



Gaël YIMEN &lt;yimengael@gmail.com&gt;

---

## Probleme avec la commande LS sur Hadoop

9 messages

Gaël YIMEN &lt;yimengael@gmail.com&gt;

22 janvier 2015 18:55

À : hadoop-user-group-france@googlegroups.com

Bonjour Chers Hadoopiens,

Je reviens sur la plateforme avec un souci qui me tracasse depuis un moment.  
Je souhaite faire une commande LS sous Hadoop afin d'en exploiter le résultat de façon optimale.

### Mon problème est le suivant:

Je souhaite exploiter le résultat de mon **LS** tel que présenté ci-dessous. Le résultat ci-dessous est obtenu sur un OS compatible POSIX.

Code:

=====

```
biadmin@bivm:~> ls -l file[1-6]*
file1.txt
file2.txt
file5.txt
file6.txt
biadmin@bivm:~>
=====
```

Par contre sur mon système Hadoop (HDFS n'étant pas compatible POSIX), je ne peux utiliser l'option **-l** comme précédemment. J'ai donc la commande suivante :

Code:

=====

```
biadmin@bivm:~/POC_BIGDATA_SCRIPT-In-Dev> hadoop fs -ls /biginsights/hive/warehouse/profil
Found 2 items
-rw-r--r--  1 biadmin biadmin    171495 2015-01-22 16:34 /biginsights/hive/warehouse/
profil/part-r-00000-profil-ext-1421940833734
-rw-r--r--  1 biadmin biadmin    9137051 2015-01-22 16:33 /biginsights/hive/warehouse/
profil/part-r-00000-profil-int-22012015163425
biadmin@bivm:~/POC_BIGDATA_SCRIPT-In-Dev>
=====
```

Ainsi, lorsque je charge le résultat de cette commande dans une variable telle que présenté dans le code suivant,

Code:

=====

```
biadmin@bivm:~> list_elt=`hadoop fs -ls /biginsights/hive/warehouse/profil`
biadmin@bivm:~> echo $list_elt
Found 2 items -rw-r--r-- 1 biadmin biadmin 171495 2015-01-22 16:34 /biginsights
/hive/warehouse/profil/part-r-00000-profil-ext-1421940833734 -rw-r--r-- 1 biadmin biadmin
9137051 2015-01-22 16:33 /biginsights/hive/warehouse/profil/part-r-00000-profil-
int-22012015163425
biadmin@bivm:~>
=====
```

En parcourant la variable **\$list\_elt**, je n'ai besoin que d'utiliser les valeurs suivantes : **part-r-00000-profil-ext-1421940833734** et **part-r-00000-profil-int-22012015163425**

La question

Y aurait-il un moyen d'obtenir le résultat de la commande **LS -1** sur Hadoop comme sur un OS compatible POSIX ?

La version de Hadoop que j'utilise est : **Hadoop\_1.1.1.**

Mon OS est : **SUSE Linux Enterprise Server 11 (x86\_64), VERSION = 11, PATCHLEVEL = 2**

Merci par avance pour vos retours. 😊 😊

Gaël.

--

---

**Aurélien GOUJET MapR** <agoujet@mapr.com>  
Répondre à : hadoop-user-group-france@googlegroups.com  
À : hadoop-user-group-france@googlegroups.com

28 janvier 2015 22:56

Bonjour Gaël,

Ca fonctionne avec notre distribution chez MapR puisque nous avons un double accès NFS/POSIX et HDFS pour ne pas locké les clients et garantir que les applis existantes fonctionnent et soient portable. Notre FS permet justement la simplification du scripting et coding autour de notre data platform.

Cette vidéo de 12 min est intéressante sur le même sujet : <https://www.mapr.com/resources/videos/comparing-mapr-fs-and-hdfs-nfs-and-snapshots>

