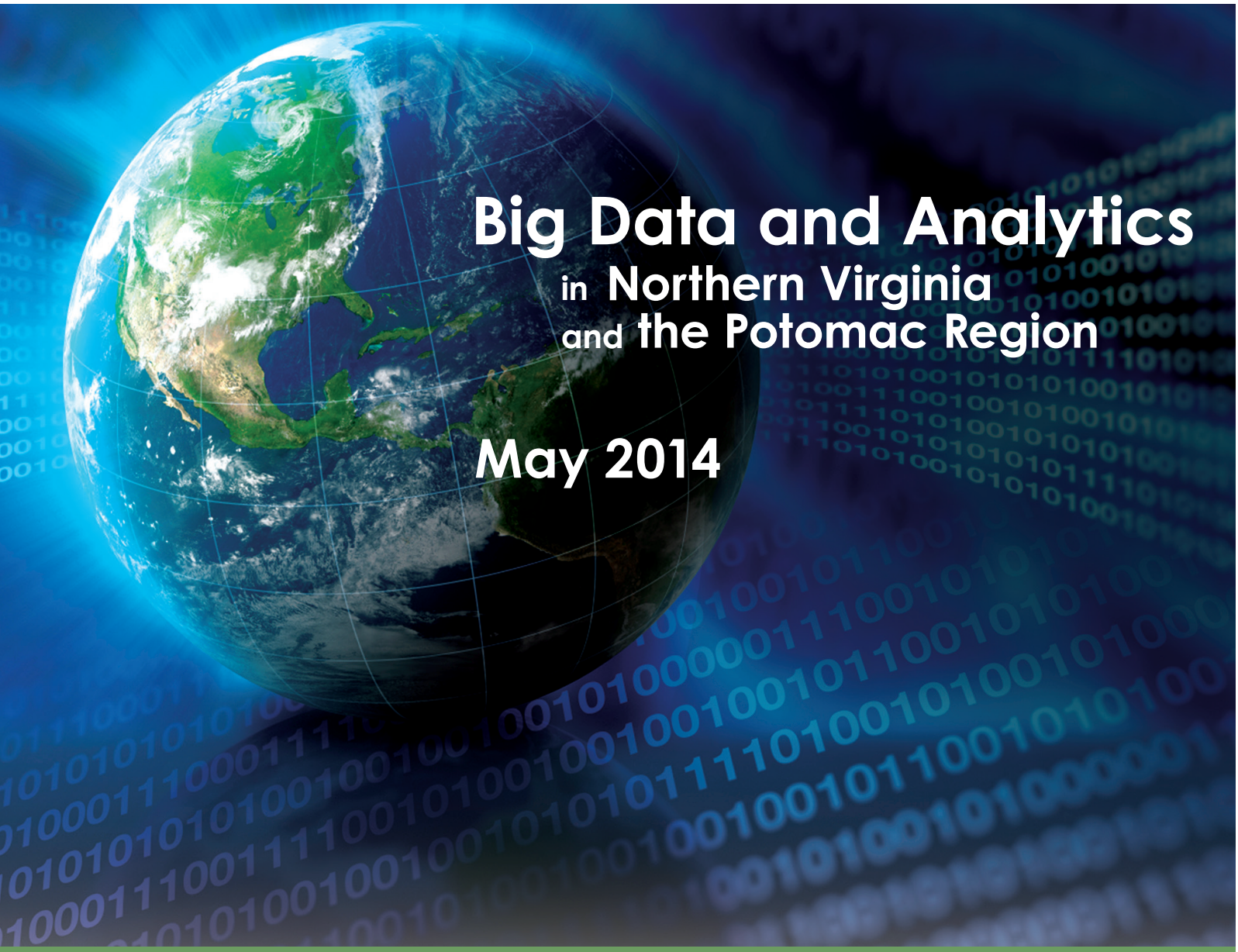


Big Data and Analytics
in Northern Virginia and the Potomac Region

May 2014

Sponsored by

nvtc NORTHERN VIRGINIA TECHNOLOGY COUNCIL

GW The George Washington University Virginia Science & Technology Campus

ATTAIN Proven Performance. Forward Thinking.

Prepared by

Chmura Economics & Analytics

About the Northern Virginia Technology Council

The Northern Virginia Technology Council (NVTC) is the membership and trade association for the technology community in Northern Virginia. As the largest technology council in the nation, NVTC serves about 1,000 companies and organizations, including businesses from all sectors of the technology industry, service providers, universities, foreign embassies, non-profit organizations and government agencies. Through its member companies, NVTC represents about 300,000 employees in the region. NVTC provides value by connecting, growing, educating and advocating for the innovative and entrepreneurial technology community in Northern Virginia and the greater region.

Learn more at **www.nvtc.org**.

About Chmura Economics & Analytics

Chmura Economics & Analytics (Chmura) specializes in applied economic consulting, quantitative research, and software solutions requiring the integration of advanced economic analysis. Our work has included economic and workforce development, site selection, and impact analysis. Our technology platform, JobsEQ®, is currently in widespread use by economic development agencies throughout the country. Chmura's national client base includes economic developers, workforce practitioners, education reformers, firms seeking geo-location labor data, as well as many other public and private entities with needs for credible information on a timely basis.

Learn more at **www.chmuraecon.com**.

About the George Washington University Virginia Science and Technology Campus

The GW Virginia Science and Technology Campus, established in Loudoun County in 1991, is a robust center for innovative research, graduate education and regional corporate partnerships. As the university's science and technology campus, it hosts 17 laboratories and institutes in critical areas such as computational biology, energy science and technology, transportation safety and high performance computing. The GW Virginia Science and Technology Campus also combines its research excellence with more than 15 exceptional graduate degree and certificate programs in business, education, and engineering and unique undergraduate programs in the health sciences. GW has been educating Virginians for more than 50 years, beginning in Hampton Roads in 1958 and expanding its service across the Commonwealth, notably in Arlington, Alexandria and Ashburn.

Learn more at **http://virginia.gwu.edu**.

About Attain, LLC

Based in Vienna, Va., Attain is a leading management, technology and strategy consulting firm comprised of innovative problem solvers who deliver tangible results to address today's complex public sector, health, and higher education challenges. Attain's transformative business and IT solutions and services deliver market-leading results to our customers in the defense, civilian, and federal health markets, as well as the higher education, academic medical center and non-profit communities. With a nationwide presence, Attain is an agile company with the experience to scale our solutions and services to meet your needs.

Learn more at **www.attain.com**.

Contents

ACKNOWLEDGEMENTS

The Northern Virginia Technology Council thanks the George Washington University Virginia Science & Technology Campus and Attain, LLC for their generous financial support of this research. In particular, NVTC recognizes the following individuals whose contributions, support, and advice were essential to the report:

The George Washington University

Dr. Steven Knapp, President

Dr. Ali Eskandarian, Dean, Virginia Science & Technology Campus

Dr. Leo Chalupa, Vice President for Research

Dr. Christopher Deering, Senior Associate Dean and Associate Provost, Virginia Science & Technology Campus

Dova Wilson, Marketing Communications Specialist, Virginia Science & Technology Campus

Andrea Barnes, Special Assistant, Office of the Dean, Virginia Science & Technology Campus

Attain, LLC

Greg Baroni, Chairman and CEO

Darlene Covington, Marketing and Communications, Corporate Services

Andrew Ritcheson, Senior Manager, Federal Services

Chmura Economics & Analytics (Chmura) and the Northern Virginia Technology Council also thank the following firms and individuals for their participation in the focus groups that provided valuable insights into Big Data and Analytics that helped shape the survey design:

Acentia	Tom Woteki	Senior Vice President and Chief Architect, Strategy and Solutions
Amazon Web Services	Tim Bixler	Federal Manager, Solutions Architecture
AOL	Jon Abramson	Sr. Director - Consumer, Membership & Revenue Analytics
CALIBRE	Carter Schoenberg	Director of Cyber Security Services
CALIBRE	Don Taylor	Vice President of Business Development
CGI Federal	Malcolm Harden	Vice President and Chief Solution Architect
CIT	Paul McGowan	VP & Managing Director Consulting Services
comScore	Magid Abraham	Executive Chairman and Co-Founder
CSC	Yogesh Khanna	CTO, North American Public Sector
IBM Federal	Frank Stein	Director, Analytics Solution Center
ICF International	Doug Beck	Senior Vice President & Director, Corporate Development
Noblis	H. Gilbert Miller	Corporate Vice President and CTO
NTELX	Rob Quartel	Chairman and CEO
Optensity	Pamela Arya	CEO
Razorsight	Suren Nathan	CTO
SAIC	Tony Moraco	CEO
Software AG	Mike Shumpert	Vice President, Industry Solutions
The George Washington University Virginia Science & Technology Campus	Chris Deering	Senior Associate Dean and Associate Provost

In addition, NVTC is grateful to the NVTC Big Data Task Force, and its co-chairs Anne Altman of IBM and Jennifer Morgan of SAP, for laying the groundwork leading up to this study.

EXECUTIVE SUMMARY

The evolution of data storage, computing power, and intelligent software has advanced our capacity to manage and manipulate large volumes of varied data. The growth of such "Big Data and Analytics" has led to new ways to capture, store, and interpret data across business, government, and research functions.

As one of the nation's premier high-tech centers, Northern Virginia and the greater Potomac region are home to a growing number of firms that both provide and use Big Data as a core competency. In addition, the region's proximity to the federal government, which produces, stores, and manages some of the largest data sets in the world; its historic workforce expertise in managing and interpreting data; and its many outstanding institutions of higher learning make it uniquely positioned to leverage the growing demand for Big Data and Analytics services. The region's higher education institutions will play a particularly central role in developing Big Data and Analytics talent and driving the competitiveness of the region as a destination for Big Data and Analytics businesses.

Based on focus groups with regional technology leaders and a survey of organizations active in the sector, *Big Data and Analytics in Northern Virginia and the Potomac Region* catalogues the breadth and depth of Big Data and Analytics experience, expertise, and assets of the region. This report outlines how the resources and knowledge of the region's providers, users, and educators will allow it to advance its leadership position in Big Data and Analytics and drive future economic development in the region.

Key Findings:

The Nation's "Data Capital" is a Natural Leader in Big Data and Analytics.

1. Northern Virginia respondent firms generate an estimated 717 terabytes of new data on an average day. Additionally, up to 70 percent of the world's Internet traffic passes through Northern Virginia each day.
2. Seventy-two percent of respondent firms working with Big Data and Analytics in the Potomac region are located in Northern Virginia
3. Eighty-seven percent of Northern Virginia respondent firms have active or planned Big Data and Analytics projects.
4. With proximity to the federal government and data centers, regional firms have unique subject matter expertise related to Big Data and Analytics in areas such as intelligence/ national security, cyber security, and healthcare.
5. Organizations in Northern Virginia are putting the flood of data moving through the region to use for both customers and their own internal projects.
6. Reflecting the region's high demand for Big Data and Analytics services, the majority of Big Data and Analytics related sales by firms in the Potomac region are provided to customers within the Potomac region.

Big Data and Analytics Jobs are Expected to Be a Key Driver of Regional Employment Growth.

7. Big Data and Analytics related occupations already make up an average of 11 percent of the workforce in respondent firms.
8. Big Data and Analytics firms in the region are positioning for an average annual employment growth rate over the next three years that is three times as fast as annual average forecasts from the U.S. Bureau of Labor Statistics for all occupations over the next decade.
9. Big Data and Analytics firms in the Potomac region expect their employment to grow by 20,500 jobs over the next three years, with 64 percent of those jobs related to Big Data and Analytics.
10. Big Data and Analytics firms in Northern Virginia expect to add more than 16,700 jobs over the next three years, and more than 73 percent of those are expected to be occupations related to Big Data and Analytics.

Together, Educational Institutions and Businesses are Producing Talent to Address a Surging Demand.

11. For almost all Big Data and Analytics related degree programs in the region, total degree awards fall short of total annual demand, indicating there are shortages of workers with those degrees and/or employees are hiring students from outside the region. In the 2011–2012 academic year, the regional and statewide higher education institutions awarded 8,844 degrees in Big Data and Analytics related fields, well short of industry demand.
12. Colleges and universities in Northern Virginia and the Potomac region are graduating more students than ever before with Big Data and Analytics related degrees and up-skilling current employees with Big Data and Analytics related training.
13. Seventy percent of respondent colleges, universities, and other educators in the Potomac region offer or intend to offer courses or programs specifically targeting Big Data and Analytics by 2015; within the next five years, 90 percent expect to offer courses or programs.
14. Sixty-six percent of respondent firms provide Big Data and Analytics training for employees. Some note using local Virginia and Maryland colleges and universities for their training, but the majority of these firms have developed their own in-house or on-the-job training, or work with local dedicated information technology (IT) training firms and partners.
15. While visualization was ranked by businesses to be the most important workforce attribute for careers in Big Data and Analytics, unstructured data ranked first in importance for educators in the region.

1. BACKGROUND

Over the last half-century, the evolution of data storage and computing power, along with the development of intelligent software, has enhanced the human capacity to manage and manipulate large volumes of varied data. Since data have become more integrated into our lives and work, the term "Big Data and Analytics" has emerged. As the name implies, there are two sides to this phrase: the actual data as it exists and the analysis of data as valuable insight.

Big Data and Analytics reflects the challenges of data that are too vast, too unstructured, and too fast-moving to manage by traditional methods. Gleaning meaningful information and competitive advantages from massive amounts of data has become increasingly important to organizations globally. From businesses and research institutions to governments, organizations now routinely generate data of unprecedented scope and complexity, capturing information about transactions, customers, suppliers, operations, and more. Thus, analytics has become inextricably vital to realizing the full value of Big Data. These entities are looking to harness the power of Big Data and Analytics to improve their business performance and increase their market share.

Big Data and Analytics applies to many industries from manufacturing to services, and by its nature creates a plethora of interdisciplinary opportunities. Not surprisingly, firms that provide Big Data and Analytics products and services rely heavily on individuals with strong science, technology, engineering, and math (STEM) skills. In addition, institutions need data savvy managers capable of asking the right questions, interpreting and challenging the results, and making effective business decisions. As Big Data and Analytics continues to evolve, these highly-educated workers will remain in high demand.

