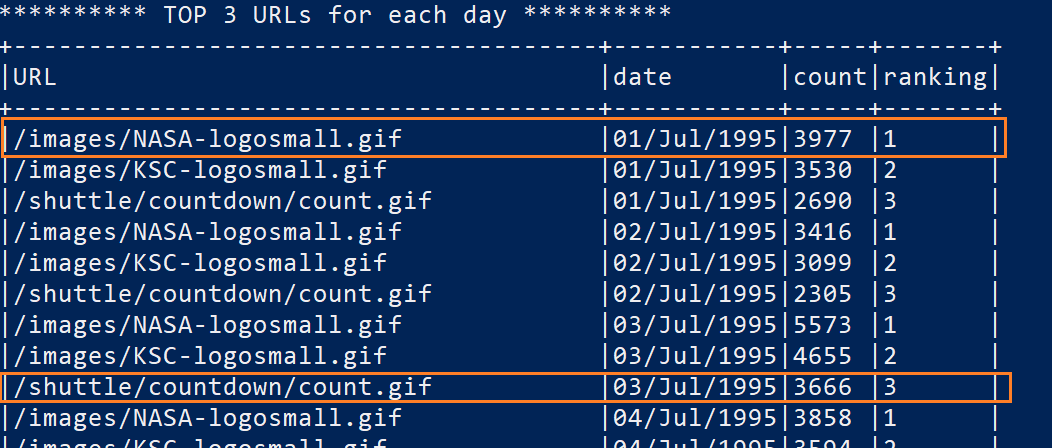
**Unit Testing for Top-N most frequent URLs:**

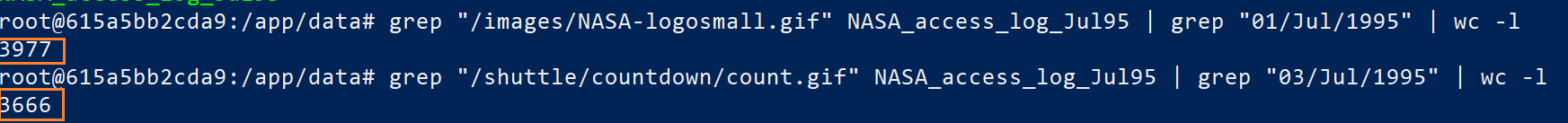
Below is some portion of the output from the spark job. We will validate the highlighted in orange counts using linux commands.



Using the below grep commands, we can see the results are matching.

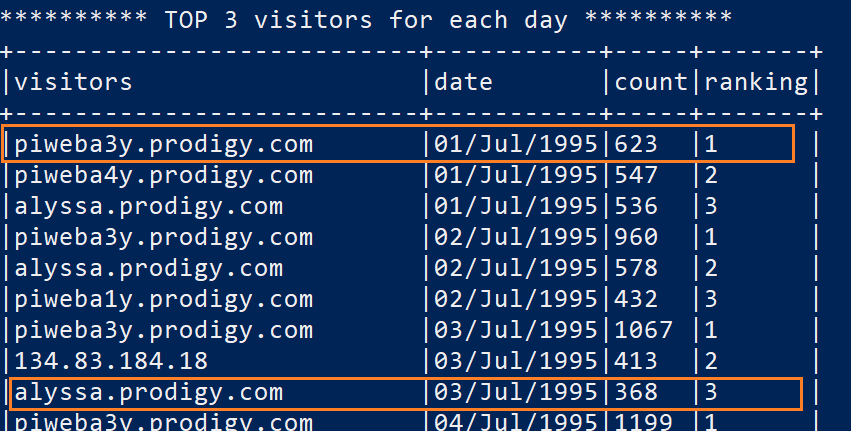
grep "/images/NASA-logosmall.gif" NASA\_access\_log\_Jul95 | grep "01/Jul/1995" | wc -l

grep "/shuttle/countdown/count.gif" NASA\_access\_log\_Jul95 | grep "03/Jul/1995" | wc -l



**Unit Testing for Top-N most frequent visitors:**

Below are few output rows from the spark job. We will validate the results highlighted in orange boxes using linux commands.



Using the below grep commands, we get the same results.

grep "piweba3y.prodigy.com" NASA\_access\_log\_Jul95 | grep "01/Jul/1995" | wc -l

grep "alyssa.prodigy.com" NASA\_access\_log\_Jul95 | grep "03/Jul/1995" | wc -l

