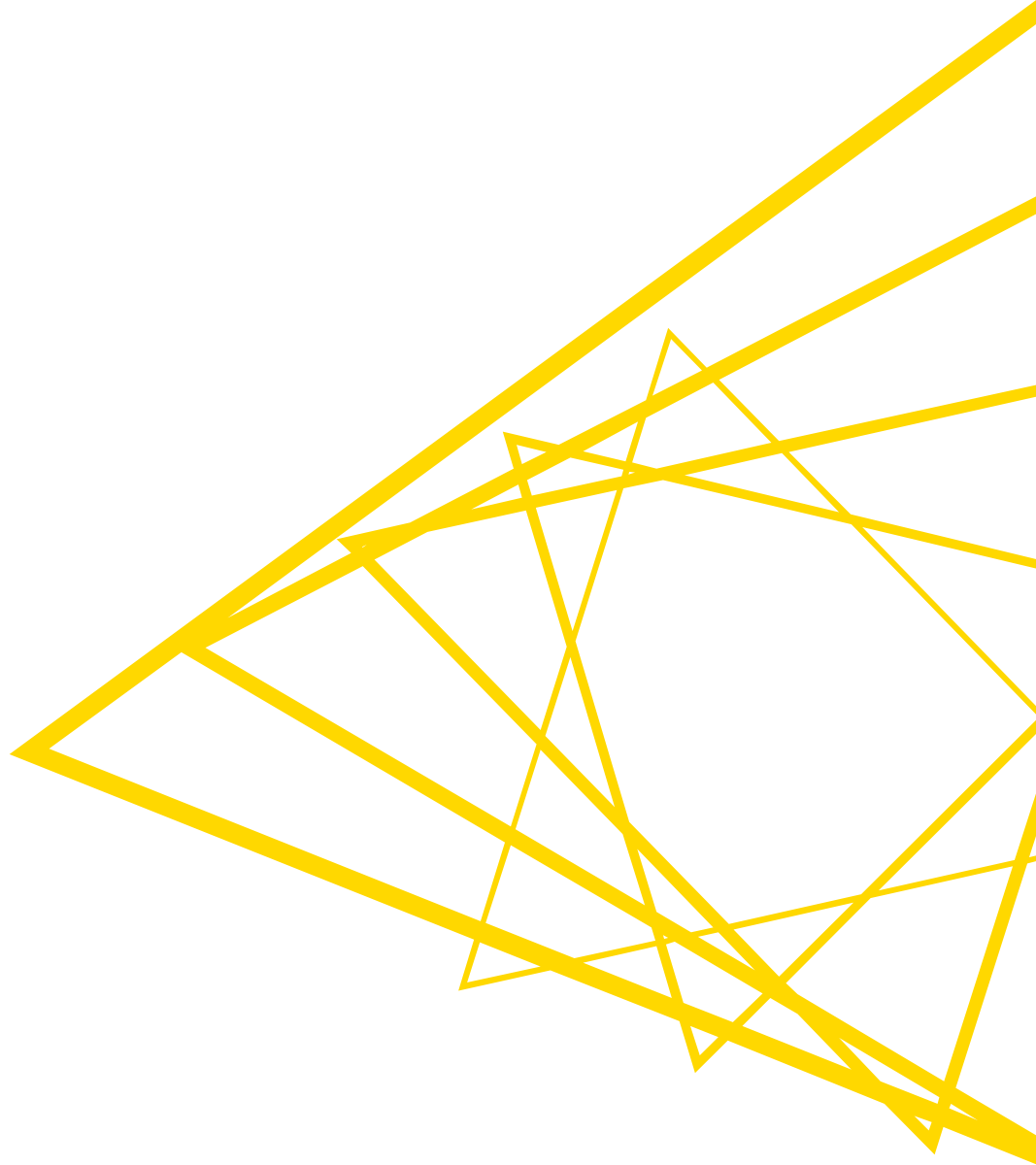




What's Cooking

Bernd Wiswedel

KNIME

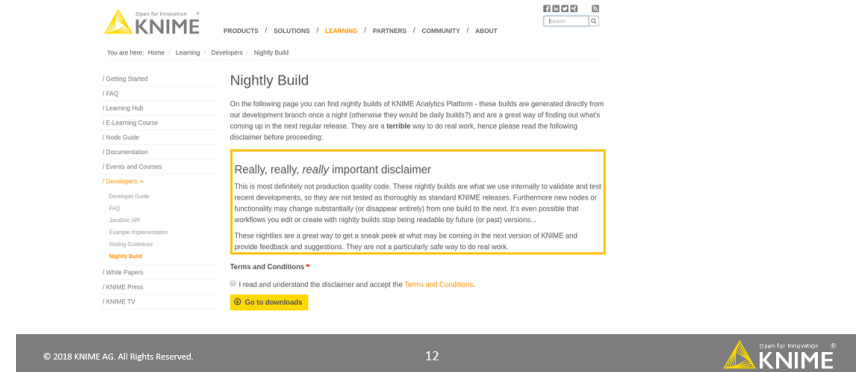


What's Cooking

- Enhancements to the software planned for the next feature release
- Actively worked on
- Available in Nightly build

KNIME Analytics Platform – Nightly Builds

- Nightly Builds publicly available:

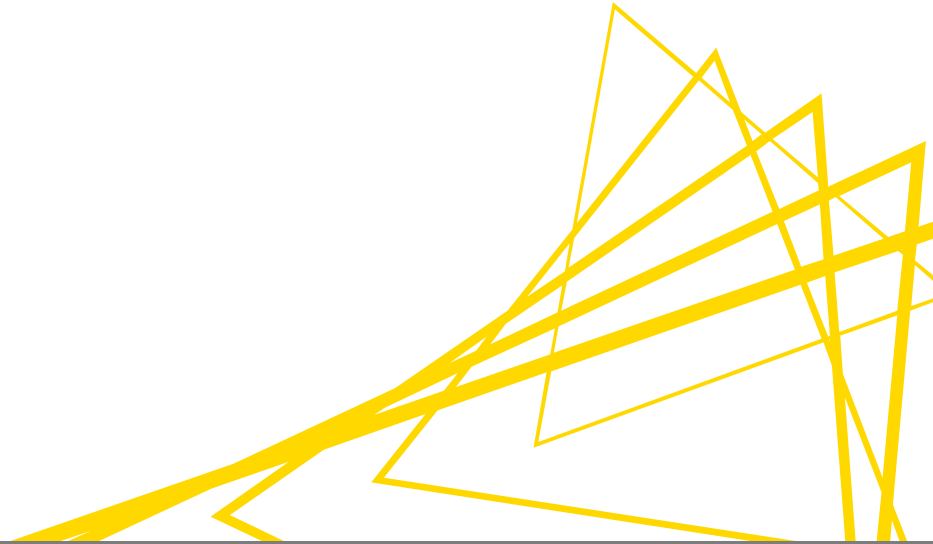


The screenshot shows the KNIME Analytics Platform website. The header includes the KNIME logo and navigation links: PRODUCTS / SOLUTIONS / LEARNING / PARTNERS / COMMUNITY / ABOUT. Below the header, there's a breadcrumb trail: You are here: Home / Learning / Developers / Nightly Build. The main content area is titled "Nightly Build" and contains a disclaimer: "Really, really, really important disclaimer. This is most definitely not production quality code. These nightly builds are what we use internally to validate and test recent developments, so they are not tested as thoroughly as standard KNIME releases. Furthermore new nodes or functionality may change substantially (or disappear entirely) from one build to the next. It's even possible that workflows you edit or create with nightly builds stop being readable by future (or past) versions. These nightly builds are a great way to get a sneak peek at what may be coming in the next version of KNIME and provide feedback and suggestions. They are not a particularly safe way to do real work." Below the disclaimer, there's a "Terms and Conditions" section with a checkbox for "I read and understand the disclaimer and accept the Terms and Conditions." and a "Go to downloads" button. The footer of the website shows "© 2018 KNIME AG. All Rights Reserved." and the page number "12".

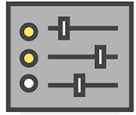
<https://www.knime.com/form/nightly-build>

KNIME Server & Cloud

— Jon Fuller —



Automate: Remote Control

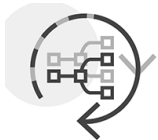


Control

See what's happening to your workflow on the Server

Edit your workflow directly on the KNIME Server

Control the Server access to protected resources

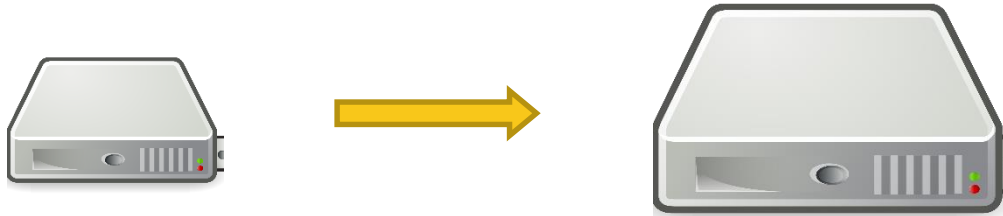


Scale

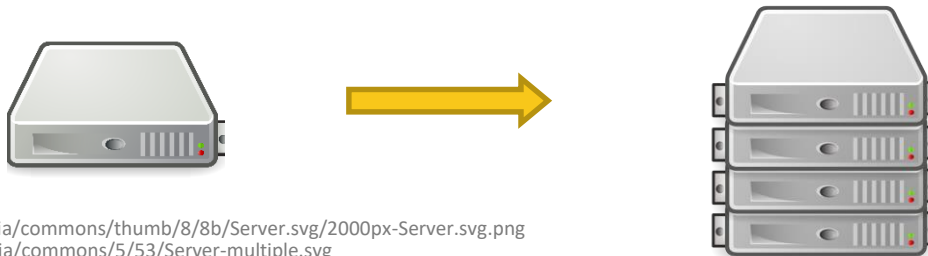
Use powerful server hardware to execute the workflow

Automate: KNIME Server Distributed Executors

- Need more workflow horsepower?
- KNIME Server currently supports ‘Scale Up’



- KNIME Server Distributed Executors allows ‘Scale Out’



<https://upload.wikimedia.org/wikipedia/commons/thumb/8/8b/Server.svg/2000px-Server.svg.png>
<https://upload.wikimedia.org/wikipedia/commons/5/53/Server-multiple.svg>

Automate: KNIME Server Distributed Executors

- Automatically train and evaluate 300,000 models using 80 executors running on AWS
- <https://www.knime.com/blog/beauty-and-the-monster>

9:30 AM - Session 3

- On Monsters and Tags...
Jeany Prinz & Greg Landrum (KNIME)
- Deploying KNIME in an Amazon Cloud Environment for High-Throughput Image Analysis
Andries Zjijstra (Vanderbilt/Nashville)
- A Data Pipeline Approach to Orphan Disease Insights
Sebastian Lefebvre (Alexion Pharmaceuticals)

Manage and Monitor (Personalisation)

The screenshot displays the KNIME Analytics Platform interface with several annotations for personalization:

- KNIME Explore Menu:** The 'Contact Support' option is highlighted in the 'KNIME Explore' menu.
- Welcome Page:** The 'Welcome to KNIME Analytics Platform!' page is shown. Annotations include:
 - Add a custom support email address:** Points to the 'Contact Support' menu item.
 - Add your company logo:** Points to the 'ACME Airlines' logo placeholder.
 - Customize the welcome page:** Points to the main content area of the welcome page.
- Preferences Dialog:** The 'Preferences' dialog is open, showing the 'General' tab. A yellow circle highlights the 'Available Software' section, which lists:
 - ACME Airlines Secret Sauce Update Site
 - KNIME Analytics Platform Nightly Update
 - KNIME Community Contributions (3.6)
 - KNIME Partner Update Site
 - KNIME Community Contributions
- Status Bar:** A yellow banner at the bottom right states: **You are using a nightly build!**

General Performance and Security Enhancements

- Explorer – now using REST, faster
- Speedups for servers running on AWS/Azure
- Implemented new recommendations for server hardening (CIS)

Edit Mount Point

Edit a resource for display in the KNIME Explorer

KNIME ServerSpace

Server name or address:

Server address:

Use REST: ☒

Server login credentials:

Username: Password (optional):

Test Connection

Enter the name of the KNIME Server (Mount ID).

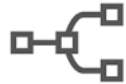
A default is fetched from the server when "Test Connection" is pressed or 'Use REST' is selected.

Mount ID:

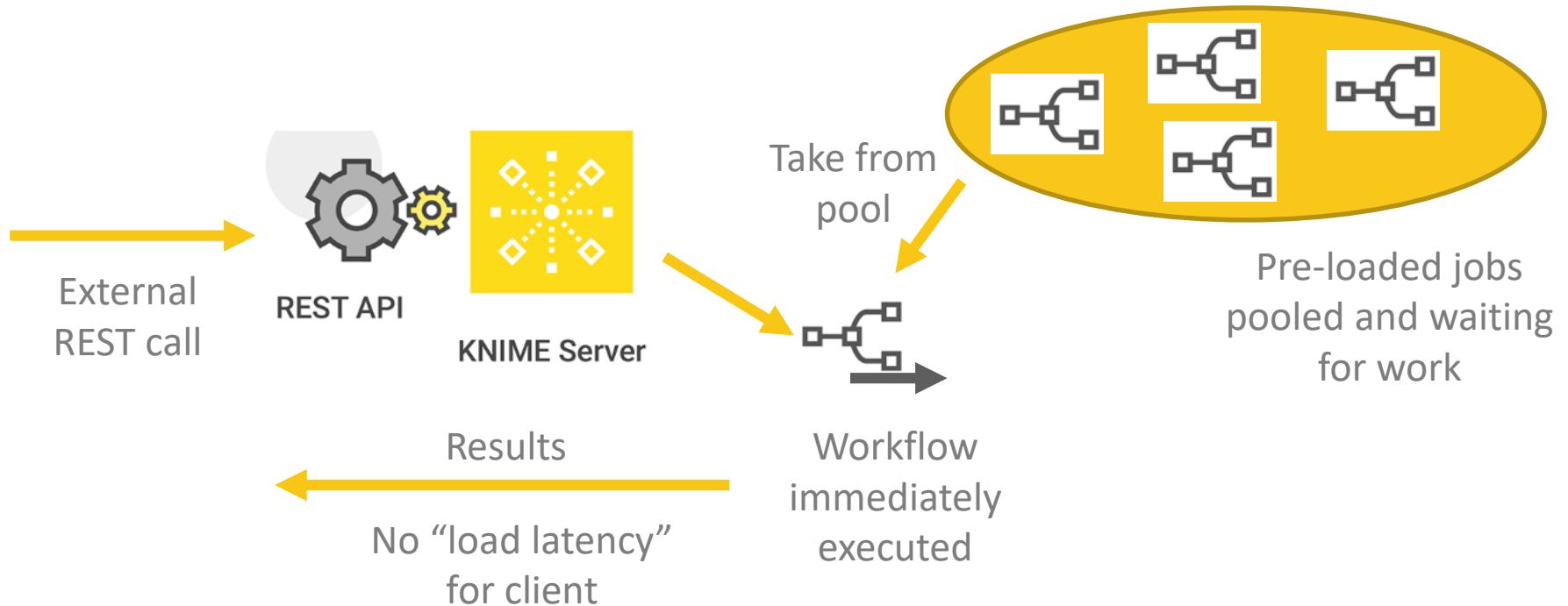
Reset Mount ID

OK Cancel

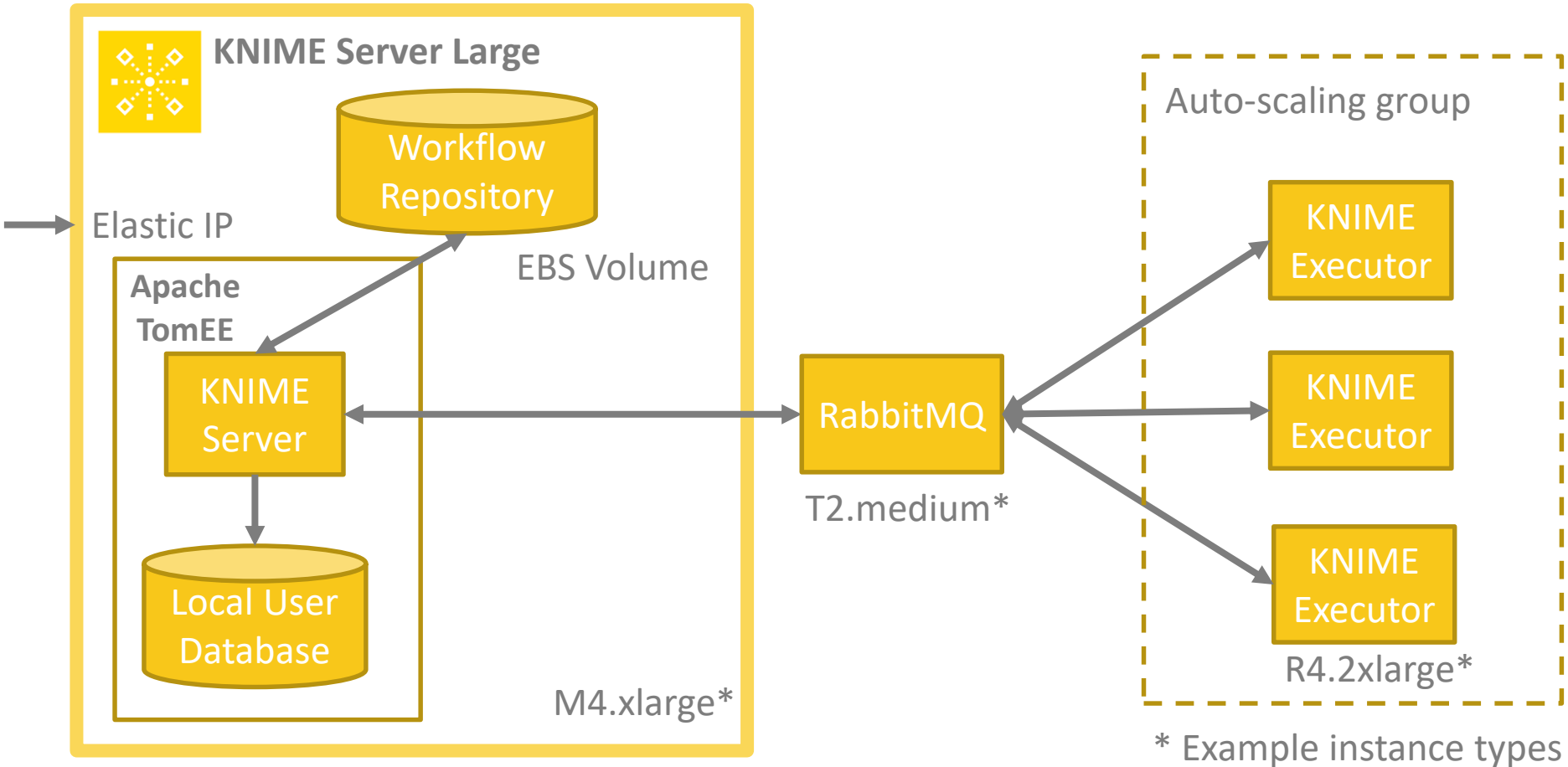
KNIME Server REST API: Job Pool



KNIME Workflow for e.g. sentiment prediction, churn score, chemical property



Manage and Monitor: KNIME Server Large for AWS

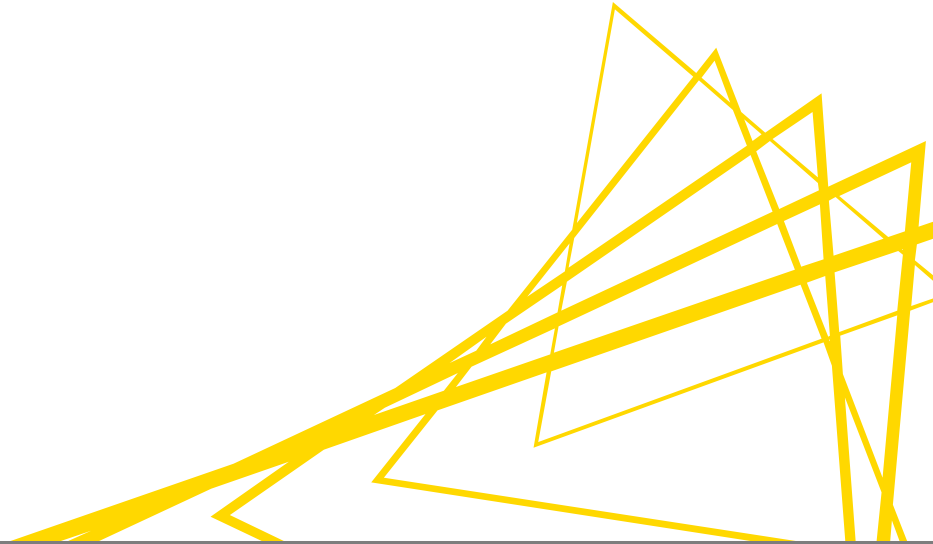


Manage and Monitor: KNIME Server Large for AWS

- Buy through marketplace
- Templated launch
- On-demand billing
- Elastic scaling for more workflow execution power

KNIME Managed Scoring Service

– Jason Tyler –



Considerations

- I have a workflow that takes data, applies an algorithm/model and returns a score.
- I need to deploy that to hundreds or thousands of end users, where there may be spikes in demand.
- I need to update the model/workflow periodically
- I don't want to think about servers, hosting, building services, etc...

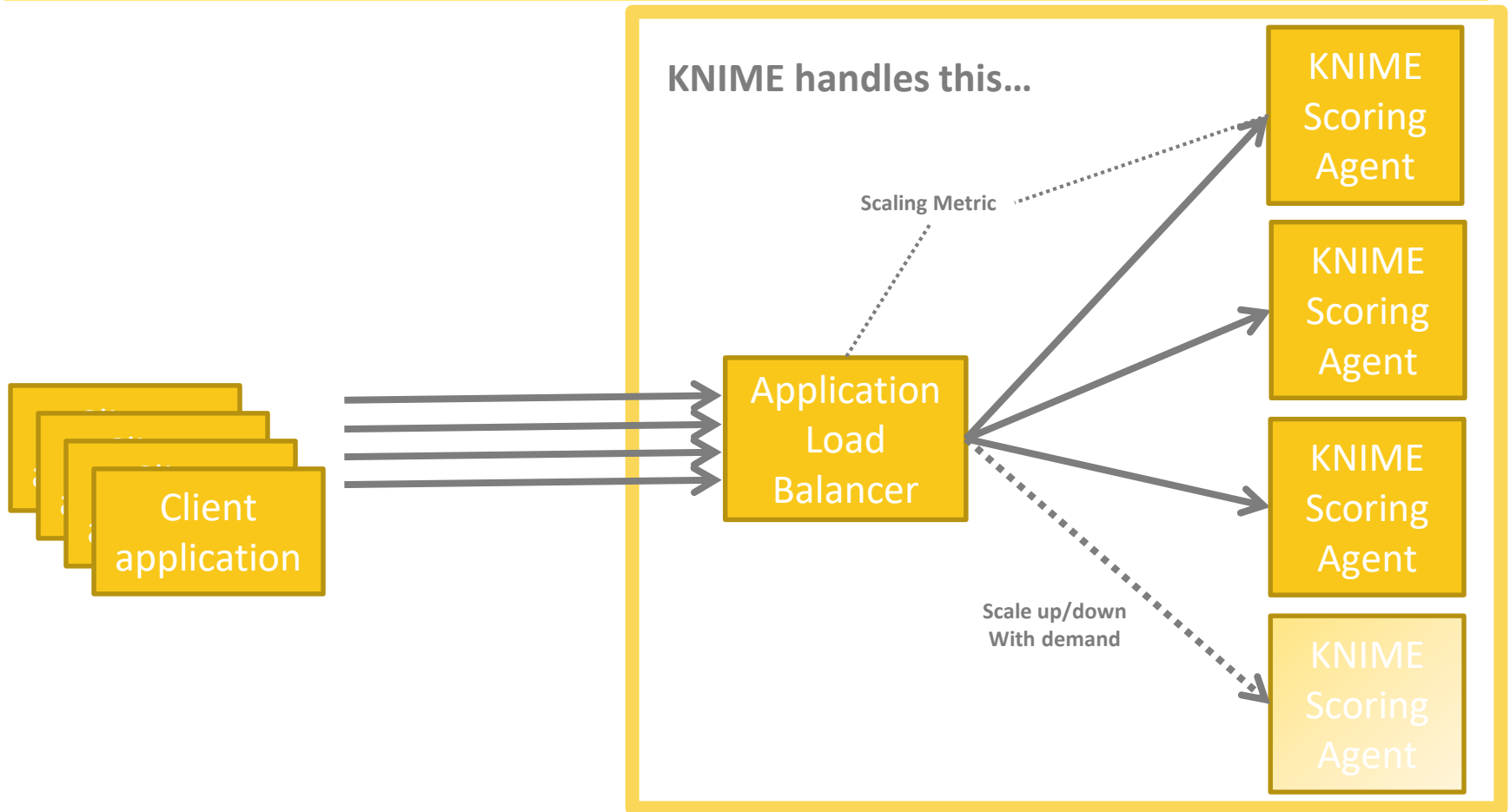
KNIME Managed Scoring Service

- The KNIME Managed Scoring Service is a hosted service that allows provisioning and consuming of score-based workflows as publicly available web APIs.

KNIME Managed Scoring Service

- Built around a lightweight agent designed for concurrently scoring individual workflows
- Wraps scoring workflows in synchronous web requests
- Modular, container based infrastructure
- Every workflow runs in its own isolated environment
- Designed for rapid scale-in/out to allow high performance, cost effective scoring

KNIME Managed Scoring Service



Sentiment Analysis Example

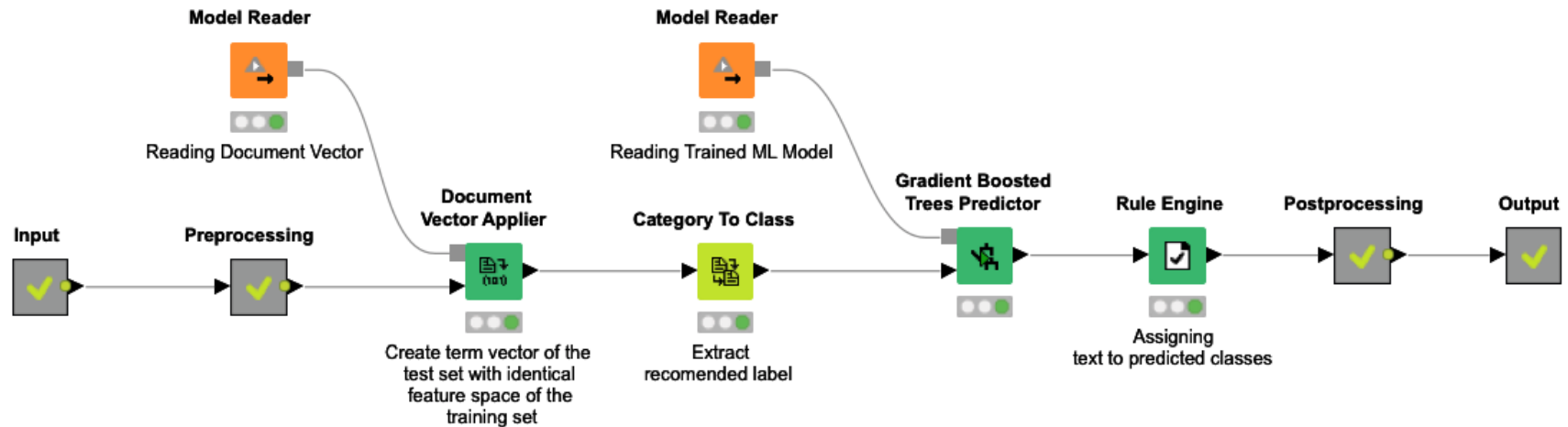
- Lets take a look at the Sentiment Analysis end point in a bit more detail:
- It's just a URL: <https://sentiment-prediction.dev.scoring.knime.com/score>
- Just POST your data and get a result
 - { “content”: [“I had a great flight! Awesome staff!”]}

Sentiment Analysis Scoring Workflow

REST API for Sentiment Analysis

This workflow showcases the Rest API capabilities of the KNIME Server. We are reusing the model trained on the airline reviews and providing sentiment analysis as a service.

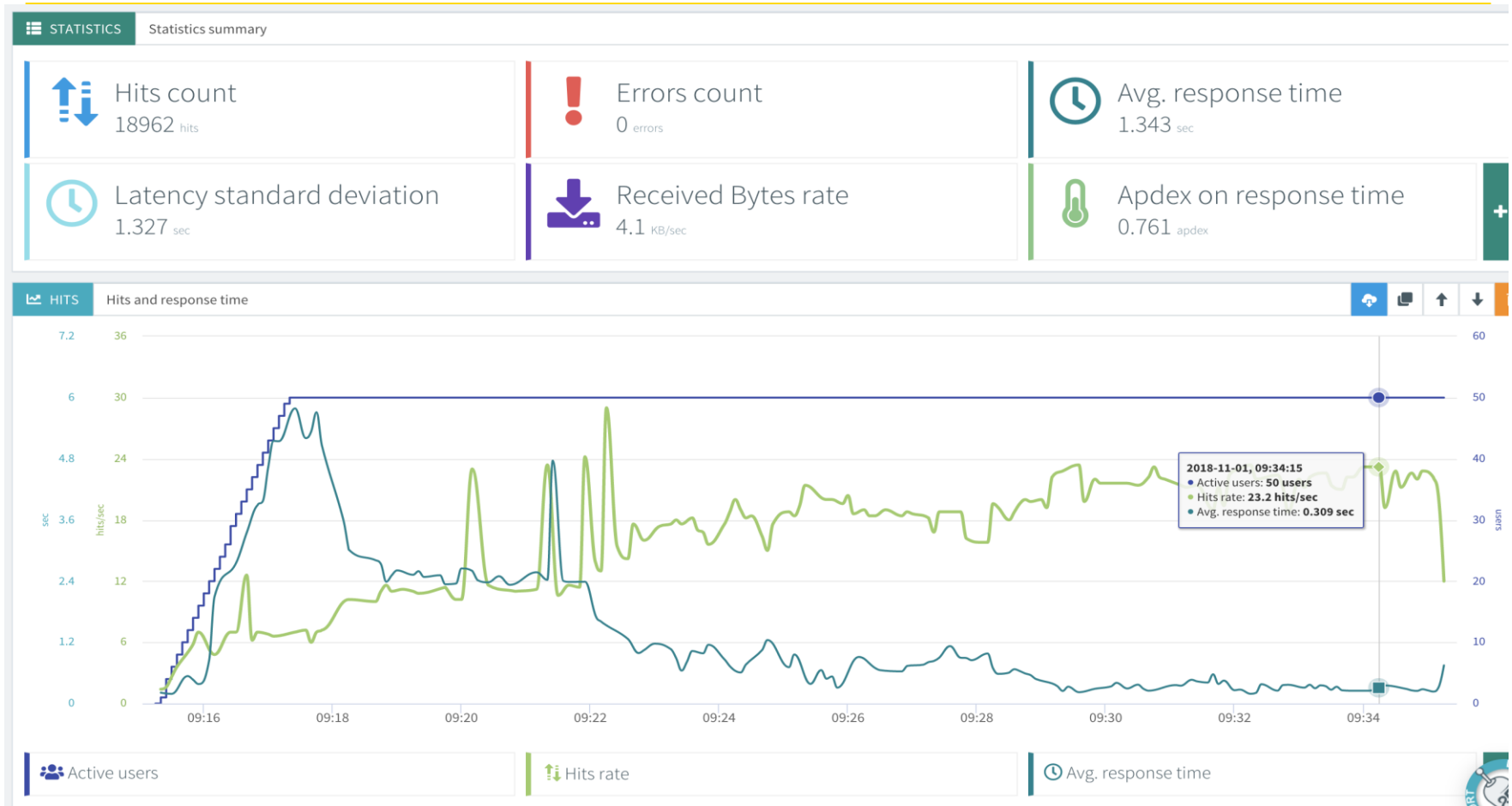
Please try our corresponding responsive, mobile ready application to see sentiment prediction in action.



Managed Scoring Service

- Demo
- Try it out yourself!
 - <https://knime.org/airline>

Latency, Scaling, Throughput

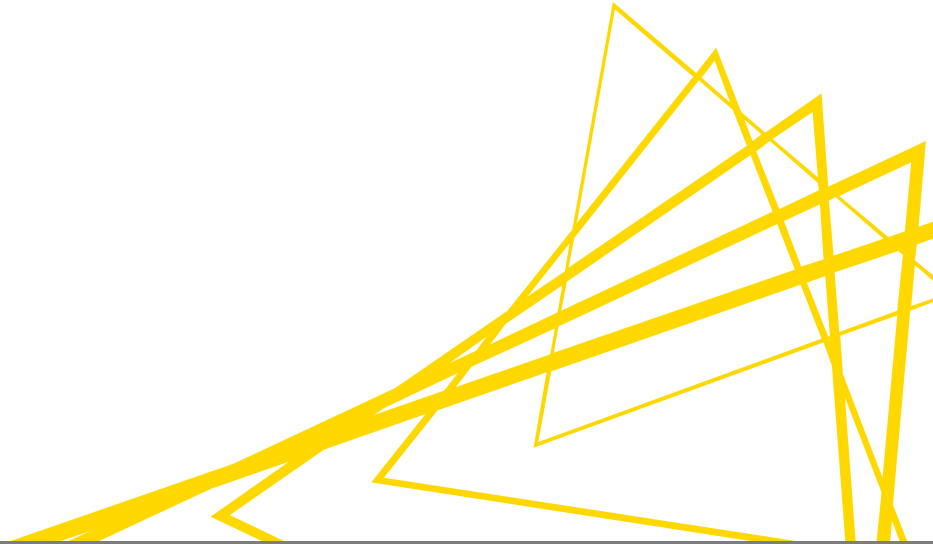


What's Coming?














- Preview launch in December
 - Open to public, look for announcements
 - Available in multiple geographies
 - Free to try!
- Who to contact for more information?
 - [Jim Falgout](#)

Big Data & Database integration

– Tobias Koetter –



PySpark Script Nodes

- ▼  Apache Spark
 - >  IO
 - >  Column
 - >  Mining
 - ▼  Misc
 - >  Java Snippet
 - >  Management
 - ▼  PySpark
 -  PySpark Script (1 to 1)
 -  PySpark Script (1 to 2)
 -  PySpark Script (2 to 1)
 -  PySpark Script (2 to 2)
 -  PySpark Script Source

**PySpark
Script Source**



Dialog - 2:1 - PySpark Script (1 to 1)

File

Script | Flow Variables | Job Manager Selection | Memory Policy

Columns

- Universe_0_0
- Universe_0_1
- Universe_1_0
- Universe_1_1
- Cluster Membership

Flow variables

- knime.workspace

```
1 # System imports
40 #Custom imports
41 #Flowvariables
42
43 #Custom globals
44 # Initialization of Spark environment
52 # Custom pySpark code
53 # SparkSession can be used with variable spark
54 # The input dataframe(s): [dataFrame1]
55 # The output dataframe(s) must be: [resultDataFrame1]
56 resultDataFrame1 = dataframe1
57
58
59 # End of user code
60 # Send data to jvm
62
```

Validate on Cluster Number of rows to validate on 50

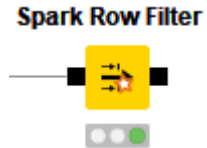
Execution finished.

resultDataFrame1(10 of 50 rows):

Universe_0_0	Universe_0_1	Universe_1_0	Universe_1_1	Cluster Membership
10.45988140338...	10.39797592433...	10.12912234078...	10.31661006779...	Cluster_01
10.63284098962...	10.42015828384...	10.18761452018...	10.52522056130...	Cluster_01

OK Apply Cancel ?

Spark Row Filter



Dialog - 0:5 - Spark Row Filter

File

Conditions Flow Variables Job Manager Selection Memory Policy

Preview

- OR
 - AND
 - TailNum RLIKE B.*
 - Year BETWEEN [2015, 2016]
 - AND
 - TailNum RLIKE A.*
 - Year BETWEEN [2017, 2018]

Edit condition

Year BETWEEN 2017 2018

Delete

Add Condition Group Ungroup Delete

OK Apply Cancel ?

Apache Livy Support

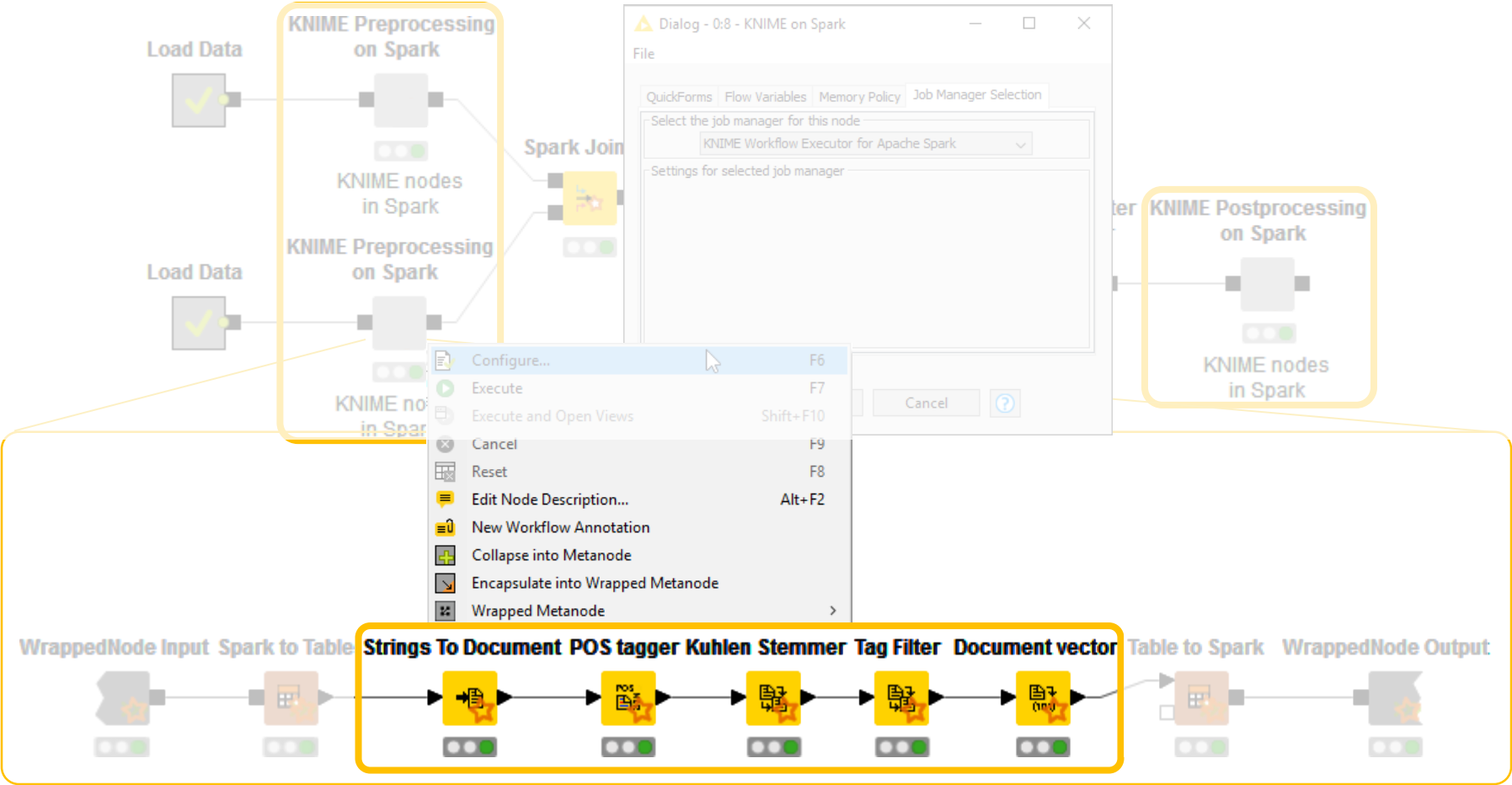
- Gives out-of-the-box Spark compatibility with:
 - Hortonworks HDP
 - Amazon EMR
 - Microsoft Azure HDInsight
- Cloudera packaging provided by KNIME
- Revised node dialog

A screenshot of a software dialog box titled "Dialog - 2:2 - Create Spark Context via Livy (preview)". The dialog has a "File" menu and several tabs: "General", "Advanced", "Flow Variables", "Job Manager Selection", and "Memory Policy". The "General" tab is active. It contains the following fields and options:

- "Spark version:" with a dropdown menu set to "2.3".
- "Livy URL:" with a text input field containing "http://<host>:8998/".
- "Authentication:" section with two radio buttons: "None" and "Kerberos" (which is selected).
- "Spark executor resources:" section with a checked checkbox "Override default Spark executor resources". Below this are "Memory:" (set to 16) and "Cores:" (set to 4), both with up/down arrows and a "GB" unit dropdown.
- Allocation options: "Default allocation", "Fixed allocation", and "Dynamic allocation" (which is selected).
- "Minimum number of executors:" set to 1.
- "Maximum number of executors:" set to 10.
- "Estimated total cluster resources:" text: "20-182 GB of memory and 5-41 cores."
- "Estimated per-container resources:" section with two bullet points:
 - one Spark driver with 2 GB of memory and 1 core(s)
 - 1-10 Spark executors, each with 18 GB of memory and 4 core(s)

At the bottom are "OK", "Apply", "Cancel", and a help icon (?) buttons.

Native KNIME Node Execution in Apache Spark (Preview)

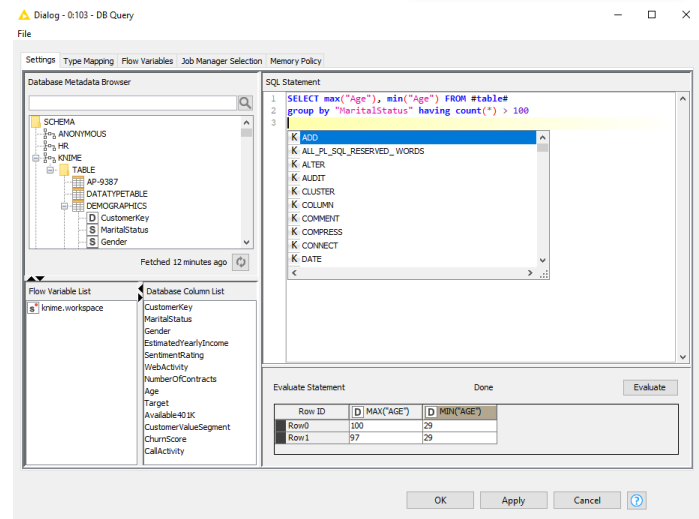
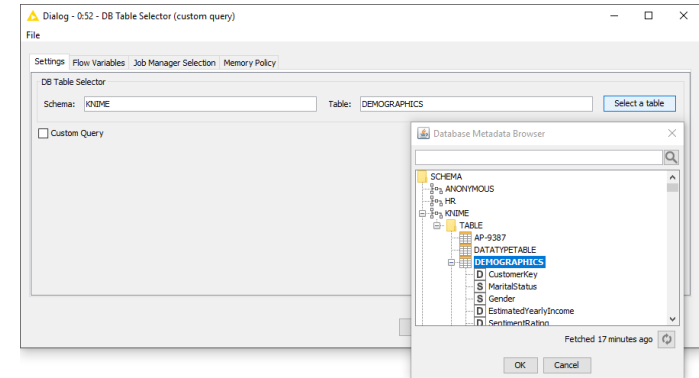
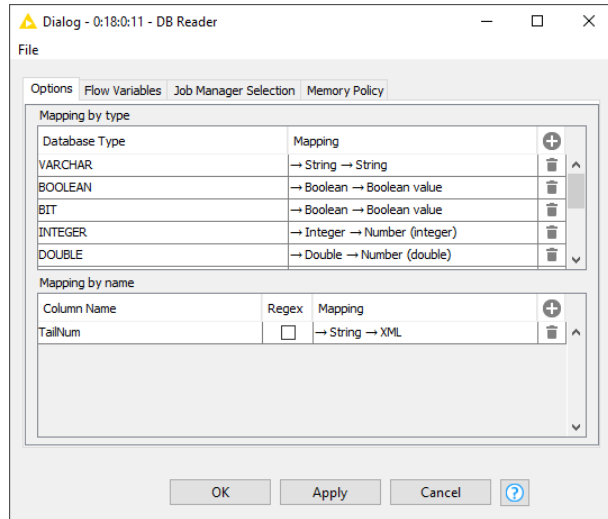


(New) Database Integration

(Preview)

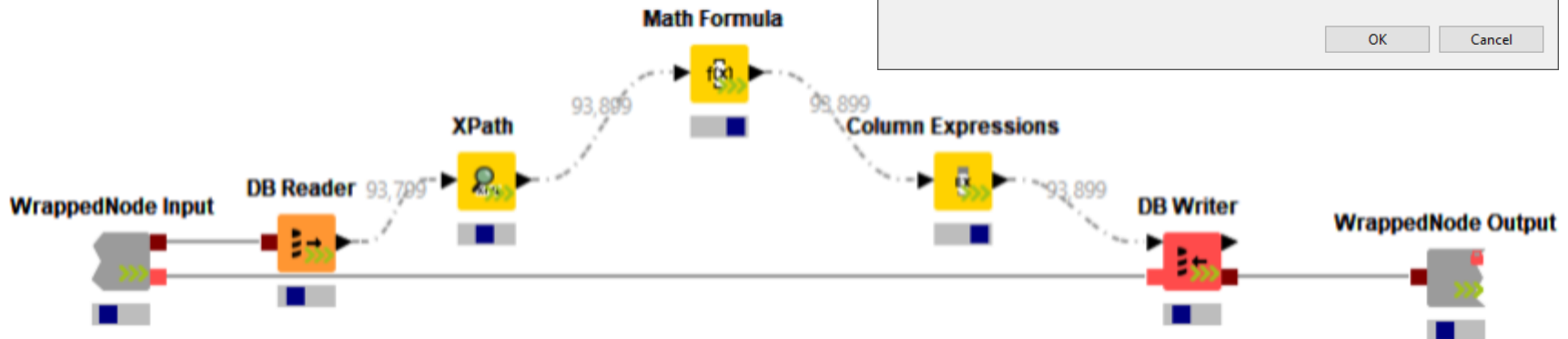
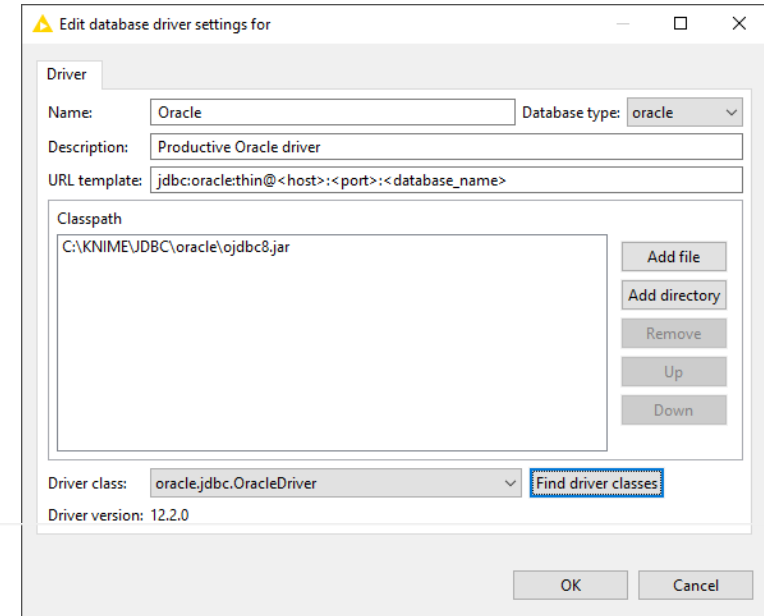
Usability Improvements

- Improved schema handling
- Flexible type handling
- Rich SQL editor

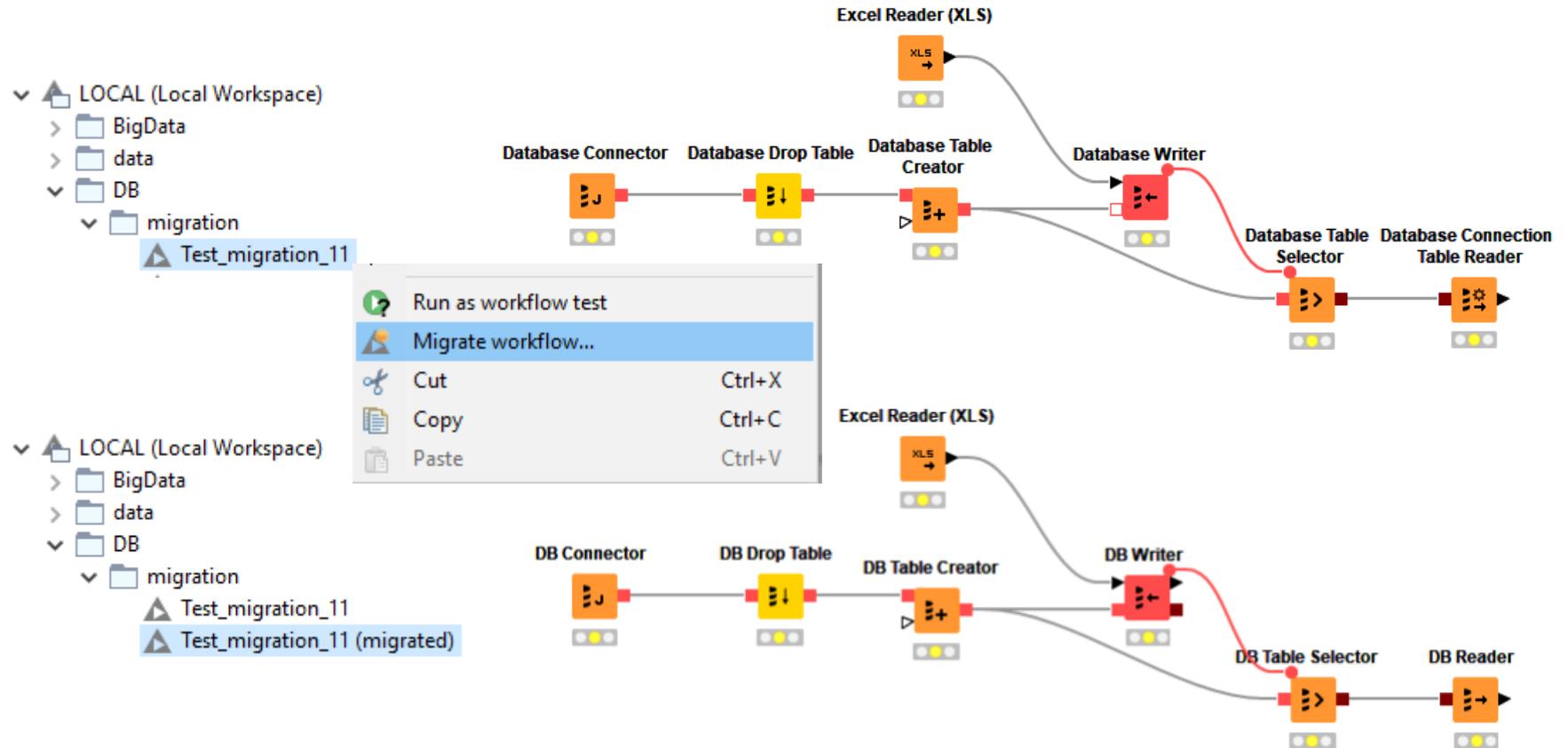


Framework Improvements

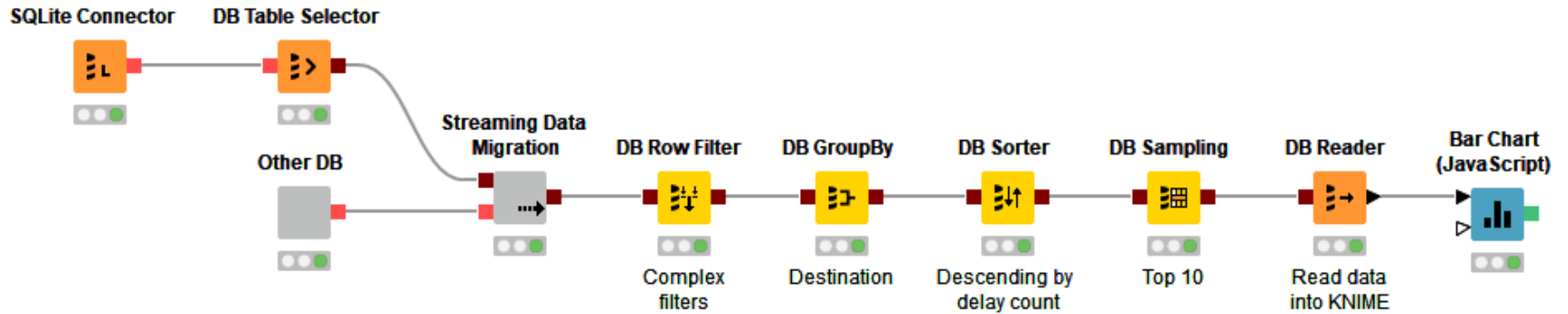
- Driver management
- Parallel execution
- Streaming execution



Workflow Migration Tool



Demo



Integrations

– Bernd Wiswedel –

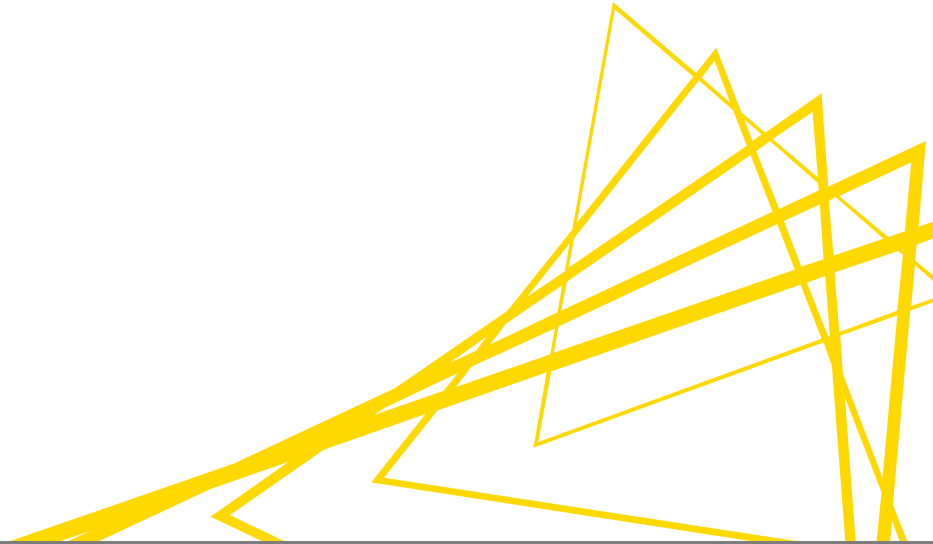
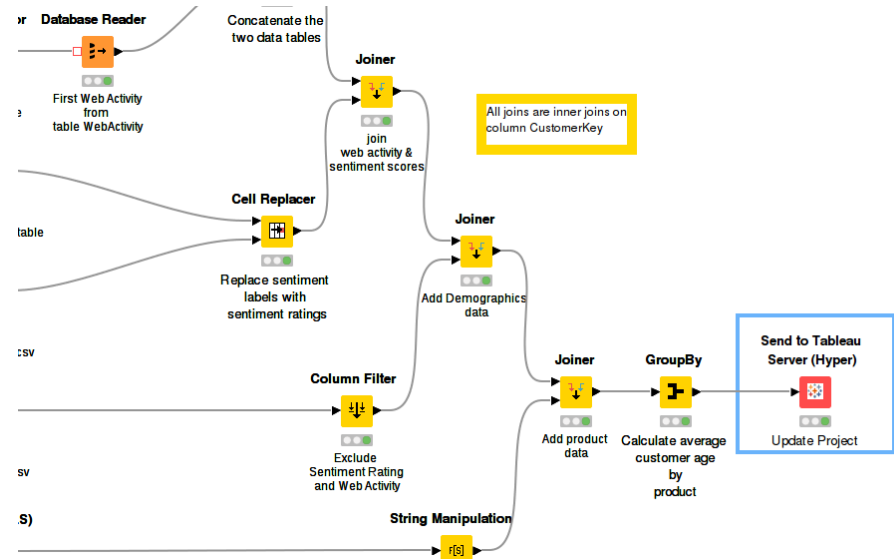
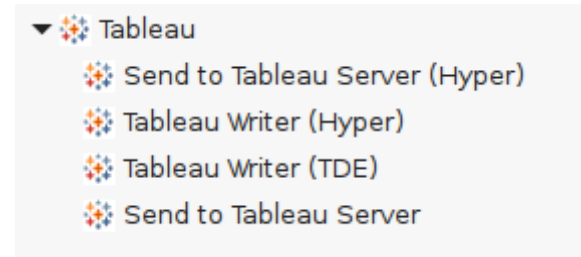


