National University of Computer and Emerging Sciences

Lahore Campus

Quiz 7

Total Marks: **12**
Time allowed: **5 Minutes**
Date: **April 22, 2025**
Section: **BCS-6A**

Q1: [2 Marks] What it means when we say that a region of a GFS file is *defined*?

Solution:
A GFS file region is defined if:

(a) That region is consistent (meaning data in that region is same at all the replicas)
(b) Application-level record(s) are written in its entirety in that region.

Q2: [10 Marks] Let's assume a GFS client application wants to read 100 Bytes starting from application Byte index of 64 billion (Billion = 10^9). Meaning client wants to read starting from 64 billionth Byte. But the client library needs to figure out in which chunk that Byte could be.

Tell an efficient way to figure out the chunk number, where client's required data could reside. Assume chunk numbers of the file in question starts from 0 and then monotonically increase (1,2, etc.)

Solution:
Assuming GFS uses 64 MB chunk sizes where M=106

First possible chunk where data could be = [64 * 10^9 /(64 * 10^6)] - 1 = 1000 -1 = 999
(We did -1 because question says chunk numbers start from 0.)

So if file had no duplicates / paddings in 999 chunks, then client's required data will start in 1000th chunk whose ID will be 999. But if there were any duplicates or paddings, client library will start reading from 1000th chunk onwards and looking at the information embedded in the application-level records to find out the required index.

So client's required data can be anywhere in chunk ID 999 or greater.