ETL Project Hints

This document has some hints that will help you while you are working on the ETL project.

The following table will help you in understanding the relationships between the various Parent and Child Tables that you will have to work within this project, along with their corresponding Foreign and Natural Keys.

Parent Table	Foreign Key	Child table	Natural Key
fact_atm_trans	'year', 'month', 'day', 'hour', 'weekday'	dim_date	'year', 'month', 'day', 'hour', 'weekday'
fact_atm_trans	'card_type'	dim_card_type	'card_type'
fact_atm_trans	'location', 'lat', 'lon', 'streetname', 'street_number', 'zipcode'	dim_location	'location', 'lat', 'lon', 'streetname', 'street_number', 'zipcode'
fact_atm_trans	'atm_number', 'atm_manufacturer', 'location_id'	dim_atm	'atm_number', atm_manufacturer', 'location_id'
dim_atm	'location', 'streetname', 'street_number', 'zipcode', 'lat', 'lon'	dim_location	'location', 'streetname', 'street_number', 'zipcode', 'lat', 'lon'

Note: Here, these relationships have been used to avoid redundancy by normalising the data. To match the record count for this assignment, please apply this left outer join in the parent-child relationship.

Following are some additional hints which will help you in the project:

1. Join the Foreign keys of the base RDS table with the natural keys of the various Dimension tables in order to fetch the surrogate keys into the Fact table.
2. In case of a Parent-Child relationship between the dimension tables, the Parent table has to be populated before the Child table.
3. In case of a Parent-Child relationship between the dimension tables, the Child table should be deleted first before deleting the Parent table.
4. The joins between the Parent and Child tables should be a Left Outer join so that the Parent data is not lost.