

Master thesis : On the Design and Implementation of an ETL Configuration DSL for Non-programming Experts

Auteur : Duchateau, Jakub

Promoteur(s) : Debruyne, Christophe; 17378

Faculté : Faculté des Sciences appliquées

Diplôme : Master en sciences informatiques, à finalité spécialisée en "computer systems security"

Année académique : 2022-2023

URI/URL : <http://hdl.handle.net/2268.2/17649>

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.

University of Liège
School of Engineering and Computer Science

On the Design and Implementation of an ETL Configuration DSL for Non-programming Experts

Jakub Duchateau

Promotors: Christophe Debruyne, Frédéric Duquenne

Academic year 2022-2023

8 June 2023

Abstract Extract-transform-load (ETL) tools are used in the business to ingest external data into their operational database. The design of an ETL process can be achieved through either code and libraries or a graphical tool featuring a graph of operations. However, the former is only accessible to programmers, while the latter lacks extensibility. This thesis proposes an approach to ETL configuration, based on a projectional domain-specific language (DSL), implemented with JetBrains MPS, and targeting ETL.NET. The ETL process is designed using a language that represents data as tables manipulated through sequences of operations. A prototype implementation is evaluated with a small user study, which shows that the DSL is accessible to non-programmers, and programmers prefer it over a tool based on graphs.

Résumé Les outils Extract-Transform-Load (ETL) sont utilisés dans les entreprises pour importer des données externes dans leur base de données opérationnelle. La conception de processus ETL, se fait principalement avec du code textuel et des bibliothèques logicielles ou avec des interfaces graphiques présentant un graphe d'opérations. La première convient bien aux programmeurs tandis que la seconde est plus difficile à étendre en fonctionnalité. Nous proposons ici une méthode intermédiaire basée sur un langage de programmation spécifique (DSL) projectionnel, implémenté avec JetBrains MPS, et intégré à ETL.NET. Le processus ETL est exprimé dans un langage qui modélise les données sous forme de tableaux qui sont transformés avec des séquences d'opérations. Le prototype est évalué avec une petite étude utilisateur, qui montre que le DSL est accessible aux non-programmeurs, et que les programmeurs le préfèrent à une interface basée sur des graphes.