Scholar Publishing Group
Distributed Processing System
https://doi.org/10.38007/DPS.2021.020204
ISSN 2790-0916 Vol. 2, Issue 2: 26-33

Distributed Computing Model Based on Three-layer B/S Mode

Ciminello Monica*

Warsaw Univ Technol, Inst Elect Syst, Nowowiejska 15-19, PL-00665 Warsaw, Poland

**corresponding author*

Keywords: Three-Tier B/S Model, Distributed Computing, Hadoop Platform, Wavelet Analysis

Abstract: With the rapid development of the Internet, the amount and form of data in various industries are becoming increasingly large and complex. The crazy growth of data volume increases the value of data mining, but also brings difficulties to professionals and scholars engaged in data mining. At this time, it is necessary to divide a computing task into multiple task nodes through a distributed computing(DC) method, and then the computer can greatly improve the computing efficiency by participating in the computing process. In this paper, a DC model based on the three-layer B/S mode is built on the Hadoop platform. In order to verify the performance of the model, the performance of the model is evaluated by indicators such as SR, ACC, and ITR. Through two sets of comparative experiments to analyze the effect of the DC mechanism, the experiments show that the SP value and ITR value of the model under parallel computing are improved, and after the WAFE, the improvement effect of ITR is more obvious.

1. Introduction

Faced with an increasingly large amount of data, researchers not only need to quickly obtain data models, but also need to timely feedback the results and predictions of data models. Only in this way can the value of data be obtained. Data analysis and mining of big data is not a simple data analysis, nor is it a simple data storage problem, but also includes technical problems such as data analysis and big data storage. In this process, researchers should not only pay attention to the algorithms and methods of data analysis, but also make the data analysis process accurate and fast.

There are many applications of DC models, and good research results have been achieved. For example, a scholar used and analyzed some evolutionary algorithms in DC, but did not consider the performance improvement space with discrete characteristics and the impact of performance evaluation with high computational load on the calculation results, so that the solutions obtained by these algorithms did not Well, the corresponding performance optimization process is also quite [1].

In terms of DC framework, some scholars proposed the Apache Spark framework, which was developed by the UCBerkeleyAPM laboratory and is a parallel computing framework similar to the Hadoop computing principle, which is applied to the fields of data mining and machine learning, especially in optimizing the performance of iterative computing. Especially prominent [2]. A scholar uses MPI to study the synchronous and asynchronous cluster computing methods of PSO algorithm. However, due to the shortcomings of MPI's own data storage and processing, which are independent and cannot deal with node failures, the parallel PSO algorithm implemented by MPI model is used. The practicality is much weaker [3]. Although DC makes full use of network computing power, distributed parallel computing needs to have good scalability to improve the performance of parallel computing.

This paper first expounds the basic situation of DC, and then proposes a WAFE algorithm, which can be used to analyze signal data. Then this paper builds a DC model based on B/S, uses the model to conduct signal analysis experiments of DC mechanism, and compares the performance of the DC model before and after parallel computing and before and after WAFE.

2. DC and Related Algorithms

2.1. Overview of DC

Due to the rapid development of science and technology, the amount of data on the Internet has increased exponentially. Using DC technology to process big data samples can better discover the value of data, improve the service quality of software manufacturers and the efficiency of government decision-making [4]. The widespread application of DC solves the problem of massive data processing, and makes full use of the idle processing power of computers. Parallel DC can be achieved through computers, or language programming can be implemented on computers [5].

2.2. Wavelet Analysis Feature Extraction(WAFE)

Wavelet changes can be analyzed at multiple resolutions, so that wavelet transforms can be used in real life. For example, in medical treatment, it can be used to reduce the time of CT or MRI, in image research, it can be used for denoising and compression, and in mathematics. In research, it can be used to solve differential equations, etc. [6]. In this paper, we use it to analyze signals with non-stationary properties. The following is the definition of continuous wavelet transform (CWT), where ε(x) is the wavelet function:

$$W_{\varepsilon} f(m,n) = |m|^{1/2} \int f(x) \overline{\varepsilon(\frac{x-n}{m})} dx,\ f \in K^{2}(R), m \neq 0 \tag{1}$$

In the above formula, the continuous wavelet transform based on ε(x). in:

$$\varepsilon_{m,n}(x) = m^{1/2} \varepsilon(\frac{x-n}{m}) \tag{2}$$

Therefore, formula (1) can be expressed as the following formula in inner product form:

$$W_{\varepsilon} f(m,n) = \left\langle f(x), \varepsilon_{m,n}(x) \right\rangle \tag{3}$$

In the above representation, it can be found that the continuous wavelet transform makes the univariate function f(x) into a bivariate function, so it is the projection of f(x) on it, which can be obtained by translating and scaling the pair, which is called the wavelet function here. basis, m is the scale factor, and n is the translation factor.

