Available online at www.sciencedirect.com

ScienceDirect

Procedia Environmental Sciences 33 (2016) 626 – 634

Procedia Environmental Sciences

The 2nd International Symposium on LAPAN-IPB Satellite for Food Security and Environmental Monitoring 2015, LISAT-FSEM 2015

Extraction, Transformation, and Loading (ETL) module for hotspot spatial data warehouse using geokettle

Winda Astriani, Rina Trisminingsih*

Department of Computer Science, Bogor Agricultural University, Kampus IPB Darmaga, Bogor 16680, Indonesia

Abstract

Spatial data warehouse technology is one solution to the problem of big spatial data. Accumulation In the process of making spatial data warehouse, extraction, transformation, and loading (ETL) process has an important role to determine the quality of data. Manual ETL process requires a long time and makes a lot of queries. Therefore, this research uses Geokettle as a spatial ETL tool to integrate spatial data. This research used hotspot dataset of Indonesia from 2006 to 2014 and administrative districts data in Indonesia. This research performed ETL modeling with the simplification, adjustment, and design of ETL scenarios. The result of this research is ETL modeling implemented using Geokettle. SpagoBI Studio was used to create multidimensional data cubes. Moreover ETL testing was conducted using Geokettle, and spatial data warehouse testing was done by comparing the total number of hotspots between SQL query result and spatial analysis hotspot result on Quantum GIS.

Peer-review under responsibility of the organizing committee of LISAT-FSEM2015

Keywords: ETL; hotspot; spatial data warehouse.

1. Introduction

Forest is a natural resource potential and has an important role on earth. Serious problems experienced by Indonesia in forest management is forest fires. One of indicators used as the possibility of forest fires is hotspots. There are several satellites and remote sensing systems that can be used to monitor hotspots from the sky. Sensors that are used to monitor hotspots occurence, especially in Indonesia are NOAA-AVHRR sensor and Terra-MODIS

* Corresponding author. Tel.: +6281-2986-0241.
E-mail address: rina.ilkomipb@gmail.com.

Peer-review under responsibility of the organizing committee of LISAT-FSEM2015
doi:10.1016/j.proenv.2016.03.117

sensor [1]. Both sensors are manufactured by the US space agency (NASA), which provides data in realtime hotspot.

Hotspot data contains a dimension of time and location, consequently data collected would be big datasets, Spatial data warehouse technology is one solution to the problem accumulation of big spatial data. In the development of spatial data warehouse, the process of extraction, transformation, loading (ETL) has an important role. ETL process is a cornerstone of a data warehouse. An ETL design well will extract the data from the source systems, maintain data quality, applying standard rules, and presenting data in a variety forms that can be used in the decision-making process [2]. In this research, the ETL process automatically using spatial ETL tool that supports vector geometry data types and provides a consistent data integration that is Geokettle. In addition to the preprocessing of data, this study also makes a model that aims to design ETL scenarios, customize and simplify the mapping between the attributes in the data source with the attributes of the data warehouse tables. ETL application of using Geokettle expected to facilitate data warehouse developers in performing preprocessing data automatically that allows regulate the insertion of new data and update data without generating a lot of queries.

2. Data and research steps

2.1 Data

The data used in this research is hotspot datasets from 2006 to 2014, as the geometric vector data in with shapefile format (.shp). This hotspot data can be obtained at http://firms.modaps.eosdis.nasa.gov. Meanwhile the administrative district data provided by Geospatial Information Agency (BIG).

2.2 Researh Steps

The research step can be seen in Fig. 1.

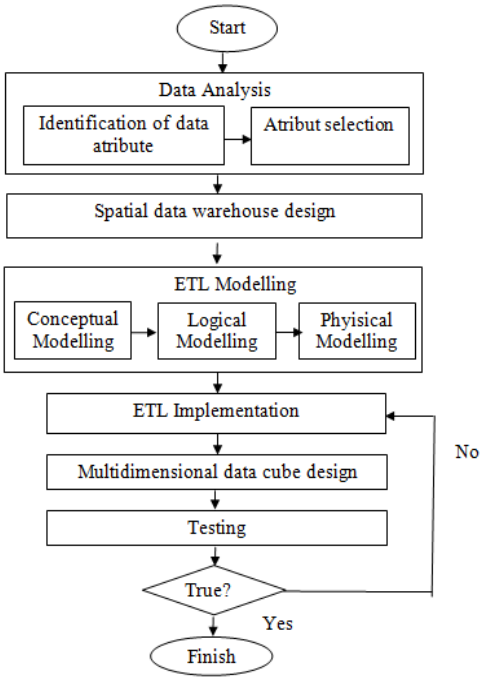

Fig. 1 Research steps