

APACHE CASSANDRA DARBINĀŠANAI NEPIECIEŠAMĀS INFRASTRUKTŪRAS IZVEIDE AWS MĀKONĪ D.Gabaliņa, R.Jankovskis

DEVOPS PAMATI IESĀCĒJIEM

2022.gada 4.jūlijs

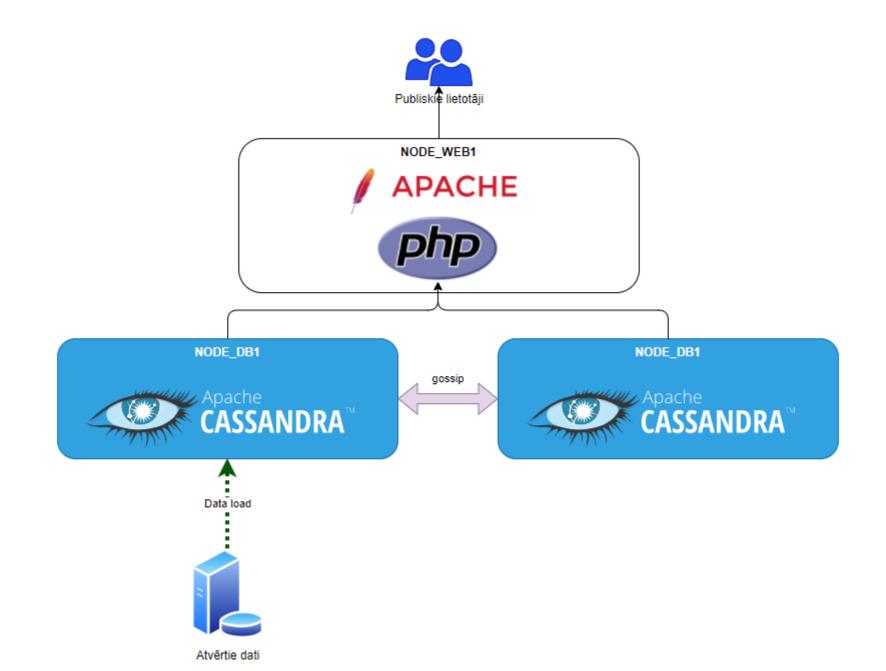
https://github.com/DitaGabalina/ devops_db_group

MĒRĶIS

Praktiski izmēģināt kursā apgūtos rīkus scenārijā, kāds tiešā veidā netika mācīts

Veidojamais risinājums

- Minimālā programma
 - Izveidot infrastruktūras risinājumu, kas uzstāda standartprogrammatūru, piestartē datu bāzi un ielādē tajā datus no kāda atvērto datu avota
- Jauki ja būtu («could have» prioritāte)
 - Apakštīkls, nodalot publisko piekļuvi un privātā tīklā datu bāzes nodes
 - Datu bāzes nodes apvienotas klāsterī un spēj izkliedēt datus ar Gossip protokolu
 - legūtie dati tiek atrādīti lietotājam, izmantojot web serveri
 - Piesaistīt AWS publiskās IP adreses, lai tās nemainās pēc serveru pārstartēšanas vai infrastruktūras pārbūves
 - Izmēģināt GitLab Runner / Github actions



IZMANTOTIE RĪKI UN RISINĀJUMI

AWS – mākoņskaitļošanas pakalpojumu piegādātājs (EC2, EIP, VPC)

GIT – pirmkoda kontrole, projekta darbi

Terraform – infrastruktūras izveide

Ansible – programmatūras instalācija uz izveidotajiem serveriem

Ubuntu – uzstādāmā OS, Linux paveids, 22.04 versija (t3.small)

OpenSSH – atslēgu ģenerēšana SSH piekļuvei





Apache Cassandra – klāsterēts datu bāzu risinājums augstai pieejamībai, CQLSH – komandrindas rīks

