

3.3 Methodology

Algorithm1: Calculate the number of deviation of the parameters of the same model

Table 2: Algorithm1

Algorithm1
Input fault vehicle detection data
Use filter() to filter data with vehicle problem 1
Use map() to set key= vehicle model, value= vehicle problem
Use reducebykey() to increase the number of vehicles appearing on the same road condition
Loops Step2, 3, 4.
Output the number of vehicle problems according to different types of vehicles

Algorithm2: The influence of driving distance on vehicle problem based on spark

Table 2: Algorithm2

Algorithm2
Input fault vehicle detection data
Use map() to set key= distance, value= vehicle problem
Use sortByKey () to sort by travel distance
Use distinct() to remove duplicate data from vehicles
Output data using map(), map set key= vehicle problem value=1
Use reducebykey() to increase the number of occurrences
Output the Influence of driving distance on different vehicle problems

4. PERFORMANCE EVALUATION

4.1 Analysis of vehicle parameters affected by vehicle type

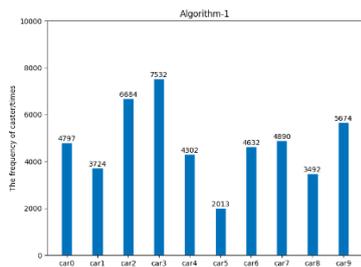


Figure 3: Analysis of Vehicle Parameters

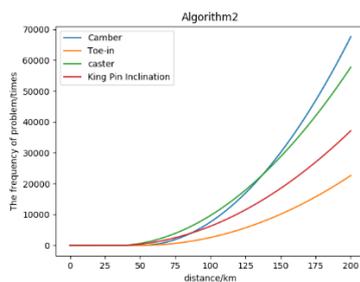


Figure 4: Analysis of Distance

The figure is the statistics of the 100,000 data, respectively, calculated the number of different models appear the number of issues. The use of the platform 100,000 data calculation time of 1 minute, the calculation results as shown. Through this method, a large number of possible shortcomings of different brands can be quickly counted.

5. CONCLUSION

This paper puts forward the analysis of mass vehicle data based on Spark platform. Under this platform, it can quickly and reliably analyze and process the vehicle data according to the existing data in Hbase, and get the expected results quickly. Has a certain degree of scalability, can dig out more effective data and information, and plays a guiding role for vehicle maintenance in the future.

ACKNOWLEDGEMENTS

This research work was supported by the National Natural Science Foundation of China (Grant No.61762031), Guangxi Key Research and Development Plan, GuangXi key Laboratory of Embedded Technology and Intelligent Information Processing.

REFERENCES

[1] Zaharia, M., Chowdhury, M., Franklin, M.J., Shenker, S., Stoica, I. 2010. Spark: cluster computing with working sets. Usenix Conference on Hot Topics in Cloud Computing USENIX Association,10-10.

[2] Vavilapalli, V.K., Murthy, A.C., Douglas, C., Agarwal, S., Konar, M., Evans, R., Graves, T., Lowe, J., Shah, H., Seth, S., Saha, B., Curino, C., O'Malley, O., Radia, S., Reed, B., Baldeschwieler, E. 2013. Apache Hadoop YARN: yet another resource negotiator. Symposium on Cloud Computing (pp.5). ACM.

[3] Chowdhury, M., Zaharia, M., Ma, J., Jordan, M.I., Stoica, I. 2011. Managing data transfers in computer clusters with orchestra. Acm Sigcomm Conference ACM, 98-109.

[4] Lee, B., Riche, N.H., Karison, A.K., Carpendale, S. 2010. SparkClouds: visualizing trends in tag clouds. IEEE Transactions on Visualization and Computer Graphics 16 (6), 1182.

[5] Ewen, S., Tzoumas, K., Kaufmann, M., Markl, V. 2012. Spinning Fast Iterative Data Flows. Proceedings of the Vldb Endowment, 5 (11), 1268-1279.

[6] Wirtz, G.P., Brown, S.D., Kriven, W.M. 2016. Ceramic coatings by anodic spark deposition. Materials and Manufacturing Processes, 6 (1), 87-115.

[7] Lu, R., Lin, X., Zhu, H., Shen, X. 2009. Spark: A New Vanet-Based Smart Parking Scheme for Large Parking Lots. INFOCOM (pp.1413-1421). IEEE.