Longest Path in a Set of Files - MapReduce Implementation in C++

Your Name(s)

December 17, 2024

1 Introduction

This report presents the implementation of a MapReduce framework to find the longest file path in a distributed set of files, using C++. The files are distributed across multiple laptops, and each file contains one path per line. The task is to identify the longest file path by length.

2 MapReduce Implementation

The system consists of two main components: the Mapper and the Reducer. The Mapper processes each input line (representing a file path), calculates its length, and outputs a key-value pair. The Reducer receives these pairs, compares the lengths, and outputs the longest file path(s).

2.1 Mapper

The Mapper performs the following steps:

1. Reads each line representing a file path.
2. Computes the length of the file path.
3. Emits a key-value pair, where the key is the length of the path, and the value is the file path itself.

2.2 Reducer

The Reducer performs the following steps:

1. Receives key-value pairs from the Mapper.
2. Compares the path lengths and keeps track of the longest path(s).
3. Outputs the longest path(s) after processing all input pairs.

3 Diagram of the Process

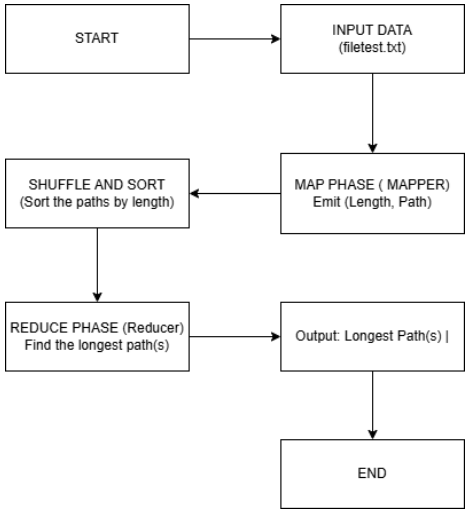

Figure 1: Flowchart of the MapReduce process for finding the longest path.

4 Conclusion

The C++ implementation of the MapReduce framework successfully identifies the longest file path by length in a distributed environment. This approach efficiently handles large datasets and provides an optimal solution to the problem.