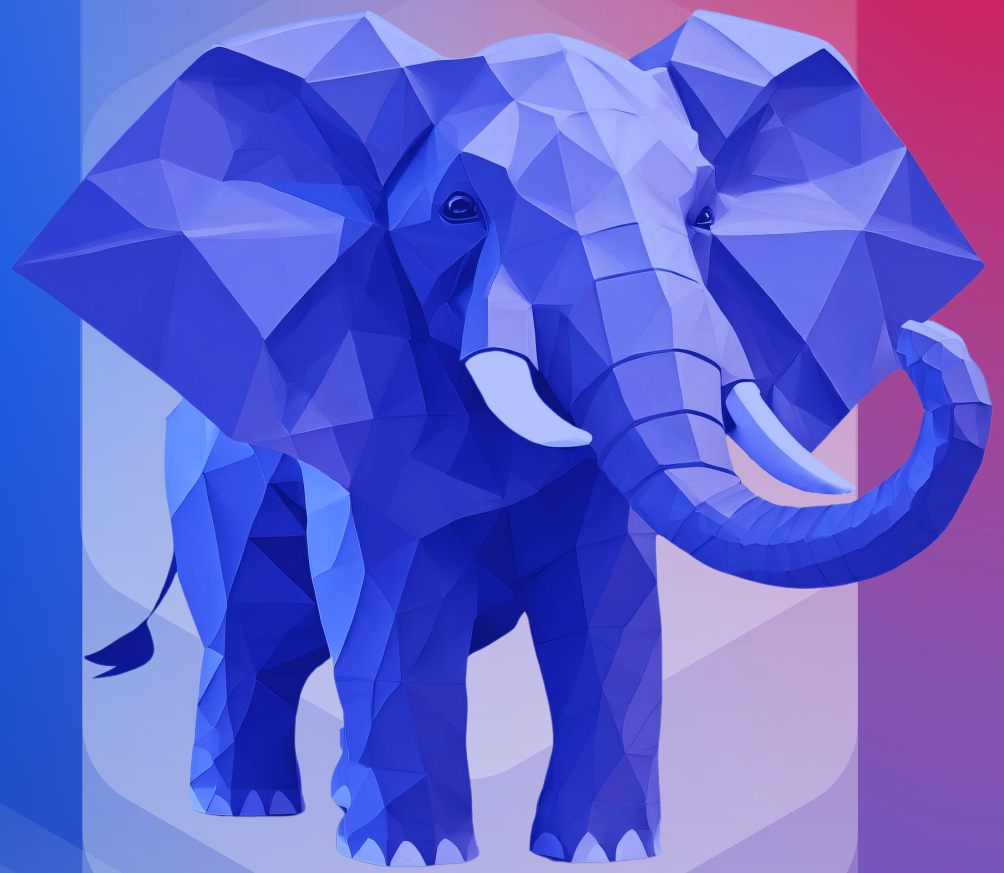
PostgresPro

Reliability features in Postgres Pro: BiHA, load balancing and backups



PGConf India 2025

Leonid Albrekht, l.albrekht@postgrespro.ru

Leonid Albrekht

PostgresPro

@LALBREKHT

Roadmap

PostgresPro

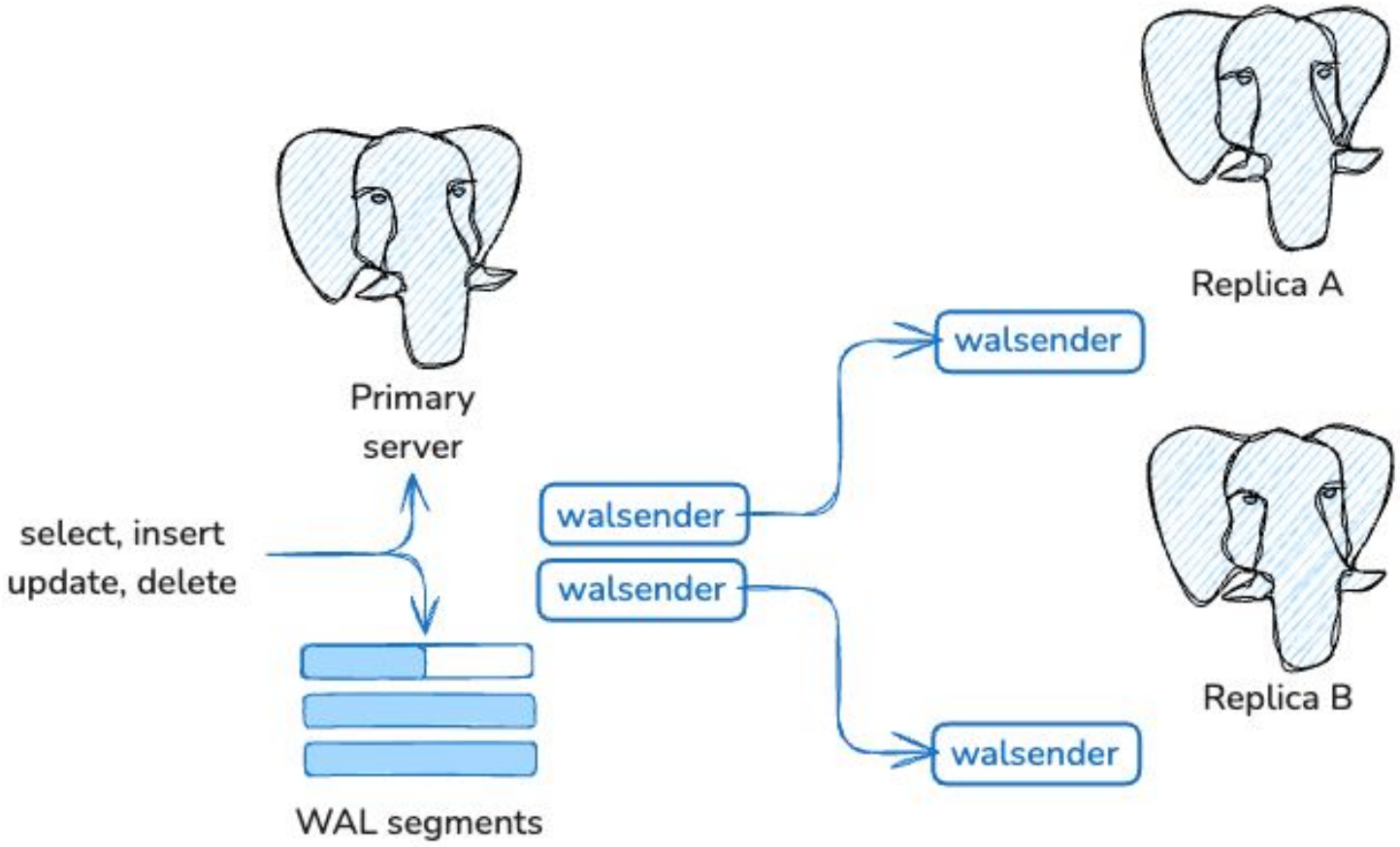
- Reminder about Postgres Physical replication
- What is Postgres Pro BiHA - Build-In High Availability
- BiHA Architecture
- Failover, switchover and network isolation
- Benefits of BiHA
- Clients connections and load balancing
- Backup options

Postgres Professional

PostgresPro

BiHA overview

Physical replication



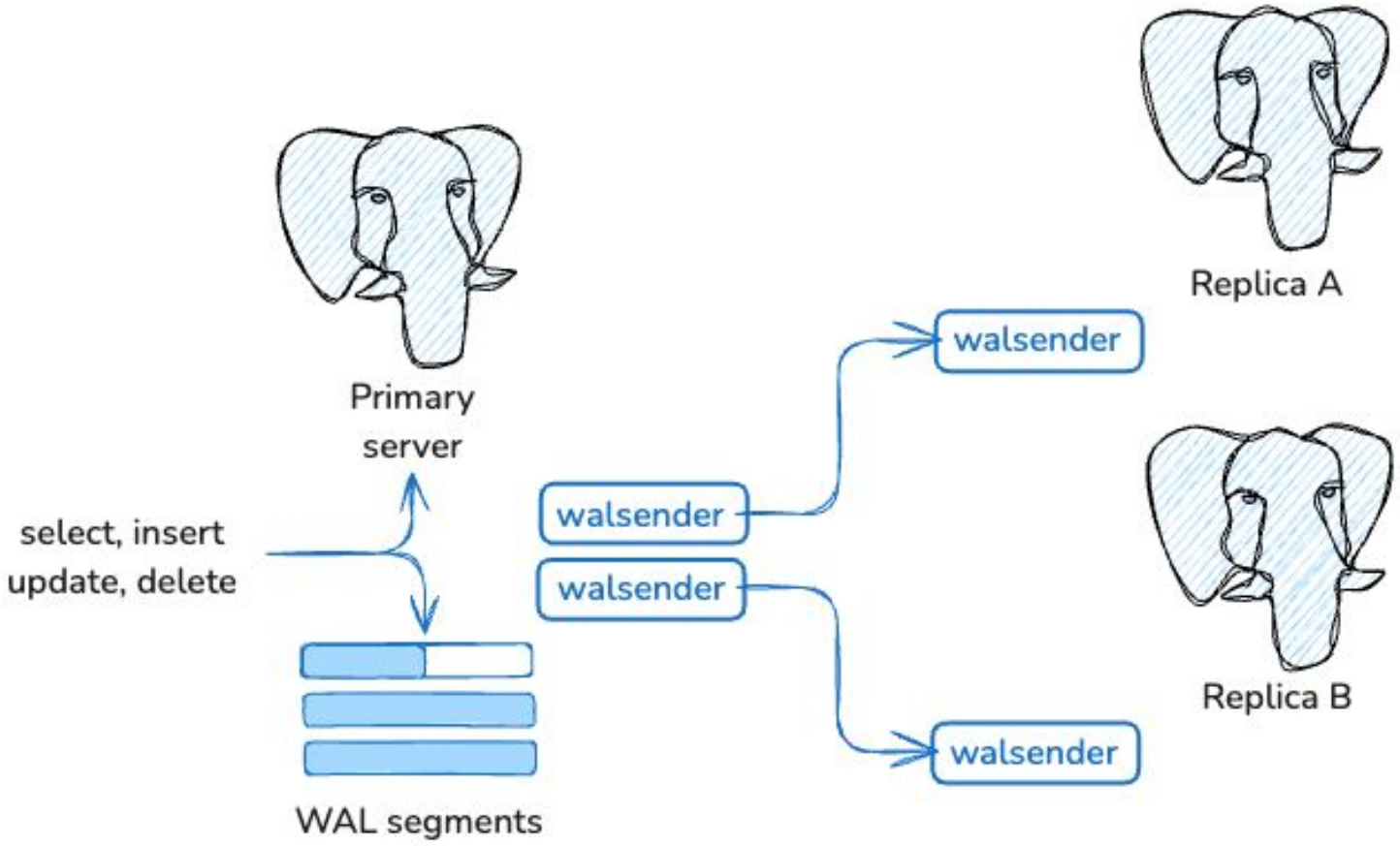
Replication

The process to synchronizing multiple copies of a database cluster on different servers

Purpose

- Reliability if one of the servers fails, the system must maintain availability
- Scalability
- Load distribution between server

