

特典ファイルの環境構築について

ここでは、本書の特典ファイル使うための環境セットアップについて説明をします。本書の特典ファイルを実行するための環境構築は、特にWindowsでは難しいため、Mac OSでの環境構築をオススメします。

Windowsの環境構築については、参考として編集部所有のWindows10（64bit）で動作確認が取れた内容について記載いたしますが、環境の違いによりエラーが出るが多いため、あくまで参考としてご覧くださいませ。このファイル記載の内容は、本書サポートの対応外となります。

- [1. Cythonを使うための準備](#)
 - [・ macOSの場合](#)
 - [・ Windows10/64bit の場合](#)
- [2. PySparkのインストール](#)
 - [2-1. Java（JDK）のインストール](#)
 - [・ macOSの場合](#)
 - [・ Windows10/64bitの場合](#)
 - [2-2. Java（JDK）のPATHの登録](#)
 - [・ macOSの場合](#)
 - [・ Windowsの場合](#)
 - [2-3. Javaが正しくインストールされたかの確認](#)
 - [2-4. Apache Sparkのインストール](#)
 - [2-4-1. Apache Sparkのダウンロード](#)
 - [2-4-2. ダウンロードしたファイルを解凍する](#)
 - [・ macOSの場合](#)
 - [・ Windowsの場合](#)
 - [2-4-3. 解凍したフォルダをコピーする](#)
 - [2-4-4. winutilのインストール（Windowsのみ）](#)
 - [2-5. Apache SparkとPySparkのPATHの登録](#)
 - [・ macOSの場合](#)
 - [・ Windowsの場合](#)
 - [2-6. findsparkのインストール](#)
 - [2-7. Sparkの起動](#)

1. Cythonを使うための準備

・ macOSの場合

Cythonは、Anacondaに含まれていますので、macOSでは特に追加でのインストールは不要です。

・ Windows10/64bit の場合

WindowsではCythonを動かすための準備が必要です。

Windowsの場合、Cythonを動かすためのコンパイラを別途インストールする必要があります。本書ではVisual Studio 2017を使う方法をご紹介します。

まず、Visual Studio 2017の公式サイト（<https://docs.microsoft.com/ja-jp/visualstudio/install/install-visual-studio?view=vs-2017>）（<https://docs.microsoft.com/ja-jp/visualstudio/install/install-visual-studio?view=vs-2017%EF%BC%89>）にアクセスします。

Microsoft Office Windows Surface Xbox お買い物商品 サポート

すべて Microsoft 製品

Docs / Visual Studio / インストール / Visual Studio のインストール

フィードバック 編集 共有 英語で読む

サインイン

① テキストの上でマウス ポインターを動かすと、ポップアップ ウィンドウに英語のテキストが表示されます。

有効にする

Visual Studio 2017 のインストール

2018/05/07 • 共同作成者

ここでは、Visual Studio の新しいインストール方法について説明します。最新バージョンでは、必要な機能だけを簡単に選んでインストールできるようになりました。また、Visual Studio の最小フットプリントも減らしているため、以前よりもシステムへの影響が少なくなり、より迅速にインストールできます。

① 注意

このトピックは、Windows 上の Visual Studio に適用されます。Visual Studio for Mac については、[Visual Studio for Mac のインストール](#)に関するページを参照してください。

このバージョンの他の新機能については、[リリースノート](#)を参照してください。

インストールの準備ができたら、各ステップを順に実行していきます。

この記事の内容

- 手順 1 - コンピューターで Visual Studio の準備ができていないことを確認する
- 手順 2 - Visual Studio をダウンロードする
- 手順 3 - Visual Studio インストーラーをインストールする
- 手順 4 - ワークロードを選択する
- 手順 5 - 個々のコンポーネントを選択する (省略可能)
- 手順 6 - 言語/パックをインストールする (省略可能)

つづいて、「手順 2 - Visual Studio をダウンロードする」にある「Visual Studio 2017をダウンロードする」ボタンをクリックします。

