Statistics Package With Spark Framework

David Alejandro Vásquez Carreño

August 2020

1   Abstract

In this paper, we are going to see a simple an understandable implementation of a web application with the Spark framework for Java, as well the advantages or disadvantages of certain libraries involved into this tool.

Previous implementation of a Statistics package will be put in into use in this new work, to show it's practical help into web environments. As a short package of collections and calculations, the big help it can give us in a good separation of work and libraries in our application is immeasurable in real developments.

2   Introduction

In the development of web application, there is a lot of tools, frameworks, engines and libraries that can be useful in making web pages with acceptable responsiveness and good user experience, but certain tools can be too big or too complex to resolve simple problems, and for that reason different solutions come up, like CMS systems, that prevent content creators of add extensive work to something is already created. We'll talk about a simple approach to make a very simple web application deployed in heroku.

The quantity of tools can be a little underwhelming at the moment of choosing the right parts, so simple solutions have arrived not just in java, but in a lot in programming languages like c++, python or ruby.

1

For every person looking for a practical and fast implementation of a web application in java, then this paper is the right choose for finding an excellent tool that can help in the future, as long as the implementation can keep it's size small.

First we are going to talk about the importance of web development in the life of every company in the world, and how the use web environments to show all it's products to the world. Second, we will see a simple implementation of a web application with Spark, and finally all the advantages and disadvantages that this can bring to the table.

3   Problem

As programmers, web developers, executives and businessmen, we must care about keeping companies relevant and competitive in the market, and the implementation of technologies is key nowadays to achieve this competitiveness. The moment of choosing what tool, architecture o language is going to be used is nothing easy, and it must be made carefully if we don't want to have unmaintainable and old frameworks into our companies.

Most of the products and services are going to be showed into websites, what open possibilities in this globalization era, where companies can not limit itself to one country. Web applications are essential in making the company visible and relevant into this new technological world.
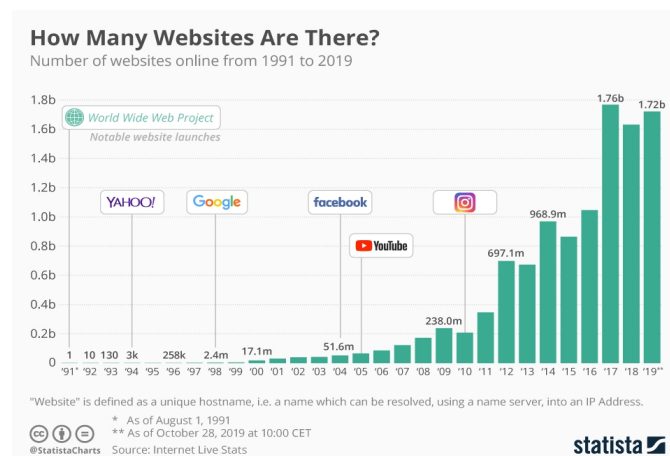

Figure 1: Website growth

All industries are making use of evolving technologies, and that makes everyone not using those, noncompetitive and irrelevant at the moment of showing and selling their organizational culture, moral values and products to the market. If we want proof of this, then we must see the most interesting companies nowadays.

Figure 2: Most Valuable Brands

Amazon is a very interesting study case, because even if their production value comes from physical products and deployment, that also support all that in their web infrastructure. Amazon has created good products and platforms like AWS, prime and recently has been involving in video games development.

4   Solution

Spark is designed to allow java know http requests and responses, and make programmerrs able to send proper data to the user by specific responses of a message with java logic. This framework gets closer frontend functions into the backend, and with some downsides, it also makes some tasks easier. Local deployment of spark involves an opened session in port 4567 by default, so an direct communication with the web browser and the localhost is established.
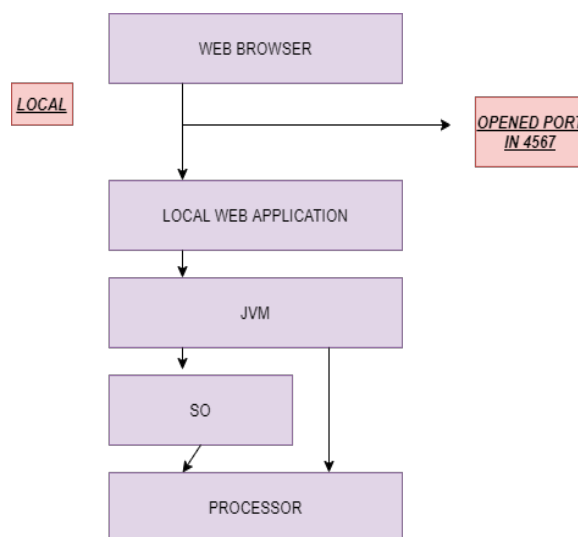
Figure 3: Local deployment

We will deploy the machine in heroku to give server responsibilities to a different machine, and also allow us to have a free domain to see how our application is going to behave in a different environment.
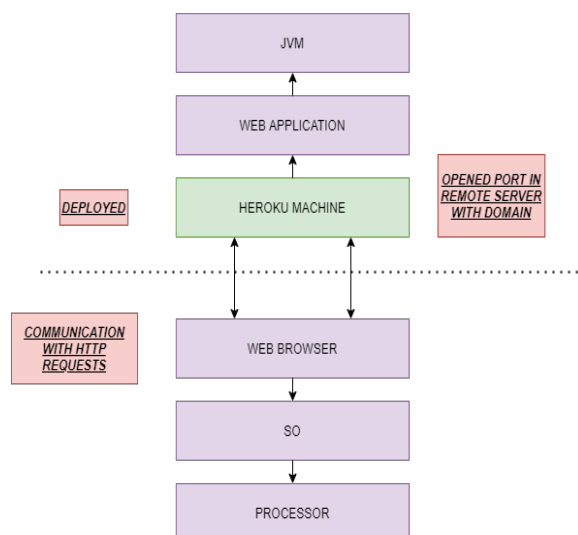

Figure 4: Heroku deployment

Every request is going to be received and will have a custom response with a message, in our case is going to be HTML files with form to set data that will be processed by the application in the future.

A downside of spark framework is to set manual response with data in memory like strings that will represent HTML files. To try to make this problem little, a new layer of files reading is added to allow developers writing custom HTML files instead of writing those in java, what is not good practice.
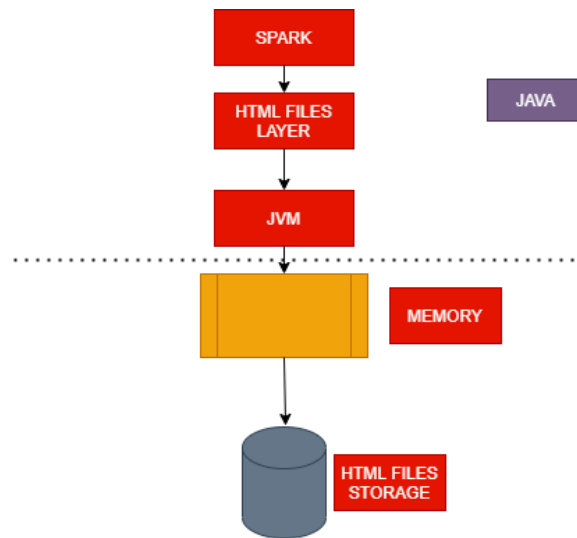

Figure 5: Memory design

This design allow us to just write HTML code in isolated files instead that writing them in java syntax. This HTML code will be sent in the responses to the client via HTTP communication with GET and POST methods. At the moment of requesting the web page, the massage will be a GET method for a resources, while at the moment of fetching parameters for the page it will be a POST method.
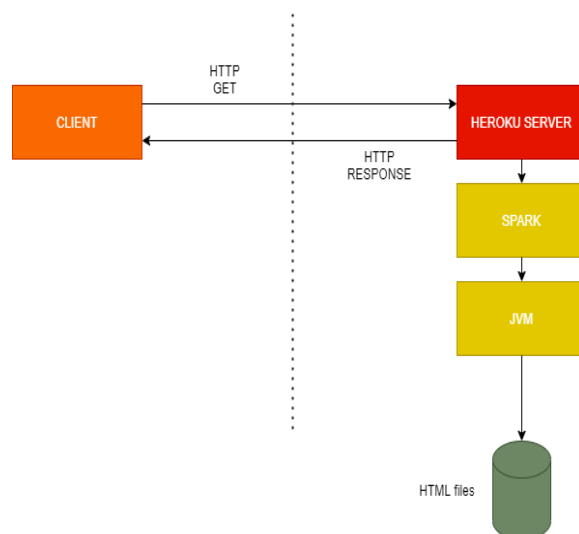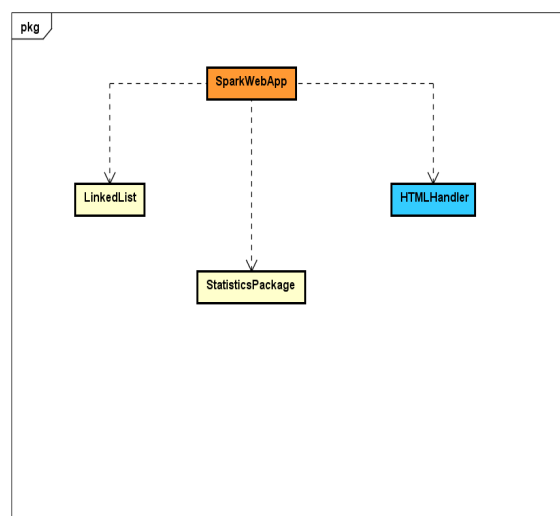
Figure 6: Communication

Figure 7: Design

Every response is handled with an specialized class in reading files, that we are using in this case to read HTML files directly from storage.

5   Evaluation

This framework adds flexibility and simplicity into a web application, but the big problem with this is the handling of HTML files into java code, that should not be connected by any means.

This implementation is fine for short systems, like a practice page, but for large environments is not that great, because is breaks the simplicity and benefits of HTML just because the backend is written in java.

6   Conclusions

We have seen fundamentals about the way java and http works in web applications and for web sites, and so benefits and downsides of this implementation. The biggest downside is of course the management of HTML files into spark is not that great, and that management should be commended to a bigger and better engine like apache tomcat.

Spark is a fun framework to learn, but not so much in a development environment, so if someone is looking for a complex handling of files and responses, then a apache tomcat is my recommendation.

References

[1]   BFGlobal500. *World's most valuable brands 2020*. 2020. URL: `https://prnoticias.com/wp-content/uploads/2020/01/b.jpg`.

[2]   StatistaCharts. *How many web sited are there?* 2019. URL: `https://cdn.statcdn.com/Infographic/images/normal/19058.jpeg`.