Open for Innovation
KNIME

Gesicherte Cluster Connection Guide

für KNIME Server

KNIME AG, Zürich, Schweiz

Version 5.7 (letzte Aktualisierung auf)

Inhaltsverzeichnis

Überblick

KNIME Server führt Workflows aus, die versuchen können, auf Kerberos gesicherte Dienste wie Apache HiveTM, Apache ImpalaTM und Apache Hadoop® HDFSTM.

Diese Anleitung beschreibt die Konfiguration von KNIME Server so, dass es **Authentizität** gegen Kerberos und dann **Impersonen** eigene Nutzer zu Kerberos-gesicherten Cluster-Diensten.

Was ist Benutzerinstanz?

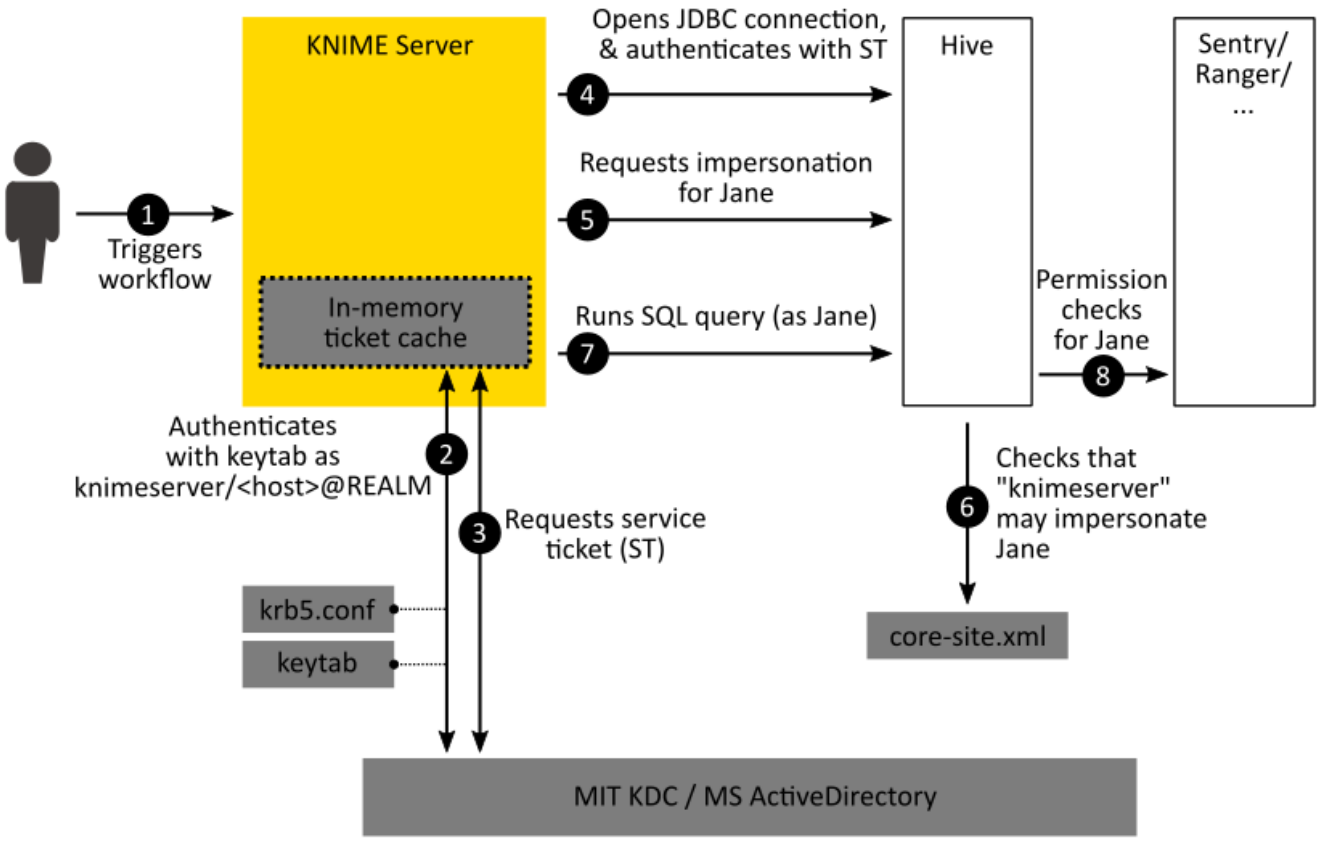
Bei Benutzereinstimmung spielt es keine Rolle, ob ein Benutzer einen Workflow in KNIME betreibt Analytics Platform oder auf KNIME Server. In beiden Fällen werden alle Operationen im Cluster durchgeführt **als besonderer Benutzer** und **gleiche Berechtigungen und Berechtigungsregeln** gelten. Dies hat folgende Vorteile:

- Workflows, die auf einen gesicherten Clusterlauf ohne Änderungen auf KNIME Server zugreifen.
- Zulassung zum Zugriff auf Clusterressourcen (Hive Tabellen, HDFS-Dateien, ...) wird verwaltet mit den üblichen Mechanismen, z.B. Apache SentryTM oder Apache RangerTM.

