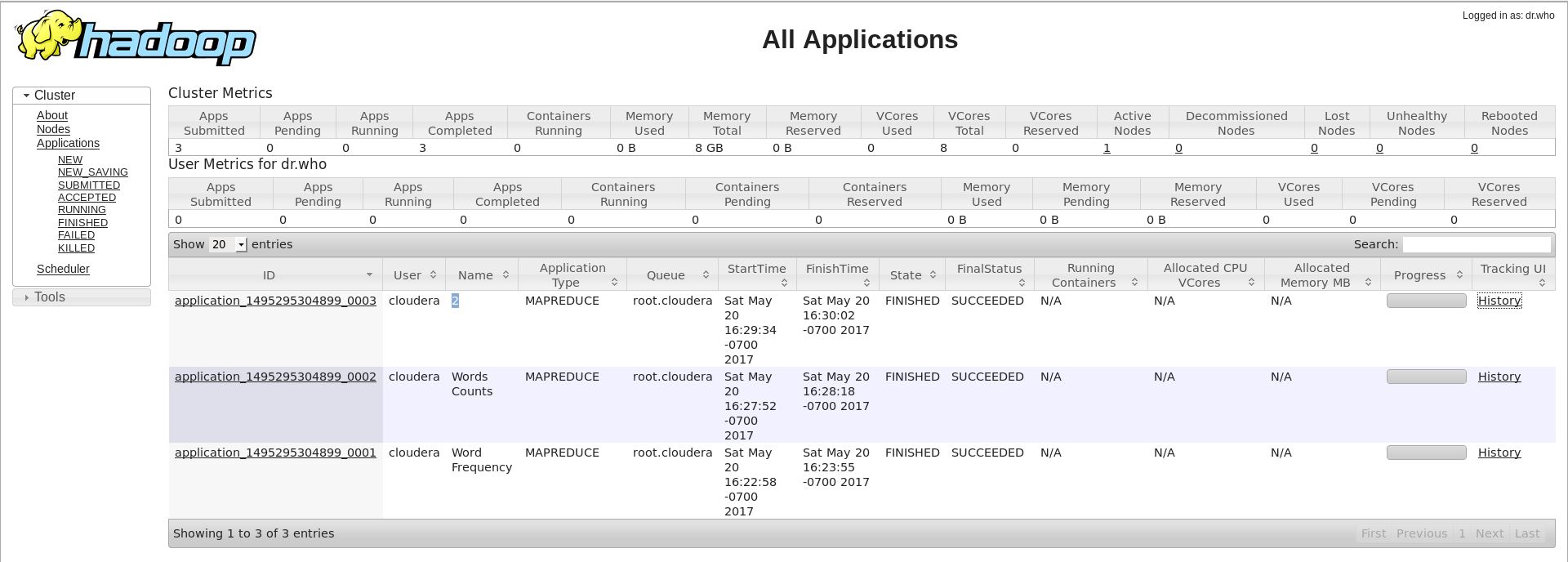
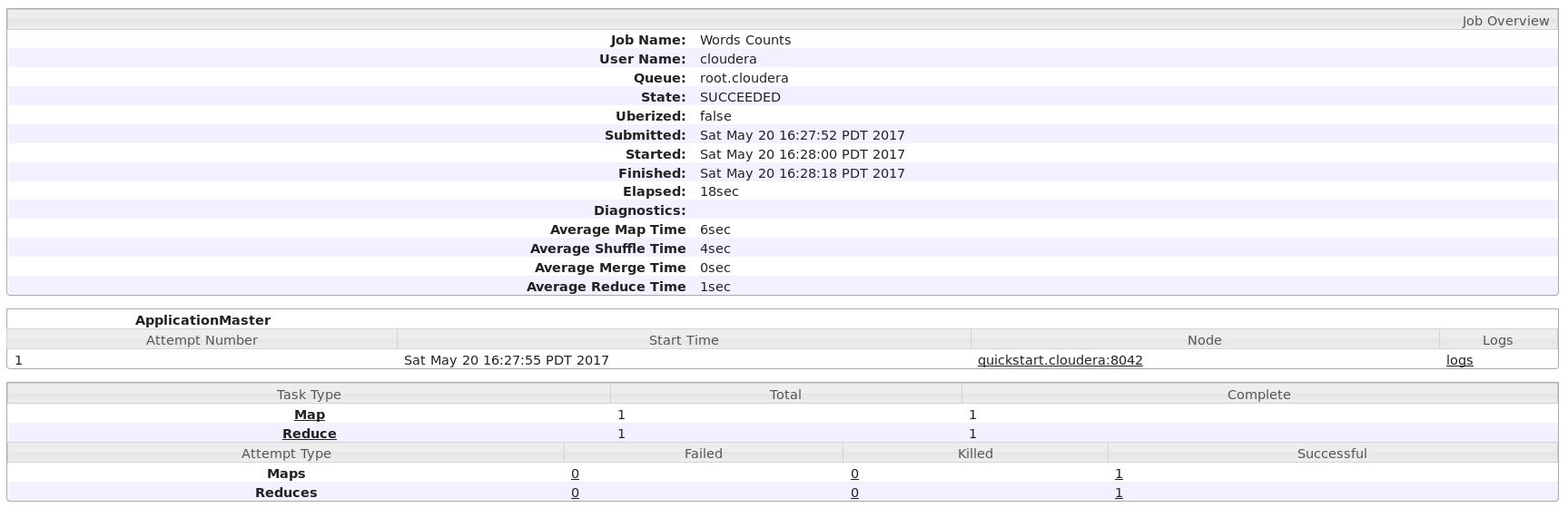
**QUESTION 4 OUTPUTS AND HADOOP RUN SUMMARIES**

