

Kivonat

A lekérdezések központi szerepet játszanak az adatvezérelt alkalmazásokban. A model-vezérelt szoftvertervezés (model-driven engineering, MDE) eszközei és transzformációi erősen támaszkodnak a modellekérdezések hatékony kiértékelésére. A szoftvermodellek mérete és komplexitása intenzíven nő, ezért a jelenlegi MDE eszközökkel gyakran komoly skálázhatósági problémák merülnek fel, amelyek csökkentik a fejlesztés hatékonyságát és növelik annak költségeit.

A skálázhatósági kérdések központi témája az adatbázis-kezelés területén végzett kutatásoknak. A NoSQL rendszerek részben megoldást kínálnak erre a problémára, de cserébe le kell mondanunk az SQL rendszerek által biztosított deklaratív ad-hoc lekérdezések erejéről. A NoSQL rendszerek model-vezérelt alkalmazásokban történő közvetlen alkalmazása jelenleg is nyitott kutatási kérdés, az ezekben futtatott lekérdezések ugyanis jelentősen bonyolultabbak, mint az általános adatbázis-kezelő alkalmazásokban használtak.

Dolgozatom célja, hogy az EMF-INCQUERY-ben alkalmazott inkrementális gráfmintaillesztő technikákat elosztott, felhőalapú infrastruktúrára implementáljam. Bemutatok egy olyan újszerű architektúrát, amely elosztott, skálázható módon alkalmas lekérdezések inkrementális kiértékelésére. Az architektúra prototípusa, az INCQUERY-D rendszer képes egyetlen számítógéptől egy többgépes fürtig skálázódni, így képes nagy modelleken komplex lekérdezések hatékony futtatására. Az INCQUERY-D további előnye, hogy a lekérdezőmotor független a mögöttes adatbázis adatmodelljétől.

Az elképzelés működőképességét mérési eredményekkel igazoltam egy RDF- és egy gráfalapú adatbázis rendszerrel. Az eredmények bizonyítják, hogy az inkrementális lekérdezési technikák képesek hatékonyan működni elosztott környezetben is.

Abstract

Queries are the foundations of data intensive applications. In model-driven software engineering (MDE), model queries are core technologies of tools and transformations. As software models are rapidly increasing in size and complexity, traditional MDE tools frequently exhibit scalability issues that decrease productivity and increase costs.

While such scalability challenges are a constantly hot topic in the database community and recent efforts of the NoSQL movement have partially addressed many shortcomings, this happened at the cost of sacrificing the powerful declarative ad-hoc query capabilities of SQL. Unfortunately, this is a critical problem for MDE applications, as their queries can be significantly more complex than in general database applications. The applicability of NoSQL databases in MDE applications is subject for future research.

In this report, I aim to address this challenge by adapting incremental graph search techniques, known from the EMF-INCQUERY framework, to a distributed cloud infrastructure. I present a novel architecture for distributed, scalable incremental query evaluation. INCQUERY-D, the prototype system can scale up from a single node to a cluster of nodes that can handle very large models and complex queries efficiently. INCQUERY-D is a backend-agnostic system, meaning that its query engine is independent from the data model of the underlying database.

The feasibility of the approach is supported by early experimental results with both an RDF and a graph database backend. The results prove that incremental query evaluation techniques can work efficiently in a distributed environment as well.