UDC 519.816
DOI: 10.15587/2312-8372.2019.182622

Mishchenko V.

DEVELOPMENT OF A METHODOLOGY FOR ASSESSING RELIABILITY OF DATABASES USED IN RADIO NETWORK MANAGEMENTS

У роботі проаналізовано задачі, які виконуються під час обробки розподілених даних. Об'єктом дослідження є процес автоматизації управління у розподілених системах та організація обробки розподілених даних. У ході дослідження використовувався метод реплікації даних, який передбачає їх дублювання в різних вузлах інформаційно-аналітичної системи. Це пов'язано з тим, що зазначений метод дозволяє привести декілька баз даних з однаковою структурою в один і той же несуперечливий стан, який супроводжується взаємним внесенням змін. Завдяки цьому забезпечується ефективне інформаційне та інтелектуальне (аналітичне) забезпечення територіально розподілених, але функціонально взаємозалежних груп посадових осіб в процесі вироблення ними управлінських рішень. Аналіз етапів елементарного циклу управління надскладними системами – збір даних обстановки, вироблення рішення, генерація управляючих команд компонент мережі – виявляє загальну для усіх етапів особливість. А саме – усі вони є складними інформаційними технологічними процесами (перші три – обмін інформацією в системі передачі даних та її обробка в системах управління даними, останній – зміна параметрів мережі генерації управляючих команд компонент мережі). Висока швидкість протікання зазначених процесів і логічна складність їх структури призвели до необхідності автоматизації управління даними процесами, коли значна частина інформаційних функцій центру управління передається технічними засобами – інформаційно-логічними пристроями автоматизованої системи управління. У результаті проведення дослідження отримані експериментальні результати, на підставі яких можливо значно підвищити рівень функціональної надійності баз даних інформаційно-аналітичних систем. Також стало можливим зменшити час на обробку та синхронізацію розподілених даних інформаційно-аналітичних ситем. Результати дослідження мають використовуватися під час розрахунків функціональної надійності розподілених баз даних інформаційно-аналітичних систем, інформаційних та телекомунікаційних мереж. Планується використовувати зазначену методику під час побудови великих масивів розподілених баз даних у інформаційних мережах, побудованих на сучасних безпроводових радіотехнологіях.

Ключові слова: *інформаційна мережа, технології реплікації даних, система управління базами даних, обробка розподілених даних.*

Received date: 12.06.2019
Accepted date: 10.07.2019
Published date: 31.10.2019

1. Introduction

Currently, the information infrastructure of modern enterprises and organizations is becoming more and more distributed. The level of managerial decisions made, control and management of information resources requires increasing decentralization [1–3]. The structure of modern organizations is a branched scheme, which includes dozens of geographically separated units, the functioning of which without each other is often impossible at all. Moreover, the quality of functioning of such a complex, corporate system largely depends on the use of the latest information technologies in its activities.

The main problem in the automation of distributed systems is the organization of processing distributed data. The use of data replication technology [4–6] is an important step in the successful work of the organization. The introduction of this technology ensures guaranteed delivery, timeliness and integrity of the transmitted data.

Currently, the development of new approaches to assessing the reliability of databases is relevant. Since this requires:

– the rapid introduction of automated control systems at the beginning of the twentieth century [7–9];
– the emergence, development and dissemination of complex systems with components such as hardware, software, communication tools and people [10];
– means of struggle in cyber space [11].

Thus, *the object of research* is the process of automation control in distributed systems and the organization of processing distributed data. And *the aim of research* is development of a methodology to increase the functional reliability of databases of information-analytical systems by the method of data replication, providing for their duplication in different nodes of the information-analytical system.

 ISSN 2664-9969

2. Methods of research

When conducting research, the following scientific methods were used:
- scientific analysis method;
- generalization method;
- comparison method;
- modeling method;
- forecasting method.

The main purpose of using these methods in research is providing the ability to process data of any structure and complexity.

This approach allows to process and analyze data without reference to their presentation and territorial location. Rapid data output for management decisions by relevant officials.

