

The last technological underpinning you need to learn about is elasticity, the transmission in the ongoing vehicle analogy.

2.1.6 Elasticity: scaling your application as demand rises and falls

The transmission smoothly adapts the speed of a car's wheels to the engine speed as you vary the accelerator position. Similarly, *elasticity* enables an application running in a cloud to smoothly expand and contract according to demand. More precisely, elasticity is the ability to have capacity as demand increases and to release that capacity when you're done with it. Many big organizations have been close to disaster or faced it because of scalability failures in times of need.

Elasticity and celebrity deaths

In July 2009, two celebrity deaths occurred on the same day. First, *Charlie's Angels* star Farrah Fawcett died, which resulted in a minor news flurry. Then, later in the afternoon, a major web storm erupted when news of Michael Jackson's death hit the social web. Unexpectedly, Twitter had major scaling issues dealing with the sudden influx of hundreds of thousands of tweets as news of Jackson's death spread. But Twitter wasn't alone.

According to TechCrunch, "Various reports had the AOL-owned TMZ, which broke the story, being down at multiple points throughout the ordeal. As a result, Perez Hilton's hugely popular blog may have failed as people rushed there to try and confirm the news. Then the *LA Times* had a report saying Jackson was only in a coma rather than dead, so people rushed there, and that site went down. (The *LA Times* eventually confirmed his passing.)"

Numerous examples exist of a news story, a product announcement, or even the infamous Victoria's Secret Super Bowl commercial, sending people directly to a web site that then crashes. Too much traffic meets with insufficient capacity and results in catastrophe. When people are directed to a site and it then breaks down, their reaction is to not try that again. These issues severely hurt a company's business. This illustrates the criticality of being able to scale as capacity dynamically grows.

Scalability is about the cloud platform being able to handle an increased load of users working on a cloud application. *Elasticity* is the ability of the cloud platform to scale up or down based on need without disrupting the business. Without this, the economies of moving a business/application to the cloud don't make sense.

The example code snippets that follow set up an EC2 application to be load balanced and auto-scaled (that is, elastic) with a minimum of 2 instances and a maximum of 20 instances. Auto-scaling in this example is configured to scale out by 1 instance when the application's average CPU utilization exceeds a threshold of 80 percent and scale in by 1 instance when it drops below 40 percent for 10 minutes.

Call `CreateLoadBalancer` with the following parameters:

```
AvailabilityZones = us-east-1a
```

```
LoadBalancerName = MyLoadBalancer  
Listeners = lb-port=80,instance-port=8080,protocol=HTTP
```

Call `CreateLaunchConfiguration` with the following parameters:

```
ImageId = myAMI  
LaunchConfigurationName = MyLaunchConfiguration  
InstanceType = m1.small
```

Call `CreateAutoScalingGroup` with the following parameters:

```
AutoScalingGroupName = MyAutoScalingGroup  
AvailabilityZones = us-east-1a  
LaunchConfigurationName = MyLaunchConfiguration  
LoadBalancerNames = MyLoadBalancer  
MaxSize = 20  
MinSize = 2
```

Call `CreateOrUpdateScalingTrigger` with the following parameters:

```
AutoScalingGroupName = MyAutoScalingGroup  
MeasureName = CPUUtilization  
Statistic = Average  
TriggerName = MyTrigger1a  
Namespace = AWS/EC2  
Period = 60  
LowerThreshold = 40  
LowerBreachScaleIncrement = -1  
UpperThreshold = 80  
UpperBreachScaleIncrement = 1  
BreachDuration = 600
```

You learned in chapter 1 that there is more than one flavor of cloud computing. Let's combine what you learned in chapter 1 about the different types of clouds with what you now know about the six critical enabling technologies in clouds to better understand how these different flavors of clouds work, what they offer, and how they differ. The next section will help you better understand which is best for you.