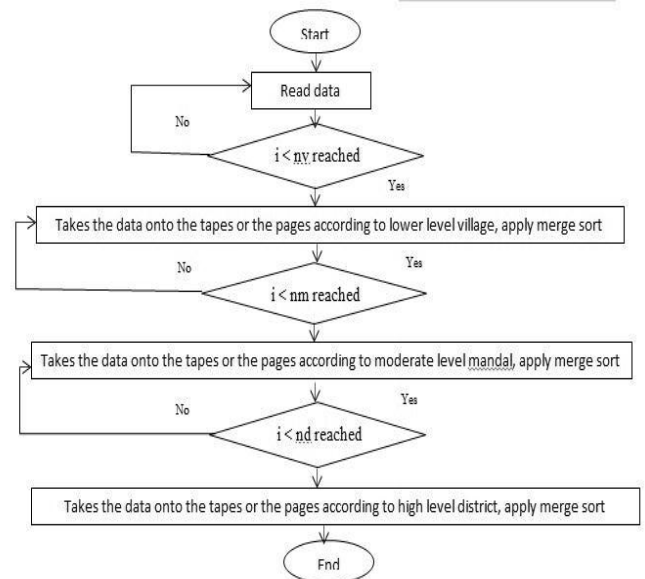


Figure 2.1: External Sorting for gas holders

Step4: Retrieving the final details using external sorting is done quickly using either k-way external sorting or poly-phase merge sorting. In application of external sorting, any duplication occurs in the fields, the data preprocessing removes them and takes or updates only one unique item in the data set.

The pictorial representation of the ESTG_EDP tool is as follows:

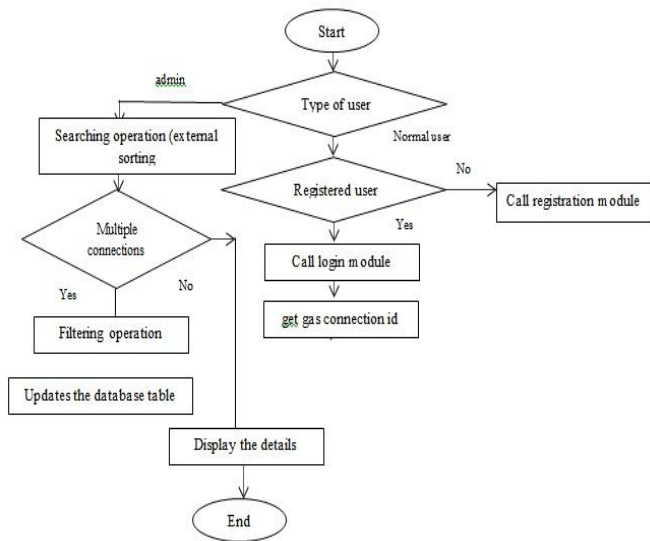


Figure 2.2: ESTG_EDP tool process

The tool processes the information efficiently using external sorting technique as background data structure. The use of this is searching and sorting became easy and became simpler.

III. Results

The screen shots of this tool are demonstrated in terms of its modules such as login, registration, searching, and filtering. The first two operations are done by the basic user customer, and third and fourth operations are done by the advanced user called admin. The order of screen shots is illustrated as below:

The first screen looks like as follows:

The second screen shot if the basic user called customer has to log in:

In above screen, authorized user enters user name as gas_id and valid password.

In case of unregistered user, click on registration link.

The fourth screen can be obtained for the customer gas_id:

The fifth screen is admin login:

The sixth screen is:

The seventh screen is:

The eighth screen is obtained as follows

Gas_id	Name	father	mother	dependents	Dependent1	Dependent2	Dependent3	address	Data of Birth

From this, the admin analyze the details. Based on address, the nearest location id fetched using GPS. The far location connections are deleted in the filtering operation. The ninth screen is as follows:

External sorting as a Tool for filtering multiple gas connections

Enter customer name:

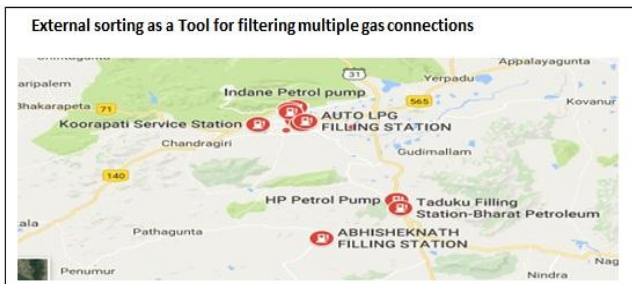
Enter gas_id1:

Enter gas_id2:

...

Enter gas_idN:

The tenth screen is :



Accordingly, the additional gas connections are specified by seeing the map, from which customer address and centers addresses are identified. The far different agencies are taken in to account, only nearest center connection is preserved.

