

KNIME Best Practices Guide

KNIME AG, Zürich, Schweiz

Version 5.7 (letzte Aktualisierung auf)



Inhaltsverzeichnis

Einleitung	1
Was ist KNIME?	2
KNIME Analyseplattform	2
KNIME Hub.	2
Workflow Design Prozess	3
Projektvoraussetzungen	4
Ziele der besten Praktiken	5
Machen Sie Ihren Workflow effizient	6
Namenskonventionen	6
Verwenden Sie relative Dateinamen	7
Dokumentieren Sie Ihre Workflows	7
Strukturieren Sie Ihre Workflows	8
Teilen Sie Komponenten	8
Komponenten für Data Science	8
Erfassung und Aufbereitung von Daten	9
Prüfung und Datenvalidierung	10
Machen Sie Ihren Workflow effizient	11
Effiziente Dateilesungen	11
Effiziente Workflows	11
Messen Sie die Leistung	11
Prozess in der Speicher	11
Ausführung optimieren	11
Columnar Tisch Backend	11
Parallelisieren Sie die Arbeit	11
Effiziente Datenbanknutzung	11
Machen Sie Ihren Workflow sicher	12
Geben Sie die Anrechte	12
Umgang mit sensiblen Daten	12
Datenanonymisierung	12
Geheimnisse: Sichere	12
Glossar.	13
Workflow-Annotationen	13
Node Label	13
Workflow Beschreibung	13

[Hubräume](#page40)[Job](#page40)[.](#page40)[Daten-App-Bereitstellung](#page41)[Zeitplan der Bereitstellung](#page41)[Service-Installation](#page41)[Workflow Service](#page41)

Einleitung

Dieser Leitfaden bietet Best Practices, um effizient mit der KNIME Analytics Platform und auf KNIME Hub. Es zeigt Ihnen, wie Sie Ihre KNIME Workflows beschleunigen, wiederverwenden und sichern können. Auch bietet beste Praktiken für die Aktualisierung und Überwachung von Workflows, die mit Ihrem Ende geteilt werden Benutzer.

Was ist KNIME?

KNIME ist eine Plattform zum Ende der Datenwissenschaft. Es bietet zwei komplementäre Produkte:

KNIME Analytics Platform und KNIME Hub.

KNIME Analytics Plattform

KNIME Analytics Plattform ist eine Open-Source-Software mit einer visuellen Schnittstelle, die gezeigt wird

unten. Es ermöglicht Benutzern, Datenwissenschaftsanalysen zu erstellen. Benutzer können Zugriff, Mischung, Analyse und die Daten visualisieren und Tools und Bibliotheken von Drittanbietern integrieren.



Abbildung 1. Benutzeroberfläche der KNIME Analytics Plattform

Wenn Sie die KNIME Analytics Platform noch nicht heruntergeladen haben, können Sie dies tun

[zum Download](#)

Seite .

KNIME Hubraum

KNIME Hub ist die kommerzielle Software für die Zusammenarbeit und Skalierung von Datenwissenschaft, angeboten

online als KNIME Community Hub oder in die private Infrastruktur eines Unternehmens als

KNIME Business Hub.

KNIME Community Hub

Zur KNIME-Gemeinschaft Hub, Benutzer können Arbeitsbeispiele Workflows der Datenwissenschaft durchsuchen Lösungen oder finden Knoten und Komponenten in ihren KNIME-Projekten. Über die Gemeinschaft Hub, Sammlungen bieten ausgewählte Workflows, Knoten und Links zu einem bestimmten, gemeinsamen Thema. Open Source Erweiterungen bieten zusätzliche Funktionalitäten wie Zugang zu und die Verarbeitung komplexer Datentypen sowie die Hinzufügung von fortschrittlichem maschinellem Lernen Algorithmen.

Darüber hinaus können kleine Gruppen oder Teams von Benutzern mit Team-Plan bezahlte Dienste teilen und an Lösungen in Privaträumen zusammenarbeiten und Workflows ad hoc als Datenapps ausführen oder automatisieren ihre Ausführung.

KNIME Business Hub

KNIME Unternehmen Hub bietet die gleichen Fähigkeiten für die Zusammenarbeit und den Austausch – aber innerhalb der dedizierten Infrastruktur eines Unternehmens. Teams können das Wissen öffentlich über ihre Organisation oder privat mit ausgewählten Benutzern. Zusätzlich können Datenexperten bereitstellen und ihre Workflows überwachen und analytische Modelle als Datenapps und -dienste mit Datenverbraucher.

Workflow Design Prozess: Von KNIME Analytics Plattform für KNIME Hub

Beim Starten Ihres Datenprojekts mit KNIME erstellen Sie einen Workflow in der KNIME Analytics Plattform, die dann auf KNIME Hub hochgeladen werden kann. Die KNIME Analytics Plattform ist wie Sie gestalten Ihren Datenprozess über einen Workflow. Sobald Ihr Workflow fertig ist, können Sie hochladen es zu Ihrer KNIME Hub-Instanz. Dort können Sie es einfach automatisieren oder bereitstellen.

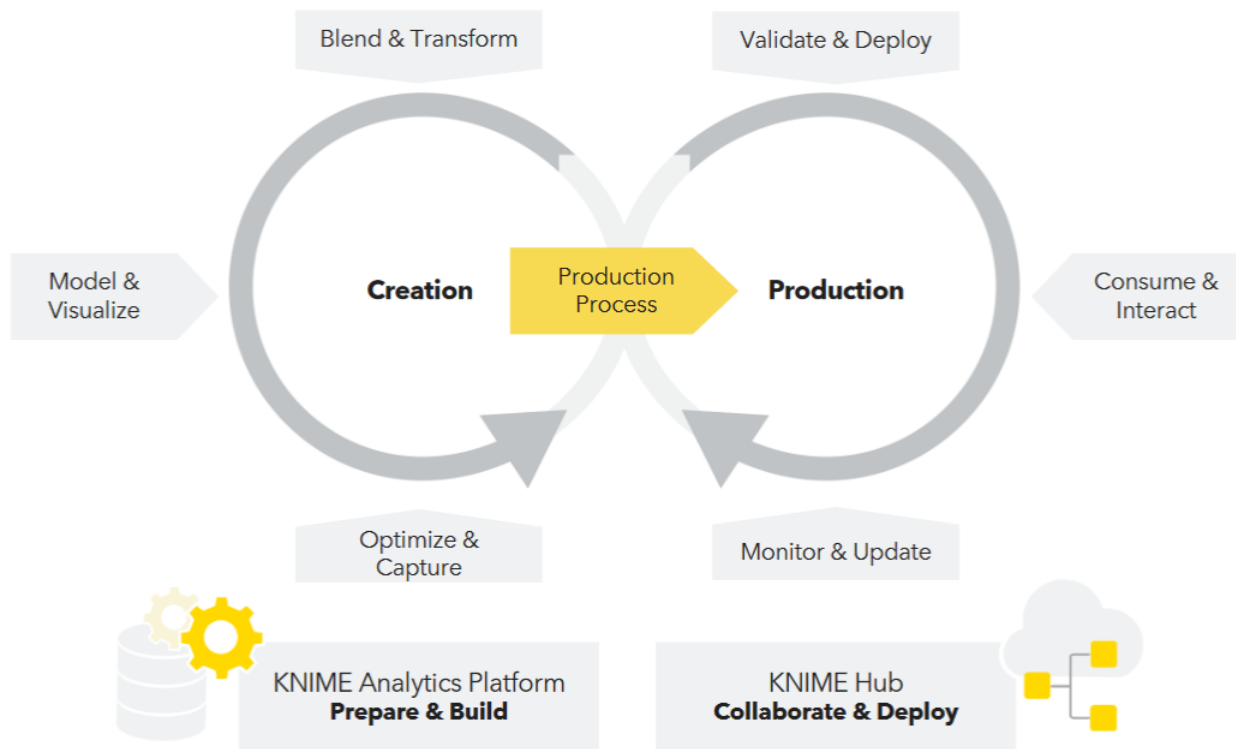


Abbildung 2. Der Lebenszyklus der Datenwissenschaft

Projektvoraussetzungen: Vor dem Bau eines KNIME Arbeitsablauf

Wie bei jedem Projekt, vor dem Aufbau Ihres Workflows, ist es wichtig, den Umfang der den Projekt- oder Anwendungsfall, mit dem Sie arbeiten. Diese Checkliste kann Sie führen:

- ☐ Definieren Sie ein Ziel. Es sollte messbar und erreichbar innerhalb einer bestimmten Frist sein. Für
Beispiel: Reduzieren Sie die Churn-Rate im kommenden Quartal um 15%.
- ☐ Zugriff auf Ihre Daten: Welche Quellen sind erforderlich? Welche Knoten für den Zugriff unerlässlich sind sie?
- ☐ Definieren Sie Ihr Publikum. Ein Workflow wird selten nur von der Person, die es gebaut. Frage stellen selbst, die die letzten Ergebnisse konsumiert. Was sollte ihr Mitnehmen von dir?
Einsichten? Definieren Sie eine Methode, um Ihre Erkenntnisse zu teilen. Es sollte geeignet sein, die Daten und Ihre daraus gezogenen Schlussfolgerungen. Beispielsweise könnte ein Verfahren sein:
 - ☐ Ein geplanter Workflow, der regelmäßig einen Bericht erstellt und sendet
 - ☐ Ein geplanter Workflow, der regelmäßig ausgeführt wird, um neue Daten zu verarbeiten oder eine Prädiktivmodell?
 - ☐ A Daten App, die auf KNIME Hub eingesetzt wird
 - ☐ Ein prädiktives Modell, das über REST API zugänglich ist
- ☐ Führen Sie die Struktur Ihres Workflows aus. Betrachten Sie die folgenden Fragen:
 - ☐ Was ist die Aufgabe jedes einzelnen Workflows, den Sie aufbauen möchten? Wenn Sie sich identifizieren mehrere Aufgaben, begrenzen einen Workflow auf eine eigenständige Aufgabe.
 - ☐ Was ist die Eingabe und Ausgabe jedes einzelnen Workflows?
 - ☐ Teilen Ihre Workflows Teile?
- ☐ Definieren Sie die Anforderungen für Ihre Workflows. Betrachten Sie die folgenden Fragen:
 - ☐ Wie schnell soll der Workflow ausgeführt werden?
 - ☐ Wie oft sollte es ausgeführt werden, etc.?
- ☐ Definieren Sie die Rahmenbedingungen. Welche Hardware und Software müssen Sie treffen die Anforderungen Ihres Workflows?

Ziele der besten Praktiken bei der Arbeit mit KNIME

Nun, da Sie Ihre Projektvoraussetzungen definiert haben, möchten Sie die beste Methode finden, um mit Ihrem Workflow.

Sie werden lernen, was die besten Praktiken sind, warum sie wichtig sind und wie Sie sie in Ihrem Projekte mit KNIME Analytics Platform und KNIME Hub. Diese Anleitung hilft Ihnen, eine Intuition darüber, was zu beachten ist, wenn Sie einen Workflow erstellen, ausführen oder teilen.

Wenn Sie jedoch mit der KNIME Analytics Platform Ihre ersten Schritte unternehmen, beginnen Sie bitte mit folgendes: [Erste Anleitung erhalten](#) .

Machen Sie Ihren Workflow wiederverwendbar

Namenskonventionen

Workflows, die in Ihrer lokalen KNIME Analytics Platform-Installation entstehen, sind genauso wie jede andere.

