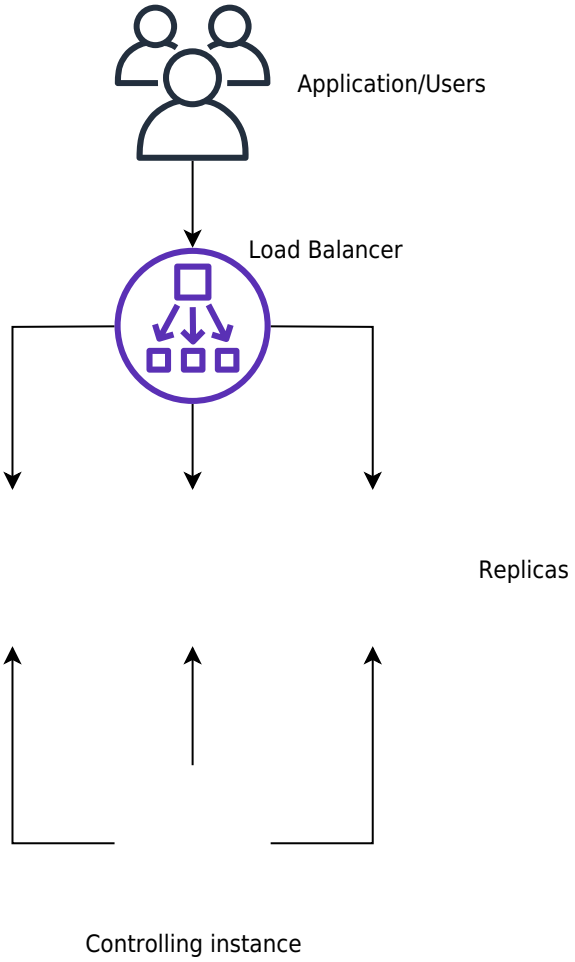
AllegroGraph Replication on Amazon’s AWS using Terraform

Introduction

In this document we describe how to setup an AllegroGraph replication cluster on AWS using the terraform program. The cluster will have one controlling instance and a set of instances controlled by an Auto Scaling Group and reached via a Load Balancer.



Creating such a system on AWS takes a long time if done manually through their web interface. We have another document that takes you through the steps. Describing the system in terraform first takes a little time but once that’s done the cluster can be started in less than five minutes.

Steps

1. Obtain an AMI with AllegroGraph and aws-repl (our support code for aws) installed.
2. Edit the terraform file we supply to suit your needs
3. Run terraform to build the cluster

Obtain an AMI with AllegroGraph and aws-repl

An AMI is an image of a virtual machine. You create an AMI by launching an ec2 instance using an AMI, altering the root disk of that instance and then telling AWS to create an AMI based on your instance. You can repeat this process until you create the AMI you need.

We have a prebuild AMI with all the code installed. It uses AllegroGraph 6.5.0 and doesn’t contain a license code so it’s limited to 5 million triples. You can use this AMI to test the load balancer or you can use this image as the starting off point for building your own image.

Alternatively you start from a fresh AMI and install everything yourself as described next.

We will create an AMI to run AllegroGraph with Replication with the following features

1. When an EC2 instance running this AMI is started it starts AllegroGraph and joins the cluster of nodes serving a particular repository.
2. When the the EC2 instance is terminated the instance sends a message to the controlling instance to ensure that the terminating instance is removed from the cluster
3. If the EC2 instance is started at a particular IP address it creates the cluster and acts as the

controlling instance of the cluster

This is a very simple setup but will serve many applications. For more complex needs you'll need to write your own tools. Contact support@franz.com to discuss support options.

The choice of AMI on which to build our AMI is not important except that our scripts assume that the initial account name of the image is ec2-user. Thus we suggest that you use one of the Amazon Linux images. If you use another kind of image you'll need to do extra work (as an example we describe below how to use a Centos AMI). Since the instance we'll build with the AMI are used only for AllegroGraph and not for other uses there's no point in running a different version of Linux that you may use in your development work.

These are the steps to build an AMI:

Start an instance using an Amazon Linux AMI with EBS support.

We can't specify the exact name of the image to start as the names change over time and depending on the region. We will usually pick one of the first images listed.

You don't need to start a large virtual machine. A t2.micro will do.

You'll need to specify a VPC and subnet. There should be a default VPC available. If not you'll have to create one.

