PURPOSE: At each stages of database changes there were script additions and changes which still persists to final product. This document is to allow future developers to understand why certain scripts are derived, and purpose they serve.

‘Starhub’ table is the first and only database we planned.

Initial plan was to have 6 columns:

1. S\_ID (Primary Key)
2. UNIX\_TIME (Primary Key)
3. SESSION (Unused)
4. LOCATION
5. LATITUDE
6. LONGITUDE

The script ‘populating master table.py’ was created to upload the 4GB master data into the database (see script definition for more information).

However it is determined that with 7 million rows of data, any select query based on unix\_time and / or location will take at least 3 minutes to run. With 74 millions rows it will be too inefficient for our application.

Thus it is decided to create a denormalised table ‘movement’ (our informal OLAP cube), and a lookup table ‘coordinates’, stripping column latitude, longitude from ‘starhub’ table.

This is when ‘populating movement table.py’ is created to populate the movement table from ‘starhub’ table

‘Coordinates’ table only had 2 columns:

1. LOCATION (primary key)
2. COORD (Concatenation of both latitude and longitude separated by comma)

This was thought to be enough, but the two coordinates was separated. The separation and filling of the new columns was done using ‘convert COORD to LatLon.py’ script.

‘Movement2’ is full 2.3 million movements generated from ‘starhub’ raw data. ‘Movement’ contains only a subset for faster demonstration / testing purposes, and contains a lot of missing uncaptured movements.

The above conclusion persisted until the current version.

To create a redundancy for future purposes all available 314 locations was saved into ‘locations.txt’.

‘Populating location table.py’ was created to read ‘locations.txt’, lookup lat lon, and populate.