



第5届全国高校大数据与人工智能教学研讨会

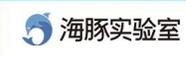
2022.05.13-2022.05.14 中国·厦门



大会官网

主办单位：教育部高等学校计算机类专业教学指导委员会

承办单位： 厦门大学 XIAMEN UNIVERSITY  天津大学 Tianjin University  中南大学 CENTRAL SOUTH UNIVERSITY  西南财经大学 SOUTHWEST UNIVERSITY OF FINANCE AND ECONOMICS  福州大学 FUZHOU UNIVERSITY  华侨大学 HUAQIAO UNIVERSITY  集美大学 JIAMEI UNIVERSITY  福建技术师范学院  人民邮电出版社 POSTS & TELECOM PRESS

协办单位： 美林数据 MERITDATA  TRANSWARP 星环科技  海豚实验室  山东省计算中心 SHANDONG COMPUTER SCIENCE CENTER  荔峰科技 LIFENG  Bai 百度  商迪3D web3di.com

大数据专业建设与科教研究融合

张祖平

中南大学计算机学院

zpzhang@csu.edu.cn



- **专业发展现状**
- **培养体系演变**
- **专业建设认证**
- **系列教材建设**
- **实验实训建设**
- **科教研究融合**

一、专业发展现状

■ 全国:

- 本科院校设立“数据科学与大数据技术”专业 (2016年,3+32+248+203+62)
- 高职院校设立“大数据技术与应用”专业 (2016年,62+208+143+406)
- 本科院校设立“大数据管理与应用”专业 (2018年, 6+25+59)

■ 中南大学大数据专业:

- 2015级 计算机科学与技术(大数据方向, 31人, 大四)
- 2016级 数据科学与大数据技术(2个班, 60人->61人+5人-2人)
- 2017级 数据科学与大数据技术(2个班, 62人->68人-8人)
- 2018级 数据科学与大数据技术(2个班, 大类招生, 60人->68人)
- 2019级 数据科学与大数据技术(2个班, 大类招生, ->88人) 需求太旺, 扩大招生
- 2020级 数据科学与大数据技术(2个班, 大类招生, ->82人) 需求太旺, 扩大招生
- 是首批申报成功数据科学与大数据专业的3所高校之一
- 2017年全国专业排名第1, 2018年排名第2, 2019年, 2020年多个排名, 排1-9

一、专业发展现状

➤ 2015级 计算机科学与技术(大数据方向, 31人, 大四)

◆ 读研14人(推免6人), 出国7人, 就业8人, 备二战2人

✓ 读研 14人

- 本校1, 北大1, 浙大2, 科学院大学2, 中科大1
- 计算所1, 北邮1, 电子科大3, 华中科大1, 航天二院1

✓ 就业 8人

- 腾讯1, 阿里1, 维沃移动通信有限公司1,
- 工商银行1, 中信行1, 深农商行1,
- 北京字节跳动科技有限公司1人, 汉得信息技术股份公司1

一、专业发展现状

■ 中南大学大数据专业_毕业班级:

➤ 2016级，64人

➤ 出国 9人

- 美国 德州大学奥斯汀分校、弗吉尼亚大学；英国 布里斯托大学、曼彻斯特大学、南安普顿大学；新加坡 新加坡国立大学；中国香港 香港城市大学、香港科技大学

➤ 读研 22人，其中保研17人

- 本校4人 北京大学 1人 中国科学院大学 1人 华中科技大学 6人 东南大学2人
- 上海交通大学 1人 同济大学 1人 武汉大学1人 西安交通大学 1人 其它 4人

➤ 就业 22人，分布在15个省市，各行各业

- 广东 7人、宁夏 2人、天津 2人
- 湖南、安徽、北京、福建、河北、江苏
- 辽宁、山东、陕西、四川、天津、浙江

➤ 自由职业 7人

➤ 待就业（备二战） 4人

一、专业发展现状

中南大学数据科学与大数据专业招生及升学就业统计

专业招生:

- 2015级 计算机科学与技术(大数据方向, 31人, 2016级, 64人; 2017级, 60人; 2018级, 68人;
- 2019级 **88人**; 2020级 **82人**; 2021级 **待专业分流**

专业升学就业:

专业人数	升学人数		本校升学		外校升学		出国、出境深造		就业人数	
	人数	占总人数比例	人数	占升学人数比例	人数	占升学人数比例	人数	占升学人数比例	人数	占总人数比例
2019-31	23	74.19%	1	4.35%	14	60.87%	8	34.78%	8	25.81%
2020-64	31	48.44%	4	12.90%	18	58.06%	9	29.03%	33	51.56%
2021-60	21	35.00%	10	47.62%	8	38.10%	3	14.29%	39	60.00%

- 专业发展现状
- 培养体系演变
- 专业建设认证
- 系列教材建设
- 实验实训建设
- 科教研究融合

二、培养体系演变

■ 学分与学时 (15级大数据方向)

课程模块类别		必修课		选修课		合计		占总学分比例
		学分	学时	学分	学时	学分	学时	
通识教育	理论教学	40	724	8	0	48	724	25.53%
	实践环节	4.5	6周			4.5	6周	2.39%
学科教育	理论教学	46.5	744	12	192	58.5	936	31.12%
	实践环节	8.5	16+8周			8.5	16+8周	4.52%
专业教育	理论教学	专业核心类		9	144	9	144	4.8%
		专业类		24.5	392	24.5	392	13%
	实践环节	27	27周			27	27周	14.3%
个性培养	课外研学	8						4.2%
总计		134.5	1316+41周	53.5	856	188	2272+41周	100%
其中:实践环节		50.5	64+41周			50.5	64+41周	26.8%

