

**EXP NO: 4****CREATE UDF IN PIG****\$start-all.sh**

```

(hadoop@kali)~$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [kali]
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
2024-09-11 04:59:16,429 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Starting resourcemanager
Starting nodemanagers

```

**\$ jps**

```

(hadoop@kali)~$ jps
14436 NodeManager
16772 Jps
13830 SecondaryNameNode
14311 ResourceManager
13597 DataNode
13471 NameNode

```

**\$wget <https://d1cdn.apache.org/pig/pig-0.16.0/pig-0.16.0.tar.gz>****\$ tar xvfz pig-0.16.0.tar.gz**

```

kali-linux-2023.4-vmware-amd64 - VMware Workstation 17 Player (Non-commercial use only)
Player
(hadoop@kali)~$ wget https://d1cdn.apache.org/pig/pig-0.16.0/pig-0.16.0.tar.gz
--2024-08-29 10:55:33-- https://d1cdn.apache.org/pig/pig-0.16.0/pig-0.16.0.tar.gz
Resolving d1cdn.apache.org (d1cdn.apache.org) ... 208.146.62.844, 151.101.2.132
Connecting to d1cdn.apache.org (d1cdn.apache.org)[208.146.62.844]:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 177279333 (169M) [application/x-gzip]
Saving to: 'pig-0.16.0.tar.gz'

pig-0.16.0.tar.gz 100%[=====] 169.07M 26.5MB/s in 6.5s

2024-08-29 10:55:41 (26.0 MB/s) - 'pig-0.16.0.tar.gz' saved [177279333/177279333]

(hadoop@kali)~$ tar xvfz pig-0.16.0.tar.gz
pig-0.16.0/
pig-0.16.0/bin/
pig-0.16.0/conf/
pig-0.16.0/contrib/
pig-0.16.0/contrib/piggybank/
pig-0.16.0/contrib/piggybank/java/
pig-0.16.0/contrib/piggybank/java/build/
pig-0.16.0/contrib/piggybank/java/build/classes/org/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/datetime/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/datetime/convert/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/datetime/diff/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/datetime/truncate/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/decode/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/math/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/string/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/util/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/evaluation/util/apacheLogParser/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/storage/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/storage/allloader/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/storage/apachelog/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/storage/avro/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/storage/hiverc/
pig-0.16.0/contrib/piggybank/java/build/classes/org/apache/pig/piggybank/storage/partition/
pig-0.16.0/contrib/piggybank/java/build/docs/
pig-0.16.0/contrib/piggybank/java/build/docs/api/
pig-0.16.0/contrib/piggybank/java/build/test/
pig-0.16.0/contrib/piggybank/java/lib/
pig-0.16.0/contrib/piggybank/java/src/

```

**\$nano ~/.bashrc**

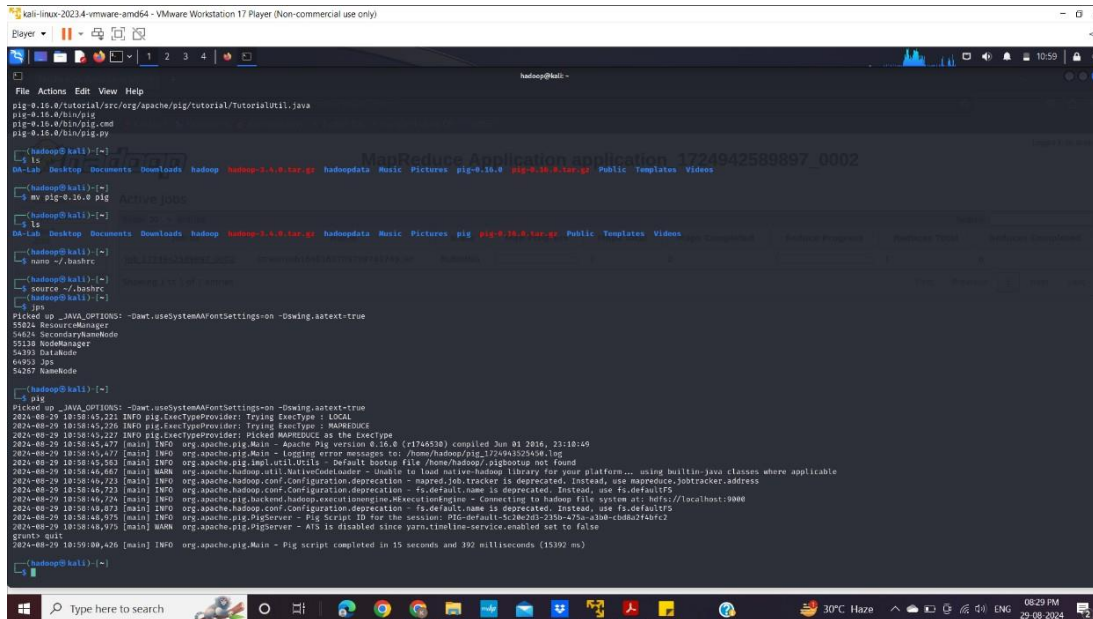
```

#PIG settings
export PIG_HOME=/home/hadoop/pig
export PATH=$PATH:$PIG_HOME/bin
export PIG_CLASSPATH=$PIG_HOME/conf:$HADOOP_HOME/etc/hadoop/
export PIG_CONF_DIR=$PIG_HOME/conf
#export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export PIG_CLASSPATH=$PIG_CONF_DIR:$PATH
#PIG setting ends

