



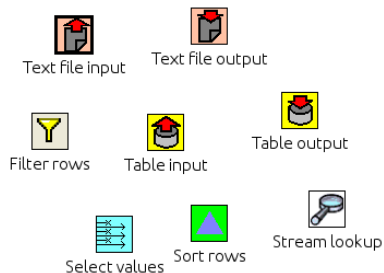
The story behind the ugly icons

From Kettle to PDI

May 13th 2014

Matt Casters

Chief Architect Data Integration at Pentaho



Feel free to

BOO!

Before the code

What I used to do 15 years ago (1999-2003)

Data Integration to enable Business Intelligence

Front-ends:

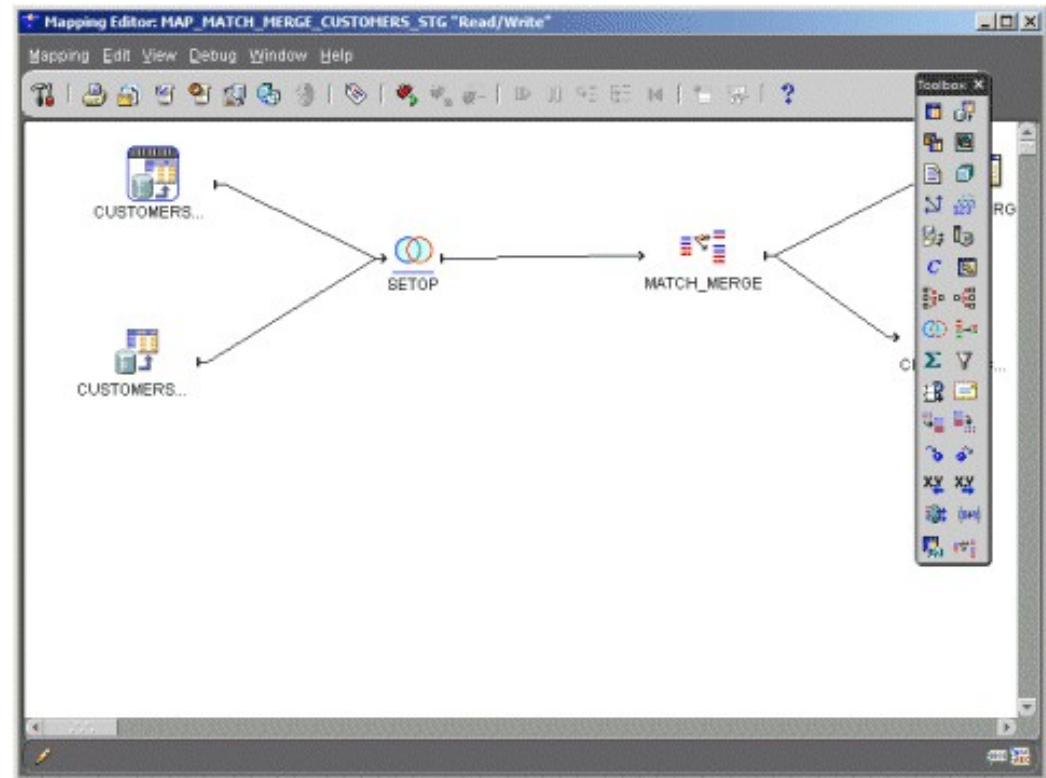
- Platinum InfoBeacon
- Business Objects
- Excel

Databases:

- Oracle
- SQL Server
- DB2, Informix, MS Access

Data Integration:

- Hand-coding C/SQL
- Data Mirror
- Oracle Datamart/Warehouse Builder
- MS DTS



What was wrong with the software?

- **Price: high \$ £ € \$B**
- **Hard to install**
- **Missing core functionality like:**
 - **Copy/Paste**
 - **Very limited data routing**
 - **Multi-data source**
- **Hard to debug**
 - **Code generation: T/SQL, PL/SQL, ...**
 - **No debugging nor preview of data**
- **Hard to maintain**
 - **Stupid naming limitations**
 - **Tedious source-to-target mapping**
- **Metadata not open, hard to read/write**

The resulting design goals?

- Open source with an Free license
- Open, readable XML format
- Open, readable relational repository format
- Open API
- Easy to set up (<2 minutes)
- Open to all kinds of databases
- Easy-to-use GUI
- Easy to pass data around
- Easy to convert data from/to any possible format

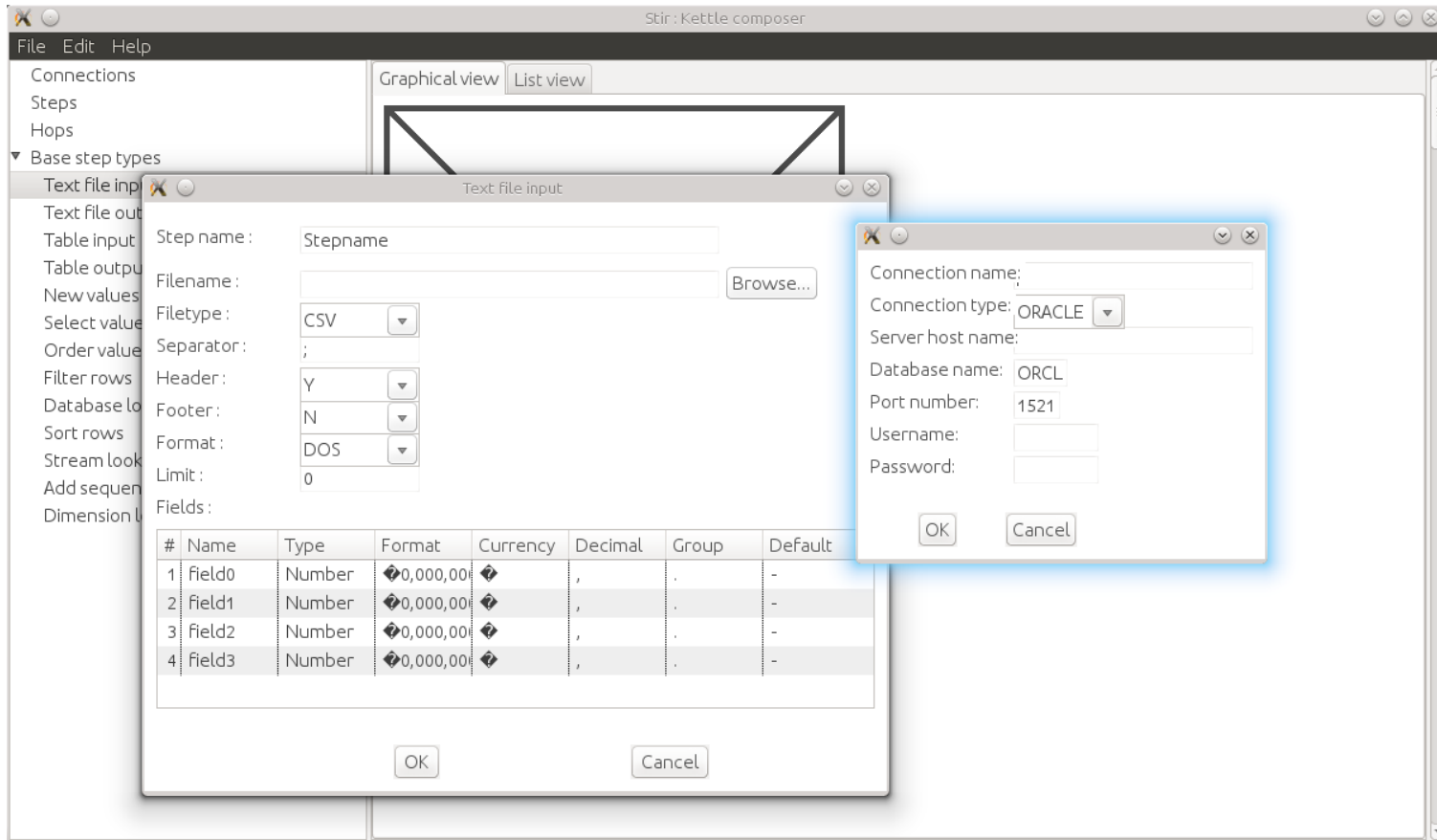


Open & Easy!

