

Software: KNIME

Gruppe: 3

Mitglieder: Jonas Bahr, Sebastian Diem, Jonas Kromolan,
Christian Pecher

Vorbereitung

Die Vorbereitung umfasst ausschließlich das Downloaden der entsprechenden KNIME-Konfiguration (64 Bit vs. 32 Bit)

Installation

Die Installation von KNIME erfolgt über das Entpacken der gedownloadeten Datei und der anschließenden Auswahl des KNIME-Workspaces.

Einarbeitung

KNIME stellt jeden Arbeitsschritt in einem Workflow mit entsprechenden Nodes dar. Diese werden per Drag and Drop angeordnet, wodurch die Lernkurve relativ gering ist. Für die Anbindung der Hive-DB kann man entweder eine JDBC-Connection wählen oder man verwendet den entsprechenden Big-Data-Connector von KNIME, der über die einfach zu installierenden Extensions zu Verfügung gestellt wird. Die Daten können sehr einfach über die entsprechenden Nodes aufbereitet werden und in jedem dieser Arbeitsschritte eingesehen werden. Dabei dauert das Finden der entsprechenden Node manchmal länger als die eigentliche Ausführung. Jedoch konnte für unsere Zwecke mit Hilfe der ausführlichen Dokumentation immer eine entsprechende Node gefunden werden. Eine Verzweigung bzw. Aufspaltung des Workflows ist ebenfalls möglich. In diesem Fall agieren die verzweigten Nodes als eigenständig und unabhängig voneinander. Das heißt, es muss nicht erneut eine Datenbank-Verbindung für jede neue Verzweigung angelegt werden. Die Nodes sind dabei recht kleingranular angelegt. So kann man in einer Node nach bestimmten Spalten filtern, um diese Spalte umzubenennen benötigt es eine weitere Node. Dadurch entstehen teils sehr lange Workflows, was möglicherweise ein Verständnisproblem auslösen kann. Die Zusammenarbeit im Team erfolgte über Github. Es ist nach Wissensstand der Autoren jedoch nicht möglich Git direkt in KNIME einzubinden. Weitere Möglichkeiten der Zusammenarbeit sind vorhanden und finden sich zum Beispiel in KNIME Teamspace.

Fazit

Knime ist ein leicht zu verstehendes Open-Source-Tool, um Daten zu integrieren, aufzubereiten, zu manipulieren und auszuwerten. Die Bedienung hat man in weniger Selbstversuchen erlernt. Der größte Aufwand liegt in der Regel darin, den korrekten Knoten für die gewünschte Operation zu finden. Darüber hinaus ist das Tool gerade im grafischen Reporting stark kompliziert und benötigt für weiterführende Grafiken komplexe Workflow-Geflächte. Soll hingegen in erster Linie auf Datenebene gearbeitet werden, wo in der Regel auf Visualisierungen verzichtet werden kann, können wir Knime nach unseren Erfahrungen uneingeschränkt empfehlen. Gerade die Workflowmechanik empfinden sehr angenehm und nachvollziehbar. Zur Datenmodifikation kann Knime mit SQL und Excel alle Male mithalten und ist an vielen Stellen sogar einfacher. Die Dokumentation im Tool ist hervorragend, wodurch wir das Programm auch für Laien empfehlen können. Wir konnten jedoch nicht das Verarbeiten riesiger Datenmengen mit Knime analysieren, jedoch sollte die Möglichkeit der Speicherverwaltung bei Knime trotzdem einen schnellen Workflow mit adäquatem Speicherverbrauch gewährleisten.