



kettle franchising factory

Spice up your kitchen NOW!

kff.kjube.be

Why oh why?

Agenda

What is KFF?

Where is this going
and how will we get there?

(The) Who?

(What) Do you think?





Jan Aertsen
Managing partner

kJube



Matt Casters
Chief Data Integration

Pentaho

Tired

... of re-inventing



Mama, ... taught us to share

Why oh why?

Agenda

What is KFF?

Where is this going
and how will we get there?

(The) Who?

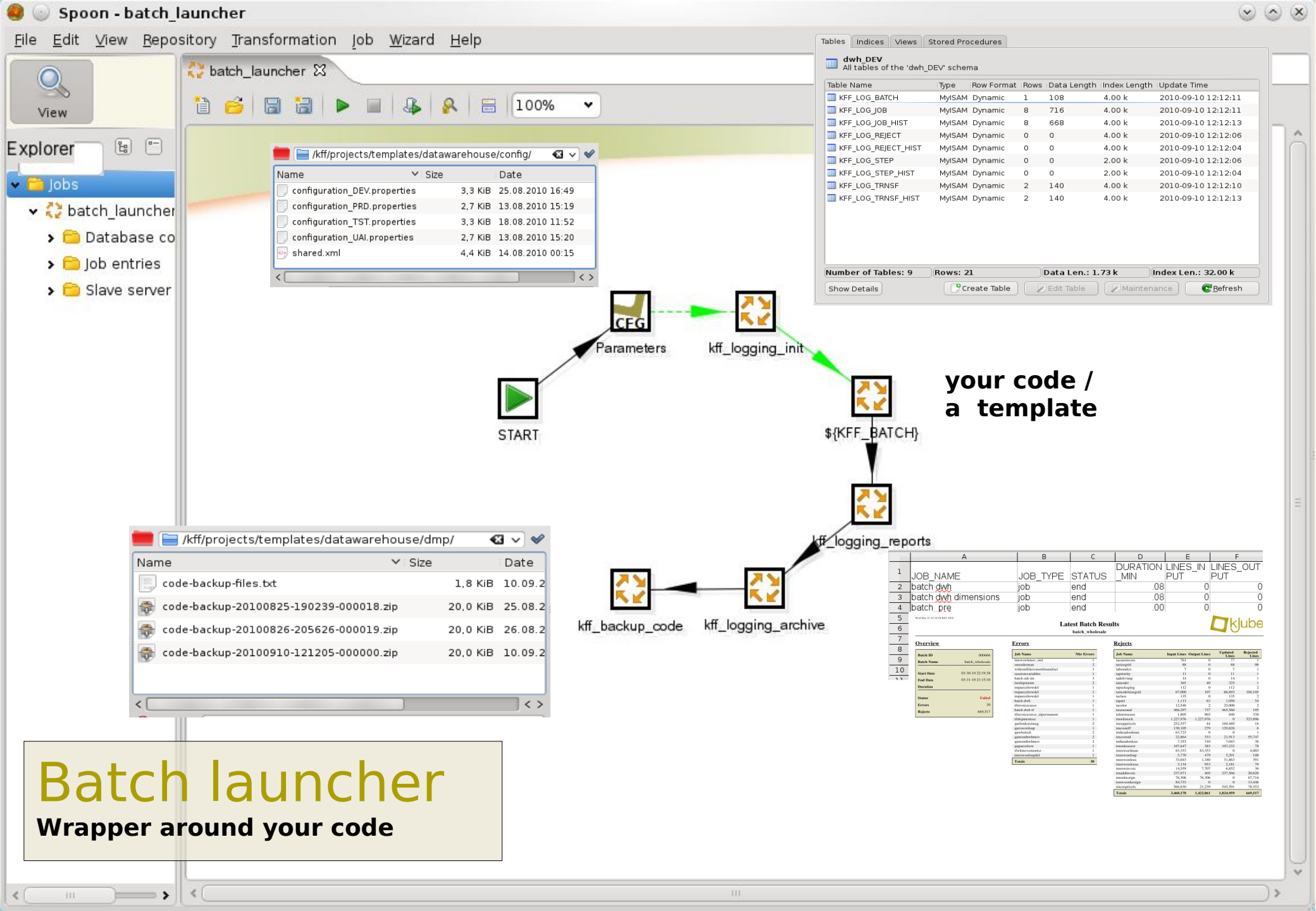
(What) Do you think?



