

Pentaho

sETLer – simple ETL engineering

André Simões aka ITXpander aka “the useless guy on IRC”

andre.simoes@xpand-it.com

<http://www.twitter.com/itxpander>

THE USUAL ETL STUFF



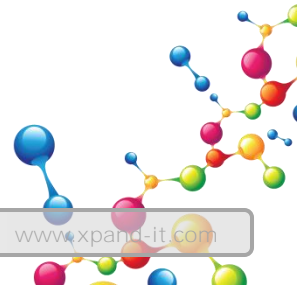
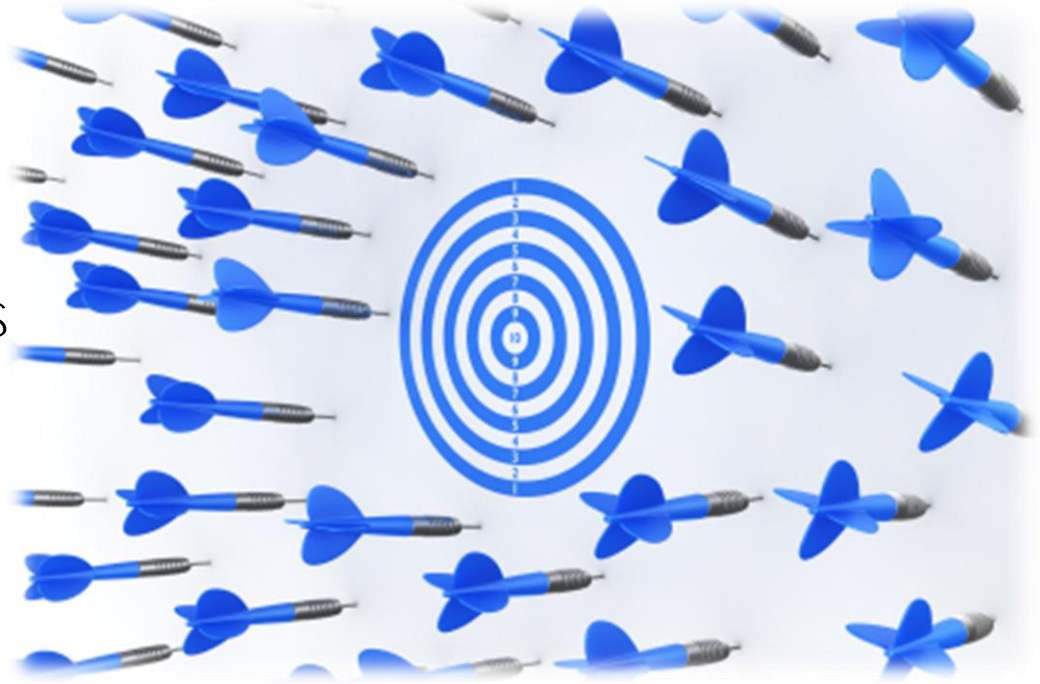
- ETL project setup effort
- ETL Maintenance
- Too much time figuring out where
- Cumbersome process chaining
- Hard to modify the ETL process/flow
- No easy way to schedule
- ...



- ... **Chaos!**



- Scheduling
- Alerts
- Execution parameters
- Execution variables
- Dependencies
- History

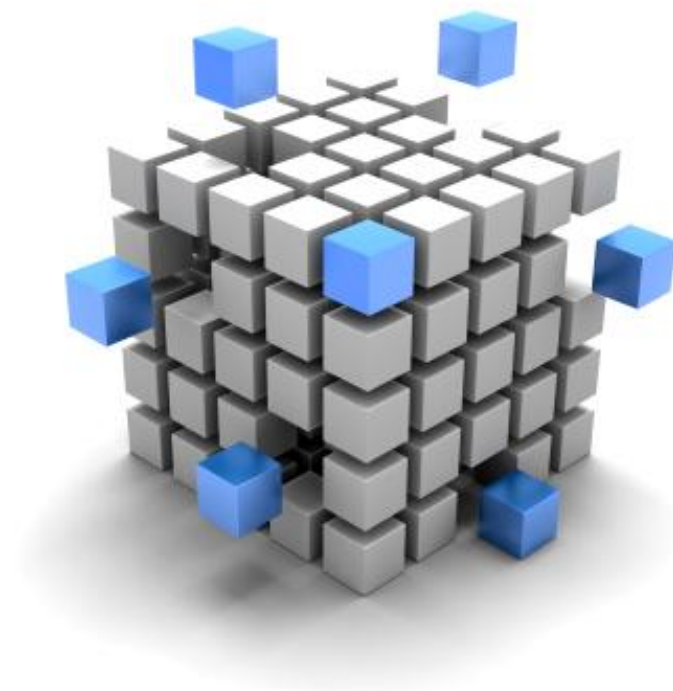


sETLer



- ETL chaining
- ETL scheduling
- ETL dependency configuration
- Dynamic ETL Loading
- Configurable
- Per-project context

- ... Great!

