

MSDS442

Project – Phase 3

Submission:

Submit your Project-Phase 3 as a SINGLE ZIP file on Canvas by the due date.

Your submitted ZIP file must have the name:

Project_Phase_3_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**, **Facebook/Prophet** package 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. Generate at least 10,000 real time transactions for OnMart
2. Create the Neo4j Graph for OnMart logistics and supplychain network (State, City, Warehouse, Distribution Center, Zip Code)
3. Forecast the number of returned orders for every warehouse
4. Forecast the number of delayed deliveries for every zip code
5. Forecast the number of delayed deliveries for every distribution center
6. Identify top 10 distribution centers with highest number of order returns
7. Identify top 10 distribution centers with lowest number of order returns
8. Identify top 10 distribution centers with lowest number of deliveries
9. To offload busiest **warehouse** with highest number of **order returns**, identify the warehouse that has the **forecast** for the highest number of order returns, and find its nearest neighboring warehouse that has the lowest number of order returns (consider 50-200 mile radius)
10. To offload the busiest **distribution center** with highest number of delayed deliveries, identify the distribution center that has the **forecast** for the highest number of delayed deliveries, and find its nearest neighboring distribution center that has the lowest number of delayed deliveries (consider 1-20 miles radius)