MAPREDUCE USING MPI & OPENMP WORDCOUNT

SURAJ SAJEEV(2018MCS2024)
AKHILESH KARRA(2015CS10233)

MAPREDUCE

A MapReduce program is composed of a map procedure,which performs filtering and sorting , and a reduce method, which performs a summary operation.

Map procedure : Distributing data(words) to different processes/threads.

Reduce method : Combining all the individual results(from diff processes/threads) into a single(final)result.

SEQUENTIAL

Read in the input file and store data in a char array.

Split using several stop words and store the words in a vector.

Create a hashmap of tuples(string,int).

Take each word from vector and map it into hashmap and update the count accordingly.

OPENMP

Here we create an array of hashmap whose size is equal to number of threads mapper[n].

Each thread gets a hashmap.

Mapping: distribution of data to each thread.

Each thread when receives a word,maps it into its corresponding hashmap.

Reduction: In critical section,we create a result hashmap,which takes data from each hashmap of mapper and maps it into itself accordingly.

MPI

Here we create hashmap for every process.

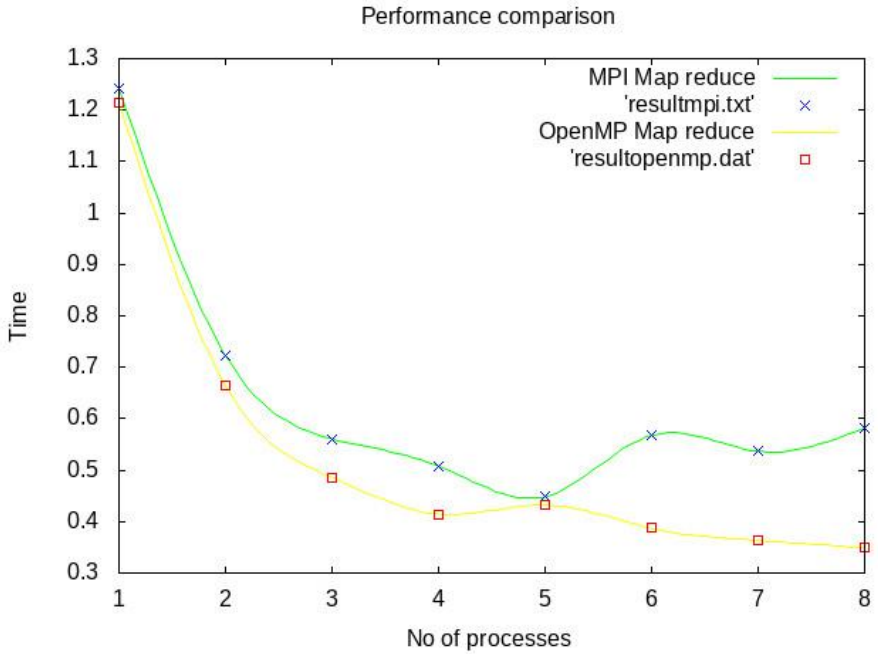
Mapping: Parent process distribute data to every other process.

Each process after receiving data,maps the data it into its corresponding hashmap.

Reduction: Every process then sends its result data to parent process.

Parent process receives data from individual processes and then combines by mapping into a single hashmap.

Timing comparison



Speedup comparison