

**Önálló laboratórium beszámoló**

**BME-TMIT**

**Készítette:** Sümeghy Tamás Pál

Neptun-kód: GFHSRE

Szak: műszaki informatikus

Szakirány: Internet és infokommunikációs alkalmazásai

E-mail cím: [schumy@sch.bme.hu](mailto:schumy@sch.bme.hu)

**Konzulens(ek):** Kardkovács Zsolt

Sárecz Lajos

Email címe(k): [kardkovacs@tmit.bme.hu](mailto:kardkovacs@tmit.bme.hu)

[lajos.sarecz@oracle.com](mailto:lajos.sarecz@oracle.com)

**Tanév:** 2007/2008. 1. félév

**Téma címe:** Üzleti intelligencia és térképi megjelenítési lehetőségei

**Feladat:**

A féléves feladat az Oracle két technológiájának, a Spatial-nek és a Business Intelligence-nek a megismerése, megértése, majd pedig gyakorlati kipróbálása. Ennek során először a megfelelő eszközöket kell telepíteni, működésüket megismerni, kipróbálni. Ezután egy mintaadatbázist kell betöltenünk, majd azon keresztül megismerni a technológiákban rejlő lehetőségeket. A mintaadatbázis lehet előre elkészített, letöltött fájl, de lehet általunk kitalált és összeállított is.

## 1. A laboratóriumi munka környezetének ismertetése, a munka előzményei és kiindulási állapota

### Bevezető/elméleti összefoglaló

A féléves munka során elsősorban arra kellett koncentrálnunk, hogy egy megfelelő elméleti alapot szerezzünk az általunk választott témakörben, kiegészítve ezt némi gyakorlati ismerkedéssel is. Erre azért van nagy szükségünk, hogy később jobban el tudjunk mélyedni a témában, továbbá ki tudjunk választani egy irányt, mellyel valóban részletesebben szeretnénk foglalkozni, amennyiben a témakört túlzottan komplexnek találánk. Én ezért főleg az elméleti háttérét próbáltam megérteni a kiválasztott feladatomban, és némileg háttérbe szorítottam emiatt a gyakorlati alkalmazási problémákat. Mindazonáltal nem szeretném ezen a szinten abbahagyni a téma vizsgálatát, így később természetesen a gyakorlati feladatokkal is részletesebben szeretnénk foglalkozni.

### Oracle

Elsőként néhány szót magáról az Oracle-ről, mint az egész feladatnak keretet adó szoftvercsomagról. Napjainkban egyre nagyobb érdeklődés mutatkozik az adatbázis-kezelők iránt. Ez nagyrészt annak köszönhető, hogy digitalizálódó világunkban ma már minden eszköz ontja magából az adatokat, és ezt a rengeteg információt nem szeretnénk veszni hagyni. Értelemszerűen tárolnunk kell, és ez a tárolás sem lehet akármilyen, mivel egy szint után áttekinthetetlené válhat az adathalmaz. Erre a problémakörre jelentenek megoldást az adatbázis-kezelők, melyek közül hosszú idő óta kiemelkedő szerepet tölt be az Oracle. Ezt a vezető pozíciót többek között a nagyfokú integrálhatóságnak és a folyamatos teljesítménybeli fejlesztésnek köszönheti. A legújabb verziója idén nyáron jelent meg, ez a 11g. Ezzel együtt az alkotóelemek is megújultak, felfrissültek, többek között a későbbiekben sokat említett Spatial kiegészítés is.<sup>[3]</sup>

### Locator / Spatial

A Locator az Oracle 8i újdonsága (Standard és Enterprise Edition) volt. Olyan helyfüggő és helymeghatározó alap funkció, amely a legtöbb felhasználói alkalmazáshoz szükséges. Ez az alap funkció az, hogy különböző térinformációkat (pl. szélességi és hosszúsági koordináták) tudunk az adatainkkal együtt, egy helyen tárolni. Ezen kívül, plusz szolgáltatásként a kiegészítő adatok analízisére is lehetőséget ad.

