

Data Integrátorok a gyakorlatban – Oracle DI vs. Pentaho DI

Fekszi Csaba – Ügyvezető

Vinnai Péter – Adattárház fejlesztő

2013. február 20.

1

Pentaho eszköztára – Data Integrator

2

Spoon felület

3

Spoon program elemek

4

Hagyományos ETL - alap komponensekből

5

ODI Architektúrája és működési elve

6

Oracle Data Integrator ETL vs ELT

7

ODI tervezési szemlélete

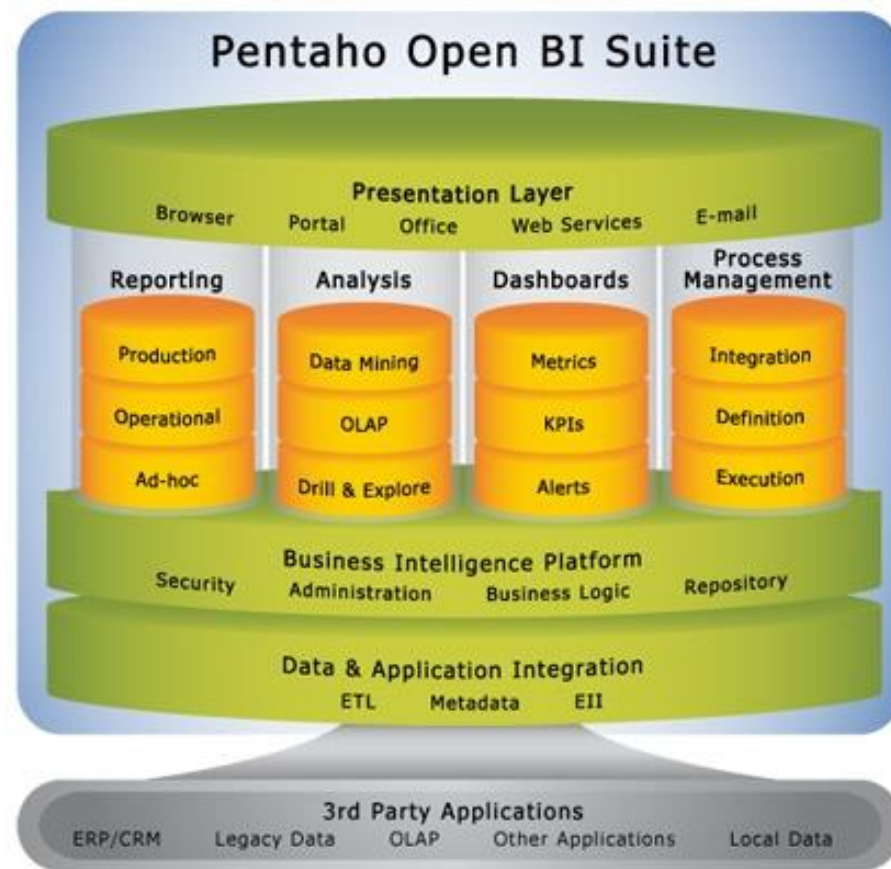
8

Knowledge modulok - Adatfolyamok

A Pentaho eszköztára

- ETL
- Pixelhelyes riportok
- Önkiszolgáló riportolás
- Analízis
- Műszerfalak
- Adatbányászat
- Big Data
- ...

Funkció	Tervező eszköz	Megjelenítő eszköz
OLAP elemzések	Schema Workbench	Analysis View, Saiku, OpenI
Ad-hoc riportok	Metadata Editor	Ad-hoc Reporting, Saiku
Pixelhelyes riportok	Report Designer	Riportok
ETL	Spoon	Kitchen, Pan, Carte



Pentaho Data Integrator bemutatása

- Legfontosabb előnyök:
 - könnyű telepíthetőség
 - platform függetlenség (100 % Java – Windows, Linux,...)
 - egyszerű használhatóság (grafikus felület, specifikus ismereteket nem igényel)
 - integrálhatóság
 - nagyfokú teljesítmény és skálázhatóság
- PDI komponensei:
 - Spoon: grafikus felületű transzformáció és job készítés
 - Kitchen: parancssori job futtatás
 - Pan: parancssori transzformáció futtatás
 - Carte: egyszerű web szerver transzformációk távoli futtatása céljából
- Repository:
 - File alapú
 - Adatbázisban

Spoon felület

- Menü
- Perspektívák
- Paletta
- Grafikus munkaterület
- Eszköztár
- Futási naplók
- Step metrika

Spoon - [EE Repository] Sample Transformation v1.20

Perspective: Data Integration

Steps

Input

- Access Input
- CSV file input
- Data Grid
- De-serialize from file
- ESRI Shapefile Reader
- Excel Input
- Fixed file input
- Generate random value
- Generate Rows
- Get data from XML
- Get File Names
- Get Files Rows Count
- Get SubFolder names
- Get System Info
- Google Analytics Input
- Google Docs Input
- LDAP Input
- LDIF Input
- Mondrian Input
- OLAP Input
- Property Input
- RSS Input
- S3 CSV Input
- Salesforce Input
- SAP Input
- Table input
- Text file input
- XBase input

Output

Transform

Utility

Flow

Scripting

Lookup

Joins

Data Warehouse

Validation

Statistics

Sample Transformation

Read Sales Data

Filter Missing Zips

Read Postal Codes

Lookup Missing Zips

Value Mapper

Prepare Field Layout

Select values

Number range

Write to Database

To test this transformation, you will need to:

- Make sure the Hypersonic sample database is running (.\pdi-ee\data-integration-server\data\start_hypersonic.bat)
- Open the Table Output step and click the SQL button to create the target output table

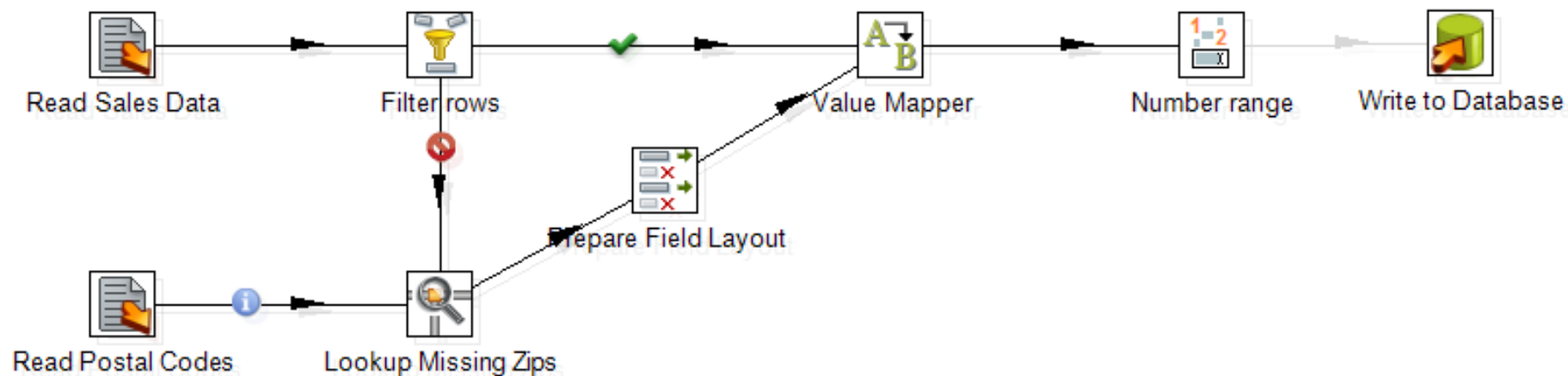
Execution Results

Execution History | Logging | Step Metrics | Performance Graph

Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	inp
1 Filter Missing Zips	0	2823	2823	0	0	0	0	0	0	Finished	0.5	6019.1
2 Lookup Missing Zips	0	21455	76	0	0	0	0	0	0	Finished	0.9	24520.0
3 Read Postal Codes	0	0	21379	21380	0	1	0	0	0	Finished	0.7	31815.4
4 Prepare Field Layout	0	76	76	0	0	0	0	0	0	Finished	0.9	85.2
5 Value Mapper	0	2823	2823	0	0	0	0	0	0	Finished	0.9	3112.4
6 Read Sales Data	0	0	2823	2824	0	1	0	0	0	Finished	0.3	8209.3
7 Select values	0	2823	2823	0	0	0	0	0	0	Finished	0.9	3112.4
8 Number range	0	2823	2823	0	0	0	0	0	0	Finished	0.9	3061.8
9 Write to Database	0	2823	2823	0	2823	0	0	0	0	Finished	1.1	2543.2

