



泰迪智能科技

服务热线：40068-40020

# 顶尖大数据工程教学实训平台 (TipDM-H8)

---

## 产 品 说 明 书

---

广州泰迪智能科技有限公司 版权所有

地址： 广州市经济技术开发区科学城 232 号

网址： <http://www.tipdm.com>

邮箱： [services@tipdm.com](mailto:services@tipdm.com)

热线： 40068-40020

企业 QQ： 40068-40020

邮编： 510663

电话： （020） 82039399

# 目 录

1	产品简介 .....	3
2	产品组件 .....	5
2.1	大数据挖掘建模平台 .....	5
2.2	大数据开发实训平台 .....	8
2.3	大数据分析教学实训教程 .....	9
2.4	大数据分析教学案例库及配套资源 .....	11
2.5	大数据挖掘教学沙盘 .....	13
2.6	大数据整合平台 .....	14
2.7	课程规划及师资培训 .....	15
2.8	学生实习实训 .....	16
3	产品优势 .....	17
4	应用案例 .....	错误!未定义书签。
4.1	主要客户 .....	错误!未定义书签。
4.2	典型示例 .....	17
4.2.1	实验室架构 .....	18
4.2.2	设备清单 .....	18
4.2.3	环境部署 .....	19
4.3	虚拟机资源 .....	19
4.3.1	硬件配置及软件部署 .....	20

# 1 产品简介

大数据的出现催生出产业人才缺口瓶颈，在大数据挖掘项目的实施方面，被调查公司普遍缺乏相关的技术能力。75%以上的公司表示在人员和培训方面存在障碍，会大数据挖掘技术的人才很热门，但是比较难找而且昂贵，会 Hadoop 技术的数据挖掘人才更是奇缺。

虽然大数据目前在国内还处于初级阶段，但是其商业价值已经显现出来，特别是有实践经验的大数据分析人才更是各企业争夺的热门。为了满足日益增长的大数据分析人才需求，很多大学开始尝试开设不同程度的大数据分析课程。“大数据分析”作为大数据时代的核心技术，必将成为高校数学与统计学专业的重要课程之一。

目前高校大数据教学的主要困境：

- ✧ 教师资源严重不足
- ✧ 尚未建立起合理的大数据教学体系
- ✧ 实践教学平台搭建困难
- ✧ 大数据实践应用案例缺乏
- ✧ 学生缺少实战机会

大数据工程教学实训平台 (TipDM-H8, 简称H8)，是将云存储资源、服务器资源和网络资源整合，然后通过Vmware等虚拟化搭建私有云平台，在私有云平台上搭建统一的数据挖掘平台和基于Hadoop的大数据分析平台。

大数据工程教学实训平台的建设能让学生既掌握必要的理论基础，又能将大数据分析理论和方法用于解决实际问题。该大数据教学平台也能解决教学过程中实验环节非常薄弱，造成学生缺少数据挖掘实践经验、缺乏实践应用能力等问题。基于Hadoop实战项目，在大数据实战环境中，通过动手实操，让学员在短时间内掌握使用平台开发Hadoop程序，以完成高效的大数据存储、清洗和分析。

大数据工程教学实训平台的建设采用四层架构，其整体架构如下图所示。



大数据工程教学实训平台底层数据处理平台采用云计算技术实现，基于云计算的基础设施层能充分保障资源的有效利用率和资源的动态伸缩性，云基础设施层上搭建了基于VMware、Hyper-v或KVM等的虚拟化层，该虚拟化层能充分保证业务的连续性和平台运行的可靠性。建立在虚拟化层上的统一数据挖掘平台TipDM-HB 能满足有大数据挖掘需求的师生利用图形化的界面进行大数据分析，而基于Hadoop 的大数据开发平台（TipDM-HD）能为大数据算法分析和大数据处理提供基础平台。位于架构最上层的大数据工程教学实训平台为培养大数据时代的数据科学家提供实训平台。

**适用对象：**

- ✧ 有大数据分析课程教学的高校、研究所和培训机构。
- ✧ 用大数据分析进行科研的高校、研究所。
- ✧ 计算机类及相关专业：软件工程、信息与计算科学、数据科学与大数据技术等。

**特别说明：**泰迪科技作为目前国内唯一Hortonworks官方同时认证的培训和系统集成签约合作伙伴，有多年的Hadoop开发及大数据分析和挖掘实战经验。TipDM团队的讲师均获得Hortonworks专业讲师资格认证。所有教材均由Hortonworks官方授权提供，课程涵盖Hadoop生态系统中所有常用组件，与美国Hortonworks College 保持同步，是进行Hortonworks Hadoop认证考试的唯一官方教程。

**建议：**如果采用泰迪科技提供的大数据工程教学实训平台建设方案，泰迪科技将派出讲师对实战中心技术人员进行专业培训（或直接派讲师进行授课），课程结束后，学员可直接参加认证考试！

## 2 产品组件

H8主要包含以下8个组件：

### 核心组件：

- 大数据挖掘建模平台（云数据挖掘引擎；云数据挖掘算法库；主机监控系统；云平台监控系统）；
- 大数据开发实训平台（基于VMware、Hyper-v或KVM等虚拟化的Hadoop开发实训平台）。

### 可选组件：

- 大数据分析教学实训教程：由校企联合打造的以解决企业大数据挖掘应用为目标的案例教程，主要包括：《Hadoop与大数据挖掘》、《Hadoop大数据分析与应用实战》；
- 大数据分析教学案例库及配套资源：所有案例均来自企业在大数据挖掘领域的典型应用；
- 大数据挖掘教学沙盘：模拟各种大数据应用案例场景的演练沙盘，如菜品智能推荐沙盘、电子商务网站用户行为分析与服务推荐沙盘、窃漏电用户自动识别沙盘、智能交通仿真沙盘、管网漏损智慧水务沙盘等；
- 大数据整合平台：包括数据管理、运维视图、接口管理、调度管理、系统管理等功能；
- 课程规划及师资培训：协助完成课程设置及教学规划，并为大数据挖掘课程教学提供师资培训；
- 学生实习实训：提供大数据挖掘实践项目，组织学生参与公司的项目实习及集中实训。

### 2.1 大数据挖掘建模平台

TipDM-HB是一套基于Hadoop架构的大大数据挖掘建模平台，支持Excel、Txt、关系型数据库及Hive数据源，支持数据源合并；提供多种数据挖掘算法进行建模，且所有算法提供API（JAR包）和WebService接口，方便业务系统进行调用；支持开源hadoop平台，同时可选支持多种商业hadoop发行平台；支持建模中间结果在线预览，分析结果能导出到Word文档及关系型数据库；屏蔽复杂的大数据分析底层技术，提供可视化数据分析用户界面，满足有大数据挖掘需求的师生利用图形化的界面进行大数据分析。

HB 主要包括：挖掘引擎、调度系统、主机监控系统、云平台监控系统、数据预处理、挖掘建模算法、统计分析方法、分类与预测算法、聚类算法、关联规则、分布式算法、时序模式、自定义算法等，具体如下表：

模块	说明
挖掘引擎	云分类引擎、云聚类引擎、云关联规则引擎、云智能推荐引擎等

调度系统	包括作业调度、作业监控和作业管理
主机监控系统	主要用来采集集群中主机的 cpu、memory、disk、process、network 等相关数据，并采用图形化的方式展示
云平台监控系统	主要用来监控云计算平台的运行指标，可以实时监控云平台的运行情况，子节点运行情况，用来为系统增加/删除/更新节点提供知道依据。它主要包括：1）分布式文件系统监控；2）作业监控；3）云平台管理等功能
数据预处理	特征提取、标准化处理、正交化处理、特征选择、异常值处理、离散处理、空值处理、数学类函数、属性选择等
挖掘建模算法	中值插值法、众值插值法、回归插值法、拉格朗日插值法、牛顿插值法、方差齐性检验、独立性样本 T 检验、二项分布检验、正态性检验、卡方检验、相关性检验、精确性检验、Granger 因果检验等
统计分析方法	1、基础统计分析：相关性分析、单因素方差分析、标识重复记录、主成分分析、因子分析等 2、高级统计分析：支持线性回归、逐步回归、逻辑回归、Cart 回归树、保序回归、C4.5 回归树、RPart 等
分类与预测算法	基于 R 语言的分类算法，如：朴素贝叶斯网络、贝叶斯信念网络、决策表、ID3 决策树、CART 决策树、C4.5 决策树、BP 神经网络、RBF 神经网络、AdaBoostM1 算法、KStar 算法、SVM 支持向量机、K-最近邻分类
聚类算法	基于 R 语言的聚类算法，如：k-Means 算法、快速 k-Means 算法、EM 最大期望算法、DBScan 密度算法、KMedoids 聚类、系统聚类
关联规则	基于 R 语言的关联规则算法，如：Apriori、HotSpot、时序模式关联
时序模式	基于 R 语言的时序模式，如：GM 灰色理论、ARIMA 模型、指数平滑等时序
分布式算法	基于 Spark 的分布式算法，如：1、SVD 奇异值分解 2、空值处理 3、时间序列插值 4、记录选择 5、异常值处理 6、离散化处理 7、PCA 主成分分析 8、FuzzyKMeans 算法 9、k-Means 算法 10、RandomForecast 算法 12、LinearRegression 算法 13、NaiveBayes 算法 14、LogisticRegression 算法 15、SVM 算法 16、Collaborative_Filtering 算法
自定义算法	1、支持自定义算法包上传、注册及发布 2、自定义算法的类型具体包括 R 算法、MR 算法、Spark 算法及 Weka 算法

运行界面：



[illegible]

Logged in as: dr.who

- Cluster
  - About
  - Nodes
  - Applications
    - NEW
    - NEW SAVING
    - SUBMITTED
    - ACCEPTED
    - RUNNING
    - REMOVING
    - FINISHING
    - FINISHED
    - FAILED
    - KILLED
  - Scheduler
- Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes
7	0	1	6	1	1 GB	24 GB	0 B	3	0	0	0	0

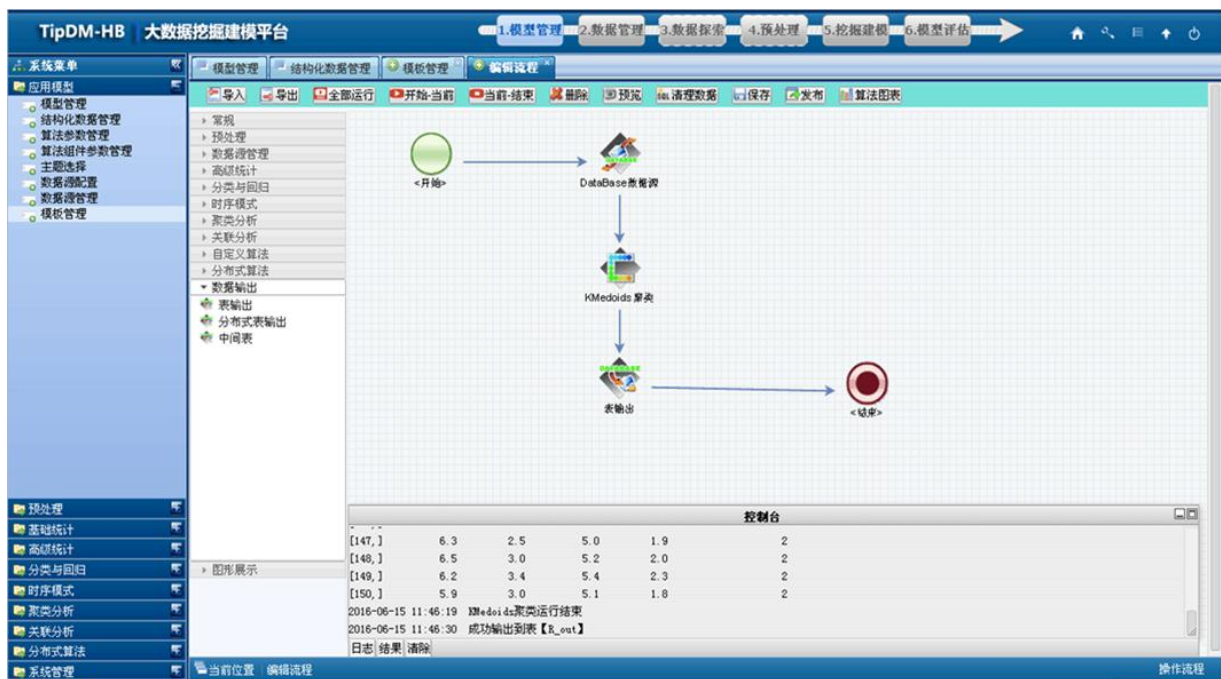
Show 20 ▾

entries

Search:

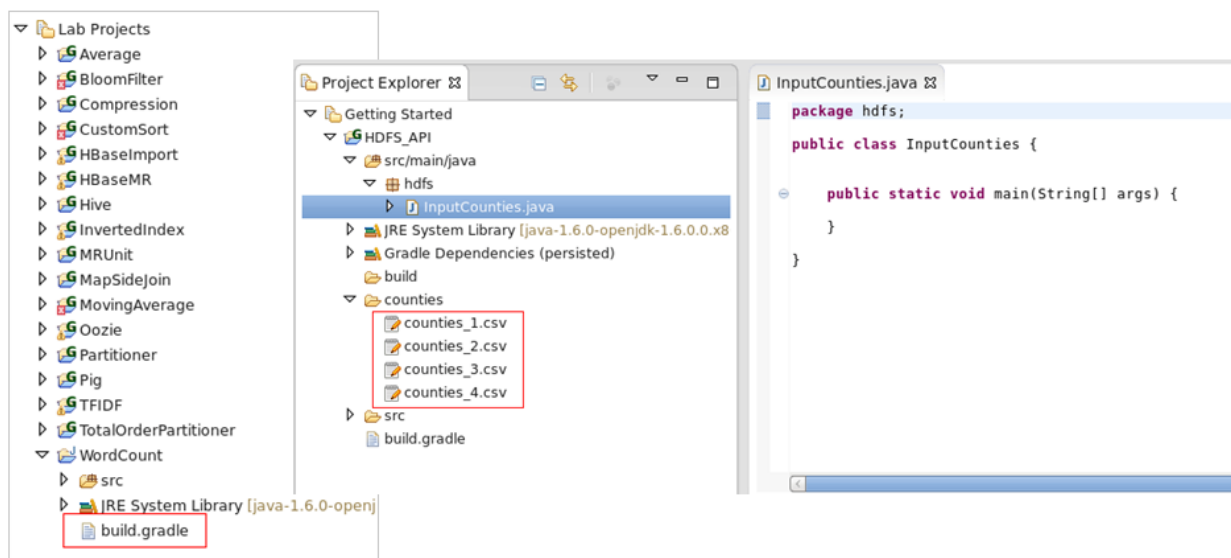
ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Progress	Tracking URL
<a href="#">application 1467852839248 0007</a>	root	kmeans algorithm run	SPARK	default	Thu, 07 Jul 2016 09:06:29 GMT	N/A	ACCEPTED	UNDEFINED	<div></div>	<a href="#">UNASSIGNED</a>
<a href="#">application 1467852839248 0006</a>	root	kmeans algorithm run	SPARK	default	Thu, 07 Jul 2016 02:56:05 GMT	Thu, 07 Jul 2016 02:56:43 GMT	FINISHED	SUCCEEDED	<div></div>	<a href="#">History</a>
<a href="#">application 1467852839248 0005</a>	root	kmeans model predict algorithm	SPARK	default	Thu, 07 Jul 2016 01:28:52 GMT	Thu, 07 Jul 2016 01:29:32 GMT	FINISHED	SUCCEEDED	<div></div>	<a href="#">History</a>
<a href="#">application 1467852839248 0004</a>	root	kmeans model predict algorithm	SPARK	default	Thu, 07 Jul 2016 01:27:27 GMT	Thu, 07 Jul 2016 01:28:06 GMT	FINISHED	SUCCEEDED	<div></div>	<a href="#">History</a>
<a href="#">application 1467852839248 0003</a>	root	kmeans algorithm run	SPARK	default	Thu, 07 Jul 2016 01:24:58 GMT	Thu, 07 Jul 2016 01:25:38 GMT	FINISHED	SUCCEEDED	<div></div>	<a href="#">History</a>
<a href="#">application 1467852839248 0002</a>	root	kmeans model predict algorithm	SPARK	default	Thu, 07 Jul 2016 00:59:04 GMT	Thu, 07 Jul 2016 00:59:44 GMT	FINISHED	SUCCEEDED	<div></div>	<a href="#">History</a>
<a href="#">application 1467852839248 0001</a>	root	kmeans	SPARK	default	Thu, 07 Jul 2016	Thu, 07 Jul 2016	FINISHED	SUCCEEDED	<div></div>	<a href="#">History</a>





## 2.2 大数据开发实训平台

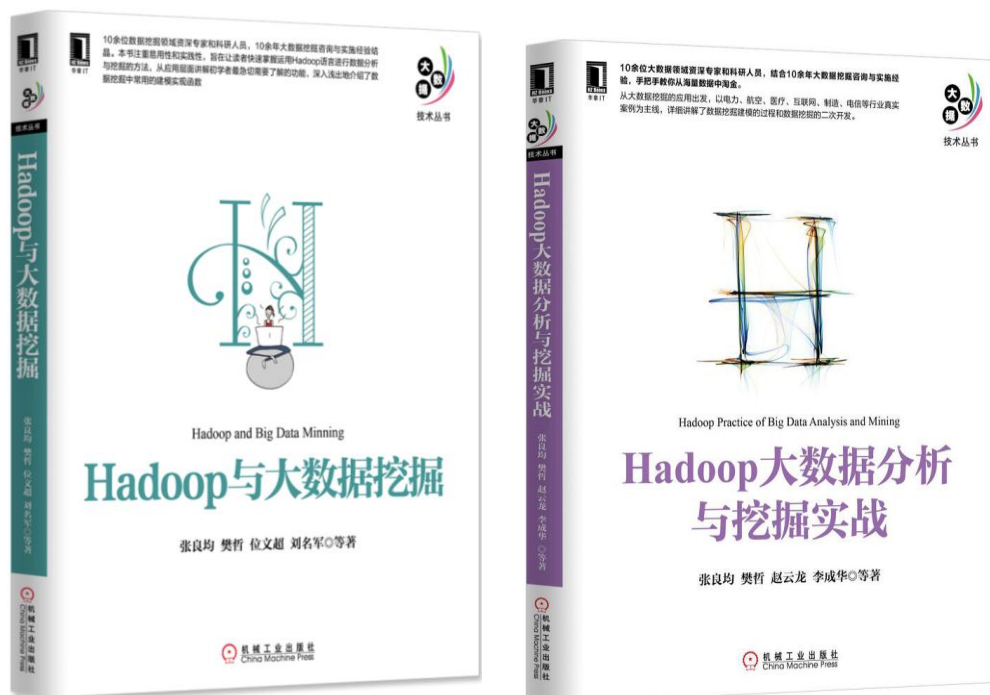
HD是一套建立在虚拟化层上的大数据开发平台，提供了一个开放、稳定和高度可扩展的框架，集成有Apache Hadoop的数据流业务与现有的数据架构。基于此平台，学员能在极短的时间内快速掌握各种Apache Hadoop项目以及Hadoop分布式文件系统（HDFS）、MapReduce、Pig、Hive、HBase、Zookeeper和其他各种组件的使用方法，并真正完成基于Hadoop框架的大数据应用开发。



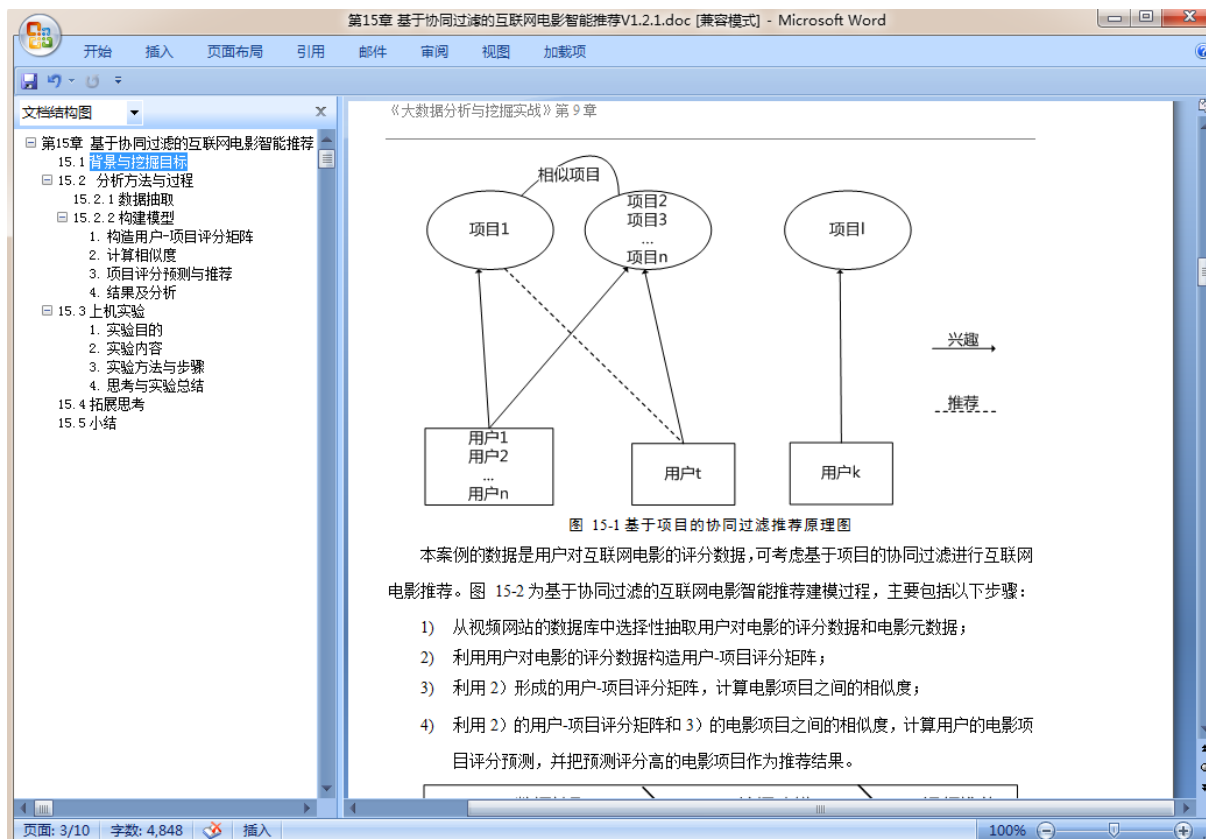
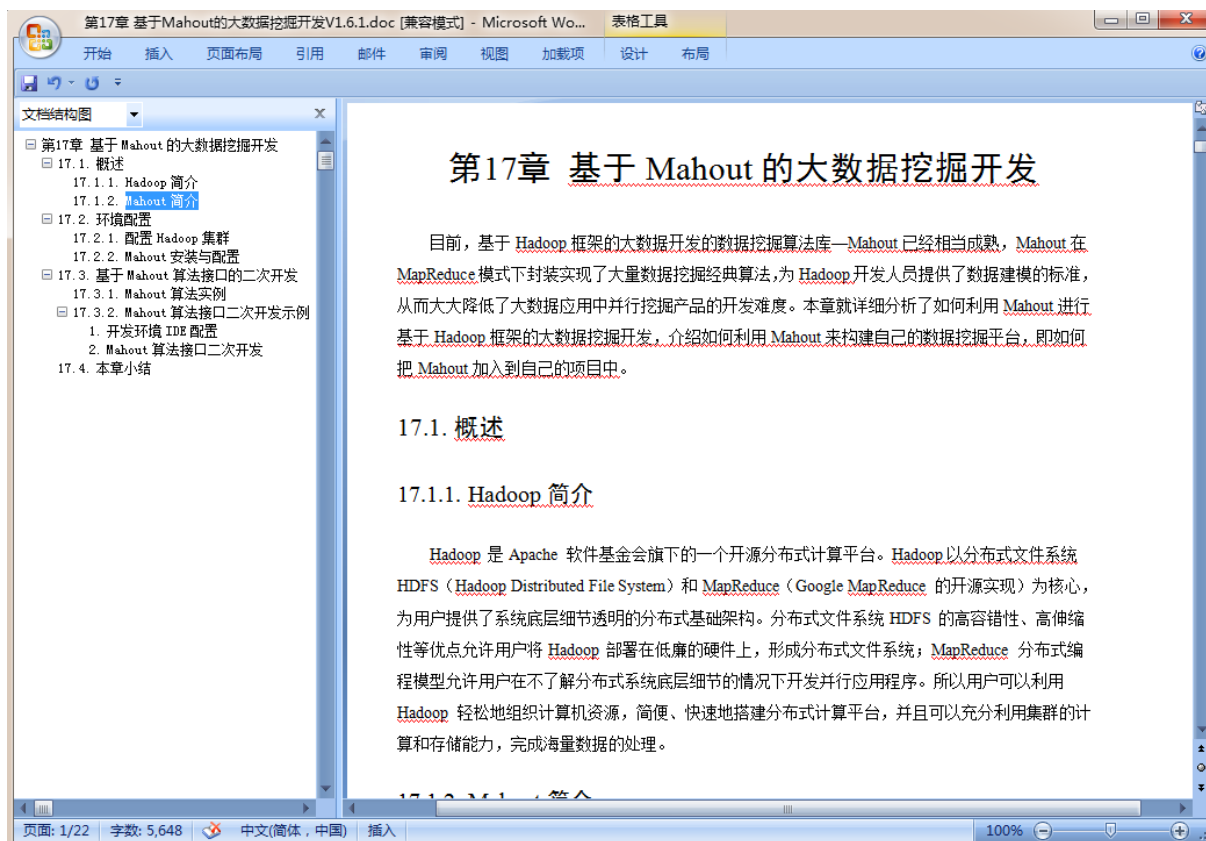


## 2.3 大数据分析教学实训教程

结合课程教学需要，H8提供《Hadoop大数据分析与应用实战》和《Hadoop与挖掘实战》辅助教学教程，方便教师授课和学生学习。



图书主要包括两部分：大数据基础理论和大数据挖掘案例分析。大数据基础理论主要包括常用算法、工具、Hadoop原理及开发；大数据挖掘案例分析详尽介绍10个经典的大数据应用案例，每个案例分别由挖掘目标的提出、分析方法与过程、建模仿真、结果评价和动手实践等部分构成，如下图所示：



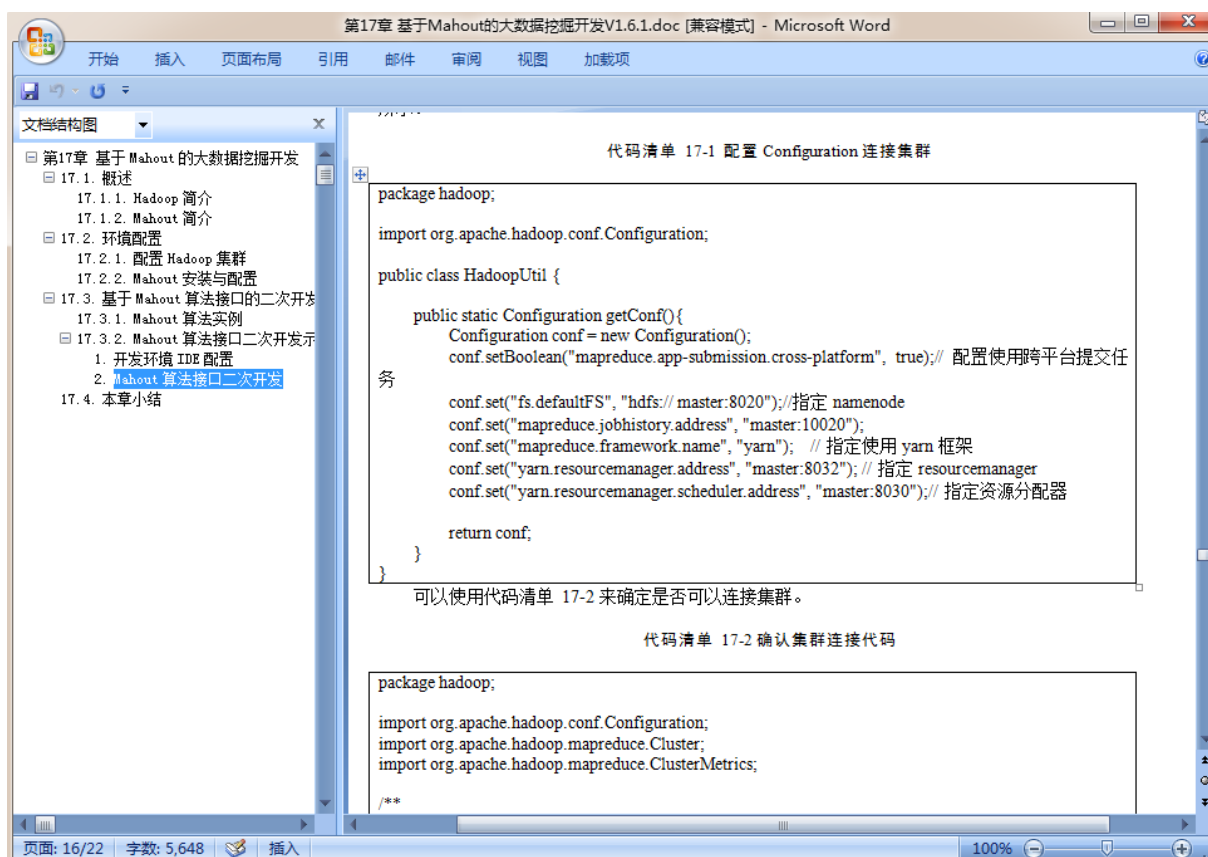
## 2.4 大数据分析教学案例库及配套资源

大数据分析教学案例库是泰迪科技在大数据挖掘领域探索10余年和高校资深讲师联合经验总结之作，内容涵盖智能电网、移动通信、医疗健康、网络舆情、电子商务、金融保险、交通运输等诸多行业，特别适合有大数据挖掘相关课程教学的高校、研究所和培训机构，也可作为个人学习大数据挖掘技术的最佳素材。

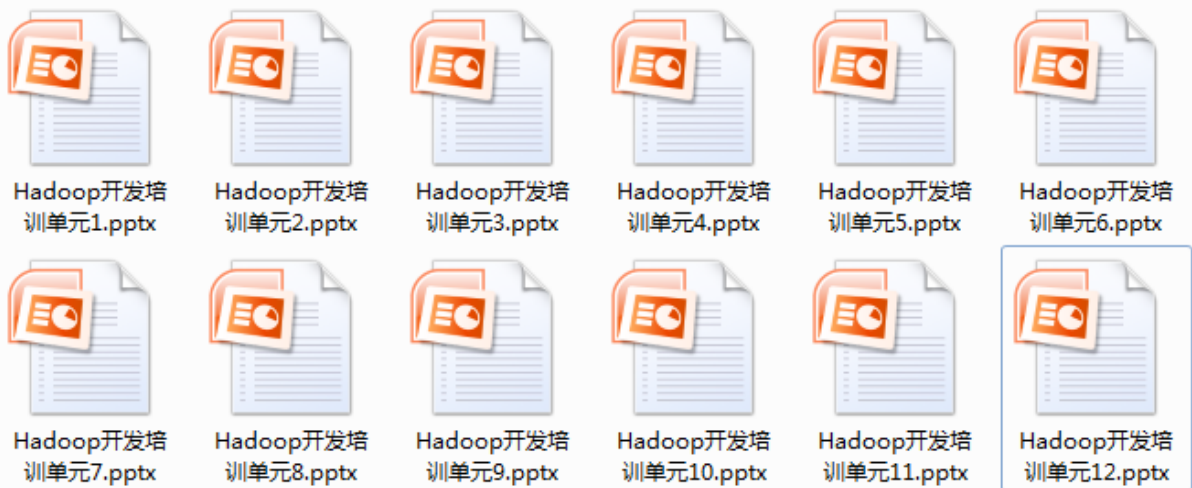
大数据挖掘案例主要包括为：

- ✧ 法律咨询数据分析与服务推荐
- ✧ 电商产品评论数据情感分析
- ✧ 航空公司客户价值分析
- ✧ 基于基站定位数据的商圈分析
- ✧ 基于协同过滤的互联网电影智能推荐
- ✧ 家电故障备件储备预测分析
- ✧ 市供水混凝投药量控制分析
- ✧ 基于图像处理的车辆压双黄线检测
- ✧ .....

案例配套资源包括：源数据及过程数据，Hadoop环境下的程序/模型。



结合课程教学需要，大数据工程教学实训平台(TipDM-H8)同时提供配套的教学用PPT：



- |   |  |   |
|---|--|---|
| 第1章 大数据分析挖掘基础.ppt<br>Microsoft PowerPoint 97-2003...<br>5.82 MB | 第2章 大数据基础原理.ppt<br>Microsoft PowerPoint 97-2003...<br>4.08 MB    | 第3章 大数据挖掘建模平台简介.ppt<br>Microsoft PowerPoint 97-2003...        |
| 第4章 数据探索.ppt<br>Microsoft PowerPoint 97-2003...<br>2.23 MB      | 第5章 数据预处理.ppt<br>Microsoft PowerPoint 97-2003...<br>2.10 MB      | 第6章 挖掘建模.ppt<br>Microsoft PowerPoint 97-2003...<br>1.39 MB    |
| 第7章 电力窃漏电用户自动识别.ppt<br>Microsoft PowerPoint 97-2003...          | 第8章 航空公司客户价值分析.ppt<br>Microsoft PowerPoint 97-2003...<br>2.34 MB | 第9章 中医证型的关联规则挖掘.ppt<br>Microsoft PowerPoint 97-2003...        |
| 第10章 基于水色图像的水质评价.ppt<br>Microsoft PowerPoint 97-2003...         | 第11章 电商平台家电设备评论信息大数据分析.ppt<br>Microsoft PowerPoint 97-2003...    | 第12章 家用电器用户行为分析及事件识别.ppt<br>Microsoft PowerPoint 97-2003...   |
| 第13章 基于基站定位数据的商圈分析.ppt<br>Microsoft PowerPoint 97-2003...       | 第13章 家用电器用户行为分析及事件识别.ppt<br>Microsoft PowerPoint 97-2003...      | 第14章 气象与输电线路缺陷关联分析.ppt<br>Microsoft PowerPoint 97-2003...     |
| 第15章 基于关联规则的网站智能推荐服务.ppt<br>Microsoft PowerPoint 97-2003...     | 第16章 基于协同过滤的互联网电影智能推荐.ppt<br>Microsoft PowerPoint 97-2003...     | 第17章 基于Hadoop的数据挖掘二次开发.ppt<br>Microsoft PowerPoint 97-2003... |



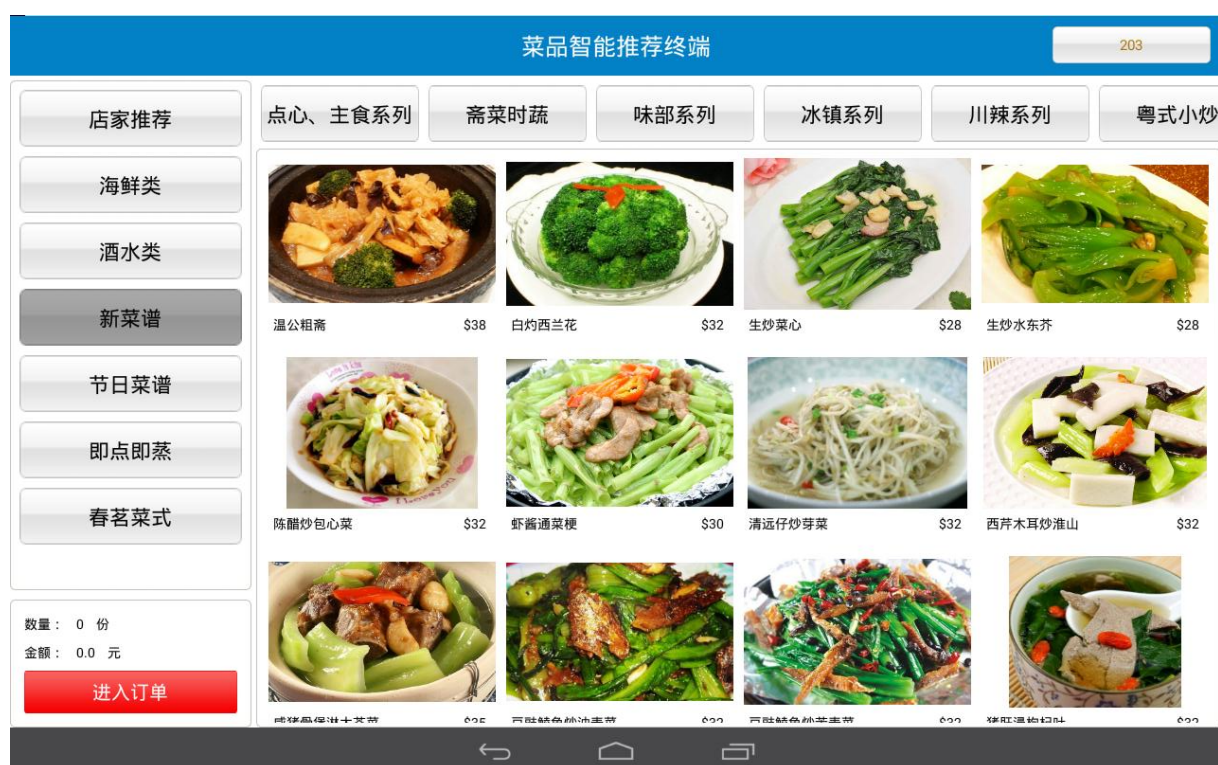


## 2.5 大数据挖掘教学沙盘

大数据挖掘教学沙盘能够让老师与学生在实验室模拟案例现场真实数据采集过程，在讲解数据挖掘算法理论的同时，通过实用的建模工具来模拟实际应用案例，在实验室环境下体验真实应用，由此弥补学员实践经验的不足。

“沙盘模拟”实验课程改变传统的教学模式，将理论与实践融为一体的设计思路，使学生在参与、体验中完成从知识到技能的转化，这种体验式教学方式完全不同于传统理论教学及案例教学，是教学方式的一大创新。沙盘模拟过程中，学生经历了一次从理论到实践再到理论的上升过程，实现从感性到理性的飞跃。在全面深刻理解分析设计、数据取样、数据清洗、模型构建、推断思考等数据挖掘建模问题。下面以餐饮菜品智能推荐沙盘为例展开说明：

餐饮菜品智能推荐教学沙盘，包括硬件与软件两部分。硬件部分为服务器和平板终端；软件部分主要为会员管理、菜品管理、销售管理及菜品智能推荐等模块，有关界面如下图所示。



确认点菜信息



名称：波士顿龙虾

单价：388

数量：- 1 +

需求：

取消选择
加入订单

203-订单详情						
图片	菜名	单价	份额		备注	删除
	波士顿龙虾	388	-	1 +		
	金牌烧鹅/半只	98.0	-	1 +		
	铁板酸菜豆腐	38.0	-	1 +		
	沙虾/斤	58.0	-	1 +		
菜品数量：4 份						
订单总额：582.0 元		<span style="background-color: #ff0000; color: white; padding: 5px 15px; font-weight: bold;">提交订单</span>				

## 2.6 大数据整合平台

TipDM-I2 基于 J2EE、Web Services、Internet 等技术，具有很强的跨平台性，部署简便，维护简单，容易使用。TipDM-I2 作为一套优秀的企业级数据整合和优化工具，适合于企业快速构建数



据中心，是企业快速实施基础数据平台的工具。在实施过程中进行开发过程的规范化、代码自动化、代码和质量可控化、调度管理统一化、监控可视化。也可与第三方 ETL 工具（如 Datastage）实现互补，增强其开发过程自动化、统一调度、全局监控等功能。

TipDM-I2 具有灵活的调度机制，可维护性强；支持可视化的多角度监控，拓展性强；提供负载均衡特性；提供二次开发接口，与第三方系统集成。TipDM-I2 主要有数据管理、运维视图、接口管理、调度管理和系统管理等功能，具体如下表：

功能	说明
数据管理	主要包括数据采集、数据分发、数据抽取、数据清洗、数据转换和数据装载等
运维视图	主要包括报表视图、数据采集分发视图、ETL 视图、硬件资源视图和接口热点视图等
接口管理	主要包括数据访问 web service 接口、数据访问 socket 接口、数据访问 HTTP 接口、数据导出 HTTP 界面接口、访问验证服务、资源访问控制和数据查询服务等
调度管理	主要包括作业定义、作业预处理、作业命令发送、作业命令接受、作业唤起
系统管理	主要包括用户管理、角色管理、用户角色管理、控制点管理、权限管理、业务参数管理、硬件资源管理、业务日志查询、数据导出模板管理和告警信息推送等

## 2.7 课程规划及师资培训

为适应大数据时代及数据挖掘实践教学需求，泰迪科技可参与到高校教师队伍建设，协助完成课程设置及教学规划，并为数据挖掘课程教学提供一对一的师资培训，培训主要包括：大数据挖掘案例培训；基于Hadoop框架的大数据开发培训。

