Workload Forecasting Based on Big Data Characteristics in Cloud Systems

Chapter · January 2020

DOI: 10.1007/978-981-15-3284-9_75

2 authors:

R Kiruthiga
SRM Institute of Science and Technology

Akila D.
Saveetha College of Liberal Arts and Sciences

Workload Forecasting Based on Big Data Characteristics in Cloud Systems

R. Kiruthiga and D. Akila

Abstract Resource allocation for big data streams in cloud systems involves selecting the appropriate cloud resources. Big data has certain precise features such as size, speed, veracity, variety, and value. In this paper, a workload forecasting system for resource allocation in big data streams is developed. In this system, the data characteristics such as data type (variety), size (volume), and deviation in data flow rate (velocity) are extracted. Based on these data characteristics, the expected workload of the next time interval is predicted using support vector machine (SVM). Followed by this, the cloud resource manager dynamically allocates the available cloud resources depending on the predicted workload. The presentation outcomes have confirmed that the proposed system has less execution time and achieves better utilization of resources, when compared to the existing tools.

Keywords Big data · Resource allocation · SVM · Workload forecast

1 Introduction

In big data uses, numerous types of enormous information are produced, treated, communicated, and deposited in each instant. Certain big data sellers afford big data as a service (BDaaS) that offers users right to use to serviceable gathered and augmented data, along with their exact modified necessities [1]. Due to the transient nature of cloud servers, management of big data applications becomes complicated [2]. Significant data handling agenda is embellishing more widespread to undertake huge quantities of data in a native or a cloud organized group [3].

Resource allocation for big data streams in cloud systems involves selecting the appropriate cloud resources. Resource allocation based on data characteristics is challenging for big data since the features of information in big data rivulets are strange [4].

R. Kiruthiga (✉) · D. Akila
School of Computing Sciences, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, Tamil Nadu, India

S.-L. Peng et al. (eds.), *Intelligent Computing and Innovation on Data Science*, Lecture Notes in Networks and Systems 118,
https://doi.org/10.1007/978-981-15-3284-9_75

In static allocation, the user specifies how many doers, interiors, reminiscence, etc., a solicitation can ensure. In dynamic allocation, some idle executors may be released to return certain revenues to the group which may also be returned later on if desired [3]. The issues are

- When a single application is executing using default resource allocation mechanism, it consumes all the resources, thereby preventing other applications from sharing the resources.
- Defaulting source distribution appliances may possibly not exert as any solicitation with a firm target might have to hold in the FIFO line.
- Unfitting source distribution in both stationary and active source distribution methods might disturb the targets.
- A virtual machine (VM) may require more CPU resources, while another VM needs more network bandwidth or memory. Such a dynamic imbalance of resources in individual VM leads to the total inefficiency of cluster resources.
- Cost, performance, and availability are the main concerns of resource allocation.

Some existing examination on cloud data distribution concentrated on only one of these constraints. But, it involves high time complexity. In this paperwork, the data characteristics are extracted based on the type, size, and deviation in data flow rate. Based on these data characteristics, the current data segment is analysed and the expected workload of next time interval is predicted. The cloud resource manager dynamically allocates available cloud resources depending on the predicted workload.

2 Related Works

Kaur et al. [4] have suggested a scheme that foretells the data features regarding size, velocity, diversity, inconsistency, and accuracy. The anticipated values are conveyed in an increase fourfold CoBa. Zhang et al. [5] have suggested a novel temporary load predicting agenda centred on big data skills. Initially, a group investigation is done to categorize regular load decorations for distinct loads by means of keen measure data. Then, a link investigation is made use to regulate acute prominent features. At that time, correct predicting replicas are selected for diverse load designs. In conclusion, the predicted entire structure load is got over an accumulation of a distinct load's predicting outcomes.

Tang et al. [6] have suggested an enhanced LSTM expectation exemplary only if its precise report and a fault back proliferation technique. The relative case studies validate that their suggested amended LSTM expectation exemplary can attain advanced accurateness and actual presentation in extensive computing structures.

Dhamodharavadhani et al. [7] have reviewed the V-characteristics of big data for knowing the evolutionary stages of big data and big data opportunities. This review summarizes that even the basic V's volume, variety, and velocity still make the best part in big data analytics, but none of these stand on their own.

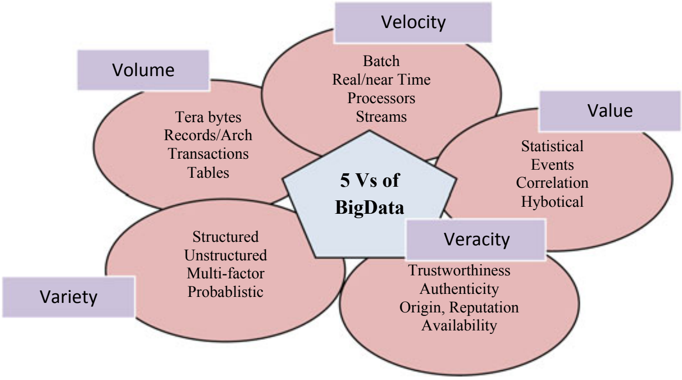

Fig. 1 5V's of big data

2.1 *Proposed Solution*

2.1.1 The 5V's of Characteristics of Big Data

Big data has been described centred on few of its features. There are five features that have been utilized to describe big data. (similarly known as 5V's) (Fig. 1).

Volume: it denotes the amount of data collected by a group.
Velocity: it denotes the growing rapidity at which this data is made.
Value: it is a significant characteristic of the data described by the added value in which the composed data can take to the anticipated procedure or trade [8, 9].

2.2 *Estimating the Volume and Data Flow Rate*

The steps involved in this process are explained in this section.

The functions of MapReduce-based framework are illustrated in Fig. 2.

In the beginning, the elementary data rivulet is arbitrarily divided into numerous map functions. The amount of map functions differs along with the data onset rate. Every map function appraises volume and velocity by means of Kalman filter. The assessed values of volume and velocity are directed to the corresponding decrease functions EVol() and EVel(), correspondingly. These two reduce functions combine the obtained values to determine the cumulative volume and velocity.

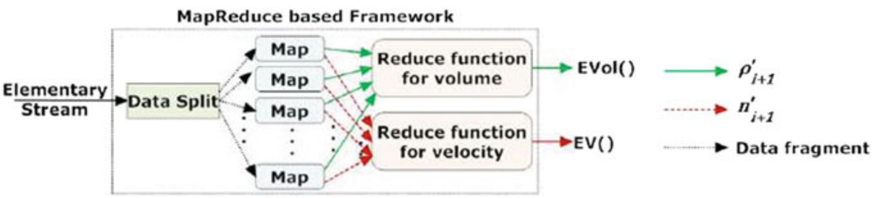

Fig. 2 MapReduce framework

The particulars of map and decrease functions are given below.

Map function: it implements the forecaster corrector calculations of Kalman filter. These calculations are applied by each map function on the fundamental data streams. Initially, the predictor estimate the size, speed, and fault covariance of $(i+1)$th data segment, afore it really attains, by means of the subsequent calculations

$$\rho'_{i+1} = \alpha_1 \rho_i + \alpha_2 \rho_{i-1} + \cdots + \alpha_q \rho_{1-q+1} \tag{1}$$

where

$$\alpha_j = \frac{\text{covariance}(\rho_i, \rho_{i-j})}{\text{variance}(\rho_j)} \tag{2}$$

$$\eta'_{i+1} = \beta_1 n_i + \beta_2 n_{i-1} + \cdots + \beta_q n_{i-q+1} \tag{3}$$

where

$$\beta_j = \frac{\text{covariance}(n_i, n_{i-j})}{\text{variance}(n_j)} \tag{4}$$

$$\Omega'_{i+1} = \Omega_i + Q \tag{5}$$

where

ρ'_i, ρ_i are the estimated and corrected volumes of ith segment.
η'_i, η_i are the estimated and corrected velocities of ith segment.

Equation (1) evaluates the capacity of $(i+1)$th data section by putting on unevenness into concern. Akin is the situation for speed evaluation in Eq. 3.

When the $(i+1)$th data segment is received, the corrector applies the following equations to modify these estimated values.

$$\rho_{i+1} = \rho'_i + K_i(y_i - \rho'_i) \tag{6}$$

$$\eta_{i+1} = \eta'_{i+1} + K_i(z_i - \eta'_i) \tag{7}$$

where

K_i is the Kalman gain at ith prediction step, given by
y_i is the ith measurement of volume
z_i is the ith measurement of velocity.

Reduce function: the EVol() reduce function estimates the average (Avg) of these entire values.