Physical replication



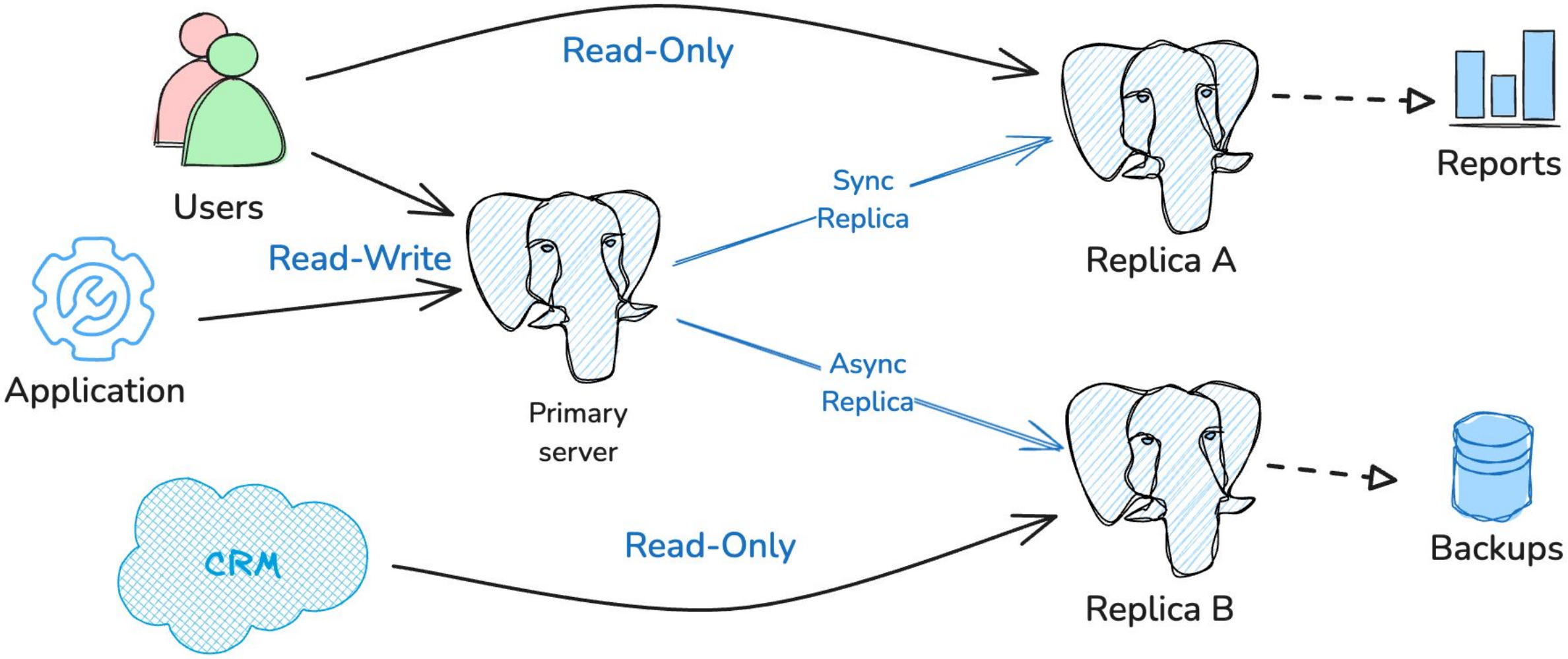
Physical replication

Primary -> Replica

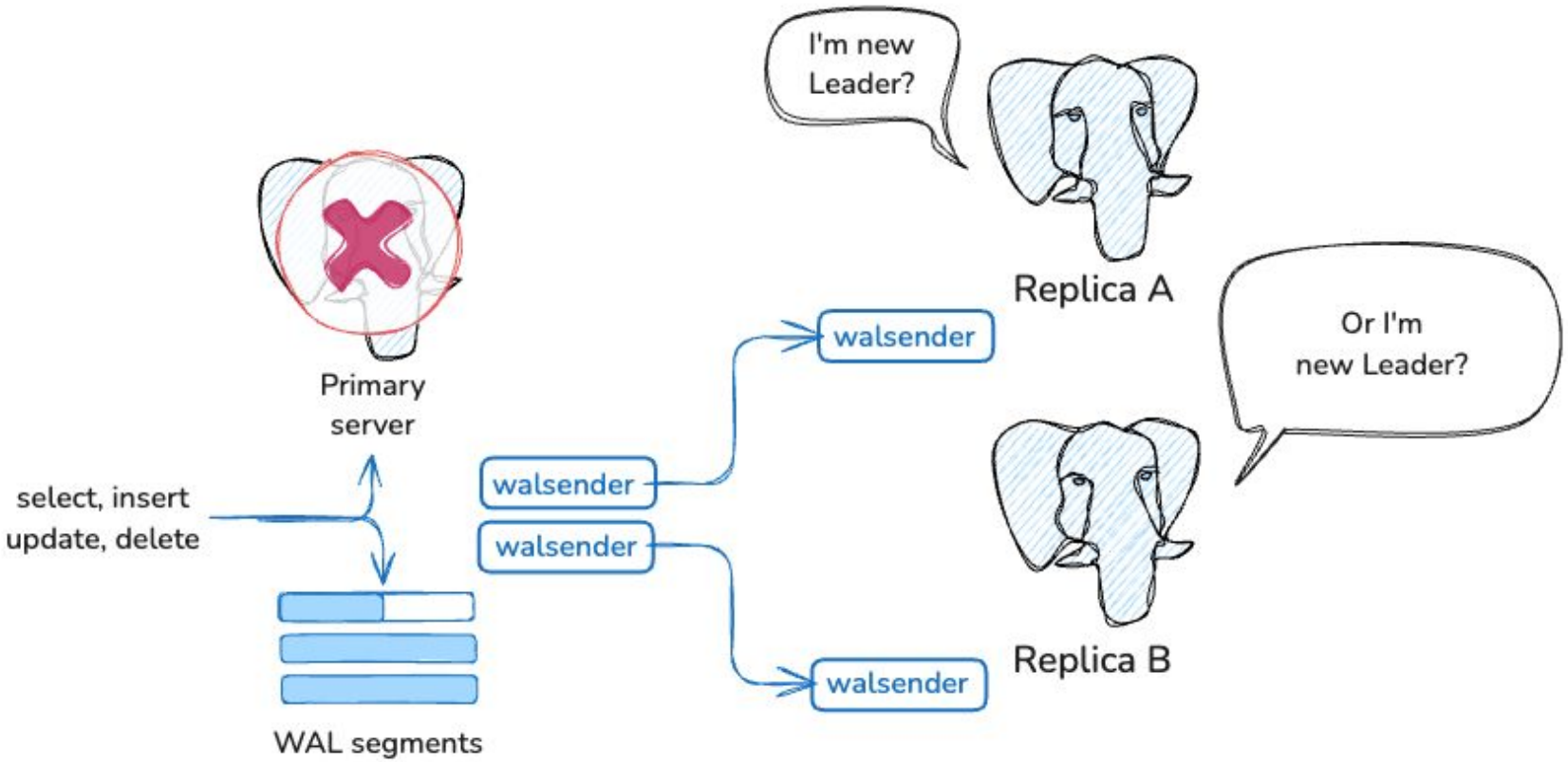
- Data flow in one direction only
- Binary server compatibility is required
- Only the the whole cluster can be replicated

Physical replication

PostgresPro



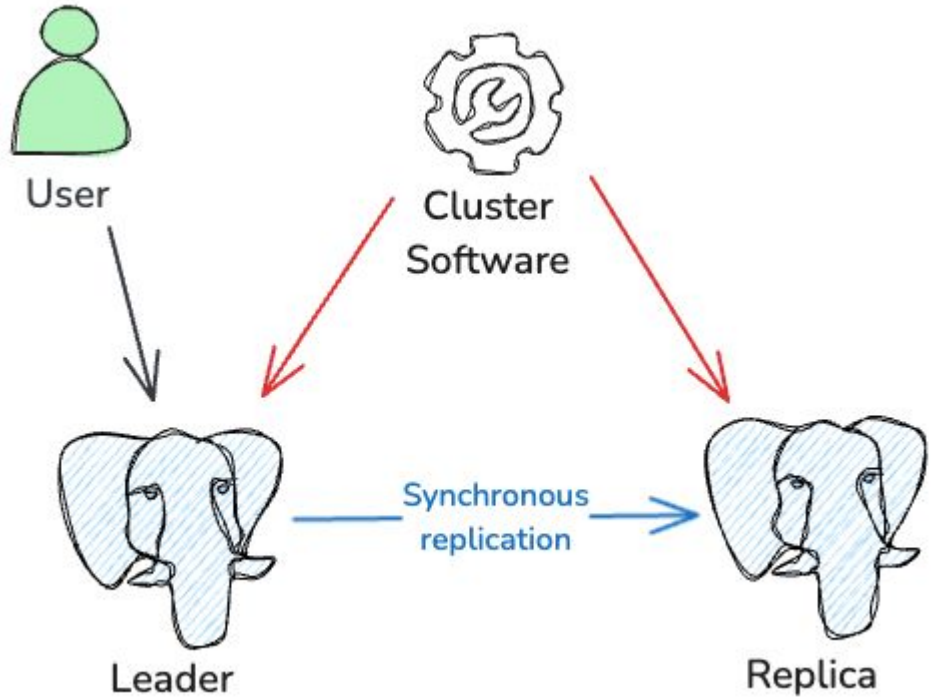
Physical replication: Switching to a replica



- None of the servers will become leaders without manual intervention
- No automation from the box
- Create procedures, call a DBA at night, etc

Physical replication: Failover

- The decision to change roles in a failover cluster when a Leadr fails can be made automatically
- The main tasks of cluster software are:
  - Detect a failure
  - Promote the Replica to a new leader
  - Prevents split-brain



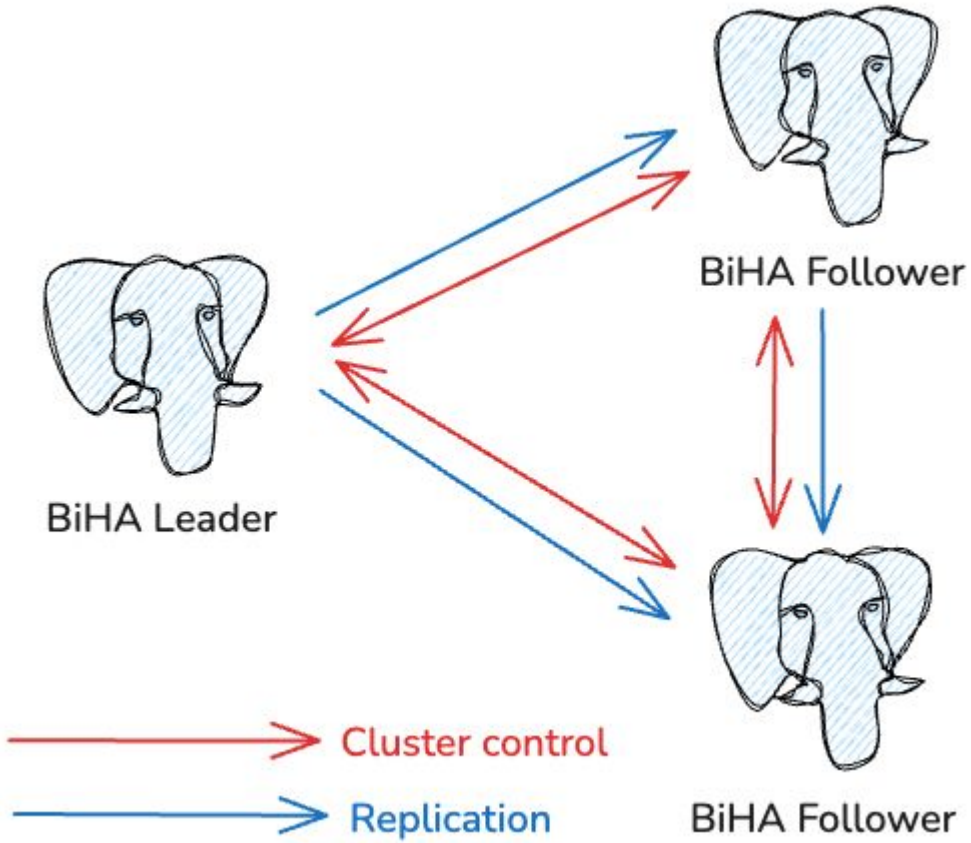
Popular cluster software: Patroni, Stolon, Corosync ….

Postgres Pro : BiHA

PostgresPro

BiHA - Built-in-High-Availability

- Build on top of physical replication
- Leader-Follower Model
- Extension in Postgres Pro Enterprise Edition (16.x and newer)
- Automatic failover
- SQL and CLI interface
- No additional external cluster software required
- ***No additional license required***



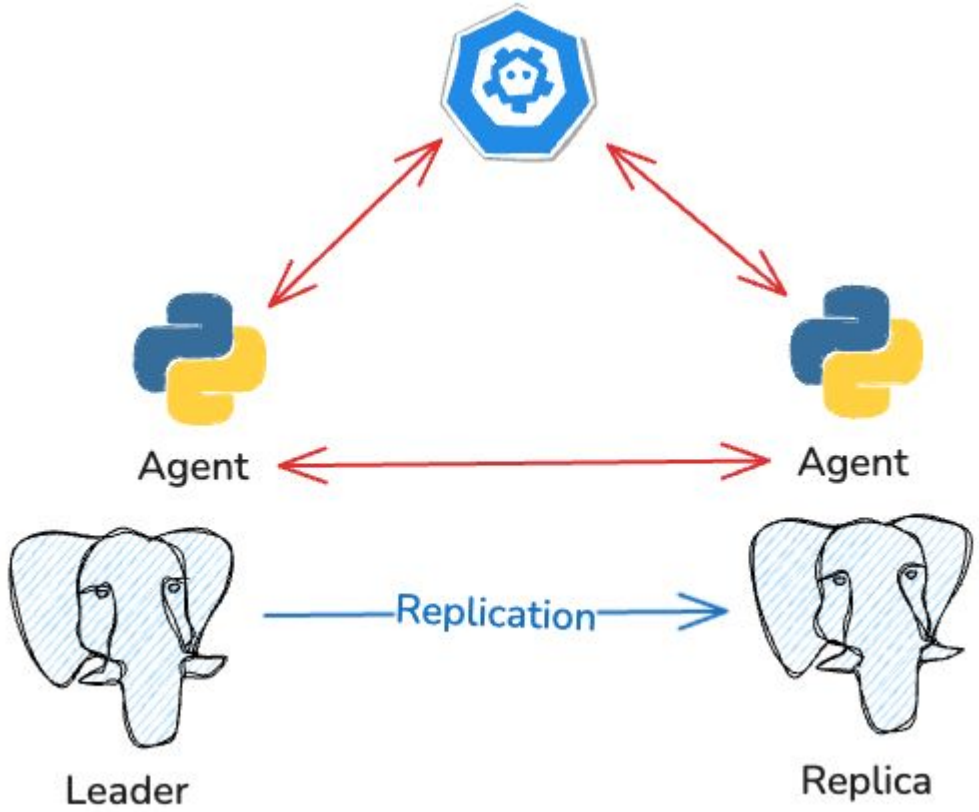
https://postgrespro.com/docs/enterprise/16/biha

Why BiHA?

PostgresPro

External cluster software requires a complicated architecture:

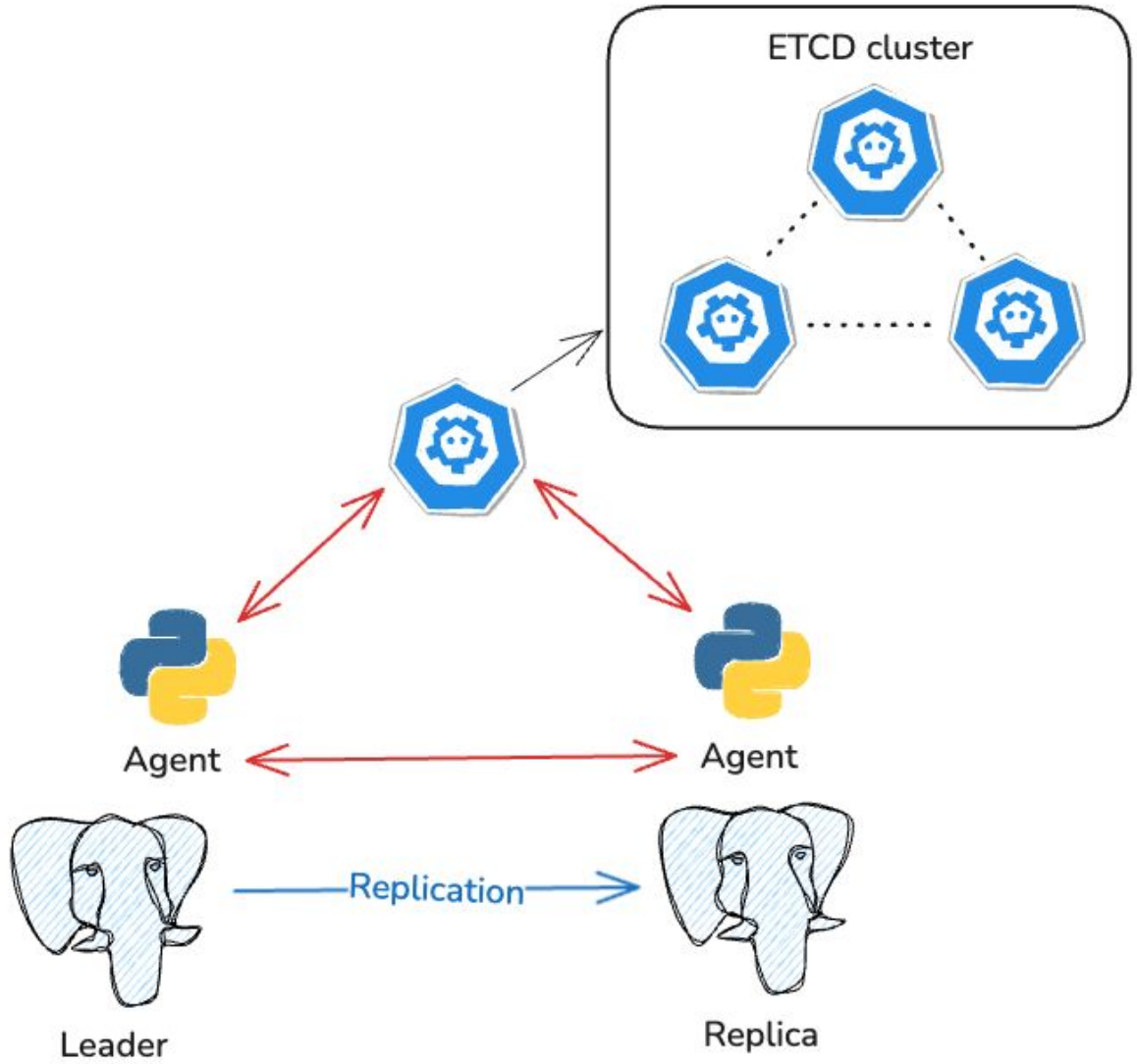
- Additional software like agents
- Additional DCS cluster
- New network channels and ports to look for



Why BiHA?

