A SUSTAINABLE DEVELOPMENT TECHNIQUE: DATA WAREHOUSE OF WATER QUALITY TO ASSESS AND MONITOR THE WATER POLLUTION IN PERIPHERAL RIVERS OF DHAKA, THE CAPITAL CITY OF BANGLADESH

Mohammad Nasim Akhtar [a,b], Mohammad Mamunur Rashid [a,] , Cyril Y. Kolybanov [a]

[a] Department of Information Technology, Moscow State Academy of Fine Chemical Technology, Moscow, Russia,

[b] Corresponding author: nasim_duet@yahoo.com

 ISSN 1923-6654 (print)
ISSN 1923-6662 (online). Available at http://www.ssrn.com/link/OIDA-Intl-Journal-Sustainable-Dev.html

Abstract: With the industrial development the environment is polluting severely. So this issue becomes the burning question of Bangladesh and it should be solved as Sustainable Development aspect. In Bangladesh, the capital city Dhaka is the nerve center of all activities. The present population of the city is around 12 million and its growth rate is about 3%. The city is nearly surrounded by a circular river system, which includes Turag, Buriganga, Dhaleswari, Balu, Lakhya and Tongi Khal.The surface water along these peripheral rivers is known to be highly polluted due to municipal and industrial untreated wastewaters that are discharged. At present moment, Dhaka Water Supply and Sewerage Authority (DWASA) is highly dependent on ground water for the water supply in the city. As a result the average ground water depletion in most area in the city reportedly around 1-3 m/year. The present rate of depletion is alarming because it can potentially cause environmental hazards such as land subsidence prolonged water logging, alteration in vegetation etc. So, more research work should be done on surface water to make it usable. In this paper the water quality data of peripheral rivers of Dhaka city of different researchers of different time periods have been collected, extracted and then stored in a data warehouse. Data warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data designed to facilitate time-series analysis, trend forecasting and decision making. This procedure brings water quality data in a common platform which gives researchers or related organizations the opportunity of processing data quickly to know the past, present and future trend of water pollution in peripheral rivers of Dhaka city. It also gives the chance of comparing the data from various researchers. Water quality data include physicochemical parameters such as DO (dissolved oxygen), BOD (biochemical oxygen demand), COD (chemical oxygen demand) and TDS (total dissolved solids) as well as concentrations of various chemical pollutants (acid ions, metal ions, ammonia etc).

Keywords: Data warehouse, Peripheral River, Surface and Ground Water, Water Pollution.

I. INTRODUCTION

Today there are many cities worldwide facing an acute shortage of water. Dhaka City (the Capital of Bangladesh) is one such city labeled as a mega city (i.e., cities with population exceeding 10 million) of the world [1],[2]. The city is nearly surrounded by a circular river system, which includes: Turag, Buriganga, Dhaleswari, Balu, Lakhya and Tongi Khal (Figure-2).The surface water along these peripheral rivers is known to be highly polluted due to municipal and industrial untreated wastewaters that are discharged. At present moment, Dhaka Water Supply and Sewerage Authority (DWASA) is highly dependent on ground water for the water supply in the city. As a result the average ground water depletion in most area in the city is reportedly around 1-3 m/year. The present rate of depletion is alarming because it can potentially cause environmental hazards such as land subsidence, prolonged water logging, alteration in vegetation etc. [1],[3],[4]. Thus, there is an urgent need to alleviate the demand on the upper aquifers and explore sustainable sources to augment the present water supply. A sustainable solution is needed. This implies a conjunctive use of groundwater and surface water in order to maintain the balance between anthropogenic demand and water's natural availability [5],[6],[7].

To find the sustainable solution, more research should be done on surface water that can be easier through water quality data warehouse. The primary concept of data warehousing is that the data stored for analysis can most efficiently be accessed by separating the data from the operational systems. In addition to producing standard reports data warehousing systems support very sophisticated online analysis including multi dimensional analysis [INM96CHA97]. In this paper the water quality data warehouse of peripheral rivers of Dhaka city has been designed and developed.

The paper is outlined as follows. Section 2 describes the characteristics of data warehouse. Section 3 describes the subject domain of data warehouse. Proposed data warehouse specially will be loaded with the water quality parameters of peripheral river of Dhaka city, Bangladesh. In this section we have shown the river system of Dhaka city. Section 4 describes the graph model of river systems which allows forming a private river subsystem using various data sources. Section 5 describes us the prerequisite data, knowledge and information, which are necessary for design issue of data warehouse. In section 6 we consider the design methodology of proposed data warehouse. We discuss here the steps of construction of data warehouse, fact and dimensions of warehouse and draw a conceptual snowflake schema of data warehouse. Section 7 describes the graph obtained by executing the SQL commands on DW. Section 8 summarizes the article.

II. A Brief Characterization Of Data Warehousing

As a data warehouse we understand a collection of data that is "subject-oriented, integrated, nonvolatile, time-variant" and it supports the management's decisions [8],[10]. Data warehouses (DW) are not database systems in traditional understanding of this term. They differ from databases among other things with e.g. data models, entering information methods, processing models, query optimization algorithms, visualization techniques [9],[11]. Transformation from data to knowledge has been presented in Figure 1.

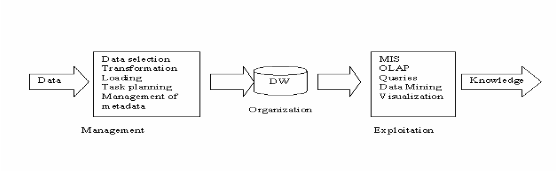

Fig. 1 The example of the managerial information system based on data warehouse

III. Subject Domain

The main rivers around the greater Dhaka comprised the river routes for this paper (Figure 1). Dhaka City is located between 23º35´ to 23º54´ North Latitude and 90º20´ to 90º33´ East Longitude [2]. The routes included (Figure 2):

i. Tongi Canal-Balu river
ii. Turag river-Buriganga river-Dhaleshwari river
iii. Shitalakhya River

The study encompassed the entire reach of the Tongi Canal, partial reach of the Turag (from the confluence of Turag and Tongi Canal to the confluence of Turag and Buriganga Rivers), partial reach of the Balu River (from the confluence of Tongi Canal and Balu River to the confluence of Balu River and Shitalakhya River at Demra), partial reach of the Dhaleshwari River (from the confluence of Buriganga and Dhaleshwari rivers to the confluence of Dhaleshwari and Shitalakhya rivers at Kalagachia) and finally, partial reach of the Shitalakhya river (from Ghorashal railway bridge to the confluence of Shitalakhya and Dhaleshwari at Kalagachia).This river system is our subject domain of construction of data warehouse.

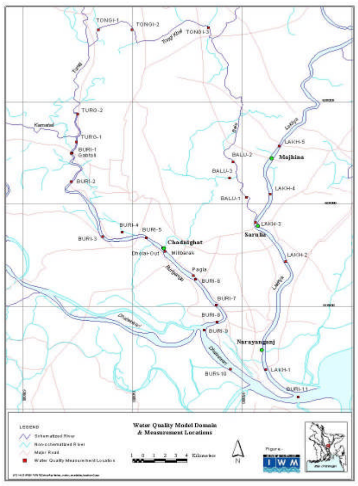

Fig. 2 Peripheral River System of Dhaka city, Bangladesh

IV. Graph Models Of River System

To describe the structure of the river system, graph model has been developed (Fig. 3).Here the vertices of a directed graph is the point of confluence of the rivers falling into the river system (the vertices X4, X5, X7, X8, X10, X11), as well as the conventional boundaries of the river, defining the boundaries of the study to metropolitan (vertices X1, X2, X3, X6, X9, X12). Each edge of the graph corresponds to a fragment of the river system.

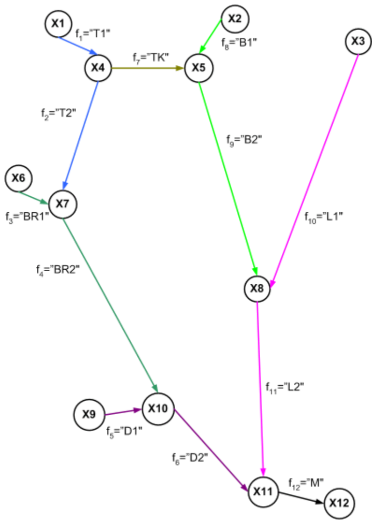

Fig. 3 Graph model of the river system

In terms of set theory we can write the following expressions:

R = (r_j) = ("Balu", ... , "Turag") - the set of rivers;

F = (f_i) = ("T1", "T2", ... , "M") - the set of river segments;

S = (s_k) - the set of river subsystems;

P = {p_l} - the set of sampling points;

$r_j \subseteq$ {fi} - every river consists of a subset of fragments;

$s_k \subseteq$ {f_i} - each river subsystem consists of a subset of fragments;

$P_l \in f_i$ - each sampling point belongs to only one fragment.

In terms of graph theory, indexing of vertices can be arbitrary. However, in this directed graph numbering of vertices is made in a special way, considering the direction of water flow in the river system. The basic rule of numbering of the vertices: Number of the vertices increases along the direction to the flow of the river, that is the vertices with greater numbers located downstream. Thus, for any edge of the graph $f_k = \left(X_i, X_j\right)$ the index of last vertex will be greater than the index of the initial vertex $\left(\forall f_k = \left(X_i, X_j\right)\right)$ $i < j$.

The resulting graph is incomplete and has some special properties that distinguish it from an arbitrary graph. Vertices of the graph, conditional boundaries of the entire river system, have a degree $d\left(X_i\right) = 1$. Graph has no vertices with degree of $d\left(X_i\right) > 3$, it means that the branches of the river at one point simultaneously can be three or less.

Joining of the fragment of river system (corresponding to individual edges of the graph) into the river subsystem (corresponding to individual sub graphs) is also strictly defined by rules. In one subsystem s = (f1, f2) may be united only successive fragments f1 = (X_i, X_j) and f_2 = (X_j, X_k), having a common vertex, which is the end of one fragment and the beginning of other, with i <j <k. It is impossible to combine the fragments which do not have common vertices, ie isolated fragments of the river system. Association of fragments of the rivers in river subsystems allows to model the carry of pollution in the direction of water flow [12].

V. Prerequisites For Designing Of Water Quality Data Warehouse

As a preliminary step for initiating the work, following tasks were first identified:

- A map of the study area showing greater Dhaka and the peripheral rivers was studied first.
- The sampling points of water quality parameter were to be detected and identified through maps.
- Total length of peripheral rivers and the length of subsystems were measured.
- A thorough study of water pollution parameters was completed.
- Water quality standards of drinking water of Bangladesh were collected.
- Water quality data of various research organizations and researchers were collected.
- The collected data need to be extracted, transformed and stored into data warehouse.

VI. Design Methodology

A. The main steps of constructing a data warehouse

- A formal description of the subject domain;
- Development of system requirements to data warehouse;
- Development of structure of data warehouse ;
- Selection of target DBMS and software implementation of data warehouse;
- Development of procedures for administration of data warehouse;
- Development of user interfaces for different group of users.

B. Facts and Dimensions

Analyzing all the source documents (primary raw data), to design the water quality data warehouse it is needed to determine what type of entities and what attributes are necessary to be taken in. The following entities have been considered during designing of data warehouse:

Source: Water quality data stored in the data warehouse have been collected from different sources, such as scientific reports of World Bank, the Department of Water and Sewerage Dhaka (DWASA), Department of Environment, Bangladesh (DOE), Institute of Water Modeling (IWM), the Asian Institute of Technology (AIT), Bangkok, Thailand etc.The information about data sources have been stored in an entity named “Source”. The entity “Source” consists of the fields such as “source_id”, “source_name”, “source_addr” , and “author_information”. The field “source_id” is used as primary key. The fields “source_name”, “source_addr”, and “author_information” have been used for storing source name, source address and information of author respectively.

Type: Water quality data stored in the data warehouse can be obtained by collecting either results of individual measurements or averaged or simulated values. Thus an entity named “Type” with “tide_type” field was included in warehouse structure to indicate this fact, where “tide_id” is primary key.

Water Layer: During the analysis of source documents, it was found that the researchers considered different water layers at the time of water quality measurements. Some researchers have measured values of the water quality parameters at the top of the surface water while some have taken values at the bottom of the surface water. To illustrate this characteristics an entity named “Layer” has been considered, which includes the fields “layer_id” and “layer_name”. “Layer_id” is used for primary key and layer_name for indicating different layer name like surface layer and bottom layer.

Season: Water quality parameters of the different months have been classified as the following seasons:

i) Dry season (December-January-February)
ii) Pre monsoon season (March-April-May)
iii) Monsoon season of rainy season (June-July-August)
iv) Post monsoon season (September-October-November)

To expose the above classification we consider an entity named “Season”. The entity has 3 fields such as “season_id”, “season_name” and “month_included”. The field “season_id” is considered as primary key. The field “season_name” stores the name of season where “month_included” shows duration of each season.

Tide: The peripheral rivers of Dhaka city experience tidal oscillations induced by Bay of Bengal. Some researchers notify the tide types as high tide and low tide while some do not mention this fact. This fact has been considered by an entity named “Tide”. The entity has two fields that are “tide_id” and “tide_type”. The field “tide_id” indicates the primary key and “tide_type” classifies the tide.

Location: The researchers performed their experiments in various locations of the peripheral river system. Water quality data warehouse should store the data collected from various locations. To store information about the location of sampling points, we had to add an entity named “Location” which consists of location_id, river_id, scientific_location_name, public_location_name, total_distance, individual_distance, river_fragment. Location_id uses as primary key. River_id determines to which river the sampling point belongs to. Scientific_location_name and public_location_name store the name of locations used in scientific reports and by local population respectively. For example, a sampling point in Buriganga river is known to the researchers as BURI_1, whereas it is known to the public as “Gabtoli Bridge”. Individual_distance indicates distance from the beginning of the river fragment to the current sampling point. Total_distance indicates distance from the very beginning of the whole river to the current sampling point.

River: Six river reaches encompass the Dhaka city which includes Tongi Khal, Turag river, Buriganga river, Dhaleswari river, Lakhya river and Balu river. Different researchers performed their experiments in different locations at different rivers. We need to locate the sampling point which belongs to particular River, so we have thought about an entity containing the river name. The field river_id uses as primary key and river_name indicates the name of the river.

Parameter: The concentration water quality parameter shows the water pollution level of a river. The water is polluted if the concentrations are beyond to the standard level of water quality. The main parameters of water quality in the river systems are the concentration of dissolved oxygen (DO), chemical (COD) and biochemical (BOD) oxygen demand, total dissolve solids (TDS), as well as concentrations of chemical pollutants acid ions(NO_3^-, Cl^-, SO_4^{--}, PO_4^{--}), metal ions (Cr^{++}, Pb^{++}, Hg^+, Zn^{++}),ammonium(NH_4^+) etc. An entity named “Parameter” was used to store parameters. This entity consists of 6 attributes such as “parameter_id”, “parameter_name”, “drinking_ws”, "household_ws", "fishing_ws" and "irrigation_ws" which correspond to primary key, parameter's name and water quality standards of Bangladesh for various uses such as drinking water, household work, fishing and irrigation.

River fragment: It is observed that different organizations have assessed the water quality in different manner. Some have assessed the water quality of a particular river while some have taken as subject area the particular fragment of the river. For this reason, the whole peripheral river system has been divided into several fragments. An entity named "River fragment" has been considered during designing of data warehouse. The entity consists of six fields. The field "river_fragment_id" uses as primary key. The field "river_fragment_name" stores the name of the river fragment. "Shortriverfragment_name" exposes the abbreviation for the fragment. Riverfragment_length indicates the length of river fragment. The fields "river_id" is used as foreign key.