Apache webserver (httpd) – lietojumserveris

CURL – datņu lejupielādes rīks

Dati.ur.gov.lv – lejupielādējamo datu avots

GitHub, GitHub wiki, GitHub projects

Telegram

IZMANTOTIE RĪKI UN RISINĀJUMI

Ditas izstrādes un testa vide

- Windows 10
- Visual Studio Code teksta redaktors, darbs ar Git
- Tortoise Git Git komandas no Windows Explorer
- Windows Terminal SSH piekļuve izveidotajiem serveriem
- UVMWARE, VM ar Oracle Linux 8 (Mēģinājums izveidot terraform un aws darbināšanai Linux vidē)
- Ditas AWS konts
- GitLab
- MkDocs (tehnisko piezīmju krāšanai)

Rolanda izstrādes un testa vide

- Windows 11
- Ubuntu 22.04/18.04 LTS
- Visual Studio Code terraform un ansible izstrāde ar tam paredzētajiem paplašinājumiem
- Visual Studio Code remote ssh host, playbook izstrādei iekš AWS remote hostiem
- Windows Terminal
- Github Desktop
- Git bash / GUI
- AWS lu-vumc-devops konts

PROJEKTĀ IZMANTOTĀS LABĀS PRAKSES UN SECINĀJUMI PIRMKODA PĀRVALDĪBA

Pirmkoda pārvaldība tika veikta Git, replicējot tās uz GitLab un GitHub attālinātajām repozitorijām

Pirmkoda pārvaldībai tika veidoti zari - pēc kolēģa veiktās koda pārskates feature zari tika sapludināti ar development zaru

GitHub tika izveidotas wiki lapas, kas apraksta projekta mērķi, veidojamo infrastruktūru, zarošanas stratēģiju

PROJEKTĀ IZMANTOTĀS LABĀS PRAKSES UN SECINĀJUMI KOMANDAS DARBA VADĪBA

Darba noslodzes dēļ ne visi komandas biedri varēja pilnvērtīgi iesaistīties kopējos darbos, (ar šādu risku jārēķinās, un tas jāvada, uzstādot prasību prioritātes un mērķus, minimālo programmu)

Palīdzēja GitHub projects dashboard - kopējais darbu saraksts (paldies Pāvelam) , kuru izmantojām lai sadalītu darbus un atsekotu to statusu

Agile un Devops ir iteratīvi procesi, tie vienmēr uzlabojas, un arī tika uzlaboti savstarpējā saziņā

Vēlams savlaicīgi vienoties par naming conventions, jo pēcāk tos būs grūti labot

PROJEKTĀ IZMANTOTĀS LABĀS PRAKSES UN SECINĀJUMI INFRASTRUKTŪRA KĀ KODS (IOC)

Izmantotie rīki (Terraform un Ansible) ļauj elastīgi veidot infrastruktūru un šim nolūkam nepieciešamo informāciju glabāt pirmkoda pārvaldības sistēmās

Komandas, kas tiek izmantotas serveru instalēšanai, vēlams pirms tam pārbaudīt un tikai tad likt IOC skriptos

Svarīgi pievērst uzmanību Linux piekļuves tiesībām (ar kādu lietotāju tiek veikta instalācija, kādas tam ir piekļuves tiesības, kam jāizmanto sudo, kādas komandas ar sudo var veikt bez atkārtotas paroles ievadīšanas, failu un mapju *permissions*)

SSH atslēgas (nesaputroties, tās var būt vairākas, netīšām neiekļaut pirmkodā)

CASSANDRA | DATU LEJUPIELADE



CASSANDRA | DATU MODELIS

```
76 lines (70 sloc) 1.89 KB
                                                                                                                                                        Blame
                                                                                                                                                  Raw
       -- Create a keyspace - replication factor should not exceed the number of nodes in your cluster
      CREATE KEYSPACE IF NOT EXISTS ur WITH REPLICATION = { 'class' : 'SimpleStrategy', 'replication_factor' : '2' };
       -- data_load\register_name_history.csv
       CREATE TABLE IF NOT EXISTS ur.register_name_history (
       regcode text PRIMARY KEY,
       name text,
       date to timestamp
  9
       );
  10
       -- data_load\ppi_delegated_entities.csv
       CREATE TABLE IF NOT EXISTS ur.ppi_delegated_entities (
       registrationNumber text PRIMARY KEY,
       name text,
       delegatedEntityName text,
       delegatedEntityRegistrationNumber text,
       delegatedEntityRegisteredOn timestamp,
       delegatedEntityRemovedOn timestamp
 19
       );
  20
```

