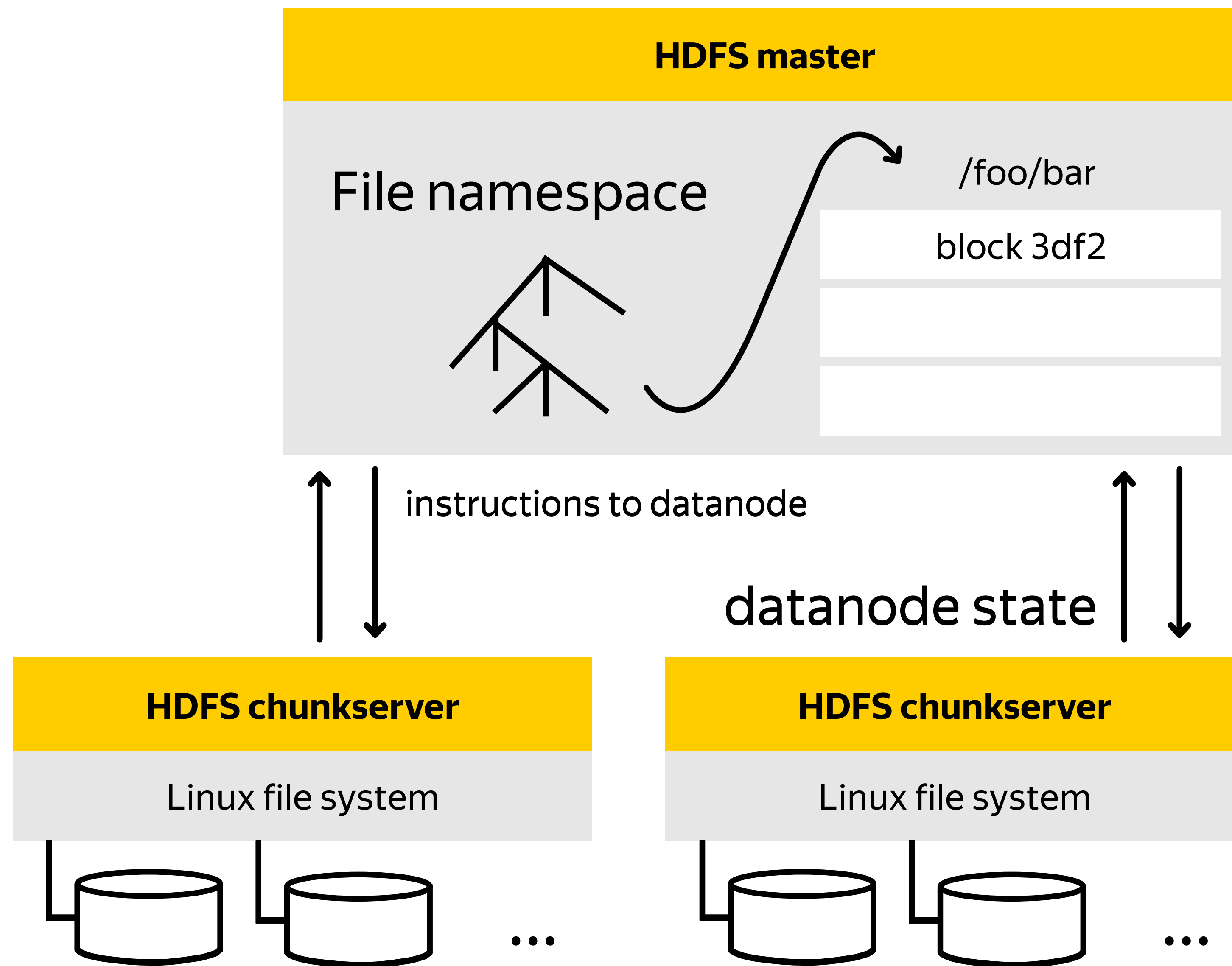
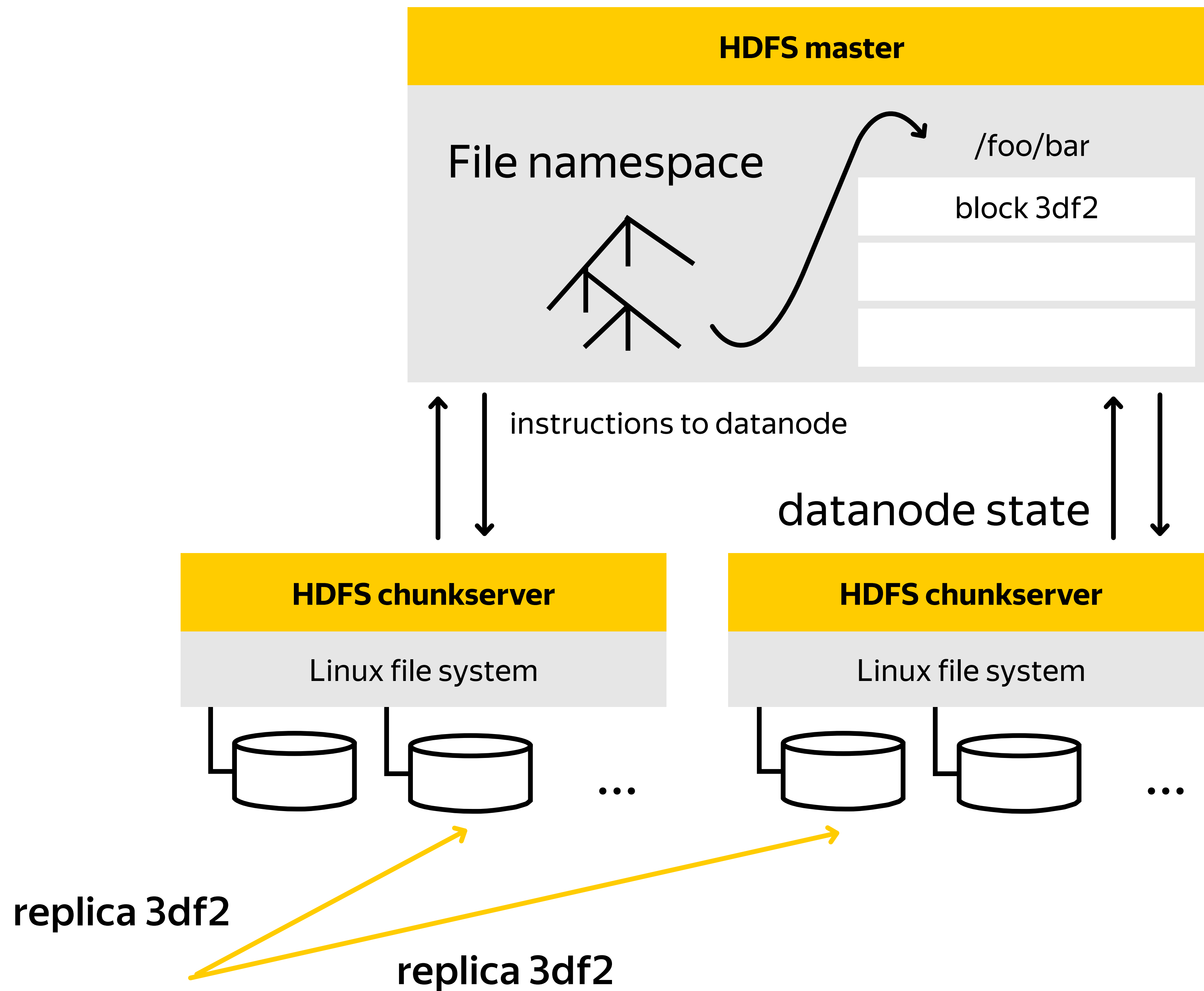


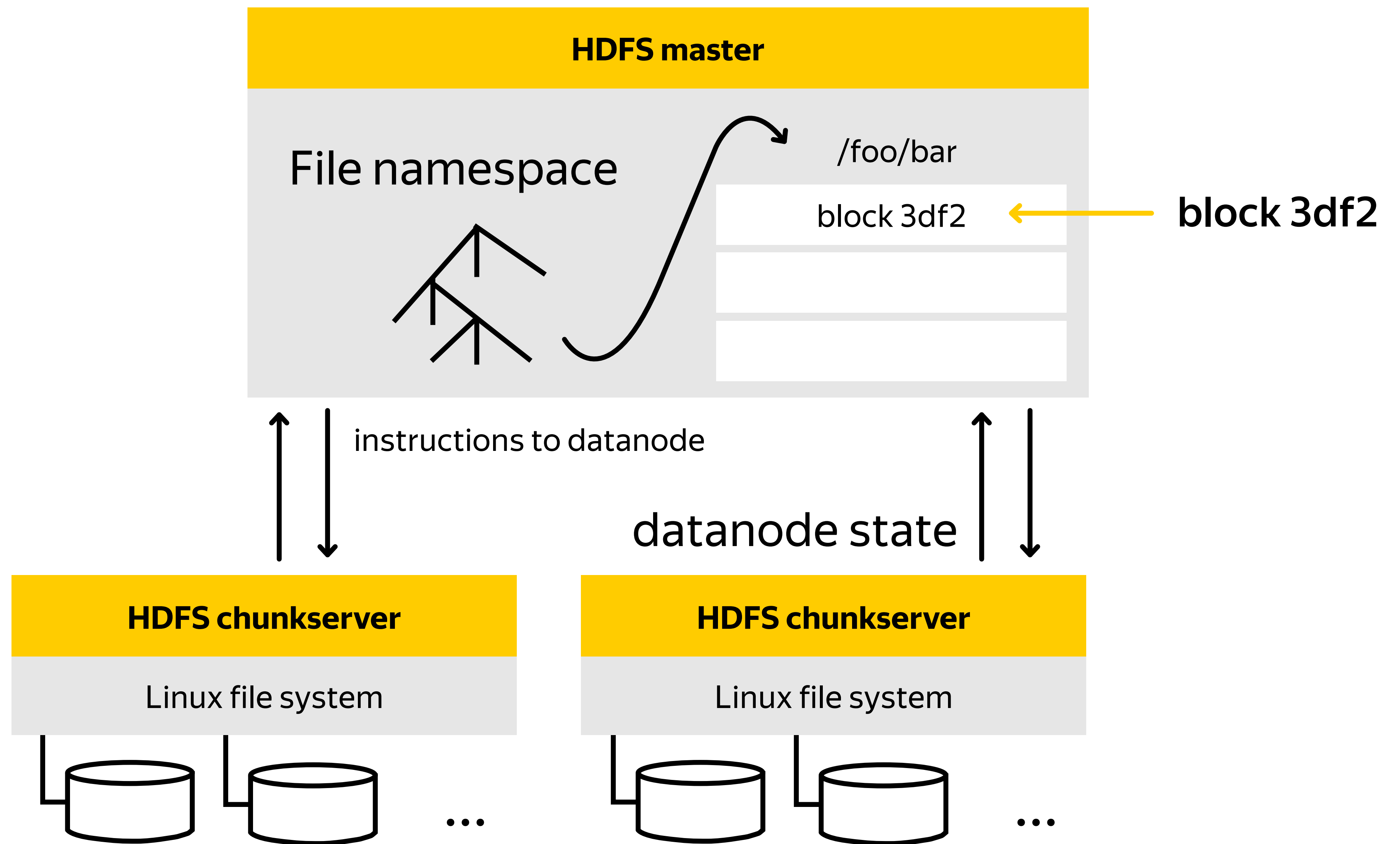
Yandex

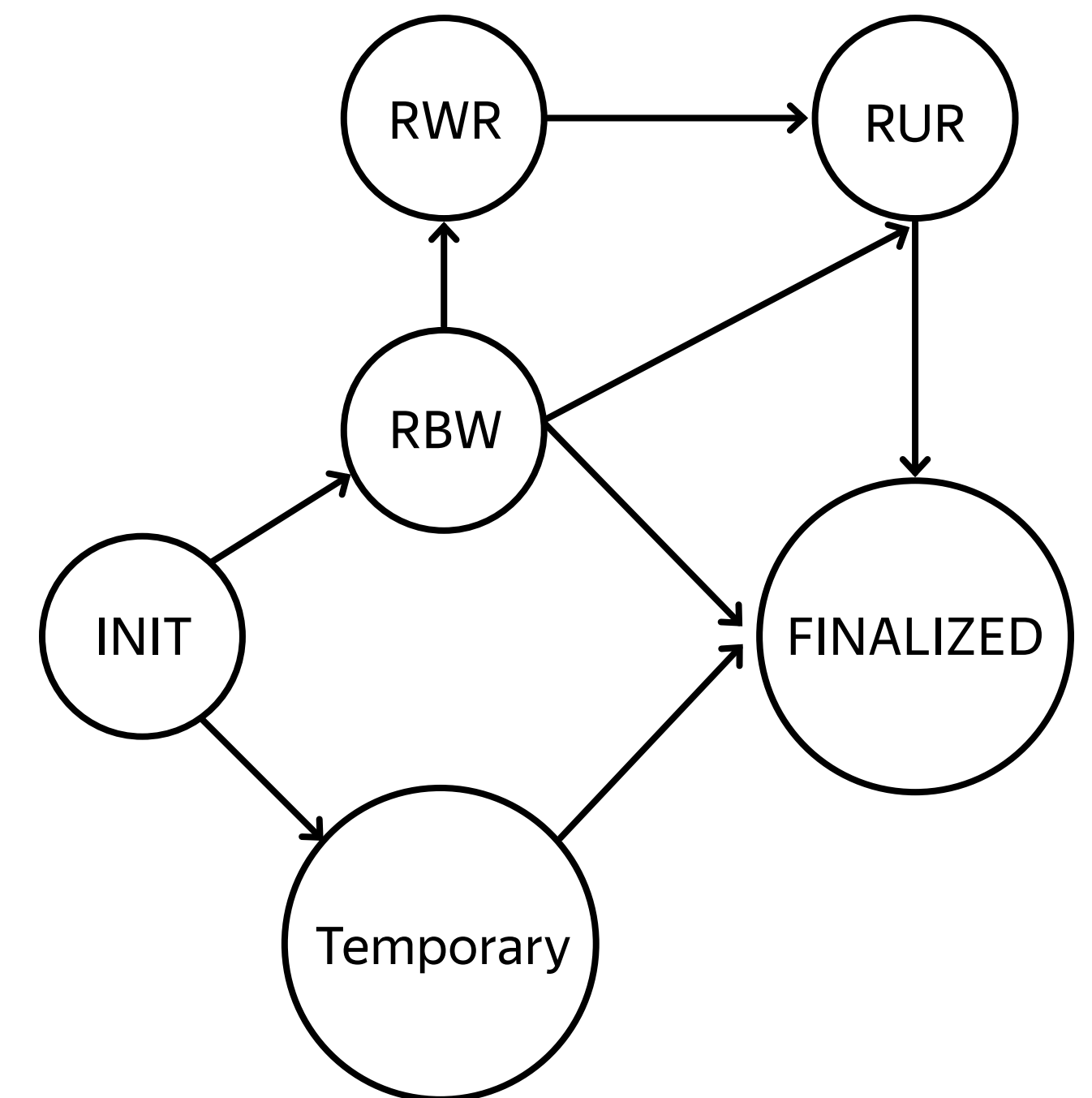
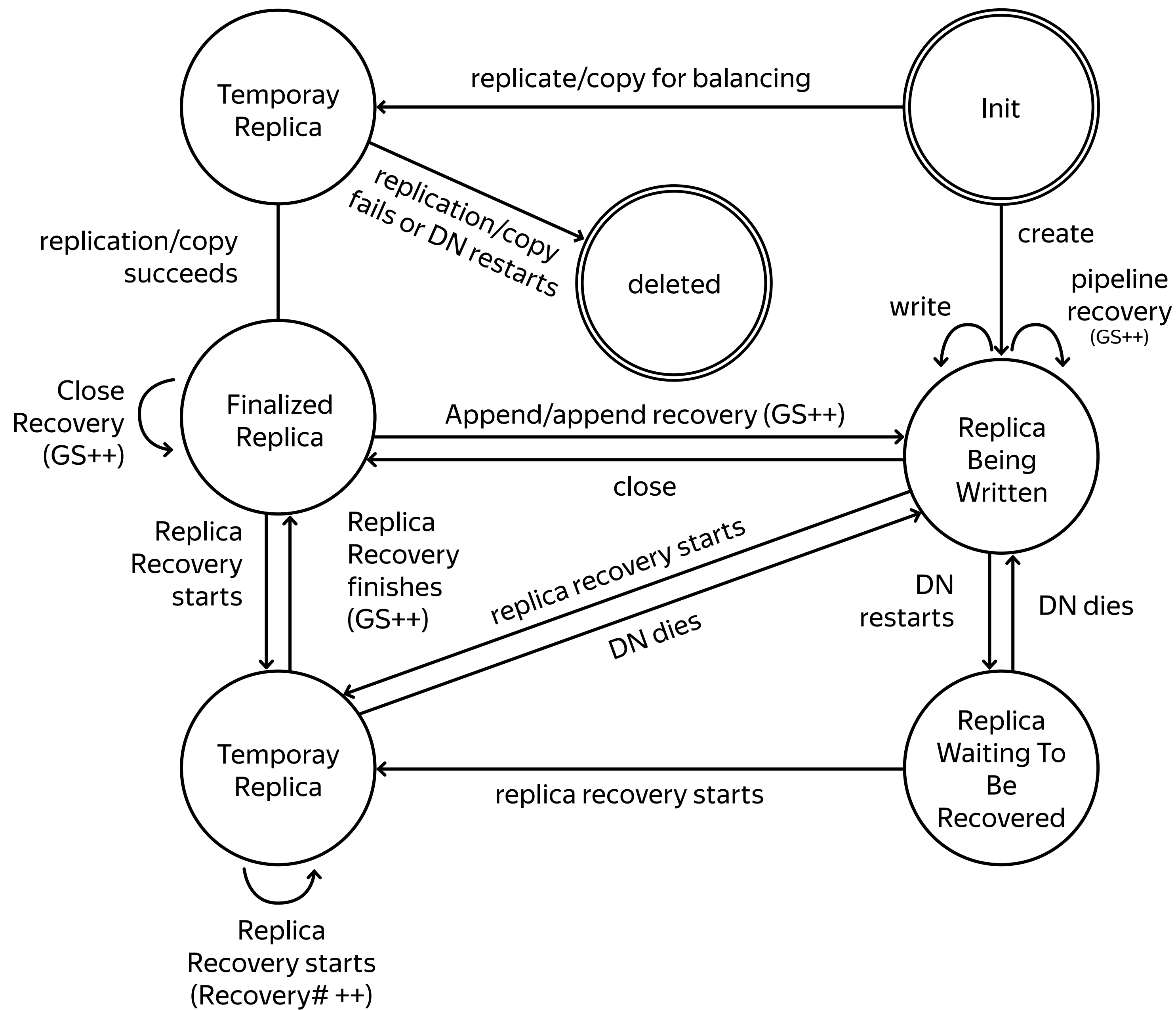
HDFS

Block and Replica States, Recovery Process

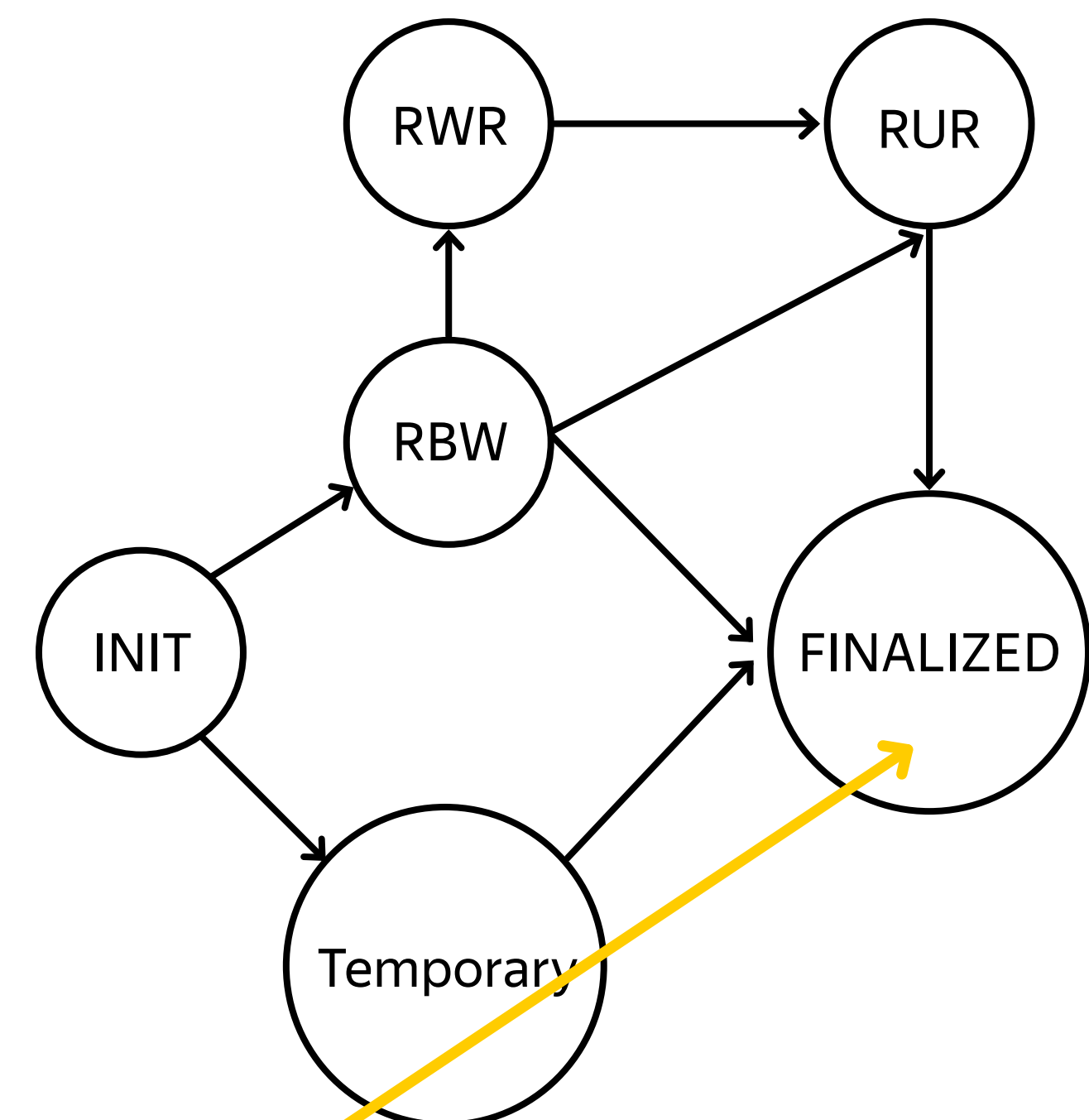
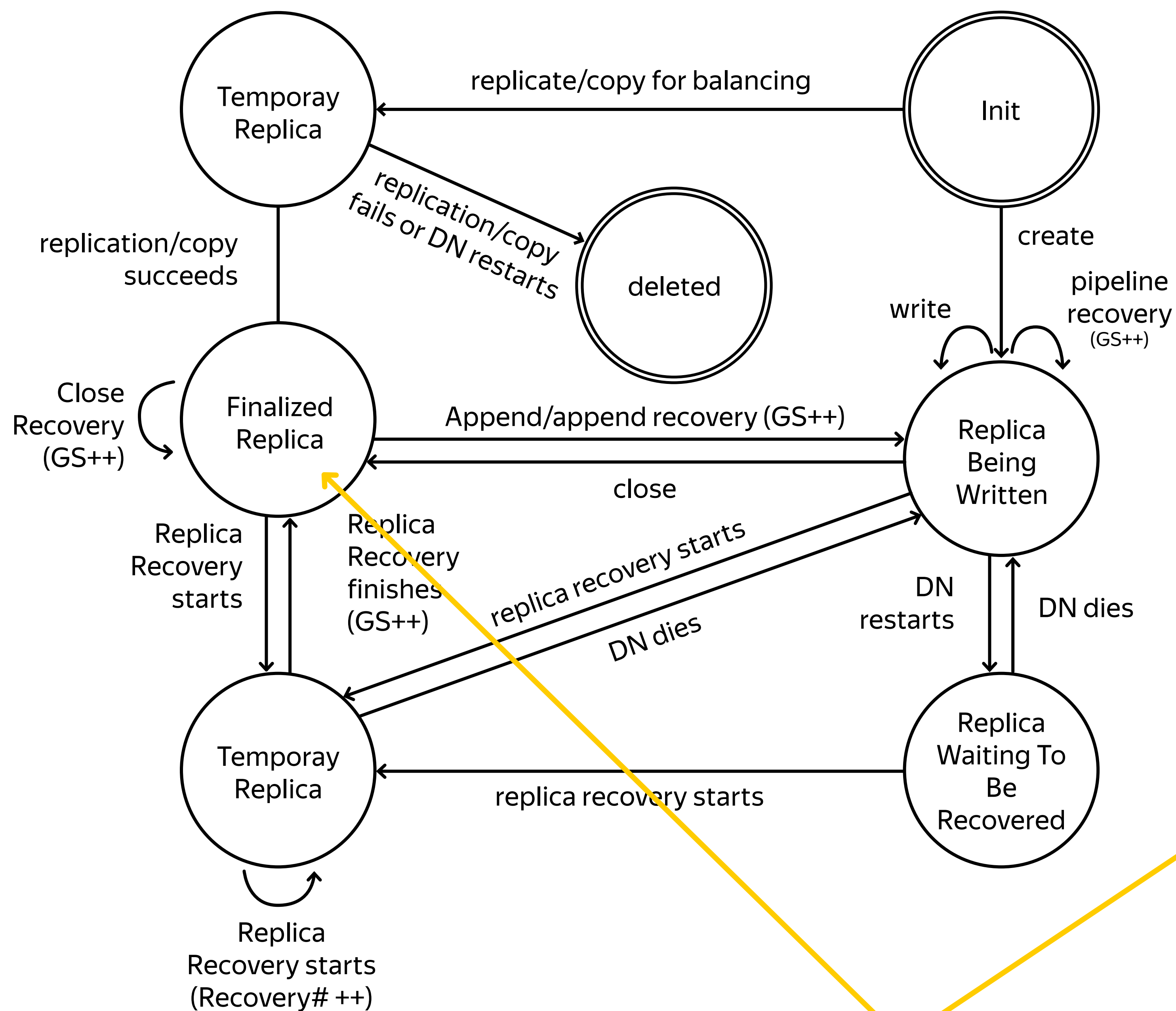








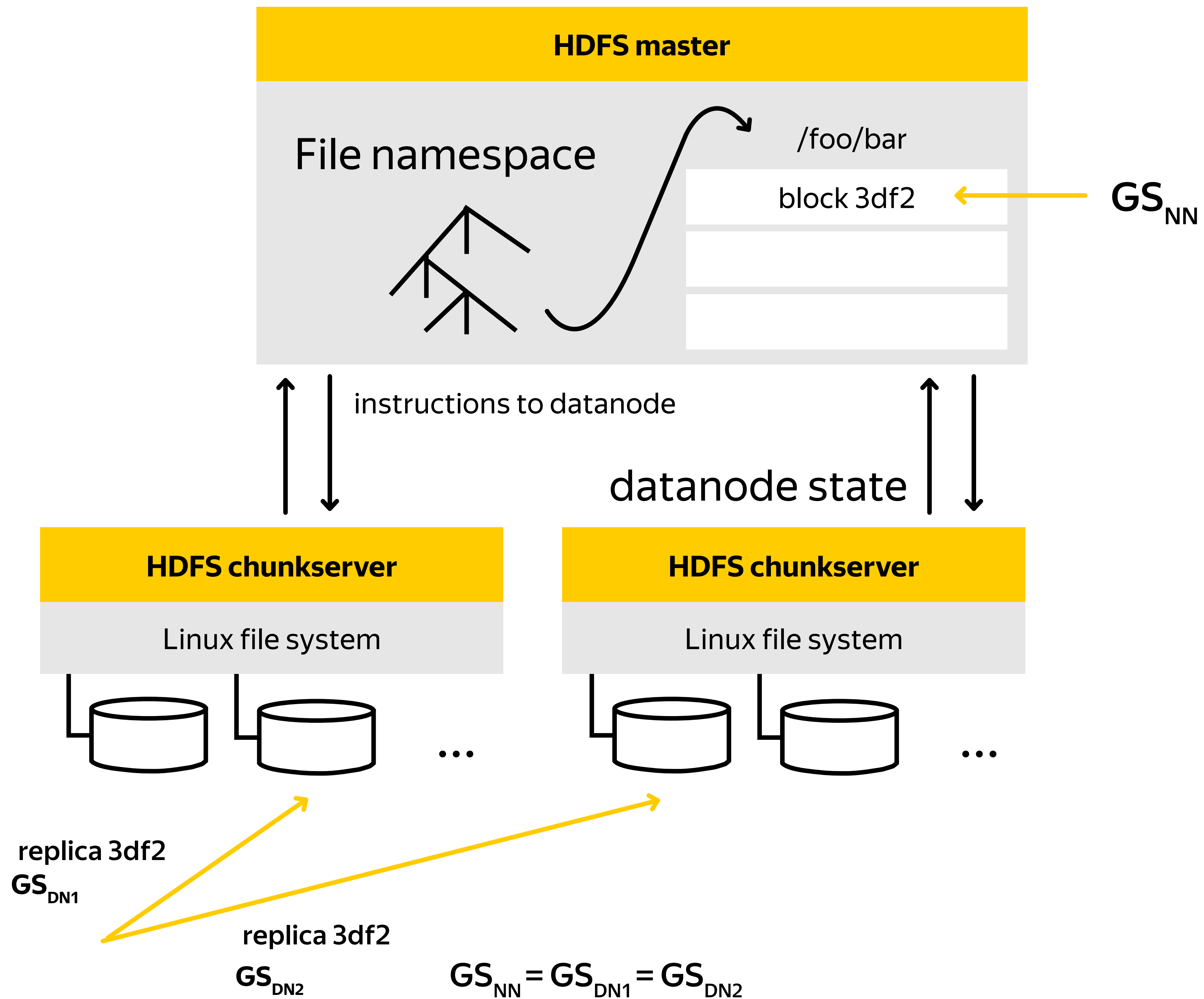
Simplified
Replica State
Transition

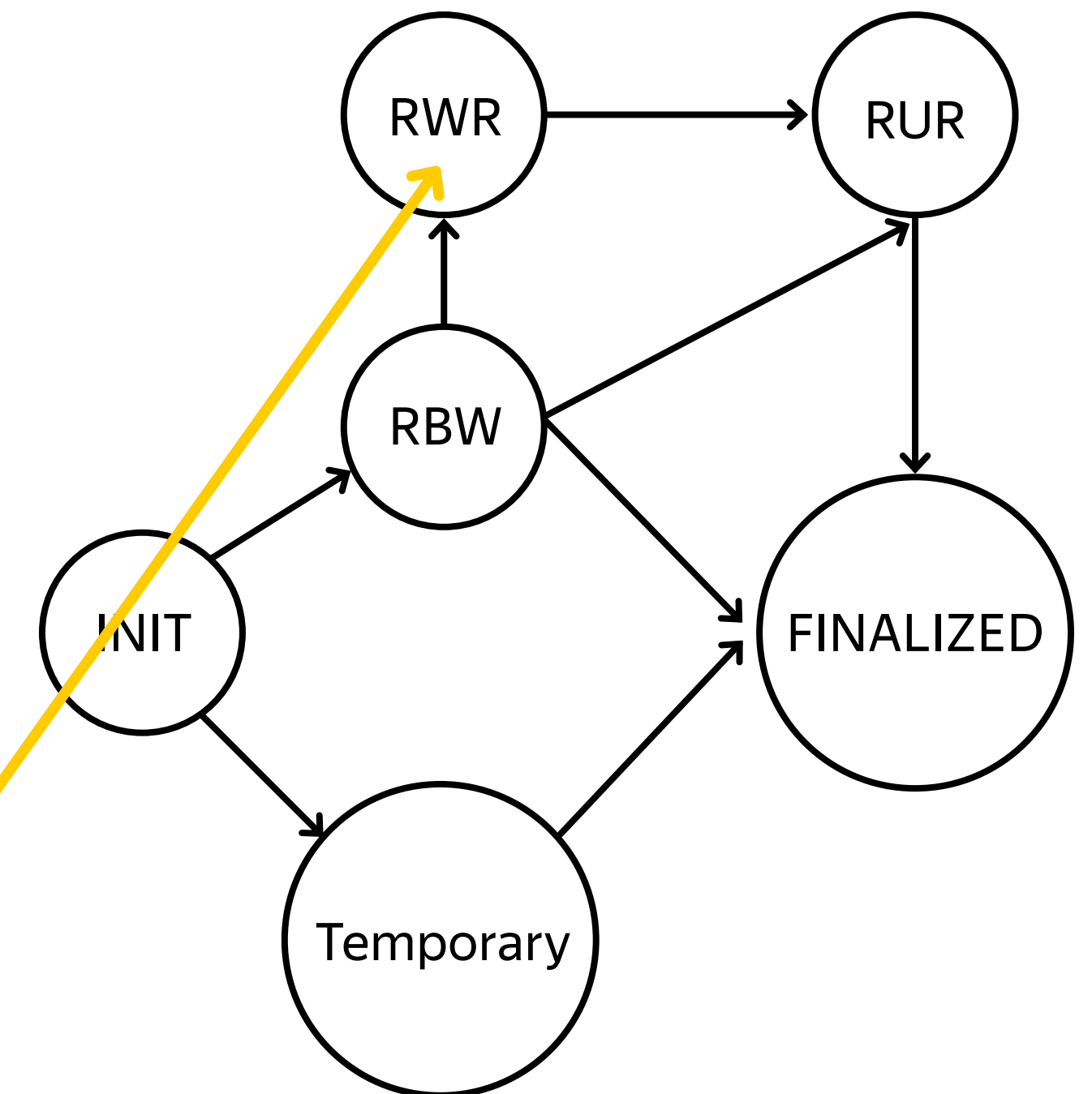
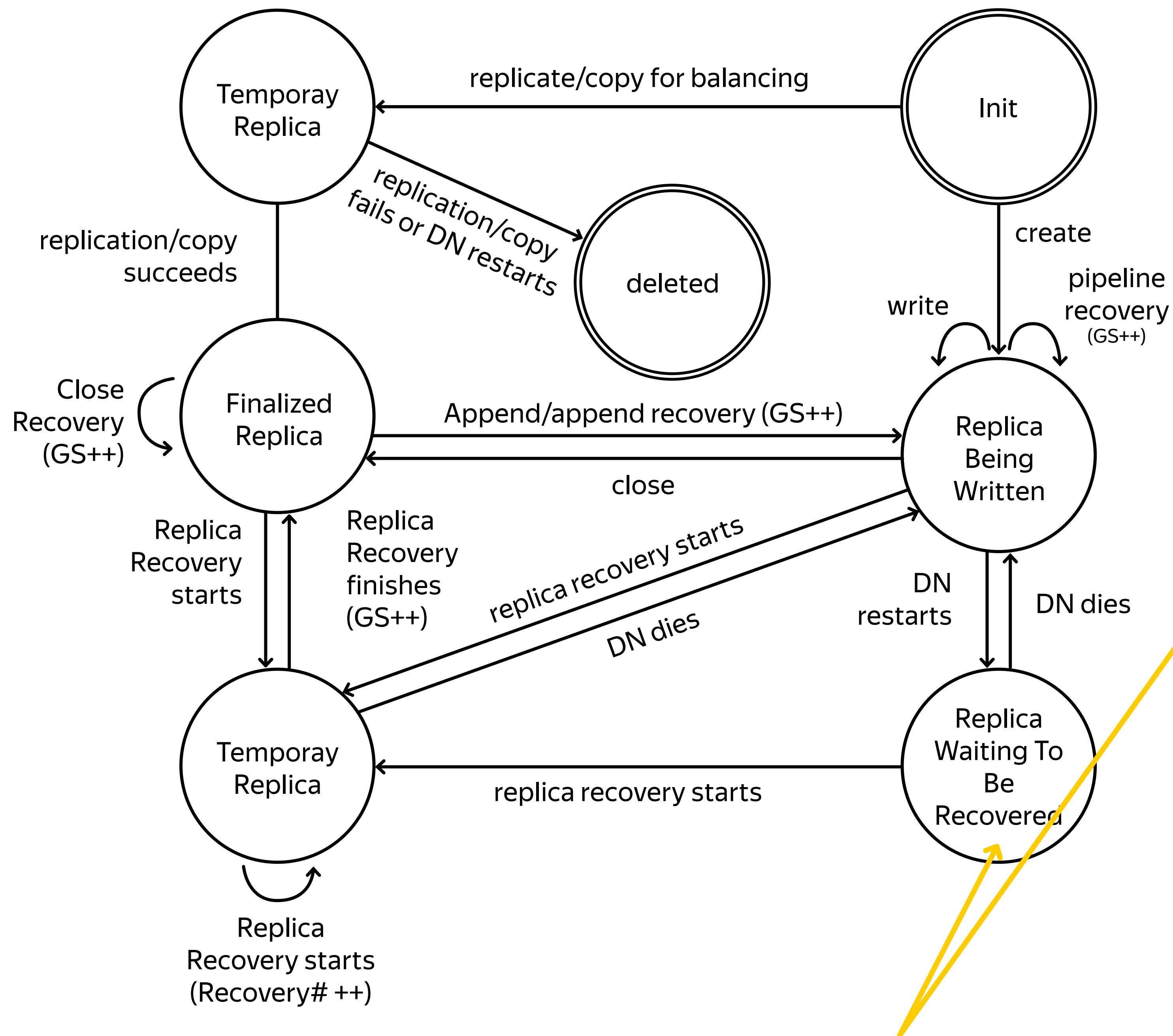