Concentration: This is the vital entity which keeps direct or indirect relations with all above mentioned entities and contains extra attributes such as "date and time" of the experiments and "concentration" levels of every water quality parameter.

C. Conceptual Schema

Most data warehouses use a star schema to represent the multidimensional data model. The database consists of a single fact table and a single table for each dimension. Snowflake schema provides a refinement of star schema where the dimensional hierarchy is explicitly represented by normalizing the dimension tables. This leads to advantages in maintaining the dimension tables. However, the renormalized structure of the dimensional tables in star schemas may be more appropriate for browsing the dimensions. Fact collections are examples of more complex structures in which multiple fact tables share dimensional tables. In our proposed DW we use the snowflake data model schema because:

- The snowflake schema is a normalized star schema. In a snowflake schema, the dimension tables are normalized.
- In some cases it may improve performance because smaller tables are joined.
- It is easier to maintain.
- It increases flexibility.

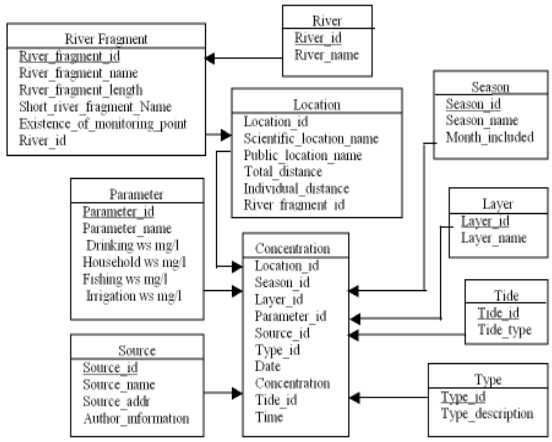

Fig. 4 Snowflake data model schema of proposed data warehouse

VII. VISUALIZATION AND RESULT DISCUSSION

As an example, we consider a river subsystem consisting of four fragments of three different rivers - Turag (T1 and T2), Buriganga (BR2) and Dhaleswari (DR2). The total length of river subsystem is L = 55 km. Dissolved oxygen (DO) refers to the volume of oxygen that is contained in water. Its limit is 4 according to the standard of Department of Environment, Bangladesh. Below of this limit it is harmful as because Dissolved oxygen plays a large role in the survival of aquatic life.

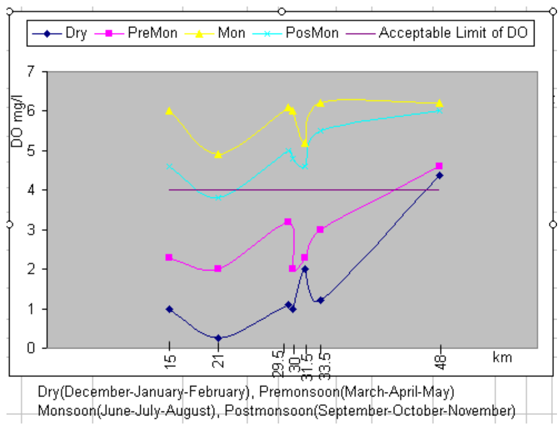

Fig. 5 Profile of DO along Turag-Buriganga-Dhaleshwari River in 2008

The above figure shows the concentration of dissolved oxygen (DO) in the considered river subsystem in 2008. We present here 7 points to which the industrial and domestic effluent of the city fall through canals.

Here it is analyzed that the most favorable conditions, ie highest concentrations of dissolved oxygen is observed during the monsoon period (June - August) and the lowest concentrations is observed during dry period (December-Feb). Data for the graph is obtained from different literature sources by SQL-query exploring our data warehouse. In total more than 11 000 records were collected and processed over the period of 1998 to 2009 years on 135 locations of river systems.

VIII. Conclusion

Measurement of water quality data is very expensive, requires skilled manpower, sophisticated instrument and well organized laboratory facilities. One of the best advantages to use our data warehouse is that users will be able to access a large amount of information, which can be used to solve a large number of problems. As our water quality data are taken from multiple sources and placed in a centralized location, an organization can analyze it in a way that may allow them to come up with different solutions than they would if they looked at the data separately.

Overall, the work has laid the foundations for the Dhaka city planners and designers to make a qualitative resource assessment of surface water. Such an assessment can eventually evolve to a long-term monitoring system of water supply sources for Dhaka City. The work also facilitates water quality modelers to choose the proper water quality model and help in formulating the strategy for water abstraction and water supply for Dhaka city.

Acknowledgment

The authors wish to thank the Water Resources Planning Division of the Institute of Water Modeling (IWM) (Dhaka, Bangladesh), Department of Environment (DOE) (Bangladesh) for data on the peripheral rivers of Dhaka. In particular, support received from Mahbubur Rahman, Head of WRP Division (IWM), Dr. Sohrab Ali, Deputy Director Natural Resource Management (DOE), Md. Abul Kalam Azad, Analyst (DOE) are gratefully acknowledged.

References

[1] Haigh , M.J., Sustainable Management of Headwater Resources: The Nairobi 'Headwater' Declaration (2002) and Beyond, Asian Journal of Water, Environment and Pollution, Vol. 1(1-2), 2004, pp.17 –28.

[2] Karn, S.K. and H. Harada, Surface Water Pollution in Three Urban Territories of Nepal, India, and Bangladesh,Environmental Management,vol. 28(4), 2001, pp. 483-496.

[3] Kabbour, B.B, and L., Zouhri Overexploitation and continuous drought effects on groundwater yield and marine intrusion: considerations arising from the modeling of Mamora coastal aquifer, Morocco, Hydrological Processes, 2005 (In press, doi: 10.1002/hyp.5842).

[4] Karami, E. and D. Hayati Rural, Poverty and Sustainability: The Case of Groundwater Depletion in Iran, Asian Journal of Water, Environment and Pollution, Vol. 2(2), 2005, pp. 51-61.

[5] Onta, P.R., A.D. Gupta, and R. Harboe, Multistep Planning Model for Conjunctive Use of Surface- and Ground-Water Resources, Journal of Water Resources Planning and Management.Vol.117(6), 1991, pp. 662-678.

[6] Ejaz M.S and R.C. Peralta R.C., Maximizing conjunctive use of surface and ground water under surface water quality constraints, Advances in Water Resources, Vol. 18(2), 1995, pp. 67-75.

[7] Emch, P.G. and W. W-G. Yeh., Management Model for Conjunctive Use of Coastal Surface Water and Ground Water, Water Resources Planning and Management, Vol. 124(3), 1998, pp. 129-139 (May/June1998).

[8] W.H. Inmon: Building the Data Warehouse. ed. Wiley. 1996

[9] S. Chaudhuri, U. Dayal: An overview of Data Warehousing and OLAP technology. ACM SIGMOD Record 26 (1). 1997

[10] Inmon, W.H., Welch, J.D., & Glassey, K.L., Managing the data warehouse. Wiley Computer Publishing, 1997.

[11] Grothe, M., & Gentsch, P., Business intelligence. Addison-Wesley, 2000.

[12] Dzhebunnahar, Kolybanov K., Morozov, OA Simulation of pollutant transport in the river Buriganga (Dhaka city, Bangladesh) on the basis of integrated water quality criteria. Proceedings MITHT, Moscow, Russia № 1, 1999.

About the autours :

Md. Nasim Akhtar is with the Ph.D. Student, Department of Information Technology, Moscow State Academy of Fine Chemical Technology, Moscow, Russia, (email: nasim_duet@yahoo.com).

Md. Mamunur Rashid is with the Ph.D. Student, Department of Information Technology, Moscow State Academy of Fine Chemical Technology, Moscow, Russia, (email: mamunst2003@yahoo.com).

Cyril Y. Kolybanov is with the Professor, Department of Information Technology, Moscow State Academy of Fine Chemical Technology, Moscow, Russia, (email: cyrk@mail.ru).