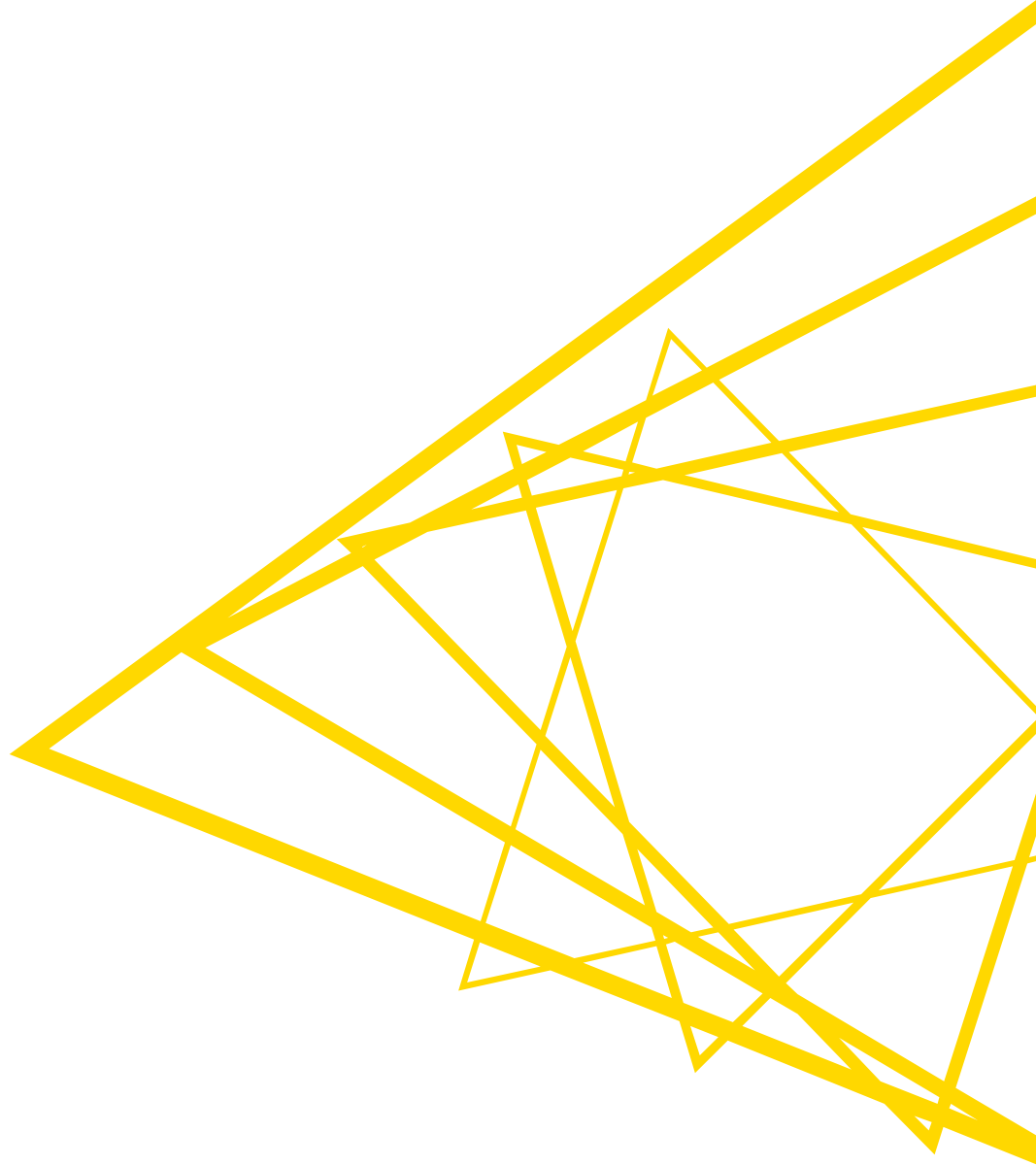




What's New

Bernd Wiswedel
KNIME



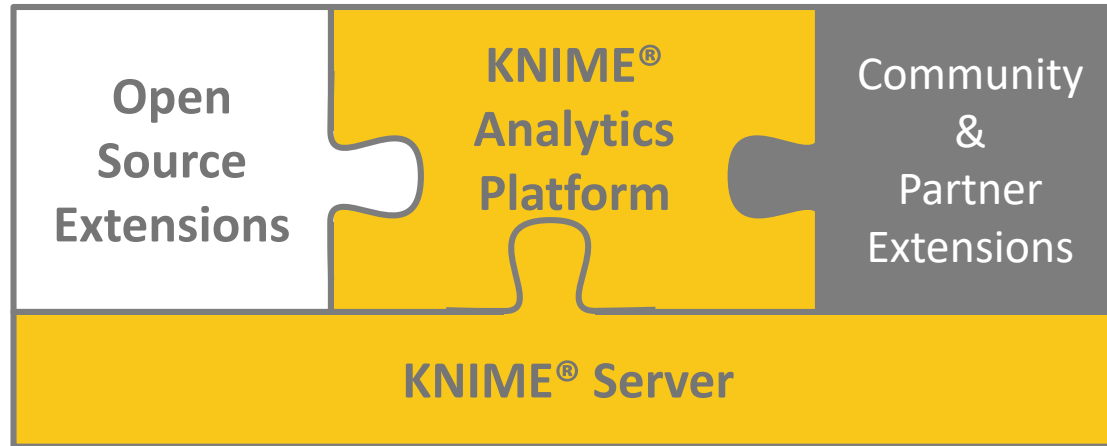
Outline

- “What’s new” presented in two use cases, presented by the team
- Questions/Discussions/Concerns: Find us!
- Demo booths in the registration area

What's new - Agenda

- Too much to show it all.
 - Some highlights:
 - Audio Processing
 - Text Processing
 - Analytics & Scripting
 - KNIME Server
 - Cloud Offerings
 - Guided Analytics
- } Use Case 1: IMDB Reviews
- } Use Case 2: Census Data

KNIME Software Pieces



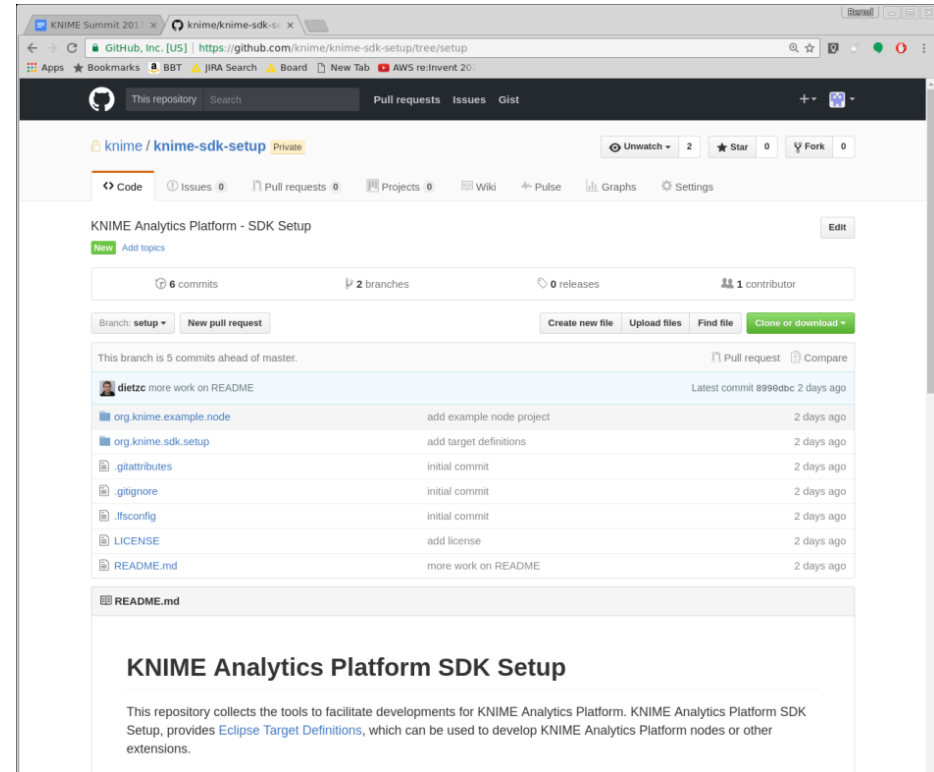
What's new - Agenda

- Too much to show it all.
 - Some highlights:
 - Audio Processing
 - Text Processing
 - Analytics & Scripting
 - KNIME Server
 - Cloud Offerings
 - Guided Analytics
- } Use Case 1: IMDB Reviews
- } Use Case 2: Census Data

KNIME as an Open (Source) Platform

KNIME Analytics Platform – Source Code on GitHub

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



<https://github.com/knime/>

SVN To Git – Fun Facts

- Switched from SVN to Git earlier the year
- ~3 weeks conversion time
(parallelized over 10 physical machines)
- 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)

KNIME Analytics Platform – Nightly Builds

- Nightly Builds publicly available:



Really, really, *really* important disclaimer

This is most definitely not production quality code. These nightly builds are based on recent developments, so they are not tested as thoroughly as stable releases. The functionality may change substantially (or disappear entirely) between builds.

Example Implementations
Noding Guidelines
Nightly Build

/ White Papers

/ KNIME Press

/ KNIME TV

Terms and Conditions *

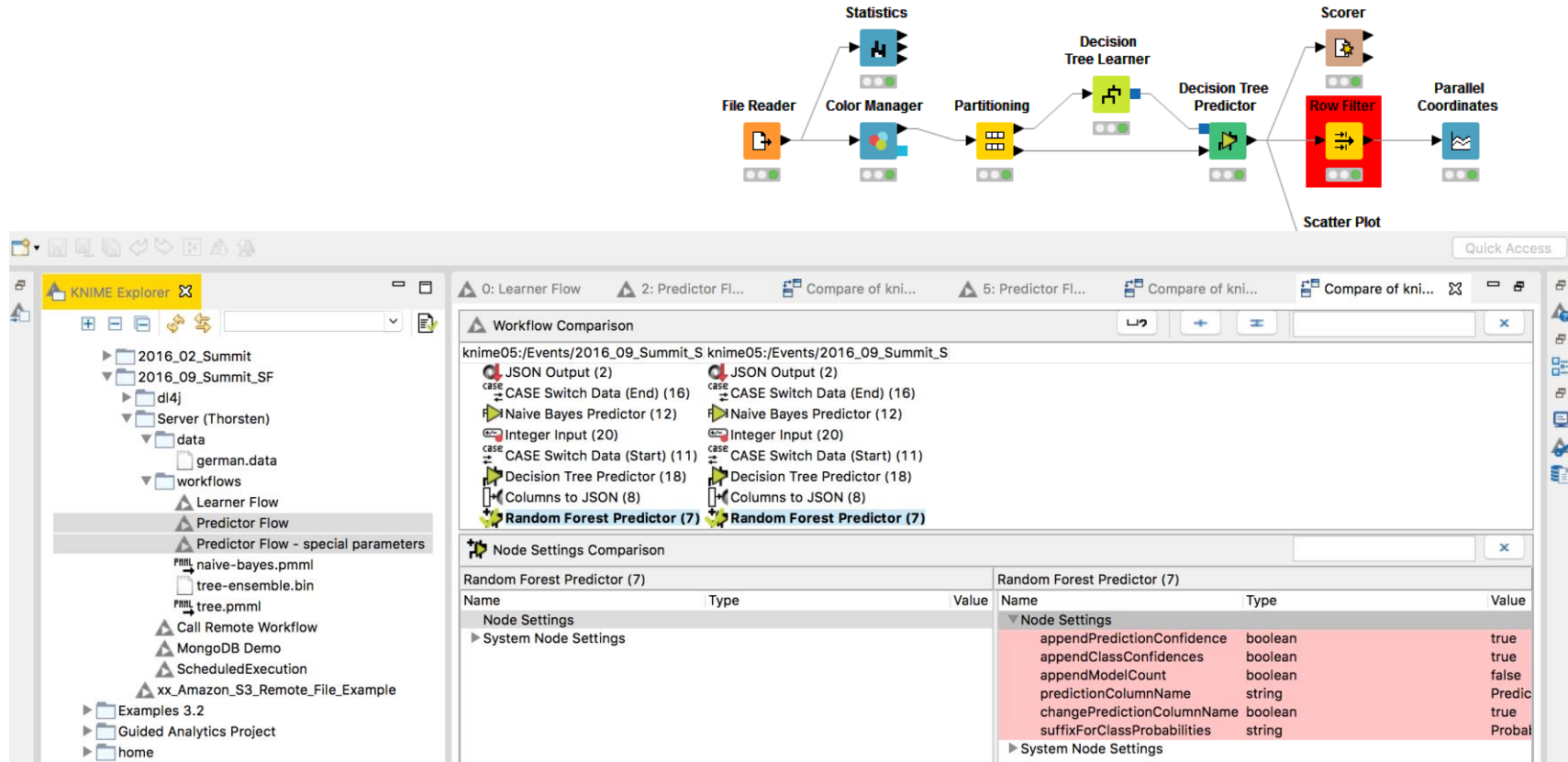
☐ I read and understand the disclaimer and accept the [Terms and Conditions](#).

