

Principles and properties	Structures
✓ Versatility	Module
✓ Conceptual integrity	✓ Dependency
Independently changeable	Process
Automatic propagation	✓ Data access
Buildability	
✓ Growth accommodation	
Entropy resistance	

CHAPTER THREE

Architecting for Scale

Jim Waldo

Introduction

ONE OF THE MORE INTERESTING PROBLEMS IN DESIGNING AN ARCHITECTURE for a system is ensuring flexibility in the scale of that system. Scaling is becoming increasingly important, as more of our systems are run on networks or are available on the Web. For such systems, the idea of capacity planning is absurd if you want a margin of error that is under a couple of orders of magnitude. If you put up a site and it becomes popular, you might suddenly find that there are millions of users accessing your site. Just as easily (and just as much of a disaster), you can put up a site and find that no one is particularly interested, and all of the equipment in which you invested now lies idle, soaking up money in energy costs and administrative effort. In the networked world, a site can transition from one of these states to the other in a matter of minutes.

The scaling problem is faced by anyone who attaches a system to a network, but it is particularly interesting in the case of massively multiplayer online games (MMOs) and virtual worlds. These systems must be capable of scaling to large numbers of users. Unlike web servers, however, where the users are requesting fairly static information and are not interacting with each other, players in an MMO or residents in a virtual world are there to interact with both the world (changing the underlying information in the world) and each other. These interplays complicate the scaling of the infrastructures for such systems, as the user interactions with the