



## **Reliability and Survivability are Complementary Concepts**

Years ago, I was talking to a VP of one of my data centers about my concern that some operating departments were complaining about the instability of hardware in the data centers. The VP assured me that they were working on the problem and had installed and practiced procedures that would enable the data center to move the workload from one system to another in two minutes. A month later I was having a meeting with a senior executive of one of those operating departments when his admin interrupted our meeting to tell us that the data center was “down”. With confidence, I waited for the admin to return with a report that everything was back up, but the admin never returned. Finally after thirty minutes, I asked the executive I was meeting with to ask his admin if the data center problem was resolved, only to be told it wasn't. I called the VP in the data center and asked “What happened to your famous two minute move”. After a pause she responded “unfortunately it takes an hour or more to make the decision to make the two minute move”. From this experience, I quickly became a proponent of reliability as the first best step in maximizing hardware stability. **I am a firm believer that prevention is always preferable to reactive response to failure, particularly when prevention is often free and reactive response is often expensive and occasionally problematic.**

I understand that regardless of how reliable hardware is, it can still fail. As management, we need to maintain the capacity to survive that failure. But survival capabilities aren't problem free. What looks easy in theory often becomes more difficult in practice. My rule has always been to do what we can to prevent the occurrence of problems first, and then ensure we survive the few problems we couldn't prevent.

Such capabilities as automated provisioning and resource pooling, in theory, enable workloads to be moved from a failing resource to backup resource, instantly and seamlessly. Some might argue that this capability therefore negates the need for greater reliability. I would argue it doesn't for the same reason that the famous “two minute move” didn't negate the fact that my data center remained down for more than an hour. The ability to move workloads instantly doesn't insulate the people on the network from the impact of hardware failures if those people are the primary sensing mechanism to determine when hardware is failing? What good is the instant ability to move resources when the delay in reporting and analyzing the problem takes at least an hour?

Think about the following: If I doubled the reliability of my hardware that has the same effect as reducing the time of reporting and analyzing the problems by half. I know how to increase the reliability of hardware and I might not know how to reduce the reporting and analyzing time. Every time I increase reliability further, it is the same as reducing the reporting and analyzing time further, except that improving reliability does not negatively impact people. And the time it takes to report, analyze and act does negatively impact people... the people experiencing the problem on the network, the people who are working to understand and resolve the problem in the data center, and the people between those people such as the people on the helpdesk. The best way to keep people from being involved in hardware failures is to prevent those failures from occurring in the first place. The best way to survive the impact of failures that do occur is to have effective survival capabilities. Reliability and survivability are not competing concepts, they are complementary concepts.