CASSANDRA | DATU IELĀDE

```
A lines (3 sloc) 212 Bytes

Raw Blame 

TRUNCATE ur.register_name_history;

-- Insert some data

COPY ur.register_name_history (regcode,date_to,name) FROM './register_name_history.csv' WITH HEADER = TRUE AND DELIMITER=';' and DATETIMEFORMAT='%Y-%m-%d';
```

CASSANDRA | REZULTĀTA PĀRBAUDE

select * from ur.register name history;

```
40002086696
             2003-09-17 00:00:00.000000+0000
                                                                                                                                           SIA "TRIALS"""
42103048376
             2010-07-14 00:00:00.000000+0000
                                                                                                                                      SIA "L Lingerie"""
              2017-01-12 00:00:00.000000+0000
                                                                                                      Sabiedrība ar ierobežotu atbildību "MP Skaistums"""
                                                                                     Ogres rajona A.Grašinas individuālais uzņēmums "BIRZGALES APTIEKA"""
             2004-02-16 00:00:00.000000+0000
40002079092
                                                                          Sabiedrība ar ierobežotu atbildību "Andas Kaires ārsta prakse zobārstniecībā"""
43603041468
             2021-08-04 00:00:00.000000+0000
40008018867
             2004-12-13 00:00:00.000000+0000
                                                                                                                                 Aizgādnu fonds "ASSOR"""
40103669322
             2015-10-22 00:00:00.000000+0000
                                                                                                                                        SIA "4 b.e.e."""
40003051329
             2004-03-03 00:00:00.000000+0000
                                                                                            Sabiedrība ar ierobežotu atbildību-kopuznēmums firma "EKJU"""
40002036661
                                                                                                Valērija Černišova individuālais uzņēmums firma "BESTS"""
             2004-12-08 00:00:00.000000+0000
44101034647
                                                                                            Mārsnēnu pagasta Ulda Pilvera zemnieka saimniecība "UPMALI"""
             2006-07-03 00:00:00.000000+0000
                                                                                    Ražošanas komercfirma sabiedrība ar ierobežotu atbildību "MD VĀRME"""
40003114200
             2003-04-02 00:00:00.000000+0000
                                                                                                            Sabiedrība ar ierobežotu atbildību "KRAUJA"""
40003087504
             2004-12-28 00:00:00.000000+0000
                                                                                                        Sabiedrība ar ierobežotu atbildību "AGROBIZNES"""
40003239883
             1995-01-09 00:00:00.000000+0000
40003898810
             2008-05-06 00:00:00.000000+0000
                                                                                                  Sabiedrība ar ierobežotu atbildību "MADONAS PROJEKTS""
                                                                                                            Sabiedrība ar ierobežotu atbildību "Š UN C"""
40003397454
             2001-10-04 00:00:00.000000+0000
46601010028
             2005-10-05 00:00:00.000000+0000
                                                                                        Limbažu rajona Ainažu pilsētas zemnieku saimniecība "AVOTKALNI"""
                                                                                                            Sabiedrība ar ierobežotu atbildību "KD Jet"""
40003791008
             2012-01-13 00:00:00.000000+0000
40008044018
             2008-10-10 00:00:00.000000+0000
                                                                                                    Latvijas Aviācijas Darbinieku Arodbiedrību Federācija
                                                                                                                        INŽENIERTEHNISKĀ FIRMA "ALKORS"""
50003263641
             1995-10-25 00:00:00.000000+0000
                                                                                                                          SIA kopuznēmums "UNITED OILS"""
40003214019
             2004-08-09 00:00:00.000000+0000
             2009-01-12 00:00:00.000000+00
                                                                                                                         V.Zikovas individuālais uznēmums
```

