

# Time Processing







```
access_log = spark_session.read.table("web.access_log")
```

```
access_log.limit(5).toPandas()
```

	http_code	ip	response_length	time	url	user_agent
0	200	109.106.133.8	21546	12/Dec/2015 :01:31:46 +0400	/id53821	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_4)...
1	200	46.31.82.254	8777	12/Dec/2015 :01:31:47 +0400	/id33929	Mozilla/5.0 (Windows NT 5.1; U; de; rv:1.9.1.6...
2	200	193.124.254.46	8731	12/Dec/2015 :01:31:48 +0400	/id35754	Mozilla/4.0 (compatible; MSIE 7.0; Windows NT ...
3	200	185.103.220.164	22776	12/Dec/2015 :01:31:48 +0400	/id78231	Mozilla/5.0 (Linux; Android 4.4.2; nb-no; SAMS...
4	200	185.103.220.164	18335	12/Dec/2015 :01:31:49 +0400	/id39395	Mozilla/5.0 (Linux; Android 4.4.2; nb-no; SAMS...

```
access_log = spark_session.read.table("web.access_log")
```

```
access_log.limit(5).toPandas()
```

	http_code	ip	response_length	time	url	user_agent
0	200	109.106.133.8	21546	12/Dec/2015 :01:31:46 +0400	/id53821	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_4)...
1	200	46.31.82.254	8777	12/Dec/2015 :01:31:47 +0400	/id33929	Mozilla/5.0 (Windows NT 5.1; U; de; rv:1.9.1.6...
2	200	193.124.254.46	8731	12/Dec/2015 :01:31:48 +0400	/id35754	Mozilla/4.0 (compatible; MSIE 7.0; Windows NT ...
3	200	185.103.220.164	22776	12/Dec/2015 :01:31:48 +0400	/id78231	Mozilla/5.0 (Linux; Android 4.4.2; nb-no; SAMS...
4	200	185.103.220.164	18335	12/Dec/2015 :01:31:49 +0400	/id39395	Mozilla/5.0 (Linux; Android 4.4.2; nb-no; SAMS...

# In this video you will learn

- how time is stored in spark
- how to parse strings with time
- how to work with time to solve our problems





```
spark_session.sql("""  
    select unix_timestamp() as unix_timestamp  
""").toPandas()
```

	unix_timestamp
0	1499760609



timestamp

```
spark_session.sql("""  
    select current_timestamp() as timestamp  
""").toPandas()
```

	timestamp
0	2017-07-09 18:51:35.198

```
spark_session.sql("""  
    select cast(current_timestamp() as bigint) as int_timestamp  
""").toPandas()
```

	int_timestamp
0	1499760619

```
spark_session.sql("""  
    select cast(current_timestamp() as bigint) as int_timestamp  
""").toPandas()
```

	int_timestamp
0	1499760619

	unix_timestamp
0	1499760609



```
access_log.withColumn("unixtime", f.unix_timestamp("time"))\  
            .limit(5).toPandas()
```



```
access_log.withColumn("unixtime", f.unix_timestamp("time"))\
    .limit(5).toPandas()
```

	http_code	ip	response_length	time	url	user_agent	unixtime
0	200	109.106.133.8	21546	12/Dec/2015 :01:31:46 +0400	/id53821	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_4)...	None
1	200	46.31.82.254	8777	12/Dec/2015 :01:31:47 +0400	/id33929	Mozilla/5.0 (Windows NT 5.1; U; de; rv:1.9.1.6...	None
2	200	193.124.254.46	8731	12/Dec/2015 :01:31:48 +0400	/id35754	Mozilla/4.0 (compatible; MSIE 7.0; Windows NT ...	None
3	200	185.103.220.164	22776	12/Dec/2015 :01:31:48 +0400	/id78231	Mozilla/5.0 (Linux; Android 4.4.2; nb-no; SAMS...	None
4	200	185.103.220.164	18335	12/Dec/2015 :01:31:49 +0400	/id39395	Mozilla/5.0 (Linux; Android 4.4.2; nb-no; SAMS...	None

```
access_log.withColumn("unixtime",  
                        f.unix_timestamp("time",  
                                         "dd/MMM/yyyy:HH:mm:ss Z"))\  
      .limit(5).toPandas()
```

