

Insert here your thesis' task.



**FACULTY
OF INFORMATION
TECHNOLOGY
CTU IN PRAGUE**

Master's thesis

Object-relational mapping for database access in JavaScript

Bc. Ladislav Louka

Department of Software Engineering

Supervisor: Ing. Jan Matoušek

March 5, 2023

Acknowledgements

THANKS (remove entirely in case you do not wish to thank anyone)

Declaration

I hereby declare that the presented thesis is my own work and that I have cited all sources of information in accordance with the Guideline for adhering to ethical principles when elaborating an academic final thesis.

I acknowledge that my thesis is subject to the rights and obligations stipulated by the Act No.121/2000 Coll., the Copyright Act, as amended, in particular that the Czech Technical University in Prague has the right to conclude a license agreement on the utilization of this thesis as a school work under the provisions of Article 60 (1) of the Act.

In Prague on March 5, 2023

.....

Czech Technical University in Prague

Faculty of Information Technology

© 2023 Ladislav Louka. All rights reserved.

This thesis is school work as defined by Copyright Act of the Czech Republic. It has been submitted at Czech Technical University in Prague, Faculty of Information Technology. The thesis is protected by the Copyright Act and its usage without author's permission is prohibited (with exceptions defined by the Copyright Act).

Citation of this thesis

Louka, Ladislav. *Object-relational mapping for database access in JavaScript*. Master's thesis. Czech Technical University in Prague, Faculty of Information Technology, 2023.

Abstrakt

V několika větách shrňte obsah a přínos této práce v českém jazyce.

Klíčová slova Replace with comma-separated list of keywords in Czech.

Abstract

Summarize the contents and contribution of your work in a few sentences in English language.

Keywords Replace with comma-separated list of keywords in English.

Contents

Introduction	1
Conclusion	5
Bibliography	7
A Acronyms	9
B Contents of enclosed medium	11

List of Figures

Introduction

Data lineage is an integral aspect of data governance as it offers a thorough understanding of the journey of data from its origin to its final destination. This knowledge is crucial for organizations to make informed decisions regarding their data and guarantee its accuracy, reliability, and security.

By tracking the lineage of data, organizations can determine the source of their data and the different modifications it experiences as it flows through various systems and applications. This information is vital for auditing and compliance with regulations as it provides a clear understanding of how data is utilized and where it is stored. Moreover, data lineage helps organizations detect any potential errors or inconsistencies in their data, allowing them to resolve these issues before they escalate.

In addition, data lineage offers valuable insights into the relationships between different data structures and systems, which helps make data-driven decisions. For instance, if an organization intends to modify its data architecture, it can evaluate the impact of these changes on other systems and processes by examining the data lineage.

Data lineage is a critical part of data governance as it provides organizations with a comprehensive understanding of the journey of their data. This knowledge is necessary for making informed decisions about data, guaranteeing its accuracy and security, and complying with regulations.

One of the tools that can analyze the data lineage is Manta. The Manta is a project that focuses on the management and analysis of data lineage, offering organizations a comprehensive view of the flow of their data across different

systems and applications. The Manta project provides metadata about the data flow and its representations, making it a valuable tool for organizations to understand the lifecycle of their data and make informed decisions about it.

This thesis aims to study the Databricks dialect of SQL language and its syntax and semantics. The implementation part of this thesis will focus on extending the Manta tool to retrieve information about the dataflow in Databricks SQL scripts through static analysis of the Databricks SQL scripts.

The thesis will be organized into five chapters, each focusing on a different aspect of our research. The first chapter will provide a theoretical background, focusing on parsing, static analysis of formal languages, automata grammar theory, and related topics. This chapter will provide a foundation for the rest of the thesis and help us understand the concepts and techniques used in our analysis.

The second chapter will analyze the Databricks SQL dialect, focusing on its syntax, semantics, and features. This chapter will provide a general overview of the language, its capabilities, and the use of Databricks technology.

The third chapter will cover the ANTLR grammar of the Databricks SQL dialect. In this chapter, we will delve deeper into the specifics of the language and explore how it is structured and parsed. We will also describe the challenges in creating comprehensive grammar for the language.

The fourth chapter will focus on implementing the module for resolving and analyzing Databricks SQL, which will retrieve data flow information from the scripts and store this information in the Manta. This chapter will describe our approach to analyzing the Databricks SQL scripts. We will also explain the implementation of our tool and its key components.

The fifth and final chapter will conclude the thesis with a discussion of the testing of our prototype. In this chapter, we will describe our approach to testing, evaluate the results, and discuss any required limitations or future work.

In conclusion, this thesis seeks to enhance the understanding of the dialect of SQL language used in the Databricks ecosystem, its syntax, and semantics, as well as extend the Manta tool with a new module for Databricks SQL analysis. The Manta project is a tool that helps organizations analyze and

manage their data lineage, providing them with valuable insights into their data journey. By studying the Databricks dialect of SQL, we aim to extend the capabilities of the Manta tool and enable it to retrieve information about the data flow in Databricks SQL scripts through static analysis.

Conclusion

Bibliography

- [1] Hopcroft, J.; Ullman, J. D. *Introduction to automata theory, languages, and computation*. CNIB, 1995.
- [2] Aho, A. V.; Sethi, R.; et al. *Compilers: principles, techniques, and tools*, volume 2. Addison-wesley Reading, 2007, ISBN 0-321-48681-1.
- [3] Melichar, B.; Češka, M.; et al. *Konstrukce překladačů*. Vydavatelství ČVUT, 1999, ISBN 80-01-02028-2.
- [4] Delta live tables SQL language reference. Available from: <https://docs.databricks.com/workflows/delta-live-tables/delta-live-tables-sql-ref.html>
- [5] What is Delta Lake? Feb 2023. Available from: <https://docs.databricks.com/delta/index.html>
- [6] Introduction to data lakes. May 2022. Available from: <https://www.databricks.com/discover/data-lakes/introduction>

Acronyms

PDA Push-down automaton

DFA Deterministic finite automata

Contents of enclosed medium

	readme.txt	the file with CD contents description
	exe	the directory with executables
	src	the directory of source codes
	wbdcm	implementation sources
	thesis	the directory of \LaTeX source codes of the thesis
	text	the thesis text directory
	thesis.pdf	the thesis text in PDF format
	thesis.ps	the thesis text in PS format