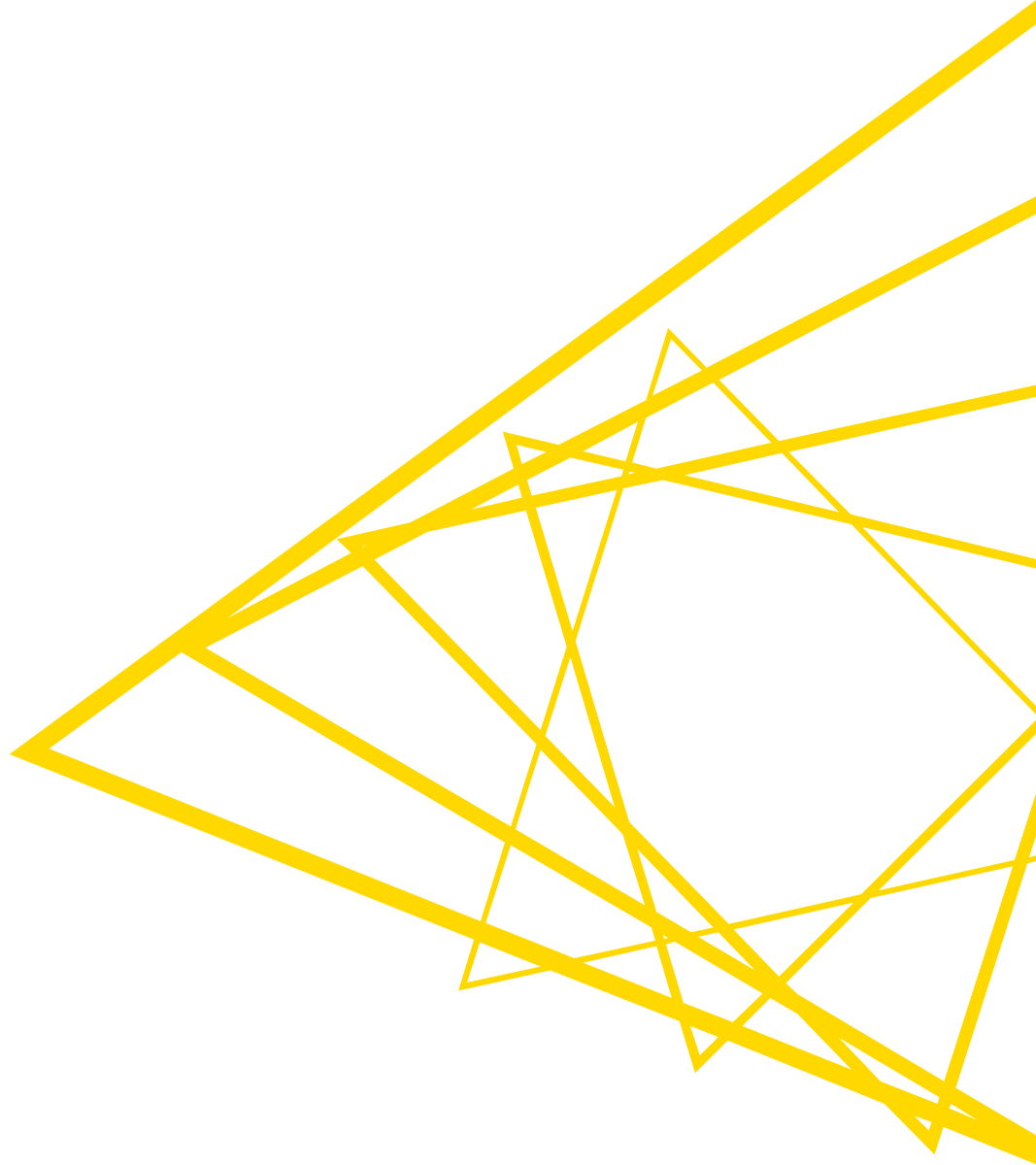




What's New

Bernd Wiswedel
KNIME

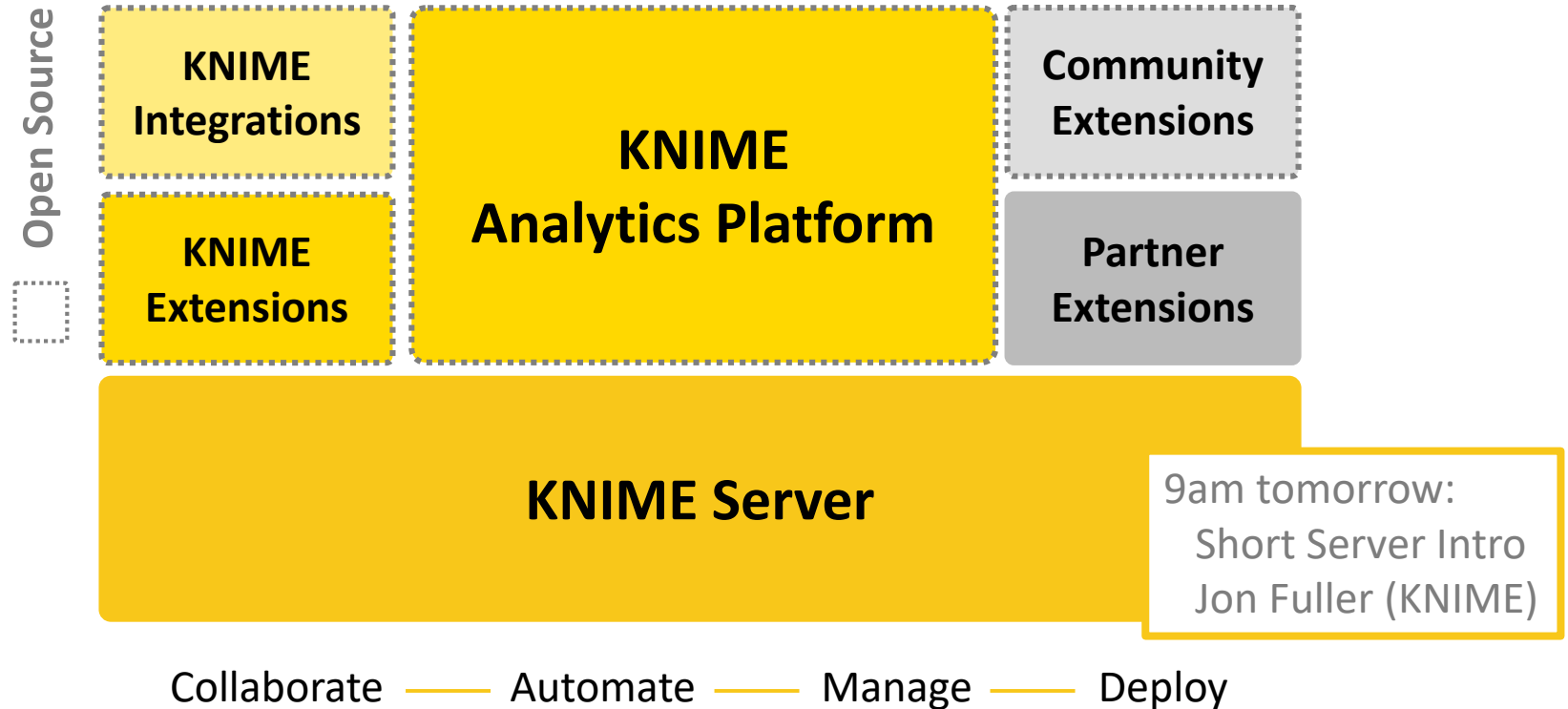


What this session is about

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

KNIME Software

Load > Integrate > Transform > Analyze > Visualize



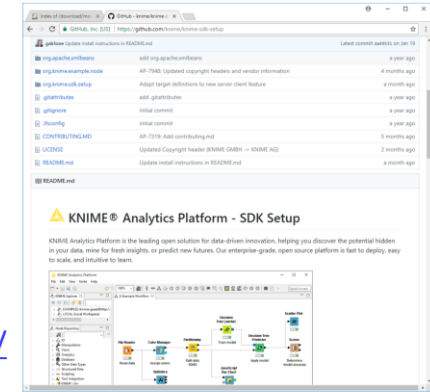
Git Client Nodes

- 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/>

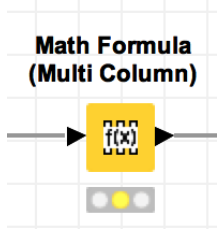


© 2018 KNIME AG. All Rights Reserved.

5



Math Formula (Multi Column)*



Math Expression Flow Variables Job Manager Selection Memory Policy

Selected Columns (assigned one by one to 'CURRENT_COLUMN')

☒ Manual Selection ☐ Wildcard/Regex Selection

Exclude

Column(s): Search

☐ Select all search hits

- sex
- citizenship status
- category of work
- english ability
- vehicle occupancy
- means of transportation to work
- marital status
- military service

☒ Enforce exclusion

Select

add >>

add all >>

<< remove

<< remove all

Include

Column(s): Search

☐ Select all search hits

- interest income
- other income
- public assistance income
- retirement income
- self employment income
- supplementary social security income
- social security income
- wage income

☐ Enforce inclusion

Math Expression

Column List	Category	Description
ROWINDEX	All	
ROWCOUNT		
CURRENT_COLUMN		
weight	Function	
age	ROWCOUNT	
quarter of birth	ROWINDEX	
income adjustment	pi	
interest income	e	
other income	COL_MIN(col_name)	
public assistance income	COL_MAX(col_name)	
retirement income	COL_MEAN(col_name)	
self employment income	COL_MEDIAN(col_name)	
supplementary social security income	COL_SUM(col_name)	
	COL_STDDEV(col_name)	
	COL_VAR(col_name)	

Flow Variable List

Expression

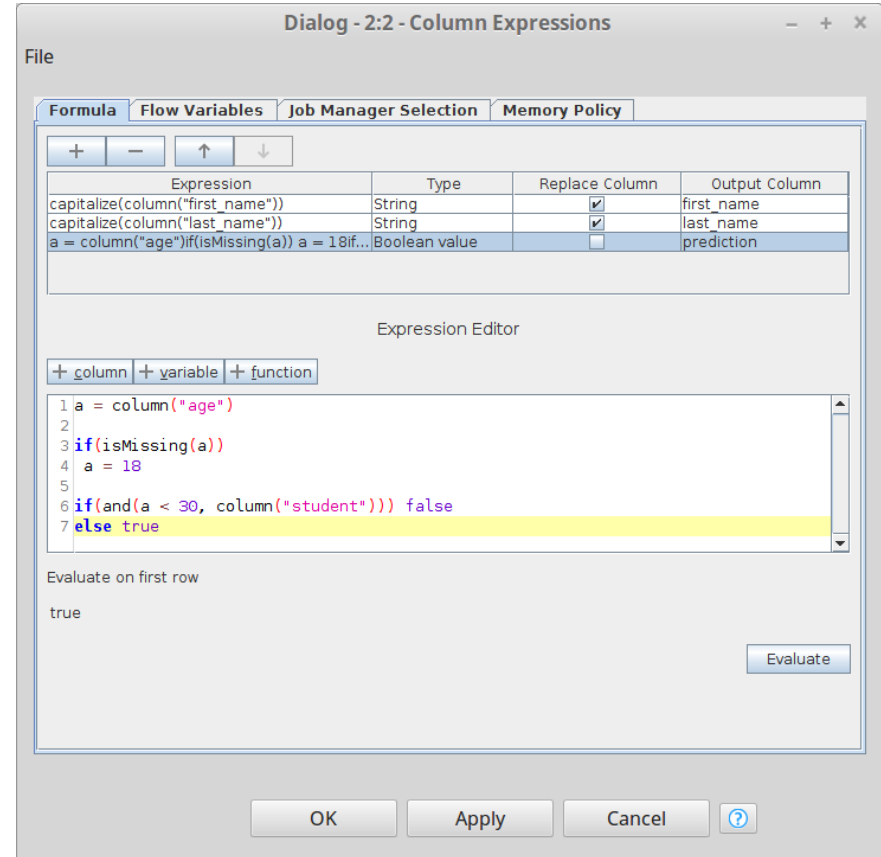
```
$$CURRENT_COLUMN$$ * $income adjustment$/1e6
```

*) This is not only for Dean.

Column Expressions

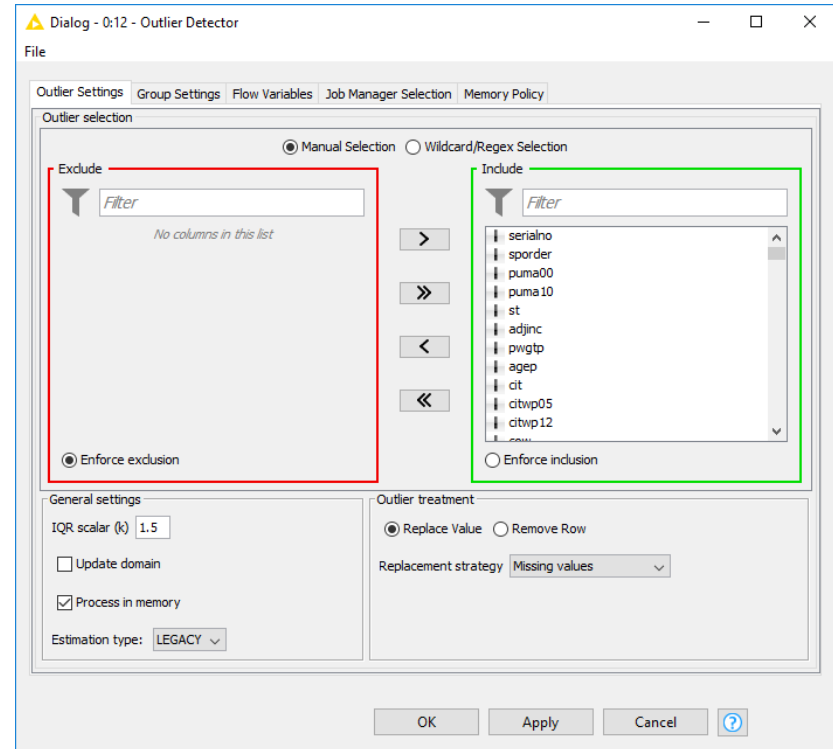
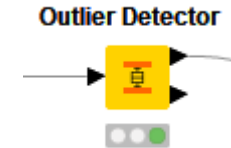
- Swiss army knife for deriving new columns
- Complete type support
- Multiple expressions

Column Expressions



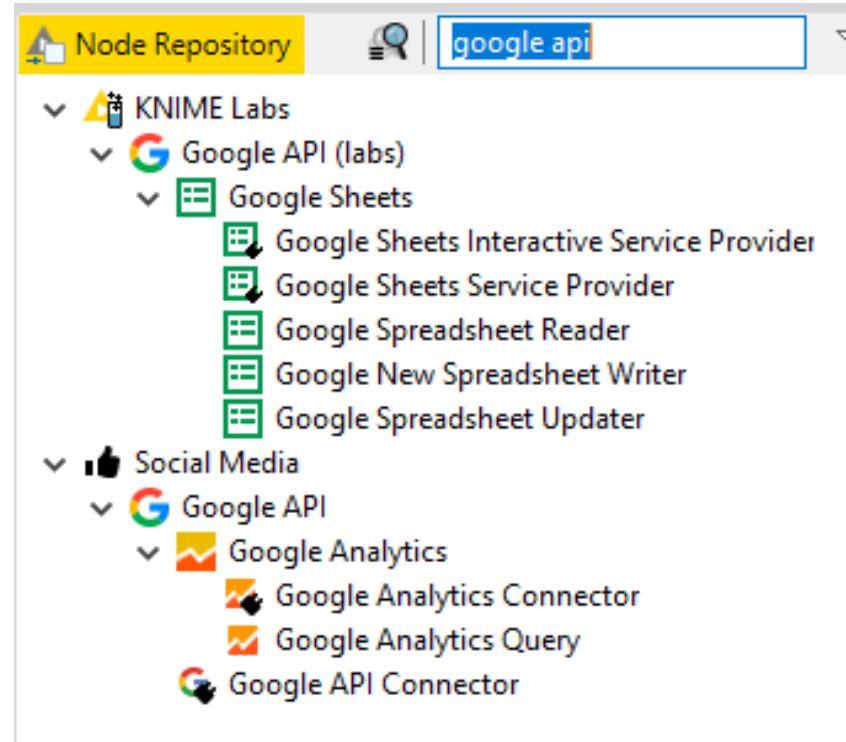
Outlier Handling

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



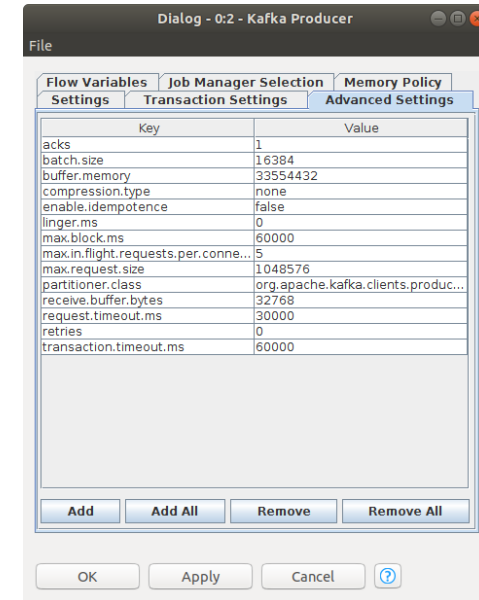
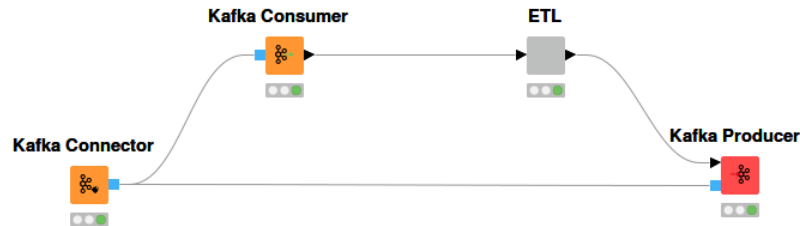
Google Connectivity (Labs)

- Read and Write Sheets in Google Drive
- Access via API Token & Interactive
- Same concepts as Google Analytics Nodes

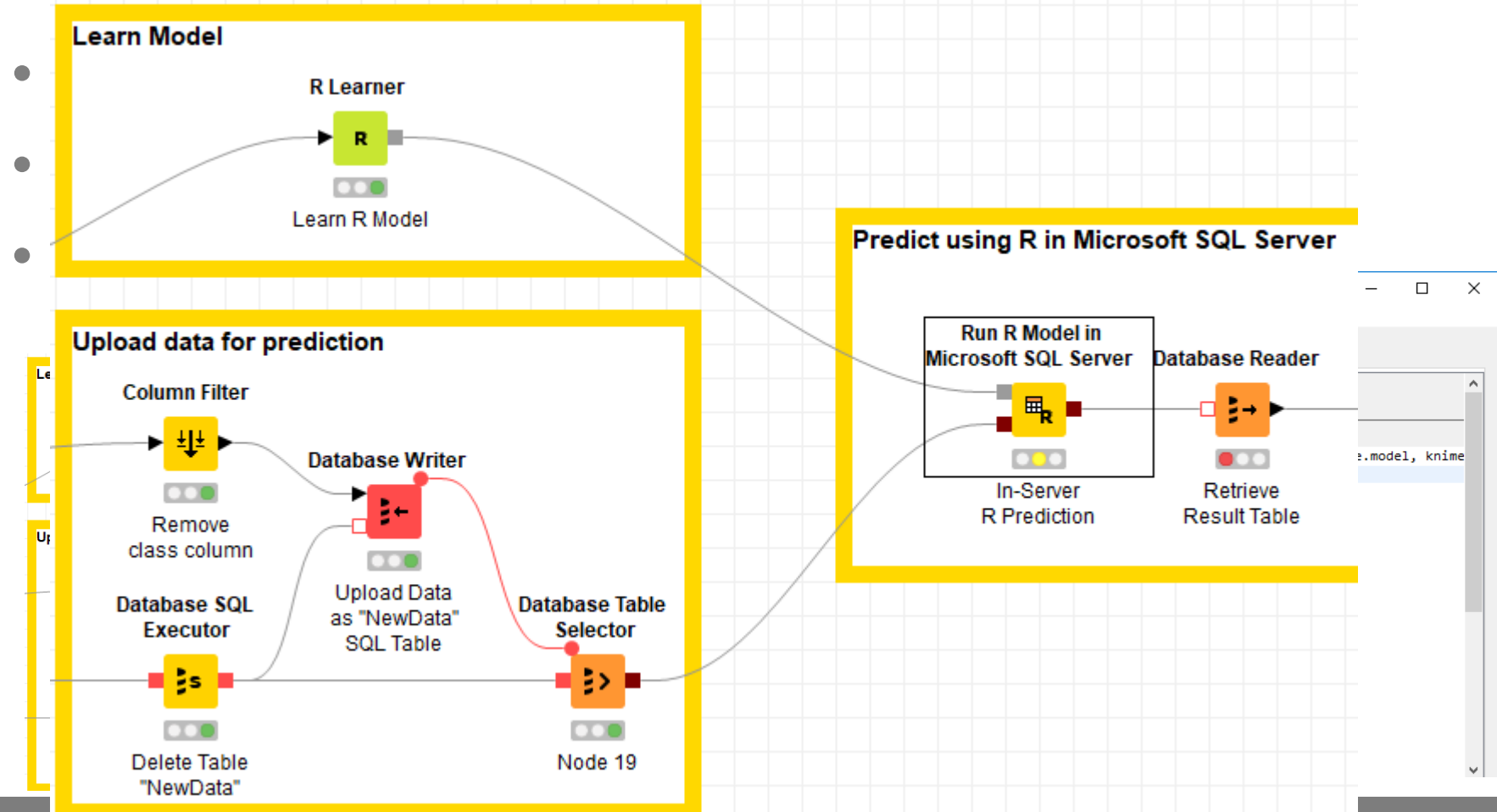


Kafka Integration (Labs)

- Apache Kafka – distributed publish-subscribe streaming platform
- KNIME Integration supporting sending and consuming messages

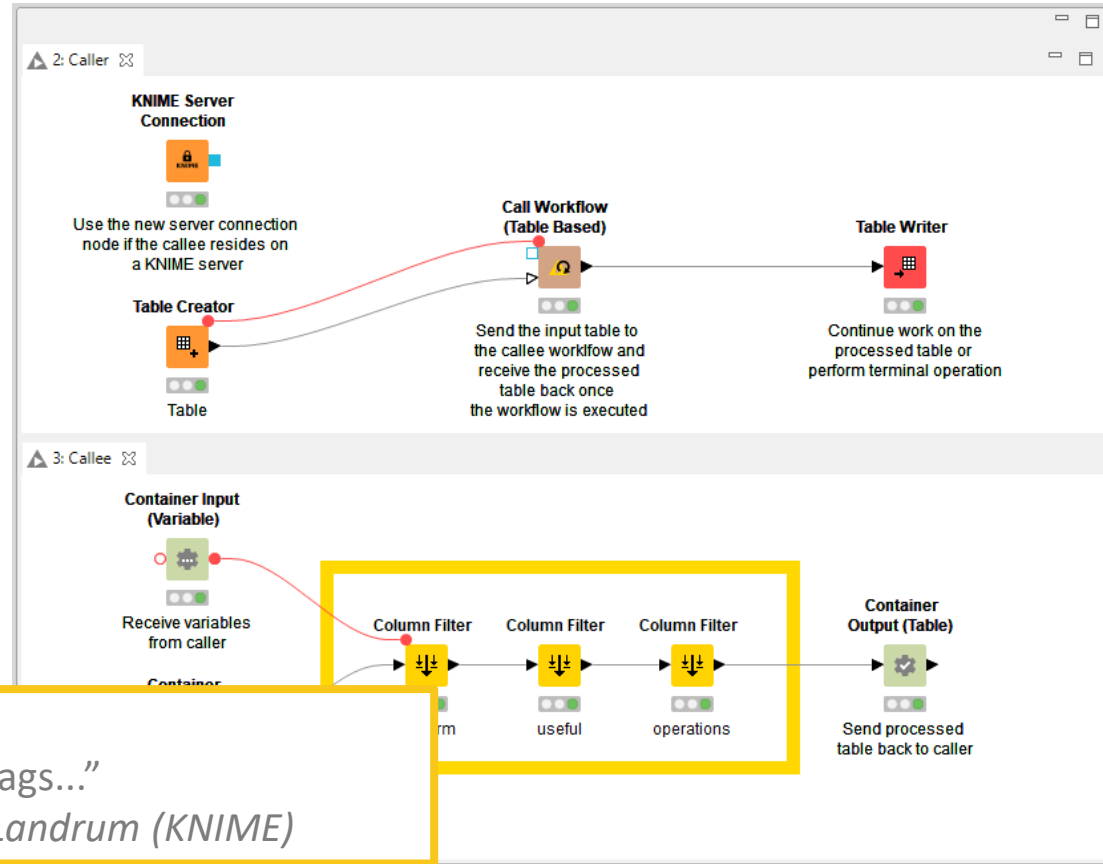


Azure/Microsoft SQL Server R Services



Workflow Automation – “Call Workflow”

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



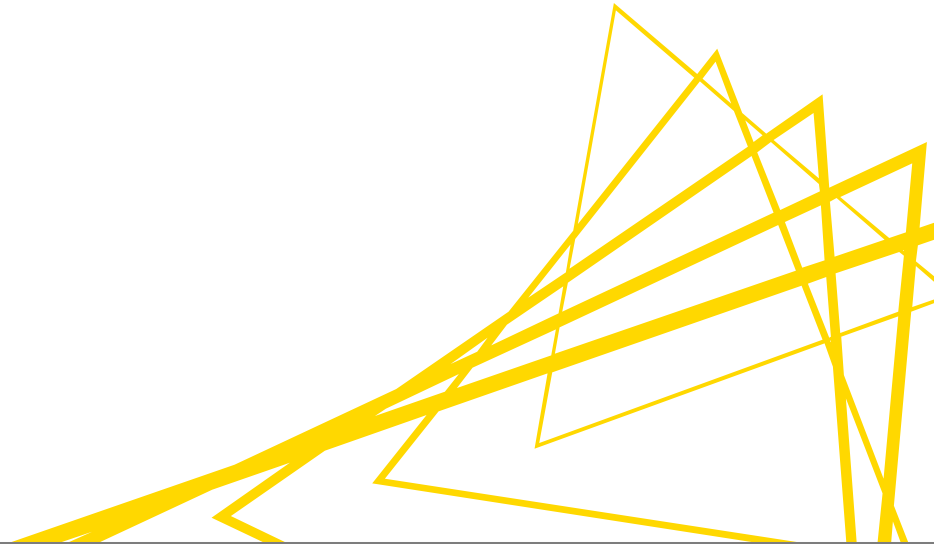
Tomorrow at 9:30
“On Monsters and Tags...”
Jeany Prinz & Greg Landrum (KNIME)

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 –



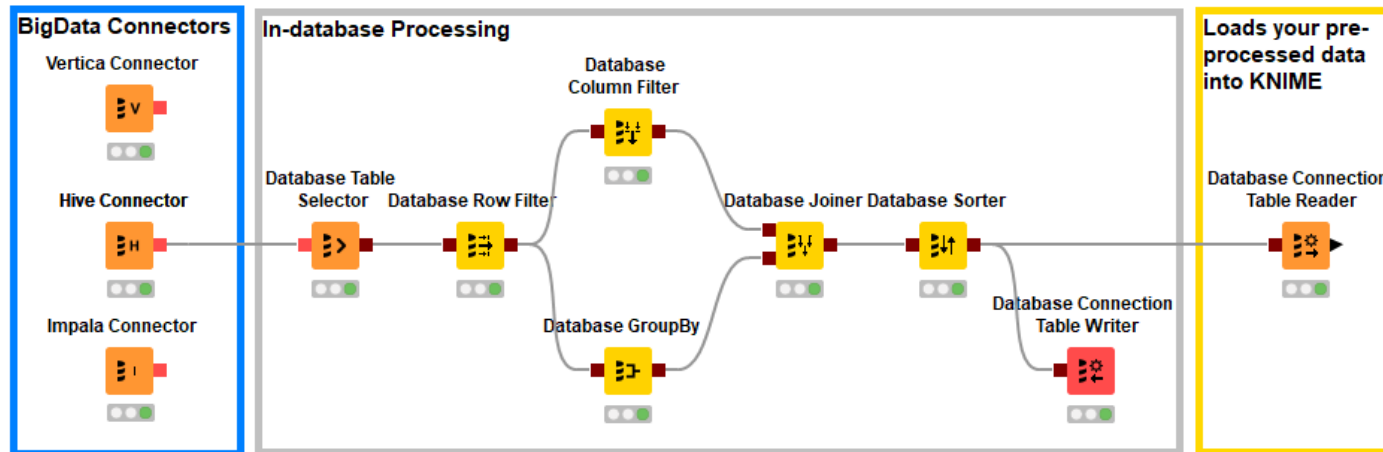
KNIME Big Data Connectors

- 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



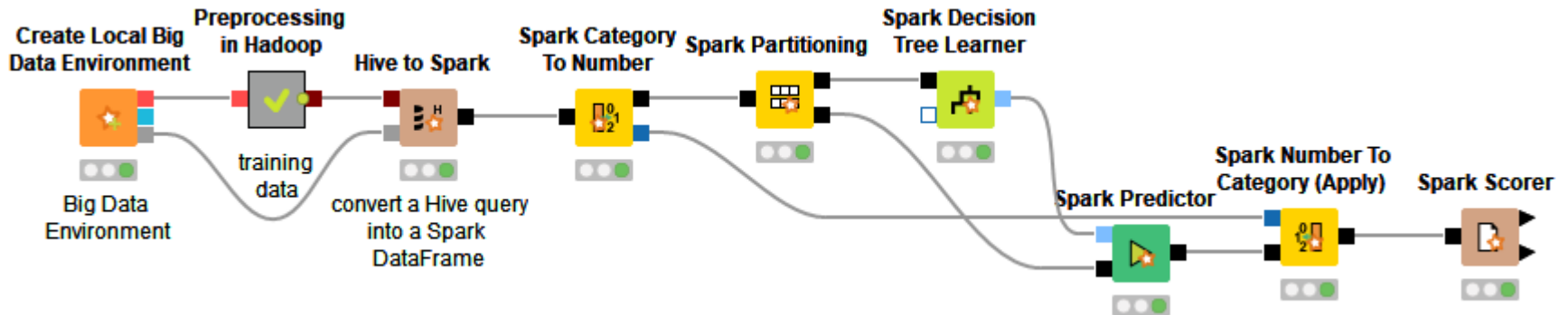
KNIME Extension for Apache Spark

- 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

New

Parquet Writer



ORC Writer



Parquet Reader



ORC Reader



Updated

Hive Loader



Impala Loader



Spark to Hive



Spark to Impala

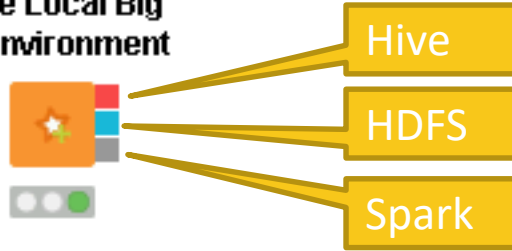


KNIME Extension for Apache Spark

- 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



Provides everything to

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

Batteries included:

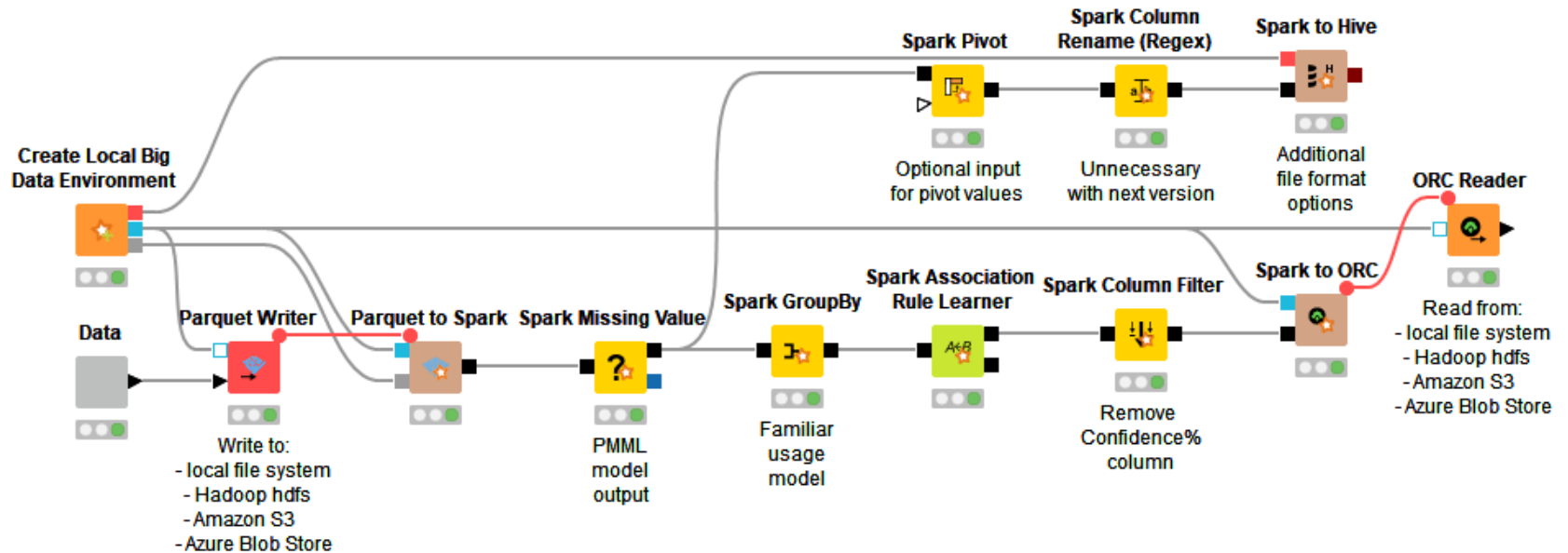
- no further software installation required
- no cluster required
- no jobserver required

A screenshot of the 'Local Big Data Environment Settings' dialog box. The dialog has a title bar with 'Local Big Data Environment Settings' and a 'Flow Variables' button. The main content area includes the following fields and options:

- Context name:** A text field containing 'krimeSparkContext'.
- Number of threads:** A spinner box set to '2'.
- Action to perform on dispose:** A group box containing three radio buttons: 'Destroy Spark context', 'Delete Spark DataFrames' (selected), and 'Do nothing'.
- Use custom Spark settings:** An unchecked checkbox.
- Custom Spark settings:** A text area containing the following text:

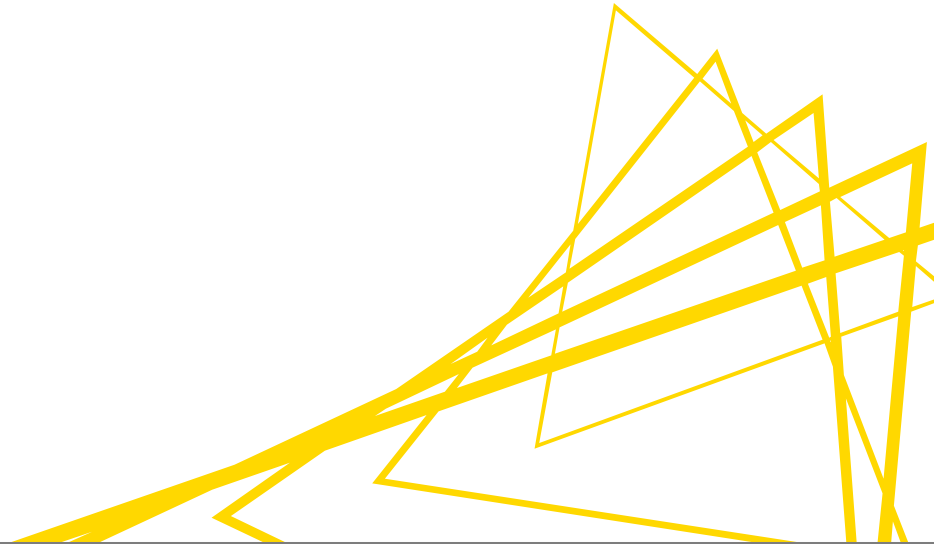
```
spark.jars: /path/to/some.jar
spark.sql.shuffle.partitions: 100
```
- SQL Support:** A group box containing three radio buttons: 'Spark SQL only', 'HiveQL', and 'HiveQL and provide JDBC connection' (selected).
- Use custom Hive data folder (Metastore DB & Warehouse):** An unchecked checkbox.
- Folder selection:** A text field with a dropdown arrow, a 'Browse...' button, and a 'V' icon.
- Hide warning about an existing local Spark context:** An unchecked checkbox.

Demo



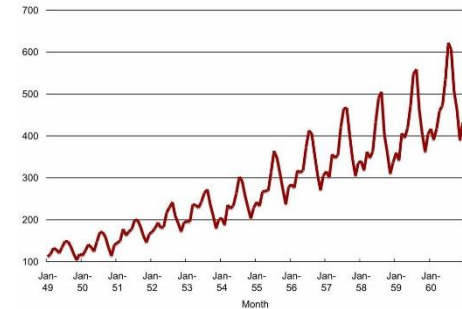
Deep Learning in KNIME

- Christian Dietz -



Deep Learning?

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

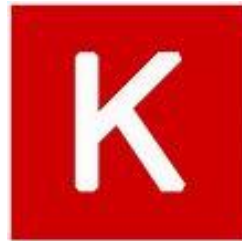


KNIME Deep Learning

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



Since 3.2

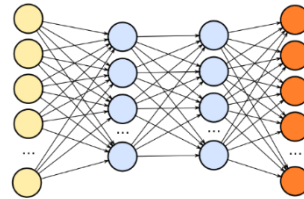


Since 3.5



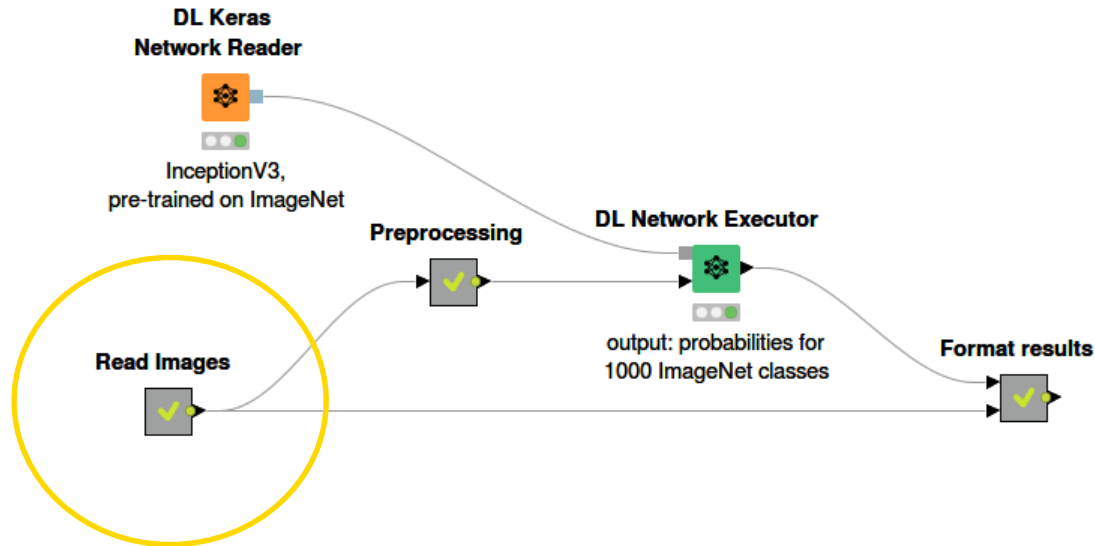
Since 3.6

KNIME Deep Learning – Executing

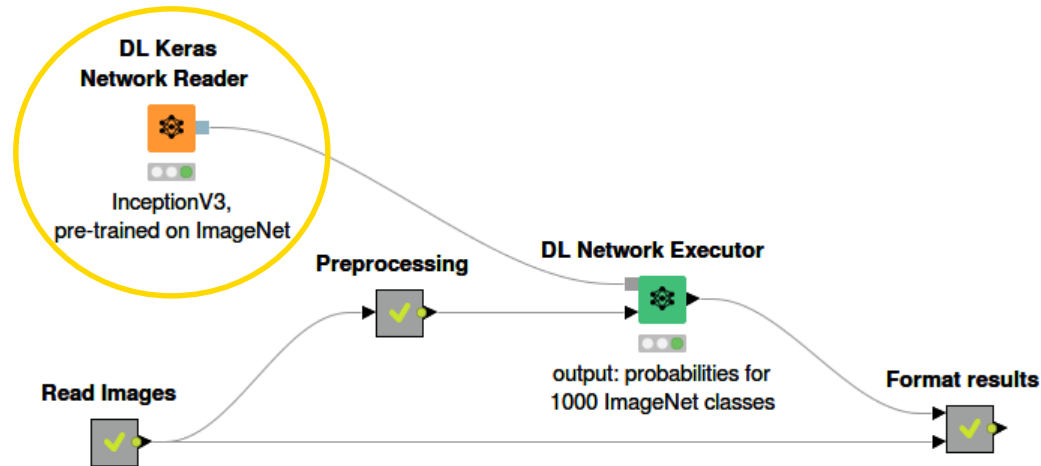


Soccer ball

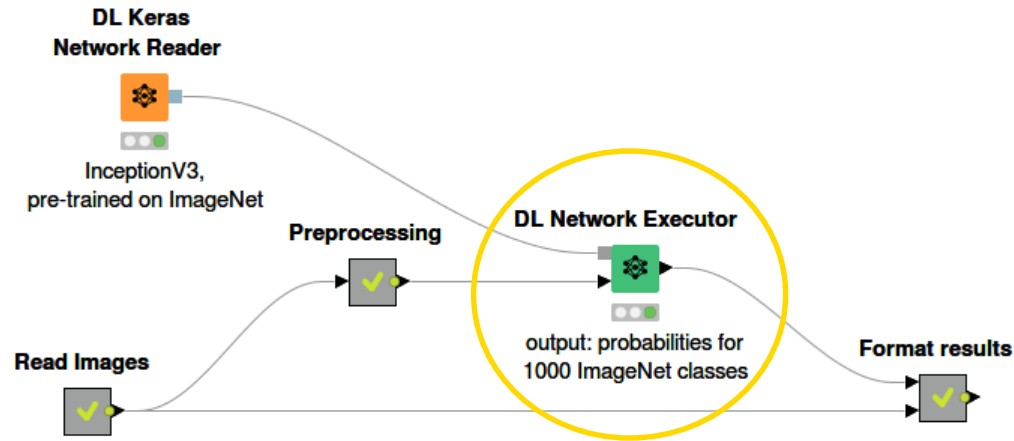
KNIME Deep Learning – Executing



KNIME Deep Learning – Executing






KNIME Deep Learning – Executing

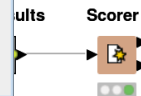
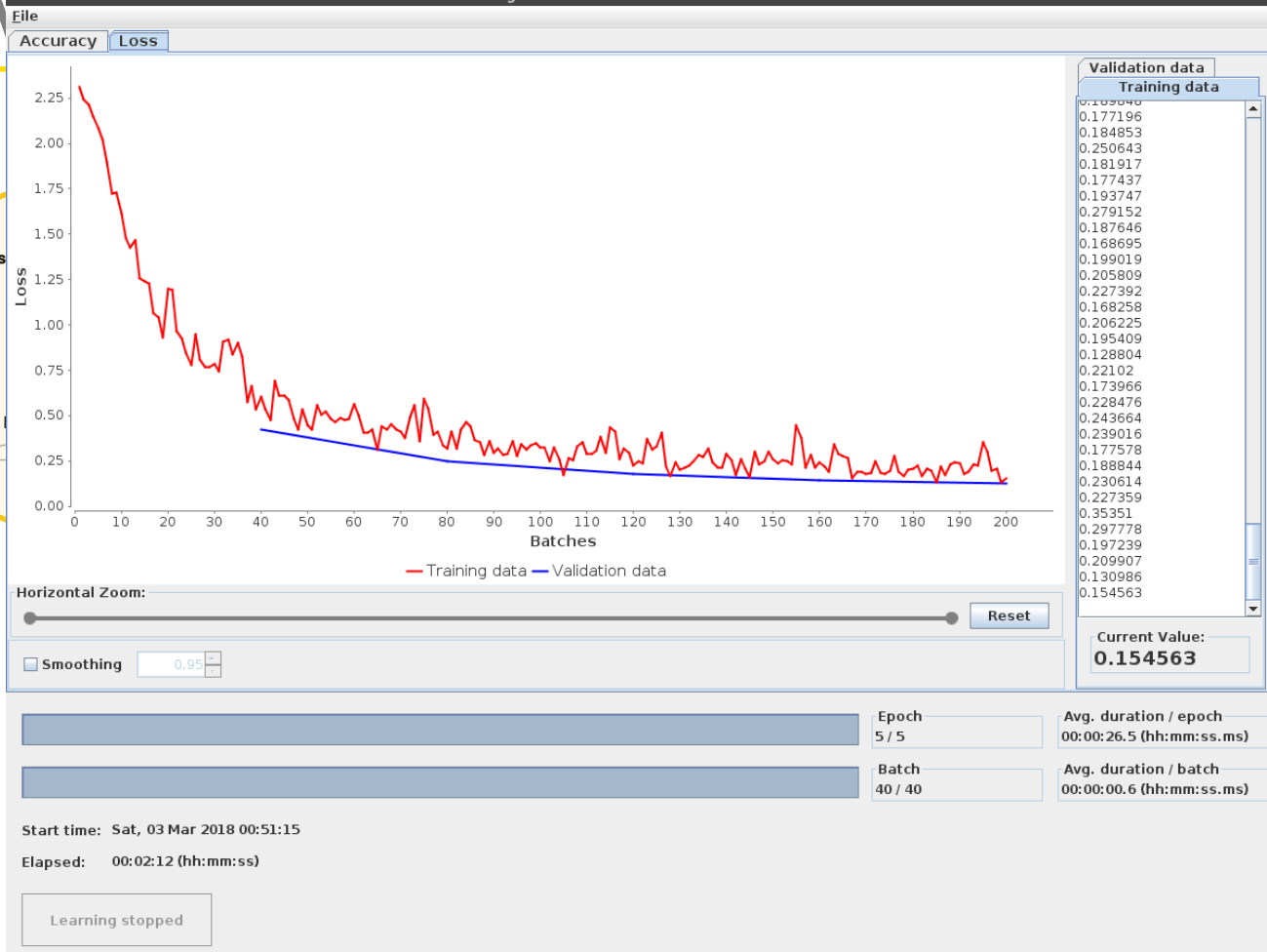


