Inhaltsverzeichnis	1
Kottle Grundlagen	
Nettle-Grundlagen	∠
Kettle-Grundlagen	2
Zielsetzung	2
Geschichte von Kettle	2
Alternativen	2
Installation und Einrichtung	2
Voraussetzungen	2
Download und Installation	2
Start nach der Installation	2
	- 3
Angemeines zu opuon	3
Spoon unter Windows	3
Spoon Einstellungen	3
Kettle und Spoon in der Praxis	4
Vokabular	4
Hallo Welt	4
Hallo Welt Design	4
Hallo Welt Protokollierung	5
Hallo Well to Etelalasgabe Hallo Well to Evel	0 6
Hallo Welt als Job	7
Hallo Welt im Batch-Modus	7
Hallo Welt Logging und Fehlerbehandlung	7
Daten aus Excel-Dokumenten uploaden	9
Projekt: Absolvent_innenlisten nach Fachbereich in Excel-Vorlagen schreiben	11
Aufbau Gesamtjob Absolvent_innenliste	11
I ransformation der Absolvent innen Absolvent insenliste Selektion der Besiedeten	11 19
Absolvent innenliste Aufbereitung der Daten	13
Absolvent_innenliste Excel-Export	13
Absolvent_innenliste PDF-Erzeugung	14
Absolvent_innenliste mit Schleife über Fakultäten	15
Autoau desamiyo Ausoveniemise ini Gomene Fakultäen / Fachbereiche emittelin	15
Absolvent_innenliste pro Fachbereich generieren	15
Absolvent_inneniste Expet Expert pro Fachbereich PDF-Expert for Fakultäten	16
Das Ergebnis	18

# Kettle-Grundlagen

# Kettle-Grundlagen

# Zielsetzung

Das Anwendungsspektrum vom OpenSource-Werkzeug zum Datenmanagement Kettle / Pentaho Data Integration reicht vom Kopieren einzelner Daten zwischen zwei Quellen bis zum professionellen "Data Warehousing". Für diesen Zweck bietet das Werkzeug eine mächtige Funktionalität und graphische Oberfläche. Die Nutzung des Werkzeugs wird unten anhand von praktischen Beispielen aus der Hochschulverwaltung vorgestellt.

# **Geschichte von Kettle**

Zeitpunkt	Was passierte
Sommer 2001	Der Belgier Matt Casters gründet eigenes Unternehmen als BI-Consultant
Ab 2003	Start der Arbeiten an Kettle. Ursprünglich war geplant, Kettle als KDE-Anwendung zu implementieren, daher der Name KDE Extraction, Transportation, Transformation and Loading Environment.
2004	Kettle Version 1.2 mit SWT statt AWT. Das belgische Verkehrszentrum setzt Kettle ein
2005	Kettle 2.0 wird veröffentlicht, mit Plugin-Architektur und SAP-Plugin. Matt Casters stellt Kettle unter OpenSource-Lizenz. In der ersten Woche hat Kettle 35.000 Downloads
2006	Pentaho wechselt von Octopus zu Kettle, Matt Casters wird angestellt. Kettle wird zu PDI (Pentaho Data Integration)
2010	Kettle / PDI 4.0 Release
2015	Pentaho wurde von Hitachi Data Systems aufgekauft.
2022	Der Kettle Fork "hop" wird ein Top-Level Produkt der Apache Foundation.

Wir sehen dass die Software eine "bewegte" Geschichte hat. Durch die ganzen Namensänderungen bleiben wir erstmal beim Namen "kettle". Weitere Details siehe Wikipedia-Eintrag.

Derzeit ist Kettle ein Produkt von Hitachi Vantara.

# Alternativen

Eine Übersicht finden Sie auf Wikipedia:



Kommerzielle Produkte im DWH-Bereich:

IBM InfoSphere

Informatica

OSS-Tools

- TalendPDI / Kettle
- Octopus

# Installation und Einrichtung

## Voraussetzungen

- · Kettle läuft unter Windows und Linux bzw. anderen Systemen mit Java Runtime
- Für Kettle Version 6-8.3 wird Java 8 vorausgesetzt, und zwar nur genau diese Versionen. Ältere Java Versionen und Java 11 werden dort nicht unterstützt.
- Ab Kettle 9.3 wird auch Java 11 und Java 17 unterstützt.

Entgegen der Dokumentation funktioniert auch ein Open Source Java, z.B. openjdk, Sie müssen nicht Oracle Java nehmen. Zur Installation von OpenJDK siehe unsere Java-Installationsanleitung.

# **Download und Installation**

Kettle oder PDI (Pentaho Date Integration) steht in seiner "Community Edition" unter Apache License V.2.0. Das konkrete Programmpaket heißt "pdi-ce-Versionsnr.zip" (Abkürzung für Pentaho Data integration Community Edition).

- Es gibt einen Community Bereich bei Hitachi, wo Sie die Software herunterladen können: https://www.hitachivantara.com/en-us/products/pentaho-platform/data-integration-analytics/pentahocommunity-edition.html
- Die Version 8.3 passt zur HISinOne-BI ab Version 2021.06
   Quellcode erhalten Sie auf github: https://github.com/pentaho/pentaho-kettle

# Start nach der Installation

Kettle beinhaltet die Anwendung Spoon (graphische Entwicklungsumgebung), sowie die Anwendungen kitchen und pan. Letztere erlauben den scriptgesteuerten Aufruf von Jobs bzw.

Transformationen. Die Anwendung carte ermöglicht den Remote-Einsatz von Kettle.

Am besten beginnen Sie mit der graphischen Entwicklungsumgebung "Spoon".

## Allgemeines zu Spoon

Das Kommandozeilen-Script "spoon.sh" bzw. "spoon.bat" liegt im obersten Verzeichnis der Auslieferung. Wenn Java 8 oder 17 installiert ist, reicht es die Datei mit Doppelklick zu starten.

Aus der Kommandozeile starten Sie mit

spoon.bat /norep

oder

spoon.sh /norep

"/norep" setzt den Parameter, dass Kettle nicht nach einen Repository Server sucht

#### Spoon Besonderheiten für Linux

Unter Linux sucht Spoon zunächst nach einer passenden Java-Runtime: entweder Sie setzen die Variable JAVA\_HOME / JRE\_HOME, oder Sie installieren oberhalb von spoon.sh im Verzeichnis java eine passende runtime. Wenn Sie nur eine spezielle Runtime verwenden wollen, setzen Sie z.B.

PENTAHO\_JAVA\_HOME=/usr/lib/java export PENTAHO\_JAVA\_HOME

Wenn sie nach dem Start die Meldung bekommen:



Dann müssen Sie zunächst das Paket libwebkitgtk installieren.

Z.B. unter Ubuntu Linux:

apt-get install libwebkitgtk-1.0-0

Danach starten Sie Spoon neu.

Hinweis: Spoon stürzt unter manchen Linux-Versionen ab sobald man auf "Preview" klickt oder eine Transformation starten will.

vermutlich ist die SWT Bibliothek in Spoon nicht kompatibel mit dem Betriebssystem.

Besorgen Sie sich eine swt-Datei z.B. aus dem aktuellen Eclipse-Paket. Es liegt im Ordner "plugins", ermittelbar dort mit

ls org.eclipse.swt\*jar

Diese Dateien kopiert man ins Spoon Verzeichnis nach

libswt/linux

Ursache:

Lösung:

(jeweils x86 und x86\_64), die alten Dateien benennt man um nach swt.old. Danach startet man Spoon neu.

## Spoon unter Windows

Unter Windows installieren Sie zunächst Java, und setzen dann in der Systemsteuerung die Umgebungsvariable JAVA\_HOME auf den Pfad, wo Java installiert ist, z.B.

C:\Program Files\Java\jre1.8.0\_45

Probieren Sie den Start mit Doppelklick auf die Datei spoon.bat. Bei Startproblemen unter Windows öffnen Sie eine DOS-Box ("Eingabeaufforderung") und gehen mit cd in das Verzeichnis, wo Kettle installiert ist (z.B. C:\Users\superx\pdi-ce-7.0.0.25\data-integration).

Führen Sie in der DOS Box das Kommando aus:

set JAVA\_HOME=C:\Program Files\Java\jre1.8.0\_45

Editieren Sie dann die Datei spoon.bat und ersetzen Sie den Aufruf "javaw" durch "java". Achtung: das geht nicht mit Windows Notepad, weil es die Unix Zeilenumbrüche nicht versteht, Sie müssen einen alternativen Editor. z. B. Windows WordPad nehmen

und starten Sie das Script

#### spoondebug.bat

Beantworten Sie alle Fragen mit "Y", am Ende wird eine Datei

SpoonDebug.txt

geschrieben, der Sie diagnostische Meldungen entnehmen können. Wenn z.B. das RAM nicht ausreicht, müssen Sie die Datei spoon.bat editieren, und den Passus "-Xmx2048m" ersetzen z.B. durch "-Xmx1024m".

Weitere Hilfen bei Startproblemen siehe den FAQ.

## Spoon Einstellungen

Als erstes sollten Sie die Sprachumgebung von Spoon einstellen:

Im Menü Tools -> Options wählen Sie im Reiter "UI Settings" die deutsche Umgebung:

Opnoral DOOK & Feel				
	Fixed width fort	Courier - 9	/	0
	ant on workspace:	Microsoft Sans Serif - 0	1	0
	Fast for notes:	Microsoft Sans Serif - 8	/	0
	Background color:		1	0
Workspan	background color:		1	0
	Tab polor:		1	0
loor	size in workspace: 22			
Linev	idh on workspace: 1			
Shadow	size on workspace: 4			
Dialog	middle percentage: 36			
	Oxid size 1			
Ca	rvas arti-aliasing?			
	Use look of 057			
Sho	chanding graphics			
	Nerred Language: Double	1		×
A8	erndve Language: English	លន		v

Achten Sie darauf, daß Sie bei "Alternative Language" English angeben, damit Menüs, die noch nicht übersetzt sind, in Englisch angezeigt werden.

# Kettle und Spoon in der Praxis

## Vokabular

Zunächst: Kettle unterscheidet Jobs und Transformationen. Jobs sind Sammlungen von Aktionen in einer definierbaren Reihenfolge, und mit Kettle ausführbar. Transformationen sind die Bausteine von Jobs, und dienen der Datenverarbeitung. Innerhalb von Jobs und Transformationen sind die einzelnen Aktionen sog. "Steps", deren Reihenfolge wiederum durch Verknüpfungen ("hops") organisiert wird.

Jobs und Transformationen werden lokal im XML Format gespeichert, die Endung ist jeweils "\*.kjb" fürs Jobs und "\*.ktr" für Transformationen.

## Hallo Welt

## Hallo Welt Design

Unsere Beispiel-Transformation erzeugt 10 Zeilen "Hallo Welt", die am Bildschirm angezeigt werden, sowie als Datei gespeichert werden. Das Beispiel soll die Grundprinzipien von Spoon veranschaulichen.

