

RAFT

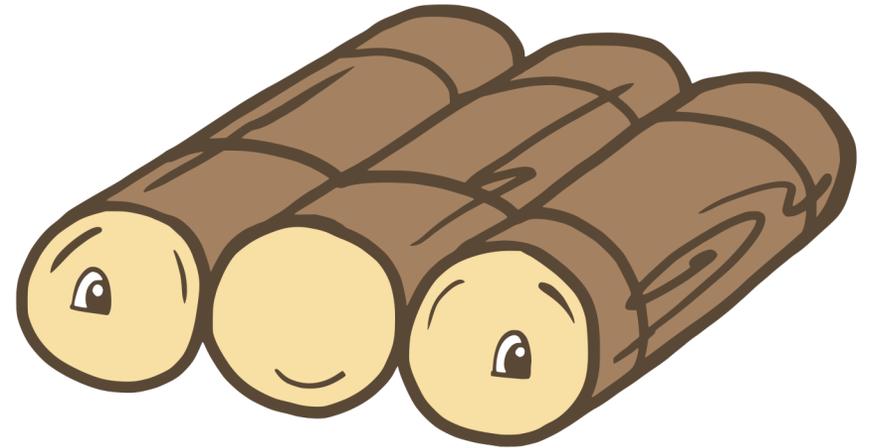
Eine Wiederholung

Majority-Consensus Algorithmen

- Werden eingesetzt in fehler-toleranten verteilten Systemen
- Um einer Variablen einen neuen Wert geben zu können, muss die Mehrheit der Server diesem neuen Wert zustimmen
- Solange die Mehrheit der Server erreichbar ist, kann das System arbeiten

Raft

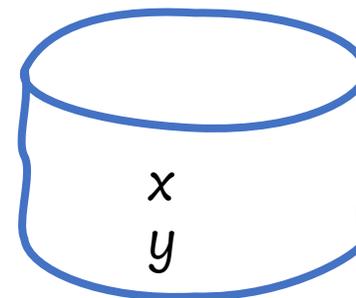
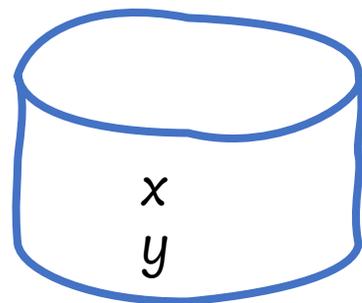
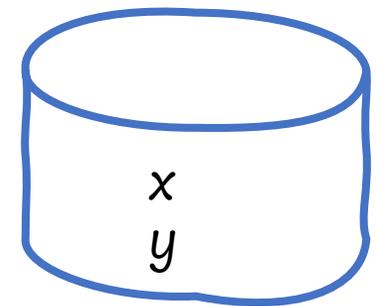
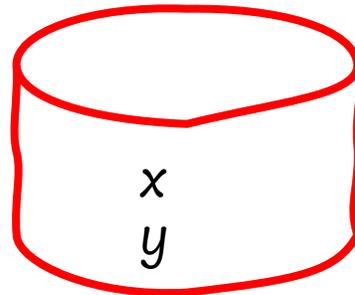
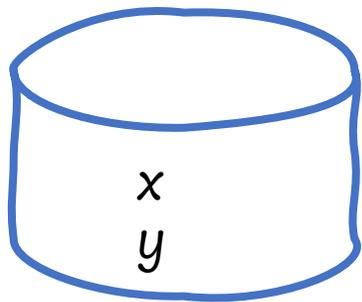
- Ein Majority-Consensus Algorithmus
- Von: Ongaro & Ousterhout (Stanford University)
- Vereinfachte Form von Paxos (Lamport)
- Unterteilung der Probleme in relativ unabhängige Komponenten
 - Wie die Stämme eines Floßes
 - Vereinfacht den Einsatz von Consensus-Algorithmen in der Praxis



Raft



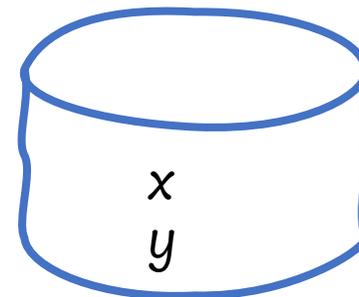
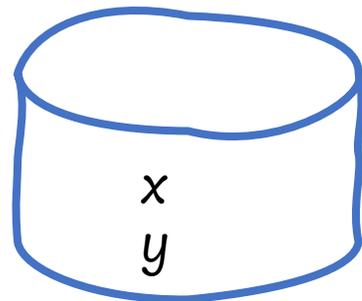
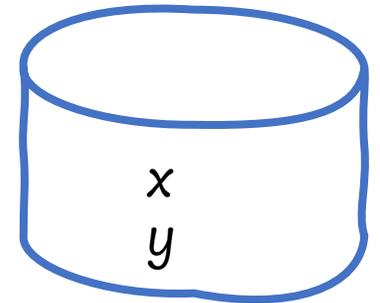
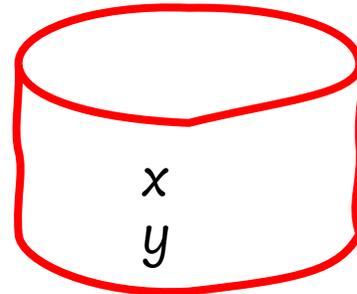
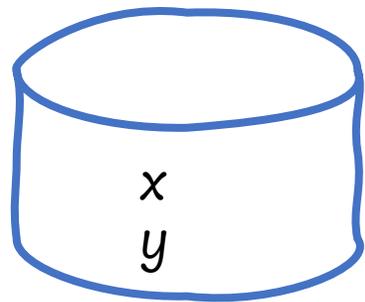
Die Stationen in diesem verteilten System replizieren die Variablen x und y .



Raft

Eine Station ist der Leader.

- Dieser Leader muss eindeutig sein
- Alle Klienten wenden sich an den Leader

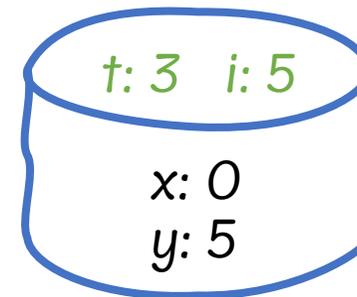
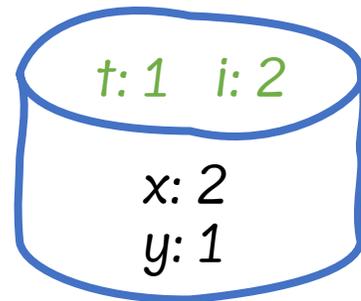
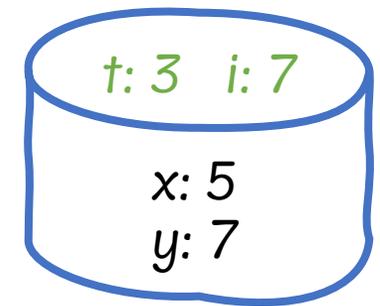
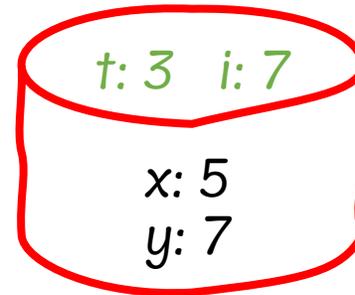
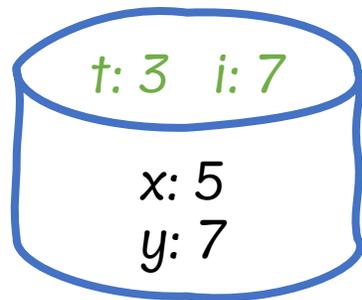


Raft

- Jede Station hält einen committeden Zustand
- Der term t („Epoche“) gibt Auskunft darüber wie aktuell der Wert der Stationen ist
- Der Index i (der Log-Index)



