Research on intelligent service scheduling model based on service level optimization

X. CHENG
Department of Management, Dalian Polytechnic University, Dalian, Liaoning Province, China

W.H. LI
Department of Computer Science and Technology, Jilin University, Changchun, Jilin Province, China

ABSTRACT: Parallel services efficiently mining frequent itemsets is the core issue of dynamic service scheduling. For the feature that service nodes split the task into reorganization service under the data transmission environment, how to arrange the services sequence of matched nodes in accordance with specific calculation sequence to ensure the best computing and communications performance and to avoid explosion in the service delivery process is a puzzle. Through in-depth study of Apriori algorithm, a mining algorithm based on Partial Depth Priority (PDP) was proposed. The algorithm has a variety of features like various service nodes distributed in the AOV network to meet the characteristics of the DAG chart, short services communication time, dynamic allocation of service priorities. Experiments prove that: the model by studying the Apriori algorithm proposed "preferential attachment-merge pruning algorithm for mining" partial depth based on MapReduce, using the "density grid attunements connected region algorithm" deep into the MapReduce technology, so as to better complete the service node partition and data merge.
KEYWORD: service scheduling; mesh refinement; node searching; transmission path

1 INTRODUCTION

The originally cluster computing system, like as MapReduce and Dryad, did resolve batch jobs for retrieve. With the expansion of the application of such a cluster system, cluster shared between multiple users, service operations with batch mixing length of a data set and short communications for interactive services increasingly become mainstream.

To improve the success rate of service and real-time scheduling information on the process of mining services become critical issues. The most classic association rule mining algorithms are Apriori algorithm, the main idea of the algorithm is to use step by step iterative method to obtain high-dimensional frequent item sets through low-dimensional frequent item sets. However, in practical application Apriori algorithm, where there is not satisfactory.

2 CLUSTER SYSTEMS COMMONLY USED SCHEDULING ALGORITHM

2.1 *FIFO with priority scheduling policy*

Hadoop default scheduling policy is FIFO with priority [1]. All user services are only submitted to a queue. Low priority in order to scan and submit chronological entire service queue, select a service execution requirements are met. FIFO achieve simpler, less scheduling overhead of the entire cluster [2-3]. FIFO scheduling algorithm biggest drawback is that in the presence of large and small-job response time is poor [4]. Productive work as long occupied cluster resources will cause another user's batch job unbearable so long response time, but also to other users becomes prolonged lack of interactive query processing.

2.2 *HOD Scheduling Policy*

HOD is able to use Torque on a shared physical cluster provides proprietary Hadoop clusters and Hadoop distributed file system instance system [5-6]. HOD small job in improving the response time has been great progress compared to FIFO. This leads to a part of the input data is not on the Map task local private cluster node, you must read the data through the network, thereby reducing the response time of the system throughput and services [7-8].

HOD scheduling algorithms in order to solve problems arising, Facebook made a fair scheduling algorithm [9]. Fair scheduling algorithm as possible to ensure that each user receives an equal share of resources [10]. In addition to providing a fair sharing method, the fair share scheduler also provides a minimal amount [11]. Each pool set a minimum slot number, when making scheduling scheduler will ensure that each pool when there is a need to get its minimum share amount.

3 APRIORI ALGORITHM ANALYSIS

Apriori algorithm uses prior knowledge of frequent item sets in nature, first identify the set 1 frequent item sets, the collection denoted by L1. L1 is used to identify a collection of frequent item sets L2, while L2 is used to find L3, and so on until you can not find the frequent set k_ items. Important properties in order to improve the efficiency of frequent itemsets produced layer by layer , the one called Apriori property applied to the algorithm[12], so look for frequent item sets can be divided into two processes:

1. Connect: To find Lk, k set of candidate itemsets generated by Lk-1 connection with yourself. The candidate set is denoted as Ck. Let L1 and L2 are Lk-1 in the item set. Sign L1 [j] represents Li item j. Perform connection Lk -1> <Lk-1, wherein Lk-1 is connected to the element, if the term of k-2 before they are identical entries, L1 and L2 generated by the connection result item set is L1 [1] L2 [2], ..., L1 [k-1] L2 [k-1].
2. pruning step: Ck is a superset of Lk . According to the Apriori property-- any non- frequent (k-1) - item set can not be a subset of the frequent k- item set. Thus, if a candidate k- itemsets (k-1) is not a subset of Lk-1, then the candidate can not be frequent, which can be removed in the Ck. Then scanning services library to determine the degree of support for each candidate Ck set to determine Lk.

