Distributed Systems: Lab Assignment 2

- In a client-server system, consider a set of N client nodes + one designated Master node
- Allocate some integer data to each of the client nodes
- Example: Node 1 contains the integers 1-100, node 2 contains the integers 101-200, node 3 contains 201-300 and so on
- The Master node should keep track of what data exists at each of the client nodes
  - This is essentially the index at the Master node

Distributed Systems: Lab Assignment 2

PART 1

- Now generate M queries based on a random distribution
  - Each query concerns the search for a single integer and the result of the query is the node ID at which that integer is stored
- The query goes to the Master node and then the Master node re-directs the query to the correct node based on its index
- The Master Node should also keep track of load at each of the client nodes
  - Load = number of queries pertaining to a given client node
  - Obviously, there could be more complex definitions of load, but the above definition is for simplicity reasons
- Output the load at each of the client nodes after all the M queries have been executed by the system

Distributed Systems: Lab Assignment 2

PART 2

- Now repeat the same, but this time, replace the random query distribution by a Zipf distribution
- The Zipf distribution essentially relates to skew
- When you use the Zipf distribution, a large percentage of the queries will get directed to some of the client nodes (hot nodes), while other nodes will receive relatively fewer queries (cold nodes)
- Output the load at each of the client nodes after all of the M queries generated by the Zipf distribution have been executed by the system

Distributed Systems: Lab Assignment 2

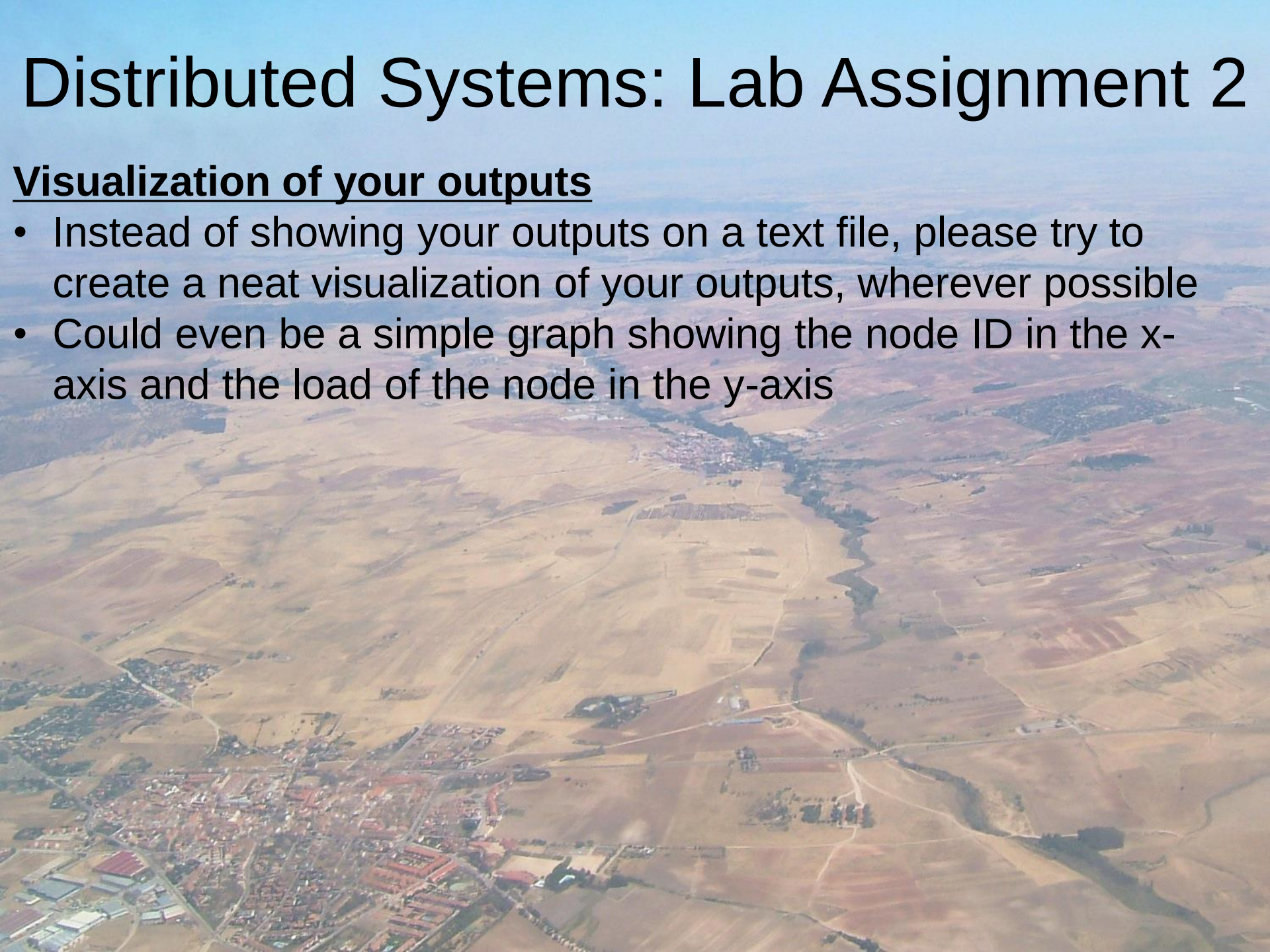
PARTs 3 and 4

- Now distribute the data to each node as before
- But this time, there should be no Master node
- Each node will store the index that was previously being stored only by the Master node
  - In essence, this is the P2P architecture
- Now the query could randomly come to any of the nodes in the system
- Output the path taken by the queries
- As before, queries would be based on a random distribution and then Zipf distribution
  - Part 3: Random distribution for query generation
  - Part 4: Zipf distribution for query generation
- More details concerning the assignment will be discussed during your lecture class

Distributed Systems: Lab Assignment 2

Visualization of your outputs

- Instead of showing your outputs on a text file, please try to create a neat visualization of your outputs, wherever possible
- Could even be a simple graph showing the node ID in the x-axis and the load of the node in the y-axis



Distributed Systems: Lab Assignment 2

Administration details and logistics

- This assignment will contribute to 10% of your course grade
- Submission deadline: April 19, 2018, 3 pm IST **(this is a hard deadline)**
- **Late submissions will incur a penalty of 3 marks, unless there are genuine excuses for late submissions and/or extenuating circumstances**
- You are allowed to use any programming language + use MPI for inter-process communication
- As usual, this assignment will be done by your lab group, not individually
- If you need any clarifications, please discuss with your TA or you can contact me