Make sure that when you specify that subnet that you want to external IP address.

Copy an agraph distribution (tar.gz format) to the ec2 instance into the home directory of ec2-user. Also copy the file aws-repl/aws-repl.tar to the home directory of ec2-user on the instance. aws-repl.tar contains scripts to support replication setup on AWS.

Extract the agraph repo in a temporary spot and run install-

agraph in it, specifying the root of the agraph distribution.

I put it in /home/ec2-user/agraph

For example:

```
% mkdir tmp
% cd tmp
% tar xfz ../agraph-6.5.0-linuxamd64.64.tar.gz
% cd agraph-6.5.0
% ./install-agraph ~/agraph
```

Edit the file ~/agraph/lib/agraph.cfg and add the line

UseMainPortForSessions yes

This will allow sessions to be tracked through the Load Balancer.

If you have an agraph license key you should add it to the agraph.cfg file.

Unpack and install the aws-repl code:

```
% tar xf aws-repl.tar
% cd aws-repl
% sudo ./install.sh
```

You can delete aws-repl.tar but don’t delete the aws-repl directory. It will be used on startup.

Look at aws-repl/var.sh to see the parameter values. You’ll see an agraphroot parameter which should match where you installed agraph.

At this point the instance is setup.

You should go to the aws console, select this instance, and from the Action menu select “Image / Create Image”. Wait for the AMI to be built. At this time you can terminate the ec2 instance.

Using a CentOS 7 image:

If you wish to install on top of CentOS then you'll need additional steps. The initial user on CentOS is called 'centos' rather than 'ec2-user'. In order to keep things consistent we'll create the ec2-user account and use that for running agraph just as we do for the Amazon AMI.

ssh to the ec2 vm as centos and do the following to create the ec2-user account and to allow ssh access to it just like the centos account

```
[centos@ip-10-0-1-227 ~]$ sudo sh

sh-4.2# adduser ec2-user
sh-4.2# cp -rp .ssh ~ec2-user
sh-4.2# chown -R ec2-user ~ec2-user/.ssh
sh-4.2# exit

[centos@ip-10-0-1-227 ~]

$
```

At this point you can copy the agraph distribution to the ec2 vm. Scp to ec2-user@x.x.x.x rather than centos@x.x.x.x. Also copy the aws-repl.tar file.

The only change to the procedure is when you must run install.sh in the aws-repl directory.

The ec2-user account does not have the ability to sudo. So this command must be run

when logged in as the user centos;

```
centos@ip-10-0-1-227 ~]$ sudo sh
sh-4.2# cd ~ec2-user/aws-repl
sh-4.2# ./install.sh
+ cp joincluster /etc/rc.d/init.d
+ chkconfig --add joincluster
sh-4.2# exit
```

```
[centos@ip-10-0-1-227 ~]

$
```

Edit the terraform file we supply to suit your needs

Edit the file agelb.tf. This file contains directives to terraform to create the cluster with load balancer. At the top are the variables you can easily change. Other values are found inside the directives and you can change those as well.

Two variables you definitely need to change are

1. “**ag-elb-ami**” – this is the name of the AMI you created in the previous step or the AMI we supply.
2. “**ssh-key**” – this is the name of the ssh key pair you want to use in the instances created.

You may wish to change the region where you want the instances built (that value is in the provider clause at the top of the file) and if you do you’ll need to change the variable “azs”.

We suggest you try building the cluster with the minimum changes to verify it works and then customize it to your liking.

Run terraform to build the cluster

To build the cluster make sure your have an ~/.aws/config file with a default entry, such as

```
[default]
aws_access_key_id = AKIAIXXXXXXXXXXXXXXX
aws_secret_access_key = o/dyrxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
```

This is what terraform uses as credentials when it contacts AWS.

In order to use terraform the first time (or any time you change the provider clause in agelb.tf) run this command

```
% terraform init
```

Terraform will download the files appropriate for the provider you specified.

After that you can build your cluster with

```
% terraform apply
```

And watch the messages. If there are no errors terraform will wait for confirmation from you to proceed. Type yes to proceed, anything else to abort.

After terraform is finished you’ll see the address of the load balancer printed.

You can make changes the agelb.tf file and again ‘terraform apply ‘ and terraform will tell you what it needs to do to change things from how they are now to what the agelb.tf file specifies.

To delete everything terraform added type the command

```
% terraform destroy
```

And type yes when prompted.