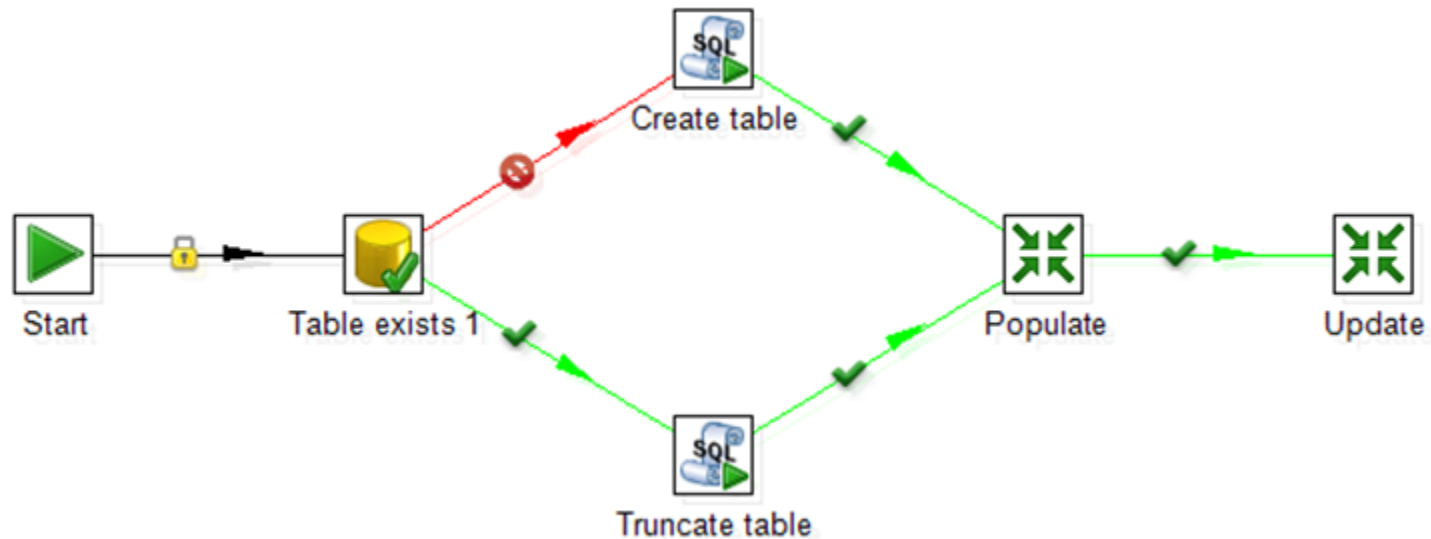
Spoon program elemek - Transzformáció

- Transzformáció (Kettle Transformation - kiterjesztés: ktr):
 - Transzformációs lépés/elem (Step)
 - „Adatfolyam”(Hop)



Spoon program elemek - Job

- Munkafolyamat (Kettle Job - kiterjesztés: kjb):
 - Elem (Job entry)
 - „Adatfolyam” (Job hop)