A Spatial szintén az Oracle 8i-ben jelent meg (Enterprise Edition), nagyban hasonlít a Locator-re, az alapjaik azonosak. Ugyanazt az alapvető objektumtípust használják (SDO\_GEOMETRY), azonos metaadatok és indexelési sémák jellemzőek rájuk. Tulajdonképpen a Locator tartalmazza a Spatial alapfunkciókat, de a fejlesztések mind a Spatial-ben kerültek implementálásra.<sup>[2]</sup>

A Spatial lehetséges definíciói:

- egy séma, ami geometriai adattípusok tárolását, szintaxisát és szemantikáját írja le
- egy térbeli indexelési mechanizmus
- operátorok, funkciók és eljárások gyűjteménye, amelyek segítségével térben összekapcsolt adatbázisokat kérdezhetünk le vagy térinformációt hordozó adatokon végezhetünk analízist
- topologikus adatmodell, hogy az adatra úgy tekinthessünk, mint csúcspontra, élre vagy felületre egy topológiában
- hálózati adatmodell, melynek segítségével tulajdonságokat vagy objektumokat modellezhetünk, mint csomópontok és a köztük lévő kapcsolatok egy hálózatban
- GeoRaster adatformátum, ami egy raszterkép és egy rácsháló, valamint a kettőt összekapcsoló metaadatok összessége<sup>[1]</sup>

A Spatial adatmodellje egy hierarchikus struktúra, amely alapelemekből, geometriai alakzatokból és rétegekből épül fel. A rétegek geometriai alakzatokból állnak, míg ezeket alapelemek alkotják. Az elemek az alapvető építőkövek, ezek a pont, a szakasz és a töröttvonal, valamint a poligon. A geometriai alakzatok az alapelemek egy rendezett halmazaként definiálhatóak. Állhatnak mindössze egyetlen alapelemből, de alapelemek homogén vagy heterogén halmazából is. Egy réteg olyan geometriai alakzatok gyűjteménye, melyeknek az attribútumhalmaza azonos. Például egy térinformatikai rendszerben egy réteg tartalmazza a topográfiai információkat, egy másik a népsűrűség adatait, míg egy harmadik az utak és hidak hálózatát.<sup>[6]</sup>

A geometriai adatokat egy általános adatszerkezetben, az SDO\_GEOMETRY-ben tárolhatjuk:

```
CREATE TYPE sdo_geometry AS OBJECT (
  SDO_GTYPE NUMBER, - alakzat dimenziója, típusa
  SDO_SRID NUMBER, - koordinátarendszer
  SDO_POINT SDO_POINT_TYPE, - pont adatszerkezet
  SDO_ELEM_INFO SDO_ELEM_INFO_ARRAY, - adatszerkezet, mely
megadja, hogy hogyan kell értelmezni a pontok koordinátáit
  SDO_ORDINATES SDO_ORDINATE_ARRAY); - pontok koordinátái
```

A geometriai metaadatok a dimenziók jellemzésére használatosak, minden egyes dimenzióhoz tárolhatjuk a magasabb és alacsonyabb szintek felé fennálló kötöttségeket illetve az adott dimenzió túrását is. Minden Spatial felhasználó két nézetet láthat ebből a táblából, ezek: *USER\_SDO\_GEOM\_METADATA* és *ALL\_SDO\_GEOM\_METADATA*. Az előbbi tartalmazza azokat a metaadatokat, melyek a felhasználó saját tábláihoz kapcsolódnak, míg a második nézet minden táblához tárolja a metaadatokat, ez a nézet lekérdezésekhez használható.<sup>[1]</sup>

A Spatial kétszintű lekérdező modellt alkalmaz a térbeli lekérdezések és összekapcsolások feloldásához. A két szintet elsődleges és másodlagos szűrőnek nevezzük.

Az elsődleges szűrő gyorsan és egyszerűen kiválasztja azokat a potenciális rekordokat, amiket a másodlagos szűrő felé továbbításra érdemesnek tart. Közelítést alkalmaz a geometriai alakzatokra, ezzel is gyorsítva a futást és csökkentve a költségeket. Az elsődleges szűrő használata azt eredményezi, hogy egy lekérdezés teljesítménye nem függ az adathalmaz méretétől, mivel egy adott lekérdezés esetén az elsődleges szűrő kimenete mindig ugyanakkora méretű származtatott adathalmaz. A másodlagos szűrő már pontos számítások alapján hozza létre az eredményhalmazt. Ezek számításigényes feladatok, de az elsődleges szűrő miatt már egy relatíve szűk halmazzal kell csak dolgoznunk, nem a teljes adatmennyiséggel.

Az elsődleges szűrő megvalósításához szükséges az adatok térbeli indexelése. Mivel az elsődleges szűrőnek igen gyorsan és hatékonyan kell az adatok egy részhalmazát kiválasztania, ez az igény meghatározza az indexelés karakterisztikáját.

