



Forest Grove
Data-Driven Decision Making

Optimising technology to
unlock insights that drive
better business decisions

Forest Grove Customer Conference 2023

www.forestgt.com.au | info@forestgt.com.au | 1300 559 827



Forest Grove

Data-Driven Decision Making

Optimising technology to
unlock insights that drive
better business decisions

[D1S3.2] KNIME functionality & value - today & tomorrow by

Presented by:

Vincenzo Tursi - Partner Director, KNIME

About the session:

Vincenzo's session will showcase some recent enhancements deployed with KNIME Developer and the latest KNIME Deployment Server technology, KNIME Business Hub.

He will also give us a window into what we can expect to see in KNIME technology over the next 12 months. KNIME is an international leader in data connectivity, data transformation and data science. As part of Vincenzo's presentation, he will provide examples of how KNIME technology is being deployed worldwide as a foundation technology to build data excellence across the business enterprise. The data partnerships Forest Grove has built locally with our customers have been enhanced considerably with the value provided by applying KNIME technology. Vincenzo's session will help communicate additional areas of data optimisation that will help drive business success.



Open for Innovation

KNIME

KNIME Functionality & Value

2023 CUSTOMER CONFERENCE
Forest Grove Technology

Vincenzo Tursi
Partner Director @ KNIME

February 19, 2023



Forest Grove – Our Partner & Friends from Australia

- Collaboration +8 years
- Delivered high quality work across public and private sector for clients utilizing KNIME Software
- Contributed to the broader open-source data science community, especially in the Asia-Pacific Region
- Our top customer success partner with many happy customers over years..



WE
LIFE WITHOUT BARRIERS
VE



Market Trends

“Smart workflows and seamless interactions among humans and machines will likely be as standard as the corporate balance sheet”

McKinsey’s prediction for 2025

26% CAGR 2021-2030

Data Science Platform Market

Grand View Research

Through 2025, Gartner estimates that the majority of CDOs will fail to foster the necessary **data literacy** within the workforce to achieve their stated strategic data-driven business goals.

Executives saw less than one-third of the value that they expected from digital transformation initiatives. Worse yet, only 7% believe they can address the skills gap in next 5 years.

McKinsey & Company, 2022

Divergent Approaches...

Building
Analytic Center
of Excellence



SKILLS



TOOLS



Upskilling
Business &
Domain Experts



- Standardize, Reuse
- Productionize, Automate
- Data Security, Governance
- Long Term Data Value

FOCUS

- Agility, Time to Value
- Immediacy of Results
- Data Access & Exploration
- Depreciating Data Value (Data Decay)

Analytics for the Entire Enterprise

Agility

Timely access to relevant insights

- Empower business with no-code analytics
- Build a data-driven organization with apps & services
- Leverage a library of pre-built solutions to get started quickly



Operationalization

Enterprise-scale support

- Centrally administer access for data security and privacy
- Support model testing, validation, deployment & monitoring
- Enable use & reuse across the enterprise

Result: Empower anyone in the organization to learn & collaborate on data science.

Support for All Data Users

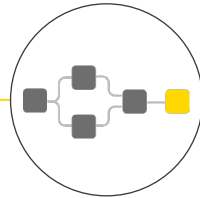
Front-line Business Users



Data apps for instant insights

- Gain analytic insight without need to build any models
- Inject domain expertise into advanced analytical models.
- Use for data input, dashboards, data exploration, IT monitoring, and more.

Data & Business Analysts



Visual workflows for deep analyses

- Perform any kind of analysis, from ETL to ML, without any coding
- Get fast access to data & insights, without IT dependence.
- Leverage library of pre-built workflows for faster start.

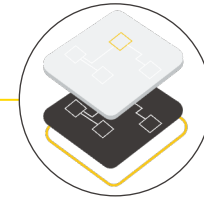
Data Scientists/ Coding Experts



Extensions for advanced analysis

- Access all popular AI/ML libraries through open ecosystem
- Script in a language of your choice through pre-built integrations.
- Create custom KNIME nodes, share & re-use

MLOps & IT

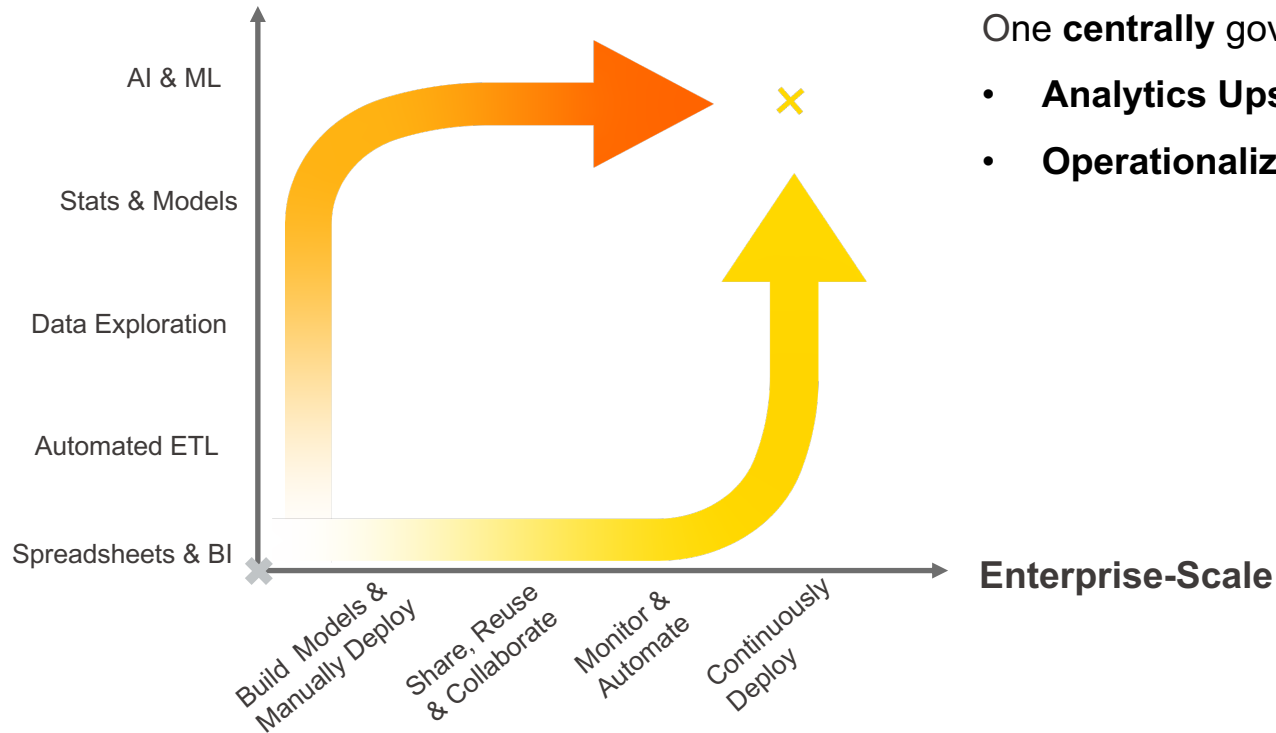


Secure deployment & centralized governance

- Test, validate, deploy and monitor models in one centralized platform
- Centrally administer access for data security & governance
- Deploy on prem or in the cloud

KNIME: Bringing People & Process Together

Analytical Depth



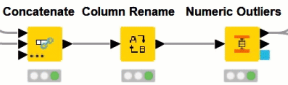
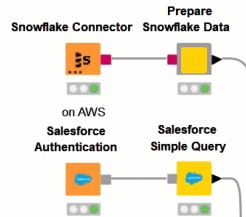
One **centrally** governed platform for:

- **Analytics Upskilling**
- **Operationalization at Scale**

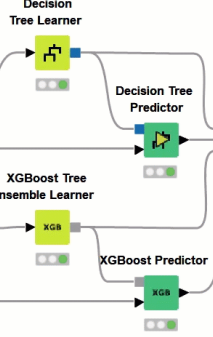
A Complete Analytics Platform for All Data Workers



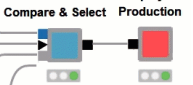
Blend & Transform



Model & Visualize



Deploy & Manage



300+ Data Sources

REST, Edge

Data Apps

BI & Report

Advanced Methods

- Automated Consolidation
- Advanced data Manipulation
- Text, Image Voice

AI/ML Libraries

Scripts/ Languages

Distributed & Cloud Execution

Intuitive Visual Programming for Data Science

The screenshot displays the KNIME software interface with a workflow titled "Demo Workflow". The workflow consists of four nodes: "File Reader" (Node 1), "Excel Reader (XLS)" (Node 2), "Joiner" (Node 3), and "Number To String" (Node 4). Node 1 and Node 2 are connected to Node 3, which is then connected to Node 4.

The "Joiner" node description is as follows:

Joiner

This node joins two tables in a database-like way. The join is based on the joining columns of both tables.

Dialog Options

Joiner settings

Join mode

If a row from the top table cannot be joined with a row from the bottom table (and vice versa) there are several options of handling this situation. After an **inner Join** only matching rows will show up in the output table. A **Left Outer Join** will fill up the columns that come from the bottom table with missing values if no matching row exists in the bottom table. Likewise, a **Right Outer Join** will fill up the columns from the top table with missing values if no matching row in the bottom table exists. A **Full Outer Join** will fill up columns from both the top and bottom table with missing values if a row cannot be joined.

Joining columns

Select the columns from the top input (left table) and the bottom input (right table) that should be used for joining. You may select a real column or the table's row ID here. You must make sure, that the type of selected columns matches. The row IDs are interpreted as StringCells.

Match all of the following: A row of the top input table and a row of the bottom input table match if they match in all specified column pairs.

Match any of the following: A row of the top

The interface also shows a "Workflow Coach" with a list of recommended nodes, a "Node Repository" with a search for "number to", and a "Console" window at the bottom right.

