Unlocking the Cloud-Native Data Layer

The enterprise guide for developers, operators, and architects to a future-proof cloud-native data layer

redislabs
HOME OF REDIS

Executive Summary

Cloud computing is all about accelerating innovation and lowering barriers to entry. Successful companies are leveraging the cloud to leverage the power of software to enable new business models across a wide variety of industries, from real-time inventory for retailers to mobile banking and fraud detection for financial services companies.

As more and more companies become software-centric businesses, they are adopting cloud-native strategies designed to deliver the innovative real-time experiences that today's online customers demand. Increasingly, those businesses realize that to get the full benefit from a cloud-native initiative, it must include hybrid cloud and multi-cloud approaches that can offer access to the latest technologies, help control costs, and avoid vendor lock in.

Not surprisingly, this massive infrastructure shift to the cloud is also driving big changes at the application layer. Applications are increasingly moving from monolithic architectures to highly distributed microservices architectures to make software releases faster and make operations more nimble. Applications have also been distributed around the globe in order to serve local customers fast and reliable experiences. And as AI, IoT, and machine learning use cases become increasingly important, companies must deal with exponentially larger data sets. These developments are putting a ton of pressure on the data layer, which must stretch to meet the new requirements of the modern cloud-native world.

Redis Enterprise is perfectly suited to power the modern cloud-native data layer. As the foundation of a primary data platform, Redis Enterprises' in-memory database delivers instant experiences in a highly reliable and scalable manner while unifying any combination of clouds, geographies, and on-premises data centers.

> To stay competitive in a software-centric world, companies of all sizes are adopting cloud-native technologies and strategies to become more efficient and speed up their innovation cycles.

Leveraging the Power of the Cloud

To keep up with increasing pressure of real-time consumer demands, companies across a wide variety of industries are becoming software-centric businesses. This has completely disrupted traditional business models as startups leverage the transformational power of the software to disrupt legacy businesses that remain stuck in the old ways of operating.

To stay competitive in this software-centric world, more and more companies of all sizes are adopting cloud-native technologies and strategies to become more efficient and speed up their innovation cycles. Increasingly, that strategy includes going far beyond the simple lift-and-shift transfer of existing applications into the cloud to include hybrid cloud and multi-cloud approaches designed to wring maximum benefit from the cloud while maintaining optimum flexibility.

This massive infrastructure shift is also driving big changes in the application layer—companies are looking to become more distributed to facilitate nimble and fast development. Companies are restructuring their existing monolithic applications into highly distributed microservices architectures, while new greenfield applications are being designed from the beginning to take advantage of a distributed structure.

Rethinking the data layer in a cloud-centric world

Critically, in dealing with this seismic shift, one area that cannot be forgotten is the data layer. Companies must rethink their data layer strategy as they migrate to the cloud.

It's not optional.

In today's digital economy, if application performance lags for even 2-3 seconds it can have enormous downstream impact on the customer experience—and ultimately the business' success. Data processing must be fast enough to keep up with the real-time business-critical applications that today's consumers demand. If the Uber app, for example, doesn't load instantly or provide a driver within a few seconds, the user is taking her business to Lyft. And as customers interact with their bank via mobile apps and place orders for fast-food restaurants online, for example, this issue is just as critical for traditional businesses as it is for digital startups.

Take Nike. For decades the sports apparel and footwear company was primarily a brick-and-mortar business. Not any more. Today it increasingly relies on direct-to-consumer business initiatives like its SNKRS app. New sneaker launches are released on the app, allowing legions of sneaker-heads to snap up the entire inventory of hot new sneaker releases in seconds. To enable this kind of innovation, Nike has shifted its infrastructure to the cloud, decentralized its developer organizations, and made its application architectures more distributed to enable the pace of innovation necessary to service real-time user needs.

