



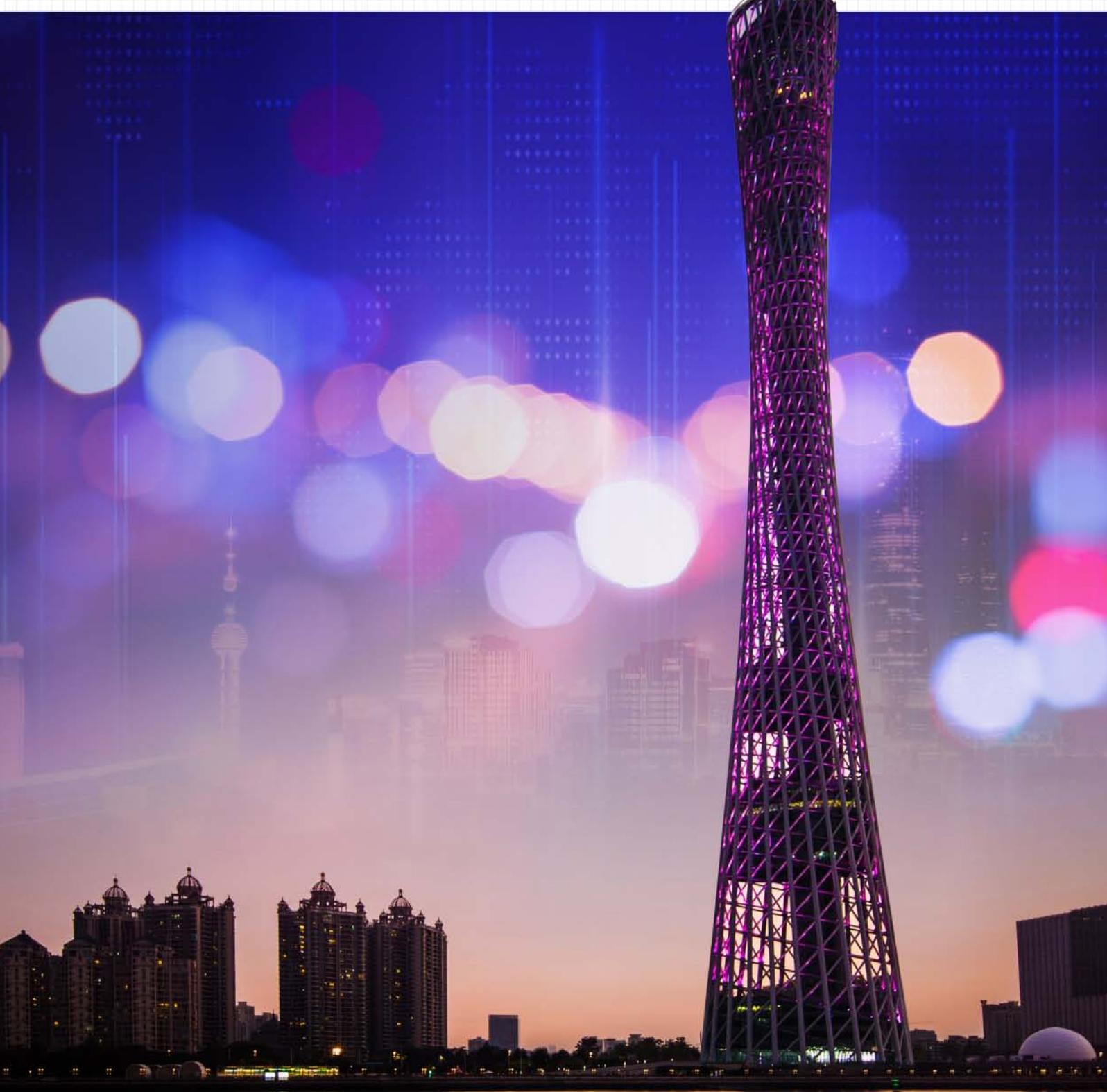
数据分析师

CHINA DATA ANALYSIS 数据分析·因你而不凡

—中国数据分析行业核心刊物—



《中国数据分析》行业特刊
2022年第02期 总第50期(季刊)
咨询热线: 400-050-6600
<http://www.chinacpda.org>



中国商业联合会数据分析专业委员会 主办

产学研融合创新模式 培养应用型数据人才

创新之道，唯在得人。随着中国数据科学产业的蓬勃发展，数据人才的匮乏，人才培养是助力大数据产业发展的重要之举，推进产学研融合，培养更多应用型的数据人才，帮助企业提升数据科学水平。

依托数字化转型，新旧动能加速转换，产学研融合将成为提高新兴产业人才培养质量的重要模式。作为数据分析行业协会，组织共建中国数据分析行业案例库的工作，就是始终秉持以数据分析思维为中心，以实践应用能力为导向的理念，专注数据人才培养，探索和践行产学研融合创新模式，助力培养高端数据创新人才，以人才赋能支持中国产业升级。

随着新兴科技日新月异，产学研融合成为转型升级的助推器、促进就业的稳定器和人才红利的催化器。从行业角度分析，共建案例库通过学科融合、实战教学、专业师资打造等方式，使人才培养紧密契合产业发展态势，企业是学术研究的稀缺资源，产业师资、案例都需要加强共创共享。行业协会利用自身的资源整合、学术研究等核心能力，将行业内的数据分析师事务所、大数据公司的资源转化为符合教育规律的知识体系，最大程度共享和优化配置产学资源，最终再回馈到企业中，助力企业数字化转型，同时帮助数据人才在真实的企业环境中锻炼实践能力，为行业发展蓄力，这本身是适应时代发展的新形势。

产学研模式的创新，是打造人才培养实战的新路径。实用人才的培养离不开“实战”的磨炼。我们基于行业应用型人才的培养特点，将发展方向对接行业标准，将培养方向对接从业过程，通过对实际案例的分析，将数据分析标准化流程、业务范围及收费标准、工作模式与项目运营管理、项目交付流程标准贯穿到实际应用中，这种模式将学习工作合二为一，在实际工作和项目中帮助分析师快速将理论知识转化为应用能力，也能通过直面行业优质人才，获得宝贵的技能和经验。

技术的赋能，也是助力人才培养模式的新突破。当前，随着大数据技能供给不断更新迭代，我们打造了可提升学习与实际应用相结合的智能分析平台，充分利用大数据新技术，通过数据洞察和分析学习数据、感知场景、与实战技能训练相结合，助力并延伸产学研模式的发展思路，构建创新生态，激发协同发展的新力量。

内化于心，才能外化于行；不忘初心，才能践行使命。行业协会历经 14 年，始终在人才培养的道路上不断深耕、创新，汇集行业各方面资源，推动产学研密切融合，推动行业资源共享，构建未来协同创新模式，培养社会应用型数据人才。我们始终以前锋者的姿态，披荆斩棘，行稳致远。

中国商业联合会数据分析专业委员会



本期目录 CONTENTS

卷首语

- 01 产学研融合创新模式 培养应用型数据人才

协会动态

- 03 4月14日数据分析行业线上分享活动：数据分析师的“面试经验”
4月28日数据分析行业线上沙龙活动：“从人事分析学两种分析思路”
04 5月19日 数据分析行业线上公益沙龙活动：“如何做数据分析报告”
04 数据分析需求登记窗口：帮助解决数据分析业务需求
06 2022 帆软 BI 数据分析大赛：推进大数据优秀人才培养，数据委助力会员单位活动
07 会员专属社群知识分享服务：PowerBI 工具讲解
08 解读《数据分析行业服务参考文件》系列采访：
——上海荣增创始人冼峰先生
10 解读《数据分析行业服务参考文件》系列采访：
——帆软数据应用研究院研究主管王文信
11 解读《数据分析行业服务参考文件》系列采访：
——北京派可数据联合创始人吕品
14 解读《数据分析行业服务参考文件》系列采访：
——行标专家指导组成员陈虹坚

政策向导

- 15 湖北省大数据产业“十四五”发展规划
22 发改委：正在牵头制定数据要素基础制度文件
23 国知局：将加快推动数据产权保护制度的完善
24 中央网信办曹淑敏：互联网企业要加大研发投入，让创新成为发展的核心竞争力

行业动态

- 25 中国信通院闫希敏：个人信息保护视域下知情同意框架的应用困境与对策探析
28 肖亚庆：构建稳定高效的大数据产业链要从两个方面重点着力
29 用好大数据，助力市场数字化升级

学“数”交流

- 30 非线性回归充电之：时间序列分解

事务所专栏

- 36 中国商业联合会数据分析专业委员会会员单位
——数据分析师事务所会员
39 数据分析师事务所创业指导“圆桌派”第一期 Q&A



主办单位

中国商业联合会数据分析专业委员会

编委成员

会员处 李苗苗

出版时间

2022年06月出版 <总第50期>

美工设计

市场处 崔峻珩

联系我们

中国商业联合会数据分析专业委员会
地址：北京市朝阳区朝外 SOHO-C 座 9 层
电话：400-050-6600 / 010-5900.0991 转 652
传真：010-59000991 转 607
官网：www.chinacpda.org

欢迎广大读者踊跃投稿，内容包括学术观点、教学体验、教学活动、学习感悟、实战经验、随笔文章等。

稿件附图格式为 JPG 或 TIFF 格式，大于 1M，分辨率在 300dpi 以上。

感谢您对《中国数据分析》的支持！ 投稿邮箱：xiehui@chinacpda.org

4月14日数据分析行业线上分享活动： 数据分析师的“面试经验”

来源 / 中国商业联合会数据分析专业委员会 编辑 / 数据委员会处 李苗苗 日期 / 2022-04

4月14日，数据分析行业公益沙龙活动如约而至，离开舒适区，选择转型数据分析，绽放自身价值，实现职业梦想。众多数据分析师一路走来，沉淀了许多宝贵的经验。

正值金三银四招聘季，中国商业联合会数据分析专业委员会邀请某制造企业中层leader-穆桐、某国有通信企业数据分析师-陈嘉祺、某世界100强公司data engineer-贾小丽以及百特投资（中国）有限公司高级BI开发工程师-王杰斯四位优秀的数据分析师，为数据分析行业会员及数据分析爱好者们带来一场关于“面试经验”的主题沙龙活动。

在一个多小时的分享里，四位分析师分别以自己的角度阐述观点，分享经验，干货满满，内容包含：企业人员招聘考核能力总结，面试技巧；面试常见问题分享，详细解答；如何利用Tableau制作数据看板；传统制造企业背景下，如何利用数据分析进行中层管理等等充满干货的主题分享，是数据分析师们工作中密切相关的问题，深受我会会员及数据分析爱好者的关注和好评。

4月28日 数据分析行业线上沙龙活动： “从人事分析学两种分析思路”

来源 / 中国商业联合会数据分析专业委员会 编辑 / 数据委员会处 李苗苗 日期 / 2022-04



4月28日，我会邀请九数云产品运营经理洪映老师为大家分享主题为“从人事分析学两种分析思路”的直播沙龙活动，颇具场景化的主题内容选题受到大家热烈关注。

直播间主持人首先对分享嘉宾洪映老师进行了简单介绍，作为九数云产品运营经理，洪映老师对大型企业数字化管理有着丰富经验。通过两个企业人事场景下的数据看板案例，洪映老师对企业人事数字化分析做什么，怎么做，前景好不好，都进行了重点讲解。

本次活动，洪映老师通过“人力资源盘点”“人事考勤

分析”两个数据看板案例介绍，为大家讲解基础的数据分析思维，非常适合入门新手以及对数据分析感兴趣的同学了解学习。讲解过程中，同学们在留言区提问不断，“数据指标怎么制定？”“BI产品工具的选择？”答疑环节未到，主持人整理出的留言问题就达到了近20个。在直播间主持人控场下，同学们就各自工作中容易遇到的典型问题和场景与洪映老师进行了请教探讨。洪映老师的精彩解答也让同学们共鸣满满，纷纷在留言区分享自己在业务处理中的经验，整场直播取得良好效果。

5月19日 数据分析行业线上公益沙龙活动： “如何做数据分析报告”

来源 / 中国商业联合会数据分析专业委员会 编辑 / 数据委员会处 李苗苗 日期 / 2022-05

5月19日，数据分析师胡红艳老师精心准备的“数据分析报告到底该怎么做”数据分析行业沙龙直播正式跟大家见面。颇具业务实战思想的内容选题受到大家热烈关注，90分钟时间里，大家积极在直播间与胡老师进行了精彩互动和话题探讨。

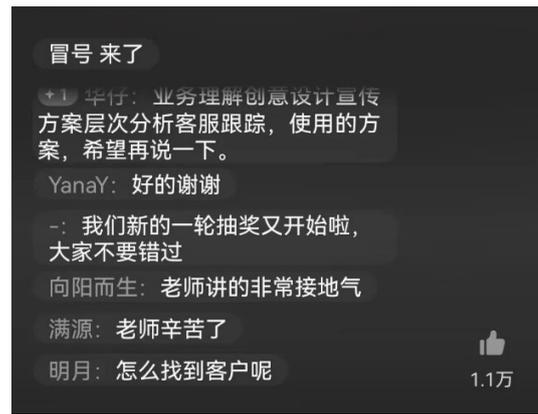


直播间主持人首先对分享嘉宾胡老师进行了简单介绍，胡红艳老师毕业于对外经济贸易大学，曾任职于北京国际鲜花港数据中心主任、中旅北京生态旅游集团数据中心副主任，长期致力于数据分析的实践应用研究。

通过本期分享案例，胡老师以幽默风趣、深入浅出的方式为大家详细梳理了数据分析报告的撰写思路，在实际工作场

景下数据分析报告撰写中可能遇到的各类障碍，希望借此能为大家在数据分析学习和实际工作中提供一些实用的帮助。

直播过程中，大家在留言区提问不断，“数据治理方案应该如何选择？”“怎么制定工作计划？”“怎样从业务角度去理解数据？”答疑环节未到，主持人整理出的留言问题就达到了十余个。



在直播间主持人控场下，大家就各自工作中容易遇到的典型问题和场景与胡老师进行了探讨。胡老师的精彩解答也让大家共鸣满满，纷纷在留言区分享自己在业务处理中的经验，整场直播共取得良好效果。

数据分析需求登记窗口： 帮助解决数据分析业务需求

来源 / 中国商业联合会数据分析专业委员会 编辑 / 数据委员会处 李苗苗 日期 / 2022-05

自数据委搭建数据分析业务需求登记平台以来，解决了不少企业中的数据分析相关问题，近日，我们就大家较关心的一些问题进行分享：

1. 企业有数据，但是体量不大，不知道是否可以登记？

这种情况协会能帮助解决吗？

可以的，无论数据体量多少，都可以登录数据分析需求登记窗口提交需求，我们会根据需求情况安排合适的会员单位对接。

2. 我们企业有服务需求，但是费用预算有限，不知道是否可以登记？

可以的，我们对预算没有具体的要求和限制，有数据分析相关业务都可以提交。会员单位将根据服务内容提供对应付费解决方案。会员单位承接业务时，需遵守数据委2022年发布的《数据分析行业服务参考文件》中业务范围与收费标准等服务内容指导。

3. 目前有项目也有预算，想了解下从数据委平台提交需求登记与在社会上寻求服务方有什么区别？或者优势？

(1) 在数据委平台提交的业务需求经审核通过后会转至数据委员会单位。数据委员会单位为专业数据分析师事务所会员、大数据公司、咨询公司组建的行业会员，会员单位严格遵守数据委《会员管理办法》、《数据分析行业自律公约》等相关要求。数据委每年对会员单位进行年检，年检考核标准从研究质量、承接业务情况、客户满意度、行业影响力等多维度制定，因此数据委员会的服务能力与质量是有一定保障的。

(2) 数据委员会单位在承接执行业务时须遵守数据委编制的《数据分析行业服务参考文件》中对数据分析业务、收费参考标准等服务标准。因此客户在寻求服务过程中会有明确的服务标准有据可循。

(3) 数据委员会单位在承接由数据委平台接收的业务时将提供低于市场的价格，在服务性价比方面给客户提供保障。

4. 如何提交业务需求？具体操作方式有吗？

我会业务需求登记窗口是为了搭建数据分析业务需求对接平台。我们将收到的客户业务需求经初步审核后发至我会会员单位，包括数据分析师事务所会员与行业会员，如有承接意愿的会员单位，会联系客户并进行沟通接洽。以下为提交业务需求流程：

(1) 登录协会官网-会员专区-数据分析需求登记窗口。

(<http://www.chinacpda.org/project/index.html>)：通过手机号验证码即可登录（后续可登录进行查询结果、修改内容等操作）。



(2) 完善基本信息并提交数据分析业务需求：

业务需求包括但不限于：咨询服务、基于数据做分析、数据调研、构建算法、解决方案、数据分析报告、研发大数据分析平台、研发数据分析产品及其他。

需注意：

- 在提交业务需求相关信息时，请对数据进行脱敏，若有

敏感信息请重新提交；

- 我会仅搭建提交业务需求的平台，对提交到该平台的信息或数据不做任何保密性承诺，请谨慎提交；当我们把需求及联系方式给到会员单位时，原则上我会服务结束，后期业务推进我会不再介入。



(3) 我会对业务需求进行初审，通过后需求转交会员单位：

符合我会各会员单位业务范围的，经我会审核通过后，将需求转交到相关会员单位，如有会员单位承接，一般3-5个工作日内会与需求方取得联系。

我会不承诺会员单位提供公益服务或一定可以解决问题，具体合作模式需要双方沟通推进。

对于提交需求审核不通过，可根据我会反馈结果进行调整后重新提交。

(4) 与会员单位沟通具体需求及合作方式：

有承接意愿的会员单位，会主动与业务需求方取得联系商讨具体业务细节，此过程我会不再参与。

我会会员在提供服务过程中，如对服务不满意，可向我会会员处反馈。

2022帆软BI数据分析大赛： 推进大数据优秀人才培养，数据委助力会员单位活动

来源 / 中国商业联合会数据分析专业委员会 编辑 / 数据委员会处 李苗苗 日期 / 2022-04



中国商业联合会数据分析专业委员会走专业型协会、靠专业引领市场已经发展近15年，为中国大数据领域发展、为广大企业数字化转型、为社会培养数万名优秀的分析师，得到广大社会的一致认同。

帆软软件有限公司于2021年6月正式加入我会，成为众多会员单位之一。帆软是国内知名的BI厂商，专注BI十六年，为超过18000家合作伙伴提供全面、专业、高效的大数据BI产品，在业界享有一定知名度。

“帆软BI数据分析大赛”是帆软公司一年一届的大型大赛，此大赛面向全社会相关领域的从业者和爱好者，集竞赛、培训、交流于一体，旨在引领开启关注数据价值、释放数据无限潜能的科技风尚。为支持会员单位办好此次大赛，帮助相关从业人员提升数据分析技能，贡献数据分析人才培养力量，我会数据分析师认证品牌——CPDA数据分析师以特别支持单位身份助力此次大赛。



