

Latency Is the Enemy

But the biggest difference in the two environments traces back to the differences in what the users are doing. In an enterprise environment, the goal is to conduct business, and some lags in processing are acceptable if the overall throughput is improved. In the MMO and virtual world environment, the goal is to have fun, and latency is the enemy of fun. So the infrastructure for an MMO or virtual world needs to be designed around the requirement of bounding latency whenever possible, even at the cost of throughput.

Online games and virtual worlds have clearly found ways to scale to large numbers of users. The current mechanisms fall into two groups. The first of these is geographic in nature. The game is designed as a group of different areas, with each area designed to be run on a single server. It might be an island or room in a virtual world or a town or valley in an online game. The design of the game tries to make each geographic area independent, and scale the geographic area in such a way that the server will not be overwhelmed by too many users occupying the area. In practice, such areas are often self-limiting, as when the server is being overwhelmed, the play becomes less responsive and less interesting. As a result, players leave for more interesting areas, which makes the formerly overwhelmed area less occupied and improves response time.

The problem with scaling by assigning geographic areas to different servers is that the decision of what areas scale to a server must be made when the game is being written. Although new areas might be able to be added to a game or world fairly easily, changing the area that is assigned to a server is something that requires changing the code. The decision of what areas are the unit of scale has to be made as part of development.

A second way of dealing with areas that are overcrowded in a game or world is known as *sharding*. A shard is a copy of the area, run on its own server and independent of other shards, that presents the same portion of the game as the original area. Thus, a shard might present a different copy of a particular room or village, allowing twice as many players to occupy that part of the world. The drawback of shards is that they do not allow players in different shards to interact with each other. As games and worlds become more social experiences than simple game play, this disadvantage can be major. The goal of players is not only to be in the virtual world, but to occupy it with their (real or virtual) friends. Sharding interferes with that goal.

Thus, another major goal of the Darkstar architecture is to allow on-the-fly scaling in a way that does not require the game logic to become involved in the scaling. The infrastructure should allow the game to dynamically react to the load rather than make such reaction part of the design of the game.