EDB Postgres Distributed Always On Architectures

AUTHORED BY:

Gianni Ciolli
VP, Solutions Architecture

Marc Linster
Chief Technology Officer

Petr Jelinek
Chief Architect, Replication and HA

POWER TO POSTGRES

Contents

INTRODUCTION

This whitepaper describes the blueprint for Always On architectures for Postgres. These architectures reflect EDB's recommended practices and help customers achieve highest possible service availability in multiple different configurations, ranging from single-location architectures all the way to complex distributed systems that protect from hardware failures and data centre failures. The architectures leverage EDB Postgres Distributed's multi-master capability and leverage its ability to achieve 99.999% availability including maintenance operations.

This whitepaper builds on **EDB Postgres Distributed: The Next Generation of Postgres High Availability** and **The End of the Reign of Oracle RAC: EDB Postgres Distributed Always On**.

EDB Postgres Distributed can be used for architectures beyond the examples described in this document. Use-case specific variations have been successfully deployed in production, but such variations have to undergo rigorous architecture review first, and EDB's standard deployment tool for Always On architectures, TPAexec, has to be enabled to support the variations before they can be supported in production environments.

1. Overview of the Always On Architectures

Overview of the Always On Architectures

EDB has identified 4 standard architectures:

- Always On Bronze: Single active location (data centre or availability zone)
- Always On Silver: Single active location with redundant hardware to quickly restore failover capability and a backup in a disaster recovery (DR) location
- Always On Gold: Two active locations
- Always On Platinum: Two active locations with additional redundant hardware in a hot standby mode

EDB Postgres Distributed consists of two key components:

- Bidirection Replication (BDR) - a Postgres extension that creates the multi-master mesh network
- High Availability Router for Postgres (HARP) - a connection router that makes sure the application is connected to the right BDR nodes.

All Always On architectures protect an increasing range of failure situations. For example Always On Bronze protects against local hardware failure, but does not provide protection from location (data centre or availability zone) failure. Always On Silver makes sure that a backup is kept at a different location, thus providing some protection in case of the catastrophic loss of a location, but the database still has to be restored from backup first which may violate recovery time objective (RTO) requirements. Always On Gold provides two active locations connected in a multi-master mesh network, thus making sure that service remains available even in case a location goes offline. Finally, Always On Platinum adds redundant hot standby hardware in both locations to maintain local high availability in case of a hardware failure.

Each architecture can provide zero recovery point objective (RPO), as data can be streamed synchronously to at least one local master, thus guaranteeing zero data loss in case of local hardware failure.

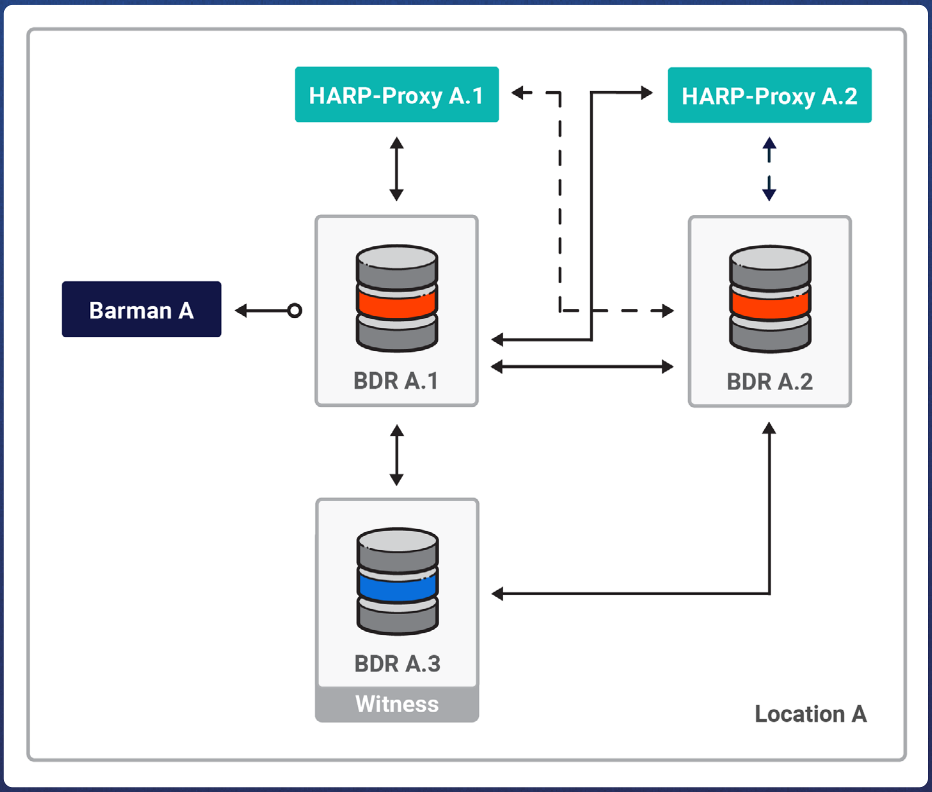
Increasing the availability guarantee always drives additional cost for hardware and licences, networking requirements, and operational complexity. Thus it is important to carefully consider the availability and compliance requirements before choosing an architecture.

2. Architecture Details

Architecture Details

Postgres Distributed uses a Raft (https://raft.github.io) based consensus architecture. While regular database operations (insert, select, delete) don't require cluster-wide consensus, Postgres Distributed benefits from an odd number of nodes to make decisions that require consensus, such as generating new global sequences, or distributed DDL operations. Thus, even the simpler architectures always have three nodes, even if not all of them are storing data. Always On Gold and Platinum, which use two active locations, introduce a fifth node as a witness node to support the RAFT requirements.

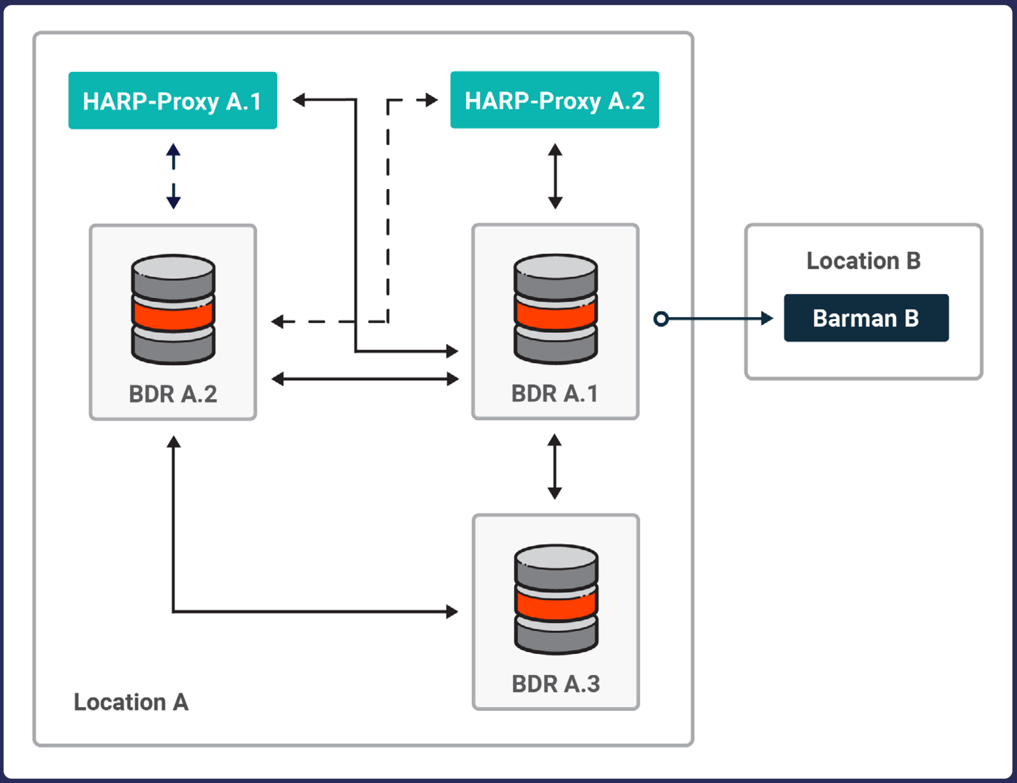
2.1 Always On Bronze



Always On Bronze (single active location)