A Spatial egy igen erőteljes és gyors indexelést használ, amit R-fának hívnak. Mint minden indexnek, egy térbeli indexnek is egy olyan mechanizmust kell adni, ami gyorsítja és korlátossá teszi a keresést. Ebben az esetben ez a mechanizmus térbeli kritériumokon alapszik.<sup>[1]</sup>

## Business Intelligence

Az üzleti intelligencia egy folyamat melynek során információt szerzünk vállaltunk működéséről, állapotáról, eredményességéről, stb. a rendelkezésünkre álló adatok alapján. Egy mai vállalat rengeteg információt gyűjt és tárol a különböző tranzakciós rendszerein keresztül. Ezek az adatok kiválóan alkalmasak az üzleti intelligencia analíziseihez.<sup>[5]</sup>

A vállalatok számára szükséges széleskörű rálátás biztosításához sok, eddig nem használt képességet használ a BI. Az információk minden felhasználó számára fontosak, így igen sok formában kell tudnunk az összegyűjtött adatokat továbbadni. Lényeges szempont, hogy az információt a kellő időben tudjuk prezentálni, ne teljen el feleslegesen sok idő, mire az összegyűjtött adatokból használható információ válik. Fontos elem, hogy a már meglévő üzleti folyamatokba, workflow-kba észrevétlenül integrálódjanak az alkalmazás elemei, ne kelljen ezeket lényegesen módosítani. A leglényegesebb azonban, hogy skálázható maradjon a rendszerünk a nagyon gyorsan növekvő adatmennyiség és a folyamatosan bővülő felhasználói igények mellett is.

Mindezek kielégítésére a BI alrendszer szinte minden rétegében bővíti az Oracle adatbázis-kezelő eredeti szolgáltatáscsomagját. A bővítmények és új fejlesztések 6 nagy kategóriába sorolhatók be:

- Prezentációs eszközök: olyan új elemek, mint például Interactive Dashboard, Answers, Reporting and Publishing, Delivery, Disconnected Analytics
- BI szerver: teljesítmény, skálázhatóság és rendelkezésre állás javítása
- Telepítési és adminisztrációs költségek: mind az időszükséglet, mind az anyagi ráfordítás mértéke csökkent
- Integrálhatóság: alacsonyabb költség és egyszerűbb megoldások
- Adatforrások integrálása: gyorsabb kapcsolódás a vállalati adatforrásokhoz
- Biztonság, azonosítás menedzsment, audit<sup>[4]</sup>

### **A munka állapota, készültségi foka a félév elején**

A félév elején gyakorlatilag semmi kézzelfogható dolog nem lehetett a birtokunkban, köszönhetően annak, hogy a vizsgált technológia világviszonylatban is igen új volt. Félév elején például maga a 11g rendszer még csak Linux operációs rendszer alatt létezett, a félév során jelent meg egyéb platformokra is. Ettől függetlenül a technológiák korábbi verziójú kiadásai, az ezekkel kapcsolatos dokumentációk, ismeretek, tapasztalatok már léteztek, így ezek alapján könnyedén neki lehetett állni a kutatásnak.

A gyakorlati alkalmazások terén lehetőségünk lett volna egy mintaadatbázis használatára, mely az Oracle honlapjáról letölthető, és a Navteq cég térképi adatbázisát tartalmazza. Én azonban részben az Oracle és a kapcsolódó rendszerek közötti beállítási problémáim, részben a már említett elméletközpontú megközelítem miatt egy saját, lényegesen egyszerűbb adatbázist hoztam létre, és ezen próbáltam ki megszerzett tudásom egy részét.

## **2. Az elvégzett munka és eredmények ismertetése**

### **Az általam konkrétan elvégzett munka bemutatása**

A félév során elsődlegesen tehát elméleti anyagok megismerésével foglalkoztam. Korábban már foglalkoztam az egyetemi oktatás keretében adatbázis-kezeléssel, de mindenképpen szükségét láttam ismereteim felfrissítésének, így első lépésben a korábban tanultakat ismételttem át. Ezzel egy időben az első feladatot is el kellett végezni, össze kellett állítanom a munkatervemet.

Ezután kezdődött az ismerkedés az elméleti háttérrel, elsősorban a Spatial-lel. Nagy segítségemre volt a <http://otn.oracle.com> oldal, ahol mind a különböző cikkekben, mind a fórumokban sok hasznos és érdekes információt tudtam meg a régebbi és a vadonatúj kiadásról is. Korábbi verziókhoz létezett magyar nyelvre fordított dokumentum is, az újhoz azonban ezek még nem álltak rendelkezésre. Ezek a dokumentációk főleg az újdonságokat, frissítéseket emelték ki, volt köztük azonban néhány általános témákkal foglalkozó is.

