Budapesti Műszaki és Gazdaságtudományi Egyetem
Méréstechnika és Információs Rendszerek Tanszék

Moduláris benchmark keretrendszer fejlesztése

Kővári Zsolt
Konzulens: Szárnyas Gábor doktorandusz
MIT, Informatikai technológiák szakirány – Rendszertervezés ágazat
Önálló laboratórium
2014/15. I. félév

Napjainkban a gyorsan növekedő adatmennyiségek következtében, elengedhetetlen feltétele lett a különböző adatbázis kezelőknek, lévén szó NOSQL, vagy relációs sémára épülő reprezentációkról, hogy a skálázhatósági, illetve a valós idejű lekérdezési kritériumoknak eleget tegyen. Ezen esszenciális feltételek, egyrészt a megnövekedett modell méret, másrészt a komplexebb lekérdezés esetében várnak el stabil és valós idejű működést az adatbázis kezelőktől.

Ezen, fent említett kritériumok betartásának megbizonyosodása érdekében született meg a Hibatűrő Rendszerek Kutatócsoport által fejlesztett, Java programozási nyelvben írt Train Benchmark keretrendszer, amely különböző adatbázis reprezentációk mérésére szolgál, és többek között olyan jellemzőket vizsgál, vet össze egymással, mint a lekérdezés kiértékeléséhez, vagy a modell beolvasásához szükséges idő, illetve a használt memória mérete.

Feladatom volt a félév során, hogy a Train Benchmark keretrendszerrel megismerkedjek, mind használati, mind fejlesztési szinten. A rugalmasabb működés elérése érdekében, további komponensekkel egészítettem ki a rendszert, amik egyrészt könnyebb vezérlést, olvashatóbb és automatizáltan feldolgozhatóbb kimeneti eredményeket jelentettek, másrészt egy új mérhető eszköz integrálását is, a Jena TDB-t.

A Train Benchmark vezérlésének és konfigurálásának újragondolása, majd ennek implementálása volt az elsődleges cél a munkám során, melyhez a Python programozási nyelvet használtam. A félév végére elértem azt, hogy szemben a korábban elérhető megoldáshoz képest, a saját megvalósításom lehetővé teszi tetszőleges és rugalmas mérési konfigurációk megadását, továbbá a benchmark folyamatának egyszerűbb és automatizáltabb vezérlését. Ezenfelül az általam beépített, konfigurációs paraméterekre futtatott validáció, még a benchmark futása előtt szűri a hibás értékeket, meggátolva ezzel a mérés helytelen működését, esetleg abortálását.

Új, Java-ban írt osztályok implementálása szükségeltetett ahhoz, hogy a mérési eredményeket szebb és átláthatóbb, utólag könnyebben feldolgozhatóbb alakba öntsem, mégpedig JSON formátumba. A benchmark minden egyes futása esetén egy, az imént említett fájl generálódik az adott mérés minden releváns eredményével. Szintén Java nyelvben programoztam, hogy az RDF alapú adatbázis kezelőkhöz elérhető Jena TDB nevű eszközt mérhetővé, tehát a keretrendszer részévé tegyem.

Végezetül, a mérési eredményeket feldolgozó komponens megírására került a sor. Az R programozási nyelvet használva, az adott mért eszközök futási időinek összevetésére, és ezek ábrázolására törekedtem. A félév végére elértem azt, hogy mindegyik mért eszköz által produkált modell beolvasási, és lekérdezés kiértékelési idői összevethetők egymással, közös diagramon, a modell méretének függvényében.