CHAPTER 2

Predictable Demands

The foundation of successful application deployment, management, and coexistence on a shared cloud environment is dependent on identifying and declaring the application resource requirements and runtime dependencies. This *Predictable Demands* pattern is about how you should declare application requirements, whether they are hard runtime dependencies or resource requirements. Declaring your requirements is essential for Kubernetes to find the right place for your application within the cluster.

Problem

Kubernetes can manage applications written in different programming languages as long as the application can be run in a container. However, different languages have different resource requirements. Typically, a compiled language runs faster and often requires less memory compared to just-in-time runtimes or interpreted languages. Considering that many modern programming languages in the same category have similar resource requirements, from a resource consumption point of view, more important aspects are the domain, the business logic of an application, and the actual implementation details.

It is difficult to predict the amount of resources a container may need to function optimally, and it is the developer who knows the resource expectations of a service implementation (discovered through testing). Some services have a fixed CPU and memory consumption profile, and some are spiky. Some services need persistent storage to store data; some legacy services require a fixed port number on the host system to work correctly. Defining all these application characteristics and passing them to the managing platform is a fundamental prerequisite for cloud-native applications.