#ASLI ENGINEERING

Two Phase Commit

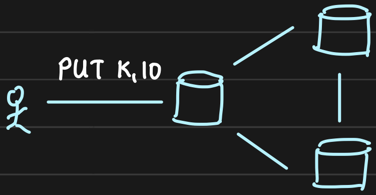
SWIPE

BY

ARPIT BHAYANI

Two Phase Commit

Say, we have a distributed database with three nodes and we want our "commit" to succeed when commit at all DBs succeed otherwise "abort"



This is a classic case of Distributed Transaction

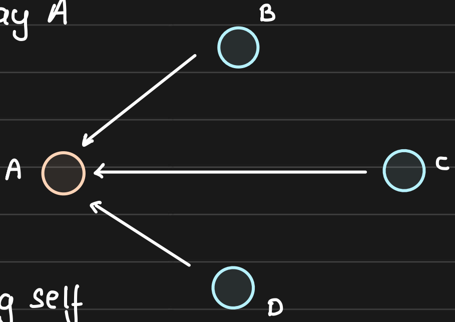
Assumption : No message loss
Process failure can happen
The graph is fully connected

No two processes can decide different values

Two Phase Commit

Say, we have N processes participating in the transaction

We choose a distinguished process, say A
↳ may be a leader / master node



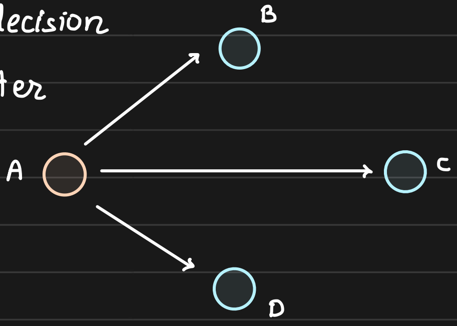
Phase 1 : All nodes send if they can commit or abort to A. (no msg = abort)

A would then gather all decisions including self
if all commit : decision commit
otherwise : decision abort

ARPIT BHAYANI

Round 2: Process A will broadcast its decision to all the nodes in the network / cluster

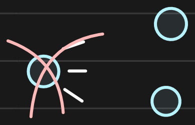
if any node, did not participate in round 1, will have to decide on the decision sent by A.



Failures

1. Co-ordinator fails before initiating phase 1

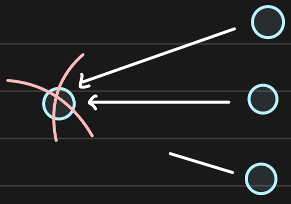
   ↳ consensus did not begin, all good



2. Co-ordinator fails after initiating phase 1

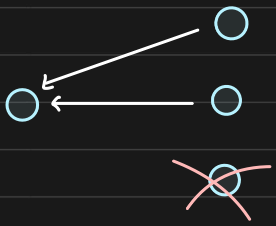
   ↳ some nodes who sent their state are blocked on co-ordinator to respond

   Perpetual starvation can be solved through timeouts + re-election



3. 2PC halts if a participant crashes before sending its preference to co-ordinator

   ↳ co-ordinator cannot proceed



4. If participant crash at phase 2

↳ co-ordinator does not know if the participant crashed before/ after applying the changes

participant does not know if it has to commit or abort (after recovering)

5. if co-ordinator and one participant die in phase 2 without other participating nodes knowing the decision

↳ Newly elected co-ordinator would not know the decision and if the crashed participating node committed/aborted

Hence, Two Phase Commit is a Blocking Protocol

↳ in case of a failure (co-ordinator / participant) no one is able to confidently recover

Complexity Analysis

Communication complexity is $2(n-1)$ and algorithm runs for 2 rounds