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

https://www.knime.com/form/nightly-build
KNIME Server & Cloud
– Jon Fuller –
Automate: Remote 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

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’
Automate: KNIME Server Distributed Executors

- Automatically train and evaluate 300,000 models using 80 executors running on AWS

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 Zijestra (Vanderbilt/Nashville)*
- A Data Pipeline Approach to Orphan Disease Insights
  - *Sebastian Lefebvre (Alexion Pharmaceuticals)*
Manage and Monitor (Personalisation)
General Performance and Security Enhancements

• Explorer – now using REST, faster
• Speedups for servers running on AWS/Azure
• Implemented new recommendations for server hardening (CIS)
KNIME Server REST API: Job Pool

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

External REST call

REST API

KNIME Server

Take from pool

Pre-loaded jobs pooled and waiting for work

No “load latency” for client

Workflow immediately executed

Results
Manage and Monitor: KNIME Server Large for AWS

KNIME Server Large

- Workflow Repository
- EBS Volume
- Local User Database
- Apache TomEE
- KNIME Server
- RabbitMQ
- T2.medium*

Auto-scaling group

- KNIME Executor
- KNIME Executor
- KNIME Executor
- R4.2xlarge*

* Example instance types

Elastic IP

M4.xlarge*
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

KNIME handles this...

Application Load Balancer

Scaling Metric

Client application

KNIME Scoring Agent

Scale up/down With demand
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](https://knime.org/airline)
Latency, Scaling, Throughput

**Statistics Summary**

- **Hits count**: 18962 hits
- **Errors count**: 0 errors
- **Avg. response time**: 1.343 sec
- **Latency standard deviation**: 1.327 sec
- **Received Bytes rate**: 4.1 KB/sec
- **Apdex on response time**: 0.761 apdex

**Hits and Response Time**

The graph shows the changes in hits, rate, and average response time over time. The x-axis represents time from 09:16 to 09:34, and the y-axis represents hits, rate, and response time. The data points indicate the active users and the corresponding hits rate and average response time.

- **Active users**: 50 users
- **Hits rate**: 23.2 hits/sec
- **Avg. response time**: 0.309 sec
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

![Image of Apache Spark and PySpark nodes]

![Image of PySpark Script Node dialog and execution results]

© 2018 KNIME AG. All Rights Reserved.
Spark Row Filter
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
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
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 –