IoT Big Data Processing

Lab work 1

Hadoop MapReduce

The assignment of the first lab session was to get familiar with Hadoop MapRaduce system. During the lab session we have installed Hadoop, followed the tutorial and saw how it works on some examples.

The last part of the first lab session was to write MapReduce program in Java by ourselves.

Task:

Make a recommendation engine for online stores. Basically, for each bought item we should make a list of items that are often both together with the current item.

Our example:

We have lists of bought items from two users, in two separate files:

file01: book12, book34, cd12, cd42, dvd32

file02: book32, book34, dvd32

Output should be: key-value pairs in which the key is the item and the value is the list of the items most commonly bought by customers who also bought this item.

The code is in /src folder.
The input files are in /src/ItemSuggestion/Input
The output is in /src/ItemSuggestion/output

The program is written in Java. And the code is in **ItemSuggestion.java** file.
The mapper and reducer functions have been implemented.
Mapper makes a key value pairs, where the key is bought item and the value is a string containing all items bought in combination with the one which is the key.
In reducer, we merge all strings with the same key and then we sort the items in that order that the items which are most often bought in the pair with current items come first. Also, when we do the sorting, we collect all items in string and output it as a value.

Output of this program looks like this:

book12	**book34 cd42 dvd32 cd12**
book32	**book34 dvd32**
book34	**dvd32 cd42 book32 book12 cd12**
cd12	**book12 book34 cd42 dvd32**
cd42	**book12 book34 dvd32 cd12**
dvd32	**book34 book12 cd42 book32 cd12**