



# UE INF2245

## Hadoop Spark: programs on the AIS-UBS data-set

*Frédéric Raimbault*

Note: the following programs will take as input a file (at most a few files) located in the Amazon S3 directory `s3://ubs-datasets/ais/clean/2021`

1. Write a program (named it `SparkTop10Ships`) displaying on the screen the TOP10 fastest ships. Add some useful counters on the execution. Writes the result as a comment into the head of your Java source.
2. Write a program (named it `SparkShipMessages`) that determines the first and the last time every ship emit a message and its total number of sent messages. Add some useful counters on the execution.
3. Write a program (named it `SparkTypeCount`) that counts and displays the number of ships per ships type. Writes the result as a comment into the head of your Java source.