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 Software

Load > Integrate > Transform > Analyze > Visualize

Open Source

- KNIME Integrations
- KNIME Extensions

KNIME Analytics Platform

- Community Extensions
- Partner Extensions

KNIME Server

Collaborate — Automate — Manage — Deploy

9am tomorrow: Short Server Intro
Jon Fuller (KNIME)
Git Client Nodes

- Extract Statistics from Git repositories
- Push, Pull, Tag, Commit History & Details

KNIME Analytics Platform – Source Code on GitHub

- Source Code available on GitHub & BitBucket

https://github.com/knime/
Math Formula (Multi Column)*

*) This is not only for Dean.
Column Expressions

- Swiss army knife for deriving new columns
- Complete type support
- Multiple expressions
Outlier Handling

- Handle extremely large or small values in numeric columns
- Replace Values or Filter Rows containing outliers
- Usual pattern of Compute-and-Apply (two nodes)
Google Connectivity (Labs)

• Read and Write Sheets in Google Drive
• Access via API Token & Interactive
• Same concepts as Google Analytics Nodes
Kafka Integration (Labs)

- Apache Kafka – distributed publish-subscribe streaming platform
- KNIME Integration supporting sending and consuming messages
Azure/Microsoft SQL Server R Services

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

Upload data for prediction

- Column Filter
- Database Writer
- Database SQL Executor
- Delete Table "NewData"
- Database Table Selector
  - Upload Data as "NewData" SQL Table

Predict using R in Microsoft SQL Server

- Run R Model in Microsoft SQL Server
- In-Server R Prediction
- Retrieve Result Table
Workflow Automation – “Call Workflow”

• Call out to other workflows
  – Locally (same KNIME instance)
  – On KNIME Server
• Orchestrate set of workflows

Tomorrow at 9:30
“On Monsters and Tags...”
Jeany Prinz & Greg Landrum (KNIME)
Promotion of KNIME Extensions out of KNIME Labs

- REST Client Nodes
- Feature Elimination
- JavaScript Views (c.f. Guided Analytics)
- Text Processing
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 all major Spark version
• 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
KNIME Big Data Connectors

Better support for big data file formats:

– Why? Efficiency & Speed
– Which formats? Parquet and ORC
KNIME Extension for Apache Spark

• Data preparation
  – Spark GroupBy
  – Spark Pivot
  – Spark Missing Value (Apply)

• Data mining
  – Frequent Item Set
  – Association Rule Learner (Apply)

• Usability
  – Destroy Spark context when workflow is closed
Local Big Data Environment

Create Local Big Data Environment

- Hive
- HDFS
- Spark

Provides everything to
- get started with big data **easily**
- rapidly prototype workflows
- leverage Spark on multi-core machines

Batteries included:
- no further software installation required
- no cluster required
- no jobserver required
Deep Learning in KNIME
- Christian Dietz -
Deep Learning?

- Very complex Multi-layer Neural Networks
- Great performance on images, sound, text or timeseries data.
KNIME Deep Learning

• Lots of frameworks out there (e.g. Keras, Tensorflow, CNTK, PyTorch, DL4J, MXnet,…).
• Make DL frameworks more accessible for non-experts.

Since 3.2

Since 3.5

Since 3.6
KNIME Deep Learning – Executing

Soccer ball
KNIME Deep Learning – Executing

- DL Keras Network Reader
  - InceptionV3, pre-trained on ImageNet
- Preprocessing
  - DL Network Executor
    - Output: probabilities for 1000 ImageNet classes
- Format results
- Read Images
KNIME Deep Learning – Executing
KNIME Deep Learning – Executing

- DL Keras Network Reader
- InceptionV3, pre-trained on ImageNet
- Preprocessing
- DL Network Executor: output probabilities for 1000 ImageNet classes
- Format results
- Read Images
### KNIME Deep Learning

<table>
<thead>
<tr>
<th>Row ID</th>
<th>Image</th>
<th>Predicted Class</th>
<th>Class ...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row2</td>
<td><img src="image1.jpg" alt="Dog" /></td>
<td>dalmatian, coach dog, carriage dog</td>
<td>0.966</td>
</tr>
<tr>
<td>Row3</td>
<td><img src="image2.jpg" alt="Soccer Ball" /></td>
<td>soccer ball</td>
<td>0.999</td>
</tr>
<tr>
<td>Row4</td>
<td><img src="image3.jpg" alt="Umbrella" /></td>
<td>umbrella</td>
<td>0.99</td>
</tr>
</tbody>
</table>
KNIME Deep Learning – TensorFlow Integration

TensorFlow Network Reader
Load TensorFlow network

Keras Network Reader
InceptionV3, pre-trained on ImageNet

Keras to TensorFlow Network Converter
Read and prepare data

DL Python Network Editor

DL Network Executor
Execute network

Format results
transform probabilities to predicted classes' labels
KNIME Deep Learning – What else?

• Customize existing networks
• Modify network architectures
• Transfer Learning
• Export and deploy networks
• ...

Make DL more approachable!
KNIME H2O Sparkling Water 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 Apache Spark
RECAP: KNIME H2O Machine Learning Integration (since 3.4)
RECAP: KNIME H2O Machine Learning Integration (since 3.4)
New: H2O Sparkling Water Integration
Scoring

Score models on tables

New: Deploy MOJO on Apache Spark
KNIME Server
– Jon Fuller –
Better Decision-making, Faster!

Load > Integrate > Transform > Analyze > Visualize

KNIME Analytics Platform

KNIME Server

Collaborate — Automate — Manage — Deploy

Open Source

KNIME Integrations

KNIME Extensions

Community Extensions

Partner Extensions
KNIME Software – On premise and in the cloud
AWS Marketplace Updates

- Latest versions of software available
- New: KNIME Server Small
- Free trial for KNIME Server Small

KNIME Server (BYOL)

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 16.04-3 LTS - 64-bit Amazon Machine Image (AMI)

KNIME Server Medium for AWS

Starting from $3.97/hr or from $28,000.00/yr (17% savings) for software + AWS usage fees

Use the KNIME Analytics Platform to automate advanced analytics, machine learning, or data prep/ETL tasks. Productionize your data science applications using the KNIME Server...

Linux/Unix, Ubuntu 16.04-3 LTS - 64-bit Amazon Machine Image (AMI)

KNIME Server Small for AWS

Starting from $1.16/hr or from $8,500.00/yr (17% savings) for software + AWS usage fees

Use the KNIME Analytics Platform to automate advanced analytics, machine learning, or data prep/ETL tasks. Productionize your data science team with collaboration and automation...

Linux/Unix, Ubuntu 16.04-3 LTS - 64-bit Amazon Machine Image (AMI)

KNIME Analytics Platform

Starting from $0.06 to $0.06/hr for software + AWS usage fees

At KNIME, we build software for fast, easy and intuitive access to advanced data science, helping organizations drive innovation. KNIME Analytics Platform brings the power...

Azure Marketplace Updates

• Latest versions of software available
• New: KNIME Server Medium (Marketplace billed)
New Feature: Show API definition

[Diagram showing KNIME Explorer with an option to show API definition and a Swagger UI window demonstrating API metadata and job-control endpoints.]
New Feature: Manage and Monitor (Preferences)

• **Features:**
  - Easier IT Operations.
  - Manage Analytics Platform preferences centrally
  - Include dependencies – e.g. driver files.
  - Deliver updates to configurations automatically.

• **Solution:** Management (Preferences) part of the Management Tools
New Feature: Manage and Monitor (Preferences)

- Different departments/teams have different requirements
- Multiple OS deployments
  - Windows 7
  - Windows 10
  - Linux
  - macOS

Marketing
- Windows 10
- Hive
- Spark

Finance
- Windows 7
- Oracle
- MS Access

R&D
- Linux and macOS
- Python
- R
New Feature: Manage and Monitor (Preferences)

Profiles can include:
- Client-profiles
  - Python-Linux
  - Python-macOS
  - R-Linux
  - R-macOS
  - Databases-Win7
  - Big Data-Win10

Profiles can include:
Preferences, drivers, and more
New Feature: Manage and Monitor (Preferences)

• Client-profiles
  • Default
    • default.epf
  • Python-Linux
    • python.epf
    • py-linux.sh
  • Python-macOS
    • python-mac.epf
    • py-mac.sh
  • R-Linux
    • R-linux.epf
  • R-macOS
    • R-mac.epf
• Databases
  • db.epf
  • oracle.jdbc
  • msaccess.jdbc
• Big Data
  • bigdata.epf

KNIME Server

client-profiles

Big Data-Win10

Databases-Win7

Python-macOS (etc)

Marketing

Finance

R&D
New Feature: Manage and Monitor (Preferences)

Recommendation:
Add default profiles to knime.ini, allow users to choose more profiles if needed
Guided Analytics
– Greg Landrum –
Guided Analytics?

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
- KNIME expert/data scientist creates workflow capturing best practices
- Typically deployed via KNIME WebPortal
- Heavy use of KNIME’s 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

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

• Heavy use of KNIME’s interactive JavaScript views for the user experience
Demo