MapReduce Education Resource

Overview

This exercise aims to teach children about the MapReduce function that is a part of the software library Hadoop. The idea of MapReduce is to decompose large tasks into smaller tasks that can be computed in parallel across multiple servers. It consists of a Map function which breaks the tasks down, and then a Reduce function which collects the outputs from Map and combines them back up into a data set.

Suitable For

Primary school children - 8+, can be adapted to suit older children by using a more complicated story, see Alice in the Wonderland and Jekyll & Hyde excerpts respectively

Key Concepts

MapReduce function that comes with Hadoop, idea of parallelism, using multiple cores, word counting

Learning Intentions

* Remember what the MapReduce function is and what it is used for.
* Apply the knowledge of MapReduce to count a series of words from a story.

How you are learning – recipe to complete a task

* Read every word and count words.
* Split up text into smaller units and count them again in parallel.
* Compare the times this took and understand why the parallel time is more likely to be faster.

Time Required

15 minutes, introductory task

Preparation

1. Print out a page from the story.docx for each group of three.
2. Print out the tally board, one for every pair of the class, and then another time for every four in the class for the second round of counting.

Prior Learning Assumed

None - new topic

Outline of Activity

1. Split the class up into pairs
2. Describe MapReduce to students
3. Say that each group has a section from a chapter of Alice in Wonderland and say that you want to count how many times the names ‘Hatter’ ‘Alice’ and ‘Dormouse’ appear in the story they are given or alternatively ‘Jekyll’, ‘Hyde’, ‘Utterson’ if used for higher years
4. Illustrate a slow and bad way:
   1. Instruct the pair to split into counter and timer.
   2. Outline that they should use their board on the desk to tally up each time they find the word in the story.
   3. Ask the counter to count how many times the words appear, the timer to note the time this takes, and that they should report these back to you, note down the times on a board.
5. Explain that they can split up this big task into smaller tasks to solve it more quickly. Put each pair with another pair, and ask them to allocate one person to be the timer the rest to be counters.
6. Ask them to cut along the dotted lines, each ‘counter’ getting a third of the A4 page. Outline that they will have to each individually count the number of time each word appears, and put the tally in the shared tally board. Make sure each student has a section from their page and that the three have one shared tally board.
7. Again, use showme boards to get how long it took them, and also the count from the results. Keep a count on the board from each group, with a ‘total’ underneath.
8. Explain that the process of splitting up and counting is called ‘mapping’ and that this is the computer counting each time the word appears individually. Explain that reporting back to the teacher is ‘reducing’ - putting back together the split up totals each section has.
9. When all groups have reported back, add up the total at the bottom and explain that you have counted up how many times each word has individually appeared in the story – and illustrate (hopefully by comparing the times) that it was much much quicker than a single person and this is how MapReduce works.