

Mongodb 复制集的原理机制

@eshujiushiwo (周李洋)

个人 qq:378013446

Mongodb 上海用户组: 192202324

Mongodb 在 DeNA 的应用



选择 MongoDB 的原因

1. 原生支持的复制集架构
2. 扩展性极强
3. Bson 文档结构

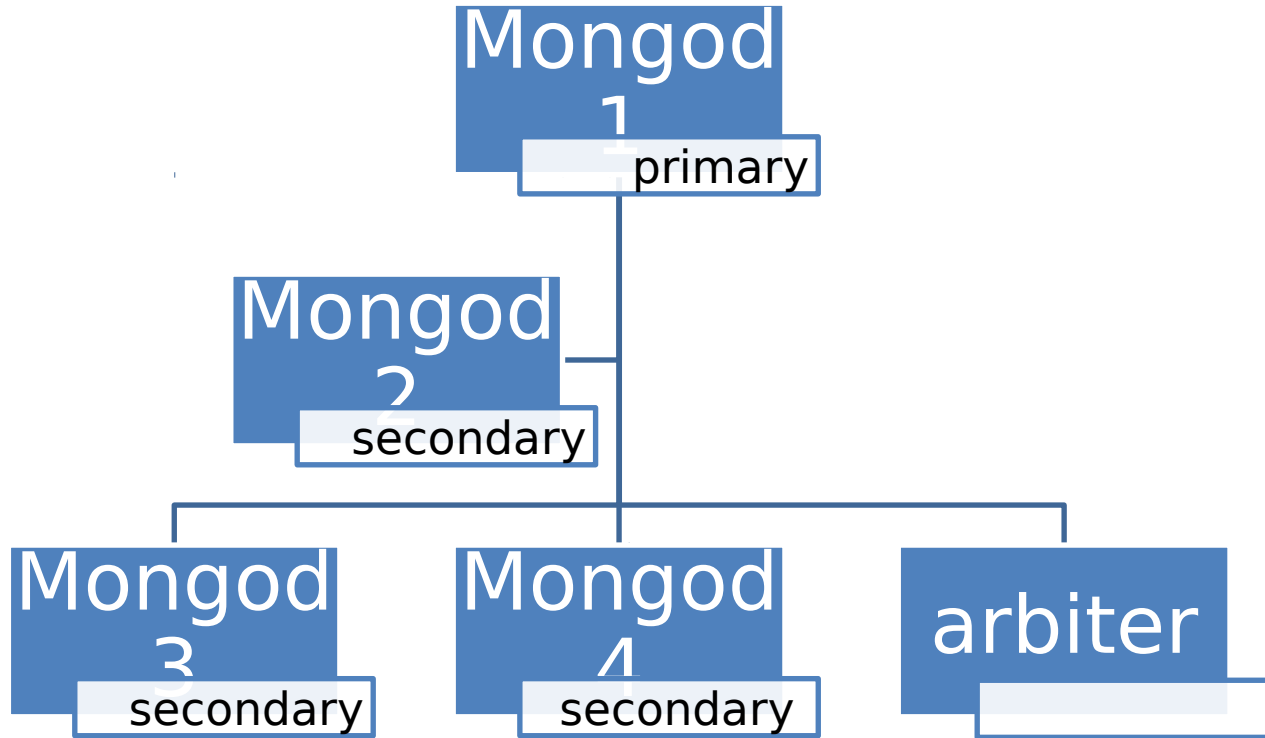
DeNA



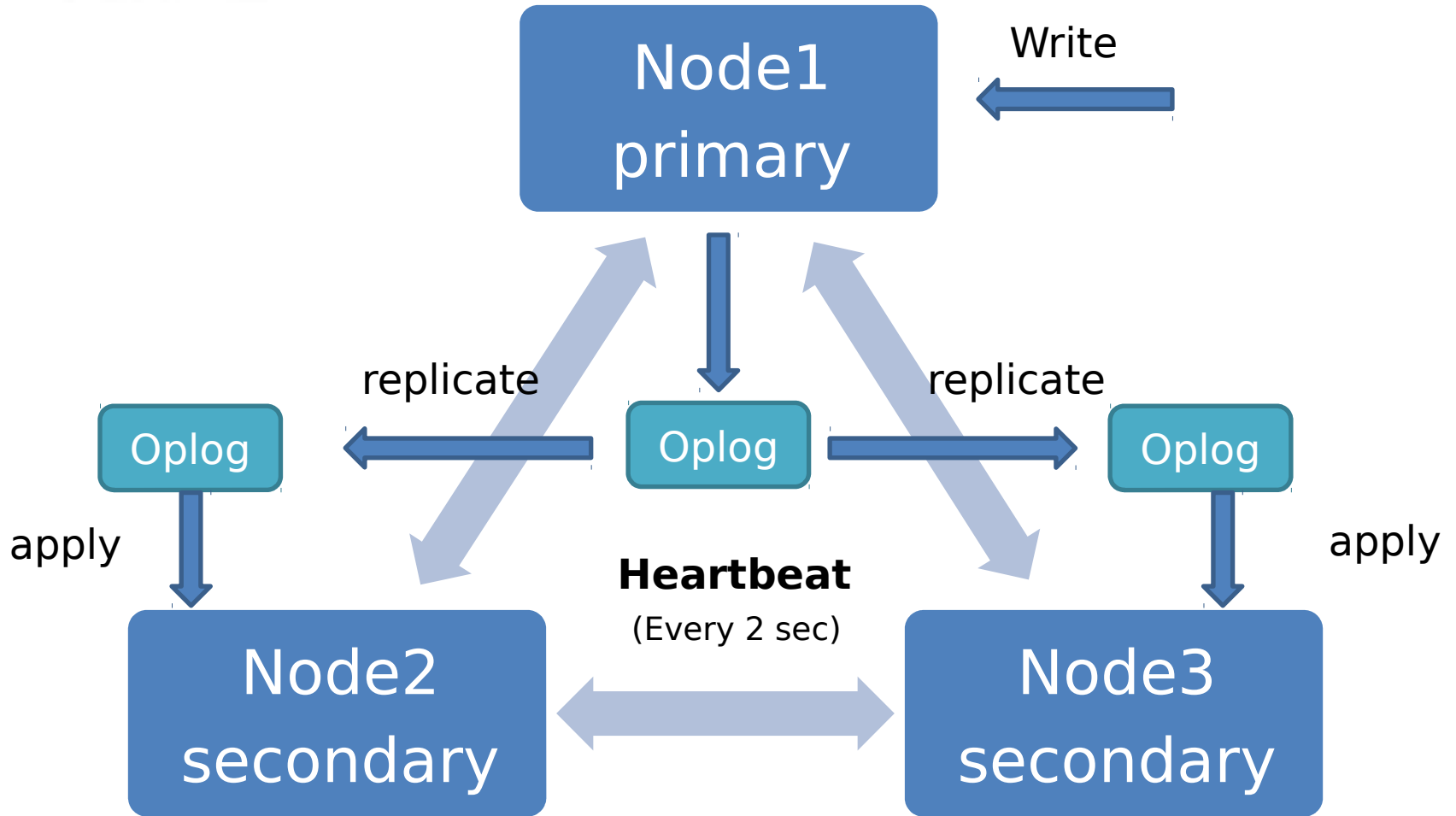
《NBA 梦之队》：
月流水：3500W+，注册用户：1600W+

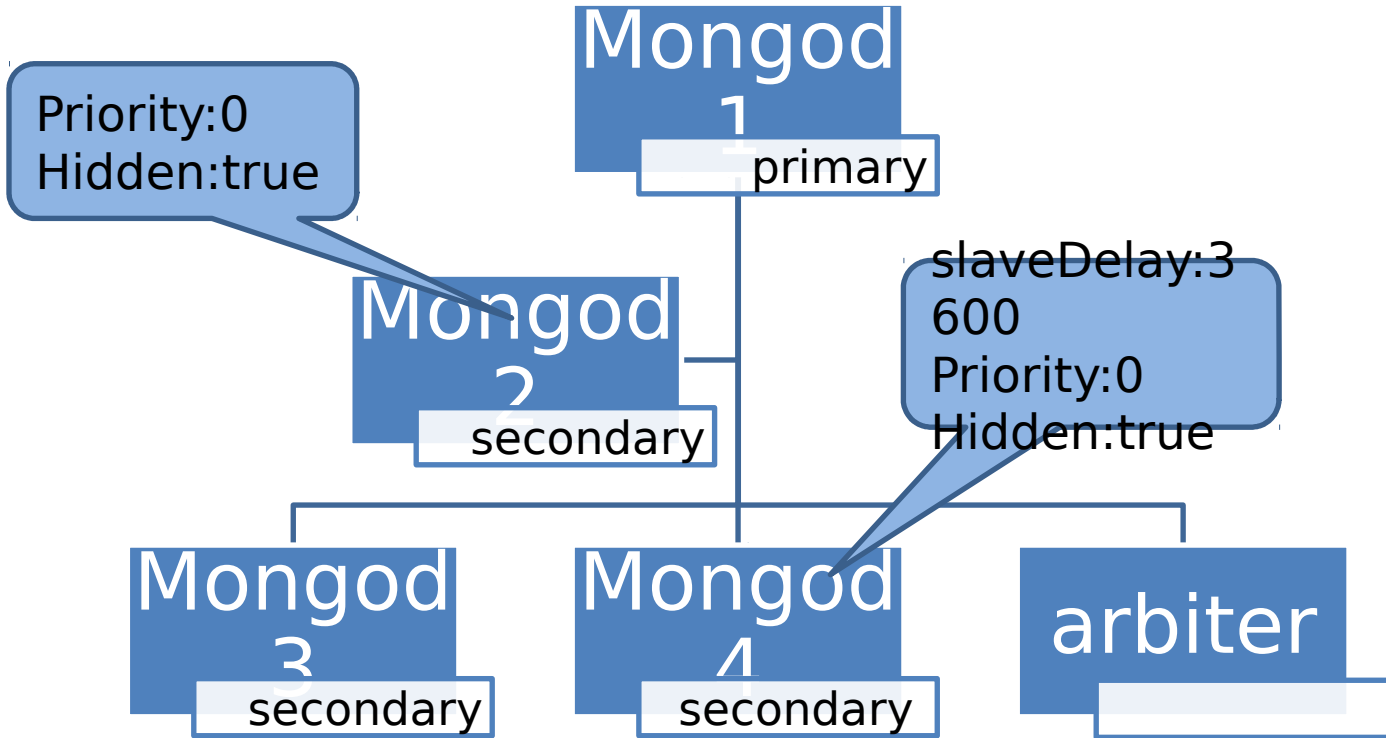
《变形金刚 崛起》：
月流水：1500W+，注册用户：500W+

《航海王 启航》：
即将启航



复制集架构





priority(优先级)

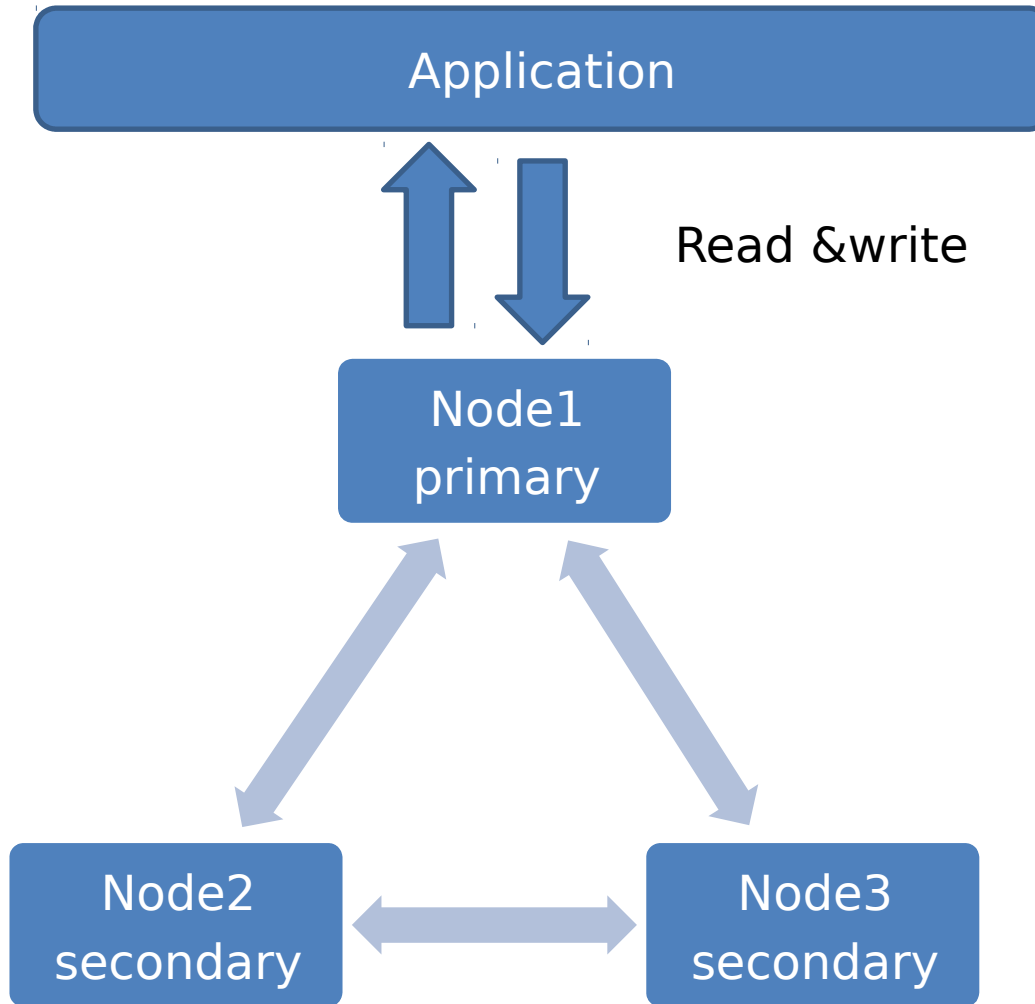
1. 优先级为 0 的节点将不能成为主节点。
2. 在选举中，选举票将优先投给 priority 更高的节点。



Hidden (隐藏与否)

当我们将 `hidden` 设置为 `true` 的时候，该节点将对应用程序不可见。（即当应用程序通过复制集读选项访问从节点的时候，请求将不会发送到该节点上。）





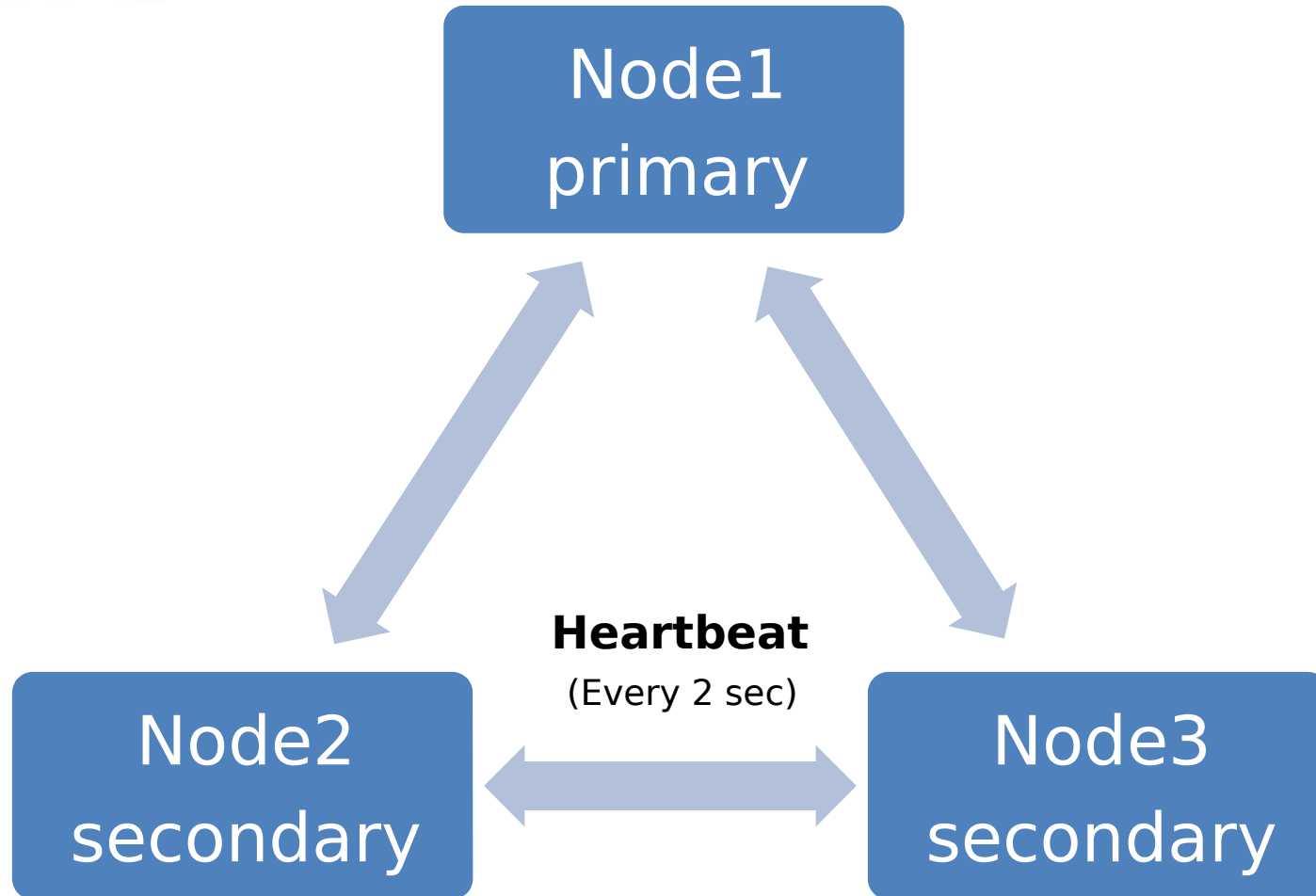
Read Preference Modes (复制集读选项)

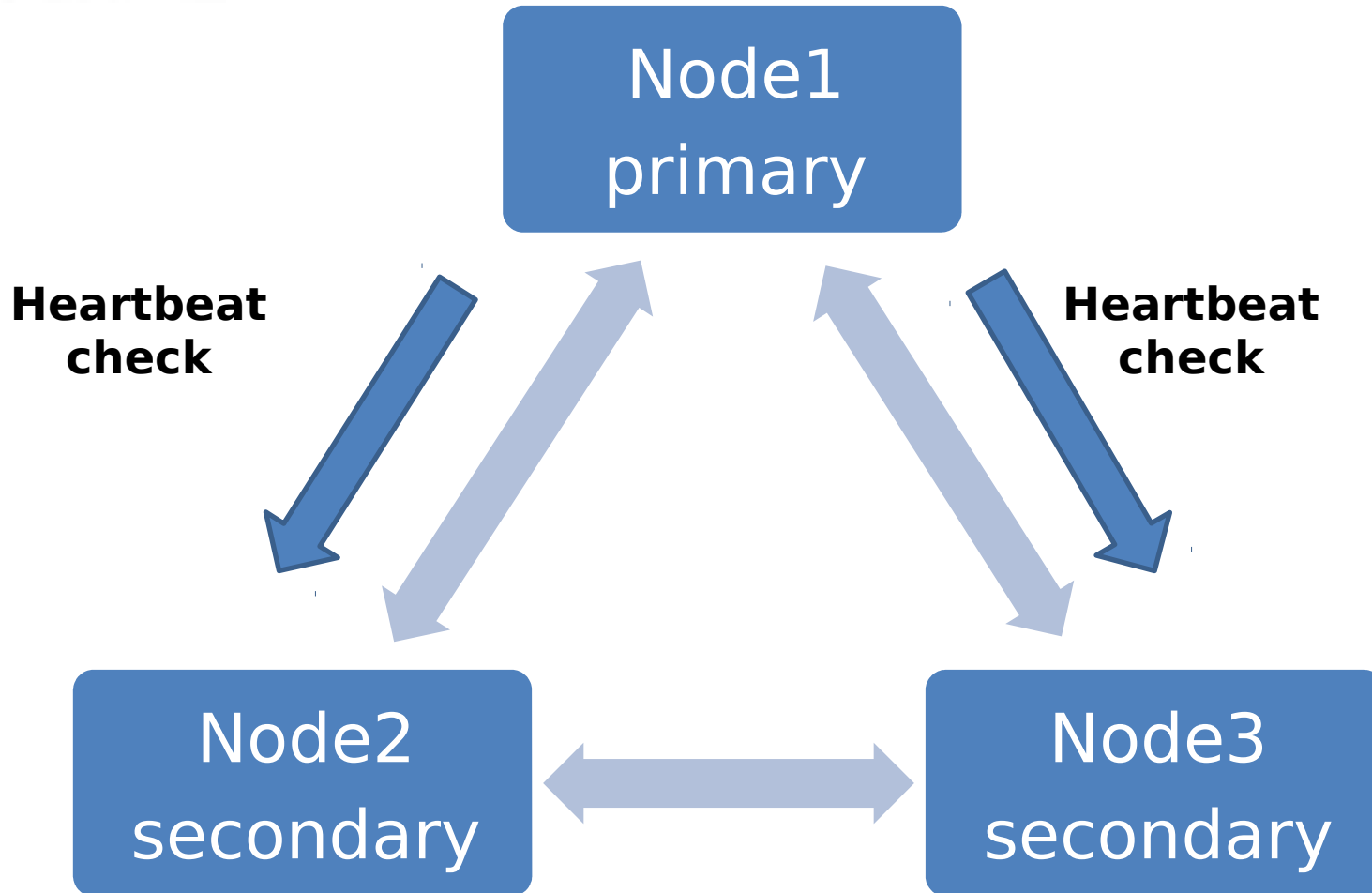
5 种模式

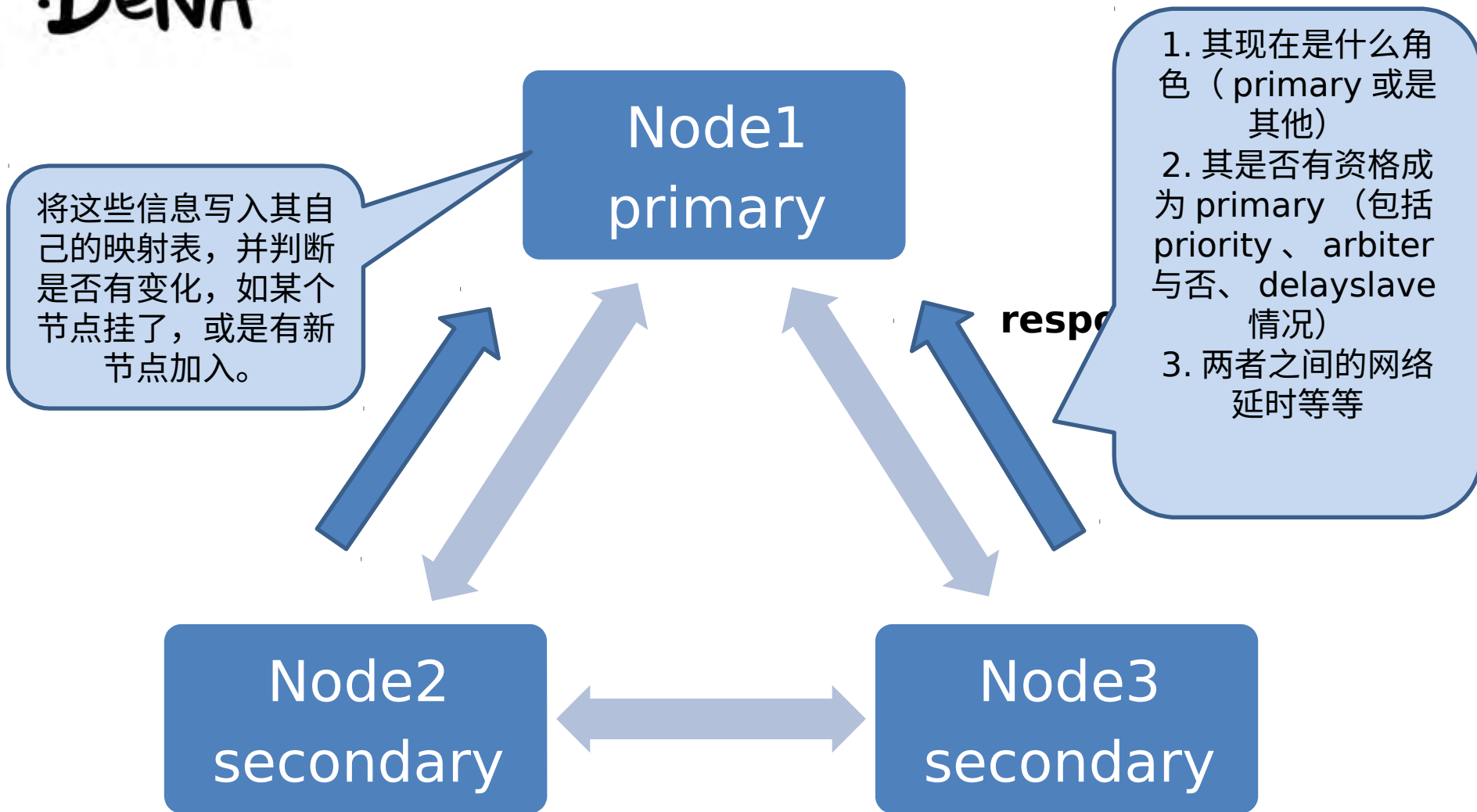
- 1.primary (only)(只在主节点上读) – Default
- 2.primaryPreferred (优先在主节点上读)
- 3.Secondary (在从节点上读)
- 4.secondaryPreferred (优先在从节点上读)
- 5.Nearest (就近读)



Election(选举)







DeNA

每隔一段时间，就会进行一次状态监测，或是当状态映射表变动了，如有节点新增或是 down 了，也会进行。

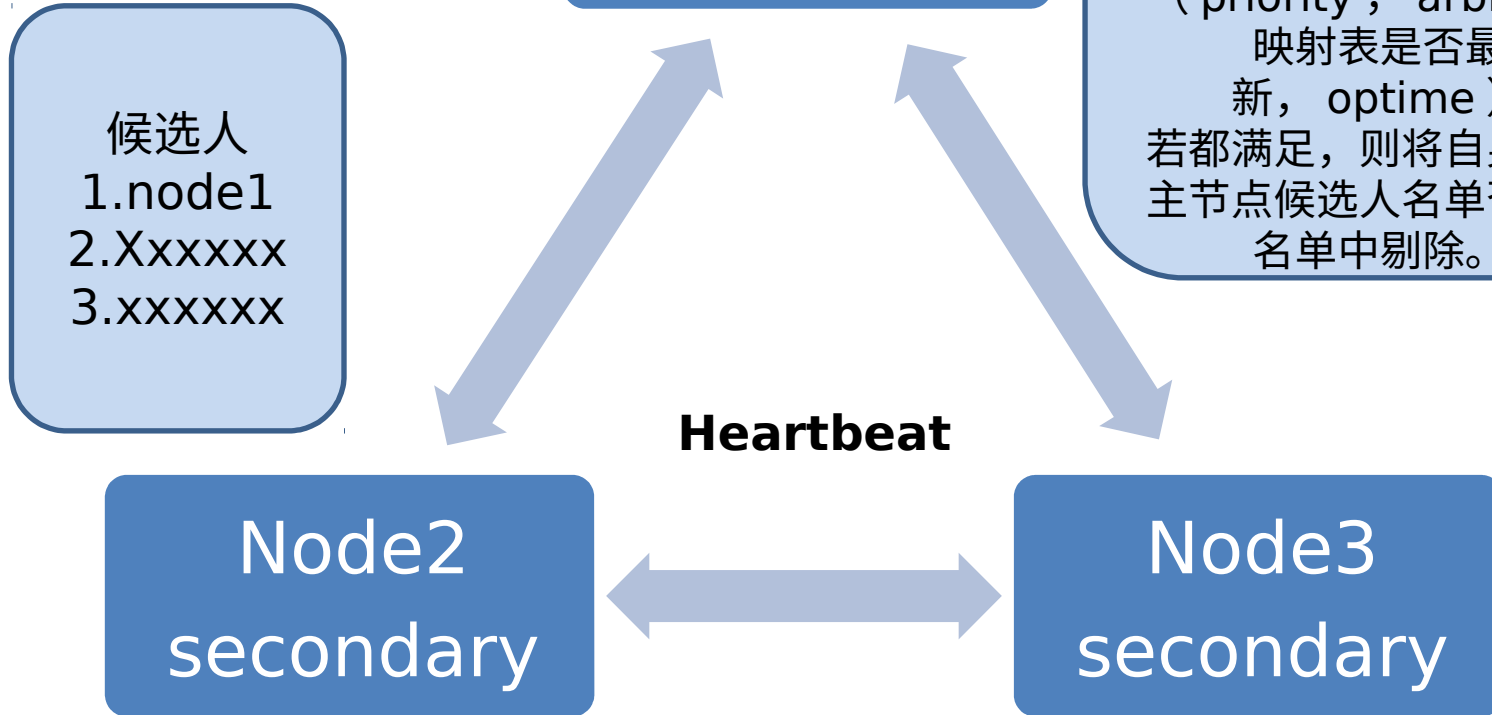
进行状态检测，具有一票

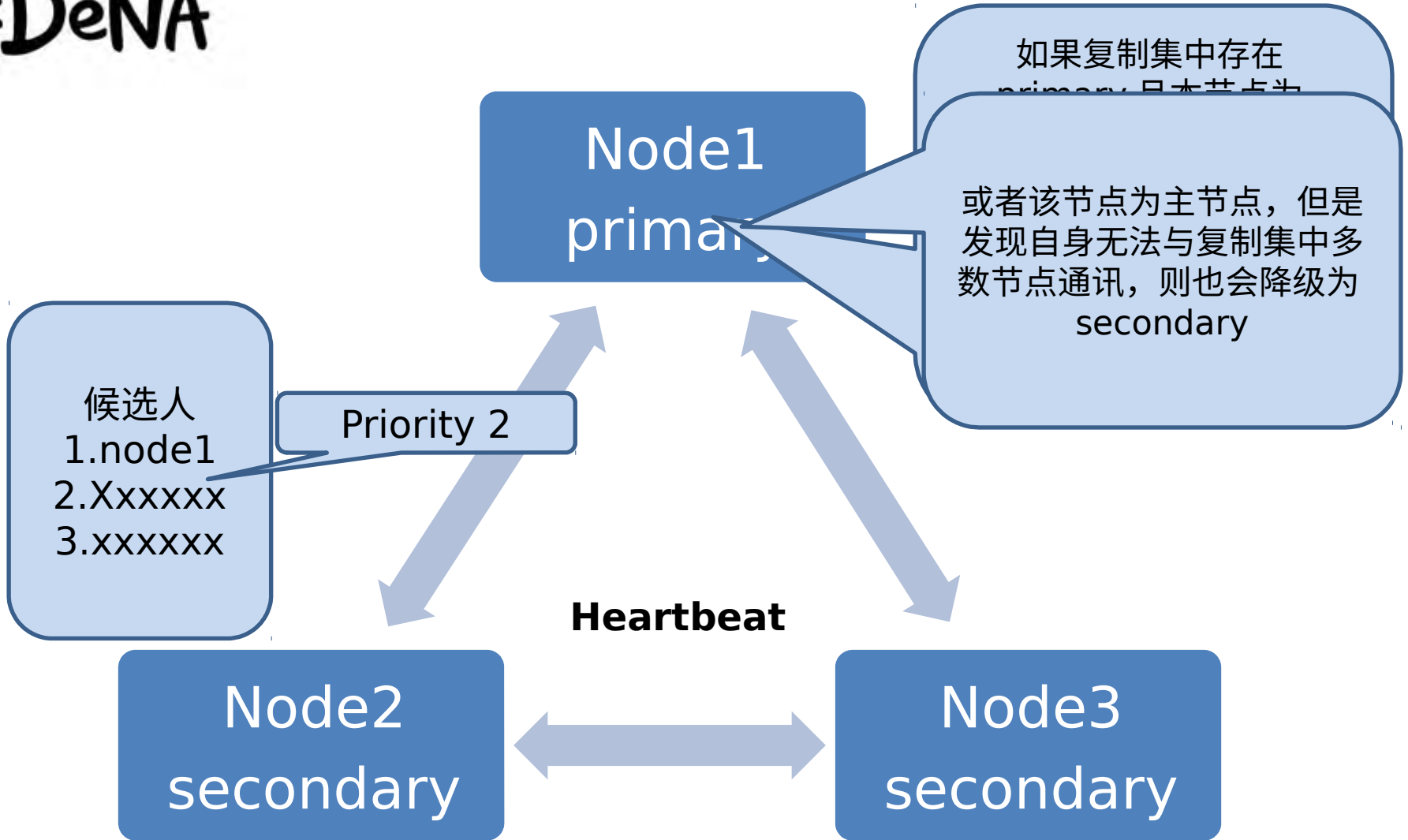
否决权，内容包括：

1. 自身是否处于选举过程
2. 是否能与复制集中多数节点进行通讯。
3. 是否满足成为主节点的条件

(priority, arbiter, 映射表是否最新, optime)

若都满足，则将自身加入主节点候选人名单否则从名单中剔除。





若在，则开始判断自己能否发起选举。判断有以下内容，具有一票否决权

1. 是否能与复制集中多数成员通讯。
 2. 是否在主节点候选人列表中。
 3. 是否满足成为主节点的条件
(priority, arbiter, slavedelay, optime 是否最新)。
 4. 30s 内是否透过赞成票。
- 若满足则会标示自己已经在选举中。并发起选举。

如果复制集中不存在 primary。

Node1
Primary

Node2
secondary

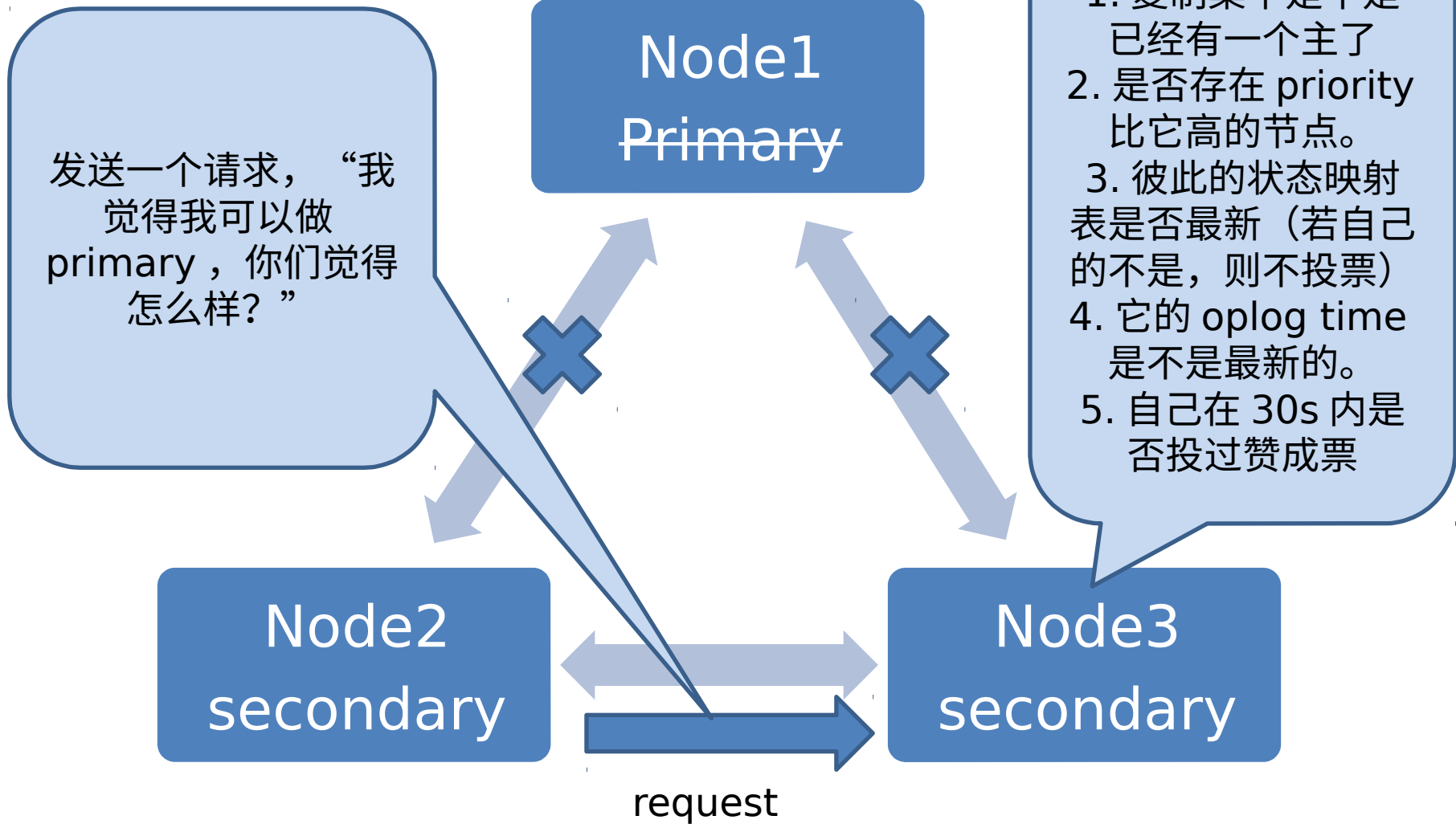
Node3
secondary

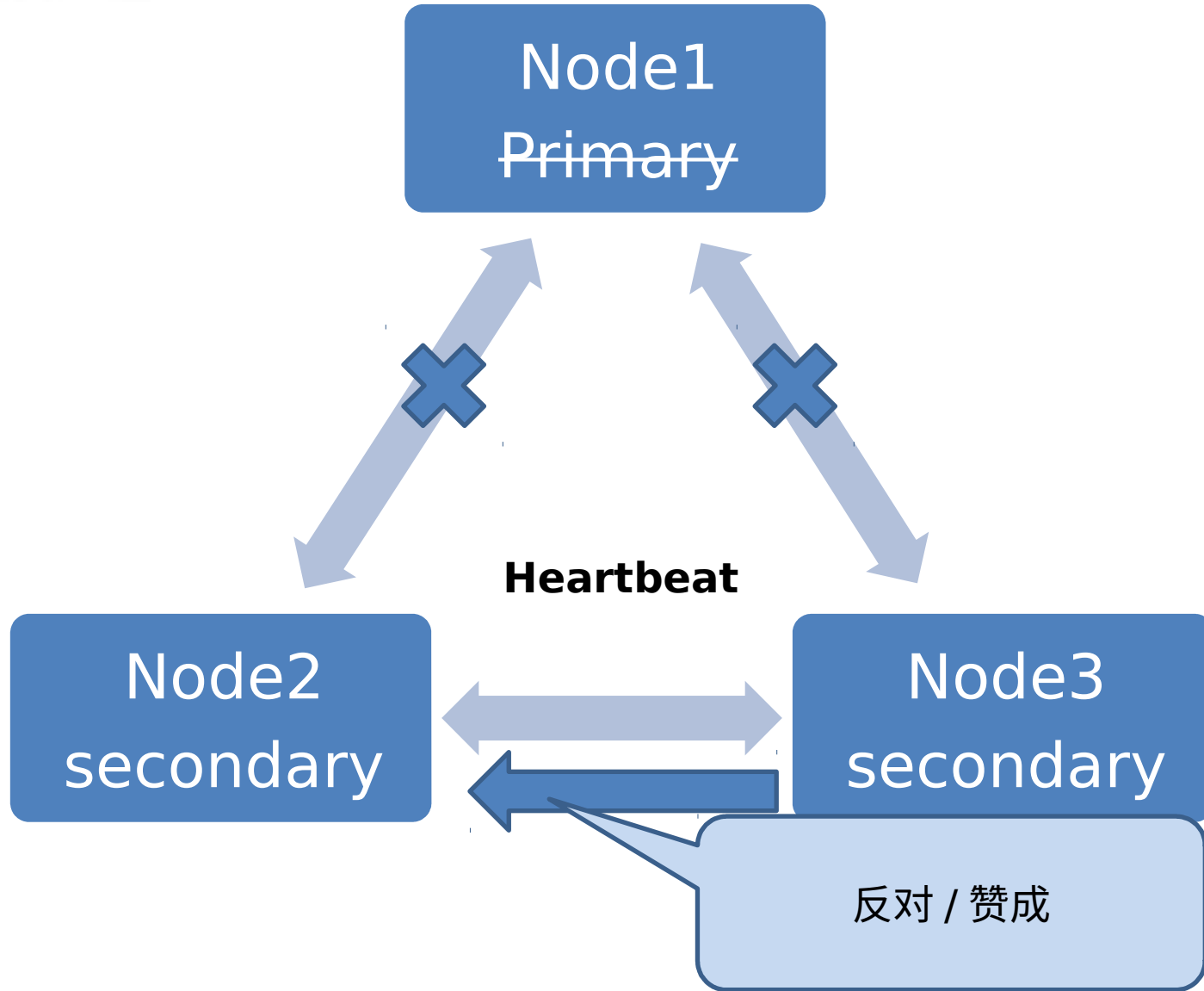
检测自己是否在主节点候选人名单中，若不在，则退出该流程

候选人
1.node3
2.node2
3.xxxxxx

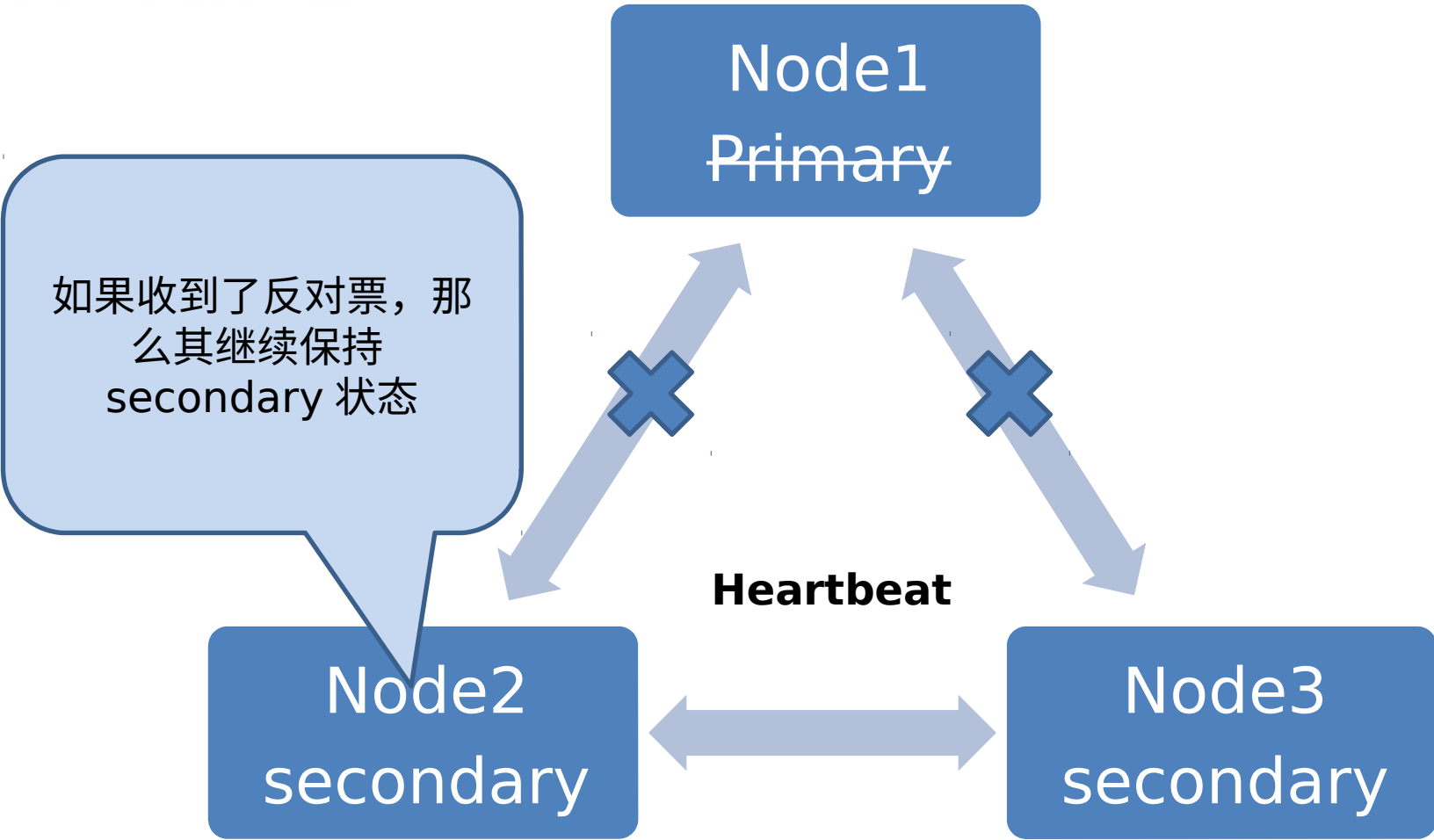
Heartbeat

DeNA

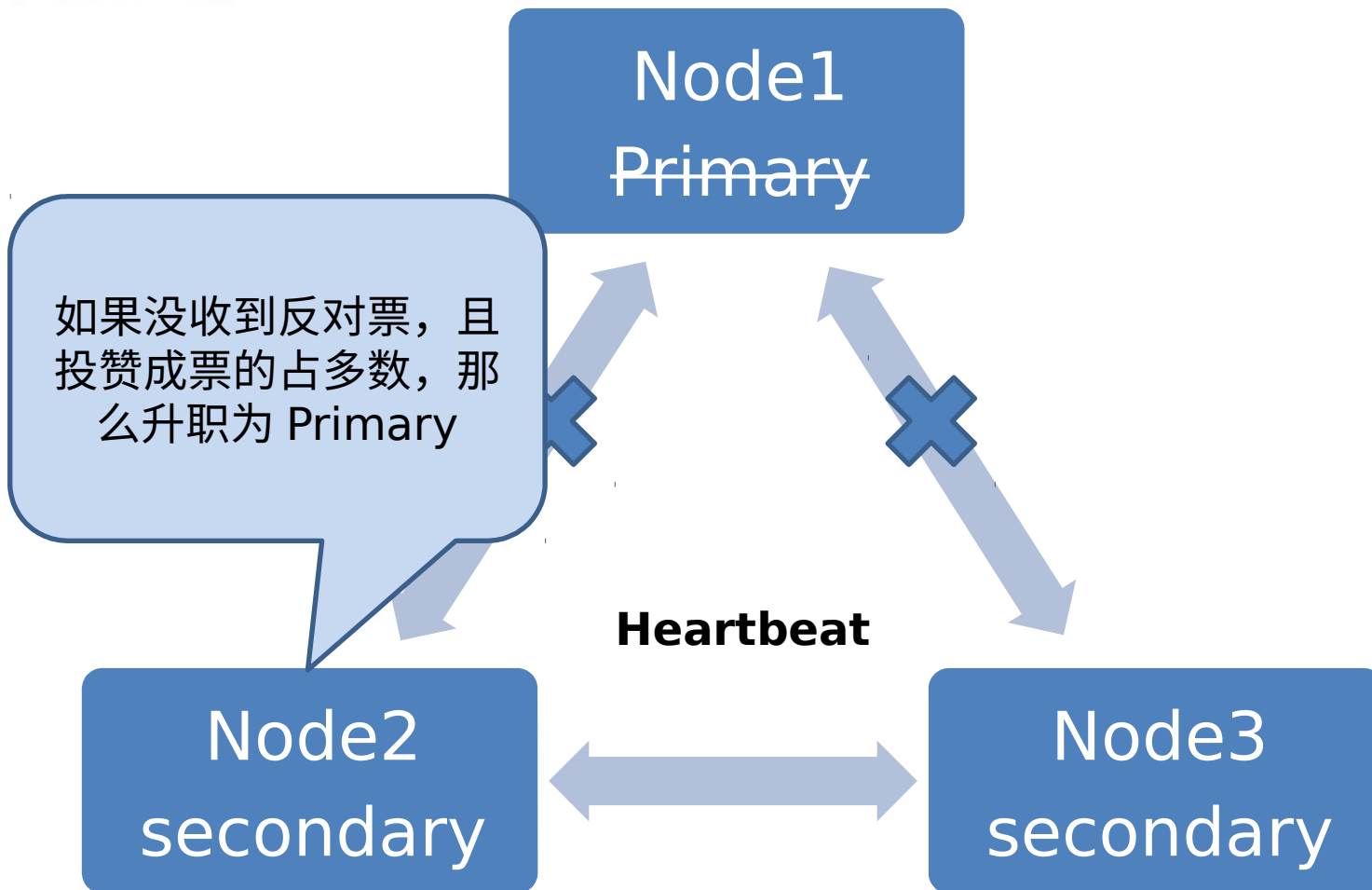