Starten Sie Spoon, und machen Sie einen Doppelklick auf "Transformations", und wählen Sie unterhalb von "Input" den Schritt "Generate Rows". Ziehen Sie die Zeile in den Entwurfsbereich. Machen Sie dann einen Doppelklick auf das Icon. Danach konfigurieren Sie den Schritt:



#### Das Icon ändert sich dann im Entwurfsbereich:



Danach ziehen Sie unter "Flow" den Schritt "Dummy" in den Entwurfsbereich. Dann markieren Sie den Schritt "Erzeuge 10 Nachrichten", und halten die Shift-Taste gedrückt. Dann ziehen Sie von dort einen Pfeil zum Dummy-Schritt:



Dann speichern Sie die Transformation mit dem Namen "hallo\_welt.ktr" und markieren im linken Menü die Transformation, und drücken die rechte Maustaste. Sie können dann einen Namen und einen Beschreibungstext vergeben:



Dann markieren Sie den Dummy-Step, und klicken Sie auf das Preview Icon:



#### Es erscheint ein Dialog:

<b>€</b> ⊖	Transformation debug dialog	000
Ecoupe 10 Nachristee	Number of revea to relative 1000	
Dureny (do nothing)	Rathere first raws (preview)	
	Pause transformation on condition	
	Broak point/ pause condition	۲
	Cear <	>
	Quick Lauroh Configure Carnel	

## Klicken Sie hier auf "Quick Launch". Es erscheint das Ergebnis:

0	Examine preview data 🛛 😒 🔿 🛞
Rows	of step: Dummy (do nothing) (10 rows)
^	#: Nachricht
	1 Hallo Welt!
	2 Hallo Welt!
	3 Hallo Welt!
	4 Hallo Welt!
	5 Hallo Welt!
	6 Hallo Welt!
	7 Hallo Welt!
	8 Hallo Welt!
	9 Hallo Welt!
	10 Hallo Welt!
L	
	Close

## Hallo Welt Protokollierung

Machen Sie einen Doppelklick auf den Schritt "Erzeuge 10 Nachrichten", und erhöhen Sie das Limit von 10 auf 10.000.

Klicken Sie dann auf "OK", und zeigen Sie über das Menü "View"-> "Execution Results" die Ausgabe an. Wählen Sie dann die Registerkarte "Logging". Klicken Sie dann in der Symbolleiste auf das "Run"-lcon:



Wählen Sie bei "Log Level" den "Row level":

	Execute a tr	ansternation	000
Lecal, remote or clustered executiv	64		
<ul> <li>Letal election</li> </ul>	<ul> <li>Excuta renot</li> </ul>	ely Execute Mathematic	
	Remote heat	👻 🕑 Pest tanakemais	
	Pass conth	remoto samor	
	<u> </u>	2 Net excellent	
		The reg of previous executions is cleared before the execution starts. Nestformati	ions
Details			
		Crable sale mode	
		Constant and percent seconds.	
Lagreen		Howkevel being detailed;	~
Replay data Investments	0	Manufaceul Deiny delta fic C	×
Replay data (population of the second	10	th burdervell freiny contrains fil	×
Poplay data (poplakild Hormons) Paramoters	10	jihawaana karrokana kata Veri shive	×.
Ripley dels (con/RM/dd Hrismes Pasanolies	Default take	Verlabies	Value
Poplay data booy/MM/64 Hitsminis Pasamolius A # Pasamolor Value 1	n) Default rake	Verlables Verlables  V	Value Parett Job Fil
Paping dak (confident det errers) Pasiereiters A # Pasiereiter Value 1	Detail salue	Verables Verables Verables Verables Verables Verables I procession Pressence Directory 2 trocssion Pressence Darace	Value Parcet Job Fit Parcet Job Fit
Replay dala (populitiki) di Hritores s Pantenolors	Detaat salue	Versibles           N         Variables           N         Variables           Important for an and participation of the state of	Value Parcet Job Fit Parcet Job Fit Parcet Job No
Nipley dak (pop/MMId H1mms) Patembes A # Patember Value 1 Nopenetts A # Argement Value	n) Detait take	Verables  verabl	Value Parett Job Fill Parett Job Fill Parett Job Ra Parett Job Ra
Replay dels boolinitid Pfrancs Paramolers A # Parameter Value 1 Noernetis A # Agameti Value 5	Oetsat solue	Weinberg Horz Konzellen     Verlachen     Verlachen     Verlachen     Verlachen     Verlachen Zwarz Krannen Directory     Verlachen Zwarz Krannen Zwarz     Verlachen Zwarz Krannen     Verlachen ZwarZeine Zwarz     Verlachen ZwarZeine Zwarz	Value Parett Job Fil Parett Job Re Parett Job Re
Representation of the second s	no Detaart solue	Resident Anapping State (     Vertable	Value Parent Job Pa Parent Job Re Parent Job Re
Augureen Richter de Bossenhildel Helsminis Parameters A # Parameter Indue 1     Noemeters A # Apparent Value 5	no) Default naker	Presentation and Constants  Versibles  Versibles  Versibles  Versibles  Versibles  Versibles Versibles Directive	Value Parent Job Pin Parent Job Pin Parent Job Pin Parent Job Re

Dann klicken Sie unten auf "Launch". Die Transformation wird gestartet, und im Register Logging sehen Sie ein Protokoll:



#### Neben dem Protokoll können Sie auch eine Ausführungsstatistik aufrufen:



# Hallo Welt Dateiausgabe

Speichern Sie obige Transformation als Datei "hallo\_welt\_dateiausgabe.ktr" und löschen Sie den Dummy-Schritt im Entwurfsbereich.

Wählen Sie dann unter Output den Schritt "Text file output", und ziehen Sie den Schritt in den Entwurfsbereich. Danach können Sie den Schritt "Erzeuge 10 Nachrichten" damit verbinden. Machen Sie dann einen Doppelklick auf den Text File Output Schritt, und vergeben Sie den Namen "Hallo Welt Ausgabe".



