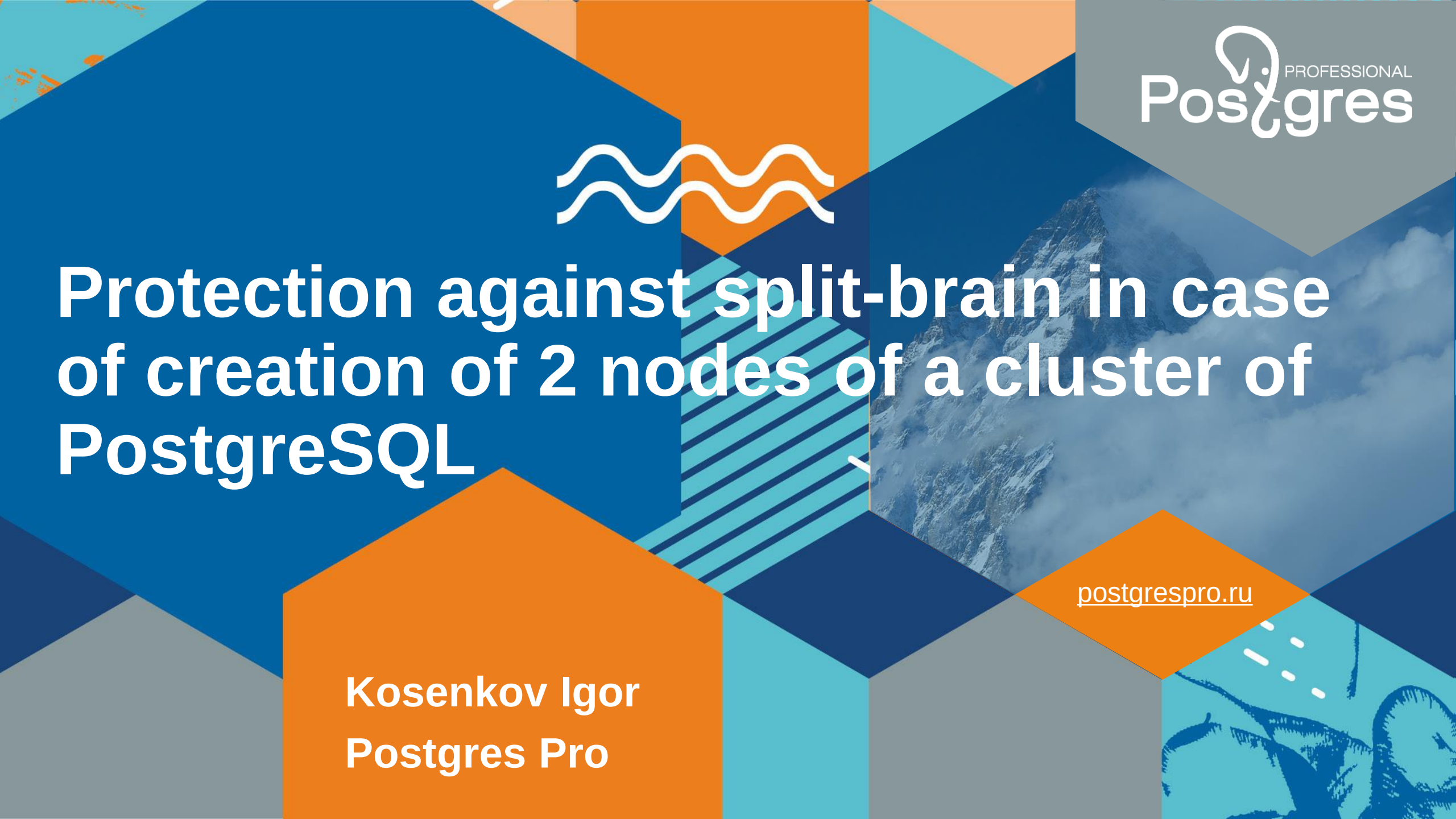
