Cloud database dynamic route scheduling based on polymorphic ant colony optimization algorithm

Chen Qing*, Yong Zhong, Liuming Xiang

Chengdu Inst. of Computer Applications, Chinese Academy of Science, People's South Road, Section 4 Chengdu, China

Received 1 March 2014, www.tsi.lv

Abstract

Big Data era spawned the development of Cloud database. As a database, which need easily scale out, how to quickly find the available nodes are focuses of the study. Ant colony algorithm is based on bionic optimization algorithm and has the characters of smart searching, global optimization, robustness, distributed computing and easily combined with other algorithms, but the algorithm is prone to premature convergence, making the results often caught local optimum. According to this, polymorphic ant colony algorithm was proposed which combined with a Cloud database; the algorithm can quickly and reasonably find the nodes in Cloud environment, reducing the load of routing, thus greatly improved the Cloud database's ability of scaling out.

Keywords: Big Data, Cloud database, Ant colony algorithm, premature convergence, scaling out

1 Introduction

As the representatives of the Sensor systems just as RFID are widely used, global data are growing explosively. These data have characters of complexity, spontaneity and randomness. How to effectively manage these big data has become a big challenge. Cloud database as a data storage system, has been studying for long. Cloud database are different from traditional distributed database and static database, how to find the most suitable nodes as soon as possible in order to improve database's ability of scale-out is the focus of research.

2 Cloud database

Cloud database system consists of a number of sites gathered together. These sites also called nodes, which are joined together in the communication network, each node is an independent database system, and they have their own database, the central processor, terminals, and their respective local database management system. Therefore Cloud database system can be seen as a series of joint centralized database system [1], and this system also known as architecture of share-nothing (SN). They logically belong to one system, but on the physical structure are distributed. Big Data Cloud database as an effective management system that has the following characteristics:

1) Highly Scalable: Cloud database nodes provide dynamic expansion and contraction capabilities, users can adjust number of nodes in the entire system according to needs, and is convenient for user's resource allocation and management.
2) High Reliability: each data node at least have multiple copies, a copy of the data nodes automatically translate to the master node and provide data access services when a master data node crashed.
3) High Stability: Cloud database system can keep running as long as there has only one node.
4) High Efficiency: data dynamically balanced allocation algorithm, fully coordinated, balanced each storage server's storage pressure to ensure that the data on the storage server stress basically average; various databases' parallel scheduling algorithm, fully coordinated each node within Cloud systems to work together, to provide users with microsecond response support.
5) Low Cost: Cloud database can be used as an extremely low-cost computer system to build Cloud node, system construction costs can be controlled at a lower level. Some explanations above Cloud database are distributed, dynamic, global balanced and scalable database.

At present, the design of Cloud Databases scaling out have two main ideas. One is the database based on key/value structural, which belongs to non-relational databases, and the database system can easily sale out due to no correlation between the data, this design greatly enhanced the level of scalability, but because of lacking of transaction management, it can only be used in a particular scene, such as Social Networking Services(SNS). The other idea of design is based on traditional distributed relational database (DDBS), because this system retains the management of transaction, how to improve the ability of scale-out to adapt Big Data environment is the difficulty in designing.

In this paper, we did not focus on the transaction level, but focus on the node-finding algorithm when scaling out which all the design of CloudDB would face to in Cloud environment.

*Corresponding author - E-mail: qing33dd@aliyun.com

3 Basic ant colony algorithm models

Ant colony algorithm is typically used for solving complex combinatorial optimization problems. In solving the problems of different properties, the ant colony algorithm model definitions are different. We take TSP [2], which has m nodes as an example to illustrate the basic ant colony algorithm model. m nodes TSP problem is to find through m nodes each time and finally back to the starting point of the shortest path.