External cluster software requires a complicated architecture:

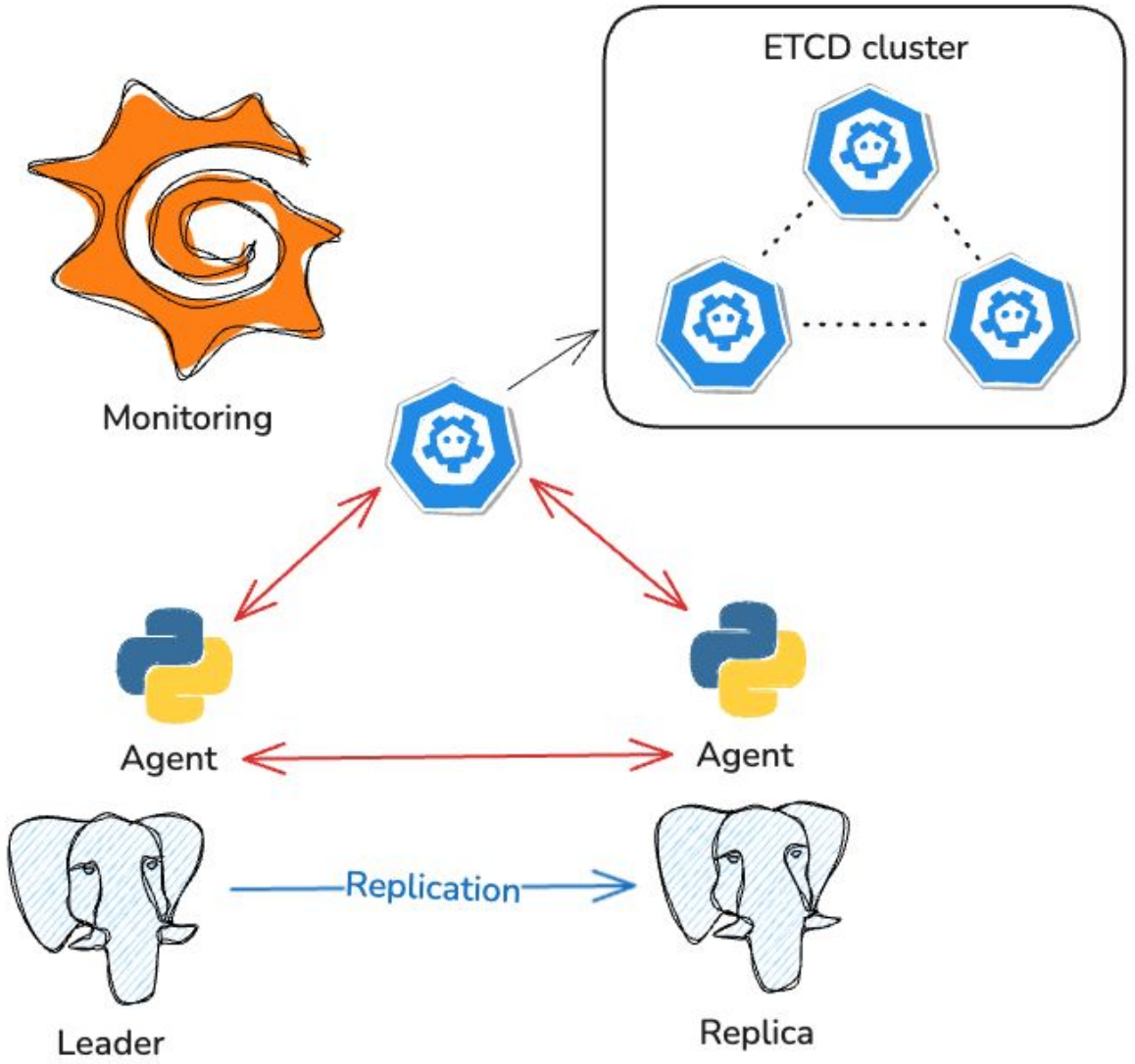
- Additional software like agents
- Additional DCS cluster
- New network channels and ports to look for
- External cluster software also require HA



Why BiHA?

External cluster software requires a complicated architecture:

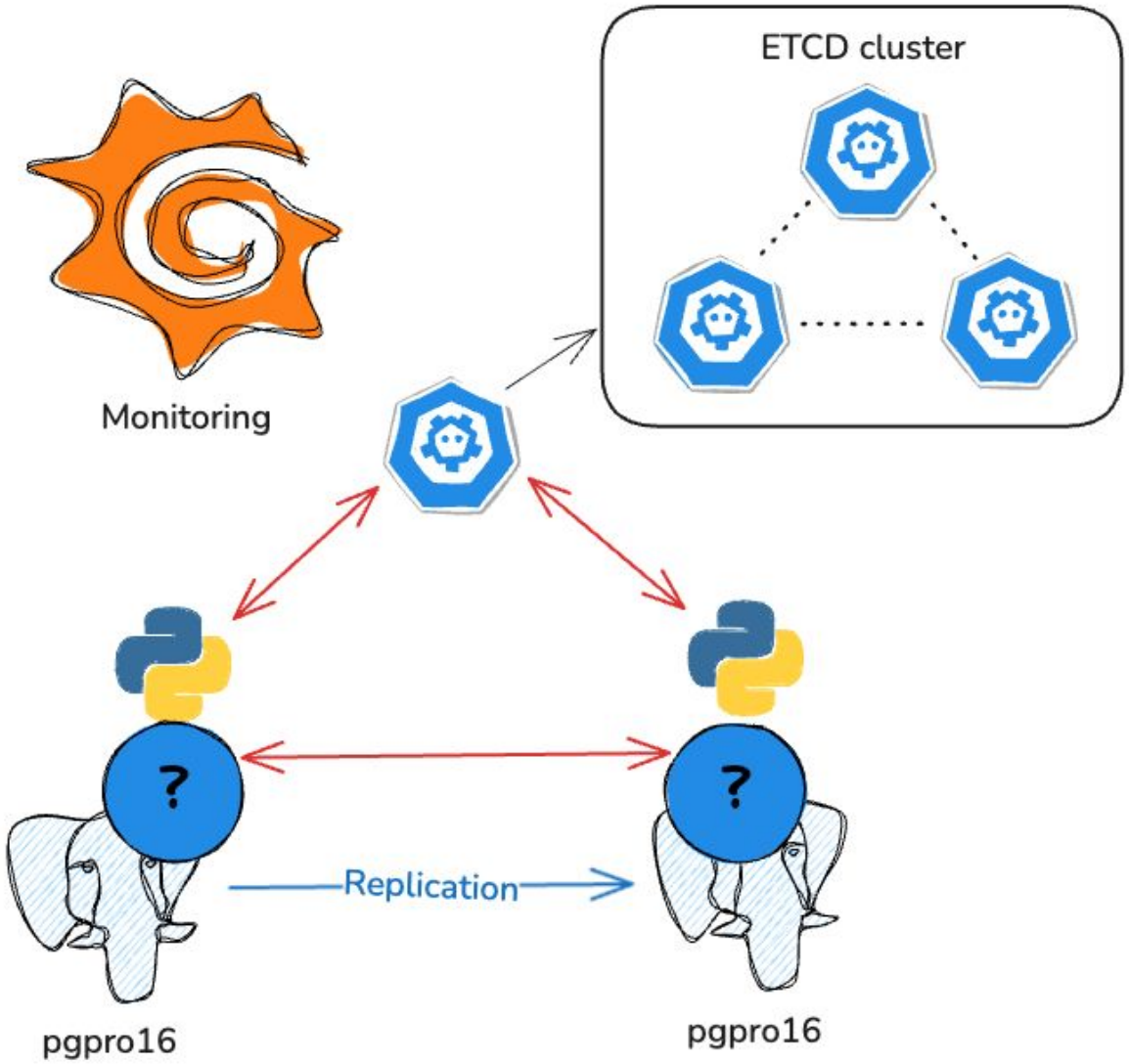
- Additional software like agents
- Additional DCS cluster
- New network channels and ports to look for
- External cluster software also require HA
- Monitoring system became more complicated



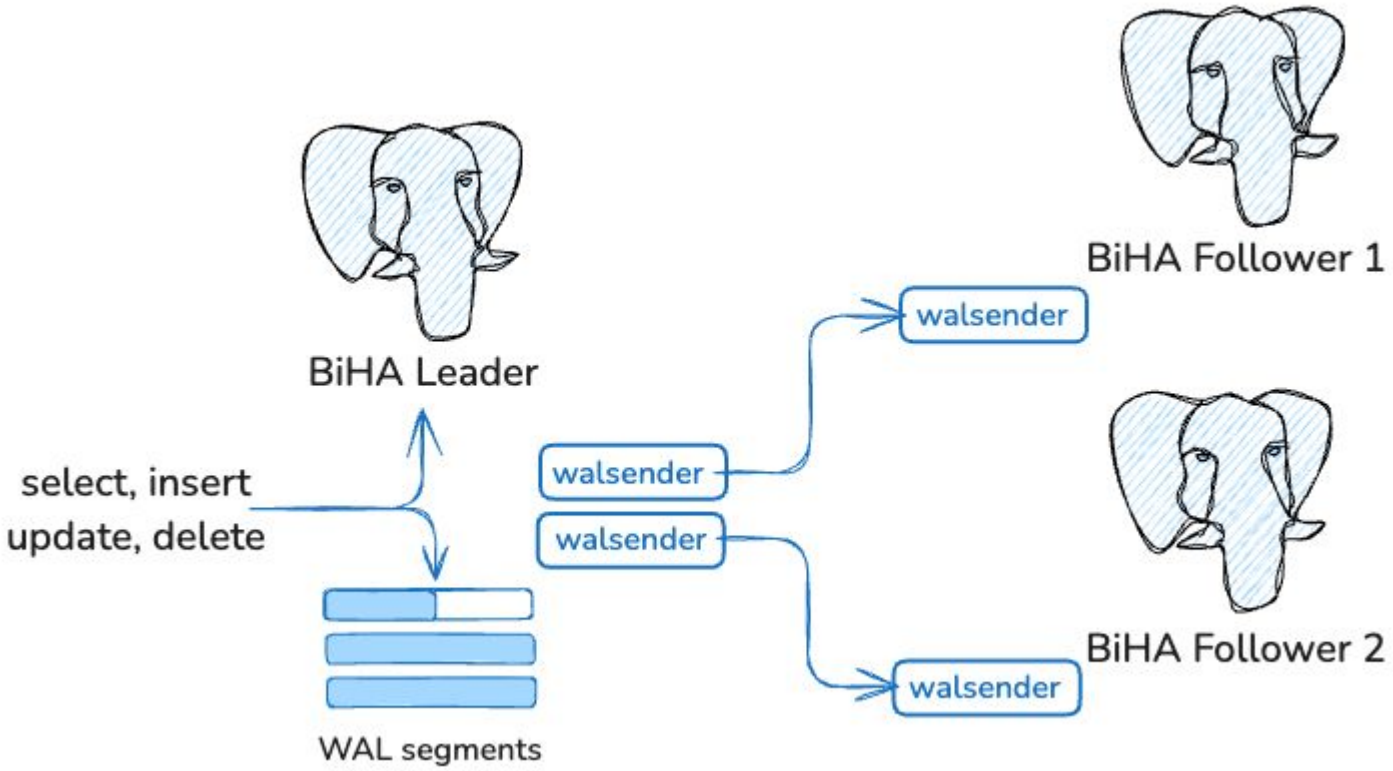
Why BiHA?

External cluster software requires a complicated architecture:

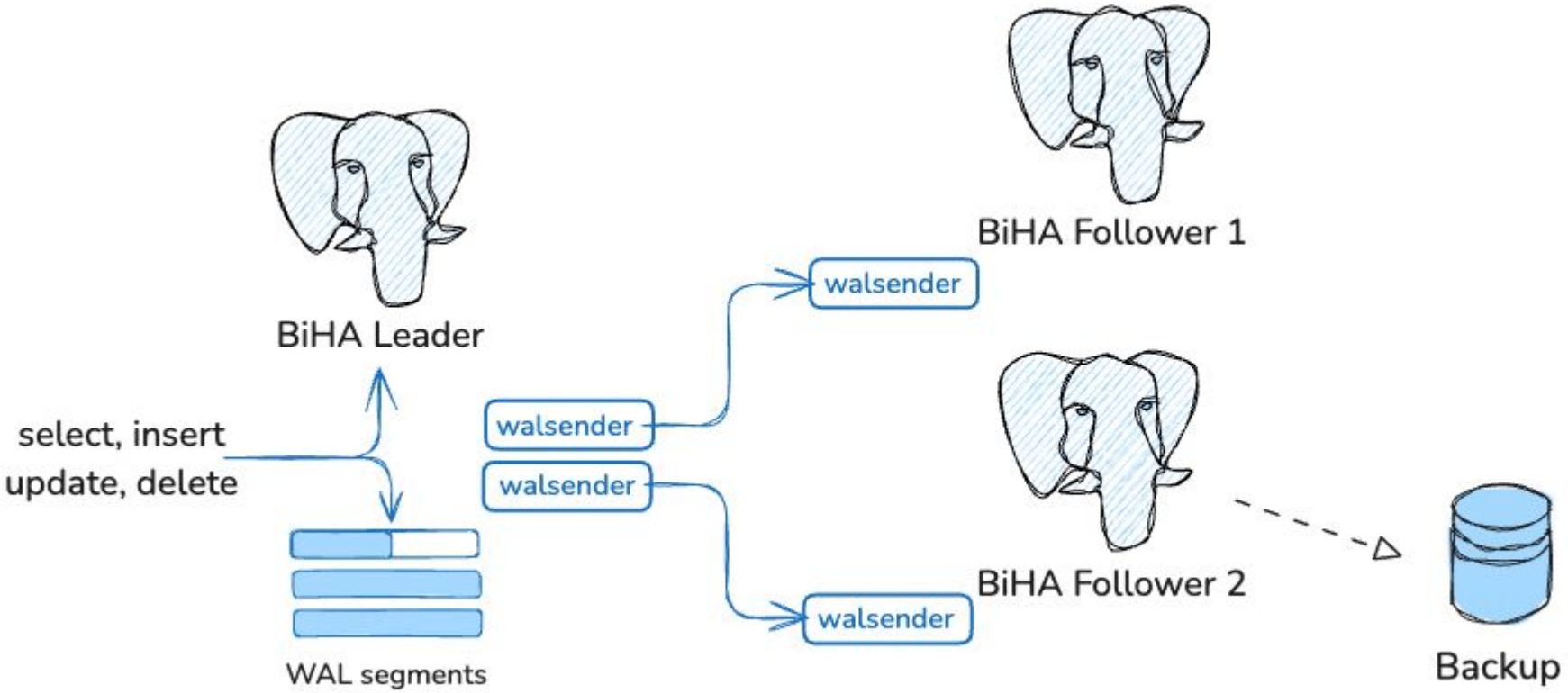
- Additional software like agents
- Additional DCS cluster
- New network channels and ports to look for
- External cluster software also require HA
- Monitoring system became more complicated
- Delays with software updates



