UNIVERSIDAD EAFIT

Tópicos Especiales en Telemática

Laboratorio 3-0

Juan Manuel Gómez - Ingeniería de Sistemas

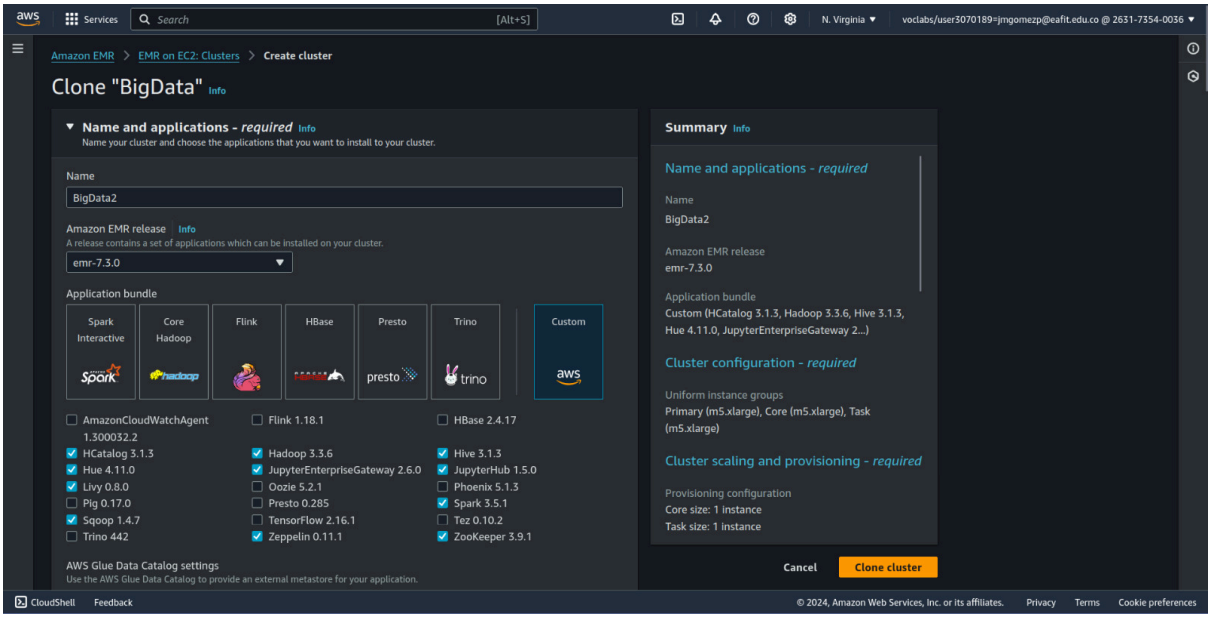
Profesor: Edwin Nelson Montoya

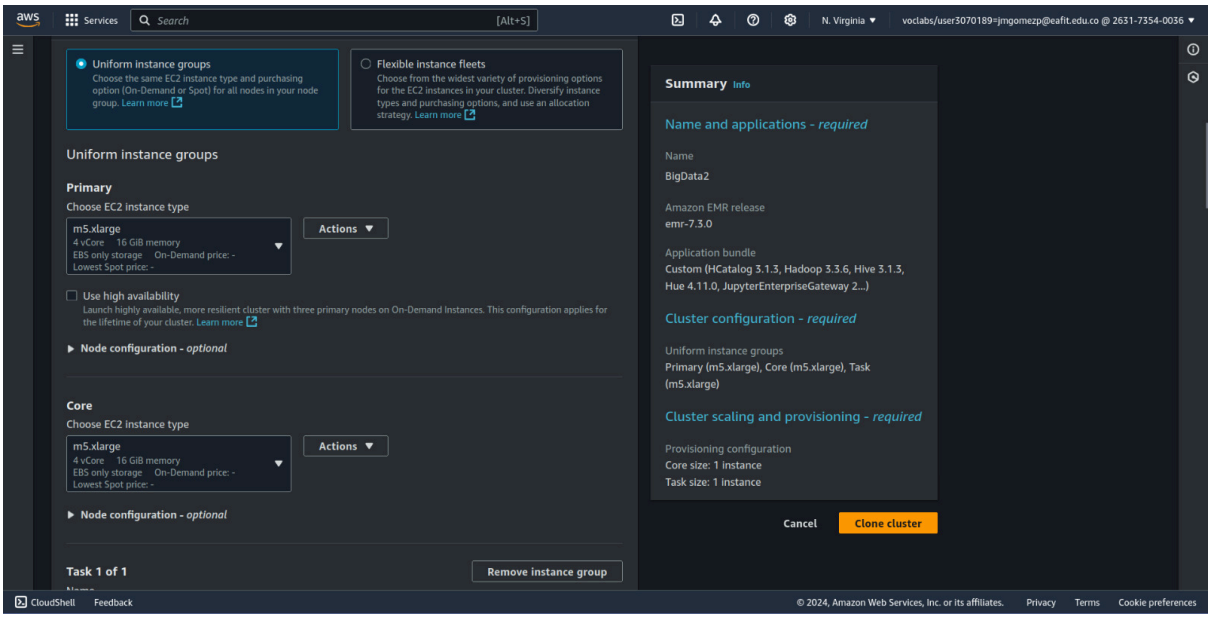
Medellín, 7 de noviembre de 2024

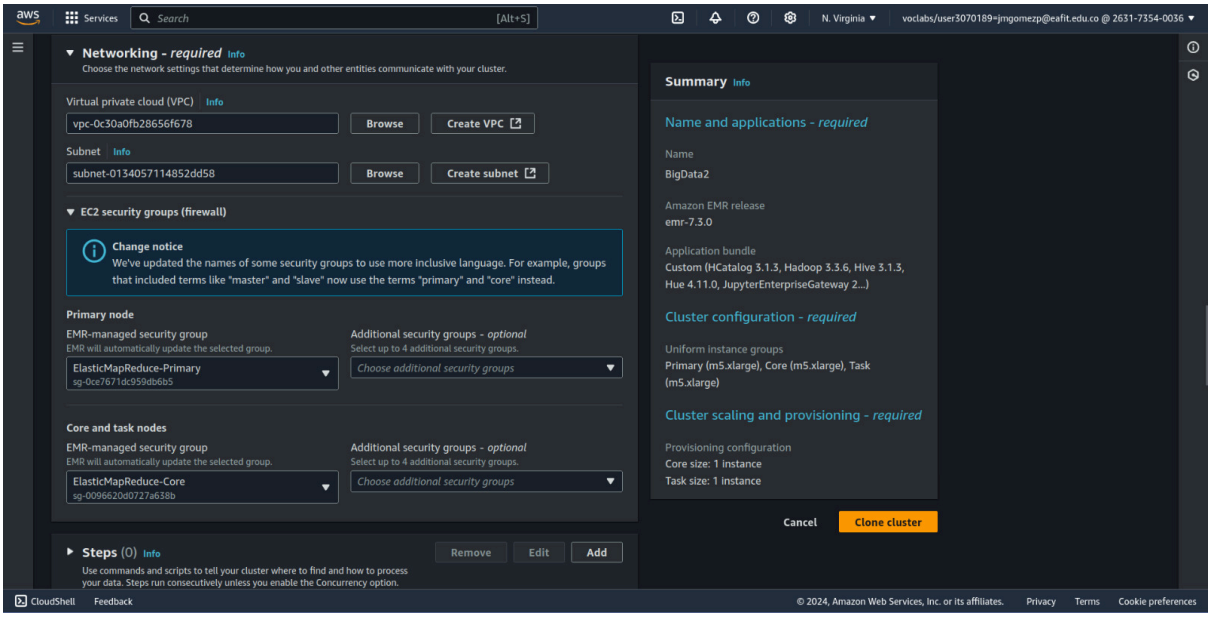
En este laboratorio se instanció un cluster de EMR con 3 nodos (cada uno m5.xlarge), donde uno hace de master y los otros dos de slaves. En este cluster se instalaron las aplicaciones necesarias para poder desarrollar los siguientes laboratorios (spark, JupyterHub, Hue, etc.).

