



**Adatmodellezés  
Data Vault alapokon  
tutorial / workshop**

## Meta Consulting Kft.

- 2002-ben alakult, magyar tulajdonú vállalat

## Tevékenységünk

- DW & BI, metaadat management, data governance
- Rendszertervezés, fejlesztés, tanácsadás, oktatás

## Ügyfeleink

- Bank, biztosító, FMCG, távközlés, média, felsőoktatás, államigazgatás, stb.
- Közép-kelet európai régióban (HU, DE, CZ, PL, SK, RO, MNE) + UK/USA
- 60+ ügyfél, 12 ország, 130+ projekt

## A Snowflake, WhereScape és Fivetran magyar partnere

# AZ ELŐADÓRÓL...

---

## Golnhofer Gábor

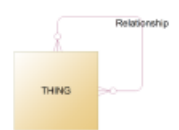
- 1992 óta nagyvállalati adatbáziskezelés
- 1996 óta adattárház és BI jellegű projektek
- adatmodellezés, modellezési tanácsadás, oktatás
- 3NF, dimenziós és Data Vault rendszerek tervezése
- iparágiak: államigazgatás, bank, biztosító, FMCG, felsőoktatás, távközlés, stb.
- technológiák: Snowflake, Oracle, Microsoft, Wherescape, stb.
- Certified Data Vault 2.0 Practitioner
- „Building a Scalable Data Warehouse with Data Vault 2.0” - technical reviewer



# Tematika

### Adatmodellezés bevezető

A "mindenség teljes modellje":



AlmásiBoris Data Vault alapok  
BPDData2023 tutorial/workshop @ 2023.06.05.


### Mi a Data Vault?

<b>Adatmodellezés</b> <ul style="list-style-type: none"><li>• Részletes adatok</li><li>• Történeltség kezelés (mindig, mindenhol)</li><li>• Egy vagy több üzleti terület támogat</li><li>• Normalizált táblák egyértelműen összekötött halmaza</li></ul>	<b>Módszertan</b> <ul style="list-style-type: none"><li>• A DW/BI legjobb gyakorlat (best-practice) alapján</li><li>• Ismételhető, konzisztens, optimalizált</li><li>• Projekttervezés, kockázatkezelés, verziókezelés, stb.</li></ul>
<b>Architektúra</b>	<b>Implementáció</b>

„The Data Vault is the optimal choice for modeling the EDW in the DW 2.0 framework.”  
Bill Inmon

AlmásiBoris Data Vault alapok  
BPDData2023 tutorial/workshop @ 2023.06.05.

### A Data Vault alap elemek Üzleti „fogalmak” (concept)



Forrás: Dan Linstedt - The Cosmos of Data Vault Modeling


AlmásiBoris Data Vault alapok  
BPDData2023 tutorial/workshop @ 2023.06.05.

### 1. gyakorlat – kiinduló üzleti modell(?)

Feladat

- 3-4 fős csoportokban dolgozzatok
- Válasszátok egy üzleti területet
- Vázoljátok fel a kulcs üzleti elemeket és a tartalmukat/definíciójukat
- Technológia: papír + ceruza + beszélgetés

Közösen megbeszéljük az eredményt!



AlmásiBoris Data Vault alapok  
BPDData2023 tutorial/workshop @ 2023.06.05.

### A Data Vault modell alapelemei

<b>HUB</b> Primary Key (NASH) Business Key Load Date & Time Record Source	<b>LINK</b> Primary Key (HASH) HUB 1 Key HUB 2 Key Load Date & Time Record Source	<b>SAT</b> Primary Key (HASH) HUB/Link Key Load Date & Time Record Source
---	--	---

Hub

- Az üzleti kulcsok (Business Key) egyedi listája

Link

- A Hub-ok közötti kapcsolatok (mindig N:M) egyedi listája

Satellite

- A leíró adatok, teljes történeltség kezelésével
- A Sat-nak egy és csak egy szülő táblája van (Hub vagy Link)

Egyéb

- Kód, Referencia, PIT, Bridge táblák


AlmásiBoris Data Vault alapok  
BPDData2023 tutorial/workshop @ 2023.06.05.

### 2. gyakorlat – kiinduló Data Vault modell

Feladat

- 3-4 fős csoportokban dolgozzatok
- Az előző feladatban elemzett üzleti terület „elemekből” kiindulva
- Vázoljátok fel a kulcs üzleti elemeket (HUB) és a kapcsolataikat (LINK)
- Technológia: papír + ceruza + beszélgetés

Közösen megbeszéljük az eredményt!



AlmásiBoris Data Vault alapok  
BPDData2023 tutorial/workshop @ 2023.06.05.

### Data Vault modell elemek – további részletek

Speciális esetek

- LINK
- SAT

Egyéb elemek

- Referencia és kód táblák
- PIT és Bridge táblák


AlmásiBoris Data Vault alapok  
BPDData2023 tutorial/workshop @ 2023.06.05.

### 3. gyakorlat – Data Vault modell kiegészítése

Feladat

- Az előző gyakorlatban elkezdett modell folytatása
- 3-4 fős csoportokban dolgozzatok
- Vázoljátok fel a leíró adatokat, melyik milyen kulcs üzleti „concept”-hez vagy kapcsolathoz tartozik
- Vannak-e Referencia/Kód táblák?
- Szükséges-e PIT v. BRIDGE táblákat létrehozni?
- Technológia: papír + ceruza + beszélgetés

Közösen megbeszéljük az eredményt!

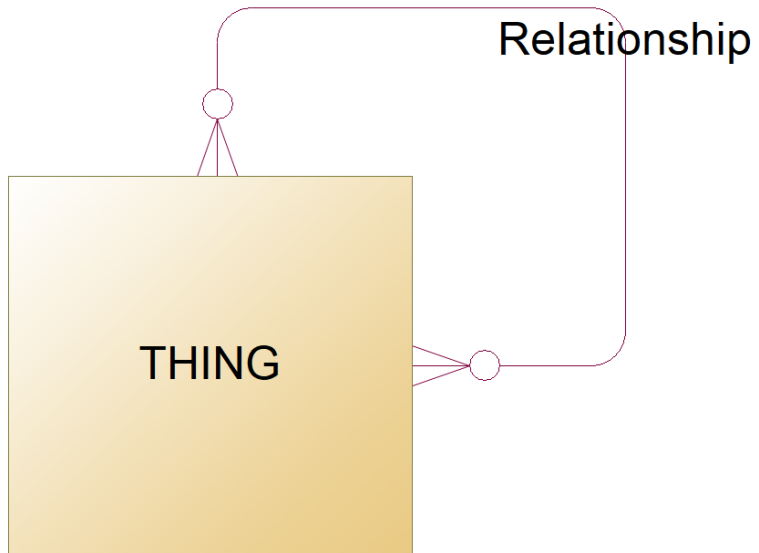


AlmásiBoris Data Vault alapok  
BPDData2023 tutorial/workshop @ 2023.06.05.

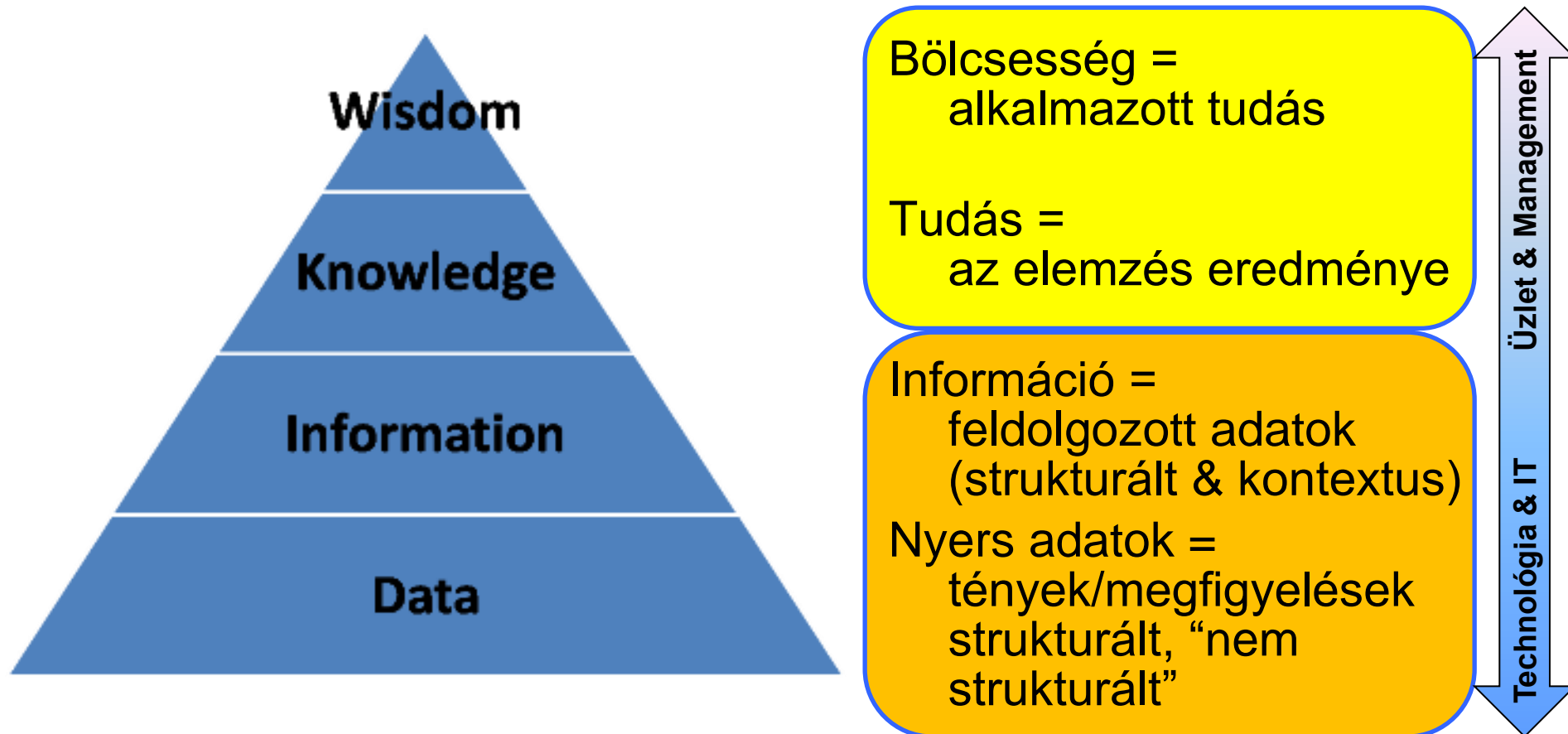
# Adatmodellezés bevezető

---

A “mindenség teljes modellje”:



# A „tudás piramis” (DIKW)



# Az EDW-vel szembeni elvárások

---

A vállalati adatvagyon integrált tárháza

Stabil, megbízható működés

Jó teljesítményű, méretezhető, bővíthető és költség hatékony

Rugalmas, agilis

- Változó környezet, adatforrások, felhasználások

Auditálható, a szabályoknak megfelelő (külső/belső)

- Adatok, folyamatok, lekérdezések/riportok
- Felhasználói módosítások (pl. csoportosítások)

# Miért kell az adatmodellel foglalkozni?

Az adatmodell a kialakítandó rendszer alapvető „térképe”

Itt kell definiálni, hogy:

- milyen területekkel, adatokkal foglalkozik
- ezek között milyen összefüggések vannak
- 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 az adatmodellel foglalkozni?

