



Adatmodellezés alapok

© 2012-2016 - Gollnhofer Gábor

Tematika

Mi az adatmodellezés?

- Mi ez? Miért kell? Alapelemek

Hogyan modellezzünk?

- Koncepcionális – logikai – fizikai adatmodell
- 5 (7) W

Adatmodellezési „módszerek”

- „Klasszikus” modellezés (egyed-kapcsolat avagy „relációs”)
- „Adattárház” adatmodellezés (Dimenziós, Data Vault)

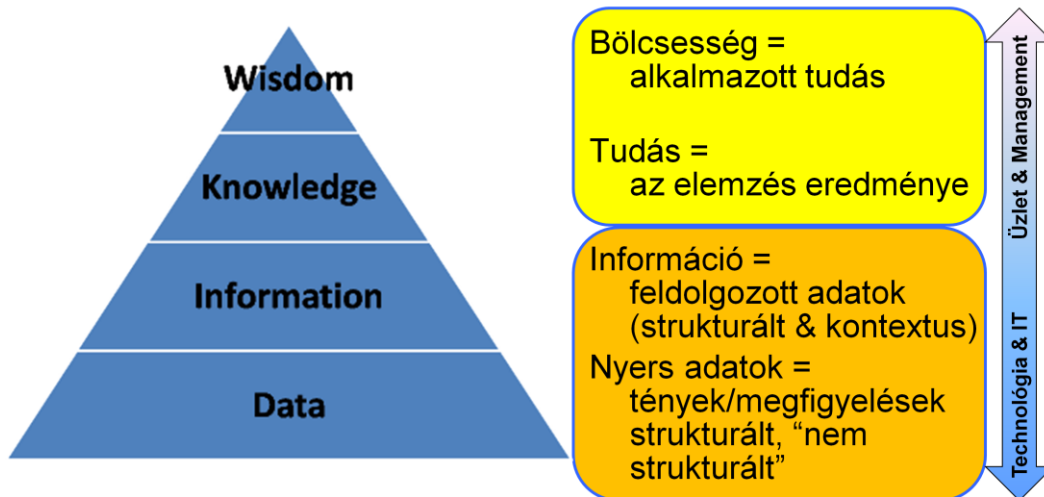
Érdekes adatmodellezési esetek

- Dátum/idő, hierarchiák

Egyéb adatmodellezési módszerek

Adatmodellezési eszközök

A „tudás piramis” (DIKW)



2016.06.14

Adatmodellezés workshop

3

Milan Zeleny
Russell Ackoff

Példa:

Időjárás és vízállásjelentés ☺

<https://www.youtube.com/watch?v=1F5fEUNrQmM>

Miért kell ezzel foglalkozni?

Az adatmodell a kialakítandó rendszer alapvető „térképe”

Itt kell definiálni, hogy:

- milyen területekkel, adatokkal foglalkozik a rendszer
- ezek között milyen összefüggések vannak / lesznek
- hogyan képezzük le a tényleges üzleti folyamatokat és mérőszámaikat

A kialakítandó rendszer alapjait határozza meg, ezért – legalábbis koncepcionális szinten – a felhasználónak is meg kell értenie és el kell fogadnia!

Miért kell ezzel foglalkozni?



Tematika

Mi az adatmodellezés?

- Mi ez? Miért kell? Alapelemek

Hogyan modellezzünk?

- Koncepcionális – logikai – fizikai adatmodell
- 5 (7) W

Adatmodellezési „módszerek”

- „Klasszikus” modellezés (egyed-kapcsolat avagy „relációs”)
- „Adattárház” adatmodellezés (Dimenziós, Data Vault)

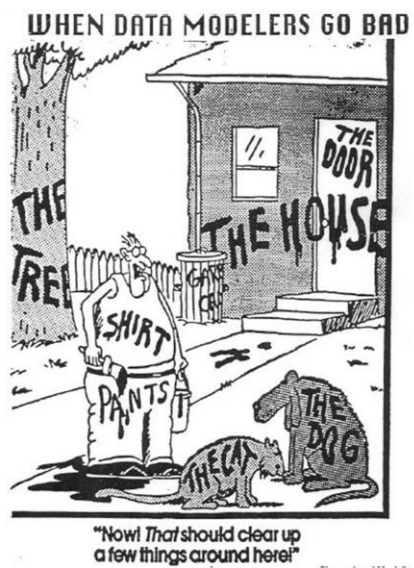
Érdekes adatmodellezési esetek

- Dátum/idő, hierarchiák

Egyéb adatmodellezési módszerek

Adatmodellezési eszközök

Hogyan modellezzünk?

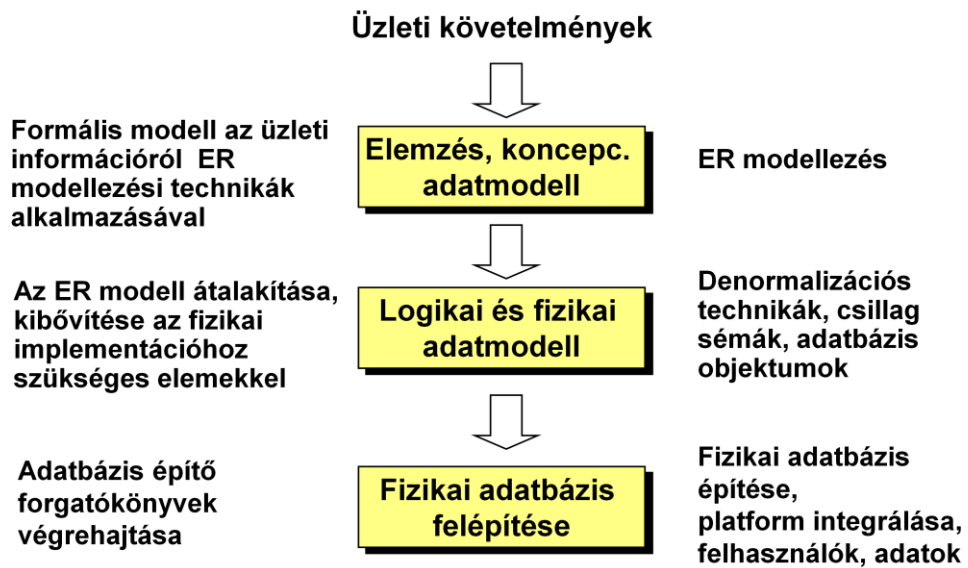


2016.06.14

Adatmodellezés workshop

7

Az adatmodellezés lépései



Koncepcionális modell

A rendszer kialakításának első lépése (elemzés).

Az üzleti folyamat(ok) és az érintett adatelemek strukturált leírása.

Azonosítja az üzleti entitásokat és a hozzájuk tartozó attribútumokat.

Kapcsolatot létesít az entitások között.

Implementáció független (nem foglalkozik a fizikai kialakítással)

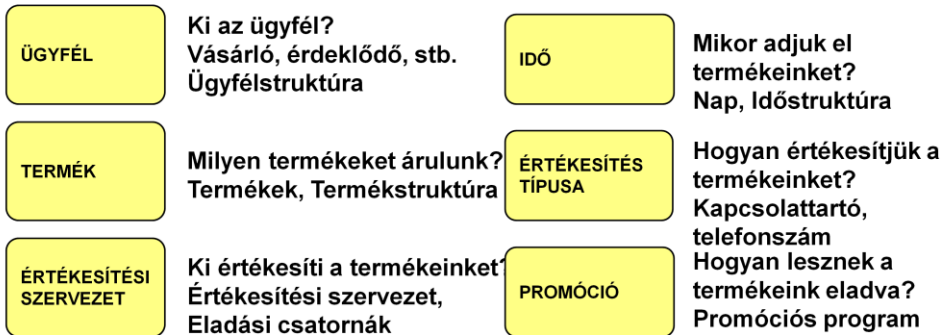
Koncepcionális adatmodell

Adatmodell, amely bemutatja az adat belső struktúráját. Független az adat egyedi alkalmazásaitól és az adat megjelenítésére és használatára alkalmazott szoftver vagy hardver mechanizmustól.

A koncepcionális adatmodell kapcsolatokkal és attribútumokkal ellátott.

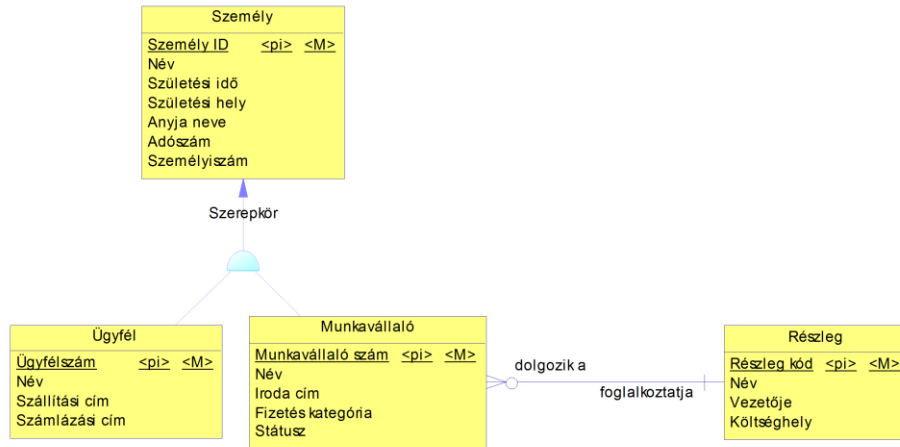
Koncepcionális adatmodell

Üzleti fogalmak



Koncepcionális adatmodell

Entitás, attribútum, kapcsolat



2016.06.14

Adatmodellezés workshop

12

Entitás - Egyed

Az entitás egy olyan „objektum”, amely a vállalat szempontjából fontos jellemzőket tartalmaz, és amelyekről információt kell tárolni (pl. személy, hely, tranzakció).

ÜGYFÉL

név
cím
születési idő
születési hely
anyja neve
stb.

Információ csoportokat definiál

Egy megnevezett dolog. Egy dolog osztálya vagy kategóriája.
Jellemzően egyes számú főnév.

Attribútum

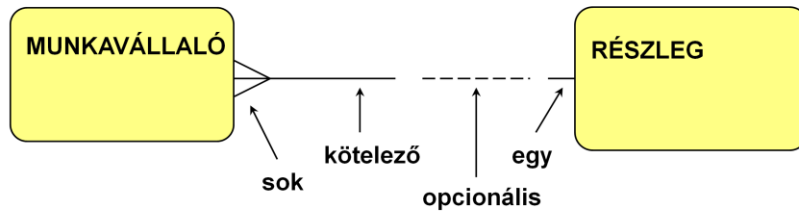
Az attribútum egy entitás tulajdonsága, fontos, specifikus információk az entitásokról (pl.: név, azonosító).
Mindegyik attribútum jellemzi az adott entitást.

MUNKAVÁLLALÓ
azonosító
név
cím
fizetési kategória
születési dátuma
státusa

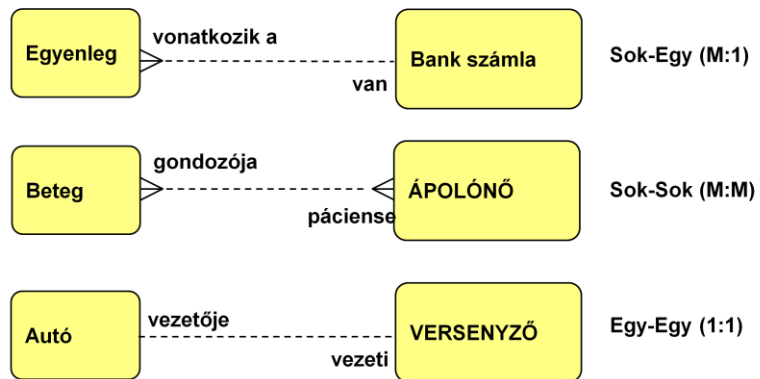
Az entításokhoz attribútumok tartoznak

Kapcsolatok

Az entitások viszonya egymáshoz.
Információ igényeket összekapcsoló üzleti szabályok.
Entitások közötti megnevezett asszociáció.



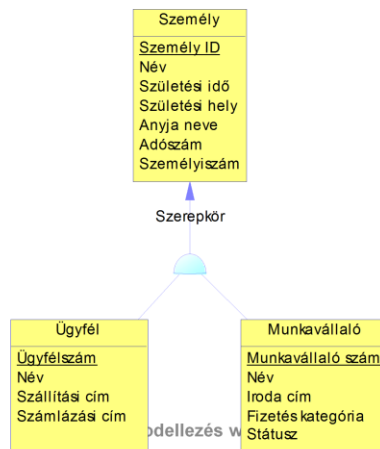
Kapcsolat típusok



Öröklés

2 vagy több entitás (gyerek) öröklí a szülő entitás attribútumait

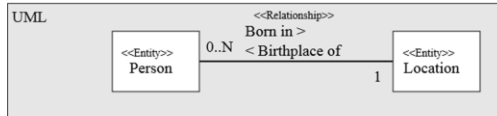
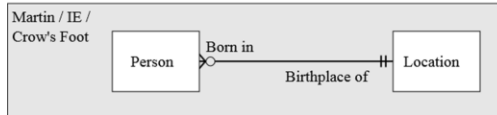
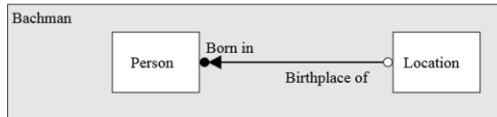
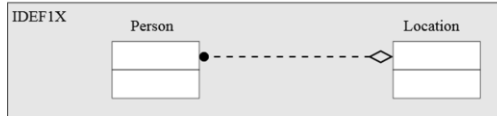
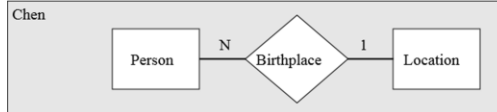
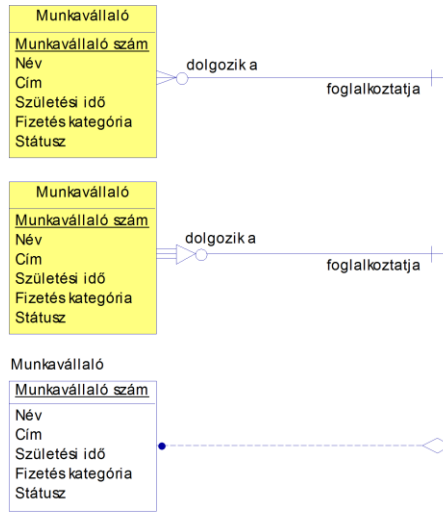
A gyerekeknek lehetnek saját attribútumai is



2016.06.14

17

Egyéb jelölés módok



2016.06.14

Adatmodellezés

Logikai adatmodell

A közbülső lépcső a koncepcionális és a fizikai adatmodell között;

Az adatbázis struktúrákat (tábla, view, index, kulcsok) definiálja;

A konkrét implementációs technológiától független;

Lehetőséget nyújt denormalizációra, stb.;

Jellemzően akkor használják, ha több adatbázis technológiát is támogatnia kell a rendszernek;

Fizikai adatmodell

A logikai (vagy koncepcionális) adatmodell rendszer specifikus megvalósítása.

Megfelel a fizikai/implementációs rendszer

- elnevezési konvencióinak,
- az ott elérhető adattípusok és hosszak használatával
- korlátozások definiálásával (pl. PK, UK, FK, check)
- index definíciók és
- nézetek alkalmazásával.

A fizikai adatmodellezés célja?

Séma kialakítása a fizikai adatbázisba történő bevezetésre

A fizikai környezet megtervezése, pl.:

- tárolási módok (partíciók, táblateretek)
- adatbázis specifikus jellemzők (identity, sequence, compressed, bitmap index)

A séma optimalizálása:

- a tervezett munkaterhelésre
- a hardver/szoftver környezetre

A koncepcionális és a fizikai adatmodell kapcsolata

Teljesen normalizált entitásokat használ

Entitás

Tábla

Az entitások táblákba vannak leképezve

Az entitásokat kapcsolatok kötik össze

Kapcsolat

Külső kulcs

A kapcsolatok külső kulcsokba vannak leképezve

Az attribútumok írják le az entitásokat

Attribútum

Oszlop

Az attribútumok oszlopokba vannak leképezve

Az EA különbözteti meg az entitás egyik példányát a másiktól

Egyedi Azonosító

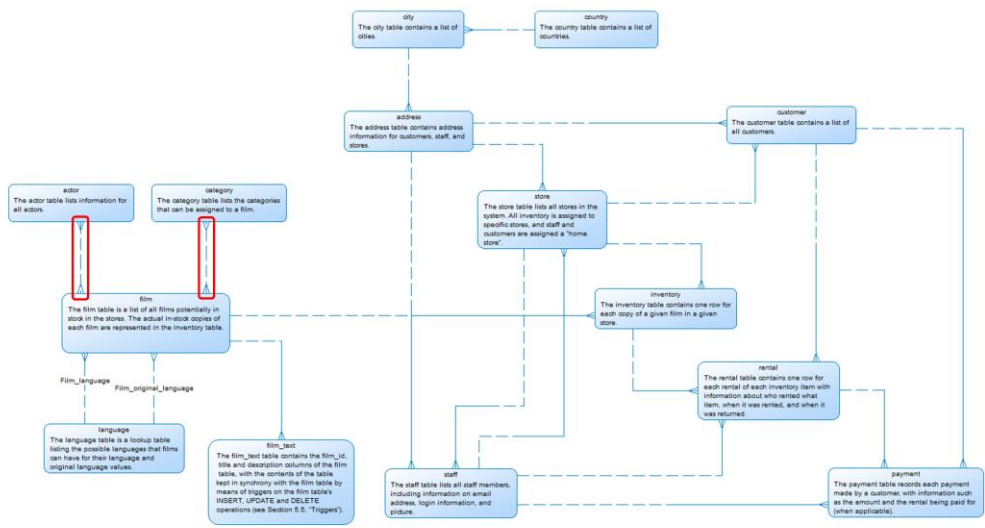
Elsődleges kulcs

Az egyedi azonosítók elsődleges kulcsokba vannak leképezve

Elsődleges kulcs (primary key) az adott tábla sorait egyértelműen azonosító kulcs (amennyiben több ilyen kulcs van, akkor azokat alternate key-nek nevezzük).

Külsőkulcs vagy idegen kulcs (foreign key) távoli táblára (elsődleges vagy alternate kulcsára) vonatkozó kulcs.

Konceptcionális adatmodell (MySQL demo: Sakila)



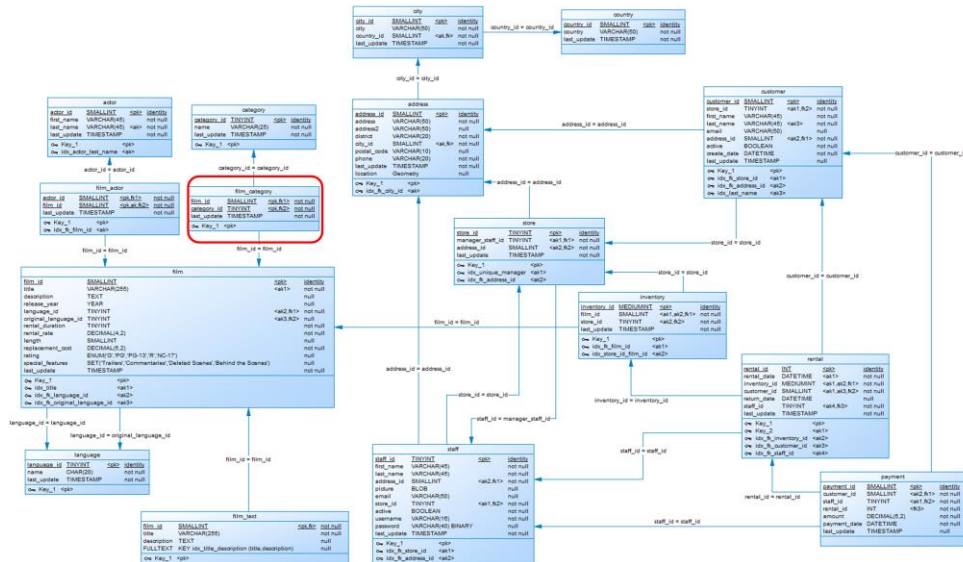
2016.06.14

Adatmodellezés workshop

23

Vegyük észre a több-többes kapcsolatokat

Fizikai adatmodell (MySQL demo: Sakila)



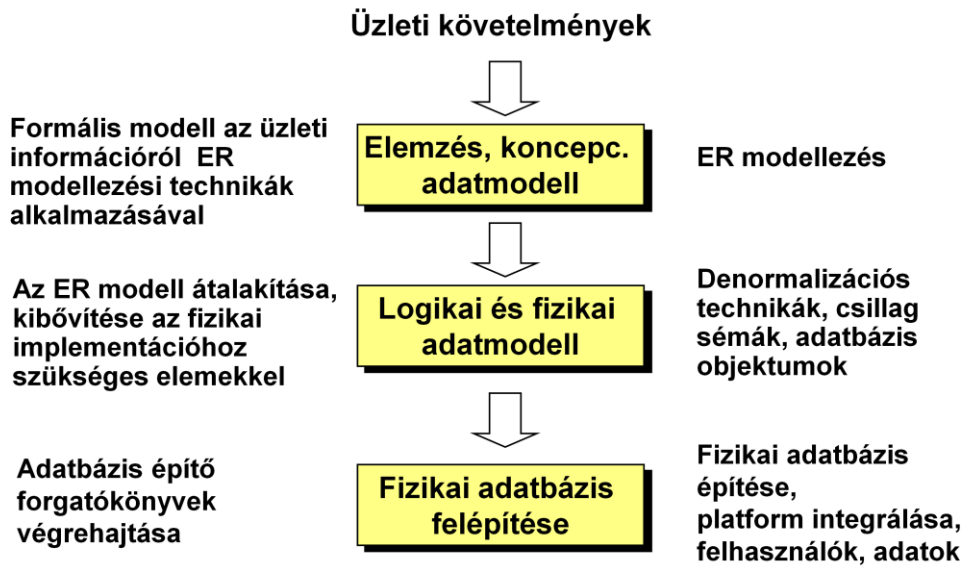
2016.06.14

Adatmodellezés workshop

25

Vegyük észre: a több-többes kapcsolatok feloldva (kapcsoló táblákkal)

Az adatmodellezés lépései



Az 5 (6-7) kérdés (5Ws [+1H])

Ki (Who)

Mit (What)

Miért (Why)

Hol (Where)

Mikor (When)

Hogyan (How)

Mennyi (how many)

Kivel/kinek (to whom)

Mivel (with what)

Rudyard Kipling: The Elephant's Child

„I keep six honest serving-men
(They taught me all I knew);
Their names are What and Why and When
And How and Where and Who”

Plusz:

http://en.wikipedia.org/wiki/Five_Ws

The BI Model Canvas		Table, Author(s), Date/Version
<p>When</p> <p>When does it happen?</p> <p>Date Time Time Zone Period Timeline: Milestones</p>	<p>How</p> <p>How does it happen? How do we know it happened? How do we uniquely identify a fact/event?</p> <p>Activity, Process, Event Effect, Outcome Transaction Type Transaction # [Degenerate Dimension] Step #</p> <p>[Granularity, Event Type: Discrete, Evolving, or Recurring]</p> <hr/> <p>How Many</p> <p>How many/much is involved? How long does it take?</p> <p>Revenues Costs Quantities, Balances Activity/Status Counts Durations</p> <p>[LoM, Fact Type: Fully Additive, Semi-Additive, Non-Additive]</p>	<p>Who</p> <p>Who does what? How do we organize them? How do they change? Who else is involved?</p> <p>Customer: Business, Consumer, Segment Employee Supplier Partner Third Party</p>
<p>Where</p> <p>Where does it happen? Where does it refer to?</p> <p>Location Brand, Store, Facility, Channel URL Map: Start, Previous, Current, Next, Last</p>	<p>Why</p> <p>Why does it happen? Can it be different? If so, why?</p> <p>Cause, Reason Promotion</p>	<p>What</p> <p>What is involved/used? How are they organized? How do they change?</p> <p>Value Proposition Product Service Resource</p>



Eredeti lásd: <http://modelstorming.com/>

Feladat #1

Bejövő adatok elemzése, modellezése

- Keressük meg az „érdekes” részeket
 - Mit tartalmaz?
 - Kulcs jellemzők?
 - Összefüggések
 - Eltérések
- Beszéljük meg közösen!

Bejövő adatok elemzése

Keressük az „érdekes” részeket!

METRO KERESKEDELMI KFT. BUDAPEST, KELETI 3. 2041 BUDAORS Tel: 23-449-100 Fax: 23-418-040 IBAN: HU72 10950009 00000002 53760139 Swift: BACCHUH3 Adószám: 10826643-2-44 Köztisztviselői adószám: HU10828663		SZÉKESFEHÉRVÁRI ÖNK. NAGY. ÁRHÁZ JÉZSUITA UTCA 1. 8000 SZÉKESFEHÉRVÁR Tel: 22-519-100 Fax: 22-519-110 Működési engedély szám: 3632		Oldal: 1 Számla dátuma: 2015.11.19. 13:53 Nyomatás dátuma: 19.11.2015. 13:54									
Számlasszám: 0/0 (021) 0004/036980 (004-025508) [*** EREDETI ***]		021/165		6635									
VÖLGYI MÁRIA JÉZSUITA UTCA 1. 8000 SZÉKESFEHÉRVÁR		Vevőszám: 21-REKORD: SC Adószám: 62713979-1-27 7250											
Cikkszám	VTSZ	Megnevezés	Csom. Db/ Egys. Csom.	Menny.	Nettó egys. súly	Egységár	Nettó ár	Afa	Bruttó ár	Kedv. kód	OP	DK	T
5998818171034	22030009	0.5 L STEFFL (DOB)	DB	24	1	0.500	3402,00		3402,00	1			4320,54
5998202640009	04012011	1L PARMALAT SZINTISZTA TEJ1,5% PD	PD	1	1	1,003	167,00		167,00	8			197,06
5998202640009	04012011	1L PARMALAT SZINTISZTA TEJ1,5% PD	PD	1	1	1,003	167,00		167,00	8			197,06
8594007980576	36200000	SHERON TELISZELVEDEK -20 SL	DB	1	1	4,942	899,00		899,00	1			1141,73
5998200744969	1806310099	6X30G PÖTTYÖS TÜRÖ RUDI NATUR CS	CS	1	1	0,180	396,00		396,00	1			502,92
5998200744969	1806310099	6X30G PÖTTYÖS TÜRÖ RUDI NATUR CS	CS	1	1	0,180	396,00		396,00	1			502,92
5998200744969	1806310099	6X30G PÖTTYÖS TÜRÖ RUDI NATUR CS	CS	1	1	0,180	396,00		396,00	1			502,92
5998200744969	1806310099	6X30G PÖTTYÖS TÜRÖ RUDI NATUR CS	CS	1	1	0,180	396,00		396,00	1			502,92
7622210145390	18063100	31G SPORT MAXI SZELET	DB	1	4	0,031	65,00		260,00	1			330,20
MINDENNAPI TERMÉKEK VERHETETLEN ÁRON - 6X30G PÖTTYÖS TÜRÖ RUDI NATUR							88,00-	1	111,76-				57243
Csoomag: 12		Összes súly: 19,792 KG		Nettó összesen:		6391,00							
				Nettó AFA %		6657,00 1+27,00%							
				AFA		334,00 8+18,00%							
				6391,00		1695,51							
				Bruttó összeg:		8086,51							
				Kerekített bruttó összeg:		8087,00							
				Bankkártya Online		8087,00							

TELJESÍTÉS DÁTUMA: 2015.11.19.

Fizetés: Azonnal

Továbbértékesítés esetén a számla a jövedéki termék szárazanyagának igazolására nem alkalmas.

