

Open for Innovation

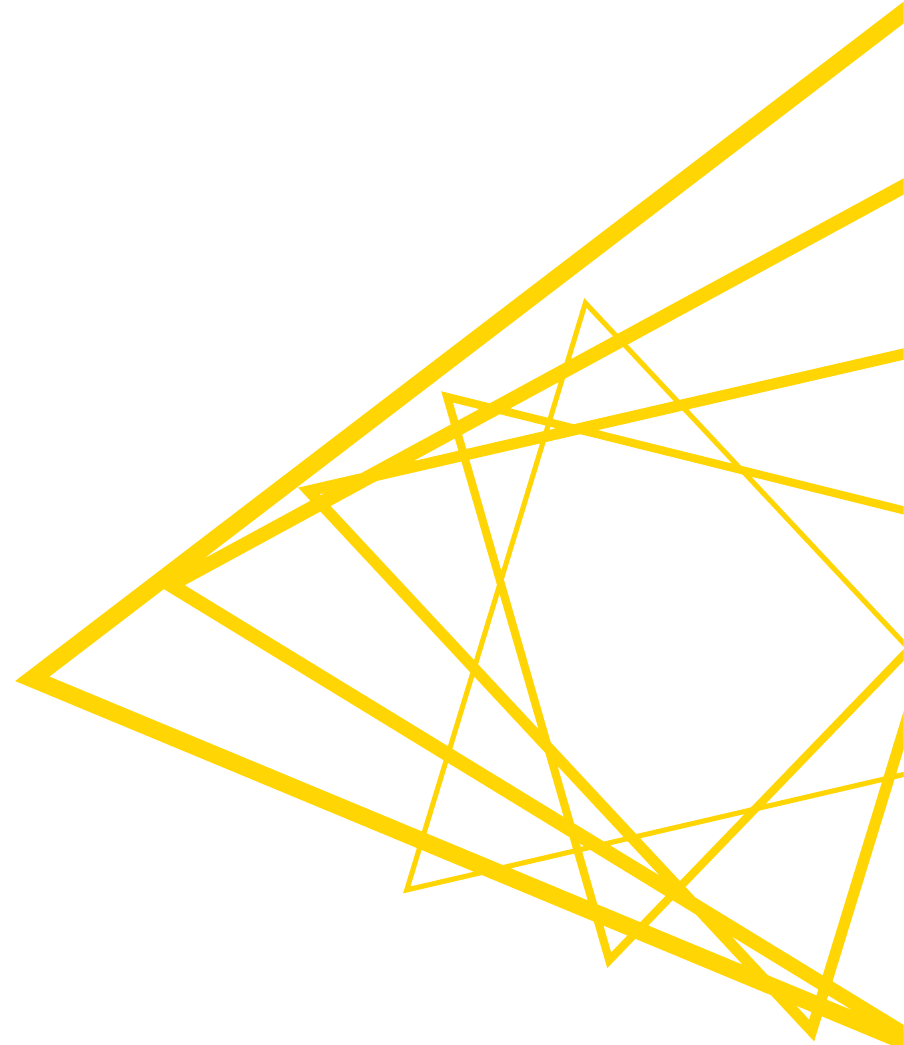
KNIME

KNIME Edge Introduction

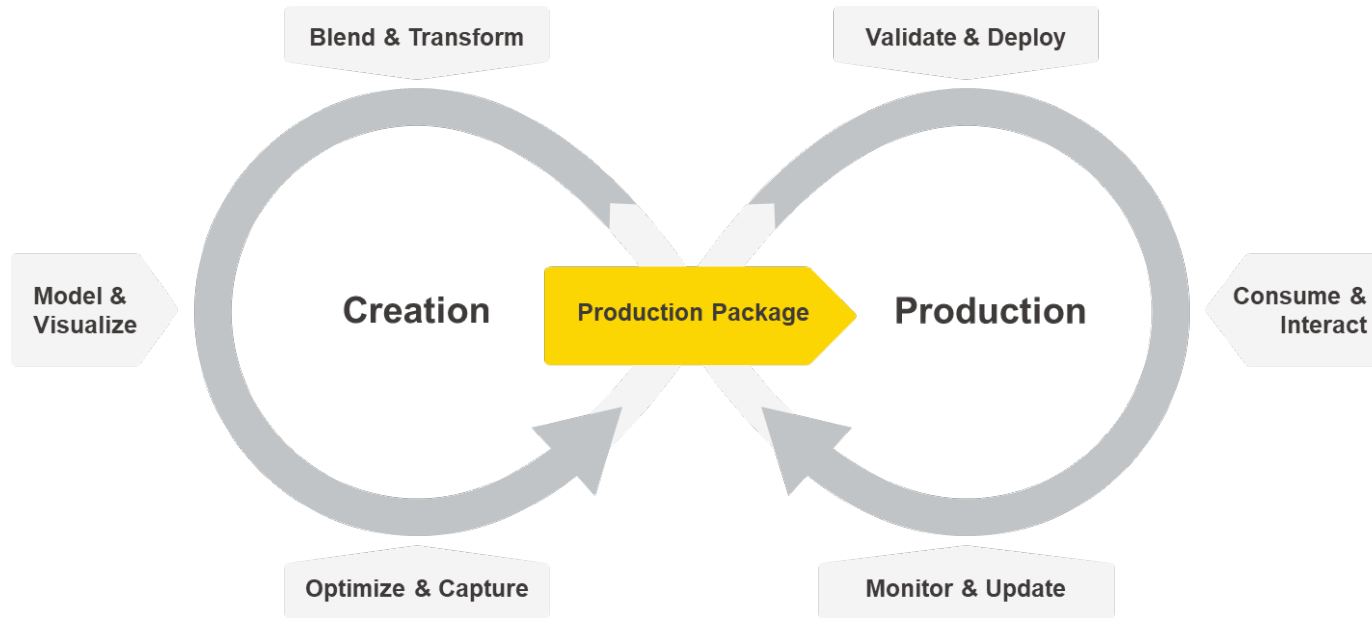
Data Science in Action

Inference as a Service

September 29, 2021



KNIME Software – One Ecosystem



KNIME Analytics Platform

KNIME Extensions	KNIME Integrations	Community Extensions	Partner Extensions
------------------	--------------------	----------------------	--------------------

KNIME Server

Team Collaboration	End User Applications	API Services	Managed Execution
--------------------	-----------------------	--------------	-------------------

Customer Use Cases - Inference

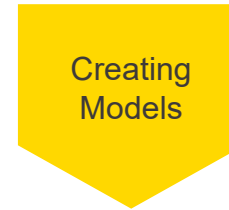
- Manufacturing – Automating Defect Detection*
 - Geographically dispersed factories
 - Defect analysis of manufactured item images using deep learning techniques
 - Models distributed to factories
 - Apply models to images of boards in manufacturing process to detect defects
 - Automating this process provides a substantial boost in defect detection
 - Fewer defects increases customer satisfaction

- Financial Services – Fraud Detection Service*
 - Serving banking and financial services companies
 - Customers make API calls into the service during transaction processing to determine the probability of fraud
 - Runs on a cloud platform with external ingress for clients
 - Requires hundreds of transactions per second to support SLA's
 - Request volume increasing as more customers are onboarded

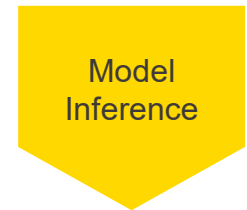
* These are two potential use cases, there are many more!

Inference on Demand (as a Service)

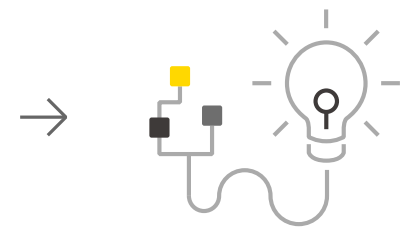
- Inference Deployments
 - Inference -> Application of a model
 - A workflow that accepts data, applies a model and returns an answer
 - Deployed as a service – API endpoint
- Deployed to many, concurrent end users
- With spikes in demand
- Needing low and consistent latency
- Ability to update models (workflows)
- Replicate deployments across geographical regions
- Controlled from a central location



Model & Visualize

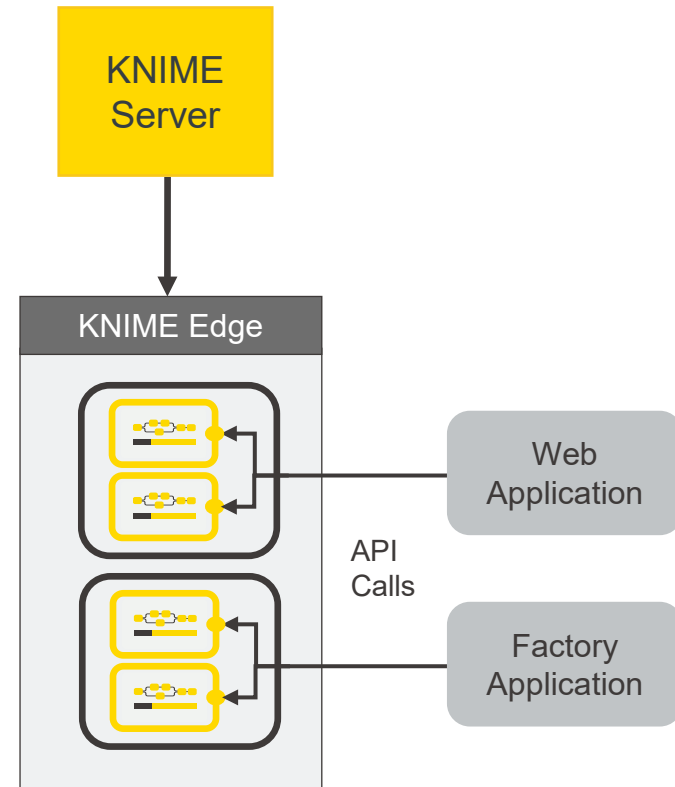


Deploy & Manage



Inference on Demand – KNIME Edge

- New extension to KNIME Server
 - **Model operations:** productionizing model deployment and monitoring
 - **Inference workflows as a Service**
 - Rolling updates with no down time
 - **High request throughput capability**
 - Low latency
 - Scalable on demand
 - **Decentralized execution platform**
 - KNIME Server as a control plane
 - **Flexible deployment**
 - Scale resources to meet demand
 - Reliable and highly available
 - Load balancing
 - Monitoring and management
- Available from today



Thank you!

simon.schmid@knime.com

