



# Spark

17/08/2017

Everton Osnei Cesario

---

# Links para Download

- ◉ <https://www.virtualbox.org/wiki/Downloads>
- ◉ <https://anaconda.org/anaconda/python>
- ◉ <https://mobaxterm.mobatek.net/download.html>
- ◉ <https://spark.apache.org/downloads.html>
- ◉ <https://github.com/bmatzelle/gow/downloads>
- ◉ <https://github.com/steveloughran/winutils/tree/master/hadoop-2.7.1/bin>
- ◉ <https://www.win-rar.com/download.html?&L=0>
- ◉ <https://notepad-plus-plus.org/download/v7.5.4.html>
- ◉ <https://www.microsoft.com/pt-br/download/details.aspx?id=353>
- ◉ <http://www.oracle.com/technetwork/pt/java/javase/downloads/jdk8-downloads-2133151.html>

# Instalações e configurações

- Instalar JDK

- C:\Jdk

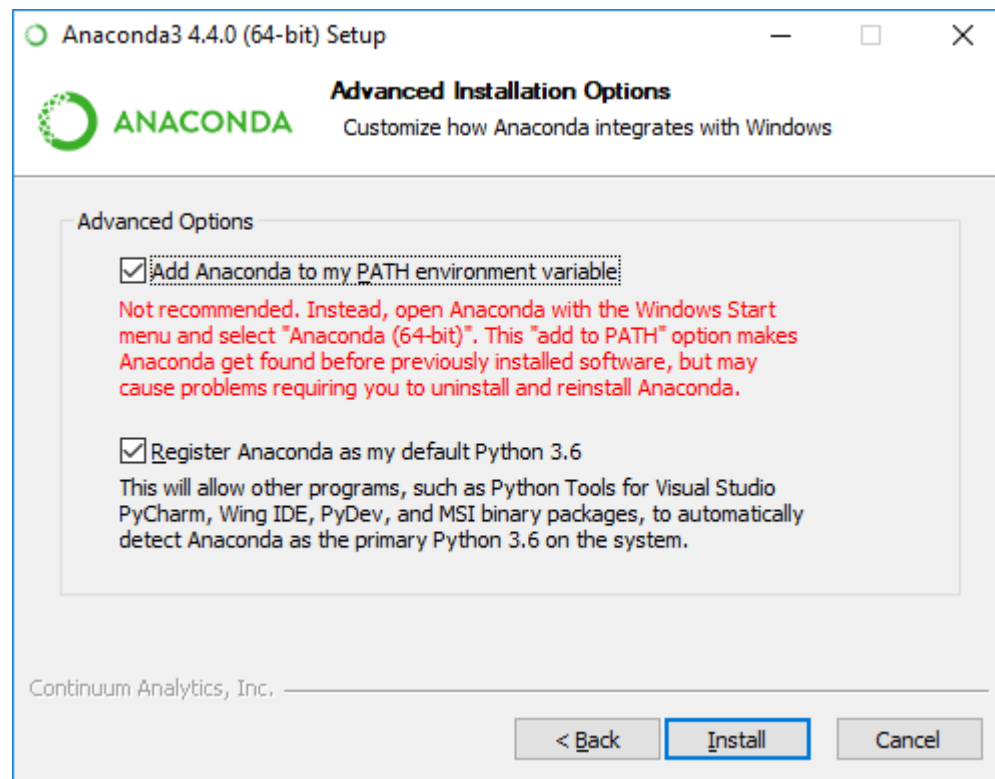
- C:\Jre

# Variáveis de ambiente - Java

- ◉ JAVA\_HOME
  - ◉ C:\jdk
- ◉ JRE\_HOME
  - ◉ C:\jre
- ◉ HADOOP\_HOME
  - ◉ c:\hadoop
- ◉ Adicionar variáveis no PATH
  - ◉ Para adicionar outras variáveis, colocar ;
  - ◉ %JAVA\_HOME%\bin
  - ◉ %JRE\_HOME%\bin
- ◉ Reinicializar a maquina

# Anaconda

- Executar instalador
  - Na tela advanced Options, selecionar opção de Path



- Reiniciar a maquina

# Anaconda

- Menu do Windows – Anaconda Navigator
- Criar pasta c:\Projetos
- Prompt de comando
  - C:\Projetos>jupyter notebook

# Spark

- Criar pasta c:\Spark
- Copiar spark-2.2.1-bin-hadoop2.7.tgz
- Extrair os arquivos na raiz da pasta
- Criar pasta c:\Hadoop\bin
- Jogar o arquivo winutils.exe na bin
- Criar variáveis de ambiente
- SPARK\_HOME
  - C:\Spark
- Adicionar SPARK\_HOME e HADOOP\_HOME no path
  - %SPARK\_HOME%\bin
  - %HADOOP\_HOME%\bin

# Configuração PySpark

- Configuração do Jupyter para abrir o PySpark
- Adicionar as variáveis de ambiente
  - PYSPARK\_DRIVER\_PYTHON
    - jupyter
  - PYSPARK\_DRIVER\_PYTHON\_OPTS
    - notebook
- Reiniciar a maquina



# Spark - Hive

- Criar a pasta c:\tmp\hive
- Criar a pasta c:\direct
- Executar o instalador directx....exe direcionando para a pasta direct
- Dentro da pasta direct, executar o DXSETUP.exe
- Prompt de comando:
  - c:\Hadoop\bin
  - winutils.exe chmod -R 777 C:\tmp\hive
- Para testar:
  - c:\Projetos > jupyter notebook

# Spark – Contexto Spark

- Prompt de comando:
  - c:\Projetos> pyspark
  - Sc
- Criar a pasta C:\Spark\projetos
- Prompt de comando:
  - c:\spark\projetos> pyspark
  - Sc