

『StuQ 大数据微课堂基础预习资料』

祝海林老师推荐

1. 语言：

大数据目前大部分主要组件都是用Java/Scala开发。下面是对应参考文档： - Scala 入门手册：<http://www.scala-lang.org/docu/files/ScalaTutorial.pdf> (英文版) http://www.scala-lang.org/docu/files/ScalaTutorial-zh_CN.pdf (中文版) - Java的话，还是推荐 【Thinking In Java】这本书了。

2. 关于 Spark

对于Spark,其实官方文档足够了，但是因为是英文，有些朋友可能看起来会有些吃力：

中文版：<https://github.com/aiyanbo/spark-programming-guide-zh-cn>

但是因为Spark发展非常快，所以更新不一定能及时。但是可以让你了解spark的体系。

- 这里再推荐几篇入门文章：
 1. <http://www.csdn.net/article/2015-07-10/2825184>
 2. <http://www.infoq.com/cn/articles/apache-spark-introduction>
- Spark官方文档：<http://spark.apache.org/docs/latest/programming-guide.html>
- Spark MLlib 文档 <http://spark.apache.org/docs/latest/mllib-guide.html> 书籍的话，因为我都没看过，不好做推荐。

3. 机器学习相关

- 吴恩达的机器学习不可错过。在网易斯坦福公开课有。
- 另外希望这个知乎帖子也能帮到你：<https://www.zhihu.com/question/20691338>。

吴雪峰老师推荐：

1. Scala 培训资料：<https://github.com/XuefengWu/ScalaTraining>
2. Spark 峰会培训资料：<https://spark-summit.org/2014/training>

StuQ 编辑推荐资料：

1. 祝海林老师专栏：http://www.jianshu.com/users/59d5607f1400/latest_articles
2. 桑文锋老师专栏：<http://zhuanlan.zhihu.com/sangwf/20496355>
3. 陈敏敏老师书籍：《Storm技术内幕与大数据实践》
4. 张丹老师博客：
 - Hadoop家族系列文章：<http://blog.fens.me/series-hadoop-family/>
 - 书籍：《R的极客理想》
5. StuQ 大数据工程师技能图谱 Github 地址：<https://github.com/TeamStuQ/skill-map>