

Experiment Plan:

The experiment plan for my project largely consists of gathering timing information. I plan to time:

- data segmentation/distribution time: how long it takes to divide the dataset into blocks and distribute those blocks to the cache nodes.
- network communication time during rendering: how much time is spent waiting for blocks to be transmitted from cache nodes to render nodes.
- rendering time per node: how long an individual render node takes to render its portion of the image.
- total rendering time: how long it takes to render the entire image.
- output image composition time: how long it takes to composite the outputs from the render nodes.

I also plan to track the number of blocks requested from each cache node. By doing this I hope to get a feel for the efficiency/usefulness of my caching strategy.

In order to get an accurate picture of my program's performance, I plan to vary the following parameters:

- number of cache nodes/render nodes: from 1->many nodes. I aim to find an optimal division of I/O and processing tasks.
- transfer function: this function determines which features of the dataset are being rendered. It can play a big part in how long an image takes to render and also how much of the dataset needs to be traversed in order to render the output.
- view transformation: zoomed in/out, rotated 360 degrees. This is the angle and scale of the dataset with respect to the view plane.
- dataset size: datasets range in size from small (~4MB) to huge (many GBs)
- in-memory cache size: from ~2MB to as much real memory as the nodes have. The more memory a node can use, the less it will need to use the network or disk.