

Deploying Postgres databases in containers

Getting started

Contents

- What is driving containers?
- Why choose containerized databases now?
- The bottom line
- How EDB accelerates containerization
- Who benefits from using containers?
- Conclusion

Introduction

IBM is committed to the value of containers—in particular with the widespread rollout of Red Hat® OpenShift® Kubernetes containers for the IBM Cloud Pak® for Data platform. In this white paper, IBM Business Partner® EDB explores the factors driving the growth of containers, as well as the opportunities and challenges of a specific container use case: databases.

Today it is safe to say that containers and container orchestration have left the hype phase and are quickly becoming mainstream application development technologies. The container growth trends are compelling:

- According to a [survey of over 600 IT decision makers](#) conducted by ClearPath Strategies for the Cloud Foundry Foundation, 30% of enterprises have already deployed containers and 42% are in the evaluation stage.
- A [Sysdig report](#) shows that the median number of containers per server now stands at 15, up 50% over the previous year.
- Containers are not only being experimented with in dev/test environments. A [Portworx survey](#) of more than 500 IT professionals revealed that in 2018, 83% of the enterprises running containers were using them in production, versus 67% in 2017.
- Diamanti's [Container Adoption Benchmark Survey](#) clearly shows that container use cases are becoming far more diverse, including new cloud-native applications (~55%), lightweight stateless apps (~39%), modernizing legacy apps (~31%), and databases (~30%).
- [451 Research](#) reports that revenues from container technologies in 2018 were \$1.53 billion, which is a little more than three times the revenues in 2015. By 2020, container technology revenues are expected to grow by another 75%, reaching \$2.7 billion.

But what is a container, exactly? A container is a resource-isolated software object that packages all the code, configurations, and dependencies needed to execute and run an application on any Linux® or Windows platform, regardless of the environment (on-premises, public or private cloud, or even bare metal). Because multiple containers can share a single operating system (OS), they are significantly more lightweight than virtual machines (VMs). Container orchestration platforms, such as Kubernetes, automatically launch clusters of containers and manage their entire lifecycle, making it easier to deploy containers at scale.

What is driving containers?

In *The Road to Abstraction*, Stephen O’Grady places containers on a continuum of virtualization that started in the 1960s with COBOL, which buffered program developers from the bits and bytes of machine language.

Another significant node on the virtualization continuum was Java Application Servers, which emerged at the turn of the millennium to provide an abstraction layer between Java apps and the underlying OS and hardware infrastructure.

In parallel, VMs—with their ability to abstract and standardize infrastructure deployments on commodity servers—became popular. VMs are still widely deployed in data centers around the globe, and will most likely be used for many years to come. However, there are numerous factors that have driven, and will continue to drive, the growth of container usage—whether alongside or instead of VMs. These drivers include:

- With more and more born-in-the-cloud apps, containers, along with microservices and serverless, are an integral part of the trend towards cloud-native technology.
- By abstracting apps completely from the OS and infrastructure layers, containers provide unprecedented agility and flexibility to support a [DevOps](#) approach of continuous integration/delivery/deployment. In addition, container images launch much faster than VMs, making them more suitable for today’s dynamic runtime environments, in which apps are expected to scale up and down on demand.
- Containers save costs throughout the application lifecycle, from dev/test to production. Some [quantified examples of these savings](#) are:
 - Reduced hypervisor licensing fees that can lower production environment costs by as much as 50%.
 - Process optimizations that not only streamline dev/test costs by as much as 70%, but also shorten time-to-market so that revenues and other business values are captured more quickly.
 - Thrifty runtime resource consumption as compared to VM deployments, since you don’t need to run as many copies of the OS—up to 80% fewer servers.
- As opposed to proprietary VM platforms, containers have been primarily open source. Enterprise development teams using containers benefit from a vibrant open-source community. They also avoid vendor lock-in, with [83% of containers](#) today running on the royalty-free Docker Engine.
- From a developer and operator standpoint, container images run exactly the same in development, staging, and production, regardless of where they are deployed. This consistency eliminates the “it works on my machine” response in which problems arise due to a discrepancy between environments.

All of these benefits have undoubtedly contributed to the enthusiastic adoption of containers.

However, EDB believes that the single most important container value proposition is **application portability**, which is alluded to in the last bullet point. With increasingly diverse and complex hybrid and multi-cloud environments becoming the norm, it is a huge benefit to be able to operate the same software in the same way across multiple public and private clouds, and even in on-premises virtualized environments. All the app deployment and infrastructure requirements are packaged in the container image, which can be launched anywhere, with full confidence that it will perform as expected.

It is also important to acknowledge that containers could not have gone mainstream without the emergence of container orchestration frameworks that manage and deploy containers across clusters, on demand. The leading framework by far is Kubernetes, the open-source pioneer of container management that originally came out of Google.

Kubernetes is also the underlying foundation for other frameworks, such as OpenShift, GKE, Amazon EKS, and IBM Cloud®, and is now being incorporated into frameworks that were not originally Kubernetes-based, such as Docker Enterprise and Cloud Foundry with PKS.

With orchestration, containers are automatically reconfigured, scaled, upgraded, updated, and migrated without disrupting applications and services. The orchestration framework monitors the health of the runtime container deployment and ensures high availability and continuity in failover or disaster scenarios. It can also monitor load and automatically scale services up and down in response to changes in demand. The container orchestrator facilitates networking among containers or to external environments. It handles data storage in general, and statefulness in particular, through the management of attached persistent or ephemeral volumes, which can be local or in the cloud.

Why choose containerized databases now?

A containerized database is an encapsulation of its DBMS server software, with access to a physical database file residing somewhere within the network.

Each DBMS is encased in its own container image. Containerizing a database, however, is not quite as straightforward as containerizing an application. Some of the database containerization challenges often cited are:

- Databases typically require high-throughput, low-latency networking. However, Docker containers do not natively provide the level of storage and network resource isolation that is necessary to achieve these requirements.
- The considerable disk space required to store large amounts of data in a containerized database makes it less agile and less relocatable.
- Databases are inherently stateful and durable, while containers are typically stateless and ephemeral. The workarounds put into place to handle persistent data storage and longer-than-usual container lifespans often detract from the key container benefit of reduced runtime resource usage.
- Databases typically have numerous tuning parameters, many of which are dynamic. Building a new immutable container image for every possible database configuration can quickly result in image sprawl. It should be noted, however, that this issue is even more of a challenge in VM deployments, since containers are considerably more lightweight than VMs.

An alternative to containerized databases is Database-as-a-Service (DBaaS), an API-based cloud service model whereby the service provider is responsible for the required database physical infrastructure and server-side DBMS resources, including performance configurations.

The bottom line

Ultimately, each project team must make its own decision about whether a containerized database or DBaaS is the best cloud database provisioning fit for the overall application architecture and its development process. The primary drivers for a DBaaS cloud deployment are typically productivity and simplified administration, while for a containerized database, they are usually portability and automation. In any case, when choosing the optimal pathway to a database in the cloud, the following containerized database characteristics should be taken into account:

- Smaller containerized databases are well-suited to the microservices architecture that is now often favored over large monolithic applications. In contrast to a DBaaS database deployment, where the database is a central resource for multiple applications, a containerized database becomes, in essence, a component of the specific application that it serves.
- When the automated/scripted deployment of containerized databases is used in conjunction with an orchestration framework, the result works together with traditional clustering to ensure high availability for both stateful and stateless systems.
- The inherent elasticity of containerized databases supports more flexible, and hence, less wasteful upfront database capacity planning and provisioning. With containers, you can approach the database as an on-demand utility.
- Similarly, because containerized databases separate storage from compute, storage performance and capacity can be scaled independently of compute resources.
- A software-defined containerized database provides a crucial missing link in high-velocity DevOps cycles, allowing development and operations teams to collaborate seamlessly.

One indication that containerized databases are trending is that Postgres, a well-known open-source relational database, is currently the **third most popular technology being run on Docker**.

Top technologies running on Docker

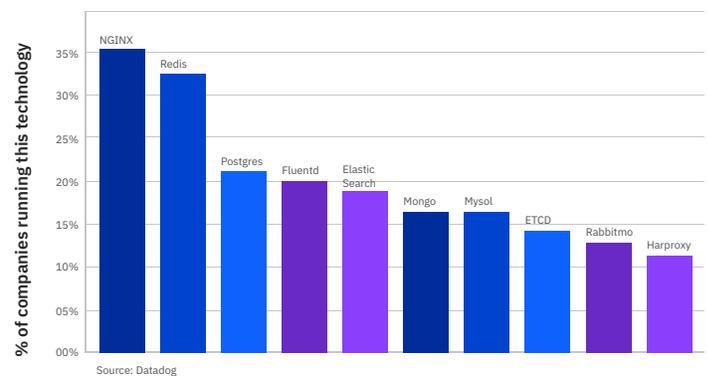


Figure 1: Top technologies running on Docker.

How EDB accelerates containerization

Before considering the architecture of a containerized Postgres deployment, it is important to understand what a standard EDB environment configured for high availability looks like (Figure 2):

1. Monitoring and administration

Each Postgres server is monitored by an agent that reports its tracked data to an administration tool.

