

Inhoudstafel

1. Versiebeheer	1
2. Voorwoord	1
3. Open source ETL tool vergelijking.....	2
3.1 pentaho data integration (Kettle – engine based).....	2
3.1.1 Sterke punten.....	2
3.1.2 Zwakke punten	2
3.2 Talend (code generator)	2
3.2.1 Sterke punten.....	2
3.2.2 Zwakke punten	3
3.3 Cloveretl (engine based).....	3
3.3.1 Sterke punten.....	3
3.3.2 Zwakke punten	3
3.4 Ketl (engine based).....	3
3.4.1 Sterke punten.....	3
3.4.2 Zwakke punten	3

1. VERSIEBEHEER

Versienummer	Datum	Auteur
1.0	20 maart 2012	Kim Van Gossum (Blue Bay)

2. VOORWOORD

Dit document beschrijft de sterke en zwakke punten van een aantal open source ETL tools die op dit moment op de markt aanwezig zijn. Hieronder verstaan we tools waarvan de broncode openbaar en beschikbaar is voor download en waaraan extra ontwikkelingen (functionaliteit) kunnen toegevoegd worden.

Van elke hieronder beschreven tool bestaat een forum waarop gebruikers en ontwikkelaars met mekaar in contact kunnen komen en waarvan updates en functionele uitbreidingen vrij toegankelijk zijn.

We beperken ons ook tot de versies die volledig vrij zijn van licensiekost. Een aantal leveranciers maken namelijk het onderscheid tussen “Community” en “Enterprise” versies. De eerste is volledig openbaar en gratis verkrijgbaar, de tweede is functioneel uitgebreider, men kan zich beroepen op professionele 1 op 1 support, maar deze versies zijn betalend. Deze analyse richt zich naar de “Community” versies.

Elk van deze tools werd door onze consultants gedownload, geïnstalleerd en vervolgens aan een aantal technische tests onderworpen. Er werden telkens een aantal extracties ontwikkeld op verschillende databank schema's, gebruik makende van verschillende types van bronnen en specifieke transformaties uit de data warehousing wereld (als mappingen voor SCD type 2).

Alle tools werden aan de hand van een uitgebreide functionele en technische vragenlijst getest. Hieronder vindt men een samenvatting terug, gebaseerd op de technische en functionele detailanalyse.

3. OPEN SOURCE ETL TOOL VERGELIJKING

3.1 PENTAHO DATA INTEGRATION (KETTLE – ENGINE BASED))

3.1.1 STERKE PUNTEN

- Databank connectiviteit:
Er is een uitgebreide lijst met mogelijke connecties en er is ook de mogelijkheid om te connecteren naar SAP ERP systeem.
Voorbeeld enkele connecties : AS/400, DB2,MS SQL Server, MySql, Oracle, PostgreSQL, Sybase, Teradata,
- Bevat een grote lijst met kant-en-klare standaard functies en transformaties.
- Mogelijkheid om te schrijven naar databanken, Excel, Access, Properties File, text file, xml file
- Visueel sterke tool, intuïtieve en gebruiksvriendelijke opmaak. Het bevat een zeer werkbare drag and drop interface.
- Snel en makkelijk te installeren en apart installeerbaar zonder andere componenten
- Ontwikkeling door meerdere concurrent developers dankzij gebruik van een repository.
- Bulk Loading:
Standaard voorzien zijn Teradata fastload bulk loader, PostgreSQL Bulk Loader, Oracle Bulk Loader, MySQL Bulk Loader , MonetDB Bulk Loader, LucidDB Streaming Loader, Ingres VectorWise Bulk Loader, Infobright loader, greenplum load, greenplum bulk loader.

3.1.2 ZWAKKE PUNTEN

- Geen SQL query builder
- Onvoldoende debug functionaliteit
- Code kan niet herbruikt worden in andere componenten (dus soms dubbele code)
- Beperkt in geheugen gebruik tijdens lookup functies (fout java heap space)

3.2 TALEND (CODE GENERATOR)

3.2.1 STERKE PUNTEN

- Databank connectiviteit:
Er is een grote lijst met mogelijke connecties en er is ook de mogelijkheid om te connecteren naar SAP ERP systeem en enkele CRM systemen (Centric, Microsoft, Sugar, Vtiger).
Voorbeeld enkele connecties : AS/400, Access, DB2,MS SQL Server, MySql, Oracle, PostgreSQL, Sybase, Teradata,
- Bevat een grote lijst met kant-en-klare standaard functies en transformaties.
- Mogelijkheid om te schrijven naar databanken, Excel, Properties File, text file, xml file, Sap, enkele crm systemen, Palo Olap cube
- Visueel sterke tool, intuïtieve en gebruiksvriendelijke opmaak, gebaseerd op

- Eclipse.
- Bevat ook een uitgebreide debug functie.
- Groot voordeel van deze tool is dat je code kunt herbruiken.
- Snel en makkelijk te installeren.
- Bulk loading:
 - Wordt ondersteund voor verschillende databases zoals MySQL, oracle, PostgreSQL, Sybase,...

3.2.2 ZWAKKE PUNTEN

- Om connecties naar databases te maken heb je telkens een bijkomende JDBC driver nodig

3.3 CLOVERETL (ENGINE BASED)

3.3.1 STERKE PUNTEN

- Databank connectiviteit:
 - Bevat een beperkte lijst van connecties namelijk : Derby, Firebird, Microsoft SQL Server, MySQL, Oracle, PostgreSQL, SQLite, Sybase
- Mogelijkheid om te schrijven naar databanken, Excel, flat files
- Visueel sterke tool, intuïtieve en gebruiksvriendelijke werking, gebaseerd op Eclipse.
- Bevat ook een uitgebreide debug functie.
- Snel en makkelijk te installeren.

3.3.2 ZWAKKE PUNTEN

- Bevat maar enkele kant-en-klare standaard functies en transformaties.
- Community editie is zeer beperkt.
 - Voorbeelden van wat niet beschikbaar is in de community editie:
 - export functies
 - lookups (enkel een simpele lookup is beschikbaar)
 - Je kan zelf geen database drivers toevoegen , je kan enkel de fixed database drivers gebruiken
- Je kan slechts 20 componenten toevoegen per ETL graph
- Bulk loading : Niet voorzien.

3.4 KETL (ENGINE BASED)

3.4.1 STERKE PUNTEN

- Geen noemenswaardige

3.4.2 ZWAKKE PUNTEN

- Databank connectiviteit:
 - Bevat een beperkte lijst van connecties namelijk : MySQL, Oracle, PostgreSQL
- Bevat geen kant-en-klare standaard functies en transformaties
- Geen visuele ontwerpmodule, geen gebruiksvriendelijke layout. Alles wordt uitgevoerd via de command line.
- Bulk loading : Niet voorzien.