Optimization for Date Redistributed System with Applications

Mădălina Văleanu, Smaranda Cosma, Dan Cosma, Grigor Moldovan, Dana Vasilescu

Mădălina Văleanu
University of Medicine and Pharmacy "Iuliu Hatieganu"
Medical Informatics and Biostatistics Department
Cluj-Napoca
E-mail: mvaleanu@umfcluj.ro

Smaranda Cosma
Babes-Bolyai University
Business Department, Faculty of Business
Cluj-Napoca
E-mail: smaranda.cosma@tbs.ubbcluj.ro

Dan Cosma
University of Medicine and Pharmacy "Iuliu Hatieganu"
Department of Pediatric Surgery and Orthopedics
Cluj-Napoca
E-mail: dcosma@umfcluj.ro

Grigor Moldovan
Babes-Bolyai University
Computer Systems Department
Cluj-Napoca
E-mail: moldovan@cs.ubbcluj.ro

Dana Vasilescu
University of Medicine and Pharmacy "Iuliu Hatieganu"
Department of Pediatric Surgery and Orthopedics
Cluj-Napoca
E-mail: dana.vasilescu@umfcluj.ro

Abstract: In this paper we intend to define a strategy for managing databases with mobile structures, taking into account their redistribution in the nodes of a computer network. The minimal cost of the redistribution is highlighted and some applications for medical and business databases are presented.

Keywords: distributed database, costs, medical databases, economical databases, wireless network

1 Introduction

The paper presents a generalization and extension for mobile environments of the context of the problem expressed in paper [1]. It also presents some applications of the problem. Let us consider the mobile databases (tables) $B_i, i = \overline{1, n}$ distributed in r nodes of a network of computer stations with own memories $S_i, i = \overline{1, r}$. Hence, we have:

$B = \{B_1, B_2, \ldots, B_n\}$ and $S = \{S_1, S_2, \ldots, S_r\}$,

where B is a distributed database.

Now we identify the nodes, that is the stations of the computer network considered with the same symbols as for the memory supports S_i in S.

The architectures of the fixed computer networks with nodes $S = \{S_1, S_2, \ldots, S_r\}$ can differ, but on the horizontal level, in general, they are modeled through a relative graph. Particular computer network architectures can exist, such as the hypercube or generalizations of the hypercube [2] [3], which provide simple and efficient routing means but whose complexity in number of nodes and restrictions make them inefficient in the end. Modeling a computer network in a graph, in general, will give a static network with fixed and well defined geographical locations.

In our previous paper [1] we have supposed that subbases B_i of the database B have a well defined location and that this location is maintained fixed during the running of a distributed application.
At present, mobile databases characterized by allocation in permanent change are known. Consequently, dynamism characterizes all their aspects. They are searched for (selected), accessed and processed from a mobile environment made up of laptops, mobile phones etc. They become stronger and stronger elements of data processing. They are connected by means of Wireless stations in special fixed points (nodes) belonging to a computer network. The traditional model of transactions migrates towards a mobile transaction model.
The following presents a strategy for this kind of databases processing.

Databases $B_i, i = \overline{1, n}$ and other soft resources (programs) shall be stored in fixed hosts, noted FH and identified through S_i (fixed host computers!), in a well established network, $S = \{S_1, S_2, \ldots, S_r\}$. The mobile environment will inherit the properties of the distributed environment. The mobile environment with ever increasing storing spaces will be able to take over the mobile databases for processing data, by duplicating procedures, and the results will turn back to a fixed host.

The fixed hosts will be those that permanently preserve the data subbases B_i. Databases B_i could be preserved in mobile hosts, but any update (modification) should end with their replication to a fixed host mentioned with a new version number. In the fixed network, it is necessary to update all duplicated subbases B_i, using the most recent version numbers allotted in order to maintain the consistency of the distributed database B. The update of the databases should occur instantly after the alteration of their version number. The update of duplicated data subbases B_i can also be achieved with the help of a jeton function periodically passing through the nodes S of the network in consideration.

It is obvious that between the mobile hosts noted MH and fixed hosts noted FH belonging to a computer network, there is a fixed interface, called mobile support station, noted MSS (mobile support station) or base station. The connection between a mobile support station (MSS) and a mobile host (MH) is wireless. Each MSS allotted to a node S_i controls a cell of mobile hosts identified by $\{i_1, i_2, \ldots, i_k\}$. A mobile host can disconnect from a MSS and possibly reconnect to another MSS just while running a distributed application. The disconnection and connection of mobile stations occurs frequently. Disconnection establishes a new distribution of subbases in the network nodes.

The figure below presents the general architecture of a mobile platform.

In the design of a mobile database, in each node S_i we shall highlight - for B_i - an often modifying and dynamic component that frequently is accessed from other nodes of the network, noted b_i and other components, noted $m_{i(s)}$ associated to mobile units, quite often accessed locally to mobile units but where the modification of data is less frequent.
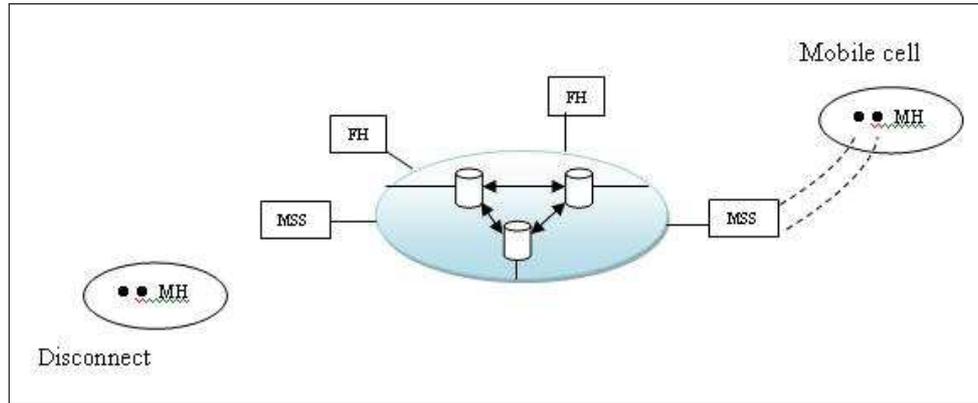
Figure 1: Mobile platform

We have

$B_i = (b_i; m_{i(1)}, m_{i(2)}, \ldots, m_{i(k)})$

where we used for indices i_p notation $i(p)$ and $p = \overline{1, k}$. We can consider that we made a decomposition of the subbases B_i by selection and/or projection operations so that by further union operations we get updated subbase B_i.

At a given moment in the computer network stations we witness a certain distribution, respectively a grouping of databases B_i, $i = \overline{1, n}$. In general, if $n \geq r$, then, more subbases B_i will exist in the memory S_i. For reasons of simplification, in our study, we will suppose that consecutively, in each station S_i of the r station, we will have d subbases S_i, hence $n = d.r$. We also suppose that we have removed, by means of a certain conveniently selected strategy, the duplications of subbases B_i, therefore B_i are distinct. A distributed application supposing programs running in the network under consideration leads to the access, from the nodes S_i, of the subbases B_j in a defined succession, until the result needed is obtained. We note the successively accessed data subbases (some of them more often accessed in a distributed application) in a vectorial form, as follows:

$B_L = (b_{m_1}, b_{m_2}, \ldots, b_{m_s})$

then

$L = (m_1, m_2, \ldots m_s); m_k \in \{1, 2, \ldots, n\}$

Remember that m_k identifies the place in the succession of accesses of the subbases in B performed.

In general, in the case of an access from S_i network node to a subbase found in S_j, a so-called penalty should also be considered (for instance: time, cost) noted with p_{ij} for all $i, j = \overline{1, r}$ which will be composed of a fixed penalty pf_{ij} established in a fixed computer network and possibly from a penalty due to the specific working manner with mobile databases noted pm_{ij}. We will obtain $p_{ij} = pf_{ij} + pm_{ij}$.

2   Grouping data subbases on the network fixed nodes

Let distributed database be $B = \{B_1, B_2, \ldots, B_n\}$. The indices of subbases build the set $I_n = \{1, \ldots, n\}$. The reorganisation of these data is defined by permutation indices

$\sigma = \begin{pmatrix} 1 \, 2 \, \ldots k \ldots n \\ i_1 i_2 \ldots i_k \ldots i_n \end{pmatrix}$

where $i_k \in I_n; k = \overline{1, n}$ are distinct.

The permutation σ is also written as $\sigma = (\sigma(1), \sigma(2), \ldots, \sigma(n))$. If we have two permutations σ and τ,

their produce $\sigma\tau$ is obtained by the composition of the two functions, so that

$\sigma\tau = (\sigma(\tau(1)), \sigma(\tau(2)), \ldots, \sigma(\tau(n)))$.

Let us mark with $\text{Supp}(\sigma)$, the permutation support σ, i.e. the set of the elements $i \in \{1, 2, \ldots, n\}$ having the property $\sigma(i) \neq j$.

A permutation σ is called cyclic of length m, $m \geq 2$ if elements $i_1, i_2, \ldots, i_m \in \text{Supp}(\sigma)$ exist so as $\sigma(i_1) = i_2, \sigma(i_2) = i_3, \ldots \sigma(i_{m-1}) = i_m, \sigma(i_m) = i_1$.

It is known that any permutation that is different from the identical one can be written as a produce of distinct cycles. [4]

Let us consider that $n = d.r$, i.e. on any S_i station in the r computer network we have d data subbases. Let us take the following reorganisation (distribution) of the n data subbases on the r workstations, successively:

$B_{\sigma^{-1}(1)}$	$B_{\sigma^{-1}(2)}$	$B_{\sigma^{-1}(d)}$	in S_1
… … … …			
$B_{\sigma^{-1}(1+(i-1)d)}$	$B_{\sigma^{-1}(2+(i-1)d)}$	$B_{\sigma^{-1}(id)}$	in S_i
… … … …			
$B_{\sigma^{-1}(1+(r-1)d)}$	$B_{\sigma^{-1}(2+(r-1)d)}$	$B_{\sigma^{-1}(rd)}$	in S_r

Note: σ is a bijective application, $\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ where $\sigma(k) = i_k$,

respectively $\sigma^{-1}(i_k) = k$; $k = \overline{1, n}$.

Consider the sequence of successive accesses $L_D = (m_1, m_2, \ldots, m_s)$ of the subbases in B and the indices of these stations are. With these notations, we define the cost of a distributed application with the relationship

$$C(S, B, L_D, \sigma) = \sum_{k=1}^{s-1} (p_{R_\sigma(m_k),R_\sigma(m_{k+1})} + q_{R_\sigma(m_k)})$$

where $q_{R_\sigma(m_k)}$ represents the cost of the activities in station $S_{R_\sigma(m_k)}$.

Let us note with a_{ij} the number of the times $(m_k, m_{k+1}) = (i, j)$, $k = \overline{1, s-1}$; $i, j \in \{1, 2, \ldots, n\}$, and with c_i , how many times $m_k = i$, $k = \overline{1, s}$; $i, j \in \{1, 2, \ldots, n\}$.

Then,

$$C(S, B, L_D, \sigma) = \sum_{i=1}^{n} \sum_{j=1}^{n} (a_{ij} p_{R_\sigma(m_k),R_\sigma(m_{k+1})} + c_i q_{R_\sigma(m_k)})$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} (a_{\sigma^{-1}(i),\sigma^{-1}(j)} p_{R_I(i),R_I(j)} + c_i q_{R_I(i)})$$

The cost of a distributed application will be:

$$C(S, B, L_D, \sigma) = \sum_{i=1}^{n} \sum_{j=1}^{n} (a_{\sigma^{-1}(i),\sigma^{-1}(j)} p_{ij}^* + c_i q_i^*), \text{ where } p_{R_I(i),R_I(j)} = p_{ij}^*; q_{R_I(i)} = q_i^*$$

A way to find out this sums can see in [5].

Notes

1. In practice, we can consider that p_{ij} is symmetrical, hence p_{ij}^* will also be symmetrical, i.e. $p_{ij}^* = p_{ji}^*$. Penalties p_{ij} can be determined, for instance, with the help of statistical data after more run-

nings of the programs of the distributed application D.

2. If permutation σ is decomposed in a produce of cyclical permutations, the formula for the application cost can be simplified accordingly.

3   Applications

3.1   Medical Databases

The medical field is one of the most important fields of social concern. It is defined by high level dynamism.

Nowadays the healthcare system from every country have to faced to many challenges of the 21st century which influenced significant the financial aspects of the organizations' activities. In these new conditions the competitions increased and pressed on the costs to acquire and maintain a high quality of the technology and capital outlay.

Today, the data needed for research are registered in medical documents, in a written form, but not in a unique form and the work necessary to retrieve and process such data is enormous.

It is for this reason that it is required to design general interest databases, for doctors, patients, researchers and health units. Such databases are characterised by the fact that they will include an enormous volume of data distributed among the nodes of a computer network.

The medical field activities can be divided into primary level care, secondary level care activities and the division can go even further.

Primary care refers to family medicine that is to the first contact and consultation point for the patient. Secondary care is the service provided by a specialist, who does not have the first contact with the patient, in general. Usually, a doctor providing secondary care services treats patients previously consulted by a family doctor.

To establish a correct diagnosis and a proper treatment it is necessary to handle the databases in question that, often mean to return to their redistribution in the nodes of the computer network. A concrete examples are given in papers [6]. In such an application it is equally important to find an answer to a query to the databases in as short a time as possible.

A valuable component of a heath care computer-based system is its capacity to work with information on the patients stored in various locations: the national health authority, the social insurance system, providers of primary and secondary care services. In order to find solutions, a distributed approach should be taken into consideration. [7] [8]

Health care information, such as medical records, X-rays, lab tests results are more are more kept and processed by computers. That is why standards to send such data away in an unambiguous manner among computers are required. In this way, uniform health care-related information shall also be available.

A standard represents a conceptual lexis common for all the stakeholders in the healthcare network. Standards of this kind are already available for many medicine subfields.

3.2 Economical Databases

The economic applications of distributed databases are diverse and numerous and their importance is also valuable. It is enough to take as an example a system in the field of international trade involving the handling of very large volumes of data relative to the varied goods distributed among shop networks. The management information system of a company becomes decisive for international corporations of the Learning-Company type that extend more and more their activities using varied international market penetration strategies.

E-commerce develops and includes larger and larger scopes. That is why it is necessary to establish standards related to marketed products, in this field too. The queries put to distributed and increasing in size databases should be made more efficient in time. It is sure that in the near future the optimising of costs (query types) of the used databases will have to be considered.

Marketing research identifies, collects, processes, analyses, interprets and communicates information relevant for a specific marketing situation to make a decision. All the data are organised in distributed databases in more locations.

Marketing research mainly aims at reducing risk and uncertainty in the conception and grounding of marketing-related decisions and at implementing and controlling the practical putting into application of these decisions. [7]

4 Conclusions

The fundamental problem, with respect to the distributed application D under consideration, consists in the determination of a permutation σ in the set of possible permutations P having elements $\{1, 2, \ldots, n\}$, indices of the B data subbases so that the cost of the use of the distributed application D programs would be minimal, i.e.
$$\min\{C(S, B, L_D, \sigma); \sigma \in P\}.$$

It is obvious that the problem relates to combination, its solution is important when the distributed application D is used repeatedly. The problem is solved once and the advantage remains operational all along the use of the respective distributed application.

The applications for this are multiple. Here are describes some applications in medical and economical domain, that can use distributed databases and where the problem of cost minimizing (time need to transferred information from one point of the network to another) is very important.

Bibliography

[1] G. Moldovan, M. Văleanu, *The performance optimization for date redistributing system in computer network,* International Journal of Computer, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844, Vol III, Supl. issue, p. 116-118, 2008.

[2] G. Moldovan, M. Văleanu, *Redistributing databases in a computer network,* Analele Univ. Bucureşti, Ser. Math.-Info., 56, 2006.

[3] G. Moldovan, I. Dzitac, *Sisteme distribuite - Modele matematice,* Ed. Univ. Agora, 2006.

[4] L. Aspinall, *Data base. Re-organisation –Algorithms,* IBM, UKSC – 0029, 1972.

[5] A. Gog, H. Grebla, *Evolutionary Tuning for Distributed Database Performance,* The 4th International Symposium on Parallel and Distributed Computing (ISPDC), Lille, France, IEEE Computer Society, 2005, p 275-281.

[6] S. Cosma, D. Cosma , A. Negrusa, M. Văleanu, G. Moldovan, D. Vasilescu, *Implementation of the communication system for clubfoot,* WSEAS Transactions on Communications, ISSN:1109-2742, ISSUE 9, Vol 7, sept 2008, p.932-941

[7] D.E. Vasilescu, D. Cosma, M. Văleanu, I. Negreanu, D. Vasilescu, *The results of the early conservative orthopedic treatment in the congenital talipes equinovarus,* Applied Medical Informatics 2004, Vol 15, p 34-43.

[8] D. Cosma, S. Cosma, M. Văleanu, D.E. Vasilescu, G. Moldovan, *Web - based guideline for clubfoot: patient - orientated materials,* Journal of International Business and Economics, ISSN 1544-8037, 8-1, 2008.