2. High availability

EDB streams incremental changes from the master to any number of read-only standby replicas. A high-availability agent watches for failures and automatically triggers failover protocols according to the nature of the detected failure.

3. Disaster recovery

EDB maintains a backup data store and automatically orchestrates enterprise-grade disaster recovery procedures as required.

4. Scalability

EDB is one of the major contributors to the pgPool proxy, which helps applications scale by load balancing read transactions to the replicas while directing write requests to the master. There is also a redundant pgPool proxy to which the application is automatically connected by the high-availability agent in a failover situation.

Typical deployment

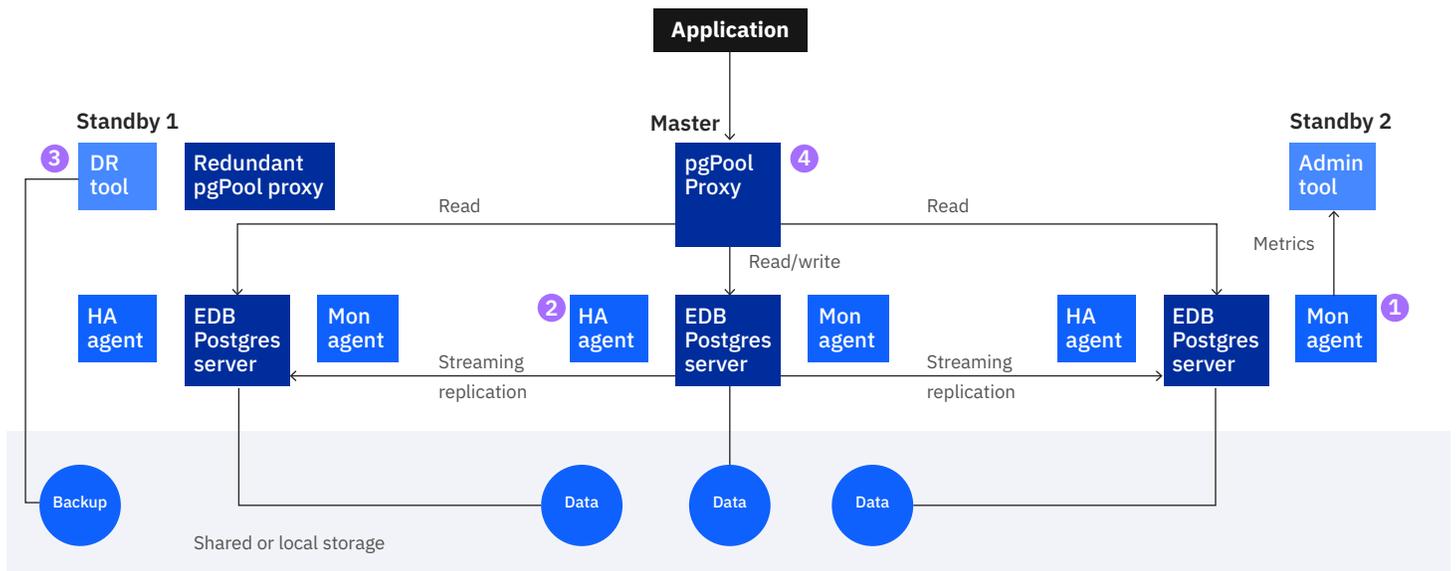


Figure 2. Standard (not containerized) high-availability Postgres database deployment

In order to create a resilient Postgres architecture that can operate at scale, the following four capabilities must be included:

- Failover management for high availability
- Database monitoring and administration management
- Backup for disaster recovery
- Query routing and load balancing for scalability

EDB Postgres is delivered as four containers, all of which are managed by Kubernetes:

1 Database administration

Postgres Enterprise Manager (PEM), packaged in this container, monitors the database, collecting performance and status data that's displayed within dashboards and analyzed for alert conditions. These alerts can be relayed to operators or to other enterprise-level management systems.

2 High availability and data management

This container includes EDB Postgres Advanced Server along with EDB Failover Manager (EFM). Users can have Kubernetes spin up several replicas of a master container so if the master fails, one of the replicas can take over automatically.

3 Disaster recovery

This container includes the EDB Backup and Recovery Tool (BART), which can back up databases in multiple different containers. This enables BART to oversee multiple deployments.

4 Read scalability

This container delivers query routing and connection pooling using pgPool. The benefit of this container is scaling read activity where pgPool acts as a load balancer routing queries to replica databases. It sits in front of the other containers and can be scaled independently of the database container.