Simplified
Replica State
Transition

Finalized

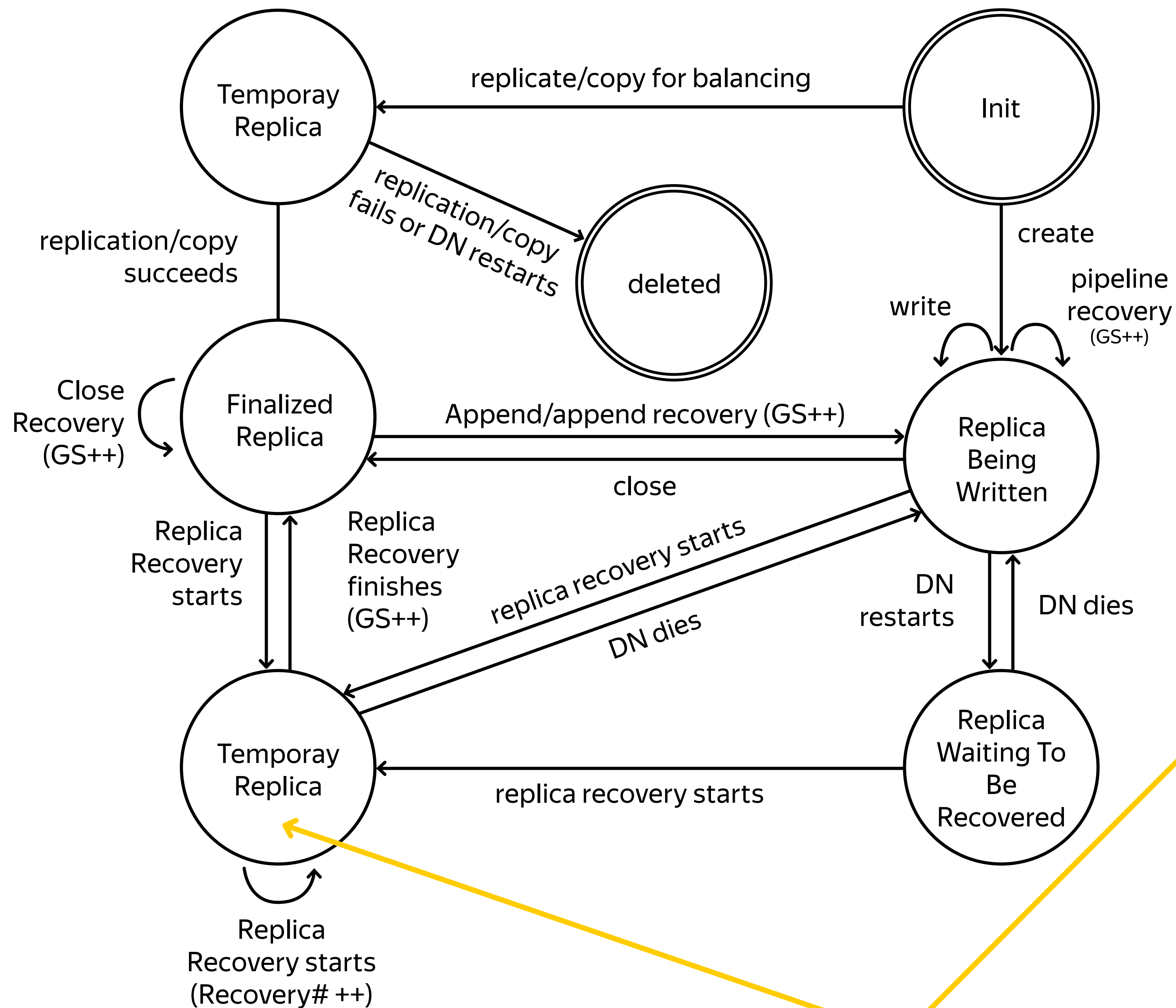
Finalized



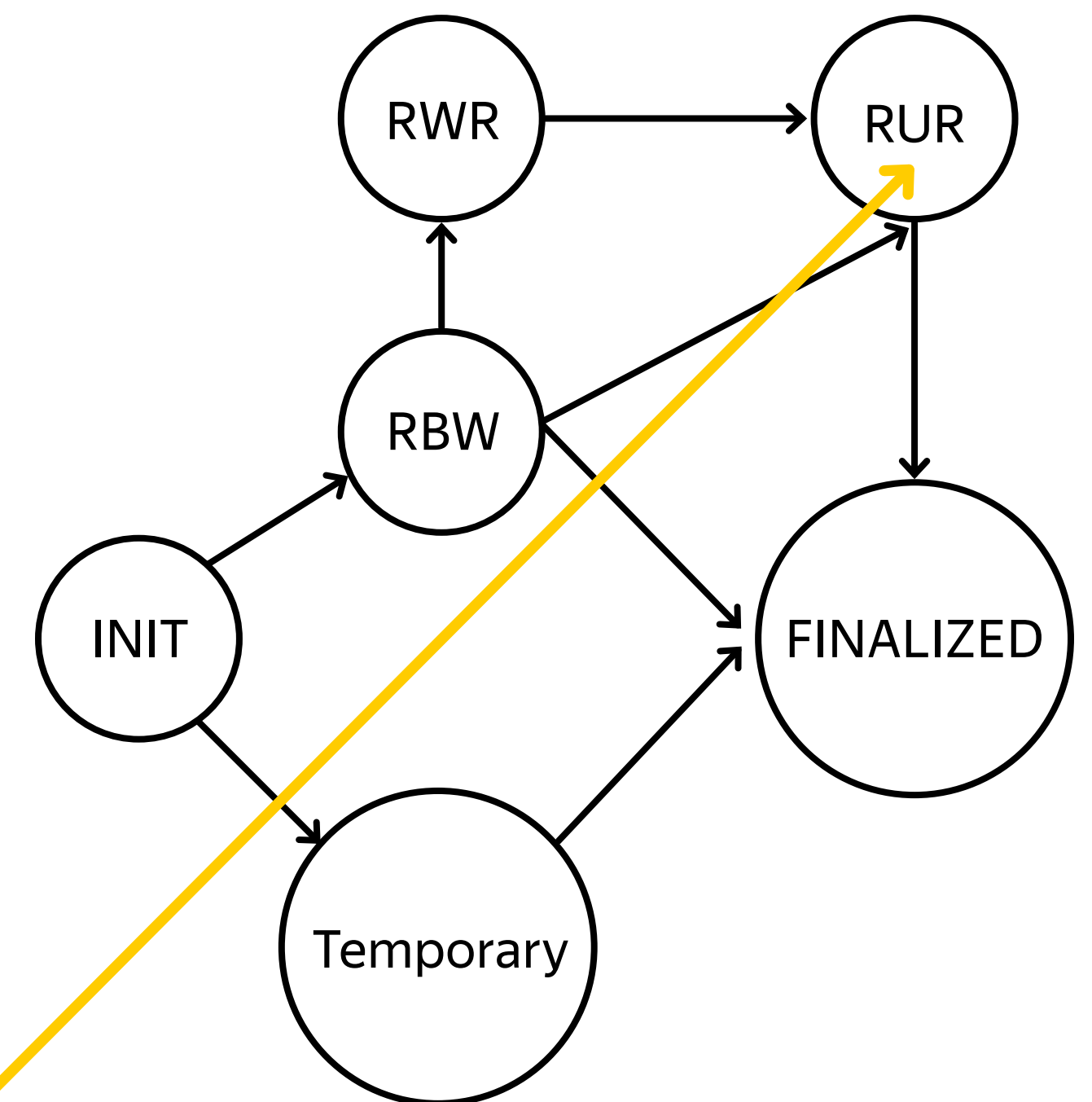


Simplified
Replica State
Transition

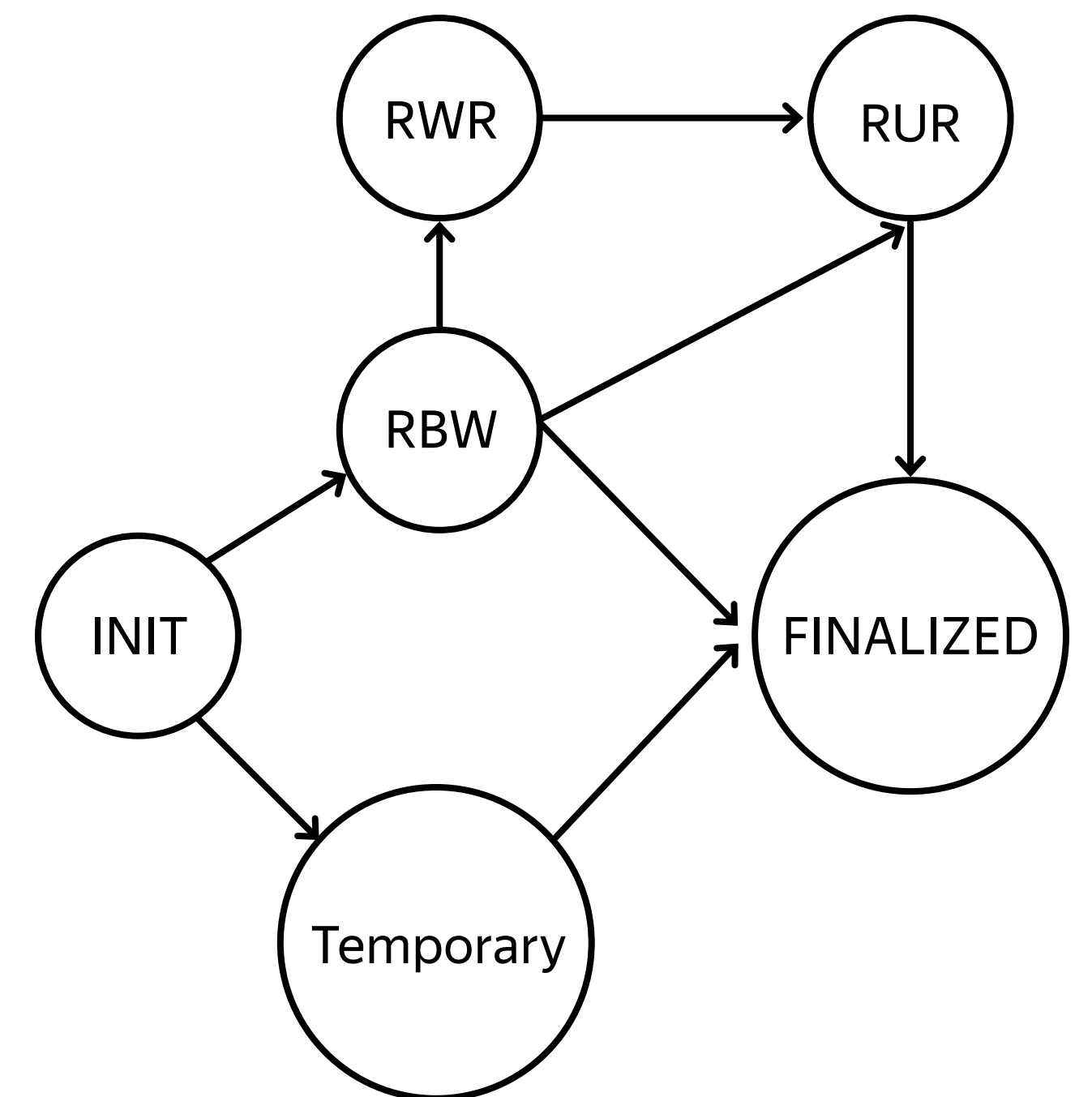
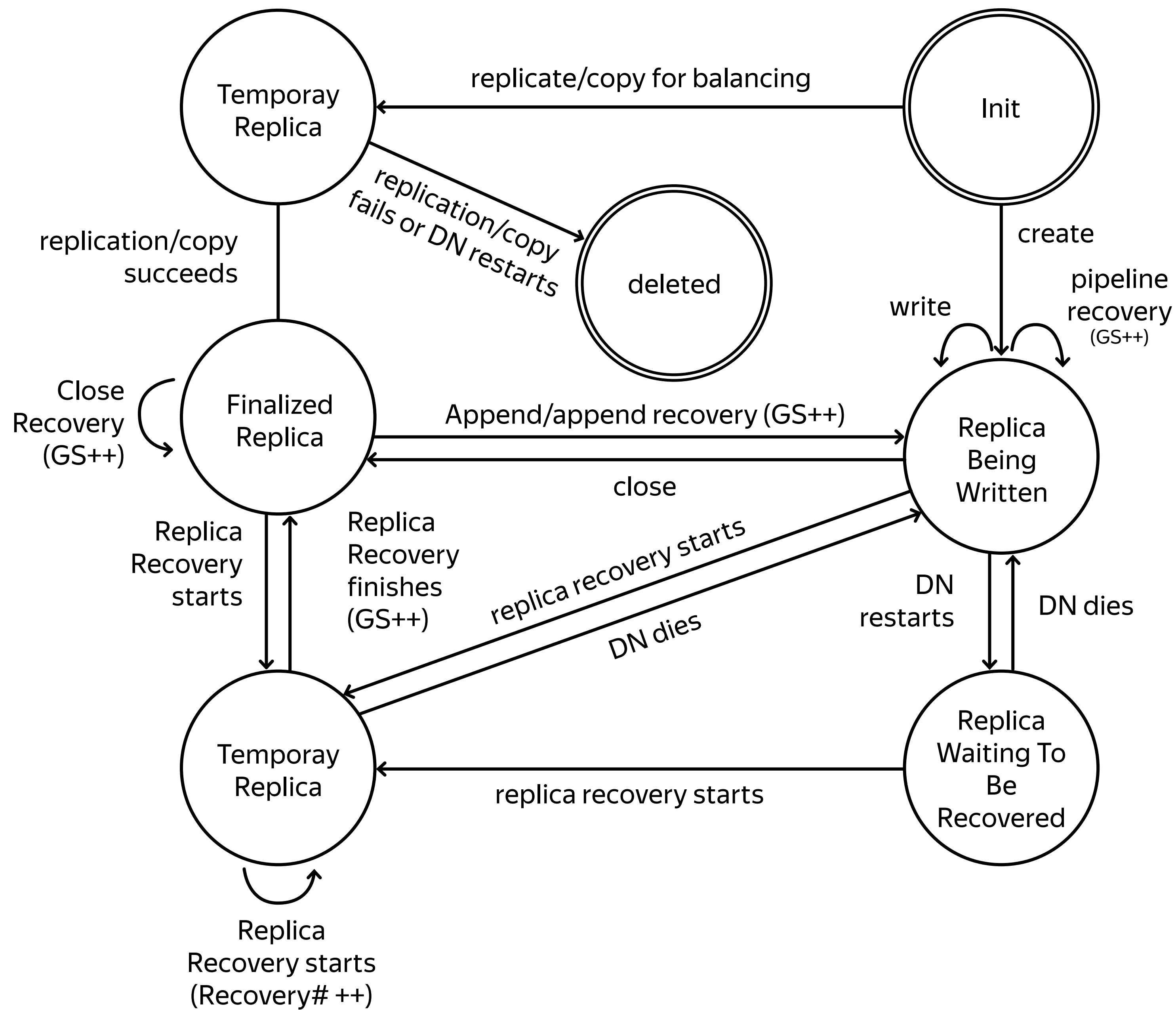
Replica Waiting to be Recovered



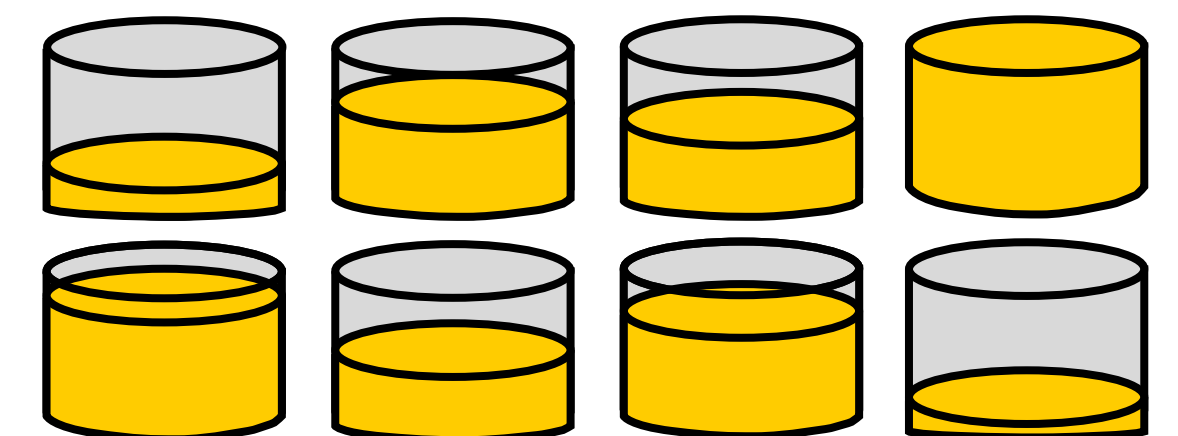
Replica Under Recovery

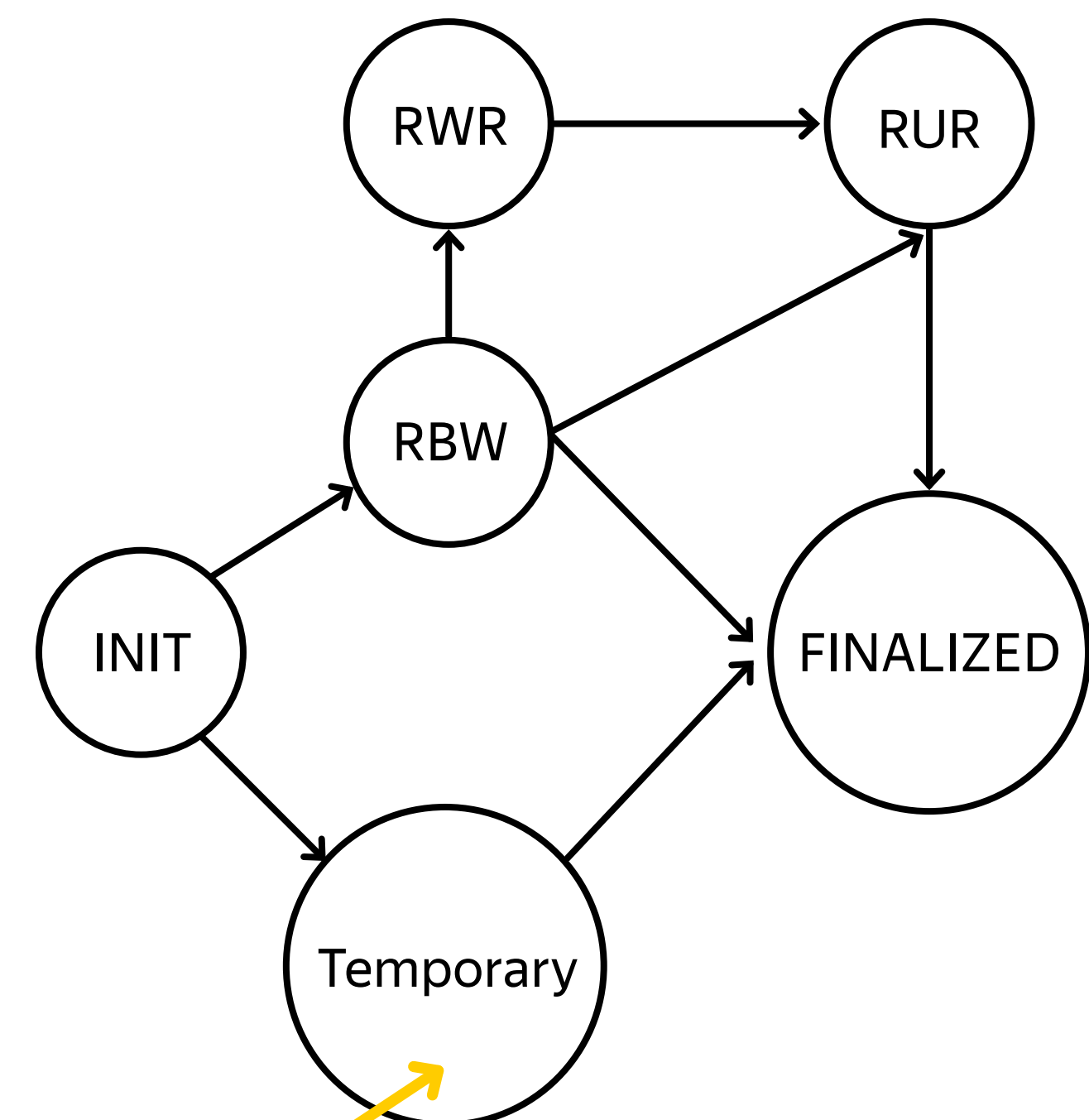
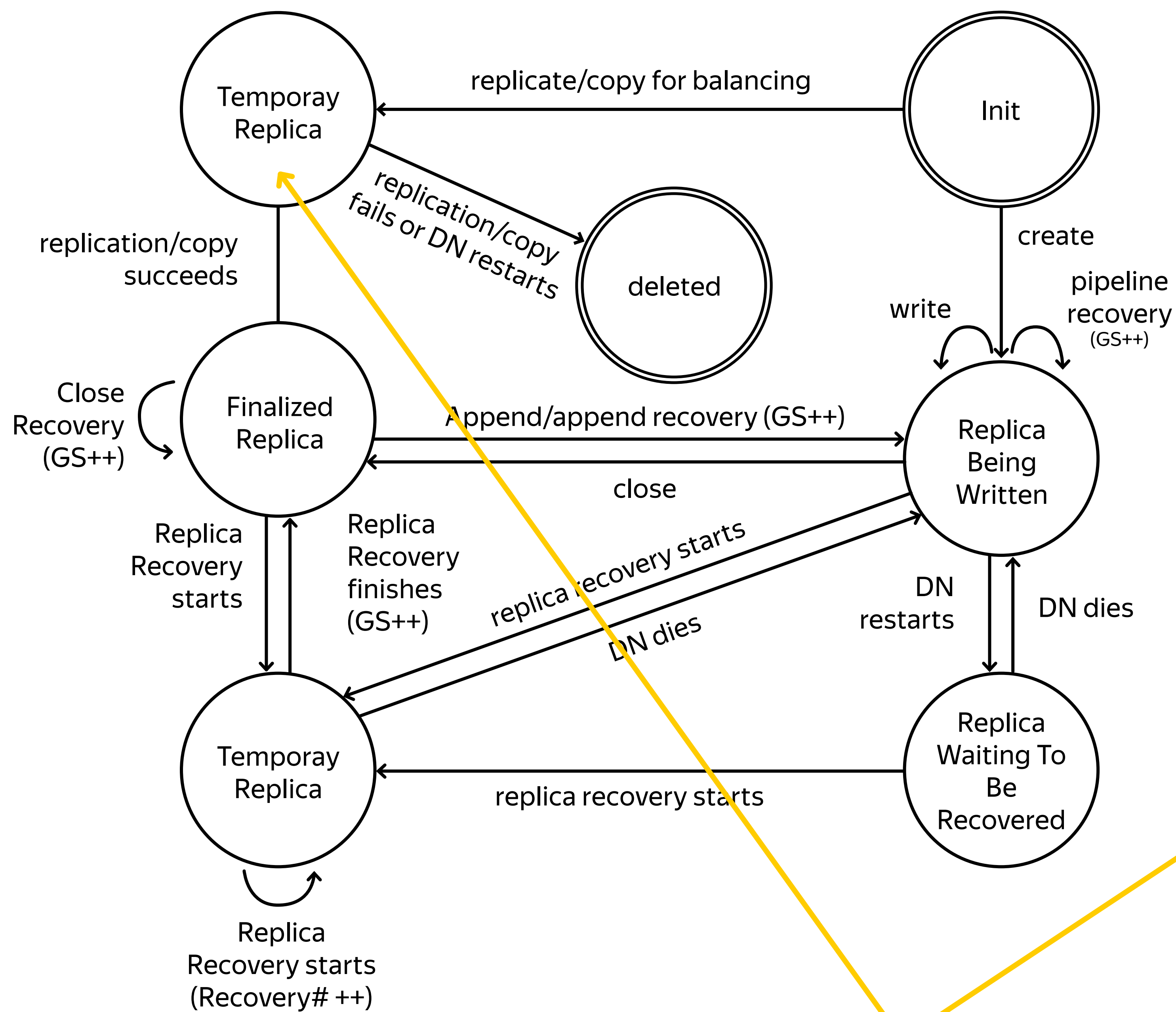


**Simplified
Replica State
Transition**

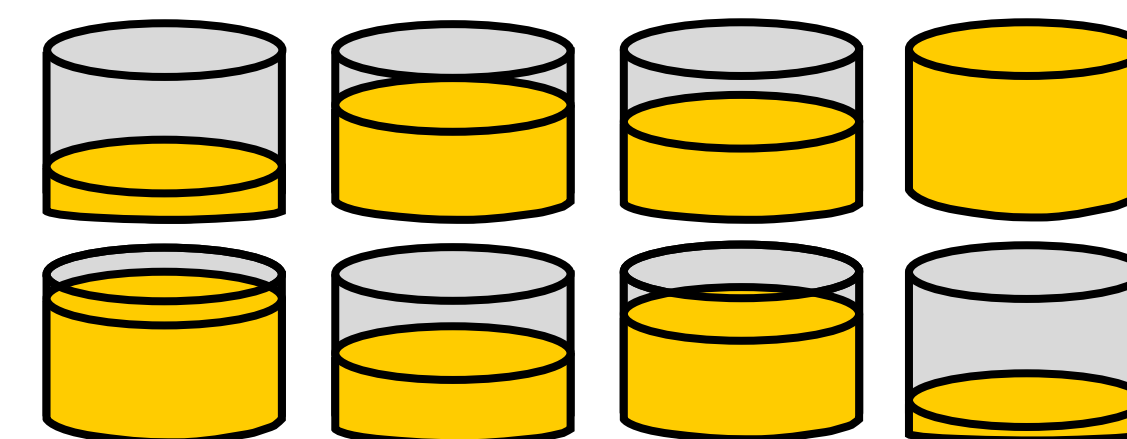


Simplified
Replica State
Transition

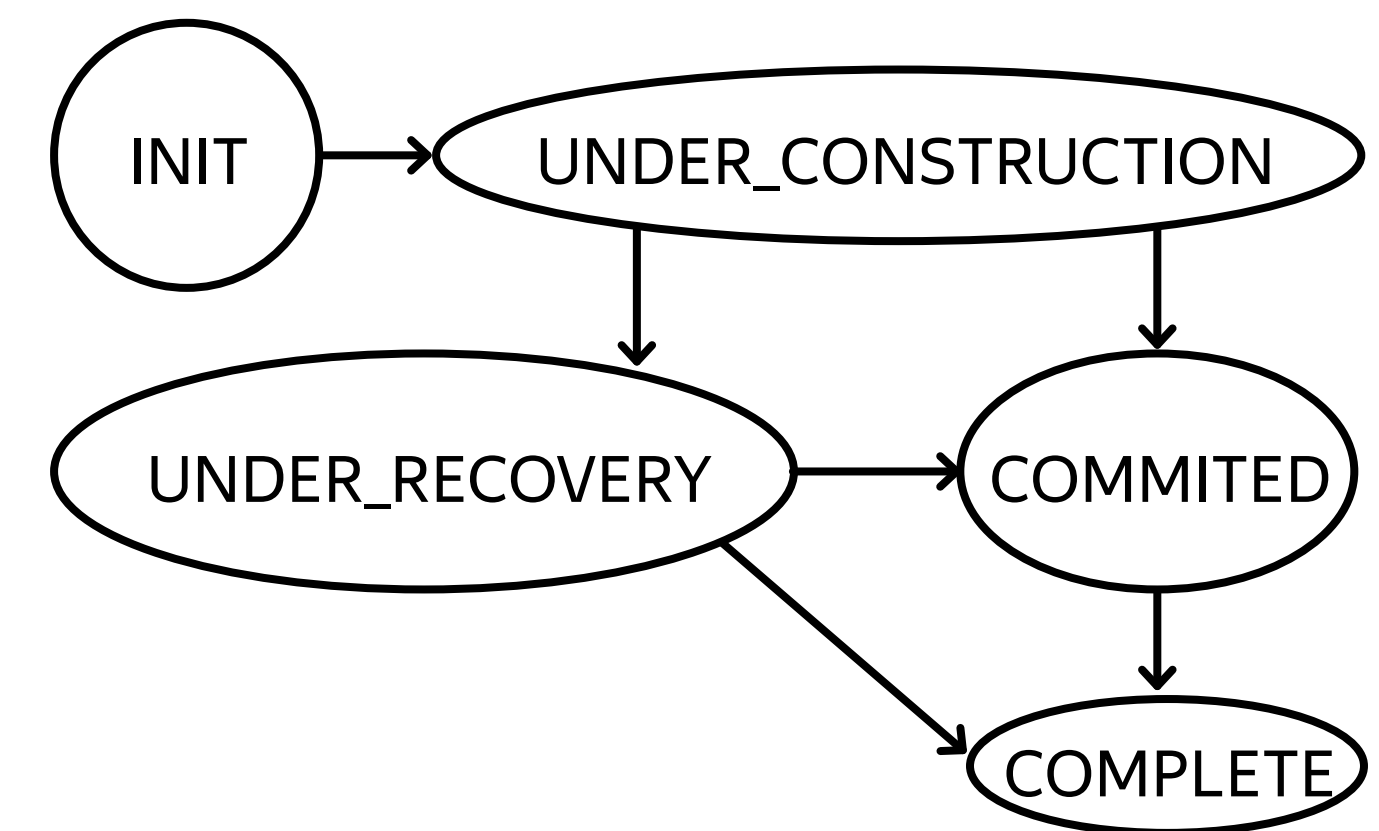
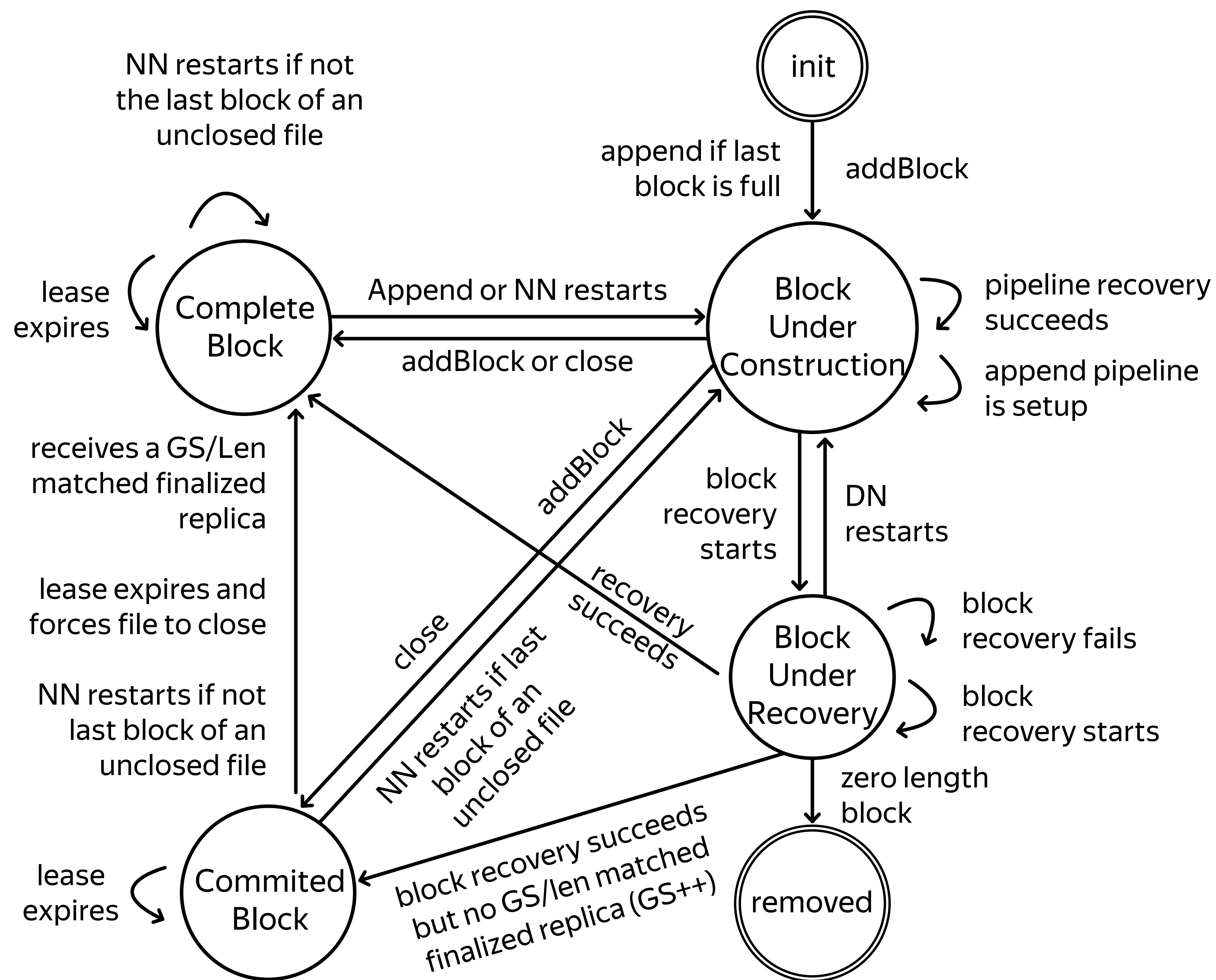




