

小木瓜瓜瓜

Win10下快速搭建spark

在win10下搭建spark，需要安装java的jdk，scala，spark，hadoop。

一、安装配置jdk

下载 jdk版本：jdk-8u151-windows-x64.exe

增加2个环境变量：

JAVA_HOME E:\Java\jdk1.8.0_151 （注意，这里把jdk安装的目录，不是C盘的默认目录，路径中不能有空格）

CLASSPATH %JAVA_HOME%\lib;%JAVA_HOME%\lib\tools.jar

在系统环境变量Path中，在原来基础上加：%JAVA_HOME%\bin

二、安装配置scala

访问官方地址<http://www.scala-lang.org/download/2.11.8.html>

下载：scala-2.11.8.msi

在系统环境变量 Path 中，增加：C:\Program Files (x86)\scala\bin

三、安装配置spark

1、下载spark

访问官方地址<http://spark.apache.org/downloads.html>

下载文件：spark-2.2.0-bin-hadoop2.7.tgz



公告

昵称：小木瓜瓜瓜
园龄：1年1个月
粉丝：1
关注：0
[+加关注](#)

2019年1月						
日	一	二	三	四	五	六
30	31	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31	1	2
3	4	5	6	7	8	9

搜索

常用链接

[我的随笔](#)
[我的评论](#)
[我的参与](#)
[最新评论](#)
[我的标签](#)

随笔分类

C语言(1)
HTML(9)
Javascript(31)
jQuery(1)
Linux(13)
MySQL(12)
PostgreSQL(1)
Python(25)
Redis(2)
SQL Server on Linux(3)
SQL Server备份与恢复(10)
SQL Server高可用、迁移(11)
SQL Server管理(42)
SQL Server开发与商业智能(43)
SQL Server论坛问题(55)
SQL Server优化(43)
SQL 编程你也行(22)
杂项(14)

随笔档案

2018年7月 (1)

2、解压tgz文件

我把文件解压到目录：D:\spark-2.2.0-bin-hadoop2.7

在这个目录下，有 bin等文件夹。

3、配置

添加系统环境变量：

SPARK_HOME D:\spark-2.2.0-bin-hadoop2.7



在系统环境变量Path增加：%SPARK_HOME%\bin

四、安装配置hadoop

1、下载hadoop

访问官方<http://hadoop.apache.org/releases.html>

可以下载 2.7.6版的binary文件



Top

Wiki

About

Releases

Download

Release Notes

Release Versioning

Mailing Lists

Issue Tracking

Who We Are?

Who Uses Hadoop?

Buy Stuff

Sponsorship

Thanks

Privacy Policy

Bylaws

Committee criteria

License

Documentation

Related Projects

Search with A

Last Publish

Apache Hadoop Releases

Download

Hadoop is released as source code tarballs with corresponding binary tarballs for convenience. The downloads are distributed via mirror site checked for tampering using GPG or SHA-256.

Version	Release Date	Tarball	GPG	SHA-256
3.0.3	31 May, 2018	source	signature	checksum file
2.8.4	15 May, 2018	source	signature	checksum file
2.9.1	3 May, 2018	source	signature	checksum file
3.1.0	6 Apr, 2018	source	signature	checksum file
3.0.2	21 April, 2018	source	signature	checksum file
2.9.0	17 November, 2017	source	signature	checksum file
2.8.3	12 December, 2017	source	signature	checksum file
2.7.6	16 April, 2018	source	signature	checksum file
2.5.3	08 October, 2016	source	signature	checksum file

To verify Hadoop releases using GPG:

但是，我在安装的时候，直接百度，找了 hadoop2.7.1的压缩文件。

在bin目录中，包含了：hadoop.dll，winutils.exe,这2个文件就够用了。

然后解压为：D:\hadoop2.7.1

2、配置

增加系统环境变量：

HADOOP_HOME D:\hadoop2.7.1

在系统环境变量Path增加：%HADOOP_HOME%\bin

3、下载winutils

下载路径：<https://github.com/steveloughran/winutils>

五、配置pyspark

在搭建spark环境之前，安装了Anaconda，包含了python，为了使用pyspark：

- 2018年1月 (1)
- 2017年12月 (3)
- 2017年5月 (1)
- 2017年4月 (1)
- 2017年3月 (24)
- 2017年2月 (7)
- 2017年1月 (32)
- 2016年12月 (26)
- 2016年11月 (6)
- 2016年10月 (9)
- 2016年9月 (2)
- 2016年8月 (3)
- 2016年7月 (5)
- 2016年6月 (2)
- 2016年5月 (9)
- 2016年4月 (5)
- 2016年3月 (1)
- 2016年1月 (34)
- 2015年12月 (2)
- 2015年11月 (2)
- 2015年10月 (1)
- 2015年4月 (2)
- 2014年12月 (1)
- 2014年9月 (2)
- 2014年7月 (1)
- 2014年5月 (3)
- 2014年4月 (1)
- 2014年3月 (20)
- 2014年2月 (3)
- 2014年1月 (11)
- 2013年12月 (4)
- 2013年11月 (13)
- 2013年10月 (18)
- 2013年9月 (6)
- 2013年8月 (5)
- 2013年5月 (2)
- 2013年4月 (1)
- 2013年2月 (2)
- 2013年1月 (5)
- 2012年11月 (2)
- 2012年10月 (6)
- 2012年9月 (6)
- 2012年8月 (18)
- 2012年7月 (9)
- 2012年6月 (3)
- 2012年3月 (1)
- 2012年2月 (1)
- 2011年12月 (1)
- 2011年11月 (1)
- 2011年10月 (5)
- 2011年8月 (8)

阅读排行榜

- 1. Win10下快速搭建spark(673)
- 2. python做量化交易干货分享(315)
- 3. Linux开机自启动目录(182)
- 4. Linux切换用户(153)
- 5. python合并多个excel表格：openpyxl模块 (三) (131)

1、把 D:\spark-2.2.0-bin-hadoop2.7\python 复制到 E:\Anaconda3\Lib\site-packages 路径下。

2、通过pip install py4j 安装 py4j。

3、修改权限 `winutils.exe chmod 777 D:\tmp\Hive`，在运行命令前先创建目录 `D:\tmp\Hive`。

4、配置

增加系统环境变

```
量: PYTHONPATH %SPARK_HOME%\python\lib\py4j;%SPARK_HOME%\python\lib
\pyspark;E:\Anaconda3;
```

在系统环境变量Path，增加：E:\Anaconda3

六、验证

启动cmd, 输入: pyspark

```

C:\管理員: C:\Windows\system32\cmd.exe - pyspark
Microsoft Windows [版本 10.0.16299.431]
(c) 2017 Microsoft Corporation. 保留所有权利。

C:\Users\Administrator>pyspark
Python 3.6.1 |Anaconda 4.4.0 (64-bit)| (default, May 11 2017, 13:25:24) [MSC v.1900 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
18/07/09 22:05:24 WARN ObjectStore: Failed to get database global_temp, returning NoSuchObjectException
Welcome to


  ____      __
 / ___ |__ /  __/
/ /___/ _// _//
/____/_//___/

version 2.2.0

Using Python version 3.6.1 (default, May 11 2017 13:25:24)
SparkSession available as 'spark'.
>>>

```

或者输入: spark-shell

```
C:\Windows\system32\cmd.exe - spark-shell
C:\Users\Administrator>spark-shell
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN"
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
18/07/09 22:11:01 WARN General: Plugin (Bundle) "org.datanucleus" is already registered. Ensure you dont have multiple JAR versions of the
me plugin in the classpath. The URL "file:/D:/spark-2.2.0-bin-hadoop2.7/jars/datanucleus-core-3.2.10.jar" is already registered, and you are
trying to register an identical plugin located at URL "file:/D:/spark-2.2.0-bin-hadoop2.7/bin/./jars/datanucleus-core-3.2.10.jar."
18/07/09 22:11:01 WARN General: Plugin (Bundle) "org.datanucleus.store.rdbms" is already registered. Ensure you dont have multiple JAR versi
as of the same plugin in the classpath. The URL "file:/D:/spark-2.2.0-bin-hadoop2.7/bin/./jars/datanucleus-rdbms-3.2.9.jar" is already regi
tged, and you are trying to register an identical plugin located at URL "file:/D:/spark-2.2.0-bin-hadoop2.7/jars/datanucleus-rdbms-3.2.9.jar."
18/07/09 22:11:01 WARN General: Plugin (Bundle) "org.datanucleus.api.jdo" is already registered. Ensure you dont have multiple JAR versions
f the same plugin in the classpath. The URL "file:/D:/spark-2.2.0-bin-hadoop2.7/jars/datanucleus-api-jdo-3.2.6.jar" is already registered, a
d you are trying to register an identical plugin located at URL "file:/D:/spark-2.2.0-bin-hadoop2.7/bin/./jars/datanucleus-api-jdo-3.2.6.jar."
18/07/09 22:11:03 WARN ObjectStore: Failed to get database global_temp, returning NoSuchObjectException
Spark context Web UI available at http://192.168.5.1:4040
Spark context available as 'sc' (master = local[*], app id = local-1531145460237).
Spark session available as 'spark'.
Welcome to
 version 2.2.0

Using Scala version 2.11.8 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_151)
Type in expressions to have them evaluated.
Type help for more information.

scala>
```

分类: Python

[好文要顶](#)
[关注我](#)
[收藏该文](#)



[+加关注](#)

« 上一篇: [快速优化SQL的5个步骤](#)

posted @ 2018-07-09 22:12 小木瓜瓜瓜 阅读(673) 评论(0) 编辑 收藏

[刷新评论](#) [刷新页面](#) [返回顶部](#)

注册用户登录后才能发表评论，请 [登录](#) 或 [注册](#)，[访问网站首页](#)。

【推荐】超50万VC++源码: 大型组态工控、电力仿真CAD与GIS源码库！

【活动】华为云12.12会员节 云产品1折起 满额送Mate20 点击抢购



相关博文:

- win10下搭建QTP测试环境
- win10下Spark的环境搭建
- WIN10下OpenGL开发环境的搭建
- 彻底去除Win10“快速访问”
- Win10下使用WampServer搭建PHP开发环境



最新新闻:

- 全世界都在指责中国污染环境，他们做出了反击
 - 全球专利实力榜：三星第二 华为位居第七
 - 今日头条意向收购锤子科技 三方同时洽购
 - 太阳系的“天涯海角”，我飞了4730天跨越65亿千米来看你
 - 2018年新开15条铁路盘点：看哪条经过你家？
- » [更多新闻...](#)