The forecasting task is in finds in the relationship between the date and resource parameters obtained from the database of the specified structure, based on the analysis of previously obtained data. The dynamics of changes in the produced capacity can be characterized with respect to some basic (usually the first) observation and the magnitude of the change in neighboring levels. The following quantities are used as statistical characteristics of the time series Y_i, $i=1...n$ arithmetic mean $Y=\frac{1}{N}\sum_{j=1}^{N}Y_i$, average absolute increase:

$$Y=\frac{Y_n-Y_1}{N-1},$$

where N – the number of levels of the series; Y_i – the levels of the series [12].

In accordance with the method of checking the materiality of the difference in the average indicators, the initial time series is decomposed into two identical (or almost identical) parts, after which the hypothesis of the materiality of the difference in the average values for the indicated parts is checked. The disadvantage of this method is the inability to correctly determine the presence of a trend in the case when the time series contains a point of trend change in the region of the middle of the series [12].

The lack of trend hypothesis is checked using auxiliary functions:

$$L=\sum_{t=2}^{N}I_t,\ I_t=U_t-V_t,$$

$$u_t=\begin{cases}1, & <Y_{t-1} \quad ...,\\ Y & 0 \quad 1,\end{cases} \quad v_t=\begin{cases}1, & >Y_{t-1} \quad ...,\\ Y & 0 \quad Y_1.\end{cases}$$

The hypothesis of the absence of a trend is not accepted if the calculated t-value is greater than the tabulated value at the selected significance level of 0.95.

Checking data homogeneity is based on the Irwin criterion, which is based on a comparison of adjacent values of the series. In accordance with it, the characteristic t is calculated:

$$t=\frac{Y_t-Y_{t-1}}{Y}.$$

The assessment of the properties is reduced to the study of the autocorrelation function of the initial and difference series. Autocorrelation analysis is performed using a graph and critical values of the coefficients established by experts.

3. Research results and discussion

Since the information and analytical system should understand the interconnected set of information resources, human intelligence, modern means of informatization and information and telecommunication technologies, these resources are combined into telecommunication and computer networks. They are also based on in-depth analysis, widespread use of information and analytical systems for decision support, electronic document management and paperwork and are intended to combine advanced information technologies [13].

Accordingly, the functional reliability of information-analytical systems is understood to mean their ability to provide a certain level of performance when performing certain functions. Effective information and intellectual (analytical) support of geographically distributed, but functionally interconnected groups of officials in the process of developing managerial decisions.

Data replication provides for their duplication in different nodes of the information-analytical system. Moreover, any database (DB) is local (both for the database management system (DBMS) and for the user working with it). That is, data is always located locally on a specific network node, where it is processed, and all transactions in the system are processed and completed locally. The effectiveness of data replication technology in a distributed system depends on what operations should be performed on the system.

Any information-analytical system (IAS) S_{IAS} is described by a relation of the type:

$$S_{IAS}\le T+D+P+U, \tag{1}$$

where T – the set of technical means of the information-analytical system; D – set of system databases; P – set of system software; U – set of users of the system.

The database is an essential component of IAS and the effective implementation of almost all system functions depends on its reliable functioning.

The use of database replication allows to bring several databases with the same structure into the same consistent state, which is accompanied by a mutual introduction of changes.

In [14, 15], replication is defined as the process of generating and reproducing multiple copies of data hosted on one or more sites. In [1], DBMS data replication is understood as bringing the databases functioning in a distributed environment up to date by identifying the changed data, transferring and applying these changes to the recipient database.

The simplest model for assessing the reliability of replicating IAS databases can be represented as a renewable system with a limited amount of spare parts (in this case, database replicas).

Consider a hot standby database system that undergoes a Poisson failure flow with intensity λ.

The specified system has n local copies, that is, replicas of the database, as well as files of changes made (local transaction logs).

In the simplest case, replica recovery is performed using the standard version of the database, and their further update is generated based on the corresponding electronic local change log.

We assume that the time required to restore some database replica using the reference version is much higher than the considered operating interval. Since it is required to transfer large amounts of data (a full copy of the reference database), as well as to make them updated.

In this case, the reliability of the database is determined by the probability that the available number of its replicas is enough to ensure stable operation of the system in the time interval $(0, t)$. Since, in case of failure of any replica of the database, the requests will be redirected to one of the replicas that remained in working condition.

It is advisable to evaluate the reliability of the IAS database as the probability that on the interval $(0, t)$ the number of database replica failures will be no more than $(n-1)$, which is calculated by the formula:

$$P_{DB}(t)=\sum_{k=0}^{n-1}P_k(t)=e^{-\lambda t}\sum_{k!}^{n-1}\frac{(\lambda t)^k}{k!}. \quad (2)$$

The indicated value can be taken as an estimate of the probability of failure-free operation of a system with «hot» redundancy, which contains n replicate IAS databases at time t.

Having designated through P_{DB}^{m} the DB reliability level, it is required as part of IAS, it is possible to determine the number n^m of DB replicas, it is required for its operation with the reliability level P_{DB}^{m} for time t, taking into account expression (2) by the formula:

$$n^m=\arg\min_{n=1,2,\ldots}\left\{\sum_{k=0}^{n-1}\frac{(\lambda t)^k}{k!}\geq P_{BD}^{m}e^{\lambda t}\right\}. \quad (3)$$

Let's consider a much more complicated, but rather real situation, when in the geographically distributed data processing nodes of the IAS, a local backup of the database replica states is performed at some a priori specified time intervals.

At this time, it is assumed that the sequence of taking backups (i. e. points) as a whole for the entire system is synchronized.

In this case, each of the database replicas is restored from its last local backup (LB), and then it is updated in accordance with the local change log.

Further construction of models designed to assess the reliability characteristics of the system under consideration will be based on the formalization of finite Markov chains [16].

Consider the following system conditions:

S_0 – in the system, all database replicas are functioning and available;

S_1 – one of the replicas is in a state of failure (inoperative), it is restored using the corresponding local LB;

S_2 – two replicas in a failed state;

S_n – all available replicas of the IAS database are in a state of failure, they are being restored. It is in this case that it is possible to talk about the failure of the entire IAS database, since none of the replicas will be functional and accessible for users.

As before, let's assume that the system undergoes a Poisson failure flow with intensity λ, and each of the replicas is restored with intensity μ.

The indicated process of state transitions of a complex system is known as the Markov chain of «death-reproduction» [17].

Let's also consider the reduced failure flow rate, which is determined by:

$$\rho=\frac{\lambda}{\mu}.$$

The basic model for calculating the reliability of the IAS database for a graph is determined through the so-called final probabilities of system states according to the Erlang formula [18], taking into account that $\mu_1=\mu$, $\mu_2=2\mu$, $\mu_3=3\mu,\ldots\mu_n=n\mu$:

$$p_0=\left(1+p+\frac{\rho^2}{2!}+\frac{\rho^3}{3!}+\ldots+\frac{\rho^n}{n!}\right)^{-1}; \quad (4)$$

$$p_1=\rho p_0;$$

$$p_2=\frac{\rho^2}{2!}p_0;$$

$$p_3=\frac{\rho^3}{3!}p_0;$$

$$\ldots;$$

$$p_n=\frac{\rho^n}{n!}p_0;$$

$$\sum_{i=0}^{n}p_i=1.$$

The probability of failure-free operation of the considered algorithm for the functioning of the IAS database is calculated using the system of equations (4), which describes the final probabilities of all possible states of the system in the framework of this model:

$$Q_{DB}=1-p_n=1-\frac{\rho^n}{n!}p_0. \quad (5)$$

Thus, this technique allows:

– bring several databases with the same structure into the same consistent state, which is accompanied by mutual changes;

– get a fairly adequate assessment of such important indicators of the reliability of the IAS database;

– provide a high probability of failure-free operation and the likelihood of working with a performance not lower than a given distributed IAS database.

4. Conclusions

During the study, experimental results were obtained, on the basis of which it is possible to significantly increase the level of functional reliability of IAS databases. It has also become possible to reduce the processing time and synchronization of distributed IAS data.

The research results will be useful in calculating the functional reliability of distributed IAS databases, information and telecommunication networks.

It is proposed to use this methodology when building large arrays of distributed databases in information networks built on modern wireless radio technologies.

