

# Jedox und R – Erweiterung von OLAP-Analysen um statistische Funktionen

OSBI Workshop Karlsruhe  
Andreas Fröhlich, Jedox AG

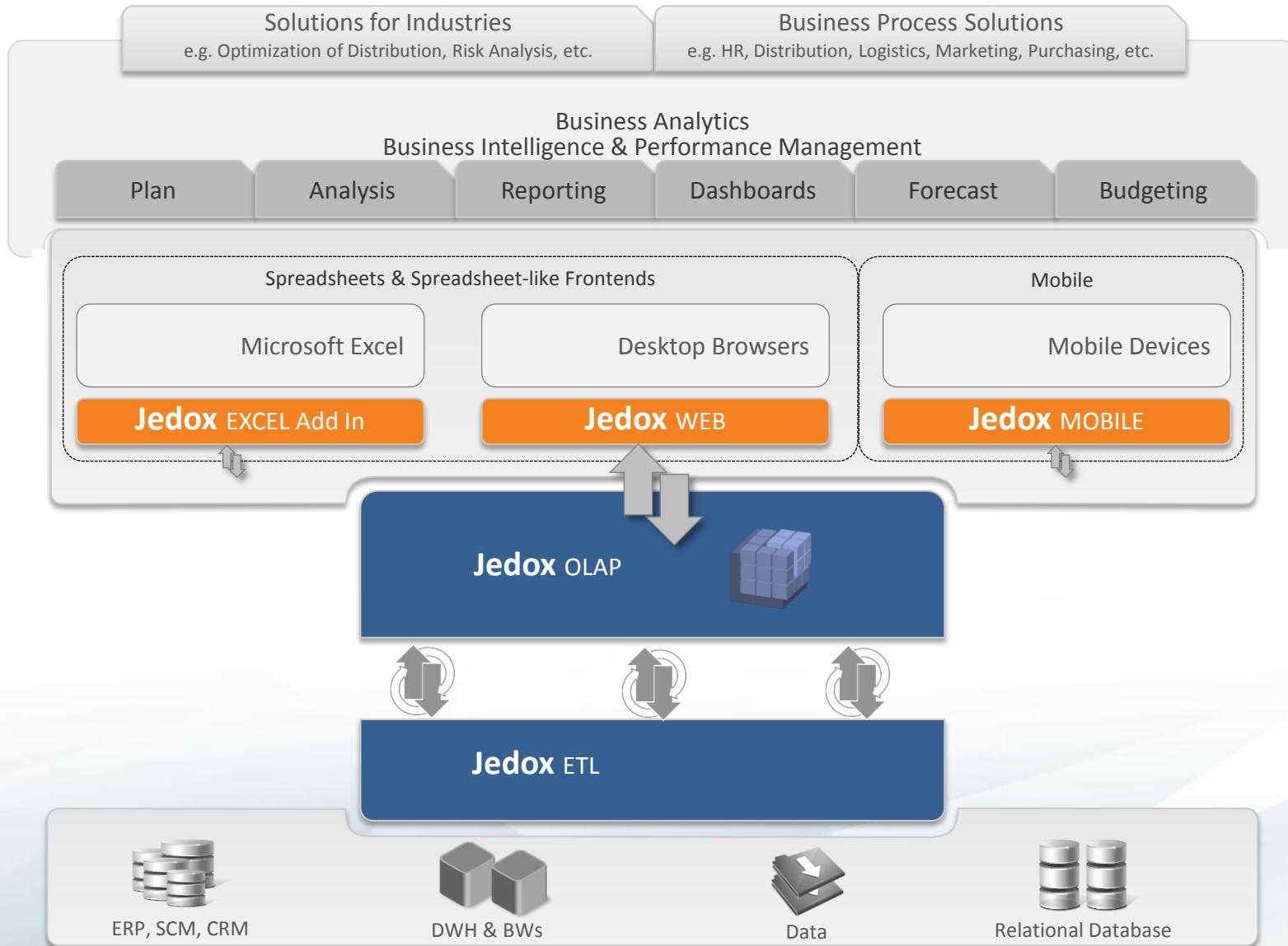
20.02.2014

# Overview

- What is Jedox, what is R ?
- Integration of Jedox and R
- Live Demo
  - Forecast
  - Clustering
- And what about Open Source ?
- Conclusion

# **WHAT IS JEDOX, WHAT IS R ?**

# What is Jedox ?



# What is R ?

- “ R is a free software programming language and software environment for statistical computing and graphics. The R language is widely used among statisticians and data miners for developing statistical software<sup>[2][3]</sup> and data analysis.<sup>[3]</sup> Polls and surveys of data miners are showing R's popularity has increased substantially in recent years. “
- “R provides a wide variety of statistical and graphical techniques, including linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering, and others. R is easily extensible through functions and extensions, and the R community is noted for its active contributions in terms of packages.”