1. If $0 \leq \text{avg}(\rho'_{i+1}) < (\mu(\rho))$, thenEvol() = "Low"
2. If $(\mu(\rho)) - \sigma(\rho)) \leq \text{Avg}(\rho'_{i+1} \leq (\mu(\rho)) + \sigma(\rho))$, thenEVol() = "Medium"

3. If $\text{Avg}(\rho'_{i+1}) > (\mu(\rho)) + \sigma(\rho))$, thenEVol() = "Low"

Similarly, the Avg of all velocities and compare with mean and std. Deviation.

2.3 SVM-Based Total Load Prediction

For diverse load arrangements, diverse support vector machine (SVM) prototypes and factors are advanced to make sure the predicting accurateness inside the essential confines. SVM is an actual method for organization and reversion difficulties. SVMs are overseen learning prototypes with related learning procedures that examine data and identify designs made use for organization and reversion examination. SVM can competently achieve a nonlinear organization by means of the kernel hoax, indirectly plotting the efforts into high-dimensional characteristic places.

The sustenance vector reversion difficulty is exposed underneath:

$$\min_{\omega,b,\xi,\xi^*} \frac{1}{2}\omega^T \omega + C\sum_{i-1}^{n}(\xi_i + \xi_i^*) \tag{8}$$

Subject to:
$y_i(\omega^T \phi(x_i) + b) \leq \varepsilon + \xi_i,$
$(\omega^T \phi(x_i) + b) \leq \varepsilon + \xi_i),$
$\xi_i, \xi_i^* \geq 0, i = 1, \ldots, n$

where $(x_1; y_1)\ldots(x_\text{n}; y_\text{n})$ are a couple of input and output trajectories, n is the numeral of models, ! is mass factor, *b* is the verge value, and *C* is fault charge.

Input models are plotted to advanced dimensional area by means of kernel function φ; ϵ_i is the higher working out fault; ξ_i^* is the minor working out fault bound by ϵ-insensitive duct.

The entire structure load is predicted depending on accumulation of a distinct load's predicting outcomes; once the predicting outcome of every consumer's load is got, the predicted whole load L_total can be intended by totalling the entire predicted distinct loads, along with route harm L_loss.

$$L_\text{total} = L_\text{loss} + \sum_{i=1}^{n} l_\text{user(i)} \tag{7}$$

3 Experimental Results

3.1 Generation of Workloads

The data stream types considered in this work are listed in Tables 1, 2, and 3 which

Table 1 Different stream types used

Stream type	Big data stream
1	Text
2	Image
3	Audio (VoIP)
4	Video

Table 2 Volumes of different workloads

Workload no.	Volume of stream (GB)			
	1	2	3	4
1	10	30	10	60
2	0	0	40	20
3	15	60	0	0
4	0	10	55	0
5	62	0	0	35
6	0	25	30	65
7	25	0	28	0
8	0	35	55	70
9	55	40	0	45
10	0	11	0	50

Table 3 Velocities of different workloads

Workload no.	Velocity of stream (number of)			
	1	2	3	4
1	100	350	120	600
2	0	0	140	220
3	150	700	0	0
4	0	130	580	0
5	450	0	0	350
6	0	225	375	720
7	210	0	200	0
8	0	300	500	700
9	580	410	0	450
10	0	125	0	800

show the volume (size) and velocity (data flow rate) of the workloads with respect to the mixture of given stream types.

3.2 Results

Primarily, the amount of work of 25 GB is served to EC2 work out enhanced c4.huge occurrences at velocity of 50. Every amount of work is served to the structure later

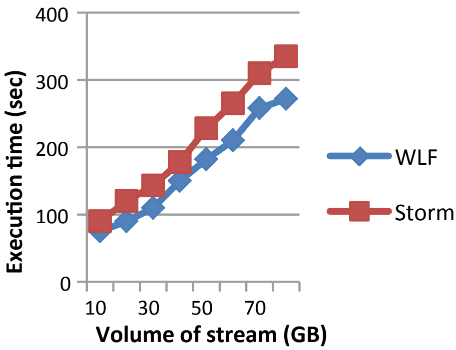

Fig. 3 Execution times for various sizes of workloads

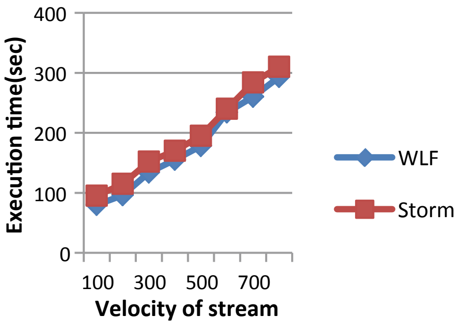

Fig. 4 Execution times for various data generation rates

every 5 min for the period of 1 h. The performance of WLF is compared with the Apache Storm 0.92 tool using the execution time and resource utilization metrics.

A. **Execution time**

In this section, the execution time for different volumes and velocities of workloads is measured and depicted in Figs. 3 and 4, respectively.

B. **Resource utilization**

In this section, the resource utilization (%) for different volumes and velocities of workloads is measured and depicted in Figs. 5 and 6, respectively.

Figures 3 and 4 show the comparison results of both the systems in resource utilization. The outcomes display that use of cloud means is advanced in event of WLF when associated to Storm.

4 Conclusion

In this paper, a work- load forecasting system for resource allocation in big data streams has been developed. In this system, the data characteristics such as the data

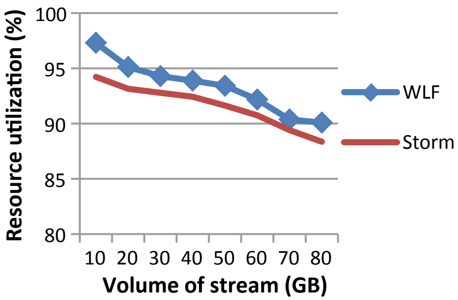

Fig. 5 Resource utilization for various sizes of workloads

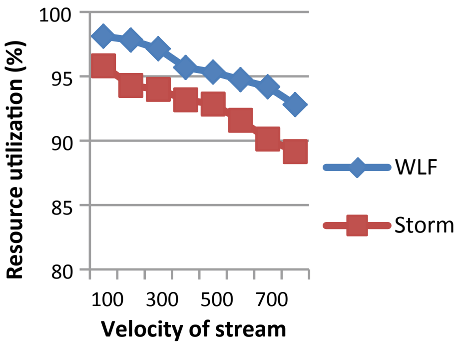

Fig. 6 Resource utilization for various data generation rates

type (variety), size (volume), and deviation in data flow rate (velocity) are extracted. Based on these data characteristics, the expected workload of the next time interval is predicted using support vector machine (SVM). Followed by this, the cloud resource manager dynamically allocates available cloud resources depending on the predicted workload. The suggested scheme is applied in a Java-based solicitation on Amazon EC2 figure improved c4.huge occurrences. The enactment of the WLF scheme is assessed by associating with the Apache Storm 0.92 device using the execution time and resource utilization metrics. Investigational outcomes display that the suggested WLF scheme has fewer execution time and achieves better utilization of resources, when compared to the existing tools.

References

1. Dai W, Qiu L, Wu A, Qiu M (2016) Cloud infrastructure resource allocation for big data applications. IEEE
2. Spicuglia S, Cheny LY, Birkey R, Binder W (2015) Optimizing capacity allocation for big data applications in cloud datacenters. In: IEEE IFIP/IEEE international symposium on integrated network management (IM), Canada

3. Islam MT, Karunasekera S, Buyya R (2017) dSpark: deadline-based resource allocation for big data applications in Apache Spark. In: IEEE IEEE 13th international conference on e-science (e-science), New Zealand
4. Kaur N, Sood SK (2017) Dynamic resource allocation for big data streams based on data characteristics (5Vs). Int J Network Mgmt
5. Zhang P, Wu X, Wang X, Bi S (2015) Short-term load forecasting based on big data technologies. CSEE J Power Energy Syst 1(3)
6. Tang X (2019) Large-scale computing systems workload prediction using parallel improved LSTM neural network. IEEE Acess
7. Dhamodharavadhani S, Gowri R, Rathipriya R (2018) Unlock different V's of big data for analytics. Int J Comput Sci Eng 6(4)
8. Ishwarappa, Anuradha J (2015) A brief introduction on big data 5Vs characteristics and hadoop technology. Preocedia Comput Sci 48:319–324 (Elsevier)
9. Hadi HJ, Shnain AH, Hadishaheed S, Haji Ahmad A (2015) Big data and five V's characteristics. Int J Adv Electron Comput Sci 2(1)