

Answers

1. Timeslots distribution

- Map: (offset, line) → [(timeslot, 1)]
- Reduce*: (timeslot, [1, 1, ...]) → [(timeslot, n)]

2. Over 1 million hits

- Map: (offset, line) → [(url, 1)]
- Combiner (url, [1, 1, ...]) → [(url, n)]
- Reduce: (url, [1, 1, ...]) → [(url, n)] if $n > 10^6$

* Combiner too

Answers

3. Links

- Map: $(url, content) \rightarrow [(link1, url), (link2, url), \dots]$
- Reduce: $(link, [url1, url2, \dots]) \rightarrow [(link, [url1, url2, \dots])]$

4. Match proportion (but, counters better)

- Map: $(offset, line) \rightarrow [(match, 1), (total, 1)]$
- Reduce*: $(type, [1, 1, \dots]) \rightarrow [(type, n)]$

5. Sort

- Map: $(offset, line) \rightarrow [((k1, k2), line)]$
- Reduce: $((k1, k2), [line1, line2, \dots])$
- $\rightarrow [((k1, k2), line1), ((k1, k2), line2), \dots]$

Answers

6. Most popular pages

- Map 1: (offset, line) → [(url, 1)]
- Reduce* 1: (url, [1, 1, ...]) → [(url, n)]
- Map 2: (url, n) → [(n, url)]
- Reduce* 2: (n, [url1, url2, ...]) → [(url1, n), (url2, n), ...]