

The Inevitability of Idle Resources



Lori MacVittie, 2014-10-03

#SDN #cloud #devops Scalability's dirty little secret: if your architecture is designed to support elasticity, you must have idle resources

Proponents of pure software infrastructures often point to rapid provisioning (like right now, push a button and blam! Instant scale) as one of the reasons to "go software." Superficially this is true. If your infrastructure is all software (or virtual, if you prefer) then when you need more X you can simply grab some available resources and provision more X. No procurement process, no PO, no approvals, no waiting.

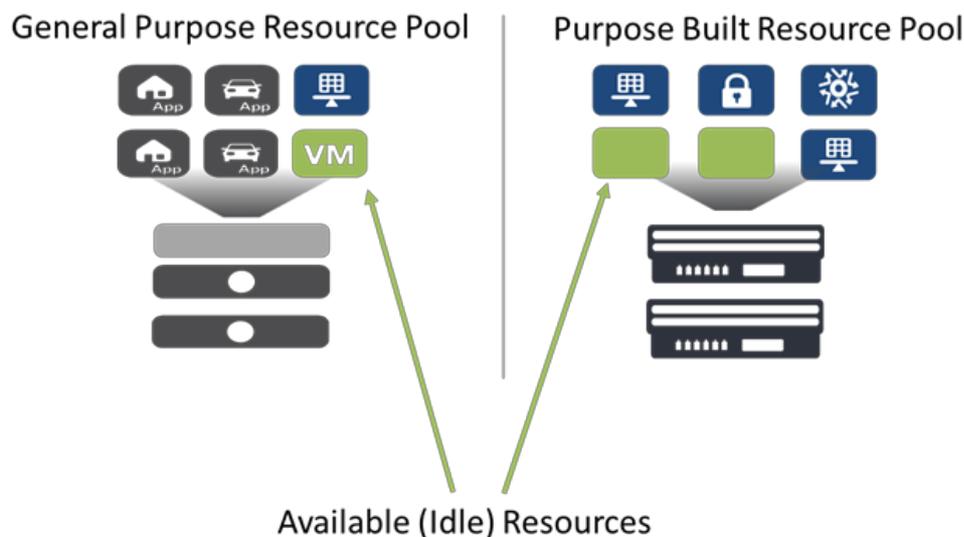
But did you notice what's ultimately required for this to happen? Available resources.

The dirty little secret of on-demand scalability is that no matter what, you need **available** resources. You need a pool of available (idle) resources from which to draw in order to provision, well, anything. Whether you need to scale server, storage, or network services, those services have to be provisioned *on* some kind of hardware.

Let's pause for a moment while I say that again: **All software requires some kind of hardware on which to execute.** For some reason this basic truth seems to get trampled under the rainbow hooves of the software-only unicorns that start dancing every time someone says SDN or NFV. I emphasize this because it's important. Software can't run without resources, and resources come from hardware.

Available resources ultimately means *idle* resources. Unless you were planning on freeing up resources on-demand by decommissioning some other service. I'm sure the consumers of that other service would have something to say about that (and what they might say probably can't be repeated in polite company).

Which ultimately means that if you think you might need to scale X, you'd best have a pool of available (idle) resources **appropriately** allocated for the service at the ready (We'll explore *that* further at a later date). Whether that comes from appropriately sized general purpose virtual machines or appropriately sized purpose built virtual instances isn't the point today. Suffice to say that you need the resources when you need them and that is the point - they need to be available, which means sitting idle when you don't need them.



Idle is idle, regardless of its source. And there must be idle resources available *somewhere* if you're going to successfully execute on a capacity-on-demand plan. The only question is whether those resources are appropriate to the task you'll put them to or not.

The moral of the story is that there is no technology or architecture today that eliminates the "waste" of idle resources without sacrificing some other data center requirement.

F5 Networks, Inc. | 401 Elliot Avenue West, Seattle, WA 98119 | 888-882-4447 | f5.com

F5 Networks, Inc.
Corporate Headquarters
info@f5.com

F5 Networks
Asia-Pacific
apacinfo@f5.com

F5 Networks Ltd.
Europe/Middle-East/Africa
emeainfo@f5.com

F5 Networks
Japan K.K.
f5j-info@f5.com

©2016 F5 Networks, Inc. All rights reserved. F5, F5 Networks, and the F5 logo are trademarks of F5 Networks, Inc. in the U.S. and in certain other countries. Other F5 trademarks are identified at f5.com. Any other products, services, or company names referenced herein may be trademarks of their respective owners with no endorsement or affiliation, express or implied, claimed by F5. CS04-00015 0113