HM Health Solutions

Highly Available Infrastructure

John McDowell, Lead Enterprise Architect 2/28/17

When you pick up the phone, you want a dial-tone every time. Business areas expect the same thing from an IT Infrastructure operation. When companies use hosted services, they are expected to work consistently. With a modern IT Infrastructure management environment, there are a variety of methods to keep systems operating continuously. This ranges from Uninterruptible Power Supplies (UPS) to features more obscure, such as the recovery time of modern data routing protocols.

As IT Infrastructure specialists, we need to build a highly available environment in a layered approach. It starts from the ground up with the selection of physical space. One needs to start by choosing a data center location that is not easily impacted by local events; the building shouldn’t be sitting by a river that floods every spring. Next, layer on redundant power feeds combined with UPS’s that will dual feed into every cabinet. The cabinets themselves need to be arranged in a manner consistent with proper air-flow, keeping in mind that detailed heating/cooling plans are needed. If components start failing in August from overheating, then it is time to rethink the cooling plan.

Once the basic facilities are planned out, you move up the stack. The selection of telecommunications partners is critical to the success of any business. If the wrong partner is selected or the design/implementation of the services is done poorly, your business will be cut-off at inopportune times. Losing communications and having service down-time will undermine credibility and is a strong driver for clients to look elsewhere. Planning to go with multiple partners or staying with a single-source is an important decision. On one side you potentially avoid the vendor-wide outage, on the other side you have to successfully keep multiple services with varied implementations running optimally. A multi-vendor plan assumes diverse paths of entry; otherwise a single telephone pole event could take them both out at once.

From here, the decisions don’t get any easier. The data networking team needs to select and implement the right routing protocols. If they don’t understand your needs, the applications could flounder while the network is rebuilding itself from a minor event. The storage teams have significant decisions to make around the type of work required. Does the business have a need to do a lot of real time transactions, historical analysis, fast intake of mass-data, immediate duplication? Understanding these requirements helps the storage team to balance between high-cost chip-speed memory to long term storage as well as the placement of systems. They also need to manage replication of data to ensure your systems recover quickly following a component outage. Compute (CPU) resource services face similar concerns about what the application needs and where it is needed.

Once the essential resources (memory, CPU, network, power) are finalized, you enter the realm of application High Availability. Thanks to the world of virtualization, most operating systems are really application resources that run on abstracted hardware resource pools. Virtual operating systems can reserve resources on any hardware pool they are allowed to reach and reattach themselves to the storage pools. The teams that manage these virtual operating systems need to understand the designs of all the resources they depend on so as not to mismatch availability plans.

These virtual operating systems already offer a default level of High Availability based on the designs of all the resource areas they utilize. A virtual OS may be able to recover itself in real-time or near real-time as it transitions from a failing hardware resource pool to a healthy pool. From an application level one may also incorporate multiple systems to handle load, High Availability pairs, live-live services, geo-redundancy, cloud based recovery options, etc. Just like the rest of the stack, if the application owners make improper assumptions about the resources they depend on, there could be an outage if resource plans are not aligned.   Cross-team planning is paramount to having a fully realized highly available solution.

So how do we best aid a business function as Infrastructure specialists in terms of High Availability? We design our services from the ground up to provide the highest level of availability through building reliable (quality components) and resilient (well designed) systems. We work with our business areas to understand their specific needs and do not try to apply a one-size fits all approach to applications.

Above all, we never take any component for granted; they all count.