Output data - 0:75:65 - Column Resorter

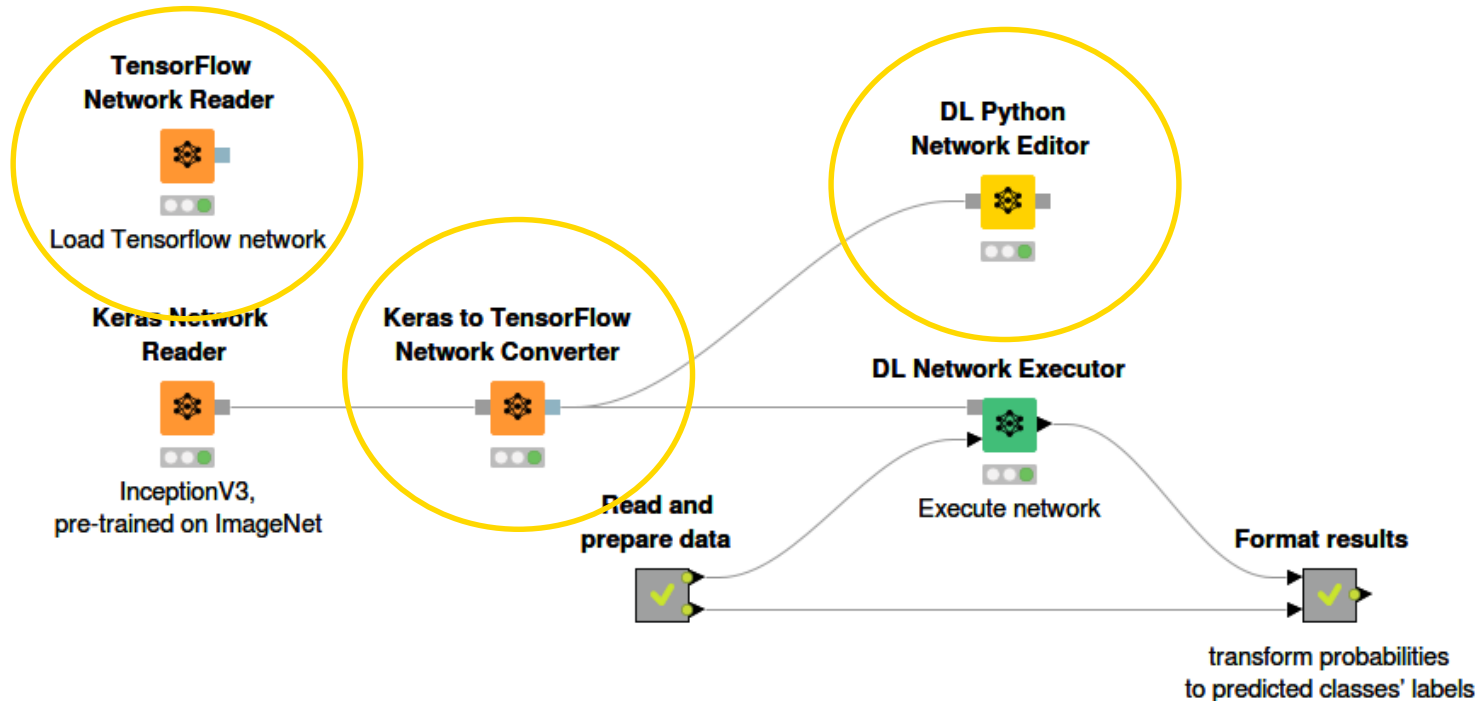
File Hilite Navigation View

Table "default" - Rows: 7 Spec - Columns: 3 Properties Flow Variables

Row ID	Image	S Predicted Class	D Class ...
Row2		dalmatian, coach dog, carriage dog	0.966
Row3		soccer ball	0.999
Row4		umbrella	0.98



KNIME Deep Learning – TensorFlow Integration



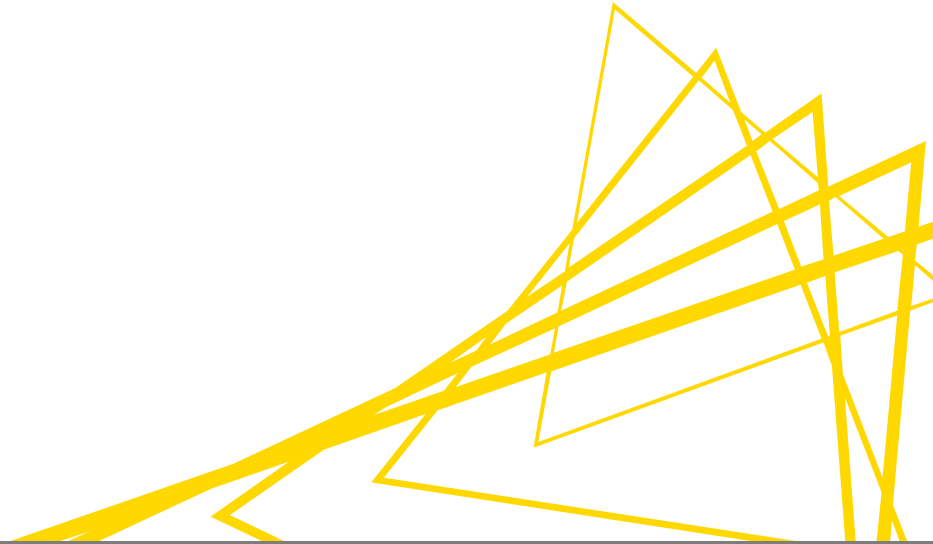
KNIME Deep Learning – What else?

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

Make DL more approachable!

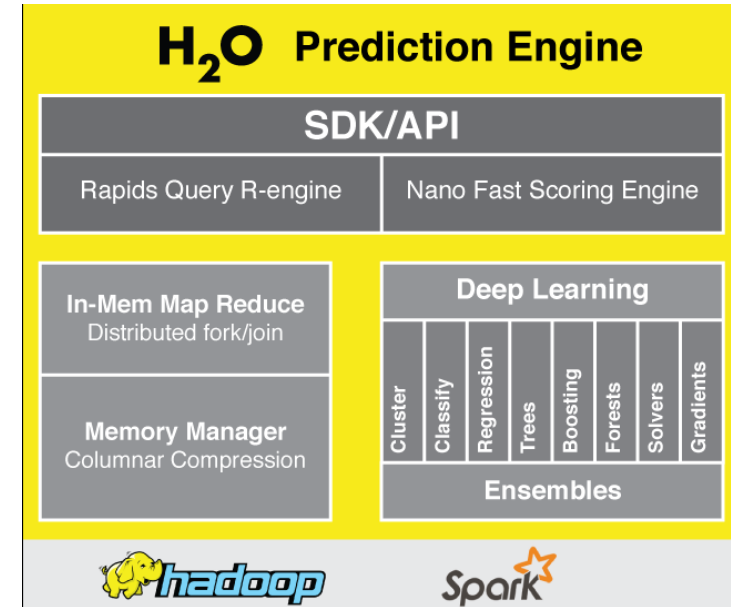
KNIME H2O Sparkling Water Integration

- Christian Dietz -



What is H2O?

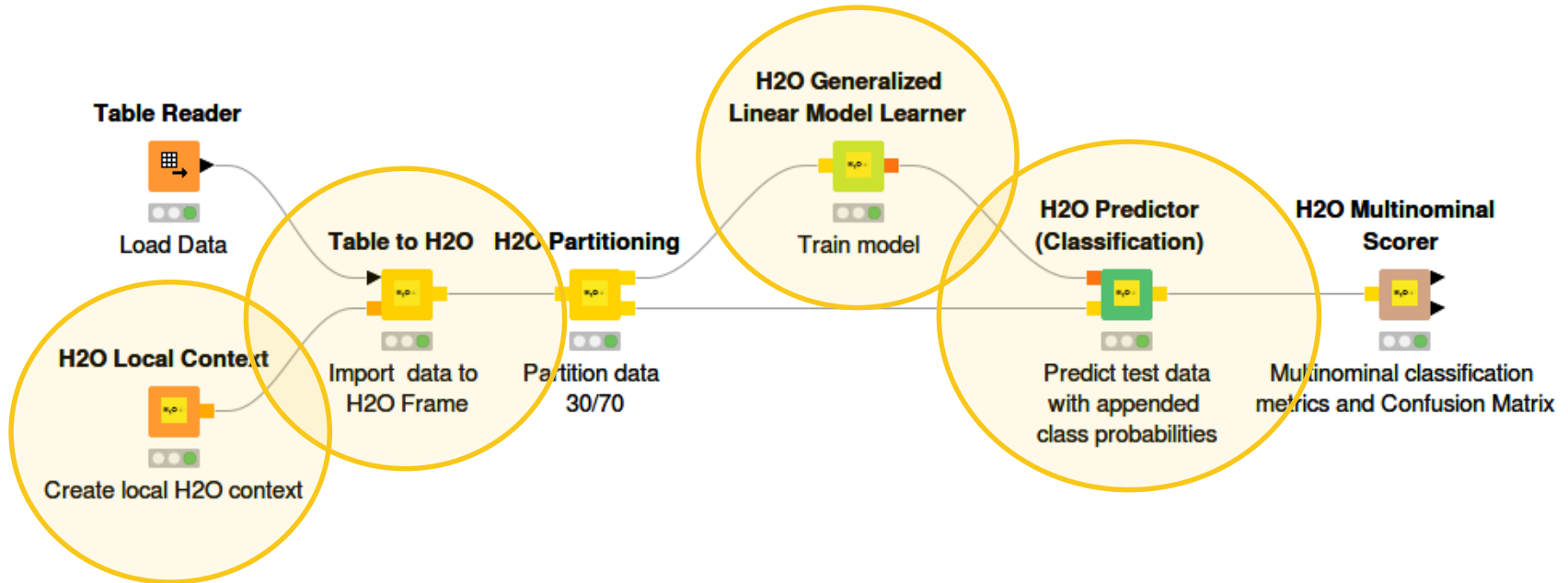
- Open-source machine learning framework
- Fast, scalable and in-memory implementations of state-of-the-art machine learning algorithms.
- Massively scalable big data analysis on for example on Apache Spark



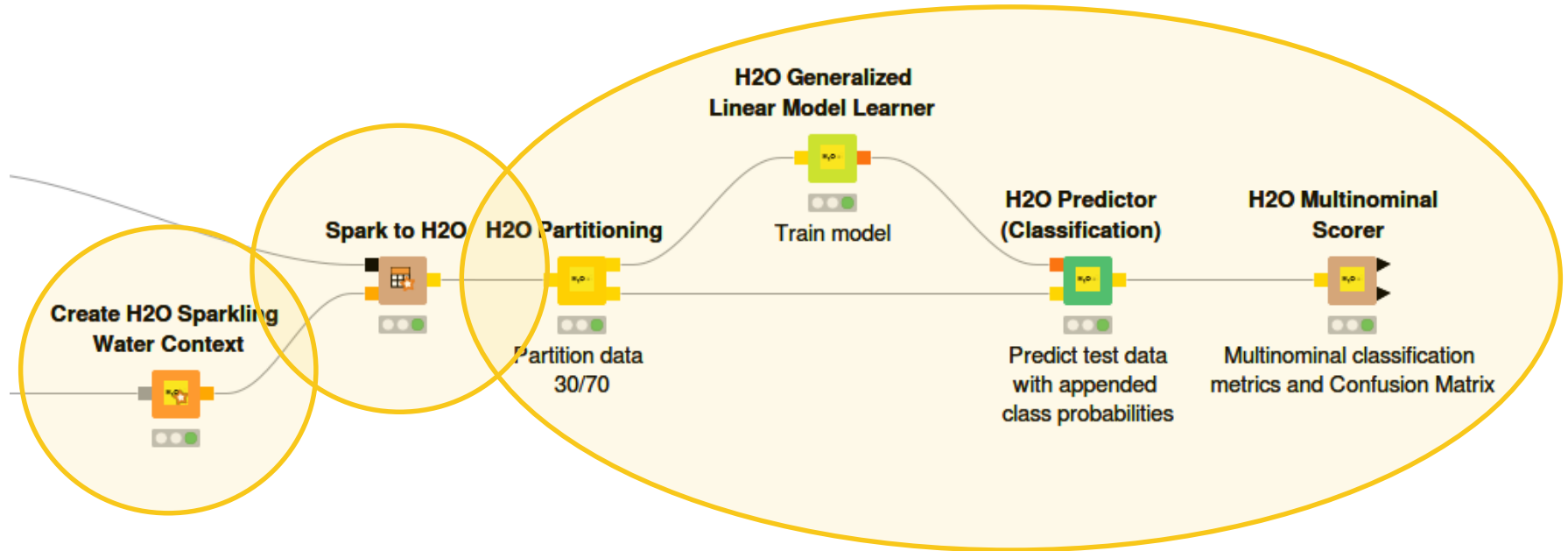
RECAP: KNIME H2O Machine Learning Integration (since 3.4)

