Two Phase Commit

Distributed transactions

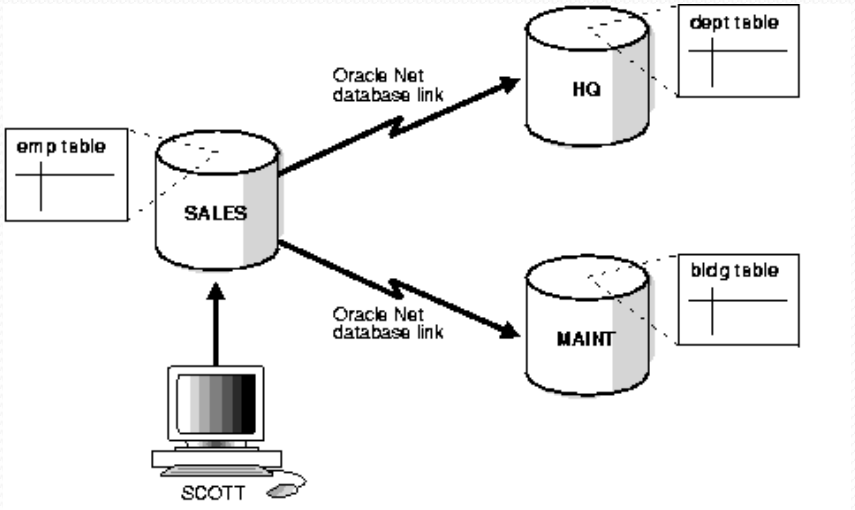
- If transaction modifies objects in multiple databases, a single "two phase commit" can commit changes in all databases, whilst retaining ACID properties.

- Distributed or remote?
  - *Note: If all statements of a transaction reference only a single remote node, then the transaction is remote, not distributed.*

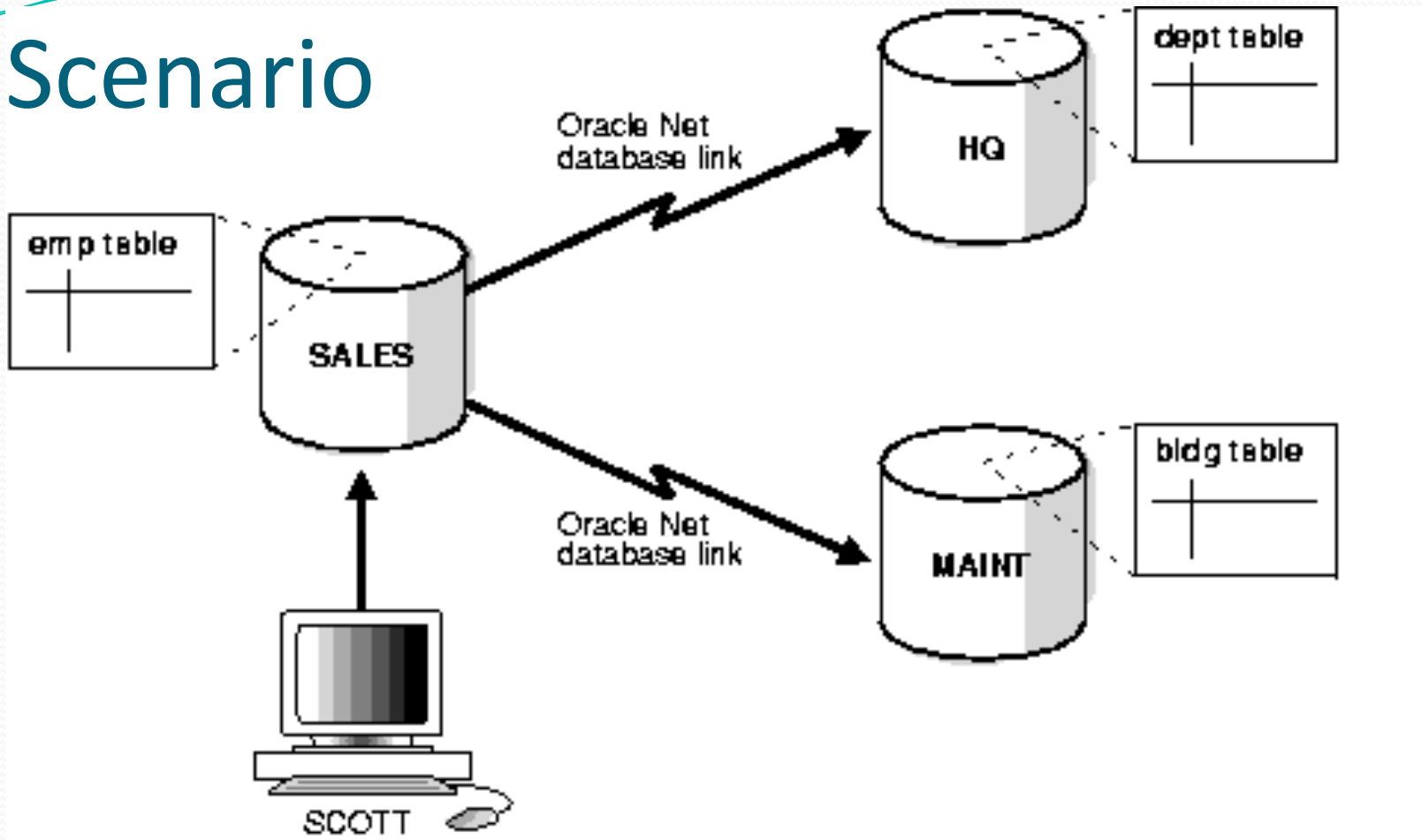
The databases involved are called nodes; the initiating node is called the 'Global Coordinator'

