

面向中国特色高水平职业院校 赋能大数据智能软件职业技能人才培养

王建民

清华大学软件学院

大数据系统软件国家工程研究中心

2022.12

主要内容

1

清华大学软件学院和职业教育协同发展的趋势

2

清华数为大数据软件栈赋能职业院校人才培养

3

清华数为大数据应用课程体系和合作成果

4

推动双高院校大数据智能方向发展的展望

清华大学软件学院赋能职业教育协同发展

清华大学软件学院是2001年经国家教育部和国家计委联合发文批准成立的首批全国示范性软件学院之一，隶属于清华大学信息科学与技术学院，其的前身是：“**清华大学应用技术学院**”。

■学生培养：落实开展大类培养 + 积极投入强基人才培养

■课程体系：**重构本科培养方案 + 建设强基课程体系**

目前开设38 门本科生课程

■精品课程：**国家级精品课两门、MOOC精品课程三门**



清华大学软件学院自成立以来，通过软件智能核心成果转化、各类各层次人才培训、企业信息化产业升级人才培养，积极赋能我国职业教育协同发展！

清华大学软件学院学科建设，夯实职业教育软件技术底层逻辑

学院面向世界科技前沿、国家重大需求和经济主战场，持续布局和优化五个方向：

- 软件科学与工程
- 可信软件与系统
- 数据科学与大数据技术
- 计算机图形学与辅助设计
- 群智感知计算与物联网



各学科方向前沿技术研究成果，奠定我国职业教育软件智能人才发展基石！

协同创新，积极响应国家发展职业技术教育战略需求

2022年5月1日：新版《职业教育法》公布

CCTV 13 新闻 新修订的《职业教育法》最受关注的方面

- “职业教育”与“普通教育”具有同等重要地位
- 国家采取措施，提高技术技能人才的社会地位和待遇
- 高等职业学校教育由专科、本科及以上教育层次的高等职业学校和普通
高校等学校实施
- 推动企业深度参与职业教育
- 职业学校学生在升学、就业、职业发展等方面与同层次普通学校学生享有
平等机会
.....



大修后的《职业教育法》，将带来哪些改变？

星期四 21:37 TV 新闻 欧洲理事会主席米歇尔20日访问乌克兰首都基辅并与乌总统泽连斯基

CCTV 13 新闻

我国已建成世界上规模最大的职业教育体系



目前共有职业学校

1.13万所



在校生

超过3000万人

但高技能人才占技能劳动者的比例还不到30%

直播连线

北京师范大学国家职业教育研究院院长 和震

新闻1+1

大修后的《职业教育法》，将带来哪些改变？

4月21日 21:43 TV 新闻 中国第25批赴刚果（金）维和医疗分队20日高标准通过联合国作战



统筹职业教育、高等教育、继续教育**协同创新**，推进职普融通、产教融合、**科教融汇**，优化职业教育类型定位。《第二十次全国代表大会报告》—习近平，2022年10月16日

提高质量。要以**国家战略需求为导向**，一体推进教育强国、科技强国、人才强国建设，推动建设世界重要人才中心和创新高地各项决策部署在高校、在教育系统的落实落地。**深化现代职业教育体系建设改革，培养更多高技能人才**。《深入学习宣传贯彻二十大精神 办好人民满意的教育》—怀进鹏，2022年12月13日

主要内容

1

清华大学软件学院和职业教育协同发展的趋势

2

清华数为大数据软件栈赋能职业院校人才培养

3

清华数为大数据应用课程体系和合作成果

4

推动双高院校大数据智能方向发展的展望

清华大学大数据研究中心，支撑大数据职业教育人才培养

- 2017年6月获批，2018年9月26日正式成立的校级跨学科交叉研究中心。
- 依托的国家、省部级科研平台：
- **大数据系统软件国家工程技术研究中心**
 - 高校：清华+国防科大+中山
+北理工+北大+人大
 - 企业：百度、阿里、腾讯
- **工业大数据系统与应用北京市重点实验室**
 - 软件学院+化工系+能动系



2017年9月11日成立仪式



2019年8月16日技术委员会会议

定位：服务国家实体经济发展需求,研发“自主可控”大数据系统软件

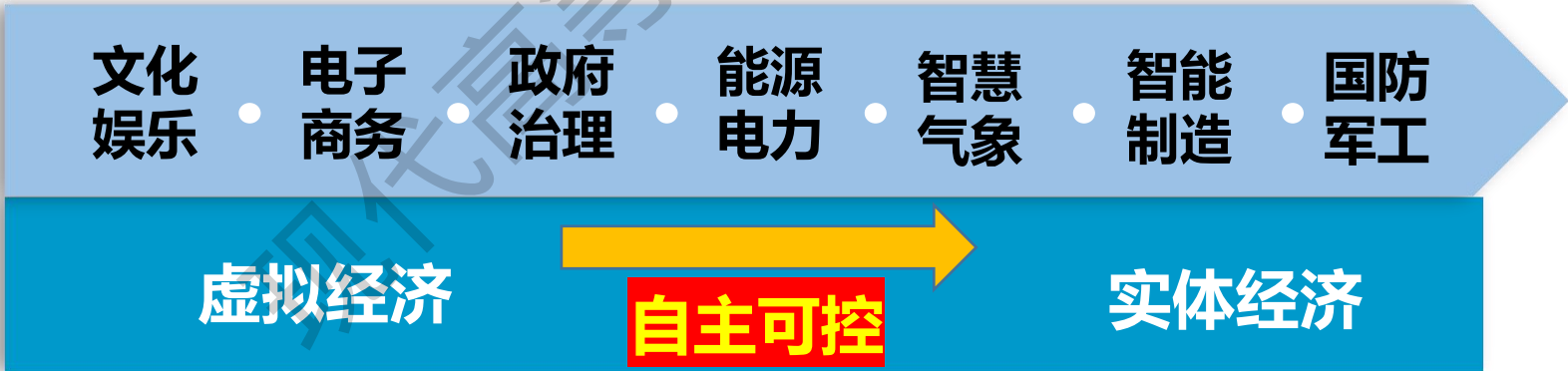
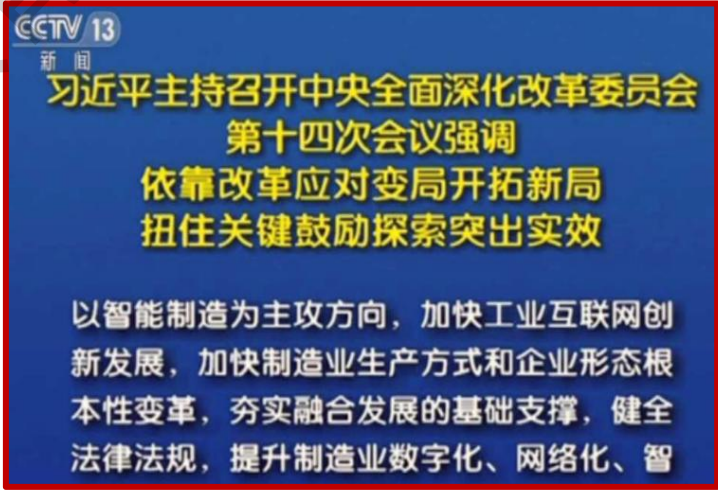
《促进大数据发展行动纲要》



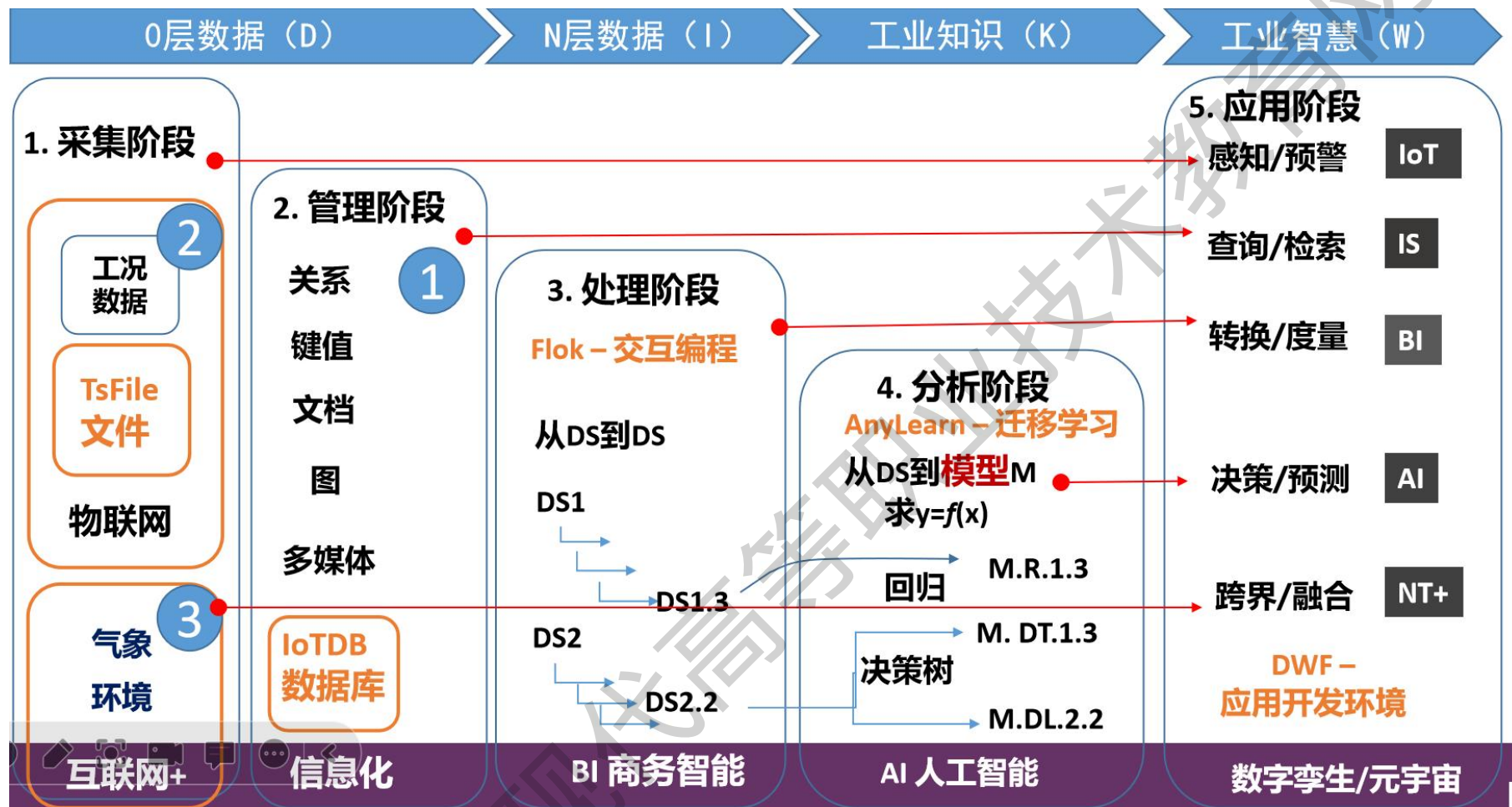
《中国制造2025》



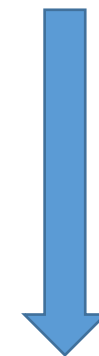
《关于深化“互联网+先进制造业”发展工业互联网的指导意见》



大数据价值创造过程



大数据
关键技术



赋能专业
职业技能

研制“数为”大数据系统软件栈，为工匠人才提供“用武之地”

大数据
应用软件

C919试飞



歼20生产



A/C型地铁列车



M310反应堆



气象雷达



其他国家
重大工程

... ..

“数为”大数据软件核心构件

方向2：
分析挖掘

Anylearn

Flok

...

方向3：
集成处理

TsClean

AutoVis

方向6：
计算框架

MaxCompute

Spark

Flink

Tensorflow

PyTorch

...

方向1：
存储管理

IoTDB

Cassandra

HDFS

PostgreSQL

Kafka

...

方向4：开发环境

DWF

领域支撑组件

应用基础组件

大数据可选组件

大数据基础组件

方向5：可信验证

Tsmart

大数据
系统软件

大数据
计算平台

异构硬件

边缘设备

私有云

公有云

赋能大批
具有工匠
精神的大
数据人工
智能软件
专业职能
技能人才

代表性成果1：软件开发环境

大数据应用软件低代码开发工具DWF

DWF是一款大数据应用软件低代码开发工具，支持基于统一模型的异构数据源接入、可视化界面组件拖拽、低代码业务逻辑开发、应用功能模块封装与复用，具备与多种工作流及规则引擎的集成能力，降低应用软件的开发门槛，满足大数据应用系统的快速开发和迭代需求。

软件开发工作量向需求端迁移，上手快，交付效果好，可快速满足需求意图！

职业技能人才需求1—大数据技术在实体经济领域应用的“根源性”挑战



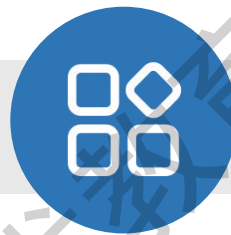
工具重

基于大数据**应用**
场景的丰富维度
和笨重的开发工
具间的矛盾



人才贵

专业化数字化技
术人才和领域复
合型**人才**之间存
在**供需的结构性**
矛盾



应用少

过剩的“平台”
产能和**稀缺的行**
业高价值应用之
间的矛盾

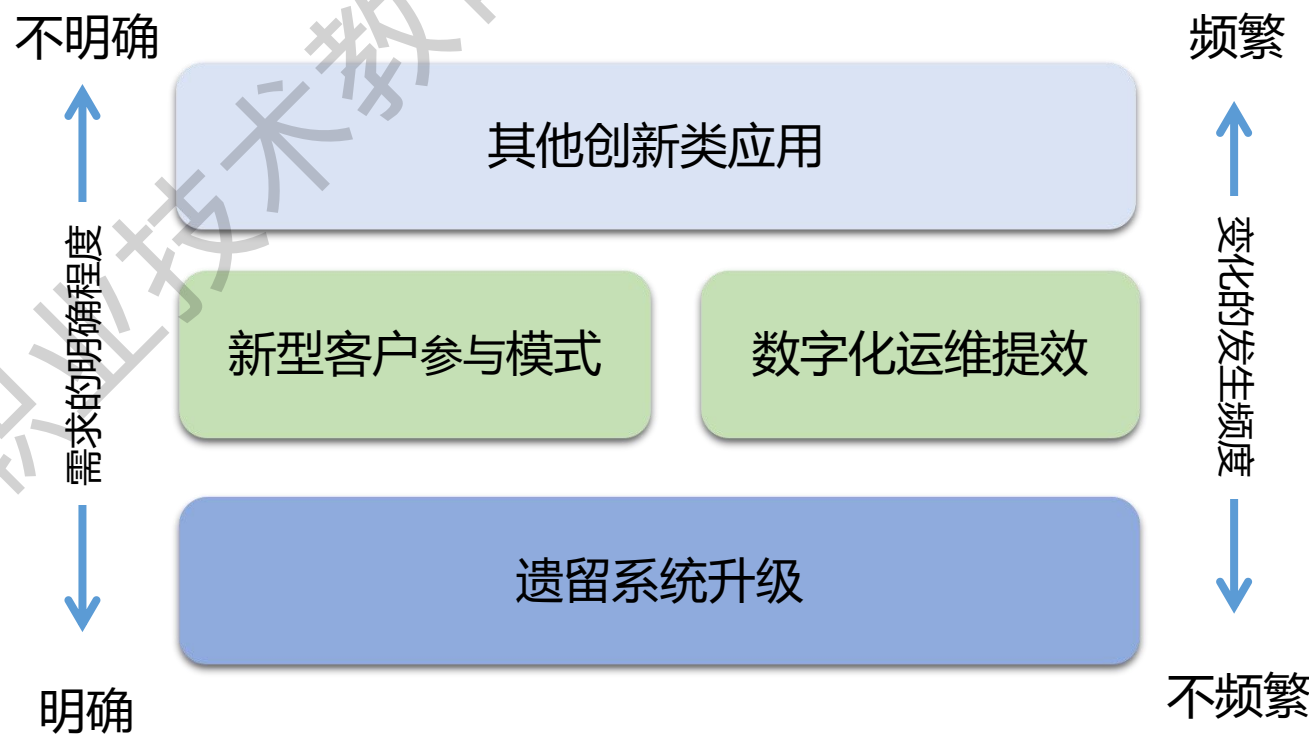


成效弱

对数字化转型的
增长期望和领域
自身内在**发展的**
客观规律之间的
矛盾

职业技能人才需求2—快速补足企业数字化转型对高技能工匠人才需求短板

- **市面上没有现成的应用程序可以直接购买**
 - 快速上云的，各类“小前台”应用迭代试错
- **现成的解决方案过于复杂和/或过于昂贵**
 - 需要购买昂贵的传统信息化软件
- **为客户提供的系统需要定制的内容较多**
 - 每个客户系统有巨大的差异性
 - 大量需求要定制开发，不得不雇佣昂贵的开发人员
- **希望低成本控制未来数字化的发展能力**
 - 例如：国家管网数字化公司



职业技能人才需求3—以低代码技术为支点，撬动高技能“技术+运营管理”复合型人才培养

- 低代码产品击中了企业在数字化转型过程中普遍存在、大量重复、功能简单的应用开发需求
- 然而，随着更加复杂的工业级数字化应用发展，低代码技术和产品还需继续向更高维度发展
 - 综合新的技术：大数据、人工智能、物联网
 - 开拓新的领域：建筑、能源、交通等
 - 形成新的应用

第一代：模型驱动代码生成
(上世纪90年代)

模型驱动代码的局部代码工程 (Code Engineering)

UML: Rose, EA...

第二代：可配置系统 (2000年初)

以核心应用功能为基础的可配置系统

SAP NetWeaver、西门子TC Foundation...

用友、金蝶BOS...

第三代：框架驱动代码生成
(2010左右)

模型驱动的代码开发框架

JeeSite、JeeCG、JeePaaS、JeePlus、牛刀云..

第四代：低代码开发 (2014年后)

• 模型直接驱动，多端定制+云服务

• Mendix, Appian, OutSystem, ServiceNow

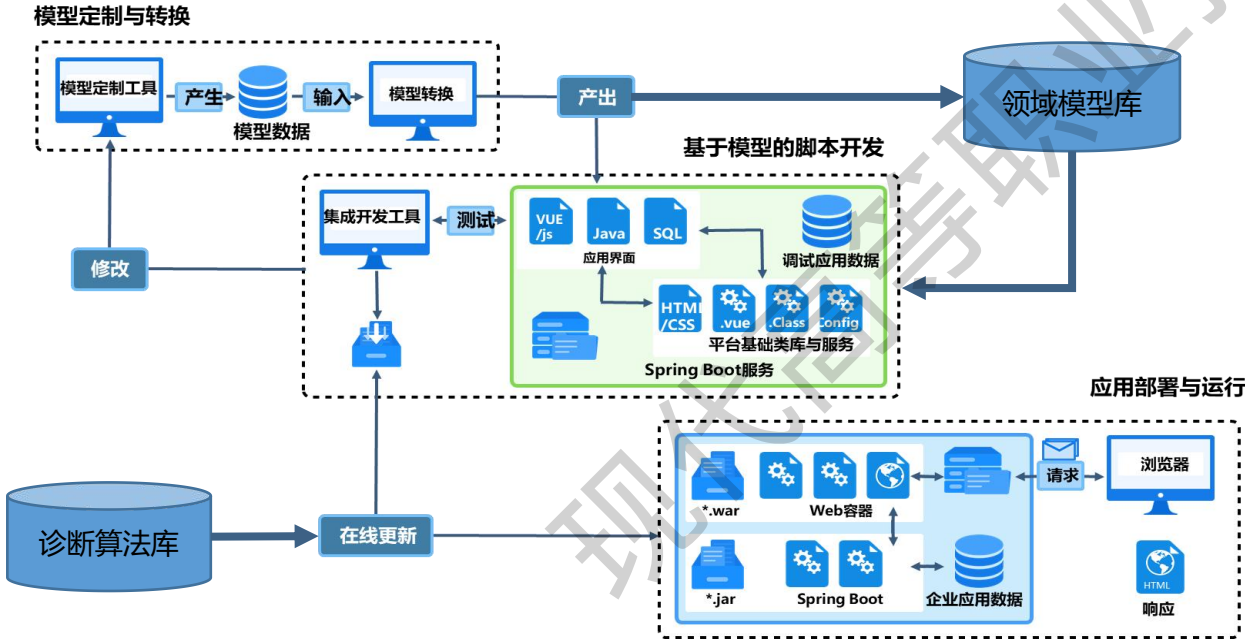
• 宜搭、简道云、搭搭云...

云计算、移动互联网带来的应用需求带动

自主研制 “数为” 大数据应用软件开发工具 DWF，打造大数据软件 “大国工匠” 利器

2019年以来，针对国产飞腾CPU和银河麒麟操作系统开展兼容性适配工作，开发了面向领域大数据应用需求的应用开发环境DataWay Framework（DWF），具有一站式、可配置、敏捷性和低码量的特点。

大数据应用快速开发



国产操作系统兼容性适配

应用证明

项目名称	基于银河麒麟操作系统的领域大数据应用开发环境建设
应用单位	湖南长城科技股份有限公司
单位注册地址	湖南长沙高新区尖山路 29 号中电软件园总部大楼
应用起止时间	2019 年 5 月至今

具体应用情况：

湖南长城科技股份有限公司（以下简称“湖南长城科技”）隶属中国电子信息产业集团，是中国电子旗下中国长城科技集团股份有限公司的全资子公司。湖南长城科技聚焦网信安全领域内的重大工程及重大项目，以维护网络和信息安全为使命，打造软硬件产品业务、信息系统集成业务、国产云计算服务、PKS（飞腾芯片+麒麟操作系统+立体安全防护）生态业务等核心主业，探索构建自主安全的信息技术体系和生态体系，为党政军以及金融、医疗、教育、交通等关键领域和重要行业提供安全可靠的国产信息技术解决方案。

自 2019 年 5 月开始，湖南长城科技开始与大数据系统软件国家工程实验室合作，针对银河麒麟操作系统开展清华数智大数据应用开发工具（Tsinghua DataWay Framework，简称 DWF）的兼容性适配工作，共同完成了《飞腾 FT1500A/16 + 银河麒麟服务器操作系统 V4》和《飞腾 FT2000 + 银河麒麟桌面操作系统 V10》的兼容性测试和适配生产环境验证，并基于湖南长城科技的信创云构建了稳定可靠、安全自主可控的领域大数据应用开发环境。

2021 年 6-7 月间，湖南长城科技利用基于信创云的 DWF 大数据应用开发环境成功支持了“2021 全球数字经济大会—清华数智为代码赛道”的竞赛活动，为全国 40 多个高校和企业的 150 多名参赛选手提供了稳定的大数据应用开发环境，扩大了基于国产操作系统的大数据应用在高校和企业的影响。

上述成果推动了国家重大科技成果—银河麒麟操作系统的工程化应用。

2021 年 8 月 6 日

项目名称：基于银河麒麟操作系统的领域大数据应用开发环境建设

共同完成银河麒麟V4和V10操作系统的兼容性测试和生产环境验证

成果推动国家重大科技成果—银河麒麟操作系统的工程化应用

代表性成果2：存储管理方向

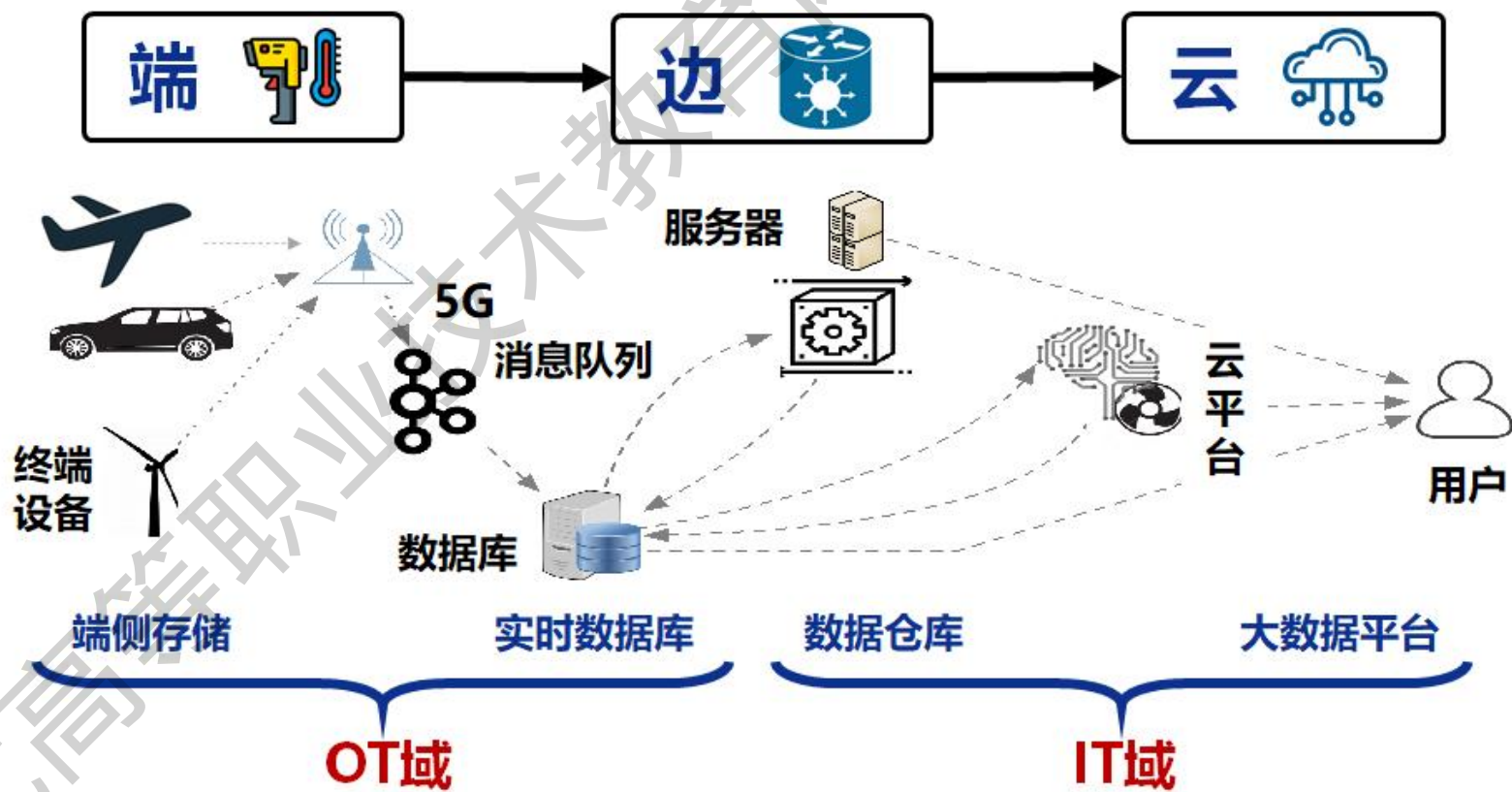
工业物联网时序数据库管理系统 IoTDB

IoTDB是一款工业物联网时序数据库管理系统，采用端边云协同的轻量化架构，支持一体化的物联网时序数据收集、存储、管理与分析，具有多协议兼容、超高压缩比、高通量读写、工业级稳定、极简运维等特点。

满足软件大数据职业技能人才可以胜任“开发+部署+管理+运维”一专多能！

工业物联网时序数据管理系统-清华数为IoTDB

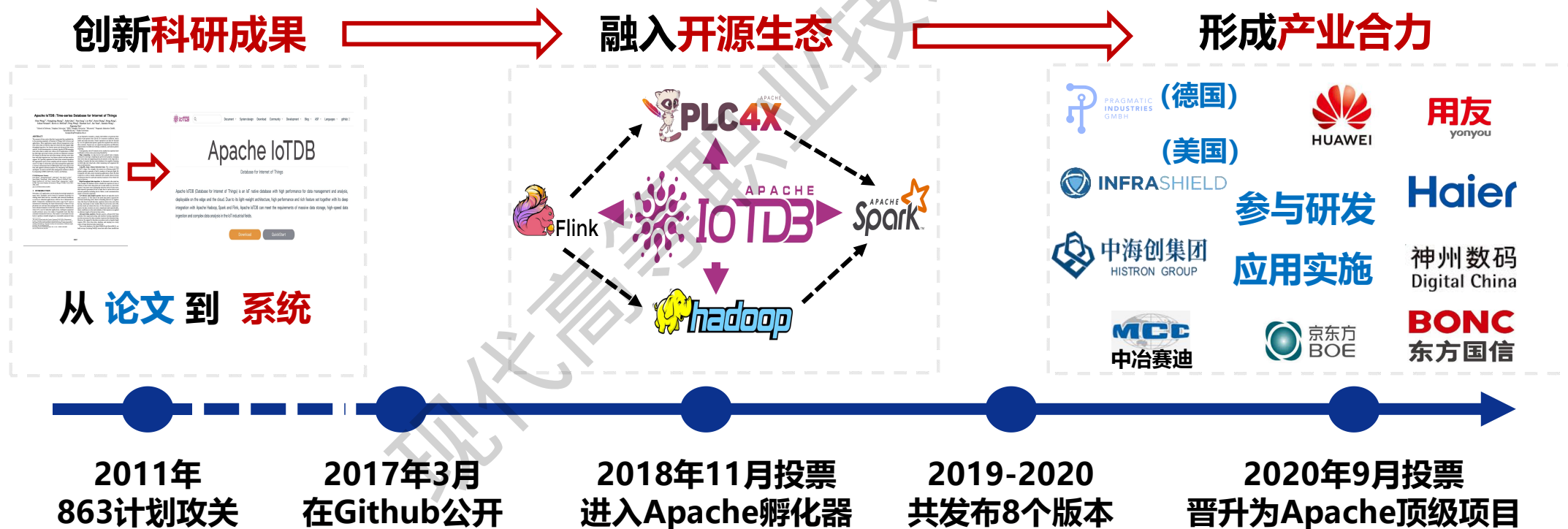
工业物联网时序数据管理需要将工业物联网看成一台计算机，因此带来了**跨越端-边-云**，融合**OT域**数据实时读写和**IT域**数据在线分析的新应用场景，带来严峻的**技术挑战**



跨越“端边云”，融合OT域实时读写和IT域在线分析

依托低代码成果开源，助力实现大数据软件产业升级转型发展

实验室研制的工业物联网数据库管理系统“清华数为”IoTDB，开源版本成为国际Apache开源社区的顶级项目，以社区为载体与华为、海尔、用友、Pragmatic等企业形成产业科技创新合力，**2020年获得北京市科技进步一等奖。**



开源模式促进IoTDB国际企业落地应用，服务“一带一路”走出去战略

IoTDB应用于德国宝马、日本小松、美国安赛乐-米塔尔、奥地利S&T、法国Grill等国际工业企业，推动全球工业数字化发展

s&t
全球最大工控机
供应商 (奥地利)

software for IoT
Semantic, IOTDB

DB 德国铁路 **BMW** 宝马

HEMA 木工机械制造商

G 法国货舱层板制造商

LAEMPE 制芯设备供应商

- BMW AG (Automotive Industry)
- Daimler AG (Automotive Industry)
- Inacore GmbH (Automotive Industry / Foundry)
- Laempe Mössner Sinto GmbH (Foundry Industry)
- Fritz-Winter Eisengießerei GmbH & Co. KG (Foundry Industry)
- The Gill Corporation (Aerospace Industry)
- Schütz GmbH & Co. KGaA (Aerospace Industry)
- Heermann Maschinenbau GmbH (Machine Manufacturing)
- Hugo Beck GmbH (Machine Manufacturing)
- DB Bahnbau Gruppe GmbH (Railway Industry)

PRAGMATIC
INDUSTRIES
GMBH

Jessingerstraße 57
73230 Kirchheim unter Teck
T: +49 7022 24 933 00
E: info@pragmaticindustries.de
W: pragmaticindustries.de

