## General Description:

The server client module is a system which was design to cluster large amounts of data in streaming format.

The system is made of 3 main parts-

The Server (server.py)

The Client (streamingClient.py)

The Worker (worker.py)

The system works in the following manner:

Client

Server

The server receives the streamed information from the client,

It then distributes the information to the workers in 'CHUNK\_SIZE'.

Each worker creates a core-set of the part it received. This is saved as a tree.

Upon request from the client, the server calls for the current corset of each worker.

It then sends all the received core-sets to the summary worker- a dedicated worker which creates the final core-set, and returns the result to the client.

## Worker:

The worker handles the creation of the core-set.

Currently it works by creating a binary tree, where each node is of size 'LEAF\_SIZE'.

The leaves are the original data, and each level above provides a smaller and better representation of the data.

The merge of the leaves is done using k-means.  
The root holds the best representation of all the leaves: the core-set.

**\_\_Init\_\_** : constructor – defines the tree building algorithm (CoresetTreeBuilder.py)

**register\_and\_handle\_summary** : handles the creation of the summary worker

**register\_and\_handle** : handles the creation of a regular worker

**get\_summary**: returns the current root node.

## streamingClient

Streaming client mimics the client.

It generates the data and streams it to the server.

**run\_client** : connects to the server

**get\_summary\_points** : sends the server a request for a summary

**\_send\_points** : sends the data to the server