International Journal of Intelligent Information and Management Science
ISSN: 2307-0692, Volume 7, Issue 2, April, 2018

Research on Data Balancing Method under High Load of Distributed Parallel Database

Jia Chen, Yingxue Cai, Song Hu, Sibo Huang, Hui Hu, Zhaoquan Cai*
Huizhou University, Huizhou, China

Abstract: In order to improve big data's access ability and reduce output error code in distributed parallel database, it is necessary to carry out equalization control. A load balancing control model of big data in distributed parallel database based on adaptive random link configuration is proposed. The coherent association detection method is used to reorganize the big data information flow in the distributed database. The output link model of the distributed database is established, and the channel impulse response of the distributed database is calculated. The optimal scheduling of the distributed database output channel is carried out by using the link random allocation method and the big data load balancing configuration is carried out by combining the data clustering method to realize the data balance processing under the high load of the distributed parallel database. The simulation results show that the proposed method can improve the load balance of distributed parallel database, and the recall of database access is better.

Keywords: distributed parallel database; load balancing; control; database access

1. Introduction

With the development of information database technology, a large amount of data is transmitted and stored through the computer database. In the database storage channel, because of the interference of information and redundant data, the unbalanced class data time series is formed in the database, the database is the core configuration of the data information management, through the database realizes the data storage, the schedule, the access and the information processing[1]. In the distributed parallel database environment, big data distribution has the characteristics of self-organization and randomness. Data access and scheduling between databases need to realize information transmission and transformation through load balancing method. It has great significance to study the distributed load balancing method in the distributed parallel database environment, so as to improve the access and information scheduling ability of the database[2].

Traditionally, the distributed load balancing method in the distributed parallel database environment mainly includes the data migration scheduling method based on matching detection. The distributed load balancing method based on particle swarm optimization in distributed parallel database environment and the distributed load balancing method based on information subspace feature extraction in distributed parallel database environment[3], in reference [4], a load-balancing method based on empirical mode decomposition was proposed. By constructing and reorganizing big data's characteristic information flow, the empirical mode decomposition of big data information flow was carried out by using statistical information processing method. The algorithm can improve the ability of data scheduling and migration, but the algorithm has a large computational overhead, which is not good for the real-time performance of database access and scheduling. In reference [5], a distributed load balancing method based on adaptive particle swarm filter in distributed parallel database environment is proposed. Particle swarm optimization method is used to reconstruct the information subspace and optimize the scheduling of data, which has good migration accuracy. But the performance of the algorithm is not good. To solve the above problems, this paper presents a load balancing control model for big data in distributed parallel database based on adaptive random link configuration. The optimal scheduling of the distributed database output channel is carried out by using the link random allocation method and the big data load balancing configuration is carried out by combining the data clustering method to realize the data balance processing under the high load of the distributed parallel database.

2. Database Structure Model and Data Detection

2.1. Distributed Database Access Channel Model

In order to realize the design of big data load balancing model in distributed parallel database, a coherent associa-

tion detection method is used to reorganize the big data information flow in distributed database, and the output link model of distributed database is established[6]. A tap delay model of distributed database channel in cloud computing environment is designed. The distributed parallel database distributed structure model under cloud computing is shown in figure 1.

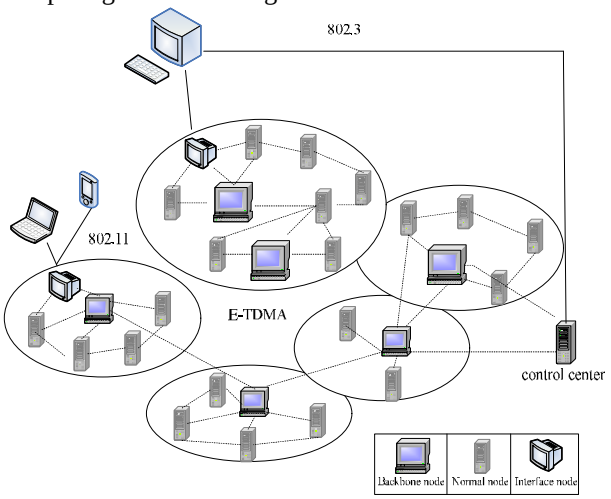

Figure 1. Distributed Parallel Database Distributed Architecture Model Under Cloud Computing.

The measure distance of calculating the routing nodes in the link layer of distributed parallel database is calculated as follows:

$$\|\boldsymbol{r}\| = \sqrt{\sum_{i=0}^{m} r_i^2 \left\|\boldsymbol{v}_{\sigma(i)}^*\right\|^2 + \|\boldsymbol{w}\|^2} \geq |r_m| \left\|\boldsymbol{v}_{\sigma(m)}^*\right\| \tag{1}$$

The congestion and smooth decision function of big data's cluster head node(SN) in distributed parallel database are as follows:

$$SL_{i`} = \begin{cases} L_i & if \ \ i=1 \\ New_{i'} & otherwise \end{cases} \tag{2}$$

Where, $New_{i`} = (e_{i`,1}, e_{i`2}, \ldots, e_{i`D})$, it represents the Source node set of distributed parallel database. The iterative equation of link offset correction for distributed parallel database communication channel is:

$$f_{ij}(n+1) = f_{ij}(n) + m_{MCMA} \frac{\partial J_{MCMA}(n)}{\partial f_{ij}(n)} \tag{3}$$

By detecting the coherence of different channel characteristics, the equalization control equation in the network link layer is obtained as:

$$\frac{\partial J_{MCMA}(n)}{\partial f_{ij}(n)} = \frac{1}{4}\left(\sum_{j=1}^{M} \frac{\partial J_{R,j}(n)}{\partial f_{ij}(n)} + \sum_{j=1}^{M} \frac{\partial J_{I,j}(n)}{\partial f_{ij}(n)} \right) \tag{4}$$

Under the condition of bandwidth constraint, the Doppler spectrum of the distributed parallel database access channel model is described as follows:

$$e_{R,j} = \left(\left|y_{R,j}(n)\right|^2 - R_{2,R}\right) \times y_{R,j}(n)^* \tag{5}$$

According to the above analysis, the coherent association detection method is used to reorganize the big data information flow in the distributed database, and the output link model of the distributed database is established.

2.2. Channel Impulse Characteristic Analysis of Distributed Database

On the basis of establishing the output link channel model of distributed parallel database, using coherent association detection method to reorganize big data information flow in distributed parallel database, the adaptive random link assignment is constructed. The link model accessed by big data in the distributed parallel database is expressed as follows:

$$\begin{aligned} R_w(l) &= E\left[w(k)\, w^H(k+l)\right] \\ &= \int_{-p}^{p} \left[d_l \cdot \frac{1}{\Delta\sqrt{2p}} e^{-(q-q_0)^2/2\Delta^2} \right] a(q)\, a^H(q)\, dq \end{aligned} \tag{6}$$

The q_0 and Δ parameters represent the phase deflection and power gain of the output big data of the distributed parallel database, respectively. The time-frequency joint characteristics of computing big data load in distributed parallel database are as follows:

$$\mathrm{Var}(T_a) = \left(bt_a^0\right)^2 \left\{ \begin{array}{l} \left(\mathrm{B}(l-2n,m) \Big/ \left(c_a^{\max}\right)^{2n} \mathrm{B}(l,m) \right) \left(\exp\left(2nm_v^a + \left(2ns_v^a\right)^2 / 2 \right) \right) \\ -\left(\left(\mathrm{B}(l-n,m) \Big/ \left(c_a^{\max}\right)^{n} \mathrm{B}(l,m) \right) \left(\exp\left(nm_v^a + \left(ns_v^a\right)^2 / 2 \right) \right) \right)^2 \end{array} \right. \tag{7}$$

The load gain V_a accessed by big data in a distributed parallel database is independent of the frequency doubling feature C_a .

3. Optimization of High Load Balancing Method For Database

3.1. Load Balancing Control

On the basis of using coherent association detection method to reorganize big data information flow and design channel model in distributed parallel database, the optimal design of big data load balancing model in distributed parallel database is carried out in this paper[7]. A load balancing model of big data in distributed parallel database based on adaptive random link configuration is proposed. The embedding dimension and average mutual information of big data load balancing configuration are obtained as follows:

$$q^w = \mathrm{E}\left(Q^w\right) = \sum\nolimits_{k \in R_w} f_k^w \ , w \in W \tag{8}$$

$$v_a = \mathrm{E}\left(V_a\right) = \sum\nolimits_{w \in W} \sum\nolimits_{k \in R_w} d_{ak}^w f_k^w \ , a \in A \tag{9}$$

$$f_k^w \geq 0, k \in R_w, w \in W \tag{10}$$

Combining with the channel balancing control method, the load balancing design is carried out, and the characteristic sequence of load balancing scheduling set in the distributed parallel database access channel is presented as follows:

$$r_1(n) = r_2(n)\exp(-jw_0T_p/2),\ n = 0,1,...,(N-3)/2 \quad (11)$$

$$r_2(n) = A\exp[j(w_0nT+q)],\ n = 0,1,...,(N-3)/2 \quad (12)$$

Adaptive random link configuration method is used to calculate the load quantity accessed by big data in the distributed parallel database.

$$R_1(k) = R_2(k)\exp(-jw_0T_p/2),\ k = 0,1,...,(N-3)/2 \quad (13)$$

$$R_2(k) = A_k\exp(jj_k),\quad k = 0,1,...,(N-3)/2 \quad (14)$$

3.2. Database Access Optimization

Frequency domain expansion method is used to realize frequency doubling of big data output channel. The energy of distributed parallel database big data accessing load information to Sink is as follows:

$$\begin{aligned} Computition(n_j) &= (E_{elec}+E_{DF})ld + E_{Tx(l,d_j)} \\ &= (E_{elec}+E_{DF})ld + lE_{elec} + le_{fs}d_j^{\ 2} \\ &= [(E_{elec}+E_{DF})d + E_{elec} + e_{fs}d_j^{\ 2}]l \end{aligned} \quad (15)$$

In the presence of channel attenuation, the output load $x_k^w(w)$ of distributed parallel database is balanced and scheduled as follows:

$$h_k^w(w) = \mathrm{E}\left(T_k^w \middle| T_k^w > x_k^w(w)\right),\ k \in R_w, w \in W \quad (16)$$

Where, $x_k^w(w)$ is expressed as:

$$x_k^w(w) = \min\left\{x \middle| \Pr\left(T_k^w \le x\right) \ge w\right\} = \mathrm{E}\left(T_k^w\right) + g_k^w(w),\ k \in R_w, w \in W \quad (17)$$

Under balanced scheduling control, big data outputs link information flow access bandwidth is:

$$x_k^w(w) = t_k^w + \Phi^{-1}(w)s_{k,t}^{\ w},\ k \in R_w, w \in W \quad (18)$$

The joint state estimation model of large data load frequency domain characteristics in distributed parallel database is calculated.

$$\Phi(w) = E[e^{jwX}] =$$

$$\begin{cases} \exp\{jmw - |sw|^a[1 - jb\ \mathrm{sgn}(w)\tan(\frac{pa}{2})]\}, a \ne 1 \\ \exp\{jmw - |sw|^a[1 + jb\ \mathrm{sgn}(w)\frac{2}{p}\ln|w|]\}, a = 1 \end{cases} \quad (19)$$

Based on the above analysis, the design of data balance under high load of distributed parallel database is realized.

4. Simulation Experiment and Result Analysis

In order to test the application performance of this method in the realization of big data load balancing scheduling and communication in distributed parallel database, the simulation experiment was carried out. The experiment was designed by Matlab 7 simulation tool and divided by SPSS 1.4 statistical software. Statistical analysis of big data in distributed parallel database shows that the size of big data sample set for distributed parallel database communication is 25Gbit. the load intensity of network access is 12dB, the initial sampling frequency is 250kHz, and the normalized termination frequency is respectively: $f_{21} = 0.45$, $f_{22} = 0.05$, $f_{23} = 0.3$. The simulation environment and parameters are set, and the data output is shown in figure 2.