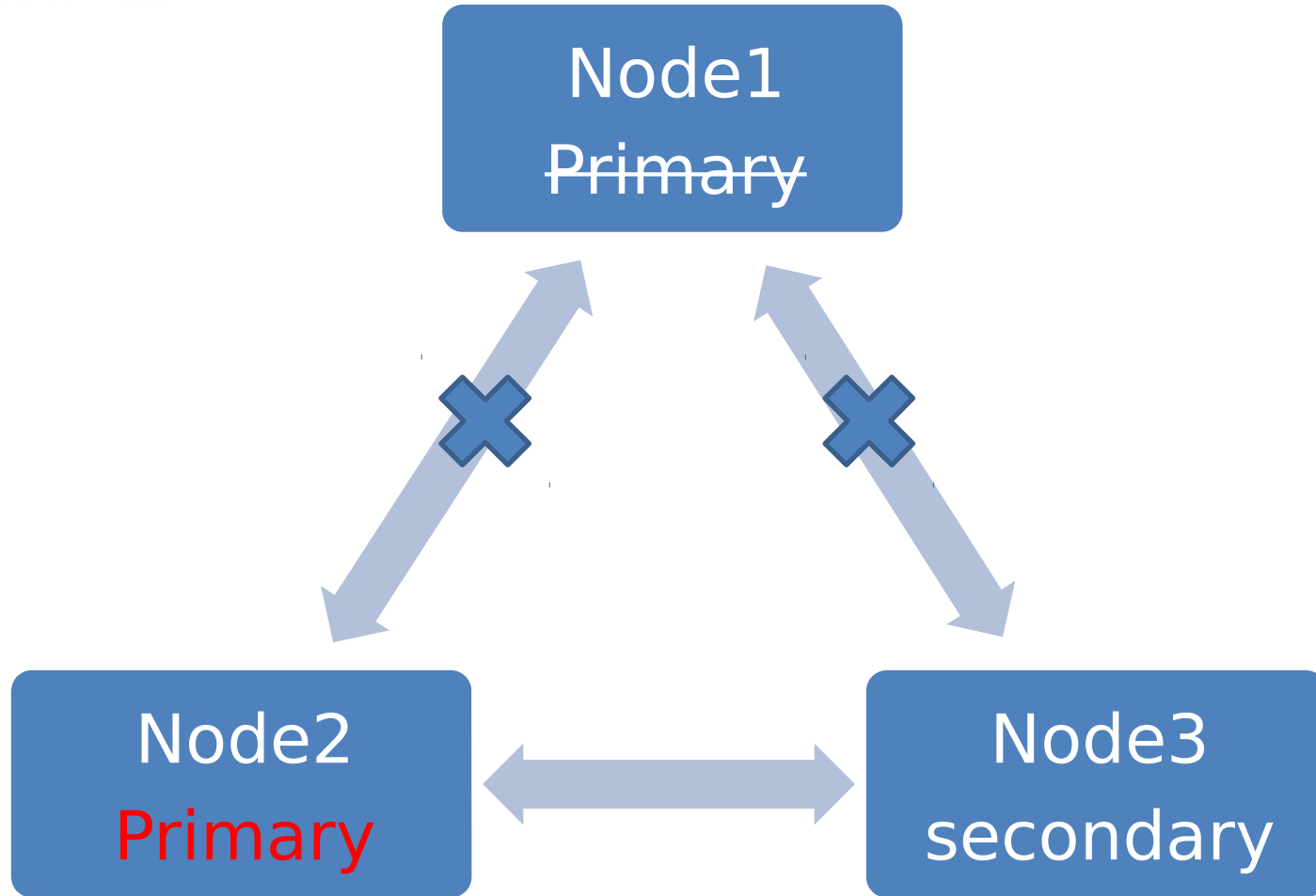


DeNA



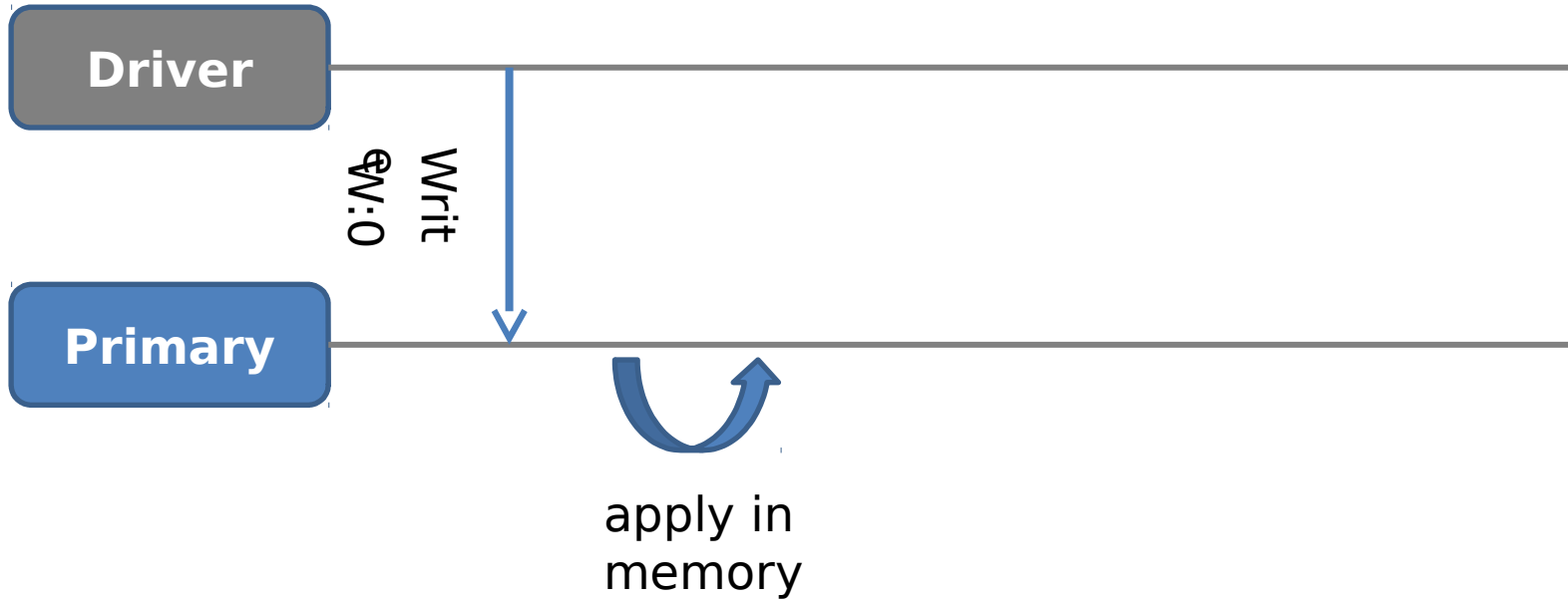
DeNA





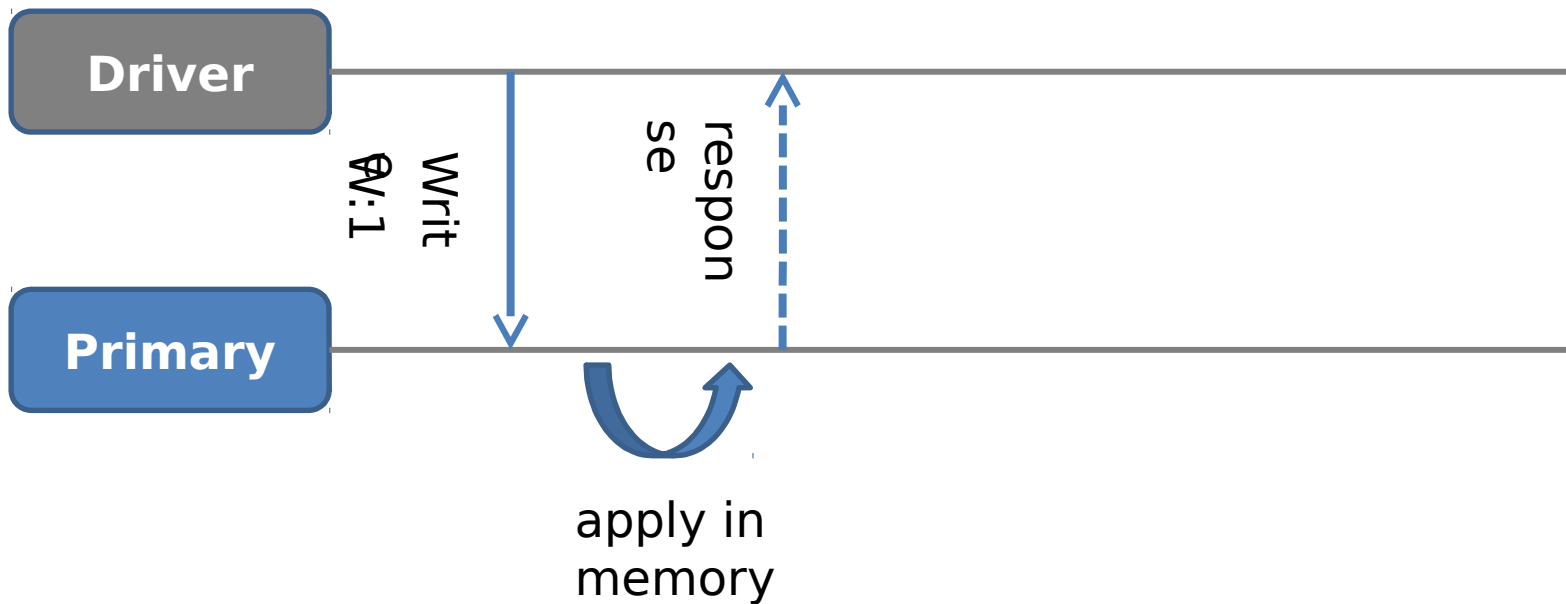
Write Concern (安全写级别)