Az adatmodell a kialakítandó rendszer alapvető „térképe”  
Itt kell definiálni, hogy:



## Larry Burns: Data Model Storytelling

„... data models help a group of people "think through" the data aspects of a certain business problem or business opportunity, allowing them to surface and resolve disagreements about data meanings, explore data opportunities, design and create data resources in support of business initiatives, and achieve consensus about data meanings.”

A „... a data model tells a story, and the story is about how a group of people come together to use data to solve a business problem (or take advantage of a business opportunity).  
The data model becomes a record of the journey from the conception of the problem to its solution.”

<https://technicspub.com/data-model-storytelling/>

# 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](http://en.wikipedia.org/wiki/Five_Ws)

# Mi a Data Vault?

---

## Adatmodellezés

- Részletes adatok
- Történetiség kezelés (mindig, mindenhol)
- Egy vagy több üzleti területet támogat
- Normalizált táblák egyértelműen összekötött halmaza

## Architektúra

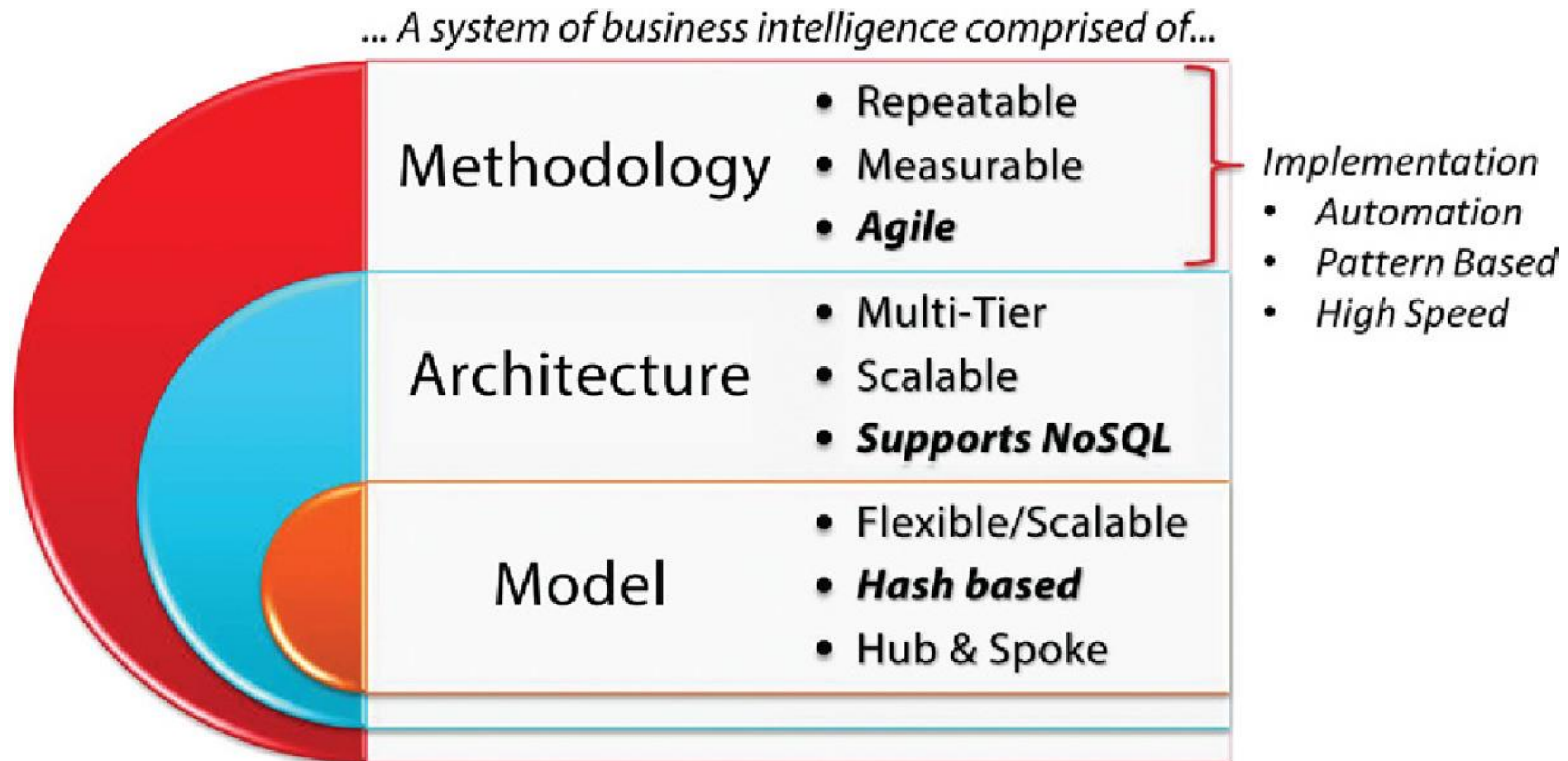
## Módszertan

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

## Implementáció

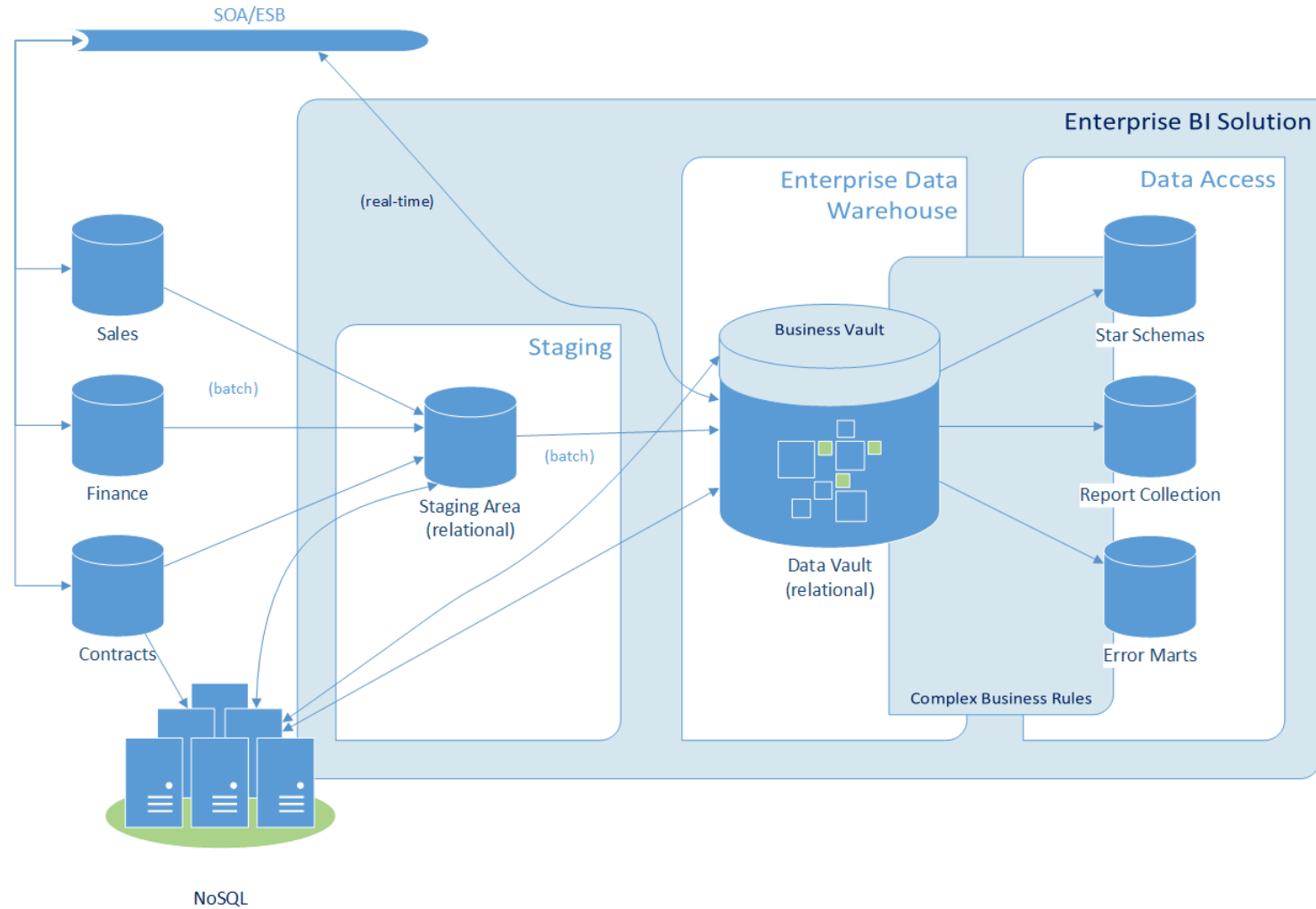
**„The Data Vault is the optimal choice for modeling the EDW in the DW 2.0 framework.”  
Bill Inmon**

# Data Vault System of business intelligence (BI)



Forrás: Data architecture: A primer for the data Scientist

# Adat architektúra

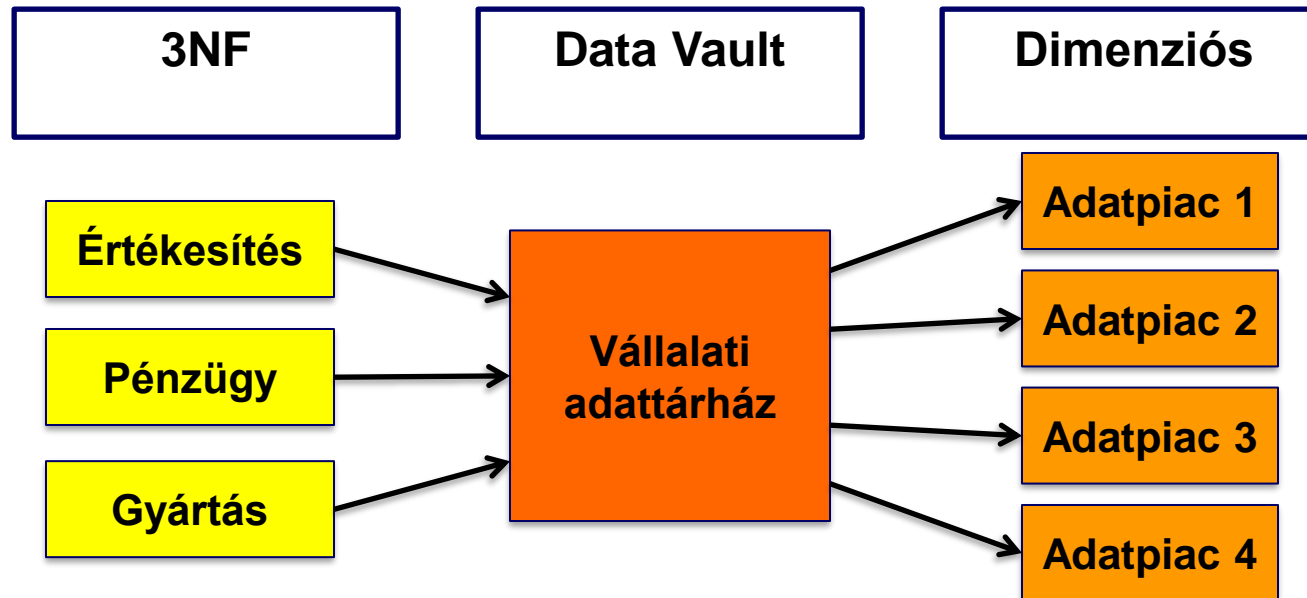


Forrás: Building a Scalable Data Warehouse with Data Vault 2.0

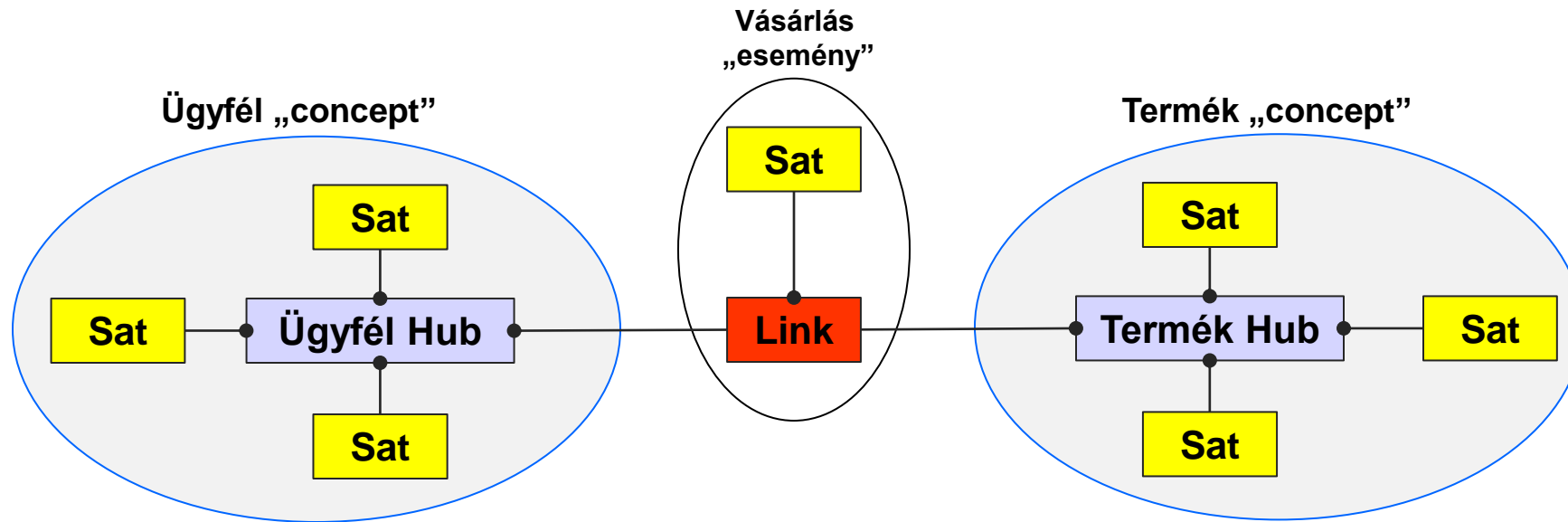
# A Data Vault adatmodellezés

Adatmodellezési módszer, mint a 3NF vagy dimenziós

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

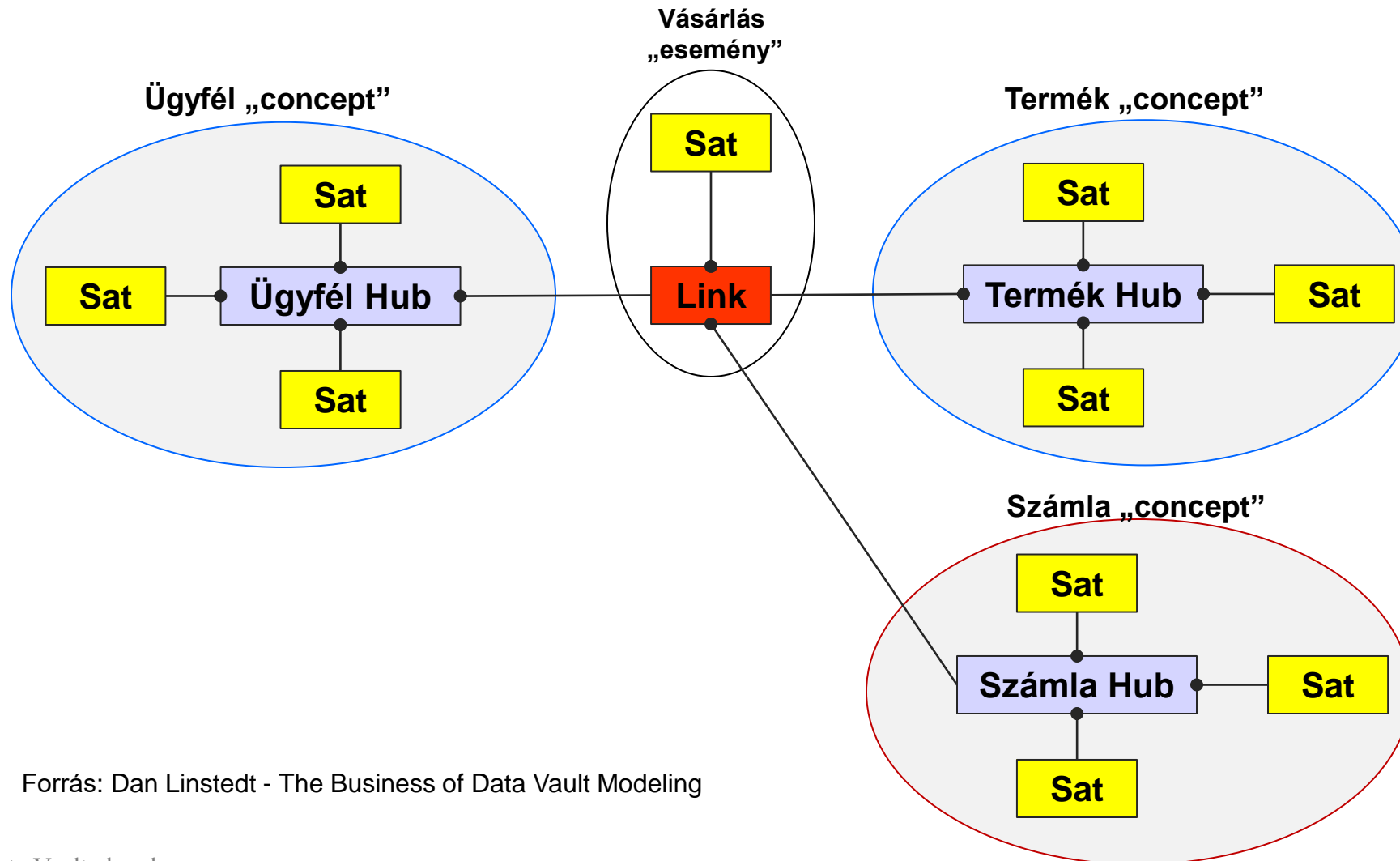


# A Data Vault alap elemek Üzleti „fogalmak” (concept)



Forrás: Dan Linstedt - The Business of Data Vault Modeling

# A Data Vault alap elemek Üzleti „fogalmak” (concept)

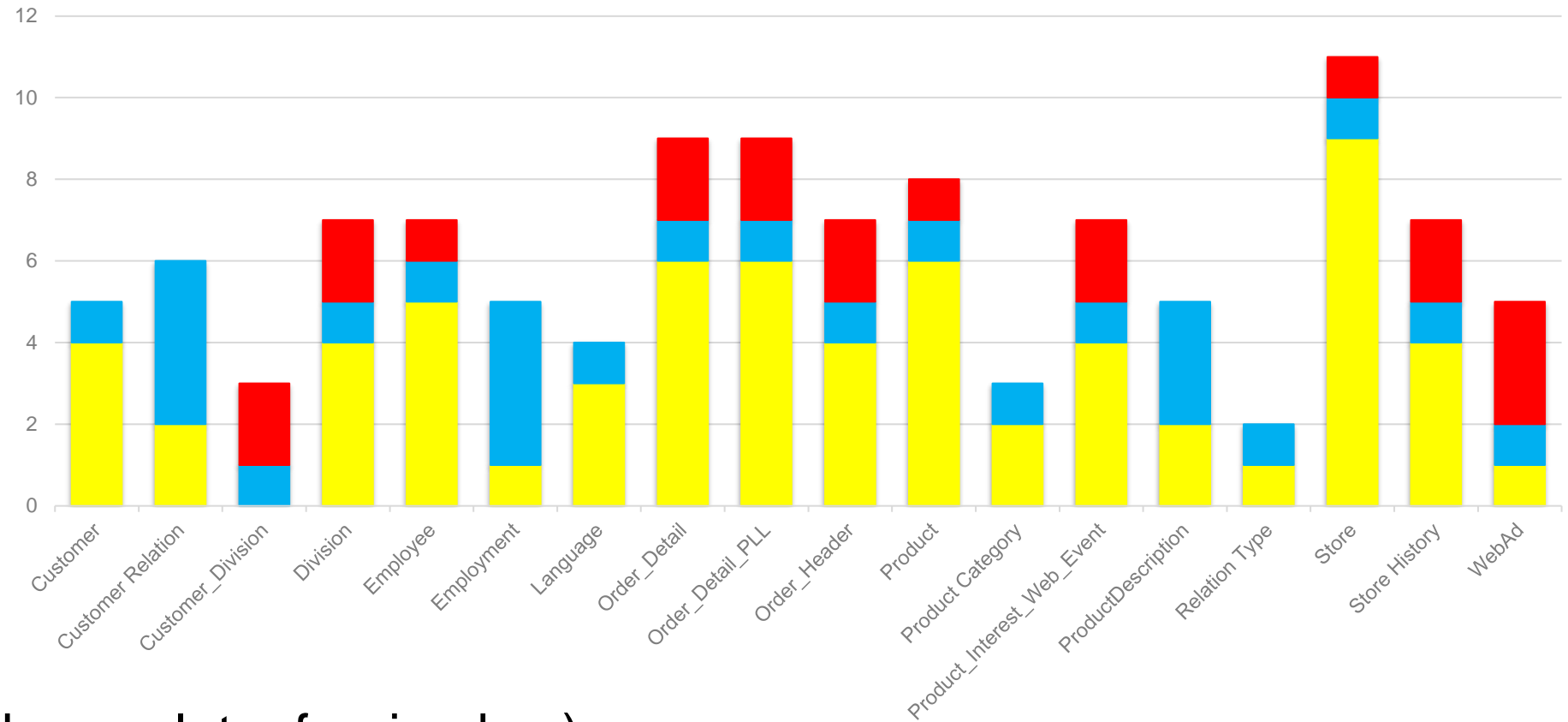


Forrás: Dan Linstedt - The Business of Data Vault Modeling



# Funkcionális dekompozíció

## Példa: oszlopok funkciók szerint



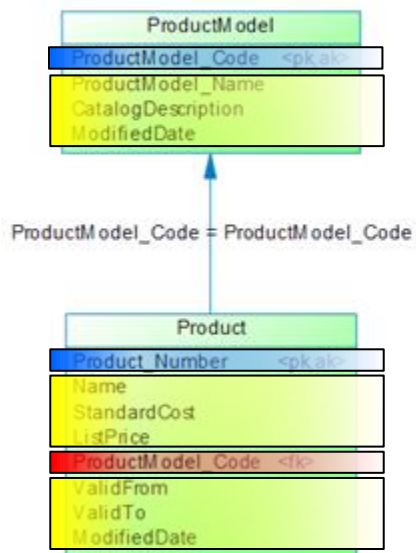
- Üzleti kulcs
- Referencia (kapcsolat – foreign key)
- Leíró adat

# Data Vault (DV) modellezés áttekintése

## Forrás elemzés, mapping

Funkcionális dekompozíció, elválasztjuk:

- az üzleti egyedeket (core business concept)
- ezek kapcsolatait
- és az ezeket leíró adatokat és történetiségüket

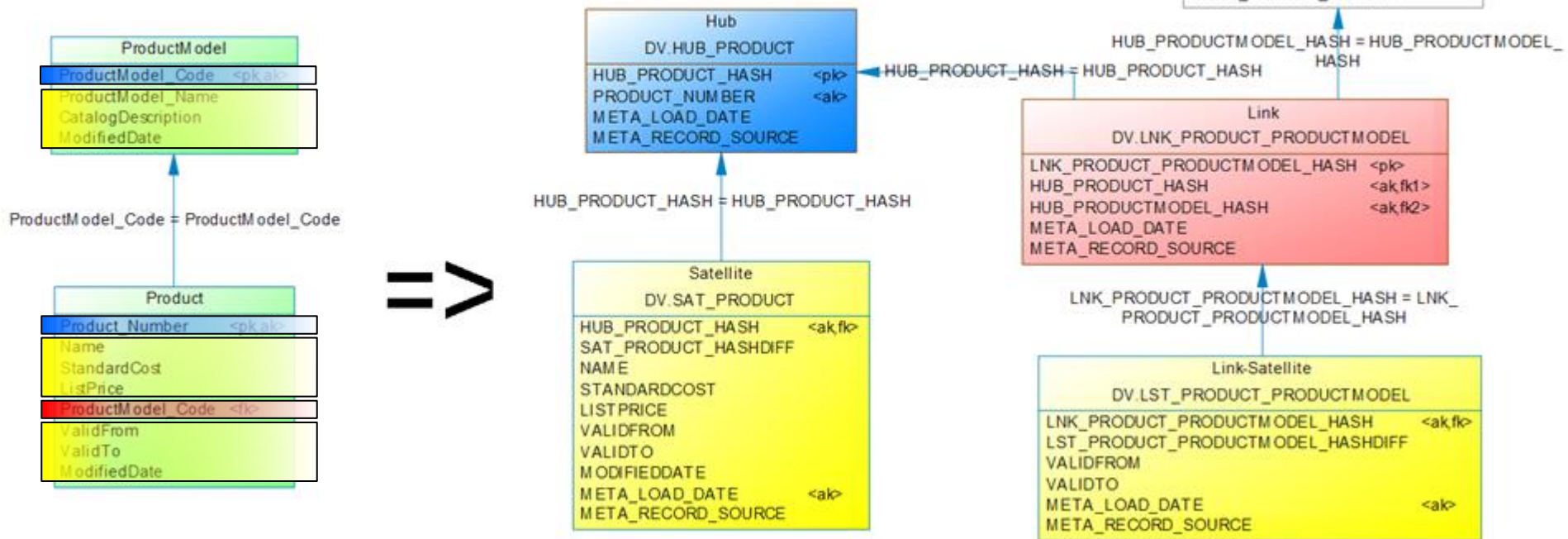


# Data Vault (DV) modellezés áttekintése

## Forrás elemzés, mapping

Funkcionális dekompozíció, elválasztjuk:

- az üzleti egyedeket (core business concept)
- ezek kapcsolatait
- és az ezeket leíró adatokat és történetiségüket



# A Data Vault modell alapelemei

---

## **HUB**

**Primary Key (HASH)**  
**Business Key**  
**Load Date & Time**  
**Record Source**

## **LINK**

**Primary Key (HASH)**  
**HUB\_1 Key**  
**HUB\_n Key**  
**Load Date & Time**  
**Record Source**

## **SAT**

**Primary Key (HASH)**  
**HUB/LINK Key**  
**Load Date & Time**  
**Load End Date & Time**  
**Record Source**

## Hub

- Az üzleti kulcsok (Business Key) egyedi listája

## Link

- A Hub-ok közötti kapcsolatok (mindig N:M) egyedi listája

## Satellite

- A leíró adatok, teljes történetiség kezeléssel
- A Sat-nak egy és csak egy szülő táblája van (Hub vagy Link)