Although Apriori algorithm for mining global mission itself has been optimized, but there are still insufficient efficiency of the algorithm, mainly in the following three aspects [13]:

1. The use of connected k itemsets generated k +1 candidate set to determine the connection conditions are too many times;
2. To determine any one of c, c$\in$ Ck of k (k-1) is a subset of all the Lk-1;
3. In order to get all Ck (k = 1, 2,…, n) candidate frequent itemsets, library services need to scan n times.

4 BASED ON MAPREDUCE LOCAL DEPTH FIRST CONNECTION - MERGE PRUNING ALGORITHM FOR MINING

During the update process, the entire D-dimensional solution space into m parts, namely, the position of vector into sub-vector, where each sub-vector corresponding to the position of the speed sub-vector interactions are represented by Zi and Vj, j = 1, ..., m.

The update process is based on the sub Map stage vector, depth-first order (from Z1 to Zm) cyclic update. Updating the first j ($1 \le j \le m$) during the sub-vector, and in the following location update interactive speed N times of iterative execution of the formula:

$$v_j^k = (a / k^p) \times r + b \times L_j^{k-1} \tag{1}$$

$$z_j^k = \begin{cases} z_j^{k-1} + v_j^k & f(x_1^k) > f(x_2^k) \\ z_j^{k-1} & f(x_1^k) \le f(x_2^k) \end{cases} \tag{2}$$

$$L_j^k = \begin{cases} L_j^{k-1} / s & f(x_1^k) \le f(x_2^k) \\ v_j^k & f(x_1^k) > f(x_2^k) \end{cases} \tag{3}$$

Among those:

$x_1^k = [z_1, z_2, \cdots, z_m]; x_2^{k-1} = [z_1, z_2, \cdots, z_m]; k = 1, 2, \cdots N.$

Based on the joint area, such as the density of the grid algorithm design attunement.

(1)The first is an irregular rectangle and add graphics to make the label shown in Figure 1:

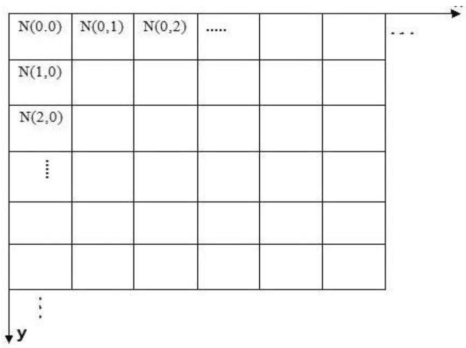

Figure 1 External matrix labeled graph

In case, each mesh, the different values of scattered randomly distributed points, start and end point is partition, the statistical value of each mesh in the scattered point, denoted as P (i, j).

(2)The grid attunement center take alternative mesh grid shown in Figure 2:

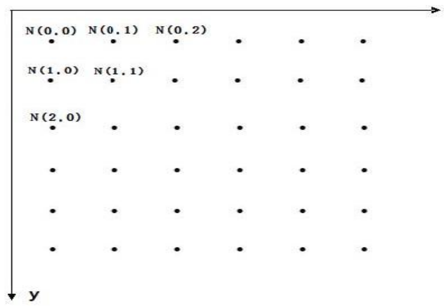

Figure 2 Schematic take to find center of grid

In this case, the introduction of the concept of β-point connectivity, is the maximum number of each point can be expanded direction outward at a certain moment.

(3)Point search partition. From the origin of the N (0,0) density started doing search, if P (0,0) <Rmin

continued to search, search destination point with the N (0,0) point adjacent transverse and take wherein the connection point of the minimum degree β, if the same two-point connectivity occurs, then you can take any.In theory, from the N (0,0) according to the connection sequence to N (0,2) of serving cell as shown in Figure 3 :

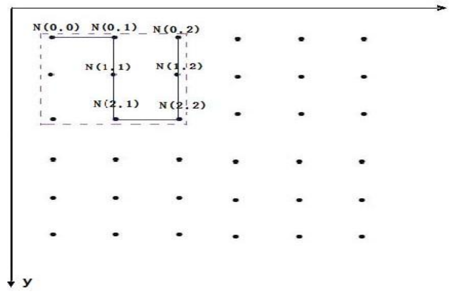

Figure 3 Schematic Zoning point search

Theorem 1: For all candidates c∈Ck, if the Lk-1 ∞Lk-1 generated and vt∈c, there are t∈Lk-1, it can also be generated by Lk-1∞L1.

Proof: If vc∈Ck, there is c∈Lk-1∞Lk-1, and c | Lk-1∞L1. Because vt∈c, there are t∈Lk-1, and because c | Lk-1∞L1, then there is ct | L1. This obviously is a length ct and c∈Ck in turn generate frequent item sets contradictory. So, Pc ∈Ck, it is c∈Lk-1∞L1.

Base on the Theorem 1, as the omission point mark, the follow-up to search the merge process, and define the starting search point for the second time, the external rectangle adjacent to the horizontal minimum point, such as Figure 4:

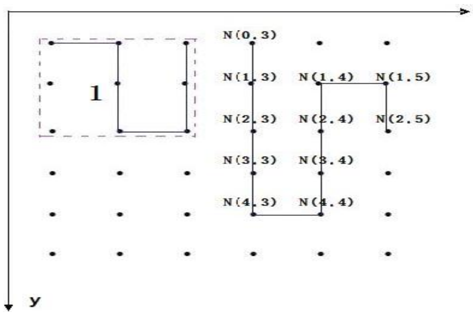

Figure 4 Schematic diagram of segmented regions

According to the above algorithm is divided again from N (0,3) as a starting point, N (2,5) end of the cell and make the circumscribed rectangular reference numerals shown in Fig.5:

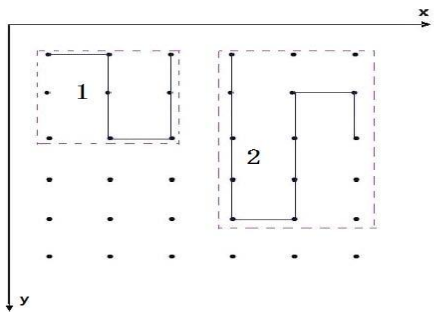

Figure 5 schematic view of divided areas numeral

(4)consolidation of the missing points:
In turn divided according to the search algorithm, such as the density of search and merge point, eventually draw N rectangular area, At this time, the missing markers combined treatment: Figure 6:to the 1、3 district,, adjacent to the missing point of N (1,0), N (2,0), N (3,1), N (3,2), when the merge process: determine P1 (plus) and P3 (plus): If: P1 (plus)> P3 (plus), will be incorporated into three zones adjacent to the missing points; P1 (plus) <P3 (plus), will be incorporated into an area adjacent to the missing point.

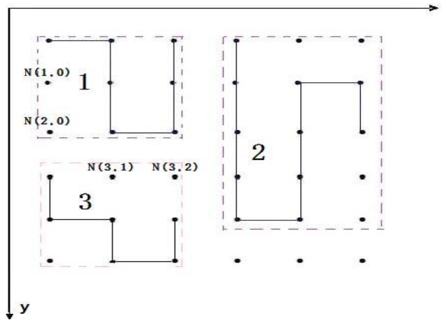

Figure 6 Schematic omitted from the consolidation point

5 TEST RESULTS AND ANALYSIS

The model takes a service platform scheduling data in a certain time period analyzed as an example in Figure 7. Search for travel route from the service node to Zone 2 and Zone 1 from full amount of statistical data serving node 2 zone to a zone, as shown in Figure 8. For data preprocessing, analysis of the service node to take the circumscribed rectangular area made, mesh refinement adjusted down to the region of the service nodes each server on the outgoing data path shown in Figure 9. Technology to reduce the use of all data points within the grid in the direction of the outgoing direction of transmission speed and the speed statistics, and considering the combined treatment based on the missing points of the colored grid, shown in Figure 10. Through 10, according to the intersection with not crossing point can be clearly seen in the case of a time period, the service node abnormal congestion, in order to avoid an explosion of abnormal state service to provide strong technical support.

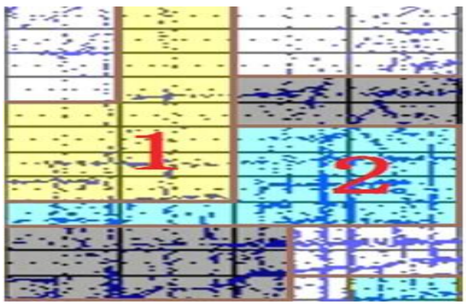

Figure 7 service node area map

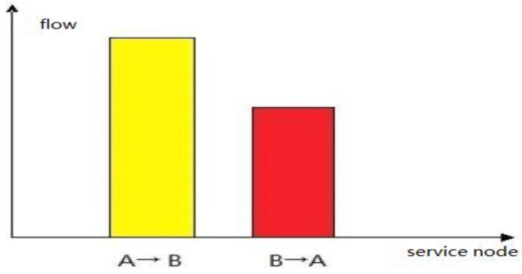

Figure 8 Flow chart data service node

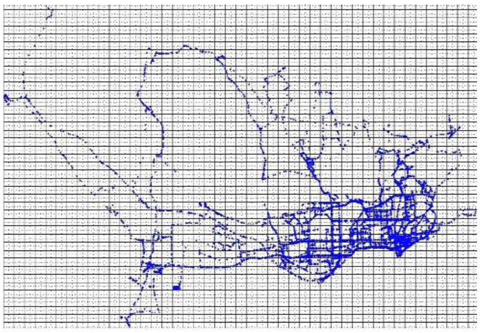

Figure 9 Outgoing data path diagram

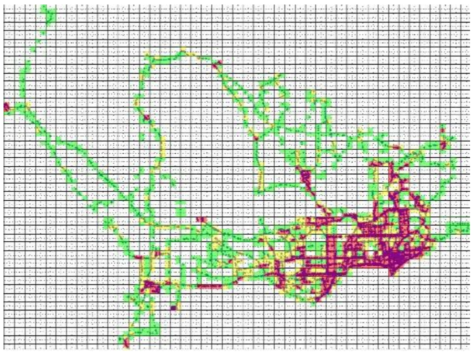

Figure 10 Data transfer speed paths merge omissions node coloring graph

6 CONCLUSION

Due to the huge amount of data collected, it is impossible to accurately each data processing, we took a fuzzy batch data improved, it will definitely affect the accuracy of the partition, the algorithm should be improved in order to improve data search and processing efficiency. The model used in the mesh, the service area boundary must smooth vertical and horizontal, will not exist in reality in this case, the improved algorithm, will more closely resemble the actual partition boundaries marginalized. This model is based on the assumption, consider the problem of disk arrays and transmission speed in all directions, you can more accurately determine the congestion in the transmission path and the service area, it will be better to avoid abnormal state service explosion. However, this model through the study of Apriori algorithm that " based on MapReduce local depth first connection -merge pruning algorithm for mining," the use of " equal density grid connection area attunement algorithm " deep into the MapReduce technology which, in order to better complete service node partitioning and data merging issues and complete service exception Miao service scheduling explosion caused the state raised strong technical support.

ACKNOWLEDGMENT

Funding information: The general project of Education Department of Liaoning province (No.L2012489); Dalian Polytechnic University Youth Fund Project (No. QNJJ201424).

REFERENCES

[1] KOLISCH R, PADMAN R.An integrated survey of deterministic project scheduling. Omega, pp.249-272, March, 2001

[2] SHOU Yong-yi, FU Ao. Multi-colony ant algorithm for multi-objective resource constrained project scheduling. Journal of Zhejiang University. pp. 51-55, Jan.2010,

[3] Labrin id is A lexandros, Roussopoulos Nick. Update propagation strategies for improving the quality of data on the Web/ /Proceeding s of the 27th International Conference on Very Large Data Bases. pp. 391-400. Apr. 2003

[4] MENG Ling-fen. Research on Job Scheduling Algorithm Based on pc Cluster. DONG Ying: China University of Petroleum, Apr. 2009.

[5] Zuo X Q，Wang C L, Tan W. Two heads are better than one:an AISTS-based hybrid strategy for job scheduling problems. International Journal of Advanced Manufacturing Technology, pp155-158, Sep, 2012.

[6] Pautz Shawn D. An algorithm for parallel Sn sweeps on unstructured meshes. Nuclear Science and Engineering , pp, 111-136,Apr,2008

[7] Li K Y, Willis R J. An iterative scheduling technique for resource-constrained project scheduling. European Journal CLF Operational Research, pp.370-379. Oct. 2002

[8] Boctor Fayez F. Some efficient multi-heuristic procedures for resource constrai-ned project scheduling. European Journal of Operational Research, pp. 3-13, Jun., 2006

[9] Thomas P, Salhi S. An investigation into t he relationship of heuristic perform-ance with network-resource characteristics. Journal of the Operational Research Society, pp.34-43, Oct, 2009

[10] SookKyong C, KwangSik C.Fault tolerance and QoS scheduling using CAN in mobile social cloud computing. Cluster Computing, pp911-926, March, 2014.

[11] Dmitry I, Boris S, Evelio A D. Integrated dynamic scheduling of material flows and distributed information services in collaborative cyber-physical supply networks. International Journal of Systems Science: Operations & Logistics, pp18-26, Jan, 2014.

[12] Byung D S, Jonghoe K, Jeongwoon Kim, et al. Persistent UAV Service: An Improved Scheduling Formulation and Prototypes of System Components. Journal of Intelligent & Robotic Systems, pp221-232, Feb, 2014.

[13] Zhang Q, Manier H, Manier M A.A genetic algorithm with tabu search procedure for flexible job shop scheduling with transportation constraints and bounded processing times. Computer&Opetations Research, pp1713-1723, July, 2012.