手順 2 - Visual Studio をダウンロードする

次に、Visual Studio ブートストラップ ファイルをダウンロードします。これを行うには、以下のボタンをクリックし、必要な Visual Studio 2017 のエディションを選択して、**[保存]**、**[フォルダーを開く]**の順にクリックします。

[Visual Studio 2017 をダウンロードする](#)



Visual Studio ブートストラップ ファイルをダウンロードして、適切な Visual Studio のエディションを選択する方法については、[こちらのビデオ](#)をご覧ください。

手順 3 - Visual Studio インストーラーをインストールする

次に、ブートストラップ ファイルを実行して、Visual Studio インストーラーをインストールします。この新しい軽量インストーラーには、Visual Studio 2017 のインストールとカスタマイズの両方に必要なすべてが含まれています。

ダウンロードのページに移動するので、「コミュニティ」の「無償ダウンロード」をクリックしてダウンロードします。

Microsoft | Visual Studio 製品 ダウンロード 表示を増やす Visual Studio 無償版

すべて Microsoft 製品

ダウンロード

Windows macOS

Visual Studio 2017

Android、iOS、Windows、Web、クラウド向けのフル機能の統合開発環境 (IDE)

コミュニティ	Professional	エンタープライズ
学生、オープンソース貢献者、個人向けの無料で強力な IDE	小規模なチームに最適な Professional IDE	あらゆる規模のチーム向けのスケーラブルなエンドツーエンドソリューション

バージョン15.9 リリースノート

エディションの比較

オフラインでインストールする方法

[無償ダウンロード](#) [無料試用版](#) [無料試用版](#)

Visual Studio コード

ニーズに適した無料で高速のオープンソース コード エディター

リリースノート

[無償ダウンロード](#)

Visual Studio Code をダウンロードおよび使用することにより、ライセンス承認およびプライバシーに関する声明に同意したものと見なされます。

製品発表

Visual Studio 2019 Preview

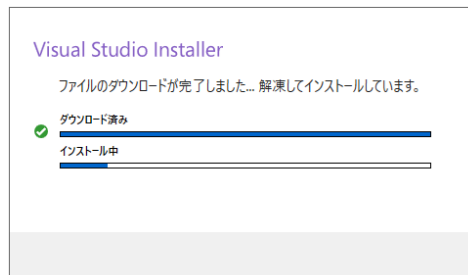
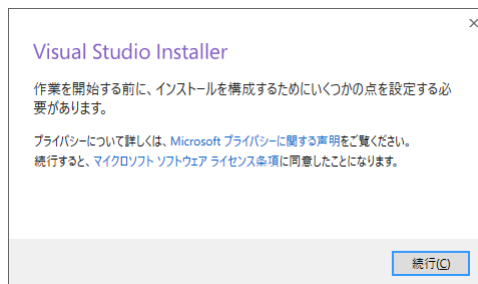
Visual Studio の未来をいち早くご体験ください

[今すぐ試す](#)

