

Data Science Applications in the Public and Private Sector

Masoud Panshiri & Schalk Gerber

Three Use Cases

to unlocking data-driven decisions in the public and private sectors

Public

Time Cost of Data Analytics



Private

Sparking Data Literacy



Public & Private

Universities & Companies



Use Case 1: VAT Fraud Detection

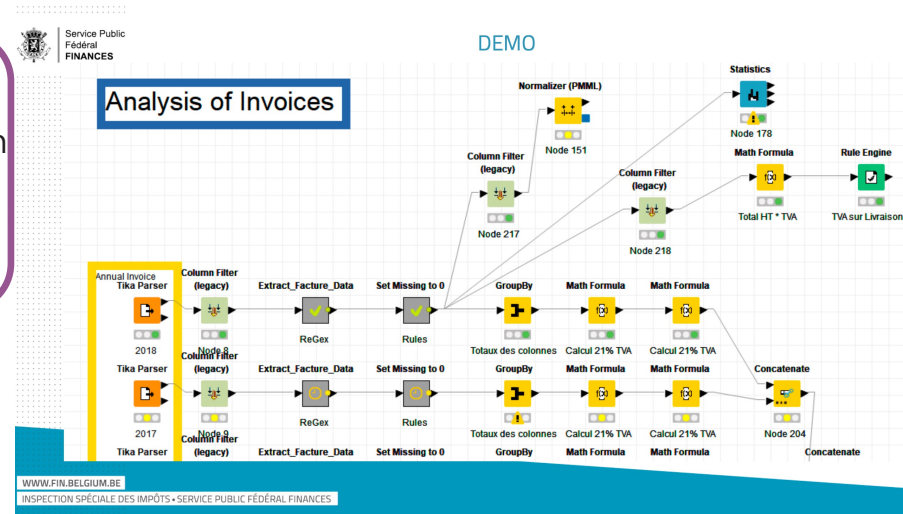
- Challenge:
 - E-audit
 - Different data sources: databases, documents, images
 - Different data types:
 - Excel files - bank transaction
 - PDF's - invoices
 - XML files
 - Example auditing VAT for one company
 - Over 3 year period
 - Up to 132 000 invoices per year
 - Previous e-audit tool
 - Work on one table - sample
 - 1 week no results
 - Repetitive task for each table and different files




Federal
Public Service
FINANCE

Use Case 1: VAT Fraud Detection

- Solution:
 - **Visual Workflows** with KNIME Analytics Platform
 - **Complete Automation** (*without code*)
 - Worked on all files



- Results
 - Processing and analyzing 3 years transactions, **240 000** Invoices, in **6 hours**
 - Calculated VAT value of **€1 065 765**
 - Visual workflows **language to communicate** with managers and lawyers
 - **Open source** (free) - easy to adopt platform



“Ministry Finance Belgium's work made it possible to refine the description of the fraud scheme in place, not only for the tax officers themselves, but also so that they could report it to the Brussels Public Prosecutor's Office.

This global and refined description convinced the judicial authorities that this was a major fraudulent network whose players were linked, not only in Belgium, but also in other EU Member States, and that a criminal file should be opened.”

- Belgian Court of Audit -

Finance

	Data Processing	Visualization & Reporting	Data Exploration	Advanced & Predictive Analytics
	Access	Hindsight	Insight	Foresight
All Banking and FSI	<ul style="list-style-type: none"> Budget Monitoring EUC Remediation Ledger Error & Compliance 	<ul style="list-style-type: none"> KPI Reporting Liquidity Tracking Regulatory Reporting 	<ul style="list-style-type: none"> Self Service Analytics Process Mining Customer Segmentation 	<ul style="list-style-type: none"> Cash Flow Prediction Anti-Money Laundering Fraud Detection
Retail Banking	<ul style="list-style-type: none"> Collection & Monitoring Know your Customer GDPR Compliance 	<ul style="list-style-type: none"> Campaign Performance Retail Shop Performance Omni-Channel Reporting 	<ul style="list-style-type: none"> Behavioral Analysis Sentiment Analysis Early Warning Signals 	<ul style="list-style-type: none"> Credit Scoring Churn Prediction Product Recommendation
Wealth Management	<ul style="list-style-type: none"> Market Data Collection Trade Processing Client Financial Reporting 	<ul style="list-style-type: none"> Deal Performance Analyst Risk Assessment Margin Analysis 	<ul style="list-style-type: none"> Key Stock Statistics Fund Manager Behavior All Security Decisions 	<ul style="list-style-type: none"> Detecting Crisis Situations Algorithmic Trading Risk Prediction
Investment Banking	<ul style="list-style-type: none"> Fund Management Client Data Aggregation Analyst Consensus 	<ul style="list-style-type: none"> Portfolio Performance Earning Reporting Trade Surveillance 	<ul style="list-style-type: none"> Research & News Historical Risk/Return Competitive Behavior 	<ul style="list-style-type: none"> Asset Allocation Plan Projections Cross-Selling
Automation & Reduction of Manual Work				
Increased Data Literacy				
Competitive Differentiation				