EZEN SZÁMLA EGYEDI ADÁSVÉTELI SZERZŐDÉSNEK MINGISÜL POLGARI JOGI SZEMPONTBÓL A PTK. 6:215. §-A SZERINT. METRO FELÍR száma: A2689410

A számla 1 oldalból áll

***** A SZÁMLA ÉRTÉKE FORINTBAN ÉRTENDŐ *****

2016.06.14

30

Bejövő adatok elemzése Az „érdekes” részek

METRO KERESKEDELMI KFT. BUDAPEST, KELETI 3. 2041 BUDAORS Tel: 23-449-100 Fax: 23-418-040 IBAN: HU72 10950009 00000002 5370139 Swift: BACCHUH3 Adószám: 10826643-3-44 Központi adószám: HU10826663	SZEKESFEHÉRVÁRI ÖNK. NAGY. ÁRPHÁZ JEZSUITA UTCA 1. 8000 SZEKESFEHÉRVÁR Tel: 22-919-100 Fax: 22-919-110 Működési engedély szám: 3632	Dátum: 2015.11.19 13:53 Nyomtatás dátuma: 19.11.2015. 13:54																																																																																																																																																																																																																																																																														
Számlaszám: 0/0 (021) 0004/036980 (004-025508) [*** EREDETI ***] 021/165 6635																																																																																																																																																																																																																																																																																
VÁSÁRLÓ: JEZSUITA UTCA 1. 8000 SZEKESFEHÉRVÁR	Vevőszám: 21-REBETI-SC Adószám: 62713979-1-27 7250																																																																																																																																																																																																																																																																															
<table border="1"> <thead> <tr> <th>Cikkszám</th> <th>VTSZ</th> <th>Megnevezés</th> <th>Csom. Db/ Egsz.</th> <th>Db/ Csom.</th> <th>Menny.</th> <th>Nettó menny. súly</th> <th>Egységár</th> <th>Nettó ár</th> <th>Afa</th> <th>Bruttó ár</th> <th>Kedv. kód</th> <th>OP</th> <th>DK</th> <th>T</th> </tr> </thead> <tbody> <tr> <td>5998202640009</td> <td>04012011</td> <td>0.5 L STEFFL (DOB)</td> <td>DB</td> <td>24</td> <td>1</td> <td>0.500</td> <td>3402,00</td> <td>3402,00</td> <td>1</td> <td>4320,54</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>5998202640009</td> <td>04012011</td> <td>IL PARMALAT SZINTISZTA TEJ1.5% PD</td> <td>I</td> <td>1</td> <td>1</td> <td>1.003</td> <td>167,00</td> <td>167,00</td> <td>8</td> <td>197,06</td> <td></td> <td></td> <td></td> <td>P</td> </tr> <tr> <td>5998202640009</td> <td>04012011</td> <td>IL PARMALAT SZINTISZTA TEJ1.5% PD</td> <td>I</td> <td>1</td> <td>1</td> <td>1.003</td> <td>167,00</td> <td>167,00</td> <td>8</td> <td>197,06</td> <td></td> <td></td> <td></td> <td>P</td> </tr> <tr> <td>599820180578</td> <td>36200009</td> <td>SERDŐS TELISZELVEGŐ - 20.5L</td> <td>DB</td> <td>1</td> <td>1</td> <td>8.882</td> <td>899,00</td> <td>899,00</td> <td>1</td> <td>1141,73</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>5998200744969</td> <td>1806310099</td> <td>6X30G PÖTTYÖS TÜRÖ RUDI) NATÜR</td> <td>CS</td> <td>1</td> <td>1</td> <td>0.180</td> <td>396,00</td> <td>396,00</td> <td>1</td> <td>502,92</td> <td></td> <td></td> <td></td> <td>57243</td> </tr> <tr> <td>5998200744969</td> <td>1806310099</td> <td>6X30G PÖTTYÖS TÜRÖ RUDI) NATÜR</td> <td>CS</td> <td>1</td> <td>1</td> <td>0.180</td> <td>396,00</td> <td>396,00</td> <td>1</td> <td>502,92</td> <td></td> <td></td> <td></td> <td>57243</td> </tr> <tr> <td>5998200744969</td> <td>1806310099</td> <td>6X30G PÖTTYÖS TÜRÖ RUDI) NATÜR</td> <td>CS</td> <td>1</td> <td>1</td> <td>0.180</td> <td>396,00</td> <td>396,00</td> <td>1</td> <td>502,92</td> <td></td> <td></td> <td></td> <td>57243</td> </tr> <tr> <td>7622210145390</td> <td>18063100</td> <td>31G SPORT MAXI SZELET</td> <td>DB</td> <td>1</td> <td>4</td> <td>0.031</td> <td>65,00</td> <td>260,00</td> <td>1</td> <td>330,20</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td colspan="10">MINDENNAPI TERMEK VERHETETLEN ÁRON - 6X30G PÖTTYÖS TÜRÖ RUDI) NATÜR</td> <td>88,00</td> <td>1</td> <td>111,76-</td> <td>57243</td> <td></td> </tr> <tr> <td colspan="6">Csoomag: 12</td> <td colspan="2">Összes súly: 19.792 KG</td> <td colspan="2">Nettó összesen:</td> <td>6391,00</td> <td colspan="4"></td> </tr> <tr> <td colspan="6"></td> <td colspan="2">Nettó</td> <td colspan="2">AFA %</td> <td>AFA</td> <td colspan="4"></td> </tr> <tr> <td colspan="6"></td> <td colspan="2">6657,00</td> <td colspan="2">1=27.00%</td> <td>1635,39</td> <td colspan="4"></td> </tr> <tr> <td colspan="6"></td> <td colspan="2">334,00</td> <td colspan="2">8=18.00%</td> <td>60,12</td> <td colspan="4"></td> </tr> <tr> <td colspan="6"></td> <td colspan="2">6391,00</td> <td colspan="2"></td> <td>1695,51</td> <td colspan="4"></td> </tr> <tr> <td colspan="10"></td> <td colspan="2">Bruttó összeg:</td> <td colspan="3">8086,51</td> </tr> <tr> <td colspan="10"></td> <td colspan="2">Kerekített bruttó összeg:</td> <td colspan="3">8087,00</td> </tr> <tr> <td colspan="10"></td> <td colspan="2">Bankkártya Online</td> <td colspan="3">8087,00</td> </tr> </tbody> </table>	Cikkszám	VTSZ	Megnevezés	Csom. Db/ Egsz.	Db/ Csom.	Menny.	Nettó menny. súly	Egységár	Nettó ár	Afa	Bruttó ár	Kedv. kód	OP	DK	T	5998202640009	04012011	0.5 L STEFFL (DOB)	DB	24	1	0.500	3402,00	3402,00	1	4320,54					5998202640009	04012011	IL PARMALAT SZINTISZTA TEJ1.5% PD	I	1	1	1.003	167,00	167,00	8	197,06				P	5998202640009	04012011	IL PARMALAT SZINTISZTA TEJ1.5% PD	I	1	1	1.003	167,00	167,00	8	197,06				P	599820180578	36200009	SERDŐS TELISZELVEGŐ - 20.5L	DB	1	1	8.882	899,00	899,00	1	1141,73					5998200744969	1806310099	6X30G PÖTTYÖS TÜRÖ RUDI) NATÜR	CS	1	1	0.180	396,00	396,00	1	502,92				57243	5998200744969	1806310099	6X30G PÖTTYÖS TÜRÖ RUDI) NATÜR	CS	1	1	0.180	396,00	396,00	1	502,92				57243	5998200744969	1806310099	6X30G PÖTTYÖS TÜRÖ RUDI) NATÜR	CS	1	1	0.180	396,00	396,00	1	502,92				57243	7622210145390	18063100	31G SPORT MAXI SZELET	DB	1	4	0.031	65,00	260,00	1	330,20					MINDENNAPI TERMEK VERHETETLEN ÁRON - 6X30G PÖTTYÖS TÜRÖ RUDI) NATÜR										88,00	1	111,76-	57243		Csoomag: 12						Összes súly: 19.792 KG		Nettó összesen:		6391,00											Nettó		AFA %		AFA											6657,00		1=27.00%		1635,39											334,00		8=18.00%		60,12											6391,00				1695,51															Bruttó összeg:		8086,51													Kerekített bruttó összeg:		8087,00													Bankkártya Online		8087,00				
Cikkszám	VTSZ	Megnevezés	Csom. Db/ Egsz.	Db/ Csom.	Menny.	Nettó menny. súly	Egységár	Nettó ár	Afa	Bruttó ár	Kedv. kód	OP	DK	T																																																																																																																																																																																																																																																																		
5998202640009	04012011	0.5 L STEFFL (DOB)	DB	24	1	0.500	3402,00	3402,00	1	4320,54																																																																																																																																																																																																																																																																						
5998202640009	04012011	IL PARMALAT SZINTISZTA TEJ1.5% PD	I	1	1	1.003	167,00	167,00	8	197,06				P																																																																																																																																																																																																																																																																		
5998202640009	04012011	IL PARMALAT SZINTISZTA TEJ1.5% PD	I	1	1	1.003	167,00	167,00	8	197,06				P																																																																																																																																																																																																																																																																		
599820180578	36200009	SERDŐS TELISZELVEGŐ - 20.5L	DB	1	1	8.882	899,00	899,00	1	1141,73																																																																																																																																																																																																																																																																						
5998200744969	1806310099	6X30G PÖTTYÖS TÜRÖ RUDI) NATÜR	CS	1	1	0.180	396,00	396,00	1	502,92				57243																																																																																																																																																																																																																																																																		
5998200744969	1806310099	6X30G PÖTTYÖS TÜRÖ RUDI) NATÜR	CS	1	1	0.180	396,00	396,00	1	502,92				57243																																																																																																																																																																																																																																																																		
5998200744969	1806310099	6X30G PÖTTYÖS TÜRÖ RUDI) NATÜR	CS	1	1	0.180	396,00	396,00	1	502,92				57243																																																																																																																																																																																																																																																																		
7622210145390	18063100	31G SPORT MAXI SZELET	DB	1	4	0.031	65,00	260,00	1	330,20																																																																																																																																																																																																																																																																						
MINDENNAPI TERMEK VERHETETLEN ÁRON - 6X30G PÖTTYÖS TÜRÖ RUDI) NATÜR										88,00	1	111,76-	57243																																																																																																																																																																																																																																																																			
Csoomag: 12						Összes súly: 19.792 KG		Nettó összesen:		6391,00																																																																																																																																																																																																																																																																						
						Nettó		AFA %		AFA																																																																																																																																																																																																																																																																						
						6657,00		1=27.00%		1635,39																																																																																																																																																																																																																																																																						
						334,00		8=18.00%		60,12																																																																																																																																																																																																																																																																						
						6391,00				1695,51																																																																																																																																																																																																																																																																						
										Bruttó összeg:		8086,51																																																																																																																																																																																																																																																																				
										Kerekített bruttó összeg:		8087,00																																																																																																																																																																																																																																																																				
										Bankkártya Online		8087,00																																																																																																																																																																																																																																																																				
TELJESÍTÉS DÁTUMA: 2015.11.19 Fizetés: Azonnal *A számla elkészítéséhez a 22. sz. a jövedéki termék szárazanyag tartalommal igazolására nem alkalmas. EZEN SZÁMLA EGYEDI ADÁSVÉTELI SZERZŐDÉSNEK MINŐSÜL. POLGÁRI JOGI SZEMPONTBÓL A PTK. 6:215. §-A SZERINT. METRO FELIR száma: A26899410 A számla kódja: 41	***** A SZÁMLA ÉRTÉKE FORINTBAN ÉRTENDO *****																																																																																																																																																																																																																																																																															

2016.06.14

31

Tematika

Mi az adatmodellezés?

- Mi ez? Miért kell? Alapelemek

Hogyan modellezzünk?

- Koncepcionális – logikai – fizikai adatmodell
- 5 (7) W

Modellezési „módszerek”

- „Klasszikus” modellezés (egyed-kapcsolat avagy „relációs”)
 - Normál formák, normalizálás és denormalizálás
- „Adattárház” adatmodellezés (Dimenziós, Data Vault)

Érdekes modellezési esetek

- Dátum/idő, hierarchiák

Modellezési eszközök

Relációs „objektumok”

- táblázatok (table, entity, relation) - egyedtípusok, állományok
- táblázat sorai (tuple) - egyedelőfordulások, rekordok,
- táblázat oszlopai (attribute) - tulajdonságtípus, mező
- táblázat elemei (domain, domén) – tulajdonság előfordulás, jellemzők lehetséges értékei, adatok
- táblázat foka (degree) - a mezők, attribútumok száma
- kulcs jelölt (candidate key) – azonosító jelölt
- logikai azonosító (primary key) - a táblázat elsőleges kulcsa (egyedi index)
- összetett azonosító (composite key) - két vagy több mezőből álló kulcs
- másodlagos kulcs (alternate key)
- idegen kulcs (foreign key)
- null érték (NULL value) - ismeretlen vagy meg nem adott mezőérték

Normál formák

A logikai tervezés célja egy redundancia mentes reláció rendszer, relációs adatbázis megtervezése.

Normalizálás: módszer a redundancia megszüntetésére.

Öt v. hat normál formát különböztetünk meg.

A különböző normál formák egymásra épülnek

A legmagasabb normál forma elérése a cél.

Az első három normál forma a funkcionális függőségekben található redundanciák, a negyedik és ötödik a többértékű függőségekből adódó redundanciák megszüntetésére koncentrálnak.

A relációs modell eredeti definíciója E.F. Codd 1970

Példák (részben): dr. Siki Zoltán – Adatbáziskezelés és szervezés

Függések

Funkcionális függés

- Legyen X és Y két egyszerű vagy összetett tulajdonságtípus.
- Ha az X tulajdonságtípus az Y tulajdonságtípus egyetlen értékét határozza meg, akkor **egyértékű függés** létezik X-ből Y-ba. Ekkor azt mondjuk, hogy Y **funkcionálisan függ** X-től.
- Ha X és Y tulajdonság-típusok között egyértékű függés áll fenn, akkor a relációs táblázatban nem lehet két olyan sort találni, ahol az X értéke megegyezik, de az Y tartalma különböző.

Többértékű függés

- Ha az X tulajdonságtípus egy értékéhez az Y értékeinek egy halmaza tartozik, akkor **többértékű függés** létezik X-ből Y-ba.

Normál formák (1-3)

1. Egy reláció első normál formában van, ha minden attribútuma egyszerű, nem összetett adat.
2. A reláció akkor és csak akkor van második normál formában, ha első normál formában van és minden nem elsődleges attribútuma teljes funkcionális függőségben van az összes reláció kulccsal.
3. Egy reláció akkor és csak akkor van harmadik normál formában, ha második normál formában van és nem tartalmaz funkcionális függőséget a nem elsődleges attribútumok között.

Első normál forma - 1NF

		Szakkörök		
tábla oszlop	Szakkör	Tanár	Diákok	
			Név	Osztály
1.sor	Számítástechnika	Nagy Pál	Kiss Rita	III.b
			Álmos Éva	II.c
2.sor	Video	Gál János	Név	Osztály
			Réz Ede	I.a
			Vas Ferenc	II.b

Ez nem 1NF!

Szakkörök			
Szakkör	Tanár	Diák	Osztály
Számítástechnika	Nagy Pál	Kiss Rita	III.b
Számítástechnika	Nagy Pál	Álmos Éva	II.c
Video	Gál János	Réz Ede	I.a
Video	Gál János	Vas Ferenc	II.b

Második normál forma – 2NF

Konferencia

Terem	Időpont	Előadás	Férőhely
B	10:00	Mitológia	250
A	8:30	Irodalom	130
B	11:30	Színház	250
A	11:00	Festészet	130
A	13:15	Régészet	130

Ez nem
2NF!

Konferencia

Terem	Időpont	Előadás
B	10:00	Mitológia
A	8:30	Irodalom
B	11:30	Színház
A	11:00	Festészet
A	13:15	Régészet

Termek

Terem	Férőhely
A	130
B	250

Harmadik normál forma - 3NF

Szakkörök

Szakkör	Tanár	Születési év
Képzőművész	Sár Izidor	1943
Iparművész	Sár Izidor	1943
Karate	Erős János	1972

**Ez nem
3NF!**

Szakkörök

Szakkör	Tanár
Képzőművész	Sár Izidor
Iparművész	Sár Izidor
Karate	Erős János

Tanárok

Tanár	Születési év
Erős János	1972
Sár Izidor	1943

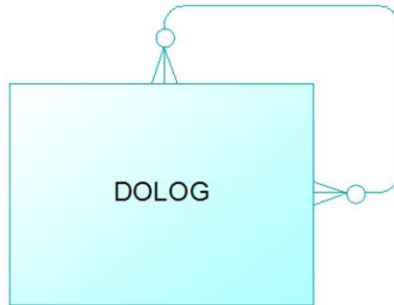
További normál formák

Általában „megelégszünk” a 3NF-el, de azért vannak még további normál formák is:

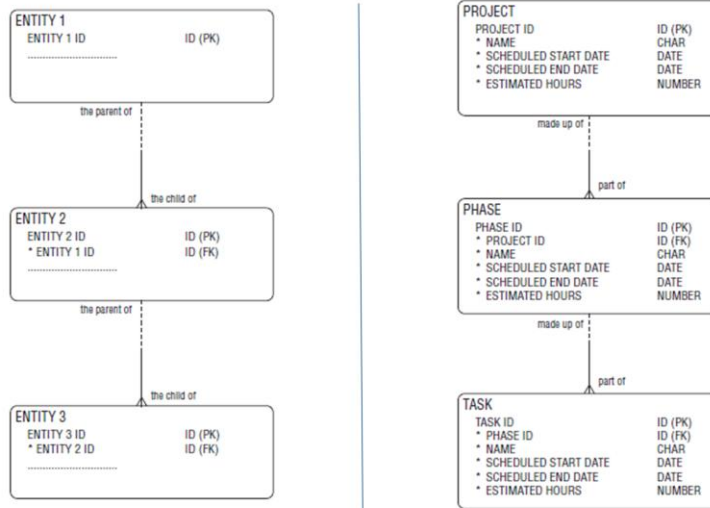
- BCNF (Boyce-Codd normál forma)
- 4NF
- 5NF
- 6NF (két „verzió”: A – temporal; B – domain/key normal form [DKNF])

Több jó modell is létezhet!!! Specifikus vs. absztrakt

A világegyetem modellje:



Specifikus vs. Absztrakt pl. rekurzív struktúrák 1. szint



Forrás:

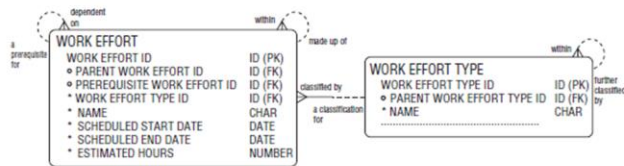
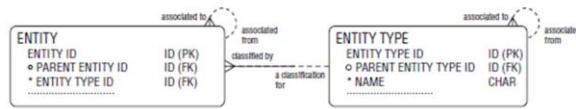
Len Silverston, Paul Agnew:
The Data Model Resource Book: Volume 3: Universal Patterns for Data Modeling

2016.06.14

Adatmodellezés workshop

42

Specifikus vs. Absztrakt pl. rekurzív struktúrák 2. szint



Forrás:

Len Silverston, Paul Agnew:

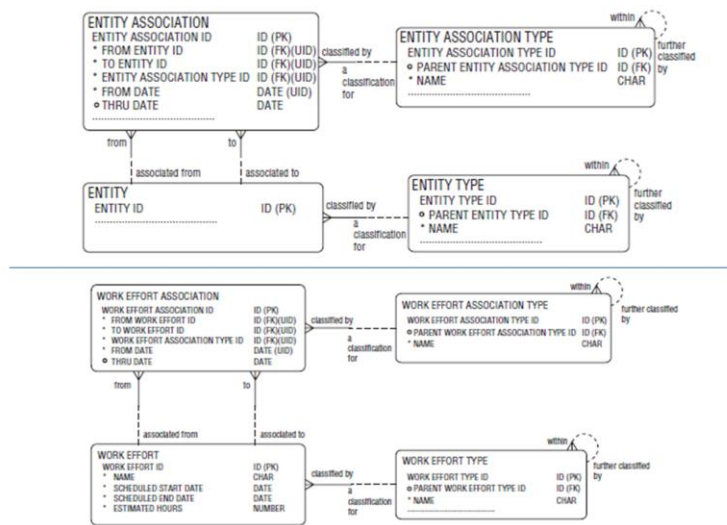
The Data Model Resource Book: Volume 3: Universal Patterns for Data Modeling

2016.06.14

Adatmodellezés workshop

43

Specifikus vs. Absztrakt pl. rekurzív struktúrák 3. szint



Forrás:

Len Silverston, Paul Agnew:

The Data Model Resource Book: Volume 3: Universal Patterns for Data Modeling

2016.06.14

Adatmodellezés workshop

44

A modell ellenőrzése #1

Tartalmi ellenőrzés

1. Mennyire jól írja le a követelményeket?
2. Mennyire teljes körű?
pl. entitások, attribútumok, kapcsolatok, definíciók!,
ID+BK („valós” üzleti kulcs)
3. Strukturálisan jó?
pl. körkörös vagy redundáns kapcsolatok, hiányzó
kulcsok, hibás hivatkozások, eltérő adattípusok
4. Mennyire használja az „általános” struktúrákat?
pl. dátum, deviza, országkód, személyi szám, stb.
de nem túl általános (pl. a világegyetem modellje 😊)
5. Mennyire konzisztens a vállalati modellel?
pl. „közös” entitások, kapcsolódási pontok

A modell ellenőrzése #2

Formai ellenőrzés

1. Mennyire olvasható?
Jó megjelenés, megfelelően tagolt/csoportosított, specifikus diagramok és részletezettség, olvasható nevek, követhető vonalak, stb.
2. Mennyire követi a névkonvenciókat?
(Ugye van névkonvenció?)
3. Mennyire jók a definíciók?
Érthető, tiszta, pontos, teljes körű definíciók
4. Mennyire illeszkedik a valós környezetbe?
Kivitelezhető? Fejlesztési idő, rendszer sebesség (indexek?), rugalmasság, hatékonyság, stb.

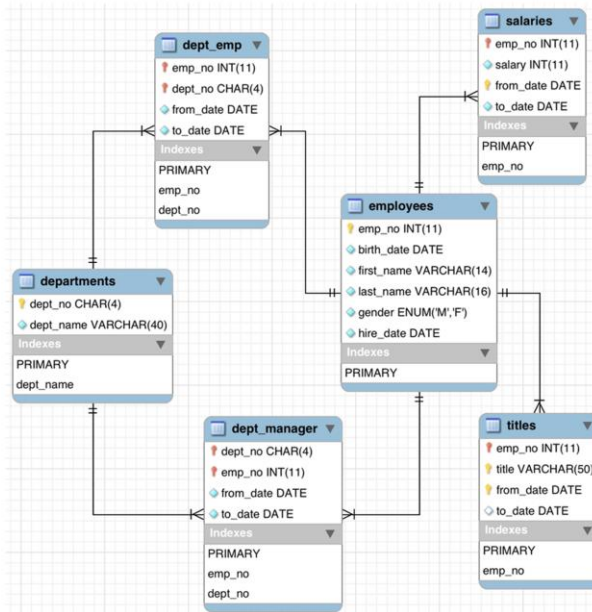
Feladat

Keress „hibákat” a modellben

- Keressük a „gyanús” részeket!
- Milyen feltételezésekkel éltél?
- Beszéljük meg közösen!

Feladat #2a

Keress „hibákat” a modellben!



2016.06.14

Adatmodellezés workshop

48

1. `emp_no` <> `dept_no` (int <> char) – nem konzisztens
2. egy employee csak egyszer léphet be a céghez
3. nincs `employees.leave_date`
4. stb.

Tematika

Mi az adatmodellezés?

- Mi ez? Miért kell? Alapelemek

Hogyan modellezzünk?

- Koncepcionális – logikai – fizikai adatmodell
- 5 (7) W

Modellezési „módszerek”

- „Klasszikus” modellezés (egyed-kapcsolat avagy „relációs”)
 - Normál formák, normalizálás és denormalizálás
- „Adattárház” adatmodellezés (**Dimenziós, Data Vault**)

Érdekes modellezési esetek

- Dátum/idő, hierarchiák

Modellezési eszközök

Denormalizálás

Fizikai adatmodellezési technika

Általában a nagyobb lekérdezési teljesítmény érdekében alkalmazzák

A normalizált táblák összevonását, attribútumok redundáns tárolását, kalkulált értékek, aggregátumok képzését jelenti.

Hátránya hogy a redundáns adatokat több helyen kell karbantartani, valamint tárolásuk is több helyet igényel

Adattárház modellezés

Problémák a 3NF modellekkel

- Tipikusan a tranzakciós rendszerekhez optimalizált
- Nagyon normalizált, akár több száz entitás
- A felhasználók számára nehezen átlátható, lekérdezhető
- Sokszor bonyolult, lassú, „drága” join-ok segítségével kérdezhető csak le

Adattárház modellezési módszerek

- Dimenziós modellezés
- Data Vault modellezés
- Anchor modelling
- Egyéb: flat & wide;

Dimenziós adatmodellek

Főleg az elemzés jellegű felhasználások támogatására kialakított modellezési technika

Céljai:

- a felhasználók számára is könnyen átlátható, érthető, lekérdezhető struktúrában tárolni az adatokat
- nagy sebességű lekérdezések, ad-hoc lekérdezések támogatása

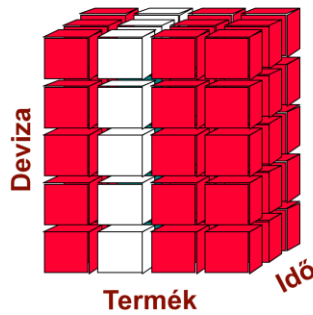
Több dimenziós nézet

A dimenziók a különböző vetületek, szempontok, „a kocka élei”

A mérőszámok az egyes kocka elemekben szereplő értékek

Dimenzió példák

- Idő
- Vállalat
- Deviza
- Termék
- Ügyfél
- Partner



Mérőszám példák

- Darab
- Összeg
- Árfolyam
- Időtartam

Dimenzió, attribútum, hierarchia

Dimenzió

- Vizsgálati szempontok (idő, termék, vevő, ...)
- A dimenziók attribútumokból, és hierarchiákból épülnek fel.

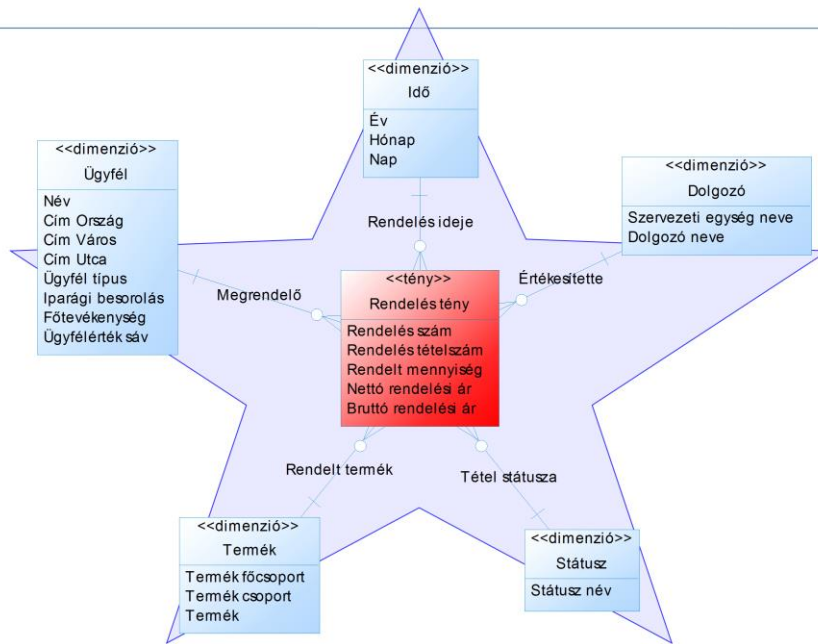
Attribútum

- A dimenzió elemek tulajdonsága (pl. hét napjai, munkanap, ünnepnap)
- Lehet egy hierarchia szintjei (év, negyedév, hónap, nap,...)

Hierarchia

- Az attribútumok láncolata (év alatta negyedév, alatta hónap, ...)
- Fajtái:
 - Természetes (pl. év-negyedév-hónap-nap)
 - Reporting (pl. szervezeti hierarchia, termék hierarchia)

Csillag séma, a dimenziós adatmodell



2016.06.14

Adatmodellezés workshop

56

Csillag séma, a dimenziós adatmodell

Csillag séma

Nagy központi tény tábla

- Az egyes üzleti esemény adatait tartalmazza

Körülötte kisebb dimenzió táblák

- Az esemény „kontextusát” írja le
- A tényadat szűrésére, szeletelésére „slice-and-dice”

Tények és dimenziók

A jó tényadatok jellemzően numerikusak és összesíthetők (additive vagy semi-additive)

- Ez fontos, mert a legtöbb elemzésben egyidejűleg több tényadat sort kérdezőnk le

A dimenziók jellemzően leíró jellegű adatokat tartalmaznak

- általában ezeken szűrünk (where) és ezek a fejlécek (select lista)
- a tipikus csillagokban 4 – 10 (15) dimenzió van

Dimenziós és 3NF modellek kapcsolata

Általában egy 3NF modell több csillaggal ábrázolható

A kialakítás/átalakítás alapvető lépései:

- A koncepcionális modellt bontsuk több modellre az önálló üzleti folyamatok mentén
- A sok-sok kapcsolatokat (akár néhány nem additív [nem PK vagy FK] tény adattal) alakítsuk ténytáblává
- A maradék táblákat denormalizáljuk, „terítsük ki” dimenzió táblákká
- A dimenzió tábláknak adjunk mesterséges elsődleges kulcsot

Normalizált adatmodell

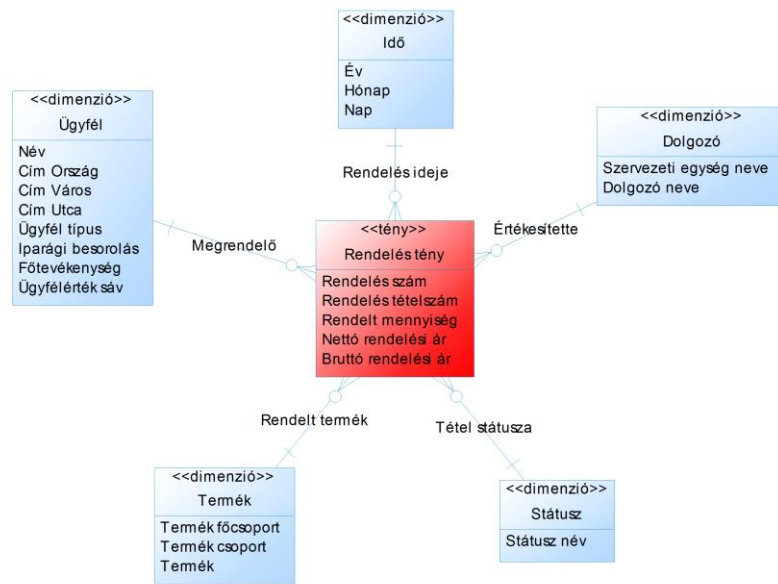


2016.06.14

Adatmodellezés workshop

60

Dimenziós adatmodell



2016.06.14

Adatmodellezés workshop

61

Dimenziók, tények, kapcsolatok

Tény -> Dimenziók

- „normál” csillag
- mindig, 1:M kapcsolat

Tény -> Tény

- tény csoportok, családok, általános – specifikus tény
- Ritkán, 1:M vagy 1:1 vagy 1:0 kapcsolat

Dimenzió -> Dimenzió

- hópehely séma
- dimenzió outrigger, hierarchia kezelés, aggregátumok

Dimenziók

Egy dimenziós tábla általában

- egy mesterséges elsődleges kulcsot (DW ID),
- leíró jellegű szöveges vagy numerikus mezőket és
- technikai/metaadatokat tartalmaz.

