

# Az adattárház-technológia előnyei és hátrányai – gyakorlati szempontok figyelembe vételével

---

*Positive and negative sides of data warehousing technology – with practical viewpoints*

## **Abstract**

Sooner or later, but every enterprise will face the idea of creating an enterprise-scale data warehouse. Just like before every IT / development / introductory project, in these cases it is important to examine, that who may be involved in producing such a system, and what positive and negative additional benefits may be brought by the introduction.

A decade ago, the existence of a data warehouse could only be afforded by the largest companies – now this fortunately changes. In the last few years many of the greater software houses have the initiative to produce a data warehousing software, which could serve the needs of small and mid-sized companies. This is worth monitoring, since the situation (breaking into the small and mid-sized company sector) was the same with business intelligence a few years ago – and as known, data warehouse technology serves as a good ground for business intelligence initiatives / business intelligence enrichment.

The article primarily focuses on collecting the positive and negative sides of the technology.

Keywords: data warehouse, data warehousing technology, decision support, enterprise information system, reporting, management information system, datamart, data procedures

## Absztrakt

Minden nagyobb vállalatnál előbb utóbb felmerül a vállalati adattárház létrehozásának ötlete. Mint minden informatikai / fejlesztési / bevezetési projekt előtt, ilyen esetben is fontos megvizsgálni, hogy kik lehetnek érintettek egy ilyen rendszer kialakításában, illetve milyen pozitív, illetve negatív hozadéka lehetnek a bevezetésnek.

Egy évtizeddel ezelőtt az adattárházak meglétét csak a legnagyobb cégek tudták megengedni maguknak – mára ez szerencsésen változóban van. Az utóbbi néhány év technológiai fejlesztései között több szoftvercégnél ott a kezdeményezés, mi szerint létre kell hozni egy kis- és közepes vállalatok által használható és elérhető adattárház technológiát megvalósító szoftvert. Erre érdemes lesz figyelni még évekig, mivel az üzleti intelligencia megoldások ugyanígy kezdtek betörni néhány éve a kisebb cégek szektorába – az adattárház technológia pedig nagyon jó táptalajként funkcionál üzleti intelligencia megoldásokhoz és azok tökéletesítéséhez.

A cikk elsősorban az adattárházak pozitív és negatív oldalainak összegyűjtésére fókuszál.

Kulcsszavak: adattárház, adattárház technológia, döntéstámogatás, vállalati információs rendszer, jelentéskészítés, vezetői információs rendszer, adatpiac, adat-műveletek

## Bevezetés

Az EDW (Enterprise Data Warehouse) magyar megnevezése vállalati adattárház. Ebben az információkötegetben megtalálható minden adat: történeti és kurrens egyaránt [1]. Bár a felhalmozott adatmennyiség gyakran túl sok lehet egyes elemzők számára, meglétének célja mégis az, hogy egy helyen lehessen elérni minden a céget érintő információt.

Azaz, az adattárház gyűjti össze a vállalat minden részlegének összes adatát [2]. A központi kezelésnek több előnye van: egyszerűbbé válik az adatkezelés (a centralizált hozzáférés-jogosultság kiosztás javít a biztonságon, a centralizált tárolás pedig olcsóbbá és könnyebbé teszi az archiválást és a biztonsági másolatok készítését és kezelését), és jelentősen egyszerűsödik az adatok beszerzésének költsége (idő takarít meg).

Az adattárház létrehozása hatalmas projekt. Össze kell szedni, hogy az egész vállalatnál milyen adatok vannak, milyen természetűek, hol és milyen formában keletkeznek, és később hol lehetnek felhasználva [1]. A tény pozitív oldala az, hogy számos módszertan [3] és ajánlás foglalkozik adattárházak építésével. Ettől függetlenül persze lépcsőzetesen célszerű végrehajtani – valamint itt is igaz az általános mérnöki igazság, mi szerint az alapos és átgondolt tervezés nagyban képes hozzájárulni a termék (itt adattárház) elkészültéhez és magas minőségéhez, jó használhatóságához [4][5][6].

Ha egy adattárház már üzemben van, nagyon jó lehetőséget ad a vállalatnak arra, hogy a meglévő adathalmazt elemezve, segítse a döntési folyamatot. Az elemzések során gyakran adatbányászati eszközök kerülnek használatra, melyek eredményei riportokban jelennek meg. Ezek a riportok aztán főként taktikai és stratégiai döntések meghozatalában segítenek. Felmerül a kérdés persze, hogy operatív döntések során lehet-e használni a technológiát, de megállapítható egyrészt, hogy rendkívüli erőforrásigénye miatt csak nehezen és költségesen valósítható meg, másrészt az ilyen elemzések számítás- és adatigényük miatt nagyon hosszúra nyúlnak – így egy operatív döntés esetén nem lehetséges a legtöbb esetben az eredmények kivárása. (Ez az elemzés mára olyan munkakörök megszületését tette lehetővé, mint a riportkészítő elemző vagy adatbányászati elemző.)

