

1 Questions

1. (a) One should use multi-leader replication in cases where
 - i. we have multiple datacenters
 - ii. we have an application that needs to continue working even in cases where its connection to the internet has dropped.
 - iii. we have a system in which multiple users need to be able to collaborate and edit in real time.

This type of replication allows for better performance, since writes can be processed locally, and then replicated asynchronously to other datacenters. Additionally, this makes for a higher tolerance of outages since each datacenter can continue operating normally even when one datacenter has failed. A replication like this will also make it so that the system tolerates network problems better, since a temporary network outage would not prevent writes from being processed.

One could use leader-based replication in single datacenters or in general in cases where for instance it is not a problem for the system that all writes must go through this one leader, and that the system won't often experience network interruption **kleppmann'2017**.

- (b) One should use log shipping as a replication means instead of just replicating the SQL statements because sending SQL statements can be unpredictable in terms of non-deterministic functions, where functions calling for a time now or a randomly generated number will be different on the different replicas. This method can also cause unpredictable scenarios in terms of auto-incrementing IDs, in addition to side effects possibly having different results on different replicas **kleppmann'2017**.