

# BEVEZETÉS AZ ADATTÁRHÁZ AUTOMATIZÁLÁSBA

Gollnhofer Gábor – JET-SOL Kft.



Nyilvántartási szám:  
503/1256-1177

# JET-SOL KFT.

- Alapadatok
  - 2003-ban alakultunk
  - Több mint 120 magasan képzett munkatárs
- Ügyfélkör
  - Nagyvállalati szektor
  - Magyarországon és Közép-Kelet Európában
- Értékeink
  - Precizitás, professzionalitás, hatékonyság

# DW & BI SZAKTERÜLETEINK

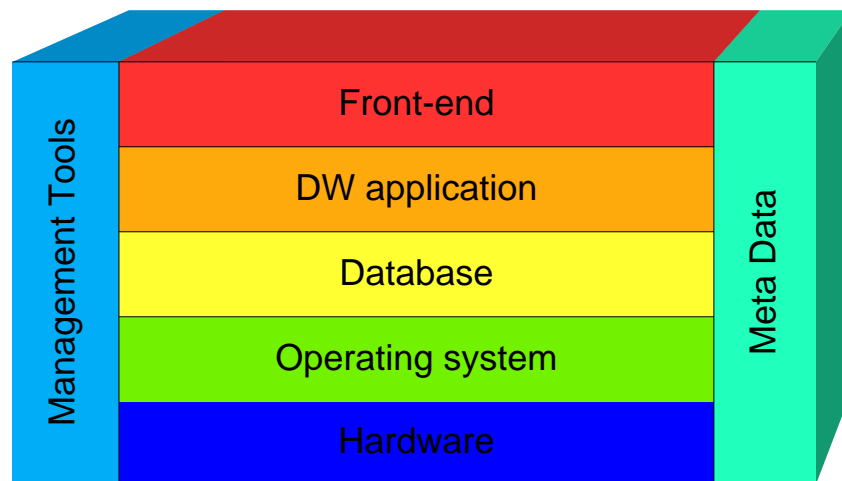
- Szakterületeink
  - Adattárház és üzleti intelligencia rendszerek
  - Adatvagyon biztosítás (Data Governance)
  - Metaadat kezelés
  - Vezetői információs rendszerek
- Technológiai partnereink
  - EMC<sup>2</sup>, Oracle, Microsoft, SAP, Teradata, AnalytixDS

# TARTALOM

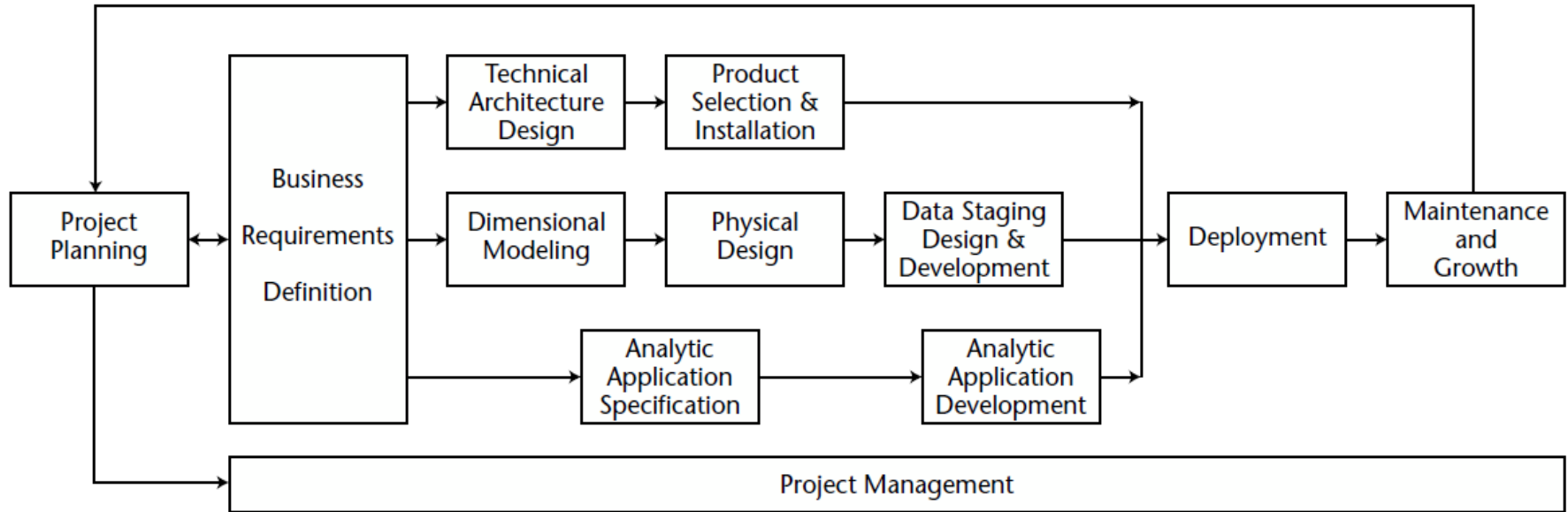
- Bemutatkozás ✓
- Hogyan építünk, használunk, üzemeltetünk adattárházat?
- Mit érdemes automatizálni?
- Mit nem érdemes automatizálni?
- Kérdések & válaszok

# A DW/BI RENDSZEREK ELEMEI

- A rendszer elemei
  - Kiaknázás, riporting
  - ETL
  - Adatbázis
  - Op. rendszer
  - Hardver
  - Metaadat kezelés
  - Menedzsment eszközök



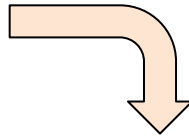
# DIMENZIÓS DW ÉLETCIKLUSA



Forrás: [The Data Warehouse Toolkit \(Ralph Kimball, Margy Ross\)](#)

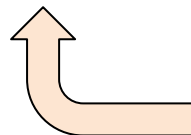
# A TERVEZÉS FOLYAMATA

Üzleti követelmények  
(követelmény felmérés)



## **Dimenziós modell**

1. üzleti folyamat
2. részletezettség (grain)
3. dimenziók
4. tények



Tényleges adatok  
(adat profilozás)

# HOGYAN ÉPÍTÜNK, HASZNÁLUNK, ÜZEMELTETÜNK ADATTÁRHÁZAT?

- Követelmények összegyűjtése
  - Üzleti és technikai követelmények
- Tervezés
  - Adatforrások felmérése, profilozás
  - Adatmodellezés
  - ETL tervezés
  - Kiaknázás tervezés (BI, analytics, operatív feedback)
- Kiépítés
  - ETL és kiaknázás fejlesztés
  - Tesztelés, élesítés, dokumentálás, oktatás, stb.
- Üzemeltetés
  - Monitorozás, adminisztráció: rendszer, DB, ETL, BI, jogosultságok



# TARTALOM

- Bemutatókozás ✓
- Hogyan építünk, használunk, üzemeltetünk adattárházat? ✓
- Mit érdemes automatizálni?
- Mit nem érdemes automatizálni?
- Kérdések & válaszok

# MIÉRT AUTOMATIZÁLJUNK?

- Egységesebb eredmények/leszállítandók → kevesebb hiba, könnyebben követhető, javítható
- Hatékonyabb emberi erőforrás kihasználás → az intelligenciát igénylő feladatokra több idő jut
- Rövidebb ciklusidők → gyorsabb reagálás a változó üzleti igényekre

# MIT NEM ÉRDEMES AUTOMATIZÁLNI?

- Az intelligenciát és interakciót igénylő feladatokat
  - Követelmények összegyűjtése és elemzése
  - Tervezés (részben automatizálható)
    - Forrásadatok elemzése (profilozás, statisztikák gyűjtése)
    - Adatmodellezés (pl. staging terület)
    - ETL tervezés (pl. 1-1 mapping source→stage)
  - Fejlesztés
    - Összetett, nehezen automatizálható feladatokat  
Pl. összetett riportok, bonyolult KPI kalkulációk

# MIT AUTOMATIZÁLJUNK?

- Ismétlődő feladatokat, sablon feladatokat, pl.:
- Tervezés, fejlesztés:
  - Adatforrás profilozás
  - Forrásrendszeri adatkinyerések
  - Stage: modellezés, ETL (általában 1:1 mapping)
  - DW: egyszerű táblák modellezése, töltése
- Üzemeltetés
  - Monitorozás, alertek (megfelelő metaadatok!)

# HOGYAN AUTOMATIZÁLJUNK? DW TERVEZÉS & FEJLESZTÉS

- Eszközök(?)
  - Profiling
  - Adatmodellező eszközök (scriptelhető, template-ek)
  - ETL eszközök (scriptelhető, template-ek)
  - Saját fejlesztésű megoldások (pl. SQL generátor)
- Módszerek(?)
  - KISS – Keep It Simple & Stupid
  - Fejlesztői szabvány, névkonvenciók, template-ek

# MIVEL AUTOMATIZÁLJUNK? DW TERVEZÉS & FEJLESZTÉS

- „Megvásárolható” eszközök, pl.
  - Analytix Mapping Manager
  - BIReady
  - Consensus
  - Qipu
  - Wherescape RED
- Saját, egyedi fejlesztés

# JET-SOL POWERDESIGNER BŐVÍTMÉNY DW MODELLEZÉS – DW TÁBLA JELLEMZŐK

PowerDesigner - [PDM Sakila DV example MSSQL, PhysicalDiagram\_1 - C:\dokum\_konyvek\_tudas\Modeling\_MetaData\Data\_Vault\DataVault\_Technical\_Book\Paid\_Book\Saki]

File Edit View Model Symbol Database Report Repository Tools Window Help

**Table: sakila\_data\_vault.sat\_category**

sat_category_id	int	<pk>
hub_category_id	int	<ak1,ak2,fk>
load_dts	timestamp	<ak1>
load_end_dts	datetime	
record_source	varchar(100)	
last_update_src	datetime	
category_id_src	int	

**Table: sakila\_data\_vault.sat\_store**

sat_store_id	int	
hub_store_id	int	
load_dts	timestamp	
load_end_dts	datetime	
record_source	varchar(100)	
last_update_src	datetime	

**Table: sakila\_data\_vault.hub\_store**

hub_store_id	int	<pk>
store_id	int	<ak>
load_dts	timestamp	
record_source	varchar(100)	

**Table: sakila\_data\_vault.link\_sat**

link_sat_id	int	<pk>
sat_category_id	int	<ak1,ak2,fk>
sat_store_id	int	<ak1,ak2,fk>
load_dts	timestamp	<ak2>

**Table: sakila\_data\_vault.hub**

hub_store_id	int	<pk>
store_id	int	<ak>
load_dts	timestamp	
record_source	varchar(100)	

**Table: sakila\_data\_vault.link\_sat**

link_sat_id	int	<pk>
sat_category_id	int	<ak1,ak2,fk>
sat_store_id	int	<ak1,ak2,fk>
load_dts	timestamp	<ak2>

Version: 2012.06.15.01

Code: sat\_category Short Table Code: SCAT

Business Description:

Business Owner:

Granularity:

Subject Area: Film Sub-Subject Area:

Definition Status: 30 - Defined Extract Spec. Status: 50 - Validated

Data Vault Table Type: Satellite

Dimensional Type: <None> SCD Type:

Refresh Method: Update-Insert Refresh Periodicity: daily Refresh deadline: Avg. Change%: 5

Data Source

Source System Name: SAKILA\_SRC

Source Module Name:

Source Table Name: Category

Load Process

Name of Load Group: LD\_SAT

Name of Load Process:

Name of Load Map:

OK Cancel

Microsoft SQL Server 2008

# JET-SOL POWERDESIGNER BŐVÍTMÉNY DW GENERÁLÁS – GENERÁLT ETL PROCEDÚRA

```
TextPad - C:\temp\meta_ddl.sql
File Edit Search View Tools Macros Configure Window Help
meta_ddl.sql
v_TableDb := '';
v_TableSchema := 'DW_DV';
v_TableName := 'HUB_SZERVEZET';

STAGE_CONTROL.PRE_LOADTODV_ACTIVITIES (p_target_system_code, p_target_instance_code, v_TableDb, v_TableSchema, v_TableName
-- out
, c_mindate, c_maxdate, c_newline, v_RecordCountBefore, v_RecordCountAfter, v_New, v_Modified, v_Closed, v_SCD2_Inserted, v_SCD2_Closed
, v_LastPeriod, v_LastModification, v_maindir, v_logdir, v_iface_cycle_proc_start_time);

insert /*+ APPEND NOLOGGING PARALLEL */ into DW_DV.HUB_SZERVEZET
(SZERVEZET_ID,
 meta_load_date,
 meta_record_source,
 meta_interface_data_id,
 meta_interface_rownum,
 meta_record_checksum
)
select DW_DV.S_SZERVEZET_ID.nextval,
 meta_load_date,
 meta_record_source,
 meta_interface_data_id,
 meta_interface_rownum,
 meta_record_checksum
from (
select
case p_initial_load_flag when 0 then i.interface_datetime else to_date('19690721','YYYYMMDD') end as meta_load_date,
sd.system_code||'/'||ssd.instance_code||'/'||il.interface_base_name as meta_record_source,
v.meta_interface_data_id as meta_interface_data_id,
min(v.meta_interface_rownum) as meta_interface_rownum,
null as meta_record_checksum
from
VIH_MAPPING_1 v
inner join dw_meta.interface_data i on i.interface_data_id = v.meta_interface_data_id
inner join dw_meta.interface_descriptor il on il.interface_descriptor_id = i.interface_descriptor_id
inner join dw_meta.source_instance_descriptor ssid on ssid.source_instance_descriptor_id = i.source_instance_descriptor_id
inner join dw_meta.system_descriptor sd on sd.system_descriptor_id = ssid.source_system_descriptor_id
group by
i.interface_datetime,
sd.system_code,
ssid.instance_code,
```



# TARTALOM

- Bemutatókozás ✓
- Hogyan építünk, használunk, üzemeltetünk adattárházat? ✓
- Mit érdemes automatizálni? ✓
- Mit nem érdemes automatizálni? ✓
- Kérdések & válaszok

# ÖSSZEFOGLALÁS

- Mit érdemes automatizálni?
  - A (viszonylag) egyszerű, ismétlődő feladatokat
  - ETL/ELT – ebből főleg az EL (extract&load)
- Hogyan érdemes automatizálni
  - KISS – Keep It Simple and Stupid
  - Fejlesztői szabályok, névkonvenciók, template-ek
  - Scriptelhető eszközöket használjunk vagy vásárolhatunk „kész” megoldásokat

# KÉRDÉSEK & VÁLASZOK

[gabor.gollnhofer@jet-sol.hu](mailto:gabor.gollnhofer@jet-sol.hu)

[www.jet-sol.hu](http://www.jet-sol.hu)



## KAPCSOLAT

1143 Budapest, Ilka utca 2-4.

Telefon: +36 1 453 0432

Fax: +36 1 430 1976

E-mail: [info@jet-sol.hu](mailto:info@jet-sol.hu)



[www.jet-sol.hu](http://www.jet-sol.hu)



Nyilvántartási szám:  
503/1256-1177