But that's only part of the story. To deliver a consistently fast, satisfying customer experience, the data layer must also be modernized. Applications are generating more and more data, and traditional databases are simply too slow to keep up. They can't meet today's web-scale demands and are not future-proof when it comes to cloud-native applications. It's no longer enough to store this data at rest for later analysis. Companies need to make decisions in real-time for prediction economy. That creates a clear need for a fast, scalable, highly available database platform to deliver instant, real-time application performance while working with huge amounts of data to vast numbers of users in real-time.

American Bankers Association

42% Online

30% Mobile banking

18% Visiting Branches

5% ATMs

> Cloud migration, decentralized developer organizations, and distributed application architectures have created a need for a new, more modern data layer.

A seismic shift in infrastructure and applications

The old world of software development was all about siloed servers with one application running per server. Then came the rise of virtual machines (VMs) and eventually today's transition to increasingly cloud-native and microservices architectures. More and more cloud-native apps are being developed across complex hybrid architectures that may include one or more public clouds, private clouds, and on-premises services and infrastructure.

Tying it all together is a new foundation of increasingly ephemeral compute resources leveraging containers and serverless architectures where resources are highly distributed and spun up and down by the dozens, or even by the thousands, as needed. A Docker container instance, for example, may run for just minutes, or even seconds. Similarly, serverless, or Function-as-a-Service (FaaS) infrastructure such as AWS Lambda can last seconds, if not milliseconds! Driving this transformation of compute resources from physical hardware to VMs, containers, and now to serverless functions is the need to speed up the development process by abstracting away operations and infrastructure concerns so developers can focus on business logic.

As infrastructure has migrated to the cloud, developer organizations have decentralized, and application architectures have become more distributed. These fundamental changes have created a need for a new, more modern data layer equipped for this brave new world of cloud-native, hybrid-cloud, and multi-cloud systems.

What a modern, cloud-native data layer needs to deliver

To underpin today's instant-gratification applications, the data layer must support a few key capabilities, including instant performance, hyper-scale, high availability, global distribution, and hybrid-cloud and multi-cloud capability.

First, a modern cloud-native data layer must deliver instant performance. Traditional databases simply cannot keep up with today's needs. Admittedly, the challenge is great: To avoid introducing delays and bottlenecks into cloud-native applications, the database must be able to respond in less than one millisecond!

Why so fast?

To deliver performance that users perceive as "instant," an app has to respond in about 100 milliseconds, far less than the blink of an eye. But the round-trip from the application to the database typically takes around 50ms, and processing takes another 50ms, leaving almost no time for the database to perform its magic.

Notably, that's significantly faster than traditional databases like Oracle, which typically take up to 100ms to deliver results. Even more-modern and cloud-based databases like Amazon DynamoDB, MongoDB, Azure Cosmos DB and others require 10 - 100ms, which can begin to introduce noticeable delays that affect the customer experience.

> "To avoid introducing delays and bottlenecks into cloud-native applications, the database must be able to respond in less than one millisecond!"

"Instant"
=
<100ms end-to-end response time

Most databases cannot meet instant requirements

> "Five-nines of uptime means your service or application cannot be down for more than 26 seconds per month!"

Second, hyper-scale capabilities in the database layer are growing increasingly critical.
IDC predicts 61% growth in data by 2025, ballooning to a mind-boggling 175 zettabytes world wide—enough to fill a stack of Blu-Ray disks tall enough to reach to the moon 23 times! Data is the lifeblood of business, and a modern, cloud-native data layer must be able to scale easily along with your organization—without sending costs soaring into the stratosphere.

Simply put, your data platform must scale linearly, which means getting 2x the performance requires roughly 2x the infrastructure, 4x performance demands approximately 4x the infrastructure, and so on. That's a real issue, because scaling many of today's database solutions require massive additional infrastructure investments as they accrue non-linear overhead in scaled-out environments.

Third, high availability (HA) of the data layer is now non-negotiable in today's five-nines economy. Five-nines of uptime means your service or application cannot be down for more than 26 seconds per month! Uptime is paramount for software-centric businesses, driven by two big trends. As more and more transactions move online, there's the obvious potential for even short outages or latency to result in huge revenue losses.

Even more important, perhaps, the rise of social media and a press corps eager to pin blame means the reputation of your business is always on the line if your application goes down. You don't want to find your company trending on Twitter or featured in a New York Times headline for the wrong reasons. The hit to your reputation could end up far more costly than the lost revenue.

To a certain extent, of course, failures are inevitable in complex software environments. To avoid these issues impacting your business, a highly available data platform must be designed with the resilience to handle a variety of failure scenarios without disrupting service. Notably, high availability goes beyond the binary distinction of whether your website or service is up or down to cover whether its performance meets customer expectations. If your data layer is making your application unbearably slow, it really doesn't matter than it's technically "up."

Fourth, a modern data layer must support applications in all corners of the world, without compromising performance or availability. The cloud has opened up the ability to run applications across the globe with a lower barrier. In the old world you would have to be a data center in Europe if you wanted to have a local data presence there, but now you can simply spin up compute in your AWS, Azure, or Google Cloud portal and have it set to run in an EU region. As a result, internet-powered businesses are no longer constrained by borders or geography, allowing them to serve customers around the globe. Along with this enormous opportunity come crucial new requirements for advanced geo-distribution to maintain application speed and uptime no matter where the customer is located.

Global distribution is essential to delivering the best possible user experience. Even at the speed of light, response times can be noticeably faster when applications are running closer to their users. Just as important, the data users see must be consistent no matter where they're located, whether it was generated next door or on the other side of the world.

Consider a financial services outfit tracking and analyzing global transactions to detect possible fraud. To detect patterns of possible fraudulent activity in time to stop them from being completed, the underlying data layer must enable local servers to ingest large numbers of local transactions, while simultaneously sharing those metrics with all the other servers around the world. They can't wait to update one location at a time. And, of course, the process must keep working—and not lose precious data or in-process transactions—even during catastrophic system or connectivity failures.

Globally distributed applications like these depend on the resources of multiple systems located around the world. In particular, as the applications themselves are becoming geo-distributed to serve worldwide customer demands, a modern data layer must be distributed and replicated across the world.

> "A highly available data platform must be designed with the resilience to handle a variety of failure scenarios without disrupting service."

> The large cloud infrastructure providers promise a world where everything runs seamlessly inside their walled gardens, but the real world doesn't always work like that.

Finally, a modern, cloud-native data layer must work with all the different flavors of cloud, from different cloud providers to hybrid-cloud and multi-cloud environments and architectures—some observers lump them all together as the "hybrid multi-cloud." While the large cloud infrastructure providers promise a world where everything runs seamlessly inside their walled gardens, the real world doesn't always work like that.

For one thing, even as the cloud continues its meteoric growth, more than 60% of corporate data still resides in on-premises data centers. Many experts see hybrid cloud as the dominant enterprise architecture for 2020 and beyond, as enterprises intentionally combine existing on-premises systems with agile, cloud-based deployments for new projects.

Meanwhile, multi-cloud environments are expected to become much more common for a number of different reasons. Some enterprises will turn to a variety of specialized clouds to run specific workloads better, faster, and more cost effectively, and improved tooling will make it easier to manage these heterogeneous environments. Other companies will work with multiple cloud vendors as a backup option or to mitigate vendor lock-in concerns. Still others will inherit multiple clouds as a kind of fall-out from mergers and acquisitions of companies with divergent technology stacks. Whatever the reasoning behind multi-cloud, a truly useful data platform must be able to run on any cloud or group of clouds, and work with replicated data across clouds and on-premises data centers.

The case for Redis Enterprise

The fastest primary database to power real-time cloud applications

As mentioned above, traditional databases can take 10s to 100s of milliseconds to respond, and that simply doesn't cut it when you need to deliver real-time data for today's applications. Redis Enterprise is a super-fast in-memory database based on a shared-nothing, distributed architecture.

This architecture lets Redis Enterprise scale linearly while maintaining sub-millisecond average latency. Replicating data in-memory makes it available for instant retrieval, whether the data is sourced from the same data center or across multiple clouds. With auto-failover capabilities, the database can recover from failures in single-digit seconds, making sure your customer-facing apps are always on and always performant.

Multi-core architecture enables hyper scalability to support growth and seasonal surges

Redis Enterprise provides linear scalability, allowing you to scale infinitely while maintaining instant performance. Deploying multiple Redis instances on a single cluster node takes full advantage of multi-core architecture. Many managed Redis services (AWS ElastiCache, etc.) rely on single-instance-to-single-cluster node mapping, forcing you to make the scale against performance trade off.

With Redis Enterprise you get the best of both worlds. You can scale to hundreds of terabytes of data while performing hundreds of millions of operations per second. And automatic re-sharding and rebalancing ensures the most efficient use of your infrastructure while promoting superior application performance at infinite scale.

> "Redis Enterprise is a super-fast in-memory database based on a shared-nothing, distributed architecture."

Five-nines uptime to ensure a reliable customer experience

Failing to recover from failures can cost companies big time, in a variety of ways: application performance, data loss, and ultimately customer satisfaction. Redis Enterprise offers high availability (HA) capabilities including diskless replication, instant failure detection with node and cluster watchdogs, and single-digit-seconds failover across racks, zones, and geographies. Redis Enterprise clusters have out-of-the-box HA profiles that let you get up and running with uninterrupted service whether you are using a noisy public cloud or a quieter private cloud or on-premises environment.

Active-Active unifies data across clouds and geographies to lower operational burden and application latency

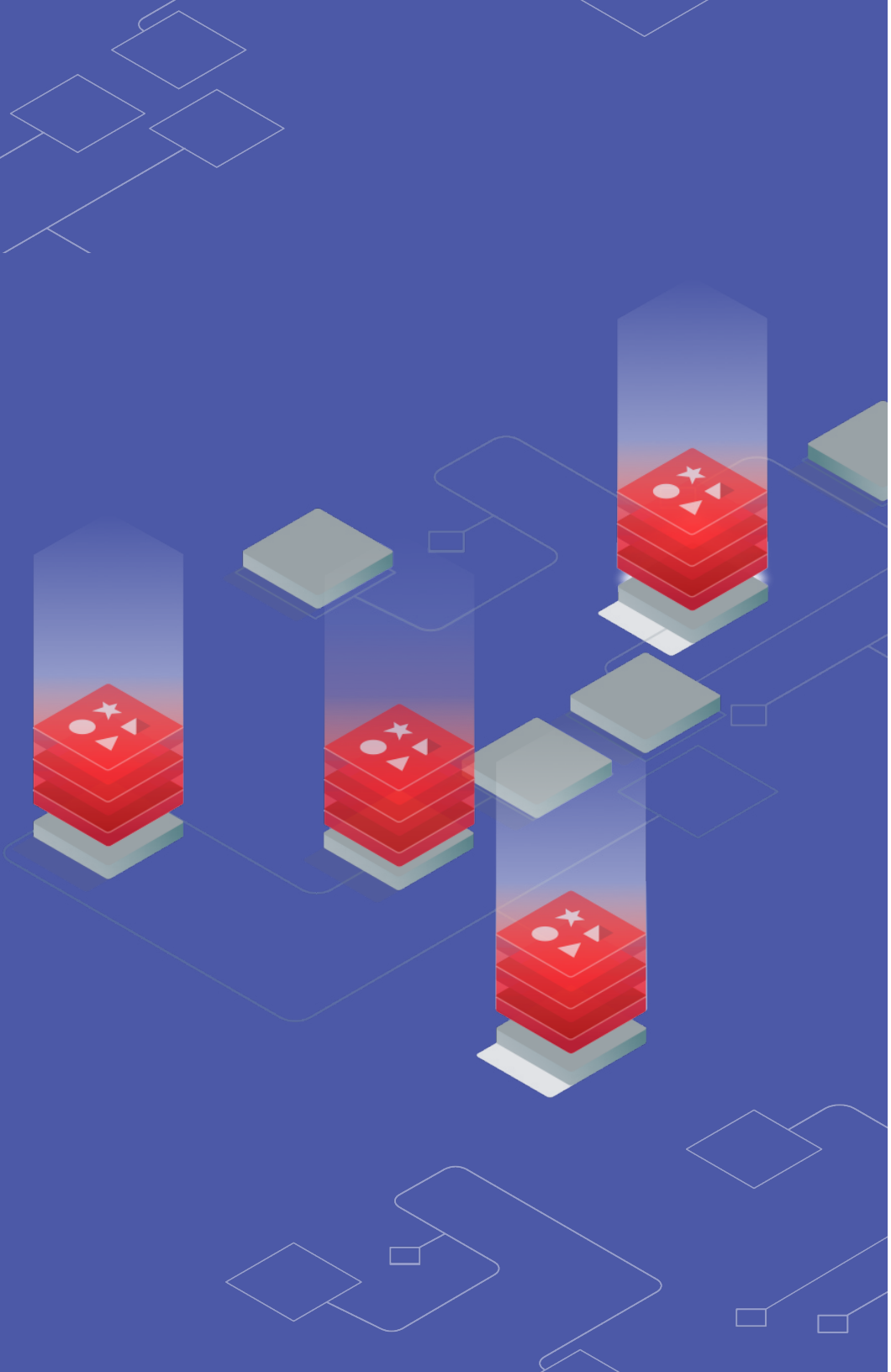
Whether your environment includes applications running on-premises, in a hybrid cloud, or on multiple clouds—or on a mix of all three—you can leverage Redis Enterprise Active-Active Geo Distribution to promote high availability and low latency. With built-in conflict-free replicated data types (CRDT) based active-active database technology, you can achieve high performance across your distributed datasets. This significantly reduces the development effort involved in building modern applications that deliver local latencies even when they need to span racks, clouds, or regions. That means you can focus on moving the needle for your business, not fighting to reduce latency.