Upskilling beyond Platform



A 300k+ strong
community of
KNIME experts



Online, self-paced
and in-class
learning plus a
Global Partner
Network



Books,
cheat-sheets,
and sample
workflows to get
you started



Pre-built nodes,
components &
extensions to
jump-start
modeling



User groups
and Data
Connects

Across industries and use cases

Verticals such as

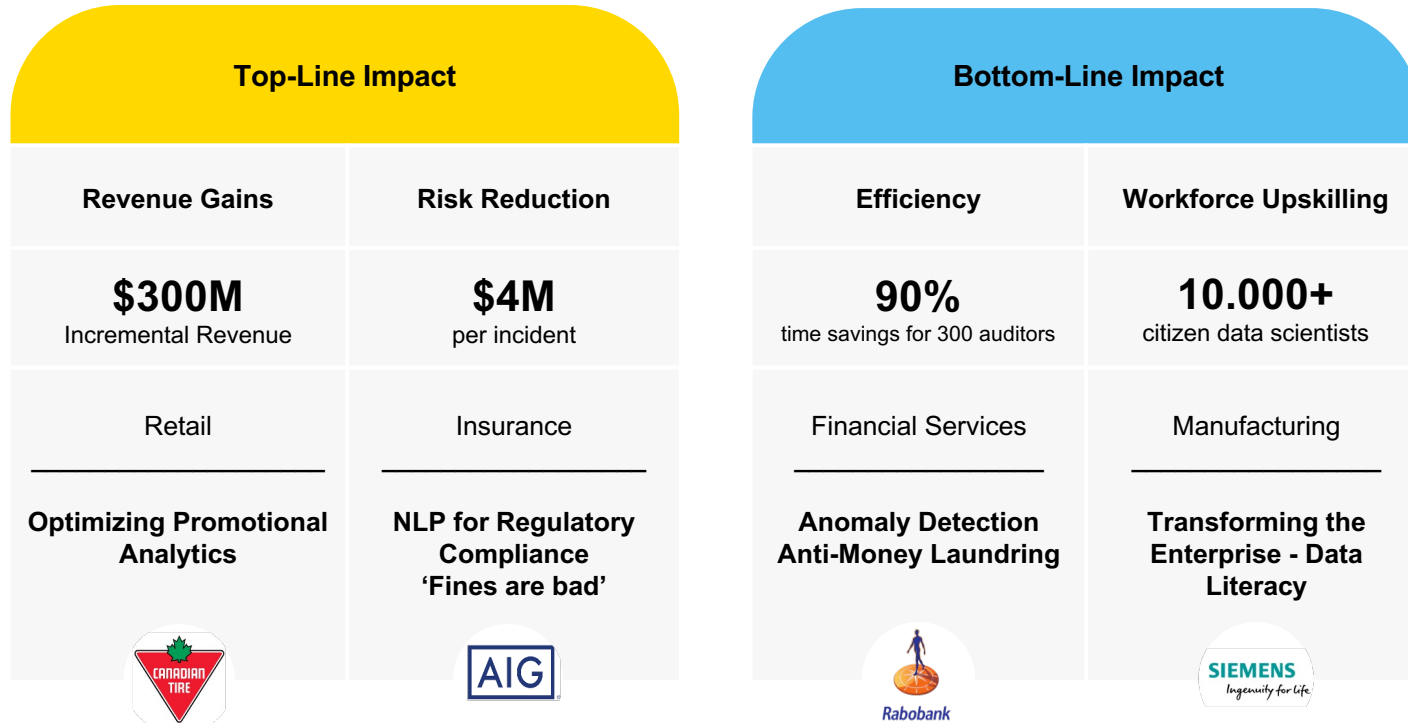
- Financial Services – e.g. Banking, Insurance
- Life Sciences – e.g. Chemistry, Bioinformatics
- Manufacturing – e.g. High Tech, Auto
- Telecommunications
- Government
- Energy
- Retail/CPG
- Etc.

Use Cases

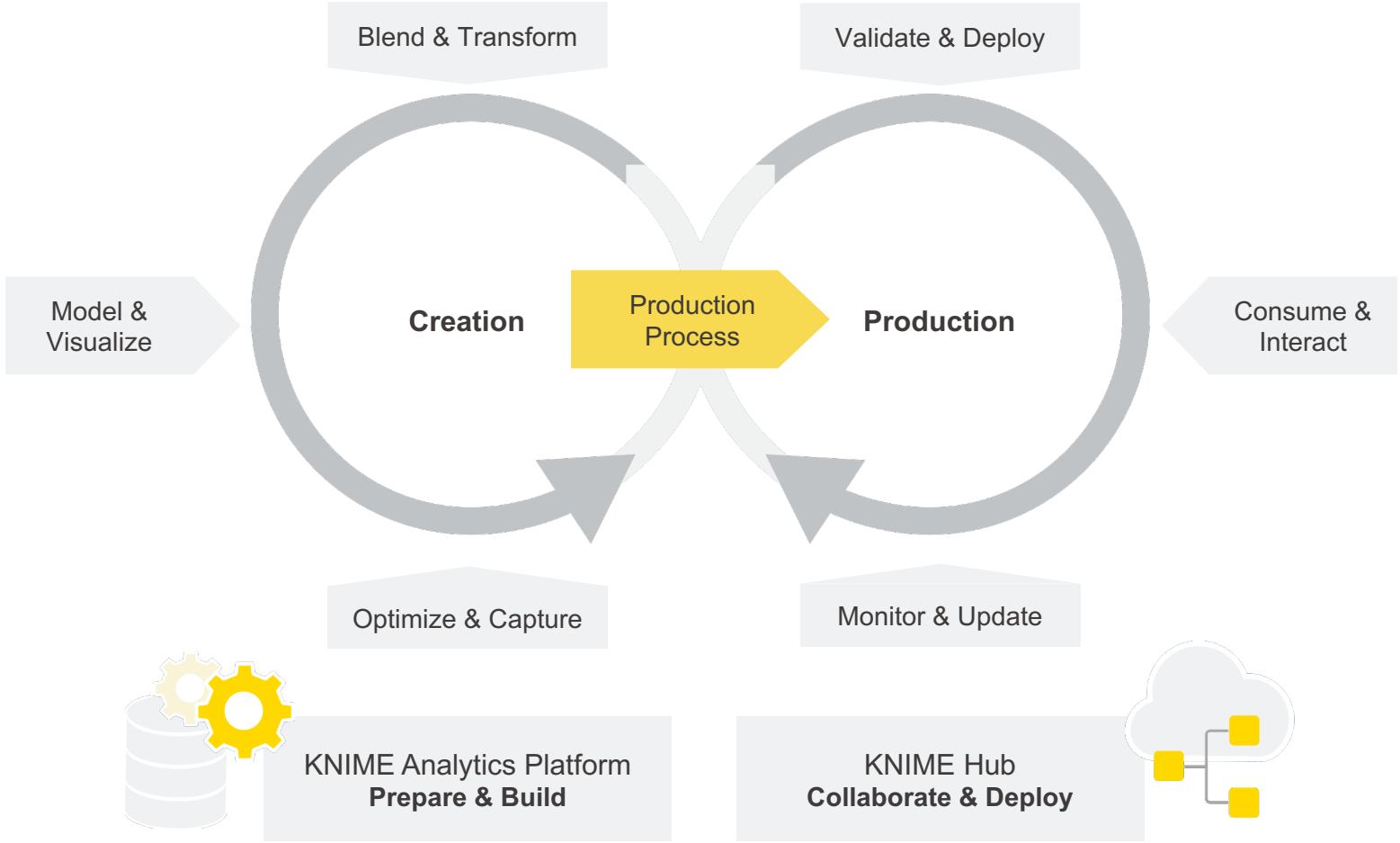
- 'Citizen Data Scientist' enablement
- Spreadsheet and ETL automation
- FP&A and Audit
- Fraud Prevention
- Predictive Maintenance
- Quality Control
- Sales & Marketing
- And many many more...



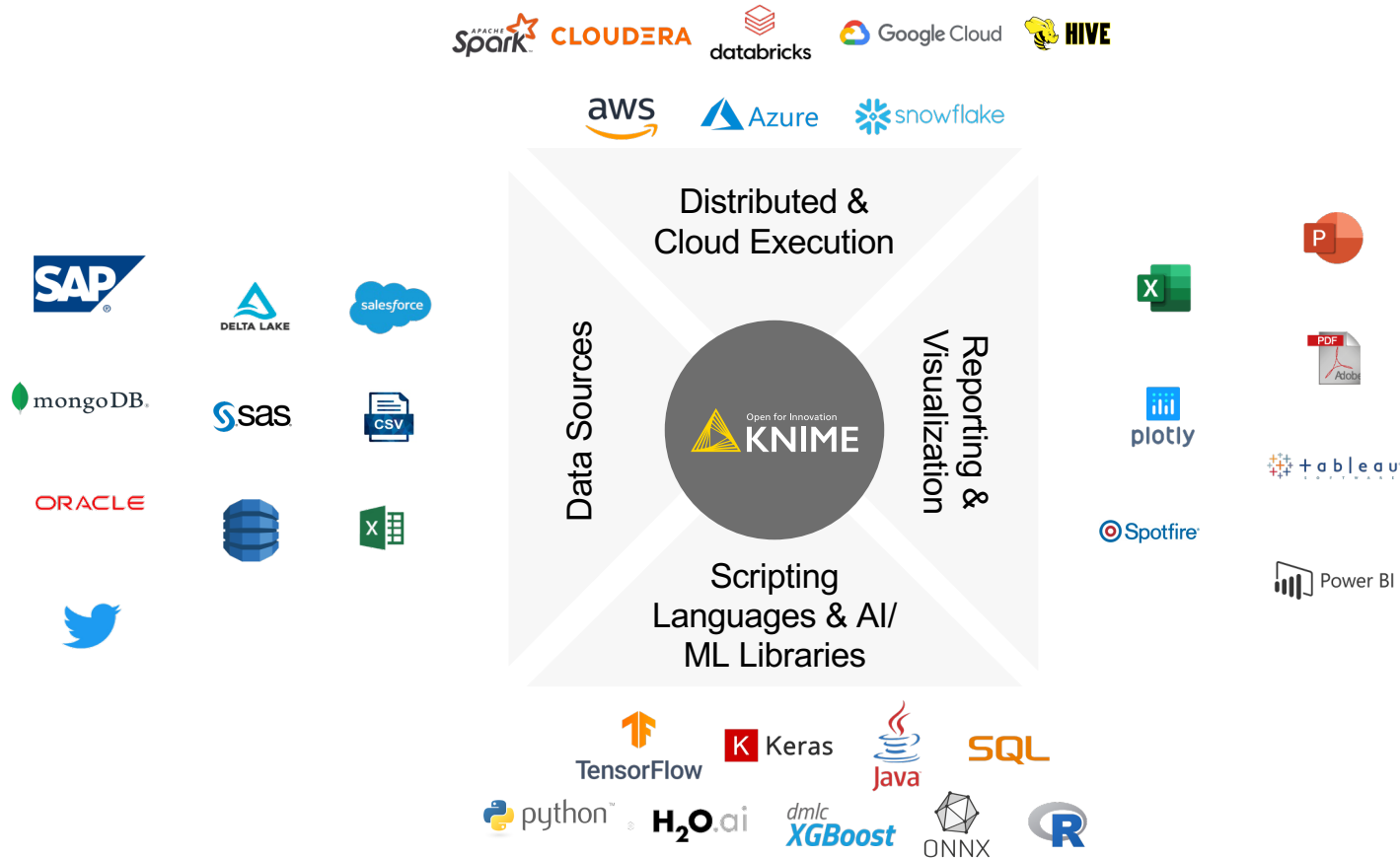
Driving Business Outcomes Across Industries



The Data Science Life Cycle & KNIME



KNIME Technology Ecosystem – Mix and Match Approach



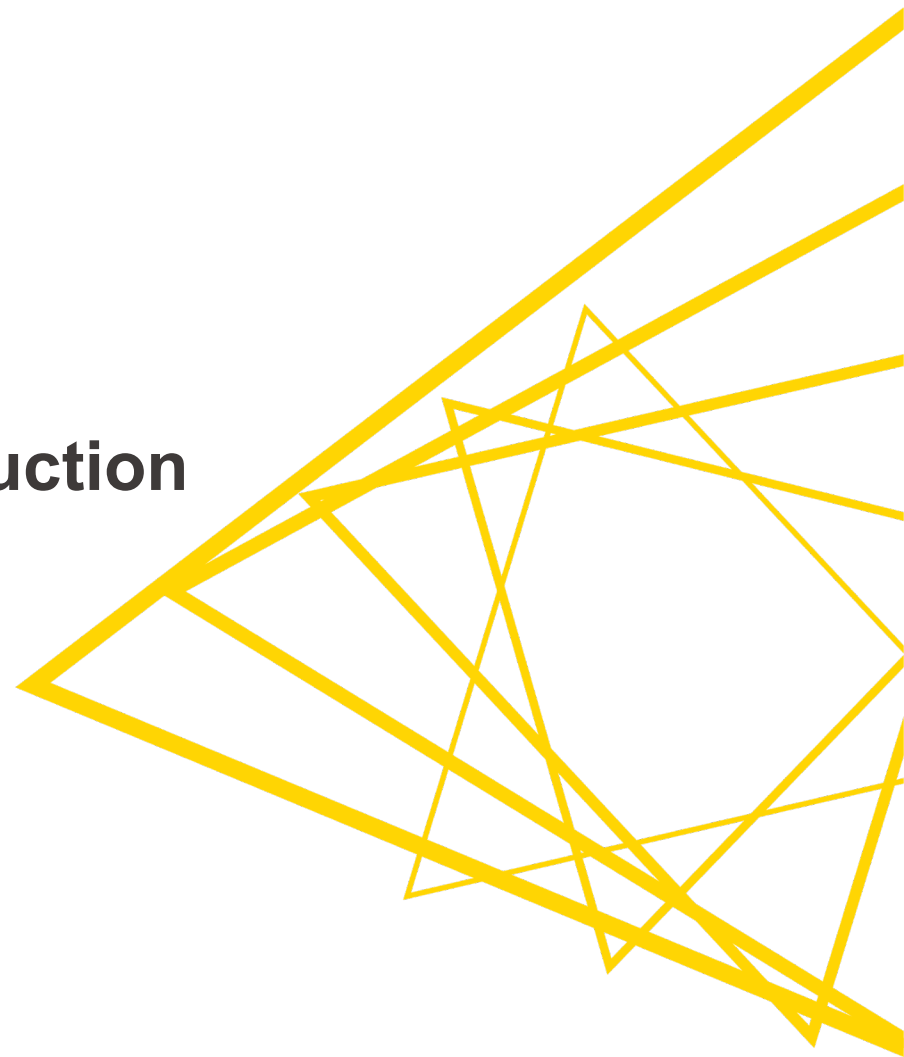


Open for Innovation

KNIME

KNIME Business Hub Introduction

Scale Data Impact in the Enterprise



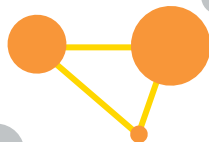


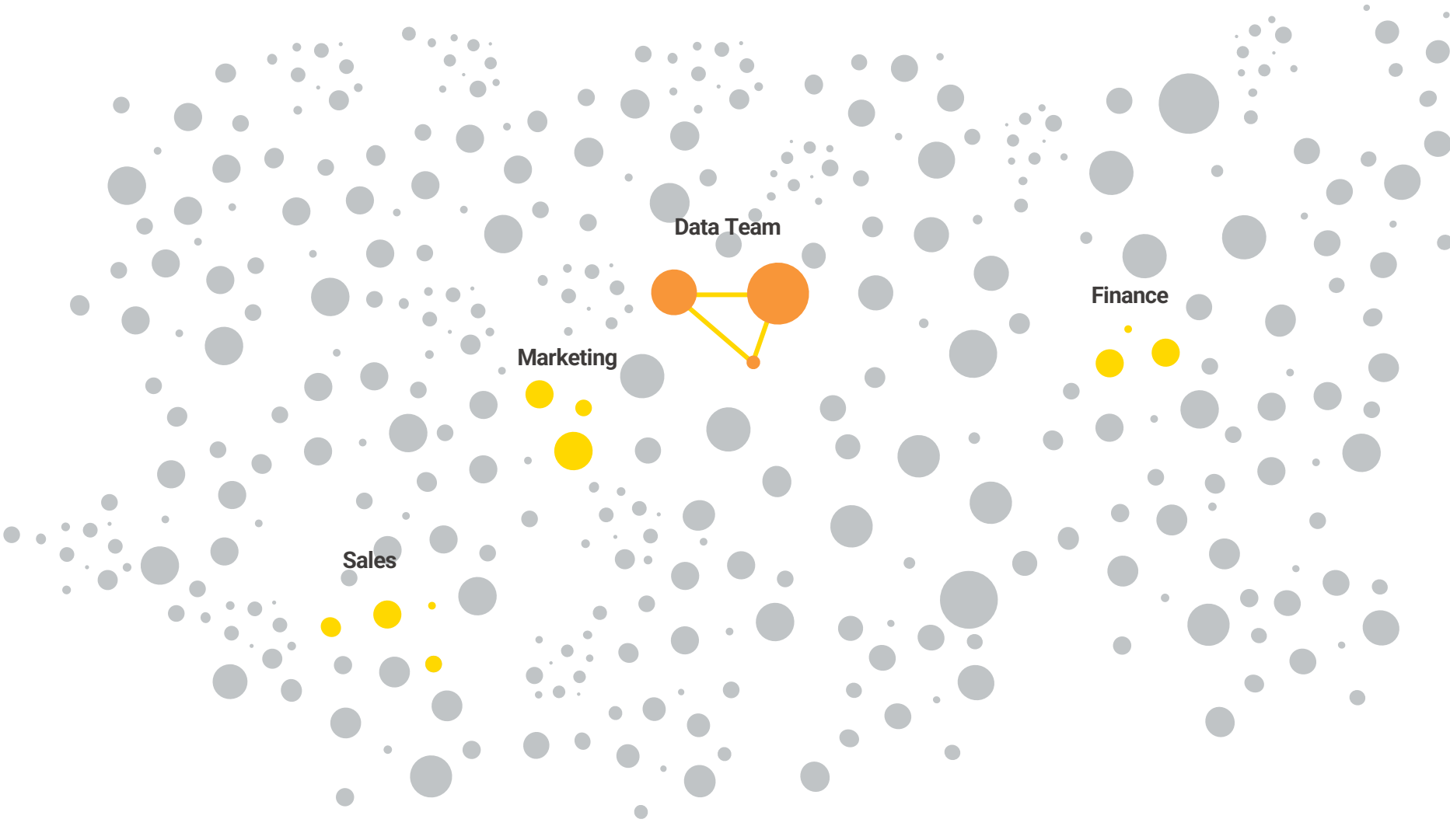
Data Scientist

Data Wrangler

Data Engineer

Data Team



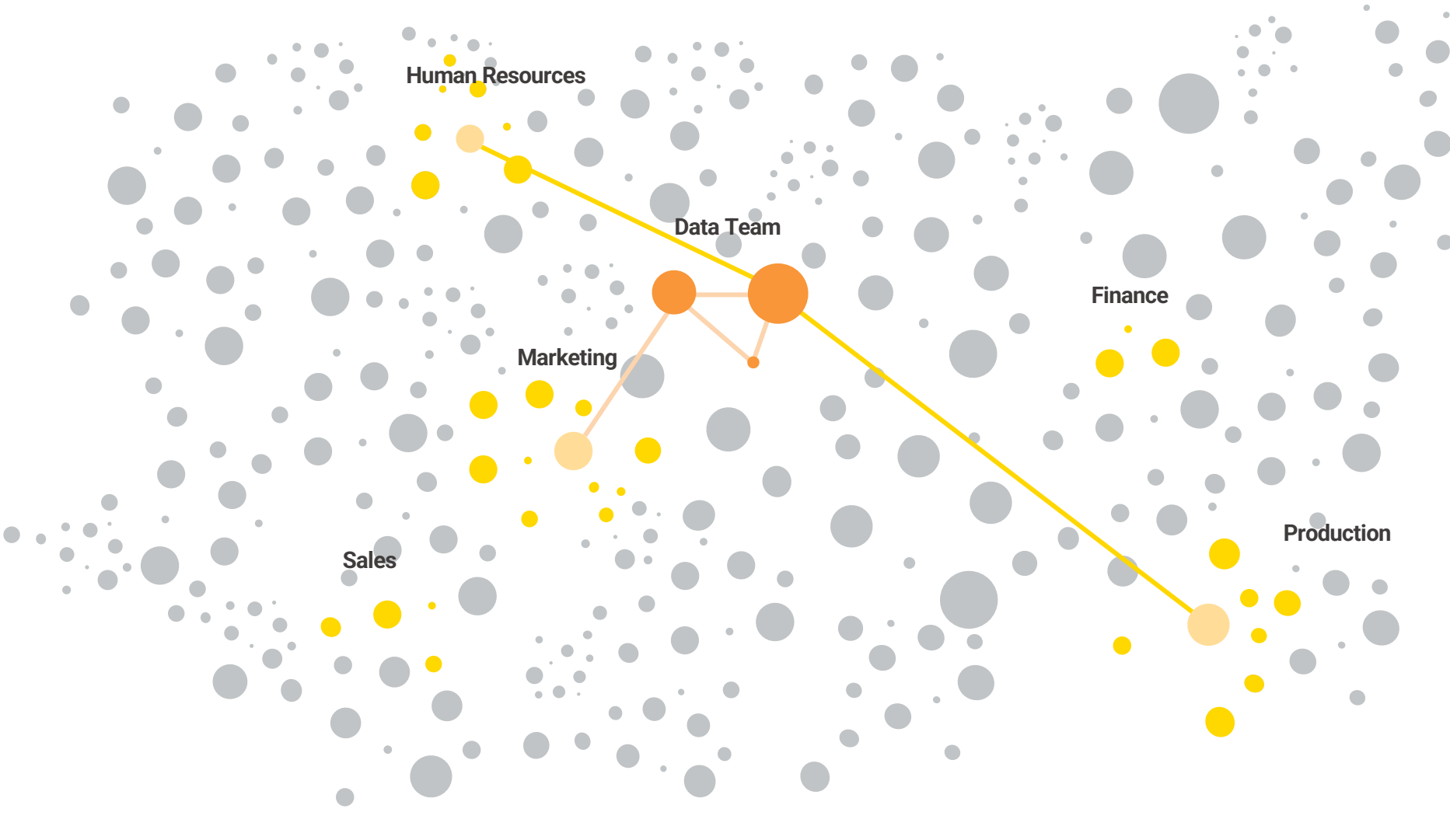


Sales

Marketing

Data Team

Finance



Human Resources

