SAMPLE

Building the Cloud Computing of Tomorrow

A Practical Guide to Designing, Operating, and Scaling Cloud-Native Infrastructure

Aiman JoeHanz

Free sample: Introduction & Chapter 1

putrabooks.github.io

Buy the full book:

Carousell carousell.sg/p/building-the-cloud-computing-of-tomorrow-ebook-pdf-1465658484

Gumroad

Ko-fi

ElevenReader

Payhip

Contents

Introduction

Chapter 1: The Current State of Cloud Computing

Chapter 2: How Today's Cloud Practices Fail

Chapter 3: Common Mistakes in Cloud Computing Programs

Chapter 4: Foundational Principles of Cloud Computing

Chapter 5: Core Cloud Architectures and Service Models

Chapter 6: Implementation Strategies for Production Cloud Systems

Chapter 7: Technology Deep Dive: From Virtual Machines to Serverless and Edge

Chapter 8: Case Studies: Real-World Cloud Computing Deployments

Chapter 9: Scaling and Future-Proofing Cloud Platforms

Chapter 10: The Future of Cloud Computing

Chapter 11: Action Plan for Organizations Building Cloud Capability

Chapter 12: Governance, Compliance, and Risks on the Horizon

Chapter 13: Trust, Ethics, and the Responsible Cloud Computing Journey

Frequently Asked Questions

Glossary

This free sample contains the Introduction and Chapter 1. The complete book contains the full contents above.

Introduction

Cloud computing has fundamentally reshaped how organizations build, deploy, and operate technology systems. What began as a novel alternative to on-premise data centers has become the dominant infrastructure model for new applications and increasingly the migration target for existing workloads. Yet despite cloud computing's maturity and near-universal adoption, many organizations struggle to use it effectively, realizing only a fraction of its potential benefits while absorbing costs and complexities they did not fully anticipate.

This book exists because the gap between cloud computing's promise and many organizations' actual experience with it remains significant. The promise is compelling: elastic scalability, pay-for-what-you-use economics, reduced operational burden through managed services, and the ability to deploy new capabilities globally in minutes rather than months. The reality, for many organizations, includes unexpected cost overruns, vendor lock-in concerns, security configurations that are easy to get wrong, and operational complexity that can exceed the on-premise systems cloud was supposed to simplify.

This book is for the cloud architect designing infrastructure for production workloads, the engineering leader responsible for building cloud capability or migrating existing systems, and the executive who needs to understand why cloud investment has not delivered the expected returns or what a well-architected cloud strategy actually requires. It assumes basic familiarity with cloud computing concepts but does not require deep expertise in any specific cloud provider's services.

Across the next thirteen chapters, you will learn the current state of cloud computing practice, the architectural patterns that distinguish well-designed cloud systems from fragile ones, the specific mistakes that waste investment and erode the benefits cloud was supposed to deliver, and the operational practices, from cost management to security to observability, that separate cloud platforms delivering sustained value from those that generate constant operational surprises. We will examine real deployment case studies, walk through building cloud capability from first principles, and close with the governance, compliance, and trust considerations that increasingly determine whether a cloud platform can be relied upon for consequential workloads.

Each chapter includes a PRACTICAL INSIGHT callout distilling the single most actionable lesson, and a KEY TAKEAWAYS summary for quick reference. This book does not promise that any specific cloud provider or service will solve your infrastructure problems; providers change, and the organizations that

succeed are those with the underlying architectural discipline to use any cloud platform well. It promises a clear-eyed, practitioner-grounded understanding of what works, what fails, and how to build cloud systems that deliver on the promise rather than just absorbing the cost.

By the final page, you should be able to evaluate whether a proposed cloud architecture is well-designed or over-engineered, recognize the operational practices a production cloud system actually requires beyond initial deployment, and explain to a skeptical stakeholder why migrating to the cloud does not automatically reduce cost or complexity and what it takes to genuinely realize cloud's benefits. Let's begin with where cloud computing practice actually stands today.

The book follows the lifecycle of cloud adoption: understanding the current landscape and its limitations, learning from common mistakes, establishing foundational principles, selecting architectures, implementing production strategies, studying real-world deployments, scaling systems, and preparing for the future. Each chapter builds on the previous, creating a complete picture of what it takes to build and operate cloud systems that deliver real value.