3月底，2022年“帆软BI数据分析大赛”正式开启报名阶段！CPDA为此次大赛最终获奖参赛企业、个人策划提供了包括CPDA数据分析师认证培训基金、数据分析实战必备课程基金等丰厚奖励，同时也希冀为参赛者提供动态行研分析与优质项目的对接机会，成为数据人才与企业需求间的桥梁，为国内数据分析行业发展助力。

会员专属社群知识分享服务：PowerBI工具讲解

来源 / 中国商业联合会数据分析专业委员会 编辑 / 数据委员会处 李苗苗 日期 / 2022-04



为了更加方便会员利用碎片时间学习、巩固大数据应用技能，中国商业联合会数据分析专业委员会会员处特别邀请深受数据分析师喜爱的李妹老师，在会员专属社群内为大家带来线上PowerBI工具讲解系列分享活动，内容涉及到数据连接、可视化视图操作、数据报表制作及实践案例等，老师采用语音图片方式带着大家一步步操作讲解，并且在群内和

大家“零距离”沟通交流。会员社群知识分享活动是我会会员众多服务中的一项，深受会员赞许和支持。此会员活动将不定期持续开展。



解读《数据分析行业服务参考文件》系列采访： 上海荣增创始人冼峰先生

来源 / 中国商业联合会数据分析专业委员会 编辑 / 数据委员会处 李苗苗 日期 / 2022-04

解读《数据分析行业服务参考文件》系列采访 · 本期嘉宾



上海荣增信息科技有限公司创始人、董事长、CEO中国商业联合会数据分析专业委员会顾问冼峰先生

冼峰先生毕业于厦门大学物理系，荷兰鹿特丹管理学院MBA。具有超过30年的丰富的健康与大数据产业运营、创业及投资经验；1989年从中国医药保健品进出口公司进口部业务员开始进入医疗行业，历任飞利浦医疗系统中国合资项目助理经理、产品经理、亚太区市场经理、中国区代理网络经理，美国史赛克中国区总经理等职，是中国数字医疗设备领域的早期市场开拓者；2002年起开始涉及医疗信息化。

曾任GE医疗信息系统大中华区总经理；2011年起进入投资领域。历任德丰杰龙脉中国基金合伙人、复星集团健康控股执行董事等职，具备同时从产业运营及金融投资双视角进行价值投资及提供增值服务的能力；2015年创立上海荣增，围绕大数据增值服务逐步建立“产融协同”的综合大数据增值服务企业集群。

关于上海荣增——上海荣增信息科技有限公司是一家拥有关键核心技术的大数据增值应用技术专业企业，其技术平台与核心技术组件已广泛应用于零售、金融、汽车、医疗、能源、网络安全等不同领域。Google对荣增零售服务的第一代技术的调研显示，采用荣增技术的案例相对于对比组可增加15%~48%的营业额；荣增自身的调研则显示，荣增技术支撑下的手机广告转化率可达10%；2019年知名院士在做区块链技术应用汇报时，特别提及由荣增团队成员发明

的数据加密技术。

冼峰先生采访记录

1. 您觉得本次行业标准产生最大的意义是什么？

冼峰先生：“目前整个社会对大数据及其应用普遍处于一种概念性认同，但却缺乏实质性理解的状态中。例如，不少人在谈到外卖平台系统的算法驱动问题时，都会怪罪于“大数据”，而没有意识到这其实是系统对数据的误用，而非数据本身的问题。”

“解决这个问题需要唤醒大众的数据意识，其中数据立法是一个很重要的支撑，但更为潜移默化的，可能正是类似本次行业标准这样的自律与引导规范。当整个行业的数据分析应用都更为规范时，不友好的数据应用比例将自然降低。我认为，本次行业标准在这个方面的尝试，是将数据分析行业往数字友好应用方向进行引领。这正是其最大的意义所在。”

2. 作为本次行业标准化工作的参与者，我们通过数据委了解到您提供了很多见解，其中您对于基础共性参考标准提了很多建议，请问您的建议依据是什么呢？

冼峰先生：“上海荣增虽然创立时间只有六年，但核心创始团队的每个成员都各自在这个领域有15年以上的行业经验，因此对这个行业有自己的深度理解。因此我们所提建议的依据，正是基于整个团队对数据应用闭环要求的经验总结。比如，我们认为大数据应用可以，也一定要，形成经济增量。这种增量不应该是来自对传统经济的压制，而更应该是对他们的提升。”

“举一个最直接的例子：目前以互联网为代表的虚拟经济与实体经济之间的矛盾关系，其实完全可以转化为协同关系。这需要理顺“什么是目标，什么是工具”的问题。数据分析行业是为甲方服务的，因此将很自然形成向甲方提供业务辅助工具的形态。这对优化整个数据应用行业的发展基础是很有意义的。”

“我们正是基于这样的闭环理解去看基础共性参考标准所应该起到的作用，提出了我们的观点与建议。”

3. 您认为本次制作的数据分析行业服务参考标准还有哪些不足吗？或者说还需要哪些对哪些方面进行探讨和补充？

冼峰先生：“需要客观地看这个问题，因为这不会是一步到位的标准，而是一个持续进化的过程。从这个角度看，

我认为目前的标准兼顾行业现实及大众的普遍理解，是一个可执行的标准。与此同时，从对中国商联数据委在编制标准的整个过程的观察，我认为这是一个可持续进化的组织。对行业的未来发展，我觉得这是一个更为重要的基础！”

4. 您认为在大数据盛行时期，数据分析人才培养从哪些方面能真正帮助您在执行项目中解决问题？

冼峰先生：“这个问题，可以从整个行业角度回答，也可以仅仅从自身业务角度来回答。”

“从行业角度，目前数据分析人才总体相当匮乏，这对数据应用行业的发展是非常不利的，也是目前我国数字经济发展既快速但又不健康的原因之一。例如，目前数字经济占比最重的部分还是产业数字化，且比重仍在持续增加中，但数字化之后的数据价值仍然得不到现金流支撑。这其中的重要原因之一，就是数据分析环节还是太弱，数据之间的逻辑关联得不到呈现，数据价值自然得不到支撑。”

“在这个背景下，回到我们的自身业务，意味着我们常常需要从最基础的培训入手对自己及客户的数据团队进行培训，这大大降低了我们的业务转化速度。因此，高质量且充足的数据分析人力资源，对企业本身乃至整个数字经济的健康发展，是非常重要的。”

5. 您觉得专业的大数据公司应如何建设自己的品牌？如何逐渐扩大自身社会影响力？

冼峰先生：“有两个方面的工作需要做，而且是持续坚持做。”

“第一个方面，在于自身的业务覆盖面。更直观的体

现，就是在销售额上。因为业务的健康发展必将促进品牌建设。”

“第二个方面，在于坚持“数字友好”，即坚持“以人为本”，使正确的价值观对品牌建设产生持续的正面影响。”

6. 您认为您的团队核心优势是什么？尤其在疫情的影响下，如何应对冲击保持自己的竞争优势？

冼峰先生：“上海荣增的核心优势，首先来自对行业的前瞻性理解，然后通过组织相关专业资源将这些理解转化成相应的技术工具，并最终在一系列不同行业的实践案例中得到验证。”

“当然，这样的优势需要通过自身的持续提升来保障。这不仅是在现在的疫情环境下，即便是没有疫情，也仍然会有各种不同的市场挑战。因此需要有积极、动态、持续地自我提升要求。”

7. 荣增团队什么时间成为的中国商联数据委员会单位？成为会员后的感受和收获有哪些？

冼峰先生：“我在2016年认识了邹东生会长，也是当年大数据峰会的演讲嘉宾。自此之后，我们一直保持很好的互动，但成为数据委员会是在2020年我们回归中国市场后。成为会员的这两年，特别是对数据委在编制标准的整个过程的观察，我认为这是一个可持续进化的组织。对行业的未来发展是非常重要的！因此我对这个行业充满信心。”



解读《数据分析行业服务参考文件》系列采访： 帆软数据应用研究院研究主管王文信

来源 / 中国商业联合会数据分析专业委员会 编辑 / 数据委员会处 李苗苗 日期 / 2022-04

解读《数据分析行业服务参考文件》系列采访
本期嘉宾



帆软软件有限公司 数据应用研究院研究主管 王文信

帆软数据应用研究院研究主管，专注于大数据BI行业趋势和企业数据化管理案例研究，《数据生产力：企业BI项目建设与运营》、《商业智能工具应用与数据可视化》书籍作者，帆软《数字化前沿》期刊主编，《商业智能白皮书》主笔，并在科技日报、钛媒体、动脉网等专业媒体机构上有多篇文章发表。

关于帆软软件有限公司——帆软软件有限公司（以下简称“帆软”）成立于2006年，是中国专业的大数据BI和分析平台提供商，专注商业智能和数据分析领域，致力于为全球企业提供一站式商业智能解决方案。帆软在专业水准、组织规模、服务范围、企业客户数量上均为业内前列，先后获得包括Gartner、IDC、CCID在内的众多专业咨询机构的认可。曾入选福布斯中国非上市企业潜力榜50强，2018年-2021年，连续4年入选中国大数据企业50强。

王文信采访记录

1. 您觉得本次行业标准产生最大的意义是什么？

王文信：“我认为本次服务参考文件发布的最大意义是完善了我国数据标准，令数据分析行业在从业过程中有规可循。它从多个维度系统的对现实生产中的数据分析活动进行全面的定义和解读，细致的为我们展现了数据分析服务从初

立项目到售后服务的一整套服务体系标准，以及实际生产过程中数据分析行业所使用的行业术语规范和从业人员分级、分类等内容。可以说是填补了市场化维度层面的空白，让行业有据可依，有理可循，对我国大数据应用和整个数据分析行业的发展有重要的参考意义。”

2. 作为本次行业标准化工作的参与者，我们通过数据委了解到您提供了很多见解，其中您对于基础共性参考标准提了很多建议，请问您的建议依据是什么呢？

王文信：“提建议的依据主要是两个方面：一方面，市面上已有的一些基础定义和相关标准、规范；另一方面是基于帆软16年在大数据BI行业的一个积累，包括我们对于自身行业和市场的深耕，对于客户需求、对于上下游厂商的一些理解和沉淀。值得一提的是，帆软也有自己的数据研究院，在融合行业最新动态及自身经验的基础上提出自己的原创性意见，可以说与本次服务参考文件的制定十分契合了。”

3. 参加本次数据分析行业服务参考标准的制作，您最大的感触是什么？

王文信：“最大的感触就是一份行业服务参考标准制作过程的不易，以及群策群力在过程中的重要性。数据分析行业涉及面广，本身较为复杂，这样一份标准需要群策群力，考虑全面才能发挥更大的价值，我们从其他行业内企业、专家老师的一些观点和思考中也学习到很多。”

4. 您认为本次制作的数据分析行业服务参考标准还有哪些不足吗？或者说还需要对哪些方面进行探讨和补充？

王文信：“数据的应用会随着信息技术的发展而发展，而且速度会越来越快，那么我们的行业服务参考标准也需要具备很强的延展性和灵活性，在未来应对行业变化时能够保障其普适性和有效性。”

5. 您认为在大数据盛行时期，数据分析人才培养从哪些方面能真正帮助您在执行项目中解决问题？

王文信：“除了数据分析人才最基础的技术能力能够在项目开发上显露价值外，数据分析人才培养过程中形成的业务能力能够在业务沟通和理解等方面对项目有很大的帮助，包括需求调研、项目实施等；另外数据分析人才的运营能力，也能够将项目成果有效地推广到更多的业务部门和企业员工，更大程度地发挥项目价值。”

6. 您觉得专业的大数据公司应如何建设自己的品牌？如何逐渐扩大自身社会影响力？

王文信：“对TO B公司而言，品牌的最终竞争力一定是落到产品和服务上的。产品为攻：优秀的产品一定是能最大化满足客户需求，为客户提供最优质体验的，这一点毋庸置疑。服务为守：如果说产品完成了从0到1的过程，那么从1-10甚至从1-100就是持续性运营的结果，大数据公司需要建立自己良好的生态和完善的服务体系，才能让产品发挥最大化的价值，服务好客户。”

“这两点一直也是帆软在追求和实际践行的准则，优秀的产品+完善的服务才能发挥1+1>2的效用，打造产品美誉度和企业口碑。”

“至于如何扩大自身影响力，我认为主要发力点就在于恪守本心，并尽到应尽的社会责任。基础是守法经营，不搞恶性竞争，推动行业的健康发展；核心是服务好客户，为客户为社会带去价值，比如目前帆软正在进行的数据分析大赛，更进一步则是积极参与行业建设与社会建设，例如参加本次标准制作，还有一些慈善、环保活动等。”

7. 您认为您的团队核心优势是什么？尤其在疫情的影响下，如何应对冲击保持自己的竞争优势？

王文信：“我认为帆软的核心优势在于优秀的产品和服务体系以及基于此建立的良好品牌，这也是拉动我们一直前行的‘三驾马车’吧，此外，帆软自身也具备了较强的组织

活力和抗风险能力，能够灵活应对突发情况，这一点在疫情的影响下是非常重要的，组织的活力、灵活性等特质会很好地削弱这些影响，化风险为机遇。”

8. 帆软什么时间成为的中国商联数据委员会单位？成为会员后的感受和收获有哪些？

王文信：“帆软成为数据委员会已有三年时间，成为会员后，我们很荣幸参与行业标准的制定以及行业报告发布等工作。最大的感受是见证了中国大数据行业的高速发展以及国产化软件的崛起，目前帆软所在的商业智能（BI）领域年增长约21%，随着云计算、移动、IoT、5G等技术的发展，未来3年全球所创造的数据量将超过过去30年中创造的数据总和，商业智能软件还将迎来更深入和更广泛的应用场景。此外，国产化软件技术不断提升，在更懂中国企业需求前提下，市场份额不断扩大。”

至于帆软，成为数据委员会，是一个提升和证明自我的机会，一方面我们可以更快了解行业动态和变化，积极做出相应调整；另一方面，数据委也为帆软提供了更大的平台去发挥证明自己，发挥BI行业领跑者的价值。”

解读《数据分析行业服务参考文件》系列采访： 北京派可数据联合创始人吕品

来源 / 中国商业联合会数据分析专业委员会 编辑 / 数据委员会处 李苗苗 日期 / 2022-04

解读《数据分析行业服务参考文件》系列采访
本期嘉宾



北京派可数据科技有限公司 联合创始人、首席数据运营专家 吕品

派可数据首席数据运营官；天善智能联合创始人；36Kr 企服联盟 | TO B 专家团成员；IT从业十五年，历经传统软件开发、架构师、商业智能BI开发、数据架构师、行业专家、微软2015-2021 MVP（全球最有价值专家）；曾服务过SHP、微软、大众金融等企业客户，熟悉企业经营管理、财务、运营等思维方式；微信视频号《吕品聊数据》作者。

1. 您觉得本次行业标准产生最大的意义是什么？

吕品：“行业标准产生最大的意义是在说明一件事情：就是这个行业逐步成熟、规范了，这个行业要接下来的发展需要有一些标准和依据来参考。”“现在有这么多的企业在大数据、数据分析领域有各种各样的诉求，但是怎么做、按照什么标准和要求去做，还是缺少一些标准和依据的。”“数据分析人才培养水平的参差不齐，在人才资源的分配上应该如何规划、提升也是需要企业来思考的。”“站

在行业的角度，有没有什么可以逐步沉淀、总结的，一些好的行业经验可以帮助更多的企业来提升数据分析的能力和价值的，这些也是行业标准制定的意义所在。”“所以，我们先解决有的问题。”“我们都知道，在任何行业，标准的制定往往是最难的，但标准的制定总是要有起步的，可以先起步，再逐步迭代和完善。”

2. 作为本次行业标准化工作的参与者，我们通过数据委了解到您提供了很多见解，其中您对于数据分析专业术语及定义、基础共性参考标准提了很多建议，请问您提建议的依据是什么呢？

吕品：“第一，我们对整个行业非常的熟悉。派可数据这家公司虽然成立的时间不长，大概也就5年左右时间，但是整个团队在这个行业服务的时间非常的长，基本上都是十几年的经验，见证了整个商业智能BI、数据分析在国内的成长过程，所以我们对这个行业非常的熟悉和了解。现在很多企业在数据类项目建设过程中所经历的所有问题，基本上我们在过去十几年的时间里都碰到过，各种各样简单或者复杂的场景我们都遇到过，知道怎么解决。所以，我们有这样的能力和经验来为这个行业做出一些基本的贡献。”

“第二，我们做了很多专业化的沉淀。我们自己也是做数据分析、商业智能BI产品的。做产品的公司大家都知道，产品本身就是一种经验的具体化和抽象化，是总结了无数经验的成果，最终将经验标准化再输出到用户，所以产品本身就是一种标准。这种标准性在市场上受到的挑战是非常大的，因为最终你要商业化。所以，可以商业化的产品或者解决方案是很厉害的，一定有它专业且被企业接受和认同的地方。很多产品的理念，或者产品本身的方法论是可以被企业所借鉴的。”

“第三，我们一直在市场一线，我们知道客户需要什么。我们整个团队十几年时间服务过大大小小几千家企业，我自己在最近的三、五年时间面对面拜访过、沟通过的企业不低于500家，基本上一直在市场的一线，对于不同行业的企业、信息化发展不同阶段的企业数字化现状还是非常了解的，想要什么、要解决什么、会遇到什么样的问题和挑战我们也是非常清楚的。我们把自己这么多年踩过的坑，走过的弯路，或者看到其他企业走过的弯路总结分享出来，帮助到更多的企业。”

“所以，这些依据更多还是来自于我们对这个行业的认知，对一线企业的了解认知，对我们自己经验的认知，最终形成一些经验沉淀并贡献给大家。”

3. 参加本次数据分析行业服务参考标准的制作，您最大的感触是什么？

吕品：“最大的感触就是尽管我们感觉自己对这个行业很熟悉了，我们的经验很丰富了，但是真正站在不同的角度思考问题后，发现还是有很多的缺失，所以这个服务参考标准不是哪一家企业、哪一个人是可以决定的，是需要大家集思广益、群力群策，一起收集、分析思考、总结出来的。还

是那句话，标准化的制定往往是最难的，但好在我们总算有一个起步了，虽然可能存在一些不完美、不完善的地方，但是至少有一个可以参考的标准了。”

“在没有参考标准之前，很多企业在这个领域可能只能做到50分，但是有了参考之后，可以做到60分、70分，这个就是一种行业的进步。”

“那我们接下来要逐步再来提升、优化这种标准，逐步的、让那些达到70分的企业慢慢可以达到80分、90分，这就是我们未来努力的方向。”

4. 您认为本次制作的数据分析行业服务参考标准还有哪些不足吗？或者说还需要对哪些方面进行探讨和补充？

吕品：“一个是数据安全方面，特别是关键数据、重要数据、跨境数据的安全管理；另一方面是对于行业中违规或违反职业道德行为的惩罚机制。”

