

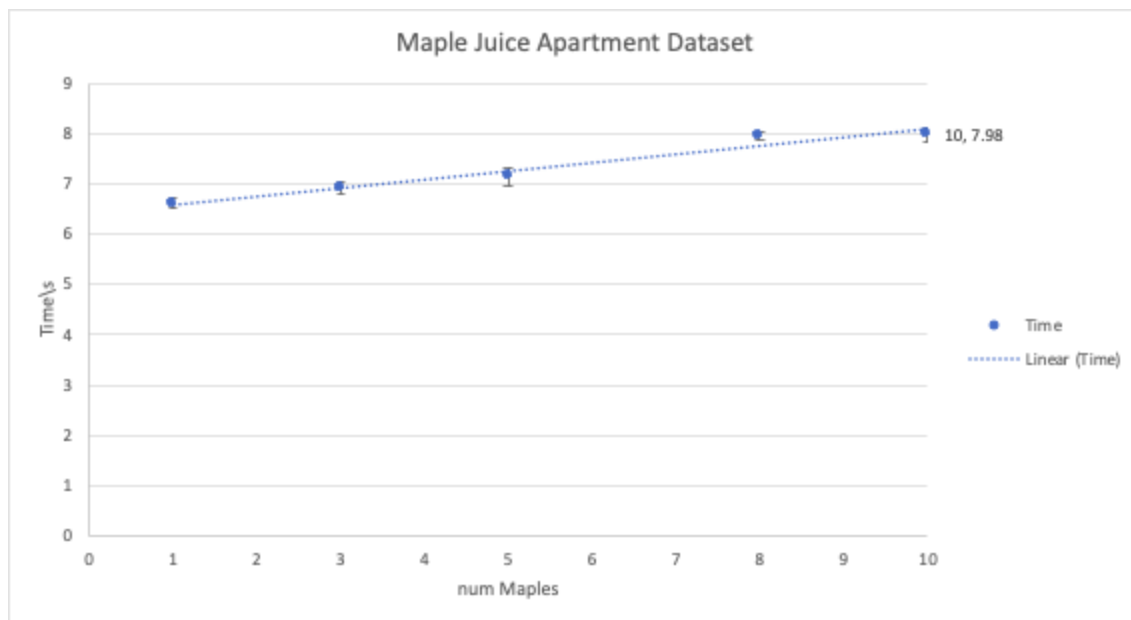
CS 425 Mp2 report (Group 10)

Design

For our MapleJuice framework we decided to make the introducer the master node. Whenever a maple or Juice task is initiated from any node in the distributed system, a maple/juice request is sent to the master along with the required information including the maple_exe and num_maples. The master then randomly decides which nodes to send a send start maple/juice request. Once a maple/juice request is received by the node, it carries out the maple/juice task after retrieving the required file from the SDFS. Once completed, the node sends a message to the master informing that the file has been processed into key value pairs and that the key files have been uploaded to the SDFS. The master waits until all the nodes have finished their assigned tasks. Once finished, the master processes the key value pair appropriately depending on whether a maple task or a juice task was initiated at the beginning.

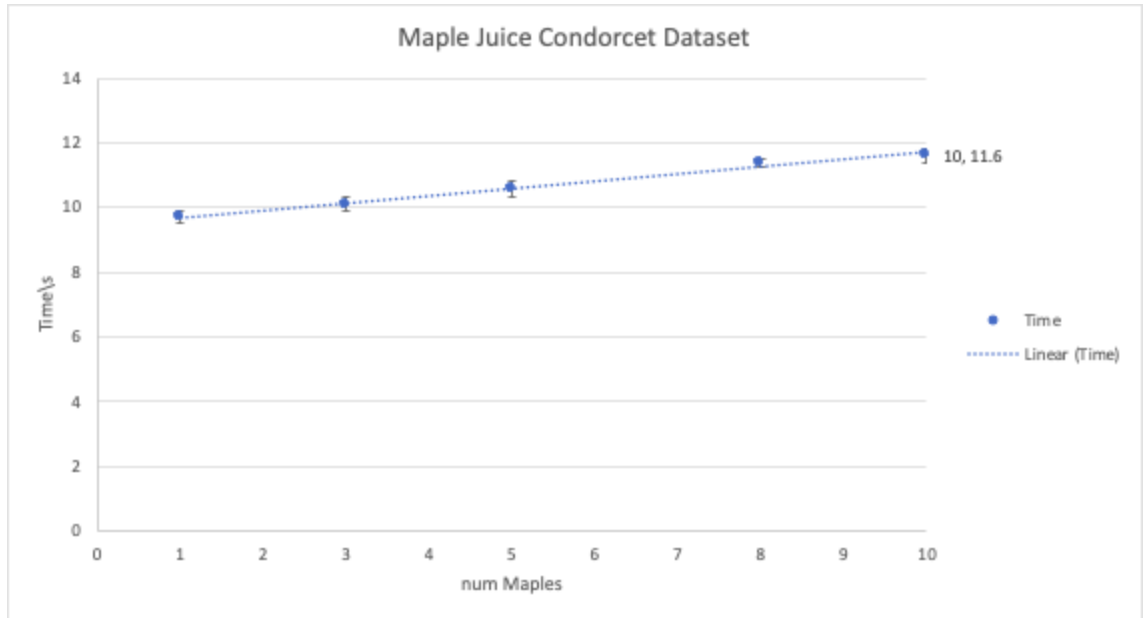
Graphs

1) Apartment Dataset (Maple Juice to find apartments with greater than 6 stories in Champaign County)



As the number of maple tasks increases, the average time to arrive at the solution also increases. This is observed because as the number of tasks increases, although the unit of work is reduced by a small amount (shards become smaller), there is a larger overhead on uploading and downloading more file shards. The time for num_maples 8 and 10 are almost the same because we have placed a cap on the number of file shards we are creating (8). Overall we expect the time to reduce if we have much bigger files and more machines to work with.

2) Condorcet Dataset (Maple Juice from HW1)



We see a similar trend here like the apartment data set except with slightly higher times. This is because here we have 2 sets of maplejuice tasks to produce our final result.