

Whenever we are designing a large system, we need to consider few things:

1. What are different architectural pieces that can be used?
2. How do these pieces work with each other?
3. How can we best utilize these pieces, what are the right tradeoffs?

Investing in scaling before it is needed is generally not a smart business proposition; however, some forethought into the design can save valuable time and resources in the future. In the following chapters, we will focus on some of the core building blocks of scalable systems. Familiarizing with these concepts would greatly benefit in understanding distributed system design problems discussed later. In the next section, we will go through Consistent Hashing, CAP Theorem, Load Balancing, Caching, Data Partitioning, Indexes, Proxies, Queues, Replication, and choosing between SQL vs. NoSQL.

Let's start with Load Balancing.