(Source: wikipedia)

# **INTEGRATION OF JEDOX AND R**

# Why integration with R ?

- R allows comprehensive data analysis e.g. in
  - Regression Analysis
  - Time Series Analysis
  - Cluster Analysis
  - Machine Learning
  - Text Mining
- R is open-source
- R is a programming language
- R is independent of User Interface
- R has a JAVA-API (rJava)
- R is integrated by other software vendors (e.g. SAP HANA)

# Statistics and OLAP

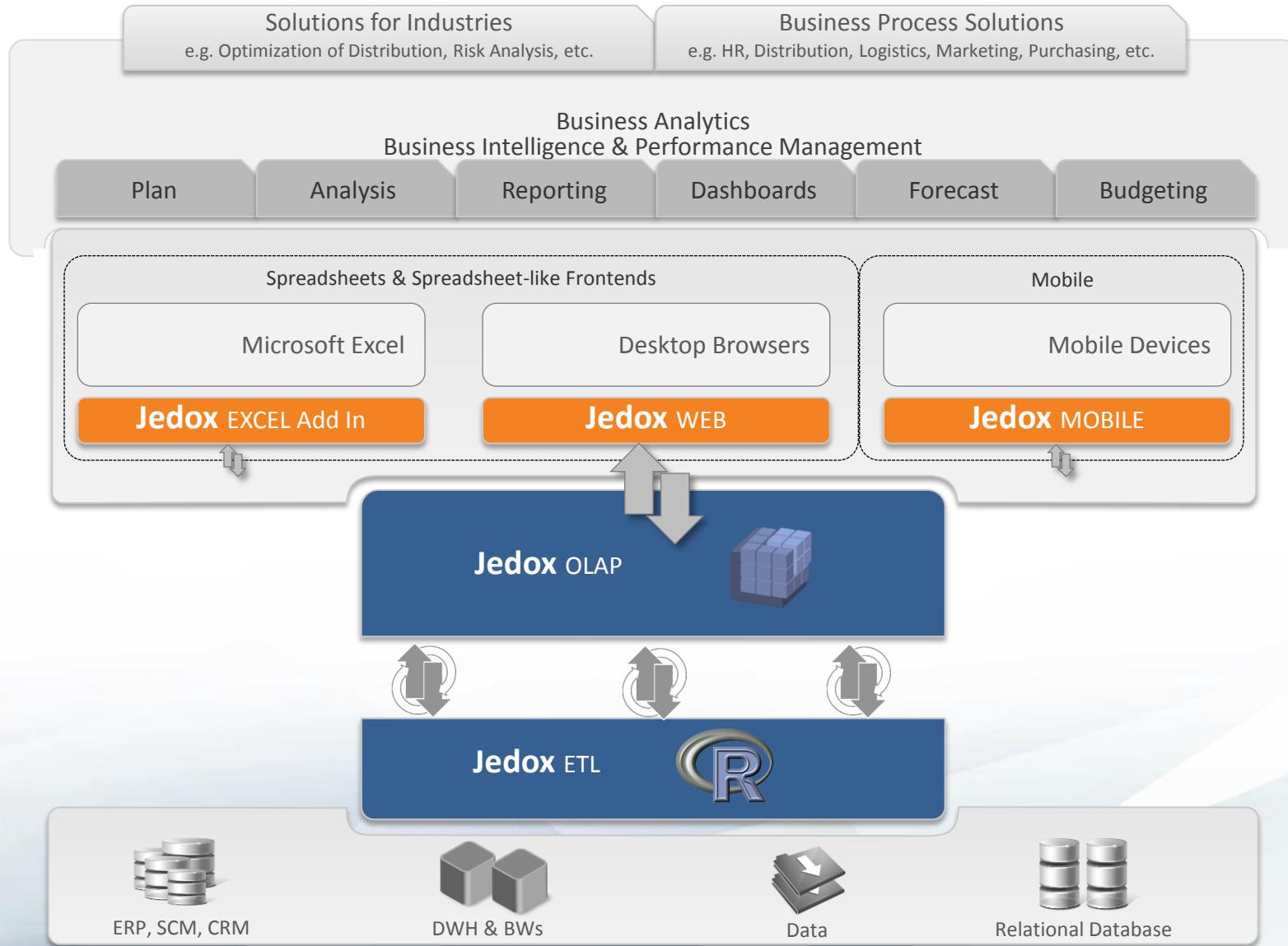
Possible variants of combining Statistical- and OLAP-analysis:

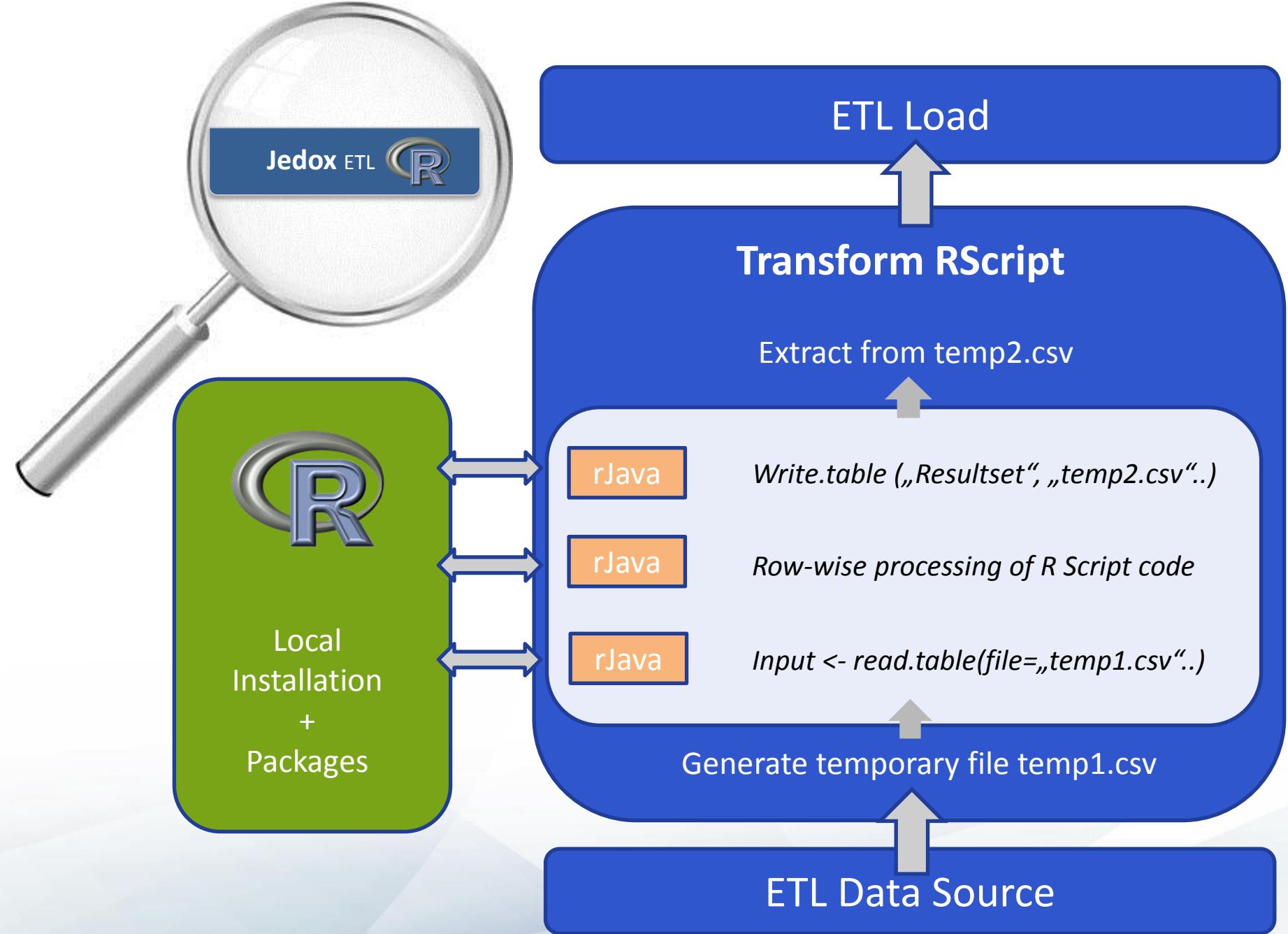
1. Online statistical calculation during OLAP analysis
2. Interactive triggering of statistical calculation during OLAP analysis
3. Scheduled statistical pre-calculation before OLAP analysis



-> Integration of R in Jedox ETL (not in OLAP)

# Integration in the Jedox-Solution





# ETL-Transform RScript

Field for the data source: From type Extract or Transform.

