Overview

This exercise aims to teach children about the MapReduce computation that comes along with 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
- Understand why it is important that the MapReduce function exists
- 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

Success Criteria

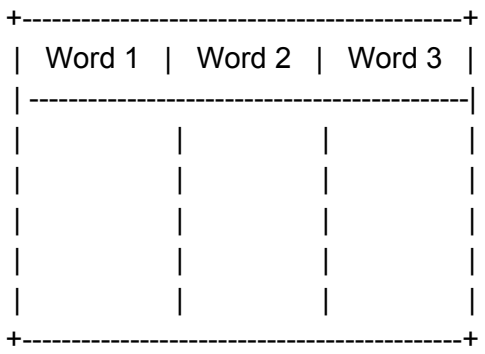
- I can understand the need for a MapReduce function when doing large menial tasks.
- I can see how a computer might use MapReduce in order to count specific words from a large section of text.

Time Required

1 period - 1 hour

Preparation

1. Printing out the story and laminating it, cutting the laminate up into separate lines but keeping them together on a board.
2. Having pieces of paper set up (one for each group) with the three words written in a grid like so:



So that students can stick their counted words in the bottom of the board.

3. Create a grid on the board similar to the one above, except with a ‘total’ row below to add up the totals.

Prior Learning Assumed

None - new topic

Outline of Activity

1. Split the class up into groups of three.
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. (Maybe don’t need to explain ‘MapReduce’ exactly - might be too complicated?)
3. Show them the story board and say that you want to count how many times the words ‘river’ ‘girl’ and ‘blue’ 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, perhaps go through this on the board yourself by picking out a helper from the class in order to demonstrate this fact.
5. Then explain that by using teamwork, you can split up this big task into a smaller tasks to solve it more quickly. Say that each member of the team is going to look for one specific word and by splitting up the story into separate lines we can find the number of each word quicker.
6. Give each group a section of the story and outline that they will have to each individually count the number of time each word appears. Make sure each member of the group knows what word they will be looking for.
7. If the children know about tally marks, outline that they should use the board on their desk to tally up each time they find the word in the story. Explain that this is called ‘Mapping’ and that this is the computer counting each time the word appears individually.
8. Leave them for about 5 minutes to count up each word, and then gather their attention again at the end.
9. Ask each group individually to say how many times they found each word in their section. Keep a count on the board from each group, with a ‘total’ underneath. Explain that this is ‘reducing’ - putting back together the split up totals each section has.
10. 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 that it was much much quicker than one person standing alone and counting up individually how many times each word appeared.