t : term / Epoche
 i : Index im Log

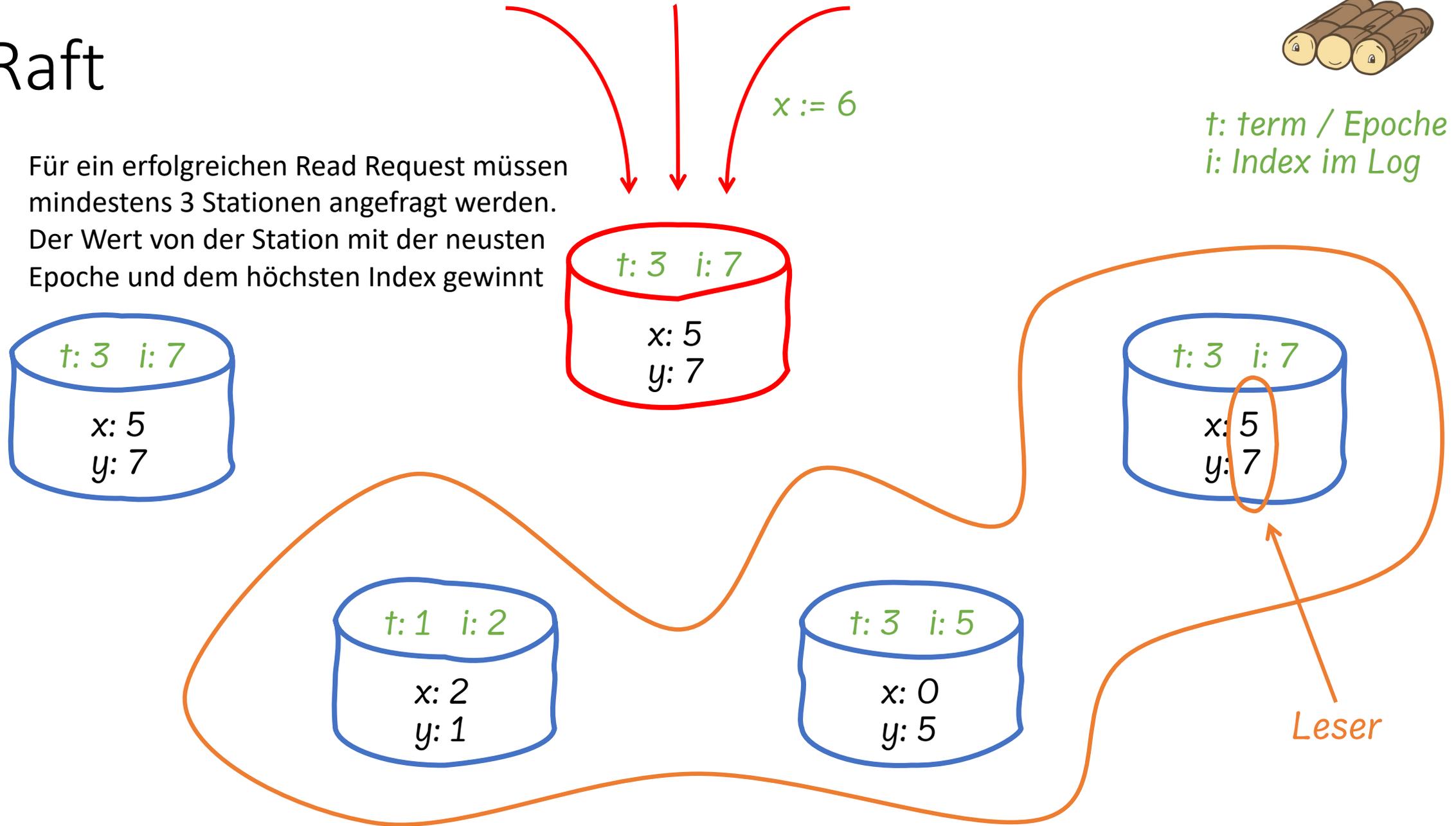


Raft

- Für ein erfolgreichen Read Request müssen mindestens 3 Stationen angefragt werden.
- Der Wert von der Station mit der neusten Epoche und dem höchsten Index gewinnt