“不同的行业、信息化发展不同阶段的企业对这种标准的感知和理解是不一样的，要求也不一样，还是存在很多优化的空间，完善也需要一段时间过程，所以这个阶段推出来后我们对它的定位还是叫做参考。”

“目前虽然定了很多的标准，但是这些标准怎么去落地，这对很多企业而言还是有些模糊的、不清晰的，恰恰这部分又是最难的，很难理论化和抽象化。所以就针对不同点再进行阐述、经验总结，不断的对外输出，所以还需要有一个经验输出的阵地，不管是通过视频、文章、社群等等，多讲讲我们怎么去落地的事情。”

5. 您认为在大数据盛行时期，数据分析人才培养从哪些方面能真正帮助您在执行项目中解决问题？

吕品：“对于数据分析人才的培养，从我们的角度来看有三个维度：数据意识、业务思维和基本技术。”

“没有数据意识是做不好数据分析的。数据意识这个点也很宽，比如大数据、小数据，先从养数据开始；如何从数据反推信息化的建设；数据如何驱动业务管理等等。像这些话题都是体现数据意识的，不仅仅是数据敏感度的问题，而是从数据产生开始，到后面的数据处理过程，到数据建模、到数据分析，再到数据应用，中间有太多的点需要涉及。简单来说，就是关于数据方面的，你能想到的是否足够多，总比别人想到的更进一步，有没有一些正确的关于数据的理念、方法论、思维方式。”

“业务思维。简单来说就是数据分析人员需要站在业务的角度去思考问题，需要了解一家企业或者这个行业的基本业务特征、行业驱动因素，企业经营方式、业务模式、各个节点的业务流程。只有了解这些基本盘，才知道分析的重点在哪里，如何从数据角度来引导业务管理层关注到这些问题，并提出一些思考的方向。数据分析是不能代替企业做任何决策的，我们只是暴露问题点，建立初步分析的结论、提供思考的方向，最终结论和下决策是管理层做的事情，数据分析做好参谋这个角色就可以了。不同层次的数据分析做不

同层次的事情，比如基础业务层、部门层、经营管理层。”

“基本技术。比如掌握一些基本的SQL、数据处理工具，有一定的数据处理能力就可以了。数据分析下层的数据处理、ETL 这些又是一个专业、纯技术驱动领域数据分析人员不需要掌握到那么深，但是要有这些基本的概念和初步应用能力，至少和其他数据团队合作的时候，能够用大家都听得懂的语言来沟通。”

“总之，数据分析领域还是一个非常专业的领域。从我的角度看，数据意识、业务思维是远远要大于技术本身的。”

6. 您认为专业的大数据公司应该如何建设自己的品牌？如何逐渐扩大自身社会影响力？

吕品：“这还是一个比较大的话题。因为像我们也是在一个创业阶段，所以我们也并没有这个能力指导整个行业，就只谈谈我们自身的理解。”

“从产品角度，还是要有创新，创新才能有进步，在产品规划层面我们是把创新摆在第一位的，这个在公司内部大家是已经取得了共识。这个创新一定是去做别人没有做过的事情，形成产品差异化路线，并且还要求要务实。所谓的务实就是不做那些虚的，赚客户眼球、吸引资本市场亮点的东西。务实是要真正了解整个行业、企业目前的真正痛点、主要矛盾，然后去解决它。”

“我们虽然在BI行业起步较晚，但是在产品差异化上已经有一些体现，也逐步获得了很多企业的认可，因为你能解决别人解决不了的问题，有创新性。这个也是我们未来一直努力的方向。”

“社会影响力我们其实还并没有真正去关注，或者想那么多，因为眼下我们还要有很多事情要去做。这个行业也不是靠我们一家公司或者几家公司去做好的，需要很多人的努力，需要很多同行企业的努力，让这个市场、这个行业更加的成熟，需要大家共同教育好这个市场。只有行业的影响力大了，在这个行业的企业才可以更好发展，这个时候就再想

想我们会有什么样的社会影响力，做了多大的贡献。”

7. 您认为你所在公司的核心优势是什么？尤其在疫情的影响下，如何应对冲击保持自己的竞争优势？

吕品：“核心优势一个是差异化的产品策略，比如我们数据仓库领域，我们是基于数据仓库模型的BI分析产品工具，这个跟国内外已有的BI产品还是有很大差异的。它能够很快速、完整、零代码的方式完成BI底层的数据仓库模型搭建，并且很注重长远的底层数据架构规划，比如三年、五年、甚至更长的时间，确保底层的数据架构是稳健的，这个能力是我们现在非常差异化并且领先的优势。”

“第二个优势就是我们的业务沉淀能力、领域沉淀能力。因为做BI不懂业务是很难做好的，所以在不同行业业务积累上，我们还是做了充分的准备。在数据仓库中，我们可以沉淀我们很多的业务分析模型，这些模型在我们和不同企业合作过程中是可以直接进行经验输出的，这个也是我们很大的优势。”

“疫情对TO B服务市场确实有比较大的影响，最直接的就是销售、顾问没有办法到处见客户出差了，很多需要面对面的沟通不太方便进行。几年疫情下来，其实大家也慢慢适应在线上沟通和交流了，居家其实也可以办公，这反而催生了各种工作协同的场景，也算是一种进步。所以，目前我们跟客户的沟通在线上比较多，感觉受到的影响也不像以前那么大了。”

8. 什么时间成为的中国商联数据委员会？成为会员后的感受和收获有哪些？

吕品：“我们去年才成为数据委员会。收获和感受就是要思考的问题更多了，更有责任感了，之前是代表一家企业或者个人，现在是要站在行业的角度去思考问题，需要更加谨慎的态度，需要更加专业了。这是一件好事情，既然对我们提出了更高的要求，那么我们就一起往这个方向去努力了。”



解读《数据分析行业服务参考文件》系列采访： 行标专家指导组成员陈虹坚

来源 / 中国商业联合会数据分析专业委员会 编辑 / 数据委员会处 李苗苗 日期 / 2022-04

解读《数据分析行业服务参考文件》系列采访
本期嘉宾



CPDA数据分析师、行标专家指导组成员 陈虹坚

工学学士、管理学硕士、高级工程师、CPDA数据分析师。长期从事技术系统运维、网络管理、数据管理等工作，致力于推动技术与业务双向融合发展，目前从事技术管理工作，担任技术部门负责人。

1. 您觉得本次行业标准产生最大的意义是什么？

陈虹坚：“最大的意义在于，它是一个从无到有的过程，规范了行业标准，让数据分析行业有据可依。”

2. 作为本次行业标准化工作的参与者，我们通过数据委了解到您提供了很多见解，其中您对于基础共性参考标准提了很多建议，请问您的建议依据是什么呢？

陈虹坚：“首先是市场需求。每个行业都有大量的数据产生，数据隐藏的价值非常巨大。它们是重要的资源，这些数据如何管理、如何使用、如何识别它的价值、如何保障它的安全，都是每个行业应该关注的问题；其次是数据分析工作目前存在的痛点。比如社会认识还不到位，数据思维尚未形成，企业掌握着数据，却看不到数据的价值，用数据决策用数据管理还只停留在口头阶段；然后是行业规范的角度。数据分析结论具有指导意义，需要数据分析师在工作中保持独立、客观和公正，实事求是，挖掘隐藏在数据背后的真相，不因任何利害关系影响客观公正。”

3. 参加本次数据分析行业服务参考标准的制作，您最大的感触是什么？

陈虹坚：“在当前，数据已经是重要的生产要素，世界各国纷纷出台大数据战略，聚力数据要素多重价值挖掘，本次数据分析行业服务参考标准非常及时，填补了我国数据分析行业暂无从业标准可参考的空白，具有里程碑式的积极意义。”

4. 您认为本次制作的数据分析行业服务参考标准还有哪些不足吗？或者说还需要对哪些方面进行探讨和补充？

陈虹坚：“一个是数据安全方面，特别是关键数据、重要数据、跨境数据的安全管理；另一方面是对于行业中违规或违反职业道德行为的惩罚机制。”

5. 您认为在大数据盛行时期，数据分析人才培养从哪些方面能真正帮助您在执行项目中解决问题？

陈虹坚：“主要是两点，一是能够提出合理的、有价值的、有意义的业务分析需求，避免无效/无理需求进而浪费资源。二是对数据的敏感度，或商业敏感度。”

6. 考取CPDA数据分析师证书，加入行业后，您对这个行业有什么感受？以及有哪些收获吗？

陈虹坚：“我觉得考取CPDA数据分析师证书的学习过程，对数据分析师的思维模式和分析师品质的形成很有帮助。首先，虽然我们学习的是数据分析技术，但是数据分析师的分析思路以及切入点，在数据分析过程中的价值远胜过技术本身；其次，作为数据分析师，我们应该对数据心存敬畏，相信在数据的背后，肯定隐藏着有价值的业务规律、市场趋势、难以察觉的现状以及风险预判等等，这需要我们客观准确地找出隐藏在数据背后的真相，不被表象误导。”



湖北省大数据产业“十四五”发展规划

来源 / 湖北省经济和信息化厅 编辑 / 数据委员会处 李苗苗 日期 / 2022-04



信息技术与经济社会的交汇融合引发了数据迅猛增长，数据正在成为与物质、能源同等重要的基础性战略资源。立足湖北通信枢纽、信息产业基础和人才资源优势，推动大数据产业发展，对促进经济转型和创新发展、提升政府治理能力、优化民生公共服务等具有重大意义。为贯彻落实国家大数据发展战略，促进大数据产业快速健康发展，依据《“十四五”大数据产业发展规划》和《湖北省国民经济和社会发展的第十四个五年规划和二〇三五年远景目标纲要》，特编制本规划。

一、发展基础

（一）发展现状

一是基础设施建设有序推进。全省在用数据中心超过80个，机架数量超过10.8万架，武钢大数据中心、国家电网基地中金大数据中心、襄阳云计算基地和宜昌三峡大数据中心等项目加快建设，初步形成了“一主两翼”的数据中心集群。截至2020年，全省5G宏基站累计超3.6万个，百兆及

以上宽带接入用户累计达到1624万户，互联网出省带宽达36.6Tbps；武汉顶级节点已接入二级节点17个，标识注册量超过48亿。

二是大数据应用领域不断拓宽。省内龙头企业积极开展大数据应用探索，在装备、汽车、电子、化工等行业建成15个重点工业互联网平台，长飞、武重等龙头企业获得全国工业互联网试点示范。湖北省政务大数据能力平台入库数据达172亿条，对接完成179个重点信息系统，省政务办“破信息共享藩篱 筑政务一窗底座”实践案例荣获政务服务创新奖。在疫情防控阻击战中，我省建成疫情防控大数据系统，有力支撑了疫情防控部署以及对流动人员的疫情监测。

三是产业发展初具实力。2020年我省大数据产业规模达510亿元，在数据采集、传输、处理和应用等环节集聚了一批优质企业，形成了从硬件到软件、从产品到服务的全产业链覆盖。中国信科集团与华为公司合作推出“长江计算”品牌鲲鹏架构服务器产品，年产能达15万台，达梦数据稳居国

产数据库市场占有率第一，武船、东风、武汉虹信、光谷信息等30个企业项目入选工信部大数据产业发展试点示范。

四是产业生态持续优化。省政府出台发布了《湖北省制造业产业链链长制实施方案（2021—2023年）》，启动实施大数据产业链链长制。阿里巴巴、腾讯、华为、浪潮、中国电子等企业区域性总部加速向我省聚集，武大、华科、武汉理工等12所高校开设大数据专业，中国信息通信研究院、中国工业信息安全中心等大数据领域的国家级智库机构在我省设立分支机构。大数据创投氛围更加活跃，创立了湖北高质量发展产业投资基金，重点支持大数据、智能驾驶等领域。

我省大数据产业十三五时期虽取得明显成效，但仍然面临着一些问题和挑战：一是运用大数据思维和数据治理能力有待提升，对发展大数据的重要性认知不够高，在大数据开发管理方面还缺乏强有力的统筹调配；二是产业规模偏小，我省大数据产业规模约为上海的1/4，缺少腾讯、阿里巴巴等能带动大数据产业发展的龙头企业；三是大数据人才匮乏，尤其是复合型人才严重匮乏是我省大数据深度融合发展的瓶颈，制约了大数据和实体经济深度融合。

（二）发展形势

从国际看，数据已经成为影响全球竞争的关键战略性资源，只有掌握和获取更多的数据资源，才能在新一轮全球话语权竞争中占据主导地位。世界主要国家把推进数字经济作为实现创新发展的重要动能，加快前瞻性战略布局。美国通过稳步实施“三步走”战略，在大数据技术研发、商业应用以及保障国家安全等方面已全面构筑起全球领先优势。围绕数据的生产要素化，欧盟加快数字规则制定，发布《通用数据保护条例》和《非个人数据在欧盟境内自由流动框架条例》，成为全球个人数据立法的典范。日本发布《面向2020年的ICT综合战略》，推动大数据技术从基础研究到应用转化。

从国内看，习近平总书记在中共中央政治局第三十四次集体学习时强调要充分发挥海量数据和丰富应用场景优势，促进数字技术与实体经济深度融合，党的十九届四中全会首次提出将数据作为生产要素参与分配。在国家战略的引领下，十三五时期我国大数据市场保持每年30%以上的快速增长，产业规模超过1万亿元，预计十四五时期仍将保持较快增长态势。目前，我国大数据产业主要位于经济发达的地区，北京、上海、广东等省市依托当地互联网及技术企业和高端科技人才等优势，充分发挥区域协同作用，带动形成了京津冀、长三角、粤港澳大湾区三大产业聚集区。贵州、重庆等通过积极吸引国内外龙头骨干企业，实现大数据产业快速发展。

当前，我省正处于重要战略机遇期和蓄积势能迸发期，转型升级势在必行，迫切需要发挥数据要素的倍增效应和大数据产业的动力变革作用，全力构建以先进制造业主导、战略性新兴产业为引领的现代产业体系。新形势下，我省要把

数字化转型放在更加突出的位置，充分发挥大数据技术和科教等方面的优势，围绕强链、补链、延链，做强做大做优大数据产业，推进大数据产业基础高端化和产业链现代化，加快大数据技术在各行业、各领域的应用，以数字技术赋能实体经济，为加快“建成支点、走在前列、谱写新篇”打造新的增长点。

二、总体要求

（一）指导思想

以习近平新时代中国特色社会主义思想为指导，认真贯彻落实国家大数据发展战略，全面贯彻党的十九大和十九届二中、三中、四中、五中、六中全会精神，按照省委十一届七次、八次、九次、十次全会部署，围绕释放数据要素价值、发挥数据倍增效应，坚持基础设施先行、坚持资源共享开放、坚持行业应用牵引、坚持产业高端发展，充分发挥大数据在经济社会中的赋能作用，助力构建全省战略性新兴产业引领、先进制造业主导、现代服务业驱动的现代产业体系，为“建成支点、走在前列、谱写新篇”提供强力支撑。



图1湖北省“十四五”大数据产业总体架构

（二）基本原则

市场主导。推动数据资源价值化，以数据交易、流通等活动来发挥市场在资源配置中的决定性作用，依靠市场力量优化数据资源布局，激发主体活力、创新发展模式。

创新引领。从急迫需要和长远需求出发，加强原创性、引领性科技攻关，推动关键核心技术自主可控。坚持创新链产业链融合，推进重点项目协同和研发活动一体化。

融合赋能。坚持以行业应用为驱动，以赋能产业转型升级为着眼点，促进大数据与传统产业在发展理念、产业体系、生产模式、业务模式等方面全面融合。

安全可控。处理好安全与发展的关系，把安全贯穿大数据发展各领域全过程，制定大数据相关制度规范，完善大数据安全支撑体系，在发展中提升数据安全保障能力。

（三）发展目标

到2025年，建成全国领先的算力基础设施和重要的数据

汇聚枢纽，打造全国大数据创新核心区、具备影响力的产业聚集区。

数据中心建设加速推进。打造全国一体化算力网络国家枢纽节点，全省数据中心机架数达到17万架以上，新建大型和超大型数据中心运行电能利用效率降到1.3以下。

数据应用持续深化。大数据在产业转型、民生服务等领域的应用深入展开，建设50个工业互联网平台、培育2—3个双跨平台，组织实施一批重点行业应用示范项目。

产业能级不断提升。高端现代化大数据产业体系初步建成，形成一批技术领先、自主可控的大数据产品和服务，大数据产业发展主体力量不断壮大，全省大数据产业规模达到1000亿元。

一级指标	二级指标	三级指标	单位	现状	2025年目标
数据基础设施	数据中心规模	标准机架数 (2.5KW/架)	万架	10.8	17
	数据中心能耗	数据中心电能利用效率 (PUE)		—	1.3
大数据应用	工业大数据应用	工业互联网标识解析二级节点	个	12	15
		工业互联网标识注册量	亿个	60	200
		工业互联网平台	个	15	50
		跨行业跨领域工业互联网平台	个	—	2—3
		DCMM3级以上企业数量	个	—	20
大数据产业	产业规模	大数据产业规模	亿元	510	1000

三、发展重点与空间布局

(一) 重点领域

做大做强大数据核心产业。夯实大数据基础技术，加强大规模数据采集、大容量存储、海量数据处理、异构数据兼容、隐私计算等关键核心技术研究。构建软硬一体的大数据产品体系，加快超融合一体机、智能终端、存储设备、服务器等大数据硬件产品产业化进程，全面升级大数据平台、数据分析工具等大数据基础软件水平，提高可视化、图像分析等大数据应用软件能力。大力推动大数据服务模式和业态创新，鼓励大数据龙头企业加快大数据服务向专业化、平台化方向发展。强化大数据安全产业发展，加强隐私计算、数据脱敏、密码等数据安全技术的研发。

推动关联产业融合发展。大力发展区块链、人工智能、工业互联网等大数据关联产业，推动大数据与新一代信息技术的集成创新。在大数据产业聚集区探索布局区块链企业，强化分布式网络、密码学、共识机制、智能合约等区块链技术与大数据技术的融合创新。充分挖掘企业在工业互联网应用中的各环节数据，全面释放数据价值。大力支持人工智能企业和大数据企业联合开展关键技术攻关，全面提升以数据为基础的感知识别、知识推理、智能计算、人机交互能力。

(二) 空间布局

支持全省各地结合资源禀赋，发挥比较优势，构建我省大数据产业“一主两翼、全域协同”发展格局。武汉充分发挥“一主引领”的带动作用，坚持产业发展和应用创新双轮驱动，打造全省大数据发展增长极；省内其他市州以大数据应用为主构建大数据服务体系，积极打造具有本地特色的大数据应用示范。

