

E-Traffic közlekedési használati előrejelző modell megalkotása

Az adatbázis logikai tervének kialakítása

Megrendelője: Hírközlési és Informatikai Tudományos Egyesület



Készítette: Comtrans Kereskedelmi és Szolgáltató Kft.



Tartalomjegyzék

Bevezetés 4

Dokumentum célja 4

Adatbázis logikai tagolása 5

Staging, beolvasás 5

SOR (System of records) a rendszer adatai 5

Indexek 5

**Foreign key** 5

Summary, riportok, kimutatások 6

Kimutatások főbb szempontjai 6

Indikátorok vizsgálata 6

Kontrolling 6

Minőségbiztosítás 6

Analysis, elemzői támogatás 7

Szempontok 8

**Indexek** 8

**Foreign key** 8

Adatbázis táblák 9

Úthálózat táblái 9

Felhasználói táblák 10

Környezeti táblák 10

Modell táblái 10

Betöltés (staging) táblái 10

Elemzői táblák 11

Adatbázis szerverrel szemben támasztott kritériumok 11

Általános irányelvek 11

Jogosultságok 11

Betöltési folyamat felhasználója 11

Riporting és megjelenítés felhasználója 12

Fejlesztői felhasználók 12

Elemzői felhasználók 12

Mentések 12

# Bevezetés

A dokumentum a Hírközlési és Informatikai Tudományos Egyesület
 mint megbízó valamint a Comtrans Kereskedelmi és Szolgáltató Kft. mint vállalkozó által létrejött 2013 Április 02. kötött szerződés alapján készült.

Jelen dokumentum a célja az „E-Traffic közlekedési használati előrejelző modell megalkotása” című projektben a térinformatikai adatbázis kialakítása, cél: az adatbázis logikai tervének kialakítása: a feldolgozandó adathalmazok minőségi jellemzőire támaszkodva és az elemzési célok szem előtt tartásával

|  |
| --- |
| * Meg kell határozni a kiszolgáló adatbázis optimális szerkezetét
 |
| * Az adattáblák és mezők típusait, értékkészletét, valamint az azonosító és kapcsoló kulcsait
 |
| * Az elemzési lekérdezésekhez illesztett indexeket
 |

# Dokumentum célja

A dokumentum célja, hogy megfogalmazza azokat az irányelveket, melyeket az adatbázis szerkezetének kialakításánál figyelembe kell venni. Megfogalmazza azokat a célokat, mellyel meghatározzák az adatbázis szerkezeti felépítését. Egy elemzői eszköz támogatását kiszolgáló táblaszerkezet adattartalmában, szerkezetében lényegesen eltér egy programot vagy modellt kiszolgáló tábla struktúrájától, adattartalmától. Ezért fontos, hogy ezek az igények meghatározásra kerüljenek, hogy a majdani tervezés és implementálás ezek igények mentén haladjanak.

# Adatbázis logikai tagolása

Az adatbázis 4 fő részre tagolódik, ezek elhelyezkedése lehet fizikailag külön schema-ban, adatbázisban, lényeg, hogy a betöltéskor, működéskor a rendszer számára mindhárom egy időben elérhető legyen.

A logikai tagolás funkcionálisan az alábbi elv szerint különíti el az adatbázisban tárolt adatokat.

## Staging, beolvasás

Adatok beolvasása: staging, jelölése a táblák elnevezésébe az stg\_ prefix. A külső rendszerek adatainak beolvasása ezekbe a táblákba történik, a beolvasás után a rendszer átalakítja, összegzi, majd a megfelelő rendszertáblákba tölti, vagy frissíti az ott lévő adatokat a beolvasás értékének megfelelően.

Ezen táblák szerkezetét tekintve az adatforrás állományoknak felelnek meg, hiszen funkcióikat tekintve csak a betöltendő adatok beolvasásának eredményeinek tárolására szolgálnak. A beolvasás oka, hogy az adatbázis-kezelő funkcionalitására épített rendszerkomponensek lényegesen gyorsabban tudják az adatok helyességét ellenőrizni vagy a megfelelő kapcsolásokhoz, normalizáláshoz szükséges adatokat előállítani.

## SOR (System of records) a rendszer adatai

A rendszer működésének biztosításához szükséges táblák és adatok. Ezen adatbázis táblák nevei a sor\_ előtaggal vannak ellátva. Ezen táblák adatai már a megszokott adatbázis reprezentációval kerülnek tárolásra, ahol a megfelelő foreign key-ek indexek minden lehetséges és szükséges helyen kialakításra kerülnek.

### Indexek

A nem egyértelmű (nem primary key vagy kereséséi feltételekben szereplő attribútumok esetén) indexek kialakítása a rendszer működésének monitorozásával a tényleges működés közben alakítható ki. Ez azt jelenti, hogy a rendszer erőforrásait nagy részben igénybevevő lekérdezések loggolásra kerülnek, majd optimalizáció szempontjából kivizsgálásra. Ha ehhez indexek kialakítására van szükség, akkor a megfelelő index típus létrehozása ezen információk alapján történik meg.

Minden referencia mezőre típusnak megfelelő index kerül, kivéve, ha ezt egy másik index feleslegessé teszi.

### **Foreign key**

Ezen táblák mindegyike normalizált, így vagy eleve rendelkezik az adat kapcsolótáblával, vagy ki kell alakítani azokat a kapcsolótáblákat, melyek mentén az adat normalizálása megtörtént.

## Summary, riportok, kimutatások

A projekt jelen fázisában a szolgáltatások kialakítása még nem kezdődött el, így ezen adattáblák még nem véglegesek, a projekt folyamán módosulhatnak, újak jöhetnek létre, illetve meglévők kerülhetnek ki a már meglévő táblák közül. Jelen fejezet célja, hogy rámutasson azon lehetőségekre, funkciókra, melyhez hasonlókra a leendő szolgáltatás támaszkodni tud. További célja, hogy összeszedje azon főbb szempontokat, melyek alapján a riporting szolgáltatás kialakítása elkezdődhet.

### Kimutatások főbb szempontjai

#### Indikátorok vizsgálata

Riportok készítése egy adott időszakról az alábbiak szerint:

* Utak, útszakaszok aktuális állapota
	+ országos riport,
	+ területi bontás (megye, kistérség).
* Utak útszakaszok fejlődése egy adott intervallumon belül (legalább 2 éves viszonylatban hasonlítja össze az úthálózatot)
* Utak, útszakaszok fejlődése paraméterek mentén:
	+ gazdaság, adó, vállalkozások száma, bevétele,
	+ népesség,
	+ infrastruktúra,
	+ oktatás,
	+ kultúra,
	+ egészségügyi infrastruktúra és ellátás,
	+ szolgáltatások.

#### Kontrolling

A rendszer működéséhez kapcsolódó riportok, minden betöltött adathalmazhoz:

* betöltés megléte,
* betöltés állapota: mely adathalmazok kerültek betöltésre, melyek hiányoznak,
* hibás sorok száma,
* elutasított sorok száma,
* javítás alatt lévő sorok száma,
* korrigált sorok száma,
* beazonosíthatatlan kapcsolók száma.

#### Minőségbiztosítás

Ezen riportok ellenőrzik azokat a minőségi indikátorokat, amelyek megléte esetén a szolgáltatás biztosítható.

Ezek az alábbiak lehetnek:

* minden indikátort sikerült a rendszernek betöltenie,
* az úthálózat frissítése megtörtént,
* forgalmi adatok az hány százalékát fedik le:
	+ automata méréssel,
	+ manuális forgalomszámlálással,
	+ mobil forgalomszámlálással,
	+ becsléssel.

## Analysis, elemzői támogatás

Elemzői táblák szerkezete eltér a megszokott normalizált adatbázis táblák szerkezetétől. Elemzési szempontból az a legfontosabb, hogy egy adott logikai egységhez tartozó információ egy adatbázis rekordban szerepeljen. Ezt a különbséget az alábbi példa jól szemlélteti:

Minden városra tároljuk iskolatípusonként a tanulók számát. Normalizálva az adat struktúrája az alábbi:

|  |  |
| --- | --- |
| id | Telepules\_nev |
| 1 | Budapest |
|  | ... |
| 6 | Eger |
|  | ... |

|  |  |
| --- | --- |
| id | Iskola\_tipus |
| 1 | Esti |
| 2 | Nappali |
| 3 | Levelező |

|  |  |  |
| --- | --- | --- |
| Telepules\_id | Iskola\_tipus\_id | Tanulok\_szama |
| 1 | 1 | 123 |
| 12 | 2 | 34 |
| 33 | 1 | 553 |

Elemzési szempontból viszont ez a struktúra lényegesen nehezebben kezelhető, mint a nem normált táblaszerkezet. Itt nem gond, ha tábla struktúrája változik, ha egy új iskolatípust kell felvenni (pl. OKJ képzés, felnőttoktatás), vagy más ehhez hasonló adatváltozás. Sokkal fontosabb, hogy a képletekben egy adott sor elemeit tudják használni, amely sokkal optimálisabb, mintha ezen értékek mindegyikét külön lekérdezéssel kellene a megkeresni. Ennek megfelelően az elemezések szempontjából a fenti adatok táblaszerkezete az alábbi:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Telepules\_id | Telepules\_nev | Esti\_t\_szama | Nappali\_t\_szama | Levelezo\_t\_szama |
| 1 | Budapest | 2312 | 435 | 543 |
| 3 | Baja | 345 | 678 | 987 |
| ... | ... | ... | ... | ... |

Az elemzés támogatását szolgáló táblák prefixe az analysis\_ előtag. További lényeges szempont, hogy ezen logikai csoportba tartozó adatok historikusan is tárolva legyenek, hogy az elemzések alatt ezek modellre gyakorolt hatását is elemezni lehessen.