```

**\$mv pig-0.16.0 pig**

**\$pig**



```
kali-linux-2023-4-vmware-amd64 - VMware Workstation 17 Player (Non-commercial use only)
hadoop@kali: ~
File Actions Edit View Help
pig-0.16.0/tutorial/src/org/apache/pig/tutorial/tutorialUtil.java
pig-0.16.0/bin/pig
pig-0.16.0/bin/pig.cmd
pig-0.16.0/bin/pig.py

(hadoop@kali)~$ mv pig-0.16.0 pig
(hadoop@kali)~$ ls
DA-Lab Desktop Documents Downloads hadoop hadoop-0.4.0.tar.gz hadoopdata Music Pictures pig pig-0.16.0.tar.gz Public Templates Videos
(hadoop@kali)~$ name ~/bashrc
(hadoop@kali)~$ source ~/.bashrc
(hadoop@kali)~$ java
Picked up _JAVA_OPTIONS: -Dwt.useSystemAWTontSettings-on -Dswing.aatext=true
55821 ResourceManager
54824 SecondaryNameNode
55133 NodeManager
54393 DataNode
64953 JPS
54267 NameNode
(hadoop@kali)~$ pig
Picked up _JAVA_OPTIONS: -Dwt.useSystemAWTontSettings-on -Dswing.aatext=true
2024-08-29 10:58:45.222 INFO pig.ExecTypeProvider: Trying ExecType: LOCAL
2024-08-29 10:58:45.226 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2024-08-29 10:58:45.477 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r1746530) compiled Jun 01 2016, 23:18:49
2024-08-29 10:58:45.563 [main] INFO org.apache.pig.impl.util.Utils - Default bootstrap file /home/hadoop/pig/bootstrap not found
2024-08-29 10:58:45.667 [main] WARN org.apache.hadoop.util.NativeCodeLoader - Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2024-08-29 10:58:46.723 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2024-08-29 10:58:46.724 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-29 10:58:46.724 [main] INFO org.apache.pig.backend.hadoop.executionengine.MRExecutionEngine - Connecting to hadoop file system at: hdfs://localhost:9000
2024-08-29 10:58:46.874 [main] INFO org.apache.pig.PigServer - Pig script ID for the session: PIG-default-5c28e283-2330-475a-a308-cdb8a274bfc3
2024-08-29 10:58:46.875 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
grunt> quit
2024-08-29 10:59:00.420 [main] INFO org.apache.pig.Main - Pig script completed in 15 seconds and 392 milliseconds (15392 ms)
(hadoop@kali)~$
```

**\$cd DA-Lab**

**\$mkdir exp4**

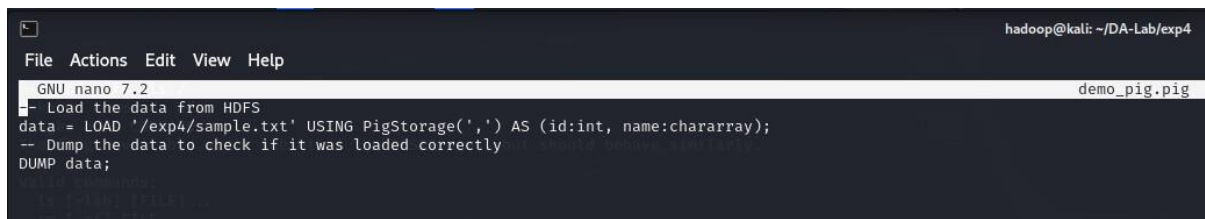
**\$cd exp4**

**\$nano sample.txt**



```
hadoop@kali: ~/DA-Lab/exp4
File Actions Edit View Help
GNU nano 7.2 sample.txt
1, John
2, Jane
3, Joe
4, Emma
```

