Hama Earvin

Bako

Big Data Programming Final Project Description

1. Website

I create two S3 buckets that will be hosting our website. The first had the name of my domain (bigdatafinalproject.com) and the second bucket([www.bigdatafinalproject.com](http://www.bigdatafinalproject.com)). The second bucket will serve as a subdomain where the traffic will be redirect to our first and main bucket.

Then we uploaded our website content (html file) into our first bucket. We also had to grant public access to our bucket so that we could add a policy and we also enabled website hosting (we use our html file as both the index document and the error document.

We also had to configure our second bucket. We went into properties to enable redirect request and redirect it to our main bucket so that we can access our website.

Our website was finally accessible through a link obtainable in the property tab of our first bucket(bigdatafinalproject.com) by clicking on static website hosting (<http://bigdatafinalproject.com.s3-website.us-east-2.amazonaws.com>).

1. Spark Streaming project

In this project, we will be using Spark Streaming to read a series of line on a local port (9999) of our computer.

I personally use putty with an Amazon EC2 instances to launch spark.

I used the same putty to launch another terminal from the same EC2 instances.

I then connected it to the same port (9999) by typing nc lk -9999 in our second terminal.

Finally, you can start the stream and start getting the readings and counting words.

Once you are done typing your lines, you can interrupt kernel and check your results.

Be careful to have that port free before running the program.