Apache IoTDB wurden initial an der Tsinghua Universität in China entwickelt und ist das führende IoT Datenmanagement System auf Grund der einfachen Geräteverwaltung, der Einsetzbarkeit in Edge und Cloud und der Zeitreihendaten Qualitätskontrolle. Zusammen mit Apache PLC4X und Apache StreamPipes bildet es die Basis des Apache Industrial IoT Ökosystems. Ausserdem lässt es sich gut mit anderen Open Source Projekten integrieren wie Eclipse Ditto oder Grafana, was das IoT Ökosystem bereichert.

pragmatic industries nimmt aktiv an der Entwicklung von IoTDB und der Community teil und integriert IoTDB in das Kernprodukt, die Plattform Digital Cockpit. Mit Hilfe des Digital Cockpit hilft die pragmatic industries GmbH mittelständischen Unternehmen aus Maschinenbau oder Fertigung bei der Nutzung ihrer Maschinendaten. Durch Apache IoTDB wird das Datenmanagement einfacher. Zum einen, weil das Datenmodell sehr gut zum Anwendungsbereich passt und weil der Betrieb im Vergleich zu anderen Zeitreihen Datenbanken wie KairosDB deutlich einfacher ist. Wir haben ausserdem einen deutlichen Rückgang in der benötigten Speicherkapazität bemerkt und sehen dass unsere Analysen schneller geworden sind.

KOMATSU
日本小松 (全球最大
工程机械制造企业之一)

2018 年清华大学软件
国际先进的水平

ArcelorMittal
美国安赛乐-米塔尔
(全球最大钢铁公司)

Replace the HBase/HDFS

BIX TECNOLOGIA
巴西工业数据
分析服务公司

I am in a project to store and analyze
wind turbines data and iotdb seems
to be a great fit.

开源文化引入职业课程体系，培育中国式开发者社区，

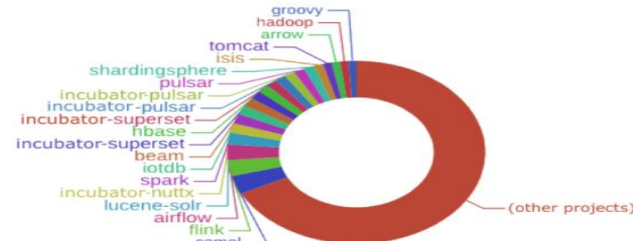
创新中国技能标准

项目活跃度

2020年在Apache 全球351个项目中 排名第七



other_projects	176549
camel	8488
flink	7407
airflow	7178
lucene-solr	5396
incubator-nuttx	4466
spark	4212
iotdb	4240
beant	4081
incub-superset	3875
hbase	3715
incub-superset	3610
incubator-pulsar	3500
incubator-pulsar	3500
pulsar	3491
shardingsphere	3412
isis	3040
tomcat	2860
arrow	2739

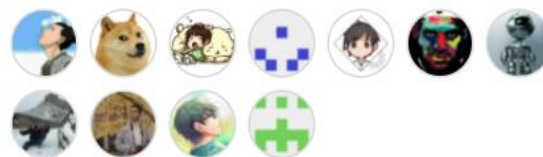


数据来源：最新Apache基金会年报

代码贡献者

来自中美德英澳
等国138人

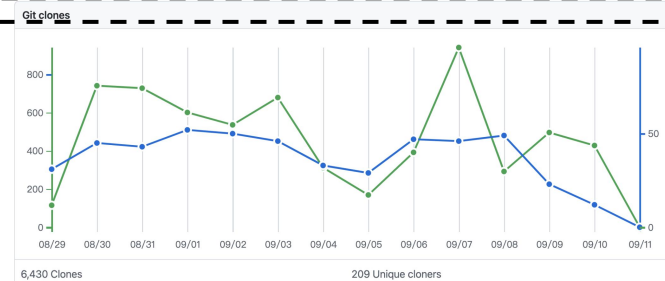
Contributors 138



信息来源: Github统计

用户
下载量

源码下载量次数
每周超6000次



信息来源: Github统计

代表性成果3：大数据处理

大数据处理流程管理系统FloK

FloK是一款大数据处理流程管理系统，通过算子拖拽实现数据处理流程交互设计，支持计算环境配置、算子规范化定义、流程断点调试，满足复杂流程搭建与迭代过程中计算框架混用、流程模型复用、算子资源预测等需求

赋能软件大数据职业技能人才在数字化全流程、全要素、多角色方面的能力体系

清华数为FLoK—从简单概念入手， 据分析人才培养难度

变“复杂”为“简约集成”，降低大数

- 数据

- **数据源**：包含数据连接信息，例如连接串、用户、密码等
- **数据集**：在数据源的基础上，包含读取数据的方式和内容

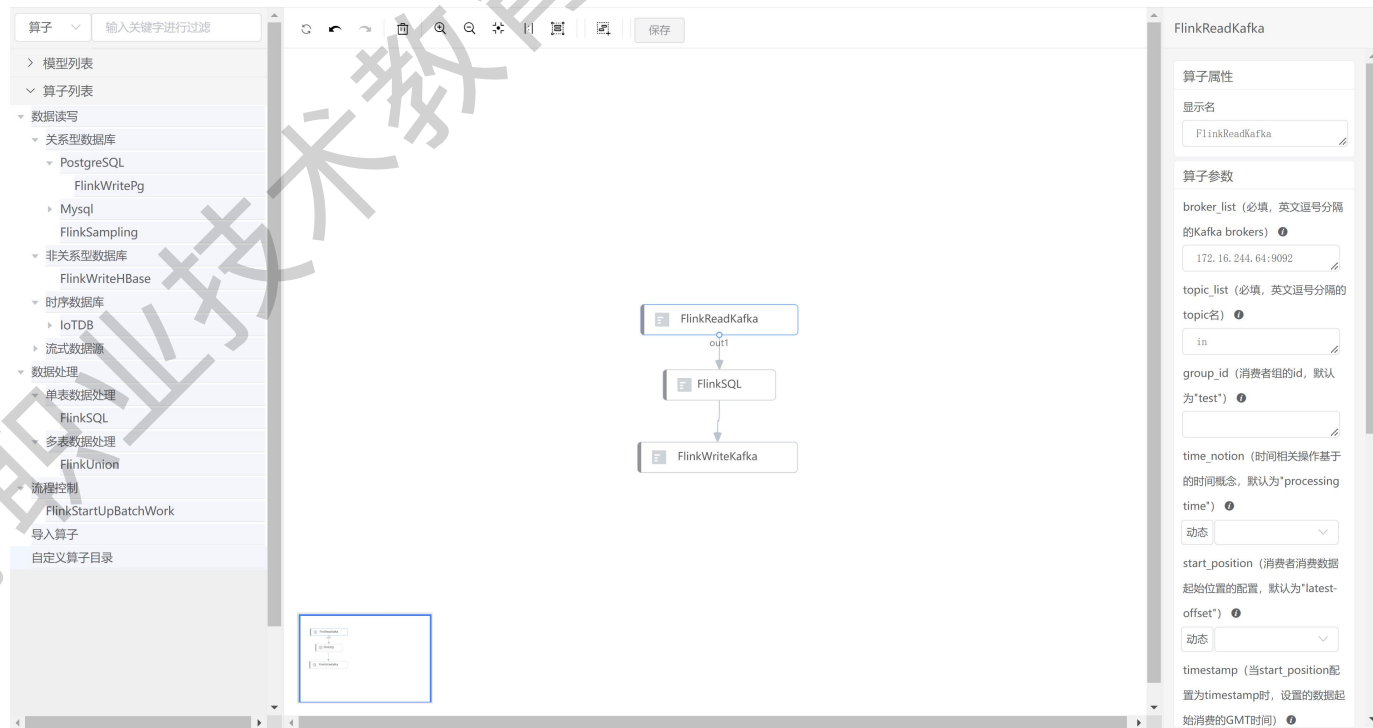
- 算子

- **输入**：算子处理前的原始数据
- **参数**：算子暴露给用户的可变部分，以提高算子通用性
- **输出**：算子处理后的结果数据

- 环境

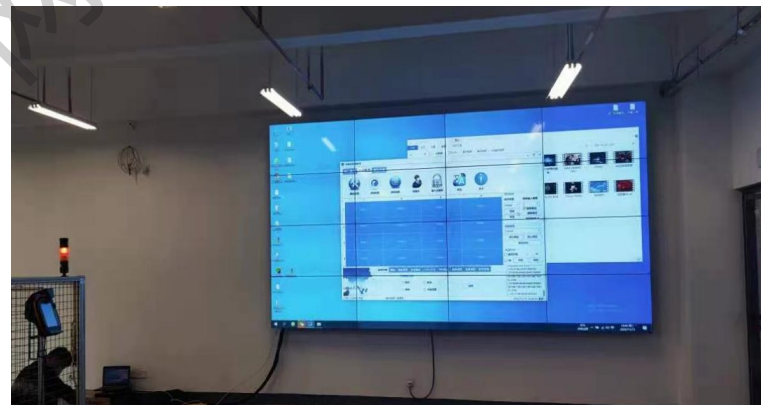
- **存储**：用于保存算子的输入、输出数据
- **计算**：将算子提交到依赖的平台去执行

