



Ferramenta Open Source de DW: Talend

Thalles Cezar Rodrigues de Lima

Introdução

- Ferramentas Open Source
 - Código aberto sob uma licença livre
 - Contras?
 - Apenas uma “alternativa barata”
 - “Inseguro”?
 - Pros
 - Uma chance de experimentar o produto
 - Acelera o desenvolvimento

Introdução

- Ferramentas Open Source de DW
 - Data Quality
 - Business Intelligence (BI)
 - Master Data Management (MDM)
 - Integração de dados

Integração de dados

- Etapa fundamental em um DW
 - “A data warehouse is a subject-oriented, ***integrated***, time-variant and non-volatile collection of data in support of management's decision making process.”
(W. Inmon)
- Combinação de processos técnicos e de negócios para combinar dados de diversas fontes.

Integração de dados

- Magic Quadrant for Data Integration 2016

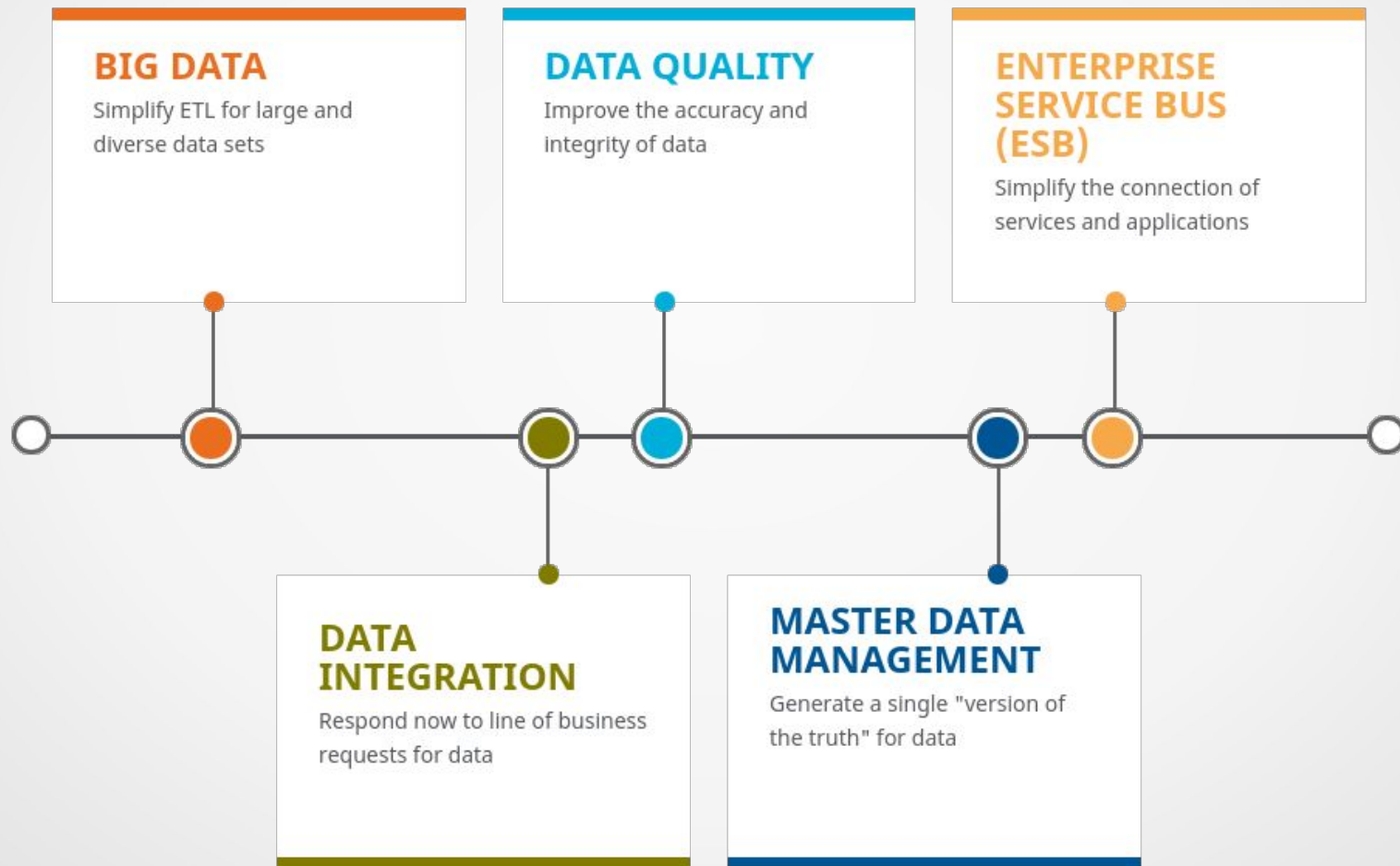


Talend

- Uma das líderes em DI
- Open Source
 - Desde que surgiu, em 2005, vem apostando na estratégia de código aberto;
 - Visionária no MQ até 2015



Talend Open Studio



Talend Open Studio

- Talend Open Studio for Data Integration
 - Principal produto da Talend, lançado em 2006;
 - Ambiente de desenvolvimento baseado em Eclipse;
 - Mais de 900 conectores para definir processos de integração complexos;
 - Grande comunidade de usuários ativos.

Talend Open Studio for DI

- GUI

The screenshot displays the Talend Open Studio for DI interface. The main workspace shows a job design for 'Job tXMLMap_S05 0.1'. The design includes a 'Customers' component connected to a 'tXMLMap_1' component, which is then connected to a 'tLogRow_1' component. A 'US Sites' component is also visible, connected to the 'tXMLMap_1' component. The job design is shown in a 'Designer' view, with a 'Code' view also available.

The 'Job tXMLMap_S05' execution log is visible in the bottom right pane. The log shows the following output:

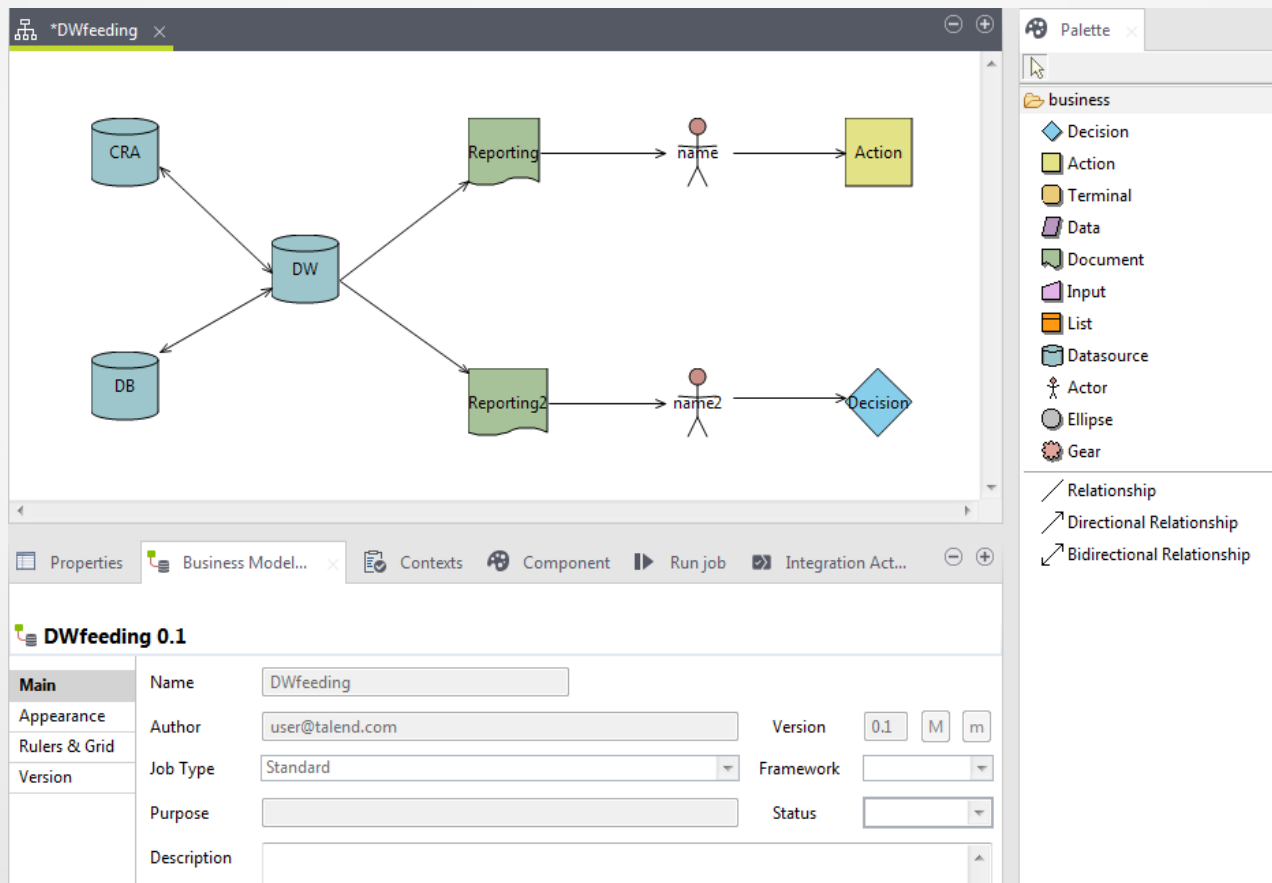
```
Starting job tXMLMap_S05 at 09:47 24/11/2014.
[statistics] connecting to socket on port 3570
[statistics] connected
<?xml version="1.0" encoding="UTF-8"?>
<customers><Customer><LabelState>Connecticut</LabelState><Name><id>1</id><CustomerName>Griffith
Paving and Sealcoat</CustomerName><Name><id>56</id><CustomerName>Glenn Oaks Office
Supplies</CustomerName><Name></Customer><Customer><LabelState>Ohio</LabelState><Name><id>2</id>
<CustomerName>Bill's Dive Shop</CustomerName><Name><id>61</id><CustomerName>DBN
Bank</CustomerName><Name></Customer><Customer><LabelState><Name><id>62</id><CustomerName>BBQ
Smith's Tex
Mex</CustomerName><Name></Customer><Customer><LabelState>Florida</LabelState><Name><id>63</id>
<CustomerName>Pivot Point College</CustomerName><Name></Customer></customers>
[statistics] disconnected
Job tXMLMap_S05 ended at 09:47 24/11/2014. [exit code=0]
```

Talend Open Studio for DI

- Criação de modelos de negócios
 - É possível representar graficamente as necessidades de implementação do projeto.
 - Modelos de negócio ajudam os operadores de TI a entender as necessidades do projeto e transformá-las em tarefas (Jobs).

Talend Open Studio for DI

- Criação de modelos de negócios

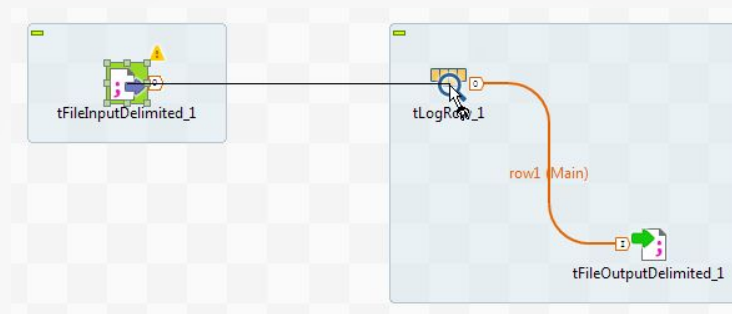


Talend Open Studio for DI

- Criação de tarefas (jobs)
 - Um job design é a camada executável de um modelo de negócios.
 - Composto por um ou mais componentes conectados, que permite a configuração e execução de processos.
 - Traduzido em código, rotinas e programas.

Talend Open Studio for DI

- Criação de tarefas (jobs)



The screenshot shows the Talend Open Studio interface with the 'Job A_Basic_Job' window open. The 'Execution' tab is active, displaying the following log output:

```
Starting job A_Basic_Job at 11:37 21/06/2015.
[statistics] connecting to socket on port 3648
[statistics] connected
-----
tLogRow_1
-----
id|CustomerName|CustomerAddress
-----
1|Griffith Paving and Sealcoat|talend@apres91
2|Bill's Dive Shop|511 Maple Ave. Apt. 1B
3|Childress Child Day Care|662 Lyons Circle
4|Facelift Kitchen and Bath|220 Vine Ave.
5|Terrinni & Son Auto and Truck|770 Exmoor Rd.
6|Kermit the Pet Shop|1860 Parkside Ln.
7|Tub's Furniture Store|807 Old Trail Rd.
8|Toggle & Myerson Ltd|618 Sheridan rd.
9|Childress Child Day Care|788 Tennyson Ave.
10|Elle Hypnosis and Therapy Cent|2032 Northbrook Ct.
-----
[statistics] disconnected
Job A_Basic_Job ended at 11:37 21/06/2015. [exit code=0]
```

At the bottom of the execution window, there are checkboxes for 'Line limit' (set to 100) and 'Wrap' (checked).

Talend Open Studio for DI

- Documentação das tarefas
 - Documentação detalhada em HTML de tarefas;
 - Propriedades do projeto que inclui a tarefa;
 - Detalhes e configurações das tarefas, além de suas representações gráficas;
 - Open Studio: documentação sob demanda

Talend Open Studio for DI

- Extract, Load and Transform (ELT)
 - É possível usar templates SQL, com *queries* pre-definidas;
 - Diferente de ETL, no modo ELT o dado é transformado depois de ser carregado;
 - O dado é migrado em *bulk* e depois transformado no DBMS de destino, reduzindo a carga na rede e trazendo um *throughput* maior;
 - Como SQL não é tão poderoso como java, as transformações são mais limitadas do que no ETL.

Avaliação Gartner

- Pontos fortes
 - Modelo de custo e flexibilidade;
 - Portfólio de produtos para DI e tecnologias relacionadas;
 - Robustez em operações de DI e compromisso com tendências do mercado.

Avaliação Gartner

- Cuidados
 - Estabilidade do produto, suporte e disponibilidade de parcerias;
 - Extensão de funcionalidades – existem outros estilos de integração de dados;

Conclusão

A ferramenta da Talend oferece uma grande liberdade de operações sobre Data Warehouse, tanto para quem não tem muito conhecimento de TI quanto para quem prefere customizar os códigos para realizar operações mais complexas.

A ascensão da Talend no *Magic Quadrant for DI* mostra que investir no modelo de Código Aberto foi uma escolha de sucesso e que pode ser seguida por outras empresas.

Referências

1. Talend Data Integration Software
2. Magic Quadrant for Data Integration Tools
3. Talend's Journey in the Gartner Magic Quadrant
4. Talend Open Studio for Data Integration – User Guide
5. Open Source DI Tool Comparison - Excella