Terraform struktūra

Terraform direktorija satur:

- root direktorijā terraform failus, kur definēti un sasaistīti visi terraform moduļi, zem /modules direktorijas
- modules direktorija ir visi nepieciešamie terraform moduļi:
 - EC2 direktorija satur 2 moduļus apache un cassandra
 - VPC modulis satur visu nepieciešamo tīkla izveidei iekš aws *apache* un cassandra moduļiem.
- assets direktorijā tiek glabāts viss pārējais, kas nepieciešams terraform, piemēram template, ssh atslēgas un skripti



Terraform palaišana

Lai palaistu Terraform nepieciešams atvērt root direktorijā esošo terraform.tfvars un nomainīt

- aws-private-key-location uz ceļu kur atrodās jūsu aws privātā atslēgu un
- aws-access-key-name uz jūsu aws ģenerētā keypair nosaukumu
- Papildus arī nepieciešams ģenerēt ssh atslēgas ansible master (privāto) un ansible host (publisko), kā arī ssh ielādēšanas skriptu
- Esot projekta root direktorijā to var izdarīt ar powershell komandām:

```
mkdir .\terraform\assets\secrets | cd
ssh-keygen -t rsa -N -q -f secret-key.pem -N '""'
ren .\secret-key.pem.pub public-key.pub
echo "#!/bin/bash" >> install-ssh_key.sh ; echo "echo
'$(cat public-key.pub)' >>
/home/ubuntu/.ssh/authorized_keys" >> install-
ssh_key.sh
```

```
terraform
terraform
assets
inventory
scripts
secrets
install-ssh_key.sh
public-key.pub
secret-key.pem
```

(ansible hosts ģenerēšana)

```
resource "local_file" "hosts_cfg" {
   content = templatefile("./assets/templates/hosts.tpl",
   {
       cassandra_node1 = aws_instance.cassandra_db_node_1.private_ip
       cassandra_node2 = aws_instance.cassandra_db_node_2.private_ip
   }
   )
   filename = "./assets/inventory/hosts.cfg"
}
```

- lekš Terraform moduļa definējam lokālā faila resursu kurā tiek norādīts lai izvada privātās ip adreses
- Privātās ip adreses tiek padotas uz hosts.tpl
 un pēc tā satura tiek izveidots fails hosts.cfg

^{*} lespējams arī ģenerēt *template*, lai tiek aizsūtītas visas moduļu instanču adreses izmantojot "for loop", automatizētai *inventory* izveidei

(ansible master konfigurēšana)

rovisioner "remote-exec" {

"sudo apt update -y", "sudo apt install ansible -y",

"mkdir ~/ansible-codes".

"sudo chmod 400 ~/.ssh/id rsa",

= "ssh"

private key = file("\${var.aws private key}") = self.public_dns

"sudo apt-add-repository ppa:ansible/ansible -y",

"sudo mv /etc/ansible/hosts /etc/ansible/hosts.bak",

"sudo mv /tmp/hosts /etc/ansible/hosts",

inline = [

type

Provisioner File caur ssh nosūta privāto atslēgu uz ansible master, node pārvaldīšanai un ģenerēto ansible inventory - hosts.cfg