支持武汉依托骨干企业，重点发展存储设备、网络通信设备、服务器、数据库、云平台、大数据应用软件、数据应用和服务等核心产业，并在政务服务、多式联运、智能制造等领域率先开展大数据应用示范。支持襄阳重点发展网络通信设备、服务器、云平台，同时加快在电子政务、智慧城市的应用，带动“襄十随神”城市群发展大数据产业。支持宜昌重点发展数据存储与云计算、数据归集、政务数据资源共享开放，加快在智慧城市、网上金融服务等领域的应用，带动“宜荆荆恩”城市群发展大数据产业。



图2 湖北省“十四五”大数据产业布局图

四、重点任务

(一) 完善数据基础设施

布局建设新型数据中心。认真贯彻落实国家碳达峰、碳中和的战略决策，加快先进绿色技术产品应用，推动绿色数据中心创建、运维和改造，加强内部能耗数据监测和管理。分类分批推动存量“老旧小散”数据中心改造升级，提高能源利用效率和算力供给能力。积极构建城市内的边缘计算中心供给体系，支撑边缘数据的计算、存储和转发，满足极低时延的新型业务应用需求。以武汉、襄阳和宜昌为核心区优

化全省新型数据中心布局，推动数据中心互联互通。

加快提升计算能力。鼓励有条件的地区积极布局人工智能、区块链算力中心，满足支撑科研探索、智慧城市和数字经济场景应用的算力需求，重点支持武汉建设运营好国内首个人工智能算力中心，加快推进武汉超算中心建设。打造具有区域影响力的云计算服务平台，以云服务等方式提供公共算力资源，降低算力使用成本，提升应用赋能作用。

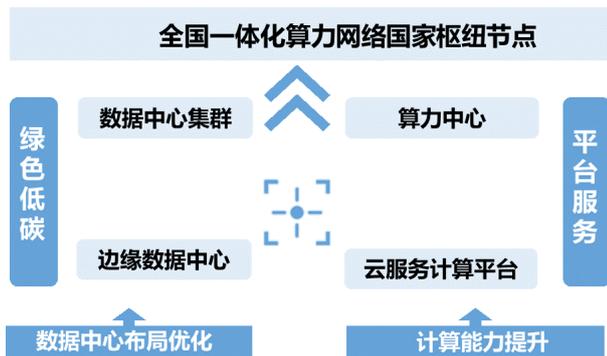


图3 湖北省“十四五”数据基础设施架构

专栏1 数据基础设施建设工程

建设数据中心集群。推进武汉数据中心重大项目建设，打造服务武汉城市圈数据中心集群。推进襄阳云谷园区建设，打造服务“襄十随神”城市群的区域性数据中心集群。推动宜昌重大数据中心项目建设，打造服务于“宜荆荆恩”城市群和国家中部地区的数据中心集群。到2025年底，全省数据中心机架数达到17万架以上。

建设云服务化算力集群。加快武汉超算中心建设，一期算力规模100P Flops HPC。建设运营好武汉人工智能算力中心，提供一站式的人工智能开发环境，为相关应用场景提供算力基础服务设施。依托武汉超算中心和人工智能算力中心，打造全国第一个集人工智能和超算融合的云服务化算力集群，创建国家超级计算武汉中心。

建设全国一体化算力网络国家枢纽节点。依托重点项目，支持宜昌积极创建全国一体化算力网络国家枢纽节点，引导数据中心集约化、规模化、绿色化发展，推动与其他国家枢纽节点之间进一步打通网络传输通道，提升跨区域算力调度水平。

（二）促进数据集聚开放

推动数据汇聚。加快全省人口、法人单位、电子证照、自然资源和空间地理、社会信用、共享材料等六大基础库和长江大保护、精准脱贫、河湖治理、污染防治、全民健康保障、全民住房保障等各类主题数据库建设，丰富政务数据服务方式和内容。完善全省“1+N+17”政务云平台布局，实现全省政务资源的集中调度和综合服务。加快国家工业互联网大数据中心湖北分中心建设，面向重点区域和重点行业，

推动骨干企业链接工业互联网大数据中心，加速全省工业数据资源高质量汇聚。支持建设数据清洗和分析应用中心，面向数据供应方和需求方，提供数据清洗、数据标注、数据分类、数据评估、数据分析等数据服务。

推进数据共享开放。按照“一数一源、多源核验”的原则，编制并维护全省统一的政务数据资源共享目录，定期发布政务数据共享、开放责任清单。完善“湖北省政务信息资源共享平台”功能，加快构建省市分级维护、协同共享的政务数据开放共享体系。面向重点行业，支持行业协会、行业龙头企业、大数据企业共建行业大数据共享平台，引导企事业单位和行业龙头企业主动共享开放有加工价值、不涉及商业秘密的生产经营管理数据。

探索开展数据交易试点。推进数据产权立法工作，加快数据所有权、占有权、支配权、使用权、收益权、处置权等权利的界定，明确数据企业、用户以及监管部门等各类数据主体的权属分配。加快利用区块链、人工智能、隐私计算等新兴技术探索数据确权新模式新路径。探索搭建基于数据行业、类别、属性等多方位因素的数据价值评估体系，面向数据共享开放程度较高的行业和地区开展数据价值评估试点。加快培育数据要素市场，支持组建湖北大数据交易集团和湖北省大数据交易联盟，构建多层次的数据交易体系。



图4 湖北省“十四五”数据开放流通架构图

专栏2 数据要素市场培育工程

建设国家工业互联网大数据中心湖北分中心。加快湖北分中心建设，推动湖北分中心与国家工业互联网大数据中心的对接，围绕汽车、现代化工及能源等支柱产业，探索建设行业分中心，加强行业数据资源管理。打造工业互联网大数据中心综合服务能力，提供工业经济和产业运行监测指挥、应急事件预警协调等服务。

开展数据交易。支持武汉等有条件的地区开展数据交易试点，推行数据信息登记制度，确定大数据资产评估定价，开展所有权交易、使用权交易、收益权交易和跨境交易等多种数据交易模式的探索，发展基于大数据资产的质押融资、保险、证券化等金融创新服务。支持通过区块链搭建数据交易平台，强化数据交易安全保障。

（三）深化行业数据应用

发展工业大数据应用。支持传统工业企业开展数字化改造，提升数字化水平。实施“云行荆楚”和“万企上云”工程，推动企业加快基础设施、业务系统、设备产品上云上平

台，支持企业积极利用云平台深挖研发设计、生产制造、管理服务、供应链优化等各个环节的数据价值。加快工业互联网平台建设，增强平台的资源集聚和数据处理能力，提高平台面向不同行业和场景的应用服务能力。以新一代信息技术（光芯屏端网）、汽车制造、现代化工及能源、大健康、现代农产品加工5个万亿级支柱产业为突破口，支持数字化基础好的企业和园区开展工业大数据应用试点，利用数据提升智能化水平。

专栏3 工业大数据应用示范工程

实施工业企业数字化网络化改造计划，引导工业企业企业开放数据接口，提升工业设备的网络互联能力。积极运用5G、新型智能网关等新技术改造建设企业内网，实现生产线装备和信息系统的互联互通，推动研发、生产、经营等全流程的数据采集。

推进标识解析国家顶级节点（武汉）建设，顶级节点围绕“域名、标识、区块链”三位一体发展，支持部署星火·链网，发展可信的工业大数据。支持工业龙头企业、工业互联网平台运营单位等牵头建设标识解析二级节点，形成重点行业重点地市深度覆盖的工业互联网标识解析服务体系。到2025年，全省二级节点数量超过15个，接入武汉顶级节点的标识注册量超过200亿。

推动工业互联网平台创新应用，支持行业龙头企业建设行业级工业互联网平台。到2025年，围绕我省特色行业领域，建设50个工业互联网平台，培育和引入2—3个跨行业跨领域工业互联网平台。

面向重点地市和重点行业探索开展DCMM评估试点示范，完善数据管理能力成熟度贯标工作方法，推动省内试点企业同步开展数据管理能力评估和建设，提升数据管理能力，到2025年，全省达到3级（稳健级）以上标准的企业突破20家。

发展农业大数据应用。整合农业、水利等方面数据资源，利用大数据技术提升农业生产、经营、管理和服务水平。加快湖北省农业农村大数据中心建设，支持武汉、黄冈、襄阳、宜昌等地建设农业农村大数据平台。开展数字农业应用示范，推进大数据在农业精准生产、环境监测、农产品产销等领域的智慧应用。支持建设新一代信息技术与传统产业融合发展示范区，支持合作共建智慧农业综合示范区。

发展服务业大数据应用。大力推动服务业利用大数据进行商业模式、服务内容和形式创新，提高行业服务水平。推动零售、外贸等企业与大数技术融合应用，加快向用户平台类企业转变，利用大数据支持品牌建立、产品定位、精准营销、认证认可、质量诚信提升和定制服务等。发展金融大数据，提升大数据在精准营销、智能风控、供应链金融、股市行情预测等场景的应用水平。发展物流大数据，加快大数据在车货匹配、运输路线优化、库存预测、设备修理预测、供应链协同等物流行业关键环节的应用。

发展政务和公共服务大数据应用。加快完善以一个门

户、一个系统、三个支撑为主体的全省一体化政务服务平台，推进政务服务“一网通办”、城市运行“一网统管”，打造现代“城市大脑”，增强信息惠民实效。着眼疫情防控、抗洪抢险、地质灾害等应急管理需要，打通安全生产、气象、水利、地质等领域数据，建立全省应急指挥信息和应急监测预警网络体系，推动风险防控关口主动前移。构建覆盖省市县乡村五级的远程医疗体系，提升大数据在健康监控、医护资源配置以及医学影像诊断等场景应用水平。

发展教育大数据，推动基于大数据的精确学情判断、个性化学习分析、智能决策支持和信息化校园管理，助力实现教育现代化。



图5 湖北省“十四五”数据行业应用领域图

（四）提升自主创新能力

推进关键技术研发及产业化。支持省内企业参与大数据关键技术研发和标准制定，重点突破分布式高速高可靠数据采集、机器学习、海量数据处理、数据真伪识别等数据采集、挖掘、展现关键技术，发展面向大数据应用的操作系统、数据库、中间件等基础支撑软件以及与工业、政府、公共服务等密切相关的应用软件及解决方案。

加强数据安全技术攻关，聚力突破加密算法、智能合约、侧链与跨链等区块链核心底层技术。加快数据共享集群、图数据库、云数据库等前沿技术攻关，推动智能化数据管理工具、非结构化管理平台的产业化进程。支持发展大数据开源项目，加强边缘计算、区块链、人工智能、量子计算等技术和大数据的融合交叉。

全面升级大数据“硬”设备和“软”服务。依托“端”产业优势，发展高性能采集设备、大容量存储介质、高性能计算机、大数据一体机等大数据硬件设备。发挥武汉“中国软件特色名城”优势，大力发展数据分析、数据挖掘、可视化工具、个性化推荐等数据服务和产品。推进国家级信创适配基地、国家级信创人才培养基地、湖北信创产业园、国家级基础软件研发中心“两基地一园一中心”建设，提升大数据软件支撑能力。

支持龙头企业从“产品为主”向“服务为主”转变，提供技术协同以及便捷服务的大数据解决方案。

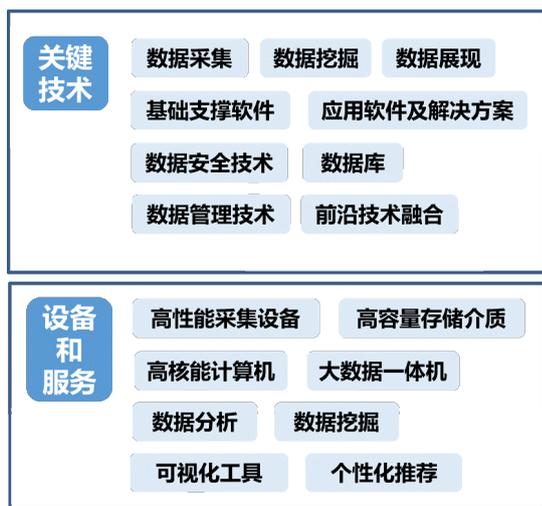


图6 湖北省“十四五”数据产业发展架构图

专栏4大数据产业高端化工程

高质量建设长江鲲鹏生态创新中心，加快鲲鹏服务器和PC整机产品研发、适配测试、生产制造以及销售，持续夯实服务器产业基础，推进高端服务器产业化，做强服务器制造。

积极发展信创经济。以信创“两基地一园区一中心”为抓手，加快推动软件产业发展。在产品适配、产业聚集、技术攻关等方面发力，依托骨干企业，打造涵盖基础硬件、基础软件、应用软件、网络安全、云服务、系统集成的信创产业链条。

(五) 壮大企业主体力量

加快国内外重点企业招引。围绕大数据产业链的重点环节，依托省内重点企业和重点园区，加快招引一批在国际国内具有行业地位和重大影响力的龙头企业在我省落地，形成

“建好一个、带来一批”的磁场效应。针对我省大数据产业链薄弱和缺失环节，开展“补充式”“填空式”精准招商，培育引进一批核心企业，带动实施一批高端补链、整体强链的重点项目。

加大本土企业培育力度。建立湖北省大数据重点企业培育库，针对发展潜力好、技术创新能力强的大数据企业，“一企一策”强化财政资金、金融支持、税收优惠、人才激励等政策支持力度。支持省内大数据龙头企业加强产业优化整合，通过联合、并购、重组等方式做大做强，形成具备全国影响力的大数据龙头企业。鼓励中小企业充分发挥创新活力，不断提高自主创新能力，增强核心竞争力，走“专精特新”发展道路。

构筑产业融通发展体系。加快大中小企业融通发展，支持龙头企业持续发挥示范引领和带动作用，培育一批推广价值高的大数据解决方案，面向中小企业共享大数据技术资源、创新能力。积极探索大中小企业创新协同、产能共享、供应链互通的新型产业创新生态，引导组织大中小企业协作配合，促进企业、行业间在产业链和价值链的分工协作，支持产业链上下游企业共享信息、互动发展和协同创新。

重点企业招引

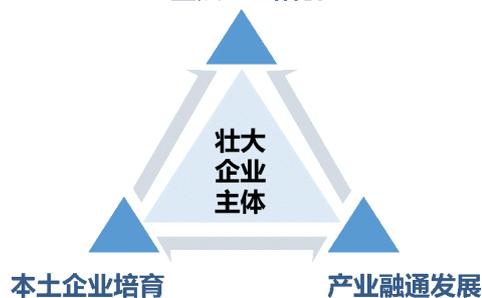


图7 湖北省“十四五”数据企业主体架构图

专栏5 大数据企业主体强化工程

实施大数据产业链链长制。坚持“招、培、创”相结合，培育一批有较强国际竞争力的链主企业。积极引导中小企业围绕头部企业和领军企业需求，提供配套产品和服务，培育一批专精特新“小巨人”、“隐形冠军”、制造业单项冠军。

推动大中小企业融通发展。支持大型工业企业、重点行业企业通过业务剥离，组建大数据骨干公司。支持建设大数据推广中心、实训基地等，开展供需对接、创业孵化、人才培养，推动中小企业通过数字化网络化智能化提高发展质量。加快大数据产业联盟、开源社区、行业协会等大数据组织建设，强化发展合力。

（六）提升安全保障能力

健全大数据安全保障机制。建立健全数据相关的法律法规和政策标准，为数据确权、个人隐私保护、数据要素化等提供依据。制定数据分级分类安全管理规则，明确各方主体的责任。探索建立信息采集和管控、敏感数据管理、数据交换标准和规则、个人隐私等大数据安全保障制度，完善大数据安全管理规则、管理模式和管理流程，切实保障数据传输安全、存储安全、审计安全和利用安全。

强化大数据安全支撑能力。加快安全态势感知平台建设，鼓励企业积极接入，增强网络空间安全防护和安全事件识别能力。开展安全监测和预警通报工作，加强大数据环境下防攻击、防泄漏、防窃取的监测、预警、控制和应急处置能力建设。支持大数据安全产业发展，加快国家网络安全人才与创新基地建设，培育数据安全龙头企业，推动安全产业提质升级。

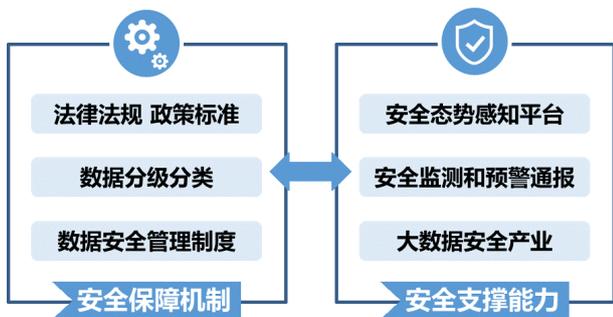


图8 湖北省“十四五”大数据安全保障架构图

专栏6 大数据安全防护工程

加强数据安全管理能力。建立数据分类分级保护制度，对数据实行分类分级保护。支持省内龙头企业实行工业数据分类分级试点，推动企业构建涵盖研发、生产、运维、管理、外部等五类数据域和高、中、低三级风险的分类分级管理体系。

建立工业信息安全态势感知平台。加快构建集监测、预

警、通报、处置为一体的工业大数据安全防护体系，推动省内工业企业接入安全态势感知平台，形成上下贯通、多方联动的安全态势监测与应急响应机制，实现闭环管理。

五、保障措施

（一）加强组织领导

实施大数据产业链链长制，组建工作专班，制定大数据产业链的推进工作方案，建立重大项目推进机制，研究出台政策举措。在链长领导下，建立“按季调度、半年通报、年度会商”机制，研究解决产业链发展重大问题，打造工作合力。在制造强省专家咨询委员会下设立大数据产业链专家咨询组，定期开展重点产业链运行监测和风险评估。

（二）加大政策支持

对接国家重大专项，积极争取国家重大计划、专项资金等各类资金对我省大数据产业长期发展的支持。充分发挥政府财政资金的导向作用，引导金融机构和社会资本等加大对大数据产业的资金投入力度，降低企业融资成本。支持各市政府资金和社会资本共同创立大数据产业基金，大力支持初创期、成长期的大数据服务企业。

（三）开展试点示范

支持省内企业积极开展国家大数据试点示范建设，围绕大数据关键技术产品研发、重点领域应用、产业支撑服务等重点方面，每年在全省范围内遴选一批大数据优秀应用案例，提升大数据对经济社会发展支撑服务能力。在全省范围内开展大数据优秀产品和应用解决方案遴选推广活动，加强交流学习开拓思路，推进大数据产业融合创新发展。

（四）强化人才支撑

做好大数据人才的培训和储备工作，将大数据领域高端人才作为重点引进方向，给予一定的人才津贴，在医疗、教育、住房等方面提供政策倾斜和支持。鼓励高校设立大数据课程，校企联合培养大数据跨界复合型人才和高端技术人才。依托高校、科研机构、企业资源等，建立一批联合教育实训基地，开展大数据关键技术和垂直行业融合相关培训。

