

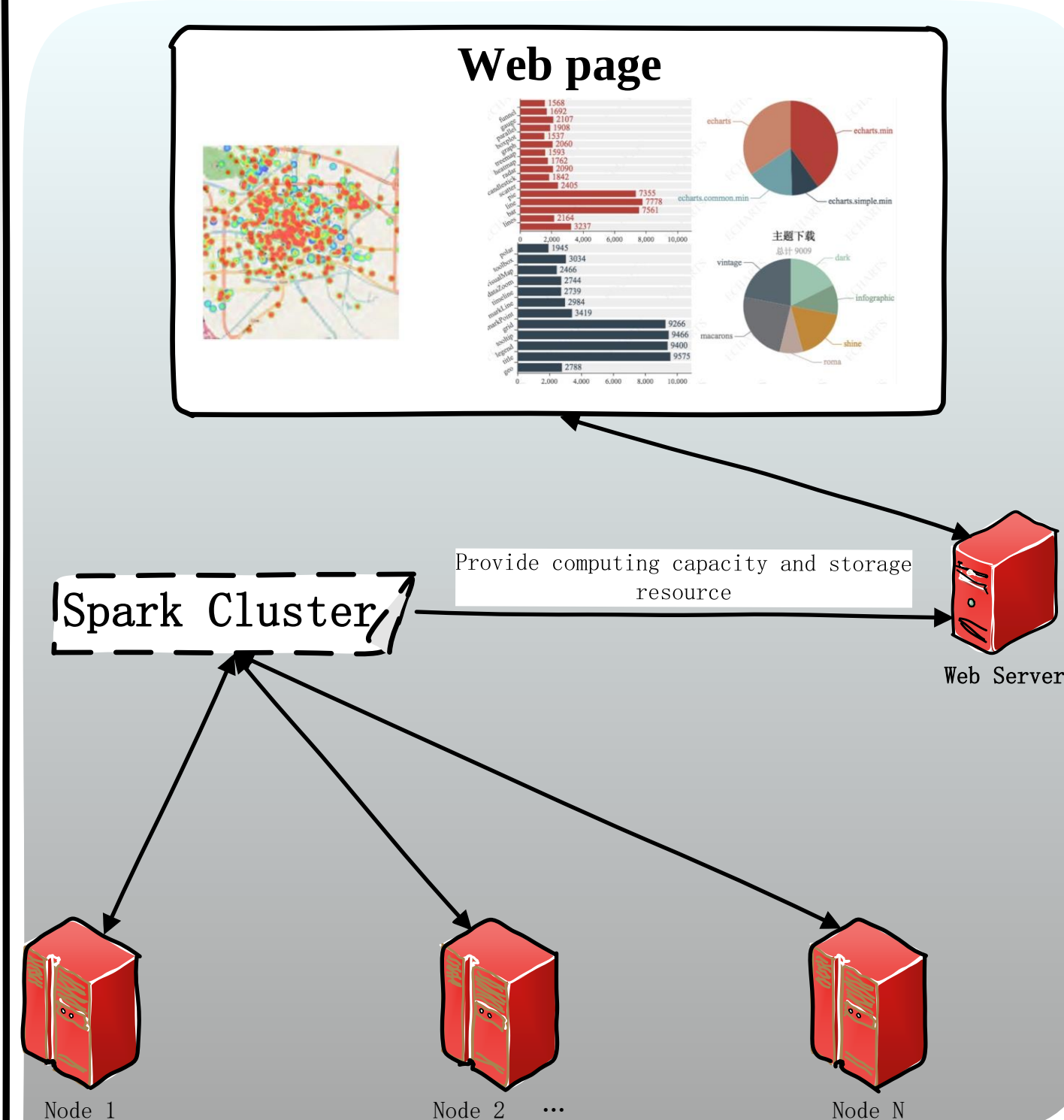
Spark-based Heatmap Visualization

Dongsheng Cao, University of Wales Trinity St. David (UNITED KINGDOM), 1805155@student.uwtsd.ac.uk

Conclusion

This summer project is to develop a web application of heatmap and data analysis based on Spark cluster in order to perform faster and to solve the problem of the large size of dataset. Using some knowledge about web technology and distributed web application that have studied in the last semester.

Graphic explanation



Abstract

In this Big Data era, facing the data scale from TB to the PB, EB and even the ZB level, it is difficult to display whole data set onto the web regarding to the heatmap of Geographic Information System (GIS). Therefore, this project aims to develop a WebGIS visualization application based on Spark to realize rapid parallel computing and storing the results on the Hadoop Distributed File System (HDFS).

Aims

The main aim of my project is to make web visualization faster and more capable facing the increasing data, realizing the heatmap visualization of WebGIS with the whole dataset using the spark cluster to improve the computing speed of a series of complex algorithm for the large size of dataset.

Background

With the increasing requirements for web visualization, there are some web visualization tools like D3, heatmap.js, the Matplotlib and Boken based on Python and so on, which almost doesn't perform well facing the large volume of data set for displaying whole data simultaneously on the web. D3 manipulates the documents of data, which is a JavaScript library to help people to make data to life by CSS, HTML and SVG (Scalable Vector Graphics). D3 actually has a variety of advantages but it is not suitable facing the data above 1 billion points [2]. Similarly, the heatmap.js using canvas to draw data also does not fit the capacity of large volume of data. In addition, the Echarts from Baidu is similar to the D3 as well.

Research and development method

In my project, I decide to use CBI model, which is explained that the steps in developing project based instructional content (CBI) That consists of three stages, those are (1) Plan, (2) Development, and (3) Evaluation. The plan stage is to make sure the major process in this summer for my project and some forums to decide what development tools will be used, what dataset will be used and what functions will be realized. The develop stage is the step to program and configure my application designed previously in practice. Besides, the final stage, evaluation, is to format my dataset to test my application's functions whether it is faster or more capable.

Some development tools will be used

Such as Spark cluster, GeoServer, Openlayer, Echart, Hbase and so on.

Objectives

1. The heatmap visualization on the web realizing 5 layers tile map.
2. Data analysis on the web.
3. Realizing the function in terms of real-time aspect.
4. Realizing some optimized functions such as using distributed database (Hbase), Ajax to dynamically transfer data of map and so on