

## Assignment-1 (Fractal - 3)

1. In this question, you will be working with a dataset of weblogs in CSV format. Each entry in the weblog contains several fields, such as IP address, timestamp, URL, and status code. Your goal is to develop a MapReduce mathematical algorithm for the mapper stage and reducer stage to perform various analyses on the dataset efficiently. (Programming is optional.) [10]

### Dataset:

The dataset can be downloaded from the following link: [Weblog Dataset](#)

### Tasks:

Your MapReduce algorithm should be able to perform the following tasks:

- Count requests by IP address
- Identify top 10 IP addresses with most requests.
- Find the top 10 URLs that were requested the most.
- Discover the top 10 status codes returned.

### Instructions:

- Download the Weblog dataset from the provided link.
  - Present your solutions mathematically by giving algorithms for the map stage and reduce stage without writing programming code. (Programming is optional.)
  - Include relevant tables or visualizations to support your results (if any).
2. Create a Hadoop multi-node cluster using either Virtual Machines or Docker. Submit a brief report detailing how you built the cluster, including step-by-step instructions and screenshots for each stage of the process. You may refer to previous class recordings or online resources to guide you in setting up the cluster. [10]