

記事

[Mihoko Iijima](#) · 2020年12月17日 4m read

Apache Spark + Apache Zeppelin + InterSystems IRIS を起動する方法

みなさん、こんにちは。

昨日、Apache Spark、Apache Zeppelin、そして InterSystems IRIS を接続しようとしたときに問題が発生したのですが、有用なガイドが見つからなかったため、自分で書くことにしました。

はじめに

Apache Spark と Apache Zeppelin とは何か、そしてどのように連携するのかを理解しましょう。

Apache Spark はオープンソースのクラスタコンピューティングフレームワークです。暗黙的なデータ並列化と耐障害性を備えるようにクラスタ全体をプログラミングするためのインターフェースを提供しています。そのため、ビッグデータを扱う必要のある場合に非常に役立ちます。

一方の Apache Zeppelin はノートブックです。分析や機械学習に役立つ UI を提供しています。組み合わせて使う場合、IRIS がデータを提供し、提供されたデータを Spark が読み取って、ノートブックでデータを処理する、というように機能します。

注意: 以下の内容は、Windows 10 で行っています。

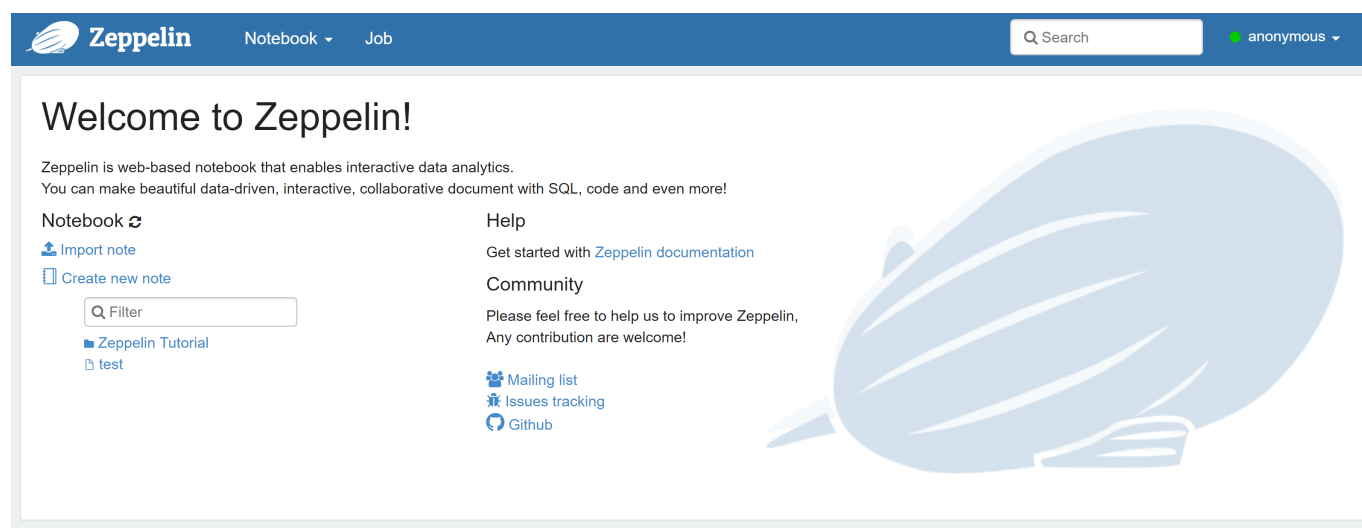
Apache Zeppelin

では、必要なすべてのプログラムをインストールしましょう。

まず、[Apache Zeppelin の公式サイト](#)から Apache Zeppelin をダウンロードします。私は zeppelin-0.8.0-bin-all.tgz を使用しました。このファイルには、Apache Spark、Scala、および Python が含まれます。

ダウンロードファイルを任意のフォルダに解凍します。解凍後、Zeppelin フォルダのルートから `bin/zeppelin.cmd` を呼び出して、Zeppelin を起動します。

「Done, zeppelin server started」が表示されたら、ブラウザで<http://localhost:8080>を開きます。すべてうまくいった場合、「Welcome to Zeppelin!」メッセージが表示されます。



注意: InterSystems IRIS がインストール済みであることを前提としています。まだインストールしていない場合は、次のステップに進む前に IRIS をダウンロードしてインストールしてください。

Apache Spark

ブラウザウィンドウに Zeppelin

ノートブックが開いている状態です。右上の「anonymous」をクリックし、「Interpreter」をクリックします。下にスクロールして「spark」を見つけてください。

「spark」の横に「edit」ボタンがあるので、それをクリックしましょう。

下にスクロールして、intersystems-spark-1.0.0.jar と intersystems-jdbc-3.0.0.jar に依存関係を追加します。

私の環境は InterSystems IRIS を C:\InterSystems\IRIS\

ディレクトリにインストールしているため、追加しなければならないものは以下の場所にあります。

私の環境でのファイルは以下の通りです。

そして保存します。

動作確認

動作確認してみましょう。新しいノートを作成し、段落に次のコードを貼り付けます。

```
var dataframe=spark.read.format("com.intersystems.spark").option("url", "IRIS://localhost:51773/NAMESPACE").option("user", "UserLogin").option("password", "UserPassword").option("dbtable", "Sample.Person").load()
```

```
// dbtable - name of your table
```

URL - IRIS アドレスを 次の書式で指定します。IRIS://ipAddress:superserverPort/namespace

プロトコル「IRIS」 は、Java 共有メモリ接続を提供する TCP/IP 経由の JDBC 接続です。

ipAddress - InterSystems IRIS インスタンスの IP アドレス。ローカルで接続している場合は、localhost の代わりに 127.0.0.1 を使用してください。

superserverPort - IRIS インスタンスのスーパーサーバーのポート番号。Web

サーバーのポート番号とは異なります。スーパーサーバーのポート番号を見つけるには、管理ポータルを開き、画面上部にある「概要」のリンクをクリックして表示される画面の「Superserver Port」を確認してください。

namespace - InterSystems IRIS インスタンスの既存のネームスペースを指定します。このデモでは、USER ネームスペースに接続しています。

この段落を実行しましょう。うまくいけば、「FINISHED」が表示されます。

私のノートブックでの実行例です。

まとめ

この記事の内容をまとめると、Apache Spark、Apache Zeppelin、および InterSystems IRIS

がどのように連携できるかがわかりました。次の記事では、データ分析についてお話しします。

リンク

- [Apache Spark の公式サイト](#)
- [Apache Spark のドキュメント](#)
- [IRIS プロトコル](#)
- [InterSystems Spark コネクタの使用](#)

[#AI](#) [#ビッグデータ](#) [#ベストプラクティス](#) [#初心者](#) [#機械学習](#) [#InterSystems IRIS](#)

ソースURL:<https://jp.community.intersystems.com/post/apache-spark-apache-zeppelin-intersystems-iris-%E3%82%92%E8%B5%B7%E5%8B%95%E3%81%99%E3%82%8B%E6%96%B9%E6%B3%95>