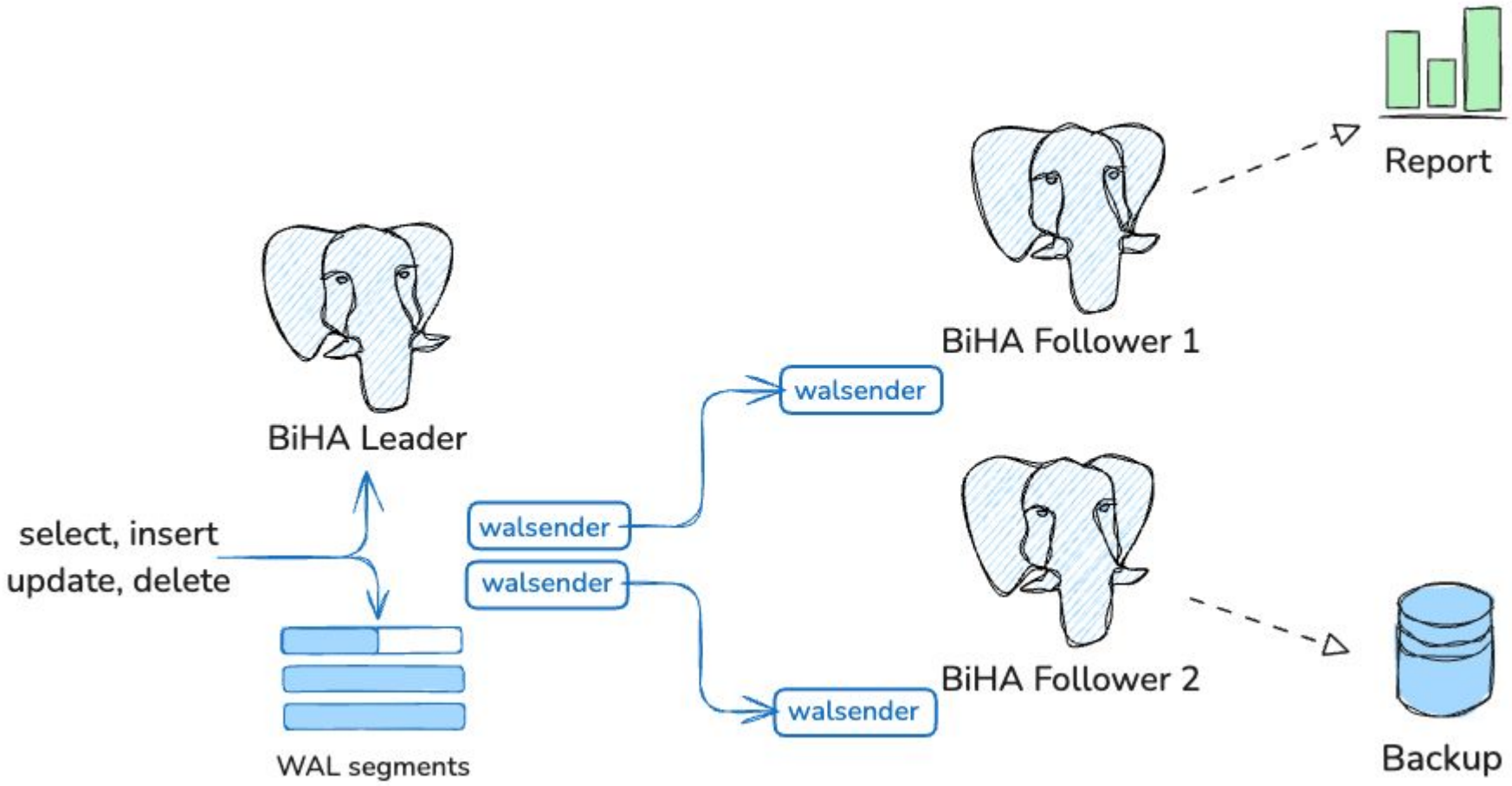
BiHA Architecture



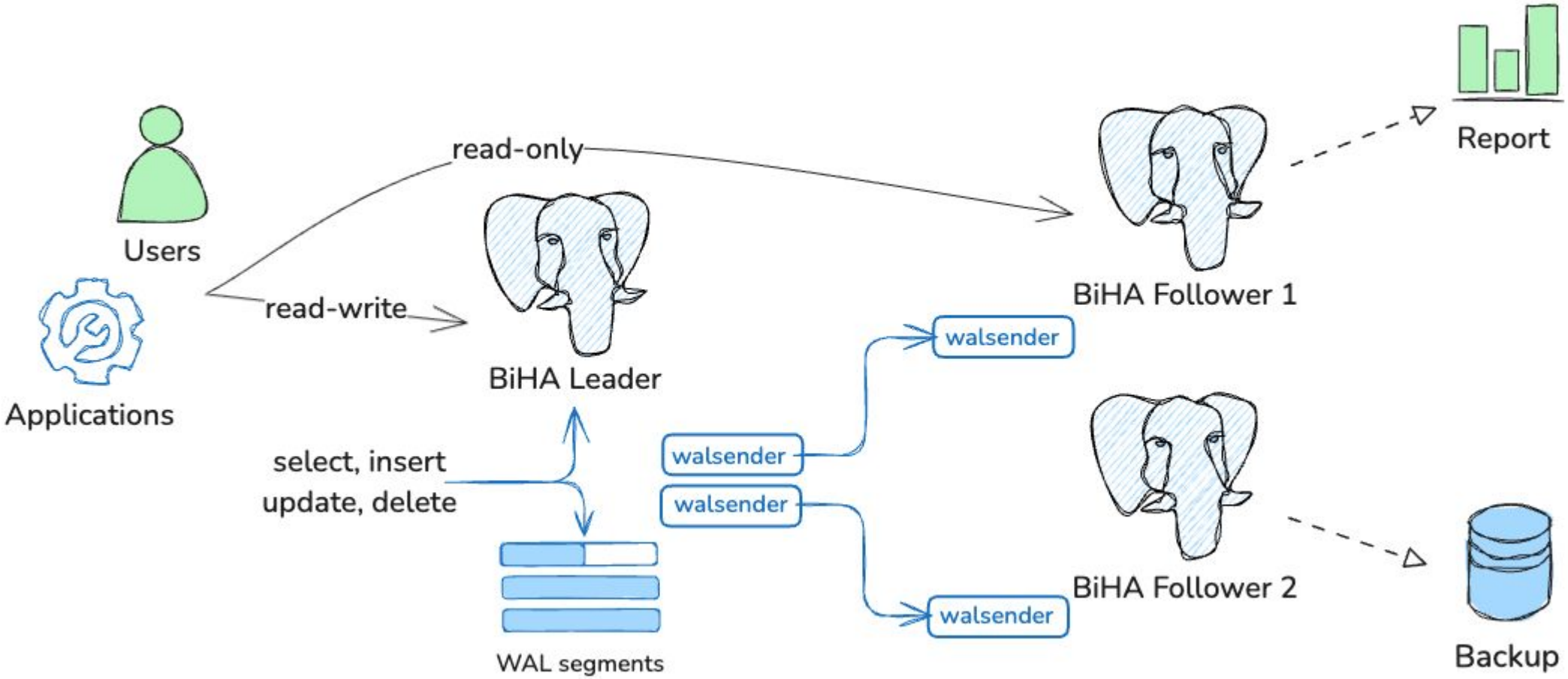
BiHA Architecture



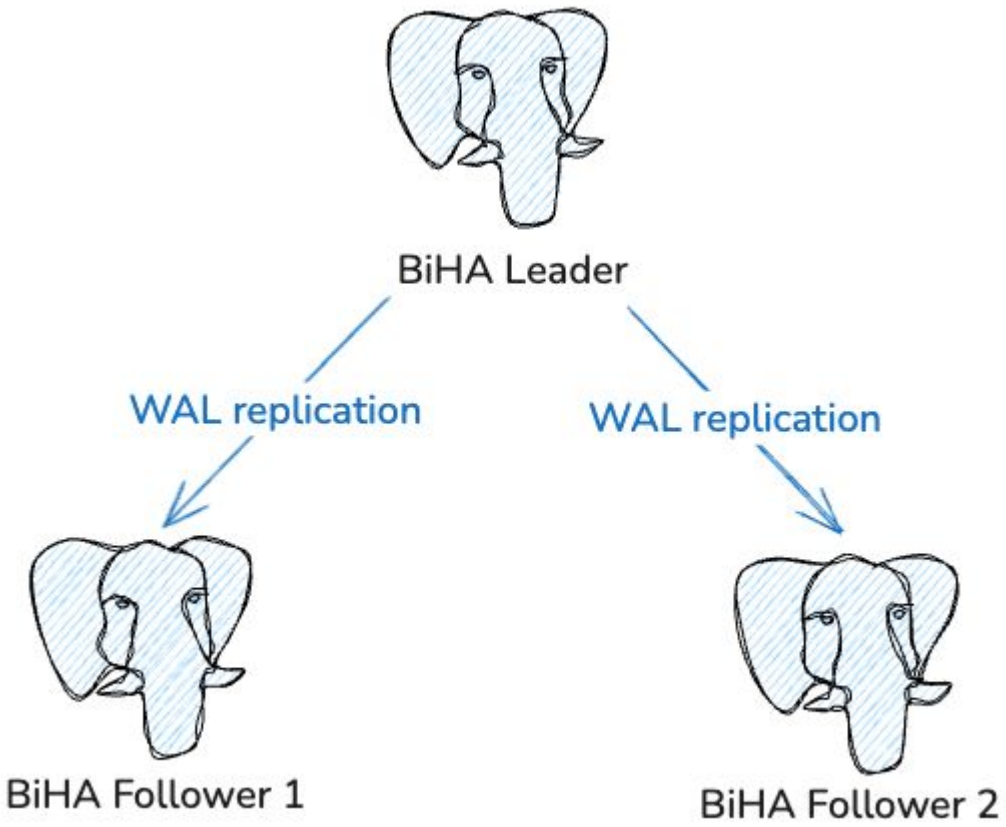
BiHA Architecture



BiHA Architecture



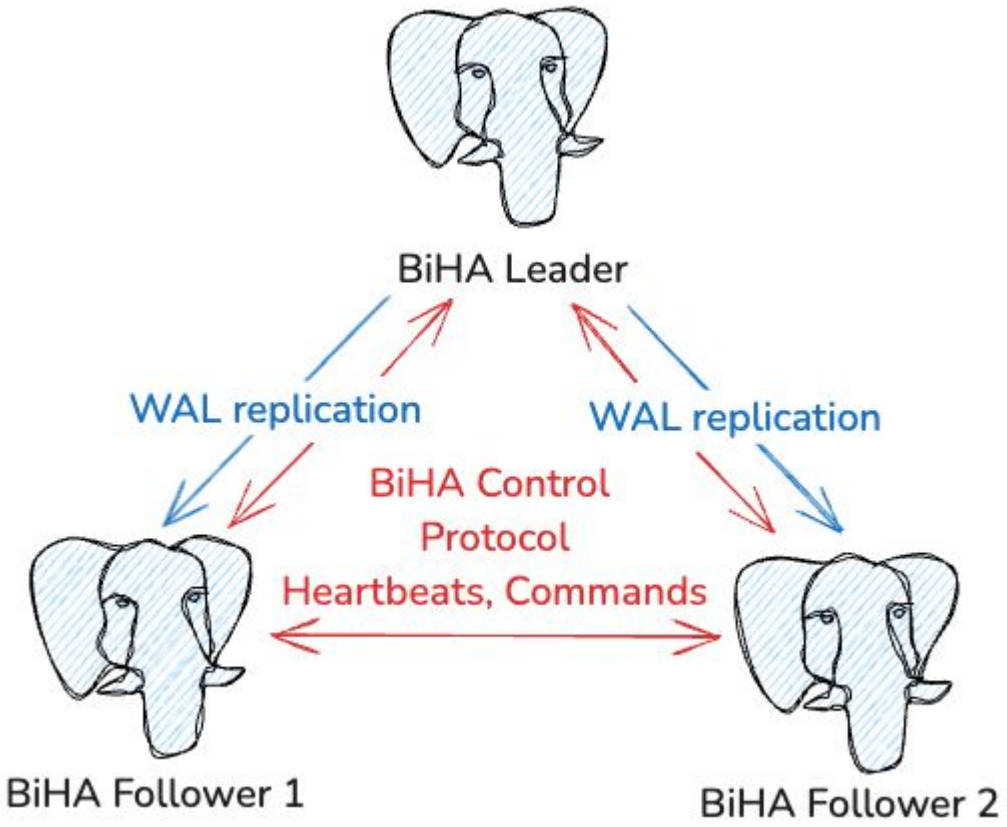
BiHA Architecture



$ ps aux|grep BiHA

postgres ..... postgres: BiHA worker: node 1

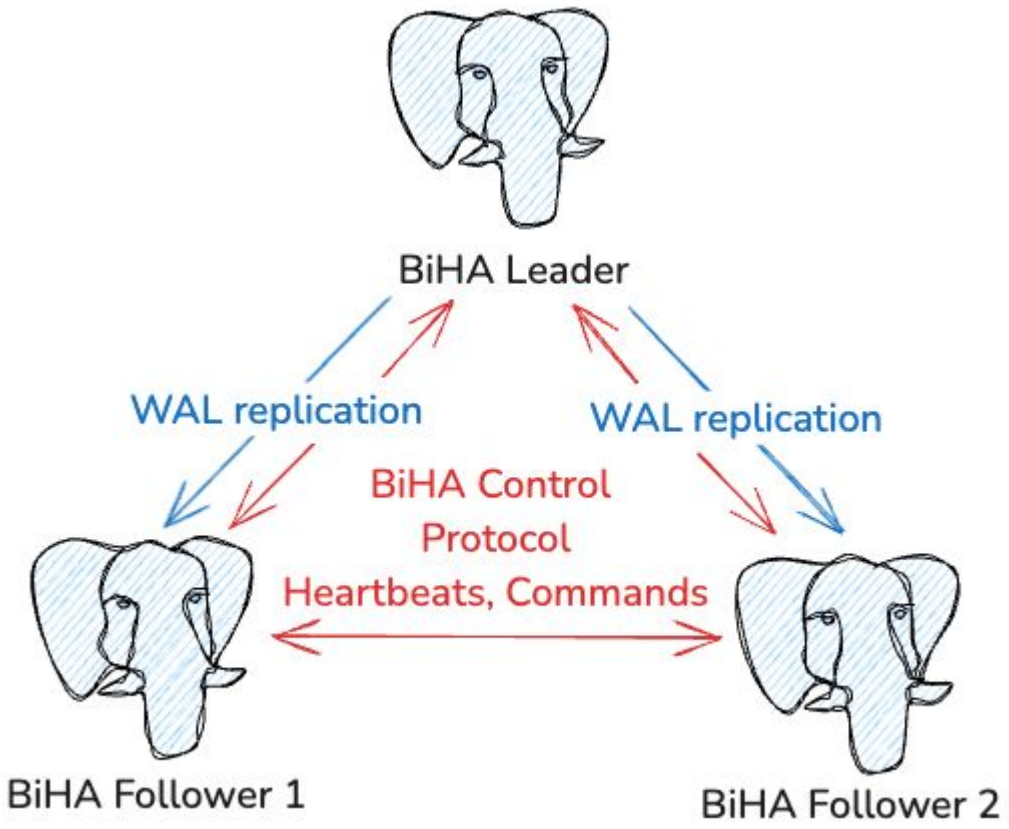
BiHA Architecture



$ ps aux|grep BiHA

postgres ..... postgres: BiHA worker: node 1

BiHA Architecture



$ ps aux|grep BiHA
postgres ..... postgres: BiHA worker: node 1

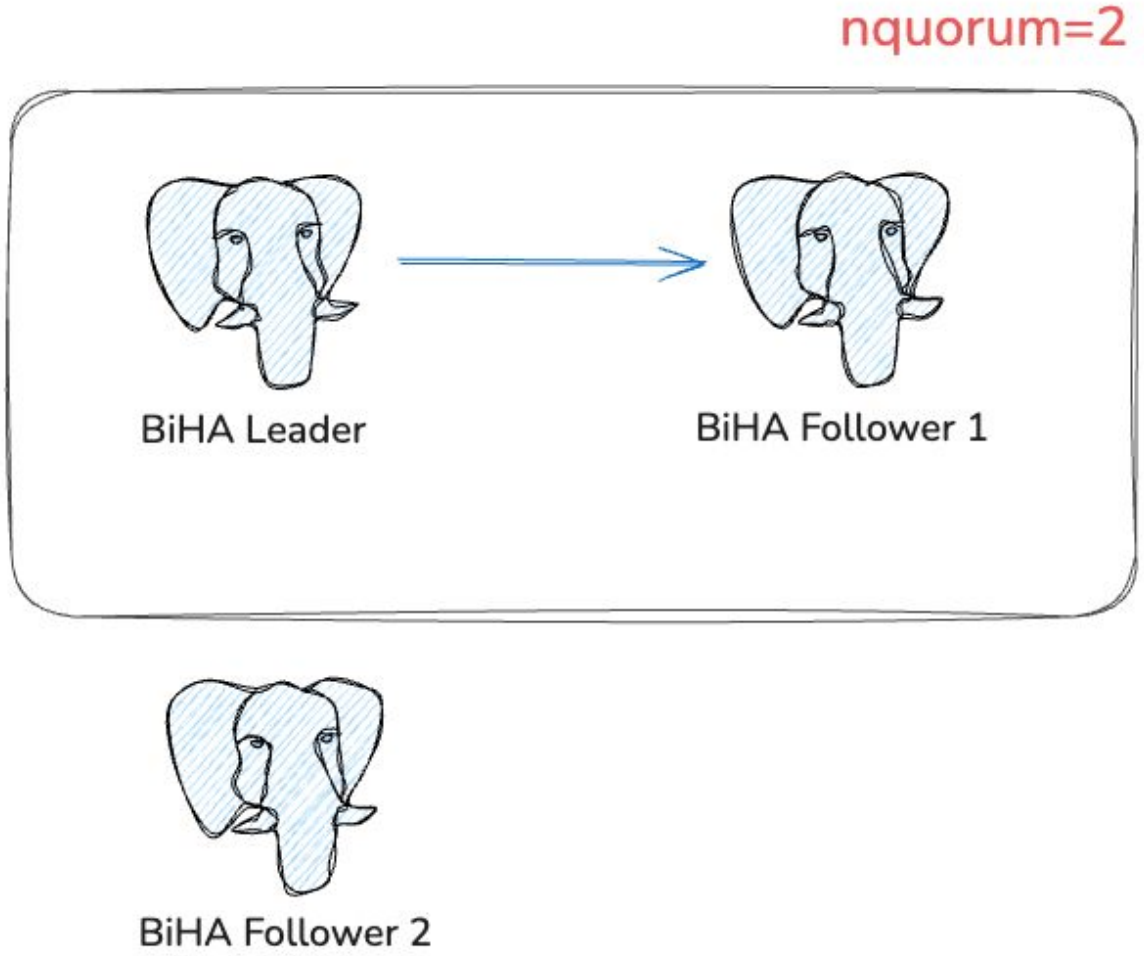
$ cat $PGDATA/pg_biha/biha.conf
node_count=3
1:bp-astra-biha-1:15432:5432:regular
2:bp-astra-biha-2:15432:5432:regular
3:bp-astra-biha-3:15432:5432:regular
crc=525348322

$ cat $PGDATA/pg_biha/biha.state
node_id=1
term=8
leader_id=1
node_error=0
next_tli=0
next_tle_begin=0
next_tle_end=0
...
crc=3842108773

BiHA
cluster quorum

Quorum is a minimum number of nodes that participate in the leader election

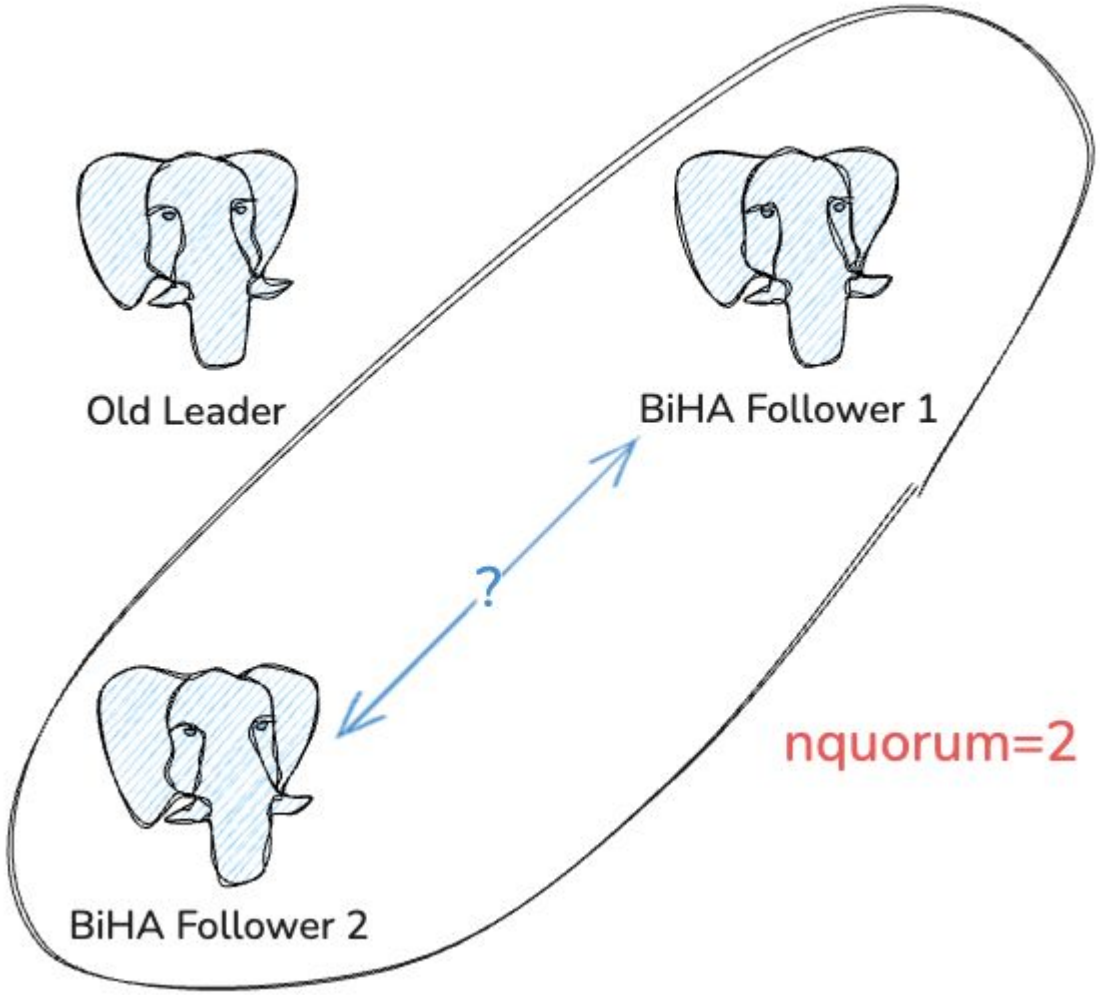
If you have a cluster with three nodes where nquorum=2 and one follower node is down, the cluster leader will continue to operate



https://developer.hashicorp.com/consul/docs/architecture/consensus

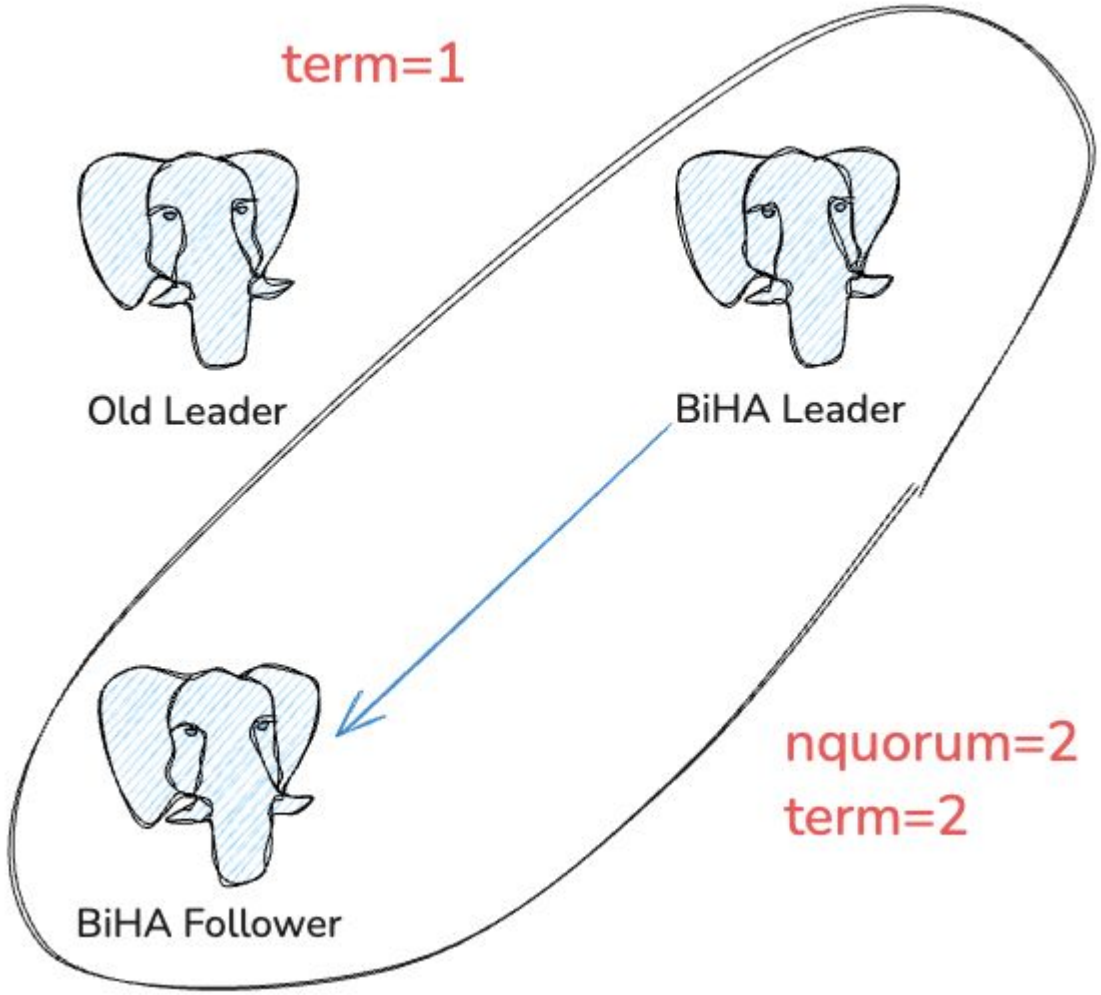
BiHA
cluster quorum

- Leader node cannot operate without quorum
- Promotion is done automatically: the follower cluster node with the most recent LSN becomes cluster Leader
- The elections are based on the cluster quorum



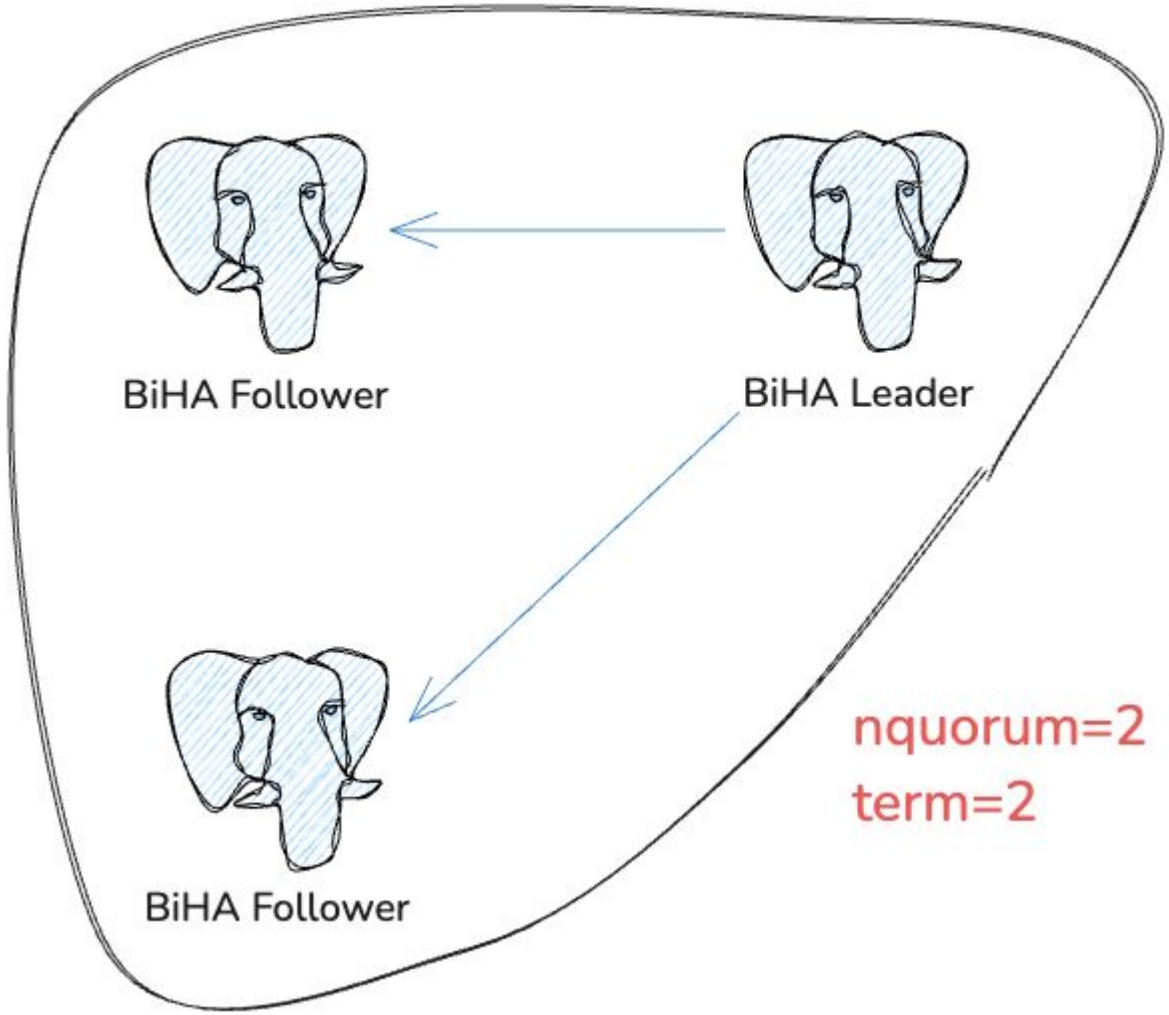
BiHA
term

- After the new leader node is elected, the term value is incremented for all cluster node
- Cluster quorum and the term concepts are implemented in BiHA based on the Raft consensus algorithm



