



# GENERALI

## Data Vault modellezés vizsgálata a Generali Biztosítóban

Csonka Zoltán

Adattárház architekt

[csonka.zoltan@general.com](mailto:csonka.zoltan@general.com)

# Miről lesz szó

- Adattárház környezet
- Üzleti elvárások
- Data Vault ígérete
- Data Vault vs. jelenlegi modell (3NF)
- Tapasztalatok

# Adattárház környezet

## Technikai háttér

- Oracle Database 11gR2 RAC, 14 TB
- Oracle Warehouse Builder ETL eszköz
- ~20 betöltő folyamat, ~5000 mapping
- ~50 forrásrendszer

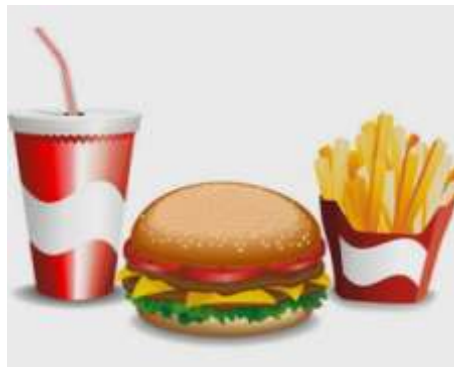
## Módszertan

- A tervezési módszertan nem egységes
- Tervezés és fejlesztés változás átvezetés lassú
- Auditálási hiányosságok

# Üzleti elvárások

- Gyorsabb reakció a változásokra
- Agilisabb fejlesztési megközelítés
- Költséghatékony fejlesztés
- Eredeti forrásadatok, „tények”, hisztorizált tárolása
- Teljes mértékben visszakövethető adatok

„Nem tudjuk pontosan, hogy mit és hogyan,  
de tudjuk miből”



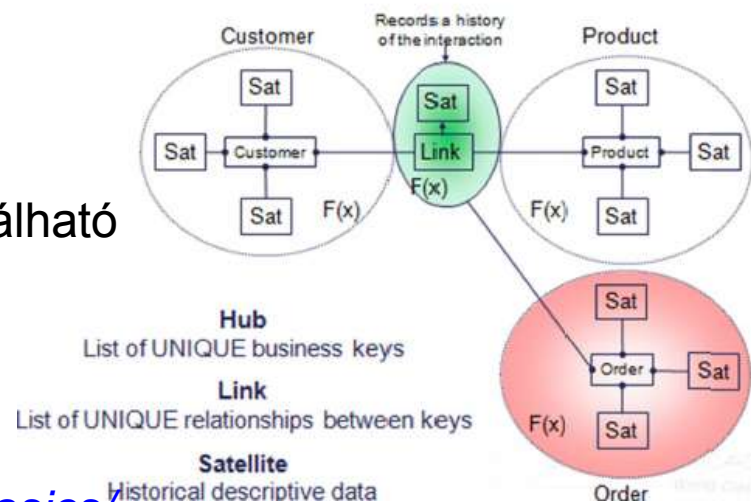
# Data Vault ígérete

„The Data Vault is a detail oriented, historical tracking and uniquely linked set of normalized tables that support one or more functional areas of business”

Dan Linstedt

DV (RDV) Adattárolásra optimalizált megoldás

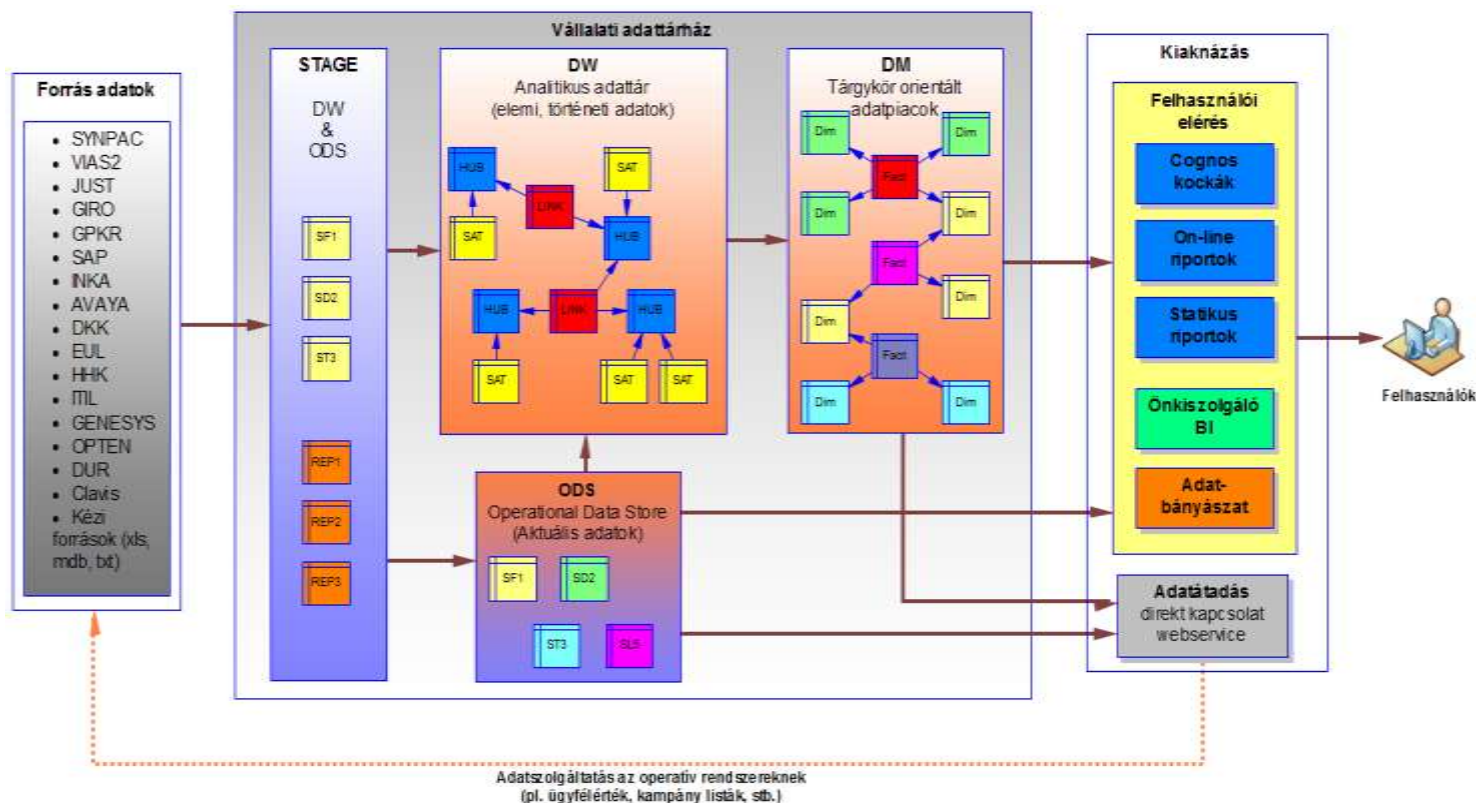
- „Single version of fact”
- A változások átvezetése nagyon egyszerű
- Kiválóan támogatja az agilis fejlesztési szemléletet
- Egyszerű komponensek, nagyon jól generálható
- Teljes mértékben auditált
- Jól párhuzamosítható töltések, folyamatok
- Valós idejű töltésekre optimalizált



<http://danlinstedt.com/solutions-2/data-vault-basics/>

# Data Vault illesztése meglévő adattárház rendszerhez

- 3 rétegű architektúrának köszönhetően (változások átvezetése felhasználó számára észrevétlen)
- Adatpiac részben virtualizált (view használata)



# Betekintés a Data Vault fejlesztésbe

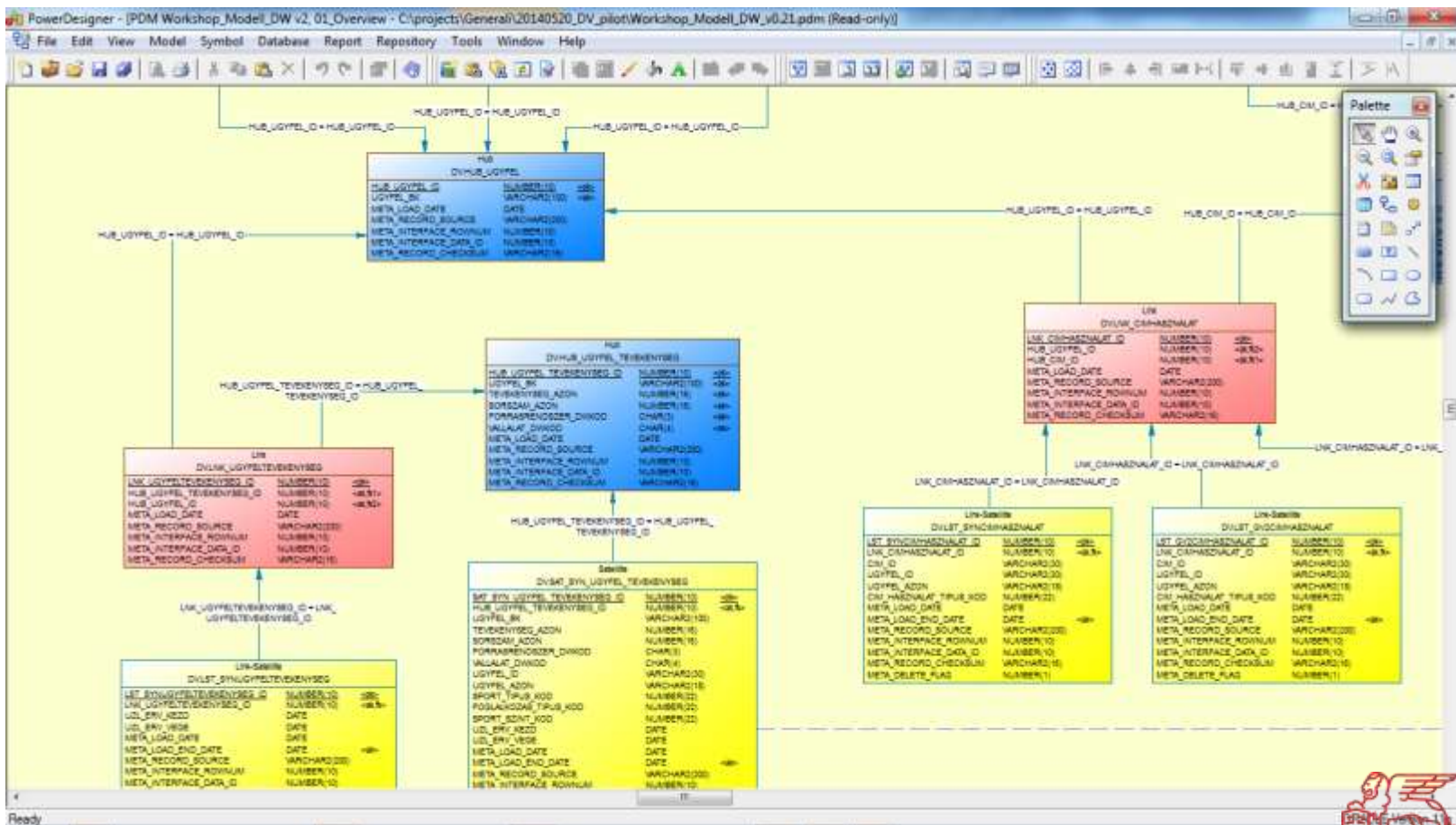
Mi a Best Practice?

Cél, hogy lássuk miképpen épülhetnek fel a betöltők és ezekből a folyamatok.

- Egy üzleti témakör töltésének átalakítása 3NF-ről Data Vault-ra
- Működő folyamat 0-ról külsős segítséggel 2 nap alatt

# DV modell részlet

A modellt és a kódot PowerDesignerből generáltuk.





# Generált SQL kód (részlet)

A névkonvenciók, kódok egységesek.

```
TextPad - C:\projects\Generali\20140520_DV_pilot\STAGE_GDV_LOADTOLST.pks
File Edit Search View Tools Macros Configure Window Help
Find incrementally Match case
STAGE_GDV_LOADTOLST.pks STAGE_GDV_LOADTOSAT.pks

merge into DW_DV.LST_CIMHASZNALAT trg
using tVUT_IDVCIMHA_GDVSVCIM src
on ( src.LST_CIMHASZNALAT_ID = trg.LST_CIMHASZNALAT_ID
)
when matched then
update set trg.meta_load_end_date = nvl( src.interface_datetime, to_date('99991231','YYYYMMDD'));

if ( SQL%ROWCOUNT > 0 ) then v_BCD2_Closed := SQL%ROWCOUNT; end if;

execute immediate ('truncate table tVUT_IDVCIMHA_GDVSVCIM');

insert /*+ APPEND NOLOGGING PARALLEL */ into DW_DV.LST_CIMHASZNALAT
(LST_CIMHASZNALAT_ID
,LNK_CIMHASZNALAT_ID
,CIM_ID
,UGYPFEL_ID
,UGYPFEL_AZON
,CIM_HASZNALAT_TIPUS_KOD
,META_LOAD_DATE
,META_LOAD_END_DATE
,META_RECORD_SOURCE
,META_INTERFACE_ROWNUM
,META_INTERFACE_DATA_ID
,META_RECORD_CHECKSUM
,META_DELETE_FLAG
)
select DW_DV.S.LST_CIMHASZNALAT_ID.nextval
,v.LNK_CIMHASZNALAT_ID
,v.CIM_ID
,v.UGYPFEL_ID
,v.UGYPFEL_AZON
,v.CIM_HASZNALAT_TIPUS_KOD
,case p_initial_load_flag when 0 then i.interface_datetime else nvl(v.META_LOAD_DATE,to_date('19990721','YYYYMMDD')) end as META_LOAD_DATE
,case p_initial_load_flag when 0 then to_date('99991231','YYYYMMDD') else nvl(v.META_LOAD_END_DATE,to_date('19490721','YYYYMMDD')) end as META_LOAD_END_DATE
,ad.system_code||'/'||ad.instance_code||'/'||i1.interface_base_name as META_RECORD_SOURCE
,v.META_INTERFACE_ROWNUM
```

# Data Vault vs. jelenlegi folyamat (3NF)

## DV előnye

- Erős szabályrendszer
- Jól párhuzamosítható töltési folyamatok
- Nagyon rugalmas, a változásokat jól tűri
- Gyors eredményeket lehet elérni
- Az inkrementális fejlesztést jól támogatja

## DV hátránya

- Sok objektum, sok join
- Kevés DV szakértő, tapasztalat

# Tapasztalatok a Generali szemszögéből

## Legfontosabb tervezési irányelvek

- Szigorú szabályok, egyértelműbb útmutatás
- Sok objektum, generátor használata erősen javasolt
- Sok join, adatpiac építése szükséges
- Üzleti szabályok RDV (Raw Data Vault) szint fölött
- Az elemek egyszerűsége segíti az egyszerűbb generálást

## Kifejezetten jól használható ahol

- Homályosak az üzleti igények
- A forrásrendszerekre kevés ráhatás van (adatminőség)
- Prototípus alapú fejlesztésnél, agilis megközelítésnél
- Fontos a jól auditált rendszer
- Több forrás, nagyobb adatmennyiség esetén

# Következő lépések

- Döntés a bevezetésről
- Belső képzés
- Generátor vásárlás, fejlesztés
- Data Vault 2.0 - június

# Köszönöm a figyelmet!

## Várom a kérdéseket!



Csonka Zoltán  
Generali Biztosító  
Adattárház architekt  
[csotka.zoltan@generali.com](mailto:csotka.zoltan@generali.com)