



ETL keretrendszer tervezése és implementálása

Gollnhofer Gábor – Meta4Consulting Europe Kft.

Tartalom

Bevezetés

ETL keretrendszer: elvárások és hogyan készítsük

A mi keretrendszerünk

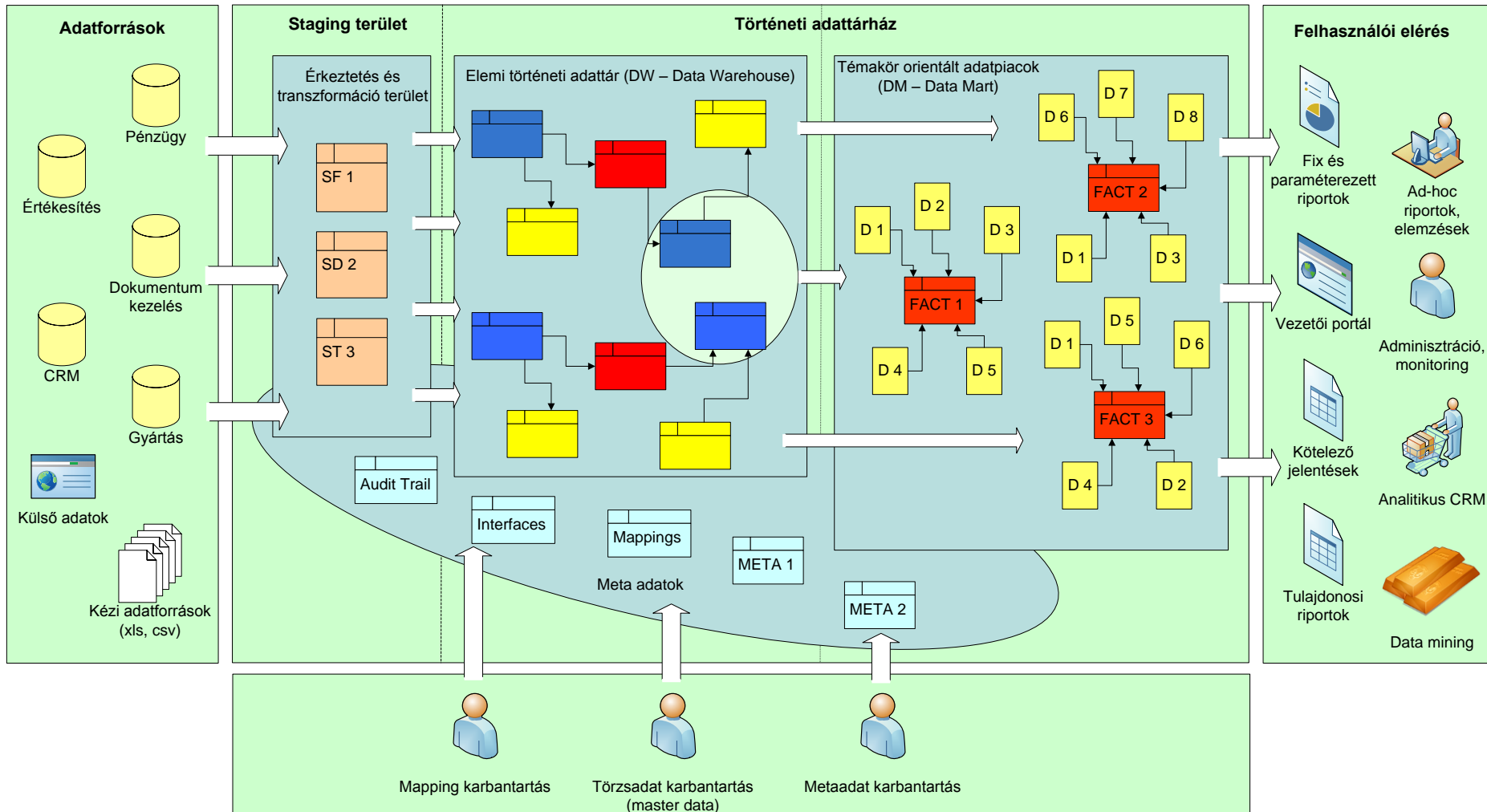
Bevezetési tanulságok

Készítsünk vagy vegyünk?

Összefoglalás

Bevezetés

Az adattárház sematikus felépítése, adatfolyamai



Bevezetés - ETL alapok

Mi az ETL? Honnan lehet szerezni ilyen?

ETL?

- Extract, Transform & Load = kinyerés, átalakítás, betöltés
- ELT?, ELTL?, stb.
- Nemcsak DW, hanem egyéb adatintegrációs feladatoknál is

Alapvető megoldási lehetőségek

- ETL eszköz használata
 - Az összes adatbázis gyártónak van saját ETL eszköze: IBM (DataStage), MS (SSIS), Oracle (OWB, ODI), SAP (BODI), Talend, Pentaho, stb.
 - Specialisták, „csak” ETL eszközt gyártanak: Attunity, Informatica, Wherescape, stb.
- „Kézzel” fejlesztett eljárások (SQL, Java, Perl, C, stb.)

Bevezetés - ETL alapok

ETL funkciók

Adatforrások, célok

- fájl (csv, xml, stb.),
- közvetlen adatbázis kapcsolat (natív, ODBC, JDBC, stb.),
- webservice,
- aszinkron kapcsolatok (messaging, replikáció)

Elérési módok

- Push: adatforrás küldi az adatot, célrendszer „csak” fogadja
- Pull: a célrendszer olvassa ki az adatot, adatforrás „csak” biztosítja

Frissítési módok

- batch, micro-batch, tranzakciós

Bevezetés - ETL alapok

Fejlesztés, üzemeltetés támogatás

Fejlesztés támogatás

- Intuitív felület, gyorsan elsajátítható
- Hatékony fejlesztést tesz lehetővé, moduláris, scriptelhető
- Csoportos fejlesztés támogatása, verziókezelés
- Impact és lineage támogatás

Üzemeltetés támogatás

- Több környezet támogatása (fejlesztői, teszt, éles)
- Folyamatok ütemezése, automatikus és kézi indítás, leállítás
- Monitorozás és statisztikák
pl. mikor, mi futott, milyen eredménnyel

Tartalom

Bevezetés



ETL keretrendszer: elvárások és hogyan készítsük

A mi keretrendszerünk

Bevezetési tanulságok

Készítsünk vagy vegyünk?

Összefoglalás

Az ETL keretrendszerrel szembeni elvárások

Funkcionális elvárások

- Tegye lehetővé a felsorolt ETL funkciók használatát
- Legyen rugalmas, könnyen kialakítható, módosítható, integrálható
- Legyen könnyen vezérelhető, karbantartható, monitorozható
- Legyen gyorsan bevezethető

Technológiai elvárások

- Támogassa mind a kézi, mind az ETL eszközökkel való fejlesztést
- Legyen technológia független
 - Op.rendszer, adatbázis, ETL eszköz, riporting eszköz

Ne kerüljön sokba!

Hogyan készítsünk ETL keretrendszert?

Követelmények pontosítása

- Támogatandó funkciók, környezet, eszközök
- Ne legyenek irreális elvárásaink!

Inkrementális kialakítás

- Pl. inputok érkeztetése, ETL eljárások monitorozása, vezérlés

Metaadat vezérelt megoldás legyen!

Scope- és változáskezelés

- Folyamatosan jönnek az új ötletek, még mi kellene, stb.

Dokumentáljunk

- Az ETL fejlesztőknek és üzemeltetőknek ismerni/használni kell a keretrendszert!

Tartalom

- Bevezetés ✓
- ETL keretrendszer: elvárások és hogyan készítsük ✓
- A mi keretrendszerünk
- Bevezetési tanulságok
- Készítsünk vagy vegyünk?
- Összefoglalás

A mi keretrendszerünk

Felépítése

- 4 modul (Alap, Vezérlés, Monitorozás, Statisztika)

Metaadat vezérelt

- Kb. 30 meta tábla, view-k és tárolt eljárások
- „Standard” függvény és eljárás hívásokkal érhető el

Technológia független

- Élesben működő verziók: Windows és Linux op.rendszer; Microsoft SQL Server és Oracle adatbázis; MS Integration Services, OWB és „kézi” SQL eljárások

A DW folyamatok teljes spektrumát képes lefedni

- Adatkinyeréstől a riport publikálásig
- Tetszés szerinti számú „layer” támogat (pl. forrás → stage → DW → DM → kocka → riport)

A mi keretrendszerünk

„Extrák”

- Lazán csatolt, aszinkron vezérlés
 - Az adatok rendelkezésre állása, a beállított prioritások, az ETL eljárások függőségei (egymástól és adatoktól) alapján
 - Paraméterben állítható párhuzamos futások
- Várható rekordszámok kezelése
 - Az ETL futásoknál várható rekordszámok és a tényleges rekordszámok tárolása, összehasonlítása, elemzése
- Historikus elemzési lehetőségek
 - Pl. mikor mi futott, mennyi ideig, mennyi adatot mozgatott
- Metaadatok karbantartása
 - SQL függvény hívásokkal és/vagy Excel táblázatból
 - „Hátulról” az ETL eszköz és az adatbázis saját metaadataiból

A mi keretrendszerünk ETL metaadatok feltöltése

Microsoft Excel - teszt_meta_data_v2012_12_15.xlsx

	A	B	G	H	I	J	K
1	Csak MAP lehet	Konkrét programnév	sp	A struktúrát azonosító egyedi név. Fájli esetén: a fájl	törzsneve, DBIink esetén: schema.tábla	STAGE, DW, DM, TRANS, stb.	A tábla pontos s
2	p_etl_type in varchar2	p_etl_program_name in varchar2		p_interface_base_name in varchar2		p_source_dw_table_database in va	p_source_dw_table_db_schema in
4	MAP	STAGE.XLD_SYN_BETEXT		SYN_BETEXT			STAGE
5	MAP	STAGE.XLD_SYN_BETEXT_BDF		SYN_BETEXT_BDF			STAGE
6	MAP	STAGE.CDC_SYN_BETEXT					STAGE
7	MAP	STAGE.ACT_SYN_BETEXT_MER					STAGE
8							
9	MAP	STAGE.XLD_SYN_INTEXT		SYN_INTEXT			STAGE
10	MAP	STAGE.XLD_SYN_INTEXT_BDF		SYN_INTEXT_BDF			STAGE
11	MAP	STAGE.CDC_SYN_INTEXT					STAGE
12	MAP	STAGE.ACT_SYN_INTEXT_DEL					STAGE
13	MAP	STAGE.ACT_SYN_INTEXT_INS					STAGE
14							
15	MAP	STAGE.DLD_VI2_VLOGVL		PRBHNGRDTA.VLOGVL			PRBHNGRDTA
16	MAP	STAGE.CDC_VI2_VLOGVL					STAGE
17	MAP	STAGE.CDC_VI2_VLOGVL					STAGE
18	MAP	STAGE.ACT_VI2_VLOGVL_INS					STAGE
19	MAP	REORG_STAGE.MP_VI2_VLOGVL_CNV					STAGE
20	MAP	REORG_STAGE.MP_VI2_VLOGVL_CNV					STAGE
21	MAP	REORG_STAGE.MP_VI2_VLOGVL_CNV					STAGE
22	MAP	REORG_STAGE.MP_VI2_VLOGVL_CNE					REORG_STAGE
23	MAP	REORG_STAGE.MP_VI2_VLOGVL_CNE					REORG_STAGE
24							
25	MAP	STAGE.DLD_INK_CLAIMDECLARATION_ROLLE		DB2ADMIN.CLAIMDECLARATION_ROLLEDCLIENT			DB2ADMIN
26	MAP	STAGE.CDC_INK_CLAIMDECLARATION_ROLLE					STAGE
27	MAP	REORG_STAGE.MP_INK_CLAIMDECLARATION_RO_CNV					STAGE
28	MAP	REORG_STAGE.MP_INK_CLAIMDECLARATION_RO_CNV					STAGE
29	MAP	REORG_STAGE.MP_INK_CLAIMDECLARATION_RO_CNV					STAGE
30	MAP	REORG_STAGE.MP_INK_CLAIMDECLARATION_RO_CNE					REORG_STAGE
31	MAP	REORG_STAGE.MP_INK_CLAIMDECLARATION_RO_CNE					REORG_STAGE
32							

Source Systems | Interfaces | ETL programs (maps, procflows) | Tables | TrqTabs - Map - SrcTabs, Ifaces | Procflow Hierarchy | CONSTANTS

A mi keretrendszerünk

Futási statisztika

Microsoft Excel - statistics_demo_v3_20120227.xlsx

UTEMEZETT_FUTAS_DATUM_ID (mind)
TENYLEGES_FUTAS_DATUM_ID 20120301

SUCCESS_FLAG
1
0
(üres)

SUCCESS_FLAG_RO...
0
1
(üres)

SYSTEM_CODE
AS400
AVAYA
BUSINESS MONI...
CONTRACTS
DKK
DUR
GENESYS
GES

SYSTEM_CODE	SUCCESS_FLAG	Db/DW	TABLE TYPE	Max/LAST_REFRESH	Max / futási idő	Max / idők
AS400	1	253		2012. 3. 1. 4:45	726,07	#SZÁM!
AVAYA	1	4		2012. 3. 1. 1:33	0,50	6:15:15
BUSINESS MONITOR	1	3		2012. 3. 1. 6:04	0,25	18:14:16
CONTRACTS	1	1		2012. 3. 1. 0:26	0,98	0:26:41
DKK	1	6		2012. 3. 1. 0:14	3,63	1:17:12
DUR	1	13		2012. 3. 1. 0:18	5,90	3:08:26
GENESYS	1	23		2012. 3. 1. 0:16	4,07	5:25:13
GES	1	12		2012. 3. 1. 0:07	3,93	1:03:32
GHIA	1	8		2012. 3. 1. 3:07	14,17	0:45:47
GPBROKER	1	6		2012. 3. 1. 3:18	472,07	2:15:39
IBA	1	6		2012. 3. 1. 0:26	1,90	2:33:59
INDEXER	1	3		2012. 3. 1. 0:26	0,48	1:18:02
INKA	1	10		2012. 3. 1. 0:10	2,95	1:32:46
NOTES	1	31		2012. 3. 1. 0:35	1,95	16:17:37
SAP FI	1	6		2012. 3. 1. 4:29	0,45	2:54:18
SHAREPOINT	1	1		2012. 3. 1. 0:09	0,05	0:09:46
SZIKE	1	38		2012. 3. 1. 1:32	1,47	10:22:12
TANTERV	1	7		2012. 3. 1. 0:09	0,37	1:06:50
TRACKING	1	8		2012. 3. 1. 0:12	6,48	0:49:07
ZSERIES	1	88		2012. 3. 1. 10:58	19,45	#SZÁM!
(üres)	1	28		2012. 3. 1. 4:03	48,20	17:52:57

Értékek

Kimutatás mezőlistája

Válassza ki a jelentésbe felvenni kívánt mezőket:

- REPORT_RUNTIME
- DW_TABLE_ID
- ETL_PROGRAM_ID
- SUBJECT_AREA_NAME
- DW_TABLE_SCHEMA
- DW_TABLE_NAME
- DW_TABLE_TYPE
- SYSTEM_CODE
- SOURCE_DW_TABLE_SCHEMA
- SOURCE_DW_TABLE_NAME
- ETL_PROGRAM_NAME
- LOAD_TYPE
- LAST_REFRESH_DATETIME
- DATUM_ID
- UTEMEZETT_FUTAS_DATUM_ID
- TENYLEGES_FUTAS_DATUM_ID
- KEZZEL_KARBANTARTOTT_FLAG
- SCHEDULE_NAME

Húzza a mezőket a lenti területek közé:

Jelentésszűrő

Utemezett...
Tenyleges...

Oszlop címkék

Σ Értékek

Σ Értékek

SYSTEM_CODE
SUCCESS_FLAG

Db/DW_TABL...
Max/LAST_R...
Max / futási idő
Max / idők

Elrendezésfrissítés elh... Frissítés

kimutatás pivot diagram adatok Munka3

Tartalom

- Bevezetés ✓
- ETL keretrendszer: elvárások és hogyan készítsük ✓
- A mi keretrendszerünk ✓
- Bevezetési tanulságok
- Készítsünk vagy vegyünk?
- Összefoglalás

Bevezetési tanulságok

Készüljünk a nem „tisztességesen” viselkedő folyamatokra, ill. fejlesztőkre

- A keretrendszer, az ETL eszköz és az adatbáziskezelő metaadatainak szinkronizálása, keresztellenőrzése

Automatizáljunk, amennyire lehet

- Induló metaadatok feltöltése „hátról” (DB és/vagy ETL)
- ETL eljárások generálása (a megfelelő naplózással – generálás és futás)

Historikus metaadatok betöltése, importálása

- Sok munkát igényel, főleg az adatok konszolidálása, integrálása

Jól kell ismerni a használt adatbázis kezelő és ETL eszköz metaadat lehetőségeit!

Tartalom

- Bevezetés ✓
- ETL keretrendszer: elvárások és hogyan készítsük ✓
- A mi keretrendszerünk ✓
- Bevezetési tanulságok ✓
- Készítsünk vagy vegyünk?
- Összefoglalás

Készítsünk vagy vegyünk?

Vegyünk

- A komoly gyártók ETL eszközei elég jól „felszereltek” adminisztráció, metaadat kezelés, riportok témakörben
- Gyors, nem kell fejleszteni, „csak” integrálni
- Általában az ETL eszköz része (néha feláras kiegészítő)

Készítsünk

- Teljesen a mi igényeinknek megfelelő lesz
- Nem homogén technológiák használata esetén előnyös
- Gyártó, technológia független megoldást szeretnénk (kezeli ha pl. ETL eszközt váltunk)
- Fejlesztést jelent, annak minden kockázatával és költségével

Tartalom

- Bevezetés ✓
- ETL keretrendszer: elvárások és hogyan készítsük ✓
- A mi keretrendszerünk ✓
- Bevezetési tanulságok ✓
- Készítsünk vagy vegyünk? ✓
- Összefoglalás

Összefoglalás

Lehet technológia független ETL keretrendszer építeni

Mielőtt belekezdünk érdemes:

- Kitalálni, hogy mire akarjuk használni
- Megnézni az ETL eszközünk képességeit
- Körülnézni a piacon hasonló megoldások után

Nekünk sikerült és nem bántuk meg 😊

Kérdések?

gabor.gollnhofer@meta4consulting.eu

blog.adatmodellezes.hu