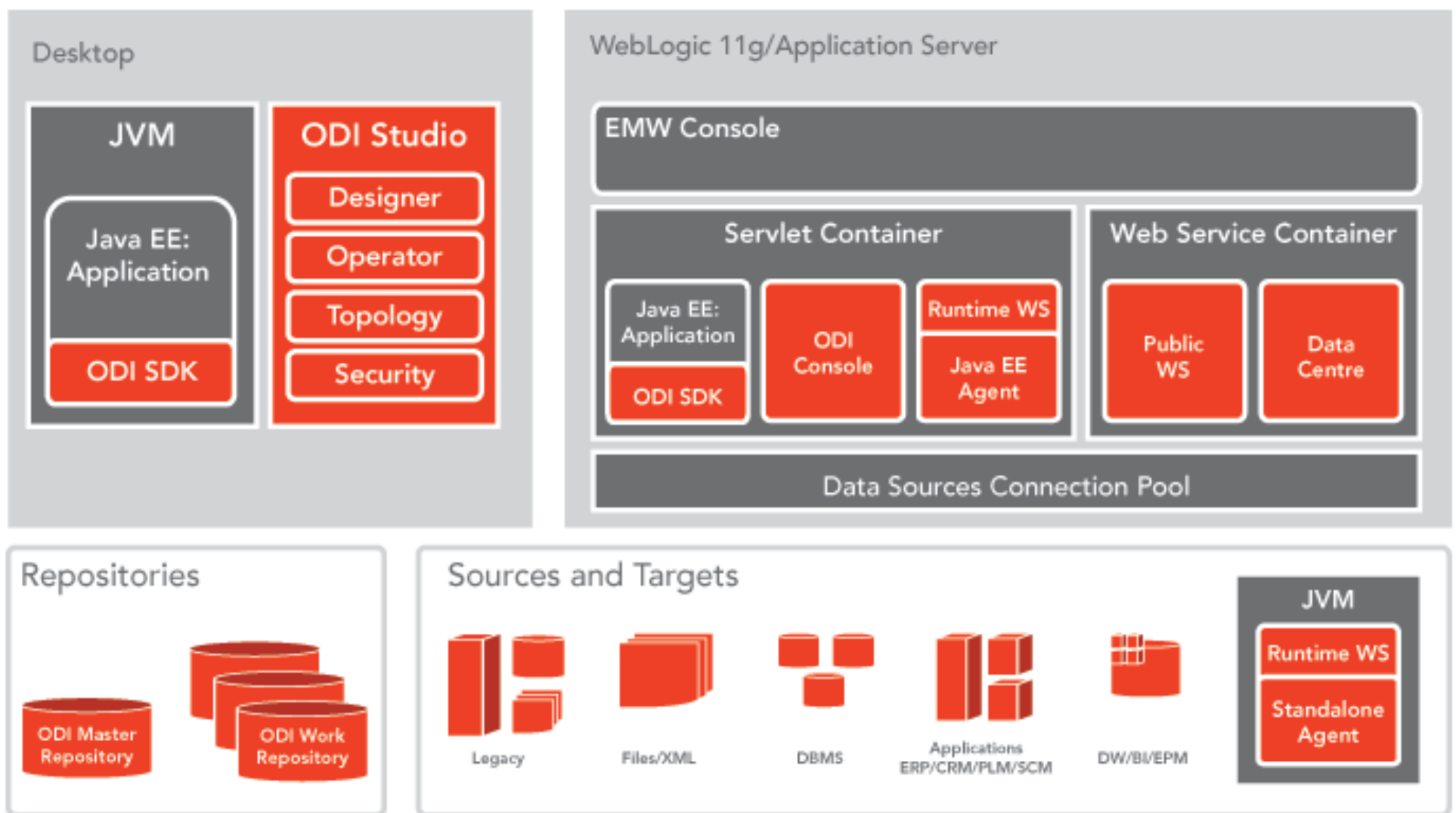
▶ Alapvető transzformációs komponensek

- Input, Output elemek (file, tábla, Excel, OLAP, LDAP, SAP)
- Adatok (sorok) generálása (Generate rows):
- Számítás (Calculator)
- Szűrés (Filter rows)
- Csoport művelet (Group by / Memory Group by)
- Sorba rendezés (Sort rows)
- Naplózás (Write to log)
- ...

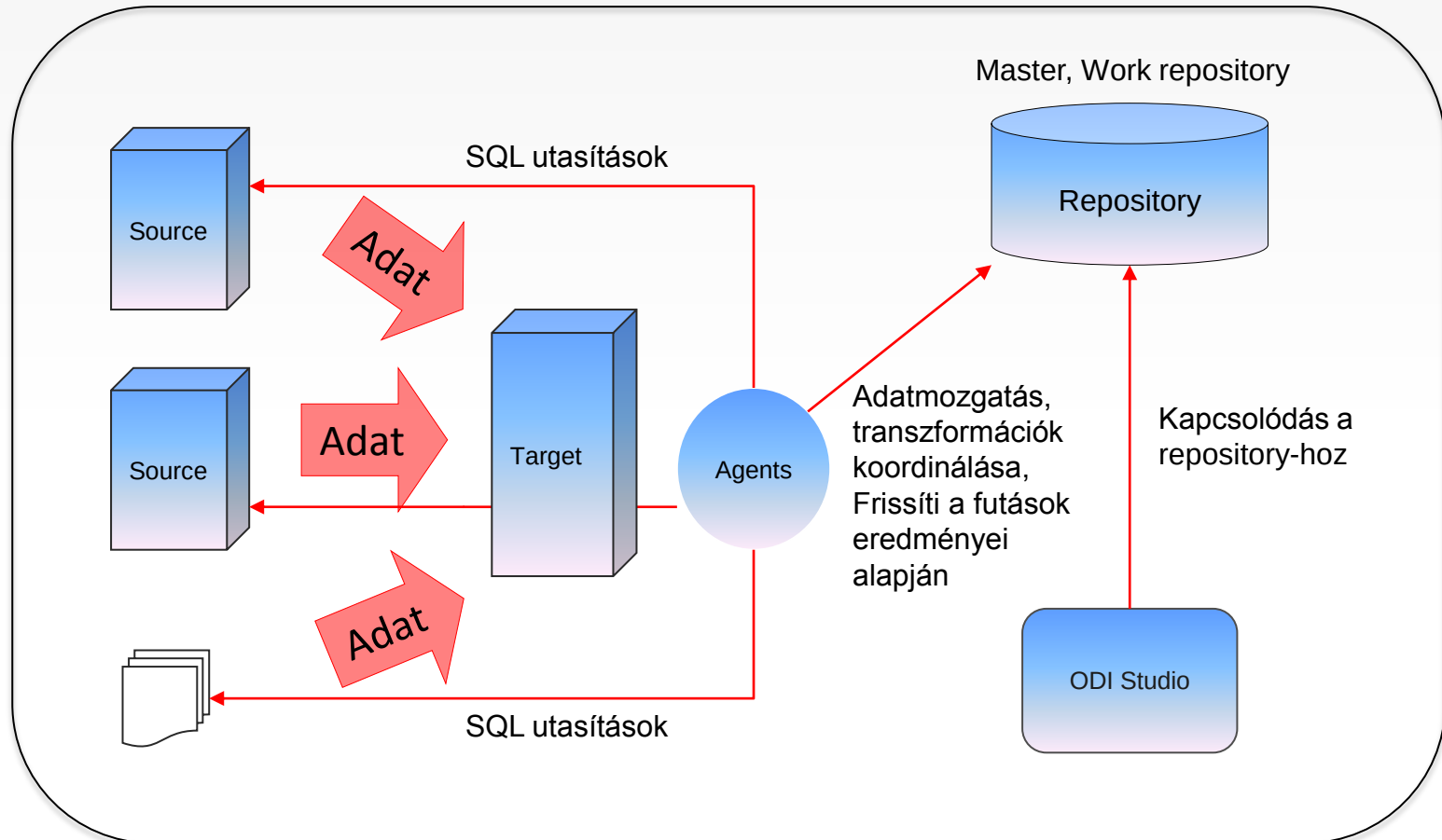
Alapvető job komponensek

- Folyamatvezérlési elemek (Start, Success, Abort, Job)
- Transzformációk hívása (Transformation)
- Feltételek (Conditions) (pl file exist, folder empty, check DB connection)
- File műveletek
- ...

ODI - Architektúra



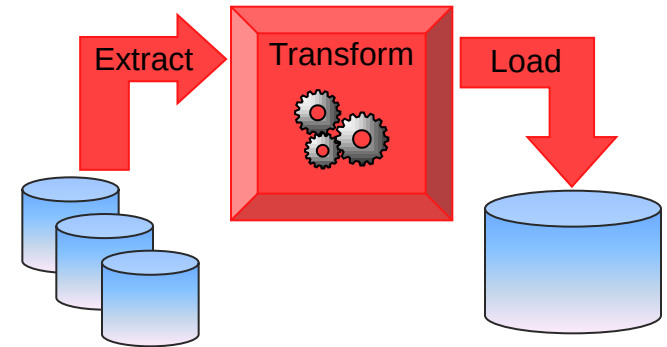
ODI – Működési elve



Oracle Data Integrátor: ETL vs E-LT

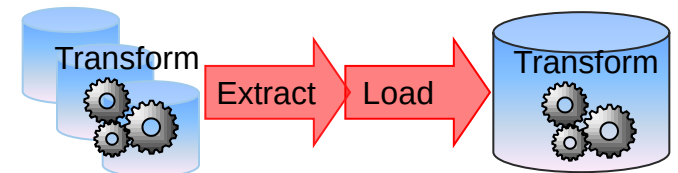
- Transzformáció külön ETL serveren
 - Gyenge teljesítmény
 - Magas költségek
- Transzformáció meglévő RDBMS-n:
 - Erőforrás jobb kihasználása
 - Hatékonyság
 - Nagy teljesítmény

Hagyományos ETL Architektúra



E-LT Architektúra

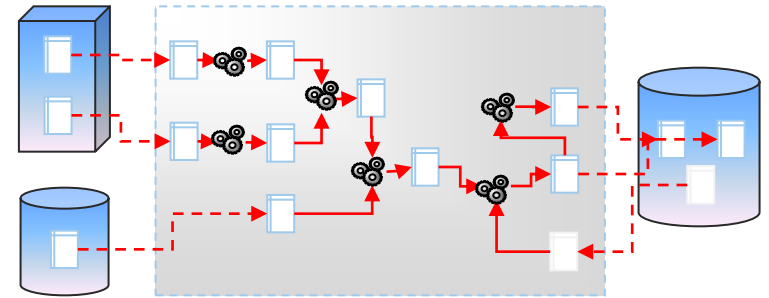
“E-LT”



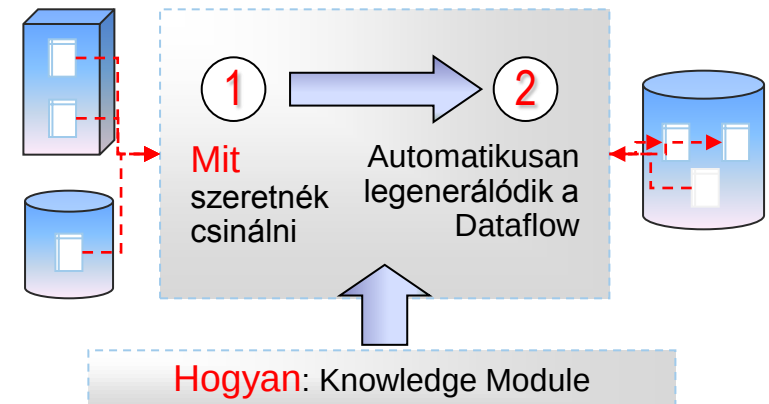
Tervezési szemlélet

- ETL Dataflow lépéseinek meghatározása:
 - minden egyes lépésénél a komplex ETL folyamat logikájának ismerete szükséges
 - speciális ETL ismeret
 - karbantartási nehézség
- Deklaratív alapú tervezés:
 - Egyszerűbb és kevesebb lépésszám
 - Automatikus Data Flow generálás, függetlenül a források és a cél platformtól
 - Rövidebb végrehajtási idő
 - Jelentősen csökkenti a tanulási időt

Hagyományos ETL tervezés



ODI deklaratív tervezés



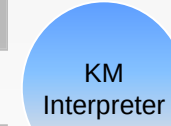
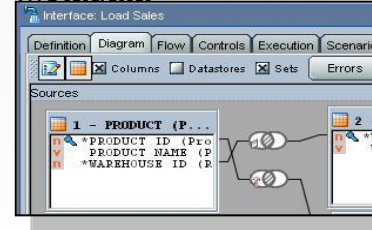
Knowledge Modulok (KM)

- A forrás és cél közötti elemi lépéseket tartalmazza
- Testreszabható template-k
- Több mint 120 előre definiált KM

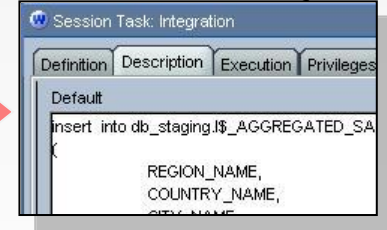
KM's Meta Code

```
insert into <%=snpRef.getTable(
    <%=snpRef.getColList("
)
select <%=snpRef.getPop("DISTI
    <%=snpRef.getColList("
<%if (snpRef.getPop("HAS_JRN").
```

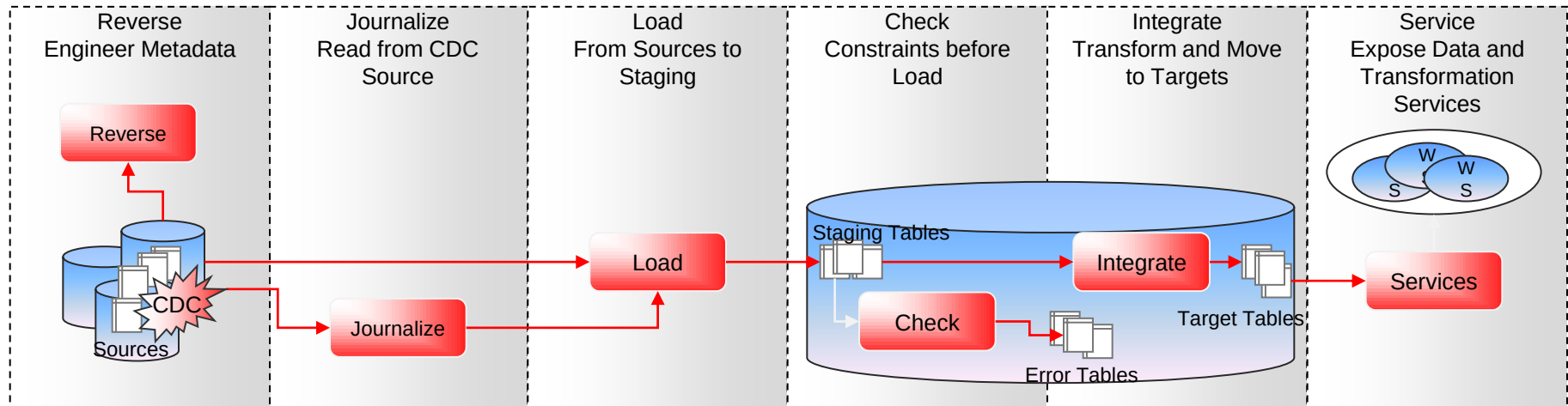
Metadata