t: term / Epoche
i: Index im Log



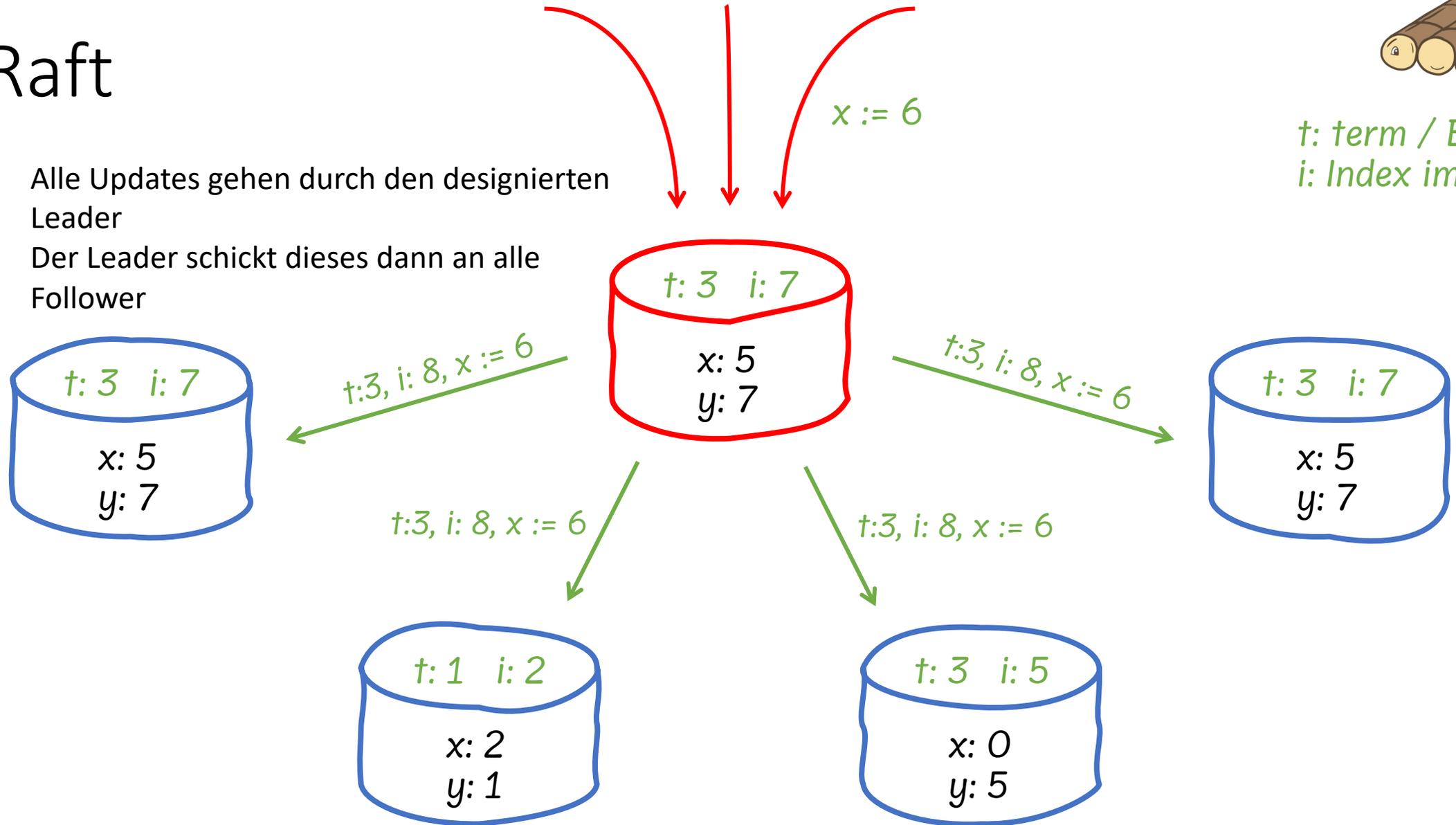
Leser

Raft



- Alle Updates gehen durch den designierten Leader
- Der Leader schickt dieses dann an alle Follower