Voici l'exemple avec MapR pour ta question :

```
+++++
[root@ip-10-68-7-26 metrics]# hadoop fs -ls /var/mapr/metrics/
Found 7 items
-rw-r--r--  3 mapr mapr    2334 2015-01-07 21:20 /var/mapr/metrics/January.07.2015
-rw-r--r--  3 mapr mapr    4262 2015-01-15 22:47 /var/mapr/metrics/January.15.2015
-rw-r--r--  3 mapr mapr    3953 2015-01-16 22:55 /var/mapr/metrics/January.16.2015
-rw-r--r--  3 mapr mapr    3494 2015-01-22 21:46 /var/mapr/metrics/January.22.2015
-rw-r--r--  3 mapr mapr    3348 2015-01-23 17:01 /var/mapr/metrics/January.23.2015
-rw-r--r--  3 mapr mapr    3828 2015-01-27 22:00 /var/mapr/metrics/January.27.2015
-rw-r--r--  3 mapr mapr    5809 2015-01-28 21:31 /var/mapr/metrics/January.28.2015
```

```
+++++
[root@ip-10-68-7-26 metrics]# hadoop fs -ls -1 /var/mapr/metrics/
-ls: Illegal option -1
Usage: hadoop fs [generic options] -ls [-d] [-h] [-R] [<path> ...]
[root@ip-10-68-7-26 metrics]#
```

```
+++++
[root@ip-10-68-7-26 metrics]# mount
localhost:/mapr on /mapr type nfs (rw,soft,intr,nolock,addr=127.0.0.1)
[root@ip-10-68-7-26 metrics]#
```

```
+++++
[root@ip-10-68-7-26 metrics]# pwd
/mapr/my.mirror.com/var/mapr/metrics
[root@ip-10-68-7-26 metrics]#
```

```
+++++
[root@ip-10-68-7-26 metrics]#
[root@ip-10-68-7-26 metrics]# ls -1 January*
```

January.07.2015  
 January.15.2015  
 January.16.2015  
 January.22.2015  
 January.23.2015  
 January.27.2015  
 January.28.2015  
 [root@ip-10-68-7-26 metrics]#

```

+++++
[root@ip-10-68-7-26 metrics]# ll
total 28
-rw-r--r-- 1 mapr mapr 2334 Jan  7 21:20 January.07.2015
-rw-r--r-- 1 mapr mapr 4262 Jan 15 22:47 January.15.2015
-rw-r--r-- 1 mapr mapr 3953 Jan 16 22:55 January.16.2015
-rw-r--r-- 1 mapr mapr 3494 Jan 22 21:46 January.22.2015
-rw-r--r-- 1 mapr mapr 3348 Jan 23 17:01 January.23.2015
-rw-r--r-- 1 mapr mapr 3828 Jan 27 22:00 January.27.2015
-rw-r--r-- 1 mapr mapr 5809 Jan 28 21:31 January.28.2015
[root@ip-10-68-7-26 metrics]#

```

Cdt

Aurélien GOUJET | SE Director Southern Europe and Benelux | MapR Technologies EMEA

Le 22 janv. 2015 à 18:55, Gaël YIMEN <yimengael@gmail.com> a écrit :

[Texte des messages précédents masqué]

--

Hadoop User Group France  
<http://hugfrance.fr> | <https://twitter.com/hugfrance>

---

Vous recevez ce message, car vous êtes abonné au groupe Google Groupes "Hadoop User Group France".  
 Pour vous désabonner de ce groupe et ne plus recevoir d'e-mails le concernant, envoyez un e-mail à l'adresse  
[hadoop-user-group-france+unsubscribe@googlegroups.com](mailto:hadoop-user-group-france+unsubscribe@googlegroups.com).  
 Pour obtenir davantage d'options, consultez la page <https://groups.google.com/d/optout>.

--

Hadoop User Group France  
<http://hugfrance.fr> | <https://twitter.com/hugfrance>

---

Vous recevez ce message, car vous êtes abonné au groupe Google Groupes "Hadoop User Group France".  
 Pour vous désabonner de ce groupe et ne plus recevoir d'e-mails le concernant, envoyez un e-mail à l'adresse  
[hadoop-user-group-france+unsubscribe@googlegroups.com](mailto:hadoop-user-group-france+unsubscribe@googlegroups.com).  
 Pour obtenir davantage d'options, consultez la page <https://groups.google.com/d/optout>.

