

# 学习大数据从哪儿下手

学习 hadoop，从哪儿开始学首先要根据你的基本情况而定。

**如果**你就一小白，没有任何开发基础，也没有学过任何开发语言，那就必须先从第一阶段基础 java 开始学起（大数据支持很多开发语言，但企业用的最多的还是 JAVA），需要学习的内容如下：

**第一阶段(加\*\*的需重点熟练掌握，其他掌握)：**

- 1、Java 基础\*\*
  - 1) 数据类型
  - 2) 运算符、循环
  - 3) 顺序结构程序设计
  - 4) 程序结构
  - 5) 数组及多维数组
- 2、面向对象\*\* 构造方法、控制符、封装
- 3、继承\*\*
- 4、多态\*\*
- 5、抽象类、接口\*\* 常用类
- 6、集合 Collection、list\*\*  
*HashSet、TreeSet、Collection*
- 7、集合类 Map\*\*
  - 1) 异常
  - 2) *File*
- 8、文件/流\*\*
  - 1) 数据流和对象流
  - 2) 线程（理解即可）
  - 3) 网络通信（理解即可）

**如果**如果你已经是脱离小白生涯，你理大数据不远了，需要学习一些额外的小知识（关系型数据库、linux 系统操作）第二阶段以夯实基础，之后就可以进入大数据学习了；

**第二阶段（必备技能）：**（推荐老男孩 linux 入门教程，高性能教程）

1. 关系型数据库（mysql,oracle 等等[java 就业班中有]）
2. 非关系性数据库（mongoDB、redis（高性能教程中有）,Hbase）
3. Linux 系统操作
4. Linux 操作系统概述

5. 安装 Linux 操作系统
6. 图形界面操作基础
7. Linux 字符界面基础
8. 字符界面操作进阶
9. 用户、组群和权限管理
10. 文件系统管理
11. 软件包管理与系统备份
12. Linux 网络配置

主要掌握 Linux 操作系统的理论基础和服务器配置实践知识，同时通过大量实验，着重

培养学生的动手能力。使学生了解 Linux 操作系统在行业中的重要地位和广泛的使用范围。

在学习 Linux 的基础上，加深对服务器操作系统的认识和实践配置能力。加深对计算机网络

基础知识的理解，并在实践中加以应用。掌握 Linux 操作系统的安装、命令行操作、用户管

理、磁盘管理、文件系统管理、软件包管理、进程管理、系统监测和系统故障排除。

掌握 Linux

操作系统的网络配置、DNS、DHCP、HTTP、FTP、SMTP 和 POP3 服务的配置与管理。为更深一

步学习其它网络操作系统和软件系统开发奠定坚实的基础。与此同时，如果大家有时间把

javaweb 及框架学习一番，会让你的大数据学习更自由一些。

**重点掌握：**

- 1) 常见算法
- 2) 数据库表设计
- 3) SQL 语句
- 4) Linux 常见命令

**如果**你已经是小牛，只不过不了解大数据而已，那么恭喜你，我们可以给你介绍大数据相关平台了。例如：hadoop 离线分析、Storm 实时计算、spark 内存计算三种计算平台，学习体系可以按照如下：

**第三阶段（大数据平台技术）：**

1. Hadoop 离线批处理平台（推荐 03hadoop => 12 期大数据 8 天）  
HDFS、MapReduce、Hive、Hbase、Sqoop 等
2. Storm 实时数据处理平台

### 3. Spark 平台

**提示：**若之前没有项目经验或 *JAVA* 基础，掌握了第一阶段进入企业，不足以立即上手做大数据

项目，企业需再花时间与成本培养；

第二阶段掌握扎实以后，进入企业就可以跟着做大数据项目了，跟着一大帮人做项目倒也不用太担

心自己能不能应付的来，当然薪资不能有太高的要求；

前两个阶段都服务于第三阶段的学习，除了熟练掌握这些知识以外，重点需要找些相应的项目去做