Auto Scaling---

Auto scale: 2 ways

Scale up—something this increases

Scale down---something this decrees

Example:

Let me take I am upcoming festival days I think Jan 26 is republic day

Jan 25-Normal day

Jan26 –republic day

Jan27—Normal day

Let as assume I am founder of the Apple Mobile phone so I will launch the new I phone to the Indian market. So some other place I will provide that information reg (mobile configuration/ colour/ screen size, Ram, Price etc)

So I will purchase the 2 server and I keep and store that information

Then Lot of the customer are looking my new mobile configuration by using amazon/flip cart/epay etc. online delivery application.

Customar open our laptop or mobile phone and search my mobile model then go to my server and response to customer then showing that configuration of mobile details in the customer laptop/mobile phone

Web traffic = Request + Response

Then Jan 25 is normal days so I have 2 servers its enough But Jan 26 not enough 2 server because it’s a special day because lot of the cumoter are looking for any special offer of this mobile. then I will give some offer in this mobile that special day so most of the customer are accessing my server that the reason web traffic is more than high so immediately I will launch one new server. It’s not possible for the physical setup but its possible for the cloud its called the Auto scalling.

Web traffic is more than high means I server is burst

I have one simple condition my cpu % is more then 90% means automatically will launch on one server its called the scale up

My CPU % is less then 40 % means automatically will terminate on the one server is called the scale downl.