二、培养体系演变

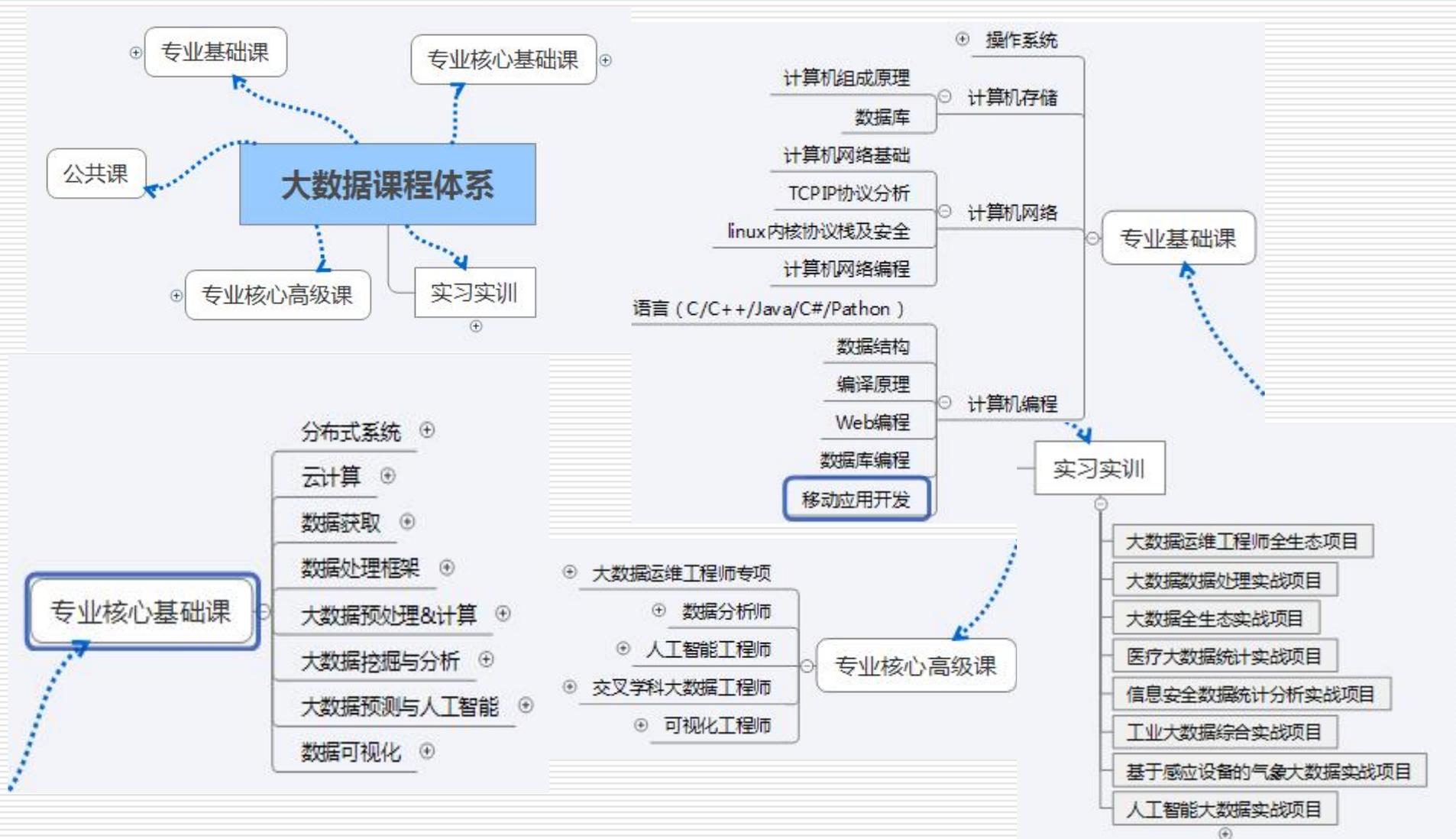
■ 学分与学时 (16版执行)

课程模块类别		必修课		选修课		合计		占总学分
		学分	学时(周)	学分	学时(周)	学分	学时(周)	比例(%)
通识教育课程	理论教学	27.5	440	8	128	35.5	568	19.7
	集中实践环节	1.5	3周	0	0	1.5	3周	0.8
学科教育课程	理论教学	25.5	408	4	64	29.5	472	16.4
	集中实践环节	0	0	0	0	0	0	0
专业教育课程	理论教学	41.5	664	25	400	66.5	1064	36.9
	集中实践环节	32	32周	7	7周	39	39周	21.7
个性培养课程	理论教学	2	32	0	0	2	32	1.1
	课外研学	0	0	6		6		3.3
总计		130	1544/35周	50	592/7周	180	2136/42周	100
其中： 实践教学	课内实践	11	176	6	96	17	272	9.4
	集中实践	33.5	35周	7	7周	40.5	42周	22.5
	课外研学	2	32	6		8	32	4.4
	合计	46.5	208/35周	13	96/7周	59.5	304/42周	36.3%

二、培养体系演变

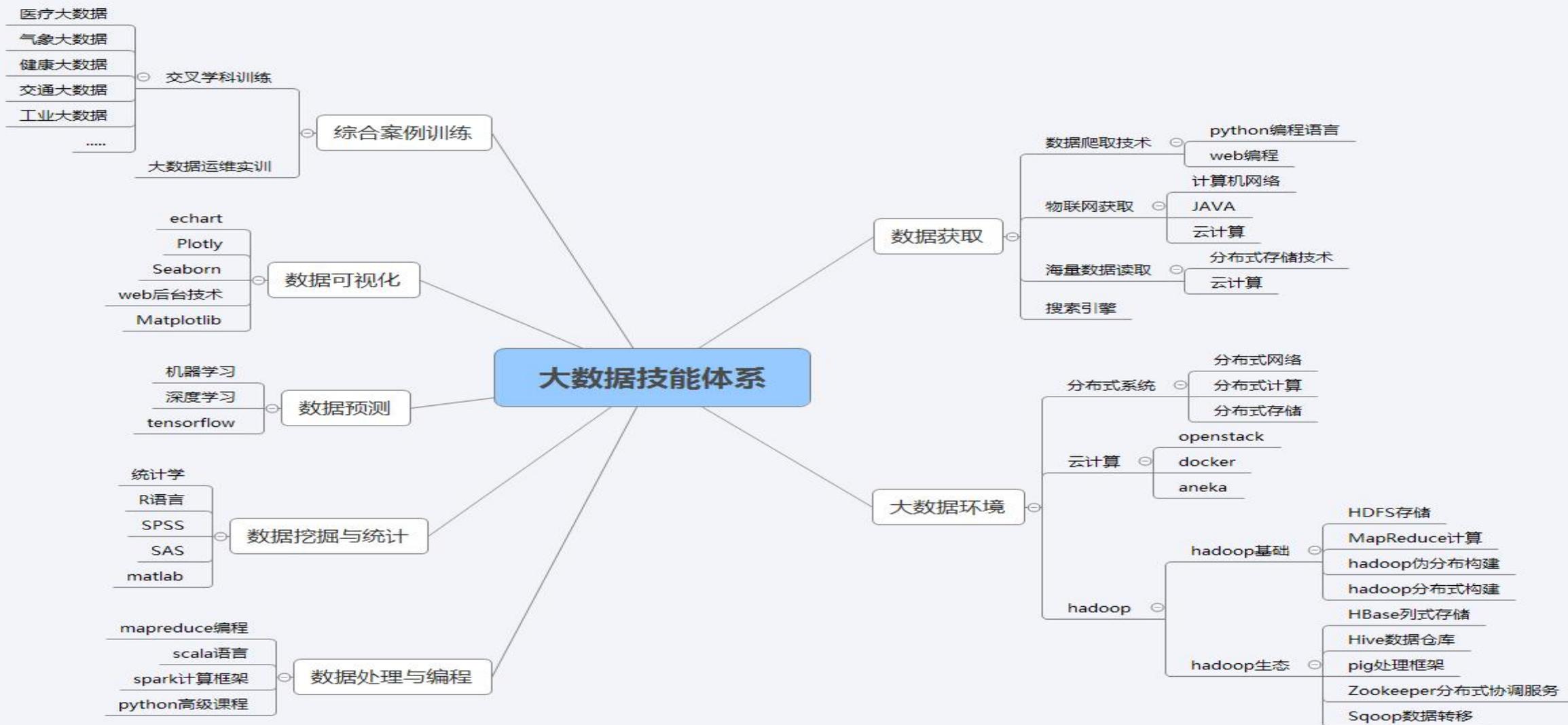
■ 学分与学时 (新版2018)

课程模块类别		必修课		选修课		合计		占总学分比例(%)
		学分	学时(周)	学分	学时(周)	学分	学时(周)	
理论教学	课堂讲授	92.2	1532+0周	28.7	460+0周	120.9	1992+0周	69.48%
	课内实践	10.8	152+3周	3.3	52+0周	14.1	204+3周	8.1%
	合计	103	1684+3周	32	512+0周	135	2196+3周	77.59%
实践教学	集中实践环节	30.5	48+29周	0	0+0周	30.5	48+29周	17.53%
	单独设课实验课	2.5	80+0周	0	0+0周	2.5	80+0周	1.44%
	个性培养	0	0+0周	6	16+5周	6	16+5周	3.45%
	合计	33	128+29周	6	16+5周	39	144+34周	22.41%
合计		136	1812+32周	38	528+5周	174	2340+37周	100%



二、培养体系演变

技能体系



二、培养体系演变

■ 实际执行课程体系

■ 通识教育课程

- ✓ 思政类、军体类、外语类、信息技术类、文化素质类

■ 学科教育课

- ✓ 公共基础课 必修**19.5**学分（编号，名称，性质，学分，学时，学期）

130702X10	高等数学A2（一）	必修	5	80	1
130702X20	高等数学A2（二）	必修	5	80	2
130711X10	线性代数	必修	2	32	2
130712X10	概率论与数理统计	必修	3.5	56	3
130201X10	科学计算与数学建模	必修	4	64	4

■ 学科教育课

✓ 学科基础课（必修36学分（编号，名称，性质，学分，学时，学期）

092101T10	新生课	必修	1	16	1
090200T10	计算机程序设计基础（C语言）	必修	4	64	1
091104X10	数字电子技术A	必修	3.5	56	1
090201X10	离散数学	必修	3	48	3
090205X10	数据结构	必修	3.5	56	2
092102Z10	数据科学与大数据技术导论	必修	2	32	3
090222Z10	计算机组成原理与汇编	必修	4	64	2
090211Z10	计算机网络	必修	3	48	3
090212Z10	数据库原理	必修	3	48	3
090213Z10	操作系统原理	必修	3	48	3
090210Z10	算法分析与设计	必修	3	48	4
090218Z10	软件工程	必修	3	48	5

■ 专业教育课

✓ 专业核心课（必修12学分（编号，名称，性质，学分，学时，学期）

092109Z10	分布式系统与云计算（强化分布式存储与数据中心技术）	必修	3	48	5
090217Z10	机器学习（强化非结构化数据分析方法）	必修	3	48	4
092105Z10	数据仓库与数据挖掘	必修	3	48	5
092113Z10	大数据编程（基于hadoop 和spark）	必修	3	48	6

