

# 贵阳市大数据人才一选一专题培训学习体会

## ——大数据时代的高校建设思考

贵阳职业技术学院 刘海波

### 一、专题培训概况

从7月4日至8月2日，我作为贵阳职业技术学院选派的代表，参加了贵阳市组织的为期21天的大数据一选一专题培训班。培训班规格之高史无前例：围绕大数据专题从国家职能机构、一流学府、大数据企业等邀请了一大批专家学者来筑授课，并在最后一周安排了在乌当区大数据产业基地进行为期一周的实习。通过本次专题培训，我认为主要从以下几个方面有所收获和提高：

1. 能够从国家战略的高度理解贵州省、贵阳市发展大数据产业的理论依据和方针政策。就为什么要发展大数据，如何发展大数据统一了思想，明白了为何需要以及如何进行全市各职能部门协同的大数据产业发展实践；
2. 通过大数据各个领域专家的授课讲解，掌握了大数据概念的核心和外延，了解了大数据的前世今生以及未来的发展方向；掌握了大数据产业的业态分布、生态系统、支撑技术以及应用领域；通过老师对实际应用案例的分析讲解，理解了大数据如何从根本上改变了传统产业的发展规律和传统的国家治理方法，堪称真正意义上的“第四次工业革命”；
3. 从贵阳市相关职能部门领导以及贵州大学诸多专家教授的报告中，了解了贵州省以及贵阳市发展大数据的战略和方针政策；了解了大数据战略在贵州省及贵阳市实施落地的具体情况，对当前发展大数据产业的方向以及存在的问题有了更深刻的认识；
4. 通过在乌当区大数据智慧产业园的实习，实地观摩了乌当区利用大数据技术提升市政管理（乌当区政府）、制造（中航工业）、物流与电子商务（康心药业）信息化智能化程度的实践，对大数据如何变革传统产业有了更深刻的体验。

下面，我想结合供职单位和自身的工作职责，就如何运用大数据提升高职高专信息化建设水平，从两个方面谈一下我的思考。

## 二、 如何运用大数据思维进行高校信息化建设

在各位专家学者授课的过程中，无一例外都会提到贵州贵阳发展大数据的优势和劣势，而对于贵州省和贵阳市发展大数据的先天不足，总结各位专家的观点，集中在“工业化、信息化程度较差以及人才匮乏”这三点上。徐昊副市长在他的报告中甚至称贵阳发展大数据是“无中生有”。

那么贵阳是不是真的可以在没有任何基础或者基础薄弱的情况下“无中生有”地发展好大数据产业呢？我的理解是，发展大数据并不意味着我们能跳过区域经济发展所必需的信息化建设的某几个特定阶段，直接进入大数据时代，而是说我们要借助发展大数据成为国家战略的历史机遇，集中资源加速布局，补强工业化，信息化基础设施薄弱的短板，实现“弯道取直”（亦即少走弯路）而不是“弯道超车”（意味着跨越式发展，超越一线城市）。那么具体如何“弯道取直”呢？

在中关村产业联盟秘书长赵国栋博士的授课中，他总结互联网时代区域经济发展的模式是：一张蓝图、两个基石、三大引擎、五个抓手、N个工程。即是以习总书记的“中国梦”中华民族伟大复兴战略为蓝图，以传统产业升级和新兴产业聚集为两个基石，以制度变革、结构优化及要素升级为三大引擎，以智库、传播、资本、联盟、园区为五大抓手，通过“6789”四个系列的N个工程的实施，实现以大数据驱动的区域经济发展。<sup>i</sup>

具体到高校的信息化建设，我认为可以从以下几点借助大数据发展战略实现“弯道取直”：

1. 大数据促进经济发展的本质还是在于对各行业经营效率的提升和交易成本的下降，而这是通过跨行业、跨地域地使用数据并发掘其中的价值来实现的，故对于高校的信息化建设而言，首先应该做的是形成大数据思维，跳出高校自身的圈子，在制定信息化发展战略、实施信息化建设的时候都要明晰传统路径与大数据路径的区别所在，运用大数据思维来考量具体问题，从战略层面与国家的大数据发展战略对齐，从而借助大数据的东风事半功倍；

2. 利用国家发展大数据产业的历史机遇，按照云计算平台的架构统筹规划高校的 IT 基础设施和应用，借助发展大数据的机遇消灭“信息孤岛”并盘活现有信息化设备及软件资产，使高校的信息化的整体性和信息化程度同步得到提升；
3. 转变高校信息化部门或网络部门的职能观念，按照大数据思维“数据即资产”的观念，逐步将信息化或网络部门提升为能够产生数据资产的“盈利”部门，通过学校形成的数据与其它职能部门和企事业单位进行的数据交换，利用数据融合产生的价值提升生产力，从而提高高校的运营管理水平。

### 三、 如何培养大数据技术人才

高校除了自身需要按照大数据思维进行信息化建设之外，还担负着培养大数据人才的任务。这里就如何进行大数据人才（主要是指技术型人才）的培养谈一下我的思考：

#### 3.1. 大数据是内容，云计算是形式

大数据是一个比较宽泛的概念，大数据之所以大，并不仅仅是因为数据的体量大，主要是大数据的 5V<sup>ii</sup>特性使得常规的数据存储、预处理、交换、分析及可视化手段无法适用，因此通常需要使用云计算（更学术化的说法是分布式计算），包括云存储的手段来处理大数据。而大数据的来源，除了传统的采集手段，还有现在方兴未艾的物联网技术，通过海量传感器采集各种物理量形成数据。与这种数据采集形式相对应的计算方式，被一些行业研究者称为雾计算（Dust Computing），其命名来源于将微小的传感器称为智能微尘（Smart Dust）<sup>iii</sup>的说法。

因此，支撑大数据发展所需要的技术型人才总体上可分为两大类，即大数据人才和云计算/雾计算人才。

大数据人才主要包括以下几类：

1. 数据科学家，主要指运用包括应用数学，统计，模式识别，机器学习，数据可视化，数据仓库，以及高性能计算等在内的知识，对数据进行价值发现的一类人才；

2. 算法工程师，能够运用计算机理论和信息化手段将大数据分析处理所需的理论和计算模型转化为计算机算法并进行实施、验证及优化的应用型人才；
3. （软件）系统架构师及软件工程师。（软件）系统架构师是软件工程师当中的一种，但需要特别指出来。原因是处理大数据所依赖的技术是大规模分布式的，这就要求需要有专门的计算机科学家或软件工程师从事系统架构的研究和实现。现有的成熟的计算平台或架构往往都不能很好的适应大数据计算的需求。这类人才的任务是为大数据寻找合适的分布式计算解决方案，而软件工程师则更偏重于运用编程技能实现这些解决方案。

云计算/雾计算人才主要包括以下几类：

1. 系统架构师及软件工程师。这里和大数据人才出现了部分的重叠，原因是系统架构师或软件工程师是一种复合型人才，他们既需要对计算大数据所使用的方法和架构非常熟悉，又需要对分布式计算平台的架构和开发、部署及运行调优非常熟悉；
2. 后端开发工程师。分布式计算的发展，对互联网应用后端的软件栈架构及开发技术提出了全新的挑战，也促进了后端架构和开发技术的巨大变革，因此，采用分布式计算的互联网应用后端需要大量的具备分布式后端架构知识及开发技能的软件开发工程师；
3. 运维工程师。分布式计算同样为大型互联网应用后端的运维带来了巨大的挑战，对运维工程师这一历史悠久的传统职业门类提出了全新的要求，主要针对分布式计算架构通常采用的大规模服务器集群技术、高可用技术、传统虚拟化技术（Xen, KVM 等，相对于 LXC, Docker 等轻量级虚拟化而言）、容器技术（LXC, Docker 等），持续集成/持续开发（CI/CD）等技术对运维工程师提出了挑战；
4. 网络工程师。云计算的形式决定了现代分布式计算中心的网络架构和传统的计算中心是迥然不同的，这就需要网络工程师这个传统职业适应大数据/云计算时代对职业技能的新挑战，不但对计算中心的网络工程了如指掌，还需要对网络所支撑的云计算架构非常熟悉，能够协同架构师、运维工程师、软件开发工程师共同保障云计算架构的实施和运作。
5. 物联网应用相关人才。由于物联网是一个相对新兴的领域，目前还没有统一标准的相关职业门类规范，所以这类人才大致包括从事物联网构建的网络工程师、

从事传感器及嵌入式系统开发的软/硬件工程师以及从事物联网应用开发、部署、运维的软件开发工程师及运维工程师等。

### 3.2. 是否需要覆盖大数据产业的全部三种业态？

大数据人才的培养，应该严格遵照大数据产业的发展规划来进行匹配，人才培养方案应该有的放矢。那大数据人才的培养，是否需要覆盖大数据产业的三种业态，即核心业态、关联业态及衍生业态呢？我认为答案是否定的，即应该避免专业设置和专业开设方向的“大而全”。例如，大数据产业的关联业态中有很多产业仅仅是对大数据技术提供必要的技术和软硬件基础设施的支撑，这些专业方向本来已经有行之有效使用经年的人才培养体系，可以从战略层面纳入到大数据人才的培养规划中，但不适宜从具体的培养方案的角度，再针对大数据技术去进行修改。

### 3.3. 做好分工协同

从前面的论述我们可以看到，大数据人才实际上包括了计算机科学、电子信息、通信技术及软件工程等多个相关学科领域的专业技能人才。而这些人才的培养对人才的基础要求差别是非常巨大的。例如，数据科学家及算法工程师对数学基础有非常高的要求，培养周期长，成材率低，通常只有基础学科实力雄厚的综合性大学有能力开设相应的专业及相应分支学科的博士点进行系统性的培养；而像系统架构师这类人才，则无法完全在学校完成培养，它的职业特性决定了这类人才只能通过大量的实际生产实践得到培养。

因此，对大数据人才的培养，从规划的角度来看，各类院校应该做好分工协同，统一筹划。基础学科实力较强的本科及以上院校应该成为数据科学、人工智能、高性能计算等领域人才培养的主力军。而专科类院校应该更多地从大数据产业落地所需要的工程型、技术型人才着手，围绕大数据企业的需求来定制培养方案，形成服务大数据核心业态的人才培养格局。

### 3.4. 拥抱开源技术及社区

开源软件技术及社区经过几十年的发展，借助古老而神圣的 C 语言及其开发环境，以及源远流长，桃李天下的 UNIX 操作系统，成为当今大数据/云计算时代

当之无愧的中流砥柱，在开源的沃土上，GNU/Linux、Android、Apache Hadoop、Spark、Node、Docker 等改变云计算面貌的革命性技术枝繁叶茂。开源文化“代言人”Eric Raymond 将开源软件运动形容成“大集市”，并且与传统的“几代人前赴后继按照宏伟蓝图修筑大教堂”的传统“闭源”软件开发商业模式相比也丝毫不落下风。<sup>iv</sup>开源软件运动是信息科学史上的奇迹，拥抱开源为我们“后发赶超”提供了千载难逢的历史契机，抓住这个机会，我们便能够突破国外“软件航母”群的封锁，实现大数据各个业态核心知识产权自主发展的目标。

## 四、 总结

大数据是人类文明史上的第四次“浪潮”，注定要全面改变人类社会的风貌。在这个历史性的时期，能够参与到贵州省贵阳市发展大数据产业的宏观战略中来，我感到十分幸运，也充满期待。我希望借助此次培训，能够更好地与全省、全市的大数据建设者们统一思想，明确目标，以点带面，推动自己所在单位的大数据建设，为贵阳市的全面发展贡献一份力量。

---

1) <sup>i</sup> 赵国栋，大数据时代区域经济发展之路，2016

2) <sup>ii</sup> Gil Press, [A Very Short History of Big Data](#), 2013 解释了 Big Data “5V”的起源

3) <sup>iii</sup> K Pister, Smart Dust: BAA97-43 Proposal Abstract, 1997 “Smart Dust”来源

4) <sup>iv</sup> Eric Raymond, 大教堂与集市, O'REILLY 机械工业出版社, 2014