

- MobaXterm Personal Edition v22.0 •
(X server, SSH client and network tools)

→ Your computer drives are accessible through the `/drives` path

→ Your `DISPLAY` is set to `192.xxx.x.xxx:x.x`


→ When using `SSH`, your remote `DISPLAY` is automatically forwarded

→ Each command status is specified by a special symbol (`v` or `x`)

- Important:

This is MobaXterm Personal Edition. The Professional edition allows you to customize MobaXterm for your company: you can add your own logo, your parameters, your welcome message and generate either an MSI installation package or a portable executable. We can also modify MobaXterm or develop the plugins you need.

For more information: <https://mobaxterm.mobatek.net/download.html>



version 2.4.7

```
scala> ls
<console>:24: error: not found: value ls
  ls
  ^

scala> exit
<console>:24: error: not found: value exit
```

```

        exit
        ^
scala>
[1]+  Stopped                  spark-shell
[everis@bigdata-srv spark]$ ls
bin  conf  data  examples  jars  kubernetes  LICENSE  licenses  NOTICE  python
R  README.m
[everis@bigdata-srv spark]$ cd sbin/
[everis@bigdata-srv sbin]$ ls
slaves.sh          start-history-server.sh          start-slave.sh
stop-master.sh
spark-config.sh    start-master.sh                  start-slaves.sh
stop-mesos-dispat
spark-daemon.sh    start-mesos-dispatcher.sh        start-thriftserver.sh
stop-mesos-shuffl
spark-daemons.sh  start-mesos-shuffle-service.sh   stop-all.sh
stop-shuffle-serv
start-all.sh      start-shuffle-service.sh         stop-history-server.sh
stop-slave.sh
[everis@bigdata-srv sbin]$ cd ..
[everis@bigdata-srv spark]$ cd conf/
[everis@bigdata-srv conf]$ ls
docker.properties.template  log4j.properties.template  slaves.template
spar
fairscheduler.xml.template  metrics.properties.template
spark-defaults.conf.template
[everis@bigdata-srv conf]$ cd ..
[everis@bigdata-srv spark]$ cd bin ..
[everis@bigdata-srv bin]$ ls
beeline            find-spark-home.cmd  pyspark2.cmd      spark-class
sparkR2.cmd
beeline.cmd        load-spark-env.cmd   pyspark.cmd        spark-class2.cmd
sparkR.cmd
docker-image-tool.sh  load-spark-env.sh    run-example        spark-class.cmd
spark-shell
find-spark-home      pyspark              run-example.cmd    sparkR
spark-shell2.
[everis@bigdata-srv bin]$ cd ..
[everis@bigdata-srv spark]$ wget
https://raw.githubusercontent.com/fivethirtyeight/data/mas
--2022-04-14 00:52:42--
https://raw.githubusercontent.com/fivethirtyeight/data/master/aven
Resolving raw.githubusercontent.com (raw.githubusercontent.com)...
2606:50c0:8001::154, 260
Connecting to raw.githubusercontent.com
(raw.githubusercontent.com)|2606:50c0:8001::154|:44
Connecting to raw.githubusercontent.com
(raw.githubusercontent.com)|2606:50c0:8000::154|:44
Connecting to raw.githubusercontent.com
(raw.githubusercontent.com)|2606:50c0:8002::154|:44
[everis@bigdata-srv spark]$ ls
bin  conf  data  examples  jars  kubernetes  LICENSE  licenses  NOTICE  python
R  README.m
[everis@bigdata-srv spark]$ wget
https://raw.githubusercontent.com/fivethirtyeight/data/mas

```

```
--2022-04-14 01:01:09--
https://raw.githubusercontent.com/fivethirtyeight/data/master/aven
Resolving raw.githubusercontent.com (raw.githubusercontent.com)...
2606:50c0:8001::154, 260
Connecting to raw.githubusercontent.com
(raw.githubusercontent.com)|2606:50c0:8001::154|:44
[2]+  Stopped                  wget
https://raw.githubusercontent.com/fivethirtyeight/data/m
[everis@bigdata-srv spark]$ cd ..
[everis@bigdata-srv opt]$ cd ..
[everis@bigdata-srv /]$ cd ~
[everis@bigdata-srv ~]$ wget
https://raw.githubusercontent.com/fivethirtyeight/data/master/
--2022-04-14 01:02:29--
https://raw.githubusercontent.com/fivethirtyeight/data/master/aven
Resolving raw.githubusercontent.com (raw.githubusercontent.com)...
2606:50c0:8003::154, 260
Connecting to raw.githubusercontent.com
(raw.githubusercontent.com)|2606:50c0:8003::154|:44
Connecting to raw.githubusercontent.com
(raw.githubusercontent.com)|2606:50c0:8000::154|:44
[3]+  Stopped                  wget
https://raw.githubusercontent.com/fivethirtyeight/data/m
[everis@bigdata-srv ~]$ sudo -u hdfs hadoop dfsadmin -safemode leave
[sudo] password for everis:
Sorry, try again.
[sudo] password for everis:
Sorry, try again.
[sudo] password for everis:
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

Safe mode is OFF
[everis@bigdata-srv ~]$ vim avengers.csv
[everis@bigdata-srv ~]$ ls
000000_0          baths  employee.txt      hive-site.xml      latitude           price
sq
avengers.csv      beds   file_teste.txt    import_sqoop.sh     longitude
sale_date         st
base_localidade.csv  city   firewall.rules    install_sqoop.sh    pokemon.sql
script_apoio      st
[everis@bigdata-srv ~]$ cat avengers.csv |more
URL,Name/Alias,Appearances,Current?,Gender,Probationary Intro,Full/Reserve
Avengers Intro,
turn2,Death3,Return3,Death4,Return4,Death5,Return5,Notes
http://marvel.wikia.com/Henry_Pym_(Earth-616),"Henry Jonathan ""Hank""
Pym",1269,YES,MALE,,
age of Ultron Vol. 1. A funeral was held.
http://marvel.wikia.com/Janet_van_Dyne_(Earth-616),Janet van
Dyne,1165,YES,FEMALE,,Sep-63,1
  Actually was sent tto Microverse later recovered
http://marvel.wikia.com/Anthony_Stark_(Earth-616),"Anthony Edward ""Tony""
Stark",3068,YES,
while under the influence of Immortus Stark committed a number of horrible acts and
was kil
```

ught him back"

[http://marvel.wikia.com/Robert_Bruce_Banner_\(Earth-616\)](http://marvel.wikia.com/Robert_Bruce_Banner_(Earth-616)), Robert Bruce Banner, 2089, YES, MALE,,

Future arc. However ""he had actually used a hidden Pantheon base to survive""

[http://marvel.wikia.com/Thor_Odinson_\(Earth-616\)](http://marvel.wikia.com/Thor_Odinson_(Earth-616)), Thor Odinson, 2402, YES, MALE,, Sep-63, 1963, 52

k because that's kind of the whole point. Second death in Time Runs Out has not yet returned

[http://marvel.wikia.com/Richard_Jones_\(Earth-616\)](http://marvel.wikia.com/Richard_Jones_(Earth-616)), Richard Milhouse Jones, 612, YES, MALE,, Sep-

[http://marvel.wikia.com/Steven_Rogers_\(Earth-616\)](http://marvel.wikia.com/Steven_Rogers_(Earth-616)), Steven Rogers, 3458, YES, MALE,, Mar-64, 1964,

r comes back.

[http://marvel.wikia.com/Clint_Barton_\(Earth-616\)](http://marvel.wikia.com/Clint_Barton_(Earth-616)), Clinton Francis Barton, 1456, YES, MALE,, May-

ee ship in Avengers Vol. 1 Issue 502. Brought back by Scarlet Witch. Dies again in House

[http://marvel.wikia.com/Pietro_Maximoff_\(Earth-616\)](http://marvel.wikia.com/Pietro_Maximoff_(Earth-616)), Pietro Maximoff, 769, YES, MALE,, May-65, 19

7. Later comes back

[http://marvel.wikia.com/Wanda_Maximoff_\(Earth-616\)](http://marvel.wikia.com/Wanda_Maximoff_(Earth-616)), Wanda Maximoff, 1214, YES, FEMALE,, May-65, 1

_14. Later comes back

[http://marvel.wikia.com/Jacques_Duquesne_\(Earth-616\)](http://marvel.wikia.com/Jacques_Duquesne_(Earth-616)), Jacques Duquesne, 115, NO, MALE,, Sep-65, 1

ought back by the Chaos King

[http://marvel.wikia.com/Hercules_\(Earth-616\)](http://marvel.wikia.com/Hercules_(Earth-616)), Heracles, 741, YES, MALE,, Oct-67, 1967, 48, Full, NO,

[http://marvel.wikia.com/T%27Challa_\(Earth-616\)](http://marvel.wikia.com/T%27Challa_(Earth-616)), T'Challa, 780, NO, MALE,, May-68, 1968, 47, Full, NO

[http://marvel.wikia.com/Vision_\(Earth-616\)](http://marvel.wikia.com/Vision_(Earth-616)), Victor Shade (alias), 1036, YES, MALE,, Nov-68, 1968,

ntually rebuilt.

http://marvel.wikia.com/Dane_Whitman, Dane Whitman, 482, NO, MALE,, Dec-69, 1969, 46, Full, NO,,,,,

[http://marvel.wikia.com/Natalia_Romanova_\(Earth-616\)#](http://marvel.wikia.com/Natalia_Romanova_(Earth-616)#), Natalia Alianovna Romanova, 1112, YES, F

and. Later revived with The Stone of the Chaste.

[http://marvel.wikia.com/Mantis_\(Earth-616\)#](http://marvel.wikia.com/Mantis_(Earth-616)#), Brandt, 160, NO, FEMALE,, Aug-73, 1973, 42, Full, YES, Y

ments of her essence were spread out all across the universe""

[http://marvel.wikia.com/Henry_McCoy_\(Earth-616\)#](http://marvel.wikia.com/Henry_McCoy_(Earth-616)#), Henry P. McCoy, 1886, NO, MALE, Jul-75, Sep-76,

[http://marvel.wikia.com/Heather_Douglas_\(Earth-616\)#](http://marvel.wikia.com/Heather_Douglas_(Earth-616)#), Heather Douglas, 332, NO, FEMALE, Jul-75, S

s_Vol_1_152. Later 'Obtained a new body.' Dies in Annihilation:_Conquest_Vol_1_3. Comes bac

--More--

[11]+ Stopped cat avengers.csv | more

[everis@bigdata-srv ~]\$ **spark-shell**

22/04/14 01:17:09 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

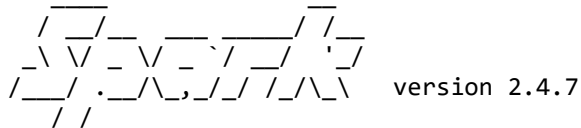
Setting default log level to "WARN".

To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).

```

22/04/14 01:18:27 WARN util.Utils: Service 'SparkUI' could not bind on port 4040.
Attemptinort 4041.
22/04/14 01:18:27 WARN util.Utils: Service 'SparkUI' could not bind on port 4041.
Attemptinort 4042.
Spark context Web UI available at http://bigdata-srv:4042
Spark context available as 'sc' (master = local[*], app id = local-1649906307602).
Spark session available as 'spark'.
Welcome to

```



```

Using Scala version 2.11.12 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_202)
Type in expressions to have them evaluated.
Type :help for more information.

```

```

scala> val insurance = spark.read.format("csv").option("sep",
",").option("header", "true").load("file://home/everis/avengers.csv")
22/04/14 01:21:07 WARN streaming.FileStreamSink: Error while looking for metadata
directory
java.lang.IllegalArgumentException: Wrong FS: file://home/everis/avengers.csv,
expected: fi:///
    at org.apache.hadoop.fs.FileSystem.checkPath(FileSystem.java:645)
    at
    org.apache.hadoop.fs.RawLocalFileSystem.pathToFile(RawLocalFileSystem.java:80)
    at
    org.apache.hadoop.fs.RawLocalFileSystem.deprecatedGetFileStatus(RawLocalFileSyst
em.jav34)
    at
    org.apache.hadoop.fs.RawLocalFileSystem.getFileLinkStatusInternal(RawLocalFileSy
stem.j:752)
    at
    org.apache.hadoop.fs.RawLocalFileSystem.getFileStatus(RawLocalFileSystem.java:52
9)
    at
    org.apache.hadoop.fs.FilterFileSystem.getFileStatus(FilterFileSystem.java:409)
    at org.apache.hadoop.fs.FileSystem.exists(FileSystem.java:1400)
    at
    org.apache.spark.sql.execution.datasources.DataSource$$anonfun$org$apache$spark$
sql$extion$datasources$DataSource$$checkAndGlobPathIfNecessary$1.apply(DataSourc
e.scala:575)
    at
    org.apache.spark.sql.execution.datasources.DataSource$$anonfun$org$apache$spark$
sql$extion$datasources$DataSource$$checkAndGlobPathIfNecessary$1.apply(DataSourc
e.scala:559)
    at
    scala.collection.TraversableLike$$anonfun$flatMap$1.apply(TraversableLike.scala:
241)
    at
    scala.collection.TraversableLike$$anonfun$flatMap$1.apply(TraversableLike.scala:
241)
    at scala.collection.immutable.List.foreach(List.scala:392)
    at scala.collection.TraversableLike$class.flatMap(TraversableLike.scala:241)

```

```

    at scala.collection.immutable.List.flatMap(List.scala:355)
    at
org.apache.spark.sql.execution.datasources.DataSource.org$apache$spark$sql$execu
tion$dsource$DataSource$$checkAndGlobPathIfNecessary(DataSource.scala:559)
    at
org.apache.spark.sql.execution.datasources.DataSource.resolveRelation(DataSource
.scala:3)
    at
org.apache.spark.sql.DataFrameReader.loadV1Source(DataFrameReader.scala:242)
    at org.apache.spark.sql.DataFrameReader.load(DataFrameReader.scala:230)
    at org.apache.spark.sql.DataFrameReader.load(DataFrameReader.scala:197)
... 49 elided

```

```

scala> val insurance = spark.read.format("csv").option("sep",
",").option("header", "true")[Stage 0:>
(0 + 1) / 1]22/04/14 01or bigdata-srv/192.xxx.x.xxx:x.x took 16057 ms
insurance: org.apache.spark.sql.DataFrame = [URL: string, Name/Alias: string ... 19
more fi

```

```

scala> insurance.show(10, false)

```

```

+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|URL|Name/Alias|
|Appearances|Current?|Gender|Probationary|Intro|Full/Reserve|Avengers I|
|ntro|Year|Years since joining|Honorary|Death1|Return1|Death2|Return2|Death3|Retu
|rn3|Death4|Return4|Death5|Return5|Notes|
|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|http://marvel.wikia.com/Henry_Pym_(Earth-616)|"Henry Jonathan ""Hank"
|" Pym"|1269|YES|MALE|null|Sep-63|
|1963|52|Full|YES|NO|null|null|null|
|null|null|null|null|Merged with Ultron in Rage of Ultron Vol. 1. A
|funeral was held.|
|
|http://marvel.wikia.com/Janet_van_Dyne_(Earth-616)|Janet van Dyne
|1165|YES|FEMALE|null|Sep-63|
|1963|52|Full|YES|YES|null|null|null|
|null|null|null|null|Dies in Secret Invasion V1:I8. Actually was se
|nt tto Microverse later recovered|
|
|http://marvel.wikia.com/Anthony_Stark_(Earth-616)|"Anthony Edward ""Tony"
|" Stark"|3068|YES|MALE|null|Sep-63|
|1963|52|Full|YES|YES|null|null|null|
|null|null|null|null|"Death: ""Later while under the influence of I
|mmortus Stark committed a number of horrible acts and was killed." This set up
|young Tony. Franklin Richards later brought him back"|

