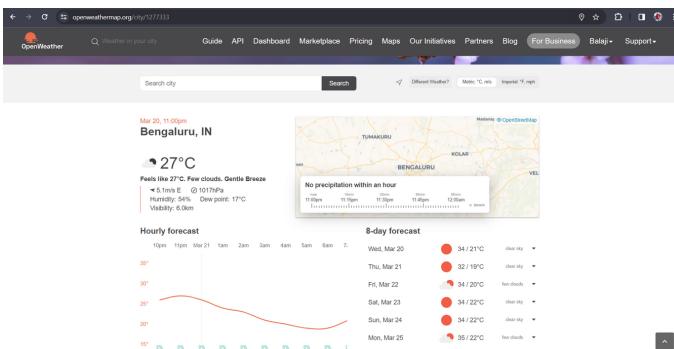# Automated Weather Data Pipeline

Developed a comprehensive automated data pipeline for real-time weather data acquisition and analysis. Leveraged Python (BeautifulSoup, pandas) for web scraping and data manipulation. Integrated PySpark for efficient processing of large datasets and Snowflake for scalable storage. Implemented dynamic schema creation, reducing manual efforts in data structuring. Provided structured, query-ready datasets facilitating advanced weather analytics for informed decision-making. Enabled seamless workflow from data collection to analysis, supporting various weather-dependent applications and industries.

Data Source & Data Stored:

ETL Process Design Diagram