



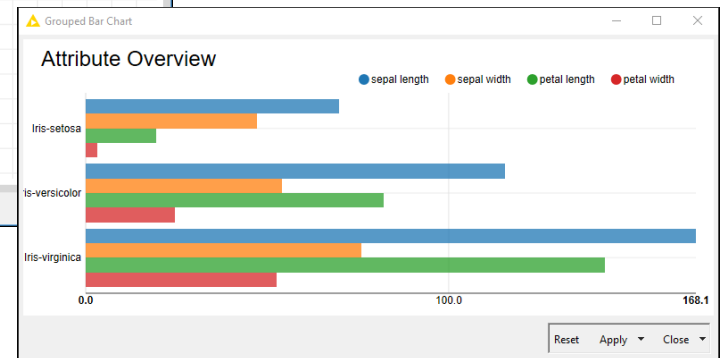
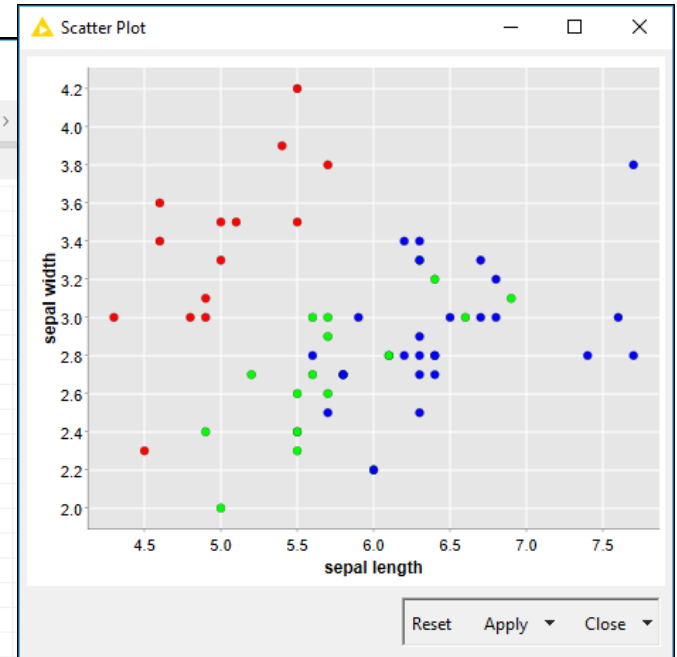
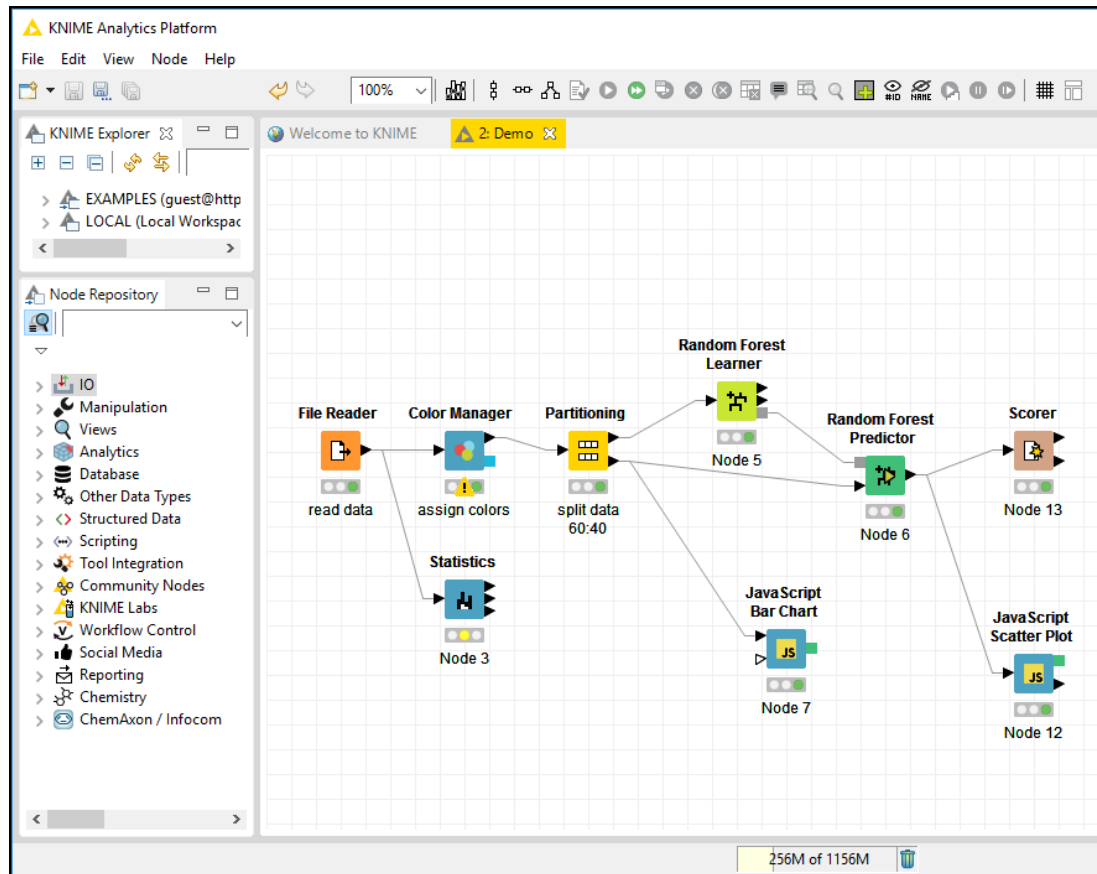
KNIME: an update

Jon Fuller, Ph.D.

ChemAxon Annual Meeting 2016

May 2016

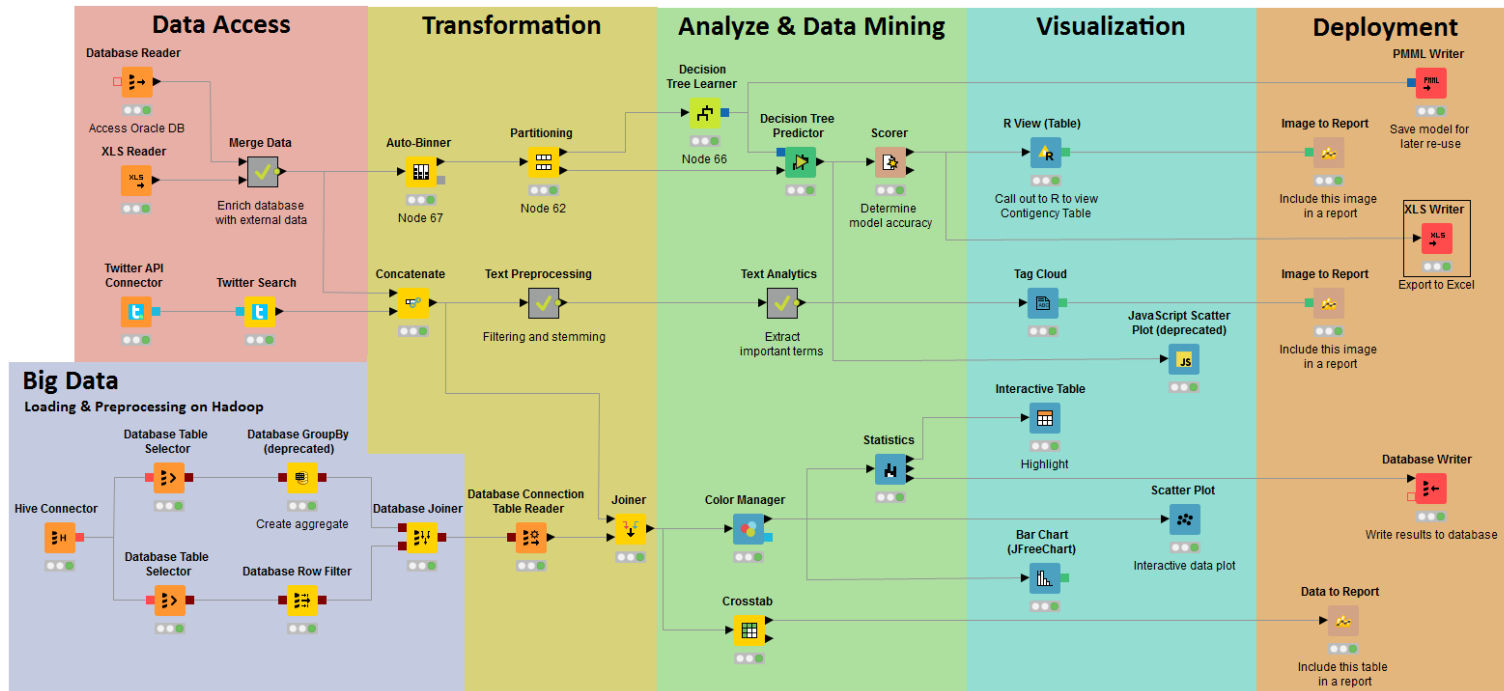
The KNIME[®] Analytics Platform



Broad Range of KNIME Application Areas & Customers



Over 1000 native and embedded nodes included:



Data Access

MySQL, Oracle, ...
SAS, SPSS, ...
Excel, Flat, ...
Hive, Impala, ...
XML, JSON, PMML
Text, Doc, Image, ...
Web Crawlers
Industry Specific
Community / 3rd

Transformation

Row,
Column
Matrix
Text, Image
Time Series
Java
Python
Community / 3rd

Analysis & Mining

Statistics
Data Mining
Machine Learning
Web Analytics
Text Mining
Network Analysis
Social Media Analysis
R, Weka, Python
Community / 3rd

