

ej\_1

June 15, 2021

```
[25]: import findspark
findspark.init()
findspark.find()
import pyspark
findspark.find()
from pyspark import SparkContext, SparkConf
from pyspark.sql import SparkSession
from pyspark.sql.functions import col,sum,avg,max

conf = pyspark.SparkConf().setAppName('appName').setMaster('local')
#sc = pyspark.SparkContext(conf=conf)
spark = SparkSession(sc)
csv_df = spark.read.option("header", "True").csv("./web.csv")
csv_df.printSchema()
sc = SparkContext.getOrCreate();
```

```
root
|-- IP: string (nullable = true)
|-- Time: string (nullable = true)
|-- URL: string (nullable = true)
|-- Staus: string (nullable = true)
```

```
[50]: import pyspark.sql.functions as F
# Grouping responses by code
csv_df.groupBy('Staus').count().where(F.length("Staus")==3).show()
```

```
+-----+-----+
|Staus|count|
+-----+-----+
| 200|11330|
| 206|   52|
| 302| 3498|
| 404|   251|
| 304|   658|
+-----+-----+
```

```
[70]: from pyspark.sql.functions import col,sum,avg,max
from pyspark.sql.types import BooleanType
import ipaddress
@F.udf(returnType=BooleanType())
def is_ip(param_ip):
    try:
        ipaddress.ip_address(param_ip)
        return True
    except:
        return False

# Grouping by host IP and number of requests.
csv_df.groupBy('IP').count().where((col("count")>10) &
↳(is_ip(col("IP"))==True)).show()
```

```
+-----+-----+
|      IP|count|
+-----+-----+
|10.131.2.1| 1626|
|10.128.2.1| 4257|
|10.130.2.1| 4056|
|10.131.0.1| 4198|
|10.129.2.1| 1652|
+-----+-----+
```

```
[106]: from pyspark.sql.functions import split, explode
@F.udf(returnType=BooleanType())
def is_command(param_comm):
    try:
        if param_comm in ['POST', 'GET', 'HEAD']:
            return True
        else:
            return False
    except:
        return False

# Grouping by command and number of requests.
split_URL = pyspark.sql.functions.split(csv_df['URL'], ' ')
url_df = csv_df.withColumn('URL_COMMAND', split_URL.getItem(0)).
↳groupBy('URL_COMMAND').count().where(is_command(col("URL_COMMAND"))).
↳show(100)
```

```
+-----+-----+
|URL_COMMAND|count|
+-----+-----+
|      POST|  682|
|      HEAD|    9|
```

```
|          GET|15098|
+-----+-----+
```

```
[144]: # Grouping by URI and number of requests

# Drop unnecessary columns
filtered_URL_status = csv_df.drop('IP').drop('Time')

# Select by Status OK
filtered_200 = filtered_URL_status.where("Staus==200")

# Drop URL, adding URI column
URIs = filtered_200.withColumn('URI', split_URL.getItem(1)).drop('URL')

# Group by URI and count
total = URIs.groupBy("URI").count().orderBy(col("count").desc()).show(500)
```

```
+-----+-----+
|          URI|count|
+-----+-----+
|      /login.php| 3290|
|/js/vendor/modern...| 1345|
|      /home.php|  486|
|/contestproblem.p...|  393|
|  /css/normalize.css|  341|
|/css/bootstrap.mi...|  332|
|/css/font-awesome...|  331|
|      /css/style.css|  328|
|      /css/main.css|  326|
|/bootstrap-3.3.7/...|  315|
|/js/vendor/jquery...|  314|
|      /contest.php|  240|
|      /archive.php|  239|
|/fonts/fontawesom...|  224|
|      /img/ruet.png|  195|
|/standings.php?id=16|  148|
|  /allsubmission.php|  140|
|/bootstrap-3.3.7/...|  140|
|/js/vendor/moment...|  132|
|/login.php?value=...|  128|
|      /sign.php|  127|
|      /|  121|
|      /compile.php|   96|
|      /compiler.php|   90|
|      /pcompile.php|   80|
|/contestsubmissio...|   79|
|/countdown.php?na...|   71|
```

/archive.php?page=2	55
/contestproblem.p...	49
/profile.php?user...	48
/contestsubmissio...	47
/details.php?id=45	46
/js/jquery.min.js	45
/js/chart.min.js	44
/details.php?id=43	40
/details.php?id=41	40
/details.php?id=42	40
/details.php?id=44	34
/description.php?...	23
/fonts/fontawesom...	23
/contestsubmit.ph...	22
/description.php?...	20
/details.php?name...	20
/details.php?id=46	19
/details.php?name...	17
/contestsubmit.ph...	17
/contestsubmissio...	15
/contestsubmissio...	14
/profile.php?user...	14
/contestsubmissio...	14
/showcode.php?id=...	13
/submit.php?id=55	12
/setproblem.php	12
/allsubmission.ph...	11
/contestproblem.php	11
/description.php?...	11
/submit.php?id=64	11
/details.php?id=33	10
/showcode.php?id=...	10
/sign.php?value=fail	9
/profile.php?user...	9
/submit.php?id=77	8
/archive.php?page=1	8
/submit.php?id=58	8
/contestsubmissio...	8
/announcement.php	8
/description.php?...	7
/standings.php?id=13	7
/contestsubmit.ph...	7
/update.php	7
/description.php?...	7
/allsubmission.ph...	7
/contestsubmissio...	6
/setcontest.php	6
/description.php?...	6

/description.php?...	6
/profile.php?user...	6
/contestsubmissio...	6
/jquery/jquery-1...	5
/description.php?...	5
/editcontestprobl...	5
/dboot/js/bootstr...	5
/dcss/bootstrap-d...	5
/contestsubmit.ph...	5
/details.php?name...	5
/description.php?...	5
/description.php?...	5
/djs/bootstrap-da...	5
/contestproblem.p...	5
/contestsubmissio...	5
/fonts/fontawesom...	4
/description.php?...	4
/showcode.php?id=...	4
/contestproblem.p...	4
/setcontestproble...	4
/contestproblem.p...	4
/contestproblem.p...	4
/description.php?...	4
/profile.php?user...	4
/details.php?name...	4
/contestsubmissio...	4
/description.php?...	4
/profile.php?user...	3
/contestsubmissio...	3
/profile.php?user...	3
/contestproblem.p...	3
/profile.php?user...	3
/submit.php?id=73	3
/fonts/glyphicons...	3
/editcontestprobl...	3
/profile.php?user...	3
/profile.php?user...	3
/contestproblem.p...	3
/contestsubmissio...	3
/index.php	3
/description.php?...	3
/contestproblem.p...	3
/contestsubmissio...	3
/submit.php?id=78	3
/contestsubmissio...	3
/contestsubmissio...	2
/profile.php?user...	2
/profile.php?user...	2

```

|/showcode.php?id=...|      2|
|/edit.php?name=sh...|      2|
|    /submit.php?id=67|      2|
|/allsubmission.ph...|      2|
|/showcode.php?id=...|      2|
|/description.php?...|      2|
|    /submit.php?id=68|      2|
|/profile.php?user...|      2|
|/description.php?...|      2|
|/allsubmission.ph...|      2|
|/profile.php?user...|      2|
|/allsubmission.ph...|      2|
|/showcode.php?id=...|      2|
|    /details.php?id=38|      2|
|/contestshowcode...|      2|
|/profile.php?user...|      2|
|/description.php?...|      2|
|/profile.php?user...|      2|
|/allsubmission.ph...|      2|
|/showcode.php?id=...|      2|
|/contestsubmissio...|      2|
|    /createadmin.php|      2|
|/description.php?...|      2|
|/contestshowcode...|      2|
|/showcode.php?id=...|      2|
|/fonts/fontawesom...|      2|
|/contestshowcode...|      2|
|/profile.php?user...|      2|
|/showcode.php?id=...|      2|
|    /standings.php?id=4|      2|
|/profile.php?user...|      2|
|/profile.php?user...|      2|
|    /details.php?id=39|      2|
|/description.php?...|      2|
|/profile.php?user...|      2|
|/editcontestprobl...|      2|
|/profile.php?user...|      2|
|    /submit.php?id=76|      2|
|/edit.php?name=as...|      2|
|/showcode.php?id=...|      2|
|/showcode.php?id=...|      2|
|/standings.php?id=12|      2|
|/fonts/glyphicons...|      2|
|/showcode.php?id=...|      1|
|    /standings.php?id=3|      1|
|/profile.php?user...|      1|
|/contestsubmissio...|      1|
|/profile.php?user...|      1|