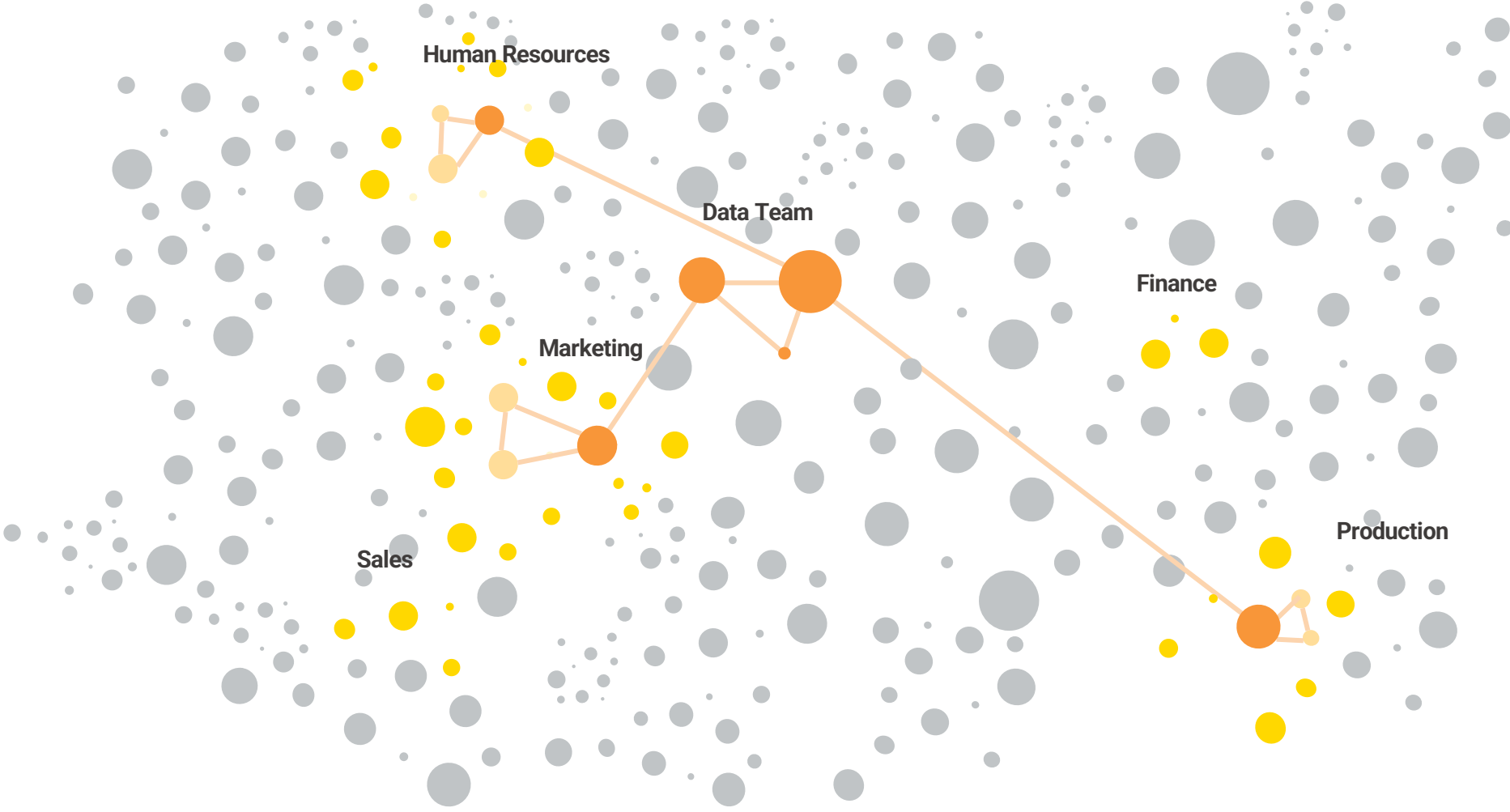
Data Team

Marketing

Sales

Finance

Production



Human Resources

Data Team

Marketing

Sales

Finance

Production

Data Engineer

Data Wrangler

Software Engineer

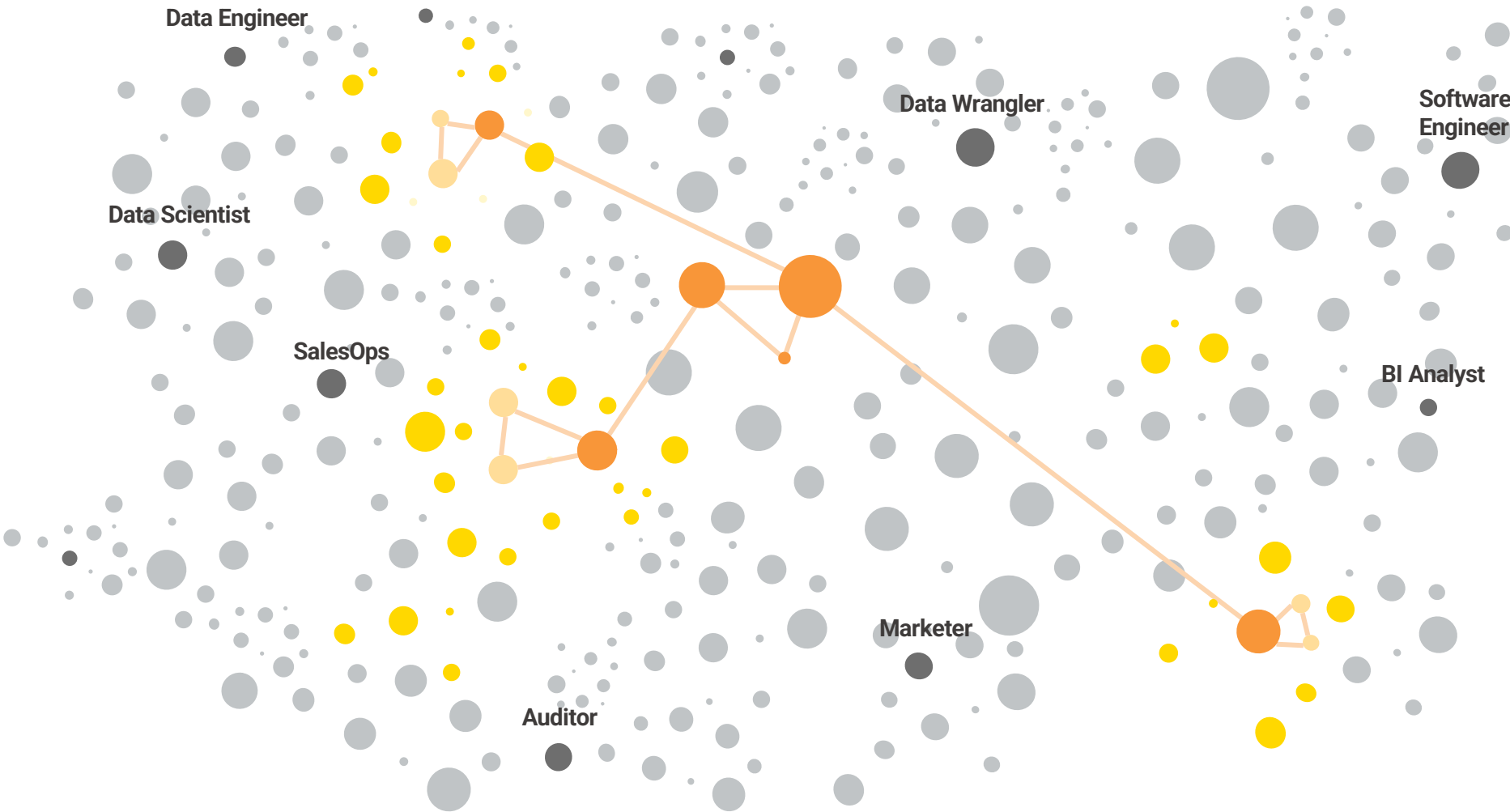
Data Scientist

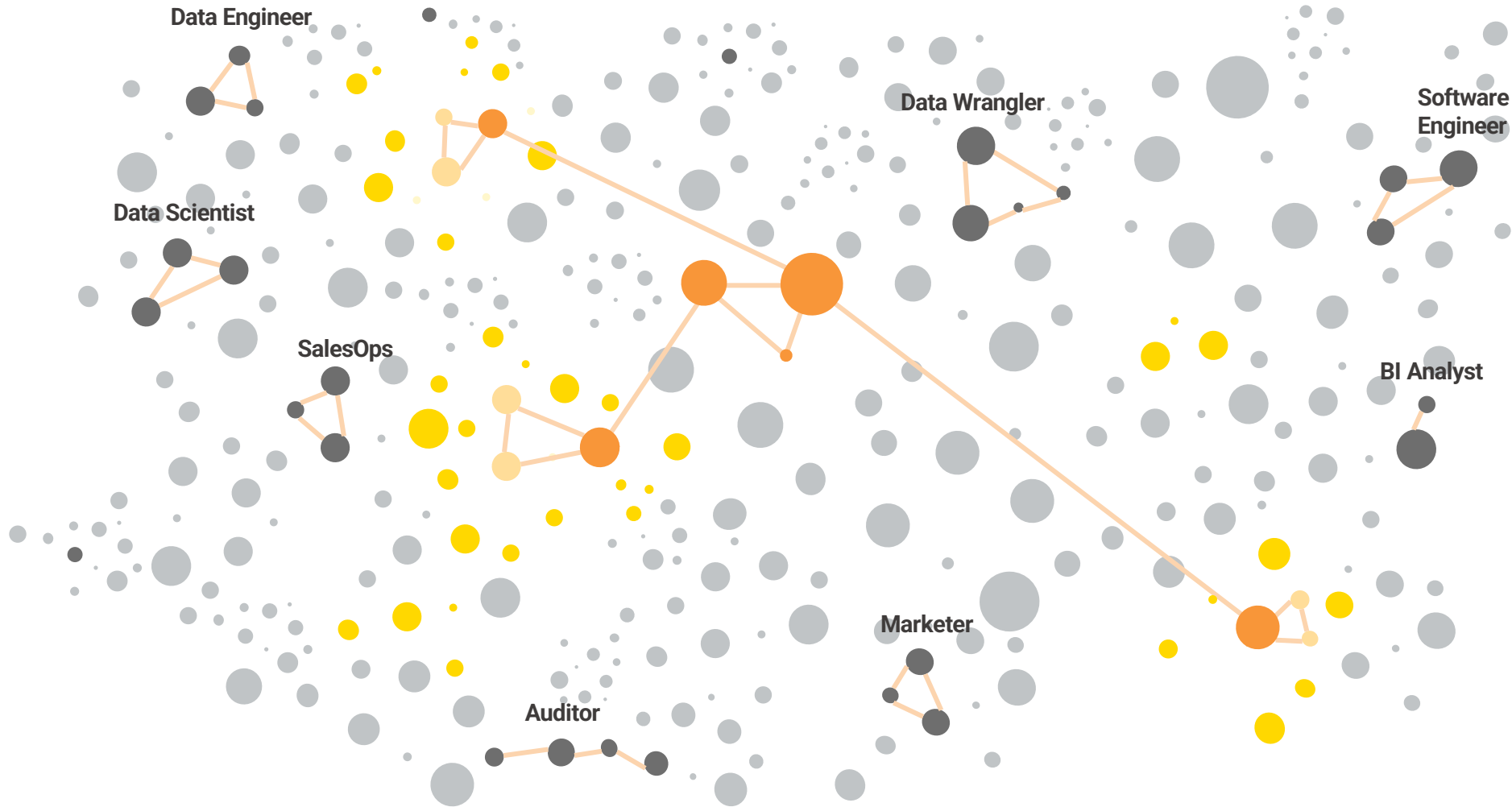
BI Analyst

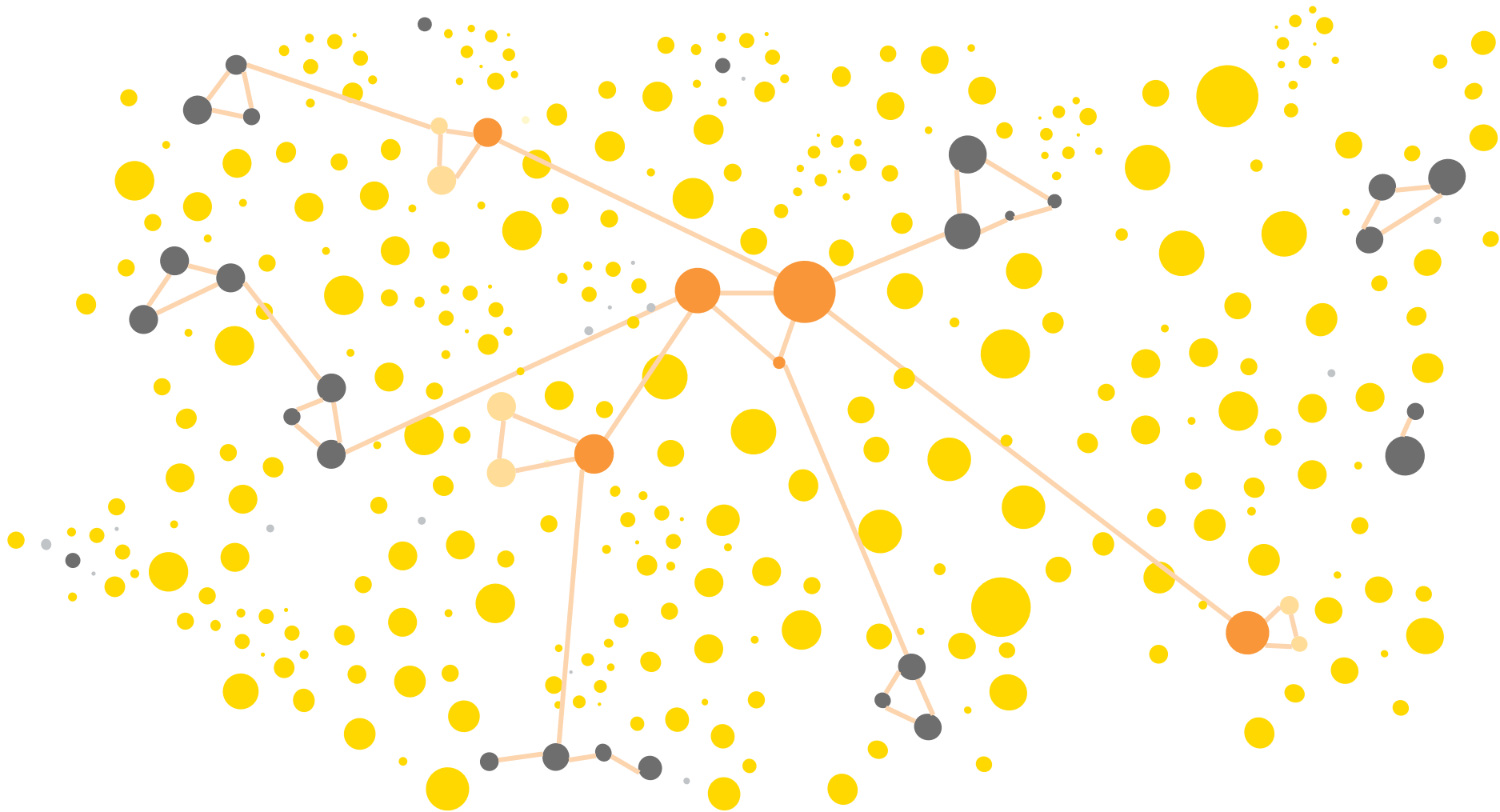
SalesOps

Marketer

Auditor







KNIME: A Single Platform for All Data Users

● Data Experts (users)

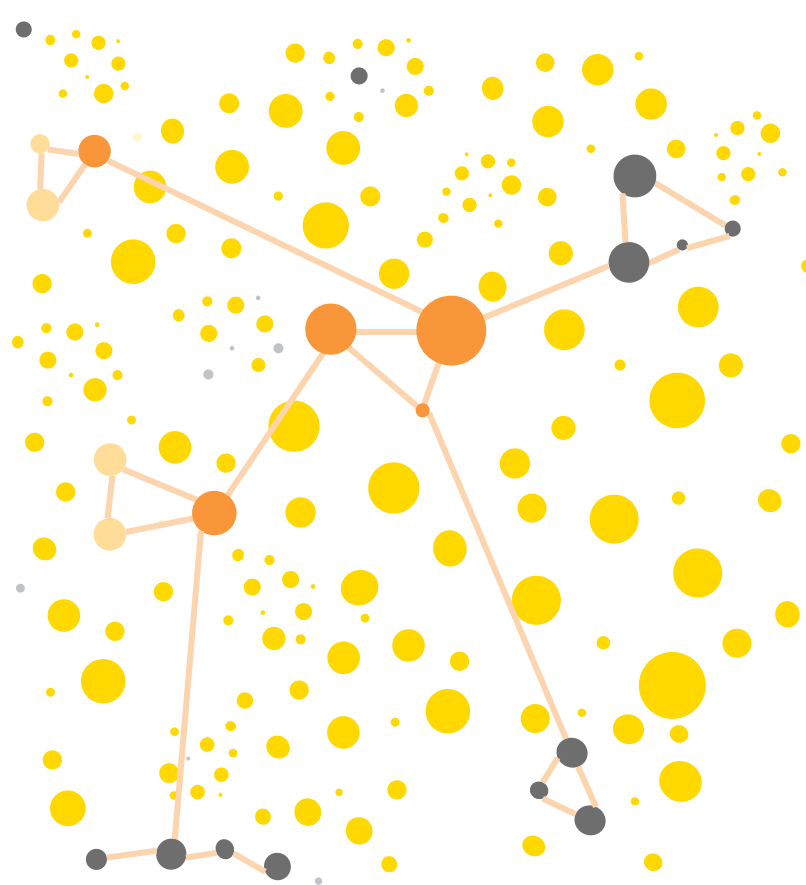
- No-code & low-code workflows
- Create custom nodes for sharing & re-use
- Extensions for advanced analysis

● Business & Domain Experts (users)

- No-code & low-code workflows
- Library of pre-built workflows for fast start
- Fast & instant access to insights, without IT

● End Users (non-users)

- Data apps for insights
- APIs, running behind-the-scenes
- Automated reports



MLOps & IT

- Secure deployment
- Centralized governance

Supporting Every Path to Data-Driven Decisioning

COLLABORATION

1

Collaborate **between skillsets**

2

Knowledge-share **across the organization**

UPSKILLING

3

Upskill **themselves**

4

Upskill **other business teams**

ENTERPRISE-WIDE SCALE

5

Deliver insights **for people**

6

Deliver insights **for systems**

Anticipating Customer Needs

COLLABORATION

1

Collaborate **between skillsets**

2

Knowledge-share **across the organization**

UPSKILLING

3

Upskill **themselves**

4

Upskill **other business teams**

ENTERPRISE-WIDE SCALE

5

Deliver insights **for people**

6

Deliver insights **for systems**

What do we mean “Collaborate between Skillsets”?

- Empowering disparate people work together on the same analytics projects
 - e.g., Data steward & financial analyst
 - e.g., Python coder & visualization specialist
 - e.g., Synthetic chemists and computational chemists

Collaboration Between Skillsets

PAIN POINTS

There's a lot of work just to communicate technical solutions—explaining, presenting, diagramming.



KNIME SOLUTION

Self-documenting, understandable reproducible KNIME workflows.

Feedback loops for analytical solutions is slow, losing value with time.



Sharing workflows via KNIME Business Hub public spaces.

Collaborate Between Skillsets

PAINPOINTS	KNIME SOLUTION
Work siloed by heterogeneous tools and languages.	KNIME workflows are intuitive for all, and serve as lingua franca—while external tools, scripting languages or REST Services can be integrated.
Advanced solutions often inaccessible to business users.	Components allow users to abstract complex parts from non experts - exposing only relevant parameters.

Anticipating Customer Needs

COLLABORATION

1

Collaborate **between**
skillsets

2

Knowledge-share
across the organization

.

UPSKILLING

3

Upskill **themselves**

4

Upskill **other business**
teams

ENTERPRISE-WIDE SCALE

5

Deliver insights **for**
people

6

Deliver insights **for**
systems

What do we mean “Knowledge Share across the Organization”?

- Empowering teams to leverage and build on the experience of others.
 - e.g., enable sharing of blueprints as examples or starting points
 - e.g., enable sharing of reusable building blocks for common tasks
 - e.g., ensure security & quality with approved & tested workflows
 - e.g., innovate more quickly by adapting or getting inspired by solutions from others

Knowledge Share Across the Organization

PAIN POINTS	KNIME SOLUTION
Lack of change tracking and reproducibility creates risk	Reproducible workflows and versioning of spaces on KNIME Business Hub
Many “front end manufacturing teams” needed similar solutions built out	Sharing commonly needed workflows and components via public spaces
Administration of solutions required too much data expert time	KNIME Business Hub for centrally administered, but self-organizing teams

Knowledge Share Across the Organization

PAINPOINTS	KNIME SOLUTION
Lack of knowledge transfer and change management creates risk	Spaces and versioning
Lack of standardization creates risk and duplicate work	Sharing components via public spaces.

Anticipating Customer Needs

COLLABORATION

1

Collaborate **between**
skillsets

2

Knowledge share
across the organization

UPSKILLING

3

Upskill **themselves**

4

Upskill **other business**
teams

ENTERPRISE-WIDE SCALE

5

Deliver insights **for**
people

6

Deliver insights **for**
systems

What do we mean “Upskill Themselves”?

- Self-motivated onboarding to increase personal efficiency, quality or sophistication of insights, or general analytics know-how
 - e.g., move from manual, difficult-to-replicate, error prone spreadsheets to low-code/no-code workflows
 - e.g., work with larger amounts of data at once
 - e.g., learn and apply more sophisticated analytics techniques to data
 - e.g., make results/insights more accessible to a wider group of people

Upskill Themselves

PAIN POINTS

Aggregating data on shift lengths, schedule types and overtime limitations is time consuming

New data means repeat work, copy & pasting between spreadsheets

Manual work in spreadsheets is error prone



KNIME SOLUTION

KNIME connectors allow users to get data out of several tools or spreadsheets at one time

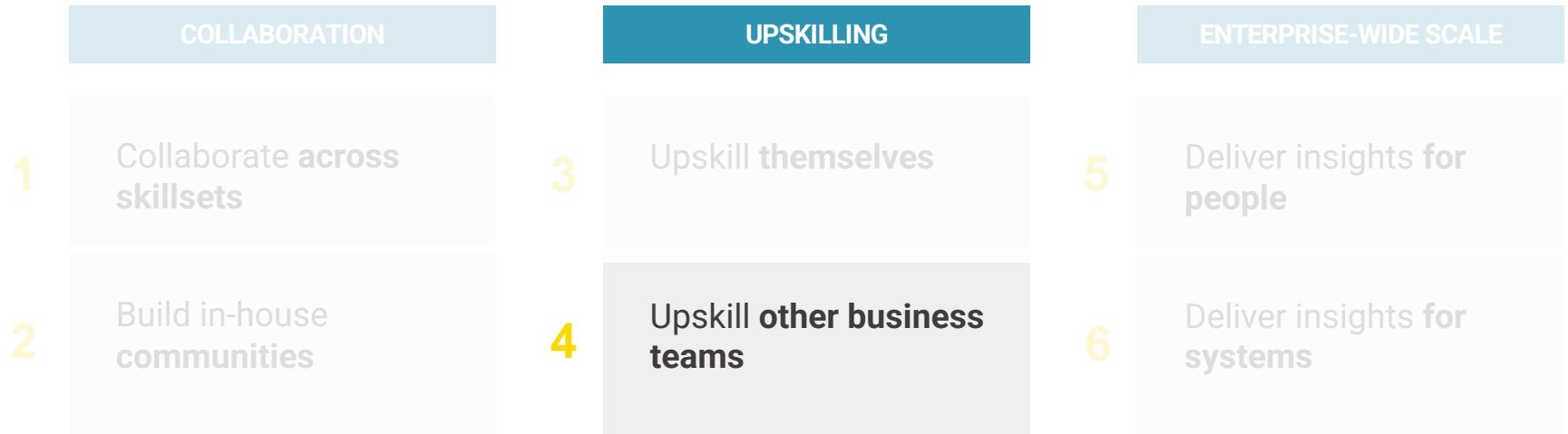
Data can then be analyzed in one click, once a workflow is set up

- Workflows reduce amount of required human intervention and guarantee reproducibility

Upskill Themselves

PAINPOINTS	KNIME SOLUTION
Aggregating data from different tools or spreadsheets is extremely time-consuming	→ 300+ KNIME connectors remove friction from data access
Spreadsheets limit data volume & type for analysis	→ KNIME can handle data sets of any size
Logic of spreadsheets is obfuscated & intransparent	→ Workflows clearly reveal and document analytic process
Most tools lack advanced analytics capabilities	→ KNIME offers more analytic sophistication than any spreadsheet or other LCNC vendor

Anticipating Customer Needs



What do we mean “Upskill Other Business Teams”

- Organization-led initiatives to upskill business teams for increased self-sufficiency
 - e.g., CoE upskilling finance analysts (spreadsheet-users) to automate data tasks
 - e.g., Data team upskilling mechanical engineers in the field of data analytics

Upskill Other Business Teams

PAIN POINTS	KNIME SOLUTION
Many more automation opportunities than data experts	Intuitive low-code/no-code make automation and reporting accessible to all
Different data users needed varying levels of complexity	Range of analytic sophistication available in AP
Difficult to onboard large #s of users around the world efficiently	◦ Sharing workflows and components with AP community on KNIME Business Hub

Upskill Other Business Teams

PAINPOINTS	KNIME SOLUTION
Data team is spending too much time on automation or basic reporting	➔ Intuitive, LCNC platform empowers self-sufficiency to a certain extent
Digitization posing risk to workforce	➔ Range of techniques enable onboarding to advanced analytics work
Specialized, expensive labor wastes time on data wrangling	➔ Automation of data tasks & reproducibility of workflows
Reliance on data experts for advanced analysis, slows time to value	➔ Intuitive low-code/no-code allows for self-sufficiency

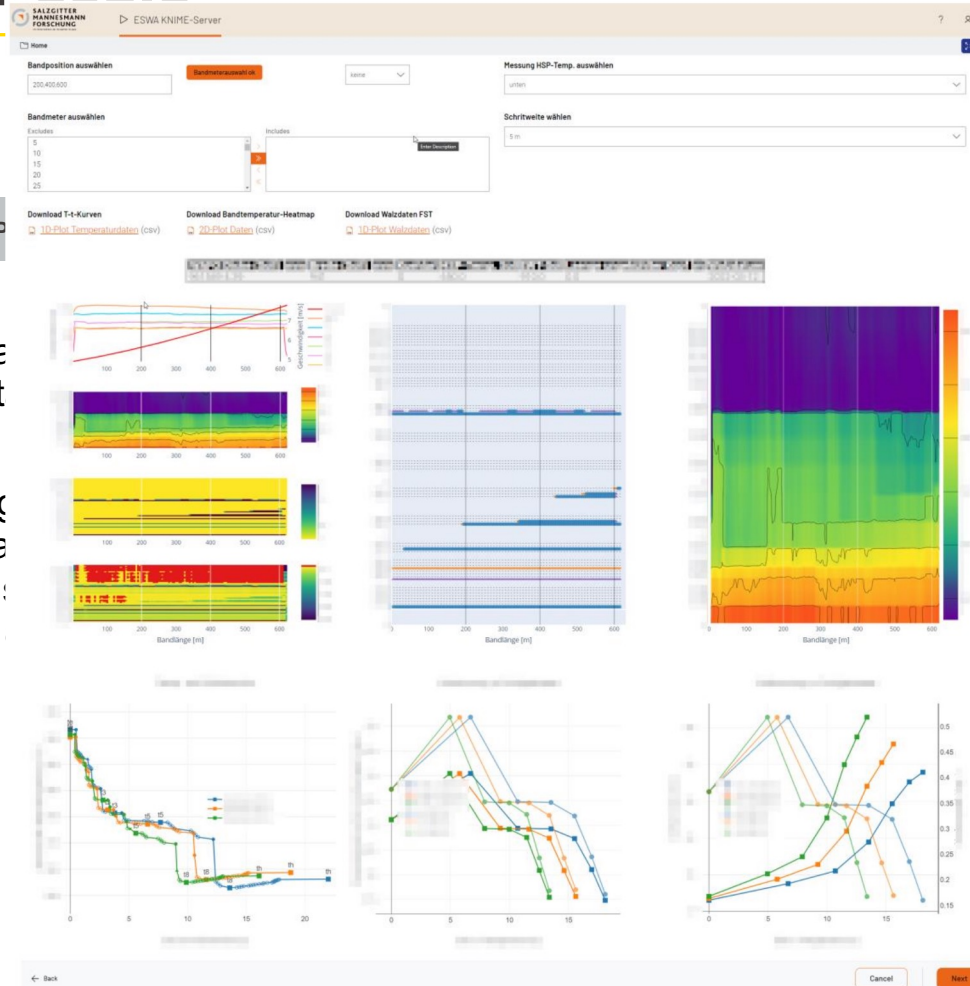
Anticipating Customer Needs



What do we mean “Deliver Insights for People”

- Make analysis available to anyone through dynamic, parameterized data apps
 - e.g., dashboard for sales people monitoring progress towards ARR goal
 - e.g., static report for HR, available on-demand, showing employees hired/month
 - e.g., self-service data app that allows nurses to properly dose medicine, based on inputs with prior medical history

Deliver Insights for People



PAIN P

Too much me
analyzing dat

Research eng
relied too hea
scientists for
& root cause

Deliver Insights for People

PAINPOINTS

Models need significant input from domain (but non-data) experts

Data science team requires data entry into analysis



KNIME SOLUTION

Data apps parameterize analysis of any level of complexity, while never revealing underlying workflow

Forms on data apps allow input from non experts

Anticipating Customer Needs



What do we mean “Deliver Insights for Systems”?

- Make analytics available as REST services that can be consumed by other applications
 - e.g., fraud detection API triggered by a web application to check credit data
 - e.g., API that triggers model to run on machine sensor data and alert on disaster
 - e.g., API to enrich CRM data every evening

Deliver Insights for Systems

PAIN POINTS

If service goes down, other systems are affected and can't operate anymore.

Many requests to service require scalability.



KNIME SOLUTION

High-availability of KNIME Edge



Scalability of KNIME Edge.