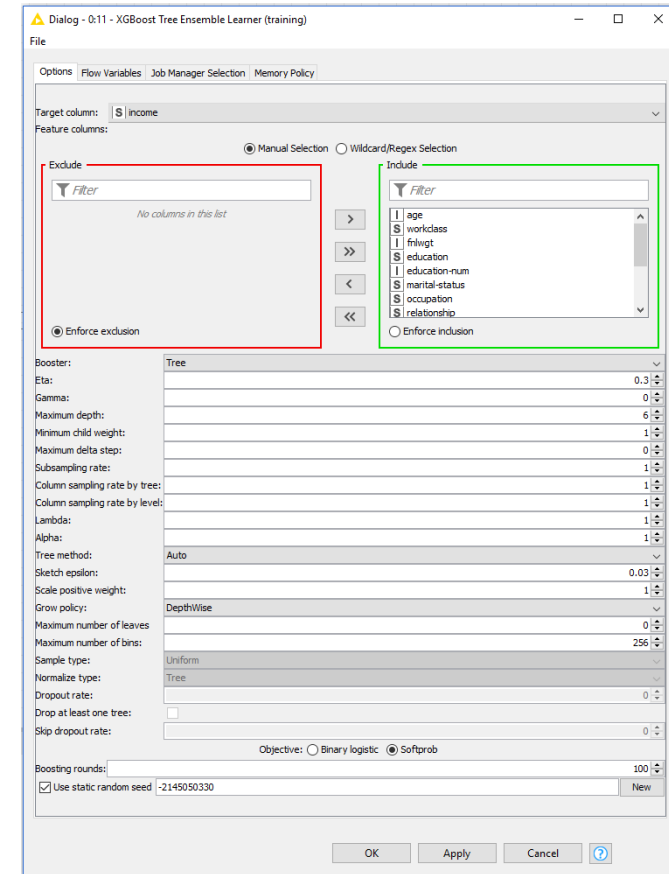
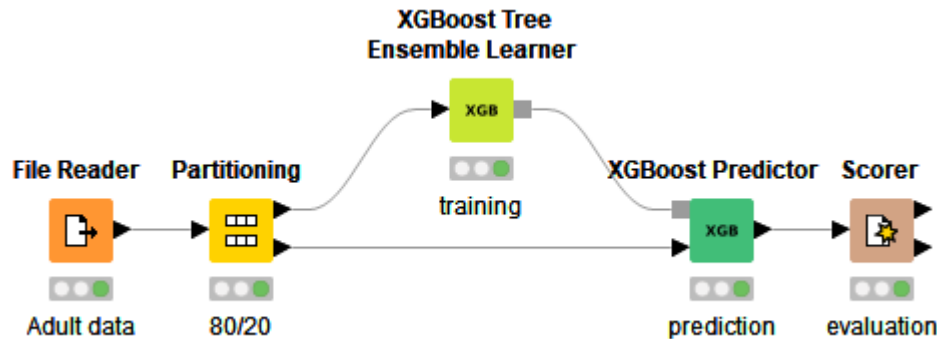
Tableau Integration – Updated

- Supporting Hyper and TDE format
- Write Extracts and push them to Tableau Server
- New capabilities
 - Appending files
 - Date & Time Support



XGBoost Integration

- Popular open-source library for optimized distributed gradient boosting
- Often used in machine learning competitions

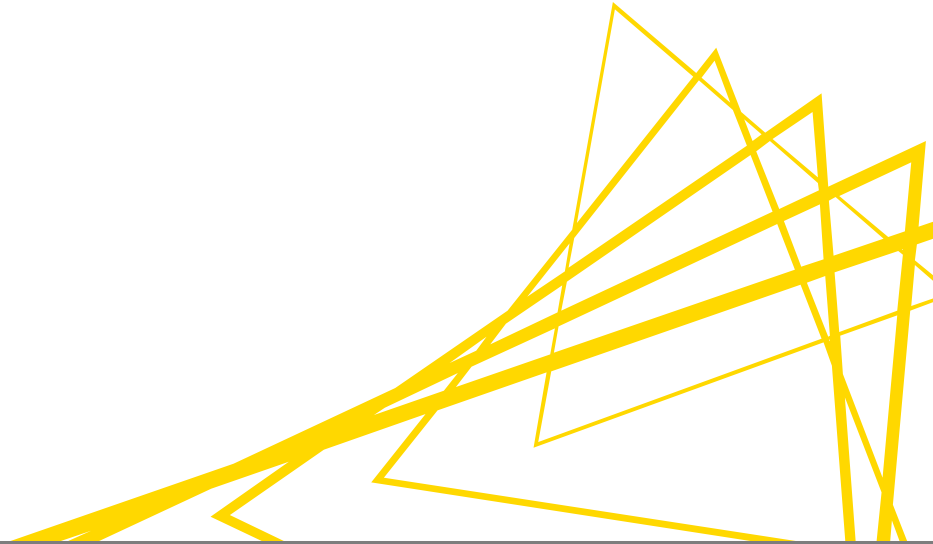


Other noteworthy changes

- Google Drive Connector
- More statistics nodes (hypothesis tests)
- Changes to underlying data format using Apache Parquet (speed-up KNIME execution)

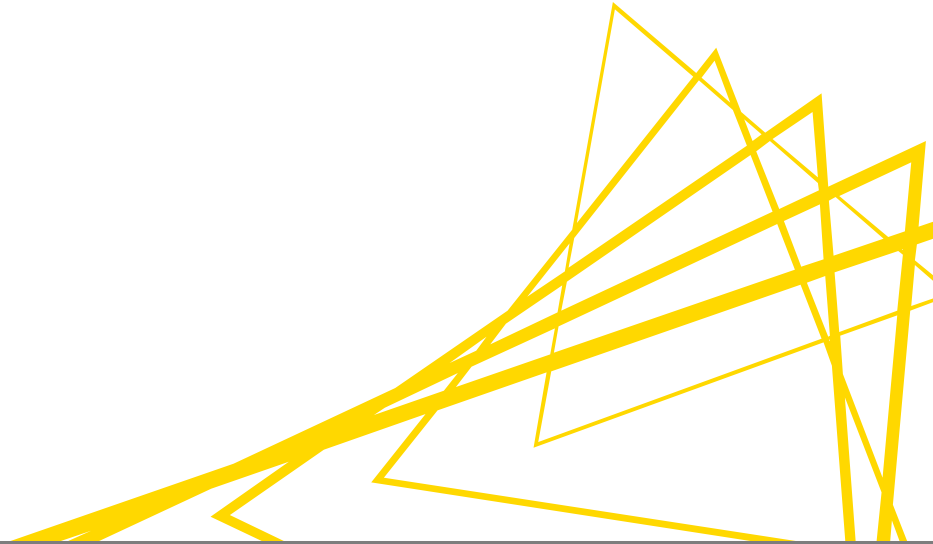
KNIME Python Integration

– Davin Potts & Greg Landrum –



Views & Guided Analytics Applications

– Greg Landrum –



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.