- 流程



支持职业技术学院跨学科大数据分析能力培养

应用落地：莆田学院工业机器人研究与创新实训平台系统数字化升级项目



代表性成果4：数据智能方向

大数据机器学习研发管理系统 Anylearn

Anylearn是一款大数据机器学习研发管理系统，支持数据集、算法族、模型库等资产管理，支持机器学习研发的过程管理、知识沉淀、模型迁移，满足资源统筹利用、团队高效协作等人工智能工程化需求。

赋能软件大数据职业技能人才在人工智能前瞻性、先进性方面的能力需求体系

机器学习技术在工业场景中应用：从复杂性到模型化泛化

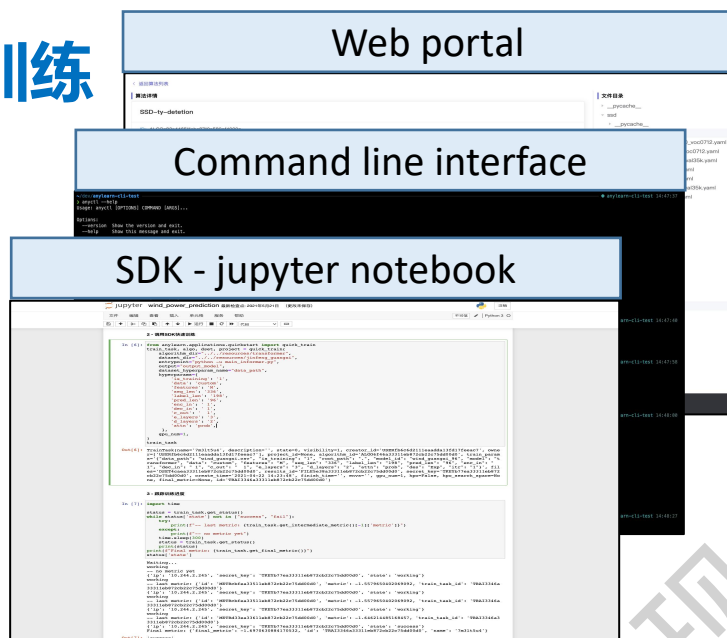
人工智能应用普遍面临“**成本高，落地难**”痛点，源自于机器学习技术的三维度复杂性

模型开发方式：交互多样化

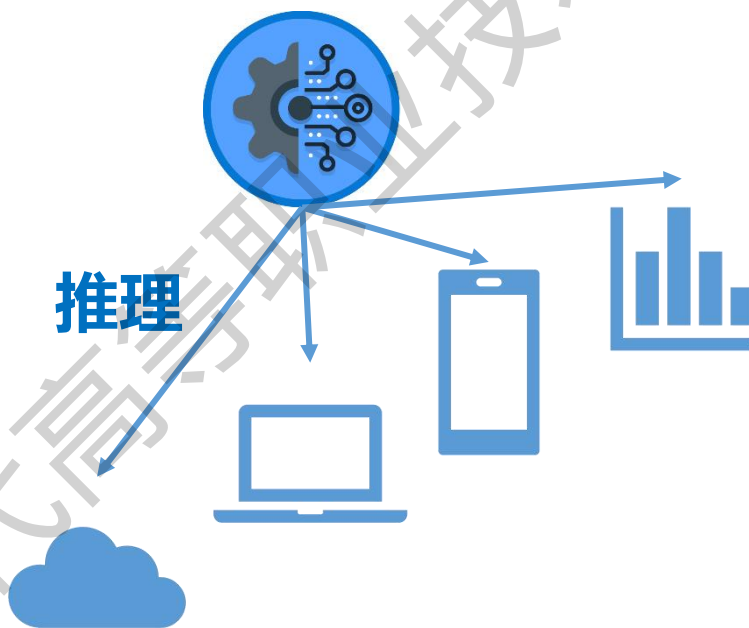
模型部署环境：端边云统一

模型泛化能力：多领域适应

训练



推理



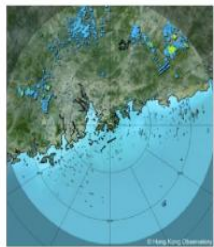
泛化



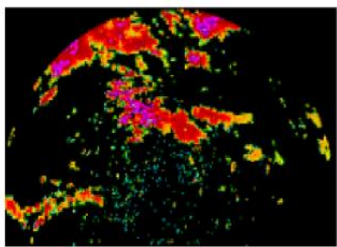
可满足软件大数据职业技能人才快速掌握、敏捷操控之需求

案例：支撑“金云工程”和“雷达专项”等国家重大工程

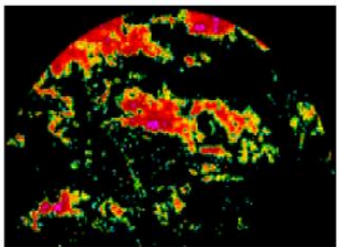
作为中国气象“金云工程”核心系统，支撑全国各级气象预报；支撑国家发改委“雷达专项”，中国区预报效果超越DeepMind技术；首个中文超大规模多模态预训练模型“文澜”效果超越OpenAI技术



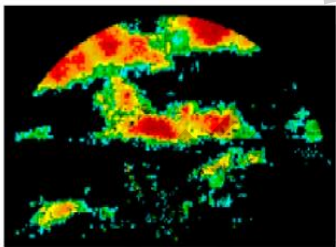
国家发改委“雷达专项”



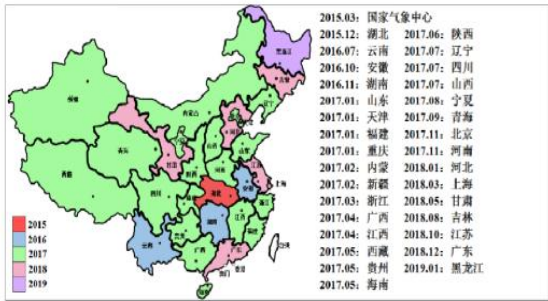
雷达实况观测



清华预报技术



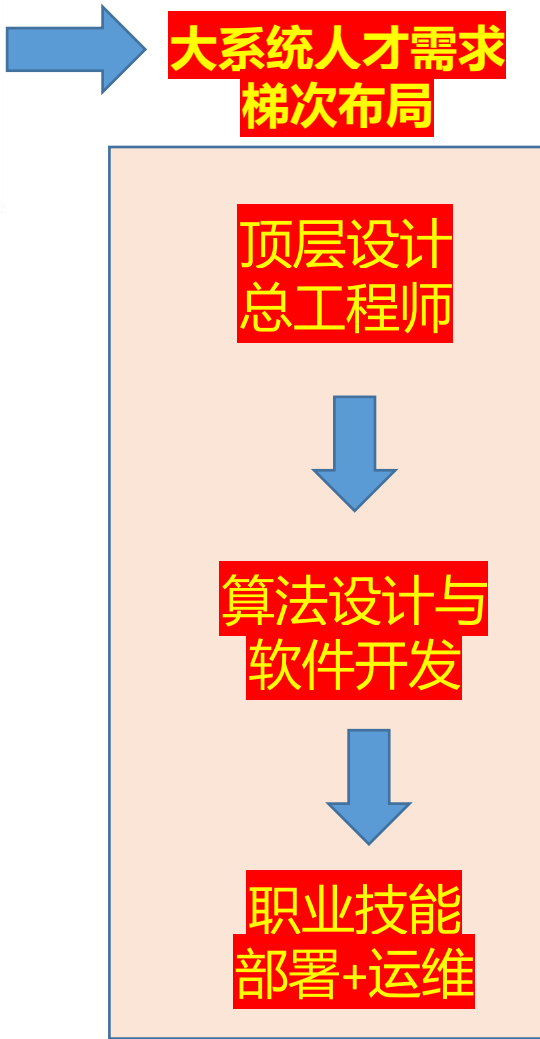
DeepMind预报技术



中国气象“金云工程”部署路线图

评测任务	图检索文			文检索图		
评测指标	R@1	R@5	R10	R@1	R@5	R@10
OpenAI CLIP (ViT-large)	63.2	88.4	94.4	49.6	79.2	88.2
Google ALIGN (EfficientNet-B7)	63.5	88.6	95.0	49.2	79.0	87.3
文澜第二代 (EfficientNet-B7)	67.4	90.0	94.8	50.3	79.5	87.9

首个中文超大规模多模态预训练模型-文澜



主要内容

1

清华大学软件学院和职业教育协同发展的趋势

2

清华数为大数据软件栈赋能职业院校人才培养

3

清华数为大数据应用课程体系和合作成果

4

推动双高院校大数据智能方向发展的展望

从2020年开始针对职业院校组织师资班培训

清华数为大数据软件栈DWF第一期培训班



2020年首期工业软件与工业互联网大数据骨干教师研修班



中国移动 14:47 西航工业设计 >

受益匪浅，不虚此行

DWF培训和学习（陈琳老师、刘英博老师）



大数据应用、前沿知识讲座

面向不同层次、不同专业要求的授课体系



入门级—适合职业技术学院跨专业的专科、本科的**小学期**课程实践

- 无需编程基础，使用零代码定制的技巧完成简单应用的开发。
- 使用DWF模型定制能力建模快速搭建应用系统，帮助形成软件开发的感性认识和基础能力。



进阶级—适合职业技术学院跨专业的专科、本科的**正式课程**

- 在专业知识基础上，了解模定制的基本思想，形成针对实际问题的分析能力。
- 掌握软件建模的具体方法，并在实训环境上能够完成定制。
- 学习基础的脚本开发技巧，掌握基于复杂数据模型的软件开发技能。



专业级—对接企业需求，定制化的专业，专科、本科高年级**实践课程**

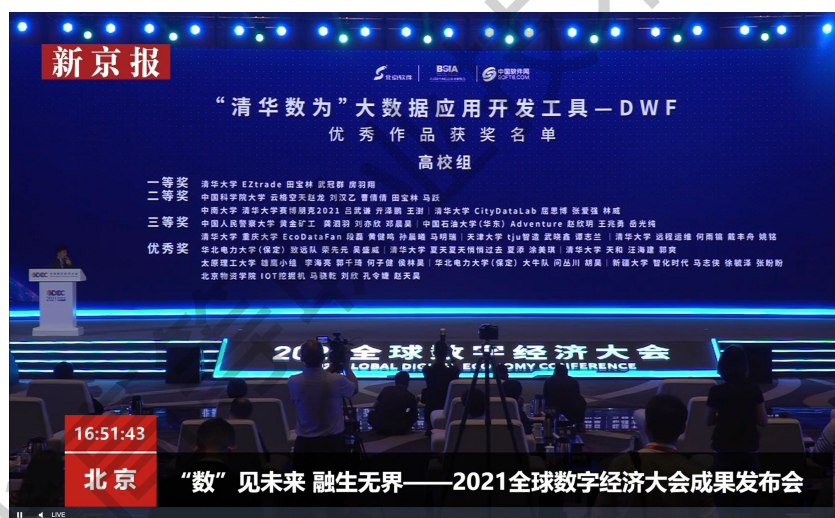
- 在具有软件专业技术能力的基础上，能够掌握复杂应用程序的扩展开发能力。
- 能够针对高级需求形成完整的技术路线，实现与大数据、人工智能、物联网的对接。
- 能够和其他专业同学形成跨学科协作团队，共同解决问题。

