State of current data storage market and development of tools for increasing data storage systems reliability

Marina Bolsunovskaya^{1*}, *Svetlana Shirokova*¹, and *Aleksandra Loginova*¹

¹Peter the Great St. Petersburg Polytechnic University, 195251, 29, Polytechnicheskaya str., St.Petersburg, Russia

Abstract. This paper is devoted to the problem of developing and application of data storage systems (DSS) and tools for managing such systems to predict failures and provide fault tolerance specifications. Nowadays DSS are widely used for collecting data in Smart Home and Smart Cities management systems. For example, large data warehouses are utilized in traffic management systems. The results of the current data storage market state analysis are shown, and the project the purpose of which is to develop a hardware and software complex to predict failures in the storage system is presented.

1 Introduction

1.1 Smart city data warehouses

Nowadays there are emerging more and more examples of “smart cities” information infrastructures, both in Europe and in Russia. At the beginning of 2019, the National competence center “Smart city” was established in our country. However successful projects of “smart cities” systems management appeared earlier, one example is the data processing center for “Smart city” in Nizhny Novgorod (since 2018), the volume of data storage of which is up to 900 Petabytes [1].

Automated traffic management systems, security systems, energy systems and others are an important part of the “smart city” infrastructure. Such systems require collecting of large amounts of data in real time, therefore, the problem of

---

* Corresponding author: marina.bolsunovskaia@spbpu.com

development data warehouses for the needs of “smart cities” and ensuring the highest standards of reliability and performance of such systems becomes urgent.

1.2. The idea of a software product for predicting failures in data storage systems

When managing a “smart city” and “smart home”, in particular, it is necessary to ensure the correct operation of electronic devices and successful coordination of the software and hardware complex. To ensure the correct functioning of the smart home, high reliability of data storage devices is necessary. This is important because the decisions are made on the basis of this data, including decisions which are made in real-time mode.

In the field of improving the reliability of data storage systems (DSS) and the prevention of critical situations by predicting failures in the operation of DSS a comprehensive solution that meets all market requirements currently does not exist [2, 3, 4].

To solve the problem of increasing the reliability of data storage systems, a software product is being developed aimed at solving the following tasks:

1. data redundancy management;
2. integration with the data consumer security system;
3. ensuring a guaranteed level of consumer access to data;
4. reduce the amount of transmitted and stored data;
5. storage of data objects with preservation of their original security attributes;
6. irrevocable guaranteed data deletion;
7. management of long-term storage and disposal of data.

The program is designed to collect the values of the storage system parameters by polling the storage software services and parsing system log messages. The collected parameter values are placed in the local database. The values of the parameters are obtained with a certain periodicity or upon request at each of the nodes included in the storage cluster. Launched instances of the program on each node of the storage cluster synchronize the accumulated data so that the collected values of parameters are placed in the database of each node for all nodes included in the cluster.

2 Justification of the need to develop a hardware and software complex for predicting failures in data storage systems

2.1. Data storage market overview

Analysis of the research conducted by IDC Perspectives showed that the data storage market accounts for approximately 23-25% of all IT infrastructure costs, and having a continuous upward trend. According to The InfoPro, Wave 11, the increase

in storage costs for medium-sized companies exceeds 50% per year. Only in 2016, the total profits of the leading suppliers and manufacturers of external storage systems – companies EMC, HP and IBM (occupying about 80% of the market) – reached USD 119 million, according to IDC. These data testify to the relevance of developments in the field of creating technologies and hardware and software systems for storing and processing information and the need for their continuous modernization and the additional costs associated with it. Because of this, the organizations operating these systems must not only build the storage infrastructure, but also reduce the costs of such upgrading, thus increasing the cost-effectiveness of owning the storage system, reducing its power consumption and service costs. The growth of data volumes, the increased requirements for storage reliability and speed of access to data imply the need to consider storage as a separate data center subsystem.

The main trends of the recent years could be distinguished:

1. The development of storages goes into the stage of deeply integrated nodes and perfect structures. The best examples of which are integrated with processors, memory of different levels and ultra-fast interfaces (propagation delay times of signals of a unit of picoseconds) Blade Systems type complexes. In fact, they are functionally complete “boxed” systems, of which storage and processing centers of any size can be created on the basis of SUN networks.

2. It should also be noted the steady growth of super-large databases and data storages of thousands of terabytes. This is a consequence of the accumulation and storage of huge amounts of data, as well as the need to fulfill regulatory requirements to increase the shelf life of control copies.

3. It is possible to note the high complexity of queries, which include analytical, operational and transactional loads in the form of point samples, lookup tables, loading and parallel processing of transactions and performing analytical tasks. According to analysts, more than 20% of the analytical tasks are loaded in the last 15 minutes and, therefore, they must be processed fairly quickly. This circumstance, in turn, tightens the requirements for reactivity, i.e. temporal characteristics of data processing in storage.

4. There is a noticeable development of the need not only for the acquisition of storage by corporate clients, but also in strict accounting, auditing and monitoring the use of expensive resources. These requirements, to a certain extent, are in contradiction with the requirements of the tough need to balance the costs of the information infrastructure and the actual effect that they can potentially bring in the business to their owners. Therefore, the owners of such systems expect not only an abstract increase in potential performance and a reduction in the total cost of ownership, but also very real opportunities with increasing load, the possibility of simplifying the deployment and administration of systems and, of course, the cumulative reduction of storage of new data units.

5. An important technological trend for the industry has been the creation of adaptable platforms for solving various analytical tasks, which include hardware component and DBMS. End users view the data warehouse as an information

service (for example, banking systems, multimedia information storage systems, etc.).

6. There is a noticeable trend of data migration to cloud resources, which is directly related to the growth in scale and increase in the increased capacity in the storage segment. The products of the world's leading manufacturers account for the bulk of sales of external disk storage systems, both in quantity and in money terms.

There are leading global storage developers on the market: Dell (PowerVault, EqualLogic, Compellent); EMC (Celerra, CLARiiON, Centera, Iomega, Symmetrix VMAX/VMAXe, VNX, VPLEX); Fujitsu (ETERNUS, FibreCAT, Quantum); Hitachi (SMS, WMS, AMS, CAP, USP, VTL); HP (StorageWorks, EVA, 3PAR); IBM (Storwize, System Storage D, System Storage T, System Storage N, SONAS); NetApp (FAS, VTL, V, E); Oracle (Sun ZFS Storage, Sun Flash, Pillar Axiom, StorageTek).

Since storage systems are inseparable from computing resources, it is not surprising that many of the world's largest manufacturers of storage systems are also leaders in the server market. Of the manufacturers listed above, only three are primarily engaged in the development of storage systems – EMC, Hitachi and NetApp. From manufacturers of storage systems represented in the Russian Federation, we note companies that belong to the class of integrators: AXUS, Buffalo, Cisco (Linksys), DLink, Dot Hill, Infortrend, Intransa, Maxtronic, Nexsan, Overland Storage, Plasmon, QNAP Systems, SGI, Thecus.

2.2. Current state of the Russian market of data storage systems

The Russian market of data storage systems is developing extremely dynamically. According to IDC, in 2016, data storage systems with a total capacity of up to 663,002 TB for a total of USD 382.77 million appeared and operate in the Russian market, that is, this market grew by 35.7% in capacity and 0.5% in de-tender terms. In the first half of 2017, external data storage systems with a total volume of more than 230.2 PB, the integral value of which exceeded USD 126.77 million, were put on the Russian market.