Datei auf Ihrem Computer. Sie können sie mit ihrem Dateinamen eindeutig identifizieren und in

Ordner. Bevor Sie in die Daten eintauchen, beachten Sie folgendes beim Speichern Ihrer Workflows:

- Verwenden Sie eine klare Namenskonvention für Ihre Workflows. Der Dateiname sollte die

Aufgabe des Workflows. Zum Beispiel:

"Lesen und Vorbereiten Kundendaten" statt

"Project 1". Sie können auch Zahlen als Präfixe einführen. Eine konsequente Benennung wird angezeigt

in [Abbildung 3](#page9)

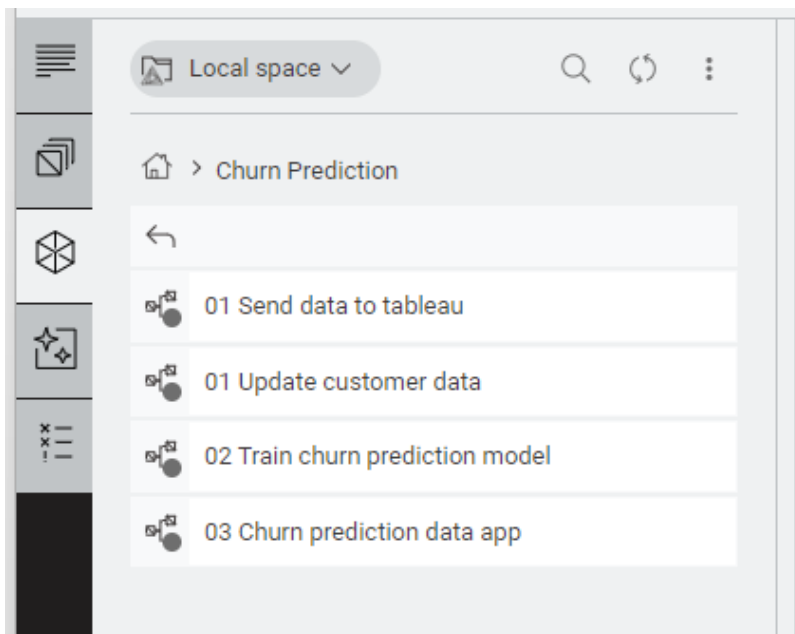


Abbildung 3. Konsistent benannte Workflows

- Denken Sie daran, einen Workflow auf eine kohärente Aufgabe zu beschränken. Wenn Ihr Anwendungsfall erforderlich ist mehrere Aufgaben, um das Endergebnis zu erreichen, gruppieren die jeweiligen Workflows in Ordner.

Wählen Sie einen Ordnernamen, der Ihr Ziel anzeigt.

Verwenden von relativen Dateipfaden

KNIME Workflows werden auf Datendateien betrieben. Wenn eine solche Datei auf Ihrem lokalen gespeichert wird Computer, sein absoluter Pfad identifiziert eindeutig den Standort Ihrer Datendatei in Ihrer Datei

System beispielsweise C:\Benutzername\Desktop unter Windows oder /home/user auf Linux.

Wenn Sie jedoch Ihren Workflow mit anderen teilen, z. [mit dem KNIME Hub als Repository](#) sie werden sie in ihrer lokalen KNIME Analytics Platform Installation ausführen.

Der Link, mit dem die KNIME Analytics Platform auf die von Ihnen zur Verfügung gestellte Datendatei zugreift mit Ihrem Workflow sollte ein relativer Pfad sein. Dies kann bedeuten, dass Ihr Weg relativ zu einem ist von:

- Ihr aktueller Workflow: Die Daten werden im gleichen Ordner wie die Workflow-Datei gespeichert, z. `/workflow_group/my_workflow`.
- Ihr aktueller Workflow-Datenbereich: Die Daten werden von der derzeit den Workflow ausführt. So werden die Daten in einem Ordner gespeichert, der sich im Inneren befindet. den Workflow-Ordner. Beispielsweise könnte ein relativer Pfad wie folgt aussehen:
Workspace > /workflow_group/my_workflow/data/. Die Daten Ordner enthält die Dateien, die die Daten enthalten. z.B. a.csv Datei. Zum Beispiel
Workspace > /workflow_group/my_workflow/data/adult.csv.
- Ihr aktueller Hubraum (angegeben, dass Sie in der jeweiligen Hub-Instanz angemeldet sind) aus Ihrem Raum Explorer).
- A [Dateisystem](#) dass Sie auf den jeweiligen Anschlussknoten zugreifen können

Alternativ können Sie eine benutzerdefinierte URL zusammen mit Ihren Daten für alle, die Sie teilen Ihr Workflow mit. Zum Beispiel

```
\https://example.com/mydata/datafile.csv
```



Mit dieser Option können Sie einzelne Dateien lesen und schreiben, aber Sie können sich nicht bewegen oder Dateien oder Ordner kopieren. Jedoch, Auflistung von Dateien in einem Ordner, d.h. Browser, ist nicht unterstützt.

Erfahren Sie mehr über die KNIME-Datei-Handling-Infrastruktur in der [KNIME Leitfaden für die Bearbeitung von Dateien](#)

Erfahren Sie mehr über das Speichern und Teilen von Artikeln auf dem KNIME Hub im [KNIME Gemeinschaft](#) [Anleitung zum Thema](#).

Dokumentieren Sie Ihren Workflow mit Knoten-Etiketten und Anmerkungen

Um Ihren Workflow wiederverwendbar zu machen, müssen Sie und Ihr Team schnell begreifen können

was jeder Teil und jeder Knoten tun. KNIME-Knotennamen sind selbsterklärend. Allerdings

um zu beschreiben, wie jeder Knoten auf Ihren spezifischen Daten arbeitet, können Sie das Knotenlabel ändern.

Wählen Sie einen Knoten im Workflow-Editor und doppelklicken Sie unter den Knoten, um sein Label zu ändern.

Hier können Sie beschreiben, welche Operation sie auf Ihren Daten durchführt.

Um eine von mehreren Knoten ausgeführte Operation zu erfassen, verwenden Sie eine Workflow-Annotation. Rechtsklick

überall im Workflow-Editor und wählen

Neue Workflow-Annotation

. Sie können eine

Workflow mit beiden Workflow-Annotationen, Knoten mit Knoten-Etiketten und einem Knoten, der

[tree diagram](#page11)

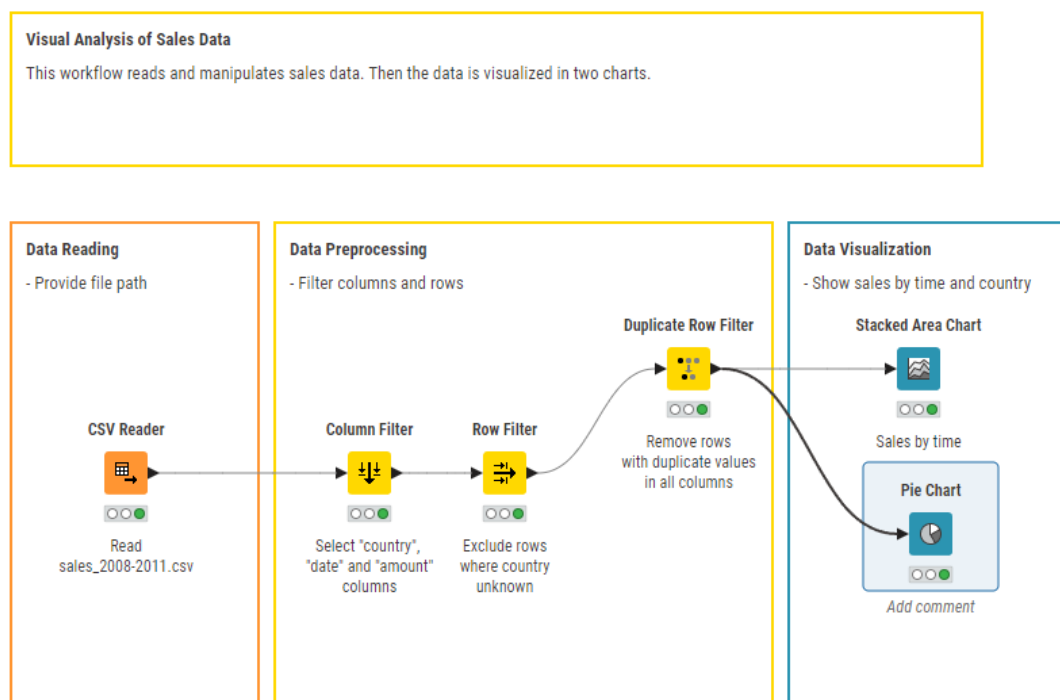


Abbildung 4. Node-Labels beschreiben Datenoperationen, die von einem Knoten durchgeführt werden. Workflow-Annotationen für mehrere Knoten dasselbe tun.

Strukturieren Sie Ihren Workflow mit Komponenten und Metanoden

Workflow-Annotations subsumieren kleinere Untergruppen von Knoten innerhalb eines Workflows. Je mehr Knoten, die Sie zu Ihrem Workflow hinzufügen, desto weniger verständlich wird es. In größeren Workflows, Sie kann Knoten in Gruppen komprimieren wollen, die durch einen einzigen Knoten dargestellt sind. Diese können entweder Komponenten oder Metanoden.

Metanodes werden ausschließlich verwendet, um Ihre Workflows klarer zu gestalten, indem viele Knoten in einen. Sie können Teile eines größeren Workflows einnehmen und in eine Metanode einbrechen, den Teil der Funktionalität des Workflows zu verstecken.

Neben den Funktionen einer Metanode kann eine Komponente auf verschiedenen Daten wiederverwendet werden, die die gleichen Operationen erfordern. Zum Beispiel, wenn Sie oft mit einem verbinden eine bestimmte Datenbank, Sie können eine Stecker-Komponente erstellen, die alle Einstellungen enthält Optionen.

Um eine Komponente oder Metanode zu erstellen, wählen Sie mindestens zwei Knoten aus, klicken Sie mit der rechten Maustaste irgendwo in der Workflow Editor, und wählen Sie entweder [Metanode erstellen](#) oder [Komponenten erstellen](#), wie in [und Abschnitt](#) der KNIME Komponentenführung.

Um eine Komponente verständlich für andere zu machen, dokumentieren Sie sie, indem Sie eine Beschreibung für die erwartete Eingabe und die verschiedenen Einstelloptionen (ähnlich einem KNIME-Knoten), wie in [dieser Abschnitt](#) der KNIME Komponentenführung.

Um die Flexibilität eines Bauteils richtig auszuschöpfen, sollte es:

- Verwenden Sie einen kleinen Satz von konsistenten Eingängen und Ausgängen.
- verlassen sich niemals auf Quellen, die für einen Datensatz spezifisch sind.
- Nicht erforderlich, dass seine Benutzer bei der Annahme von Knoteneinstellungen auf der Innenseite ändern.

Sie können Ihre Komponente auf verschiedene Daten anpassen, indem Sie [benutzerdefinierte Konfiguration Dialoge](#) [Zugang zu](#) [Durchflussgrößen](#) innerhalb des Bauteils. Dies ermöglicht Ihnen und andere Benutzer Ihrer Komponente, um eine Komponente einfach anpassen, um variable Namen von verschiedenen Datensätze als Eingaben. Insofern arbeiten Komponenten wie regelmäßige KNIME-Knoten: Du geben Sie ihnen Daten, an die Sie die Knoteneinstellungen anpassen. Letztendlich, Komponenten machen häufige Datenoperationen schneller und einfacher.



Einige Komponenten werden wiederholt ausgeführt, z.B. weil es sich um Datenapps handelt. Um ihre Ausführung zu beschleunigen, führen Sie schwere Berechnungen, wie die Verarbeitung große Datenmengen außerhalb der Komponente.

Als Knotenuntergruppen sollten Komponenten und Metanoden wie kleine Workflows behandelt werden in Bezug auf ihren Anwendungsbereich. Sie sollten eine einzelne, in sich geschlossene und kohärente

Funktion zu einer Komponente oder Metanode.