The book is designed for engineers, architects, and technical leaders who are responsible for building, migrating to, or operating cloud systems. It assumes familiarity with basic computing concepts but does not assume deep cloud expertise. The examples are grounded in real production scenarios, and the recommendations are based on patterns that have been proven in practice. Whether you are migrating an existing application to the cloud, building a new cloud-native application, or optimizing an existing cloud deployment, the principles and practices in this book will help you make sound architecture and operational decisions.

Chapter 1: The Current State of Cloud Computing

Cloud computing has become the default infrastructure choice for new applications and a growing proportion of existing workloads, but understanding what cloud computing actually means in practice today, beyond vendor marketing, requires examining its current state with the same honest assessment applied to any mature technology.

The shift from on-premise data centers, where organizations owned and operated their own servers, networking, and storage, to cloud infrastructure provided as a service represents one of the most significant shifts in how organizations consume computing resources. This shift has transferred much of the physical infrastructure management, hardware procurement, data center facilities, and basic infrastructure operations to cloud providers, allowing organizations to focus on their applications rather than the infrastructure beneath them.

This transfer is not absolute, however: organizations operating in the cloud still bear responsibility for configuration, security hardening, application deployment, cost management, and operational monitoring, with the specific division of responsibility depending on which cloud services they use. Understanding this shared responsibility model, discussed in detail in Chapter 4, is essential for organizations that incorrectly assume moving to the cloud means outsourcing all operational responsibility to the cloud provider.

The cloud infrastructure market is dominated by three major providers, each with distinct strengths, service ecosystems, and pricing models. This concentration has benefits: all three have the scale to invest in infrastructure and security at levels most organizations could never achieve independently, and their competition drives continuous innovation in services and pricing. It also creates considerations around vendor lock-in, discussed throughout this book, as organizations investing deeply in one provider's services may find migration to another provider increasingly difficult over time.

Organizations should choose a primary cloud provider based on their specific requirements, including geographic availability, service catalog, pricing structure, and existing organizational expertise, rather than assuming any single provider is universally superior. Multi-cloud strategies, discussed in Chapter 5, can mitigate lock-in risk but introduce their own complexity and should be adopted deliberately rather than by default.

Cloud providers increasingly offer managed services that abstract away

infrastructure management entirely: managed databases, managed container orchestration, managed message queues, and serverless computing platforms where the provider handles scaling, patching, and basic operations. This shift toward managed services allows organizations to focus on application logic rather than infrastructure operations, potentially reducing the operational expertise required to run production systems.

The tradeoff with managed services is reduced control and increased provider dependency: managed services typically offer less customization than self-managed alternatives, and migrating away from a provider's managed service can be more complex than migrating from self-managed infrastructure. Organizations should evaluate managed services per use case, adopting them where the operational benefits outweigh the loss of control and lock-in risk, and using self-managed alternatives where maximum control or portability is required.

Cloud-native application patterns, including microservices, containers, and serverless architectures, have emerged as the natural design approach for applications built to run in the cloud from the outset. These patterns leverage cloud infrastructure's elasticity and managed services in ways that traditional monolithic application architectures cannot, enabling independent scaling of application components, faster deployment cycles, and better fault isolation.

Cloud-native patterns introduce their own complexity, however: distributed system challenges around service-to-service communication, data consistency, and observability that monolithic architectures handle more simply. Organizations adopting cloud-native patterns should do so because their application's scale, team structure, or deployment frequency genuinely benefits from these patterns, not because cloud-native is the currently fashionable approach, a theme explored throughout this book.

Cloud computing's economic model, pay for what you use rather than investing in fixed infrastructure capacity, fundamentally changes how organizations budget for and consume computing resources. This model provides flexibility: organizations can scale resources up during peak demand and down during quiet periods, and can launch new services without upfront infrastructure investment. It also introduces risks: without active cost management, discussed throughout this book, cloud costs can grow unpredictably and exceed what equivalent on-premise infrastructure would have cost.

Understanding cloud economics requires understanding the difference between variable and fixed costs, the implications of different pricing models (on-demand, reserved, spot), and the often-overlooked costs of data transfer,

storage growth, and managed service premiums that can accumulate significantly over time.

Organizations that approach cloud with an on-premise cost mindset, treating infrastructure as a fixed investment rather than a variable expense, consistently encounter budget surprises.

Early concerns about cloud security, primarily around whether data would be safe in infrastructure an organization did not physically control, have largely been resolved for most use cases through the major providers' security capabilities, compliance certifications, and the shared responsibility model. In practice, well-configured cloud infrastructure is often more secure than typical on-premise infrastructure, which is harder to keep consistently patched and configured.