	http_code	ip	response_length	time	url	user_agent
0	200	109.106.133.8	21546	12/Dec/2015 :01:31:46 +0400	/id53821	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_4)...
1	200	46.31.82.254	8777	12/Dec/2015 :01:31:47 +0400	/id33929	Mozilla/5.0 (Windows NT 5.1; U; de; rv:1.9.1.6...
2	200	193.124.254.46	8731	12/Dec/2015 :01:31:48 +0400	/id35754	Mozilla/4.0 (compatible; MSIE 7.0; Windows NT ...
3	200	185.103.220.164	22776	12/Dec/2015 :01:31:48 +0400	/id78231	Mozilla/5.0 (Linux; Android 4.4.2; nb-no; SAMS...
4	200	185.103.220.164	18335	12/Dec/2015 :01:31:49 +0400	/id39395	Mozilla/5.0 (Linux; Android 4.4.2; nb-no; SAMS...

```
access_log.withColumn(«unixtime»,  
                      f.unix_timestamp(«time»,  
                                       «dd/MMM/yyyy:HH:mm:ss Z»))\  
      .limit(5).toPandas()
```

	http_code	ip	response_length	time	url	user_agent
0	200	109.106.133.8	21546	<u>12/Dec/2015</u> :01:31:46 +0400	/id53821	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_4)...
1	200	46.31.82.254	8777	12/Dec/2015 :01:31:47 +0400	/id33929	Mozilla/5.0 (Windows NT 5.1; U; de; rv:1.9.1.6...
2	200	193.124.254.46	8731	12/Dec/2015 :01:31:48 +0400	/id35754	Mozilla/4.0 (compatible; MSIE 7.0; Windows NT ...
3	200	185.103.220.164	22776	12/Dec/2015 :01:31:48 +0400	/id78231	Mozilla/5.0 (Linux; Android 4.4.2; nb-no; SAMS...
4	200	185.103.220.164	18335	12/Dec/2015 :01:31:49 +0400	/id39395	Mozilla/5.0 (Linux; Android 4.4.2; nb-no; SAMS...

```
access_log.withColumn(«unixtime»,  
                      f.unix_timestamp(«time»,  
                                       «dd/MMM/yyyy:HH:mm:ss Z»))\  
      .limit(5).toPandas()
```

	http_code	ip	response_length	time	url	user_agent
0	200	109.106.133.8	21546	<u>12/Dec/2015</u> <u>:01:31:46</u> +0400	/id53821	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_4)...
1	200	46.31.82.254	8777	12/Dec/2015 :01:31:47 +0400	/id33929	Mozilla/5.0 (Windows NT 5.1; U; de; rv:1.9.1.6...
2	200	193.124.254.46	8731	12/Dec/2015 :01:31:48 +0400	/id35754	Mozilla/4.0 (compatible; MSIE 7.0; Windows NT ...
3	200	185.103.220.164	22776	12/Dec/2015 :01:31:48 +0400	/id78231	Mozilla/5.0 (Linux; Android 4.4.2; nb-no; SAMS...
4	200	185.103.220.164	18335	12/Dec/2015 :01:31:49 +0400	/id39395	Mozilla/5.0 (Linux; Android 4.4.2; nb-no; SAMS...

```
access_log.withColumn(«unixtime»,  
                      f.unix_timestamp(«time»,  
                                       «dd/MMM/yyyy:HH:mm:ss Z»))\  
      .limit(5).toPandas()
```

	http_code	ip	response_length	time	url	user_agent
0	200	109.106.133.8	21546	<u>12/Dec/2015</u> <u>:01:31:46</u> <u>+0400</u>	/id53821	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_4)...
1	200	46.31.82.254	8777	12/Dec/2015 :01:31:47 +0400	/id33929	Mozilla/5.0 (Windows NT 5.1; U; de; rv:1.9.1.6...
2	200	193.124.254.46	8731	12/Dec/2015 :01:31:48 +0400	/id35754	Mozilla/4.0 (compatible; MSIE 7.0; Windows NT ...
3	200	185.103.220.164	22776	12/Dec/2015 :01:31:48 +0400	/id78231	Mozilla/5.0 (Linux; Android 4.4.2; nb-no; SAMS...
4	200	185.103.220.164	18335	12/Dec/2015 :01:31:49 +0400	/id39395	Mozilla/5.0 (Linux; Android 4.4.2; nb-no; SAMS...



```

access_log_unixtime = access_log\
    .withColumn("unixtime",
                 f.unix_timestamp("time",
                                   "dd/MMM/yyyy:HH:mm:ss Z"))\

access_log_unixtime.limit(5).toPandas()

