National University of Computer and Emerging Sciences, Lahore Campus

Course:	Parallel and Distributed Computing	Course Code:	CS-3006
Program:	BS (Computer Science)	Semester:	Spring 2023
Duration:	20 Minutes	Total Marks:	10
Paper Date:	18-May-2023	Weight	3.34%
Section:	BCS (6E-6F)	Page(s):	2
Exam:	Quiz 3	Roll No.	

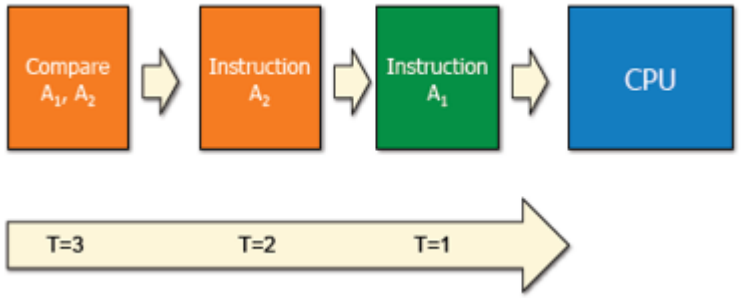
Name & Section:

Attempt all questions on the question paper. Rough sheets can be used but it should not be attached. If you think some information is missing then assume it and mention it clearly.

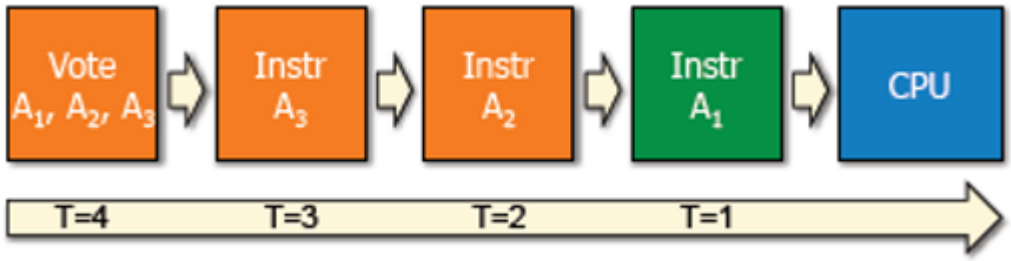
Question # 1: [4 marks, CLO # 3]

Explain time redundancy with the perspective of fault tolerance techniques.

- More time than that needed for processing an input is used
- The additional time is devoted to detect and possibly correct errors occurred during the processing



- The same instruction is executed twice, and a comparison detects the occurrence of errors during the computations
- Error Correction:

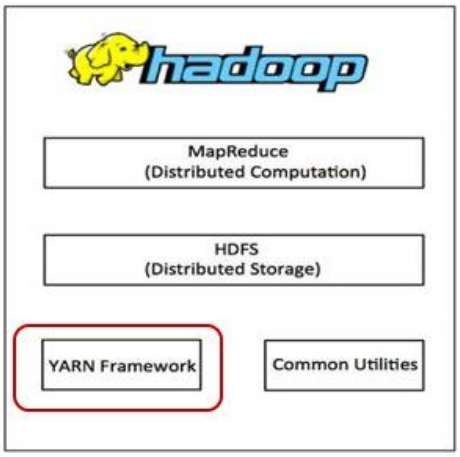


- Cost:
- Area: almost negligible
- Time: ~3x for detection, ~4x for correction

Question # 2: [3 + 3 marks, CLO # 3]

Provide a relationship between HDFS, MapReduce and Yarn Framework. For what purpose, these components are being used?

- **In Hadoop framework, HDFS is used to store data, MapReduce process the same data, and Yarn operates as the resource manager of the Hadoop framework.**



- **HDFS:**
- **HDFS is a distributed file system written in Java that is fault tolerant and scalable.**
- **HDFS is the primary distributed storage for Hadoop applications.**
- **There are two types of machines in a HDFS cluster.**
    - **NameNode is the heart of an HDFS filesystem, it maintains and manages the file system metadata. E.g; what blocks make up a file, and on which datanodes those blocks are stored.**
    - **DataNode where HDFS stores the actual data, there are usually quite a few of these.**
- **MapReduce:**
- **MapReduce is a programming model for efficient distributed computing**
    - **Processing unit of Hadoop, used by Google**
- **MapReduce fine grained Map and Reduce tasks, provides automatic re-execution on failure and locality optimizations.**
- **Yarn:**
- **YARN is the prerequisite for Enterprise Hadoop**
- **YARN provides resource management and a central platform to deliver consistent operations, security, and data governance tools across Hadoop clusters.**