（五）营造良好氛围

大力开展大数据产业对接、技术交流等活动，加强与国内外前沿技术团队的合作，充分利用国内外创新资源，推动我省大数据产业发展。积极策划组织我省大数据产业、企业和产品的宣传推介大会，向国内外企业和用户宣传推介我省投资环境、产业载体、优秀企业和重要产品，提升我省大数据产业的影响力。支持举办大数据创新应用大赛、大数据高峰论坛等品牌活动，营造大数据发展的良好氛围。

发改委：正在牵头制定数据要素基础制度文件

来源 / 湖北省经济和信息化厅 编辑 / 数据委员会处 李苗苗 日期 / 2022-04



发改委消息，5月17日，全国政协在京召开专题协商会，围绕“推动数字经济持续健康发展”进行协商议政。对于全国政协委员在会议上的建议，国家发展改革委副主任林念修从四方面作了简要回应。

关于超前部署新型基础设施，夯实数字经济发展基础的问题。目前发改委正在牵头加快推进新型基础设施建设，部署全国一体化大数据中心体系，实施“东数西算”工程。下一步，将重点推进两方面工作。一方面，强化统筹布局。引导大型、超大型数据中心向国家算力枢纽节点集聚，形成国家数据中心集群。加快推进国家枢纽节点间的直联网络建设，促进东西部算力高效互补和协同联动。另一方面，坚持绿色发展。强化能源布局联动，提升绿色能源使用比例，着力建设绿色节能数据中心。

关于深度挖掘数字要素价值，拓展数字经济发展空间的问题。按照中央部署，发改委正在牵头制定数据要素基础制度文件。下一步，将加快推动文件出台，着力建设四方面制度。一是建立保障权益激活价值的产权制度。二是建立合规高效、场内外结合的数据要素流通交易制度。三是建立

体现效率、促进公平的数据要素收益分配制度。四是建立安全可控、弹性包容的数据要素安全治理制度。

关于加快数字化转型，推动实体经济高质量发展的问題。去年，发改委牵头起草并报请国务院印发了《“十四五”数字经济发展规划》，明确了“十四五”时期推动数字经济发展的蓝图。下一步，将围绕规划落实，重点推动三方面工作。一是研究起草推动数字化转型的指导意见。二是组织专项工程，提升行业整体数字化水平。三是布局数字化转型促进中心，提供数字化转型公共服务。

关于探索完善政策环境，促进平台经济规范健康发展的问題。去年，发改委会同有关部门联合印发了《关于推动平台经济规范健康持续发展的若干意见》，下一步，将坚持规范和发展并重，重点推动两方面工作。一方面，健全平台经济治理体系，推动构建稳定可预期的政策环境。另一方面，推动平台经济高质量发展，鼓励平台企业增强创新能力，赋能实体经济，提升国际化发展水平。

国知局：将加快推动数据产权保护制度的完善

来源 / 通信世界 编辑 / 数据委员会处 李苗苗 日期 / 2022-04



数据被称作信息时代的新能源，与土地、资本、技术等传统要素并列，是数字经济发展的基础。我国是一个数据资源大国和数字经济大国，2020年，以数据为支撑的数字经济增加值规模已经达到39.2万亿元，占GDP的比重达到38.6%，位居全球第二。但要想让这些数据资源合理流动起来，充分利用起来，有效保护起来，就需要很好解决数据的产权问题，完善相关制度设计。

《知识产权强国建设纲要（2021—2035年）》和《“十四五”国家知识产权保护和运用规划》都对构建数据产权保护规则作出部署，要求实施数据知识产权保护工程，深入开展相关理论和实践研究。国家知识产权局局长申长雨在近日召开的2021年中国知识产权发展状况新闻发布会上讲道，国家知识产权局高度重视数字知识产权，按照中央有关部署，正结合部门职能和工作实际，加快推动相关研究工作，并专门成立了工作专班，与有关部门、专家学者、市场主体等进行了广泛深入的讨论，也形成了一些初步的成果和原则性思路。

一是要坚持安全为先、发展为要，充分考虑数据的安

全、公共利益的保障和个人信息的保护。

二是要充分把握数据的特有属性和产权制度的发展规律，实现数据保护模式与数据资源特性的高度契合，数据产权保护和数据有效利用的有机统一。

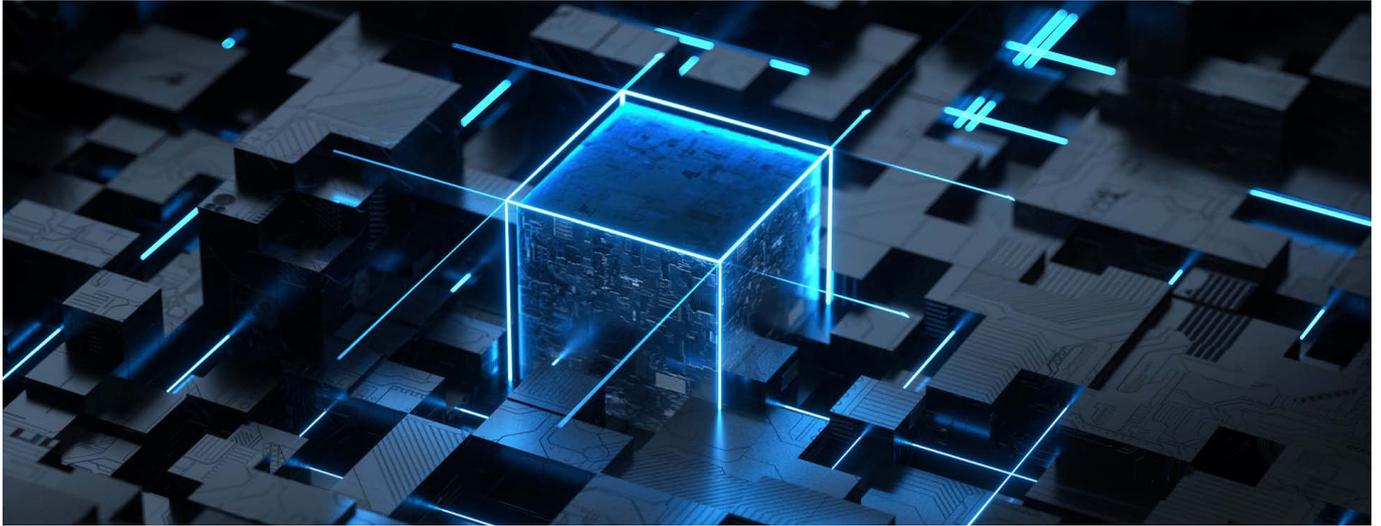
三是要充分尊重数据处理者的创造性劳动和资本投入，承认和保护数据处理者的合理收益。

四是要有利于数据的流动和交易，能够促进繁荣数据市场，壮大数字产业，促进产业数字化转型，支撑经济高质量发展。

目前，知识产权局已经在浙江、上海、深圳等地开展了数据知识产权保护试点，力争在立法、存证登记等方面取得可复制可推广的经验做法，为后续制度设计提供实践基础。其中，浙江已经建立了数据知识产权公共存证平台，并开始面向市场主体提供存证服务。下一步，知识产权局将继续加强与各方的合作，加大理论研究和实践探索，加快推动我国数据产权保护制度的建立和完善。

中央网信办曹淑敏：互联网企业要加大研发投入， 让创新成为发展的核心竞争力

来源 / 通信世界 编辑 / 数据委员会处 李苗苗 日期 / 2022-04



5月26日，2022中国国际大数据产业博览会线上开幕。在此次大会上，中央网络安全和信息化委员会办公室副主任、国家互联网信息办公室副主任曹淑敏指出，我国是数据资源大国，在数据资源开发利用初见成效，但我国数据领域还存在一系列问题，需要加快建立基础制度，推动数据开放开发和价值释放，助力高质量数字经济发展。

为推动大数据产业发展，曹淑敏提出四点建议。

一是强化顶层设计，加快数字中国建设。

大数据是数字化发展的基础和主要特征，也是数字中国建设的核心要素。当前，中央网信办正在积极推动数字中国整体布局，中央网信办将以数据资源为核心，以数字基础设施为支撑，以数字技术为驱动，以数字治理和数字安全为保障，将数字化发展全面融入“五位一体”总体布局，推动建设高质量的数字经济，高效协同的数字政府，自信繁荣的数字文化，普惠包容的数字社会，绿色智慧的数字生态文明，以体制机制改革整体驱动生产生活治理方式数字化变革。

二是加快核心技术创新，实现高水平科技自立自强。

互联网企业是网信领域的创新主体，据统计，2021年我国67家上市互联网企业研发投入同比增长32%，呈现良好发展态势。但是也要看到不足，从研发占营收比例看，国内上市互联网企业平均为6.9%，与国外同行差距明显，需要进一步加大研发投入，让创新成为企业持续健康发展的核心竞争力。同时，我国要加大对集成电路、操作系统、工业软件等

关键核心技术的协同攻关，营造充满活力的创新环境，建设互利共赢的产业生态。要继续强化企业创新主体地位，加强原创性、引领性科技攻关。构建开放创新生态，激发人才创新活力，助力实现高水平科技自立自强。

三是激活数据要素潜能，打造高质量数字经济。

数据是数字经济的关键生产要素。我国是数据资源大国，据调查，2021年我国共生产数据6.6ZB，约占全球10%。近年来，我国数据资源开发利用初见成效。近5年，全国省级公共数据开放平台从5个增至24个，开放的有效数据集从8398个增长至近25万个，但同时也要看到，我国数据领域还存在权属制度缺失、流通规则不明、数据治理缺乏统筹等问题，需要加快建立完善数据权属、流通、交易等基础制度，推动数据开放开发和价值释放，助力高质量数字经济发展。

四是推动数据惠民便民，建设普惠包容数字社会。

“十四五”规划纲要就“加快数字社会建设步伐”明确要求，提供智慧便捷的公共服务，建设智慧城市和数字乡村，构筑美好数字生活新图景。近年来，中央网信办会同有关部门，深入开展人工智能社会实验和数字乡村建设，建设了一批国家智能社会治理实验基地和数字乡村试点。下一步，中国网信办要继续聚焦教育、医疗等重点领域，发挥试点示范作用，推动数字化服务普惠。

中国信通院闫希敏：个人信息保护视域下 知情同意框架的应用困境与对策探析

来源 / 中国信通院 编辑 / 数据委员会处 李苗苗 日期 / 2022-04



在大数据时代，个人信息成为推动数字经济快速发展的资源性要素。互联网信息服务提供者（以下简称“企业”）通过对个人信息进行处理分析，向用户提供精准化便捷化的互联网信息服务。在实际应用中，企业以知情同意框架——即告知用户其个人信息的收集处理规则并获得同意的方式，构建收集和处理个人信息的合理性基础。该框架目前虽被全球立法广泛采纳，但是其自身应用困境在大数据时代也日趋明显。本文将阐述知情同意框架的应用困境，深入剖析其产生原因，并进一步探析对策建议。

1 知情同意框架的基本情况

知情同意源起医疗语境，医生充分告知患者相关治疗方案的风险，患者在充分知情的前提下，选择某种治疗方案或者作出是否接受某种治疗方案的决定。后这一理念应用于个人信息保护领域，强调用户在知悉企业收集其个人信息的目的、方式、内容、数量、后续处理情况等的前提下，选择是否同意企业收集处理其个人信息。在一定程度上，知情权为弥补用户和企业之间在个人信息收集与处理方面存在的信息势差提供了基础，同意权提升了用户对其个人信息的掌控能力。实践中，企业和用户常常通过隐私协议等形式缔结合约，用

户通过阅读相关协议内容，了解其个人信息被收集处理以及后续流转情况，并通过同意相关协议，授予企业收集处理其个人信息的权力。

当前，知情同意框架已经成为个人信息保护相关法律的基本原则。2012年出台的《关于加强网络信息保护的决定》在国内首次提出知情同意原则。2013年出台的《电信和互联网用户个人信息保护规定》，明确强调企业在收集、使用用户个人信息时必须遵守知情同意原则。《网络安全法》《民法典》《个人信息保护法》均以知情同意原则作为企业个人信息收集处理行为的正当性基础。

2 现实应用困境

尽管通过同意隐私协议进而授权企业收集处理用户的个人信息的做法十分普遍，几乎成为互联网时代所有应用在个人信息收集处理方面的“公因式”，但在实际操作中，知情同意效果欠佳，对其的批评不绝于耳。当前知情同意的现实应用困境主要包括以下方面。一是隐私协议被阅读率低，几乎沦为形式主义产物。用户使用某应用，大多不会仔细阅读相关隐私协议，而是直接选择同意后继续获取服务。二是隐私协议条款较为模糊，指向性不明显。企业往往在隐私协

议中向用户告知较为宽泛的信息收集使用情况，涵盖范围较广，语义模糊。三是隐私协议多设置打包式同意模式。隐私协议条款众多，包含个人信息从收集、传输到使用、共享、销毁全流程的内容，但仅提供一个整体“同意”或“不同意”选项。用户可能对某些条款持不同意态度，但无法单独反馈该部分意见，只能选择全部同意或全部不同意。

以上知情同意的现实应用困境，使隐私协议等收集使用个人信息规则成为平衡专业化规避法律风险与普及式告知用户相关概况的产物，仅从形式上约束企业收集处理个人信息行为，一旦其文字（例如涵盖收集个人信息的目的、方式、范围）及表达形式（例如易访问、易阅读）层面达到合规要求，即成为企业对安全使用个人信息的承诺书，普通用户既无核实企业个人信息保护情况的能力，在给出同意意见后也缺乏据此维权的依据，即知情同意框架难以约束企业的个人信息收集处理行为。

3 应用困境的影响因素分析

用户的知情同意是用户就本人事务进行自决的行为，只有在各有关主体在经济、社会等诸多方面力量均衡的条件下才能保证约束效果。知情同意框架的应用困境根本原因在于用户和企业的信息不平等地位，这种不平等体现在三方面：

一是信息获取的不平等，用户与企业相比，更难获悉其个人信息的处理情况；

二是信息掌控力的不平等，用户对其个人信息的掌控力欠佳，作为实际处理者的企业却可在授权下决定海量个人信息的处理方式；

三是责任承担的不平等，个人信息一旦泄露或滥用，用户需承担其对个人生活甚至人身财产安全造成的风险，企业的相关风险却微乎其微。当前，诸多因素共同决定了用户和企业不平等的信息地位，因此用户难以真正地在知情和理解的基础上自主行使对于其信息采集和后续处理过程的自决权利。

3.1 用户侧因素在实践中，用户不会按照既定期望行为行事，即认真阅读隐私协议→做出自主同意与否决定→使用该互联网信息服务或更换其他服务。用户通常以不辨别内容即同意隐私协议为代价，换取选定的互联网信息服务的快速即刻使用。该情况的出现源于以下两方面原因。

一是用户的专业性意识不够，对其个人信息保护的必要性认识不充分。例如有些用户的想法是“收集了一些不敏感的个人不会有太大影响”“手机号都是公开的没有问题”“普通人的个人信息不会被关注”等，受非理性、乐观意见引导，未能深刻认识到在万物互联的数字时代无数个普通人个人信息的泄露将可能对个人和对社会带来怎样的严重后果。

二是用户的个人信息保护专业性能力不足，难以仅通过隐私协议等文件明白其个人信息被处理可能存在的风险。隐私协议同时兼任法律责任明确作用，故相关表述字句专业化

程度较高，对于普通用户而言，快速阅读并准确理解难度很大，遑论据此进行同意与否的意见表达。

3.2 企业侧因素在实践中，企业作为营利性机构，一般选择更有利于其经济利益发展的做法，也就是说，在大数据时代，企业更有动力去获取大量用户个人信息，通过分析挖掘进而充分赋值数据。从企业角度，以下原因造成了知情同意框架的现实应用困境。

一是企业利用相对于用户的信息不对等地位，更易制定利于己方的隐私协议。企业了解需要收集哪些必要个人信息，但是可出于其他目的（例如多收集信息以备不时之需、掌握更多个人信息进行数据挖掘分析等）在隐私协议里列出多于其必要收集范围的个人信息，而这一方式在知情同意框架下成为收集这些额外个人信息的最好方式，反向赋予企业行为合理性——即企业基于用户同意这一条万能理由，即可把不合理的收集范围合理化。

二是某些互联网信息服务市场内存在市场占有率极高的行业巨头企业，用户几乎不具备服务选择权。一些互联网服务垂直领域内已形成事实上的垄断或寡头现象，用户面临非此即彼的服务选择或几乎没有可选择范围。行业巨头在其服务领域内具备显著的缔约优势，用户拒绝隐私协议则不能使用服务，需承担丧失便利生活、被隔绝于网络社交之外的风险，其知情同意权几乎形同虚设。

三是业内企业的知情同意框架设计向龙头企业看齐，而非向规则看齐。在实际的调研中发现，即便是在一些新兴的互联网信息服务领域内，业内企业的知情同意框架也是参照其他互联网信息服务龙头企业的方式进行设计，且企业以此作为相关机制合理性的支持依据。一旦龙头企业的知情同意框架设计存在不足之处，将导致一大批企业的知情同意机制存在缺陷。

3.3 客观因素此外，仍有一些客观因素，导致知情同意框架存在应用困境。

一是针对隐私协议的部分要求存在内在的矛盾和冲突。简短明了的内容难以负载繁多的告知事项，追求授权程序的全面带来不便捷的用户体验，而规范专业的文字表述通常导致理解门槛的提升，这些矛盾点结合在一起，导致保障用户知情权的隐私协议难以兼顾各方需求，只能寻找相对平衡点。

二是人工智能算法等技术的学习能力增强个人信息后续处理的不可预测性。随着人工智能、大数据分析等新技术和用户画像、个性化推荐等新应用的不断发展更新，个人信息的后续处理存在一定的不可预知性，即便是企业，对于收集而来的个人信息如何处理分析、应用在哪些方面、涉及哪些主体等，也很难在用户同意提供个人信息前进行明确并且保证后续不会变动。

尽管企业隐私协议会不断完善，内容更加完整、表达形式更加生动，但是以上因素将导致其保护个人信息的象征意

义大于实际意义。

4 对策探析

近两年来,相关监管机构充分关注用户个人信息保护,出台一系列管理举措。其中,在个人信息收集环节,基于知情同意框架,对于隐私协议条款、提醒方式、同意方式、权限授予方式及频度等进行了详细的规定,其隐含两层含义:一是按照这样的要求对用户进行告知,用户能够达到“知情”水平,并充分给予“同意”的自主性表达。二是默认企业达到这样的形式要求后,其个人信息收集处理行为即完全按照公开的隐私协议进行。其实这对用户行为的理性期待和对企业行为的合规期待过高。

以现实情况来看,仅以一些强制手段维系用户的知情同意权意义极为有限,为实现用户和企业双方的信息地位均衡,可尝试第三方力量干预、介入的路径,在用户和企业就信息收集和处理行为的潜在角力中,补充用户这一弱勢方对其个人信息的控制能力。本文从降低用户和企业的信息势差,增加用户对其个人信息收集处理的控制力出发,探析对知情同意框架的修改与完善。