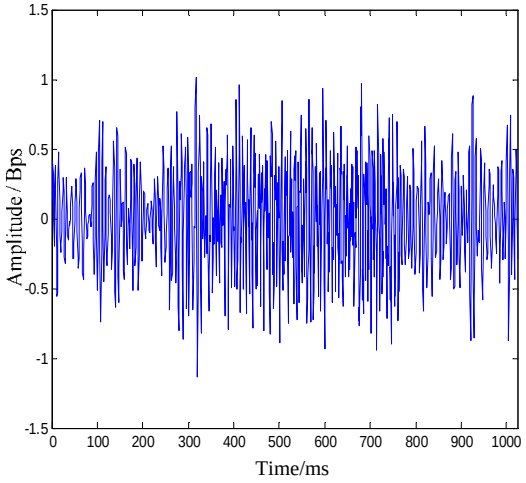

Figure 2. Data Output.

Taking the data of figure 2 as the sample, the load balancing scheduling analysis is carried out, and different methods are used to test the recall of database access. The comparison results are shown in figure 3.

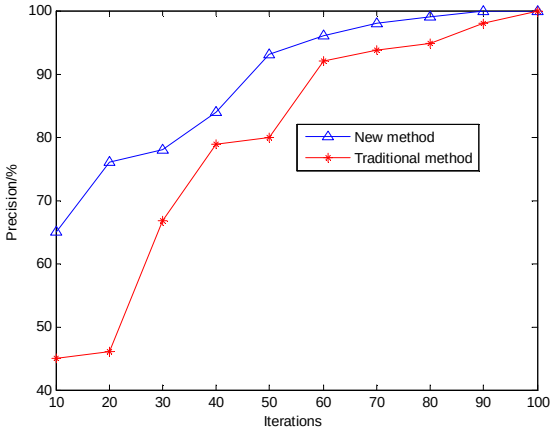

Figure 3. Recall Comparison of Database Access.

The simulation results show that the proposed method can improve the load balance of distributed parallel database, and the recall of database access is better.

5. Conclusions

In this paper, a load balancing control model of big data in distributed parallel database based on adaptive random

link configuration is proposed. The coherent association detection method is used to reorganize the big data information flow in the distributed database. The output link model of the distributed database is established, and the channel impulse response of the distributed database is calculated. The optimal scheduling of the distributed database output channel is carried out by using the link random allocation method and the big data load balancing configuration is carried out by combining the data clustering method to realize the data balance processing under the high load of the distributed parallel database. It can improve the load balance of distributed parallel database, and the recall of database access is better, it has great application value.

6. Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 61772225);the Foundation for Distinguished Young Talents in Higher Education of Guangdong (No. 2015KQNCX153);Science and Technology Program of Huizhou (No. 2015B010002002, No. 2016X0431046, No.2016X0434049, No. 2016X0432047, No.2017c0406022,No.017c0407023, No. 2017c0414030).

References

[1] CHANDRASEKHAR R, KAMARAJU M, SAIRAM M V S, et al. PAPR reduction using combination of precoding with Mu-law companding technique for MIMO-OFDM systems[C]. IEEE International Conference on Communications and Signal Processing, Melmaruvathur, India, 2015: 0479-0483.

[2] HE Y, ZHANG C S, TANG X M, et al. Coherent integration loss due to pulses loss and phase modulation in passive bistatic radar[J]. Digital Signal Processing, 2013, 23(4):1265-1276.

[3] HAO H. Multi component LFM signal detection and parameter estimation based on EEMD-FRFT[J]. Optik-International Journal for Light and Electron Optics, 2013, 124(23):6093-6096.

[4] GOVONI M A, LI H, KOSINSKI J A. Range-Doppler resolution of the linear-FM noise radar waveform[J]. IEEE Transactions on Aerospace and Electronic Systems, 2013, 49(1):658-664.

[5] XIA Mengying, SU Weimin, and GU Hong. The research on Doppler ambiguity for dual pulse repetition frequencies radar[C]. CIE International Conference on Radar, Guangzhou, 2016: 1090-1093.

[6] SI Hongwei and LI Zhaohui. An optimization design scheme for multiple PRT and multi-frequency coherent Doppler velometry[C]. 2016 MTS/IEEE OCEANS'16 Conference, Shanghai, 2016: 1-6.

[7] PAN Ying, TANG Yong, and LIU Hai. Access control in very loosely structured data model using relational databases[J]. Acta Electronica Sinica, 2012, 40(3): 600-606.