Let n be the number of ants in ant colony, d_{ij} (i, j = 1,2,...,m) is the distance between node i and node j, $\tau_{ij}(0)$ is the concentration of pheromone of node i and node j at time t. At initial time, concentration of pheromone is the same on each path, set $\tau_{ij}(0) = C$ (C is constant). Ant k(k=1,2,...,n) In the process of movement, concentration of the pheromone in each path decides the direction, $p_{ij}^{k}(t)$ is the probability of Ant k transfer from the node i to node j at time t, which is calculated as (1):

$$p_{ij}^{k} = \begin{cases} \dfrac{\tau_{ij}^{\alpha}(t)\eta_{ij}^{\beta}(t)}{\sum\limits_{s \notin tabu_k} \tau_{ij}^{\alpha}(t)\eta_{is}^{\beta}(t)'} & j \notin tabu_k \\ 0, & j \in tabu_k \end{cases}. \quad (1)$$

$tabu_k$ (k=1, 2, ..., n) is the set of nodes which ant k has passed. In the beginning, $tabu_k$ has only one element, which is beginning node, along with the evolution, elements in $tabu_k$ are increasing. As time goes on, the pheromone in the path gradually disappeared. Parameter (1-Q) represents the degree of pheromone's volatilization, all the ants to complete a circulation; each concentration of pheromone is adjusted according to the formula (2),

$$\left.\begin{array}{l} \tau_{ij}(t+1) = \rho * \tau_{ij}(t) + (1-\rho) * \Delta\tau_{ij} \quad \rho \in (0,1) \\ \Delta\tau_{ij} = \sum\limits_{k=1}^{n} \Delta\tau_{ij}^{k} \end{array}\right\}. \quad (2)$$

$\Delta\tau_{ij}^{k}$ is the concentration of pheromone in path ij left by ant k in circulation, and $\Delta\tau_{ij}^{k}$ is the sum of concentration of pheromone in path ij left by ant k in circulation.

Dorigo has given three different models, referred to ant cycle system, ant quantity system, ant density system; the difference lies in their different calculation expression of $\Delta\tau_{ij}^{k}$ are different.

In the model of ant cycle system,

$$\Delta\tau_{ij}^{k} = \begin{cases} Q/L_K, & \text{if ant k passed node i,j from time t to t+1} \\ 0, & \text{el se} \end{cases}. \quad (3)$$

L_k is the distance walked by ant k in this circulation.

In ant quantity system and ant density system, $\Delta\tau_{ij}^{k}$ means:

$$\Delta\tau_{ij}^{k} = \begin{cases} Q/d_{ij}, & \text{if ant k passed node i,j from time t to t+1} \\ 0, & \text{el se} \end{cases}, \quad (4)$$

$$\Delta\tau_{ij}^{k} = \begin{cases} Q, & \text{if ant k passed node i,j from time t to t+1} \\ 0, & \text{el se} \end{cases}, \quad (5)$$

In the three models above, the first model uses global information, the other two use local information, so the model of ant quantity system is usually used as basic model. Parameters can use experimental methods to determine the optimal combination and also can be obtained by evolutionary learning. The calculation ceased when the evolutionary trend is not obvious.

4 Ant colony algorithms

P The literature [3] shows that real ant colony society is organized, and the work is divided. The ants in ant colony are divided into labour ants, scout ants, soldier and worker ants, etc. They carry out their duties, and also collaborate, forming as an organic one, and each ant contact with each other with hormones secreted, the class of hormones are divided into warning class, attracting class, convening class, marking region class etc. While a single ant is small, but by forming a group, it reflects the high degree of organization and society, so that the entire ant colony has extraordinary abilities. Based on this, Xu Jingmin [4] et proposed the based on polymorphic ant colony, where the "polymorphic" refers to the ant society has a variety of the state and pheromone.

4.1 COMPARING WITH OTHER SEARCHING ALGORITHMS

Comparing with evolutionary computation [5-7], ant colony algorithm is similar within two aspects; First, two algorithms used group to express solution of the problem; second, the new groups are created by the knowledge related to old groups. And main difference between two algorithms are that in evolutionary computation knowledge of all issues are included in the current groups, while in ant colony algorithm, all the knowledge in trail of pheromones.

