

集群搭建

1. 环境准备

- 每个小组成员（这里是三人）的电脑都需要安装Java和Spark。
- 确保所有电脑在同一个网络下，以便相互通信。

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

- 在其他两台电脑（Worker 节点）上配置 Spark，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）

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（可不作）

- 在任一节点上提交上节课的程序，指定 Master URL。在集群上提交程序时，需要在创建SparkSession时候指定Master URL。