

KNIME Große Datenerweiterungen Admin

Leitfaden

KNIME AG, Zürich, Schweiz

Version 5.7 (letzte Aktualisierung auf)



Inhaltsverzeichnis

Überblick
Allgemeine Kompatibilität
Die neue Version Kompatibilität
Die Welt der Welt Kompatibilität
Cloudera HDP Kompatibilität
Amazon EMR Kompatibilität
Phase der Unterstützungsfür
Apache Livy Setup
Cloudera CDP
Cloudera CDH
Cloudera HDP
Amazon EMR
Downloads
Apache Livy herunterladen

Überblick

KNIME Big Data Extensions integrieren Apache Spark und das Apache Hadoop Ökosystem mit KNIME Analytics Platform.

Dieser Leitfaden richtet sich an IT-Experten, die die KNIME Analytics Platform mit einer bestehende Hadoop/Spark-Umgebung.

Die Schritte in diesem Leitfaden sind erforderlich, damit Benutzer der KNIME Analytics Platform Spark laufen

Workflows. Beachten Sie, dass die Nutzung von Spark-Workflows auf KNIME Server erforderlich ist

zusätzlich Schritte

in [Gesicherter Cluster Connection Guide für KNIME Server](#)

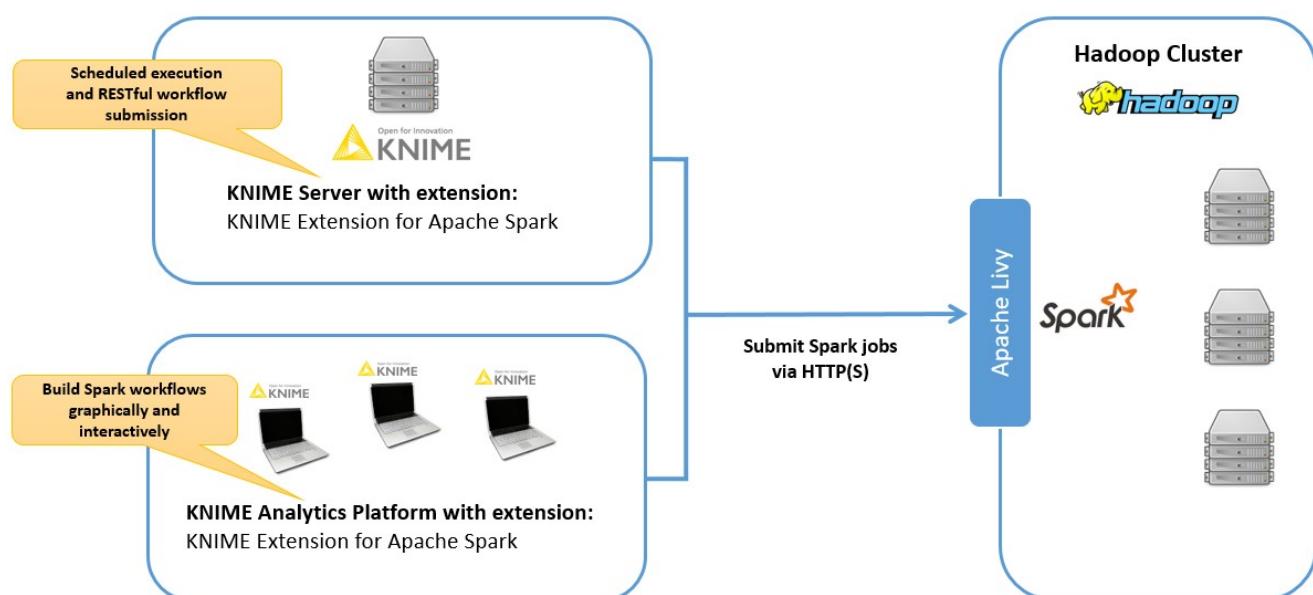


Abbildung 1. Gesamtarchitektur

KNIME Erweiterung für Apache Spark erforderlt [Apokalypse](#) als REST-Service auf einem Rand/Frontknoten des Clusters. Siehe die Anforderungen in den Kompatibilitätslisten unten und [Installation von Livy](#page4). Wie man

Allgemeine Kompatibilität

KNIME Erweiterung für Apache Spark ist kompatibel mit

- Spark 2.x - 3.5
- Livy 0.4 - 0.7

Die neue Version Vereinbarkeit

KNIME Erweiterung für Apache Spark ist kompatibel mit

- Funk 3.3 auf CDP 7.1.8, wie vorgesehen [Cloudera CDs 3.3](#)
- Spark 3.2 auf CDP 7.1.7 wie vorgesehen [Cloudera CDs 3.2](#)
- Spark 3.1 auf CDP 7.1.6 wie vorgesehen [Cloudera CDs 3.1](#)
- Spark 3.0 auf CDP 7.1.5 wie vorgesehen [Die Welt der Welt 3](#)
- Spark 2.4 wie in CDP 7 enthalten

Die Welt der Welt Vereinbarkeit

KNIME Erweiterung für Apache Spark ist kompatibel mit

- [Spark 2.x auf CDH 5 wie vorgesehen](#) [Die Welt der Welt](#)
- Spark 2.x wie in CDH 6 enthalten



Cloudera CDH 5/6 enthält nicht Livy, daher bietet KNIME < a href="#page4" style="color:#0000ff;text-decoration:none;outline:none;cursor:pointer;">CDH 5/6 ist nicht mit der KNIME Erweiterung für Apache Spark kompatibel

Die neue Version Vereinbarkeit

KNIME Erweiterung für Apache Spark ist kompatibel mit

- Spark 2.x wie in HDP 2.6.3 - 2.6.5 enthalten
- Spark 2.x wie in HDP enthalten 3.0.0 - 3.1.5

Amazon EMR Vereinbarkeit

KNIME Erweiterung für Apache Spark ist kompatibel mit

- EMR 7.x mit Spark 3.5 und Livy 0.7 - 0.8
- EMR 6.x mit Spark 3.x und Livy 0.6 - 0.7
- EMR 5.9+ mit Spark 2.x und Livy 0.4 - 0.7

Phase aus Unterstützung für H2O Sparkling Water

Beginnend mit der KNIME Analytics Platform Version 5.3 die Unterstützung für [H2O Sparkling Water](#) ist Phased out, und keine Unterstützung wird für kommende Spark-Versionen hinzugefügt. Das heißt, Sparkling Water wird nicht mehr auf Clustern mit Spark 3.4 oder höher unterstützt, z. Databricks. Derzeit wird nur noch der Knoten "Create Local Big Data Environment" unterstützt, Diese Unterstützung wird jedoch auch in naher Zukunft entfernt werden.

Apache Livy Setup

Die neue Version

Cloudera Runtime 7.0 - 7.1 enthält Spark 2.4 und Livy als Service. Cloudera povids Spark 3.x als benutzerdefinierter Service-Deskriptor, der mit der enthaltenen Spark-Version koexistiert. Vgl. Installation von CDS Angetrieben von Apache Spark in der [CDS 3.3](#) oder [CDS 3.2](#) oder [CDS 3.1](#) oder [CDS 3](#) Cloudera-Dokumentation für weitere Informationen. Beachten Sie, dass Livy im Spark 3.x CSD verwendet 28998 statt der üblichen 8998 als Standardport.

Wenn Sie Spark Workflows auf KNIME Server ausführen möchten: Bitte konsultieren Sie die [Gesicherter Cluster Connection Guide für KNIME Server](#) KNIME erlauben Server zu verkörpern.

Die Welt der Welt

Für Cloudera CDH bietet KNIME eine CSD und ein Paket, damit Livy als Add-On installiert werden kann. im Dienst. Die aktuelle Version von Livy für CDH von KNIME ist 0,5.0.knime3

Die folgenden Schritte beschreiben die Installation von Livy als Managed Service durch Cloudera Manager mit einem Paket. Wenn im Zweifel, bitte auch den offiziellen [Cloudera-Dokumentation zum Umgang mit Paketen](#)

Voraussetzungen

- Ein Cluster mit CDH 5.8 und neuer oder CDH 6.0 und neuer
 - Nur auf CDH 5 : Spark 2.2 oder höher als Add-on-Service (erstellt von [In den Warenkorb](#) [CDS](#))
- Root Shell Zugriff (z.B. über SSH) auf dem Gerät, in dem Cloudera Manager installiert ist.

- Voller administrativer Zugriff auf den Cloudera Manager WebUI.

Installationsschritte

In einer Root Shell auf der Maschine, in der Cloudera Manager installiert ist:

ANHANG Cloudera für einen öffentlichen Internetzugriff
bis /opt/cloudera/csd/ auf der

Maschine, wo Cloudera Manager installiert ist.

2. Nur wenn Cloudera Manager nicht auf das öffentliche Internet zugreifen kann:
Matching .parcel und .sh1 Datei aus
/opt/cloudera/parcelrepo
Download / Kopieren der
bis

3. Neues aus der Welt Manager aus der Befehlszeile, beispielsweise mit:

```
systemctl restart cloudera-scm-server
```

Im Cloudera Manager WebUI:

ANHANG Navigieren Sie zum Parcel-Manager und finden Sie die LIVY Paket.

2. Download (weniger bereits manuell gemacht), Verteilen und Aktivieren der LIVY Paket.

3. Fügen Sie den Livy Service zu Ihrem Cluster hinzu (siehe die offizielle Cloudera-Dokumentation über [Hinzufügen Dienstleistungen](#))

L 347 vom 20.12.2013, S. 1). Navigieren Sie zur HDFS-Servicekonfiguration und fügen Sie die folgenden Einstellungen hinzu
Clusterweite Advanced Configuration Snippet (Sicherheitsventil) für core-site.xml :

hadoop.proxyuser.livy.hosts = *

hadoop.proxyuser.livy.groups = *

5. Wenn Ihr Cluster verwendet wird [HDFS transparente Verschlüsselung](#) : Navigieren Sie zum KMS-Service

Konfiguration und Hinzufügen der folgenden Einstellungen Key Management Server erweitert

Konfiguration Snippet (Sicherheitsventil) für kms-site.xml :

hadoop.kms.proxyuser.livy.hosts = *

hadoop.kms.proxyuser.livy.groups = *

6. Wenn Sie Spark Workflows auf KNIME Server ausführen möchten: Bitte konsultieren Sie die [Gesicherter Cluster Anschlussanleitung für KNIME Server](#) KNIME erlauben Server, um Benutzer zu identifizieren.

7. Starten Sie alle von Ihrer Konfiguration betroffenen Dienste neu.

Die neue Version

HDP enthält bereits kompatible Versionen von Apache Livy und Spark 2 (siehe [Vereinbarkeit](#page3))

Bitte folgen Sie der jeweiligen Hortonworks-Dokumentation, um Spark mit

die Livy für Spark2 Server

Komponente:

- [Installation von Spark mit Ambari \(HDP 2.6.5\)](#)
- [Installieren Sie Spark mit Ambari \(HDP 3.1\)](#)

KNIME Erweiterung für Apache Spark nur unterstützt

Livy für Spark2 Server

die

Verwendung von Spark 2. Die Livy für Spark Server

Bauteil wird nicht unterstützt, da es

basierend auf Spark 1.

Amazon EMR

Amazon EMR enthält bereits kompatible Versionen von Apache Livy und Spark 2 (siehe [EMR Vereinbarkeit](#page3))

einfach sicherstellen, dass Sie Libyen in der Softwarekonfiguration Ihres

Cluster.

[EMR Vereinbarkeit](#page3)

Downloads

Apache Livy downloads

CSDs für Cloudera CDH

- [CSD für CDH 5](#)
- [CSD für CDH 6](#)

Pakete für Cloudera CDH

Links für CDH 5 herunterladen:

RHEL/CentOS	RHEL 7: Paket ↗ !	RHEL 6: Paket / !	RHEL 5: Paket / !
SIC!H	SLES 12: Paket ↗ !	SLES 11: Paket / !	

Ubuntu	Ubuntu 16 (Xenial): Paket / !	Ubuntu 14 (Trusty): Paket / !	Ubuntu 12 (Precise): Paket / !
Debian	Debian 8 (Jessie): Paket / !	Debian 7 (Wheezy): Paket / !	

Links für CDH 6 herunterladen

RHEL/CentOS	RHEL 7: Paket / !	RHEL 6: Paket / !	RHEL 5: Paket / !
SICH	SLES 12: Paket / !	SLES 11: Paket / !	
Ubuntu	Ubuntu 16 (Xenial): Paket / !	Ubuntu 14 (Trusty): Paket / !	Ubuntu 12 (Precise): Paket / !
Debian	Debian 8 (Jessie): Paket / !	Debian 7 (Wheezy): Paket / !	



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