And comparing with the simulated annealing algorithm (SA) [8], the search strategy is almost the same in essential. In SA, the process of calculating the energy Ei at state i in a solid is same with ants in a "travel", they are both sampling the solution of space; "annealing" and "secrete pheromones" are both using accumulated information to enhance the subspace search; and "Metropolis Standard" and "random state transition rules" are both trying to use the algorithm to escape from local optima, and accept the worse within a certain range, and search for a new sub-space.

So From the perspective of ant colony algorithm, we can see that: the scale of ants actually affect the frequency of updates, so if the scale is large, the gap between different state become widen when updating the pheromone; and when the scale is small, only a small amount of pheromone will be updated in order to avoid premature .and moreover, because both algorithms are essentially the same, some improvements and variations in SA can be used directly on the ant he ant colony algorithm to improve its performance.

Ant colony also can be considered as connection system, and the most representative example of connect system is neural networks (Neural Network, referred NN) [9, 10]. From a structural point of view, ant colony algorithm and neural networks usually have a similar parallel mechanism; every state ant visited corresponds to the neural networks of neurons. Ant itself through the neural network can be seen as concurrent input signal. Ant colony algorithm learning rule can be interpreted as an acquired rule, that the signal are more strength in good quality solutions than in bad solutions.

4.2 SHORTAGE OF EXISTING ANT COLONY ALGORITHM

Although the ant colony algorithm has many advantages, but there are also some drawbacks of this algorithm. Compared with other methods, the algorithm generally require a longer search time, and complexity of the algorithm can reflect that , and also the algorithm is prone to stagnation , just like that when searching at a certain time, all the solutions become consistent, and cannot search further in solution space for a better solution.

In the ant colony algorithm, the ants are always dependent on other ants feedback to reinforce learning , rather than to consider their own experience, that herd behaviour, easily lead to precocious, stagnation , so that the algorithm converges slowly . Based on this, scholars have proposed a new ant colony algorithm to improve the algorithm. Dorigo M proposed an ant colony algorithm called Ant-Q System; German scholar Stutzle T and Hoos H proposed an algorithm called "max-min ant system" (MAX-MIN ant system, MMAS). Wu Qinghong [5] inspired by mutation operator of the genetic algorithm, and used a reversal mutation mechanism, and proposed an ant colony algorithm having mutation characteristics, and which is the earliest improvements in ant colony algorithm made by domestic scholars. Since then there have been many scholars making improvement in ant colony algorithm , such as an ant colony algorithm with characteristics of sensation and perception [11] , self-adaptive ant colony algorithm [12], an ant colony algorithm based on pheromone diffusion [13], an ant colony algorithm based on hybrid behaviour [14], and an ant colony algorithm based on pattern-learning [15, 16].

4.3 POLYMORPHIC ANT COLONY ALGORITHM

In polymorphic ant colony algorithm, the ant colony society is divided into three categories: scouting ants, working ants and searching ants. Scouting ant's task is to use the node as starting point for local reconnaissance, and mark the results by hormones in order to leave the message for coming ants; searching ant's task is doing global searching, choosing the next node by hormones and searching until finding the best (shortest) route, and mark the route for working ant; working ant task is to feed back the food to the nest on the marked route which is shortest one. In algorithm designing, because working ant not responsible for routing, so we only need to control scouting ant and searching ants' pheromone.

Algorithm of scouting ants: put m ants in m nodes, and using node as centre to scout other (m-1) nodes, and combined results with MAXPC used as pheromone, marked s[i][j], which means the route from node I to node j. Calculated as follows:

$$s[i][j]=\begin{cases} d/d_{ij}, & \text{if node j in the scope of nodei's MAXPC} \\ 0, & \text{else} \end{cases} . \quad (6)$$

d_{ij} is the shortest distance from node i to other m-1 nodes. And according to this, the amount of information on each path the initial time set as follows:

$$\tau_{ij}(0)=\begin{cases} c*s[i][j], & \text{if } s[i][j]\neq 0 \\ c*d/d_{ij}, & \text{else} \end{cases} , \quad (7)$$

d_{ij} is the longest distance from node i to other m-1 nodes, and C is concentrations of pheromone on the roads at initial time, and pheromone is support for calculating of transition probability and provide support for adjusting the concentration of pheromone on each path.