Example

- A **distributed transaction** updates data on two or more distinct nodes of a distributed database.
- This examples has three databases, one local, and two remote.
- The user is updating tables on all three databases using a distributed transaction.



Scenario



- User Scott updates the dept table on the remote HQ database, the local emp table and the remote bldg table on the MAINT database, in one transaction.

Example

- This distributed transaction run by scott updates
  - the remote hq database
    - dept table
  - the local sales database
    - emp table
  - the remote maint database
    - bldg table
- Then commits

```
UPDATE
scott.dept@hq.us.acme.com
   SET loc =
   'REDWOOD SHORES'
   WHERE deptno = 10;
UPDATE scott.emp
set deptno = 11
Where deptno = 10;
UPDATE
scott.bldg@maint.us.acme.com
   SET room = 1225
   WHERE room = 1163;
 COMMIT;
```

The two phases

- Prepare phase
  - The global coordinator (initiating node) asks participants to prepare (to promise to commit or rollback the transaction, even if there is a failure)
- Commit Phase
  - If all participants respond to the coordinator that they are prepared, the coordinator asks all nodes to commit the transaction.
  - If all participants cannot prepare, the coordinator asks all nodes to roll back the transaction.

Prepare Phase

- By preparing, a node:
  - **Logs the transaction locally.**
  - Places a distributed **lock** on modified tables, which prevents reads
  - **Responds with:**
    - Prepared
    - Read-only or
    - Abort

What happens next?

- The prepared nodes then wait until a COMMIT or ROLLBACK request is received from the global coordinator.
- After the nodes are prepared, the distributed transaction is said to be **in-doubt** until all changes are either committed or rolled back.

