CSE 598: Data-Intensive Systems For Machine Learning

Report: Assignment 8
Dense Matrix Multiplication Comparison: Handcoded, Breeze, Spark

Muskan Singhal
1217107320

Testing Machine Hardware Configuration:

Model Name	MaccBook Pro
Processor Name	Intel Core i5
Processor Speed	1.4 GHz
Number of Processors	1 processor
Total Number of Cores	4 cores
L2 Cache (per core)	256 KB
L3 Cache	6 MB
Memory	8 GB

Performance Comparison:

As we can see, Handcoded Scala is the slowest as compared to Spark and Breeze.

	Handcoded Scala	Spark	Breeze
N (Total iterations)	1000 iterations	1000 iterations	1000 iterations
Total Execution Time	3835.63 seconds	19.70 seconds	32.43 seconds
Average Execution Time	3.83 seconds	0.019 seconds	0.032 seconds

Observations:

- Have implemented m*n matrix multiplication where m = 1000 and n = 1000. Also, matrix multiplication has been repeated 1000 times in scala, spark and breeze.
- After implementing and running the code, Scala is the slowest execution wise as it takes almost 3 seconds to multiply two matrices of size 1000*1000 whereas Apache Spark implemented using Scala takes almost 0.019 seconds for the same.
- When it comes to implementations, it is easier to code in Scala as compared to Apache Spark it requires us to take care of the spark context created. Even though matrix multiplication is fast in Spark, it takes longer time in overall execution.
- To work in breeze you need to first install sbt (Scala Build Tool) and run commands using sbt.

Things learnt:

Scala
- My take-away from the assignment was to learn how Scala works and how to install Apache Spark and Breeze locally in the machine.
- Scala is an object-oriented and a functional programming language and coding in Scala is very much similar to Java.

Spark
- Apache Spark is an open-source distributed general-purpose cluster-computing framework. It run workloads 100 times faster. We can write applications using Java, Scala, Python, R and SQL.
- Spark-shell is an environment where we can the run our Scala implemented code and see the output upon execution of the code. We can also specify the number of cores to be used by Spark for running our program.
- In Spark, we have to create a configuration which is used to set various Spark parameters as key-value pairs.

Breeze
- To install Breeze locally, we have to first install "Scala Build Tool" (SBT) to be able to run breeze using scala.
- Breeze is hosted on GitHub.
- Breeze is a library for numerical processing, machine learning and natural language processing. It aims to be generic, clean and powerful without sacrificing efficiency.
- It is the merger of ScalaNLP and Scalala Projects.
- It is a library for fast linear algebra and manipulation of data arrays, as well as many other features necessary for scientific computing.