一是以法律规范明确个人信息收集处理规则和处罚措施,制约企业对于个人信息收集处理的自我合理化能力。政府可借助第三方独立机构的力量,深度介入专业领域,推动出台同时具有专业性和约束力的个人信息收集处理规则,明确核心业务功能、所需最小化个人信息范围、频次,厘清相关业务的必要性收集和非必要性收集范畴,并明确相关处罚措施。

例如,《常见类型移动互联网应用程序必要个人信息范围规定》的印发有助于打破企业在个人信息收集处理领域的解释权垄断,使得企业在存在不合理收集处理信息行为时难以依靠技术和信息优势进行自我行为合理性证明。

二是深入企业内部,强化个人信息处理后端环节的监督检查力度。以现有的个人信息保护专项工作来看,主要以远程测试方式对个人信息的收集、传输环节的违法违规行为进行检查,并且取得一定成效,但无法确定处理、存储、删除

等后端环节的企业实际情况是否符合相关要求,未来亟需以现场问询、资料查看、系统演示、样本测试等方式,在企业内开展稳定、持续的个人信息安全保护能力检查工作,以监督检查压力促使企业提升个人信息保护能力。

三是推进信息披露市场化,引入竞争机制约束企业行为。借鉴金融领域的上市公司财务报告和信用评级制度,推进互联网领域的个人信息收集处理相关信息披露的市场化,要求企业将个人信息收集处理情况以年报形式进行公开披露,接受公开监督,鼓励第三方独立机构作为隐私协议的真实读者群和企业信息行为的评估者,基于隐私协议和收集处理行为等出具对于相关企业的个人信息收集处理行为的“评级”报告,推动用户优先选择个人信息保护举措到位、没有泄露滥用信息记录的企业推出的相关应用,借助市场的力量推动企业自发提升其个人信息保护水平。

四是推动企业探索建立个人信息收集处理的分层动态同意机制,为用户提供更大选择范围。一方面,探索为用户提供针对不同功能、不同个人信息处理环节的分别告知和获取同意方式,改善当前以单个隐私协议的总体一次同意授予企业永久收集使用个人信息权利的现状;另一方面,设置“同意自动过期”或“用户选择退出”等保障用户被遗忘权实现的功能,畅通用户收回同意的通道,并借助监督检查工作推动个人信息的删除销毁工作落实到位。

五是推进事后救济能力建设,降低用户个人信息保护维权成本。推进相关个人信息维权救济组织建立,作为隐私协议的真实读者群,在用户个人信息泄露和滥用事件或者企业不合理信息行为发生后,能够有效提供个人信息不当收集利用的完整证据链条,帮助用户提升索赔成功率,以个人信息侵权诉讼的压力推动企业加强个人信息保护及提升相关信息行为的规范性。

5 结束语

知情同意框架在个人信息保护法理基础方面具有不可替代的作用,为充分利用该框架约束企业收集处理个人信息行为,可尝试从降低用户和企业的信息势差角度出发,引入政府和第三方机构等其他力量进行干预,降低企业利用其信息优势地位不合理获取个人信息的可能性,使知情同意框架更好地实现其应有之义。



肖亚庆：构建稳定高效的大数据产业链要从两个方面重点着力

来源 / 通信世界 编辑 / 数据委员会处 李苗苗 日期 / 2022-05

工业和信息化部部部长肖亚庆在5月26日举办的“2022年中国国际大数据产业博览会”上指出，我国要构建稳定高效的大数据产业链，要从两个方面着力发展：一坚持适度超前建设、以建带用、以用促建，推动数字基础设施建设与应用场景协同发展。二要开展大数据技术、产品和服务协同创新，补齐关键技术短板。

2022年中国国际大数据产业博览会主题为“抢数字先机享数字价值”。“十四五”时期，中国数字经济以消费互联网为主要增长点的“上半场”蓬勃发展，以产业互联网为主要增长点的“下半场”势头强劲，数字经济大有可为。随着5G等新一代信息技术加速普及，新一轮超大规模数据浪潮已经到来，中国经济社会数字化转型正在积累从量变到质变的势能。

肖亚庆指出，当前我国大数据产业发展取得明显成效，产业规模快速增长。“十三五”时期，我国大数据产业年均复合增长率超过30%，2021年产业规模突破了1.3万亿元。大数据产业链初步形成，一批龙头企业快速崛起。目前大数据应用从互联网、金融、电信等领域逐步向智能制造、数字社会、数字政府等领域拓展，极大丰富了我国数据资源，催生一批新场景新模式新业态，行业融合逐步深入。

“中央出台一系列政策文件，地方探索制定数据条例等地方性法规，为大数据产业营造了良好的发展环境。目前我国已设立了大数据管理机构、组建数据交易所，数据确权、定价、交易等标准制度也不断完善。”肖亚庆说道。

“我们要顺应数字化发展趋势，坚定不移实施国家大数据战略，统筹发展和安全，以释放数据要素价值为导向，协同推进基础设施、技术创新、融合应用和治理体系建设，加快数字产业化、产业数字化，为制造强国、网络强国、数字中国提供有力支撑。”肖亚庆说到。

对于今后我国大数据产业发展，肖亚庆建议从4个方面重点发展。

一是加快建设数字基础设施。坚持适度超前建设、以建带用、以用促建，推动数字基础设施建设与应用场景协同发展。全面推进5G网络和千兆光网建设，加快工业互联网、车联网等布局，促进数据高效采集和传输。统筹布局数据与算力设施，加快建设国家工业互联网大数据中心，合理部署超级计算中心，打造若干国家大数据枢纽节点和区域中心集群。



二是着力拓展融合应用场景。围绕促进产业转型升级、提高政府治理效能，加快数字社会建设。开展大数据技术、产品和服务协同创新，补齐关键技术短板，构建稳定高效的大数据产业链。以制造业数字化转型为引领，实施工业大数据价值提升行动，构建多层次工业互联网平台体系，培育一批专业化、场景化大数据解决方案，发展一批面向细分场景的工业APP，推广一批基于数据驱动的新技术新业态新模式。

三是推进数据资源整合和开放共享。采好数据、管好数据、用好数据是大数据产业发展壮大的关键。将加强部门联动、省部协同，加快数据确权、交易流通、跨境传输和安全等基础制度规范建设，探索多种形式的数据交易模式，鼓励互联网企业、电信运营商、工业企业等有序开放数据，实施数据管理能力国家标准，继续支持数据交易所建设、完善运营体系，培育一批数据交易服务商。

四是深化国际务实合作。将坚定不移贯彻对外开放的方针，支持国内外大数据企业在技术研发、标准制定、产品服务、知识产权等方面深入合作。支持国内企业“走出去”开拓国际市场，支持跨国公司、科研机构在国内设立大数据研发中心、教育培训中心。

用好大数据，助力市场数字化升级

来源 / 人民网 编辑 / 数据委员会处 李苗苗 日期 / 2022-05



随着我国数字经济不断发展壮大，对市场的影响程度越来越深、辐射范围越来越广。以“抢数字先机，享数字价值”为主题的2022中国国际大数据产业博览会（简称“数博会”）线上活动，将于5月26日举行，届时会发布和展示众多领域大数据相关的领先科技成果。在数博会即将召开之际，4月10日国家又发布了《中共中央、国务院关于加快建设全国统一大市场的意见》。其中，提到要推进市场设施高标准联通。因此，要用好大数据，助力市场数字化升级尤为重要。

用好大数据，助力市场数字化转型。一方面要加强优化商贸流通基础设施，比如加快公共资源交易平台数字化转型升级。通过深化资源、整合共享用好大数据，改变过去数据市场乱、资源散、流通缓、交易小、治理慢等现象，推进公共资源交易平台数字化进程覆盖面广，商贸流通模式多的业态新模式。另一方面是市场数字化转型升级可向农村地区延伸。用好大数据，建立健全城乡融合、区域联通、安全高效等市场数字化基础设施网络，以此服务于乡村，发展振兴乡村。

用好大数据，助力市场注入新活力。一是用好大数据能进一步推动线上线下融合发展，改变过去线上线下融合发展

不完善、交易平台不健全的情形；二是促进市场供需关系转型升级，用好大数据可以给企业提升现有技术，以及产品服务的供给能力，同时也能大幅度提高企业的生产制造、经营管理、商贸流通等环节的效率。另外，通过大数据分析消费者需求，实现有效细分客户、防控风险、评价信用等，对企业的产品研发更具有针对性和导向性，能有效激活线上线下消费需求，有效释放更大更多的消费潜力，从而为市场注入新活力。

用好大数据，助力市场信息交互新渠道。一是用好大数据能推动各领域、各企业间的技术与资本，各地区劳动力，以及城乡土地等生产要素的大力变革和优化重组，解决市场经营信息不充分、不对称的痛点问题，有效提升生产效率，实现公共资源信息互通共享，以此促进跨网、跨地区、跨企业数据交互，推动产业链和价值链互联互通。二是各领域市场公共资源信息互通共享，又会倒逼加快推进同类型及同目的信息认证平台统一接口建设，完善接口标准，促进市场信息流动和高效使用，最终实现市场信息交互新渠道。

非线性回归充电之：时间序列分解

作者 / CPDA数据分析师 刘程浩 编辑 / 数据委员会处 李苗苗 日期 / 2022-05

从最早学《统计学》、《金融工程》乃至后来在数据分析师的教材知识里，都有介绍对时间序列的分解。不过我发现国内大多只是介绍what和How，最重要的why没有说，或者模棱两可说两句，看了后没有深入理解。

时间序列分解的经典模型，最早诞生于上个世纪20年代，后来成熟应用于上个世纪50年代，而后被美国人口普查得以深度应用并持续改进，产生了非常多的迭代。例如X11, X11-ARIMA, X12, X13……不同版本。

除了人口普查之外，很多经济现象和经济数据分析也深度应用到了这个知识领域。就以人口普查为例。人口普查的战略指导应用很多，对于每个国家来说都很重要。

那为啥美国这么重视人口普查？首先美国众议院435个席位是根据各州人口分配的。每10年美国都要展开人口普查，并重新调整这些席位的分配。例如加利福尼亚州和德克萨斯州，1920年时只有11个和18个席位；随着人口增长到了2020年，席位数增长到了52个和38个了！

<BBC>说2020年加州是52个，但其他网站例如<维基百科>则说有54个，但大体上差不多。

而到了美国总统选举时，每个州的众议院席位数量就起到作用了，因为决定总统宝座花落谁家的选举人票数，其中有435票就来自众议院的这435个席位。

加利福尼亚州的人口最多，因而选举人票也最多，有52张；而一些人口稀少的州，比如怀俄明、阿拉斯加和北达科他（还有华盛顿特区）等，选举人票数量只有3个。

等坐上了总统宝座后，每个州的众议院席位数量也会起到作用。因为总统的施政纲领未来落地是否顺利，也是要看支持自己的“兄弟手足”在参众两院的席位多少。

以上等等，说到底人口数量和总统饭碗能不能端上，以及以后能不能端稳有如此紧密的联系，也难怪人家这么在乎人口普查了。

介绍完了美国人口普查的花絮之后，以下为正文：

● 为什么要做时间序列的分解？

以经典的时间序列为例

经典的时间序列分解，是将一条时间序列看成3个部分组成：

$$y_t = S_t + T_t + R_t \quad \text{或} \quad y_t = S_t \times T_t \times R_t$$

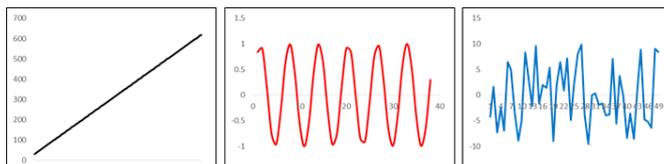
其中， S_t ：季节波动部分（短周期波动）；

T_t ：趋势部分；

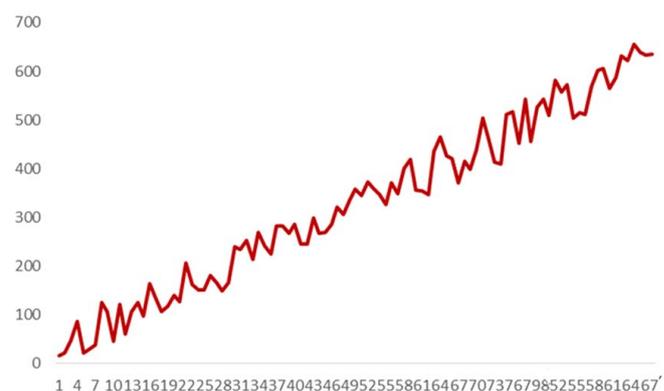
R_t ：随机扰动部分。

选择加法或者乘法做分解，关键是看原始数据的呈现特点。

我随机设定了一条直线方程，一个正弦波，还有一组白噪声数



首先，把这3种元素通过线性加法加起来，得到下面的结果



这个结果的数据，看上去就是一个有固定倾斜角的波浪线，但每处的波浪变化都和其他处有一些不同。

并且整体上看，从头到尾波浪线的振幅都相对稳定，没有逐渐放大或者逐渐缩小。

接着，把这3种元素通过乘法计算，得到下面的结果



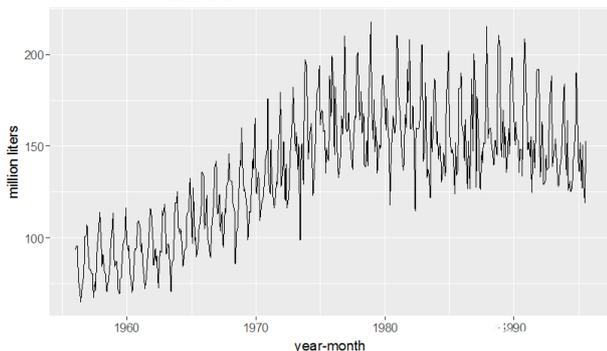
这个结果的数据，看上去是一个逐渐放大振幅的波浪线，和之前加法的波浪线相比，随着时间推移，不确定性变

得越来越大。

所以，选择用加法还是乘法去分解一个时间序列，关键的选择还是看数据的特点表现。

在这里，为了更接地气一些理解为什么要采用时间序列分解来分析数据（why），我列举了澳大利亚的在上个世纪近50年来的月度啤酒产量的数据。

澳大利亚啤酒产量历史数据



如果你是一个刚到澳大利亚留学的学生，准备在课余时间兼职卖啤酒勤工俭学；

或者你是一个刚进入啤酒销售行业的新人，从上图的历史数据来看，你能得到哪些信息？

我正好也从没有做过啤酒行业，因此我尝试了一下：

首先，澳大利亚啤酒的产量在1955-1975这段时间的销量一直是上升的；

接下来，从1975-1995这20年啤酒产量似乎有些稳中下滑，但下滑的幅度看不太清楚；

然后，每年的不同月份的啤酒产量波动也很大，但波动也有些规律，那应该是和天气冷暖有关。

最后，1975-1995这段期间，的啤酒产量的波动，要略比1955-1975这段期间大些。

以上4个观点中，之所以第2个语句里我用红色字体标记，主要是因为啤酒产量因为受到季节和随机因素的影响，波动太大，它的变化趋势的确没那么容易捕捉到。

而这也是很多经济学家看到经济数据时所共同关心的地方。

因为很多时候专家们想首先通过数据的趋势变化，去假设和解释这些变化的原因。但由于季节和随机因素的影响，造成了直观上捕捉趋势变化比较困难。

另外还有一种原因，就是专家们对季节性变化并不关注，他们知道有淡旺季，但就是想剔除淡旺季的影响去看整体趋势的变化。

例如，国外研究就业率的机构，他们知道毕业季大量的学生一下子没有找到工作，因此就业率在毕业季会明显下降，但就业率很快也会随着时间的推移而回升。这些机构想了解的是社会的就业率整体的变化趋势是上升还是下降，就必须剔除掉季节变化的影响。

也就是说，时间序列分解，给研究人员提供了对单一的

时间序列数据的多种分析选择。在这里以加法分解作为一个例子：

$$y_t = S_t + T_t + R_t \text{ 研究整体数据；}$$

$$y_t = T_t + R_t \text{ 或 } y_t = T_t \text{ 只研究趋势变化情况；}$$

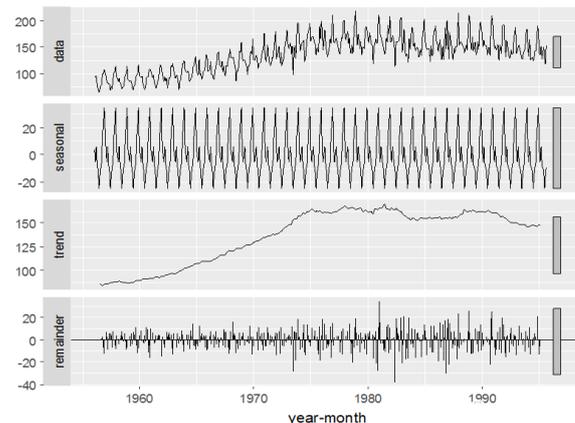
$$y_t = S_t + R_t \text{ 或 } y_t = S_t \text{ 只研究波动周期；}$$

$$y_t = R_t \text{ 只研究随机/特定事件的影响；}$$

以上内容，回答了为什么要对时间序列做分解的why！

为了更直观的展现啥叫时间序列分解，以下就是以经典的方法为例，对澳大利亚啤酒产量做的分解

澳大利亚啤酒产量序列分解



上图就是我们想要看到的序列分解了。

第一个图是原始数据，第二个图是只看趋势，不考虑季节波动和随机扰动因素的情况下数据的变化。

仅从第二个图我们可以清楚的看到，1975年之后，澳大利亚的啤酒产量的趋势是停止增长了，并且维持到1983年左右。之后有几年的降幅，但也还稳定；但到了1990年后，开始下降。

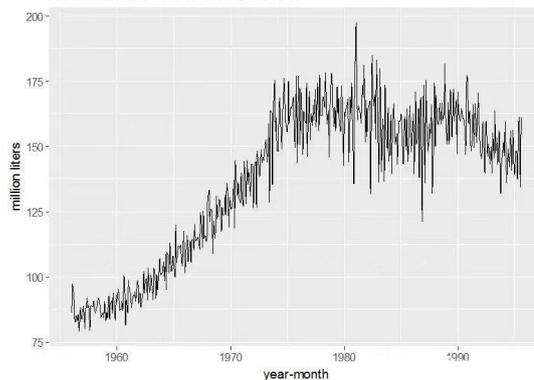
这样，就可以清晰地回答刚才前文红色字体的难题了。

接下来就可以根据趋势的变化，想办法去找到解释这一变化的原因，做更深入的研究。

那假如我只想看剔除掉季节变化，仅考虑趋势+随机扰动的情况如何？

可以做成下图进行分析

澳大利亚啤酒产量-趋势和随机扰动



只看趋势和随机扰动，也能比较清晰的看到和之前一样的结论，因为季节性波动的影响，导致的数据曲线振幅过大影响阅读的不利因素被大大弱化了。

而且，我们发现1981年初和1987年初，似乎随机扰动导致的数据变化非常大，这样我们可以尝试寻找历史的事件，去解释这些变化，看看是否做别的变量引入。

写到这里，我心中一直憋了很久的疑问就解决了。但我并不打算就此收笔，想再稍微深入研究下。

●如何做时间序列分解？以经典的模型为例

（一）首先，如何做趋势T？

经典模型的方法就是做移动平均数。

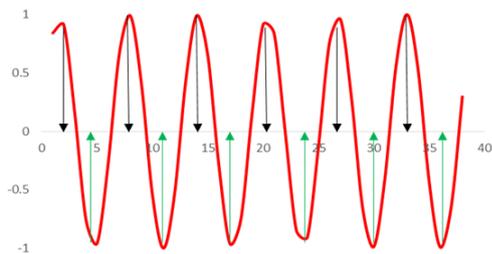
有人说，移动平均还不简单，小学或初中都学过，像什么移动3期平均啊，移动5期平均啊。