Technology leaders in the Potomac region are keen on leveraging the area's significant existing Big Data and Analytics resources to ensure the region remains positioned at the forefront of Big Data and Analytics development. For that reason, the Northern Virginia Technology Council (NVTC) retained Chmura Economics & Analytics (Chmura) to conduct a regional Big Data and Analytics asset/inventory study, with an emphasis on Northern Virginia. This report, Big Data and Analytics in Northern Virginia and the Potomac Region, presents the findings of that study.

The analysis of Big Data and Analytics in the region is based on focus groups with regional technology leaders and a survey of organizations active in the sector, details of which are provided in Section 4 of this report.

Purpose

Big Data and Analytics in Northern Virginia and the Potomac Region catalogues the depth and breadth of experience, expertise, and assets the region possesses. This research can help the region's business and academic communities to achieve the following economic development goals:

- Stimulate Big Data and Analytics-driven economic development
- Demonstrate to Big Data and Analytics related businesses, investors, and customers why they should invest in the region
- Grow the region's Big Data and Analytics workforce and its workforce development capabilities

Approach

This report was completed in three phases. Phase 1 included focus group sessions, which focused on the definition of Big Data and Analytics, given the relative novelty of the phrase. In Phase 2, primary data were collected to identify the number of Big Data and Analytics firms in the region and to inventory their assets. Finally, the survey results were analyzed in Phase 3. Further discussion of Chmura's approach to producing this report can be found in **Appendix 2**.

Together, the three phases of this project provide evidence that Northern Virginia has the suppliers, users, educators, and expertise needed to advance its leadership position in Big Data and Analytics and drive economic development in the coming years.

Northern Virginia has the suppliers, users, educators, and expertise needed to advance its leadership position in Big Data and Analytics and drive economic development in the coming years.

Organization of Report

Section 1. Background

Section 2. Defining Big Data and Analytics: This section summarizes Chmura's comprehensive literature review, including studies from academic journals, industry reports, and other internet sources, to address the definition, market components, and potential economic value of Big Data and Analytics. This section also focuses on the specific markets and underlying industries with which regional Big Data and Analytics firms interact and summarizes focus group discussions among the region's subject matter experts.

Section 3. Workforce Skills and Education: When an industry or group of industries are growing rapidly, it is not unusual for a shortage of workers to occur until educational institutions and training organizations build the capacity to teach more individuals, and more people are attracted to the needed occupations. This section summarizes workforce issues related to Big Data and Analytics, including knowledge and skills needed by Big Data and Analytics workers, as well as education and training offered by the area's academic institutions.

Section 4. Big Data and Analytics Assets: Northern Virginia and the Potomac Region: In February 2014, Chmura surveyed businesses in the Potomac region to identify firms' characteristics, their current and expected Big Data and Analytics related workforce needs, and the role education and training is playing to fill those needs. A total of 270 organizations responded to the survey. This section analyzes the survey results and their implications for the region's Big Data and Analytics assets.

2. DEFINING BIG DATA AND ANALYTICS

Few terms in the common business lexicon are as variably defined and differently understood as Big Data and Analytics. Moreover, defining the market for Big Data and Analytics is difficult from an economic perspective because it crosses into many different industries, as defined by the North American Industry Classification System (NAICS), but is not, however, fully represented by any one particular NAICS code. Yet some shared definition, even of the most rudimentary concepts of Big Data and Analytics, is central to cataloguing the regional assets tied to it as a market.

This section summarizes Chmura's comprehensive literature review, including studies from academic journals, industry reports, and other internet sources,[1] to address the definition, market components, and potential economic value of Big Data and Analytics. The literature review also focuses on the specific markets and underlying industries with which regional Big Data and Analytics firms interact. This section also summarizes the focus group discussions among the region's subject matter experts. For further background on the definitions of Big Data from the literature review, **see Appendix 1**.

Technical Definition

Neither a comprehensive literature review regarding Big Data definitions, nor comments from focus group participants[2] could yield a widely accepted, uniform definition of the term. Instead, Big Data and Analytics experts continue to provide diverse and sometimes contradictory definitions, with some emphasizing data size and complexity, while others highlight the business value of the Big Data and Analytics.[3]

"The idea that Big Data firms must add value was the mantra of most focus group participants."

Most focus group participants, however, acknowledged that the 3Vs Gartner proposed in 2001 are useful in interpreting Big Data: volume, velocity, and variety.[4] Focus group participants also agreed with Gartner's expanded definition from 2012 which included veracity, representing requirements about trust and uncertainty pertaining to data and the outcome of data analysis.[5]

In a 2012 report, IDC defined an additional V as value—highlighting that Big Data applications need to bring incremental value to businesses.[6] This concept resonated strongly with the focus group.

The idea that Big Data firms must add value was the mantra of most focus group participants. Based on this perspective, the focus groups determined that references to Big Data should be expanded to "Big Data and Analytics." As one participant said, "Big Data and Analytics is part of a delivery chain in providing value to the client." The expanded name is more inclusive and acknowledges that analytics are an integral part of creating value from Big Data.

[1] Keywords searched included the following: Big Data definition, Big Data market, Big Data industry, and Big Data skills. Only literature germane to this study is included in this report.
[2] Notes from the focus group event are available upon request. Names are not associated with any of the comments to maintain anonymity.
[3] The full literature review is found in **Appendix 1** of this report.
[4] Source: Data Management: Controlling Data Volume, Velocity, and Variety. Gartner, 2001.
[5] Source: The Importance of big data: A Definition. Bu M.A. Beyer and D. Laney, Gartner, 2012.
[6] Source: Worldwide Big Data Technology and Services, 2012-2015 Forecast, IDC Market Analysis. Available at: http://download.microsoft.com/download/7/B/8/7B8AC938-2928-4B65-B1B3-0B523DDFCDC7/IDC%20Report-Worldwide%20Big%20Data%20Technology%20and%20Services.pdf.

In conclusion, the general consensus of subject matter experts in the Potomac region was that Big Data is most fully understood when each of its characteristics—volume, velocity, variety, veracity, and value—is considered. Moreover, they agree that V for value implies that the term "Analytics" is inextricable from the purpose or meaning of Big Data.

Defining the Big Data and Analytics "Industry" in the Potomac Region

To better define the region's Big Data and Analytics market, including the number of firms and employment, focus group participants discussed using the North American Industry Classification System (NAICS). However, they determined it was not a viable option. Although using NAICS would enable comparisons between the Potomac region and other regions around the country, including growth over time, it would likely overstate total employment related to Big Data and Analytics because there is not a clear mapping from NAICS to all of the industries Big Data and Analytics in some way overlaps.

This report classifies the Potomac region's Big Data and Analytics industry as both the providers and users of Big Data and Analytics. Since many companies and institutions in the Potomac region do not derive all of their revenue from Big Data and Analytics market segments, and a reliable NAICS mapping does not exist, Chmura determined that a survey was the best approach to identify the Big Data and Analytics assets in the Potomac region.

The survey of regional Big Data and Analytics subject matter experts asked recipients to identify staff and business operations associated with Big Data and Analytics. The results of this survey are found in Section 4 of this report.

Overview of Global Providers, Users, Market Size, and Trends

From an economic perspective, the value of Big Data and Analytics lies not in the terabytes or petabytes of data per se, but in how Big Data and Analytics can generate insights and revenue for businesses. Big Data and Analytics can generate revenue for two broad groups of market participants: (1) providers that supply hardware, software, and services that facilitate collection, storage, and analysis of Big Data and Analytics; and (2) users, which are companies applying Big Data and Analytics to both improve their business intelligence and increase their market share and profitability.

Providers

Broadly speaking, Big Data and Analytics providers include suppliers of a new generation of technologies and architectures designed to extract value from very large volumes of a variety of data by enabling high-velocity capture, discovery, and analysis. According to IDC,[7] Big Data suppliers can be classified into three major segments—hardware, software, and services. Hardware includes external storage systems, servers, and datacenter networking infrastructure. Software includes data organization and management software, analytics and discovery software, and decision support and automation software. Finally, Big Data and

[7] Source: Undefined by Data: A Survey of Big Data Definitions, by Jonathan Stuart Ward and Adam Barker, School of Computer Science, University of St. Andrews, UK, 2013. Available at: http://arxiv.org/pdf/1309.5821v1.pdf. --

Analytics services include business consulting, IT outsourcing, IT support, and training services. Other studies on Big Data and Analytics markets follow similar classifications of hardware, software and services. [8]

Users

Big Data and Analytics can offer unprecedented insight, improved decision-making skills, and untapped sources of profit.[9] As a result, many firms across nearly all industries are working to harvest the business value provided by Big Data and Analytics.

In July 2011, McKinsey Global Institute published the study "Big Data: The next frontier for innovation, competition and productivity."[10] This report focused on the viable applications of Big Data to transform industries, and highlighted five industries with high potential: (1) the healthcare sector in the United States, (2) public sector administration in the European Union, (3) the retail sector in the United States, (4) global manufacturing, and (5) global personal location data.

For those industry sectors identified, the McKinsey report asserts that Big Data can:

- Improve transparency
- Enable experimentation to discover needs, expose variability, and improve performance
- Segment populations to customize solutions
- Replace/support human decision making with automated algorithms
- Underpin innovation in new business models, products, and services

Many Big Data and Analytics firms in Northern Virginia will be leveraging their expertise and capabilities to transform the departments, agencies, and organizations in the U.S. federal government.

Within the Potomac region, Big Data and Analytics also has the potential to transform the government. Undoubtedly, many Big Data and Analytics firms in Northern Virginia will be leveraging their expertise and capabilities in this area to transform the departments, agencies, and organizations in the U.S. federal government. In 2012, the TechAmerica Foundation Big Data Commission released a report concluding that Big Data can be a transformational force in improving the government sector.[11] The report concluded that Big Data can enable government organizations to be smarter, improve their productivity, and serve the needs of their stakeholders by improving decision making in individual agencies

[8] The Wikibon report, for example, used the same hardware, software, and service classification: Big Data Market Size and Vendor Revenues, by Jeff Kelly, David Vellante, and David Floyer, April 2013. Available at: http://wikibon.org/wiki/v/Big_Data_Market_Size_and_Vendor_Revenues. A study by Transparency Market Research used slightly different groupings—it listed storage as a separate category, but combined software and service together: Big Data Market, Global Scenario, Trend, Industry Analysis, Size, Share and Forecast, 2013. Available: http://www.academia.edu/3072476/Big_Data_Market_-_Global_Scenario_Trends_Industry_Analysis_Size_Share_And_Forecast_2012_-_2018.

[9] Source: The Big Data Conundrum: How to Define It. MIT Technology Review, October 2013. Available at: http://www.technologyreview.com/view/519851/the-big-data-conundrum-how-to-define-it/.

[10] Source: McKinsey Global Institute, 2011. Available at: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.

[11] Source: Demystifying Big Data: A Practical Guide to Transforming the Business of Government, Prepared by TechAmerica Foundation's Federal Big Data Commission. Available at: http://www.techamerica.org/Docs/fileManager.cfm?f=techamerica-bigdatareport-final.pdf.

and across the government ecosystem. The report highlighted several areas where Big Data can be the most effective, such as: healthcare, transportation, education, cybersecurity, and weather. The report highlighted several Big Data in government case studies, such as the National Archive and Records Administration (NARA) Electronic Records Archive Program, which provides electronic records archives and an online public access system for U.S. records. Another example is Center for Medicare & Medicaid Services (CMS) Medical Records Analytics program, which aims to protect the health of all Americans and ensure complaint processing of insurance claims. Other agencies such as International Revenue Service and National Aeronautics and Space Administrations have also started to harness the power of Big Data and Analytics.

A more recent survey by Gartner found that Big Data investments in 2013 continued to rise, with 64 percent of surveyed organizations investing or planning to invest in Big Data technology, compared to 58 percent in 2012.[12] The study found that industries leading Big Data investment are media and communications, banking, and services. (**See Appendix 1** for additional information on users of/markets for Big Data.)

Market Size and Trends

While the phrase "Big Data and Analytics" is only now becoming more commonplace, firms and institutions have used Big Data for a long time. Today's increased amounts of available data, in part due to streaming data from social media, and the ease and speed of processing offered by cloud computing are two possible reasons why Big Data and Analytics is seen as an emerging and fast growing market.

Big Data and Analytics is a multi-billion dollar global market that is expected to expand rapidly through at least 2017. IDC estimated that the worldwide Big Data market amounted to $4.8 billion in 2011, with distribution of profits at 42 percent from services, 30 percent from software, and 29 percent from hardware. A report from Wikibon showed a similar composition of three categories in 2012: total Big Data revenue was $11.5 billion, with 44 percent from services, 25 percent from software, and 31 percent from hardware.[13]

IDC and Wikibon differed in projecting future market growth, even though both project explosive increases. IDC forecasts that from 2010 to 2015, the worldwide Big Data market will grow by 39.4 percent per year, reaching $16.9 billion in 2015. Wikibon is more optimistic, forecasting that the Big Data market will grow 58 percent per year from 2012 to 2017 and will reach $32.1 billion in 2015 and $53.4 billion in 2017.

[12] Source: Gartner Survey Reveals That 64 Percent of Organizations Have Invested or Plan to Invest in Big Data in 2013. September 2012. Available at: http://www.gartner.com/newsroom/id/2593815.

[13] Source: Worldwide Big Data Technology and Services, 2012-2015 Forecast, IDC Market Analysis. Available at: http://download.microsoft.com/download/7/B/8/7B8AC938-2928-4B65-B1B3-0B523DDFCDC7/IDC%20Report-Worldwide%20Big%20Data%20Technology%20and%20Services.pdf.

3. WORKFORCE SKILLS AND EDUCATION

With the significant projected revenue growth in Big Data and Analytics in the nation and the Potomac region, talent acquisition will undoubtedly be essential to extracting its full value. Thus, it is important for the region's Big Data and Analytics providers and educational institutions to determine what types of skills and training are needed to fill Big Data and Analytics related jobs, as well as how to recruit talent and continue creating Big Data and Analytics related jobs. These considerations are addressed in this section using secondary data from a proprietary resource: JobsEQ® Technology Platform[14] and primary data, detailed in Section 4, collected from regional firms and educational institutions associated with Big Data and Analytics.

Studies show that talent to support Big Data and Analytics firms already is, and will continue to be, in high demand in the nation. For example, a McKinsey study found that by 2018 there could be a shortage of 140,000 to 190,000 individuals in the United States with deep analytical talent needed for Big Data and Analytics.[15] Northern Virginia and the Potomac region show evidence of shortages as well based on the findings in this report, which will be examined in more detail later in this section.

Big Data and Analytics Skills

Human capital and skills needed to analyze Big Data combine strong analytics, software proficiency, and a business acumen to ask the right questions and guide analyses toward driving value. In the "1st Industry Trends Report," Mark Whitehorn, chair of analytics and professor at the School of Computing, University of Dundee, explained that the necessary skills be differentiated into four areas:[16]

- **Business Intelligence (BI):** understanding of relational and multi-dimensional design and structures, database programming languages, and strong verbal and written communication skills
- **Analytics:** backgrounds in data, data mining, statistics, and algorithm design
- **Software:** proficiency in NoSQL systems and accompanying languages
- **Experience:** knowledge of the data field and creative minds for new ideas and analysis techniques that generate value

[14] The JobsEQ® Technology Platform is a software system in widespread use across the country that provides access to historical and forecasted national and regional economic data for economic and workforce development and education, including industry, occupations, and curriculum.

[15] Source: McKinsey Global Institute. Big Data: The next frontier for innovation, competition, and productivity, 2011. Available at: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation. See also 1st Industry Trends Report, Big Data Insight Group, April 2012, page 8; available at: http://www.thebigdatainsightgroup.com/site/article/1st-big-data-insight-group-industry-trends-report. And "e-skills UK is the Sector Skills Council for Business and Information Technology working on behalf of employers to develop the software, internet, computer gaming, IT services and business change expertise necessary to thrive in today's global digital economy." Source: SAS, November 2013. Available at: http://www.sas.com/offices/europe/uk/downloads/bigdata/eskills/eskills.pdf.

[16] Big Data Insight Group, page 13.

The authors of "Analyzing the Analyzers, An Introspective Survey of Data Scientists and Their Work" point out the importance of effective communication of job descriptions when firms are hiring talent.[17] Job titles are very important in recruitment, as "data scientist" is often too vague. Their survey broke down the data scientist role into the following four clusters using self-identification:

- **Data Developers:** developers and engineers
- **Data Researchers:** researchers, scientists, and statisticians
- **Data Creatives:** jacks of all trades, artists, and hackers
- **Data Businesspeople:** leaders, business people, and entrepreneurs

While these skills and talent segments are the necessary building blocks for Big Data and Analytics, firms also must have domain expertise capable of telling a story from the underlying data. It is not enough for firms to have the talent needed to work with Big Data and Analytics; they must have domain-specific knowledge and the capability of understanding the unique nature of the underlying data. This insight drives value creation.

Potomac region focus group participants voiced difficulty in finding the skilled workers needed for Big Data and Analytics tasks. The following skills and degrees were identified as needs: analytical expertise, data interpreter, data analyst, data scientist, information technology professionals, mathematicians, and statisticians. Several focus group members mentioned that the best data scientists usually come from a science background, and are grounded in creating and testing hypotheses. In addition to scientists and information technology workers, the need for domain expertise was recognized in areas such as healthcare, finance, and other fields.

Big Data and Analytics Workforce Development

The Skills Gap

Chmura examined the number of Big Data and Analytics related graduates in the 2011–2012 academic year in the Potomac region.[18] Overall, the regional and statewide higher education institutions awarded 8,844 degrees in Big Data and Analytics related fields.[19] Among those, 1,528 were two-year degrees or certificates, 4,671 were bachelor's degrees, and 2,645 are postgraduate degrees. Chmura also identified the annual regional demand for those degrees based on its proprietary JobsEQ®[20], as well as the current employment in the region. The difference between total awards and total annual demand can be used as a measure of the gap between the supply (awards) and demand for Big Data and Analytics related talent.

[17] Source: Analyzing the Analyzers, An Introspective Survey of Data Scientists and Their Work, by Harland Harris, Sean Murphy, and Marck Vaisman, published by O'Reilly, 2013. Available at: http://www.oreilly.com/data/free/analyzing-the-analyzers.csp. This classification is similar to those in the McKinsey Global Institute report, with the exception of an additional group of data creatives.

[18] This is the latest degree award data from National Center of Education Statistics (NCES). Some programs will not be reflected as they are relatively new. Please see **Appendix 2** for detailed degree data.

[19] The institutions analyzed in this report include all those located in the Potomac Region, plus University of Virginia, Virginia Tech, and Virginia Commonwealth University. The latter three, despite not physically located in the Potomac Region, supply a large number of graduates to the area businesses. They are referred as statewide institutions in this report.

[20] JobsEQ® is a national labor-related database created by Chmura Economics & Analytics. It provides data related to industry clusters, occupation demand and supply, and education.

The overall supply and demand pattern confirms the widespread belief that there is a talent gap in Big Data and Analytics fields. For almost all Big Data and Analytics related programs, total degrees fall short of total annual demand, indicating that there are shortages of workers with those degrees and/or businesses will need to bring graduates in from outside the region to fill positions. The talent gap surrounding Big Data and Analytics cannot be bridged by any one method. It is being addressed by several strategies—each approach with different benefits. The two major supply sources of Big Data and Analytics talent are graduates from colleges and universities with Big Data and Analytics related degrees, and existing workers who are up-skilled with Big Data and Analytics related training.

The overall supply and demand pattern confirms the widespread belief that there is a talent gap in Big Data and Analytics fields.

Virginia's higher education institutions will be playing a central role in developing Big Data and Analytics talent in the state. They not only supply new graduates with Big Data and Analytics related degrees and certificates, but also can help Virginia's business community train and up-skill their existing workforce. Higher education institutions also can help drive the competitiveness of the region and address the national shortfall of Big Data and Analytics talent by producing a disproportionate share of the data analysts and data managers that businesses nationally require. In doing so, they will enhance the attractiveness of Northern Virginia as a destination for Big Data and Analytics businesses.

Growing New Talent: Graduates with Big Data and Analytics Related Degrees

Many academic institutions in the nation and in the Potomac region are offering degrees related to Big Data and Analytics. One- and two-year master's degrees are the norm for full-time, comprehensive programs. These programs vary widely based on specializations (information management, business-meets-analytics, statistical and operations, etc.) and concentration areas (marketing, insurance, financial services). [21]

Virginia's higher education institutions will be playing a central role in developing Big Data and Analytics talent in the state.

Eight universities in the region and Commonwealth of Virginia awarded almost 72 percent of all Big Data and Analytics related degrees in the 2011–12 academic year (**Table 3.1**). Among those, public universities and colleges—University of Maryland-University College, George Mason University, and University of Maryland-College Park, Virginia Tech and Northern Virginia Community Colleges—are important suppliers of graduates, and each awarded more than 500 Big Data and Analytics related degrees. University of Virginia and Virginia Commonwealth University awarded more than 300 degrees each. Two private universities are also playing important roles, with the George Washington University awarding 440 degrees and Georgetown University awarding 142 degrees. All other colleges awarded less than 28 percent of Big Data and Analytics related degrees in the region. However, University of Maryland-University College offers online degrees. As a result, some of those receiving degrees from this college may come from outside the Potomac region and are less likely to participate in the regional workforce without relocation. [22]

[21] Source: Information Week, Big Data Analytics Master's Degrees, available at http://www.informationweek.com/big-data/big-data-analytics/big-data-analytics-masters-degrees-20-top-programs/d/d-id/1108042?page_number=1.

[22] Several for-profit colleges in the region also offer online programs. As a result, degree awards and gap data should be interpreted with caution.

Table 3.1: Big Data and Analytics Related Degrees by Colleges

Institutions	Number of Awards
University of Maryland-University College	1,817
George Mason University	1,077
University of Maryland-College Park	928
Virginia Tech	728
Northern Virginia Community College	504
The George Washington University	440
University of Virginia	364
Virginia Commonwealth University	325
Georgetown University	142
Others	2,519
Total	**8,844**

Source: JobsEQ

One example degree program is a Big Data and Analytics degree offered by the George Washington University (GW). In 2012, GW's Department of Decision Science developed a master's of science degree with a concentration in analytics. In fall 2013, the GW business school, in partnership with IBM, launched a new master's of science degree in business analytics.[23] Elsewhere in the Potomac region, Georgetown University plans to establish a Massive Data Institute in its public policy school, using Big Data to improve public policy decision making.[24] George Mason University also offers a graduate certificate and Master's degree in data analytics in its Volgenau School of Engineering.[25]

Other universities across Virginia are also offering Big Data and Analytics related curriculums. The University of Virginia's Big Data Institute[26] has created a master's degree program in data science and is developing both Ph.D. and undergraduate programs as well.[27] Virginia Commonwealth University offers a master's degree in business with a concentration in decision analytics.[28] The College of William and Mary is integrating Big Data into undergraduate studies through their EXTREEMS initiative,[29] funded by the National Science Foundation.[30] Virginia Tech also launched the Center for Business Intelligence and Analytics to support research and teaching in the growing fields of Big Data.[31]

[23] Source: GW Partners with IBM for New 'Big Data and Analytics' Degree, at: http://gwtoday.gwu.edu/gw-partners-ibm-new-%E2%80%98big-data%E2%80%99-degree.

[24] Source: Georgetown Receives $100M to Create New Public Policy School, at: http://www.georgetown.edu/mccourt-school-public-policy-announced.html.

[25] Source: George Mason University Catalog, available at: http://catalog.gmu.edu/preview_program.php?catoid=22&poid=21887.

[26] Source: University of Virginia, MS in Data Science, available at: http://people.virginia.edu/~jjh2b/msds/.

[27] Source: University of Virginia, UVA Today, available at: https://news.virginia.edu/content/uva-appoints-engineering-professor-don-brown-lead-new-big-data-institute.

[28] Source: VCU Business, available at: http://business.vcu.edu/graduate/dsba.html.

[29] Source: William and Mary News & Events, available at: http://www.wm.edu/news/stories/2013/extreems-mathematics-initiative-takes-dead-aim-at-big-data123.php.

[30] Source: National Science Foundation, available at: http://www.nsf.gov/awardsearch/showAward?AWD_ID=1331021&HistoricalAwards=false.

[31] Source: Virginia Tech News, available at: http://www.vtnews.vt.edu/articles/2014/04/042114-pamplin-centers.html.

Up-skilling Existing Workers: Big Data and Analytics related Training

McKinsey's study[32] suggests that while developing specific graduate programs and importing new talent are important pieces to closing the talent gap in the Big Data workforce, the real challenge will be to retain and up-skill the current talent. Businesses need to integrate Big Data and Analytics with existing expertise through up-skilling portions of their current workforce. To protect workforce investments, it is important to maintain high retention rates among that talent. This is also a cost-effective way of developing a Big Data and Analytics team with domain expertise specific to the business. Up-skilling the workforce to tackle Big Data and Analytics is most often accomplished through one of two paths: specific skill-set supplements through local colleges or online courses, or comprehensive masters' degree programs.

While Big Data and Analytics is a new field to teach, there are many different programs available to up-skill the workforce. Online education companies such as Udacity, edX, and Coursera have developed free or inexpensive educational courses by partnering with established top universities (examples include Caltech, Stanford, Massachusetts Institute of Technology, Harvard, Oxford) as well as with well-respected Big Data and Analytics related companies (examples include Cloudera and SAS). Most of these programs require a basic knowledge of statistics and are geared towards individuals with bachelor's degrees.

Some businesses have been successful in developing deep analytical talent in-house. McKinsey's study highlights a well-respected financial services company for its talent development strategy. The company developed relationships with nearby institutions and has a strong presence during on-campus recruiting; they also use the institutions to up-skill their employees. This cost-effective strategy increases the firm's competitiveness through institutional networking. Another talent up-skilling method appealing to larger firms is developing a fully customized internal training institute.

Big Data and Analytics Related Academic Programs

Many higher education institutions are still thinking through the curriculum design and offerings in this area. At this time, there is no standard Classification of Instructional Programs (CIP) code for Big Data and Analytics degrees or certificate programs.[33] For example, some programs in the study region, such as that at the George Washington University is housed in its business school, while George Mason's program is located in its engineering school. National surveys show that some universities' Big Data and Analytics programs are affiliated with the departments of computer science, statistics, or engineering.[34] Those various classifications of Big Data and Analytics related programs reflect different specializations (information management, business analytics, statistics and operational research, etc.) and concentration areas (marketing, insurance, financial services) of different universities.

---

[32] Source: McKinsey Global Institute. Big Data and Analytics: The next frontier for innovation, competition, and productivity, 2011. Available at: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.

[33] **See Appendix 3** for degree Awards in Big Data and Analytics related-academic programs.

[34] For example, the Big Data program at the University of Illinois is in the Department of Statistics, University of California-Berkley's is in the School of Engineering, and Columbia University's is in the Computer Science Department. Source: http://www.informationweek.com/big-data/big-data-analytics/big-data-analytics-masters-degrees-20-top-programs/d/d-id/1108042?page_number=22.

In the remainder of this section, Chmura presents current degree offerings in Big Data and Analytics related instructional programs in the Potomac region in the area of statistics, computer science, engineering, and business analytics. Not all of these degrees are specifically for Big Data and Analytics. This broader definition of Big Data and Analytics related programs is presented—rather than a specific Big Data and Analytics-focused program—for three reasons. First, because Big Data and Analytics is a new field and different schools are using different CIP codes, it is not easy to separate Big Data and Analytics-focused degree awards without a school-specific survey. Second, many programs are new, and their information were not captured by the National Center for Education Statistics (NCES). Third, and perhaps most important, since there is a large gap for Big Data and Analytics talent now, the graduates from Big Data and Analytics-focused degree programs will not be able to meet the business demand. Big Data and Analytics businesses in the region will recruit graduates with degrees in other related fields, such as management of information system or statistics, and develop them into Big Data and Analytics talents through additional training. Examining the broad array of Big Data and Analytics related programs allows regional businesses to evaluate the potential source of Big Data and Analytics talent that is educated in the region.

Computer Science

Regional institutions graduated 1,328 students with bachelor's or higher degrees in computer and information sciences, general (CIP code 11.0101), yet the regional demand for such degrees is estimated to be 3,042 per year. The degree awards only accounted for 44 percent of the business demand. Shortages for other degrees in computer science programs are even more severe. For example, regional institutions could only satisfy 2 percent of demand for degrees in computer system analysis/analyst (CIP code 11.0501), and 7 percent of demand in data modeling/warehousing and database administration (CIP code 11.0802).[35]

Engineering

For Big Data and Analytics related programs under the umbrella of engineering programs, severe shortages exist in computer engineering (CIP code 14.0901) and computer software engineering (CIP code 14.0903), where degree awards were about 10 percent of demand for those degrees. Graduates with mathematics and statistics degrees can also be attracted to the Big Data and Analytics careers, but additional training is needed. There are talent gaps for such degrees as well in the region, as degree awards accounted for 71–75 percent of the demand in recent years.

Business Analytics

Several regional universities have started Big Data and Analytics-focused programs in their business schools (e.g. the George Washington University). Those programs typically bear the name of Business Analytics, Data Analytics, or Data Science. Under the current CIP system, those programs are normally classified under management information systems or management science categories. Talent shortages exist for many such programs. For example, regional and state-wide institutions could only satisfy 14 percent of demand for degrees in management information systems, general (CIP code 52.1201), 13 percent of the demand for information resources management (CIP Code 52.1206), and 12 percent of the demand for management science (CIP code 52.1301).

[35] There are overlaps in total annual demand in those figures. For example, a computer programmer can have different degrees. As a result, the demand for such occupations is included in degree demand for multiple instructional programs.

Even though exact data for Big Data and Analytics-focused programs are not available, the above analysis of the Big Data and Analytics related programs in the Potomac region indicates that there is a large gap in Big Data and Analytics talent in the area based on demand from the current mix of industries in the region. Investing in Big Data and Analytics programs will help meet the business demand in the region. This is important if the region wants to maintain its position as a leader in the Big Data and Analytics market. Without that talent, businesses will have to recruit from higher education institutions in other areas—increasing recruitment costs. The lack of Big Data and Analytics talent may also affect the ability of businesses to expand their operations in the area.

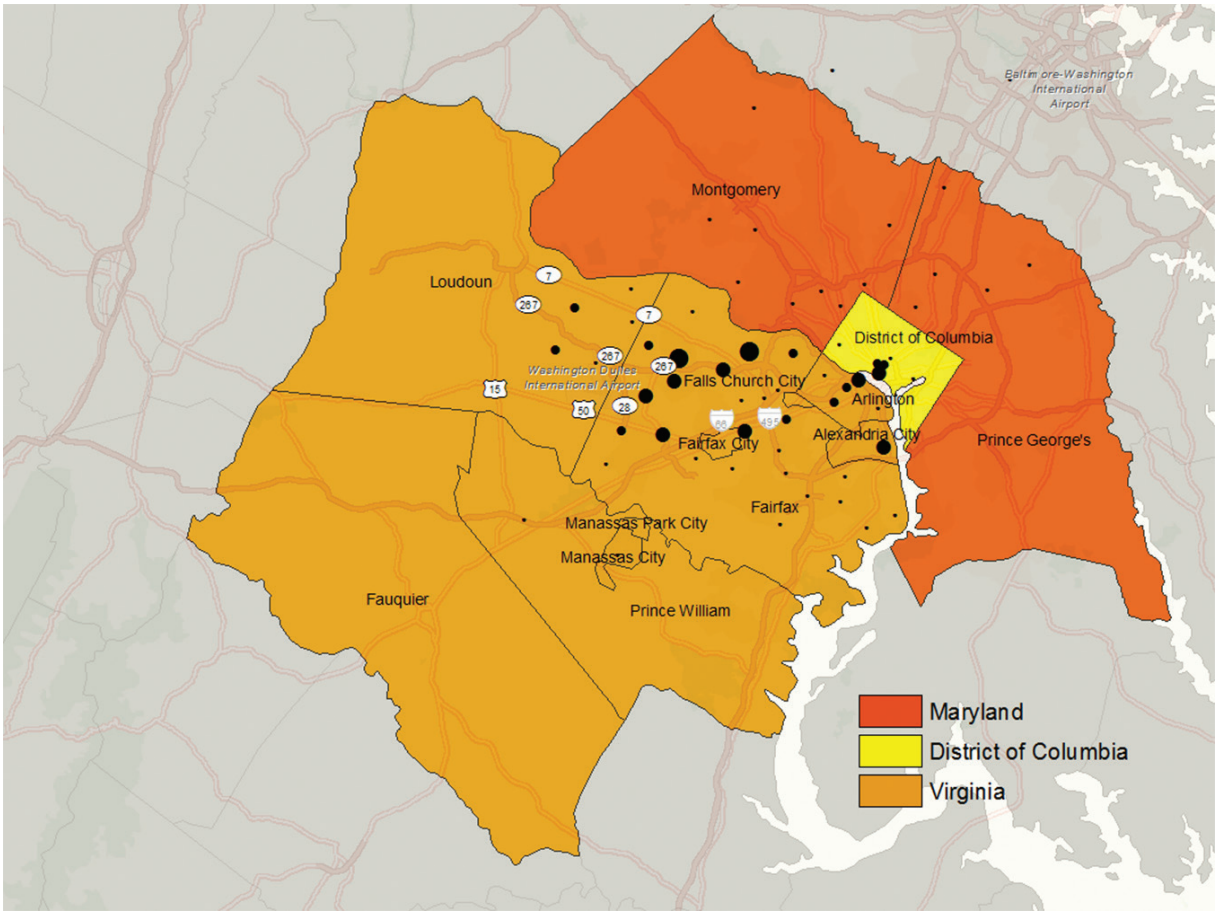
4. BIG DATA AND ANALYTICS ASSETS: NORTHERN VIRGINIA AND THE POTOMAC REGION

Location of Firms

In February 2014, Chmura surveyed businesses in the Potomac region to identify firms' characteristics, their current and expected Big Data and Analytics related workforce needs, and the role education and training is playing to fill those needs. A total of 270 organizations responded to the survey, with 162 identified as Big Data and Analytics firms.

The map in **Figure 4.1** shows the concentration of survey respondents associated with Big Data and Analytics by zip code. The size of the dot represents the number of responses received from that zip code. Seventy-two percent of all survey respondents in the Potomac region are located in Northern Virginia. It is not surprising that a large cluster of Big Data and Analytics firms are found along the Dulles Corridor, which is known as the "high-tech corridor" for its high-tech and innovative firms, as well as the nation's capital, where federal agencies are undoubtedly among the largest Big Data and Analytics clients in the world.

Figure 4.1:
Big Data and Analytics Respondents by Zip Codes

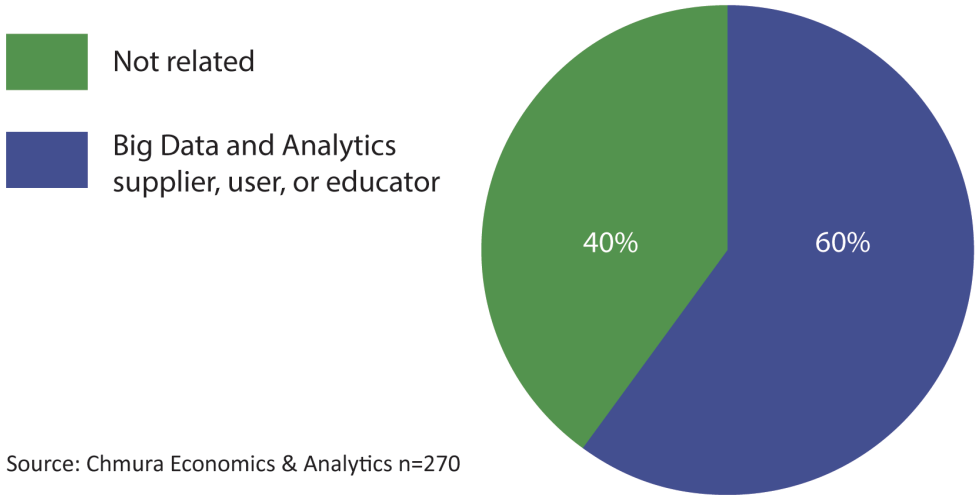


Source: Chmura Economics & Analytics

Relationship of Firms to Big Data and Analytics

As shown in **Figure 4.2**, 60 percent of the respondents in the Potomac region said they had some relation to Big Data and Analytics while 40 percent said they had no relation.

Figure 4.2: Relationship Between Respondents and Big Data and Analytics



Source: Chmura Economics & Analytics n=270

The majority of Big Data and Analytics firms responding to the survey identified themselves as providers rather than users of Big Data and Analytics. As shown in **Table 4.1** below, 63 percent of respondents described their organization as a provider of Big Data and Analytics related services, 28 percent as providers of Big Data and Analytics software, and 7 percent as providers of Big Data and Analytics hardware (note that respondents were asked to select all categories that applied). Forty-one percent described their organization as a user of Big Data and Analytics. Twenty-seven percent of respondents said they were both suppliers and users of Big Data and Analytics while 16 percent selected "none of the above." The most common reasons given for selecting "none of the above" were that the respondent worked in consulting or was not currently using Big Data and Analytics but was interested in using it in the future.

The distribution of categories in Northern Virginia is similar to the Potomac region. Among the Northern Virginia respondents, 68 percent supply Big Data and Analytics related services, 29 percent supply Big Data and Analytics related software, 8 percent supply Big Data and Analytics related hardware, and 35 percent are users of Big Data and Analytics. Twenty-five percent said they were both suppliers and users of Big Data and Analytics.

Table 4.1: Self-Categorization of Respondents

	Potomac Region		Northern Virginia	
Category	**Number**	**Percent**	**Number**	**Percent**
Services Suppliers	102	63%	79	68%
Software Suppliers	46	28%	34	29%
Hardware Suppliers	12	7%	9	8%
Users	66	41%	41	35%
University/Educator	10	6%	5	4%
None of the Above	26	16%	20	17%
Total	**262**		**188**	
Source: Chmura Economics & Analytics	n=162		n=117	

Importance of Big Data and Analytics to Respondent Firms

Among Big Data and Analytics providers and users in the Potomac region, the median firm started working with Big Data and Analytics five years ago, compared with four years ago in Northern Virginia. As shown in **Figure 4.3** below, when asked how important Big Data and Analytics was to their organization, 50 percent responded that it is "mission-critical," 47 percent stated that it is "somewhat important," and 3 percent reported that it is "not important." In Northern Virginia, a somewhat higher percentage (53 percent) responded that Big Data and Analytics is "mission-critical," 44 percent that it is "somewhat important," and 3 percent said it is "not important."

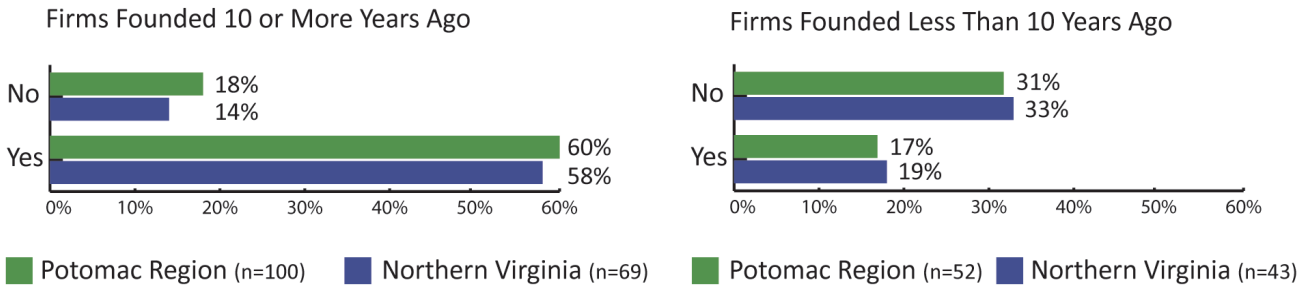
Figure 4.3: Importance of Big Data and Analytics to Respondent Firms

Source: Chmura Economics & Analytics

Survey results confirm that the use of Big Data and Analytics in the region is not limited to new startups. Only 22 percent of respondent firms in the Potomac region, and 21 percent in Northern Virginia report working with Big Data and Analytics since their founding, indicating that mature organizations in the region are seeking to leverage the latest hardware, software, and data skills to grow their businesses.

As indicated in **Figure 4.4**, 60 percent of respondent firms in the Potomac region that were founded more than 10 years ago have worked with Big Data and Analytics for more than five years (58 percent in Northern Virginia), implying that more mature firms are incorporating Big Data into their existing operations. Among firms founded less than 10 years ago, 31 percent in the Potomac region that work with Big Data and Analytics and 33 percent in Northern Virginia have worked with Big Data and Analytics since their founding.

Figure 4.4: Experience with Big Data & Analytics



Source: Chmura Economics & Analytics

Big Data and Analytics Employment and Firm Size

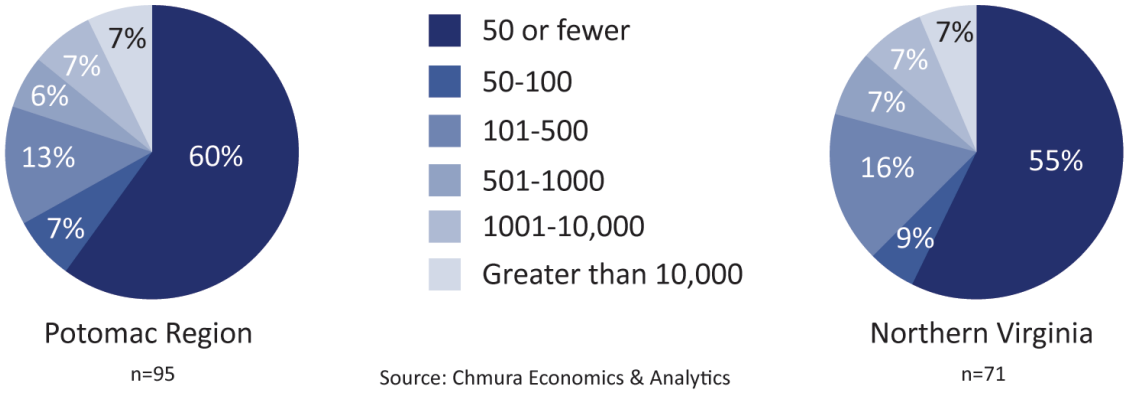
The respondent firms in the Potomac region that are involved in Big Data and Analytics said they employ 140,380 Big Data and Analytics workers, which accounts for an average 11 percent of their workforce.[36] Similarly, in Northern Virginia respondents said they employ a total of 123,458 Big Data and Analytics workers, which also accounts for an average 11 percent of their workforce.

Both large and small firms are engaged in the Big Data and Analytics marketplace. As shown in **Figure 4.5**, 60 percent of the Big Data and Analytics firms in the Potomac region employ 50 or fewer people compared with 55 percent in Northern Virginia. Although firms with greater than 10,000 employees make up only 7 percent of the Big Data and Analytics firms in the Potomac region, they employ 96 percent of the Big Data and Analytics employees. In Northern Virginia, 6 percent of the Big Data and Analytics firms have more than 10,000 employees and account for 94 percent of the Big Data and Analytics employment in the region.

The nation's "Data Capital" is a natural leader in Big Data and Analytics.

[36] Respondents did not specify whether they were providing global workforce numbers or only their workforce in the study area.

Figure 4.5: Size of Big Data and Analytics Firms, by Employment



Big Data and Analytics related employment accounts for a larger share of employment within small firms. As **Table 4.2** indicates, 63 percent of total employment in firms employing 10 or fewer individuals is related to Big Data and Analytics in the Potomac region, compared with 70 percent in Northern Virginia. The Potomac region's mix of large and small firms is otherwise similar to Northern Virginia. Smaller firms in the Big Data and Analytics marketplace are more focused on particular segments of the market. While lack of product and customer diversification implies more exposure to market risk, the potential for fast or breakthrough growth for those companies is also significant. Those small companies may also need more financial and technical support from regional business and community leaders.

Table 4.2: Big Data and Analytics Firms and Employment

Potomac Region				
Firm size	**Number of Firms**	**Total Employment**	**Big Data and Analytics Employment**	**% Big Data and Analytics Employment**
Less than or equal to 10	47	225	142	63%
11–50	44	1,103	488	44%
51–100	11	916	119	13%
101–500	20	4,298	1,305	30%
501–1,000	9	6,849	937	14%
1,001–10,000	11	37,849	3,321	9%
Greater than 10,000	10	1,201,500	134,138	11%
Total	**152**	**1,252,740**	**140,451**	**11%**

Northern Virginia				
Firm size	**Number of Firms**	**Total Employment**	**Big Data and Analytics Employment**	**% Big Data and Analytics Employment**
Less than or equal to 10	32	146	102	70%
11–50	29	677	288	42%
51–100	10	816	109	13%
101–500	18	4,073	1,263	31%
501–1,000	8	5,850	737	13%
1,001–10,000	8	27,549	1,821	7%
Greater than 10,000	7	1,123,500	119,138	11%
Total	**112**	**1,162,611**	**123,458**	**11%**

Source: Chmura Economics & Analytics

Big Data and Analytics Revenue

For those firms that provided information about their revenue from Big Data and Analytics, 21 percent reported that revenue from Big Data and Analytics accounts for more than half of their revenue, confirming that the Big Data and Analytics market is critical for them. Sixteen percent of firms reported that Big Data and Analytics revenue accounts for less than 10 percent of their current revenue, though many of them still consider Big Data and Analytics to be mission-critical and a source of future growth opportunities. Twelve percent are fully dependent on Big Data and Analytics products and services, reporting that 100 percent of their sales are directly related to Big Data and Analytics. Big Data and Analytics related revenue of respondent firms is summarized in **Table 4.3** below.

For those firms in Northern Virginia that provided information about revenue, 18 percent reported that revenue from Big Data and Analytics accounts for more than half of their revenue while 17 percent of firms reported that Big Data and Analytics accounts for less than 10 percent of their current revenue. A slightly higher percentage of firms in Northern Virginia are completely dependent on Big Data and Analytics revenue—13 percent reported that 100 percent of their total sales are directly related to Big Data and Analytics.

Overall, an average of 44 percent of respondent organizations' total sales is directly related to Big Data and Analytics products and services in the Potomac region—similar to the 45 percent in Northern Virginia. Perhaps reflecting the importance of the federal government as a customer, the majority of respondents' sales and purchases stay within the region. On average, 63 percent of total sales from Potomac region firms directly related to Big Data and Analytics are sold to customers in the Potomac region compared with 64 percent for Northern Virginia firms.

Big Data and Analytics Customers and Domain Expertise

Big Data and Analytics firms in the Potomac region tend to rely on other firms in the region for purchases of Big Data and Analytics products and services. As **Table 4.3** summarizes below, on average, 58 percent of respondent firms' total purchases are purchased from providers in the Potomac region compared to an average 52 percent in Northern Virginia.

An average of 44 percent of respondent organizations' total sales is directly related to Big Data and Analytics products and services indicating a strong and growing Big Data and Analytics ecosystem.

Those relatively higher percentages of intra-regional sales and purchases indicate that a Big Data and Analytics ecosystem is forming in the Potomac region.[37] This can be interpreted as a competitive advantage for the region, as it implies that regional Big Data and Analytics businesses offer a diverse set of products and services, and they can meet the majority of demand of other businesses in the region. This will be appealing for the region to recruit other Big Data and Analytics firms or start-ups, as those firms tend to locate in a region where they can easily acquire necessary products, services, and talent. On the other hand, the fact that more than 40 percent of products and services are purchased from outside the region also points to an opportunity for the region. Understanding what products and services are currently not available in the region points to potential targets for future business recruitment efforts.

[37] In economic development literature, the close-knit suppliers and purchasers are often called clusters.

Table 4.3: Big Data and Analytics Revenue

Potomac Region						
Big Data and Analytics Share of Total Revenue	**Number of Firms**	**Percentage of Firms**	**Total Employment**	**Percentage of Total Sales Directly Related to Big Data and Analytics** (Average)	**Percentage of Big Data and Analytics Sales that are Sold to Customers in Potomac Region** (Average)	**Percentage of Big Data and Analytics Purchases from Potomac Region (Average)**
Less than or equal to 10%	18	16%	626,479	5%	41%	38%
11-25%	12	11%	14,065	19%	67%	●
26-50%	9	8%	309	40%	54%	45%
51-75%	4	4%	341	71%	68%	0%
76-99%	5	5%	111	84%	89%	85%
100%	13	12%	1,056	100%	73%	●
Not Available	49	45%	582,403	●	83%	50%
Total	**110**	**100%**	**1,224,764**	**44%**	**63%**	**58%**
			n=110	n=61	n=110	n=110

Northern Virginia						
Big Data and Analytics Share of Total Revenue	**Number of Firms**	**Percentage of Firms**	**Total Employment**	**Percentage of Total Sales Directly Related to Big Data and Analytics** (Average)	**Percentage of Big Data and Analytics Sales that are Sold to Customers in Potomac Region** (Average)	**Percentage of Big Data and Analytics Purchases from Potomac Region (Average)**
Less than or equal to 10%	15	17%	608,169	5%	47%	50%
11-25%	5	6%	13,728	18%	58%	●
26-50%	8	9%	299	40%	59%	45%
51-75%	3	3%	331	73%	90%	●
76-99%	2	2%	42	85%	83%	80%
100%	11	13%	1,006	100%	76%	●
Not Available	42	49%	533,361	●	78%	25%
Total	**86**	**100%**	**1,156,936**	**45%**	**64%**	**52%**
Source: Chmura Economics & Analytics			n=86	n=44	n=86	n=86

* No respondents within the range of revenue share provided an answer to this question.

As shown in Table 4.4 below, Big Data and Analytics respondent firms report domain expertise in a wide range of fields, reflecting the cross-industry applicability of Big Data and Analytics and the difficulty of categorization. When asked to provide their company's top three areas of expertise, 54 percent of the firms in the Potomac region chose government followed by 49 percent with information technology, and 28 percent chose intelligence/national security. Rounding out the top five domains and are cyber security (21 percent) and healthcare (18 percent). Similarly, in Northern Virginia 54 percent of firms chose information technology, followed by 50 percent government, 28 percent intelligence/ national security, 20 percent cyber security, and 19 percent healthcare. For firms responding "other," unique areas of expertise listed included nonprofit, consulting, energy, advertising/marketing analytics, geographic information systems, technology, industrial engineering, and telecommunications.

Table 4.4: Big Data and Analytics Domain Expertise

	Potomac Region		Northern Virginia	
Domain Expertise	Number of Firms	% of Respondents	Number of Firms	% of Respondents
Government	82	54%	58	50%
Information Technology	75	49%	63	54%
Intelligence/ National Security	42	28%	33	28%
Cyber Security	32	21%	23	20%
Healthcare	28	18%	22	19%
Financial Services	18	12%	16	14%
Education	15	10%	9	8%
Networking Infrastructure	11	7%	10	9%
Logistics	10	7%	6	5%
Manufacturing	8	5%	5	4%
Law	7	5%	2	2%
Real Estate	5	3%	4	3%
Other (Represents 34 unique areas of expertise in Potomac region and 21 in Northern Virginia)				
Source: Chmura Economics & Analytics	n=152		n=117	

Respondent firms in Northern Virginia generate an estimated 716.6 terabytes of new data on average daily.

Northern Virginia's role as a leader in data generation, data storage, computing power, and innovation in Big Data and Analytics has impacted several industries in the region. Respondent firms in Northern Virginia generate an estimated 716.6 terabytes of new data on average daily, and up to 70 percent of the world's internet traffic passes through Northern Virginia on a daily basis.[38] With their proximity to the federal government and data centers, Northern Virginia firms have the unique subject matter expertise in areas such as government, intelligence/national security, and cyber security necessary to analyze Big Data.[39] Northern Virginia firms have expertise in a wide variety of sectors, hire top employees in their fields, and supply hardware, software, and services to diverse and innovative users.

[38] Source: Loudoun Virginia Economic Development, Business & Industry Stats, available at: http://www.biz.loudoun.gov/index.aspx?NID=93.

[39] Please see **Appendix 4** for a summary of domain expertise, employment and services of Big Data and Analytics firms in Northern Virginia.

It is not surprising that, outside of information technology, more regional firms have domain expertise in government and intelligence/national security. This is the biggest advantage for the Potomac region, compared with other high-tech regions such as Silicon Valley. The literature review in **Appendix 1** shows that government is one of the large potential areas for big data utilization, and Northern Virginia is poised to be the dominant player in this domain. It is interesting to note that for firms with expertise in government, only a small percentage of them are in hardware, while the majority is in Big Data and Analytics services. This seems to suggest that government utilization of Big Data and Analytics are in the stages of software and analytics development based on existing IT infrastructure.

It is not surprising that, outside of information technology, more regional firms have domain expertise in government and intelligence/national security. This is the biggest advantage for the Potomac region, compared with other high-tech regions such as Silicon Valley.

Healthcare and financial services have higher percentages of firms as Big Data and Analytics users than other industry domains. Both of them also have few firms specialized in hardware. Those are two industries that are keener on the "analytics" part of the Big Data and Analytics, or the "value" component of the 5Vs of Big Data. Those two industries also have high demand for workers in business intelligence or consultants. A number of regional universities have started business analytics degree programs, which can supply talent to those industries.

Data are flowing into the Potomac region at a rapid pace and, as seen in **Table 4.5**, organizations in the region are putting it to use for customers and their own internal projects. Eighty-three percent of respondents in the Potomac region report active or planned Big Data and Analytics projects. Sixty-one percent have active customer projects involving the processing or use of Big Data, and 41 percent have currently active internal projects (some companies reported both internal and customer projects). Ten percent have no active projects but planned customer projects, while 7 percent have no active projects but planned internal Big Data projects. Compared to the Potomac region overall, a higher percentage of respondents in Northern Virginia report active or planned projects across all categories.

Table 4.5: Firms With Active or Planned Big Data & Analytics Projects

	Potomac Region Firms	Percent	Northern Virginia Firms	Percent
One or more active internal projects	63	41%	47	42%
One or more active customer projects	93	61%	70	63%
No active internal projects, but one or more planned	10	7%	9	8%
No active customer projects, but one or more planned	15	10%	14	13%
No active or planned projects	26	17%	15	13%
Source:Chmura Economics & Analytics	n=152		n=112	

Amount and Types of Data Analyzed/Processed

As seen in **Tables 4.6** and **4.7** below, an estimated 1.03 petabytes (1 petabyte = 1015 bytes) of new data flow into respondent organizations in the Potomac region on an average day in the form of transactions, operational data, human-generated social data, etc.[40] On an average day, respondent organizations analyze/process and save an estimated 42.6 petabytes of data, including seven organizations that report saving between one and 10 petabytes.

An estimated 716.6 terabytes (1 terabyte = 1012 bytes) of new data flows into respondent organizations in Northern Virginia on an average day in the form of transactions, operational data, human-generated social data, etc. On an average day, respondent organizations analyze/process and save an estimated 25.2 petabytes of data, including four organizations that report saving between one and 10 petabytes (1 petabyte = 1015 bytes).

Table 4.6: Amount of Data Analyzed/Processed and Saved on an Average Day

	Percent of Respondents in the Potomac Region	Percent of Respondents in the Northern Virginia
Less than 1 terabyte	24%	28%
1 to less than 10 terabytes	16%	15%
10 to less than 50 terabytes	7%	7%
50 to less than 100 terabytes	4%	6%
100 to less than 250 terabytes	3%	3%
250 terabytes to less than 1 petabyte	3%	3%
1 to less than 10 petabytes	5%	4%
10 petabytes or more	0%	0%
Don't Know	37%	35%
Source: Chmura Economics & Analytics	n=147	n=109

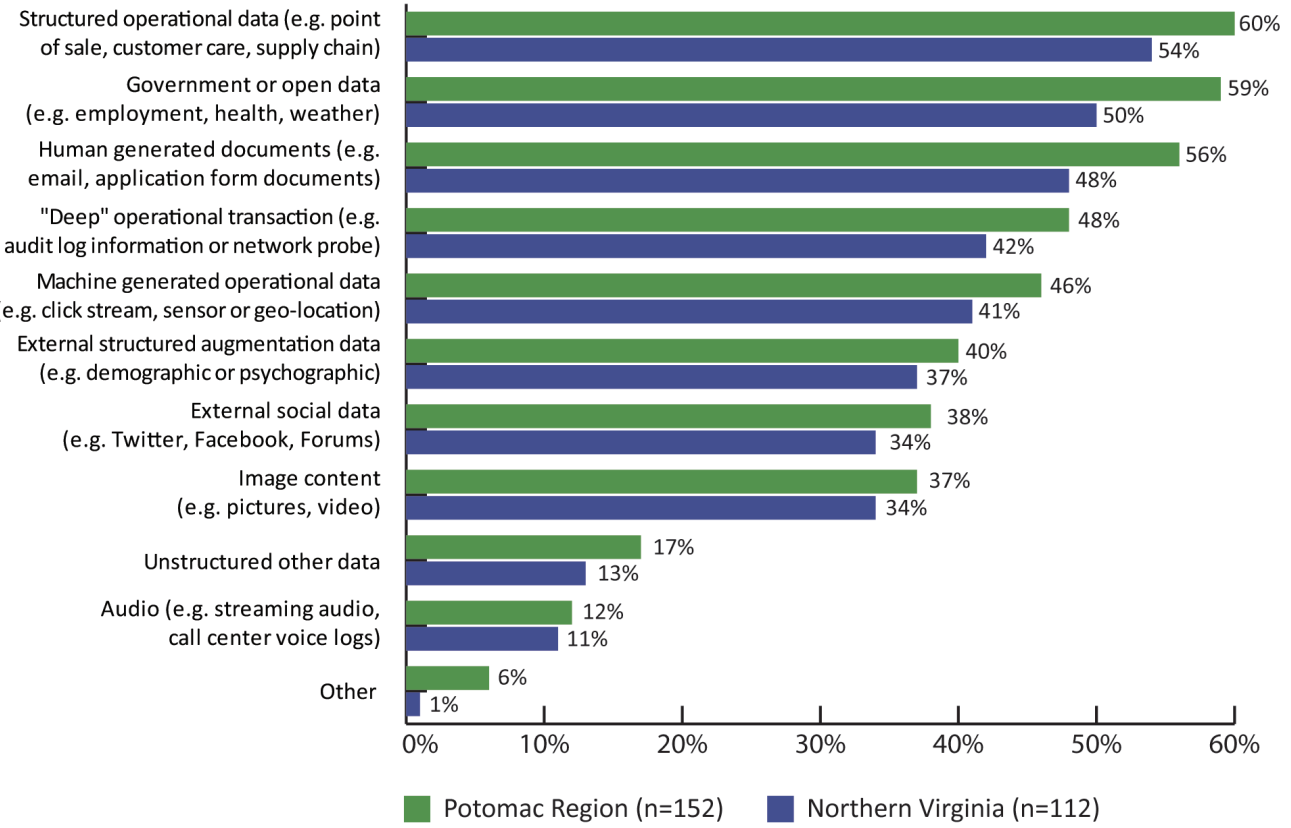
Table 4.7: Amount of New Data Flowing Into Firms on an Average Day

	Percent of Respondents in the Potomac Region	Percent of Respondents in the Northern Virginia
Less than 10 gigabytes	27%	28%
10 gigabytes to less than 100 gigabytes	12%	11%
100 gigabytes to less than 500 gigabytes	3%	4%
500 gigabytes to less than 1 terabyte	4%	6%
1 to less than 5 terabytes	5%	5%
5 to less than 10 terabytes	1%	2%
10 to less than 50 terabytes	2%	2%
50 to less than 100 terabytes	1%	1%
100 to less than 250 terabytes	1%	0%
250 terabytes or more	1%	2%
Don't Know	43%	41%
Source: Chmura Economics & Analytics	n=147	n=109

[40] Those figures represent data flow into surveyed firms, and do not include data flow into government agencies. So, the overall data flow is likely much larger than reported here.

Figure 4.6 below illustrates that the most common type of data used by organizations is structured operational data, or data that fits a defined model and relates to the operations of a company. The popularity of each dataset seems to correspond to the ease in which it can be used to glean meaningful business information.

Figure 4.6: Data Sources Currently Used by Respondent Organizations

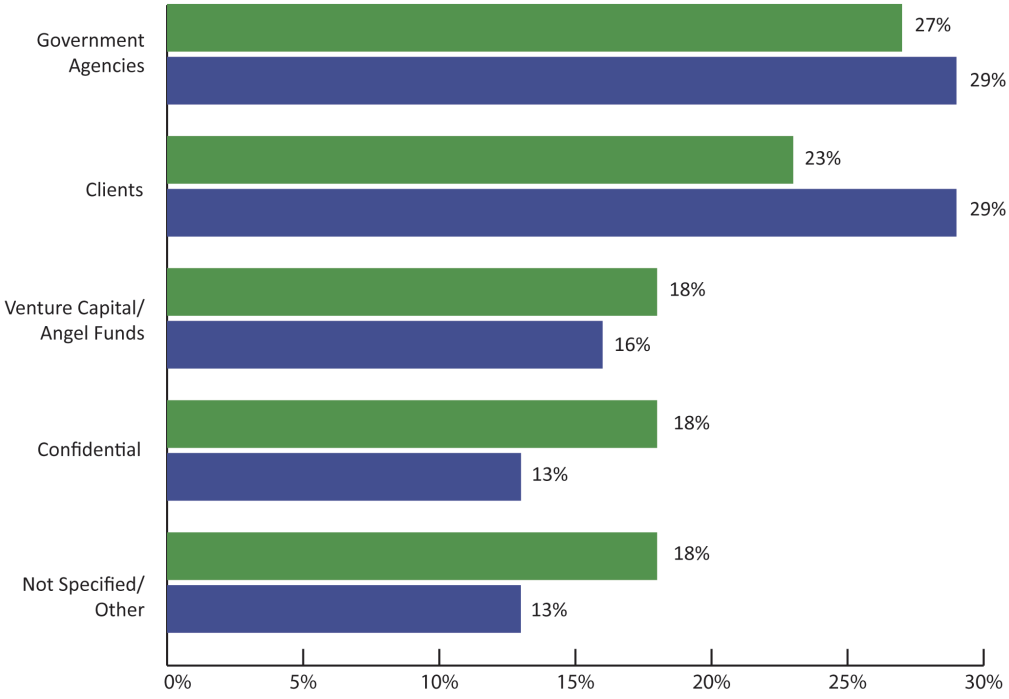


Source: Chmura Economics & Analytics

Funding for Big Data and Analytics

Sources of funding are important, particularly in small firms, to foster future innovation. It is not surprising that government agencies provide the largest percentage of firms with funding both in the Potomac region and Northern Virginia. In the Potomac region, 58 percent of respondents are engaged in the commercialization of Big Data and Analytics related products or services. Among them, 35 percent receive funding from outside sources—with federal agencies providing the largest percentage. As indicated in **Figure 4.7** below, the top three funding sources are government agencies, with 27 percent of Potomac region firms engaged in commercialization receiving funding from this source; clients (23 percent); and venture capital/angel funds (18 percent). The same is true in Northern Virginia—of those respondents that are engaged in the commercialization of Big Data and Analytics related products or services and received funding from outside sources, 29 percent received funding from government agencies, 29 percent from clients and 16 percent from venture/angel funders.

Figure 4.7: Funding Sources for Commercialization of Big Data and Analytics Products and Services



Source: Chmura Economic & Analytics

The majority of respondents in the Potomac region that receive funding are relatively small firms, with 32 percent of firms employing 10 or fewer and another 32 percent employing between 11 and 50, as shown in **Table 4.8** below. This concentration of funding among smaller firms reflects venture capital funds, angel investors, and government funds investing in promising Big Data and Analytics companies early in their growth, while larger firms are more likely to report clients and government funds as major sources of funding.

Table 4.8: Size of Firms Receiving Funding for Big Data and Analytics Related Products and Services

Firm Size	Potomac Region	Northern Virginia
Less than or equal to 10	32%	33%
11-50	32%	24%
51-100	0%	0%
101-500	19%	24%
501-1000	3%	5%
1001-10,000	3%	5%
Greater than 10,000	10%	10%
Source: Chmura Economics & Analytics	n=162	n=117

In addition, 74 percent of respondents in the Potomac region that obtain outside funding are federal contractors, and 58 percent have engaged in a Big Data and Analytics project with a government client. In Northern Virginia, 71 percent of respondents that receive outside funding for commercialization are federal contractors, and 62 percent have engaged in a Big Data and Analytics project with a government client.

Workforce Needs

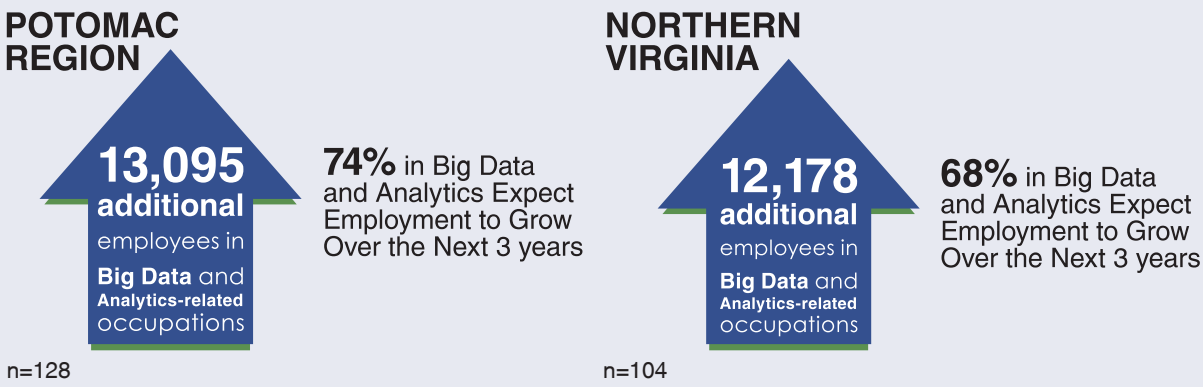
Respondent Big Data and Analytics organizations in Northern Virginia and the Potomac region are optimistic about their growth prospects over the next three years. Of the 128 respondents having some percentage of employees currently involved in Big Data and Analytics work in the Potomac region, 74 percent report that they expect Big Data and Analytics employment to grow. Twenty-five percent responded that they don't know how their organization's employment will change over the next three years, 10 percent expected no change, and only 1 percent (1 respondent) expected employment to decline. By comparison, in Northern Virginia, 68 percent of the 104 respondents with some percentage of employees currently involved in Big Data and Analytics work report that they expect employment to grow. Twenty-three percent responded that they don't know how their organization's employment will change over the next three years, 9 percent expected no change in employment, and none of the respondents expected employment to decline. Regional employers' growth expectations are summarized in **Figure 4.8** located on the next page.

In Northern Virginia, nearly 73 percent of new jobs over the next three years are expected to be in occupations working with Big Data and Analytics.

Organizations in the Potomac region expect to add over 20,500 jobs in total, with the median firm expecting 20 new employees over the next three years. Over half of those jobs (13,095) are expected to be in occupations working with Big Data and Analytics, while the median firm predicts adding 10 new employees working with Big Data and Analytics. In Northern Virginia, organizations expect to add over 16,700 jobs in total, with the median firm expecting 30 new employees over the next three years. Nearly 73 percent of those jobs (12,178) are expected to be in occupations working with Big Data and Analytics, while the median firm predicts adding 10 new employees working with Big Data and Analytics.

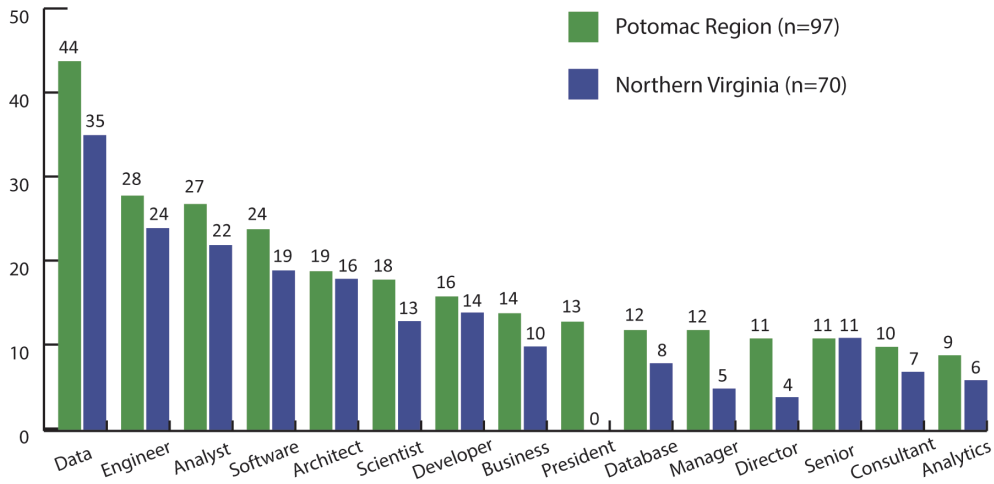
This level of employment growth translates into an average 3.0 percent growth rate over the next three years for Big Data and Analytics occupations in the Potomac region and 3.2 percent in Northern Virginia. These projected increases in Big Data employment in the region are significantly higher than forecasted employment growth for the nation overall. The Bureau of Labor Statistics forecasts national employment to grow an average 1.0 percent over the ten years from 2012 to 2022.[41]

Figure 4.8: Employer Growth Expectations



Amid widespread expectations of growth, not all occupations are equally in demand. When asked to list their top five occupations most involved in Big Data and Analytics related work in their organizations, 97 respondents in the Potomac region provided nearly 200 unique job titles while 70 respondents in Northern Virginia provided about the same amount of unique titles. The most common word appearing in the job titles was "data," followed by "engineer" and "analyst." Titles such as "president," "manager," "director," "senior," and "chief" also appeared frequently, indicating the importance of Big Data and Analytics in the hierarchy of respondent organizations. The 15 words that were used most frequently in job titles are summarized in **Figure 4.9**.

Figure 4.9: Frequency of Terms in Reported Big Data & Analytics Job Titles



[41] Source for BLS projections: http://www.bls.gov/news.release/ecopro.t03.htm.

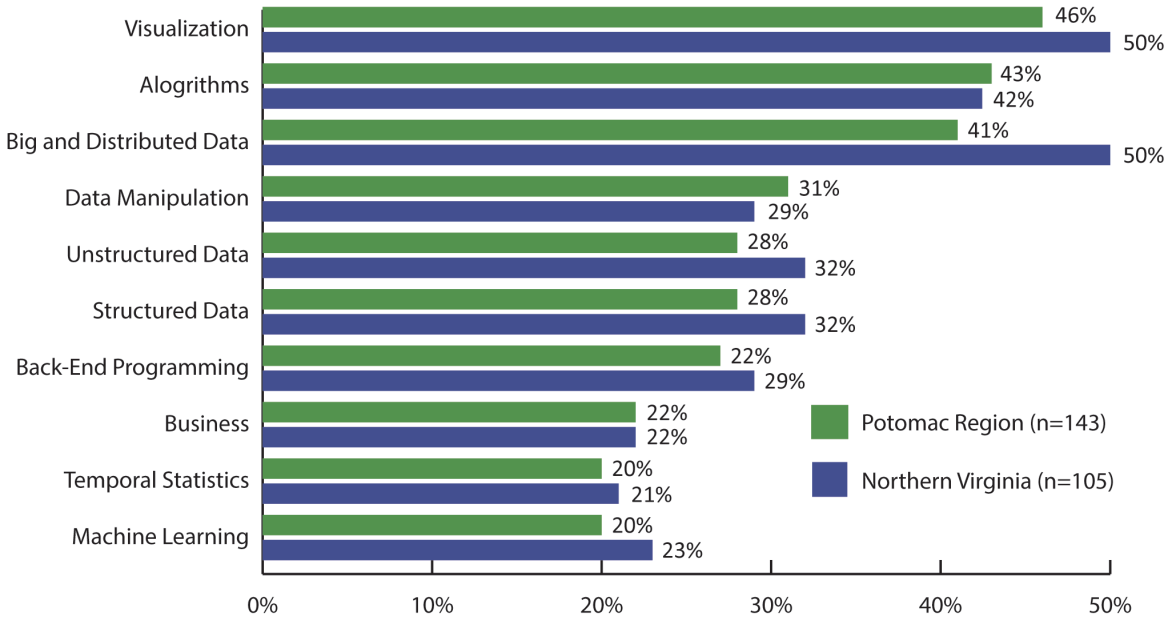
Respondents were also asked to estimate the number of individuals currently employed in their organization in each of the occupations they listed. **Table 4.9** details the 15 occupations with the largest number of employees reported by two or more Big Data and Analytics employers in the Potomac region and Northern Virginia. Some of the occupations with the largest numbers of employed individuals (including information management and analytics practitioner, financial analyst, and healthcare data analyst) were only provided by one large organization, and were excluded from the table in favor of occupations demanded by multiple employers.

Table 4.9: Job Titles with Largest Reported Number of Employees, Potomac Region

Job Title	Number of Employees
Business Analyst	2,254
Chief Technology Officer	1,706
Data Scientist	613
Software Developer	254
Engineering	200
Data Analyst	189
Analyst	52
Engineer	50
Operations	50
Software Engineer	27
Economist	20
Data Visualization	17
Business Intelligence Consultant	16
Software Development	13
ETL (Extract, Transform, Load) Developer	11
n= 97	
Source: Chmura Economics & Analytics	

When given a list of 23 knowledge areas and skills and asked to select up to five that they considered to be the most important for a career in Big Data and Analytics, nearly half of respondents chose visualization (46 percent) in the Potomac region and exactly half choose it in Northern Virginia as shown in **Figure 4.10** below. The second and third most important skills were juxtaposed in the two regions. Algorithms (43 percent) and big and distributed data (41 percent) were the next two most important skills in the Potomac region. In Northern Virginia, big and distributed data was identified as an important skill by 50 percent of the respondents followed by algorithms (42 percent).

Figure 4.10: Most Important Knowledge and Skills Chosen by Big Data and Analytics Businesses

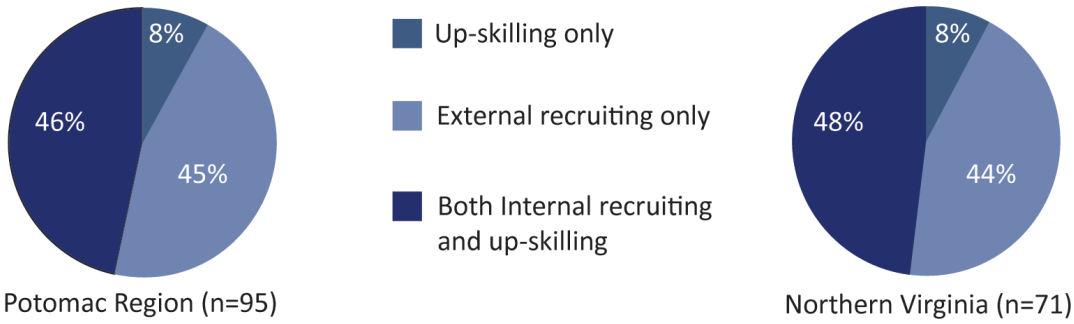


Source: Chmura Economics and Analytics

Education and Training

How do businesses plan on finding employees with the education and skills they need? **Figure 4.11** below illustrates the distribution of employer responses that are similar in the Potomac region and Northern Virginia. Only eight percent of respondents in both the Potomac region and Northern Virginia expect to rely solely on up-skilling current employees to increase the number of individuals in their organization working with Big Data and Analytics. Forty-five percent in the Potomac region expect to rely on external recruiting of new employees alone compared with 44 percent in Northern Virginia. In the Potomac region, 46 percent will use both external recruiting of new employees and up-skilling of current employees compared with 48 percent in Northern Virginia.

Figure 4.11: Respondents' Expectations for Filling Big Data and Analytics Positions



Source: Chmura Economics & Analytics

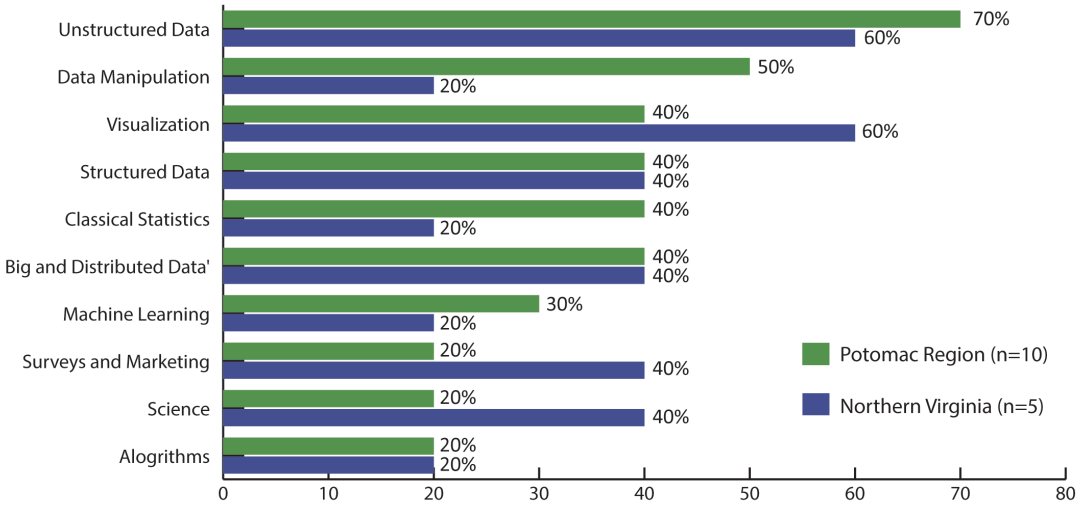
Sixty-six percent of respondent firms provide Big Data and Analytics training for employees, and some note using local Virginia and Maryland colleges and universities for their training. However, the majority has developed their own in-house or on-the-job training or work with local dedicated IT training firms and partners.

Recognizing the increasing need for training in this area, 90 percent of colleges, universities, and other educators who responded to the survey already offer or intend to offer courses or programs specifically targeting Big Data and Analytics within the next five years. Seventy percent intend to offer courses or programs within the next year. Currently, 40 percent of respondents are offering courses or programs, including minors, master's, and certificate programs in data science, as well as Hadoop, MongoDB, and Amazon Web Services courses. New programs expected to be added are primarily at the graduate level, including a master's of science program in data science and a graduate program in data analytics, though one university plans to offer comprehensive data management training within a year and another is planning on adding a Ph.D. in data science within two years.

As indicated in **Figure 4.12** below, educators in the Potomac region value many of the same skills businesses noted as important, though with a different emphasis on certain knowledge and skill sets. Visualization, algorithms, big and distributed data, data manipulation, unstructured data, structured data, and machine learning all made it into the top ten most important knowledge and skills from the educator perspective. While visualization was the most popular choice for Big Data and Analytics businesses, unstructured data ranked first for educators in the region.

Educational institutions and businesses are producing talent to address a surging demand.While visualization was the most popular choice for Big Data and Analytics businesses, unstructured data ranked first for educators in the region.

Figure 4.12: Most Important Knowledge and Skills Chosen by Educators



Source: Chmura Economics & Analytics

In addition to the courses and programs currently offered and expected in the next few years, 70 percent of respondent educators in the Potomac region and 80 percent of those in Northern Virginia are already engaged in Big Data and Analytics research. Funding sources include the National Science Foundation, Defense Advanced Research Projects Agency (DARPA), the Department of Defense (DoD), the National Institutes of Health (NIH), health systems, private foundations, and other external firms. In both the Potomac region and Northern Virginia, 40 percent of respondent firms are engaged in the commercialization of Big Data and Analytics

related products or services with outside businesses in the region, with industry partners in government, information technology, healthcare, education, intelligence/national security, cyber security, networking infrastructure, manufacturing, and internet commerce. Thirty-six percent of businesses in the Potomac region and 35 percent in Northern Virginia reported working with university research partners across the nation and around the world on a wide variety of initiatives, including: supporting collaboration on research to reduce health disparities internationally; advising university partners on teaching curriculums; partnering on interoperability of intelligent tutoring systems; government intelligence; funding post-doctoral research and hiring graduate students; and supplying general Big Data and Analytics software and expertise.

Local universities listed within the survey as research or commercialization partners or resources for training or recruiting skilled employees are summarized in **Table 4.10** below, along with programs and courses related to Big Data and Analytics as provided on the university's website.

Table 4.10: Local Universities Listed by Businesses in Northern Virginia and Big Data and Analytics Program Offerings

University	Big Data and Analytics Courses and Programs
American University	Business Analytics, Business Intelligence, and Database and Big Data courses
George Mason University	Data Analytics Graduate Certificate and Master of Science in Data Analytics Engineering
The George Washington University	Master of Science in Business Analytics
Georgetown University	Graduate Certificate in Data Analytics
University of Virginia	Master of Science in Data Science through Data Science Institute
Virginia Commonwealth University	Master of Science in Business with a concentration in Decision Analytics
Virginia Polytechnic Institute and State University	Has announced a new Center for Business Intelligence and Analytics
University of Maryland	Has announced a Master's program in Marketing Analytics starting in 2013 in Business School

Source: Chmura Economics & Analytics, University websites

APPENDIX 1: LITERATURE REVIEW OF TECHNICAL DEFINITION AND MARKET TRENDS

Big Data Definition

Chmura reviewed a comprehensive list of studies on Big Data and Analytics through academic journals, industry reports, and other internet sources.[42] Though some data analysis techniques and data collection/storage methods have existed for years, the phrase Big Data and Analytics has only recently emerged.

Since Big Data and Analytics is a relatively new and evolving phrase, there is no uniform definition; various stakeholders have provided diverse and sometimes contradictory definitions as shown in **Table A.1**. One of the first widely-quoted definitions of Big Data resulted from the Gartner report of 2001. Gartner proposed that Big Data be defined by three Vs: volume, velocity, and variety.[43] Gartner expanded its definition in 2012 to include veracity, representing requirements about trust and uncertainty pertaining to data and the outcome of data analysis.[44] In a 2012 report, IDC defined the 4th V as value—highlighting that Big Data applications need to bring incremental value to businesses.[45]

Gartner's 3V definition has been influential since 2001. Many other definitions generally follow three Vs with different emphasis. Some definitions focus on the volume or the size of data. For example, in a 2012 study, Intel linked Big Data to organizations "generating a median of 300 terabytes (TB) of data weekly."[46] A 2012 McKinsey study also focused on size, defining Big Data as "datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze."[47]

Markets and Users of Big Data

A study by TATA Consultancy found that telecommunication, travel, and finance spent the most on Big Data investment, while energy and utility industries expect the highest return from Big Data investments.[48]

Studies indicate that almost all corporations can benefit from Big Data and Analytics, as it can revolutionize business operations across the board. A Skytree survey of businesses found that of all business operations, even though marketing (54 percent) stood out as the biggest benefactor of Big Data, it was clear that all areas of an organization could benefit, including operations (49 percent), sales (39 percent), IT (44 percent), and finance (25 percent).[49]

---

[42] Key words searched included the following: Big Data definition, Big Data market, Big Data industry, and Big Data skills. Only literature germane to this study is included in this report.

[43] Source: Data Management: Controlling Data Volume, Velocity, and Variety. Gartner, 2001.

[44] Source: The Importance of big data: A Definition. Bu M.A. Beyer and D. Laney, Gartner, 2012.

[45] Source: Worldwide Big Data Technology and Services, 2012-2015 Forecast, IDC Market Analysis. Available at: http://download.microsoft.com/download/7/B/8/7B8AC938-2928-4B65-B1B3-0B523DDFCDC7/IDC%20Report-Worldwide%20Big%20Data%20Technology%20and%20Services.pdf.

[46] Source: The Big Data Conundrum: How to Define It. MIT Technology Review, October 2013. Available at: http://www.technologyreview.com/view/519851/the-big-data-conundrum-how-to-define-it/.

[47] Source: McKinsey Global Institute, 2011. Available at: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.

[48] Source: The Emerging Big Returns on Big Data, TATA Consultancy Services, 2013. Available at: http://www.tcs.com/big-data-study/Pages/default.aspx.

[49] Source: Big Data Analytics, 2013, by Skytree, 2013. Available at: http://www.skytree.net/resources/2013-big-data-analytics-report/.

Consistent with the Skytree survey, IBM released a study in 2012 concluding that Big Data represents an enormous opportunity for marketers.[50] Big Data can drive decisions by accurately delivering the right message to the right person at the right time for the right price. The travel industry has recognized the importance of Big Data and Analytics in transforming its industry. Amadeus IT Group concluded in its study that Big Data and Analytics provides significant benefits for travel companies by offering better decision support, new products and services, and better customer relationships.[51]

Table A.1: Big Data Definitions

Date of Study	Author	Characteristics	Definition
2001	Gartner	Volume, Velocity, Variety	The three Vs (volume, velocity, and variety) were later expanded as shown below
2012	Gartner	Veracity	Veracity addresses the questions of trust and uncertainty with regards to data and the outcome of data analysis
2012	IDC	Value	Highlights that applications need to bring incremental value to businesses
2012	Intel	Volume or Size	Generating a median of 300 terabytes (TB) of data weekly
2012	McKinsey	Volume or Size	Datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze
2007	MIKE2.0 [52]	Complexity or Variety	"…big refers to big complexity rather than big volume."
2012	Oracle	Traditional plus unstructured new data	The derivation of value from traditional relational databases driving business decision making, augmented with new sources of unstructured data
2013	National Institute of Standards and Technology	Technology pioneering new information	Big Data exceed the capacity or capability of current or conventional methods and systems, and enable novel approaches to frontier questions previously inaccessible or impractical using current or conventional methods
2013	Ward and Baker	Size, Complexity, and Technology	Big Data is a term describing the storage and analysis of large and/or complex data sets using a series of new techniques

Source: Chmura Economics & Analytics

Some definitions do not consider size as the critical factor. Instead, it is complexity that matters. In these cases, the definitions are consistent with the variety components of Gartner's 3Vs. For example, the Method for an Integrated Knowledge Environment (MIKE2.0) project claimed that, in Big Data, "big refers to big complexity rather than big volume."[53]

The idea of complexity and variety is incorporated into other definitions. An Oracle study contended that Big Data is the derivation of value from traditional relational databases, driving business decision making augmented with new sources of unstructured data.[54]

[50] Source: Moving up the digital marketing maturity with big data analysis. A Thought Leadership White Paper, IBM, 2012, Available at: http://www.ibmbigdatahub.com/sites/default/files/whitepapers_reports_file/Moving_up_digital_marketing_maturity-IMW14658USEN.pdf.

[51] Source: At the Big Data Crossroads: turning towards a smarter travel experience. By Thomas H. Davenport, 2013. Available at: http://www.amadeus.com/web/binaries/blobs/60/112/Amadeus_Big_Data.pdf.

[52] MIKE2.0 Methodology was originally created by BearingPoint in 2007. It was handed over to Mike2.0 Governance Association in 2009. Source: http://mike2.openmethodology.org/.

[53] Sources: http://mike2.openmethodology.org/, and http://mike2.openmethodology.org/wiki/Big_Data_Definition.

[54] Source: Undefined by Data: A Survey of Big Data Definitions, by Jonathan Stuart Ward and Adam Barker, School of Computer Science, University of St. Andrews, UK, 2013. Available at: http://arxiv.org/pdf/1309.5821v1.pdf.

Another dimension of the Big Data definition involves technology. Big Data is not only large and complex, but it requires innovative technology to analyze and process. In 2013, the National Institute of Standard and Technology (NIST) Big Data workgroup proposed the following definition of Big Data that emphasizes application of new technology:

> Big Data exceed the capacity or capability of current or conventional methods and systems, and enable novel approaches to frontier questions previously inaccessible or impractical using current or conventional methods.[55]

This definition points to the velocity components of Gartner's 3V definition. It suggests that data are "big" relative to current standards of computation. Big Data are too fast for the prevailing data management technology. In the current marketplace, a set of technologies that are frequently associated with Big Data includes NoSQL data-stores, analysis tools and methods including MapReduce, text mining, NLP, statistical programming, machine learning, and information visualization.[56]

In the paper "Undefined by Data: A Survey of Big Data Definition," Jonathan Ward and Adam Baker proposed a definition of Big Data that incorporates three critical components: size, complexity, and technology: "Big Data is a term describing the storage and analysis of large and/or complex data sets using a series of new techniques." This definition summarizes several key aspects of Big Data.

[55] Source: http://bigdatawg.nist.gov/home.php.
[56] Source: Undefined by Data: A Survey of Big Data Definitions, by Jonathan Stuart Ward and Adam Barker, School of Computer Science, University of St. Andrews, UK, 2013. Available at: http://arxiv.org/pdf/1309.5821v1.pdf.

APPENDIX 2: RESEARCH PROCESS

The first task of Phase 1 was to provide a definition of Big Data and Analytics. Some firms and research institutions, such as banking, insurance, manufacturing supply-chain management, and higher education, have used large-scale data analysis techniques and data collection/storage methods for years. Even so, Big Data and Analytics is a relatively new phrase that reflects the confluence of market needs for data-driven insights, as well as the technological developments that makes collecting and analyzing large amounts of data possible.

There is no standard definition for Big Data and Analytics. Unlike industry definitions that simply reflect a group of products and services—such as chemical manufacturing or medical services—Big Data and Analytics refers to technologies and processes that can be applied in a wide range of industries, including healthcare, transportation, national security, and scientific research.

To provide an understanding of Big Data and Analytics, a review of literature about Big Data and Analytics definitions and market trends is provided in Section 3 along with outcomes from two focus group events.[57] The focus group discussions centered on how to define Big Data and Analytics, the markets they serve, and the education and skills that are needed to support the markets. Based on the literature review and input from the focus groups, Chmura and NVTC concluded that a survey-based definition at the firm level was the best approach to identify Big Data and Analytics in the region.

When an industry or group of industries are growing rapidly, it is not unusual for a shortage of workers to occur until educational institutions and training organizations build the capacity to teach more individuals, and more people are attracted to the needed occupations. Section 2 summarizes workforce issues related to Big Data and Analytics, including knowledge and skills needed by Big Data and Analytics workers, as well as education and training offered by the area's academic institutions. Even though exact data for Big Data and Analytics-focused programs are not available, the analysis of the Big Data and Analytics related programs in the Potomac region indicates that there is a large gap in Big Data and Analytics talent in the area based on demand from the current mix of industries in the region. Investing in Big Data and Analytics programs will help meet the business demand in the region. This is important if the region wants to maintain its position as the leader in the Big Data and Analytics market.

Phase 2 of the project involved primary data collection to identify firms that engage in Big Data and Analytics markets—either as a supplier of Big Data and Analytics related hardware, software, and services, or as Big Data and Analytics users—including businesses in finance, healthcare, and manufacturing. With input from NVTC, Chmura designed a firm-level survey and distributed it to 727 NVTC member organizations. It was supplemented with additional contacts identified at venture-backed Big Data companies[58] as well as federal contractors within the study region that were registered in the Central Contractor Registration (CCR)

[57] Chmura led two focus group discussions with attendees from the region's Big Data and Analytics footprint as well as a university expert on December 13, 2013.

[58] Source: CB Insights, available at http://www.cbinsights.com/datastore/product/list-big-data-companies/.

system.[59] A total of 9,722 surveys were sent via email.[60] The survey identified the characteristics of Big Data and Analytics firms in the Potomac region, their current and expected workforce needs, and the role education and training is playing to fill those needs.

Phase 3 of the project focused on data analysis and the results of the survey. Phase 3 findings are presented in Section 4 of this report, which identifies the overall number of Big Data and Analytics firms and employment in both the Potomac region and Northern Virginia. It also analyzes both the domain expertise and workforce needs of Big Data and Analytics firms in the region.

[59] The CCR database provides contact information for firms that obtain federal government contracts. It was used as a source of company names because it was postulated that some federal contractors in the region that did not belong to the NVTC were Big Data and Analytics companies.
[60] A copy of the survey is available upon request.

APPENDIX 3: DEGREE AWARDS IN BIG DATA RELATED-ACADEMIC PROGRAMS

Below, Chmura presents current degree offerings in Big Data and Analytics related instructional programs in the Potomac region, plus programs from statewide universities such as University of Virginia, Virginia Tech, and Virginia Commonwealth University. The programs included are in the area of statistics, computer science, engineering, and business analytics. Not all of these degrees are specifically for Big Data and Analytics. The table below presents Big Data and Analytics related degree awards (2-year, bachelor's, and postgraduate degrees) in the 2011–2012 academic year in the Potomac region. The table also identifies the annual regional demand (based on net growth and replacement, which reflects retirements or individuals moving to a new occupation) for those degrees based on Chmura's proprietary JobsEQ®, as well as the current employment in the region. The difference between total awards and total annual demand can be used as a measure of the gap between the supply (awards) and demand for Big Data and Analytics related talents.

Big Data and Analytics Related Degree Awards and Demand

		Awards				Related Occupations			
CIP Code	Instructional Programs	Certificates and 2-year Awards	Bachelor's	Postgraduate Awards	Total Awards	Annual Replacement Demand	Annual Growth Demand	Total Annual Demand	Current Employment
Computer Science Programs									
11.0101	Computer and Information Sciences, General	310	776	242	**1,328**	1,186	1,856	3,042	70,230
11.0103	Information Technology	411	334	167	**912**	1,425	3,231	4,656	92,995
11.0199	Computer and Information Sciences, Other	0	3	0	**3**	22	25	47	1,359
11.0201	Computer Programming/ Programmer, General	54	1	3	**58**	1,339	2,233	3,572	75,033
11.0202	Computer Programming, Specific Applications	9	0	0	**9**	1,121	1,962	3,083	63,311
11.0203	Computer Programming, Vendor/ Product Certification	6	0	0	**6**	457	279	736	15,833
11.0301	Data Processing and Data Processing Technology/Technician	8	0	0	**8**	228	58	286	12,448
11.0401	Information Science/Studies	45	1,217	822	**2,084**	639	1,077	1,716	42,141
11.0501	Computer Systems Analysis/ Analyst	0	28	7	**35**	611	1,134	1,746	35,518
11.0701	Computer Science	82	87	106	**275**	1,779	2,846	4,626	102,486
11.0801	Web Page, Digital/Multimedia and Information Resources Design	59	11	0	**70**	344	261	604	10,816
11.0802	Data Modeling/Warehousing and Database Administration	9	5	0	**14**	86	108	194	4,280
11.0803	Computer Graphics	2	35	0	**37**	644	362	1,006	20,897
11.0901	Computer Systems Networking and Telecommunications	280	46	106	**432**	658	1,296	1,954	38,197
11.1001	Network and System Administration/Administrator	2	2	0	**4**	661	1,068	1,729	40,508
11.1002	System, Networking, and LAN/WAN Management/Manager	0	12	0	**12**	161	281	441	9,551
11.1003	Computer and Information Systems Security/Information Assurance	54	366	37	**457**	748	1,175	1,923	44,787
11.1005	Information Technology Project Management	0	3	39	**42**	1,075	674	1,749	38,057
11.1006	Computer Support Specialist	125	0	0	**125**	430	725	1,155	25,642
11.1099	Computer/Information Technology Services Administration and Management, Other	0	19	0	**19**	0	0	0	0
11.9999	Computer and Information Sciences and Support Services, Other	3	51	90	**144**	0	0	0	0

APPENDIX 3:
DEGREE AWARDS IN BIG DATA RELATED ACADEMIC PROGRAMS

Big Data and Analytics Related Degree Awards and Demand

		Awards				Related Occupations			
CIP Code	Instructional Programs	Certificates and 2-year Awards	Bachelor's	Postgraduate Awards	Total Awards	Annual Replacement Demand	Annual Growth Demand	Total Annual Demand	Current Employment
Engineering Related Programs									
14.0901	Computer and Information Sciences, General	0	179	95	**274**	975	2,029	3,004	61,978
14.0903	Information Technology	0	22	227	**249**	797	1,777	2,574	52,841
14.1001	Computer and Information Sciences, Other	0	394	315	**709**	407	264	671	16,919
Mathematics & Statistics Programs									
27.0101	Mathematics, General	0	521	105	**626**	596	282	878	22,127
27.0301	Applied Mathematics, General	0	1	36	**37**	129	92	221	5,384
27.0501	Statistics, General	0	55	121	**176**	133	102	235	5,940
27.0502	Mathematical Statistics and Probability	0	0	3	**3**	69	69	137	2,681
27.0503	Mathematics and Statistics	0	0	31	**31**	123	89	212	5,233
27.0599	Statistics, Other	0	11	0	**11**	123	89	212	5,233
27.9999	Mathematics and Statistics, Other	0	7	0	**7**	90	57	147	4,191
Business Analytics Related Programs									
52.1201	Management Information Systems, General	46	37	92	**175**	643	616	1,259	28,008
52.1206	Information Resources Management	2	64	0	**66**	186	337	522	12,175
52.1299	Management Information Systems and Services, Other	14	0	0	**14**	0	0	0	0
52.1301	Management Science	7	254	1	**262**	1,102	1,024	2,127	54,123
52.1399	Management Sciences and Quantitative Methods, Other	0	130	0	**130**	0	0	0	0
	Total	1,528	4,671	2,645	**8,844**				

Source: Jobs EQ

APPENDIX 4: BIG DATA AND ANALYTICS DOMAIN EXPERTISE IN NORTHERN VIRGINIA

Domain Expertise in Northern Virginia

	Number of Respondent Firms	Big Data and Analytics Employment	Examples of Reported Job Titles	Estimated Data Stored (petabytes)	Hardware / Software / Services / User
Information Technology	63	15,109	• Information Management & Analytics Practitioner • Data Scientist • Software Developer • Senior Systems Administrator	**23.4**	6% 22% 48% 24%
Government	58	3,411	• Business Intelligence Expert • Mathematician/ Statistician • Domain Expert • Analyst	**12.5**	2% 17% 55% 25%
Intelligence/ National Security	33	3,319	• Systems Engineer • Systems Analyst • Government Analyst • Database Administrator • Data Visualization	**1.2**	10% 24% 46% 20%
CyberSecurity	23	9,143	• Cyber Security Architect • Analytics Specialist • Senior Systems Integrator • Data Scientist	**11.2**	9% 23% 46% 23%
Healthcare	22	4,389	• Healthcare Data Analyst • Domain Expert • Software Developer • Business Intelligence Consultant • Data Architect	**11.8**	3% 17% 51% 29%
Financial Services	16	4,440	• Chief Technology Officer • Senior Consultant • Business Intelligence • ETL Consultant	**5.6**	0% 10% 62% 29%
Networking Infrastructure	10	3,203	• Big Data Infrastructure Engineer • Network Engineer • Database Administrator • Database Engineer	**5.7**	8% 23% 46% 23%

Source: Chmura Economics & Analytics. Note that respondents were asked to select up to three areas of domain expertise and total number of firms, employment, and data stored are counted in full under each selected domain. n=86

Northern Virginia Technology Council

2214 Rock Hill Road
Herndon, Virginia 20170

703.904.7878 (phone)
703.904.8009 (fax)

info@nvtc.org
www.nvtc.org