Use Case 2: Sparking Data Literacy

Challenge

- Time-consuming data processing & reporting due to complete reliance on Excel.

Solution

- Automated manual processes such as planning, reporting, Kanban boards, deadline tracking, supply chain warnings with KNIME.
- Data integration and data wrangling done independently by business users in KNIME.
- KNIME rolled out company-wide for 1000+ business users across Controlling, Tax, HR, R&D, Production Planning, etc.

Results

Month-end controlling tasks shortened from 2 days to 30 minutes



Ability to run an entire plant budgeting process at a tenfold precision level



Serial number traceability to R&D, plant, and quality users as self-service-providing answers in minutes instead of days

“

From a business perspective, the open source model was the only commercially feasible option to rapidly spread data analytics competence throughout the organization”

- Arne Beckhaus, Head of Data Architecture Automotive, Continental.

Manufacturing

Automation

Data Literacy

Use Case 2: Sparking Data Literacy

Extension

Continental Nodes for KNIME

v 1.4.1

The Continental Nodes for KNIME Extension provides data processing and reporting capabilities intended for business users of the KNIME Analytics Platform. This extension comprises the XLS Formatting nodes.

Included nodes

- Network Component Splitter**
This node analyzes a list of relations between nodes for unconnected network components. ...
- XLS Background Colorizer**
The XLS Background Colorizer node changes the background color and/or pattern fill for se...
- XLS Border Formatter**
The XLS Border Formatter node changes cell borders of a given range (not necessarily rect...
- XLS Cell Commenter**
The XLS Cell Commenter node adds comments to cells. This node defines a formatting instru...

Overview
Included nodes
Related workflows
Legal & update site

Advance Shipment Notification (ASN) Use Case

NodeType	NodeCount
Rule Engine	18
MetaNode	16
Column Rename	15
Constant Value Column	15
Concatenate	11
Multi Formula	9
GroupBy	9
Joiner	8
Excel Reader_XLS	7
Missing Value	6
Column Filter	5
Row Splitter	4
Transpose	3
Column Resorter	3
Insert Column Header	3
String Manipulation	2
String To Number	2
Unpivoting	2
Excel Writer_XLS	2
Sorter	1
Column Splitter	1
	Σ 142

Advance Shipment Notification (ASN) Use Case

perspective	aggregation	MonthAggregation	goods receipt	green	green (relativ)	yellow (relativ)	red (relativ)	grey (relativ)
BU	KN	08.2018	4760	2729	0.57322244	0.28698041	0.08987246	0.05002509
BU	KN	09.2018	4859	2887	0.59405246	0.33328711	0.09233838	0.01983314
BU	KN	10.2018	4787	2870	0.59950189	0.32819088	0.10354448	0.11611254
BU	KN	11.2018	4758	2946	0.533698781	0.32522858	0.08759428	0.072275975
BU	KN	12.2018	3913	2042	0.52032272	0.29399041	0.19274645	0.01090401
BU	BER	08.2018	7626	3430	0.449512129	0.29258379	0.17541791	0.20961132
BU	BER	09.2018	7868	3552	0.449912122	0.34652209	0.17388397	0.057780458
BU	BER	10.2018	7908	3630	0.459237979	0.27945427	0.087782526	0.042846227
BU	BER	11.2018	7314	3447	0.473002727	0.28434172	0.18124837	0.18448213
BU	BER	12.2018	2964	1409	0.476333289	0.32552849	0.137863785	0.153718489
Plant	54	08.2018	2630	888	0.33693617	0.28423858	0.05631202	0.34081512
Plant	54	09.2018	2880	979	0.33891054	0.32523091	0.08722444	0.27013244
Plant	54	10.2018	2964	993	0.33910243	0.33609317	0.08893483	0.265487117
Plant	54	11.2018	2890	958	0.33488807	0.27578242	0.09257409	0.21482542
Plant	54	12.2018	2234	808	0.36131811	0.28322246	0.07717208	0.188428797
Plant	72	08.2018	9130	5367	0.58	0.33099117	0.10537884	0.01813998
Plant	72	09.2018	9227	5711	0.62033816	0.28322246	0.146997913	0.059486479
Plant	72	10.2018	9133	5611	0.615338313	0.29477178	0.08898901	0.061788913
Plant	72	11.2018	9071	5689	0.62838839	0.27578242	0.16787799	0.03285949
Plant	72	12.2018	3871	2238	0.58099133	0.30544884	0.08211484	0.027497148
Lead Division D	08.2018	54862	32299	0.592200915	0.32522846	0.09022389	0.10967777	0.0817521
Lead Division D	09.2018	54673	31232	0.56980779	0.32522846	0.10079628	0.109666184	0.100396184
Lead Division D	10.2018	54782	33887	0.592940616	0.31	0.304128022	0.103889992	0.103889992
Lead Division D	11.2018	54627	34256	0.590999953	0.32320079	0.084294264	0.031020253	0.031020253
Lead Division D	12.2018	21874	55314	0.253773949	0.29238379	0.063811384	0.031762112	0.031762112
Lead Division F	08.2018	39883	11134	0.380883973	0.32322284	0.097892424	0.20495791	0.20495791
Lead Division F	09.2018	39375	13673	0.44	0.300770421	0.09047	0.118786379	0.118786379
Lead Division F	10.2018	39389	12129	0.399012461	0.288989431	0.10531189	0.210188318	0.210188318
Lead Division F	11.2018	39387	11911	0.4200986	0.310797442	0.10351189	0.100463818	0.100463818
Lead Division F	12.2018	12424	50147	0.400779971	0.29946718	0.075943975	0.220428718	0.220428718

(all data are displayed in randomized form)

Continental Nodes for KNIME extension!

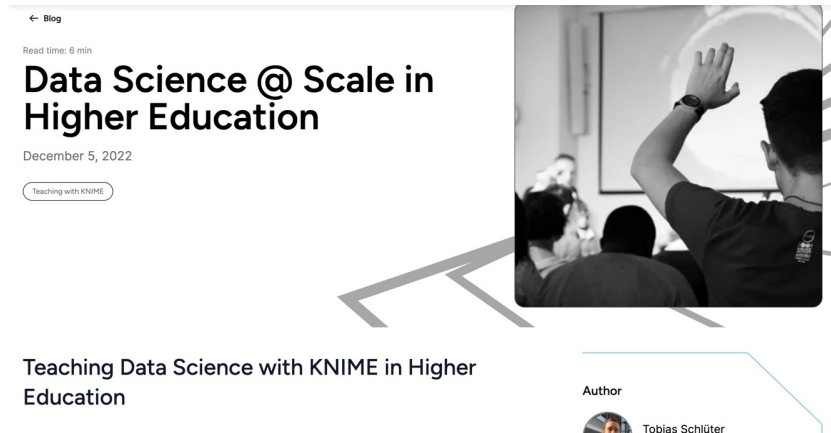
Manufacturing

		Data Processing	Visualization & Reporting	Data Exploration	Advanced & Predictive Analytics
		Access	Hindsight	Insight	Foresight
End-to-End Process Optimization	R&D	<ul style="list-style-type: none"> • Life Cycle Management • Machine Monitoring 	<ul style="list-style-type: none"> • Cost Management • Procurement 	<ul style="list-style-type: none"> • Reducing Time to Market • Connected Factory 	<ul style="list-style-type: none"> • Quality Prediction • Resource Optimization
	Production	<ul style="list-style-type: none"> • Capacity Planning • Disaster Avoidance 	<ul style="list-style-type: none"> • Equipment Monitoring • IIoT Remote Analytics 	<ul style="list-style-type: none"> • Root Cause Analysis • Inventory Risk Assessment 	<ul style="list-style-type: none"> • Predictive Maintenance • Yield Prediction
	Logistics	<ul style="list-style-type: none"> • Inventory Management • RFID Correlation 	<ul style="list-style-type: none"> • Traceability • Allocation Management 	<ul style="list-style-type: none"> • Warehouse Planning • Supply Chain Optimization 	<ul style="list-style-type: none"> • Demand Forecasting • Delay Forecasts
		<ul style="list-style-type: none"> • EH&S Risk Management • Revenue Growth Management 	<ul style="list-style-type: none"> • Energy Consumption • Sustainable Manufacturing 	<ul style="list-style-type: none"> • Automated Security • Process Mining 	<ul style="list-style-type: none"> • Price Forecasting • ML-Powered Six Sigma Analysis
Automation & Reduction of Manual Work					
Increased Data Literacy					
Competitive Differentiation					

