

Reaching Zen in Your Cluster Coordination Layer

Philipp Krenn

@xeraa



Developer 🥑



elasticsearch

Concepts

**Cluster, Node, Index, Shard,
Document, ID**

Cluster Coordination?

Cluster State?

== Cluster Metadata

Cluster Settings

Index Metadata

Lots more

GET `_cluster/state`

Only move forward

Do **not lose data**

Demo

Cluster

Cluster State

3 Main Components

Discovery

Master Election

Cluster State Publication

Zen

Zen to Zen2

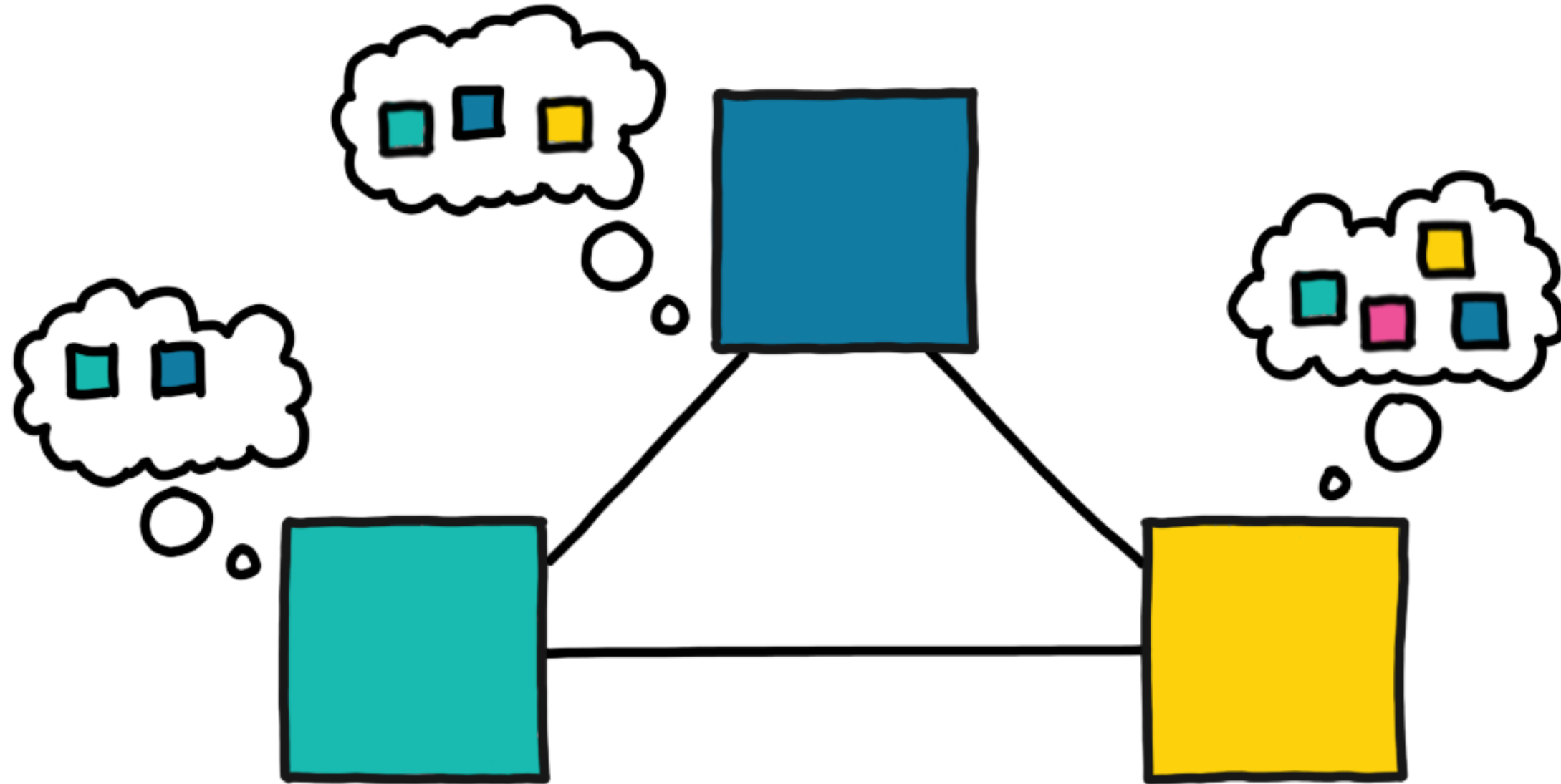
Not pluggable



Discovery

Where are master-eligible nodes?

Is there a master already?



Settings

`discovery.zen.ping.unicast.hosts -> discovery.seed_hosts`
static

`discovery.zen.hosts_provider -> discovery.seed_providers`
dynamic (File, EC2, GCE,...)

Master Election

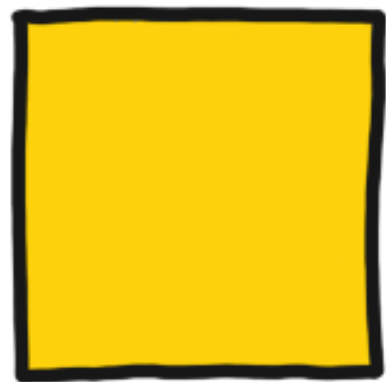
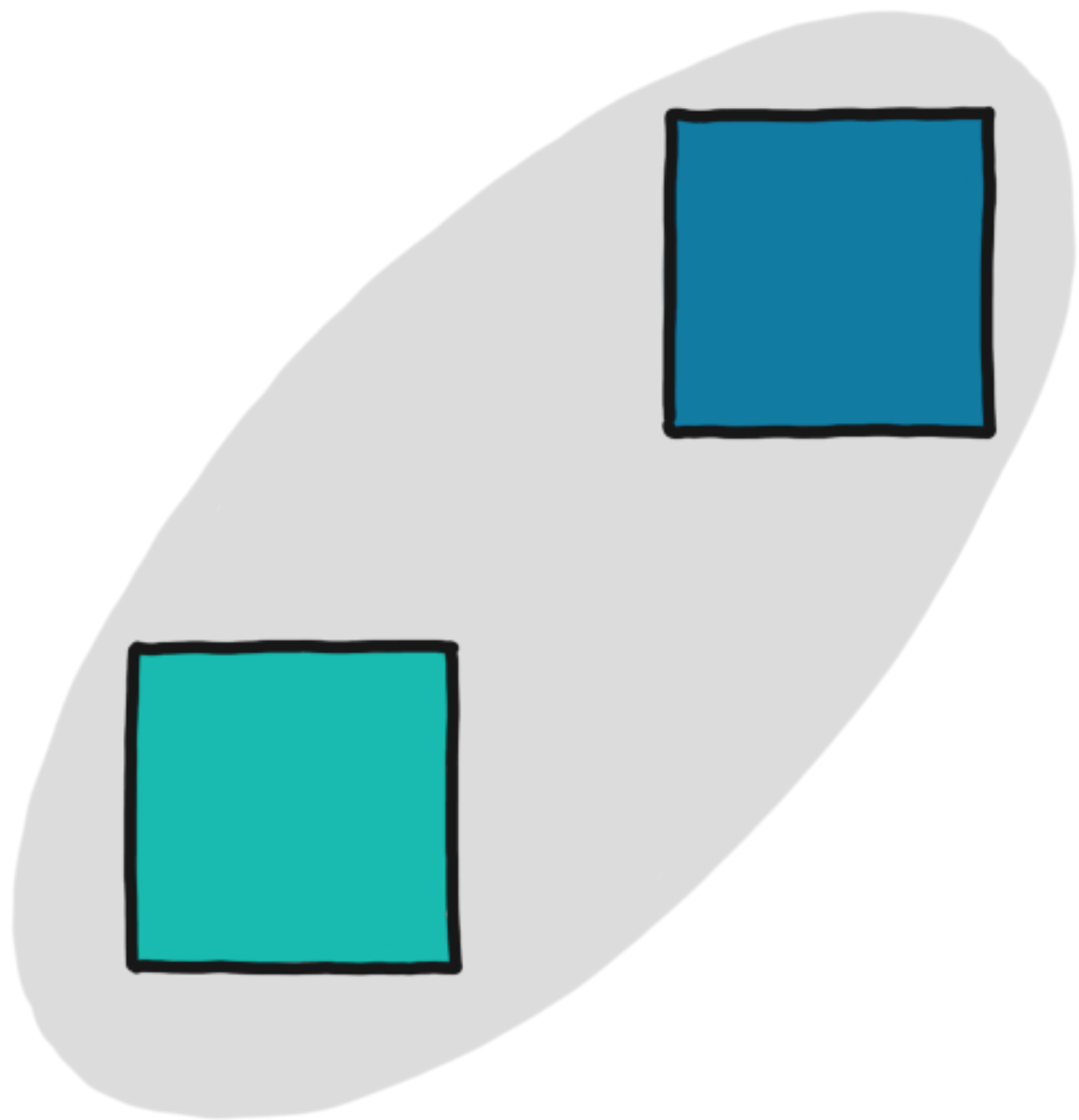
Agree which node should be master

