

Neuedu



东软-大数据方向项目综合培训方案- 适用于实训三



一、实训对象

广西科技大学启迪数字学院 2018 级学生

二、实训周期

2021 年 7 月起至 2022 年 1 月止（具体开始时间视放假情况决定）

三、实训地点

东软广州软件园（地址：广州市黄埔区科学城揽月路 8 号，近地铁 21 号线科学城站（距离 500 米）），是目前东软在华南地区集软件研发、服务外包的研发基地，也是大学生实训和软件人才培养为一体的人才培养基地；可容纳 2000 多人同时进行培训，每年可为 10000 多名大学生和软件人才提供实训，具备国家级工程实践教育中心资质。



四、实训费用支付方式

每位同学所支付的费用应是公司报价减去优惠费用

公司报价：20800 元；

缴费方式：

①一次性缴费，费用 20800 元减去优惠费用；

②贷款缴费：以费用 20800 元为例，按照“8+16”的分期支付模式，即前 8 个月（付利息 156 元）+后 16 个月（付本金+利息，1518.4 元）的方式支付，注：可提前还款，提前还款则不需要支付还款后续期数的利息；

③可部分一次性缴费+剩余费用贷款缴费的方式；

注：请关注蓝色字体区（优惠政策）

在上述报价的基础上，有如下优惠：

优惠政策一、该报价含学院的补贴 2000 元；

优惠政策二、通过大学英语四级学生-立减 300 元

通过大学英语六级学生-立减 500 元

优惠政策三、校级奖学金（一、二、三等奖）获得者：立减 300 元

国家奖学金或国家励志奖学金或社会捐赠类奖学金或定向奖学金获得者：立减 500 元

特别提醒：

①第一项中补贴为直接减免，每人均享受该项政策；

②第二项中四六级证书均以 425 分以上获减免资格；

③上述每项优惠政策只以所获最高奖项做免额计算；该三项优惠政策可以累加；每位同学所支费用应为报价减去优惠费用

④举例：A 同学（大学英语六级考试 500 分、国家励志奖学金获得者）则需付培训费为 17800 元（即 20800 减 2000 减 500 元减 500 元）

五、实训目标

5.1 培养目标

- ✧ 掌握 JavaSE、JavaEE 体系所有开发技术，掌握 OOAD 思想、编码规范，具备开发 Java 企业级项目的能力；
- ✧ 掌握 SQL、NoSQL 数据库开发技术，具备 Oracle、MySQL、Redis 数据库的使用与设计能力；
- ✧ 掌握 LINUX、HDFS、YARN、MR、HADOOP 等数据架构技术，具备大数据平台开发能力；
- ✧ 精通 HBASE、HIVE、HUE、PIG、FLUME、SQOOP、ZOOKEEPER 开发框架技术，具备提高项目开发效率与软件质量的能力；
- ✧ 掌握 SCALA, SPARK, KAFKA 等技术的使用，具备企业项目开发工程过程管理工具使用能力；
- ✧ 跟随主流技术的需求动态调整实训内容
- ✧ 具备项目需求理解、文档编写、模块化思维、单元测试等项目工程能力；
- ✧ 具备职业素养与就业能力，良好的学习能力、沟通能力、解决问题能力；

5.2 就业岗位

Java 开发工程师、大数据开发工程师、Hadoop 工程师、大数据分析工程师、JavaWeb 开发工程师、大数据运维工程师、数据库工程师、测试工程师

六、实训环境

6.1 教学环境

教室：中央空调，投影仪，白板，麦克及音箱，电脑；

6.2 住宿环境

6 人/间(一间公寓 50-60m²，独立洗手间，独立阳台，提供空调、热水器)；

住宿方面需要收取 400 元/月/人的住宿费，水电费舍友平摊。

6.3 配套设施

食：内有食堂，周边有其他商铺；

行：周边公交配套完整；

购物：园区内有多家超市；周边有多家小卖部；

取钱：园区内有 ATM 机；

活动：园区内为学员提供乒乓球，足球，篮球，排球，羽毛球等活动场所。同时园区内有健身活动中心供学员选择使用；

【学习环境】广州东软软件园实景图



【宿舍所在园区机房图片】



【宿舍】宿舍场景



【食堂】食堂场景



七、实训教学模式与就业保证

7.1 教学模式

①真人教学，非视频教学；

②教育理念：培养体系完全基于“产业融入教育”经营理念，采用产学研融合式人才培养模式。人才培养体系完全基于东软 20 多年产业实践经验进行设定，引入东软在互联网、金融、政府、医疗、大数据、汽车电子等行业的真实产业项目，将产业项目转化为教学案例和实训项目，并合东软的 CMMI、RUP、敏捷精益软件开发方法学，形成特有的产业项目驱动教学的教学模式，真正的培养出符合软件开发企业所需技术标准的人才。

③师资团队介绍：公司的教学队伍有百余人，均具备项目组实际开发经验、具备很强的实操性，在职教师融合了东软教育体系的教学理念和集团项目实际开发应用的实践性，另东软在研究和开发领域，成立东软研究院，致力于课程体系打造与融合、新技术的学习和推广等；

④实训平台（免费提供给每位在司培训学员）：将企业真实软件项目开发全流程信息化管理与高校实训教学管理相融合的信息化平台，教师可便捷地进行教学设置，包括项目分组、项目阶段设置、学习任务安排、开发任务分配、互评设置、组长权限设置、日报时间设置等功能；系统集成了 Gitlab 进行版本管理，为每个项目组分配专属代码仓库。系统可从代码量、提交时间、贡献者等不同维度统计学生项目代码贡献度（代码行）。下图是，该平台上创建的实训项目示例。

[全部项目](#) [我创建的项目](#) [新建项目](#)

自建

东软交通时空大数据分析
技术方向: JavaWeb方向 项目等级: ★
创建人: neuedutea01 项目周期: 2周
模块数: 0 任务数: 0

完整

2020丹尼测试项目
技术方向: JavaWeb方向 项目等级: ★
创建人: neuedutea01 项目周期: 1周
模块数: 2 任务数: 2

完整

测试项目
技术方向: JavaWeb方向 项目等级: ★★
创建人: neuedutea01 项目周期: 2周
模块数: 1 任务数: 1

自建

55555
技术方向: JavaWeb方向 项目等级: ★

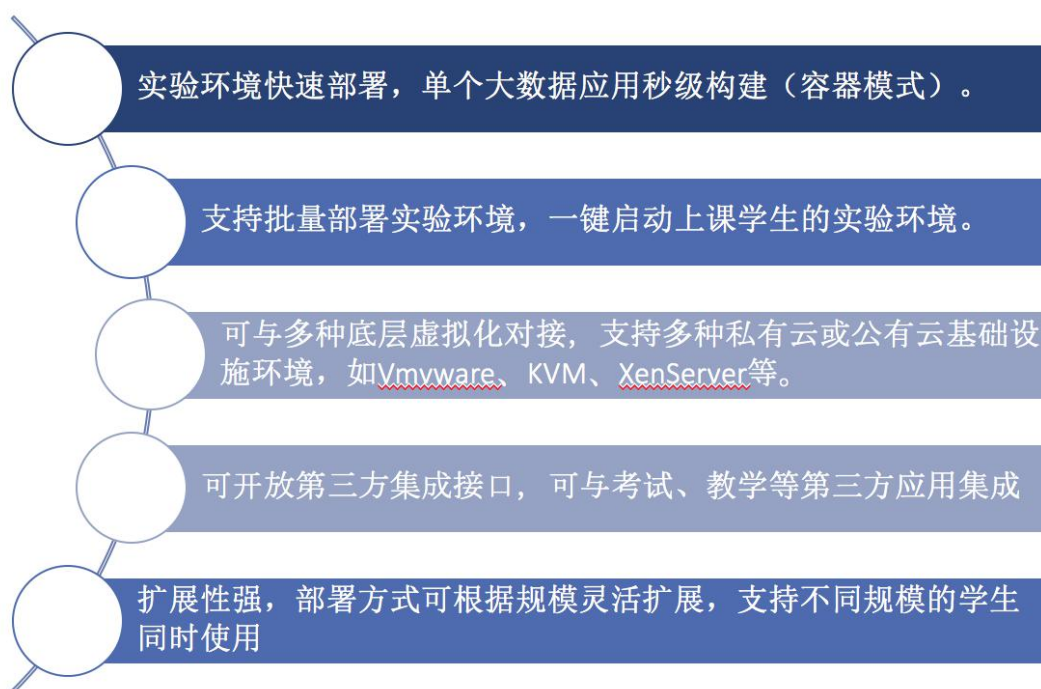
自建

测试项目
技术方向: JavaWeb方向 项目等级: ★
创建人: neuedutea01 项目周期: 1周
模块数: 0 任务数: 0
适合人数: 1-1
创建时间: 2020-03-19

完整

HTML5 网页制作
pythonweb后端
技术方向: PythonWeb... 项目等级: ★★★

核心优势：



7.2 就业方面

①公司拥有完整而严密的就业推荐管理体系

公司具备专业的就业团队，学生自参加培训起即可加入东软人才储备资源库，可享有就业推荐的机会，并将持续跟踪学员3年的就业情况，提供更优质的服务，为学员在每个重要的发展阶段提供支持力量；

就业团队职责：

I 负责东软集团各项目组人员岗位情况整理（学生可优先入职东软各项目组）；

II 负责引进优质企业入司招聘；

III 负责学生职业素质提升，如商务礼仪、简历制作、面试技巧等。

②就业联动机制

公司在全国有7个区域总部，各区域均有就业团队，在学员就业时，可打破地域限制，按学员意愿推动就业，保障就业工作顺畅进行。

③具体实施情况（招聘会现场）和实际就业案例（广西科技大学学生就业情况）



方欣科技招聘现场



宇信科技招聘现场



东华信息招聘现场



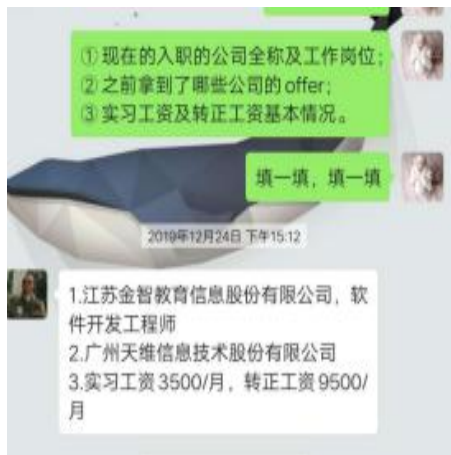
东软集团招聘现场



深圳光启招聘现场



神州泰岳招聘现场



创新教育 源动中国

互联网+时代高校创新信息服务提供商

金智教育为全国800余所高校提供信息化平台建设服务

邀 请 函

尊敬的 李海枫 先生:

欢迎您加入金智教育大家庭, 共同实现“创新教育、源动中国”的发展蓝图。

经公司与您的沟通, 双方已达成合作意向。具体如下:

所在岗位: 软件开发工程师

工作地点: 深圳

岗位薪酬:

(1) 月度基本薪酬: 试用期 8500 元, 转正 9500 元;

