The third reason why DevOps takes so long—the essential complexity of this problem—is that there is a genuinely long checklist of tasks that you must do to prepare infrastructure for production. The problem is that the vast majority of developers don't know about most of the items on the checklist, so when they estimate a project, they forget about a huge number of critical and time-consuming details. This checklist is the focus of the next section.

The Production-Grade Infrastructure Checklist

Here's a fun experiment: go around your company and ask, "What are the requirements for going to production?" In most companies, if you ask this question to five people, you'll get five different answers. One person will mention the need for metrics and alerts; another will talk about capacity planning and high availability; someone else will go on a rant about automated tests and code reviews; yet another person will bring up encryption, authentication, and server hardening; and if you're lucky, someone might remember to bring up data backups and log aggregation. Most companies do not have a clear definition of the requirements for going to production, which means each piece of infrastructure is deployed a little differently and can be missing some critical functionality.

To help improve this situation, I'd like to share with you the *Production-Grade Infrastructure Checklist*, as shown in Table 8-2. This list covers most of the key items that you need to consider to deploy infrastructure to production.

Table 8-2. The Production-Grade Infrastructure Checklist

Task	Description	Example tools
Install	Install the software binaries and all dependencies.	Bash, Ansible, Docker, Packer
Configure	Configure the software at runtime. Includes port settings, TLS certs, service discovery, leaders, followers, replication, etc.	Chef, Ansible, Kubernetes
Provision	Provision the infrastructure. Includes servers, load balancers, network configuration, firewall settings, IAM permissions, etc.	Terraform, CloudFormation
Deploy	Deploy the service on top of the infrastructure. Roll out updates with no downtime. Includes blue-green, rolling, and canary deployments.	ASG, Kubernetes, ECS
High availability	Withstand outages of individual processes, servers, services, datacenters, and regions.	Multi-datacenter, multi-region
Scalability	Scale up and down in response to load. Scale horizontally (more servers) and/or vertically (bigger servers).	Auto scaling, replication
Performance	Optimize CPU, memory, disk, network, and GPU usage. Includes query tuning, benchmarking, load testing, and profiling.	Dynatrace, Valgrind, VisualVM
Networking	Configure static and dynamic IPs, ports, service discovery, firewalls, DNS, SSH access, and VPN access.	VPCs, firewalls, Route 53
Security	Encryption in transit (TLS) and on disk, authentication, authorization, secrets management, server hardening.	ACM, Let's Encrypt, KMS, Vault
Metrics	Availability metrics, business metrics, app metrics, server metrics, events, observability, tracing, and alerting.	CloudWatch, Datadog
Logs	Rotate logs on disk. Aggregate log data to a central location.	Elastic Stack, Sumo Logic
Data backup	Make backups of DBs, caches, and other data on a scheduled basis. Replicate to separate region/account.	AWS Backup, RDS snapshots