

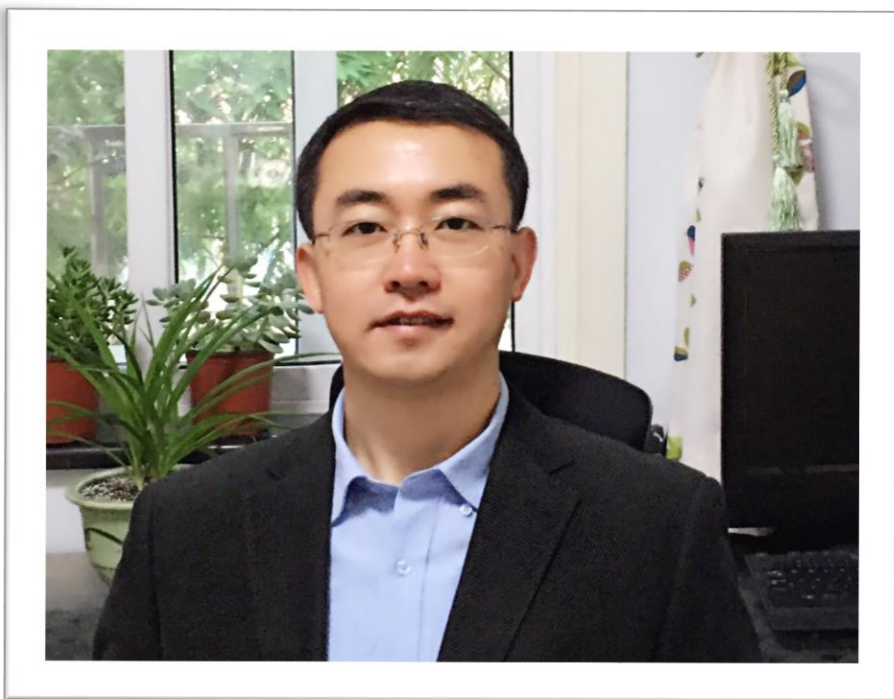
请扫码加入《大数据理论与实践》课程微信群



《大数据理论与实践》课程介绍

范颖捷 | 2018年9月

1 教师简介



范颖捷 博士，毕业于复旦大学计算机学院，星环高级技术专家，大数据架构师。

先后就职于军队院校和大数据公司，长期从事计算机软件的研发、科研和教学工作，具有较扎实的理论功底和较宽广的专业视野，对大数据技术体系有比较全面、深刻的理解。在军事、政务、教育、金融等领域积累了丰富的的大数据解决方案实施经验。

具有22年项目经验，17年教学经验，主持或参与国家和国际标准2项、科研课题和软件开发项目28项，参编专著3部，撰写学术论文10多篇。

2 教学目标

➤ 掌握开源大数据技术的基础理论

- 文件系统：HDFS
- 资源管理框架：YARN
- 计算框架：MapReduce、Spark Core
- ETL工具：Sqoop、Flume、Kafka
- 数仓和SQL引擎：Hive
- NoSQL数据库：HBase
- 全文搜索引擎：ElasticSearch

➤ 掌握星环TDH平台的基本原理和使用方法

- 数仓和SQL引擎：Inceptor（基于Hive和Spark）
- 流处理引擎：Slipstream（基于Spark）
- NewSQL数据库：Hyperbase（基于HBase）
- 全文搜索引擎：Search（基于ElasticSearch）



夯实基础！！！！

3 课程安排

时间	主题	课程名称
第1周	大数据技术概览	大数据技术综述
第2周	Hadoop生态圈：HDFS	分布式文件系统HDFS
第3周	Hadoop生态圈：YARN	分布式资源管理系统YARN
第4周	Hadoop生态圈：MR & Spark	分布式计算框架MapReduce & Spark
第5周	Hadoop生态圈：Sqoop、Flume和Kafka	分布式ETL工具Sqoop
		分布式数据采集工具Flume
		分布式消息队列Kafka
第6周	大数据平台：安装部署实验	大数据平台的功能介绍
		大数据平台的软硬件环境要求
		大数据平台的角色分配

3 课程安排

时间	主题	课程名称
第6周	Hadoop生态圈：实验	HDFS基本操作（实验）
		YARN基本操作（实验）
		Sqoop基本操作（实验）
		Flume基本操作（实验）
		Kafka基本操作（实验）
第7周	大数据平台组件：Inceptor	分布式数据仓库与SQL引擎Inceptor
第8周	大数据平台组件：Slipstream	实时流处理引擎Slipstream
第9周	大数据平台组件：Hyperbase	分布式NewSQL数据库Hyperbase
第10周	大数据平台组件：Search 大数据技术概览	分布式全文搜索引擎Search
		大数据技术总结与展望

3 课程安排

时间	主题	课程名称
第11周	大数据平台组件：实验	SQL集成开发工具Waterdrop的安装与使用（实验）
		Inceptor使用方法（实验）
		Hyperbase使用方法（实验）
		Slipstream使用方法（实验）
		Search使用方法（实验）
第12周	考试	笔试（开卷）