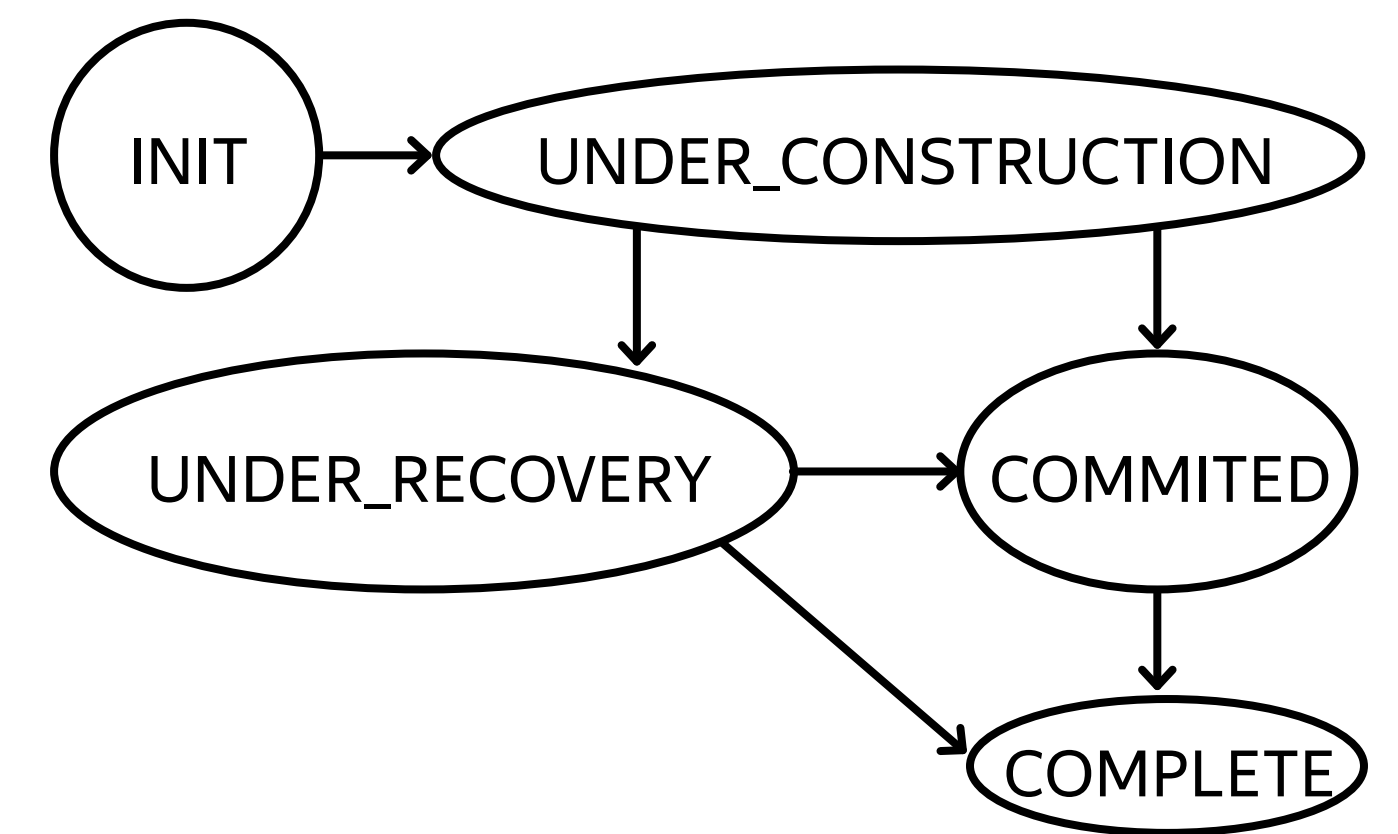
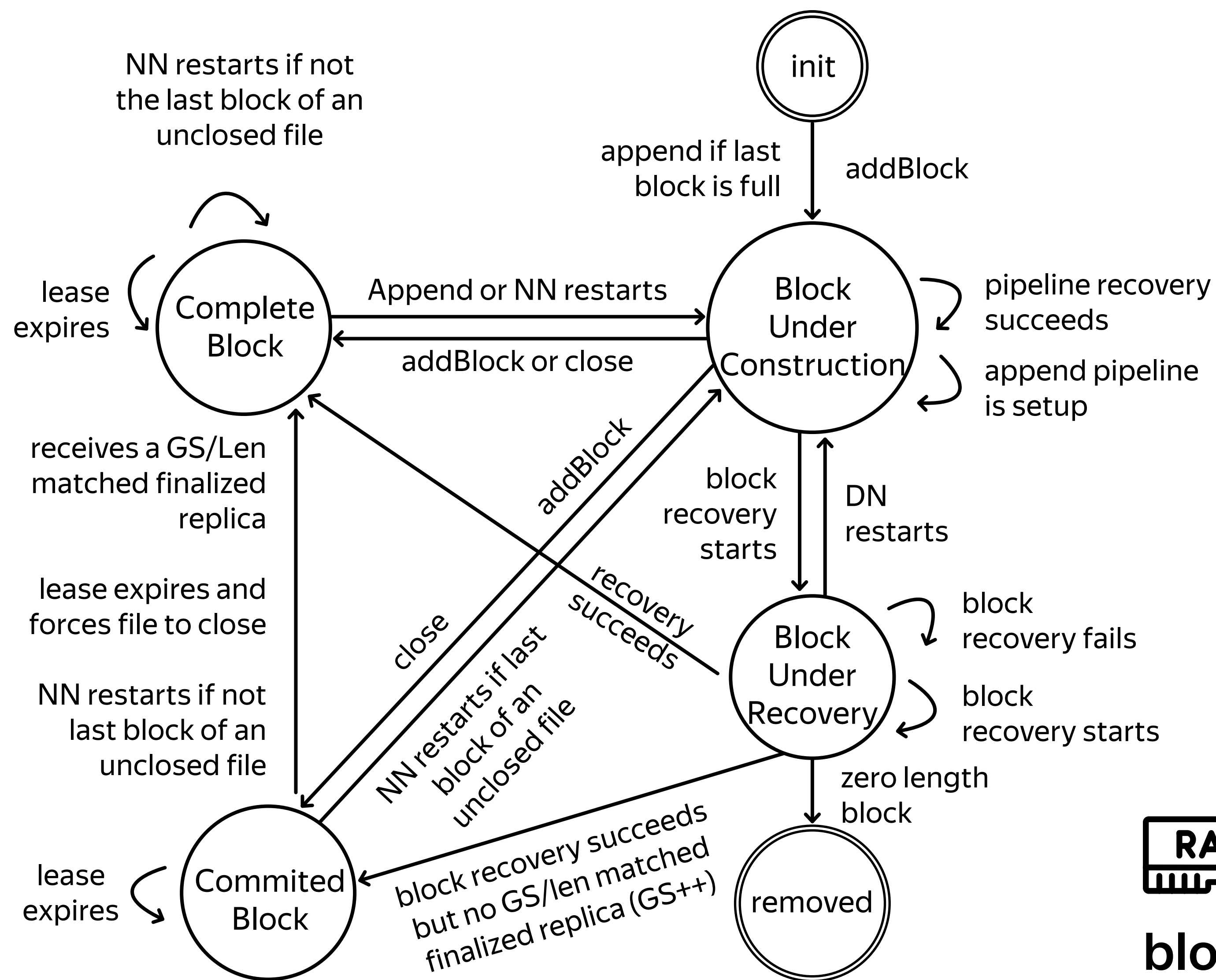
Simplified
Replica State
Transition



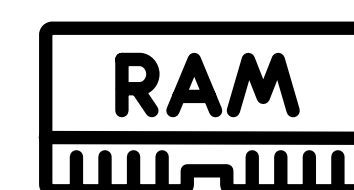
Temporary



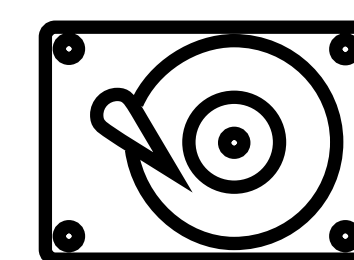
Simplified
Block State
Transition



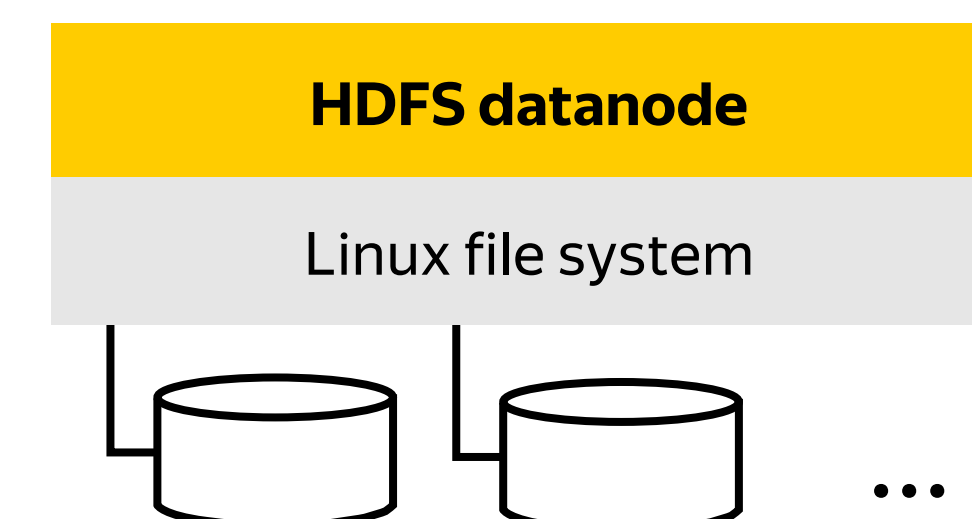
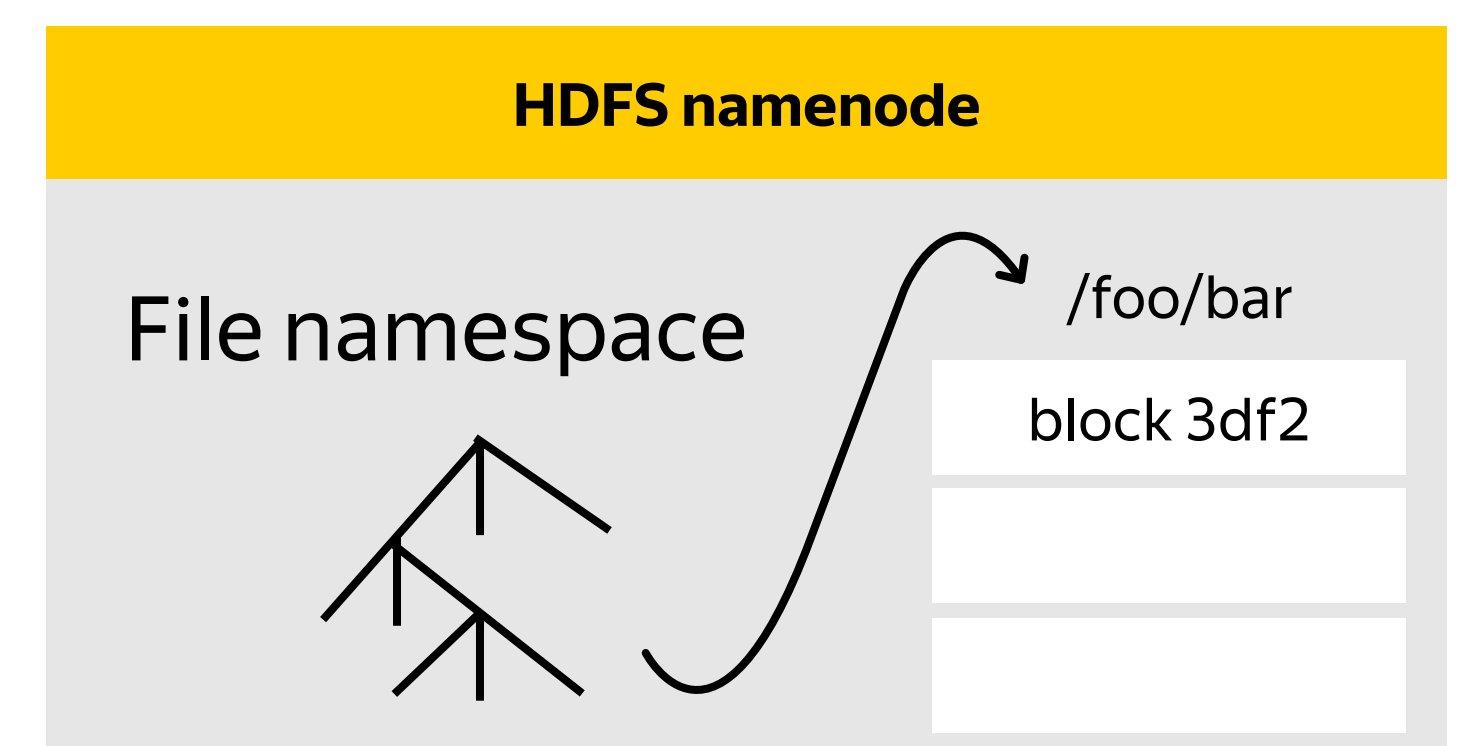
Simplified Block State Transition

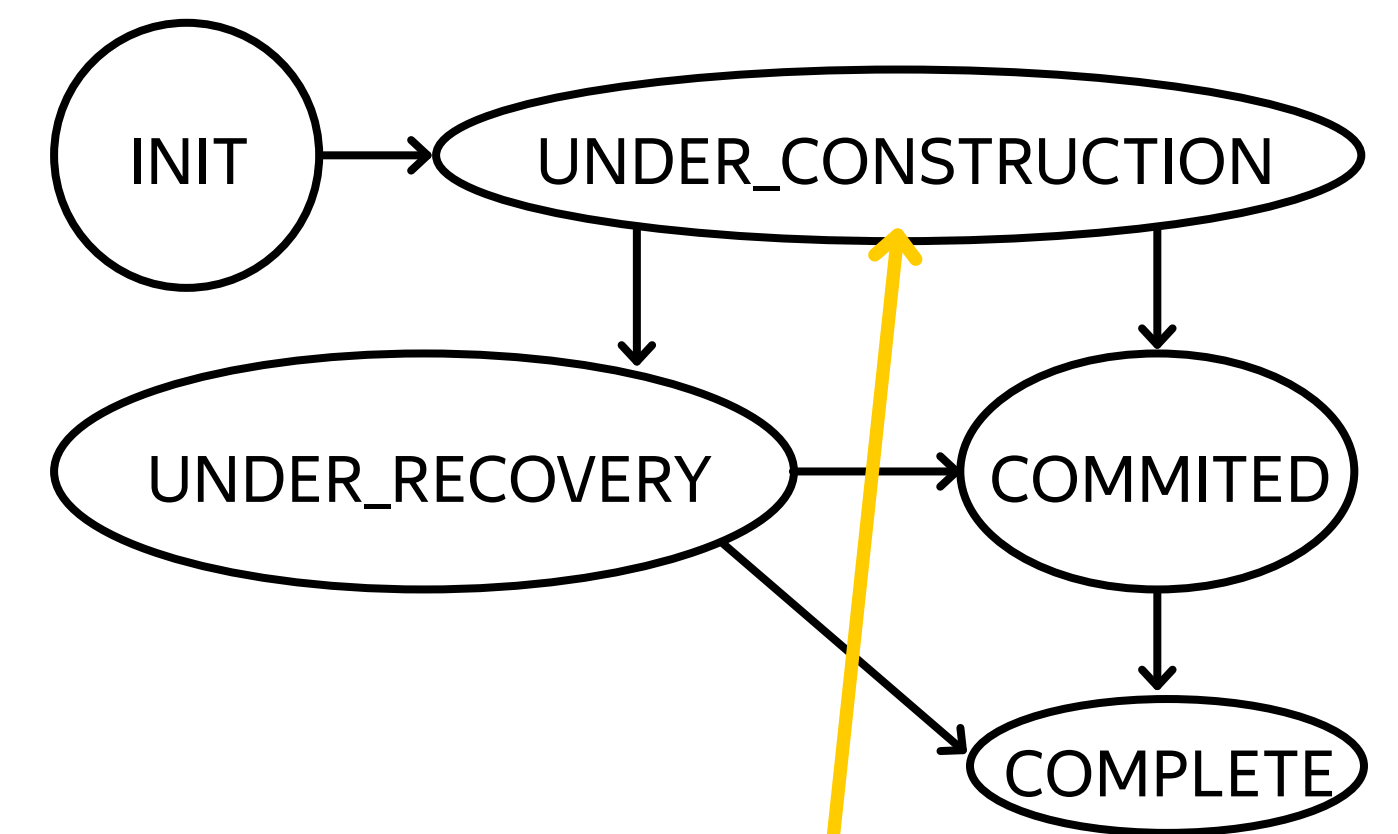
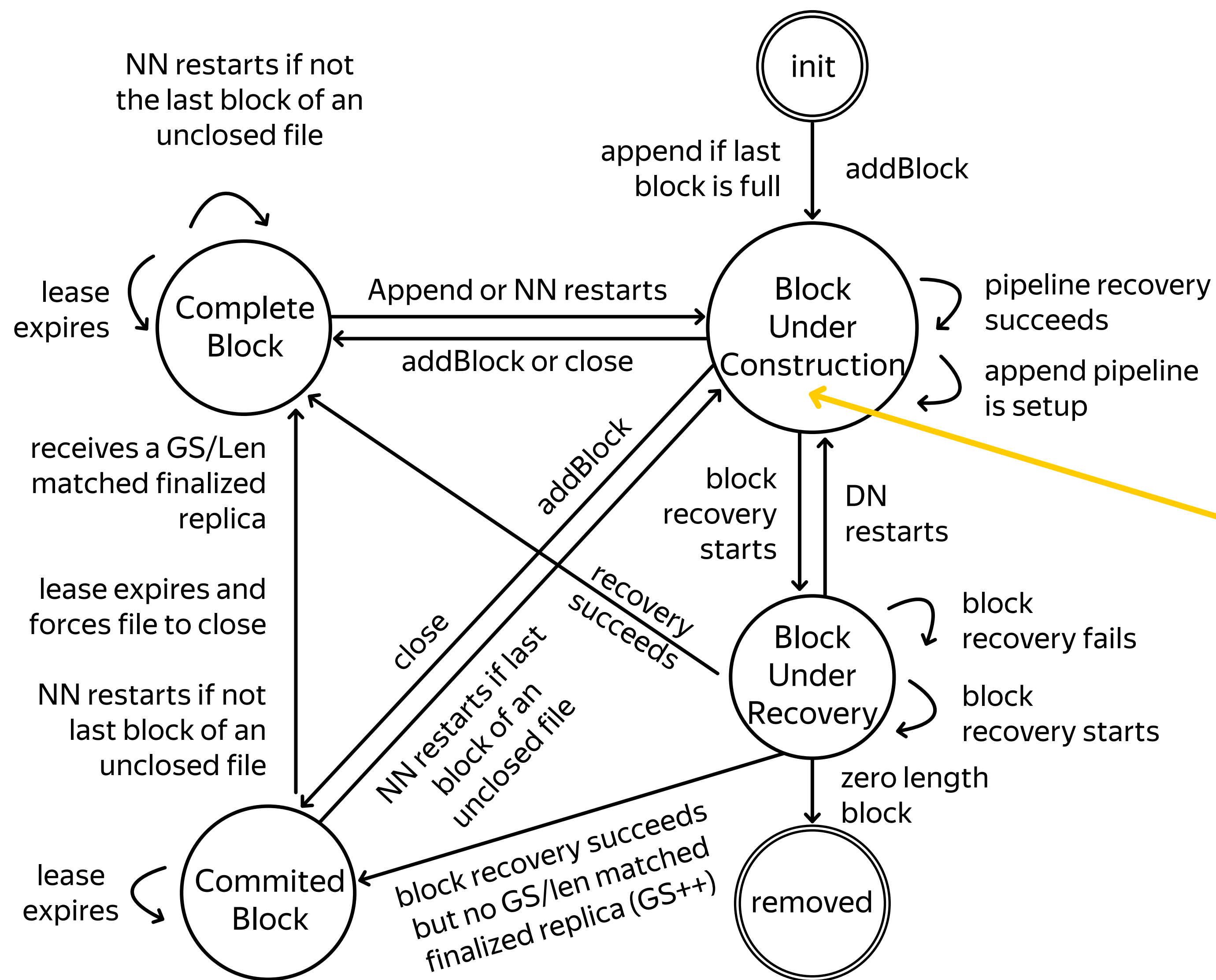


**block
state**



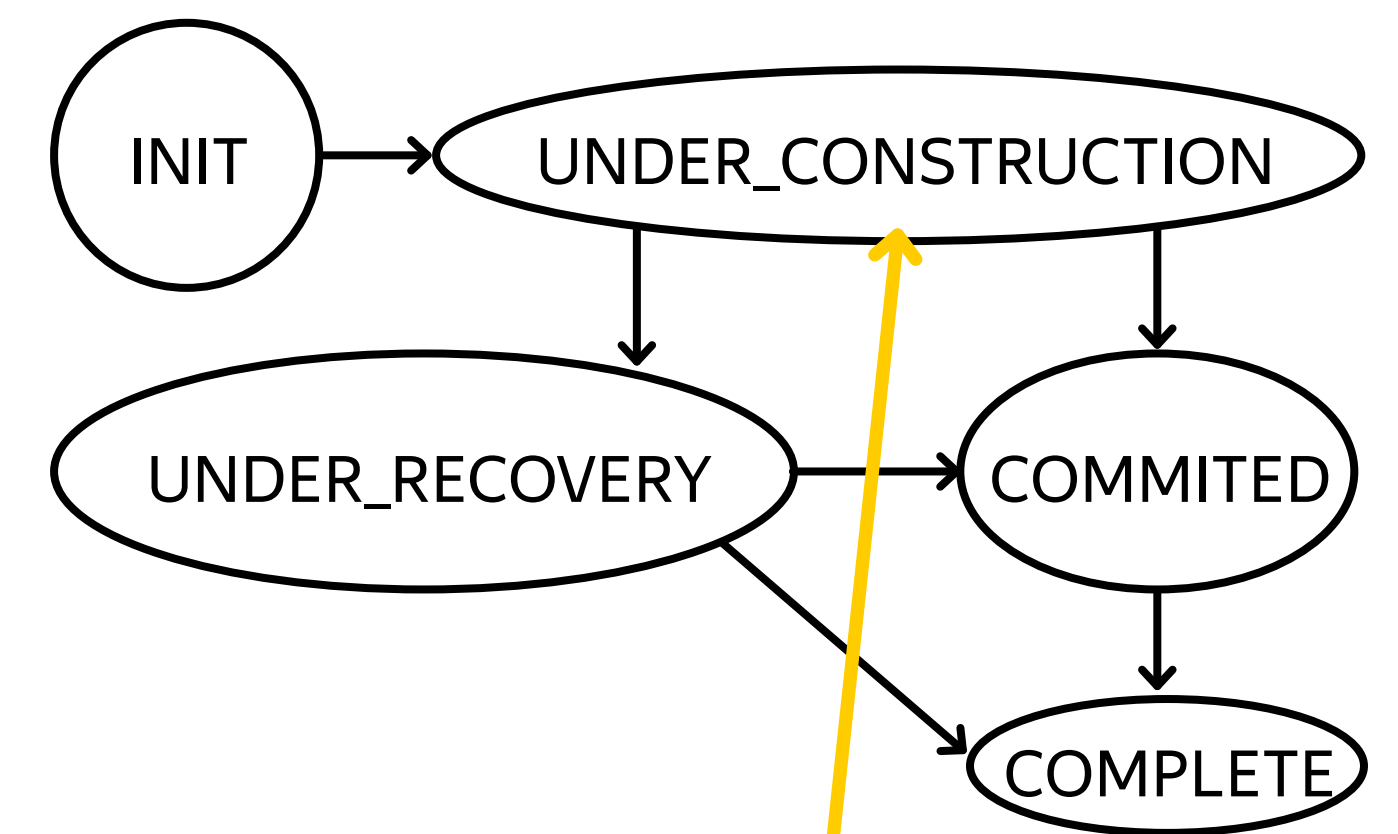
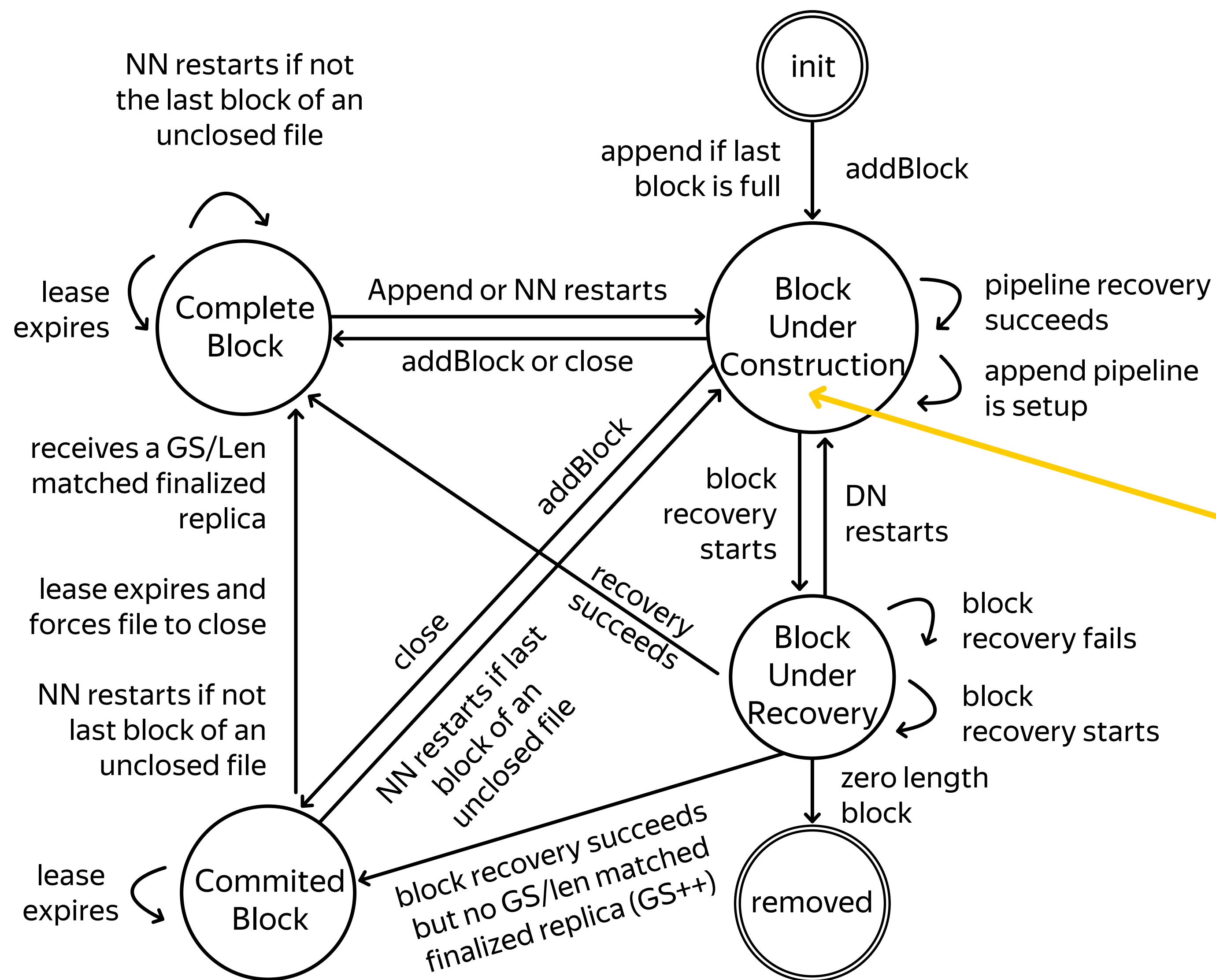
**replica
state**





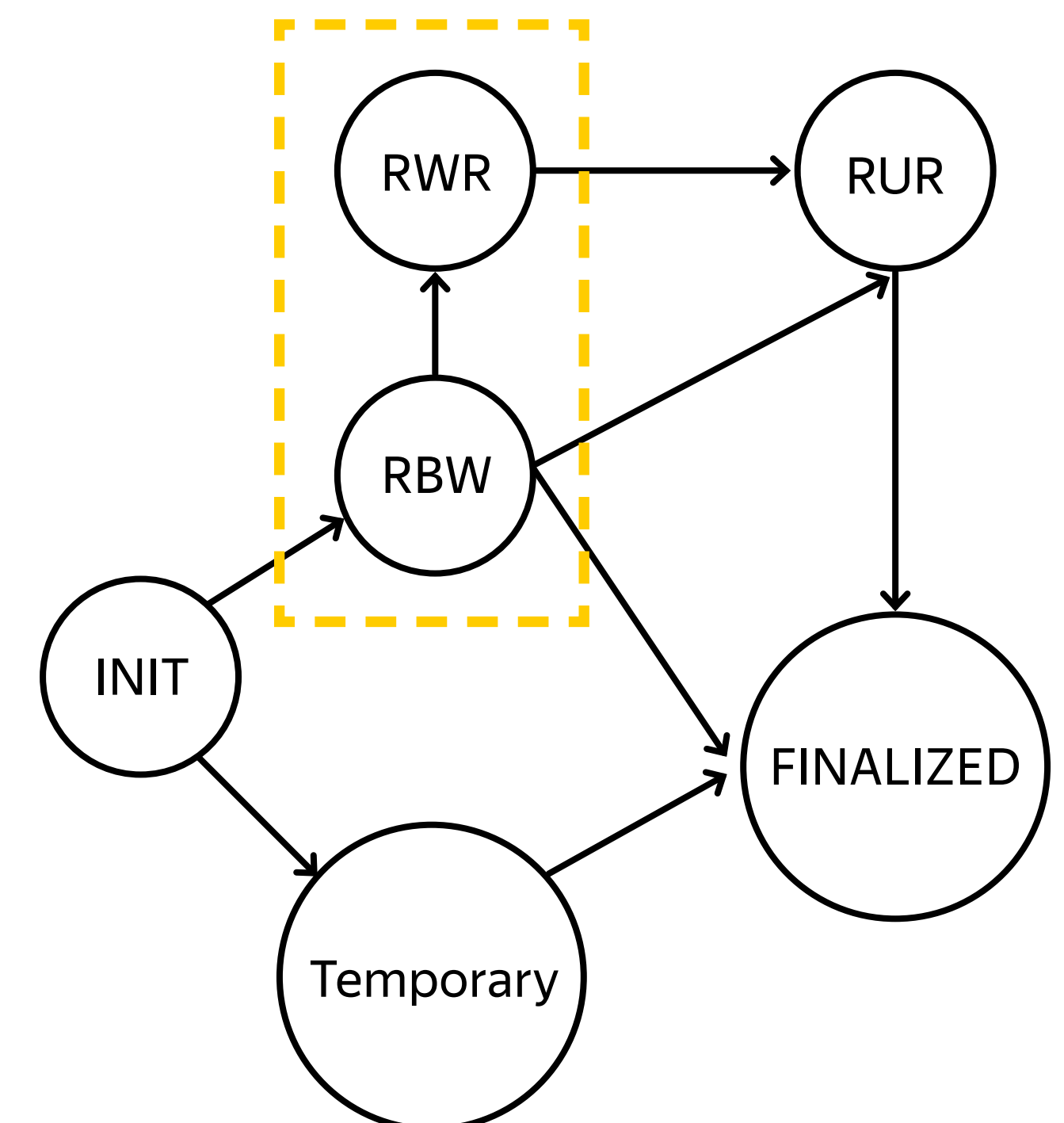
Simplified
Block State
Transition

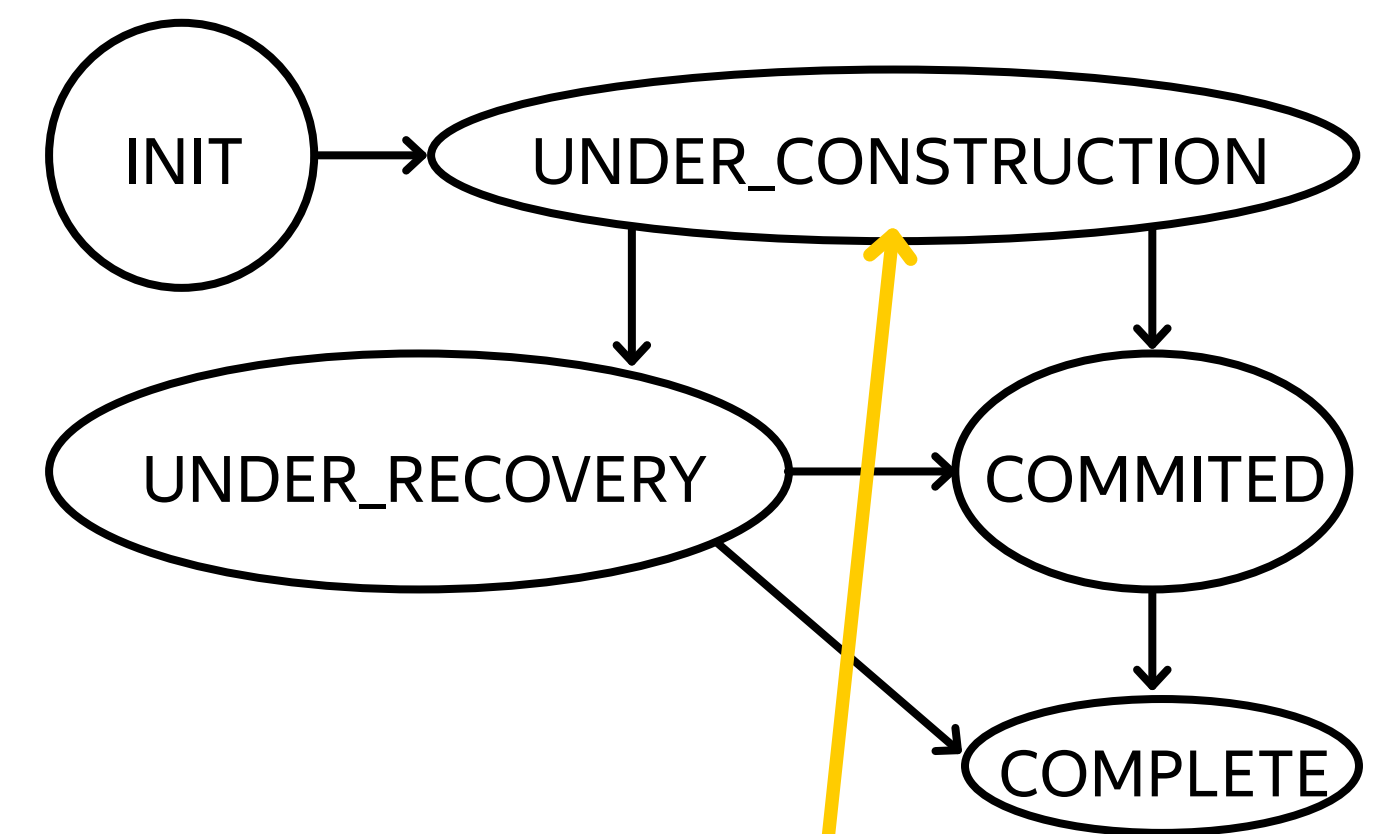
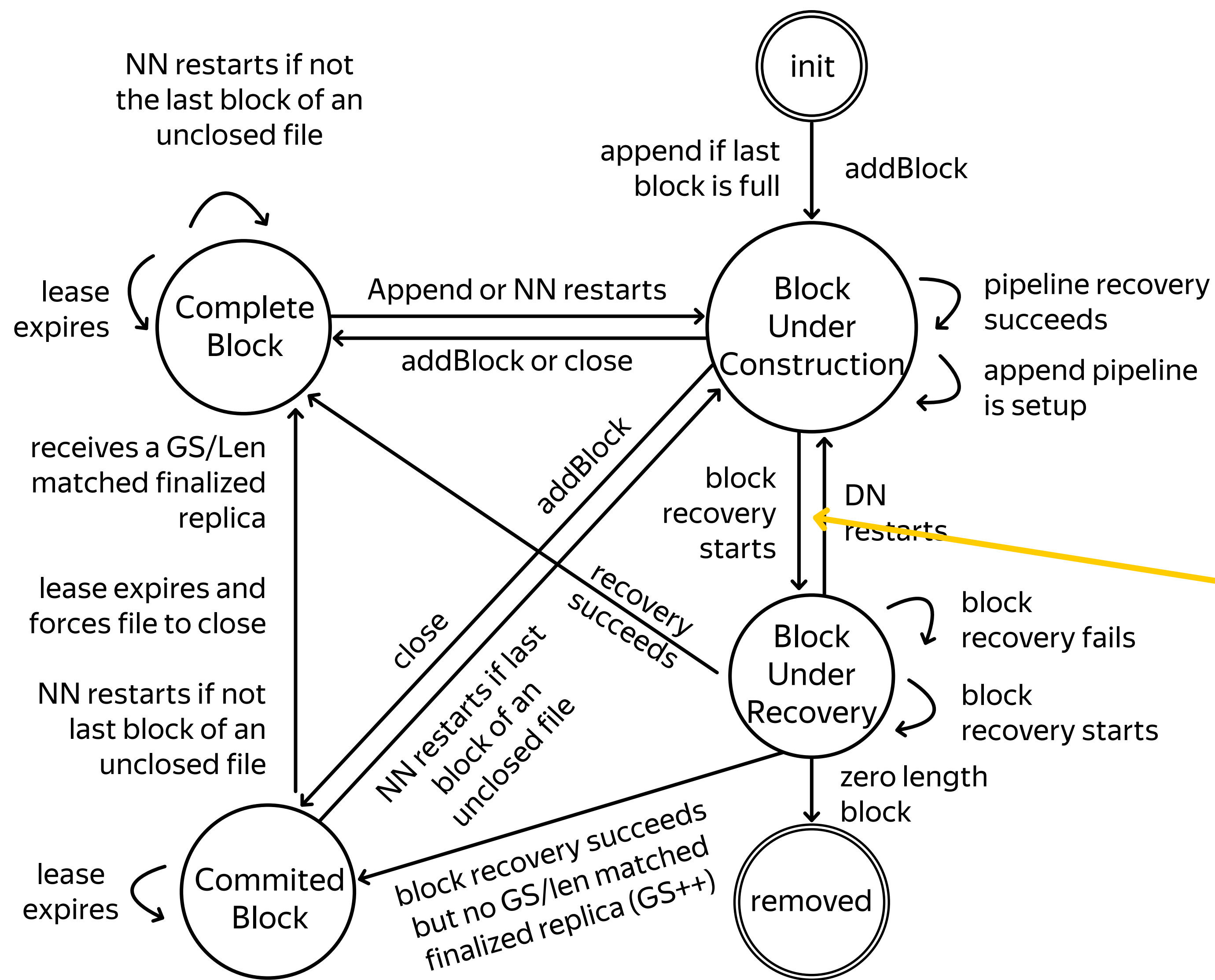
under_construction