Bei den Eigenschaften des Schrittes Ausgabe können Sie nun einen Dateinamen angeben. Wählen Sie hier keinen absoluten Pfad, sondern drücken Sie im Feld die Taste STRG-Leerzeichen. Es erscheint eine Auswahlliste, wo Sie den Pfad

\${Internal.Transformation.Filename.Directory}

auswählen. Ergänzen Sie dann noch den Dateinamen "hallo\_welt":



# Hallo Welt in Excel

Für den Export nach Excel erzeugen Sie eine neue Transformation hallo\_welt\_dateiausgabe\_excel.ktr



Die Eigenschaften des Excel Exports sind selbsterklärend.

## Hallo Welt als Job

Wählen Sie in Spoon im Menü "New" den Eintrag "Job". Drücken Sie dann CTRL-j, um die Eigenschaften des Jobs zu definieren:



Im Entwurfsbereich ziehen Sie zunächst unter "General" das Icon "Start" herein. Ziehen Sie dann unter "General" das Icon "Transformation" in den Entwurfsbereich, und verbinden Sie beide Schritte:



In den Eigenschaften der Transformation wählen Sie im Feld "Transformation Filename" wieder STRG-Leerzeichen, und wählen die Variable "\${Internal.Job.Filename.Directory}" und als Unterverzeichnis den Dateinamen der Transformation:



Sie müssen den Job dann im gleichen Verzeichnis wie die Transformation speichern. Mit dem "Run"-Icon können Sie den Job testen.

## Hallo Welt im Batch-Modus

Um den Hallo-Welt-Job über die Kommandozeile aufzurufen, wählen Sie folgenden Befehl:

kitchen.sh /file:<

Damit wird der Job über die Kommandozeile ausgeführt. Der Schalter "norep" bewirkt, daß nicht versucht wird, eine Repository-Datenbank zu nutzen.

## Hallo Welt Logging und Fehlerbehandlung

Am Beispiel des HalloWelt-Jobs wollen wir die Fehlerbehandlung erläutern:

Jeder Schritt in einem Job kann einen Ausgabe- und einen Fehlerkanal ansteuern. Die Art des Sprungs kann variiert werden. Hier ein Beispiel:



Die Transformation "Hallo Welt Transf." kann einen Fehler enthalten. In diesem Fall wird eine Mail an den Schritt Mail->Mail versendet. Wir nennen diesen Schritt "Fehlermail". Der Sc

- Im Reiter Addresses und Server geben Sie die Adreßdaten und dem SMTP-Server an.
- Im Reiter "EMail Message" können Sie Inhalt und Betreffzeile spezifizieren:

1 G	Details des Job-Eintrages (Mail)	80
Name des Job-Eintrages:	Fehremalt	
Accresses Server EMaTNOSSage Atta	ched Files	
Message Settings		
Datum in Nachricht einbeziehen?	1	
Nur Kommertar is Nachricht senden?		
Use HTML format in mail body?		
Encoding	UTF-8	~
Managa priority		
Priority	Normal	*
Importance	Normal	~
Message		
Betreff	Ketle mail Fehler	
Kammentar		8
		U.
	<	
	gK Babrechen	

#### Wir wählen im Betreff "Kettle mail Fehler".

Im Reiter "Attached Files" wählen wir die Logdatei vom Typ "Allgemein":

	Ostalis des Job-Eintrages (Mal)	00	8
Name des Job-Eintrages:	Fehlermail		٦
Addresses Server EMail Message (Atact	wd Fies		
Files added in result filesname			
Log-Datei(en) an Nachricht anhängen	~		
Dateityp auswählen:	Alastish		
	Lig Peblorailon Peblor Wamangen		
Dateien in Zip-Archir packen?			
Name des Zip-Archives:		*	
Embedded images			
Filmana	Add Browset	les.	
ContentID	•	_	
Embedded images			
▲ # Image : Content ID			
1		Davida	
		Edit	
l	gic Akkrechen		

Man kann bei Dateityp auswählen auch mehrere Zeile markieren.

Den normalen Logging-Kanal konfigurieren wir ebenfalls als Mail-Schritt. Dann produzieren wir einen Fehler in der Transformation, z.B. daß die Zieldatei nicht geschrieben werden kann



## Daten aus Excel-Dokumenten uploaden

Mit Kettle lassen sich elegant Excel-Dateien hochladen und somit beliebige Daten in die Datenbank integrieren. Der Vorteil gegenüber dem CSV-Format ist, dass Sie bspw. Formeln n

Der Kettle-Job besteht lediglich aus einem Start- und Success-Step, sowie einer Transformation. Trotzdem benötigen wir einen Job, da nur Jobs im System ausgeführt werden können.



#### Die Transformation besteht aus einem "Microsoft Excel Input"-Step und einem "Table Output"-Step.



Im "Microsoft Excel Input"-Step ist es wichtig im Reiter Files die Variable \${PATH\_TO\_UPLOADFILE} anzugeben. Diese muss zuvor dem Job als Parameter bekanntgemacht werden.

Charles and a first start	and the set for the set of the se		с
	D,	Microsoft Excel input – o	×
	Step name	Microsoft Excel Input	
	Files Sheets Content Error Handling Fields Additional out	put fields	
R.	Spread sneet type (engine)	Excel 2007 XLSX (Apache POI)	1
Microsole Excertingue	Regular Expression	Add Browse.	**
	Exclude Regular Expression	·	
	Selected files:	File/Directory     Wildcard (RegExp) Exclude wildcard Required Include	
		1 \${PATH_TO_UPLOADFILE} N N Delet	ce.
		Edit	
		Accept filenames from previous steps Accept filenames from previous step	
		Step to read filenames from	ŕ
		Field in the input to use as filename	·
		Show filename(s)	
	O Help	OK Preview rows Cancel	

Um Parameter zu setzen Doppelklicken Sie im Job auf das freie Feld.



Im Reiter Sheets wird angegeben wie das Tabellenblatt der Excel-Datei heißt, welches die Daten enthält. Zudem wird die Spalte in die Zeile angegeben, ab welcher die Daten erfasst

D,		Microsof	t Excel inpu	ıt	- 0
Step name	Mi	crosoft Excel Input			
File: Sheets Content Error Handling Fields Additional out	put	fields			
List of sheets to read		Sheet name	Start row	Start column	
	1	prof_bewertung	3	0	

Im Reiter Content wird angegeben, dass die Excel-Datei eine Kopfzeile besitzt, bevor die hochzuladenden Daten folgen. Dies können Sie nutzen, um Spaltenüberschriften zu setzen.

D,	Microsoft Excel input -	e x
Step name	Microsoft Excel Input	
Files Sheets Content Error Handling Fields Additional ou	tput fields	
Header No empty rows		
Stop on empty row		
Encoding	0	
Parult filanamar	UTF-8	$\sim$
Add filenames to resul		

Im Excel beginnt die Kopzeile in Zelle A4. Darunter folgen die hochzuladenden Daten. Die übrige Exceldatei kann beliebig gestaltet werden.

	coups server re-	rusvinger tasers taken terras					
	N X D H -	* * >+C+R*	1 <b>1</b> 1 4 4 4 4 4	S 10 17 13.	• S 9 0	10	<b>1</b>
w	39 H * F	KU・ユ・■・日 平 日	T 🕈 🗄 💷	E E E • %	이번 전	9 I I I -	S - 🖸 - 🗛 -
· · · · · ·							
A		-	0	(	1		
Professo	ren Bew	/ertuna					
Notiz: Note Imm	er etwas bess	er machen.	Faktor	0.9			
Name	Fakultät	Studiengang	Note				
Lenhoff	Fak1	Mathe	2,16				
Plessner	Fak1	Physik.	1,89				
Schorr	Fak1	Informatik	1,98				
Wimmer	Fak2	Humanmedizin	2,79				
Wagener	Fak2	Klinische Pflege	1,98				
Leier	Fak2	Neurowissenschaften	0,9				
1 Ptennig	Fak2	Zahnmedizin	2.52				
Schomberg	Fak3	Archäologie	3.6				
Reisinger	Fak3	Deutsch	1,17				
Mosbacher	Fak3	Englisch	3,33				
Wollschlager	Fak3	Geschichte	1,71				
-							
0							
8							
4							

Im Reiter Fields werden die eingelesenen Felder definiert.

			114	-		-					
					osoft Excelling						
500	Sheets Conte	et Error Handling	Fields Adda	ional output fi	olds						
-	Name	Type	Length	Precision	Trim Type	Repeat	Format	Carrency	Decireal	tirouping	
2	Name	String	-1	-1	none	N					
2	Rekultät	String	-1	-1	0000	N					
3	Studiengang	String	-1	-1	0000	N					
4	Nic6e	Number	-1	-3	0000	N					
		A STATE OF									

Im "Table-Output"-Step wird die Datenbankverbindung und Zieltabelle angegeben. Zudem erfolgt die Zuordnung der Felder aus dem Kettle-Stream, welche zuvor im "Microsoft Excel Inc



## Projekt: Absolvent\_innenlisten nach Fachbereich in Excel-Vorlagen schreiben

Das folgende Projekt an der Uni Wuppertal zeigt, wie man auf der Basis von HISinOne eine Absolvent\_innenliste für ein bestimmtes Semester erzeugt. Kettle schreibt die Daten dabe Um die Schleifenfunktion zu demonstrieren, wird dann diese Liste noch einmal fakultätsweise erzeugt.

# Aufbau Gesamtjob Absolvent\_innenliste



#### Mit Semester als Parameter:



## Transformation der Absolvent\_innen

Hier die gesamte Transformation:



#### Absolvent\_innenliste Selektion der Basisdaten

Wählen Sie einen "Table Input" Step:

E.*	Table input				~ ^ 8
Step name	getAbsolverter				
Connection	Visitone	~	Edit.	New.	Wizard .
squ			GetS	QL selec	statement
	no. 10. exemptition_id.0. i.0 10. 10. exemptition_id.0. i.0 10. 10. exemptition_id.0. i.0				
Replace variables in script?	2				~
Execute for each row?					
Limit size	0				•
OHalp	OK. Preview Cancel				

Der Parameter ist das Semester. Hier der SQL zum Kopieren:

select C.orgunit\_lid as fakultaet\_nr, PE.sumame as nachname, PE.firstname as vorname, S.registrationnumber as mtknr, C.degree\_lid as abschluss/lid, C.subject\_lid as studiengang\_lid, E.date\_of\_work as abschlussatum, N.grade as Gesamtnote, null:decimal as bewertungabschlussarbeit, :varchar as titelabschlussarbeit, :varchar as betreuerabschlussarbeit, DP.studysemester as anzahlsemester, :varchar as plz, :varchar as ort, :varchar when PE.k\_gender\_id=2 then 'Frau' else null::varchar(255) end as Anrede, current\_date as heute from hisinone.unit\_studies US, hisinone.course\_of\_study C, hisinone.degree\_program\_progress DP, hisinone.degree\_program D, hisinone.period P, hisinone.student S. hisinone.person PE, hisinone.examplan E. hisinone.term\_type T, hisinone.unit U , hisinone.examrelation R left outer join hisinone.examresult N on (N.examrelation\_id=R.id) where U.id=US.unit\_id and C.lid=US.course\_of\_study\_lid and E.unit\_id=U.id and E.default examrelation id=R.id and D:dealDP:degree\_program\_id and DP:course\_of\_study\_id=C.id and DP.period\_id=P.id and D.student\_id=S.id and PE.id=S.person id and PE.id=S.person\_id and P.term\_type\_id=E.term\_type\_id and P.term\_year=E.term\_year and S.person\_id=E.person\_id and T.id=E.term\_type\_id and Lid=E.term\_type\_id and U.official\_statistics =1 --Hauptprüfung

and (to\_number(' ' || E.term\_year || T.termnumber,'99999') ) = \${Semester} order by 1,2,3;