- Uz Ansible Master tiek uzstādītas pareizās tiesības privātajai ssh atslēgai, uzstādīts ansible, pareizi instalēts sagavaotais inventory fails.
- Tad tiek klonēts repozitorijs un startēti sagatavotie playbooki

```
user
                                                                                            = "ubuntu"
                                                                              private key = file("${var.aws private key}")
                                                                              host
                                                                                            = self.public dns
                                                                         rovisioner "file" {
                                                                        source = "./assets/inventory/hosts.cfg"
                                                                        destination = "/tmp/hosts"
                                                                          connection {
                                                                                            = "ssh"
                                                                              type
                                                                                            = "ubuntu"
                                                                              user
                                                                              private key = file("${var.aws private key}")
                                                                                            = self.public_dns
git clone -b development https://github.com/DitaGabalina/devops_db_group.git ~/ansible-codes/devops_db_group",
"ansible-playbook ~/ansible-codes/devops db group/ansible/cassandra node-setup.yaml --ssh-common-args='-o StrictHostKeyChecking=accept-new'",
ansible-playbook ~/ansible-codes/devops_db_group/ansible/cassandra_main_node-data_load.yaml",
"ansible-playbook ~/ansible-codes/devops db group/ansible/cassandra-cluster-creation.yaml",
"ansible-playbook ~/ansible-codes/devops db group/ansible/apache server-setup.yaml --ssh-common-args='-o StrictHostKeyChecking=accept-new'"
```

provisioner "file" {

connection {

type

source = "./assets/secrets/secret-key.pem" destination = "/home/ubuntu/.ssh/id rsa"

= "ssh"

(ansible master konfigurēšana)

rovisioner "remote-exec" {

"sudo apt update -y", "sudo apt install ansible -y",

"mkdir ~/ansible-codes".

"sudo chmod 400 ~/.ssh/id rsa",

= "ssh"

private key = file("\${var.aws private key}") = self.public_dns

"sudo apt-add-repository ppa:ansible/ansible -y",

"sudo mv /etc/ansible/hosts /etc/ansible/hosts.bak",

"sudo mv /tmp/hosts /etc/ansible/hosts",

inline = [

type

Provisioner File caur ssh nosūta privāto atslēgu uz ansible master, node pārvaldīšanai un ģenerēto ansible inventory - hosts.cfg

- Uz Ansible Master tiek uzstādītas pareizās tiesības privātajai ssh atslēgai, uzstādīts ansible, pareizi instalēts sagavaotais inventory fails.
- Tad tiek klonēts repozitorijs un startēti sagatavotie playbooki

```
user
                                                                                            = "ubuntu"
                                                                              private key = file("${var.aws private key}")
                                                                              host
                                                                                            = self.public dns
                                                                         rovisioner "file" {
                                                                        source = "./assets/inventory/hosts.cfg"
                                                                        destination = "/tmp/hosts"
                                                                          connection {
                                                                                            = "ssh"
                                                                              type
                                                                                            = "ubuntu"
                                                                              user
                                                                              private key = file("${var.aws private key}")
                                                                                            = self.public_dns
git clone -b development https://github.com/DitaGabalina/devops_db_group.git ~/ansible-codes/devops_db_group",
"ansible-playbook ~/ansible-codes/devops db group/ansible/cassandra node-setup.yaml --ssh-common-args='-o StrictHostKeyChecking=accept-new'",
ansible-playbook ~/ansible-codes/devops_db_group/ansible/cassandra_main_node-data_load.yaml",
"ansible-playbook ~/ansible-codes/devops db group/ansible/cassandra-cluster-creation.yaml",
"ansible-playbook ~/ansible-codes/devops db group/ansible/apache server-setup.yaml --ssh-common-args='-o StrictHostKeyChecking=accept-new'"
```

provisioner "file" {

connection {

type

source = "./assets/secrets/secret-key.pem" destination = "/home/ubuntu/.ssh/id rsa"

= "ssh"

(ansible host konfigurēšana)

 Izmantojot user-data komandu tiek palaists sagatavots skripts kas uzkopē publisko atslēgu pie authorized keys.