Kettle Franchising Factory

a framework for rapid deployment of kettle solutions

Standards for kettle set-up (stand-alone DI)
Multi-environment / Multiple customer set-up
Standardize logging, scheduling, restartability
Project templates
Plug-ins (where needed)



**your code /
a template**

**Batch launcher
Wrapper around your code**

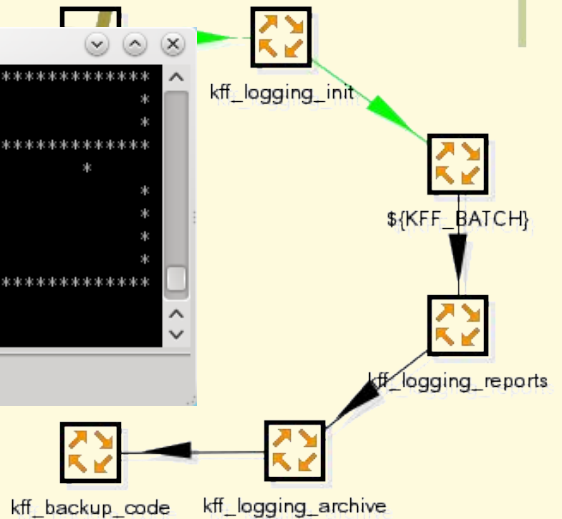
/kff/projects/templates/datawarehouse/config/

Name	Size	Date
configuration_DEV.properties	3,3 KiB	25.08.2010 16:49
configuration_PRD.properties	2,7 KiB	13.08.2010 15:19
configuration_TST.properties	3,3 KiB	18.08.2010 11:52
configuration_UAI.properties	2,7 KiB	13.08.2010 15:20
shared.xml	4,4 KiB	14.08.2010 00:15

spoon

```

datawarehouse : bash
*****
* Launching Pentaho Data Integration
* Choose your PDI environment!
*****
* templates - datawarehouse
* P - PRD
* U - UAI
* T - TST
* D - DEV
*****
Enter your choice: T
  
```



kitchen

kjube configurator

Job entry name: Parameters

Parameters

Customer parameter:

Application parameter:

Lifecycle parameter:

Config file path:

Buttons: OK, Verify/Add parameters, Cancel

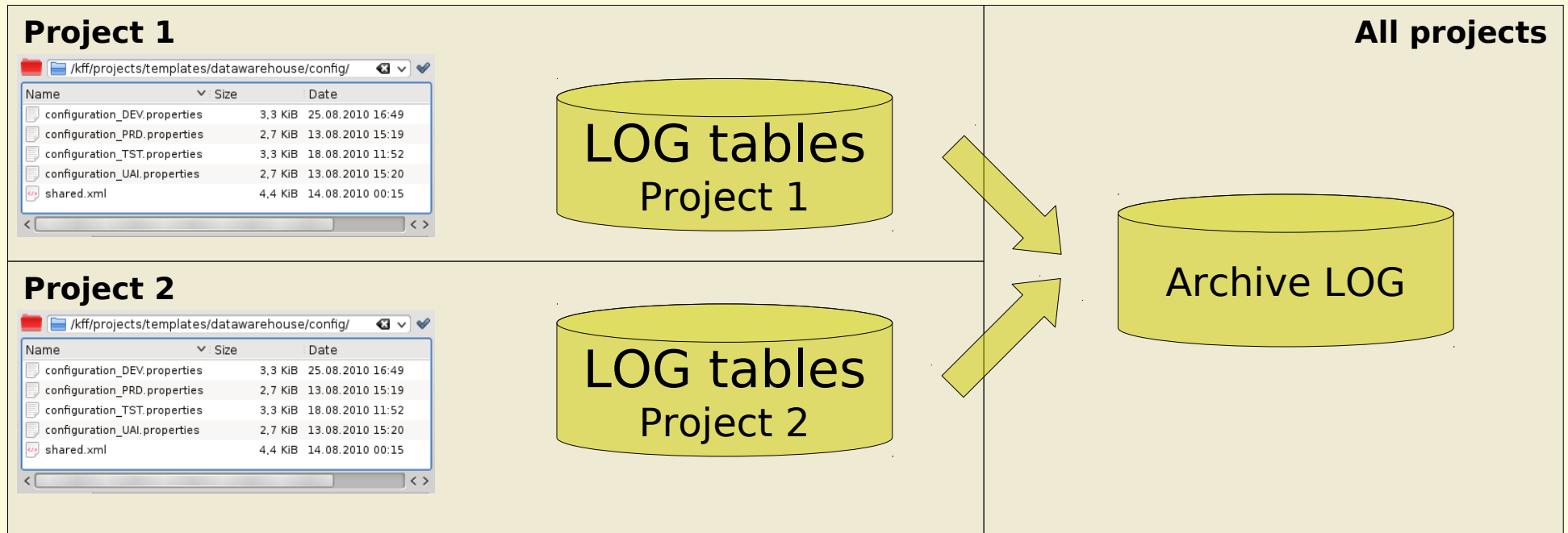
Why automatic environment configuration?