- 3 BDR nodes (one is a witness only and does not hold data)
- Barman node for backup and recovery can be offsite
- 2 HARP nodes
- BDR nodes could also be spread across availability zones

2.2 Always On Silver

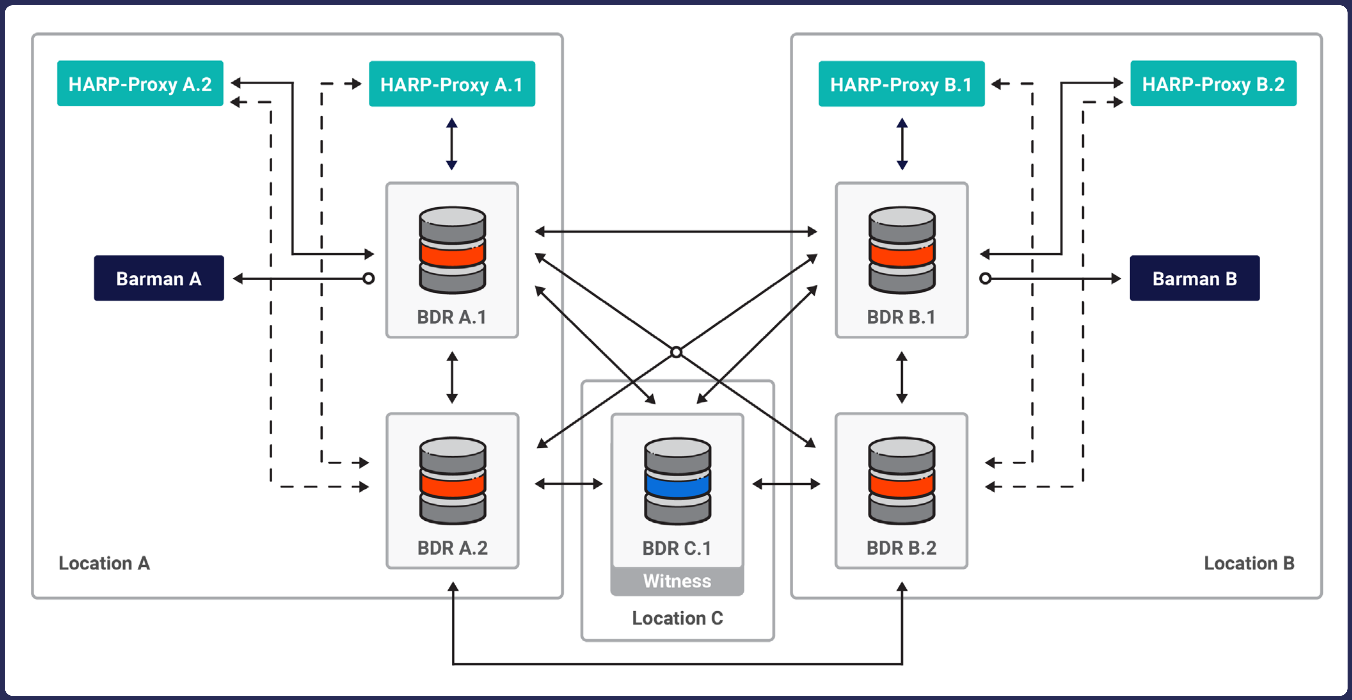


Always On Silver (single active location, backup in DR location)

- 3 BDR nodes, each node has data
- Redundant hardware to quickly restore failover capability
- Barman offsite (offsite is optional, but recommended)
- 2 HARP nodes

BDR nodes could also be spread across AZs, and Barman could be located in Location A

