**Formation of New Clean Attendance system**

**Candidate Name:** Manikandan Ganesh

**Email:** [karthickgm27@gmail.com](mailto:karthickgm27@gmail.com)

**Phone:** 312-874-8660

**LinkedIn:** <https://www.linkedin.com/in/manikandan-ganesh-02919963>

**GitHub:** <https://github.com/manikandanganesh2791>

**Objective**

The purpose of this task is to create a clean attendance record dataset which is likely to be loaded into a new sole system.

The word” **clean**” indicates three things here:

1. The final data set must be simple and all the department in the company should find it easy to use.
2. The structure of my design must be flexible so that, if the business logic changes, the design can be changed easily.
3. The solution must open doors for more data.
4. The data set must be consistent across all the regions.

Making these assumptions, I decided to work on the given unclean data set from different systems (c, k, s) and transform them into a single clean data set satisfying the **ACID** properties in the context of database.

**Data Cleaning:**

I first started off by looking at the data set of all the given three different systems and I could observe the following inconsistency and unclean records in the data sets.

Records from the Attendance\_S system had the following structure: