# Week 8 Contribution - Tanmay Munigala – 12742995

## ETL Pipeline Requirements

1. Utilize Node and Express to write a spatial data ETL.

2. The project’s ETL target sources are mainly HTTP, FTP, raw files and end imports

3. The ETL’s targets are Postgres, which is running on the team’s database as a RDS instance in Amazon AWS.

4. Error handling is one of the main challenges. Using typescript will make it easier to identify issues due to compiler level error handling but will require more time as both source and target data types will have to be explicitly defined.

5. Scheduling and scalability are the main concerns.

In terms of scheduling, running daily a job every 24 hours will be sufficient for our project’s use case to keep the data in our database up to date.

Scalability wise, the ability to handling a minimum of 1 GB on each scheduled run of the ETL pipeline will be required as the data sources will increase with time.