The screenshot shows the configuration interface for an RScript Transform. It consists of three main sections:

- Header:** Shows the Type as "RScript Transform" and the Name as "Duplicate of RS\_descriptive\_statistics".
- General Settings:** Contains a "Data sources" tree view with "E\_Cube" selected. Below it, the "Name of result set" is set to "Result".
- RScript:** Displays the R code:

```
1 data <- E_Cube
2 data_Value <- data$value
3
4 Variance <- var(data_Value)
5 Result <- data.frame(data[,],Variance)
6
```

Annotations with blue arrows point to specific fields:

- An arrow points from the text "Field for the data source: From type Extract or Transform." to the "Data sources" tree view.
- An arrow points from the text "Name of the variable containing the data frame with the results." to the "Name of result set" field.
- An arrow points from the text "Field for the R-Code." to the R code editor.

# LIVE - DEMO

# Forecasting (1)

Jedox Web - Mozilla Firefox  
[localhost:99/ui/wss/?appm=user&prvw&node=fgrp1-h1-n10&sheet=b92d2d96-3111-4e09-9640-2eb8e0930cfb#hrncz3mo](http://localhost:99/ui/wss/?appm=user&prvw&node=fgrp1-h1-n10&sheet=b92d2d96-3111-4e09-9640-2eb8e0930cfb#hrncz3mo)

File Edit View Data Help  
 Refresh XLSX Snapshot XLSX OLAP Snapshot Print Preview PDF Close

R Report ForecastDemo 51 20131216 1100

	Actual		Budget			Planung vs Predictive Analysis					
	2012	2014	Budget R	Lower Bound	Upper Bound	Budget R	Lower Bound	Upper Bound	Budget R	Lower Bound	Upper Bound
		2013	2013	2013	2014	2014	2014	2015	2015	2015	
<b>All Products</b>	56,718,049	0	54,555,562			52,593,518			50,660,453		
<b>Bikes</b>	38,612,203	0	36,910,425	32,882,853	40,937,997	35,268,513	26,304,797	44,232,228	33,626,601	19,286,793	47,966,408
<b>Components</b>	10,046,831	0	9,281,021	8,592,393	9,969,650	8,525,773	6,612,332	10,439,215	7,775,104	4,058,680	11,491,527
<b>Clothing</b>	153,559	0	166,998	160,463	173,533	155,499	148,559	162,439	168,402	158,899	177,904
<b>Accessories</b>	7,905,456	0	8,197,118	6,021,814	10,372,421	8,643,733	4,962,336	12,325,129	9,090,347	4,000,978	14,179,717

**Planung vs Predictive Analysis**

**Bikes**

**Components**

**Clothing**

Job Status Completed successfully

# Forecasting (2)

Navigation

- Report Manager
- File Manager
- OLAP Manager
- Task Manager
- ETL Manager
- ETL Projects
  - Duplicate of sampleR\_Cluster
  - ETLTasks
  - New Project
  - sampleBiker
  - sampleR\_Cluster
  - sampleR\_Cluster dynamic
  - sampleR\_Forecast
    - Variables
    - Connections
    - Extracts
    - Transforms
      - FieldTransform
      - RScript
        - RS\_RForecast\_ARIMA
        - RS\_RForecast\_BestModel
        - RS\_RForecast\_HW
        - RS\_RForecast\_LM
        - RS\_RForecast\_LMwithCycle
  - TableAggregation
  - TableJoin
  - TableLoop
  - TableNormalization
  - TableView
- System Manager

ETL Manager

sampleR\_Forecast

Save | Test | Data Preview | Insert Variable | Layout

Type: RScript Transform

Name: RS\_RForecast\_LM

Description:

General Settings

Data sources

Data source Tree format

E\_Cubedata

Name of result set: lm.summary

RScript

```
1 suppressPackageStartupMessages(require(forecast))
2 data=E_Cubedata
3 data <- data[with(data,order(Years)),]
4 Sales <- data$Value
5 t <- 1:${_InputMonths}
6 linReg <- lm(Sales~t)
7 lm.summary <- summary(linReg)
8 coef <- as.data.frame(lm.summary$coef)
```

