课时一：概念综述
1、大数据的定义由来和原因
2、大数据的 6V 特征
3、从数据库，数据仓库到大数据
4、大数据相关技术和处理

课时二：Hadoop 生态圈、spark 生态圈、搜索引擎概述
1、hadoop：HDFS、Map-Reduce、Hbase、Hive 等
2、spark：scala、spark-SQL、spark-Streaming 等
3、搜索引擎：lucene（solr）、ES
4、并发的机器学习工具：R-hadoop、spark-MLLIB、 spark-R、pyspark

课时三：存储在 hbase 中的数据
1、NoSQL（key-value）
2、Hbase：安装
3、行键与列簇
4、如何利用 Hbase 的特点存储数据
5、应用程序如何访问 Hbase 中的数据
6、数据迁移：sqoop
7、Hbase 的应用场景

课时四：Hive：为用 SQL 的开发者留的活路
1、Hive：安装（单用户与多用户）
2、Hive：基本操作
3、Hive：与典型的关系型数据库的区别
4、如果“想慢”，你可以这样…（不恰当使用 hive 的案例介绍）
5、Hive 的应用场景

课时五：Spark 各组件在卫生领域的应用
1、Hadoop 最大的特点是什么？
2、Spark 概述与安装
3、Scala：你可以一直“点”下去
4、RDD：“映射”、“转换”解决一切
5、spark-SQL
6、spark-streaming
7、spark 的其他组件
8、应用场景

课时六：机器学习算法介绍—I
1、综述（人工智能、数据挖掘、机器学习、机器智能、大数据：这些词的确切含义）
2、监督学习、无监督学习与强化学习
3、工具：R、Python 等
4、决策树详解（熵、贪心法、连续的和离散的）
5、神经网络详解（神经元、激励函数、前馈神经网络的 BP 算法，其他神经网络）

课时七：机器学习算法介绍—II

1、关联规则详解（频繁项集、Apriori、支持度、置信度）

2、聚类详解（k-means、k-medoid）

3、常见算法的简述（Naïve-Bayes、k-NN、HMM、SVM 等）