The evolution of the code

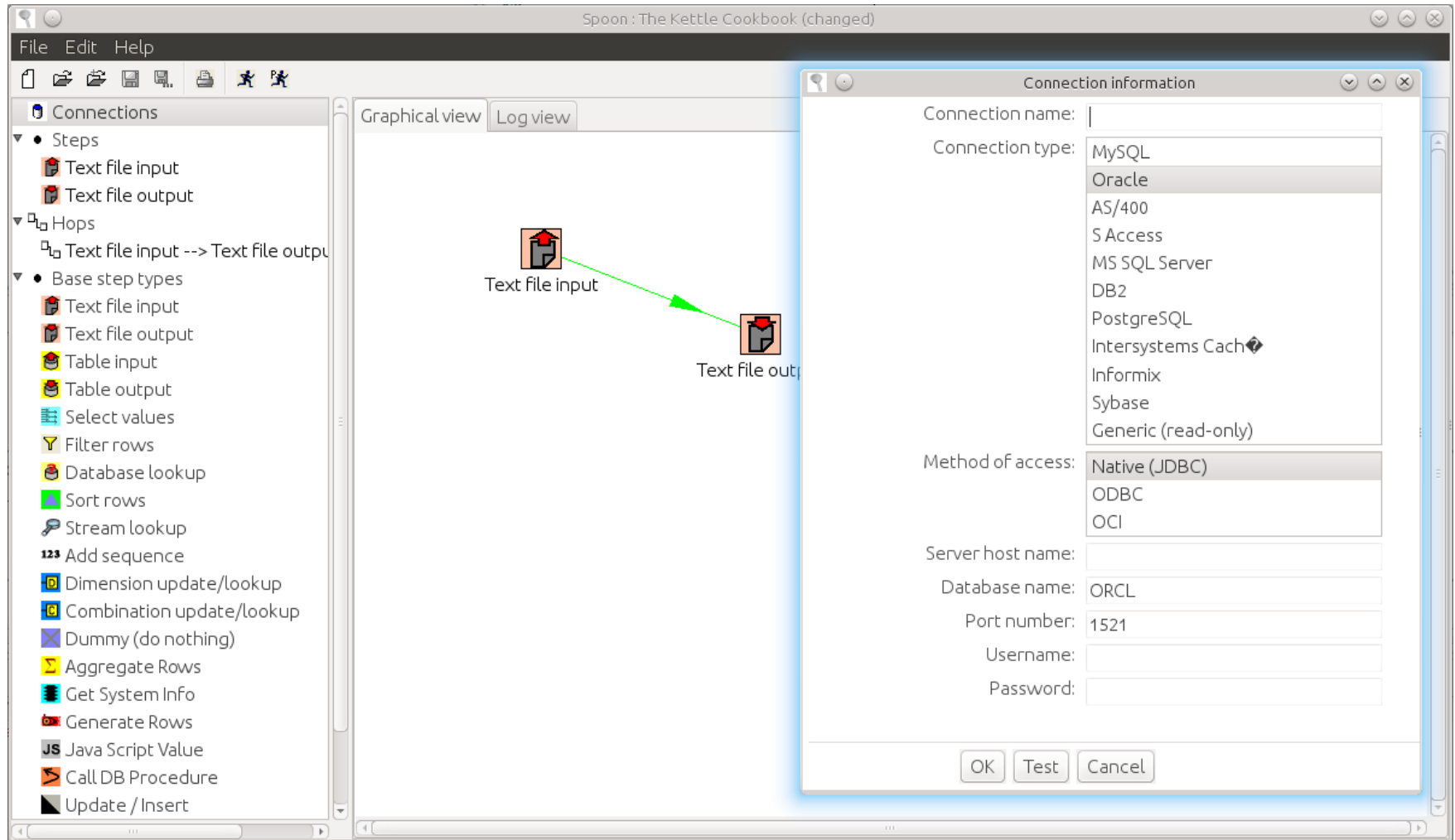
Code metrics & Demo

March 2003 v0: 15,000 lines of code (demo)



Code metrics & Demo

- **December 2003 v1: 50,000 lines of code (demo)**

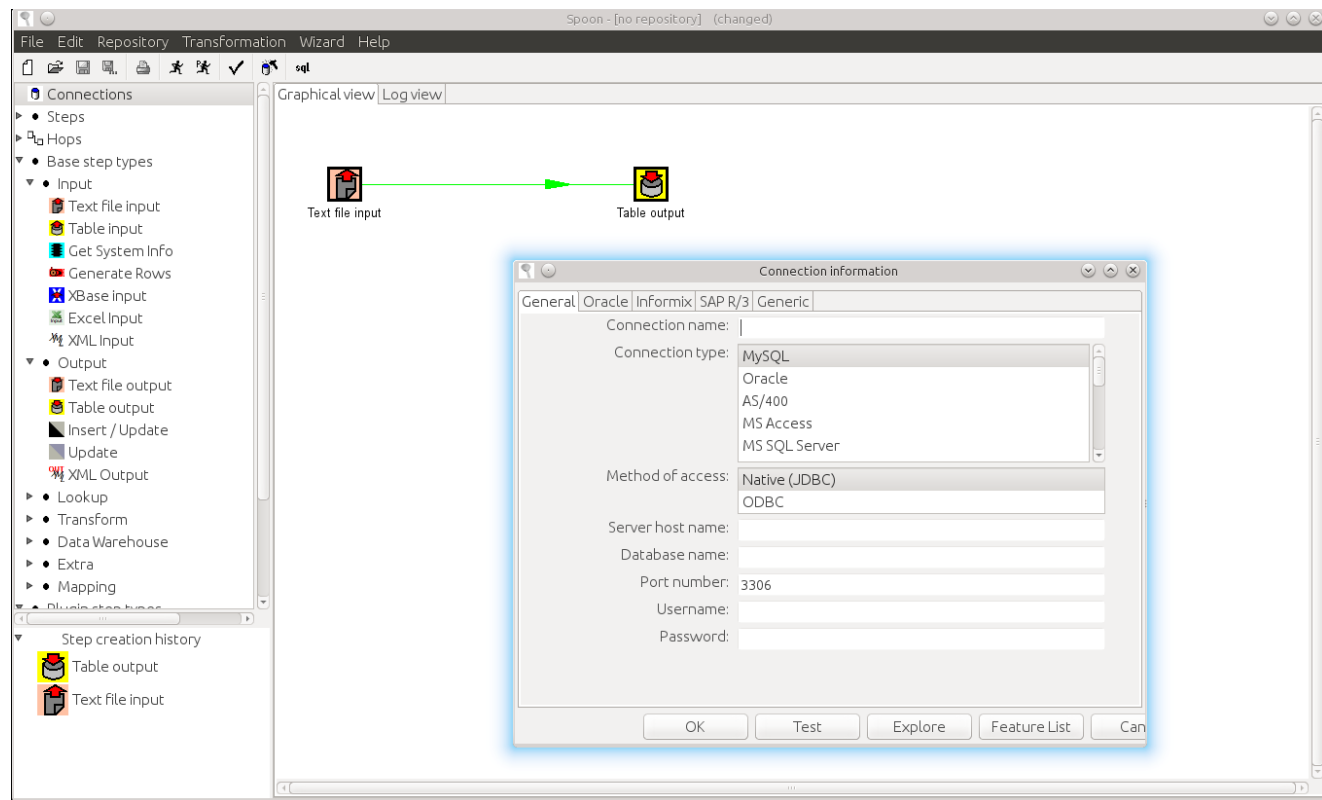




KETTLE® PRESENTATION 2004

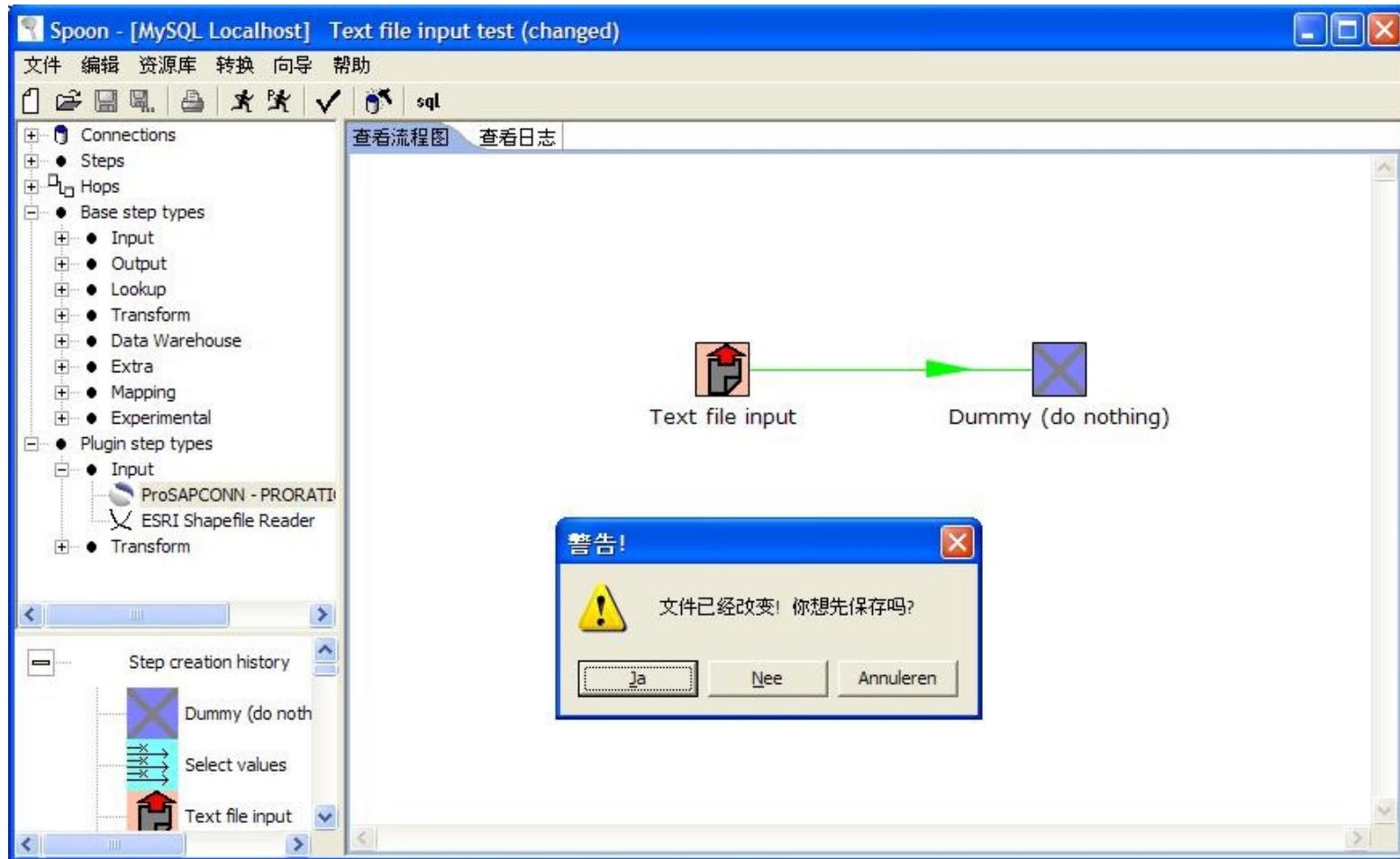
Code metrics & Demo

- **December 2005 v2.2.2: 177,000 lines of code**
- **Open sourcing of Kettle: LGPL**
- **Demo**



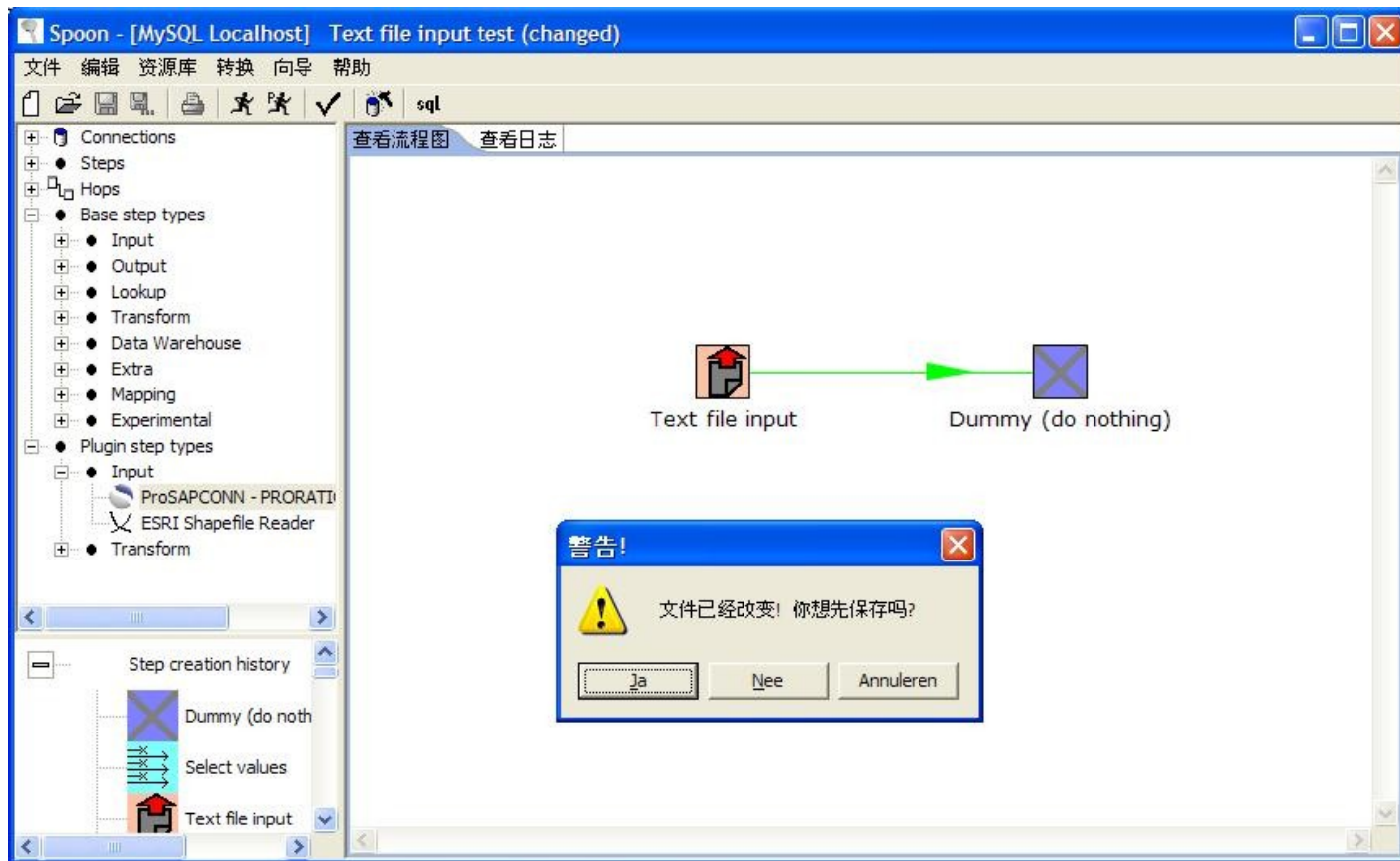
Code metrics & Demo

- March 2006 v2.3.0: 180,000 lines of code
- i18n



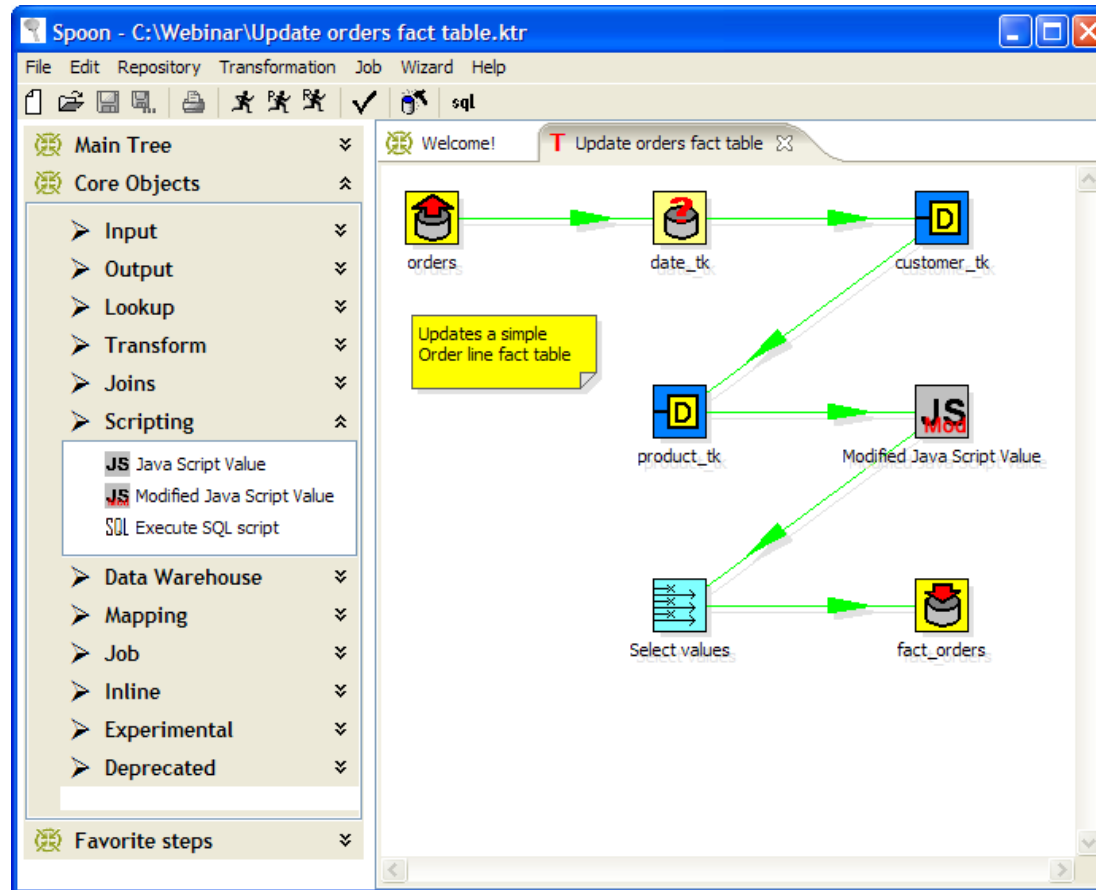
Code metrics & Demo

- March 2006 v2.4.0: 260,000 lines of code
- Clustering & Partitioning
- Still single document



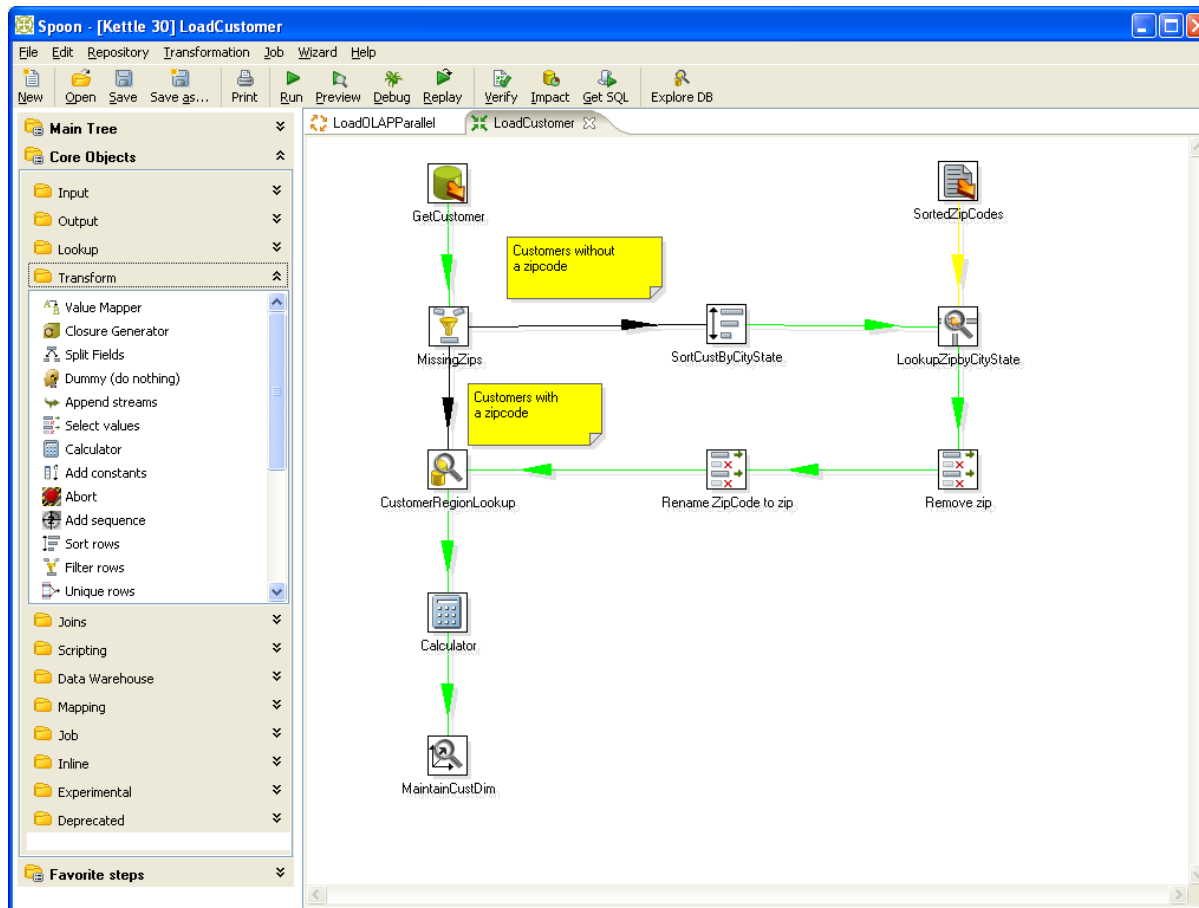
Code metrics & Demo

- March 2006 v2.5.0: 290,000 lines of code
- Multi-document (no more Chef tool)



Code metrics & Demo

- December 2007 v3.0.0 : 345,000 lines of code
- Finally: new icons!
- Multi-document, ...



Code metrics & Demo

- **Memorable version v3.2.5**
- **Updated until late 2011!**

Code metrics & Demo

- **v4.0.0 : 630,000 lines of code**
- **Perspectives**

Code metrics & Demo

- **v5.0.0 : 900,000 lines of code**
- **Too many cool toys to mention**

Thanks to the community!

2008 MAINZ



2008 MAINZ



2009 BARCELONA



2009 BARCELONA



2010 CASCAIS



2011 FRASCATI



**PENTAHO COMMUNITY GATHERING
ROME (FRASCATI) 2011**

2012 FORTALEZA



2012 AMSTERDAM



2013 FORTALEZA

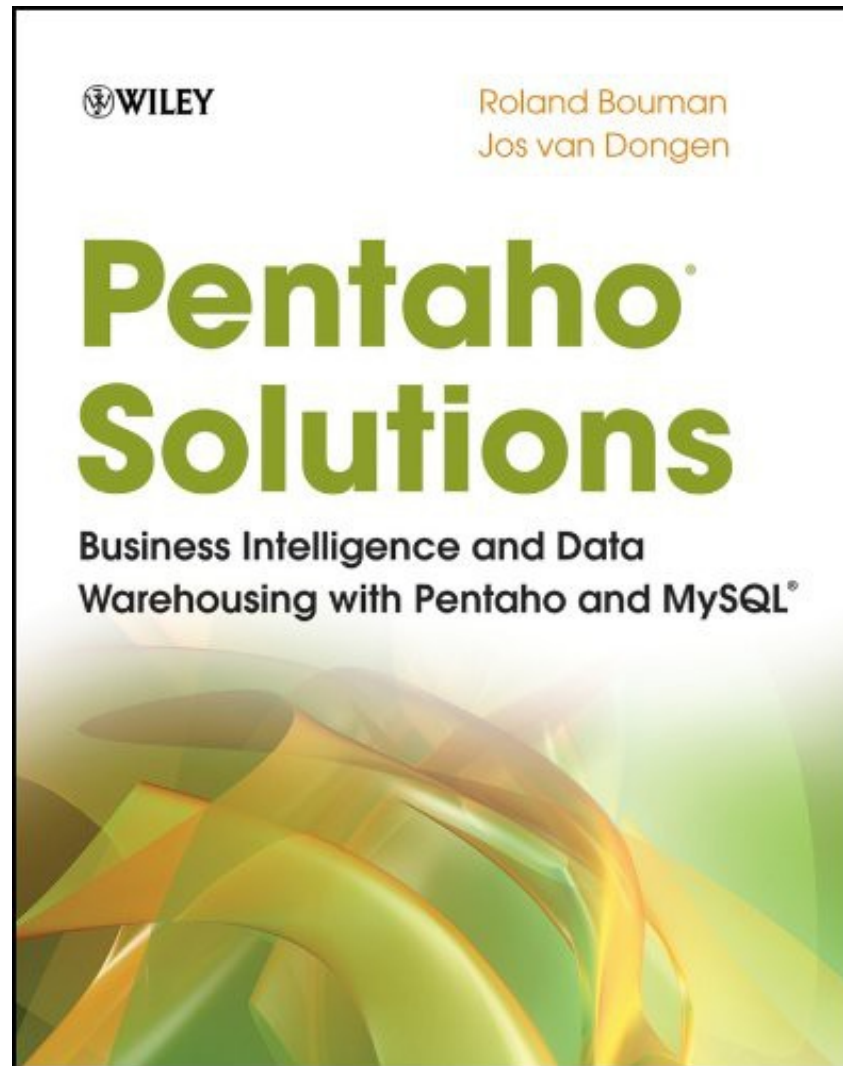


2013 SINTRA

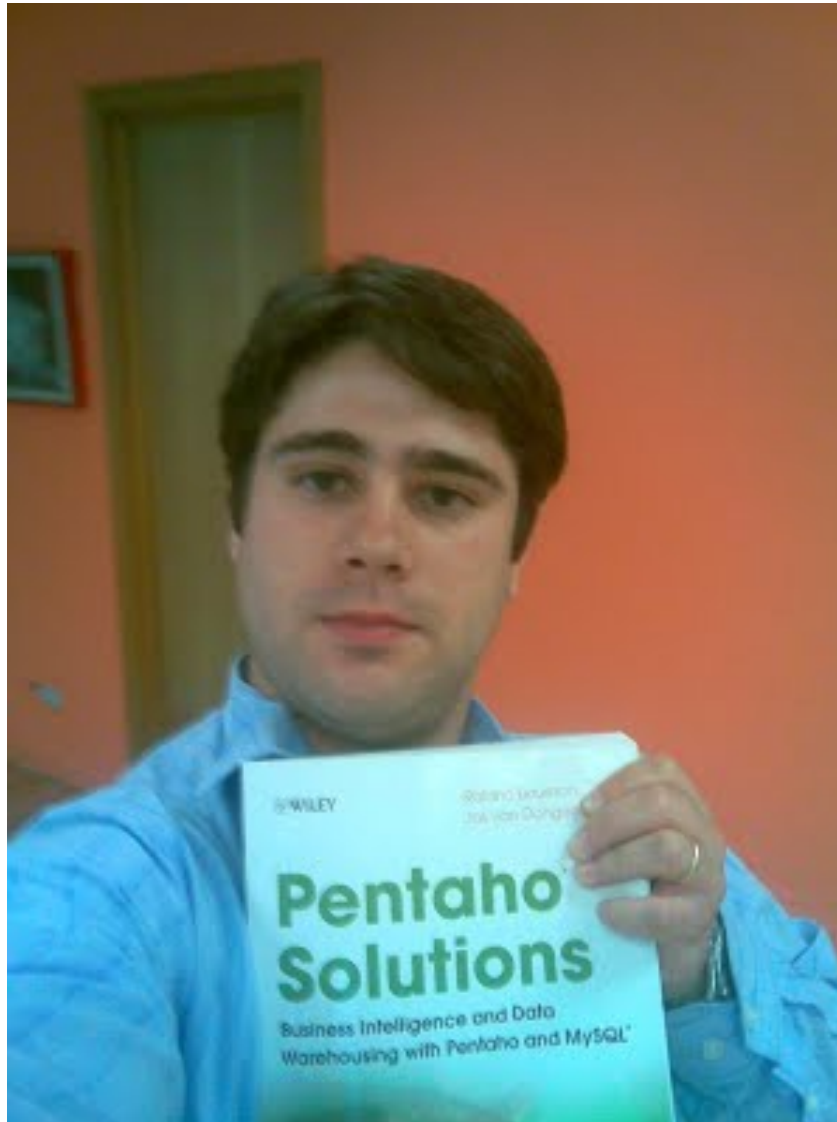


And books...

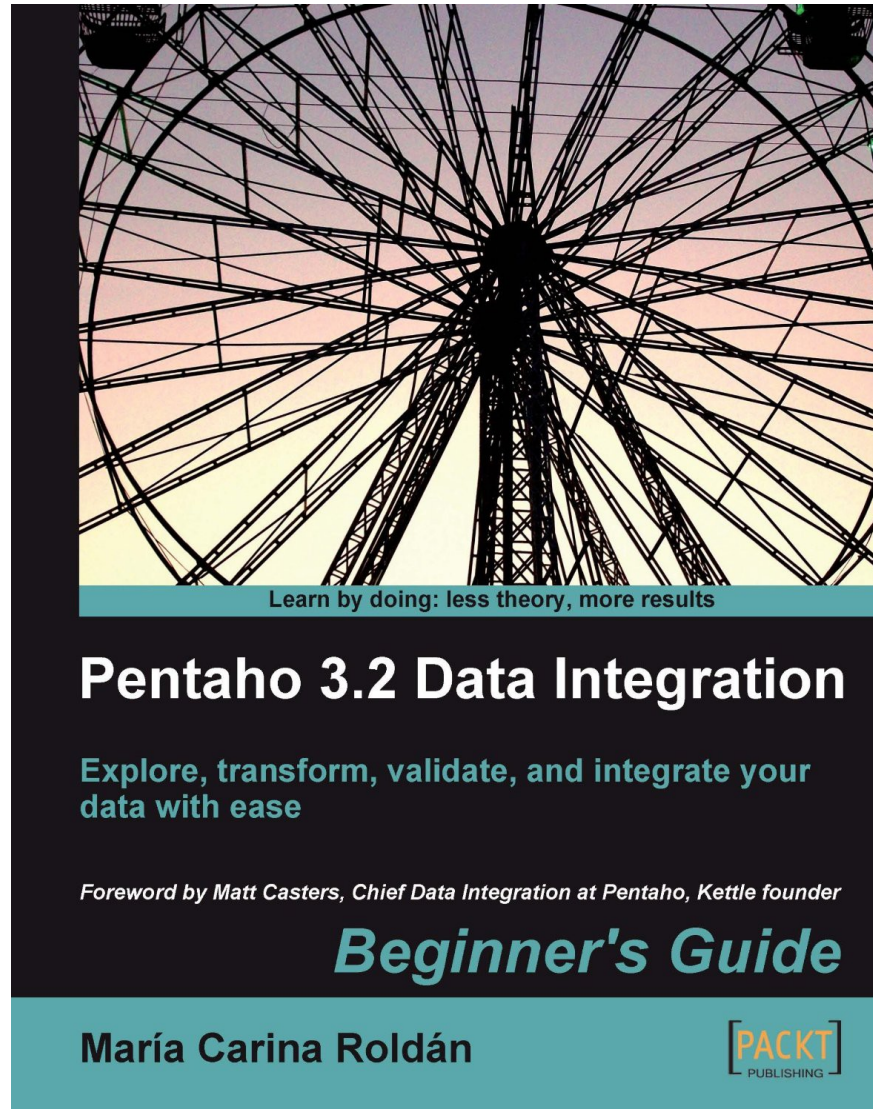
Books...



