

## Bevezetés

Az elmúlt egy, másfél évtizedben sok projekt tűzte ki céljául adattárházak építését, üzletiintelligencia-rendszerek készítését. A tapasztalat azt mutatja, hogy többféle megközelítés, értelmezés létezik az ilyen jellegű projektek megvalósításával kapcsolatban. Az alapfogalmak lassan bevonultak az informatikai köztudatba, de úgy érezzük, hogy továbbra is szükség van egy átfogó technológiai kézikönyvre, amely a teljes folyamatot ismerteti, részletesen tárgyalja a dimenzionális modellezés technikáját, bemutatja a különböző szoftvereszközöket.

Így a könyv írásakor a következő célokat tűztük ki:

- ◆ Szeretnénk a különböző adattárház-készítési megközelítéseket bemutatni és összehasonlítani;
- ◆ az adattárház és BI projektek során előforduló fejlesztési lépéseket, fogalmakat definiálni;
- ◆ a dimenzionális modellezés technikáját részletesen ismertetni, kiegészítve azt számos saját fejlesztési megfontolással;
- ◆ a teljes adattárház építési folyamat során használatos szoftvereszközök típusait bemutatni és a velük szemben támasztott elvárásokat definiálni, s végül, de nem utolsó sorban
- ◆ az Oracle eszközeit megismertetni az Olvasókkal az előző kategorizálás szerint.

A könyvet egyaránt használhatják kezdő adattárház-építők és már gyakorló szakemberek. Azt gondoljuk, hogy az adattárház-készítési folyamatban részt vevő bármely szakember számára szól a könyv olyan módon, hogy lefedi a teljes fejlesztési életciklust és technológiai folyamatba rendezi a lépéseket. Ugyanakkor nagyon jó lenne, ha üzleti felhasználók is forgatnák a könyvet, és tudnák, hogy milyen elvárásaik lehetnek egy jó adattárházzal szemben. Szeretnénk, ha megismernék azokat a fantasztikus analitikus elemzési lehetőségeket, amelyeket egy jól megtervezett dimenzionális modell szolgáltat egy nagy tudású végfelhasználói elemző szoftver egyszerű alkalmazásával. Számukra is szeretnénk hangsúlyozni azt a tényt, hogy milyen kihívásokkal kell megküzdeni egy projektnek, hogy jó minőségű adatokat biztosítson az adattárházban.

A könyv több részre tagolódik.

- ◆ Az I. rész foglalkozik a technológia kialakulásával, a különböző adattárház-definíciókkal, ismerteti a használt fogalmakat, a fejlesztési folyamatot.
- ◆ A II. rész fejezetei szólnak a dimenzionális modellek készítésének alapttechnikájáról, míg
- ◆ a III. rész az adattárházak egyik sarkalatos tervezési elemét, a dimenziókat részletezi.
- ◆ A IV. rész a tény táblák hatékony tervezését tűzte ki célul alapszintű tény adatok tárolásától az összegzett információk kezeléséig.
- ◆ Az V. rész az ETL folyamat legfontosabb lépéseinek, feladatainak leírását tartalmazza.

- ♦ A VI. részben a BI eszközök általános végfelhasználói és adminisztratív lehetőségeit ismertetjük, és a könyv központi esettanulmányán mutatunk be néhány elemzést, riportálást végző felületet.
- ♦ Az 1. melléklet ismerteti az Oracle-szoftvertermékeit, amelyeket a fejlesztés különböző céljaira használhatunk a projekt megfelelő fázisaiban.

Az első hat rész fizikai környezettől független technológiát ismertet, csak az 1. melléklet tartalmaz eszközspecifikus leírásokat. Ez alól kivétel némiképp a VI. rész, ahol az elemzési, riporting felületek bemutatása során Oracle eszközöket használtunk.

A könyv végén Név- és tárgymutató található. Fogalomtárat nem mellékelünk, mert a fogalmak definíciói a kötetben megtalálhatóak, a tárgymutató segítségével könnyen el lehet érni azokat. Mellékelünk viszont egy speciális Esettanulmány-összefoglalót és indexet. Segítségével oldottuk meg a központi esettanulmánnyal foglalkozó részekre való hivatkozást a konkrét üzleti téma és az alkalmazott technika megnevezésével. Célunk az volt, hogy az esettanulmány modelljének fejlődését könnyen követhessük a különböző fejezetekben.

A témában használatos fogalmakat nem teljesen következetesen használjuk magyarul. Számos kifejezésnek, például data warehouse, datamart, használatos a magyar megnevezése (adattárház, adatpiac), számos fogalom azonban az angol kifejezésével terjedt el. A staging area helyett például erőltettnek éreznénk a – fogalmat egyébként jól kifejező – felvonulási terület kifejezés használatát.

A fejezetek végén rövid összefoglalás található a különösen fontos vagy hangsúlyozni kívánt közlendők tömör megismétlésével.

A modellezési technikák és módszerek szemléltetéséhez több környezetből veszünk példákat. Erre azért van szükség, mert nehéz lenne olyan témát találni, amelyen minden tervezési szempontot könnyen tudnánk magyarázni. Ugyanakkor egy egyszerű, de komplett esettanulmány, amely egy képzeletbeli Pizzafutár cég üzleti elemzéseiről szól, végigvonul a fejezeteken.

A modellezés során **konzekvensen elkülönítjük az elemzési és tervezési szinteket**. A tényleges fizikai megvalósítás kérdéseivel általában nem foglalkozunk, hiszen azok függhetnek a megvalósítási eszközöktől. Többször kitérünk azonban hasonló környezetekben alkalmazott tipikus megoldásokra.

Az elemzés és tervezés során előálló dimenzionális modellek ábrái az Oracle Designer grafikus komponenseinek jelölésrendszere szerint készültek: a fogalmi modellek ábrázolása az Entity Relationship Diagrammer, míg a tervezési modellek a Design Editor stílusában. Azoknak az Olvasóknak, akik nem jártasak ezen technikák jelölésrendszerében, a könyv 2. Mellékletében leírtuk a szükséges tudnivalókat. A modelleken az attribútumok illetve oszlopok jellemzőit (kötelezőség, adattípus) igény szerint tüntettük fel: néha elégnak láttuk az oszlop nevének használatát, máskor további tulajdonságait is célszerűnek tartottuk ábrázolni. Az egyed-kapcsolat modelleken szereplő fogalmak megnevezésénél a magyar, míg a fizikai adatbázisstervek esetén természetesen az angol ábécé betűit használtuk.