A weighted value scheduling algorithm based on Hadoop computer platform

Feng Liu

Information Engineering Institute,
Yulin University, Yulin 719000,Shanxi,China

Wengang Zhou*

School of Computer Science and Technology,
ZhouKou Normal University, Zhoukou 466001, Henan,China

Yang Zhao

Department of Electronic and Information Technology,
Jiangmen Polytechnic,Jiangmen 529090,Guangdong,China

**Corresponding author is Wengang Zhou, email:zhouwengang@zknu.edu.cn*

Abstract
Although Hadoop at the core of computing systems , but the efficiency is very low, ,and due to the Map Reduce scheduling system is not well on the process for priority classification, This article from the efficiency point of view, allocate weight values of task priority, In the weighting process, According to the Container resource request amount, expect a resource located node, the number of Container, submitted to the four main aspects of time for modeling Container, Because the running priority, making the system to improve the operational efficiency.
Key words: HADOOP, CONTAINER, SCHEDULING ALGORITHM,WEIGHTED VALUE, PRIORITY

Introduction

1980s can be said to be a milestone in the development of computer network, because of the emergence of the computer local area network, connected to an AC joint operation of multiple computers and information [1]. Computer systems are becoming increasingly complex.

With the advent of cloud computing systems, Hadoop in large-scale processing of data obtained in a wide range of applications [2], due to the Hadoop has many features, such as increasing the number of cluster points that can be edited, extended cluster size. Hadoop itself has a central role as a computing platform. However Hadoop also has some drawbacks, the most important is the Hadoop appear in low efficiency of the scheduling process. Since the process for the job priority is not given a better determination, leading to Hadoop reduced efficiency in the process [3].

We all know, use of the network, has been inseparable from contemporary college students , college students learn everyday life, chat, messaging, email , news, video , and many other embodies the importance of networking , the paper got the rate of change of various types of network application usage 2010-2011, shown in Fig.1. Along with the growing popularity, the use of computers to have higher requirements to the performance, for Hadoop, the existing scheduling algorithm including delay scheduling method [4], Newest algorithm [5], etc.

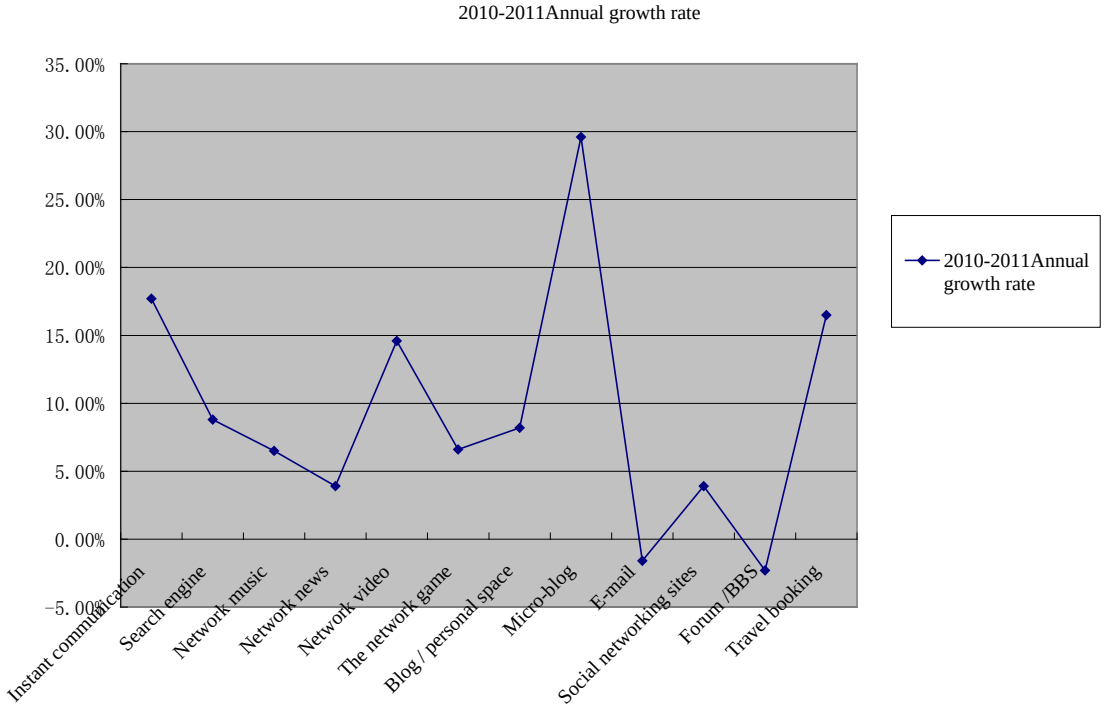

Figure 1. The rate of change of various types of network application usage 2010-2011

The Hadoop task scheduling

The resource allocation process is shown in Figure 2.

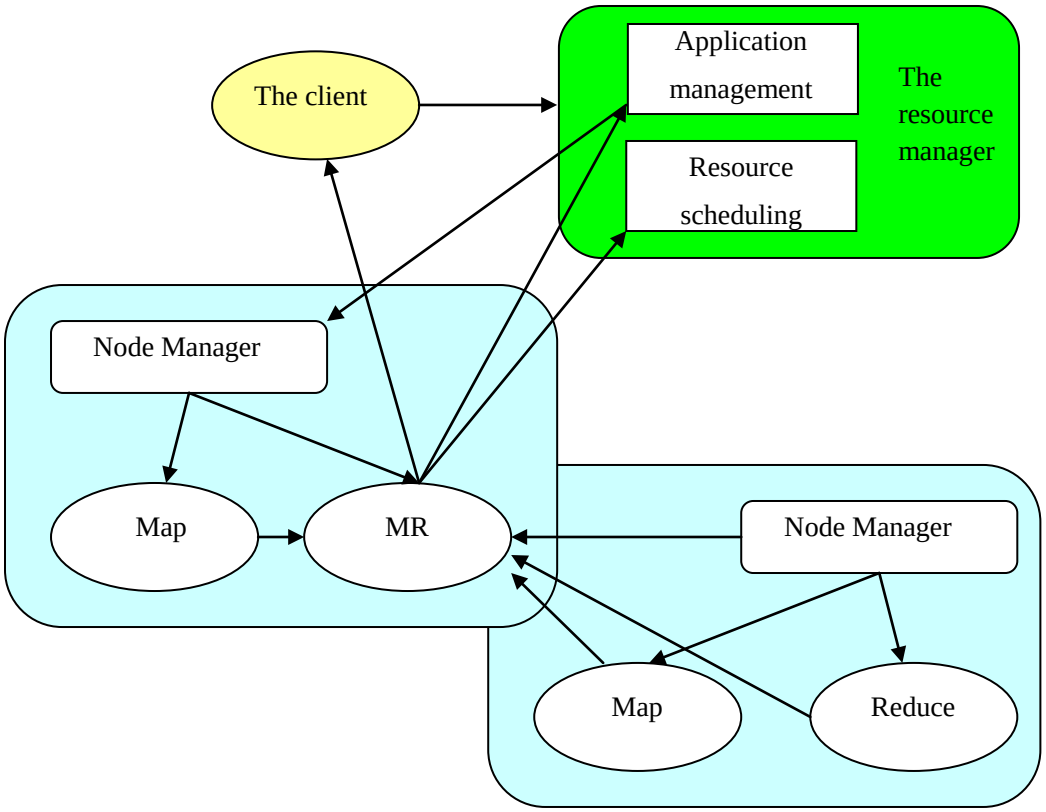

Figure 2. The resource allocation process scheduler

The resource allocation process is:

(1) Node manager finishing node information through the periodic response;

(2) Resource manager and node manager returns the periodic response, release the Container list information;

(3) Resource manager receives a node manager information trigger node update event;

(4) Resource scheduler in the receive node updates after the incident, according to certain strategy will be the strategy of resource allocation for each application node;

(5) Application to the resource manager sends periodic signal, to receive the newly allocated Container;

(6) Resource manager receives the application lifecycle information after the allocation of Container response to periodic returns;

(7) The application receives the new distribution Container distribution list, internal task.

Establish class hierarchy

(1) Establish the factors set U ,let $U=\begin{pmatrix}U_1 & U_2 & \cdots & U_k\end{pmatrix}$.

(2) The establishment of evaluation set V ,let $V=\begin{pmatrix}V_1 & V_2 & \cdots & V_n\end{pmatrix}$.Under the general assessment system to determine selection level domain, $V=\{V_1,V_2,V_3,V_4\}$.

(3) The establishment of fuzzy evaluation matrix mapping from U to V , the fuzzy relation matrix obtained as follows,First of all, on any one u_i made an evaluation of the $f(u_i)(i=1,2,\cdots,n)$, we can get from U to V fuzzy relations, that is, $f:U\to F(U)$

$$u_i \mapsto f(u_i)=(r_{i1},r_{i2},\cdots,r_{im})\in F(V) \quad (1)$$

Where, $\sum_{i=1}^{n} r_{ij} \quad j=1,2,3,\cdots,m$.

Fuzzy relation can be get from fuzzy mapping,that is, $R=\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix}$.Each line of fuzzy relation R, reflects the degree of influence factors on the object of the line judge, at the same time.Each column of R, row reflects the extent to which factors influence the judgment of the object.

(4) Establish weight set, $A=(a_1,a_2,\cdots,a_n)\in F(U)$, satisfy the conditions :

$$\sum_{i=1}^{n} a_i = 1 \quad a_i \geq 0$$

$$B = A \cdot R$$

$$= (a_1, a_2, a_3, \cdots, a_n) \cdot \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix} \quad (2)$$

$$= (b_1, b_2, b_3, \cdots, b_n)$$

Establish goals, rule, and program level relationships.Target level: The effect of Container selection.Factors programs : the criterion level [6], c_1 is container resource request amount、c_2 is the number of Container、c_3 is expect the node resources 、c_5 is container commits time. As shown in Figure 3.

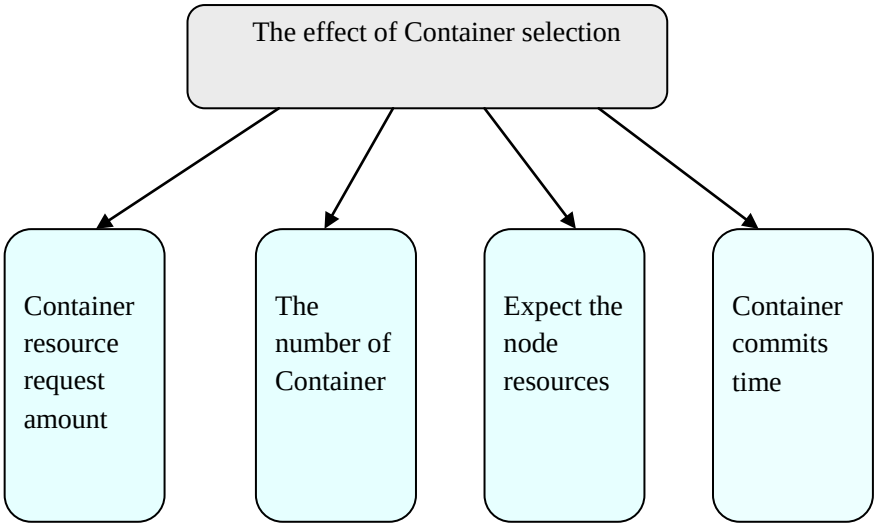

Figure 3. Class hierarchy

Constructing judgment matrix (paired comparison)

In order to make pair wise comparison between the various factors to be quantified judgment matrix, the introduction of a scale of 1 to 9, shown in Table 1.

First solve the judgment matrix, based on the above principles; the reference scale is set from 1 to 9 and according to experts and a lot of experience and reference documents, obtained pair wise comparison matrix under four criteria, the results of Table 2.

Table 1. 1 to 9 scale table

Scale a_{ij}	Define
1	i and j factors are equally important
3	i factor is more important than j factor slightly
5	i than j factors are more important
7	i than j factors is very important
9	i factor than j is absolutely important
2,4,6,8	The middle scale values
Reciprocal	If i and j factors compare, obtained value is, $a_{ji} = 1/a_{ij}$, $a_{ii} = 1$

Table 2. Comparison Matrix

	c_1	c_2	c_3	c_4
c_1	1	8	5	3
c_2	1/8	1	1/2	1/6
c_3	1/5	2	1	1/3
c_4	1/3	6	3	1

Hierarchical single sorting and consistency check

Tested with consistency index:

$$CI = \frac{\lambda_{\max} - n}{n - 1} \quad (3)$$

Where λ is the maximum Eigen values comparison matrix, n is the order of the comparison matrix. The greater the value of CI can be determined by the end of the consistency of the matrix. On the contrary, the judgment matrix has sex small.

Hierarchy total sorts and consistency check

Comparison Matrix is $A=\begin{Bmatrix} 1 & 8 & 5 & 3 \\ 1/8 & 1 & 1/2 & 1/6 \\ 1/5 & 2 & 1 & 1/3 \\ 1/3 & 6 & 3 & 1 \end{Bmatrix}$,

Column vector normalization is $\begin{Bmatrix} 0.603 & 0.470 & 0.526 & 0.667 \\ 0.075 & 0.059 & 0.053 & 0.037 \\ 0.121 & 0.118 & 0.105 & 0.074 \\ 0.201 & 0.353 & 0.316 & 0.222 \end{Bmatrix}$,

Row vector summation is $\begin{Bmatrix} 2.266 \\ 0.224 \\ 0.418 \\ 1.092 \end{Bmatrix}$,

When they are Normalized,

$$W^{(0)}=\begin{Bmatrix} 0.567 \\ 0.056 \\ 0.104 \\ 0.273 \end{Bmatrix}$$

$$AW^{(0)}=\begin{Bmatrix} 1 & 8 & 5 & 5 \\ 1/8 & 1 & 1/2 & 1/6 \\ 1/5 & 2 & 1 & 1/3 \\ 1/3 & 6 & 3 & 1 \end{Bmatrix}\begin{Bmatrix} 0.567 \\ 0.056 \\ 0.104 \\ 0.273 \end{Bmatrix}=\begin{Bmatrix} 2.554 \\ 0.225 \\ 0.422 \\ 1.110 \end{Bmatrix}$$

$$\lambda_{\max}^{(0)}=\frac{1}{4}\left(\frac{2.354}{0.567}+\frac{0.225}{0.056}+\frac{0.422}{0.104}+\frac{1.110}{0.273}\right)=4.073$$

$$w^{(0)}=\begin{pmatrix} 0.156 \\ 0.367 \\ 0.204 \\ 0.273 \end{pmatrix}$$

Tested with consistency index : $CI=\frac{\lambda_{\max}-n}{n-1}$, $CR=\frac{CI}{RI}$, RI values can be get from Table 3.

Table 3. RI values

n	1	2	3	4	5	6	7	8	9	10	11
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49	1.51

where $\lambda^{(0)}{}_{\max}=4.073, RI=0.9$,

$CI=\frac{4.073-4}{4-1}=0.24$,

$CR=\frac{CI}{RI}=\frac{0.024}{0.90}=0.027<0.1$.The judgment matrix A represents the degree of inconsistency within the allowable range , available at this time instead of the eigenvectors of A weight vector .

The use of a hierarchy chart to get the results shown in Figure 4:

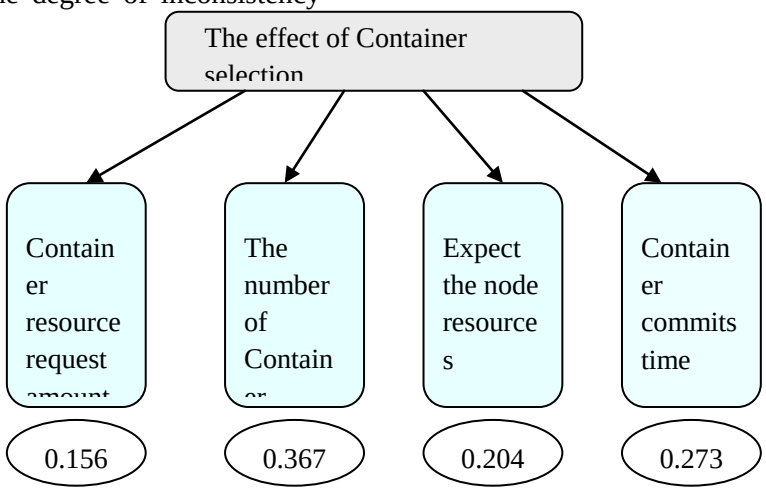

Figure 4. Hierarchical structure

Empowerment results obtained: Container resource request to obtain the weights for the amount of 0.156, the number of Container gain weight is 0.367, Container expected resource where nodes obtain weight is 0.204, and Container submission time gets the weights for the 0.273.

Conclusions

Job scheduling system is the core Hadoop platform, largely affected the performance and efficiency of the system platform. This paper uses the AHP empowerment scheduling algorithm in computing power on the basis of the algorithm, based on the amount of resource requests, according to the Container , expect the node resources , Container Number , Container weights submit time , when the system is idle high priority scheduling weights operations. To enable the system to adjust the order of execution of operations, reduce waiting time, and improve the operating efficiency of the system.

Acknowledgements

This work was supported by the Science and Technology Planning Project of Shanxi Provincial Department of Education(2013Jk1168);Yulin Municipal Bureau of Science and Technology Research Project (2014cxy-09);the Research Projects of Henan Province under Grant (142400411058, 132300410276).

References

1. Wenzheng Li (2007)*Modern computer network technology*. China Water Power Press: Beijing.
2. Chengxi Dong(2014)*Hadoop technology insider: in-depth analytical YARN architecture design and implementation of the principle*. Mechanical Industry Press: Beijing.
3. Dean J, Ghemawat S.(2008) Map Reduce: simplified data processing on large clusters. *Communications of the ACM*, 51(1), p.p.107-113.
4. Zaharia M, Borthakur D, Sen Sarma J(2010). Delay scheduling: a simple technique for achieving locality and fairness in cluster scheduling.*Proc. Conf. on Computer Systems*. ACM,p.p.265 -278.
5. Hu Dan, Yu Jiong(2014)Improve the newest scheduling algorithm of Hadoop platform. *Computer Engineering and Application*, 50(4),p.p.86-89.
6. Jayalath C, Stephen J, Eugster P.(2014) From the Cloud to the Atmosphere: Running MapReduce across Data Centers.*IEEE Transactions on Computers*,63(1), p.p.74-87.