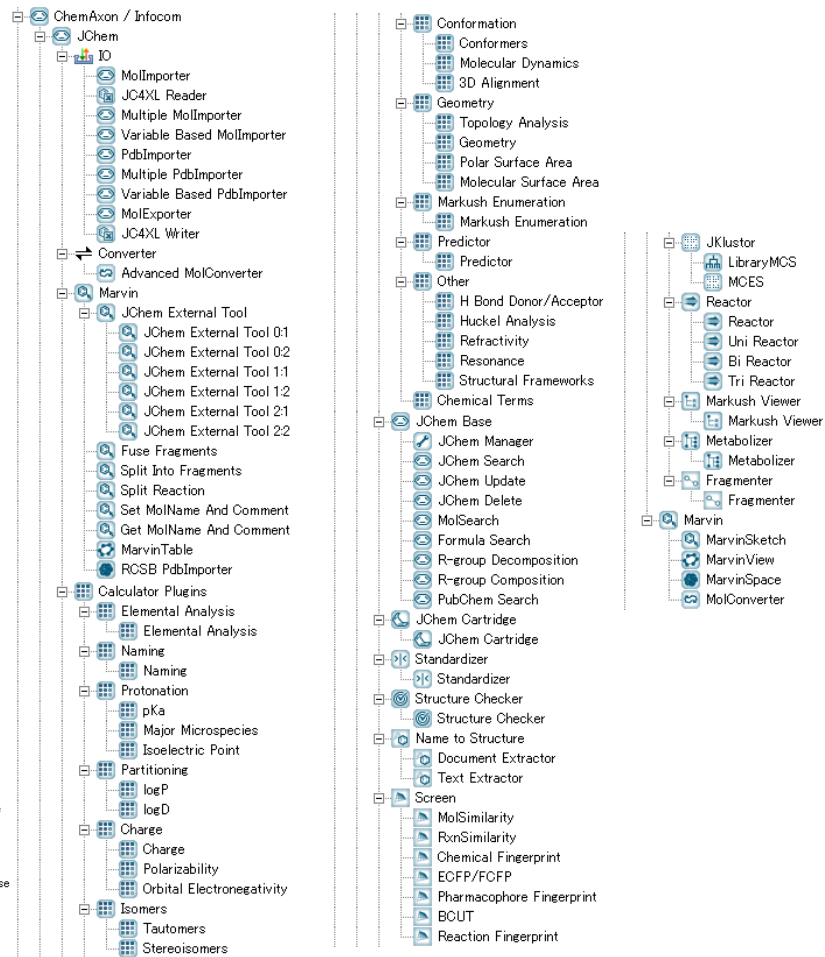
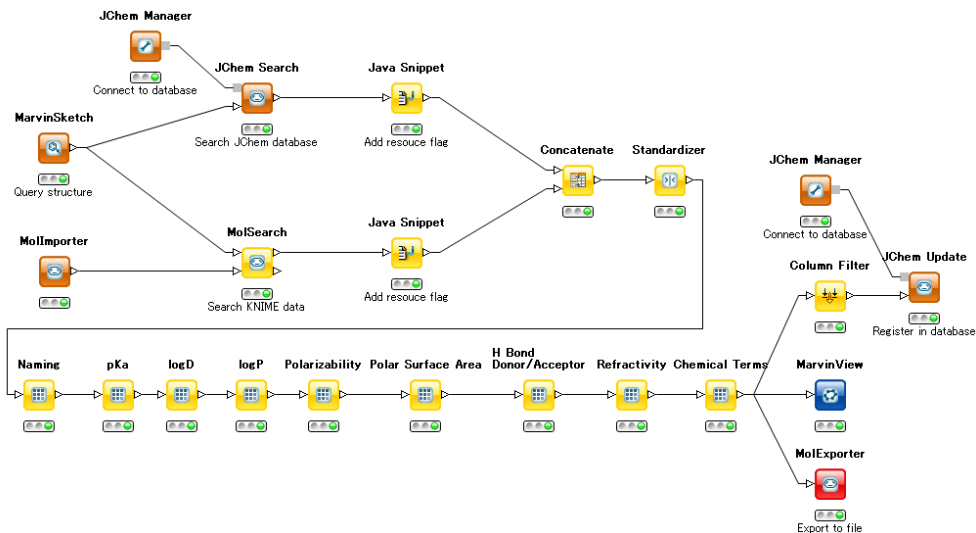
Visualization

R, Python
JFreeChart
JavaScript
Community / 3rd

Deployment

via BIRT
PMML
XML, JSON
Databases
Excel, Flat, etc.
Text, Doc, Image
Industry Specific
Community / 3rd