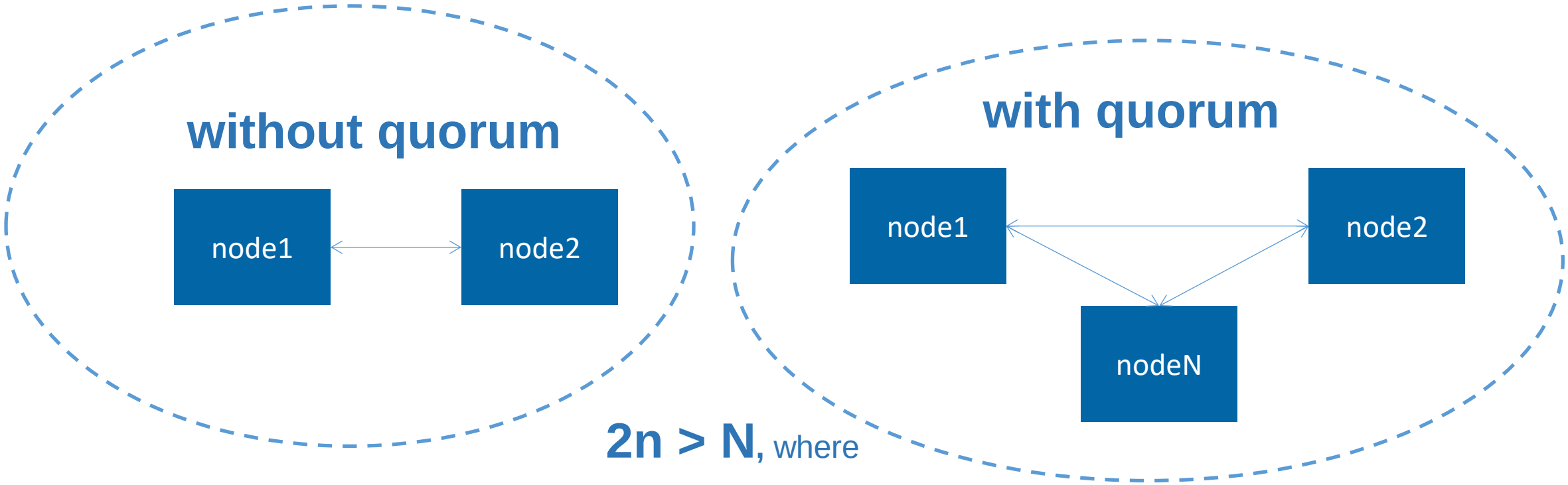


