MapReduce Education Resource

Overview

This exercise aims to teach children about the MapReduce function that is a part of the software library Hadoop. The main 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 output 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.

Key Concepts

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

Learning Outcomes

* Understand what the MapReduce function is used for
* Remember what the MapReduce function is
* Apply the knowledge of MapReduce to count a series of words from a story.
* Understand how it maps (splits up) the tasks and reduces (collates) them back together through the use of a wordcount example.
* Understand how a computer might count (specific) words in a book or a story.
* Analyse the need for MapReduce in a computer system.

Success Criteria

* I can understand the need for a MapReduce function when doing large menial tasks.
* I can see how a computer might use MapReduce in order to count specific words from a large section of text.

Time Required

1 period - 1 hour

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.
3. Make sure each group has a show me board and pens, and a timer of sorts.

Prior Learning Assumed

None - new topic

Outline of Activity

1. Split the class up into pairs, if there are any students left over just add them to a pair.
2. Outline to the students that you are going to be studying an important computing concept called MapReduce. It allows you to break up huge tasks into smaller more manageable ones in order to be completed quicker, and that it is very helpful in looking at how Google manages to search for the result you’re looking for in a very fast time. Clearly explain that this is how search can be done, i.e. students want to look up if there is a new book by their favorite author. So they search in a database which is stored on many computers, each computer searches within their fragment of the database (map) and returns the result, which is then summarised into a nice list (reduce)
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.
4. Say that a really slow and bad way of counting would to have one person go through the entire story and pick out the individual words and count them by themselves.
   1. Get the pair to identify one to be a timer, and one to be the counter.
   2. If the children know about tally marks, 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 go through their page, and count how many times the words appear, the timer should write down the time taken on the show me board and hold it up when they are done, for the teacher to note down times on the board. Keep all the times from this round on one side of the board, perhaps under a table heading of ‘pairs’.
5. Then explain that by using teamwork, you 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, three to be counters:
   1. It might be best to allocate pairs who have similar reading levels together, to avoid some students who are faster/slower readers changing the results.
6. Ask that they cut along the dotted lines, each member of the ‘counters’ 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 has 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. Once each group has 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 show (hopefully working by comparing the times) that it was much much quicker than one person counting up each occurrence of the word themselves, and this is how MapReduce works.