1. Unacknowledged
2. Acknowledged
3. Wait for journal sync
4. Wait for replication(2,majority,all)



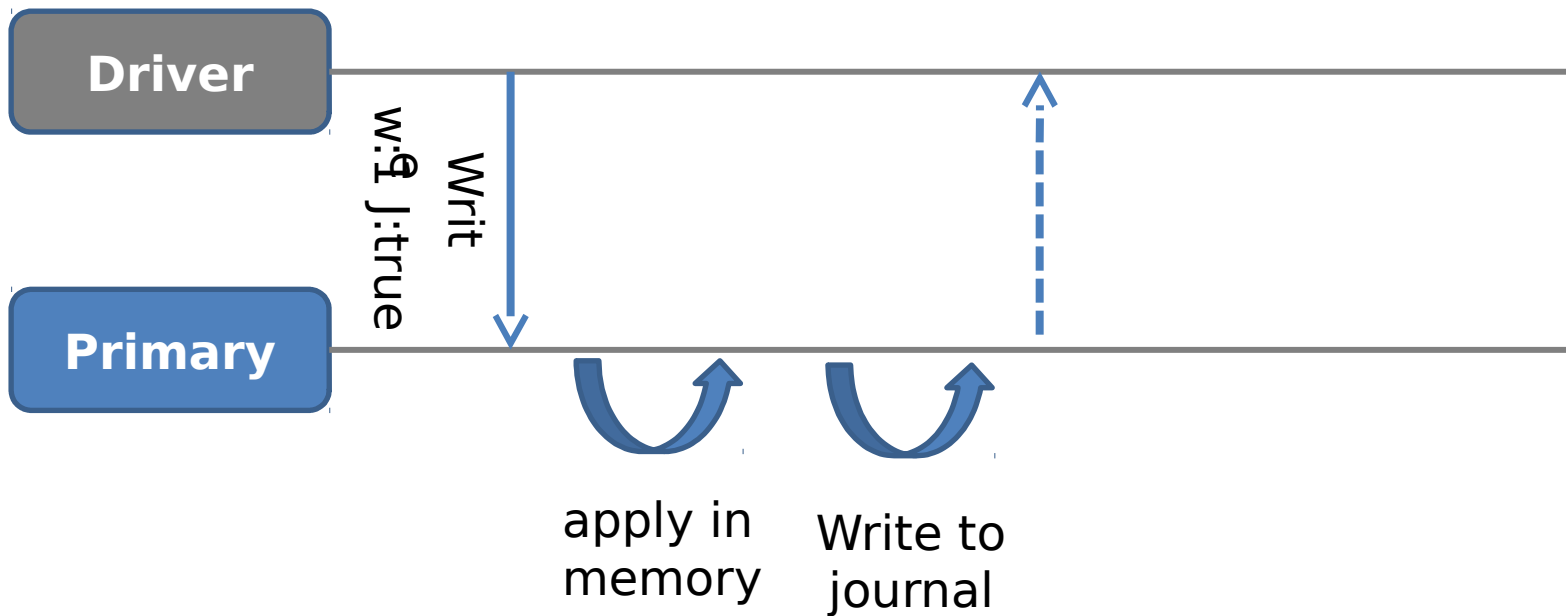
UNACKNOWLEDGED





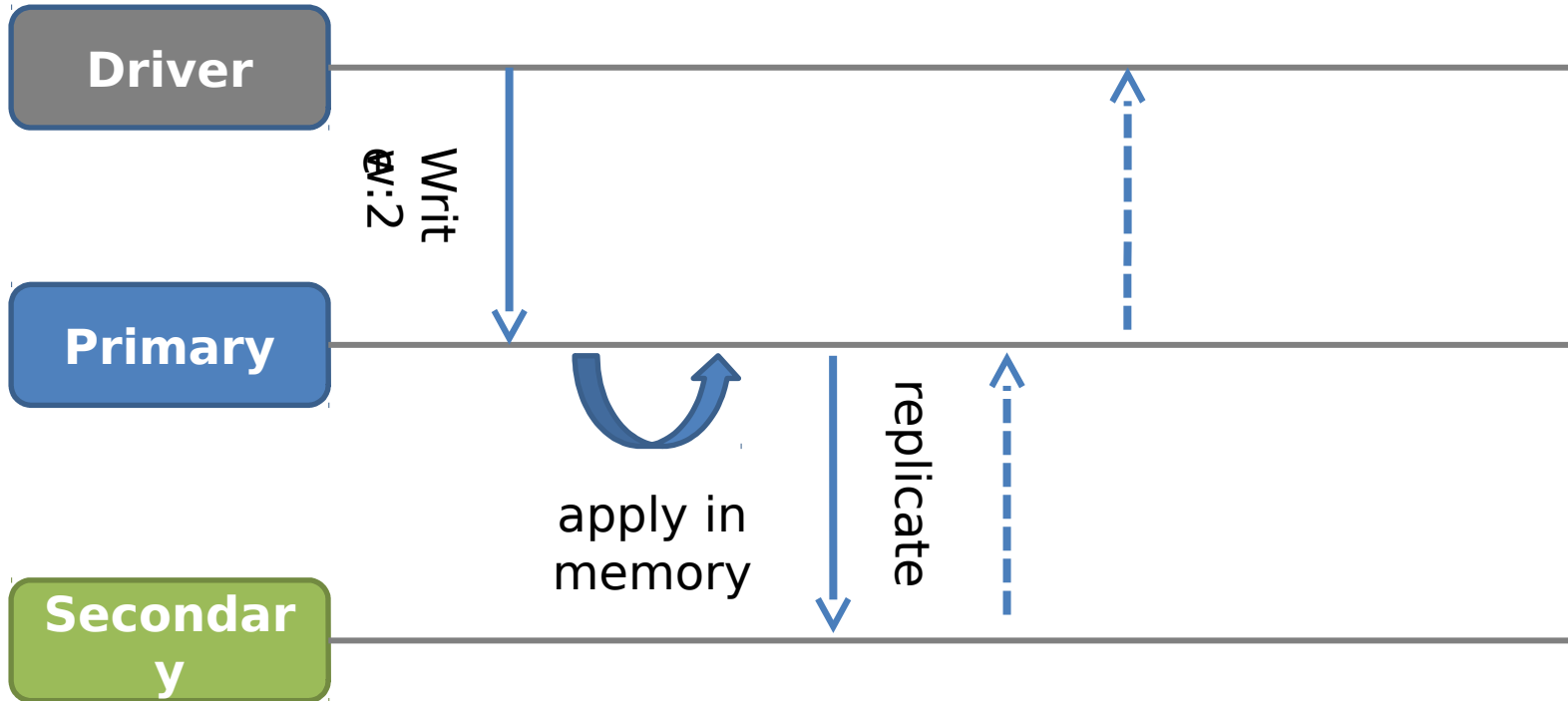
ACKNOWLEDGED





WAIT FOR JOURNAL SYNC

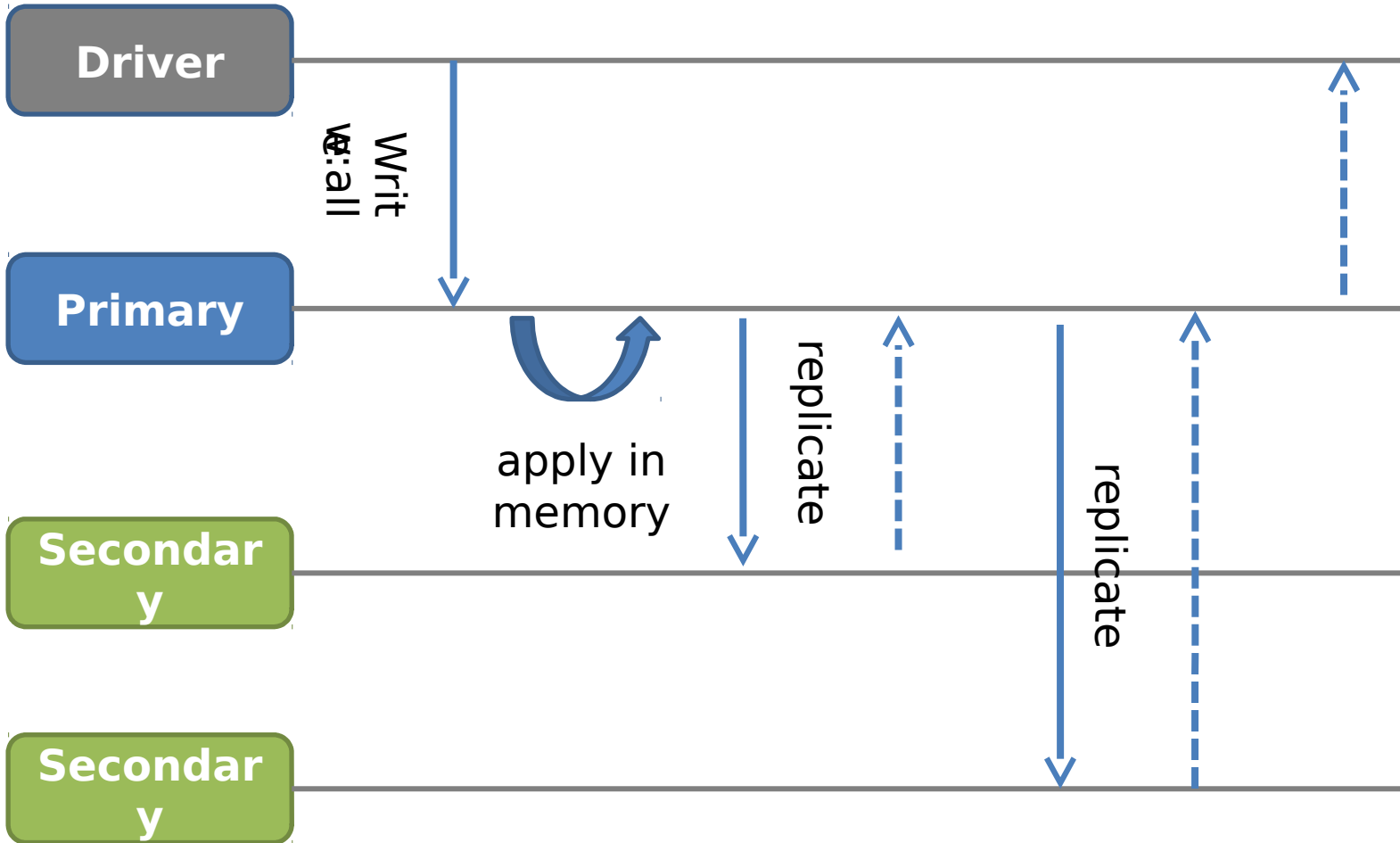




WAIT FOR REPLICATION



DeNA



:DeNA

END