Nagy (és főként régóta üzemelő) adattárházak esetén a legnagyobb problémát a megfelelő teljesítmény nyújtása adja [5]. Nem ritka, hogy az adattárházban található adatbázisok összmérete több 10 terabájtos értéket ér el. Egy ekkora adatmennyiséget persze célszerű elosztani, hogy egy része háttértáron, egy (aktuálisabb rész) a memóriában tárolható

– ekkor a friss adatok kezelése és lekérdezése gyors lesz ugyan, de a régebbi adatok elérése lassúvá válik, és ez különösen akkor nyújtja nagyon hosszúra az adatok elemzését, ha az sok adatot érint [7].

Az olyan elemzések gyorsíthatóak fel, melyek kifejezetten a vállalat egy területéről származó adatokkal foglalkozik – ebben nyújt segítséget az adatpiacok alkalmazása. (Elengedhetetlen persze az adatpiacok és az adattárház közötti megfelelő sávszélességű hálózati kapcsolat.)

## **Tények és nézőpontok**

Minden nagyobb vállalatnál előbb utóbb felmerül a vállalati adattárház létrehozásának ötlete. Mint minden informatikai / fejlesztési / bevezetési projekt előtt, ilyen esetben is fontos megvizsgálni, hogy kik lehetnek érintettek egy ilyen rendszer kialakításában, illetve milyen pozitív, illetve negatív hozadéka lehetnek a bevezetésnek [12][18][21].

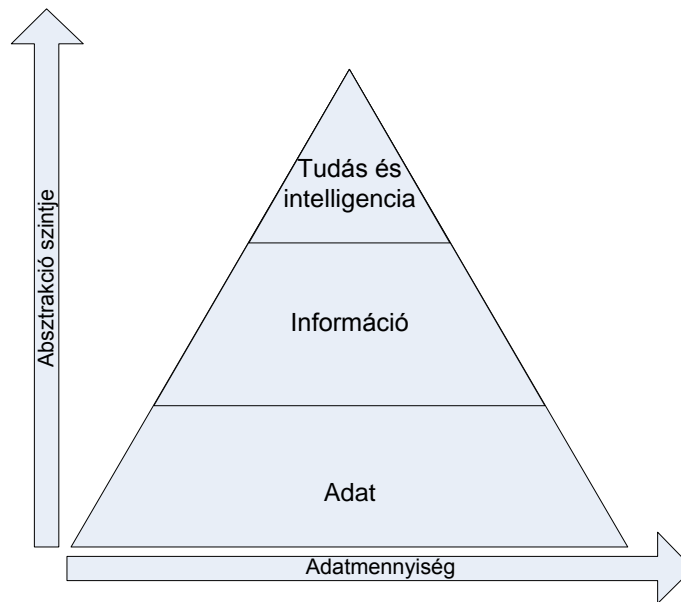
Egy évtizeddel ezelőtt az adattárházak meglétét csak a legnagyobb cégek tudták megengedni maguknak – mára ez szerencsésen változóban van. Az utóbbi néhány év technológiai fejlesztései között több szoftvercégnél ott a kezdeményezés, mi szerint létre kell hozni egy kis- és közepes vállalatok által használható és elérhető adattárház technológiát megvalósító szoftvert. Erre érdemes lesz figyelni még évekig, mivel az üzleti intelligencia megoldások ugyanígy kezdtek betörni néhány éve a kisebb cégek szektorába – az adattárház technológia pedig nagyon jó táptalajként funkcionál üzleti intelligencia megoldásokhoz és azok tökéletesítéséhez.

A cikkben elsősorban az adattárházak bevezetésének / létrehozásának előnyei és hátrányai fogalmazódnak meg.

## **A vitathatatlan előnyök**

Az üzleti világban szokták mondani: az idő pénz – mára kiegészítve azzal, hogy az információ ugyancsak pénz. Minden jó üzleti intelligencia rendszer számára előnyös, ha adatait egy adattárházból nyeri – így biztosítható, hogy jó minőségű adatokból jó minőségű elemzések és jelentések születnek, gyorsabban és megbízhatóbban. Éppen ezért, mindenkinek, aki a vállalaton belül a BI kezdeményezések hatósugarába kerül, fontos értenie, hogy igenis olyan információkat kell gyűjteni, ami később a vállalat működését jobbá teheti [8].

