Project 4 (Total points: 20)
Due time: Dec 15th, 2023 11:59pm

After you submit your project, **download it and make sure you submit it successfully and correctly**.

Reimplement the **Biological Data Counting** in Spark and run it on the department Spark cluster (spark.cs.ndsu.edu). The requirements are the same as the project 3. For graduate students, in your report please compare the run time of your Hadoop and Spark implementation and discuss other differences you find between these two.

Note: The deadline is a hard one. No late project is accepted because I also have a hard deadline for your final grade. How to calculate your final grade, please check out the syllabus. (You can also check out the "TOTAL" on the Blackboard.)