TALLINN UNIVERSITY OF TECHNOLOGY

Faculty of Information Technology

Department of Informatics

IDU70LT

Alla Tšornenkaja IAMP132383

DATA LINEAGE TRACING PROCESS IMPROVEMENT IN FINANCIAL INSTITUTION'S DATA WAREHOUSE

Master's thesis

Supervisor:	Eduard Ševtšenko
	Ph.D.
	Associate Professor
Co-supervisor:	Igor Artemtšuk
	M.A.
	Developer

Tallinn 2016

TALLINNA TEHNIKAÜLIKOOL

Infotehnoloogia teaduskond

Informaatikainstituut

IDU70LT

Alla Tšornenkaja IAPM132383

ANDMETE ELUTSÜKLI JÄLGIMISPROTSESSI PARENDAMINE FINANTSETTEVÕTTE ANDMEAIDAS

Magistritöö

Juhendaja:	Eduard Ševtšenko
	Ph.D.
	Dotsent
Kaasjuhendaja:	Igor Artemtšuk
	M.A.
	Arendaja

Tallinn 2016

Summary

This thesis presents one possible solution for the financial data lineage tracing in the existing Enterprise Data Warehouse environment in the Financial Institution N. Current research gives an overview of different lineage types and existing prototype solutions presented in other scientific works, and formalizes the possibilities and constraints of applying current approaches in the specific data warehousing environment. The benefit of this work is to simplify the impact analysis for financial metrics calculated in the data warehousing environment.

The main task of this thesis is to research the existing data lineage tracing approaches and develop the prototype solution as well as possible.

As the result of this thesis, we have adopted some ideas found in other research efforts as well as proposed our own solutions for appearing constraints, and implemented them in the data lineage tracing prototype. We have created the prototype solution based on the requirements provided, and identified the metadata usage constraints, which should be improved. The prototype solution has been tested according to the requirements and the outcome has been described in details. Additionally, a graphical visualization has been done in order to give a better overview of the current work contribution.

We have suggested the different directions for the future prototype improvements, among which the most important at the moment is to supplement the missing relations for the ETL processes to the corresponding objects.

Current prototype solution will be developed further within the organization, and therefore, the complexity of the pattern can grow over time. The final aim is to have a universal pattern which will be applicable for every classifier lineage tracing.