

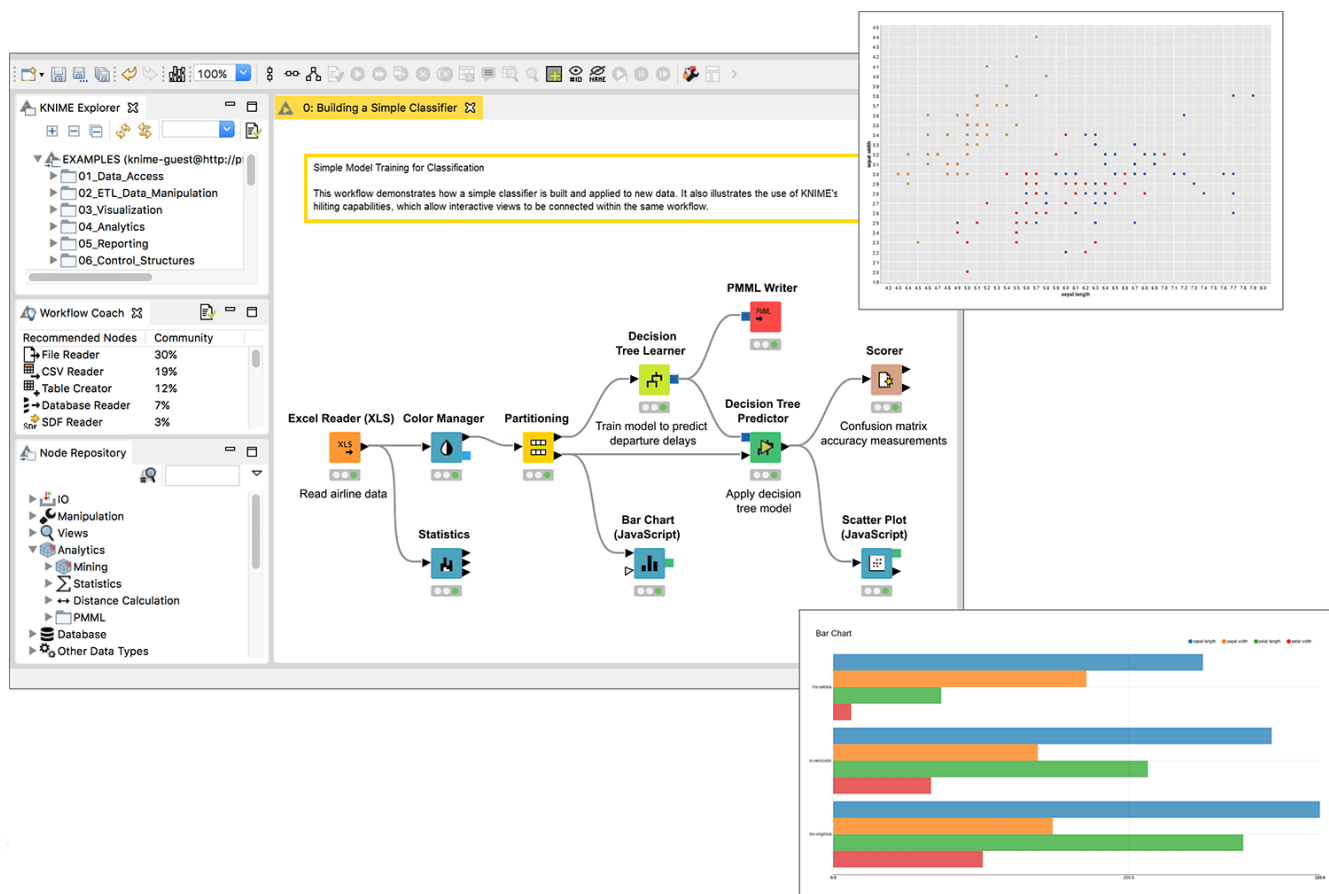
KNIME

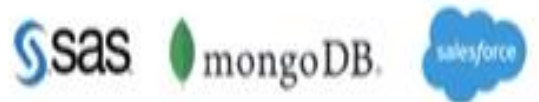
Аналитическая платформа KNIME

Открытая, интуитивная, интегративная наука о данных.



Аналитическая платформа KNIME
KNIME Analytics Platform - это программное обеспечение с открытым исходным кодом для создания приложений и сервисов для обработки данных. Интуитивно понятный, открытый и постоянно интегрирующий новые разработки, KNIME обеспечивает понимание данных и разработку рабочих процессов в области науки о данных и повторно используемых компонентов, доступных каждому.





External
Data



ML / DL
Libraries



Reporting &
Visualization

Scripting
Languages



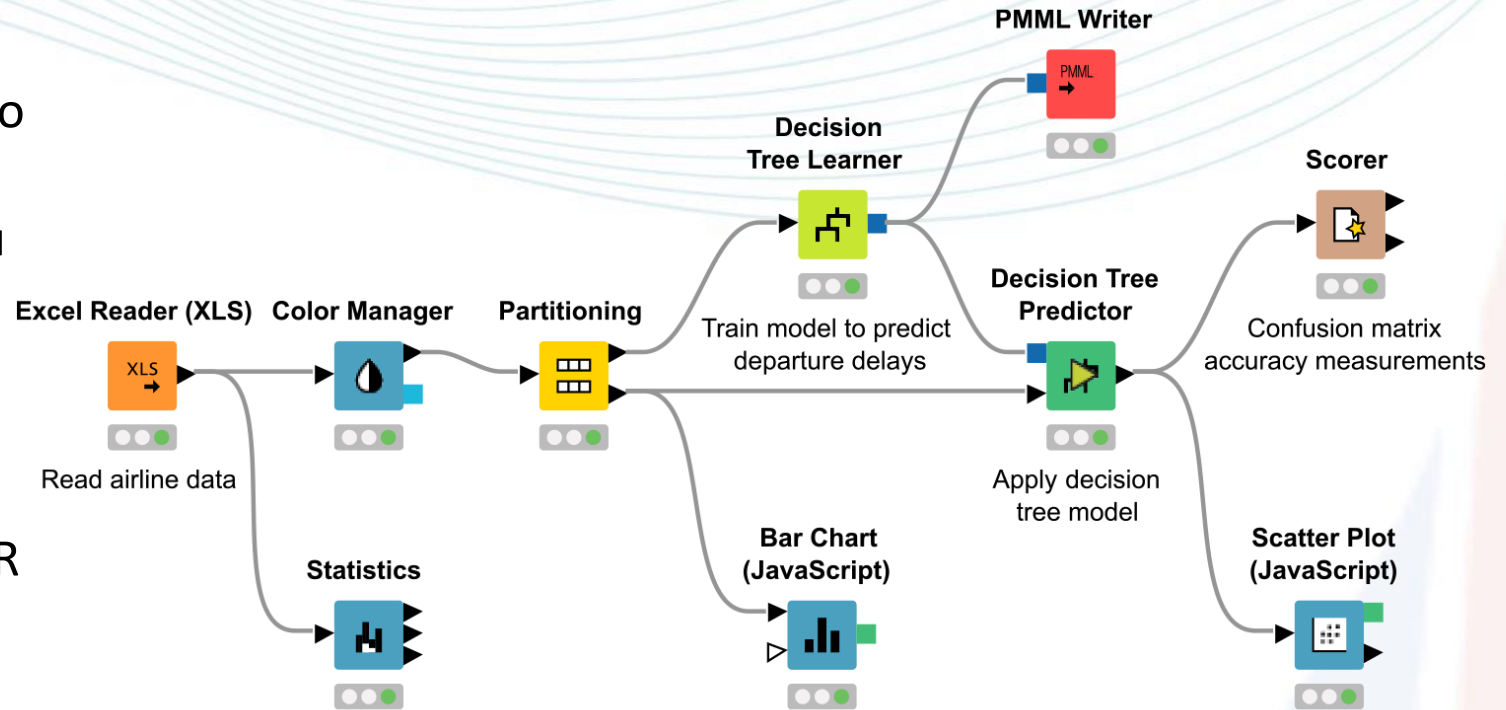
Distributed /
Cloud Execution



Создание сквозных рабочих процессов в науке о данных

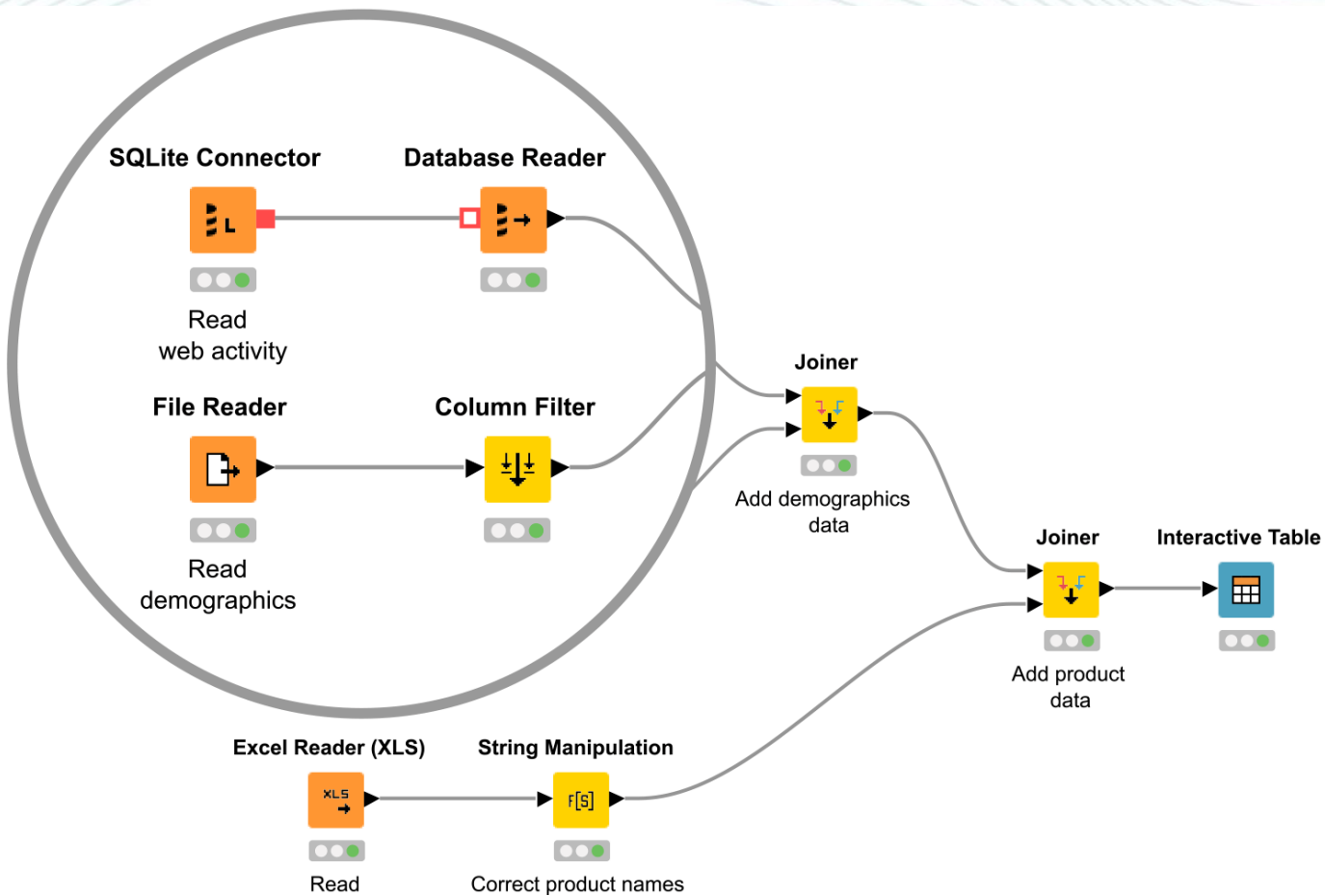
Создание рабочих процессов в интуитивно понятным графическим интерфейсе в стиле перетаскивания без необходимости кодирования.

Смешивайте разные инструменты с собственными узлами KNIME в одном рабочем процессе, включая сценарии на R & Python, машинное обучение или соединители с Apache Spark.



Выберите из более чем 2000 модулей («узлов») для создания вашего рабочего процесса. Смоделируйте каждый шаг вашего анализа, контролируйте поток данных и убедитесь, что ваша работа всегда актуальна. Выберите один из сотен общедоступных примеров рабочих процессов или используйте встроенный инструктор рабочих процессов, который поможет вам в создании рабочего процесса.



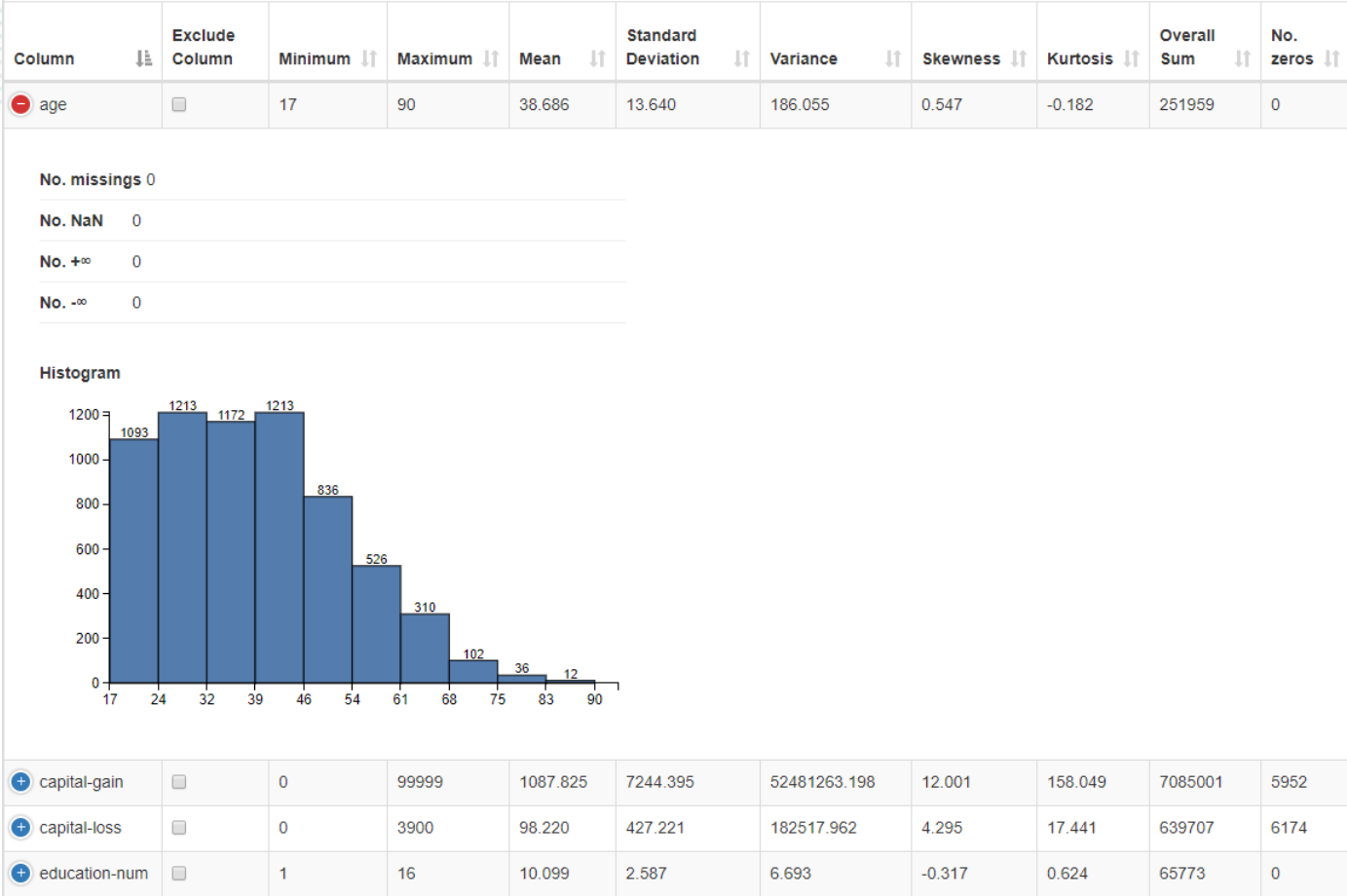


Смешайте данные из любого источника

Открывайте и комбинируйте простые текстовые форматы (CSV, PDF, XLS, JSON, XML и т. д.), неструктурированные типы данных (изображения, документы, сети, молекулы и т. д.) или данные временных рядов.

Подключайтесь к множеству баз данных и хранилищ данных для интеграции данных из Oracle, Microsoft SQL, Apache Hive и других. Загрузите файлы Avro, Parquet или ORC из HDFS, S3 или Azure. Доступ и получение данных из таких источников, как Twitter, [AWS S3](#), Google Sheets и [Azure](#).





Формируйте свои данные

Получите статистику , включая среднее значение, квантили и стандартное отклонение, или примените статистические тесты для проверки гипотезы. Интегрируйте сокращение размерности, корреляционный анализ и многое другое в свои рабочие процессы. Агрегируйте, сортируйте, фильтруйте и объединяйте данные на локальном компьютере, в базе данных или в распределенных средах с большими данными.

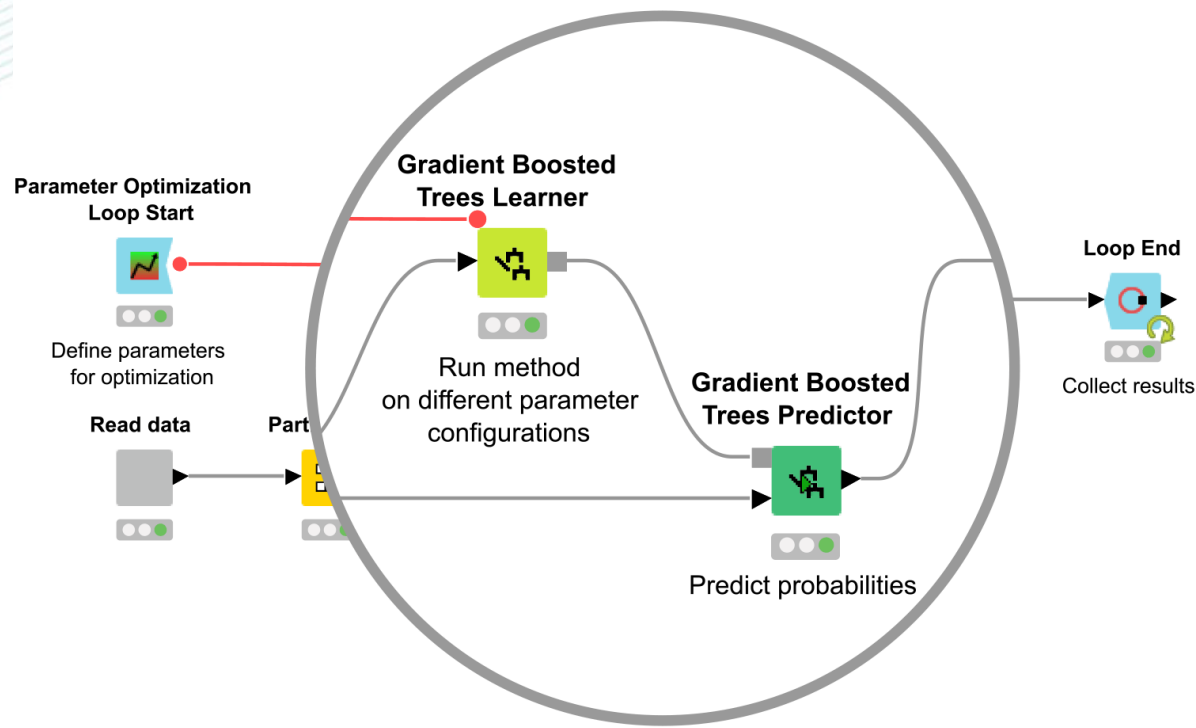
Очистите данные с помощью нормализации, преобразования типов данных и обработки пропущенных значений. Обнаружение значений вне диапазона с помощью алгоритмов обнаружения выбросов и аномалий.

Извлеките и выберите функции (или создайте новые), чтобы подготовить ваш набор данных для машинного обучения. Управляйте текстом, применяйте формулы к числовым данным и применяйте правила для фильтрации или маркировки образцов.



Использование машинного обучения и искусственного интеллекта

Создавайте модели машинного обучения для классификации, регрессии, сокращения размерности или кластеризации, используя передовые алгоритмы, включая глубокое обучение, древовидные методы и логистическую регрессию.



Оптимизируйте производительность модели с помощью оптимизации гиперпараметров, повышения, суммирования, укладки или создания сложных ансамблей.

Проверяйте модели, применяя показатели производительности, включая точность, R^2 , AUC и ROC. Выполните перекрестную проверку, чтобы гарантировать стабильность модели.

Делайте прогнозы, используя проверенные модели напрямую или с помощью ведущих в отрасли PMML, в том числе в Apache Spark.



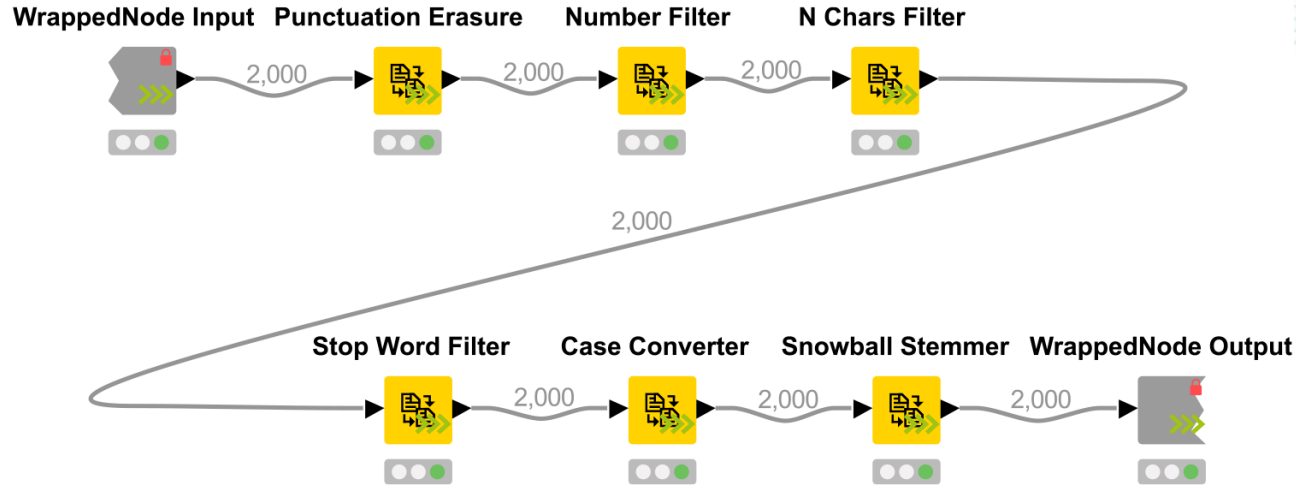


Узнайте и поделитесь своими мыслями
Визуализируйте данные с помощью классических (гистограмма, точечная диаграмма), а также расширенных диаграмм (параллельные координаты, солнечные лучи, график сети) и настраивайте их в соответствии с вашими потребностями.
Отобразите сводную статистику о столбцах в таблице KNIME и отфильтруйте все, что не имеет значения.

Экспортируйте отчеты в формате PDF, Powerpoint или другие форматы для представления результатов заинтересованным сторонам.

Храните обработанные данные или результаты аналитики во многих распространенных форматах файлов или базах данных.





Шкала исполнения с требованиями

Создание прототипов рабочих процессов для изучения различных подходов к анализу. Проверяйте и сохраняйте промежуточные результаты, чтобы обеспечить быструю обратную связь и эффективное обнаружение новых креативных решений.

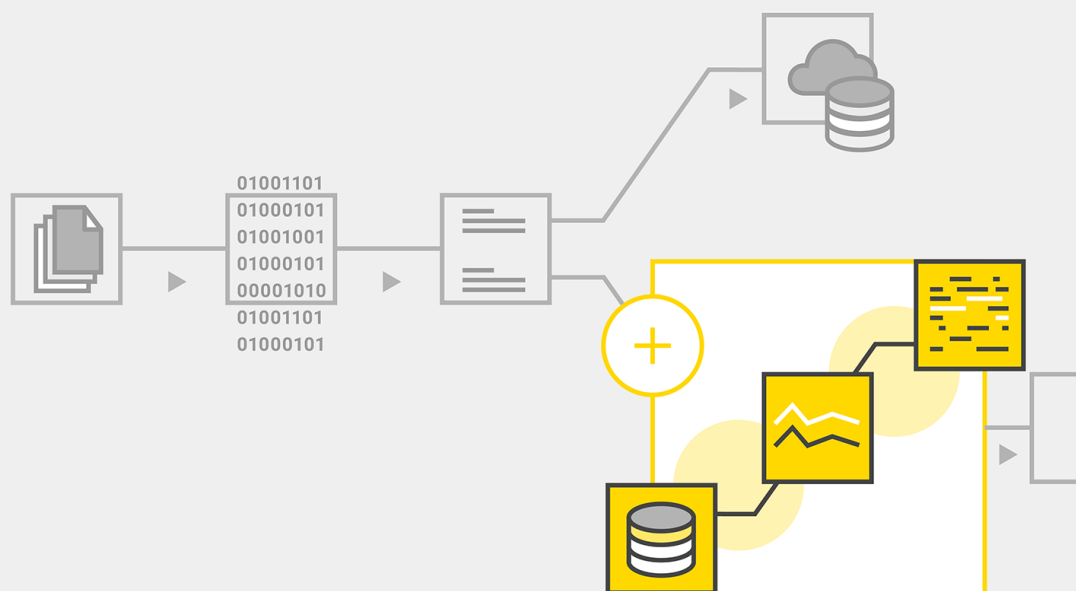
Повышение производительности рабочего процесса за счет потоковой передачи в памяти и многопоточной обработки данных.

Используйте возможности обработки в базе данных или распределенных вычислений в Apache Spark для дальнейшего повышения производительности вычислений.



Расширения KNIME

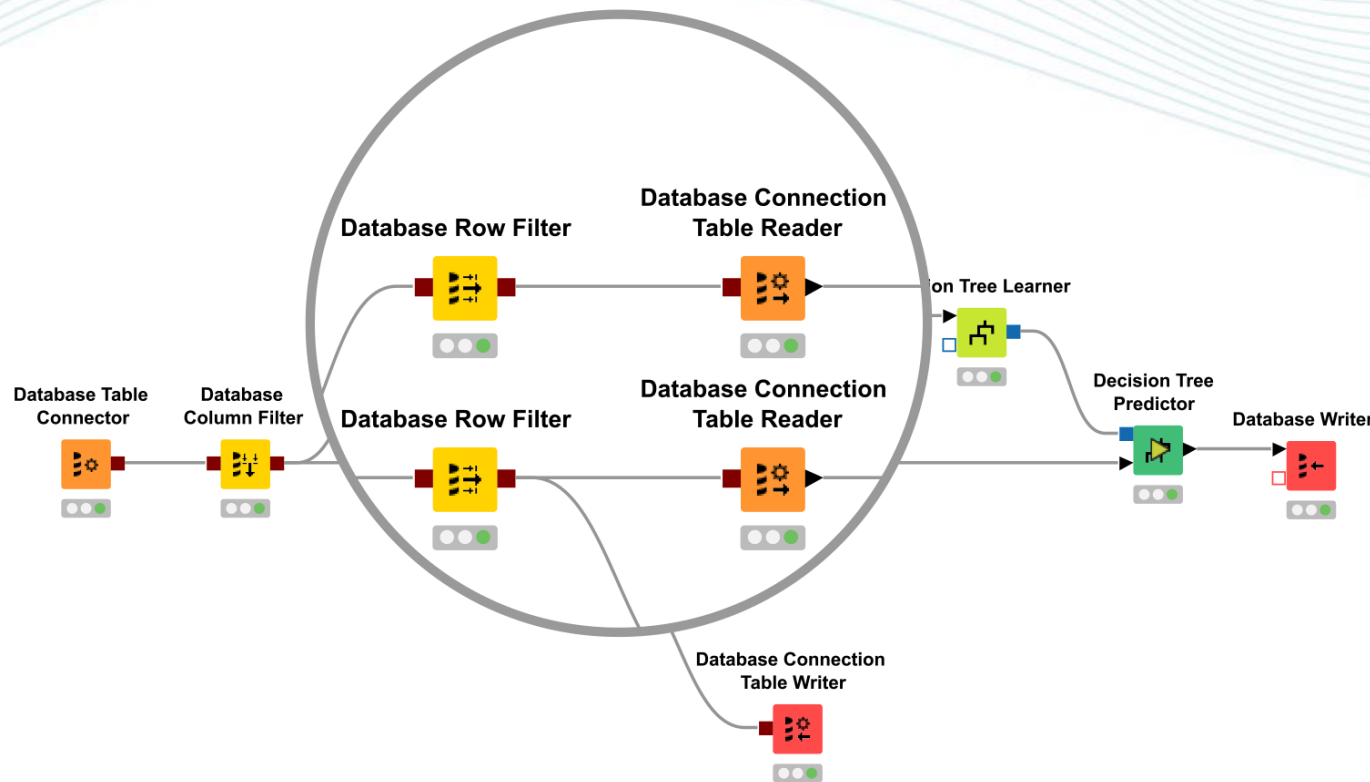
Быстрый и гибкий способ расширить вашу платформу для обработки данных.



Расширения KNIME

Расширения с открытым исходным кодом для платформы KNIME Analytics разрабатываются и поддерживаются KNIME и предоставляют дополнительные функциональные возможности, такие как доступ к сложным типам данных и их обработка, а также добавление расширенных алгоритмов машинного обучения.





Базы данных

Интегрируйте данные из Amazon Athena и Redshift, H2, Hive, Impala, Microsoft SQL, MySQL, PostgreSQL, Vertica и многих других.

Удобно добавлять дополнительные коннекторы для JDBC-совместимых баз данных.

Организовывайте сложные операторы SQL визуально и содержательно, используя парадигму визуального программирования в KNIME.

Смешивайте и сопоставляйте рабочие процессы обработки в базе данных с собственными узлами KNIME и инфраструктурами больших данных, включая Apache Spark.

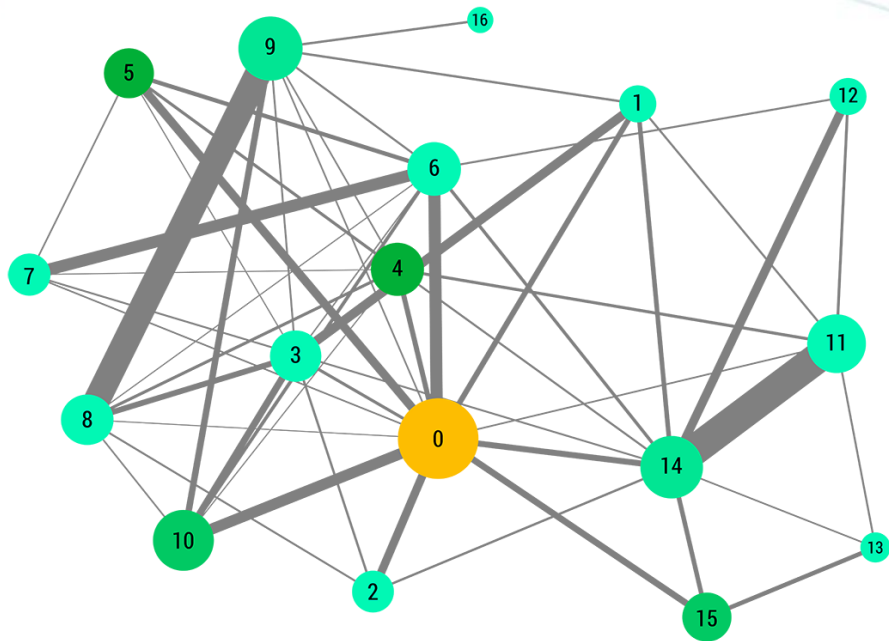


Загрузка документов из различных форматов , включая PDF, DOCX, DOC, PubMed, DML, и многое другое.

Обогащайте текстовые данные распознаванием и маркировкой именованных объектов.

Извлекайте числовые описания из документов, используя различные методы (Word2Vec, Document Hashing и т. Д.), чтобы применить машинное обучение к документам.





Найти отношения в сети

Извлечение информации из атрибутивных графиков, таких как социальные сети, соавторские сети и т.д.

Работа с большими сетями с узлами для создания, создания, управления, анализа и визуализации этих сетей.

Интерактивный контроль сетей, а также отправка и получение сетей из внешних программ и из них.

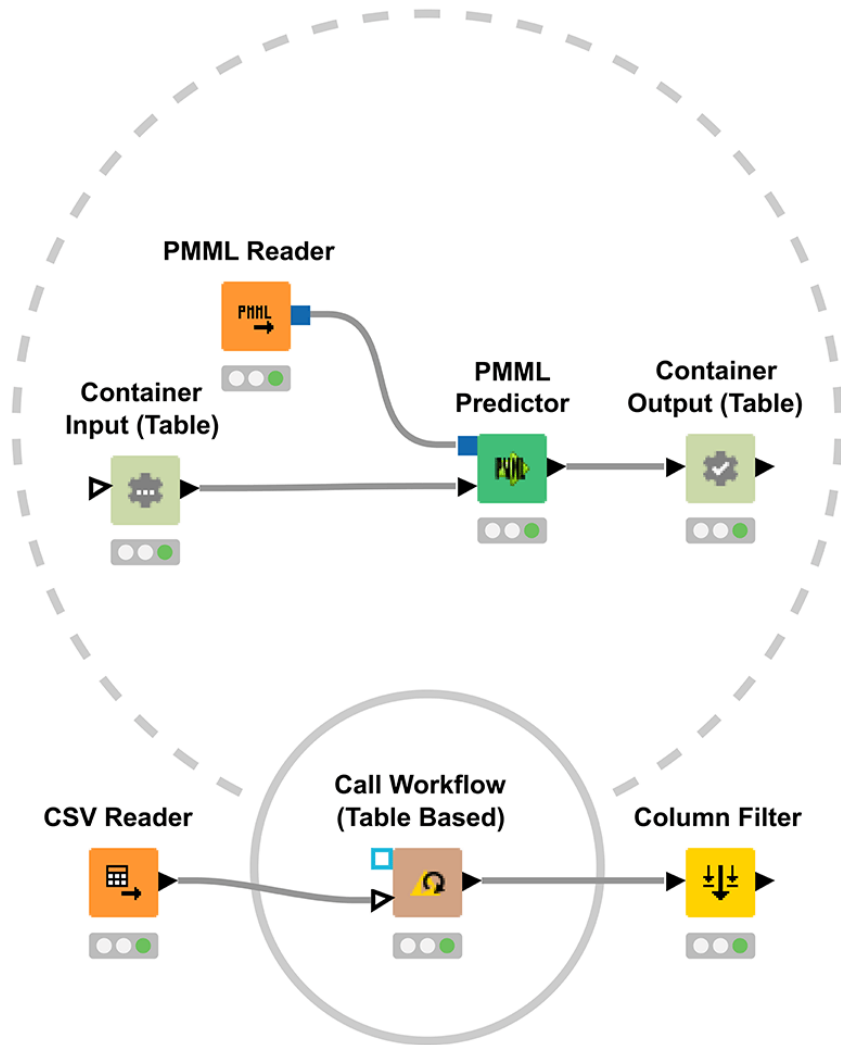


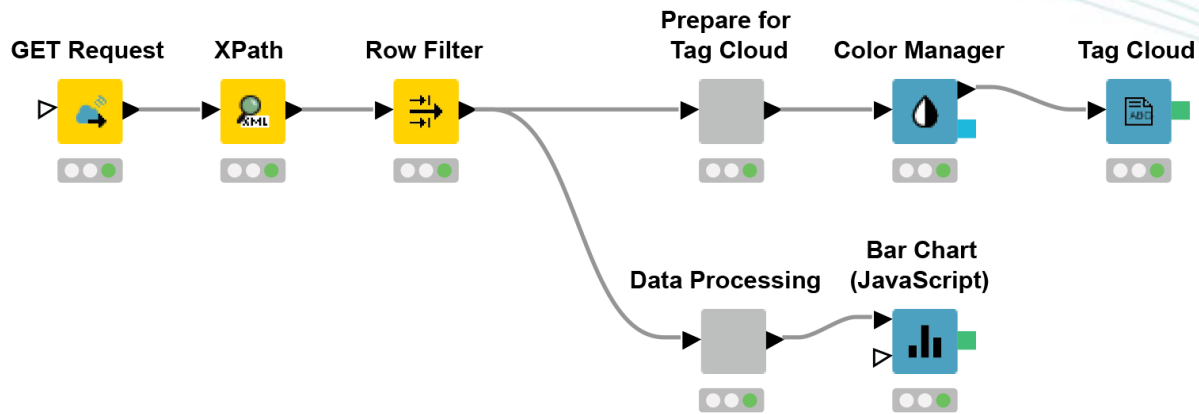
Производительность

Сравните разные версии ваших рабочих процессов, чтобы точно определить любые изменения.

Организируйте повторно используемые части рабочих процессов в метаноды и сохраните их в локальном рабочем пространстве. Повторно используйте по желанию и синхронизируйте их между рабочими процессами.

Управляйте и автоматизируйте локальные рабочие процессы, вызывая рабочие процессы из рабочих процессов, добавляя еще один уровень гибкости в ваш инструментарий.





И больше...

Отправляйте запросы в службы REST и интегрируйте их ответы в рабочий процесс.

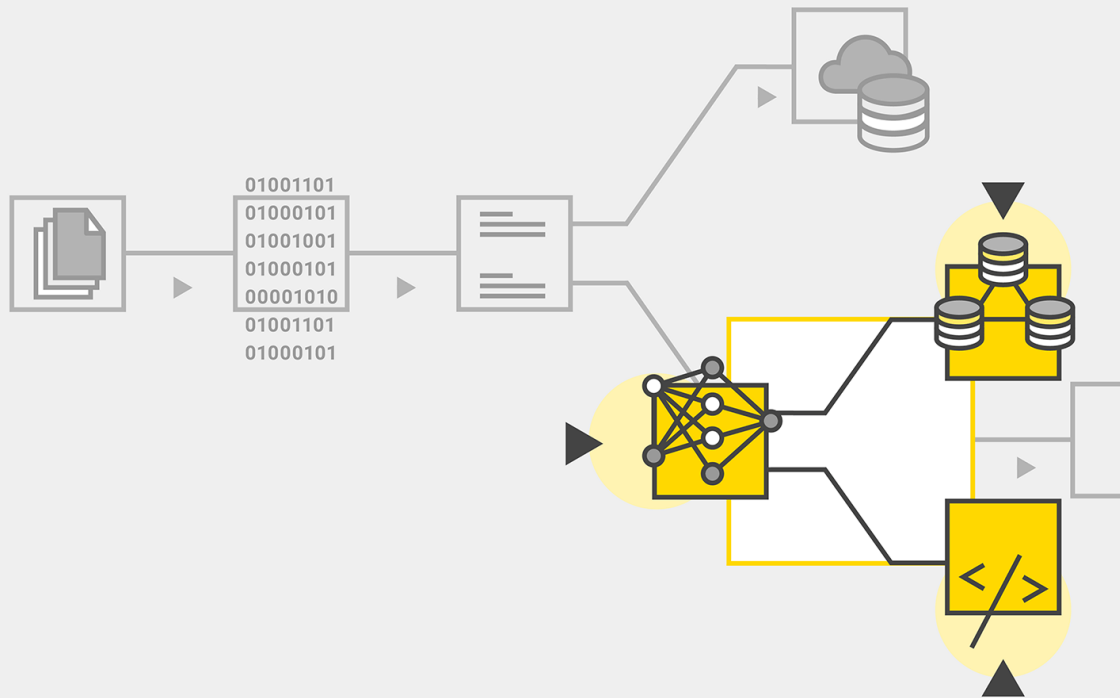
Импорт и экспорт моделей PMML для высокопроизводительной оценки и гарантированного повторного использования. Вы даже можете комбинировать их для создания сложных ансамблей.

Обработывайте, анализируйте и прогнозируйте данные временных рядов с помощью скользящего среднего, ARIMA, оконных функций и многого другого.



KNIME Integrations

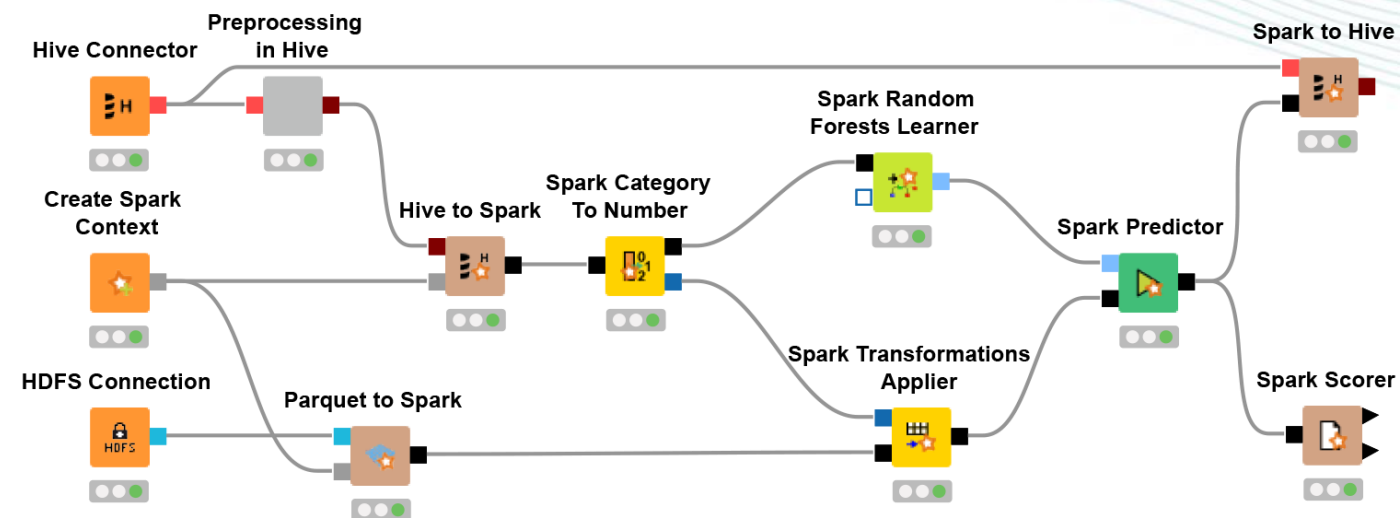
Интеграция больших данных, машинного обучения, искусственного интеллекта, сценариев и многого другого.



KNIME Integrations

Интеграции с открытым исходным кодом для платформы KNIME Analytics (также разработанной и поддерживаемой KNIME) обеспечивают беспрепятственный доступ к крупным проектам с открытым исходным кодом, таким как Keras для глубокого обучения, H2O для высокопроизводительного машинного обучения, Apache Spark для обработки больших данных, Python и R для сценариев, и больше.





Большие данные

Импортируйте, экспортируйте и получайте доступ к данным в Hive, Impala или HDFS с помощью KNIME Analytics Platform.

Создавайте и запускайте приложения Apache Spark на платформе KNIME Analytics. Визуальное программирование позволяет проводить анализ больших данных без кода, а сценарии заданий позволяют осуществлять подробный контроль при необходимости.

Проводите прогнозную аналитику и оценку в Apache Spark с использованием моделей PMML и интегрируйте сложную статистику и машинное обучение с Spark MLlib.

Смешивать и сопоставлять локальное выполнение и выполнение Hadoop в одном рабочем процессе.



Расширения больших данных

Расширения больших данных KNIME объединяют возможности Apache Hadoop и Apache Spark с платформой KNIME Analytics и сервером KNIME. Наше программное обеспечение устраняет путаницу с большими данными, делая их доступными в нашей знакомой аналитической среде. Эти расширения состоят из двух дополнительных библиотек узлов:

- **Коннекторы больших данных KNIME** позволяют импортировать / экспортировать данные HDFS и выполнять аналитику SQL в Hive и Impala.



- **Расширение KNIME для Apache Spark** позволяет создавать и запускать приложения Apache Spark на платформе KNIME Analytics или на сервере KNIME, используя возможности масштабируемой аналитики. Чтение / запись данных в HDFS, Hive и Impala из Apache Spark.



KNIME Analytics Platform or KNIME Server

KNIME Big Data Connectors

KNIME Extension for Apache Spark

HDFS
DataHive / Impala
Queries via
JDBCApache
Spark Jobs
via HTTP(S)

Hadoop Cluster

Мощная комбинация

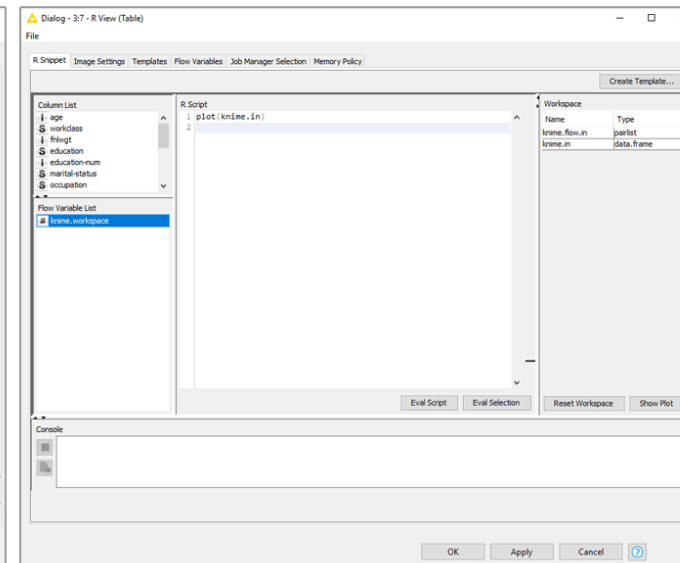
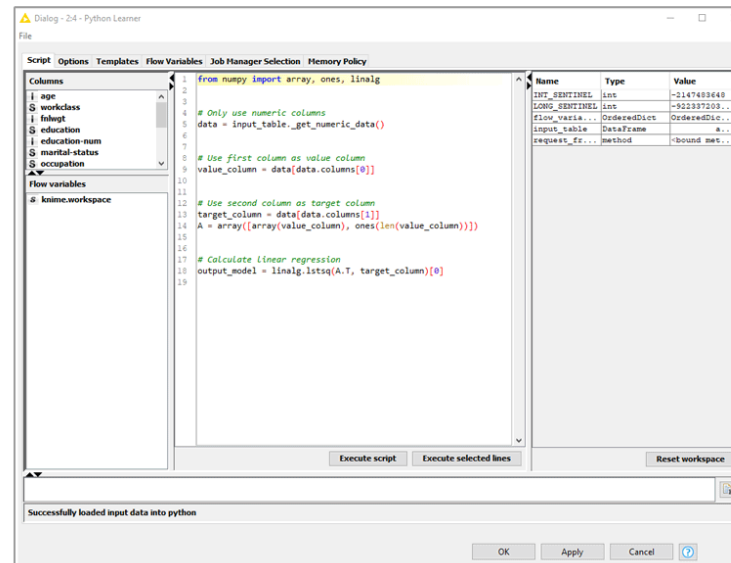
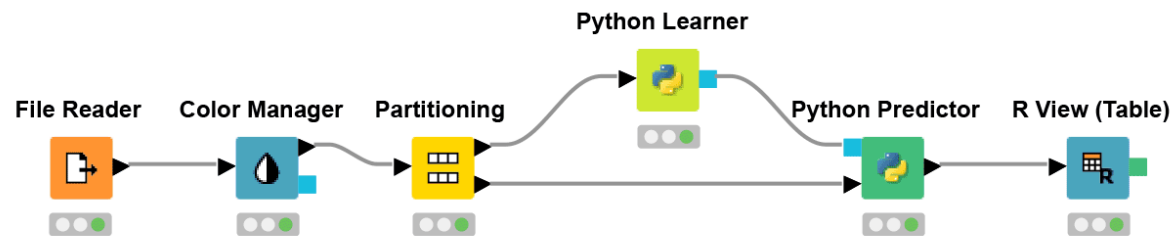
Расширения KNIME для больших данных предоставляют знакомый, простой в использовании графический подход к проблемам больших данных. Эти библиотеки объединяют возможности KNIME Analytics Platform и Hadoop, чтобы расширить преимущества обоих.

- Запросы больших данных в стиле SQL
- Сложный интеллектуальный анализ данных
- Расширенная прогностическая аналитика
- Обработка в памяти
- Обширный дополнительный функционал

Расширенные возможности

- Подключайтесь к популярным дистрибутивам Hadoop
- Интегрируйте Apache Spark с более чем 2000 собственных узлов KNIME, используя знакомые рабочие процессы KNIME
- Смешивать и сочетать удаленные и распределенные вычисления по мере необходимости
- Импортируйте прогностические модели в Apache Spark с моделями PMML, сгенерированными из рабочих процессов KNIME
- Включить популярный набор алгоритмов машинного обучения через интеграцию MLlib



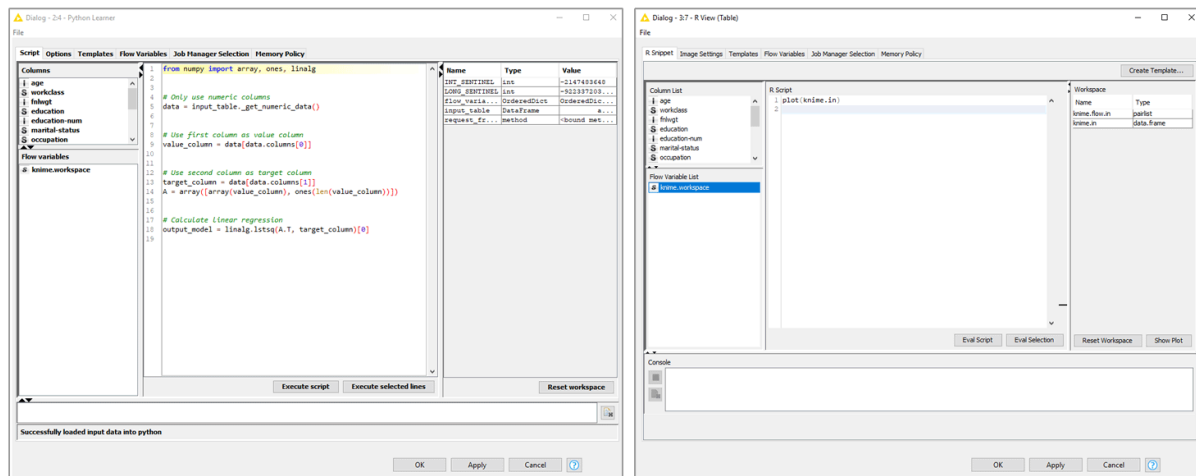
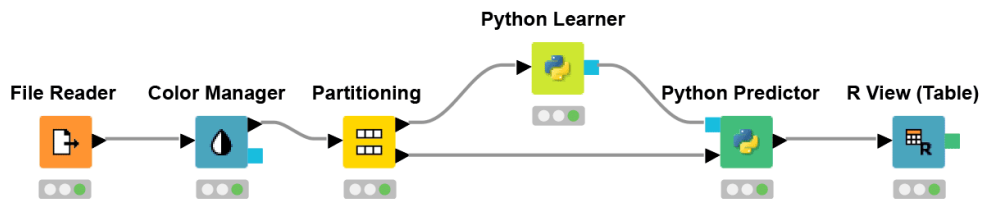


R и Python Scripting

Добавляйте пользовательские функциональные возможности с родным R, Python (версии 2 и 3) и возможностями сценариев Java - от пользовательских заданий Apache Spark до визуализаций или расширенной аналитики и машинного обучения.

Без проблем запускайте сценарии в сочетании с другими узлами KNIME в рамках одного рабочего процесса. Документируйте отдельные шаги, позволяющие осуществлять масштабное развертывание





R и Python Scripting

Платформа KNIME уже давно имеет хорошую интеграцию с Python. В Knime есть две особенно интересные функции, связанные с Python:

1. Можно использовать код Python, из блокнотов **Jupyter**, в узлах сценариев Python в Knime.
2. И наоборот вы можете выполнять рабочие процессы Knime напрямую из Python. Если вы работаете в блокноте **Jupyter**, вы также можете получить (статический) вид рабочего процесса в блокноте.

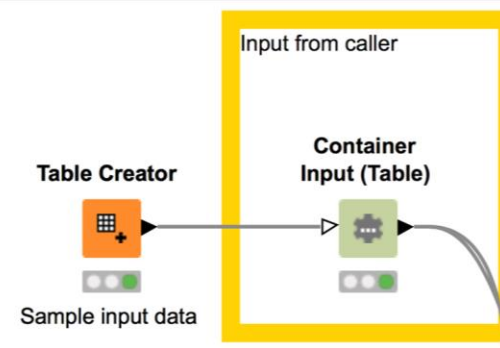
```
In [1]: import pandas as pd
import knime
knime.executable_path = "c:/Users/glandrum/Desktop/KNIME_3.7/knime.exe"
```

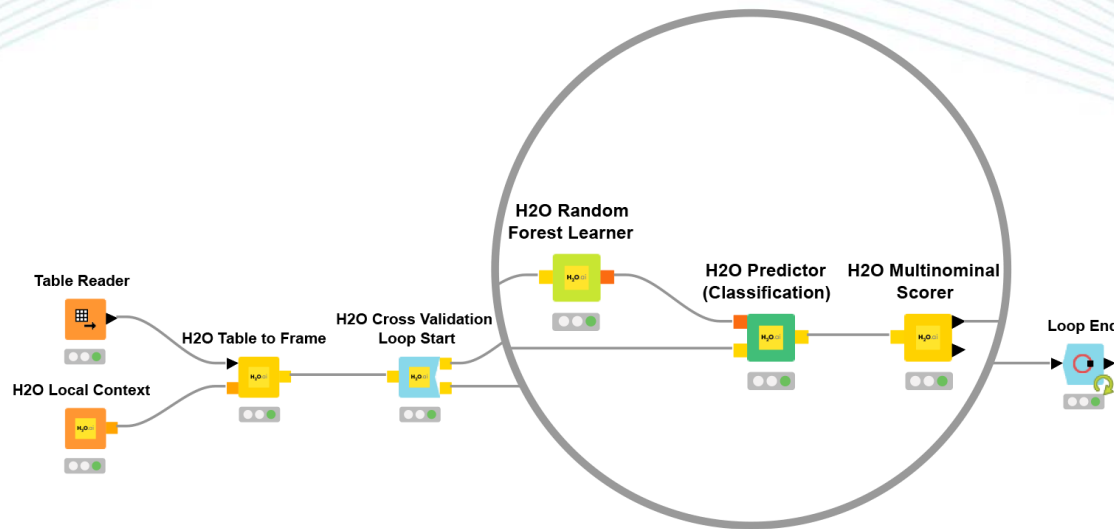
```
In [2]: workspace = "c:/Users/glandrum/KNIME workspaces/KNIME nightly"
```

Now pick a workflow and make sure it's the right one by displaying an SVG of the workflow:

```
In [3]: workflow = "KNIME Stories/Jupyter KNIME Jupyter/Fetch ChEMBL Target Data"
knime.Workflow(workflow_path=workflow,workspace_path=workspace)
```

Out[3]:





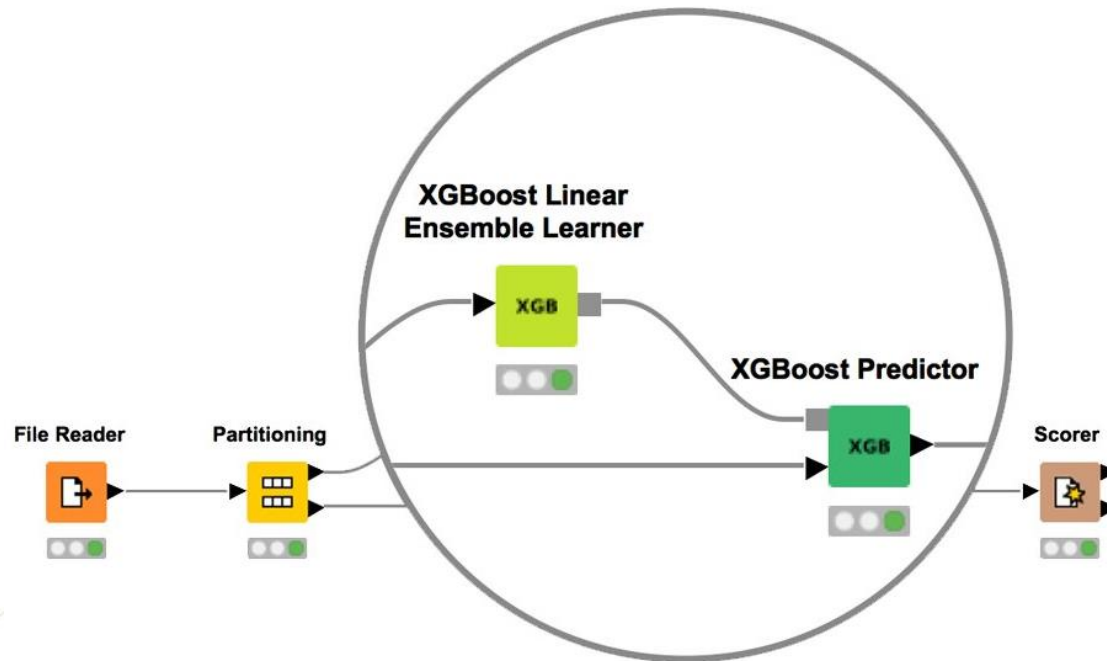
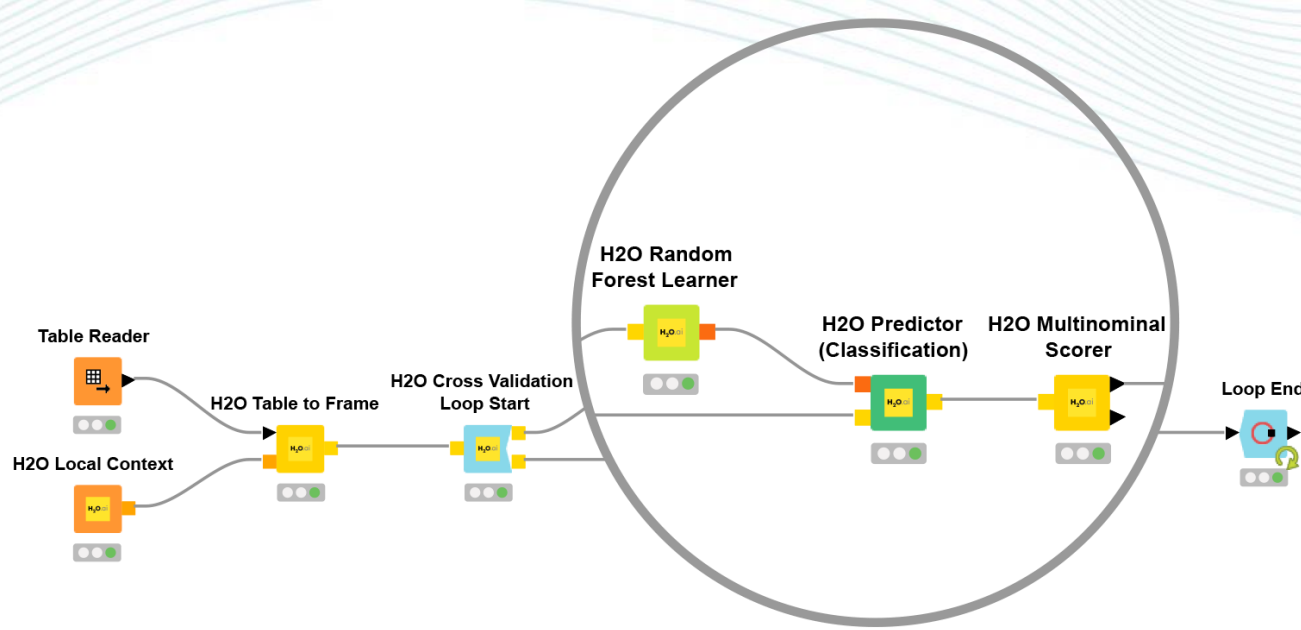
Машинное обучение H2O

Воспользуйтесь преимуществами машинного обучения H2O и выберите один из множества высокопроизводительных алгоритмов (деревья с градиентным усилением, обобщенные линейные модели, случайный лес и т. д.).

Обучайте и проверяйте модели в H2O, используя split данных, перекрестную проверку, биномиальную и полиномиальную оценку.

Интеграция с существующими узлами KNIME для подготовки и очистки данных, визуализации и(или) оптимизации гиперпараметров, объединяя их напрямую с функциональностью H2O.

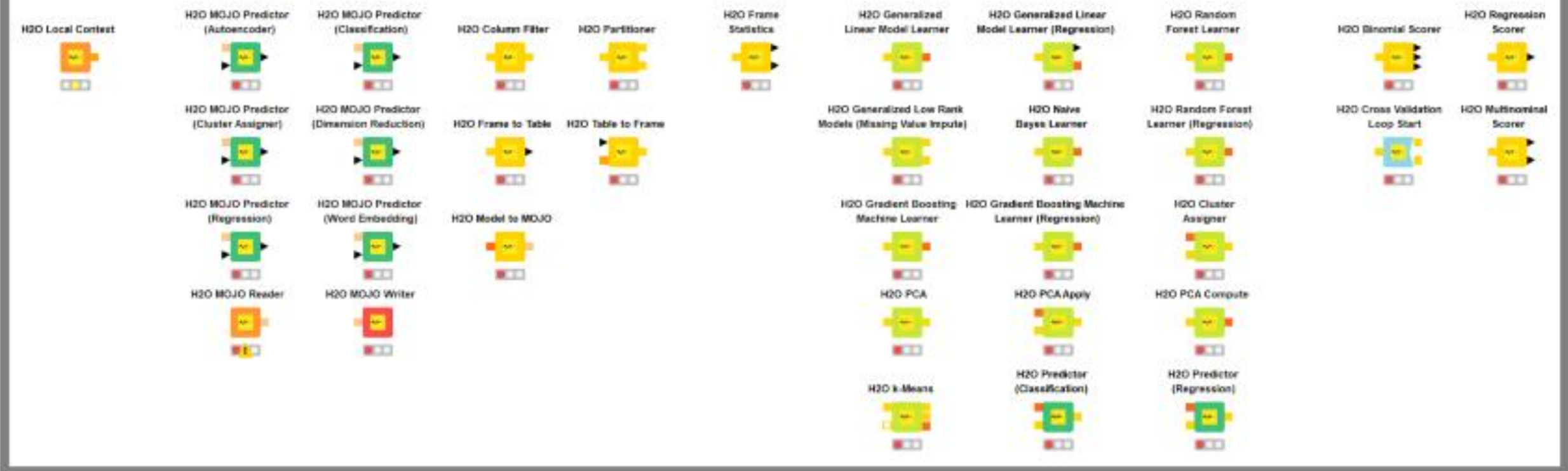




За счет интеграции с приложениями с открытым исходным кодом, можно воспользоваться преимуществами машинного обучения популярных библиотек машинного обучения, таких как экстремальный градиентный бустинг XGBoost и H2O и выбирать один из множества высокопроизводительных алгоритмов (деревья с градиентным бустингом, обобщенные линейные модели, случайный лес и другие). Интеграция H2O и XGBoost с узлами KNIME для подготовки, очистки данных и визуализации обеспечивает наилучшие результаты.

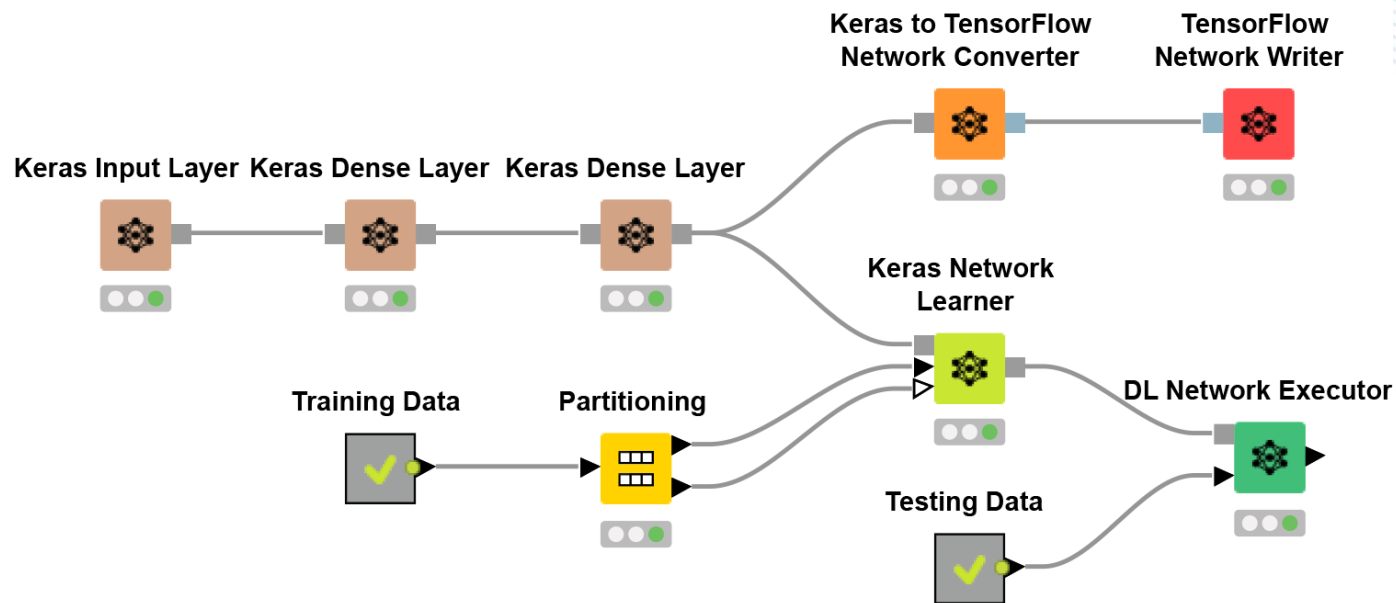


Available nodes of the KNIME H2O Extensions



Все доступные узлы в расширении KNIME H2O





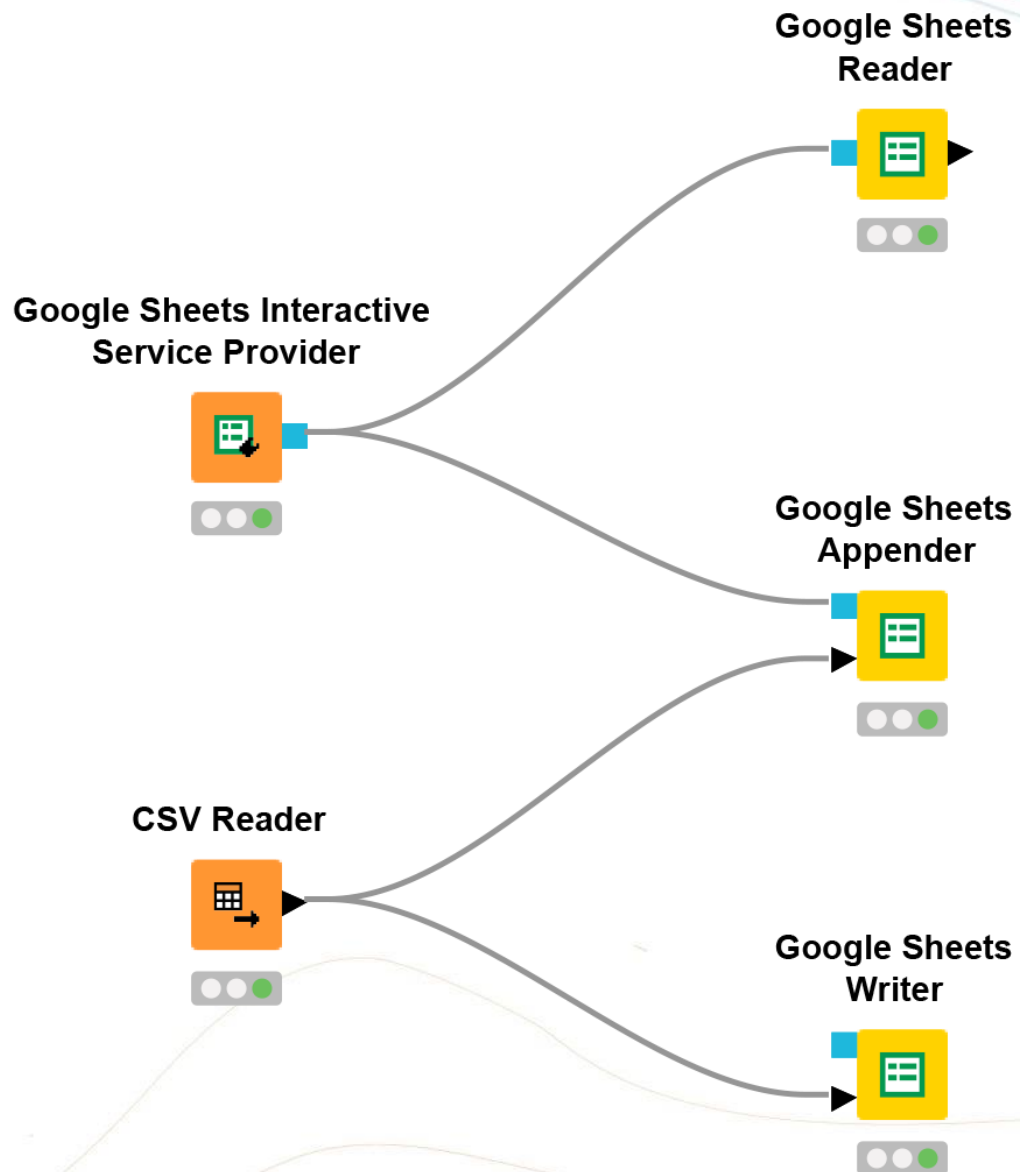
Глубокое обучение

Загружайте, создавайте, редактируйте, обучайте и моделируйте глубокие нейронные сети в рамках KNIME Analytics Platform.

Получите доступ ко множеству самых современных сред глубокого обучения, таких как TensorFlow, CNTK и другие, через Keras или DeepLearning4J.

Точная настройка обученных сетей для вашей задачи анализа. Богатый выбор неструктурированных (текст, изображения и т. д.) и структурированных типов данных может быть непосредственно использован для обучения и прогнозирования.





Google Sheets

Считывание данных из Google Sheets, записывать информацию на новые листы или изменять существующие листы. Выполняйте различные задачи, такие как чтение или добавление заголовков, подстановка пропущенных значений и автоматическое открытие Google Sheets. Войдите непосредственно из конфигурации узла или предоставьте учетные файлы (если предпочитаете).