[Go to downloads](#)

KNIME Personal Productivity – free & open-source

- Part of the KNIME Analytics Platform 3.4
- What is it?
 - Workflow Diff Viewer
 - Metanode Templates
 - Workflow Linking
 (“Call Local Workflow”)
 - “Local” Workflow Coach

KNIME Personal Productivity – Workflow Diff



The image displays the KNIME software interface, showing a workflow diagram at the top and a comparison window below it.

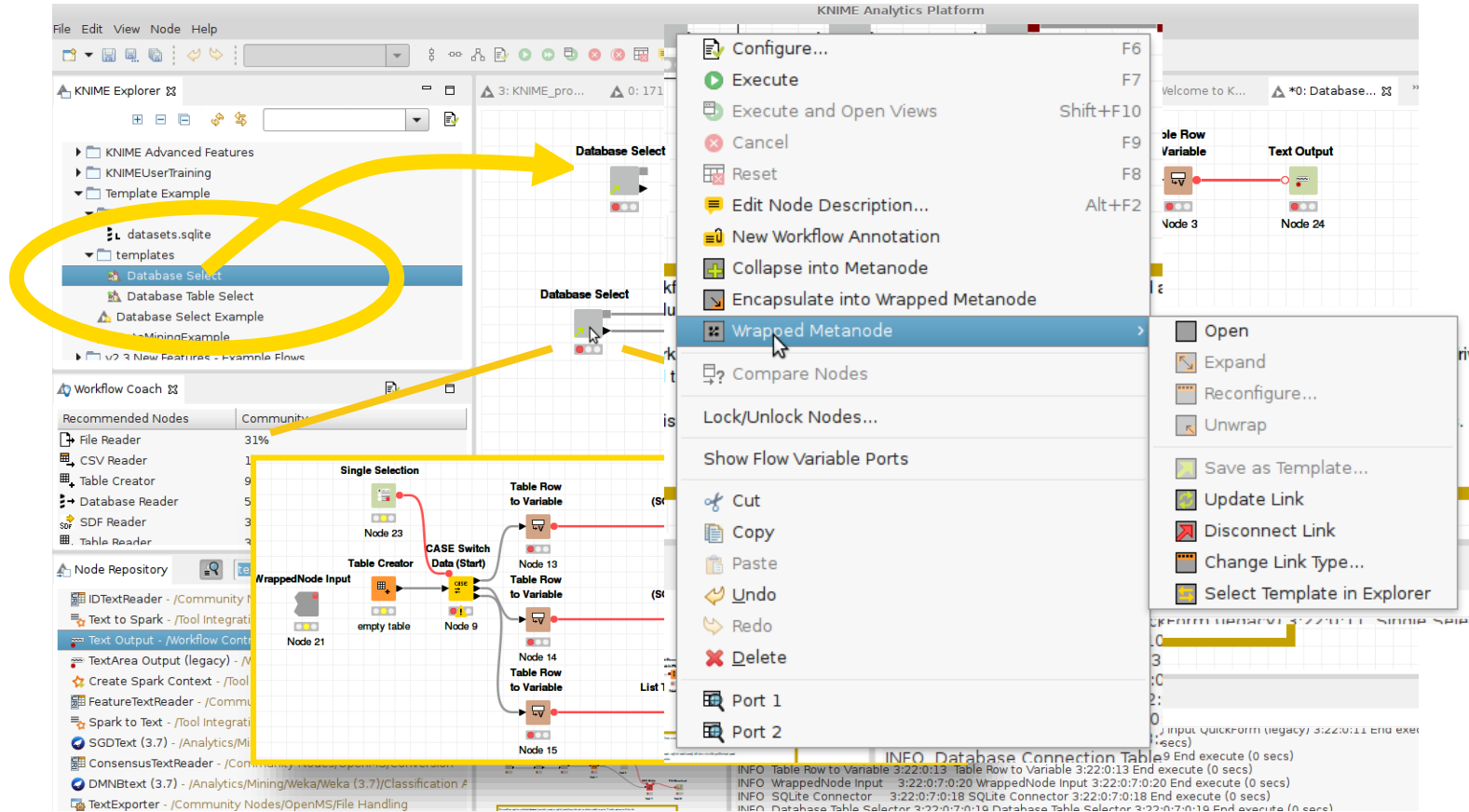
Workflow Diagram: The workflow starts with a **File Reader** node, which feeds into a **Color Manager** node. The **Color Manager** node outputs to a **Statistics** node and a **Partitioning** node. The **Partitioning** node feeds into a **Decision Tree Learner** node, which then feeds into a **Decision Tree Predictor** node. The **Decision Tree Predictor** node feeds into a **Scorer** node, which then feeds into a **Row Filter** node. The **Row Filter** node feeds into a **Parallel Coordinates** node. A **Scatter Plot** node is also shown, connected to the **Row Filter** node.

Comparison Window: The window is titled "Workflow Comparison" and shows a comparison between two workflows. The left workflow is "knime05:/Events/2016_09_Summit_S" and the right workflow is "knime05:/Events/2016_09_Summit_S". The comparison shows differences in the "Random Forest Predictor (7)" node settings.

Node Settings Comparison: The comparison shows the following settings for the "Random Forest Predictor (7)" node:

Name	Type	Value
appendPredictionConfidence	boolean	true
appendClassConfidences	boolean	true
appendModelCount	boolean	false
predictionColumnName	string	Predic
changePredictionColumnName	boolean	true
suffixForClassProbabilities	string	Probal

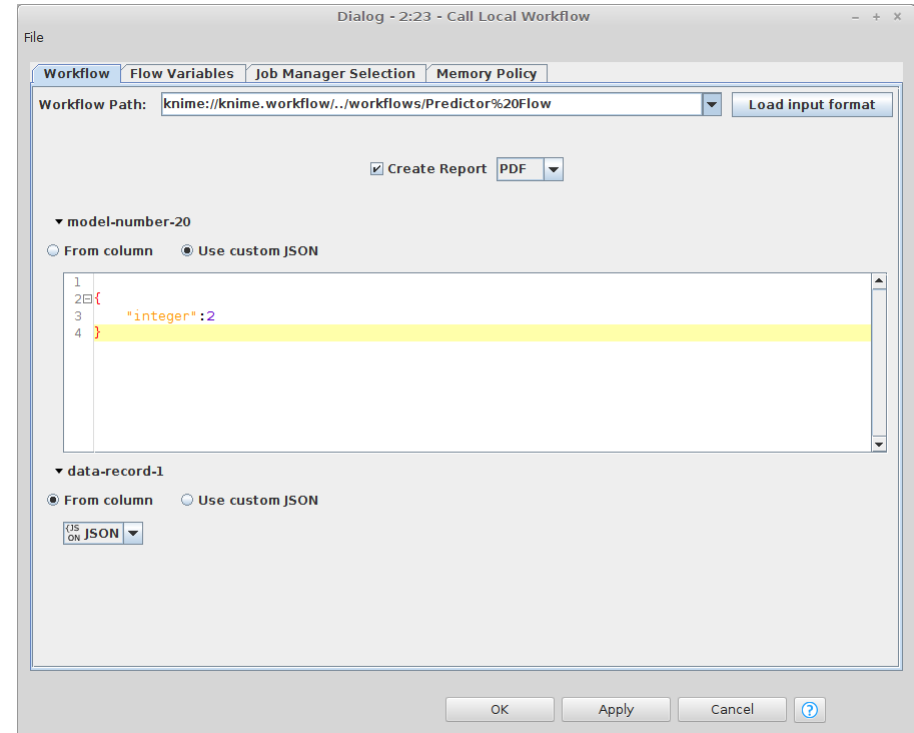
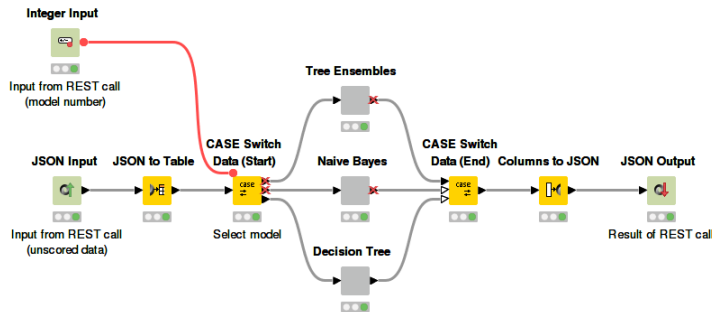
KNIME Personal Productivity – Metanode Templates



KNIME Personal Productivity – Workflow Linking

- Workflow Orchestration by calling out to other workflows

Call Local Workflow

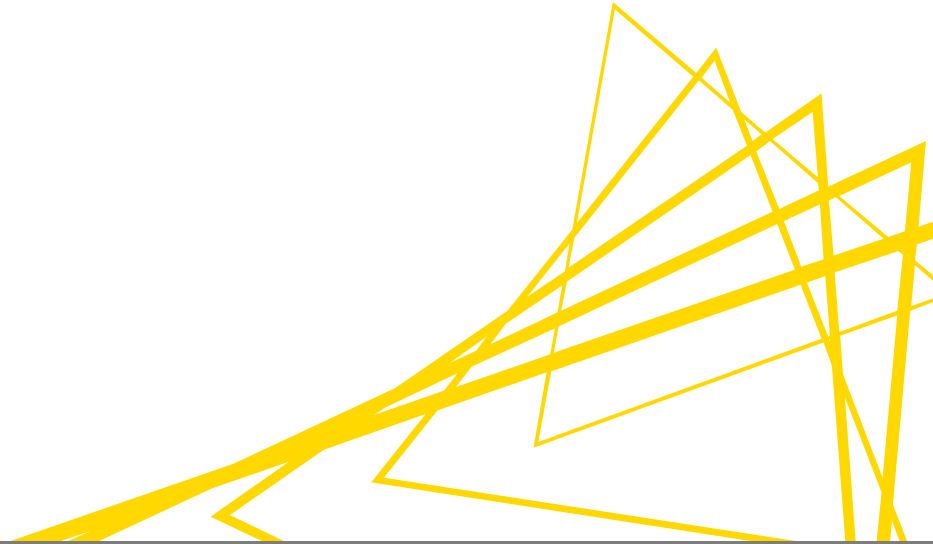


KNIME Personal Productivity – free & open-source

- Part of the KNIME Analytics Platform 3.4
- Big Data Extensions in December (3.5)

Speech Recognition

– Tobias Koetter –



KNIME Audio Processing

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

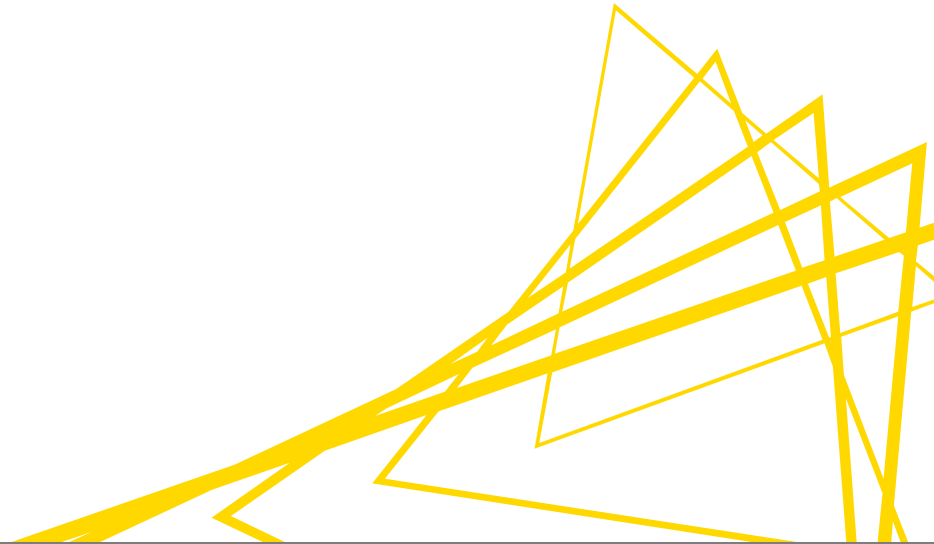


<https://pixabay.com/en/microphone-silver-metal-sound-waves-1074362/>

<https://www.youtube.com/watch?v=1L6zwLMvA-U>

Text Processing

– Kilian Thiel –



Example Use Case

- Sentiment classification of movie reviews
 - PDF parsing
 - Feature vector creation
 - Predictive modeling

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. ...”

A yellow callout box with a black border and a pointed top, containing the text "Positive or negative?".

Positive or negative?

Data Set

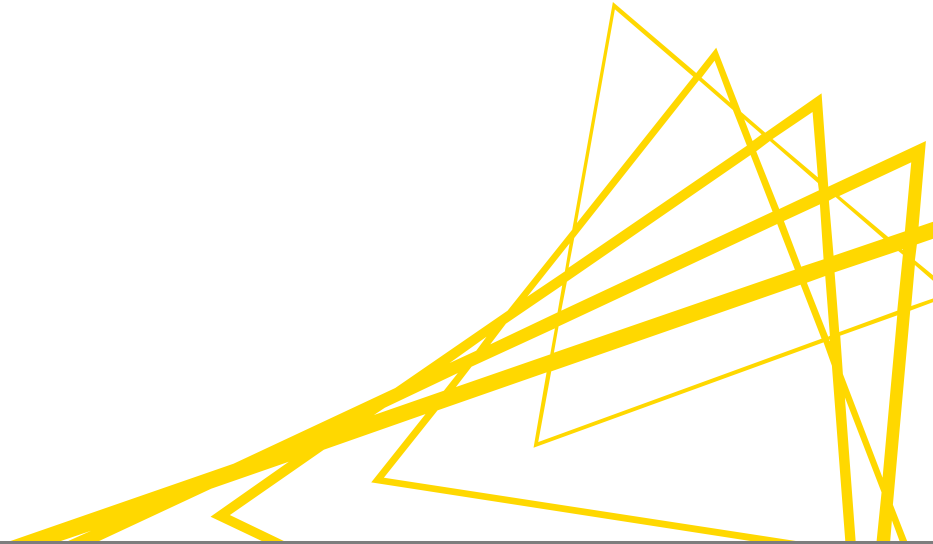
- The Large Movie Review Dataset v1.0
 - English movie reviews
 - Associated sentiment labels “positive” and “negative”
 - <http://ai.stanford.edu/~amaas/data/sentiment/>
- 2000 PDFs
 - 1000 positive reviews
 - 1000 negative reviews

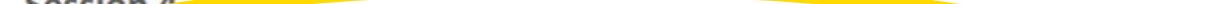
Textprocessing

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

Analytics & Scripting

– Bernd Wiswedel –



- L  new

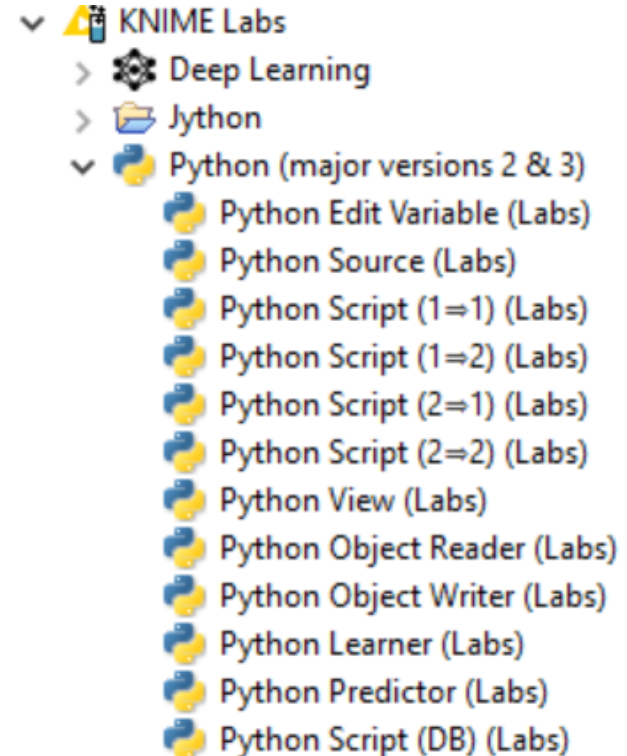
- 11:30 am KNIME Deep Dive: Deep Learning (and some Images)

11:30 am	KNIME Deep Dive: Deep Learning (and some Images) Christian Dietz (KNIME)
12:00 pm	Leveraging KNIME for Fire Analysis and Public Safety Mohammed Ayub (National Fire Protection Association)

12:30 pm	Guided Analytics in Action: Census and Sensibilities Greg Landrum (KNIME)
1:00 pm	Lunch Break - Picnic Ziler Buffet: Tejas Conference Dining

Scripting – Python (Labs)

- Supports Python 3
- Major Speedup
- Used for new DeepLearning Integration



Type Extensions in Java Snippet

- Java Snippet: Swiss Army knife for data manipulation
- Read/write support for CSV, JSON, XML, SVG, XMI

The screenshot displays the KNIME Java Snippet editor interface. On the left, a workflow diagram shows a 'File Reader' node connected to an 'RDKit From Molecules' node, which is then connected to a 'convert to RDKit molecule' node. The main area shows a code snippet for processing SMILES strings with RDKit. The code includes comments for variables and custom variables, followed by a Java code block that calculates the number of atoms and unspecified chiral centers in a molecule. The code is as follows:

```
// Enter your code here:  
long numAtoms = c_smilesRDKitMol.getNumAtoms();  
RDKitFuncts.assignStereochemistry(c_smilesRDKitMol, false, true, true);  
long nUnspecifiedChiralCenters=0, nSpecifiedChiralCenters=0;  
for (int i = 0; i < c_smilesRDKitMol.getNumAtoms(); ++i) {  
    if (c_smilesRDKitMol.getAtomWithIdx(i).getChiralTag() != Atom.ChiralType.CHI_UNSPECIFIED) {  
        ++nSpecifiedChiralCenters;  
    } else if (c_smilesRDKitMol.getAtomWithIdx(i).hasProp("_ChiralityPossible")) {  
        ++nUnspecifiedChiralCenters;  
    }  
}  
  
numAtoms = numAtoms;  
nSpecifiedChiralCenters = nSpecifiedChiralCenters;  
nUnspecifiedChiralCenters = nUnspecifiedChiralCenters;
```

Below the code, there is a note: 'used on the C++ backend. This is not strictly necessary. Usage down when working with large tables.'

On the right side of the editor, there is a 'Create Template...' button and a table with columns 'Java Field' and 'Type'. The table contains the following entries:

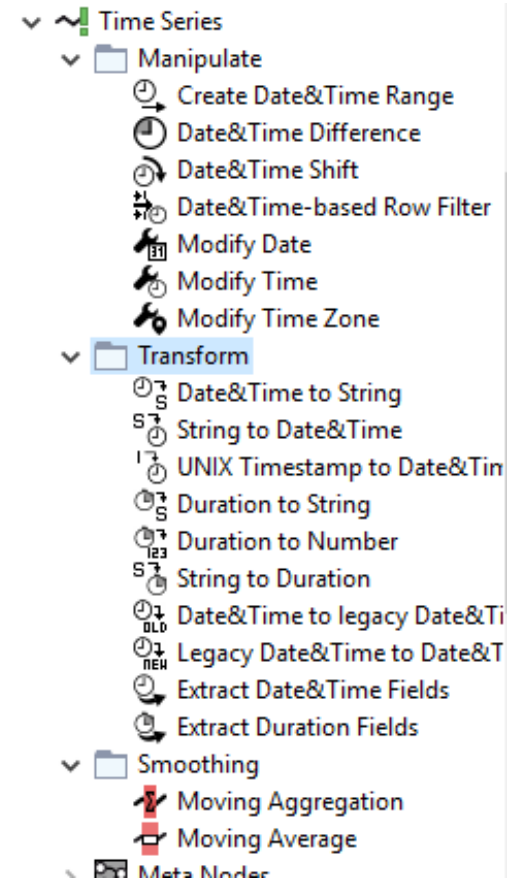
Java Field	Type
out_numAtoms	Long
out_numSpecifiedChiral...	Long
out_num_unspecifiedCh...	Long

Buttons 'Add' and 'Remove' are present next to the table.

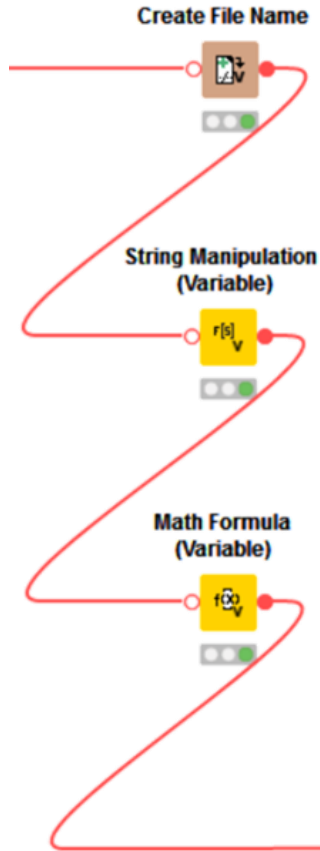
<https://www.knime.org/docs/using-custom-data-types-with-the-java-snippet-node>

Date & Time Handling Nodes revised

- Richer data types
- Timezone support
- Duration “math”



Three new helpers!!