A dimenziós táblák:

- tartalmazhatnak viszonylag nagy számú sort is (akár tízmillió nagyságrend),
- de a táblák mérete általában nem jelentős.
- A dimenziós tábla „mindig” nagyságrendekkel kisebb, mint a ténytáblák (1:N kapcsolat)

Dimenziók kulcsai

Mesterséges kulcsok

- Mesterséges azonosító mező (surrogate keys)
 - Általában futó sorszám (integer, number) – nincs belső „értelme”
 - Fontos a kis méret és a gyors generálhatóság
- Célja:
 - Függetlenséget biztosít a forrásrendszeri azonosítóktól
 - Támogatja a változáskezelést
- „Minden” dimenzióknak az elsődleges kulcsa ilyen legyen
 - Kivéve <<Dátum>> dimenzió
- A ténytáblának is lehet, de nem feltétlenül szükséges

Természetes kulcsok

- A dimenziós táblában jellemzően attribútumként jelennek meg
- A dimenzió karbantartás során használhatjuk (lásd. történetiség kezelés)

Dimenzió történetiség kezelés típusok

A dimenzió rekordok történetiségének kezelése
SCD = Slowly Changing Dimension, avagy
lassan változó dimenzió

- SCD1 – felülírás
- SCD2 – új rekord felvétele
- SCD3 – oszlop duplikálás
- Vegyes

Lehetnek gyorsan változó dimenzió jellemzők is

SCD1 - felülírás

Kezelése:

- Mindig az utolsó értéket látjuk
- Az adott jellemzőt egyszerűen felülírjuk/update

Előnye:

- Könnyű és gyors a karbantartás
- Nem kell kulcsokat módosítani, a táblák érintetlenek maradnak

Hátránya:

- A múltbeli állapotokról nem tartalmaz információt

Alkalmazása:

- Ha értelemszerűen nem módosulhat (pl. születési dátum)
- Ha üzleti szempontból érdektelen a korábbi értéke (pl. személyi igazolvány száma)

SCD2 – új rekord felvétele

Kezelése:

- Új rekordot veszünk fel, új mesterséges kulccsal
- Az előző rekord érvényességét lezárjuk (valid_from, _to, _flag)
- A rekordokat a természetes/üzleti kulcs köti össze
- A régi tények a régi, az újak az új dimenzió rekordra mutatnak

Előnye:

- Megmarad a történeti adat
- Minden tény rekord a megfelelő dimenzió rekordra mutat (?)

Hátránya:

- Növekszik a dimenzió tábla mérete
- Nagy és gyorsan változó dimenzió spec.kezelést igényel

Alkalmazása:

- Ha üzleti szempontból érdekes a történetiség („alapértelmezett”)

SCD3 – duplikált oszlop

Kezelése:

- Új oszlopot veszünk fel, az előző érték tárolására
- A régi oszlop értékét felülírjuk
- Pl. „termék név” => „előző termék név”

Előnye:

- Nem kell tény rekordot módosítani

Hátránya:

- Ritkán használt
- Csak korlátozottan követi a változásokat (aktuális és előző érték)
- Csak a jól tervezhető/ritka változásoknál célszerű

Alkalmazása:

- Alternatív valóságok támogatása (akkor és most)

Vegyes – SCD 6 (1+2+3 v. 1*2*3)

Kezelése:

- Az SCD2-es táblában SCD2-es oszlopban tároljuk a történeti adatot és plusz SCD1-es oszlopot veszünk fel az aktuális érték tárolására
- Új rekordot veszünk fel, új mesterséges kulccsal
- Az előző rekord érvényességét lezárjuk (valid_from, _to, _flag)
- Az SCD1-es oszlop értékét felülírjuk minden állapotban

Előnye:

- Alternatív valóság könnyen kezelhető

Hátránya:

- Csak a jól felkészült elemző tudja használni/érteni meg
- Sok ilyen oszlopnál nagyon széles lehet a dimenzió tábla

Alkalmazása:

- Alternatív valóságok támogatása (akkor és most), teljes történetiség kezeléssel

Tények

Az adott üzleti folyamatra vonatkozó adatok

- Kontextus (dimenziókra mutató idegen kulcsok)
- Degenerált dimenziók (lásd. 2.blokk)
- Mérőszámok
- Technikai/metaadat mezők

Lehet nagyon nagy méretű

- Sok sorral sok oszloppal
- A tábla szélessége és fizikai paraméterei is fontosak (adattípusok, indexek, particionálás, tömörítés, stb.)

Tények

Részletezettség (grain)

- A tényadatok felbontásának jellemzője
- Általában a dimenziók határozzák meg
- Egy tényben ne keverjük a különböző felbontású sorokat
- Ugyanazt a folyamatot több felbontásban is ábrázolhatjuk pl. egyedi tranzakciók, nap végi egyenleg

Tények

Mérőszámok

- Általában numerikusak
- Általában összeadható (additive)
 - Pl. rendelt termék ára
- Vagy részben összeadható (semi-additive)
 - Pl. bankszámla egyenleg
- Vagy nem összeadható (non-additive)
 - Pl. hőmérséklet, csapadék mennyiség

Tények

Kontextus

- Dimenziókra mutató idegen kulcsok
 - A dimenziók mesterséges DW ID-jára mutat
 - Nem a természetes (üzleti, forrásrendszeri kulcsra)
 - Általában definiált, de nem kikényszerített
 - Oracle: rely, disable, novalidate
- Speciális dimenziók
 - A dátum és idő
 - Sok esetben többször is kapcsolódik, szerepjátszó dimenzió

Tény típusok és történetiség kezelése

Tény típusok

- Tranzakció, esemény (Transaction)
- Pillanat felvétel (Periodic Snapshot)
- Gyűjtött pillanat felvétel (Accumulating Snapshot)

Az egyes típusok kiválasztása az adatok és az elemzési igények függvényében történik

Egy adott üzleti folyamathoz több típusra is szükség lehet

Tény típusok összefoglalása

	Transaction	Periodic Snapshot	Accumulating snapshot
Periodicitás	Konkrét tranzakció v. esemény	Előre ütemezett időszakonként	Folyamat előrehaladásától függő időtartamonként
Részletezettség (grain)	1 sor tranzakciónként ill. tran.soronként	1 sor időszakonként és dimenzióként	1 sor folyamat lépésenként
Dátum dimenzió(k)	Tranzakció dátuma	Snapshot dátuma	Több dátum, az egyes lépésekhez igazodóan
Tények	Tranzakció mértékei	Az időszakra összegzett mértékek	Az egyes lépésekre jellemző mértékek
Kitöltöttség	Ritka v. sűrű az aktivitástól függően	Tervezhetően sűrű	Ritka v. sűrű a folyamat eseményektől függően
Tény módosítások	Nincs (csak hibajavítás)	Nincs (csak hibajavítás)	Van, minden folyamat esemény bekövetkeztekor

Dimenziós modell tervezése

Felmérés, koncepcionális modell

- 7 kérdés, busz mátrix

Csillag séma tervezése

- Üzleti folyamat kiválasztása (busz mátrix sor)
- A mérendő tényezők azonosítása
- A részletezettség (grain) meghatározása
- A dimenziók azonosítása és tervezése
- A mértékek azonosítása és tervezése

Iteratívan végig a busz mátrix sorain

Az 5 (6-7) kérdés (5Ws [+1H])

Ki (Who)

Mit (What)

Miért (Why)

Hol (Where)

Mikor (When)

Hogyan (How)

Mennyi (how many)

Kivel/kinek (to whom)

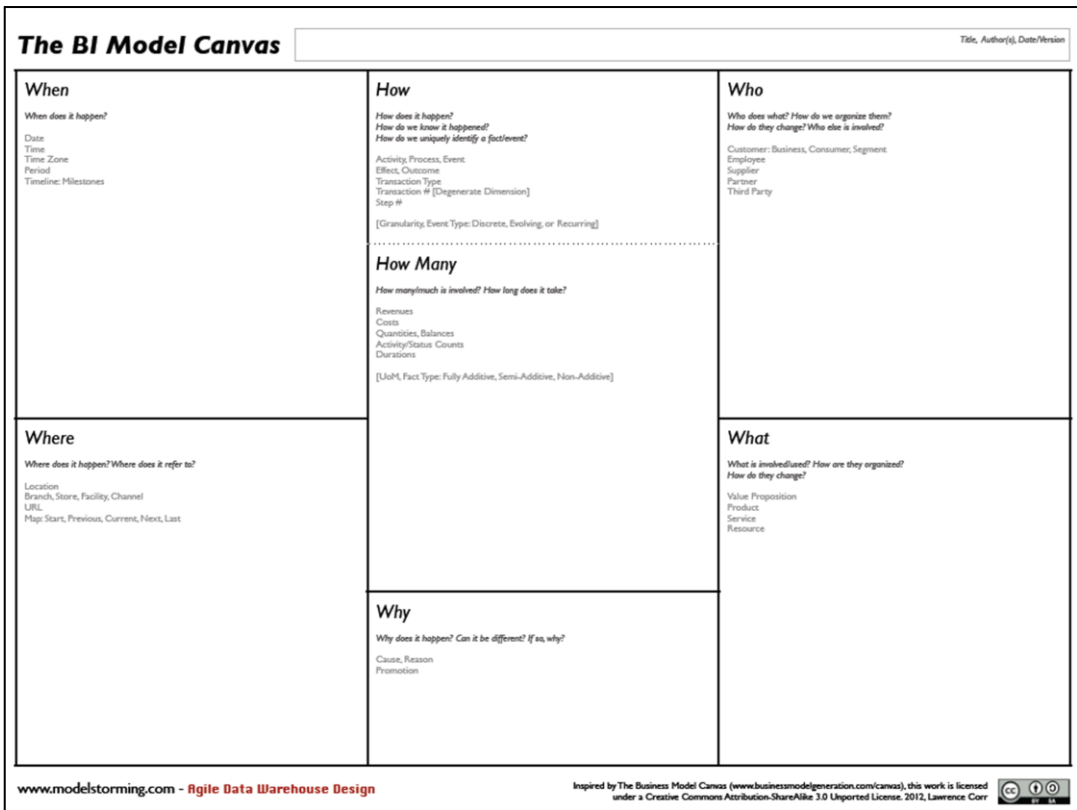
Mivel (with what)

Rudyard Kipling: The Elephant's Child

„I keep six honest serving-men
(They taught me all I knew);
Their names are What and Why and When
And How and Where and Who”

Plusz:

http://en.wikipedia.org/wiki/Five_Ws



Eredeti lásd: <http://modelstorming.com/>

Busz mátrix

Számítógép rendszer-sín (adatbusz) analógiájára

- Folyamatok/mértékek (busz)
- Leíró jellemzők/dimenziók (csatlakozó elemek)

Üzleti folyamatok	Dimenziók							
	Dátum	Termék	Bolt	Ügyfél	Raktár	Gyártó	Szállító	Csatorna
Bolti értékesítés	X	X	X	X				
Bolti készletezés	X	X	X					
Bolti kiszállítás	X	X	X					
Raktári készletezés	X	X			X			
Beszerezések	X	X			X	X	X	
Ügyfél panaszok	X	X	X	X				X

Tervezési „szabályok” - 1

Az összes metaadat oszlop legyen benne minden táblában

- Legfeljebb fix értékekkel lesz töltve

Tény tábláknál is kell:

- „last_update_time” és „valid_flag”

Névkonvenciók

- prefixelés: D_, F_

Érvényesség kezelés (dátumok)

- zárt-nyitott intervallum, azaz {kezdet} >= és < {vége}

Tervezési „szabályok” – 2

Oszlop szerepkör típusok

Szerepkör	Leírás
ID	mesterséges azonosító (v. DWID)
HASH	mesterséges azonosító (alternatív)
BK	business key
DAT	dátum (ID)
ATTR	leíró attribútum
MEAS	mérőszám (measure)
DEG	degenerált dimenzió
PK	elsődleges kulcs (primary key)
FK	ldegen kulcs (foreign key)
FK,DAT	foreign key, dátum (ID)
META	metaadat
META,DAT	metaadat, dátum (ID)
META,ID	metaadat, azonosító
TECH	technikai
TECH,DAT	technikai, dátum (ID)
TECH,ID	technikai, azonosító

Tematika

Mi az adatmodellezés?

- Mi ez? Miért kell? Alapelemek

Hogyan modellezzünk?

- Koncepcionális – logikai – fizikai adatmodell
- 5 (7) W

Modellezési „módszerek”

- „Klasszikus” modellezés (egyed-kapcsolat avagy „relációs”)
 - Normál formák, normalizálás és denormalizálás
- „Adattárház” adatmodellezés (Dimenziós, **Data Vault**)

Érdekes modellezési esetek

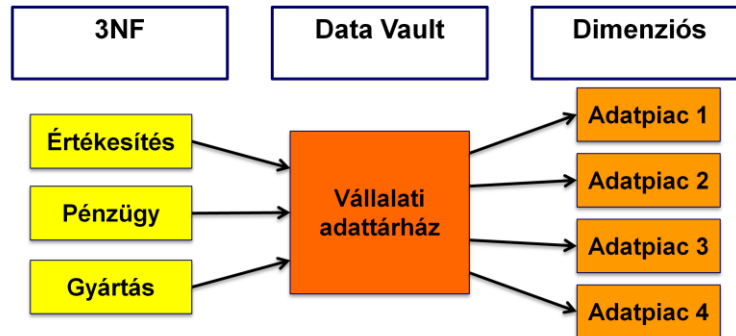
- Dátum/idő, hierarchiák

Modellezési eszközök

A Data Vault adatmodellezés

Adatmodellezési módszer, mint a 3NF vagy dimenziós

- 3. normál forma (3NF) – operatív rendszerek
- Dimenziós – data mart / OLAP / elemzések
- Data Vault – vállalati adattárház



2016.06.14

Adatmodellezés workshop

83

Mi a Data Vault?

