



Supervisor's statement of a final thesis

Supervisor: Ing. Jan Trávníček, Ph.D.
Student: Bc. Ondřej Hlaváč
Thesis title: Data flow analysis of scripts in SAP Hana SQL dialect
Branch / specialization: Software Engineering
Created on: 30 May 2022

Evaluation criteria

1. Fulfillment of the assignment

- ▶ [1] assignment fulfilled
- [2] assignment fulfilled with minor objections
- [3] assignment fulfilled with major objections
- [4] assignment not fulfilled

The task was to analyze SAP Hana's scripting language (extension of SQL, but mostly SQL compliant), to design and to implement a tool that is able to perform a static analysis of scripts in the language capable of generating dataflow among tables, views, and other structures in SAP Hana database. The implementation fulfils the practical part of the assignment fully; the text complements the implementation in theoretical and analytical aspects of the assignment well.

2. Main written part

72/100 (C)

The text is written in the English language on a good level, however, it would still benefit from additional proofreading. There are missing articles and some sentences are strangely worded.

Factual issues:

Despite this thesis being a software engineering one, the task is heavily based on the theory of formal languages, including grammars and automata. However, some topics there seem at least partially disconnected from each other and from the rest of the thesis; For instance, it is not mentioned how pushdown automaton and finite automaton is used in the implementation of a parser and lexer, respectively. Moreover, some notions are not defined precisely enough for the thesis text to be self-contained enough.

The text should mention that in the context of Manta, whitespace tokens are retained (albeit hidden) for the further possible reconstruction of the input.

"It may be helpful to show what transformations were used on the data in data lineage." - I presume, this is here to tease some further information, however, there is no link and thus it does not make sense at the place the sentence is written; most importantly the

information is never expanded.

In the description of case expressions, only the version, commonly known as "matched", is presented, and the "searched" one is completely missing.

Function calls are described rather strangely. "Functions expressions are references to functions. The process of referencing a function is called calling."

Select into clause description is confusing as it mixes "select into variables" (where the limitation to a single row makes sense) and "select into table" (where no such limitation is needed).

It is strange to see generated dataflow results in the analysis chapter.

In the design chapter, in the description of the data flow generator, the statement on the common data flow generator only raises a question about how much of a code is actually created as part of the proof of concept implementation regarding the data flow generator. The text specifies what are the priorities of arithmetic operators but doesn't do so for logical ones.

With some examples of the lexer rules, one would expect some parser rule examples to be shown too, however even though parsing is a lot more complex than lexing, both are described with roughly the same amount of text.

Typography issues:

Text on two occasions overflows the right edge of its bounding box.

Typesetting of math formulae is sometimes incorrect -- a command is printed as is, instead of its execution, missing whitespaces around some constructs, and symbols are sometimes typeset inside a math environment and sometimes not.

There are few forward references.

The images representing dataflow examples could have been inserted as vectors (they were most likely generated by Graphviz).

3. Non-written part, attachments

88 /100 (B)

The attached proof of concept implementation is in Java programming language.

It features the artefacts created as part of the implementation and it requires other common artefacts to execute.

The code satisfied the criteria to be mainlined into the codebase at Manta and as mentioned it already has been released as part of the regular release of the tool.

Each change underwent a mandatory code review. There naturally are some issues like unfinished dataflow generation for not often used constructs and a few imprecisions or small amounts of known bugs, however, all main statements are covered and the dataflow generated for them seems correct.

Notable issue:

- The DELETE statement does not process correctly a where clause if it contains a subselect.

Some more issues occurred on the customer side, however, these were caused by undocumented syntax.

4. Evaluation of results, publication outputs and awards

95 /100 (A)

Most importantly, the proof of concept implementation may serve as a minimum viable product implementation, and it actually already does so. That said, the code is now to be gradually improved with either support for more language constructs or bug-fixing.

5. Activity of the student

- [1] excellent activity
- ▶ [2] **very good activity**
- [3] average activity
- [4] weaker, but still sufficient activity
- [5] insufficient activity

Throughout the time of working on the thesis, the student continuously developed the implementation.

6. Self-reliance of the student

- [1] excellent self-reliance
- ▶ [2] **very good self-reliance**
- [3] average self-reliance
- [4] weaker, but still sufficient self-reliance
- [5] insufficient self-reliance

Throughout the time of working on the thesis, the student required consultations only occasionally.

The overall evaluation

85 /100 (B)

All in all, the quality of the implementation is very good and the text is good. Without repeating what is presented above, I can only state that I recommend the thesis for defence and I recommend evaluating it with 85 points, i.e. grade B.

Instructions

Fulfillment of the assignment

Assess whether the submitted FT defines the objectives sufficiently and in line with the assignment; whether the objectives are formulated correctly and fulfilled sufficiently. In the comment, specify the points of the assignment that have not been met, assess the severity, impact, and, if appropriate, also the cause of the deficiencies. If the assignment differs substantially from the standards for the FT or if the student has developed the FT beyond the assignment, describe the way it got reflected on the quality of the assignment's fulfilment and the way it affected your final evaluation.

Main written part

Evaluate whether the extent of the FT is adequate to its content and scope: are all the parts of the FT contentful and necessary? Next, consider whether the submitted FT is actually correct – are there factual errors or inaccuracies?

Evaluate the logical structure of the FT, the thematic flow between chapters and whether the text is comprehensible to the reader. Assess whether the formal notations in the FT are used correctly. Assess the typographic and language aspects of the FT, follow the Dean's Directive No. 52/2021, Art. 3.

Evaluate whether the relevant sources are properly used, quoted and cited. Verify that all quotes are properly distinguished from the results achieved in the FT, thus, that the citation ethics has not been violated and that the citations are complete and in accordance with citation practices and standards. Finally, evaluate whether the software and other copyrighted works have been used in accordance with their license terms.

Non-written part, attachments

Depending on the nature of the FT, comment on the non-written part of the thesis. For example: SW work – the overall quality of the program. Is the technology used (from the development to deployment) suitable and adequate? HW – functional sample. Evaluate the technology and tools used. Research and experimental work – repeatability of the experiment.

Evaluation of results, publication outputs and awards

Depending on the nature of the thesis, estimate whether the thesis results could be deployed in practice; alternatively, evaluate whether the results of the FT extend the already published/known results or whether they bring in completely new findings.

Activity of the student

From your experience with the course of the work on the thesis and its outcome, review the student's activity while working on the thesis, his/her punctuality when meeting the deadlines and whether he/she consulted you as he/she went along and also, whether he/she was well prepared for these consultations.

Self-reliance of the student

From your experience with the course of the work on the thesis and its outcome, assess the student's ability to develop independent creative work.

The overall evaluation

Summarize which of the aspects of the FT affected your grading process the most. The overall grade does not need to be an arithmetic mean (or other value) calculated from the evaluation in the previous criteria. Generally, a well-fulfilled assignment is assessed by grade A.