^{*} Šajā scenārijā remote-exec un file provideris nav pieļaujams, kā ansible master, jo instances atrodās privātā tīklā un tām pa tiešo nevar pieslēgties

(ansible master Apache2 un php uzstādīšana)

- Norādīts hosts localhost, lai izpildītos uz ansible master
- Uzstāda apache2 un php uz master node

```
hosts: localhost
become: true
tasks:
  - name: Update and upgrade apt packages
    apt:
      upgrade: yes
      update cache: yes
  - name: Install Apache
    package:
      name: apache2
      state: latest
  - name: Start theApache service
    service:
       name: apache2
       state: started
  - name: Install Php
    package:
      name: php
      state: present
```

(ansible node Apache Cassandra uzstādīšana ar dependencies un java test)

- uzstāda nepieciešamās apt pakas
- instalē php atslēgas priekš Cassandra repozitorija
- uzstāda Cassandru
- iestata env mainīgo JAVA_HOME un to pārbauda

```
- hosts:
   - cassandra nodes
 become: true
   - name: Update and upgrade apt packages
       upgrade: yes
       update_cache: yes
   - name: Install MC
       state: latest
   - name: Install openjdk-11-jre-headless
        name: openjdk-11-jre-headless
        state: present
   - name: Install apt-transport-https
       name: apt-transport-https
       state: present
   - name: Add Key for cassandra repo
       url: https://www.apache.org/dist/cassandra/KEYS
       state: present
```

```
name: Add cassandra debian repository
apt_repository: repo='deb http://www.apache.org/dist/cassandra/debian 40x main'
name: Install cassandra
   name: cassandra
  state: present
name: make sure cassandra is started
 service: name=cassandra state=restarted
- name: Set JAVA HOME
   dest: /etc/environment
  state: present
   regexp: '^JAVA_HOME'
   line: 'JAVA HOME=/usr/lib/jvm/java-11-openjdk-amd64'
name: JAVA HOME ielasisana
shell: printf $JAVA HOME
register: output

    name: JAVA_HOME Parbaude

  msg: "JAVA_HOME Veiksmigi registrets"
 when: output.stdout is match ("/usr/lib/jvm/java-11-openjdk-amd64")
```

(ansible datu lejupielāde un imports iekš main node)

- Izveido directory datu saglabāšanai
- Lejupielādē Github repo uz cassandra nodes
- No lejupielādētajiem datiem palaiž sagatavotos skriptus datu lejupielādei un importēšanai

```
- cassandra main node
- name: Create External data dir
 file:
    path: ~/external_data
   state: directory

    name: Download devops db group from the GitRepo

   repo: 'https://github.com/DitaGabalina/devops_db_group.git'
   dest: "~/devops db group/"
    version: development
    force: yes
- name: Download UR Data
 shell: chmod +x ~/devops db group/data load/download ur data.sh && ~/devops db group/data load/download ur data.sh
- name: Create Data Model
 shell: cqlsh -f ~/devops db group/data load/create cassandra data model.cql
  args:
    chdir: ~/external_data
- name: Copy to cassandra
 shell: cqlsh -f ~/devops db group/data load/copy to cassandra.cql
  ignore errors: yes
  args:
   chdir: ~/external_data
```

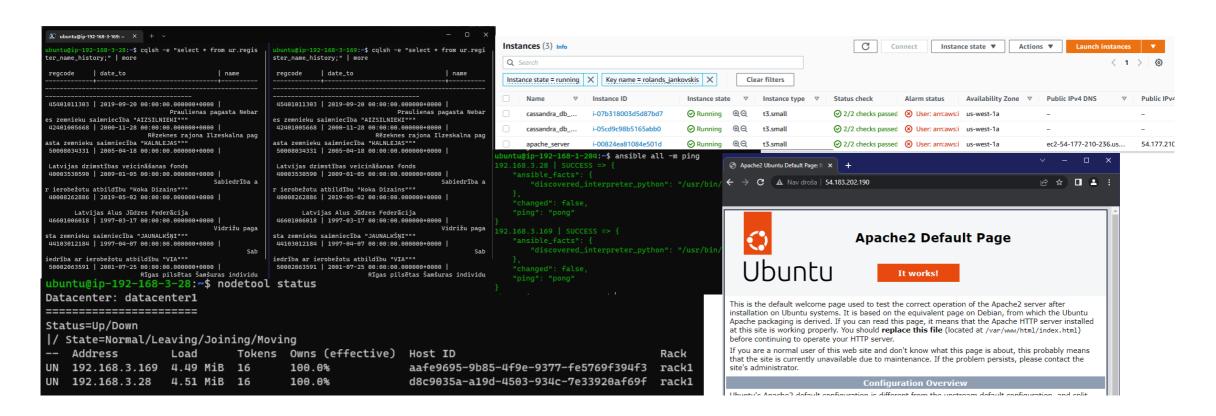
(ansible Cassandra node cluster izveide)

