M.Sc. in IT – Field of Specialization Enterprise Application Development
2025 – Semester 01
SE5090 – Distributed Computing

Assignment

Objective	:	The objective of this assignment is to demonstrate the knowledge and the expertise of distributed design patterns and algorithms

Percentage of overall	:	40%
Duration	:	08 Weeks *(Take home assignment)*
Assignment Type	:	Individual
Deadline	:	19th April 2025
Viva Date	:	27th April 2025
Assignment created by	:	Lasitha Petthawadu

Assignment Submissions

- The assignment report needs to be submitted as a softcopy to moodle.
  - Report should include
    - The way you got the usecase from your index number.
    - Outline of approach.
    - Solution design
    - Sidecar implementation details.
    - Sequence diagram of communication flow.
    - Failure scenarios and fault tolerance.
- Students need to attend the viva with a demonstrable solution.

Assignment Task:

Use case

In a distributed environment we are planning on grouping counts of words by their starting letter in a book, pdf, or large article. there can be any number of distributed nodes.

There can be 1 coordinator node. multiple proposer nodes,acceptor nodes and one learner node. The coordinator node will understand how many proposers are available and how many acceptors and learner nodes are available.

The coordinator will broadcast this information to all current participants of the cluster.
Then the coordinator node will assign multiple letter ranges to proposer nodes. eg(a-c for one node d-g to another likewise). In your solution report you should outline the algorithm used to decide how the characters are divided.

Next the coordinator will read the document line by line and then multicast it to all proposers. Proposers will count how many words there are based on the range given to them. You need to document the algorithm using necessary diagrams to outline the logic executed in the properser. Next the counted result is passed to 2 or more acceptors and the acceptor nodes validates whether it's true, and passes the result to the learner node. The learner node will count the final result and will give it as below.

Starting letter	Count	Words
A	5	apple,aid,avid
B		bat,ball,base,beach
C		cat,call,could,cold
..	…	…

Design of Nodes and other constraints

- All nodes should be running the same codebase.
- There should be separate distributed processes.
- Nodes can be implemented in any programming language.
- Nodes need to run in their own individual process.
- Nodes should have a side-car proxy which would handle logging and also encapsulate talking to other nodes.
- You are free to use any communication protocol, HTTP, GRPC, MQTT and can use libraries to ease out the communication implementation in the sidecar.
- Should be able to spawn new nodes anytime (up to any number) and the cluster should work accordingly.

Marks breakdown

Component	Marks (out of 100)
Solution Report	**30%**

Completeness of the implementation and fault tolerance
    - Sidecar implementation (Logging of requests for analyzing)
    - Distributed Algorithms
    - Scalability
    - Fault tolerance
    - Data Consistency
	60%

Presentation skills	**10%**

Constraints:

1. Should any plagiarism be identified during viva, the assignment will be zeroed out!

2. Vivas are held strictly duringA the date communicated, if a student is unable to attend the viva they should inform 1 week prior to the date of the viva. No rescheduling would be done after the viva or a day or two prior to the viva.

3. Helping and sharing knowledge is encouraged. If you helped any student specify the students index number and mention what sort of help you did in the report.