The security challenge in cloud computing is not the physical security of the underlying infrastructure but the configuration of cloud services, which is entirely the customer's responsibility under the shared responsibility model. Misconfigured cloud resources, not provider security breaches, are the leading cause of cloud security incidents, making cloud security primarily a configuration and governance challenge rather than an infrastructure security concern.

As cloud computing has grown to represent a significant portion of global computing infrastructure, its environmental impact, particularly the energy consumption and carbon emissions of the data centers powering cloud services, has become an increasingly important consideration. Major cloud providers have invested heavily in renewable energy and efficiency improvements, and many now offer tools for customers to track and reduce the carbon footprint of their cloud usage.

This sustainability consideration is becoming a factor in cloud provider selection and architecture decisions, particularly for organizations with their own sustainability commitments or regulatory requirements around environmental impact reporting. Organizations should expect sustainability to become an increasingly prominent consideration in cloud computing practice, alongside the traditional considerations of cost, performance, and reliability.

The competition among major cloud providers has driven continuous innovation in services, pricing, and geographic expansion, benefiting customers through better capabilities and lower costs over time. This competition has also led to increasing convergence in basic service offerings, as providers match each other's capabilities, while differentiation has shifted to higher-level managed services, AI capabilities, and specialized offerings for

specific industries or use cases.

This competitive dynamic means organizations choosing a cloud provider should evaluate not just current capabilities but the provider's trajectory: investment in new services, geographic expansion plans, and commitment to open standards. A provider that is strong today may not remain the best choice if its investment trajectory does not align with the organization's future needs.

Cloud computing has affected not just how organizations consume technology but how they structure their technology teams. The shift from infrastructure management to service composition, from capacity planning to cost optimization, and from hardware procurement to architecture design has changed the skills and roles organizations need, leading to new team structures like platform engineering teams and site reliability engineering practices.

This organizational impact should be anticipated during cloud strategy planning, not discovered after migration when the existing team structure proves poorly suited to cloud operations. Organizations that plan for organizational change alongside technical migration consistently transition more smoothly than those that focus solely on the technical aspects.

Beyond the major providers' own services, a large ecosystem of third-party tools has emerged for cloud management, monitoring, security, cost optimization, and deployment. This ecosystem provides capabilities that provider-native tools may not offer, but it also adds complexity in tool selection, integration, and management. Organizations should evaluate third-party tools based on whether they address genuine gaps in provider-native capabilities and whether their value justifies the additional complexity and cost.

The distinction between cloud operations and traditional on-premise IT operations is blurring as organizations adopt hybrid approaches and as on-premise infrastructure increasingly adopts cloud-like operational models, including self-service provisioning, infrastructure as code, and API-driven management. This convergence means the skills and practices that were specific to cloud operations are becoming relevant across all IT operations, not just cloud-deployed workloads.

This convergence benefits organizations that can apply consistent operational practices across both cloud and on-premise environments, but it also means organizations can no longer treat cloud operations as a separate discipline with its own team and practices. Operational excellence increasingly requires

fluency across both paradigms.

As organizations deepen their use of a specific cloud provider, their ability to negotiate pricing decreases because the cost and disruption of switching providers increases. This dynamic means that organizations should consider not just current pricing but the long-term pricing trajectory when choosing a provider and should structure their cloud usage to maintain some negotiating leverage, whether through multi-cloud optionality or by ensuring critical components could be migrated if pricing becomes unfavorable.

Cloud computing is often positioned as a key enabler of broader digital transformation, providing the infrastructure flexibility and managed service capabilities that allow organizations to build new digital experiences, automate operations, and leverage data more effectively. While this positioning is broadly correct, organizations should understand that cloud computing is an enabler of digital transformation, not a substitute for it. Moving to the cloud without rethinking how the organization uses technology to serve its mission produces cloud-hosted versions of the same legacy processes, not genuine digital transformation.

Organizations pursuing digital transformation should view cloud migration as one component of a broader strategy that also includes application modernization, data strategy, organizational change, and process redesign. Cloud computing provides the infrastructure foundation for transformation, but the transformation itself comes from how the organization uses that infrastructure to change how it operates and serves its customers.

Open source software has been fundamental to the development of cloud computing, from the Linux operating systems that power most cloud infrastructure to the container runtimes, orchestration platforms, and infrastructure-as-code tools that have become standard practice. The open source ecosystem has enabled innovation in cloud computing that would not have been possible if all innovation had to come from the major cloud providers alone.