### Szempontok

A fent felsorolt strukturális kritériumokon túl fontos, hogy az elemzői táblák egyértelműen legyenek szeparálva. A bontást inkább a képletekben együtt szereplő adatok adják, mintsem az adatok logikai struktúrája vagy adatköre. Abban az esetben amennyiben kétséges, vagy többféle számítás használja ugyanazon adatokat, akkor az adatok kerüljenek egy táblába, így az elemzés során az adatok helyét nem kell az elemzőknek keresniük.

### **Indexek**

Indexek ezeken a táblákon csak azokra a külső kulcsokra tehetők, melyek egy másik tábla elsődleges kulcsára referálnak vagy az sok számítás, keresés vagy szűrés indokolja.

### **Foreign key**

Minden elsődleges index, amely a táblában szerepel, külső kulcsként is meg kell jelenjen az adat konzisztencia biztosításának érdekében.

# Adatbázis táblák

## Úthálózat táblái

Tárolni kell:

* az úthálózat csomópontjait,
* a csomópontok közötti útszakasz egyszerűsített levágott szakaszait,
* az egyszerűsített útszakasz pontjait az ábrázoláshoz,
* az útszakaszok forgalmi információit.

Ezen információk az idők folyamán változnak, ezek a változások a modellezés szempontjából lényegesek, így fel kell készülni a változások kezelésére, tárolására.

## Felhasználói táblák

A felhasználóhoz tartozó információk az alábbiak:

* a felhasználói adatok, információk,
* a felhasználó által létrehozott utak, útszakaszok,
* a felhasználó által módosított környezeti és gazdasági változók,
* a felhasználói modellezés eredményei.

A felhasználók a jogot a szolgáltatás igénybevételére egy adott időszakra kapják meg, amelyet regisztrációval és valószínűleg előfizetés vásárlásával szereznek meg. Fontos, hogy az adatbázis alkalmas legyen ezen hozzáférési korlátok kezelésére, mind a beregisztrálására, meghosszabbítására.

## Környezeti táblák

A modellezés szempontjából fontos változók tárolása. Ennek kialakítása a projekt ezen vagy későbbi szakaszában történik, így ezen dokumentáció jelenleg nem képes meghatározni ezen adatok körét.

## Modell táblái

Fel kell készülni a különböző gazdasági, népességügyi, infrastrukturális és egyéb statisztikai adatok tárolására. Az adatbázis modellezéshez kapcsolódó adattáblái két részre tagolhatóak:

1. a számítások elvégzéséhez szükséges adatok,
2. a számítások eredményét tartalmazó táblák, melyekkel a rendszer a modellezést végzi.

Mindkét csoportba tartozó táblaszerkezetek a projekt ezen fázisaiban kerülnek kialakítására, majd a projekt későbbi szakaszaiban folyamatosan bővülnek és változnak az ismeretanyag bővülése által.

## Betöltés (staging) táblái

A rendszer működése szempontjából fontos adatforrások adatainak adatbázisba töltéséhez kialakított táblák, melyek szerkezetei megegyeznek a forrás-file adatainak szerkezetével, illetve értékkészletével. A rendszer ezen táblákon végzi el a szükséges transzformációkat, melyekből a modellezést és az úthálózat megjelenítését támogató adattáblákat tölti fel.

## Elemzői táblák

Minden betöltött adatforrás adatainak meg kell jelennie egy az elemzők által elérhető és az elemzéseket támogató adatbázistáblákban. Az elemzések a projekt szempontjából kiemelt fontosságúak, hisz a modellezés pontosságát, annak validálását ezen adatokra támaszkodva végzik.

# Adatbázis szerverrel szemben támasztott kritériumok

## Általános irányelvek

A projektre dedikált adatbázis szervernek 7/24-ben elérhetőnek kell lennie. A változások kezelésére ki kell dolgozni egy együttműködést, hogy a project résztvevőinek sem a munkája, sem az eredményessége ne legyen veszélyeztetve. Minden módosítást, nagyobb terheléssel járó műveletet, vagy bejelentés, egyeztetés után lehet futtatni, véghezvinni, vagy erre külön dedikált szerveren kell a módosításokat, terheléseket lefuttatni.

Az ilyen jellegű adatbázis klónozáshoz, rendelkezésre kell, álljon egy, az eredetivel megegyező, adatbázis másolat, illetve egy betöltési dokumentáció.

## Jogosultságok

A projekt jelen állásában a rendszer elérését az alábbiak szerint korlátozzuk:

* betöltési folyamat felhasználója
* riporting és megjelenítés felhasználója
* fejlesztői felhasználók
* elemzői felhasználók

### Betöltési folyamat felhasználója

Olvasási jog az összes adatbázis táblára, a felhasználó táblák kivételével írási jog az összes adatbázis táblára.

### Riporting és megjelenítés felhasználója

Olvasási jog az összes adattáblára, írási jog a felhasználói táblákra.

### Fejlesztői felhasználók

Teljes hozzáférés az adatbázishoz és az összes táblához.

### Elemzői felhasználók

Hozzáférés olvasásra az összes adattáblához.

## Mentések

Minden jelentősebb változás (elemzői modell felülírása, adatfrissítések) előtt az adatbázist menteni kell. Az adatbázis mentéséről, disaster recovery plan-ről gondoskodni kell.