Az információk hatékony gyűjtése, feldolgozása és értékelése nem egyszerű feladat. Mint ahogyan az absztrakciós piramis is rámutat, a rendelkezésre álló adatok mennyiségétől függ a megszerezhető információk és megszerezhető tudás mennyisége, de minden adag így elérhetővé tett tudáshoz rengeteg adatra van szükség [18][21].



1. ábra: Absztrakciós piramis

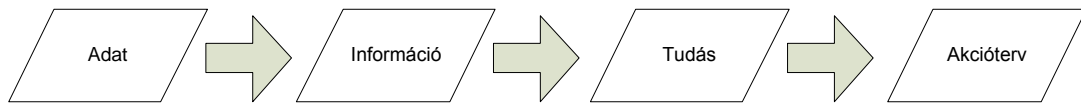
Az absztrakciós piramis [1. ábra] értelmezéséhez [9] tisztázni kell néhány alapvető fogalmat:

Az adat lehet bármilyen mérési eredmény, észlelés, ismérv, hír vagy közlemény [10].

Az információ az adatok értelmezéséből adódik. Megfogalmazható úgy is, hogy az adatok gyűjtésének és oly módú rendezésének eredménye, mely az adatok közti kapcsolatokra fényt derít, azokat tartalommal és értelmezéssel ellátva.

A tudás valamilyen tárgykör ismeretét jelenti, beleértve az információkat, tényeket és leírásokat. A tudás megszerzése persze történhet oktatás vagy tapasztalat által – jelen esetben a tapasztalat által megszerezhető tudásra kell koncentrálni. A tudás megléte annyit tesz, hogy a tárgykört elméleti és gyakorlati oldalról egyaránt értjük. Tudás továbbá az információk megértésén és azokban minták felismerésén alapuló elv, mely mély belelátást biztosít az információkba.

A folyamat végeredményeként előálló tudás [2. ábra] hasznosításának módja az, ha akcióttervet készít a vállalat, majd meg is valósítja azt.



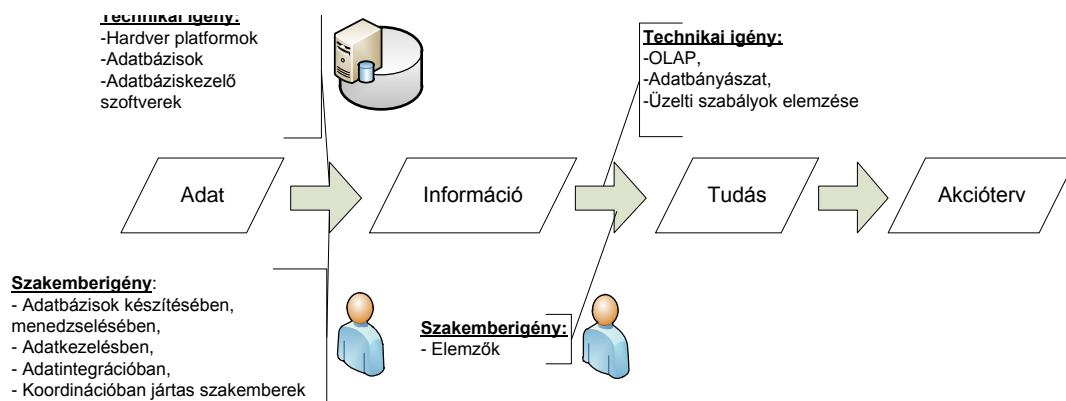
2. ábra: Az üzleti intelligencia folyamata

Az a folyamatrész, mely során az adatokból információkat képezünk, megfogalmazható úgy is, hogy meg kell határozni, melyik adatot, milyen összefüggésben kell gyűjteni, illetve kezelni. A környezetből kiemelve (vagy környezet nélkül) az adatok kevésbé értékesek (ilyen lehet egy vásárlói névhalmaz), ám megfelelő környezetben gyűjtve, azokat értelmezéssel ellátva sokkal nagyobb értéket nyernek (vásárlói adatok összegyűjtve: nevek, hozzájuk kapcsolódó vásárlások, egyéb értékek) [11].

Elmondható, hogy a BI-folyamatok közül főként az adatok kezeléséről és megjelenítéséről van szó, mely fizikai megjelenésüket tekintve főként hardverplatformokat, relációs adatbázisokat és azokat kezelő szoftvereket jelentenek – funkciójuk tekintetében lekérdezések futtatására és jelentések készítésre van szükség. Ehhez a fázishoz az emberigény elsősorban adatkezelésben, adatintegrációban és koordinációban jártas szakemberekben merül ki [13][14].