**\$nano demo\_pig.pig**



```
hadoop@kali: ~/DA-Lab/exp4
File Actions Edit View Help
GNU nano 7.2 demo_pig.pig
-- Load the data from HDFS
data = LOAD '/exp4/sample.txt' USING PigStorage(',') AS (id:int, name:chararray);
-- Dump the data to check if it was loaded correctly
DUMP data;
```

**\$hdfs dfs -mkdir /exp4**

**\$hdfs dfs -copyFromLocal ~/DA-Lab/exp4/sample.txt /exp4**

**\$pig demo\_pig.pig**

[illegible]

## \$nano uppercase\_udf.py

```
hadoop@kali: ~/DA-Lab/exp4
File Actions Edit View Help
GNU nano 7.2
def uppercase(text):
    return text.upper()
if __name__ == "__main__":
    import sys
    for line in sys.stdin:
        line = line.strip()
        result = uppercase(line)
        print(result)
```

uppercase\_udf.py



**\$hdfs dfs -copyFromLocal ~/DA-Lab/exp4/uppercase\_udf.py /exp4**

```

(hadoop@kali)-[~/hadoop/bin]
$ ./hdfs dfs -ls /exp4
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
2024-09-21 00:26:01,736 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
Found 3 items
drwxr-xr-x  - hadoop supergroup          0 2024-08-30 05:07 /exp4/output
-rw-r--r--  1 hadoop supergroup        27 2024-08-30 04:43 /exp4/sample.txt
-rw-r--r--  1 hadoop supergroup       172 2024-08-30 05:02 /exp4/uppercase_udf.py

```

**\$nano udf\_example.pig**

```

GNU nano 7.2 udf_example.pig
-- Register the Python UDF script
REGISTER 'hdfs:///exp4/uppercase_udf.py' USING jython AS udf;
-- Load some data
data = LOAD 'hdfs:///exp4/sample.txt' AS (text:chararray);
-- Use the Python UDF
uppercased_data = FOREACH data GENERATE udf.uppercase(text) AS uppercase_text;
-- Store the result
STORE uppercased_data INTO 'hdfs:///exp4/output';

```

**\$pig -f udf\_example.pig**

```

kali-linux-2023.4-vmware-amd64 - VMware Workstation 17 Player (Non-commercial use only)
Player
(hadoop@kali)-[~/DA-Lab/exp4]
$ nano uppercase_udf.py
(hadoop@kali)-[~/DA-Lab/exp4]
$ cd ~/hadoop/bin
(hadoop@kali)-[~/hadoop/bin]
$ ./hdfs dfs -put ~/DA-Lab/exp4/uppercase_udf.py /exp4
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
2024-08-30 05:07:54,341 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
(hadoop@kali)-[~/hadoop/bin]
$ cd ~/hadoop/exp4
(hadoop@kali)-[~/DA-Lab/exp4]
$ nano udf_example.pig
(hadoop@kali)-[~/DA-Lab/exp4]
$ pig -f udf_example.pig
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
2024-08-30 05:08:18,591 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2024-08-30 05:08:18,601 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
2024-08-30 05:08:18,601 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2024-08-30 05:08:18,836 [main] INFO org.apache.pig.Main - Apache Pig version 0.16.0 (r174538) compiled Jun 81 2016, 23:10:49
2024-08-30 05:08:18,836 [main] INFO org.apache.pig.Main - Logging error messages to: /home/hadoop/DA-Lab/exp4/pig17250687/main.log
2024-08-30 05:08:19,685 [main] WARN org.apache.hadoop.util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2024-08-30 05:08:20,127 [main] INFO org.apache.pig.impl.util.Utils - Default bootstrap file /home/hadoop/pigbootstrap not found
2024-08-30 05:08:20,282 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2024-08-30 05:08:20,283 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 05:08:20,383 [main] INFO org.apache.pig.backend.hadoop.executionengine.MExecutionEngine - Connecting to Hadoop file system at: hdfs://localhost:9000
2024-08-30 05:08:22,438 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 05:08:22,513 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-udf_example.pig-1d0c8ccc-3e11-47e3-a9e8-561b5b6b38a1
2024-08-30 05:08:22,513 [main] WARN org.apache.pig.PigServer - AIO is disabled since yarn.timeline-service.enabled set to false
2024-08-30 05:08:22,725 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 05:08:24,298 [main] INFO org.apache.pig.scripting.jython.JythonScriptEngine - created tmp python.cachedir=/tmp/pig_jython.3781812655894475147
2024-08-30 05:08:26,844 [main] INFO org.apache.pig.scripting.jython.JythonScriptEngine - pig.cmd.args.reminders is empty. This is not expected unless on testing.
2024-08-30 05:08:27,274 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 05:08:27,867 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 05:08:28,407 [main] INFO org.apache.pig.scripting.jython.JythonFunction - No schema defined for function 'uppercase' in /tmp/pig2151342629392931085tmp/uppercase_udf.py
2024-08-30 05:08:28,583 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 05:08:38,743 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.textoutputformat.separator is deprecated. Instead, use mapreduce.output.textoutputformat.separator
2024-08-30 05:08:38,838 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNKNOWN
2024-08-30 05:08:38,858 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2024-08-30 05:08:38,982 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set.... will not generate code.
2024-08-30 05:08:39,132 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - [RULES_ENABLED:[AddFileCache, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushdownForEachFlatten, PushupFilter, SplitFilter, StreamTypeCastInserter]]
2024-08-30 05:08:39,403 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (PS Old Gen) of size 699138848 to monitor. collectionUsageThreshold = 489396640, usageThreshold = 489396640
2024-08-30 05:08:39,484 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MRCompiler - File concatenation threshold: 384 optimizer: False
2024-08-30 05:08:39,858 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2024-08-30 05:08:39,859 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapreduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2024-08-30 05:08:39,904 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

