Lesson 4 Demo 17: Scale Down an Application with Less Replicas

This section will guide you to:

* Scale down an application with less replicas

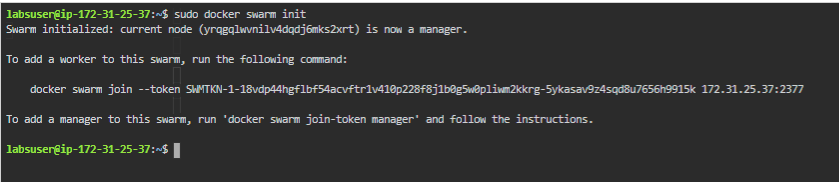
This lab has one sub-section, namely:

1. Scaling down an application with less replicas

**Note:** If you don’t have an existing Kubernetes cluster, refer to the Demo 1.1 of Lesson 1.

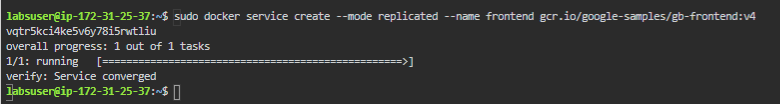
**Step 1:** Scaling down an application with less replicas

* Start the kubernetes cluster in the lab
* You can install docker with **sudo apt install docker**
* Let's start with scaling a single service. As the first step, start **sudo docker *swarm init*** as shown below to access docker services:



* Create a service using docker commands as shown below:

*sudo docker service create --mode replicated --name frontend gcr.io/google-samples/gb-frontend:v4*



* The docker scale command enables you to scale one or more replicated services either up or down to the desired number of replicas. This command cannot be applied on services which are in the global mode. The command will return immediately, but the actual scaling of the service may take some time. To stop all the replicas of a service while keeping the service active in the swarm, you can set the scale to 0
* Let's now scale down a single service, as shown below:

*sudo docker service scale frontend=10*

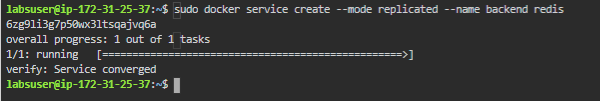
* Now let's scale down the service. As the first step, let's see the available ReplicaSets for existing services as shown below:

*sudo docker service ls*

* Now we will scale down the replica for service *frontend* to 4 as shown below:

*docker service scale frontend=4*

* Let's create another service *backend* and work on multiple services as shown below:

*sudo docker service create --mode replicated --name backend redis  
*

* The docker service scale command allows you to set the desired number of tasks for multiple services at once. The following example scales down both the backend and frontend services and lets you scale down the multiple services at once as shown below:
* You can verify the services ReplicaSet using the docker command as shown below:

*sudo docker service scale backend=3 frontend=2*

* This is how you can scale down an application with less replicas
* You can remove the services with **sudo docker service rm frontend backend**