■ 专业教育课

✓ 专业课（选修第一组共9门，至少选14学分）

092103Z10	大数据采集与融合技术	选修	2	32	3
092104Z10	信息组织理论与技术	选修	3	48	4
090228Z10	大型数据库技术	选修	2	32	5
092106Z10	Python数据处理编程	选修	2	32	5
092107Z10	R语言数据分析编程	选修	2	32	5
092111Z10	智能搜索引擎技术	选修	2	32	6
092115Z10	深度学习	选修	2	32	5
090242Z10	可视化技术	选修	2	32	6
090267Z10	信息与网络安全（强化大数据安全与隐私保护）	选修	2	32	6

二、培养体系演变

■ 实际执行课程体系

■ 专业教育课

✓ 专业课（选修第二组共12门，至少选18.5学分

090203Z10	面向对象程序设计C++	选修	3	48	2
090207Z10	JAVA语言与系统设计	选修	3	48	3
090220Z10	Web技术	选修	2	32	3
090219Z10	Linux系统及应用	选修	2	32	4
090234Z10	多媒体原理与系统设计	选修	2	32	6
090232Z10	移动应用开发	选修	2	32	6
090245Z10	计算机仿真与建模	选修	1.5	24	6
090248Z10	生物信息学	选修	2	32	6
092112Z10	医疗大数据	选修	2	32	6
090236Z10	并行计算	选修	2	32	7
090241Z10	人机交互	选修	1.5	24	7
090244Z10	电子商务	选修	2	32	7

■ 专业教育课

✓ 集中实践环节（必修31学分，选修2学分）

090206X11	程序设计实践（C++和数据结构）	必修	2	2周	2期末，3期分散进行
090215Z11	应用基础实践一（网络+数据库+Java）	必修	2	2周	3期末，4期分散进行
092110Z11	数据处理方法课程设计	必修	2	2周	5期分散进行
092114Z11	大数据综合应用实践（基于医疗大数据）	必修	3	3周	6
092108Z11	信息组织课程设计	选修	2	2周	4期末，5期分散进行
090250Z11	IT项目管理培训	选修	2	2周	7
090253Z11	IT项目开发案例分析	选修	1	16	7
092116Z11	生产实习	必修	4	4周	7
090273Z11	认识实习	必修	2	2周	2
092117Z11	毕业实习、毕业设计	必修	16	16周	8

■ 个性培养课

✓ 创新创业课（必修2学分）

430601G10 创新创业导论 必修 2 32 5

✓ 课外研学

- 选修6学分（其中须修2学分创新创业实践，1学分实验室技术安全与环境保护知识学习培训与考核）

- 专业发展现状
- 培养体系演变
- **专业建设认证**
- 系列教材建设
- 实验实训建设
- 科教研究融合

工程教育认证申请书

(2022 版)

中国工程教育专业认证协会秘书处：

根据《工程教育认证办法》有关认证申请资格的规定，我校以下专业满足申请条件，现申请参加工程教育认证。

申请认证学校：中南大学

申请认证专业：数据科学与大数据技术

我校保证本申请书及所有附件材料完全属实，保证按照《工程教育认证办法》及各项文件要求，规范参与认证，并严格遵守认证工作各项纪律要求，不探听评审相关信息，不拜访专家或以任何形式请托关照，愿意承担现场考查期间（含在线方式考查）各项费用支出。

■ 三、本专业培养目标和毕业要求（不超过 2000 字）

- ✓ 最近一届毕业生使用的培养方案为 2016 版本。目前在校学生使用的都是 2018 版本。
- ✓ 2018 版本是为适应大类招生在 2016 版的基础上进行的微调。
- ✓ 明年进行专业认证时的 2022 届毕业生也使用的是 2018 版，因此申请书主要以 2018 版培养方案进行描述。

■ 中南大学数据科学与大数据技术专业本科生的培养目标具体包括以下 5 个方面：

- ✓ 培养目标 1：掌握数学、自然科学、计算机科学和数据科学与大数据技术的基础知识与专业知识，并用于分析和解决大数据复杂工程问题；
- ✓ 培养目标 2：能综合考虑技术、社会、法律、文化等因素，设计和开发复杂大数据系统，并表现创新意识；
- ✓ 培养目标 3：具有良好的人文素养、社会责任感和道德品质，遵守职业规范和社会伦理，德智体美劳全面发展，是社会主义事业合格建设者和可靠接班人；
- ✓ 培养目标 4：具有团队合作精神和组织管理能力，能与国内外同行、客户和公众有效沟通；
- ✓ 培养目标 5：能始终坚持学习和自我完善，紧跟数据科学与大数据技术领域的理论与技术发展趋势

■ 从知识体系角度布局课程体系模块，7大模块

① **通识教育知识模块**：国家有统一规定，应遵照执行。

② **自然科学知识模块**：对接工程认证需求，应予充分考虑。

③ **数理统计学科知识模块**：是大数据领域建模与大数据分析的基础；是对已有典型模型理解、应用、并尝试改进的基础。

④ **计算机学科知识模块**：是理解大数据平台工作原理的基础；是分析各类平台适应范围与性能优劣的基础；是分析平台演进趋势与革新的基础。

⑤ **数据挖掘与人工智能知识模块**：是熟练应用典型算法解决实际问题，以及面向大数据问题域设计大数据核心处理算法的基础。

⑥ **大数据平台类知识模块**：是了解各类平台适应范围，熟练掌握平台使用方法，高效开发大数据产品的基础。

⑦ **专业交叉融合知识模块**：除了上述知识模块以外，更重要的是提供灵活的机制，方便学生根据兴趣深入学习其它专业领域知识，熟悉相关专业业务，运用大数据专业知识、方法、技术与基本技能解决面向特定领域的问题。

■ 从能力形成角度厘清课程依赖关系

① 培养对大数据专业的基本认知能力的课程设置与依赖关系示例：数据科学与大数据技术导论→大数据企业认识实习（学数据知识到自己感知数据）

② 培养通用问题求解能力的课程设置与依赖关系示例：离散数学→数据结构→算法分析与设计（学会逐步用计算思维来表达并解决问题）

③ 培养大数据领域问题求解能力的课程设置与依赖关系示例：科学计算与数学建模→数据仓库与数据挖掘→机器学习与深度学习（现实世界的抽象到求解再有效求解）

④ 培养对大数据平台工作原理理解能力与平台架构设计能力的课程设置与依赖关系示例：计算机组成原理与汇编→操作系统原理→数据库原理→计算机网络→分布式系统与云计算→软件工程（架起原理与应用的桥梁）

⑤ 培养大数据应用能力与大数据产品开发能力的课程设置与依赖关系示例：程序设计类课程与程序设计实践→应用基础实践（网络+数据库+Java）→数据处理方法课程设计→大数据编程→大数据综合应用实践→大数据企业生产实习（从易到难强力强化）

■ 从人才需求角度确定毕业基本要求

数据科学与大数据技术专业学生的毕业要求

毕业要求1	知识要求	具备数学、自然科学、计算机科学基础知识、以及大数据工程专业知识，用于描述和分析大数据系统、大数据应用工程、大数据科学研究等相关复杂问题。
毕业要求2	知识要求	了解国家发展战略规划、产业政策、法律法规、正确认识、理解、评价大数据工程对经济、社会、环境、健康、安全、文化的影响，保持经济增长、社会和谐、环境友好的协调发展。
毕业要求3	知识要求 能力要求	具有对大数据系统、大数据应用及相关复杂工程问题进行建模、设计、分析、研究、验证等工程综合知识和实践能力，并表现出创新意识。
毕业要求4	能力要求	熟练运用主流大数据平台（如Hadoop 或Spark）、典型深度学习系统（如TensorFlow），设计、开发、生产面向特定行业的大数据产品。
毕业要求5	能力要求	具有分享包容的心态、沟通与协作的愿望、规范化组织与管理意识，能熟练运用一门以上外语进行国际交流，具有较强的口头和书面表达能力。
毕业要求6	素质要求	具有科学人文素养、强烈的社会责任感、理解并遵守职业伦理。
毕业要求7	素质要求	了解信息学科前沿发展趋势，关注本专业与其他学科交叉融合的新理论、新方法和新技术，具有开放意识和全球视野。
毕业要求8	素质要求	具有探索新事物的兴趣，能保持上进心、自主学习和持续更新核心知识以适应专业或职业发展的能力。

■ 从社会适应角度奠定长期发展基础

依据国家社会需求、行业产业需要、学校定位及发展目标，本专业致力于培养适应不断演化的经济与社会发展需要，注重大数据科学与工程领域与医学医药、轨道交通、有色金属行业交叉融合的复合型高级工程技术人才。

① **能适应行业大数据应用发展需要，融会贯通**数学与自然科学基础知识、计算机科学基础知识、大数据科学与工程专业知识，提出复杂大数据工程项目的系统性解决方案。

② **能够跟踪大数据科学与工程领域的前沿技术，具备一定的大数据工程创新能力**、大数据分析价值挖掘能力，能够从事应用驱动的大数据产品的设计、开发和生产。

③ 具备良好的职业道德精神、社会责任感，**理解法律、环境、发展的相互关系，在工程项目实施中坚持绿色发展理念、能够注重经济与社会效益的协调。**

④ **具备健康的身心，拥有科学的人文精神、创新创业精神、团队精神**，具备良好的人际沟通与协调能力、有效的工程项目管理能力。

⑤ **能够从全球视野思考问题**，主动应对不断变化的国内外形势，具备自主学习能力、批判思维能力和国际交流能力。

■ 一、课程达成目标

- 数据科学与大数据技术导论是一门面向本专业的导论性课程，旨在让学生在大学入学最初阶段对本专业的发展历史、知识结构、培养目标与要求及数据科学与大数据技术相关的基础知识、典型技术、具体应用等有一个直观的认识
- 区别于新生课的普识性介绍，相关内容偏专业，目标是让学生对本专业的知识及培养要求有一个相对全面而直观的了解
- 概述性地介绍对应学科相关内容及典型人物，用以激发学生的学习兴趣，进一步了解设置新专业的历史背景与总体要求。

■ 二、课程的基本要求

— 知识

- 较好掌握数据科学与大数据技术的发展历史及相关典型概念如5V特性；
- 了解典型的大数据分析环境所包括的技术体系如Hadoop；
- 了解计算机典型的基础概念如数据、算法；
- 了解专业所需要掌握的知识体系及课程要求；
- 对大数据技术的典型应用有相对直接的了解并能联想生活中的大数据技术应用场景。

— 能力

- 主要培养学生对本专业的课程体系有区别与选择的能力；
- 对典型的大数据分析环境的技术体系有一定的判别与选择的能力；
- 对应用系统是否要用到大数据平台有一定的判别能力，对整个专业的知识体系有一定提前的预判与认知。

