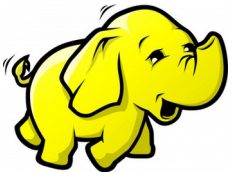


Cloudera Administrator Apache Hadoop

Parte 04-1 Segurança



Marco Reis
<http://marcoreis.net>

Agenda

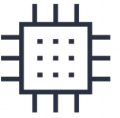


- Segurança no CDH
 - Kerberos
 - Autenticação
 - Autorização
 - Kerberos no CDH
 - Operação com usuários autenticados

Kerberos



- Kerberos é um protocolo de rede para autenticação baseado em criptografia de chave privada
- Permite que os usuários acessem recursos na rede de forma simplificada
 - Single sign-on (SSO): o usuário se autentica apenas uma vez
- O MIT desenvolveu uma versão gratuita que pode ser usada no CDH
- O Kerberos funciona bem em ambientes distribuídos, como em um cluster, protegendo a rede de usuários maliciosos
- As senhas não trafegam pela rede, evitando o ataque de sniffers
- As informações de autenticação ficam centralizadas no servidor



Componentes

- Key Distribution Center (KDC): uma instalação Kerberos (KDC) é composta dos seguintes componentes:
 - Authentication Service (AS): realiza as operações de autenticação com login e senha do usuário e retorna um Ticket Granting Ticket (TGT) que será armazenado localmente na máquina do usuário
 - Database: base de dados das chaves encriptadas dos usuários
 - Ticket Granting Server (TGS): valida o ticket do usuário no serviço solicitado

Conceitos



- Realm: conjunto de servidores que compartilham uma base de dados (LAB.BIGDATA.COM)
- Principal: usuário ou serviço autenticado no Kerberos
- Ticket Granting Ticket (TGT): identificação encriptada válida por um período de tempo
- Keytab: arquivo que armazena localmente as chaves criptografadas do Kerberos



Instalação do KDC

- O servidor do KDC pode ser uma máquina fora do cluster
 - `$ apt update`
 - `$ apt upgrade -y`
 - `$ apt install krb5-kdc krb5-admin-server krb5-config -y`
- Quando solicitado, preencher os parâmetros com os valores a seguir
 - `# Default realm: LAB.BIGDATA.COM`
 - `# Kerberos servers: localhost`
 - `# Admin server: localhost`
- Para criar um novo Realm
 - `$ krb5_newrealm`
 - `# Master password: kerberos`
- Alterar o arquivo `/etc/krb5kdc/kadm5.acl` para habilitar o usuário administrador
- Adicione as linhas:
 - `*/admin *`
 - `*/admin@LAB.BIGDATA.COM *`
- Alterar o arquivo `/etc/krb5.conf` com os parâmetros do CDH de autenticação e renovação de tickets
- Adicionar as linhas na seção principal
 - `max_life = 1d`
 - `max_renewable_life = 7d`
 - `kdc_tcp_ports = 88`
- Reiniciar os serviços para validar as configurações
 - `$ service krb5-kdc restart`
 - `$ service krb5-admin-server restart`

Adicionando usuários



- Para adicionar um usuário
 - `kadmin.local -q "addprinc -pw cloudera-scm cloudera-scm/admin"`
 - `kadmin.local -q "addprinc -pw dataengineer dataengineer"`
 - `kadmin.local -q "addprinc -pw datascientist datascientist"`
- O usuário do Kerberos deve ter um usuário correspondente no Linux em cada máquina

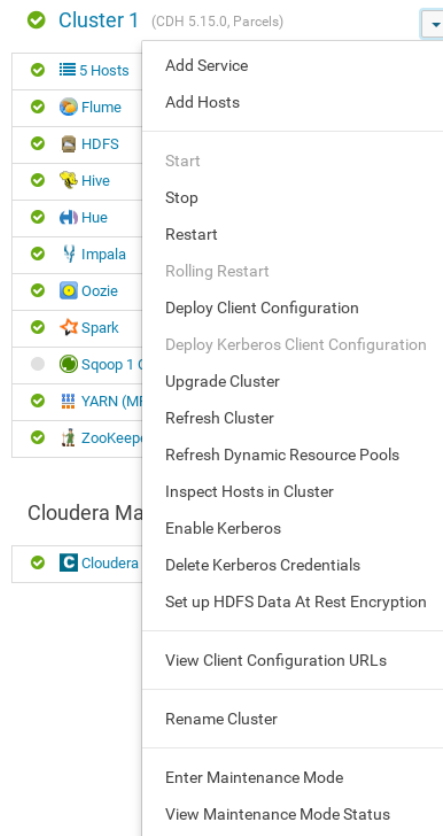
Instalando os clientes



- A autenticação no cluster será feita a partir do Kerberos
- Todas as máquinas com acesso ao cluster precisam do cliente Kerberos
- A instalação é feita pelo pacote:
 - `$ apt install krb5-user -y`
- Preencher os dados com os parâmetros do Kerberos
- Todas as máquinas devem ter o JCE para a criptografia selecionada (aes-256)
 - Os arquivos são `local_policy.jar` e `US_export_policy.jar`
 - `$ scp *.jar root@nome-da-maquina:/usr/lib/jvm/java-7-oracle-cloudera/jre/lib/security/`
- Para autenticar com o usuário no Kerberos
 - `$ kinit dataengineer`
 - `$ klist -e`
- Para destruir o ticket
 - `$ kdestroy`

Habilitar Kerberos no CDH

- O Kerberos é o protocolo padrão para segurança no CDH
- Para habilitar o Kerberos no CDH use a opção Enable Kerberos
- A seguir veremos as configurações necessárias



Início do processo

- Verifique se todas as configurações foram atendidas

Enable Kerberos for Cluster 1

Getting Started

This wizard walks you through the steps to configure Cloudera Manager and CDH to use Kerberos for authentication. All services in the cluster, as well as the Cloudera Management Service, are restarted as part of the wizard. Before proceeding with the wizard, read the [documentation](#) about enabling Kerberos.

Before using the wizard, ensure that you have performed the following steps:

Set up a working KDC. Cloudera Manager supports MIT KDC and Active Directory.

☒ Yes, I have set up a working KDC.

The KDC should be configured to have non-zero ticket lifetime and renewal lifetime. CDH will not work properly if tickets are not renewable.

☒ Yes, I have checked that the KDC allows renewable tickets.

OpenLdap client libraries should be installed on the Cloudera Manager Server host if you want to use Active Directory. Also, Kerberos client libraries should be installed on ALL hosts.

☒ Yes, I have installed the client libraries.

Cloudera Manager needs an account that has permissions to create other accounts in the KDC.

☒ Yes, I have created a proper account for Cloudera Manager.

KDC

- Nossa instalação usa o MIT KDC para autenticação
- A encriptação usa aes256
- O endereço do KDC Server/Admin deve estar configurados em todos os hosts no /etc/hosts

Enable Kerberos for Cluster 1

Setup KDC

Specify information about the KDC. The properties below are used by Cloudera Manager to generate principals for CDH daemons running on the cluster.

KDC Type

☐ Active Directory

☒ MIT KDC



Kerberos Encryption Types

aes256-cts-hmac-sha1-96



Kerberos Security Realm
default_realm

LAB.BIGDATA.COM



KDC Server Host
kdc

kdc.lab



KDC Admin Server Host
admin_server

kdc.lab



Domain Name(s)



Maximum Renewable Life
for Principals

5

day(s)



Manage krb5.conf

- Selecione a opção Manage krb5.conf para que o Cloudera Manager cuide da configuração do Kerberos

Enable Kerberos for Cluster 1

Manage krb5.conf

Specify the properties needed for generating the krb5.conf file for the cluster. You can use the Advanced Configuration Snippet to specify configuration of an advanced KDC setup; for example, with cross-realm authentication.