那问题来了，为什么要选择移动平均，选择移动平均的期数有什么讲究？

其实选择移动平均，主要就是要抹掉季节变化的影响。

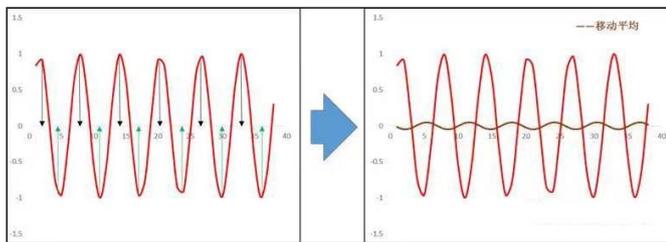
很多书本和教材都这么写，一般人看了就看了，就这么记着。但我却要多想深入一点，为什么说移动平均能够消除季节波动影响。

一种方法是不看广告看疗效，回到之前我随机的正弦波例子：



其实求平均值，就是削峰填谷。例如有一个段：我和姚明的平均身高是2米，就是把他身高的峰来填我身高的谷。用上图来看，就是黑色箭头和绿色箭头出发点的数据往均值0的位置变化。

执行平均值一次，只会削峰填谷一次，如果不断的按照固定期数往后“移动地”平均，那么就会不断的削峰填谷，从而得到一个比较平滑的曲线，就如下图右侧的结果一样。



那第二个问题，移动平均的期数是如何定的？

这其实和周期波动的周期长度和时间序列颗粒度有关。

这个就很多场景了，如果时间颗粒度是天，波动或峰谷期是一周内，那么就选7，例如旅游景点的门票收入或人流量；

如果时间颗粒度为月，波动的峰谷期是要跨完一整年，那么就选12。例如羽绒服厂，这次说的啤酒的销售数据。

但，出于对称性的要求，一般能选择奇数期次就选奇数。

因为看下面的式子，假如波动的周期是5，那么1~5期就会有有一个中心数3，第3期前后各自的观测值就会各自为波峰和波谷，

$$\frac{y_{-2} + y_{-1} + y_0 + y_1 + y_2}{5}$$

所以，在执行移动平均数计算时，第一次计算（也就是削峰填谷）的结果，应对应放置在第3个观测值位置。执行最后一次移动平均数的时候，计算结果应对应放置在倒数第三个观测值位置。这样，整个时间序列前后各2期的观测值位置，刚好缺失了对应位置的移动平均值。因此就显得对称。

如果，周期波动次数是偶数，为了保证对称性，要做2次偶数次的移动平均。

看下面的式子，假如波动的周期是4，那么1~4期没有整数的中心数，在执行移动平均数计算时，第一次计算（也就是削峰填谷）的结果，只能放置在第2个或第3个观测值位置。

$$\frac{y_{-2} + y_{-1} + y_1 + y_2}{4}$$

而最后一次计算的结果，同样也只能放置倒数第3个或第2个位置。这样就不能对称。

如果再做一次偶数次的移动平均，结果就不一样了。

例如，再做一次移动2次平均

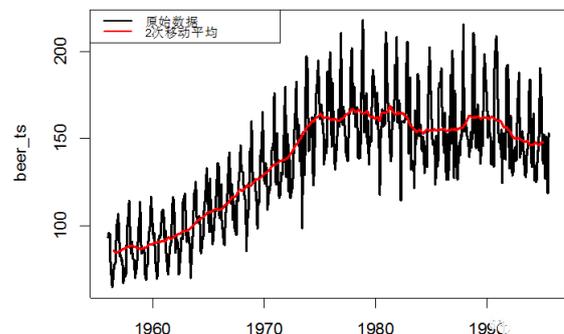
$$\frac{\left(\frac{y_1 + y_2 + y_3 + y_4}{4} + \frac{y_2 + y_3 + y_4 + y_5}{4} \right)}{2}$$

$$= \frac{1}{8}y_1 + \frac{1}{4}y_2 + \frac{1}{4}y_3 + \frac{1}{4}y_4 + \frac{1}{8}y_5$$

这样，就变成了1~5期的加权平均数，并且有一个中心期数。

还是以澳大利亚啤酒产量为数据，计算经典模型下的12期移动平均，之后再做2期移动平均

澳大利亚啤酒产量2次移动平均





不过这个是一个简单的特例，下面我再花点儿时间证明一下一般性。

感兴趣的读者可以一直往下看，对证明不感兴趣的读者可以跳过这个部分。

假设有偶数M和N，其中M是第一次移动平均期数，N是第二次移动平均期数。

若M=N，2次移动平均的计算过程如下：

$$\frac{1}{N} \left[\begin{array}{l} \frac{1}{M}(y_1 + y_2 + y_3 + \dots + y_M) \\ + \frac{1}{M}(y_2 + y_3 + \dots + y_M + y_{M+1}) \\ + \frac{1}{M}(y_3 + y_4 + \dots + y_M + y_{M+1} + y_{M+2}) \\ \dots \\ + \frac{1}{M}(y_N + y_{N+1} + y_{N+2} + \dots + y_{N+M-1}) \end{array} \right]$$

由于M=N，整理得

$$\frac{1}{M} \left[\begin{array}{l} \frac{1}{M}(y_1 + y_2 + y_3 + \dots + y_M) \\ + \frac{1}{M}(y_2 + y_3 + \dots + y_M + y_{M+1}) \\ + \frac{1}{M}(y_3 + y_4 + \dots + y_M + y_{M+1} + y_{M+2}) \\ \dots \\ + \frac{1}{M}(y_M + y_{M+1} + y_{M+2} + \dots + y_{M+M-1}) \end{array} \right]$$

$$= \frac{1}{M^2} \left[\begin{array}{l} y_1 + y_2 + y_3 + \dots + y_M \\ + y_2 + y_3 + \dots + y_M + y_{M+1} \\ + y_3 + \dots + y_M + y_{M+1} + y_{M+2} \\ \dots \\ + y_M + y_{M+1} + y_{M+2} + \dots + y_{M+M-1} \end{array} \right]$$

仔细观察上式右侧，可以得到以下：

$$\frac{1}{M^2} \left[\begin{array}{l} y_1 + 2y_2 + 3y_3 + \dots + (M-1)y_{M-1} \\ + My_M \\ + (M-1)y_{M+1} + (M-2)y_{M+2} + \dots + 2y_{M+M-2} + y_{M+M-1} \end{array} \right]$$

上式中[]内的式子有奇数个计算元素，因此[]内的式子呈现出对称性。

现在考虑M≠N时的场景

若M>N时，2次移动平均的计算公式如下

$$\frac{1}{N} \left[\begin{array}{l} \frac{1}{M}(y_1 + y_2 + y_3 + \dots + y_N + \dots + y_M) \\ + \frac{1}{M}(y_2 + y_3 + \dots + y_N + \dots + y_M + y_{M+1}) \\ + \frac{1}{M}(y_3 + y_4 + \dots + y_N + \dots + y_M + y_{M+1} + y_{M+2}) \\ \dots \\ + \frac{1}{M}(y_N + y_{N+1} + y_{N+2} + \dots + y_{N+M-1}) \end{array} \right]$$

上式整理得

$$\frac{1}{NM} \left[\begin{array}{l} y_1 + y_2 + y_3 + \dots + y_N + \dots + y_M \\ + y_2 + y_3 + \dots + y_N + \dots + y_M + y_{M+1} \\ + y_3 + \dots + y_N + \dots + y_M + y_{M+1} + y_{M+2} \\ \dots \\ + y_N + \dots + y_M + y_{M+1} + y_{M+2} + \dots + y_{N+M-1} \end{array} \right]$$

同样的，仔细观察上式可以得到以下：

$$\frac{1}{NM} \left[\begin{array}{l} y_1 + 2y_2 + 3y_3 + \dots + (N-1)y_{N-1} \\ + Ny_N + Ny_{N+1} + \dots + Ny_M \\ + (M-1)y_{M+1} + (M-2)y_{M+2} + \dots + y_{N+M-1} \end{array} \right]$$

上式[]内的式子计算元素的个数奇数还是偶数的讨论：

[]内第一行和第三行计算元素个数是一样的，2者加起来必然是偶数个；

[]内第二行计算元素个数为从N(含)到M的个数，由于包含了N，那么总个数为奇数个。因为M-N+1为奇数。

这样一来，[]内计算元素总个数为偶数+奇数=奇数个，

因此[]内的式子呈现出对称性。

最后讨论M<N时的情况

$$\frac{1}{N} \left[\begin{array}{l} \frac{1}{M}(y_1 + y_2 + y_3 + \dots + y_M) \\ + \frac{1}{M}(y_2 + y_3 + \dots + y_M + y_{M+1}) \\ + \frac{1}{M}(y_3 + y_4 + \dots + y_M + y_{M+1} + y_{M+2}) \\ \dots \\ + \frac{1}{M}(y_M + y_{M+1} + y_{M+2} + \dots + y_{M+M-1}) \\ \dots \\ + \frac{1}{M}(y_N + y_{N+1} + y_{N+2} + \dots + y_{N+M-1}) \end{array} \right]$$

上式整理得：

$$\frac{1}{NM} \left[\begin{array}{l} y_1 + y_2 + y_3 + \dots + y_M \\ + y_2 + y_3 + \dots + y_M + y_{M+1} \\ + y_3 + y_4 + \dots + y_M + y_{M+1} + y_{M+2} \\ \dots \\ + y_M + y_{M+1} + y_{M+2} + \dots + y_{M+M-1} \\ + y_{M+1} + y_{M+2} + \dots + y_{M+M-1} + y_{M+M} \\ \dots \\ + y_N + y_{N+1} + y_{N+2} + \dots + y_{N+M-1} \end{array} \right]$$

同样的，仔细观察上式可以得到以下：

$$\frac{1}{NM} \left[\begin{array}{l} y_1 + 2y_2 + 3y_3 + \dots + (M-1)y_{M-1} \\ + My_M + My_{M+1} + \dots + My_N \\ + (M-1)y_{N+1} + (M-2)y_{N+2} + \dots + y_{N+M-1} \end{array} \right]$$

同前，对上式[]内的式子计算元素的个数奇数还是偶数的讨论：

[]内第一行和第三行计算元素个数是一样的，2者加起来必然是偶数个；

[]内第二行计算元素个数为从M(含)到N的个数，由于包含了M，那么总个数为奇数个。因为N-M+1为奇数。

这样一来，[]内计算元素总个数为偶数+奇数=奇数个，因此[]内的式子呈现出对称性。

所以，2次偶次期数的移动平均的结果可以保证对称性。

(二) 接下来，求出趋势T后，如何求周期S？

刚才经典的模型分解出来了趋势T，而加法模型为例，结构及变形是这样：

$$y_t = S_t + T_t + R_t$$

$$S_t + R_t = y_t - T_t$$

这样，我们用原始数据减去趋势T，就能得到带有随机扰动和季节因素的部分。

其中，一月份的“随机+季节”值，就等于1月份的原始数据-1月份的T值，这样就会有若干个1月份的结果；

二月份的“随机+季节”值，就等于2月份的原始数据-2月份的T值，这样就会有若干个2月份的结果；

以此类推，就得到了1-12月的所有“随机+季节”值。

然后，将所有1月份“随机+季节”求均值，2月份的“随机+季节”求均值……依次类推。即每个月的“随机+季节”求均值为

$$\frac{\sum_{i=1}^n (y_i - T_i)}{n}$$

得到了12个月的“随机+季节”均值求出来后，相加的结果和0做个对比，如果和0很接近，那么说明这个周期提取的比较成功。因为一个周期，周而复始，每个值围绕着均值0上下波动，因此理论上所有的波动求和=0。

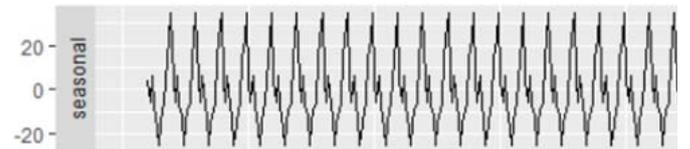
然而因为存在着随机扰动的影响，导致了12个月的“随机+季节”均值求和不等0，因此需要稍加处理。

经典模型的处理是这样，每个月修正后的季节因素= 每个月“随机+季节”均值-所有月份“随机+季节”均值的均值，也就是

$$\hat{S}_t = S_t - \frac{\sum_{i=1}^{12} (S_i + R_i)}{12}$$

本次澳大利亚的啤酒产量的例子，就是得到这样的结果，在R里面他用季节指数

```
$figure
[1] 3.643172 [2] -5.500204 [3] 6.215633 [4] -
7.319126
[5] -11.503226 [6] -25.05169 [7] -15.4917 [8] -
9.438772
[9] -
6.113986 [10] 13.536121 [11] 22.602360 [12] 34.421484
```



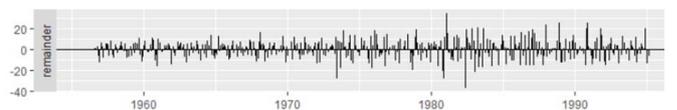
(三) 求出R

这个就一句话的事儿了：

$$y_t = S_t + T_t + R_t \text{ 求出了 } T_t \text{ 和 } S_t \text{ 的替代值 } \hat{S}_t$$

那么

$$R_t = y_t - T_t - \hat{S}_t$$



(四) 经典模型的问题

主要来自于求趋势时采用的移动平均。

因为移动平均计算时，原始数据的首尾若干数无法计

算，这样也就造成了趋势项T无法每一期都计算到；

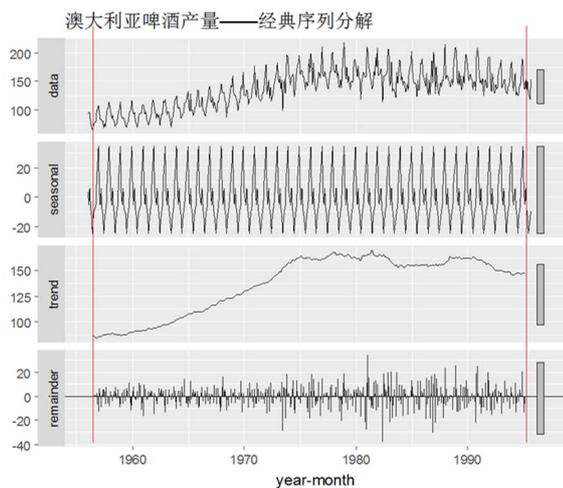
另外经典时间序列分解法假设季节项每年是重复的。对于很多序列来说，大环境稳定时这是合理的，但是对于更长的时间序列来说这还有待考量。例如，欧洲居民的用电量，在很多地方几十年前的时候，各个季节中冬季是用电高峰（用于供暖加热），但是现在夏季的用电需求最大（由于开空调）。经典时间序列分解法无法捕捉这类的季节项随时间变化而变化。

还有就是有一些突发事件对时间序列中一些时期的值影响大，但这个临时事件，不一定会按照规律复现。例如，美国的工会就经常组织和资方对抗，具体来说航空客运量可能会受到工会罢工的影响，使得罢工时期的客运量与往常十分不同。处理这类异常值，经典时间序列分解法通常不够稳健。

●经典的时间序列分解之外的其他的一些方法

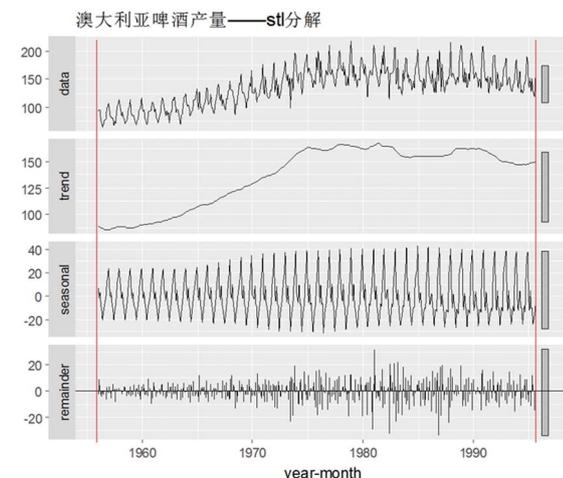
正因为经典时间序列的分解有这些问题，后来为了弥补，催生了其他的方法。

例如stl, X11, steat, X11-ARIMA, X12...等。



例如上图经典分解模型分解下，趋势项T和随机扰动项R的左边和右边，和原始数据相比，某些期次有缺失。

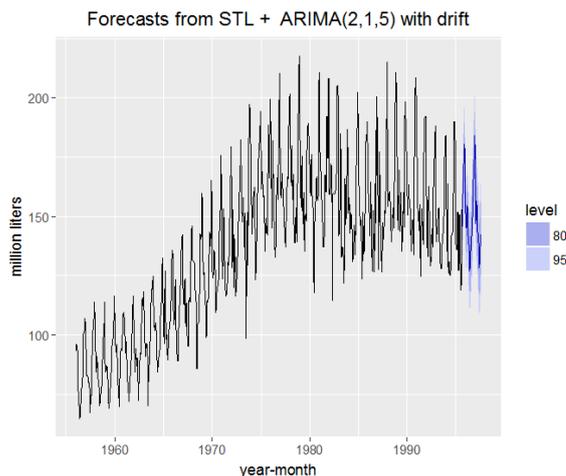
而stl分解法则通过一些算法修正解决了这个问题



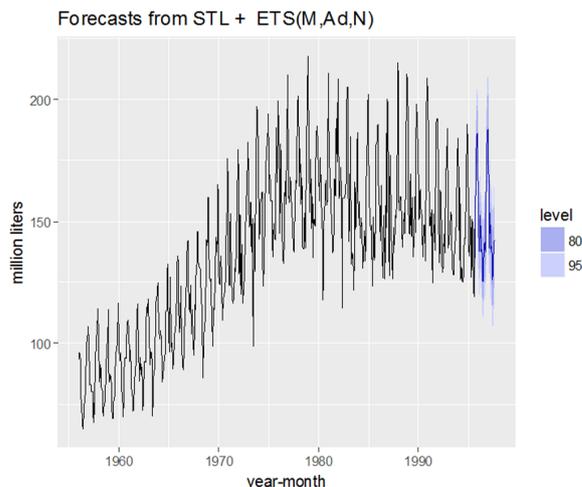
另外，在周期分解上，stl就打破了经典模型对周期分解固定不变的计算方式。因为从历史数据上看，周期影响造成的数据波动的振幅，是逐渐增加的，而stl也直接反映出来。

另外，这些新方法也提供了预测，而经典方法则难以预测。因为你首先难以找到一个合适的线性方程去作为整个数据的趋势，并预测趋势。

而新方法则解决了这个问题，例如stl 通过使用带arima算法进行未来的预测



又如stl 通过使用ETS算法进行预测



小结：今天回炉再充电关于时间序列分解的知识，主要是解决我一直以来关心的但没有人回答的问题：why，为什么要用时间序列分解来分析问题。并且将经典模型做了深入剖析，搞清楚了它的方法论计算步骤。

中国商业联合会数据分析专业委员会会员单位 ——数据分析师事务所会员

作者 / CPDA数据分析师 刘程浩 编辑 / 数据委员会处 李苗苗 日期 / 2022-05



一、数据委介绍

中国商业联合会数据分析专业委员会 (China Data Analysis Committee ,China General Chamber Of Commerce, 缩写CDAC) (以下称“数据委”)是中国商业联合会下属分支机构。

中国商业联合会(原中华人民共和国商业部、原中华人民共和国国家国内贸易部)成立于1994年,是由从事商品生产、商品流通业、生活服务业的企业事业单位,社会组织及个人自愿结成的全国性、行业性社会团体,是非营利性社会组织。中国商业联合会现设有工作机构13个,分支机构33个;直接会员4500多家,间接会员8万多家;接受政府委托,管理14个企事业单位,代管40个全国性专业协会、学会和研究会,主管32家国内外公开发行的报刊。中国商联分别是亚太零售商协会联盟、国际零售论坛的成员。

数据委成立于2008年4月,是经国务院国有资产监督管理委员会审核同意、中华人民共和国民政部正式批准和登记的数据分析行业组织。以数据分析师及数据分析师事务所等从事与数据分析行业相关的团体与个人自愿组成的

数据分析行业组织,是较早坚持发展专业型、靠专业引领市场的行业组织。

数据委的服务宗旨是搭建专业平台,促进国内外同行及业界内外的交流;制定行业标准,规范行业市场;加强内外协调,监督行业自律;服务于会员和行业的市场需求,维护会员的合法权益;加强行业之间的协作,维护行业的正当有序竞争;积极推动数据分析的普及和应用,培养专业人才,促进数据分析行业健康发展。

2015年中共中央办公厅、国务院办公厅印发《行业协会商会与行政机关脱钩总体方案》后,2016年数据委成为前三批脱钩协会之一,脱钩后正式由中国商业联合会直接监管。此政策的出台,更加坚定了数据委走专业型、靠专业引领市场的发展之路。

二、数据分析行业发展历程

我国的数据分析行业从无到有,直至今今天不断发展壮大,主要经历了以下几个阶段:

- 2003年原信息产业部确立CPDA数据分析师信息技术人才培养计划。

● 2004年我国首批“CPDA项目数据分析师”经过考核，在深圳诞生。

● 2005年第一家数据分析师事务所经工商局审批在陕西成立。此后，由数据分析师组建的事务所开始在北京、西安、深圳、成都等地纷纷诞生。数据分析专业事务所的出现，是我国数据分析行业的一个里程碑，我国数据分析行业开始兴起，从此进入不断发展的新时期。

● 2006年至2007年，数据分析师专业培训管理机构已培养出大量学员，并在多个省市组建了几十家专业的数据分析师事务所。这一时期，我国的数据分析行业已经成型，数据分析师和专业的事务所开始在数据分析所涉及的各个领域发挥着巨大的作用。

● 2008年经国务院国资委审核同意，国家民政部正式批准数据分析行业的行业组织——中国商业联合会数据分析专业委员会的成立。数据委的成立，标志着中国数据分析行业由此开始走向组织化、规范化、标准化的发展道路。

● 2009年数据分析行业的第一个行业标准在行业专家及全体事务所的支持下正式发布。

● 2010年在国家发改委及相关领导的支持和监督下，数据分析师事务所代表共同签署了行业自律宣言，并由数据委牵头启动了行业公益服务的大数据分析服务平台，开始面向社会开放公益性服务职能。

● 2013年正式启动“企业经营决策服务年”活动，以此推动我国大数据领域的发展。随着媒体对“大数据”的关注，政府、企业以及公众开始认知“大数据”，大数据元年正式到来。

● 2015年各地区“项目数据分析师事务所”正式更名为“数据分析师事务所”。

● 2016年“CPDA项目数据分析师”正式更名为“CPDA数据分析师”，由此CPDA品牌向着更广阔的市场空间进军。

● 2017年中国商业联合会数据分析专业委员会发布《中国大数据人才培养体系标准》旨在为大数据应用人员、企业提供一套具有实用价值、可衡量的评估体系。

● 2018年携手万宝盛华共同推进大数据人才战略合作，同期，数据分析智能大数据分析平台Datahoop2.0正式上线。

● 2019年CPDA数据分析师课程体系完成第八次改革，同期，CPDA考试正式委托第三方ATA组织实施，进一步提升数据分析行业教育考试的公正性、规范性和严谨性。

● 2020年疫情期间，数据分析师事务所助力企业数字化转型，同期，在国家大力主导大数据产业发展及大数据人才培养的背景下，经过对高校设立“数据科学与大数据技术”专业深入评估，同意此专业毕业生可直接报名参加CPDA数据分析师考试。同年9月，CPDA数据分析课程体系完成第九次课改。

● 2021年组织数据分析行业标准化制定工作，协同大数据领域相关企事业单位共同编制《数据分析行业服务参考文件》，为数据分析领域从业提供行之有效的指导意见。

三、数据分析师事务所介绍

数据分析师事务所（以下称“事务所”）是全新的第三方独立服务机构。是在工商局注册备案后成立，经数据委审批通过成为数据分析行业会员，并授予中国数据分析行业会员执业资质证书，接受数据委的监督和管理。事务所统一的服务标准规范，是数据分析行业走向规范和自律的中流砥柱，是促进数据分析行业健康发展的中坚力量。

事务所由专业的数据分析师人才组成，可为企事业单位提供与大数据相关的专业服务。目前，我国数据分析师事务所会员遍布在全国各省，服务范围涉及数据分析相关领域，随着社会数字化转型的需求，广大企业对数据价值的认可，数据分析师事务所必将有更广阔的发展空间。

四、事务所发展前景

目前百余家专业的数据分析师事务所遍布在各省市，业务基本围绕着数据的深度分析、业务场景构建、深层次的咨询，以帮助企业实现数据变现的能力。近两年随着政府和企业对大数据的重视，数据会越来越多，技术门槛会越来越低，但是大量的数据进行深层次的分析就成为企业竞争的核心、成为企业大数据变现的核心，这对数据分析师事务所来说是一个商业价值巨大的朝阳行业。

五、2022年申请事务所会员可享受服务政策

1、协助品牌宣传推广：依托数据委资源平台，通过广告推广、公关推广、网络推广等方式，帮助会员宣传企业，提升品牌影响力，展示品牌形象；

2、帮助提升大数据专业能力：提供不低于四次大数据相关技能培训学习；

3、帮助解决数据分析师人才不足问题：提供数据分析师招聘平台，以解决会员分析师应用人才短缺问题；

4、给予会员资质在中国数据分析行业官网上备案、查询：可通过数据委在线会员查询系统进行资质查询，可为广大客户提供自助搜索查询功能，同时有效杜绝非会员单位伪造数据委颁发会员执业资质承接业务现象；

5、优先报名参加数据委举办各种线上下活动：数据委举办的各类公益沙龙、创业指导等活动，会员单位可优先参加，并可从中挖掘商机不断提升自身发展能力；

6、提供Datahoop智能大数据分析平台高级版使用权限：Datahoop智能大数据分析平台高级版费用890元/年，数据分析师事务所会员在会籍有效期内可免费使用高级版权限；

7、搭建会员单位交流服务平台，加强会员单位之间沟通联系：通过数据委平台，建立会员单位（包括数据分析师事务所、行业会员）之间联系，促进各方合作机制，实现共赢；

8、享受会员专属执业教育培训：以网络远程课程+线上测评形式，内容以理论结合实操进行大数据相关内容培训；

六、事务所主要业务范围

事务所主要业务包括：数据分析咨询顾问服务、企业经营类数据分析服务、搭建大数据业务场景服务、综合解决方案服务、行业数据分析与解决方案服务等有关数据分析的业务。其中：

数据分析咨询顾问服务主要包括业务能力咨询、数据能力咨询、技术能力咨询、数据洞察能力咨询、解决方案咨询等咨询顾问服务等。

企业经营类数据分析服务包括财务分析、市场数据分析、客户数据分析、销售数据分析、质量数据分析、采购数据分析、人力资源数据分析等有关企业经营行为的数据分析等。

搭建大数据业务场景服务主要指模拟客户的业务场景，帮客户构建商业运营数据模型。

综合解决方案服务包括数据产品、数据平台建设等综合解决方案等。

七、申请事务所会员条件

(一) 事务所会员入会条件

- 1、机构名称中需包含“数据分析师事务所”字样；
- 2、申请者是数据分析师,成立事务所需要持证分析师或具备数据分析师水平的人员人数不少于3人，否则需不少于5人；
- 3、从事数据分析相关业务，有成熟的业务资源与一定的企业经营经验；
- 4、注册资金不低于100万；
- 5、办公场所不低于100平米。

(二) 事务所会员命名规则

X省/X市+特征名+数据分析师事务所+有限公司/有限责任公司/合伙人企业，如：海南智企数据分析师事务所有限公司。

(三) 事务所经营范围

须含有数据咨询、数据处理、数据分析或研究等相关业务。

八、事务所入会申请材料

- 1、《数据分析师事务所成立预备案及入会申请表》；
- 2、事务所运营方案；
- 3、法人身份证、学历证明复印件，简历，一寸白底照片一张；
- 4、数据分析师证书、身份证、学历证明复印件，简历，一寸照片1张；
- 5、数据分析师证书原件。

九、入会流程

- 1、申请者向数据委员会处递交事务所会员预备案申请

电子版相关资料进行审核；

2、申请材料审核通过后，会员处通知申请人交纳预备案评审费；

3、确认收到预备案评审费，会员处将申请材料上报秘书处审核；

4、秘书处审核通过后，会员处向申请人出具《预备案通知书》；

5、申请人收到《预备案通知书》后，向会员处提供相关纸质资料；

6、会员处安排专人实地考察(费用由事务所自行承担)；

7、会员处组织召开事务所评审会；

8、评审通过后,会员处通知申请机构交纳入会费用；

9、会员处向申请机构出具批复函、颁发《会员执业资质证书》。

十、事务所会员执业资质证书和授权铜牌示意图



十一、联系方式

数据委官网: www.chinacpda.org

会员处电话: 010—59000056-652、18101312526 (同微信)

电子邮箱: xiehui@chinacpda.org

联系人: 李老师

邮寄地址: 北京市朝阳区朝外大街乙6号朝外SOHO-C座-931室 会员处

数据分析师事务所创业指导“圆桌派”第一期Q&A

作者 / CPDA数据分析师 刘程浩 编辑 / 数据委员会处 李苗苗 日期 / 2022-03

3月17日，数据分析师事务所创业指导“圆桌派”第一期活动中，我们现场及在活动后，收到很多人咨询成立事务所的相关问题。

我们陆续把问题整理并请两位分享老师为大家解答欢迎大家和我们的事务所进行交流。

01 “数据分析的业务主要来源于哪些企业，或者企业的哪类部门？”

我们服务涉及网络安全、零售、医疗、航空、房地产、政府、电商等行业，为大量企业和政府提供了各类“咨询+技术”服务。不同企业找的部门不同，要看企业组织管理结构，一般数据部门或业务部门多一些，有的也有总经理助理。

02 “现在数据安全立法了，我们帮客户分析数据，需要哪些流程？”

数据分析流程方面与立法前后都一样，对技术分析来讲不会有影响，不过要对数据的来源合法与客户进行充分地沟通，数据分析师事务所以客户提供的数据，与业务相结合提供数据分析服务，同时结合分析师的实战分析经验进行综合分析。

03 “想问下老师们是怎么找到第一个客户的，且怎么说服第一个客户的？”

要给客户自己对于业务的数据分析理解和客户痛点分析。客户比较看重的是找谁解决问题分析问题更划算。获得客户渠道有很多种，每个人的第一个客户都不一样，每个分析师最好要有一技之长的分析能力，以专业度获得认同，对我们事务所而言，第一个客户是营销渠道引流，通过客户对比后，与我们进行商业合作的。

04 “爬取之外，具体数据是通过购买还是通过调查得到？”

提供数据不是数据分析事务所要做的事，数据建议与客户多沟通，共同协商解决，让客户来做决定以什么方式参与数据分析合作，数据分析师事务所的主要方向是做好分析工作，数量来源尽量与客户协调解决，避免风险。

05 “请问老师，初期投入都有什么？多少运作资金？”

初期具体投入多少资金，每个人的做事风格不一样，我建议量力而行，视自己可承受的经济范围投入资金运作。没有具体的数据参考建议，这里没有可比性，因人而异。

06 “多数企业是没有什么有价值的数据，都想着怎么能乙方提供更多有价值的数据和数据产品给自己用，乙方如何提供海量有价值的数据？”

希望明确数据分析师事务所的经营方向和定位，也就是经营企业的最基础的条件是在法律允许的范围内从事合法经营的行为。所以海量有价值的数据，要明确的是数据分析师事务所核心竞争力是助力客户分析有价值的决策，不是提供海量有

效数据给客户从事相关经营工作，涉及到数据隐私和法律合规性，要明确定位服务在可控的法律合规风险范围内。

07 “分析是我们的亮点，我目前接触的数据公司最多做到数据可视化。问下老师，数据分析和统计事务所有什么区别和共同点呢？”

核心区别在于数据分析的专业度上。数据分析师主要体现在对于数据价值的挖掘和分析能力、数据分析结论的准确度，以及能否对决策提供有力支持，为客户创造更多经济效益和数据价值。

08 “您们给客户做报表是用什么软件呢？”

EXCEL、SPSS、SAS、PYTHON、DATAHOOP等，没有一概而论的，不同业务对应不同软件。

09 “老师有没有服务项目的报告分享一下？”

这个我们和客户都签订了保密协议，涉及到客户的隐私无法对外分享，希望理解。

10 “如果客户想做数字化转型，但又不知道自己想要啥，要怎么办？”

可以询问客户为什么做数字化转型，业务瓶颈在哪里？痛点、痒点、兴奋点是什么，帮助客户梳理想法并可落地执行。只有深度了解客户问题与需求，才能帮客户提出有效的解决方案，被客户认可后，才有机会推进合作。

11 “数据分析师事务所里的分析师收入怎么样？”

数据分析师事务所分析师收入主要看数据分析经验与项目承接量，收入按照数据分析师级别不等，有一定的实际从业分析经验专业级别的数据分析师年薪50万以上。

12 “数据分析师是不是得依托某个专业领域？”

不需要依托专业领域，但是一定要有数据分析师的分析能力和与客户沟通、帮助客户解决问题的能力，当然最好是建议有一技之长，某个领域的数据专家更受欢迎些。

13 “请问老师，事务所是怎么拓展业务的，通过什么途径和方法发展客户？哪个行业相对比较好开发？哪类客户对事务所接受度比较高？”

事务所建议走合作模式，建立营销网络渠道，开展业务合作模式。开发行业和客户没有比较好开发和接受度比较高的说法，还是要优化事务所自身的数据分析实力，要学会与客户双向沟通，多用心去了解客户的真实需求和预算，签订合同后一定要全力以赴地认真做，建立长期稳定的合作关系，因为开发一个新客户成本远比留住老客户成本高很多。

14 “是否需要在做过这个行业后才能帮相应的企业做项目？比如说必须有银行行业经验才能接银行相关的项目吗？”

数据分析是团队合作的过程，每个数据分析师事务所为客户提供项目服务是要经过很多分析流程的，包括前期与客



中国商业联合会数据分析专业委员会
China Data Analysis Committee

数据分析师事务所 创业指导圆桌派 <第一期>



户做调研沟通、数据分析能力与数据挖掘价值判断，数据分析专业更多的体现在技术专业度，客户行业经验只是一个方面，可以定为加分项，但不是绝对项，所以具体要看每个项目决策者对数据分析师的选择。

15 “没有投资方支持，数据分析师事务所初期业务不多，要怎么活下来？”

创立数据分析师事务所前提就要把这个问题想清楚。极端情况下，或者在在业务不多的情况下该如何应对？一般要准备自有资金注入，不要有依靠“金主”的思维，这世界上没有免费的午餐，但从天上掉下的“陷阱”却很多，所以脚踏实地学会“自生长营收”，快速开拓客户。

16 “黄老师说的产品是数据软件或者平台吗？”

数据分析交付客户方式有很多种，比如数据软件、数据平台、数据分析报告、数据决策研究等等。具体要与客户多沟通，了解客户需要我们提供哪些具体可用的数据服务。

17 “可以去黄老师的事务所那边参观学习吗？”

欢迎来我所参观学习，交流经验。需要我所解决问题的，我们会提供力所能及的帮助。

18 “为客户企业做数据分析时，客户数据是连库吗？”

数据分析技术细节要视不同客户需求而决定，没有一概而论的。

19 “客户数据会不会进行脱敏处理？”

数据脱敏处理方式视不同客户需求而决定，不会有一概而论的处理方式。不管处不处理，我们都会跟客户进行严格的合作保密约定。

20 “我们的报告方案最终是不是还需要闭环，然后再复盘分析呢？”

这个没有标准的答案，我建议是与客户需求相对应，我们数据分析师事务所的方案是通过数据专家过会的。也就是报告方案后，数据专家们会在会议上共同商讨，确保每份报告方案的定稿分析专业度。

End

在大数据应用价值不断凸显的今天，数据分析师事务所作为一个全新的第三方服务机构，围绕数据的深度分析、业务场景构建，为国内企业数字化转型产生了巨大推力，在企业成长的各个阶段起到重要作用。数据委借创业指导“圆桌派”活动形式为大家带来创业分享，引导大家踊跃思考、交流热点问题。希望此系列活动后续能为大家带来更多工作帮助。



咨询入会：会员处
010-59000056转651、652

CPDA® 数据分析师
CERTIFIED PROJECTS DATA ANALYST. SINCE 2003

恒

心

”为学需刚与恒
不刚则堕落
不恒则退“



数据分析 · 因你而不凡!

www.chinacpda.com | www.cpda.cn
TEL. 400-050-6600