



Budapesti Műszaki és Gazdaságtudományi Egyetem
Méréstechnika és Információs Rendszerek Tanszék

Train Benchmark kiegészítése az OreintDB adatbázis-kezelővel

Selmezi Zoltán (M2IZQE)

Konzulens: Szárnyas Gábor doktorandusz

MIT, Informatikai technológiák szakirány, Rendszertervezés ágazat

Önálló laboratórium (BSc) összefoglaló

2014/15. II. félév

Az önálló laboratóriumi fejlesztés során a Train Benchmark tanszéki fejlesztésű keretrendszer kiegészítésén dolgoztam. A félév elején csak egy gráfadatbázis-kezelő implementálása volt megtalálható a keretrendszerben, így a motivációm az volt, hogy egy másik gráfadatbázis-kezelőt is hozzáadjak, és a végén összehasonlítsam a teljesítményüket. A kiválasztott eszköz az OrientDB lett, mivel ugyanannak a keretrendszernek (TinkerPop) az elemeit használja.

A félév első felében megismerkedtem a Train Benchmark keretrendszerrel, az OreintDB-vel, valamint a TinkerPop lekérdező nyelvével a Gremlinnel. Fontos feladat volt, hogy a Train Benchmark lekérdezéseit elkészítsem Gremlin nyelven, mivel csak így tudtam letesztelni az OrientDB teljesítményét. A félév első felében el is készültem ezekkel a lekérdezésekkel, és le is teszteltem helyes működésüket.

A félév második felében jöhetett az OrientDB implementálása a keretrendszerbe Java nyelven. Először a gráfgenerálót kellett kiegészíteni, hogy az OreintDB-nek megfelelő gráfot generáljon. Szerencsére a graphml fájlban csak egy sort kellett hozzáadni, így elég hamar megoldódott ez a probléma.

Következhett a benchmark megvalósítása, ami egy új Maven projektet jelentett. A pom.xml fájlban meg kellett adni az OrientDb legfrissebb verziójának függőségeit. A projektben ugyanolyan fájlokat kellett létrehozni, mint a többi eszköz esetén, és az egyszerűbbek tartalma is hasonló lett, csak az eszköz nevét kellett OrientDb-re átírni. Az összetetteknél viszont használni kellett az OrientDB Java API-jának függvényeit is.

Fontos volt a tranzakciók pontos megvalósítása, ugyanis a Train Benchmark legfőbb tulajdonsága, hogy a tranzakciók végrehajtásának a teljesítményét is méri. Ezen implementációk lekérdezésenként eltértek.

Az önálló laboratórium utolsó szakaszában a feladatom a tesztek elvégzése volt, valamint az eredmények összevetése. A legfőbb eredmény az volt, hogy a Neo4j (a már létező megvalósítás a Train Benchmarkban) Cypher lekérdezéseket használva minden esetben gyorsabb volt, mint az OrientDB Gremlin lekérdezések használva.