Simplified
Block State
Transition

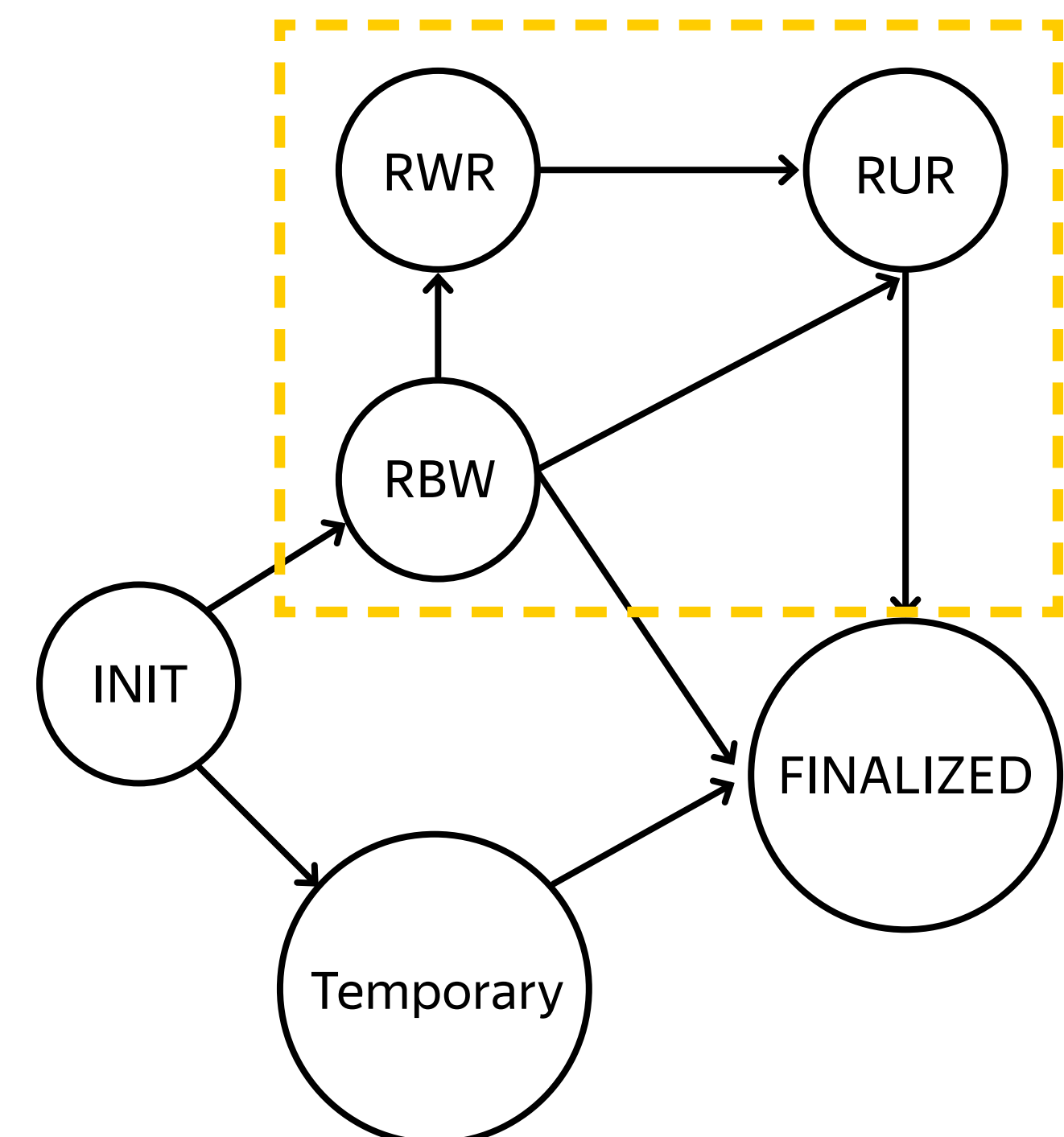
under_construction

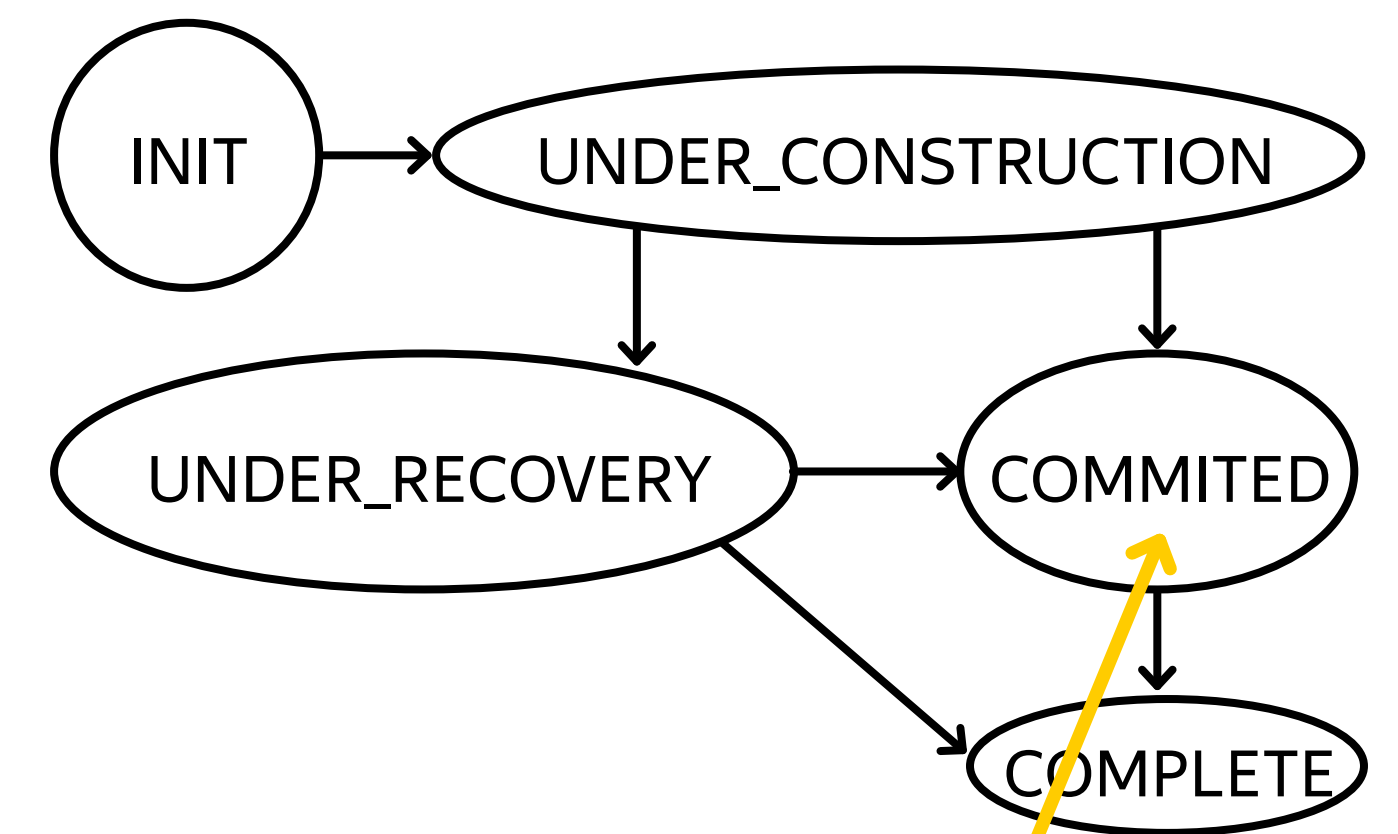
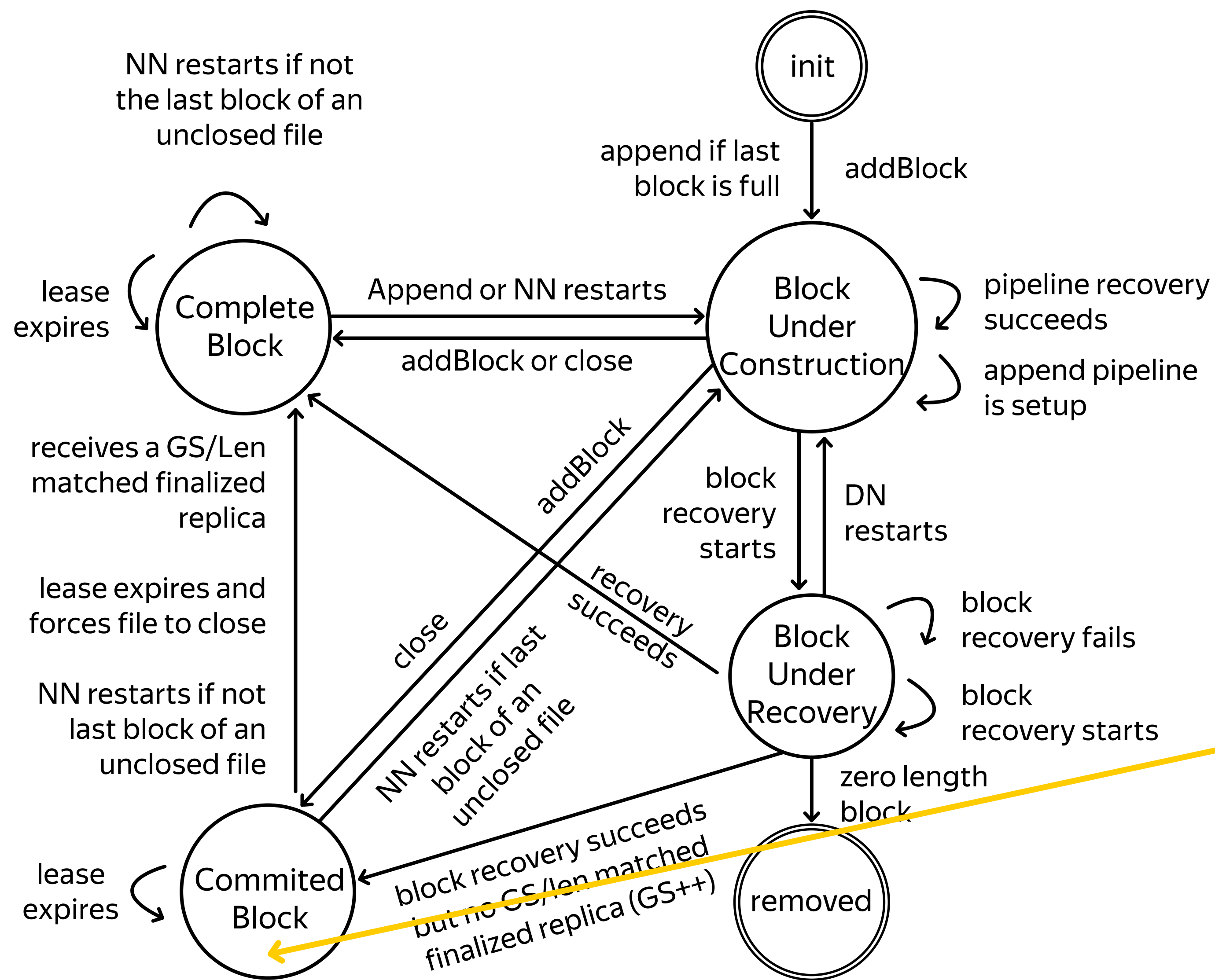




Simplified
Block State
Transition

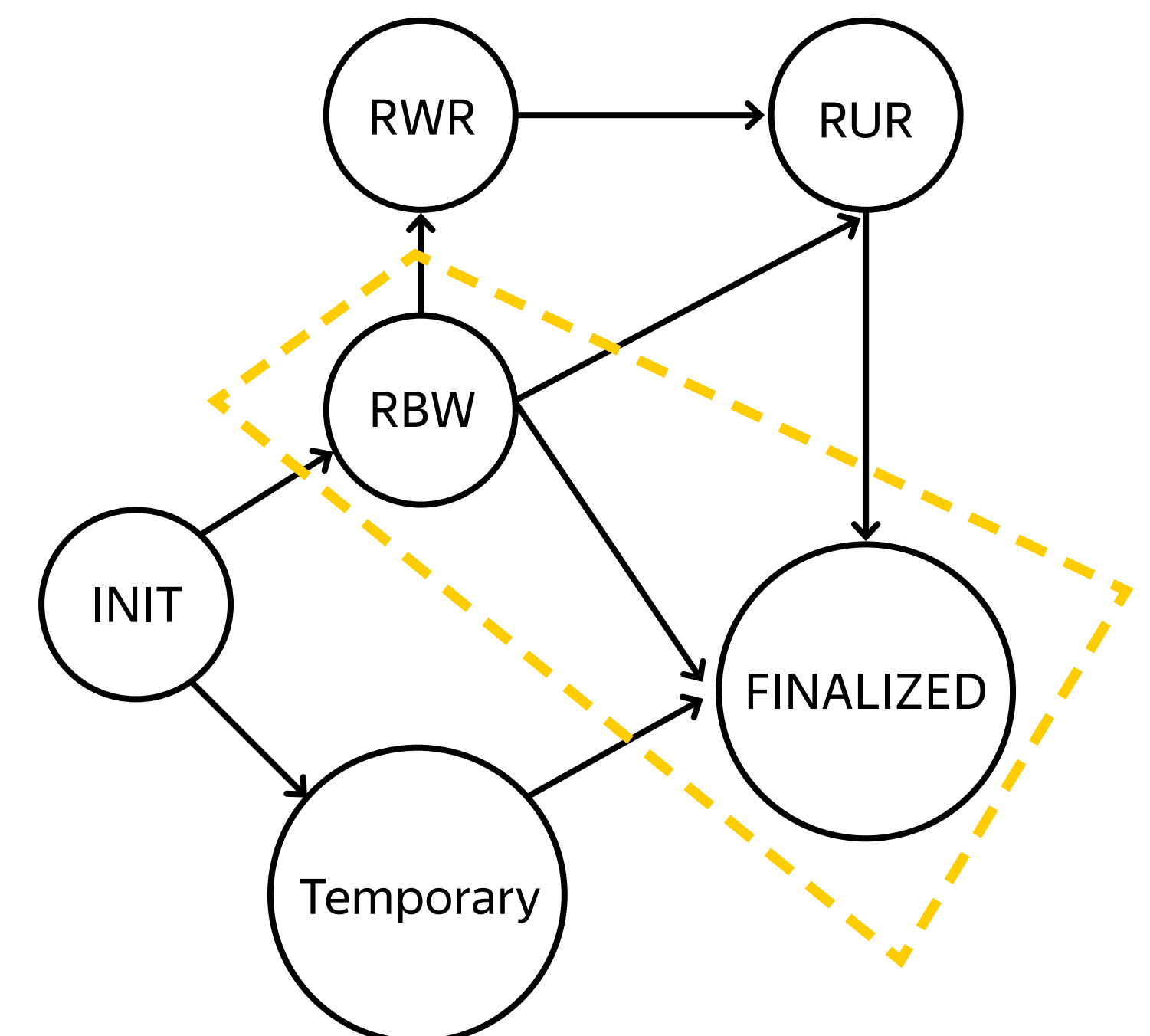
under_recovery

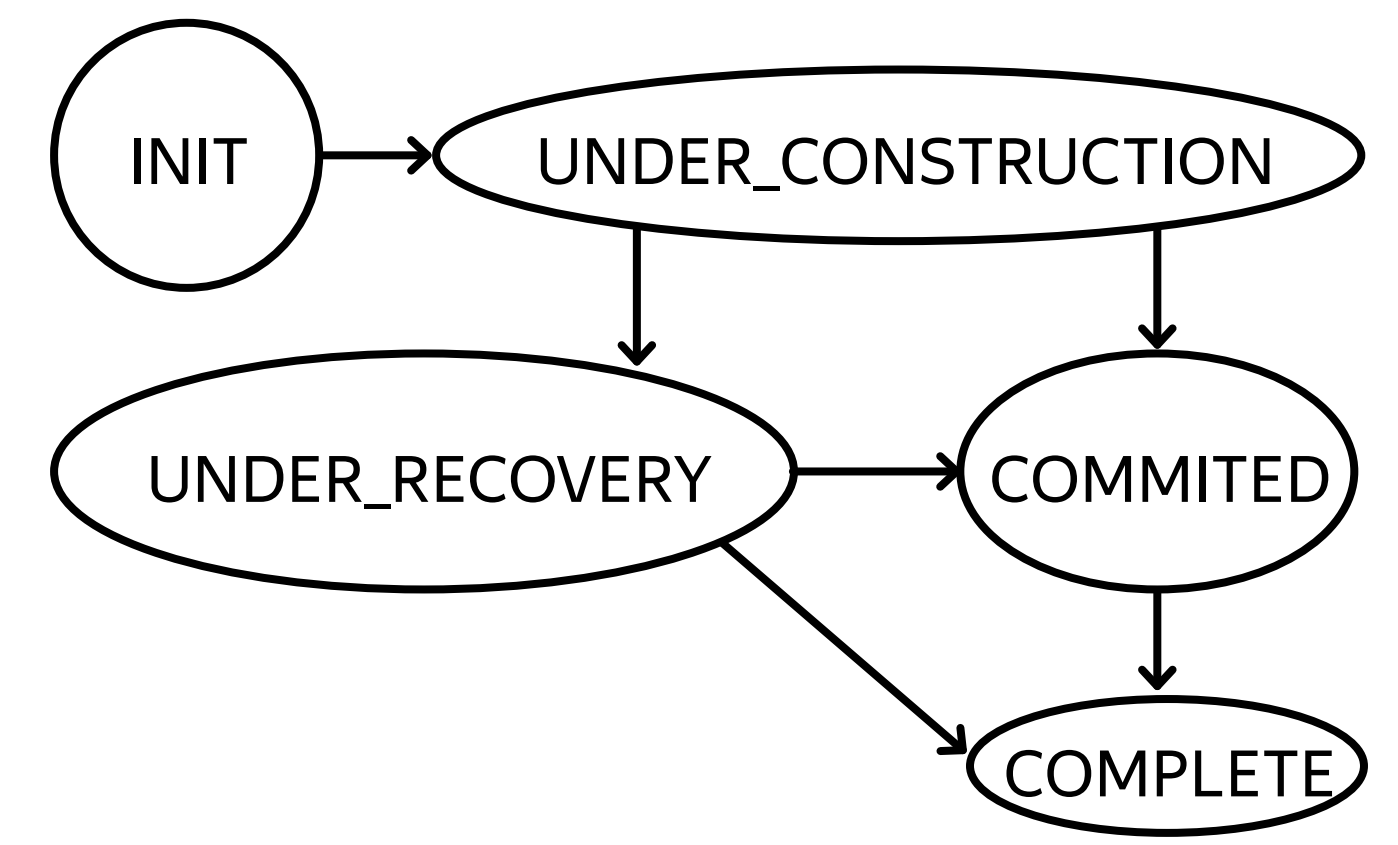
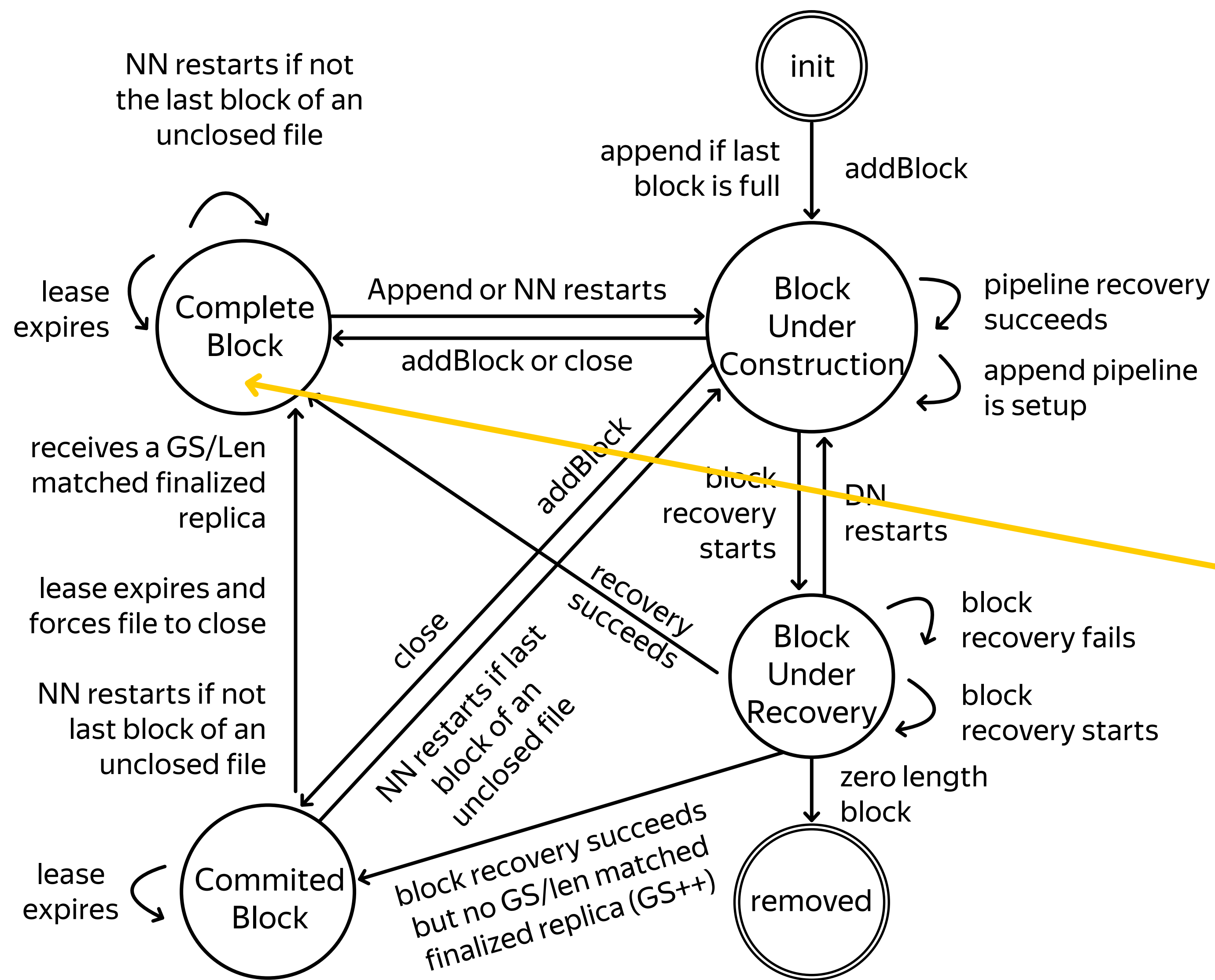




Simplified
Block State
Transition

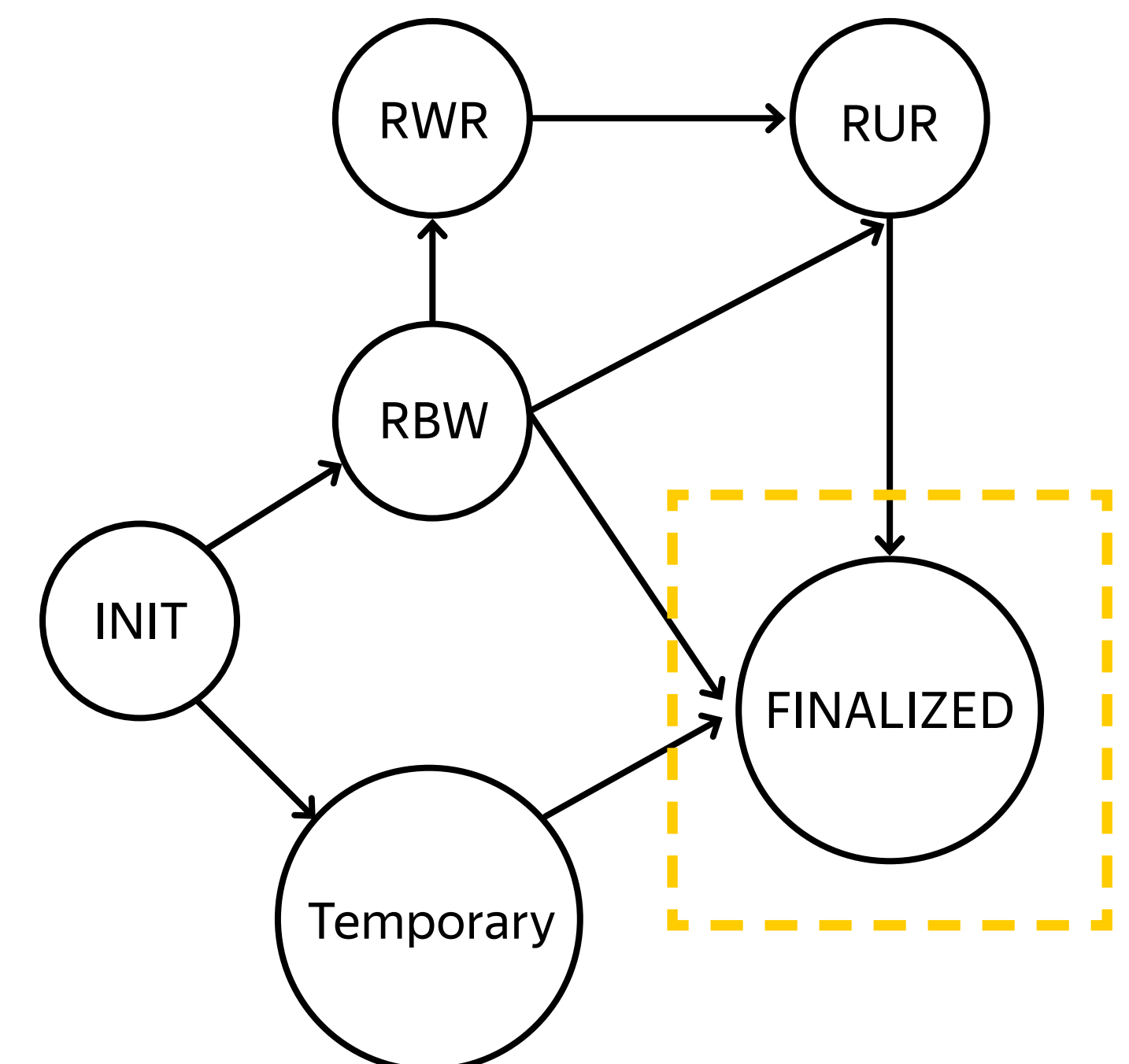
committed





Simplified
Block State
Transition

complete



Recovery

Recovery

- › Replica Recovery

Recovery

- › Replica Recovery
- › Block Recovery

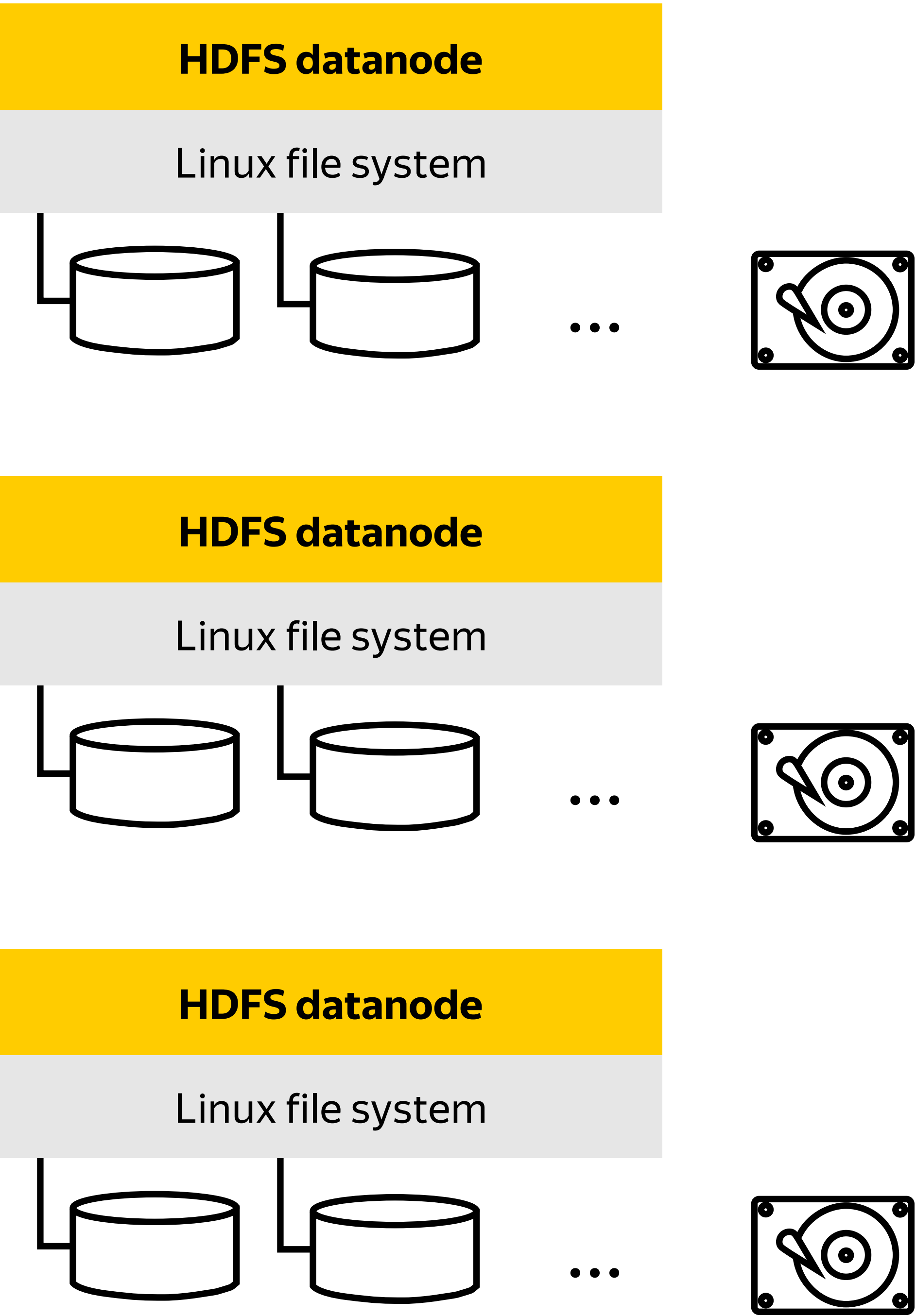
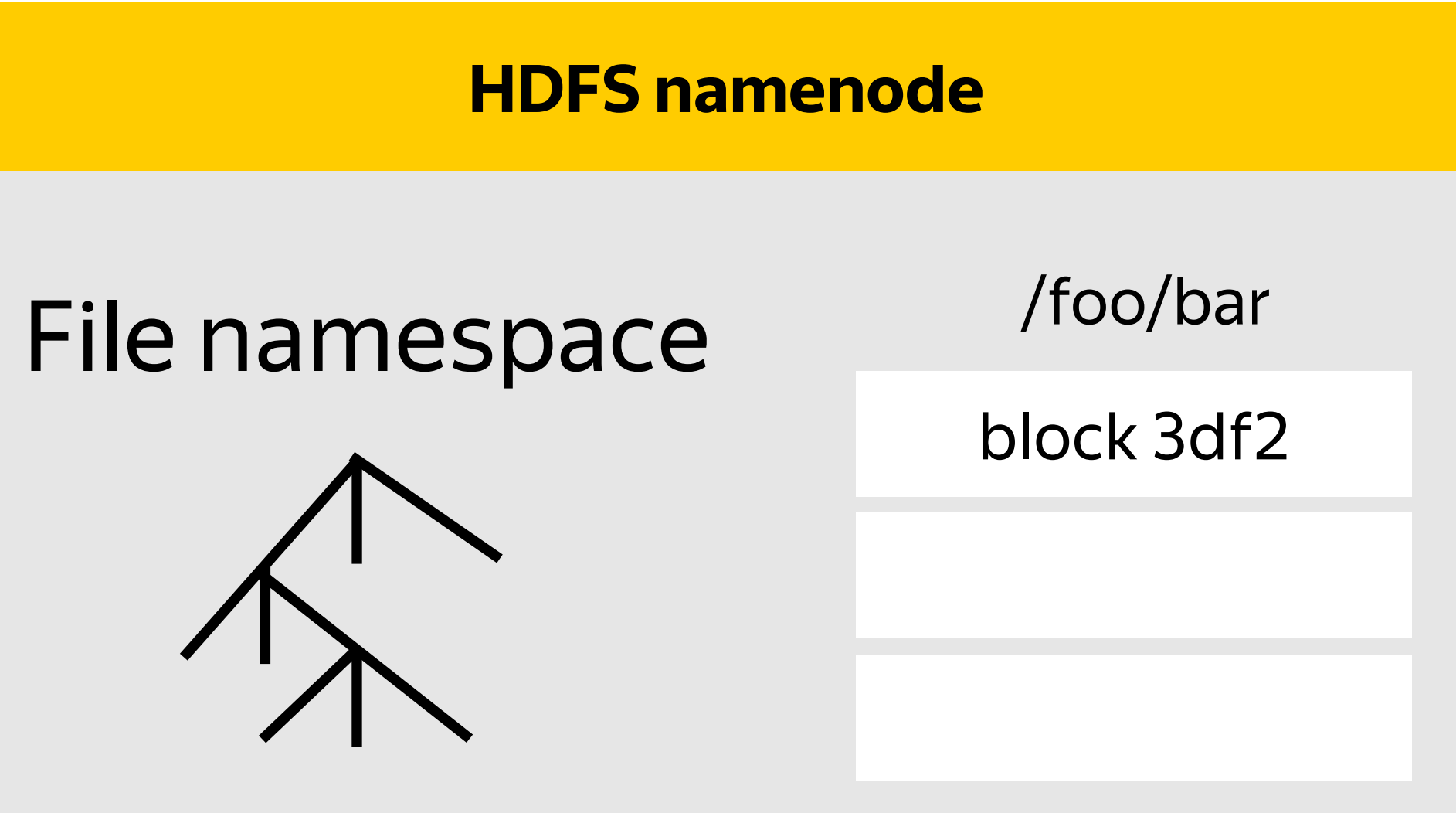
Recovery