A mélyebb megértés érdekében a 10g-s Spatial-hez készült felhasználói útmutatót tanulmányoztam át, itt olyan témákba kaphattam részletes betekintést, mint az

adatszerkezetek, metaadatok, indexelési és lekérdezési mechanizmusok, koordinátarendszer-kezelés, geokódolás.

A Business Intelligence témakörre némileg kevesebb energiát fordítottam, úgy éreztem, a jövőben inkább a térinformatikai alkalmazások felé szeretnék orientálódni, ezért arra helyeztem a nagyobb hangsúlyt. Ennek ellenére itt is több dokumentációt olvastam át, megpróbálva ezekből kiszűrni a lényegeset, illetve azokat a részeket, melyekre később nekem is szükségem lehet.

A gyakorlati alkalmazáshoz elengedhetetlenül szükséges a rendszerek megfelelő telepítése. A munkám elején nagyon sok gondom adódott ezzel kapcsolatban, Mivel sokáig csak Linux alatt volt elérhető a 11g, megpróbálkoztam ennek telepítésével. Ritkán használok Linuxot, így sokadik alkalomra sem sikerült a telepítemem. Ezután jelent meg a Windows-os változat, melynek telepítése már sokkal gördülékenyebben ment, így elkezdhettem az érdemi munkát is.

Az ismerkedés és az alapfunkciók kipróbálása nem okozott problémát. Azonban a letölthető mintaadatbázist nem tudtam elkezdni használni, valószínűleg beállítási hibáim miatt. Ezért egy saját, de lényegesen egyszerűbb adatbázist hoztam létre, melyhez egy rövid leírással a valósághoz kapcsolódó tartalmat is tudtam adni.

Az adatbázis létrehozását, feltöltését, majd a lekérdezéseket egy szkript fájlban gyűjtöttem össze, a könnyebb tesztelhetőség kedvéért. Szintén elmentettem a kimenetet is, így a későbbiek során könnyen ellenőrizhető az adatbázis.

## Összefoglalás

A félév során szerzett új tudásom, tapasztalatok, a gyakorlati alkalmazás eredményei:

- általános ismeretek az Oracle térinformációs modelljéről, gyakorlatáról
- a Spatial fogalma, létrehozásának indítókai
- specifikus ismeretek a Spatial adatkezelési elvéről, adatformátumokról, metaadatokról
- a térinformációs adatok feldolgozási lehetőségei, lekérdezési és indexelési algoritmusok
- néhány érdekesebb témakör megismerése, mint például a koordinátarendszerek
- geokódolás elméleti háttere
- Business Intelligence fogalma, az ehhez használható elméleti háttér, megvalósításbeli példák
- Oracle webes kezelőfelületének, telepítési és menedzselési filozófiájának megismerése
- kapcsolódó szoftverek (oc4j, MapViewer) érintőleges megismerése
- gyakorlati alkalmazásként egy közel 300 soros sql szkript fájl megírása

### 3. Irodalom, és csatlakozó dokumentumok jegyzéke

#### A tanulmányozott irodalom jegyzéke:

- [1] Spatial User's Guide
- [2] <http://www.oracle.com/technology/pub/articles/lokitz-spatial-geoserver.html>
- [3] <http://otn.oracle.com> – fórumok, cikkek
- [4] New Release in Oracle Business Intelligence Suite Enterprise Edition – Oracle White Paper
- [5] Oracle Business Intelligence Technical Overview – Oracle White Paper
- [6] ORACLE SPATIAL 11g - Advanced Spatial Data Management for the Enterprise – Oracle Data Sheet
- [7] ORACLE LOCATOR ÉS ORACLE SPATIAL OPCÍÓ - Térinformatika az Oracle Database 10g Release 2-ben – Oracle Adatlap
- [8] ORACLE LOCATOR - Location Features in Oracle Database 11g – Oracle Feature Overview

#### Csatlakozó egyéb elkészült dokumentációk / fájlok / stb. jegyzéke:

A félév során az általunk létrehozott blog-ban adtuk közre munkánk eredményét. Ez a blog teljesen nyilvános, bárki által elérhető, továbbá fellelhető itt az összes általam készített dokumentáció és egyéb fájl is. A blog címe: <http://sumeghy-onlab.blog.hu>.

#### Elkészített dokumentációk:

- munkaterv.pdf
- onlab\_071009.pdf
- onlab\_071009mod.pdf
- onlab\_071105.pdf
- onlab\_071129.pdf
- onlab\_071204.pdf
- onlab\_ossz.pdf

#### A gyakorlati alkalmazás során készített fájlok:

- halozat\_sql.txt
- halozat\_ki.txt