What’s New

Bernd Wiswedel
KNIME
What this session is about

• Presenting (and demo’ing) enhancements added in the last year
• By the team

• Questions? See us at the booth.
KNIME as an Open (Source) Platform
Open Sourcing KNIME Analytics Platform Extensions

• Personal Productivity Extensions in July (version 3.4)
  – Workflow Diff Viewer
  – Metanode Templates
  – Workflow Linking ("Call Local Workflow")
  – "Local" Workflow Coach
KNIME Personal Productivity – Workflow Diff
KNIME Personal Productivity – Metanode Templates
KNIME Personal Productivity – Workflow Linking

- Workflow Orchestration by calling out to other workflows

- Call Local Workflow

- 13:30 - Session 1
  - H2O in KNIME: Integrating High Performance Machine Learning
    Jo-Fai Chow (H2O.ai) and Christian Dietz (KNIME)
  - KNIME in Action: The Monster Model Factory
    Greg Landrum (KNIME)
  - Predictive Analytics of Hair Colorations
    Georg Knübel (Henkel Beauty Care) and Iris Adä (KNIME)
Open Sourcing KNIME Analytics Platform Extensions

- Personal Productivity Extensions in July (version 3.4)
  - Workflow Diff Viewer
  - Metanode Templates
  - Workflow Linking
    (“Call Local Workflow”)
  - “Local” Workflow Coach

- Big Data Extensions in December (version 3.5)
  - Big Data Connectors
  - KNIME Executor for Apache Spark
KNIME Big Data Extensions

- Connect to Big Data Stores: Hive, Impala, HDFS
- Run Computation in Apache Spark
KNIME Analytics Platform – Source Code on GitHub

- Source Code available on GitHub & BitBucket

https://github.com/knime/
Really, really, really important disclaimer

This is most definitely not production quality code. These nightly builds are recent developments, so they are not tested as thoroughly as stable versions. Functionality may change substantially (or disappear entirely) at any time without notice.
The KNIME Software Ecosystem

KNIME

Supported Extensions

KNIME Analytics Platform

Community Extensions

KNIME Extensions

Partner Extensions

KNIME Server
Outline – What’s new

• Integrations / Utility Nodes
• Deep Learning in KNIME
• Big Data Integration
• Education and Evangelism
• KNIME Server
• Guided Analytics
Google Connectivity

- Read and Write Sheets in Google Drive
- Access via API Token & Interactive
- Same concepts as Google Analytics Nodes
Google Connectivity - Demo
Connectivity: Amazon Redshift

- Fully managed data warehouse
- SQL interface
- Distributed and parallelized queries

- Amazon Redshift connector integrates seamlessly with existing KNIME Database connectors
Connectivity: Amazon Athena

- Serverless, interactive query service
- SQL interface

- Amazon Athena integrates seamlessly with existing KNIME Database connectors
Connectivity: Amazon Athena
# Azure/Microsoft SQL Server Connector

## Subscription Table

<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
Azure/Microsoft SQL Server R Services

- Run R Script (+ model) in SQL Server
- Learn local (in KNIME), deploy in database
- In-database Analytics

Diagram of Azure/Microsoft SQL Server R Services

- Learn Model
  - R Learner
  - Learn R Model

- Predict using R in Microsoft SQL Server
  - Run R Model in Microsoft SQL Server
  - Database Reader

- Upload data for prediction
  - Column Filter
    - Remove class column
  - Database Writer
    - Upload Data as "NewData" SQL Table
  - Database Table Selector
    - Delete Table "NewData"
    - Node 19

© 2018 KNIME AG. All Rights Reserved.
Audio Processing nodes

- New audio cell type
- Listen to audio files within KNIME
- Extract acoustic features
- Speech recognition (Bing, IBM Watson, and other)
Date & Time Handling Nodes revised

- Time zone support
- Multiple columns support
- “Duration Math”
- Flow variable support
Promotion of KNIME Extensions out of KNIME Labs

• REST Client Nodes
• Feature Elimination
• JavaScript Views (c.f. Guided Analytics)
• Text Processing
Logistic Regression Nodes revised

- More Scalable
- Additional output (learning properties)
Math Formula (Multi Column)*

*) This is not only for Dean.
Math Formula (Multi Column) – Demo
Scripting – Python

• Supports Python 3
• Major Speedup
• Used for new Deep Learning Integration
Deep Learning in KNIME
- Christian Dietz -
Deep Learning?

- Multi-Layer Neural Networks for learning tasks on images, videos, text or sound
- Representation Learning
- Regression and Classification, Unsupervised Learning, Reinforcement Learning, …

https://research.googleblog.com/2016/03/train-your-own-image-classifier-with.html
Typical application areas

• Computer Vision
• Speech Recognition
• Natural Language Processing
• Cheminformatics
• Internet of Things
• Timeseries analysis
• …
Recap: KNIME DeepLearning4J Integration (since 3.3)
DeepLearning4J - Basic workflow

Create Network Architecture

Train Network

Predict
However...

- Lots of other frameworks out there (e.g. Keras, Tensorflow, CNTK, PyTorch, MXnet).
- Many state-of-the-art frameworks offer pre-trained models which we want to reuse.
- Support for more advanced techniques: recurrent neural networks, GAN execution, multi-layer input and output etc.
- Enable scripters to intervene.
New: KNIME Deep Learning (since 3.5)
KNIME Deep Learning

• General framework for executing, training, fine-tuning, building, deploying deep learning networks.
• Independent (whenever possible) from actual deep learning backend.
• First implementation of KNIME Deep Learning based on Keras with Tensorflow or CNTK as backend.
KNIME Deep Learning – A simple MNIST model
KNIME Deep Learning – A simple MNIST model

Live Demo
KNIME Deep Learning – Transfer learning

Elephant
Soccer
Ball
...
Umbrella

Cat
Dog
KNIME Deep Learning – Transfer learning
KNIME Deep Learning – Transfer learning
KNIME Deep Learning – Transfer learning
KNIME Deep Learning – Transfer learning
Why Deep Learning with KNIME?

• Integrating state-of-the-art functionality from multiple frameworks rather than reinventing the wheel
• Make DL frameworks more accessible for non-experts
  – Visual model building
  – Data pre- and postprocessing
  – Fine-tuning and training of existing architectures
  – Model deployment
• Make DL more approachable
Big Data Extensions
– Tobias Kötter –
KNIME Big Data Connectors

• Preconfigured connectors
  – Hive
  – Impala
  – (secured) HDFS, webHDFS, httpFS
• Runs on Hadoop
• Support for Kerberos secured cluster
• Extends the database and remote file handling integration
Database Extension

- Visually assemble complex SQL statements
- Connect to almost all JDBC-compliant databases
- Harness the power of your database within KNIME
- Operations are performed within the database
KNIME Extension for Apache Spark

- Based on Spark MLlib
- Scalable machine learning library
- Runs on Hadoop
- Algorithms for
  - Classification (decision tree, naïve bayes, ...)
  - Regression (logistic regression, linear regression, ...)
  - Clustering (k-means)
  - Collaborative filtering (ALS)
  - Dimensionality reduction (SVD, PCA)
- Supports Spark version 1.2, 1.3, 1.5, 1.6, 2.0, 2.1 and 2.2
- Support for Kerberos secured cluster
Familiar Usage Model

• Usage model and dialogs similar to existing nodes
• No coding required
• Spark nodes start and manage Spark jobs
• Data stays within your cluster
Big Data Extensions – What’s New

• Free & open source since December 2017

• Cloud connector support
  – Amazon S3
  – Microsoft Azure Blob Store

• New Nodes
  – Spark Missing Value
  – Spark GroupBy

• Usability
  – Destroy Spark context when workflow is closed
Spark 2.x
Reasons to upgrade to Spark 2.x

Reason 1: More functionality and fixes

More than 4600 resolved Spark tickets in total
Reasons to upgrade to Spark 2.x

Reason 1: More functionality and fixes
Reason 2: Parallel installation
Reasons to upgrade to Spark 2.x

Reason 1: More functionality and fixes
Reason 2: Parallel installation
Reason 3: Up to 10x faster
Migration from Spark 1.6 to Spark 2.x

- Major version change 😞
  - Adapt to backward incompatible changes in the Spark API
- Update Spark ecosystem libraries
- Migrate to DataFrames to get better performance
Migrating Your KNIME Workflows
Migrating Your KNIME Workflows

Step 1: Change Spark version in Create Spark Context node
Migrating Your KNIME Workflows

Step 1: Change Spark version in *Create Spark Context* node
Step 2: Execute the workflow
Migrating Your KNIME Workflows

Step 1: Change Spark version in *Create Spark Context* node
Step 2: Execute the workflow
Step 3: SMILE ☺
Migrating Your KNIME Workflows

Step 1: Change Spark version in Create Spark Context node
Step 2: Execute the workflow
Step 3: SMILE 😊
Education and Evangelism

– Rosaria Silipo –
Courses
What can you learn?

Available Courses:

• Introduction to KNIME
• Advanced KNIME
• Text Mining
• Big Data
• KNIME Server
Where can you learn?

Germany
- Berlin: May 14-18
- Jul 02-06
- Nov 19-23
- Stuttgart: Apr 16-20
- Jun 11-15
- Oct 08-12
- Munich: May 17

US
- Austin: Apr 23-27
- Nov 06-07

Switzerland
- Zurich: May 22-25
- Nov 12-16
- Bern: Oct 15-19

Italy
- Rome: Apr 19-20

https://www.knime.com/courses
Courses in your own time

Online Course

- “Introduction to KNIME” with Kathrin
- With teacher - 4 x 1h/week
- With exercises and exercise corrections
- Location: Webex
- Cost: EUR 250
- Next: “Advanced KNIME”
- https://www.knime.com/about/events/online-course-basic-functionalities-of-knime-analytics-platform-may-2018

E-Learning Course

- “Introductory Course to Data Science”
- With video, no teacher - 15 chapters
- With exercises and solutions
- Location: Website
- Cost: EUR 0
- More units to come ...
- https://www.knime.com/knime-introductory-course
Books from KNIME Press
New Books and Book Series at KNIME Press

KNIME Beginner’s Luck
To start ...

The “Will they blend?” Collection
All you need to know on Data Access

The KNIME Cookbook
To deepen ...

From Words To Wisdom
For Text Processing
From Words to Wisdom: Premiere

- Available at KNIME Press: https://www.knime.com/knimepress
- Select book “From Words to Wisdom”
- Use Promotion Code: Text_Mining_Book_Premiere
Meetups
What is a Learnathon?

Lately we have introduced **Learnathon** events

A Learnathon is:

- In between a hackathon and a workshop
- 3-4 hours
- 70-80 attendees per event so far
Upcoming Meetup Events
US Roadshow

KNIME®
Data Science Roadshow 2018
March 19 – April 19

Boston
Cincinnati
New York
Seattle
San Francisco
San Diego
Denver
Houston

© 2018 KNIME AG. All Rights Reserved.
On the Web
Innovation Notes

KNIME Innovation Notes are:
• Practical
• Always working with a KNIME example

https://www.knime.com/solutions/innovation-notes

They include:
• Topic / Business Issue
• High Level Summary
• Detailed Description
• Achievable Results
• KNIME Output / Graphic / Interface
• Overview Solution Workflow
• KNIME Software used
• Fully working KNIME workflow location
The new KNIME Forum
Social Media

KNIMEanalytics

@KNIME

KNIME

KNIMETV
KNIME Server
— Jon Fuller —
The KNIME Software Ecosystem

- KNIME Supported Extensions
- KNIME Extensions
- KNIME Analytics Platform
- Community Extensions
- Partner Extensions
- KNIME Server

Collaboration:
- Compliance
- Best Practices
- Sharing Expertise

Automation:
- Scheduling
- (Model) Management

Deployment:
- to Applications
- to Humans
KNIME Server

Shared Repositories

Access Management

Web Enablement

Flexible Execution

Nodes & Metanodes

Nodes, Files, Applications

Content, Services

Data

Partitioning

My Predictor

Remote, Scheduled

Workflows

© 2018 KNIME AG. All Rights Reserved.
KNIME Server – Admin made easy

- Easier log file download
- Upload new license files via AdminPortal
- Log File analysis workflow
- List scheduled jobs in WebPortal
Deploy to Server
Recap: Convert workflow to REST enabled workflow