Manage krb5.conf through
Cloudera Manager



Kerberos Ticket Lifetime
ticket_lifetime

day(s) ▼



Kerberos Renewable
Lifetime
renew_lifetime

day(s) ▼



DNS Lookup KDC
dns_lookup_kdc



Forwardable Tickets
forwardable



KDC Timeout
kdc_timeout

second ▼



KDC Account

- O Cloudera Manager cria todos os usuários necessários no Kerberos
- É necessário ter um usuário com acesso de administrador no Kerberos
 - Username: cloudera-scm/admin

Enable Kerberos for Cluster 1

Setup KDC Account

Enter the credentials for the account that has permissions to **create** other users. Cloudera Manager will store the credentials in encrypted form and use them whenever new principals need to be generated.

Username

cloudera-s

@

LAB.BIGDATA.COM



Password

Import Credentials

- O CM faz a criação e autenticação dos usuários no Kerberos de forma transparente

Enable Kerberos for Cluster 1

Import KDC Account Manager Credentials Command

Status  **Finished**  Sep 5, 1:50:20 PM  5.06s

Successfully imported KDC Account Manager credentials.

Configure Principals

- Os usuários do Kerberos são geralmente definidos pelos usuários do próprio serviço, como pode ser conferido ao lado

Enable Kerberos for Cluster 1

Configure Principals

Specify the Kerberos principal used by each service in the cluster. Additional steps may be required if you decide to change these principals from their default values. Please read the [documentation](#) about custom principals before making changes on this page.

Kerberos Principal	
Flume (Service-Wide)	<input type="text" value="flume"/>
HDFS (Service-Wide)	<input type="text" value="hdfs"/>
Hive (Service-Wide)	<input type="text" value="hive"/>
Hue (Service-Wide)	<input type="text" value="hue"/>
Impala (Service-Wide)	<input type="text" value="impala"/>
Oozie (Service-Wide)	<input type="text" value="oozie"/>
Spark (Service-Wide)	<input type="text" value="spark"/>
YARN (MR2 Included) (Service-Wide)	<input type="text" value="yarn"/>
ZooKeeper (Service-Wide)	<input type="text" value="zookeeper"/>

Configure Ports

- As portas podem continuar com o padrão e, em seguida, vamos reiniciar o cluster para que a configuração de segurança seja habilitada

Enable Kerberos for Cluster 1

Configure Ports

Configure the privileged ports required by DataNodes in a secure HDFS service.

DataNode Transceiver Port	<input type="text" value="1004"/>
Port for DataNode's Xceiver Protocol. Combined with the DataNode's hostname to build its address.	
DataNode HTTP Web UI Port	<input type="text" value="1006"/>
Port for the DataNode HTTP web UI. Combined with the DataNode's hostname to build its HTTP address.	

The cluster needs to be restarted for the changes to take effect.

☒ Yes, I am ready to restart the cluster now.

Enable Kerberos

- O último passo do processo de habilitação do Kerberos é a reinicialização dos serviços
- Observação: faça o download e atualize a client configuration em todos os hosts não gerenciados pelo CM









Enable Kerberos for Cluster 1

Enable Kerberos Command

Status ☒ Running Context [Cluster 1](#)  Sep 5, 9:00:41 PM [Abort](#)

▼ [Completed 7 of 8 step\(s\).](#)