---

**matt Lieber** <matt.lieber@gmail.com>

28 janvier 2015 23:07

Répondre à : [hadoop-user-group-france@googlegroups.com](mailto:hadoop-user-group-france@googlegroups.com)

À : "[hadoop-user-group-france@googlegroups.com](mailto:hadoop-user-group-france@googlegroups.com)" <[hadoop-user-group-france@googlegroups.com](mailto:hadoop-user-group-france@googlegroups.com)>

Oui malheureusement les options de 'ls' ne sont pas toutes dispo sur hadoop. (et en tout cas pas sur la plateforme IBM :-);

Je crois qu'il va te falloir faire un petit script bash.. Beaucoup d'exemples ici, ca ne devrait pas prendre beaucoup de temps..:

<http://stackoverflow.com/questions/8518750/to-show-only-file-name-without-the-entire-directory-path>

Matt

2015-01-22 9:55 GMT-08:00 Gaël YIMEN <yimengael@gmail.com>:

[Texte des messages précédents masqué]

--

Hadoop User Group France  
<http://hugfrance.fr> | <https://twitter.com/hugfrance>

---

Vous recevez ce message, car vous êtes abonné au groupe Google Groupes "Hadoop User Group France".  
Pour vous désabonner de ce groupe et ne plus recevoir d'e-mails le concernant, envoyez un e-mail à l'adresse [hadoop-user-group-france+unsubscribe@googlegroups.com](mailto:hadoop-user-group-france+unsubscribe@googlegroups.com).  
Pour obtenir davantage d'options, consultez la page <https://groups.google.com/d/optout>.

[Texte des messages précédents masqué]

---

**Gaël YIMEN** <yimengael@gmail.com>  
À : hadoop-user-group-france@googlegroups.com

28 janvier 2015 23:18

Merci Aurélien, je comprends très bien la manip. Dans la distro entreprise de IBM, il existe le GPFS qui est POSIX-compliant. Il est certainement possible de faire fonctionner toutes les options de "ls". Malheureusement, je n'ai pas encore accès à cette distro.

Toutefois, la solution que me propose Matt semble répondre à mon problème. Je vais regarder le lien sur STACKOVERFLOW et je te fais un retour Matt (A tous les hadoopiens). Je souhaite mettre en place un script shell qui puisse me permettre de faire du "ls -l" sous HDFS.

Merci à tous pour vos retours.

Vive le HUG France :) :) :)

Gaël.

--

Eng, OCADB  
[yimengael@gmail.com](mailto:yimengael@gmail.com)  
skype id : yimen.gael  
(m) +33 7 81 67 78 23  
<http://yimengael.blogspot.fr/>

Le 28 janvier 2015 23:07, matt Lieber <matt.lieber@gmail.com> a écrit :

[Texte des messages précédents masqué]

---

**Aurélien GOUJET MapR** <agoujet@mapr.com>  
Répondre à : hadoop-user-group-france@googlegroups.com  
À : hadoop-user-group-france@googlegroups.com

28 janvier 2015 23:22

Avec plaisir !

A plus  
Aurélien

---

Le 28 janv. 2015 à 23:18, Gaël YIMEN <yimengael@gmail.com> a écrit :

[Texte des messages précédents masqué]

[Texte des messages précédents masqué]

---

**Raphaël Luta** <raphael.luta@aptiwan.com>

29 janvier 2015 10:12

Répondre à : [hadoop-user-group-france@googlegroups.com](mailto:hadoop-user-group-france@googlegroups.com)

À : [hadoop-user-group-france@googlegroups.com](mailto:hadoop-user-group-france@googlegroups.com)

Si tu as beaucoup de manipulations de ce type, il peut être plus simple pour toi de monter ton HDFS comme un filesystem Unix et de travailler directement avec des outils Unix.

Au delà de l'implémentation MapR, qui est certainement la plus poussée techniquement sur ce point, il existe pas mal de pilotes FUSE pour HDFS (<https://wiki.apache.org/hadoop/MountableHDFS>).

Sinon pour répondre à ton besoin direct, awk est ton ami :)

```
hadoop fs -ls /biginsights/hive/warehouse/profil | awk '{print $8}'
```

te donnera uniquement le contenu de la huitième colonne de la sortie de hadoop ls, ie le nom de fichier

-- raphael

> Le 28 janv. 2015 à 23:18, Gaël YIMEN <[yimengael@gmail.com](mailto:yimengael@gmail.com)> a écrit :

>

[Texte des messages précédents masqué]

[Texte des messages précédents masqué]

Pour plus d'options, visitez le site <https://groups.google.com/d/optout> .