Note that this process is sustainable due to the explosive growth of data volumes, and the fleet of new systems simply replaces outdated and inefficient old storage systems. Huawei showed the maximum growth in deliveries in monetary terms in the first half of 2017 with an increase of 242.4%, it also became the leader in growth in total supplied capacity – by 45.8% compared with the results of the first half of the year earlier. In addition, it should be noted that the majority of Russian companies purchasing external storage systems are guided by the significant technological advantages of new systems built exclusively on flash memory arrays without the use of hard drives. Deliveries of such systems also increased by 86.5% in monetary terms, and their share was about 24% of the market in the first half of the year. The rapid growth of data is increasingly forcing domestic companies to acquire external disk storage systems. This is largely due to the traditional trend of reducing the cost of IT components. If earlier external storage systems were

perceived only as an attribute of large organizations, now even small companies do not reject the need for these systems. At the same time, IDC noted the changed preferences of customers who bought the systems of the middle price segment much more often than the initial and higher systems. The top five largest suppliers for the year included Dell Technologies, Hewlett Packard Enterprise, Hitachi Data Systems, IBM and NetApp. Many domestic manufacturers of disk storage systems (for example, DEPO Computers (DEPO Electronics)) build their systems based on components from foreign manufacturers, including Microsemi (formerly Adaptec), Chenbro, Falconstore, Intel, LSI Logic, Luster and others. In general, it can be noted that Russian-made storage systems are supplied predominantly for Russian enterprises.

3 Overview of the suggested approaches to building a hardware and software complex benefits

3.1. A competitive advantages of the proposed software product

The significant competitive advantages of the software products developed for Russian market are: cost, reliability of the finished product, independence from the western markets, including spare parts markets and markets of analogous products.

It is an important fact that the hardware and software complex to predict failures in storage systems presented in this article is based on the new mathematical models of DSS reliability:

1. a new estimation method based on the Huygens theorem is presented;
2. a mathematical model of scrubbing is proposed, taking into account unresolved bit errors and the process of continuous verification of checksums. Based on the principles of simulation modeling, a computational model of failures in an ultra-large distributed data repository was developed that takes into account hidden bit errors, data checking and recovery, designed to statistically describe the working properties of the repository, including the operation time before data loss.
3. based on the Markov chain theory, a family of new analytical models was developed to assess the reliability of large data storage systems;
4. a qualitative coordination of the reliability estimates calculated by the developed Markov chains analytical model with full-scale imitational calculations is shown, the following properties of the analytical model are confirmed: adequacy, versatility, accuracy, efficiency;
5. developed practical recommendations in the development of methods, algorithms and technologies to improve the reliability and efficiency of storing large amounts of data.

The results of the project are designed to develop basic software technologies to improve the reliability of DSS and the subsequent commercialization of these technologies as part of innovative software solutions for managing storage and ensuring the reliability of storing large amounts of data.

The product being developed is a software package for predicting storage system failure (DSS). This software package is designed to diagnose and predict the status of storage systems and their components in real time.

3.2. Tools for improving a storage system reliability

Describing the storage management software market (including forecasting and evaluating their performance and reliability), it is important to note that the design of reliable drives has been solved for decades by large storage companies such as TDK Corporation, Seagate Technology, Hitachi Global Storage Technologies and others [5]. Research groups in different countries around the world are working on the problems of predicting disk failures in large-scale data storage systems, and models of storage reliability have been proposed [6-8].

The most well-known solutions aimed at improving the reliability of storage are duplication (mirroring) of data, multi-level redundancy coding for digital data archives, deduplication systems for parallel data backup [9,10]. The nature of problems in storage systems is studied taking into account the age of the system [11]. The concept of "system survivability" in relation to storage systems is becoming popular [12]. At the same time, the issue of storage power management is being studied [13].

The software package for predicting storage failures is designed for solving such storage management tasks as [7]:

- 1) transmission of telemetric information on the state of the storage system in real time;
- 2) detection of failures, failures and critical situations in storage based on the proposed classification algorithms for failures, failures and critical situations;
- 3) prediction of faults, failures and critical situations during storage operation;
- 4) decision support for the prevention of critical situations in the storage, implemented in real time;
- 5) provision of a wide range of health zones of the developed decision support system.

Potential similar software systems are:

- 1) enterprise-producers of storage systems and their components;
- 2) enterprises using storage systems of various architectures;
- 3) scientific and educational laboratories.

Currently, the leading manufacturers of data storage systems in the world are: Dell EMC, IBM, Hewlett Packard Enterprise (HPE), NetApp, Pure Storage, Hitachi Vantara, Huawei, and others. Among the foreign vendors named, HPE are the leaders in the Russian market. Hitachi and Huawei. Among domestic developers and

integrators, the following should be noted: Kraftway; Yadro (KNS Group LLC, as part of the National Computer Corporation, NCC); MCST; INEUM named after I.S. Brooke; T-Platforms; National Center for Informatization (NCI); “National Technologies”.

Features of the program: use of the concept of “system survivability” and models of system dynamics in the development of algorithms. Technical characteristics of the developing complex: platform consists of a set of storage controllers, a PCI-express factory (PCIe-factory controllers), a disk chassis and disk drives. It is possible to create a system for storing and managing large (from 1 PB) and super-large data volumes (from 10 PB); scaling distributed storage; providing adjustable redundancy on the client side. Uninterrupted client access to storage is supported (hardware-software storage management system is designed to ensure availability of at least 99.99%).

4 The main trends of the storage systems market

Despite the presence of recognized leaders in the Russian market supplying storage systems and their components, it can be noted that the market continues to saturate (there is a growing demand for storage systems for Russian companies due to the growth of volumes of corporate data presented in digital form), which contributes to increasing competition between manufacturers and suppliers of storage solutions.

Consumers of storage systems are enterprises in various industries (both manufacturing and services, including the banking and insurance sector). Competition between storage vendors involves finding more reliable, flexible, and relatively inexpensive data warehouse design solutions.

At the same time, there is a tendency of growth in the need for storage capacities of large organizations. According to experts, in such organizations the cost of data storage systems can be up to 25% of the cost of IT infrastructure. At the same time, the needs of companies in increasing storage sizes (and, consequently, the total cost of storage) continue to grow. One of the reasons is a change in legislation, for example, the emergence of laws obliging to store data. In Russian practice, such laws appeared relatively recently and entailed changes in the IT infrastructure of companies.

But in order to reduce the cost of owning an IT infrastructure, some companies are turning to new types of services when building storage systems. An important trend of the modern storage market is a massive transition to the use of cloud technologies, including such forms of service as SaaS (software as a service); also software on demand.

An example of the implementation of cloud technologies in the field of storage are the so-called software-defined storage (eng. Software-Defined Storage, SDS). This method of implementing storage systems was proposed in the early 2000s and is now becoming increasingly popular, as it can significantly reduce the cost of IT infrastructure. The Storage Networking Industry Association (SNIA) defines SDS as a virtualized storage environment with a service management interface, which

should include automation, standard APIs, virtualization of data access paths, and ensure the scalability of IT infrastructure and its “transparency”. To implement SDS, a standardized management interface is needed – the SNIA Storage Management Initiative Specification (SMI-S), which is part of the concept of software-defined data centers (SDDC). The use of the SDS concept gives companies such advantages as: abstraction from the lower level (hardware platform), scalability, simplified storage infrastructure, relatively low cost of solutions. According to Gartner, by 2020, 70–80% of unstructured data will be stored on systems managed by SDS, and by 2019, 70% of existing storage arrays will be available in a fully software version.

Currently, software-defined storage on the Russian market is offered by many vendors, including Dell EMC (Dell Nexenta, EMC ScaleIO solutions), HPE (StoreVirtual VSA solution), IBM (Spectrum Storage solution), NetApp (ONTAP Select solution), VMware (solution vSAN), Red Hat (Red Hat Storage solution), StoneFly (SCVM, SDUS solutions), DataCore (SANsymphony solution), etc. These solutions can be grouped into three categories: “classic” (CEPH, Red Hat Storage Server, EMC ScaleIO), based on traditional storage systems (NetApp ONTAP Select, HPE StoreVirtual VSA), solutions as part of computer systems (VMware vSAN). Some manufacturers, such as Huawei and Dell EMC, offer both integrated SDS solutions and software SDS.

In 2016-2017 The global storage market as a whole showed insignificant fluctuations in demand, but the structure of corporate customer spending on storage implementation was changing: the cost of traditional external arrays continued to decline, while the costs of implementing all-flash systems (based on flash memory) showed significant growth. In the first quarter of 2017, server solutions accounted for \$ 2.7 billion in revenue in the market (13.7% less than in 2016). At the same time, the total SHD revenues of ODM manufacturers (supplying equipment directly to data centers) increased by 78.2% in 2017, exceeding \$ 1.2 billion. The share of these companies in the market reached 13.2%.

Software (software) for storage management should support flexible organization of data storage, deduplication and data replication, dynamic memory allocation, file system snapshots.

Any storage system involves the use of software that supports its work. Other examples of storage management software with an extended feature set include:

1) CloudIQ is a free application for users of Dell EMC mid-tier storage systems operating in the form of SaaS. This application is a cloud service, thanks to which users can get intelligent analytical data on the general state of the storage system (a comprehensive assessment of the state of systems without interruption), including predictive analytics;

2) VMware IT Business Management Suite (developed by VMware Corporation). The average price on the Russian market is from \$ 95,000. The software package is released in three editions (Standard, Advanced, Enterprise) and is available as software installed and managed inside the organization or in the form of SaaS (software as a service). The software package provides control over the costs and quality of the IT infrastructure (it is possible to estimate costs, quality, user composition and consumption), as well as IT infrastructure transformations. The

product provides the analytical data necessary for understanding the return on investment and total cost of ownership of projects.

5 Comparative analysis of the technical characteristics of the proposed solution

A comparative analysis of the basic technical and cost characteristics of storage systems of foreign vendors that are popular in the domestic market has been carried out. The cost of storage systems varies greatly depending on their functional capabilities, technical characteristics and scale of business for which these solutions may be suitable. In general, in Russia, the storage market behaves stably, but, as in the rest of the world, there is a decline in demand for traditional disk storages in favor of solutions on flash drives.

Among the competitive advantages of products created with the use of RID ("Program for collecting parameters of the data storage system") should be noted:

1) the compliance of decisions with the requirements of modern Russian legislation;

2) high reliability of solutions;

3) use of original data protection mechanisms;

4) independence from the western market of spare parts and analog products;

5) the relatively low cost of solutions, and therefore their availability to a wide range of users.

In general, large Russian storage manufacturers are now able to compete with foreign vendors. A national policy aimed at import substitution and changes in domestic legislation will also contribute to the gradual growth of domestic developments among Russian users.

Table 1 contains the results of the SWOT analysis studies of the state of the internal and external environment the Russian storage systems developers. It is important for Russian developers to maintain relations with the national scientific community, including creation of clusters with participation of academic institutions and industrial partners.

Table 1. SWOT analysis of the state of the internal and external environment of Russian storage systems developers.

Strengths	Weakness
1. Relations with the scientific community and the availability of scientific reserve 3. Having contacts with state organizations, ministries and departments 4. Availability of own developments, confirmed by patents and certificates 5. Availability of necessary resources for the production of high-tech products 6. Well-established distribution channels for products 7. The average age of employees - about 30 years.	1. High production cost 2. Difficult technological cycle and high requirements for personnel training 3. Entry into the market with a new product under development
Opportunities	Threats
1. Increasing demand on DSS in Russia 2. State support 3. Information technology development 4. The presence of effective demand in the market	1. Strong competition from foreign storage manufacturers 2. Changes in domestic legislation 3. The decrease in effective demand 4. The emergence of resource-intensive technologies in the field of DSS

6 The developed project of the program complex developing for forecasting of data storage system failures

The program complex for forecasting of data storage system failures has been developing by the scientific group from Peter the Great St. Petersburg Polytechnic University. The program complex was designed to diagnose and predict the state of DSSs and their components in real time [14]; the model of the complex was based on the idea of the system dynamics model.

The program complex was initially designed for DSSs running on the platform “YADRO TATLIN” in various configurations. Structural scheme of the program complex for predicting DSSs failures and other characteristics of the complex are presented in Reference.

7 Conclusion

Since the storage and associated software is expected to be released in several versions, it is planned to use a flexible pricing system: for enterprises of different scale, the capabilities of the storage system and its price will vary. There is also the possibility of expanding the system at the request of the user.

Based on the competitive advantages of the products identified in the work, the results will be used to conduct theoretical and experimental research and further development of the mathematical models developed at the first stage, the diagnostic

algorithm and the stands for software testing to ensure the development of predictive analytics algorithms for storage systems data of various configurations, taking into account the variability of system parameters and modes of operation, the models are increased I reliability storage systems and hardware and software to predict failures in the storage system.

The developed models and the diagnostic algorithm will be used when conducting tests of manufactured experimental samples of DSS for compliance with the requirements.

Developed stands in conjunction with the design and software documentation developed for them should be used when conducting experimental studies of software tools for predicting and simulating data storage systems and software systems for improving data reliability in the “smart home” system.

The presented technical and economic assessment shows the possibility of achieving an exceptional position in the markets for goods (services) or technologies of products created using the developed software for the “smart home”.

The research is carried out with the financial support of the Ministry of Science and Higher Education of Russian Federation within the framework of the Federal Program “Research and Development in Priority Areas for the Development of the Russian Science and Technology Complex for 2014-2020”. The unique identifier is RFMEFI58117X0023.

References

1. V. Ya. Kopp M. V. Zamorenov Yu. E. Obzherin M. Yu. Larin *S.P.P.U.J. Engi. Sci. Tech. Inform.Telecommu.Contr.* ISSN 1994-2354 **3** (247) 23-34 (2016)
2. I. P. Simakov, P. V. Kholodnykh, ITM-2016 244-254 (2016)
3. R. Wood, J. Magne. Magne. Mater. **321**(6), 555–561 (2009)
4. Z. Bingpeng, IPDPS, 1–5 (2013) DOI: 10.1109/MSST.2013.6558427
5. V. Venkatesan, *Reliability Analysis of Data Storage Systems: These pour l'obtention du grade de docteur es sciences* (2012)
6. V. Venkatesan, I. Iliadis, IPDPS, 209–219 (2012)
7. Y. Tianming, IPDPS, 1–12 (2010) DOI:10.1109/IPDPS.2010.5470468
8. A. Wildani, Proc. Internat. Symp. Mod., Anal. Simu. Comp. Telecom. Syst. 1–11. (2009) DOI: 10.1109/MASCOT.2009.5366825
9. T. J. E. Schwarz, E. L. Miller Proc. Internat. Symp. Mod., Anal. Simu. Comp. Telec. Syst, 125–134 (2005)
10. J. J. Wylie, Sch. Compu. Sci. (2001)
11. Abdullah-al Mamun PACC, 3206 – 3211 (2016)
12. A. S. Makarov, M. V. Bolsunovskaya, S. V. Shirokova, Proc. Int. Conf. SCM-2018, 61-64 (2018)
13. V. N. Volkova, A. A. Denisov *Systems theory and systems analysis* (2017)
14. M. V. Bolsunovskaya, S. V. Shirokova, A. V. Loginova, M. B. Uspenskiy Proc. 2nd Inter. Scie. Confe. DTMIS (2018) DOI: 10.1088/1757-899X/497/1/012024