Use Case 3: Universities & Companies

■ Challenge

- One side: **Companies are adopting KNIME** for internal use
 - Looking to hire students with KNIME skills
- Other side: **Universities adopting KNIME**
 - to teach data science, marketing analytics, etc., because of **Low-code/No-code** platform with **fast learning curve**
- KNIME has network of universities using KNIME to teach



The image shows a screenshot of a blog post. On the left, the text reads: '← Blog', 'Read time: 6 min', 'Data Science @ Scale in Higher Education', 'December 5, 2022', and a tag 'Teaching with KNIME'. On the right, there is a black and white photograph of a person in a classroom pointing at a screen. Below the photo, the text says 'Teaching Data Science with KNIME in Higher Education' and 'Author Tobias Schlüter' with a small profile picture.

■ Read more: <https://www.knime.com/blog/teaching-data-science-higher-education>

Use Case 3: Sponsored Student Challenges

- Solution:
 - Connect universities with companies
 - Create internship infrastructure

← Blog

Read time: 3 min

Students at Oklahoma State University competing for internships at Siemens Healthineers

Participants are building predictive models of injury severity in automobile accidents using low-code tools

November 6, 2023

Teaching with KNIME

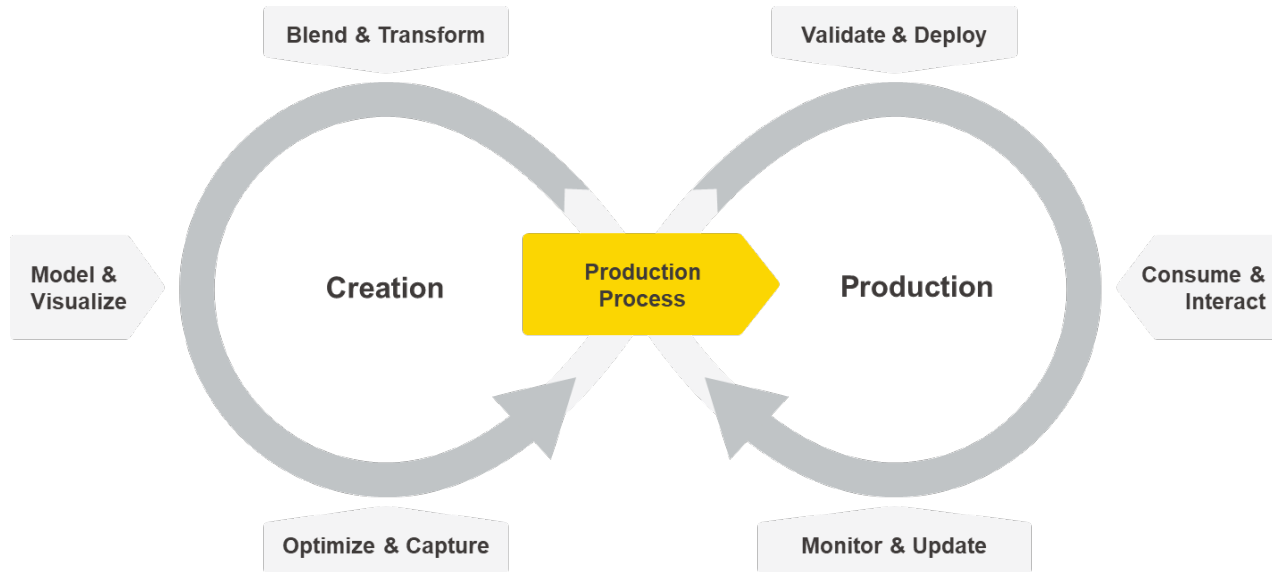
Company news



- Results
 - Increase and supply the demand for KNIME skills in job market
 - Connect universities and companies using KNIME directly

