

Article

[Yuri Marx](#) · Oct 19, 2021 2m read

[Open Exchange](#)

Using SQL (Apache Hive) into Hadoop Big Data Repositories

Hi Community,

The InterSystems IRIS has a good connector to do Hadoop using Spark. But the market offers other excellent alternative to Big Data Hadoop access, the Apache Hive. See the differences:

HIVE	SPARK
Hive is a database which stores data in the form of tables like RDBMS databases.	Spark is not a database; it is a data analytics framework which can perform complex data analytics in-memory on large volumes of data sizing up to petabytes.
Data can be extracted from Hive using its own SQL engine called HiveQL. The data can be extracted only by using SQLs.	Spark performs data analytics using Complex SQLs and also uses the MapReduce mechanism. It supports analytics frameworks written in Java, Scala, and Python.
Hive operates on top of Hadoop.	Spark does not have its own dedicated storage. In fact, it extracts data from external distributed data stores like Hive, HBase running on Hadoop, and MongoDB.
Hive is a data warehousing database.	Spark is best for performing complex and faster in-memory data analytics and live data streaming.
Hive is a best suited for those applications performing DWH operations on RDBMS databases that need a scale-out database.	Spark is best suited for applications performing big data analytics requiring a solution faster than the MapReduce mechanism.

Source: <https://dzone.com/articles/comparing-apache-hive-vs-spark>

I created a PEX interoperability service to allows you use Apache Hive inside your InterSystems IRIS apps. To try it follow these steps:

1. Do a git clone to the iris-hive-adapter project:

```
$ git clone https://github.com/yurimarx/iris-hive-adapter.git
```

2. Open the terminal in this directory and run:

```
$ docker-compose build
```

3. Run the IRIS container with your project:

```
$ docker-compose up
```

4. Open the Hive Production of the project (to execute a hello sample):

<http://localhost:52773/csp/irisapp/EnsPortal.ProductionConfig.zen?PRODUCTION=dc.irishiveadapter.HiveProduction>

5. Click Start to run to production

6. Now we will test the App!

7. Run your REST Client app (like Postman) the following URLs and command in the body (using POST verb):

7.1 To create a new table in the Big Data: POST <http://localhost:9980/?Type=DDL>. In the BODY: CREATE TABLE helloworld (message String)

7.2 To insert in the table: POST <http://localhost:9980/?Type=DDL>. In the BODY: INSERT INTO helloworld VALUES ("hello")

7.3 To get the result list from the table: POST <http://localhost:9980/?Type=DML>. In the BODY: SELECT * FROM helloworld (P.S.: Type is DML here)

Now you have 2 options to use Big Data into IRIS: Hive or Spark. Enjoy!!

[#Big Data #Interoperability #InterSystems IRIS #InterSystems IRIS for Health](#)
[Check the related application on InterSystems Open Exchange](#)

Source URL: <https://community.intersystems.com/post/using-sql-apache-hive-hadoop-big-data-repositories>