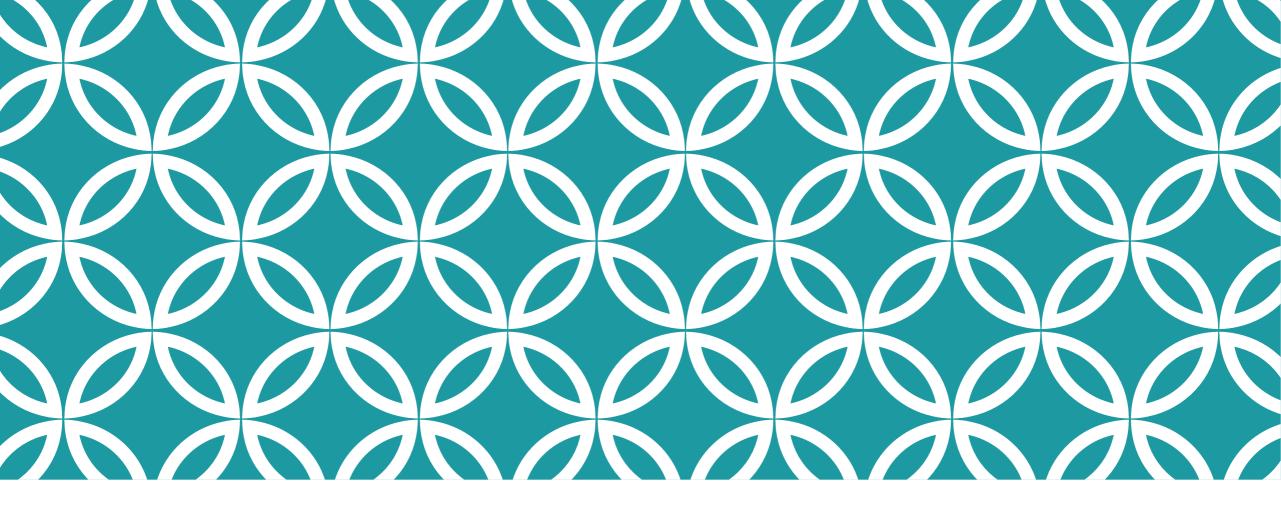
- lelasa galvenās nodes ip adresi uz mainīgo seeds
- Aizstāj konfigurācijas failu cassandra.yaml uz ansible hostiem pielietojot sagatavoto šablonu un mainīgos
- Restartē cassandru, pagaida kamēr tā iestartējas un tad palaiž repair kommandu lai dati tiktu replicēti.

```
# Setting listen_address to 0.0.0.0 is always wrong.
#
listen_address: {{ inventory_hostname }}
```

Rezultāts pēc Terraform apply

- Pilns Terraform apply aizņem aptuveni 11 min un 40 sek, no kurām 5 min ir datu lejupielāde no UR
- Terraform automatizēti uzstāda ansible struktūru un apmaina ssh keys starp hostiem
- Terraform uzstāda VPC tīklu ar 2 privātiem un 2 publiskiem subnetiem kas ir pārdalīti pa 2 dažādiem avaliablity zones
- Terraform/ansible rezultātā tiek uzstādītas 3 instances 2 Cassandra nodes privātā tīkla un 1 Apache node publiskā tīklā
- Ansible uzstāda Apache server, Cassadra un tās dependencies un ielādē datus no atvērtiem datiem
- Ansible apvieno abas Cassandra nodes vienā klasteri ar datu replikāciju





PALDIES PAR UZMANĪBU!