

MSDS442

Project – Phase 5

Submission:

Submit your Project-Phase 5 as a SINGLE ZIP file on Canvas by the due date.

Your submitted ZIP file must have the name:

Project_Phase_5_Your_LastName.zip

Deliverables:

Your ZIP file for the exercise submission must include the following:

- All source code that you wrote, compiled and built on your personal computer.
- Panopto video recording of a live run of your code on your personal development computer.

Requirements specification:

Utilize **Kafka**, **Java**, **Python**, **SurpRise**, **TensorFlow Recommenders** and **Neo4j GDS** library to implement the following requirements based on the requirements specification and architecture document for the OnMart Superstore real-time data streaming application that you reviewed in Phase 1:

1. After you create the CSV log file for at least 10,000 real-time transactions, perform the following tasks on the data log
 - 1) Filter the products with more than 100 ratings.
 - 2) Filter the customers that have rated more than 10 products.
 - 3) Filter the products with more than 100 ratings in a certain city.
 - 4) Filter the customers that have rated more than 10 products in a certain zip-code.
 - 5) Exclude products that are in the top 5 highest product returns from the dataset used to build the recommender system.
2. Utilize popular recommendation algorithms from **SurpRise** and **TensorFlow Recommenders** packages in order to execute two experiments to develop and evaluate **two** recommender systems and provide a formal evaluation for the performance and product recommendation accuracy of these algorithms when running these algorithms on CPUs and GPUs; note that not all packages (algorithm implementation) are developed to run on GPU.