```

```

|http://marvel.wikia.com/Robert_Bruce_Banner_(Earth-616)|Robert Bruce Banner
|2089      |YES      |MALE  |null      |Sep-63
|1963|52      |Full   |YES  |YES  |null |null |null |null
|null |null |null |null |"Dies in Ghosts of the Future arc. However ""h
e had actually used a hidden Pantheon base to survive""
|
|http://marvel.wikia.com/Thor_Odinson_(Earth-616)      |Thor Odinson
|2402      |YES      |MALE  |null      |Sep-63
|1963|52      |Full   |YES  |YES  |YES  |NO   |null |null
|null |null |null |null |Dies in Fear Itself brought back because that'
s kind of the whole point. Second death in Time Runs Out has not yet returned
|
|http://marvel.wikia.com/Richard_Jones_(Earth-616)      |Richard Milhouse Jones
|612       |YES      |MALE  |null      |Sep-63
|1963|52      |Honorary|NO   |null |null |null |null |null
|null |null |null |null |NA
|
|http://marvel.wikia.com/Steven_Rogers_(Earth-616)     |Steven Rogers
|3458      |YES      |MALE  |null      |Mar-64
|1964|51      |Full   |YES  |YES  |null |null |null |null
|null |null |null |null |Dies at the end of Civil War. Later comes back
.
|
|http://marvel.wikia.com/Clint_Barton_(Earth-616)      |Clinton Francis Barton
|1456      |YES      |MALE  |null      |May-65
|1965|50      |Full   |YES  |YES  |YES  |YES  |null |null
|null |null |null |null |Dies in exploding Kree ship in Avengers Vol.
1 Issue 502. Brought back by Scarlet Witch. Dies again in House of M Vol 1 Issu
e 7. Is later brought back.
|
|http://marvel.wikia.com/Pietro_Maximoff_(Earth-616)   |Pietro Maximoff
|769       |YES      |MALE  |null      |May-65
|1965|50      |Full   |YES  |YES  |null |null |null |null
|null |null |null |null |Dies in House of M Vol 1 Issue 7. Later comes
back
|
|http://marvel.wikia.com/Wanda_Maximoff_(Earth-616)    |Wanda Maximoff
|1214      |YES      |FEMALE|null      |May-65
|1965|50      |Full   |YES  |YES  |null |null |null |null
|null |null |null |null |Dies in Uncanny Avengers_Vol_1_14. Later comes
back
|
+-----+-----+
+-----+-----+
+-----+-----+
+-----+-----+
+-----+-----+
+-----+
only showing top 10 rows

```

```

scala> insurance.select("URL").show(1, false)
+-----+
|URL      |
+-----+
|http://marvel.wikia.com/Henry_Pym_(Earth-616)|

```

```
+-----+
only showing top 1 row
```

```
scala> dataset2 = insurance.select("URL")
<console>:27: error: not found: value dataset2
val $ires6 = dataset2
              ^
<console>:25: error: not found: value dataset2
      dataset2 = insurance.select("URL")
      ^
```

```
scala> val dataset2 = insurance.select("URL")
dataset2: org.apache.spark.sql.DataFrame = [URL: string]
```

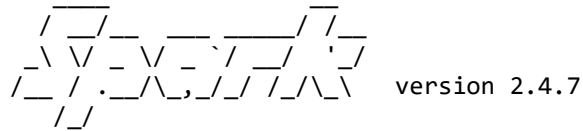
```
scala> insurance.select("URL").show(1, false)
+-----+
|URL|
+-----+
|http://marvel.wikia.com/Henry_Pym_(Earth-616)|
+-----+
only showing top 1 row
```

```
scala> dataset2.show(10, false)
+-----+
|URL|
+-----+
|http://marvel.wikia.com/Henry_Pym_(Earth-616)|
|http://marvel.wikia.com/Janet_van_Dyne_(Earth-616)|
|http://marvel.wikia.com/Anthony_Stark_(Earth-616)|
|http://marvel.wikia.com/Robert_Bruce_Banner_(Earth-616)|
|http://marvel.wikia.com/Thor_Odinson_(Earth-616)|
|http://marvel.wikia.com/Richard_Jones_(Earth-616)|
|http://marvel.wikia.com/Steven_Rogers_(Earth-616)|
|http://marvel.wikia.com/Clint_Barton_(Earth-616)|
|http://marvel.wikia.com/Pietro_Maximoff_(Earth-616)|
|http://marvel.wikia.com/Wanda_Maximoff_(Earth-616)|
+-----+
only showing top 10 rows
```

```
scala>
[12]+  Stopped                  spark-shell
[everis@bigdata-srv ~]$ pyspark
Python 3.7.9 (default, Jan 14 2021, 10:35:13)
[GCC 4.8.5 20150623 (Red Hat 4.8.5-44)] on linux
Type "help", "copyright", "credits" or "license" for more information.
22/04/14 01:34:36 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLeve
l(newLevel).
22/04/14 01:35:42 WARN util.Utills: Service 'SparkUI' could not bind on port 4040
. Attempting port 4041.
22/04/14 01:35:42 WARN util.Utills: Service 'SparkUI' could not bind on port 4041
```



```
. Attempting port 4042.
22/04/14 01:35:42 WARN util.Utils: Service 'SparkUI' could not bind on port 4042
. Attempting port 4043.
Welcome to
```



```
Using Python version 3.7.9 (default, Jan 14 2021 10:35:13)
SparkSession available as 'spark'.
```

```
>>> insurance.show()
```

```
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+
|          URL|          Name/Alias|Appearances|Current?|Gender|Probation
ary Intro|Full/Reserve Avengers Intro|Year|Years since joining|Honorary|Death1|
Return1|Death2|Return2|Death3|Return3|Death4|Return4|Death5|Return5|
Notes|
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+
|http://marvel.wik...|"Henry Jonathan "...|          1269|          YES|  MALE|
null|          Sep-63|1963|          52|          Full|  YES|
NO| null| null| null| null| null| null| null| null|Merged with
Ultro...|
|http://marvel.wik...|          Janet van Dyne|          1165|          YES|FEMALE|
null|          Sep-63|1963|          52|          Full|  YES|
YES| null| null| null| null| null| null| null| null|Dies in Secr
et In...|
|http://marvel.wik...|"Anthony Edward "...|          3068|          YES|  MALE|
null|          Sep-63|1963|          52|          Full|  YES|
YES| null| null| null| null| null| null| null| null|"Death: ""La
ter w...|
|http://marvel.wik...| Robert Bruce Banner|          2089|          YES|  MALE|
null|          Sep-63|1963|          52|          Full|  YES|
YES| null| null| null| null| null| null| null| null|"Dies in Gho
sts o...|
|http://marvel.wik...|          Thor Odinson|          2402|          YES|  MALE|
null|          Sep-63|1963|          52|          Full|  YES|
YES| YES| NO| null| null| null| null| null| null|Dies in Fear
Itse...|
|http://marvel.wik...|Richard Milhouse ...|          612|          YES|  MALE|
null|          Sep-63|1963|          52|Honorary| NO|
null| null| null| null| null| null| null| null| null|
NA|
|http://marvel.wik...|          Steven Rogers|          3458|          YES|  MALE|
null|          Mar-64|1964|          51|          Full|  YES|
YES| null| null| null| null| null| null| null| null|Dies at the
end o...|
|http://marvel.wik...|Clinton Francis B...|          1456|          YES|  MALE|
null|          May-65|1965|          50|          Full|  YES|
```

YES	YES	YES	null	null	null	null	null	null	Dies in exploding...
http://marvel.wik...	Pietro Maximoff	769	YES	MALE					
null	May-65 1965	50	Full	YES					
YES	null	null	null	null	null	null	null	null	Dies in House of ...
http://marvel.wik...	Wanda Maximoff	1214	YES	FEMALE					
null	May-65 1965	50	Full	YES					
YES	null	null	null	null	null	null	null	null	Dies in Uncanny_A...
http://marvel.wik...	Jacques Duquesne	115	NO	MALE					
null	Sep-65 1965	50	Full	YES					
YES	null	null	null	null	null	null	null	null	Dies in Avengers_...
http://marvel.wik...	Heracles	741	YES	MALE					
null	Oct-67 1967	48	Full	NO					
null	null	null	null	null	null	null	null	null	
NA									
http://marvel.wik...	T'Challa	780	NO	MALE					
null	May-68 1968	47	Full	NO					
null	null	null	null	null	null	null	null	null	
NA									
http://marvel.wik...	Victor Shade (alias)	1036	YES	MALE					
null	Nov-68 1968	47	Full	YES					
YES	null	null	null	null	null	null	null	null	Dies in Avengers_...
http://marvel.wik...	Dane Whitman	482	NO	MALE					
null	Dec-69 1969	46	Full	NO					
null	null	null	null	null	null	null	null	null	
NA									
http://marvel.wik...	Natalia Alianovna...	1112	YES	FEMALE					
null	May-73 1973	42	Full	YES					
YES	null	null	null	null	null	null	null	null	Killed by The Han...
http://marvel.wik...	Brandt	160	NO	FEMALE					
null	Aug-73 1973	42	Full	YES					
YES	null	null	null	null	null	null	null	null	"Dies in Silver_S...
http://marvel.wik...	Henry P. McCoy	1886	NO	MALE					
Jul-75	Sep-76 1976	39	Full	NO					
null	null	null	null	null	null	null	null	null	
NA									
http://marvel.wik...	Heather Douglas	332	NO	FEMALE					
Jul-75	Sep-76 1976	39	Full	YES					
YES	YES	YES	null	null	null	null	null	null	"Dies in Defender...
http://marvel.wik...	Patsy Walker	557	NO	FEMALE					
Nov-75	Sep-76 1976	39	Full	YES					
YES	null	null	null	null	null	null	null	null	Died during time ...

```


+-----+-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+-----+
-----+

```

only showing top 20 rows


```
-- Death4: string (nullable = true)
-- Return4: string (nullable = true)
-- Death5: string (nullable = true)
-- Return5: string (nullable = true)
-- Notes: string (nullable = true)
```

```
scala> [everis@bigdata-srv ~]$ spark-shell
22/04/14 02:26:33 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Spark context Web UI available at http://bigdata-srv:4040
Spark context available as 'sc' (master = local[*], app id = local-1649910428391).
Spark session available as 'spark'.
Welcome to
```



```
Using Scala version 2.11.12 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_202)
Type in expressions to have them evaluated.
Type :help for more information.
```

```
scala> df.show(16, false)
<console>:24: error: not found: value df
    df.show(16, false)
    ^
```

```
scala> val df = spark.read.format("csv").option("sep", ",").option("header",
"true").load("ris/avengers.csv")
[Stage 0:> (0 + 1) /
1]22/04/14 02nt.TransportClientFactory: DNS resolution for
bigdata-srv/192.xxx.x.xxx:x.x:37811 took 16062 ms
df: org.apache.spark.sql.DataFrame = [URL: string, Name/Alias: string ... 19 more
fields]
```

```
scala> df.show(16, false)
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|URL                                           |Name/Alias
|Apt?|Gender|Probationary Intro|Full/Reserve Avengers Intro|Year|Years since
joining|HonoraryDeath2|Return2|Death3|Return3|Death4|Return4|Death5|Return5|Note
s
|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
```

```


-----+-----
|http://marvel.wikia.com/Henry_Pym_(Earth-616)      |"Henry Jonathan ""Hank""
Pym" |12 |MALE |null |Sep-63 |1963|52
|Full null |null |null |null |null |null |null |null |Merged with
Ultron in Rage of Ufuneral was held.
|
|http://marvel.wikia.com/Janet_van_Dyne_(Earth-616) |Janet van Dyne
|11 |FEMALE|null |Sep-63 |1963|52
|Full null |null |null |null |null |null |null |null |Dies in
Secret Invasion V1:I8. t tto Microverse later recovered
|
|http://marvel.wikia.com/Anthony_Stark_(Earth-616)  |"Anthony Edward ""Tony""
Stark"|30 |MALE |null |Sep-63 |1963|52
|Full null |null |null |null |null |null |null |null |"Death:
""Later while under themortus Stark committed a number of horrible acts and was
killed.' This set up young Tony. s later brought him back"|
|http://marvel.wikia.com/Robert_Bruce_Banner_(Earth-616)|Robert Bruce Banner
|20 |MALE |null |Sep-63 |1963|52
|Full null |null |null |null |null |null |null |null |"Dies in
Ghosts of the Future a had actually used a hidden Pantheon base to survive""
|
|http://marvel.wikia.com/Thor_Odinson_(Earth-616)  |Thor Odinson
|24 |MALE |null |Sep-63 |1963|52
|Full YES |NO |null |null |null |null |null |null |Dies in Fear
Itself brought bac kind of the whole point. Second death in Time Runs Out has not
yet returned
|
|http://marvel.wikia.com/Richard_Jones_(Earth-616)  |Richard Milhouse Jones
|61 |MALE |null |Sep-63 |1963|52
|Honorarynull |null |null |null |null |null |null |null |NA
|
|http://marvel.wikia.com/Steven_Rogers_(Earth-616)  |Steven Rogers
|34 |MALE |null |Mar-64 |1964|51
|Full null |null |null |null |null |null |null |null |Dies at the
end of Civil War. L
|
|http://marvel.wikia.com/Clint_Barton_(Earth-616)  |Clinton Francis Barton
|14 |MALE |null |May-65 |1965|50
|Full YES |YES |null |null |null |null |null |null |Dies in
exploding Kree ship in Issue 502. Brought back by Scarlet Witch. Dies again in House
of M Vol 1 Issue 7. Is late
|
|http://marvel.wikia.com/Pietro_Maximoff_(Earth-616)|Pietro Maximoff
|76 |MALE |null |May-65 |1965|50
|Full null |null |null |null |null |null |null |null |Dies in
House of M Vol 1 Issue ack
|
|http://marvel.wikia.com/Wanda_Maximoff_(Earth-616) |Wanda Maximoff
|12 |FEMALE|null |May-65 |1965|50
|Full null |null |null |null |null |null |null |null |Dies in
Uncanny_Avengers_Vol_1_back
|
|http://marvel.wikia.com/Jacques_Duquesne_(Earth-616)|Jacques Duquesne
|11 |MALE |null |Sep-65 |1965|50
|Full null |null |null |null |null |null |null |null |Dies in
Avengers_Vol_1_130. Bro Chaos King

```

http://marvel.wikia.com/Hercules_(Earth-616)							Heracles		
74	MALE	null			Oct-67		1967	48	
Full	null	null	null	null	null	null	null	NA	
http://marvel.wikia.com/T%27Challa_(Earth-616)							T'Challa		
78	MALE	null			May-68		1968	47	
Full	null	null	null	null	null	null	null	NA	
http://marvel.wikia.com/Vision_(Earth-616)							Victor Shade (alias)		
10	MALE	null			Nov-68		1968	47	
Full	null	null	null	null	null	null	null	Dies in Avengers_Vol_1_500. Is lt.	
http://marvel.wikia.com/Dane_Whitman							Dane Whitman		
48	MALE	null			Dec-69		1969	46	
Full	null	null	null	null	null	null	null	NA	
http://marvel.wikia.com/Natalia_Romanova_(Earth-616)# Romanova							Natalia Alianovna		
	11	FEMALE	null		May-73				
1973	42			Full	null	null	null	null	
null	null		Killed by The Hand. Later revive of the Chaste.						

only showing top 16 rows

```
scala> df.select("field1", "field2").show([everis@bigdata-srv ~]$ spark-shell
22/04/14 02:50:19 WARN NativeCodeLoader: Unable to load native-hadoop library
for yourtform... using builtin-java classes where applicable
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use
setLogLevel(newLevel)
Spark context Web UI available at http://bigdata-srv:4040
Spark context available as 'sc' (master = local[*], app id = local-1649911882147).
Spark session available as 'spark'.
Welcome to
```



version 2.4.7

```
Using Scala version 2.11.12 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_202)
Type in expressions to have them evaluated.
Type :help for more information.
```

```
scala> val df = spark.read.format("csv").option("sep", ",").option("header",
"true").load("csv")
df: org.apache.spark.sql.DataFrame = [URL: string, Name/Alias: string ... 19 more
fields]
```

```
scala> df.groupBy("age").count().show()
org.apache.spark.sql.AnalysisException: Cannot resolve column name "age" among
(URL, Name/Apearances, Current?, Gender, Probationary Introl, Full/Reserve Avengers
Intro, Year, Years ining, Honorary, Death1, Return1, Death2, Return2, Death3,
Return3, Death4, Return4, Death55, Notes);
    at org.apache.spark.sql.Dataset$$anonfun$resolve$1.apply(Dataset.scala:223)
    at org.apache.spark.sql.Dataset$$anonfun$resolve$1.apply(Dataset.scala:223)
    at scala.Option.getOrElse(Option.scala:121)
    at org.apache.spark.sql.Dataset.resolve(Dataset.scala:222)
    at org.apache.spark.sql.Dataset$$anonfun$groupBy$2.apply(Dataset.scala:1628)
    at org.apache.spark.sql.Dataset$$anonfun$groupBy$2.apply(Dataset.scala:1628)
    at
scala.collection.TraversableLike$$anonfun$map$1.apply(TraversableLike.scala:234)
    at
scala.collection.TraversableLike$$anonfun$map$1.apply(TraversableLike.scala:234)
    at
scala.collection.mutable.ResizableArray$class.foreach(ResizableArray.scala:59)
    at scala.collection.mutable.ArrayBuffer.foreach(ArrayBuffer.scala:48)
    at scala.collection.TraversableLike$class.map(TraversableLike.scala:234)
    at scala.collection.AbstractTraversable.map(Traversable.scala:104)
    at org.apache.spark.sql.Dataset.groupBy(Dataset.scala:1628)
    ... 49 elided
```

```
scala> df.groupBy("Honorary").count().show()
[Stage 2:> (0 + 1) /
1]22/04/14 02tClientFactory: DNS resolution for
bigdata-srv/192.xxx.x.xxx:x:33085 took 16060 ms
+-----+-----+
| Honorary|count|
+-----+-----+
| Honorary|    16|
| Probationary|    2|
|      Full|   138|
| Academy|    17|
+-----+-----+
```

```
scala> df.groupBy("alias").count().show()
org.apache.spark.sql.AnalysisException: Cannot resolve column name "alias" among
(URL, Nameurrent?, Gender, Probationary Introl, Full/Reserve Avengers Intro, Year,
Years since joinineturn1, Death2, Return2, Death3, Return3, Death4, Return4, Death5,
Return5, Notes);
    at org.apache.spark.sql.Dataset$$anonfun$resolve$1.apply(Dataset.scala:223)
    at org.apache.spark.sql.Dataset$$anonfun$resolve$1.apply(Dataset.scala:223)
    at scala.Option.getOrElse(Option.scala:121)
    at org.apache.spark.sql.Dataset.resolve(Dataset.scala:222)
    at org.apache.spark.sql.Dataset$$anonfun$groupBy$2.apply(Dataset.scala:1628)
    at org.apache.spark.sql.Dataset$$anonfun$groupBy$2.apply(Dataset.scala:1628)
    at
scala.collection.TraversableLike$$anonfun$map$1.apply(TraversableLike.scala:234)
    at
scala.collection.TraversableLike$$anonfun$map$1.apply(TraversableLike.scala:234)
    at
scala.collection.mutable.ResizableArray$class.foreach(ResizableArray.scala:59)
    at scala.collection.mutable.ArrayBuffer.foreach(ArrayBuffer.scala:48)
```

```

at scala.collection.TraversableLike$class.map(TraversableLike.scala:234)
at scala.collection.AbstractTraversable.map(Traversable.scala:104)
at org.apache.spark.sql.Dataset.groupBy(Dataset.scala:1628)
... 49 elided