Erfahren Sie mehr über Komponenten und Metanoden in der [KNIME Komponentenführung](#).

Komponenten auf dem KNIME Hub teilen

Nachdem Sie eine Komponente in einem Workflow erstellen, wenn Sie sie in einem anderen Workflow wiederverwenden möchten, Sie kann es kopieren und in den neuen Workflow einfügen. Auf diese Weise ändert sich jedoch die Eine Komponente in einem Arbeitsablauf wird von den anderen nicht aufgenommen. Um die volle Nutzung eines Das Wiederverwendbarkeitspotenzial der Komponente können Sie auf dem KNIME Hub teilen. Zum Teilen einer Komponente, Rechtsklicken, wählen Komponente → Anteil . Dadurch können Sie Ihre Mitarbeiter aufrufen Ihre Änderungen an einer Komponente und auch ihre.

Zusätzlich, von [Verknüpfung einer gemeinsamen Komponente](#) Sie können Änderungen von anderen in Ihre lokale Installation. Wann immer Ihre Mitarbeiter die Komponente auf dem KNIME aktualisieren Hub, Ihre KNIME Analytics Platform Installation fordert Sie auf, die Komponente zu aktualisieren. Zu Anpassung der Update-Richtlinie, navigieren, Vorlieben → KNIME . Unter Komponentenupdates , du kann die Standard-Update-Richtlinie festlegen Nie , Immer , oder Prompt .

Erfahren Sie mehr über [Teilen von Komponenten im KNIME Components Guide](#).

Bitte beachten Sie, dass sie, wenn Eingabedaten enthalten sind, für jeder, der Zugang zu der Komponente selbst hat. gründliche Prüfung [die Art des Hubraums](#) Sie speichern Ihre Komponenten und ob Ihre Daten enthält sensible Informationen.

Erfahren Sie, wie Sie Workflows behandeln, die vertrauliche Daten beinhalten, in der [Führung](#)

Einige Komponenten ermöglichen Streaming Execution. Lesen Sie mehr darüber, wie diese Funktion macht [Workflow-Streaming](#)

Komponenten für Datenapps

Sie können Komponenten und ihre zusammengesetzte Ansichten Seiten in der Web-Anwendung definieren Workflows. Nach dem Upload auf KNIME Hub können sie als Data Apps eingesetzt werden.

Um direkt mit der Data App UI im Browser zu interagieren, müssen Sie Komponenten implementieren mit Widgets. Widgets ermöglichen es den Verbrauchern Ihrer Data App, zwischen ihren Seiten zu navigieren. Jede Seite entspricht einer Komponente auf der Wurzelebene des Workflows, enthaltend Widget Knoten und Knoten anzeigen. Dies bedeutet, wenn Sie die zusammengesetzte Ansicht in KNIME inspizieren Analytics Platform, Sie können die Parameter im Fenster einstellen, das öffnet.



Eine Komponente, die eine Daten-App darstellt, soll wiederholt ausgeführt werden.

Um seine Ausführung zu beschleunigen und unerwünschte Folgen zu vermeiden, führen Sie schwere Berechnungen außerhalb des Bauteils.

Erfahren Sie mehr über Daten-Apps in der

[jeweiliger Abschnitt des KNIME Komponentenführers](#)

, oder

lernen, wie eine Daten-App Schritt für Schritt in der

[KNIME Daten Apps Anfänger Anleitung](#)

.

Workflows erfassen und aufrufen

Workflow Segmente erfassen

Viele Teilprozesse in einem Arbeitsablauf zu haben, macht es langsam und schwierig, zu warten und zu testen.

[Strukturieren Sie Ihren Workflow](#page12) die Vorteile von

Es werden wiederverwendbare Workflow-Segmente beschrieben. Eine weitere Option zur automatischen Erfassung einer

Auswahl von Knoten aus einem Workflow ist die [KNIME Integrierte Bereitstellungserweiterung](#).

Diese Funktion ermöglicht es Ihnen:

- Erfassen Sie Kernsegmente Ihres Workflows zur Wiederverwendung
- Speichern Sie erfasste Segmente automatisch als Workflows mit allen relevanten Einstellungen und Transformationen enthalten Infolgedessen benötigen erfasste Segmente keine manuelle Arbeiten oder Änderungen, die wiederverwendet werden sollen. Dies erleichtert die Bereitstellung und Prüfung relevanter Teile eines Workflow, z.B. benutzerdefinierte Datenvorverarbeitung.

Sie können die KNIME Integrated Deployment Extension für verschiedene Anwendungsfälle verwenden, z.B.:

- Bereitstellung von Datenfernleitungen
- Modelleinsatz
- Ad-hoc-Tests
- Zusammenfassung der Arbeitsabläufe

Im Allgemeinen profitieren Sie von:

- Sparzeit
- Fehler reduzieren
- Erhöhung der Compliance
- Optimierung von Prozessen

Wenn Sie mehr Möglichkeiten zur Optimierung der Workflow-Performance erfahren möchten, lesen Sie bitte die

[Effizienz](#page24)

Um ein Workflow-Segment zu erfassen, d.h. eine Auswahl von Knoten, stellen Sie eine [Workflow Start erfassen](#)

Knoten vor dem ersten Knoten und [Workflow-End erfassen](#) nachdem der letzte Knoten erfasst werden soll, wie dargestellt [Abbildung 5](#page16)

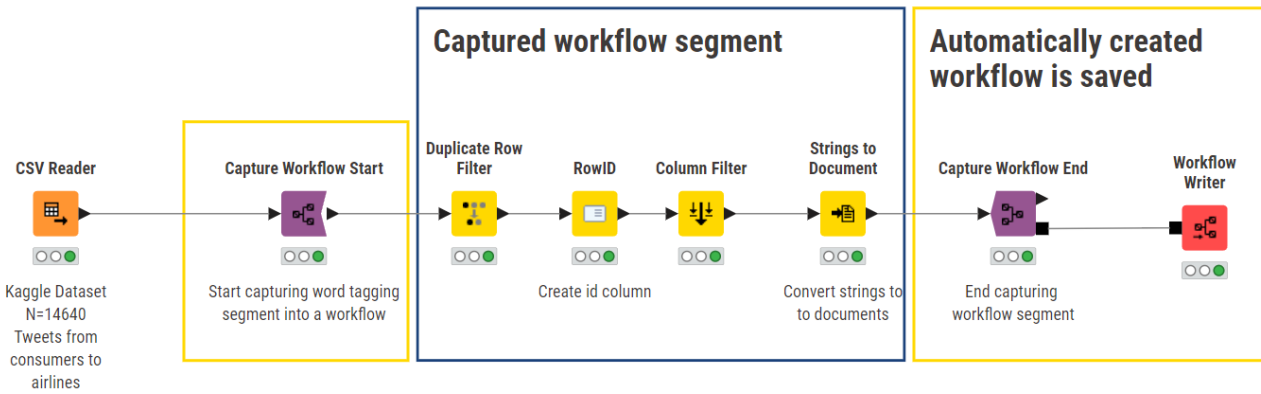


Abbildung 5. Ein von Capture Workflow Start und Capture Workflow gekapseltes Workflow-Segment Endknoten.

Das gesamte Workflow-Segment im Rahmen dieser beiden Knoten steht dann am Workflow-Ausgangsanschluss des Workflow-End erfassen Knoten. Im Konfigurationsdialog der Workflow-End erfassen node, Sie können Eingabedaten speichern, die standardmäßig verwendet werden, wenn Führen des erfassten Workflow-Segments, es sei denn, Sie geben eine andere Eingabe. Sie können auch propagate-Variablen, die innerhalb des erfassten Segments definiert sind.

[

Knoten, die ausgehende Verbindungen zu einem Knoten haben, der Teil des Geltungsbereichs ist, sind nicht Teil des Umfangs selbst als statische Eingänge dargestellt, sondern nicht gefangen.

Um einen erfassten Workflow zu speichern, haben Sie folgende Optionen:

1. ANHANG Schreiben Sie den Workflow in einen lokalen Ordner mit dem Workflow Writer Knoten, wie in [die jeweilige Sektion im KNIME Integrated Deployment Guide](#).
2. Schreiben Sie den Workflow an ein angeschlossenes Dateisystem, indem Sie einen Datenbanksteckerknoten hinzufügen. Siehe [jeweilige Sektion im KNIME Integrated Deployment Guide](#) für Details.
3. Lesen Sie den Workflow in einem anderen Workflow mit dem [Workflow Reader](#) Knotenpunkt. Es liest ein Ein einziger Workflow in ein Workflow-Port-Objekt. Da der Workflow nicht ausgeführt wird dieser Knoten, fügen Sie ein Workflow-Ausschnitt das zu tun. Siehe [jeweiliger Abschnitt der KNIME Integrierter Bereitstellungsleitfaden](#) für Details.

Die Workflow Reader node ermöglicht es Ihnen, ein Workflow-Objekt zu lesen, auch wenn Sie nicht wollen Bringen Sie es sofort aus. Dies ist nützlich, um den Workflow, z.B. mit dem [Zusammenfassung des Ablaufs Extrakt](#) Knotenpunkt oder, wenn Sie dieses Workflow-Objekt einem nachfolgenden Knoten zur Verfügung stellen möchten oder als Eingabe. Die Behandlung des Workflow-Objekts wird an den Knoten oder eine Komponente. Die Workflow Reader nicht auf den Datenbereich des Workflow zugreifen können, wird gelesen, oder liest es teilweise ausgeführte Workflows. Der gesamte Workflow wird zurückgesetzt. Die [Workflow-Ausschnitt](#) Knotenpunkt kann nützlich sein, um ein Workflow-Objekt auszuführen, das die Ausgabe eines Knoten oder Komponente. Zum Beispiel könnten Sie den Ausgangsport des [Arbeitsablauf](#)

Kombinationer^{Knotenpunkt}, die die Konfiguration von Ein-/Ausgabe-Port-Verbindungen zwischen Arbeitsflussegmente zu den Workflow-Ausschnitt. Die Workflow-Ausschnitt nur einen Workflow ausführen, der vom gleichen Workflow erfasst wird, den der Workflow-Executor Knoten ist Teil von.



Man kann KNIME nicht lesen

Komponenten

mit dem Workflow Reader Node.

Schließlich, wenn Sie erfasste Workflows sofort nach dem Lesen in einen anderen verwenden möchten

Workflow, die Nutzung des Call Workflow Services Nodes statt. Siehe [Dienstleistungen](#) für Details.

Erfahren Sie mehr über die Erfassung von Segmenten eines Workflows in der

[KNIME Integrierte Bereitstellung](#)

[Leitfaden](#).

Workflow Services: Invoke Workflows ineinander

Workflow Services bieten Knoten, die es einem Caller-Workflow ermöglichen, andere Workflows einzufordern.

Die genannten Workflows sind Callee Workflows.

Da ein idealer Workflow nur eine Aufgabe erfüllt, können Sie Ihre Modularisierung

Workflows. Letztlich macht es Workflows schneller auszuführen und einfacher zu verstehen und

halten. Sie können Workflow Services für:

- Datentabellen, Text, Modelle und andere Objekte zwischen Workflows teilen
- Erstellen Sie modulare Workflows, von denen jeder ein Diagramm für einen Bericht erstellt, die von einem Caller Workflow inszeniert, der einen Bericht erstellt, der alle umfasst

Anstatt standardisierte (aber limitierende) JSON-basierte APIs einzufordern, ist es möglich, zu erstellen und zu erstellen.

verbinden Sie direkt mit KNIME native API Endpunkte. Dies bedeutet, dass man mehr teilen kann

KNIME-Datentypen, über das, was mit JSON möglich ist, z.B. Textdokumente, Entscheidungsbäume,

tiefe Lernmodelle und vieles mehr. Dies ist besonders nützlich bei der Workflow-Orchestrierung.