Infocom implements it with the support of ChemAxon



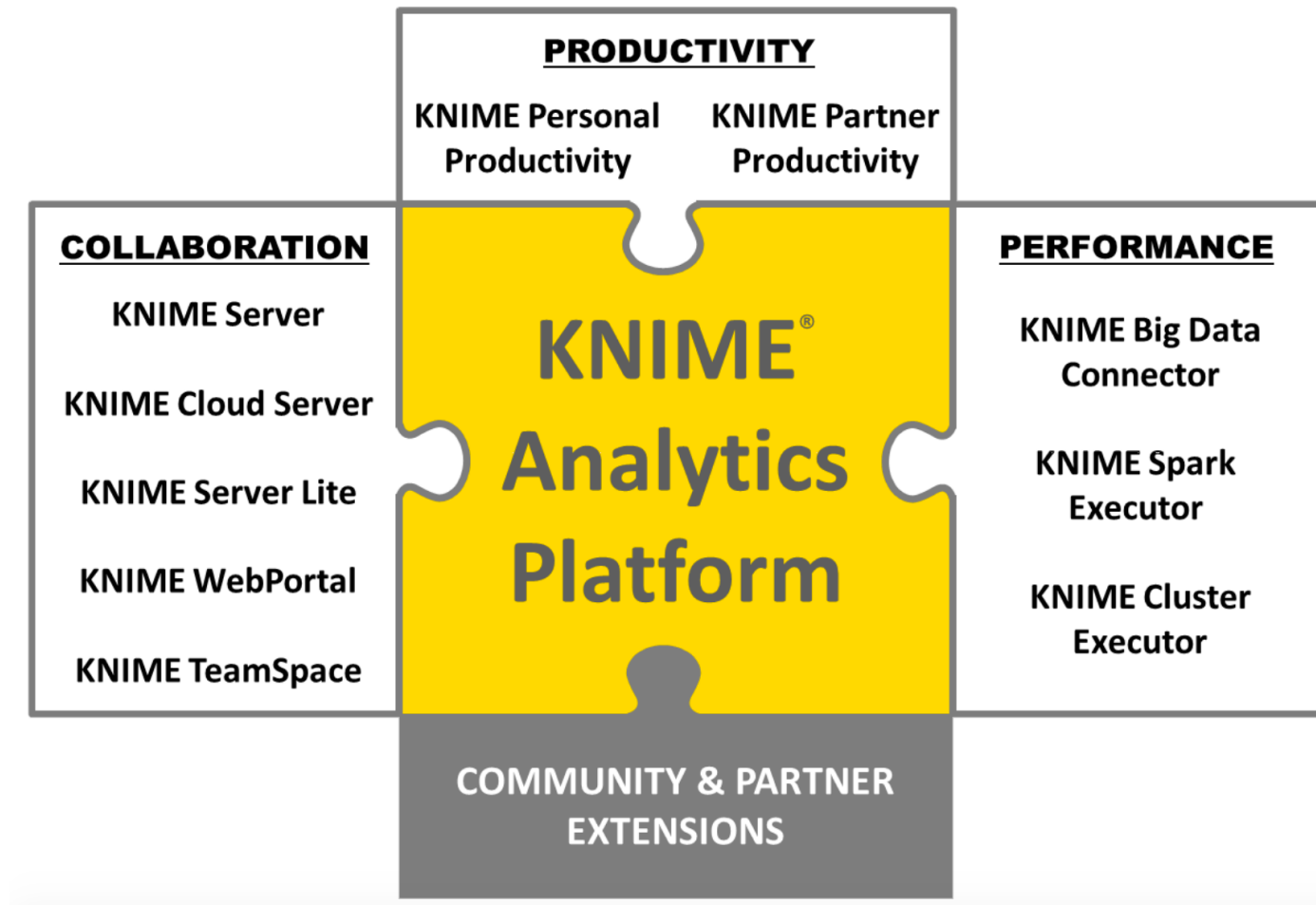
Some typical drug-discovery use cases

- Building, validating, and applying predictive models (e.g. QSAR)
- Virtual screening
- Chemistry experiment planning
- NGS data analysis
- HCS data analysis

The KNIME Commercial Extensions



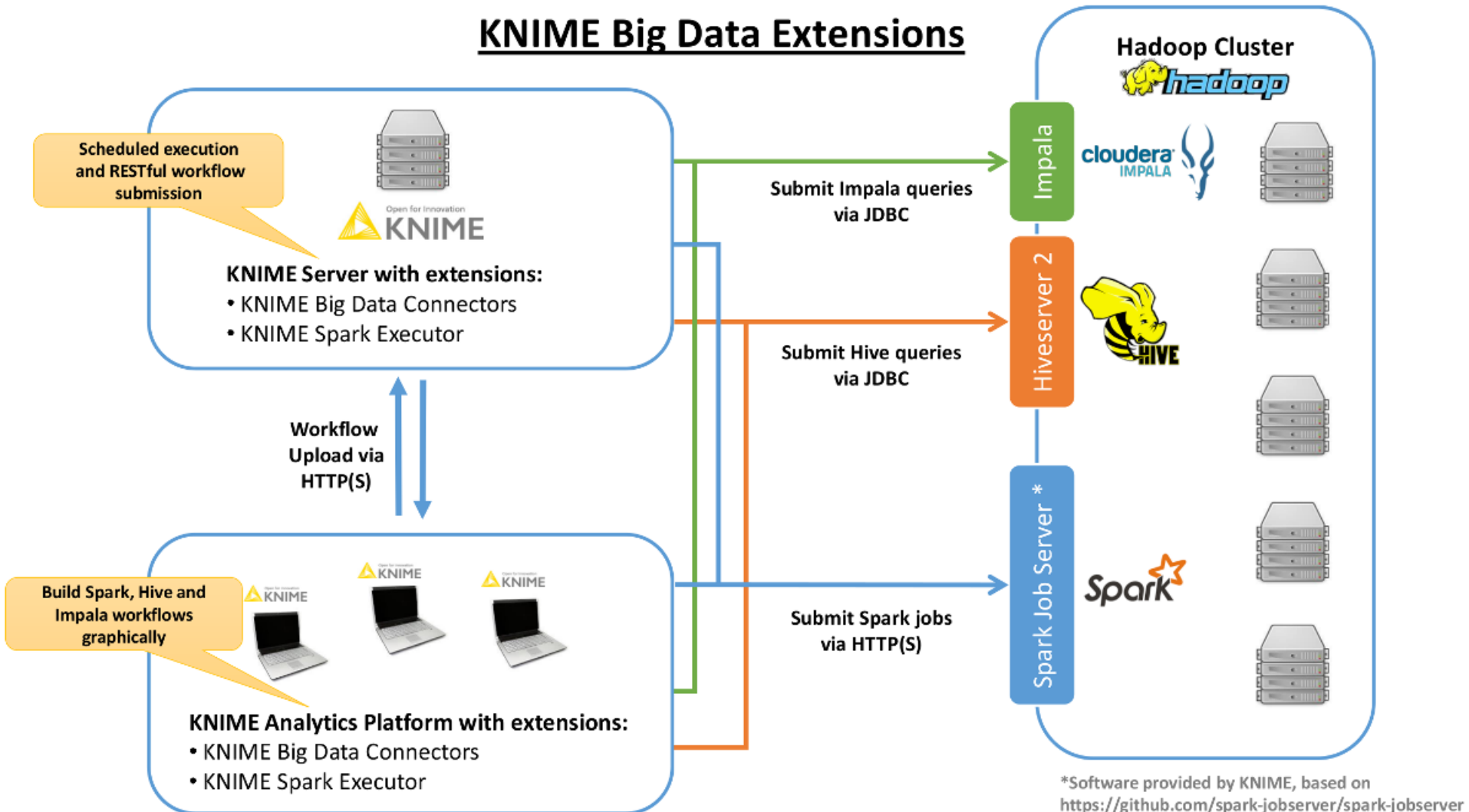
KNIME Software



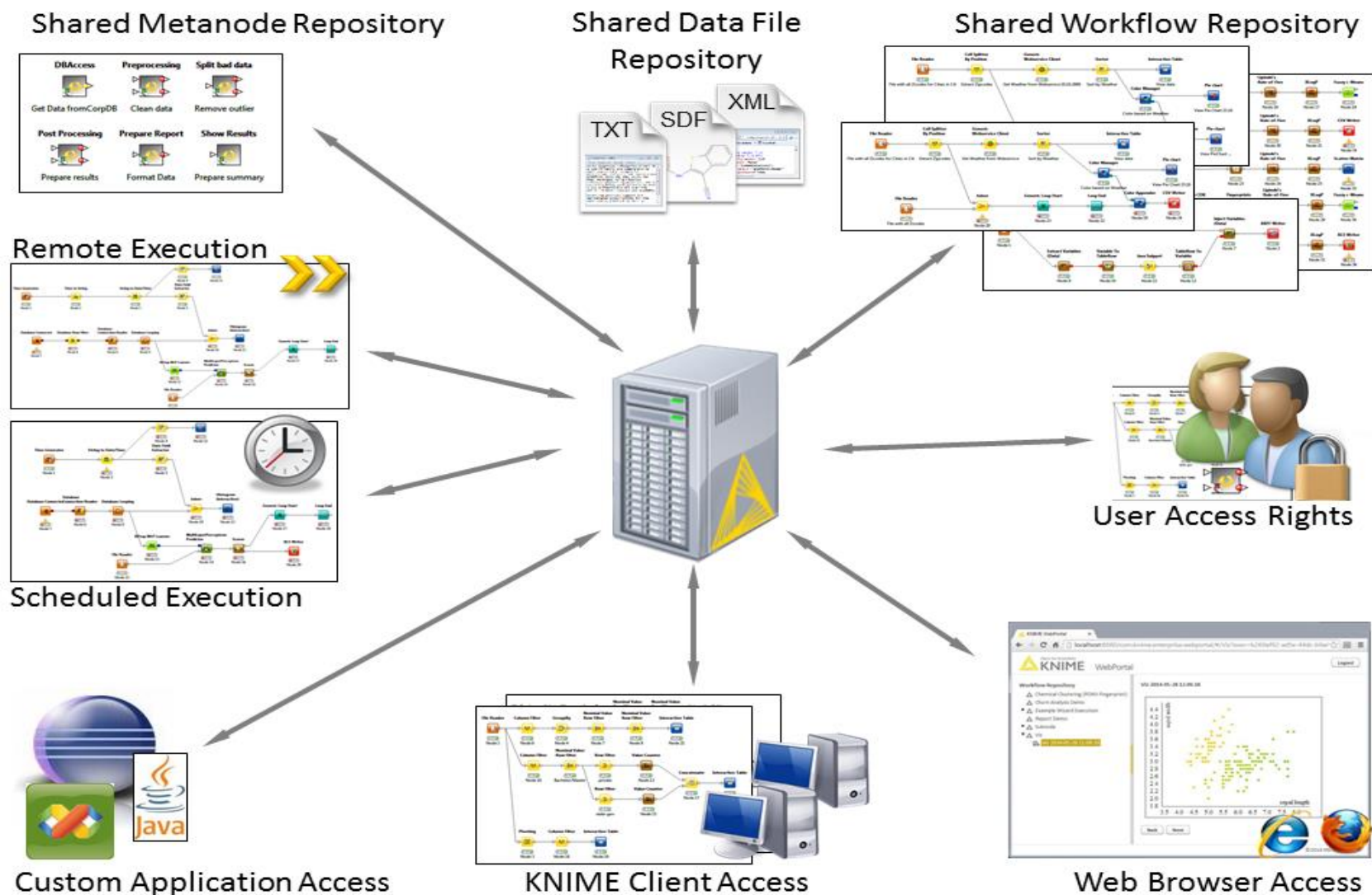
KNIME commercial extensions to the platform for **collaboration, productivity, performance**

KNIME Big Data Architecture

KNIME Big Data Extensions



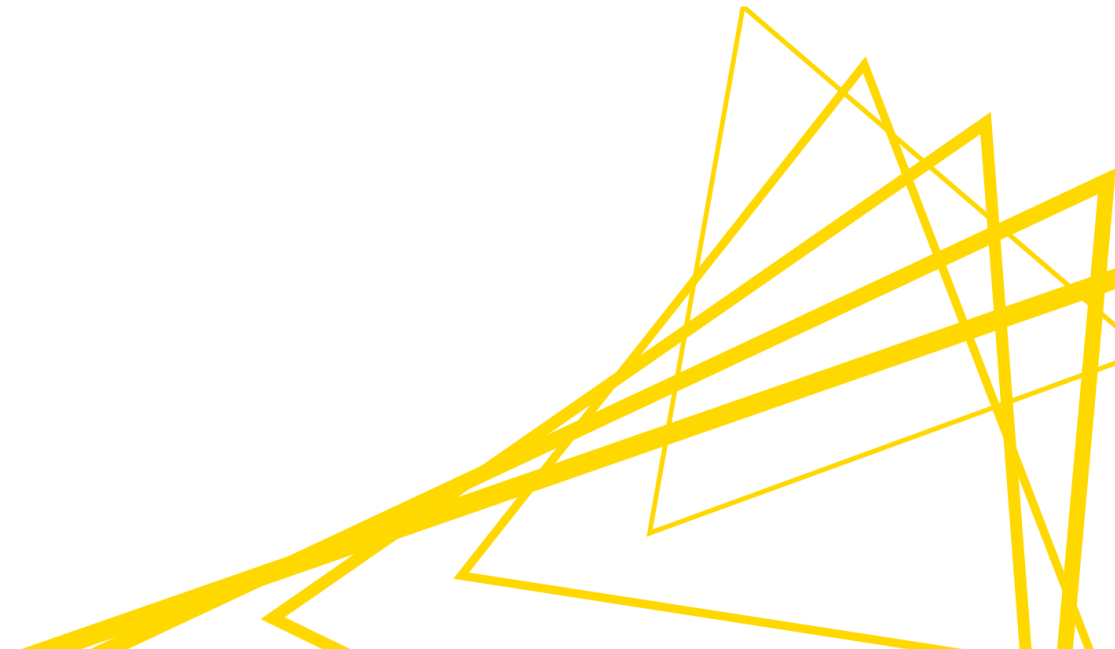
KNIME Server



KNIME Server

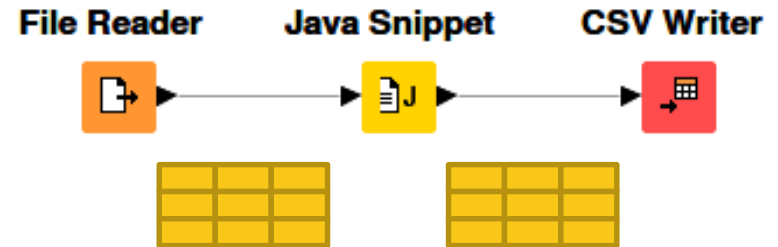


Some recent additions



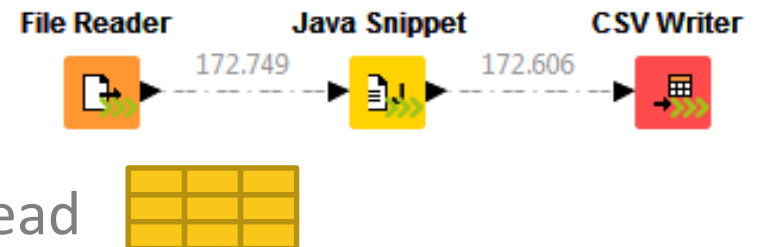
Streaming

- Default Execution



- Streaming Execution

- Row-wise
- Process, pass & forget
→ Faster with less I/O overhead
- Concurrent execution



Advantages

Less I/O overhead (process, pass & forget)
Parallelization

Disadvantages

No intermediate results, no interactive execution
Not all nodes can be streamed

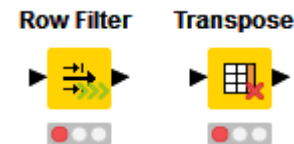
Streaming – Pros and Cons

Advantages

- Less I/O overhead (process, pass & forget)
- Parallelization

Disadvantages

- No intermediate results, no interactive execution
- Not all nodes can be streamed



WebPortal

Page 3

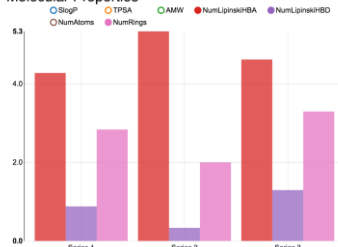
Page 1a

Upload an SDF file:

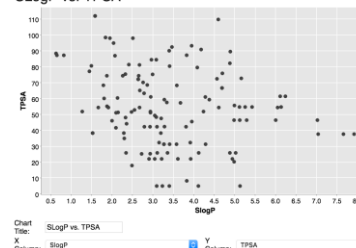
Select File

<no file selected>*

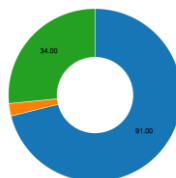
Molecular Properties



SlogP vs. TPSA

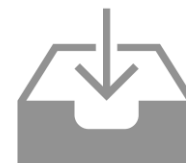


Compounds in Series



Row ID	Molecule	Name	OSM_ID	CHEMBL	Series Tag	Link
Row0		(Z,Z)-5-((1-(3-fluorophenyl)-2,5-dimethyl-1H-pyrazol-3-yl)methylene)-3-phenyl-2-phenylimino[1,2,4]oxadiazole-4-one	OSM-L-01	dep1	Series 1	?
Row1		Molecule	OSM-S-1	?	Series 1	http://lm
Row2		1-(4-fluorophenyl)-2,5-dimethyl-1H-pyrazole-3-carbaldehyde	OSM-S-2	dep1	Series 1	http://lm
Row3		ethyl 1-(4-fluorophenyl)-2,5-dimethyl-1H-pyrazole-3-carboxylate	OSM-S-3	dep1	Series 1	http://lm

Page 5



Page 1: Select path

Page 1a: Select File to Upload

Calculate Properties

Node 99

Page 2: Choose descriptors

Joiner

Page 3: Visualize results

Row Filter

Page 4: View selected compounds

Page 5: File Download

Page 1

Page 2

WebPortalChemistryExample

Data Source

- ☐ Upload SDF file
- ☒ Example data

Back

Next

Excludes

Back

Next

Includes

SlogP
TPSA
AMW
NumLipinskiHBA
NumLipinskiHBD
NumAtoms
NumRings

Page 4

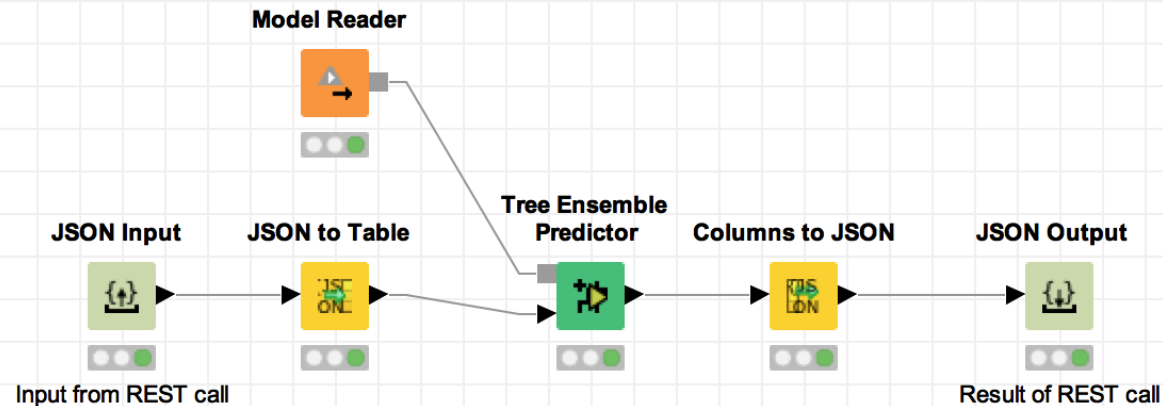
Row ID	Molecule	Name	OSM_ID	CHEMBL	Series Tag	Link
Row0		(Z,Z)-5-((1-(3-fluorophenyl)-2,5-dimethyl-1H-pyrazol-3-yl)methylene)-3-phenyl-2-phenylimino[1,2,4]oxadiazole-4-one	OSM-L-01	dep1	Series 1	?
Row1		Molecule	OSM-S-1	?	Series 1	http://lm
Row2		1-(4-fluorophenyl)-2,5-dimethyl-1H-pyrazole-3-carbaldehyde	OSM-S-2	dep1	Series 1	http://lm
Row3		ethyl 1-(4-fluorophenyl)-2,5-dimethyl-1H-pyrazole-3-carboxylate	OSM-S-3	dep1	Series 1	http://lm

KNIME Server REST API

Predictor workflow that is exposed as RESTful web service.

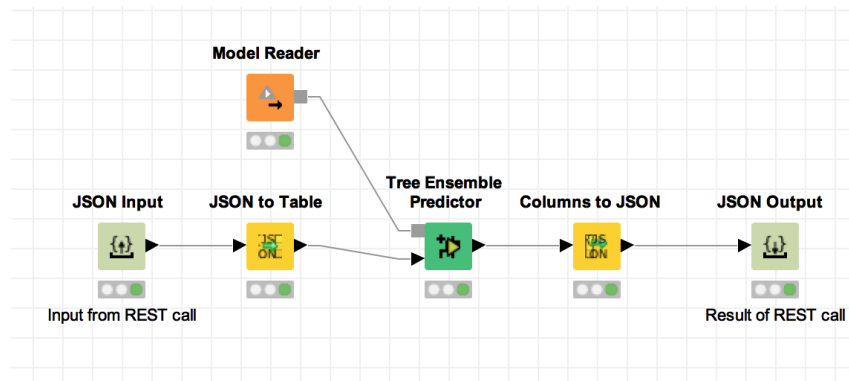
Typically an instance of this workflow would be instantiated once (or multiple times to increase performance) and then re-used for new data as it becomes available.

The JSON In/Output nodes define the endpoints. Quickform nodes can be used to define additional parameters (not used here).



ChemAxon tools and REST API

KNIME Server
HTTP POST

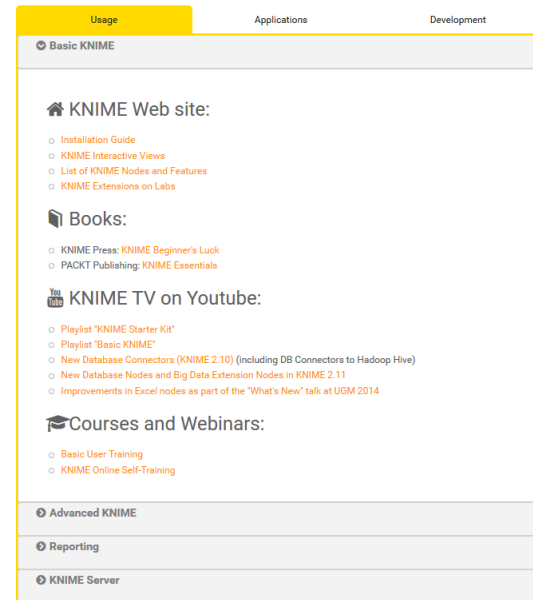


Result

Great for:
New method development
Rapid Prototyping
Integrating heterogeneous data

KNIME Analytics Platform: Try it Now!

1. Download from
www.knime.com
2. Browse the KNIME Learning Hub at
www.knime.com/learning-hub
3. Check out KNIME Press
“KNIME Beginner’s Guide”
4. Become active on our Forum!



The KNIME® trademark and logo and OPEN FOR INNOVATION® trademark are used by KNIME.com AG under license from KNIME GmbH, and are registered in the United States. KNIME® is also registered in Germany.