What works?

**On page** **load**

* scrapy starts scraping, puts data into database and retrieves it from the database in the desired format to generate the input file for the mapper.
* This page gets uploaded to the S3 bucket and the streaming step begins (Mapper + Reducer).
* Reducer output gets downloaded in the current directory as output\_folder and merge.py merges them to create locMergeFile.tsv

**When Location Based Search is clicked:**

* D3 page is called with field = Location and locMergeFile.tsv.
* If you click on the graph you get the corresponding listing in tabular format and simultaneously in the background the input file for mapper\_skill is uploaded to S3.

Till this stage everything has been tested.

**In process:**

* On clicking the button “Filter based on Skills” connect\_EMR\_phase2.py should be called.
* D3 for skills generated.
* Back buttons….

**Problems:**

1. **Link format**

(/pagead/clk?mo=r&ad=-6NYlbfkN0CQ7aaSVhIXdrXw9Mn\_GOE0BpUPVgeY1NIb8C-tcTbMKZb-E0TP4TvUV5wPQ9AVbt8-SYG6ToERcBi7WHqgiFQh2-Us6oZBx4jalslJjAmn1xvu7ixgWSUqikPE5z5OO\_krjPDRYOIbZw0e8QMhZSGoMHsO0kAiVByFJ6pBXGvW3cQGWRwNo-knVvSacGCKTM2YLRPiQw-V0wBIZEewVh5KUf2hoARkNbOKmcn0Z7AkH\_2RDPAqtejrjjAAFNLNyUuodCXUT0qqaNQcEBKUSBIHNqe8onZBelu\_73-fufEoQksJAja-zPPtOBu2zE3X8DgQoRhJV5lY7tKrWpXDQdYOt7sFez0yWw12yDE53geJYtCO22CqBqm8lseRKzy53BACjeeQYg0DMxpVsDHOV3TqWuBP\_KKQFHhZwIxparo9Ew==&p=1&sk=36968f8dc1b483d4&fvj=1

2**. Mapper output format**. (Old one being used.) (current o/p: company::skills::……; desired output: Location::company::skills::…….)

3. **Splitting the tsv** or use the parts. (Due to the huge tsv and unimaginably large no. of cities the D3 graph scale is not what we want.

     Possible solutions: 1) Have multiple pages ( Prev…Next) where each page takes decent no. of cities (We can also use the part files(before merging).

                                    2) Just show top 10(decent no.) locations.

                                    3) Make the mapper country based rather than cities.