External sorting as a Tool for filtering multiple gas connections

Enter gas_id1

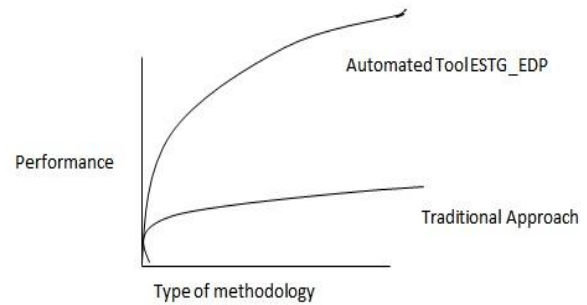
Enter gas_id2:

The eleventh screen shows unique gas connections of the customers in which eight screen having users having multiple connections is refined and eleventh screen is outputted having unique gas connections:

External sorting as a Tool for filtering multiple gas connections

Gad_id	Name	father	mother	dependents	Dependent1	Dependent2	Dependent3	address	Data of Birth

The following graph shows the efficiency of the tool over traditional approaches:



The performance of the traditional approach is low compared to the automated approach called the developed tool.

V. Conclusion

In this, There are many benefits of this tool called ESTG_EDP such as reducing time because it is an automated tool, searching and identifying the multiple gas connections of certain customers according to area selected, remove duplicate gas connections using filtering, update the records and preserve unique customer name records. The results chapter illustrates many screen shots which shows the navigation of this tool right from login credentials to the searching, and then to the filtering. This increases reliability in the sense integrity of the customer details, readability in the sense user friendly in the viewing details, and efficiency is much better in terms of updating the details quickly, and retrieving is also obtained in less time although the data is accessed from huge storage. The external sorting is made as a tool in which searching can be done efficiently and filtering can be applied on the multiple connections. In future, the scope can be updated to the area wise higher level from districts i.e. to the states. Better security methods can be adapted to the admin in doing these operations.

REFERENCES

- [1]. Hrshikesava Raju S. , Nagabhushana Rao M.,“ Improvement of Time Complexity on Pattern Matching using One -Time Look Indexing and Data Preprocessing”, International Journal of computer sciences and Engineering, Vol.4(11), pp.82-86, 2016, E-ISSN:2347-2693.
- [2]. Hrshikesava Raju S.,Swarna Latha T. ,“Dynamic Pattern Matching: Efficient Pattern Matching using Data Preprocessing with help of One time look indexing method”, International Journal of Advanced Research in Computer Engineering and Technology, Vol.2(2),592-599,2013,ISSN:2278-1323.
- [3]. Hrshikesava Raju S. , Nagabhushana Rao M., “Pattern Matching Using Data Preproc-Essing With The Help Of One Time Look Indexing Method”, International Journal of Pharmacy and Technology, Vol.8(3),pp.18395-18407,2016, ISSN:0975-766X.

- [4]. Hrushikesava Raju S. , Nagabhushana Rao M.,“A Review on Specific Data Structures Using Data Preprocessing and Refinement of Existing Algorithms in Order to Improve Time Complexities”,International Journal of computer sciences and Engineering, Vol.4(9),pp.146-151,2016.
- [5]. HweeHwa Pang , Michael Carey J. , Miron Livny ,” Memory-Adaptive External Sorting“, pp.1-12. Proceedings of the 19th VLDB Conference Dublin, Ireland, 1993.
- [6]. John Yiannis, Justin Zobel, “Compression Techniques for Fast External Sorting”, Proc. British National Conference on Databases, A.James (ed),Coventry, UK, pp.115-130, July2003.
- [7]. Young Sik Lee, Luis Cavazos Quero, Youngjae Lee,Jin-Soo Kim, Seungryoul Maeng, “Accelerating External Sorting via On-the-fly Data Merge in Active SSDs”, pp.1-5,2014.
- [8]. Cormen Th. H., Leiserson Ch. E., Rivest R. L., and Stein C.,”Introduction to Algorithms”, 2nd edition, MIT Press ,2001.
- [9]. Knuth D. E., “The Art of Computer Programming: Volume 3 (Sorting and Searching)”,2nd edition, Addison-Wesley ,1998.
- [10].Garcia Molina H., Ullman J. D., and Widom J., “Database Systems: The complete Book”,International edition, Prentice Hall , 2002.
- [11].Ramakrishnan R.,Gehrke J.,”Database Management Systems”, Second Edition,pp.301-315,1997.

Author’s Profile

Mr. S. Hrushikesava Raju, working as a Professor in the Dept. of CSE, Siddharth Institute of Engineering and Technology(SIETK), Narayanavanam Road, Puttur. He is pursuing Ph.D from Rayalaseema University in the stream of CSE. His areas of interest are Data Mining, Data Structures, and Networks.



Dr. M.Nagabhushana Rao, working as Professor in the Dept. of CSE, K L University, Vijayawada, A.P. He had completed Ph.D from S.V. University in the area of Data mining. He is presently guiding many scholars in various disciplines.

