

# Execution Engine for Apache Hadoop

Last Updated: 2024-12-11

Version: 5.1.0

Experience: Cloud Pak for Data watsonx™

## Description [↗](#)

The Execution Engine for Apache Hadoop service integrates the Watson Studio service with your remote Apache Hadoop cluster. 12.1

Data scientists can use this service for the following tasks.

- Browse remote Hadoop data through connections.
- Cleanse and transform remote Hadoop data with Data Refinery.
- Run Data Refinery and jobs on the Hadoop Spark cluster.
- Run a notebook session on the remote Hadoop system.
- Access Hadoop systems with basic utilities from RStudio® and Jupyter Notebooks.

The Execution Engine for Apache Hadoop service includes the following capabilities:

- Services that establish secure connections between Watson Studio and Hadoop.
- Integration with Hadoop for Data Refinery and Notebooks.
- A high availability configuration to the remote Hadoop system.
- Utilities that connect Watson Studio and Hadoop.

The service requires a service user who has the necessary privileges to submit requests on behalf of the Watson Studio users to WebHDFS, Spark, and YARN. The service generates a secure URL for each Watson Studio cluster that is integrated with the Hadoop cluster.

## Hadoop ecosystem services [↗](#)

Table 1. Services that interact with a Hadoop cluster in Watson Studio

Service	Purpose
WebHDFS	Browse and preview HDFS data.
Jupyter Enterprise Gateway	Submit jobs to JEG on the Hadoop cluster.

Service	Purpose
Livy for Spark	Submit jobs to Spark on the Hadoop cluster.

## Licensing information [↗](#)

This service is included in the following licenses:

- IBM Cloud Pak® for Data Enterprise Edition
- IBM Cloud Pak for Data Standard Edition
- IBM® watsonx.ai™

For more information, see [Licenses and entitlements](#).

## Quick links [↗](#)

- [Install](#): Install the service
- [Use](#): Work with the service
- [What's new](#): See a list of new features
- [Known issues](#): View limitations
- [Administer](#): Manage and maintain the service

## Integrated services [↗](#)

Table 2. Prerequisite services. This service requires the following prerequisite services to be installed.

Service	Capability
<a href="#">Watson Studio</a>	Prepare, analyze, and model data in a collaborative environment with tools for data scientists, developers, and domain experts.

Table 3. Related services. The following related services are often used with this service and provide complementary features, but they are not required.

Service	Capability
<a href="#">Watson Machine Learning</a>	Build, train, and deploy machine learning models with a full range of tools.

Service	Capability
<a href="#">RStudio Server Runtimes</a>	Access the RStudio IDE.