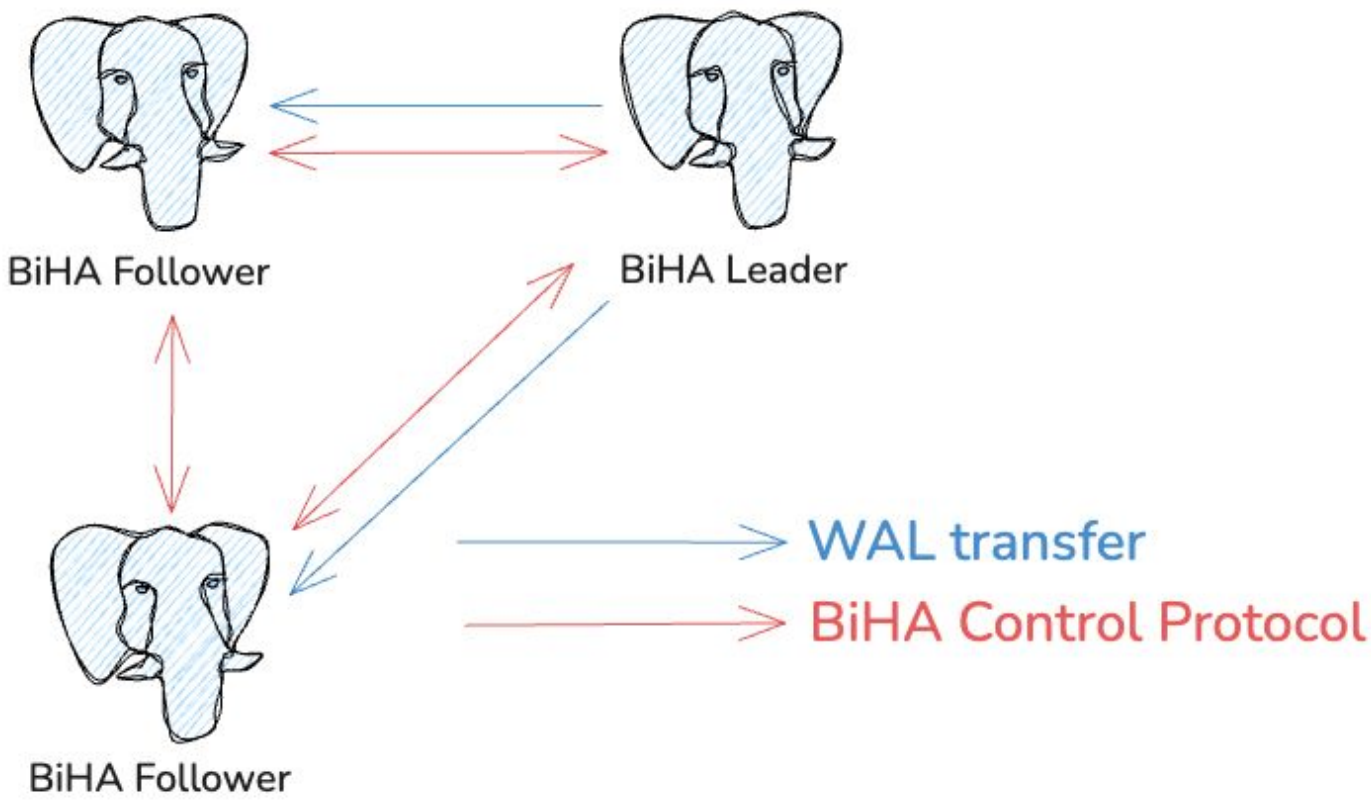
BiHA
term

- If, after elections of the new leader, the old leader returns to the cluster, it's TERM is lower than new clusters TERM.
- So, in order to protect from split brain it cannot be a leader and switches to a follower



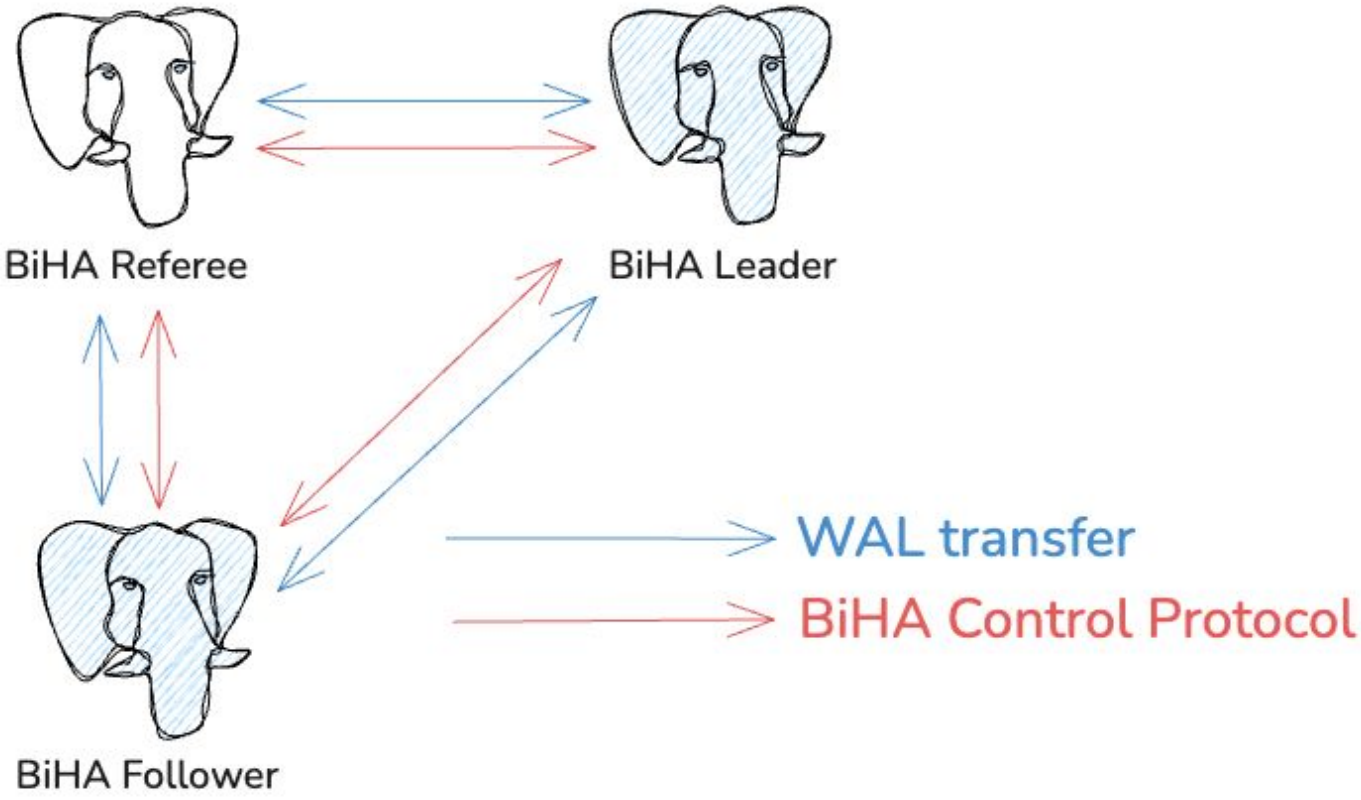
BiHA control protocol

- The control channel is used to exchange service information between the nodes
- Every node has BiHA worker process that uses BiHA control protocol (in red)
- Continuous monitoring of the status of cluster nodes



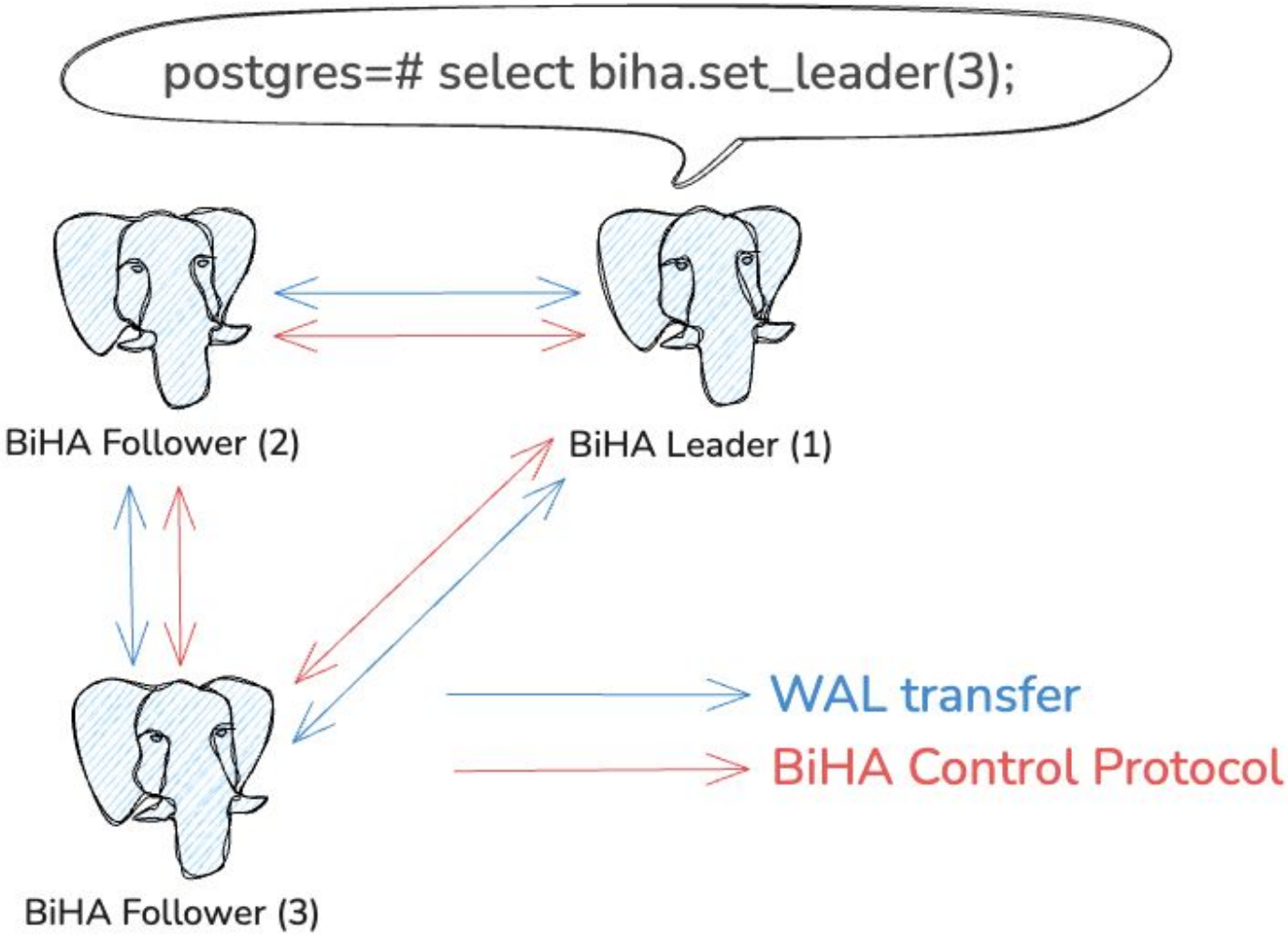
BiHA
referee – configuration 2+1

- Setup referee to avoid split-brain situations
- Referee participates in elections, but not contain any user data
- Referee with WAL receive the entire WAL from the Leader node
- In some circumstances Follower can tries to get missing WAL files from the Referee



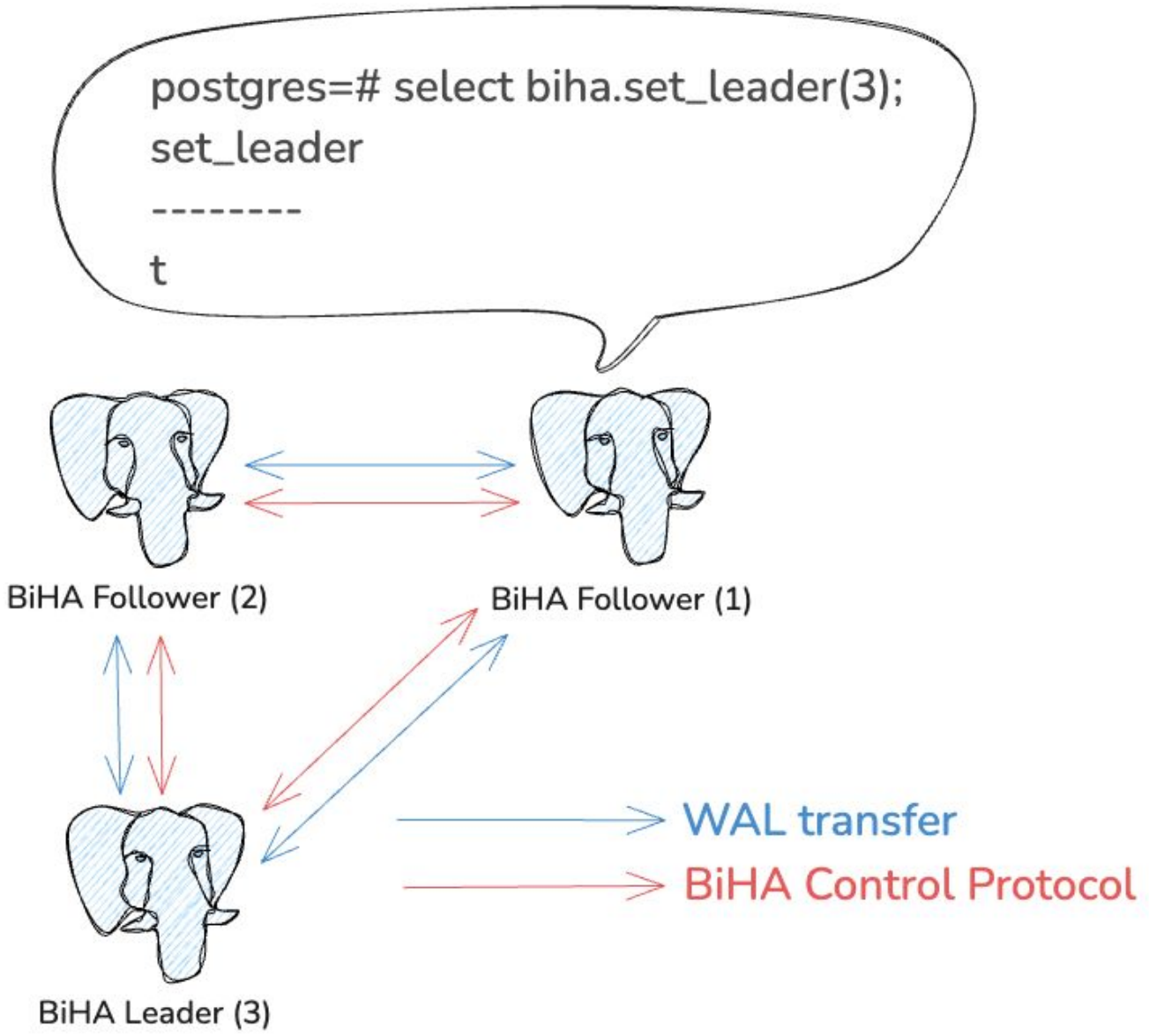
BiHA
switchover

- To put the Leader into maintenance mode
- To assign a leader to a preferred host
- After returning the old leader