```

```
scala> df.groupBy("Return2").count().show()
```

```

+-----+-----+
|Return2|count|
+-----+-----+
|    null|   157|
|     YES|     8|
|     NO |     8|
+-----+-----+

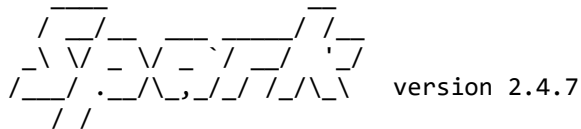
```

```
scala> [everis@bigdata-srv ~]$ spark-shell
```

```

22/04/14 03:00:01 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLeve
l(newLevel).
Spark context Web UI available at http://bigdata-srv:4040
Spark context available as 'sc' (master = local[*], app id = local-1649912471839).
Spark session available as 'spark'.
Welcome to

```



```

Using Scala version 2.11.12 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_202)
Type in expressions to have them evaluated.
Type :help for more information.

```

```
scala> val df = spark.read.format("csv").option("sep", ",").option("header",
"true").load(":///home/everis/avengers.csv")
```

```
df: org.apache.spark.sql.DataFrame = [URL: string, Name/Alias: string ... 19 more
fields]
```

```
scala> df.withColumn("Return02", col ("Return2")).show()
```

```

+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|URL|Name/Alias|Appearances|Current?|Gender|Probationary Intro|Full/Reserve Avengers
Intro|Year|Years since joining|Honorary|Death1|Return1|Death2|Return2|Death3|Return3|Death4|Return4|Dea
th5|Return5|Notes|Return02|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|http://marvel.wik...|"Henry Jonathan "...|1269|YES|MALE|

```


NO null null null null null null null null null	Sep-63 1963	52	Full	YES
Ultro... null				merged with
http://marvel.wik...	Janet van Dyne	1165	YES	FEMALE
null	Sep-63 1963	52	Full	YES
YES null null null null null null null null				Dies in Secret
In... null				
http://marvel.wik... "Anthony Edward "...		3068	YES	MALE
null	Sep-63 1963	52	Full	YES
YES null null null null null null null null				"Death: ""Later
w... null				
http://marvel.wik... Robert Bruce Banner		2089	YES	MALE
null	Sep-63 1963	52	Full	YES
YES null null null null null null null null				"Dies in Ghosts
o... null				
http://marvel.wik...	Thor Odinson	2402	YES	MALE
null	Sep-63 1963	52	Full	YES
YES YES NO null null null null null null				Dies in Fear
Itse... NO				
http://marvel.wik... Richard Milhouse ...		612	YES	MALE
null	Sep-63 1963	52	Honorary	NO
null null null null null null null null null				
NA null				
http://marvel.wik...	Steven Rogers	3458	YES	MALE
null	Mar-64 1964	51	Full	YES
YES null null null null null null null null				Dies at the end
o... null				
http://marvel.wik... Clinton Francis B...		1456	YES	MALE
null	May-65 1965	50	Full	YES
YES YES YES null null null null null null				Dies in
exploding... YES				
http://marvel.wik...	Pietro Maximoff	769	YES	MALE
null	May-65 1965	50	Full	YES
YES null null null null null null null null				Dies in House of
... null				
http://marvel.wik...	Wanda Maximoff	1214	YES	FEMALE
null	May-65 1965	50	Full	YES
YES null null null null null null null null				Dies in
Uncanny_A... null				
http://marvel.wik...	Jacques Duquesne	115	NO	MALE
null	Sep-65 1965	50	Full	YES
YES null null null null null null null null				Dies in
Avengers_... null				
http://marvel.wik...	Heracles	741	YES	MALE
null	Oct-67 1967	48	Full	NO
null null null null null null null null null				
NA null				
http://marvel.wik...	T'Challa	780	NO	MALE
null	May-68 1968	47	Full	NO
null null null null null null null null null				
NA null				
http://marvel.wik... Victor Shade (alias)		1036	YES	MALE
null	Nov-68 1968	47	Full	YES
YES null null null null null null null null				Dies in
Avengers_... null				

http://marvel.wik...	Dane Whitman	482	NO	MALE
null	Dec-69 1969	46	Full	NO
null	null	null	null	null
NA	null			
http://marvel.wik...	Natalia Alianovna...	1112	YES	FEMALE
null	May-73 1973	42	Full	YES
YES	null	null	null	Killed by The
Han...	null			
http://marvel.wik...	Brandt	160	NO	FEMALE
null	Aug-73 1973	42	Full	YES
YES	null	null	null	"Dies in
Silver_S...	null			
http://marvel.wik...	Henry P. McCoy	1886	NO	MALE
Jul-75	Sep-76 1976	39	Full	NO
null	null	null	null	null
NA	null			
http://marvel.wik...	Heather Douglas	332	NO	FEMALE
Jul-75	Sep-76 1976	39	Full	YES
YES	YES	YES	null	"Dies in
Defender...	YES			
http://marvel.wik...	Patsy Walker	557	NO	FEMALE
Nov-75	Sep-76 1976	39	Full	YES
YES	null	null	null	Died during time
...	null			

