Constructing a Course-Planning Framework for Talent Development of Big Data Analysts in Taiwan

Hsu, Shun-Fa, Lee, Yi-In, and Hsu, Ching-Tzu

Abstract— Big data has been a rapid-growing technological field in recent years, and also is the foundation of "Industry 4.0" among Germany, the paragon of manufacturing industries. In addition, omni-channel, which is formed by two main factors of O2O model in commerce and service sector- online electronic commerce stores and offline retailers, is getting attention. In the trend of growing amount of data, fastening updates, and varying sources, the demand of big data analysts has been dramatically increased. No matter mathematical/ statistical talents, information technology talents, or people with applied abilities, all play important roles in big data groups.
At present, the demand of big data analysts is increasing in Taiwan; therefore, big data related training courses appear as training institutes try to meet the trend, and some programs or degree programs are established in universities as well. But how and what the courses and programs should go on does not have a standard as the competency indicators for big data analysts have not been constructed yet, still waiting for further investigation. Based on the current situation, this research makes inquiries into training courses and degree programs of big data analysts, and constructs a course-planning framework for talent development of big data analysts in Taiwan.

Index Terms—Big Data Analyst, Course-Planning Framework, Training Courses, Talent Development

I. INTRODUCTION

As time goes by, social, cloud, mobile, and big data become four important factors for forcing the industrial development of technology. As the world becomes more digitized, the accumulated amount of data from internet rapidly expands. Lots of enterprises and experts find out customer behaviors and consuming habits analyzed from numerous data could be important indicators for gaining market share and acquiring new customers. Big data, derived from social, cloud and mobile, can create enormous commercial value by data capturing, mining, and analyzing. Thus, more and more enterprises invent on big data analysis, the former "trash" is now "treasure". Big data can interpret what happened before, and also can predict what customers will do while enough amount of data. In this era of digital convergence and "multi-screens and a cloud", one key to success in entrepreneurship is to find commercial potential by big data analysis.

According to International Data Corporation (IDC), the big data technology and services market will grow to 41.5 billion US dollars through 2018.

But in the talent market, there are still some problems remaining, like McKinsey & Company estimated at 2011 in their report "Big Data: The Next Frontier for Innovation, Competition, and Productivity" that by 2018 "the United States alone faces a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts to analyze big data and make decisions based on their findings." [1] Canada's Big Data Consortium also estimated that "Canada's Big Data Talent Gap is between 10,500 and 19,000 professionals with deep data and analytical skills, such as those required for roles like Chief Data Officer, Data Scientist, and Data Solutions Architect" in their 2015 report. They further estimated the gap for professionals with solid data and analytical literacy to make better decisions at 150,000, such as the Business Manager and Business Analyst roles. The consortium also pointed out the needs to improve labor market clarity. Due to the absence of common professional definitions and career pathways, they found that employers struggle to clearly express their big data and analytics talent needs, and that prospective talent finds it difficult to assess their suitability, interest, and candidacy in the field. To close the big data talent gap, they believed that labor market clarity between prospective employers and prospective talent is essential. [2]

At present, the demand for talents of big data analytics is increasing in Taiwan; therefore, big data related training courses appear as training institutes for enterprises try to meet the trend, and some programs or degree programs are established in universities as well. But how and what the courses and programs should go on does not have a standard as the competency indicators for big data analysts have not been constructed yet, still waiting for further investigation. Based on the current situation, this research conducts an investigation into training courses and degree programs of big data analysts, and constructs a course-planning framework for talent development of big data analysts in Taiwan.

Hsu, Shun-Fa is with the Chung-Chou University of Science and Technology(CCUT), Chang-Hua County, Taiwan(R.O.C.) (e-mail: alpha168@ gm.ccut.edu.tw).
Lee, Yi-in is with Chung-Chou University of Science and Technology(CCUT), Chang-Hua County, Taiwan(R.O.C.) (e-mail: yiin@ gm.ccut.edu.tw). Corresponding Author.
Hsu, Ching-Tzu is with National Taiwan University, Taipei City, Taiwan (R.O.C.) (e-mail: b01208008@ntu.edu.tw).

II. BIG-DATA ANALYST

Big data is relevant for all components of our society. Industry is using big data for shifting business intelligence from reporting and decision support to prediction and next-move decisions. This use of big data emphasizes that big data is critical for obtaining actionable knowledge.[3] Governments are also interested in using big data and predictive analytics to improve decision making and transparency, to engage citizens in public affairs, to improve national security.[4]

Big data analytics has four stages: data capture, data process, data analysis and value added. In the whole process, one should be capable of handling data, integrating information technology (IT) and big data, and offering value-added strategies.
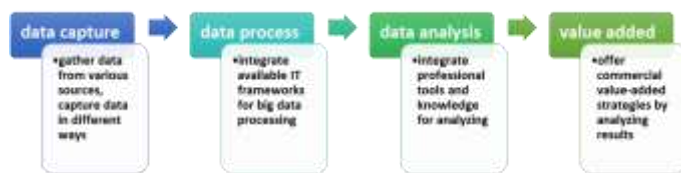

Fig. 1 Four stages for big-data analytics

According to the job specification of big data analysts from the Ministry of Economic Affairs, R.O.C, big data analysts will have to use the knowledge of commercial analyzing, industry consulting, communicating, and information technology at once, and know how to acquire the most important answer from enormous unstructured information. Big data analysts should have expertise in many aspects, including business studies, statistics, information management, logistics, etc. to form the correct and testable hypothesis from the original question. Besides, in order to use the most efficient computer techniques for managing the huge amount of data, they have to possess enough knowledge of computer and programming. At last, big data analysts have to interpret the results with nontechnical languages, they need to have not only the ability of analyzing but also the ability of interpretation.

Talents in big data analytics are mainly data scientists, data analysts, data engineers, and software engineers. With regard to the employment agency, many kinds of companies, such as retailers, medical companies, telecommunication companies, financial companies, entertainment companies, have job openings for analysts and database managers with job descriptions about dividing customer groups and offering customized services and products by generalizing useful information from big data. Online-learning website "Udacity" also offers "Ultimate Skills Checklist for Your First Data Analyst Job" for learners to easily arrange their training courses.

The official employment agency website of Taiwan "Taiwan-Jobs" provides near 2,000 job openings related to big data, for instance, Trade-Van company recruits 50 people for data analysis and database management. In the job description of big data analysts, big data analysts should possess the ability of industrial knowledge, statistical analysis,

computer science, database analysis and integration regarding the complexity of works.

Fig. 2 The job description of big data talents

III. ANALYSIS OF TRAINING COURSES AND CURRICULUMS

As the need of big data talents rising sharply, more and more training institutes start to offer courses related to big data, table 1 collects the courses. Also, these courses can be divided into three types:

1. Introduction to Big Data Analytics: Introduce the basic knowledge of big data analytics and practical scenarios.

2. Instruction to Big Data Analytics Software: Mostly R and Hadoop.

3. Text Mining: courses specifically for non-numeric data.

TABLE I:
COURSES IN TRAINING INSTITUTES

Institutes	Courses
Industrial Technology Research Institute	1. Implementations of big data analysis 2. Implementations of text mining and data analyzing 3. Data mining and Practical applications with R II
Institute for Information Industry	1. Implementations of text mining 2. Financial data analysis with R 3. Big data analysis using Spark 4. Big data analysis using RHadoop I 5. Implementations of data mining with Python 6. Data visualization with R 7. Practical process and analysis of big data

	8. Implementations of R 9. Data mining with R 10. Practice of distributed system- Elasticsearch 11. Web data capture with R 12. Big data analysis using RHadoop II	
Etu	1. Hadoop of software developer 2. R and Hadoop of data analysts 3. Software developer approaching Pig 4. Software developer approaching Hive 5. Software developer approaching NoSQL-HBase	
DSP	1. A2: Advanced data analyzing 2. E1: Basic data engineering 3. Data ETL & Analysis with Python 4. Analytical thinking team training 5. Data Engineer training program 6. Data thinking team training	
Gjun information	Practical analysis and applications of big data	
SAS	SAS Enterprise Guide	
Agilearning .IO	R Crawler 101	

From the angle of higher education, more and more universities and colleges open programs related to big data analytics to cultivate talents for industries, mostly in the way of cross-departmental collaboration.

For instance, National Chiao Tung University combines the faculties of institute of statistics and department of computer science, offers "NCTU Big Data Analysis Credits Program", and points out that a successful big data analyst, aside from statistics, should (1) have expertise in some science fields to form a firm and testable hypothesis for the questions, (2) have enough knowledge of computer and programming to manage the enormous amount of data in the most efficient processing techniques, (3) interpret the results with nontechnical languages to form useful communication. Courses are divided into "Prerequisite", "Basic Core Courses" and "Advanced Select Courses". Prerequisites include the experience of programming, like C, C++, Java, Python, or R, and the course of "Fundamental Statistics." Other programs in universities and colleges are listed in table 2.

TABLE II:
PROGRAMS AND BASIC CORE COURSES IN UNIVERSITIES

Units in Universities	Programs	Basic Core Courses
Institute of Statistics, National Chiao Tung University Department of Computer Science, National Chiao	Big Data Analysis Credits Program	Data Mining Statistical Methods Introduction to Data Science and Software Practice Big Data Analytics Techniques and Applications Science Project
Department of Management Information Systems, National Chengchi University	Big Data Analysis Credits Program	Introduction to Big Data Analytics Statistical Analysis Regression Analysis I Database Management Systems Management Information Systems
		Business Analytics with SAS/R
Department of Information Management, National Sun Yat-sen University	Big Data Analysis Credits Program	Introduction to Big Data Analysis Statistics Database Management Data Structure
Institute of Statistics, National Tsing Hua University	Data Science Program	Probability Statistics Programming Data Structure Statistical Data Analysis Database Management Systems Introduction to Big Data Analytics
Department of Statistics, National Cheng Kung University (2015)	Big Data Analysis Credits Program	Regression Analysis Multivariate Analysis Programming Data Structure
Department of Computer Science and Information Engineering, National Taipei University	The Program of Big Data Mining	Computer Programming Data Mining Linear Algebra Regression Analysis Big Data Analysis Probability/ Engineering Statistics/ Statistics
Graduate Institute of Statistics, National Central University	Big Data Analysis Credits Program	Independent Study of big data
National Chung Cheng University Ministry of Finance, R.O.C.(2016)	(Scheduled) Big Data Analysis Credits Program	-
National Taitung University	Interdisciplinary Program of Green and Information Technology	Programming Object-Oriented Programming Language Introduction to Green Technology Introduction to Nanoscience Data Structure Energy Management
Department of Computer Science and Information Engineering	The Program of Big Data	Big Data Analysis Data Mining Machine Learning Internet Security/ Cloud Computing Security Statistics
School of Big Data Management, Soochow University (2015)	Big Data Analysis Credits Program	Introduction to Data Science and Big Data Introduction to Statistics Data Mining with Applications on Business Applied Statistic
College of Business, Chung Yuan Christian University	Bachelor Program in Commercial Big Data	Workshop in Big Data Analytics Principles of Database System Applied Statistics and Analytics

		Business Models and Value Creation Big Data Platform Deployment Big Data Management and Applications Practice of Big Data Analytics Applied Machine Learning Business Intelligence and Data Analytics
Department of Transportation Technology and Logistics Management, Chung Hua University HP Inc. Genesis Technology,Inc.	Program for Big Data Analysis on Transportation	Introduction to Big Data Analysis Statistics (Big Data Analysis and Software Application, Core Techniques of Big Data)
Department of Industrial and Business Management, Chang Gung University HP Inc. VSSeurity, Inc.	Program for Big Data and Applications on Business	Introduction to Big Data Database Management Special Topics on Big Data Applications Big Data System/Big Data Analysis Techniques/ Applied Statistics
College of Management, Fu Jen Catholic University	Big Data Industrial Intelligence Program	Data Mining Applications for Statistical software/ Biostatistics Database Systems/ Business Knowledge/ Health Industry
College of Engineering, Tunghai University	Big Data Analysis Credits Program	Special Topics on Big Data Applications Data Mining/ Internet Applications/ Multivariate Statistical Analysis
Department of Information Management, Chinese Culture University SAS Institute Inc.	Big Data Analysis Credits Program	Statistics Big Data Analysis Industrial Intelligence Data Mining
Lunghwa University of Science and Technology	Big Data Analysis Credits Program	Applied Statistics I, II Big Data Analysis Introduction to Industry 4.0
Department of Statistics, Feng Chia University SAS Institute Inc.	Finance Analytics Credited Program	Financial Management Time Series Analysis Financial Engineering Statistical Computing Statistical Data Analysis Using Computer Software Data Mining

From the aspect of industrial talent training, as the need of big data analysis talents increasing, more and more training institutes offer courses, for example, "Courses of Cultivating Big Data Analysts" formed by institute for information industry(III) is in total 670 hours, and mainly concerns about Java, Hadoop, and business intelligence. It includes 162 hours of programming techniques, 51 hours of database, 78 hours of

techniques for analyzing business intelligence, 155 hours of big data analyzing, 224 hours for applications in enterprises' intelligence systems. In addition, III offers "Holiday Courses for Big Data Analysts" in total of 56 hours for cultivating the abilities of data processing, data analyzing, and data visualization. The content of courses include Basic R Language, MySQL, Data Mining, and Database Designing. Other training courses are mainly in three types:

1. Introduction to big data analysis: introduce essential concepts of big data analysis and applications.

2. Big data analytics software: mostly R and Hadoop.

3. Text mining and data visualization: mainly for non-numeric data.

Apart from training institutes, universities and research organizations also establish big data research centers to solve practical problems in industries by academy-industry collaboration. Table 3 lists big data research center in universities and research organizations.

TABLE III:
THE LIST OF BIG DATA RESEARCH CENTER IN UNIVERSITIES AND RESEARCH ORGANIZATIONS

Organizations	Name of Research Center
Department of Statistics and Information Science, Fu Jen Catholic University	Big Data Research Center
National Chiao Tung University	Big Data Research Center
Yuan Ze University	Innovation Center for Big Data and Digital Convergence
National Kaohsiung First University of Science and Technology	Big Data Research Center
Taipei Medical University	Big Data Research Center
National Sun Yat-sen University	The Multidisciplinary and Data Science Research Center
National Taitung University	Industrial IoT and Big Data Research Center
National Chengchi University	IoT and Big Data Research Center
Asia University	Big Data Research Center
Chung Chou University of Science and Technology	Intelligent Cloud-computing & Big-data Center
Chung Hua University	Big Data Research Center
Industrial Technology Research Institute	Computational Intelligence Technology Center

IV. THE COURSE-PLANNING FRAMEWORK

As the needs of big data analysts raising rapidly, more and more people would like to take part in this fast-growing industry. Present big data analysts are mostly from people originally had the abilities of statistical analysis or information system, and then improved other capabilities. Besides, according to "Talents of Big Data Analysis and Applications" in the program "Course Plans for Cultivating Talents in Productivity 4.0", which is generated by Executive Yuan, R.O.C., Ministry of Education, R.O.C. suggests the future course plans should include (1)the introduction to Business 4.0, (2)the knowledge of business, (3)techniques of data analyzing

and abilities of applications. Among these suggestions, course content and targets of courses for techniques of data analyzing and abilities of applications are listed in table 4:

TABLE IV:
COURSE CONTENT AND TARGETS OF TECHNIQUES OF DATA ANALYZING AND ABILITIES OF APPLICATION COURSES

Course content	Cultivating abilities	Course targets
Data processing and languages	Data preprocessing: data type, importing, database links, missing values processing Data management: data filtering, clipping, text managing Reports: charts drawing, reports generating	Knowledge: introduce data preprocessing, missing values processing, statistical graphs, and R languages and packages. Techniques: analyze data and generate results by R. Attitude: accomplish analyses and find out problems.
Data models and analysis	Data mining: Data Pattern & Sequence, Text Mining Statistics: Regression, cluster analysis Visualization: 2D,3D, machine learning	Knowledge: introduce data analysis models, including statistical analysis, data visualization, machine learning, data mining, and relating R packages Techniques: analyze data and generate results by R. Attitude: accomplish analyses and solve problems.

There are plenty of names describing big data talents, and their main job content is showed in fig. 3: depend on the complexity of works, data analysts should have the knowledge of industries and the ability of statistical analysis, data engineers should possess the knowledge of computer science and database usages, data scientists need to connect both – find out important management meanings in big data by information systems.
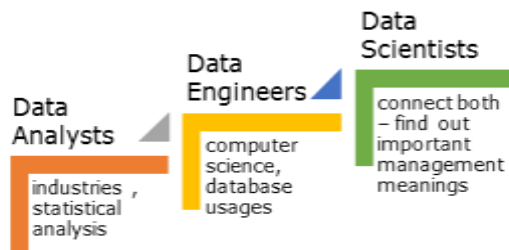

Fig. 3 Main job content of big data talents

To sum up, the courses of big data analysts emphasize the combination of theoretical knowledge and practical applications. Courses focus on how to collect qualitative and quantitative data from various platforms, by the help of information technologies, and process them to provide enterprise value-added service. Based on the background knowledge of trainees, different suitable competency development routes can be provided. For example, a trainee who has business and statistics background can choose the courses of big data software developers, and acquire the authentication of big data analysts after accomplishing the courses; a trainee who has the base of information technology choose to improve the competency of big data managers, and acquire the authentication of big data analysts after improving. This research generates the competency development route map of big data analysts as fig. 4, also provides further description:

1. Big Data Managers: Based on business and statistics, managers should have the competencies of data collecting, data distinguishing, statistical analysis and data interpretation. Basic big data managers focus on theoretical knowledge, advanced big data managers emphasize practical ability on data management.

2. Big Data Software Developers: Based on information science, software developers should have the competencies of programming languages like R and Hadoop, database usage, and data visualization. Software developers have to combine different big data analytics software to solve problems. Basic big data software developers just operate the software, while advanced big data software developers generate the solutions by software.

3. Big Data Analysts: Possess both the abilities of data analyzing, interpretation and software operation, big data analysts can solve problems, accomplish programs, and display results independently.
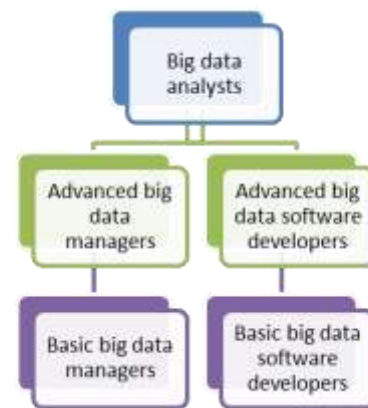

Fig. 4 Competency development route map of big data analysts

V. CONCLUSION

In accordance with "2016-2018 Talents Research and Estimation Report" generated by Manpower Supply and Demand Information Platform of National Development Council, R.O.C.,[5] the industry of cloud service and big data is one of the key industries, big data analysts are focal talent needs as well. Talent development of big data analysts is the inevitable trend while the world keep going digitized, Taiwan,

always maximize in hardware production in the world industrial value chain, still can maintain the advantage in this cloud service and big data industry as long as cultivating enough technical and managing talents. This research will continuously cooperate with training institutes to develop competency indicators of big data analysts, as well as the training courses with e+C blended learning model, expecting to improve the abilities of people who would like to learn and enlighten other training institutes on the need of cultivating big data talents, and in consequence, narrow the talent gap of big data analysts of Taiwan.

ACKNOWLEDGMENT

REFERENCES

[1] *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, McKinsey Global Institute, 2011.

[2] *Closing Canada's Big Data Talent Gap,* Canada's Big Data Consortium October 2015, Toronto, Canada, 2015.

[3] E. Bertino, "Big Data - Opportunities and Challenges", In *Proceedings of the 2013 IEEE 37th Annual Computer Software and Applications Conference*, pp. 479-480.

[4] T. Murdoch and A. Detsky, "The Inevitable Application of Big Data to Health Care", *JAMA*, vol. 309, no. 13, pp.1351-1352, April 2013.

[5] *2016-2018 Talents Research and Estimation Report*, Manpower Supply and Demand Information Platform of National Development Council, R.O.C., June 2016.

Hsu, Shun-Fa, born in Taiwan, is an Assistant Professor of in Department of Information Management, and the director of "Intelligent Cloud-computing & Big-data Center" at Chung-Chou University of Science and Technology (CCUT), Changhua County, Taiwan (R.O.C.). He is also the member of IAE. He graduated with honors from National Changhua University of Education (NCUE) with a Ph.D. degree in business education. His major research fields are business education, HRD & talent training, e-learning & Competency-Based Training (CBT) and EC & internet marketing, etc.

He is the Consultant of TTQS, one project of Workforce Development Agency, Department of Labor, R.O.C., which helps companies and training institutes knowing the talent training management system and ensures the quality of training activities in organizations. He helps some organizations awarding "TTQS Silver Medal", such as "Kits Co., Ltd.", "Commercial Association of Nantou Culinary", "Ding-Wang International Food and Beverage Co., Ltd.", "Howards Hotel (Taichung)" and "PRO Machinery Co., Ltd.", etc.

Dr. Hsu is also the Director of "Small & Medium Enterprises Diagnosis & Marketing Center" in College Tourism and Management at CCUT, and hosted several projects for Ministry of Education, R.O.C. to bridge the industrial value chains of Apps through the alliance of university-enterprise cooperation, all participants of these projects like companies, instructors, and students got benefits.

Lee, Yi-In received the MS degree in statistic from the Fu Jen Catholic University in 1999 and received the PhD degree in Management from the National Yunlin University of Science and Technology, Taiwan (R.O.C.), in 2006.

She joined Chung Chou University of Science and Technology in 2002. She is currently an Assistant Professor of Information Management at the Faculty of Management. She is also the Chief Operating Officer of the Operation Center and the Chief Advisor of Intelligent Cloud-computing & Big-data Center. She has been in the teaching profession for 15 years. She has extensive experience in teaching data analysis, marketing research, customer relationship management, human resource management. Her research interests include big data analysis, service quality management, training and education. She has published over 20 papers in international journals and conference proceedings, and hosted over 70 projects in the past decade.

Hsu, Ching-Tzu, born in Changhua County, received the bachelor degree in geography from National Taiwan University (NTU), Taipei, Taiwan, R.O.C. in 2016.

She is the Research Assistant in Academia Sinica, Taiwan (R.O.C.). She had joined in the short-term exchange program to Kanazawa University, Japan, involving geological research, and cultivated students of Kang Chiao Bilingual School for iGeo competence for four years.