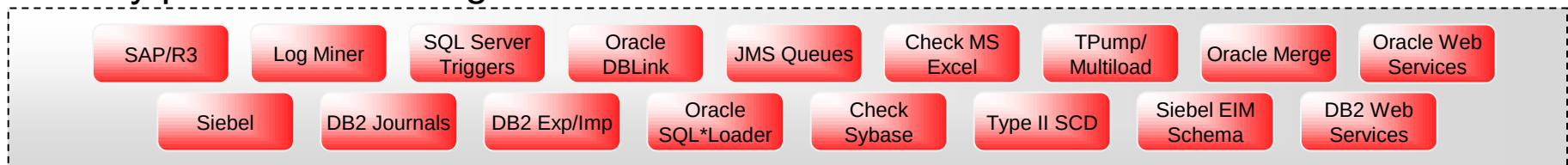
Executed Code - SQL



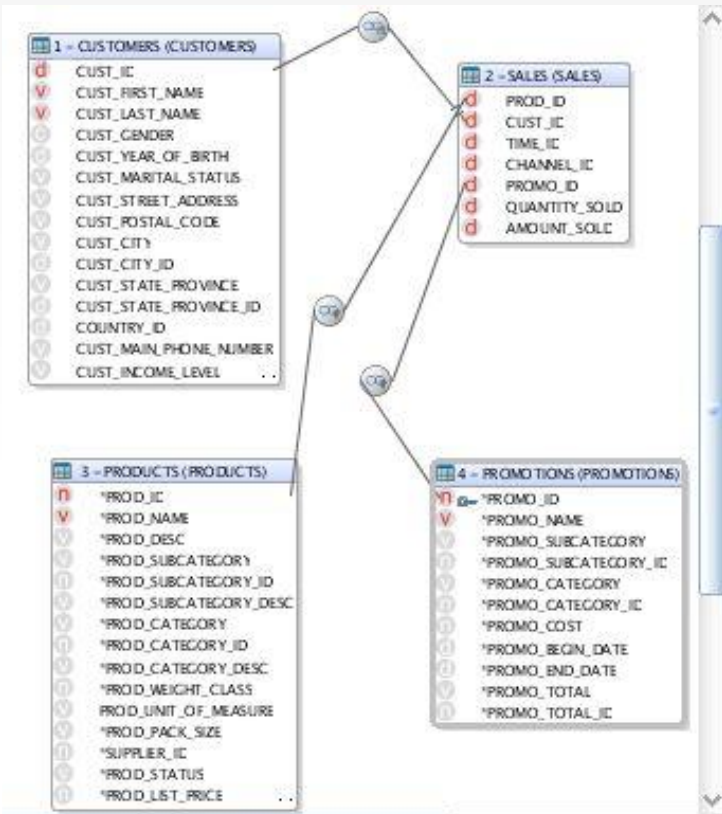
Knowledge Modules Architecture



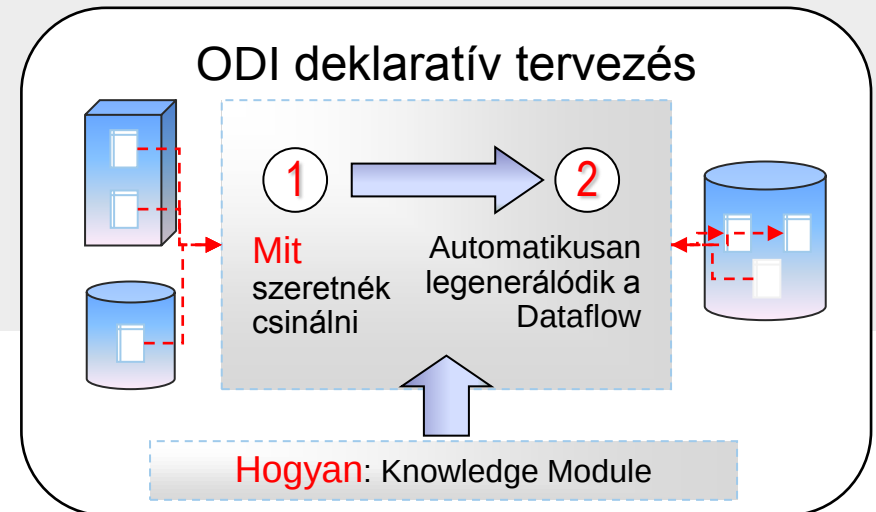
Néhány példa a Knowledge Modules-ra



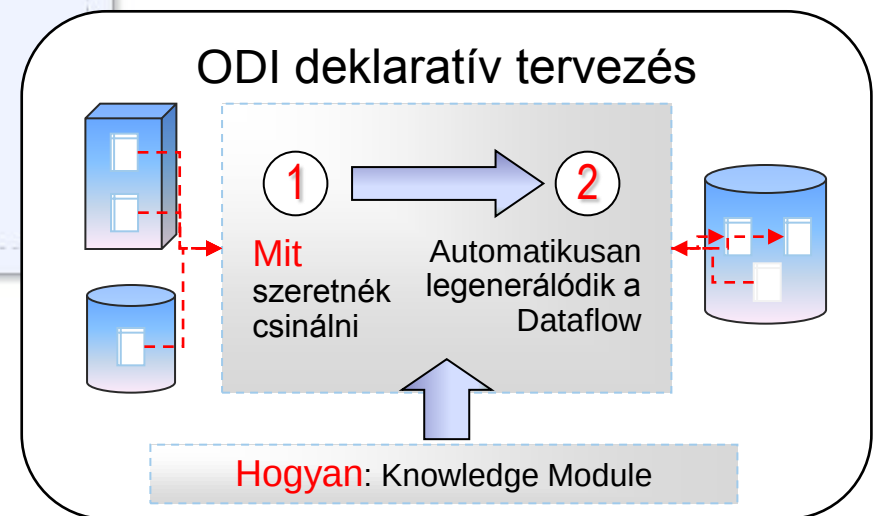
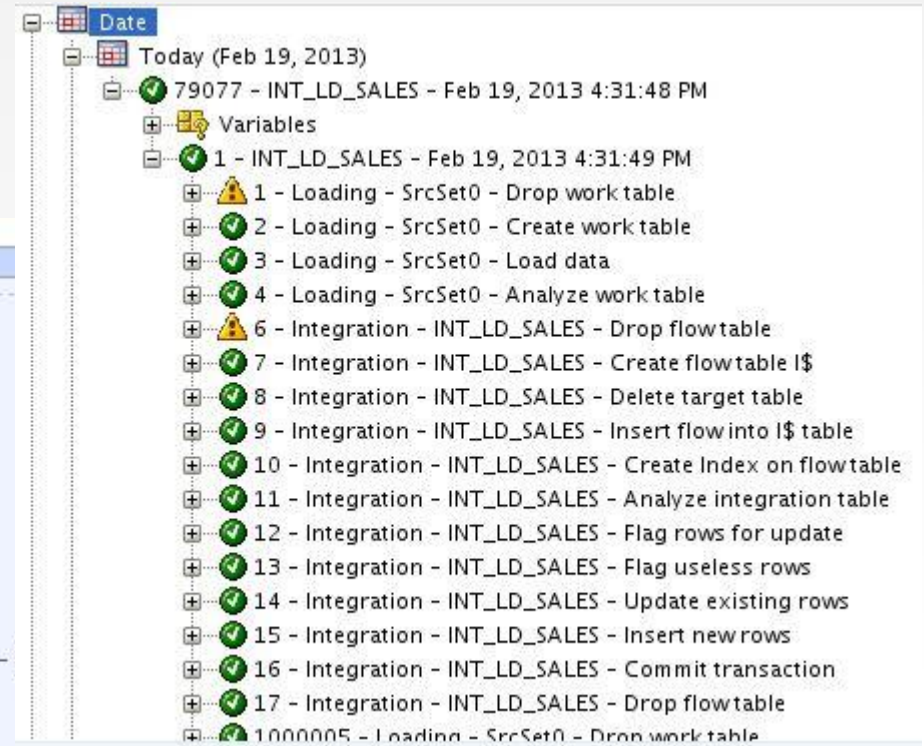
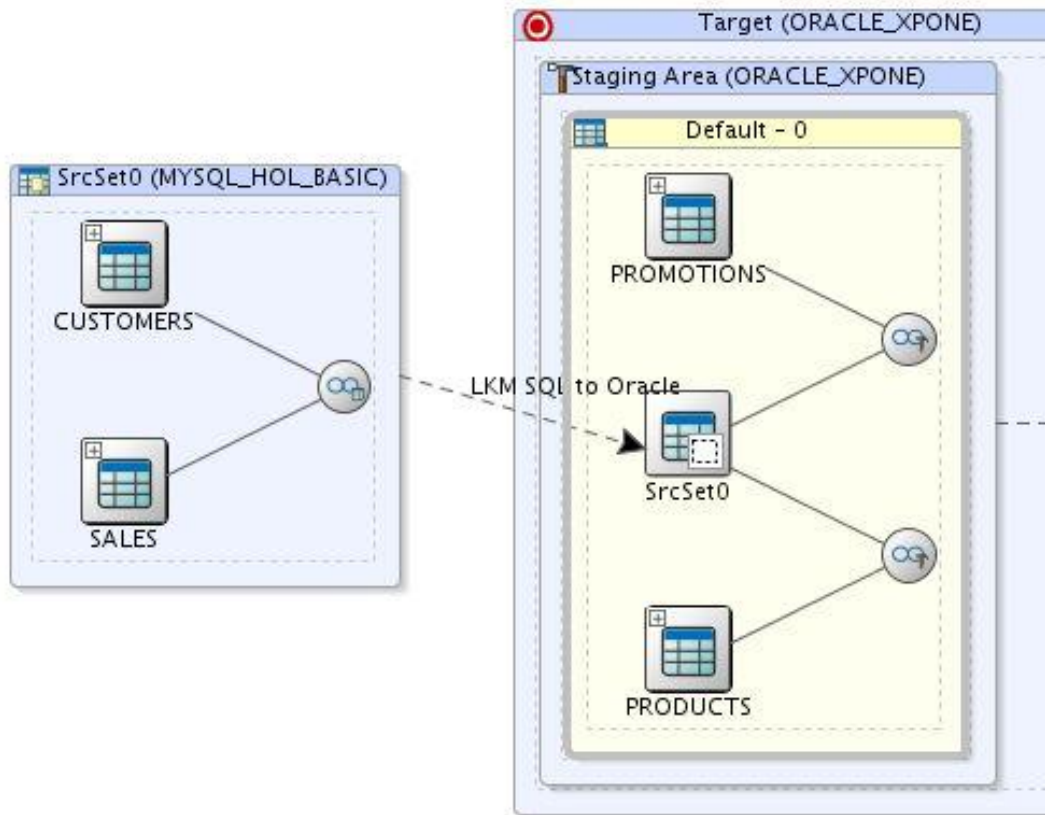
ODI – Mapping – Mit?



Target Datastore - SALES			
Position	Indicators	Name	Mapping
1		*PROD_ID	SALES.PROD_ID
2		PRODUCT	PRODUCTS.PROD_NAME
3		*CUST_ID	CUSTOMERS.CUST_ID
4		CUSTOMER	CUSTOMERS.CUST_FIRST_NAME ' ' CUSTOMERS.CUST_LAST_NAME
5		*TIME_ID	SALES.TIME_ID
6		*CHANNEL_ID	SALES.CHANNEL_ID
7		*PROMO_ID	SALES.PROMO_ID
8		PROMOTION	PROMOTIONS.PROMO_NAME
9		*QUANTITY_SOLD	SALES.QUANTITY_SOLD
10		*AMOUNT_SOLD	SALES.AMOUNT_SOLD



ODI – Dataflow – Hogyan?





▶ Köszönöm a figyelmet!