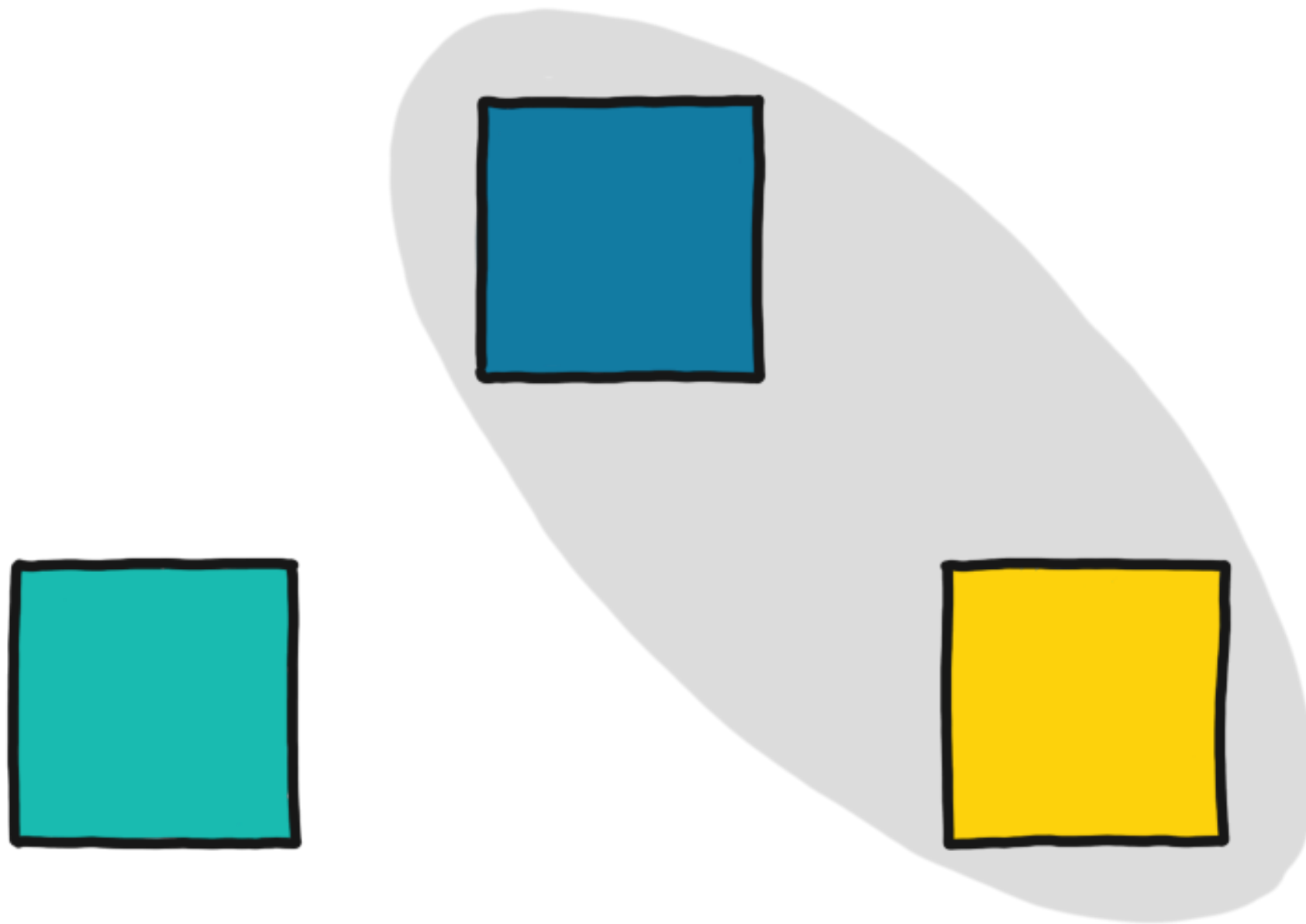
Form a cluster

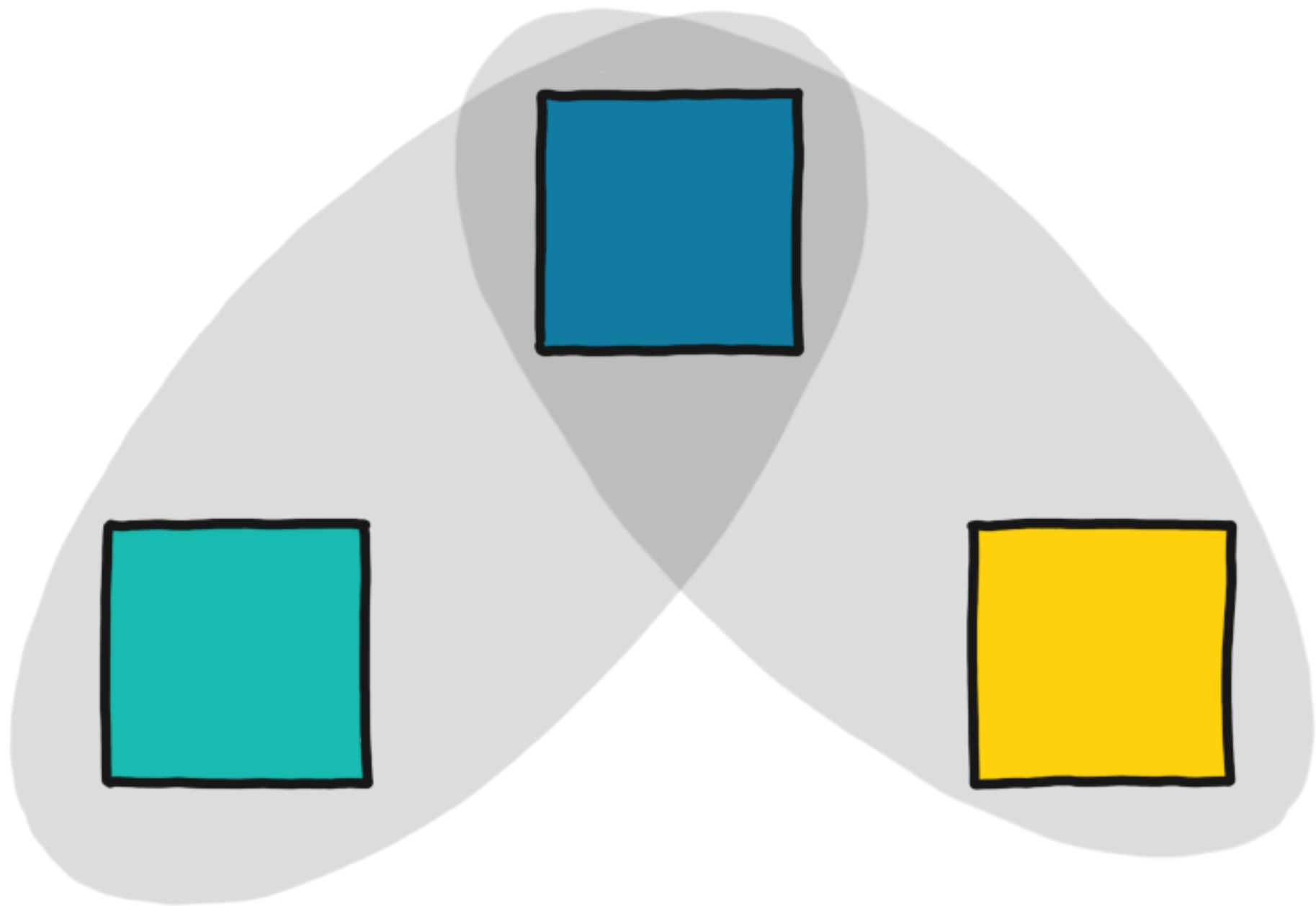


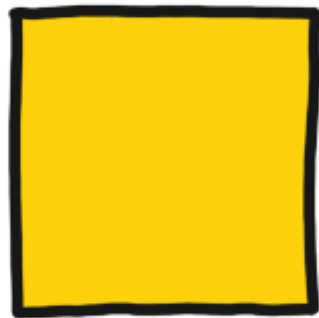
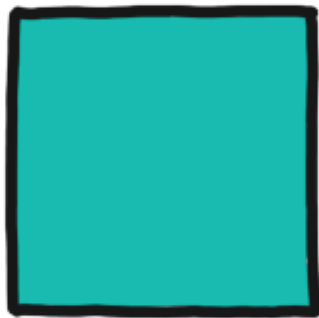
FOLLOW THE LEADER

