

Distributed Subtrajectory Clustering > [download here](#)

An open source implementation of the solution proposed in [1].

Trajectory clustering is an important operation of knowledge discovery from mobility data. Especially nowadays, the need for performing advanced analytic operations over massively produced data, such as mobility traces, in efficient and scalable ways is imperative. However, discovering clusters of complete trajectories can overlook significant patterns that exist only for a small portion of their lifespan. Here, we address the problem of Distributed Subtrajectory Clustering (DSC) in an efficient and highly scalable way.

Implementation Details

This is a MapReduce solution in Java that has been implemented and tested against Hadoop 2.7.2. The only external library used here is [cts](#) for coordinate transformation.

Input Data

The input is an hdfs directory containing csv files (comma delimited) of the form <obj_id, traj_id, t, lon, lat>, where

- obj_id and traj_id are integers (traj_id might be omitted if not available)
- t is an integer corresponding to the unix timestamp
- lon and lat are the coordinates in WGS84

[>> MORE INFO](#)