What’s New

Bernd Wiswedel
KNIME
Outline

• High-level motivations
• “What’s new“ presented in two use cases, presented by the team

• Questions/Discussions/Concerns: Find us!
• Demo booth in the back
Remember this?

Business Analyst

Data Scientist

Some Interface

MoreExplanations

Start here

Analytics with Interaction Points

Feedback
Visual Data (and Tool!) Blending
What’s new - Agenda

• Too much to show it all.
• Some highlights:
  – User Experience
  – Text Processing
  – Analytics
  – KNIME Server
  – Cloud Offerings
  – Guided Analytics

  Use Case 1: IMDB Reviews
  Use Case 2: Census Data
What’s new - Agenda

• Too much to show it all.
• Some highlights:
  – User Experience
  – Text Processing
  – Analytics
  – KNIME Server
  – Cloud Offerings
  – Guided Analytics

Use Case 1: IMDB Reviews
Use Case 2: Census Data
User Experience

– Iris Adä –
Workflow Coach!

NEW since 3.2: Workflow Coach recommends matching nodes.

- Three options
  - Based on community statistics
  - Analyze your workspace*
  - From local server**

* Requires KNIME Personal Productivity License
** Requires KNIME Server License
Curved Connections
Automated installation of features

• If you open a workflow with a missing node
  
  • The correct plugin is automatically proposed for installation
Easier import/export of workflows

New File Extension for workflows and groups

Use the KNIME protocol for starting workflows
Node Guide

Search for examples

Data Loading Example

This workflow shows how to use the File and CSV Reader nodes to read text files. The resulting data tables are joined and written out to a CSV file.

Text File Reading, Joining and CSV Writing.
This workflow reads two text files, joins them and writes the result out as another text file.

File Reader
read workflow local file
Joiner
Column Ressorter
CSV Writer
write file out to workflow location.

Resources
EXAMPLESERVER:
"01_Data_Access/01_Common_Type_Files/01_Data_Loading_Example"
Download a zip-archive

KNIME protocol
Open Workflow locally

Download as zip file
Three new helpers!!

- Generate a file path based on folder, file name and extension type
- Simple changes to string based variables
- ... and for number variables as well.
Text Processing
– Kilian Thiel –
Example Use Case

- Sentiment classification of movie reviews
  - PDF parsing
  - Feature vector creation
  - Predictive modeling
  - Semantic analysis of word vectors (tomorrow)
Example Use Case

Goal

• Build predictive models to predict sentiment labels of reviews, “positive” or “negative”.

• “Ah, Moonwalker, I'm a huge Michael Jackson fan …”

• “This film has a very simple but somehow very bad plot. …”

Positive or negative?
Data Set

• The Large Movie Review Dataset v1.0
  – English movie reviews
  – Associated sentiment labels “positive” and “negative”
  – http://ai.stanford.edu/~amaas/data/sentiment/

• 2000 PDFs
  – 1000 positive reviews
  – 1000 negative reviews
Textprocessing

- Tika integration
- StanfordNLP NER Learner & Tagger nodes
- Document Vector Hashing
- Word2Vec integration
Tika Integration

- Tika Parser
- Tika Language Detector
- Tika Parser URL Input
StanfordNLP NER Integration

- StanfordNLP NER Learner
- StanfordNLP NER tagger
Document Vector Creation

• Document Vector Hashing
• Document Vector Adapter
• Word2Vec
Document Vector Creation

- Document Vector Hashing
- Document Vector Adapter
- Word2Vec
Analytics

• Deep Learning in KNIME: DeepLearning4J Extension

• Tree Ensembles and Gradient Boosted Trees

• (Forward) Feature Selection
Analytics

Deep Learning in KNIME: DeepLearning4J Extension
What is Deep Learning?

- State-of-the-art algorithms for learning tasks on images, videos, text or sound
- Multi-Layer Neural Networks
- Regression and Classification, Unsupervised Learning, Reinforcement Learning, ...
What is DeepLearning4J?

• Open-source Deep Learning framework
• Supports state-of-the-art network architectures
• GPU/CPU support
• Distributed computations on Apache Spark and Hadoop
• Word2Vec for Text Mining
DeepLearning4J in KNIME

• Visually assemble networks using KNIME nodes
• Integrates with other KNIME extensions, e.g. KNIME Image Processing / KNIME Text Mining
• Networks can be trained and executed on GPU and CPU
Demo: Deep Learning
Analytics

Tree Ensembles and Gradient Boosting
The general idea is to take advantage of the “wisdom of the crowd”: combining predictions from a large number of weak predictors leads to a more accurate predictor.