- 公司根据每年发展情况、相关规章制度以及您个人的表现情况, 对薪酬标准进行相应的调整;

- 福利按照公司相关标准执行

- 在您拿到毕业证之前, 我们将统一安排实习, 实习期薪资为 3500 元, 其中包含 1000 元实

绩效考核。实习时间将由人力资源部统一安排

(2) 公司将根据您的岗位所对应的绩效管理办法, 结合岗位绩效考核结果执行相关绩效奖金的计

(3) 工资属公司高度机密, 请您严格履行薪酬保密承诺。

五险一金、个税:

公司按照相关规定为您代扣代缴个人所得税及社会保险、住房公积金个人应缴纳的部分。

劳动合同: 首期劳动合同约定为 4 年, 试用期 3 个月。

正式入职日期: 2020 年 7 月 1 日。

本邀请函的有效期限为发出后 3 个工作日。请您在收到后, 在有效期内以书面或邮件

方式与我们确认, 我公司将为您办理必要的入职前准备工作。真诚期待您的加盟!

相关说明请见附件, 若有疑问请联系我们。

安硕信息录用通知书

尊敬的赵女士:

您好!

感谢您对我公司的信任和大力支持! 非常荣幸的通知您, 您已顺利通过我司录用评估, 并与公司达成三方签约意向, 我们欢迎您的加入。

一、职位及工作地点:

职位: JAVA 工作地点: 深圳

二、薪资待遇:

实习期:	正式入职:
实习薪酬: 3000 元/月	合同期限壹年, 其中含试用期为叁个月
享有公司实习生相应福利。	试用期基本工资为: 6000 元/月
	转正后基本工资为: 7500 元/月
	另享有补贴, 补贴以 HR 披露为准。

陆*荣 同学:

祝贺您通过烽火星空2021届校园招聘*聘。

欢迎加入烽火星空，挑战PB海量数据挑战，助力网络安全、大数据及AI的未来！

无尽高峰，邀你攀登，并肩前行，开启网络安全大数据新视界

如您签约本公司，视为同意以下声明：您提交的个人资料全部属实，如有不实信息，属于严重违反诚信原则，

则，公司有权随时拒绝录用且不承担任何责任。

请仔细阅读：

一、JD、薪酬福利

1、JD：Java软件工程师，城市：广州市，转正后税前薪酬 9000 元/月，试用期按照80%发放。

2、缴纳五险一金、补充医疗保险。

3、根据国资委关于中央企业管理规定，享有的福利，包括但不限于：

- a、base南京：按出勤餐补13元/天+餐补15元/天，自有餐厅。
- b、base其他城市：依城市不同，转正后按出勤餐补35-40元/天。
- c、差旅：交通住宿依据标准实报实销，差旅补贴从80元/天到100元/天不等。
- d、高原和边远地区：高原和边远地区补贴。
- e、体检关怀、工会活动经费、重大节假日礼品。
- f、生日、结婚、生育礼金券或慰问金。

广西科技大学部分就业案例（注：表格中月薪以就业时实习和就业薪资来统计）

学院	姓名	公司	年级	月薪
启迪数字学院	陆*荣	烽火众智	2017 级	9k（转正，如上截图）
	吴*	腾讯瑞德铭	2017 级	5000（实习）
	李*超	金智教育	2017 级	8k+补贴（转正）
	杨*婷	宇信科技	2017 级	8.4k（转正）
	赵*莲	安硕	2017 级	7.5k（转正）
	赵*岸	碧桂园总部	2016 级	7k+补贴（转正）
	黄*	易诚互动	2016 级	3.5k(实习)
	李*枫	江苏金智教育信息股份有限公司	2016 级	9.5k（转正）
计通学院	宁*思	广州市云景信息科技有限公司	2017 级	3.5k（实习）
	李*辉	广东海聊科技有限公司	2017 级	3.5k（实习）
	张*金	深圳联友科技	2017 级	8.2k+补贴（转正）
	罗*杰	烽火众智	2017 级	8k+补贴（转正）
	黄*欣	广州市蓝岭信息技术有限公司	2017 级	4.5k（实习）
	覃*发	东软	2017 级	4k（实习）
	杨*程	神州信息	2016 级	7.5k（转正）
	梁*钰	泛微软件	2016 级	3.5k（实习）
	曾*发	玉柴	2015 级	8k（转正）
理学院	许*杰	东软医疗	2017 级	3.5k（实习）
	卢*杰	浪潮国际	2017 级	6.5k+补贴（转正）
	郑*钰	宏景科技	2017 级	3.5k+补贴（实习）

	黄*欢	麦迪斯顿医疗科技股份有限公司	2016 级	7.5k+补贴（转正）
	韦*明	望海康信（北京）科技股份公司	2016 级	7k+补贴（转正）
	李*瑶	金蝶医疗	2016 级	3.5k（实习）
	阮*芸	优诺科技	2015 级	8.5k（转正）

八、实训纪律与成绩评定

在学员的管理方面，严格按照企业员工的管理规范进行，采用科学、多样的管理方法，营造出紧张有序的工作氛围，并按照项目组管理方式对学员进行真实项目的管理。

东软睿道教育拥有国内领先的全流程实训质量保证及过程改善体系，整个实训过程严格依据标准进行实施。

通过考试了解学员的能力水平，形成具有针对性的培养计划，并按照严谨的教学计划进行理论结合实践的培训和模拟项目开发，合格学员将按照标准流程向优秀企业推荐录用。每个实训活动都有标准化的策划、筹备、实施和总结活动，在各阶段中还有严格的里程碑管理、例会等项目管理活动，以确保学员在实训活动中得到真正的能力提升。

针对教学内容，以项目组为单位对实训学员进行技术考核，得出技术分数，结合整个实训过程中的出勤情况、教师评价等综合评定，得出学员的实训成绩，汇总得出《实训成绩报告》，成绩合格者颁发《东软实训证书》。

考核成绩=理论成绩*40%+项目成绩*60%-违纪扣分+特别加分

其中项目成绩=团队成绩*20%+个人成绩*80%

项目考核内容：

专业能力						行为能力					评语
编 码 能 力 (30%)	详 细 设 计 (25%)	开 发 工 具 与 调 试 技 巧 (20%)	单 体 测 试 (10%)	项 目 文 档 撰 写 (10%)	配 置 管 理 (5%)	沟 通 能 力 (20%)	学 习 能 力 (20%)	团 队 合 作 (20%)	主 动 性/责 任 心 (20%)	问 题 判 断 与 解 决 (20%)	

违纪扣分：

针对学员出勤情况及在实训期间出现的违反公司制度情况而设置。

特别加分：

针对学员在实训期间表现出的卓越的职业化素养并造成一定影响力的行为和事件而设置。如技术创新性、责任感、团队合作、提出合理化建议等等。

九、实训课程

9.1 学习路径

详情请见章节十 实训安排

9.2 部分实训项目展示（注：仅列举部分案例）

东软内训 OJT 大数据解决方案是面向大数据存储、处理、展现全环节、一体化的项目产品。

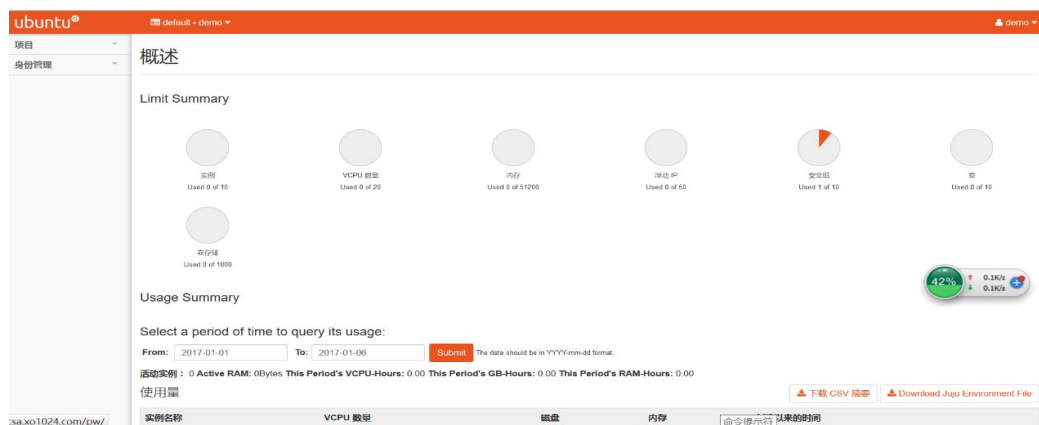
总体结构如下图：



通过睿道大数据集群软件对底层计算资源、存储资源、网络资源进行统一的管理，为学生分配大数据教学使用的资源，提供的功能如下：

- 用户管理模块：

高校教师可以以班级为单位，为每个班级设置项目，在项目中设定可以使用的资源，包括存储资源、计算资源、网络资源；为每个班级分配一个账号，每个班级设置一个学生管理员，由该学生管理员登录本班级的账号对项目中的资源进行分配，创建教学使用的云主机。在教学开始及教学结束后，及时的开启和关闭云主机，以保证云资源的共享。



- **权限管理模块：**

为每个班级的管理员设定相应的权限，便于进行本班级资源的统一管理。

- **角色管理模块：**

在系统中，存在各种操作权限，为便于权限的管理，可以通过设定角色来简化管理。

- **计算管理模块：**

系统可以规划并管理大量虚拟机，每个班级的管理员可以按照预先分配的总的计算资源，根据班级学员的数量及实际教学的需要为每个学员批量分配计算资源，从而按需提供计算资源；

The screenshot shows the 'Instances' (实例) page of the Ubuntu Cloud dashboard. It features a table with columns: 实例名称 (Instance Name), 映像名称 (Image Name), IP 地址 (IP Address), 大小 (Size), 密钥对 (Key Pair), 状态 (Status), 可用域 (Availability Zone), 任务 (Task), 电源状态 (Power State), 创建以来的时间 (Time since creation), and Actions. The table lists four instances: slave3, slave2, slave1, and master. A sidebar on the left contains navigation links like 项目 (Project), 管理员 (Admin), 系统 (System), 概述 (Overview), 管理程序 (Hypervisor), 主机集群 (Host Cluster), 实例 (Instances), 卷 (Volumes), 快照 (Snapshots), 网络 (Network), 路由表 (Routing Table), 默认值 (Defaults), 元数据定义 (Metadata Definitions), 系统信息 (System Information), and 身份管理 (Identity Management).

实例名称	映像名称	IP 地址	大小	密钥对	状态	可用域	任务	电源状态	创建以来的时间	Actions
slave3	hadoop模板	10.25.57.202	ubuntu16.04	-	活动	nova	无	正在运行	1 日, 20 小时	创建实例
slave2	hadoop模板	10.25.57.201	ubuntu16.04	-	活动	nova	无	正在运行	1 日, 20 小时	创建实例
slave1	hadoop模板	10.25.57.200	ubuntu16.04	-	活动	nova	无	正在运行	1 日, 21 小时	创建实例
master	ubuntu16.04_snap	10.25.57.199	ubuntu16.04	-	活动	nova	无	正在运行	2 日	创建实例

- **存储管理模块：**

大数据集群可以为云应用提供所需的存储资源，考虑到性能及价格需求，可以根据实际需要可提供可配置的存储功能；

创建卷

卷名称

vol3

描述

卷来源

映像

使用映像作为源

cirros (12.7 MB)

类型

无卷类型

大小 (GiB)

1

可用域

nova

描述:

卷是可连接至实例的块设备。

卷类型描述:

如果选择“无卷类型”，那么将创建没有卷类型的卷。

卷限度

总大小 (GiB) (3 GiB)

1000 GiB 可用

卷数量 (2)

10 可用

取消

创建卷

● 镜像管理模块:

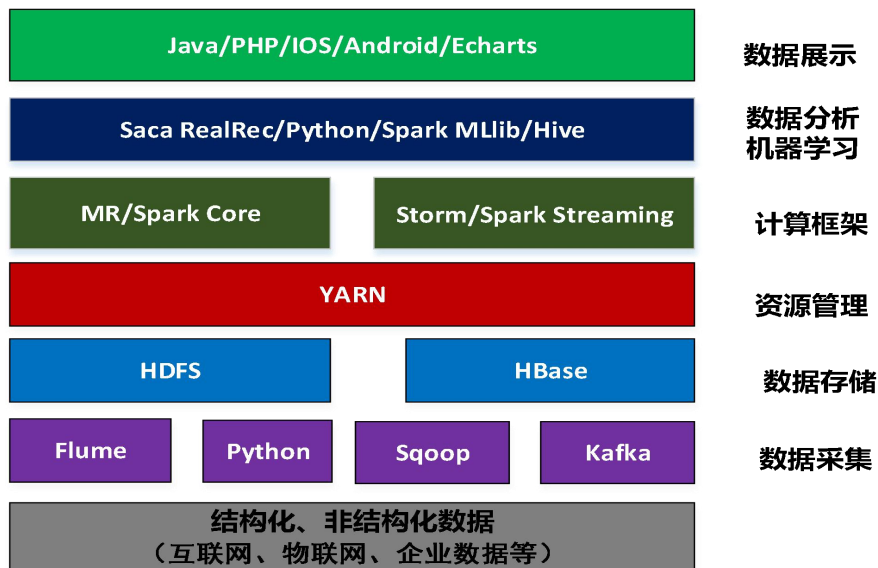
为了能够快速构建云主机及hadoop集群环境，在镜像管理模块中，预先定义了大量的镜像资源，包括linux操作系统镜像、hadoop 管理节点镜像、hadoop 数据节点镜像、hadoop hbase等等，使学员能够节省大量时间来进行编程方面的训练。

● Hadoop 集群构建模块

在教学过程中，大数据分布式软件系统安装、配置难度大，而且实验环境容易被学生破坏，Hadoop 集群构建模块可以快速的重新构建新的大数据实验环境，满足教学需要。

● 大数据教学产品

大数据是个开放的生态系统，包含的组件及工具更是多达上百种，东软根据实际的主流技术进行了筛选，在东软大数据平台中，进行了如下分层：



● 数据采集层

数据采集层包括 Flume、sqoop、kafka、Python 爬虫，通过这些工具将大量 web 页面数据、日志数据或关系型数据库中的数据采集到 hdfs 或 NoSQL 数据库中，为后续的大数据应用提供数据保证。

● 数据存储层

数据存储层包括 HDFS 和 Hbase，考虑到性价比及高吞吐的应用，将海量数据保存到 HDFS

分布式文件系统中；如果需要对数据进行随机读操作或者随机写操作，以及大数据上高并发操作，可以将数据保存到 Hbase 分布式数据库中。

HDFS 可以通过 web 接口进行统一管理监控：

Hadoop	Overview	Datanodes	Datanode Volume Failures	Snapshot	Startup Progress	Utilities
--------	----------	-----------	--------------------------	----------	------------------	-----------

Overview 'master:9000' (active)

Namespace:	testCluster
Namenode ID:	nn1
Started:	Fri Jun 02 08:31:51 CST 2017
Version:	2.7.2, rb165c4fe8a74265c792ce23f546c64604acf0e41
Compiled:	2016-01-26T00:08Z by jenkins from (detached from b165c4f)
Cluster ID:	CID-54ce317c-1b6b-4d80-bbc8-182467ae672b
Block Pool ID:	BP-707574761-10.25.57.199-1496286933350

Summary

Security is off.

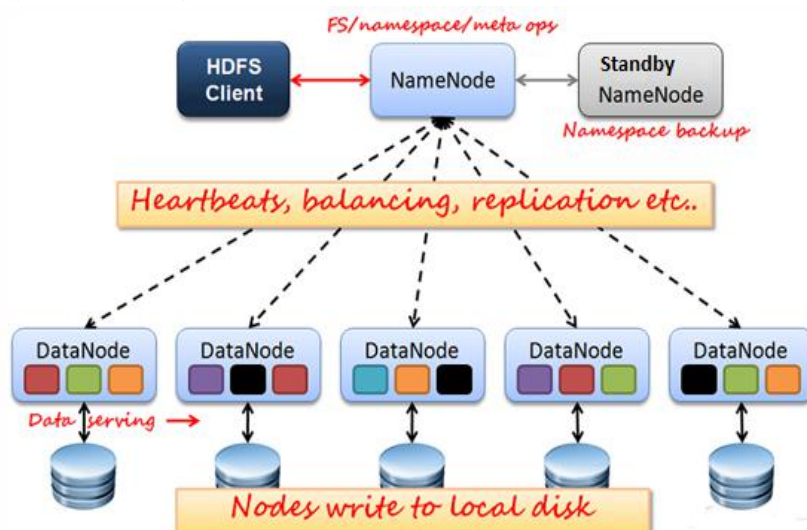
Safemode is off.

1 files and directories, 0 blocks = 1 total filesystem object(s).

Heap Memory used 38.49 MB of 50.83 MB Heap Memory. Max Heap Memory is 966.69 MB.

Non Heap Memory used 43.09 MB of 44.47 MB Committed Non Heap Memory. Max Non Heap Memory is -1 B.

通过运用 Zookeeper、Journalnode 等组件，确保 HDFS 的 NameNode 节点高可用。每个学员可以在分配的多个云主机下搭建 hadoop 高可用集群环境，在此基础上进行大数据的开发，完全模拟真实的企业环境。



- 资源管理层

随着集群资源的增加，以及多种计算框架（包括 MapReduce、spark 等）应用，为了便于资源（多台机器）在不同的计算集群中统一管理和调度，以充分利用资源，在资源管理层我们使用 YARN 进行集中管理。



All Applications

Logged in as: dr.who

Cluster

About

Nodes

Node Labels

Applications

NEW

NEW SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes
1	0	1	0	19	20 GB	24 GB	0 B	19	24	0	3	0	0	0	0

Scheduler Metrics

Scheduler Type		Scheduling Resource Type		Minimum Allocation		Maximum Allocation					
Capacity Scheduler		[MEMORY]		<memory:1024, vCores:1>		<memory:8192, vCores:8>					
Show 20 entries		Search:									
ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Progress	Tracking UI	Blacklisted Nodes
application_1496627918377_0001	hadoop	word count	MAPREDUCE	default	Mon Jun 5 16:17:15 +0800 2017	N/A	RUNNING	UNDEFINED	<div></div>	ApplicationMaster	0
Showing 1 to 1 of 1 entries											

- 计算框架层

计算框架层包含批量离线计算 MapReduce、内存计算 Spark 以及流式计算 Storm、Spark Streaming 等，根据不同的业务场景使用不同的计算框架，使每个学员掌握在大数据集群环境下不同计算框架的开发应用。

Retired Jobs

Show 20 entries Search:											
Submit Time	Start Time	Finish Time	Job ID	Name	User	Queue	State	Maps Total	Maps Completed	Reduces Total	Reduces Completed
2017.06.10 09:18:30 CST	2017.06.10 09:19:14 CST	2017.06.10 09:20:21 CST	job_1497057336012_0001	word count	hadoop	default	SUCCEEDED	28	28	1	1
2017.06.05 16:17:15 CST	2017.06.05 16:18:00 CST	2017.06.05 16:19:08 CST	job_1496627918377_0001	word count	hadoop	default	SUCCEEDED	28	28	1	1
2017.06.05 09:23:30 CST	2017.06.05 09:24:17 CST	2017.06.05 09:25:52 CST	job_1496622715752_0005	word count	hadoop	default	SUCCEEDED	28	28	1	1
2017.06.05 08:52:05 CST	2017.06.05 08:53:04 CST	2017.06.05 09:05:33 CST	job_1496622715752_0003	word count	hadoop	default	SUCCEEDED	28	28	1	1
Submit Time	Start Time	Finish Time	Job ID	Name	User	Queue	State	Maps Total	Maps Completed	Reduces Total	Reduces Completed
Showing 1 to 4 of 4 entries First Previous 1 Next Last											

- 数据分析与机器学习层

数据分析与机器学习层包括 Hive、Spark MLlib、Python 机器学习、SACA RealRec。

Hive 是构建在 Hadoop 之上的数据仓库，使用一种类似 SQL 的查询语言 HQL，一般用于进行离线的数据处理，进行简单的统计分析。

Python 机器学习和 Spark MLlib 是个可扩展的机器学习和数据挖掘库，提供了各种各样的算法，包括分类、回归、聚类、协同过滤等，通过运用这些算法来发掘海量数据的价值。

- 数据展示层

使用不同的工具，比如图表秀、Echarts、Java 图表、android 相关插件等，将大数据分析的结果通过柱状图、饼状图、分布图、折线图等多种图表的方式进行数据的展示。



十、实训进度安排

授课时长：4.5 个月，外加毕设辅导时长半个月，故该课程预计实施 5 个月（遇国家法定节假日则顺延）

上课时间安排：

	周一	周二	周三	周四	周五	周六	周日
上午 8.30-11.30	上课	上课	上课	上课	上课	自习	休息
下午 13.30-17.30	上课	上课	上课	上课	上课		
晚上 18.30-21 点	自习	自习	自习	自习	休息	休息	

课程安排（授课天数仅为工作日，不含周六日）：

序号	内容类型	课程名称	课程概述	天数	先导课程
1	专业基础课	linux 基础	本课程内容主要包括：linux 安装、linux 基本命令、文本编辑工具、用户与用户组管理、文件与目录管理、linux 网络基础、文件安装、文件系统、磁盘管理、iptables 防火墙、raid 技术、shell 编程、字符串匹配、正则表达式、Linux 备份恢复、SAN 与 NAS 存储、LINUX 通过 iSCSI 挂接存储、linux 系统监测与维护等等	6	
2	专业基础课	linux 配套服务课程	APT 服务器配置、NTP 服务器配置、SSH 免密码访问、VNC 访问、DHCP 服务器配置、NFS 服务器配置、ftp 服务器配置、Apache 服务器配置、rsync 数据复制、DNS 服务	4	linux 基础

