Northeastern University
College *of* Engineering

# Data Warehousing & Integration
# IE 6750
# FALL 2024

# [Formula 1 Data Pipeline]

**Milestone 5**

Group 10


Avani Kala

Sri Sai Prabhath Reddy Gudipalli

gudipalli.s@northeastern.edu

kala.a@northeastern.edu


**Submission Date: 11/14/2024**

# Problem Statement

Formula 1 racing is a data-rich sport with a long history. Despite this wealth of information, extracting meaningful insights to improve team performance, fan engagement, and strategic decision-making can be challenging.

**Goal:** To create a centralized data warehouse that integrates and harmonizes F1 data from raw data sources, enabling comprehensive analysis and unlocking valuable insights to enhance team performance and strategic planning.

Data Sources

We will utilize 2 separate data sources for our project. We will use FastF1 API to collect real-time data about drivers and races, and the following Kaggle link for historical data.

- Fast F1: https://docs.fastf1.dev/
- Kaggle link: https://www.kaggle.com/datasets/rohanrao/formula-1-world-championship-1950-2020

# Facts and Dimensions

Here is the Fact table:

- Contains the core data about Formula 1 races, including:
    - driverID (foreign key to driver dimension)
    - raceID (foreign key to race dimension)
    - constructorsID (foreign key to Constructor dimension)
    - countriesID (foreign key to Countries dimension)
    - circuitID (foreign key to Circuits dimension)
    - timeID (foreign key to Time dimension)
    - position
    - point

**Dimension Tables**

- **Driver:** driver key, reference, number, forename, surname, date of birth, nationality, code.
- **Circuit:** circuit key, reference, location, country, latitude, altitude, and name.
- **Constructor:** constructor key, name, nationality, and constructor reference.
- **Country:** country key and country name.
- **Time:** race time, date and year.
- **Race:** race key, name and round.

# Strategic and Operational Insights

By leveraging the data warehouse, we can gain valuable insights to optimize team performance and inform strategic decision-making. Some potential insights include:

**Strategic Insights:**

- **Historical Performance Analysis:** Identify trends in team and driver performance over time
- **Circuit-Specific Strategies:** Analyze past performance at specific circuits to develop tailored strategies for each race.

**Operational Insights:**

- **Real-Time Race Strategy:** Use real-time data to make informed decisions during races, such as pit stop timing, tire strategy, and overtaking maneuvers.
- **Driver Performance Analysis:** Evaluate individual driver performance metrics to identify areas for improvement and optimize driver pairings.

# Final Remarks

By harnessing the power of data and advanced analytics, Formula 1 teams can gain a competitive edge, improve decision-making, and ultimately achieve greater success on the track.