☒ Show All Steps ☐ Show Only Failed Steps ☐ Show Running Steps

>  Stop cluster	Cluster 1	Sep 5, 9:00:41 PM	30.71s
>  Stop Cloudera Management Services	Cloudera Management Service	Sep 5, 9:01:12 PM	9.55s
>  Deploy krb5.conf	Cluster 1	Sep 5, 9:01:21 PM	17.63s
>  Configure all services to use Kerberos	Cluster 1	Sep 5, 9:01:39 PM	35ms
>  Wait for credentials to be generated		Sep 5, 9:01:39 PM	16.94s
>  Deploy client configuration	Cluster 1	Sep 5, 9:01:57 PM	17.91s
>  Start Cloudera Management Services	Cloudera Management Service	Sep 5, 9:02:14 PM	24.71s
>  Start cluster	Cluster 1	Sep 5, 9:02:40 PM	Abort

Autenticando como dataengineer

- Com a segurança, a utilização do cluster é limitada a usuários autenticados autenticados no Kerberos
- Para autentica no Kerberos com o usuário dataengineer:
 - \$ kinit dataengineer
- Para listar as configurações do ticket:
 - \$ klist -e
- Caso um usuário não autenticado tente acessar o cluster é disparada a exceção vista na imagem

```
Exception in thread "main" java.io.IOException: Failed on local exception: java.io.IOException: javax.security.sasl.SaslException: GSS initiate failed [Caused by GSSException: No valid credentials provided (Mechanism level: Failed to find any Kerberos tgt)]; Host Details : local host is: "hadoopclient.lab/127.0.1.1"; destination host is: "headnode.lab":8020;
    at org.apache.hadoop.net.NetUtils.wrapException(NetUtils.java:772)
    at org.apache.hadoop.ipc.Client.call(Client.java:1508)
    at org.apache.hadoop.ipc.Client.call(Client.java:1441)
    at org.apache.hadoop.ipc.ProtobufRpcEngine$Invoker.invoke(ProtobufRpcEngine.java:230)
    at com.sun.proxy.$Proxy16.getFileInfo(Unknown Source)
    at org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolTranslatorPB.getFileInfo(ClientNamenodeProtocolTranslatorPB.java:788)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)
    at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
    at java.lang.reflect.Method.invoke(Method.java:606)
    at org.apache.hadoop.io.retry.RetryInvocationHandler.invokeMethod(RetryInvocationHandler.java:258)
    at org.apache.hadoop.io.retry.RetryInvocationHandler.invoke(RetryInvocationHandler.java:104)
    at com.sun.proxy.$Proxy17.getFileInfo(Unknown Source)
    at org.apache.hadoop.hdfs.DFSClient.getFileInfo(DFSClient.java:2168)
    at org.apache.hadoop.hdfs.DistributedFileSystem$20.doCall(DistributedFileSystem.java:1266)
```



Autenticando como superuser

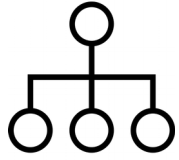
- Operações de administração e manutenção do HDFS são permitidas apenas com autenticação do superuser (hdfs)
- Além da senha, é possível autenticar no Kerberos com o keytab, um arquivo que guarda a senha criptografada do usuário
 - O usuário hdfs não tem senha definida
- No CDH, a keytab está gravada no diretório `/run/cloudera-scm-agent/process/`
- No exemplo, vamos autenticar o usuário hdfs a partir da keytab
 - `$ find /run/cloudera-scm-agent/process/ -name hdfs.keytab`
 - `$ su - hdfs` # opcional
 - # Anote o caminho da keytab
 - `$ kinit -kt $arquivokeytab hdfs/nome-da-maquina@LAB.BIGDATA.COM`
- Para verificar se o usuário está com poderes de superuser execute o `fsck` com o usuário hdfs, que é o único habilitado para tarefas administrativas
 - `$ hdfs fsck /`

Usuários e grupos do cluster



- Usuários e grupos
 - `hdfs:hdfs` – superusuário (já existente no Hadoop)
 - `dataengineer:dataengineer` - amplo acesso
 - `datascientist:datascientist` - amplo acesso
 - `diego:marketing` – acesso local e limitado
 - `mariana:vendas` – acesso local e limitado
- Objetivo: os usuários do marketing não podem acessar os dados do grupo vendas