**Standard Workflow**

**REST enabled Workflow**
Recap: Convert workflow to REST enabled workflow

REST enabled Workflow
Recap: Convert workflow to REST enabled workflow

```json
{
    "age": 42,
    "workclass": "State-gov",
    "fnlwgt": 77516,
    "education": "Bachelors",
    "education-num": 12,
    "marital-status": "Never-married",
    "occupation": "Adm-clerical",
    "relationship": "Not-in-family",
    "race": "White",
    "sex": "Male",
    "capital-gain": 2174,
    "capital-loss": 0,
    "hours-per-week": 40,
    "native-country": "United-States"
}
```
Recap: Convert workflow to REST enabled workflow
Show API definition
Social Workflow Repository (Preview)

• Preview functionality.
• You can use it now.
• Know KNIME Labs? It’s the same kind of idea...
  – Useful functionality
  – More functionality planned
  – Feedback from users very welcome
  – Functionality and UI likely to change
Social Workflow Repository (Preview)

If you’ve got to get something done, and quickly...

...wouldn’t it be great if someone else had already done this?

– See EXAMPLES server and Node Guide
– But what about something that’s relevant just for my team, lab, group, company...
Social Workflow Repository (Preview)

If you’re managing a team, and are responsible for reproducible, reusable data science...

...wouldn’t it be great if my team didn’t need to reinvent the wheel each time we start a new project...

  – What happens if one of my team members leaves?
  – How can I foster a sharing culture within my organisation?
Welcome to the KNIME Spring Summit in Berlin! Check out the Social Workflow Repository...

**News**

- **KNIME Summit Server Workshop**
  Find the KNIME Server workshop in room Berlin-London on Friday March 9th.
- **KNIME Summit**
  KNIME Summit in Berlin is from March 5th to March 9th.
- **KNIME Server 4.6.1 released**
  KNIME Server 4.6.1 is now available for download.

**Most Recent Workflows**

- **10_DiseaseGenes**
  uploaded by KNIME on 02.03.2018 10:36:15
- **2. Predict Results on a Schedule**
  uploaded by Emil on 28.02.2018 11:04:32
- **1. Train model (on schedule)**
  uploaded by Alex on 28.02.2018 10:44:54
- **3. Predict Results Using REST API**
  uploaded by Emil on 28.02.2018 10:24:58
- **01 - Test Basic Workflow - Data Blending**
  uploaded by knimeadmin on 05.12.2016 17:44:24
Guided Analytics
Guided Analytics

– Greg Landrum –
Reminder of what we’re talking about here

Someone who needs data / analytics

KNIME expert
What often happens

Someone who needs data / analytics

KNIME expert
What often happens

Someone who needs data / analytics

KNIME expert
What often happens

Someone who needs data / analytics

KNIME expert
What often happens

Someone who needs data / analytics

KNIME expert
What often happens

Someone who needs data / analytics

KNIME expert
What often happens

Someone who needs data / analytics

KNIME expert
What we’d like to provide

Someone who needs data / analytics

Guided Analytics workflows

KNIME expert
Guided Analytics workflows

- Enable people who are not KNIME experts to work with/explore/analyze/learn from data
- Typically deployed via KNIME WebPortal
- Heavy use of interactive JavaScript views for the user experience
Guided Analytics workflows

- Enable people who are not KNIME experts to work with/explore/analyze/learn from data
- Typically deployed via KNIME WebPortal
- Heavy use of interactive JavaScript views for the user experience

- Bonus for everyone: the JavaScript views that we’re building for Guided Analytics are also really useful in “normal” KNIME workflows.
So what’s new?

• JavaScript Views:
  – Network view
  – Stacked area chart
  – Sunburst chart
  – Tag cloud
  – Data explorer
  – OPTICS nodes
  – Table editor

• Interactivity: Selection and filtering
  – Accessing views in wrapped metanodes, composite view
I like making slides, but demos are even better...
What this session is about

• Presenting (and demo’ing) enhancements added in the last year
• By the team
• Questions? See us at the booth.