Part 1.

Top 100 Words:

```
1. Word: said, Count: 16958954
2. Word: I, Count: 9961168
3. Word: , Count: 6982257
4. Word: would, Count: 5787877
5. Word: one, Count: 5015462
6. Word: also, Count: 4477917
7. Word: people, Count: 4015800
8. Word: year, Count: 3908563
9. Word: He, Count: 3772036
10. Word: Mr, Count: 3729679
11. Word: It, Count: 3722954
12. Word: In, Count: 3718041
13. Word: two, Count: 3642120
14. Word: first, Count: 3590245
15. Word: years, Count: 3575937
16. Word: last, Count: 3571939
17. Word: time, Count: 3500072
18. Word: new, Count: 3462921
19. Word: A, Count: 3369396
20. Word: could, Count: 3335796
21. Word: US, Count: 3242110
22. Word: like, Count: 2908097
23. Word: him, Count: 2745707
24. Word: We, Count: 2742413
25. Word: says, Count: 2387393
26. Word: get, Count: 2286265
27. Word: million, Count: 2228848
28. Word: back, Count: 2216942
29. Word: government, Count: 2216286
30. Word: made, Count: 2112044
31. Word: told, Count: 2088780
32. Word: three, Count: 2076528
33. Word: make, Count: 2062872
34. Word: many, Count: 2033137
35. Word: percent, Count: 2016333
36. Word: This, Count: 2015765
37. Word: New, Count: 1990427
38. Word: way, Count: 1896643
39. Word: say, Count: 1845459
40. Word: even, Count: 1825259
41. Word: much, Count: 1818380
42. Word: since, Count: 1797599
43. Word: And, Count: 1794800
44. Word: well, Count: 1783449
45. Word: still, Count: 1782164
46. Word: Its, Count: 1766166
47. Word: may, Count: 1765668
48. Word: per, Count: 1763626
49. Word: week, Count: 1760487
50. Word: work, Count: 1725326
51. Word: take, Count: 1717389
```

Top 100 Words with more than 6 characters:

```
1. Word: million, Count: 2228848
2. Word: government, Count: 2216286
3. Word: percent, Count: 2016333
4. Word: including, Count: 1587810
5. Word: company, Count: 1441241
6. Word: another, Count: 1304675
7. Word: country, Count: 1247074
8. Word: President, Count: 1195606
9. Word: business, Count: 1134627
10. Word: American, Count: 1125997
11. Word: officials, Count: 1090915
12. Word: support, Count: 1050791
13. Word: billion, Count: 1041505
14. Word: European, Count: 1017653
15. Word: children, Count: 1010708
16. Word: according, Count: 1002939
17. Word: president, Count: 990651
18. Word: yearold, Count: 976865
19. Word: political, Count: 975737
20. Word: Tuesday, Count: 968395
21. Word: without, Count: 962312
22. Word: whether, Count: 945231
23. Word: already, Count: 936733
24. Word: Wednesday, Count: 919779
25. Word: financial, Count: 916291
26. Word: reported, Count: 912274
27. Word: expected, Count: 912176
28. Word: Thursday, Count: 909748
29. Word: something, Count: 871349
30. Word: military, Count: 848191
31. Word: information, Count: 845618
32. Word: economic, Count: 832171
33. Word: companies, Count: 828339
34. Word: British, Count: 812232
35. Word: statement, Count: 802312
36. Word: security, Count: 800012
37. Word: members, Count: 796375
38. Word: economy, Count: 794501
39. Word: campaign, Count: 781054
40. Word: several, Count: 778289
41. Word: important, Count: 767666
42. Word: working, Count: 737672
43. Word: countries, Count: 727760
44. Word: players, Count: 724315
45. Word: decision, Count: 723288
46. Word: different, Count: 714556
47. Word: National, Count: 707963
```

Part 2

1. Section A:

```
+-------------+-------------------------+----------+
|Day in a Week|first(endpoint)          |max(count)|
+-------------+-------------------------+----------+
|Thursday     |/images/NASA-logosmall.gif|46920    |
|Wednesday    |/images/NASA-logosmall.gif|37598    |
|Tuesday      |/images/NASA-logosmall.gif|33685    |
|Monday       |/images/NASA-logosmall.gif|30380    |
|Friday       |/images/NASA-logosmall.gif|29139    |
|Saturday     |/images/NASA-logosmall.gif|16168    |
|Sunday       |/images/KSC-logosmall.gif |15218    |
+-------------+-------------------------+----------+
```

```
+----+-----+
|Year|count|
+----+-----+
|1995|20899|
+----+-----+
```

2. Section B:

```
Start
Getting Data and appending=0
Getting Data and appending=1
Getting Data and appending=2
Getting Data and appending=3
Getting Data and appending=4
Getting Data and appending=5
Getting Data and appending=6
Getting Data and appending=7
Getting Data and appending=8
Getting Data and appending=9
Getting Data and appending=10
```

```
+---------------+------+-----------------+--------+------+------------+-------------------+
|           host|method|         endpoint|protocol|status|content_size|               time|
+---------------+------+-----------------+--------+------+------------+-------------------+
|  10.82.240.107|   GET|/Archives/edgar/d...|HTTP/1.0|  304|        1638|2118-12-26 12:00:00|
|169.174.119.238|DELETE|/Archives/edgar/d...|HTTP/1.0|  303|       40118|2118-12-26 12:00:00|
| 116.82.220.168|  POST|/Archives/edgar/d...|HTTP/1.0|  403|        5856|2118-12-26 12:00:00|
| 99.149.129.109|  POST|/Archives/edgar/d...|HTTP/1.0|  500|       25206|2118-12-26 12:00:00|
|  167.134.232.8|   PUT|/Archives/edgar/d...|HTTP/1.0|  200|       29727|2118-12-26 12:00:00|
| 128.171.81.194|   GET|/Archives/edgar/d...|HTTP/1.0|  304|        3695|2118-12-26 12:00:00|
|  74.59.229.245|   GET|/Archives/edgar/d...|HTTP/1.0|  200|       37021|2118-12-26 12:00:00|
|198.123.235.207|DELETE|/Archives/edgar/d...|HTTP/1.0|  403|       46602|2118-12-26 12:00:00|
|   77.26.34.100|   PUT|/Archives/edgar/d...|HTTP/1.0|  502|       50169|2118-12-26 12:00:00|
|  83.211.223.129|  POST|/Archives/edgar/d...|HTTP/1.0|  403|       39593|2118-12-26 12:00:00|
+---------------+------+-----------------+--------+------+------------+-------------------+
only showing top 10 rows
```

```
23/06/06 14:46:16 WARN TaskSetManager: Stage 4 contains a task of very large size (5318 KiB). The maximum recommended task size is 1000 KiB.
23/06/06 14:46:19 WARN MemoryManager: Total allocation exceeds 95.00% (906,992,014 bytes) of heap memory
Scaling row group sizes to 96.54% for 7 writers
23/06/06 14:46:19 WARN MemoryManager: Total allocation exceeds 95.00% (906,992,014 bytes) of heap memory
Scaling row group sizes to 84.47% for 8 writers
23/06/06 14:46:22 WARN MemoryManager: Total allocation exceeds 95.00% (906,992,014 bytes) of heap memory
Scaling row group sizes to 96.54% for 7 writers
Saving File
```

3. Section C:

```
         requesting_host  cluster
0          in24.inetnebr.com      0
1            uplherc.upl.com      0
2            uplherc.upl.com      0
3            uplherc.upl.com      0
4            uplherc.upl.com      0
...                    ...    ...
1555715   gatekeeper.uccu.com      2
1555716   gatekeeper.uccu.com      2
1555717  cys-cap-9.wyoming.com      2
1555718    www-c8.proxy.aol.com     2
1555719  cindy.yamato.ibm.co.jp     2

[1555720 rows x 2 columns]

Cluster statistics:
cluster
0     526820
2     520682
1     508218
Name: count, dtype: int64
```