— 素质

- 提升自主学习和终身学习的意识，形成不断学习和适应发展的素质。

■ 第1章 专业概论

- 专业发展现状历史、紧密相关的专业、专业特点及要求
- 专业课程模块及技术体系，专业出路与就业

■ 第2章 数据科学与大数据基本概念

- 信号，数据，信息，知识；数据处理，数据管理，数据库，大数据
- 数据科学，数据统计，数据挖掘

■ 第3章 大数据核心技术

- 数据采集、数据存储与管理，数据预处理
- 数据清洗、数据挖掘、数据可视化分析

■ 第4章 大数据环境与技术

- 大数据技术生态; Hadoop, HDFS, YARN, SPARK

■ 第5章 大数据应用系统

- 交通大数据、医学大数据

■ 培养方案解读

■ 实验20个，自己做！

导论课不再虚！

养成自己独立做实验的习惯！

- 专业发展现状
- 培养体系演变
- 专业建设认证
- 系列教材建设
- 实验实训建设
- 科教研究融合

四、系列教材建设

■ 教材选用面临的挑战

■ 系列教材建设

- ✓ 大数据时代前沿技术丛书
- ✓ 中南大学出版社，学校给政策鼓励
- ✓ 教材的定位：
 - 实际可用，好用！
 - 成系列！
- ✓ 目前已正式出版8本

作者	书名
张祖平	数据科学与大数据技术导论
龙 军	数据仓库与数据挖掘
高建良	Spark大数据编程基础(Scala版)
刘丽敏	大数据采集与融合
高 琰	智能搜索引擎技术
余腊生	分布式系统与云计算—原理、技术与应用（上册，下册）
鲁鸣鸣	深度学习
龙 军	数据仓库与数据挖掘实践教程
高建良	大数据编程实践教程
鲁鸣鸣	深度学习实践教程
刘丽敏	大数据采集与融合实践教程
桂劲松	数据处理方法课程设计
桂劲松	大数据综合应用实践
夏佳志	可视化技术

- 专业发展现状
- 培养体系演变
- 专业建设认证
- 系列教材建设
- 实验实训建设
- 科教研究融合

五、实验实训建设

■ 实验室建设总体规划 投入600多万元

■ 实验室首期建设 投入96万元。

✓ 招标：采用北京以利天诚冰山大数据解决方案

✓ 大数据控制服务器*1台

- 处理器：2颗8核Intel Xeon E5-2630 v3,2.4G;
- 内存：128GB 硬盘：2*2TB
- 网卡：标配6个PCI-E*8千兆以太网控制器;

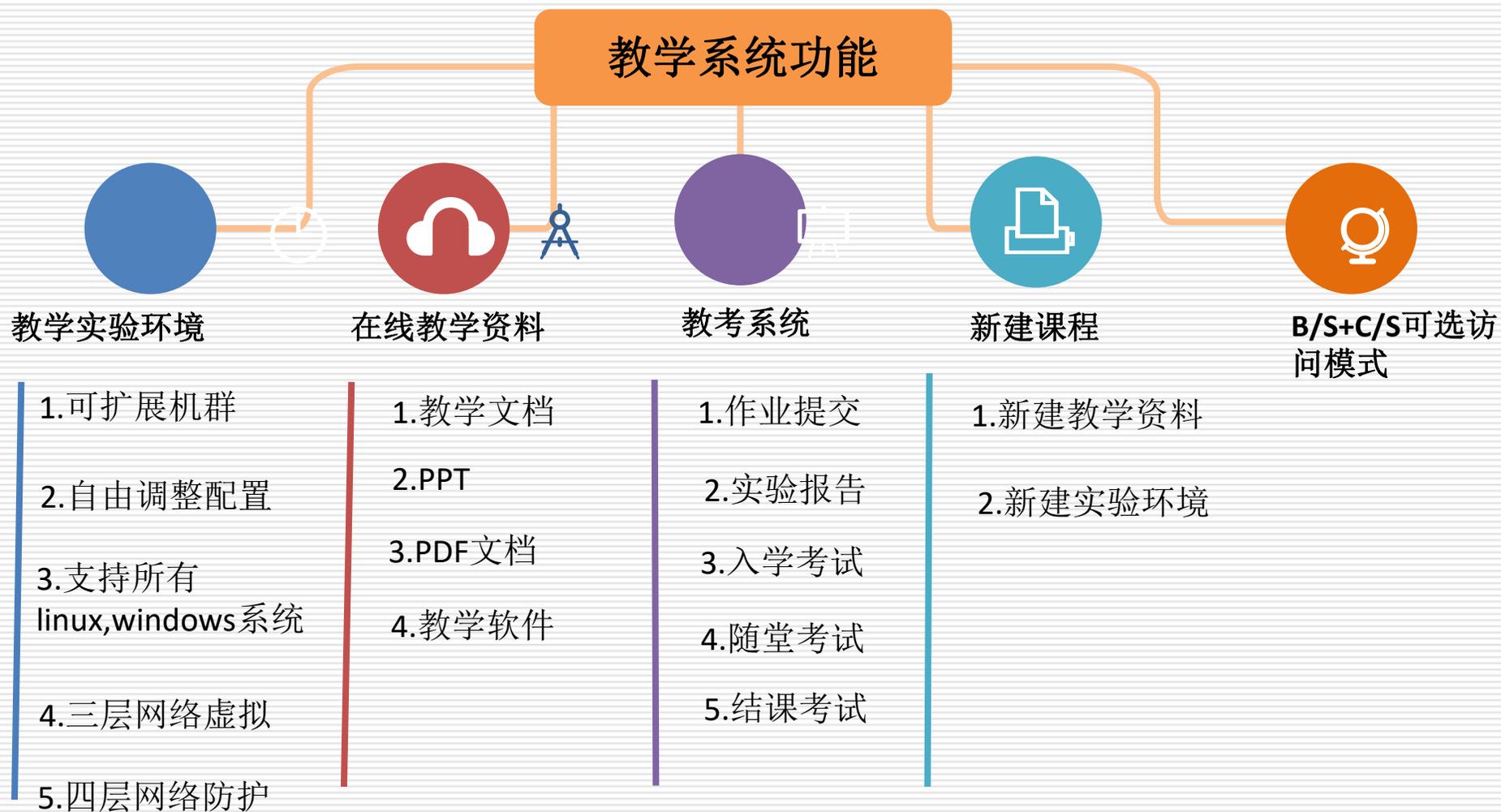
✓ 大数据计算服务器*3台

- 处理器：2颗8核， Intel Xeon E5-2630 v3,2.4G;
- 内存：128GB 硬盘：2*2TB
- 网卡：标配6个PCI-E*8千兆以太网控制器;