```

```

|/profile.php?user...| 1|
|/allsubmission.ph...| 1|
|/profile.php?user...| 1|
|/contestsubmissio...| 1|
|/edit.php?name=Aslan| 1|
| /details.php?id=34| 1|
|/profile.php?user...| 1|
|/profile.php?user...| 1|
|/description.php?...| 1|
| /details.php?id=37| 1|
|/fonts/fontawesom...| 1|
|/allsubmission.ph...| 1|
|/showcode.php?id=...| 1|
|/profile.php?user...| 1|
|/profile.php?user...| 1|
|/profile.php?user...| 1|
| /edit.php?name=bala| 1|
|/showcode.php?id=...| 1|
|/profile.php?user...| 1|
|/edit.php?name=ze...| 1|
|/description.php?...| 1|
|/profile.php?user...| 1|
|/allsubmission.ph...| 1|
|/details.php?name...| 1|
|/editcontestprobl...| 1|
|/edit.php?name=ham05| 1|
|/description.php?...| 1|
| /edit.php| 1|
|/showcode.php?id=...| 1|
| /submit.php?id=63| 1|
|/profile.php?user...| 1|
|/showcode.php?id=...| 1|
|/contestsubmissio...| 1|
|/editcontestprobl...| 1|
|/edit.php?name=ks...| 1|
| /details.php?id=20| 1|
|/profile.php?user...| 1|
|/details.php?name...| 1|
| /submit.php?id=71| 1|
|/countdown.php?na...| 1|
|/profile.php?user...| 1|
|/edit.php?name=Ta...| 1|
|/showcode.php?id=...| 1|
|/contestsubmissio...| 1|
|/contestsubmissio...| 1|
|/showcode.php?id=...| 1|
|/contestsubmissio...| 1|
|/contestsubmit.ph...| 1|

```

```

|/showcode.php?id=...|    1|
|/showcode.php?id=...|    1|
|/profile.php?user...|    1|
|/standings.php?id=14|    1|
|/contestsubmissio...|    1|
|  /details.php?id=35|    1|
|/contestproblem.p...|    1|
|/contestsubmissio...|    1|
|/showcode.php?id=...|    1|
|  /details.php?id=3|    1|
|/editcontest.php?...|    1|
|/allsubmission.ph...|    1|
|/edit.php?name=ma...|    1|
|/contestsubmissio...|    1|
|/showcode.php?id=...|    1|
|/contestsubmissio...|    1|
|/contestsubmissio...|    1|
|/profile.php?user...|    1|
|/profile.php?user...|    1|
|/edit.php?name=Sh...|    1|
|  /details.php?id=36|    1|
|/profile.php?user...|    1|
|/contestsubmissio...|    1|
|/profile.php?user...|    1|
|/profile.php?user...|    1|
|/contestproblem.p...|    1|
|/profile.php?user...|    1|
|  /details.php?id=9|    1|
|/edit.php?name=m...|    1|
+-----+-----+

```

[ ]: