What’s Cooking

Bernd Wiswedel
KNIME
Outline

• KNIME as an open (source) platform
• What’s Cooking
  – Speech Recognition
  – H2O Integration
  – Cloud Connectors & Offerings
  – Guided Analytics
KNIME Analytics Platform – soon on GitHub

- Source Code available on GitHub & BitBucket
- Ongoing effort, enough to get started

https://github.com/knime/
KNIME Analytics Platform – Nightly Builds

• Nightly Builds will be publicly available

• Will then also include enhancements shown in the following

• Disclaimer: Bleeding edge; not meant for production use!
SVN To Git – Fun Facts

• Recently switched from SVN to Git
• Split into 70 repositories (60 open source)
• Code Base:
  – ~20GB (includes text models, R, etc)
  – ~1.7 Mio Lines of Code (open source only)
  – 1500 integration tests per version (excl. unit tests)
• ~3 weeks conversion time (parallelized over 10 physical machines)
KNIME Personal Productivity – free & open-source

• Part of the KNIME Analytics Platform 3.4 (due in summer ‘17)
• What is it?
  – Workflow Diff Viewer
  – Metanode Templates
  – Workflow Linking (“Call Local Workflow”)
  – “Local” Workflow Coach
KNIME Personal Productivity – Workflow Diff
KNIME Personal Productivity – Workflow Linking

• Workflow Orchestration by calling out to other workflows
KNIME Personal Productivity – free & open-source

• Part of the KNIME Analytics Platform 3.4 (due in summer ‘17)
• Want to use it NOW?

Available in KNIME Store free of charge
What’s Cooking

• Speech Recognition
• H2O Integration
• Cloud Connectors & Offerings
• Guided Analytics
Speech Recognition
– Tobias Koetter –
KNIME Audio Processing

• New audio cell type
• Listen to audio files within KNIME
• Extract acoustic features
• Speech recognition


https://www.youtube.com/watch?v=1L6zwLMvA-U
H2O Integration
– Christian Dietz –
What is H2O?

- Open-source machine learning framework
- Fast, scalable and in-memory implementations of state-of-the-art machine learning algorithms.
- Massively scalable big data analysis on for example on Hadoop and Spark
H2O in KNIME

• Integrate H2O machine learning algorithms
• Tool Blending: Mix and match with existing KNIME Analytics Platform extensions
  – Basic KNIME Analytics Platform functionality
  – KNIME Big-Data Connectors
  – Text Mining, Image Processing, ChemInformatics, ...
  – and more!
H2O Algorithms

- Tree Ensembles
- Clustering
- Generalized Linear Modeling
- Grid Search
- Ensemble Stacking
- Deep Learning
Demo: H2O in KNIME
Cloud Connectors & Offerings

– Jon Fuller –
Overview

• More in the marketplace

• More connectors

• More deployment options

• Better integration
• KNIME Cloud Analytics Platform

• KNIME Server (5-User)

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

• KNIME Server (+ Big Data Extensions) BYOL
More connectors: Amazon Redshift

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

• Amazon Redshift connector will seamlessly integrate with existing KNIME Database connectors
S3/Blob Store support for Spark Data Source Nodes
Cloud Resources

• Data storage
  – MS Azure Blob Storage, Amazon S3

• Databases
  – Azure SQL Server, Amazon RDS (MySQL, PostgreSQL etc)

• Data warehouse
  – Amazon Redshift

• Hadoop/Spark
  – Azure HDInsight, Amazon EMR
Cloud Resources

What are we missing? Let us know!
Cloud Deployment Templates

• To allow deployment of a complete environment for Analytics using KNIME Software.

• KNIME Server + EMR/HDInsight + Managed DB

• AWS CloudFormation templates

• Azure ARM templates
AWS CloudFormation template

Amazon VPC

Internet gateway

KNIME Server on EC2

EMR Cluster

Workflows

Data

KNIME Server from AWS Marketplace
Azure ARM template
Deploy and architecture from Github!

• Create (or fork) an ARM template

• Click the blue button

• Add details in the portal, and launch your architecture

• https://github.com/jfkotw/azure-quickstart-templates/tree/master/hdInsight-apache-spark
Guided Analytics
– Greg Landrum –
What’s cooking?

• **JavaScript Views:**
  – Network viewer
  – Stacked Area chart

• **Interactivity:**
  – Composite (wizard) view in Analytics Platform
  – Layout editor
JavaScript network viewer demo
The workflow behind the demo
The workflow behind the demo
JavaScript Stacked Area Chart demo
The workflow behind the demo:
The workflow behind the demo:
... and the KNIME Server?

2:30pm - Session – KNIME in the Enterprise

- Deploying KNIME to the Enterprise: Reshaping Data & Architecture for Healthcare
  *James Lumley (Eli Lilly and Company)*

- Digital factories with CloudFoundry, KNIME, and BOSH
  *Jan Pospisil (Siemens), Stefan Zabor & Noah Ispas (Anynines)*

- Tech Briefing: Distributed Workflow Executors
  *Thorsten Meinl (KNIME)*

3:45pm - Afternoon Break

4:15pm - Special Session

- The KNIME Model Factory or "Getting your life back"
  *Iris Adä & Phil Winters (KNIME)*

5:00pm - End of Day 1
Other noteworthy changes

• Date & Time Handling Nodes revised
  – Richer data types
  – Timezone support
  – Duration “math”

• R Integration reviewed
  – Adding support for Microsoft R Open & R Server
  – Professional Support by Microsoft
  – (Check out the booth for details)
Thanks

• Functionality presented here part of next feature release (due in summer)
• Want to try it earlier? Watch out for nightly builds as they become available
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.