■ 目前建设了4期，实际投入了600多万元，还与计算机专业共享云平台，已基本形成了面向大数据专业教学实验的平台，但不足以支持实训



五、实验实训建设



五、实验实训建设

■ 实训平台建设总体规划

- ✓ 校内基地与校外基地相结合
- ✓ 学科建设与专业建设相结合
- ✓ 公开数据资源引进与自产数据脱敏相结合

■ 实训平台建设现状

- ✓ 医学大数据综合实训平台
 - 基于医学大数据国家工程实验室
 - 基于医疗大数据湖南省协同创新中心
- ✓ 智慧城市沙盘系统
 - 从调度控制算法的演示到数据产生获取、交通优化、数据分析展示
- ✓ 与企业单位、教育机构等社会力量合作共建实训基地 30万元/个

- 专业发展现状
- 培养体系演变
- 专业建设认证
- 系列教材建设
- 实验实训建设
- 科教研究融合

六、科教研究融合

■ 需求

- ✓ 学科评估中教学成果需要，新教学模式下需要研究型
- ✓ 人尽其才，有人喜欢偏教学；
- ✓ 需要科研的人参与教学
- ✓ 一个中心，两个中心，还是显式成果为中心

■ 模式

- ✓ 以平台为依托，项目为支撑，多种方式相结合
 - 教学科研实验平台一体化
 - 科研平台的成果与数据服务于教学；利用合作的力量
- ✓ 政策引导
 - 加大教学成果的奖励与关键环节挂勾；经费的投入倾斜，精品立项，孵化意识

六、科教研究融合

交通大数据_总体目标

■ 交通大数据平台的建设:

- ✓ 实现对多源数据资源的汇集融合;
- ✓ 海量数据的分析碰撞
- ✓ 分析成果的可视化展示和精准推送

为交管数据信息归集的最大化、交通分析研判的科学化、交通组织管理的智能化、交通指挥调度的精准化和交通综合治理的一体化提供情报支撑。



体目标

■ 大数据计算与服务环境

CPU 40 Cores, RAM128GB, HD 4TB 数十台

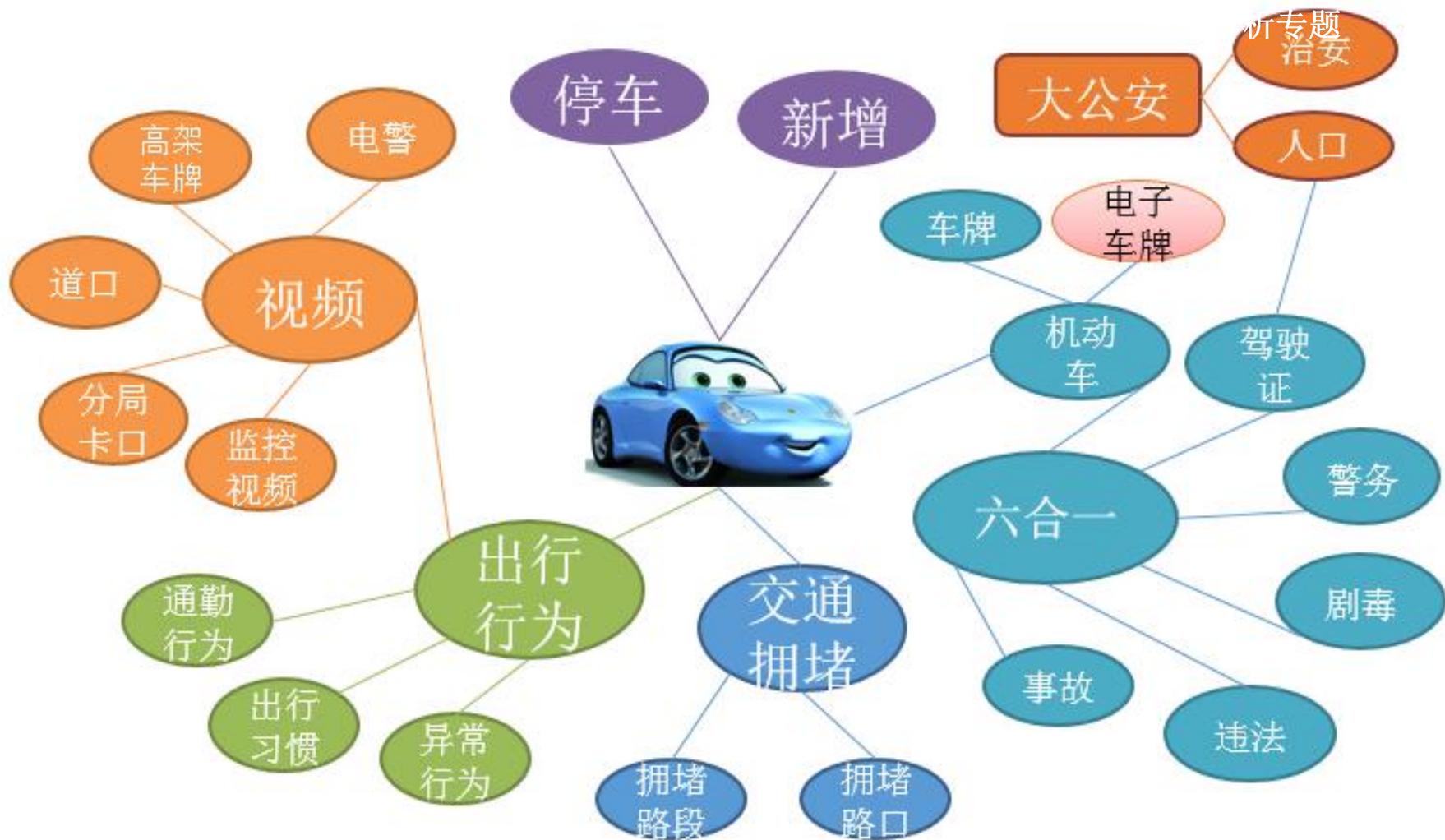
■ 大数据技术生态

- hadoop、hdfs、oozie、spark、sqoop、hive、hbase
- Yarn, kylin
- Echarts
- Oracle –ETL--Hbase