- › Replica Recovery
- › Block Recovery
- › Lease Recovery

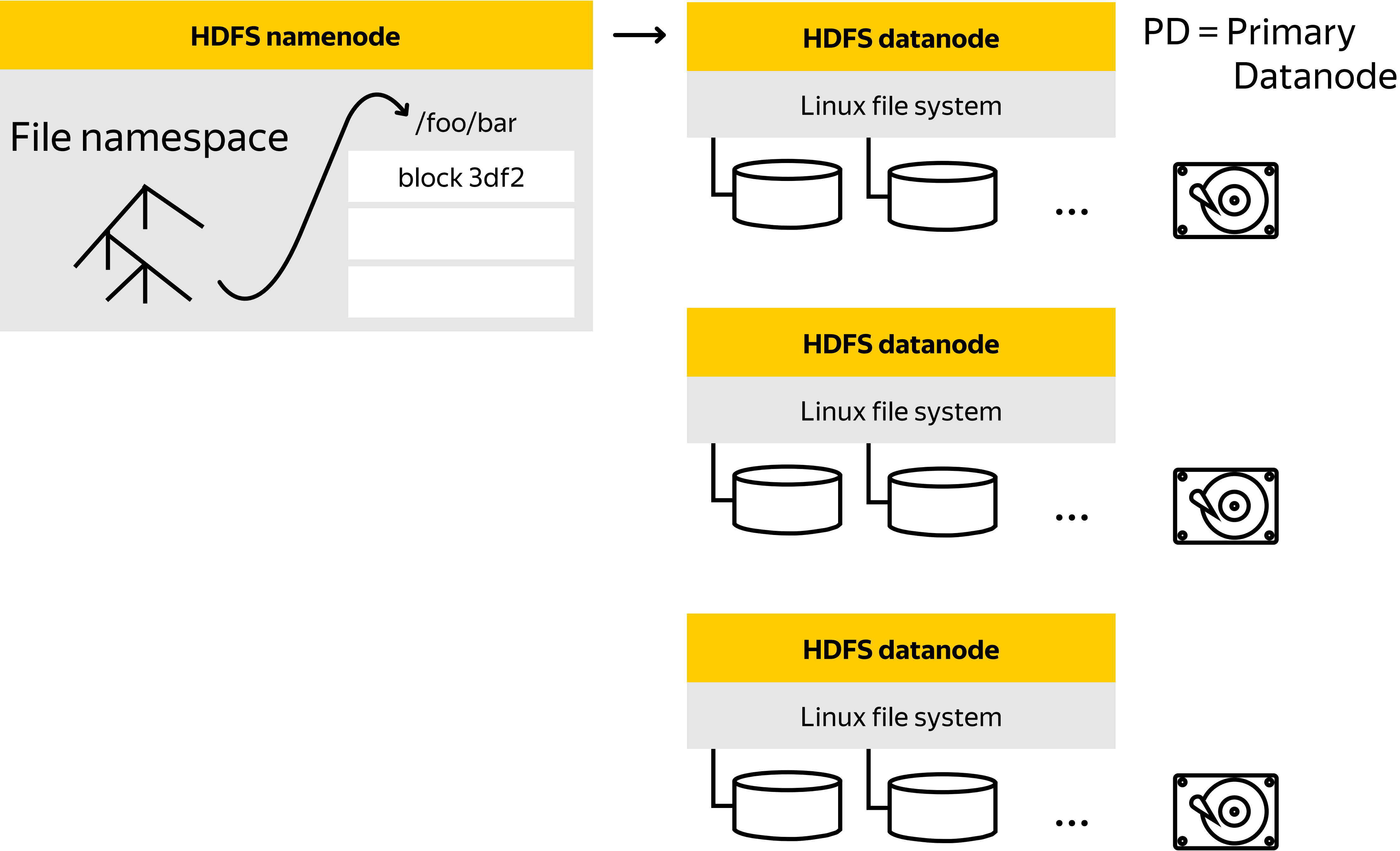
Recovery

- › Replica Recovery
- › Block Recovery
- › Lease Recovery
- › Pipeline Recovery

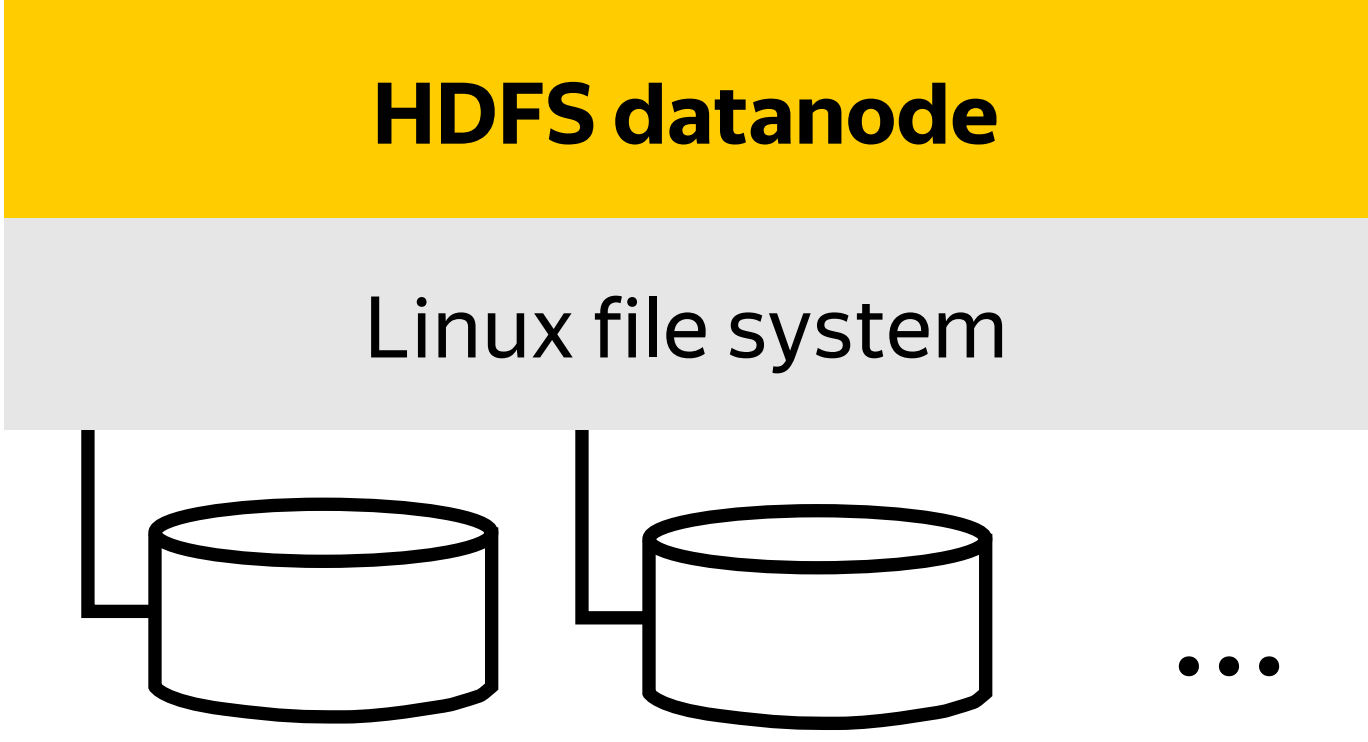
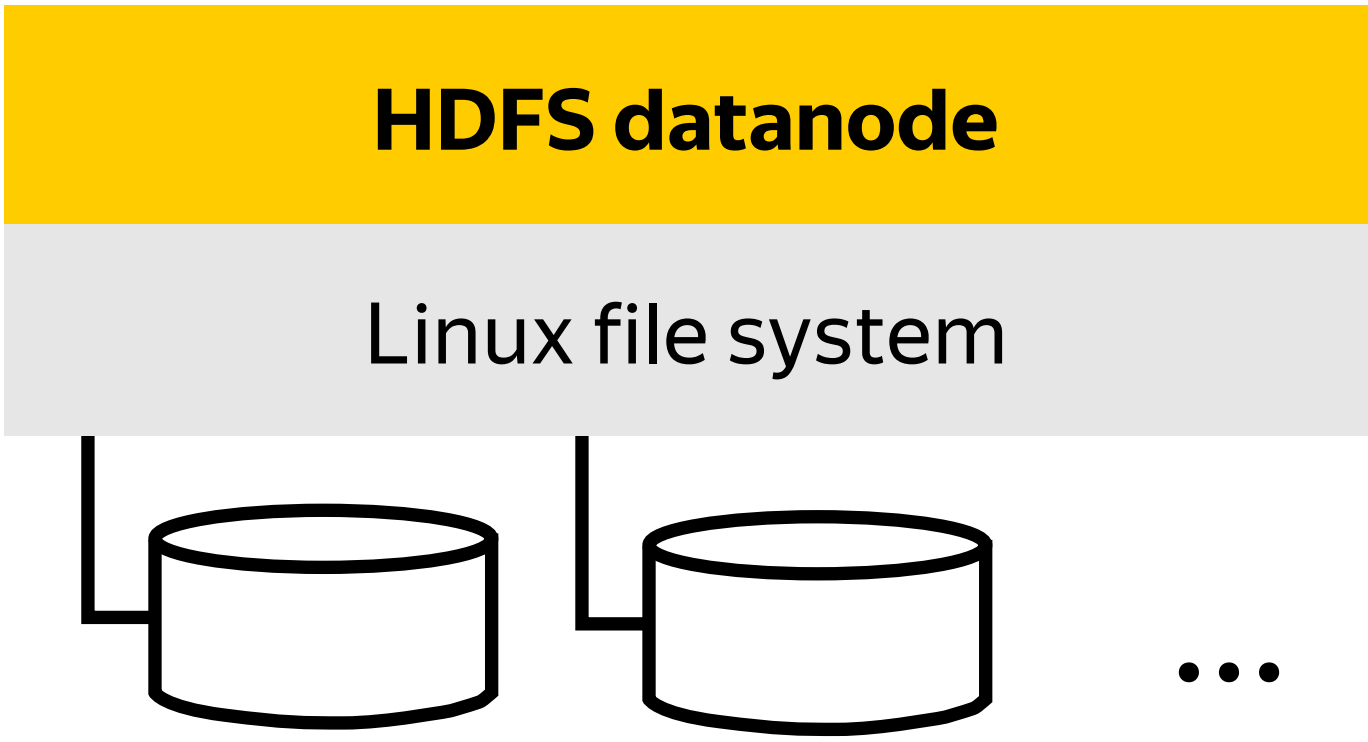
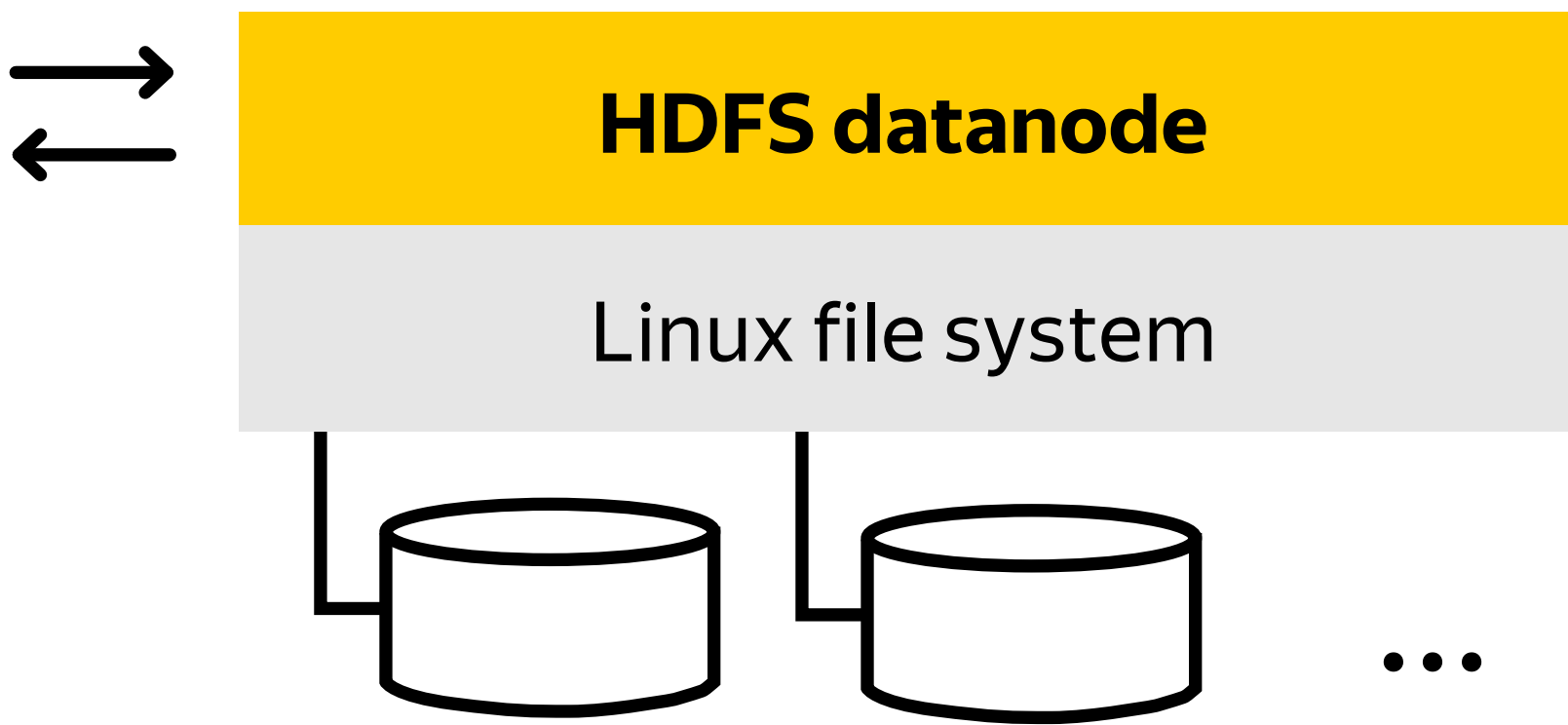
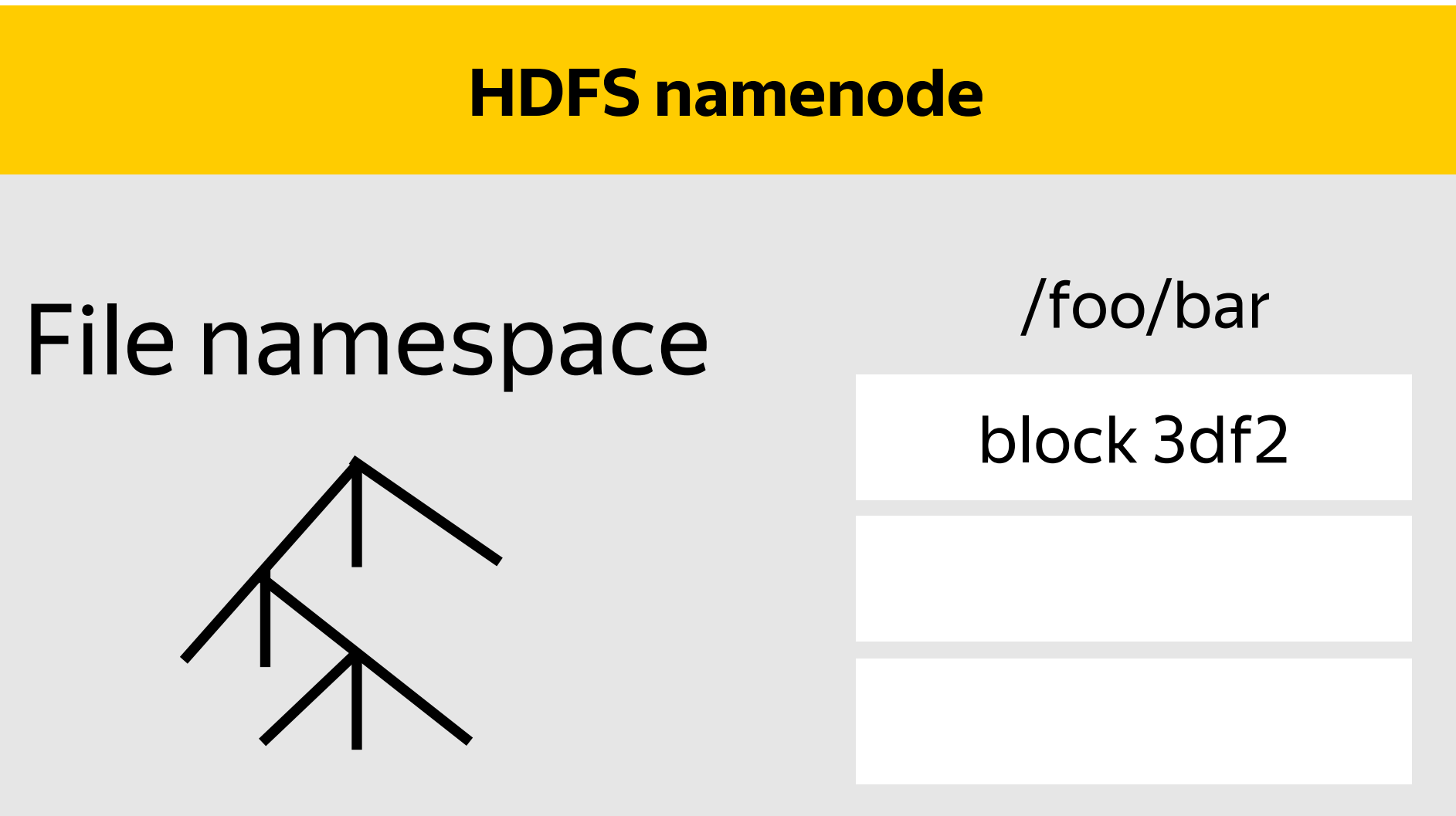
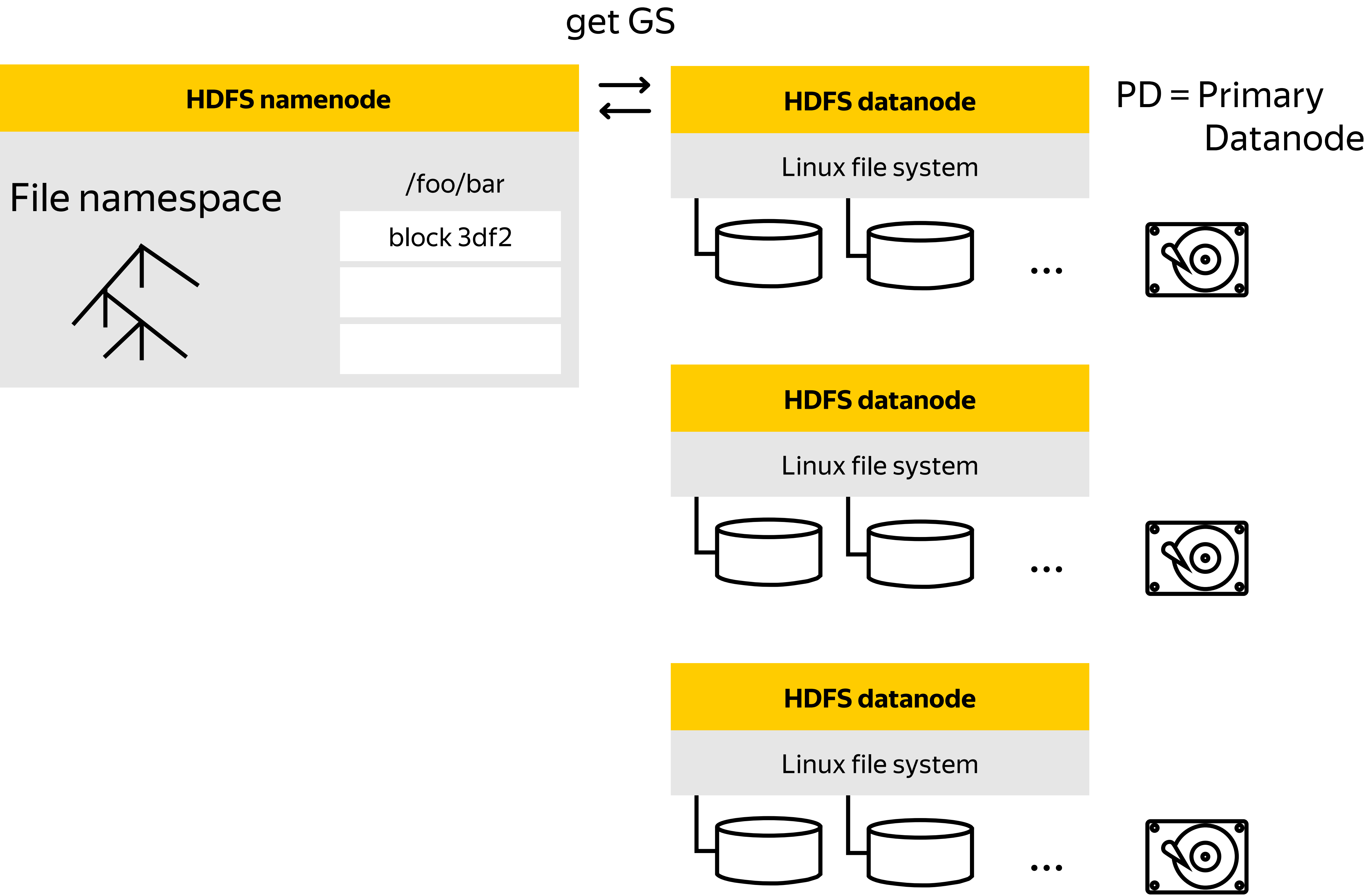
Block Recovery



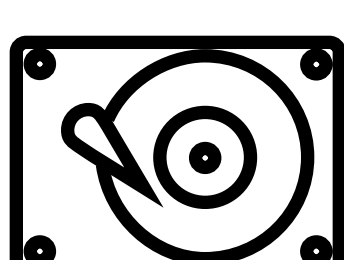
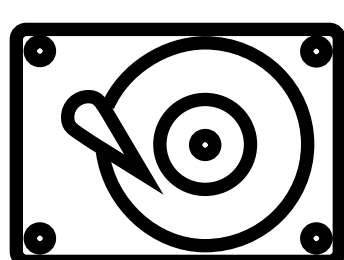
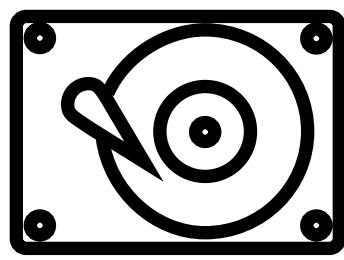
Block Recovery



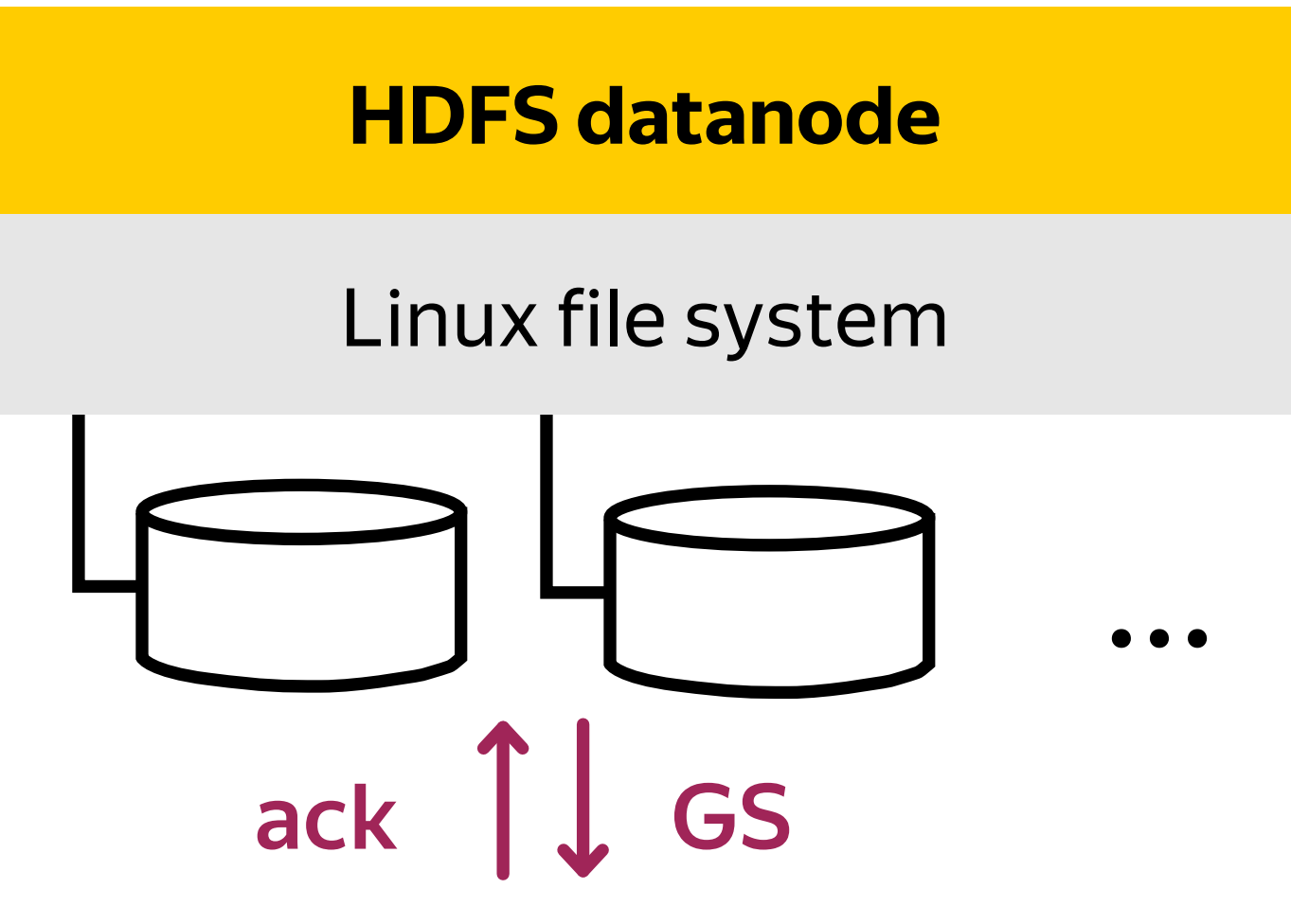
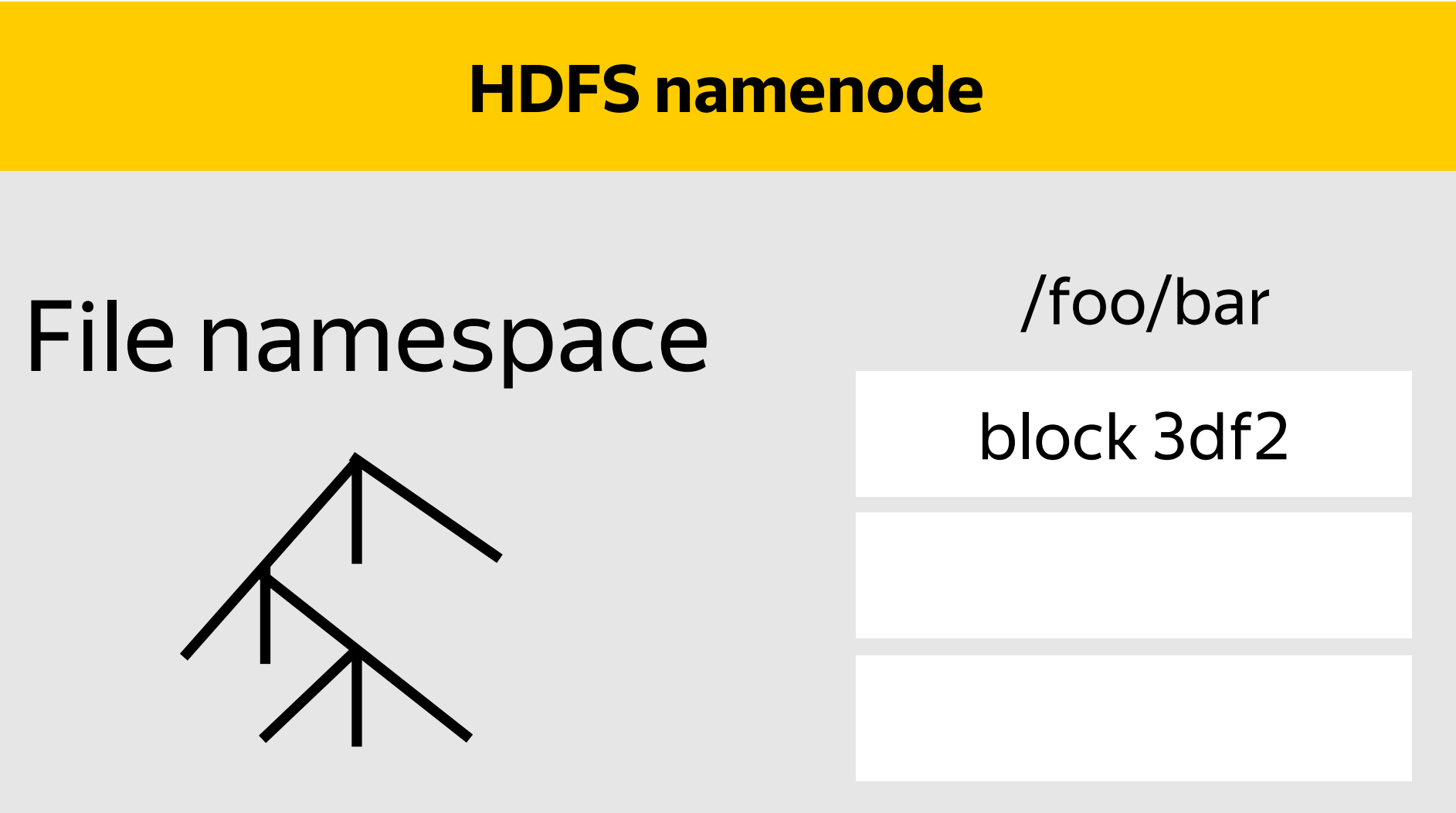
Block Recovery



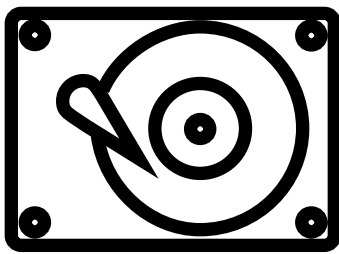
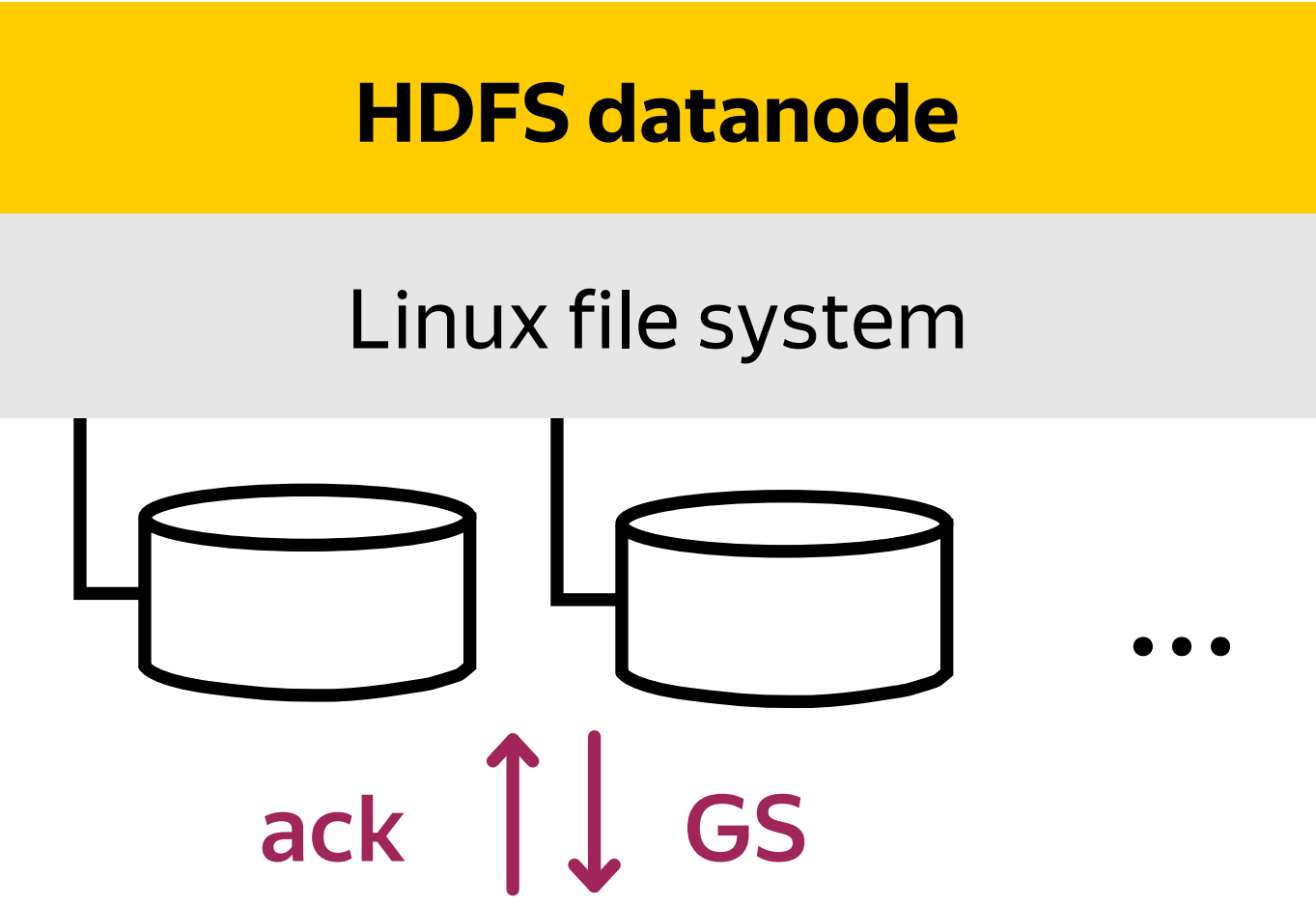
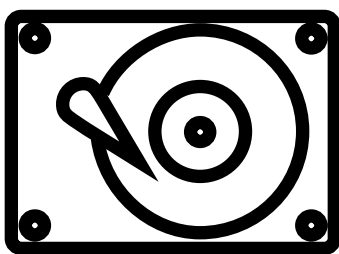
PD = Primary Datanode



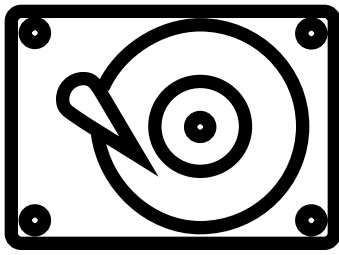
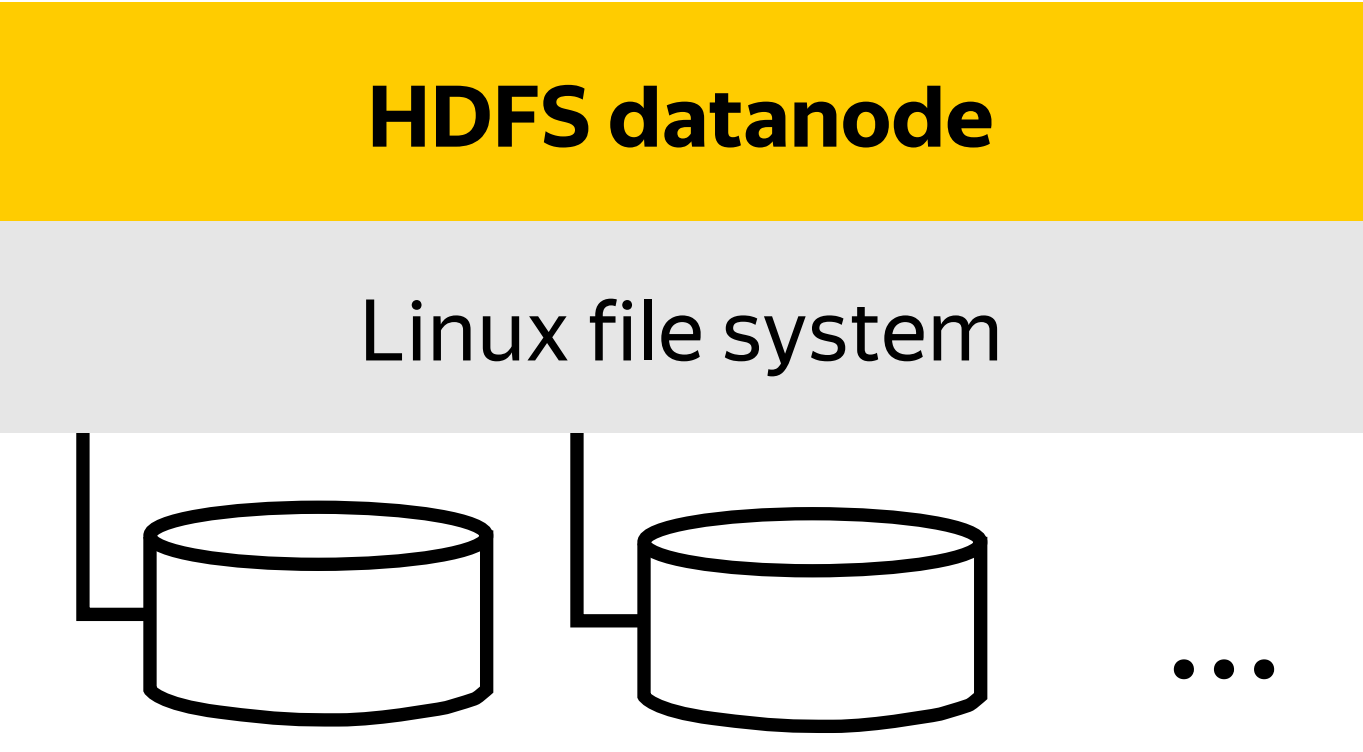
Block Recovery



PD = Primary
Datanode

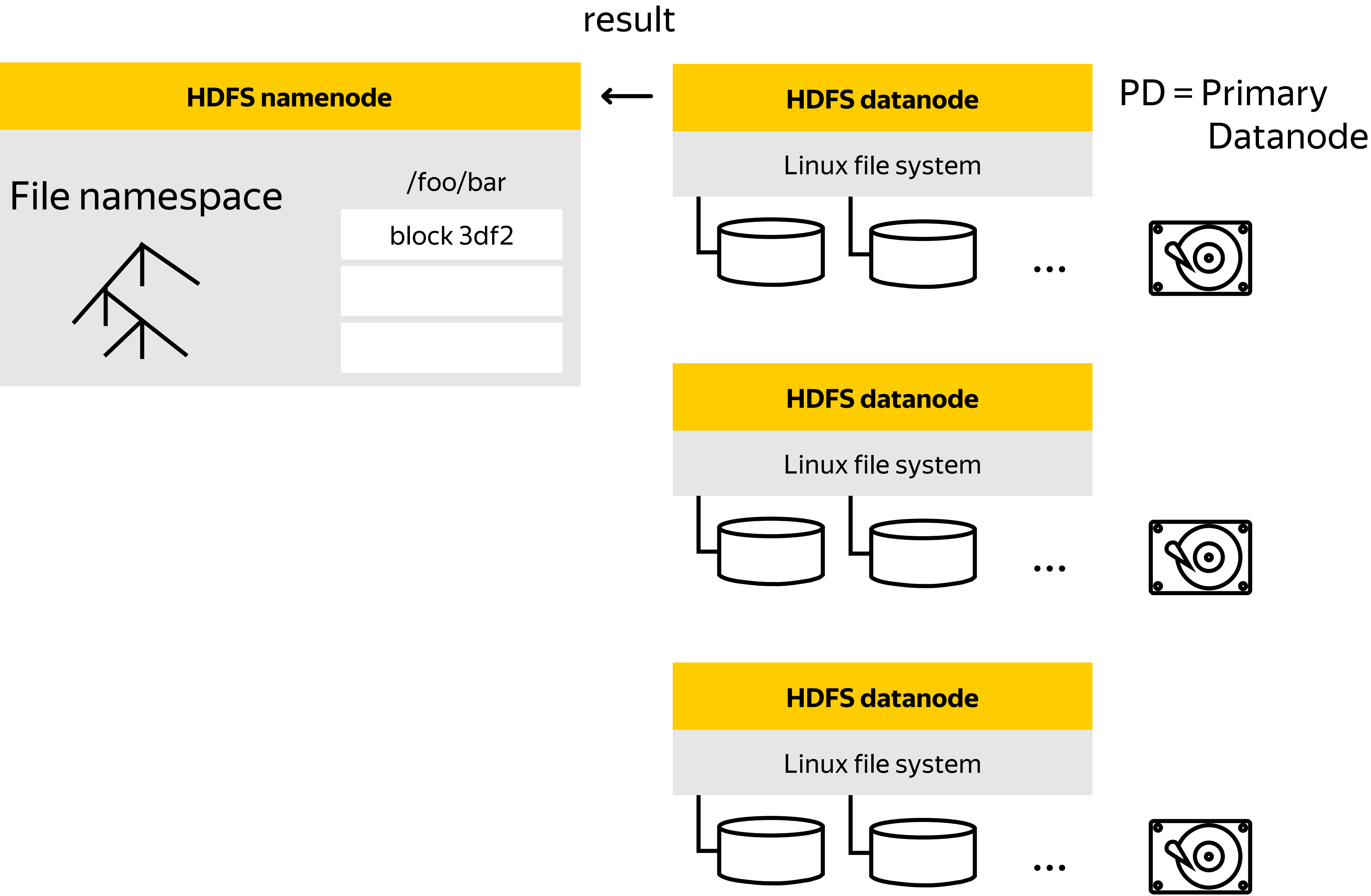


replica
recovery



replica
recovery

Block Recovery



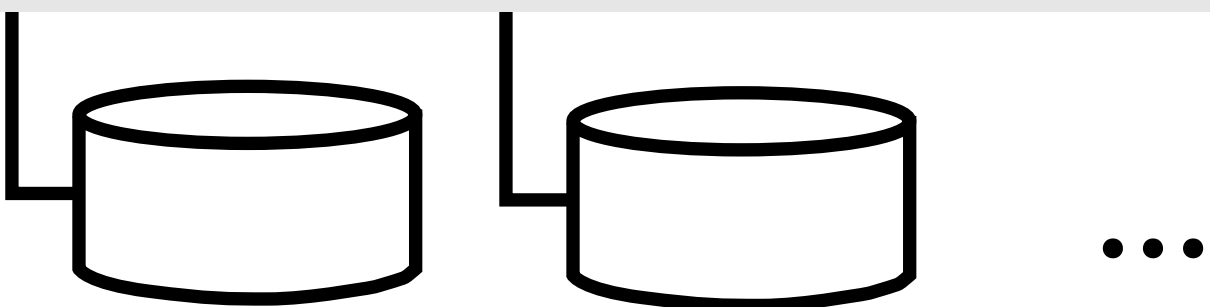
File namespace

/foo/bar

block 3df2

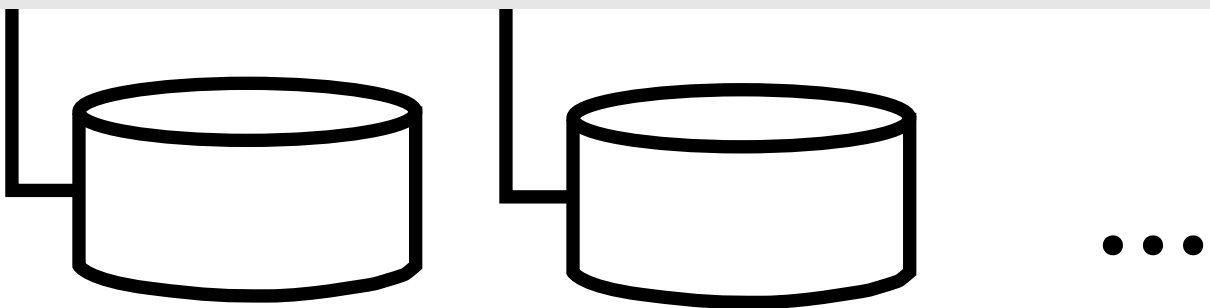
HDFS datanode

Linux file system



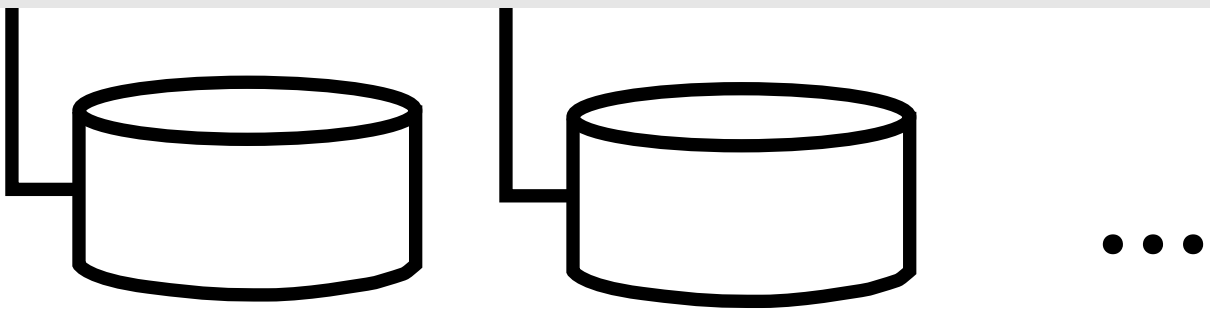
HDFS datanode

Linux file system

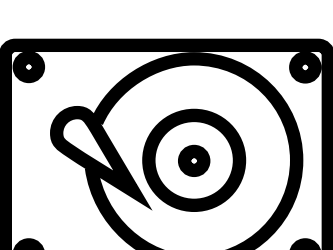
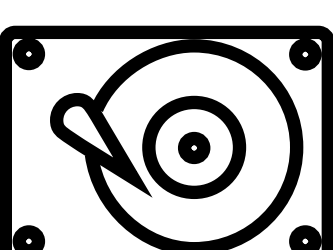
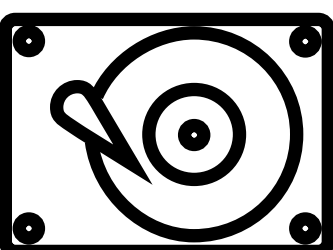


HDFS datanode

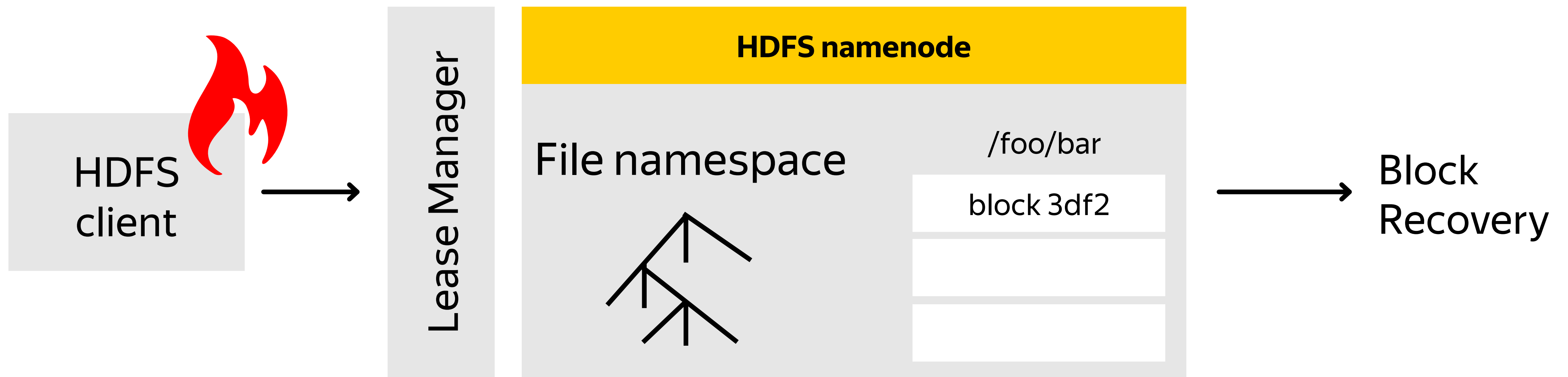
Linux file system



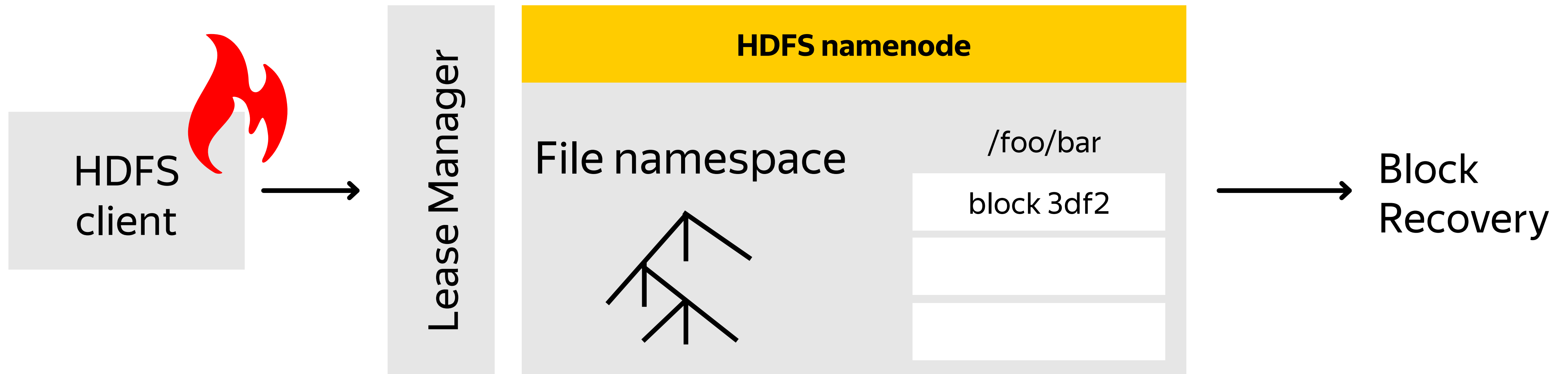
PD = Primary Datanode



Lease Recovery

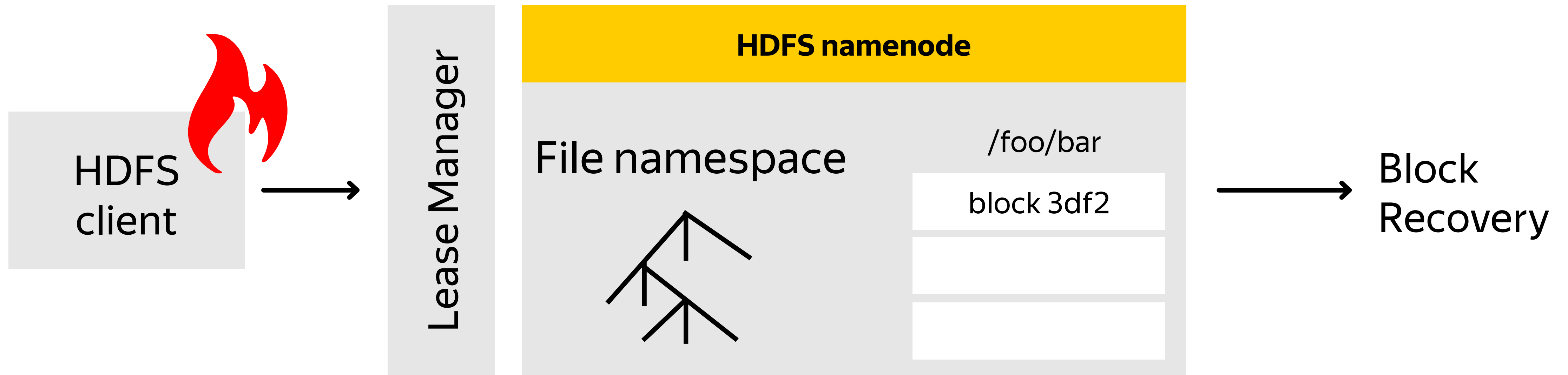


Lease Recovery



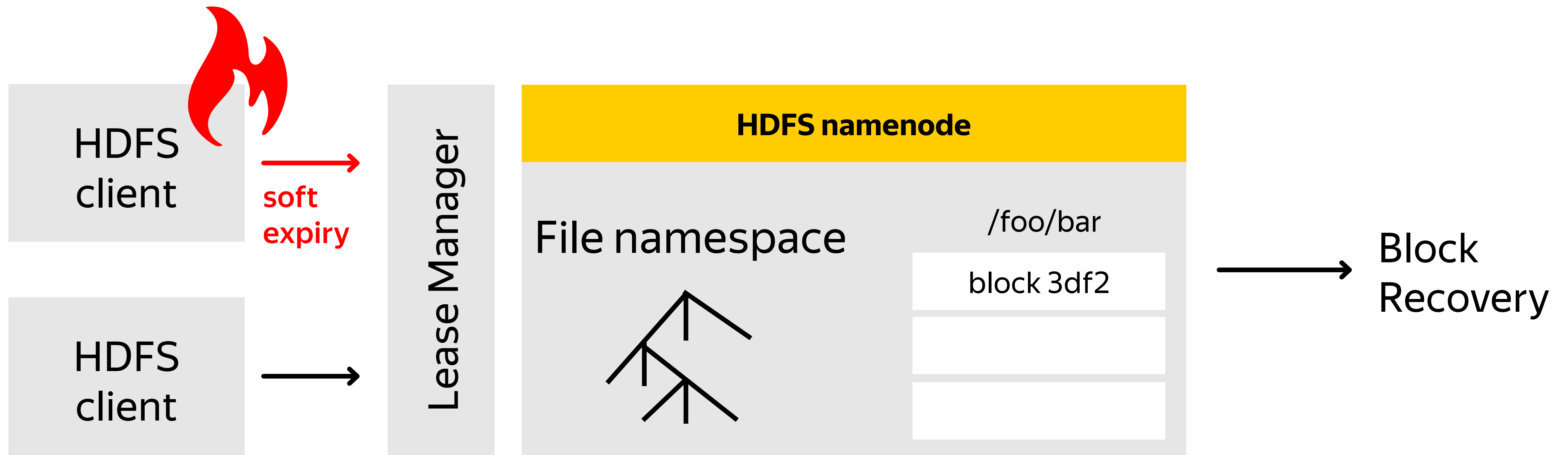
user₁, /path/1, lease
user₁, /path/2, lease
user₂, /path/3, lease
user₃, /path/4, lease
...

Lease Recovery

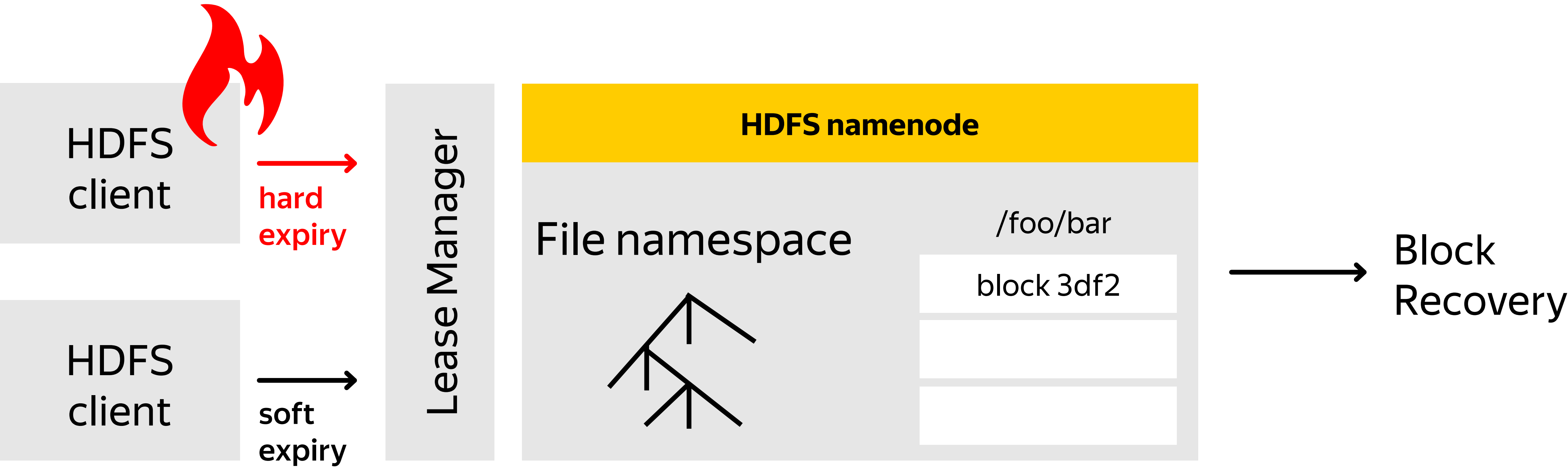


user₁, /path/1, lease **(soft + hard)**
user₁, /path/2, lease (soft + hard)
user₂, /path/3, lease (soft + hard)
user₃, /path/4, lease (soft + hard)
...

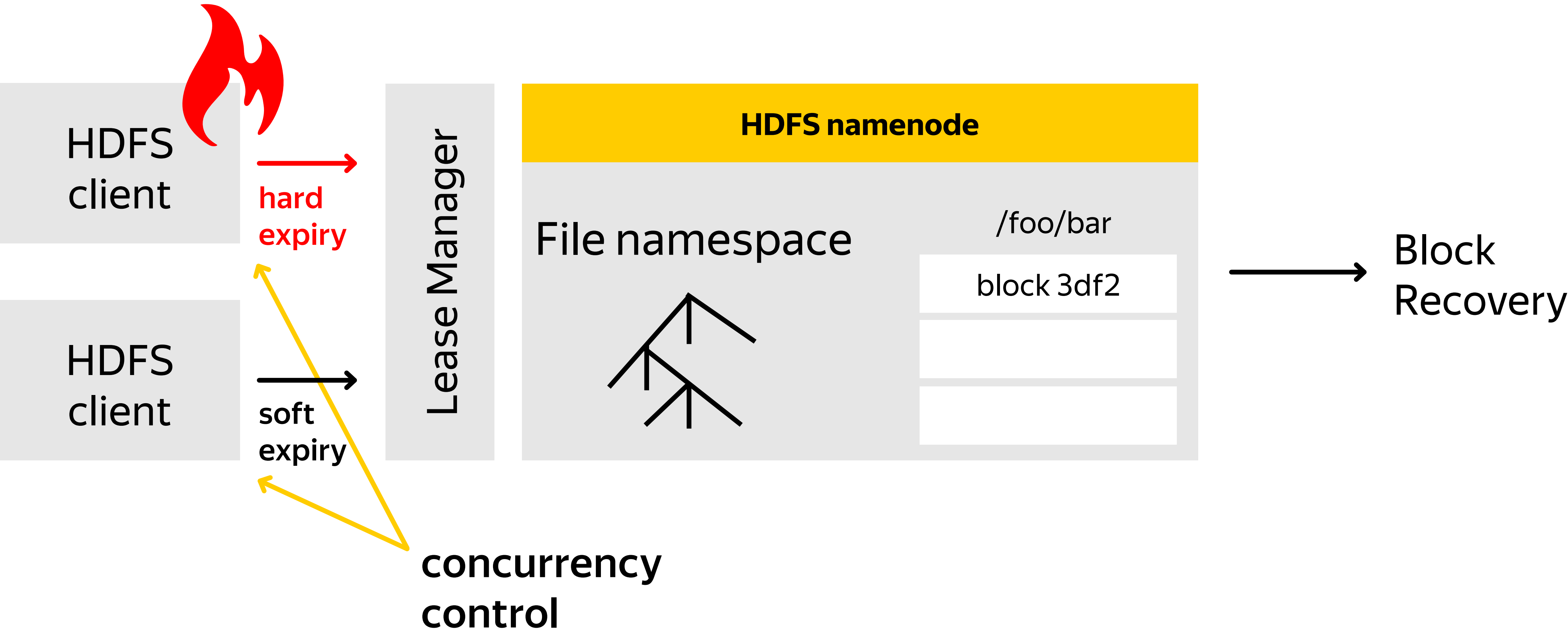
Lease Recovery



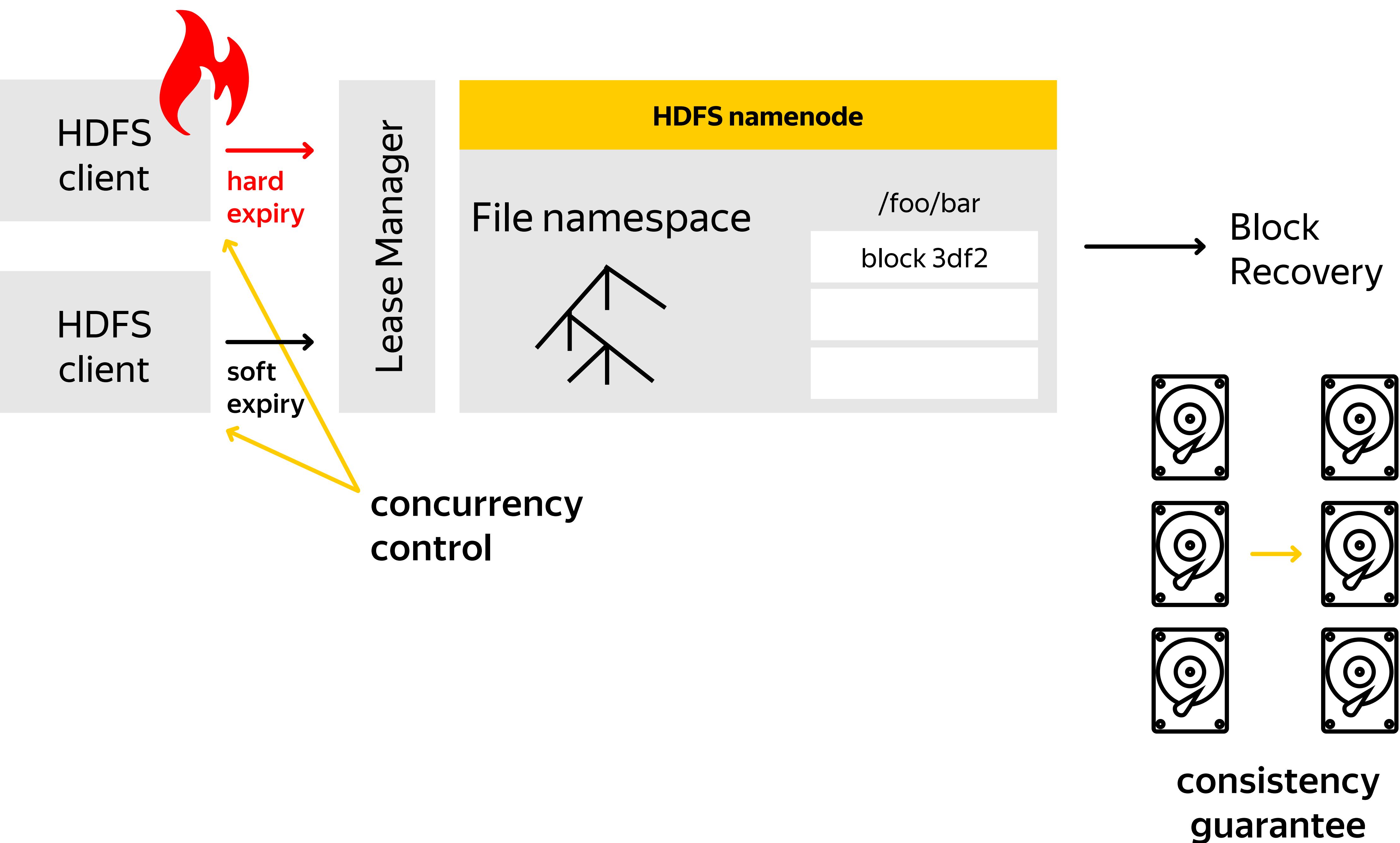
Lease Recovery



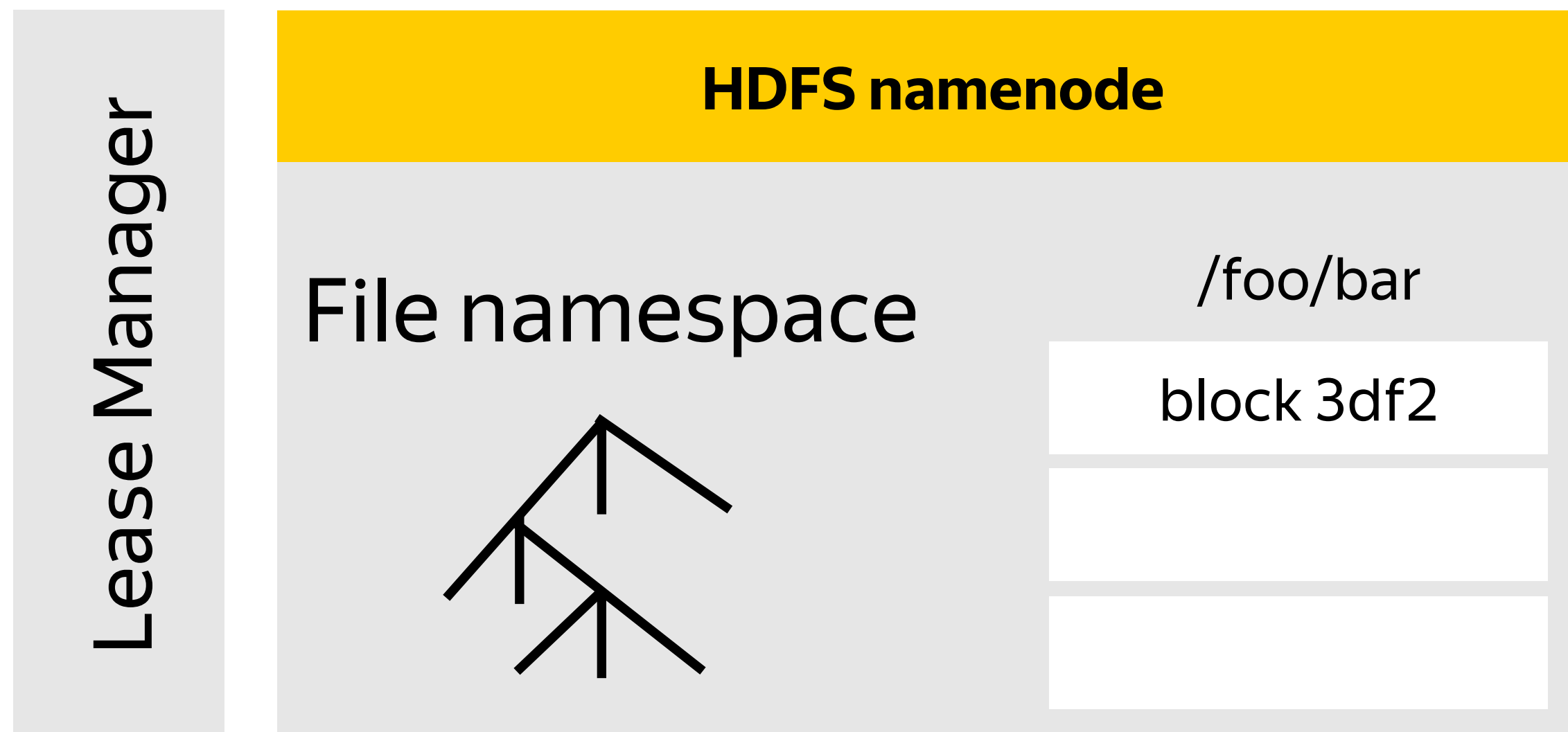
Lease Recovery



Lease Recovery



Lease Recovery

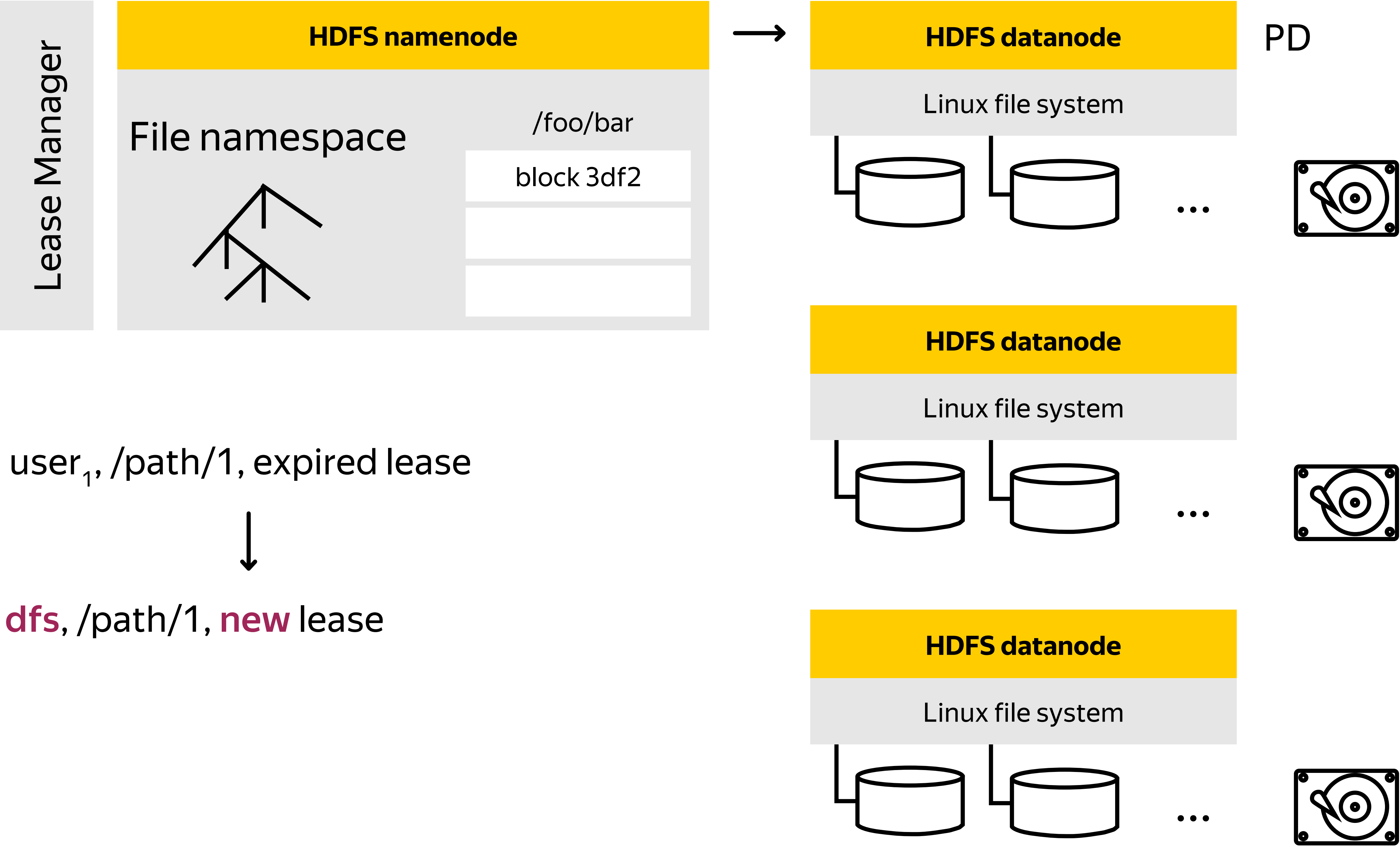


user1, /path/1, expired lease

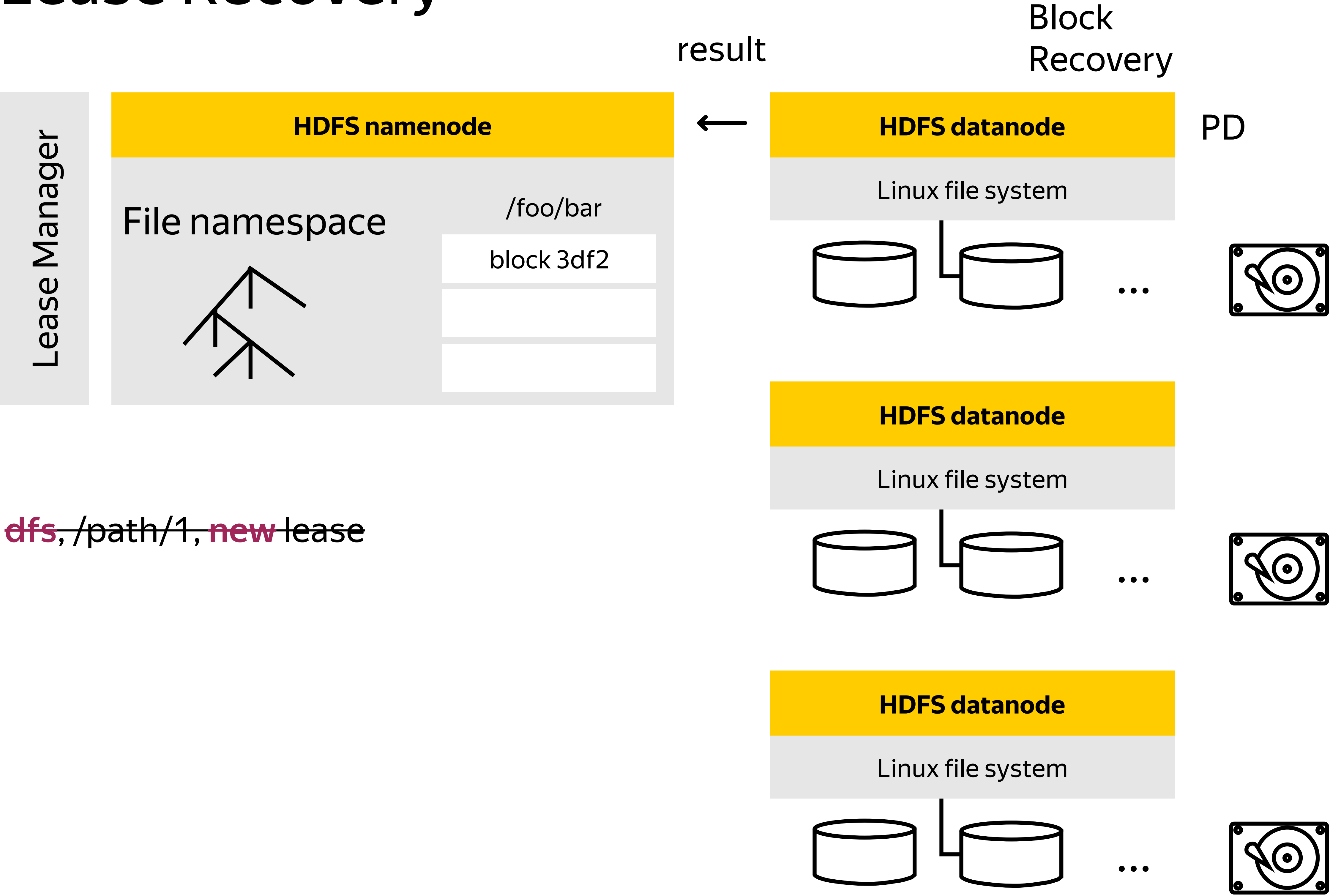


dfs, /path/1, **new** lease

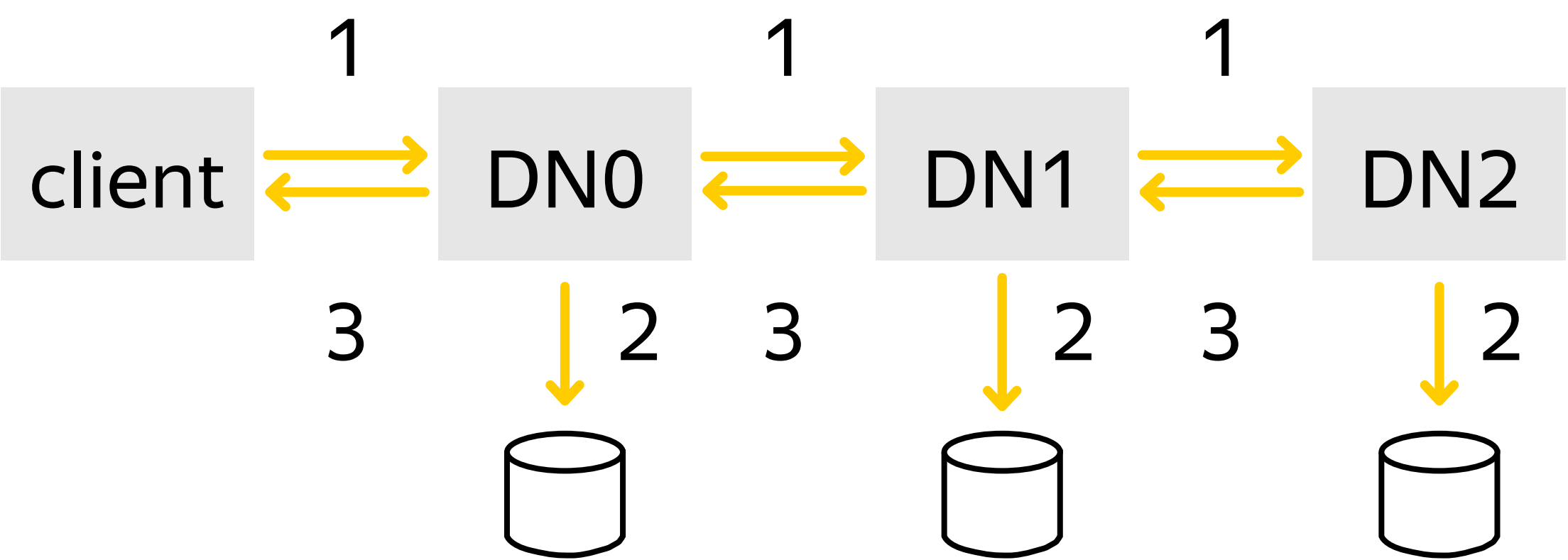
Lease Recovery



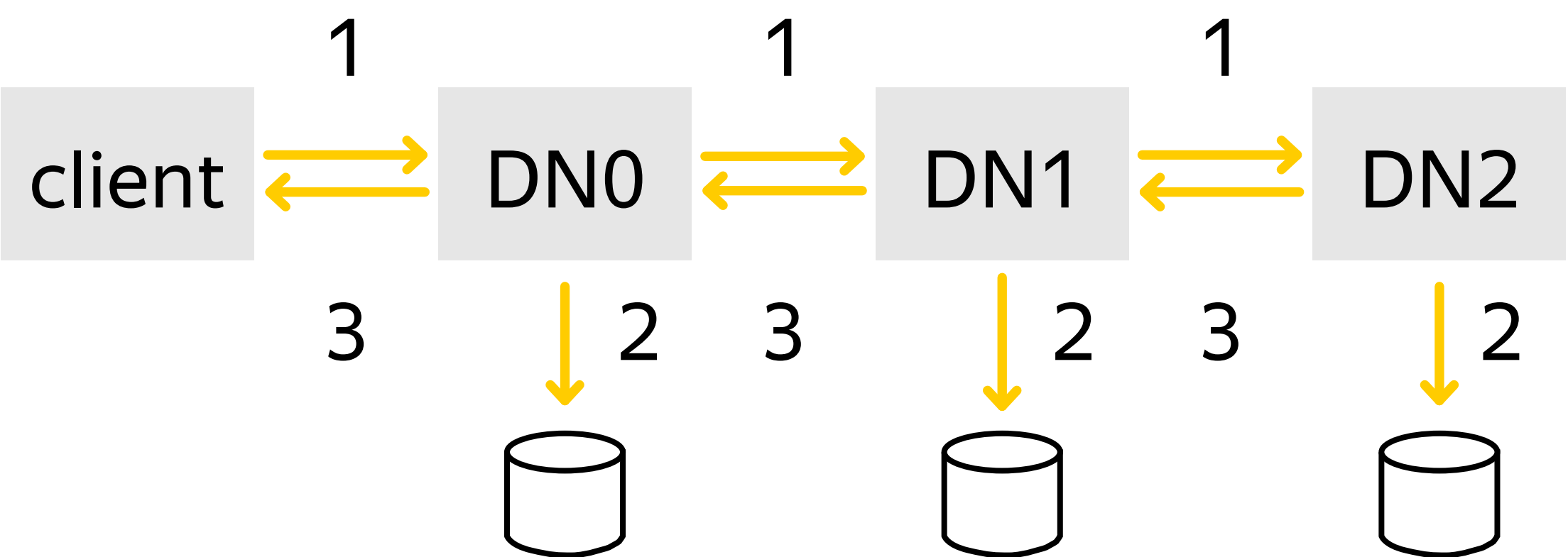
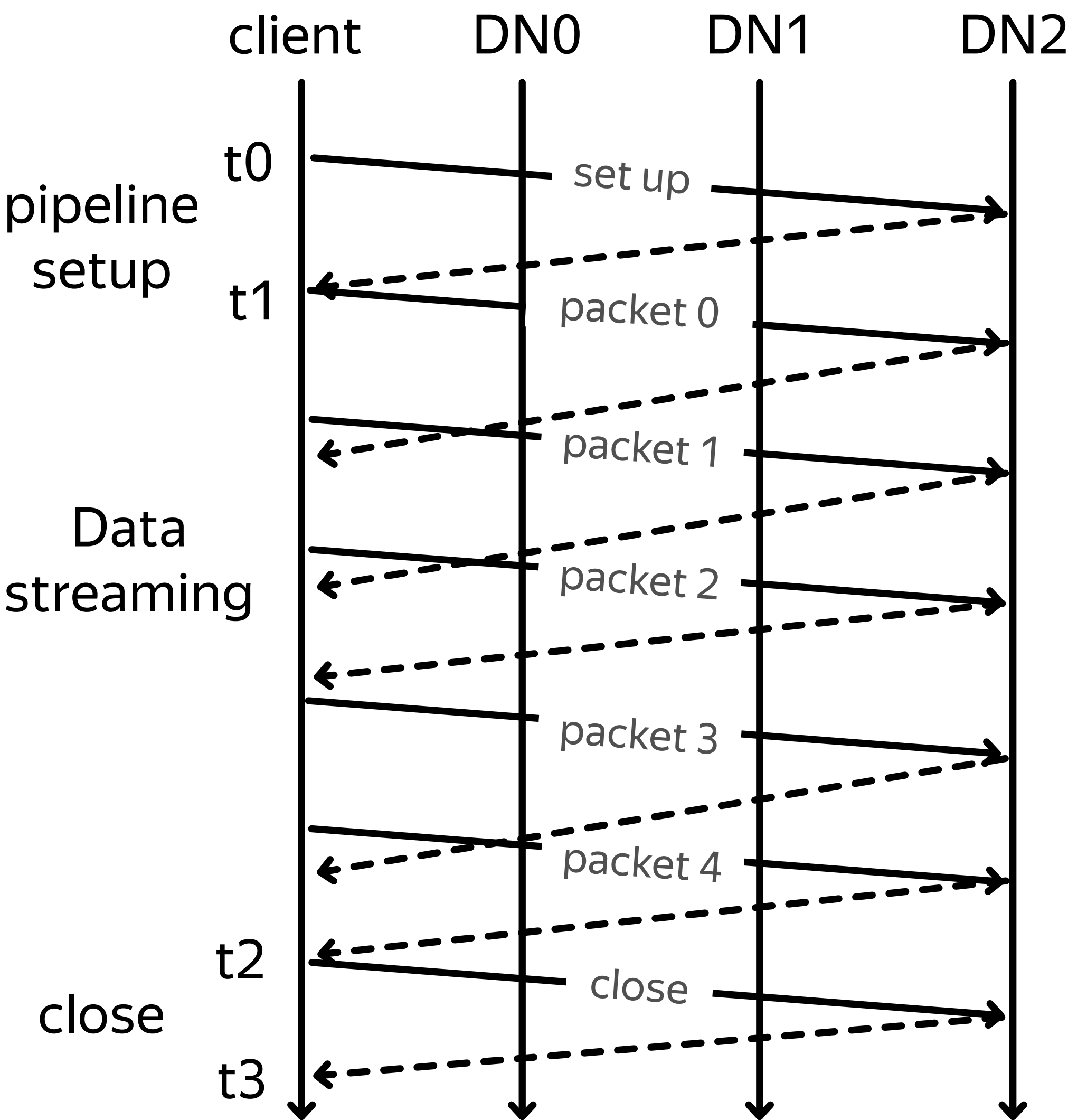
Lease Recovery



