

贾西贝：我有创业和冒险的基因

文/本刊记者 杨淼

贾西贝，第十二批创业“千人计划”专家，深圳市华傲数据技术有限公司CEO。



说起大数据，人们想到的是庞大的数据量、瞬息变化的积累和盘根错节的内容，但这就是大数据的全部吗？Informatica高级副总裁Dennis Moore告诉我们：“大数据的核心问题不是数量大，而是质量高，而主数据管理确保数据的质量和真实性，企业通过主数据管理可以很好地洞悉业务数据中所隐藏的价值。”但是在世界范围内都存在着数据不正确、缺乏一致性和完整度等质量问题，更有需要证明“我妈是我妈”的数据孤岛问题。在此情况下，华傲数据在中国乃至世界数据质量基础软件创下首例，对数据管理具有跨越式的发展，填补中国在数据质量基础软件方面的空白，通过先进数据清洗技术推动中国大数据产业发展。为此，本刊特邀请华傲数据技术有限公司CEO贾西贝，探索这个被称为和百度、腾讯一起引领中国大数据发展的企业有何与众不同的基因。

大数据有创造价值的基因

今天提及大数据，更多想到的是风云诡谲的市场竞争和日新月异的技术突破，但是回溯2011年，大数据还是一个不为人知的概念。大数据一词起源于2004年，直至2006年大数据主流技术才开始起步，之后一直在技术领域内发展。但是2012年可以说是大数据发展中的一个关键节点，因为美国和联合国的重视使得大数据逐渐被大众所熟知，并且在国庆的时候广东省省委书记汪洋专门写了一封关于广东省发展大数据建议信，这标志着大数据正式被国内所熟知认可。

贾西贝在2011年的时候创立了华傲数据，可以说是中国大数据行业发展的见证者。他认为虽然大数据引入中国只有十年不到的时间，但是它所创造的价值却远远超越了任何想象：“在过去的几年，高科技行业出现了很多新的名词和概念，让人应接不暇，有些甚至昙花一现，但是大数据行业相较很多行业会显得更加平稳。这一点来说最明显的表现就是一批著名大学开设了数据科学与工程专业，有的还成立了大数据学院。要知道在电商、物联网和云计算崛起时，国内著名的大学并没有开设相关专业，更不用说是成立专门的学院。另外大数据不仅带来了一种新的学科，也带来了例如数据科学家、数据分析师等一批新兴职业，这是和很多新技术不同的地方。”可以说大数据正在脱离一种概念的束缚，而真正开始服务于人们的生活、改造现有的生产环境，技术和生活的边界也变得越来越模糊。在金融征信、精准营销以及健康大数据、工业大数据这四个典型大数据应用场景中，这样的变革更加明显。贾西贝介绍金融征信中通过电商、支付数据分析信用已经具有比较成熟的商业模式和技术支撑了，但是在这之外大数据服务商还在进行探索，比如利用社交数据进行信用分析；在健康领域基因数据、可穿戴设备产生的数据、健康档案等医疗健康数据的应用，正在让人们变得更健康；在工业4.0里面大数据是一个非常关键的环节，这在我国一些老工业基地中正在尝试创新，大数据结合例如3D打印一类的新型制造方式会体现出更多的价值。

2015年9月1日国家发布了《大数据行动纲要》，代表着大数据上升到国家政策层面，文件首次系统地提出了大数据概念以及规划大数据发展的路径。



大数据正在脱离一种概念的束缚，而真正开始服务于人们的生活、改造现有的生产环境，技术和生活的边界也变得越来越模糊。在金融征信、精准营销以及健康大数据、工业大数据这四个典型大数据应用场景中，这样的变革更加明显。



不同于2011年对于大数据的观望和怀疑态度，在政策加持和隐私观念转变的影响下大数据现在俨然一副生机盎然的景象，但是贾西贝认为认识大数据不单只看现在，更要关注的是未来。“其实大数据是一个历久弥坚的行业，就像一个老中医。阿里巴巴的蚂蚁金服就是靠阿里成立十几年的数据积累才能做成的，并不是一蹴而就。如果金融征信行业能够积累一百年的数据，那么就可以完整记录一个人一生的数据，就能够更好地通过大数据对一个人做健康分析，能够更好地判断一个人的消费偏好，也能够更好地分析一个人的信用；人的生命周期之外，一个行业、一个国家如果能够积累几百年数据，那就能科学指导行业发展和国家决策。”另外大数据虽然深入各行各业，但是在数据交易方面不是特别成熟，要发展到证券交易所这样的规模，贾西贝建议首先需要解决的是数据交易的合法性、数据估值的合理性、数据产权保护的有效性等问题，而这些涉及到立法、政策制定、观念转变、产业生态方方面面，对于数据交易来说还有很长的路要走。

华傲有数据的基因

2013年，《哈佛商业评论》一篇名为“谁来引领中国大数据的发展？”的报道中，华傲与百度、腾讯共同被评为“中国大数据产业的领军公司”。彼时百度与腾讯已经加入BAT三角巨头之中，然后华傲仅仅只是一个初创公司。是什么让《哈佛商业评论》这样青睐一家初创公司？对此贾西贝回答：“华傲是一家基因里就是数据的公司。”BAT的发展路径是从业务到数据，华傲则是从数据到数据，是一个纯粹的数据公司。他认为BAT三家并不是典型的纯粹大数据公司，但是他们在社交、搜索、电商领域利用大数据并且成为了享誉世界的公司，所以《哈佛商业评论》认为他们在大数据领域仍然是有很大的上升空间。“华傲进入这个排行榜，一方面代表《哈佛商业评论》对华傲专业技术的认可，另一方面代表大数据行业发展其实存在着不平衡的情况：有些领域很早就引起了注意，目前竞争非常激烈；而有的领域对技术的要求比较高，所以只能是专业的公司才能找到其中的价值。”

在大数据的产业链中，BAT所处数据产生与收集环节，这个领域多为强业务的巨头公司；在数据存储与管理环节，IBM等我们熟知的传统的数据软件公司早已重点布局；在最后的数据分析、应用领域小公司众多，百花齐放。而大数据中最薄弱的就是华傲涉及到的数据清理环节：“大数据时代业务系统或软件每天都会产生大量的‘排泄物’，我们在做的事情就是通过数据整合、清理、质量提升的工作将‘排泄物’转变为数据资源。”

最近几年里，华傲的数据清理业务已经在互联网、金融、电信和政府领域得到了很好的推广和市场化应用。华傲在深圳市做了四年的“织网工程”就是



一个很好的例子：“在我们的业务场景内，政府面临的碎片化数据问题最为棘手。因为政府信息系统多是纵向的业务系统、烟囱林立、格式不统一、模型不一致、实体不关联等导致一个人的信息没有办法在整个政府系统中流通，这才会出现‘证明我是我、我爸是我爸、我妈是我妈’的笑话。将政府系统中的成百上千的业务系统中的数据汇聚成为一个大的数据资源库，对新型智慧城市的建设意义非凡。中国建设智慧城市的城市就有400多个，但是他们都没有享受到大数据时代带来的便利，没有形成城市数据资源。”通过复制“织网工程”，可以在全国推行这样的模式，目前在沈阳已经开始在智慧城市中引入织网工程一样的大数据城市理念与实践，下一步将会在贵阳、海南、盐城等城市进行推广。

贾西贝在采访中一直认为每个企业都有一个基因，企业需要结合强项来发展业务。正是因为华傲有着数据的基因，所以无论是在大数据不为人所知的2011年还是到大数据炙手可热的2014年、2015年，华傲都不会改变企业技术

的方向。在技术之外，华傲紧抓市场的变化，使得商业模式随着大数据产业的发展而发生着改变：“在创业初期，我们首先尝试的是做软件产品公司，在发现了软件产品公司在中国缺乏一个好的土壤环境之后，我们转变战略走DT服务道路。发展至今，我们同样在尝试数据运营、大数据风控等不仅需要创新技术、还结合了创新商业模式的业务。这其实就是华傲发展的两条线：在领域技术上一直处于不断深耕求精的过程；在商业模式上是一个不断探索提高的过程。”

我有不安分的基因

1975年出生的贾西贝刚进入“不惑”之年，就已在国际顶级和一流数据库会议及学术期刊刊发表10余篇论文，在大数据领域申报了两项美国专利）、87项中国发明专利、40项国际专利（PCT），曾获IEEE国际数据工程大会（ICDE）最佳论文奖。即使在同龄人当中，也算是比较“牛气”的人物了。

喜爱登山、跑步、自行车、划船之类冒险性活动的贾西贝在采访中自我剖

析为“有一颗不安分的心”：“我认为创业者都是天生喜爱冒险的人，无论我是选择出国还是创业，很大一部分原因都在于我个人的经历。”在北大结束学习生涯后的两年后，贾西贝才选择了去英国攻读爱丁堡大学博士学位，他回忆起这段学习经验，更多的是感激：“英国很重视基础研究和国际交流，我在国外的经历接触了很多前沿技术，锻炼了自己的创新能力和独立思考能力，这些都为创业打下很好的基础。从北大到英国，我觉得首先是培养了我对数据的热爱，我认为创业首先需要的是有热爱这个领域的情怀，这样才能让创业者锲而不舍地坚持下来。另外从事技术行业需要专业的训练，在英国的经历也让自己的专业能力得到了很大程度的提升。而我内心的不安分来自于我的经历，我在学校和企业都有经验，这使我看到了很多市场机会，因此我就选择了创业而非留在学术界。”

2011年贾西贝选择落户深圳，贾西贝说这是因为他在国外的经验使得他认识到了内心对国内有情感和文化的认同。“创业不仅是需要市场和技术，还有一个很重要的因素就是认同。我的人生经历更多是在国内，对国内市场和用户都有更为贴切的理解。选择落户深圳也是基于了解的基础：深圳有很好的创业创新土壤，深圳市委市政府很重视海外高层次人才，除去千人计划以外，还有本地的孔雀计划、珠江人才计划等一系列人才计划，我们是慕名来的深圳。”

今天的中国，吸引了更多像贾西贝这样的优秀人才，也为他们提供了广阔的发展空间。对此贾西贝深有感触，他认为今天中国处在快速上升的阶段，能够将自己的经历和想法与社会的需求相结合，是自豪感和归属感的重要来源。