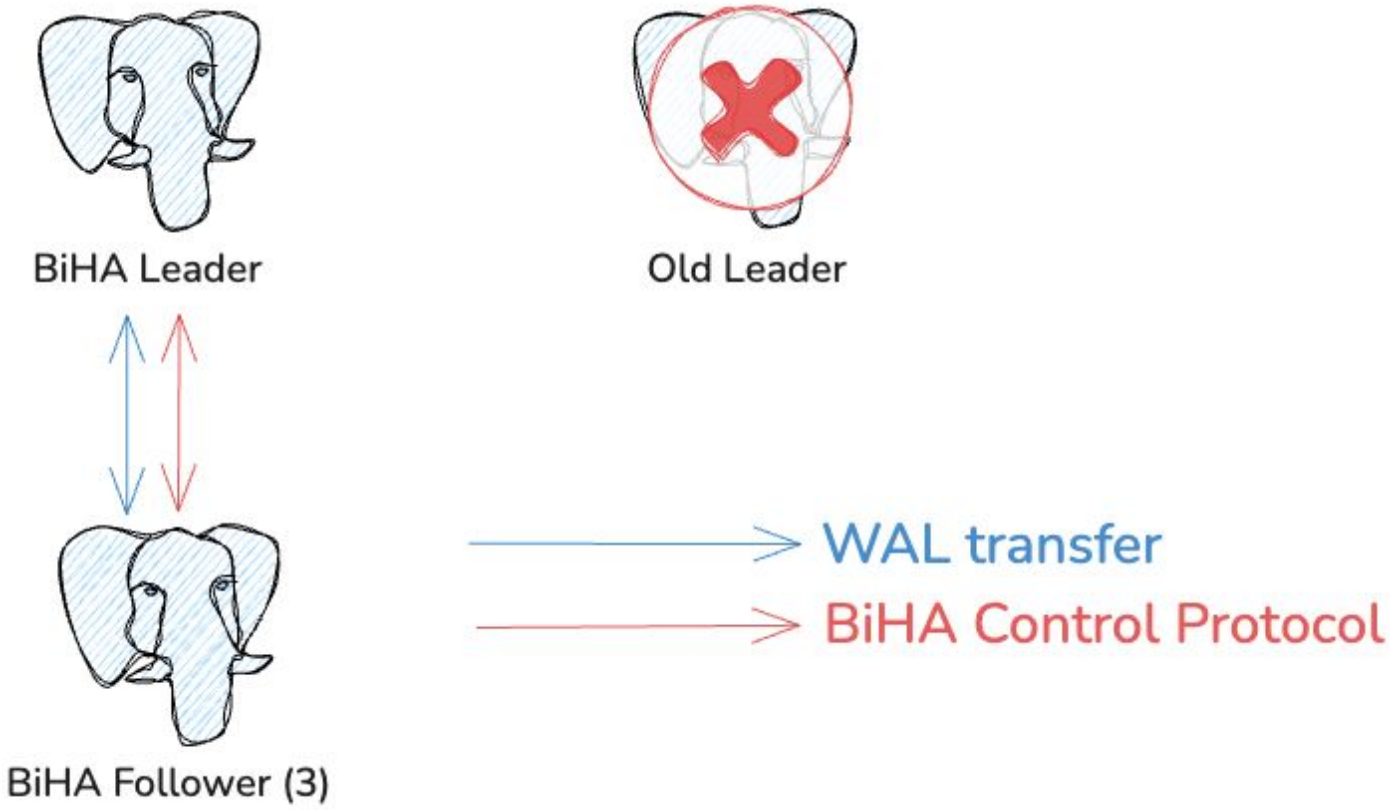
BiHA
switchover

- All attempts to perform elections are blocked and the timeout is set
- The current leader node becomes the follower node
- The newly selected node becomes the new leader
- If the switchover process does not end within the established timeout, the selected node becomes the follower and new elections are performed to choose the new cluster leader



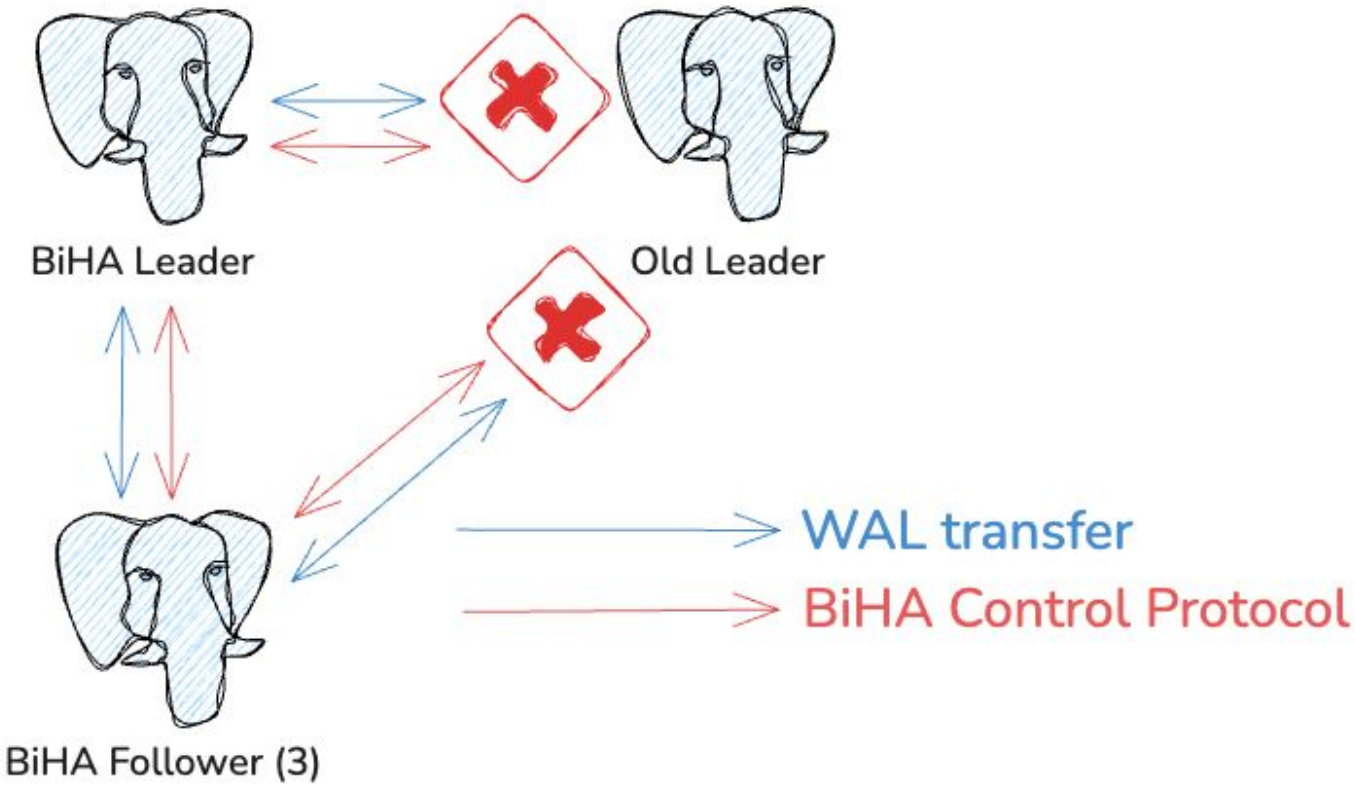
BiHA
failover

- Failover or automatic change of leader happens in emergency situations
- When a leader fails, the follower organize a voting process to select a new leader
- The follower node with maximus LSN (the most nearest to old Leader) becomes the new leader



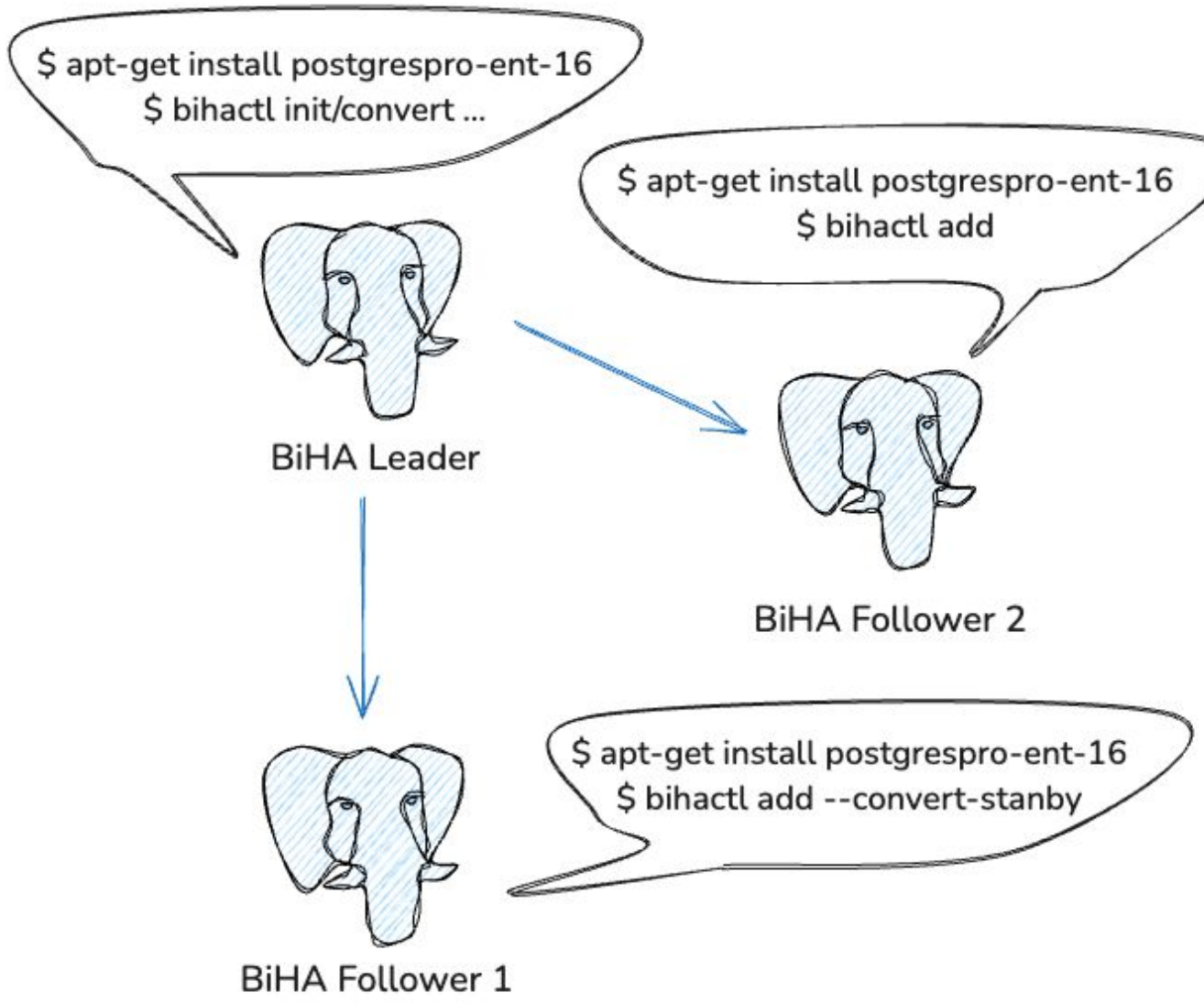
BiHA
Network failure

- Network failure between the Leader node and follower nodes
- The old Leader becomes read-only
- This protection ensures that any operation that modify data are prohibited to prevent recording to several Leaders simultaneously (Split-brain)



Benefits of BiHA

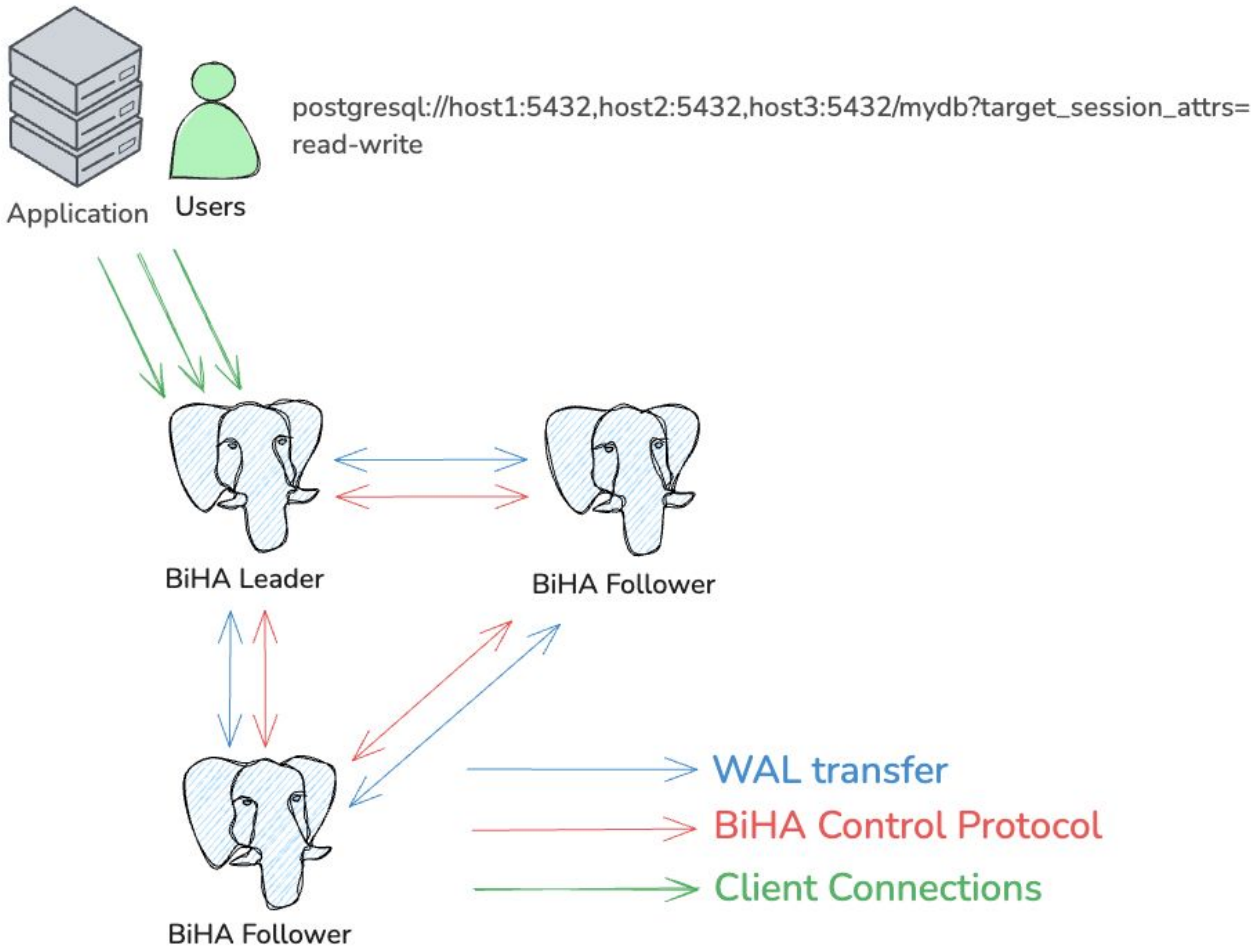
- Build-in to the Postgres Pro
- Easy installation and configuration
- No any additional software required
- No any additional nodes required
- Operational updates without delays



Clients request
&
load balancing

libpq

- Standard libpq C library operator
- The goal is to find the leader via `target_session_attrs` parameter
  - read-write
  - read-only
  - primary
  - https://www.postgresql.org/docs/current/libpq-connect.html
- Same operators are applicable for languages like Python, Java, Go, etc.
- In case of failure the client will automatically reconnect to the new Leader



HAProxy (with external script)

