

5<sup>th</sup> International Conference on

# Big Data Analysis and Data Mining

June 20-21, 2018 | Rome, Italy

## Big Data Analytics for Junior Citizens

**Lakshmi Prayaga**

University of West Florida, Department of Information Technology, Pensacola, Florida, USA

It is estimated that 2.5 quintillion bytes of data is being generated each day[Ref 1], and the need for data analysts to interpret this data is also increasing at an exponential rate[Ref 2]. It is this gap in the workforce that needs to be addressed to take full advantage of the power of big data analytics. It has been documented that early exposure to STEM disciplines influence young students' choice of their careers. Given this context, we worked within our local school district to introduce Big Data Analytics to junior students in their 10th grade Information technology course. Presented below are some pertinent points related to our research project –Big Data Analytics for Junior Citizens”.

The goal of this project was to inquire if it feasible to introduce the Big Ideas related to Big Data Analytics to Junior Citizens? If so what are the mechanisms of introducing these Big ideas. Specifically, were students able to appreciate the power of parallel processing and how Big data is generated. Student Demographics include 18 10th grade students. There were 7 girls and 11 boys. 30% of the students were African Americans and minority. The remaining students were Caucasian. The methods used were offering a refresher on java concepts by having students write programs that used reading and writing from and to files, arrays, lists, search and count features. Students followed up with writing a word count program in java using a sequential method. The same program was later written using the map reduce technique from Hadoop environment. Faculty mentor and a graduate student guided students through this process. Students finally used different datasets as files to frame their research questions, and write a sequential code that read data from a file, counted the number of instances (examples: how many flights were delayed between two dates, two locations, how many crime rates occurred in a particular zip code or state etc.) and display the results and the time it took to execute their code. Following this they wrote the program using the map reduce technique. We used the Hadoop cluster at the University of West Florida to run the code for the student projects.

Results and Observations include eight students completed the projects on their own with very little help from the faculty mentors. Six students needed some help in writing the final map reduce phase, and four students needed help both in the sequential version of java and the map reduce version. All students noticed that the map reduce technique executed faster than the sequential version of the java file. This also demonstrated the fact that execution speed was greater because of parallel processing used by the map reduce technique. Additionally, students were also able to suggest multiple examples of how Big data is generated from multiple sources such as multimedia, social networking, wearable devices etc. Students presented their projects to school district officials and employers. School district officials were also engaged in the presentation and asked questions such as “What is the one thing common in all your presentations?”. This prompted students to think and answer that it was the speed of execution using the map reduce technique. Employers also appreciated the project and were interesting in offering internships to interested students.

However, some of our observations include good working knowledge of problem decomposition and programming is very helpful in easily translating a sequential file to a map reduce technique based file. And it is possible to introduce Big ideas related to Big Data Analytics to Junior citizens following a structured design. We will be happy to share our experiences and projects with attendees at the conference.

lprayaga@uwf.edu