

# KNIME Große Datenerweiterungen Admin

## Leitfaden

KNIME AG, Zürich, Schweiz

Version 5.7 (letzte Aktualisierung auf )



## Inhaltsverzeichnis

[Überblick . . . . .](#page2) [Allgemeine Kompatibilität](#page2)

[Die neue Version Kompatibilität](#page3)

[Die Welt der Welt Kompatibilität](#page3)

[Cloudera HDP Kompatibilität](#page3)

[Amazon EMR Kompatibilität](#page3)

[Phase der Unterstützung](#page4)

[Apache Livy Setup . . .](#page4) [Cloudera CDP. . .](#page4)

[Cloudera CDH. . .](#page4) [Cloudera HDP. . .](#page6)

[Amazon EMR . . .](#page6) [Downloads . . . . .](#page6)

[Apache Livy heruntergeladen](#page6)

# 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 eine 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](#).

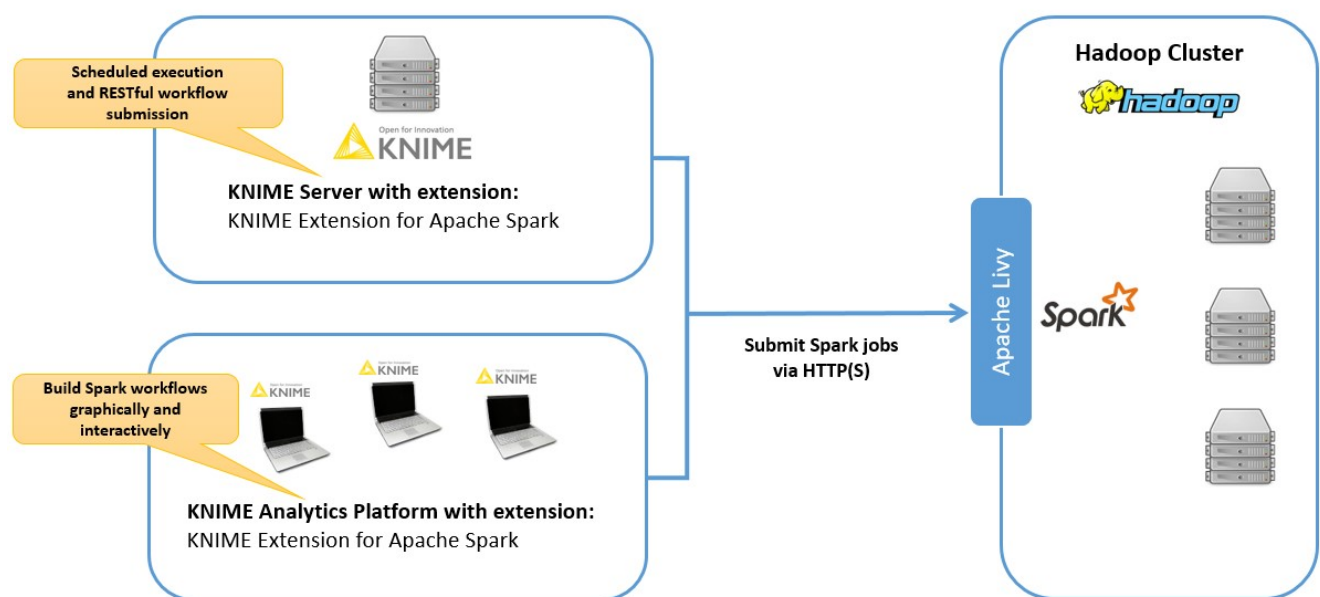


Abbildung 1. Gesamtarchitektur

KNIME Erweiterung für Apache Spark erfordert [Apokalypse](#) als REST-Service auf einem

Rand/Frontknoten des Clusters. Siehe die Anforderungen in den Kompatibilitätslisten unten und

Wie man [Installation von Livy](#page4)

## Allgemeine Kompatibilität

KNIME Erweiterung für Apache Spark ist kompatibel mit

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

ist kompatibel mit

ist kompatibel mit

Cloudera CDH 5/6 enthält nicht Livy, daher bietet KNIME

ist kompatibel mit

ist kompatibel mit

## 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 A-1 Cloudera Manager installieren](#page6)

Maschine, wo Cloudera Manager installiert ist.

2. Nur wenn Cloudera Manager nicht auf das öffentliche Internet zugreifen kann:

Matching .parcel und .sha1 Datei aus  
/opt/cloudera/parcelrepo .

[Download / Kopieren der](#page6)

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. DiLivy 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.

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: <a href="#">Paket</a> / <a href="#">!</a>	RHEL 6: <a href="#">Paket</a> / <a href="#">!</a>	RHEL 5: <a href="#">Paket</a> / <a href="#">!</a>
SICH	SLES 12: <a href="#">Paket</a> / <a href="#">!</a>	SLES 11: <a href="#">Paket</a> / <a href="#">!</a>	

<b>Ubuntu</b>	Ubuntu 16 (Xenial): <a href="#">Paket</a> / <a href="#">!</a>	Ubuntu 14 (Trusty): <a href="#">Paket</a> / <a href="#">!</a>	Ubuntu 12 (Precise): <a href="#">Paket</a> / <a href="#">!</a>
<b>Debian</b>	Debian 8 (Jessie): <a href="#">Paket</a> / <a href="#">!</a>	Debian 7 (Wheezy): <a href="#">Paket</a> / <a href="#">!</a>	

#### Links für CDH 6 herunterladen

<b>RHEL/CentOS</b>	RHEL 7: <a href="#">Paket</a> / <a href="#">!</a>	RHEL 6: <a href="#">Paket</a> / <a href="#">!</a>	RHEL 5: <a href="#">Paket</a> / <a href="#">!</a>
<b>SICH</b>	SLES 12: <a href="#">Paket</a> / <a href="#">!</a>	SLES 11: <a href="#">Paket</a> / <a href="#">!</a>	
<b>Ubuntu</b>	Ubuntu 16 (Xenial): <a href="#">Paket</a> / <a href="#">!</a>	Ubuntu 14 (Trusty): <a href="#">Paket</a> / <a href="#">!</a>	Ubuntu 12 (Precise): <a href="#">Paket</a> / <a href="#">!</a>
<b>Debian</b>	Debian 8 (Jessie): <a href="#">Paket</a> / <a href="#">!</a>	Debian 7 (Wheezy): <a href="#">Paket</a> / <a href="#">!</a>	



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