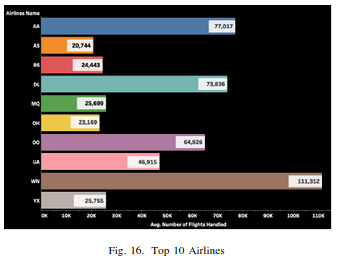
**Hadoop MapReduce in Aviation**

By: Saytu Singh

Hadoop and MapReduce is highly and widely used in the aviation industry. These two analytical tools that data scientists use are very critical when it comes to aviation operations and data collecting. Many people across the globe are booking tickets every millisecond and getting tickets from various airlines and carriers. But out of curiosity, how do we collect such kind of data and how is that data easily readable and delivered to people? And how does software such as Hadoop Mapreduce plays a significant role when it comes to data in the aviation industry? Suppose we want to know and find out what are the airlines that passengers prefer for to choose when flying within the United States itself, and the objective is to obtain the top ten airlines favored for travelling. This is where Hadoop Mapreduce comes to its advantage. TheMapreduce Java program was taken into consideration where the data set from HDFS is utilized. With the mapper, airline names can be taken as the input to the mapper which is obtained from a MySQL table. And the reducer is the code that contains the counts of the column as well as the filtration which provides the keywords that are “top.size” “10”, and that outputs the result of only top ten airlines that are being fetched. By depicting the output of the top ten airlines, the information would be useful for the airline industry increasing the number of flights for fulfilling customer demands.



Another problem on what the aviation industry also face, which many people and data scientists have to always resolve and gather data, are which factors cause significant delays for a specific time frame/month? As Big Data engineers, they use the ingestion tools to retrieve the big data on which time frame record has the most delays, compared to the times where there is the least amount of delays. With the aid of Hadoop Mapreduce, we are able to gather that piece of data. The mapper The mappper will display the four different types of delays that affects the aircraft performance are obtained. Then a program that does a word count is written for a column in a MySQL database labeled as the cancellation-code for the input. As for the reducer, the code displays the word count for each distinct alphabet present in the particular column. From the data image provided below, it is understood that major delays are cause by severe weathers, which contributes to 67.5% of the delays. The factor with the least affecting cause of delay is security delay.

Results like these are very important to the aviation industry to minimize or eliminate delays and reduce operating cost. And above all, other airline companies can understand their competitors strategy as well.