3	专业基础课	云计算基础	介绍虚拟化基础知识、虚拟化的常见类型、服务器虚拟化构建技术、虚拟化应用、服务器虚拟化的性能分析、云计算的服务模式、云计算的关键技术、云数据中心构建、云应用、KVM 虚拟化技术等内容	1	linux 基础
5	专业基础课	Java 编程基础	Java 语言基本语法、面向对象编程思想、面向对象高级特性、继承、封装、多态、接口、抽象类、引用数据类型的转换、访问权限修饰符的使用、final 关键字、方法的覆盖、异常处理等	8	
6	专业基础课	web 前台技术	HTML5 基本标记、表单、视频与音频、本地存储、离线应用、地理定位、JavaScript 基础语法、chrom 等浏览器调试工具、数组、Object、Function、String 和正则表达式、BOM 模型、事件、location 对象、history 对象、HTML DOM 编程等	8	Java 编程基础
7	专业基础课	Java Web 编程	Tomcat 与 eclipse 的集成开发、Servlet3.0、JSP、JSTL+EL、MVC、分层架构、Session、Cookie、连接池、文件的上传下载、过滤器、反射、jQuery 框架介绍安装、jQuery 选择器、jQuery 筛选、jQuery 事件、jQuery Ajax 等前端开发框架。Ajax 工作原理、JS 异步交互案例、XML、JSON	8	Java 编程基础、web 前台技术
8	专业基础课	mysql 数据库基础	mysql 介绍、mysql 安装、编写简单的查询语句、限制数据和对数据排序、单行函数、多表查询、分组函数、子查询、DML、表和约束、用户及权限管理 mysql 日志、数据导入导出、MySQL 体系结构、MySQL 存储引擎	4	
9	专业核心课	大数据技术基础	本课程介绍大数据的整体概况，使学生对大数据的课程有一个整体的把握；大数据平台 Hadoop 的生态系统，基本组件，Hadoop 发行版本、比较、选择，单机及集群环境的准备（系统、基本配置、规划等），搭建企业 Hadoop 大数据集群平台；HDFS 基本架构及原理、HDFS Shell 命令、Java AP；MapReduce 工作机制、执行过程、序列化机制和数据类型、TopN、数据去重、排序等	6	Java 编程基础、linux
10	专业核心课	分布式数据库（数据仓库）	介绍数据仓库理论；HBase 数据模型、物理存储、架构、Write-Ahead-Log、HBase 存储格式、应用、配置管理、HBase Java API 使用；Hive 功能、体系结构、使用场景、Hive 环境搭建、Hive 常见查询、Hive 中的内部表、外部表、分区表；Sqoop 功能、使用原则、将 RDBMS 数据导入到 Hive 表中、将 HDFS 上文件导出到 RDBMS 表中；mysql 分布式数据库、数据路由（mycat）、mongodb 数据库	8	hadoop 大数据
11	专业核心课	数据可视化工具	介绍数据可视化工具的功能、安装配置、使用及程序的编写等	3	web 前台技术
12	专业核心课	大数据流式计	内容包含 kafka 功能、体系结构、kafka 通信协议、安装部署、shell 操作、java 操作、producer 和	6	hadoop 大数据

		算	consumer 开发; Storm 架构及原理、Storm 集群安装部署、API 介绍及开发、Storm 事务、Storm 分区、Storm Trident; Spark Streaming 流式计算等		
13	专业核心课	Redis 内存数据库编程	介绍 redis 的功能、安装部署、配置、数据类型、redis 常用命令、发布订阅、事务、脚本等	2	linux 基础
14	专业核心课	大数据内存计算技术	基础语法、条件控制与循环、函数、数组、集合、面向对象编程、类型参数、模式匹配、正则表达式、异常处理、文件操作等; Spark 基本概念及原理、RDD 技术原理与操作编程、Spark 程序设计、Spark Core 模块、Spark SQL 等	7	hadoop 大数据 Java 基础
16	专业核心课	数据采集与数据预处理	课程包含使用大数据爬虫采集互联网数据、使用 Sqoop 采集关系型数据库数据、使用 Flume 采集日志数据, 使用 Kettle 工具进行数据的 ETL 处理等内容	4	大数据编程
19	专业核心课	大数据项目 东软汽车电子大数据分析项目	东软与某城市交管局合作, 研发乘用车大数据分析系统, 进行该市人群乘用车销管分析, 解决该市汽车销管问题。乘用车大数据分析系统需要使用离线通信数据进行分析, 东软与某省相关部门进行合作, 东软负责提供分析模型并将分析模型部署到该省的的系统机房内, 该省相关部门负责提供省市信令数据, 共同完成乘用车销管分析并将最终分析结果提供给某城市交管局。	7	大数据编程基础
20	专业核心课	Docker 课程	介绍 Linux 下的 Docker 技术的基本情况; 基于 Docker 容器的快速部署、维护、开发; Docker 如何进行运维、监控、优化, 以及常见故障的处理;	4	linux 基础

十一、毕业设计指导

11.1 毕业设计采用双师型指导方案:

- 1)、学生自主选题
- 2)、企业老师对选题及技术选型进行指导
- 3)、学校老师对课题进行评审及对文档编写进行指导与评审
- 4)、企业老师对课题的实现提供相应的技术支持与指导
- 5)、争取在去实习之前将毕设初稿完成

11.2 毕业技术与论文指导流程

培训内容	课程名称	课程内容概述	培养方式	备注
准备部分	论文选题	1、企业老师介绍各个题目内容 2、根据学生前期实训成绩确认题目	面授	
	任务书制作	1、根据题目企业指导老师下达任务书并且与学生确认内容进度	面授	

培训内容	课程名称	课程内容概述	培养方式	备注
	开题报告制作	1、学生根据任务书，搜集资料撰写开题报告、明确模块内容和创新点	面授	
	需求分析及设计	1、使用思维导图 Xmind 进行需求分析和任务列表 2、通过使用 team-oschina 进行团队进度管理		
编码部分	论文成果物制作	根据任务书中功能任务进行编码，企业导师指导学员完成技术能点公关，控制整个进度，定期完成进度墙更新，向相关人员进行通报。		
	论文格式审核	企业指导老师审核学员所有需要提交的文档和论文格式内容		

11.3 部分毕设选题参考：

1、公安厅大数据案例分析的设计与实现：

【本课题目标】公安厅大数据案例分析的设计与实现

【本课题任务和要求】近年来违法犯罪活动从数量到质量都在发生着变化，犯罪的组织化、职业化、智能化日趋明显。公安部门希望基于现有的海量犯罪数据进行分析与挖掘，预测犯罪的趋势与动态，为公安决策部门制定有针对性的措施和打击犯罪活动提供辅助决策。。
建议开发语言：大数据技术。以吸毒嫌疑人预测为切入点，通过使用 SaCa RealRec 数据科学平台对吸毒人员行为进行分析、特征提取与建模，实现了对吸毒嫌疑人的精准预测功能，流程方法也同样适用于其他犯罪嫌疑人预测，如电信诈骗、盗抢等。

这里选择分类模型对吸毒嫌疑人进行预测，主要步骤如下：

1. 原始数据预处理，提取吸毒嫌疑人相关的特征，包括：性别、民族、年龄、出生地、文化程度、婚姻状况、职业类别，机动车违章频率等。过滤原始数据中关键特征缺失的数据，并将各特征进行规范化处理，按比例分配建模数据和测试数据
2. 多维特征分析，分析各特征数据与吸毒嫌疑人的关联程度及各特征间的相互作用情况。
3. 利用基本人口统计学信息及人员行为数据进行分类模型构建，当不断的增加数据特征时，模型预测的准确率从最初的 60%逐渐开始增加。可见当数据特征越来越多时，少量特征造成的欠拟合问题开始得到解决，通过对模型不断的优化最终达到了 85%的准确率。

2、银联防欺诈案例分析的设计与实现

【本课题目标】银联防欺诈案例分析的设计与实现

【本课题任务和要求】现在有来有多的人使用金融服务，也催生了很多欺诈交易，所谓的欺诈交易，是指涉嫌使用虚假身份获取银联卡，或冒用他人银联卡（或账户）获取商品或服务的欺骗性交易行为。银联的目标是，通过以往积累的数据，通过大数据以及机器学习的方法建立一套有效的预测欺诈行为的体系，为个人以及团体的财产安全作出保障。

建议开发语言：大数据技术。基于以往的欺诈交易的信息，通过大数据科学平台对欺诈的信息进行数据清洗，特征提取，完成建模找寻关系以及预测欺诈等功能。

这里完成两大功能，第一进行数据分析，第二进行欺诈预测

数据分析判定如下的步骤：

1. 数据占比的分析，通过占比分析可以知道，这里的高风险的交易占总体少数，大部分都是合法的交易。这个功能可以作为抽样或是调整数据比例作出依据。

2. 关联分析，高风险的交易，和其他的属性的关系，由于其他属性比较多，选取最可能的一些属性进行关联性分析，我们选取的是高风险交易，高风险资金转移，大额资金转移，高风险的省份。

风险的资金转移往往是大额资金转移，当有大额资金转移时那么就要对资金转移进行预警和防范了，高风险交易与高风险省份有关，可见地理位置对高风险交易的影响明显。

3、交通大数据案例分析的设计与实现

【本课题目标】交通大数据案例分析的设计与实现

【本课题任务和要求】由于轨道建设具有工期长、投资高、运营成本大等特点，因此有必要进行轨道客流预测，才能从定量的角度进行经济、环境和社会等各个角度的分析和评价，进而评价轨道建设项目的可行性和实施效益。另外，在轨道交通系统的运营过程中，对客流的动态变化进行实时跟踪和系统分析，掌握客流变化规律也是轨道交通系统运输组织工作得以顺利进行的前提。

建议开发语言：大数据技术，对轨道交通业务数据进行分析、特征提取、用户建模，实现对个体卡用户进行实时进出站精准预测的功能，从而能够基于个体预测做某个站点的宏观客流预测。基于数据科学提供的丰富大规模机器学习算法库，通过多种算法进行离线评估比较，最终选择决策树模型对卡用户进站站点进行预测，其主要流程分为两步：

一是对原始数据进行特征抽取工程，基于数据科学提供的多维特征分析功能组件，可以方便的过滤原始数据中记录不足的用户，并将连续数值处理为离散特征，如“时间”特征的离散化。二是基于特征抽取过程中的构建的多维特征向量，利用个体用户出行记录进行模型构建，由于个体用户出行行为具有较强的个性化规律化特点，基于数据科学强大的分布式计算能力，可以为每一卡用户构建个性化的预测模型，最终实现大规模增强学习模型。

另：其他可选择选题（部分）：

海洋数据监测及可视化分析	基于 Hadoop 的学业预警系统的设计与应用
基于 Hadoop 技术的电影推荐系统的设计与实现	基于图像的课堂人数统计系统设计与实现
基于大数据的企业竞争情报系统设计与实现	基于 spark 的物流园区交通拥堵系统的设计与实现
微博舆情分析系统的设计与实现	高速路视频交通事件智能检测系统应用探究
基于 Hadoop 技术的电力大数据平台的分析与应用	基于 Hadoop 技术的地下车库智能照明系统的开发与设计
智慧城市大数据平台的设计与实现	基于大数据技术的音乐个性化推荐系统设计与实现
基于大数据技术的病症种类统计系统的设计与实现	基于大数据技术的手机应用推荐系统的设计与实现