only showing top 20 rows

```
scala> df.withColumn("Return002", col ("Return02").cast("long")).show()
org.apache.spark.sql.AnalysisException: cannot resolve 'Return02' given input
columns: [Rn1, Return3, Gender, Honorary, URL, Death3, Return2, Probationary Introl,
Notes, Full/Reservengers Intro, Death4, Death5, Return5, Death2, Years since
joining, Year, Return4, Currenteath1, Appearances, Name/Alias];;
'Project [URL#10, Name/Alias#11, Appearances#12, Current?#13, Gender#14,
Probationary Intro, Full/Reserve Avengers Intro#16, Year#17, Years since joining#18,
Honorary#19, Death1#20, rn1#21, Death2#22, Return2#23, Death3#24, Return3#25,
Death4#26, Return4#27, Death5#28, Ret#29, Notes#30, cast('Return02 as bigint) AS
Return002#163]
+-
Relation[URL#10,Name/Alias#11,Appearances#12,Current?#13,Gender#14,Probationary
Introl#111/Reserve Avengers Intro#16,Year#17,Years since
joining#18,Honorary#19,Death1#20,Return1#2ath2#22,Return2#23,Death3#24,Return3#2
5,Death4#26,Return4#27,Death5#28,Return5#29,Notes#30]

at
org.apache.spark.sql.catalyst.analysis.package$AnalysisErrorAt.failAnalysis(pack
age.sc:42)
at
org.apache.spark.sql.catalyst.analysis.CheckAnalysis$$anonfun$checkAnalysis$1$$a
nonfunly$3.applyOrElse(CheckAnalysis.scala:111)
at
org.apache.spark.sql.catalyst.analysis.CheckAnalysis$$anonfun$checkAnalysis$1$$a
```

```
nonfunly$3.applyOrElse(CheckAnalysis.scala:108)
  at
org.apache.spark.sql.catalyst.trees.TreeNode$$anonfun$transformUp$1.apply(TreeNo
de.sca80)
  at
org.apache.spark.sql.catalyst.trees.TreeNode$$anonfun$transformUp$1.apply(TreeNo
de.sca80)
  at
org.apache.spark.sql.catalyst.trees.CurrentOrigin$.withOrigin(TreeNode.scala:69)
  at
org.apache.spark.sql.catalyst.trees.TreeNode.transformUp(TreeNode.scala:279)
  at
org.apache.spark.sql.catalyst.trees.TreeNode$$anonfun$3.apply(TreeNode.scala:277
)
  at
org.apache.spark.sql.catalyst.trees.TreeNode$$anonfun$3.apply(TreeNode.scala:277
)
  at
org.apache.spark.sql.catalyst.trees.TreeNode$$anonfun$4.apply(TreeNode.scala:328
)
  at
org.apache.spark.sql.catalyst.trees.TreeNode.mapProductIterator(TreeNode.scala:1
86)
  at
org.apache.spark.sql.catalyst.trees.TreeNode.mapChildren(TreeNode.scala:326)
  at
org.apache.spark.sql.catalyst.trees.TreeNode.transformUp(TreeNode.scala:277)
  at
org.apache.spark.sql.catalyst.trees.TreeNode$$anonfun$3.apply(TreeNode.scala:277
)
  at
org.apache.spark.sql.catalyst.trees.TreeNode$$anonfun$3.apply(TreeNode.scala:277
)
  at
org.apache.spark.sql.catalyst.trees.TreeNode$$anonfun$4.apply(TreeNode.scala:328
)
  at
org.apache.spark.sql.catalyst.trees.TreeNode.mapProductIterator(TreeNode.scala:1
86)
  at
org.apache.spark.sql.catalyst.trees.TreeNode.mapChildren(TreeNode.scala:326)
  at
org.apache.spark.sql.catalyst.trees.TreeNode.transformUp(TreeNode.scala:277)
  at
org.apache.spark.sql.catalyst.plans.QueryPlan$$anonfun$transformExpressionsUp$1.
apply(yPlan.scala:93)
  at
org.apache.spark.sql.catalyst.plans.QueryPlan$$anonfun$transformExpressionsUp$1.
apply(yPlan.scala:93)
  at
org.apache.spark.sql.catalyst.plans.QueryPlan$$anonfun$1.apply(QueryPlan.scala:1
05)
  at
org.apache.spark.sql.catalyst.plans.QueryPlan$$anonfun$1.apply(QueryPlan.scala:1
05)
```

```

    at
org.apache.spark.sql.catalyst.trees.CurrentOrigin$.withOrigin(TreeNode.scala:69)
    at
org.apache.spark.sql.catalyst.plans.QueryPlan.transformExpression$1(QueryPlan.sc
ala:10)
    at
org.apache.spark.sql.catalyst.plans.QueryPlan.org$apache$spark$sql$catalyst$plan
s$Quern$$recursiveTransform$1(QueryPlan.scala:116)
    at
org.apache.spark.sql.catalyst.plans.QueryPlan$$anonfun$org$apache$spark$sql$cata
lyst$p$QueryPlan$$recursiveTransform$1$2.apply(QueryPlan.scala:121)
    at
scala.collection.TraversableLike$$anonfun$map$1.apply(TraversableLike.scala:234)
    at
scala.collection.TraversableLike$$anonfun$map$1.apply(TraversableLike.scala:234)
    at scala.collection.immutable.List.foreach(List.scala:392)
    at scala.collection.TraversableLike$class.map(TraversableLike.scala:234)
    at scala.collection.immutable.List.map(List.scala:296)
    at
org.apache.spark.sql.catalyst.plans.QueryPlan.org$apache$spark$sql$catalyst$plan
s$Quern$$recursiveTransform$1(QueryPlan.scala:121)
    at
org.apache.spark.sql.catalyst.plans.QueryPlan$$anonfun$2.apply(QueryPlan.scala:1
26)
    at
org.apache.spark.sql.catalyst.trees.TreeNode.mapProductIterator(TreeNode.scala:1
86)
    at
org.apache.spark.sql.catalyst.plans.QueryPlan.mapExpressions(QueryPlan.scala:126
)
    at
org.apache.spark.sql.catalyst.plans.QueryPlan.transformExpressionsUp(QueryPlan.s
cala:9)
    at
org.apache.spark.sql.catalyst.analysis.CheckAnalysis$$anonfun$checkAnalysis$1.ap
ply(CheckAnalysis.scala:108)
    at
org.apache.spark.sql.catalyst.analysis.CheckAnalysis$$anonfun$checkAnalysis$1.ap
ply(CheckAnalysis.scala:86)
    at org.apache.spark.sql.catalyst.trees.TreeNode.foreachUp(TreeNode.scala:126)
    at
org.apache.spark.sql.catalyst.analysis.CheckAnalysis$class.checkAnalysis(CheckAn
alysis.scala:86)
    at
org.apache.spark.sql.catalyst.analysis.Analyzer.checkAnalysis(Analyzer.scala:95)
    at
org.apache.spark.sql.catalyst.analysis.Analyzer$$anonfun$executeAndCheck$1.apply
(Analyzer.scala:108)
    at
org.apache.spark.sql.catalyst.analysis.Analyzer$$anonfun$executeAndCheck$1.apply
(Analyzer.scala:105)
    at
org.apache.spark.sql.catalyst.plans.logical.AnalysisHelper$.markInAnalyzer(Analy
sisHelper.scala:201)
    at

```

```
org.apache.spark.sql.catalyst.analysis.Analyzer.executeAndCheck(Analyzer.scala:105)
    at
org.apache.spark.sql.execution.QueryExecution.analyzed$lzycompute(QueryExecution.scala
    at
org.apache.spark.sql.execution.QueryExecution.analyzed(QueryExecution.scala:56)
    at
org.apache.spark.sql.execution.QueryExecution.assertAnalyzed(QueryExecution.scala:48)
    at org.apache.spark.sql.Dataset$.ofRows(Dataset.scala:78)
    at
org.apache.spark.sql.Dataset.org$apache$spark$sql$Dataset$$withPlan(Dataset.scala:3412)
    at org.apache.spark.sql.Dataset.select(Dataset.scala:1340)
    at org.apache.spark.sql.Dataset.withColumns(Dataset.scala:2258)
    at org.apache.spark.sql.Dataset.withColumn(Dataset.scala:2225)
    ... 49 elided
```

```
[everis@bigdata-srv ~]$ vim FL_insurance_sample.csv
```


```
[everis@bigdata-srv ~]$ vim FL_insurance_sample.csv
```

000000_0	file_teste.txt	longitude	street
avengers.csv	firewall.rules	pokemon.sql	tmp
base_localidade.csv	FL_insurance_sample.csv	price	type
baths	hive-site.xml	sale_date	zip
beds	import_sqoop.sh	script_apoio	
city	install_sqoop.sh	sq_fit	
employee.txt	latitude	state	

```
22/04/14 03:13:22 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Setting default log level to "WARN".
```

```
Spark context Web UI available at http://bigdata-srv:4040
```

Spark session available as 'spark'.



version 2.4.7

Using Scala version 2.11.12 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_202)
Type in expressions to have them evaluated.

Type :help for more information.

```
scala> import org.apache.spark.sql.SparkSession
import org.apache.spark.sql.SparkSession
```

```
scala> val spark = SparkSession.builder().appName("Spark SQL basic
example").getOrCreate() e()
22/04/14 03:16:53 WARN sql.SparkSession$Builder: Using an existing SparkSession;
some spar k core configurations may not take effect.
spark: org.apache.spark.sql.SparkSession =
org.apache.spark.sql.SparkSession@6779959d
```

```
scala> val spark = SparkSession.builder().appName("Spark SQL basic
example").getOrCreate()
22/04/14 03:17:22 WARN sql.SparkSession$Builder: Using an existing SparkSession;
some spar k core configurations may not take effect.
spark: org.apache.spark.sql.SparkSession =
org.apache.spark.sql.SparkSession@6779959d
```

```
scala> val peopleDFCsv =
spark.read.format("csv").option("sep", "," ).option("header",
"true").load("file:///home/everis/FL_insurance_sample.csv")
[Stage 0:> (0 + 1) /
1]22/04/14 03:21:36 WARN client.TransportClientFactory: DNS resolution for
bigdata-srv/192.xxx.x.xxx:x.x:43515 took 16053 ms
peopleDFCsv: org.apache.spark.sql.DataFrame = [[everis@bigdata-srv ~]$ vim
FL_insurance_sample.csv: string]
```

```
scala> peopleDFCsv.withColumn("eq_site_limit",
col("eq_site_limit").cast("long")).groupBy("construction",
"county").avg("eq_site_limit").show()
org.apache.spark.sql.AnalysisException: cannot resolve '`eq_site_limit`' given
input columns: [[everis@bigdata-srv ~]$ vim FL_insurance_sample.csv];;
'Project [[everis@bigdata-srv ~]$ vim FL_insurance_sample.csv#10,
cast('eq_site_limit as bigint) AS eq_site_limit#12]
+- Relation[[everis@bigdata-srv ~]$ vim FL_insurance_sample.csv#10] csv
```

```
at
org.apache.spark.sql.catalyst.analysis.package$AnalysisErrorAt.failAnalysis(pack
age.scala:42)
at
org.apache.spark.sql.catalyst.analysis.CheckAnalysis$$anonfun$checkAnalysis$1$$a
nonfun$apply$3.applyOrElse(CheckAnalysis.scala:111)
at
org.apache.spark.sql.catalyst.analysis.CheckAnalysis$$anonfun$checkAnalysis$1$$a
nonfun$apply$3.applyOrElse(CheckAnalysis.scala:108)
at
org.apache.spark.sql.catalyst.trees.TreeNode$$anonfun$transformUp$1.apply(TreeNo
de.scala:280)
at
org.apache.spark.sql.catalyst.trees.TreeNode$$anonfun$transformUp$1.apply(TreeNo
de.scala:280)
at
org.apache.spark.sql.catalyst.trees.CurrentOrigin$.withOrigin(TreeNode.scala:69)
at
```

```
org.apache.spark.sql.catalyst.trees.TreeNode.transformUp(TreeNode.scala:279)
  at
org.apache.spark.sql.catalyst.trees.TreeNode$$anonfun$3.apply(TreeNode.scala:277)
)
  at
org.apache.spark.sql.catalyst.trees.TreeNode$$anonfun$3.apply(TreeNode.scala:277)
)
  at
org.apache.spark.sql.catalyst.trees.TreeNode$$anonfun$4.apply(TreeNode.scala:328)
)
  at
org.apache.spark.sql.catalyst.trees.TreeNode.mapProductIterator(TreeNode.scala:186)
  at
org.apache.spark.sql.catalyst.trees.TreeNode.mapChildren(TreeNode.scala:326)
  at
org.apache.spark.sql.catalyst.trees.TreeNode.transformUp(TreeNode.scala:277)
  at
org.apache.spark.sql.catalyst.trees.TreeNode$$anonfun$3.apply(TreeNode.scala:277)
)
  at
org.apache.spark.sql.catalyst.trees.TreeNode$$anonfun$3.apply(TreeNode.scala:277)
)
  at
org.apache.spark.sql.catalyst.trees.TreeNode$$anonfun$4.apply(TreeNode.scala:328)
)
  at
org.apache.spark.sql.catalyst.trees.TreeNode.mapProductIterator(TreeNode.scala:186)
  at
org.apache.spark.sql.catalyst.trees.TreeNode.mapChildren(TreeNode.scala:326)
  at
org.apache.spark.sql.catalyst.trees.TreeNode.transformUp(TreeNode.scala:277)
  at
org.apache.spark.sql.catalyst.plans.QueryPlan$$anonfun$transformExpressionsUp$1.apply(QueryPlan.scala:93)
  at
org.apache.spark.sql.catalyst.plans.QueryPlan$$anonfun$transformExpressionsUp$1.apply(QueryPlan.scala:93)
  at
org.apache.spark.sql.catalyst.plans.QueryPlan$$anonfun$1.apply(QueryPlan.scala:105)
  at
org.apache.spark.sql.catalyst.plans.QueryPlan$$anonfun$1.apply(QueryPlan.scala:105)
  at
org.apache.spark.sql.catalyst.trees.CurrentOrigin$.withOrigin(TreeNode.scala:69)
  at
org.apache.spark.sql.catalyst.plans.QueryPlan.transformExpression$1(QueryPlan.scala:104)
  at
org.apache.spark.sql.catalyst.plans.QueryPlan.org$apache$spark$sql$catalyst$plans$QueryPlan$$recursiveTransform$1(QueryPlan.scala:116)
  at
org.apache.spark.sql.catalyst.plans.QueryPlan$$anonfun$org$apache$spark$sql$cata
```

```
lyst$plans$QueryPlan$$recursiveTransform$1$2.apply(QueryPlan.scala:121)
  at
scala.collection.TraversableLike$$anonfun$map$1.apply(TraversableLike.scala:234)
  at
scala.collection.TraversableLike$$anonfun$map$1.apply(TraversableLike.scala:234)
  at scala.collection.immutable.List.foreach(List.scala:392)
  at scala.collection.TraversableLike$class.map(TraversableLike.scala:234)
  at scala.collection.immutable.List.map(List.scala:296)
  at
org.apache.spark.sql.catalyst.plans.QueryPlan.org$apache$spark$sql$catalyst$plan
s$QueryPlan$$recursiveTransform$1(QueryPlan.scala:121)
  at
org.apache.spark.sql.catalyst.plans.QueryPlan$$anonfun$2.apply(QueryPlan.scala:1
26)
  at
org.apache.spark.sql.catalyst.trees.TreeNode.mapProductIterator(TreeNode.scala:1
86)
  at
org.apache.spark.sql.catalyst.plans.QueryPlan.mapExpressions(QueryPlan.scala:126
)
  at
org.apache.spark.sql.catalyst.plans.QueryPlan.transformExpressionsUp(QueryPlan.s
cala:93)
  at
org.apache.spark.sql.catalyst.analysis.CheckAnalysis$$anonfun$checkAnalysis$1.ap
ply(CheckAnalysis.scala:108)
  at
org.apache.spark.sql.catalyst.analysis.CheckAnalysis$$anonfun$checkAnalysis$1.ap
ply(CheckAnalysis.scala:86)
  at org.apache.spark.sql.catalyst.trees.TreeNode.foreachUp(TreeNode.scala:126)
  at
org.apache.spark.sql.catalyst.analysis.CheckAnalysis$class.checkAnalysis(CheckAn
alysis.scala:86)
  at
org.apache.spark.sql.catalyst.analysis.Analyzer.checkAnalysis(Analyzer.scala:95)
  at
org.apache.spark.sql.catalyst.analysis.Analyzer$$anonfun$executeAndCheck$1.apply
(Analyzer.scala:108)
  at
org.apache.spark.sql.catalyst.analysis.Analyzer$$anonfun$executeAndCheck$1.apply
(Analyzer.scala:105)
  at
org.apache.spark.sql.catalyst.plans.logical.AnalysisHelper$.markInAnalyzer(Analy
sisHelper.scala:201)
  at
org.apache.spark.sql.catalyst.analysis.Analyzer.executeAndCheck(Analyzer.scala:1
05)
  at
org.apache.spark.sql.execution.QueryExecution.analyzed$lzycompute(QueryExecution
.scala:58)
  at
org.apache.spark.sql.execution.QueryExecution.analyzed(QueryExecution.scala:56)
  at
org.apache.spark.sql.execution.QueryExecution.assertAnalyzed(QueryExecution.scal
a:48)
```



```

    at org.apache.spark.sql.Dataset$.ofRows(Dataset.scala:78)
    at
org.apache.spark.sql.Dataset.org$apache$spark$sql$Dataset$$withPlan(Dataset.scala:3412)
    at org.apache.spark.sql.Dataset.select(Dataset.scala:1340)
    at org.apache.spark.sql.Dataset.withColumns(Dataset.scala:2258)
    at org.apache.spark.sql.Dataset.withColumn(Dataset.scala:2225)
    ... 49 elided

```

```

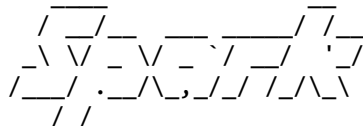
scala> peopleDFCsv.withColumn("eq_site_limit",
col("eq_site_limit").groupBy("construction",).show()

```

```

[15]+  Stopped                  spark-shell
[everis@bigdata-srv ~]$ spark-shell
22/04/14 03:28:54 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use
setLogLevel(newLevel)
22/04/14 03:30:19 WARN util.Utills: Service 'SparkUI' could not bind on port 4040.
Attemptinport 4041.
22/04/14 03:30:19 WARN util.Utills: Service 'SparkUI' could not bind on port 4041.
Attemptinport 4042.
22/04/14 03:30:19 WARN util.Utills: Service 'SparkUI' could not bind on port 4042.
Attemptinport 4043.
22/04/14 03:30:19 WARN util.Utills: Service 'SparkUI' could not bind on port 4043.
Attemptinport 4044.
Spark context Web UI available at http://bigdata-srv:4044
Spark context available as 'sc' (master = local[*], app id = local-1649914219564).
Spark session available as 'spark'.
Welcome to