- ▼ H2O Machine Learning
 - ▼ IO
 - ▼ Read
 - Spark to H2O
 - Table to H2O
 - ▼ Write
 - H2O To Spark
 - H2O to Table
 - Create H2O Sparkling Water Context
 - H2O Local Context
 - ▼ MOJOs
 - ▼ Spark
 - H2O MOJO Predictor (Autoencoder)
 - H2O MOJO Predictor (Classification)
 - H2O MOJO Predictor (Cluster Assigner)
 - H2O MOJO Predictor (Dimension Reduction)
 - H2O MOJO Predictor (Regression)
 - H2O MOJO Predictor (Word Embedding)
 - H2O MOJO Reader
 - H2O MOJO Writer
 - ▼ Manipulation
 - H2O Column Filter
 - H2O Model to MOJO
 - H2O Partitioning
 - ▼ Misc
 - H2O Statistics
 - ▼ Models
 - ▼ Generalized Linear Model
 - H2O Generalized Linear Model Learner
 - H2O Generalized Linear Model Learner (Regression)
- ▼ Generalized Low Rank Models
 - H2O Generalized Low Rank Models (Missing Value Impute)
- ▼ Gradient Boosting Machine
 - H2O Gradient Boosting Machine Learner
 - H2O Gradient Boosting Machine Learner (Regression)
- ▼ Naive Bayes
 - H2O Naive Bayes Learner
- ▼ PCA
 - H2O PCA
 - H2O PCA Apply
 - H2O PCA Compute
- ▼ Random Forest
 - H2O Random Forest Learner
 - H2O Random Forest Learner (Regression)
- ▼ k-Means
 - H2O k-Means
 - H2O Cluster Assigner
 - H2O Predictor (Classification)
 - H2O Predictor (Regression)
- ▼ Scoring
 - H2O Binomial Scorer
 - H2O Cross Validation Loop Start
 - H2O Multinomial Scorer
 - H2O Regression Scorer

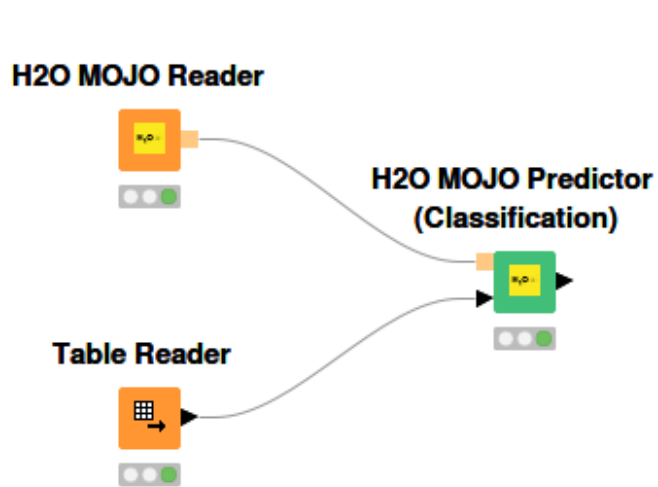
RECAP: KNIME H2O Machine Learning Integration (since 3.4)



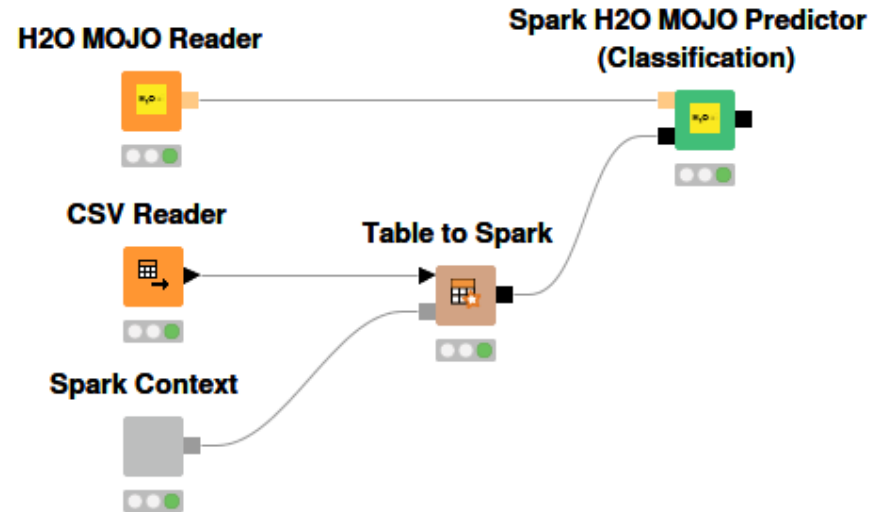
New: H2O Sparkling Water Integration



Scoring



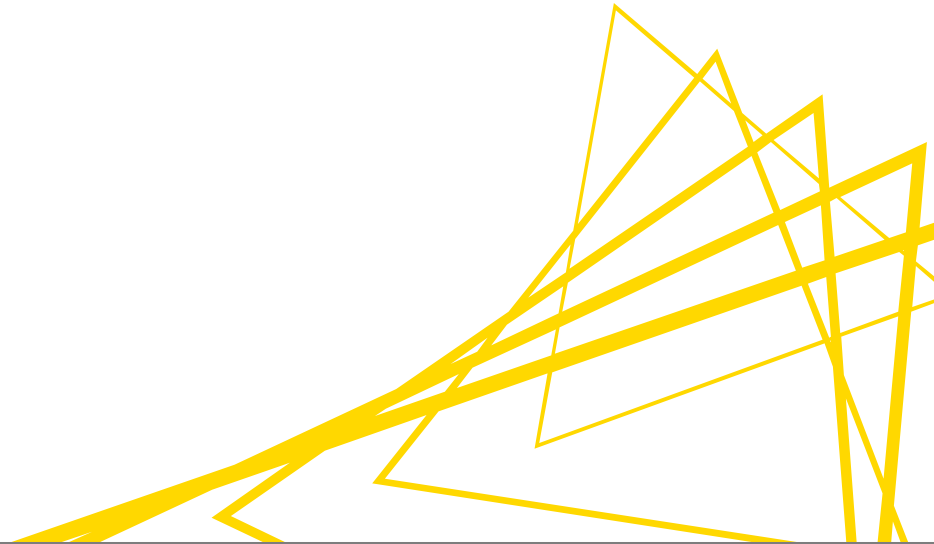
Score models on tables



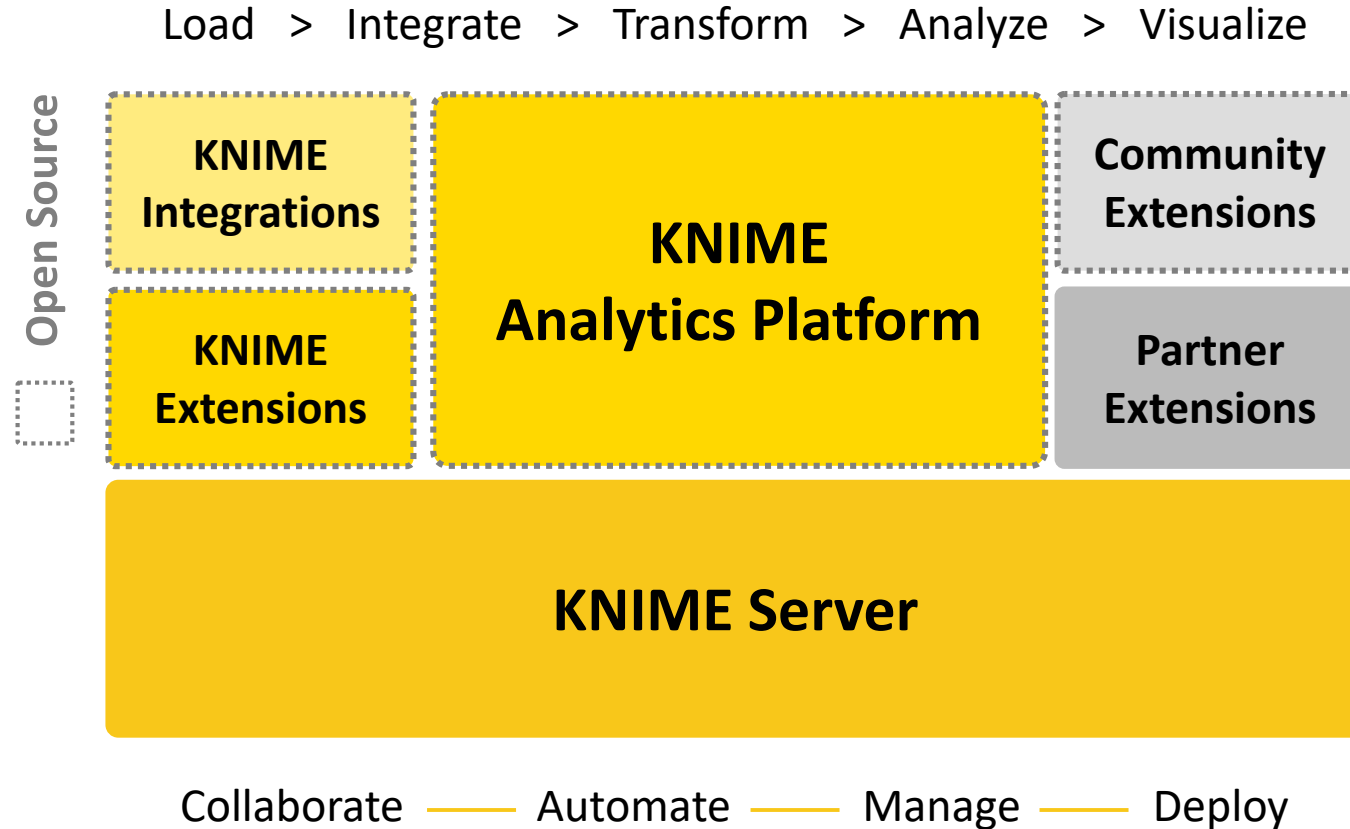
New: Deploy MOJO on Apache Spark

KNIME Server

— Jon Fuller —



Better Decision-making, Faster!