Adatmodell

- Részletes adatok
- Történetiség kezelés (mindig, mindenhol)
- Egy vagy több üzleti területet támogat
- Normalizált táblák egyértelműen összekötött halmaza
- „The Data Vault is the optimal choice for modeling the EDW in the DW 2.0 framework.” – Bill Inmon

Módszertan

- A DW/BI legjobb gyakorlat (best-practice) alapján
- Ismételhető, konzisztens, optimalizált
- Projekttervezés, kockázatkezelés, verziókezelés, stb.

Adatmodellezés

Data Vault vs. többiek

Dimenziós	Történet kezelt 3NF	Data Vault
Dimenzió táblák (SCD 1,2,3) (Konform dimenziók / DW bus)	Normalizált (Timestamp) (Start/end dátum)	Hub (kulcsok) és Satellite (történetiség)
Tény táblák		Link és Link-Sat
Bridge, helper táblák		Stand-alone (Reference)
Terület/témakör specifikus modell	Vállalati / üzleti modell	Üzleti témakör alapú, „hibrid” modell
Érthető, egyszerűen lekérdezhető	Teljes körű	Rugalmas, egyszerűen bővíthető
Bonyolult töltés (ETL)	Bonyolult lekérdezés (sok join)	Egyszerű töltés & Business Vault v. DM
Ralph Kimball	Bill Inmon	Dan Linstedt

A Data Vault modell elemei

Hub

- Az egyedi **üzleti** kulcsok táblája
- Az adott üzleti elem (fogalom) egyedi üzleti kulcsait tartalmazza

Link

- Kapcsolatok a Hub-ok között (mindig N:M)

Satellite

- Leíró jellegű adatok
- Mind a hub-ok, mind a link-ek leírására

Egyéb

- Referencia táblák, kódtáblák

A Data Vault elemei - HUB

A Hub az adott entitáshoz tartozó egyedi üzleti kulcsokat (természetes kulcsokat) tartalmazza

- „Soft” integrációs pont a különböző forrásokból származó szemantikailag azonos elemekre.

Struktúrája:

- DV HASH - mesterséges kulcs (MD5/SHA/stb. hash a BK-n)
- Forrás rendszeri kulcs (üzleti kulcs – Business Key)
Lehet összetett kulcs is
- Betöltés timestamp (Load Date TS - LOAD_DTS)
amikor először „láttuk” ezt a kulcsot a DW-ben (forrásadatban)
- Rekord forrás jelölő (RECORD_SRC)
nyomon követhetőséghez és spec. elemzésekhez
- Utolsó timestamp (Last Seen Date TS) - Opcionális
amikor utoljára „láttuk” ezt a kulcsot a DW-ben (forrásadatban)

A Data Vault elemei - HUB

A Hub az adott entitáshoz tartozó egyedi üzleti kulcsokat (természetes kulcsokat) tartalmazza.

HUB Ügyfél		
Ügyfél Hash	char(32)	<pk>
Ügyfélszám	char(10)	<ak>
Load DTS	datetime	
Record Source	varchar(20)	
DV ID		<pk>
Business Key		<ak>

Ügyfél HUB			
Ugyfel_Hash	Ugyfelszam	Load_DTS	Record_Src
7f2208d6f5c634e5bb5591d277ce8d1f	CR1A2560	2011.01.01	CRM
28666b283ce0a0bfa26df2339674a769	CR1A2561	2011.01.01	CRM
3c9ebb086069ce63856d9df46e0b4ab5	CC5A2562	2011.03.10	SALES
9e2208d6f5c634e5bb5591d277cea67c	CC3A2563	2013.12.22	SALES
86aa6b283ce0a0bfa26df2339674bc7f	CR1A2564	2014.01.08	SALES
5d0dbb086069ce63856d9df46e0b3ecd	CR2A2565	2015.02.03	SALES
4a3208d6f5c634e5bb5591d277ce97ea	CC5A2566	2015.03.11	FINANCE
13ef6b283ce0a0bfa26df23396745bde	CC3A2567	2015.05.06	SALES
01a2bb086069ce63856d9df46e0b9875	CR2A2568	2015.06.22	SALES
a3d208d6f5c634e5bb5591d277ce19bc	CR1A2569	2015.08.23	CRM

A Data Vault elemei - HUB

Jellemző HUB területek:

- Dolgozó
- Termék
- Ügyfél
- Szállító
- Érdeklődő
- Címek
- Telefonszámok
- Vevői rendelés
- Szállítói rendelés

A Data Vault elemei - LINK

A Link az üzleti kulcsok közötti kapcsolatokat jelenti

- A Link részletezettsége (grain) a szülő kulcsai által definiált
- Minden Link egy adott Unit Of Work-öt (UOW) testesít meg
- A Link célja, hogy az adatelemek közötti legrészletesebb kapcsolatot tárolja (múlt, jelen és jövőbeli kapcsolatot is)

Szabályai:

- Egy vagy több Hub vagy Link kulcsot (DV HASH) tartalmaz (FK a HUB/LINK felé) – minimum egy FK kötelező
- Lehet saját satellite táblája/táblái
- A Load date-től a Last seen date-ig érvényes, vagy ahogy a link-satellite meghatározza
- Az adatpiacban Tény táblák lesznek belőle (jellemzően)

A Data Vault elemei - LINK

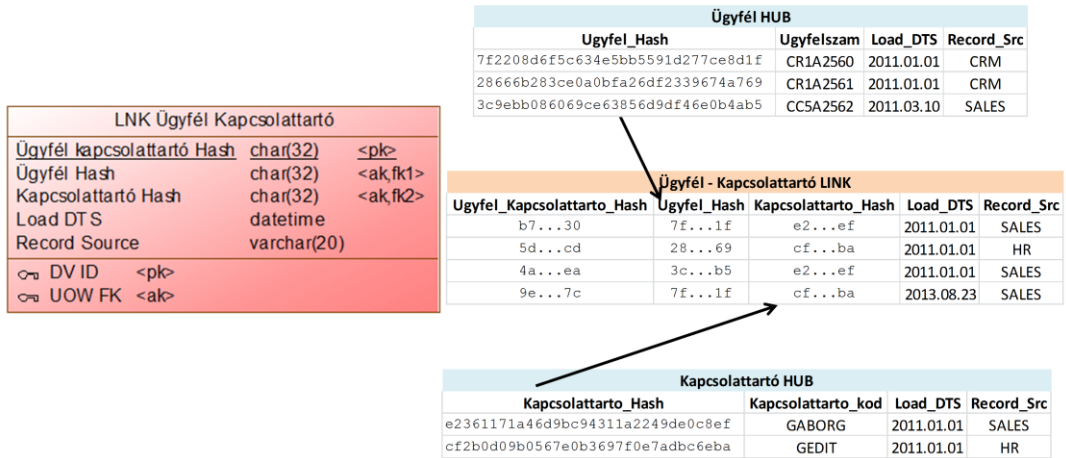
A Link több-a-többes (N:M) kapcsolatot tesz lehetővé!

Megengedett struktúrája:

- DV HASH – mesterséges kulcs (HASH)
- Hub/Link kulcs(ok) – idegen kulcs (FK: HASH), a kapcsolódó elemeket határozza meg (UOW és unique idx)
- Betöltés timestamp (Load Date TS) amikor először „láttuk” ezt a kapcsolatot
- Rekord forrás jelölő (RECORD_SRC) nyomon követhetőséghez és spec. elemzésekhez
- Utolsó timestamp (Last Seen Date TS) - Opcionális amikor utoljára „láttuk” ezt a kulcsot a DW-ben (forrásadatban)

A Data Vault elemei - LINK

A Link az üzleti kulcsok közötti kapcsolatokat jelenti



A Data Vault elemei - LINK

A Link-ek definiálásánál fontos:

- Link típus (sima, same-as, hierarchical)
- Link részletezettsége (Unit-of-Work – UOW)
- Link meghatározó kulcsa (driving key)

A Data Vault elemei - SATELLITE

A Satellite írja le a hub vagy link jellemzőit
Történetiséget kezel, hasonlóan a Type-2 dimenzióhoz

Megengedett struktúrája:

- Hub/Link kulcs (FK a Hub/Link PK-ra; HASH)
- Betöltés timestamp (Load Date TS) – érvényesség kezdete
- Utolsó timestamp (Load End Date TS) – érvényesség vége
- Rekord forrás jelölő (RECORD_SRC)
nyomon követhetőséghez és spec. elemzésekhez
- Hash diff (Opcionális; HASH alapú CDC-hez használható)
- Leíró/tartalmi jellegű attribútumok

A Data Vault elemei - SATELLITE

Egy Hub v. Link Satellite-ja több Sat-ra is szétosztható:

- Oszlopok adattípusa alapján (vertical partitioning)
- Változási gyakoriság alapján
- Adatforrásai alapján (EZ A BEST PRACTICE!)

SAT-ok felhasználási módjai:

- Effectivity Sat (mitől meddig érvényes a hivatkozott Link)
- Record Tracking Sat (mikor, mit, honnan kaptunk)
- Status Tracking (adott Hub/Link BK státuszát követi: I/U/D)
- Computed Sat (ha „véletlenül” a DV-ben kellene kalkulálni, pl. hogy ugyanazt kapja minden DM és csak egyszer futtassuk)

A Data Vault elemei - SATELLITE

A Satellite írja le a hub vagy link jellemzőit és történetét

Ügyfél HUB			
Ugyfel_Hash	Ugyfelszam	Load_DTS	Record_Src
7f2208d6f5c634e5bb5591d277ce8d1f	CR1A2560	2011.01.01	CRM
28666b283ce0a0bfa26df2339674a769	CR1A2561	2011.01.01	CRM
3c9ebb086069ce63856d9df46e0b4ab5	CC5A2562	2011.03.10	SALES

SAT Ügyfélnév		
Ügyfél Hash	char(32)	<pkfk>
Load DTS	datetime	<pk>
Load End DTS	datetime	
Record Source	varchar(20)	
Hashdiff	char(32)	
Családnév	varchar(50)	
Keresztnév	varchar(50)	
☞ Sat PK	<pk>	

Ügyfélnév SAT							
Ugyfel_Hash	Load_DTS	Load_End_DTS	Record_Src	Hashdiff	Csaladnev	Keresztnev	
7f2208d6f5c634e5bb5591d277ce8d1f	2011.01.01	2014.06.30	CRM	7f22...8d1f	Fülig	Jimmy	
28666b283ce0a0bfa26df2339674a769	2011.01.01	2011.01.02	CRM	2866...a769	Pizskos	Alfréd	
3c9ebb086069ce63856d9df46e0b4ab5	2011.03.10	null	SALES	3c9e...4ab5	Szőke	Ciklon	
28666b283ce0a0bfa26df2339674a769	2011.01.02	null	CRM	9034...bd57	Pizskos	Fred	
7f2208d6f5c634e5bb5591d277ce8d1f	2014.06.30	2014.07.31	CRM	a2ef...85b0	don Fülig	Jimmy	
7f2208d6f5c634e5bb5591d277ce8d1f	2014.07.31	null	CRM	7f22...8d1f	Fülig	Jimmy	

2016.06.14

Adatmodellezés workshop

96

A Data Vault elemei - SATELLITE

Jellemző SAT adatok:

- Státuszok
- Mennyiségek
- Dátumok
- Szöveges jellemzők

Egyéb tábla típusok

Referencia táblák

Referencia táblák

- Nem változó kódok és jellemzőik tárolására
- Pl. országkódok (ISO 3166), igen/nem, férfi/nő, stb.
- A DV része, a HUB-SAT struktúrákhoz hasonlóan épül fel
- Közvetlenül kapcsolódik a SAT-okhoz, nem építünk LINK-et

Egyéb tábla típusok PIT és Bridge táblák

PIT (Point-In-Time) táblák

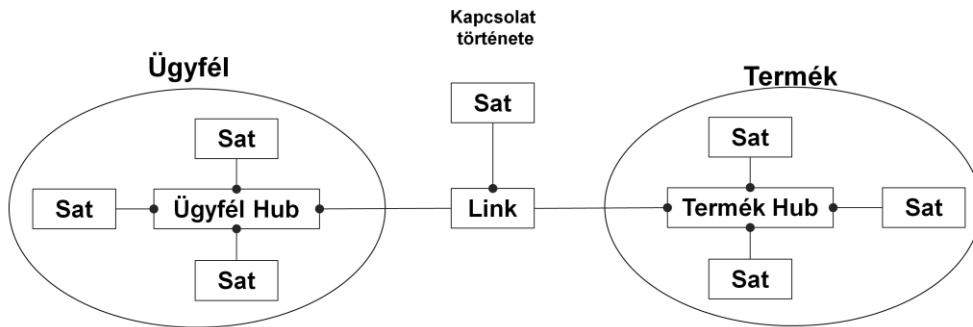
- A DV lekérdezést és DM töltést segítő táblák
- pl. HUB-LINK-HUB-SAT táblák egy adott időpillanatban érvényes ID-jait köti össze
- Limitált időtávig őrizzük csak meg az adatait

Bridge táblák

- A DV lekérdezést és DM töltést segítő táblák
- A LINK-LINK kapcsolatokat egyszerűsíthetjük ki velük
- A HUB-LINK-LINK-HUB ID-kat köti össze

A Data Vault alap elemei

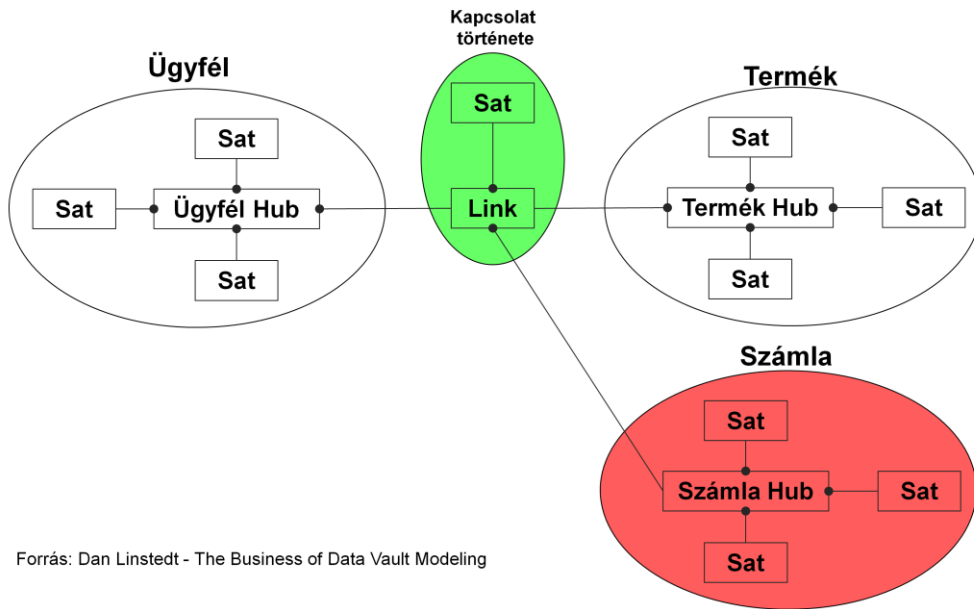
Rugalmas adatmodell – kiinduló állapot



Forrás: Dan Linstedt - The Business of Data Vault Modeling

A Data Vault alap elemei

Rugalmas adatmodell – bővített állapot



Forrás: Dan Linstedt - The Business of Data Vault Modeling

Data Vault alapú adattárház kialakítása/áttérés #1

DV modell kialakítása

- Üzleti vagy természetes „kulcsok” - HUB
 - Megkeresése, elemzése
 - A redundanciák kiszűrése, konszolidálás
- Kapcsolatok - LINK
 - Az ezek közötti kapcsolatok megkeresése, elemzése
- Leíró jellegű adatok – SAT
 - A „maradék” adatok elemzése,
 - Konszolidálás, csoportosítás (forrás, változási sebesség, stb.)
 - Hub vagy Link leíró adatok?