Recommended containerized deployment

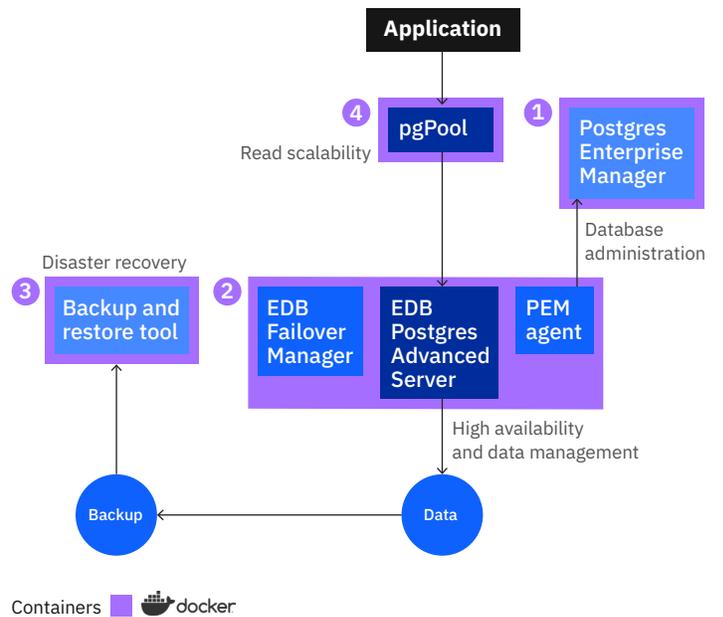


Figure 3. Recommended containerized deployment of EDB

Who benefits from using containers?

Anyone who wants to modernize their infrastructure

EDB is platform agnostic and can be deployed on any Kubernetes platform such as Red Hat OpenShift, Google Kubernetes Engine, Pivotal Cloud Foundry, Docker, IBM Cloud, or Amazon Web Services (AWS).

Figure 4 shows a typical configuration of EDB in a production environment, with a master and two standby pods for high availability and failover. Standby pods are a way to manage a deployment so that redundant database servers can be on separate physical hardware.

Management and monitoring is done through PEM, with shared storage being the preferable data storage practice in production for the database itself, as well as for backup and recovery.

