

集群搭建

JAVA 21对集群支持不够，需要卸载当前JAVA 21 版本，并安装JAVA 11版本

Java SE subscribers have more choices

Also available for development, personal use, and to run other licensed Oracle products.

[GraalVM for JDK 17](#) has moved and is now available through [My Oracle Support \(MOS\)](#).

[Java 17](#) [Java 11](#) [Java 8](#) [Java 8 Enterprise Performance Pack](#)

Java SE Development Kit 11.0.28

Java SE subscribers will receive JDK 11 updates until at least **January 2032**.

These downloads can be used for development, personal use, or to run Oracle licensed products. Use for other purposes, including production or commercial use, requires a [Java Universal Subscription](#) or another Oracle license.

Commercial license and support are available for a low cost with [Java SE Universal Subscription](#).

JDK 11 software is licensed under the [Oracle Technology Network License Agreement for Oracle Java SE](#).

JDK 11.0.28 [checksums](#) and [OL 9 GPG Keys](#) for RPMs

[Linux](#) [macOS](#) [Solaris](#) [Windows](#)

| Product/file description | File size | Download |
|--------------------------|-----------|---|
| x64 Installer | 141.86 MB |  jdk-11.0.28_windows-x64_bin.exe |
| x64 Compressed Archive | 159.60 MB |  jdk-11.0.28_windows-x64_bin.zip |

*注意安装后需要修改环境变量JAVA_HOME

集群搭建

建立集群前先确认Java 版本

```
PS C:\Users\pxyfang> java -version
java version "11.0.28" 2025-07-15 LTS
Java(TM) SE Runtime Environment 18.9 (build 11.0.28+12-LTS-279)
Java HotSpot(TM) 64-Bit Server VM 18.9 (build 11.0.28+12-LTS-279, mixed mode)
PS C:\Users\pxyfang> .
```

Power shell里输入java -version 确认java的版本是11

集群搭建

1. 环境准备。二人一组（也可以一个人一组），每个人的电脑都需要安装同版本的Java，Spark以及winutils（hadoop）。并且要让所有电脑处于同一个网络下，以便相互通信。在同一个网络下的标准是能够相互ping通。Ping不通的时候可以尝试关闭下防火墙。同一个网络可以用手机做热点连接。
2. 设定 Master 节点。选择一个电脑作为 Master 节点。在这台电脑上启动 Spark Master 服务。记录 Master 节点的 IP 地址和端口号（默认是 7077）。
3. 配置 Worker 节点。在其他电脑（或者同一台电脑，如果你是一人成组的情况）上配置 Spark的worker节点，使其能够连接到 Master 节点。启动 Worker 服务，并指定 Master 节点的 URL（spark://{Master_IP}:7077）

```
Using Spark's default log4j profile: org/apache/spark/log4j2-defaults.properties
25/09/19 22:37:39 INFO Master: Started daemon with process name: 11832@DESKTOP-ROOD39U
25/09/19 22:37:40 INFO SecurityManager: Changing view acls to: pxyfang
25/09/19 22:37:40 INFO SecurityManager: Changing modify acls to: pxyfang
25/09/19 22:37:40 INFO SecurityManager: Changing view acls groups to:
25/09/19 22:37:40 INFO SecurityManager: Changing modify acls groups to:
25/09/19 22:37:40 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: pxyfang; gr
25/09/19 22:37:40 INFO Uti ls: Successfully started service 'spark25/09/19 22:38:19 INFO SecurityManager: Changing view acls to: pxyfang'
25/09/19 22:37:40 INFO Master: Starting Spark master at spark://25/09/19 22:38:19 INFO SecurityManager: Changing modify acls groups to:
25/09/19 22:37:40 INFO Master: Running Spark version 3.5.6 25/09/19 22:38:19 INFO SecurityManager: Changing modify acls groups to:
25/09/19 22:37:40 INFO JettyUtils: Start Jetty 0.0.0.0:8080 for 25/09/19 22:38:19 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: pxyfang; gr
25/09/19 22:37:40 INFO Uti ls: Successfully started service 'Master25/09/19 22:38:19 INFO Worker: Worker decommissioning not enabled.
25/09/19 22:37:40 INFO MasterWebUI: Bound MasterWebUI to 0.0.0.0:8080 25/09/19 22:38:20 INFO Worker: Starting Spark worker 192.168.0.186:80187 with 4 cores, 8.0 GiB RAM
25/09/19 22:37:41 INFO Master: I have been elected leader! New 25/09/19 22:38:20 INFO Worker: Running Spark version 3.5.6
25/09/19 22:38:20 INFO Worker: Spark home: C:\spark-3.5.6-bin-hadoop3
25/09/19 22:38:20 INFO ResourceUtils: =====
25/09/19 22:38:20 INFO ResourceUtils: No custom resources configured for spark.worker.
25/09/19 22:38:20 INFO ResourceUtils: =====
25/09/19 22:38:20 INFO JettyUtils: Start Jetty 0.0.0.0:8081 for WorkerUI
25/09/19 22:38:20 INFO Uti ls: Successfully started service 'WorkerUI' on port 8081.
25/09/19 22:38:20 INFO WorkerWebUI: Bound WorkerWebUI to 0.0.0.0, and started at http://DESKTOP-ROOD39U:8081
25/09/19 22:38:20 INFO Worker: Connecting to master 192.168.0.186:7077...
25/09/19 22:38:20 INFO TransportClientFactory: Successfully created connection to /192.168.0.186:7077 after 33 ms (0 ms spent in bootstrap)
25/09/19 22:38:20 INFO Worker: Successfully registered with master spark://192.168.0.186:7077
```

MASTER节点的输出

WORKER节点的输出

集群搭建

4. 验证集群状态。访问 Master 节点的 Web UI（默认是`http://{Master_IP}:8080`）检查是否所有 Worker 都已连接。
5. 运行应用程序。运行附件所提供的 *集群上运行最小示例.html* 中的代码，需要在创建SparkSession时候指定Master的URL。

 **Spark Master at spark://192.168.0.186:7077**

URL: spark://192.168.0.186:7077

Alive Workers: 1

Cores in use: 4 Total, 0 Used

Memory in use: 8.0 GiB Total, 0.0 B Used

Resources in use:

Applications: 0 Running, 3 Completed

Drivers: 0 Running, 0 Completed

Status: ALIVE

Workers (1)

| Worker Id | Address | State | Cores | Memory | Resources |
|---|---------------------|-------|------------|----------------------|-----------|
| worker-20250919215619-192.168.0.186:63396 | 192.168.0.186:63396 | ALIVE | 4 (0 Used) | 8.0 GiB (0.0 B Used) | |

Running Applications (0)

| Application ID | Name | Cores | Memory per Executor | Resources Per Executor | Submitted Time | User | State | Duration |
|----------------|------|-------|---------------------|------------------------|----------------|------|-------|----------|
|----------------|------|-------|---------------------|------------------------|----------------|------|-------|----------|

Completed Applications (3)

| Application ID | Name | Cores | Memory per Executor | Resources Per Executor | Submitted Time | User | State | Duration |
|-------------------------|---------|-------|---------------------|------------------------|---------------------|---------|----------|----------|
| app-20250919220804-0002 | MyApp | 4 | 1024.0 MB | | 2025/09/19 22:08:04 | pxyfang | FINISHED | 44 s |
| app-20250919220854-0001 | SparkPi | 4 | 1024.0 MB | | 2025/09/19 22:06:54 | pxyfang | FINISHED | 10 s |
| app-20250919220229-0000 | MyApp | 4 | 1024.0 MB | | 2025/09/19 22:02:29 | pxyfang | FINISHED | 3.3 min |

提示

- 使用spark-class命令先后启动Master和Worker节点（该命令所用参数自己探索）。
- 额外给的 集群上运行最小示例.html 文件中的master的Ip地址要
根据自己集群中master的地址进行修改。

```
from pyspark.sql import SparkSession
import random

# 连接到 Spark 集群
spark = (
    SparkSession.builder
        .appName("SparkPi")
        .master("spark://192.168.0.186:7077") # 这里假设你的master在192.168.0.186
        .getOrCreate()
)

sc = spark.sparkContext

# ----- 计算 Pi -----
def inside(_):
    x, y = random.random(), random.random()
    return 1 if x*x + y*y < 1 else 0

num_samples = 1000000

# 分发到集群上做并行计算
count = sc.parallelize(range(0, num_samples), numSlices=10) \
    .map(inside) \
    .reduce(lambda a, b: a + b)

pi = 4.0 * count / num_samples
print(f"Pi is roughly {pi}")

spark.stop()
```

这里要进行修改