Sharing and Deploying Data Science with KNIME Server

Roland Burger
Marten Pfannenschmidt
KNIME
KNIME Server

• Collaboration
  – Share Expertise and Templates/Blueprints

• Automation
  – Schedule, Monitor, Update

• Deployment
  – To Humans: Analytical Applications ("Guided Analytics")
  – To Machines: via REST and custom APIs

• Management
  – Manage distributed setups
  – Integrate with existing enterprise setup (authentication)
Delivering Data Science

Data Science

- Analytics as a service
- Guided Analytics
- Analytical Apps & Dashboards

- Data Access
- Data Blending
- Transformation
- In-database
- “In” Big Data

Marketing and Sales
- Decisions

Financial
- Decisions

Manufacturing
- Decisions

Product R&D
- Decisions

IT
- Decisions

Data Blending

Data Analytics

Predictive Analytics

Machine Learning

Artificial Intelligence

- Statistics
- Mining
- AI/ML/DL Models
- R/Python Scripts
- Legacy Languages
Creating Data Science: KNIME Analytics Platform

Data Science
- Data Blending
- Data Analytics
- Predictive Analytics
- Machine Learning
- Artificial Intelligence

KNIME Analytics Platform
- KNIME Integrations
- KNIME Extensions
- Community Extensions
- Partner Extensions

Roles:
- Data Engineers
- Data Analysts
- Data Scientists
- ML/AI Engineers
- App Developers
Productionizing Data Science: KNIME Server

Collaborate  Automate  Manage  Deploy

KNIME Server

Data Science Teams  IT Teams  Business Teams

Guided Analytic Apps

- REST API
- Cloud-based Services
- Mobile Applications
- Web Applications

Apps

- Marketing and Sales
- Financial
- Product R&D
- Manufacturing
- IT
Better Decision-making, Faster!

KNIME Analytics Platform

Load > Integrate > Transform > Analyze > Visualize

Open Source

KNIME Integrations

KNIME Extensions

Community Extensions

Partner Extensions

KNIME Server

Collaborate — Automate — Manage — Deploy
Collaboration

• Connect to KNIME Server from local KNIME Analytics Platform
  – Authentication possible via LDAP / Active Directory
  – Centralized administration and security

• Interact with KNIME Server:
  – Upload workflows for sharing and collaboration with colleagues
  – Execute workflows, create schedules, set permissions, ...
  – Organize your work in workflow groups (directories)
Deploy workflows to KNIME Server

- Right-click -> Deploy to Server
  - Share workflows, data, etc. for collaboration
  - Continuous delivery in all steps of workflow life cycle
KNIME Workflow Hub

• A place to share knowledge about Workflows.
  – Share
  – Discover
  – Search
  – Rate
  – Comment
  – Version
Metanode Templates

- Create template from metanode and store on server
- Templates are linked -> check for new version and apply updates automatically
- Increase efficiency through reusable workflow components
- Ensure compliance with company policies
Remote Execution

- Execute workflows directly on KNIME Server
- Scale beyond limitations of desktop machines
Scheduled Execution

- Automate workflow execution in regular intervals
- Orchestrate job pipeline by using Call Workflow nodes
Remote Workflow Editor

- Monitor progress of execution on server (almost) live
- Very helpful for longer-running jobs
Remote Workflow Editor

- Change node configurations directly in the remote job
  - Resume execution if there was an error
  - Try different configuration options without re-executing the full job
  - Add nodes, change node connections, all live and in the secure server environment
KNIME WebPortal

- All workflows on KNIME Server available as web apps
- Step-by-step execution of workflows from any browser
- Simple, clean interface for end users
  - Guided Analytics
- Customize layout to match corporate design
Guided Analytics: Interactive Data Science

Interaction Points

User Guided Data Loading
- Select Data File
- File Reader
  - Read data from CSV

User Guided Analytics
- Define Cluster Parameters
  - Model configuration and feature selection
- k-Means
  - Perform k-Means clustering and produce PMML model

User Guided Deployment
- Color Manager
  - Color by cluster
- PMML Writer
  - Write clustering model to disk to deploy in production
- Display Cluster Result

© 2019 KNIME AG. All rights reserved.
Inside an Interaction Point

Clustering Results

- Results of k-Means clustering with k=3 clusters
- Data source: customerData.csv
- PMML model stored in knime/~/knime.workflow/./.02_models/cluster_model.pmml

<table>
<thead>
<tr>
<th></th>
<th>Day Mins</th>
<th>Eve Mins</th>
<th>Night Mins</th>
<th>Indi Mins</th>
<th>Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>167.0</td>
<td>134.6</td>
<td>242.2</td>
<td>7.4</td>
<td>cluster_0</td>
</tr>
<tr>
<td>2</td>
<td>170.9</td>
<td>132.3</td>
<td>112.9</td>
<td>11.6</td>
<td>cluster_2</td>
</tr>
<tr>
<td>3</td>
<td>183.2</td>
<td>126.8</td>
<td>203.3</td>
<td>11.2</td>
<td>cluster_0</td>
</tr>
<tr>
<td>4</td>
<td>172.7</td>
<td>133.4</td>
<td>259.5</td>
<td>9.8</td>
<td>cluster_1</td>
</tr>
<tr>
<td>5</td>
<td>161.5</td>
<td>130.7</td>
<td>186.5</td>
<td>8.5</td>
<td>cluster_2</td>
</tr>
<tr>
<td>6</td>
<td>180.1</td>
<td>127.5</td>
<td>237.4</td>
<td>7.5</td>
<td>cluster_2</td>
</tr>
<tr>
<td>7</td>
<td>192.0</td>
<td>127.6</td>
<td>155.6</td>
<td>7.5</td>
<td>cluster_2</td>
</tr>
<tr>
<td>8</td>
<td>197.1</td>
<td>130.1</td>
<td>78.1</td>
<td>7.4</td>
<td>cluster_2</td>
</tr>
</tbody>
</table>
KNIME Server REST API

- Enables external integration
- Build applications around KNIME Server
- e.g. for Microservices and real time scoring

```
"Day Mins":191.5,
"Eve Mins":175.2,
"Night Mins":220.3,
"Int1 Mins":0.0
```
KNIME Server REST API: Job Pool

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

External REST call

Take from pool

Pre-loaded jobs pooled and waiting for work

Results

No “load latency” for client

Workflow immediately executed

KNIME Executor

KNIME SERVER
KNIME Server REST API

- Swagger UI: Explore API definition of workflow in browser
- Provides all information needed to interact with API
KNIME Server REST API

- Programmatically control KNIME Server
  - Upload/download/delete resources
  - Staging from Dev to Prod
  - Upload licenses
  - Empty trash/restore items
  - Execute workflows
  - Schedule jobs
  - Set permissions
  - Create users & groups, etc.

All server functionality available via REST API
Version Control

- Create snapshots of workflow revisions
- Allow roll-back to previous versions
- Full workflow history available in client and on Workflow Hub
Workflow Difference

- Inspect differences between versions of the same workflow
Permission Management

- Manage permissions for workflows and data on KNIME Server
- Control which groups can access, modify, or execute a workflow
- Use LDAP / Active Directory groups for centralized administration
Managing Preferences – local

How do I configure Python?

How do I install database drivers?

KNIME Website
Internal knowledgebase
Managing Preferences – KNIME Server

- Different departments/teams have different requirements

- Multiple OS deployments
  - Windows 7
  - Windows 10
  - Linux
  - macOS

Windows 10
Hive
Spark

Windows 7
Oracle
MS Access

Linux and macOS
Python
R
Managing Preferences – KNIME Server

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
Client Customization Profiles

ADD CUSTOM SUPPORT E-MAIL ADDRESS

Welcome to KNIME Analytics Platform!

ADD YOUR COMPANY’S LOGO

CUSTOMIZE THE WELCOME PAGE

ADD AN UPDATE SITE

Where to go from here
- Create new workflow
- Learning Hub
- Browse example workflows
- Get additional nodes
- Go to my workflows
- Mount KNIME Server
- ACME Airlines Workflow Hub

Tips & Tricks
- Data access

© 2019 KNIME AG. All rights reserved.
KNIME Server – Admin made easy
Distributed Executors

• Need more workflow horsepower?

• KNIME Server currently supports ‘Scale Up’

• KNIME Server Distributed Executors allows ‘Scale Out’
Distributed Executors

Requests to execute workflows

KNIME Server

RabbitMQ

Executor 1
Executor 2
Executor 3

KNIME Server + Distributed Executors
Distributed Executors

Request to execute workflow

KNIME Server

RabbitMQ

Executor 1

Scripted launch on AWS CloudFormation

KNIME Server + Distributed Executors
Distributed Executors

More requests to execute workflows

KNIME Server

RabbitMQ

Executor 1

AWS Auto-scaling Group

Executive 2

CPU load increases

Launching

Executor 2

KNIME Server + Distributed Executors
Distributed Executors

Requests to execute workflows

KNIME Server

RabbitMQ

Executor 1

Executor 2

AWS Auto-scaling Group

KNIME Server + Distributed Executors
KNIME Software – On premise and in the cloud

On-Premise

KNIME Analytics Platform

KNIME Server

Cloud

KNIME Analytics Platform

KNIME Server

AWS partner network

Azure

KNIME Analytics Platform

KNIME Server

Open for Innovation
KNIME Server

Collaboration
- Workflow Hub
- Workflows
- Metanodes, Nodes
- Data
- Shared Repository, Versioning

Automation
- Provisioning
- Remote, Client Control
- Distributed Executors

Deployment
- Web Front-ends
- Guided Analytic Apps
- RESTful Services

Management
- Nodes, Files, Applications
- Security
- Configuration & Versioning
- Client & User Preferences

Ready to get started with KNIME Server? Contact us for a free trial:
contact@knime.com