

**Data Warehousing – Lab 1**



**Group members :**  
Huy Pham  
Vivian Nakaya  
Kushmi Anuththara

# **Introduction**

In this lab, we explored a Data Warehouse (DW) containing sales data for a grocery store. The main goal was to understand and analyze the data using pivot tables, applying various concepts like roll-up, drill-down, slicing, and pivoting. The dataset follows a star schema, with one fact table (Sales Fact) and four dimensions (Time, Product, Store, and Promotion). Our task was to answer specific business-related questions by analyzing the data.

# **Task Results and Analysis**

## **1. How many products did Store No. 19 sell in 1994?**

After analyzing the dataset, we found that **Store No. 19** sold a total of **14,016 units** in the year 1994. This was calculated by summing up the unit\_sales for this specific store and year.

## **2. Which type of promotion provides the highest amount of unit sales?**

The promotion with the highest unit sales is **"No Promotion"**, with a total of **399,010 units sold**. This indicates that even without specific promotions, certain products or customer preferences drive high sales volumes.

## **3. Which are the top 10% best stores in terms of dollar sales in 1995?**

To identify the top 10% best-performing stores in 1995 based on dollar sales, we calculated the 90th percentile threshold and identified stores exceeding it. The top stores are:

* **Store No. 8**: $21,563.91
* **Store No. 12**: $22,647.59

These stores significantly outperformed others in terms of revenue.

## **4. Which one of the stores has the highest amount of customers?**

The store with the highest number of customers is **Store No. 8**, serving a total of **21,843 customers**. This reflects either a larger customer base or higher foot traffic compared to other stores.

## **5. Which product is the most lucrative?**

The most lucrative product in terms of profit (calculated as the difference between dollar sales and dollar cost) is **Buffalo Jerky**, generating a profit of **$11,022.52**. This highlights its high profitability and possible popularity among customers.

## **6. What was the most lucrative day, month, and year?**

* **Day**: December 10, 1994, with a profit of **$899.97**.
* **Month**: November, with a profit of **$37,367.45**.
* **Year**: 1994, with a total profit of **$54,282.79**.

These results indicate significant sales spikes during the holiday season and overall strong performance in 1994.

# **Conclusions**

**Advantages of Using a Data Warehouse:**

1. **Centralized Data:** Combines data from multiple sources for consistent analysis.
2. **Flexibility:** Allows slicing and dicing of data across different dimensions (e.g., time, product, store).
3. **Scalability:** Handles large datasets effectively, enabling detailed insights.
4. **Ease of Use:** Tools like pivot tables simplify data exploration and visualization.

**Challenges and Limitations:**

1. **Data Complexity:** Requires familiarity with schema designs and tools.
2. **Static Structure:** Pre-defined schema may limit dynamic data analysis.
3. **Cost:** Implementing and maintaining a DW can be expensive.

Overall, the Data Warehouse provides invaluable support for decision-making by enabling quick and detailed data analysis. This lab helped us understand how to extract meaningful business insights using structured datasets and pivot table functionalities.

# **Recommendations**

For future analyses, focusing on trends in promotions and regional sales could offer further actionable insights. Additionally, combining external data sources (e.g., market trends, competitor analysis) with the DW could enhance the decision-making process.