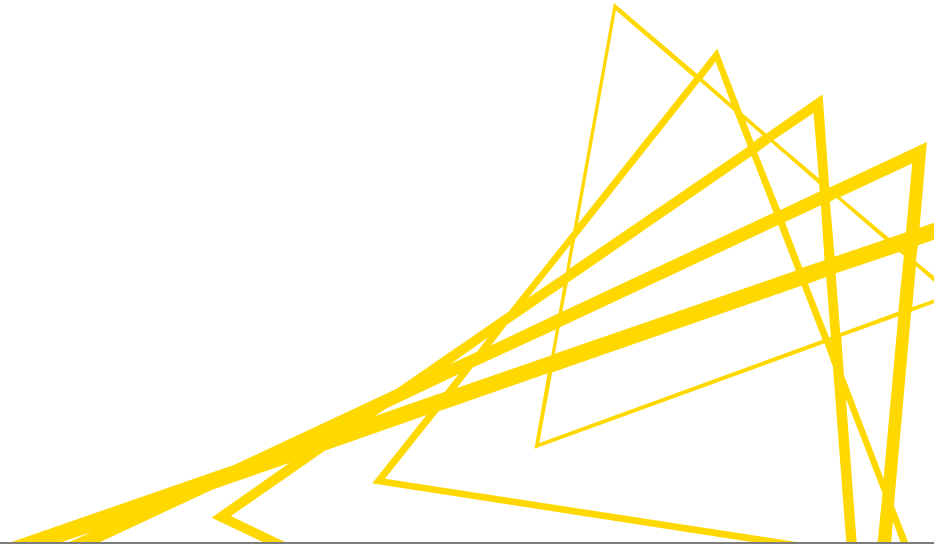
- Generate a file path based on folder, file name and extension type
- Simple changes to string based variables
- ... and for number variables as well.

... and now for something completely different

– Rosaria Silipo –



E-Learning



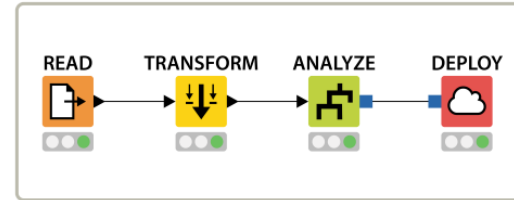
Free E-Learning Course: Web Page

- Hands-on e-learning course
- Data Access, ETL, **Analytics, Control Structures, Visualization**
- Around 50 small units
- ... with exercises
- ... and with solutions on the EXAMPLES server
- Final exercises to test your knowledge!

Check Section 1.4!

<https://www.knime.org/knime-introductory-course>

Introductory Course to Data Manipulation and ETL



KNIME® Analytics Platform is the leading open solution for data-driven innovation, helping you discover the potential hidden in your data, mine for fresh insights, or predict new futures. Our enterprise-grade, open source platform is fast to deploy, easy to scale, and intuitive to learn.

With more than 1000 modules, hundreds of ready-to-run examples, a comprehensive range of integrated tools, and the widest choice of advanced algorithms available, KNIME Analytics Platform is the perfect toolbox for any data scientist. Our steady course on unrestricted open source is your passport to a global community of data scientists, their expertise, and their active contributions.

KNIME Analytics Platform is easy to use, however if you are not yet very familiar with data manipulation and graphic programming, a few hints can help. That is why we have devised a series of short courses to get started with KNIME Analytics Platform and to show you how you can use it for data manipulation, ETL, data analytics, data blending, reporting – basically everything that falls into the category of data science.

In this particular course “Introductory Course to Data Access, Data Manipulation, and ETL”, we begin by exploring the basics of data access and manipulation. The goal is to build a small workflow, which takes a sales file including a number of contracts with different customers, extracts the active customers, measures their loyalty, ranks them by purchase value, and assigns them to the correct country sales manager.

Are you ready? Let's get started then!

- ▷ Chapter 1. KNIME Analytics Platform: an Overview
- ▷ Chapter 2. Data Access: Text and Binary Files, Web Services, Databases, Big Data, and more
- ▷ Chapter 3. ETL and Data Manipulation: Filtering, Aggregation, Join, and Concatenation
- ▷ Chapter 4. Exporting Data: to Text and Binary Files, to Databases, to Reports, and more
- Test your Knowledge. Assembling an ETL Workflow

[Chapter 1. KNIME Analytics Platform: an Overview >](#)

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.

KNIME Text Mining 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](#).



Copyright © 2017 KNIME AG

The KNIME® Analytics Platform

Copyright © 2017 KNIME AG

Visual KNIME Workflows

NODES perform tasks on data

Nodes are combined to create **WORKFLOWS**

Copyright © 2017 KNIME AG

Data Access

- Databases
 - MySQL, PostgreSQL
 - any JDBC (Oracle, DB2, MS SQL Server)

E-Books

At the **KNIME Press**:

<https://www.knime.com/knimepress>



Will They Blend? The Blog Post Collection edited by Rosaria Silipo

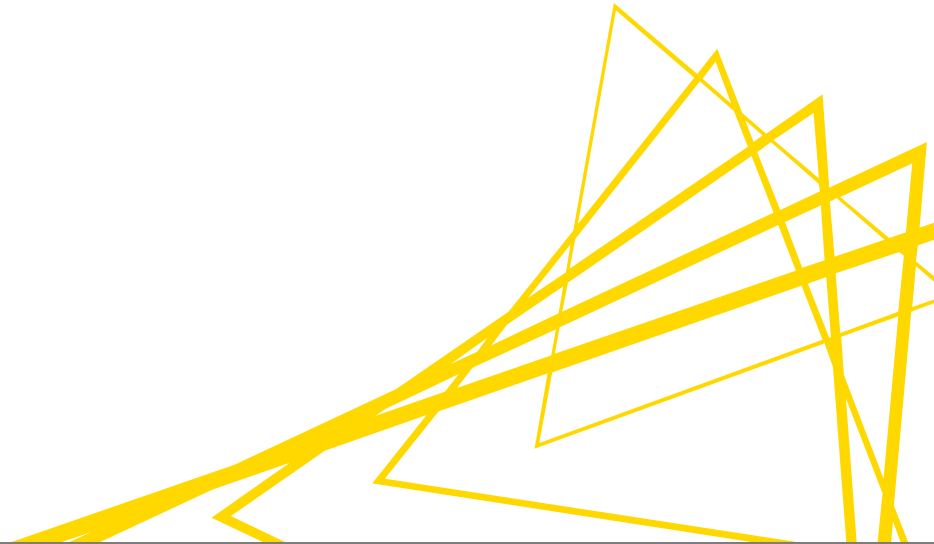


Beginner's Luck written by Rosaria Silipo



The KNIME Cookbook written by Rosaria Silipo and Michael P. Mazanetz

On-line Learning



On-line Courses

- On-Line Courses vs. On-Site Courses with Teacher
- First Experiment with Joe Porter's group in November/December
- Basic Course on KNIME Analytics Platform
- With Homework and Homework Correction

Send us an email: education@knime.com

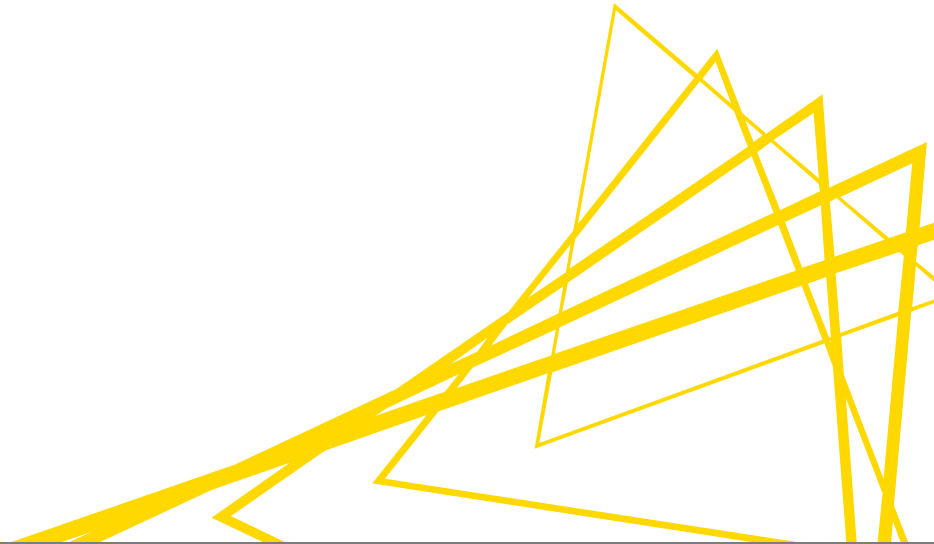
Webinars

In 2018 Series of Webinars:

- Data Science Cycle
- Introduction to KNIME Analytics Platform
- Data Access
- Database
- Logistic Regression
- Ensemble Models
- Text Processing
- ...

Keep an eye on our web site www.knime.com

Learning @Meetups



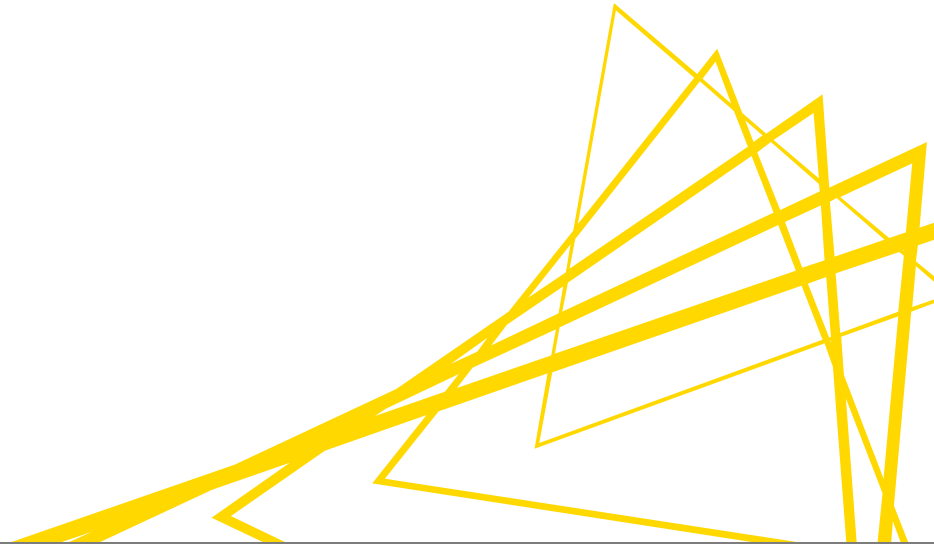
Learnathons

Data Science Cycle: ETL, Model Training, Deployment

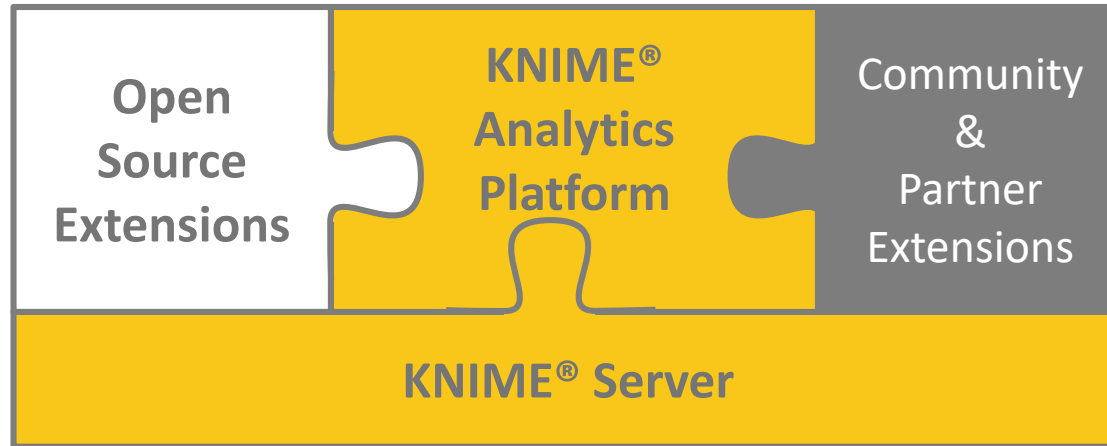


KNIME Server

– Jon Fuller–

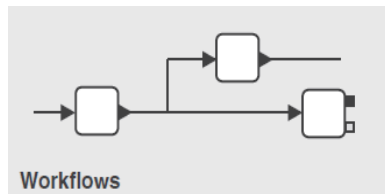
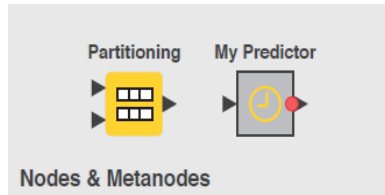
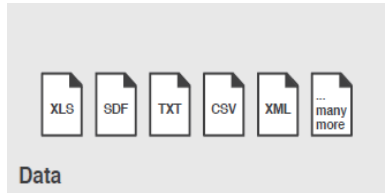


KNIME Software Pieces

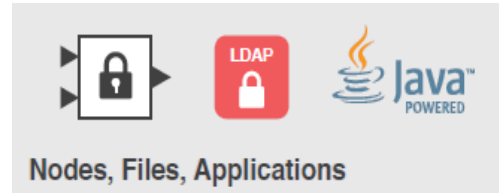


KNIME Server

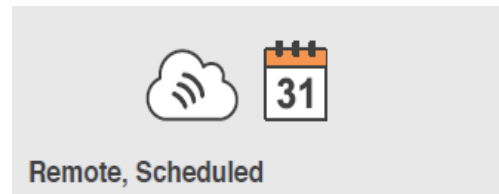
Shared Repositories



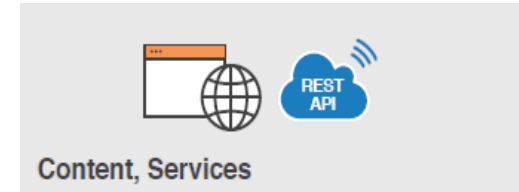
Access Management



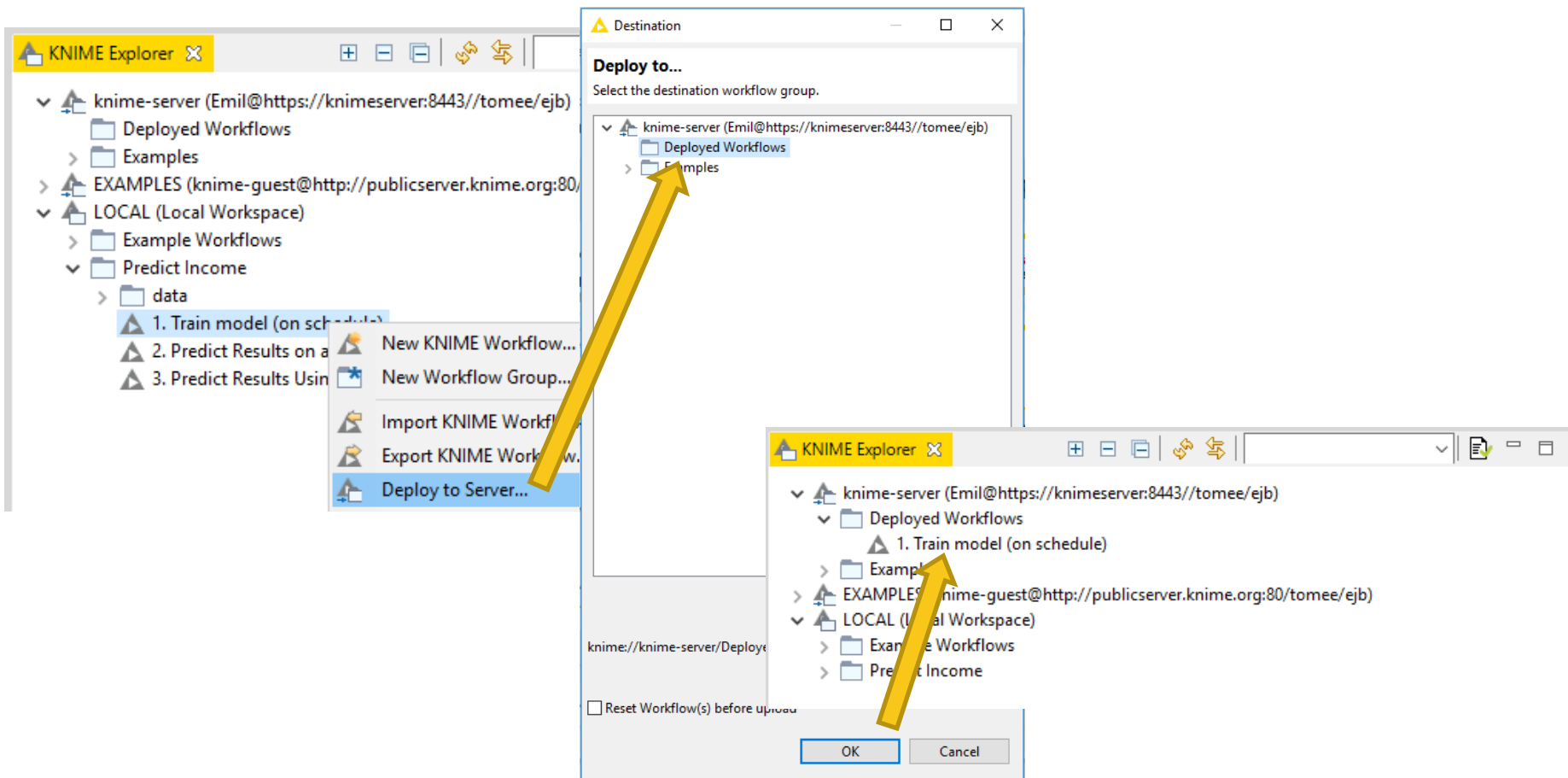
Flexible Execution



Web Enablement



Deploy to Server



Open in WebPortal

The image illustrates the process of opening a workflow in the KNIME WebPortal. On the left, the KNIME Explorer window shows a tree view of workflows. The 'WebPortal' folder is expanded, and the context menu for the workflow '01 - WebPortal Data Mining (Basic)' is open. The 'Open in WebPortal' option is highlighted. A yellow arrow points from this option to the WebPortal web interface on the right. The WebPortal interface shows the workflow details, including a description and a 'Start' button.

KNIME Explorer Tree View:

- knime-server (Emil@https://knimeserver:8443//tomee/ejb)
 - Deployed Workflows
 - 1. Train model (on schedule)
 - Examples
 - Orchestration
 - Reporting
 - REST
 - Shared Metanodes
 - Supplementary Data (really only the data)
 - Test Workflows (add your own for databases)
 - WebPortal
 - 01 - WebPortal Data Mining (Basic)
 - 02 - WebPortal Customer Segmen
 - 03 - Webportal IF statements
- EXAMPLES (knime-guest@http://publicserve)
- LOCAL (Local Workspace)
 - Example Workflows
 - Predict Income
 - 1. Train model (on schedule)
 - 2. Predict Results on a Schedule
 - 3. Predict Results Using REST API

Context Menu Options:

- New Workflow Group...
- Deploy to Server...
- Execute...
- Open in WebPortal
- Delete...
- Show history
- Create snapshot
- Rename...

WebPortal Interface:

Open for Innovation
KNIME WebPortal
Settings Logout

Workflow Repository

- Deployed Workflows
- Examples
 - Orchestration
 - Reporting
- Test Workflows (add your own)
- WebPortal
 - 01 - WebPortal Data Mining (Basic)
 - 02 - WebPortal Customer Segmen
 - 03 - Webportal IF staten

01 - WebPortal Data Mining (Basic)

This workflow shows how to upload a file and train a simple decision tree. The example file 'CurrentDetailData.table' can be downloaded from the Supplementary Data (really only the data)/Basics directory, using the KNIME Analytics Platform.

☐ Mail notification on completion

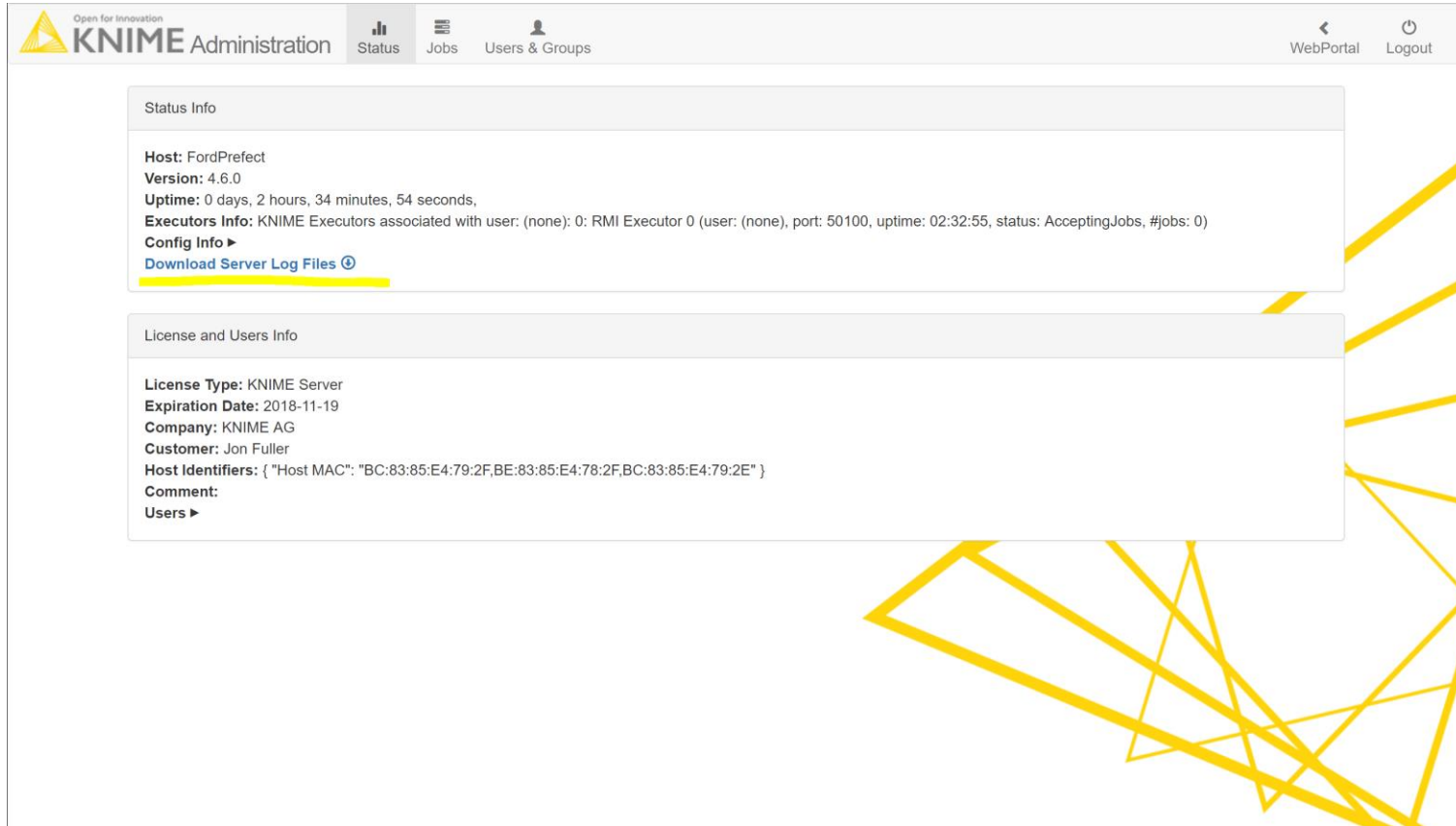
Start ▶

© 2017 KNIME.com AG - Version 4.6.0

Authentication

- Authentication with LDAP/Active Directory.
- **Single Sign On (SSO) with Kerberos**

KNIME Server – Admin made easy



The screenshot displays the KNIME Administration web interface. The top navigation bar includes the KNIME logo with the tagline "Open for Innovation", the title "KNIME Administration", and three main menu items: "Status" (with a bar chart icon), "Jobs" (with a list icon), and "Users & Groups" (with a person icon). On the right side of the navigation bar are links for "WebPortal" and "Logout".

The main content area is divided into two sections:

- Status Info**: This section provides system details:
 - Host:** FordPrefect
 - Version:** 4.6.0
 - Uptime:** 0 days, 2 hours, 34 minutes, 54 seconds,
 - Executors Info:** KNIME Executors associated with user: (none): 0: RMI Executor 0 (user: (none), port: 50100, uptime: 02:32:55, status: AcceptingJobs, #jobs: 0)
 - Config Info** ▶
 - [Download Server Log Files](#) ⓘ
- License and Users Info**: This section displays license details:
 - License Type:** KNIME Server
 - Expiration Date:** 2018-11-19
 - Company:** KNIME AG
 - Customer:** Jon Fuller
 - Host Identifiers:** { "Host MAC": "BC:83:85:E4:79:2F,BC:83:85:E4:78:2F,BC:83:85:E4:79:2E" }
 - Comment:**
 - Users** ▶

Decorative yellow geometric lines are overlaid on the right side of the interface.

KNIME in the Cloud

— Jon Fuller —



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)

The screenshot shows the AWS Marketplace interface. At the top, there's a search bar with 'AMI & SaaS' selected. Below the header, the KNIME logo is prominently displayed with the tagline 'Open for Innovation'. The 'About KNIME' section describes the platform as a leading open solution for data-driven innovation. The 'KNIME Products (3)' section lists three offerings: 'KNIME Server + Big Data Extensions', 'KNIME Server (BYOL)', and 'KNIME Server'. Each product listing includes a star rating, version information, and pricing details. For example, 'KNIME Server + Big Data Extensions' is priced at \$6.805/hr or \$49,680.00/yr (17% savings) for software + AWS usage fees.

The screenshot shows the Azure Marketplace interface. At the top, there's a search bar with 'Search Marketplace' text. Below the header, the KNIME logo is prominently displayed with the tagline 'Open for Innovation'. The 'KNIME® Cloud Analytics Platform' is the main heading. The 'Overview' section describes the platform as a powerful open analytics solution with on-demand infrastructure. It highlights that KNIME Cloud Analytics Platform brings the power and flexibility of the open source KNIME Analytics Platform to the cloud. The 'Learn more' section provides links to 'Support' and 'KNIME Cloud Analytics Platform'.

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 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](https://www.knime.com).

KNIME

[Visit the KNIME Website](#)

KNIME Products (3)

showing 1 - 3



KNIME Server + Big Data Extensions

★★★★★ (0) | Version 4.4.0 with Big Data Extensions | Sold by [KNIME.com](#)

Starting from **\$6.805/hr** or from **\$49,680.00/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)

★★★★★ (0) | 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

★★★★★ (0) | Version 4.4.0 | Sold by [KNIME.com](#)

Starting from **\$1.726/hr** or from **\$12,600.00/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

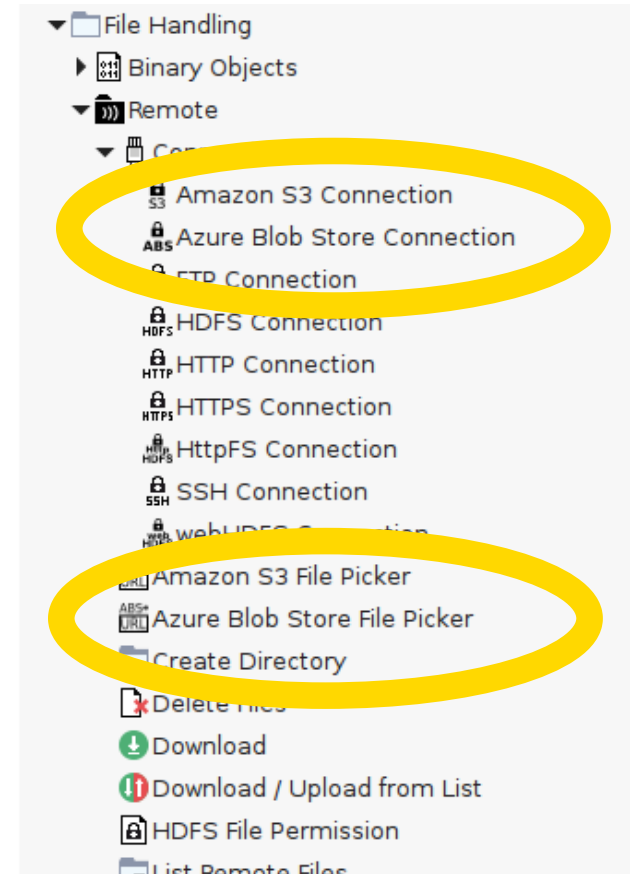


- SQL Server

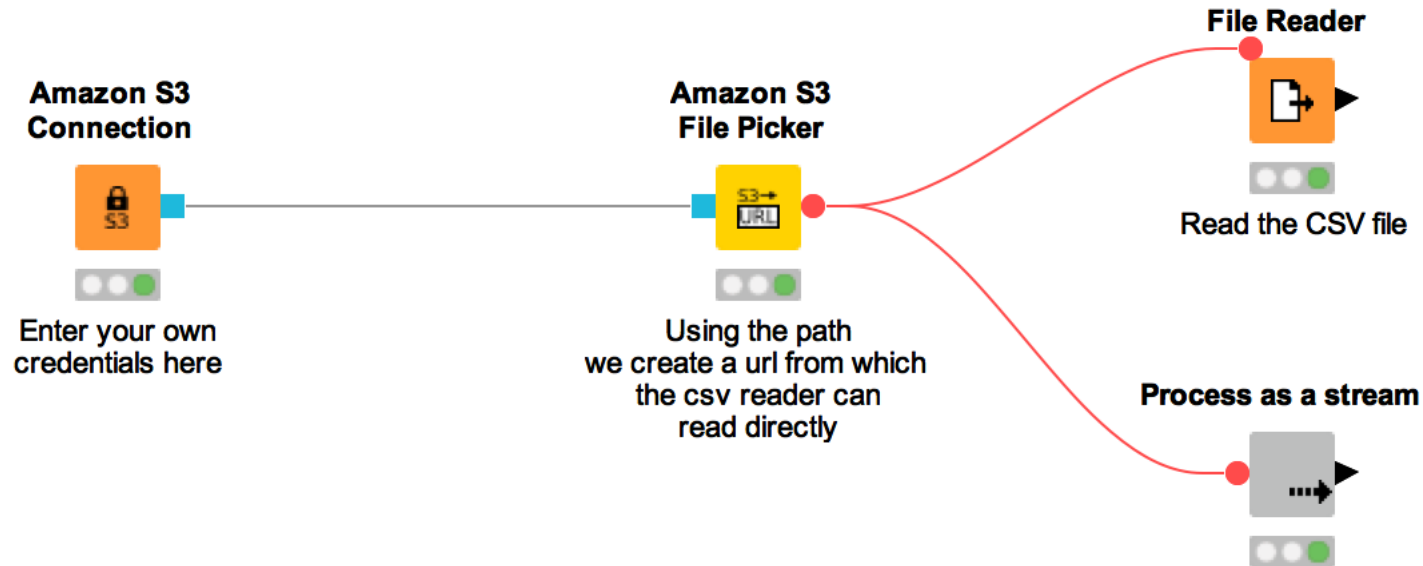


KNIME Cloud Connectors

- 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



Azure SQL Server Connector

+ Add Columns Refresh


Subscriptions: All 2 selected

Filter by name...

All subscriptions

▼

1 items

NAME ▼	STATUS	REPLICATION ROLE	SERVER	PRICING TIER	LOCATION ▼	SUBSCRIPTION ▼
 demodb	Online	None	knime	Basic	West Europe	Microsoft Azure Spons... ⋮

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

More connectors: Amazon Redshift

- Fully managed data warehouse
 - SQL interface
 - Distributed and parallelized queries
-
- Amazon Redshift connector integrates seamlessly with existing KNIME Database connectors

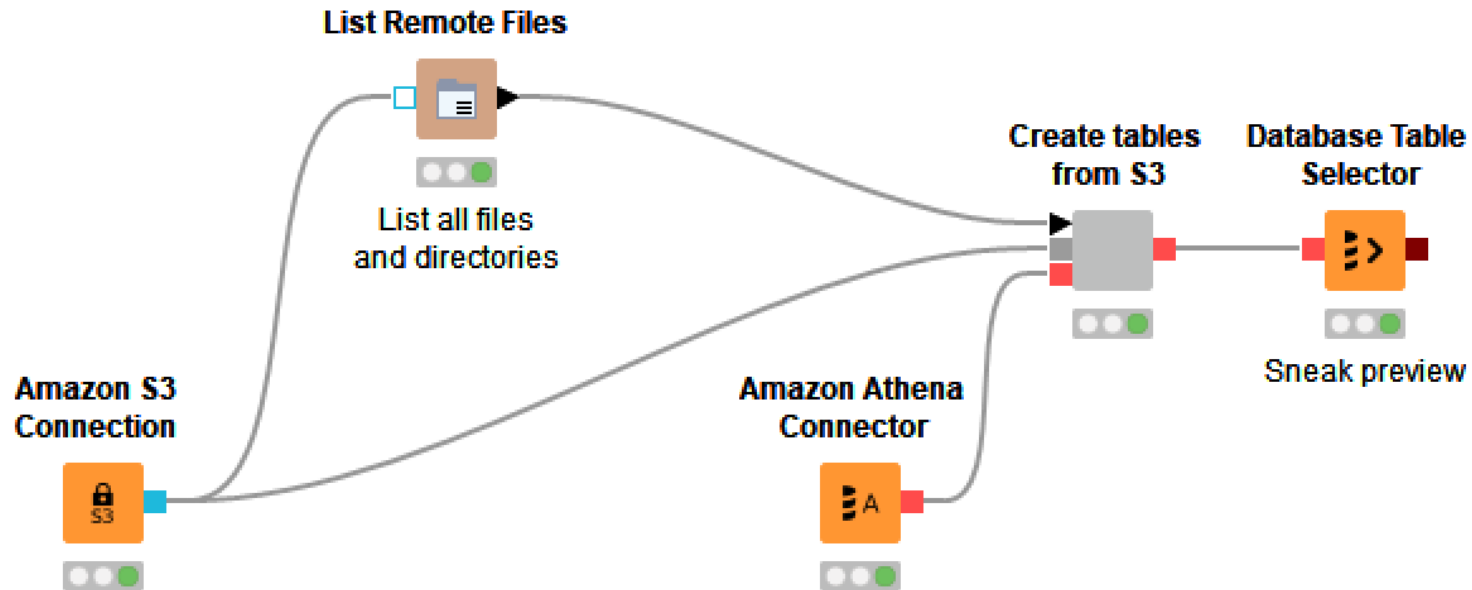


Amazon Athena

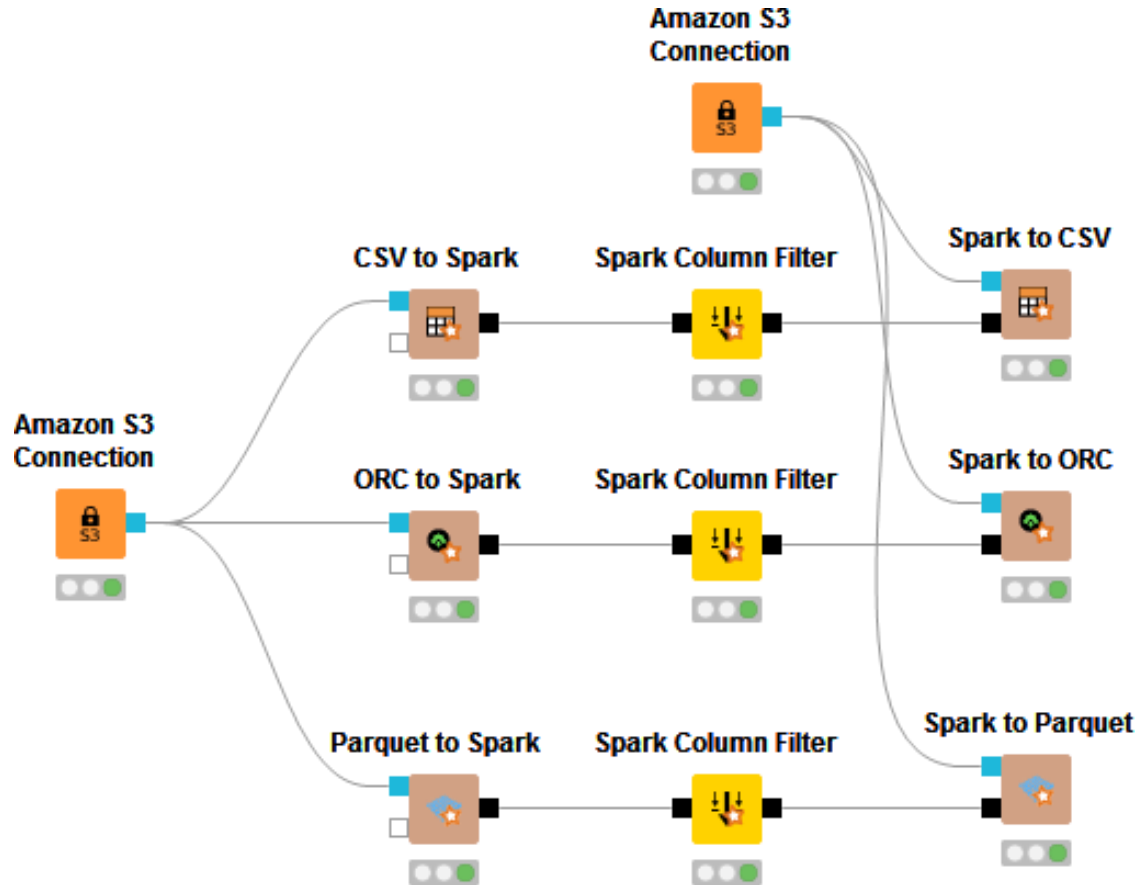
- Serverless, interactive query service
 - SQL interface
-
- Amazon Athena integrates seamlessly with existing KNIME Database connectors



Amazon Athena

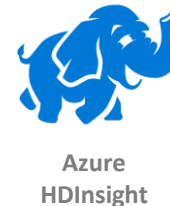
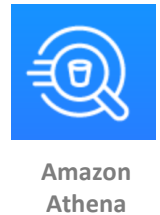
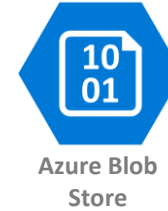


S3/Blob Storage support for Spark Data Source Nodes



Cloud Resources

- Data storage
 - MS Azure Blob Storage, Amazon S3
- Databases
 - Azure SQL Server, Amazon RDS (MySQL, PostgreSQL etc)
- Data warehouse
 - Amazon Redshift, Amazon Athena, Azure SQL-DW
- Hadoop/Spark
 - Azure HDInsight, Amazon EMR

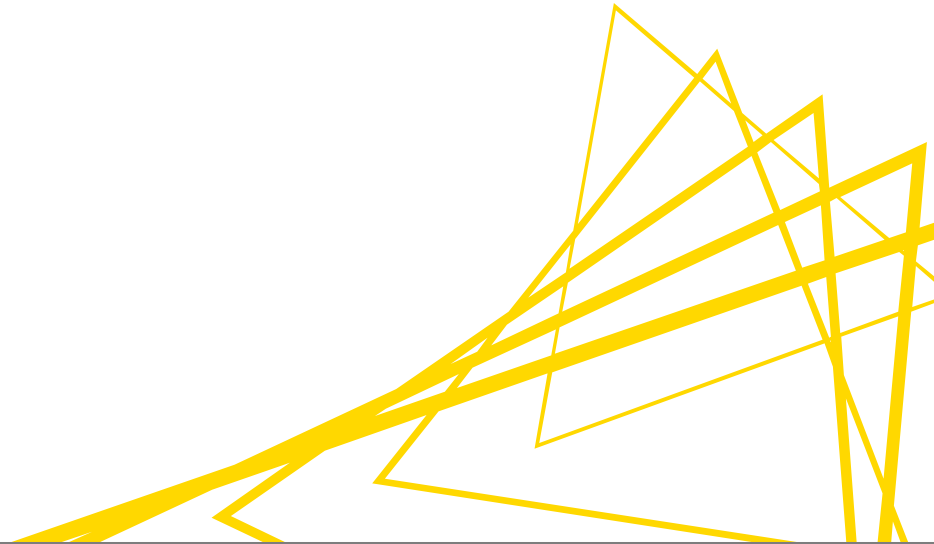


KNIME in the Cloud: Summary

- KNIME Analytics Platform available on Azure Marketplace
- KNIME Server available on Azure and AWS marketplaces
- Cloud connectors for S3, Blob Store, Azure SQL server

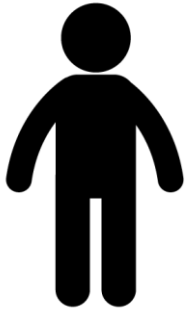
Guided Analytics

– Greg Landrum –



Reminder of what we're talking about here

Someone
who needs
data /
analytics



KNIME expert



What often happens

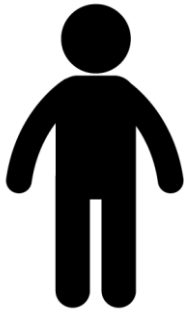
Someone
who needs
data /
analytics

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



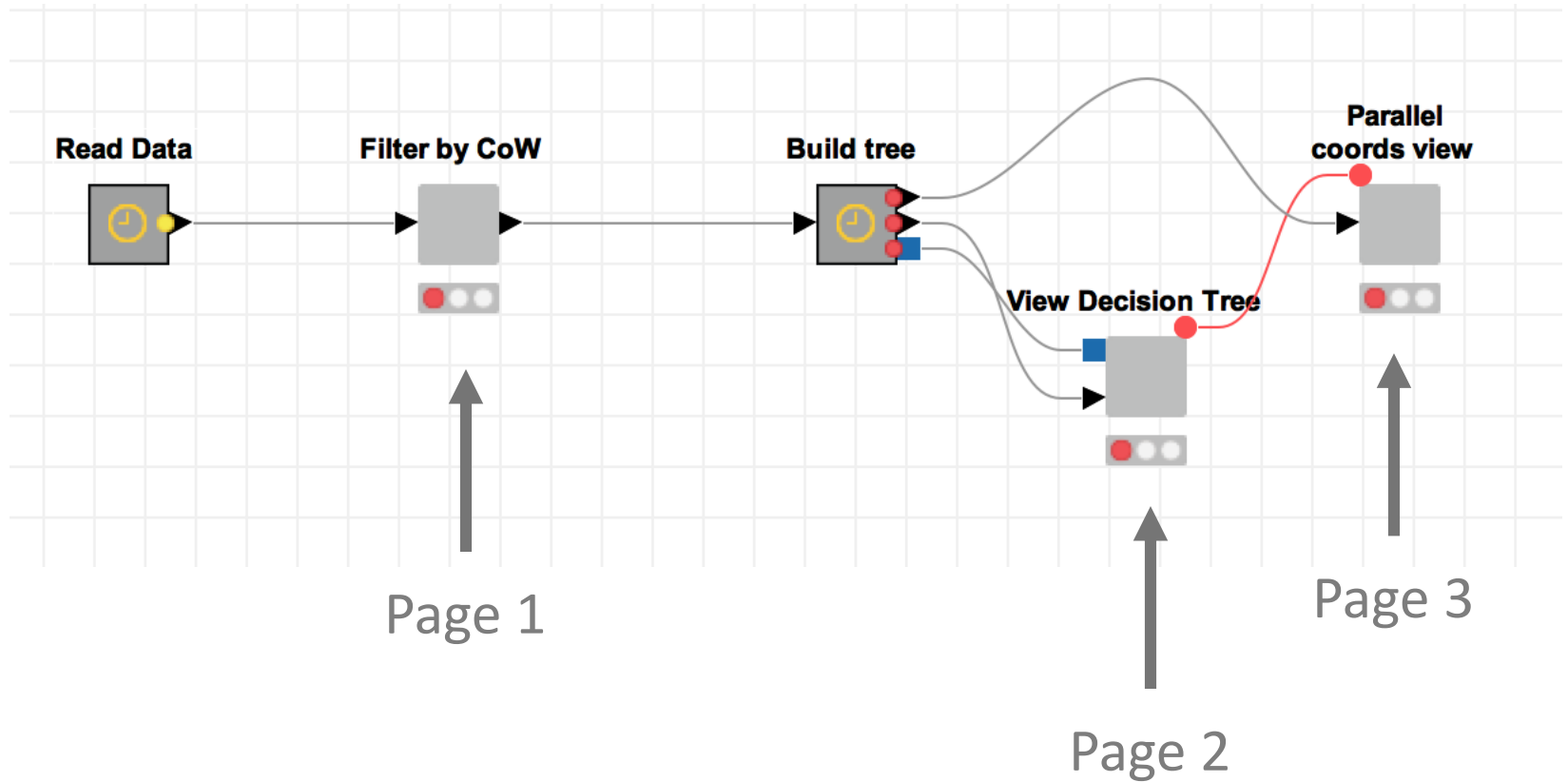
- Enable people who are not KNIME experts to work with/explore/analyze/learn from data
- Typically deployed via KNIME WebPortal
- Heavy use of interactive JavaScript views for the user experience
- Bonus for everyone: the JavaScript views that we're building for Guided Analytics are also really useful in "normal" KNIME workflows.

So what's new?

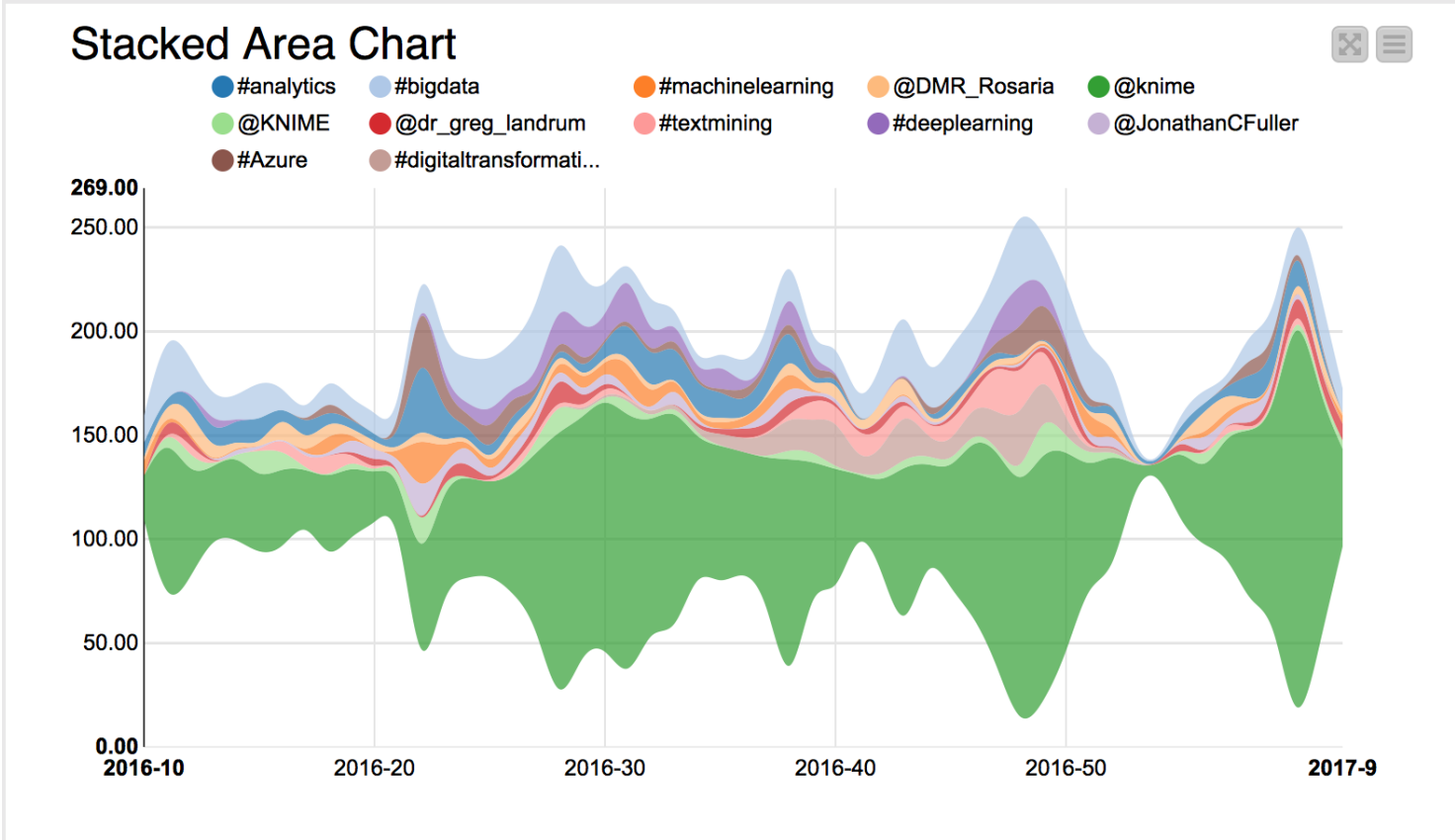
- JavaScript Views:
 - Parallel coordinates plot
 - Table view
 - Decision tree view
 - Network view
 - Stacked area chart
 - Sunburst chart
- Interactivity: Selection and filtering
 - Range filter widgets
 - Linked views
 - Accessing views in wrapped metanodes, composite view
 - Layout editor

But let's start with a demo

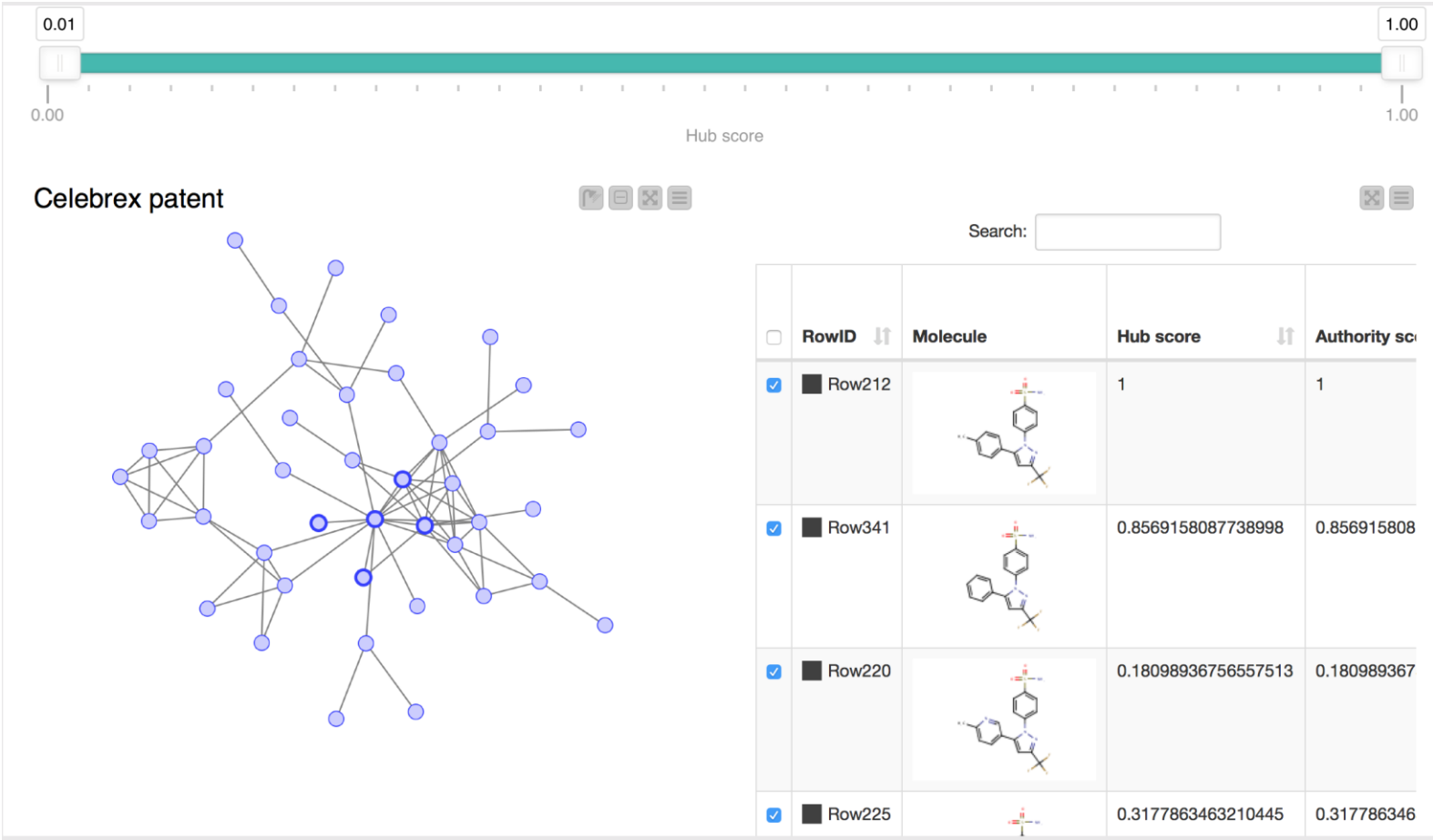
The workflow behind the demo:



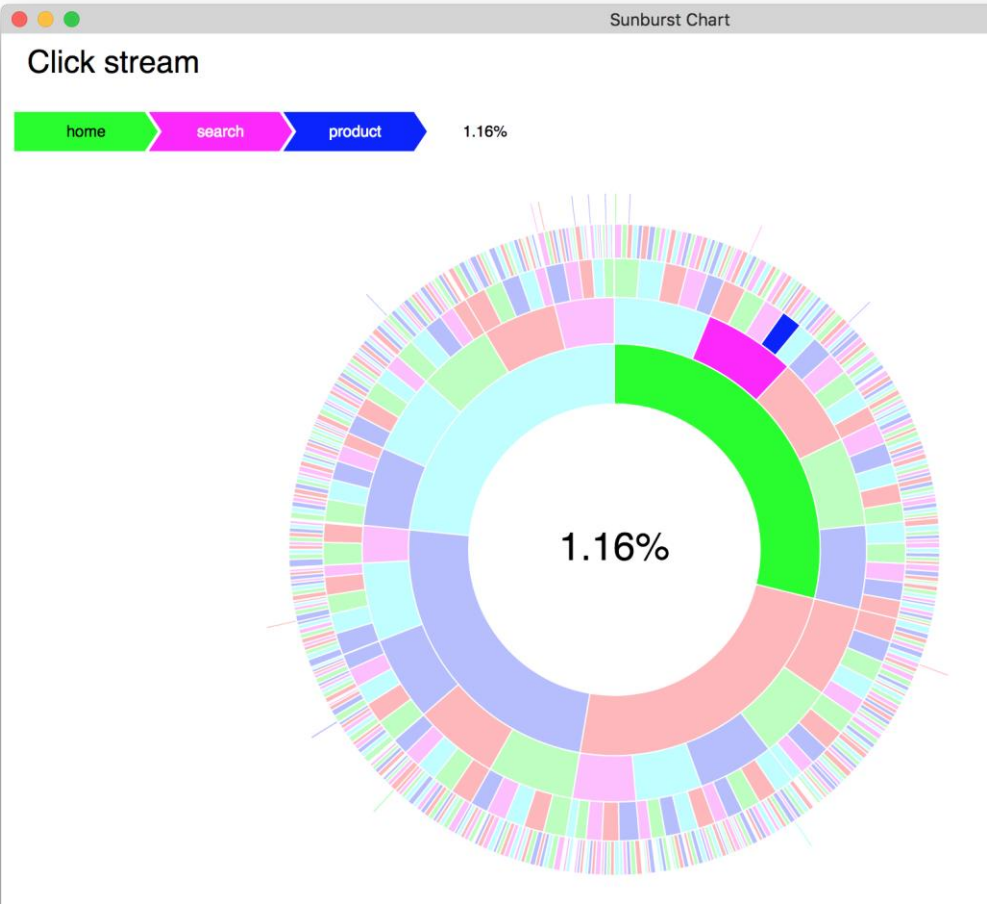
JavaScript Stacked Area Chart



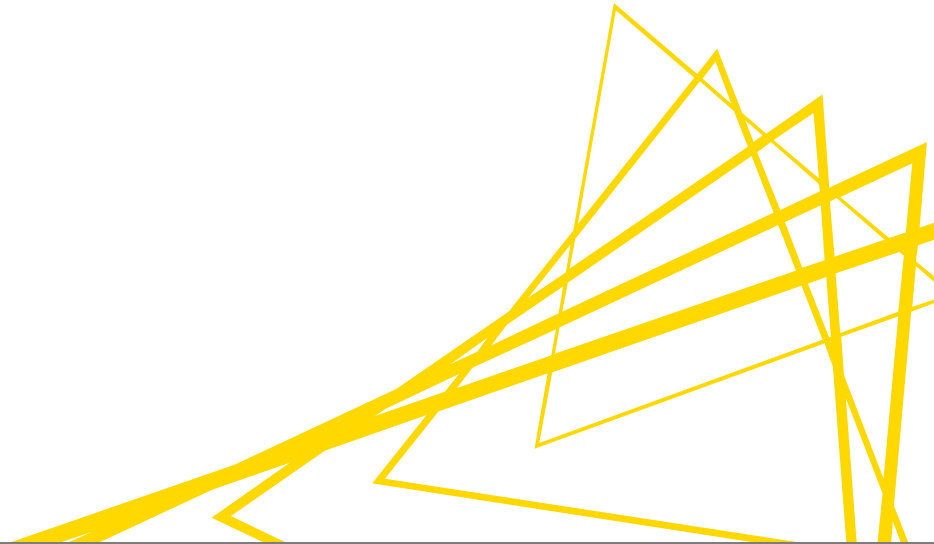
JavaScript Network Visualization



JavaScript Sunburst Chart



(Short) Coffee Break
- all of us -



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.