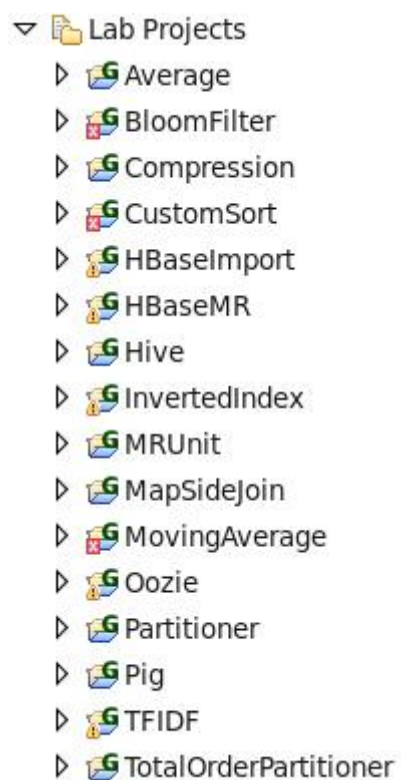
基于Hadoop框架的大数据开发培训使用TipDM提供的培训资料，帮助开发者了解Hadoop平台，理解开发Hadoop程序的基本原理，最后可以使培训者能自主开发Hadoop程序，同时进行培训后，学员可以有更大的机会通过HortonWorks开发者认证，其认证在业界是比较有含金量的。

课程安排内容如下：

大纲	课程
Hadoop基础	培训准备：环境配置 单元1：理解Hadoop2.0和HDFS 单元2：开发MapReduce程序
Hadoop进阶	单元3：Map端流程—聚合 单元4：分区和排序
Hadoop进阶	单元5：Hadoop输入输出格式 单元6：优化MapReduce任务

	单元7：高级MapReduce特性
Hadoop应用	单元8：Hadoop单元测试编程 单元9：HBase编程 单元10：Pig编程
Hadoop应用	单元11：Hive编程 附录A：Oozie工作流 其他：交流

与课程相关的代码资源：



## 2.8 学生实习实训

结合课程教学需要，可组织学生到泰迪公司进行集中实训或实习，实习项目为从在建的企业项目中剥离出部分大数据应用子课题，让学生真切参与到企业大数据项目建设中，实习实训包括校内和校外（在泰迪公司现场）两种模式。

### 3 产品优势

H8与其它同类产品相比，具有如下明显优势：

- ✧ 提供完整的Hadoop生态链产品，同时包含HDFS、Yarn、Hive、HBASE、Zookeeper等组件；
- ✧ 提供作业调度、作业监控和作业管理等大数据作业调度系统；
- ✧ 提供大数据分析平台监控系统，能实时监控大数据分析平台的运行指标；
- ✧ 提供数十个基于Mahout、Spark开发的分布式算法；
- ✧ 提供大数据开发实训平台，使用学生能真正完成基于Hadoop框架的大数据应用开发；
- ✧ 自带数据导入模块，可导入结构化和非结构化数据进行分析。

### 4、应用案例

#### 4.1 主要客户

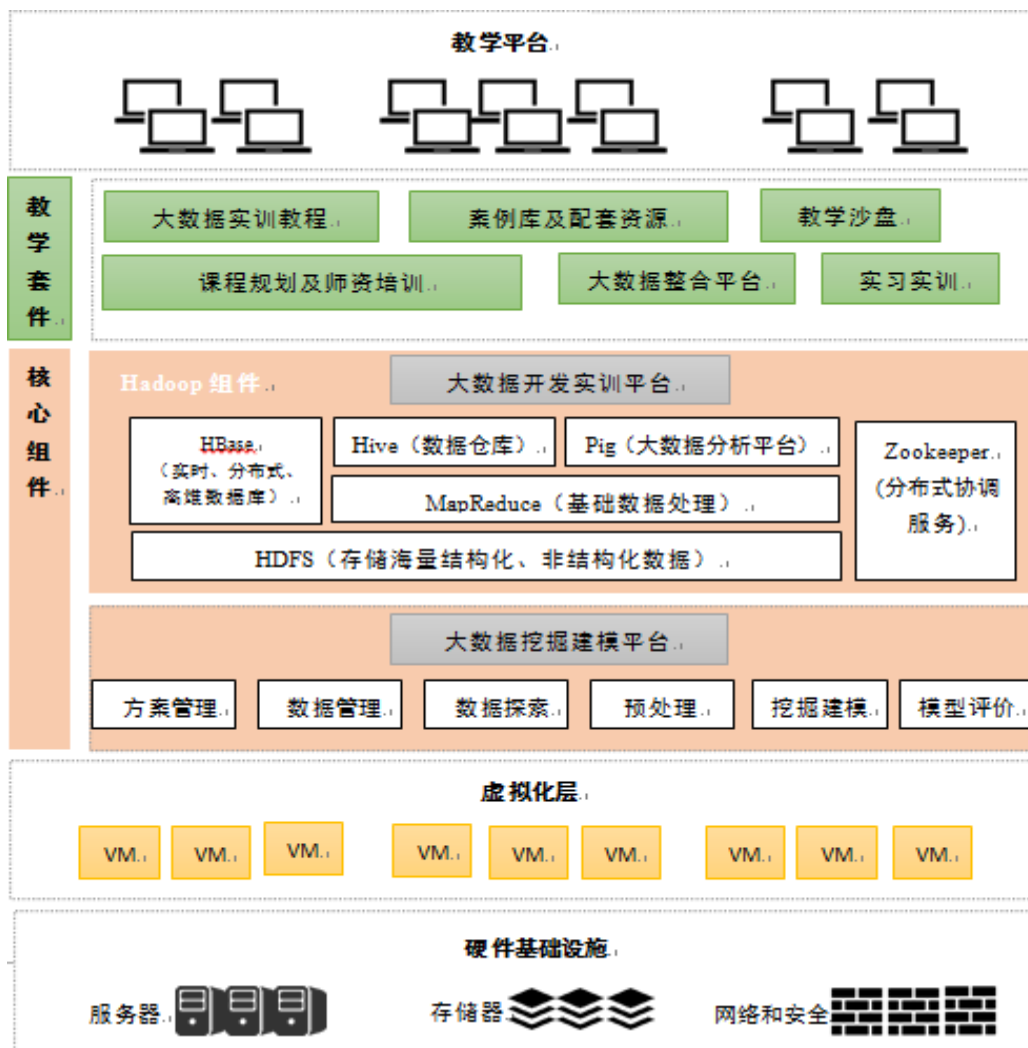
- ✧ 韩山师范学院
- ✧ 广东技术师范学院
- ✧ 北京城市学院

#### 4.2 典型示例

2014年，广东某高校数学与统计学院意识到自身面临师资资源不足、实验条件缺乏、实习实践条件缺失，导致学生就业困难，于是决定对症下药，与泰迪科技合作，共建大数据工程分析实验室。