Data Vault alapú adattárház kialakítása/áttérés #2

EDW kialakítása

- Alap ETL eljárások, „hard business rule”-ok, stb.
- Raw DV, Quality DV, Error Mart, adatelemzés, stb.

Data Martok (vagy Business Data Vault) kialakítása

- Adatstruktúrák, eljárások („soft business rule”-ok)
- Virtualizált vagy fizikai
- Riportok, stb.

Data Vault kialakítási alapelvek

Célok

- Gyorsaság, rugalmasság, agilitás
- Megbízhatóság és compliancy („All the data all the time”)
- Jó teljesítmény, méretezhetőség

Egyszerűsítés (maximalizáljuk az el nem végzett munkát 😊)

- KISS (Keep It Simple & Stupid)
- A lehető legkevesebb kézi kódolás (használjunk kódgenerátorokat)
 - Adatmodellező eszközök, ETL generátorok (pl. PowerDesigner, ODI); van open source eszköz is (Quipu)
 - Nincsenek szintaktikus hibák! 😊
 - Egyszerűbb módosítani (csak újra kell generálni)

Data Vault kialakítási alapelvek

Agilis módszerek támogatása

- Kis lépésekben is felépíthetjük a vállalati DW-t
- Rövid, gyors ciklusok (modellezés, táblák kialakítása, ETL kialakítása, feltöltés)
- FONTOS! Priorizáljuk az adatigényeket.

Üzleti szabályok (soft) csak a DV utáni rétegben

- Hub - a hiányzó üzleti kulcsokat automatikusan felveszi
- Link - mindig N:M kapcsolatok (nem gond, ha változik a szabály)

Párhuzamos feldolgozások

- Hub-ok egyszerre -> Link-ek egyszerre -> Sat-ok egyszerre

A Data Vault adatmodell kialakítása 5 lépésben

1. Teljeskörű forrás/üzleti modell(ek) kialakítás ("Technical-Functional" modell)
2. A kulcsok és oszlopok analizálása és klasszifikálása
3. Az entitások és kapcsolatok klasszifikálása
4. A 2-3 lépések eredményének integrálása
5. Data Vault modellre átalakítás

A Data Vault adatmodell kialakítása 5 lépésben

1. Teljeskörű forrás/üzleti modell(ek) kialakítás
(“Technical-Functional” modell)
 - PK (Primary Key), FK (Foreign Key), SK (Surrogate Key) és AK (Alternate Key)
2. A kulcsok és oszlopok analízálása és klasszifikálása
 - Melyik PK lesz BK (Business Key)
 - Melyik PK igazából SK (Surrogate Key)
 - Melyik kulcs az entitás/attribútumok meghatározó kulcsa („driving key”)?
 - A linkekhez a Unit-of-work (UOW) megkeresése

A Data Vault adatmodell kialakítása 5 lépésben

3. Az entitások és kapcsolatok klasszifikálása

- Melyik forrás entitásból lesz HUB?
- Melyik forrás entitásból lesz Link?
- Melyik forrás entitásnak vannak nem kulcs attribútumai -> Satellite?

4. A 2-3 lépések eredményének integrálása

5. Data Vault modellre átalakítás

- BK-k -> HUB-ok
- UOW -> LINK-ek
- Attribútumok -> SAT-ok
- "Denormalizálni" a LINK-LINK kapcsolatokat

A Data Vault modellezés előnyei

Teljes körű

- „All the data all the time”
- Auditálható, „compliance”

Rugalmas

- Gyorsan kialakítható, módosítható
- Rugalmas, egyszerű struktúrák
- Egyszerűen tölthető (generálható ETL eljárások)
- Segíti az agilis IT-t

Teljesítmény

- Gyors, parallel töltések
- Könnyen bővíthető, jól skálázható

A Data Vault modellnek is vannak „hátrányai”, de...

Sok join kell a lekérdezésekhez...

- DE a modern parallel adatbázisok és skálázható szerverek jól kezelik ezt a problémát

Nem támogatja az ad-hoc felhasználói lekérdezéseket...

- a DV „csak” az EDW központi része;
- DE az elemzés típusú felhasználást (megfelelő képzés mellett) támogatja/megengedi

Nem ad megoldást a Data Mart-ok modellezésére...

- DE az architektúrában dedikált DM réteg van, ami lehet dimenziós, OLAP, vagy „Business Vault” struktúrájú is
- a DM réteg lehet fizikai vagy logikai (pl. view-k)

Tematika

Mi az adatmodellezés?

- Mi ez? Miért kell? Alapelemek

Hogyan modellezzünk?

- Koncepcionális – logikai – fizikai adatmodell
- 5 (7) W

Adatmodellezési „módszerek”

- „Klasszikus” modellezés (egyed-kapcsolat avagy „relációs”)
- „Adattárház” adatmodellezés (Dimenziós, Data Vault)

Érdekes adatmodellezési esetek

- Dátum/idő, hierarchiák

Egyéb adatmodellezési módszerek

Adatmodellezési eszközök

Feladat #3 - Modellezés

Naptár témakör modellezése

- Tervezz adatmodellt naptár adatok kezelésére
 - Entitás(ok) v. tábla(k)
 - Kulcs jellemzők
 - Hierarchiák
 - Kapcsolatok (ha van)
 - Milyen előfeltételezésekkel éltél a modellezéskor?
 - Milyen modellezési módszert választottál és miért?
- Beszéljük meg közösen!

Speciális adatmodellezési kérdések

Törzsadatok

Dátum & idő kezelése

- Normalizált és denormalizált
- Alternatív naptárak
- Speciális idők (pl. időzónák, műszakok)

Hierarchiák kezelése

- Fix mélységű vs. Dinamikus
- Fa vs. háló

Törzsadatok

Szabványokban definiált adatok (pl. ISO/MSZ)

- Célszerű ezeket használni, pl.:
- ISO 639 – nyelv kódok és nevek
- ISO 3166 – ország kódok és nevek
- ISO 4217 – pénznem kódok és nevek
- ISO 8601 – dátum és idő formátumok
- ITU E.123 és E.164 – telefon számformátum és ország hívókódok
- TEÁOR kódok, SZJ kódok
- Közterület jegyzék, irányítószám jegyzék

Dátum kezelése

Az egyik legfontosabb dimenzió(k)

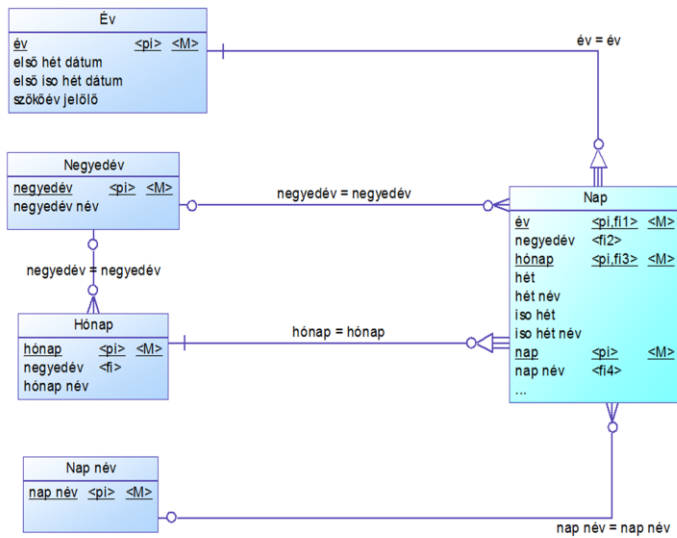
Dátum

- Normalizált modell
 - Naptári nap -> Hónap -> Negyedév -> Év
 - Naptári nap -> Hét -> Év (?)
 - Naptári nap -> pénzügyi naptár? egyéb naptár?
- Denormalizált modell
 - Egy táblában az attribútumok
- Általában szerepjátszó dimenzió

Dátum kezelése – standard naptár

Normalizált modell

Dimenziós modell



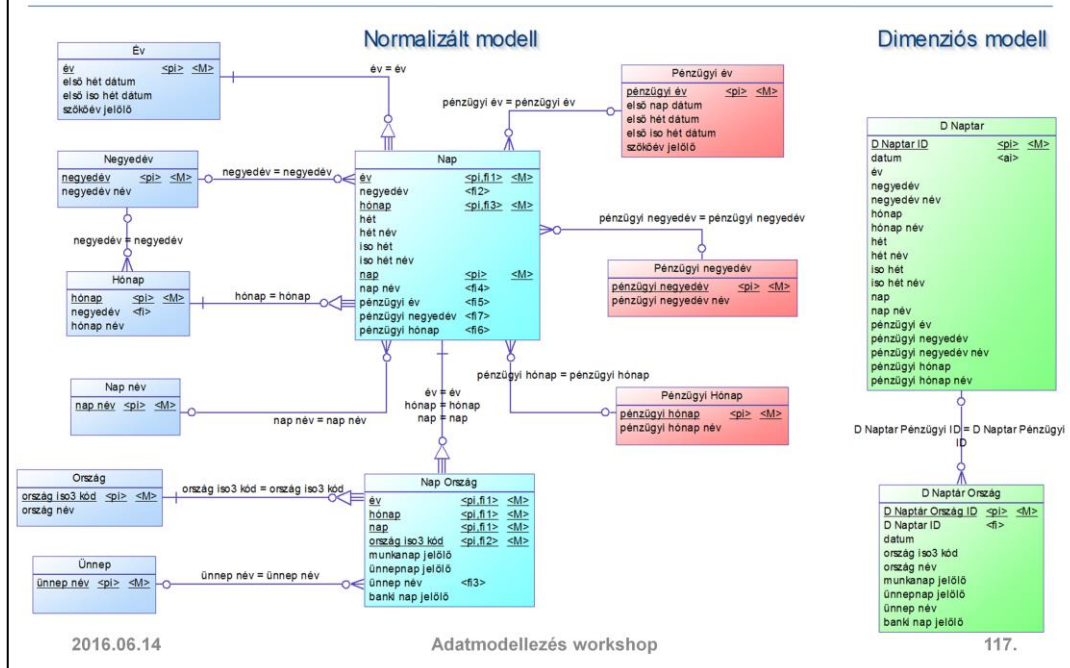
D Naptár Alap		
D Naptár ID	<pi>	<M>
datum	<a1>	
év		
negyedév		
negyedév név		
hónap		
hónap név		
hét		
hét név		
iso hét		
iso hét név		
nap		
nap név		

2016.06.14

Adatmodellezés workshop

116.

Dátum + pénzügyi + ország specifikus naptár



Ország specifikus naptárak az eltérő munka és ünnepnapokra, vagy más speciális napokra.

Egyéb kiegészítő attribútumok lehetnek:

- első és utolsó napok (évben, negyedévben, hónapban, hétben) – minden rekordhoz duplikálva a konkrét dátum, illetve flagként is
- fix dátum („világ kezdete”) óta a napok sorszáma (x időre visszamenőlegesen kell az adatokat adni)
- éven belül a munkanap sorszám
- ünnephez közeli nap jelölés (pl. karácsonyi időszak, húsvéti időszak, stb.)
- naptári egységek elnevezései (pl. nyári szünet, téli szünet)
- nem naptári napok (pl. 13. hónap)

Idő kezelése

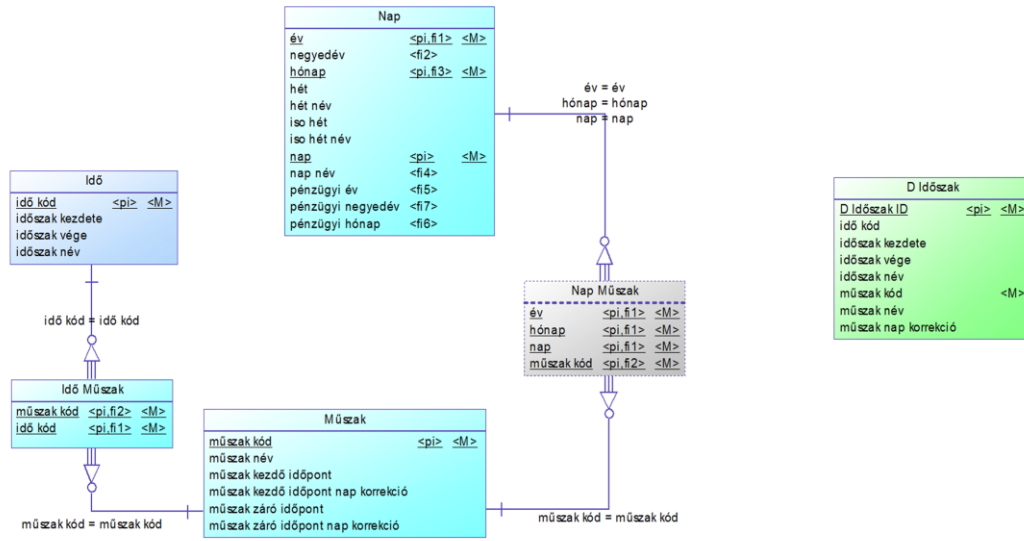
Idő

- Célszerű a dátumtól leválasztani
- Fontos a részletezettsége, felbontása:
 - Perc => $60 \cdot 24 = 1440$ sor
 - Másodperc => $60 \cdot 60 \cdot 24 = 86400$ sor
- Szerepe:
 - Időpont: általában szerepjátszó dimenzió
 - Időtartam: általában tény (mutatószám)
- Lehetnek napon átnyúló időszakok
 - pl. műszak => önálló vagy dátum & idő összevont dimenzió
- Lehetnek időzónák (pl. helyi idő + UTC idő)

Idő kezelése – idő + műszak/napszak

Normalizált modell

Dimenziós modell



2016.06.14

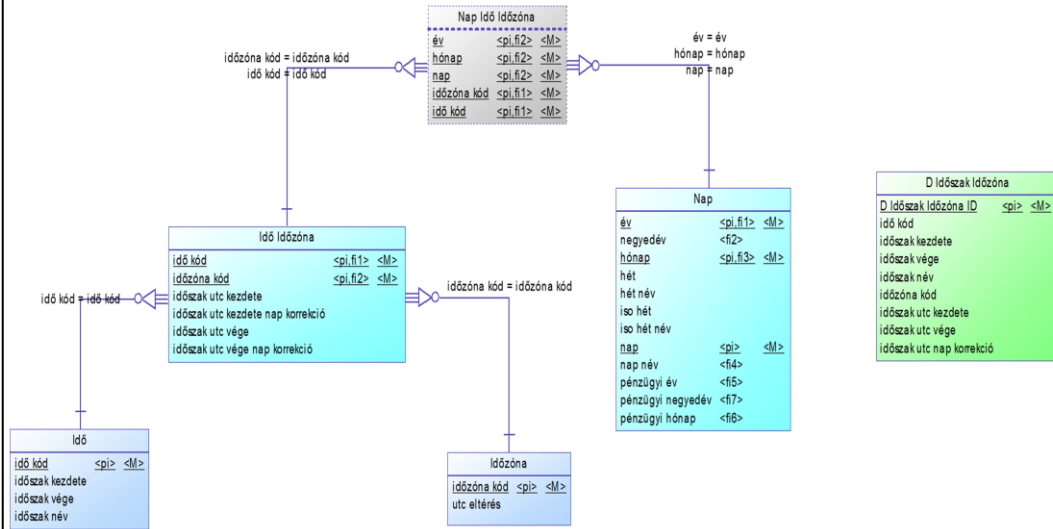
Adatmodellezés workshop

119.