t: term / Epoche
i: Index im Log

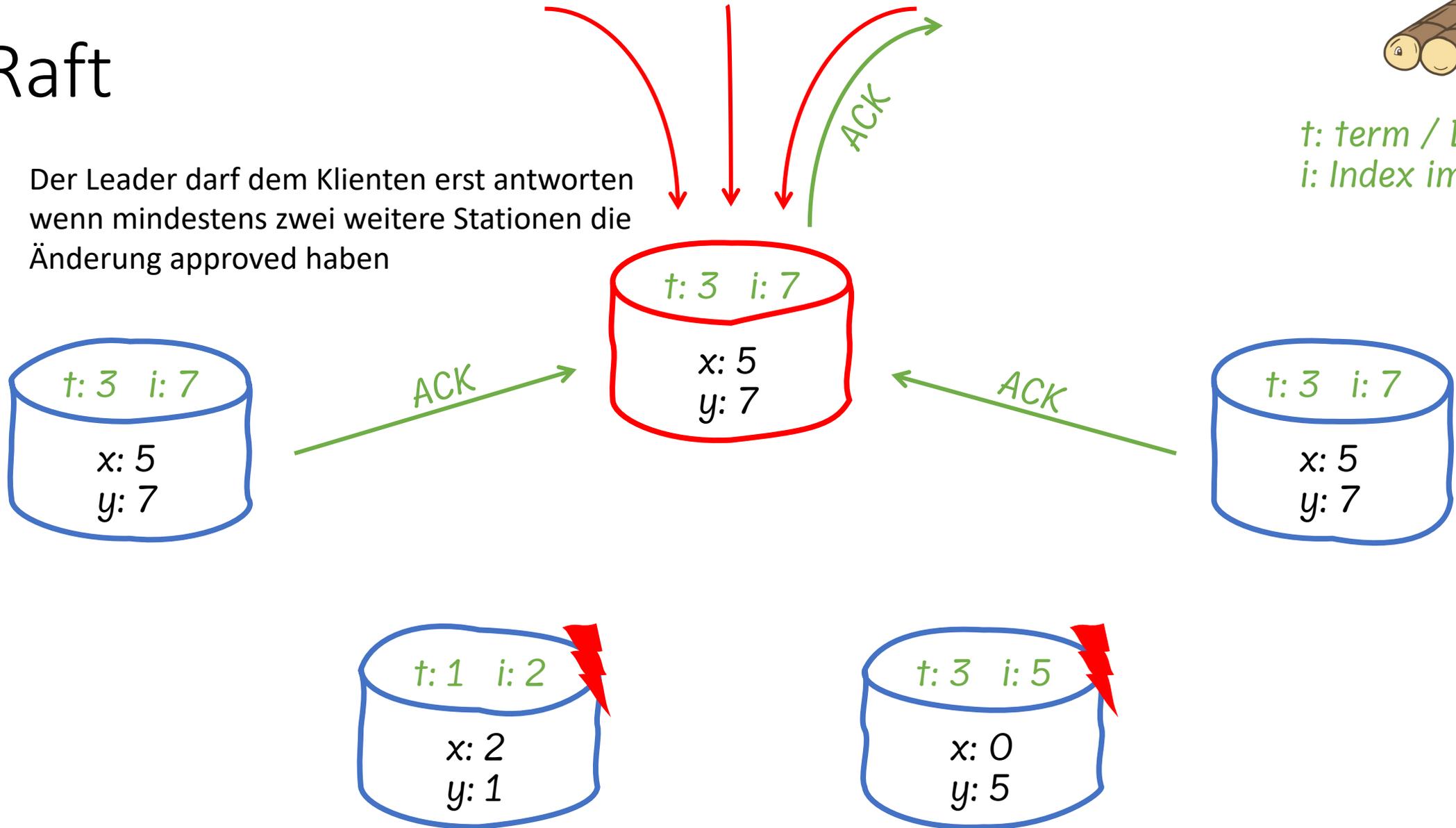


Raft



t: term / Epoche
i: Index im Log

- Der Leader darf dem Klienten erst antworten wenn mindestens zwei weitere Stationen die Änderung approved haben

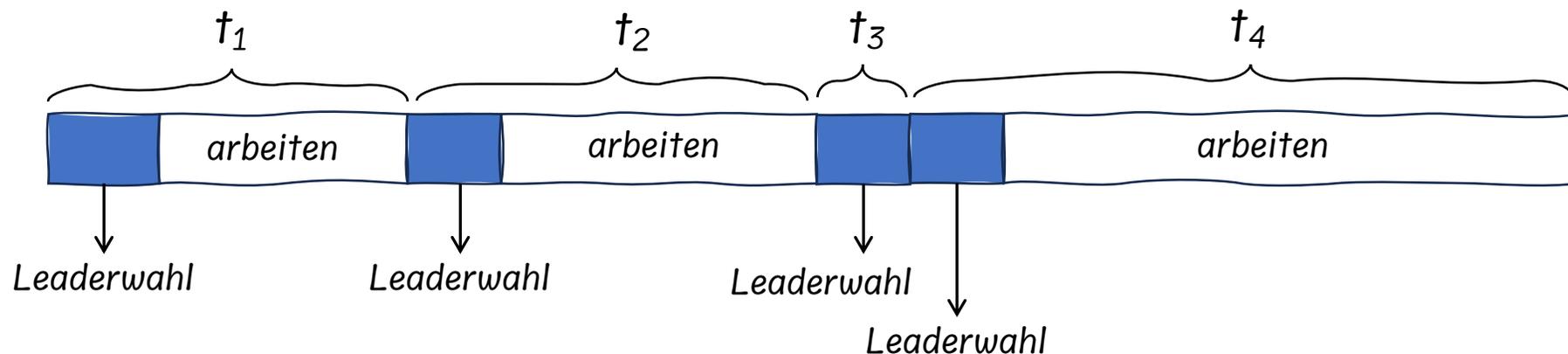


Raft



Garantien in Raft

- Es gibt nur einen Leader
- Alle Follower werden über kurz oder lang auf den aktuellen Stand gebracht
- Sollte der Leader nicht mehr erreichbar sein, beginnt eine neue Epoche (ein neuer Term) und ein neuer Leader wird gewählt
- Wenn zwei Kandidaten gleichzeitig versuchen Leader zu werden und beide keine Mehrheit bekommen, gilt die Wahl als gescheitert (wie in Epoche 3)
 - Die Wahl wird wiederholt. Die beiden Kandidaten, die verloren haben, warten eine unbestimmte Zeit (random) und versuchen es erneut

