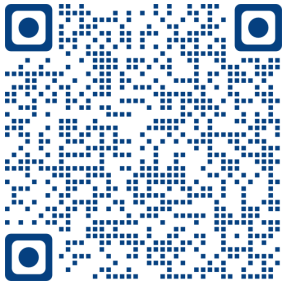


大数据技术基础及应用

陈一帅

yschen@bjtu.edu.cn

北京交通大学电子信息工程学院网络智能实验室



A 课程信息

- [分组信息表](#)
- Lab 提交文档链接（腾讯文档）
 1. [Lab 1 云计算平台调研](#)
 2. [Lab 2 微服务系统开发入门](#)
 3. [Lab 3 Spark编程入门](#)
- 本站链接: <https://yishuai.github.io/bigdata/bigdata.html>
- 本站材料Github下载: <https://github.com/yishuai/bigdata>

B 课程材料

一、大数据的信息基础设施

从计算、存储和网络的基础设施的演进，引出驱动大数据发展的底层物理技术因素。

1. 简介: [网页](#), [PDF](#), [Lab 1 云计算平台调研](#)
2. 存储模型: [网页](#), [PDF](#)
3. 计算虚拟化: [网页](#), [PDF](#), [Lab 2 微服务系统开发入门](#)

二、大数据编程模型和使用技巧

本节介绍大数据技术平台的编程模型和使用技巧。要求学生进行 Mapreduce、SQL 类查询、Spark 的编程实践。要求学生通过实践编程理解大数据计算模型和实际编程技术，为后面的分析打下基础。

1. Hadoop/YARN: [网页](#), [PDF](#)
2. Spark: [网页](#), [PDF](#), [Lab 3 Spark编程入门](#)
 - 参考资料
 - 王一行: [Python 和 JupyterNotebook 环境安装指南 \(DOC\)](#)
 - 斯坦福CS231n: [Jupyter Notebook 入门指南 \(IPYNB\)](#)

- 斯坦福CS231n: [Python Numpy 编程入门指南 \(PDF\)](#)
 - 陈一帅, [Spark原理与实践](#), 2020年上课材料, 包括 B站视频, PPT, 实验指南等
3. 流式计算: [网页](#), [PDF](#)

三、大数据信息处理技术

本节介绍大数据技术平台的信息处理技术, 包括文本处理, 时间序列处理。

1. Rayid Ghani, 芝加哥大学, Text Analytics 101, [PDF](#)