

## **CONCEPTION OF STATE INFORMATION RESOURCES STORAGE OF AZERBAIJAN**

**Ismayil Ibragimov<sup>1</sup>, Geray Kengerlinskiy<sup>2</sup>**

Institute of Cybernetics of ANAS, Baku, Azerbaijan

<sup>1</sup>kengerli@mail.ru, <sup>2</sup>kengerli@rambler.ru

In [1] the interim total (2002-2007) of realization of National strategy of IKT in Azerbaijan is summed up, indubitable successes are marked, unsolved questions concerning organizational-methodical and normative-legal side of the matter are revealed. At the same time, development perspectives of IKT make relook through and some technico-technological aspects of creation, especially of IK-infrastructure of the republic and first of all, the storage system of state information resources. Its reason is for the tendency as a whole, in accordance with which the growth of data volume subjected to the storage passes ahead the rate of cost reduce of volume unit of stores (ribbon drives, and also magnetic and optic disks). The break between them increases monotonily, and the expenses for the data storage systems (DSS) increase, and the personnel expenses connected with the storage controlling reach 40-80% of total operating costs. Otherwise, the reserves of extensive the storage volume widening are reduced, and the only opportunity for the present day the reduction of the indicated expenses is connected with the productivity increase of using the storage resources. And it is doubtful that this tendency won't pass by Azerbaijan. On the contrary, taking into consideration the present rate of IKT we'll have to come into collision with this in the nearest future. Because of this the key task of information of the republic becomes the formation on the conceptual and the realization on the practical level of the common strategy of the store creation of state information resources, oriented to the perspective from point of view effective use of storage resources, the technology of massifs and data control.

It is known [2. p 30] that in the medium Unix/Linux it is available to operate in fact 35-40% of the nominal volume of carriers, but in the medium Windows-20-40%, and only 20% of stored data makes interest for the users, and the rest hopelessly became old, and nevertheless are subjected to the full reserve copying and resuming. That is what caused the forcible argument for processing and practical realization (as a dominant of common strategy of the development of DSS) of the ideology of the effective storage the base component of which are virtualization, information lifecycle management (ILM), automatization of data storage controlling and integration of various carriers in the frame of heterogeneous medium. However, but only the virtualization and ILM are, strictly saying, oriented on the task solution of the increase of storage productivity. Integrated heterogeneous medium creates only different variants, necessary for the choice of optimal solution in realization of virtualization mechanisms and ILM, but the automatization releases the personnel from performing the routine operation of administrating.

The idea of virtualization arose to the answer of the demand of the main principle of working out of the data: the submitted resources should be corresponded to the task. But in realization in the frame of the traditional approach stipulating as the initial division of all existing resources proceeding from maximum necessities of supplements and further strict fixing "own" share for each of them, the productivity of use of the resources leaves much to be desired. The virtualization stipulates the opposite approach to the problem, and just-the initial consolidation and the collective use of the resources by means of their dynamic distribution among the suggestions, proceeding from the current necessities of the last.

The virtualization procedure used to DSS (data storage system) provides automatic representation of physical carriers on the logical level not connected with the data displacement on the concrete carriers, that's, creates metadata, with which one can control as a single virtual resources, and is realized by three ways. In the first of them the representation is realized by the volume managers in the operating systems, and the application transparently

mount logical volumes on the storage equipments. The second is organized on the storage level and is oriented for the use of the built in functions of hardware of the disk massifs and net controllers. And at last, the third way is built in the net level due to the formation software and hardware "interim layer" of virtualization. And for the technologies of virtualization it is practically indifferent if they deal with the stores of one producer or various, with uniformal and different equipment; if DSS is one or multilevelled, and migration mechanism volumes support data exchange between DSS without stopping the operating in spite of the difficulty of putting into complicated and valuable technology demanding a higher level of preparation of administrators, the sphere of usage of virtualization is widening, as far as in all additional excesses of putting it into the systems, the store volume of which is hundred terabyte, allows to reduce the expenses for about 24% and more than 16% for the equipment and software, correspondingly, and administration costs for 20% cheaper [3. p 42].

In contrast to ILM virtualization is a particularly specialized conception, the field of appendix of which doesn't exceed the limits of data storage. The essence of the matter is in the thesis: the actuality of the information from the creation moment to the destruction must be correlated with expenditure for storing. The logic of such an assertion is unstained. As 20% of the stored data is actual, and they just must be located on the fast, though valuable carriers, and while the rest 80% can quite be satisfied with less faster, but on the other hand with rather cheaper massifs. And the term "information actuality" is treated in the broad sense, designating, depending on the context, its status, claiming, importance, value, significance... The actuality reduction is associated with the duration of information storage, with the change of the priorities of appendixes or storage regulations.

The procedure of the realization of the idea of ILM brings to four consecutive stages:

- the formation of a code of signs, characterizing the actuality of the information and the classification of the initial massif;
- the task of rules of the information displacement onto those levels, the expenses for storing in which correspond its actuality;
- displacement and placement of the information on the level, the expenses for storing in which reply to the set for it the rule;
- elimination of the information according to the set rules.

In such generalized interpretation it differs in virtualization and multistaged storage medium, with the possible exception of specific rules of information transformation set out on the basis of fundamental understanding of information actuality. The possibility of sizing of this qualified understanding, set in the conception of ILM, promotes more rational, in comparison with the virtualization, and even optimal, for example, as a criterion of the minimal total expenditure, the use of storage resources. From this point of view, the virtualization is only the stage preceding the use of ILM. But, on the other hand, one shouldn't forget that ILM isn't more than a conception realized by the technologies of virtualization.

In spite of indubitable priority, allowing to optimize the expenditures, to increase the flexibility and storage controlling, to regulate the work with the information, DSS, realizing the conception of ILM in the total volume haven't become the mass product yet. The essential reason consists of the active propaganda of this idea that is not supported by ready-made decisions enveloping of the stages of ILM: the classification and the task of rules are performed automatically for the documents; but for the e-mail these procedures are only automatized, that's the part is done manually. The second, but one of the most important reasons is a high cost, reaching hundred thousands dollars for larger decisions. Such expenditures are economically realized in creation or modernization only large (with the storing capacity of no more than 20 T byte) and dynamically growing (the growth of data volume no more 50% a year) storages. And one more deterrent one is the lag of normative-technical provision on rate of the realization of ILM ideology in the concrete items.

Comparatively not long ago, the heterogeneity was considered as unwished, but inescapable caused, mainly to preserve maximally the embedded investment in the

infrastructure in modernization or placement of the resources. However, as far as the development of virtualization technologies, various stores that Possess different resource capacity, productivity with mutually completing functional opportunities and, the main, greatly differing in cost were purposely brought in composition of DSS. And it always allows to have "under hand" the device with corresponding, from point of view of effective usage, opportunities. For example, the price of storing 1 Z byte information on the tape costs only 0,5-3,0 dollars. The price of 1 K byte: of a disk massif of a corporative level increases to 40-70 dollars, of a disk massif of an average level and optic disks varies within 15-35 dollars; but the price of the cheapest disk space varies within 3-15 dollars. [2. p 39]. On the other hand, the expensiveness of magnetic and optic disks are compensated with a quick access to the data and with a low cost of the equipment. All of these gives an opportunity in each concrete case to choose the optimal variant of displacement of the information on various stores depending on its actuality. And there summed up all the precondition for integration such stores onto multilevelled heterogeneous medium using the ILM ideas and virtualization mechanism, and finally the minimization of total expenses for information storage.

Such opportunities are fully realized by three-levelled heterogeneous medium that foresees the consecutive automatic displacement of the data according to the scheme D<sub>2</sub>D<sub>2</sub>T (Disk-to-Disk-to-Tape) and includes:

- quick and expensive disk massifs on senior level;
- less quicker, but cheaper and two times more capacious than the valuable ones in capacity, disks of intermediate stores of average level;
- slower and the cheapest ribbon stores of junior level.

It is obvious, that the optimization opportunity increase by growing the level of the heterogeneous medium, the integration of which is integrated in the broad sense, as far as involves the decisions and the products of various firm-producers including already set in the operating systems. And so the paradigm of the organization of data storage comes to be the transition, equipment and software comprising the whole-the net integrated infrastructure of the storage.

In Azerbaijan the storage of numbered state information resources is dispersed to the data base of authorized information systems, for which comparatively little for present measures, capacity and little (for technical and organization reasons) efficiency of the storage are typical. Meanwhile, one can bravely forecast that expected in the nearest future our republic's becoming the member of World Trade organization, the planned transition to the system of insurance medicine, widening the sphere of usage of "the single window" in serving the population and etc., will set off an explosion in information.

The volume, composition, data expiration date will increase for some orders. Especially it concerns archive, normative and inquiry information, subjected to the storage in new for the republic, digital format, the volumes of which will reach hundreds of Pet byte, but the duration of the regulated storage is tens years. That's why one must not rely on the splicing of store capacity of authority system for location of additional volumes of national resources. Rather even, on the contrary the fragmentation will make difficulties for the integration of authorized data base and joint usage of their resources at interdepartmental interrelation, will bring to the decrease of total efficiency of storage due to the information duplication.

The solution of such a task is feasible only for the centralized storages, the productivity and capacity of which provides the profitability of expenses on the use in total volume of all the advantages of rather expensive technologies of effective storage, as virtualization, ILM, brief information including reduplication of data. Such storages form the nucleus of net integrated infrastructure of the storage, and the located in them national information resources should be oriented on one or a group of subject branches. And the uniformed data of some sources are previously integrated (are coordinated with

indicators, units of measurement, structure and physical attributes), are transformed (are brought to a single sight) and aggregated. The non-contradiction of (per each moment of time) located data is provided by chronology support. All the base functions of the storage are restricted by importating the initial data (after coordination and verification) and by access to them, that's in future these data aren't renewed, remain unchangeable and used only in the regime of reading.

And correspondingly the status of such storages will change. From organizational point of view they will become autonomous (interdepartmental) and will be specialized not only for suggesting the stored the total governmental information resources, but also later the storage resources at the disposal of the clients either it is an organization or a citizen. However, in any case the services should be done commercially, and the service fare isn't performed according to mythically said "spent machine time", but according to the volume of concrete automatically followed and tariffed storage resources in the course of performing each order. In the technical plan, the centralized storages (differing from the stores of departmental systems, being, in fact, the servers appendage) should be considered as independent object of the information infrastructure like the centers of data processing and telecommunicational system.

Taking into consideration that the realization of rather a large project will start from a zero level, it is necessary to invest means to the most contemporary decisions, orienting on flexible and consolidated net heterogeneous medium with dynamic revealing the storage resources, with their effective use and splicing. Centralization storing of the total governmental information resources will provide, with the interdepartmental interrelation, joint use of full, uncontradicted and reliable data, will allow to eliminate the inlet duplicating and the storage of the uniformed information in the departmental bases and will increase the total efficiency of the storage.

### **Literature**

1. Ibraghimov I.A., Kengherlinskiy G.A. "Realization of information projects of Azerbaijan", successes and studies of elementary stage, *Izv.NANA, series phys.-tech. and math. sciences*, vol.XXVI, 2007, No. 2-3, pp. 127-133. (in Russian)
2. Zhilkina N. "Data storing on the tape recorder". *Journal of net decisions LAN*, July, 2005, pp. 30-39. (in Russian)
3. Orlov S. "Virtualization of new ware". *Journal of net decisions LAN*, July, 2006, pp. 32-42. (in Russian)  
[www.osp.ru/lan/2006/07/2549614/](http://www.osp.ru/lan/2006/07/2549614/)