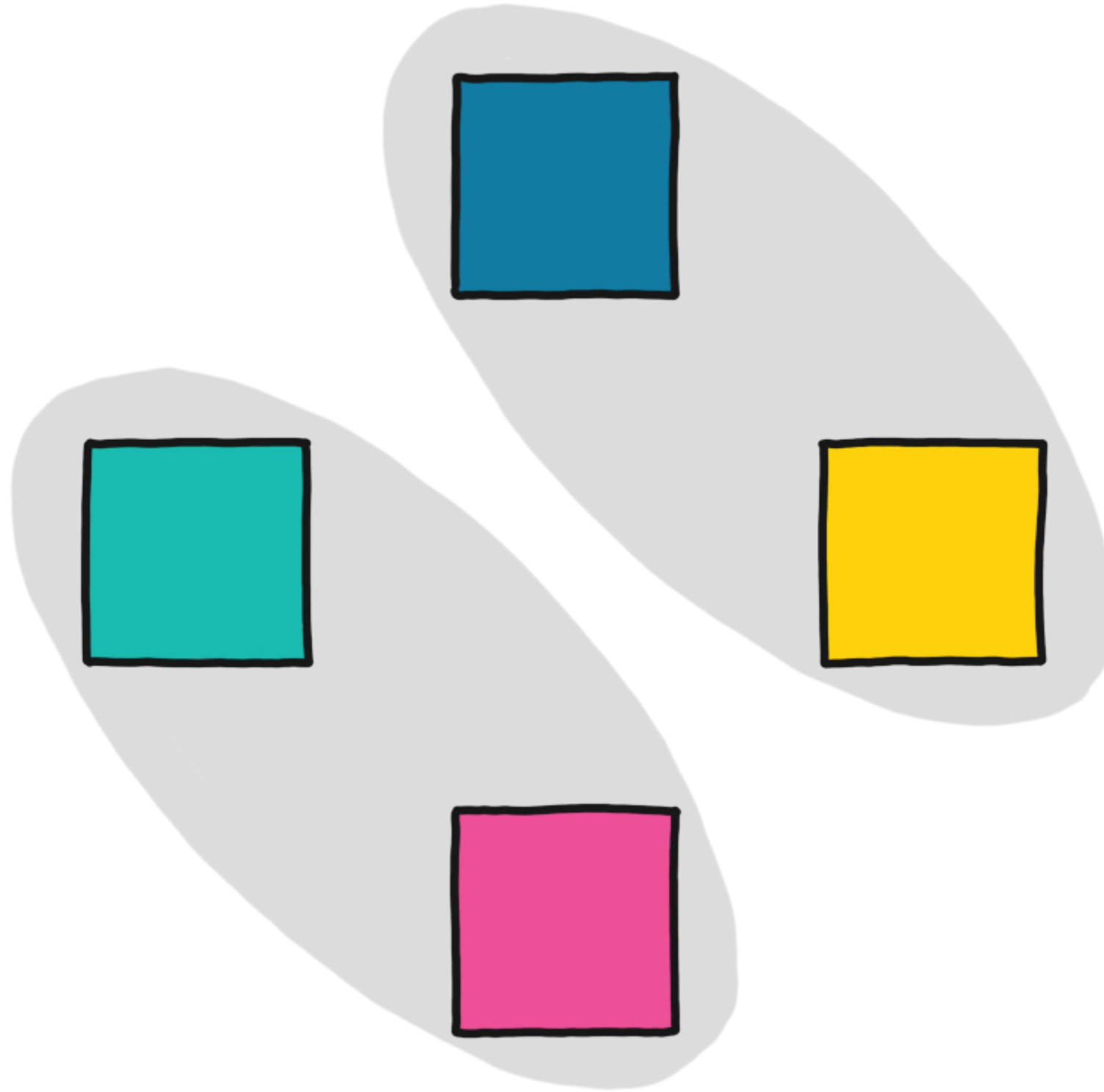












discovery.zen.

minimum_master_node

s

Trust users?

Demo

Upgrade

6.7 -> 7.0, 6.8 -> 7.1

Demo

Full Cluster Restart

[CTRL] + [C]

cluster.

initial_master_node
s

Nodes for the very first election

List of node names

**OK to set on multiple nodes as long
as they are all consistent**

**Ignored once node has joined a
cluster even if restarted**

**Unnecessary when joining new node
to existing cluster**

Upgrade 6 to 7

Full cluster restart: Set

`cluster.initial_master_nodes`

Rolling upgrade:

`cluster.initial_master_nodes` **not
required**

Demo

Cluster Rebuild

```
elasticsearch2 | {"type": "server",  
  "timestamp": "2019-05-24T14:02:51,173+0000",  
  "level": "WARN",  
  "component": "o.e.c.c.ClusterFormationFailureHelper",  
  "cluster.name": "docker-cluster",  
  "node.name": "elasticsearch2",  
  "message":
```

```
"master not discovered yet,  
this node has not previously joined a bootstrapped (v7+) cluster,  
and [cluster.initial_master_nodes] is empty on this node:  
have discovered [  
  {elasticsearch1}{pSUJ60tSRWSrcWkRevLfyA}{_jIaabgyTQOHA0jcwUruIQ}  
    {192.168.112.3}{192.168.112.3:9300}  
    {ml.machine_memory=1073741824, ml.max_open_jobs=20, xpack.installed=true},  
  {elasticsearch3}{ngaTCze8QHSHydCXsttXyw}{mbIad-A4SL0JvP7Ava5dEw}  
    {192.168.112.4}{192.168.112.4:9300}  
    {ml.machine_memory=1073741824, ml.max_open_jobs=20, xpack.installed=true}  
];  
discovery will continue using [192.168.112.3:9300, 192.168.112.4:9300] from hosts providers and [  
  {elasticsearch2}{iANt64LESxqjJv8tHV5KKw}{KobYEuQ2Tnamsi0efTUXgQ}  
    {192.168.112.2}{192.168.112.2:9300}  
    {ml.machine_memory=1073741824, xpack.installed=true, ml.max_open_jobs=20}  
]  
from last-known cluster state;  
node term 0, last-accepted version 0 in term 0"
```