```



version 2.4.7

Using Scala version 2.11.12 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_202)
Type in expressions to have them evaluated.
Type :help for more information.

```

scala> val df = spark.read.format("csv").option("sep", "," ).option("header",
"true").load("le:///home/everis/avengers.csv")
[Stage 0:> (0 + 1) /
1]22/04/14 03:TransportClientFactory: DNS resolution for
bigdata-srv/192.xxx.x.xxx:x:37334 took 16057 ms
df: org.apache.spark.sql.DataFrame = [URL: string, Name/Alias: string ... 19 more
fields]

```

```

scala> df.createOrReplaceTempView("av")

```

```

scala> df.printSchema
root
|-- URL: string (nullable = true)
|-- Name/Alias: string (nullable = true)

```

```

|-- Appearances: string (nullable = true)
|-- Current?: string (nullable = true)
|-- Gender: string (nullable = true)
|-- Probationary Introl: string (nullable = true)
|-- Full/Reserve Avengers Intro: string (nullable = true)
|-- Year: string (nullable = true)
|-- Years since joining: string (nullable = true)
|-- Honorary: string (nullable = true)
|-- Death1: string (nullable = true)
|-- Return1: string (nullable = true)
|-- Death2: string (nullable = true)
|-- Return2: string (nullable = true)
|-- Death3: string (nullable = true)
|-- Return3: string (nullable = true)
|-- Death4: string (nullable = true)
|-- Return4: string (nullable = true)
|-- Death5: string (nullable = true)
|-- Return5: string (nullable = true)
|-- Notes: string (nullable = true)

```

```

scala> spark.sql("SELECT Appearances FROM av WHERE URL LIKE '%IRON%'").show(10,
false)
22/04/14 03:35:04 WARN metastore.ObjectStore: Version information not found in
metastore. hn is not enabled so recording the schema version 1.2.0
22/04/14 03:35:04 WARN metastore.ObjectStore: Failed to get database default,
returning NoS
22/04/14 03:36:02 WARN shortcircuit.DomainSocketFactory: The short-circuit local
reads featdoop cannot be loaded.
22/04/14 03:36:02 WARN metastore.ObjectStore: Failed to get database global_temp,
returning
[Stage 1:> (0 + 1) /
1]22/04/14 03:36:02 WARN ntFactory: DNS resolution for bigdata-srv/192.xxx.x.xxx:x:37334 took
16505 ms
+-----+
|Appearances|
+-----+
+-----+

```

```


scala> spark.sql("SELECT Appearances FROM av").show(10, false)
+-----+
|Appearances|
+-----+
|1269
|1165
|3068
|2089
|2402
|612
|3458
|1456
|769
|1214
+-----+

```

```

[everis@bigdata-srv ~]$ spark-shell
22/04/14 03:39:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use
setLogLevel(newLevel)
[everis@bigdata-srv ~]$ spark-shell
22/04/14 03:41:40 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLeve
l(newLevel).
[everis@bigdata-srv ~]$ spark-shell
22/04/14 03:43:20 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for y  our platform... using builtin-java classes where applicable
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use
setLogLevel(newLev el).
22/04/14 03:44:35 WARN util.Utills: Service 'SparkUI' could not bind on port 4040.
Attemp ting port 4041.
22/04/14 03:44:35 WARN util.Utills: Service 'SparkUI' could not bind on port 4041.
Attemp ting port 4042.
22/04/14 03:44:35 WARN util.Utills: Service 'SparkUI' could not bind on port 4042.
Attemp ting port 4043.
22/04/14 03:44:35 WARN util.Utills: Service 'SparkUI' could not bind on port 4043.
Attemp ting port 4044.
Spark context Web UI available at http://bigdata-srv:4044
Spark context available as 'sc' (master = local[*], app id = local-1649915075571).
Spark session available as 'spark'.
Welcome to

```



```
Using Scala version 2.11.12 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_202)
Type in expressions to have them evaluated.
Type :help for more information.
```

```
scala> val df = spark.read.format("csv").option("sep", ",").option("header",
"true").load("file:///home/everis/avengers.csv")
df: org.apache.spark.sql.DataFrame = [URL: string, Name/Alias: string ... 19 more
fields]
```

```
scala> df.createOrReplaceTempView("av")
```

```
scala> df.printSchema
```

```
root
|-- URL: string (nullable = true)
|-- Name/Alias: string (nullable = true)
|-- Appearances: string (nullable = true)
```

```
scala> spark.sql("SELECT Appearances FROM av WHERE URL =  
'http://marvel.wikia.com/Anthon y Stark_(Earth-616)']").show(10, false)  
22/04/14 03:45:59 WARN shortcircuit.DomainSocketFactory: The short-circuit local  
reads feature cannot be used because libhadoop cannot be loaded.  
22/04/14 03:46:08 WARN metastore.ObjectStore: Failed to get database global_temp,  
returning NoSuchObjectException
```

```
scala> df.show(10, false)
```

URL	Name/Alias
Appearances	Current?
Gender	Probationary
Intro1	Full/Reserve
Avengers	Intro
Year	Years since
joining	Honorary
Death1	Return1
Death2	Return2
Death3	Return3
Death4	Return4
Death5	Return5
Notes	

|null |null |null |null |Merged with Ultron in Rage of Ultron Vol. 1. A funeral was held.

|
|http://marvel.wikia.com/Janet_van_Dyne_(Earth-616) |Janet van Dyne
|1165 |YES |FEMALE|null |Sep-63
|1963|52 |Full |YES |YES |null |null |null |null
|null |null |null |null |Dies in Secret Invasion V1:I8. Actually was sent tto Microverse later recovered

|
|http://marvel.wikia.com/Anthony_Stark_(Earth-616) |"Anthony Edward ""Tony"" Stark"|3068 |YES |MALE |null |Sep-63
|1963|52 |Full |YES |YES |null |null |null |null
|null |null |null |null |"Death: ""Later while under the influence of Immortus Stark committed a number of horrible acts and was killed.'" This set up young Tony. Franklin Richards later brought him back"|
|http://marvel.wikia.com/Robert_Bruce_Banner_(Earth-616)|Robert Bruce Banner
|2089 |YES |MALE |null |Sep-63
|1963|52 |Full |YES |YES |null |null |null |null
|null |null |null |null |"Dies in Ghosts of the Future arc. However ""he had actually used a hidden Pantheon base to survive"""

|
|http://marvel.wikia.com/Thor_Odinson_(Earth-616) |Thor Odinson
|2402 |YES |MALE |null |Sep-63
|1963|52 |Full |YES |YES |YES |NO |null |null
|null |null |null |null |Dies in Fear Itself brought back because that's kind of the whole point. Second death in Time Runs Out has not yet returned

|
|http://marvel.wikia.com/Richard_Jones_(Earth-616) |Richard Milhouse Jones
|612 |YES |MALE |null |Sep-63
|1963|52 |Honorary|NO |null |null |null |null |null
|null |null |null |null |NA

|
|http://marvel.wikia.com/Steven_Rogers_(Earth-616) |Steven Rogers
|3458 |YES |MALE |null |Mar-64
|1964|51 |Full |YES |YES |null |null |null |null
|null |null |null |null |Dies at the end of Civil War. Later comes back.

|
|http://marvel.wikia.com/Clint_Barton_(Earth-616) |Clinton Francis Barton
|1456 |YES |MALE |null |May-65
|1965|50 |Full |YES |YES |YES |YES |null |null
|null |null |null |null |Dies in exploding Kree ship in Averngers Vol. 1 Issue 502. Brought back by Scarlet Witch. Dies again in House of M Vol 1 Issue 7. Is later brought back.

|
|http://marvel.wikia.com/Pietro_Maximoff_(Earth-616) |Pietro Maximoff
|769 |YES |MALE |null |May-65
|1965|50 |Full |YES |YES |null |null |null |null
|null |null |null |null |Dies in House of M Vol 1 Issue 7. Later comes back

|
|http://marvel.wikia.com/Wanda_Maximoff_(Earth-616) |Wanda Maximoff
|1214 |YES |FEMALE|null |May-65
|1965|50 |Full |YES |YES |null |null |null |null
|null |null |null |null |Dies in Uncanny_Avengers_Vol_1_14. Later comes back

|
+-----+
-----+-----+-----+-----+-----+-----+
|

only showing top 10 rows

```
scala>
```