

# KNIME Benutzerhandbuch der Azure Integration

KNIME AG, Zürich, Schweiz

Version 5.7 (letzte Aktualisierung auf )



## Inhaltsverzeichnis

<a href="#page2" style="color: #000000; text-decoration: underline;">Überblick . . . . .</a>	<a href="#page2" style="color: #000000; text-decoration: underline;">Überblick . . . . .</a>
<a href="#page2" style="color: #000000; text-decoration: underline;">Azure Storage . . . . .</a>	<a href="#page2" style="color: #000000; text-decoration: underline;">Azure Storage . . . . .</a>
<a href="#page3" style="color: #000000; text-decoration: underline;">Azure Blob Storage</a>	<a href="#page3" style="color: #000000; text-decoration: underline;">Azure Blob Storage</a>
<a href="#page3" style="color: #000000; text-decoration: underline;">Azure Data Lake Storage</a>	<a href="#page3" style="color: #000000; text-decoration: underline;">Azure Data Lake Storage</a>
<a href="#page4" style="color: #000000; text-decoration: underline;">Azure SQL Database .</a>	<a href="#page4" style="color: #000000; text-decoration: underline;">Azure SQL Database .</a>
<a href="#page4" style="color: #000000; text-decoration: underline;">Verbinden Sie mit Azure SQL</a>	<a href="#page4" style="color: #000000; text-decoration: underline;">Verbinden Sie mit Azure SQL</a>
<a href="#page7" style="color: #000000; text-decoration: underline;">Azure HDInsight . . . .</a>	<a href="#page7" style="color: #000000; text-decoration: underline;">Azure HDInsight . . . .</a>
<a href="#page7" style="color: #000000; text-decoration: underline;">Cluster Setup . . . .</a>	<a href="#page7" style="color: #000000; text-decoration: underline;">Cluster Setup . . . .</a>
<a href="#page7" style="color: #000000; text-decoration: underline;">Verbinden Sie mit HDInsight</a>	<a href="#page7" style="color: #000000; text-decoration: underline;">Verbinden Sie mit HDInsight</a>
<a href="#page8" style="color: #000000; text-decoration: underline;">Apache Hive in Azure HDInsight</a>	<a href="#page8" style="color: #000000; text-decoration: underline;">Apache Hive in Azure HDInsight</a>

## Überblick

KNIME Analytics Platform enthält eine Reihe von Knoten, um Azure Cloud-Dienste zu unterstützen. Die [Azure HDInsight](#), [Azure Lagerung von Kugeln](#), [Azure Data Lake Storage Gen2](#), und [A.S. SQL](#).

Das KNIME Azure Cloud Connectors Erweiterung ist auf [KNIME Hubraum](#).

## Azure Storage

[KNIME Azure Cloud Connectors](#) Erweiterung bietet Knoten zur Verbindung mit Azure-Speicher (Blob Lagerung und Data Lake Storage Gen2).

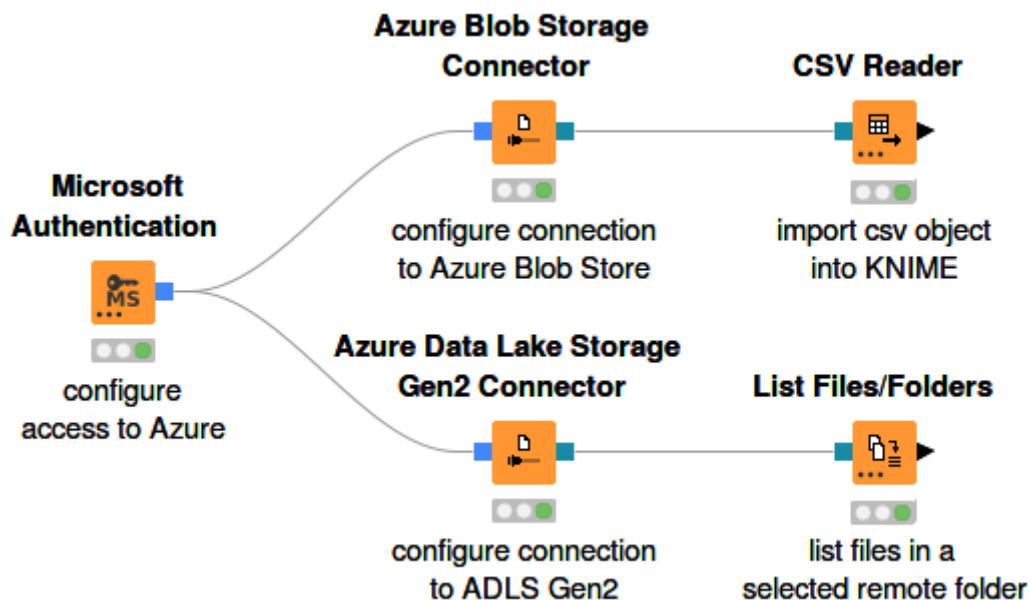


Abbildung 1. Verbinden mit und arbeiten mit Azure-Speicher

[Abbildung 1](#) zeigt ein Beispiel für die Verbindung mit einem Azure-Speicher und die Arbeit mit der Fernbedienung

Dateien.

[Microsoft Authentifizierung](#) node bietet eine Authentifizierung auf Microsoft Azure. Die

Folgende Authentifizierungsmodi werden unterstützt:

- Interaktive Authentifizierung
- Benutzername/Passwort-Authentifizierung
- Geteilte Schlüsselauthentifizierung (nur Azure Storage)
- Geteilte Zugriffssignatur (SAS) Authentifizierung (nur Azure Storage)



## A.S. SQL Datenbank

KNIME Analytics Platform enthält eine Reihe von [Datenbankknoten](#) die Verbindung und Arbeiten mit [A.S. SQL Datenbank](#) .

Die Einrichtung der KNIME Analytics Platform for Azure SQL hat folgende Voraussetzungen:

ANHANG Ein aktives Azure Abonnement. Für weitere Informationen zum Erstellen einer, bitte überprüfen

! [Dokumentation der Azure](#) .

2. Erstellen Sie die SQL-Datenbank (z.B. einzelne Datenbank). Weitere Informationen zur Erstellung

eine SQL-Datenbank auf dem Azure-Portal, folgen Sie bitte der [Dokumentation der Azure](#) .

### Verbindung mit Azure SQL Datenbank

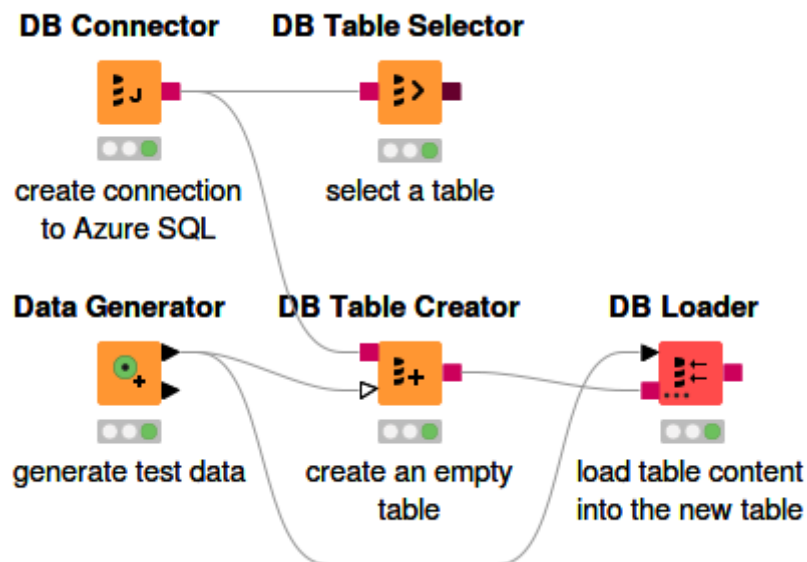


Abbildung 2. Verbindung mit Azure SQL-Datenbank mit DB Connector-Knoten

Die [DB Connector](#) [Knoten](#) oder [Microsoft SQL Server Connector](#) Knoten kann verwendet werden, um zu verbinden zu Azure SQL. Der erste Schritt ist die Installation der [offizieller Treiber für Microsoft SQL Server](#) in KNIME Analyseplattform. Bitte folgen Sie dem Tutorial, wie Sie den Microsoft SQL Server installieren JDBC-Treiber in der KNIME Analytics Platform [KNIME Datenbankdokumentation](#) .

Der Standard [TDS für Microsoft SQL Servertreiber](#) die mit dem Microsoft SQL Server Connector-Knoten unterstützt nicht einige Funktionen, wie der DB Loader-Knoten.

[Abbildung 2](#)

zeigt, wie man mit dem DB Connector-Knoten die Azure SQL-Datenbank verbindet. Der Knoten [Microsoft SQL Server Connector](#) unterstützt nicht einige Funktionen, wie der DB Loader-Knoten. Die Datenbank-URL sollte wie folgt aussehen:

```
jdbc:sqlserver://.database.windows.net;databaseName=
```

Wo:

- Insgesamt ist der Hostname oder der Name des Servers, der während der [Datenbank Schöpfung](#)
- ist der Datenbankport. Der Standardwert beträgt 1433.
- ist der Name der erstellten Datenbank.

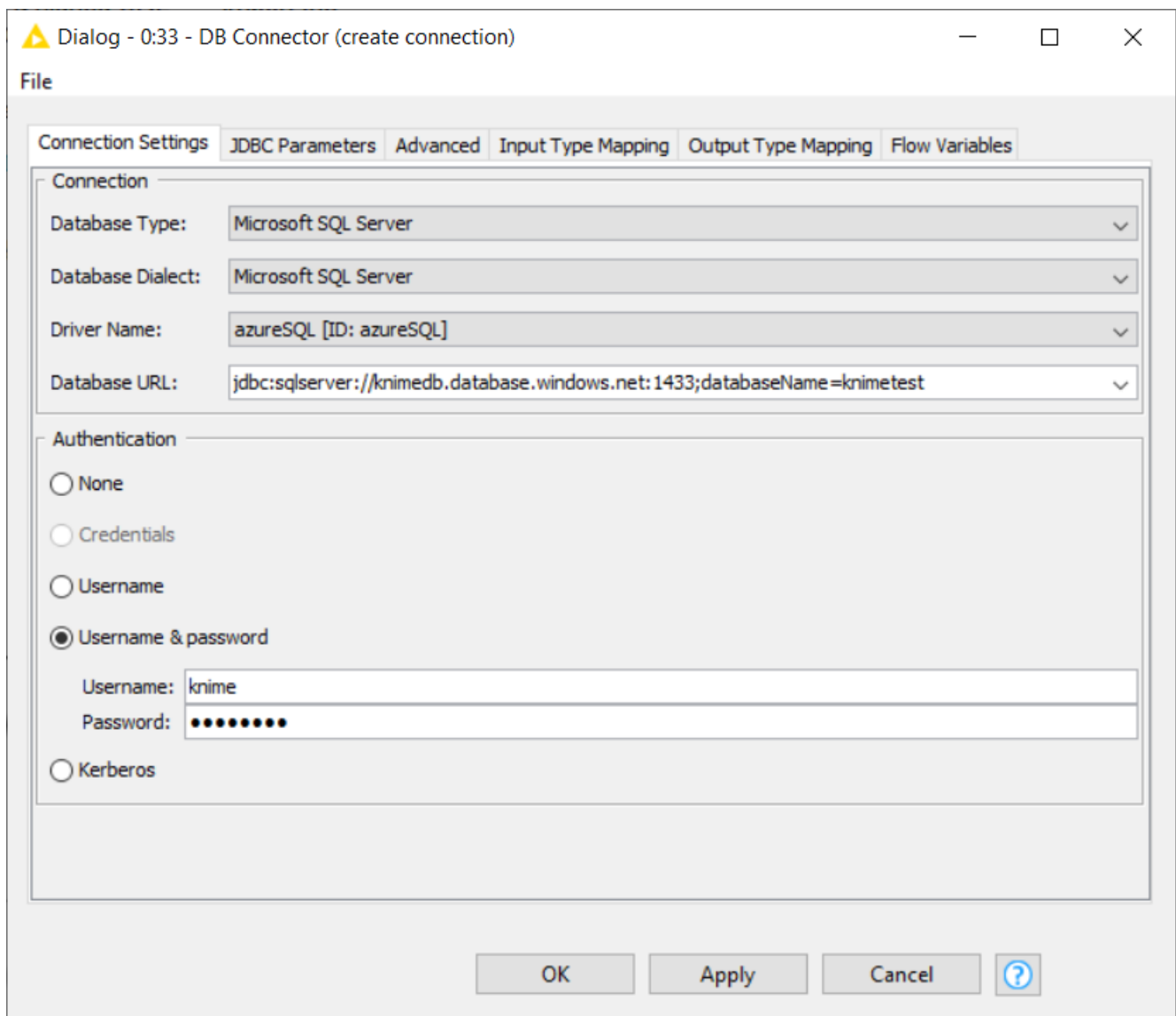


Abbildung 3. DB Connector Node Konfiguration Dialog

Die Datenbankauthentifizierung ist die Serverauthentifizierung, d.h. Server-Admin-Login und Passwort.



Für weitere Informationen über die JDBC Parameter Tab oder Erweiterte Tab in der Knotenkonfigurationsdialog des DB Connector-Knotens, bitte überprüfen Sie den [KNIME Dokumentation](#) .

Durch die Ausführung dieses Knotens wird eine Verbindung zur Azure SQL Datenbank erstellt und Sie können jede [KNIME Datenbankknoten](#) Ihre SQL-Anweisungen visuell zusammenstellen.



Weitere Informationen zu KNIME-Datenbankknoten finden Sie in der [KNIME Datenbankdokumentation](#) .



Ein Beispiel-Workflow zum Nachweis der Nutzung des Microsoft SQL Servers Verbindungsknoten, um AzureSQL von der KNIME Analytics Plattform zu verbinden ist verfügbar auf [KNIME Hubraum](#) .

## Azure HDInsight

### Cluster Setup

Um einen Azure HDInsight Cluster mit dem [Azure Portal](#), folgen Sie der Schritt-für-Schritt-Führung [bereitgestellt durch Dokumentation der Azteken](#). Bei der Cluster-Erstellung sind die folgenden Einstellungen wichtig:

- Cluster-Anmeldeinformationen. In diesem Abschnitt müssen Sie Anmeldeinformationen zum Zugriff und den Cluster verwalten. Bitte beachten Sie den Benutzernamen und das Passwort des Cluster-Login, die später benötigt wird, um sich über die KNIME Analytics Platform mit ihr zu verbinden.
- Speicherkonto. Ein Speicherkonto enthält alle Ihre Azure Storage-Objekte. Die Speicherkonto bietet einen einzigartigen Namensraum für Ihre Azure Storage-Daten, die von überall zugänglich, darunter ein HDInsight-Cluster.
- Clustertyp. Der Cluster-Typ definiert die Dienste, die für Ihre Cluster. Wählen Sie zum Beispiel Apache Spark aus, um die Spark-Verarbeitung auf dem Cluster zu ermöglichen.

HDInsight-Cluster stellen nur drei Ports öffentlich aus: 22, 23, und 443. Für mehr Informationen zu den Ports, die von Apache Hadoop-Diensten mit HDInsight verwendet werden Cluster, bitte überprüfen Sie die [Dokumentation der Azteken](#).

### Verbinden mit HDInsight Cluster

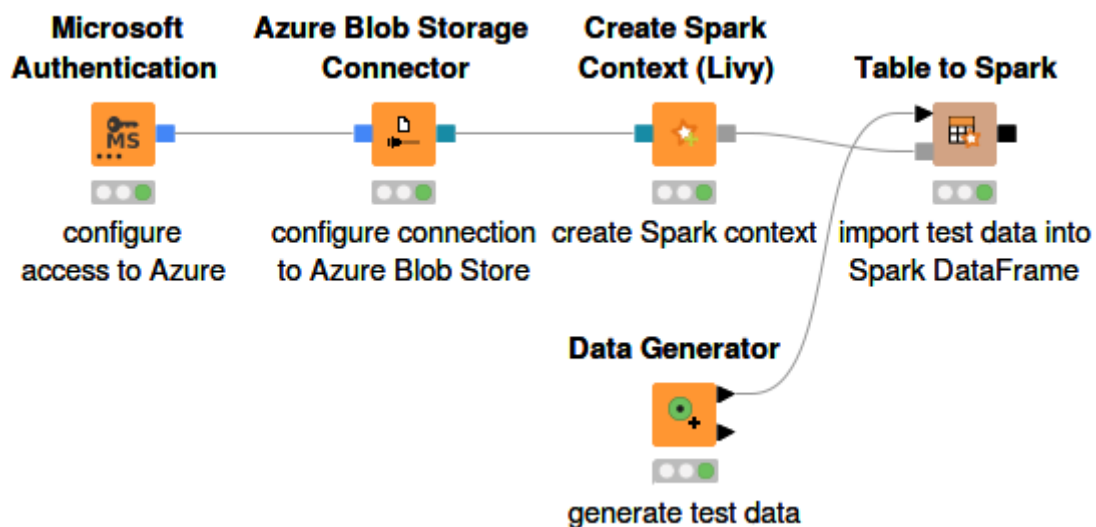


Abbildung 4. Verbindung mit HDInsight Cluster

[Abbildung 4](#) zeigt, wie eine Verbindung zu einem laufenden HDInsight Spark Cluster über KNIME aufgebaut wird

Analyseplattform. Der Microsoft Authentication-Knoten bietet Authentifizierung für den Zugriff



Dienstleistungen von Azure. Der Azure Blob Storage Connector Knoten wird verwendet, um eine Verbindung mit Azure Blob Aufbewahrung. Weitere Informationen finden Sie in der [Führung](#).

Die [Spark Context \(Livy\) erstellen](#) node erstellt einen Spark-Kontext [Apokalypse](#). Im Inneren der Knotenkonfiguration Dialog, die wichtigsten Einstellungen sind:

- Funkversion. Bitte achten Sie darauf, dass die Spark-Version der Spark-Version entspricht den Cluster.

- Die Livy URL. Es hat das Format `http://name.azurehdinsight.net:443/livy` wenn ist der Name des HDInsight Cluster.

- Authentifizierung. Geben Sie in diesem Feld den Benutzernamen und das Passwort des Cluster-Login ein. Bitte! [Cluster Setup](#) überprüfen [Abschnitt, um mehr Informationen über die Cluster-Anmeldeinformationen zu finden.](#)

- Unter [Erweiterte](#) Tab, es ist obligatorisch, die [Inszenierungsbereich für Spark Jobs](#). Die Inszenierung Fläche, die sich im angeschlossenen Azure Blob-Speichersystem befindet, wird verwendet, um temporäre Dateien zwischen der KNIME Analytics Platform und dem Spark-Kontext austauschen.

Die restlichen Einstellungen können nach Ihren Bedürfnissen konfiguriert werden. Für weitere Informationen den Spark Context (Livy)-Knoten erstellen, bitte überprüfen Sie die [KNIME Amazon Web Services Benutzerhandbuch](#).

Sobald der Spark-Kontext erstellt ist, können Sie eine beliebige Anzahl der KNIME Spark-Knoten von die [KNIME Erweiterung für Apache Spark](#) Ihre visuelle Montage Funkanalysestrom auf dem Cluster ausgeführt.

## Apache Hive in Azure HDInsight

Dieser Abschnitt beschreibt, wie eine Verbindung zu Apache Hive™ auf Azure HDInsight in KNIME Analytics Platform.

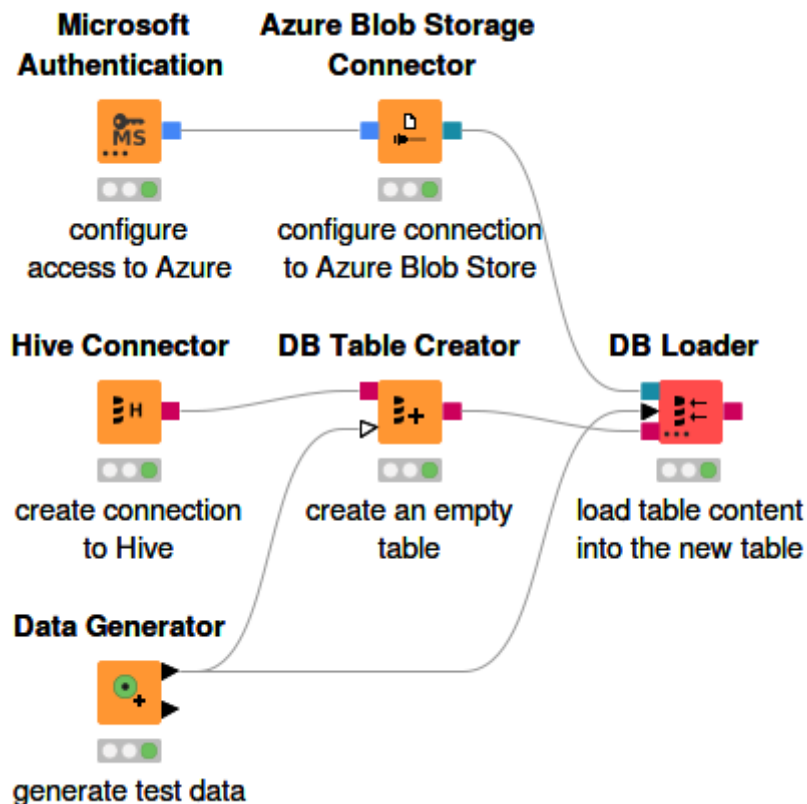


Abbildung 5. Verbinden Sie mit Hive und erstellen Sie einen Hive Tisch

[Abbildung 5](#page9) zeigt, wie sich Hive auf einem HDInsight-Cluster anschließt und wie man ein

Hive Tisch.

Der erste Schritt ist die Registrierung des Hive JDBC Treibers mit einer benutzerdefinierten JDBC URL. Folgen Sie der Führung

[über die Registrierung eines Hive JDBC Treibers in](#)

[KNIME Dokumentation](#)

. Jedoch für Hive auf Azure

HDInsight, geben Sie die folgende URL-Vorlage ein (siehe

[\)](#page10)

```
jdbc:hive2://:/default;transportMode=http;ssl=1;httpPath=/hive2
```

**Edit database driver settings for AzureHive**

**Driver**

ID:  Database type:

Name:

Description:

URL template:

▶ URL template syntax information ?

**Classpath**

Driver class:

Driver version: 2.6.0

Abbildung 6. Hive JDBC URL Vorlage

Sobald der Hive JDBC-Treiber registriert ist, können Sie den Hive Connector-Knoten konfigurieren. Die

[11](#page11), wo der Hostname der HDInsight Cluster ist

URL und die Anmeldeinformationen sind der Benutzername und das Passwort des Clusters (siehe [7](#page7) HDInsight Cluster Abschnitt für weitere Details). Es ist hier sehr wichtig, den Hafen auf 443 einzustellen.

anstatt der üblichen Hive Port 10000 oder 10001.

Für weitere Informationen über die Konfiguration der Einstellungen im Knotenkonfigurationsdialog,

Bitte beachten Sie die [KNIME Dokumentation](#) . Durch Ausführen des Knotens wird eine Verbindung zu [Apache Hive](#) und Sie können jede [KNIME Datenbankknoten](#) um Ihren SQL visuell zu montieren Aussagen.

Schlussfolgerung

Bitte stellen Sie sicher, dass Sie den Port auf 443 setzen, weil alle Verbindungen zum Cluster werden über ein sicheres Gateway verwaltet. Das bedeutet, Sie können nicht verbinden direkt Hive Server auf den Ports 10001 oder 10000, weil sie nicht exponiert sind an die Außenseite des virtuellen Netzwerks von Azure.

Dialog - 0:16 - Hive Connector (create connection)

File

Connection Settings | JDBC Parameters | Advanced | Input Type Mapping | Output Type Mapping | Flow Variables

Configuration

Database Dialect: Hive

Driver Name: AzureHive [ID: AzureHive]

Location

Hostname: knimetest.azurehdinsight.net Port: 443

Database name: default

Authentication

☐ Credentials

☐ Username

☒ Username & password

Username: admin

Password: ●●●●●●●●

☐ Kerberos

OK Apply Cancel ?

Abbildung 7. Hive Connector Node Konfiguration Dialog



Ein Beispiel Workflow, um die Nutzung von HDInsight von innen zu demonstrieren  
KNIME Analytics Platform ist auf [KNIME Hubraum](#) .

KNIME AG  
Talacker 50  
8001 Zürich, Schweiz  
[www.knime.com](http://www.knime.com)  
[Info@knime.com](mailto:Info@knime.com)