Books...



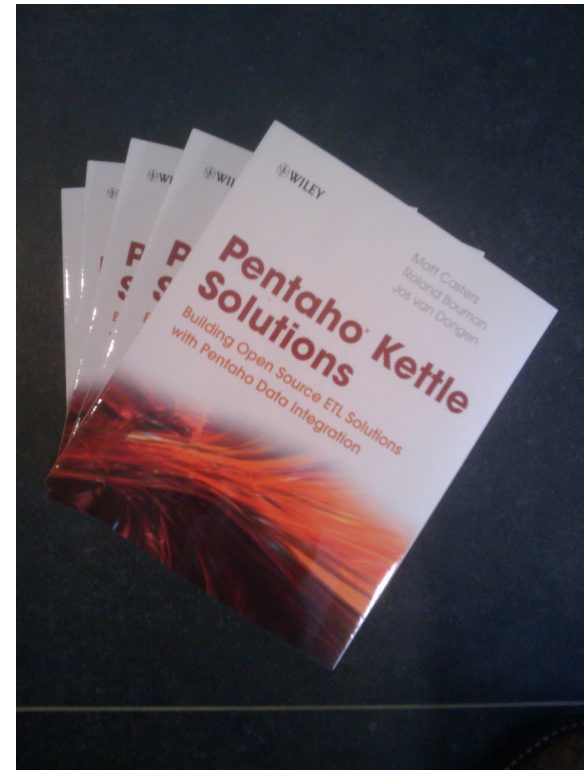
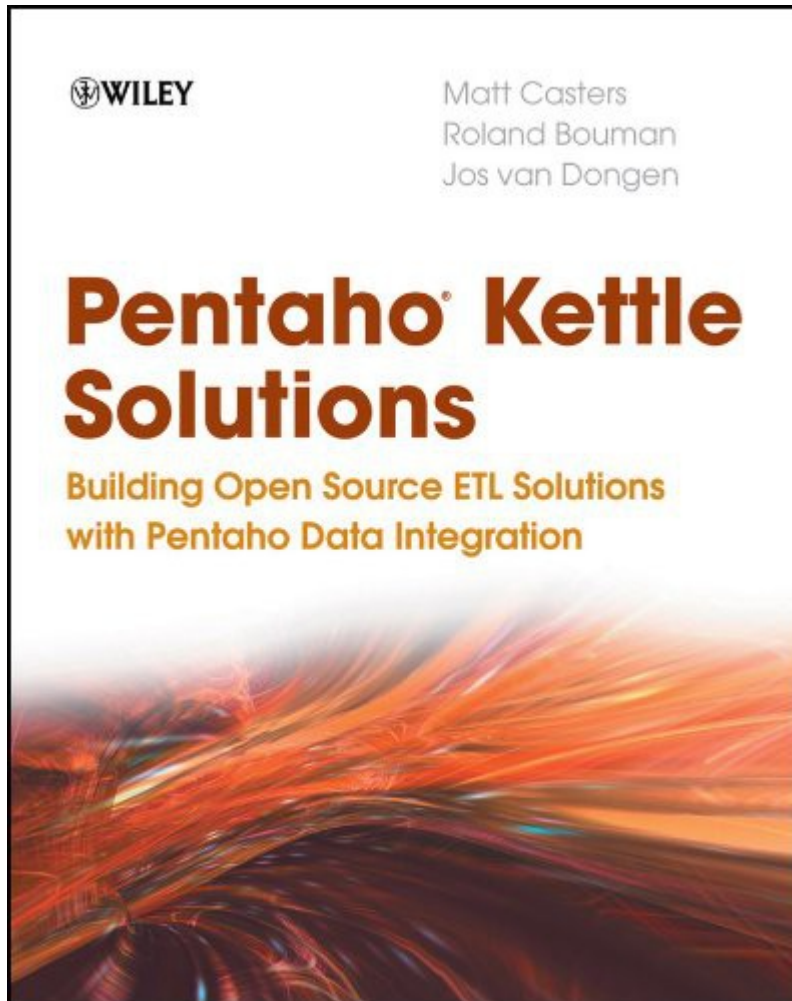
Books...



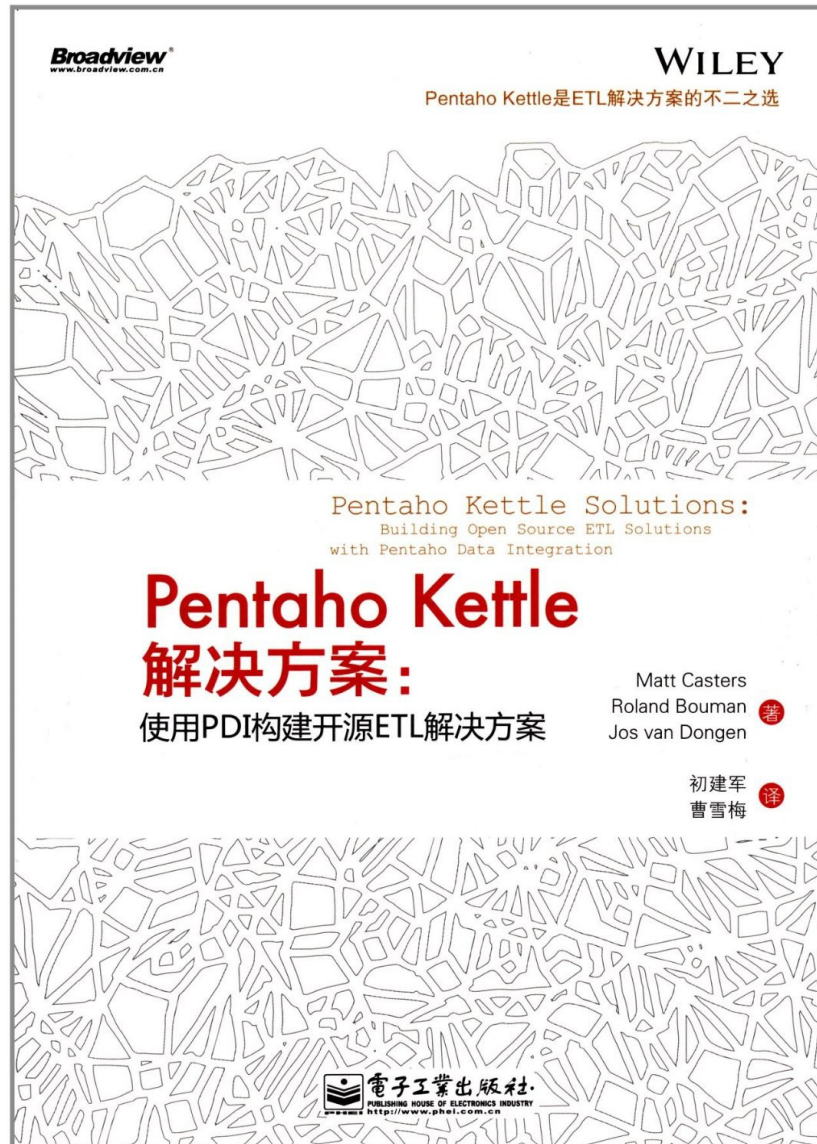
Books...



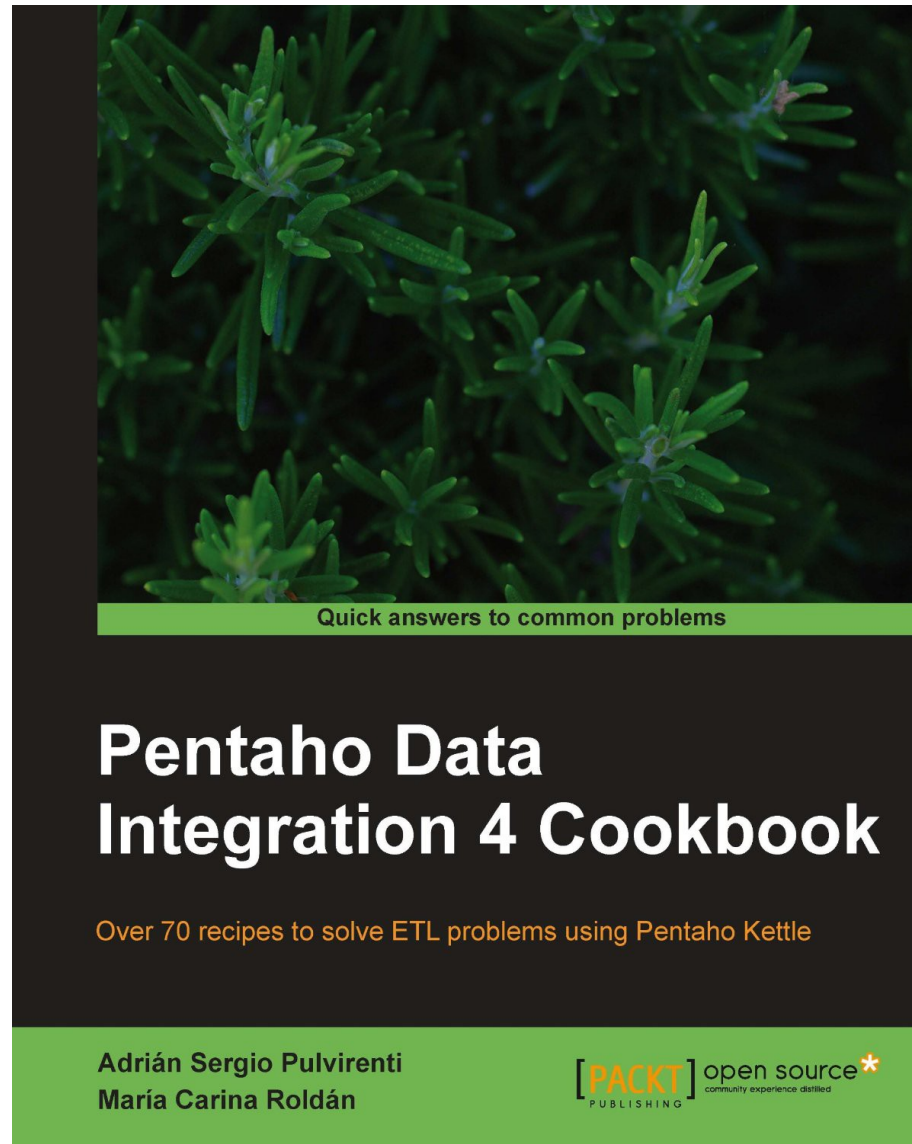
Books...



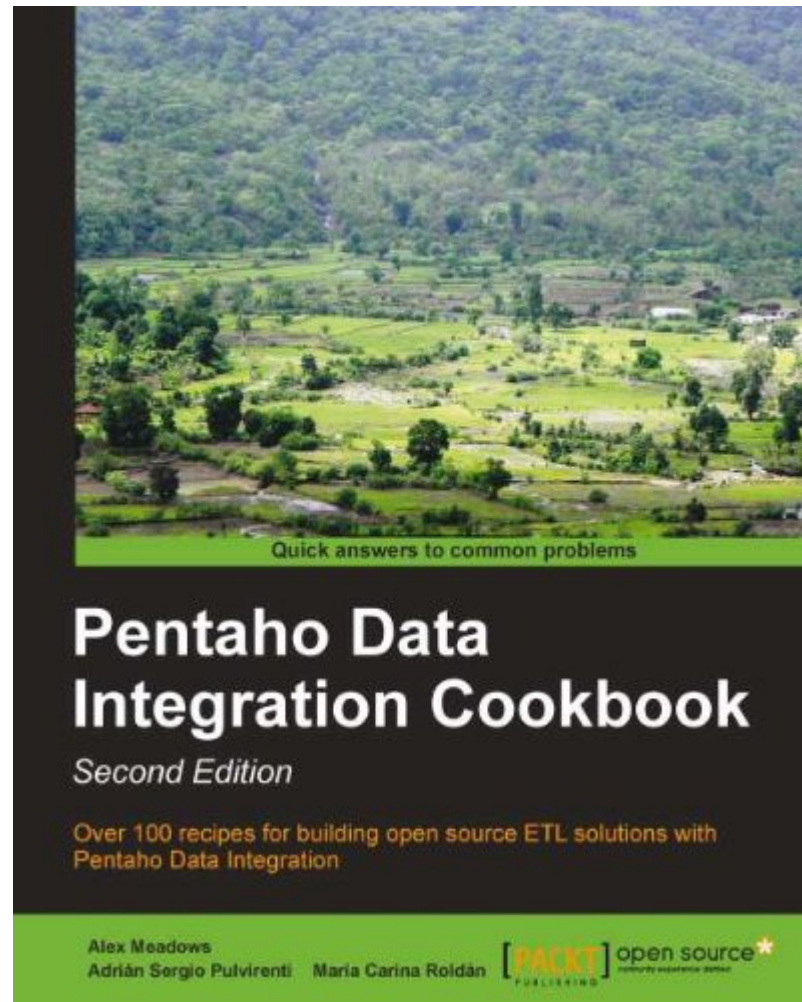
Books...