Zum Aufruf von Workflows von externen REST-Clients nutzen Sie bitte

Containereingang/Ausgang

Knoten. Siehe

[jeweilige Sektion im KNIME Workflow Invocation Guide](#)

für

Details.

Die wichtigsten Knoten in Bezug auf Workflow Services sind die folgenden aus

[KNIME](#)

[Persönliche Produktivität Erweiterung der Werkzeuge](#) :

- Workflow Service Input Node: Es erhält eine Tabelle oder jedes Port-Objekt von einem Workflow, nennt diesen Workflow, also einen Caller Workflow. Verschiedene Porttypen sind verfügbar.

- **Workflow Service Output** Node: Es sendet ein Objekt an einen Anrufer-Workflow. Verschiedene Ports Typen sind verfügbar.
- **Call Workflow Service** Node: Es nennt andere lokale und Remote-Workflows. Mehrere Ports und verschiedene Ports sind verfügbar.

Schlussfolgerung **Workflow Input** und **Ausgangsleistung** Knoten werden verwendet, um Workflows zu erstellen, die für aus einem anderen Workflow (d.h. mit dem **Call Workflow Service**-Knoten) aufgerufen werden.

Die **Call Workflow Service** Knoten wiederum erhält die Ergebnisse für die weitere Verarbeitung in der Arbeitsablauf. Es kann jederzeit verwendet werden, wenn wir wollen, dass ein Anrufer Workflow einen Callee Workflow anruft. Die Workflow kann Eingänge über **Workflow Service Input** Knoten und Rückgabeausgänge welche dann die **Workflow Service Output**-Knoten sammeln. Die Ports des **Call Workflow Service** Knoten werden automatisch nach dem im Konfigurationsdialog.

Die Vorteile von Workflow Services sind:

- Sie können ausreichend auf den Workflow-Datenbereich des Callee-Workflows zugreifen, der unterscheidet sie von der **Workflow Reader** Knoten.
- Sie nutzen die Infrastruktur des KNIME Business Hub, was zu einer besseren Skalierbarkeit führt.
- Sie sind effizienter als die Verwendung **Workflow Reader** und **Workflow-Ausschnitt** Knoten, wenn die direkte Workflow-Ausführung das Hauptziel ist. Wenn Sie mehr Möglichkeiten lernen möchten, [<a href="#page24" style="color: #ff6600; text-](#page24) Optimierung der Workflow-Leistung, bitte beachten Sie die

Erfahren Sie mehr über Workflow Services im [KNIME Workflow Invocation Guide](#).

Prüfung und Datenvalidierung

Beginnen Sie mit warum und wie zu testen

Um sicherzustellen, dass ein KNIME-Workflow, eine Komponente oder ein Workflow-Segment als erwartet, sollten Sie Ihre Workflows regelmäßig testen. Letztlich macht das Ihre Workflows weniger fehleranfällig und einfacher zu pflegen. Dieser Abschnitt erklärt, wie KNIME als Testrahmen.

Ihre Workflows werden nicht immer mit der idealen Eingabe konfrontiert, noch werden sie in der Richtig. Dadurch sollten Ihre Workflows robust sein, um von der Norm abweichende Eingaben zu erreichen. kontrollierbar brechen. Dies trägt dazu bei, dass alle künftigen Änderungen weiterhin alle passieren der gleichen Tests.

Vor der Prüfung sollten Sie Ihre Ziele klar definieren. Bei Workflow-Tests können Sie formulieren Sie Testfälle, die jede gegebene Situation, die durch Ihren Workflow auftritt, d.h. was passiert, wenn gültige Daten in den Workflow eingespritzt werden und welche Fehler angehoben werden, wenn ungültige Daten werden bereitgestellt.

Mögliche Probleme, die bei der Prüfung auftreten können, sind:

- Failing aus unerwarteten Gründen
- Unerwartete Fehlermeldungen
- Falscher Ausgang

Die Gründe könnten sein:

- Invalide Eingaben, z.B. von Endbenutzern
- Konfigurationsänderungen
- Änderungen durch ein Update
- Ändern der externen Dienste

Test und Validierung in der KNIME Analytics Platform

In der KNIME Analytics Platform erstellen und automatisieren Sie einen Testrahmen für die Zukunft Bereitstellungen. Sie können sogenannte Testflows erstellen, die KNIME Workflows sind, die gesteuerte Testeingänge und -ausgänge zu einem Workflow, Bauteil oder Workflow-Segment. Prüfung in und Ausgänge sind Gold-Standard-Daten, d.h. sie stellen kleine Proben der Eingänge dar und Ergebnisse, die Sie erwarten, dass Ihr Workflow zu handhaben.

Wenn Testausgänge vorgesehen sind, überprüfen die Testabläufe dann, ob die Ausgänge mit der

erwartete Ergebnisse aus den Testeingängen. Wenn dies nicht der Fall ist, sollten die Testläufe eine Alarm, die eine Antwort auf die Abweichung von Eingaben oder unerwarteten Fällen ist. Sie haben bereits definiert die Fälle als Fehlertestfälle, z.B. fehlende oder ungültige Eingaben, unerwartete Fehler usw.

Als beste Praxis sollten Sie nie:

- Verarbeiten Sie die goldenen Daten außerhalb Ihres Testflusses, als wären es regelmäßige Daten. Dies kann führen unbekannte Fehler, die Sie identifizieren müssen.
- Ändern Sie die goldenen Daten.

Es ist am besten, goldene Daten in einem separaten Ordner zu halten und Berechtigungen einzuschränken, wenn möglich. KNIME Analytics Platform bietet Knoten, mit denen Sie die Ausgabe Ihrer Testläufe testen können. Obwohl es viele mehr gibt, sind die am häufigsten verwendeten Knoten unten aufgeführt.

- [Tabelle Gültig](#)

Ermöglicht eine bestimmte Tabellenstruktur und Tabelleninhalt (Schema). Dadurch wird die Ausgangstabelle Struktur sollte identisch mit der Spezifikation sein, die der Benutzer in der Konfiguration definiert Dialog. Dieser Knoten ist Teil der KNIME Base-Knoten.

- [JSON Systemsteuerung](#)

Validiert JSON-Werte basierend auf einem bestimmten Schema. Es sammelt falsche Werte, so dass Sie können sie inspizieren.

- [Tabelle Unterschiedskontrolle](#)

Bestimmt, ob es einen Unterschied zwischen zwei Tabellen gibt. Es vergleicht die von Ihrem erstellte Tabelle Testflow mit einer Referenztable, z.B. einem goldenen Tisch.

- [Tabelle Unterschied Finder](#)

Erkennt und gibt Domänen- und Wertunterschiede für jede Zeile und Spalte in den Tabellen zurück. Im Gegensatz zum Tabelle Difference Checker scheitert der Knoten nicht sofort, wenn die Tabellen unterschiedlich, aber der Konfigurationsdialog bietet eine Option, um auch ein Ausfall zu ermöglichen.

- [Datei Unterschied Checker](#)

Der Knoten wird zwei Dateien, die durch die Flussvariable gegeben werden, vergleichen und scheitert mit einer Ausnahme, wenn die Dateien sind nicht gleich. Viele ähnliche Knoten existieren auch für andere Arten von Dateien, z.B. [BildDifferenz](#)

[Prüfer](#) . Sie finden sie auf dem KNIME Community Hub, indem Sie nach den Keywords suchen ["Differenz-Checker"](#) .

Im Folgenden werden Sie über zwei Ansätze zur Erstellung von Testläufen in KNIME Analytics lesen Plattform. Die oben aufgeführten Knoten können in beiden verwendet werden.

In der KNIME Analytics Platform können Sie entweder [KNIME Integrierte Bereitstellungserweiterung](#) oder [KNIME Testing Framework UI Extension](#). Beide Ansätze sind im die folgenden: Sie können wählen, welche man am besten Ihren Anwendungsfall passt.

KNIME Integrierte Bereitstellung

Die erste Möglichkeit besteht darin, Testflows zu erstellen, indem Workflow-Segmente erfasst und ausgeführt werden

Verwendung von [KNIME Integrierte Bereitstellungserweiterung](#), die in der [invoke Workflows](#)

Diese Option ist nützlich für Ad-hoc-Tests, d.h. Ausführen der erfassten Workflow-Segmente in den gleichen Workflow beim Testen gegen neue Daten. Zur Ausführung können Sie den Workflow nutzen

Executor. Fügen Sie zuerst das Segment ein, das Sie testen möchten [Workflow Start erfassen](#) und [Workflow-End erfassen](#) [Workflow Start erfassen](#) und [Workflow-End erfassen](#) von

diese Anleitung. Dann erstellen im gleichen Workflow mehrere Workflow-Zweigstellen. Jeder von ihnen

stellt einen anderen Testfall dar, indem verschiedene Dateneingänge in den erfassten Workflow injiziert werden [Workflow Start erfassen](#) und [Workflow-End erfassen](#)

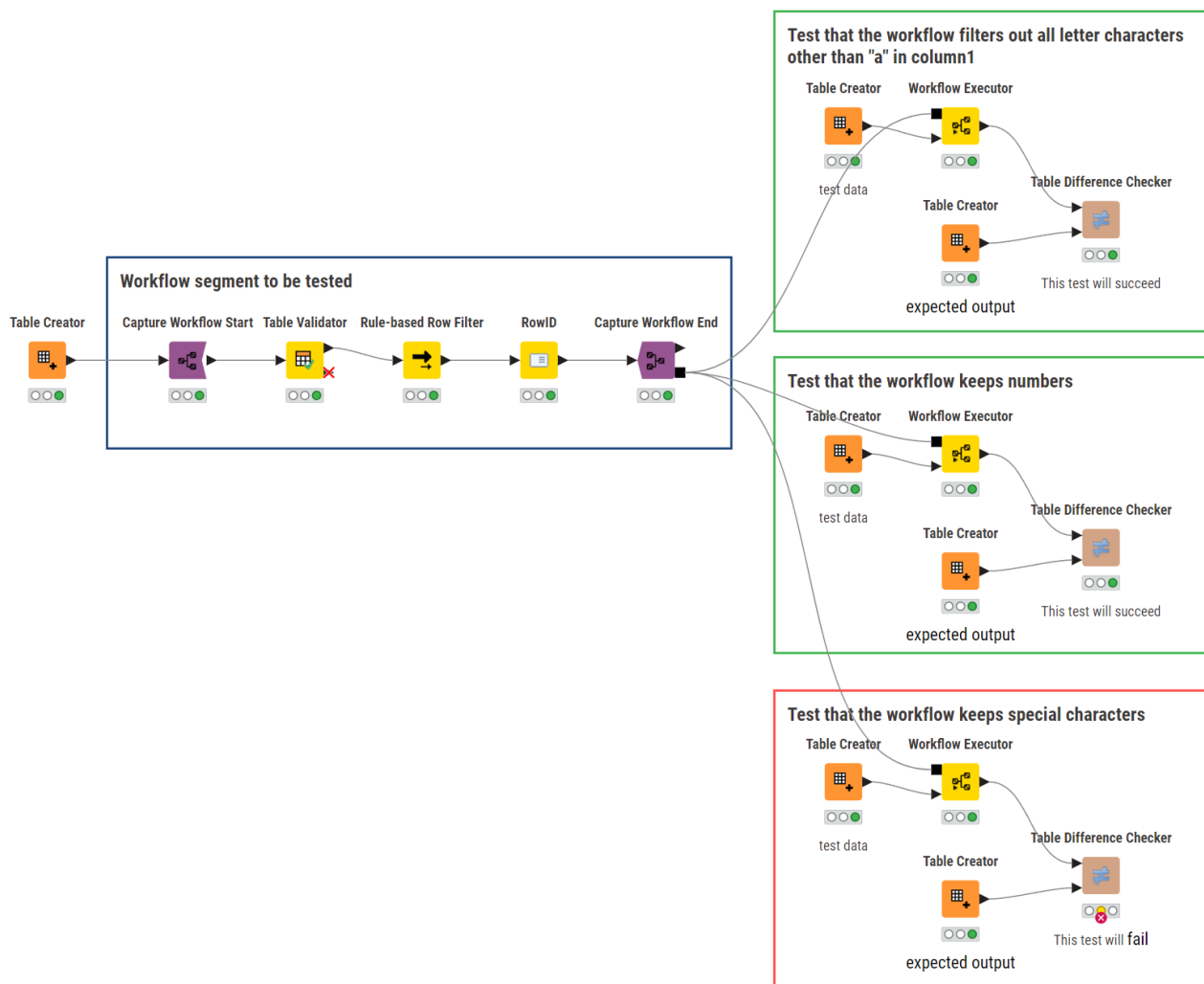


Abbildung 6. Ein Workflow mit Workflow-Segmenten, die verschiedene Testfälle darstellen.



Wenn Sie nur testen möchten, ob Ihr Workflow ausgeführt wird oder nicht, können Sie Einsatz Workflow-Ausschnitt Knoten ohne zusätzliche Eingabe, aber die erfasst Arbeitsflussegment.

Um einen Anwendungstest durchzuführen, d.h. um einen gesamten Workflow zu testen, empfehlen wir, die [Workflow-Service-Knoten](#) statt.

KNIME Prüfrahmen UI

Die zweite Option zum Testen Ihrer Workflows ist die

[KNIME Testing Framework UI Extension](#).

Es ist eine komplexere Methode der Erstellung und Durchführung von Testabläufen. Sie können Berichte erstellen die Sie innerhalb der KNIME Analytics Platform sehen können.

Ihr Testlauf kann so einfach oder komplex sein, wie Sie wählen. Der Test selbst wird von einem

Validierungsknoten wie z.B. [Tabelle Gültig](#), die die von Ihnen erzeugte Leistung vergleicht

Workflow gegen einen Goldstandard. Um festzustellen, welche Bedingungen für die

Workflow, um zu scheitern oder erfolgreich auszuführen, können Sie die [Testflow Konfiguration](#) Knoten.

Testflow Automation

Um mehrere Testabläufe automatisch auszuführen, können Sie eine [Call Workflow Service](#) Knoten in eine Schleife. Sie iterieren einfach über die Testflow-Namen, präzisieren die Ergebnisse der berichteten

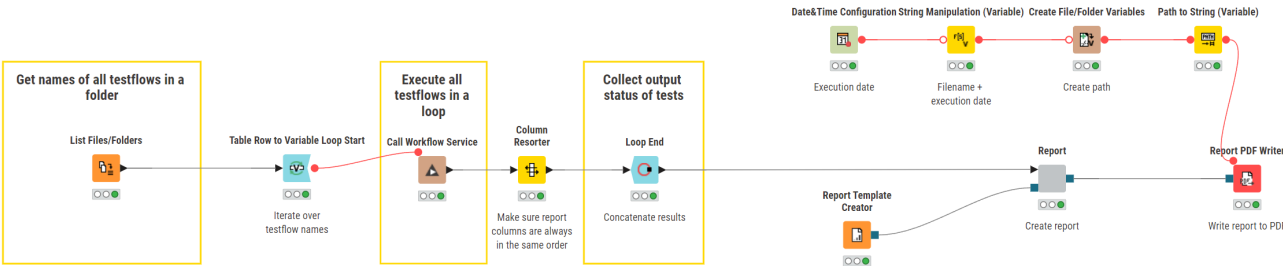


Abbildung 7. Ein Workflow, der Testflow-Dateien in einer Schleife ausführt und ihre Ausführungszeit an eine Bericht.

Erfahren Sie mehr über die Erstellung von Berichten in der [KNIME Reporting Guide](#).

Dies ist erforderlich, da die Erkennung von Fehlern mit Testläufen nicht automatisch bedeutet, dass Die Verantwortlichen werden informiert. Der Testablauf könnte entweder stillschweigend scheitern, oder der Prozess würde sogar erfolgreich abgeschlossen werden, während fehlerhafte Eingabedaten verarbeitet wurden.

Um diesen Fehlern bewusst zu werden, können Sie ausfallende Knoten in einem Try-Catch-Framework einbetten, um alle Fehler einholen. Es erfasst Informationen über ausfallende Knoten, die dann in [den Bericht](#). Auf der [KNIME Community Hub](#) Sie können verschiedene Knoten finden, um dies zu implementieren den Rahmen.

Probieren Sie Knoten einfach den Beginn des Try-Catch-Frameworks an. Stellen Sie sie vor der erste Knoten, den Sie einfügen möchten und mit seinem Eingangsport verbinden. Die [Fangfehler](#) Knoten, auf der anderen Seite, wird nach dem letzten Knoten platziert, den Sie in Ihren Try-Catch einfügen möchten Struktur. Durch Anschließen des letzten Knotens an den Eingangsport des [Fangfehler](#) Knoten, es wird die Eingabe vom ersten Port weiterleiten, wenn die Ausführung erfolgreich war. Wenn Ausführung oben Ast fehlgeschlagen (und ein passender Testknoten wurde vor dem ausfallenden Knoten angeschlossen) dann die Eingabe vom zweiten Hafen von [Fangfehler](#) port wird weitergeleitet, und die zweite Variable output enthält Informationen über den beobachteten Fehler.

Nachdem Sie die Fehler gesammelt haben, können Sie aus einer Vielzahl von KNIME-Knoten zur Nachricht wählen die Verantwortlichen über die von Ihnen identifizierten Probleme.

Machen Sie Ihren Workflow effizient

Effiziente Dateilesung

Um Ihren Workflow zu beschleunigen, beachten Sie folgende Richtlinien:

- Vermeiden Sie das mehrfache Lesen derselben Datei, d.h. verwenden Sie nicht jeden Leseknoten, wenn Sie die gleiche Datei für verschiedene Datenoperationen benötigen. Sie können mehrere Verbindungen verwenden den gleichen Leseknoten mit mehr als einem nachfolgenden Knoten verbinden, wie in [8](#page24)

[8](#page24)

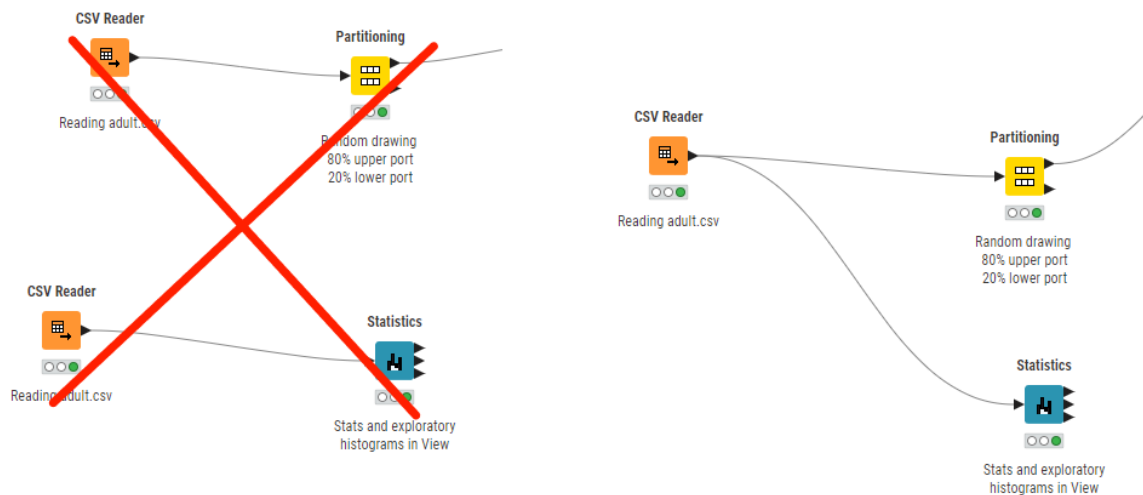


Abbildung 8. Ein Knoten, der eine Datei liest, kann mehrere Verbindungen an seinem Ausgangsport haben

- Lesen Sie mehrere Dateien mit der gleichen Struktur mit einem Leseknoten. Öffne die Konfiguration eines Leseknotens und wählen Sie die Option "Files in Ordner" in der Registerkarte Einstellungen.
- Entfernen Sie unnötige Daten. KNIME Quellknoten, wie die Reader-Knoten, bieten eine Option zur Entfernung ungenutzter oder redundanter Spalten. Öffnen Sie die Konfiguration eines Leseknotens und wählen Sie die Transformation Tab.

Effiziente Workflow-Design

Bei der Gestaltung eines Workflows gibt es nur wenige Fehler, um absolut zu vermeiden, aber viele praktische Praktiken, die ihre Nutzung erleichtern. Beide werden im folgenden vorgestellt.

Ein effizienter Workflow enthält nie:

- Trennknoten oder Komponenten. Vor allem, wenn Sie diesen Workflow auf einem KNIME Hub-Instanz, oder wenn Sie wollen
- Arbeitsabläufe, die nicht aktiv benötigt werden.
- Große, ungenutzte Dateien in Form von Komponenten, Knoten oder Metanoden. Sie verlangsamen sich Laden, Ausführen und Speichern des Workflows. In einigen Fällen verursachen sie sogar die Arbeitsablauf zu scheitern.

Um Ihren Workflow schneller laufen zu lassen, beachten Sie folgendes:

- Einige Datenmanipulationsoperationen sind rechnerisch teurer als andere.
KNIME Analytics Platform bietet Knoten an, um unnötig teure Operationen zu vermeiden.
Zum Beispiel:
 - ❑ **Mitglied**: Zwei Tabellen mit dem Joiner-Knoten verbinden, um dann für Kriterien zu filtern ist langsam. Stattdessen, ersetzen Sie ein paar Zeilen mit der [Wertesuche](#) [Knotenpunkt](#), das vergleicht und entspricht Ihren Daten zu einem Wörterbuch, das Ihre Kriterien enthält.
 - ❑ **Loops** : Abhängig von Ihren Daten und Iterationen können Schleifen Ihre Arbeitsablauf. So verwenden Sie Schleifen nur, wenn absolut erforderlich. Stattdessen, versuchen Sie Multi Column Knoten auf Strings in mehreren Spalten auf einmal zu bedienen.
- Einige Datentypen verwenden mehr Speicherplatz auf Ihrem Computer als andere. Zum Beispiel Strings belegen mehr Raum als ganze Zahlen. Überprüfen Sie Spalten mit dem falschen Datentyp, z.B., wo Zahlen als Zeichenketten gespeichert werden.
- Redundante Daten dienen in Ihrem Datenbetrieb nicht. Spalten filtern, die konstante Werte oder Informationen enthalten, die Sie später nicht verwenden.
- Der schnellste Weg, wiederkehrende Prozesse zu laufen, ist zu erfassen, zu speichern und zu replizieren Workflow auf (neuen) Daten. Um zu lernen, wie, verweisen Sie auf die [Abschnitt dieser Führung](#)

Messleistung mit KNIME-Knoten

Wie jedes gute Projekt ist ein guter Workflow skalierbar.

Um zu messen, welche Knoten am längsten in Ihrem Workflow ausgeführt werden, können Sie die **Timer Info** Knoten. Es meldet individuelle und aggregierte Timing- und Ausführungsinformationen für alle Knoten. Dadurch können Sie Schwachstellen in Ihrem Workflow identifizieren und Ihren Fortschritt im Laufe der Zeit verfolgen.

Die **Timer Info** [Knotenpunkt](#) ist in der KNIME-Basisinstallation enthalten. Einfach überall hineinziehen den Workflow, den Sie bewerten und ausführen möchten.

Prozess in der Speicheroption

In der KNIME Analytics Platform [Gruppe](#) und [Pivot](#) nodes bieten das "Process in Memory" Option. Während es einen signifikanten Speicher verwendet, bringt es auch eine bemerkenswerte Leistung Verbesserung durch Reduzierung von Scheibenwaps, wenn diese Knoten ausgeführt werden. Um die "Process in Memory" Option, öffnen Sie den Konfigurationsdialog der [Gruppe](#) oder [Pivot](#) Knoten. Unter den Tabs [Einstellungen](#) > [Gruppen](#) navigieren [Erweiterte Einstellungen](#) und wählen Sie die Checkbox [Prozess im Speicher](#) .

Ausführung optimieren

Einige KNIME-Knoten und Komponenten ermöglichen eine Streaming-Ausführung. Sie können sie mit die [KNIME Erweiterung der Execution \(Beta\) auf KNIME Hub](#).

Während die Standardausführung node-by-node betreibt, sind bei Streaming-Ausführungsknoten gleichzeitig ausgeführt. Jeder Knoten übermittelt Daten zum nächsten, sobald er verfügbar ist, d.h. bevor der Knoten vollständig ausgeführt wird. Insbesondere werden alle Daten in eine Anzahl von Chargen aufgeteilt die von einem durch einen gestreamt werden. Insbesondere beim Lesen oder Vorbereiten von Daten, Streaming Ausführung beschleunigt Laufzeit.



Beim Streamen der Ausführung sind Zwischenergebnisse nicht verfügbar, da keine Daten wird geätzt.

Um von Standardeinstellung zu Streaming-Ausführung zu wechseln, erstellen Sie eine Komponente und öffnen Sie ihre Konfigurationsdialog. Öffne die **Job Manager Auswahl** Tab in der Bauteilkonfiguration Dialog. Wenn die Streaming-Ausführungsoption nicht für den Knoten verfügbar ist, sehen Sie nur den **Standard** Option im Menü. In Komponenten, die eine Streaming-Ausführung ermöglichen, wählen Sie die "Simple Streaming" Option. Das stromfähige Bauteil ist nun mit einem Pfeil in seinem unten links Ecke.

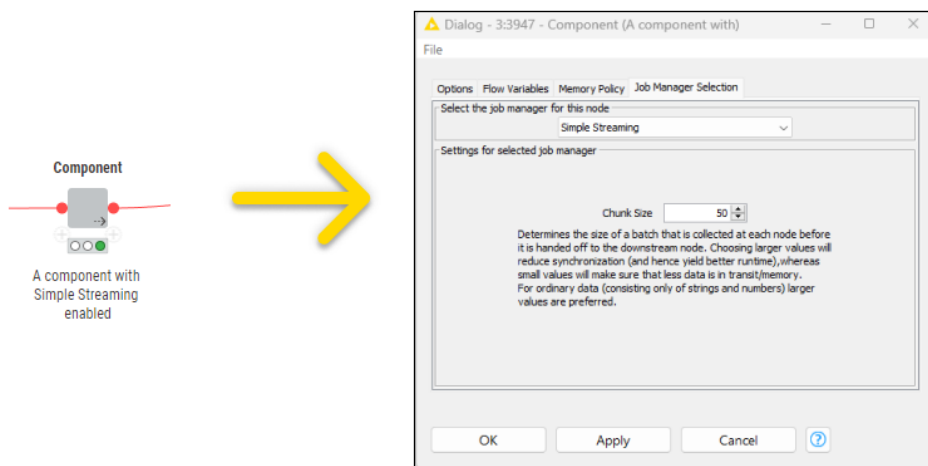


Abbildung 9. Eine Komponente mit Simple Streaming-Ausführung aktiviert



Streaming-Ausführung ist noch nicht optimiert, um mit dem Columnar zu arbeiten

Backend. Sie können Ihren Standard-Tabellen-Backend wie in der

[Columnar Tisch Backe](#)

Weitere Informationen zu Streaming Execution finden Sie in der

[KNIME Komponentenführung](#)

[jeweiliger Abschnitt in der](#)

Erfahren Sie mehr darüber, wie Sie Ihren Workflow mit Komponenten im

[Wiederverwendbarkeit dieser Füh](#)

Columnar Tisch hinten


Sie können die [KNIME Columnar Table Backend](#) Erweiterung des Workflows Ausführung. Es optimiert die Nutzung des Hauptspeichers in der KNIME Analytics Platform durch Überprüfung der zugrunde liegende Datendarstellung, wobei Zellelemente in einer Tabelle durch Java dargestellt werden Objekte. Seine nun unterschiedliche zugrunde liegende Datenschicht wird durch [Apokalypse](#) die auf eine Säulendarstellung.

Die Vorteile sind:

- **Einfachheit** : Einfache Zellelemente (z.B. ganze Zahlen) sind nicht als Speicher dargestellt - intensive Objekte.
- **Kompaktheit** : Die Daten werden kompakter gehalten. So passen mehr Daten ins Haupt Gedächtnis.
- **Kontrolle** : Kleinere Datenmengen werden in-memory ("cached") gehalten, wodurch die Speicherfußabdruck von KNIME viel stabiler. Auch die Daten leben "off-heap" (eine Region wo, für Sie als Benutzer, ist es steuerbarer.
- **Vereinbarkeit** : Es ist ein erster Schritt, den gemeinsamen Speicher mit anderen Programmierungen zu aktivieren Sprachen wie Python, was zu mehr Leistungsverbesserungen führt. und damit mehr Leistungsverbesserungen.

Sie können das Columnar Backend aktivieren, indem Sie navigieren [Vorlieben > Zurück zur Übersicht](#) .

- Setzen Sie Ihren Tabellen-Backend in Spalte [Vorlieben](#) nur beeinflussen Workflows erstellt nach der Anpassung. bestehende Workflows, einschließlich die aktuell in Ihrer KNIME Analytics Platform eröffneten Workflows müssen manuell aktualisiert werden.

Um das Tisch-Backend für einen bestimmten Workflow einzustellen, z.B. einen Workflow, der vor Ändern Sie Ihr Standard-Tabellen-Backend, navigieren Sie auf die Node Beschreibung Tab in der Seitentafel Navigation. Klicken Sie auf die  Icon. Im Fenster, das öffnet, im [Zurück zur Übersicht](#) Tab, wählen Die Welt der Welt im Dropdown-Menü. Die Parameter bezüglich der Speichernutzung der Columnar Backend kann unter [Vorlieben > Tabelle Backend > Columnar Backend](#) .

- Der Columnar Backend ist noch nicht optimiert, um mit Streaming zu arbeiten Ausführung. Sie können Ihren Standard-Tabellen-Backend wie in der [>Durchführungsabschn](#page28)

Weitere Einzelheiten finden Sie in der [Bevorzugter Abschnitt der KNIME Analytics Plattform Benutzerhandbuch](#)

Parallelisieren Sie die Abzweigung mit Schleifen

Eine andere Möglichkeit, die Ausführung des Workflows zu beschleunigen, ist Parallel Computing. Sie können dies implementieren mit Schleifen. Sie iterieren über einen bestimmten Teil des Workflows, z.B. über einen Workflow-Verzweig. Jedes Iteration, ein wiederholter Vorgang, hat unterschiedliche Eingabe. Unter anderem kann dieser Eingang aus chunks der Daten, die Sie in der KNIME Analytics Platform verarbeiten möchten. Diese Stücke werden jeden parallel an getrennten Arbeitsabläufen verarbeitet werden.

Um zu beginnen, müssen Sie die [KNIME Parallel Chunk Loop Nodes Erweiterung](#) installiert.

[

Um die Parallelisierung der Workflow-Ausführung zu nutzen, müssen Sie zuerst
[Ihr Workflow im KNIME Hub](#)

[Ausführung](#)

Nach dem Lesen der Daten und vor Ihrer ersten Datenmanipulation können Sie die Ich weiß nicht.
Start Knoten. Es spaltet Daten in gleich große Stücke für jede Iteration. Dann führt es die
Manipulationen im Schleifenkörper an jedem Stück parallel. Sie können entweder die
Anzahl der Stücke oder die Anzahl der Reihen pro Stück. Am Ende eines parallelen Blocks,
den Parallel Chunk End-Knoten hinzufügen. Sie sammelt und verdichtet die Ergebnisse der Parallelen
Arbeitsabläufe. Wenn das Datensegment in der Schleife Tabellen mit wiederholten RowIDs generiert,
die Parallel Chunk Endknoten fügt der RowID einen Blockindex hinzu.

Weitere Informationen zu Schleifen finden Sie in der [Loops Abschnitt der KNIME Flow Control](#)
[Leitfaden.](#)

Effiziente Datenbanknutzung

KNIME Analytics Platform ermöglicht es Ihnen, auf große Mengen zuzugreifen, zu manipulieren und zu teilen Daten. Die sicherste und effektive Weise, diese Daten zu speichern, ist in einer (relativen) Datenbank.

Mit [KNIME Datenbankerweiterung](#), die bereits Teil jeder KNIME Analytics Platform ist Installation können Sie die volle Funktionalität einer JDBC-konformen Datenbank innerhalb von KNIME nutzen. Dazu gehören die folgenden Datenbankoperationen:

- Verbindung zur Datenbank
 - Lesen aus der Datenbank
 - Abfragen in der Datenbank generieren
 - Verwaltung der Datenbankstruktur, d.h. Erstellung und Löschung von Tabellen in der Datenbank
 - Verwalten von Elementen in einer Datenbanktabelle, z.B. Löschen oder Aktualisieren von Werten
 - Typ Mapping, d.h. Festlegung von Regeln für die Zuordnung von Datenbanktypen zu KNIME-Typen und Vize
- Viel Spaß. Lesen Sie mehr dazu in [der jeweilige Abschnitt der KNIME Datenbankerweiterung Leitfaden](#).

Um mit der Datenbank Ihrer Wahl zu verbinden, bietet KNIME Connector-Knoten an, die gefunden in verschiedenen Erweiterungen. Es gibt drei Möglichkeiten, eine Datenbank in KNIME zu verbinden:

1. ANHANG Verbinden Sie mit einer vordefinierten Datenbank, die einen eigenen dedizierten Knoten in der [KNIME Datenbankerweiterung](#). Unter anderem PostgreSQL, MySQL, Microsoft SQL Server oder SQLite hat dedizierte Knoten.
2. Verbinden Sie eine vordefinierte Datenbank mit einem eigenen dedizierten Knoten in einem anderen dedizierten Knoten KNIME Erweiterung, was z.B. für die KNIME Amazon Athena der Fall ist Connector[Amazon Athena Connector Extension] oder die [KNIME Integration von Snowflake](#)
3. Verbinden Sie mit einer anderen JDBC-konformen Datenbank mit dem generischen DB Connector-Knoten.



Die datenbankspezifischen Verbindungsknoten enthalten bereits die notwendigen JDBC Treiber und einen Konfigurationsdialog zur Verfügung stellen, der auf die spezifischen Datenbank. Es wird empfohlen, diese Knoten anstelle der generischen DB zu verwenden Verbindungsknoten, wenn möglich. Erfahren Sie mehr über Datenbank-Konnektivität in der [jeweiliger Abschnitt des KNIME Extension Guide](#).

Sobald Sie angeschlossen sind, können Sie die Beziehungen mit dem DB Tabellenauswahl Node und lesen aus dem Datenbank mit Hilfe der DB Reader Knoten.

Neben dem Lesen aus der Datenbank können Sie in-Datenbank-Verarbeitung durchführen, d.h. Daten

Manipulation direkt auf der Datenbank mit der [KNIME Datenbank](#) Knoten, z. DB Joiner oder DB Gruppe und formulieren Sie Ihre Anfrage.

Ihre Konfiguration ist ähnlich wie die KNIME Manipulationsknoten. Die DB Query Nodes erfordern keine SQL-Codierung, da die Knoten selbst eine SQL-Abfrage konstruieren und ausgeben.

Schließlich können Sie Daten über die Datenbank schreiben Nodes erstellen, aktualisieren und löschen

Ihre Wahl. Die [DB Verbindungstisch Schreiber](#) node ist nützlich für Effizienz und Datenleitung

Wartung, weil es eine neue Datenbanktabelle basierend auf der Eingabe auf der SQL-Abfrage erstellt.

Dadurch ist es nicht erforderlich, die Daten in die KNIME Analytics Platform selbst zu lesen.

Zusätzlich, die [DB Loader](#) node können Sie Zeilen in eine bestehende DB-Tabelle einfügen, obwohl diese wird nicht von allen Datenbanken unterstützt.

KNIME Analytics Platform unterstützt auch NoSQL und nicht-relationale Datenbanken wie

[MongoDB](#) (documentorientiert), [DynamoDB](#) (dokumentorientiert und schlüsselwertbasiert) und [Neo4j](#) (Grafik)

Um Ihren Workflow in Bezug auf Datenbanken zu beschleunigen, beachten Sie folgende Richtlinien:

- Vermeiden Sie die Verwendung des [DB Connector](#) Knoten mehr als einmal bei der Verbindung mit dem gleichen Datenbank. Es sei denn, Sie müssen parallele Verbindungen öffnen, um die Ausführung von Parallelzweige, diese Praxis verbraucht mehr Rechenressourcen.
- Verarbeiten Sie Ihre Daten nicht lokal. Machen Sie es direkt auf der Datenbank, wann immer möglich.
- Verwenden Sie immer [DB Anschluss schließen](#) Knoten am Ende Ihrer Datenbankverarbeitung. Es eine DB-Session während der Workflow-Ausführung schließt und Datenbankressourcen freigibt, sobald sie vom Workflow nicht mehr benötigt werden.

Die KNIME Datenbankerweiterung umfasst Knoten zur Simulation von Datenbanktransaktionen, wie

wie [DB Transaction Start](#) und [DB Transaction End](#) . Diese Knoten helfen Gruppe multiple Datenbank Operationen in eine Einheit, die sicherstellen, dass entweder alle Operationen ausgeführt werden oder keine. Wenn Transaktion ist erfolgreich, es endet mit einem Commit in die Datenbank. Wenn es nicht erfolgreich ist, ist es zurückgerollt, die Datenbank zu seinem Zustand zu Beginn der Transaktion zurückgegeben.

Dies erhöht die Effizienz durch:

1. ANHANG Ermöglichen einer zuverlässigen Wiederherstellung von Fehlern, Beibehaltung der Datenbankkonsistenz, auch wenn Ausführung stoppt vorzeitig, Vermeidung unvollständiger Operationen und unklare Zustände.
2. Sicherstellung der Isolation zwischen gleichzeitigen Programmen, die auf die Datenbank zugreifen, potenzielle Fehler in Programmergebnissen.

Erfahren Sie mehr über Datenbank-Fähigkeiten mit KNIME und was Sie für verschiedene

Datenbankanbieter in der [KNIME Leitfaden für die Erweiterung](#) .

Machen Sie Ihren Workflow sicher

Anmeldeinformationen entsprechend ausfüllen

Bei der Arbeit mit vertraulichen Daten ist es wichtig, die besten Sicherheitsmaßnahmen zu treffen, sie vor Verletzungen und Missbrauch zu schützen. Der Schutz sensibler Informationen ist für regulatorische Compliance und Aufrechterhaltung des Vertrauens von Kunden und Interessenvertretern.

KNIME Analytics Platform hilft, verschiedene Sicherheitsfunktionen bereitzustellen. Es umfasst Werkzeuge Daten verschlüsseln, Benutzerzugriff verwalten und die Datenintegrität sicherstellen, damit Sie empfindlich umgehen können Informationen.

Erstellen von Knoten

Um sicherzustellen, dass vertrauliche Daten nicht versehentlich gespeichert oder geteilt werden, stellen Sie sicher, dass Sie keine hart kodierte Informationen in Ihren Knoten und Workflows speichern. Im einzelnen bedeutet dies die folgendes:

- Speichern Sie keine Passwörter im Klartext, da sie einfach zu kopieren sind.
- Verschlüsseln Sie Ihre Anmeldeinformationen nicht in Knoten.
- Setzen Sie Ihren Workflow zurück, bevor Sie ihn überall teilen.

Sie können vermeiden, Passwörter als Teil Ihres Workflows auf Festplatte standardmäßig zu speichern. Navigieren

die [knime.ini](#) [Datei in der Wurzel Ihrer lokalen KNIME-Installation](#), und gesetzt

`Dknime.settings.passwords.forbidden = true`.

Eine andere Möglichkeit, Anmeldeinformationen zu sichern, ist, sie zu verschlüsseln. Die Konfigurieren von Anmeldeinformationen der Einsteiger Widget nodes ermöglichen es dem Benutzer, Anmeldeinformationen zu Laufzeit anzugeben.

Konfigurationsknoten liefern Eingabeparameter an andere Knoten im Workflow. Sie tun das durch Durchflussgrößen. Dies bedeutet, dass wenn der variable Ausgang des Konfigurationsknotens an einen anderen Knoten angeschlossen ist, der auf die in der Konfigurationsknoten.

Widget-Knoten hingegen werden nur in den zusammengesetzten Ansichten als Widgets angezeigt. Das bedeutet, dass Sie bei der Betrachtung der Gesamtansicht in der KNIME Analytics Platform Einstellen der Parameter im Fenster, das öffnet.

Mit Hilfe von Widget- und Konfigurationsknoten, um Anmeldeinformationen zu handhaben, bedeutet der Benutzer nicht muss auf den Knoten zugreifen, der die Aktion durchführt, für die die Anmeldeinformationen benötigt werden. Stattdessen kann der Benutzer von außen auf die erforderlichen Einstellungen zugreifen, ohne dass

die Konfiguration des Knotens selbst ändern.

Sie können Konfigurations- und Widget-Knoten verwenden, um den Datenbankstecker sicher zu bedienen, Authentisierung und REST-Knoten. Unten ist ein Beispiel dafür, wie ein Datenbank-Connector-Knoten Zugriffe in einem Konfigurieren von Anmeldeinformationen-Knoten.

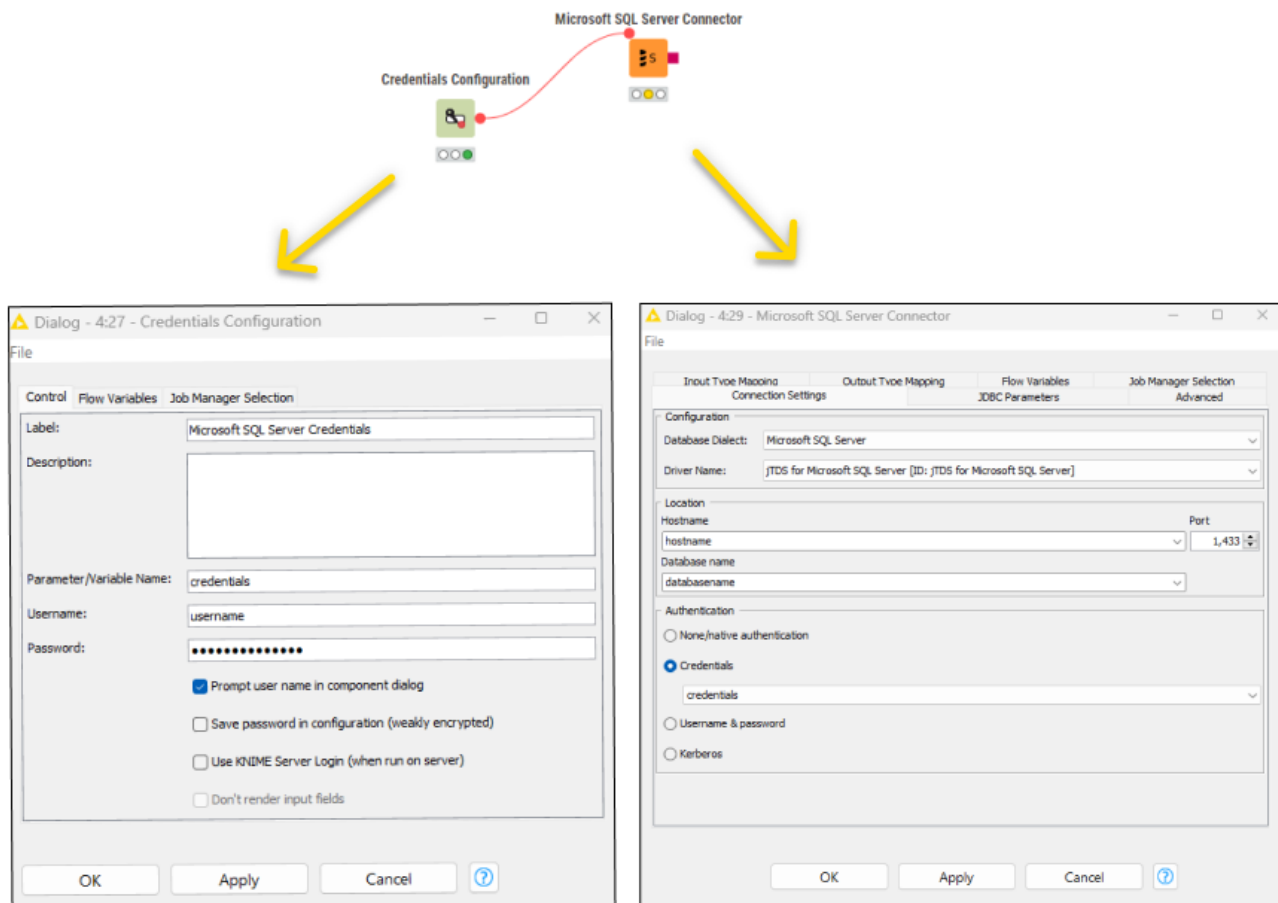


Abbildung 10. Umgang mit Anmeldeinformationen mit dem Anmelde-Konfigurationsknoten

Erfahren Sie mehr über Configuration nodes in the [jeweilige Sektion in den KNIME Komponenten Leitfaden](#). Die gleiche Führung bietet auch [einen Abschnitt auf Widget-Knoten](#). Auch.

Cloud-Anmeldeinformationen

Um Ihre Workflows flexibel zu gestalten, können Sie auf Remote-Dateisysteme zugreifen. Bei Verbindung mit einem Cloud-Service wie Microsoft, Google, etc., sollten Sie entsprechende Connector-Knoten verwenden, um sicher authentifizieren. Im Konfigurationsdialog eines Authentisierungsknotens können Sie die Zugangsmöglichkeiten, die für diese Verbindung gewährt werden. Sie sollten die Bereiche auswählen auf der Ebene des Zugriffs, den Sie benötigen. Die AnwendungsbereichSchaltfläche erlaubt es Ihnen, hinzuzufügen zusätzliche Reichweiten im Cloud-Service. Neben der Standardliste der Anwendungsbereiche können Sie mehrere benutzerdefinierte Bereiche hinzufügen.



Um den entsprechenden Scope unter der Standardliste der Scopes zu sehen, Sie muss die Erweiterung zuerst installiert haben.

Umgang mit sensiblen Daten

Wenn Sie einen Workflow teilen, können Sie wählen, um Daten zu speichern, die zusammen mit ihm verarbeitet wurden. Zwischenergebnisse von Datenmanipulationsoperationen sind dann auch nach dem Der Arbeitsablauf ist geschlossen. Insbesondere bei der Verarbeitung vertraulicher Daten innerhalb Ihres Workflows, Sie sollten sicherstellen, dass keine Informationen enthalten, die Sie nicht teilen möchten. Wann mit sensiblen Daten zu arbeiten, beachten Sie folgendes:

- Speichern Sie nicht vertrauliche Daten innerhalb Ihres Workflows
- Setzen Sie Ihren Workflow zurück, bevor Sie ihn teilen, es sei denn, Sie haben einen bestimmten Grund nicht. Das ist [beschrieben in der jeweilige Abschnitt des KNIME Community Hub User Guide](#).



Artikel, die Sie auf einen Platz auf dem KNIME Hub hochladen, werden für jeden verfügbar sein mit Zugang zu diesem Raum. Seien Sie daher sehr vorsichtig mit den Daten, die Sie teilen.

Datenanonymisierung

Sensitive Daten müssen mit Vorsicht behandelt werden. DSGVO (Datenschutzgrundverordnung) sieht vor, dass nur anonymisierte Daten weitgehend verwendet werden können. Als Folge können Sie Ihre vertraulichen Daten anonymisieren möchten, während Sie später die Person erneut identifizieren können Datenpunkte im Datensatz. Dies wird als Pseudo-Anonymisierung bezeichnet. Zusätzlich die Risiken der De-Anonymisierung bewerten, können Sie die [Redfield Privacy Nodes Erweiterung für KNIME](#) die verschiedene Algorithmen für die Anonymisierung und die Bewertung von Re-Identifikationsrisiken.

Für die grundlegende Anonymisierung (pseudo-anonymization) können Sie die [Anonymisierung Knoten](#). In den Konfigurationsdialog, wählen Sie einfach die identifizierenden Attribute, d.h. die Spalten in Ihrem Tabelle, die sensible Informationen enthalten. Die [Anonymisierung](#) die Werte in der Spalten mit SHA-1. Sie können auch zwischen verschiedenen Salzstrategien wählen. Eine Vorteil dieses eher unkomplizierten Ansatzes ist, dass Sie auf die ursprünglichen Daten zurückgehen können, da die Übersetzungstabelle am zweiten Ausgabeport verfügbar ist.

Für eine fortgeschrittene Anonymisierung sollten Sie sich auf Hierarchien verlassen. Sie können sie bauen, indem Sie definieren komplexe binning-Regeln, die die unmodifizierten Daten in vollständig unterdrückte Daten umwandeln. Sie können verschiedene Arten von Hierarchien mit den [Hierarchie erstellen Knoten](#). mehrere Hierarchien können nachträglich auf die Daten angewendet werden. Um mehr über die Hierarchietypen zu erfahren, Bitte beachten Sie die [Gebäudehierarchie Abschnitt des folgenden Artikels: Anonym bleiben mit Redfields Privacy Extension](#).

Um nun die auf die ursprünglichen Daten gesetzten Regeln anzuwenden, müssen Sie sie in der [Hierarchie Anonymisierung Knoten](#).

Um endlich zu beurteilen, welche Risiken durch Ihre Anonymisierungstechniken entstehen,

Anonymitätsbewertung Knoten. Um mehr über die Arten von Risiken zu erfahren, die der Knoten

kennzeichnet, bitte verweisen auf die

[Knotenseite des Knotens Anonymity Assessment auf dem KNIME](#)

[Community Hub](#) .

Secrets: Sichere Informationen zum KNIME Hub

Verwalten und verwenden Sie Geheimnisse

Secrets bieten einen Weg, um zentral zu speichern und verwalten Logins zu anderen Systemen. Zum Beispiel a geheim könnte Anmeldeinformationen sein, um sich in eine externe Datenbank, Dateisystem oder Dienst einzuloggen. Geheimnisse werden von einem Benutzer oder Team verwaltet und verwaltet. Benutzergeheimnisse sind für die Verwaltung von persönlichen Anmeldungen, z.B. john.smith. Team-Geheimnisse auf der anderen Seite sind für gemeinsame Anmeldungen bestimmt manchmal als technische oder Service-Nutzer bezeichnet, z.B. hr_read_only, die mit mehrere Benutzer.

Um Ihre persönlichen Geheimnisse zu verwalten navigieren auf Ihre Kontoseite und wählen [Geheimnisse von](#) Menü links. Auf Ihrer Secrets Seite können Sie Ihre persönlichen Geheimnisse erstellen, bearbeiten und löschen. Ebenso, um die Geheimnisse Ihres Teams zu verwalten, navigieren Sie auf die Seite des Teams und wählen Sie [Geheimnisse](#) aus dem Menü links. Auf deinem [Geheimnisse](#)Seite, die Sie erstellen, bearbeiten und löschen können [Geheimnisse, wie in der](#) [jeweiliger Abschnitt des KNIME Secrets User Guide](#).

□ Team-Geheimnisse können nur von Team-Admins erstellt werden.

Secrets ermöglichen es Ihnen, sensible Informationen wie Passwörter und API-Schlüssel zu verwalten. Jedes Geheimnis hat einen einzigartigen Namen, eine optionale Beschreibung, einen geheimen Typ und einen Authentifizierungstyp. Ein Geheimnis bearbeiten Änderungen an seinen Werten, aber nicht an seinem geheimen oder authentifizierten Typ, neues Geheimnis muss geschaffen werden. Die Löschung eines Geheimnisses erfordert die Bestätigung des Namens des Geheimnisses, um zu vermeiden versehentliche Löschungen. Einige geheime Typen wie OAuth2, benötigen Sie, um sich in Ihr Konto einzuloggen eine Sitzung zu erwerben und zu aktualisieren. Wenn Sie nicht eingeloggt sind, werden diese Geheimnisse markiert die Geheimnisse Tisch als Nicht verwertbar in der Statusspalte. Teamgeheimnisse können mit andere durch die Verwaltung von Zugriffsrechten, die auf "Benutzen" oder "Bearbeiten" gesetzt werden können. Diese Zugangsrechte Kontrollieren, ob Benutzer nur das Geheimnis in Workflows verwenden können oder auch ändern und löschen. Persönliche Geheimnisse können jedoch aus Sicherheitsgründen nicht mit anderen geteilt werden.

[Erfahren Sie mehr über Secrets in the](#) [KNIME Secrets Benutzerhandbuch](#).

Sie können auf Secrets in Ihren KNIME Workflows durch die [Secrets Retriever](#) Knoten, Teil von die [KNIME Hub Zusätzliche Connectivity \(Labs\) Erweiterung](#).

□ Die [Secrets Retriever](#) node benötigt eine Verbindung zu einem KNIME Business Hub.

Der Knoten ermöglicht die Auswahl mehrerer Geheimnisse, die Einstellung seiner Ausgangsports basierend auf dem Geheimnis Typen. Für die Sicherheit werden Geheimnisse nicht mit dem Workflow gespeichert und müssen neu ausgeführt werden Zeit. Der Knoten ruft Geheimnisse mit den Berechtigungen des ausführenden Benutzers ab, und wenn der Benutzer fehlt ein Fehler auftritt. Die Geheimnisse werden durch interne Kennungen referiert, um Namensänderungen zu gewährleisten keine Verbindungen stören.

Geheime Typen

Geheime Typen in KNIME sorgen für vielfältige Authentifizierungsanforderungen, Verbesserung der Workflow-Sicherheit und Flexibilität. Credentials können von verschiedenen Arten sein, wie Benutzernamen, Passwörter oder Passwort-Nur. So können Geheimnisse verschiedene Authentifizierungsszenarien wie Datenbank aufnehmen. Anmeldungen, etc. Der Box-Geheimtyp erleichtert die Verbindung zu Box zur Dateiverwaltung, unterstützt OAuth2-basierte Benutzerauthentifizierung. Ebenso ermöglicht der Google-Geheimtyp die Verbindung zu Google-Dienste wie Drive oder BigQuery. Microsoft geheime Typen erweitern Kompatibilität zu Microsoft/Azure Cloud-Plattformen, die Benutzern den Zugriff auf Office 365, SharePoint, Azure ermöglichen Dienstleistungen und mehr. Jeder Geheimtyp bietet spezifische Authentifizierungsmethoden, die verschiedene Anwendungsfälle. Stellen Sie sicher, den geeigneten geheimen Typ für interaktive Anmeldungen oder automatisierte Workflows. Darüber hinaus sorgt die Flexibilität, Zugriffsbereiche festzulegen, für feingekörte Kontrolle über Ressourcenberechtigungen, Verbesserung der Sicherheit.

Erfahren Sie mehr über Secrets auf dem KNIME Hub in der

[KNIME Secrets Benutzerhandbuch](#) .

Es bietet vielfältige Anleitungen, die systematische Verfahren zur Einbeziehung

Authentifizierung und Zugriffsberechtigungen in Microsoft Azure-Diensten in KNIME-Workflows.

Glossar


Workflow-Annotation

Eine Box mit farbigen Grenzen und optionalem Text, der in einen Workflow gesetzt werden kann, um einen und weitere Informationen zur Verfügung stellen. Klicken Sie mit der rechten Maustaste auf einen leeren Platz in einem Workflow und Wählen Sie „Neue Workflow-Annotation“ aus, um an dieser Position eine neue Workflow-Annotation zu erstellen.

Node Label

Der Text unter einem Knoten. Wählen Sie den Knoten aus und klicken Sie auf den Text. Dann können Sie es bearbeiten und Fügen Sie eine Beschreibung dessen hinzu, was der Knoten tut.

Beschreibung des Workflows

Die Beschreibung eines Workflows, der vom Benutzer bearbeitet werden kann. Klicken Sie auf die  an der Spitze rechts um es zu bearbeiten. Die Workflow-Beschreibung wird auch im KNIME Hub angezeigt, wenn ein Benutzer die Die Seite des Workflows.

Hubraum

Der Ort am KNIME Hub, an dem Workflows, Komponenten und Datendateien gespeichert werden. Finden mehr in der [Benutzerhandbuch des Community Hub](#) .

Stellen

Jedes Mal, wenn ein Workflow ad hoc ausgeführt wird oder eine Bereitstellung ausgeführt wird, wird ein Job erstellt KNIME Hub. Mehr Informationen zu Jobs finden [Hier](#) .

(Gelöscht)

Ein Workflow-Snippet, der in einen einzigen Knoten eingewickelt ist. Eine Komponente kann eine eigene Konfiguration aufweisen Dialog und kann auch eine Ansicht zugeordnet werden. Komponenten ermöglichen es Ihnen, die Funktionalität für eine einen bestimmten Teil des Workflows und ermöglichen das Teilen und Wiederverwenden. Für weitere Informationen siehe [KNIME Komponentenführung](#) .

Bereitstellung von Daten-Apps

Ein Workflow, bestehend aus einer oder mehreren Komponenten, die Ansichts- und Widget-Knoten enthalten kann mit im Browser interagieren. Eine Data App wird auf dem KNIME Business Hub eingesetzt. Eine Anleitung auf [Gebäude Daten Apps finden Sie](#) [Hier](#).

Aufstellung

Mit einem Zeitplan können Sie einen Workflow automatisch zu ausgewählten Zeiten einrichten. Mehr Informationen sind verfügbar:

- [Hier](#) für KNIME Business Hub.
- [Hier](#) für Teamplan zum KNIME Community Hub.

Bereitstellung von Service

Mit einer Service-Bereitstellung können Sie den Workflow als API-Endpunkt nutzen. Für mehr Informationen beziehen sich auf [KNIME Benutzerhandbuch des Business Hub](#).

Workflow Service

Ein Workflow mit [Workflow Service Input](#) und [Workflow Service Output](#) Knoten, die sein können von anderen Workflows mit [Call Workflow Service](#) Knoten.

KNIME AG
Talacker 50
8001 Zürich, Schweiz
www.knime.com
Info@knime.com