Lower total cost of ownership coupled with better performance

As an in-memory database, Redis is built to leverage DRAM's unmatched speed. But to maximize performance at the lowest possible cost, Redis Labs has long been a leader in tiered-database storage, supporting Redis on Flash since 2016 and Intel Optane DC Persistent Memory in 2019. Mobile game developer Etermax, for example, used DRAM/Flash combination saved 70% in infrastructure costs, including cloud hosting fees, with a latency of just 100 milliseconds.

Redis Enterprise works on all the clouds to preserve operational flexibility

Redis Enterprise can be provided as a managed service or as software to run on your own cloud infrastructure—no matter which clouds you use. Running Redis Enterprise as a managed service is the fastest way to deploy Redis Enterprise and get immediate time to value. Redis Enterprise is available via your cloud vendor's marketplace, including AWS Marketplace, Microsoft Azure Marketplace, and Google Cloud Marketplace.



Accelerate and optimize your cloud-native data layer with Redis to drive business outcomes

As the world becomes increasingly cloud centric, it's time for forward looking enterprises to rethink the role of the data layer. This seismic shift in enterprise infrastructure means that a modern, cloud-native data layer needs to deliver radically new and expanded levels of performance, scalability, availability, and global distribution–all with the flexibility to work with any combination of public and private clouds and on-premises data centers.

Redis Enterprise, the most-loved database by developers, is uniquely positioned to meet all those needs—with less management toil and lower overhead for the IT team. Getting started with Redis Enterprise Cloud is easy. You can get your feet wet with free trial with Redis Enterprise Cloud Essentials, or even try out Redis Enterprise Cloud Pro for free. But if you want to scale out Redis Enterprise immediately, Redis Enterprise Cloud Ultimate is your best bet. Learn more about our cloud plans and get pricing here!