UiO / IFI / PUT
Spring 2025
Questions on Replication

1. The replication model differs from device-side caching in a number of key ways. Mention two of them and explain their meaning.

2. Give some examples of applications that illustrate the use of replication.

3. Compare the approaches of “Remote Data Access” and “Master Replication”.

4. In a master replication model, what component is responsible for detecting when two devices produce conflicting updates? Can all such conflicts be solved automatically?

5. How do you compare the several solutions for accessing data (remote access, device-caching, master replication, etc..) with respect do devices, data, reads and updates?

6. In a replicated system, what is consistency?

7. Regarding consistency, what is the notion of: weak/best effort/eventual/causal/bounded /VFC/session?

8. As introduced by Bayou, what is a session? Does a session corresponds to an atomic transaction?

9. Consider the session guarantee **Monotonic Writes** (or any other of those presented). At what instants (Write or Read operation) is the session state: i) updated, and ii) checked. Justify your answer.

10. Take into account the notion of best effort consistency. Indicate four situations in which, even with reliable delivery, replicas will not converge.

11. Consider the various session guarantees that are possible under a distributed system with a database replicated on multiple servers (e.g. Bayou). Such system is used in a scenario of a company that sells several models of tires and whose sellers, each one using a mobile device, can connect to any server at any moment:
    - there is a table in the database where each record is related to a tire model and its characteristics;
    - each record corresponds to a specific tire model;
    - each field of each record in the table corresponds to a product characteristic (e.g. price, existing units, etc.);
    - each one of the vendors sells only one tire model which is distinct from the others;

- the database is modified by the various vendors, such that it reflects the number of units already sold, i.e. each seller decrements/increments the number of existing units taking into account the units sold/bought of the tire model for which he is responsible.

Which **session guarantee** should be applied in order to ensure that each seller always knows the exact value of existing units of the model he sells? Indicate the less demanding guarantee and justify your answer.

12. Consider the following sentence and say if you agree (or not) and why: "*Eventually consistent systems make no guarantees whatsoever about the freshness of data returned by a read operation*". Give an example that illustrates your answer.

13. Consider the following sentence and say if you agree: *the session guarantee "Read your Writes" guarantees that, in every copy of the database, Writes made during the session are ordered after any Writes whose effects were seen by previous Reads in the session*. Justify your answer.

14. Consider a user editing a replicated file in his laptop. When a file is saved, it is stored at a server; then, the system is responsible to propagate the new file version to other servers. The system should ensure that if the user saves version N of the file and later saves version N+1, then version N+1 will replace version N at all servers. Thus, the following situation must be prevented: version N is written to some server and version N+1 is written to a different server, and both versions get propagated such that version N is applied after N+1. What is the most adequate session guarantee?

15. One of the possible techniques to deal with network connection loss (or weak connectivity) is based on replication and it is designated staging. Taking into account this technique, how much does it depend on the correctness of the working set that is replicated? How does this technique deal with the case in which the user tries to access a file that is not staged? Describe what security aspects are considered and how they are dealt with. (Draw a figure that shows how the system works to better explain your answer.)

16. Consider the session guarantee *Read Your Writes*. Does this session guarantee affect the users outside the session under consideration?

17. Consider the following scenario:

    A user's appointment calendar is stored online in a replicated database where it can be updated by both the user and automatic meeting schedulers. The user's calendar program periodically refreshes its display by reading all of today's calendar appointments from the database. If it accesses servers with inconsistent copies of the database, recently added (or deleted) meetings may appear to come and go unless a session guarantee prevents that from happening.

    What is the weakest session guarantee that should be used?

18. Consider the four session guarantees presented from Bayou. Explain each one and provide an example.

19. To ensure that the Bayou guarantees are meet, what can the servers do?

20. In the Bayou system do you agree with the following? “The term “database” is not meant to imply any particular data model or organization, nor are the techniques specific to any data model”. Justify your answer.

21. Bayou assumes that the underlying replicated system provides eventual consistency and thus includes mechanisms to ensure two properties. What properties are these?

22. The implementations of the session guarantees require only minor cooperation from the servers that process Read and Write operations. Specifically, what information must a server be willing to provide?

23. For each session, the session manager maintains two sets of WIDs; which sets are these?

24. Regarding the support for “Read your Writes”, there are two basic steps. Which are these two? Where could these check s be done?

25. Regarding the support for “Monotonic Reads”, there are two basic steps. Which are these two?

26. Regarding the support for “Writes Follows Reads”, there are two basic steps. Which are these two?

27. Regarding the support for “Monotonic Writes”, there are two basic steps. Which are these two?

28. What is a version vector?

29. What version vectors is maintained in each server?

30. How can a set of WIDs be replaced by version vectors?

31. Describe how a client can find a suitable server?