

Distributed File Systems

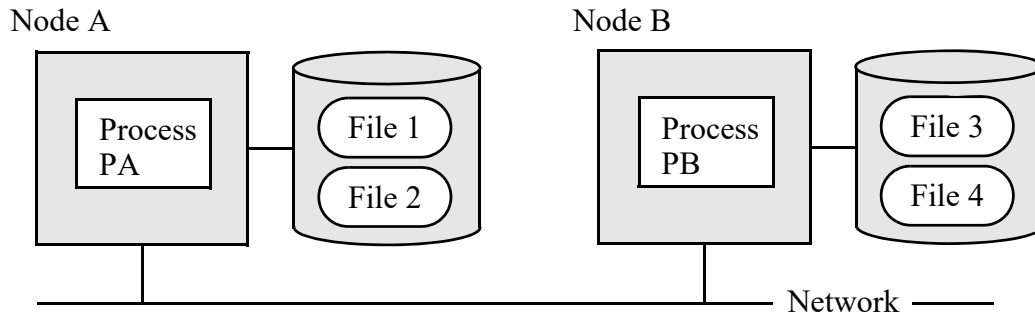
Exercise 1: File System Structures

- Given:
 - Four groups of files:
 - "database":
db.exe (1.4 MB), data.dat (300 KB)
 - "prog":
cc.exe (500 KB), p1.c (30 KB), p2.c (60 KB)
 - "inet":
mail.exe (400 KB), ftp.exe (250 KB), www.exe (150 KB)
 - "texts":
word.exe (750 KB), letter.doc (220 KB)
 - A file tree:
 - Main directory "users"
with subdirectories "user1" and "user2" for the files of two users
 - Three (very small) file server nodes in the network:
 - node_A: disk size 1 MB
 - node_B: disk size 1.5 MB
 - node_C: disk size 2 MB
 - A client node (i.e. a fourth node in the network) from where users access the files
- Do the following:
 - Place the files on the nodes
→ Specify which file shall be stored on which server node.
→ Try to place all files of a group on the same node.
 - Sketch the file trees
→ One diagram for each of the three techniques, as seen in the lecture.
 - Let user1 (who is working on the client node) access the file "p1.c"
→ For each of the three techniques: State the file path and name
and (if necessary) the additional operations.

p.t.o.

Exercise 2: Replication

- Given: A configuration of network nodes and disks with files



- The number of file accesses per minute is given by the following table:

File accesses per minute	... to File 1	... to File 2	... to File 3	... to File 4
Reads from PA ...	20	5	4	3
Reads from PB ...	8	30	1	4
Writes from PA ...	2	3	2	18
Writes from PB ...	2	1	15	2

- Assume that a local file access needs L milliseconds, a remote access R milliseconds. L and R are constants.
- Do the following:
 - Calculate the "total costs" K per minute, K being the sum of the durations of all file accesses in a minute:
 - Set up a formula that is based on a.) the symbols L and R and b.) the concrete number of accesses, as given by the table.
 - Use the formula to calculate the numerical costs K for $L=15$ ms and $R=40$ ms.
 - Assume now that all files are replicated, i.e. each of the two disks stores copies of all four files. Again, set up a formula for the total costs and calculate its value for $L=15$ ms and $R=40$ ms. Have in mind that a read operation needs only to access the local disk but that a write operation must access both local and remote disk in order to keep the copies consistent. Local and remote accesses can be done in parallel, i.e. their total duration is the maximum of L and R .
 - Does it pay off (for $L=15$ ms and $R=40$ ms) to replicate all four files? If so, why? If no, which file should better not be replicated?
 - Assume now that you only have the alternative to replicate *all four* files or *none* of them. From which ratio R/L on does it pay off to replicate all files (based on the given table and the formulae from above)?