This document indicates scope for 43 projects. These 43 projects are grouped into three main classes: DHTs, File Systems, and Algorithms.

**DHTs**

One of the good projects in distributed systems is to implement various kinds of DHTs. In all of these projects, the goal is to implement the DHT as per the original design and test the system at scale. To this end, the system should be able to store millions of objects over thousands of machines. The search method and a possible application is what is studied for performance and scalability in addition to processing the join/leave of machines. In most cases, the entire system design is available as part of existing papers.

On top of the baseline system, the project should attempt to add at least one feature that is not part of the baseline system. For instance, one can think of replication, fault tolerance, consistency, heterogeneity, and the like.

The project essentially is therefore a choice of an existing system and a choice of an enhancement. This allows for multiple projects to be supported from the following list.

Systems:

1. DynamoDB
2. Chord
3. CAN
4. Tapestry
5. Napster
6. Gnutella

Enhancement:

1. Replication
2. Fault tolerance
3. Consistency
4. Heterogeneity
5. Your own idea here.

Project numbers are marked as follows: Projects 1a to 1e get numbers 1 to 5. Projects 2a to 2d get numbers 5 to 10, and so on. So, this is a collection of 25 projects.

**File Systems**

Another set of projects that are a good way to understand distributed systems in practice is file systems. These offer scope for setting up a basic system and propose additional enhancements. In that spirit, we studied two simple file systems in class: GFS and Haystack. For each of these, we could think of various enhancements some of which are listed below.

1. GFS:
   1. dynamic replication
   2. exactly once semantics for RecordAppend
   3. replicas by geography
   4. multiple master nodes
   5. Your own idea here
2. Haystack
   1. Dynamic replication
   2. Link to similar photos (use some similarity measure to link from one photo to another very similar photo, either in the same collection or across collections)
   3. Your own idea here.

Like in the case of DHTs, the projects on GFS get numbers 26 to 30 and the projects on Haystack get numbers 31 to 33.

**Graph Algorithms:**

Another popular project topics in distributed systems is to study distributed algorithms on a variety of distributed computing models. Typically, graph problems are studied in this context. In the following, we propose three graph problems and five different models/platforms for implementing algorithms for these graph problems. Each choice of problem and platform create a project.

Graph Problems:

1. Connected Components
2. Minimum Spanning Tree
3. All-pairs-shortest-paths

Model/Platform

1. Erlang
2. MPI
3. Map-Reduce
4. K-Machine
5. Massively Parallel Computing (MPC) linear regime
6. Massively Parallel Computing (MPC) sub-linear regime

As earlier, these give rise to another 18 projects.

**Project 43:** Create your own project – to send me a description for scoping

**Reference Material:**

For all the projects, there is enough reference material that is easily available on the web. If you face any questions, please contact me by email or in person. I can suggest some good references.

**Programming and Testing:**

Each project will involve some programming component and extensive and scalable testing. In particular, testing should be done on large datasets/ graphs as the case may be. For graphs, one useful collection of datasets is the University of Florida Dataset.

**Platform:**

Except for specific projects that use MPI or Map-Reduce, others can be simulated on a single machine. For projects on MPI etc., we will provide the required platform once we know the groups who are executing such projects.

**Timelines and Deliverables:**

Each project group will have the following deliverables.

1. A 2-page report on the project scope, problem being solved, basic ideas, and the like to be submitted by October 15, 2024 on moodle. While you do not have to wait till then to start working on your project, this extended time is provided keeping in mind that HW4 is just out. This submission counts as Homework 5.
2. A mid-evaluation of the project based on another 2-page submission will be due early November 2024. Date will be announced soon.
3. Final project evaluation will be an in-person assessment with each group getting a 15 minute slot. The assessment will require a presentation, a report to be uploaded prior to the assessment, all the software developed and plots produced, etc. to be uploaded prior to the assessment. The date/slots for the assessment will be announced closer to the date of the final examinations. These slots will be just after the week of the exams.