```

```
kali-linux-2023.4-vmware-amd64 - VMware Workstation 15 Player (Non-commercial use only)
Player
[Icons] [Windows Taskbar] [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18] [19] [20] [21] [22] [23] [24] [25] [26] [27] [28] [29] [30] [31] [32] [33] [34] [35] [36] [37] [38] [39] [40] [41] [42] [43] [44] [45] [46] [47] [48] [49] [50] [51] [52] [53] [54] [55] [56] [57] [58] [59] [60] [61] [62] [63] [64] [65] [66] [67] [68] [69] [70] [71] [72] [73] [74] [75] [76] [77] [78] [79] [80] [81] [82] [83] [84] [85] [86] [87] [88] [89] [90] [91] [92] [93] [94] [95] [96] [97] [98] [99] [100] [101] [102] [103] [104] [105] [106] [107] [108] [109] [110] [111] [112] [113] [114] [115] [116] [117] [118] [119] [120] [121] [122] [123] [124] [125] [126] [127] [128] [129] [130] [131] [132] [133] [134] [135] [136] [137] [138] [139] [140] [141] [142] [143] [144] [145] [146] [147] [148] [149] [150] [151] [152] [153] [154] [155] [156] [157] [158] [159] [160] [161] [162] [163] [164] [165] [166] [167] [168] [169] [170] [171] [172] [173] [174] [175] [176] [177] [178] [179] [180] [181] [182] [183] [184] [185] [186] [187] [188] [189] [190] [191] [192] [193] [194] [195] [196] [197] [198] [199] [200] [201] [202] [203] [204] [205] [206] [207] [208] [209] [210] [211] [212] [213] [214] [215] [216] [217] [218] [219] [220] [221] [222] [223] [224] [225] [226] [227] [228] [229] [230] [231] [232] [233] [234] [235] [236] [237] [238] [239] [240] [241] [242] [243] [244] [245] [246] [247] [248] [249] [250] [251] [252] [253] [254] [255] [256] [257] [258] [259] [260] [261] [262] [263] [264] [265] [266] [267] [268] [269] [270] [271] [272] [273] [274] [275] [276] [277] [278] [279] [280] [281] [282] [283] [284] [285] [286] [287] [288] [289] [290] [291] [292] [293] [294] [295] [296] [297] [298] [299] [300] [301] [302] [303] [304] [305] [306] [307] [308] [309] [310] [311] [312] [313] [314] [315] [316] [317] [318] [319] [320] [321] [322] [323] [324] [325] [326] [327] [328] [329] [330] [331] [332] [333] [334] [335] [336] [337] [338] [339] [340] [341] [342] [343] [344] [345] [346] [347] [348] [349] [350] [351] [352] [353] [354] [355] [356] [357] [358] [359] [360] [361] [362] [363] [364] [365] [366] [367] [368] [369] [370] [371] [372] [373] [374] [375] [376] [377] [378] [379] [380] [381] [382] [383] [384] [385] [386] [387] [388] [389] [390] [391] [392] [393] [394] [395] [396] [397] [398] [399] [400] [401] [402] [403] [404] [405] [406] [407] [408] [409] [410] [411] [412] [413] [414] [415] [416] [417] [418] [419] [420] [421] [422] [423] [424] [425] [426] [427] [428] [429] [430] [431] [432] [433] [434] [435] [436] [437] [438] [439] [440] [441] [442] [443] [444] [445] [446] [447] [448] [449] [450] [451] [452] [453] [454] [455] [456] [457] [458] [459] [460] [461] [462] [463] [464] [465] [466] [467] [468] [469] [470] [471] [472] [473] [474] [475] [476] [477] [478] [479] [480] [481] [482] [483] [484] [485] [486] [487] [488] [489] [490] [491] [492] [493] [494] [495] [496] [497] [498] [499] [500] [501] [502] [503] [504] [505] [506] [507] [508] [509] [510] [511] [512] [513] [514] [515] [516] [517] [518] [519] [520] [521] [522] [523] [524] [525] [526] [527] [528] [529] [530] [531] [532] [533] [534] [535] [536] [537] [538] [539] [540] [541] [542] [543] [544] [545] [546] [547] [548] [549] [550] [551] [552] [553] [554] [555] [556] [557] [558] [559] [560] [561] [562] [563] [564] [565] [566] [567] [568] [569] [570] [571] [572] [573] [574] [575] [576] [577] [578] [579] [580] [581] [582] [583] [584] [585] [586] [587] [588] [589] [590] [591] [592] [593] [594] [595] [596] [597] [598] [599] [600] [601] [602] [603] [604] [605] [606] [607] [608] [609] [610] [611] [612] [613] [614] [615] [616] [617] [618] [619] [620] [621] [622] [623] [624] [625] [626] [627] [628] [629] [630] [631] [632] [633] [634] [635] [636] [637] [638] [639] [640] [641] [642] [643] [644] [645] [646] [647] [648] [649] [650] [651] [652] [653] [654] [655] [656] [657] [658] [659] [660] [661] [662] [663] [664] [665] [666] [667] [668] [669] [670] [671] [672] [673] [674] [675] [676] [677] [678] [679] [680] [681] [682] [683] [684] [685] [686] [687] [688] [689] [690] [691] [692] [693] [694] [695] [696] [697] [698] [699] [700] [701] [702] [703] [704] [705] [706] [707] [708] [709] [710] [711] [712] [713] [714] [715] [716] [717] [718] [719] [720] [721] [722] [723] [724] [725] [726] [727] [728] [729] [730] [731] [732] [733] [734] [735] [736] [737] [738] [739] [740] [741] [742] [743] [744] [745] [746] [747] [748] [749] [750] [751] [752] [753] [754] [755] [756] [757] [758] [759] [760] [761] [762] [763] [764] [765] [766] [767] [768] [769] [770] [771] [772] [773] [774] [775] [776] [777] [778] [779] [780] [781] [782] [783] [784] [785] [786] [787] [788] [789] [790] [791] [792] [793] [794] [795] [796] [797] [798] [799] [800] [801] [802] [803] [804] [805] [806] [807] [808] [809] [810] [811] [812] [813] [814] [815] [816] [817] [818] [819] [820] [821] [822] [823] [824] [825] [826] [82
```

## \$hdfs dfs -cat /exp4/output/\*

```

(hadoop@kali)-[~/hadoop/bin]
$ ./hdfs dfs -cat /exp4/output/*
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
2024-09-21 00:33:32,731 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
1,JOHN
2,JANE
3,JOE
4,EMMA

```