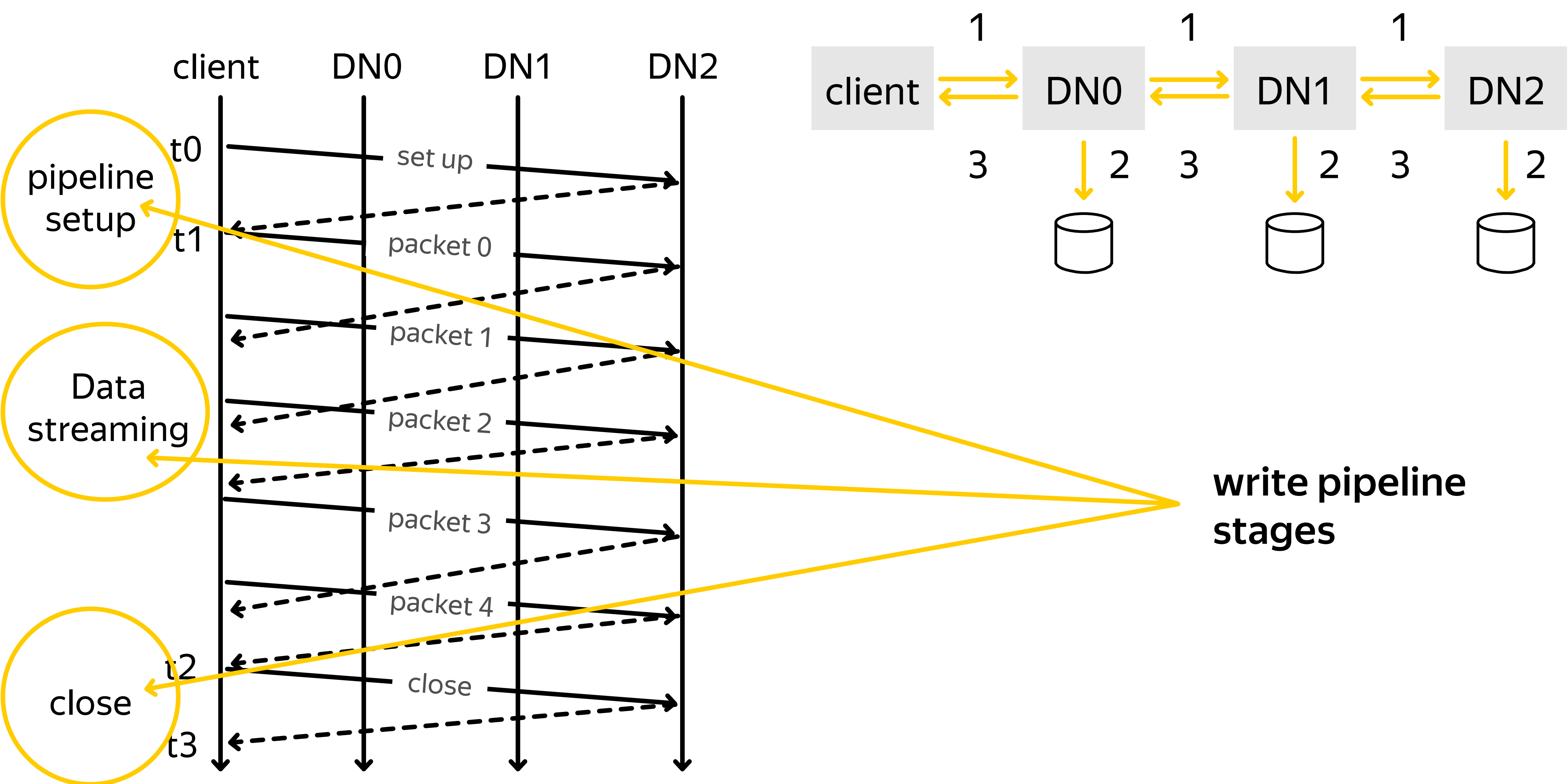
Pipeline



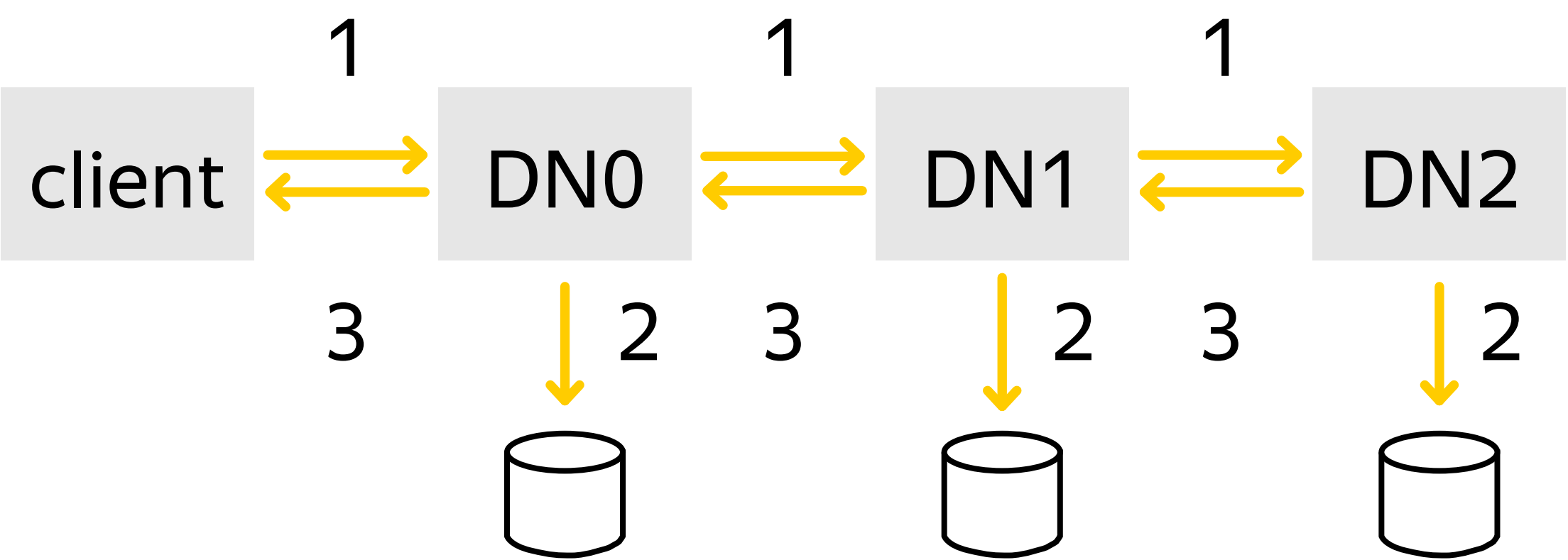
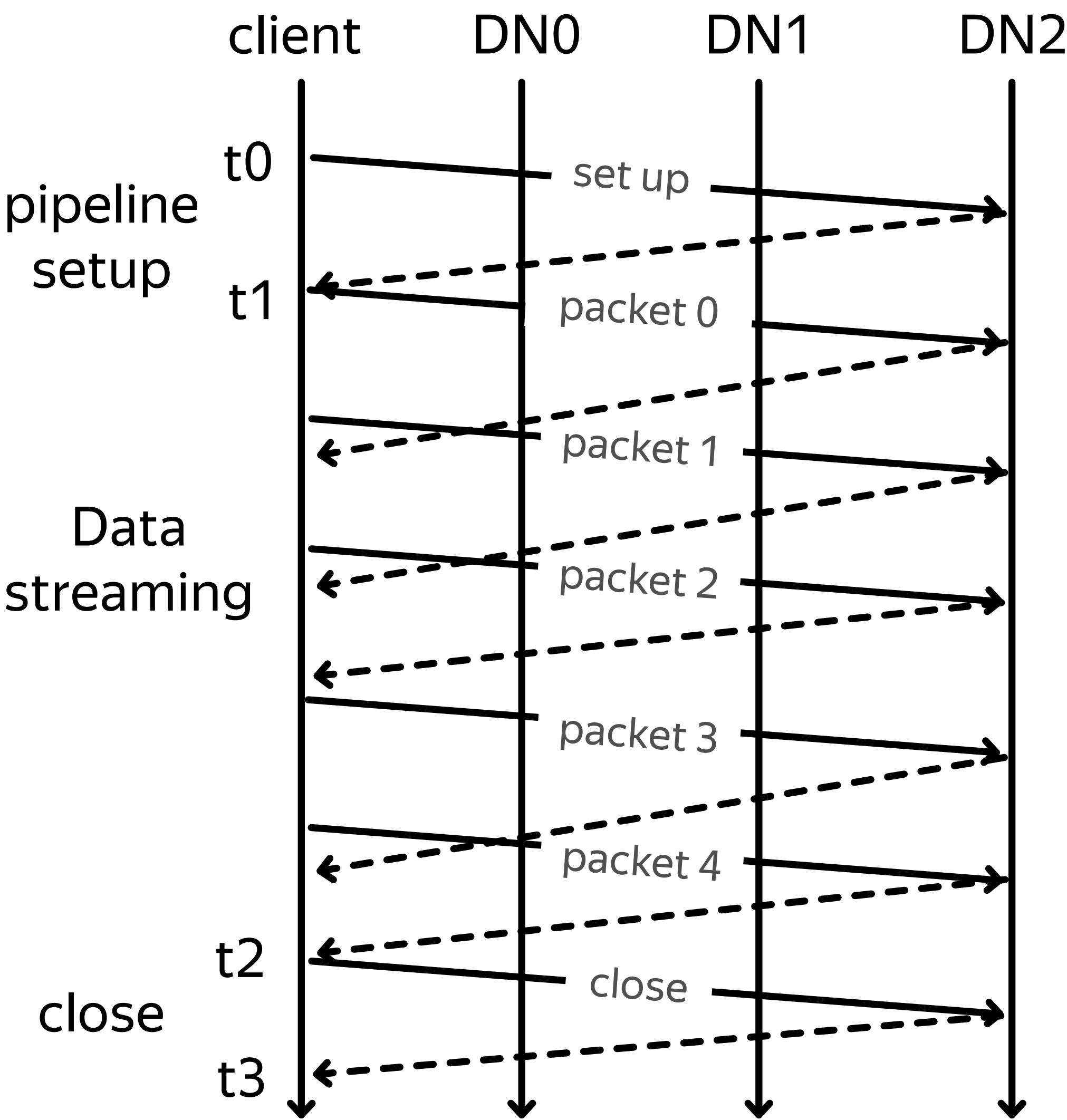
Pipeline



Pipeline

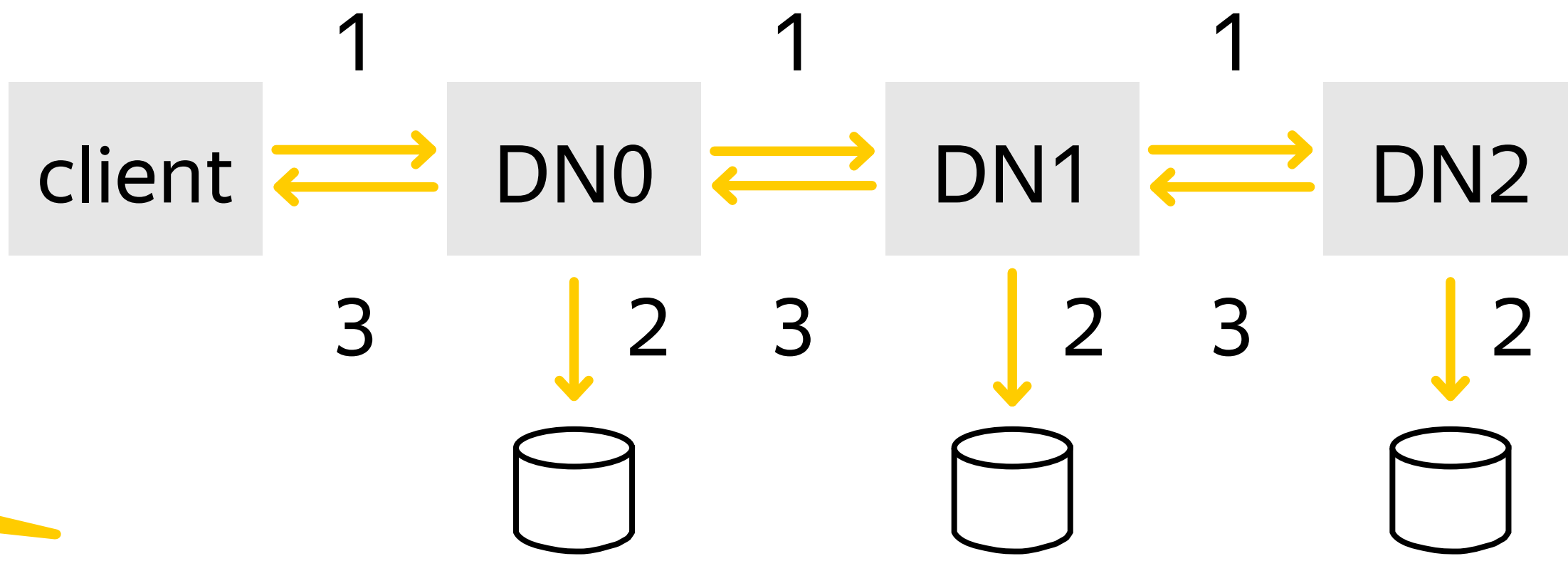
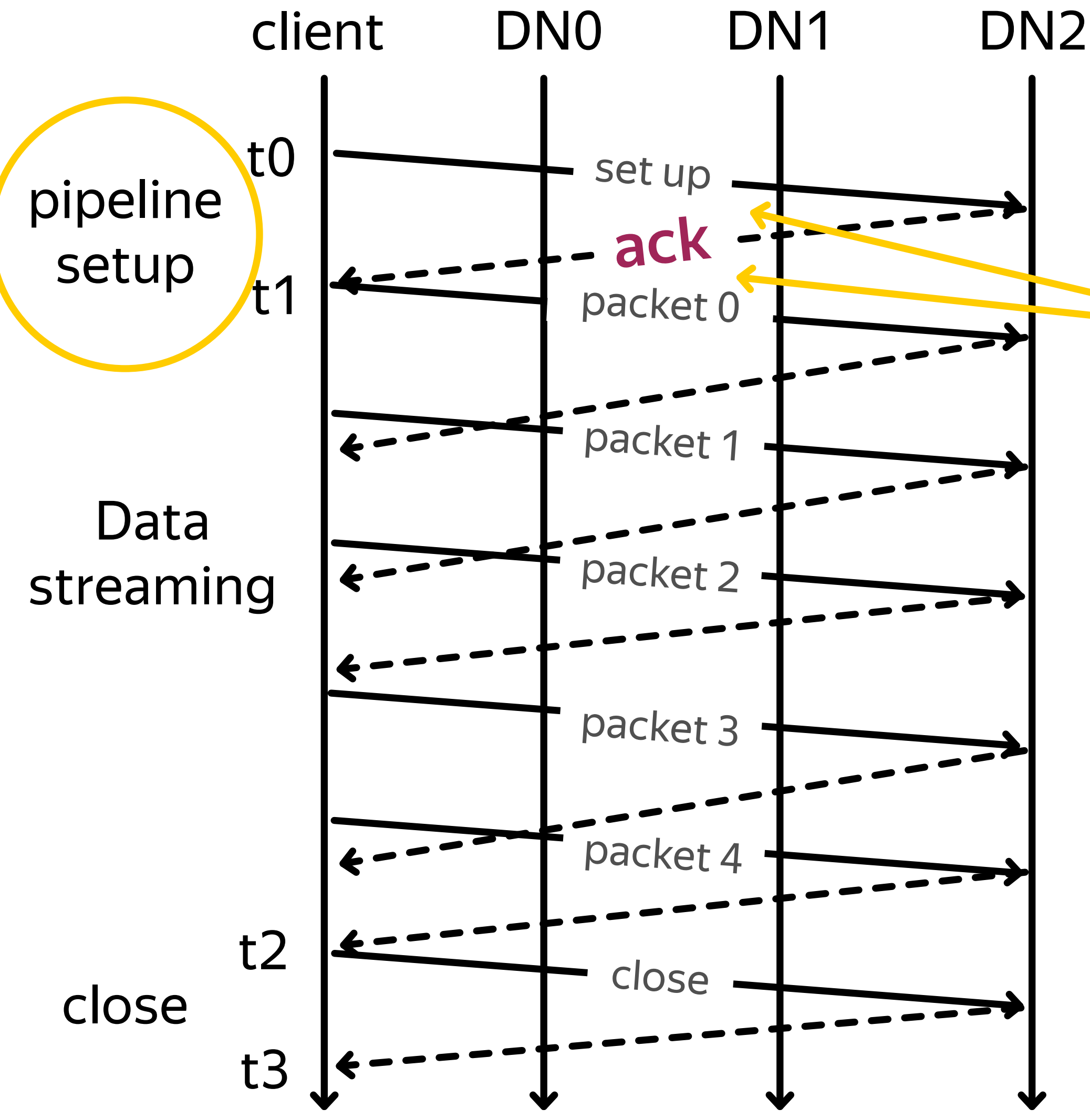


Pipeline

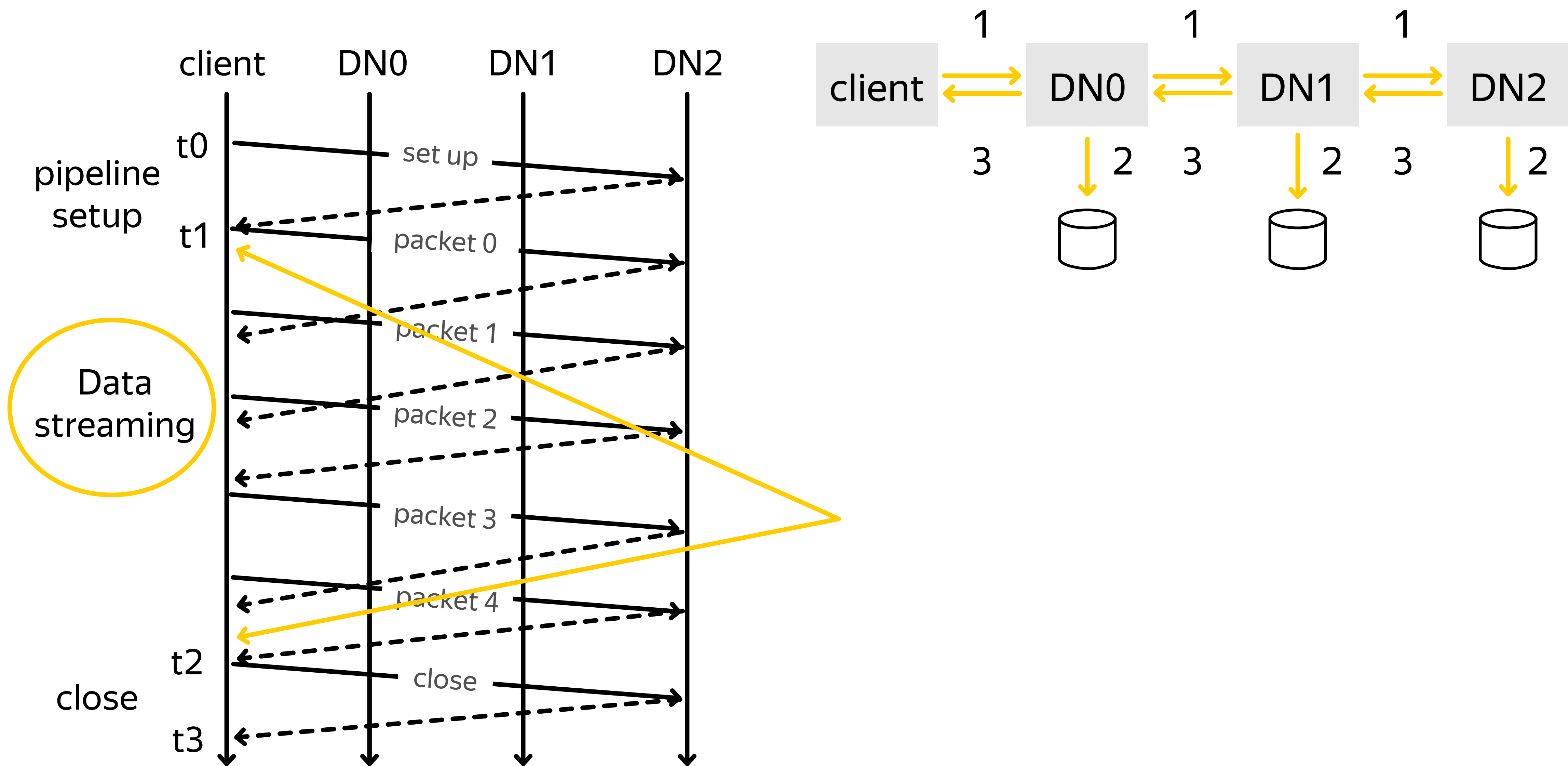


Pipeline

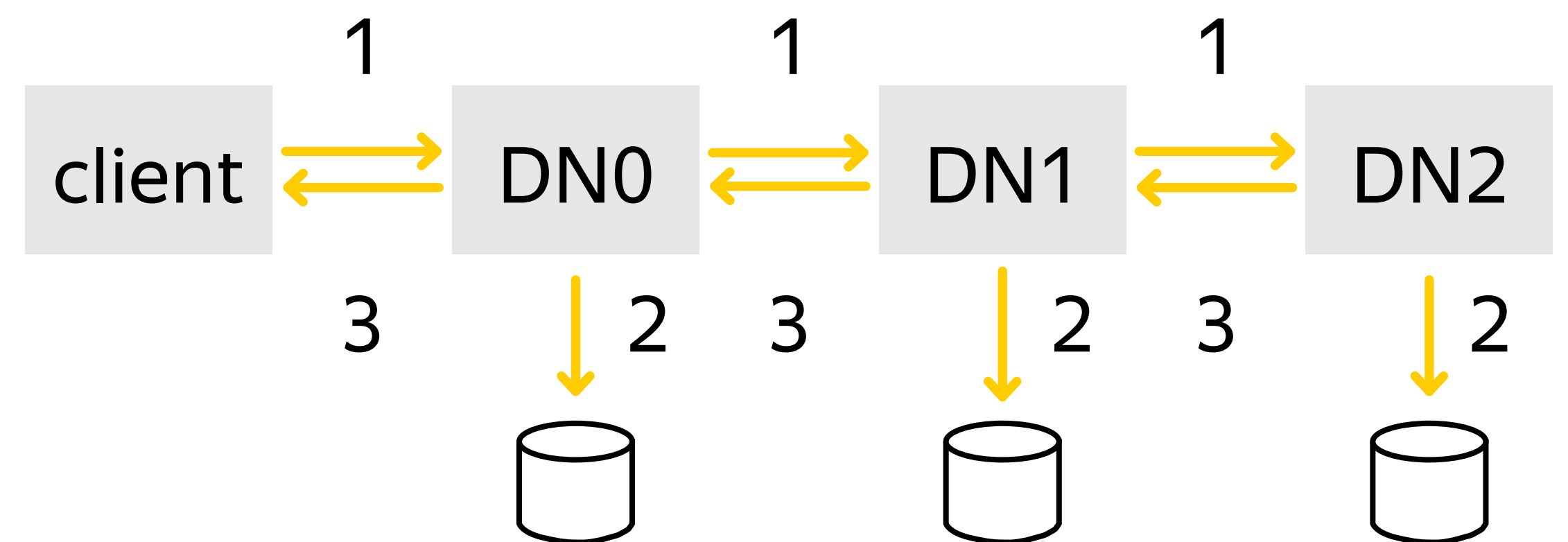
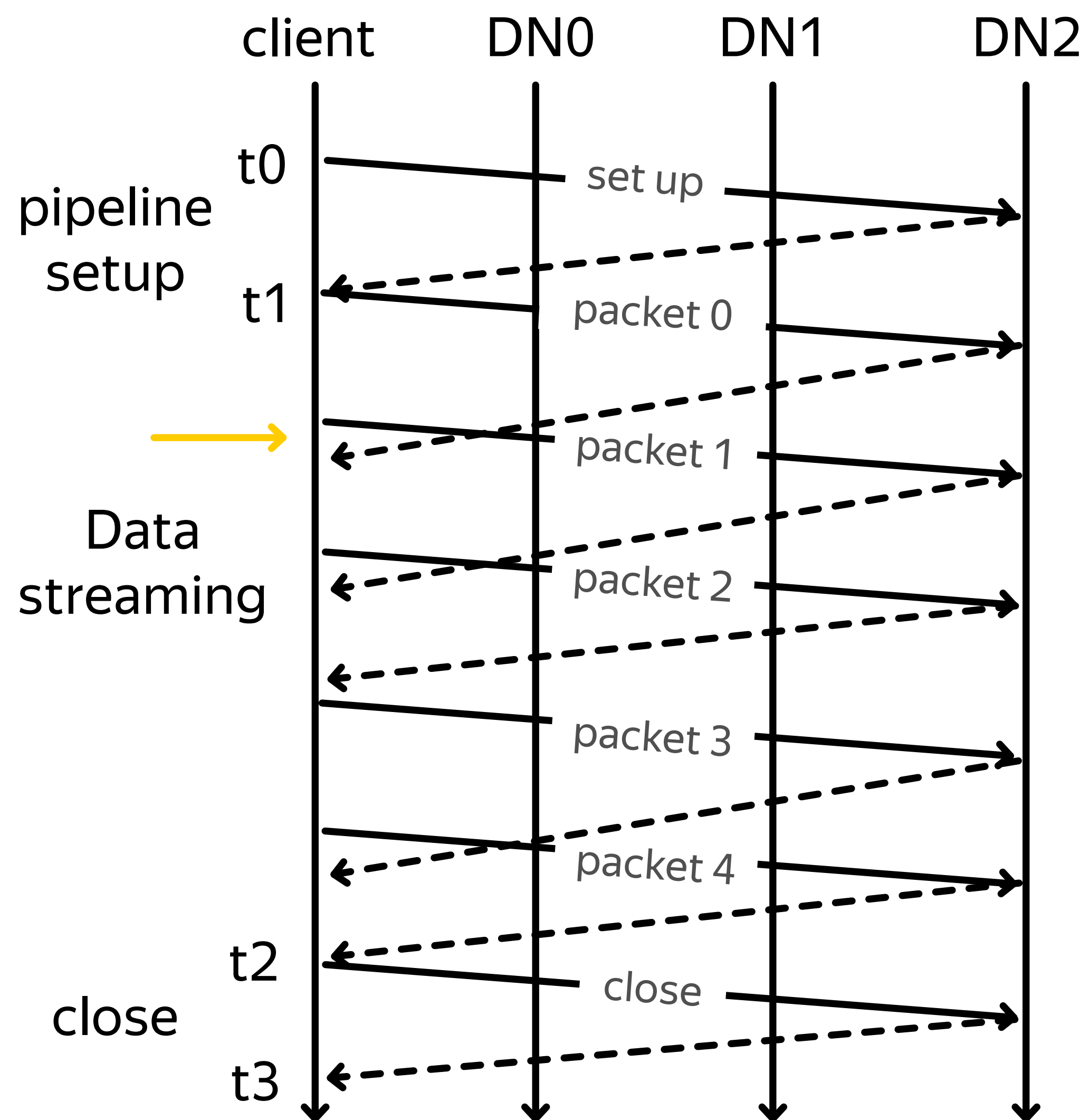
pipeline
setup



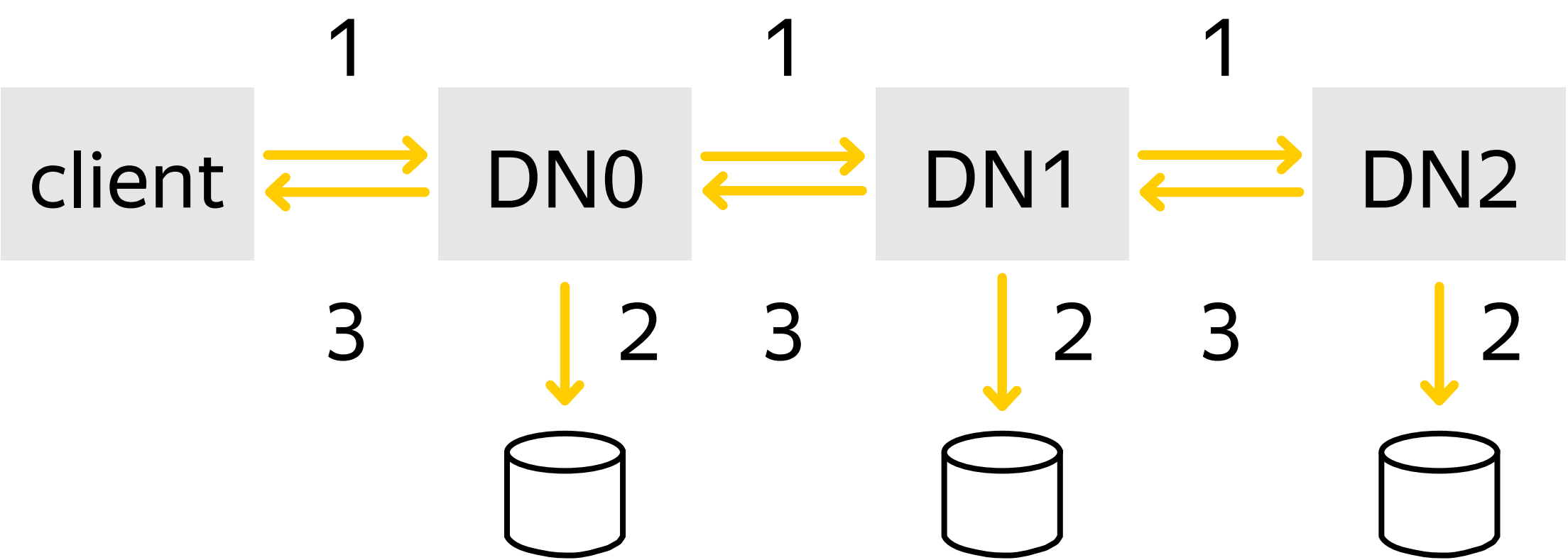
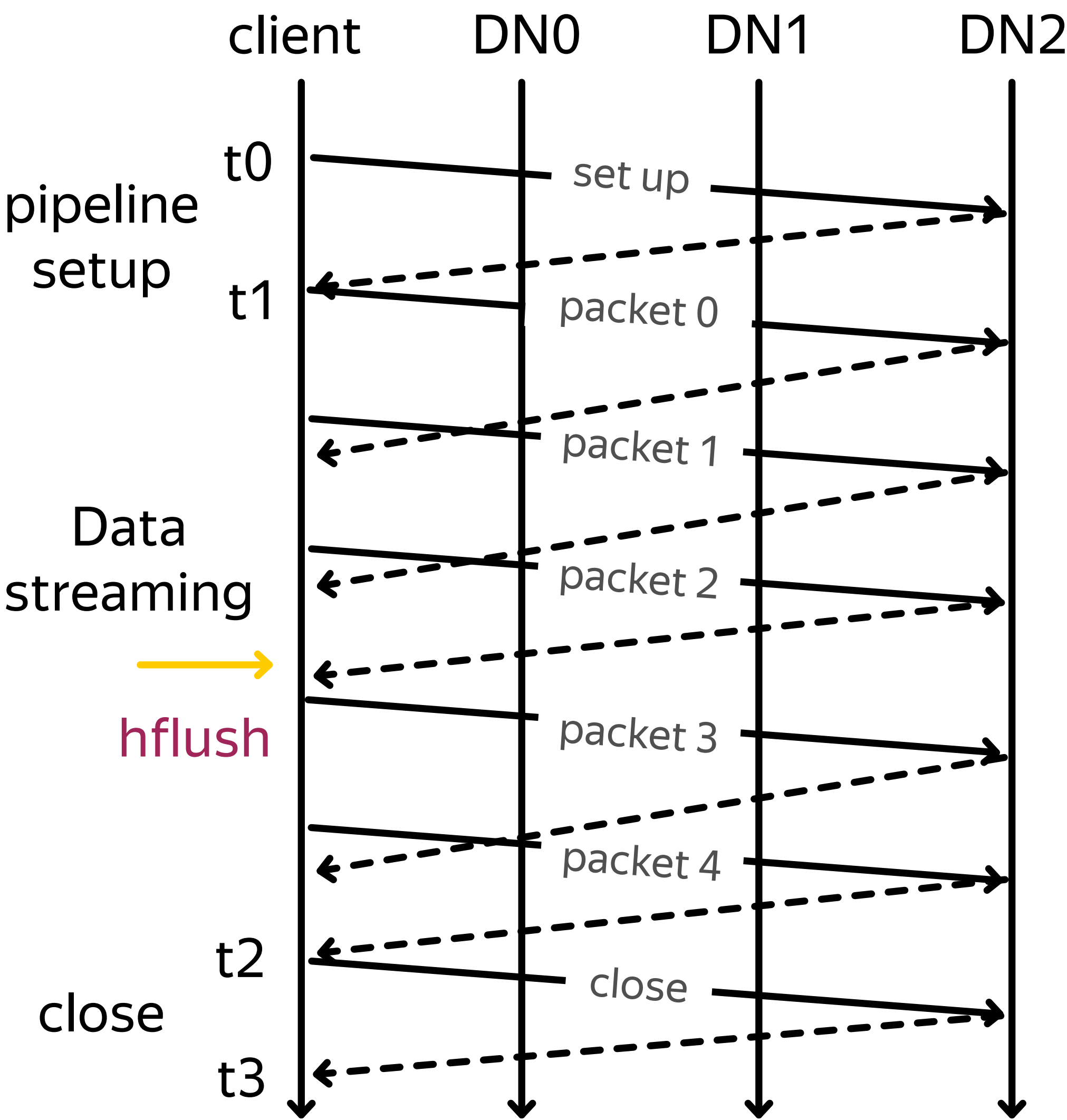
Pipeline



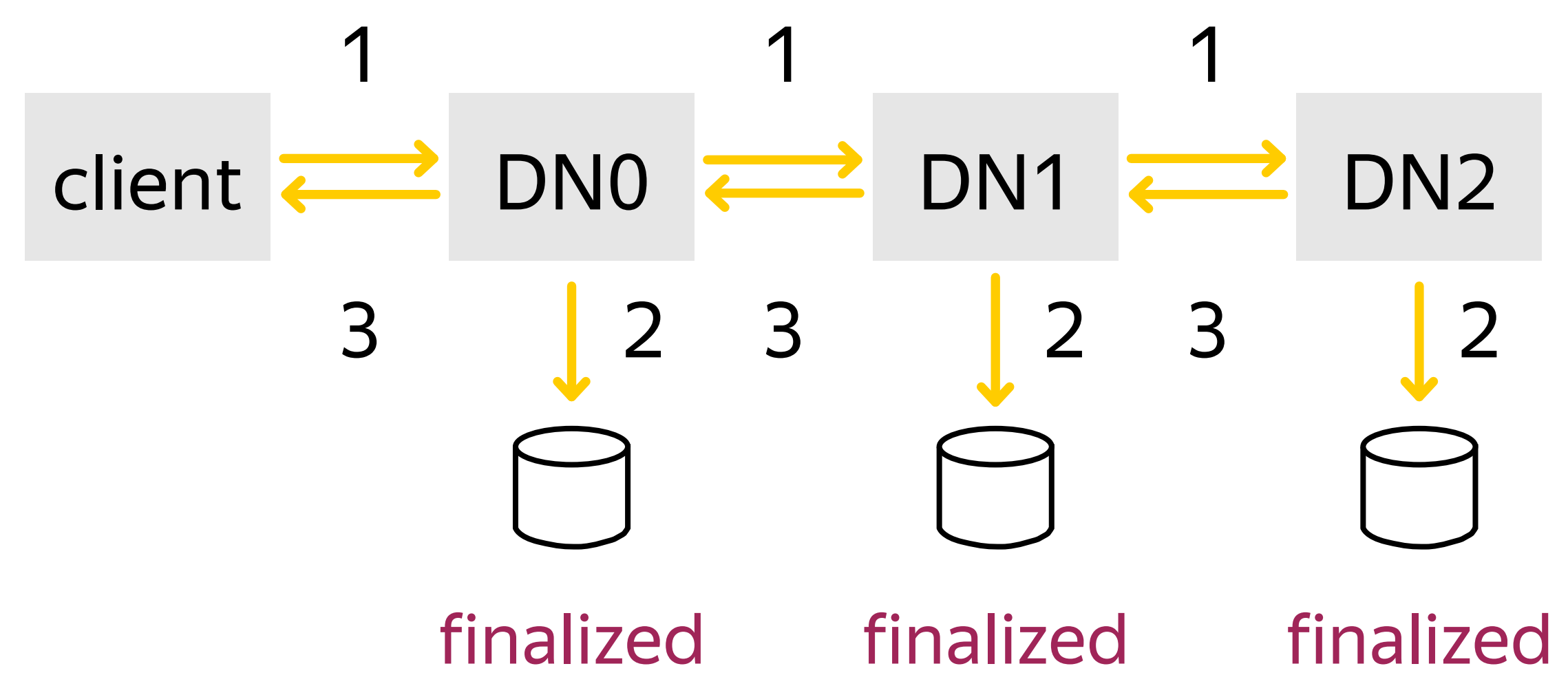
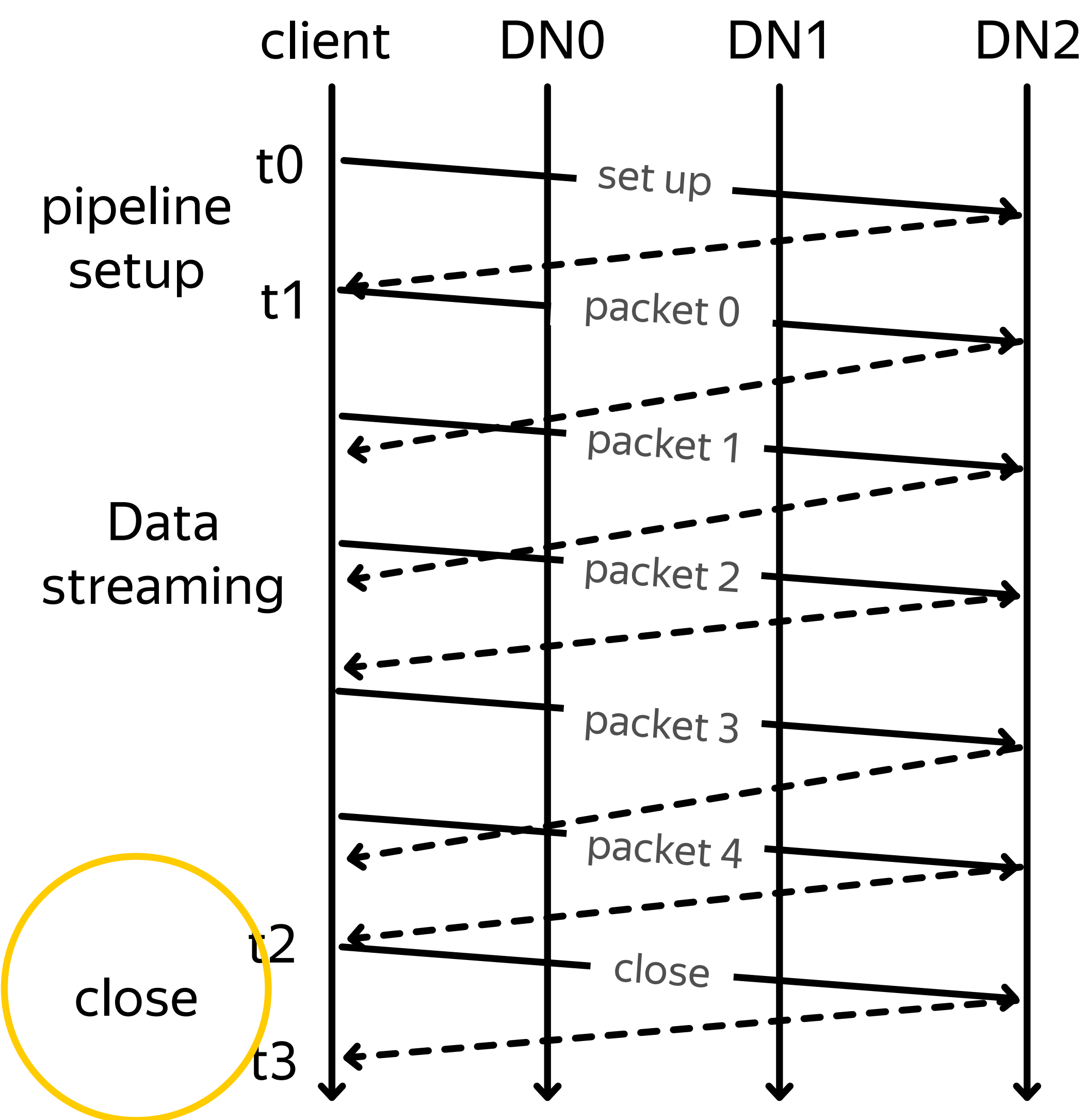
Pipeline



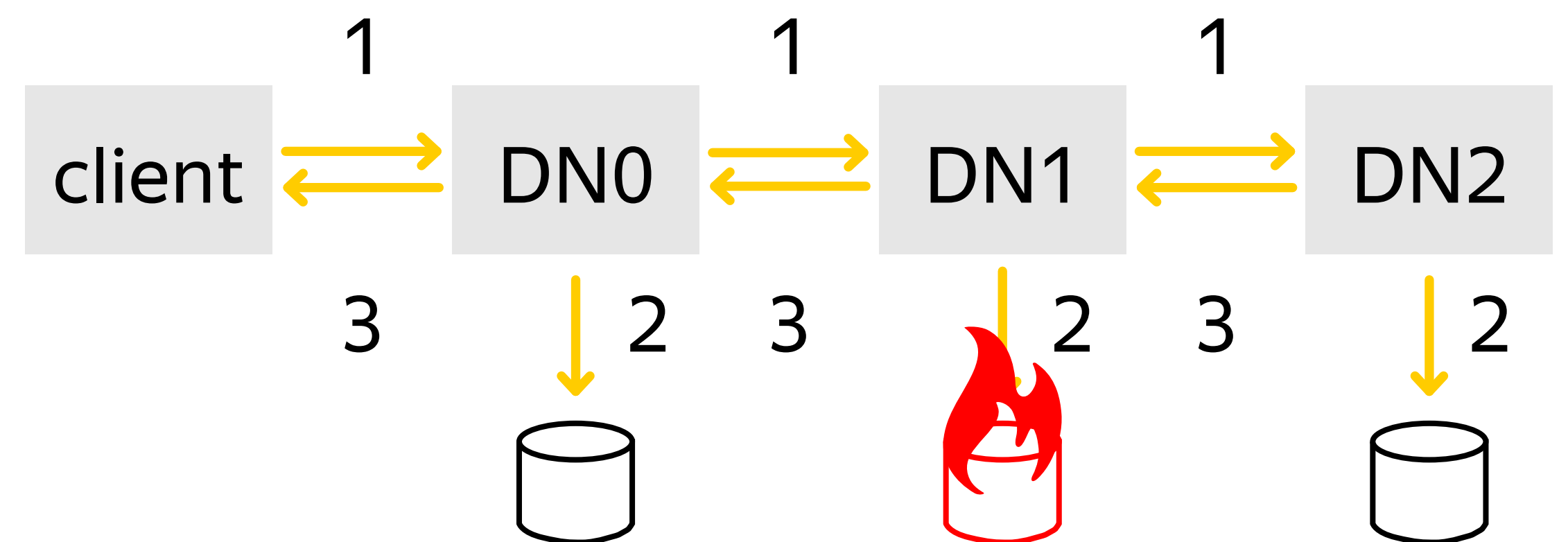
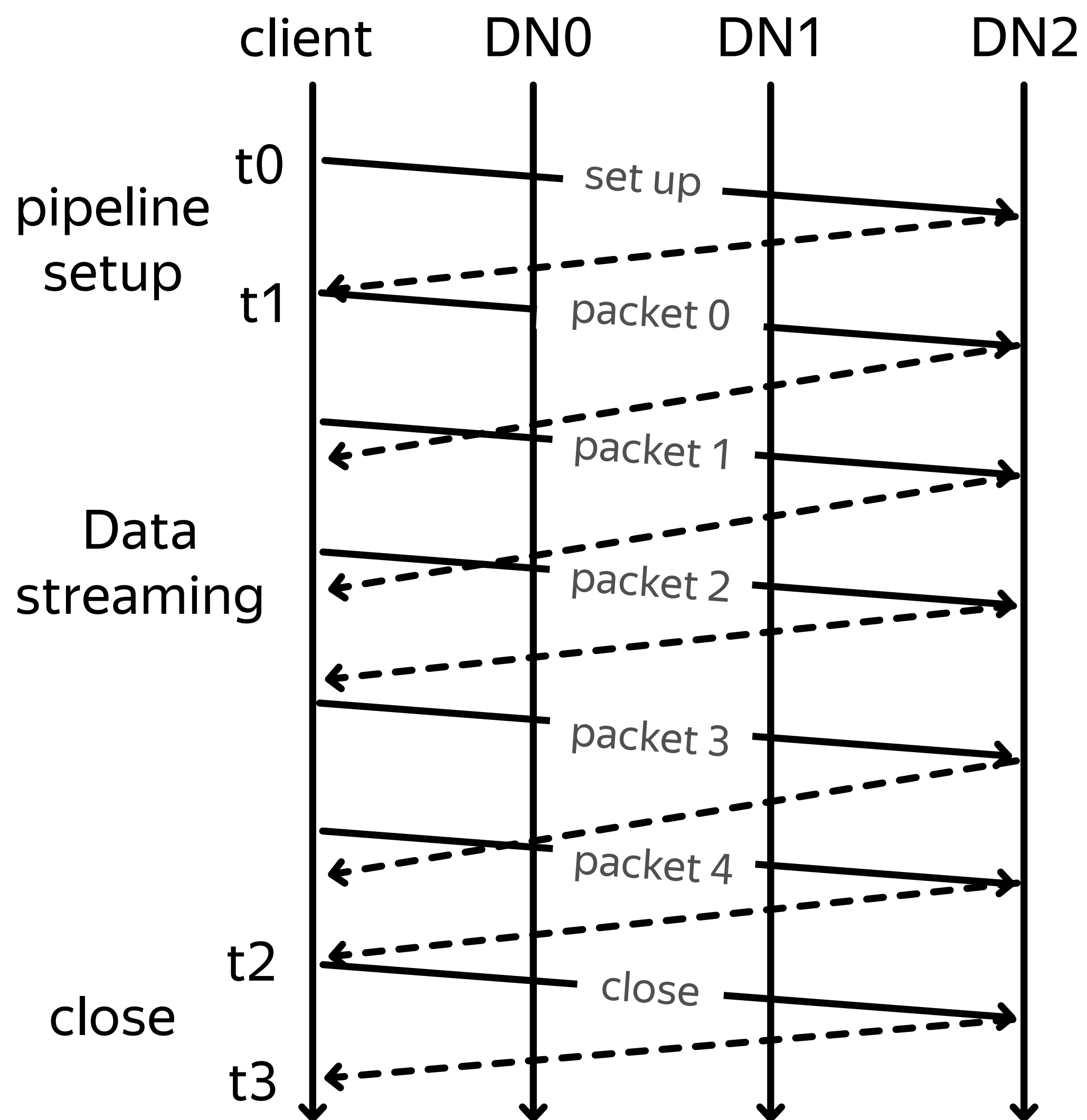
Pipeline



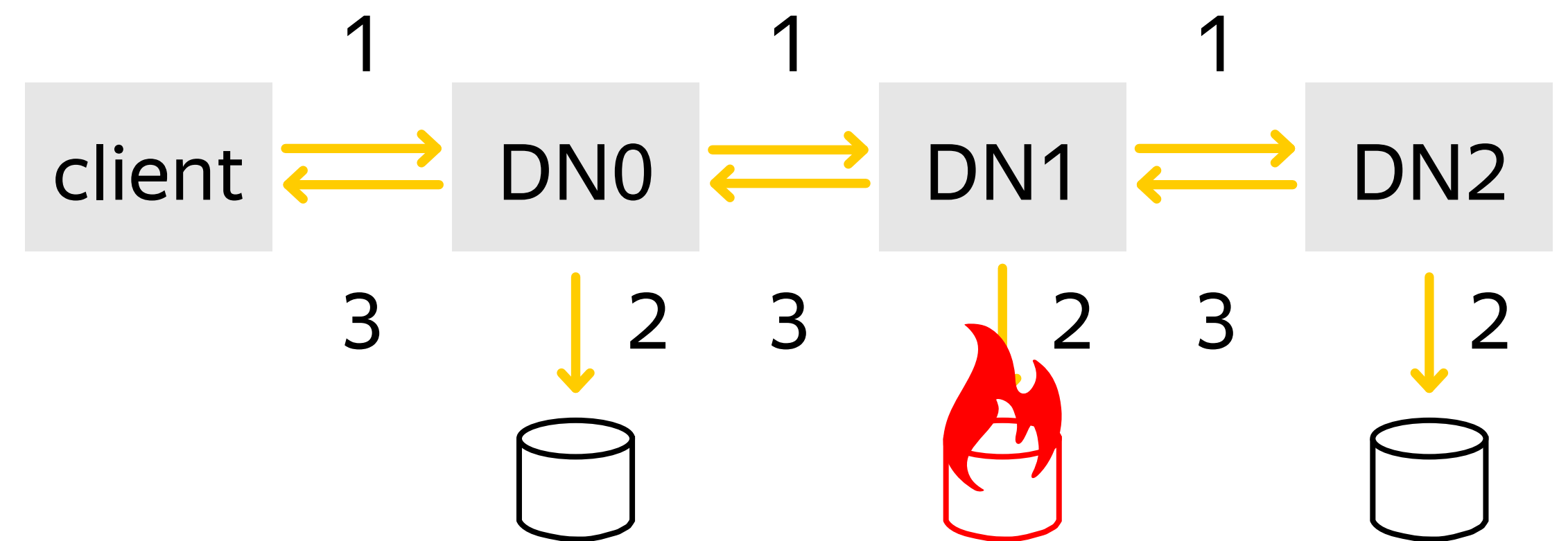
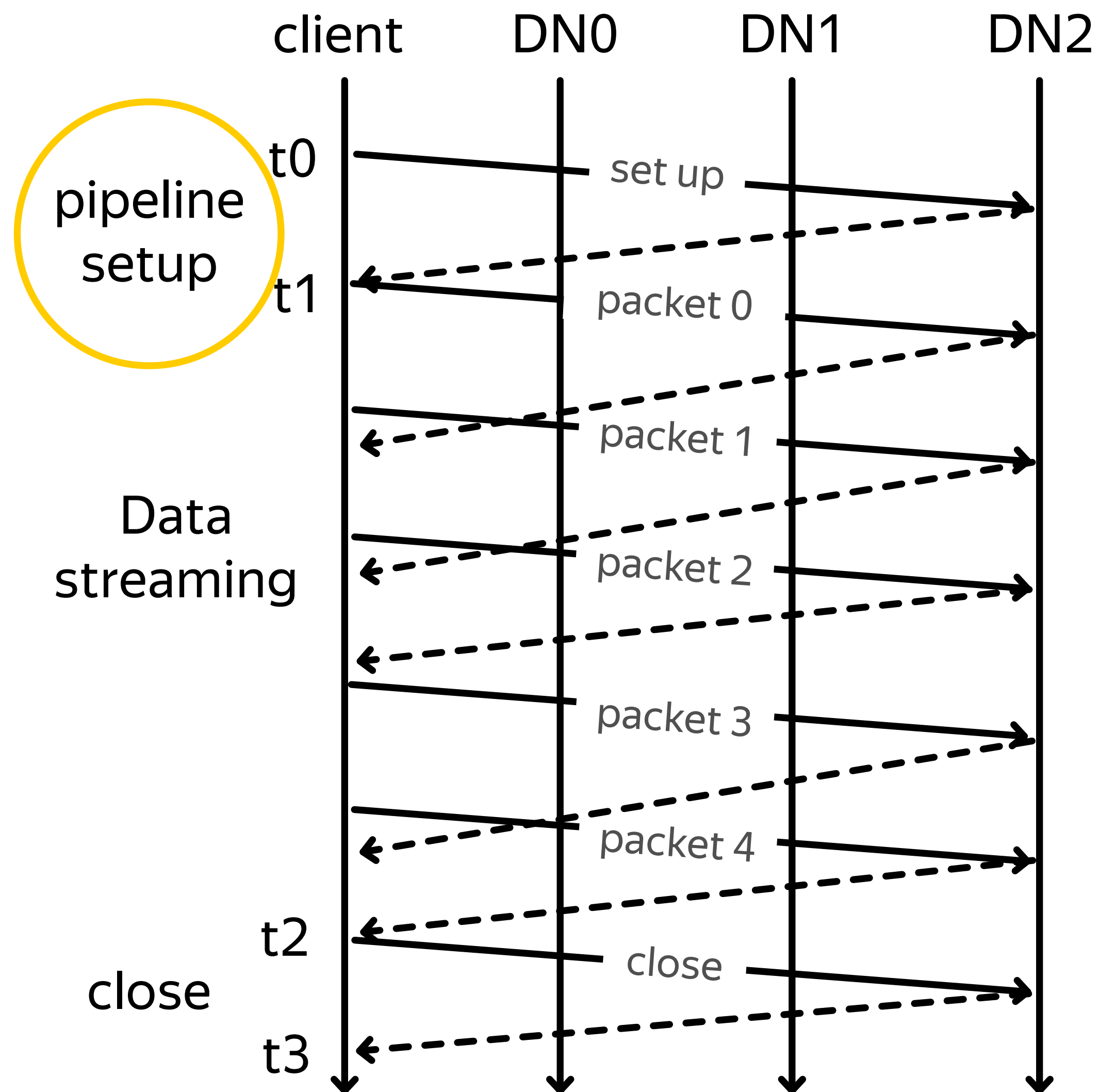
Pipeline



Pipeline Recovery

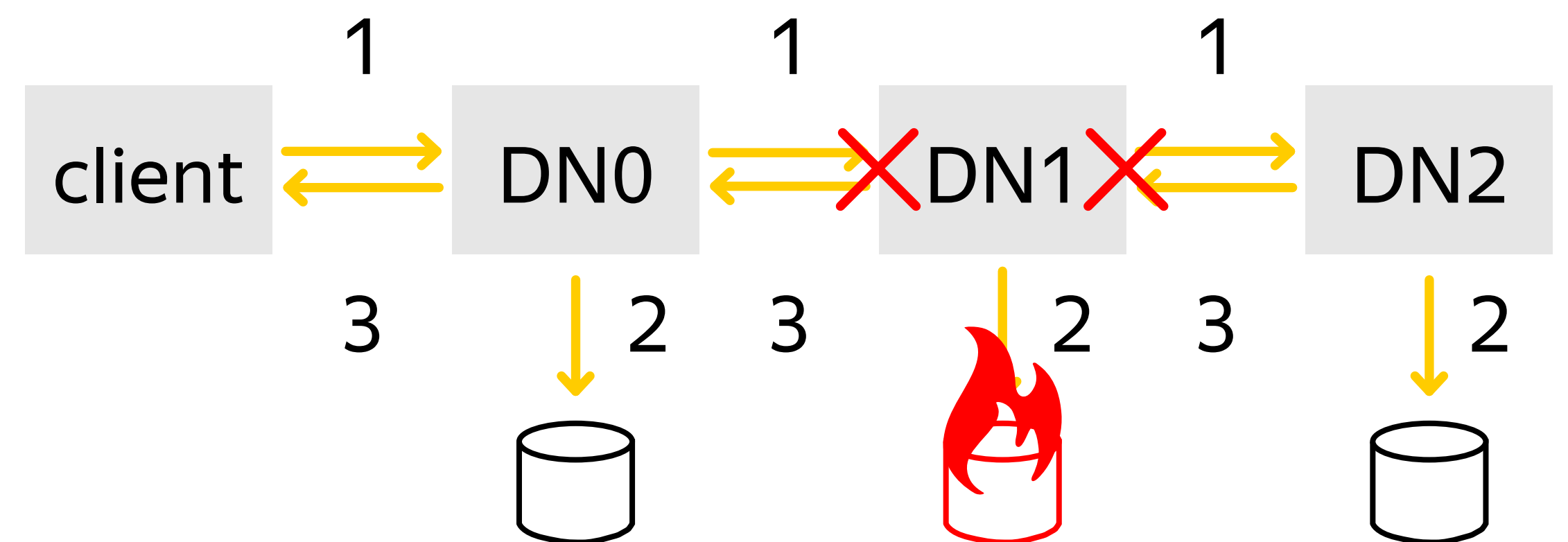
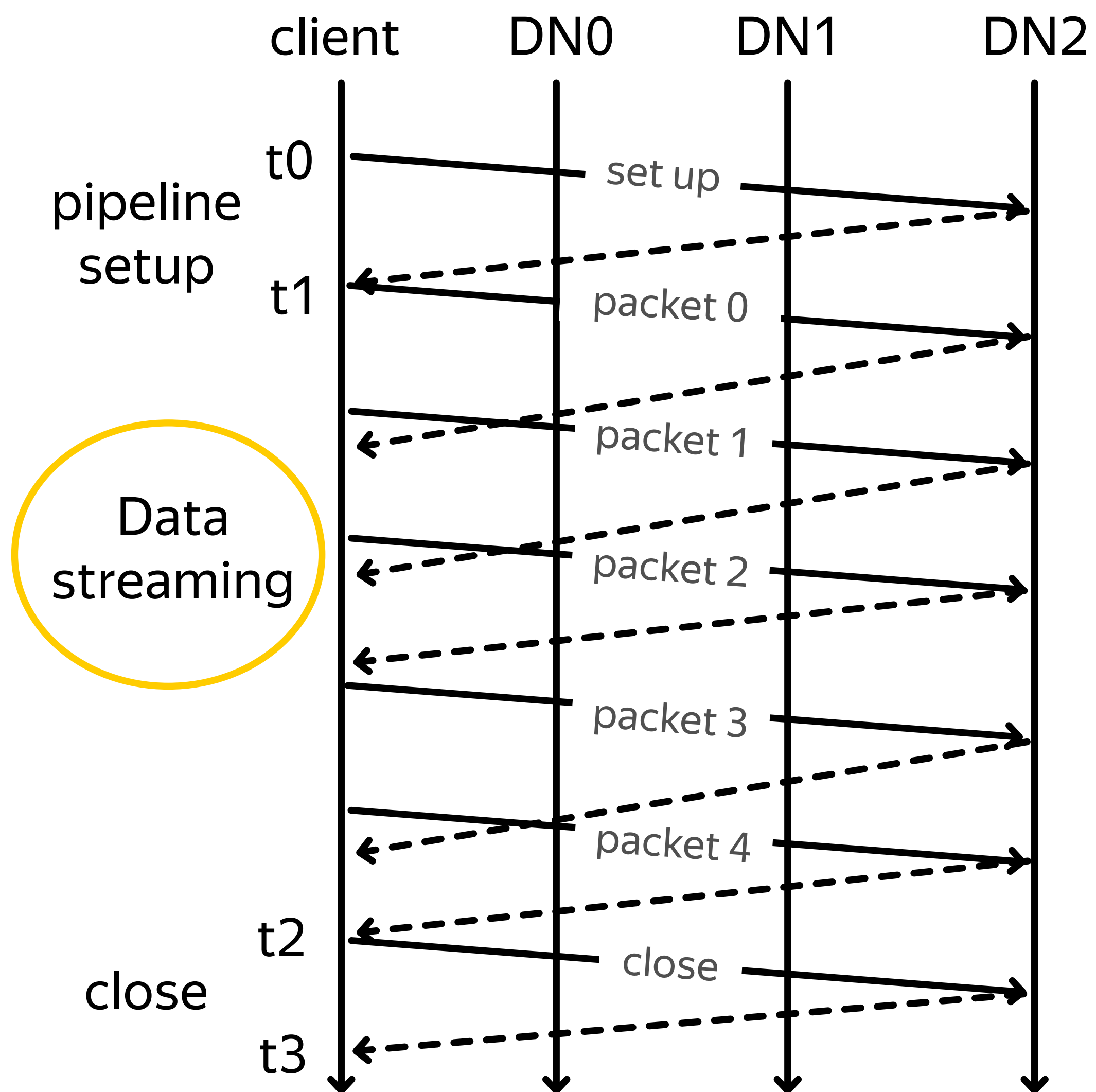


Pipeline Recovery

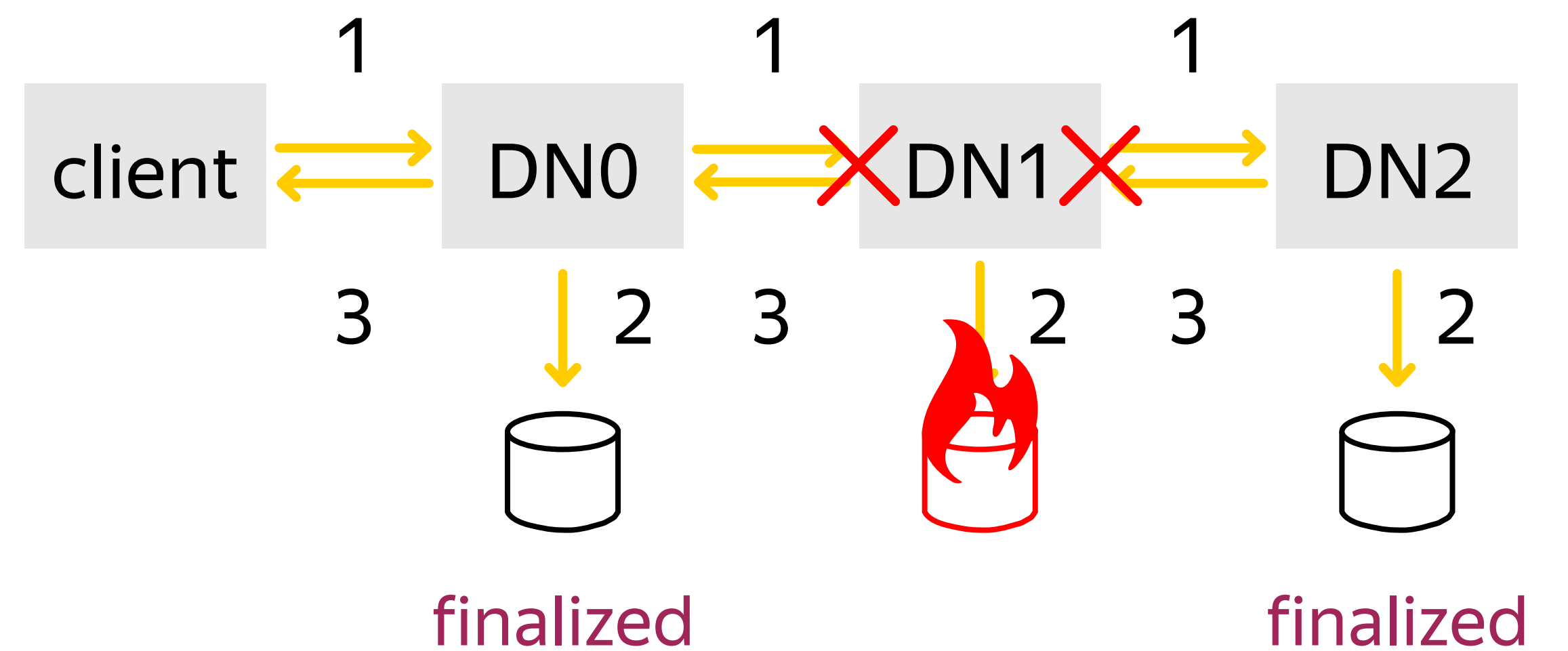
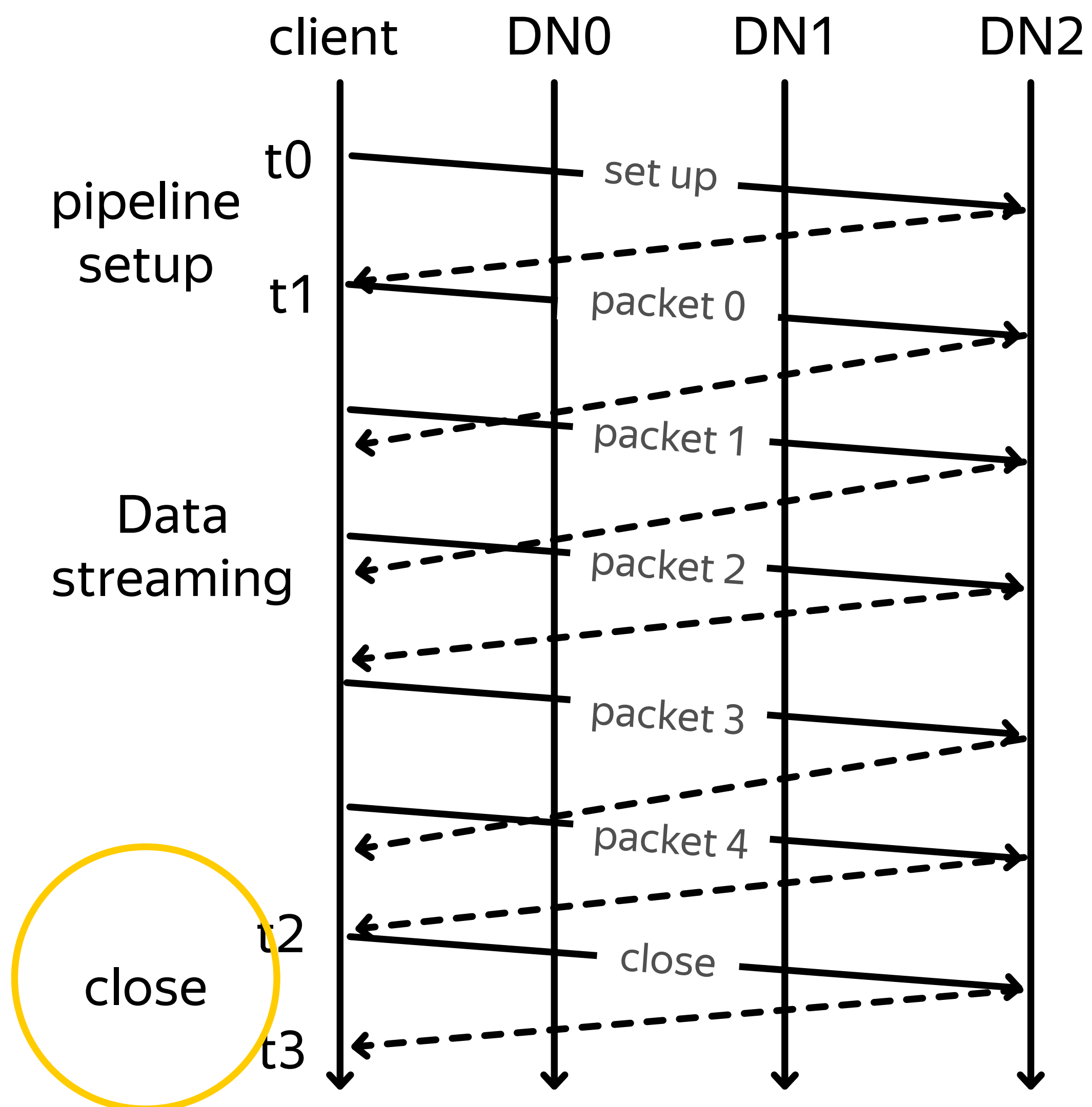


- cases:
1. new file
 2. append mode


Pipeline Recovery



Pipeline Recovery



Pipeline Recovery

 Hadoop HDFS / HDFS-8344

NameNode doesn't recover lease for files with missing blocks

Agile Board

Export

Details

Type: Bug

Priority: Major

Affects Version/s: 2.7.0

Component/s: namenode

Labels: None

Target Version/s: 2.9.0

Release Note: Allow a configuration to specify the maximum number of recovery attempts for blocks under construction.

Status: PATCH AVAILABLE

Resolution: Unresolved

Fix Version/s: None

People

Assignee: Ravi Prakash

Reporter: Ravi Prakash

Votes: 0 Vote for this issue

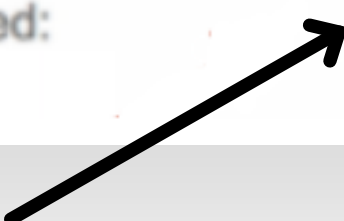
Watchers: 29 Start watching this issue

Dates

Created: 07/May/15 17:17

Updated: 10/Mar/17 16:10

2017



Summary

Summary

- › you can **draw** State block and replica transition tables

Summary

- › you can **draw** State block and replica transition tables
- › you can **identify** write pipeline stages and associated recovery process

Summary

- › you can **draw** State block and replica transition tables
- › you can **identify** write pipeline stages and associated recovery process
- › you can **list** 4 recovery processes and **explain** their purpose (lease recovery → block recovery → replica recovery; pipeline recovery)

Summary

- › you can **draw** State block and replica transition tables
- › you can **identify** write pipeline stages and associated recovery process
- › you can **list** 4 recovery processes and **explain** their purpose

Design Document:

<https://issues.apache.org/jira/secure/attachment/12445209/appendDesign3.pdf>

BigDATAteam