

Artigo

[Larissa Prussak](#) · Out. 21, 2021 2min de leitura

[Open Exchange](#)

## Usando SQL (Apache Hive) em repositórios de Big Data do Hadoop

Olá comunidade,

O InterSystems IRIS tem um bom conector para fazer Hadoop usando Spark. Mas o mercado oferece outra alternativa excelente para o acesso ao Big Data Hadoop, o Apache Hive. Veja as diferenças:

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.

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

Eu criei um serviço de interoperabilidade PEX para permitir que você use o Apache Hive dentro de seus aplicativos IRIS da InterSystems. Para experimentar, siga estas etapas:

1. Faça um git clone para o projeto iris-hive-adapter:

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

2. Abra o terminal neste diretório e execute:

```
$ docker-compose build
```

3. Execute o contêiner IRIS com seu projeto:

```
$ docker-compose up
```

4. Abra o Hive Production do projeto (para executar um exemplo de olá):

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

5. Clique em Iniciar para ir para a produção

6. Agora vamos testar o App!

7. Execute seu aplicativo REST Client (como Postman) as seguintes URLs e comando no corpo (usando o verbo POST):

7.1 Para criar uma nova tabela no Big Data POST <http://localhost:9980/?Type=DDL>. No CORPO: CREATE TABLE helloworld (message String)

7.2 Para inserir na tabela: POST <http://localhost:9980/?Type=DDL>. No CORPO: INSERT INTO helloworld VALUES ("hello")

7.3 Para obter a lista de resultados da tabela: POST <http://localhost:9980/?Type=DML>. No CORPO: SELECT \* FROM helloworld (P.S.: Type is DML here)

Agora você tem 2 opções para usar Big Data no IRIS: Hive ou Spark. Aproveite!!

[#Big Data #Interoperabilidade #InterSystems IRIS #InterSystems IRIS for Health](#)  
[Confira o aplicativo relacionado no InterSystems Open Exchange](#)

URL de origem: <https://pt.community.intersystems.com/post/usando-sql-apache-hive-em-reposit%C3%B3rios-de-big-data-do-hadoop>