Project: sampleR\_Forecast  
Transform: RS\_RForecast\_LM  
14.02.2014 15:47

Transform: RScript  
RS\_RForecast\_LM

Extract: Cube  
E\_Cubedata

Connection: JedoxGlobal  
OlapDatabase

```
graph TD; C[Connection: JedoxGlobal OlapDatabase] --> E[Extract: Cube E_Cubedata]; E --> T[Transform: RScript RS_RForecast_LM]
```

# Clusteranalyse (1)

File Edit View Data Help

Refresh XLSX Snapshot XLSX OLAP Snapshot Print Preview PDF Close

Clusteranalyse

## Clusteranalysis

Version: C5 Kmeans Clustering on Customers (Partitions predefined) predefined Clusters: 5 estimated Clusters: 5 Avg Silhouette Width: 0.463

Cluster! delete Cluster!

View: All Cluster

Scatter plot showing customer data points clustered into 5 groups (Cluster 1 to Cluster 5). The x-axis ranges from 0k to 20k, and the y-axis ranges from -10000 to 20000.

Legend: All Cluster, Cluster 1, Cluster 2, Cluster 3, Cluster 4, Cluster 5

Job Status: Completed with Warnings

Table:

Year	Customer	Sold Units	Gross Profit	Sales
2010	Bike Universe	3,030.95	83,823.86	327,259.63

Line chart showing quarterly performance metrics over four quarters (Qtr.1 to Qtr.4).

Legend: Sold Units, Gross Profit, Sales, Cost of Sales

# Clusteranalysis (2)

Navigation

- Report Manager
- File Manager
- OLAP Manager
- Task Manager
- ETL Manager
- ETL Projects
  - Duplicate of sampleR\_Cluster
  - ETLTasks
  - New Project
  - sampleBiker
  - sampleR\_Cluster
  - sampleR\_Cluster dynamic
  - sampleR\_Forecast
    - Variables
    - Connections
    - Extracts
    - Transforms
      - FieldTransform
      - RScript
        - RS\_RForecast\_ARIMA
        - RS\_RForecast\_BestModel
        - RS\_RForecast\_HW
        - RS\_RForecast\_LM
        - RS\_RForecast\_LMwithCycle
  - TableAggregation
  - TableJoin
  - TableLoop
  - TableNormalization
  - TableView
- System Manager

ETL Manager

sampleR\_Forecast

Save | Test | Data Preview | Insert Variable | Layout

Type: RScript Transform  
Name: RS\_RForecast\_LM  
Description:

General Settings

Data sources

Data source Tree format

E\_Cubedata

Name of result set: lm.summary

RScript

```
1 suppressPackageStartupMessages(require(forecast))
2 data=E_Cubedata
3 data <- data[with(data,order(Years)),]
4 Sales <- data$value
5 t <- 1:_InputMonths
6 linReg <- lm(Sales~t)
7 lm.summary <- summary(linReg)
8 coef <- as.data.frame(lm.summary$coef)
```

Project: sampleR\_Forecast  
Transform: RS\_RForecast\_LM  
14.02.2014 15:47

Transform: RScript  
RS\_RForecast\_LM

Extract: Cube  
E\_Cubedata

Connection: JedoxGlobal  
OlapDatabase

```
graph TD; A[Connection: JedoxGlobal  
OlapDatabase] --> B[Extract: Cube  
E_Cubedata]; B --> C[Transform: RScript  
RS_RForecast_LM]
```

# And what about Open Source ?

## Components of Jedox - R Integration :



- R

### • Jedox Backend



- Jedox Olap Server
- Jedox ETL Server



### • Jedox Frontend

PREMIUM

- Jedox Web

PREMIUM

- ETL Manager

FREE

- Jedox Base



- Open Office Plug-In (PalOOCa)

# Conclusion

Integration of Jedox and R is interesting...

- For current Jedox users :
  - Variety of algorithms for statistics and data mining
  - Applications e.g. for Time-Series analysis and clustering
  - Interactive and pre-calculated analysis possible
  - Neat integration in Jedox Suite
- For current R users :
  - Browser as frontend
  - OLAP analysis on results of R calculation
  - Access to new source- and target-systems via Jedox ETL
  - Scheduled calculations
  - R-calculations „in the loop“ on set of input data