# Absolvent\_innenliste Aufbereitung der Daten

Der Name des Abschlusses wird über einen Lookup Step geholt:

					00100070 1	not cosmop				
			Step name	getAb	schluss					
			Connection	hisinor	e		~	Edit.	New	Wizard
		Loo	kup schema	hisino	ne				•	Browne
		U	ookup table	decre						Browse
		En	able carbe?	_						
	keels) to look	un the value/e):	a morn nacio							
-	Table field	Comparator	Field1		Field2					
1	1d		abschlus	s_Nd						
2	valid_from	<=	heste							
3	valid_to	>*	heute							
ah.	es to return fro	m the lookup tabl	e:							
×k ^	es to return fro Field	m the lookup tabl	e: Default	Type						
ak ^ 1	es to return fro Field defaulttext	m the lookup tabl New name abschluss	e: Default	Type None						
ak 1	Field defaulttext Do not p	m the lookup tabl New name abschluss ass the row if the Fail on mult	e : Default ookup fails pie results?	Type None						

Man müßte hier noch weitere Lookups ausführen, z.B. für den Studiengangnamen, oder die Note der Abschlussarbeit.

Danach wird die Ergebnisliste um nicht mehr benötigte Felder bereinigt:



# Absolvent\_innenliste Excel-Export

Im Excel Export wird eine Datei "Absolventenliste.xls" aus der "vorlage.xls" erzeugt:

Step name	ExportVachExcel	
We & Sheet Content		
F80		
Filename	Spreamal Transformation Primame Directory (ribsolventeritiste	Growss.
Exension	vis (Excel 97 and above)	~
Stream XSLX date		
Split every data rows	0	
include stepper in filename?		
include date in filenary of		
Include time in filename?		
Specify Date time format		
Date time format	dd MM yyyr	
	Show Hename(s)	
if satpat file exists	replace with new output Bie	~
Mait for first row before creating file		
Add filenames to result	~	
Sheet		
Shart name (mar. 21 characters)		
Make that the active street	8	
IT STREET ENDER IN ENDERTING	write to existing sheet	×
Protect sheet? (RLS format only)		
Protected by user		
Password		
Territre		
Line termint when creating one files	2	
Yemplatefie	Enternal Taxanformatics Electron Standard Incluse de	A Browne
	and a second sec	•
Concompany www.creating.new soless		
100,000,000	vthige	•
Hide Template Sheet	×	
	(W) Carrol	
() Help	LON. Laborat	

#### Dabei werden die Basisdaten am Zelle A9 eingefügt:

a /	Excel milter Step v1.2 by type-exit.org	× ^
Step name	Report the children	
We & Sheet Contest		
File		
Filename	Spreamal Transformation Primarie Directory; (Non-iventeniste	Ø Growps.
Exercice	via [Depel 97 and above]	~
Stream XSLX data		
Split every data rows	0	
include steps: in filesame?		
include date in Nenares'		
include time in filename?		
Specify Date time format		
Date time format	dd MMX yyyy	
	Show Network()	
if sutput file exists	replace with new output file	~
Wait for first raw before creating file		
Add filenames to result	8	
Sheet		
Shart name (mar. 21 character)		
	STeel1	0
Make this the active sheet	8	
It sheet exists in carput the	write to existing sheet	*
Protect sheet? (RLS format only)		
Pretected by use		•
Password		0
Template		
Use template when creating new files	8	
Y emplote file	Spreamal Transformation Filename Directory/vorlage.vls	Growes
Use template when creating new shoets	2	
T emplate sheet	Volage	•
Hide Templato Sheet	2	
(1) Help	UK Carow	

## Absolvent\_innenliste PDF-Erzeugung

Kettle kann selbst kein PDF erzeugen, man kann aber mit OpenOffice bzw. Libreoffice über Kommandozeile PDF generieren. Dazu wählt man im Job den Step "Scripting" -> "Shell".



## Dies erzeugt die Kommandozeile:

/usr/sbin/soffice --headless --convert-to pdf Absolventenliste.xls

#### Absolvent\_innenliste mit Schleife über Fakultäten

#### Aufbau Gesamtjob Absolventenliste mit Schleife

Der Gesamtjob ermittelt zuerst die Fakultäten, und erzeugt dann in einer Schleife je eine Absolventenliste in Excel. Diese werden dann auf einen Schlag nach PDF exportiert.



#### Fakultäten / Fachbereiche ermitteln

Die Fakultäten werden aus der Datenbank ermittelt:

<u>= / () () () () () () () () () () () () () </u>	table vijaz	× ^ (
5kp nam	4 24LB	
Connection	Distort	V Edt. New. Wirard.
iq.		Ort EQ, select statement.
Line 5 Column 10	-	
ine 5 Column 10 Enable lazy convenin	и 🗆	
ine 5 Column 10 Enable larg conveni Pisplace watables in scop	n - 9 -	
ane 5 Column 10 Enable large conversi Phylice wantables in scrip Ionard data from all	κ - δ	
ine S Column 10 Enable large community Replaces standards in s cop Inward data france in Decande for another	н — р р	~
ine S Golume 10 Enable larg convens Replace valuellen in sop Ineard data from die Denable for eich tron Lard tot	α	

#### Hier der SQL zum Kopieren:

SELECT O.lid as orgunit\_lid,O.uniquename as fb\_nr,O.defaulttext as fb\_name from hisinone.orgunit O, hisinone.k\_orgunittype K where K.id=O.k\_orgunittype\_id and K.hiskey\_id=4 and current\_date between O.valid\_from and O.valid\_to order by 2;

## Absolvent\_innenliste pro Fachbereich generieren

Die Fachbereiche werden zu Parametern kopiert, und die Transformation getAbsolventen wird in einer Schleife über jeden Fachbereich ausgeführt:



Die Transformation bekommt Parameter:

Para	meters :		ź	
-	Parameter	Default Value	Description	
1	Semester			
2	fb_nr			
3	orgunit_lid			

## und der "Table Input"-Step wertet diesen aus:



## Absolvent\_innenliste Excel Export pro Fachbereich

Der Dateiname der Excel-Datei enthält jetzt auch den Fachbereichsschlüssel:

51	top name	audi FB Ni		
le & Sheet Content				
580				
	Féesame	S(Internal Transformation Filesame Directory)/Wbs alvesterilisteS(BL_R)	•	Depwise
	Extension	ts [Ercel 97 and above]	~	
52 years	2757,X data			
Split every	deta raws	0		
Include stepher in	Merara?			
Include date in	ferane?			
Include time in	Mename?			
Specify3ute1	iree format			
2825	ree tormat			
		thow filename(s)		
if outpu	t filo exists	replace with new output file		~
Wait for first row before o Add Siename	reating file   es to result	2		
Sheet				
Sheet name (max, 31 c	(ereschers)	(heat)		-
Maker this the a	ctive sheet	~		
If shoet exists in	output file	write to existing sheet		~
Protect sheet? (XLS fo	ernat.cely)			
Protect	ted by user			
	Password			•
Template				
Use template when creating	g new files (	2		
Те	mplatefile	Sfinternal Transformation Filename Directory//vorlage.xis	0	trouse.
Use template when creating a	ev sheets	2		
Terry	plate sheet	Volage	•	

## Die Fakultätsnummer wird im ersten Excel-Writer geschrieben:

1 /		Excel V	Witer Step v1.2 by type-exit.org			~ ^
	Step name	Excel FB	NI			
le & Sheet Content						
Content options						
	Start writing at cell	cs				•
	When writing rows	OVERATE	e existing cells			~
	Write Header					
	Write Footer					
	Auto size columns					
	Force formula recalculation					
Leave style	es of existing cells unchanged					
When writing to existin	ig sheet					
Start writing at e	ed of sheet (appending lines)					
	Offset by rows	0				
	Danin hy setting arrest lines					_
		0				
	Omit header					
Fields						
+ Name	Type Format			Style from cell	Field title	Header/F
1 fb_nummer	String				fb_name	
0			OK Gannel			
C Help			000000			

Dann die Absolvent\_innenliste.

Die Transformation darf aber nicht gleichzeitig in die Excel-Datei schreiben. Dazu wird ein "Block until steps finish"-Step eingerichtet:



Der zweite Excel Writer ergänzt dann nur noch die Absolvent\_innenliste:

*			Excel Writer Step v1.2 by type-exit.org	~ ^
		Step name	AbsolventenNachExcel	
le & Sheet Co	ontent			
File				
		Filename	\${Internal.Transformation.Filename.Directory)/Absolventenliste\${fb_nr}	Browse
		Extension	xls [Excel 97 and above]	
		Stream XSLX data		
	Spl	lit every data rows	0	
	Include	stepnr in filename?		
	Inclu	de date in filename?		
	Includ	de time in filename?		
	Spec	ify Date time format		
Daid	lie Excel Da	tei zur Laufz	dd MM yyyy	~
bere	ite Exect Du		Show filename(s)	
oviet	tiort		onor inclaime(y)	
CAISI	uert	If output file exists	Use existing file for writing	~
	Wait for first row	/ before creating file		
	Ade	d filenames to result		
Sheet				
	Sheet name (	(max. 31 characters)	Sheet1	
	Make	this the active sheet	<u>ح</u>	
	If shee	t exists in output file	write to existing sheet	×
	Brotost shoet	·	white to existing sheet	
	Frotect snee	Protected by user		
wird nu	ir noch eras	Password	mplate mehr genutzt	0
Template	innoenerge	inzt, kein rei		
<u> </u>	I lea template whe	an creating new files		
L	ose template unt	Template file	Startarnal Transformation Ellename Directory) (variance via	Browse
	a a tamplata wh		Quiterial Hanstoffiator.Frenance.bilectory)/Voltage.bis	
0	se tempiate when	Template sheet		
			Vonage	
	н	lide Template Sheet		

## PDF-Export pro Fakultäten

Da keine Wildcards unterstützt werden, müssen wir ein Shellscript erzeugen, z.B. für Linux:

<u>, x</u>		Execute a shell script	~ ^
	Job entry name:	PDF	
General Script			
	Insert script	D	
	Script file name:	\$(Internal.Job Filename.Directory)/createpdf.x	Browse
	Working directory:	\${Internal.Job.Filename.Directory}	
Logging settings			
	Specify logfile?		
	Append logfile?		
	Name of logfile:		0
	Extension of logfile:		•
	Include date in logfile?		
	Include time in logfile?		
	Loglevel:	Basic	$\sim$
Cop	by previous results to args?		
E	execute for every input row?		
Fields:			
▲ Argument			
1			
Quale		OK Cancel	

Hier der Inhalt des Shellscriptes createpdf.x:

#l/bin/bash soffice --headless --convert-to pdf A\*.xls

## Das Ergebnis

Im Ergebnis erhalten wir eine Dateiliste mit n Absolvent\_innenlisten jeweils in Excel und PDF:

🖙 🖻 absolventenListe 🔕 🖻 Bilder ⊗	
Name	
Absolventenliste122.pdf	
-> Absolventenliste121.pdf	
Absolventenliste09.pdf	
→ Absolventenliste08.pdf	
Absolventenliste07.pdf	
→ Absolventenliste06.pdf	
Absolventenliste05.pdf	
→ Absolventenliste04.pdf	
Absolventenliste03.pdf	
Absolventenliste0203.pdf	
Absolventenliste0203010101.pdf	
Absolventenliste0202.pdf	
Absolventenliste0201.pdf	
Absolventenliste01.pdf	
Absolventenliste122.xls	
Absolventenliste121.xls	
Absolventenliste09.xls	
Absolventenliste08.xls	
Absolventenliste07.xls	
Absolventenliste06.xls	
Absolventenliste05.xls	
Absolventenliste04.xls	
Absolventenliste03.xls	
Absolventenliste0203.xls	
Absolventenliste0203010101.xls	
Absolventenliste0202.xls	
Absolventenliste0201.xls	
Absolventenliste01.xls	
— 🌵 masterjob.kjb	
— ♦ getFB.ktr	
—	
— I getAbsolventen_in_schleife.ktr	
— I masterjob_mit_schleife_fak.kjb	
─>_ createpdf.x	
─ ≻ vorlage.pdf	
vorlage.xls	

Hier eine Vorschau in Excel:

	Ars	dre Einfüg	en Eorma	t Egnas I	Dagen H	ende. F	iffe																
<b>.</b>	1 8	5   <b>5</b>	5 0	хD	٥·	1 I S	• < •	9. A	87		BE	47 <b>I</b> L	$\mathbf{t}_{\downarrow}$	₩   8	2 Ji	3 <b>3</b>	Ω °	: 🖓				1.	
And	2	14	ΒZ	<u>U</u> <u>T</u>	- 📖	E	÷ 3	5	T	F <u>1</u>	\$	% 0,0	1	oq .o		3	۰ ت	≡ • <u>∎</u>		. =	<b>s</b> :	÷.	
de a	A1		∼ f	. Σ =	Bergi	sche Unive	stat wupp	etal															¥
10 10 10 10 10 10 10 10 10 10 10 10 10 1		A	в	c		0	ε			6	н					×		L	M	N		0	P
		Bergisc	he Unit	versität V	Vuppe	ertal								Aust	wertu	ng au	s HIS	POS-C	X				1
Derection:	2 8 4	Abso	vente	en und	Abs	solve	ntinne	en								ir	n						
la OLE-Objek Komment:	5	Fakultät	ા	01												Stand	t:						
Zeichenob	5		_						_			_			_				-	-	_		_
		Hachnarre	Vername	Mikur		bachlass	Stadiença	ng Absch ss-	au Gen	6 19	wertu (	Arbeit	chiuse-	E .		Annahi Sernes	Ansch	rift	PLZ	Ort	hi i	ede	
	8		i F	-	<b>T</b>	-		-	. <u>1</u>		····· 👻	1		-	" 🖃	-	-		<b>-</b>	-	<b>T</b>	<b>T</b>	
	8	Firaus	Isabel	11230527	•	taster gt siktice	1	15.11	2016 1.3					-		3	-		-	-	Fra		-
	8 0 10	Kraus Simon	Jana Jana	11230627	2	laster gi cistos laster gi	1	15.11	2014 1,3					-	. 12	3	-		-	-	Fra	▼ ″	
	8 9 10 11	Kraus Simon	lested Jana	11230527	N N	laster gt sistes laster gt	1	15.11	2016 1.3		un Ç			V IIII		3	-		<b>T</b>	-	Fra	v ,	
-	8 10 11 12	Kraus Simon	Jans	11230527	V S	laster gi siktisk	1	15.11	2014 1.2		un					3	-			*	Fra	v ,	
	6 10 11 12	Rraus Simon	lsabeli Jena	11230627		laster gt pierce haster gt	1	15.11	5014 1'3							3				*	Fra	v ,	
	8 10 11 13 13	Fraus Simon	Jama	11230627	V N N	laster gl SKESK Nater gl	1	15.11.	2014 1.3							5				-	Fra Her	r	
	8 10 11 13 13 14 15 16	Fraus Simon	Jana	11230627	V V	laster gl cierce laster gl	1	15.11	2014 1.2							3				-	Fra Har	r	
	8 9 10 11 13 13 14 15 16 17	Kraus Simon	Jens	11230627		auster gf Distor Inater gf	1	15 11	5014 1.3							5					Fra	·	
asolventon v	8 10 11 13 13 13 13 13 13 15 15 16 17 10 <	Kraus Simon	Jana	11230627	2	auster gf Sister laater gf	1	15.11	2014 1.3							2					Fra	·	

und hier in PDF:

a basishi	olana las		where Davids	and the		Absolv	entenfiste01.pdf - Okula									
n Areicet : i > Nichini	beta beber	iceionen E	d < E	ogen mine Vadialises 🖓 Ve	epiten 🛔	Narigation	() tape [2] Assesti									
Bergisc	he Univ	ersität	Wupperta	ป				1.1	Auswertung aus HIS-POS-GX							
Abso	lvente	n un	d Abso	olventin	nen					im						
/ 1000	- Critic	in an	071000	JIV CITCIT												
Fakultät		01							Stand:							
_																
Nachname	Vorname	Mtkor	Abschluss	Studiengang	Abschluss- Datum	Cesamt -note	Bewertung Titel A Abschluss- Arbeit Arbeit	oschluss-	Betrever Abschluss - Arbeit	Anzahi Semester	Anschrift	PLZ	Ort	Anre		
Kraus	Isabell	1E+07	Master of Science	1	15.11.2016	1,3				3				Frau		
Simon	Jens	1E+07	Master of	1						5				Herr		