ECE 590 Introduction to Big Data Technologies

This course covers the principles of the Big Data ecosystem architecture and provides students with in depth, hands-on experience by learning the installation and usage of the latest, most advanced, open-sourced Big Data technologies. Students will learn how to create and manage their own virtual machines (VM) in Azure and/or AWS cloud. Such scalable VM machine(s) will be used to build a version of the Big Data compute platform and use such compute environment to address use-cases that require analytics on varied datasets. More specifically, students will gain hands-on experience with the installation and use of Apache Spark, one of the most popular, general-purpose data analytical platform. Other technologies, such as Zeppelin and Jupyter notebooks will be installed and integrated throughout the class. Advanced topics in this class include learning about the Apache Spark Machine Learning (ML) libraries and how to build ML pipeline to address AI use-cases. Furthermore, through the integration of Anaconda3 into the Big Data architecture, students will learn how to install and use Tensorflow to build Deep Learning applications. Because the fore mentioned big data technologies are cloud ready, the experience gained can be directly applied toward any cloud computing environment that exists today.

Figure 1. 590 Big Data Architecture