- 15 customers/projects x 4 environments vs. 1 kettle.properties just ain't fair
- I don't really wake up until 11AM, but all re-runs need to be done before 8AM.
- Avoid having to buy more computers to avoid mixing up projects

Why create logging on the fly?

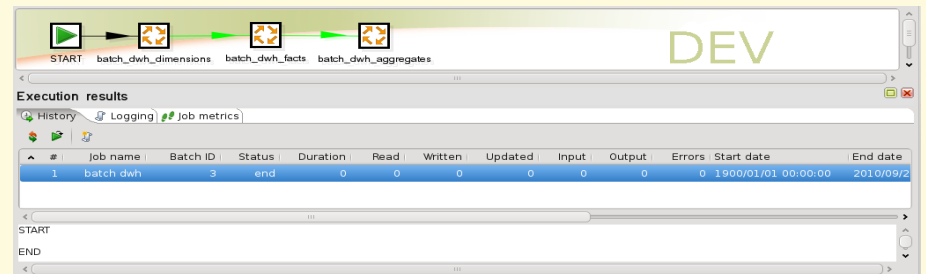
- Quicker to get started with KFF
- Quicker in case you need to migrate
- Allows consistent logging over parallel runs

```
# KFF-logging tables #####
KFF_LOGGING_CONNECTION=DWH_LOG
KFF_LOGGING_SCHEMA=
KFF_LOGGING_BATCH_TABLE=KFF_LOG_BATCH
KFF_LOGGING_JOB_TABLE=KFF_LOG_JOB
KFF_LOGGING_TRNSF_TABLE=KFF_LOG_TRNSF
KFF_LOGGING_STEP_TABLE=KFF_LOG_STEP
KFF_LOGGING_REJECT_TABLE=KFF_LOG_REJECT
KFF_LOGGING_JOB_TABLE_HIST=KFF_LOG_JOB_HIST
KFF_LOGGING_TRNSF_TABLE_HIST=KFF_LOG_TRNSF_HIST
KFF_LOGGING_STEP_TABLE_HIST=KFF_LOG_STEP_HIST
KFF_LOGGING_REJECT_TABLE_HIST=KFF_LOG_REJECT_HIST
```



Why archive logging?

- Keeps kettle interface “readable”



Never forget again to configure logging (injection avant la lettre)

Step name: inject logging into job

Java script functions:

- Transform Scripts
- Transform Constants
- Transform Functions
- Input fields
- Output fields