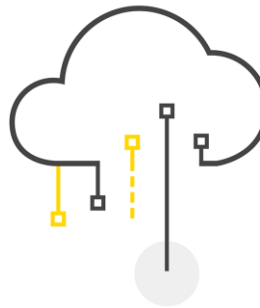
KNIME Software – On premise and in the cloud



On-Premise

**KNIME
Analytics
Platform**

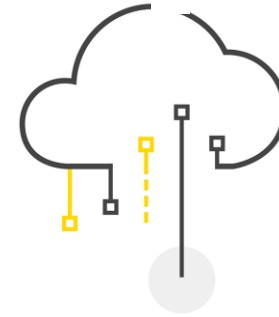
KNIME Server



aws partner network

**KNIME
Analytics
Platform**

KNIME Server



Azure

**KNIME
Analytics
Platform**

KNIME Server

AWS Marketplace Updates

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



KNIME Server (BYOL)

★★★★★ (0) | Version 4.7.2 | 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 16.04-3 LTS - 64-bit Amazon Machine Image (AMI)



KNIME Server Medium for AWS

★★★★★ (0) | Version 4.7.2 | Sold by [KNIME.com](#)

Starting from **\$3.97/hr** or from **\$29,000.00/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

★★★★★ (0) | Version 4.7.2 | Sold by [KNIME.com](#)

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

★★★★★ (0) | Version 3.6.1-a | Sold by [KNIME.com](#)

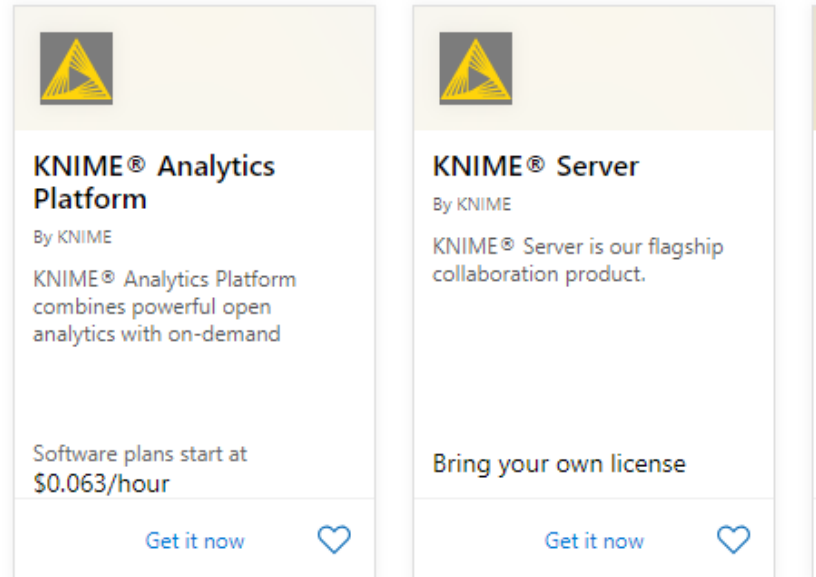
Starting from **\$0.06 to \$0.06/hr** for software + AWS usage fees

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...

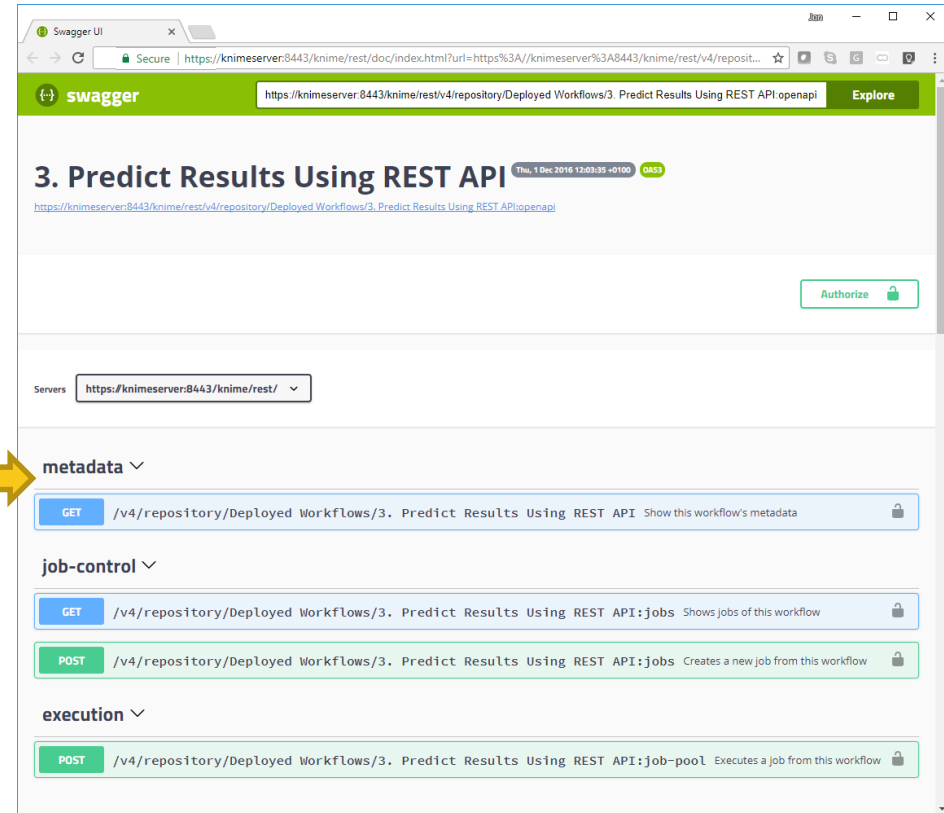
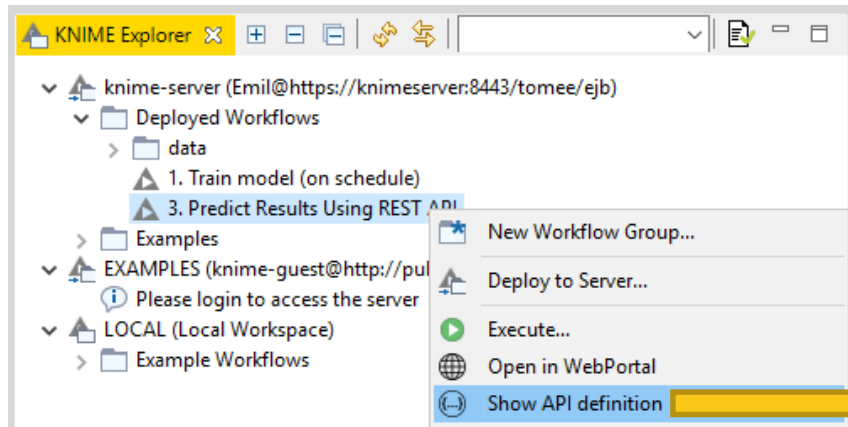
Windows, Windows Server 2016 Base Windows Server 2016 Base - 64-bit Amazon Machine Image (AMI)

Azure Marketplace Updates

- Latest versions of software available
- New: KNIME Server Medium (Marketplace billed)



New Feature: Show API definition

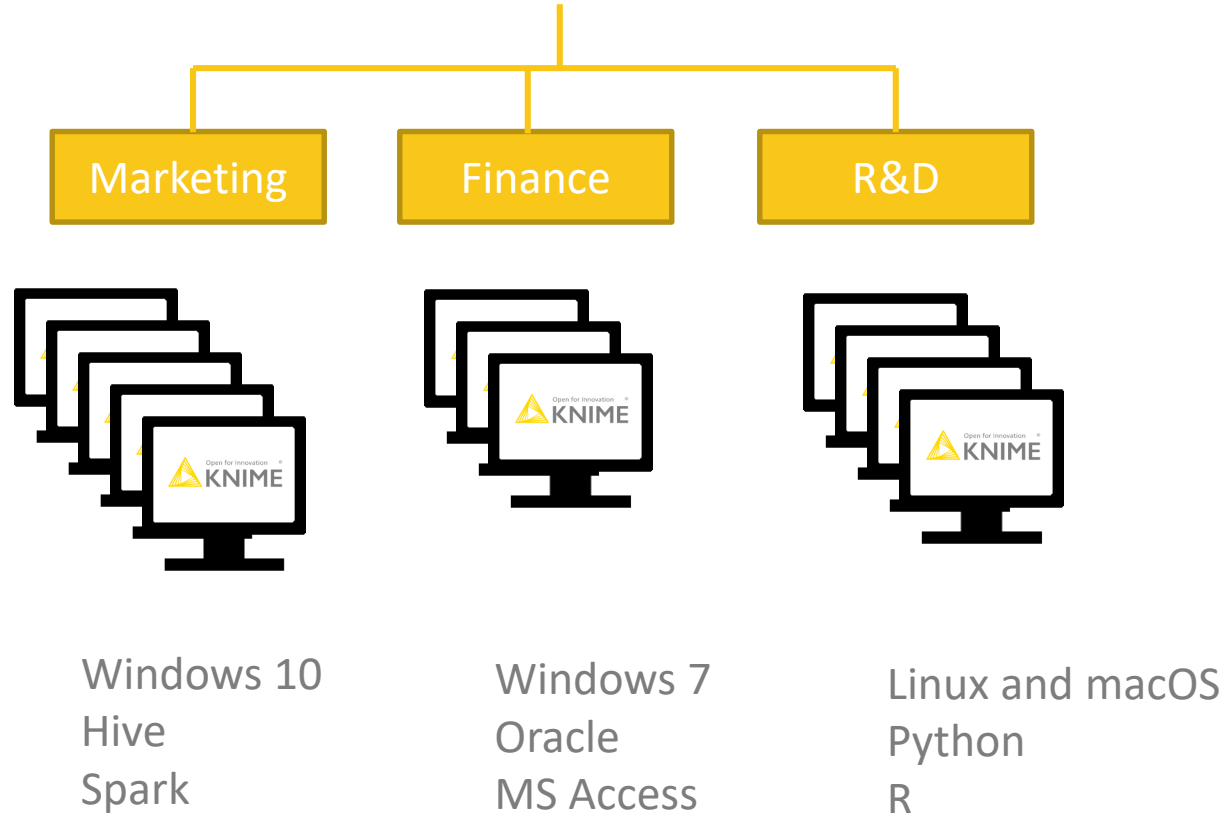


New Feature: Manage and Monitor (Preferences)

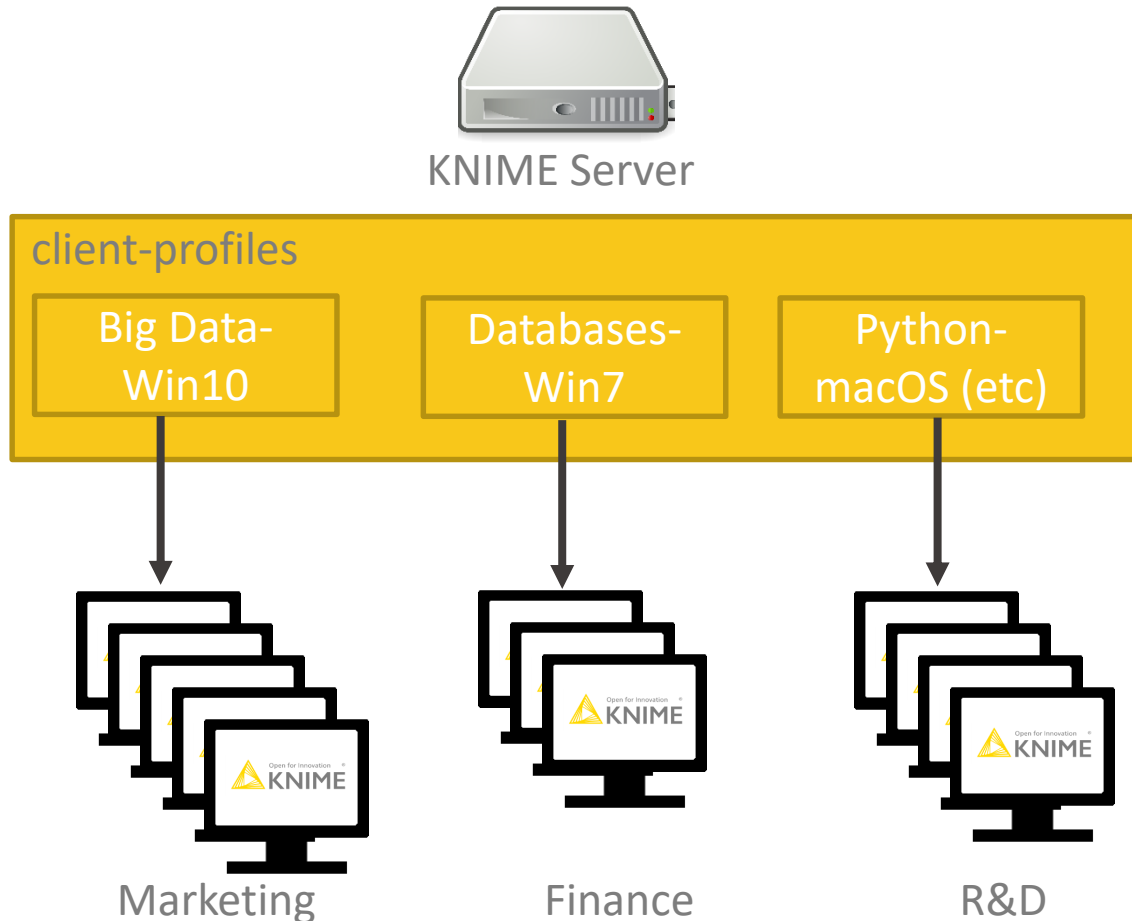
- **Features:**
 - Easier IT Operations.
 - Manage Analytics Platform preferences centrally
 - Include dependencies – e.g. driver files.
 - Deliver updates to configurations automatically.
- **Solution:** Management (Preferences) part of the Management Tools

New Feature: Manage and Monitor (Preferences)

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



New Feature: Manage and Monitor (Preferences)

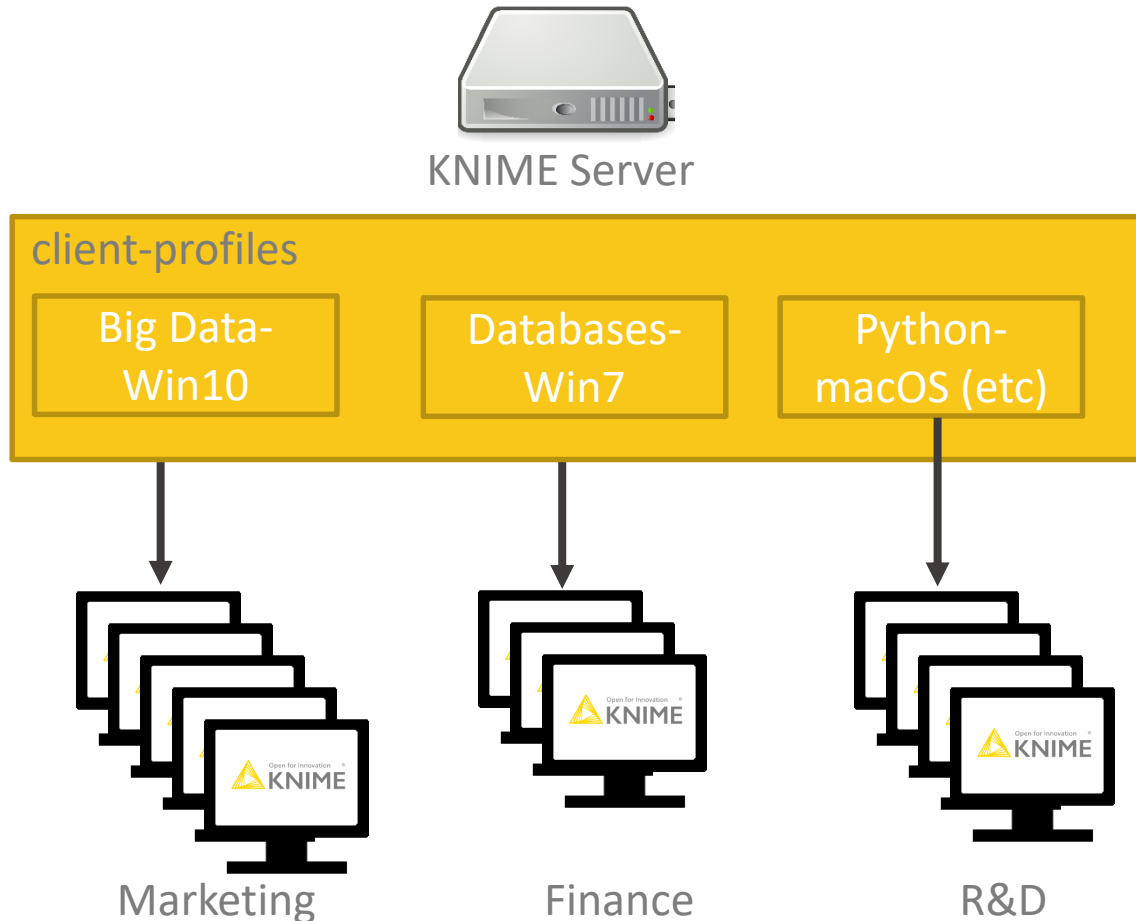


- Client-profiles
 - Python-Linux
 - Python-macOS
 - R-Linux
 - R-macOS
 - Databases-Win7
 - Big Data-Win10



Profiles can include:
Preferences, drivers, and more

New Feature: Manage and Monitor (Preferences)

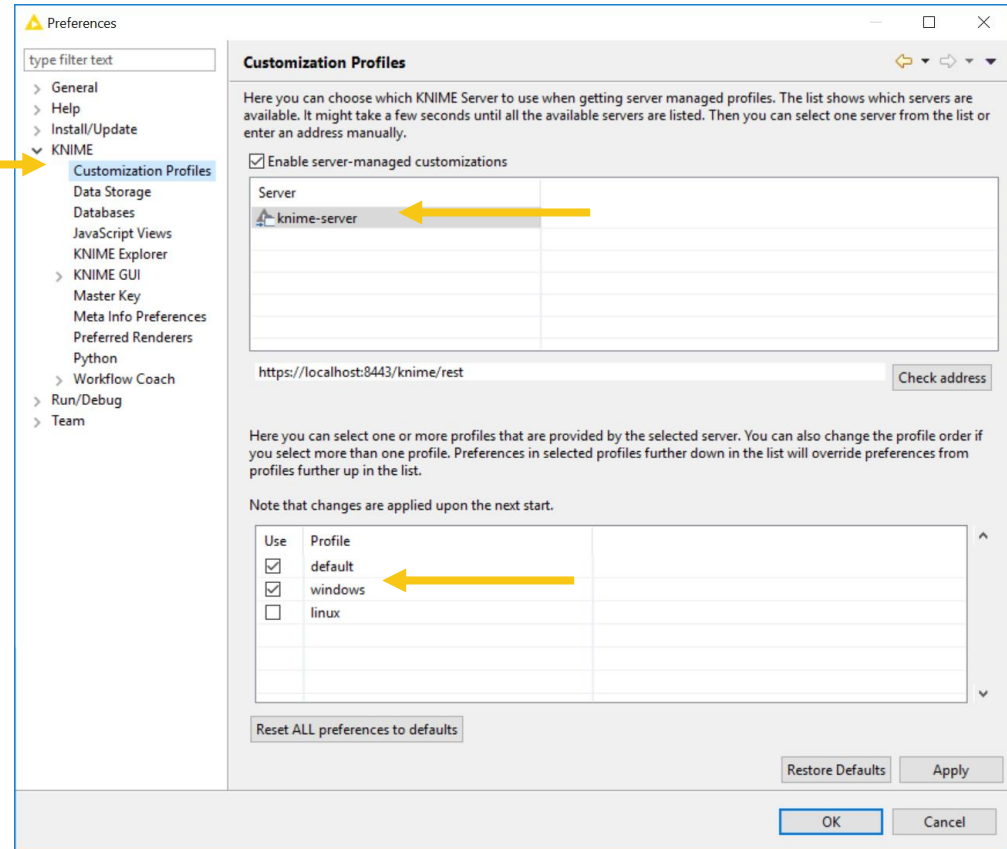


- Client-profiles
 - Default
 - default.epf
 - Python-Linux
 - python.epf
 - py-linux.sh
 - Python-macOS
 - python-mac.epf
 - py-mac.sh
 - R-Linux
 - R-linux.epf
 - R-macOS
 - R-mac.epf
 - Databases
 - db.epf
 - oracle.jdbc
 - msaccess.jdbc
 - Big Data
 - bigdata.epf

New Feature: Manage and Monitor (Preferences)



Recommendation:
Add default profiles
to knime.ini, allow
users to choose
more profiles if
needed



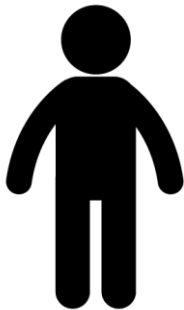
Guided Analytics

– Greg Landrum –



Guided Analytics?

Someone
who needs
data /
analytics



KNIME expert



What often happens

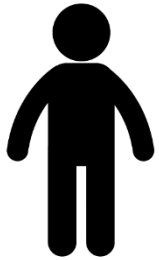
Someone
who needs
data /
analytics

KNIME expert



What often happens

Someone
who needs
data /
analytics



KNIME expert



What often happens

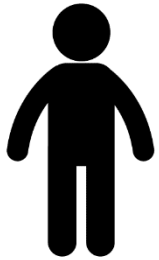
Someone
who needs
data /
analytics

KNIME expert



What often happens

Someone
who needs
data /
analytics



KNIME expert



What often happens

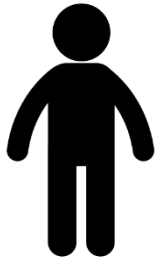
Someone
who needs
data /
analytics



KNIME expert

What often happens

Someone
who needs
data /
analytics

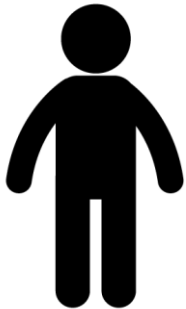


KNIME expert



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

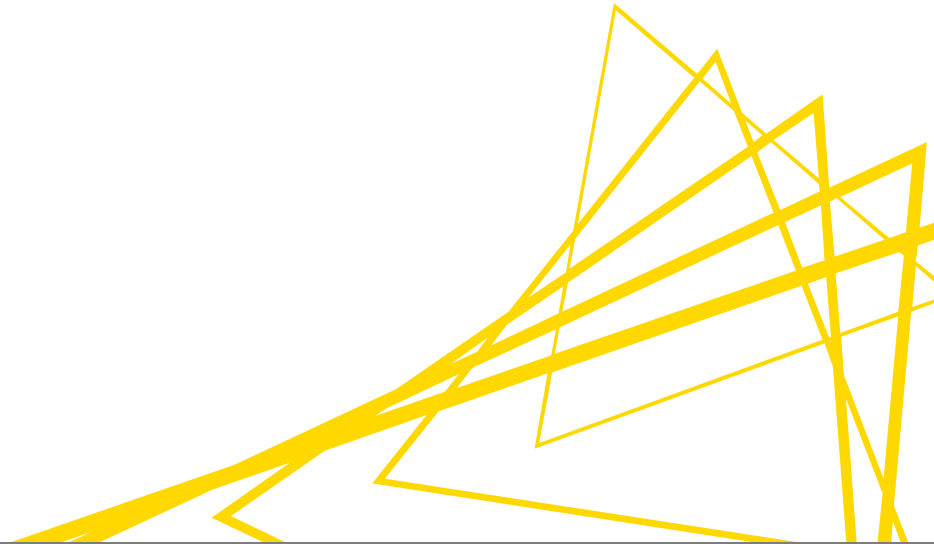


- Enable people who are not KNIME experts to work with/explore/analyze/learn from data

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



The KNIME® trademark and logo and OPEN FOR INNOVATION® trademark are used by KNIME AG under license from KNIME GmbH, and are registered in the United States. KNIME® is also registered in Germany.