FACULTY OF INFORMATION TECHNOLOGY CTU IN PRAGUE

Supervisor's statement of a final thesis

Student:	Bc. Kyrylo Bulat
Supervisor:	Ing. Jan Trávníček, Ph.D.
Thesis title:	Dataflow analysis of Google BigQuery scripts
Branch of the study:	Web and Software Engineering

Date: 20. 1. 2021

Evaluation criterion:	*The evaluation scale: 1 to 4.*
1. Fulfilment of the assignment	***1 = assignment fulfilled,*** *2 = assignment fulfilled with minor objections,* *3 = assignment fulfilled with major objections,* *4 = assignment not fulfilled*

Criteria description:
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.

Comments:
The thesis's objective was to study the SQL language dialect called Google BigQuery; to analyze it so that it can be said whether and how to extract dataflow from scripts in this SQL dialect. Another objective was to implement a tool to perform the dataflow analysis and to extract metadata needed for the dataflow analysis. All objectives of the thesis were fulfilled.

Evaluation criterion:	*The evaluation scale: 0 to 100 points (grade A to F).*
2. Main written part	*92 (A)*

Criteria description:
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. 26/2017, 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.

Comments:
I appreciate that the thesis is written in very good English with very little to no language issues. The text is easy to read; the chapters contain relevant information.

Even though the thesis is defended at the Department of Software Engineering and therefore more focused on the software engineering aspects, I would appreciate a bit deeper introduction to parsing, a definition of a finite automaton and of a context-free grammar, etc.

In the Metadata extraction section, the SDK approach references the REST API approach that is described later. The final decision on which approach to use is BigQuery API, but it is not clear which of the presented options it is. I can only assume it is the REST API.

Few typographic issues:
- Bullet lists sometimes start at the very beginning of a page.
- Section 2.3.2 strangely starts with a table.
- It would make sense to break down the section 2.3.3.2 into subsections for individual literals.
- Some paragraphs would benefit from splitting into more.
- I found a few instances of text exceeding the right edge of the text bounding box.
- Figures 5.1 and 6.1 are inserted as bitmaps even though vectors should be relatively easily obtainable.

Evaluation criterion:	*The evaluation scale: 0 to 100 points (grade A to F).*
3. Non-written part, attachments	*98 (A)*

Criteria description:
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.

Comments:

In general, the proof of concept tool's implementation follows standard maintained in the Manta code base. The code is well documented with Javadoc and well structured.

I found just a few places where the code could be simplified:
- BigQueryApiCaller::makeHttpGetListRequest
- BigQueryExpressions.g:426, BigQueryMain.g:344,347,786 - not needed rewrite rules

Evaluation criterion:	*The evaluation scale: 0 to 100 points (grade A to F).*
4. Evaluation of results, publication outputs and awards	*100 (A)*

Criteria description:
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.

Comments:

The proof of concept implementation resulting from the thesis is already part of the Manta's codebase. It is still to be gradually improved; however, the analyzed subset of the language is big enough to allow for the tool's inclusion in the official release of the Manta software soon.

Evaluation criterion:	*The evaluation scale: 1 to 5.*
5. Activity and self-reliance of the student	*5a:* ***1 = excellent activity,*** *2 = very good activity,* *3 = average activity,* *4 = weaker, but still sufficient activity,* *5 = insufficient activity* *5b:* ***1 = excellent self-reliance,*** *2 = very good self-reliance,* *3 = average self-reliance,* *4 = weaker, but still sufficient self-reliance,* *5 = insufficient self-reliance.*

Criteria description:
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 (5a). Assess the student's ability to develop independent creative work (5b).

Comments:

The student very quickly and well adopted the principles maintained at Manta, and he was consulting his progress regularly.

Evaluation criterion:	*The evaluation scale: 0 to 100 points (grade A to F).*
6. The overall evaluation	*98 (A)*

Criteria description:
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.

Comments:

The text of the thesis is, in my opinion, excellent and I very much appreciate that it is written in English. The implementation is already a part of the codebase at Manta. Based on those facts, I recommend to accept the thesis for defence and I recommend to grade it with 98 points, i.e. grade A (excellent).

Signature of the supervisor: