

Original data

	user	timestamp	hours	project
notes	One-to-many One user can have multiple projects they are working on	Start time for when a job log was made Different time formats	How many hours a user worked on a project	Many-to-many A project could be worked on by many people

How we can clean the data

- User
 - redundant data in the table
 - we can make the data cleaner by using a particular id associated with each user
- Timestamp
 - recorded different datetime formats
 - we can make the data cleaner by using the same time format throughout our data
- Hours
 - we can make the data cleaner by computing for the endtime using this field, that way, no additional computations needed when we need to present data
- Project
 - same with user

New database schema

users

Column name	Column type	Remarks
id	bigint	Primary key of the user
username	varchar	Mapped to the original “user” column

projects

Column name	Column type	Remarks
id	bigint	Primary key of the project
projectname	varchar	Mapped to the original “project” column

worklogs

Column name	Column type	Remarks
id	bigint	Primary key of the job log
userid	bigint	Foreign key - user that logged the job
projectid	bigint	Foreign key - project associated with the user’s job
starttime	datetime	Starttime of job
endtime	datetime	Endtime of job