Algorithm of searching ants: probability of ant k (k= 1, 2 … n) moving from node i to node j at t time, $p_{ij}^{k}(t)$ is:

$$p_{ij}^{k}=\begin{cases} \dfrac{\tau_{ij}^{\alpha}(t)\eta_{ij}^{\beta}(t)}{\sum\limits_{s\notin tabu_k}\tau_{ij}^{\alpha}(t)\eta_{is}^{\beta}(t)}, & j\notin tabu_k,\ and\ s[i][j]\neq 0 \\ 0, & \text{else} \end{cases} . \quad (8)$$

According to this formula, the searching ant reduces the scale of the search due to the pheromone.

If all the ants complete a circulation, each concentration of pheromone will be adjusted as follows:

$$\tau_{ij}(t+1)=\begin{cases} \rho*\tau_{ij}(t)+(1-\rho)*\Delta\tau_{ij}, & \text{if } s[i][j]\neq 0 \\ \rho*\tau_{ij}(t), & \text{else} \end{cases} . \quad (9)$$

$\Delta\tau_{ij}$ is the sum of the pheromone left by ant moving on the path ij in a circulation. $\Delta\tau_{ij}^{k}$ is the pheromone left by ant k moving on the path ij from time t to time t+1 in a circulation, and $\Delta\tau = \sum_{k=1}^{K} \tau_{ij}^{k}$ is as follows:

$$\Delta\tau_{ij}^{k} = \begin{cases} Q*(d/d_{ij})/L_k, & \text{if ant k passed ij, and } s[i][j]\neq 0 \\ 0, & \text{else} \end{cases} . \quad (10)$$

The processes of Polymorphic ant colony algorithm can be described as follows:

1) Initialize Q, C and maximum evolution algebra;
2) Put m scouting ants in m nodes, each ant scouts other m-1 nodes and calculates the pheromone using formula(6), and put the results in S[i][j]; Two different styles of referencing are in common use: the Harvard alphabetical system and the Vancouver numerical system. Use the Vancouver numerical system as described below.
3) Using formula (7) to initial information in each path;
4) Set NC = 0;
5) Randomly selecting each scouting ant's initial position and put the position in table tabu;
6) Using formula (8) to calculate each scouting ant's position to be transferred;
7) Calculating the value of the objective function ants search L k (k = 1, 2,… n), and record the best solution at current;
8) If the algebras or the solution have not got any answer, go to step 10); otherwise modify the concentration of pheromone according to formula (9);
9) Set = 0, and set table tabu Null, NC←NC+1, go to step 5) ;
10) Output the optimal answer.

5 Experiment

5.1 EXPERIMENTAL MODEL

We choose NS-2 which developed by South California University as platform [17], and use the throughput and delay of transmission time as indicators, and select the algorithm OSPF(Open Shortest Path First) [18], SPF (Shortest Path First) and BF for comparison.

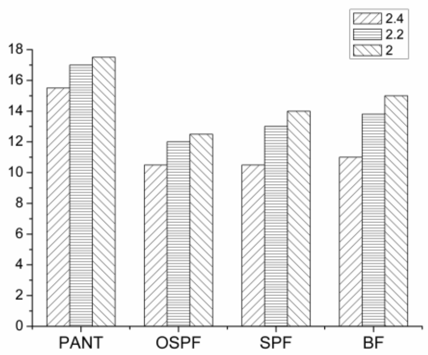


FIGURE 1 Average throughput

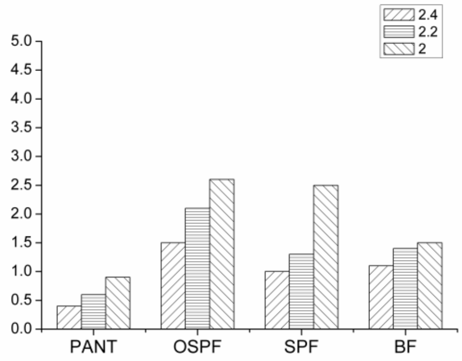


FIGURE 2 The Performance of packet delay

From FIGURE 1 and FIGURE 2, we can find that the polymorphic ant colony algorithm performances better than the other three in throughput and delay of transmission time. And through the experiment, we can also find that in the algorithm based on polymorphic ant colony ,if the inspired factor α is too small, convergence will be slow and, easily fall into local optimum; if inspired factor α is too large, pheromone' weight will be heavy in scouting, and causes premature convergence. If expected factor β is too small, the ant colony will lead into purely random searching, and difficult to find the optimal solution; if β is too large, the speed of convergence be faster, but convergence tends to be bad. If pheromone evaporation factor ρ is too large, previously searched path had possibility of researching, and it will affect the algorithm's randomness and global searching capability; if pheromone enhancement factor Q is larger, accumulation of pheromone will be faster, and it can enhance capability of the positive feedback in searching, and fasten the convergence; when Q is too large, the algorithm of the global search capability will deteriorate, easy to fall into local optimal solution, and cause the loops.

6 Conclusions

In this paper, we combined polymorphic ant colony algorithm with Cloud database. The algorithm improved previous basic ant colony algorithm, and meanwhile solving the problem of node dynamic plugging when the Cloud database needs to scale out. In experiment, the throughput and delay of transmission time were used as indicators to evaluate algorithm's capability; and using NS-2 which developed by South California University as platform. The results were compared with other similar algorithms, and shows Cloud Database can quickly find the node in Big Data's environment using polymorphic ant colony algorithm.

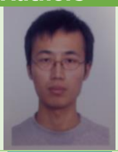
Acknowledgments

This research is supported by Science and Technology Department of Sichuan Province.

References

[1] http: //soft.ccw.com.cn/ it / Oct 2013
[2] Du D P 1990 *Logistics Graph Theory* Buaa. Press: Beijing 240-54
[3] Wu J, Wang CH L 1995 *China Ant* Chinese Forestry Press: Beijing 1-12
[4] Xu J M et al 2005 *Journal of University of Science and Technology of China* **35**(1)
[5] Wang Z Z, Bo T 2000 *Computional Evolution* National University of Defence Technology Press: Changsha
[6] Li Q M, Kong J S, Lin D, Li SH Q 1995 *The Basic Theory and Applications of Genetic Algorithms* Science Press: Beijing
[7] Liu Y, Kang L S, Chen M P 1995 *Parallel Numerical Algorithms: Genetic Algorithms* Science Press: Beijing
[8] Kang L S, Xie Y, Liu SH Y, Luo ZH H 1995 *Non-numerical Parallel Algorithms: Simulated Annealing Algorithm* Science Press: Beijing
[9] Zhang N Y, Yan P F 2000 *Neural Networks and Fuzzy Control* Tsinghua University Press: Beijing
[10] Jin F 2000 *Fundamentals of Neural Computational Intelligence Methods* Southwest Jiaotong University Press: Chengdu
[11] Chen L, Qin L, Chen H J, Xu X H 2004 *Journal of System Simulation* **33**(2) 983-9
[12] Wang Y, Xie J Y 2002 *Journal of System Simulation* **14**(1)
[13] Huang G R, Cao X B, Wang X F 2004 *ACTA Electronica SINICA* **32**(5)
[14] Hu X B, Huang X Y 2005 *Control and Decision* **20**(1) 69-71
[15] Xiao Y S, Li B Y, Wu Q D 2004 *Control and Decision* **19**(8) 885-88
[16] Mi J S, Leung Y, Wu W Z 2005 *Int J Gen Syst* **34**(1) 77-90
[17] www.isi.edu/nsnam/ns/ Oct 2013
[18] Moy J 1991 OSPF version 2. RFC 1247

Authors



Chen Qing, ChengDu, China

Current position, grades: ChengDu, PhD studies
University studies: Chinese academy of Science
Scientific interest: Big data, Cloud Database
Publications: A Seamless Software Development Approach using DCI

Liuming Xiang, born in November, 1989, ChangSha, China

Current position, grades: ChengDu, Master studies
University studies: Chinese academy of Science
Scientific interest: Big data, Cloud Database