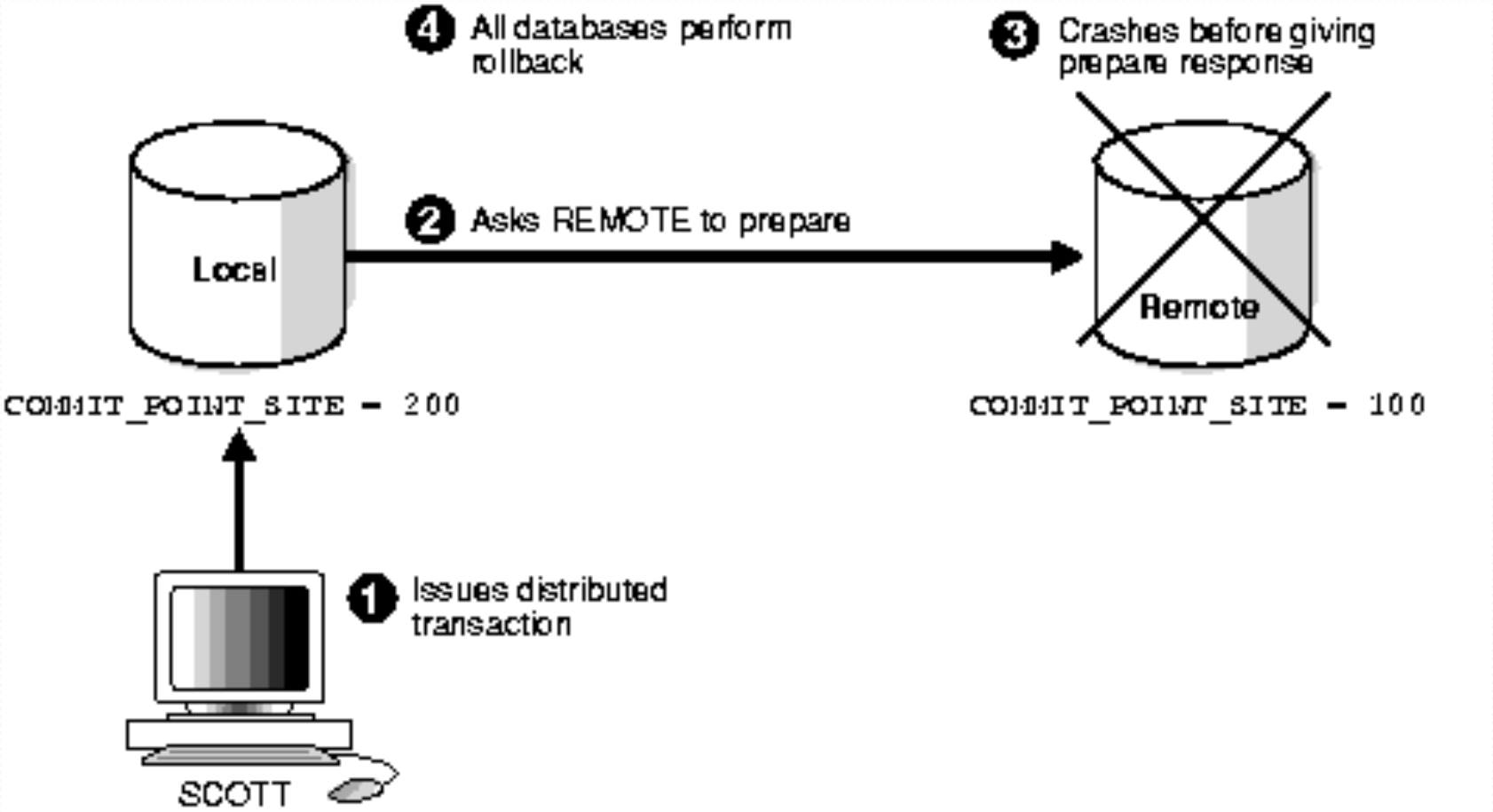
Steps in the Commit Phase

- The commit phase consists of the following steps:
  1. The global coordinator issues an order to commit.
  2. At each node, the local portion of the distributed transaction is commited and locks are released.
  3. The participating nodes notify the global coordinator that they have committed.
- When the commit phase is complete, the data on all nodes of the distributed system is **consistent**.

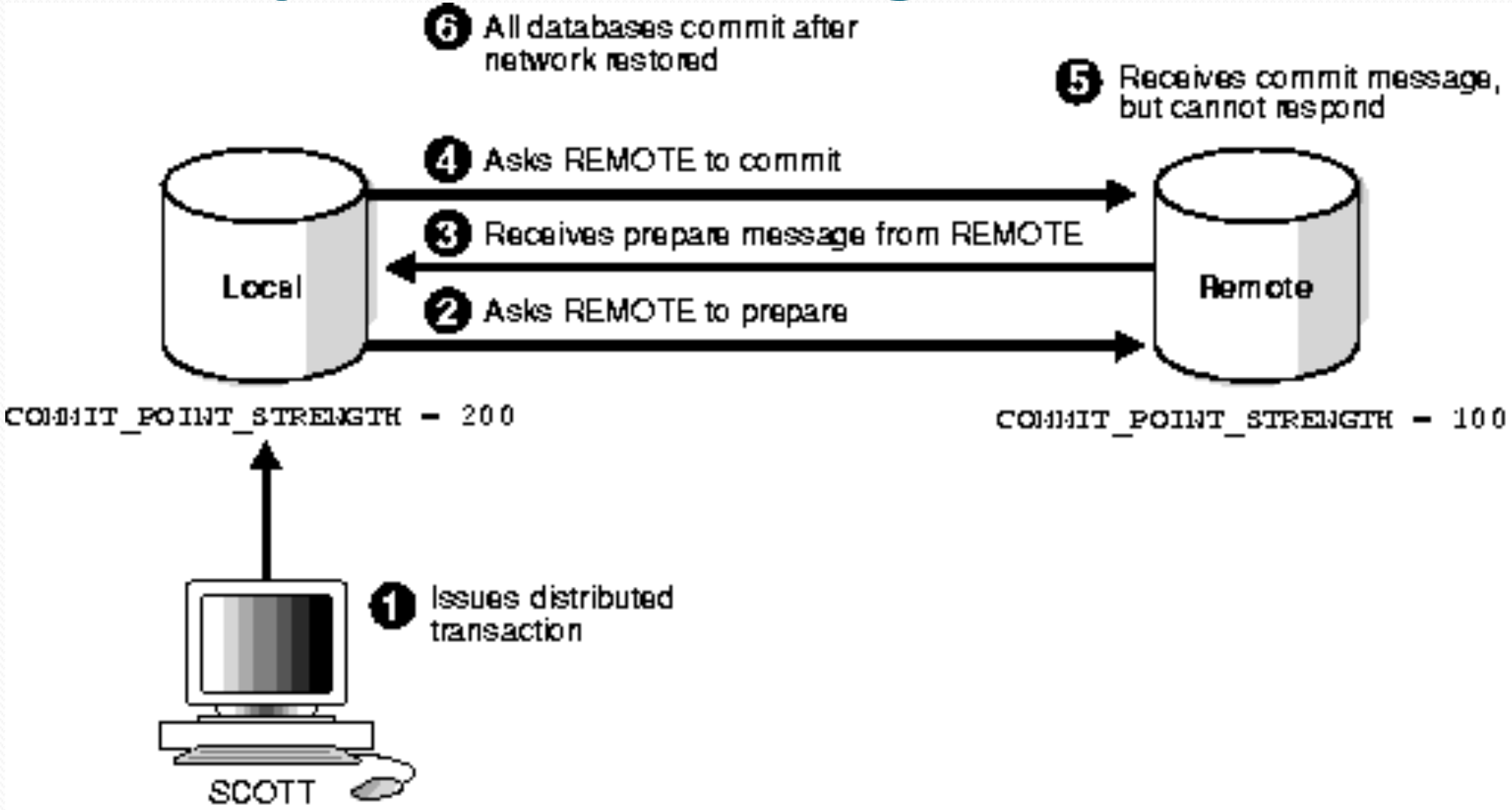
In-Doubt Transactions

- The two-phase commit mechanism ensures that all nodes either commit or perform a rollback together.
- What happens if any of the phases fails because of a system or network error?
- The transaction becomes in-doubt.

Example 1 – during PREPARE



Example 2 – during COMMIT



Trade-offs

CAP theorem

- There are three core systemic requirements that exist in a special relationship when designing and deploying distributed applications.
  - Consistency,
  - Availability and
  - Partition Tolerance

Partition Tolerance

- Partition tolerance in CAP means tolerance to a network partition.
  - i.e. when two nodes can't talk to each other.
  - A CA system guarantees strong consistency, at the cost of not being able to process requests unless all nodes are able to talk to each other.
  - An AP system is able to function during the network split, while being able to provide various forms of eventual consistency.

Eventual consistency

- Weak eventual consistency
  - This is where a write might not be consistent across the network partition, meaning that it is not possible to read them.
  - BASE (Basically Available, Soft state, Eventual consistency)
- ‘Read – your – writes’
  - Rather than being sure that all reads are consistent, this method reads from N replicated copies and if W writes (< N) agree, then that is considered to be the correct answer.