Idő kezelése – idő + időzóna

Normalizált modell

Dimenziós modell



2016.06.14

Adatmodellezés workshop

120.

Tematika

Mi az adatmodellezés?

- Mi ez? Miért kell? Alapelemek

Hogyan modellezzünk?

- Koncepcionális – logikai – fizikai adatmodell
- 5 (7) W

Modellezési „módszerek”

- „Klasszikus” modellezés (egyed-kapcsolat avagy „relációs”)
- „Adattárház” adatmodellezés (Dimenziós, Data Vault)

Érdekes modellezési esetek

- Dátum/idő, hierarchiák

Egyéb modellezési módszerek

Modellezési eszközök

Anchor modelling

6NF alapokra épül:

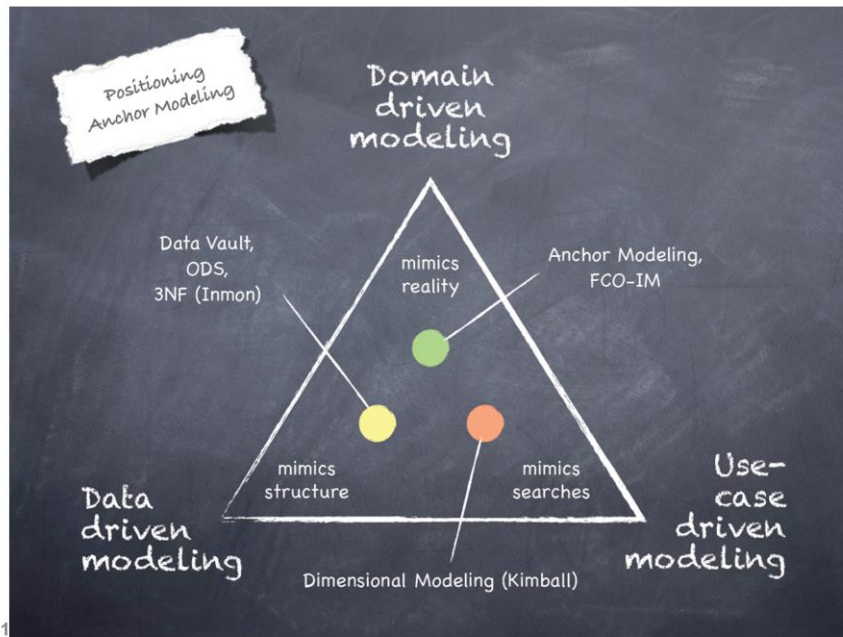
- Minden nem-kulcs attribútum önálló relációba kerül.
A 6NF tábla a kulcs plusz max. 1 oszlop.

Alapelemei:

- Anchor (horgony) – az alapvető entitások tartalma: mesterséges kulcs + metaadatok
- Attribute (attribútum/jellemző) – a leíró adatok tartalma: FK a kapcsolódó Anchor-ra, 1 leíró érték, historizációs és metaadatok
- Tie (kapcsolat) – a kapcsolatokat írja le tartalma: FK a kapcsolódó Anchorokra (n darab), historizációs és metaadatok

Historizáció: csak insert & csak FromDate

DW-s modellezési módszerek pozícionálása (forrás: Lars Rönnbäck)



Big data (NoSQL) modellezés

Kell ez?

- Schema-on-read vs. Schema-on-write
- Agilitás vs. Konzisztencia (integritás és governance)
- Dedikált adatmodell vs. A kódba „beégetett” modell

Relációs vs. NoSQL (sarkítva)

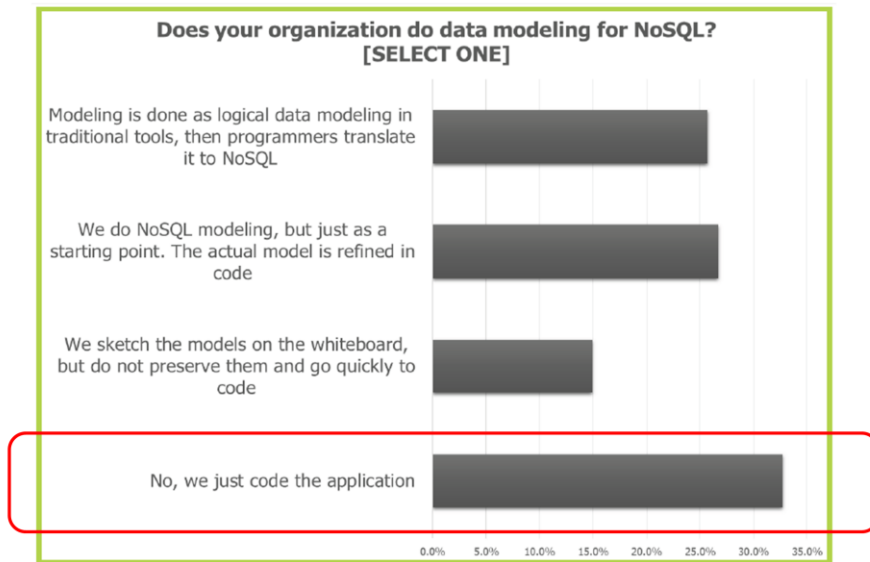
Relációs	NoSQL
SQL	Sok különböző programozási nyelv – akár többféle library-val is
Állandó adatstruktúrák	„Változó” adatstruktúrák
Előre definiált séma (schema on write) – nem rugalmas („eléggé”)	Séma nélküli, dinamikus sémák (schema-on-read) – Nagyon rugalmas
Vertikálisan skálázható (általában)	Horizontálisan skálázható, elosztott rendszerek
Az ACID tranzakciókon a hangsúly	Fizikai tárolási módtól függően, nagy teljesítmény és rendelkezésre állás az ACID tranzakciókért „cserébe”
Összetett lekérdezési lehetőségek	A fizikai tárolási módtól is függ. Eleinte inkább csak egyszerűbb lehetőségek voltak, de ez folyamatosan bővül.
Eszközök és alkalmazások széles skálája	Sok különböző lehetőség, de sok még mindig nem elég „kiforrott”
Egyszerű adatmodell; az adatok a táblák soraiban és oszlopaiban	Sokféle „adatmodell” – többnyire a „modellek” a kódban vannak
Kiforrott támogatás	Rendszer függő támogatottság
Normalizáció	Denormalizáció
Egyszerű adattípusok, nagy tranzakciósám és komplex lekérdezések	„Nem strukturált” és „részben strukturált” adatok kezelése, egyre több lehetőség az egyszerű lekérdezésekre
Strukturált adatok kezelésére tervezve	„Mindenféle” adatot jól kezel

2016.06.14

Adatmodellezés workshop

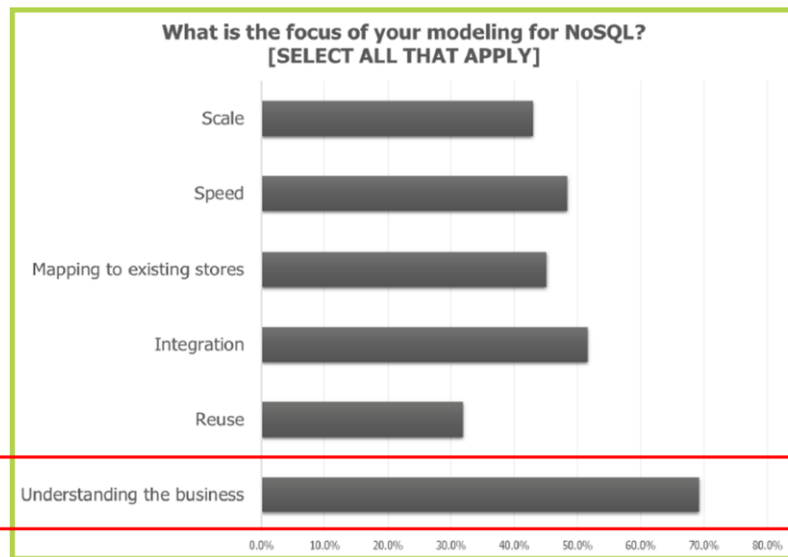
125

NoSQL modellezés?



Forrás: Insights into Modeling NoSQL A DATAVERSITY 2015 Report

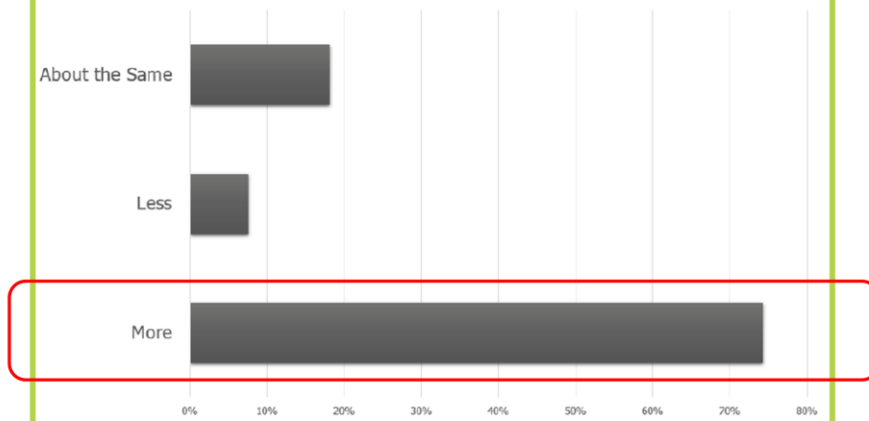
Mi a modellezés célja?



Forrás: Insights into Modeling NoSQL A DATAVERSITY 2015 Report

Több vagy kevesebb modellezés?

With Big Data an important issue for organizations in terms of massive amounts of data needing to be analyzed, do you foresee that you will have more, less, or the same number of specialist NoSQL Data Architects and Data Modelers in your organization over the next 5 years?



Forrás: Insights into Modeling NoSQL A DATAVERSITY 2015 Report

Big data (NoSQL) modellezés

NoSQL „típusok”:

NoSQL type	Horizontal Scalability	Flexibility in Data Variety	Appropriate Big Data Types	Performance	Complexity of Operation	Functionality
Key-Value stores	High	High	Yes	High	Low	Key-Value
Column stores	High	Moderate	Partially	High	Low	Column store
Document stores	Variable (high)	High	Likely	High	Low	Document store
Graph databases	Variable	High	Maybe	Variable	High	Graph theory
Relational databases	Variable	Low		Variable	Moderate	Relational algebra

Forrás: Handbook of Research on Cloud Infrastructures for Big Data Analytics

Milyen modellezés kell:

- Konceptcionális
 - „üzleti” modell,
 - kontextus
- Fizikai
 - a konkrét NoSQL „típus”-ra optimalizált

Tematika

Mi az adatmodellezés?

- Mi ez? Miért kell? Alapelemek

Hogyan modellezzünk?

- Koncepcionális – logikai – fizikai adatmodell
- 5 (7) W

Adatmodellezési „módszerek”

- „Klasszikus” modellezés (egyed-kapcsolat avagy „relációs”)
- „Adattárház” adatmodellezés (Dimenziós, Data Vault)

Érdekes adatmodellezési esetek

- Dátum/idő, hierarchiák, gráfok, „big data” és „változó struktúrák”

Egyéb adatmodellezési módszerek

Adatmodellezési eszközök

Adatmodellezési eszközök

Egyszerű „rajzoló” eszközök:

- Papír & ceruza (és/vagy flipchart, digitális tábla)
- Excel (PowerPivot?)
- Visio

Ingyenes:

- Oracle SQL Developer Data Modeler (SDDM)
- Open source eszközök (pl. PowerArchitect, MySQL, stb.)

Webes/cloud:

- Vertabelo, stb.

Professzionális:

- SAP PowerDesigner, CA Erwin, SystemArchitect, ER/Studio

Kérdések & válaszok



gollnhofer.gabor@datagovernance.hu

Olvasnivalók

Könyvek:

- [Az adattárház-készítés technológiája](#)
- [Building a Scalable Data Warehouse with Data Vault 2.0](#)
- [Data Model Patterns: Conventions of Thought](#)
- [The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling](#)
- [A Check List for Doing Data Model Design Reviews](#)
- [Data Modeling Made Simple & Data Modeling for MongoDB](#)
- [Business unIntelligence: Insight and Innovation beyond Analytics and Big Data](#)

Blog, stb.:

- [bi.hu](#)
- [LearnDataVault.com](#)
- [Kimball Design Tips](#)
- [Highly Scalable Blog](#)
- [Kent Graziano](#)
- [Steve Hoberman](#)

Egyéb:

- [www.acm.org](#)
- [www.tdwi.org](#)

Az előadóról

Gollnhofer Gábor

- 1992 óta foglalkozom nagyvállalati adatbáziskezelőkkel
- 1996 óta dolgozom adattárház és BI jellegű projekteken
- adatmodellezés, modellezési tanácsadás, oktatás
- 3NF, dimenziós és Data Vault rendszerek tervezése
- iparági tapasztalatok: államigazgatás, bank, biztosító, FMCG, távközlés, oktatás, stb.
- technológiák: Oracle, SQL Server, Sybase ASE & IQ, DB2