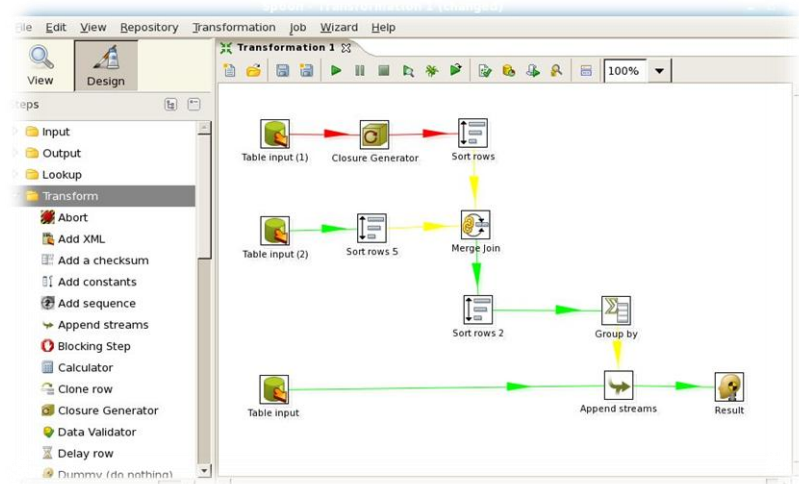


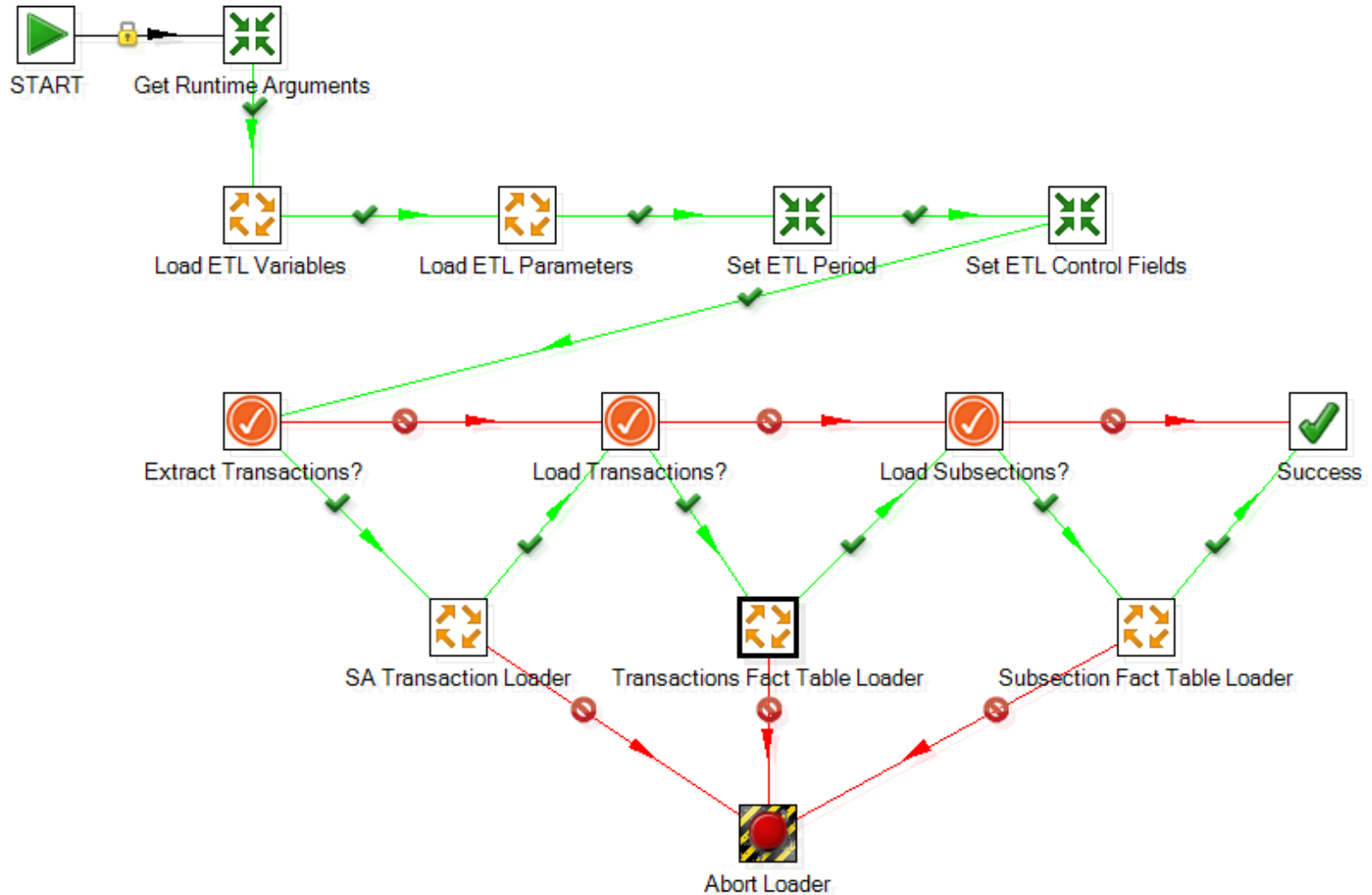
CONCEPTS



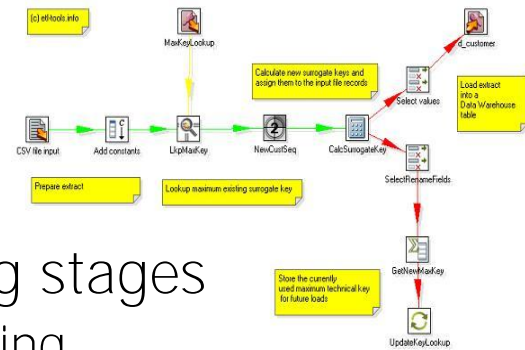
- Global
 - Stored in a pre-defined database schema
 - (name, value)
 - Example: table names, last execution time, ...
- Runtime
 - Passed as command line arguments
 - Named parameters
 - Computed in the etl process/script
 - Example: start and end processing dates
- Should be available for all the processes to use freely!

- ETL entry point
- Sets up the ETL process
 - Parameters
 - Variables
 - Command line options
- Dependency checking stage
- Checkpoints
- Contains the calls for the ETL processes
 - ... Coming up next.
- Examples: Daily aggregates load, Monthly fact tables loading, ...

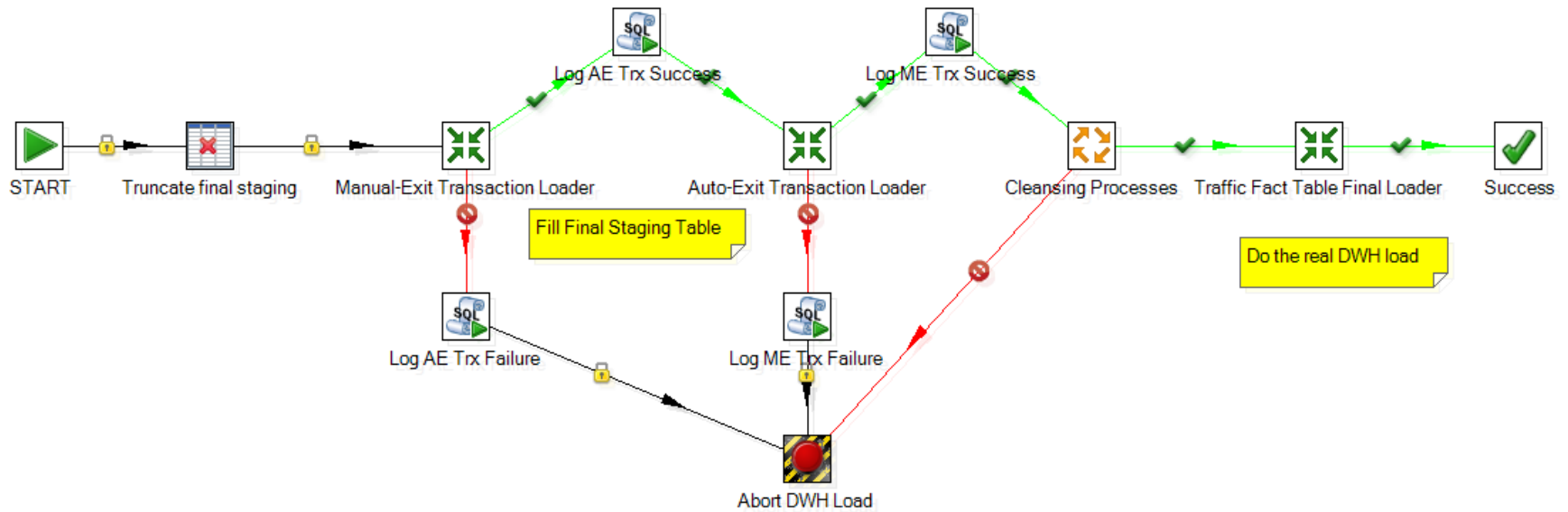




- Independent ETL process
- Must be self contained
 - Can be ran independently
- Should (!) contain well defined processing stages
 - Extraction, Transformation, Loading, Cleansing, ...
- Used on the chains
- Can contain sub-checkpoints
- Examples: Revenue Fact Table load, Customer dimension load, ...



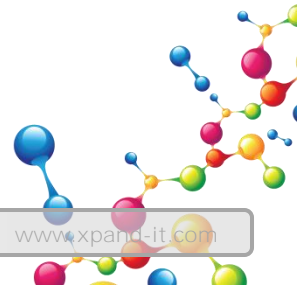
LOADER – PDI EXAMPLE

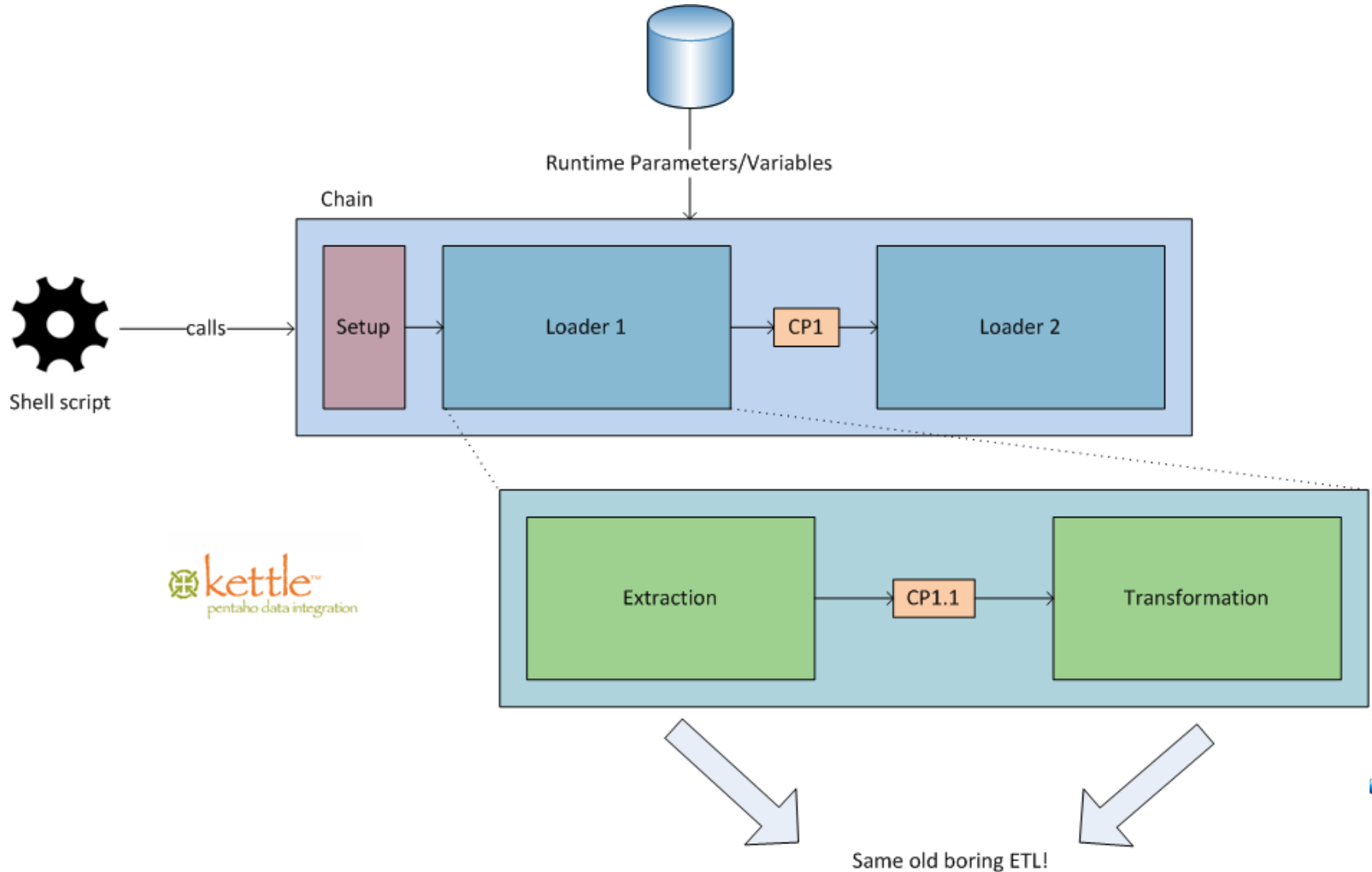


- ETL stage where you can resume from
 - Staging area load, data transformation, ...
- Mapped as PDI named parameters
 - Passed through the commandline
 - Easy to remember
 - Easy to document (with kettle cookbook!)
- Example:

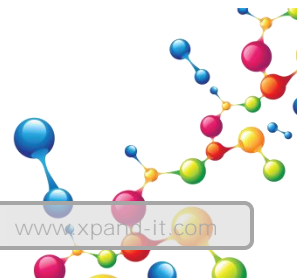
```
./kitchen.sh -Dfile="..." -param: extract.staging=n -param: transform.data=y
```

- Batch/Shell script templates
- Maintain per-project context
 - KETTLE_HOME
 - PDI installation location
- Cron safe
- Variable memory job/transformation allocation





- Publicly available
- Define templating for the usual ETL phases
 - Lookups, extraction, ...
- Merge with KFF?
- ... Whatever... !


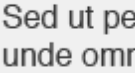


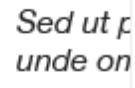





BUT WAIT... DO YOU USE CONFLUENCE?



WHAT IF YOU COULD INCLUDE YOUR REPORTS?

Select Macro

All	 Page Tree Search Provides a search box that searches a hierarchy of pages (Page Tree) from a specified root page.	 Panel Displays a block of text within a customisable panel.
Administration		
Communication		
Confluence Content		
Development		
External Content	 PentahoReport Generate Pentaho Reporting	 Popular Labels Generates a list or 'heatmap' of the most popular labels.
Formatting		
Navigation	 Quote Formats a block of text consisting of one or more paragraphs into a quote.	 Recently Updated Lists the most recently changed content within Confluence.
Reporting		
Visuals & Images	 Recently Updated Dashboard Lists the most recently changed Confluence content in a dashboard view style.	 Recently Used Labels Lists the most recently used labels in a predefined scope of spaces.

Cancel

Insert 'PentahoReport' Macro

Preview

 Refresh

Generate Pentaho Reporting

This macro does not provide any parameter information. If the available information does not help, you may find its documentation at [Confluence Extensions](#).

{pentahoReport:

What's New in Pentaho

New Report Designer

Completely New Interface

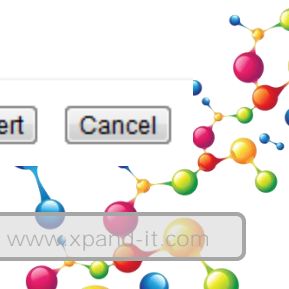
Report Tab Workspace (Design

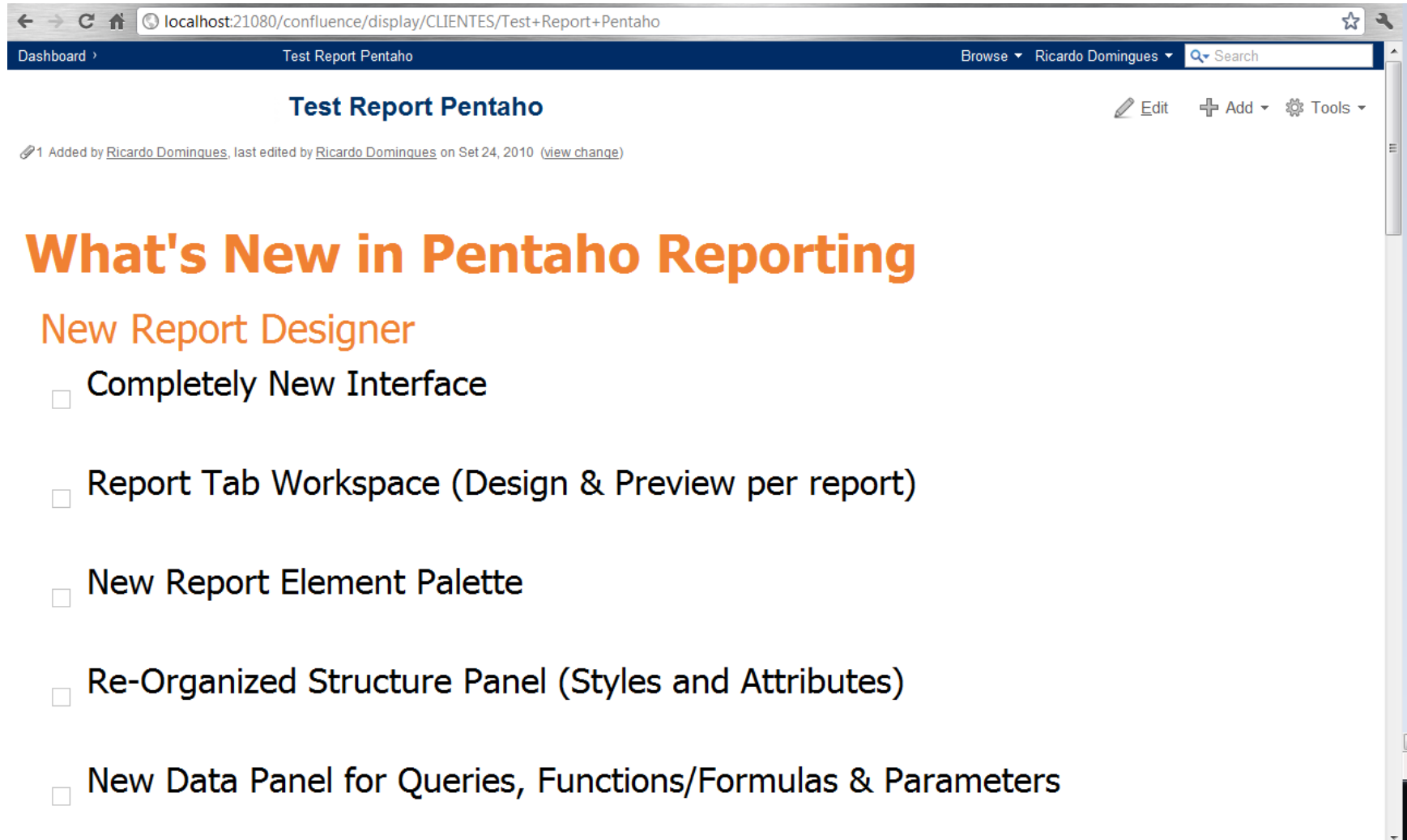
New Report Element Palette

Back

Insert

Cancel





The screenshot shows a web browser window displaying a Confluence page. The address bar shows the URL: localhost:21080/confluence/display/CLIENTES/Test+Report+Pentaho. The page title is "Test Report Pentaho". Below the title, there is a metadata line: "1 Added by Ricardo Domingues, last edited by Ricardo Domingues on Set 24, 2010 (view change)". The main content of the page is a list of new features in Pentaho Reporting, each preceded by a square checkbox.

Test Report Pentaho

1 Added by Ricardo Domingues, last edited by Ricardo Domingues on Set 24, 2010 (view change)

What's New in Pentaho Reporting

New Report Designer

- Completely New Interface
- Report Tab Workspace (Design & Preview per report)
- New Report Element Palette
- Re-Organized Structure Panel (Styles and Attributes)
- New Data Panel for Queries, Functions/Formulas & Parameters

- Still in a very early development phase
- Will be available as open-source