

**MASTER'S PROJECT (SUMMER 2014)****TOPIC:** *Big Data Analysis of Airline Data for Flight Delay Statistics***PRESENTOR:** Safia A. Habib**ADVISOR:** Dr. Haiping Xu**DATE & TIME:** Wednesday, June 18, 2014, 11:00 AM**LOCATION:** Dion 303**COMMITTEE MEMBERS:** Dr. Ramprasad Bala and Dr. Jan Bergandy**ABSTRACT**

As air travel becomes one of the major public transportation methods, millions of people travel by commercial aircraft every year. However, for the airline passengers, delay in domestic flights has been a common phenomenon in the US. With big data analysis, it now becomes possible to efficiently mine the major factors that are related to flight delay by analyzing the large amount of poly-structured data generated in the airline industry. In this project, we collect the airline data from the Bureau of Transportation Statistics, which includes commercial flight information of 14 carriers in the US. We select the airline data during the recent three years (2011-2013) to illustrate our big data analysis approach for flight delay statistics. Our goal is to find the month with the most weather delay and the aircraft carrier with the maximum carrier delay. The challenge of extracting value from big data is in many ways similar to the age-old problem of distilling business intelligence from transactional data. Apache Hadoop has emerged as the de facto standard for managing big data. In order to efficiently process airline data and draw useful conclusions, we wrote Hadoop MapReduce jobs and deployed them on a cluster of 4 nodes to extract information from the raw airline data and performed analysis. As one of the major advantages of using MapReduce, it allows us to divide the airline data into multiple datasets and process them in parallel, without the overhead of programming for the complex details of intra-cluster communication, task monitoring and failure handling. Finally, we compare our big data analysis approach with the conventional sequential processing approach, and show that our approach is more efficient in drawing useful conclusions for flight delay.