



Open for Innovation

KNIME

Closing the Gap in Data Science

- supporting the complete life cycle -

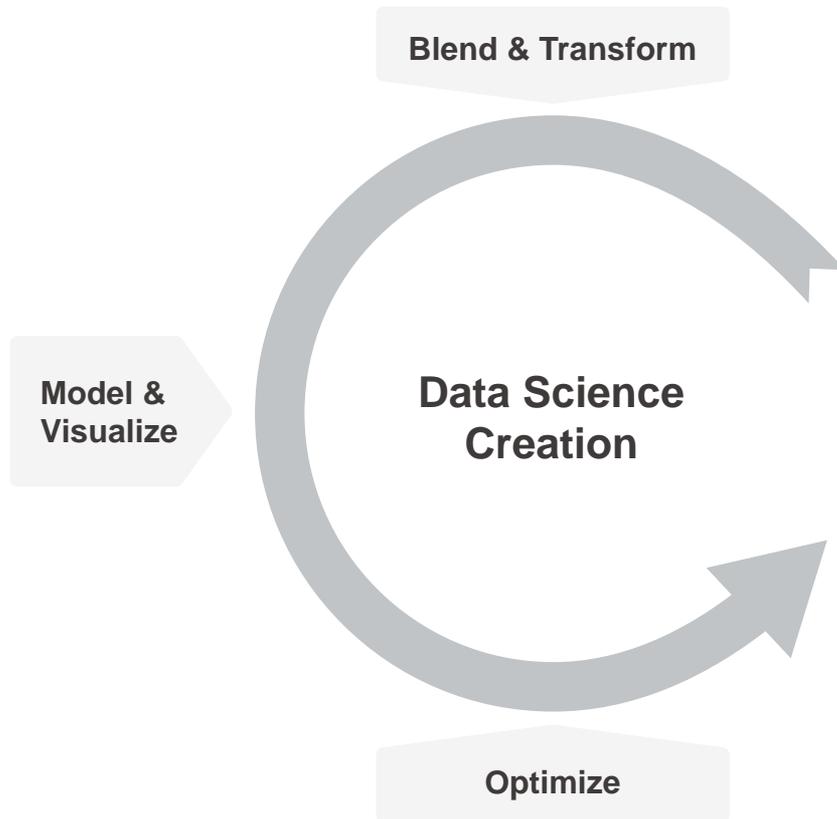
Michael Berthold

(virtual) KNIME Fall Summit 2020

November, 18, 2020



Data Science Is All About Finding the Right Model – Right?



- Data Access
- Data Integration & Transformation
- Exploration (classic “BI”)
- Model Building
 - (Statistical) Summarization
 - Segmentation
 - Prediction Model(s)
- Optimization
 - Improve Performance
 - Adjust ETL part and Model (Choice)

(often referred to as “End-to-End Data Science”)

Productionizing Data Science?

It's not:

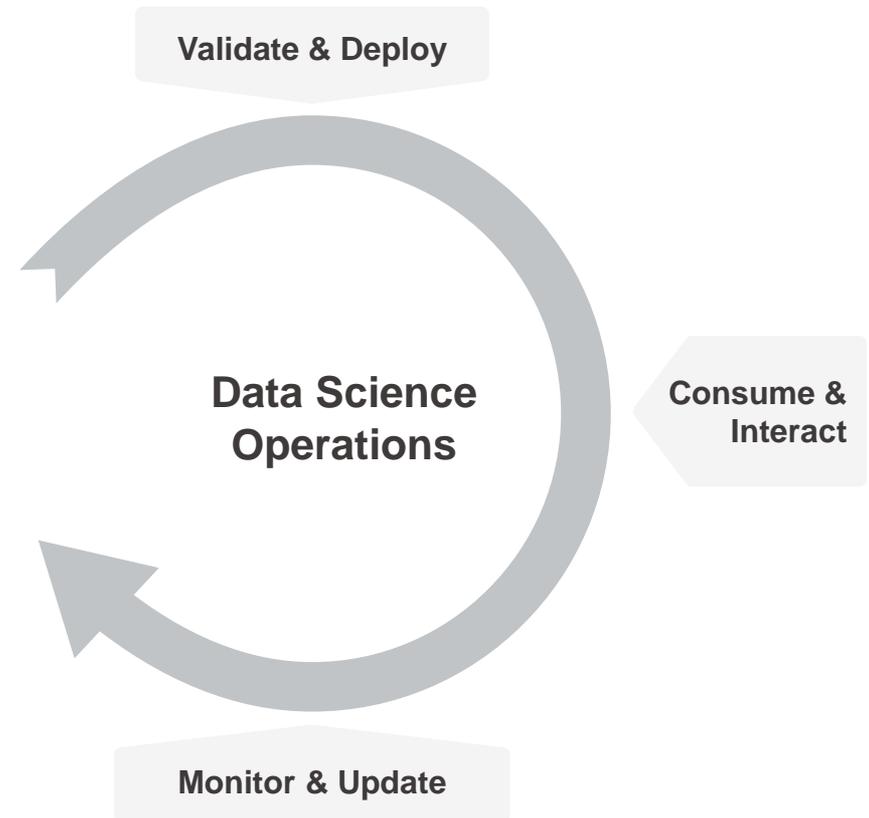
- Creating a financial data aggregate
- Segmenting customer data
- Creating and Optimizing a Churn Prediction Model (automated or not)

It is:

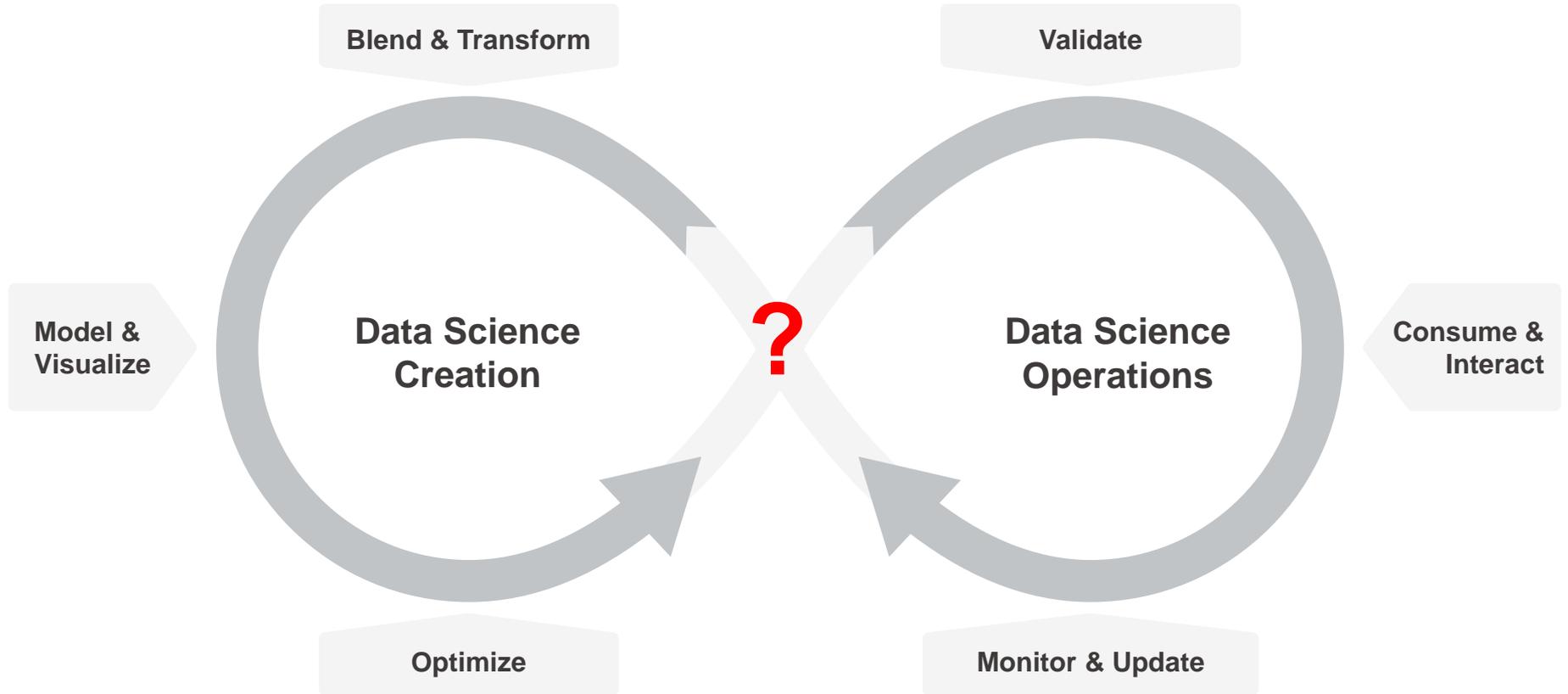
- Automatically send the resulting reports to users
- Put insights inspired by the data segmentation to actual use
- Allowing tools and business users to use the model predictions regularly

Data Science in Production

- **Validate**
 - test (e.g. on past data)
 - governance (privacy and other checks...)(often via staged environments: Dev, Test, Prod)
- **Deploy & Consume**
 - for other tools (“...as a Service”)
 - as (web) interfaces
- **Interact**
 - Incorporate User Feedback
(not just Slice&Dice)
- **Monitor**
 - Detect Failures
 - Identify Changes in Models - or Reality!
- **Update**
 - Adjust to Changes automatically
 - Alert Data Science team...



Creating and Productionizing Data Science

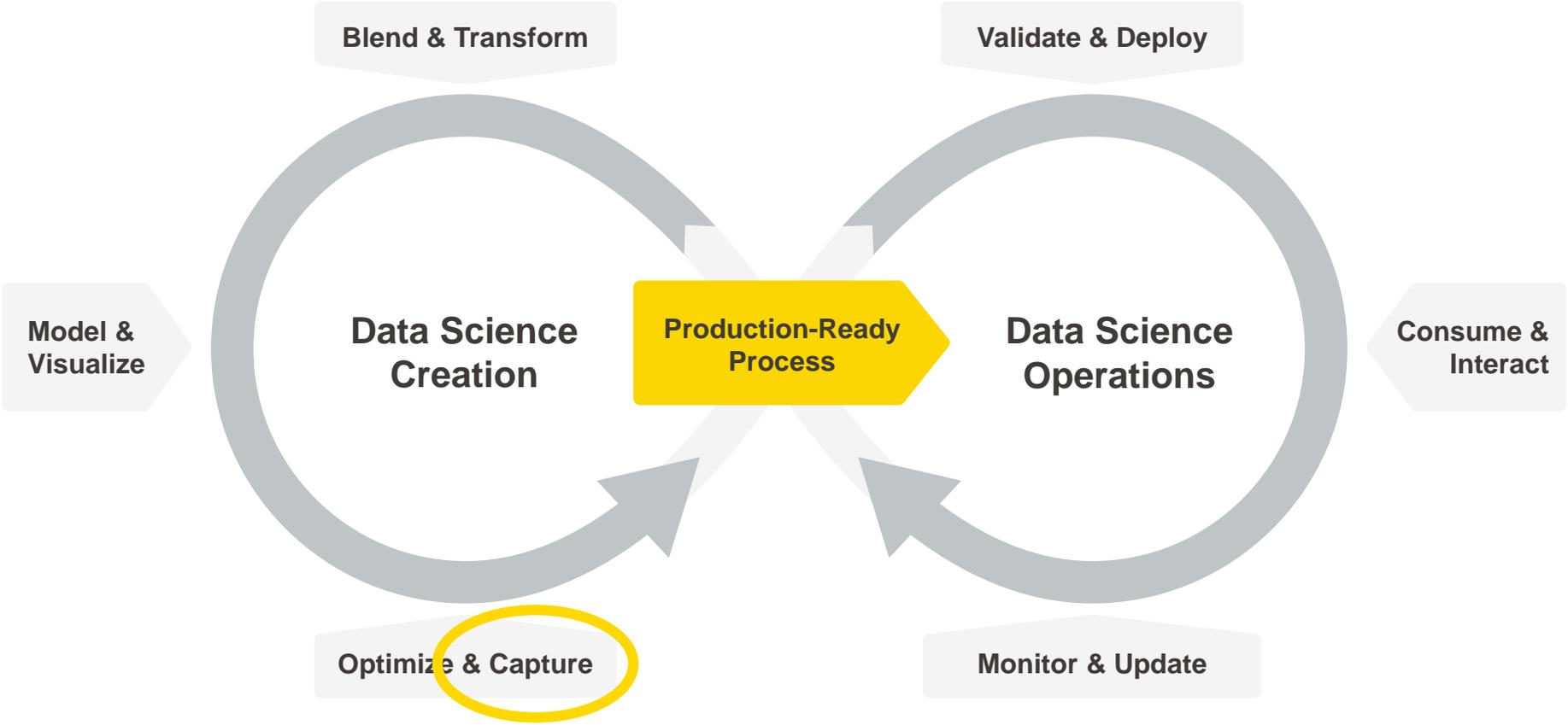




What do we move into Production?

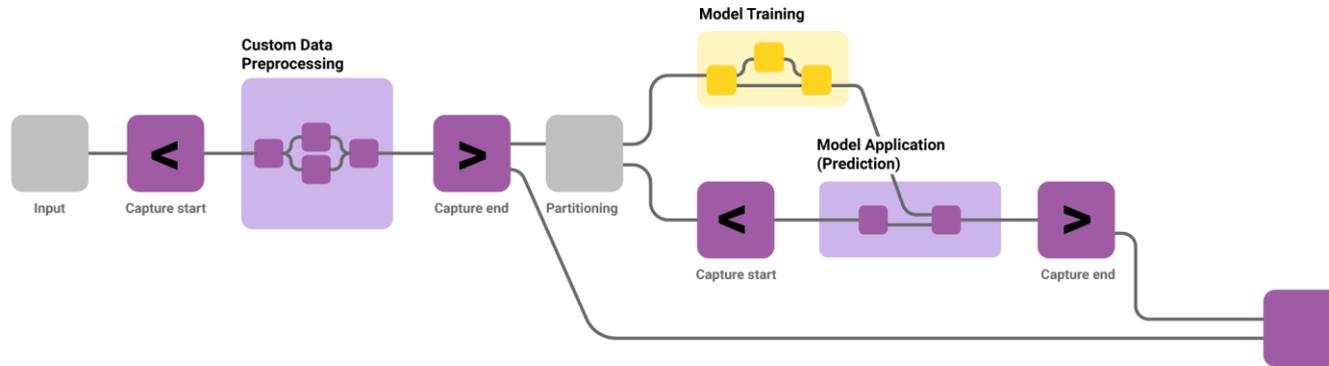
- For a Report – it's the report creating workflow
- For a simple model – it's the library using the trained model
- For serious Data Science – it's an entire process...