Wie funktioniert die User-Imitation?

Nehmen wir an, dass ein Benutzer Jane einen Workflow auf KNIME Server betreibt. Der Workflow soll eine Hive-Abfrage ausführen.

© 2025 KNIME AG. Alle Rechte vorbehalten.



Die folgende Reihenfolge der Ereignisse findet nun statt:

ANHANG Sie beginnt einen Workflow, der sich mit Hive verbindet. Dieser Workflow wird nun auf KNIME ausgeführt Server, nicht Janes Maschine.

2. Wenn Hive Connector Knoten im Workflow wird ausgeführt, KNIME Server erste Überprüfungen für einen TGT (Ticketbewilligung) in einem eigenen Ticket-Cache. Wenn es kein TGT gibt, liest es die krb5.conf Konfigurationsdatei, verbindet sich mit dem KDC und authentisiert sich. Anstatt Janes Anmeldedaten verwendet die auf dem KNIME Server konfigurierten Anmeldeinformationen, d.h. einen Dienst. für Kimeserver/@REALM und eine Keytab-Datei. Die TGT wird in einen In-Memory-Ticket-Cache.

3. Um eine JDBC-Verbindung zu Hive herzustellen, wird der Hive JDBC-Treiber auf dem KNIME Server noch erfordert ein ST (Service-Ticket), das es jetzt vom KDC anfordert. Die ST ist nur gültig für Verbindungen zwischen KNIME Server und der Hive-Instanz.

L 347 vom 20.12.2013, S. 1). Nun öffnet der Hive JDBC-Treiber eine Verbindung zu Hive und authentisiert sich mit dem ST Kimeserver/@REALM .

5. Da der Workflow von Jane gestartet wurde, sagt der JDBC-Treiber Hive, dass alle Operationen wird durchgeführt **als Benutzer Jane** .

6. Hive berät die Hadoop Core-Site.xml zu überprüfen, ob KNIME Server ist in der Tat erlaubt Jane zu verkörpern. Wenn nicht, wird es einen Fehler zurückgeben.

7. Nun legt der Workflow eine SQL-Abfrage über die JDBC-Verbindung vor. Die Abfrage ist

© 2025 KNIME AG. Alle Rechte vorbehalten.

ausgeführt auf dem Cluster **als Benutzer Jane** .

8. Hive prüft, ob der Benutzer Jane die erforderlichen Berechtigungen hat, um die Abfrage auszuführen. Es verwendet seinen üblichen Berechtigungsprüfmechanismus, z.B. Apache SentryTM oder Apache RangerTM. Die Abfrage wird gelingen oder scheitern, je nachdem, ob **Jane.** hat die nötige Berechtigungen.

Voraussetzungen

Einrichtung von KNIME Server für die Kerberos-Authentifizierung und Benutzereinstimmung die folgenden Voraussetzungen.

- Für Kerberos:
  - Ein vorhandenes Kerberos KDC wie MIT Kerberos oder Microsoft ActiveDirectory
  - Ein Service-Primär für KNIME Server. Das empfohlene Format ist Kimeserver/@ , wo
    - Insgesamt ist der vollqualifizierte Domainname der Maschine, in der KNIME Server läuft,
    - ist das Kerberos Reich.
  - Eine Keytab-Datei für den KNIME Server Service Principal.
  - Eine Kerberos Client-Konfigurationsdatei (krb5.conf) Die empfohlene Weise zu erhalten diese Datei, ist zu kopieren /etc/krb5.conf aus einem Knoten im Cluster. Alternativ, die Datei kann manuell erstellt werden (siehe Einrichtung krb5.conf)
- Für den Cluster:
  - Ein Kerberos-versicherter Cluster.
  - ein Konto mit administrativen Privilegien in der Cluster-Management-Software, die Clusterdienste konfigurieren und neu starten können. Auf Cloudera CDH bedeutet dies Cloudera Manager-Konto.
- Für KNIME Server:
  - Eine bestehende KNIME Server Installation.
  - Ein Konto mit administrativen Privilegien auf der Maschine, wo KNIME Server ist installiert. Diese Konten müssen in der Lage sein, die KNIME Server Konfiguration zu bearbeiten Dateien neu starten und KNIME Server neu starten.

© 2025 KNIME AG. Alle Rechte vorbehalten.

Unterstützte Clusterdienste

KNIME Server unterstützt die Kerberos-Authentifizierung und Benutzer-Imitation für Verbindungen zu
die folgenden Dienste:

- Apache Hive
- Apache Impala
- Apache Hadoop HDFS (einschließlich HttpFS)
- Apache Livy

Einrichten der Kerberos-Authentifizierung

Dieser Abschnitt beschreibt, wie man KNIME Server einrichten kann, um sich gegen Kerberos zu authentifizieren.

Kerberos Client Konfiguration (krb5.conf)

Das KNIME Server-Executor muss die krb5.conf Datei während der Kerberos Authentifizierung.
Ein gültigerkrb5.conf eine Datei wird benötigt. Das KNIME Server-Executor überprüft mehrere Standorte für
die krb5.conf Datei. Der Abschnitt Mögliche Standorte krb5.conf beschreibt das Verfahren nach
die KNIME Server-Executor ordnet die krb5.conf Datei. Falls der Standort der Datei ist
unbekannt oder die Datei nicht verfügbar ist, kontaktieren Sie bitte den lokalen Administrator.

Als Alternative, die krb5.conf die Datei kann manuell erstellt werden. Bitte konsultieren Sie den Abschnitt
Einrichtung krb5.conf von Kerberos Admin Guide für weitere Informationen und ein Beispiel
wie man eine einfache krbf5.conf Datei.

> Für Hadoop, wenn der Benutzer im selben Kerberos-Bereich wie der Hadoop-Cluster ist,
> dann /etc/krb5.conf Datei kann direkt von einem Cluster-Knoten heruntergeladen werden,
> z.B. mit WinSCP oder pscp aus PuTTY.

Kerberos Anpassungsprofile

Kerberos Konfigurationen, wie die krb5.conf Standort-, Service-Haupt-, Keytab-Werte und
viele andere, werden in einer Präferenzdatei (.epf Datei), die an alle verteilt werden kann
angeschlossene KNIME Serverausführungen über Anpassungsprofile .

Bitte konsultieren Sie die Kerberos Konfigurationstabelle für eine Liste aller unterstützten Kerberos
Konfigurationsoptionen und wie man sie in einer Präferenzdatei schreibt. Für einen ausführlichen Leitfaden

© 2025 KNIME AG. Alle Rechte vorbehalten.

um ein Anpassungsprofil zu erstellen, um Kerberos Präferenzen an alle KNIME Server zu verteilen
Executors, bitte überprüfen Sie die Kerberos Admin Guide .

Um Kerberos zu beheben, überprüfen Sie bitte die Fehlerbehebung Abschnitt der Kerberos Admin Guide .

Einrichtung eigener JDBC-Treiber (optional)

Jedes KNIME Server Executor ist eine kopflose Instanz von KNIME Analytics Plattform. Wenn KNIME Big Data Connectors Erweiterung installiert, KNIME Server Der Executor umfasst einen vollfunktionalen eingebetteten JDBC-Treiber für Hive und Impala. Ist die Verwendung dieses Fahrers bevorzugt, so kann dieser Abschnitt übersprungen werden.

Die **eingebettet** Apache Hive JDBC Treiber für Impala **nicht unterstützt** **Verkörperung** . Für die Verkörperung, wenn Sie mit Impala verbinden, bitte einrichten Eigener Fahrer .

Der aktuell eingebettete JDBC-Treiber ist der Open Source Apache HiveTM JDBC Treiberversion 1.1.0-cdh5.13.0 (Veröffentlichungshinweise) Der Fahrer wurde überprüft, mit CDH 5.3 und später kompatibel sein.

Wenn der Aufbau eines proprietären Cloudera JDBC Treibers für Hive/Impala **(empfohlen)** gewählt wird, Bitte konsultieren Sie die folgenden Abschnitte für einen Schritt-für-Schritt JDBC Fahrerregistrierungsführer je nach spezifischem Hadoop-Anbieter:

- Registrieren Hive Cloudera JDBC Treiber auf KNIME Server
- Registrieren Impala Cloudera JDBC Treiber auf KNIME Server

Das oben beschriebene JDBC-Treiberregistrierungsverfahren schafft auch eine Anpassungsprofil, um den Treiber an alle angeschlossenen KNIME Server zu verteilen Executors. Ist ein Profilordner bereits während der Authentifizierung Schritt, dann ist die Schaffung eines neuen in diesem Schritt nicht notwendig.

Einrichtung von Benutzer-Imitation

In diesem Abschnitt wird beschrieben, wie beide Enden der User-Imitation aufgebaut werden, was Konfiguration auf zwei Seiten: KNIME Server **und** den Cluster.

© 2025 KNIME AG. Alle Rechte vorbehalten.

Benutzereinstimmung auf KNIME Server

Standardmäßig, KNIME Server versucht, seine Benutzer auf Kerberos-gesicherten Verbindungen zu verkörpern zu folgenden Cluster-Diensten:

- HDFS (einschließlich httpFS)
- Apache Livy
- Apache Hive

Imperson für HDFS und Apache Livy wird automatisch durchgeführt und erfordert keine weitere Einrichtung. Verbindungen zu Apache Hive erfordern weitere Setup-Schritte abhängig von der verwendeten JDBC Treiber.

Schlussfolgerung

Die **eingebettet** Apache Hive JDBC Driver (für Impala) **nicht unterstützt** **Verkörperung** . Für die Verkörperung bei der Verbindung mit Impala bitte die Einrichtung Eigener Fahrer .

Wenn die Anweisungen zur JDBC-Treiberregistrierung in KNIME Server Anleitung für Hive oder Impala im voraus , die Benutzer-Impersonation Aktivierung ist bereits enthalten. Bitte! zum nächsten Abschnitt.

Abschnitt

Für **eingebettet** Apache Hive JDBC Treiber, bitte folgen Sie der Anleitung in der Benutzer Imitation auf Hive Abschnitt.

Prüfen Sie die Treiberdokumentation für den entsprechenden Identitätsparameter, wenn ein Drittel party JDBC driver ist im Einsatz.

Benutzerinfektion auf Apache HadoopTM und Apache HiveTM

Apache HadoopTM und Apache HiveTM beraten Core-Site.xml Datei zu bestimmen, ob KNIME Server ist erlaubt, Benutzer zu verkörpern.

Ändern der Core-Site.xml Datei muss über Ambari (auf HDP) oder Cloudera erfolgen Manager (auf CDH). Ein Neustart der betroffenen Hadoop-Dienste ist erforderlich.

Bitte fügen Sie die folgenden Einstellungen zum Hadoop hinzu Core-Site.xml auf dem Cluster:

© 2025 KNIME AG. Alle Rechte vorbehalten.

```
hadoop.proxyuser.knimeserver.hosts 1
*

hadoop.proxyuser.knimeserver.groups 1
*
```

1 Wenn ein Service-Prinzip für KNIME Server erstellt wurde, außer Kimeserver/@ , dann muss der Immobilienname entsprechend angepasst werden.

Benutzerinstitution auf Apache ImpalaTM

Apache ImpalaTM benötigt eine Konfigurationseinstellung, um festzustellen, ob KNIME Server den Benutzern erlaubt.

> Es wird empfohlen, auch Apache SentryTM Autorisierung in Apache aktivieren Impala . Andernfalls führt Impala alle Lese- und Schreibvorgänge mit dem Vorrechte der Impala Benutzer.

Die erforderlichen Schritte sind ähnlich Impala-Delegation für Hue konfigurieren . In Cloudera Manager, navigieren **Impala** > **Konfiguration** > **Impala Daemon Kommandozeile Argument Erweiterte Konfiguration Snippet (Sicherheitsventil)** und die folgende Zeile hinzufügen:

```
-autorisiert_proxy_user_config = 'hue = *;knimeserver = * '
```

Dann klicken Sie Speichern und alle Impala-Daemons neu starten.

Bitte beachten Sie:

- Dies wird Farbe und Das ist nicht möglich die einzigen Dienste, die Benutzer in Impala. Wenn andere Dienstleistungen das gleiche tun dürfen, müssen sie einbezogen werden Auch hier.
- Wenn ein Service-Primär für KNIME Server außer Kimeserver/@ war erstellt, dann die obige Einstellung entsprechend anpassen.

© 2025 KNIME AG. Alle Rechte vorbehalten.

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

Die Marke KNIME® und das Logo OPEN FOR INNOVATION® werden von der KNIME AG unter Lizenz verwendet von der KNIME GmbH, und sind in den USA registriert. KNIME® ist auch in Deutschland registriert.