- Read more: <https://www.knime.com/blog/students-oklahoma-state-university-competing-internships-siemens-healthineers>

Data Science Life Cycle



KNIME Analytics Platform

KNIME Extensions

KNIME Integrations

Community Extensions

Partner Extensions

KNIME Business Hub

Team Collaboration

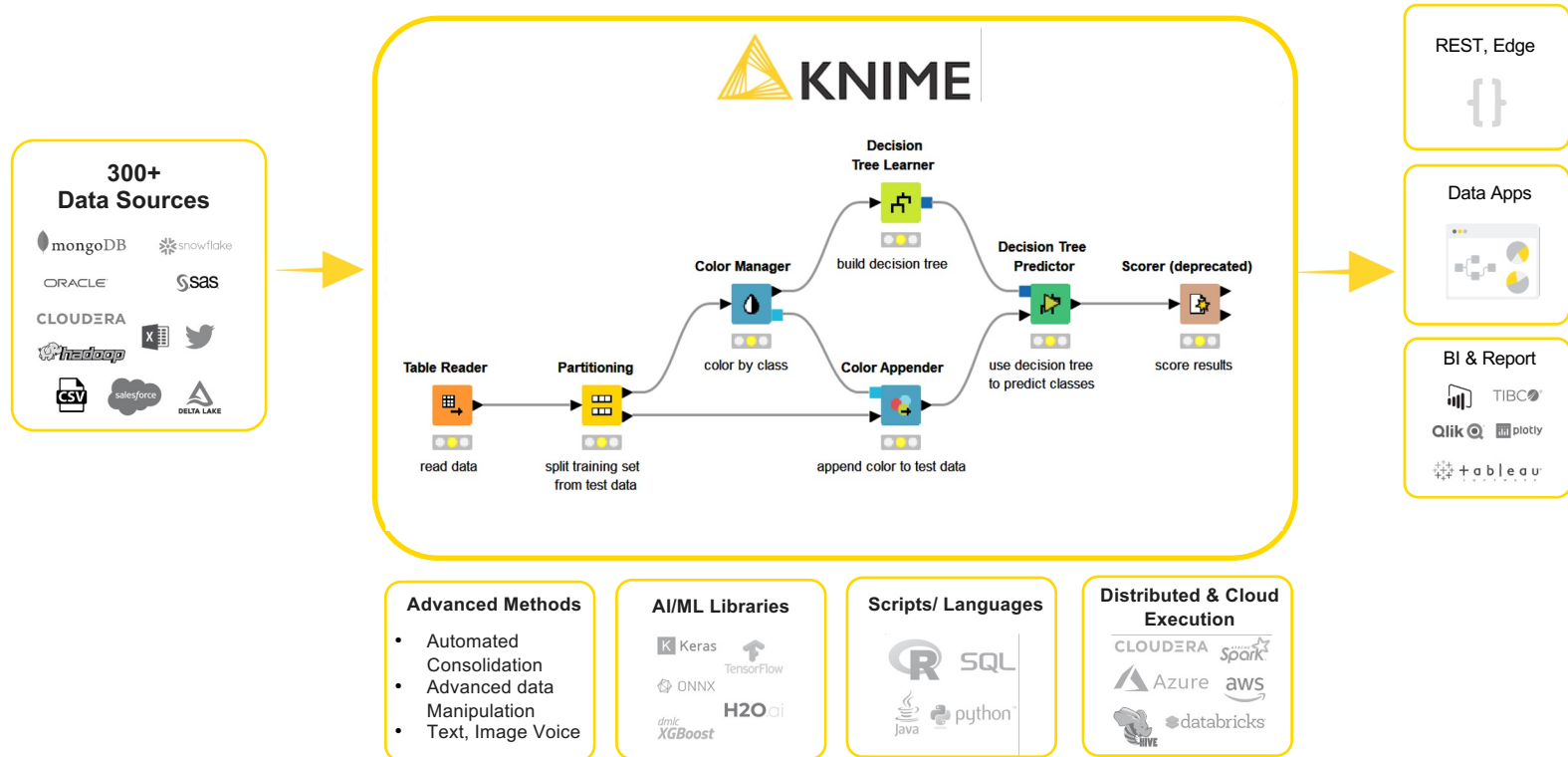