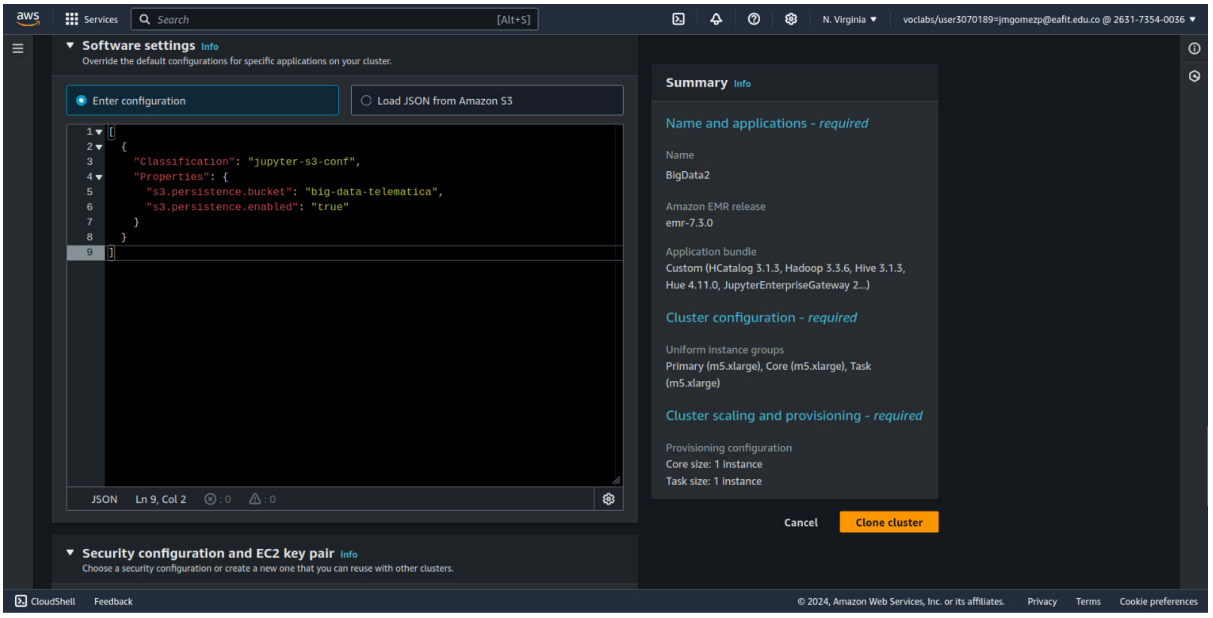
A partir de este primer cluster se crearon los otros usando la opción de clonar para ahorrar tiempo al momento de seguir creando clusters (ya que estos son temporales).

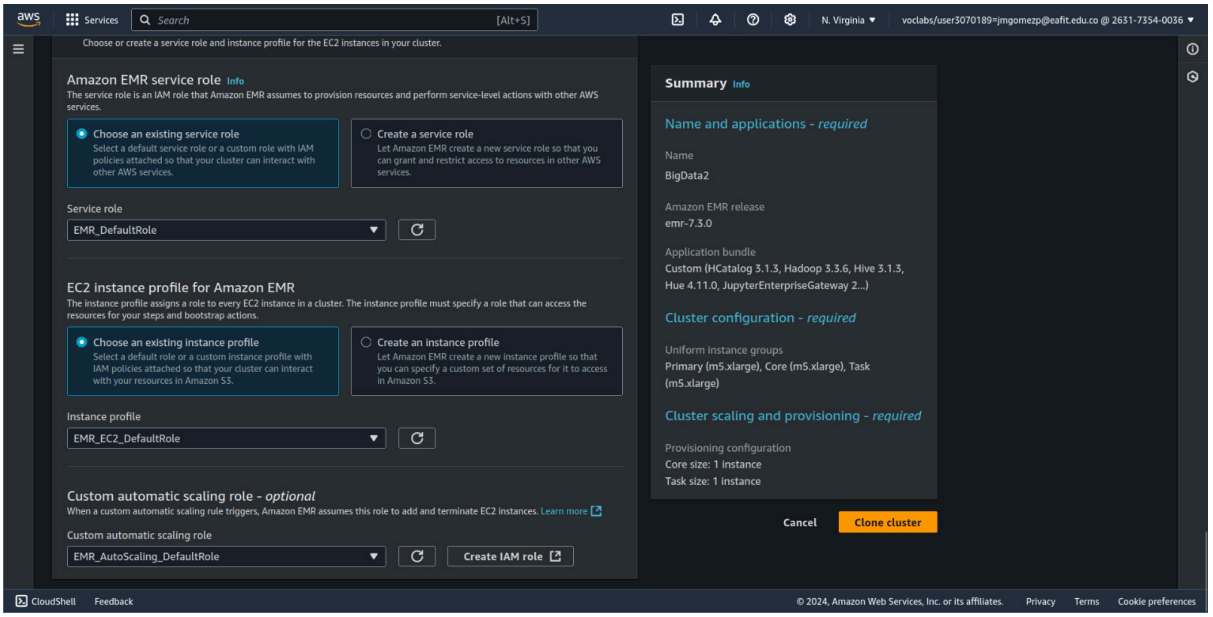
A continuación se muestra el proceso de creación del cluster paso a paso:






Services
Search
[Alt+S]
N. Virginia
voclabs/user3070189=jmgomezp@eafit.edu.co @ 2631-7354-0036
Networking - required Info
Choose the network settings that determine how you and other entities communicate with your cluster.
Virtual private cloud (VPC) Info
vpc-0c30a0fb28656f678
Browse
Create VPC
Subnet Info
subnet-0134057114852dd58
Browse
Create subnet
EC2 security groups (firewall)
Change notice
We've updated the names of some security groups to use more inclusive language. For example, groups that included terms like "master" and "slave" now use the terms "primary" and "core" instead.
Primary node
EMR-managed security group
EMR will automatically update the selected group.
ElasticMapReduce-Primary
sg-0ce7671dc959db6b5
Additional security groups - optional
Select up to 4 additional security groups.
Choose additional security groups
Core and task nodes
EMR-managed security group
EMR will automatically update the selected group.
ElasticMapReduce-Core
sg-0096620d0727a638b
Additional security groups - optional
Select up to 4 additional security groups.
Choose additional security groups
Steps (0) Info
Use commands and scripts to tell your cluster where to find and how to process your data. Steps run consecutively unless you enable the Concurrency option.
Remove
Edit
Add
Summary Info
Name and applications - required
Name
BigData2
Amazon EMR release
emr-7.3.0
Application bundle
Custom (HCatalog 3.1.3, Hadoop 3.3.6, Hive 3.1.3, Hue 4.11.0, JupyterEnterpriseGateway 2...)
Cluster configuration - required
Uniform instance groups
Primary (m5.xlarge), Core (m5.xlarge), Task (m5.xlarge)
Cluster scaling and provisioning - required
Provisioning configuration
Core size: 1 instance
Task size: 1 instance
Cancel
Clone cluster
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences


Services
Search
[Alt+S]
N. Virginia
voclabs/user3070189=jmgomezp@eafit.edu.co @ 2631-7354-0036
Software settings Info
Override the default configurations for specific applications on your cluster.
Enter configuration
Load JSON from Amazon S3
[
{
"Classification": "jupyter-s3-conf",
"Properties": {
"s3.persistence.bucket": "big-data-telematica",
"s3.persistence.enabled": "true"
}
}
]
JSON
Ln 9, Col 2
Summary Info
Name and applications - required
Name
BigData2
Amazon EMR release
emr-7.3.0
Application bundle
Custom (HCatalog 3.1.3, Hadoop 3.3.6, Hive 3.1.3, Hue 4.11.0, JupyterEnterpriseGateway 2...)
Cluster configuration - required
Uniform instance groups
Primary (m5.xlarge), Core (m5.xlarge), Task (m5.xlarge)
Cluster scaling and provisioning - required
Provisioning configuration
Core size: 1 instance
Task size: 1 instance
Cancel
Clone cluster
Security configuration and EC2 key pair Info
Choose a security configuration or create a new one that you can reuse with other clusters.
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences


Services
Search
[Alt+S]
N. Virginia
voclabs/user3070189=jmgomezp@eafit.edu.co @ 2631-7354-0036
Choose or create a service role and instance profile for the EC2 instances in your cluster.
Amazon EMR service role Info
The service role is an IAM role that Amazon EMR assumes to provision resources and perform service-level actions with other AWS services.
Choose an existing service role
Select a default service role or a custom role with IAM policies attached so that your cluster can interact with other AWS services.
Create a service role
Let Amazon EMR create a new service role so that you can grant and restrict access to resources in other AWS services.
Service role
EMR_DefaultRole
EC2 instance profile for Amazon EMR
The instance profile assigns a role to every EC2 instance in a cluster. The instance profile must specify a role that can access the resources for your steps and bootstrap actions.
Choose an existing instance profile
Select a default role or a custom instance profile with IAM policies attached so that your cluster can interact with your resources in Amazon S3.
Create an instance profile
Let Amazon EMR create a new instance profile so that you can specify a custom set of resources for it to access in Amazon S3.
Instance profile
EMR_EC2_DefaultRole
Custom automatic scaling role - optional
When a custom automatic scaling rule triggers, Amazon EMR assumes this role to add and terminate EC2 instances. Learn more
Custom automatic scaling role
EMR_AutoScaling_DefaultRole
Create IAM role
Summary Info
Name and applications - required
Name
BigData2
Amazon EMR release
emr-7.3.0
Application bundle
Custom (HCatalog 3.1.3, Hadoop 3.3.6, Hive 3.1.3, Hue 4.11.0, JupyterEnterpriseGateway 2...)
Cluster configuration - required
Uniform instance groups
Primary (m5.xlarge), Core (m5.xlarge), Task (m5.xlarge)
Cluster scaling and provisioning - required
Provisioning configuration
Core size: 1 instance
Task size: 1 instance
Cancel
Clone cluster
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences