**Examining Horizontal Scaling: Google+ Hangout Recap**

*By Alan Bush*

<https://www.youtube.com/watch?v=VF0BzRFNLKE>

Last week on the Cloud Office hours, we talked about another one of the Pillars of Cloudiness, horizontal scaling. These hangouts have been inspired by the [blog series created by Wayne Walls](http://www.rackspace.com/blog/pillars-of-cloudiness-no-3-scaling-horizontally/) and I would encourage you to check it out if you are getting started in the cloud. Some of the highlights from the hangout are listed below, and you can also watch the [full video on horizontal scaling](http://www.rackspace.com/blog/pillars-of-cloudiness-no-3-scaling-horizontally/) at the end of the post.

* We talked about the differences between vertical and horizontal scaling. Drew [defined vertical scaling](http://youtu.be/VF0BzRFNLKE?t=11m50s) as taking a server that you have and increasing the resources on it. This traditional way of scaling is often done by adding more RAM or hard drives to a particular server. [Horizontal scaling](http://youtu.be/VF0BzRFNLKE?t=12m22s) is like adding more members to your team. Rather than having one server handle the load, you could add in multiple servers to the pool to help do the work.
* There are [three different ways that people can scale horizontally](http://youtu.be/VF0BzRFNLKE?t=22m10s). One option is a manual preemptive scaling, where you proactively scale out your configuration in anticipation of a massive amount of traffic for a predefined event. For example, our customer [FLO Cycling](http://flocycling.blogspot.com/2013/03/flo-cycling-rackspace-saves-day.html) knows that they better have a beefy configuration to support a product launch or event. This style of scaling works best for businesses that know when major events will occur.
* Another way is [manual reactive scaling](http://youtu.be/VF0BzRFNLKE?t=24m11s), and often occurs after an event has occurred, or when your application or site all the sudden go viral. This occurs when you have a huge surge of traffic and you need to bring on additional resources to help serve up the data, often in reaction to a major event or breaking news. The key to manual reactive scaling is having a run book where you know the exact steps needed when it is time to pull the trigger to scale.
* And finally you have the option that most people get excited about, automatic scaling. With this option the environment scales on demand as the traffic increases. Having your app scale automatically does require some work inside your application, and you will need to hook into our APIs. Fortunately, you can code with one of our cloud SDKs so that your software can actively control the infrastructure.
* There are a couple of considerations that you need to incorporate in your application to help make it scale horizontally. First, you need to have “[dumb servers](http://youtu.be/VF0BzRFNLKE?t=30m44s),” or applications that are able to work on each node. Having the ability to deploy your application directly from GitHub, or code repository, helps achieve this. You should also look into [data consistency](http://youtu.be/VF0BzRFNLKE?t=36m50s), [session persistence](http://youtu.be/VF0BzRFNLKE?t=42m49s), [developing in an agile fashion](http://youtu.be/VF0BzRFNLKE?t=45m) and having a [message queue](http://youtu.be/VF0BzRFNLKE?t=53m44s) to help grab requests as they come in.

Those are the highlights from the horizontal scaling Office Hours Hangout. We covered a lot, and I would encourage you to watch the entire video if you have further questions. Remember, we will be back to host the next show this Thursday at 1:00 p.m. CST on the next Pillar of Cloudiness: Modular Design. We will post more information on the [Rackspace Google+ events page](https://plus.google.com/u/0/events/c7c3vl7rii0tmu6h2o5mdc797cc) for this next hangout, looking forward to seeing you then!