Result: Safer Cluster Operations

Dynamic Cluster Scaling

Master-ineligible: same as before

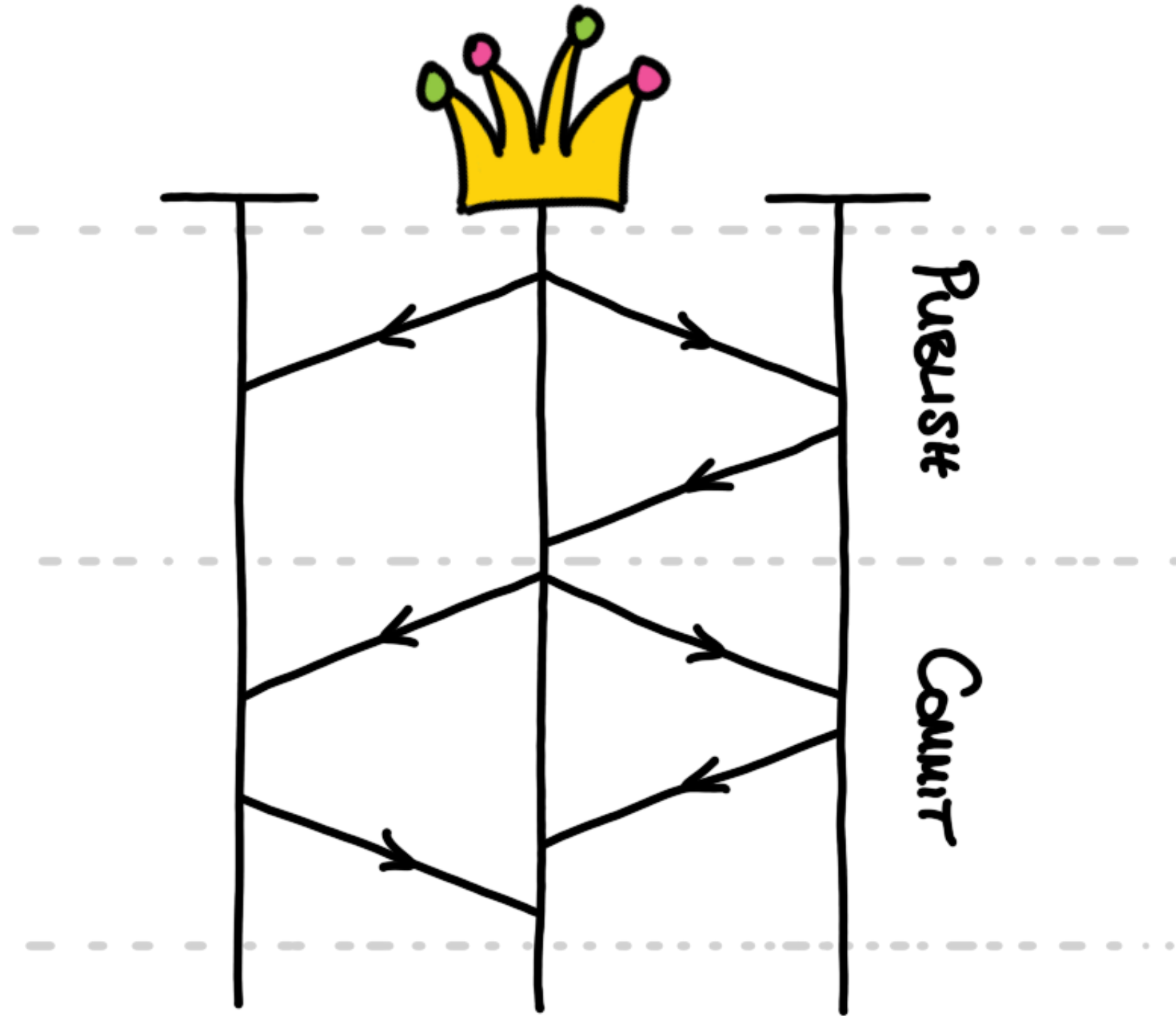
Adding master-eligible: Just do it

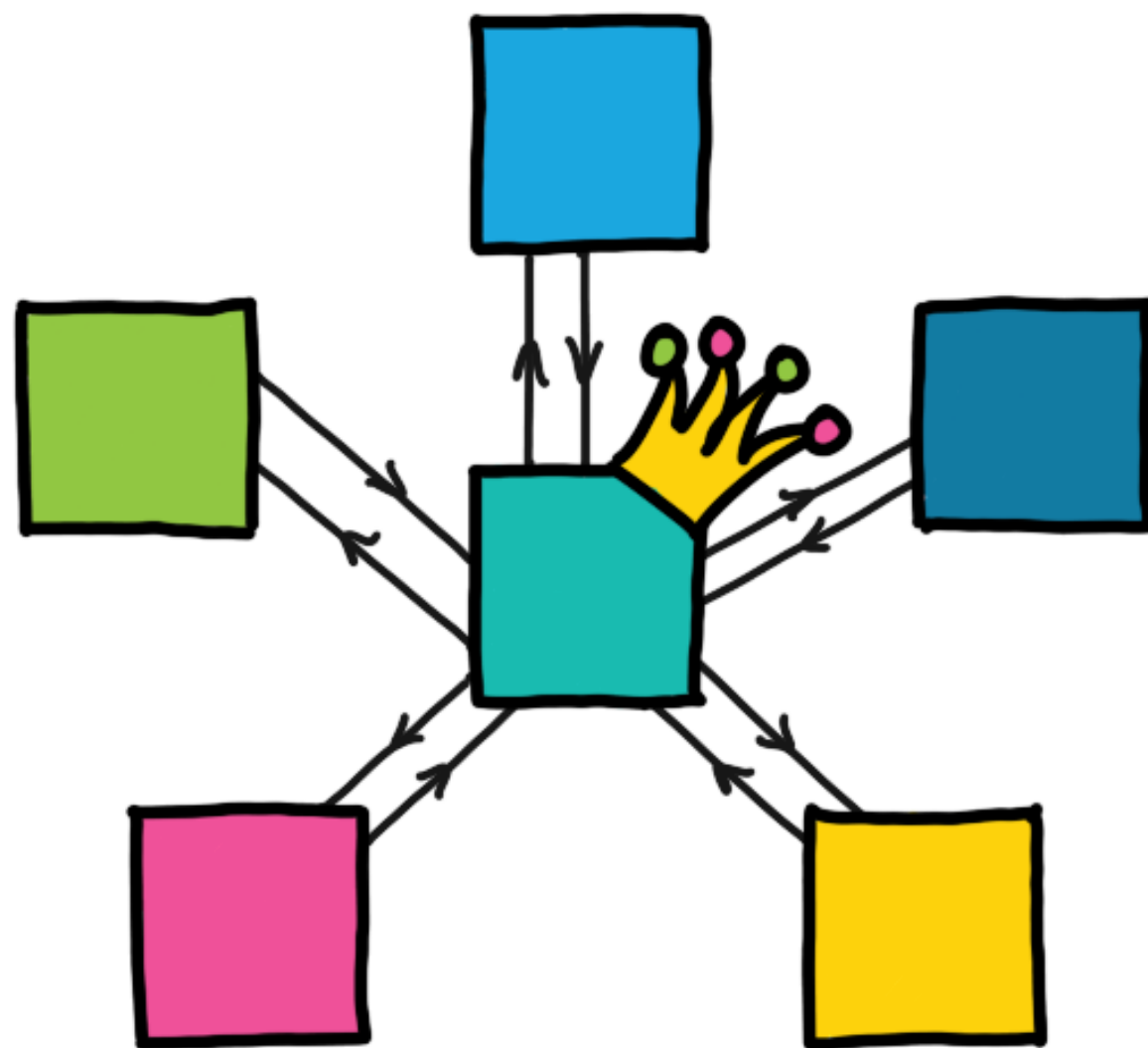
**Removing master-eligible: Just do it,
as long as you remove less than half**

Cluster State Publication

Agree on cluster state updates

Broadcast updates to all nodes





Conclusion

Zen to Zen2

Faster, safer, more debuggable

Reaching Zen in Your Cluster Coordination Layer

Philipp Krenn

@xeraa