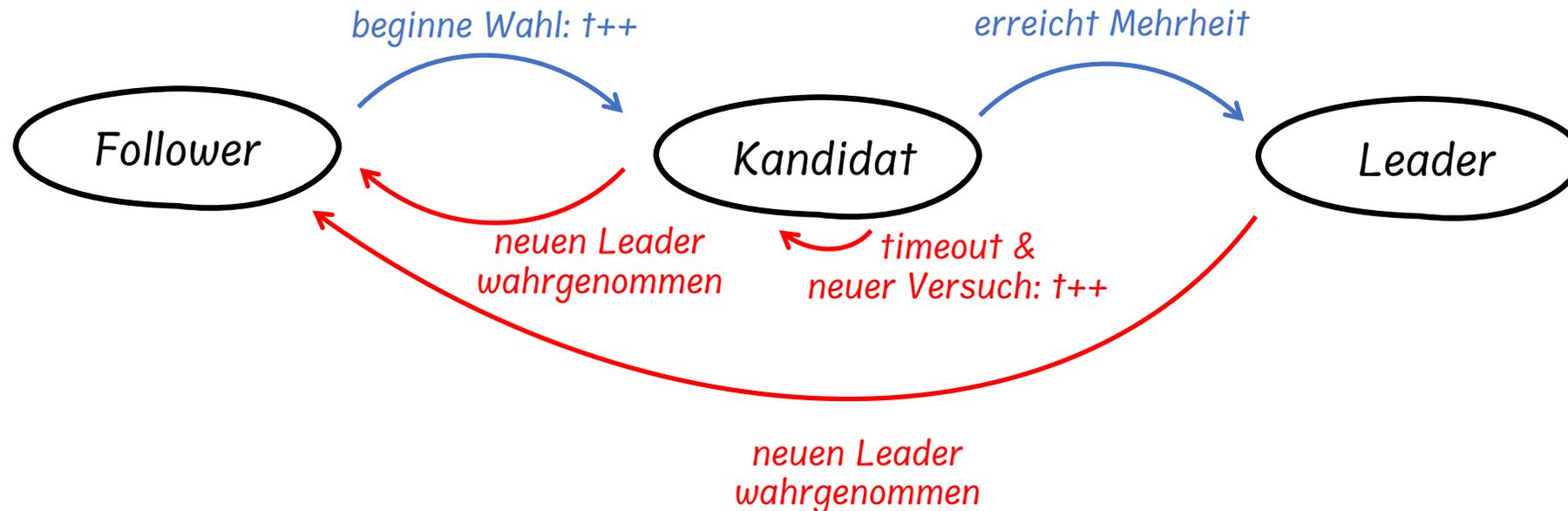


Raft



Die Leaderwahl

- Stationen sind erstmal Follower
- Sollte eine Station den Eindruck haben, der Leader ist nicht mehr da (Heartbeat kam nicht mehr), dann werden sie zum Kandidat und erhöhen die Epoche



Raft

Das Log

- Im Log wird neben dem Update auch der Index (i) sowie der Term (t) protokolliert
- Die Follower können sich nach dem Wiederaufwachen an den Leader wenden und den aktuellen Stand anfordern.
 - Die Werte von i und t zeigen dem Follower ab welchem Punkt er nicht mehr erreichbar war und welche Einträge noch fehlen.



Committede Version

Leader

i	1	2	3	4	5	6	7	8
t	1	1	1	2	3	3	3	3
	x = 3	y = 1	y = 9	x = 2	x = 0	y = 7	x = 5	x = 6

Follower

i	1	2	3	4	5
t	1	1	1	2	3
	x = 3	y = 1	y = 9	x = 2	x = 0

i	1	2	3	4	5	6	7
t	1	1	1	2	3	3	3
	x = 3	y = 1	y = 9	x = 2	x = 0	y = 7	x = 5

i	1	2	3
t	1	1	1
	x = 3	y = 1	y = 9

i	1	2	3	4	5	6	7
t	1	1	1	2	3	3	3
	x = 3	y = 1	y = 9	x = 2	x = 0	y = 7	x = 5