PostgreSQL Split-Brain

What is it and why does it happen?

Author: RadixTrie Open-Source Team

The Split-Brain Syndrome.

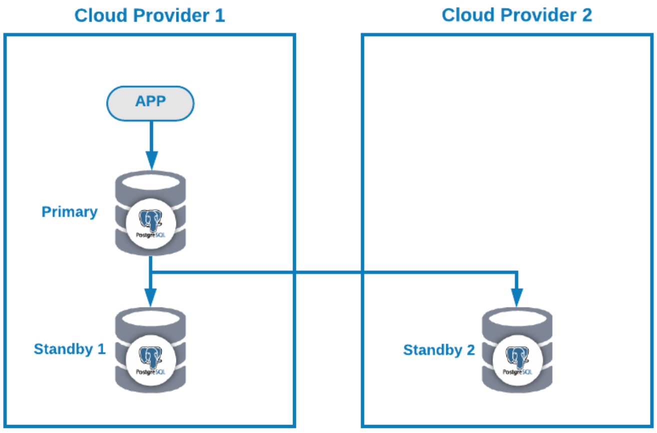
Segal's law states that "A man with a watch knows what time it is. A man with two watches is never sure." Similar considerations apply to databases,

What is Split-Brain?

In the PostgreSQL world, split-brain occurs when more than one primary node is available at the same time (without any third-party tool to have a multi-master environment) that allows the application to write in both nodes. In this case, you'll have different information on each node, which generates data inconsistency in the cluster. Fixing this issue could be hard as you must merge data, something that sometimes is not possible.

What causes Split-Brain?

Common Topology:



If your primary node is down, one of the standby nodes should be promoted as a new primary and you should change the IP address in your application to use this new primary node. There are different ways to make this an automatic way.

You can use a virtual IP address assigned to your primary node and monitor it. If it fails, promote one of the standby nodes and migrate the virtual IP address to the new primary node. This can be done using your own script or tool.

If your old primary node comes back, you must make sure you won't have two primary nodes in the same cluster at the same time.

The most common methods to avoid this situation are:

- STONITH: Shoot The Other Node In The Head.
- SMITH: Shoot Myself In The Head.

You can improve your topology by adding a Load Balancer (HAProxy), which you can also do using ClusterControl.

Monitoring.

When HAProxy detects that one of the nodes, either primary or standby, is not accessible, it automatically marks it as offline and does not take it into account for sending traffic to it. This check is done by health check scripts that are configured by ClusterControl at the time of deployment. These check whether the instances are up, whether they are undergoing recovery, or are read-only.

If your old primary node comes back, ClusterControl will also avoid starting it, to prevent a potential split-brain in case you have a direct connection that is not using the Load Balancer, but you can add it to the cluster as a standby node in an automatic or manual way using the ClusterControl UI or CLI, then you can promote it to have the same topology that you had running before the issue.

Summary.

It is important to note that in the event of a Split-Brain the longer both masters are up with read-write capabilities the more data will need to be synchronized to avoid any data loss.

In the read-only port, you have both the primary and the standby nodes online. In this way, you can balance the reading traffic between your nodes but you make sure that at the time of writing, the read-write port will be used, writing in the primary node that is the server that is online.

Disclaimer: It is always recommended to log a support request with the vendor for any errors, anomalies you may encounter in your environment.