3. DC Model Construction

3.1. Three-Layer B/S Mode

The three-tier design adds an intermediate tier, also known as partitioning, between the client and the database. The application software of the B/S three-tier system needs to deal with the work in the middle layer, including data analysis, business conversion, data verification, etc. [7]. In general, the client needs to go through the middle layer when it communicates with the database, that is, first establish a channel between the middle layer and COM/DOOM, and then communicate with the database through the middle layer. The core technology of B/S architecture is Web technology, coupled with the browser's powerful code analysis capabilities as support, B/S has a wider range of applications [8]. With the help of the B/S structure, the web browser can have the powerful functions of the original software, which greatly reduces the development and operation costs of the system. In the three-tier architecture, different servers are independent of different programs, that is, they show different functions on different computers. The business service interacts with the data service, handles complex data analysis business, and returns the analysis result to the user from the presentation layer [9-10]. The application mode of the three-layer B/S structure is shown in Figure 1.

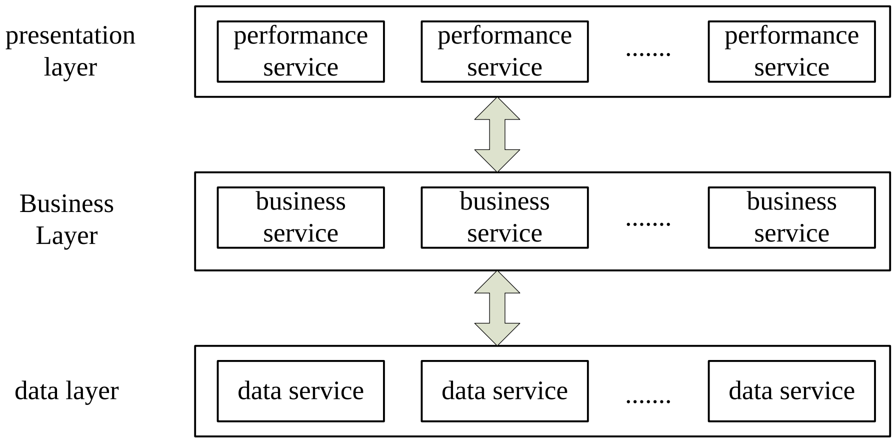

Figure 1. B/S mode

3.2. Hadoop Platform

On the Hadoop platform, users can build a DC platform according to their own needs. The system structure of Hadoop is designed to include the following advantages.

Strong scalability. The strong scalability is reflected in the process of running the Hadoop platform, the user can control the number of computing nodes, and when the number of computing nodes changes, the platform will automatically back up the existing data sets. During the backup process, the stored data format and data processing method will not be changed. In addition, its strong scalability is also reflected in the amount of data processed, and ordinary users can process PB-level data with the help of the Hadoop platform [11-12].

Efficiency. During data processing, parallel operations are performed between each node, and the processing speed is very fast.

The programming model is simple and easy to implement. When writing code for data processing tasks, the processing logic is relatively straightforward. The Map stage is responsible for data classification, and the Reduce stage is responsible for the summary of classification results [13].

In statistics, data analysis of various models in programming languages such as R and Python has become more and more popular. However, with the increase in the amount of model data, the time cost of data analysis using various programming languages is also increasing. In order to solve the problem of excessive time cost, researchers introduced the idea of DC framework into programming languages such as R and Pyhton [14].

3.3. DC Model Based on Three-Layer B/S Mode

Building a DC model on the Hadoop platform requires the use of a Storm cluster. The coordination and cooperation between Storm nodes is completed by ZooKeeper, which records all the states of each node, and if it stops, it is recorded in the local disk [15]. It is because of this feature that Storm has a high stability in practical applications and can well cope with various emergencies. DC utilizes this feature of Storm to enable DC to have a platform capable of real-time big data analysis and batch processing of data. In the case of grouped input data sources, the grouped data is stored in a file system or database, although the input data for computational streams is usually obtained from the Internet [16]. However, in practical business applications, for some real-time services, the data is stored in the database, but since the application always records the real-time performance data of the objects using the Internet, the database will be updated frequently [17]. For users, whether the task requires big data processing or real-time monitoring of computing flow, it can be easily and directly displayed and run on the platform, and users can obtain different task models when creating or modifying data.

In the commonly used network parallel computing environment, such as PVM and MPI, there are two typical structures, namely SPMD and master/slave mode. SPMD mode, also called peer-to-peer mode, corresponds to data parallel processing, in which each process processing unit performs roughly the same work. The master/slave mode, also known as the master-slave mode, can be regarded as a combination of data parallelism and function parallelism. Which parallel processing method is adopted in a specific application needs to be comprehensively considered [18-19].

Figure 2 shows the B/S-based DC model built on the Hadoop distributed system.

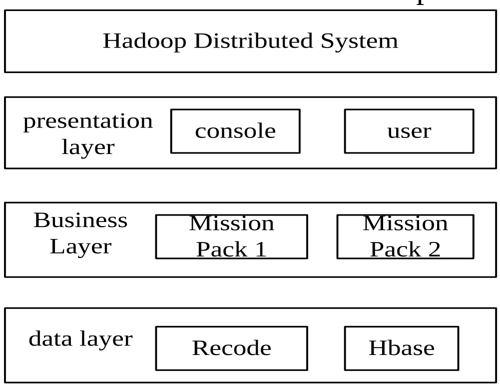

Figure 2. DC model system

4. Application of DC Model

4.1. Experiment Content

In this paper, the proposed DC mechanism is applied to the signal analysis experiment, and a control experiment is performed with the traditional situation without any other mechanism or algorithm, and the difference in performance between the two is compared, and the DC mechanism is evaluated through the control experiment effect to come. In this experiment, there are two groups of control experiments. One is the case where only the SWLDA classifier is used, in which case the parallel mechanism is used for comparison, and the other is the WAFE method added in the case of the SWLDA classifier, in this case A parallel mechanism was used for comparison with or without points.

4.2. Experimental Results

In this experiment, the signal transmission sequence is divided into 4 data sets, and the traditional computing experiments with WAFE and the parallel DC experiments using SWLDA as a classifier are completed for the 4 data sets. All completed the character spelling task. Record the data and results of each dataset experiment, and compare the performance of the two in terms of SR (selection rate) value, spelling accuracy rate ACC, and information transmission rate ITR, as shown in Table 1.

Table 1. Performance indicators of whether parallel DC is used under the dataset SWLDA classifier

	SR(selections/min)		ACC(%)		ITR	
	Tradition	parallel	Tradition	parallel	Tradition	parallel
1	1.47	2.65	96.37	91.53	5.94	12.43
2	1.47	3.17	92.54	86.49	5.65	11.87
3	1.47	3.22	97.82	83.64	6.08	9.86
4	1.47	2.83	99.16	88.77	7.34	10.58
Average	1.47	2.97	96.47	87.61	6.24	11.19

As can be seen from Table 1, since the traditional method is a fixed number of sequence flashes, the SR of the traditional P300 is all 1.47. The parallel P300 DC uses a threshold algorithm, which can dynamically adjust the number of sequences, so its SR is not fixed. From the performance of SR, it can be seen that in the four datasets, the SR value of each subject's parallel DC is better than the traditional computing model, and the final average is 2.97 selections/min, that is, the average per The output of 2.97 characters can be completed in minutes, which is 2.02 times the SR value of 1.47 of the traditional computing model.

On the other hand, in terms of spelling accuracy, due to the obvious improvement in spelling speed of the parallel P300, there will inevitably be a slight decrease in accuracy, and the accuracy of the four data sets has decreased to varying degrees compared with the traditional P300.. The final average accuracy has dropped to 87.61% compared to the traditional 96.47%. Although there has been a certain drop, it is generally acceptable.

Finally, on the ITR that combines the two factors of speed and spelling accuracy, the ITR of the four datasets is improved. Among them, the improvement effect of data set 1 is the most obvious, from 5.94bits/min to 12.43bits/min, an increase of nearly 2.09 times. The improvement of data set 4 is not very obvious, and it also reached 10.58bits/min from the previous 7.34bits/min. From the overall average effect, the ITR value has increased by nearly 79.33% from 6.24bits/min to 11.19bits/min. Figure 3 shows the performance histograms of ITR under different methods for 4 datasets.

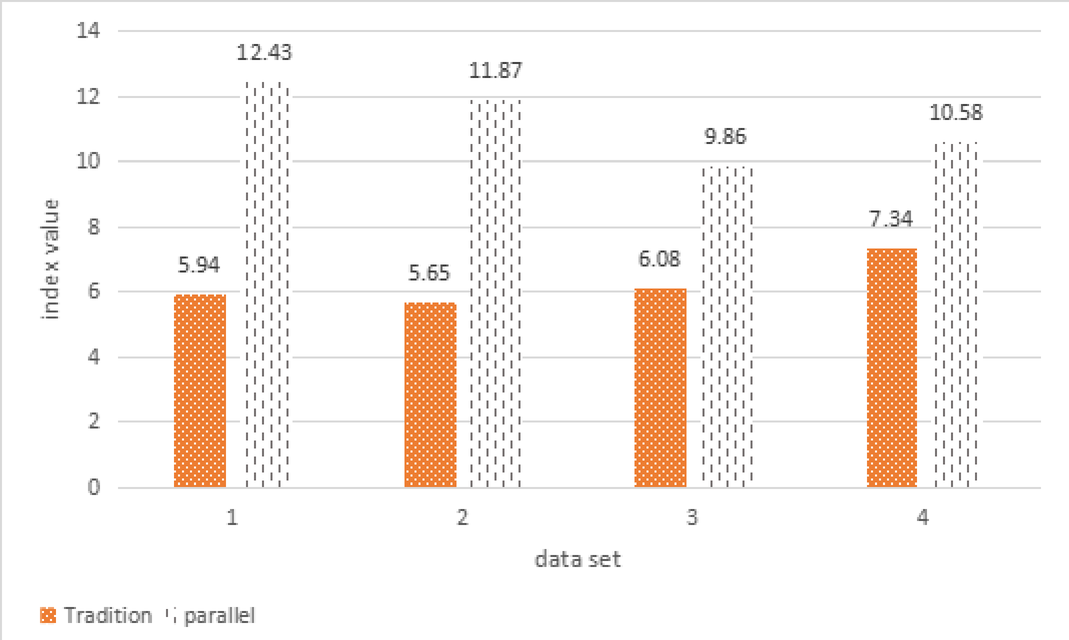

Figure 3. ITR performance

It can be seen intuitively from Figure 3 that the four datasets have improved ITR after using parallel DC.

Now compare the same 4 datasets in the case of using parallel DC to verify the impact of the use of WAFE on performance, the results are shown in Table 2.

Table 2. Performance indicators after adding wavelet extraction under parallel DC

	SR(selections/min)		ACC(%)		ITR	
	no wavelet	with wavelets	no wavelet	with wavelets	no wavelet	with wavelets
1	2.65	3.53	91.53	92.45	12.43	13.31
2	3.17	3.84	86.49	87.69	11.87	12.56
3	3.22	4.27	83.64	85.33	9.86	11.45
4	2.83	4.05	88.77	88.92	10.58	12.21
Average	2.97	3.92	87.61	88.60	11.19	12.38

It can be seen from Table 2 that the WAFE method has brought about an improvement in speed. It has increased the SR value from only 2.97selections/min to 3.92selections/min, and the accuracy has also increased slightly. 87.61% rose to 88.60%. On the whole, when the WAFE is used, the ITR value is increased from the original 11.19bits/min to 12.38bits/min. We can consider that the speed and accuracy are combined. The WAFE The method plays a role in improving the overall performance of the DC model.

5. Conclusion

In this paper, a DC model based on B/S mode is constructed. In order to analyze the performance of the model, a parallel DC mechanism experiment is carried out, and the performance of the model after parallel computing and wavelet feature extraction is verified by comparing methods. Through control experiments, we verified the performance impact of the DC model in the case of using the SWLDA classifier whether to use parallel DC and whether to perform WAFE under DC. The results show that parallel DC can improve the ITR value of the DC model in this paper. Therefore, the parallel computing method proposed in this paper can indeed effectively improve the performance of the model. In addition, whether the results of feature extraction are obtained by using wavelet analysis, this method also plays a role in improving the performance of DC models.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Ghosh S, Rawat K. Hybrid Analog/Digital Continuous Class B/J Mode for Broadband Doherty Power Amplifiers. IEEE Access, 2019, PP(99):1-1.