**signature.asc**

1K

---

**Gaël YIMEN** <[yimengael@gmail.com](mailto:yimengael@gmail.com)>

29 janvier 2015 10:52

À : [hadoop-user-group-france@googlegroups.com](mailto:hadoop-user-group-france@googlegroups.com)

Bonjour Raphael,

Tu as raison. Voici le résultat que j'obtiens

=====

```
biadmin@bivm:~> hadoop fs -ls /pocbigdataIBM/landingzone/source_definitive/DOC
```

```
Found 02 items
```

```
-rw-r--r--  1 hdfs  biadmin    66842 2015-01-24 04:56 /pocbigdataIBM/landingzone/source_definitive/DOC/AB.docx
```

```
-rw-r--r--  1 hdfs  biadmin    72298 2015-01-24 04:56 /pocbigdataIBM/landingzone/source_definitive/DOC/AN.docx
```

```
biadmin@bivm:~> hadoop fs -ls /pocbigdataIBM/landingzone/source_definitive/CV/DOC | awk '{print $8}'
```

```
/pocbigdataIBM/landingzone/source_definitive/DOC/AB.docx
```

```
/pocbigdataIBM/landingzone/source_definitive/DOC/AN.docx
```

```
biadmin@bivm:~>
```

=====

Ce que je souhaite obtenir c'est uniquement le nom du fichier ( \$basename ). C'est a dire ce qui suit :

=====

AB.docx

AN.docx

=====

Tu vois ce que je veux dire ? :) :)

Gaël.

--

Eng, OCADB

[yimengael@gmail.com](mailto:yimengael@gmail.com)

skype id : yimen.gael

(m) +33 7 81 67 78 23

<http://yimengael.blogspot.fr/>

Le 29 janvier 2015 10:12, Raphaël Luta <[raphael.luta@aptiwan.com](mailto:raphael.luta@aptiwan.com)> a écrit :

[Texte des messages précédents masqué]

---

**Raphaël Luta** <[raphael.luta@aptiwan.com](mailto:raphael.luta@aptiwan.com)>

29 janvier 2015 10:56

Répondre à : [hadoop-user-group-france@googlegroups.com](mailto:hadoop-user-group-france@googlegroups.com)

À : [hadoop-user-group-france@googlegroups.com](mailto:hadoop-user-group-france@googlegroups.com)

hadoop fs -ls /pocbigdataIBM/landingzone/source\_definitive/CV/DOC | awk '{print \$8}' | xargs basename

:)

-- raphael

> Le 29 janv. 2015 à 10:52, Gaël YIMEN <[yimengael@gmail.com](mailto:yimengael@gmail.com)> a écrit :

[Texte des messages précédents masqué]

---

 **signature.asc**  
1K

---

**Gaël YIMEN** <[yimengael@gmail.com](mailto:yimengael@gmail.com)>

29 janvier 2015 11:52

À : [hadoop-user-group-france@googlegroups.com](mailto:hadoop-user-group-france@googlegroups.com)

Hi Raphael,

J'ai utilisé la commande avec l'utilitaire AWK de la façon suivante :

====

```
local PROG_HADOOP=`which hadoop`
suff_name=`date +%d%m%Y%H%M%S`
for i in $(($PROG_HADOOP fs -ls $files_to_import | awk '{print $8}'))
do
    if $PROG_HADOOP fs -test -e "$i"; then
        #suff_name=`date +%d%m%Y%H%M%S`
        $PROG_HADOOP fs -mv "$i" "$HIVE_WAREHOUSE/data/$(basename $i)-data-$suff_name"
    fi
done
====
```

Et ça marche parfaitement !!!

Néanmoins pour améliorer et éviter de faire le test dans la boucle for, lorsque j'utilise ta commande, j'ai le message suivant:

====

```
biadmin@bivm:~> hadoop fs -ls /pocbigdataIBM/landingzone/source_definitive/CV/DOC | awk '{print $8}' | xargs
basename
basename: opérande supplémentaire « /pocbigdataIBM/landingzone/source_definitive/CV/DOC/AP.docx »
Saisissez « basename --help » pour plus d'informations.
biadmin@bivm:~>
====
```

Merci encore pour le support :) :)

Gaël.

--

Eng. OCADB

[yimengael@gmail.com](mailto:yimengael@gmail.com)

skype id : yimen.gael  
(m) +33 7 81 67 78 23  
<http://yimengael.blogspot.fr/>

Le 29 janvier 2015 10:56, Raphaël Luta <[raphael.luta@aptiwan.com](mailto:raphael.luta@aptiwan.com)> a écrit :

[Texte des messages précédents masqué]