

# The Parallel Processing of ECG signal based on Hadoop Framework

Xuecan Huang, Huhe Dai, and Ye Li\*, *Member, IEEE*

**Abstract**—Recent advances in healthcare and bioscience technologies and proliferation of portable medical devices are producing massive amounts of multi-modality data. The need of parallel processing is apparent for minimizing these data sets. In order to enhance the system processing capability, in this paper, we developed an ECG analysis model that based on Map-Reduce parallel framework. Experiments show that the improved platform’s processing facility is significantly enhanced compared with the existing serial one.

## I. INTRODUCTION

THE Health Services with a Novel Efficient Cloud (HCloud), as one part of the Low Cost Health Research, has become more and more popular in China [1]. Taking into account the future increasing throughput and large data processing, it is necessary to introduce parallel computing into the platform, in which the ECG analysis cost most computation time as well as will make the system storing more and more data. Hadoop allows for the distributed processing of large datasets across multiple computer nodes, supports big data scaling (HDFS), and enables fault-tolerant parallelized analysis (Map-Reduce). Compared with OpenMp and MPI, we do not need construct low-level functional modules and high-level data analysis algorithms, only considering the execution of the task at a higher level which is from the perspective of key-value pairs to the function, thus it can be achieved automatically in parallel, and with higher reliability.

## II. THE ECG PROCESSING FEATURES

We have test the execution speed of the original serial program with the raw data which is recorded by Mini-Holter, a device that can record human dynamic electrocardiography. The test results showed that filtering baseline wander took about 70 percent of the whole ECG processing time in our existing HCloud platform, and thus it was considered to be the first part computing in parallel on Map-Reduce framework. The filtering baseline wander iteration formula is as follows,  $y_1(i) = x(i) - x(i-90) + y_1(i-3)$ ,  $y_2(i) = y_1(i) - y_1(i-90) + y_2(i-3)$ ,  $y(i) = x(i-87) - y_2(i)/900$ , where  $x$  is the input signal and  $y$  is the output signal. Obviously, the formula is characterized by data dependency. It means that the following calculations depend on the previous results. However, if the cutting of data is large enough, the result is equivalent to filtering baseline wander of the segmenting data separately, which does not have much influence on the data processing results.

Ye Li is with Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China (e-mail: hh.dai@siat.ac.cn).

## III. THE MAP-REDUCE FRAMEWORK

According to the analysis in section II, map function can be designed to handle part of the whole data, while reduce function is to merge the output produced by the map function, and then output the entire filtered results. The framework process is shown in Fig.1. Map-Reduce works by breaking the processing into two phases: the map phase and the reduce phase. Each phase has key-value pairs as input and output, the types of which in the filtering baseline wander parallel programming is that each split represents a segment of data, in which key stands for the data fragmentation flag, while value for original ECG signal data.

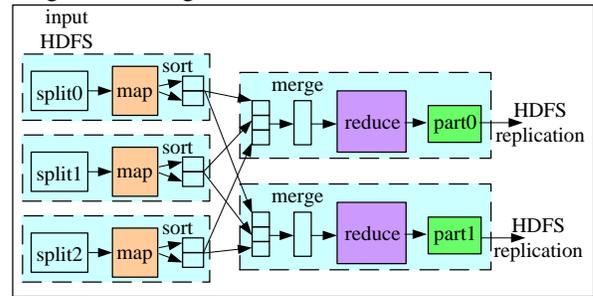


Fig.1 The MapReduce Framework

## IV. THE RESULTS AND CONCLUSIONS

We run the parallel programming on the Hadoop platform constructed by three servers, all of which are Xeon E5606, 4cores, 16processors, 2.40GHz. Table I shows the results. It can be seen that the running time is essentially the same as the data increases, which means increasing the number of nodes can guarantee the stability of time. The running time of serial processing is much longer than that of the parallel one for the same data, besides, it varies linearly with the size of data.

TABLE I  
RUNNING TIME OF PARALLEL AND SERIAL PROGRAMMING OF FILTERING

Data Length	Load Map Tasks	Parallel Programming	Serial Programming
8h	1	18s	74s
16h	2	17s	146s
32h	4	18s	292s

HCloud is oriented to people all around China. That means the amount of data analyzing by HCloud would be huge. Therefore, the introduction of Hadoop is a wise choice.

## REFERENCES

- [1] Chenguang He, Xiaomao Fan, Ye Li. “Toward Ubiquitous Healthcare Services With a Novel Efficient Cloud Platform” . IEEE Transactions on Biomedical Engineering, vol. 60, NO. 1, 2012, pp.230-234