实验室建设总共60个实验机位，要求每个学生至少拥有1个独立主机的实验环境，以完成大数据挖掘建模分析、基于Hadoop的大数据开发实训。

## 4.2.1 实验室架构



## 4.2.2 设备清单

按60人规模配置。

1、台式计算机（学生用机） 联想 扬天 R系列（60台）

CPU: Intel 酷睿 i5

内存容量: 4GB DDR3 1600MHz

硬盘容量: 1TB

2、服务器 IBM x3650 M4（16台）

具体配置见 4.3.1

3、液晶投影仪 明基 MP782ST（2台）

2500 流明以上

分辨率: 1280\*800

#### 4、网络交换机 千兆以太网交换机 华为 S5700-24TP-SI (1 个)

应用层级：三层

传输速率：10/100/1000Mbps

背板带宽：256Gbps

包转发率：36Mpps

### 4.2.3 环境部署

使用10台实体服务器分别安装Windows7 64位操作系统或者Linux (Redhat 6、CentOS 6都可) 64位操作系统，每台机器安装Vmware软件用于部署培训使用虚拟机。

每台实体服务器建立6个虚拟机，其虚拟机参数设置如下：

硬件	要求	备注
服务器数量	7（集群）+1（数据库服务器）+1（应用服务器）	
CPU	厂家：Intel/AMD 主频：3000MHz+ 核数：12+ L3 缓存：15MB+ 线程数量：12+	针对所有服务器
内存	16G+（集群） 8G+（数据库服务器） 8G+（应用服务器）	最少 8G
硬盘	SATA/SAS 1T+200G，2 个磁盘驱动器，每个 500G+2 个 磁盘驱动器每个 100G（用作系统盘）（集群） 1T+200G（数据库服务器） 300G+200G（应用服务器）	最少 50G
网络	千兆以太网	固定 IP
宿主机	支持虚拟化/64 位系统	
系统	CentOS6.4 64 位	可为虚拟机系统

说明：若需保持60个客户端同时可操作，则每台实体机开启6个虚拟机，这时需要客户端安装Vmware相关软件保证可以远程桌面到虚拟机。

### 4.3 虚拟机资源

虚拟机资源列表如下：

软件	版本	备注
系统	CentOS6.4	64 位
TipDM-HD Hadoop 平台	HDP2.0	包括服务： HDFS Yarn

		MapReduce Hive HBase Pig Ganglia Nagios ZooKeeper
Ambari	1.5	HDP 调度软件
Eclipse	4.3	Java 开发 IDE
Java	1.6	
数据		包含培训使用各种数据

HD 平台包括的服务有：



#### 4.3.1 硬件配置及软件部署

硬件清单：

硬件使用 16 台 IBM 机器，分为两部分：1) 大数据挖掘建模平台（TipDM-HB）使用 6 台；2) 大数据开发实训平台（TipDM-HD）使用 10 台。其主要硬件要求以及说明如下表。

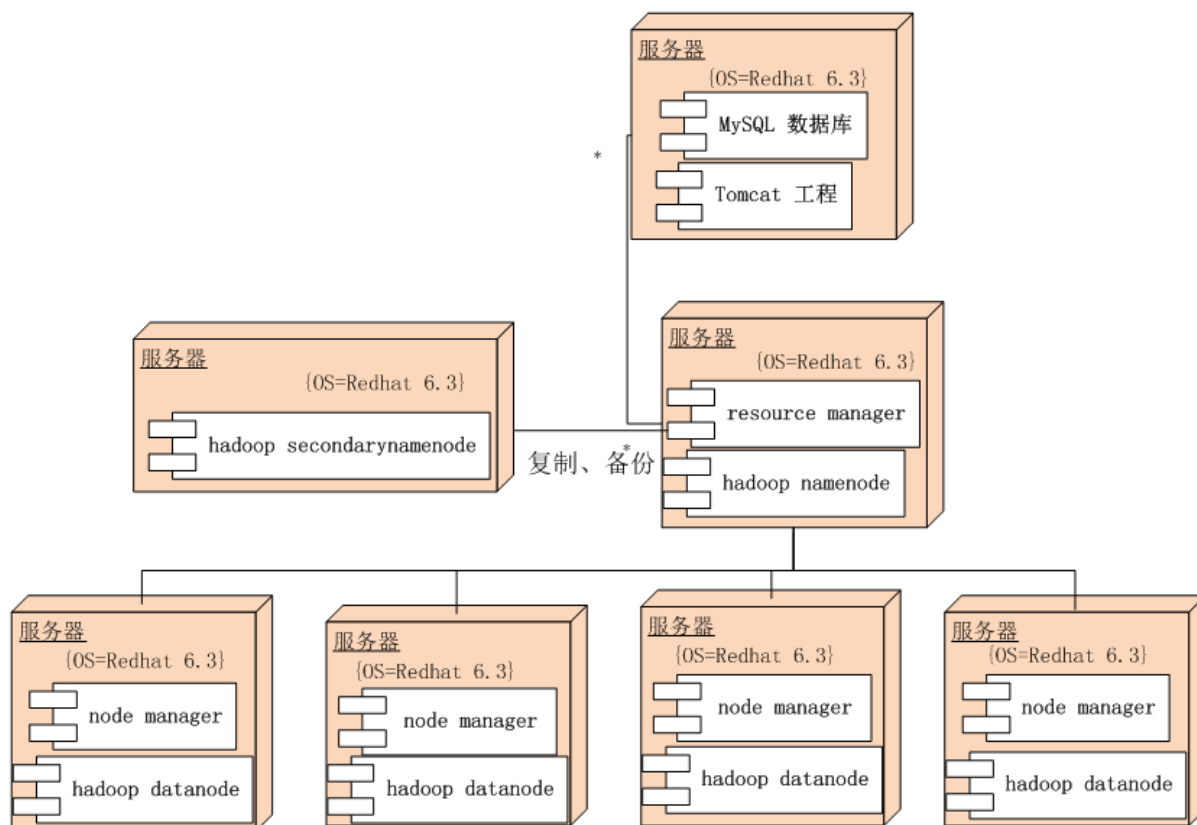
服务器名称	服务器型号	服务器数量	CPU	内存	硬盘	备注
IBM 机	x3650 M4	6	Xeon E5-2650	每 16G/个	SAS 硬盘/2 个 每 200G 个	TipDM-HB 使用
IBM 机	X3650 M4	10	Xeon E5-2650	每 60G/个	SAS 硬盘/2 个 每 500G 个	TipDM-HD 培训使用/支持虚拟化



以上硬件配置供参考，同类配置亦可。

### 大数据挖掘建模平台（TipDM-HB）配置：

部署 TipDM-HB 共 6 台机器，其中云平台占用 5 台机器，数据挖掘平台主工程部署在 1 台机器，其拓扑图如下所示：



### 大数据开发实训平台（TipDM-HD）机器配置：

大数据开发实训平台机器配置主要是指针对学员使用的环境进行配置。初步估计 60 个客户端接入，那么 10 台 IBM 机器每台配置 6 个虚拟机即可。同时需保证每台虚拟机占有专属的 IP，每个客户端可以远程桌面到虚拟机。