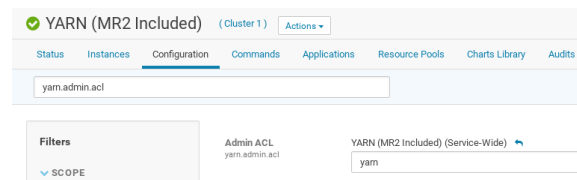
Permissões no HDFS



- Kerberos
 - `$ kadmin.local -q "addprinc -pw diego diego"`
 - `$ kadmin.local -q "addprinc -pw mariana mariana"`
- Linux (todos os hosts)
 - `$ groupadd marketing`
 - `$ groupadd vendas`
 - `$ useradd -m diego -s /bin/bash`
 - `$ usermod -a -G marketing diego`
 - `$ useradd -m mariana -s /bin/bash`
 - `$ usermod -a -G vendas mariana`
- Usuários HDFS (como usuário hdfs)
 - `$ hdfs dfs -mkdir /user/diego`
 - `$ hdfs dfs -mkdir /user/mariana`
 - `$ hdfs dfs -chown -R diego /user/diego/`
 - `$ hdfs dfs -chown -R mariana /user/mariana/`
- Grupos HDFS (como usuário hdfs)
 - `$ hdfs dfs -mkdir /user/marketing/`
 - `$ hdfs dfs -mkdir /user/vendas/`
 - `$ hdfs dfs -chgrp -R marketing /user/marketing/`
 - `$ hdfs dfs -chgrp -R vendas /user/vendas/`
 - `$ hdfs dfs -chmod -R 770 /user/marketing/`
 - `$ hdfs dfs -chmod -R 770 /user/vendas/`

Permissões no YARN

- O padrão do YARN é permitir que todos os usuários submetam aplicações ao cluster
- Vamos limitar para que apenas uma lista de usuários/grupos esteja autorizada para submissão
- No serviço YARN, selecione Configuration e a propriedade yarn.admin.acl, alterando seu valor para yarn, ou seja, apenas o usuário yarn poderá submeter aplicações
 - Esse valor será redefinido nas configurações a seguir



Permissões no Dynamic Pool

- Podemos configurar a lista de usuários/grupos cada um dos pools
- Selecione o menu superior Cluster → Dynamic Resource Pool Configuration → selecione o primeiro Resource Pool (root) → clique no botão Edit
- Na aba Submission Access Control:
 - Usuários: yarn, dataengineer e datascientist
 - Grupos: marketing
- Administration Access Control:
 - Usuário: yarn
- Clique no botão Refresh Dynamic Resource Pools
- Para validar as permissões tente rodar uma aplicação com o usuário mariana, que não está entre os usuários habilitados
- Espera-se que o cluster informe a mensagem de erro:
 - Failed to submit application_1536628374499_0007 to YARN : User mariana cannot submit applications to queue root.users.mariana
- Observação 1: cada Resource Pool pode ter uma lista de usuários/grupos habilitados
- Observação 2: apenas o usuário yarn está habilitado para forçar a parada das aplicações
 - Isso previne que um usuário possa parar a execução das aplicações de outros usuários

The image displays two screenshots of the 'Edit Resource Pool' configuration page in a web interface. The top screenshot shows the 'Submission Access Control' tab, where the 'Allow these users and groups to submit to this pool' option is selected. The 'Users' field contains 'yarn,dataengineer,datascientist' and the 'Groups' field contains 'marketing'. The bottom screenshot shows the 'Administration Access Control' tab, where the 'Allow these' option is selected. The 'Users' field contains 'yarn' and the 'Groups' field is empty, with a placeholder text 'Comma-separated list of groups.' Both screenshots have 'Cancel' and 'Save' buttons at the bottom right.

Dúvidas?

Marco Reis
<http://marcoreis.net>