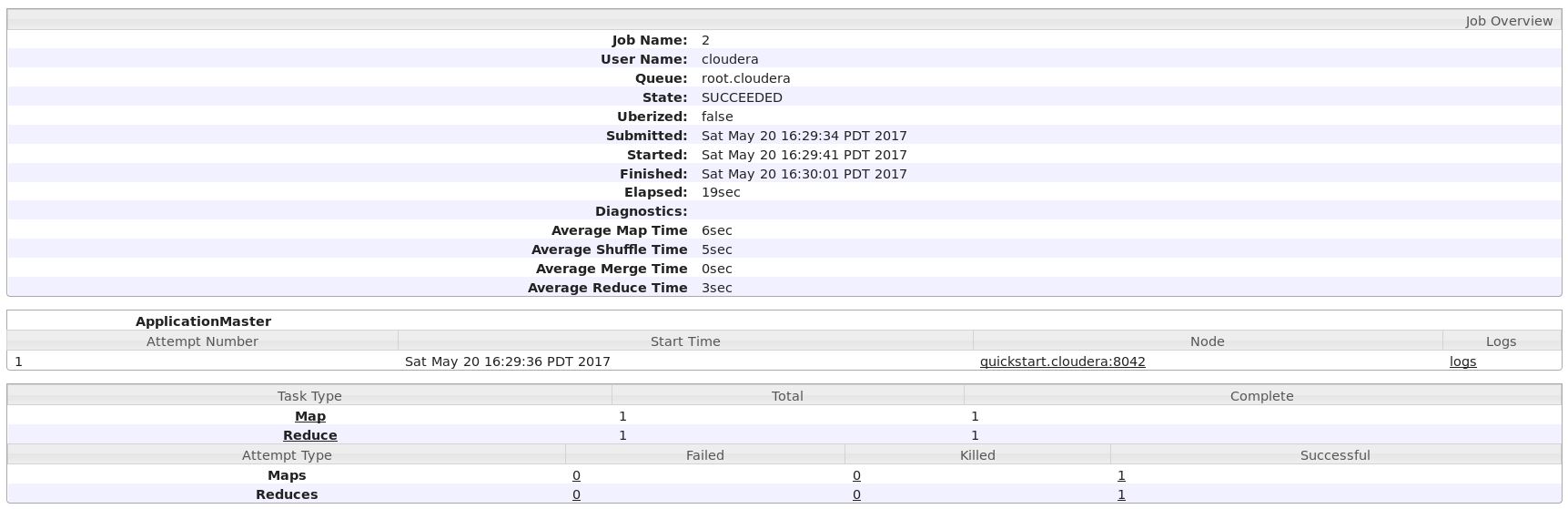
Q4.1. Top 20 TF IDF

|  |  |
| --- | --- |
| the@callwild | 0.07055 |
| the@robinsoncrusoe | 0.04842 |
| and@callwild | 0.04744 |
| i@robinsoncrusoe | 0.04229 |
| and@robinsoncrusoe | 0.03953 |
| to@robinsoncrusoe | 0.03539 |
| of@robinsoncrusoe | 0.02899 |
| of@callwild | 0.027 |
| he@callwild | 0.02524 |
| was@callwild | 0.02152 |
| to@callwild | 0.0209 |
| a@callwild | 0.02044 |
| a@robinsoncrusoe | 0.01872 |
| my@robinsoncrusoe | 0.01759 |
| his@callwild | 0.01737 |
| in@callwild | 0.01669 |
| was@robinsoncrusoe | 0.01655 |
| in@robinsoncrusoe | 0.0159 |
| that@robinsoncrusoe | 0.01552 |
| it@robinsoncrusoe | 0.01515 |
|  |  |









Q4.2. Top 10 PageRank

PageRank Node

|  |  |
| --- | --- |
| 47.07101058959961 | 18 |
| 21.750492095947266 | 4415 |
| 20.142515182495117 | 737 |
| 17.534427642822266 | 790 |
| 17.32414436340332 | 1753 |
| 17.248205184936523 | 143 |
| 17.167720794677734 | 1719 |
| 15.091687202453613 | 136 |
| 15.033056259155273 | 751 |
| 13.03643798828125 | 118 |



