Virtual Machines at Hyper-Scale
Mark Russinovich, BS ECE ’89, PhD ECE ‘94
Chief Technology Officer - Microsoft Azure

Learn more about Azure virtual machine capabilities and architecture. This includes a look at how the platform offers various VM sizes and availability constraints, how VMs are provisioned to servers, and how they make use of both server-local storage for caches and ephemeral data, as well as remote block storage for storing persistent data.

Additionally, this talk will deep dive into a specific problem related to VM management at hyper-scale: how to place VMs (or containers) belonging to different Azure services onto the many servers in Azure data centers. The task is to satisfy various placement constraints and to achieve optimization goals that have great impact on utilization, availability, and other key measures. This problem is not only algorithmically interesting, but also tightly interconnected with platform design issues as well as workload characteristics. You’ll learn more about our data-driven state-of-the-art multi-priority VM placement algorithm that allow us to run Azure clusters at high utilization.

Time
Thursday, September 24th, 5:00PM

Location
Rashid Auditorium (GHC 4401)
5000 Forbes Ave,
Pittsburgh, PA 15213

microsoft.com/university