DWF支持全球数字经济大会和全国高校教师创新大赛

- **已陆续开展5届培训班，支持3次全国性疾病，总计培训人次达到855人次**
- **同时，配套发展了1套实训管理平台，可实现任课教师自助服务完成实训**



2020全国高校教师创新大赛

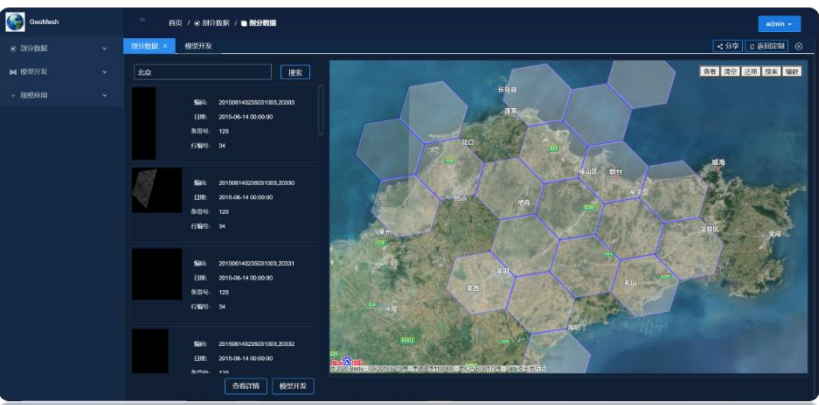


2021全球数字经济大会



2022全球数字经济大会

竞赛学生和企业的代表性作品



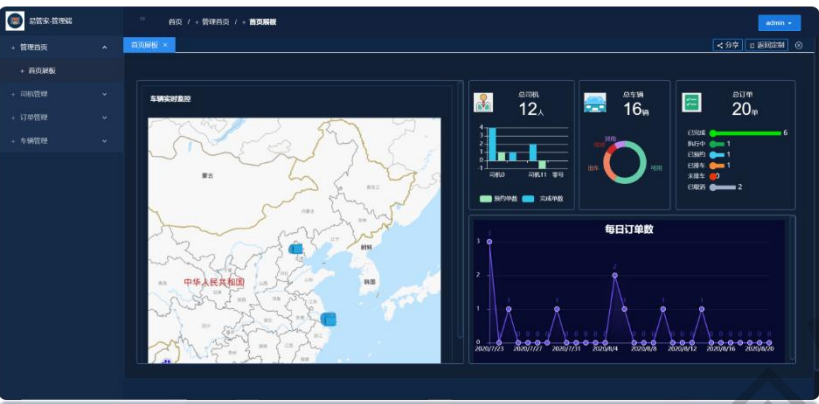
陆地卫星数据集成
中科院遥地所



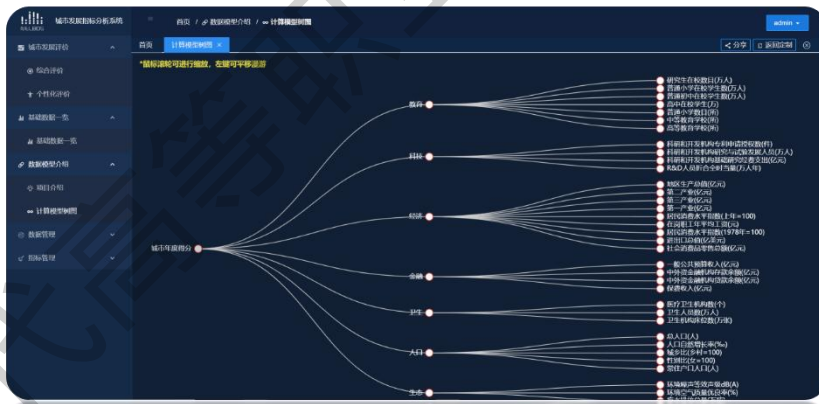
智慧工地建材质量检测
广西建院



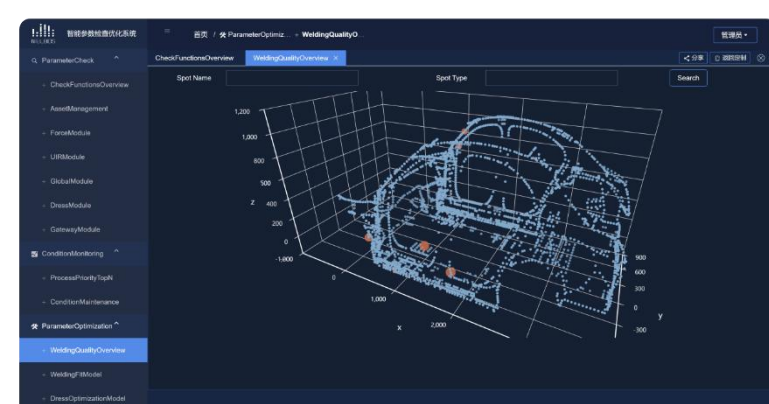
数控机床联网健康评估
天津大学机械系



校园车队智能管理系统
清华大学公共管理学院



城市整体指标评估
清华大学工物系



汽车点焊数据综合评判
北京奔驰

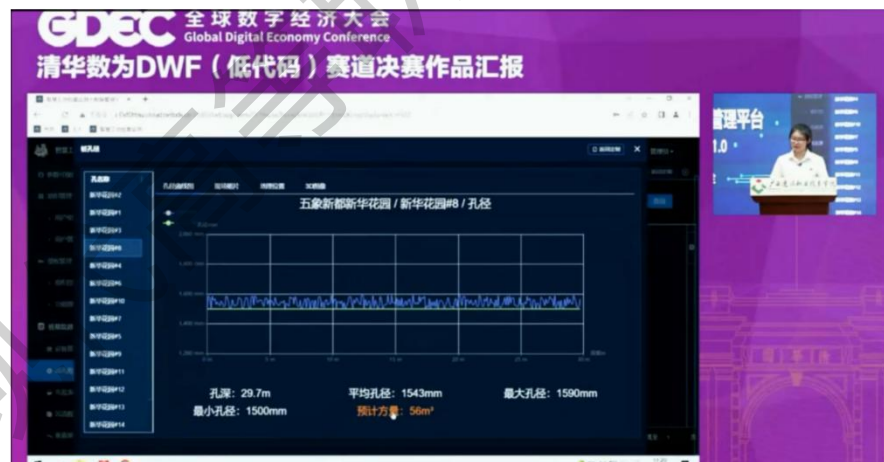
2022年：广西建职院获奖作品

- 2022年全球数字经济大会比赛，后发先至，特色突出，得到评委的一致认可



三是应用作品质量高，此次大赛的许多参赛作品，均体现出了较高的技术性、创新性、应用性。以清华数为DWF赛道为例，一等奖应用作品《智慧工地建材构件质量检测评估与溯源大数据管理平台》、二等奖应用作品《智慧工地桩基监测大数据管理平台--DWF PileTesting Sys V1.0》均获得评审专家一致认可，具有创新与应用价值，参赛得分在90分以上。

作为2022全球数字经济大会北京数字经济



主要内容

1

清华大学软件学院和职业教育协同发展的趋势

2

清华数为大数据软件栈赋能职业院校人才培养

3

清华数为大数据应用课程体系和合作成果

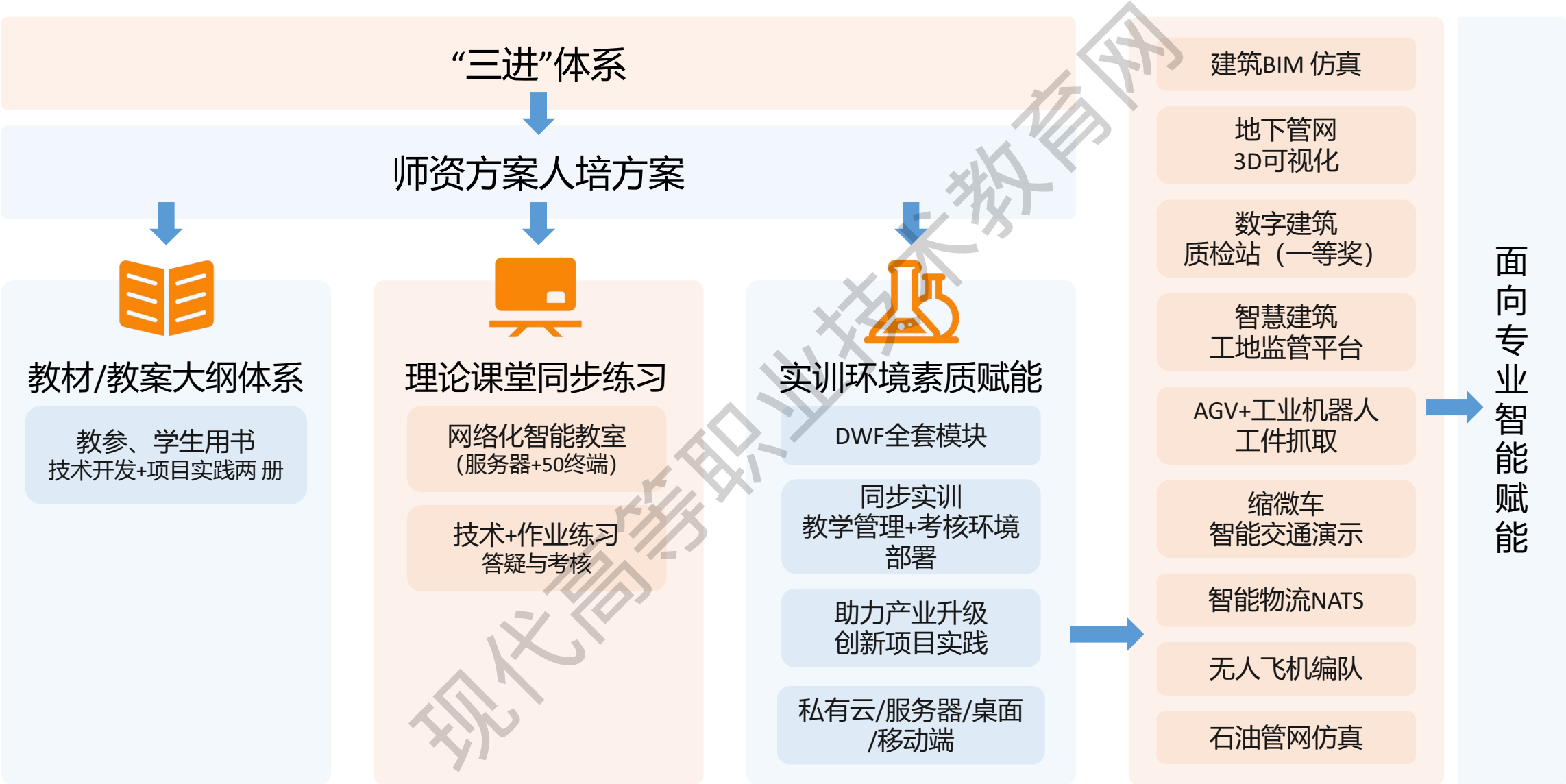
4

推动双高院校大数据智能方向发展的展望

工作展望部署：

**DWF低代码开发技术课程“三进”体系统筹
及助力产业升级技能人才培养**

工作展望部署：DWF低代码开发技术课程“三进”体系统筹及助力产业升级技能人才培养



DWF低代码开发技术课程“三进”体系统筹

助力产业升级创新人才需求...

应用教学：广西建设职业技术学院 DWF低代码开发技术课程



实施单位：广西建设职业技术学院信息工程学院

培养学生：人工智能专业方向（2022级）

理论教学：应用基础 + 低代码模块 + 扩展脚本

实训作业：大数据智能低代码开发（大作业）



感谢聆听，谢谢！

清华数为：数据之路，有所作为！