Organizations should understand and engage with the open source ecosystem that underlies their cloud infrastructure, both to make informed decisions about tool selection and to contribute back where they can. The open source community provides an alternative to full dependence on any single provider's proprietary ecosystem, and organizations that engage with it consistently make better-informed cloud technology decisions.

Cloud computing has been a critical enabler of the shift toward remote and distributed work, providing the infrastructure that allows teams to access

applications, data, and development environments from anywhere with an internet connection. This enabling role was demonstrated dramatically during the global shift to remote work, where organizations with cloud-hosted applications and cloud-based development environments transitioned far more smoothly than those dependent on on-premise infrastructure that required physical office access.

This experience has permanently changed expectations about where and how work can happen, and organizations should recognize that cloud infrastructure is not just a technical choice but an enabler of organizational flexibility that affects talent recruitment, business continuity, and the ability to operate across geographic boundaries.

PRACTICAL INSIGHT Before beginning a cloud migration or significant new cloud deployment, conduct an honest assessment of your organization's current cloud readiness, including existing cloud skills, operational practices, and cost management capability. Organizations that migrate to the cloud without this readiness assessment consistently discover gaps in skills and practices after migration, when addressing them is more costly and disruptive than building them proactively would have been.

KEY TAKEAWAYS - Cloud computing has become the default infrastructure for new applications, but the shared responsibility model means organizations retain significant operational responsibility - The cloud market is dominated by three major providers, and provider choice should be based on specific requirements rather than assumed universal superiority - Managed services reduce operational burden but increase provider dependency and should be adopted per use case, not universally - Cloud-native patterns leverage cloud's elasticity but introduce distributed system complexity that should be adopted deliberately, not by default - Cloud's pay-per-use economics provide flexibility but require active cost management discipline that on-premise budgets did not require

- Cloud security is primarily a configuration challenge, not an infrastructure concern, and misconfiguration is the leading cause of cloud security incidents

The cloud computing market is dominated by three hyperscale providers — AWS, Microsoft Azure, and Google Cloud — each with distinct strengths and strategic positions. AWS, the market leader, offers the broadest service catalog (200+ services), the largest partner ecosystem, and the deepest operational experience. Azure's strength is its enterprise integration — Active Directory, Microsoft 365, and enterprise agreement pricing make it the natural choice for organizations already invested in the Microsoft ecosystem. Google Cloud's strength is in data analytics and AI/ML, with BigQuery, Vertex AI, and

TensorFlow support that attract data-intensive workloads.

The competitive differentiation among the three providers has narrowed over time — most core services (compute, storage, databases, networking) are comparable in capability and price. The differentiation is in higher-level services: AI/ML (Google Cloud), enterprise integration (Azure), and breadth and ecosystem (AWS). For most organizations, the provider choice is driven by existing relationships (Microsoft shops choose Azure, data-centric organizations choose Google Cloud, greenfield deployments often choose AWS for its ecosystem), not by technical superiority. The architect should evaluate providers based on the organization's specific requirements rather than on market share or brand reputation.

Beyond the three hyperscalers, alternative cloud providers are carving out niches: DigitalOcean and Linode (now Akamai) serve small businesses and developers with simpler pricing and easier UX than the hyperscalers. Oracle

Cloud targets enterprise database and ERP workloads with aggressive pricing for Oracle software customers. IBM Cloud focuses on regulated industries (financial services, government) with industry-specific compliance offerings. Regional providers (Alibaba Cloud, Tencent Cloud in China; OVH in Europe; LocalStack in various countries) serve data sovereignty requirements that the hyperscalers cannot.

The alternative providers are viable for specific use cases: DigitalOcean for simple web applications and startups, Oracle Cloud for Oracle database workloads, regional providers for data sovereignty. But they lack the service breadth and global scale of the hyperscalers, which limits their applicability for complex, global workloads. The architect should consider alternative providers when they offer a specific advantage (simpler UX, lower cost for specific workloads, data sovereignty) but should evaluate the long-term viability of the provider — a provider that goes out of business or changes strategy can leave customers stranded.

End of free sample. The full book is available at putrabooks.github.io

ABOUT THE AUTHOR

Aiman JoeHanz writes about the technologies that are quietly reshaping the world. His Building of Tomorrow series examines the infrastructure behind artificial intelligence -- from blockchain and cloud computing to agentic systems and token optimization -- and asks how these technologies will evolve over the coming decade.

His books are built on technical research and engineering documentation, written for readers who want to understand systems rather than trends. He covers what is being built today and what it will take to build what comes next.

He believes that understanding infrastructure is the difference between watching the future happen and helping to create it.