Recommended Kubernetes deployment

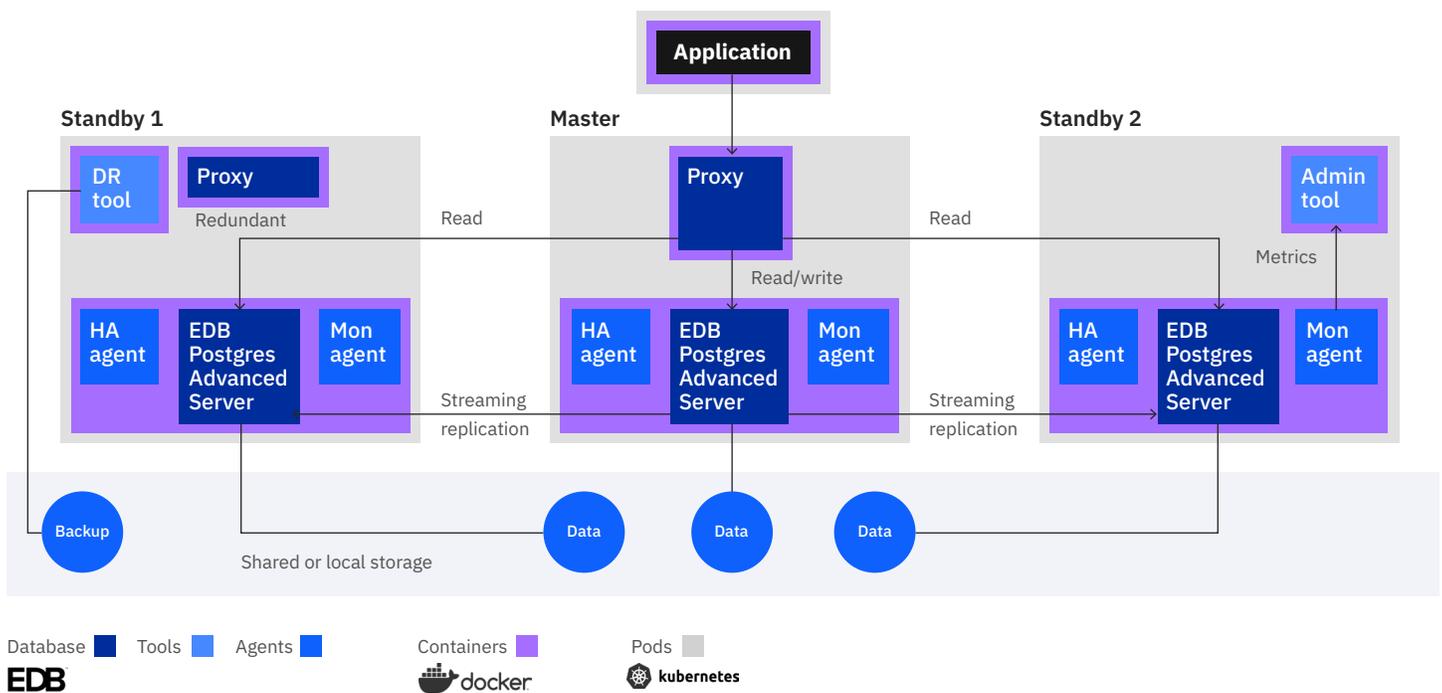


Figure 4. Recommended containerized deployment of EDB

EDB customers can get immediate out-of-the-box benefits from containerized Postgres databases. Typically, customers start with simple applications and move to more complex systems over time. EDB works closely with its customers, sharing with them best practices for effectively leveraging containers in next-generation, microservice-based applications.

Through its [Architectural Roadmap and Solution Blueprint Service](#), EDB solution architects help their customers design a customized container roadmap for getting from where they are today (struggling with large and cumbersome databases), to a new world of small nimble deployments that are fully aligned with their strategic business objectives. Leveraging a reference library of proven blueprints, full-stack tool sets, and API integration designs, EDB customers can accelerate their digital transformation initiatives, model a robust open-source-based data architecture, and deploy rapidly across complex environments.

Conclusion

Containers and container orchestration have matured to the point that they are now positioned at the very core of cloud-native initiatives. Like VMs, containers abstract the application layer from the compute/storage/network infrastructure layer. Going one step further, however, containers also abstract the operating system layer. The result is a lightweight, resource-optimized, self-contained application that performs consistently across a wide range of diverse platforms. Enterprises around the globe are using containers for their dev/test and production workloads to rapidly develop and deploy born-in-the-cloud products and services that scale effortlessly.

In addition to containerized applications, containerized databases have emerged as part of the paradigm shift from large monolithic applications to applications composed of microservices. Rather than a large centralized database that serves multiple applications, containerized databases have become an on-demand utility that is an integral part of the application itself. Despite these advantages, containerized databases have also raised a unique set of challenges in terms of high data availability, backup and recovery, and other critical database performance and compliance requirements.

However, enterprise-grade database containerization solutions from IBM and EDB allow enterprises to benefit from containerized databases without forfeiting critical database administration and monitoring requirements. For example, using native Postgres Docker containers orchestrated with Kubernetes, EDB ensures high data availability, as well as seamless failover, scalability and load balancing, automated monitoring and tuning, and robust backup and recovery. And deploying database solutions such as [IBM® Db2® and Db2 Warehouse](#) as a microservices-based data platform via IBM Cloud Pak for Data can help you organize and analyze your data, ultimately infusing AI capabilities throughout your enterprise.

Next steps

Visit the [EDB website](#) to learn more and try EDB Postgres on Kubernetes—a containerized database management system orchestrated by Kubernetes.

Try [IBM Cloud Pak for Data](#), a containerized platform that includes solutions from the Db2 family of products for database and data warehouse needs.



© Copyright IBM Corporation 2020

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the United States of America
August 2020

IBM, the IBM logo, ibm.com, IBM Cloud Pak, IBM Business Partner, IBM Cloud, and Db2 are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at www.ibm.com/legal/copytrade.shtml.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

The registered trademark Linux® is used pursuant to a sublicense from the Linux Foundation, the exclusive licensee of Linus Torvalds, owner of the mark on a worldwide basis.

Red Hat and OpenShift are registered trademarks of Red Hat, Inc. or its subsidiaries in the United States and other countries. This document is current as of the initial date of publication and may be changed by IBM at any time. IBM Business Partners set their own prices, which may vary. Not all offerings are available in every country in which IBM operates.

The performance data discussed herein is presented as derived under specific operating conditions. Actual results may vary. It is the user’s responsibility to evaluate and verify the operation of any other products or programs with IBM products and programs. THE INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

EDB is not an IBM product or offering. EDB is sold or licensed, as the case may be, to users under EnterpriseDB terms and conditions, which are provided with the product or offering. Availability, and any and all warranties, services and support for EDB is the direct responsibility of, and is provided directly to users by EnterpriseDB.

WGY501DD