Deliver Insights for Systems

PAINPOINTS

Integration with other applications oftentimes requires custom code

Deployment of analytics as REST often requires IT

KNIME SOLUTION



KNIME can deploy workflows as REST interfaces



KNIME automates deployment, removing reliance on software team

Anticipating Customer Needs

COLLABORATION

1

Collaborate **between skillsets**

2

Knowledge-share **across the organization**

UPSKILLING

3

Upskill **themselves**

4

Upskill **other business teams**

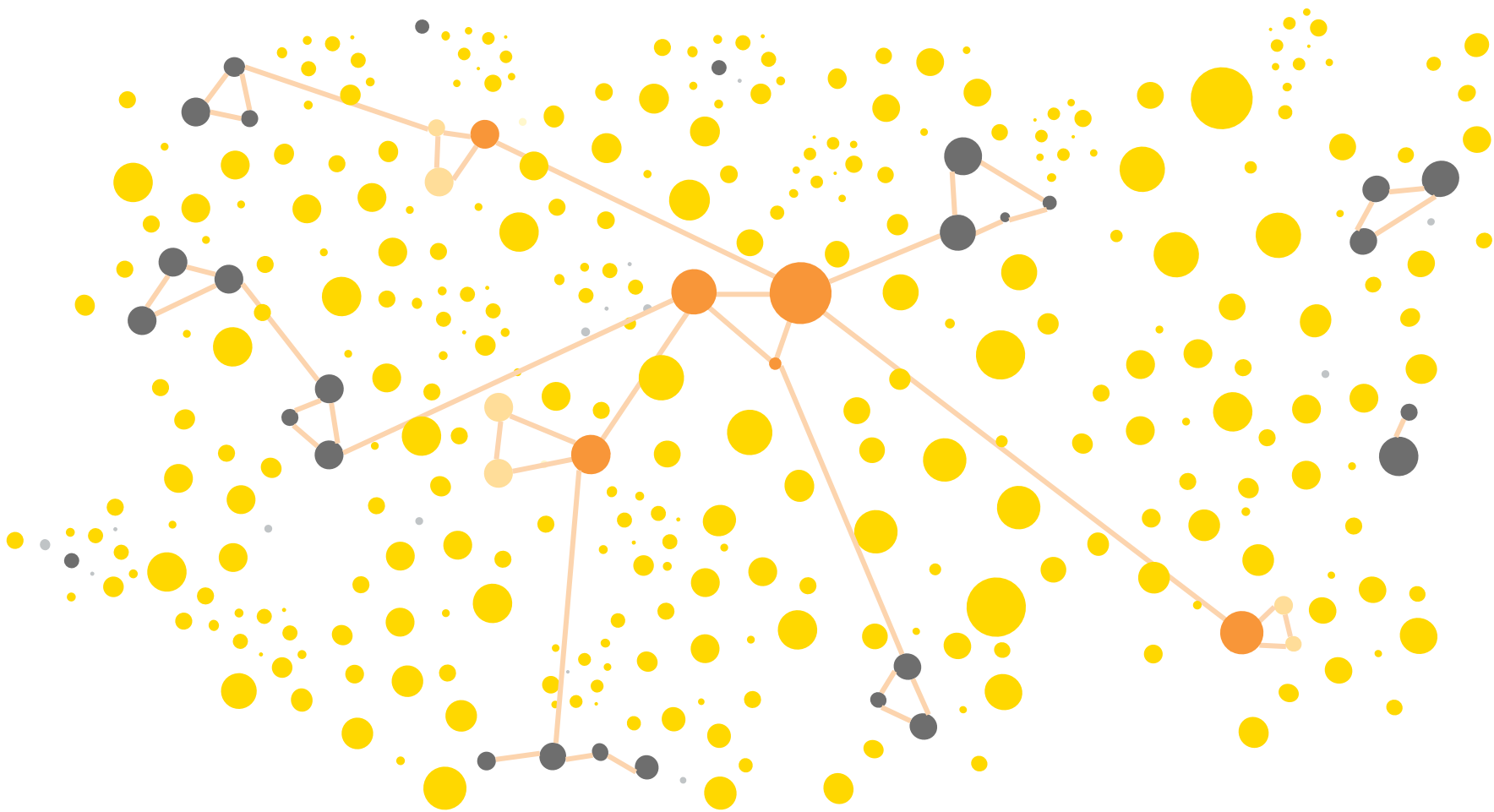
ENTERPRISE-WIDE SCALE

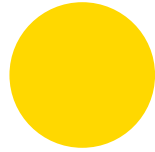
5

Deliver insights **for people**

6

Deliver insights **for systems**







**Research
Institutions**



Manufacturers



Scientists



**Pharma
Companies**



**Data
Scientists**



**Financial
Institutions**



Research Institutions



Scientists



Data Scientists



Manufacturers



Pharma Companies



Financial Institutions





UPPSALA
UNIVERSITET



UNIVERSITY OF
ILLINOIS
URBANA-CHAMPAIGN



GEMMACON



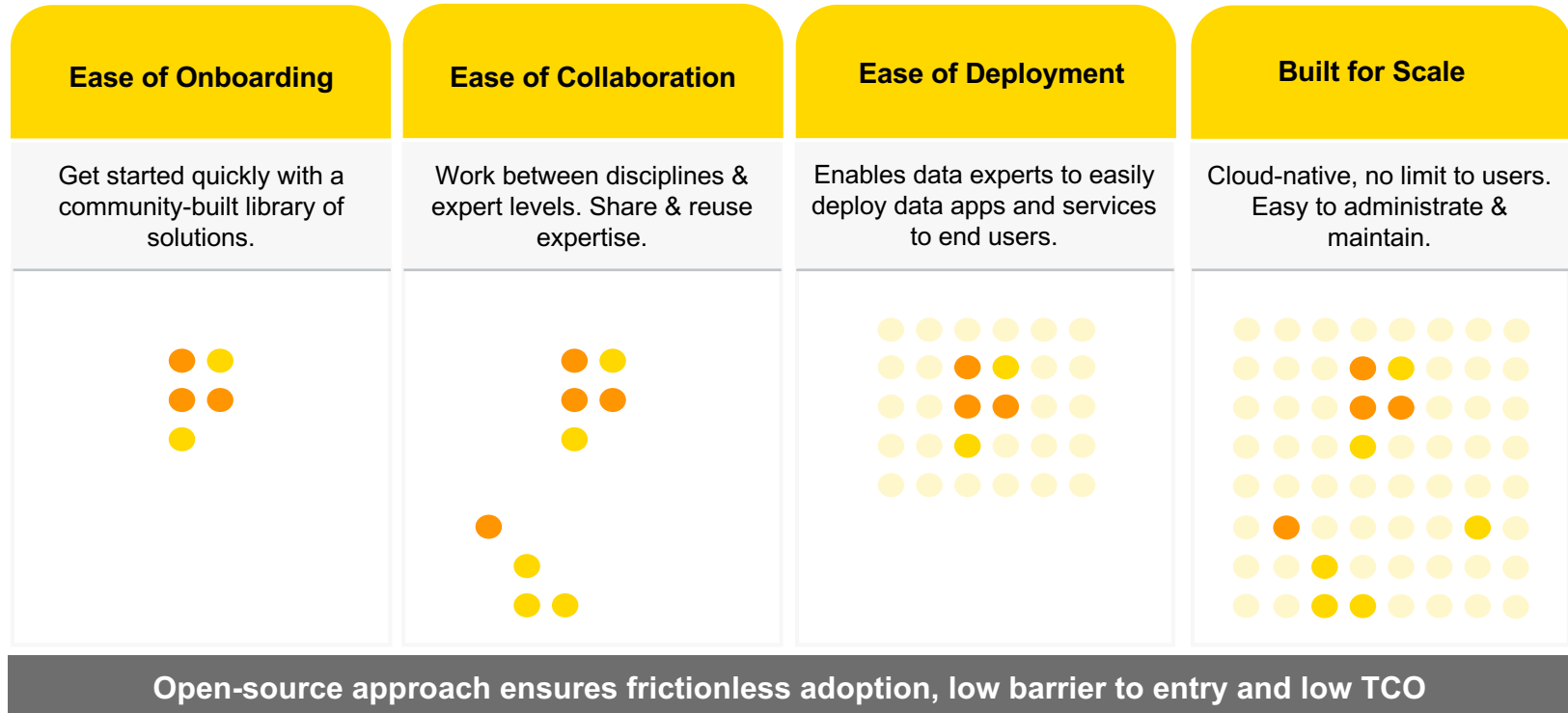
Welcome to the
KNIME Community Hub



Genentech



How KNIME Hub Scales Data Science Impact



KNIME Hub Key User Concepts

Relevant to
KNIME Business Hub *

Relevant to
KNIME Community Hub

- **User** builds data science solutions & shares them
- **Teams** are units of users. They collaborate on spaces.
- **Spaces** store a team's files, components & workflows
 - Spaces can be **public** or **private** to a team
 - Spaces are versioned
- **Execution resources (vCores)** are specific to teams
 - Explicit deployments of data apps and REST services
 - Ad hoc execution, scheduled execution
 - Execution context to configure where workflows are executed

Individuals

Free.

What's Included:

- ✓ Browse, download & share workflows & components with community
- ✓ Store workflows in private spaces
- ✓ Collaborate with others

Teams

(3 users)

Subscription on Community Hub

What's Included:

- ✓ Browse, download & share workflows & components with community
- ✓ Store workflows in private spaces
- ✓ Collaborate with team members in public spaces or private spaces
- ✓ Centralized billing, on team level
- ✓ Option to extend disk space for your spaces

Basic

(5 users, 4 vCores)

Installed into customer infrastructure

What's Included:

- ✓ Single Team
- ✓ Browse, download & share workflows & components just with your team in a dedicated instance
- ✓ Deploy and manage workflows as scalable apps and services
- ✓ Single execution context
- ✓ Enterprise authentication via e.g., LDAP or OIDC

Standard

(5 users, 8 vCores)

Installed into customer infrastructure

What's Included:

- All from Basic, plus
- ✓ Supports up to three teams
- ✓ Collaborate with other teams through public spaces
- ✓ Unlimited Data App and REST API Access for consumers
- ✓ Advanced permission management (support for internal / external groups)
- ✓ Unlimited execution context
- ✓ Edge for highly scalable REST API deployments

Enterprise

(20 users, 16 vCores)

Installed into customer infrastructure

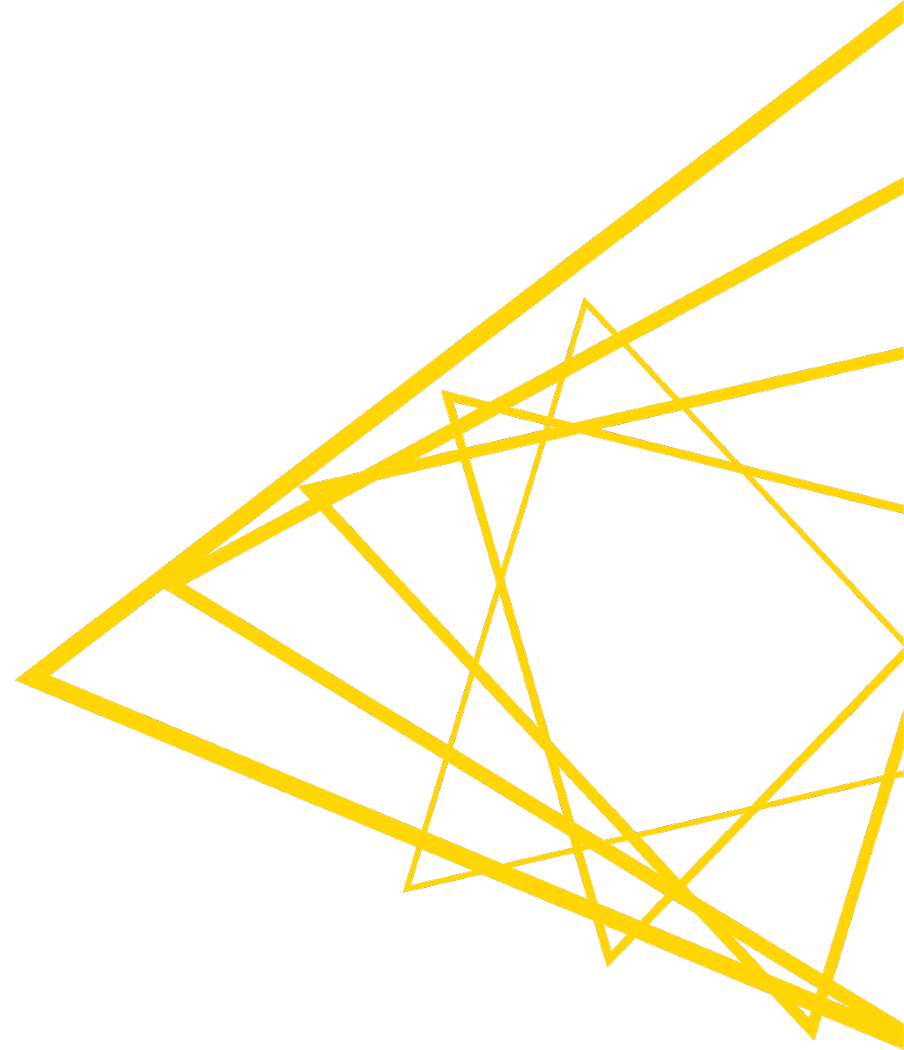
What's Included:

- All from Standard, plus
- ✓ Anonymous (read-only) user support
- ✓ Supports unlimited teams
- ✓ Several KNIME Hub installs possible
- ✓ Installation into existing Kubernetes clusters

KNIME-managed
KNIME Community Hub

Customer-managed
KNIME Business Hub

**And What is New in
KNIME Analytics Platform?**



4.7 Modern UI Preview

KNIME Analytics Platform - /Users/ps/knime-workspace-4-6-3-modern-ui

KNIME Explorer

- My-KNIME-Hub (api.hub.knime.com)
- EXAMPLES (knime@api.hub.knime.com)
- LOCAL (Local Workspace)

Workflow Coach

Recommended Nodes	Community
Excel Reader	27%
CSV Reader	20%
Table Creator	12%
File Reader	9%
DB Reader	4%
DB Query Reader	4%
DB Table Selector	3%
List Files/Folders	3%
Table Reader	2%
Create Date&Time Range	2%
File Reader (Complex Format)	1%
Database Reader (legacy)	1%
DB Connector	<1%
Microsoft SQL Server Connector	<1%
Variable Creator	<1%
Read Excel Sheet Names	<1%
Microsoft Authentication	<1%

Node Repository

- IO
- Manipulation
- Views
- Analytics
- DB
- Other Data Types
- Structured Data
- Scripting
- Tools & Services
- KNIME Labs
- Workflow Control
- Workflow Abstraction
- Reporting

Workflow Diagram: Building a Simple Classifier

Simple Model Training for Classification

This workflow demonstrates how a simple classifier is built and applied to new data. It also illustrates the use of KNIME's hitting capabilities, which allow interactive views to be connected within the same workflow.

Task: Predict the income group from demographic attributes of the adult data set (census data).

Find more information on KNIME's Learning Hub at <http://www.knime.org/learning-hub> (tutorials, videos, white papers, many more workflows).

Workflow Steps:

- Data Reading:** CSV Reader (Reading adult.csv)
- Graphical Properties:** Color Manager (Read for income "<=50K" Blue for income ">50K")
- Data Partitioning:** Partitioning (Random drawing 80% upper part 20% lower part)
- Train a Model:** Decision Tree Learner (Train to predict class "income")
- Apply the Model:** Decision Tree Predictor (Apply decision tree model to test set)
- Score the Model:** Scorer (Node 12)
- Descriptive Statistics:** Statistics (Stats and exploratory histograms in View)
- Interactive Table:** Interactive Table (Local) (Show entire data as table)
- Visualize:** Scatter Plot (Age vs. number-hours color-coded by income)

Right-hand Pane:

Building a Simple Classifier

Title: Simple Model Training for Classification

Description: Task. From adult data set (census data) predict income group from demographic attributes.

Tags: model training, classification, training, testing

Links: Learning Hub (Website)

Creation Date: 2016-7-25

Author: KNIME

Outline: (Summary of workflow steps)

Console / Node Monitor: Node: n/a, State: no node selected, n/a. Load data

Extra: Views

4.6

Bar Chart



Node 1

Line Plot



Node 2

Table View



Node 4

Scatter Plot



Node 3

4.7

Box Plot



Node 5

Density Plot



Node 6

Stacked Area Chart



Node 9

Histogram



Node 7

Pie Chart



Node 8

Statistics



Node 10

Extra: Views

Dialog - 4:30 - Statistics

Rows: 5 | Columns: 22

Name	Type
sepal.length	Number (double)
sepal.width	Number (double)
petal.length	Number (double)
petal.width	Number (double)
variety	String

Statistics

Excluded Values

- sepal.length
- sepal.width
- petal.length
- petal.width
- variety

Included Values

- sepal.length
- sepal.width
- petal.length
- petal.width
- variety

Displayed Statistics

Excludes

- # Unique val...
- Minimum
- Maximum
- 10 most co...
- 1% Quantile
- 5% Quantile
- 10% Quantile

Includes

- Name
- Type

View

Title: Statistics

Buttons: Cancel, OK

Dialog - 4:10 - Pie Chart

lupOP

Data

Dimension: population

Aggregation

- None
- Occurrence count
- Sum
- Average

Aggregate small categories

Threshold: 15

Plot

Title: lupOP

Size: 50

Donut chart

Sort categories by

- Size
- Name
- Insertion order

Buttons: Cancel, OK

Dialog - 4:27 - Stacked Area Chart

Produced carbon emissions

Data

Horizontal dimension: Year

Aggregation

- None
- Occurrence count
- Sum
- Average

Frequency dimensions

Excludes

- Solid fuel
- Liquid fuel
- Gas fuel
- Cement produ...
- Gas flaring

Max rows: 2500

Plot

Title: Produced carbon emissions

Buttons: Cancel, OK

Dialog - 4:13 - Density Plot

Density Plot

Data

Dimension column: sepal.width

Condition column: None

Plot

Title: Density Plot

Kernel bandwidth

- Rule of thumb
- Fixed

Adjustment factor: 0.4

Dimension axis label: Dimension

Density axis label: Density

Buttons: Cancel, OK

Dialog - 4:14 - Histogram

Histogram of sepal.length

Data

Dimension: sepal.length

Binning

Binning Type

- Equal width
- Equal frequency
- Custom cutoffs

No. bins: 20

Enforce integer cutoffs

Fix lower bound

Fix upper bound

Special value handling

- Ignore special values
- Show single bar aggregating all invalid values
- Show separate bar for each type of special values

Buttons: Cancel, OK

Dialog - 4:12 - Box Plot

Please execute the node to see the preview.

