



Budapesti Műszaki és Gazdaságtudományi Egyetem  
Méréstechnika és Információs Rendszerek Tanszék

## Szuperskálázható modellezés

**Szárnyas Gábor (U944EQ), II. évf., MSc mérnök inf. szakos hallgató**  
**Konzulens: Izsó Benedek, Ráth István, Varró Dániel, MIT**  
**Szolgáltatásbiztonságra tervezés szakirány**  
**Diplomatervezés 1. összefoglaló**  
**2012/13. II. félév**

A diplomatervezés során végzett munkám egy hosszabb projekt része, amelynek célja egy elosztott modelltároló rendszer kialakítása.

A modellvezérelt tervezés egyre szélesebb körű alkalmazásának következménye, hogy a modellek komplexitása egyre nő, mind a kézzel készített, mind a generált modellek esetén. A növekedésnek jelenleg komoly határt szab, hogy a modelltároló rendszerek tipikusan csak egyetlen számítógépen képesek működni, ezért csak korlátozott méretű modellek tárolására alkalmasak. A problémára egy olyan horizontálisan skálázható rendszer jelentené a megoldást, amely újabb számítógépekkel bővítve egyre nagyobb méretű modellek tárolására lenne képes. Az elosztott tárolás mellett megfogalmazódik a modell párhuzamos feldolgozásának és biztonságos tárolásának igénye is.

A rendszer alapját egy adatbázis-kezelő rendszer képi, amely képes az elosztott működésre. Az elmúlt években a jól bevált relációs adatbázis-kezelők mellett tucatjával jelentek meg ún. NoSQL elosztott adatbázis rendszerek. Ezek specializált architektúrájukkal a Web 2.0-s alkalmazások igényeinek egy-egy aspektusát próbálják minél hatékonyabban lefedni, a relációs adatbázisokhoz képest gyakran egyszerűbb adatmodellt, gyengébb konzisztenciát, de magasabb rendelkezésre állást, jobb skálázhatóságot nyújtanak. A NoSQL adatbázis rendszerek gyakran egy-egy használati esetre jellemző terhelésre optimalizáltak. A modellalapú tervezőrendszerek tipikus felhasználásai: a modell eseményvezérelt szerkesztése, frissítése, modelltranszformációk és jólfomáltsági validációk futtatása.

A munkám előzménye az Izsó Benedek, Ráth István és Szatmári Zoltán által készített „TrainBenchmark” teszt sorozat, amely különböző – pl. EMF- és RDF-alapú – tárolórendszerekbe betöltött modelleken futtat jólfomáltsági kényszereket kiértékelő lekérdezéseket [1]. Az előző félévek során TrainBenchmarkhoz készült generátort kiegészítettem egy tulajdonsággráfokat támogató modulal, amely képes GraphML formátumú példánymodellek generálására.

A félév során folytattam a TrainBenchmark mérés sorozaton végzett munkát. Az EMF-IncQuery Rete Visualizer eszközével megvizsgáltam a RouteSensor lekérdezés Rete hálóját. A Rete háló alapján egy olyan deszkamodellt készítettem, amely a Rete hálót egy több számítógépből álló fürtön futtatja [4].

A deszkamodellben Neo4j gráfadatbázis kezelőt és Akka aszinkron kommunikációs keretrendszert használtam. A méréseink alapján az inkrementális kiértékelés legalább 1–2 nagyságrenddel gyorsabb a tisztán Neo4j-t használó nem-inkrementális lekérdezéshez képest.

Mivel a Neo4j nem támogat valódi elosztott működést, valóban skálázható rendszer megvalósításához más tároló technológiára lesz szükségünk. A félév során megismerkedtem a Titan fürtözött gráfadatbázis rendszerrel, amely képes valódi elosztott működésre az Apache Cassandra és az Apache HBase adatbázis kezelők használatával. A Titan mintájára elkészítettem a saját Cassandra-alapú implementációm, amely képes elosztott fájlrendszeren tárolt GraphML fájlok párhuzamos betöltésére is.

A témát TDK dolgozat és a Diplomatervezés 2. tárgy keretében fogom folytatni további mérésekkel, és egy elosztott modelltároló rendszer fejlesztésével.

[1] <https://trac.inf.mit.bme.hu/Ontology/wiki/TrainBenchmark/>

[2] <http://viatra.inf.mit.bme.hu/incquery/>

[3] <https://trac.inf.mit.bme.hu/Ontology/wiki/DistributedGraphProcessing>

[4] <https://svn.inf.mit.bme.hu/projects/ontology/source/incquery-d/>