COOPERATION

MONDO

Scalable Modelling and Model Management on the Cloud

(Temporary Logo)

Project Number 611125

TR – Train Benchmark

Version 0.1
29 January 2014
Draft

Public Distribution

BME

Project Partners: **ARMINES**, **Autonomous University of Madrid**, **BME**, **IKERLAN**, **Soft-Maint**, **SOFTEAM**, **The Open Group**, **UNINOVA**, **University of York**

Project Partner Contact Information

ARMINES Massimo Tisi Rue Alfred Kastler 4 44070 Nantes Cedex, France Tel: +33 2 51 85 82 09 E-mail: massimo.tisi@mines-nantes.fr	**Autonomous University of Madrid** Juan de Lara Calle Einstein 3 28049 Madrid, Spain Tel: +34 91 497 22 77 E-mail: juan.delara@uam.es
BME Daniel Varro Magyar Tudosok korutja 2 1117 Budapest, Hungary Tel: +36 146 33598 E-mail: varro@mit.bme.hu	**IKERLAN** Salvador Trujillo Paseo J.M. Arizmendiarrieta 2 20500 Mondragon, Spain Tel: +34 943 712 400 E-mail: strujillo@ikerlan.es
Soft-Maint Vincent Hanniet Rue du Chateau de L'Eraudiere 4 44300 Nantes, France Tel: +33 149 931 345 E-mail: vhanniet@sodifrance.fr	**SOFTEAM** Alessandra Bagnato Avenue Victor Hugo 21 75016 Paris, France Tel: +33 1 30 12 16 60 E-mail: alessandra.bagnato@softeam.fr
The Open Group Scott Hansen Avenue du Parc de Woluwe 56 1160 Brussels, Belgium Tel: +32 2 675 1136 E-mail: s.hansen@opengroup.org	**UNINOVA** Pedro Maló Campus da FCT/UNL, Monte de Caparica 2829-516 Caparica, Portugal Tel: +351 212 947883 E-mail: pmm@uninova.pt
University of York Dimitris Kolovos Deramore Lane York YO10 5GH, United Kingdom Tel: +44 1904 32516 E-mail: dimitris.kolovos@york.ac.uk	

Contents

Document Control

Version	Status	Date
0.1	Document outline	27 January 2014

Executive Summary

TODO

1 Overview

Scalability issues in model-driven engineering arise due to the increasing complexity of modeling workloads. This complexity comes from two main factors: (i) *instance model sizes* are exhibiting a tremendous growth as the complexity of systems-under-design is increasing, (ii) increasing *feature sophistication* in toolchains, such as complex model validation or transformations.

One of the the most computationally expensive tasks in modeling applications are *model queries*. While there are a number of existing benchmarks for queries over relational databases [6] or graph stores [3, 5], modeling tool workloads are significantly different. Specifically, modeling tools use much more complex queries than typical transactional systems, and the real world performance is affected by response time (i.e. execution time for a specific operation such as validation or transformation) than throughput (i.e. the amount of parallel transactions).

Overview To address this challenge, the Train Benchmark [7, 1] is a macro benchmark that aims to measure batch and incremental query evaluation performance, in a scenario that is specifically modeled after *model validation* in (domain-specific) modeling tools: at first, the entire model is validated, then after each model manipulation (e.g., the deletion of a reference) is followed by an immediate re-validation. The benchmark records execution times for four phases:

1. During the *read* phase, the instance model is loaded from hard drive to memory. This includes the parsing of the input as well as initializing data structures of the tool.
2. In the *check* phase, the instance model is queried to identify invalid elements. This can be as simple as reading the results from cache, or the model can be traversed based on some index. By the end of this phase, erroneous objects need to made available in a list.
3. In the *edit* phase, the model is modified to simulate effects of manual user edits. Here the size of the change set can be adjusted to correspond to small manual edits as well as large model transformations.
4. The re-validation of the model is carried out in the *re-check* phase similarly to the *check* phase.

The Train Benchmark computes two derived results based on the recorded data: (1) *batch validation time* (the sum of the *read* and check phases) represents the time that the user must wait to start to use the tool; (2) *incremental validation time* consists of the *edit* and *re-check* phases performed 100 times, representing the time that the user spent waiting for the tool validation.

Instance models The Train Benchmark uses a domain-specific model of a railway system that originates from the MOGENTES EU FP7 project, where both the metamodel and the well-formedness rules were defined by railway domain experts. This domain enables the definition of both simple and more complex model queries while it is uncomplicated enough to incorporate solutions from other technological spaces (e.g. ontologies, relational databases and RDF). This allows the comparison of the performance aspects of wider range of query tools from a constraint validation viewpoint.

The instance models are systematically generated based on the metamodel and the defined complex model queries: small instance model fragments are generated based on the queries, and then they are placed, randomized and connected to each other. The methodology takes care of controlling the number of matches of all defined model queries. To break symmetry, the exact number of elements and cardinalities are randomized.

This brings artificially generated models *closer to real world instances*, and *prevents query tools from efficiently storing* or caching of instance models. During the generation of the railway system model, errors are injected at random positions. These errors can be found in the check phase of the benchmark, which are reported, and can be corrected during the edit phase. The initial number of constraint violating elements are low (<1% of total elements).

Queries and transformations Queries are defined informally in plain text (in a tool independent way) and also formalized using the standard OCL language as a reference implementation (available on the benchmark website [1]). The queries range from simple attribute value checks to complex navigation operations consisting of several (4+) joins.

The functionally equivalent variants of these queries are formalized using the query language of different tools applying tool based optimizations. As a result, all query implementations must return (the same set of) invalid instance model elements.

In the *edit* phase, the model is modified to change the result set to be returned by the query in the re-check phase. For simulating manual modifications, the benchmark always performs hundred random edits (fixed low constant) which increases the number of erroneous elements. An edit operation only modify single model elements at once - more complex model manipulation is modelled as a series of edits.

Evaluation of measurements The Train Benchmark defines a Java-based framework and application programming interface that enables the integration of additional metamodels, instance models, query implementations and even new benchmark scenarios (that may be different from the original 4-phase concept). The default implementation contains a benchmark suite for queries implemented in Java, Eclipse OCL and EMF-IncQuery.

Measurements are recorded automatically in a machine-processable format (CSV) that is automatically processed by R [2] scripts. An extended version of the Train Benchmark [4] features several (instance model, query-specific and combined) *metrics* that can be used to characterize the "difficulty" of benchmark cases numerically, and – since they can be evaluated automatically for other domain/model/query combinations – allow to compare the benchmark cases with other real-world workloads.

References

[1] The train benchmark website. https://incquery.net/publications/trainbenchmark/full-results, 2013.

[2] The R project for statistical computing. http://www.r-project.org/, 2014.

[3] Christian Bizer and Andreas Schultz. The Berlin SPARQL benchmark. *International Journal on Semantic Web & Information Systems*, 5(2):1–24, 2009.

[4] Benedek Izsó, Zoltán Szatmári, Gábor Bergmann, Ákos Horváth, and István Ráth. Towards precise metrics for predicting graph query performance. In *2013 IEEE/ACM 28th International Conference on Automated Software Engineering (ASE)*, pages 412–431, Silicon Valley, CA, USA, 11/2013 2013. IEEE, IEEE. Acceptance Rate: 23%.

[5] Michael Schmidt, Thomas Hornung, Georg Lausen, and Christoph Pinkel. SP2Bench: A SPARQL performance benchmark. In *Proc. of the 25th International Conference on Data Engineering*, pages 222–233, Shanghai, China, 2009. IEEE.

[6] Transaction Processing Performance Council (TPC). TPC-C Benchmark, 2010.

[7] Zoltán Ujhelyi, Gábor Bergmann, Ábel Hegedüs, Ákos Horváth, Benedek Izsó, István Ráth, Zoltán Szatmári, and Dániel Varró. EMF-IncQuery: An integrated development environment for live model queries. *Science of Computer Programming*, 2014. Accepted.