**大数据技术师资培训班**

**教学大纲和教学计划**

为了推动高校数据科学与大数据技术专业课程建设，促进中国大数据人才的培养，夯实高校大数据教育人才的储备，提高高校教师的大数据技术教学改革与创新的业务水平和能力，由教育部与工业和信息化部指导的信息技术新工科产学研联盟教师培训工作委员会和江苏省地方高校产教融合工作委员会，联合举办系列大数据技术师资培训班。

本次培训邀请来自南京大数据研究院、金陵科技学院、南京信息工程大学、南京云创大数据科技股份有限公司等大数据领域的课程专家和技术专家现场授课，为有志于在高校开展大数据教育工作、培养大数据人才的青年教师提供深入培训及交流机会。

1. **教学计划**

|  |  |  |
| --- | --- | --- |
| **培训日期** | **培训时间** | **培训内容** |
| 第一天上午 | 08:00-09:00 | 签到 |
| 09:00-09:30 | 大数据师资培训班开班仪式 |
| 09:30-10:30 | 大数据人才培养体系 张燕副校长 |
| 10:30-12:00 | 大数据理论框架与应用实践 刘鹏教授 |
| 第一天下午 | 13:30-15:00 | 《深度学习》全书内容串讲，赵海峰博士 |
| 15:00-15:10 | 茶歇 |
| 15:10-15:50 | 深度学习 专家讲座，刘光灿教授 |
| 15:50-16:10 | 《云计算》课程介绍，刘鹏教授 |
| 16:10-16:30 | 《大数据》课程介绍，刘鹏教授 |
| 16:30-16:50 | 《大数据库》课程介绍，刘鹏教授 |
| 第二天上午 | 09:00-10:30 | 《数据挖掘》全书内容串讲，施建强教授 |
| 10:30-10:40 | 茶歇 |
| 10:40-11:20 | 数据挖掘 专家讲座，业宁教授 |
| 11:20-11:50 | 人工智能实验平台介绍及演示，云创工程师 |
| 第二天下午 | 13:30-15:00 | 《大数据可视化》全书内容串讲，何光威教授 |
| 15:00-15:10 | 茶歇 |
| 15:10-15:50 | 大数据可视化 专家讲座，毕硕本教授 |
| 15:50-16:20 | 大数据实验平台介绍及演示，云创工程师 |
| 16:20-16:40 | 《大数据实验手册》课程介绍，沈大为总监 |
| 第三天上午 | 09:00-10:30 | 大数据Hadoop实验，云创工程师 |
| 10:30-10:40 | 茶歇 |
| 10:40-12:10 | 大数据Spark实验，云创工程师 |
| 第三天下午 | 13:30-15:00 | 数据挖掘实验，云创工程师 |
| 15:00-15:20 | 大数据师资培训班结业典礼 |

**二、教学大纲**

|  |  |  |
| --- | --- | --- |
| **课程名称** | **课时** | **主要内容** |
| 大数据课程体系 | 4 | 大数据人才培养与课程体系介绍，云计算与大数据课程系列课程介绍，包括Google云计算、Amazon云计算、微软云计算，Hadoop、虚拟化、OpenStack，大数据采集与预处理、数据挖掘算法与工具，以及相关实验。 |
| 大数据实验 | 8 | Hadoop、Spark、数据挖掘配置、部署、实验 |
| 深度学习 | 4 | 深度学习的基本概念、数学基础和主流模型，深度学习的应用框架，以及在图像、文本、语音中的应用，以及前沿发展 |
| 数据挖掘 | 4 | 经典数据挖掘算法，大数据环境下常用数据挖掘算法的优化，大数据新常态下催生的数据分析方法 |
| 大数据可视化 | 4 | 大数据可视化概述、可视化的类型与模型、数据可视化基础、数据可视化的常用方法、大数据可视化的关键技术、可视化交互、大数据可视化工具、大数据可视化系统及大数据可视化的行业案例 |

**二、教材：**

1、《深度学习》

本书从深度学习的发展历史入手，系统地介绍了深度学习的基本概念、数学基础和主流模型，以夯实读者的理论基础。同时，针对当前流行的主流框架，循序渐进，一步一步引导读者了解深度学习的使用过程，增强动手能力。在此基础上，通过具体例子介绍了深度学习在图像、语音、文本中的应用，还简要地介绍了增强学习、迁移学习、深度学习硬件实现等前沿知识，扩展了深度学习的内容。每章都附有相应的习题和参考文献，以便感兴趣的读者进一步深入思考。“让学习变得轻松”是本书的基本编写理念。



2、《数据挖掘》

本书的定位是大数据挖掘技术与应用。本书系统地介绍了数据挖掘算法理论与方法、工具和应用，包括经典数据挖掘算法，大数据环境下常用数据挖掘算法的优化，大数据新常态下催生的数据分析方法（如推荐系统、链接分析与网页排序、互联网信息抽取、日志挖掘与查询分析）、工具与应用。



3、《大数据可视化》

本书是针对当前大数据应用、可视化分析研究和应用的新形势，专门为大数据专业本科生编写的大数据可视化教材。本书全面诠释了大数据可视化的内涵与外延，详细介绍了大数据可视化概述、可视化的类型与模型、数据可视化基础、数据可视化的常用方法、大数据可视化的关键技术、可视化交互、大数据可视化工具、大数据可视化系统及大数据可视化的行业案例等内容。为了便于学习，每章都附有习题，题型设计也充分考虑到大数据可视化课程教学的需要。



4、《云计算》

本书是被绝大多数高校采用的教材《云计算（第三版）》的最新升级版，是中国云计算专家咨询委员会秘书长刘鹏教授团队的心血之作。在应对大数据挑战的过程中，云计算技术日趋成熟，拥有大量的成功商业应用。本书追踪最新技术，相比第二版更新了60%以上的内容，包括大数据与云计算、Google云计算、Amazon云计算、微软云计算、Hadoop 2.0及其生态圈、虚拟化技术、OpenStack开源云计算、云计算数据中心、云计算核心算法和中国云计算技术等。刘鹏教授创办的中国云计算（chinacloud.cn）、中国大数据（thebigdata.cn）网站和刘鹏微信公众号（lpoutlook）为本书学习提供技术支撑。



5、《大数据》

本书是国内大多数高校采用的知名教材《云计算》（1～3版）的姊妹篇，是中国大数据专家委员会刘鹏教授联合国内多位专家历时两年的心血之作。大数据领域一直缺乏一本合适的教材，希望本书能够填补空白。本书系统地介绍了大数据的理论知识和实战应用，包括大数据采集与预处理、数据挖掘算法与工具、深度学习以及大数据可视化等，并深度剖析了大数据在互联网、商业和典型行业的应用。本书的全部实验可以在大数据实验平台（https://bd.cstor.cn）上远程开展，也可在高校部署的BDRack大数据实验一体机上本地开展。



6、《大数据库》

本书是全国高校标准教材《大数据》的姊妹篇，是中国大数据专家委员会刘鹏教授联合国内多位专家历时两年的心血之作。本书系统地介绍了大数据库的理论知识和实战应用，包括大数据库概述、分布式数据库HBase、数据仓库工具Hive、大数据查询系统Impala、内存数据库Spark、Spark SQL、键值数据库、流式数据库和大数据应用托管平台Docker等。本书紧跟大数据的发展前沿，既有理论深度，又有实践价值。刘鹏教授创办的网站中国大数据（thebigdata.cn）、中国云计算（chinacloud.cn）和微信公众号刘鹏看未来（lpoutlook）将免费提供PPT和其他资料，为本书的学习提供技术支撑。



7、《大数据实验手册》

本书是中国大数据专家委员会刘鹏教授团队的心血之作。针对高校大数据相关专业实践教学以及个人提升大数据动手能力的需求，带领大数据研发团队，经过反复实践、提炼和验证形成本书。本书主要内容包括HDFS实验、YARN实验、MapReduce实验、Hive实验、Spark实验、ZooKeeper实验、HBase实验、Storm实验、MongoDB实验、LevelDB实验、Mahout实验和综合实战等。每个实验呈现详细的实验目的、实验内容、实验原理和实验流程。本书所有实验均在向公众开放的大数据实验平台（https://bd.cstor.cn）和BDRack大数据实验一体机上验证通过。