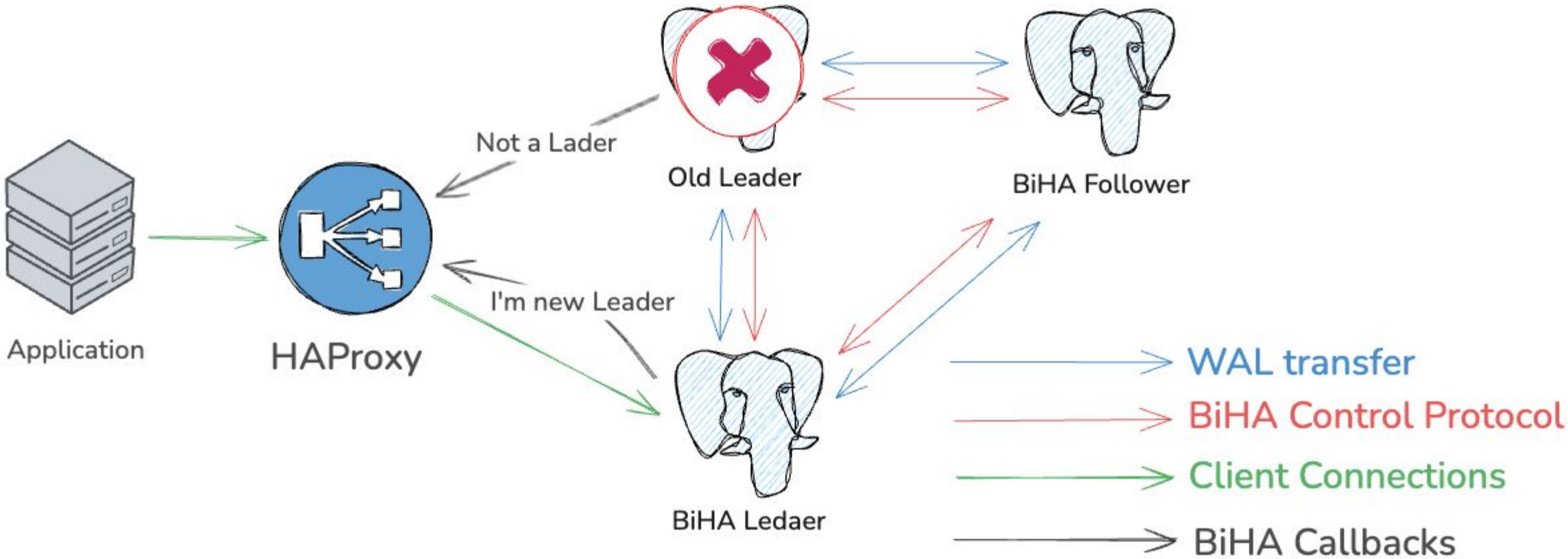
PostgresPro

- The most popular util that can handle some serious logic
- Easy to install, configure and maintain
- Has a GUI and prometheus format metrics
- And the first option on how to find the Leader is to use an external script

```
backend biha-rw from unnamed_defaults_1
  mode tcp
  option external-check
  external-check command /etc/haproxy/check_biha_leader.sh
  timeout queue 5s
  default-server inter 500ms downinter 1s fall 3 rise 2 on-marked-down shutdown-sessions
  server biha-leader vp-bihatest-pgproee-2.l.postgrespro.ru:5435 check
```

HAProxy with Data Plain API

- HAProxy backends can be defined dynamically via Data Plain API
- It's done by running callbacks scripts in BiHA cluster
- Fast and furious



Other custom solutions

PostgresPro

- TCP proxy solutions like NGINX
- keepalived
- Hashicorp Consul service discovery + Consul template + NGINX/HAProxy
- Own script, tool, etc…

Conclusion

PostgresPro

- libpq - and engineer team are not bothered on how application find the target session
- haproxy - more flexible solution, but requires configuration and tuning. And maybe it’s own HA
- KeepaliveD - not so flexible as HAProxy but done its job well
- Other solutions - not recommended ;)

How to backup BiHA?

Backup utils

PostgresPro

- pg_basebackup - standard postgresql backup tool. Fast, flexible, but needs custom wrappers to operate
- WAL-G - fast and has S3-like storage compatibility from the box
- pgbackrest - another good utility with S3 compatibility
- Barman - old, but good tool
- pg_probackup - multitool with multi-purposes

pg_probackup

PostgresPro

Key features:

- S3 compatibility out of the box
- Integrity control
- Well compressed incremental backups
- Most processes backup, archive, restore, delete, validate, etc - can be runned in parallel mode
- Local and remote operation mode
- Partial restore
- Synchronization of delayed Replica

pg_probackup 3

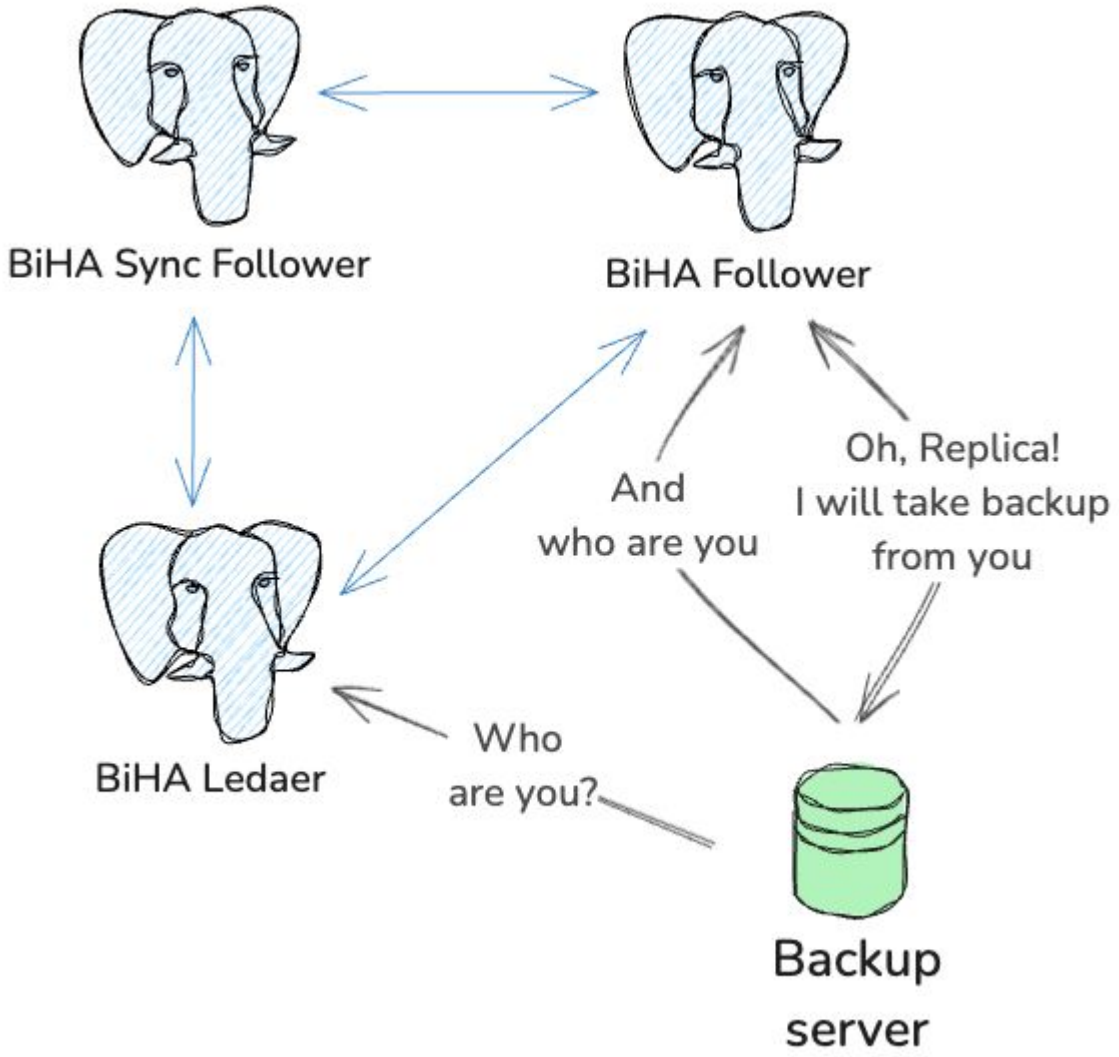
PostgresPro

More powerful:

- New replication protocol
- Remote mode don’t need SSH anymore
- Improved partial database restore
- FUSE filesystem features
- All this is compatible with backups created with pg_probackup 2

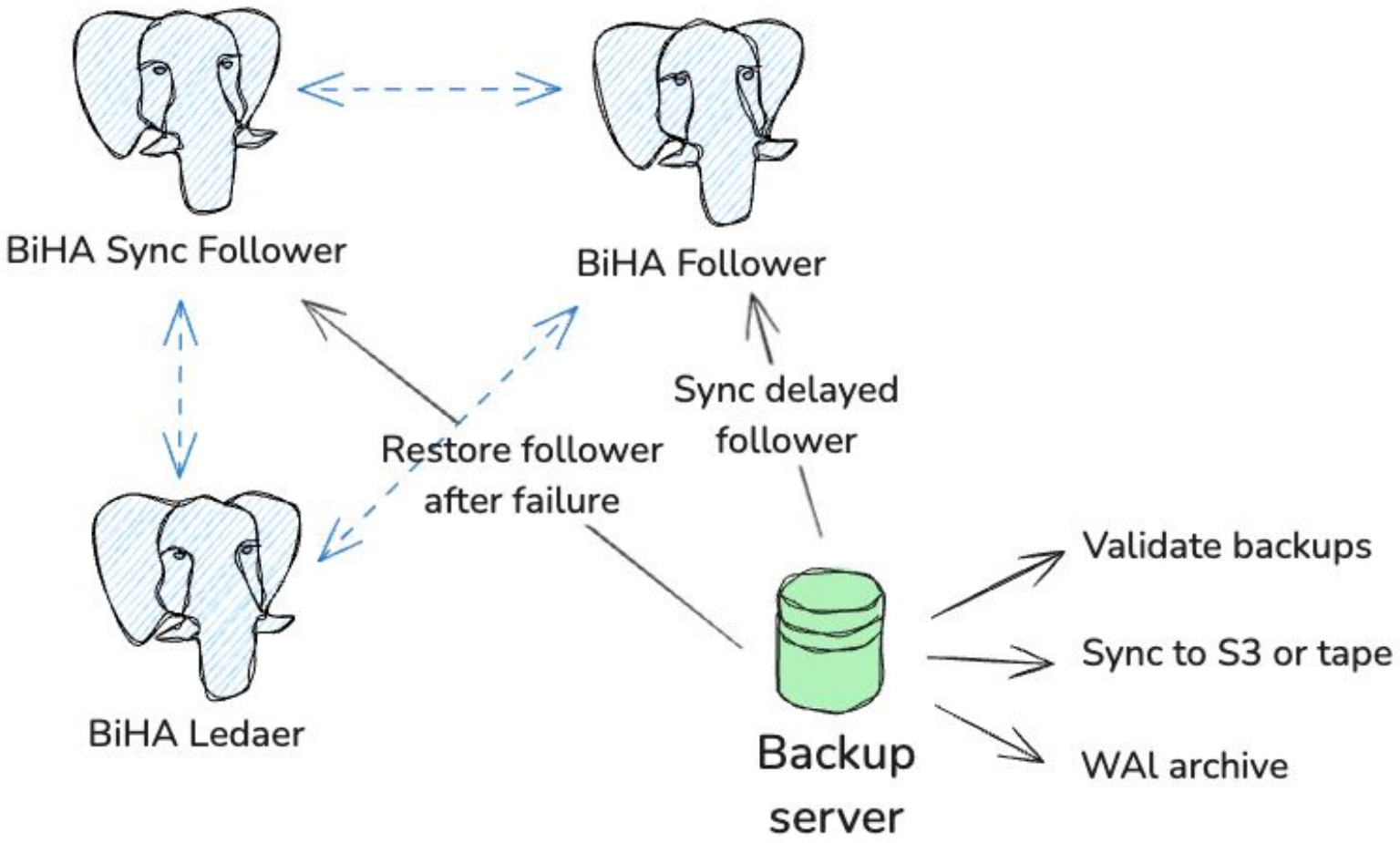
So, how to backup BiHA

- As usual, but we need to find the desired node
  - If we have a HAProxy - let’s use it!
  - SQL interface
  - Bash script
  - Different tools, but the goal is one - find the desired node in cluster



What else

- Easy follower return to the cluster
- Fast sync of severely lagging replica
- Validate your backups
- Sync to S3 or tape
- Work with WAL archive (PiTR)



Conclusion

BiHA - Built-in-High-Availability

- Does not have disadvantages of external cluster software
- Simplifies setup and configuration of physical replication
- Automatically elects the new Leader in case of failures
- Does not require additional infrastructure: nodes, software and network channels
- Integrated to the Postgres Pro Enterprise 16+

Useful links
https://postgrespro.com/docs/enterprise/16/biha
https://postgrespro.com/docs/enterprise/16/bihactl

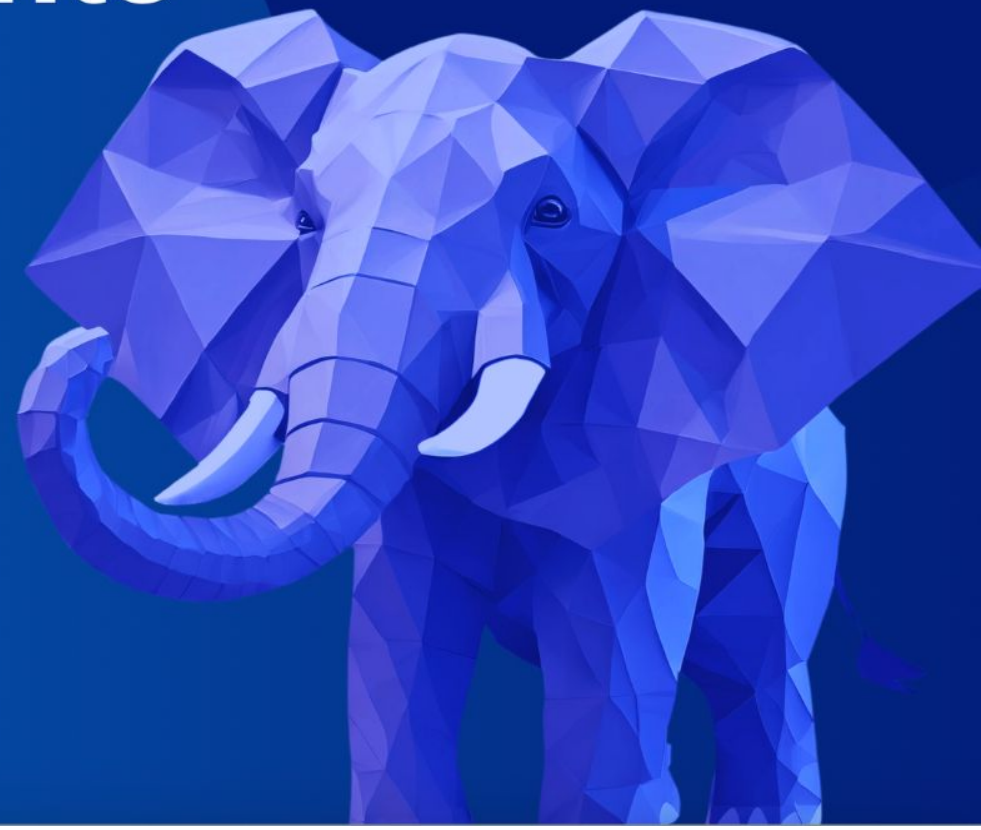
PostgresPro

Q&A

PostgresPro

Reliable DBMS for extremely high-workload requirements

For the latest news join our community at LinkedIn



Email:

india@postgrespro.com