End User Applications

API Services

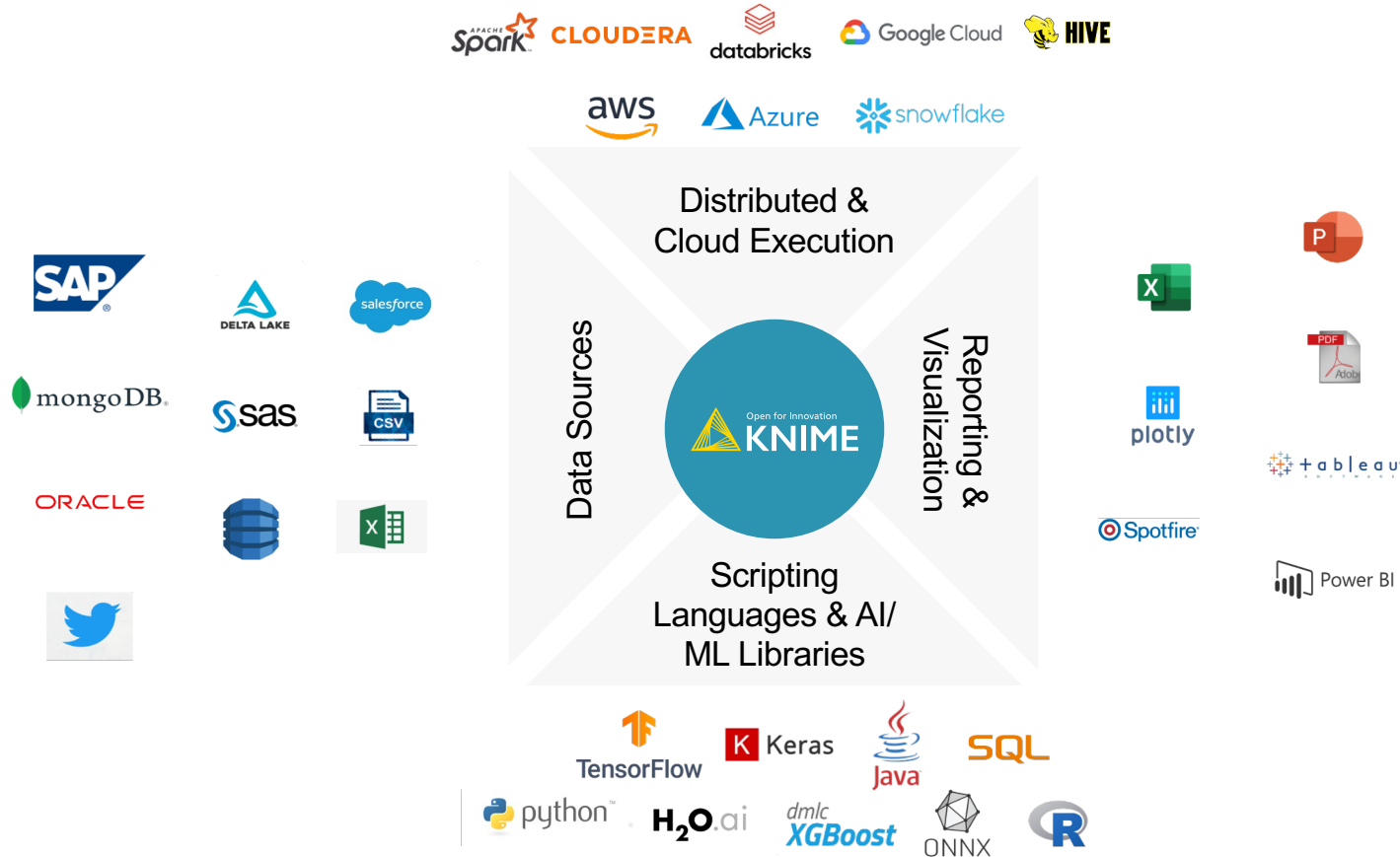
Managed Execution

One Platform for End-to-End Data Science

A Complete Analytics Platform for End-to-End Data Science



KNIME Technology Ecosystem



Let's stay in touch!

KNIME Hub



masoud.panshiri@knime.com, schalk.gerber@knime.com



[Masoud](#), [Schalk](#)