

Resilient Time-Series Data and Visualization Infrastructure using High-Availability TimescaleDB

Prem Kumar Kirti Mahajan

Institute for Plasma Research, Gandhinagar, India

prem.kumar@ipr.res.in

Second Visualizing Offline and Live Data with AI (VOLDA) Workshop
CIEMAT, Madrid, 2025

What is High Availability?

- Definition: Ability of a system to remain operational with minimal downtime
- Key Concepts:
 - **Redundancy**: Having backup components (hardware, software, or data) that can take over if the primary component fails.
 - **Fault Tolerance**: The system's ability to continue operating despite errors or failures in its components.
 - **Failover**: The process of automatically switching to a redundant component when the primary component fails.

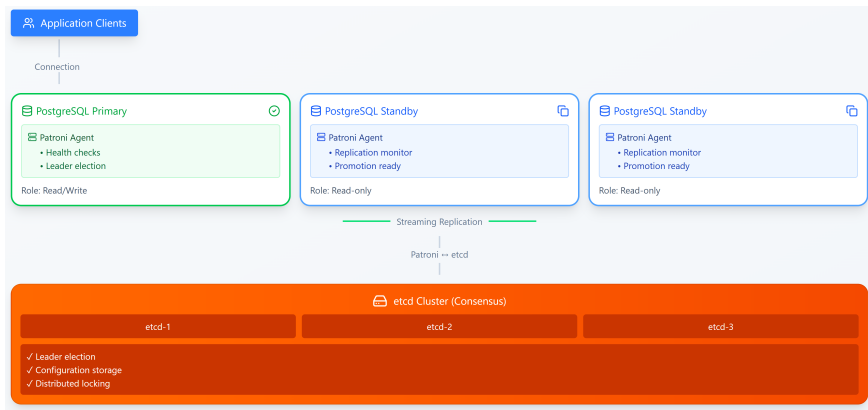
Motivation

- Tokamak operations generate large volumes of diagnostic and machine health data.
- Reliable data storage and availability are essential for post-discharge and predictive analysis.
- Traditional single-node databases suffer from lack of failover and limited scalability.
- Need for a resilient, containerized, and high-availability time-series database infrastructure.

System Overview

- Architecture integrates:
 - **TimescaleDB with Postgres** for storing time-series data
 - **Patroni** for replication and fault tolerance
 - **etcd** for distributed coordination
 - **HAProxy** for routing and load balancing
 - **Keepalived** for virtual IP failover management

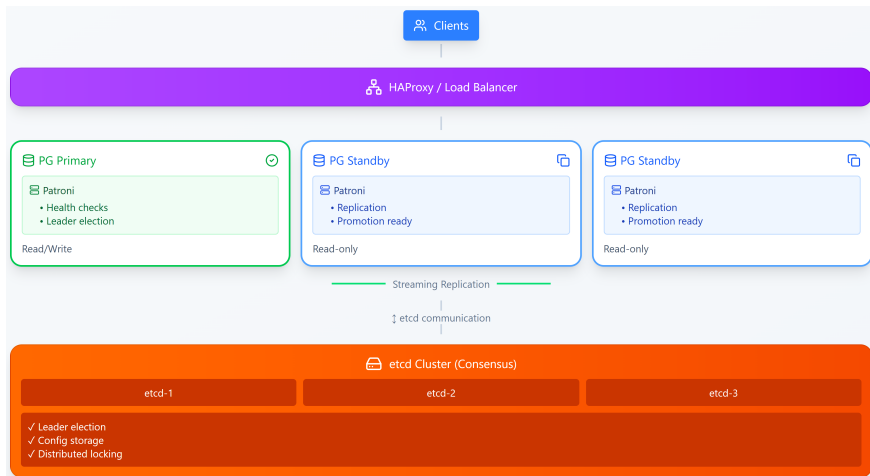
Achieving Redundancy and Fault Tolerance



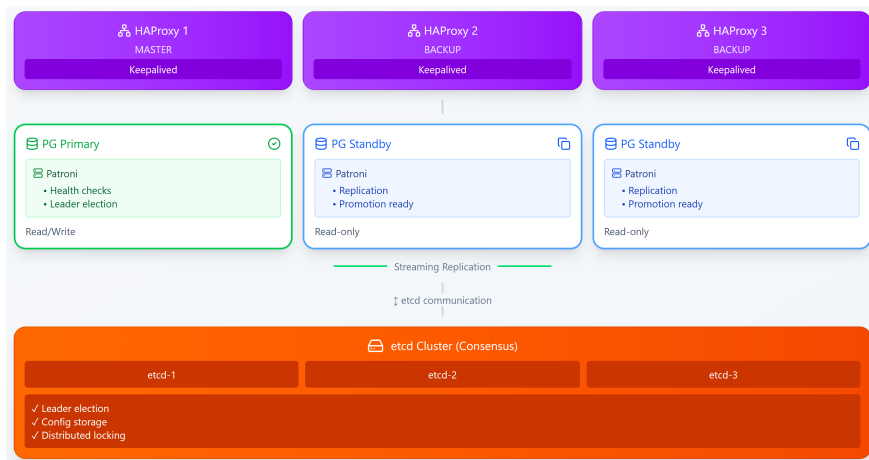
Handling failure



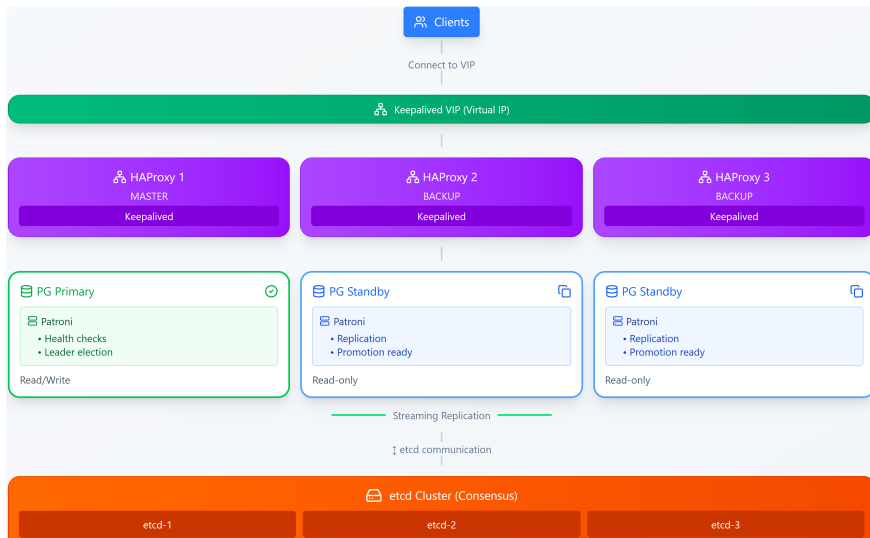
Achieving Failover



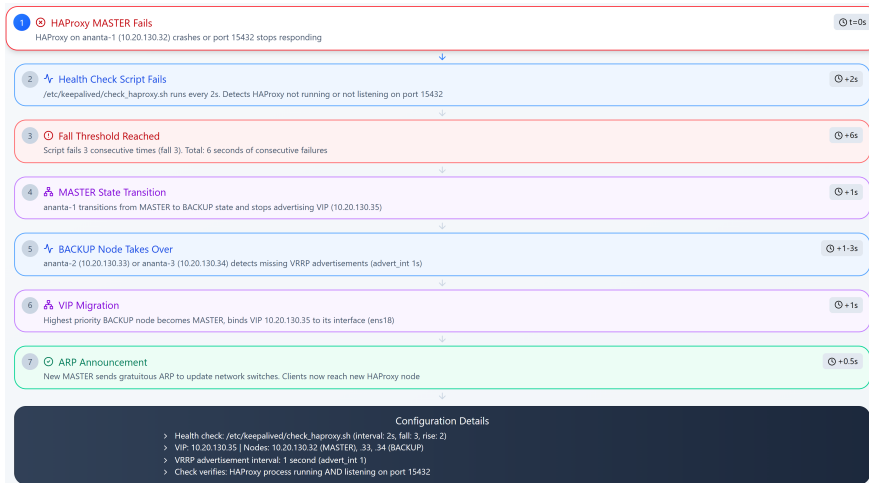
Achieving Failover: Continued



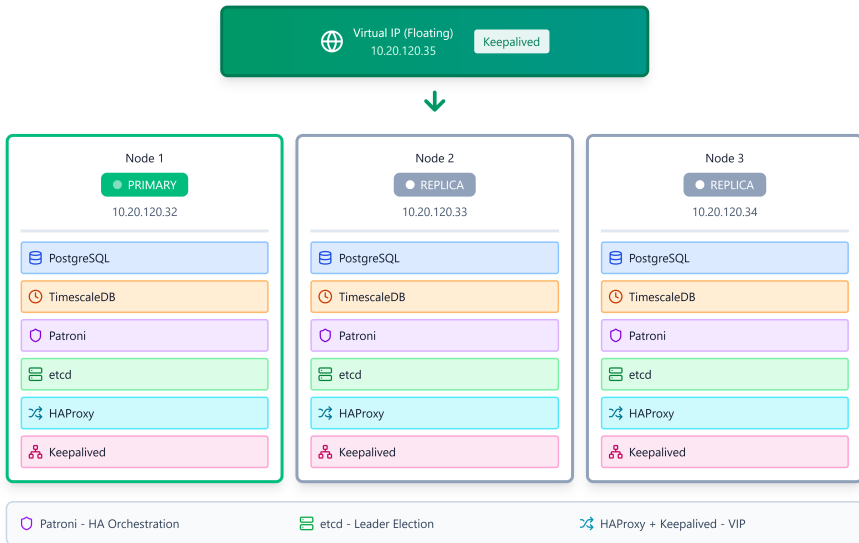
Achieving Failover: Continued



Handling failure in HAProxy and Keepalived



Current Architecture



Performance Benchmarking

- Test dataset(Synthetic): 560 channels, 1 Hz frequency, 15 days of data.
- 1.29 million rows, 3.3 GB table size.
- Queries benchmarked across generic PostgreSQL and HA PostgreSQL+TimescaleDB.

Operation	PostgreSQL (ms)	HA TimescaleDB (ms)	Improvement (%)
5 min data	2754	49	98.2
10 min data	2211	78	96.5
1 hr data	2125	449	78.8
24 hr data	8491	10579	-24.6

Key takeaway: TimescaleDB provides near real-time responsiveness for short-interval queries essential in fusion diagnostics.

Conclusion

- HA TimescaleDB architecture provides fault (hardware/network issues) tolerance and reliability.
- Load distribution for read-heavy workloads
- TimescaleDB enhances the performance for short queries.
- Grafana enables intuitive visualization and trend analysis.

Future Work

- Migration to **Kubernetes** for scalability and self-healing.
- **Natural Language Querying** through LLM-based Grafana plugins.
- **Annotated Graphs** with AI-generated explanations.

Thank You

prem.kumar@ipr.res.in