■ 数据

卡口表大约18.5亿条数据，每天新增1000w多数据，大约3.5G，三期完全建成后，卡口数据每天新增3000w左右数据,大约10G

全息数据库业务模型



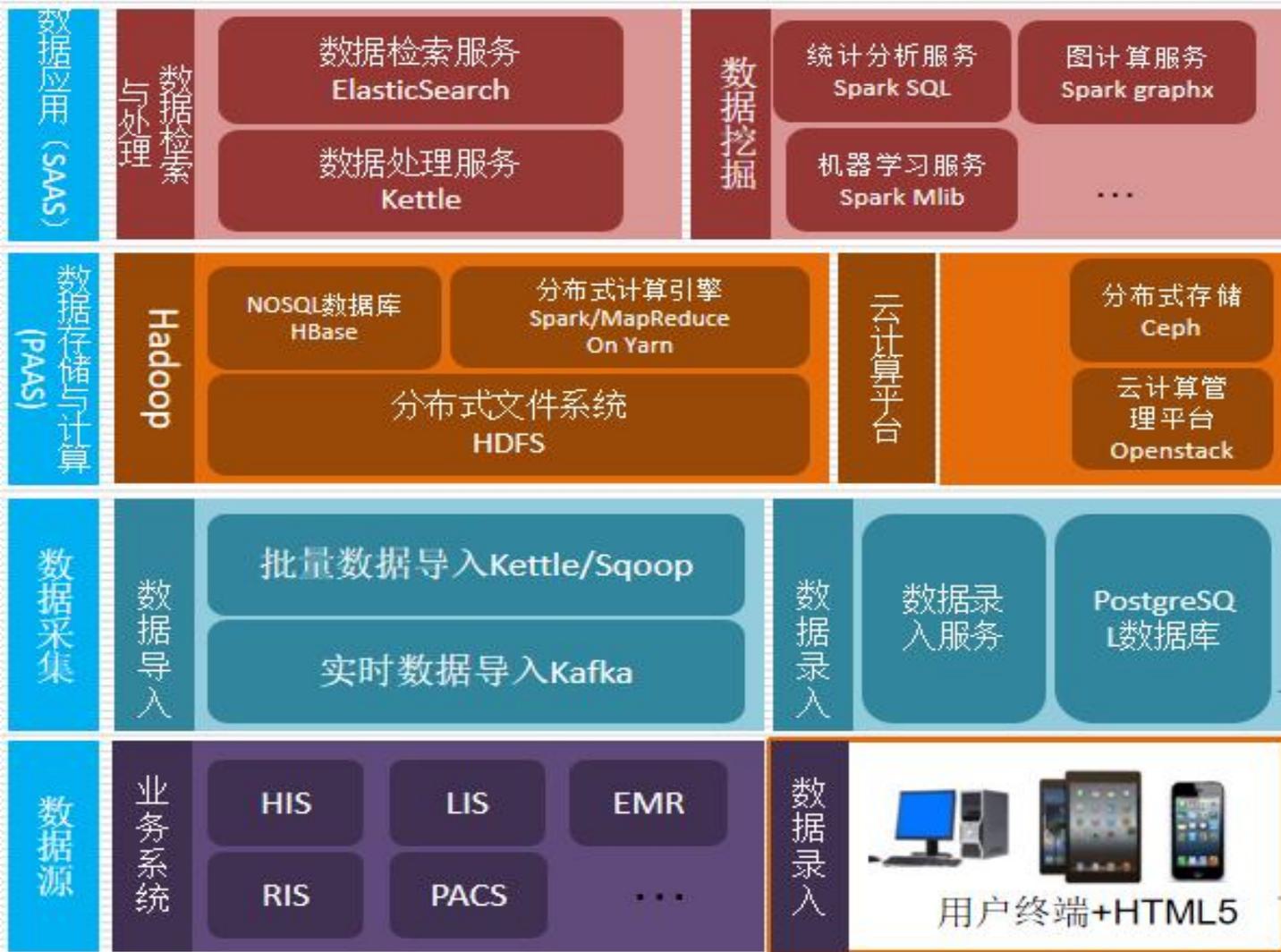
六、科教研究融合

交通大数据-两客一危一校分析



六、科教研究融合

医学大数据_软件架构



- 数据检索服务采用elasticSearch
- 数据处理服务采用ETL工具Kettle实现
- 数据挖掘相关服务采用Spark的SQL、graphx、Mlib实现，为科研人员进行数据挖掘提供基础软件包和计算环境。
- 基于Hadoop,采用spark、mapReduce分布式计算框架和资源调度服务Yarn组合作为平台的计算引擎
- 基于Openstack搭建云计算平台，分布式存储系统Ceph为云计算平台提供块、对象的存储服务
- 用户终端录入的数据服务管理的数据录入服务，保存到PostgreSQL数据库，然后同步到Hadoop平台
- 数据来源主要来自医院的业务系统和医护人员的录入
- 数据录入通用用户终端和HTML5实现。

体目标

- 基于深度学习的中文电子病历命名实体识别研究
- 面向大数据的汽车行业精准营销算法研究与应用
- 基于图神经网络的疾病基因预测方法研究
- 自然场景下的视频车辆跟踪研究
- 面向电力行业的信息实体抽取方法研究与实现
- 面向知识图谱的汽车智能化推荐研究与实现
- 汽车服务域元图谱构建研究
- 轨迹知识图谱的构建方法研究与实现
- 基于深度学习的非小细胞肺癌亚型预测
- 面向微博短文本的情绪分析研究

谢谢!



个人主页