Ha a vállalat odáig eljutott, hogy rengeteg információval rendelkezik, akkor azokat többféle módon kell elemezni, hogy abból tudás nyerhető legyen. A nyert tudás felhasználható arra, hogy akciótervek készüljenek belőle, majd azokat megvalósítva megoldást nyújtsanak valamilyen problémára.

Itt elmondható, hogy a BI-folyamatok közül itt főként az elemzést támogató komponensekre van szükség [3. ábra], mint amilyenek az adattárházak, az analitikai feldolgozás (OLAP), az adatbányászat, az üzleti szabályok elemzése, az adatok körvonalazása és minősítése. Itt a szakemberigény elsősorban olyan szakemberekben nyilvánul meg, akik ezeket az informatikai eszközöket hatékonyan tudják kezelni, hogy kell a megfelelő eredményeket kinyerni – e nélkül semmit sem érnek a drága szoftverek [16][17].



3. ábra: Üzleti intelligencia igényei

Valószínűleg az utolsó fázis a legfontosabb lépés, itt keletkezik a valódi, kézzel fogható és alátámasztható tartalom (tudásból akcióterv), valamint ez az amire a vezetés is vágyik. Az egész üzleti intelligenciára alapozott vállalati stratégia kulcsfontja a reagálás képessége. Ezen keresztül lesz megállapítható az információtechnológiára fordított beruházás megtérülése is. Az akciók segítségével növelhető lesz az üzleti hatékonyság, az eladások, javítható a vásárlók megcélzása, csökkenthetőek a szolgáltatások költségei, felismerhetőek a csalások és összességében a költségek csökkentésével nőhet a profit.

## A megfogalmazható hátrányok

Az adattárházakkal kapcsolatban minden forrásból szinte kizárólag az előnyöket kaphatjuk meg. A technológia alkalmazásának hátrányai azonban bizonyos idő után szembetűnőek – ezek a megfogalmazható hátrányok alapvetően három csoportba sorolhatóak:

- Sok esetben nincs tényleges szükség egy adattárházra az ügymenet gyorsítása szempontjából,
- Az adattárházhoz való alkalmazkodás jelentősen meg tud bonyolítani üzleti folyamatokat, illetve
- A magas fejlesztési és fenntartási költségek, valamint a lassú megtérülésből származó hátrányok.

Az tény, hogy az adattárház technológia népszerű, és emiatt kelendő is. Ám nem minden cég számára elengedhetetlen egy adattárház léte. Olyan helyeken, ahol nem fontosak a történeti adatok, nem kell azokat elemezni (nem alkalmaznak semmiféle üzleti

intelligenciát), és nem fontos, hogy azok azonnal rendelkezésre álljanak, ott nem szükséges az adattárház kiépítése.

Ahol mégis kiépítésre kerül egy adattárház (mert igazolható, hogy szükséges, megtérül és segít az ügymenet gyorsabbá és hatékonyabbá tételében), ott is számolni kell azzal, hogy lesznek olyan üzleti folyamatok, melyek bonyolultabbá válnak azáltal, hogy a napi rutin részévé válik az adattárház. Itt elsősorban az adatok beszerzését és frissítését érintő folyamatokra kell gondolni – hozzáférés, engedélyezés. Ezen felül az sem mindegy, hogy az adattárház milyen módon tárolja az adatokat: a tervezési fázis legelején kell eldönteni, hogy dimenziós vagy normálformás technika kerül alkalmazásra. Míg az előbbi könnyebb adatelérést és könnyebb kezelhetőséget biztosít olyan emberek számára is, akik nem elsősorban informatikai végzettségűek – cserébe az adatok hozzáadása bonyolultabb lesz, addig az utóbbi esetében az adatok hozzáadása könnyebb, ám lekérdezések / riportok készítése lesz bonyolultabb – sok esetben informatikai tudást igényelve.