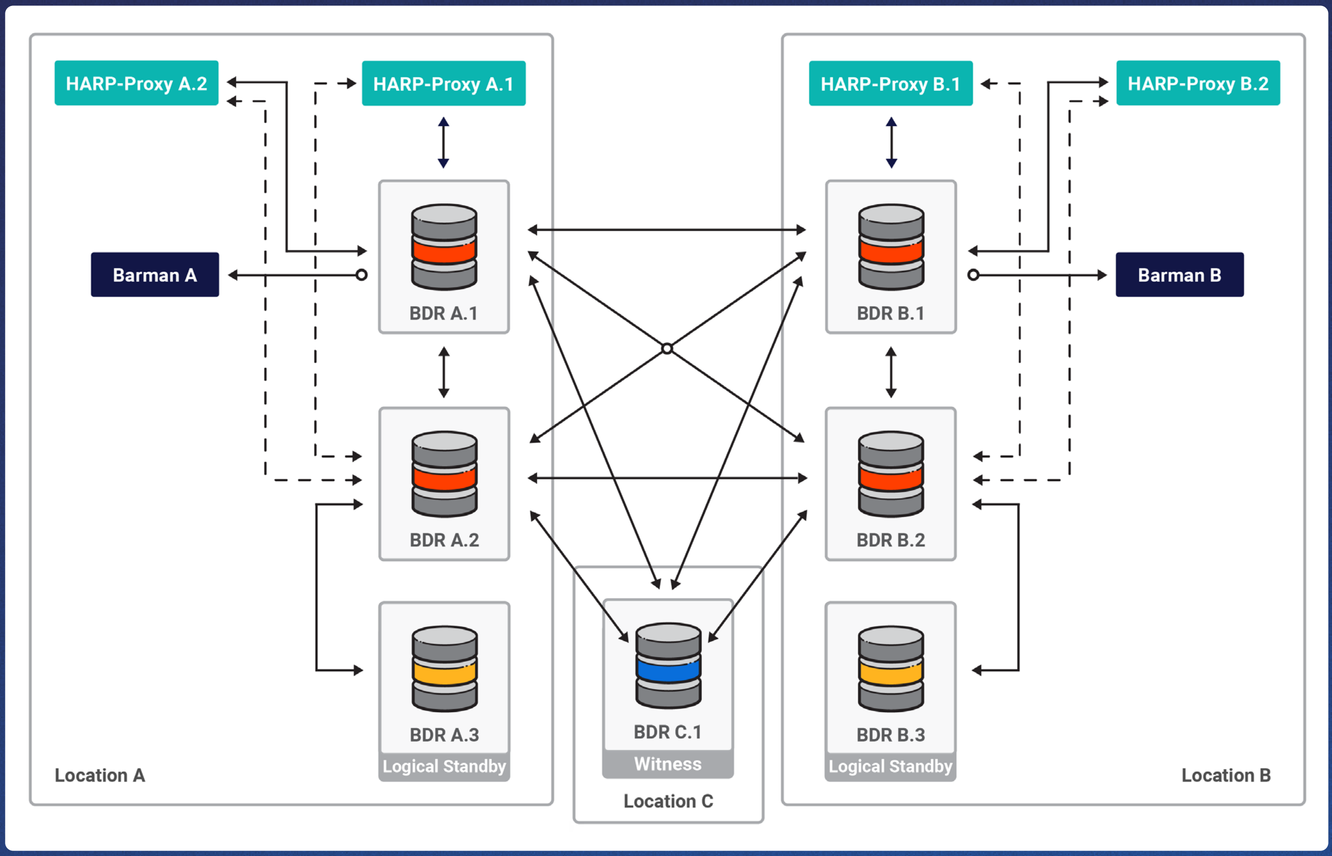
2.3 Always On Gold 3DC, 2 Lead Masters



Always On Silver (single active location, backup in DR location)

- 4 BDR nodes (2 in location A, 2 in location B)
- 2 Barman nodes (1 in location A, 1 in location B)
- 1 witness node in location C (optional, but recommended)
- 4 HARP (2 in location A, 2 in location B)

2.4 Always On Platinum



Always On Platinum (two locations; fast HA restoration)

- 4 BDR nodes (2 in location A, 2 in location B)
- 2 logical standbys (1 in location A, 1 in location B)
- 2 Barman nodes (1 in location A, 1 in location B)
- 1 witness node in location C (optional, but recommended)
- 4 HARP nodes (2 in location A, 2 in location B)

Both locations have redundant hot standby hardware to maintain local high availability in case of a hardware failure without waiting to restore a BDR node from backup.