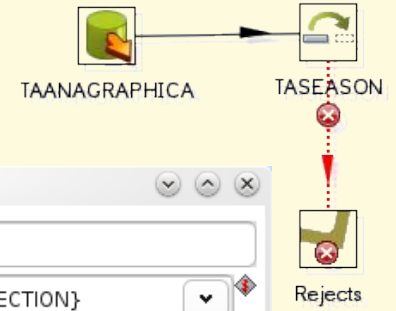
Java script:

```
Script 1 ⌘  
//Script here  
  
var logConnection = getVariable("KFF_LOGGING_CONNECTION", "??DWH_LOG??");  
var logTable = getVariable("KFF_LOGGING_JOB_TABLE", "???JOB_LOG???");  
  
var log = Packages.org.pentaho.di.core.logging.LogWriter.getInstance();  
var jobMeta = new Packages.org.pentaho.di.job.JobMeta(log, path, null);  
var db = jobMeta.findDatabase(logConnection);  
  
jobMeta.setLogConnection(db);  
jobMeta.setLogTable(logTable);  
  
var dos = new Packages.java.io.DataOutputStream(  
    Packages.org.pentaho.di.core.vfs.KettleVFS.getOutputStream(path, false,  
    );  
dos.write(Packages.org.pentaho.di.core.xml.XMLHandler.getXMLHeader().getBytes(  
    );  
dos.close();  
}
```

Fields

#	Fieldname	Rename to	Type	Length	Precision	Replace value 'Fieldname' or 'Rename to'
1						

Buttons: OK, Cancel, Get variables, Test script



Step name: Rejects

Rejected records connection: \${KJUBE_LOGGING_CONNECTION}

Rejected records schema: \${KJUBE_LOGGING_SCHEMA}

Rejected records table: \${KJUBE_LOGGING_REJECT_TABLE}

Error count field: \${KJUBE_ERROR_COUNT_FIELD}

Error descriptions field: \${KJUBE_ERROR_DESCRIPTIONS_FIELD}

Error fields field: \${KJUBE_ERROR_FIELDS_FIELD}

Error codes field: \${KJUBE_ERROR_CODES_FIELD}

Fields

#	Key Field (optional)
1	IDSEASON
2	IDLANGUAGE

Buttons: OK, SQL, Cancel

An extra log table: rejects

Some other plug-ins

Trim strings

Trim string fields

Step name: Trim strings

Trim all string fields:

Trim type: Right

Exclude these fields from trim:

Fields

#	Field	Trim type
1	MyField	Left

Left
Right
Both
None

OK Get Fields Cancel

Date time calculator

Date-time calculator

Step name: Date-time calculat

Ignore date errors (convert to null):

Date-time fields (H,M,S not required):

#	Result	Century	Year					
1	PCG_PresentationDATETIME	20	10					
2	PCG_BeerDateTime	20	10	9	25	19	0	0

OK Cancel

Table compare

kjube Table Compare

Step name: Table Compare

Reference connection: [dropdown]

Reference schema field: [dropdown]

Reference table field: [dropdown]

Compare connection: [dropdown]

Compare schema field: [dropdown]

Compare table field: [dropdown]

Key fields field: [dropdown]

Exclude fields field: [dropdown]

Number of errors field: nrErrors

Number of reference table records field: nrRecordsReferenceTable

Number of compare table records field: nrRecordsCompareTable

Number of left join errors field: nrErrorsLeftJoin

Number of inner join errors field: nrErrorsInnerJoin

Number of right join errors field: nrErrorsRightJoin

Error handling key description input field: [dropdown]

Error handling reference value input field: [dropdown]

Error handling compare value input field: [dropdown]

Project templates

Datawarehouse

Typical datawarehouse project, with staging area, ODS, and data marts. All connections preconfigured. Standard dimensions to be added.

Data vault generator (Edwin Weber)

Recent contribution allows to generate data vault from configuration files.

Campaign manager

Template project to manage mailing lists + responses.

Database comparison

Pre-configured project template (using table compare step) to simplify acceptance testing (comparing UAI database against PRD).

Logging-monitoring

On the fly creation / initialization
Inject logging settings into job/trf
Archive logging
Generate logging report
Mail results to administrator

Archiving

Archive your code

Project templates / Best practices

Data warehouse template
Data vault (generator) *
Database compare *
Campaign manager *

Multi-environment / Multi-customer

Standardize across environments
- naming conventions
- directory structure
Automatically use right configuration

Plug-ins

To support the above
Environment Configurator
Table compare

To speed up development
Data Grid
Trim strings
AS400 Date/time

...

* To be released as soon as we've cleaned up some code 13



Why oh why?

Agenda

What is KFF?

Where is this going
and how will we get there?

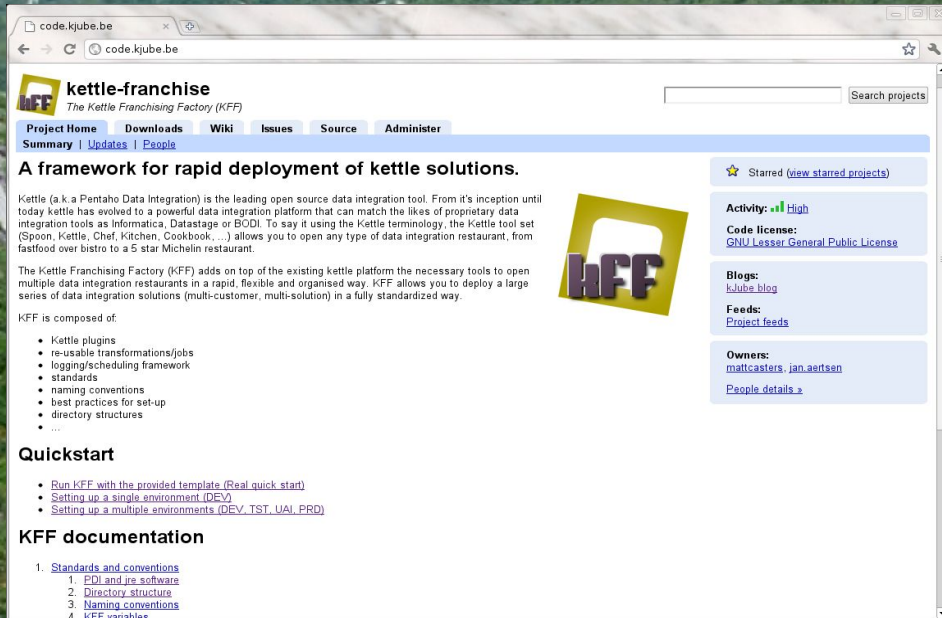
(The) Who?

(What) Do you think?



How will we move forward?

Share share share share



The screenshot shows a web browser window with the URL code.kjube.be. The page title is "kettle-franchise" and the subtitle is "The Kettle Franchising Factory (KFF)". The page content includes a search bar, navigation tabs (Project Home, Downloads, Wiki, Issues, Source, Administer), and a main heading "A framework for rapid deployment of kettle solutions." Below this, there is a description of Kettle and KFF, a list of KFF components, a quickstart section, and KFF documentation links. A sidebar on the right contains metadata such as "Activity: High", "Code license: GNU Lesser General Public License", "Blog: Kettle blog", "Feeds: Project feeds", and "Owners: matcasters, jan_aertsen".

Make KFF install as simple as Kettle install

Add to kettle what belongs in kettle

Make plug-ins if that simplifies the job

Integrate with other kettle projects as cookbook etc ...

Release schedule:
Aug -> Dec 2010

Why oh why?

Agenda

What is KFF?

Where is this going
and how will we get there?

(The) Who?

(What) Do you think?





Jan Aertsen
Managing partner

kJube



Matt Casters
Chief Data Integration

Pentaho



Greta
Marketing



Enthusiast (or unaware)
but always happy
kJube customers

... and hopefully soon also:



Why oh why?

Agenda

What is KFF?

Where is this going
and how will we get there?

(The) Who?

(What) Do you think?

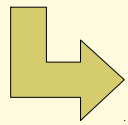


A copy of the presentation?
Need to vent your opinion?
Have marketing ideas for Greta?



SMS us on +32485329828
with the following message structure:

XX your@email.address



X1) Presentation quality? 1=Low 9=High

X2) Usability of KFF? 1=Low 9=High

Mail us at jan.aertsen@kjube.be
and write all the proza you want to share.

*Do not go where the path may lead,
go instead where there is no path
and leave a trail.*

- Ralph Waldo Emerson (1803 - 1882)

Thank you.