Az adattárházak fejlesztése (vagyis igények és követelmények felmérése, tervezés és prototípuskészítés az új verzióból, annak bevezetése és üzemeltetése / fenntartása) – függetlenül attól, hogy egy iteratív folyamat és ebből kifolyólag relatíve gyorsan jelentkezik használható eredmény – hosszú és bonyolult, sok embert igénybe vevő folyamat: Nem ritkák fél évnél hosszabb adattárház projektek, amelyekben (valamilyen mértékben) a cég munkatársainak több mint harmada részt kell vegyen (a hosszú és sokemberes fejlesztések pedig mindig drágák). Az ETL folyamatok tervezése során ráadásul a fejlesztőknek két irányba kell figyelniük: egyfelől meg kell határozni, hogy melyik adat, honnét (melyik forrásrendszerből), milyen formában (tisztítással és szűréssel) és hová kerüljön az adattárházba, milyen rendszerességgel; másfelől vizsgálni kell a kialakuló adatfüggéseket, azaz, ha valamin változtatni kell (az adaton, típusán vagy akár a betöltés folyamatán), akkor az milyen hatásokkal járhat. A fenntartási költségek ugyancsak magasak lesznek (más rendszerekkel összehasonlítva), ráadásul sok adattárház projekt folytonos fejlődés alatt áll, azaz a fejlesztési munkák folyamatosan jelen vannak, igaz, csak néhány személyt érintő mértékben. Végül pedig: a hosszú fejlesztések miatt és a sok esetben évről évre növekvő ráfordításigény miatt nehéz feladat megtérülést számolni, illetve ha mégis sikerül, akkor kijelenthető: egy adattárház beruházás nagyon lassan térül meg [12][20].



## Konklúzió

Az persze vitathatatlan, hogy a felsorolt előnyök messze felülmúlják az említett hátrányokat, ám fontos kiemelni – mivel a legtöbb kudarcba fulladt projekt ennek köszönheti bukását – hogy a megfelelő igazolás és business case hiányában felelőtlen dolog adattárház fejlesztésbe kezdeni.

Ugyancsak kiemelendő, hogy a tapasztalatok azt mutatják, hogy az adattárház megléte és használata önerősítő folyamat: idővel egyre több részleg fog az adattárházhoz fordulni információ-kéréssel, riport-igénnyel, így az egyre inkább részévé válik majd a vállalatnak.

## Irodalomjegyzék

- [1] Dyché, Jill: e-Data: turning data into information with data warehousing. Addison-Wesley, 2000.
- [2] Hammergren, Thomas C. – Simon, Alan R.: Data Warehousing for Dummies. For Dummies, 2009.
- [3] Moss, Larissa Terpeluk – Atre Shaku: Business intelligence roadmap: the complete project lifecycle for decision-support applications. Addison-Wesley, 2003.
- [4] Jarke, Matthias: Fundamentals of data warehouses. Springer, 2003.
- [5] Khan, Arshad: SAP and BW Data Warehousing: How to Plan and Implement: iUniverse, 2005.
- [6] Galli Richard Péter: Software technology: Specialized development / launch method for BI applications, XX. SZÁMOKT 2010- Számítástechnika és oktatás Konferencia, Románia, Szatmárnémeti, 2010.
- [7] Westerman, Paul: Data warehousing: using the Wal-Mart model. Morgan Kaufmann, 2001.
- [8] Hosni, Yasser A. – Khalil, Tarek M.: Management of technology: opportunities and challenges for developed and developing regions of the world : selected papers from the Eleventh International Conference on Management of Technology. Emerald Group Publishing, 2004.
- [9] Kovács László Dr.: Adatbázisok tervezésének és kezelésének módszertana, Computerbooks Kiadó , 2004.
- [10] Raffai Mária: Információrendszerek fejlesztése és menedzselése. Novadat, 2003.
- [11] Davenport, Thomas – Prusak, Laurence: Working knowledge: how organizations manage what they know. Harvard Business Press, 2000.
- [12] Reeves, Laura L.: A Manager's Guide to Data Warehousing. Wiley, 2010.
- [13] Inmon, W. H.: Building the data warehouse. Wiley, 2002.
- [14] Hoberman, Steve: Data Modeler's Workbench. Wiley, 2002.
- [15] Ponniah, Paulraj: Data warehousing fundamentals – A Comprehensive Guide for IT

professionals. Wiley, 2001.

- [16] Chan, Yupo – Talburt, John – Talley, Terry M.: Data Engineering – Mining, Information and Intelligence. Springer, 2010.
- [17] Wang, John: Encyclopedia of Data Warehousing and Mining. Idea Group, 2006.
- [18] Schlukbier, Alan: Implementing Enterprise Data Warehousing – A Guide for Executives. Lulu.com, 2007.
- [19] Ullrey, Bruce Russel: Implementing a Data Warehouse – A Methodology That Worked. AuthourHouse, 2007.
- [20] Prabhu, C. S. R.: Data Warehousing: Concepts, Techniques, Products and Applications. PHI Learning, 2004.
- [21] SCN Education B.V.: Data Warehousing: The Ultimate Guide to Building Corporate Business Intelligence. BirkHauser, 2001.