Applications connect to the standard Always On architectures via multi-host connection strings, where each HARP server is a distinct entry in the multi-host connection string. Other connection mechanisms have been successfully deployed in production, but they are not part of the standard Always On architectures.

3. How to Choose the Right Architecture

How to Choose the Right Architecture

This section discusses criteria that help in selecting the appropriate Always On Architecture.

	Always On Bronze	Always On Silver	Always On Gold	Always On Platinum
Hardware failure protection	Yes	Yes	Yes	Yes
Location failure protection	No (unless Barman is moved offsite)	Yes - Recovery from backup	Yes - instant failover to fully functional site	Yes - instant failover to fully functional site
Failover to DR or full DC	DR (if Barman is located offsite); NA otherwise	DR (if Barman is located offsite);	Full DC	Full DC
Zero downtime upgrade	Yes	Yes	Yes	Yes
Support of AZs in public/ private cloud	Yes	Yes	Yes	Yes
Fast local restoration of high availability after device failure	No; time to restore HA: (1) VM prov + (2) approx 60 min/500GB	Yes; three local BDR nodes allow to maintain HA after device failure	No; time to restore HA: (1) VM prov + (2) approx 60 min/500GB	Yes; logical standbys can quickly be promoted to full BDR nodes
Cross data centre network traffic	Backup traffic only (if Barman is located offsite); none otherwise.	Backup traffic only (if Barman is located offsite); none otherwise.	Full replication traffic	Full replication traffic
BDR license cost	2 BDR nodes	3 BDR nodes	4 BDR nodes	4 BDR nodes 2 logical standbys

4. Deployment and sizing considerations

Deployment and sizing considerations

For production deployments, EDB recommends a minimum of 12 cores for each BDR server and logical standbys. Witness nodes do not participate in the data replication operation and do not have to meet this requirement. Logical standbys should always be sized exactly like the BDR nodes to avoid performance degradations in case of a node promotion. In production deployments HARP nodes require a minimum of four cores. EDB recommends detailed benchmarking of performance requirements, ideally in close collaboration with EDB's Professional Services team.

For development purposes, BDR nodes should not be assigned less than two cores. The sizing of Barman nodes depends on the database size and the data change rate.

BDR nodes, logical standbys, Barman nodes, and HARP nodes can be deployed on virtual machines or in a bare metal deployment mode. However, anti-affinity between BDR nodes and HARP nodes needs to be maintained:

- Multiple BDR nodes should not be on VMs that are co-located on the same physical hardware, as that reduces resilience.
- Multiple HARP nodes should not be on VMs that are co-located on the same physical hardware, as that reduces resilience.
- Single HARP nodes can be co-located with single BDR nodes when deployed as VMs.
- When HARP nodes are colocated with BDR nodes, then separate VMs should be used to assure proper CPU and memory resource assignment.

5. Conclusion

Conclusion

EDB Postgres Distributed provides four standard architectures to support a varying degree of availability requirements, ranging from single-location architectures to multi-location architectures that include redundant hardware components. These architectures have been proven in production and leverage Postgres Distributed's logical replication and mesh-based architecture to achieve industry-leading Postgres availability of 99.999% in public and private cloud deployments.

About EDB

EDB provides enterprise-class software and services that enable businesses and governments to harness the full power of Postgres, the world's leading open source database. With offices worldwide, EDB serves more than 1,500 customers, including leading financial services, government, media and communications and information technology organizations. As one of the leading contributors to the vibrant and fast-growing Postgres community, EDB is committed to driving technology innovation. With deep database expertise, EDB ensures high availability, reliability, security, 24x7 global support and advanced professional services, both on premises and in the cloud. This empowers enterprises to control risk, manage costs and scale efficiently. For more information, visit **www.enterprisedb.com**.

EDB Postgres Distributed Always On: **The End of the Reign of Oracle RAC**

EnterpriseDB Corporation
34 Crosby Drive
Suite 201
Bedford, MA 01730