Data Warehousing Assignment—Part III

Toon Calders Hatem Haddad

Deadline: 21 December 2016

1 Practical information

Deadline:	21 December 2016
Groups:	Same groups as for assignment parts I and II
How to submit:	Upload solution at `uv.ulb.ac.be`

Please read this complete document carefully before starting your assignment. In case of doubts, after carefully reading the document, do not hesitate to contact the lecturer. The complexity of the assignment is intended to be in the execution of it, not in its interpretation.

2 Objectives

The goals of this assignment is creating ETL scripts for updating the database in SSIS.

3 Assignment

In order to test the benefits of the data warehouse, the company is asking for a proof of concept, based on 4 historical snapshots of the database that can be taken from backup. Your task is now to define a SSIS package to incrementally load these snapshots into the data warehouse. In the assignment package you will find 4 snapshots (in directory "snapshots"; each snapshot corresponds to a specific day: 04 December 2016, 05 December 2016, 06 December 2016 and 07 December 2016). The snapshots represent the state of the database at one point in time. As such, next to new data, the snapshots will also contain customers, staff, stores, rentals, etc. that were already in earlier snapshots or even in the original load. Therefore, in order to correctly update the content of the datawarehouse based on a snapshot, you first need to figure out what has changed since the last snapshot. In order to facilitate this operation, in the tables Customer, store and staff, you have to use the attribute "last_update" that contains the last date on which the tuple was changed. Table rental already has a "rental_date" attribute indicating when the rental was entered. Table payment already has a "payment_date" attribute indicating when the payment was entered. These attributes can be used to correctly set the valid times for the versioned dimensions in the data warehouse and to test which tuples are new. The schema of the snapshot databases is available in Figure 1. The Table Payment of snapshot 07 December 2016 is empty.

4 Deliverables

You should deliver the following elements.

1. A report (as a.pdf), containing (length indication is purely indicative):
   (a) A **cover page** with the list of group members, including student ID,
   (b) Figures showing all your data flows and control flow with a succinct explanation whenever needed (length depends on your ETL flow)
2. The ETL package for the incremental load of the data into the data warehouse that it should load the snapshots one by one and the package should end when all snapshots have been loaded.

Submit both files as a single .zip or .rar file on the université virtuelle course website.

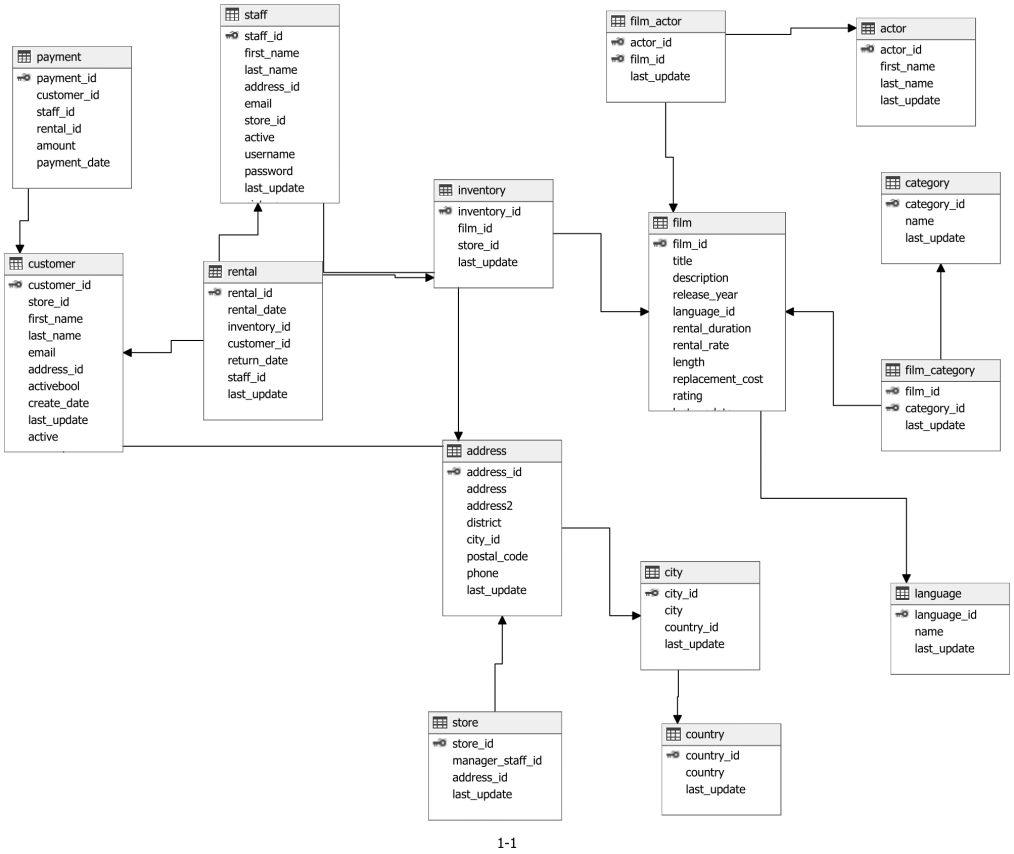

Figure 1: Schema of the snapshot databases