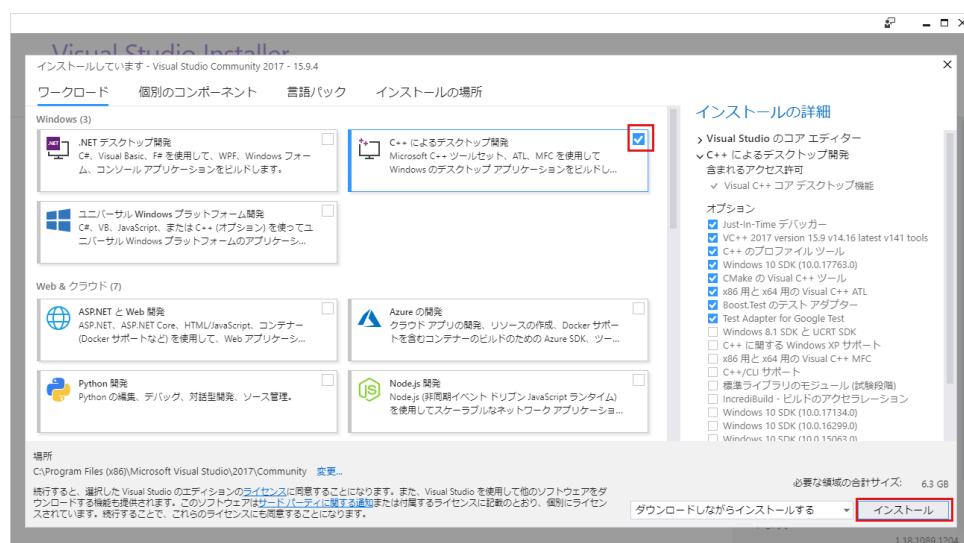
Feedback

Search all downloads

ダウンロードしたファイルをダブルクリックして起動し、表示にそってインストールを進めます。



インストールの途中で以下のような画面が開くので、ここで「C++によるデスクトップ開発」にチェックを入れ、画面右下の「インストール」をクリックしてインストールします。



インストールが終了したら、Cythonを使用する準備は完了です。

インストール完了後、Visual Studio Communityの起動画面やサインインの画面が表示されますが、そのまま閉じて問題ありません。

2.PySparkのインストール

次は、PySparkを使うための環境準備方法を紹介します。なお、Jupyter Notebookは使える状態であることを前提とします。以下の順番で進めていきます。

- Java (JDK) のインストール
- Apache Sparkのインストール
- 動作確認

2-1.Java (JDK) のインストール

1.JDKのダウンロードは以下のページから行います。

<https://www.oracle.com/technetwork/java/javase/downloads/index.html>
(<https://www.oracle.com/technetwork/java/javase/downloads/index.html>)

アクセスすると以下のような画面が表示されるので、「Java Platform (JDK) XX」をクリックします。「XX」のところは番号が変わることがありますので最新版をクリックしてください。



2.次の画面ではOSに合ったインストーラーをダウンロードします。「Accept License Agreement」にチェックを入れて、インストーラーをクリックします。どの形式でも構いませんが、以下ではmacOS版として「dmg」形式、Windows版として「exe」形式をダウンロードした前提で話を進めます。

See also:

- [Java Developer Newsletter](#): From your Oracle account, select **Subscriptions**, expand **Technology**, and subscribe to **Java**.
- [Java Developer Day hands-on workshops \(free\) and other events](#)
- [Java Magazine](#)

[JDK 11.0.1 checksum](#)

Java SE Development Kit 11.0.1

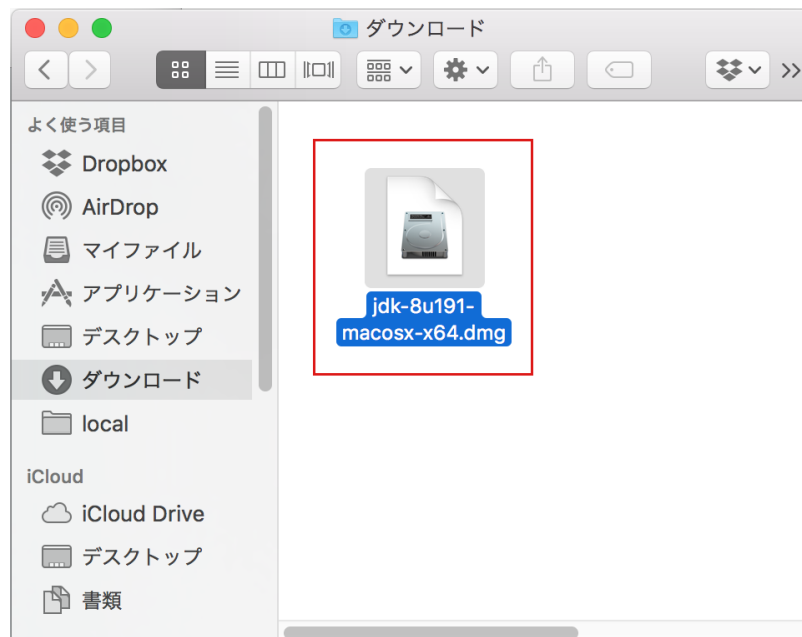
You must accept the [Oracle Technology Network License Agreement](#) for Oracle Java SE to download this software.

☐ Accept License Agreement ☒ Decline License Agreement

Product / File Description	File Size	Download
Linux	147.4 MB	jdk-11.0.1_linux-x64_bin.deb
Linux	154.09 MB	jdk-11.0.1_linux-x64_bin.rpm
Linux	171.43 MB	jdk-11.0.1_linux-x64_bin.tar.gz
macOS	166.2 MB	jdk-11.0.1_osx-x64_bin.dmg
macOS	166.55 MB	jdk-11.0.1_osx-x64_bin.tar.gz
Solaris SPARC	186.8 MB	jdk-11.0.1_solaris-sparcv9_bin.tar.gz
Windows	150.98 MB	jdk-11.0.1_windows-x64_bin.exe
Windows	170.99 MB	jdk-11.0.1_windows-x64_bin.zip

・ macOSの場合

ダウンロードしたファイルをダブルクリックしてインストールを開始します。画面に沿ってインストールを完了します。一般的な環境では、JDKは、/Library/Java/JavaVirtualMachines/ にインストールされます。



・ Windows10/64bitの場合

ダウンロードしたファイルをダブルクリックしてインストールを開始します。画面に沿ってインストールを完了します。途中でJDKをインストールするフォルダが表示されるので、これをメモしておいてください。



2-2.Java (JDK) のPATHの登録

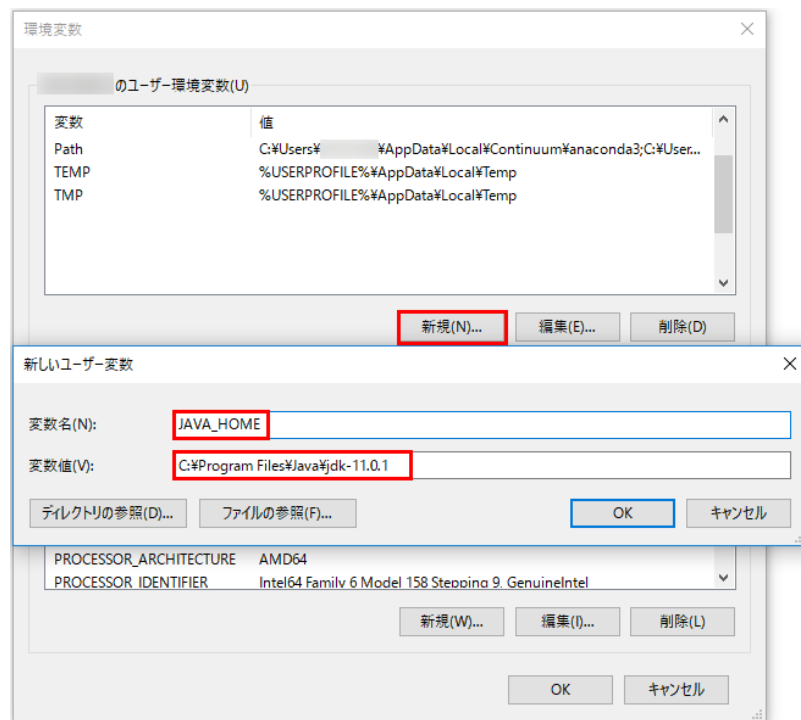
・ macOSの場合

macOSでは、JDKをインストールするとPATHの登録も行われます。

・ Windows10/64bitの場合

インストールが完了したら、PATHの登録を行います。「Anacondaをインストールする」の手順「4」「5」に沿って、「環境変数」の画面を表示します。

「環境変数」の画面上部の「<ユーザー名>のユーザー環境変数」の「新規」をクリックし、「変数名: JAVA_HOME」、「変数値」には、「1.Java (JDK) のインストール」の手順「2」でメモしたJDKをインストールしたフォルダを登録して、[OK] をクリックします。



2-3.Javaが正しくインストールされたかの確認

Windowsではコマンドプロンプト、macOSではターミナルを起動して、以下のように入力し、出力にJavaのバージョンが表示されればOKです。

```
$ java -version
```

2-4.Apache Sparkのインストール

次に、Sparkのインストールを実行します。

2-4-1.Apache Sparkのダウンロード

まずは以下のURLサイトにアクセスします。

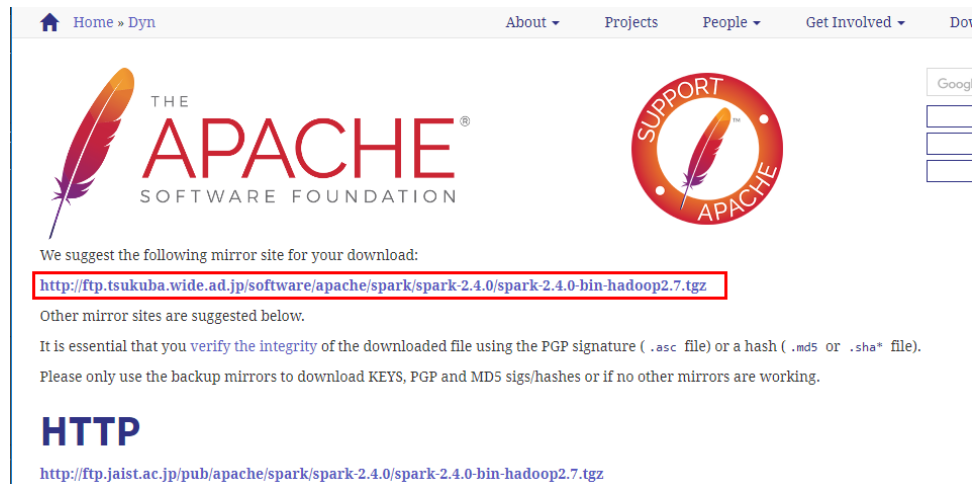
<http://spark.apache.org/downloads.html> (<http://spark.apache.org/downloads.html>)

「Download Apache Spark」の「1.Choose a Spark release」で、バージョンを選びます。最新バージョンを選んでください。

次に「2.Choose a package type」を選びます。「Pre-built for Apache Hadoop 2.7 and later」を選んでください。

そうすると「3.Download Spark」のところに、パッケージファイルが表示されますので、これをクリックします。

すると次の画面で、おすすめのダウンロードURLが表示されますので、これをクリックして、ファイルをダウンロードします。



2-4-2.ダウンロードしたファイルを解凍する

・ macOSの場合

macOSでは、ターミナルでダウンロードしたファイルの場所まで移動し、まずchmodコマンドで実行権限を付与します。

```
$ chmod u+x <ダウンロードしたファイル名>
```

そして、以下のコマンドでファイルを解凍します。

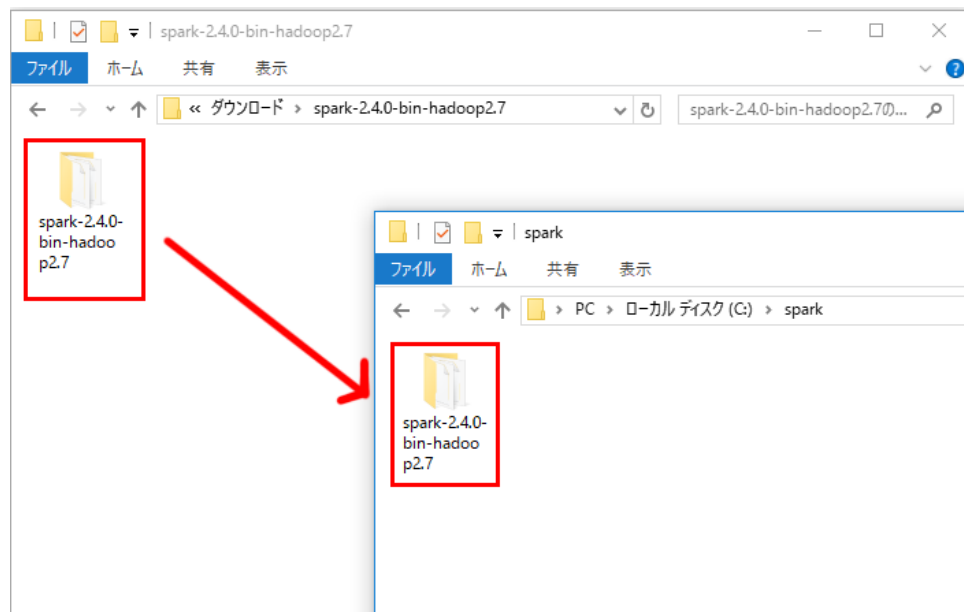
```
$ tar -xf <ダウンロードしたファイル名>
```

・ Windows10/64bitの場合

ダウンロードしたファイルを解凍します。Windowsでは「tgz」形式のファイルは、「Lhaca デラックス版」などの解凍用ツールを使用すると解凍できます。

2-4-3.解凍したフォルダをコピーする

続いて、ドライブの直下などに「spark」フォルダを作り、そこに上記で解凍したフォルダの中身をコピーしてください。

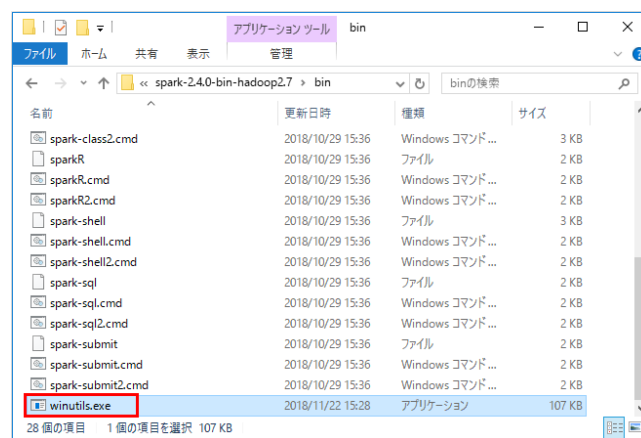
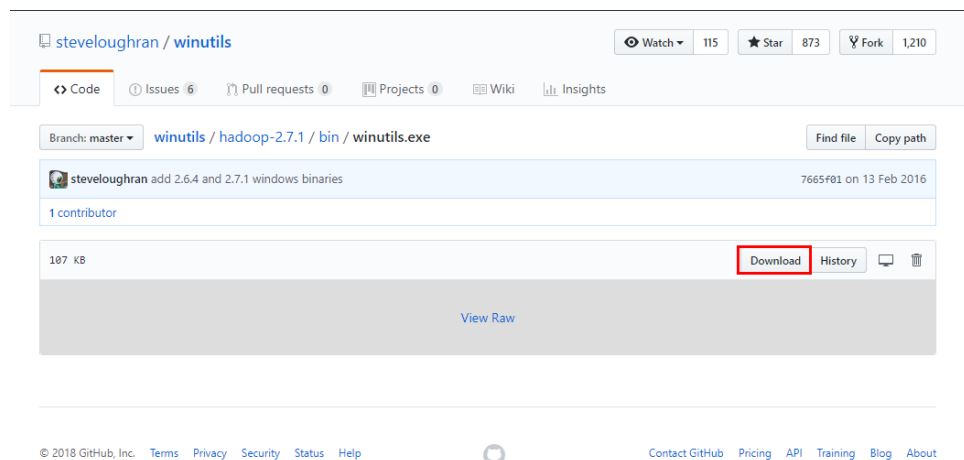


2-4-4.winutilのインストール (Windowsのみ)

ここで、Windowsではwinutilをインストールします。macOSでは必要ありません。以下のサイトにアクセスしてください。

<https://github.com/steveloughran/winutils/blob/master/hadoop-2.7.1/bin/winutils.exe>
(<https://github.com/steveloughran/winutils/blob/master/hadoop-2.7.1/bin/winutils.exe>)

画面右側にある「Download」をクリックして、ファイルを保存します。保存先は、さきほどドライブの直下などに作成した「spark」フォルダの「spark-X.X.X-bin-hadoopX.X」→「bin」の下にします。



2-5.Apache SparkとPySparkのPATHの登録

・ macOSの場合

ターミナルを起動して、以下のように入力して、.bash_profileファイルの編集画面を開きます。

```
#vi ~/.bash_profile
```

.bash_profilecファイルに以下のように追加します。

```
export HADOOP_HOME=%spark%<sparkのフォルダ名>
```

```
export SPARK_HOME=%spark%<sparkのフォルダ名>
```

```
export PYSARK_PYTHON=<Anacondaのインストールフォルダ>
```

```
export PYSARK_DRIVER_PYTHON="jupyter"
```

```
export PYSARK_DRIVER_PYTHON_OPTS="notebook"
```

```
export PATH=$SPARK_HOME/bin:$PATH
```

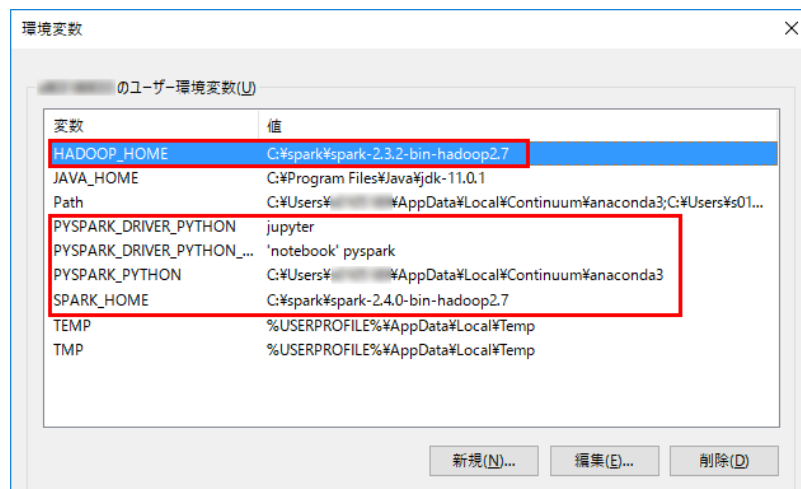
追加できたら、「ZZ」コマンドで保存してファイルを閉じます。

・ Windowsの場合

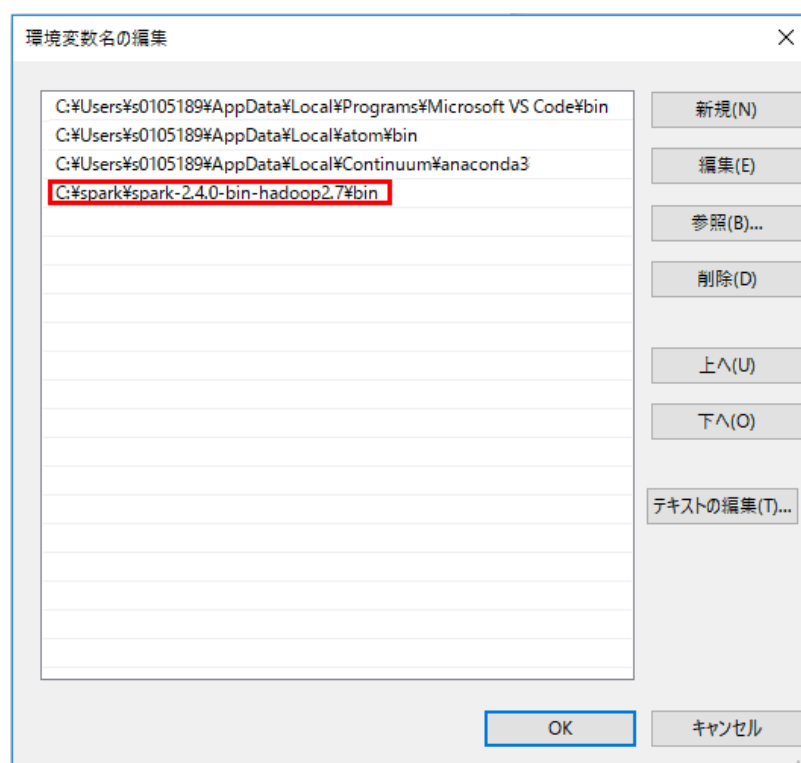
「2.Java（JDK）のPATHの登録」の手順を参考に、「環境変数」の画面を表示します。

「<ユーザー名>のユーザー環境変数」の「新規」をクリックして、それぞれ以下のように環境変数を設定します。

- ・変数名「HADOOP_HOME」、変数値「<ドライブ名>:\spark\<sparkのフォルダ名>」
- ・変数名「SPARK_HOME」、変数値「<ドライブ名>:\spark\<sparkのフォルダ名>」
- ・変数名「PYSPARK_PYTHON」、変数値「<ドライブ名>:<Anacondaのインストールフォルダ名>」
- ・変数名「PYSPARK_DRIVER_PYTHON」、変数値「jupyter」
- ・変数名「PYSPARK_DRIVER_PYTHON_OPTS」、変数値「'notebook' pyspark」



続いて、変数「Path」を選択し「編集」をクリックします。続いて「新規」をクリックして「<ドライブ名>:\spark\<sparkのフォルダ名>\bin」という値を追加します。



2-6.findsparkのインストール

続いてfindsparkのインストールを行います。macOSはターミナルを起動し、Windowsはコマンドプロンプトを起動して、以下のように入力します。

```
$ python -m pip install findspark
```

2-7.Sparkの起動

ここまで終了したら一度パソコンを再起動します。そして、ターミナルまたはコマンドプロンプトを起動し、以下のように入力します。

```
$ spark-shell
```

そうするとsparkが実行され、以下のような画面がでます（画面とはバージョンが異なる場合があります）。もしWindowsで「setting default log level to "WARN"」と表示されるようなら、管理者としてコマンドプロンプトを実行してください。

```
Welcome to
  ____  ____  ____  ____  ____  ____  ____  ____  ____  ____
 /_ _/_/ _/_/ _/_/ _/_/ _/_/ _/_/ _/_/ _/_/ _/_/ _/_/ _/_/
version 2.4.0

Using Scala version 2.11.12 (Java HotSpot(TM) Client VM, Java 1.8.0_71)
Type in expressions to have them evaluated.
Type :help for more information.

scala>
```

無事にSparkが起動したら、その後Jupyter Notebookを起動します。そして、右上から「New→Python3」をクリックしてノートブックを作成し、以下のように入力します。

```
In [1]: import findspark
findspark.init()
import pyspark
sc = pyspark.SparkContext(appName="myAppName")
```

scを実行して、以下のような結果が表示されたらOKです。これで準備は整いました。

```
In [7]: sc
```

```
Out[7]: SparkContext
```

[Spark UI \(http://localhost:4041\)](http://localhost:4041)

Version

v2.4.0

Master

local[*]

AppName

myAppName

出力された文字の「Spark UI」はリンクになっており、これをクリックすれば、以下のような画面をみることができます。ここではパフォーマンスなどをチェックすることができます。

Spark Jobs (?)

User: mynavi

Total Uptime: 1.3 min

Scheduling Mode: FIFO

▼ [Event Timeline](#)

☐ Enable zooming

