

Dashboard Project: Milestone III

Data Warehousing and Business Intelligence - Computing Lab

Ivan Vallejo Vall, Saurav Poudel & Daniel Velásquez Vergara

November 11, 2016

1 Overview

The dataset for this project is composed of telecommunication indicators from the International Telecommunication Union (ITU), World Bank (WB) macroeconomic data, and Google broadband pricing data.

ITU and WB data contain one observation per country, year and indicator. Google data include multiple observations per year and country, i.e. all the broadband plans offered by the operators in the given country for the given year.

Before being transferred into the database, data have been processed according to the schema in Figure 1. Data processing included cleaning incomplete or erroneous records, normalizing the data (e.g. from local currency to USD, speeds and caps into common units) and formatting (e.g. coding of blank records, treatment of mixed variables including numeric and text values).

1.1 Database dscription

Figure 2 shows the structure of the database. It is made of eight tables that contain the data obtained from the three different sources:

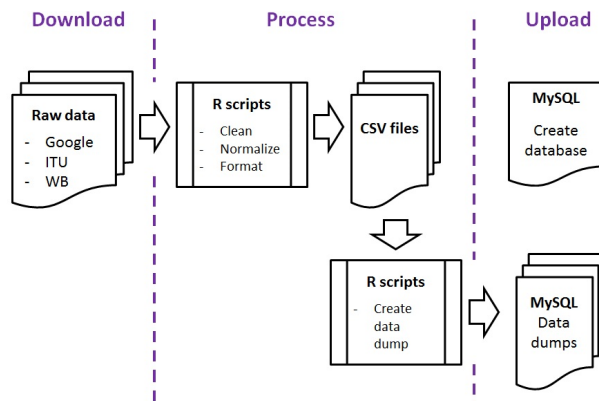


Figure 1: Data processing diagram

- The macroeconomic indicators by country sourced from the World Bank open data website <http://data.worldbank.org> are structured into two tables in the database. On the one hand, the table `wb_indicators` contains the identification code and definition of 1449 macroeconomic indicators. On the other hand, the table `wb_data` lump together data for 695 indicators from 211 countries, between the years 2000 and 2015.
- ITU data are organized in two tables: `itu_data` contains the indicator value per year per country (three indicators, 228 countries, 2000-2015), and `itu_indicators` the indicator code and name.
- Google's price data are stored into two separate tables: `mobilebb_prices` (11906 plans, 171 countries, period 2012-2015) and `fixedBB_prices` fixed-broadband prices (7254 plans, 153 countries, period 2012-2015), since the descriptive data for the two types of plans differ. For instance, mobile-broadband prices include a variable on the type of contract (prepaid or postpaid), whereas all fixed-broadband prices correspond to postpaid plans, so no variable is needed.
- An additional table (`LCCs`) indicates which of the 228 ITU countries were classified as "least connected countries (LCCs)" in 2008, 2010, 2011, 2012, 2013 and 2015. Countries are either LCCs, not LCCs or unknown (NULL). The latter are the ones to be classified as LCCs or not.

1.2 Available Information

All the scripts related to the project are available at the following repository: https://github.com/veldanie/dwbi_project. The folder `data` includes the data processing programs. On the other hand, the folder `db` contains all the *dapa dump files* and the script `telecom.sql` that creates the database and the required tables.



Figure 2: Telecom Database