Creating and Productionizing Data Science

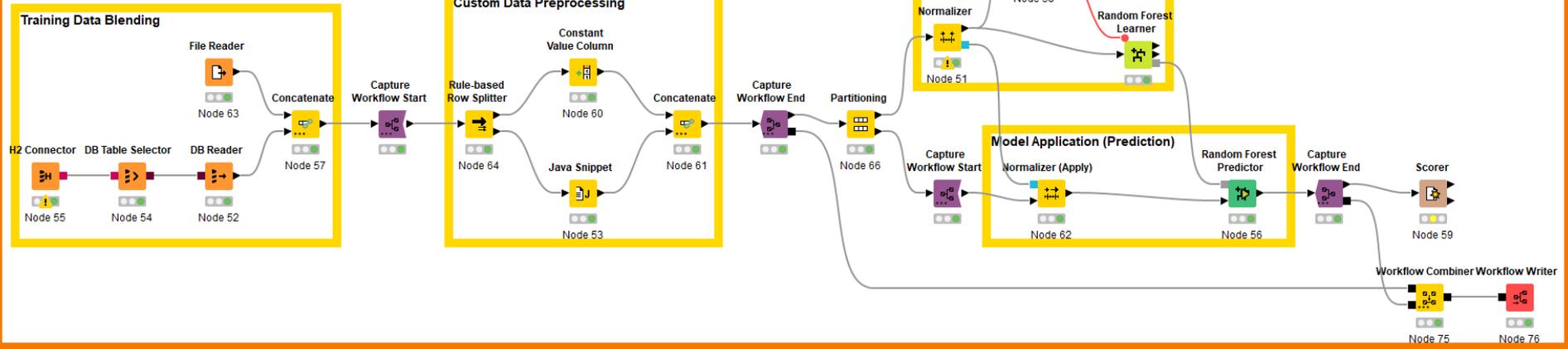


Minding the Gap the KNIME Way: Integrated Deployment

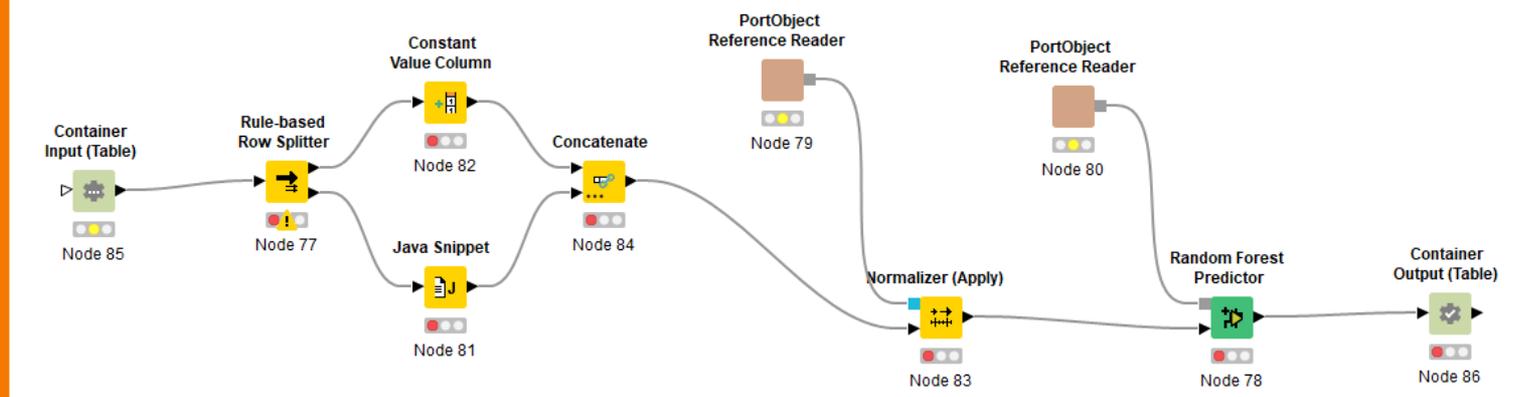
Creating Prediction Model



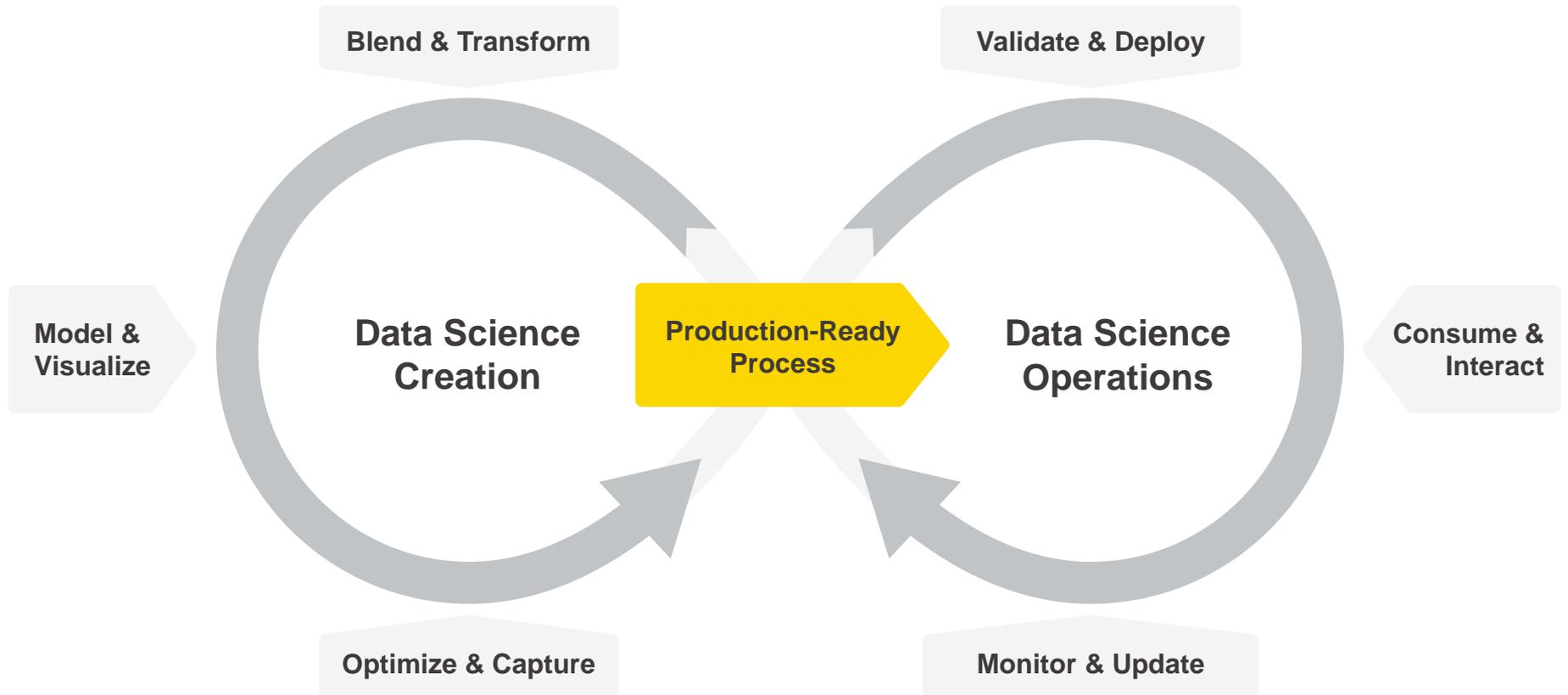
Creating Productionizable Data Science



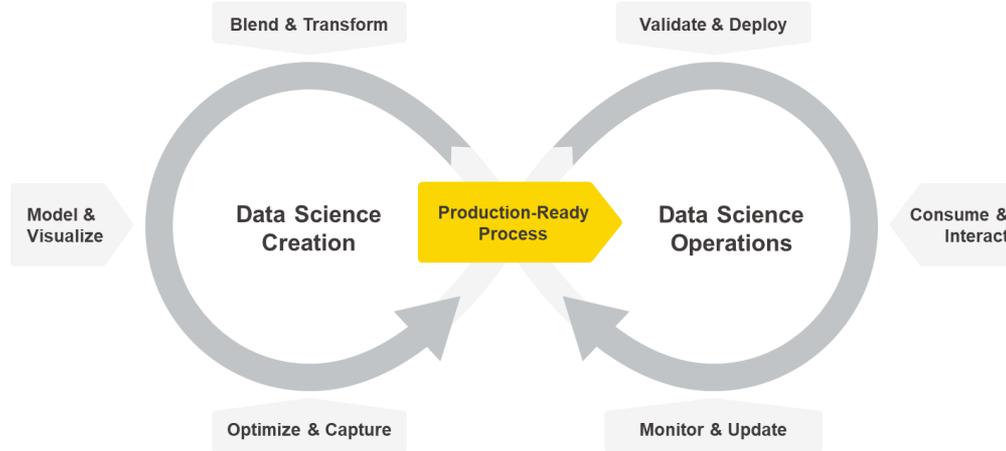
Automatically Created Production Workflow



Creating and Productionizing Data Science



Data Science in Production



Data Science Practices increasingly “grown up”:

- Collaboration: IT, Data Engineer, ML/AI Scientist, Domain Expert.
- Governance
- Reproducibility
- Continuous Delivery of new/updated Data Science Applications & API Services

KNIME Software – One Ecosystem