Protection against split-brain in case of creation of 2 nodes of a cluster of PostgreSQL

postgrespro.ru

Kosenkov Igor
Postgres Pro

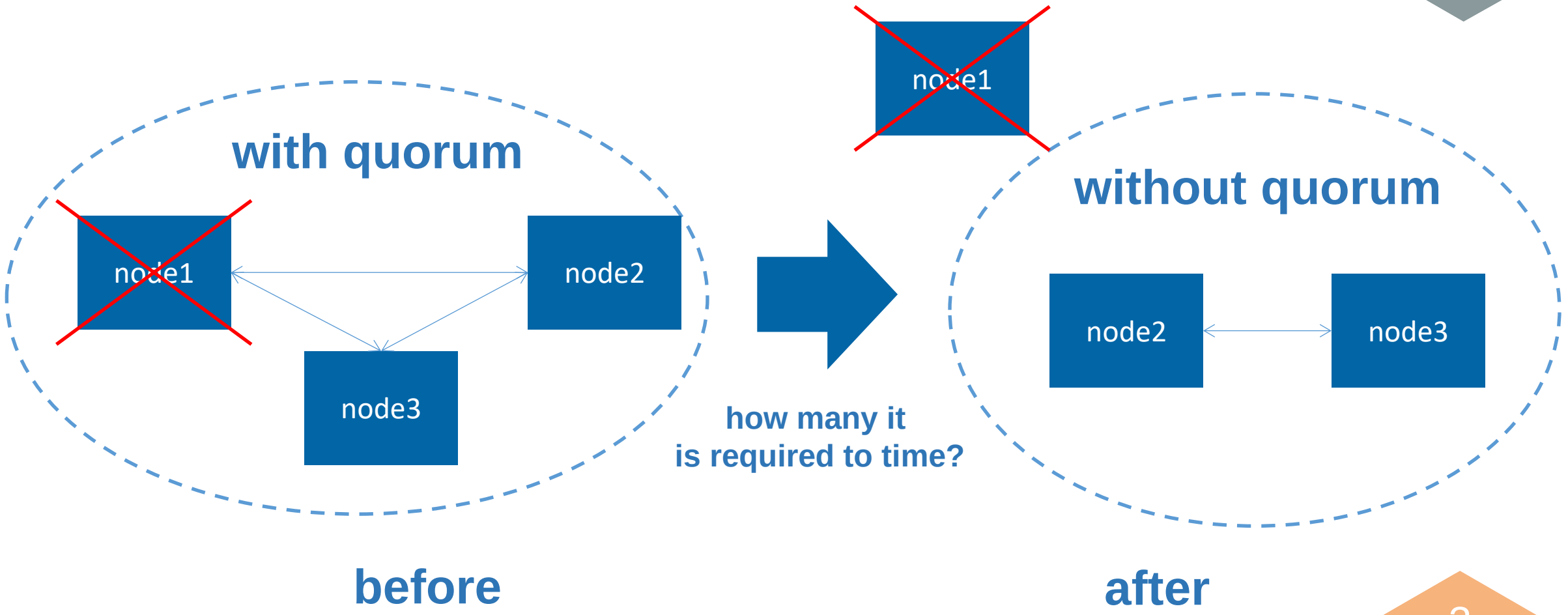
Types of failover clusters on the basis of Pacemaker&Corosync



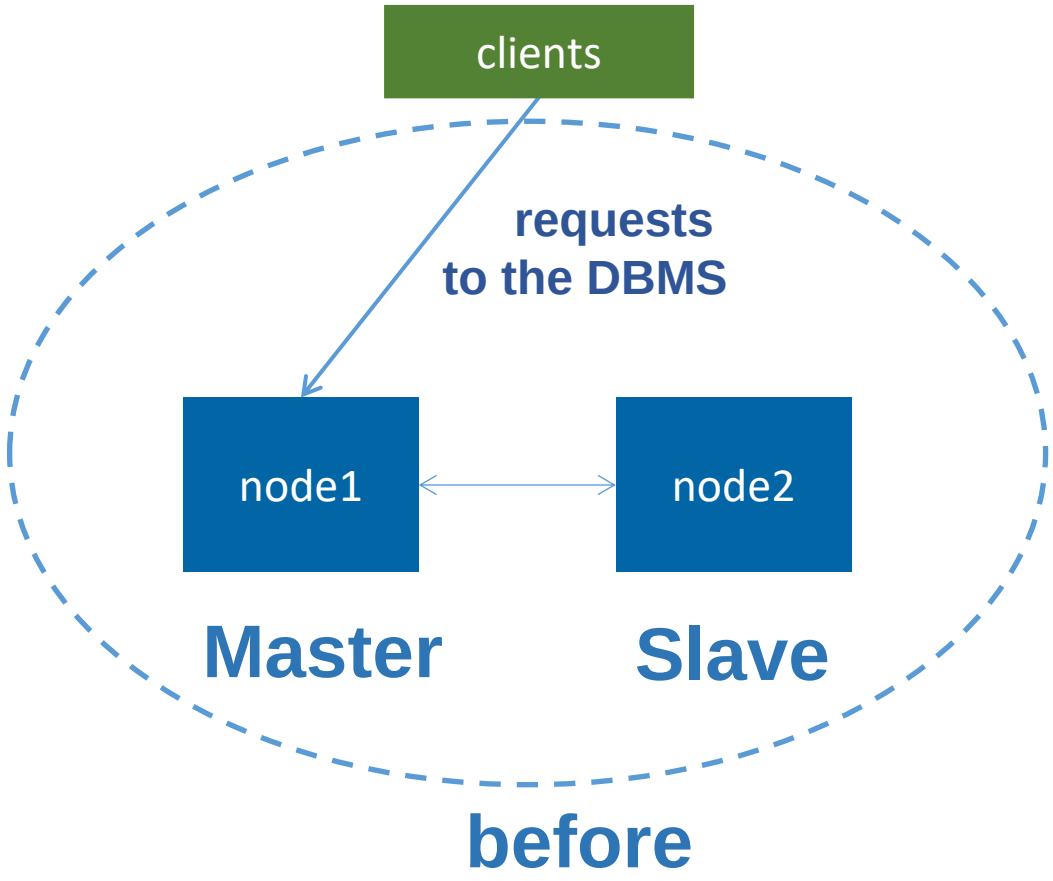
2n > N, where

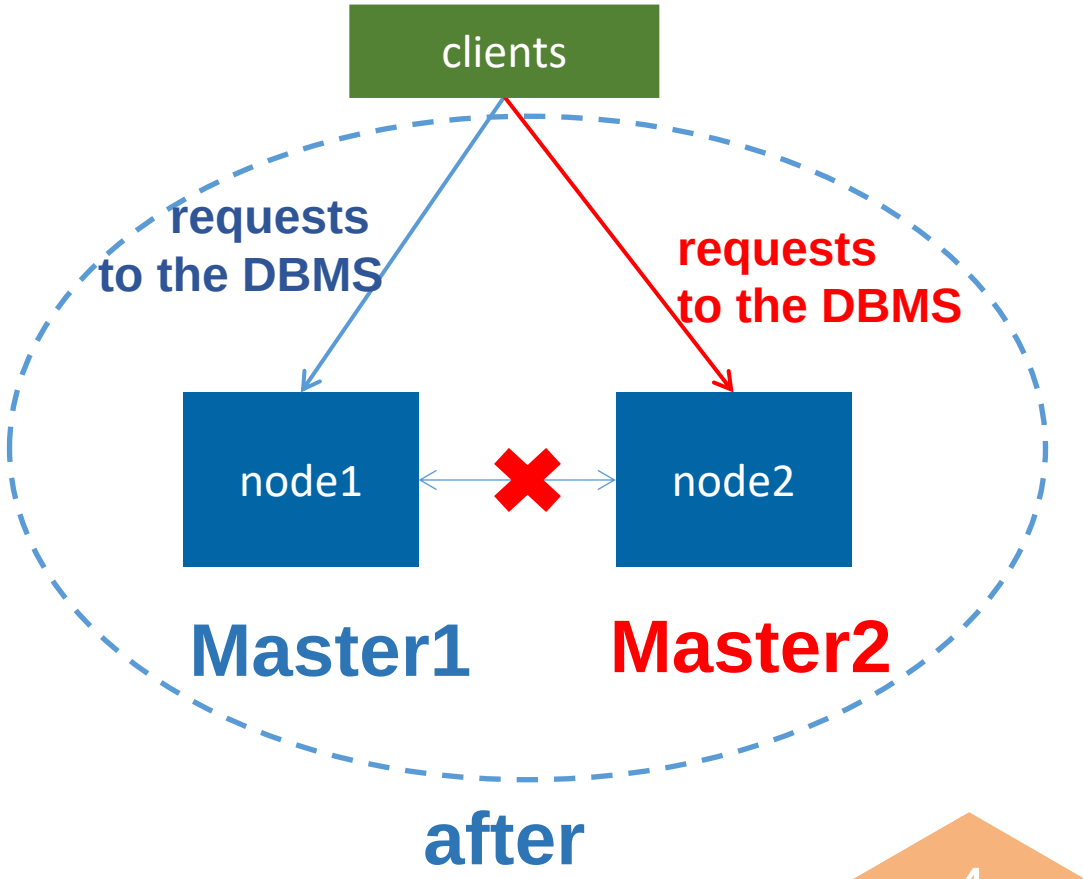
N – total of nodes in a cluster
n – the number of live nodes in a cluster

Practical application 2-nodes cluster

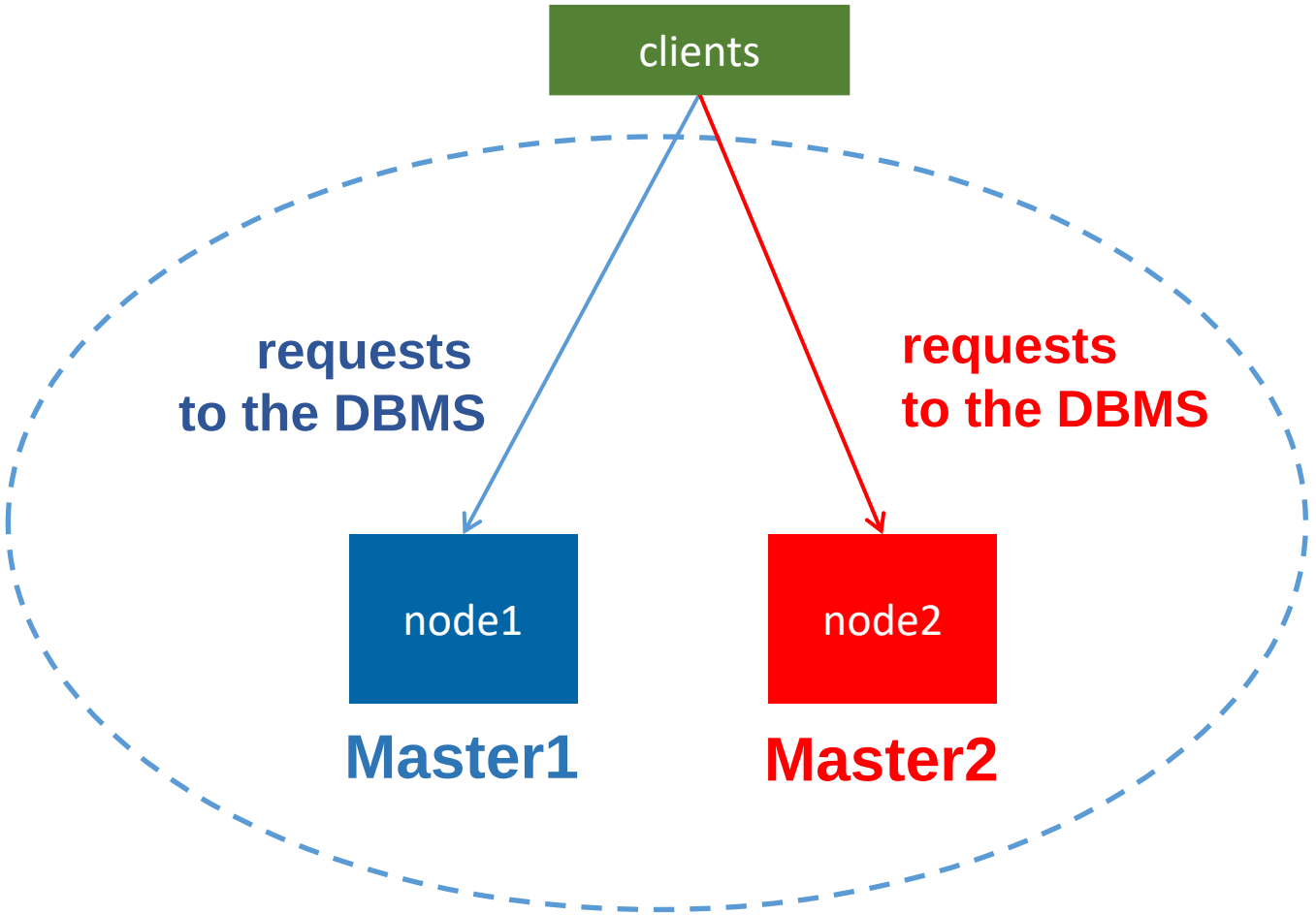


Loss of network connectivity between nodes





Aftermath of a split-brain



What do we have?

- **Two timelines of the DBMS**
- **Two identical virtual IP**

Known methods of protection against split-brain

PROFESSIONAL Postgres

Mechanism STONITH

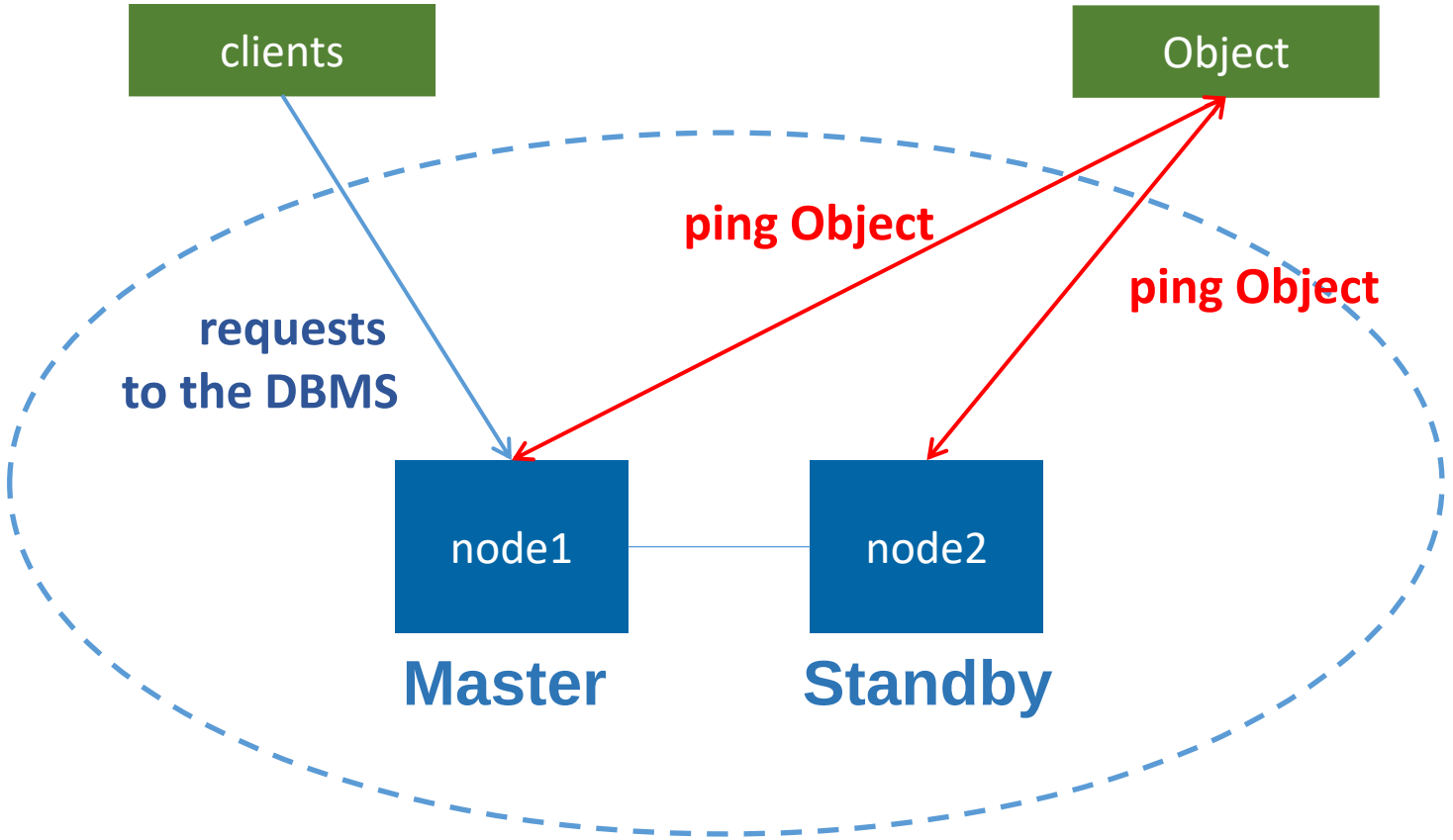
(Shot-The_Other-Node-In-The-Head)

Shortcoming – surely physical servers with the IPMI or ILO function

Switch-off of all resources of a cluster in case of loss of a network between nodes

Shortcoming – a failure in service

Protection against split-brain in case of creation of 2 nodes of a cluster of PostgreSQL



Description of protection against split-brain

Add a resource the Ping type – “default_ping_set “
The resource has type a clone

Change value resource-stickiness to 500
Value by default = INFINITY

Specify a role at failure
When a node in isolation, the default_ping_set value is equal to 111
Behavior of the node: If default_ping_set =111 then role=slave and score=-INFINITY

Normal state of a cluster crm_mon –Afr output

```
Last updated: Wed Oct 18 17:29:01 2017          Last change: Wed Oct 18 17:28:38 2017 by root via crm_attribute
on pgsql01
Stack: corosync
Current DC: pgsql02 (version 1.1.13-10.el7_2.4-44eb2dd) - partition with quorum
2 nodes and 6 resources configured

Online: [ pgsql01 pgsql02 ]

Full list of resources:

 Clone Set: pingCheck1-clone [pingCheck1]
     Started: [ pgsql01 pgsql02 ]
 Master/Slave Set: msPostgresql [pgsql]
     Masters: [ pgsql01 ]
     Slaves: [ pgsql02 ]
 Resource Group: master-group
     vip-master (ocf::heartbeat:IPaddr2):       Started pgsql01
     vip-rep    (ocf::heartbeat:IPaddr2):       Started pgsql01

Node Attributes:
* Node pgsql01:
    + default_ping_set                  : 333
    + master-pgsql                      : 1000
    + pgsql-data-status                 : LATEST
    + pgsql-master-baseline             : 00000024BD000098
    + pgsql-status                      : PRI
* Node pgsql02:
    + default_ping_set                  : 333
    + master-pgsql                      : 100
    + pgsql-data-status                 : STREAMING|SYNC
    + pgsql-status                      : HS:sync

Migration Summary:
* Node pgsql02:
* Node pgsql01:
```

Loss of network connectivity between nodes (output node1)

```
Last updated: Wed Oct 18 17:32:30 2017          Last change: Wed Oct 18 17:32:07 2017 by root via crm_attribu
te on pgsql01
Stack: corosync
Current DC: pgsql01 (version 1.1.13-10.el7_2.4-44eb2dd) - partition with quorum
2 nodes and 6 resources configured

Online: [ pgsql01 ]
OFFLINE: [ pgsql02 ]

Full list of resources:

 Clone Set: pingCheck1-clone [pingCheck1]
     Started: [ pgsql01 ]
     Stopped: [ pgsql02 ]
 Master/Slave Set: msPostgresql [pgsql]
     Stopped: [ pgsql01 pgsql02 ]
 Resource Group: master-group
     vip-master (ocf::heartbeat:IPaddr2):       Stopped
     vip-rep    (ocf::heartbeat:IPaddr2):       Stopped

Node Attributes:
* Node pgsql01:
    + default_ping_set                  : 111           : Connectivity is degraded (Expected=333)
    + master-pgsql                      : -INFINITY
    + pgsql-data-status                 : LATEST
    + pgsql-status                      : STOP

Migration Summary:
* Node pgsql01:
```

Loss of network connectivity between nodes (output node2)

```
Last updated: Wed Oct 18 17:33:09 2017          Last change: Wed Oct 18 17:32:06 2017 by root via crm
_attribute on pgsql02
Stack: corosync
Current DC: pgsql02 (version 1.1.13-10.el7_2.4-44eb2dd) - partition with quorum
2 nodes and 6 resources configured

Online: [ pgsql02 ]
OFFLINE: [ pgsql01 ]

Full list of resources:

 Clone Set: pingCheck1-clone [pingCheck1]
     Started: [ pgsql02 ]
     Stopped: [ pgsql01 ]
 Master/Slave Set: msPostgresql [pgsql]
     Masters: [ pgsql02 ]
     Stopped: [ pgsql01 ]
 Resource Group: master-group
     vip-master (ocf::heartbeat:IPaddr2):       Started pgsql02
     vip-rep    (ocf::heartbeat:IPaddr2):       Started pgsql02

Node Attributes:
* Node pgsql02:
    + default_ping_set                  : 222           : Connectivity is degraded (Expected=333)
    + master-pgsql                      : 1000
    + pgsql-data-status                 : LATEST
    + pgsql-master-baseline             : 00000024BD0001E0
    + pgsql-status                      : PRI

Migration Summary:
* Node pgsql02:
```

After restoration of a network between nodes

```
Last updated: Wed Oct 18 17:33:45 2017          Last change: Wed Oct 18 17:32:06 2017 by root via crm
 attribute on pgsql02
Stack: corosync
Current DC: pgsql02 (version 1.1.13-10.el7_2.4-44eb2dd) - partition with quorum
2 nodes and 6 resources configured

Online: [ pgsql01 pgsql02 ]

Full list of resources:

 Clone Set: pingCheck1-clone [pingCheck1]
     Started: [ pgsql01 pgsql02 ]
 Master/Slave Set: msPostgresql [pgsql]
     Masters: [ pgsql02 ]
     Slaves: [ pgsql01 ]
 Resource Group: master-group
     vip-master (ocf::heartbeat:IPaddr2):       Started pgsql02
     vip-rep    (ocf::heartbeat:IPaddr2):       Started pgsql02

Node Attributes:
* Node pgsql01:
    + default_ping_set                 : 333
    + master-pgsql                     : -INFINITY
    + pgsql-data-status                : DISCONNECT
    + pgsql-status                     : HS:alone
* Node pgsql02:
    + default_ping_set                 : 333
    + master-pgsql                     : 1000
    + pgsql-data-status                : LATEST
    + pgsql-master-baseline            : 00000024BD0001E0
    + pgsql-status                     : PRI
* Node pgsql02:* Node pgsql01:
```

Recovery of a cluster after failure

Following steps:

1.To stop a cluster on a failure node a command:

sudo pcs cluster stop

2.To clean a directory $PGDATA (run as user «postgres»)

3. To copy database directory contents from the Master-server the utility of pg_basebackup (run as user «postgres»)

4. To launch a cluster on a failure node a command:

sudo pcs cluster start

Postgres Professional

http://postgrespro.ru/

+7(495)1500691

info@postgrespro.ru

postgrespro.ru