```

	http_code	ip	response_length	time	url	user_agent	unixtime
0	200	109.106.133.8	21546	12/Dec/2015 :01:31:46 +0400	/id53821	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_4)...	1449869506
1	200	46.31.82.254	8777	12/Dec/2015 :01:31:47 +0400	/id33929	Mozilla/5.0 (Windows NT 5.1; U; de; rv:1.9.1.6...	1449869507
2	200	193.124.254.46	8731	12/Dec/2015 :01:31:48 +0400	/id35754	Mozilla/4.0 (compatible; MSIE 7.0; Windows NT ...	1449869508
3	200	185.103.220.164	22776	12/Dec/2015 :01:31:48 +0400	/id78231	Mozilla/5.0 (Linux; Android 4.4.2; nb-no; SAMS...	1449869508
4	200	185.103.220.164	18335	12/Dec/2015 :01:31:49 +0400	/id39395	Mozilla/5.0 (Linux; Android 4.4.2; nb-no; SAMS...	1449869509



```
access_log_unixtime.groupby("ip")\  
    .agg(f.min("unixtime").alias("begin"),  
         f.max("unixtime").alias("end"))\  
    .limit(5).toPandas()
```

	ip	begin	end
0	62.187.94.9	1449778811	1449870627
1	78.68.118.92	1449870844	1449870849
2	217.197.14.94	1449954184	1449954204
3	37.128.131.92	1449954429	1449954456
4	193.84.177.216	1449955145	1449955199



```
access_log_unixtime.groupby("ip")\  
    .agg(f.min("unixtime").alias("begin"),  
         f.max("unixtime").alias("end"))\  
    .select("ip", (f.col("end")-f.col("begin")))\  
    .limit(5).toPandas()
```

	ip	(end-begin)
0	62.187.94.9	91816
1	78.68.118.92	5
2	217.197.14.94	20
3	37.128.131.92	27
4	193.84.177.216	54

```
access_log_unixtime.groupby("ip")\
    .agg(f.min("unixtime").alias("begin"),
         f.max("unixtime").alias("end"))\
    .select("ip", (f.col("end")-f.col("begin")).alias("seconds_cnt"))\
    .limit(5).toPandas()
```

	ip	seconds_cnt
0	62.187.94.9	91816
1	78.68.118.92	5
2	217.197.14.94	20
3	37.128.131.92	27
4	193.84.177.216	54

```
access_log_unixtime.groupby("ip")\
    .agg(f.min("unixtime").alias("begin"),
         f.max("unixtime").alias("end"))\
    .select("ip", (f.col("end")-f.col("begin")) .alias("seconds_cnt"))\
    .select("ip", f.col("seconds_cnt")/60.0
                                                    /60.0
                                                    /24.0)\
    .limit(5).toPandas()
```

	ip	$((\text{seconds\_cnt} / 60.0) / 60.0) / 24.0$
0	62.187.94.9	1.062685
1	78.68.118.92	0.000058
2	217.197.14.94	0.000231
3	37.128.131.92	0.000313
4	193.84.177.216	0.000625

```
access_log_unixtime.groupby("ip")\
    .agg(f.min("unixtime").alias("begin"),
         f.max("unixtime").alias("end"))\
    .select("ip", (f.col("end")-f.col("begin")) .alias("seconds_cnt"))\
    .select("ip", f.col("seconds_cnt")/60.0
                                                    /60.0
                                                    /24.0+1)\
    .limit(5).toPandas()
```

	ip	$((\text{seconds\_cnt} / 60.0) / 60.0) / 24.0 + 1$
0	62.187.94.9	2.062685
1	78.68.118.92	1.000058
2	217.197.14.94	1.000231
3	37.128.131.92	1.000313
4	193.84.177.216	1.000625



Date diff

```
access_log_unixtime\  
    .withColumn("timestamp",  
                f.col("unixtime").astype("timestamp"))\  
    .limit(5).toPandas()
```

	http_code	ip	response_length	time	url	user_agent	unixtime	timestamp
0	200	109.106.133.8	21546	12/Dec/2015:01:31:46+0400	/id53821	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_4)...	1449869506	2015-12-11 21:31:46
1	200	46.31.82.254	8777	12/Dec/2015:01:31:47+0400	/id33929	Mozilla/5.0 (Windows NT 5.1....	1449869507	2015-12-11 21:31:47
2	200	193.124.254.46	8731	12/Dec/2015:01:31:48+0400	/id35754	Mozilla/4.0 (compatible; MSIE 7.0; Windows NT ...	1449869508	2015-12-11 21:31:48
3	200	185.103.220.164	22776	12/Dec/2015:01:31:48+0400	/id78231	Mozilla/5.0 (Linux; Android 4.4.2;...	1449869508	2015-12-11 21:31:48
4	200	185.103.220.164	18335	12/Dec/2015:01:31:49+0400	/id39395	Mozilla/5.0 (Linux; Android 4.4.2;...	1449869509	2015-12-11 21:31:49

```
access_log_ts = access_log_unixtime\  
                .withColumn("timestamp",  
                             f.col("unixtime").astype("timestamp"))  
access_log_ts.limit(5).toPandas()
```

	http_code	ip	response_length	time	url	user_agent	unixtime	timestamp
0	200	109.106.133.8	21546	12/Dec/2015:01:31:46+0400	/id53821	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_4)...	1449869506	2015-12-11 21:31:46
1	200	46.31.82.254	8777	12/Dec/2015:01:31:47+0400	/id33929	Mozilla/5.0 (Windows NT 5.1....	1449869507	2015-12-11 21:31:47
2	200	193.124.254.46	8731	12/Dec/2015:01:31:48+0400	/id35754	Mozilla/4.0 (compatible; MSIE 7.0; Windows NT ...	1449869508	2015-12-11 21:31:48
3	200	185.103.220.164	22776	12/Dec/2015:01:31:48+0400	/id78231	Mozilla/5.0 (Linux; Android 4.4.2;...	1449869508	2015-12-11 21:31:48
4	200	185.103.220.164	18335	12/Dec/2015:01:31:49+0400	/id39395	Mozilla/5.0 (Linux; Android 4.4.2;...	1449869509	2015-12-11 21:31:49



```
access_log_ts.groupby("ip")\  
    .agg(f.min("timestamp").alias("begin"),  
         f.max("timestamp").alias("end"))\  
    .limit(5).toPandas()
```

	ip	begin	end
0	62.187.94.9	2015-12-10 20:20:11	2015-12-11 21:50:27
1	78.68.118.92	2015-12-11 21:54:04	2015-12-11 21:54:09
2	217.197.14.94	2015-12-12 21:03:04	2015-12-12 21:03:24
3	37.128.131.92	2015-12-12 21:07:09	2015-12-12 21:07:36
4	193.84.177.216	2015-12-12 21:19:05	2015-12-12 21:19:59

```
access_log_ts.groupby("ip")\  
    .agg(f.min("timestamp").alias("begin"),  
         f.max("timestamp").alias("end"))\  
    .select("ip", (f.datediff("end", "begin")).alias("days_cnt"))\  
    .limit(5).toPandas()
```

	ip	days_cnt
0	62.187.94.9	1
1	78.68.118.92	0
2	217.197.14.94	0
3	37.128.131.92	0
4	193.84.177.216	0

```
access_log_ts.groupby("ip")\  
    .agg(f.min("timestamp").alias("begin"),  
         f.max("timestamp").alias("end"))\  
    .select("ip", (f.datediff("end", "begin")+1).alias("days_cnt"))\  
    .limit(5).toPandas()
```

	ip	days_cnt
0	62.187.94.9	2
1	78.68.118.92	1
2	217.197.14.94	1
3	37.128.131.92	1
4	193.84.177.216	1

# Now you know

- how time is stored in spark
- how to parse strings with time
- how to work with time to solve our problems