

# Kivonat

A lekérdezések alapvető szerepet játszanak az adatvezérelt alkalmazásokban. A lekérdezések hatékony kiértékelésének módja a végrehajtás optimalizációja. A modellvezérelt szoftvertervezés (model-driven engineering, MDE) folyamatai nagymértékben támaszkodnak a modellek hatékony lekérdezésére. Azonban a modellek több 10-100 milliós méretűre növekedésével a modell lekérdezések és transzformációk komplexitása gyakran jelentős skálázhatósági problémát okoz a jelenlegi MDE eszközöknek, amelyek már kevésbé tudják támogatni a fejlesztés hatékonyságát.

A skálázhatósági problémára megoldást nyújtanak az elosztott, felhőalapú modell lekérdező rendszerek, melyek hatékony működéséhez elengedhetetlen a lekérdezések optimalizációja. Az elosztott környezetben történő lekérdezés optimalizáció a klasszikus egygépes adatbázis-kezelő és modell lekérdező rendszerekhez képest új aspektusokat és kihívásokat rejt az egyes erőforrások korlátai, az adatok hálózaton történő átvitele és az elosztott rendszer költségei miatt.

Dolgozatom célja, hogy bemutassak és az IncQuery-D elosztott modell és gráf lekérdező rendszerhez implementáljak olyan optimalizációs módszereket, amelyek speciálisak az elosztott inkrementális gráf lekérdezésekre. Ezen módszerek a modellek mérete és a lekérdezések jellege alapján heurisztikák segítségével allokálják a rendszer számítási csomópontjait a rendelkezésre álló erőforrásokra úgy, hogy az többféle szempont szerint optimalizált végrehajtást eredményezzen.

Az optimalizáció rendszer teljesítményére gyakorolt hatását mérési eredményekkel kívánom igazolni, melyek áttekintésében segítséget nyújt az általam készített monitorozó rendszer. További eredményem, hogy az IncQuery-D rendszer használatához fejlesztői környezetbe integrált eszközöket nyújtok.

# Abstract

Queries are the foundations of data-driven applications and query optimization is essential for their efficient evaluation. Fast model queries are of primary importance in the workflows of model-driven software engineering (MDE). As big software models have already reached 10-100 million nodes in size, traditional MDE tools often encounter scalability issues because of the complexity of queries and transformations. These issues limit the efficiency of the development process.

Distributed, cloud-based model query engines address to solve the scalability issues. However these systems require query optimization in order to evaluate queries fast and efficiently. Query optimization in distributed environments brings new aspects and challenges - including the capacity limit of resources, network communication between computers and the cost of the infrastructure - compared to the traditional single workstation model and database query engines.

In this report, I aim to address the challenges of these optimization problems and implement such methods, that are special for distributed incremental graph queries, for the IncQuery-D distributed model and graph query engine.