ISSN 2664-9969

References

1. Date, C. J. (2003). *An Introduction to Database Systems*. Pearson, 1040.
2. Tretiak, V. F., Holubnychyi, D. Yu., Chelenko, Yu. V. (2005). Tyrazhyrovanye dannikh v systeme upravlenyia bazamy dannikh. *Upravlinnia rozvytkom, 3,* 94–95.
3. Tretiak, V. F., Prykhodko, V. M., Holubnychyi, D. Yu. (2005). Aktualnist pobudovy grid system. *Persha naukovo-tekhnichna konferentsiia Kharkivskoho universytetu Povitrianykh Syl*. Kharkiv: KhU PS, 240–241.
4. Tretiak, V. F. (2004). Tekhnolohyia replykatsyy v raspredelennikh systemakh upravlenyia bazamy dannikh. *Informatsiino-keruiuchi systemy na zaliznychnomu transporti, 1,* 7–10.
5. Tretiak, V. F., Holubnychyi, D. Yu., Zolotarova, I. O. (2007). Problemy rozvytku prohramnoho zabezpechennia seredovyshcha rozpodilenykh obchyslen GRID. *Materialy tretoi naukovoi konferentsii Kharkivskoho universytetu Povitrianykh Syl imeni Ivana Kozheduba*. Kharkiv: KhUPS, 7.
6. Golubiatnikov, I. V., Matveev, Iu. V., Sergeev, I. V. (2001). Sozdanie programmnogo kompleksa replikacii SUBD-nezavisimykh dannykh. *Sistemnii analiz, informatika i optimizaciia*. Moscow: RZITLP, 106–116.
7. Liashenko, I. O. (2009). Merezhotsentryzm u viiskovii spravi. *Suchasni informatsiini tekhnolohii u sferi bezpeky ta oborony, 2 (5),* 78–81.
8. Liashenko, I. O. (2009). Evoliutsiia rozvytku kontseptsii vedennia zbroinoi borotby. *Suchasni informatsiini tekhnolohii u sferi bezpeky ta oborony, 3 (6),* 23–25.
9. Liashenko, I. O. (2010). Zastosuvannia perspektyvnykh informatsiinykh tekhnolohii u zbroinii borotbi. *Trudy universytetu, 1 (94),* 144–150.
10. Liashenko, I. O. (2010). Kibernetychnyi prostir – yak evoliutsiia rozvytku informatsiinykh tekhnolohii. *Suchasni informatsiini tekhnolohii u sferi bezpeky ta oborony, 7,* 18–21.
11. Liashenko, I. O., Kyrylenko, V. A. (2010). Kibernetychni operatsii – maibutnia forma zbroinoi borotby. *Zb. nauk. pr. Seriia: viiskovi ta tekhnichni nauky, 53,* 162.
12. Boliubash, Yu. Ya. (2017). *Metody ta zasoby opratsiuvannia informatsiinykh resursiv velykykh danykh v systemakh terytorialnoho upravlinnia*. Lviv, 178.
13. Broido, V. L. (2003). *Vychislitelnye sistemy, set i telekommunikacii*. Saint Petersburg: Piter, 688.
14. Ciciani, B., Dias, D. M., Yu, P. S. (1992). Analysis of Concurrency-Coherency Control Protocols for Distributed Transaction Processing Systems with Regional Locality. *IEEE Transactions on Software Engineering, 18 (10),* 899–914. doi: http://doi.org/10.1109/32.163606
15. Connolly, T., Begg, C. (2000). D*atabase Systems: A Practical Approach to Design, Implementation, and Management*. Addison Wesley, 1236.
16. Kemeny, J. G., Snell, J. L. (1983). *Finite Markov Chains: With a New Appendix «Generalization of a Fundamental Matrix» (Undergraduate Texts in Mathematics)*. Springer, 226.
17. Wentzel, E. S. (1983). *Operations Research*. Mir Publishers, 256.
18. Beiz, G., Kuk, D. (1990). *Kompiuternaia matematika*. Moscow: Nauka, 383.

***Mishchenko Volodymyr,** Department of Radio Reception and Signal Processing, National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Ukraine, e-mail: mischenko_w@ukr.net, ORCID: http://orcid.org/0000-0002-7578-1759*