Wyatt Blair

EMT-678-WS: Big Data Technologies

10/3/2024

**Class 3 - Read sections on “Transformations” and “Actions”**

This documentation details the available “Transformations” and “Actions” in Apache Spark. The first transformation and the first action defined are map and reduce, respectively. With these two functions alone, the user can make use of the MapReduce algorithm structure.

Beyond the standard map function, Spark has several other useful transformations defined such as filter, distinct, sample, union, intersect etc. With these useful functions, the user can collect data highly efficiently in memory and prepare the data for any desired action.

The Spark library also has a few additional actions defined beyond the reduce function. These functions include: collect, count, take, forEach etc. These functions are designed to use data loaded in memory and efficiently compute some desired output.