Data

Dimension columns

Excludes

- sepal.length
- sepal.width
- petal.length
- petal.width

Condition column: variety

Plot

Title: Box Plot

Dimension axis label: Dimension

Value axis label: Value

Buttons: Cancel, OK

Geospatial Analytics for All



Geospatial Analytics Extension for KNIME

- Support for most common vector data e.g., points, lines, polygons, collections
- Joined development with the Center for Geographic Analysis from Harvard
- We will continue to work together to add more functionality and improve the usability
- Easy installation via drag and drop from KNIME Community Hub
 - Batteries included - no additional setup and installation steps required
- Feedback is always welcome via [KNIME Forum](https://forum.knime.com/c/community-extensions/geospatial)^o or [Github](https://github.com/spatial-data-lab/knime-geospatial-extension)^{*}

^o<https://forum.knime.com/c/community-extensions/geospatial>

^{*}<https://github.com/spatial-data-lab/knime-geospatial-extension>

End to End Geospatial Analytics



Spatial IO



GeoFile Reader



GeoFile Writer



GeoPackage Reader



GeoPackage Writer



Open Datasets



OSM Boundary Map



OSM POIs



OSM Road Network



US ACS 5-Year Estimates



US2020 Census Data



US2020 TIGER Map



Spatial Conversion



GeoJSON to Geometry



Geometry to GeoJSON



Geometry to Lat/Long



Geometry to WKT



Lat/Lon to Geometry



WKT to Geometry

End to End Geospatial Analytics



Spatial Calculation

- Area
- Bounding Box
- Bounds
- Convex Hull
- Coordinates XYZ
- Length
- Total Bounds
- Unary Union



Spatial Manipulation

- Buffer
- Clip
- Create Grid
- Dissolve
- Euclidean Distance
- Haversine Distance
- Multiple Ring Buffer
- Nearest Join
- Overlay
- Simplify
- Spatial Join



Spatial Transformation

- Geometry To Point
- Line To MultiPoint
- Multipart To Singlepart
- Points To Line
- Polygon To Line
- Projection

End to End Geospatial Analytics



Exploratory Spatial Data Analysis



Global Geary's C



Global Getis-Ord G



Global Moran's I



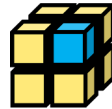
Local Getis-Ord G



Local Moran's I



Spatial Weights



Spatial Modelling



2SLS with Spatial Test



GWR Model



GWR Predictor



MGWR Model



OLS with Spatial Test



Spatial Error Model



Spatial Error Panel Model



Spatial Lag Model



Spatial Lag Panel Model



Location Analysis



LSCP



MCLP



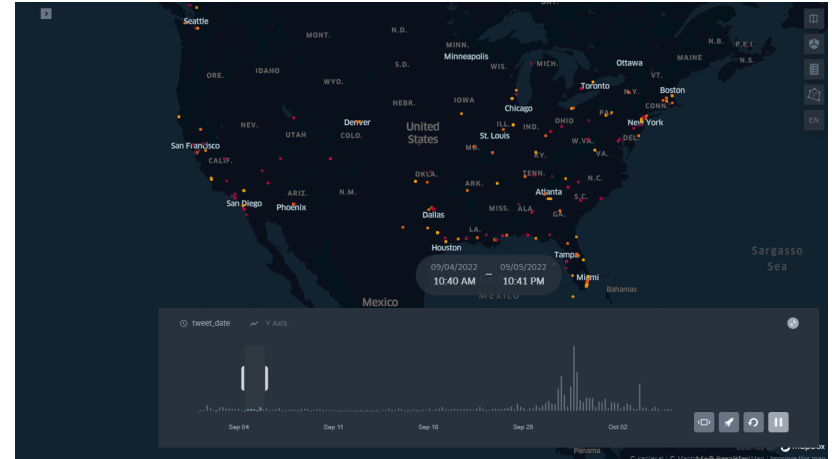
P-median

End to End Geospatial Analytics

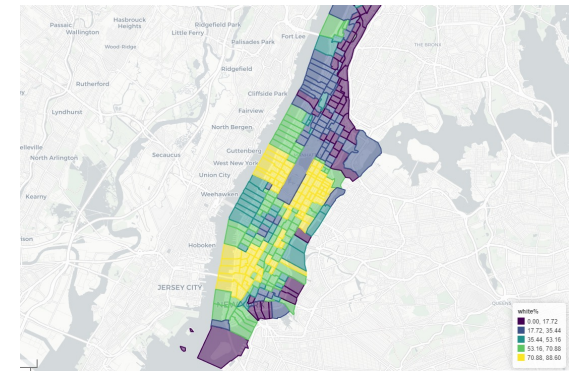
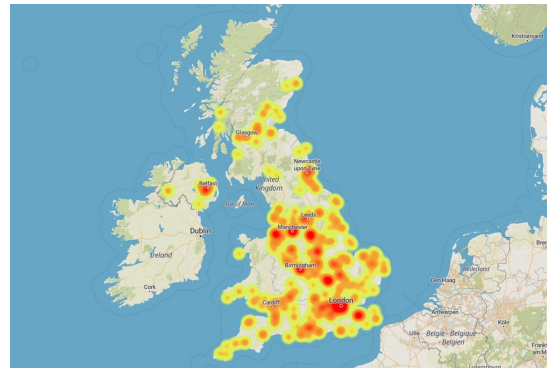
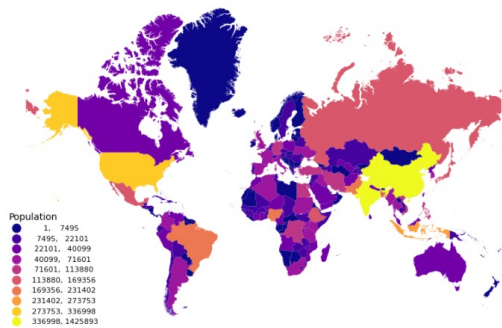


Spatial Visualization

- Geospatial View
- Geospatial View Static
- Kepler.gl Geoview
- Spatial Heatmap



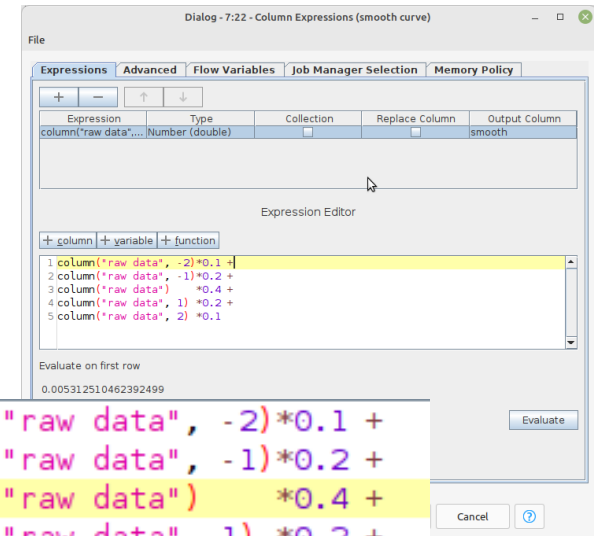
Population by Country using Mercator Projection



Column Expressions - Multi-row Formula

- Swiss army knife to run custom snippets on table rows (combining *String Manipulation*, *Math Formula*, *Rule Engine*)
- Powerful but – until 4.5 – limited in ability to access data outside the current row
- New in 4.6: Access to previous/subsequent rows (multi-row formula)
- Crucial for "between rows" calculations, e.g.:
 - Time Series Calculation
 - Diffs to previous/subsequent rows
 - Data Smoothing

Column Expressions



```
1 column("raw data", -2)*0.1 +
2 column("raw data", -1)*0.2 +
3 column("raw data") *0.4 +
4 column("raw data", 1) *0.2 +
5 column("raw data", 2) *0.1
```

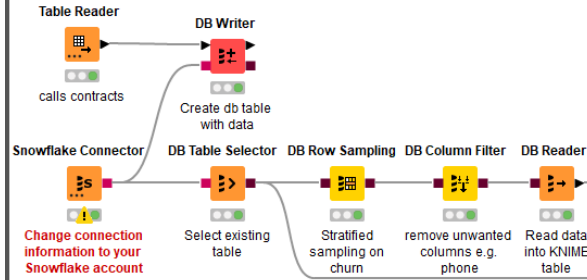

Other Noteworthy Additions

- Extended Microsoft Azure Services
 - Azure Synapse Analytics, SharePoint Online List, Azure Active Directory
- DB Framework Enhancements
 - DB Concatenate, DB Looping, DB Delete, DB Date Spec Extract
 - Updated JDBC drivers (Redshift, H2, MS Access, MySQL, PostgreSQL, SQLite)
 - Enhanced Oracle DB Support

Snowflake Machine Learning Model Push-Down

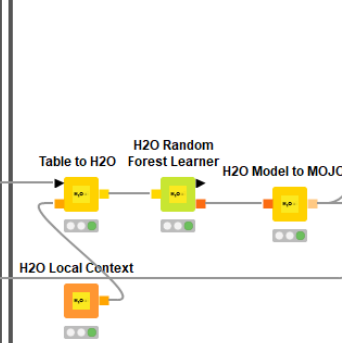
1. Data Preparation.

Sample and filter data within Snowflake before loading it into a KNIME table.



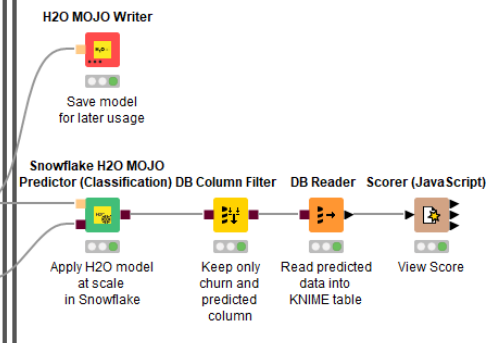
2. Model Learning.

Learns a Random Forest model and converts it into a transportable MOJO model.



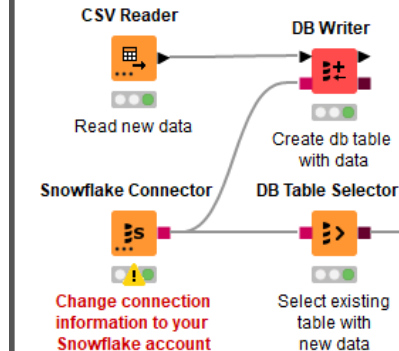
3. Model Evaluation and Storing.

Send the MOJO model to Snowflake and predict the remaining data in the Snowflake data table. The predicted and filtered result is read into KNIME to evaluate the model accuracy.



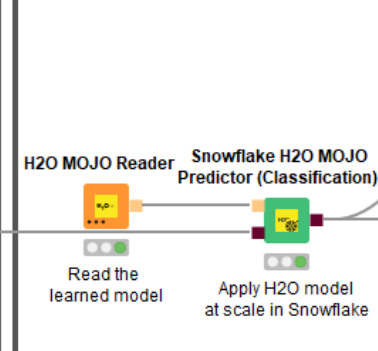
1. Selection.

Selects the "calls_contracts" table to use for prediction.



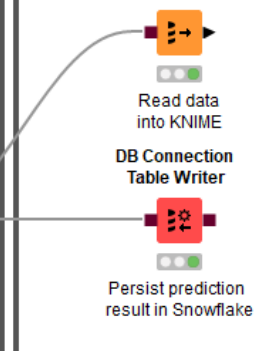
2. Deployment and Execution.

Convert the customer churn prediction model into a User-Defined Function in Snowflake.



3. Persisting.

The prediction result can either be read back into KNIME or persisted as new database table within Snowflake.



<https://www.knime.com/blog/churn-predictor-knime-snowflake>



Thank you!

