Database Systems
Exercise 2:

Background: In this exercise, you will implement OLAP database operators, columnar representation operators and distributed hash-join algorithm using scale-out framework Apache Spark. The exercise contains 4 tasks. Each task is described in a separate file. For the tasks that require Spark you may use - if needed - a VM containing an installation of Spark (download here, username: *user*, password: *user*, Spark installation under /home/user/spark).

Deadline: 30th of April @ midnight

Task 1 (20%). Implement the Skyline operator over Java.
- code should be implemented and tested **in Oracle Java** (jdk 7 or jdk 8).
- You are **not** allowed to use any third-party libraries.

Task 2 (30%). Columnar representation over Apache Spark

Task 3 (10%). Dictionary encoding on columnar representation over Apache Spark

Task 4 (40%). Implement distributed hash-join over Apache Spark.

Keep the following directory format for the answers:
/YourName/ExerciseNumber/TaskNumber/
e.g. for gaspar username "test" and exercise1 submit: /test/exercise1/task1/ etc.

Implementation specifications:
- Do not include test code in your submission; the submission will be timed and the checks will slow down your code.