[2] Salloum S, Huang J Z, He Y. Random Sample Partition: A Distributed Data Model for Big Data Analysis. IEEE Transactions on Industrial Informatics, 2019, 15(11):5846-5854. https://doi.org/10.1109/TII.2019.2912723

[3] Hafaiedh I B. A generic formal model for the comparison and analysis of distributed job-scheduling algorithms in grid environment. Journal of Parallel and DC, 2019, 132(OCT.):331-343. https://doi.org/10.1016/j.jpdc.2019.05.002

[4] Prabhu J, Malar M. Trust based authentication scheme (tbas) for cloud computing environment with Kerberos protocol using distributed controller and prevention attack. International Journal of Pervasive Computing and Communications, 2021, 17(1):78-88. https://doi.org/10.1108/IJPCC-03-2020-0009

[5] Dushkin R, Mokhov A. DC model for the organization of a software environment that provides management of intelligent building automation systems. Computer Research and Modeling, 2021, 13(3):557-570.

[6] Ramu N, Pandi V, Lazarus J D, et al. A Novel Trust Model for Secure Group Communication in DC. Journal of Organizational and End User Computing, 2020, 32(3):1-14. https://doi.org/10.4018/JOEUC.2020070101

[7] Asmaa A, Abbassia D E, Hassan B, et al. Model-Based Application Deployment on Cloud Computing. International Journal of Distributed Systems and Technologies, 2019,

10(2):110-127.
[8] Nabi S W, Vanderbauhede W. FPGA design space exploration for scientific HPC applications using a fast and accurate cost model based on roofline analysis. Journal of Parallel and DC, 2019, 133(Nov.):407-419. https://doi.org/10.1016/j.jpdc.2017.05.014
[9] Battula S K, Oareilly M M, Garg S, et al. A Generic Stochastic Model for Resource Availability in Fog Computing Environments. IEEE Transactions on Parallel and Distributed Systems, 2021, 32(4):960-974. https://doi.org/10.1109/TPDS.2020.3037247
[10] Hosseinpour F, Naebi A, Virtanen S, et al. A Resource Management Model for Distributed Multi-Task Applications in Fog Computing Networks. IEEE Access, 2021, PP(99):1-1.
[11] Kaur R, Ali A. A Novel Blockchain Model for Securing IoT Based Data Transmission. International Journal of Grid and DC, 2021, 14(1):1045-1055.
[12] Bhardwaj K, Lin C, Sartor A, et al. Memory- and Communication-Aware Model Compression for Distributed Deep Learning Inference on IoT. ACM Transactions on Embedded Computing Systems, 2019, 18(5s):1-22. https://doi.org/10.1145/3358205
[13] Ehikioya S A, Olukunle A A. A Formal Model of Distributed Security for Electronic Commerce Transactions Systems. International Journal of Networked and DC, 2019, 7(2):68-84. https://doi.org/10.2991/ijndc.k.190326.003
[14] Sudharsan B, Patel P, Breslin J, et al. Toward Distributed, Global, Deep Learning Using IoT Devices. IEEE Internet Computing, 2021, 25(3):6-12. https://doi.org/10.1109/MIC.2021.3053711
[15] Bertrand N. Model checking randomized distributed algorithms. ACM SIGLOG News, 2020, 7(1):35-45. https://doi.org/10.1145/3385634.3385638
[16] Mishra A R, Panchal V K, Kumar P. Similarity Search based on Text Embedding Model for detection of Near Duplicates. International Journal of Grid and DC, 2020, 13(2):1871-1881.
[17] Jarraya A, Bouzeghoub A, Borgi A, et al. DCR: A new distributed model for human activity recognition in smart homes. Expert Systems with Application, 2020, 140(Feb.):112849.1-112849.19. https://doi.org/10.1016/j.eswa.2019.112849
[18] Almeida R B, Junes V C, Machado R, et al. A Distributed Event-Driven Architectural Model based on Situational Awareness applied on Internet of Things. Information and Software Technology, 2019, 111(JUL.):144-158. https://doi.org/10.1016/j.infsof.2019.04.001
[19] George A K, Yeboah D, Ye L. New Paradigm of Computing (DC). International Journal of Science and Research (IJSR), 2020, Volume 9(Issue 9,):1272-1276.