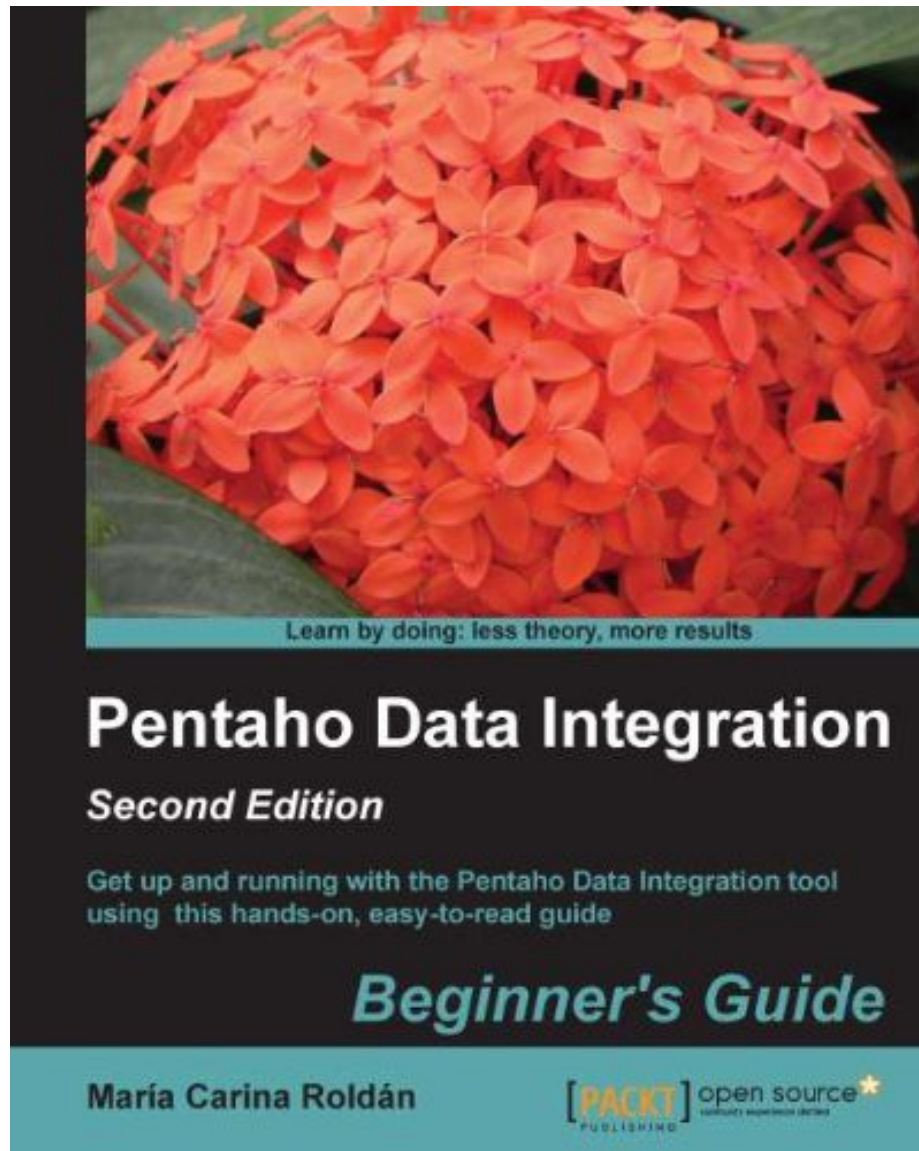
Books...



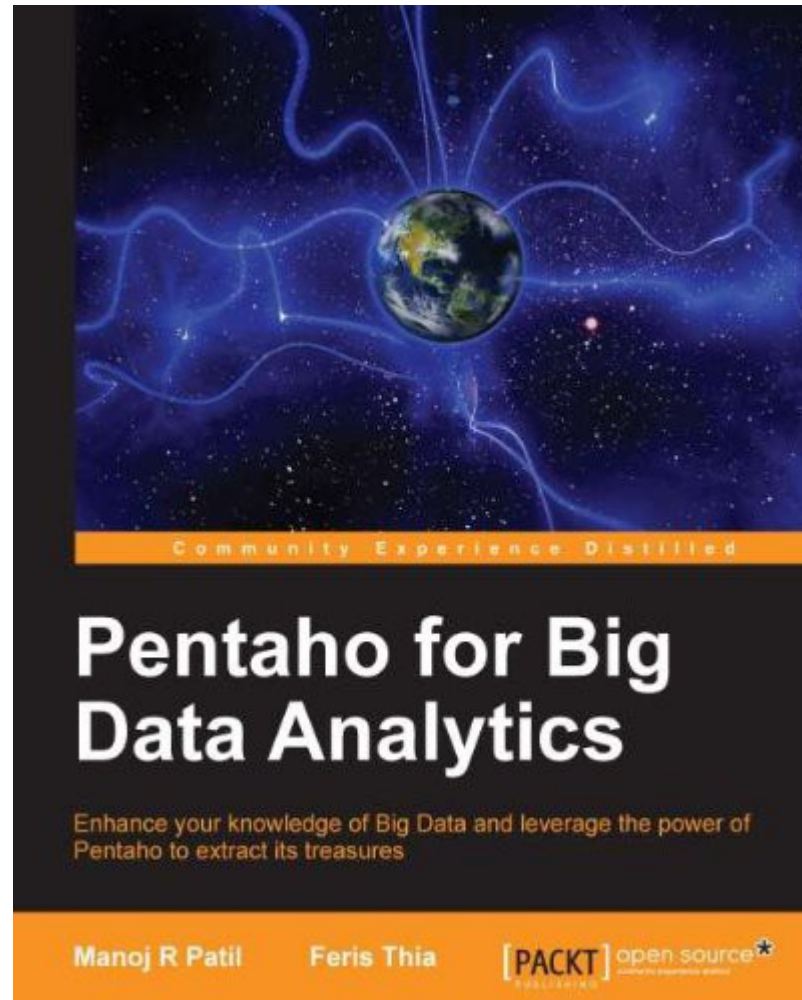
Books...



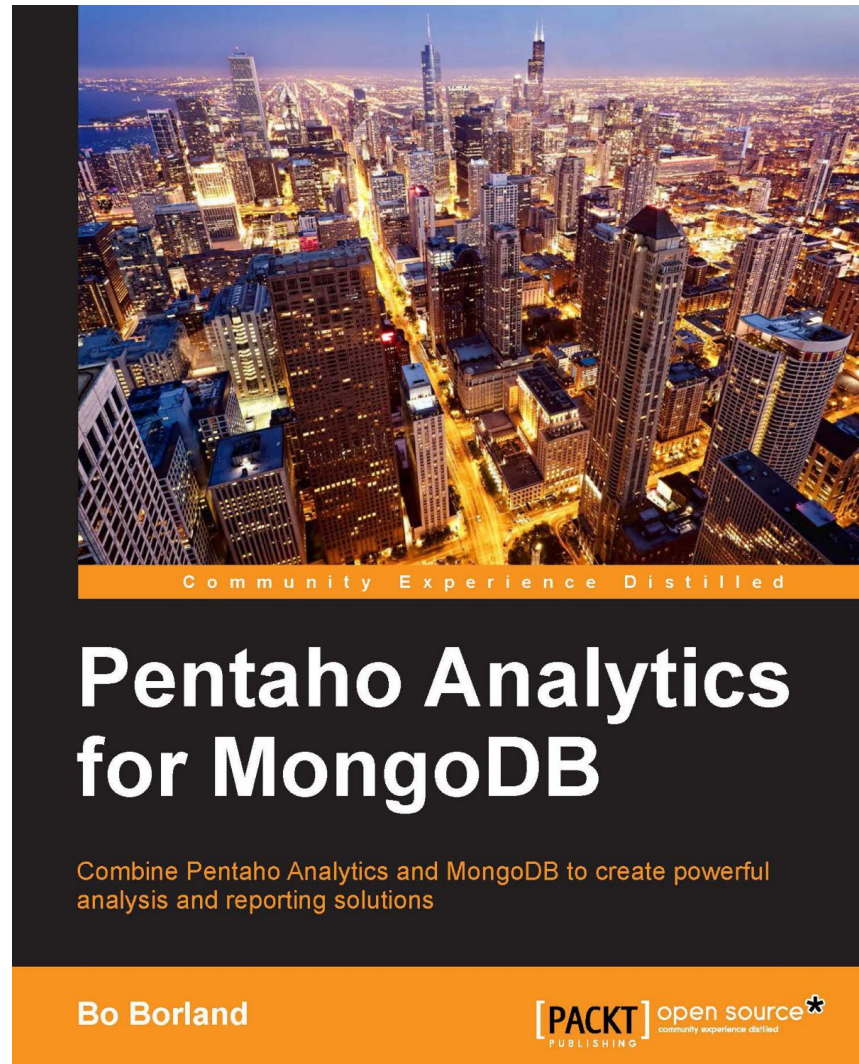
Books...



Books...



Books...



Only one thing to say...

Thank You

JOIN THE CONVERSATION. YOU CAN FIND US ON:



blog.pentaho.com



[Facebook.com/Pentaho](https://www.facebook.com/Pentaho)



[@Pentaho](https://twitter.com/Pentaho)



[Pentaho Business Analytics](#)