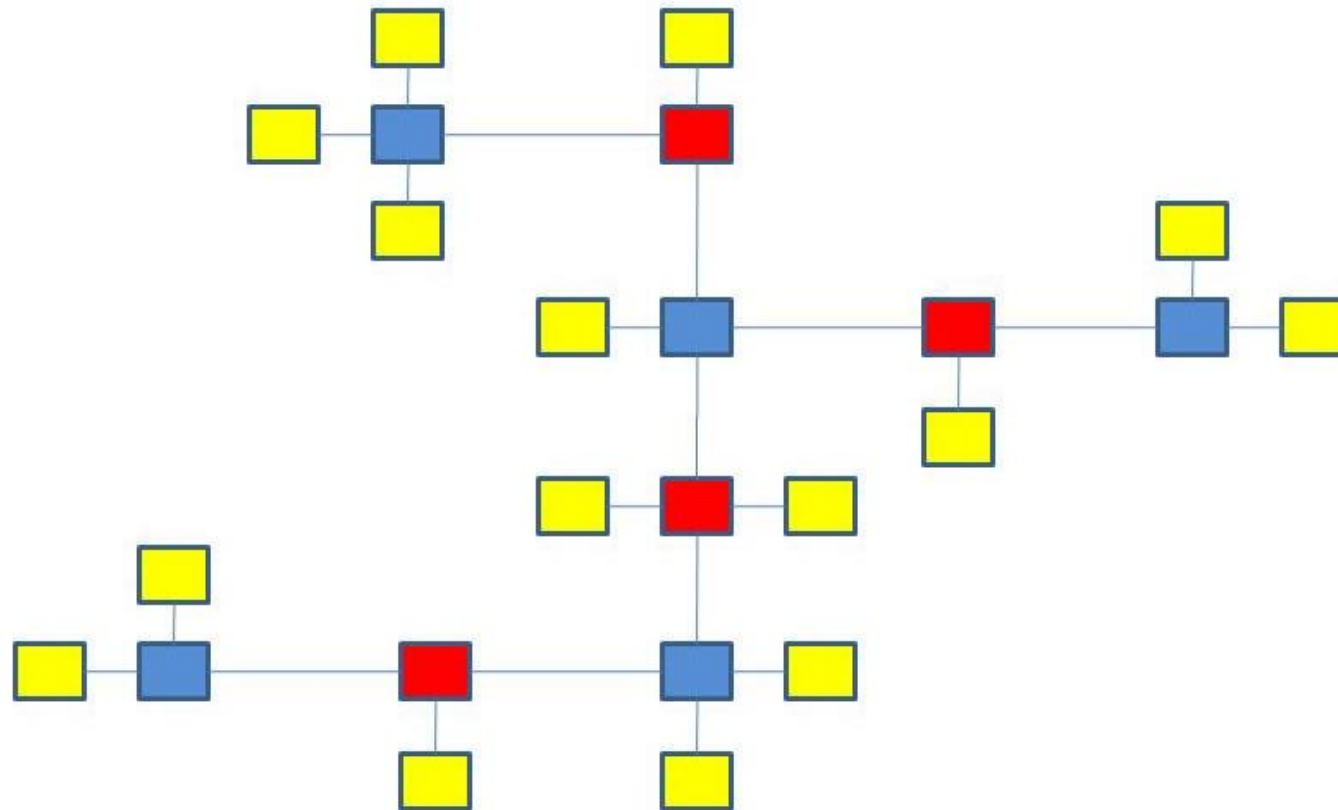
## Egyéb

- Kód, Referencia, PIT, Bridge táblák

# Adatmodellezés

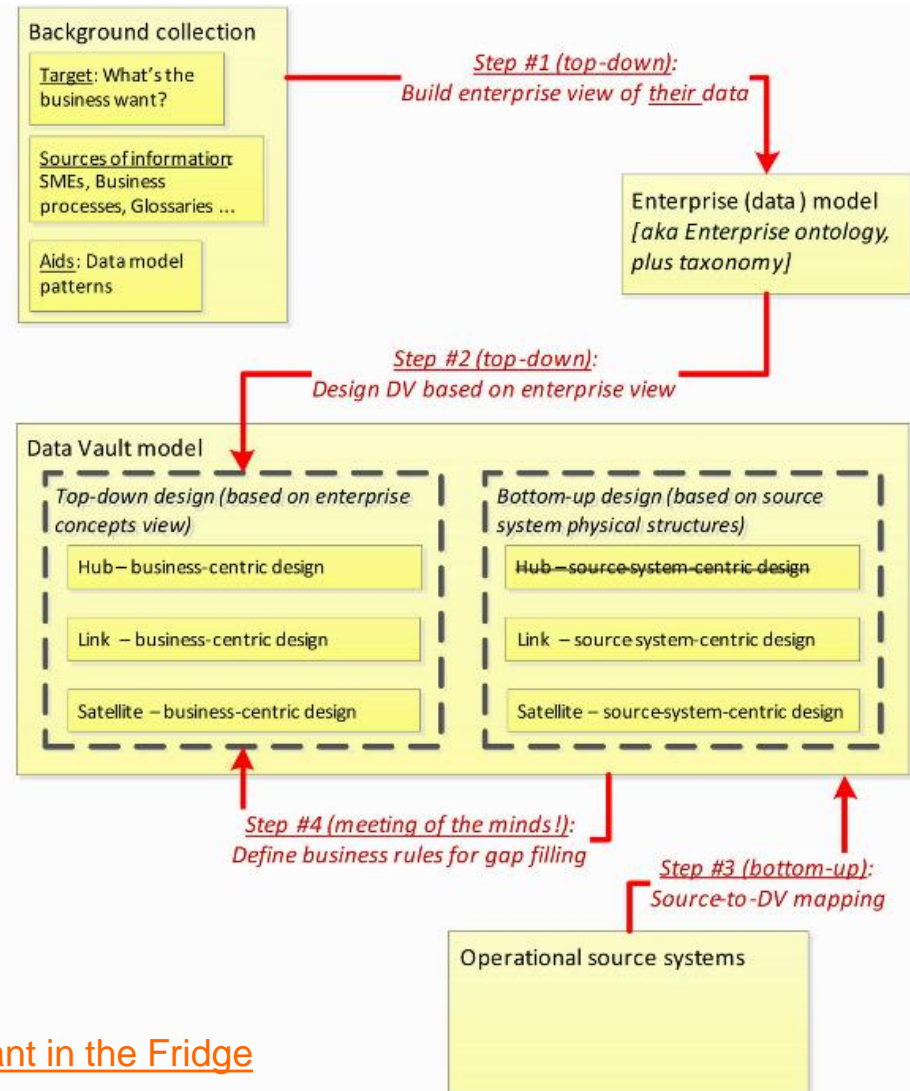
## Data Vault „csontváz”

Data Vault – Hubs / Links / Satellites



Forrás: <http://www.datavaultglobaltour.com/>

# Dan Linstedt: "Data Vault modeling was, is and always will be about the business..."



Forrás:

[John Giles: The Elephant in the Fridge](#)

# Data Vault kialakítása/áttérés

---

## DV modell kialakítása

- Top-Down „rész”
- Üzleti-koncepcionális modell (ontológia)
- 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?

## EDW rendszer kialakítása

- Bottom-up „rész”
- DW szint kialakítása
  - Adatforrások felmérése, elemzése
  - Mapping
  - Raw DV, Quality DV, Error Mart, stb. struktúrák
  - Alap ETL eljárások, „hard business rule”-ok, stb.
- Data Martok (és/vagy Business Vault) kialakítása
  - Adatstruktúrák (BDV, DM, flat&wide, csillagok)
  - ETL eljárások („soft business rule”-ok),
- Felhasználói elérés, riportok, adat küldések stb.



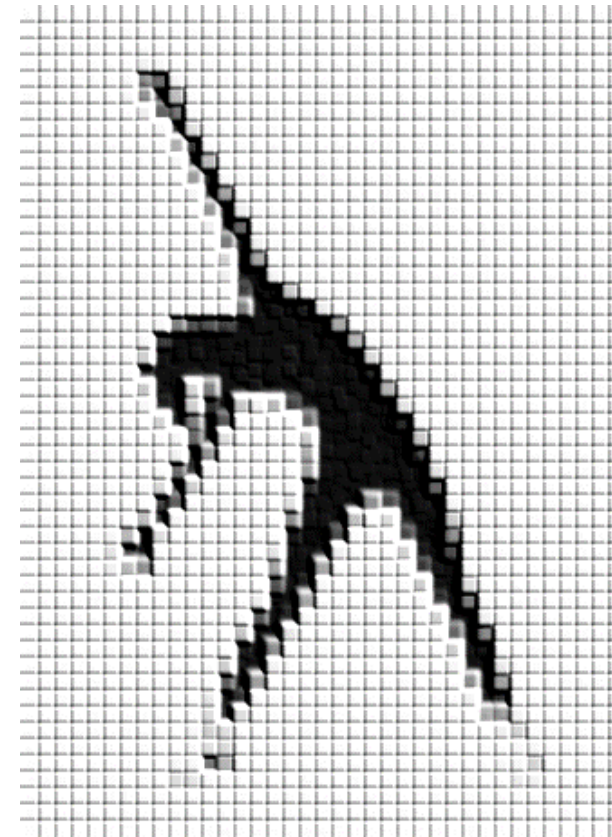
# 1. gyakorlat – kiinduló üzleti modell(?)

---

## Feladat

- 3-4 fős csoportokban dolgozzatok
- Válasszatok egy üzleti területet
- Vázoljátok fel a kulcs üzleti elemeket és a tartalmukat/definíciójukat
- Technológia: papír + ceruza + beszélgetés

Közösen megbeszéljük az eredményt!





# Kulcs üzleti fogalom (Core Business Entity/Concept)

Név	Típus				
(Kulcs üzleti fogalom)	Esemény	Személy / Szervezet	Hely	Dolog	Egyéb (mi?)

# 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](http://en.wikipedia.org/wiki/Five_Ws)

# Példa - Kulcs üzleti fogalom (Core Business Entity/Concept)

Név	Típus				
	Esemény	Személy / Szervezet	Hely	Dolog	Egyéb (mi?)
(Kulcs üzleti fogalom)					
Konferencia					
Helyszín					
Terem					
Résztevő					
Előadó					
Előadás/workshop					
Szekció					
...					

# A Data Vault modell alapelemei

---

## **HUB**

**Primary Key (HASH)  
Business Key  
Load Date & Time  
Record Source**

## **LINK**

**Primary Key (HASH)  
HUB\_1 Key  
HUB\_n Key  
Load Date & Time  
Record Source**

## **SAT**

**Primary Key (HASH)  
HUB/LINK Key  
Load Date & Time  
Load End Date & Time  
Record Source**

## Hub

- Az üzleti kulcsok (Business Key) egyedi listája

## Link

- A Hub-ok közötti kapcsolatok (mindig N:M) egyedi listája

## Satellite

- A leíró adatok, teljes történetiség kezeléssel
- A Sat-nak egy és csak egy szülő táblája van (Hub vagy Link)

## Egyéb

- Kód, Referencia, PIT, Bridge táblák

# 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) [opcionális]
- 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

# 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:

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

# A Data Vault elemei - LINK

---

A Link az üzleti kulcsok közötti kapcsolatok egyedi listája („kombinációk”)

- 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 adatalemek közötti legrészletesebb kapcsolatot tárolja (múlt, jelen és jövőbeli kapcsolatot is)

Szabályai:

- Kettő vagy több Hub kulcsot (hash?) tartalmaz (FK a HUB felé)
- Lehet saját satellite táblája/táblái
- A Load date-től érvényes (technikai érvényesség), vagy ahogy a link-satellite meghatározza (technikai és/vagy üzleti érvényesség)
- 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) [opcionális]
- Link összetett kulcs =  $\Sigma$  Hub kulcsok (hub kulcsok = 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

# A Data Vault elemei - LINK

A Link az üzleti kulcsok közötti kapcsolatokat jelenti

LNK Ügyfél Kapcsolattartó		
<u>Ügyfél kapcsolattartó Hash</u>	char(32)	<pk>
Ügyfél Hash	char(32)	<ak, fk1>
Kapcsolattartó Hash	char(32)	<ak, fk2>
Load DT S	datetime	
Record Source	varchar(20)	
🔑 DV ID	<pk>	
🔑 UOW FK	<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

Ügyfél - Kapcsolattartó LINK				
Ugyfel_Kapcsolattarto_Hash	Ugyfel_Hash	Kapcsolattarto_Hash	Load_DTS	Record_Src
b7...30	7f...1f	e2...ef	2011.01.01	SALES
5d...cd	28...69	cf...ba	2011.01.01	HR
4a...ea	3c...b5	e2...ef	2011.01.01	SALES
9e...7c	7f...1f	cf...ba	2013.08.23	SALES

Kapcsolattartó HUB			
Kapcsolattarto_Hash	Kapcsolattarto_kod	Load_DTS	Record_Src
e2361171a46d9bc94311a2249de0c8ef	GABORG	2011.01.01	SALES
cf2b0d09b0567e0b3697f0e7adbc6eba	GEDIT	2011.01.01	HR

# 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 (elavult!)~~
- 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 Ügyfélnev		
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>		

Ügyfélnev 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	2011.01.01	CRM	2866...a769	Piszkos	Alfréd	
3c9ebb086069ce63856d9df46e0b4ab5	2011.03.10	2011.03.10	SALES	3c9e...4ab5	Szóke	Ciklon	
28666b283ce0a0bfa26df2339674a769	2011.01.02	null	CRM	9034...bd57	Piszkos	Fred	
7f2208d6f5c634e5bb5591d277ce8d1f	2014.06.30	2014.07.31	CRM	a2ef...85b0	don Fülig	Jimmy	
7f2208d6f5c634e5bb5591d277ce8d1f	2014.07.31	null	CRM	7f22...8d1f	Fülig	Jimmy	

# A Data Vault elemei - SATELLITE

---

## Jellemző SAT adatok:

- Státuszok
- Mennyiségek
- Dátumok
- Szöveges jellemzők

## 2. gyakorlat – kiinduló Data Vault modell

---

### Feladat

- 3-4 fős csoportokban dolgozzatok
- Az előző feladatban elemzett üzleti terület „elemekből” kiindulva
- Vázoljátok fel a kulcs üzleti elemeket (HUB) és a kapcsolataikat (LINK)
- Technológia: papír + ceruza + beszélgetés

Közösen megbeszéljük az eredményt!



# 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:**

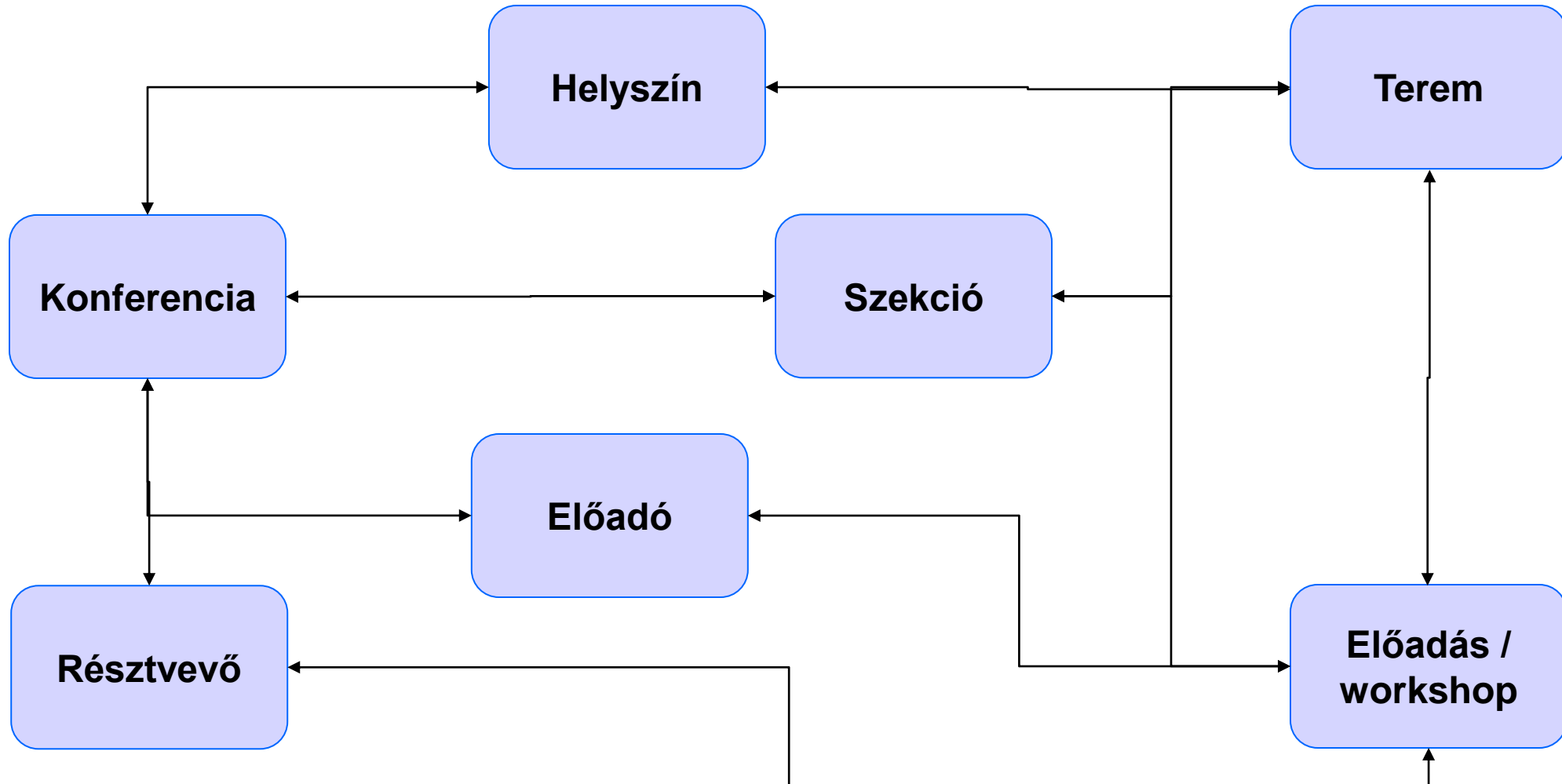
[http://en.wikipedia.org/wiki/Five\\_Ws](http://en.wikipedia.org/wiki/Five_Ws)



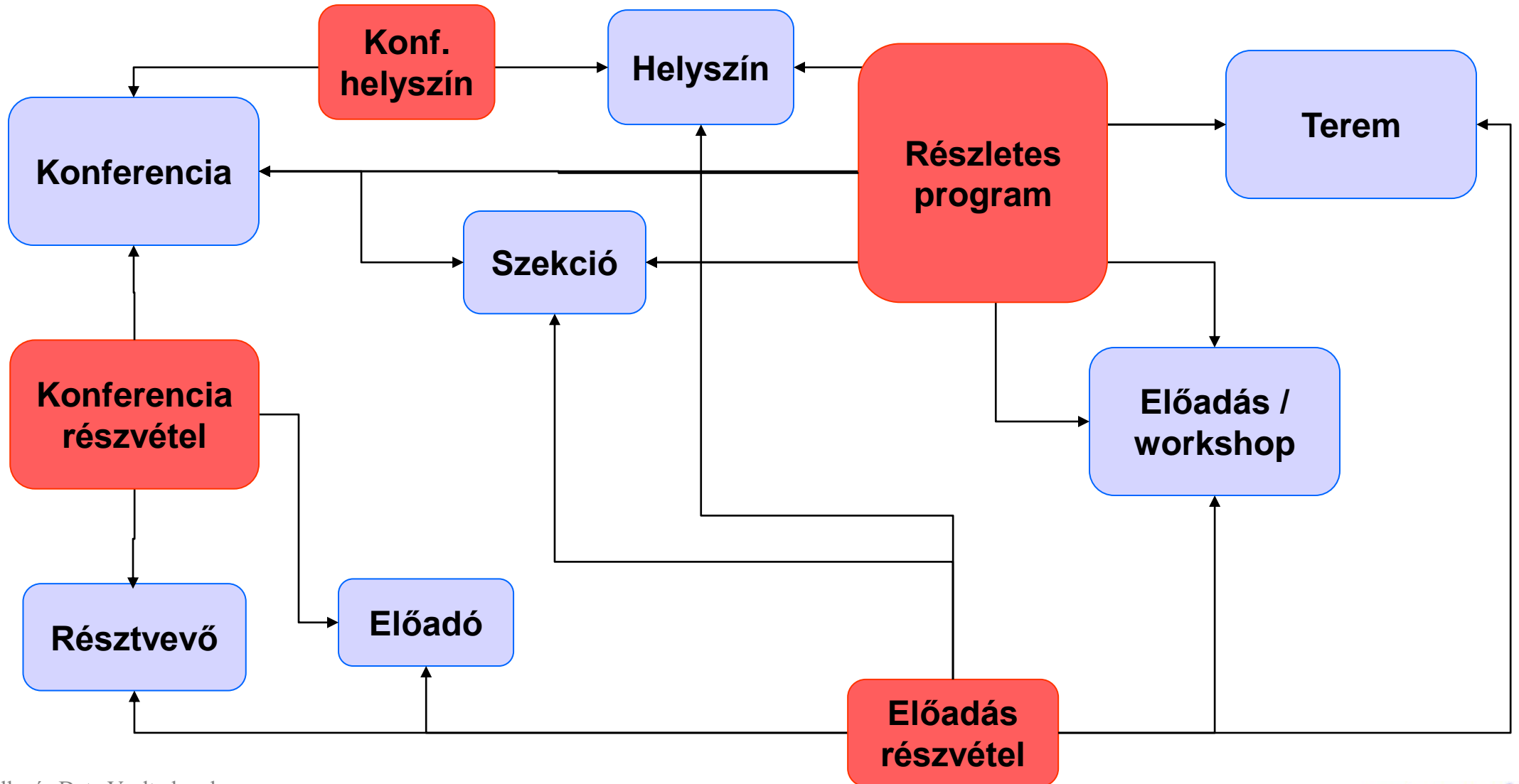
# Példa - Kulcs üzleti fogalom (Core Business Entity/Concept)

Név	Típus				
	Esemény	Személy / Szervezet	Hely	Dolog	Egyéb (mi?)
(Kulcs üzleti fogalom)					
Konferencia	X				
Helyszín			X		
Terem			X		
Részvevő		X			
Előadó		X			
Előadás/workshop	?			?	?
Szekció			?		
...					

# Kiinduló Data Vault modell „csontváz” – 1. verzió (Példa)



# Kiinduló Data Vault modell „csontváz” - 2. verzió (Példa)



# Data Vault modell elemek – további részletek

---

## Speciális esetek

- LINK
- SAT

## Egyéb elemek

- Referencia és kód táblák
- PIT és Bridge táblák

# A Data Vault elemei - LINK

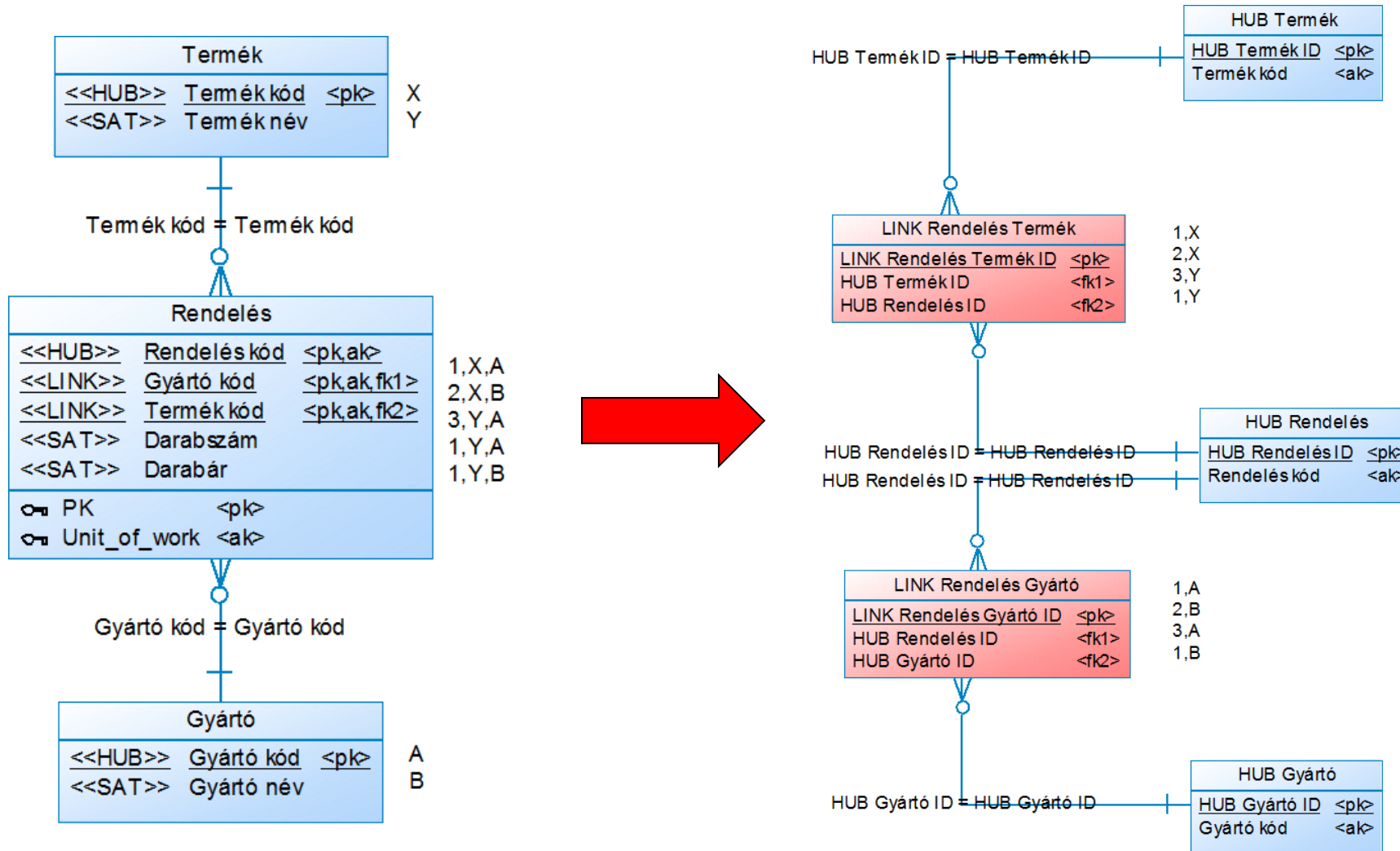
---

A Link-ek definiálásánál fontos:

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

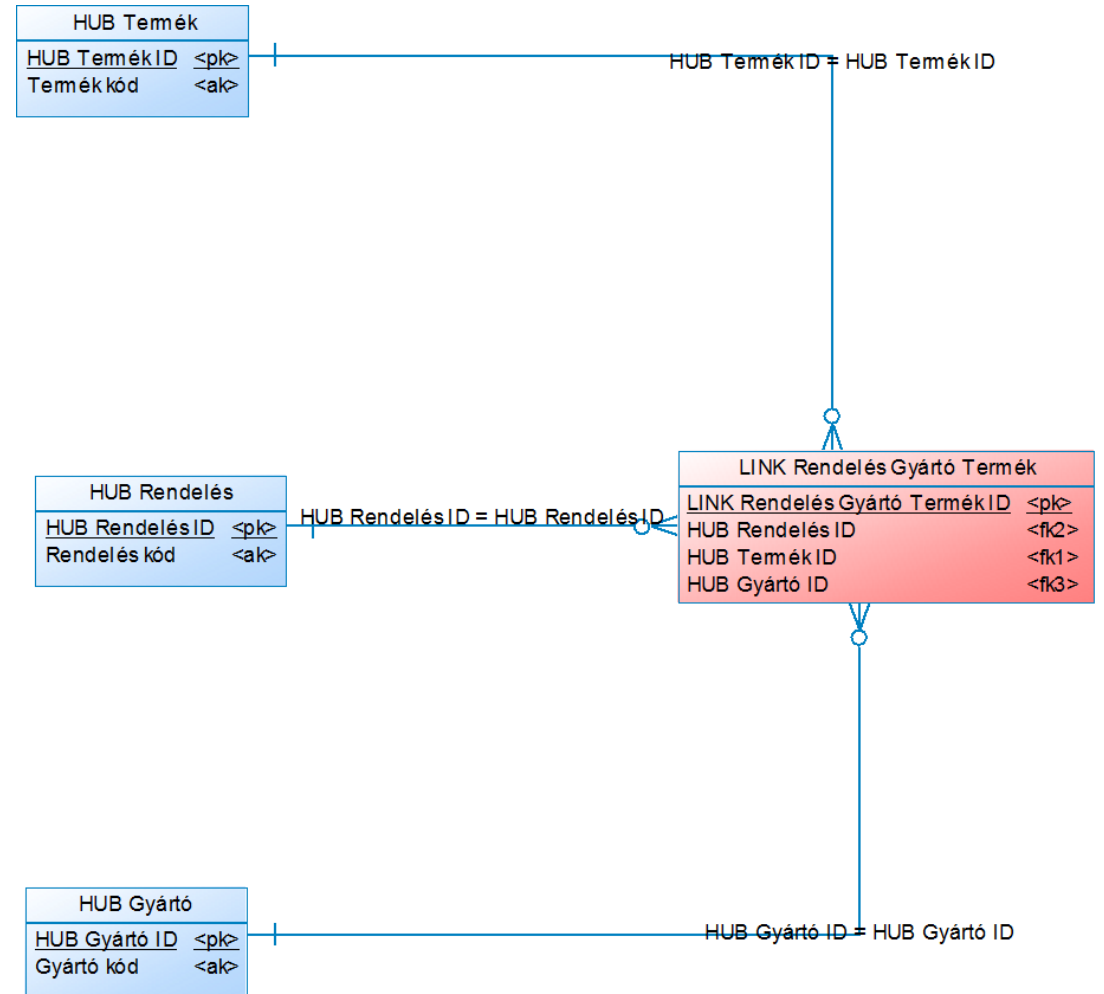
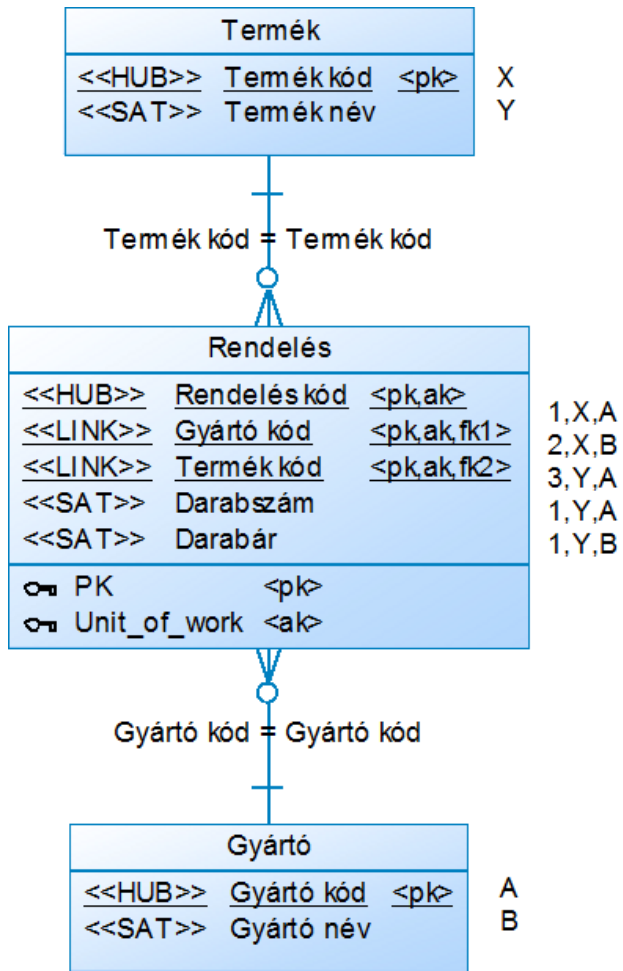
# LINK – Unit-of-Work (UoW)

## 1. Kapcsolatok felbontása (linkek létrehozása)



# LINK – Unit-of-Work (UoW)

## 2. Kapcsolatok felbontása (helyesen)



# LINK meghatározó kulcsa (Driving key)

## Szabályok/feltételezések:

- A Link kulcsa összetett kulcs
- De a Link üzleti PK-ja lehet szűkebb!!!  
Pl. egy dolgozó egy időben csak egy osztályon dolgozhat
- A Link történetét a SAT-ok kezelik

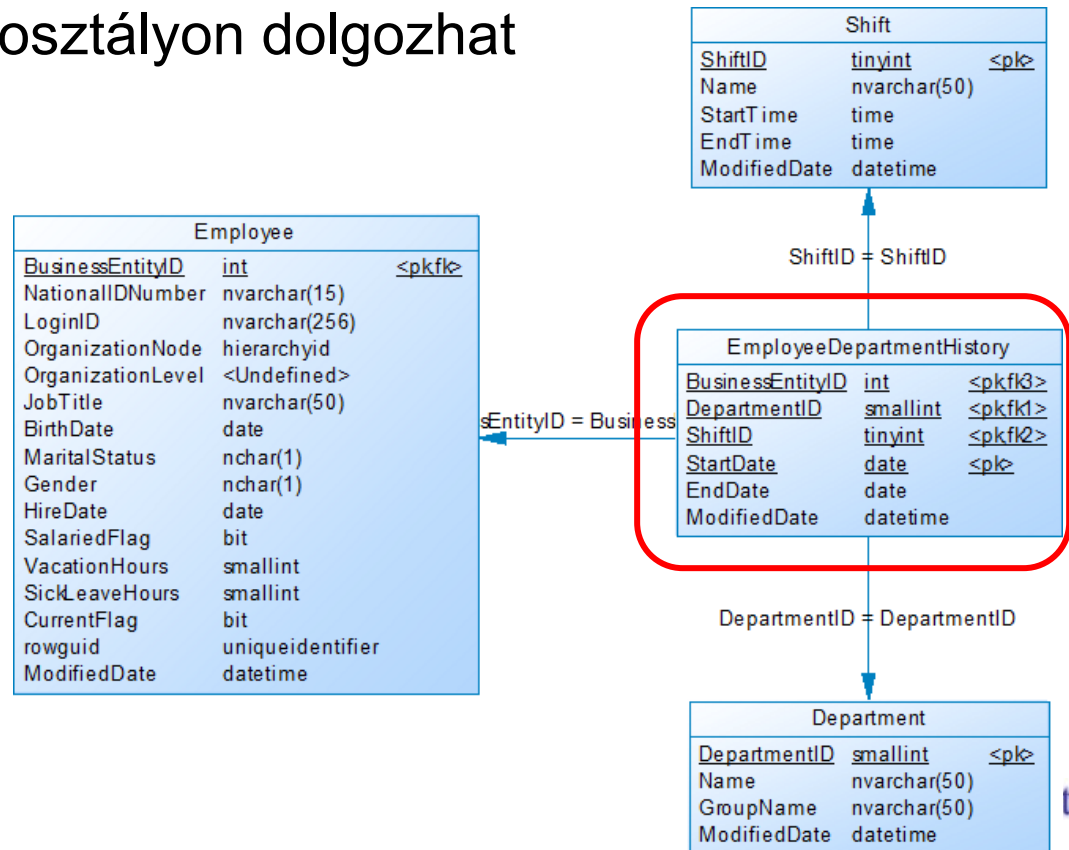
## ■ Példa:

Dolgozó 1; Osztály SY

majd átmegy máshová

Dolgozó 1; Osztály DW

Hogyan követjük le?

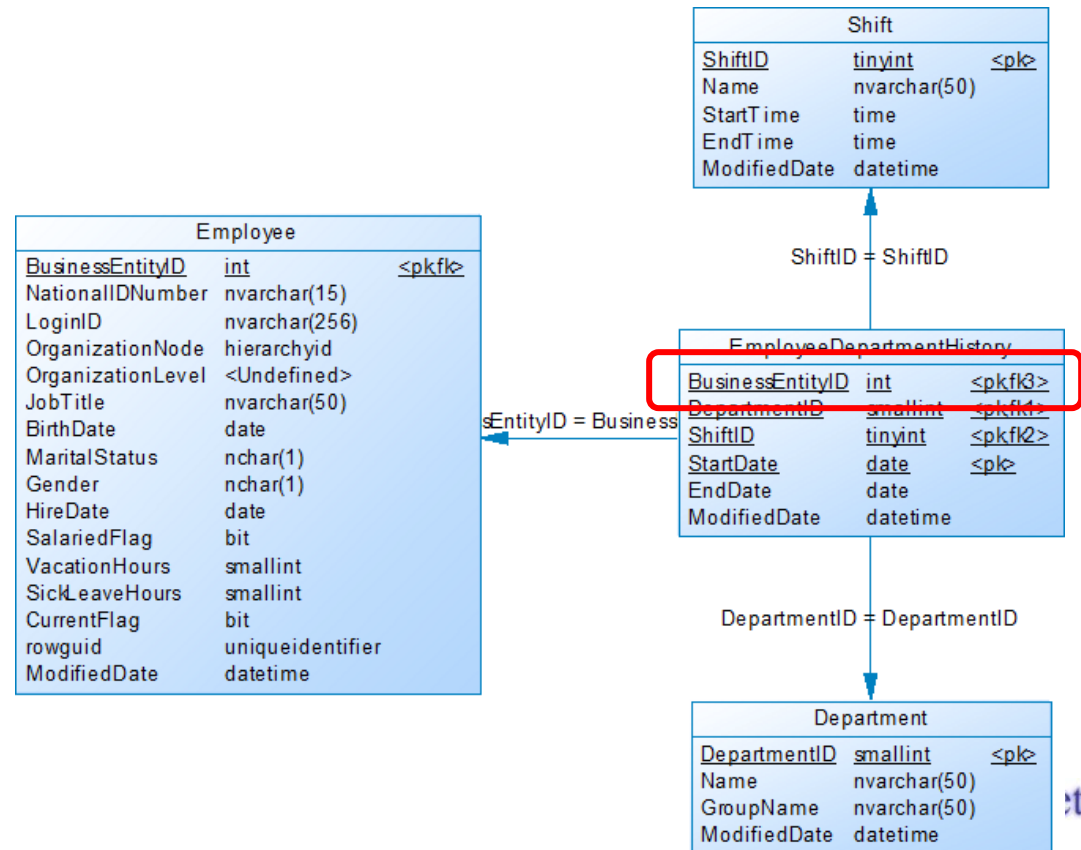




# LINK meghatározó kulcsa (Driving key)

Példa:

- Hiába összetett a kulcs, csak a „Dolgozó azonosítót” (BusinessEntityID) követjük.
- Azaz, új LINK rekordot veszünk fel,
- Új SAT rekordot veszünk fel hozzá,
- és, a régi LINKhez tartozó SAT rekordot **lezárjuk**.
- A történetiség kezelést ez a kulcs vezérli (driving key)



# A Data Vault elemei - SATELLITE

---

## Satellite-ek típusai:

- Sima
- Multi-valued v. Multi-active (advanced \*)
- Status tracking (advanced \*\*)
- Effectivity (advanced \*\*)
- Record tracking (advanced \*\*)

# Multi-active SATELLITE

Egy HUB kulcshoz, egy időben több érvényes/aktív SAT rekord is tartozik  
(pl. ügyfél → telefonszám)

A SAT PK-ja a HUB\_PK+LOAD\_DATE + még valami! leíró jellemző  
[esetleg sub-sequence]

Satellite		
>> <None>		
DV.SAT_PRODUCT		
<u>HUB_PRODUCT_HASH</u>	char(32)	<pk,akfk>
<u>META_LOAD_DATE</u>	datetime	<pk,ak>
<del>META_LOAD_END_DATE</del>	<del>datetime</del>	
META_RECORD_SOURCE	varchar(200)	
META_INTERFACE_ROWNUM	int	
META_INTERFACE_DATA_ID	int	
SAT_PRODUCT_ID	int	
PRODUCT_CODE	char(8)	
PRODUCT_ID	numeric(10)	
PRODUCT_NAME	varchar(100)	
PRODUCT_DESCRIPTION	varchar(200)	
PRODUCT_PICTURE	image	
AVAILABILITY_DATE_FROM	datetime	
AVAILABILITY_DATE_TO	datetime	
SAT_PRODUCT_HASHDIFF	char(32)	

Link-Satellite		
<<MLST>>		
>> <None>		
DV.LST_ORDER_DETAIL		
<u>LNK_ORDER_DETAIL_HASH</u>	char(32)	<pk,akfk>
<u>META_LOAD_DATE</u>	datetime	<pk,ak>
<del>META_LOAD_END_DATE</del>	<del>datetime</del>	
META_RECORD_SOURCE	varchar(200)	
META_INTERFACE_ROWNUM	int	
META_INTERFACE_DATA_ID	int	
LST_ORDER_DETAIL_ID	int	
ORDER_LINE_ID	numeric(16)	
<u>ORDER_LINE_NUMBER</u>	<u>numeric(4)</u>	<pk,ak>
ORDERED_AMOUNT	numeric(8)	
LIST_PRICE	money	
DISCOUNT_PCT	numeric(3)	
LINE_ITEM_PRICE	money	
LST_ORDER_DETAIL_HASHDIFF	char(32)	

# Egyéb tábla típusok

## Referencia táblák

---

### Referencia táblák

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

# Egyéb tábla típusok

## PIT (Point-In-Time) táblák

---

- A DV lekérdezést és DM töltést segítő táblák
- Ha több SAT tartozik egy HUB/LINK táblához, akkor ezeknek eltérő lehet a története  
→ időszeleteket kell lekérdezni
- A HUB-SAT-SAT-SAT... táblák egy adott időpillanatban érvényes sorait köti össze
- A virtualizált (view) DM megoldás egyik kulcs eleme!
- Néha csak limitált időtávig őrizzük csak meg az adatait

### Megengedett struktúrája:

- HUB/LINK hash key (FK)
- Snapshot Datetime
- SAT1 hash key, SAT1 Load Date ...
- SATn hash key, SATn Load Date

# Egyéb tábla típusok

## Bridge táblák

---

- A DV lekérdezést és DM töltést segítő táblák
- Az összetett pl. HUB-LINK-HUB-LINK-HUB kapcsolatokat egyszerűsíthetjük ki velük
- A hivatkozott táblák kulcsait köti össze

### Megengedett struktúrája:

- Bridge szekvencia (ID)
- Snapshot Date
- HUB/LINK1 hash key (FK)
- ...
- HUB/LINKn hash key (FK)
- opcionális: leíró jellemzők (SAT) [ha azonos a grain!]

# 3. gyakorlat – Data Vault modell kiegészítése

---

## Feladat

- Az előző gyakorlatban elkezdett modell folytatása
- 3-4 fős csoportokban dolgozzatok
- Vázoljátok fel a leíró adatokat, melyik milyen kulcs üzleti „concept”-hez vagy kapcsolathoz tartozik
- Vannak-e Referencia/Kód táblák?
- Szükséges-e PIT v. BRIDGE táblákat létrehozni?
- Technológia: papír + ceruza + beszélgetés

Közösen megbeszéljük az eredményt!



# KÉRDÉSEK & VÁLASZOK

---



[info@metaconsulting.hu](mailto:info@metaconsulting.hu)



# Olvasnivalók 😊

---

## Adatmodellezés:

- [Data Modeling Made Simple - http://technicspub.com/core/](http://technicspub.com/core/)
- [Enterprise Model Patterns - http://technicspub.com/patterns/](http://technicspub.com/patterns/)

## Data Warehouse & Data Vault:

- [Building a Scalable Data Warehouse with Data Vault 2.0](#)
- [The Elephant in the Fridge](#)
- [The Data Vault Guru: a pragmatic guide on building a data vault](#)
- [Super Charge Your Data Warehouse](#)
- [Modeling the Agile Data Warehouse with Data Vault](#)
- [The Official Data Vault Standards Document \(Version 1.0\)](#)