Autonomous Datacenter Operations

Given current advances in AI and robotics, we expect it will be possible to automate nearly all of the physical human labor associated with datacenter operations within the next 20 years. As the system administration and DevOps skill-set needed to maintain modern IT infrastructures has increased over time, skilled IT administrators and computer/software engineers have become more averse to performing low-level hardware-oriented tasks (i.e. server diagnosis and repair), as these operations are viewed as uninteresting and detrimental to career growth and job portability. The current solution has been to hire dedicated hardware technicians who specialize in these low-level tasks. However, the advent of cloud computing and disposable/low-cost consumer grade computer hardware has lowered the demand and interest in these positions, making competent hardware technicians difficult to find, and expensive to hire. From various studies, it has been shown that DOE national laboratory-operated datacenters are less costly to utilize than cloud providers for a number of large-scale/intensive scientific compute and storage applications, such as those from existing High Energy Physics and Nuclear Physics experiments. As a result, it is likely that local DOE national laboratory datacenters will continue to exist for many years to come in support of these experiments. By automating low-level physical datacenter operations, the expense of numerous full time technicians can be eliminated, and the cost of operating one's own datacenter can be greatly reduced.

We envision a future where AI-controlled robots move fully populated server/storage racks into place in the datacenter, and connect them to appropriate power and network distribution points without human intervention. Neural networks developed/trained for self-driving cars or warehouse operation robotics could likely be ported with relatively few changes to this use case. Robotic equipment will also be developed to automatically replace components in servers, storage, network and cooling/power distribution equipment when failures are detected. This would likely require the industry to adopt a more standardized/modular design for equipment, but given the benefits, it's likely this could be achieved.

Besides reducing labor costs, eliminating human technicians from physical datacenter operations can reduce the potential for injury. It can be quite dangerous to physically move dense racks of compute servers and storage. These racks can easily weigh over one thousand pounds each and typically have a high center of gravity, creating potential tipping hazards. This is particularly the case when a datacenter is equipped with removable floor tiles for ventilation/wiring. Developing robotic equipment to roll racks off of delivery trucks and into position in the datacenter, without physical human involvement, would greatly reduce the safety risks associated with this operation. Datacenters also pose electrical and noise hazards to humans: limiting the time people spend in these areas reduces their exposure to these hazards.

Finally, there are other possible efficiency gains to be made by creating an autonomous datacenter. If humans don't need to access a datacenter to perform installation/repairs, it may be possible to fill the datacenter space with an inert gas, or fully immerse all equipment in non-conductive/coolant liquid, eliminating the chance of fire, and increasing cooling efficiency. It would also be possible to operate datacenters at higher temperatures that would be uncomfortable for humans, and to more densely pack rows of equipment if people do not access them directly. Machines will also be able to address IT hardware failures more quickly than technicians, as they are available to make the repairs immediately at any time of day/night.