Typically: for classification the individual models vote and the majority wins; for regression, the individual predictions are averaged.
KNIME’s Tree Ensemble models

- Parameter tuning is relatively easy
- No data normalization required
- Nominal and numerical attributes
- Scalability
- Missing Value Handling
New: Gradient Boosted Trees

- Another algorithm for creating ensembles of decision trees
- Starts with a tree built on a subset of the data
- Builds additional trees to fit the residual errors
- Shows great performance on structured data
Trees and Tree Ensembles: Changes “under the hood”

- Support of binary splits for nominal attributes
- Missing value handling
- Support of byte vector data (high-dimension count fingerprints)
- Code optimization
  - Runtime
  - Memory
Demo: Gradient Boosted Trees
Analytics

Feature Selection
Feature Selection

• “Feature selection is the process of selecting a subset of relevant features (variables, predictors) for use in model construction” [Wikipedia]

• Why?
  – Better generalization
  – Simplification of the model
  – Shorter training times

• Dozens of methods to do that...
Feature Selection

- Backward Feature Elimination:
  Start with full feature set, iteratively remove ‘worst’ feature

- Forward Feature Selection:
  Start with empty feature set, iteratively add ‘best’ feature
Feature Selection nodes

• Same loop structure as former Backward Feature Elimination nodes

• Different strategies
  – Forward selection
  – Backward elimination

• Uses Flow Variable as score
  – Flexibility

• Preconfigured meta nodes for both strategies
A few more...
Semantic Web

• Access the wealth of Semantic Web from within KNIME
• Create your own Semantic Web with the Memory Endpoint
• Read/write support for Semantic Web file formats
• Manipulate triple stores via SPARQL
• Usage model similar to database integration
KNIME RESTful Web Service Client Nodes

• REST Client nodes: Get / Post / Put / Delete Resources
• Follow-up of famous KREST community extensions
• Integrating with KNIME’s XML/JSON Processing nodes
• Powerful Configuration
• Extensible (e.g. custom auth types)
KNIME Tableau Integration

• Tableau: Popular (commercial) visualization and dashboard application

• New KNIME nodes to:
  – Write native Tableau files (TDE)
  – Send data to Tableau server
Type Extensions in Java Snippet

- Java Snippet: Swiss Army knife for data manipulation
- Read/write support for 3rd party types: SVG, XML, JSON, Images, Molecules, …

... and now for something completely different

– Rosaria Silipo –
Where can I learn more?

• About KNIME:
  
  • **New!** Online Course at [https://www.knime.org/knime-introductory-course](https://www.knime.org/knime-introductory-course)
  • **New!** Course slides (pdf) downloadable from [https://www.knime.org/training](https://www.knime.org/training)
Free Online Course: Web Page

- Hands-on online course
- Introductory (Reading, ETL, Writing)
- ... but still growing
- Small units
- Learning path with:
  - Unit Tree
  - Links to Next and Previous unit
- 36 short videos!
- Exercises
- All solutions on the EXAMPLES server
- Final exercise to test your knowledge!

https://www.knime.org/knime-introductory-course
Looking for early adopters this March!

https://www.knime.org/knime-introductory-course
Free Online Course - Credits

Oleg Yasnev – Publishing
Vincenzo Tursi – Recording, Editing
Heather Fyson – Dubbing, Recording, Editing
Rosaria Silipo – Art Director
KNIME Course Slides – now available as pdf

KNIME Text Mining Training

This one day course is an intensive training focused on the processing and mining of textual data with KNIME using the Textprocessing extension. Learn how to read textual data in KNIME, enrich it semantically, preprocess, and transform it into numerical data, and finally cluster it, visualize it, or build predictive models. Text mining experience is not necessarily required for this training.

Duration: 1 day
Costs: € 750 per user
Topics:
1. Introduction to KNIME
2. Reading and Importing Textual Data
3. Text Preprocessing, Semantic Enrichment, and Transformation
4. Text Classification
5. Visualization
6. Text Clustering

PDF files for KNIME Text Mining Training material can be [downloaded here.](#)
KNIME Server
KNIME Software

KNIME Analytics Platform

PRODUCTIVITY
- KNIME Personal Productivity
- KNIME Partner Productivity

COLLABORATION
- KNIME Server
- KNIME Cloud Server
- KNIME Server Lite
- KNIME WebPortal
- KNIME TeamSpace

PERFORMANCE
- KNIME Big Data Connector
- KNIME Spark Executor
- KNIME Cluster Executor

COMMUNITY & PARTNER EXTENSIONS
KNIME Server

Shared Repositories

Access Management

Web Enablement

Flexible Execution

Data

Nodes, Files, Applications

Content, Services

Partitioning

Flexible Execution

Nodes & Metanodes

Remote, Scheduled

Workflows
KNIME Server – Extended REST API

• Integrate KNIME Server functionality with IT infrastructure
• Execute workflows, check server status, and more

• See Blog Posts for detailed tutorials:
  – https://www.knime.org/blog/giving-the-knime-server-a-rest
  – https://www.knime.org/blog/the-knime-server-rest-api
  – https://www.knime.org/blog/creating-restful-services-with-knime-an-update
Execute workflow via REST API

• Add Quickforms to define workflow API
Workflow orchestration via REST API

• Calling a remote workflow

This Friday: KNIME Server Workshop
**KNIME Server Installer**

Step-by-step guided Server installation (Windows and Linux)
KNIME Server – Admin made easy

• KNIME Administrator is often not a KNIME Analytics Platform user

• Make tasks like user administration easier

• Get an overview of the KNIME Server health
Go to the administration portal
KNIME Server – Admin made easy
### License Info

**License Type:** KNIME Server  
**Expiration Date:** 2016-07-08  
**Company:** KNIME.com AG  
**Customer:** Jon Fuller  
**Host Identifiers:** { "Host IP": "10.0.0.26" }  
**Comment:** Demo license

#### Users

**Desktop**  
- limit: 5  
- current: 0

**Webportal**  
- limit: 50  
- current: 1  
  - **active Users**  
    - **name:** admin  
    - **last login:** 2016-06-09T10:23:32.268

**Webservice**  
- limit: 50  
- current: 1  
  - **active Users**  
    - **name:** admin  
    - **last login:** 2016-06-09T10:29:26.476
KNIME Server – Admin made easy
## KNIME Server – Admin made easy

### Users

<table>
<thead>
<tr>
<th>User</th>
<th>Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>knimeadmin</td>
<td>admin, knime</td>
</tr>
<tr>
<td>easilycreatesnewuser</td>
<td>knime</td>
</tr>
</tbody>
</table>

### Groups

<table>
<thead>
<tr>
<th>Group</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>admin</td>
<td></td>
</tr>
<tr>
<td>knime</td>
<td></td>
</tr>
</tbody>
</table>
KNIME Server – Admin made easy
KNIME Server – License Distribution

- License files no longer required in client installation
- Checked out from KNIME Server
- Centrally managed, less configuration required
Why cloud?

• Bring analytics to the data

• Work fast, work agile

• Pay for what you need

• New ways to work
Cloud Platforms

- AWS and Azure
- The two leaders in Cloud (see Gartner 2016)
KNIME Cloud Products (Marketplace)

• KNIME Cloud Analytics Platform

• KNIME Server (5-User)

• KNIME Server + Big Data Extensions (5-User)

• KNIME Server (+ Big Data Extensions) BYOL
KNIME Cloud Analytics Platform

- KNIME Analytics Platform 3.3
- Running on Windows Server 2016
- Connect via Remote Desktop
- Pre-configured and optimized for Azure

Why?
- Sometimes need to process a large dataset (faster hardware).
- Want to do Deep Learning (using a powerful GPU)
KNIME Cloud Analytics Platform

• Celebrity Face Prediction with AlexNet (DL4J)

Model Training Time (mins)

- Laptop: 49 mins
- Azure CPU: 31 mins
- Azure GPU: 14 mins
KNIME Server in the Cloud

About KNIME

At KNIME, we build software for fast, easy and intuitive access to advanced data science, helping organizations drive innovation. Our KNIME Analytics Platform is the leading open solution for data-driven innovation, designed for discovering the potential hidden in data, mining for fresh insights, or predicting new futures. Organizations can take their collaboration, productivity and performance to the next level with a robust range of commercial extensions to our open source platform. For over a decade, a thriving community of data scientists in over 60 countries has been working with our platform on every kind of data: from numbers to images, molecules to humans, signals to complex networks, and simple statistics to big data analytics. KNIME’s headquarters are based in Zurich, with additional offices in Konstanz, Berlin, and Silicon Valley. We are open for innovation, so visit us at KNIME.com.

Visit the KNIME Website

KNIME Products (3)

**KNIME Server + Big Data Extensions**

*Version 4.4.0 with Big Data Extensions* | Sold by KNIME.com

Starting from $6,805/hr or from $449,600/yr (17% savings) for software + AWS usage fees

Regardless of whether you use the KNIME Analytics Platform for advanced analytics, machine learning, business intelligence, or ETL tasks, you can use the KNIME Server to extend...

Linux/Unix, Ubuntu 16.04 LTS - 64-bit Amazon Machine Image (AMI)

**KNIME Server (BYOL)**

*Version 4.4.0* | Sold by KNIME.com

Regardless of whether you use the KNIME Analytics Platform for advanced analytics, machine learning, business intelligence, or ETL tasks, you can use the KNIME Server to extend...

Linux/Unix, Ubuntu 16.04 LTS - 64-bit Amazon Machine Image (AMI)

**KNIME Server**

*Version 4.4.0* | Sold by KNIME.com

Starting from $1,725/hr or from $12,600/yr (17% savings) for software + AWS usage fees

Regardless of whether you use the KNIME Analytics Platform for advanced analytics, machine learning, business intelligence, or ETL tasks, you can use the KNIME Server to extend...

Linux/Unix, Ubuntu 16.04 LTS - 64-bit Amazon Machine Image (AMI)
Cloud Connectors

- Enable working directly with data on the AWS or Azure clouds
  - S3
  - Blob Storage
  - SQL Server
KNIME Cloud Connectors

- Amazon S3/Azure Blob Storage: Save (large) unstructured data in the cloud
- Connect through KNIME’s file handling nodes
  - List
  - Upload/Download
  - Delete
  - URL access
Cloud Connectors Workflow Example

Amazon S3 Connection
Enter your own credentials here

Amazon S3 File Picker
Using the path we create a URL from which the CSV reader can read directly

File Reader
Read the CSV file

Process as a stream
Cloud Connectors – Will they blend?

Will They Blend? Amazon S3 meets MS Blob Storage ... and as usual also a few Excel files.

The challenge here is to blend S3 formatted data from the Amazon Cloud with Blob Storage formatted data from the MS Azure Cloud. As a secondary challenge we want to blend the cloud data with Excel data. Will they blend?

Data is the new CENSUS data downloadable from [link]. We report the average travel time to work vs. the English proficiency and we discover that, if you are a woman, the less you speak English the closer to home your work place is. If we then analyze the size of the statistical sample, we also discover that the average is based on only 21 not English speaker females. Probably not enough to consider the results as reliable. Always maintain a healthy degree of skepticalism towards your results!

By the way ... yes they blend!

https://www.knime.org/blog/S3_meets_BlobStorage
# Azure SQL Server Connector

## Subscriptions:
- **All 2 selected**

<table>
<thead>
<tr>
<th>Name</th>
<th>Status</th>
<th>Replication Role</th>
<th>Server</th>
<th>Pricing Tier</th>
<th>Location</th>
<th>Subscription</th>
</tr>
</thead>
<tbody>
<tr>
<td>demodb</td>
<td>Online</td>
<td>None</td>
<td>knime</td>
<td>Basic</td>
<td>West Europe</td>
<td>Microsoft Azure Spons...</td>
</tr>
</tbody>
</table>

## Diagram:

- **SQL Server Connector**
  - Connect to Azure SQL DB

- **Database Table Selector**
  - Select Customer Table

- **Database Row Filter**
  - Filter only customers with suffixes in their names

- **Database Connection Table Reader**
  - Read data from database to KNIME Analytics Platform
Comparing on-premises vs. cloud architecture

• So now possible to work with resources hosted in the cloud

• S3, Blob storage, managed databases, Hadoop/Spark

• Need to think about how the architectures differ
  – To minimise data-transfer (maximise speed, minimise cost)
Status Quo

Laptop processing power limited

Limited Transfer Speed

Increased Transfer Cost

Analytics Platform

Your Azure Cloud

Azure SQL Server

Azure Blob Store
A Cloud Architecture

- Workflows small c.f. data
- Remote Desktop
- Your Azure Cloud
- Azure Blob Store
- Azure SQL Server
- Azure Marketplace
- Fast data transfer
- No data transfer cost
- KNIME Cloud Analytics Platform
- Get started fast via Marketplace purchase
KNIME in the Cloud: Summary

• KNIME Analytics Platform available on Azure Marketplace
• KNIME Server available on Azure and AWS marketplaces
• Cloud connectors for S3, Blob Store, Azure SQL server
Guided Analytics
– Greg Landrum –
Reminder of what we’re talking about here

Business Analyst

Data Scientist

Some Interface

Start here
So what’s new?

• JavaScript Views:
  – Parallel coordinates plot
  – Table view
  – Decision tree view

• Interactivity: Selection and filtering
  – Range filter widgets
  – Linked views
  – Accessing views in wrapped metanodes
But let’s start with a demo
The workflow behind the demo:
Interested in more?

• More details and examples available online:
  – Release notes / blog posts
  – example workflows server ...

• Meet us at the booth for demos or follow-up discussions
The KNIME® trademark and logo and OPEN FOR INNOVATION® trademark are used by KNIME.com AG under license from KNIME GmbH, and are registered in the United States.