Some Applications of MapReduce (3)

Count of URL access frequency

- − Input: Log of accessed URLs, e.g., from proxy server
- − Output: For each URL, % of total accesses for that URL

- − Map – *Process web log and outputs <URL, 1>*
- − Multiple Reducers - *Emits <URL, URL_count>*

(So far, like Wordcount. But still need %)

- − Chain another MapReduce job after above one
- − Map – *Processes <URL, URL_count> and outputs*   *<1, (<URL, URL_count>)>*
- − 1 Reducer – Does two passes. In first pass, sums up all *URL_count's* to calculate overall_count. In second pass calculates %'s

  Emits multiple <URL, URL_count/overall_count>