**Business Problem:**

A FMCG company entered into the instant noodles business two years back. Their higher management has noticed that there is a mismatch in the demand and supply. Where the demand is high, supply is pretty low and where the demand is low, supply is pretty high. In both ways it is an inventory cost loss to the company; hence, the higher management wants to optimize the supply quantity in each and every warehouse in the entire country.

**Goal & Objective**: The objective of this exercise is to build a model, using historical data that will determine an optimum weight of the product to be shipped each time to the warehouse.   
Also try to analysis the demand pattern in different pockets of the country so management can drive the advertisement campaign particular in those pockets.

**File:** FMCG\_Data.csv

**Data Description:**

| **Variable** | **Business Definition** |
| --- | --- |
| Ware\_house\_ID | Product warehouse ID |
| WH\_Manager\_ID | Employee ID of warehouse manager |
| Location\_type | Location of warehouse like in city or village |
| WH\_capacity\_size | Storage capacity size of the warehouse |
| zone | Zone of the warehouse |
| WH\_regional\_zone | Regional zone of the warehouse under each zone |
| num\_refill\_req\_l3m | Number of times refilling has been done in last 3 months |
| transport\_issue\_l1y | Any transport issue like accident or goods stolen reported in last one year |
| Competitor\_in\_mkt | Number of instant noodles competitor in the market |
| retail\_shop\_num | Number of retails shop who sell the product under the warehouse area |
| wh\_owner\_type | Company is owning the warehouse or they have get the warehouse on rent |
| distributor\_num | Number of distributors works in between warehouse and retail shops |
| flood\_impacted | Warehouse is in the Flood impacted area indicator |
| flood\_proof | Warehouse is flood proof indicators. Like storage is at some height not directly on the ground |
| electric\_supply | Warehouse have electric back up like generator, so they can run the warehouse in load shedding |
| dist\_from\_hub | Distance between warehouse to the production hub in Kms |
| workers\_num | Number of workers working in the warehouse |
| wh\_est\_year | Warehouse established year |
| storage\_issue\_reported\_l3m | Warehouse reported storage issue to corporate office in last 3 months. Like rat, fungus because of moisture etc. |
| temp\_reg\_mach | Warehouse have temperature regulating machine indicator |

| approved\_wh\_govt\_certificate | What kind of standard certificate has been issued to the warehouse from government regulatory body |
| --- | --- |
| wh\_breakdown\_l3m | Number of time warehouses face a breakdown in last 3 months. Like strike from worker, flood, or electrical failure |
| govt\_check\_l3m | Number of time government Officers have been visited the warehouse to check the quality and expire of stored food in last 3 months |
| product\_wg\_ton | Product has been shipped in last 3 months. Weight is in tons |

**MapReduce Problem Statements**

Here are specific MapReduce problem statements that can be solved using MapReduce streaming and Python programming. Each problem statement includes the objective, the dataset fields required, and a brief description of how to approach the problem using MapReduce.

**Task 1: Demand-Supply Mismatch Analysis**

**Objective:** Identify zones and regional zones with the highest mismatch between demand and supply.  
Required Fields: zone, WH\_regional\_zone, product\_wg\_ton

Description:

Map: For each warehouse, emit the zone and regional zone as the key and the product weight shipped in the last three months as the value.

Reduce: Aggregate the product weight by zone and regional zone to calculate the total supply. Compare this with known demand data to identify mismatches.

**Task 2: Warehouse Refill Frequency Correlation**

**Objective:** Determine the correlation between warehouse capacity and refill frequency.  
Required Fields: WH\_capacity\_size, num\_refill\_req\_l3m

Description:  
Map: Extract the number of refill requests (num\_refill\_req\_l3m) and warehouse capacity size (WH\_capacity\_size) for each warehouse. (For each warehouse, emit the capacity size and the number of refill requests as the value)  
  
Reduce: Aggregate the refill requests by capacity size and calculate the correlation.

**Task 3. Transport Issue Impact Analysis**

**Objective:** Analyse the impact of transport issues on warehouse supply efficiency.

Required Fields: transport\_issue\_l1y, product\_wg\_ton

Description:

Map: For each warehouse, emit whether a transport issue was reported and the product weight

shipped.

Reduce: Aggregate the product weight by transport issue status to assess the impact.

**Task 4. Storage Issue Analysis**

**Objective:** Evaluate the impact of storage issues on warehouse performance.

Required Fields: storage\_issue\_reported\_l3m, product\_wg\_ton

Description:  
Map: For each warehouse, emit whether a storage issue was reported and the product weight shipped.

Reduce: Aggregate the product weight by storage issue status to assess the impact.