UDC 004.052

I. SLIZOVSKAYA

National aerospace university named after N.E.Zhukovsky "KhAI", Ukraine

CLASSIFICATION OF THE INFORMATION INTEGRITY RECOVERY METHODS IN THE SYSTEMS OF VARIOUS COMPLEXITY

The main methods of data recovery are considered, the efficiency analysis according to the cost criteria, recovery speed and ability to apply it to various tasks was carried out .

information integrity, reliability , recovery, fault tolerance, hardware and software combined

Introduction

Currently the prominent analysts in the sphere of information technologies state [1] the fast increase in the information stream volumes that results in necessity to create more powerful systems of processing and storing data. In its turn, the development of new systems is connected with solving certain problems, among which the problem to ensure reliability and information recovery after failures [4], takes one of the most important places. Today, the following methods of data storing and data recovering became wide-spread [9]:

Software-based Method

This method uses the software which allows conducting an operation of data archive copying for their further recovery. Such software tools as BACKUP [7] can be supplied as an example. The advantage of this method is high adaptively allowing to store type safe information that causes the reduction of time access to it. The weightiest disadvantage is discreteness in work leading to reliability decrease. Taking into account the fact that the processes of data storing are conducted in parallel with the main task, it might happen that at a failure a part of information will not be stored due to the lack of time necessary for accomplishing it and as a result, it will be impossible to recover it. This feature is particularly apparent at large volumes of data streams.

Hardware and software combined method

This method realizes in the system of reserved data in cluster depositories under the management of a special program code the task of which is to serve these clusters and high-speed communication channels that connect them. Among such realizations we may distinguish Microsoft Windows 2003 Server and NAS. The advantage of this method is automated switch among clusters if the possibility of information loss appears. At such a realization there is no necessity of recovery procedure, as all the system clusters store the same information [11]. As a disadvantage of this method we may indicate its high cost in its both hardware and software part.

Hardware Method

The basis of this method is application of RAID information storage technologies. The technologies of apparatus doubling of data (RAID-0), recording of information in excessive view (RAID-5 и RAID-6), and mixed technology (RAID-5 + RAID-0) [3] became wide-spread. The advantage of this method is high speed and ability to process large volumes of information. In addition, it is necessary to note the

steadiness to the apparatus fault tolerance with the possibility of its partial replacement without causing problems to system productivity. The main disadvantage of this method is high cost both of equipment and its exploitation.

Apparently, to ensure data safe-keeping, the choice of a required method must be stipulated by many criteria. The target purpose of the information system (billing system, corporate system etc.) defines the control frequency and volumes of the inquired information that specifies the choice of the platform for its realization.

In this connection, the basis of the systems with critical approach to the time of data processing is expensive high-speed hardware solutions allowing to execute high quantity of information operations during a certain period of time. The example is billing system of real time the defining criterion for which is high speed and, for instance, environmental support system of a satellite in which reliability and response time define the ability of an object to survive.

In opposite, for the systems of data processing that do not require frequent data retrieval, less speedy (correspondingly, less expensive) methods are used, because the criterion of speed is not defining for them. In these cases, the hardware-based or purely software methods are applied, if the volumes of information are not large. These methods include doubling, archiving and cashing.

Another criterion stipulating the choice of the method of storage and data recovery is its cost. Due to the multifactor feature of this criterion, for general estimation we may assume that the cost is formed accor-ding to the following parameters:

$$C = \left(C_a + C_p\right) + \left(p_b + p_R\right) < P_L, \qquad (1)$$

where C_a – the cost of hardware part of realization;

C_p – the cost of software part of realization;

p_b – storage expanses;

p_R – recovery expanses ;

P_L – loss data retrieval expenses.

The ratio suggests that, from the economical point of view, the method the cost of which is less than P_L expenses for retrieving loss data should be applied.

The study case results of publications and technical documentation about modern data control systems show that efficiency of this or that method of storage depends on the following parameters:

Level of automatization

Any of reviewed methods considers some degree of automatization. Following methods exist: automatic, semiautomatic and manual. Automatic methods have maximum speed because from localization of cause of failure to the moment when it is fixed, all is done automatically.

Semiautomatic methods are characterized by need of decision to be taken by operator, who is choosing suitable scenario after problem cause determination. These methods are cheaper than automatic and have less speed.

During usage of manual methods solution of all tasks is done by operator who does the restore. These methods are the slowest ones and applicable for systems which are not time-sensitive.

The ratio of speed and cost

Complex relative criteria which defines economical effectiveness of method when cost of data processing is known is:

$$\overline{C} = \frac{C}{Q}, \qquad (2)$$

where C – cost of the method, defined by dependency (1);

Q – quantity of data processed by the method per time unit.

As we can see from (2), this criteria lets us evaluate economical effectiveness of different methods. Thus, for example, for big volumes of critical data to which expensive methods of providing of reliability are applied, criteria of relative cost will be similar to less expensive methods applied for processing of less critical systems. In most of cases this criteria lets us determine understanding (definition) of method cost in context of volume of data processed.

Considering aspects of reviewed methods of storing and restoring of data and offered criteria we shall highlight main attributes for classification.

Because each of methods is oriented to provide reliability of the system for solution of such kind of tasks, it usually has specific hardware and software realization. It is defined by dedication of the system and in general can be defined as sensitivity to reaction time for failure situations. Among such target systems, we can define telecommunication billing systems, systems for life conditions in Space, systems of providing of non-interrupting processes (energetic industry, metallurgy, etc.), information systems of general usage(knowledge bases, data archives, etc.)

Let us introduce attribute of method belong to architecture realization types. In such way, methods are realized as parts of multi-platform RAID systems, real-time cluster systems and classical types of server systems managed by RDBMS. For each type of mentioned realization following characteristics are related: speed and volume of processed data flows. Cost of technical solutions depends directly on them: the higher data processing speed is and the higher demands for reliability and data volumes means higher is the cost.

Schema of classification of methods of data storage and restore is shown on fig.1.

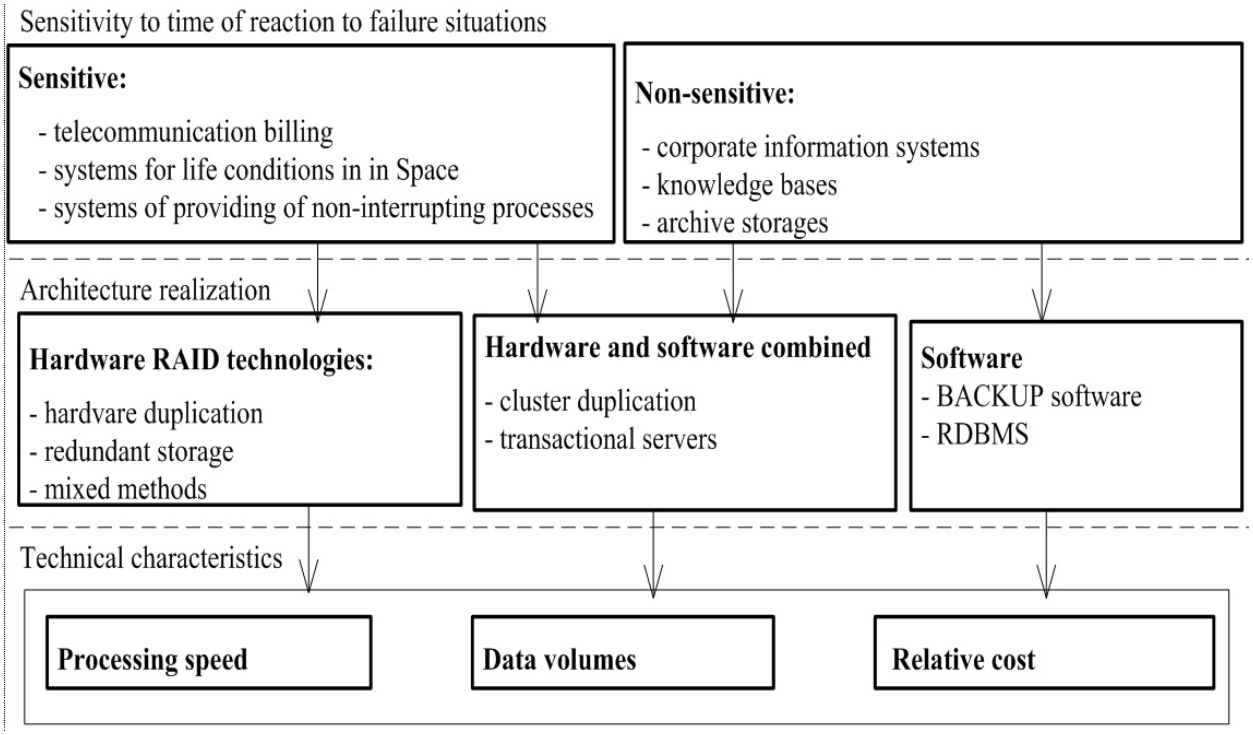

Fig. 1 Classification of methods of data storage and restore

Analyze of the most used methods of data storage ant restore gives us possibility to define general attributes and criteria, basing on which was done general classification of reviewed methods.

Further researches will be targeted to more detailed learning of methods used in telecommunication billing systems. Aim of these researches will be to form a concept of new method which will include all the best features of known solutions and excluding their handicaps.

Literature

1. Соммервилл И. Инженерия программного обеспечения, 6-е издание.: Пер. с англ. – М.: Издательский дом "Вильяме", 2002. – 624 с.

2. Липаев В.В, «Надежность программных средств». – М.: СИНТЕГ, 1998.

3. Dondolossa G. Formal Methods in the deve-lopment of safety critical knowledge - based components // ACM. – 2005. – P. 23-35.

4. Bernstein P.A., Hadzilacos V., Goodman N. Concurrency Control and Recovery in Database Systems, 1987. – 370 p.

5. Andersson B Static Priority Scheduling in Multiprocessors. PhD Thesis, Department of Comp.Eng., Chalmers University, 2003. 18 p.

6. Baruah S. Robustness Results Concerning EDF Scheduling upon Uniform Multiprocessor. Euromicro Conf. on Real-Time Systems, 2002. P. 48-51.

7. Baruah S. Optimal Utilization Bounds for the Fixed-Priority Scheduling of Periodic Tasks Systems on Identical Multiprocessors // IEEE Transactions on Computers, 2004. – P. 781-784.

8. Baruah S., Scheduling Periodic Tasks on Uniform Multiprocessor, Euromicro Conference on Real-Time Systems, June 2000. P. 334-337.

9. Baruah S. Deadline-based Scheduling of Periodic Task Systems on Multiprocessor // Information Processing Letters. – 84 (2). November 2002. – P. 93-98.

10. Stankovic V., Popov P. Improving DBMS Performance through Diverse Redundancy. – London : Centre for Software Reliability, 2006. – 11 p.

11. Харченко В.С. Гарантоспособность и гарантоспособные системы: элементы методологии // Радиоэлектронные и компьютерные системы. – № 5 (17). – С. 7–19.

Надійшла до редакції 22.02.2008

Рецензент: д-р техн. наук, проф. О.Є. Федорович, Національний аерокосмічний університет ім. М.Є. Жуковського «ХАІ», Харків.