

Trajectory ETL Modeling

Assawer Zekri^{1(✉)}, Marwa Massaâbi¹, Olfa Layouni¹, and Jalel Akaichi²

¹ Computer Science Department, BESTMOD, High Institute of Management,
University of Tunis, Tunis, Tunisia
assawer.zekri@gmail.com, massaabi.marwa@gmail.com,
layouni.olfa89@gmail.com

² College of Computer Science, King Khalid University, Abha, Saudi Arabia
jalel.akaichi@kku.edu.sa

Abstract. Extraction-Transformation-Loading (ETL) tools are pieces of software responsible for the extraction of data from heterogeneous sources, their cleansing, customization and insertion into data warehouse. In fact, ETL are the key component of data warehousing process because incorrect or misleading data will produce wrong business decisions. Therefore, a correct design of the ETL process at early stages of data warehouse project is absolutely necessary to improve data quality. So, it is essential to overcome the ETL modeling phase with elegance in order to produce simple and understandable models. Many researches are dealt with the modeling of ETL process but today with the advent of trajectory data we need interactive modeling that manages the mobility aspect. In this paper, we extend existing model by incorporating a trajectory as a first class concept. So, we propose to model the ETL workflows as directed acyclic graphs that comprise three main components which are: the data sources, all the activities and the data stores.

Keywords: Trajectory data · ETL · Modeling · Graph · Activity · Moving object

1 Introduction

Extraction Transformation Load (ETL) plays an important role in data warehouse process. In fact, extraction is the retrieval of data from various sources. The second important phase in the ETL is Transformation. It is the change that must be made on the data source before it's loading on the data warehouse. The last phase is loading which aims to load the cleaned and transformed data to the warehouse. To summarize, the overall process consists of three major steps which are the identification of the source, the cleaning and transformation of data and finally moving resultant to the target location. In the last few years there has been a growing interest in trajectory data generated by sensors, mobile computing techniques and positioning technologies, there becoming available in massive and huge quantities. They describe the mobility of diversity moving objects, such as people, vehicles, animals, ships and airplanes. They are characterized by their complex structure including spatiotemporal attributes that values represent a data stream generated in a continuous way and at high speed. So, the vastness of data